UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

Dottorato di Ricerca in Biologia Applicata XXII ciclo

Curriculum in Microbiologia

Structural and biochemical studies of MCM proteins

Tutor: Prof. Maurilio De Felice

Supervisor: Dr. Silvia Onesti

Dottorando:

Barbara Medagli

2007-2011

Abstract

The eukaryotic MCM2-7 replicative helicase is a hetero-complex composed by 6 homologues belonging to the AAA+ ATPase superfamily. These proteins form a ring-like structure which is essential for the initiation and the progression of the replication fork. In order to better understand the cellular role and mechanism of action of these proteins, I carried out structural and functional studies on two different systems: the archaeal MCM proteins, that forms a homomeric ring representing a simplified model of the eukaryotic complex, and the less well studied human MCM8, whose exact role in eukaryotic cells has not been elucidated.

MCM interacts in two ways with the DNA: the canonical "loaded" mode where the MCM protein is encircling the DNA, and a "associated" mode where the DNA is wrapped around the enzyme. Using the protein from the archaeon *Methanothermobacter thermautotrophicum,* I biochemically shown that binding of MCM proteins affect the degree of supercoiling of dsDNA and that subdomain A is crucial for this interaction. To determine the structure of the C-terminal domain of the MCM proteins, the less studied region, I produced three different constructs corresponding to the C-terminus of the MCM proteins from *M. thermoautotrophicum* and *Sulfolobus solfataricus*. Despite a large effort using “high throughput” crystallisation techniques, no well diffracting crystals could be reproducibly obtained. An alternative structural biology approach has been adopted, using NMR analysis. Preliminary data are compatible with the bioinformatics prediction of a winged helix domain, but further work on isotope-labelled proteins is necessary for a full 3D characterisation.

Two other MCM proteins have been more recently described (MCM8 and MCM9): they are present in a variety of eukaryotic organisms, do not interact with the MCM2-7 complex and their function in the cell is still unknown. These protein are co-evolutionary related and it is conceivable that they have a functional link. I cloned and expressed in *E. coli* cells a large number of hMCM8 fragments: two of them (the N-terminus and the AAA+ domain) can be successfully expressed in soluble form

and have been purified and used for biochemical assays. Both proteins forms dimers, bind ssDNA and the AAA+ catalytic domain displays ATPase activity. Further effort will be done to optimise the purification protocol to obtain proteins suitable for structural studies. In order to assess the hypothesis of a functional relationship between MCM8 and MCM9 proteins, a similarly extensive cloning strategy was applied to human MCM9, with the aim of obtaining constructs expressing a variety of fragments, to be co-expressed with the equivalent hMCM8 domains, and to test the formation of functional complexes.

Ringraziamenti

Ringrazio il mio tutor il Prof. Maurilio De Felice per la pazienza dimostrata e per il vivace scambio di emails avuto in questi anni.

Grazie al mio Supervisor la Dr. Silvia Onesti per la possibilità di lavoro che mi ha dato, per l'incoraggiamento costante nell'andare avanti, per la sua disponibilità anche al di fuori del laboratorio (continuo a sostenere che è il miglior capo che si possa avere). Grazie per avermi dato la possibilità di vivere una delle più belle esperienze della mia vita: vivere a Londra....Grazie Silvia ho finalmente "imparato" l'inglese.

Grazie a tutte le persone che ho incontrato in questi anni....

Ringrazio tutti i membri del Biophysics group, per le pinte il venerdì sera, per le lunghe pause caffè e per aver interpretato il mio inglese sgangherato. Grazie a Marta per le lunghe skypate fino a tardi quando sono tornata.

Grazie anche ai "biologi famosi", anche se lontani, sempre in contatto quando possibile ed un grazie anche al mio vecchio laboratorio al CNR per le ospitate di tanto in tanto durante i miei trasferimenti o per gli esperimenti "proibiti".

Grazie anche alle persone che ho trovato qui a Trieste e mi hanno aiutato ad ambientarmi: Paola, Cristina, Ivet...grazie per avermi aspettato per i caffè al mattino e per le pause merenda al pomeriggio...oltre che alla consulenza scientifica ovvio = P

Ivet...grazie per aver sopportato questo terremoto in laboratorio e per le lunghe chiacchierate vicino alla lavastoviglie, per le "ore piccole" in laboratorio fatte assieme, per i passaggi offerti e per gli inoculi di yogurt =).

Grazie alla mia famiglia e a Marco per tutto il supporto datomi.... e nel caso di Marco anche per il supporto tecnologico, informatico, morale, per avermi scarrozzato in giro e per avermi seguito in questa avventura.

Con tutto il mio affetto...GRAZIE

Contents

Abbreviations

AAA+	ATPases associated with various cellular activities
AEBSF	4-(2-aminoethyl)benzenesulfonylfluoride
BSA	bovine serum albumin
Cdc	Cell-division cycle
CDK	Cyclin-dependent kinase
CV	column volume
DDK	Dpb4-dependent kinase
dsDNA	double-stranded DNA
DUE	Duplex Unwinding Element
EM	Electron Microscopy
h2i	Helix 2 insertion
HEPES	N-(2-hydroxyethyl)piperazine-N´-(2-ethanosulphonic acid)
HTH	Helix-turn-helix
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria Bertani broth
MCM	Minichromosome-maintenance
Mka	*Methanopyrus kandleri*
Mth	*Methanothermobacter thermautotrophicus*
MW	Molecular weight
PCR	polymerase chain reaction
pI	Isoelectric point
Pre-IC	pre-initiation complex
Pre-LC	pre-loading complex
Pre-RC	pre-replication complex
PS1BH	Pre-sensor 1 beta hairpin
PVDF	polyvinylidene difluoride
RPC	Replisome progression complex
SDS	Sodium-dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
ssDNA	single-stranded DNA
Sso	*Sulfolobus solfataricus*
TB	Terrific broth
TEV	Tobacco Etch Virus
Tris	Tris(hydroxymethyl)aminomethane
WH	Winged helix-turn-helix
wt	Wild type
βME	2-mercaptoethanol

ΔC	Domain-C deletion mutant
ΔsA	Subdomain-A deletion mutant
ΔsB	Subdomain-B deletion mutant

1 Introduction.

The replication of the genome must to be an exact process and requires a large number of proteins. Due to the large size of eukaryotic genomes, DNA replication starts at multiple sites along the chromosomes (called “origins of replication”). The entire process is tightly regulated to allow each origin to be used once for cycle and each fragment of the chromosome to be replicated only once. Errors in the control leads to disastrous consequence and can produce a large number of human genetic diseases, developmental abnormalities and cancer. The initiation step of the DNA replications is regulated by cell-cycle dependent kinases and a key target of these enzymes is the MCM2-7 complex (mini-chromosome maintenance). This protein assembly shares some functional and structural similarities with bacterial and viral helicases, as well as other initiation factors found in the three domain of life.

My research project focuses on the biochemical and structural analysis of the eukaryotic MCM proteins and its simpler archaeal homologues.

The MCM proteins are present only in proliferating cells and are highly expressed in malignant human cancer cells and pre-cancerous cells undergoing malignant transformation. Since they are not expressed in differentiated somatic cells that have been withdrawn from the cell cycle, MCM proteins are ideal diagnostic markers for cancer and possibly targets for anti-cancer drug development (Blow and Hodgson, 2002).

Here an overview of the DNA replication in bacteria, archaea and eukarya will be presented, followed by a summary of the recent functional and biochemical characterization of the eukaryotic MCM protein and their archaeal orthologues.

1.1 The DNA replication begins.

In all organisms DNA replication starts at particular genomic regions called "origins", that are recognized by protein initiator factors. The origins are defined by specific DNA sequences in prokaryotes, usually A-T rich zones prone to melting

called Duplex Unwinding Elements (DUE), or by secondary structure like in most of the eukaryotes.

The protein factors that recognize the origin are also involved in the recruiting of other enzymes responsible for the firing of the replication fork and its progression. Most of the proteins that participate to this process are AAA+ ATPases (ATPases associated with various cellular activities), a family of enzymes which use the energy produced by ATP hydrolysis to perform mechanical work. The AAA+ proteins tend to form oligomers (often hexamers) with the nucleotide binding site at the interface of two adjacent subunits, so that elements from both subunits are involved in the ATP binding and hydrolysis process.

1.1.1 In Bacteria

The bacterial genome consists of a single circular chromosome with a single replication origin (oriC). An origin recognition protein (DnaA) recognizes consensus sequences repeats (DnaA boxes) in the origin, whose copy number varies from the organism to organism. The DnaA box/protein interaction has a 1:1 stoichiometry and the simultaneous recognition of adjacent DnaA boxes allows the protein oligomerization. DnaA has been shown to bend the DNA (Funnell *et al*., 1987; Schaper *et al*., 2000) and, under precise biochemical conditions, to mediate the melting of the DUE region (Speck and Messer, 2001). This model was based on direct topology footprinting evidence (Mizushima *et al*., 1996), and was confirmed by other biochemical and structural studies, suggesting that the ATP-bound DnaA stabilises positive supercoils (Erzberger *et al*., 2006). After DNA melting DnaB, the replicative helicase, is recruited at the origin by DnaC, the helicase loader. The ATP hydrolysis of DnaC after the correct loading of the helicase on the single stranded DNA (ssDNA), triggers the loader release and DnaB starts her unwinding work in a 5'→3' direction.

1.1.2 In Eukarya

More complex is the picture in the Eukarya; due to their large genomes, eukaryotic replication must begin simultaneously at multiple origins in order to be promptly performed. As an extreme example, during the early cleavage divisions of the *Xenopus laevis* embryos ~300,000 origins are activated ("fired") which are ~10 kb from each other (Blow, 2001).

The budding yeast *Saccharomyces cerevisiae* possesses well defined origins of replication, known as Autonomous Replicating Sequences, or ARS, bearing a conserved primary structure. This represents an exception, as in many cases the eukaryotic origins are without identifiable specific sequence (Sclafani and Holzen, 2007), but seem to be defined by topology and dynamic chromatin structures (Rampakakis, 2009). Origins can be activated at different times on the S-phase of the cell cycle (early/late origins), in a CDK (Cyclin-dependent kinase) dependent and DDK (Dbf4-dependent kinase) dependent manner.

In Eukarya, the origin are labeled by the Origin Recognition Complex (ORC), a protein complex composed by six different subunits (Orc1-6). Most of the ORC proteins belong to the AAA+ family (except for Orc6). In budding yeast ORC is bound to the origin throughout the cell cycle in a ATP-dependent manner, whereas in other organisms the binding is regulated by CDK. In the early G1 phase the Cdc6 initiation factor (a AAA+ protein that share some similarity with Orc1) lands onto the ORC platform and is required for helicase loading. It has been proposed that a complex including Cdt1 and MCM2-7 is loaded onto ORC-Cdc6 complex during the initiation, forming the pre-replication complex (pre-RC) (Remus *et al*. 2009). At this point the origin is "licensed" and is ready to "fire" in the S-phase. After the dissociation of Cdc6 and Cdt1, ATP hydrolysis by ORC completes the MCM2-7 loading. Sequential hydrolysis events, catalysed in turn by Cdc6 and ORC are required for binding and loading of multiple MCM2-7 complexes on each origin (Randell *et al*., 2006).

Once the pre-RC is formed, a number of other factors, like for example Sld3, Cdc45, MCM10 and all the proteins that compose the pre-loading complex (Sld2, Dpb11, GINS and Polε) are recruited onto the origin to form the pre-Initiation Complex (pre-IC); S-CDK (S phase CDK) and DDK act locally at the origins during the S phase and play a major role in modulating the formation of the complex. First, MCM10 (this proteins is not homologue of the MCM2-7 family) is recruited, leading to the stimulation of DDK activity (Lee *et al*., 2003), which phosphorylates the MCM2-7 complex (Sclafani *et al*., 2002; Zou and Stillman, 2000). This post-translational modification event is believed to represent one of the major switches which initiates DNA replication. In yeast, a mutant of MCM5 (MCM5 *bob1* mutant) which is able to bypass this event and promoting premature firing of the origin has been identified (Fletcher and Chen, 2006).

S-CDK activity triggers the formation of the pre-loading complex (pre-LC) by phosphorylating Sld2 and Sld3 and promoting their binding to the phospho-binding BRCT domains of Dpb11

Phosphorylated Sld3 not only binds the N-terminal BRCT domain of Dpb11, but also Cdc45: a factor which is required for the loading onto the replication bubble of polymerases and of the RPA protein (Replication Factor A, the eukaryotic single-stranded binding protein, Aparicio *et al*., 1999). The role of pre-LC can be to recruit GINS and the leading DNA polymerase Polε at the origin and allow the formation of the Cdc45/MCM/GINS (CMG) complex (Tanaka and Araki, 2010). GINS (go, ich, nii, san complex) is a complex of four polypeptides comprising Sld5 and its partners Psf1, Psf2, Psf3. GINS association with the origins depends on Sld3-Dpb11 binding. What is the exact role of GINS in the CMG complex is still unclear: a speculative model has been proposed where GINS is acting like a ploughshare to sterically separate the two DNA strands (MacNeill 2010). MCM10 has been shown to directly interact with the Polα primase and to stimulate its catalytic activity (Fien and Hurwitz, 2006; Ricke and Bielinsky, 2004).

Upon the formation of the pre-initiation complex (pre-IC), the MCM helicase is activated and, together with Cdc45, GINS and various polymerases, moves with the

fork. More factors are recruited to form the RPC (replication progression complex), as for example Top1 topoisomerase, the fork protection complex (composed by Mrc1/Claspin, Tof1/Tim and Csm3/Tipin.), Ctf4/And-1, etc.

What has been discussed so far is an overview of the budding yeast *S. cerevisiae* initiation and elongation steps (Figure 1.1). Although this represents a good model, there are differences between various eukaryotic organisms and for some of the players the counterparts in higher eukaryotes have not been yet identified. In fission yeast as in metazoans the replication origins lack consensus sequence, but they contain often AT-rich region. In *Schizosaccharomyces pombe* the ORC complex has an extra DNA binding domain (located at the N-terminal of Orc4) that seems to be important for the recognition of this AT-rich elements (Sun and Kong 2010). *S. pombe* ORC does not require ATP to bind the DNA.

Dpb11 homologous are found in fission yeast (Cut5) as well as plants and metazoans (human TopBP1 and Drosophila Mus101). These homologues are more complex and possess more BRTC domain. In *Xenopus* Cut5 (*x*Cut5) is required for the transition from pre-RC and pre-IC. A possible functional analogue of Sld2 in human cells may be the RecQL4 helicase, whose N-terminal region shares some sequence homology with Sld2. This protein seems to be essential for replication, and in *Xenopus* is able to interacts with *x*Cut5. However, unlike Sld2, RecQL4 appears to function after Cdc45 and GINS loading onto replication origins (Pospiech *et al.*, 2010).

More open is the debate as to the analogue of Sld3 as, in the last year, three different proteins have been suggested as putative homologues of Sld3 in metazoans: DUE-B (DNA Unwinding Element Binding), Treslin/Ticrr (TopBP1-interacting replication-stimulating protein), and GEMC1. Sanchez-Pulido and colleagues (2010) show that significant sequence homology does exist between Treslin/Ticrr and Sld3, suggesting an orthologous evolutionary relationship between these two proteins. Like Sld3, all three proteins can be found in a complex with TopBP1, have been connected to CDK activity and they all function specifically at the Cdc45 loading step. At least one of these (GEMC1) was found to be a bona fide CDK target. However, the binding of Treslin/Ticrr to pre-RC itself is not CDK-dependent, as for Sld3. One possibility is

that in higher eukaryotes the functions of Sld3 in replication initiation has been taken up by several proteins, not necessarily all related by sequence to Sld3. This would provide to metazoan organisms more CDK targets in replication initiation, allowing for additional regulatory inputs into this crucial event. Another possibility is that DUE-B or GEMC1 performs Sld2's role in CMG assembly. In this regard it is interesting that the recruitment of DUE-B to pre-replication complexes is Cdk2-dependent, which also appears to be the case for Sld2 (as reviewed by YV. Fu and JC. Walter, 2010). Moreover, in metazoans it has been demonstrated that Cdc45 can directly interacts with Dbp11/Cut5/TopBP1 (Pospiech *et al*., 2009).

Cell cycle regulation of DNA replication

All the steps described above are strictly regulated in a cell cycle dependent manner. The assembly of pre-RC takes place at late M and early G1 phase when CDK activity is low, whereas high CDK activity (in S, G2 and early M phase) prevents the formation of the preRC as most of the proteins that forms the pre-RC can be phosphorylated and phosphorylation inhibits their activity.

The phosphorylation of Sld2 and Sld3 by S-CDK (see previous paragraph) is necessary and sufficient for the establishment of the replication fork (Tanaka and Araki, 2010). The essential target(s) of DDK is still unknown.

The expression profile or the location of the essential factors in the cell can be a further way to control the initiation of replication. In *S. cerevisiae* Cdc6 is only expressed in G1 and is degraded in the other phases of the cell cycle, whereas Cdt1 and MCM protein are present in the nucleus only in late M, G1 and S. After replication is initiated those proteins are exported into the cytosol. In fission yeast Cdc18, the homologue of Cdc6, and Cdt1 are present only in G1 and are then degraded (Sun and Kong, 2010).

In metazoans there are some extra regulation processes to inhibits the pre-RC formation in S and G2 phase, as for example accumulation of the replication inhibitor Geminin, a protein which sequesters the Cdt1 factor, therefore preventing MCM2-7 from being loaded (McGarry and Kirschner, 1998).

In G1-phase MCM2-7 is transported at the origin in a complex with Cdt1, to form the pre-RC. Upon the S-phase the helicase is activated by DDK, that phosphorylates the MCM2, MCM4 and MCM6; this post-transcriptional modifications may trigger the conformational change needed to activate the helicase. In *S. cerevisiae* a MCM5 mutant (*bob1*) that is able to bypass the phosphorylation event, promoting premature firing of the origin replication, has been identified: probably this mutant (involving a Pro->Leu mutation at position 83) mimics the "active configuration" of the enzyme.

The MCM complex can also be modulated during the elongation step, as the collapse of the replication fork due to double-strand breaks triggers a S-phase checkpoint cascade that blocks the elongation until the problem is fixed. How the mechanism works is unclear, but the interaction of the helicase with the "fork protection complex" seems to be indispensable to stop the DNA unwinding.

1.1.3 In Archaea

Despite their morphological similarity to Bacteria, the DNA replication apparatus of Archaea shares lots of similarities with the eukaryotic replisome. Archaea are therefore used as a eukaryotic model, because their enzymes involved in DNA metabolism resemble a simpler version of the eukaryotic ones (Kelman and Kelman, 2003). For example the eukaryotic helicase MCM2-7 is an heterohexamer composed by 6 paralogues, whereas in the majority of Archaea there is only one gene encoding for a MCM-like protein, generating a homo-multimeric functional helicase.

All archaeal genomes contain at least a gene that encodes for a protein which shares homology with both eukaryotic Cdc6 and Orc1 and is variably referred to as Cdc6, Orc1 or Cdc6/Orc1 protein. In this work I will use the Cdc6 nomenclature.

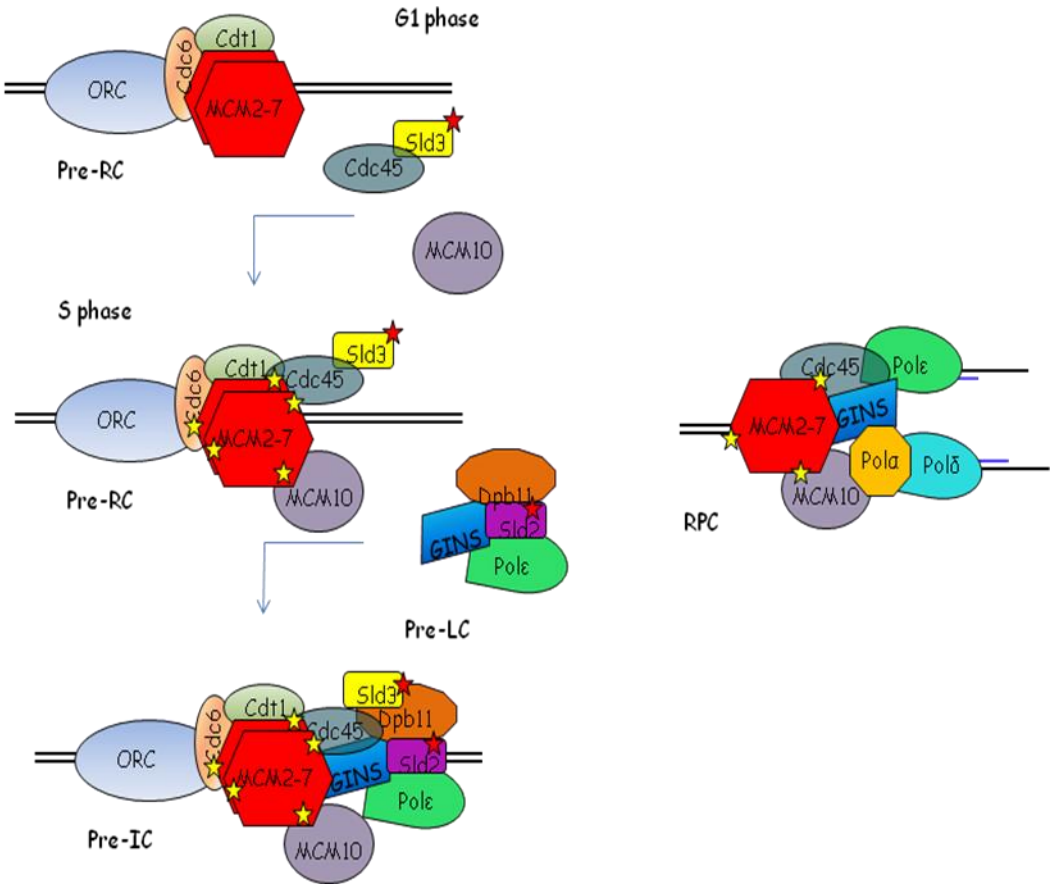


Figure 1.1 Schematic representation of the initial steps of eukaryotic replication. During the G1 phase of the cell cycle the pre-RC (constituted by ORC, Cdc6, Cdt1 and MCM) is recruited at origins of replication. Upon DDK and CDK mediated activation, a number of additional factors are recruited, to form the pre-initiation complex. These include MCM10, Cdc45, GINS, Sld2, Sld3 and Dpb11. The spatial organization and the stoichiometry of these factor is unclear as observation made in different species do not give a consistent picture. During the elongation phase of DNA replication a complex made of MCM, GINS and Cdc45 has been found to travel together with the replisome.

The archaeal genome is a single chromosome with varying number of DNA replication origin, ranging from one up to five like for *Haloferas volcanii* (Norais *et al*., 2007). In a study on *Sulfolobus solfataricus*, a number of specific DNA sequences (Origin Recognition Boxes, or ORB) have been identified within three origins of replication (named oriC1-3), which are recognized by different Orc1/Cdc6 factors: Cdc6-1, Cdc6-2 and Cdc6-3, (Robinson *et al* 2004). The ORB elements are usually found flanking the DUE sequence and the interaction with the Cdc6 may impose a topological stress that facilitates the melting of the AT rich zone. Cdc6

bound to the DNA will be the landing platform for the other proteins involved in the formation of the pre-RC.

The crystals structures of various archaeal Cdc6 orthologues are available: a monomeric structure of the *Pyrobaculum aerophilum* Cdc6 and *Aeropyrus pernix Cdc6-2* (Liu *et al*., 2000; Singleton *et al*., 2004). These structures highlight that the Cdc6 proteins comprise two candidate regions for the interaction with DNA: an helical insertion (ISM) in between the Walker A and Walker B sequence motifs and a winged helix (WH) C-terminal region. Later on the crystallographic structures of an heterodimeric complex from *S. solfataricus* composed by Cdc6-1/Cdc6-3 (Dueber *et al*., 2007) and a monomeric *A. pernix* Cdc6 (Gaudier *et al*., 2007) bound to nucleic acid (an ORB element) were published. In both structure the proteins bound an ORB elements, and the Cdc6 initiator factors clamp around and distort the DNA, contacting the double helix via the two motives previously described.

In Archaea the Cdc6 proteins may can also act as the helicase loader; biochemical studies show that they have properties similar to the bacterial DnaC interacting with helicase. In *Methanotermobacter thermoautotrophicum* the C-terminal WH of Cdc6 interacts with the N-terminal region of MCM destabilizing the ring and probably in this way promoting the loading of the helicase onto the origin. This interaction cannot be detected using the full length MCM protein (Kasiviswanathan *et al*., 2005), possibly suggesting that a conformational change is needed (which occurs more easily in the N-terminal domain alone), or that another initiator factor is required (Gautier *et al*., 2007, Kasiswanathan *et al*., 2005). A potential candidate is the WhiP initiator factor, a divergent Cdt1 homologue discovered in *S. solfataricus*, whose binding to origins of replication is stimulated by Cdc6-1 and Cdc6-2 (Robinson and Bell, 2007). The WhiP protein may works in a way similar to the bacterial plasmid initiator RepA, suggesting that this factor may play a critical role in the stabilization of the archaeal initiator/helicase heteromeric complex (Figure 1.2).

The interaction with Cdc6-like proteins may be a mechanism of MCM regulation, as in *M. thermoautotrophicum* and in *S. solfataricus* the helicase activity of MCM is affected and inhibited by Ccd6 binding, that can destabilize the hexameric ring. This

can be also the way for promote the MCM loading at the origin. On the other hands in *Thermoplasma acidophilum* MCM activity (helicase and ATPase) is stimulated by Cdc6 interaction (Sakakibara *et al.*, 2009).

How DNA replication is initiated is not yet understood and only in the last few years archaeal homologues of the GINS proteins have been identified by sequence analysis (Makarova *et al*., 2005). In all archaea a protein with sequence similarity to both Sld5 and Psf1 can be identified (GINS15), whereas a gene coding for a protein homologous to Psf2 and Psf3 (GINS23) is found in the Crenoarchaeota (such as *S. solfataricus*) and in a subset of Euryarchaeota (such as the Thermococcales, i.e. *Pyrococcus furiousus*). When two GINS proteins are found, biochemical studies show that two copies of each seem to be present, generating a quaternary structure similar to the eukaryotic tetramer. No information has been published on the stoichiometry of the archaeal single GINS15 proteins. No obvious archaeal homologues of Cdc45, Sld2, Sld3, Dpb11 and Mcm10 have been detected to date (MacNeill, 2010).

Yeast two-hybrid experiments using *S. solfataricus* proteins have shown that GINS23 interacts with the N-terminal region of the MCM complex and with the DNA primase (Marinsek *et al*., 2006); whereas the *P. furiosus* GINS complex interacts with a Cdc6 protein (Yoshimochii *et al*., 2008), suggesting that GINS is involved in the initiation as well as the elongation step of the DNA replication.

The biochemical effect of the GINS/MCM interaction in Archaea may be species specific, as in *P. furiosus* GINS was shown to stimulates the MCM helicase activity, whereas no stimulation is detectable in *S. solfataricus*.

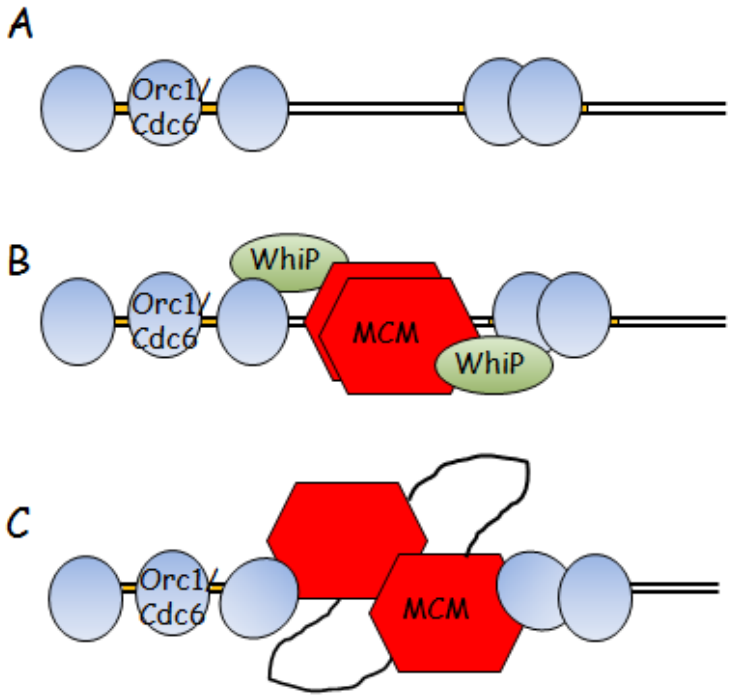


Figure 1.2. Schematic representation of the putative steps in the initiation of archeal DNA replication. A) the ORB elements (orange boxes) are recognized by Orc1/Cdc6 protein, B) the MCM complex is loaded at the origin and Whip is stabilizing the interaction between Cdc6 and MCM, C) the topological stress produced by the proteins/dsDNA interaction may contribute to the formation of the replicative bubble.

1.2 What we know about MCM protein.

The minichromosome maintenance (MCM) proteins were first identified in budding yeast as temperature sensitive mutants defective in the maintenance of minichromosomes, and were subsequently shown to play a crucial role in plasmid replication and cell cycle progression (Tye, 1999; Forsburg, 2004). Eukaryotic cells contain six conserved MCM polypeptides, named MCM2, MCM3, MCM4, MCM5, MCM6 and MCM7, which form a multiprotein complex (MCM2-7). MCMs are AAA+ proteins ranging from 600 to 1200 residues, and they are essential for the initiation and progression of the replication fork (Figure 1.3).

The eukaryotic MCM proteins have been shown to be chromatin-bound in the G1 phase, displaced during replication in S phase, and absent from chromatin in the G2 phase (Diffley and Labib, 2002; Stillman, 2005). The regeneration of replication competence is associated with a new binding event of MCMs to chromatin at the end of mitosis (Blow and Dutta, 2005). Although a large body of evidence suggests that MCM2-7 is involved in DNA unwinding, for a long time no helicase activity was detectable for the MCM complex, with only a weak unwinding activity associated with the MCM4/6/7 sub-complex. This sub-complex, forming an hexameric ring, as suggested by size exclusion chromatography and EM imaging (Bochman and Schwacha, 2007), has an ATP-dependent 3'→5' helicase activity and a preference for ssDNA than dsDNA.

A stable complex including MCM2-7 as well as Cdc45 and GINS, can be purified from *Xenopus* embryo extracts and displays an ATP-dependent helicase activity (Moyer *et al.*, 2006), thus suggesting that Cdc45 and GINS are essential cofactors. This study has been recently confirmed by the assembly of a catalytically active *Drosophila* CMG complex in baculovirus-infected insect cells (Ilves *et al*. 2010). In a parallel effort, a biochemical study identified specific buffer conditions in which the helicase activity of the recombinant human MCM2-7 complex alone is measurable. The presence of glutamate facilitates the Mcm2-7 helicase activity, with glutamate likely displacing inhibitory anions, such as chloride ions. Since DNA is itself negatively charged, helicase stimulation cannot involve the binding of

glutamate to DNA. Moreover, since glutamate changes the ability of Mcm2-7 to bind different topological forms of ssDNA rather than changing its affinity for ssDNA, glutamate likely facilitates a conformational change in the complex. The finding that glutamate mimics many of the effects of ATP, a known Mcm2-7 substrate, further suggests the physiological relevance of the effects of glutamate (Bochman and Schwacha 2008). Although chloride is commonly present in many in vitro enzymatic reactions, it is not the major cellular anion in eukaryotes (see Leirmo *et al.*, 1987 and references therein)

In archaeal genomes there is at least one gene encoding for a functional MCM paralogue, which can form homoligomers, and has been extensively studied as a model to understand how the replicative helicase works. The archaeal complex displays a robust 3'→5' helicase activity dependent on Mg^{2+} and ATP hydrolysis (Kelman *et al.*, 1999; Chong *et al.*, 2000; Shechter *et al.*,2000 Carpentieri *et al.*, 2002). The *S. solfataricus* MCM, as well as the MCM from *Archeoglobus fulgidus* and *A. pernix*, has a preference for binding bubble or Y-shaped DNA substrates. A large number of biochemical studies have been carried out on those proteins, elucidating the role of the separate domains (Kasiviswanathan *et al.*, 2004; Barry *et al.*, 2007; Pucci *et al.*, 2007), of various functional elements (McGeoch *et al.*, 2005; Jenkinson and Chong, 2006; Sakakibara *et al.*, 2008; Barry *et al.*, 2009, Brewster *et al.*, 2010), as well as specific residues (Pucci *et al.*, 2004; Moreau *et al.*, 2007; Jenkinson *et al.*, 2009) in DNA binding, ATPase and helicase activity (Sakakibara *et al.*, 2009).

Other two MCM proteins have been described in higher eukaryotes and called MCM8 and MCM9; more detailed bioinformatics analysis demonstrated that these proteins are actually widespread in eukaryotic organisms.

Another protein with some weak homology to the MCM family, named MCM-BP (MCM binding protein), was shown to co-precipitate with and MCM3-7 complex in human cells, therefore substituiting MCM2, but nothing is known about its function (Sakwe *et al.*, 2007).

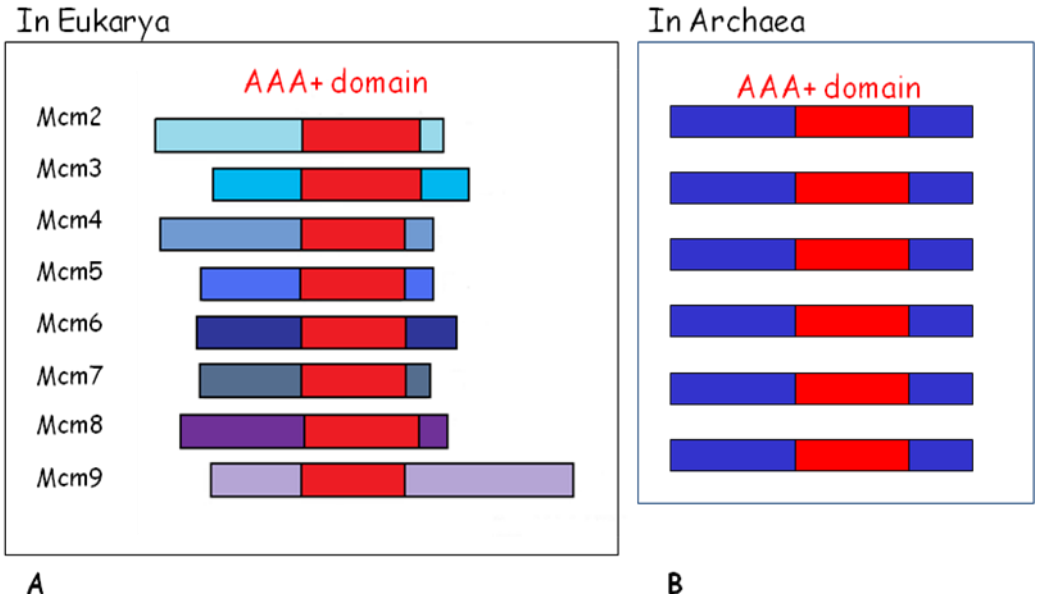


Figure 1.3. Schematic representation of eukaryotic and archaeal MCM protein. In Eukarya eight different genes are found, encoding for eight paralogues with a highly conserved AAA+ catalytic domain, whereas in the majority of Archaea there is a single gene encoding for a functional MCM protein, forming a homo-oligomer.

1.2.1 MCM structures

A canonical MCM protein can be divided into three regions: the N-terminal (N-term) domain, the AAA+ catalytic domain and a winged helix (WH) C-terminal domain (Figure 1.4).

Biochemical data show that the N-term binds both single stranded and double stranded DNA, is a determinant for the hexamerization and influence the processivity and the polarity of the helicase. Two crystal structure of this region are available: one from the MCM of the euryarchaeon *M. thermoautotrophicum* (*Mth*MCM-N) (Fletcher *et al*.,2003) and the second from the crenarchaeon *S. solfataricus* (*Sso*MCM-N) (Liu *et al*., 2008).

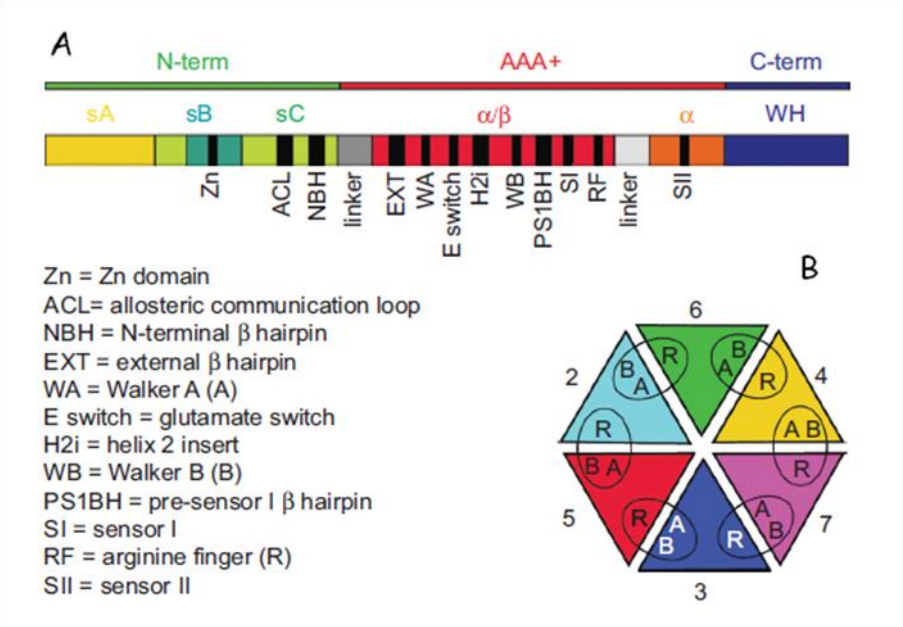


Figure 1.4 The architecture of MCM proteins. A) The primary structure of a tipical MCM protein can be divided into a N-terminal domain (green), a AAA+ domain (red) and a C-terminal domain (blue). The N-terminal domain can be further subdivided into three subdomains and includes functional elements, such as the Zn motif, N-terminal β-hairpin (NBH) and ACL loop. The AAA+ domain folds into an α/β subdomain (red) and an α domain (orange); the former contains the Walker A, Walker B, sensor I and R finger motifs found in AAA+ protein, as well as the unique insertions (EXT, h2i and PS1BH) which characterize MCM; the latter contains the sensor II motif. The helical linkers connecting the N-terminal-AAA+ domains, and the α/β - α domains are shown in dark gray and light gray, respectively. The C-terminal domain is predicted to fold into a winged helix (WH) motif. B) Schematic diagram of the putative MCM2-7 hexamer, showing the relative position of MCM proteins, as derived from biochemical data (Crevel *et al.*, 2001; Davey *et al.*, 2003; Bochman *et al.*, 2008).

The first one crystallises as a double hexamer with a head-to-head orientation, whereas the second as a single hexamer (Figure 1.5). Both present a central positively charged channel that can accommodate a single strand of DNA (ssDNA), but the channel of *Mth*MCM-N is large enough to also encircle a double stranded DNA (dsDNA). It is possible that the N-term region preferentially binds ssDNA or that the *Sso*MCM-N has been trapped in an inactive conformation.

The N-term can be divided into three subdomains (sA, sB and sC) and their biochemical functions have been extensively characterized. Subdomain A is the peripheral belt of the ring and is involved in the regulation of the helicase activity, so

that when this region is missing MCM is a more processive helicase. Subdomain B, at the interface of the double ring in the dodecameric *Mth*MCM-N structure, contains the Zn motif that contributes to DNA binding, as deletion of this domain impair ssDNA binding but not helicase activity. Subdomain C is responsible for the hexamerization and contains a β-hairpin (NBH) that protrudes towards the center of the channel and also participate to DNA binding (Kasiviswanathan *et al*.,2004; Barry *et al*., 2007). Subdomain C also comprises the ACL loop (allosteric communication loop, Sakakibara *et al*. 2008; Barry *et al*., 2009), highly conserved in archaeal and eukaryotic MCMs; this loop has an important role in the communication between an N-term and the adjacent AAA+ domain in the hexameric ring.

Various lines of evidence, based on structural analysis of the N-term suggest that this domain may undergo large conformational changes. The crystal structure of a single mutant in the *Mth*MCM (Proline 61 to Leucine) that mimics the MCM5 *bob1* in yeast (Fletcher *et al*., 2003), mutation that can bypass the DDK phosphorylation event for the activation, has a more open sA. This small movement is also confirmed by the 3D EM reconstruction of the N-terminal domain alone, where the sA is drastically open. A less dramatic swing out movement of sA is also evident when superimposing the two N-term crystals structure, with the *Sso*MCM-N having sA in a more "open" configuration (Costa *et al*., 2008).

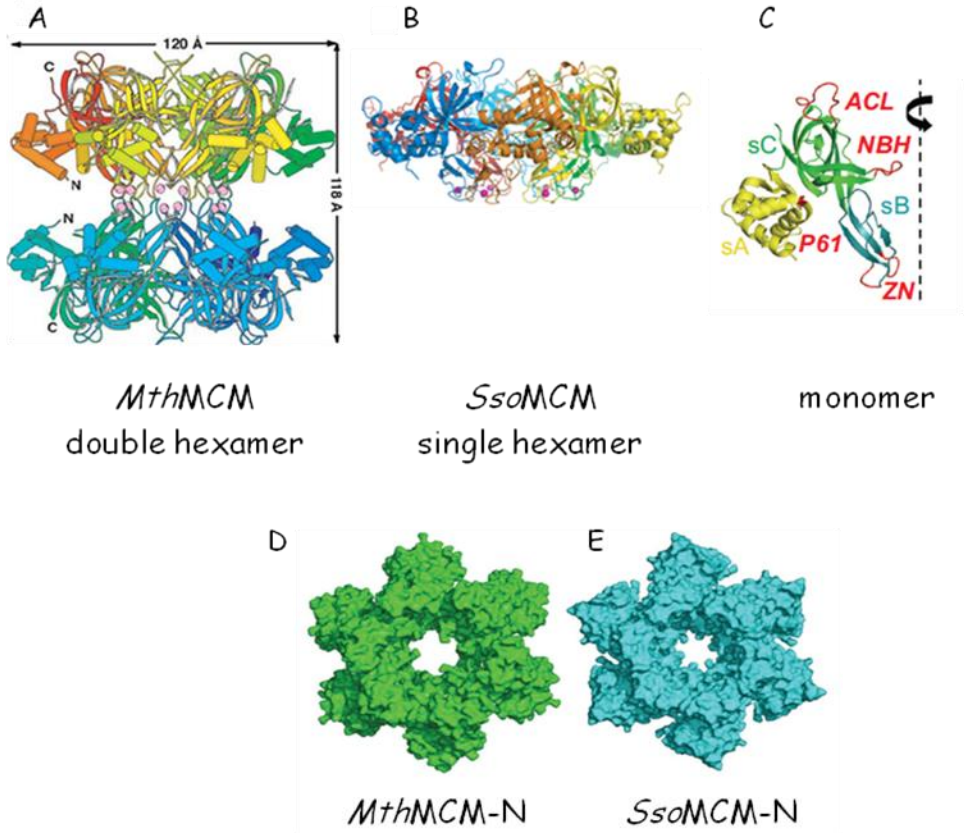


Figure 1.5 The crystal structure of the N-terminal domain. A) A side view of the dodecameric structure of *M. thermautotrophicus* MCM (Fletcher *et al.*, 2003) and B) of the hexameric structure of *S. solfataricus* MCM *(*Liu *et al.*, 2008). C) Each monomer folds into three subdomains: subdomain A (sA, yellow) is located in the outer belt of each ring; subdomain B (sB, in teal), coordinates a zinc atom that is located at the interface between the two hexameric rings; subdomain C (sC, in green) is responsible for the inter-subunit contacts within a hexamer. The structural elements that are important for the activity of the protein are highlighted in red. A dashed line indicates the position of the six-fold axis around which the hexameric complex forms. D) and E) Surface representations of the *Mth*MCM-N and *Sso*MCM-N hexamers (in green and cyan, respectively) highlight the differences between the two structures, especially regarding the size of the internal channel.

The central AAA^+ domain folds into two subdomains: an α/β subdomain and a α domain (also called "lid" domain). This region alone can form an hexamer and possesses DNA helicase activity. The catalytic domain contains the Walker A (WA), Walker B (WB), sensor I, sensor II and Arginine finger signature sequence motifs typical of AAA+ ATPases. All these motives are required for the ATP binding and hydrolysis (Figure 1.4).

In the last year structural information for two full-length MCM proteins has become available, providing a model for the AAA+ domain. One is the structure, solved at

high resolution (1.9 Å), of an unusual, inactive and monomeric MCM from *Methanopyrus kandleri* (*Mka*MCM2) (Bae *et al.*, 2009); this protein is lacking the majority of the functional motives, including subdomain B, WA, WB, sensor I and R finger as well as the C-term region. The second structure is a monomeric (almost) full length *Sso*MCM, since for the C-term region only weak and partial electron density was detectable; this structure was solved at much lower resolution (4.35 Å) but gives us a picture of a functional MCM (Brewster *et al.*, 2008). Both monomeric structure can be superposed on the *Mth*MCM-N to construct the hexameric ring (Figure1.6).

As predicted from structural comparisons (Erzberger and Berger, 2006) and biochemical studies on the trans nature of the sensor II (Moreau *et al*., 2007), due to the insertion of a helix at the end of the α/β-domain, α/β domain and α domains present a different relative arrangement compared to a classical AAA+. In this protein the α domain is acting in trans and interacting with the ATP of the neighbour active site: in fact the sensor II protrudes in the active site of the adjacent subunit (Figure 1.7).

MCM proteins are characterised by an insertion in the α/β subdomain occurring before the sensor I motif and forming a β-hairpin. This insertion has been called the pre-sensor-1 β-hairpin (PS1BH) and defines a superclade of the AAA+ superfamily. Within the PS1BH superclade, a number of AAA^{+} proteins, including MCM, contain an additional insertion disrupting the continuity of helix 2 (Iyer *et al*., 2004). The helix 2 insertion (h2i) and the PS1BH are critical for DNA unwinding (Jenkinson and Chong, 2006; Barry *et al*., 2007).

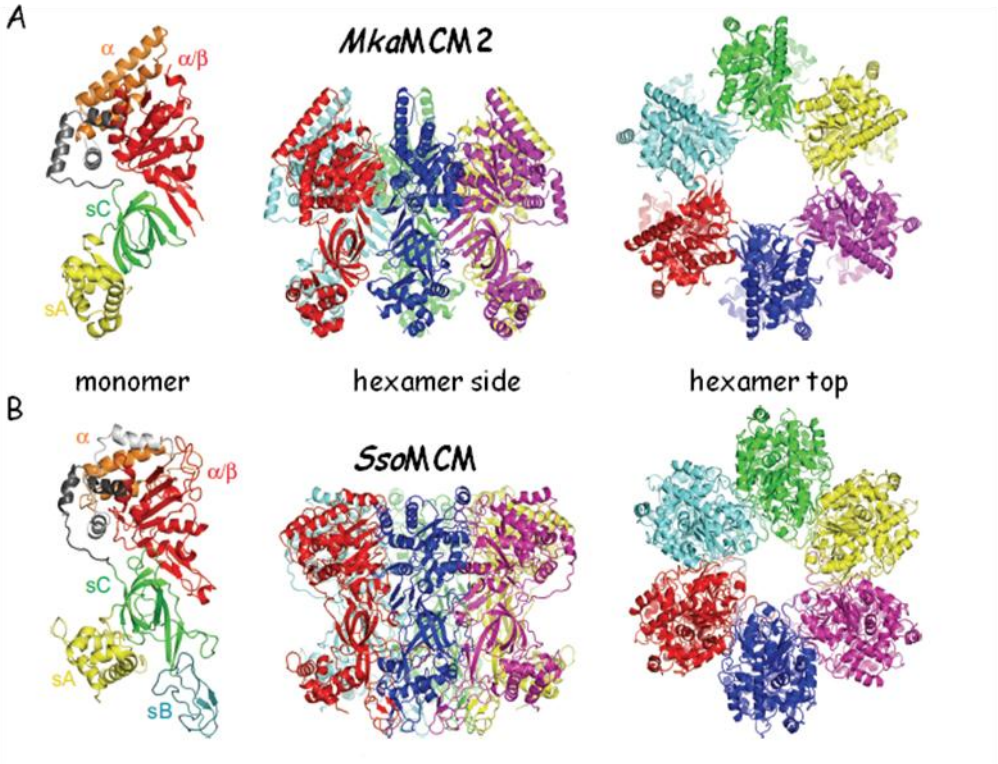


Figure 1.6 Structural models of the archaeal MCM helicases. A) The crystal structure of the MCM2 of *Methanopyrus kandleri* (Bae *et al.*, 2009). The sequence is unusual for an MCM protein, as it does not have subdomain B and the C-terminal domain, and lacks a number of conserved elements, such as ACL, NBH, WA, WB, PS1BH and R finger. The protein is a monomer in solution and in the crystal. The AAA+ domain folds into two subdomains (α/β in red and α in orange) and is connected to the N-terminal domain through a long loop wrapping around the AAA+ domain and a couple of helices (dark gray): the N–C connection is intertwined with another helical extension (light gray – partially disordered in the structure) that links the α/β and αdomains. A hexamer can be built based on the symmetry of the *Mth*MCM-N structure B) The crystal structure of the MCM of *Sulfolobus solfataricus* (Brewster *et al.*, 2008). The C-terminal domain is disordered in the crystal and no atomic model could be built. The color code for the monomer is the same as in figure 4. Although the protein can form a hexamer in solution, a monomer is present in the crystal; a hexamer can be built based on the symmetry of the *Mth*MCM-N structure.

The EM maps (Bae *et al.*, 2009) and the modelled hexameric ring also shows the putative interaction between the PS1BH of one subunit and the ACL loop of the adjacent subunit, interaction that is responsible both for the communication between N-term and AAA+ domain and between adjacent subunits.

The crystal structure of the *Sso*MCM reveals a conserved external loop (EXT) at the lateral hole and mutants of this loop are defective in the helicase activity and DNA binding (Brewster *et al.*, 2008; Brewster *et al.*, 2010).

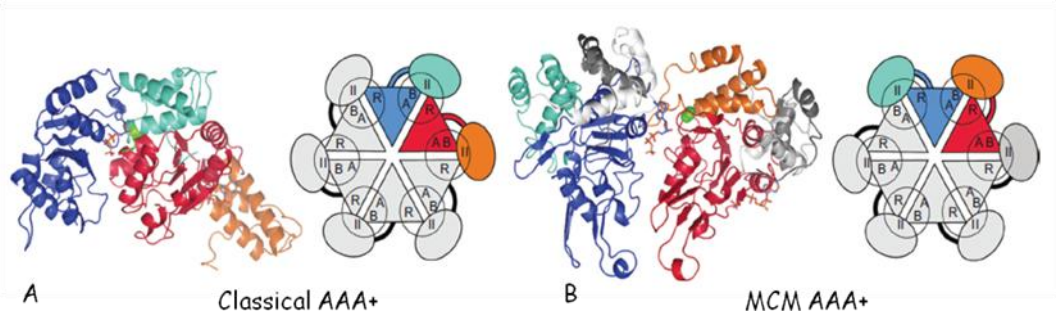


Figure 1.7 The unusual architecture of the MCM AAA+ hexamer. A) The interface between two subunits within a hexamer as seen in the crystal structure of the bacterial transcriptional enhancer PspF bound to ATP (Rappas *et al.*, 2006) shows the arrangement seen in the majority of AAA+ proteins. In one subunit the α/β domain is in red, with the α domain in orange, whereas in the adjacent subunit the the α/β domain is in blue, with the α domain in cyan; the sensor II arginine highlighted in green; the ATP molecule at the interface between subunits is shown. As schematically illustrated in the diagram, the ATPase active site at the interface between the red and blue subunit is made up of residues mostly belonging to the blue subunit, including Walker A (A), Walker B (B), sensor I, and sensor II (II, in green) acting in *cis*, whereas the arginine finger (R) is contributed in *trans* by the α/β domain of the next-neighboring subunit. B) The same inter-subunit interface is shown for a hexameric model derived from the crystal structure of *Sso*MCM-FL (Brewster *et al.*, 2008) overlapped on the *Mth*MCM N-terminal domain structure (Liu *et al.*, 2008). The same color code as for PspF is used; in addition the connection between the N and C-terminal domain is shown in dark gray, whereas the MCM peculiar insertion between the α/β and α domains is shown in light gray. The effect of this helical insertion is to dramatically reposition the α/β domain so that it is now located at the adjacent interface. The diagram illustrates how the active site now comprises the Walker A, Walker B, sensor I residues from one subunit (in *cis*) and the R finger and sensor II from the adjacent subunit (in *trans*).

Similarly to the N-term region, also the AAA+ domain may be flexible. The conformational changes of the subunits can be caused by ATP or DNA binding, as well s ATP hydrolysis.

Due to the lack of crystallographic data, most of the evidence for these conformational changes comes from single particle electron microscopy (EM). Analysing the 3D EM reconstruction of the *Mth*MCM with different substrate is possible to visualise changes in the stoichiometry of the ring (seven subunits in presence of nucleotides or six subunits in presence of a fragment of dsDNA of 100 base pairs). In the presence of dsDNA the central channel is partially occupied by peaks of electron density. By fitting the AAA+ domain structure into the EM map, this electron density can be modelled by the h2i protruding into the middle of the channel (Figure 1.8). The change in the thermostability of the *Mth*MCM when the protein is bound to the substrate is another evidence for substrate-induced

conformational changes (Sakakibara *et al*., 2009). Mutagenesis studies of the PS1BH in *Sso*MCM and of the h2i in *Mth*MCM (McGeoh *et al*., 2005; Jenkinson and Chong 2006) demonstrate that those motives are crucial for the DNA binding and the helicase activity; it has been suggested that the h2i protruding in the central channel could interact with the nucleic acid and allow the separation of the strands using a conserved aromatic residue as ploughshare (Costa and Onesti, 2009).

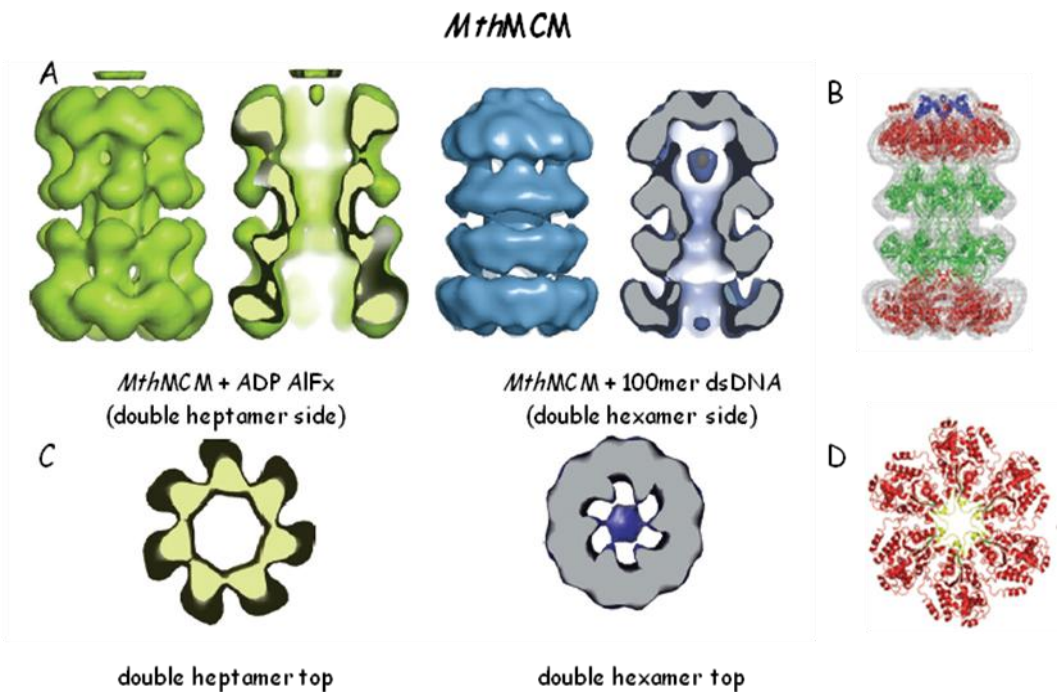


Figure 1.8 3D EM reconstruction of *Mth*MCM. A) Surface rendering of double-ring structures of *Mth*MCM from negatively stained samples in the presence of substrates (Costa *et al.*, 2006a). On the left (in yellow) a side view of the ADP·AlFx treated double heptamer showing a head-to-head double-ring configuration with lateral holes, and a slice through the center, showing the large uninterrupted channel through the molecule. On the right a side view and a slice-through of the protein treated with 100 bp of dsDNA (in blue). The protein forms a double hexamer with a strong asymmetry between the two rings. The upper ring is closed on top by a cap, and has the central channel obstructed by a peak of electron density, whereas the bottom ring is hollow and uncapped. Seven-fold and six-fold symmetry, respectively, were applied to the reconstructions. C) A section through the middle of the upper rings in both structures shows the striking features of the dsDNA-treated protein, with a peak of electron density connected to the ring by six protrusions. B) and D) Fitting of atomic models to the double hexameric EM map. The dodecameric *Mth*MCM-N atomic coordinates (Fletcher *et al.*, 2003) were fitted to the central part of the molecule (in green). For the AAA+ tier of the lower ring the *Sso*MCM-FL structure (Brewster *et al.*, 2008) arranged into a hexamer based on the *Mth*MCM-N model matches well the electron density (in red). C) and D) To fit the upper ring, a rigid body rotation needs to be applied to the AAA+ domain. The reorientation of the AAA+ domain in the top ring brings the helix-2 insertions (in yellow) towards the central channel, fitting into the lateral protrusions. The PS1BH elements are shown in yellow.

Sequence analyses predicts a winged helix (WH) motif in the C-terminal region of the archaeal enzymes. This prediction was recently confirmed, as the C-term region of the human MCM6 has been determined by NMR confirming a winged-helix fold made up of three helices and two β-strands (Figure 1.9). Winged helix proteins belong to the HTH family, typically including DNA-interacting domains, but no biochemical evidences for DNA binding by the C-terminal domain were found for *Sso*MCM (Barry et *al*., 2007) nor for the MCM6 C-terminal domain (Wei *et al*., 2010), although the deletion of this region leads to reduction of the DNA binding (Kasiviswanathan *et al.*, 2004, Costa *et al.*, 2008). C-term deletion mutants are also better helicase and ATPase, suggesting a possible regulatory activity but the exact role of this region is still unknown. However its importance is clear, since *in vivo* studies on the MCM of *H. volcanii* showed that most of the mutations involving the C-terminal domain were lethal (McNeill, unpublished result).

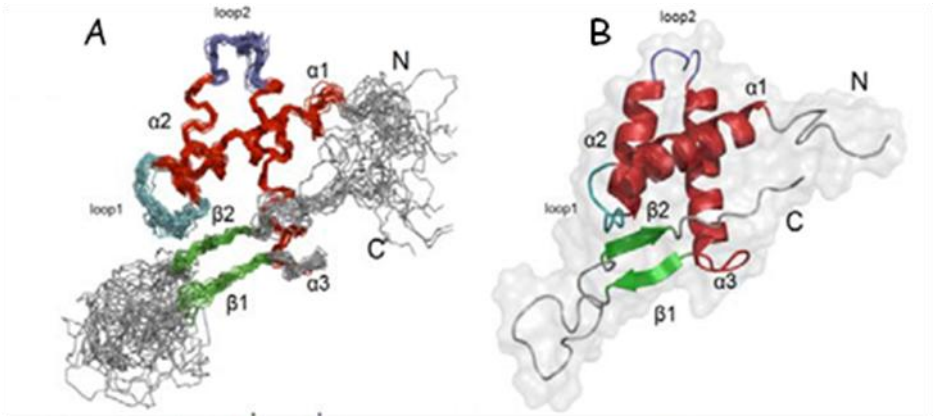


Figure 1.9 NMR structure of hMCM6-CBD. A) The backbone superposition of the 20 lowest-energy NMR structures of Mcm6-Cdt1 Binding Domain. Secondary structural elements are indicated by color coding: α helices (red), β-strands (green), and loops (cyan, gray). N-terminal and C-terminal ends are indicated as N and C. B) a ribbon representation of the same structure of CBD using the coordinates of the lowest energy structure

Different locations have been proposed for the C-terminal domain. In a double hexamer 3D reconstruction electron density occupying the central channel was suggested to be due to the C-term (Gomez-Llorente et *al*., 2005). A more recent and detailed study from our laboratory, comparing EM data for the full-length and a C-terminal deleted MCM protein, shows that the C-term region is likely to be located on top of the AAA^+ domain (Costa et *al*., 2006, Costa *et al*., 2008) (Figure 1.10).

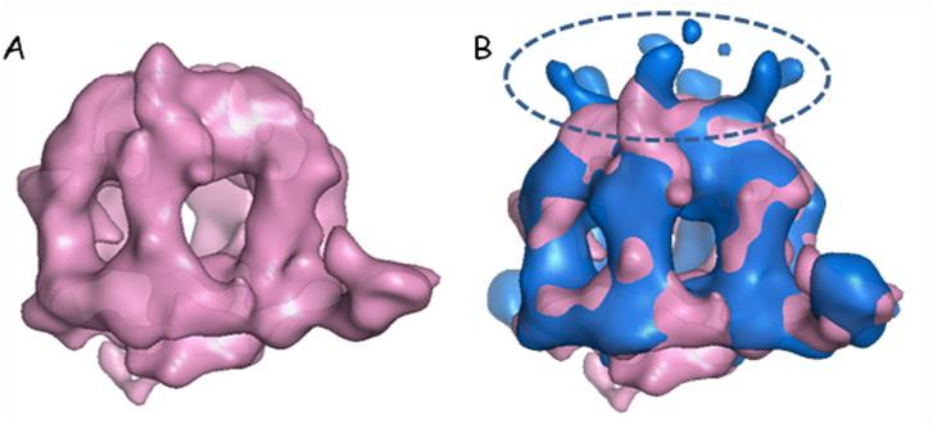


Figure 1.10 Location of C-term domain in *Mth*MCM. On the left, in magenta, the cryoEM structure of the C-terminal deletion mutant, on the right the superposition of the wt (blue) and ΔC (magenta) maps (Costa *et al.*, 2008): the two are very similar, except for the protrusions on the top of the molecule (highlighted by the dashed blue line) which are only present in the wt map

The MCM protein are really polymorphic; a good example is the *Mth*MCM. Different 3D EM reconstruction demonstrated that MCM can form single hexamers (Pape *et al*., 2003) or heptamers (Yu *et al*., 2002), as well as double hexamers or double heptamers (Gomez-Llorente *et al*., 2005; Costa *et al*., 2006a). The presence of substrates controls the stability and the stoichiometry, with short dsDNA fragments triggering the formation of head-to-head double hexamers, and nucleotides the formation of double heptamers (Costa *et al*., 2006a&b, Jenkinson *et al*. 2009). Double rings are seen in size exclusion chromatography, but they are probably unstable and not visible by EM when the protein is free from substrate.

But this protein can be also found in fibrous helical structures, where the orientation of the particles is "head-to-tail". Deletion of the N-terminal subdomain A strongly encourages the formation of helices (Chen *et al*., 2005) (Figure 1.11). Biochemical analysis have underlined as a certain degree of polymorphism may be dependent on the temperature and salt concentration (Shin *et al*., 2009).

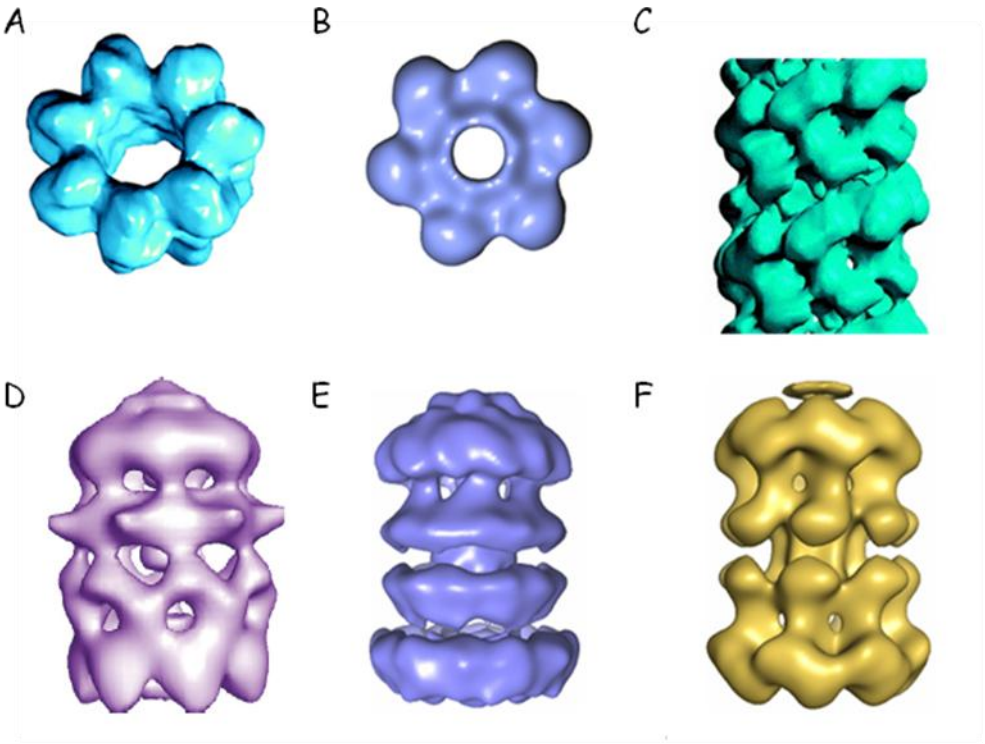


Figure 1.11 Polymorphism of *Mth*MCM as visualised by electron microscopy studies: A) single heptamer (Yu *et al*., 2002); B) single hexamer (Pape *et al*., 2003); C) helical fiber (Chen *et al*., 2005); D) R137A mutant double heptamers (Jenkison *et al*., 2009); E) double hexamer in the presence of 100 bp dsDNA; F) double heptamer in the presence of the nucleotide analogue ADP-AlF_x (Costa *et al*., 2006).

A recent 3D EM reconstruction of the *S. cerevisiae* MCM2-7 loaded onto DNA shows a structure compatible with the reconstruction of the *Mth*MCM complexed with dsDNA (Figure 1.12) (Remus *et al*., 2009). In this paper the authors are able to reproduce *in vitro* the assembly of the pre-RC and the helicase is loaded as a double hexamer (resistant to high salt washes) and able to translocate along a small plasmid. This strongly suggests that eukaryotic MCM are loaded onto the origin as double hexamers, with a head-to-head orientation.

As in the archaeal orthologue, lateral holes are visible in the MCM2-7 structure, as well as a central channel that can encircle a dsDNA (Remus *et al*., 2009). The length of the complex is consistent with footprinting experiments carried out on the MCM4/6/7 complex indicating that the protein protects a DNA region corresponding to the size of a double hexamer (You and Masai, 2005) as well as with the size of the pre-replicative footprinting (Santocanale and Diffley, 1996; Labib *et al*., 2001).

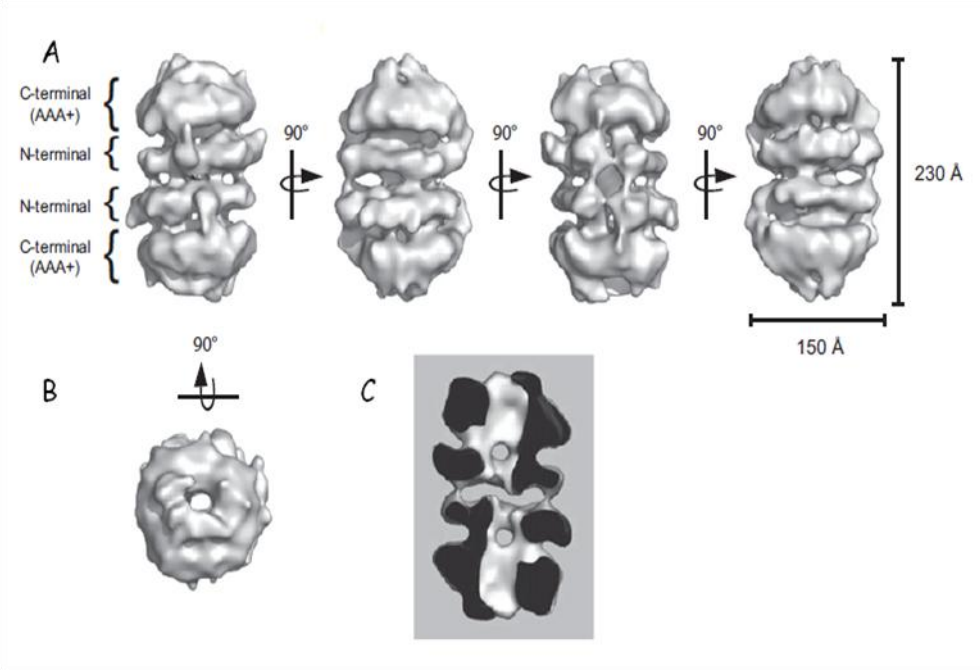


Figure 1.12 Three-Dimensional Reconstruction of the Mcm2–7 Double Hexamer. A-B) Surface representations of a 3D reconstruction of the Mcm2–7 double hexamer filtered to 30 Å at the predicted molecular weight for the Mcm2–7 double hexamer. C) Cut-open side view (from Remus *et al.*, 2009).

1.2.2 Mechanism of action

All the MCM proteins studied can bind and translocate along duplex DNA. It is not yet clear whether the protein is active as a single or double hexamer.
Whether the double hexamer has simply a role in the initial stage of DNA replication (helicase loading, initial melting), or is necessary for the helicase activity, remains to be established.

Still unknown is the physiological significance of the heptameric and double heptameric rings that are seen in the archaeal MCMs. One possibility is that the heptamer is an *in vitro* artefact, however, the formation of double heptameric rings is systematically triggered in the *Mth*MCM system by the addition of ATP analogues, such as AMP-PNP and ADP·ALFx (Costa *et al.*, 2006b). Heptameric rings are also stabilized by mutations at the very bottom of subdomain B; in particular the R137A mutant shows a double heptamer configuration with a remarkable swing-out of subdomain A in one of the two rings (Jenkinson *et al.*, 2009, Figure 1.11 D) . It is possible that the mutation "traps" a more labile transient intermediate state that is

necessary for the loading of the complex onto DNA. Although it is more difficult to extend this model to the eukaryotic hexameric MCM2-7, initial data on EM analysis of MCM complexes purified from eukaryotic cells shows hints of heptameric symmetry. This observation, although still very preliminary, has been made independently by two different groups (Remus *et al*., 2009; S. Onesti, A. Patwardhan and J. Blow, unpublished observations) working on two diverse eukaryotic systems. Further work is required to confirm the result and establish the biochemical nature of the putative heptamer.

Role in the initiation of DNA replication

MCM2-7 proteins have been shown to interact with chromatin in two different ways: as a labile "associated complex" that is disrupted by high salt washes, and as a "loaded complex" that is salt-resistant (Donovan *et al.*, 1997; Bowers *et al.*, 2004). The identification of an unexpected modality of contact between MCM and dsDNA, with the nucleic acid wrapping around the N-terminal half of the protein (Costa *et al.*, 2008) may provide a structural model for the associated complex.

The topological changes in the DNA that are a consequence of this interaction could assist the first steps in origin melting by changing the degree of supercoiling of the adjacent DNA regions, and hints at the possibility that MCM proteins may be actively involved in the initiation of DNA replication (Costa and Onesti, 2008). *Mth*MCM proteins have also been reported to form fibrous and helical assemblies (Chen *et al.*, 2005), which seem to be dependent on the presence of DNA (Costa *et al.*, 2008), reminiscent of the mechanism proposed for the bacterial initiator DnaA (Erzberger *et al.*, 2006), involving a helical structure around which the initiator DNA can be wrapped.

Role in fork elongation

Although the function of MCM proteins as replicative helicases is generally accepted, how the helicase unwinds the DNA duplex is still an open question, with a number of different models being proposed. The steric-exclusion model represents the simpler mechanism (Enemark and Joshua-Tor, 2008), where a helicase ring

encircles a single DNA strand and translocates along it; the other strand is displaced by being excluded from the ring. This mechanism probably applies to the hexameric helicases belonging to the SF3 and SF4 families (Singleton *et al.*, 2007), including the bacterial DnaB-like helicases, the bacteriophage T7 helicase (Singleton *et al.*, 2000) and the papillomavirus E1 helicase. The crystal structure of E1 bound to ssDNA and ADP strongly suggests a sequential mode of ATP hydrolysis linked to DNA translocation via the correlated movements of the PS1BH hairpins (Enemark and Joshua-Tor, 2006; Takahashi *et al*. 2005) have pointed out that in eukaryotes the helicase is loaded onto chromatin much before the helicase activity is actually needed (i.e. in the G1 rather than the S phase of the cell cycle). As presumably the loaded complex is topologically linked to DNA, the lack of evidence for origin melting before the beginning of S phase suggests that the complex is loaded onto dsDNA. However it cannot be ruled out that after initial loading onto dsDNA, MCM may undergo a rearrangement in such a way as to encircle ssDNA upon activation.

The peculiarity of the initiation of DNA replication has therefore led to a number of different suggestions which are all linked by having a MCM complex encircling (or partially encircling) dsDNA. A scheme was derived from the proposed model for the SV40 LTag helicase. In contrast with the model proposed for the homologous papillomavirus E1, this protein was suggested to remain bound to the viral origin of replication as a double hexamer and melt the DNA by extruding ssDNA through lateral holes (Li *et al.*, 2003), The last model proposed for MCM function advocates a hexameric helicase moving along dsDNA dragging a “ploughshare” that physically separates the two strands (Takahashi *et al.*, 2005). The LTag mechanism can be classified in this category if the two rings that compose the double hexamer come apart.

The available biochemical results are compatible with multiple models, showing that MCM complexes can displace a 59 nucleotides tail, translocate along dsDNA, or unwind a heterologous junction (Shin *et al.*, 2003; Kaplan and O’Donnell, 2004). MCM proteins can bind to both dsDNA and ssDNA (the latter with slightly higher affinity), but tend to have a higher affinity for forked and bubble substrates (Grainge

et al., 2003; You *et al.*, 2003; De Felice *et al.*, 2004). On the other hand the considerations summarized in Takahashi (2005) make a ploughshare model the most likely candidate to describe the activity of MCM proteins. Although the ploughshare could be an element of the MCM proteins or a separate protein (such as GINS or Cdc45), the fact that both archaeal and eukaryotic proteins display helicase activity in the absence of accessory factors argues in favor of an "internal" ploughshare (Figure 1.13).

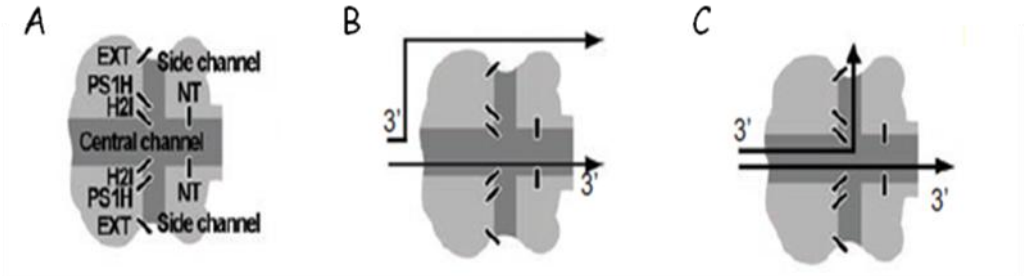


Figure 1.13 Mechanism of action of MCM proteins. A) schematic representation of a MCM protein and its loop B) Schematic diagrams illustrating the steric model, in wich the protein is moving on a ssDNA estruding the complementary strand, C) a diagram illustrating the ploughshare model with a strand of DNA coming out of the lateral hole and the second via the N-term region of the protein.

Of the two loops that can be modelled to project into the channel (PS1BH and h2i), the strongest contender for the ploughshare is the helix-2 insertion. An alternative model is that the h2i loop works by translocating dsDNA inside the channel (with a mechanism similar to the PS1BH-mediated translocation of ssDNA in the E1 helicase) and the "ploughshare" is simply represented by the lateral holes through which the ssDNA is extruded.

A model in which one strand exits through a lateral hole while the other emerges through the N-terminal hexamer is consistent with biochemical data, showing that the NBH has an important role in the regulation of helicase activity (Barry *et al.*, 2009) and that the N-terminal domain has a marked preference for ssDNA (Liu *et al.*, 2008). Moreover the position, sequence conservation and mutagenesis of the EXT loops (Brewster *et al.*, 2008; Brewster *et al.*, 2010) support a role of the holes in helicase activity.

From EM single particle work (Costa *et al.*, 2008), the lateral holes in MCM are often blocked by an isthmus of electron density that can be modelled by the inter-subunit interaction between the PS1BH and ACL loop. This interaction is disrupted in the DNA-bound double hexamer structure, where presumably DNA binding has caused a re-orientation of the h2i and PS1BH towards the center of the channel, with the consequent disengagement of the PS1BH from the ACL, possibly providing a mechanism to "free" the hole for the exit of the ssDNA.

1.3 The "new" eukaryotic MCM proteins.

In the last few years two new eukaryotic proteins belonging to the MCM family have been identified and called MCM8 and MCM9 (Gozuacik *et al.*, 2003; Johnson *et al.*, 2003; Lutzmann *et al.*, 2005; Yoshida, 2005); these two proteins share a much higher homology to each other than to the other MCM2–7 proteins (Figure 1.14). Initially found only in higher eukaryotes, a more recent in-depth bioinformatics analysis demonstrates that these proteins are widely spread in all eukaryotic taxa, with random losses in some species. An interesting observation is that these two proteins seem to share an evolutional correlation, with both of them usually present or absent in the organisms analysed, possibly reflecting a functional correlation (Liu *et al.*, 2009).

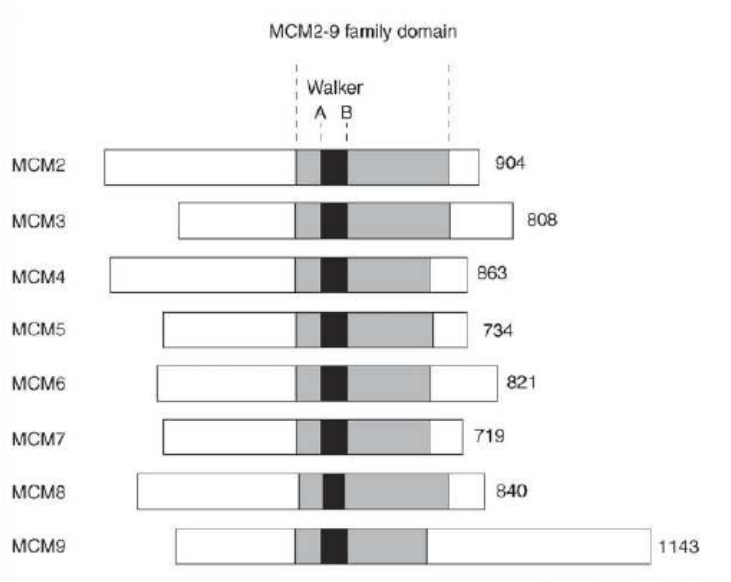


Figure 1.14 Schematic representation of the MCM family protein. The MCM8 and MCM9 contain all the characteristics of the other MCM protein. But the last to share much homology between each other. (From Maiorano *et al.*, 2006)

An exception is represented by *Drosophila: D. melanogaster* Mcm8 (REC) protein is present despite the absence of Mcm9 and is highly divergent. The divergence of REC from other Mcm8 sequences may be related to the absence of Mcm9 and a radical change in the functional role. In fact, in addition to the role in DNA replication (Crevel *et al.*, 2007), an implication in meiotic recombination has been identified for

REC. A model has been proposed in which REC drives crossover formation by acting at the repair synthesis step of meiotic recombination (Blanton *et al*., 2006). *Drosophila* was the only eukaryote analysed which has Mcm8 but lacks Mcm9 suggesting, if the co-function and distinct co-ancestry hypothesis is correct, that REC may have assumed the function of both proteins or acquired a novel function that does not require Mcm9.

Despite a few studies, the exact role of MCM8 in the cell is not clearly understood. This protein comprises a full set of the canonical motifs that have been shown to be essential for the helicase activity of the archaeal and/or eukaryotic MCM complexes. The protein does not associate with the MCM2-7 helicase complex. As reviewed in Maiorano *et al*. (2006) the information about the function of this protein are discordant. Studies carried out in human cells or in *X. laevis* egg extract demonstrate that this protein is interacting with chromatin after MCM3 during the S phase of the cell cycle. *X. laevis* MCM8 (xMCM8) colocalizes with RPA and in its absence the chromatin binding levels of RPA and Pol α-primase decrease. The recombinant xMCM8, produced in baculovirus-infected insect cells, shows helicase activity and ATPase activity stimulated by ssDNA; these results suggest that xMCM8 may act as a helicase during the elongation step. Other studies in human cell, however, suggest that human MCM8 (hMCM8) is needed for the recruitment of hCdc6 at the origin; hMCM8 results bound to the chromatin throughout the cell cycle and its downregulation by RNA interference leads to a delay in S phase entry and a lower level of hCdc6 bound to chromatin.

Very little is known about MCM9, which is the largest MCM protein known to date, with a really sizeable 500 amino-acid C-terminal domain. This protein has a direct interaction with Cdt1 and is indispensable for its recruitment at the origin and for the formation of the pre-RC (Luztmann *et al*., 2008).

1.4 My project

Archaeal C-terminal region: WH motif

When I started my work the C-terminal region of MCM proteins was still structurally uncharacterized and the only information available was based on bioinformatics sequence alignments. To confirm the presence of a WH fold within the C-terminal region of MCM and provide an atomic model to define the biological role of this domain, I cloned, expressed and purified the C-terminal domains of *Mth*MCM as well as *Sso*MCM, with the aim to carry out structural studies. The proteins have been obtained at high concentration and used for extensive crystallization experiments. In a parallel effort, preliminary NMR analysis has been performed in collaboration with Prof.ssa Henriette Molinari, Università di Verona.

hMCM8 and hMCM9

As demonstrated by biochemical studies the MCM8 proteins do not interact with the MCM2-7 helicase. This suggests that hMCM8 may act as a homo-oligomer and is therefore more similar to the archaeal model. I focused my work on this protein in order to get biochemical and structural information. As the full length MCM proteins have been rather resilient to large scale production and crystallization, I applied a different approach and separately express and purify a variety of sub-complexes corresponding to the different domains of the MCM8 (and possibly MCM9) complex.

2 Materials and Methods

PfuTurbo DNA polymerase and PfuTurbo Buffer were obtained from Stratagene, restriction endonucleases, restriction Buffers, BSA for restriction digestions and T4 ligase and T4 ligase Buffer were from New England Biolabs. IPTG (dioxan free) were from Carbosynth. Lysozyme (from chicken egg white), Maltose, PMSF, βME, BSA (for Western Blots) were all purchased from Sigma. Inhibitor cocktails EDTA free were purchased from Roche.

All other Buffers, salts and reagents used for the preparation of solutions were purchased from Sigma

2.1 Cloning of recombinant proteins in *E. coli* cells

2.1.1 Expression vectors used for cloning.

One commercially available bacterial expression vector and two other bacterial expression vectors kindly supplied by EMBL were used to produce the DNA constructs for the expression of recombinant proteins in *E. coli*. The main features of these vectors are summarized in Table 2.1.

2.1.2 Primers

The primers used for this work are listed in Table 2.2 and were provided by Sigma Genosys and resuspended in MilliQ water as manufacture datasheet indicates to reach a concentration of 100 µM.

Vector	Promoter	Antibiotic resistance	Tag	Cleavable with	Size (bp)	Restriction endonuclease used	Supplier
pProEX-Hta	Trc	Ampicillin	N-His (6H)	TEV	4779	NcoI/XhoI for wHTH cloning BamHI/XbaI for hMCM8 cloning	Gibco-BRL
pETM 30	T7	Kanamycin	6H-GST	TEV	6346	NcoI/XhoI for wHTH cloning	EMBL
pETM 41	T7	Kanamycin	6H-MBP	TEV	5389	NcoI/NotI for hMCM8 cloning	EMBL
pETSumo / CAT	T7	Kanamycin	6H-Sumo	Sumo protease	6307	Restriction Free cloning	Invitrogen

Table 2.1 List of vector used for the cloning of wHTH , hMCM9 and hMCM8 construct. The vectors used and their characteristics: promoter, antibiotic resistance, fusion tag and protease to use, the size and their supplier. Finally the restriction enzymes that will be used for the cloning.

2.1.3 Polymerase chain reaction (PCR).

DNA sequences were amplified by PCR, following a standard PCR protocol which is described in Appendix 4 (Protocol A). The plasmid used as template in the PCR reaction are pET21-*Mth*MCM kindly provided by Dr J.P. Gautier (, Schechter *et al*, 2000), pET19-*Sso*MCM (Carpentieri *et al*., 2002) and pTrcHisC-hMCM8 provided by Dr. F.M. Pisani and a pTrcHisa-hMCM8 where the NcoI site has been eliminated by site-directed mutagenesis (see Section 2.1.5). For the amplification of plasmid PCR targets, the plasmid templates used in the PCR reactions were previously prepared using the QIAprep Spin Miniprep kit (Qiagen), according to the manufacturer's instructions.

2.1.4 Preparation of DNA constructs

The amplified PCR products were analysed by agarose gel electrophoresis (1% agarose in TBE Buffer) and 95 µl of the PCR reaction were purified using the QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions and eluted in 50 µl MilliQ water.

Name of primer	Sequence (5' to 3')	Restriction site	Gene name and species
MthHTH NcoI	AAATTTCCATGGGCGAGACAGGCAAG	NcoI	*Mth*MCM
MthHTH XhoI	GGGTTTCTCGAGTCAGACTATCTT	XhoI	*Mth*MCM
SsoHTH NcoI	AAATTTCCATGGGCACAATAATGACTG	NcoI	*Sso*MCM
SsoHTH XhoI	GGGTTTCTCGAGTCAGACTATCTTAAG	XhoI	*Sso*MCM
SsoE605 NcoI	GGGTTTCCATGGGCGAAAGTGGAAAAATA	NcoI	*Sso*MCM
M1 BamH1	CCTCTGTCCCAAGCGGATCCGATGAATGGAGAG	BamH1	hMCM8
S67 BamH1	ACCCCACAGGATCCGATGTCAACATTGGATCG	BamH1	hMCM8
S373 BamH1	AGGACAGAAGGATCCGATGTCTGAGGATGGGTG	BamH1	hMCM8
P203 BamH1	GGGTTTGGATCCGATGCCTTTGACACAGCTC	BamH1	hMCM8
G380 BamH1	GGGTTTGGATCCGATGGGAATGTTGATGGAG	BamH1	hMCM8
Q765 BamH1	GGGTTTGGATCCGATGCAGCATGGTTCTGGA	BamH1	hMCM8
End Xba	GGTGAATCTAGATTACATAGTTTGAAGCTGGTAAAC	Xba	hMCM8
S372 Xba	GGGTTTTCTAGAATCACTCTTTGTTTTCTGTCC	Xba	hMCM8
Q368 Xba	GGGTTTTCTAGAATCCTGTCCTTTGCTATT	Xba	hMCM8
S764 Xba	GGGAAATCTAGAATCGGATCGCTCAAAATGTAG	Xba	hMCM8
MBPS67F	GGGTTTCCATGGCATCAACATTGGATCG	NcoI	hMCM8
MBPP203F	GGGTTTCCATGGCACCTTTGACACAGCTC	NcoI	hMCM8
MPBS764 not	GGGTTTGCGGCCGCATCGGACTGCTCAAAATCTAG	NotI	hMCM8
MPBEnd not	GGGTTTGCGGCCGCTTACATAGTTTGAAGCTGGTA	NotI	hMCM8
RFSumoM1	ATTGAGGCTCACAGAGAACAGATTGGTATGAATAGCGATCAAGTTACACTG		hMCM9
RFSumoL109	ATTGAGGCTCACAGAGAACAGATTGGTGGTCTGGTGAGGGAACACATACCTAAAA		hMCM9
RFSumoI278f	ATTGAGGCTCACAGAGAACAGATTGGTATCATCATGGATGAGGAGGTCCAA		hMCM9
RFSumoQ627	TTTGCGCCGAATAAATACCTAAGCTTGTCTTTACTGCTCTCCAGGGTTTTCAGGAA		hMCM9
RFSumoP665	TTTGCGCCGAATAAATACCTAAGCTTGTCTTTATGGTTGGGATTGGTGCACACTCTG		hMCM9
RFSumoI278r	TTTGCGCCGAATAAATACCTAAGCTTGTCTTTAGATCCCTGGGGACTGCTCATTATT		hMCM9

Table 2.2 List of primer used for the cloning. The sequences underlined correspond to the restriction site sequence, whereas the red sequences of hMCM9 correspond to the complementary sequence of the pETSumo plasmid that will be used for the cloning using a Restriction Free (RF) cloning technique.

Restriction endonuclease digestion of DNA

All restriction endonuclease digestions were performed using restriction enzymes, Buffers and BSA (required for some digestions, according to the manufacturer's instructions) from New England Biolabs.

Newly synthesized PCR products were digested with their respective restriction enzymes, in order to produce DNA fragments (insert DNA) to be used in subsequent ligation reactions. Typically, 48 µl of the purified PCR products were used for restriction endonuclease digestion. The reaction contained 3 µl of restriction enzyme for single digestions (one enzyme) and 2 µl of each restriction enzyme for double digestions (two different enzymes), 10x Buffer and BSA when required. Deionised water was added to a final volume of 60-70 µl and the reaction was incubated for 3 hr at 37ºC.

Restriction digestion of existing plasmids was performed to produce either vector (V) or insert (I) DNA fragments to be used in subsequent ligations. Typically, 8-10 µg of plasmid DNA were used for each digestion, in a total volume of 50-60 µl for single restriction enzyme digestions and 60-70 µl for double digestions. The reaction contained 4 µl of restriction enzyme for single digestions and 2.5 µl of each restriction enzyme for double digestions, 10x Buffer and BSA when required. The reaction was incubated for 2½ hr at 37ºC. All digested DNA fragments were subjected to agarose gel electrophoresis (1% agarose in TBE Buffer) and purified from agarose slices using the QIAquick Gel extraction kit (Qiagen), according to the manufacturer's protocol. The purified DNA fragments were eluted in 40 µl BufferMilliQ water.

Ligation of plasmid vector and insert DNA

Ligation reactions for DNA to be inserted into plasmid vectors were performed using T4 DNA ligase (NEB), according to the manufacturer's protocol. The amount of total DNA (insert and vector) used for each ligation reaction was 160-200 ng. The molar ratio of vector to insert DNA was typically 1:5 and the final reaction volume was 10 µl, including 1 µl T4 DNA ligase (NEB). The ligation reaction mixture was

incubated for overnight at 16°C and then either used directly for the transformation of competent cells or stored at -20ºC.

Transformation of bacteria with plasmid DNA

E. coli DH5α chemically competent cells (Stratagene), prepared according to protocol B (Appendix 4) were used for maintaining the plasmids. For the transformation, 1 µL of DNA plasmid at ~25 ng/µL was added to a 25 µL aliquot of DH5α cells. The mixtures were incubated on ice for 30 minutes, heat shocked at 42°C for 45 seconds and incubated on ice for 2 additional minutes. LB (Luria Bertani) medium, preheated to 42°C, was added to the mixture tubes to a final volume of 250 µL. The transformation reactions were then incubated for 1 hour at 37°C, shacking at 225-250 rpm. A volume of 100 µL of each reaction was streaked on LB agar plates containing either kanamycin or ampicillin depending on the resistance gene carried by the vector, at a final concentration of 50µg/µL and 100 µg/µL, respectively.

Preparation of plasmid minipreps

Single colonies from agar plates were inoculated into 5 mL LB broth medium supplemented with the appropriate antibiotics and the cultures grown for 16 hours at 37°C with shaking at ~225 rpm. The cells were pelleted by centrifugation at 3,500 g for 15 minutes at 4°C. Plasmid DNA was purified from cell pellets using the QIAprep Spin Miniprep kit (Qiagen), according to the manufacturer's protocol, eluted with 30 µL of MilliQ H2O (Qiagen) and stored at -20°C. The typical yield of plasmid DNA was ~100 ng/µL, as checked spectrophotometrically by measuring the absorbance at a wavelength of 260 nm (Table 2.3).

Construct name	Vector	Enzyme used	Sequenced
*Mth*WH	pProExHta	NcoI/XhoI	Yes
*Sso*WH	pProExHta	NcoI/XhoI	Yes
*Sso*E605	pProExHta	NcoI/XhoI	Yes
*Mth*GSTWH	pETM 30	NcoI/XhoI	Yes
*Sso*GSTWH	pETM 30	NcoI/XhoI	Yes
hMCM8_1	pProExHta	BamH1/Xba1	Yes
hMCM8_2	pProExHta	BamH1/Xba1	Yes
hMCM8_3	pProExHta	BamH1/Xba1	No
hMCM8_4	pProExHta	BamH1/Xba1	No
hMCM8_5	pProExHta	BamH1/Xba1	Yes
hMCM8_6	pProExHta	BamH1/Xba1	Yes
hMCM8_7	pProExHta	BamH1/Xba1	Yes
hMCM8_8	pProExHta	BamH1/Xba1	Yes
hMCM8_9	pProExHta	BamH1/Xba1	Yes
hMCM8_10	pProExHta	BamH1/Xba1	Yes
hMCM8_11	pProExHta	BamH1/Xba1	Yes
hMCM8_12	pProExHta	BamH1/Xba1	Yes
hMCM8_13	pProExHta	BamH1/Xba1	Yes
hMCM8_14	pProExHta	BamH1/Xba1	Yes
hMCM8_15	pProExHta	BamH1/Xba1	Yes
hMCM8_16	pProExHta	BamH1/Xba1	Yes
hMCM8_17	pProExHta	BamH1/Xba1	Yes
MBPhMCM8_7	pETM41	NcoI/NotI	No
MBPhMCM8_8	pETM41	NcoI/NotI	Yes
MBPhMCM8_11	pETM41	NcoI/NotI	Yes
MBPhMCM8_12	pETM41	NcoI/NotI	Yes

Table 2.3 List of constructs used for this work. Here are listed the constructs produced, the vector where have been insert and with enzymes have been used for the cloning.

2.1.5 Site-directed mutagenesis

Site-directed mutagenesis was used to mutate the R87 in Ala, Glu or Leu, and to mutate S342 and K344 in two Ala residues in *Mth*MCM; a silent point-mutation is performed in hMCM8 to remove a NcoI cutting site in position 1597 on the clone. All reactions were performed with the QuickChange XL Site-Directed Mutagenesis Kit (Stratagene), following the manufacturer's instructions. The method utilises double-stranded plasmid DNA as a template, and a pair of complementary oligonucleotide primers, incorporating the desired mutation. The template plasmid DNA, as well as the primer pair used for the construction of each mutant expression vector are listed in Table 2.4.

Name of mutant	Template DNA plasmid	Primer F (5'→3)	Primer R (5'→3')
*Mth*MCM R87A	pET21-*Mth*MCM	R87Af GTGGACCTCAACATAGCGTT CAGCGGGATCAG	R87Ar CTGATCCCGCTGAACGCTAT GTTGAGGTCCAC
*Mth*MCMR87E	pET21-*Mth*MCM	R87Ef GAATGTGGACCTCAACATAG AATTCAGCGGGATCAGCAAC	R87Er GTTGCTGATCCCGCTGAATTC TATGTTGAGGTCCACATTC
*Mth*MCMR87L	pET21-*Mth*MCM	R87Lf GTGGACCTCAACATACTGTT CAGCGGGATCAG	R87Lr CTGATCCCGCTGAACAGTAT GTTGAGGTCCAC
*Mth*MCMS342A	pET21-*Mth*MCM	S342Af CAGGGGGATATACACCGCCG GTAAGGGTACCTCAG	S342Ar CTGAGGTACCCTTACCGGCG GTGTATATCCCCCTG
*Mth*MCMK344A	pET21-*Mth*MCM	K344Af GATATACACCAGCGGTGCAG GTACCTCAGGGGTC	K344Ar GACCCCTGAGGTACCTGCAC CGCTGGTGTATATC
*Mth*MCMS342A/ K344A	pET21-*Mth*MCMS 342A	S342AK344Af CAGGGGGATATACACCGCCG GTGCAGGTACCTCAG	S342AK344Ar CTGAGGTACCTGCACCGGCG GTGTATATCCCCCTG
pTrcHisA-hMCM8 no NcoI	pTrcHisA-hMCM8	hMCM8 PM F GCCTTGTTGGAAGCTATGGA GCAGCAAAG	hMCM8 PM r CTTTGCTGCTCCATAGCTTCC AACAAGGC

Table 2.4 List of the primers used for the Quickchange system. In red are showed the mutations introduced.

Mutated constructs were generated using a modified PCR reaction protocol, which is detailed in Appendix 4 (Protocol C). Following temperature cycling, the reactions were digested with 1 µl of the DpnI restriction enzyme at 37ºC, for 1.5 hr, so as to digest the methylated parental DNA. A 4 µl aliquot from the reaction mixture was then used to transform 45 µl DH5α competent cells (Stratagene) and the cell mixture

was streaked on LB agar plates, containing ampicillin. After overnight incubation at 37ºC, preparation of plasmid DNA from isolated colonies proceeded as previously. Mutant DNA constructs were sequenced (BMR Genomics, Padova) to confirm the introduction of the correctly mutated bases. For the double mutant a second PCR is performed with a second couple of primer.

2.1.6 RT PCR

The production of the cDNA of hMCM9 was performed using the M-MLV Reverse Transcriptase (Invitrogen). The total RNA of different proliferating cell line (Table 5) was used and the first strand (FS) reaction synthesis was processed with a $d(T)_{16}$ as recommended by the producer of the Transcriptase (see appendix 4, protocol E). The second strand synthesis consisted of a PCR reaction (see section 2.1.3) using FS as template and carried out with the primers listed in Table 2.5.

Cell line	**Tissue**
MCF10A	Mammary gland
HepG2	Liver hepatocellular carcinoma
HeLa	Cervical cancer
A549	Lung adenocarcinoma
SK-N-AS	Neuroblastoma

Table 2.5 List of the cell line used in this work.

2.1.7 Restriction Free Cloning

Restriction Free (RF) cloning methodology is part of the Ligation Independent Cloning (LIC) processes which, were developed for the directional cloning of PCR products without restriction enzyme digestion or ligation reactions. The RF procedure is a universal cloning method allowing a precise insert of a DNA fragment into any desired position within a circular plasmid. The RF cloning is a twosteps process (Figure 2.1). Initially, the gene of interest is amplified along with primers containing complimentary sequences (25-30 bp) to the flanking sites of integration in the target vector. The two strands of the PCR product obtained in the first stage are used as mega primers for the second stage of amplification. In the second stage the PCR product and the target vector are mixed and following amplification reaction the

gene of interest is integrated into the circular vector to a pre-defined positionThe first PRC step is performed as described in Appendix 4 (protocol A). The PCR are cleaned using the Qiaquick PCR purification kit (Qiagen) as manufacture suggests and the DNA eluted in 40 µl of deionised water.

This DNA will be used as megaprimer for the RF PCR (Appendix 4, protocol F), which is a linear amplification and the 10 µl of product digested with Dpn1. These are then used to transform 90 µl of DH5α.

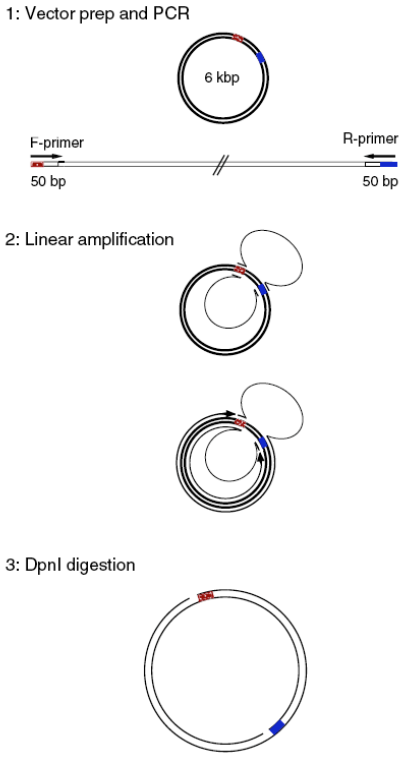


Figure 2.1 Schematic representation of the restriction-free (RF) cloning (from van den Ent F 2006,)

2.2 Protein preparation

2.2.1 *Mth*MCM and mutants expression and purification.

A construct containing the wt *Mth*MCM gene cloned into a pET21b (Novagen) expression vector (a gift from J.P. Gautier, Columbia University) was used to produce the full-length protein with a non-cleavable C-terminal His-tag. A construct containing a *Mth*MCM gene lacking subdomain A cloned into a pET21a expression vector (a gift of Zvi Kelman, University of Maryland), was used to produce a C-terminal His-tagged protein missing residues 1-91. All the site-direct mutants constructs were obtained as described previously.

Protein expression

Transformation of BL 21 strain of *E coli* cells was performed using 1 μl of plasmid miniprep added to a 25 μl aliquot of competent cells. The transformation proceeded following a protocol similar to that outlined in Paragraph 2.1.4, and cells were plated on LB agar plates containing ampicillin. Few colonies of transformed E. coli competent cells were grown at 37ºC overnight in 80 ml of LB medium containing ampicillin. The overnight culture was used to inoculate a 20-fold volume of fresh LB medium supplemented with ampicillin (i.e. 40 ml per litre, for a total of 4 liters). The culture was grown at 37ºC with shaking at 225 rpm until an OD600 of approximately 0.8 was reached, at which point protein expression was induced with 0.5 mM IPTG and cells were grown for an additional 4 hours. Cells were harvested by centrifugation and the pellet was stored at -80ºC.

Purification consisted of nickel-affinity, followed by heparin-affinity chromatography and gel filtration chromatography. All protein purification steps were carried out at 4ºC, using an ÄKTA Fast Protein Liquid Chromatography (FPLC) apparatus (GE Healthcare). The data were collected using the Unicorn Data system (ÄKTA FPLC) and edited using Microsoft Excel.

Protein extraction

Each cell pellet was thawed and re-suspended in 30 ml of re-suspending Buffer (20 mM Tris-HCl pH 7.9, 500 mM NaCl, 10 mM imidazole, 10% glycerol) supplemented with 1 mg/ml lysozyme and 1 mM AEBSF protease inhibitor. The suspension was incubated at 4°C under magnetic stirring for 30 min to allow the lysozyme to act, followed by sonication on ice. Twenty 15 sec pulses with 45 sec in ice between each pulse were performed at 40% amplitude (Soniprep 150). The cell lysate was clarified by centrifugation for 45 min at 26,000 g and the supernatant was filtered through a 0.22 μm filter before applying to affinity columns.

Protein purification

For the purification of *Mth*MCM and its mutants, a HiTrap chelating column (5 ml column volume (CV), GE Healthcare) was used. The column was charged with Ni^{2+} ions by 4 CV of 0.1 M $NiSO_4$ (charge solution), washed with 3 CV deionised water and subsequently equilibrated with 5 CV of Nickel Buffer A (20 mM Tris-HCl pH 7.9, 500 mM NaCl, 10 mM imidazole, 10% glycerol). The filtrate was then loaded on the column, followed by an extensive wash with Nickel Buffer. The column was washed with 3 CV of 50 mM imidazole in Nickel Buffer and then a linear gradient from 50 mM to 500 mM imidazole was applied to elute the protein of interest. Fractions containing the target protein product, as assessed by SDS-PAGE, were pooled and dialysed overnight in Heparin Buffer A (20 mM Hepes, 50 mM NaCl, 10% glycerol). The dialysed sample was loaded onto a Heparin HiTrap column (5 ml GE Healthcare) equilibrated in Heparin Buffer. The bound protein was eluted with a 3 CV linear gradient from 50 mM to 1 M NaCl in the same Buffer. This step was critical in removing contaminating DNA. The fractions containing the protein were pooled and concentrated on Amicon Ultra 15 (Millipore) centrifugating the tubes as recommended by the producer. The sample was then spun down and filtered to remove any precipitate and loaded on a Tricorn Superose 6 10/30 gel-filtration column (GE Healthcare) in a Buffer composed by 30 mM Hepes pH 7.5, 150 mM NaCl (GF Buffer A). Fractions corresponding to the size of a double ring were collected.

2.2.2 Archaeal C-terminal Winged Helix (WH) expression and purification.

Protein expression

As for the *Mth*MCM, also WH construct were transformed in BL 21 *E. coli* cell using 1.5 µl of the appropriate plasmid (pProExHta-*Mth*WH, pPRoExHta-*Sso*WH, pProExHta-*Sso*E605). The transformation proceeded following a protocol similar to that outlined in Paragraph 2.1.4, and cells were plated on LB agar plates containing ampicillin. Few colonies of transformed *E. coli* competent cells were grown at 37ºC overnight in 80 ml of LB medium containing ampicillin. The overnight culture was used to inoculate a 20-fold volume of fresh TB medium supplemented with ampicillin (i.e. 40 ml per litre, for a total of 6 liters). The culture was grown at 37ºC with shaking at 225 rpm until an OD600 of approximately 1 was reached, at which point protein expression was induced with 1 mM IPTG and cells were grown for an additional 5 hours. Cells were harvested by centrifugation and the pellet was stored at -80ºC.

Protein purification

The protein extraction was performed as already described for the *Mth*MCM protein, as well as the first step of the purification using a 5ml HiTrap chelating column charged with Ni^{2+} ions. Fractions containing the target protein product, as assessed by Tris-Tricine SDS-PAGE, were pooled and dialysed overnight in TEV cleavage Buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.5 mM EDTA and 10% glycerol). His-tagged rTEV protease was produced as described in Protocol D (Appendix 4) and added to the dialysed protein (a total of 1 mg of TEV protease for ~50 mg of His-tagged protein) kept in a reducing environment by βME at 2 mM. The endoprotease digestion was typically carried out overnight at 4ºC. The digested sample was loaded onto the Ni^{2+} column to remove the His-tagged TEV and any trace of uncleaved target protein (TEV cleanup step), after that the sample was concentrated and further purified using size-exclusion chromatography with a HiLoad Superdex-75 16/60 gel-filtration column (GE Healthcare). Prior to use, the

column was pre-equilibrated with ≥2 CV of GF Buffer A (30mM Hepes pH 7.5, 150 mM NaCl) or in a NMR filtration Buffer (30 mM phosphate Buffer pH 6.5, 150 mM NaCl) according to the manufacturer's instructions. Each run was performed at a flow rate of 0.5 ml/min, and a 5ml loop was used to inject the sample onto the column. Sample were typically purified by isocratic elution with 1 CV (~125 ml) of gel-filtration Buffer. During elution, 0.5 ml fractions were collected and assessed by Tris-Tricine SDS-PAGE for protein purity and concentration. The protein were then concentrated to be treated as described in section 2.6.

2.2.3 hMCM8 protein expression and purification

Small-scale expression and solubility assays

Before attempting large-scale expression of any newly synthesized DNA construct the expression of the target protein was verified by the analysis of whole-cell extracts, as described below. Once expression of the target protein was verified, determination of the protein solubility and optimization of growth conditions were carried out using the small-scale batch purification procedure described in the following Section.

Analysis of whole cell extracts and Small-scale batch purification

Single colonies of *E. coli* cells harbouring the plasmid(s) being tested were used to inoculate 10ml of LB or TB media (containing appropriate antibiotics), in 50 ml tubes to allow a sufficient aeration during growth. Different bacterial strains were used for the test (Table 2.6) to increase the chances of success.

***E. coli* (DE3) Strain name**	**Resistence**	**Provider**
BL 21		Stratagene
BL 21 codon+ RIP	chloramphenicol	Stratagene
Bl 21 codon + RIPL	chloramphenicol	Stratagene
BL 21 Gold		Stratagene
BL 21 Star		Invitrogen
Rosetta 2	chloramphenicol	Novagen
B 834		Novagen

Table 2.6 *E. coli* (DE3) strain tested. The strain cell used for the small scale expression and their antibiotic resistence.

The cultures were grown at 37ºC, shaking at 225 rpm until reaching an optical density of 0.6-0.7 at 600 nm for the cells in LB broth, or 2-2.5 OD for those in TB broth. Protein expression was induced by addition of IPTG at different concentration ranging from 0.1 to 1 mM, at different temperature and times (37°C for 5 hours, 25°C overnight/24 hours, 18°C for 48 hours). Autoinduction media were also tested at 25°C for 24 hours. Aliquots were removed from the culture prior to addition of

IPTG (pre-induction sample) and prior to harvesting (post-induction sample). The aliquots were pelleted by centrifugation at 16,000 g for 2 min at 4ºC, the supernatant was removed and each cell pellet was resuspended in 20 µl 2x SDS Sample Buffer and heated at 95ºC for 10 min to fully lyse the cells and denature the proteins. The pre-induced and post-induced SDS-whole cell lysates were analysed by denaturing SDS-polyacrylamide gel electrophoresis (SDS-PAGE, Section 2.12.3), followed by Instant blue staining to verify expression of the target recombinant proteins by comparing the pre- and post-induction whole-cell extracts from each culture.

Cell pellets were thawed on ice and resuspended in 1 ml BufferNickel Buffer A: 20 mM Tris-HCl pH 7.9, 500 mM NaCl, 10 mM imidazole, 10% glycerol, supplemented with 1 mg/ml lysozyme and 1 mM PMSF protease inhibitor and the suspension was incubated on ice for 30 min. Cell lysis was performed by 3-4 cycles of freeze (-80°C) and thawed (+37°C). The lysates were centrifuged for 30 min at 26,000 g to remove cellular debris and the supernatant (soluble fraction) was transferred to a fresh 1.5 ml microcentrifuge tube.

Ni-NTA fastflow resin (Qiagen) was used for the analysis of His-tagged target proteins. The resins were prepared according to the manufacturers' instructions as 50% slurries Nickel Buffer A prior to use. A 50 µl aliquot of resin was added to the supernatant and the tube was incubated with gentle rotation at 4°C for 45 min. The suspension was centrifuged at 500 g for 5 min to sediment the resin. The supernatant was removed and the resin was washed with 1 ml Buffer by resuspending with a pipette, centrifuging and discarding the supernatant, for a total of three washes. The third wash was performed with Nickel Buffer A supplemented with 50 mM imidazole, to reduce unspecific binding of proteins to the resin. 25 µl of 2x SDS Sample Buffer were added to the sedimented resin and the suspension was heated at 95ºC for 10 min to elute the bound proteins from the resin. The sample was centrifuged at 16,000 g for 1 minute and 15-20 µl of the resin was analysed by SDS-PAGE with Instant blue staining.

Once the optimal condition for the each construct were found, large scale expression experiments (i.e. 6 liters culture) were performed.

Protein extraction

The cell pellet from 1 L of culture was thawed and resuspended in 20 mL (for cells grown in LB) or 40 mL (for cells grown in TB) of affinity binding Buffer (MBP Buffer or Nickel Buffer depending on the protein to be purified), supplemented with 1 mg/mL of lysozyme (Sigma) and 2 mM of the protease inhibitor AEBSF (4-(2-aminoethyl) benzenesulfonyl fluoride, Sigma). The resuspended solution was incubated on ice for 20 minutes, stirring with a magnetic bar, in order to allow the lysozyme to act. The suspension was then sonicated on ice with 30 short pulses of 15 seconds each, at 40% amplitude (Soniprep 150). The cell lysates were clarified by centrifugation for 45 minutes at 26,000 g and the supernatant filtered through a 0.22 mm cut-off filter before applying to affinity columns.

Purification of MBP-tagged proteins

The protein expressed as fusion with Maltose-binding protein (MBP) were subjected to a first purification step using a MBPtrap column (GE Healthcare, with a volume of 1 or 5 ml, depending on the volume of culture). After a rinse in deionised water the column was pre-equilibrated in MBP Buffer (30 mM Tris pH 7.4, 300 mM NaCl, 5% Glycerol, 1 mM EDTA). The clarified cell extract was loaded onto the column and the unbound proteins were washed out with MBP Buffer until the UV baseline was reached. The elution of the protein of interest was carried out with 5 CV of MBP Buffer containing 10 mM Maltose. Then the protein was concentrated and injected onto a size exclusion column (GE Healthcare, Tricorn Superose 6 10/30) using 200 or 500 µl loop. The Buffer used contains 30 mM Hepes pH 7.5, 150 mM NaCl.

Purification of His-tagged proteins

For all the expressed proteins that possessed a hexa-histidine tag, the first purification step was metal-affinity chromatography using a HiTrap Chelating FF crude column (with a volume of 1ml or 5 mL depending on the volume of culture, GE Healthcare). The column was washed with deionised filtered water, charged with ions by 0.1 mM $NiSO_4$ or $CoSO_4$ and then processed as already described prior to equilibration with 5 CV of Nickel Buffer (20 mM Tris-HCl, pH 7.9, 0.5 M NaCl,

10% glycerol). The clarified cell extract was then loaded onto the column and the unbound protein washed with Nickel Buffer until a stable UV baseline was recorded. The column was washed with a Nickel Buffer containing 10 mM ATP/$MgCl_2$ to remove contamination by GroEL, and a second washing step with high salt concentration (1 M) was added to remove nucleic acid contamination. The column was then washed with 50 mM imidazole to remove proteins that bound non-specifically. The His-tagged protein was finally eluted by applying a linear gradient from 50 mM to 500 mM imidazole in Nickel Buffer. The eluted fractions were collected and their content analysed on a SDS-PAGE gel and visualised using Instant Blue stain (Biosense).

For the hMCM8_2 construct the second purification step consisted in a Heparin HiTrap column (5 ml, GE Healthcare), equilibrated in Heparin Buffer C (50 mM Tris pH 7.4, 500 mM NaCl, 10% Glycerol). The elution of the protein was carried out applying a linear gradient of NaCl, from 500 mM to 2 M.

A purification step involving size-exclusion chromatography was used for the hMCM8_14 construct using a HiLoad Superdex 200 (~120 mL CV, GE Healthcare) gel-filtration column. The column was washed with 2 CV of deionised water and then equilibrated with 2 CV of the appropriate size exclusion chromatography Buffer (30 mM Hepes pH 7.5, 300 mM NaCl, 10% Glycerol; GF Buffer B). Proteins were concentrated at 4°C, using the appropriated molecular weight cut-off concentrators (Millipore) according to the manufacturer's instructions, and injected onto the column using a 5 mL loop. The samples were purified by 1.5 CV isocratic elution and the 0.5 mL fractions assessed by SDS-PAGE.

2.3 SDS-PAGE

2.3.1 Tris-Glycine-SDS

Protein samples were analysed by electrophoresis under denaturing, reducing conditions using discontinuous SDS-polyacrylamide gels (4% for the stacking gel and a range between 8-15% for the separating gel, depending on the protein of interest). Samples were prepared by addition of 2x SDS sample Buffer and heating at

95ºC for 5 min. Electrophoresis was carried out at a constant voltage (200V) for 45-55 min in 1x Tris-glycine-SDS Buffer. Gels were stained with Instant blue (Biosense) for visualisation.

2.3.2 Tris-Tricine-SDS

Small protein samples were analysed by electrophoresis using discontinuous SDS-polyacrylamide gel (4% in 0.75 M Tris pH 8.5 fo the stacking and 16% in 1M Tris, pH 8.5 for the separating). The samples was prepared as described above and electrophoresis was carried out at a constant voltage (150V) for 50 min using 1X Cathode Buffer and 1X Anode Buffer (recipe Appendix 3). Gels were either stained with Instant blue (Biosense) for visualisation.

2.4 Determination of theoretical protein parameters

Theoretical values for the isoelectric point (pI), molecular weight (MW) and extinction coefficient at 280 nm of the different proteins used in this study were calculated using the Protparam tool at the Expasy Server: http://www.expasy.ch/tools/protparam.html.

2.5 Determination of protein concentration

Protein concentrations were determined spectrophotometrically by measuring the UV absorbance at a wavelength of 280 nm (NanoDrop Spectrophotometer, Thermo Scientific) and applying the Lambert-Beer law with the theoretical extinction coefficient calculated as indicated above.

2.6 Protein crystallization

2.6.1 Crystallization screening

Crystallization trials were carried out using the sitting drop vapour-diffusion method. Peak fractions from the gel-filtration purification step were pooled and concentrated using 10 kDa molecular weight cut-off concentrators (Amicon 15 Ultra Millipore) to a final protein concentration of 10 to 90 mg/ml (depending on the protein solubility).

The sample was filtered through a 0.22 μm microcentrifuge filter (Centrisart-C4, Sartorius) and immediately used for crystallisation using the Mosquito liquid handling robot (TTP LabTech's). Crystallisation drops (sitting drop) contained either 1:1 or 2:1 volume ratios of protein to reservoir respectively, with final drop volume of 200 or 300 nL. Initial crystallisation trials were set up at room temperature, at 4°C or 37°C using commercially available screens: Hampton Crystal Screen 1 and 2, Hampton PEG/Ion and Natrix, Index Screen, SaltRX (all from Hampton Research), Wizard 1, 2 (Emerald Biosystems), Classic, Classic II, pH clear, pH clear II, comPAS, PACT and JCSG+ (Qiagen).

2.6.2 Optimization

Optimisation of the crystallisation conditions involves the manipulation of the crystallisation phase diagram with the aim of leading crystal growth in the direction that will produce the desired results. Crystal growth can also be controlled by separating the processes of nucleation and growth. This can be achieved by bypassing the nucleation zone by inserting crystals, crystal seeds or other nucleants directly into the metastable zone.

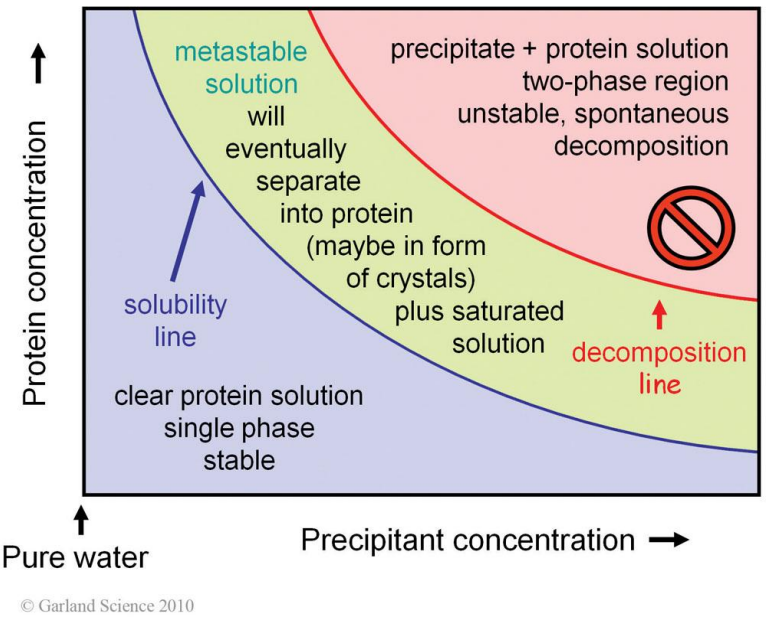


Figure 2.2 A schematic phase diagram for a protein crystallization experiment

At high protein and high precipitant concentrations, precipitation occurs as shown. At low protein and low precipitant concentration, the solution remains clear. Near the bottom of the precipitation zone, there is an area where protein crystal nuclei form and grow (red curve, Figure 2.2). Below this is an area often called the metastable zone where crystals will grow, but no nucleation occurs. A preformed crystal (e.g. from the nucleation zone, or from other screening) placed into this area it will grow.

If the nucleation zone is small, crystallization may be a rare event. By adding seeds to the experiments, crystals can grow in experiments that never leave the metastable zone. The crystal transfer can be performed with different technique.

To optimse crystal growth and morphology I used the following:

Streak-seeding or touch-seeding: a natural fiber is used to touch a crystal and transfer crystal particles in a 2 µl drop containing a 1:1 ratio protein/reservoir.

Micro-seeding: a crystal is smashed and the microcrystals are transferred by manual pipetting, upon serial dilutions.

In both technique the microcrystals are transferred in the same reservoir that gave positive results, but scaling up the final volume of the reaction.

Cross-seeding: a crystal is smashed and the diluted microcrystal are used in 96 well screening. The drops are obtained in the same way described for the preliminary screening, but using a ratio of crystal-solution/protein/reservoir of 1:2:1.

2.7 Protein biochemical assays

Preparation of radiolabelled oligonucleotides

DNA oligonucleotides were used for the nucleic acid binding assays; sequences of all the probes tested are listed in Table 2.7. DNA oligonucleotides were purchased from sigma Genosys. DNA fork was obtained by putting a poly(dT) tail at complementary ends of each oligonucleotides.. For the radiolabelling reaction, 1 µl of the probe at 10 µM concentration was mixed with 1.5 µl [γ32P]ATP (10 µCi/µL, 3000 Ci/mmol, Perkin Elmer), 1 µl T4 Polynucleotide Kinase (PNK, Roche), 2 µl

10× T4 PNK Buffer and RNase-free deionised water to a final volume of 20 µl. The mixture was incubated for 1 hour at 37°C. The radiolabelled oligonucleotides were then purified using BioSpin 6 desalting columns (Biorad) according to the manufacturer's protocol and eluted in Tris-HCl Buffer pH 8.0. To prepare the double stranded substrates, the labelled oligonucleotide was mixed with a 2 fold molar (for dsDNA) or 3 molar fold (for ForkDNA) excess of the cold complementary strand in presence of 0.3M NaCl before passing through the desalting column.

Name of primer	Sequence (5' to 3')	Final concentration
41mer f	AATCATAGATAGTATCTCCGTGCAAGATAATCACGAGTATC*	0.1 µM
41mer r	GATACTCGTGATTATCTTGCACGGAGATACTATCTATGATT	0.1 µM
61 mer Fork f	AATCATAGATAGTATCTCCGTGCAAGATAATCACGAGTATCTTTTTTTTTTTTTTTTTTTT*	0.1 µM
61 mer Fork r	TTTTTTTTTTTTTTTTTTTTGATACTCGTGATTATCTTGCACGGAGATACTATCTATGATT	0.1 µM

Table 2.7 Rabiolabeled oligonucleotides. List of the oligonucleotides used for the biochemical assays, alone or in combination, so as to obtain dsDNA and Fork substrates. * indicates the strand labeled with 32P γ-ATP.

2.7.1 Electrophoretic mobility shift assays (EMSA)

MthMCM proteins

Complexes formed between the proteins and DNA were detected by a gel mobility shift assay in reaction mixtures (10 µl) containing 25 mM Hepes-NaOH (pH 7.5), 50 mM Na acetate, 10 mM MgCl2, 1 mM ATP, 1 mM dithiothreitol, 100 µg/ml bovine serum albumin, 50 fmol labeled oligonucleotide, indicated amount of 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% bromphenol blue, 50% glycerol) was added to stop the reaction. Aliquots of the reaction mixture were electrophoresed for 1.5 h at 100V through a 4% polyacrylamide gel containing 6 mM magnesium acetate and 5% glycerol in 0.5X TBE (45 mM Tris, 45 mM boric acid, 1 mM EDTA).

hMCM8

Complexes formed between the proteins and DNA were detected by a gel mobility shift assay in reaction mixtures (10 µl) containing 25 mM Hepes-NaOH (pH 7.5), 25 mM Na acetate, 10 mM Mg acetate, 1 mM dithiothreitol, 100 µg/ml bovine serum albumin, 50 fmol labeled oligonucleotide and increasing amount of protein. The DNA binding is also tested in presence of nucleotides (ATP). After incubation at 37 °C or 25 °C for 60 min, 5 µl of 5X loading Buffer (0.1% xylene cyanol, 0.1% bromphenol blue, 50% glycerol) was added to stop the reaction. Aliquots of the reaction mixture were electrophoresed for 1.5 h at 100V through a 4% polyacrylamide gel containing 6 mM magnesium acetate and 5% glycerol in 0.5X TBE (45 mM Tris, 45 mM boric acid, 1 mM EDTA).

2.7.2 Helicase assay

MthMCM proteins

DNA helicase activity was measured in reaction mixtures (10 µl) containing 20 mM Tris-HCl (pH 8.5), 10 mM MgCl2, 2 mM dithiothreitol, 100 µg/ml bovine serum albumin, 5 mM ATP, 10 fmol of labeled DNA substrate, and the various MCM mutant proteins. After incubation at 60 °C for 1 h, reactions were stopped by adding 5µl of 5X loading Buffer (100 mM EDTA, 1% SDS, 0.1% xylene cyanol, 0.1% bromphenol blue, and 50% glycerol). Aliquots were loaded onto an 8% native polyacrylamide gel in 0.5X TBE and electrophoresed for 1.5 h at 200V. The helicase activity was visualized and quantitated by phosphorimaging.

hMCM8

DNA helicase activity was measured in reaction mixtures (10 µl) containing 25 mM Hepes-NaOH (pH 7.5), 25 mM Na acetate, 10 mM Mg acetate, 1 mM dithiothreitol, 100 µg/ml bovine serum albumin, 50 fmol labeled oligonucleotide and increasing amount of protein increasing amount of protein. After incubation at 37 °C or 25 °C for 60 min reactions were stopped and loaded as describe in the *Mth*MCM section.

2.7.3 ATPase assay

MthMCM proteins

ATPase activity was measured in reaction mixtures (10 µl) containing 25 mM Hepes-NaOH (pH 7.5), 5 mM MgCl2, 1 mM dithiothreitol, 100 µg/ml bovine serum albumin, 1.5 nmol of ATP containing 2.5 µCi of [γ-32P]ATP and X or Y pmol of proteins (as monomers) in the presence or absence of 50 ng of ssDNA. 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 and 0.5 M LiCl. The extent of ATP hydrolysis was quantitated by ImageQuant (GE Healthcare) analysis.

hMCM8

ATPase activity was measured in reaction mixtures (10 µl) containing 25 mM Hepes-NaOH (pH 7.5), 25 mM Na acetate, 10 mM Mg acetate, 1 mM dithiothreitol, 100 µg/ml bovine serum albumin the presence or absence of 50 ng of ssDNA. After incubation at 37 °C or 25 °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 and 0.5 M LiCl. The extent of ATP hydrolysis was quantitated by ImageQuant (GE Healthcare) analysis.

2.7.4 Treatment of the hMCM8 protein samples with nucleotide analogues and DNA substrates.

A protein sample at a concentration of 3.6 mg/ml was incubated in Buffer 25 mM Hepes-NaOH (pH 7.5), 150 mM Na acetate, 10 mM Mg acetate, 1 mM dithiothreitol, in the presence of 0.5 mM ATP, ssDNA or dsDNA. Incubation was carried out for 15 minutes at room temperature (1 hour in the case of DNA). The sample was then spun down and filtered to remove any precipitate and loaded on a Superdex 200 PC 3.2/30 gel-filtration column (GE Healthcare).

2.7.5 Topology footprinting assay

Topology footprint assays were carried out in 15 μl of reaction Buffer (50 mM Tris, pH 7.5, 1 mM DTT, 5 mM $MgCl_2$, 30% (v/v) glycerol) containing 150 ng of DNA (5.7 nM). The DNA was incubated for 15 min at 37°C with increasing amount of *Mth*MCM or with the same amount of ΔsA mutant, followed by the addition of 20 units of wheat germ Topoisomerase I. Reactions were stopped by incubating the reaction mixture with 5 μl of Stop Buffer (1 mg/ml Proteinase K, 100 mM EDTA, 1% SDS, 50% glycerol) for 15 min at 37°C. The samples were loaded on a 1.2% agarose gel in TBE 1× and electrophoresis was performed for 90 min at 11V/cm. Results were visualized after ethidium bromide staining.

2.8 Electroblotting to PVDF membranes for protein sequencing

Proteins separated via PAGE were transferred from a SDS-polyacrylamide gel to an PVDF membrane using the iBLot dry blotting system (Invitrogen). The electroblotting to the PVDF membranes was performed according to the guidelines provided by Proteomics Facility, University of Leeds (Leeds, UK).

2.9 Electron microscopy sample preparation and data collection

Samples were prepared by incubating for 10 mins at room T a solution containing both protein and dsDNA at a final concentration of 12.5 μg/ml in Buffer containing 30mM Tris HCl pH 7.5, 50 mM NaCl, 5mM $MgCl_2$, 5 mM β-mercaptoethanol. Negative-stain preparations were performed as follows: a sample of 4 μl was incubated for 1 min on glow-discharged continuous carbon coated formvar copper or formvar nickel grids. The grid was washed with a 40 μl drop of 2% uranyl-acetate solution, blotted and air-dried. Negative-stain EM of the wt and mutant nucleoprotein preparations was performed on either a Tecnai12 electron microscope (FEI) at an accelerating voltage of 120 kV, or a CM200 (FEI) field-emission gun electron microscope at an accelerating voltage of 200 kV. To enhance low-frequency components typical of the long stretches of dsDNA, a magnification of 11,500-

27,500× and 5-10 μm nominal defocus were used. For fibre diffraction studies, defocal pair micrographs were imaged on a CM200 electron microscope under low-dose conditions, at a magnification of 50,000× and at 0.5 and 5 μm nominal defocus.

3 Results and Discussion

In this chapter the results obtained during my PhD project will be discussed. The chapter will be organized in two different sections.

In the first part all the results of the work done with the archaeal proteins will be reported: the novel dsDNA/MCM interaction and the domain involved on it; the cloning, purification and initial biochemical characterization of the Glutamate switch (E switch) mutants and finally the work done with the C-terminal region of the archaeal MCM.

In the second part all the work aiming to obtain soluble human protein will be described, and particularly the cloning and the test expression of hMCM8 fragments and the characterization of the biochemical properties of two fragments encompassing the N-terminal domain and the catalytic core, respectively.

3.1 The Archaeal MCM proteins

3.1.1 *Mth*MCM/DNA: a novel interaction

Studies in vivo on *S. cerevisiae* described two types of dsDNA/MCM2-7 interaction: an "associated" interaction, sensitive at high salt washes, and a more stable "loaded" interaction (Browers et *al*., 2004). Whereas various canonical models for the mechanism of hexameric helicases involved a ring encircling the DNA and can therefore easily account for the stable "loaded" complexes, no structural evidence was available for the associated complexes.

Using Cryo-electron microscopy techniques A. Costa, in our laboratory, has carried out a 3D reconstruction of *Mth*MCM proteins bound to a dsDNA substrate of 5600 base pairs. Due the inability of the protein to load onto a long dsDNA segment in the absence of the appropriate loading factors, he managed to visualise a novel type of DNA/MCM interaction, with the dsDNA wrapping around N-terminal region of the ring, as opposed to the canonical binding of DNA within the central channel (Costa

et al., 2008; Figure 3.1). This interaction requires a conformational change of subdomain A (sA) located on the outer part of the MCM ring.

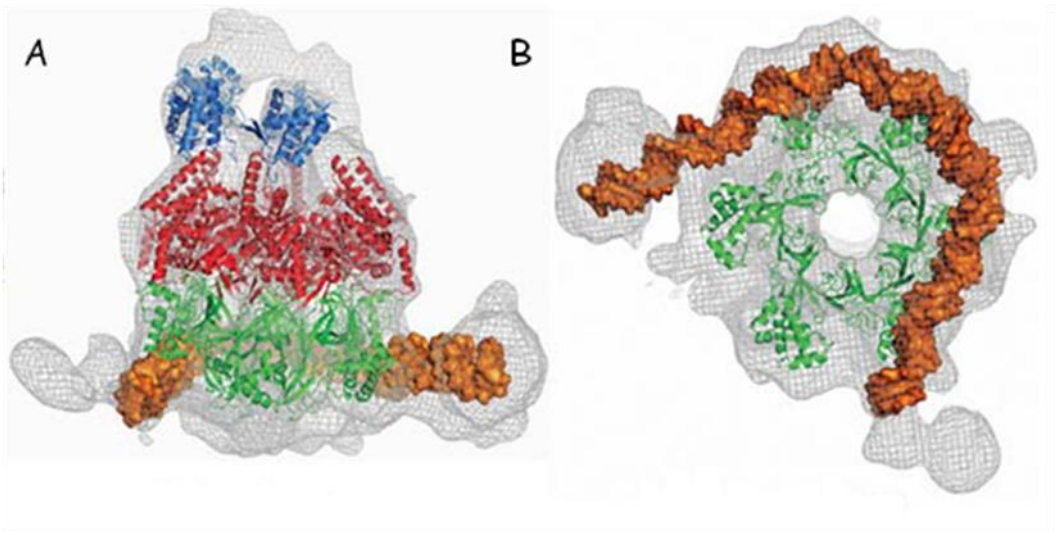


Figure 3.1 Cryo-EM 3D reconstruction of *Mth*MCM and long dsDNA. A) side and B) bottom view of the EM electron density map, fitted with the corresponding atomic models (N-term: green, AAA^+ : red, C-term: blue, dsDNA: orange). The N-terminal region is encircled by the dsDNA and a conformational change of subdomain A is required to account for the interaction.

We suggested that this model may provide a structural framework for the initial ("associated") interaction (Figure 3.2), prior to the loading and activation of the complex to function as a helicase at the fork (Costa *et al*., 2008).

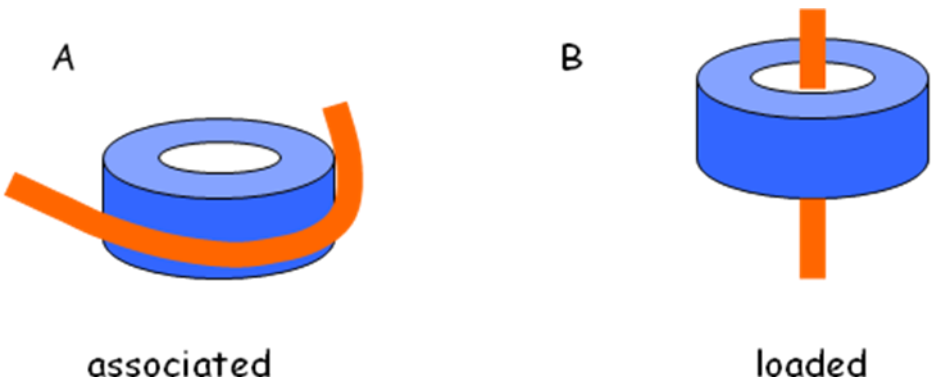


Figure 3.2 Schematic representation of two types of DNA/protein interaction. A) associated: the dsDNA wrapped around the protein; B) the protein is encircling the dsDNA, and the complex is therefore "loaded" ready for the helicase activity.

In parallel to the structural biology work, I carried out some biochemical studies on the *Mth*MCM system, to provide additional evidence for this proposal.

As the novel DNA/MCM interaction described is expected to have some effect on the dsDNA topology, I performed a topology footprinting assay, to test if incubation with *Mth*MCM affects the degree of writhing of a negatively supercoiled plasmid (pUC18) (Figure 3.3). The same assay was performed with a deletion mutant missing subdomain A (ΔsA), to assess whether this part of the N-term region is involved in this protein-DNA interaction.

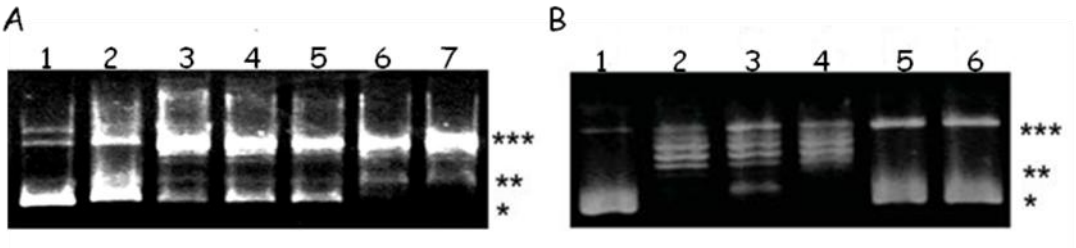


Figure 3.3 Topology footprint assay. Plasmid DNA relaxation at saturating concentration of Topoisomerase I and in the presence *Mth*MCM and in presence of ΔsA. In panel A the plasmid is incubated with increasing amount of *Mth*MCM. Lane 1: 5.7 nM pUC18; lane 2: pUC 18 incubated with 1.8 µM (as hexamer) of *Mth*MCM Lane 3: pUC18 incubated with 10 units of wheat germ Topoisomerase I. Lane 4-7; pUC18 incubated with increasing amount of *Mth*MCM (0.09 µM-1.8 µM) for 15 minutes, followed by a 60 min incubation with Topoisomerase I. In panel B Lane 1: 5.7 nM pUC18.. Lane 2: pUC18 incubated with 20 units of wheat germ Topoisomerase I. Lane 3: pUC18 incubated with 1µM of *Mth*MCM, for 15 minutes, followed by a 60 min incubation with Topoisomerase I. Lane 4: the same experiment as in lane 3, where the ΔsA mutant is used instead of *Mth*MCM. Lanes 5-6: pUC18 incubated with 1 µM of *Mth*MCM or ΔsA, respectively. Different topological states can be visualised: (*) negatively supercoiled DNA; (**) partially relaxed DNA (***) totally relaxed DNA

As we can observe from the gel in Figure 3.3, after incubation with *Mth*MCM, a band corresponding to a partial relaxed (**) plasmid appeared. This effect is not detectable after incubation with the ΔsA mutant. These results confirms that the binding of MCM proteins influence the degree of supercoiling of dsDNA and that subdomain A is involved in this association with DNA (Costa *et al*., 2008).

To further analyse the possible role of subdomain A, the DNA binding properties of the ΔsA mutant with DNA substrates of various lengths were measured (Figure 3.4). Although the ΔsA mutant is an active helicase, deletion of the sA impairs the binding of both ssDNA and dsDNA. Although ssDNA binding can be partly recovered by increasing the amount of protein (data not shown), no binding to short stretches of dsDNA (50 bp) is observed even at high protein concentration. It is possible to detect

some dsDNA binding for the ΔsA mutant when using longer dsDNA segments (250 bp) (Costa *et al*., 2008).

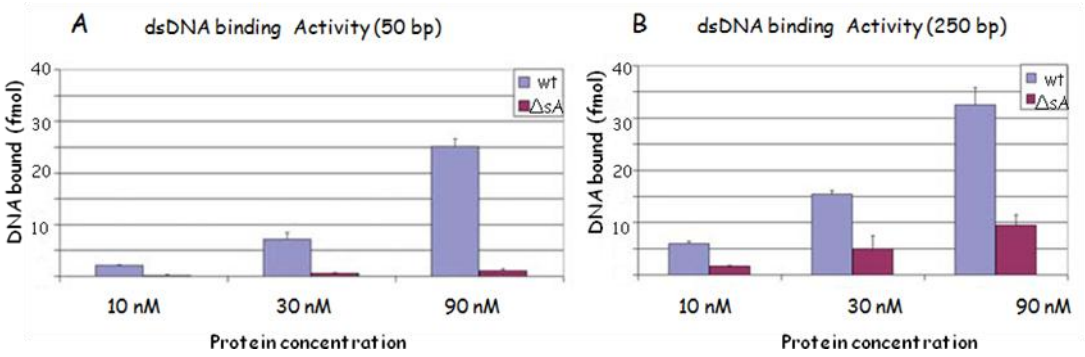


Figure 3.4 DNA binding activity. DNA binding properties of the ΔsA mutant: filter binding assays were performed using both the wt and ΔsA mutant protein, in the presence of A) a 50bp dsDNA substrate and (C) a 250 bp dsDNA substrate. The ΔsA mutant shows a decreased binding to both dsDNA (250 bp), whereas the binding to short stretches of dsDNA (50 bp) remains virtually undetectable

Secondary structure comparison allowed to identify some similarity between the sA and a large number of dsDNA binding protein containing a HTH motif. All these proteins contact the dsDNA through a recognition helix and a lysine or arginine residue. The same cluster of helices and a similarly positioned arginine (R87) have been found in the sA (Figure 3.5).

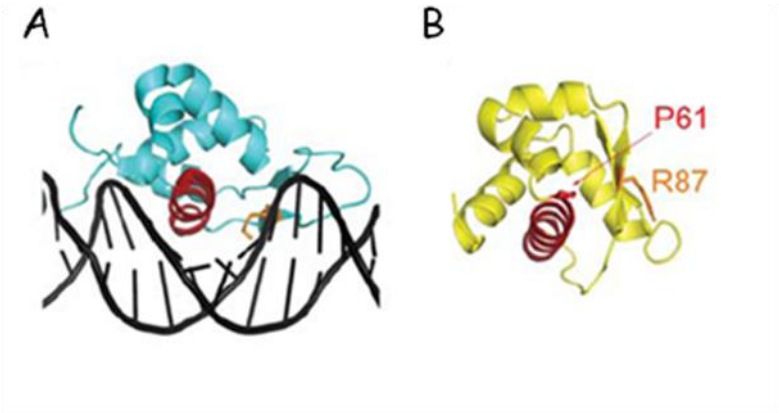


Figure 3.5 The crystal structure of the Genesis transcription factor bound to dsDNA (Jin *et al*., 1999, PDB entry 2HDC). A) The recognition helix is highlighted in red, and the lysine of the basic loop in orange. B) The structure of subdomain A in the same orientation: the atomic models of proline 61 and arginine 87 are shown in red and orange, respectively.

To assess the potential role of R87, I cloned, expressed and purified two mutants, where this residue was changed to an alanine or to a glutamic acid. The rational for the choice was that the mutation R87A is easier to be accommodated by the protein, but less likely to completely disrupt the dsDNA interaction; the reverse-charge mutation R87E may have some serious structural effect on the N-term structure, but if not disrupting the structure it is more likely to dramatically affect dsDNA binding. Both mutants could be expressed at high yield and displayed a similar behaviour to the wt in size-exclusion chromatography, confirming that the overall integrity of the structure was not affected (Figure 3.6 C).

Biochemical experiments showed a decrease in the DNA binding properties of both the R87A and the R87E (with the Glu mutant showing a more marked effect, as expected), but the decrease was not dramatic (Figure 3.6 B). It has to be stated that in the presence of a protein showing multiple types of interaction with DNA, is very difficult to properly design an experiment to selectively dissect the effect of residues involved in one specific interaction.

When we visualised both mutants in the presence of the 5,600bp dsDNA substrate, we observe that the proteins tend to form double rings or fibres and although DNA seems to engage with the protein, it is not as tense as in the presence of the wt protein (Figure 3.6 A).

Biochemical assays (ATPase, helicase and DNA binding, topology footprinting) simply showed a general decrease of all the protein functions, suggesting that probably R87 is not a crucial residue for the dsDNA recognition or interaction, as predicted from the modeling studies, but it is still showing a significant effect.

3.1.2 The Glu switch

In addition to the canonical AAA+ sequence motifs, a feature called the “glutamate switch” (E switch) has been described that may provide a way to regulate the ATPase activity of this superfamily of proteins (Zhang and Wigley, 2008). A polar residue (typically an asparagine) can form a hydrogen bond with the glutamate residues of the Walker B motif (DExD), which in turn is critically involved in

orienting and polarizing a water molecule for the nucleophilic attack on the phosphate of the ATP substrate. This ensures that the ATPase activity is switched off unless the system is correctly assembled. The RuvBlike and HslUV/ClpX families possess an unusual version of the E switch, involving a Thr/Ser residue shifted by two residues from the standard Asn position.

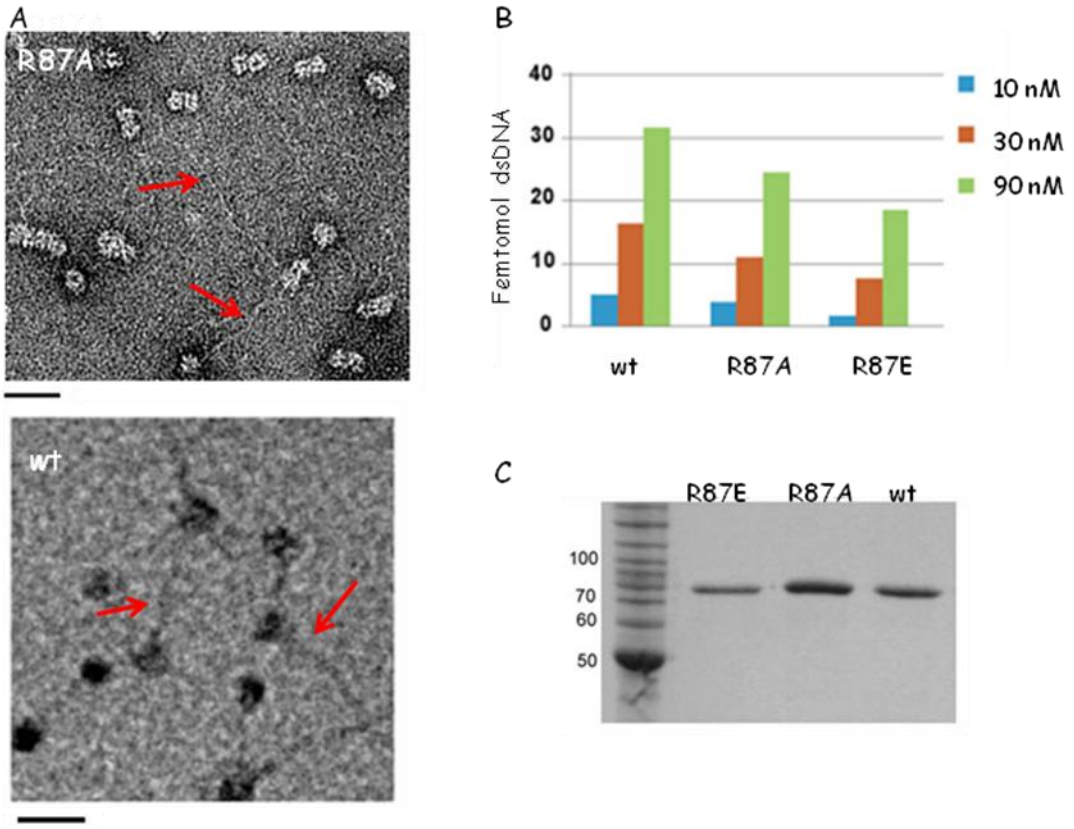


Figure 3.6 Biochemical and structural characterization of the R87 mutants. A) EM micrographs of negative stained grids, to visualise the MCM/dsDNA interactions. The upper micrograph shows the R87A mutant with a 5600 bp dsDNA substrate. The lower micrograph shows the wt *Mth*MCM with the same substrate. The red arrows indicate the DNA. Note that the DNA fibers are tense in presence of the wt protein, contrary to what is happening with the mutant, suggesting a different modality of DNA interaction. B and C) dsDNA binding properties of the *Mth*MCM proteins, 12% SDS-Page of the proteins after the exclusion size purification step.

A three-dimensional model based on the high resolution MCM structure from *M. kandleri* (Bae *et al.*, 2009) suggests the intriguing possibility that the switch may be even more complex in MCM proteins, involving a conserved Thr/Ser residue located at the end of the second strand of the AAA+ fold core, in the canonical position of the usual Asn, as well as an additional Arg/Lys located two residues

further along the chain, where the RuvB-like switch is positioned. This residue is highly conserved either as an arginine or lysine, with the exception of MCM2 proteins which have a conserved glutamine (Appendix 1). The loop containing both the Thr/Ser and the Arg/Lys residues immediately precedes helix 2 and the h2i that is postulated to interact directly with the DNA inside the channel, providing a possible route to switch on the ATPase activity upon substrate binding. (Figure 3.7).

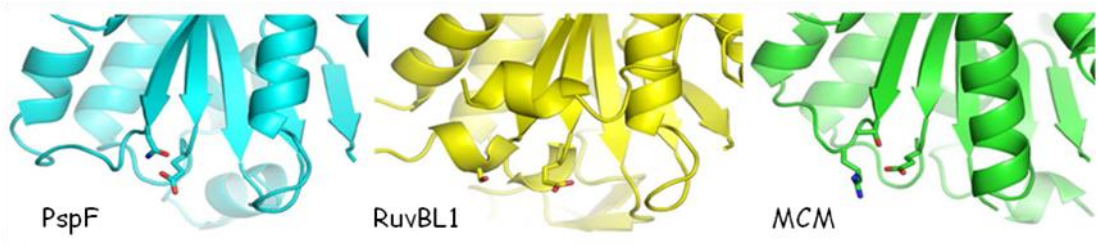


Figure 3.7. The glutamate switch in a variety of AAA+ ATPases (Zhang and Wigley, 2008). On the left panel the canonical switch as seen in the crystal structure of PspF (cyan) bound to the ATP substrate (Rappas *et al.*, 2006); the switch consists of a conserved asparagine (located at the very end of the β-strand following the Walker B motif) holding the Walker B glutamate (DExD) in an inactive conformation in the absence of the substrate. In the RuvB-like family (yellow) the switch involves a threonine or serine residue located two residues downstream from the usual Asn position, as illustrated by the crystal structure of the human RuvBL1 protein (Matias *et al.*, 2006). The *Mka*MCM2 crystal structure (green) shows that the position corresponding to the E switch in PspF is occupied by a residue that is a threonine or serine in the MCM family, whereas an arginine or lysine is located two residues downstream, in the position of the RuvBL1 E switch (Bae *et al.*, 2009). Since the inactive *Mka*MCM2 does not have a Thr/Ser in the first position and has a histidine substituting the Walker B glutamate, the corresponding residues have been mutated in the figure to illustrate the situation found in the majority of the MCM sequence.

Biochemical studies carried out in *S. sulfolobus* MCM have shown that both these residues have an effect on the degree and regulation of the helicase activity. The positively charged residue (K366) has been mutated in *Sso*MCM and the mutation completely abolishes the helicase activity and displays a modest reduction in the rate of ATP hydrolysis (Moreau *et al.*, 2007). This is consistent with a role in regulating the helicase activity and coupling ATP hydrolysis with substrate binding. On the other hand the T364 mutant shows lower DNA binding and helicase activities compared with the wild type (wt), but the ATPase activity is higher (Mogni *et al.*, 2009), again arguing for a coupling role.

To experimentally confirm these structural insights, I cloned, expressed and purified two single mutants of *Mth*MCM where both residues were mutated to alanine, S342A (SA) and K344A (KA), as well as a double mutant S342A/K344A (SKAA). All of them behave as the wt *Mth*MCM in size exclusion chromatography, confirming that they don't affect the overall structure and stoichiometry (Figure 3.8). Unfortunately the proteins were heavily contaminated by DNA and therefore as yet unsuitable for the biochemical analysis.

With hindsight the slight shift of the S342A mutant on the size exclusion chromatography profile (Figure 3.8) does correlate with a relatively lower level of DNA contamination.

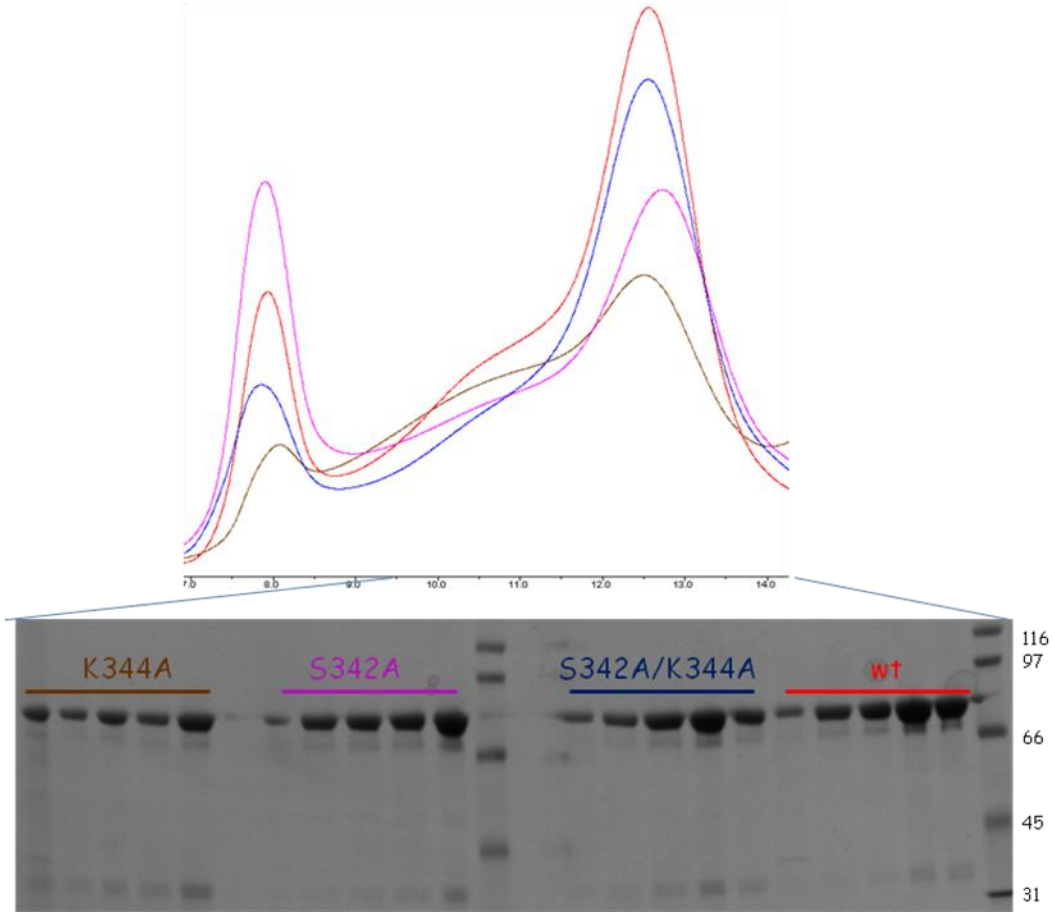


Figure 3.8 Exclusion size chromatography profiles. The chromatographic profiles of the wt *Mth*MCM (red), the double mutant (blue), the S342A (magenta) and the K344A (brown). All proteins show the same behavior forming higher molecular weight complex (first peak, generally corresponding to the helical aggregates). The fractions corresponding to a double ring (second peak) were collected.

3.1.3 The C-terminal winged helix (WH) domain

One of the aims of my project was to carry out structural and biochemical studies on the C-terminal region of the MCM proteins. When I started my work, no structure of this region was available, beside bioinformatics analysis detecting the putative presence of a WH motif in the archaeal sequences. Towards the end of my PhD, the NMR structure of the C-terminal domain of human MCM6 has been determined, showing indeed the presence of a WH domain (Wei *et al*., 2010).

In order to get soluble recombinant protein I cloned, expressed and purified the C-term of the two best studied archaeal MCMs: *Mth*MCM and *Sso*MCM. For *Mth*MCM I subcloned the region from residue 582 to residue 666 (*Mth*WH), to produce a 10 kDa protein, and for SsoMCM the region between residue 612 and 686 (*Sso*WH) generating a protein of 8 kDa, as previously characterised by Barry *et al.* (2007). Each fragment was cloned in both the pProExHTa and pETM 30 expression vectors, to generate fusion proteins with either a 6His and a 6His-GST tag (Figure 3.9), both cleavable with the recombinant Tobacco Etch Virus protease (TEV). Due to the higher levels of expression, the constructs with the Histidine tag at the N-terminus have been used for this work.

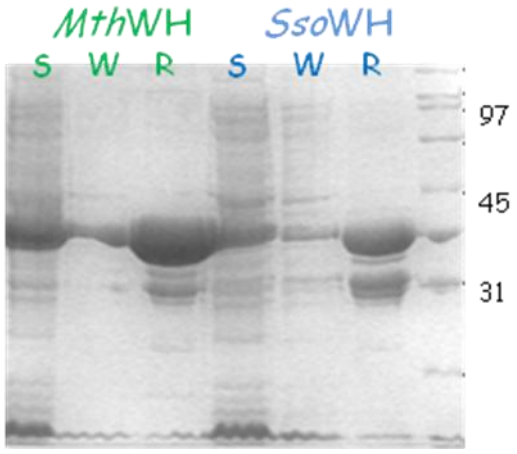


Figure 3.9 SDS-page analysis of small scale expression tests. The small scale expression of 6His-GST tagged C-term of *Mth*MCM and *Sso*MCM; S: soluble fraction, W: wash, R. Ni^{2+} resin. The molecular weight of both constructs is nearly 36KDa.

Analysing the sequence alignment of various archaeal MCM, I realized that *Sso*WH as described in Barry *et al.* (2007) was missing a stretch of residues well conserved in the archaeal sequences (Figure 3.10); I therefore cloned a new slightly longer fragment (*Sso*E605) that includes these residues. As for the others, this fragment was inserted into ProExHTa

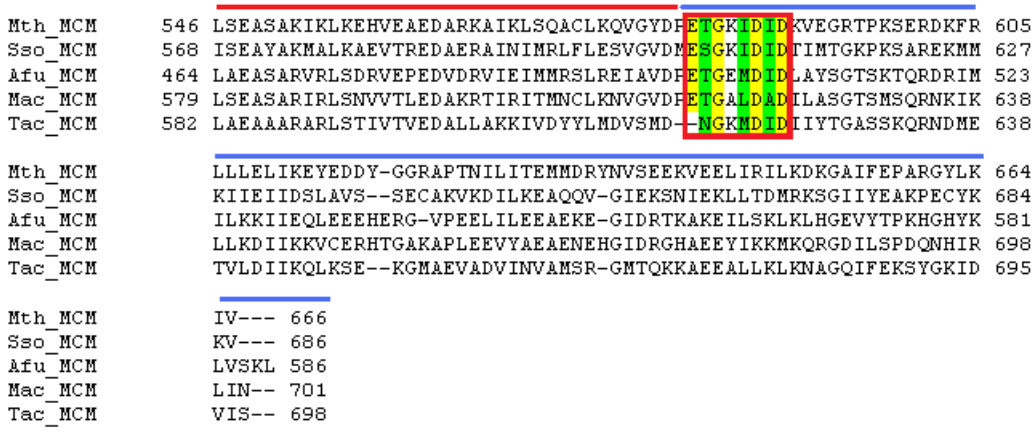


Figure 3.10 Sequence alignment of archaeal MCM C-terminal domains. The C-terminal region (blue bar) of the archaeal MCM proteins were aligned. The red box indicates the conserved residues: in yellow are the identical amino acid, whereas in green the similar.. *M. thermautotrophicum (Mth), S. solfataricus (Sso), A. fulgidus (Afu), M. acetivorans (Mac) T. acidophilum (Tac)*

For all the three proteins (*Mth*WH, *Sso*WH and *Sso*E605) , large scale expression (6 liters) was carried out using *E. coli* BL 21 cells grown in TB media; once the OD_{600} value reached 1, protein expression was induced with 1 mM IPTG for 5 hours, and the over-expressed proteins purified by Nickel affinity chromatography. After cleavage of the 6His-tag using TEV protease, the protease as well as the uncleaved protein were removed by a second Nickel column. An ionic exchange step was initially included in the purification to remove any DNA contamination, but all the samples were clean after the two Nickel columns (as tested by gel analysis or absorbance at 260 nm), so this step was omitted. Finally the samples were analyzed on a size exclusion column (Superdex 75 10/30 or HiLoad Superdex 75 16/60).

All the proteins eluted at a volume similar to the Ribonuclease A marker (Figure 3.11), a protein of 13.7 KDa, and the profile does not change upon changing the concentration of the samples (data not shows), suggesting that they are likely to be monomeric.

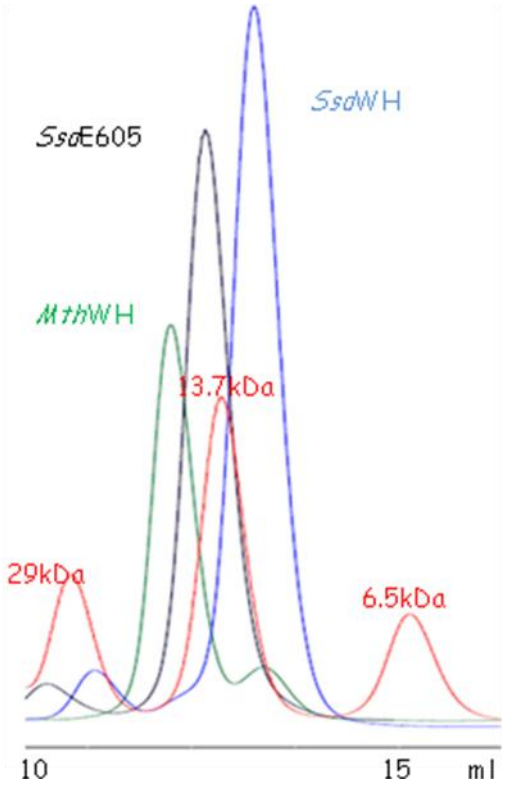


Figure 3.11 Size exclusion chromatography analysis of the C-terminal domains. The profiles of the proteins under study (*Mth*WH green, *Sso*WH blue, *Sso*E605 black) are compared to the molecular weight marker (red profile**)**; Aprotin 6.5 kDA, Ribonuclease A 13.7 kDA, Carbonic anhydrase 29 kDa**.**

A DNA binding test was carried out and showed no DNA interaction with either ssDNA or dsDNA, as already reported in literature (Barry *et al*., 2007; Wei *et al*., 2010).

All the proteins were extremely soluble and it was possible to reach very high protein concentration (up to 90 mg/ml). The proteins were routinely concentrated to the desired level in 30 mM Hepes-NaOH pH 7.5, 150 mM NaCl, 10% Glycerol. After extensive crystallisation trails, the initial condition were modified by removing the glycerol, reducing the salt concentration (down to zero) and finally changing the pH of the solution (30 mM phosphate buffer pH 6.5, 50 mM NaCl).

Crystallization trials were carried out via the sitting drop vapour-diffusion method using the Mosquito liquid handling robot (TTP LabTech's). Crystallisation drops contained either 1:1 or 2:1 volume ratios of protein to reservoir solution, respectively, with final drop volume of 200 or 300 nl (Figure 3.12)

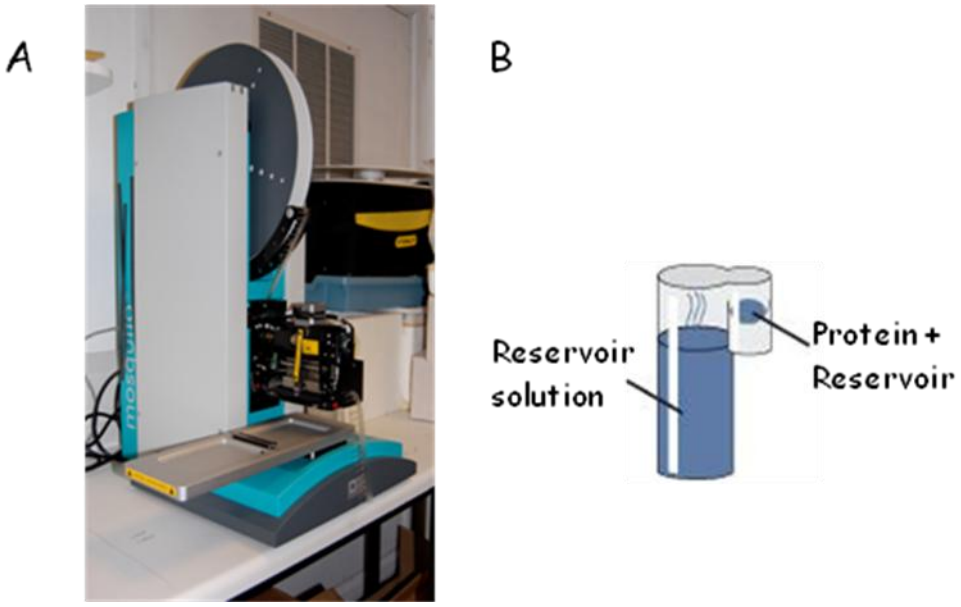


Figure 3.12 Crystallization trials. A) Mosquito liquid handling robot (TTP LabTech's) available in our laboratory. B) schematic representation of the sitting drop vapour diffusion technique used in all mine screening.

Initial crystallisation trials were set up at room temperature, at 4°C and 37°C using commercially available screens: Hampton Crystal Screen 1 and 2, Hampton PEG/Ion, Index Screen, SaltRX (all from Hampton Research), Wizard 1, 2 (Emerald Biosystems), Classic, Classic II, pH clear, pH clear II, comPAS, PACT and JCSG+ (Qiagen). These allowed the screening of many thousands of conditions for each protein.

An initial irregularly-shaped crystal of *Sso*E605 (at a concentration of 35 mg/ml) was obtained in 0.2M Ammonium acetate, 0.1M Sodium acetate pH 4.6, 30% PEG 4000 (Figure 3.13). A very large number of optimization attempts were carried out in order to reproduce the crystal and improve its morphology, using different seeding techniques, additive screens or "ad hoc" screens, with not satisfying results. It was impossible to test the diffraction property of the crystal as it dissolved during harvesting.

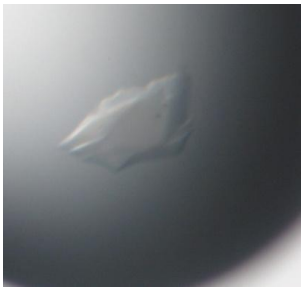


Figure 3.13 Irregular crystal obtained for *Sso*E605. The crystal (about 70 µm in lenght) was obtained in the following conditions: 0.2M Ammonium acetate, 0.1M Sodium acetate pH 4.6, 30% PEG 4000

In the meantime, preliminary NMR studies were started in collaboration with the University of Verona (Figure 3.14 A and B). Two preliminary tests have been carried out so far: a 2D TOCSY (Total Correlation Spectroscopy) experiment, which shows the correlation of H atoms separated by two or three covalent bonds and is therefore useful for the correlation of the side chains of proteins and peptides, and a 2D NOESY (Nuclear Overhauser Effect Spectroscopy) experiment that shows the correlation between H atoms separated by up to ~5 Å in space and therefore provides insights into the 3D structure.

The numbers of the peaks and their position indicate the possible presence of helices and short stretches of β-sheets for all the constructs, structures consistent with a WH motif. The protein will be enriched with the proper isotopes for further NMR analysis.

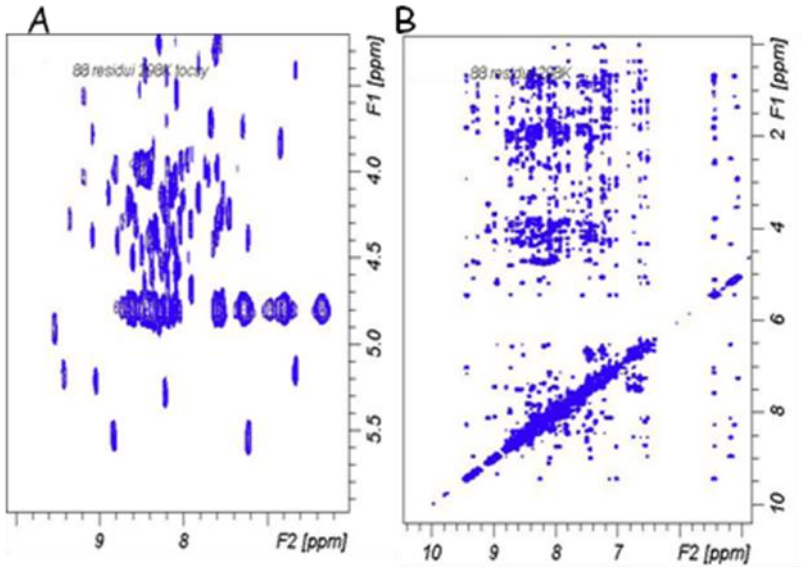


Figure 3.14 NMR and Crystallization results. The panels A and B are the TOCSY and 2D NOESY of the *Mth*WH, 0.5 mM protein in 30 mM phoshate buffer pH 6.5, 50 mM NaCl have been used for the analysis with a Bruker avance spectrometer at 600.13 MHz.

3.2 The eukaryotic MCM proteins

3.2.1 hMCM8 protein

Analysing the sequence alignment of various MCM8 with the structurally characterised archaeal MCM proteins (see Appendix 1 and 2), the putative boundaries of the domains were identified. The corresponding fragments were cloned into a pProExHta expression vector in order to obtain proteins with a 6 histidines-tag at the N-terminus (Figure 3.15).

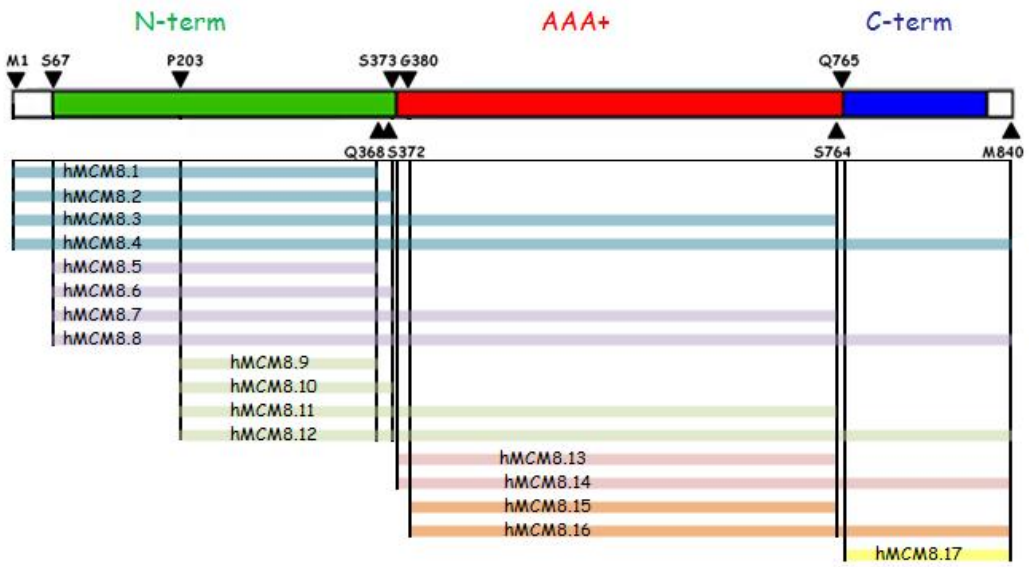


Figure 3.15 Schematic representation of the primary sequence of hMCM8 and the constructs produced during this work. The putative location of the domain boundaries, as inferred from the structure of archaeal MCM proteins is shown (green: N-term; red: AAA+; blue: C-term). The start and end points that have been used in the cloning strategy are indicated with triangles and labelled by the residues at the beginning or end of the cloned fragment (start points are shown above, end points below). Their combination generate 17 different constructs, as shown below

Using the classical cloning strategy with restriction enzymes (with BamH1 and Xba1) the fragments were generated by PCR, using as template the pTrc-HisC-hMCM8 plasmid kindly provided by Dr F.M. Pisani (IBP-CNR, Naples). The fragments and the vector have been digested with the appropriated restriction enzymes and then ligated using the T4 ligase. A colony PCR screening was performed to identify the positive clones, that were subsequently sequenced (Table 3.1). Constructs have been used to transform *E. coli* cells and a systematic screening

for optimal expression has been carried out, by testing the effect of different bacterial strains (BL 21 , BL 21 codon+, BL 21 gold, BL 21 star, BL 21 Rosetta2, B834), expression temperatures (18°C, 24°C, 30°, 37°C), expression times (5 hs, O.N, 24 hs, 48 hs), growing media (LB, TB) and induction methods (IPTG or auto-induction).

Construct Name	Starting and ending residues	Molecular weight (kDa)
hMCM8_1	M1-Q368	42.5
hMCM8_2	M1-S372	43.1
hMCM8_3	M1-S764	86
hMCM8_4	M1-M840	94.7
hMCM8_5	S67-Q368	35
hMCM8_6	S67-S372	35.4
hMCM8_7	S67-S764	78.4
hMCM8_8	S67-M840	87.1
hMCM8_9	P203-Q368	19.4
hMCM8_10	P203-S372	19.8
hMCM8_11	P203-S764	62.9
hMCM8_12	P203-M840	71.5
hMCM8_13	S373-S764	44.2
hMCM8_14	S373-M840	55.9
hMCM8_15	G380-S764	43
hMCM8_16	G380- M840	54.7
hMCM8_17	Q765- M840	9.6

Table 3.1 Histidine-tagged constructs summary. The table summarises all the hMCM8 constructs and their molecular weights

Despite the wide screening, most of the constructs were not expressed at high levels or insoluble in the examined conditions and only 2 of them were soluble: hMCM8_2 (1-372 aa) corresponding to the N-terminus of the protein, and hMCM8_14 (373-840 aa) that includes the AAA+ and the C-terminal domain of the protein.

The optimised conditions for the two successful constructs are the following:

- hMCM8_2 (1-372 aa): BL 21 Codon + RIPL cells, grown in LB broth, induced with 0.5 mM IPTG, then grown for 24 hs at 25°C

- hMCM8_14 (373-840 aa): Rosetta 2 cells, grown in TB Broth, induced with 0.5 mM IPTG, then grown for 24 hs at 25°C

As fragments hMCM8_7 (S67-S764), hMCM8_8 (S67-M840), hMCM8_11 (P203-S764), hMCM8_12 (P203-M840) were systematically insoluble (Figure 3.16), a new cloning strategy has been adopted: I decided to clone these fragments with Maltose Binding Protein (MBP) as fusion partner, as the MBP tag has been shown to help the folding of recombinant proteins, thereby increasing their solubility.

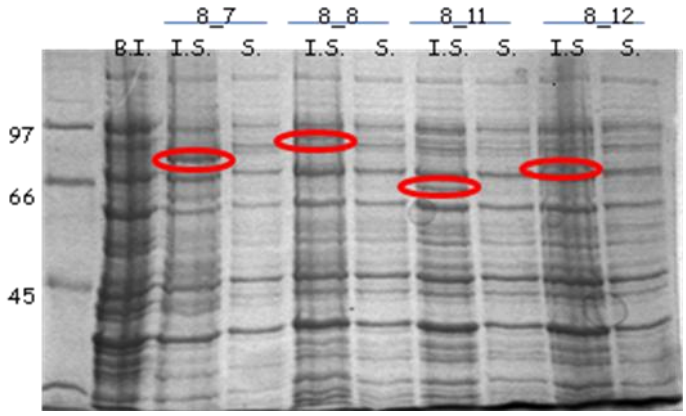


Figure 3.16 8% SDS-Page of a small scale expression test. The small scale expression test performed with BL21 codon + RIPL *E coli* strain. For all the constructs tested the insoluble (I.S) and soluble (S) fractions are loaded on the gel and compared with a totat extract before the IPTG induction (B.I.). The red circles show the proteins of interest in the insoluble fractions

As the vector chosen (pETM41, EMBL) absolutely required that the restriction site of the N-terminal primer was NcoI, the endogenous NcoI restriction site present in the hMCM8 DNA sequence (position 1597) was removed with the Quick-change point mutation system (Stratagene). The new construct, called hMCM8-noNco, was then used as template for the PCR. The restriction enzymes selected for the cloning are NcoI and NotI. In Table 3.2 the information on the MBP-fusion constructs are summarised.

Construct Name	Starting and ending residues	Molecular weight (kDa)
MBPhMCM8_7	S67-S764	121.4
MBPhMCM8_8	S67-M840	130.1
MBPhMCM8_11	P203-S764	105.9
MBPhMCM8_12	P203-M840	114.5

Table 3.2 MBP-tagged construct summary. The table summarises all the MBPhMCM8 constructs and their molecular weights (including the 6H-MBP tag of 43 kDa)

Small scale expression tests were performed with the MPB constructs and in all cases the best results were obtained using BL 21 codon + RIPL *E. coli* strain in an autoinduction TB broth (TB broth, 0.8% Glycerol, 0.5% Lactose, 0.015% Glucose, 2 mM $MgSO_4$) (Figure 3.17). Small scale purification tests were performed using MBP affinity column and the elution steps carried out with 10 mM Maltose. Similarly to the truncated forms (see below), these constructs are extremely sensitive to proteolysis.

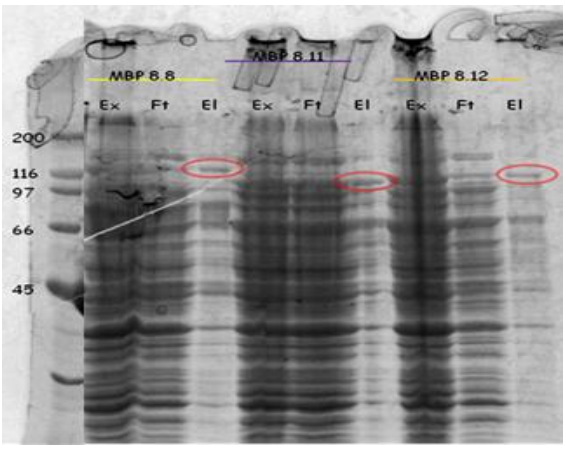


Figure 3.17 8% SDS-Page small scale expression of MBP construct. The pellet from 50 ml of culture has been resuspended in 5ml of a buffer containing 20 mM Tris pH 7.4, 200 mM NaCl, 1 mM EDTA, 5% Glycerol, 2 mg/ml lysozyme, 2.5 mM PMSF. After sonication and centrifugation, the supernatant is loaded on 1 ml MBPhitrap column. The elution performed with 10 mM Maltose. The red circles show the proteins of interest fused with MBP (43kDa). For each protein the Total Extract (Ex) Flow through (Ft) and the Elution (El) have been loaded.

In the meantime, large scale expression tests (1 Liter) were carried out for the soluble His-tagged 8_2 (corresponding to the N-terminal region of the protein) and the 8_14

(containing both AAA+ domain and C-terminal region) fragments. The first step consisted in an affinity Nickel column, a second step involved ionic exchange chromatography and a final size exclusion column (Superdex 200) was used to check the oligomeric state of the protein.

Both fragments were very sensitive to salt concentration, precipitating when the concentration of NaCl was lower than 300 mM. After the last step of purification both proteins contained a significant number of bands corresponding to impurities and/or proteolysis products (Figure 3.18).

N-terminal sequence analysis performed by Dr. Keen at the Proteomics Facility of Leed University, helped to understand the nature of the additional protein bands present in the samples. Bands running on an SDS gel at a molecular weight of approximately 45 kDa and 33 kDa (Figure 3.18, bands in the red boxes) corresponded to a protein with the same N-terminal amino acids as the unproteolysed 8_14 construct, indicating the loss of a C-terminal fragment. A 25 kDa band (Figure 3.18, green box) present in both samples is due to contamination by an *E. coli* protein. This contaminant is a chaperone protein (the peptidyl-prolyl cis-trans isomerase FKBP): containing a His-rich domain with high affinity for Ni^{2+} ions, it is a typical contaminant of proteins purified by Nickel affinity chromatography.

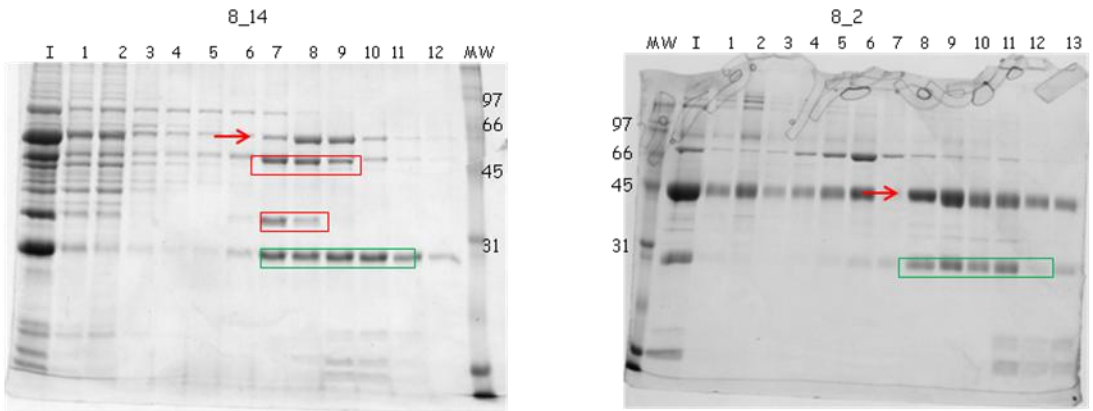


Figure 3.18 12% SDS-Page of fractions from size exclusion chromatography. After a gel filtration run the fractions have been assayed by SDS-Page. The red arrows indicate the protein of interest (hMCM8_14, 52KDa and hMCM8_2, 43Kda). Both proteins samples include contaminants (green boxes) and proteolytic fragments (red boxes).

In order to obtain proteins with a higher level of purity, a new purification protocol was devised. The metal affinity purification step was modified by using Co^{2+} (which has a lower affinity for histidine) instead of Ni^{2+} to increase selectivity and therefore remove the more loosely bound FKBP protein. The addition of Mg^{2+} and ATP to the purification buffer was introduced to avoid any GroEL contamination. An extra washing step with high salt concentration (1 M NaCl) was carried out to reduce nucleic acid contamination.

Preparative size exclusion chromatography was used as a second step for the hMCM8_14 deletion mutant (Figure 3.19).

A heparin affinity column was used for hMCM8_2 (the protein eluted at high salt concentration, close to 2M NaCl), followed by size exclusion chromatography. Although the 8_2 construct tends to form aggregates, both protein eluted in a dimeric form.

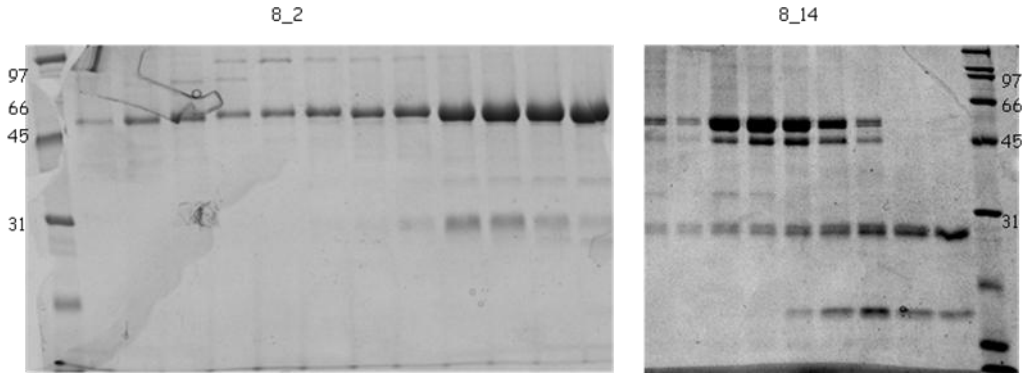


Figure 3.19 12% SDS-Page of fractions from size exclusion chromatography. The new purification strategy allows to remove the nucleic acid contamination as well as the GroEL contamination, unfortunately the protein are still proteolysed.

The new purification protocol allowed the removal of traces of chaperones and nucleic acid contamination. Unfortunately, despite increasing the concentration of protease inhibitors (Figure 3.20), I haven't been able to avoid proteolysis, especially in the case of hMCM8_14.

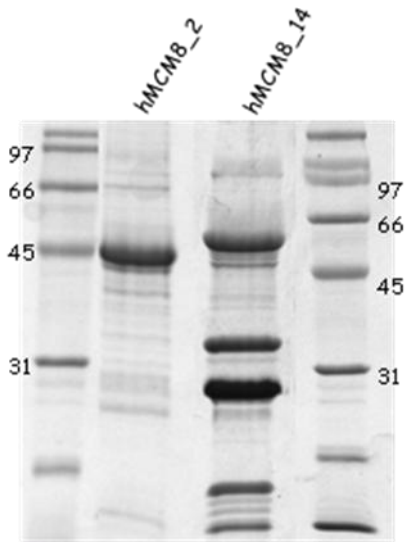


Figure 3.20 12% SDS-Page of the His-tagged hMCM8 soluble constructs. The SDS-Page of both hMCM8 deletion mutant after all the purification steps. The hMCM8_14 construct appears really sensitive to proteolysis.

A preliminary biochemical characterization for the both truncated forms was performed.

First, the oligomeric state of both constructs was assessed in presence of different substrates (nucleic acids or nucleotides) in 25 mM Hepes-NaOH (pH 7.5), 150 mM Na acetate, 10 mM Mg acetate, 1 mM dithiothreitol. The protein was incubated for one hour with ssDNA or dsDNA, at a DNA/protein ratio of 1.5/1 in molarity (at room T) and then the samples were injected in a Superdex 200 3.2/30 column. For the hMCM8_14 which comprises the AAA+ domain, incubation with 0.5 mM ATP for 15 minutes at room T was also performed. In all instances the size exclusion chromatography profile did not reveal any indication of substrate-induced oligomerisation, as the proteins maintained the dimeric state (Figure 3.21). The small shoulder of about 44 kDa visible on the chromatogram in figure 3.21 (green arrow) probably suggests the presence of dynamic equilibrium between the monomeric and dimeric forms of the protein or may be due to the formation of dimers of the proteolytic fragments present in the preparation.

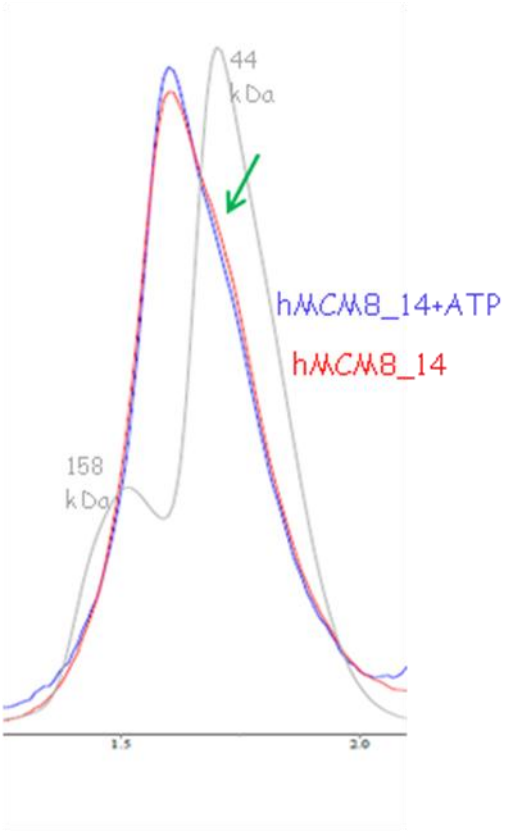


Figure 3.21 Analytical exclusion site profile of hMCM8_14 in the presence of ATP. The oligomeric state of the proteins of interest is tested with different substrate. The protein (30 µM) was incubated for 15 min a room temperature with nucleotide (ATP) or with DNA substrates for 1hour in 25 mM Hepes-NaOH (pH 7.5), 150 mM sodium acetate, 10 mM MgCl2, 1 mM dithiothreitol. Here are compared the profiles of the hMCM8_14 (red) and the profile of hMCM8_14+ATP (blue) and molecular weight marker (grey, including Ovalbumin: 44 kDa and Aldolase: 158 kDa.

The experiment described above has a number of possible pitfalls. Given the relatively low affinity of MCM proteins for DNA, as shown by the large excess of protein needed to shift the DNA in most EMSA experiments carried out on archaeal MCM proteins (Kasiviswanathan *et al*., 2004; Barry *et al*., 2007), it is possible that hexamerization or in general oligomerisation may be detected only in the presence of a large excess of substrate, not feasible in the above described experiment. Due to the inherent instability of the protein, it was not possible to reach higher concentration and the limits of the instrument may not permit the detection of very small amounts of oligomers.

Electrophoretic mobility shift assays (EMSA) were performed to evaluate the DNA binding properties of both the N-terminal and C-terminal domain. The results show

that both proteins bind radiolabeled single stranded DNA (ssDNA) forming multiple DNA/protein complexes, possibly due to the presence of different proteolytic fragments and oligomeric states (Figure 3.22)

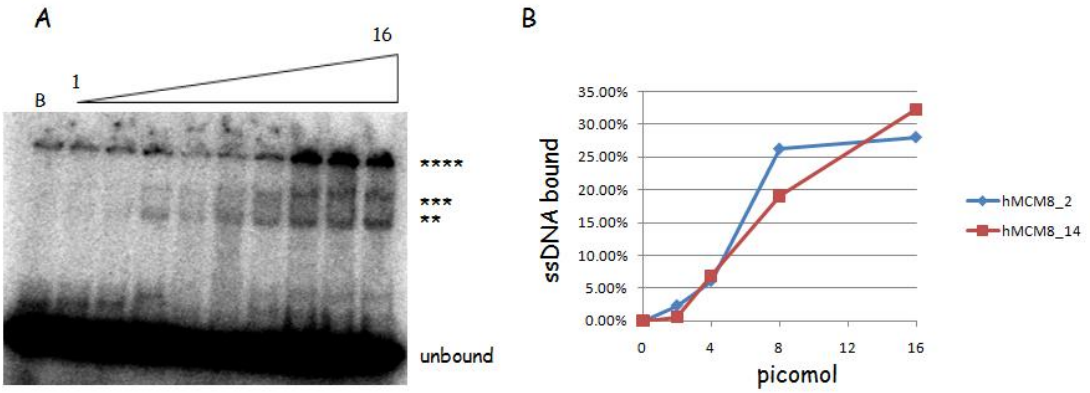


Figure 3.22 EMSA assay with a ssDNA. A) Increasing amount of hMCM8_2 (from 1 to 16 pmol as monomer) were incubated with 50 femtomol of ssDNA (41bases) for 1h at room temperature in a DNA binding buffer (25 mM Hepes pH 7.5, 25 mM NaAcetate, 10 mM MgAcetate, 1 mM BME, 0.1 mg/ml BSA, 10 µl reaction). The sample were loaded onto a native 8% polyacrylamide gel, run in 0.5X TBE buffer for 40 min @ 100V. The (*) indicates the DNA/protein complexes. B) The ssDNA ability of both truncated form of the hMCM8 has been quantified, showing a very similar behaviour for both hMCM8_2 (blue) and hMCM8_14 (red).

Generally MCM deletion mutants tend to show a relatively low affinity for DNA, loosing the synergic effect due to the presence of multiple DNA binding domains. For example, systematic studies of the DNA binding properties of various deletion mutant of *Mth*MCM and *Sso*MCM (;) showed that the N-terminus of *Mth*MCM is at least 8 times less efficient of the wt in shifting a ssDNA substrate (Kasiviswanathan *et al*., 2004), and the truncated form of *Sso*MCM even less efficient (a *Sso*MCM construct including the AAA+ and the C-terminal domain, equivalent to hMCM8_14 is 100 times less efficient compared to the full-length protein, Barry *et al*., 2007). The DNA binding affinity of the two deletion mutants of hMCM8 presented in this study is therefore in line with the above results. Preliminary experiments also show that both proteins bind dsDNA and fork-like substrates with comparable affinities.

Since the 8_14 construct contains the ATPase domain, assays to measure the ATPase catalytic activity have been performed. As shown in Figure 3.23, the 8_14 deletion mutant displays a significant ATPase activity and such activity is stimulated by ssDNA, as expected for a helicase. As in AAA+ proteins the ATP binding pocket is

located at the interface between subunits and requires residues from two adjacent monomers, these results are consistent with the dimeric state of the protein.

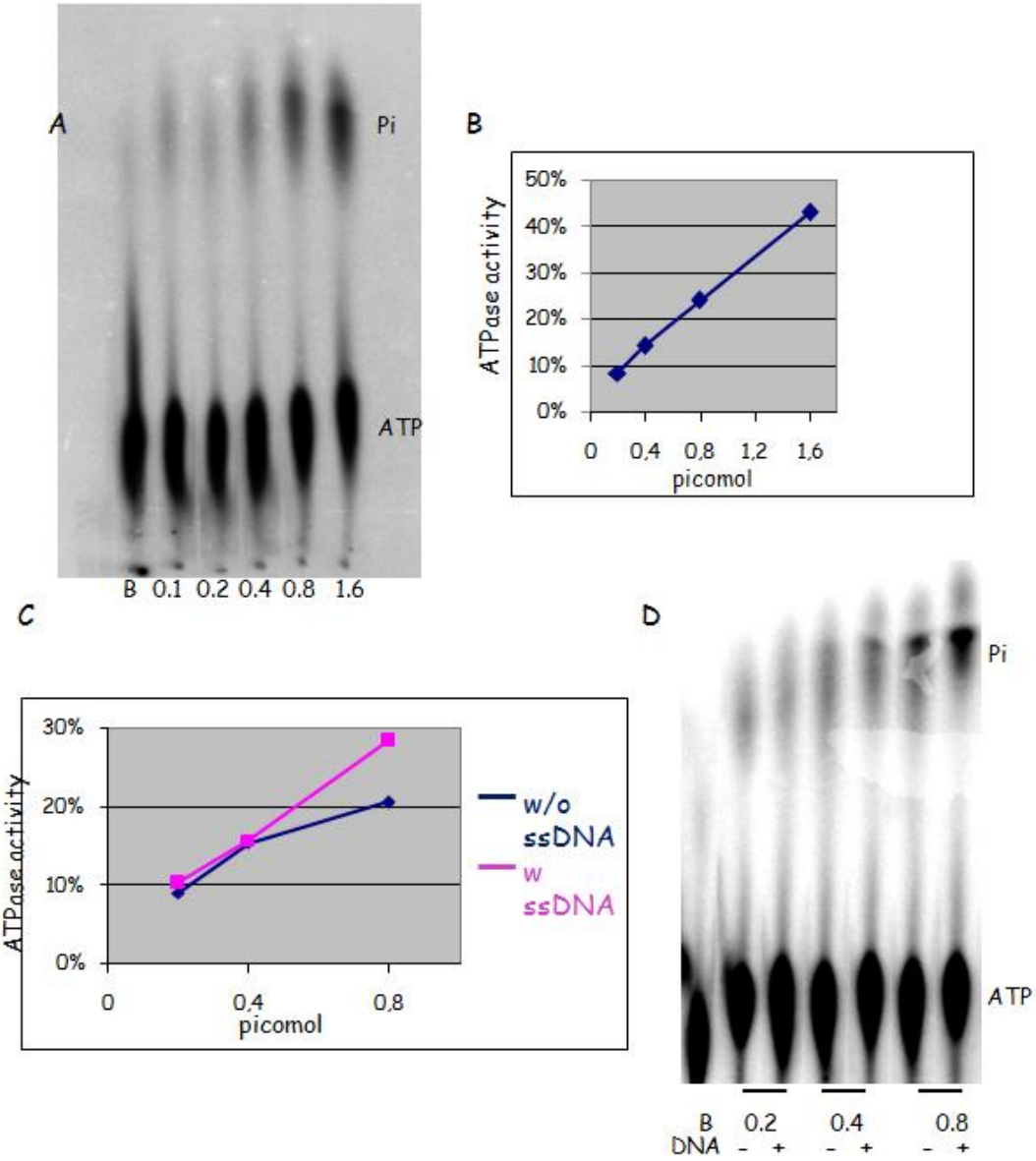


Figure 3.23 8_14 ATPase assay. A) Increasing amounts of hMCM8_14 protein (shown below the gel in pmol as monomer) have been incubated with 5 mM ATP for 30 min @ 37°C in a 50 mM Tris pH 7.9, 10 mM MgAcetate, 0.5 mg/ml BSA,2 mM BME (10µl). B) ATP hydrolysis activity of the hMCM8_14. C) the same experiment described in A) in presence or absence of 200 fmol of a ssDNA (41bp oligonucleotide) in a 50 mM Tris pH 7.9, 10 mM MgAcetate, 0.5 mg/ml BSA,2 mM BME. D) comparison of the ATP hydrolysis activity of the hMCM8_14 without (blue) or with (magenta) ssDNA

Due to time limitations, the experiments described above need to be considered as preliminary, and more work is needed to optimise the assays for a full

characterization of both truncated forms of hMCM8 and to better investigate their biochemical properties.

For example no helicase activity has so far been detected. A possible explanation may lie in the need of optimisation of the experimental conditions. Since the helicase activity should require an intact hexamer, another possibility is that the lack of DNA unwinding is due to the fact that the deletion fragments loose the ability of oligomerise, as I haven't been able to detect any hexamerization (for both deletion mutants) in the presence of substrates such as ATP and DNA. A more interesting hypothesis is that the complex may require the presence of additional physiological partners.

Although the information that I have so far obtained are still preliminary, my constructs represent the first deletion mutants of hMCM8 (obtained in *E. coli* cells, so without post-transcriptional modification) showing biological activities. Site directed mutants for each domain have been designed and further biochemical and structural investigations will be performed.

3.2.2 Human MCM9 protein

The bioinformatics analysis performed by Liu *et al.*, (2009) demonstrated that both MCM8 and MCM9 are evolutionary linked, being both present or absent in each organism, with the exception of *Drosophila*, as previously discussed. This also suggests a functional link between the two MCM protein, as both of them do not interact with the helicase 2-7 complex.

In order to analyse this potential interaction, I focused my attention on human MCM9 (hMCM9). As no cDNA is commercially available for the full-length protein (only a cDNA encoding for the N-terminal part of the protein from the residue 1 to 391 – described in Yoshida (2005) - can be purchased) I tried to produce cDNA starting from messenger RNA present in the cells using Reverse Transcriptase (RT) PCR.

No information about the specific expression patterns of hMCM9 in tissues and/or cell-cycle stages is available. As generally the MCM protein are only expressed in

proliferating cell, I used total RNA from different cancer cell lines (Table 3.3) kindly provided by Prof. G. Del Sal.

Cell line	**Tissue**
MCF10A	Mammary gland
HepG2	Liver hepatocellular carcinoma
HeLa	Cervical cancer
A549	Lung adenocarcinoma
SK-N-AS	Neuroblastoma

Table 3.3 List of cell lines used for hMCM9 RT PCR. The total RNA extract from these cell lines has been used for the first strand synthesis reaction using a $d(T)_{16}$ oligonucleotide. The PCR reaction are then performed with gene specific primers.

After many attempts using both oligo d(T) as well as gene-specific 3' UTR primers, it has only been possible to produce the cDNA encoding for the N-terminal region of the protein (Figure 3.24). Whether this is due to technical problems or to the possible presence of isoforms within the cell lines examined is still unclear.

To overcome these problems, the synthetic hMCM9 gene has been purchased (GeneArt) and the cloning of different domains is in progress using a restriction free (RF) cloning technique. RF cloning is a two-steps process. The two strands of the PCR product obtained in the first stage are used as “mega primers” for the second stage of amplification. In the second stage, the PCR product and the target vector are mixed and, following amplification reaction, the gene of interest is integrated into the circular vector into a pre-defined position.

A similar strategy to the cloning of deletion fragments of hMCM8 was implemented, using the structural information from the archaeal homologues to infer the domain boundaries of hMCM9 (Figure 3.24)

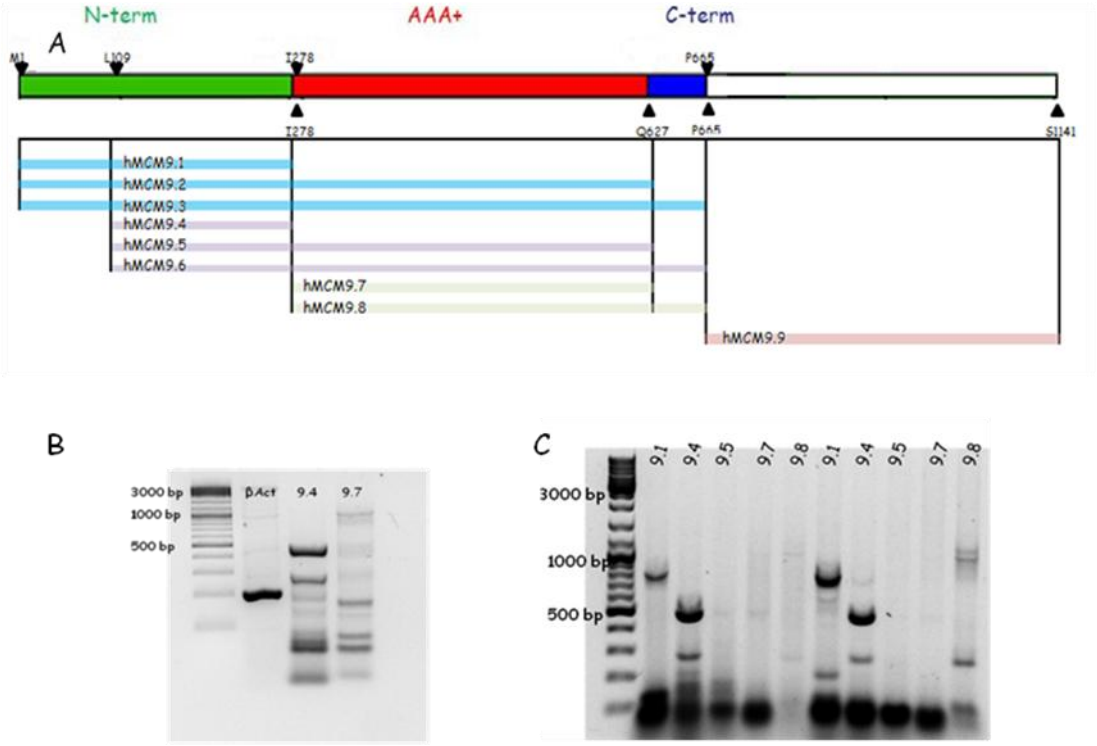


Figure 3.24 hMCM9 cDNA synthesis. A): schematic representation of hMCM9 compared with the structure of *Mth*MCM (colored part). Numbers and arrows indicate start and end point of the domain that are used in the cloning strategy; the combination of all of them allows the formation on 17 constructs. B and C): from total RNA of HeLa cells cDNA synthesis of hMCM9. A pair of primer for β-actin is used as positive control. The number of the constructs are referred to the scheme in panel A.

All the fragments are being cloned into a pETSumo vector (Invitrogen), expressing the protein as a fusion with the protein SUMO, under the control of the antibiotic kanamycin. The choice of the vector was based on the following considerations:

1. the presence of the SUMO-tag has been shown to improve the solubility of "difficult" proteins.

2. the kanamycin resistance permits the co-expression of each hMCM9 protein fragment with the equivalent fragment from hMCM8 constructs (pProExHTa, ampR); the presence of a physical interaction could be detected by pull-down assays and may strongly contribute to enhance the solubility of both potential partners.

Currently a number of constructs have been obtained (Figure 3.25) and they will soon be tested for protein expression and solubility, either alone or in co-expression with hMCM8 equivalent fragments.

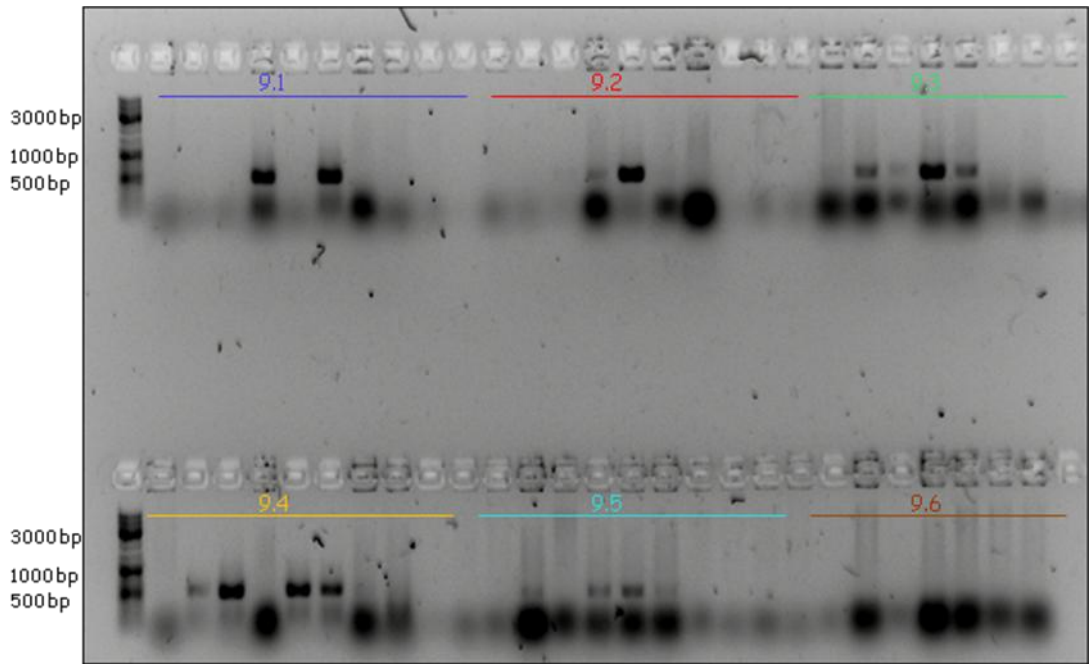


Figure 3.25 colony PCR screening of hMCM9 constructs. A colony PCR screening was performed to check positive clones. All the PCR were performed using the L109/I278r as pair of primers. The number of the construct is the same used in the Figure 3.24 A.

4 Conclusions and Future plans

4.1 Archaeal MCM proteins

The archaeal MCM proteins are good model to study the more complex eukaryoticorthologues. MCM proteins have been characterized over the years structurally and biochemically, but there are still many questions open: for example, how the protein is loaded onto the replication origin or how the helicase is unwinding the DNA duplex.

Following the results obtained in our laboratory by A. Costa, who obtained a 3D EM reconstruction of a novel mode of MCM/dsDNA interaction, I carried out a biochemical study to confirm and support the structural work. I have been able to demonstrated the effect of this interaction on the topology of the DNA and to confirm the essential role of subdomain A (Costa *et al.*, 2008). This topological stress may represent the contribution of the MCM protein to the melting of the origin and the consequent formation of the replicative bubble. An intriguing hypothesis is that MCM helices that have been previously visualised and studies by EM (Chen *et al.*, 2005) may have a physiological role similarly to the one proposed for the bacterial initiator DnaA. Preliminary work from our laboratory (Costa *et al.*, 2008) suggested that such helices are more prevalent in samples that are highly contaminated with endogenous DNA. At the beginning of my PhD I have been involved in the attempt to characterise from a structural point of view such MCM/DNA helical fibers, but the lack of reproducibility of the samples hindered progress. More work directed at the identification of experimental conditions that allow the preparation of reproducible samples may help to shed light on the putative role of MCM proteins in the initiation of DNA replication.

The C-terminal region of the MCM protein was the less characterized of the enzyme domains. To obtain structural and biochemical data on this region I have produced three expression constructs (one to produce the C-terminal region of *Mth*MCM and two corresponding to the C-terminus of *Sso*MCM) and optimise the expression and

purification protocols to obtained up to hundreds of milligrams of highly purified proteins. The DNA binding proteins of the three proteins have been examined, but no evidence for an interaction has been detected. These recombinant proteins have been used for "high throughput" crystallisation screenings using a robotic platform, to obtain high-resolution diffracting crystals. Despite the initial appearance of an irregularly shaped crystal, all the extensive attempts to reproduce and optimize the shape and the diffracting qualities have so far been unsuccessful. Due to the extension and breath of the crystallisation and optimisation experiments that have been carried out over more than two years, we reached the conclusion that the intrinsic flexibility of the C-terminal domain is probably incompatible with the formation of a regular ordered three-dimensional lattice.

We have therefore decided to adopt an alternative structural biology approach and started a collaboration with Prof. H. Molinari, at the University of Verona. The proteins is now been used for NMR experiments. Preliminary results are compatible with a winged helix structure as predicted for the archaeal sequences and as demonstrated also for the human orthologue from MCM6 (Wei *et al*., 2010). Future work include the expression of the protein in isotope-enriched media for more detailed NMR analysis, enabling the determination of the 3D structure of the C-terminal domain of archaeal MCM.

AAA+ proteins need to harness the energy from ATP hydrolysis to carry out mechanical work; to ensure their efficiency, it is therefore important to avoid the useless turn-over of ATP in the absence of the suitable substrate. This task is generally achieved via what has been described as the "Glu switch": an amino-acid residue (typically an asparagine) that holds the catalytic Walker B glutamate in an off position, unless the substrate is present (Zhang and Wigley, 2008). An examination of the structure and sequences of MCM proteins suggests the presence of a more complex 3-residue switch, involving a conserved positively charged residue (Arg/Lys) and a polar residue (Thr/Ser).

To confirm this hypothesis, I have mutated the putative Glu switch residues to alanine in *Mth*MCM, producing three mutant forms: Ser 342 -> Ala (S342A), Lys

344 -> Ala (K344A) and a double mutant, including both changes (SKAA). The three mutant proteins have been expressed and purified to homogeneity, showing a similar behaviour to the wild-type, confirming that the integrity of the complex has not been affected. However, a preliminary biochemical analysis suggested a high (and variable) level of DNA contamination in the sample preparation, which does not allow the proper comparison of the relative properties of the mutants. A new purification protocol will be set up in order to remove the nucleic acids contamination, to verify the role of the above mentioned residues in coordinating DNA binding and ATPase activity.

4.2 Human MCM8

In addition to the universally conserved eukaryotic MCM2-7 complex, two MCM proteins (MCM8 and MCM9) are present in a variety of eukaryotic organisms, but hardly any biochemical or cell biology data is available to elucidate their function within the cell.

I therefore embarked into a long and extensive project to obtain soluble truncated form of hMCM8, by systematically cloning a large number of MCM8 fragments (17) into a variety of bacterial expression vectors, and testing their ability to produce soluble proteins under a range of experimental conditions, including different cell strains, expression temperature and media. I have manage to obtain soluble protein from two deletion mutant: one corresponding to the N-terminal domain of the protein and one that contains the catalytic domain and the C-terminal domain.

As a preliminary step the biochemical properties of the fragments have been determined. Both proteins possess ssDNA binding activity comparable to the DNA binding affinity of archeal MCM properties. The catalytic core shows a ATPase activity stimulated by DNA. Both proteins tends to form dimeric structures and the stoichiometry is not changing in presence of nucleotide or DNA. Although no helicase activity or formation of hexameric structures have been detected so far, these two proteins are the first hMCM8 fragment produced in *E. coli* cells and showing biochemical activities. The lack of evidence for an hexameric structure

(which is expected to be essential for the MCM helicase activity) may be due to a lack of synergic effects in the separated domains but also to the absence of potential partners.

Further work will be carried out on the MCM8 deletion mutants, with the aim of :

1) further characterise and dissect the biochemical activity of the soluble fragments, by designing a variety of site-directed mutants targeting the residues which have been shown to be involved in DNA binding, ATP binding, ATP hydrolysis, helicase activity in the well-studied archaeal MCM orthologues. Such mutants have been already designed and the primers for the site-directed mutagenesis experiments already purchased.

2) optimise the expression protocols for some other constructs and fragments

3) optimise the purification protocol in order to control proteolysis and increase the purity of the protein sample to the level required for structural studies.

Analyzing the evolution of the MCM8 proteins and MCM9, which are either both present or both absence in almost all organisms, it seems evident that these proteins are covelutionary related, making it likely that are also functionally related. An intriguing possibility is that the two proteins form a hetero-hexameric ring, and such ring may be essential for the helicase activity.

In order to assess this hypothesis I started the cloning and the test expression of a large number of hMCM9 fragments and full-length protein. In addition to test the production of soluble hMCM9 fragments and to characterise them, the constructs have been designed to allow the co-expression with the hMCM8 fragments described above. All these proteins will therefore be co-expressed with hMCM8 to test the formation of functional complexes.

.

Appendix 1 Sequence alignment of MCM proteins

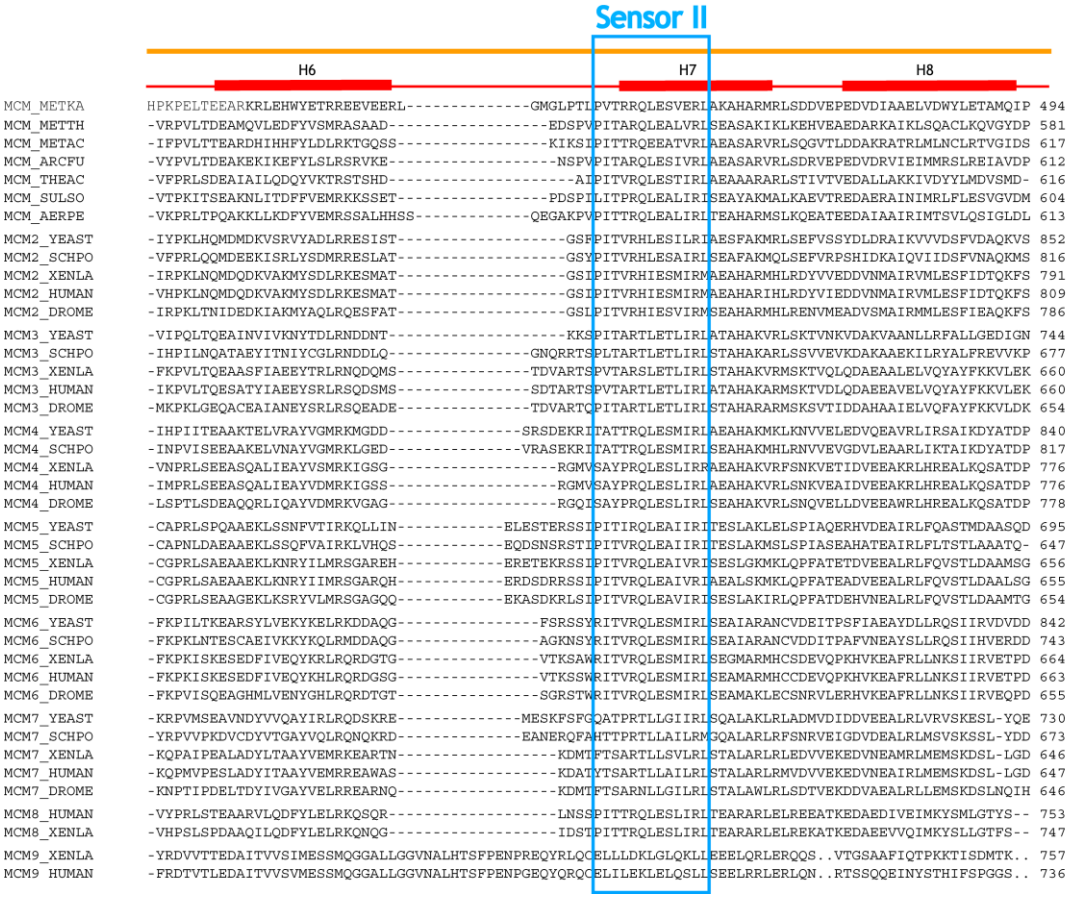
Figure A.1 Sequence alignment of MCM homologues.

Protein sequences of the MCM2 homologue from Methanopyrus kandleri (metka), the MCM homologues from Methanothermobacter thermautotrophicus (metth), Archaeoglobus fulgidus (arcfu), Thermoplasma acidophilum (theac), Sulfolobus solfataricus (sulso), Aeropyrum pernix (aerpe), the MCM2-7 homologues from Saccharomyces cerevisiae (yeast), Schizosaccharomyces pombe (schpo), Xenopus laevis (xenla), Homo sapiens (human), Drosophila melanogaster (drome), and the MCM8-9 homologues from H. sapiens and X. laevis were aligned using the program ClustalW and manually adjusted to obtain a structure-based alignment, using the atomic structure of the N-terminal domain of MthMCM (PDB entry 1LTL) and the AAA+ domain of the MkaMCM as reference. Secondary structure elements are shown as rectangles (α-helices) and arrows (β-strands) and are colour coded as belonging to the N-terminal domain (green) or the AAA+ domain (red). A coloured line above indicated the subdomain structure, with subdomain A in yellow, subdomain B in teal, subdomain C in green, subdomain α/β in red, subdomain α in orange. The location of some of the biochemically important motifs is indicated (NBH, Walker A and B, h2i, PS1BH, Sensor I and II, Arg finger). Highlighted in cyam are the residues coordinating the Zn atom, in yellow the proline and arginine discussed in the text (Pro61 and Arg87 in MthMCM)


α4
β2
β3
β4
β5
β6
β7
β8
MCM_METKA
MCM_METTH
MCM_METAC
MCM_ARCFU
MCM_THEAC
MCM_SULSO
MCM_AERPE
MCM2_YEAST
MCM2_SCHPO
MCM2_XENLA
MCM2_HUMAN
MCM2_DROME
MCM3_YEAST
MCM3_SCHPO
MCM3_XENLA
MCM3_HUMAN
MCM3_DROME
MCM4_YEAST
MCM4_SCHPO
MCM4_XENLA
MCM4_HUMAN
MCM4_DROME
MCM5_YEAST
MCM5_SCHPO
MCM5_XENLA
MCM5_HUMAN
MCM5_DROME
MCM6_YEAST
MCM6_SCHPO
MCM6_XENLA
MCM6_HUMAN
MCM6_DROME
MCM7_YEAST
MCM7_SCHPO
MCM7_XENLA
MCM7_HUMAN
MCM7_DROME
MCM8_HUMAN
MCM8_XENLA
MCM9_XENLA
MCM9_HUMAN


NBH
β9
β10
WA
H01
H02
H03
H04
S1
H1
S2

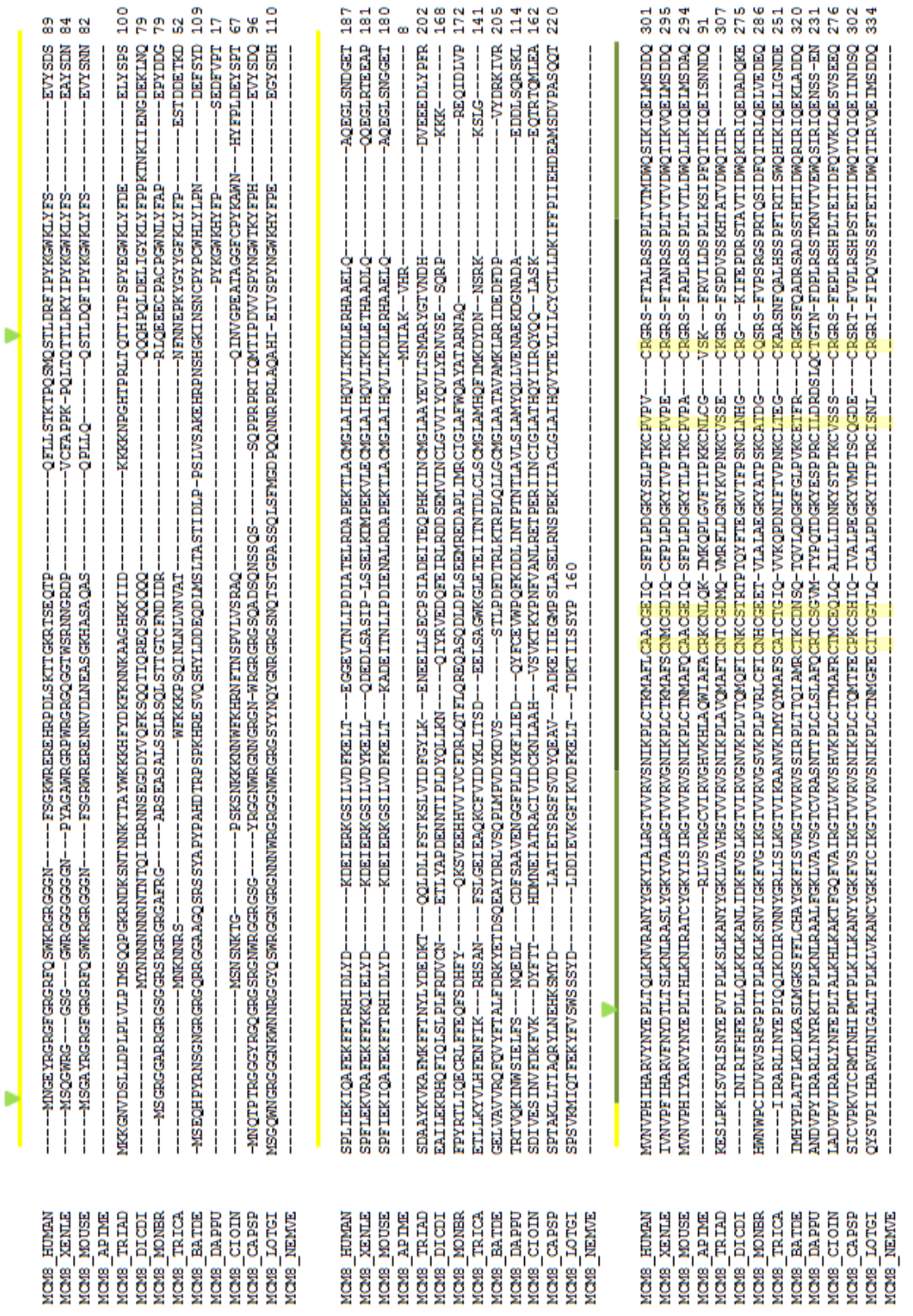

h2i
WB
PS1BH
Sensor I
H2
B1
B2
H2
S3
H3
B3
B4
S4
H05
R
H4
S5
H06

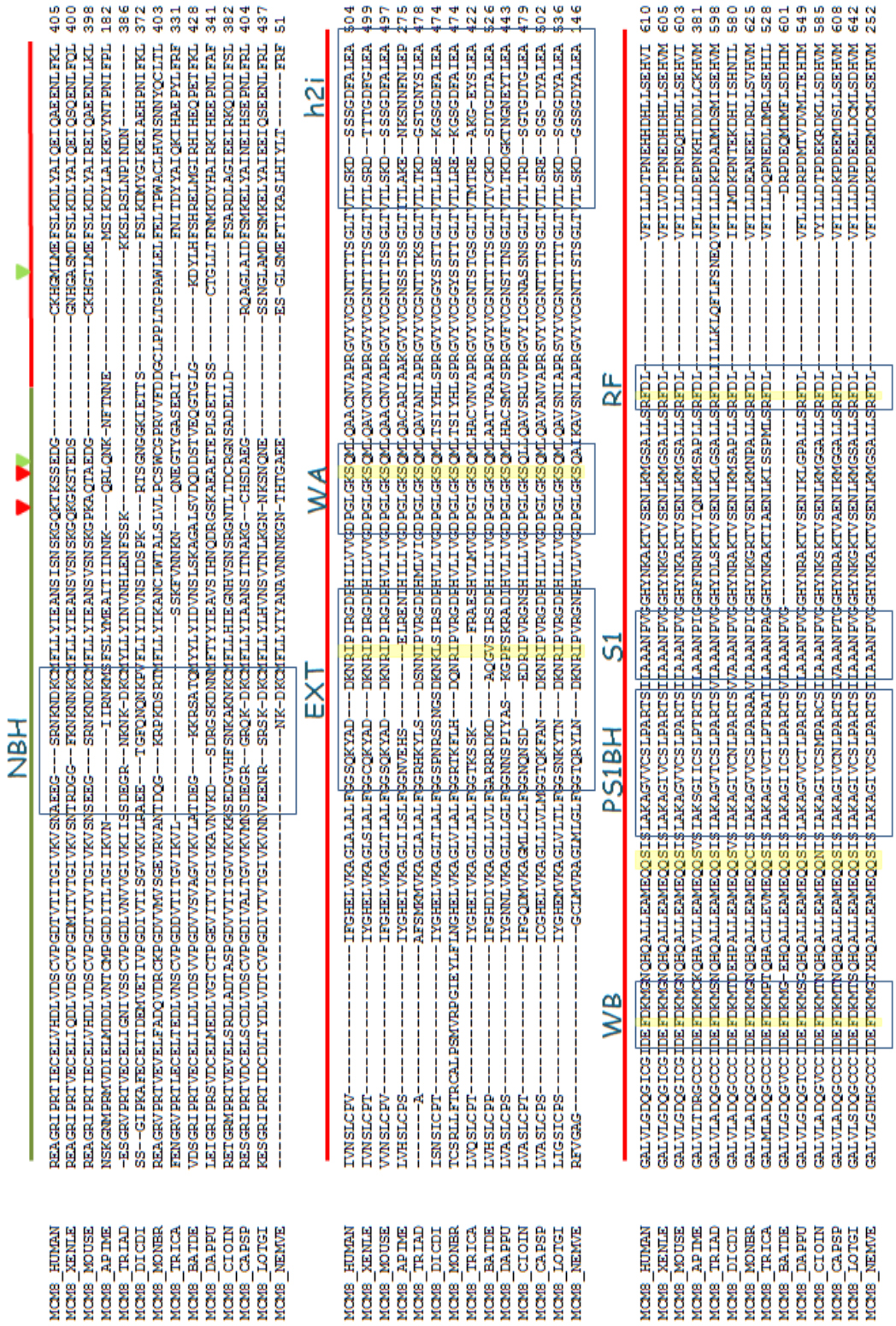

Sensor II
H6
H7
H8
MCM_METKA HPKPELTEEARKRLEHWYETRREEVEERL----------GMGLPTLPVTRRQLESVERLAKAHARMRLSDDVEPEDVDIAAELVDWYLETAMQIP 494
MCM_METTH -VRPVLTDEAMQVLEDFYVSMRASAAD----------EDSPVPITARQLEALVRLSEASAKIKLKEHVEAEDARKAIKLSQACLKQVGYDP 581
MCM_METAC -IFPVLTTEARDHIHHFYLDLRKTGQSS----------KIKSIPITTRQEEATVRLAEASARVRLSQGVTLDDAKRATRLMLNCLRTVGIDS 617
MCM_ARCFU -VYPVLTDEAKEKIKEFYLSLRSRVKE----------NSPVPITARQLESIVRLAEASARVRLSDRVEPEDVDRVIEIMMRSLREIAVDP 612
MCM_THEAC -VFPRLSDEAIAILQDQYVKTRSTSHD----------AIPITVRQLESTIRLAEAAARARLSTIVTVEDALLAKKIVDYYLMDVSMD- 616
MCM_SULSO -VTPKITSEAKNLITDFFVEMRKKSSET----------PDSPILITPRQLEALIRISEAYAKMALKAEVTREDAERAINIMRLFLESVGVDM 604
MCM_AERPE -VKPRLTPQAKKLLKDFYVEMRSSALHHSS----------QEGAKPVPITTRQLEALIRLTEAHARMSLKQEATEEDAIAAIRIMTSVLQSIGLDL 613
MCM2_YEAST -IYPKLHQMDMDKVSRVYADLRRESIST----------GSFPITVRHLESILRIAESFAKMRLSEFVSSYDLDRAIKVVVDSFVDAQKVS 852
MCM2_SCHPO -VFPRLQQMDEEKISRLYSDMRRESLAT----------GSYPITVRHLESAIRLSEAFAKMQLSEFVRPSHIDKAIQVIIDSFVNAQKMS 816
MCM2_XENLA -IRPKLNQMDQDKVAKMYSDLRKESMAT----------GSIPITVRHIESMIRMAEAHARMHLRDYVVEDDVNMAIRVMLESFIDTQKFS 791
MCM2_HUMAN -VHPKLNQMDQDKVAKMYSDLRKESMAT----------GSIPITVRHIESMIRMAEAHARIHLRDYVIEDDVNMAIRVMLESFIDTQKFS 809
MCM2_DROME -IRPKLTNIDEDKIAKMYAQLRQESFAT----------GSIPITVRHIESVIRMSEAHARMHLRENVMEADVSMAIRMMLESFIEAQKFS 786
MCM3_YEAST -VIPQLTQEAINVIVKNYTDLRNDDNT----------KKSPITARTLETLIRLATAHAKVRLSKTVNKVDAKVAANLLRFALLGEDIGN 744
MCM3_SCHPO -IHPILNQATAEYITNIYCGLRNDDLQ----------GNQRRTSPLTARTLETLIRLSTAHAKARLSSVVEVKDAKAAEKILRYALFREVVKP 677
MCM3_XENLA -FKPVLTQEAASFIAEEYTRLRNQDQMS----------TDVARTSPVTARSLETLIRLSTAHAKVRMSKTVQLQDAEAALELVQYAYFKKVLEK 660
MCM3_HUMAN -IKPVLTQESATYIAEEYSRLRSQDSMS----------SDTARTSPVTARTLETLIRLATAHAKARMSKTVDLQDAEEAVELVQYAYFKKVLEK 660
MCM3_DROME -MKPKLGEQACEAIANEYSRLRSQEADE----------TDVARTQPITARTLETLIRLSTAHARARMSKSVTIDDAHAAIELVQFAYFKKVLDK 654
MCM4_YEAST -IHPIITEAAKTELVRAYVGMRKMGDD----------SRSDEKRITATTRQLESMIRLAEAHAKMKLKNVVELEDVQEAVRLIRSAIKDYATDP 840
MCM4_SCHPO -INPVISEEAAKELVNAYVGMRKLGED----------VRASEKRITATTRQLESMIRLSEAHAKMHLRNVVEVGDVLEAARLIKTAIKDYATDP 817
MCM4_XENLA -VNPRLSEEASQALIEAYVSMRKIGSG----------RGMVSAYPRQLESLIRRAEAHAKVRFSNKVETIDVEEAKRLHREALKQSATDP 776
MCM4_HUMAN -IMPRLSEEASQALIEAYVDMRKIGSS----------RGMVSAYPRQLESLIRLAEAHAKVRLSNKVEAIDVEEAKRLHREALKQSATDP 776
MCM4_DROME -LSPTLSDEAQQRLIQAYVDMRKVGAG----------RGQISAYPRQLESLIRLSEAHAKVRLSNQVELLDVEEAWRLHREALKQSATDP 778
MCM5_YEAST -CAPRLSPQAAEKLSSNFVTIRKQLLIN----------ELESTERSSIPITIRQLEAIIRITESLAKLELSPIAQERHVDEAIRLFQASTMDAASQD 695
MCM5_SCHPO -CAPNLDAEAAEKLSSQFVAIRKLVHQS----------EQDSNSRSTIPITVRQLEAIIRITESLAKMSLSPIASEAHATEAIRLFLTSTLAAATQ- 647
MCM5_XENLA -CGPRLSAEAAEKLKNRYILMRSGAREH----------ERETEKRSSIPITVRQLEAIVRISESLGKMKLQPFATETDVEEALRLFQVSTLDAAMSG 656
MCM5_HUMAN -CGPRLSAEAAEKLKNRYIIMRSGARQH----------ERDSDRRSSIPITVRQLEAIVRIAEALSKMKLQPFATEADVEEALRLFQVSTLDAALSG 655
MCM5_DROME -CGPRLSEAAGEKLKSRYVLMRSGAGQQ----------EKASDKRLSIPITVRQLEAVIRISESLAKIRLQPFATDEHVNEALRLFQVSTLDAAMTG 654
MCM6_YEAST -FKPILTKEARSYLVEKYKELRKDDAQG----------FSRSSYRITVRQLESMIRLSEAIARANCVDEITPSFIAEAYDLLRQSIIRVDVDD 842
MCM6_SCHPO -FKPKLNTESCAEIVKKYKQLRMDDAQG----------AGKNSYRITVRQLESMIRLSEAIARANCVDDITPAFVNEAYSLLRQSIIHVERDD 743
MCM6_XENLA -FKPKISKESEDFIVEQYKRLRQRDGTG----------VTKSAWRITVRQLESMIRLSEGMARMHCSDEVQPKHVKEAFRLLNKSIIRVETPD 664
MCM6_HUMAN -FKPKISKESEDFIVEQYKHLRQRDGSG----------VTKSSWRITVRQLESMIRLSEAMARMHCCDEVQPKHVKEAFRLLNKSIIRVETPD 663
MCM6_DROME -FKPVISQEAGHMLVENYGHLRQRDTGT----------SGRSTWRITVRQLESMIRLSEAMAKLECSNRVLERHVKEAFRLLNKSIIRVEQPD 655
MCM7_YEAST -KRPVMSEAVNDYVVQAYIRLRQDSKRE----------MESKFSFGQATPRTLLGIIRLSQALAKLRLADMVDIDDVEEALRLVRVSKESL-YQE 730
MCM7_SCHPO -YRPVVPKDVCDYVTGAYVQLRQNQKRD----------EANERQFAHTTPRTLLAILRMGQALARLRFSNRVEIGDVEALRLMSVSKSSL-YDD 673
MCM7_XENLA -KQPAIPEALADYLTAAYVEMRKEARTN----------KDMTFTSARTLLSVLRLSTALARLRLEDVVEKEDVNEAMRLMEMSKDSL-LGD 646
MCM7_HUMAN -KQPMVPESLADYITAAYVEMRREAWAS----------KDATYTSARTLLAILRLSTALARLRMVDVVEKEDVNEAIRLMEMSKDSL-LGD 647
MCM7_DROME -KNPTIPDELTDYIVGAYVELRREARNQ----------KDMTFTSARNLLGILRLSTALAWLRLSDTVEKDDVAEALRLLEMSKDSLNQIH 646
MCM8_HUMAN -VYPRLSTEAARVLQDFYLELRKQSQR----------LNSSPITTRQLESLIRLTEARARLELREEATKEDAEDIVEIMKYSMLGTYS-- 753
MCM8_XENLA -VHPSLSPDAAQILQDFYLELRKQNQG----------IDSTPITTRQLESLIRLTEARARLELREKATKEDAEEVVQIMKYSLLGTFS-- 747
MCM9_XENLA -YRDVVTTEDAITVVSIMESSMQGGALLGGVNALHTSFPENPREQYRLQCELLLDKLGLQKLLEEELQRLERQQS..VTGSAAFIQTPKKTISDMTK.. 757
MCM9_HUMAN -FRDTVTLEDAITVVSVMESSMQGGALLGGVNALHTSFPENPGEQYQRQCELILEKLELQSLLSEELRRLERLQN..RTSSQQEINYSTHIFSPGGS.. 736

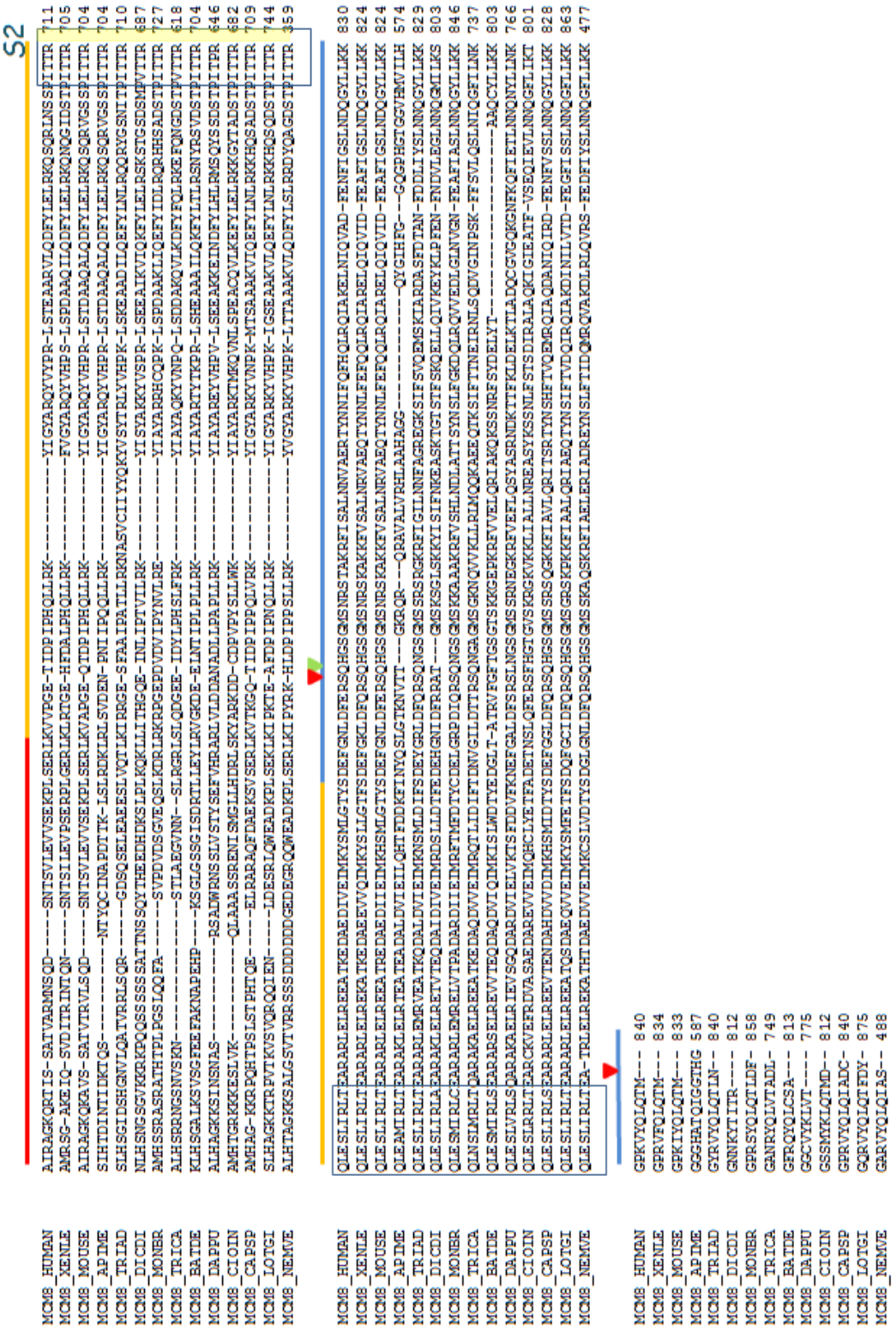
Appendix 2 Sequence alignment of MCM8 and MCM9 proteins

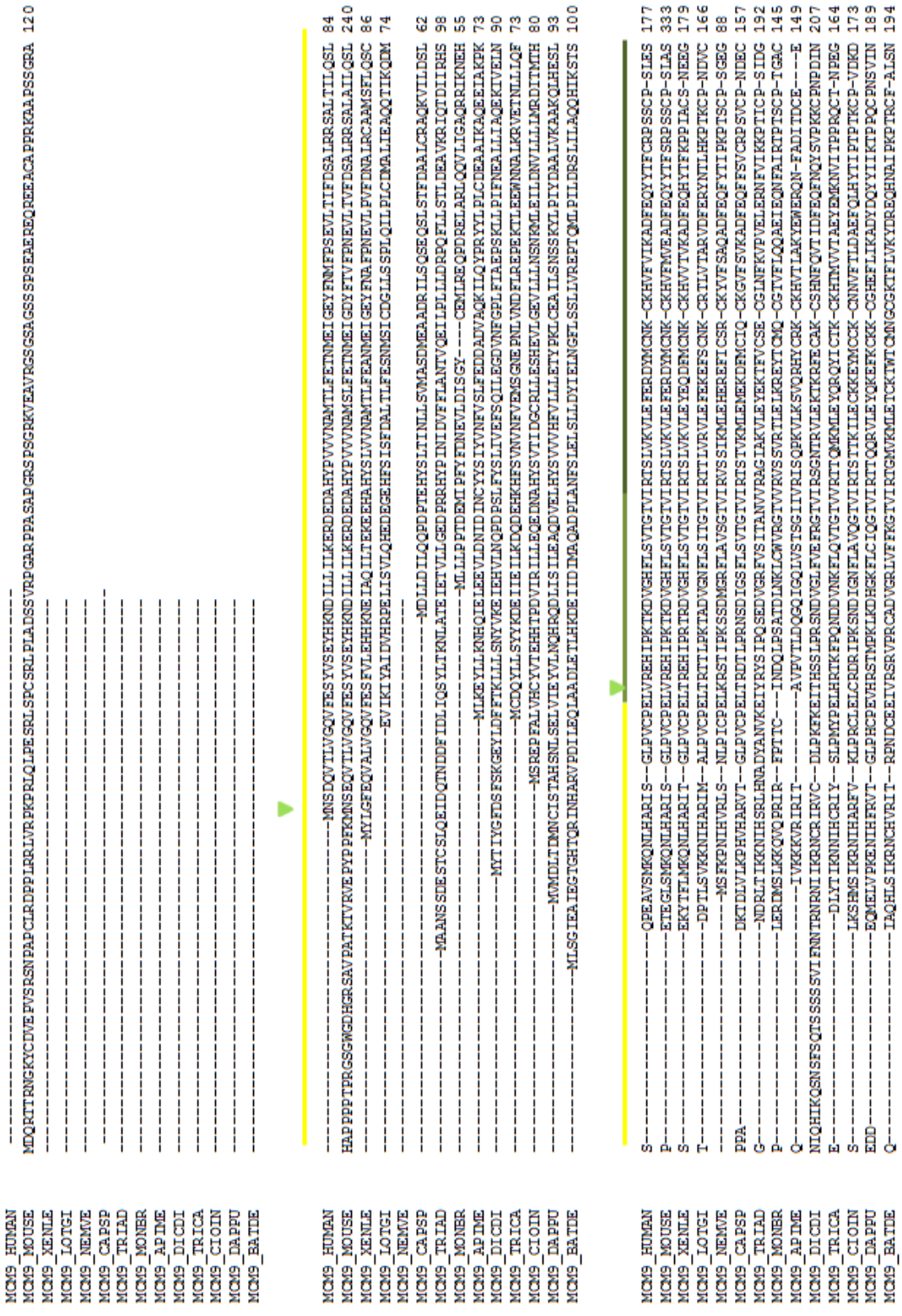
Figure A.2 Sequence alignment of MCM8 and MCM9 homologues.

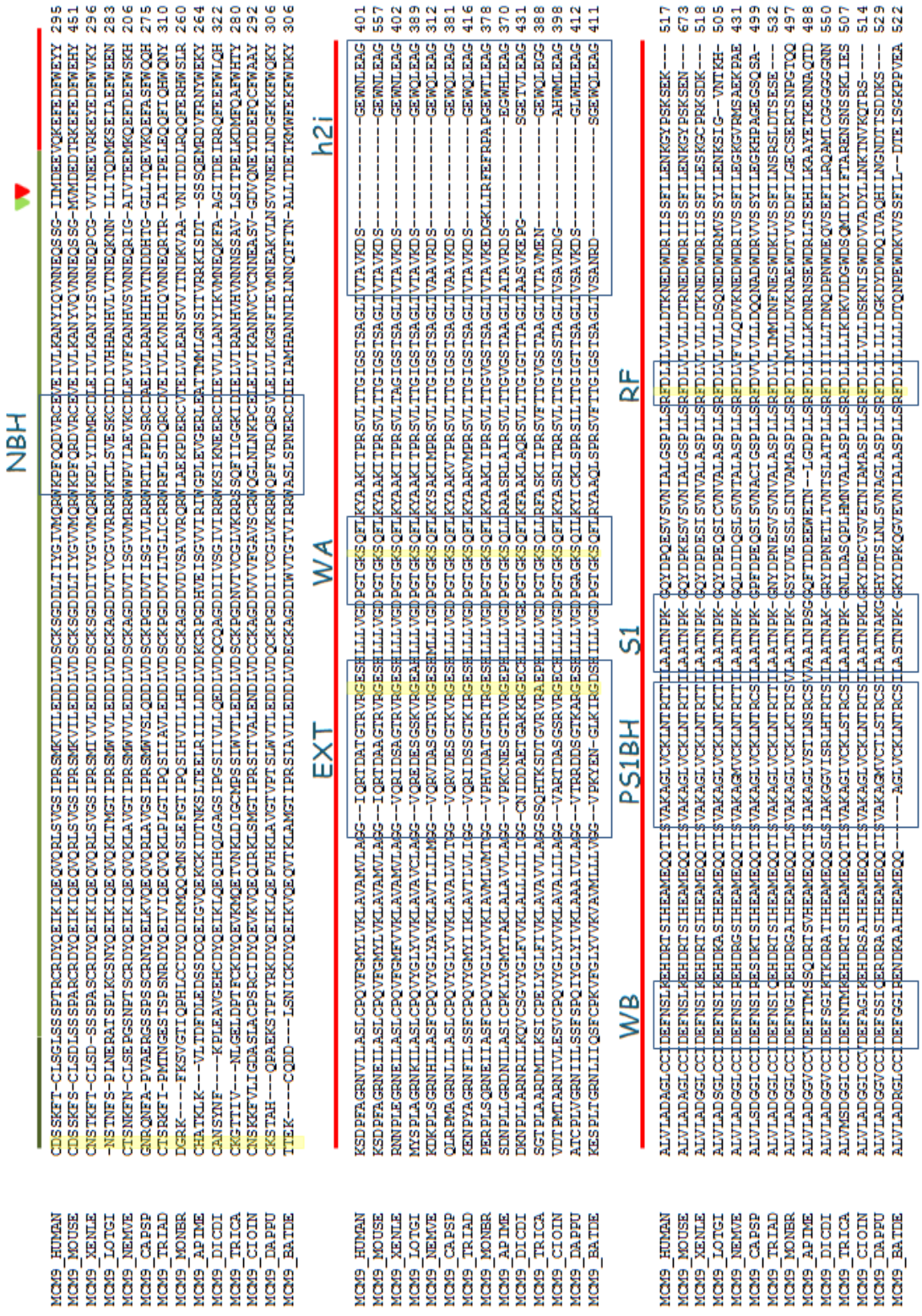
Protein sequences of the MCM8 and MCM9 orthologues from human (HUMAN), Xenopus laevis (XENLE), mouse (MOUSE), Apis mellifera (APIME), Trichoplax adhaerens(TRIAD) Dictyostelium discoideum AX4(DICDI), Monosiga brevicollis MX1 (MONBR), Tribolium castaneum (TRICA), batrachochytrium dendrobatidis (BATDE), Daphnia pulex (DAPPU), Ciona Intestinalis (CIOIN), Capitella sp (CAPSP), lottia gigantea (LOTGI) Nemotostella vectensis (NEMVE) were aligned using the program ClustalW and manually adjusted to obtain a structure-based alignment, using the atomic structure of the N-terminal domain of MthMCM (PDB entry 1LTL) and the AAA+ domain of the MkaMCM as reference. The domain structure is indicated by the same scheme described in Appendix 1. Green triangles indicate the starting point of the construct produced for this work, whereas the red triangles indicate the ending point. In yellow are highlighted crucial residues,. The canonical motifs shared by all MCM proteins are enclosed by boxes.

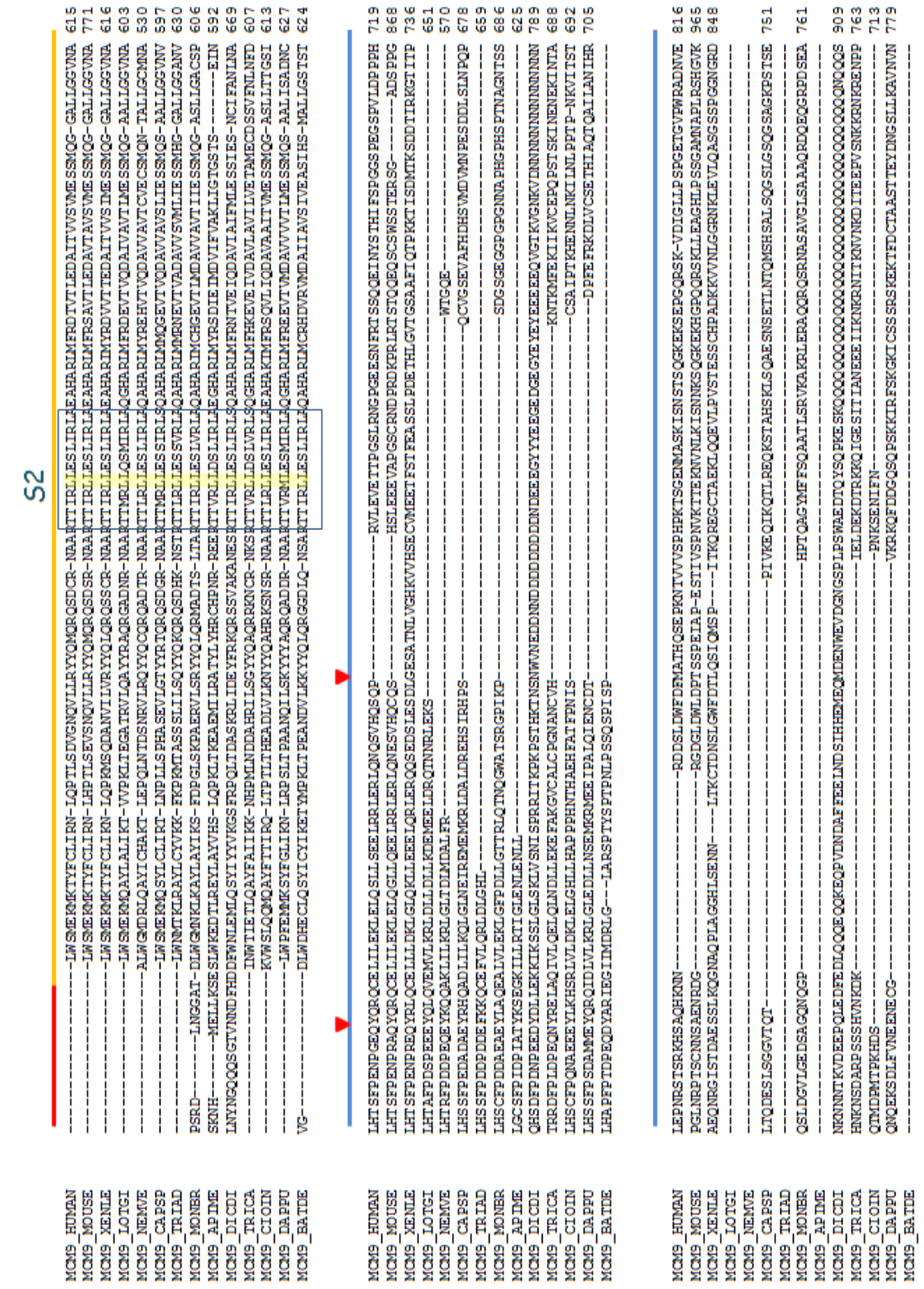



NBH
EXT
WA
h2i
WB
PS1BH
S1
RF
MCM8_HUMAN
MCM8_XENLE
MCM8_MOUSE
MCM8_APIME
MCM8_TRIAD
MCM8_DICDI
MCM8_MONBR
MCM8_TRICA
MCM8_BATDE
MCM8_DAPPU
MCM8_CIOIN
MCM8_CAPSP
MCM8_LOTGI
MCM8_NEMVE


S2




NBH
EXT
WA
h2i
WB
PS1BH
S1
RF
MCM9_HUMAN
MCM9_MOUSE
MCM9_XENLE
MCM9_LOTGI
MCM9_NEMVE
MCM9_CAPSP
MCM9_TRIAD
MCM9_MONBR
MCM9_APIME
MCM9_DICDI
MCM9_TRICA
MCM9_CIOIN
MCM9_DAPPU
MCM9_BATDE


S2

Appendix 3 Reagents, buffers and solutions

Reagents used

PfuTurbo DNA polymerase and PfuTurbo buffer were obtained from Stratagene. dNTPs, restriction endonucleases, restriction buffers, BSA for restriction digestions and Calf Intestinal Phosphatase (CIP) were from New England Biolabs. The Rapid DNA Ligation Kit was purchased from Roche. IPTG (dioxan free) and DTT were from Melford Laboratories. Lysozyme (from chicken egg white), glutathione (reduced form), thrombin (lyophilised powder), PMSF, βME, BSA (for Western Blots) and gel-filtration molecular weight markers were all purchased from Sigma. Benzonase was from Merck.

All other buffers, salts and reagents used for the preparation of solutions were purchased from Sigma.

Composition of solutions

Media

LB medium (LB): LB powder (Lennox L Broth, Sigma) was dissolved in deionised water (20 g per L of water) and autoclaved for 20 min at 121ºC. The medium was stored at room temperature and the necessary antibiotics were added to the specified final concentrations prior to use.

Terrific broth (TB): Terrific broth (Modified, Sigma) was dissolved in deionised water (47 g per L of water) and 8 ml glycerol per L of medium was also added prior to autoclaving (20 min at 121ºC). The medium was stored at room temperature and the necessary antibiotics were added to the specified final concentrations prior to use.

Autoinduction Terrific broth (TB): Terrific broth (Modified, Sigma) was dissolved in deionised water (47.6 g per L of water) and 8 ml glycerol, 5gr of Lactose, 0.15gr of Glucose, 0.49 gr of $MgSO_4$ heptahydrate per L of medium was also added prior to

autoclaving (20 min at 121ºC). The medium was stored at room temperature and the necessary antibiotics were added to the specified final concentrations prior to use.

LB Agar and agar plates: 5.25 g of LB agar powder (Lennox L agar, Sigma) were added to 150 ml deionised water and autoclaved for 20 min at 121ºC. The medium was allowed to cool to 50ºC before adding antibiotics to the specified final concentrations. 20 ml of medium were poured into 85 mm petri dishes. If necessary, the surface of the medium was flamed with a Bunsen burner to eliminate bubbles. The medium was left to harden for one hour and the agar plates were stored at 4ºC for ≤ 1 month.

Antibiotic stock solutions (1000x)

Ampicillin: 1 gr of ampicillin sodium salt (Melford Laboratories) were dissolved in 10 ml deionised water, filter-sterilised and stored at -20ºC. Ampicillin was added to growth medium to a final concentration of 50 μg/ml.

Kanamycin: 50 mg of kanamycin monosulphate (Melford Laboratories) were dissolved in 10 ml deionised water, filter-sterilised and stored at -20ºC. Kanamycin was added to growth medium to a final concentration of 25 μg/ml.

Chloramphenicol: 340 mg chloramphenicol (Sigma) were dissolved in 10 ml ethanol and the aliquots were wrapped in foil and stored at -20ºC. Chloramphenicol was added to growth medium to a final concentration of 34 μg/ml.

Buffers used for the preparation of competent *E. coli* cells

TFB1: 100 mM RbCl, 50 mM $MnCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2$, 15% glycerol, pH 5.8 (adjusted with acetic acid). The solution was filter-sterilized and stored at 4ºC.

TFB2: 10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$, 15% glycerol, pH 6.8 (adjusted with KOH). The solution was filter-sterilized and stored at 4ºC.

Buffers used for the purification and crystallisation of recombinant proteins

All buffers used for the purification of recombinant proteins were filtered through a 0.22 μm membrane filter prior to use.

Affinity chromatography buffers

MBP Buffer : 30 mM Tris-HCl pH 7.4, 300 mM NaCl, 5% Glycerol, 1 mM EDTA

Nickel A Buffer: 20 mM Tris-HCl pH 7.9, 500 mM NaCl, 10% glycerol,

Nickel B Buffer: 20 mM Tris-HCl pH 7.9, 500 mM NaCl, 10% glycerol, 1 M imidazole

charge solution: 0.1 M $CoSO_4$ or 0.1 M $NiSO_4$

Ion exchange chromatography buffers

Heparin Buffer A: 20 mM Hepes-NaOH pH 7, 50 mM NaCl

Heparin Buffer B: 20 mM Hepes-NaOH pH 7, 1 M NaCl

Heparin Buffer C: 50 mM Tris-HCl pH 7.4, 500 mM NaCl, 10% Glycerol

Heparin Buffer D: 50 mM Tris-HCl pH 7.4, 2 M NaCl, 10% Glycerol

Size exclusion chromatography buffers

GF Buffer A: 30 mM Hepes-NaOH pH 7.5, 150 mM NaCl

GF Buffer B: 30 mM Hepes pH 7.5, 300 mM NaCl, 10% Glycerol

NMR filtration Buffer: 30 mM phosphate buffer pH 6.5, 150 mM NaCl

TEV cleavage Buffer: 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.5 mM EDTA and 10% glycerol

6x DNA loading Buffer: 50% (v/v) glycerol, 100 mM $Na_2EDTA \cdot 2H2O$, 1% (w/v) SDS, 0.1% (w/v) bromophenol blue

SDS-PAGE buffers

2x SDS Sample Buffer: 0.125 M Tris-HCl, pH 6.8, 20% (v/v) glycerol, 0.04% (w/v) SDS, 10% (v/v) βME, 0.1% (w/v) bromophenol blue

6x SDS Sample Buffer: 0.35 M Tris-HCl pH 6.8, 10% SDS, 30% glycerol, 9.3% DTT

Tris Glycine SDS Separating gel solution: 0.375 M Tris-HCl, pH 8.8, 0.001% lauryl sulphate (SDS), , 10-15% Acrylamide/Bis, 0.0005% ammonium persulphate (APS) and 0.0001% N,N,N',N'-tetramethylethylenediamine (TEMED).

Tris Glycine SDS Stacking gel solution: 0.125 M Tris-HCl, pH 6.8, 0.001% SDS, , 4% Acrylamide/Bis, 0.0005% APS and 0.0001% TEMED.

10x Tris-glycine Buffer: 0.25 M Tris Base, 1.92 M Glycine, 10% SDS. The buffer was diluted 10-fold to use for electrophoresis.

Tris Tricine SDS Separating gel solution: 1 M Tris-HCl, pH 8.5, 0.001% lauryl sulphate (SDS), , 10-15% Acrylamide/Bis, 0.0005% ammonium persulphate (APS) and 0.0001% N,N,N',N'-tetramethylethylenediamine (TEMED).

Tris Tricine SDS Stacking gel solution: 0.75 M Tris-HCl, pH 8.5, 0.001% SDS, , 4% Acrylamide/Bis, 0.0005% APS and 0.0001% TEMED

10X Cathode Buffer: 1 M Tris Base, 1 M Tricine, 1 % SDS

10X Anode Buffer: 2 M Tris, pH 8.9

Agarose gel buffer

10x TBE: 890 mM Tris Base, 890 mM Boric Acid, 20 mM EDTA pH 8**5x DNA Sample Buffer:** 0.1% xylene cyanol, 0.1% bromphenol blue, 50% glycerol

Other

Primers: The cloning primers used in this study (Table 2.2) were supplied as desalted, lyophilised powders (Sigma-Genosys, UK). The primers used for site-directed mutagenesis (Table 2.6) were purchased from MWG-Biotech (Germany). Primers were resuspended in deionised water to a final concentration of 100 μM. A further dilution was made to produce a working stock of 10 μM. Stock solutions were stored at -20ºC.

IPTG: IPTG powder (Carbosynth) was dissolved in deionised water to a final concentration of 1 M. The solution was filter sterilised, wrapped in foil and stored at -20ºC.

PMSF: PMSF (Sigma Aldrich) powder was dissolved in ethanol to a final concentration of 100 mM. The stock was stored at 4ºC.

AEBSF: PMSF (Sigma Aldrich) powder was dissolved in water to a final concentration of 200 mM. The stock was stored at -20ºC.

DTT: DTT powder (Sigma Aldrich) was dissolved in deionised water to a final concentration of 1 M. The solution was filter sterilised, divided in 1.5 ml aliquots and stored at -20ºC.

Appendix 4 Experimental protocols

Protocol A – PCR

Experimental conditions for the amplification of the genes and gene fragments listed in Table

Reagent	**Volume**	**Stock concentration**
Template DNA	1µl	100-250ng/µl gen. DNA, 20-25ng/µl of plasmid DNA
dNTP mix	1µl	
10X Pfu Turbo polymerase buffer	5µl	
Primer reverse	1µl	10µM
Primer forward	1µl	10µM
Pfu Turbo polymerase	1µl	
Milli Q H_2O	40µl	

PCR program

5 min initial denaturation at 94°C
Then 10 cycles:
45" denaturation 94°C
45" annealing 45°C
1'/kB extension 72°C
Then 35 cycles: 45" denaturation 94°C
45"-1' annealing 56°-68°C
2-3' extension 72°C
then the final 8-10 min extension

Protocol B – preparation of competent E. coli cells

The following protocol was used for the preparation of the following E. coli strains from commercial original stocks: XL1-Blue (Stratagene), BL21(DE3) (Stratagene), BL21-CodonPlus(DE3)-RIL (Stratagene) and B834(DE3) (Novagen) competent cells. The original stocks for each bacterial strain were kept in aliquots at -80ºC.

All media used for the preparation of XL1-Blue, BL21 (DE3) and B834 competent cells, were not supplemented with antibiotics. For the preparation of BL21-CodonPlus(DE3)-RIL cells, however, 34 μg/ml chloramphenicol was added to all media. The compositions of buffers TFB1 and TFB2 that were used in this protocol are provided in Appendix 3.

A trace of competent cells was removed from a stock vial with a sterile tip and streaked out on an agar plate and incubated at 37ºC overnight. Next, a single colony from the agar plate was inoculated in 10 ml LB medium and the culture was grown overnight at 37ºC with shaking at approximately 225 rpm. On the following day, 1 ml overnight culture was used to inoculate 100 ml prewarmed LB medium in a 250 ml flask. The cells were grown until an A_{600nm} of 0.5 was reached (typically 4 hr for XL1-Blue cells and 2 hr for the other three strains). The culture was cooled on ice for 5 min and then transferred to a sterile 50 ml Falcon tube. The cells were pelleted by centrifugation at 4000 g for 5 min at 4ºC. The supernatant was carefully decanted and the cells were resuspended gently in 30 ml cold (4ºC) TFB1 buffer and incubated on ice for 90 min. The cells were collected by centrifugation at 4000 g for 5 min at 4ºC and the supernatant was carefully decanted. Using chilled pipettes, the cells were gently resuspended in 4 ml ice-cold TFB2 buffer. 200 μl aliquots of cells were prepared using prechilled sterile tubes. The aliquots were then frozen in a dry-ice/ethanol bath and stored at -80ºC.

Protocol C – PCR-Based Site-Directed Mutagenesis

All reactions were performed using the reagents provided in the QuikChange XL Site-Directed Mutagenesis Kit (Stratagene), except for the dNTP mix, which was prepared from dNTPs purchased from New England Biolabs. The experimental conditions for the mutagenesis reactions which produced the mjElp3 cysteine mutants are the following:

QuikSolution 3 μl

reaction buffer 10x 10 μl

primer #1 10 μM 125 ng

primer #2 10 μM 125 ng

dNTPs mix (12.5 mM)* 1 μl

template dsDNA† 25 ng

Deionised water added to a final volume of 50 µl

PfuTurbo polymerase 1 µl (2.5 U)

The reactions were cycled using the following cycling parameters: 1 cycle at 95°C, for 1 min; 18 cycles comprising 50 sec at 95°C, 50 sec at 60°C and 1.5 min/kb extension time at 68°C; 1 cycle and 68°C for 7 min.

Protocol D – preparation of TEV protease

The gene encoding the TEV protease was cloned into the expression vector pET24a (Novagen) that confers kanamycin resistance to the host cell, resulting in a recombinant protein with an C-terminal His-tag. The construct was used to transform *E coli* BL21-(DE3) Codon + RIPL cells. Cells were grown in Autoinduction TB medium (supplemented with kanamycin). Cells were grown for 24 hr at 25°C and then harvested by centrifugation at 3,000 g for 20 min at 4°C. The pellet was then stored at -80°C.

Cell resuspension and protein purification were carried out according to the purification protocols described in Sections 2.2.1 and 2.2.2, using a Ni^{2+} chelating column (5 ml, GE Healthcare). Fractions containing the rTEV protease were pooled and then the buffer was exchanged against 20 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM βME and 10% glycerol. Protein samples were stored at -80°C. The final yield of purified rTEV protease was approximately 3 mg of protein per L of cell culture

Protocol E – RT PCR

First-Strand cDNA Synthesis

A 20-µl reaction volume used for 1 ng–5 µg of total RNA

1 µl oligo $(dT)_{16}$

1 µg total RNA

1 µl 10 mM dNTP Mix (

Sterile, distilled water to 12 µl

2. Heat mixture to 65°C for 5 min and quick chill on ice. Collect the contents of the tube by brief centrifugation and add:

4 µl 5X First-Strand Buffer

2 µl 0.1 M DTT

1 µl RNaseOUT Recombinant Ribonuclease Inhibitor (40 units/µl)

Mix contents of the tube gently and incubate at 37°C for 2 min.

Add 1 µl (200 units) of M-MLV RT and mix by pipetting gently .

Incubate 50 min at 37°C.

Inactivate the reaction by heating at 70°C for 15 min.

The cDNA can now be used as a template for amplification in PCR (protocol A)

Protocol F – RF PCR

Reagent	**Volume**
dNTP mix	1µl
5X Phusion polymerase buffer (NEB)	10µl
Megaprimer (2 µM)	1µl
Vector (200 pM)	1µl
Phusion polymerase	0.8 µl
$MgCl_2$ (50 mM)	2 µl
Milli Q H_2O	Up to 50 µl

PCR program

30" initial denaturation at 95°C
Then 30 cycles:
30" denaturation 95°C
1 min annealing 60°C
5 min extension 72°C
then the final 7 min extension

5 References

Aparicio T., Guillou E., Coloma J., Montoya G., Mendez J. (2009). The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication. Nucleic Acids Res. 37, 2087-2095.

Aparicio, O. M., Stout, A. M., and Bell, S. P. (1999). Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. Proc Natl Acad Sci U S A 96, 9130-9135.

Aparicio, O. M., Weinstein, D. M., and Bell, S. P. (1997). Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. Cell 91, 59-69.

Araki H. (2010). Regulatory mechanism of the initiation step of DNA replication by CDK in budding yeast. Biochim. Biophys. Acta 1804, 520-523.

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

Bae B., Chen Y.H., Costa A., Onesti S., Brunzelle J.S., Lin Y., Cann I.K., Nair S.K. (2009). Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog. Structure 17, 211-222.

Barry ER, Lovett JE, Costa A, Lea SM and Bell SD. (2009). Intersubunit allosteric communication mediated by a conserved loop in the MCM helicase. Proc Natl Acad Sci USA 106:1051–1056.

Barry, E. R., McGeoch, A. T., Kelman, Z., and Bell, S. D. (2007). Archaeal MCM has separable processivity, substrate choice and helicase domains. Nucleic Acids Res 35, 988-998.

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

Blanton HL, Radford SJ, McMahan S, Kearney HM, Ibrahim JG and Sekelsky J. (2005). REC, Drosophila MCM8, drives formation of meiotic crossovers. PLoS Genet 1:e40.

Blow JJ and Dutta A. (2005). Preventing re-replication of chromosomal DNA. Nat Rev Mol Cell Biol 6:476–486

Blow, J. J. (2001). Control of chromosomal DNA replication in the early *Xenopus* embryo. EMBO J 20, 3293-3297.

Blow, J. J., and Hodgson, B. (2002). Replication licensing-defining the proliferative state? Trends Cell Biol 12, 72-78.

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

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

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

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

Brewster A.S., Wang G., Yu X., Greenleaf W.B., Carazo J.M., Tjajadia M., Klein M.G., Chen X.S. (2008). Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase. Proc. Natl. Acad. Sci. U.S.A. 105, 20191-20196.

Brewster AS, Slaymaker IM, Afif SA, Chen XS. (2010) Mutational analysis of an archaeal minichromosome maintenance protein exterior hairpin reveals critical residues for helicase activity and DNA binding. BMC Mol Biol. 18:11:62.

Carpentieri, F., De Felice, M., De Falco, M., Rossi, M., and Pisani, F. M. (2002). Physical and functional interaction between the mini-chromosome maintenance-like DNA helicase and the single-stranded DNA binding protein from the crenarchaeon *Sulfolobus solfataricus*. J Biol Chem 277, 12118-12127.

Chen, Y. J., Yu, X., and Egelman, E. H. (2002). The hexameric ring structure of the *Escherichia coli* RuvB branch migration protein. J Mol Biol 319, 587-591.

Chen, Y. J., Yu, X., Kasiviswanathan, R., Shin, J. H., Kelman, Z., and Egelman, E. H. (2005). Structural polymorphism of *Methanothermobacter thermautotrophicus* MCM. J Mol Biol 346, 389-394.

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.

Chou D.M., Elledge S.J. (2006). Tipin and Timeless form a mutually protective complex required for genotoxic stress resistance and checkpoint function. Proc. Natl. Acad. Sci. U.S.A. 103, 18143-18147.

Costa A., Onesti S. (2008). The MCM complex: (just) a replicative helicase? Biochem. Soc. Trans. 36, 136-140.

Costa A., Onesti S. (2009). Structural biology of MCM helicases. Crit. Rev. Biochem. Mol. Biol. 44, 326-342.

Costa A., van Duinen G., Medagli B., Chong J., Sakakibara N., Kelman Z., Nair S.K., Patwardhan A., Onesti S. (2008). Cryo-electron microscopy reveals a novel DNA-binding site on the MCM helicase. EMBO J. 27, 2250-2258.

Costa, A., Pape, T., van Heel, M., Brick, P., Patwardhan, A., and Onesti, S. (2006a). Structural basis of the *Methanothermobacter thermautotrophicus* MCM helicase activity. Nucleic Acids Res 34, 5829-5838.

Costa, A., Pape, T., van Heel, M., Brick, P., Patwardhan, A., and Onesti, S. (2006b). Structural studies of the archaeal MCM complex in different functional states. J Struct Biol 156, 210-219.

Crevel G, Hashimoto R, Vass S, Sherkow J, Yamaguchi M, Heck MM and Cotterill S. (2007). Differential requirements for MCM proteins in DNA replication in *Drosophila* S2 cells. PLoS ONE 2:e833.

Crevel G, Ivetic A, Ohno K, Yamaguchi M and Cotterill S. (2001). Nearest neighbor analysis of MCM protein complexes in *Drosophila melanogaster*. Nucleic Acids Res 29:4834–4842.

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

DeLano, W. L. (2002). The Pymol Molecular Graphics system, DeLano Scientific, San Carlos, CA, USA.

Diffley JF and Labib K. (2002). The chromosome replication cycle. J Cell Sci. 115:869–872.

Donovan S, Harwood J, Drury LS and Diffley JF. (1997). Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. Proc Natl Acad Sci USA 94:5611–5616.

Dueber, E. L., Corn, J. E., Bell, S. D., and Berger, J. M. (2007). Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. Science 317, 1210-1213.

Enemark EJ and Joshua-Tor L. (2006). Mechanism of DNA translocation in a replicative hexameric helicase. Nature 442:270–275.

Enemark EJ and Joshua-Tor L. (2008). On helicases and other motor proteins.Curr Opin Struct Biol 18:243–57.

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

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

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.

Fien, K., and Hurwitz, J. (2006). Fission yeast Mcm10p contains primase activity. J Biol Chem 281, 22248-22260.

Fletcher, R. J., and Chen, X. S. (2006). Biochemical activities of the BOB1 mutant in *Methanobacterium thermoautotrophicum* MCM. Biochemistry (Mosc) 45, 462-467.

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

Fletcher, R. J., Shen, J., Gomez-Llorente, Y., Martin, C. S., Carazo, J. M., and Chen, X. S. (2005). Double hexamer disruption and biochemical activities of *Methanobacterium thermoautotrophicum* MCM. J Biol Chem 280, 42405-42410.

Forsburg, S. L. (2004). Eukaryotic MCM proteins: beyond replication initiation. Microbiol Mol Biol Rev 68, 109-131.

Fu YV, Walter JC. (2010) DNA replication: metazoan Sld3 steps forward. Curr Biol. 22;20:R515-7.

Funnell, B. E., Baker, T. A., and Kornberg, A. (1987). In vitro assembly of a prepriming complex at the origin of the *Escherichia coli* chromosome. J Biol Chem 262, 10327-10334.

Gambus A., Jones R.C., Sanchez-Diaz A., Kanemaki M., van Deursen F., Edmondson R.D., Labib K. (2006). GINS maintains association of Cdc45 with MCM

in replisome progression complexes at eukaryotic DNA replication forks. Nat Cell Biol 8, 358-366.

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

Gaudier, M., Schuwirth, B. S., Westcott, S. L., and Wigley, D. B. (2007). Structural Basis of DNA Replication Origin Recognition by an ORC Protein. Science 317, 1213-1216.

Gomez-Llorente, Y., Fletcher, R. J., Chen, X. S., Carazo, J. M., and San Martin, C. (2005). Polymorphism and double hexamer structure in the archaeal minichromosome maintenance (MCM) helicase from *Methanobacterium thermoautotrophicum*. J Biol Chem 280, 40909-40915.

Gozuacik D, Chami M, Lagorce D, Faivre J, Murakami Y, Poch O, Biermann E, Knippers R, Brechot C and Paterlini-Brechot P. (2003). Identification and functional characterization of a new member of the human Mcm protein family: hMcm8. Nucleic Acids Res 31:570–579

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.

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.

Haugland, G. T., Shin, J. H., Birkeland, N. K., and Kelman, Z. (2006). Stimulation of MCM helicase activity by a Cdc6 protein in the archaeon *Thermoplasma acidophilum*. Nucleic Acids Res 34, 6337-6344.

Hoang, M. L., Leon, R. P., Pessoa-Brandao, L., Hunt, S., Raghuraman, M. K., Fangman, W. L., Brewer, B. J., and Sclafani, R. A. (2007). Structural changes in

Mcm5 protein bypass Cdc7-Dbf4 function and reduce replication origin efficiency in *S cerevisiae*. Mol Cell Biol. 27:7594-602

Hwang, D. S., and Kornberg, A. (1992). Opening of the replication origin of Escherichia coli by DnaA protein with protein HU or IHF. J Biol Chem 267, 23083-23086.

Ilves I., Petojevic T., Pesavento J.J., Botchan M.R. (2010). Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. Mol. Cell 37, 247-258.

Im J.S., Ki S.H., Farina A., Jung D.S., Hurwitz J., Lee J.K. (2009). Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins. Proc. Natl. Acad. Sci. U.S.A. 106, 15628-15632.

Ishimi, Y. (1997). A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex. J Biol Chem 272, 24508-24513.

Iyer L.M., Leipe D.D., Koonin E.V., Aravind L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. J. Struct. Biol. 146, 11-31.

Jenkinson E.R., Costa A., Leech A.P., Patwardhan A., Onesti S., Chong J.P. (2009). Mutations in subdomain B of the minichromosome maintenance (MCM) helicase affect DNA binding and modulate conformational transitions. J. Biol. Chem. 284, 5654-5661.

Jenkinson, E. R., and Chong, J. P. (2006). Minichromosome maintenance helicase activity is controlled by N- and C-terminal motifs and requires the ATPase domain helix-2 insert. Proc Natl Acad Sci U S A 103, 7613-7618.

Johnson EM, Kinoshita Y and Daniel DC. (2003). A new member of the MCM protein family encoded by the human MCM8 gene, located contrapodal to GCD10 at chromosome band 20p12.3-13. Nucleic Acids Res 31:2915–2925

Kaplan DL and O'Donnell M. (2004). Twin DNA pumps of a hexameric helicase provide power to simultaneously melt two duplexes. Mol Cell 15:453–465

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.

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.

Kelman, L. M., and Kelman, Z. (2003). Archaea: an archetype for replication initiation studies? Mol Microbiol 48, 605-615.

Kelman, Z., Lee, J. K., and Hurwitz, J. (1999). The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum DeltaH* contains DNA helicase activity.

Labib K., Diffley J.F. (2001). Is the MCM2-7 complex the eukaryotic DNA replication fork helicase? Curr. Opin. Genet. Dev. 11, 64-70.

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

Lee, C., Hong, B., Choi, J. M., Kim, Y., Watanabe, S., Ishimi, Y., Enomoto, T., Tada, S., and Cho, Y. (2004). Structural basis for inhibition of the replication licensing factor Cdt1 by geminin. Nature 430, 913-917.

Leirmo S, Harrison C, Cayley DS, Burgess RR, Record MT Jr. (1987) Replacement of potassium chloride by potassium glutamate dramatically enhances protein-DNA interactions in vitro. Biochemistry 21;26:2095-101.

Li, D., Zhao, R., Lilyestrom, W., Gai, D., Zhang, R., DeCaprio, J. A., Fanning, E., Jochimiak, A., Szakonyi, G., and Chen, X. S. (2003). Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. Nature 423, 512-518.

Liu W., Pucci B., Rossi M., Pisani F.M., Ladenstein R. (2008). Structural analysis of the *Sulfolobus solfataricus* MCM protein N-terminal domain. Nucleic Acids Res. 36, 3235-3243.

Liu Y, Richards TA and Aves SJ. (2009). Ancient diversification of eukaryotic MCM DNA replication proteins. BMC Evol Biol 9:60.

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.

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

Lutzmann M, Maiorano D and Mechali M. (2005). Identification of full genes and proteins of MCM9, a novel, vertebrate-specific member of the MCM2-8 protein family. Gene 5:362:51-6

MacNeill S.A. (2010). Structure and function of the GINS complex, a key component of the eukaryotic replisome. Biochem. J. 425, 489-500.

Maiorano D, Cuvier O, Danis E and Mechali M. (2005). MCM8 is an MCM2- 7-related protein that functions as a DNA helicase during replication elongation and not initiation. Cell 120:315–328.

Maiorano D, Lutzmann M and Mechali M. (2006). MCM proteins and DNA replication. Curr Opin Cell Biol 18:130–136.

Makarova K.S., Wolf Y.I., Mekhedov S.L., Mirkin B.G., Koonin E.V. (2005). Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. Nucleic Acids Res. 33, 4626-4638.

Marinsek, N., Barry, E. R., Makarova, K. S., Dionne, I., Koonin, E. V., and Bell, S. D. (2006). GINS, a central nexus in the archaeal DNA replication fork. EMBO Rep 7, 539-545.

McGarry, T. J., and Kirschner, M. W. (1998). Geminin, an inhibitor of DNA replication, is degraded during mitosis. Cell 93, 1043-1053.

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.

Mizushima, T., Katayama, T., and Sekimizu, K. (1996). Effect on DNA topology by DnaA protein, the initiation factor of chromosomal DNA replication in *Escherichia coli*. Biochemistry (Mosc) 35, 11512-11516.

Mogni ME, Costa A, Ioannou C, Bell SD. (2009) The glutamate switch is present in all seven clades of AAA+ protein. Biochemistry. 22;48:8774-5.

Moreau M.J., McGeoch A.T., Lowe A.R., Itzhaki L.S., Bell S.D. (2007). ATPase site architecture and helicase mechanism of an archaeal MCM. Mol. Cell 28, 304-314.

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

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.

Norais, C., Hawkins, M., Hartman, A. L., Eisen, J. A., Myllykallio, H., and Allers, T. (2007). Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. PLoS Genet 3, e77.

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.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605-1612.

Pospiech H., Grosse F., Pisani F.M. (2010). The initiation step of eukaryotic DNA replication. Subcell. Biochem. 50, 79-104.

Pucci, B., De Felice, M., Rocco, M., Esposito, F., De Falco, M., Esposito, L., Rossi, M., and Pisani, F. M. (2007). Modular organization of the *Sulfolobus solfataricus* mini-chromosome maintenance protein. J Biol Chem 282, 12574-12582.

Rampakakis E, Arvanitis DN, Di Paola D, Zannis-Hadjopoulos M. (2009) Metazoan origins of DNA replication: regulation through dynamic chromatin structure. J Cell Biochem. 1:106:512-20.

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

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

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

Ricke, R. M., and Bielinsky, A. K. (2004). Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. Mol Cell 16, 173-185.

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.

Sakakibara N, Kasiviswanathan R, Melamud E, Han M, Schwarz FP and Kelman Z. (2008). Coupling of DNA binding and helicase activity is mediated by a conserved loop in the MCM protein. Nucleic Acids Res 36:1309–1320.

Sakakibara N, Kelman LM and Kelman Z. (2009). Unwinding the structure and function of the archaeal MCM helicase. Mol Microbiol 72:286–296.

Sakwe, A. M., Nguyen, T., Athanasopoulos, V., Shire, K., and Frappier, L. (2007). Identification and characterization of a novel component of the human minichromosome maintenance complex. Mol Cell Biol 27, 3044-3055.

Sanchez-Pulido L, Diffley JF, Ponting CP. (2010) Homology explains the functional similarities of Treslin/Ticrr and Sld3. Curr Biol. 2010 22;20:R509-10

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

Schaper, S., Nardmann, J., Luder, G., Lurz, R., Speck, C., and Messer, W. (2000). Identification of the chromosomal replication origin from *Thermus thermophilus* and its interaction with the replication initiator DnaA. J Mol Biol 299, 655-665.

Sclafani, R. A., Tecklenburg, M., and Pierce, A. (2002). The mcm5-bob1 bypass of Cdc7p/Dbf4p in DNA replication depends on both Cdk1-independent and Cdk1-dependent steps in Saccharomyces cerevisiae. Genetics 161, 47-57.

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.

Shin JH, Heo GY and Kelman Z. (2009). The *Methanothermobacter thermautotrophicus* MCM helicase is active as a hexameric ring. J Biol Chem 284:540–546.

Shin JH, 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.

Singleton, M. R., Dillingham, M. S., and Wigley, D. B. (2007). Structure and Mechanism of Helicases and Nucleic Acid Translocases. Annu Rev Biochem 76, 23-50.

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.

Singleton, M. R., Sawaya, M. R., Ellenberger, T., and Wigley, D. B. (2000). Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. Cell 101, 589-600.

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.

Stillman B. (2005). Origin recognition and the chromosome cycle. FEBS Lett 579:877–884.

Sun J, Kong D. (2010) DNA replication origins, ORC/DNA interaction, and assembly of pre-replication complex in eukaryotes. Acta Biochim Biophys Sin (Shanghai). 2010 42:433-9.

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.

Tanaka S, Araki H. (2010) Regulation of the initiation step of DNA replication by cyclin-dependent kinases. Chromosoma. 2010 119:565-74.

Tye BK. (1999) MCM proteins in DNA replication. Annu Rev Biochem. 68:649-86.

van den Ent F, Löwe J. (2006) RF cloning: a restriction-free method for inserting target genes into plasmids. J Biochem Biophys Methods. 30;67:67-74

Wei Z., Liu C., Wu X., Xu N., Zhou B., Liang C., Zhu G. (2010). Characterization and structure determination of the Cdt1 binding domain of human minichromosome maintenance (Mcm) 6. J. Biol. Chem.

Yoshida K. (2005). Identification of a novel cell-cycle-induced MCM family protein MCM9. Biochem. Biophys Res Commun 331:669–674.

Yoshimochi T., Fujikane R., Kawanami M., Matsunaga F., Ishino Y. (2008). The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. J. Biol. Chem. 283, 1601-1609.

You, Z., and Masai, H. (2005). DNA binding and helicase actions of mouse MCM4/6/7 helicase. Nucleic Acids Res 33, 3033-3047.

You, Z., Ishimi, Y., Mizuno, T., Sugasawa, K., Hanaoka, F., and Masai, H. (2003). Thymine-rich single-stranded DNA activates Mcm4/6/7 helicase on Y-fork and bubble-like substrates. EMBO J 22, 6148-6160.

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.

Zhang X and Wigley DB. (2008). The 'glutamate switch' provides a link between ATPase activity and ligand binding in AAA+ proteins. Nat Struct Mol Biol 15:1223–1227.

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