Molecular Characterization of Mth203 protein

Chitvan Bochiwal

PhD Biology

August 2011

Author's declaration

I, Chitvan Bochiwal confirm that the work presented in this thesis is my own. Where information has been derived from the other sources, I confirm that this has been indicated in the thesis.

……………………………………………….

(Chitvan Bochiwal)

ABSTRACT

In contrast to the wealth of molecular and biochemical information available concerning eukaryotic and bacterial replication, less is known about the molecular basis of replication initiation in archaea. In general, the archaeal proteins are simplified versions of their eukaryotic counterparts. Therefore, they are potentially simple model systems to understand the conserved events in DNA replication and other processes.

Methanothermobacter thermautotrophicus, a thermophilic archaeon has a circular genome and a single origin of replication. It's replication initiation proteins MthCdc6-1 and MthCdc6-2 are homologues of the eukaryotic Cdc6 and ORC proteins. It also contains a single minichromosome maintenance protein (MthMCM) homologue which forms a hexameric complex and acts as a DNA helicase. What loads the archaeal MCM helicase and how DNA replication initiation is regulated is still unknown.

Mth203, a putative RNA helicase in *M. thermautotrophicus* has been shown to interact with MthCdc6-1 in a yeast two-hybrid screen and His-tagged full-length protein pull-downs. Mth203 is a SF2 family helicase. The proteins of this superfamily perform a wide array of functions in DNA/RNA processing and often are a part of multi-protein complexes. The Mth203 homologue (Mmp0457) in *Methanococcus maripaludis* has been co-purified with one of the MCM proteins (McmA) in this organism. Thus, it is likely that this protein interacts with the replicative machinery.

The aim of this study was to further characterize Mth203 structure and function. The results demonstrate that Mth203 is a dimeric protein and possesses NTP-dependent RNA helicase activity and RNA independent ATPase activity. Mth203 also inhibits MthMCM DNA helicase activity. Fluorescence anisotropy assays have shown that Mth203 binds non-specifically to short DNA sequences and specifically to long origin sequences. Mth203 was found to be expressed in periodic manner throughout the cell cycle and MALDI-TOF analysis has revealed that Mth203 interacts with MthCdc6-1 and ribosomal proteins in protein pull-down assays from whole cell extracts of *M. thermautotrophicus*. Further gene knock-out studies with

Mth203 homologue (Mmp0457) in *M. maripaludis* have shown that the protein is probably essential for cell survival and mmp0457 overexpression resulted in larger cell size with less DNA content. It is not yet clear whether Mth203 is involved in DNA replication and its specific role in RNA metabolism remains unknown. However, these findings provide a valuable insight into the potential role of Mth203.

Table of contents

List of Tables

List of Figures

List of Abbreviations

aa   amino acid
EDTA   Ethylenediaminetetraacetic acid
EtOH   Ethanol
g   G-force
h   hour
rpm   Revolutions per minute
RT   Room temperature
Tris   2-Amino-2-(hydroxymethyl)-1,3-propanediol
v/v   Volume per volume
w/v   Weight per volume

Acknowledgements

I want to express my gratitude towards my supervisor Dr. James Chong for his guidance, patience and encouraging support, without his persistent help this thesis would not have been possible. Also, I would like to express my gratefulness to my training advisory panel members Prof. Dale Sanders and Dr. Daniela Barilla, who continually encouraged me and gave suggestions, which were instrumental in successful completion of research work.

I would also like to thank Dr. Andrew Leech for the help in anisotropy assays, CD spectra analysis and SEC-MALLS. I would like to thank Dr. Fred Anston for allowing me to use the crystallography facilities in YSBL. I am very thankful to Dr. Callum Smits and Maria Chechik for productive discussions and their help in Mth203 crystallization experiments.

I would like to thank Dr. Richard Parker, Dr. Alison Walters, Mrs. Rubab Satti, Mrs. Huszalina Hussin, Miss Carrie O'Malley and Mrs. Juliette Borgia for making my time spent in the lab memorable.

Further, I would like to thank Dr. Harsh Amin for his constant support and motivation. I would also like to thank Dr. Kamran Haider and Miss Masridah Mahmud for being such a wonderful friends and housemates. Last but not the least I owe a great debt of appreciation to my mum, Dr. Laxmi Bochiwal and brother, Saurabh Bochiwal for their support throughout all these years.

1   Introduction

1.1.   The three domains of life

Until the first half of 20th century, it was believed that all living organisms could be classified as either bacteria or eukaryotes and for a long time there was no evolutionary framework for the phylogenetic classification of prokaryotes (Doolittle and Brown, 1994). In brief, prokaryotes included all the unicellular life forms with no properly defined nucleus and organelles whereas eukaryotes were the unicellular and multicellular organisms with membrane bound nucleus and organelles. With the advancement of biochemical, genetic and phylogenetic techniques, the prokaryotic kingdom could be classified and divided into closely related species based on 16S rRNA sequences (Zukerland and Pauling, 1965).

In 1977, Woese's studies on the differences in 16S rRNA fingerprinting revealed that several microbial groups including extreme halophiles and methanogens belong to a separate lineage, which diverges from the more familiar bacterial groups at the deepest phylogenetic levels (Woese and Fox, 1977, Woese, 1978). Woese proposed a third domain termed 'Archaeabacteria' and suggested a 'Woesian tree of life' (Figure 1.1.) consisting of three domains: Bacteria, Archaeabacteria and Eukarya (Woese and Fox, 1977). Later the term 'Archaebacteria' was designated a misnomer so it was shortened to 'archaea' (Woese *et al.,* 1990). In addition to the 16S rRNA signatures, archaeal genomes containing over 350 gene clusters exclusive to archaea has been presented (Graham *et al.,* 2000). These archaea-specific genes comprise 15% of archaeal genome and further support the theory that archaea are an independent deeply rooted domain. Nowadays, with the availability of whole genome sequences signature insertions/deletions (indels) can be identified which would assist in future classification (Gupta and Shami, 2011, Jarrell *et al.,* 2011).

1.1.1   Archaea

The term 'Archaea' is derived from the Greek word 'arch', which means old or primitive. Archaea are classified in five kingdoms (Figure 1.1): Euryarchaeota, Crenarchaeota, Korarchaeota, Nanoarchaeota and Thaumarchaeota (Woese *et al.,*
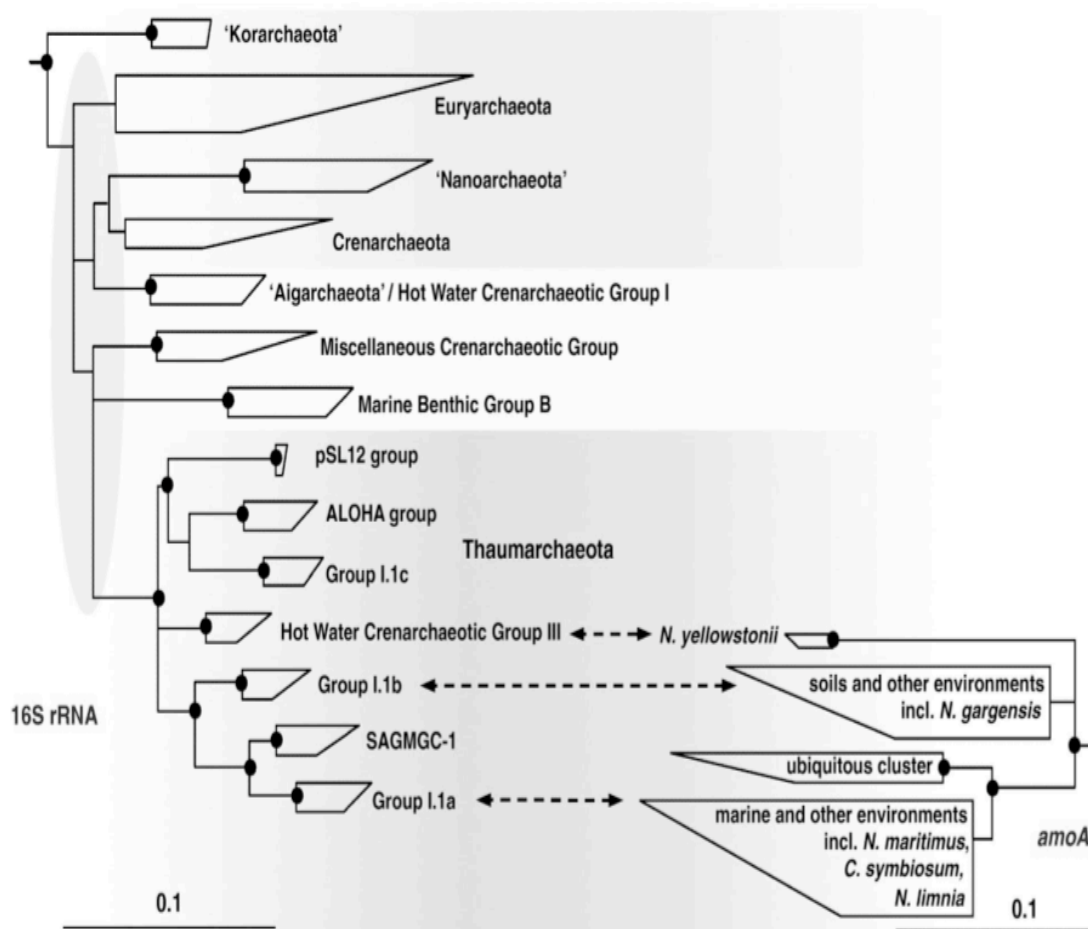
Figure 1.1. The phylogenetic tree showing the domain Archaea (from Pester et al., 2011). The phylogenetic classification on the basis of 16rRNA and ammonia oxidising (amoA) genes divides the domain archaea in five kingdoms: *Euryarchaeota, Crenarchaeota, Korarchaeota, Nanoarchaeota* and *Thaumarchaeota*. The dots in the figure indicate the bootstrap values above 80%, and the scale bar represents 10% estimated sequence divergence.

1990, Garrity and Holt, 2001, Huber et al, 2002, Brochier-Armanet et. al, 2008). Euryarchaeota consist of halophiles, hyperthermophiles, methanogens and thermophilic methanogens. Crenarchaeota include hyperthermophiles, psychrophiles and thermoacidophiles. The phylum *Thaumarchaeota* includes organisms previously classified as mesophillic crenarchaea (Brochier-armanet et. al, 2008, Gupta and Shami, 2011). The diversity in the *korarchaeal* and *nanoarchaeal* kingdoms is currently unclear, as very few members have been identified (Woese et al, 1990, Huber et al, 2002).

1.1.2   Archaea share features with both eukaryotes and bacteria

Archaea possess a mix of eukaryotic and bacterial features along with some unique aspects, which are characteristic of this domain of life (Brown, 2001, Robinson and Bell, 2005, Yutin *et al.,* 2008, Cox *et al.,* 2008, Brown and Doolittle, 1997).

Archaea and bacteria are together classified as 'prokaryotes' because they possess similar physiological characteristics i.e. size, absence of nuclear membrane and organelles, circular chromosomes, and common metabolic pathways (Brown, 2001). Many archaeal genes are organised in bacteria-like operons that are transcribed as long non-capped mRNA with short poly(A) tails (Keeling *et al.,* 1994, Ramirez *et al.,* 1993). L11, L10, S10 ribosomal protein, streptomycin, spectinomycin operon are few of the examples of gene organization similar to eukaryotes (Keeling *et al.,* 1994). Additionally, a small-RNA based defence mechanism CRISPR (clustered regularly interspaced short palindromic repeats) against phages and plasmids has also been found in both bacteria and archaea (Karginov and Hannon, 2010).

Although archaea and bacteria appear to have similar genome organization, the archaeal genes show great similarity to eukaryotic homologues. Archaea and eukaryotes share homology in the proteins involved in DNA metabolism and processing (DNA replication and repair), transcription, and translation (Edgell and Doolittle 1997). Studies on DNA binding proteins, initially on HMf in *Methanothermus fervidus* suggest the DNA is condensed into nucleosome like structures by DNA scaffolding proteins like histones in eukaryotes (Sandman *et al.,* 1990, Starich *et al.,* 1996, Reeve *et al.,* 1997). Eukaryotic histones are known to form heterodimers (H2A-H2B and H3-H4), archaeal histone are known to assemble as

both homo- and hetero-dimers. These histone-encoding genes have been identified in several members of Euryarchaea (Grayling *et al.,* 1994, Reeve *et al.,* 1997). Furthermore, bacterial DNA-binding proteins (HU) are also found in some archaea (Bianchi, 1994). In transcription machinery, archaea possess homologues of eukaryotic RNA polymerase II subunits (Puhler *et al.,* 1989, Zillig *et al.,* 1978), TATA-box like motifs, and some transcription factors (TATA binding proteins (TBP); and TFB, a homologue of eukaryotic TFIIB (transcription factor II B)) (Rowlands *et al.,* 1994, Wettach *et al.,* 1995, Langer *et al.,* 1995, Werner and Weinzierl, 2002, Werner, 2007). Archaea and eukaryotes also share a pathway of isoprenoid biosynthesis that involves the synthesis of mevalonate from 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) by the enzyme HMG-CoA (Bochar *et al.,* 1999, Lam *et al.,* 1992).

Archaea also possess features that are unique to this domain of life. The archaeal cell envelopes are made up of archaeal-specific polymer pseudomurien (Kandler and Konig , 1978), and the cytoplasmic membranes are made up of ether-linked lipids (diphytanylglycerol ethers or diphytanyldiglycerol tetraethers) (Jarrell *et al.,* 2011, Woese, 2004, Zillig, 1991). Archaea possess archaea-specific appendages like hami (Moissl *et al.,* 2005) and cannulae (Nickell *et al.,* 2003) in addition to bacteria like appendages i.e. flagella and pili. There is a lot of metabolic diversity in archaea, which has contributed towards study of new metabolic pathways in biochemistry. Methanogens are the only known organisms to use carbon fixation process methanogenesis to generate energy (DiMarco *et al.,* 1990, Thauer *et al.,* 2008, Weiss and Thauer, 1993) and several archaea are known to carry out glucose metabolism through variations of Embden-meyerhof (EM) and Entner-Doudoroff (EM) pathways (Siebers and Schonheit, 2005, Verhees *et al.,* 2003). The histone homologues in archaea lack N- and C- terminal extensions, which are sites of post-translation modifications in eukaryotes (Jarrell et. al., 2010). Furthermore, archaea possess variety of secondary and tertiary modifications of tRNA molecules (Woese *et al.,* 1980).

1.2   Cell cycle

In all life forms both prokaryotic and eukaryotic, the cells go through a cell cycle involving chromosome replication, chromosome segregation and cell division. In
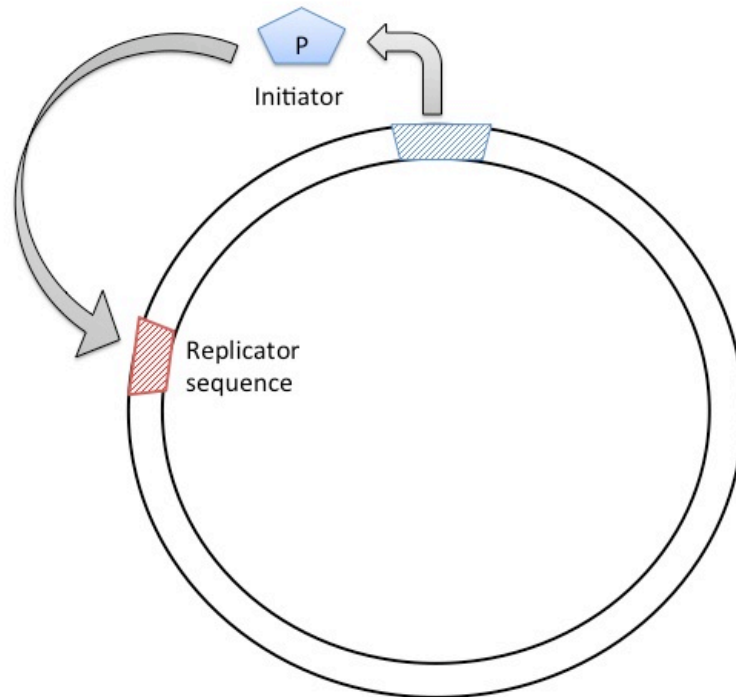
Figure 1.2. A diagrammatic representation of the replicon model to explain DNA replication and regulation (Jacob et al., 1963). A trans-acting initiator protein recognises a cis-acting DNA sequence (replicator sequence) and initiates DNA replication.

i) Slow growing cells

d_0 B i_1 C t_1 D d_1 B i_2 C t_2 D d_2

Age of cell at initiation

IIT (inter initiation time)

New born cell

o t

ii) Fast growing cells

i_1 C t_1 D d_1

i_2 C t_2 D d_2

IIT

i_3 C t_3 D d_3

IIT

i_4 C t_4 D d_4

IIT

Age of cell at initiation
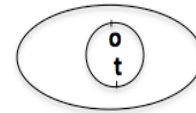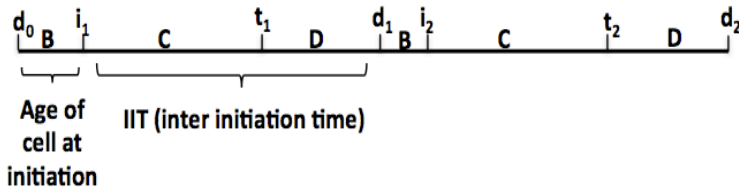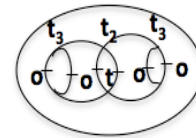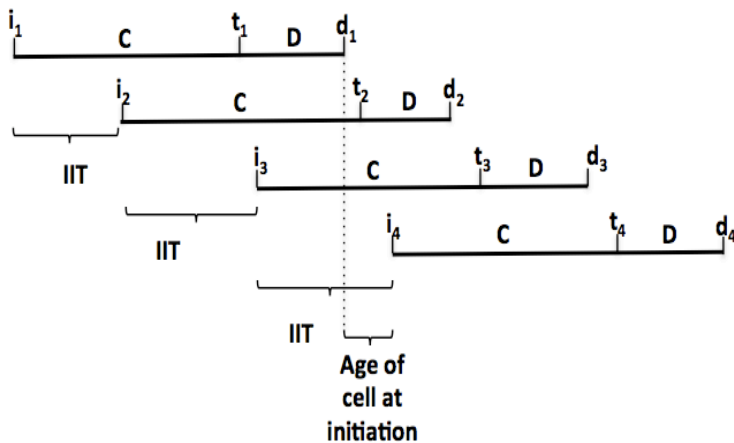
New born cell after d_1

t_3 t_2 t_3 o o t o

Figure 1.3. Schematic representation of cell cycle in slow-growing and fast growing E.coli cells (from Zyskind and Smith, 1992). B, C and D represent time interval for cell to go through pre-initiation, DNA replication (initiation (i), termination (t)) and cell division (d) respectively. The Inter initiation time (IIT) is the time difference between the start of two consecutive cell cycles. For slow growing cells, IIT is same as the length of a cell cycle and the newborn cells contain single origin and termination regions. For fast growing cells the IIT is much smaller as the second cycle of DNA replication initiates before the end of first cell cycle. The newborn cells thus produced contain multiple origins of replication forks and termini.

every organism this cell cycle is precisely controlled. Jacob *et al.* (1963) proposed a replicon model to explain the regulation of DNA replication in bacteria, which is applicable to all three domains of life: an initiator factor (trans-acting protein) would act at a replicator sequence (cis-acting sequence) in the chromosome to control and facilitate DNA replication (Figure 1.2) (Jacob *et al.,* 1963).

Cooper and Helmstetter (1968) described the process for bacteria in their pioneering work on the *E.coli* cell cycle (Figure 1.3). According to the cell cycle theory (Cooper and Helmstetter (1968), assuming that the time taken for DNA replication (C) is the time period between initiation (i) and termination (t) of the cell cycle. For slow growing cells, the growth rate (D) is longer than DNA replication (C), as there is a time difference between cell division (d) and DNA replication termination (t). Thus, DNA replication (C) plus the growth rate (D) is less then doubling time and there is usually a gap between initiation and termination (B). The time period for the whole cell cycle can be defined as the inter-initiation time (IIT). However in fast growing cells, C+D is greater than the doubling time, replication reinitiates before the completion of the previous round of replication and the IIT is much less than the length of cell cycle. Such new-born cells contain multiple origins and one terminus of replication (Figure 1.3). This research demonstrated that although the slow growing cells and fast growing cells are going through the same cell cycle, there is a marked difference in the number of origins of replications per new-born cell due to a decreased cell division time (Helmstetter, 1987, Zyskind and Smith, 1992).

1.3 DNA replication

For the propagation of genetic material, error free DNA replication is an essential step in the cell cycle. The DNA replication is a complex process requiring numerous proteins and enzymes to duplicate the genetic information present in the form of chromosomes (Grabowski and Kelman, 2003). Initiation is a key phase in DNA replication, as most of the regulation of cell cycle occurs in this step. It is imperative for DNA replication to be restricted to once every cell cycle to prevent any adverse effects of re-replication before the end of cell division. Hence, it is important to understand the mechanisms and regulation of the processes involved in initiation of DNA replication.

1.3.1 Initiation of DNA replication in bacteria

Different bacteria have replication origins of widely differing sizes but all contain several DnaA boxes and an AT-rich region (Messer, 2002). For the purpose of simplification *E.coli* DNA replication initiation is discussed here (Figure 1.4). DNA replication is initiated typically at a single origin of replication (*oriC*). *oriC* contains multiple repetitive consensus segments called DnaA boxes and an AT-rich region to facilitate DNA melting (Fuller *et al.,* 1984). The number and location of DnaA boxes varies between different bacterial species. Electrophoretic mobility shift assays (EMSA) and surface plasmon resonance assay have demonstrated that in *E.coli*, DnaA protein binds specifically to the consensus sequence 5'-TT(A/T)NCATNCACA (called a "strong box") and can also bind to the relaxed consensus sequence 5'-(T/C)(T/C)(A/T/C)T(A/C)C(A/G)(A/C/T)(A/C) (Schaper and Messer, 1995, Messer, 2002). DnaA belongs to the AAA$^+$ family of ATPases. The AAA+ (ATPases associated with a variety of cellular activities) family consists of chaperone-like proteins involved in the assembly, operation, and disassembly of diverse protein machines (Kunau *et al.,* 1993, Neuwald *et al.,* 1999). The proteins belonging to this class possess conserved N- terminal RecA-type fold containing NTPase binding site formed by a Walker-A motif (GxxxxGKT, x = any residue) and a Walker-B motif (hhhhDEXX, h = hydrophobic residue) connected to a α-helical domain and a conserved motif C (or sensor-1) (Walker *et al.,* 1982, Gorbalenya and Koonin, 1989, Koonin 1993, Saraste *et al.,* 1990). Walker A-motif is required for ATP binding and Walker B-motif helps in ATP hydrolysis (Walker *et al.,* 1982).

In order to initiate DNA replication the presence of a "strong box" (DnaA box R1 and R4 in *E.coli*) is essential (Messer, 2002). The other DnaA binding sites have relatively low affinity for DnaA and are collectively called ATP-DnaA preferential low-affinity sites (ADLAS) (Kawakami and Katayama, 2010). A DnaA monomer binds to a DnaA box, which acts as an anchor for co-operative binding of ATP-DnaA to other DnaA boxes at the *oriC*. The binding causes a sharp bend into the binding site leading to torsional stress, which unwinds the DNA in the AT-rich region next to the *oriC*, forming an "Open complex" (Figure 1.4) (Messer, 2002).

Another DNA binding protein, DiaA, is shown to form homotetramers, and is involved in stimulating the assembly of multiple ATP–DnaA molecules on *oriC*
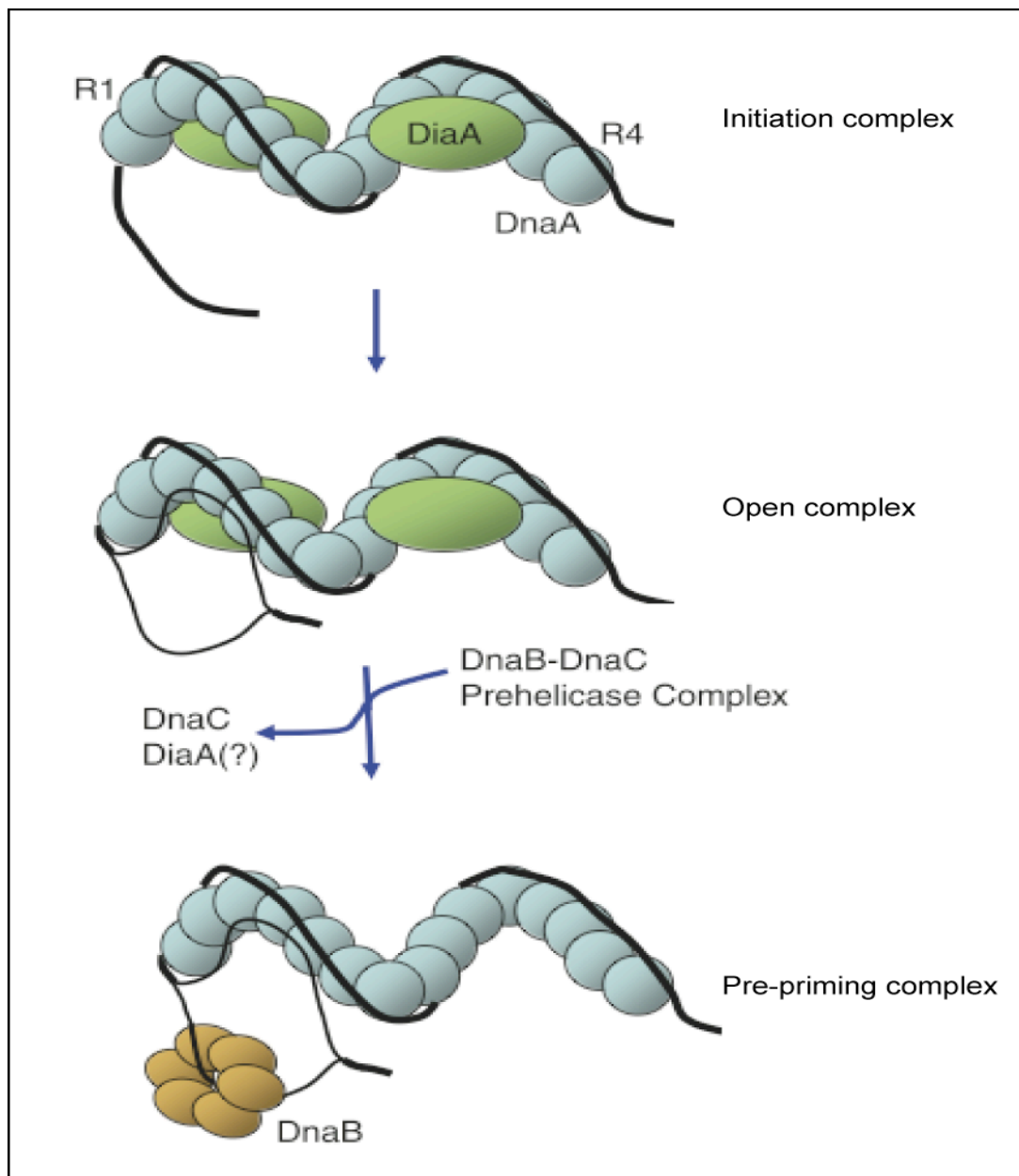
Figure 1.4. Schematic representation of initiation of DNA replication in bacteria (E.coli) (from Kawakami and Katayama, 2010). A DnaA monomer binds to a DnaA box (R1, R4), which acts as an anchor for co-operative binding of ATP-DnaA to other DnaA boxes at the *oriC*. The binding causes a sharp bend into the binding site leading to torsional stress, which unwinds the DNA in the AT-rich region next to the *oriC*, forming an "Open complex" (Messer, 2002). Another DNA binding protein, DiaA, is shown to form homotetramers, and is involved in stimulating the assembly of multiple ATP–DnaA molecules on *oriC* (Keyamura *et al.,* 2007). DnaC (helicase loader) loads DnaB (DNA helicase) onto this open complex leading to formation of a pre-priming complex ready for initiation of DNA replication.

(Keyamura *et al.,* 2007). DiaA positively regulates a conformational change of the DnaA multimer–*oriC* complex to an initiation-competent state (Ishida *et al.* 2004, Keyamura *et al.,* 2007). The DiaA tetramer can bind several DnaA molecules and serves as a bridge, bringing two DnaA molecules together and facilitating inter-DnaA interactions.

Cooperative binding to the double-stranded AT-rich region is therefore presumably the limiting step in the initiation reaction, followed by unwinding (Bramhill and Kornberg, 1988, Speck and Messer, 2001). Although both ADP-DnaA and ATP-DnaA can bind to the DnaA box, ATP-DnaA can recognise an additional six base pair element close to the DnaA box, thus adding further control over initiation of DNA replication (Robinson and Bell, 2005).

Once the DNA is unwound, DnaC (the helicase loader, and a AAA$^+$ family ATPase protein) loads DnaB (a homohexameric DNA helicase) on the open complex, forming a pre-priming complex (Kornberg and Baker, 1992). DnaC is released by ATP hydrolysis, furthermore, DnaG and DNA polymerase are recruited onto this region leading to initiation of DNA replication (Robinson and Bell, 2005).

In a steady-state culture, intervals between initiation events in a cell are equal to the doubling time of cell number (Kitagawa *et al.,* 1998). In *E.coli*, the cell takes 60 minutes for a round of replication and subsequent cell division at 37°C. Thus, in rapidly growing cells, in order to maintain rapid growth the cell has to initiate new rounds of replication while previous rounds are still in progress (Figure 1.3). Under such conditions, all origins fire essentially synchronously and, therefore, cells always contain 2n and 2n+1 origins (where n is a positive integer; Skarstad *et al.* 1986).

For partitioning equal number of chromosomes in each daughter cell, bacteria need to regulate that each copy of *oriC* fires once per cell cycle and thus regulate the length of the inter-initiation time (IIT; Figure 1.3). This is carried out by rapidly changing the level of DnaA in the cells (Zyskind and Smith, 1992). In a replication initiation study on *E.coli*, when *dnaA* gene expression was induced the DNA content per mass unit increased, suggesting that increased synthesis of DnaA causes increased initiation of DNA replication (Skarstad *et al.,* 1986). In *E.coli,* re-initiation, is prevented by three mechanisms (a) sequestering of *oriC* and the *dnaA*

promoter region by SeqA, (b) binding of DnaA to *datA locus* that provides a sink for DnaA and (c) regulatory inactivation of DnaA at the end of initiation cycle (RIDA).

The replication origin and *dnaA* promoter region contains an unusually high number of GATC sequences known as Dam methytransferase recognition sites (Messer *et al.,* 1985, Smith *et al.,* 1985). These sites are methylated by Dam methytransferase by transferring methyl group at N^6 position of adenine in GATC sites (Lacks and Greenberg, 1977). A recently replicated DNA is hemi-methylated and requires Dam methyltransferase to methylate the freshly replicated strand. It has been observed that after replication initiation the methylation of origin sequences and *dnaA* promoter is much slower (one third of replication time) as compared to rest of the genome (1 min of replication). This prolonged hemimethylation state of the origin and *dnaA* promoter sequence is caused by binding/sequestering of these sequences to the cell membrane and hence preventing re-initiation of replication (Ogden *et al.,* 1988, Kitagawa *et al.,* 1998). The sequestering is assisted by SeqA protein, which has high affinity for hemi-methylated DNA (Lu *et al.,* 1994, von Freiesleben *et al.,* 1994). SeqA competes with DnaA for binding to these sites, and prevents re-initiation and suppresses *dnaA* expression (Kitagawa *et al.,* 1998, Messer, 2002).

A locus *datA* (DnaA titration) binds large number of DnaA molecules and participates in negative control of replication initiation by reducing the amount of free DnaA (Kitagawa *et al.,* 1998, Fuller *et al.,* 1984, Ogawa *et al.,* 2002). The *datA* locus spans about 950 bp between the glyVXY and amiB-mutL operons at 94.7 min on the genetic map and could be replicated during *oriC* sequestration. The 1 kb DNA segment contains five repeated sequences matching the DnaA box sequences (Ogawa *et al.,* 2002). Usually, *datA* competes with *oriC* for DnaA binding, but after replication initiation there is more DnaA binding on *datA* due to the absence of SeqA competition at these *oriC* sites. *datA* null mutants showed asynchronous initiations and initiation frequency was increased as compared to wild type cells (Kitagawa *et al.,* 1996, 1998, Ogawa *et al.,* 2002). Thus, *datA* is a sink for DnaA protein which regulates the availability of DnaA and hence replication initiation.

The ATP bound form of DnaA is the active form involved in DNA replication initiation (Sekimizu *et al.,* 1987). Once DNA polymerase initiates DNA synthesis, ATP
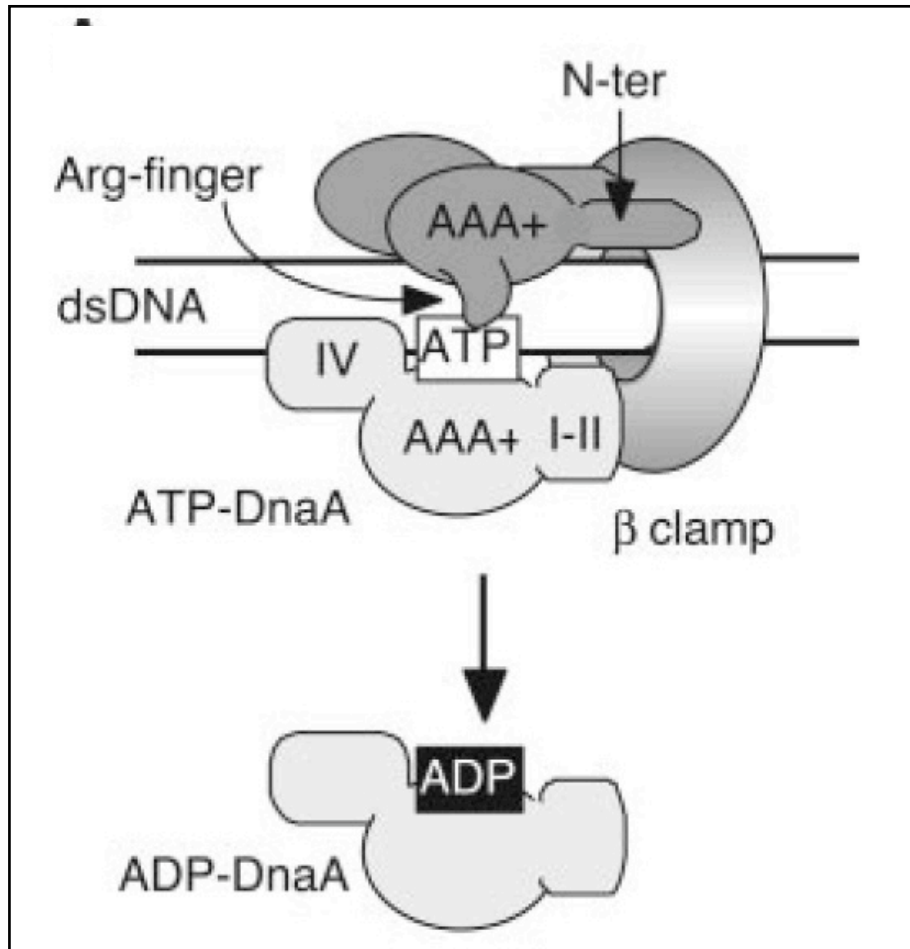
Figure 1.5. Schematic representation of RIDA reaction (as proposed by Su'etsugu et al., 2008). One or two Hda dimers form a stable complex with the hydrophobic pocket of the clamp that remains on DNA after completion of an Okazaki fragment. ATP-bound DnaA interacts with this complex in a manner depending on interaction with double-stranded DNA (dsDNA) flanking the clamp. The Hda AAA$^+$ domain (AAA+) interacts with the DnaA AAA$^+$ domain (domain III), leading to the formation of an ATP hydrolysis catalytic centre in which Hda Arg168 (the Box VII motif arginine finger) and DnaA R334 (the Box VIII motif Sensor-2 arginine) participate. ADP-Hda inactivates ATP-DnaA by converting it to ADP-DnaA.

bound to DnaA is efficiently hydrolyzed to yield the ADP-bound inactivated form. This negative regulation of DnaA occurs through interaction with the β-subunit sliding clamp configuration of the polymerase mediated by Hda (AAA$^+$ chaperone-like ATPase family protein). Hda is essential for this regulatory inactivation of DnaA *in vitro* and *in vivo* by a process termed 'Regulatory Inactivation of DnaA' (RIDA) (Katayama *et al.,* 1998, Kato and Katayama, 2001). The inactivation is caused by interaction of ADP-Hda with ATP-DnaA mediated by the DNA polymerase-loaded clamp and causes hydrolysis of ATP-DnaA to ADP-DnaA catalysed by Hda (Figure 1.5) (Su'etsugu *et al.,* 2008).

Recently, the presence of a system actively recycling ADP-DnaA to ATP-DnaA in *E.coli* cells was demonstrated (Fujimitsu *et al.,* 2009), wherein, ADP-DnaA molecules form multimers at specific sites on the DNA called 'DnaA-Reactivating Sequences' (DARSs) 1 and 2. This causes a reduced affinity of DnaA for ADP and promotes the binding of ATP (Fujimitsu *et al.,* 2009, Kawakami and Katayama, 2010).

1.3.2   Initiation of DNA replication in eukaryotes

Eukaryotic cell cycle has four phases: G1 (preparatory phase where cell prepares for DNA replication), S (synthesis phase, DNA replication), G2 (preparatory phase for cell division) and M phase (mitosis, nuclear division followed by cytokinesis) (Figure 1.6). In 1970, cell fusion experiments were carried out in order to understand eukaryotic DNA replication (Rao and Johnson, 1970) which showed that the chromosomes are prepared for initiation of DNA replication in G1 phase, however, replication is not initiated due to the absence of activators triggering the entry into S phase of cell cycle. Later, on the basis of experiments carried out on *Xenopus* egg extracts, Blow and Laskey (1988) proposed the 'Licensing model' suggesting that in G1 phase a licensing factor required for replication initiation is bound to the DNA, thus replication origins acquire replication competence but remain inactive (Licensing of origins), and in G1/S phase transition this factor is inactivated, thus origins can be activated leading to DNA replication but cannot be relicensed in order to inhibit re-initiation of replication and the licensing factor remains inactive until cells pass through mitosis (Blow and Laskey, 1988).
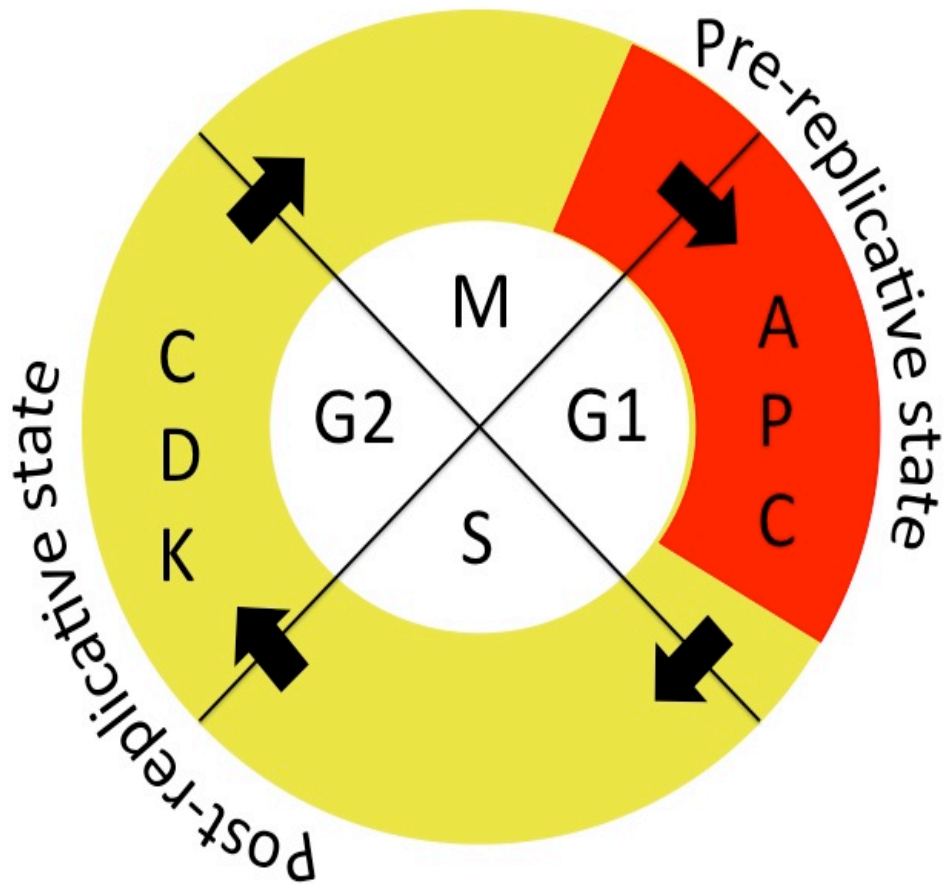
Figure 1.6. Replication licensing and cyclin-dependent kinase (CDK) activity through the cell cycle. The presence of licensed origins along with high concentration of the anaphase promoting complex/cyclosome (APC/C) is shown in red, and the presence of active cyclin-dependent kinases (CDKs) is shown in yellow.

Further, using *in-vivo* footprinting experiments in yeast, Diffley (1994) proposed that eukaryotic cell cycle involves transition between two distinct states: a pre-replicative state (G1 phase) and a post-replicative state (S, G2 and M phase) (Figure 1.6) (Diffley *et al.,* 1994, 2004). Cell cycle block and release experiments suggest that the conversion between these states occur during the M → G1 transition (post-Replicative→ pre-Replicative) and during the G1 → S transition (pre-Replicative → post-Replicative) (Dutta and Bell, 1997). The cyclin dependent kinases (CDKs) and the anaphase promoting complex/cyclosome (APC/C) are the key cell-cycle regulators, which tightly regulate the complex formation and the cell cycle phase transition (Broek *et al.,* 1991, Hayles *et al.,* 1994, Moreno and Nurse, 1994, Bell and Dutta, 2002, Walter and Newport, 2000, Tanaka *et al.,* 2006, Zegerman and Diffley, 2006). The replication licensing occurs by assembly of pre-replicative complex on the origin and only occurs when CDK activity is low and APC/C activity is high (Figure 1.6). Origin firing, however, can only occur when the APC/C is inactivated and CDKs become active. This two-step mechanism ensures that no origin can fire more than once in a cell cycle (Diffley, 2004, Broek *et al.,* 1991, Stillman, 2005).

The most well understood eukaryotic replication model is that of the yeast *Saccharomyces cerevisiae* (Figure 1.7). Unlike bacteria, eukaryotic DNA replication is initiated at multiple origins of replication at the same time. In this system origins of replication consist of conserved sequences called autonomously replicating sequences (ARS), which are 100-200 bp in length. The ARS consists of conserved ARS-consensus sequences (ACS) or A elements 5'-((T/A)TTTA(T/C)(A/G)TTT(T/A)), and divergent motifs called B-elements (for example, ARS1 on chromosome IV has elements B1 5'-(TTTTATGCTTG), B2 5'-(AATACTTAAAT), and B3 5'-(TTTGCTATTTT)) , that are collectively required for initiation of DNA replication (Brewer and Fangman, 1987, Kawakami and Katayama, 2010).

The replication pre-initiation complex forms when the origin recognition complex (ORC), a six subunit hetero-hexameric complex (containing ORC1-6 proteins), binds to the conserved ACS in an ATP-dependent manner and the ORC ATPase activity is inhibited by ORC-ARS binding (Figure 1.7) (Bell and Dutta, 2002). ORC is conserved in all eukaryotes and the ORC1p is even conserved in some species
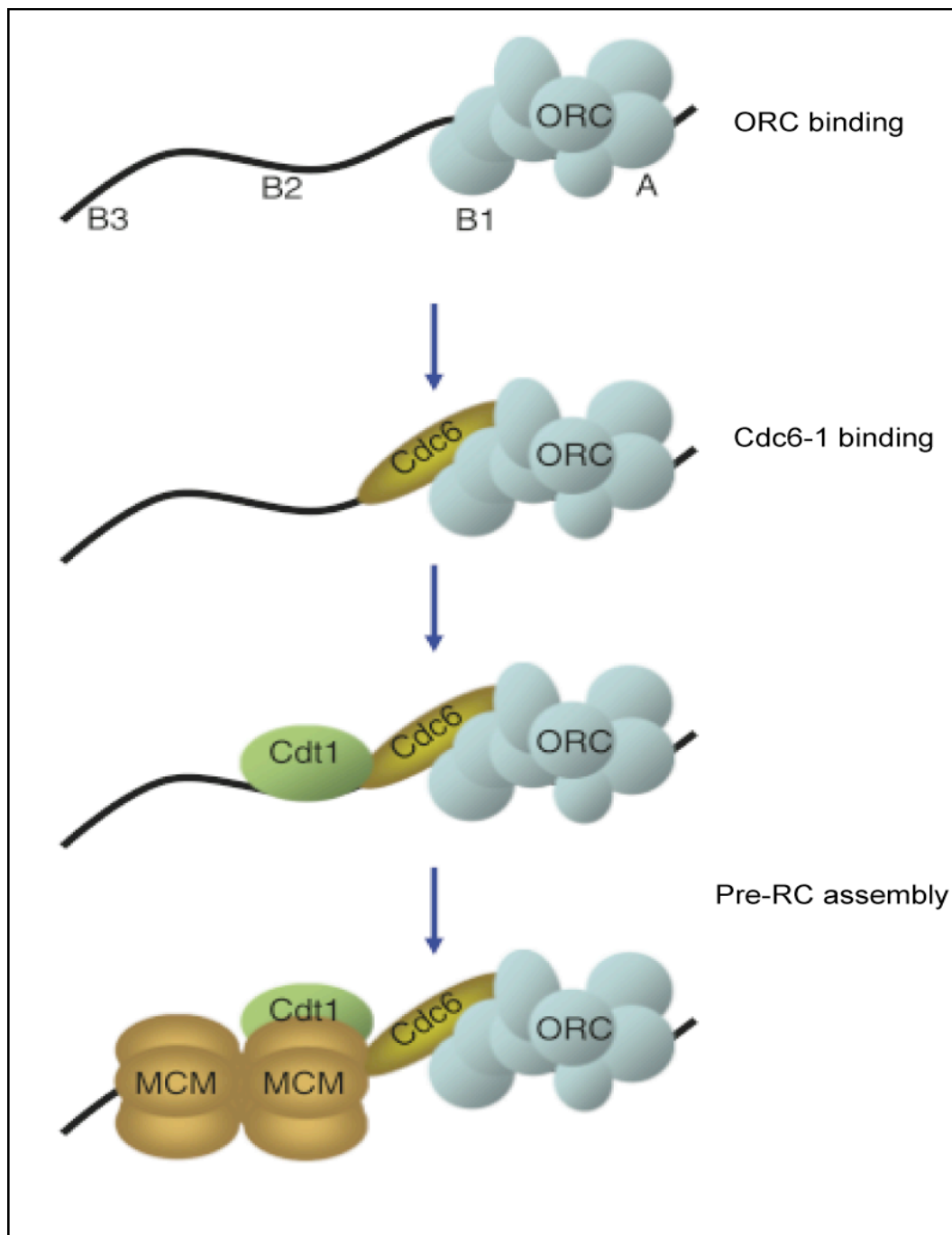
***Figure 1.7. Schematic representation of initiation of DNA replication in eukaryotes
(S. cerevisiae) (from Kawakami and Katayama, 2010).*** DNA replication is initiated at
multiple origins of replication (ACS sites) by binding of origin recognition complex
(ORC) in an ATP-dependent manner. Cdc6 protein binds to the ORC-ACS complex,
inducing a structural change in the ORC and further increasing DNA binding affinity
and specificity of ORC. Cdt1 is then loaded onto the DNA. ORC, Cdc6 and Cdt1
together are required to load MCM, the replicative DNA helicase forming the pre-
replication assembly ready for initiation of DNA replication.

of archaea (Bell and Dutta, 2002). ORC consists of a number of AAA[+] family of proteins (Orc1p-Orc6p) and the ATPase activity is inhibited by dsDNA and activated by ssDNA by inducing conformational change in the protein structure (Lee et al, 2000). In yeast, ORC remains bound to the chromatin at origins of replication through-out the cell cycle and plays an important role in origin determination as well as acting as a landing pad for the assembly of a series of cell-cycle-regulated protein complexes (Diffley *et al.,* 1994, Liang and Stillman, 1997, Stillman 2005, Mizushima *et al.,* 2000). Cdc6 (Cell division cycle 6 protein), also contains nucleotide binding Walker A and B motifs, characteristic of AAA[+] proteins (Zhou et al 1989, Kelly *et al.,* 1993, Coleman *et al.,* 1996). This region of Cdc6 shows homology with the nucleotide-binding domain in Orc1p, which is responsible for the ORC ATPase activity and ATP-dependent DNA binding (Bell *et al.,* 1995, Klemm *et al.,* 1997). In addition to these domains Cdc6 also contains an additional domain, domain III, with a winged helix (WH) type fold commonly found in dsDNA binding proteins (Singleton *et al.,* 2004). Cdc6 binds to the ORC-ACS complex, inducing a structural change in the ORC and further increasing DNA binding affinity and specificity of ORC (Figure 1.7) (Speck *et al.,* 2005). The assembly of the human Orc1 to 6 proteins, to form ORC *in vitro*, is dependent on ATP binding, however, it is unknown how ORC binds to specific sites on the chromosomal DNA in human cells (Siddiqui and Stillman, 2007).

Cdt1, a coiled coil domain protein, is then loaded onto the DNA in an ATP-dependent association with ORC (Figure 1.7) (Maiorano *et al.,* 2000, Devault *et al.,* 2002). ORC, Cdc6 and Cdt1 together are required to load MCM (Figure 1.7), the replicative DNA helicase, at the origin sequence (Bell and Dutta, 2002). Cdc6 causes ATP hydrolysis resulting in release of Cdt1 from the complex (Randell *et al.,* 2006). The complex thus formed is termed as the pre-replication complex (Pre-RC complex) and is a key event in DNA replication termed 'licensing of origins' (Blow and Laskey, 1988, Blow and Hodgson, 2002, Lei and Tye, 2001, Gillespie *et al.,* 2001, Nguyen *et al.,* 2001, Chong *et al.,* 1995, Tanaka *et al.,* 1997). The pre-RC complex thus formed is activated at the G1-S phase transition by phosphorylation of MCM complex carried out by Cdc7-Dbf4 kinase, (Sheu and Stillman, 2006), Sld2 and Sld3 by cyclin-dependent kinases (CDKs) (Zegerman and Diffley, 2006). In addition, MCMs must further interact with Cdc45 and GINS in the presence of cyclin dependent kinase,

CDK, Sld3, DDK (Dbf4 and Drf1 dependent kinase), Mcm10, Dbp11 and Cdc7 (Kamimura *et al.,* 2001, Takayama *et al.,* 2003, Kanemaki and Labib, 2006).

In other eukaryotes the origins of replication are rather elusive and no known conserved sequences have been identified. In budding yeast *S. pombe* the origin regions are AT rich and ORC recognises origins via an "AT-hook" DNA binding domain in the Orc4 subunit (Chuang *et al.,* 2002, Chuang and Kelly 1999, Lee *et al.,* 2001, Moon *et al.,* 1999, Houchens *et al.,* 2008). In *Drosophila* and *Xenopus* egg/ embryonic systems, any DNA fragment can act as an origin of replication and in mammalian cells, DNA replication appears to be initiated randomly from multiple chromosomal sites within large zones of initiation (Mechali and Kearsey, 1984, Dijkwe *et al.,* 1994, Mesner *et al.,* 2003, Arias and Walter, 2005, Vashee *et al.,* 2003, Chesnokov *et al.,* 1999, Houchens *et al.,* 2008). However, all the sites have been found to contain a multitude of AT-rich regions (Robinson and Bell, 2005).

To prevent re-replication, multiple mechanisms are employed by all eukaryotes (CDK inhibition is summarised in Table 1.1). In *S. cerevisiae*, Cdc6 is inhibited by three mechanisms (a) autophosphorylation leading to ubiquitylation by E3 ligase leading to proteolytic degradation by proteasome (Drury *et al.,* 1997), (b) CDKs inhibit Cdc6 transcription by blocking the nuclear membrane import of the transcription factor, Swi5 (Moll *et al.,* 1991), and (c) N- terminal phosphorylation of Cdc6, thus inducing stable binding with mitotic CDK, Clb2-Cdc28 and inhibits the licencing activity of Cdc6 (Mimura *et al.,* 2004, Arias and Walter, 2007). CDK phosphorylation of MCM and Cdt1 leads to their export to the cytoplasm in order to prevent re-replication (Labib *et al.,* 1999, Nguyen *et al.,* 2000, Liku *et al.,* 2005, Tanaka and Diffley, 2002). ORC is also inhibited by phosphorylation of Orc2 and Orc5 by CDK and Orc6 binding to Clb5 after initiation of DNA replication (Weinberg *et al.,* 1990, Nguyen *et al.,* 2001, Wilmes *et al.,* 2004).

Table 1.1: Pathways for cyclin-dependent kinase (CDK) inhibition of origin licensing. The different mechanisms by which CDKs can inhibit the activity of ORC, Cdc6, Cdt1 and Mcm2–7 in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe, Xenopus* and human (Arias and Walter, 2007).

Organisms	ORC	Cdc6	Cdt1	Mcm2-7
S. cerevisiae and S. pombe	Inhibition by CDK-dependent Orc2, 5 phosphorylation	CDK-dependent proteolysis CDK-regulated transcription	CDK-dependent proteolysis CDK regulated transcription	CDK regulated nuclear exclusion
Human and *Xenopus*	CDK-dependent complex disassembly or removal from chromatin?	CDK-dependent nuclear export of soluble protein (although chromatin-bound protein persists). CDK-dependent proteolysis?	Inhibition by Geminin (regulated by CDK-dependent proteolysis). CDK-dependent proteolysis? Cell-cycle dependent expression?	No clear evidence of CDK regulation

In other eukaryotes, Cdt1 is negatively regulated by Geminin, a cell cycle regulation protein and prevents MCM loading on the DNA (Kawakami and Katayama, 2010). In addition, like RIDA dependent DnaA regulation in bacteria, Cdt1 is regulated by ubiquitin-mediated proteolysis carried out by PCNA, a processivity factor for DNA polymerase (Arias and Walter 2007, Nishitani *et al.,* 2006).

After cell division the initiation of DNA replication is brought about in the cell during the transition from M-G1 phase. This is achieved in yeast through a temporary cascade of protein degradation of cyclin dependent kinases like Clb5, Dbf4 (Arias and Walter, 2007). Furthermore, pre-RC activators and components like Cdc6 are resynthesized as the cells enter G1 phase.

1.3.3 Initiation of DNA replication in archaea

In contrast to plenty of molecular and biochemical information available about the eukaryotic and bacterial origins of replication, less is known about the molecular basis of replication initiation in archaea. Bioinformatics, biochemical and structural studies have revealed that the archaeal initiation process is a combination of bacterial and eukaryotic processes (Stillman, 1994) (Table 1.2, Figure 1.8). Earlier, analysis of the cell cycle in archaea was limited to physiological experiments and most of these experiments were performed on genus *Sulfolobus solfataricus.* It was observed that, when these stationary phase cells were diluted into fresh medium, no initiation of DNA replication was observed until the preceding segregation and division events had been completed, suggesting the resting phase of *S. solfataricus* is post-replicative G2 phase (Hjort and Bernander 1999, Bernander, 2007). In addition, archaea like eukaryotes may possess important cell cycle regulatory features.

The origin of replication was not identified for a long time in archaea as no common features between bacteria and eukaryotes were found in archaeal genomes. In 1998, *in silico* skew analysis performed by several scientists identified a single origin in some archaea (e.g. *Methanothermobacter thermautotrophicus, Methanosarcina mazei* and *Pyrococcus furiosus*) and multiple origins in others (e.g. *Methanocaldoccocus jannaschii* and *Halobacterium*) (Zhang and Zhang, 2002, 2004, Lopez *et al.,* 1999, Kelman and Kelman, 2003, Matsunaga *et al.,* 2001). In some
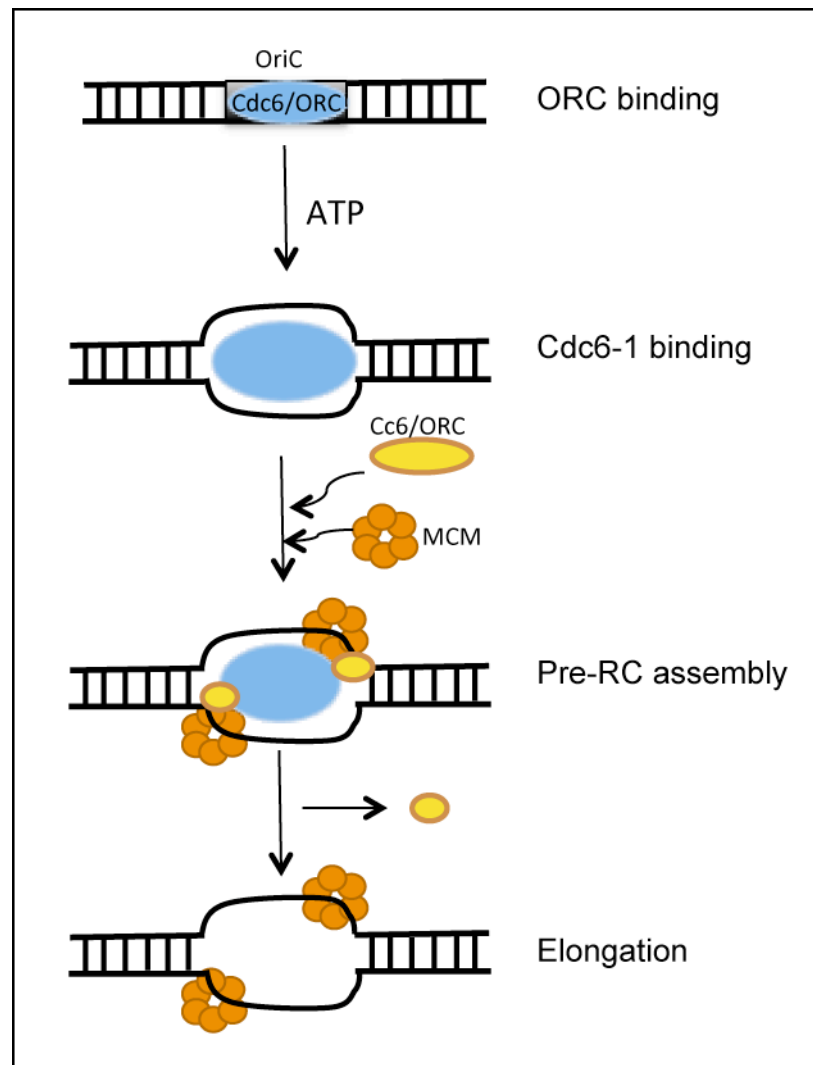
Figure 1.8. Hypothetical representation of initiation of DNA replication in archaea (M. thermautotrophicus). DNA replication in archaea is initiated at single or multiple origins of replication (ORB sites) by binding of origin recognition complex (ORC/Cdc6) and other factors (?) in an ATP-dependent manner and formation of an open-complex. Cdc6 protein binds to the complex, and MCM, the replicative DNA helicase is loaded on the DNA by helicase loader (?) forming the pre-replication (pre-RC) assembly ready for initiation of DNA replication.

organisms (e.g. *Sulfolobus solfataricus*) no origin was clearly identified (Lopez *et al.,* 1999, Grigoreiv *et al.,* 1998, Salzberg *et al.,* 1998, Kelman and Kelman, 2004). Sequence analysis showed the presence of short inverted repeat elements called origin recognition boxes (ORBs) in the sequence of origins of replication which are well conserved across many archaeal species, bringing forth the hypothesis that like bacteria, archaeal origins are also defined by specific sequence elements (Barry and Bell, 2006). The archaeal origins are A/T rich and contain conserved 13 bp repeats and one or two long stretches surrounding a putative duplex unwinding element (DUE) which are characteristic of the archaeal origin of replication (Lopez *et al.* 1999, Matsunaga *et al.,* 2003). With few exceptions like *Thermococcus kodakarensis,* most origins also contain long inverted repeats (IR) sequences at both the ends with several shorter IRs between them (Grabowski and Kelman, 2003, Fukui *et al.,* 2005). Under supercoiling conditions these IRs could form cruciform structures, which may act as potential protein binding sites for processes like DNA replication (Kelman and Kelman, 2003).

Archaea have circular chromosomes like bacteria and DNA replication is carried out bi-directionally from the origin (Myllikalio and Forterre, 2000). The identified origins of replication are often present upstream of genes encoding homologues of the eukaryotic Cdc6 protein (Lopez *et al.,* 1998, Myllikalio and Forterre, 2000). *S. solfataricus* contains three Cdc6 protein homologues (ssCdc6-1, ssCdc6-2 and ssCdc6-3), origin of replication *oriC*1 and *oriC*2 are found in close vicinity of ssCdc6-1 and ssCdc6-3, and a third origin of replication *oriC*3 is found around 50 Kb from ssCdc6-2 (Lundergen *et al.* 2004, Robinson *et al.,* 2004). Two-dimensional gel electrophoresis, marker frequency analysis and whole genome microarray studies have demonstrated that DNA replication originates from all three origins in all the cell cycles (Lundgren *et al.,* 2004, Duggin *et al.,* 2006).

With the exception of a few methanogenic archaea, like *Methanococcus maripaludis, Methanocaldoccocus jannaschii, Methanopyrus kandleri* all archaeal genomes show the presence of at least one gene encoding a protein with homology to Orc1 and Cdc6 proteins of eukaryotes (Myllikalio and Forterre, 2000). These proteins belong to AAA$^+$ protein family like the origin recognition proteins in bacteria and eukaryotes. Although all archaeal Orc/Cdc6 genes contain regions of homology

to both ORC and Cdc6 genes, in different archaeal genome sequences they are generally annotated as either Orc-x or Cdc6-x. It is not yet clear whether each homologue is able to perform the characteristic function of ORC or Cdc6 or both in the respective organisms. The ssCdc6-1-3 proteins from *S. solfataricu*s, Cdc6 protein from *Pyrococcus abyssi*, ORC1 and 2 in *Aeropyrum pernix*, MthCdc6-1-2 proteins from *M. thermautotrophicus*, have been shown to bind the origin ORB elements (Robinson et al, 2004, Matsunaga *et al.,* 2001, Grainge *et al.* 2006, Capaldi and Berger, 2004, respectively). Additionally, *in vitro* studies have shown that *Sulpholobus* ssCdc6-1 could bind to origin sequences of *Pyrococcus abyssi* and *Halobacterium NRC1* (Robinson *et al.,* 2004).

The crystal structures of Cdc6/Orc1 homologues from *Pyrobacterium aerophilum* and *Aeropyrum pernix* have revealed the presence of a N- terminal RecA-type fold containing a nucleotide-binding pocket connected to an α-helical domain, which suggests that the archaeal ORC/Cdc6 proteins, like their eukaryal homologs, belong to the AAA+ family of ATPases (Ogura and Wilkinson 2001, Liu *et al.* 2000, Singleton *et al.* 2004). The C- terminal of the protein contains domain III, which is structurally related to a Winged Helix domain (WH domain) and bears similarities to the C- terminus of ORC subunits in eukaryotes. The protein binds to DNA by its conserved Winged Helix domain (WH), and the binding is ATP dependent (Grabowski and Kelman, 2001, Kelman and Kelman, 2003, Liu *et al.,* 2000, Singleton *et al.,* 2004).

Using mutational analysis studies, the interaction between MthCdc6-1 and DNA was shown to involve a conserved Arg residue and a guanine in the ORB sequence (Majernik and Chong, 2008). In *Aeropyrum pernix*, Grainge *et al.* (2006) showed that the ORC1 (one of the two ORC/Cdc6 homologues) protein binds to the ORB as a dimer and once all the four ORBs in the *oriC* are bound by ORC1, it causes alterations in both topology and superhelicity of DNA. Further, digestion of ssDNA using the DNA-specific P1 endonuclease protection assay displayed periodicity suggesting that perhaps the wrapping of DNA on ORC1 created distortions in the outer surface of the DNA and the protein-DNA binding was further confirmed by crystallization of ORC1 with ORB sequence (Grainge *et al.,* 2006).

Also, the ssCdc6-1 and ssCdc6-3 co-crystallization with *oric2* DNA in *S. solfataricus* shows that the proteins form a heterodimer and bind to DNA (Gaudier *et al.,* 2007, Dueber *et al.,* 2007). These studies have also revealed that Cdc6 binds to DNA by both N- terminal AAA$^+$ domain and C- terminal WH domains, causing structural distortion of the bound DNA. This distortion probably triggers the unwinding of the duplex DNA to form a replication initiation complex and start DNA replication (Matsunaga *et al.,* 2010). In *P. furiosus,* the Cdc6/Orc1 binds to *oriC* and promotes localized melting of the DNA duplex in the 12 bp A/T rich sequence in the *oriC* region, in the absence of other replication proteins (Matsunaga *et al.,* 2010).

The Cdc6 proteins from *P. abyssi* and *M. thermautotrophicus* show autophosphorylation similar to that of the *S. pombe* Cdc6 homologue (Cdc18) (Grabowski and Kelman, 2004). Cdc6 autophosphorylation is inhibited in the presence of DNA in archaea but DNA has no effect on Cdc18 of *S.pombe*, the significance of this activity is still unknown (Grabowski and Kelman, 2004).

Although Cdc6 is a sequence homologue of eukaryotic ORC/Cdc6 proteins, it also shows strong structural homology to the bacterial DnaA protein (Figure 1.9 a) (Erzberger *et al.,* 2002). Both the proteins contain a different C- terminal DNA binding domain, Cdc6 has Winged-Helix domain (a DNA binding domain formed by a 3 helical bundle and 3-4 strand β-sheet (wing)) whereas, DnaA has Helix-turn-Helix domain (composed of two α-helices joined by a short strand of amino acids) (as shown in Figure 1.9 b, c)) (Erzberger *et al.* 2002, Barry and Bell, 2006, Grabowski and Kelman, 2003). DnaA binds and oligomerizes on the origin sequence containing 9 bp long DnaA consensus box, similarly recent data has shown that archaeal Cdc6 binds and oligomerizes on 13 bp consensus sequences on archaeal *oriC* causing torsional stress that unwinds DNA (Matsunaga *et al.,* 2010). It is therefore highly likely that despite the lack of sequence homology both the proteins function in similar ways.

The archaeal Cdc6 proteins have been shown to have multiple functions including origin recognition, MCM loading, DNA polymerase switching and DNA repair and may act as core regulator of DNA replication and coordinator between origin selection and cell cycle regulation. In *S. sulfolobus* Cdc6-2 protein interacts with other Cdc6 proteins, PolB1 and MCM proteins, similar to Cdc45 protein
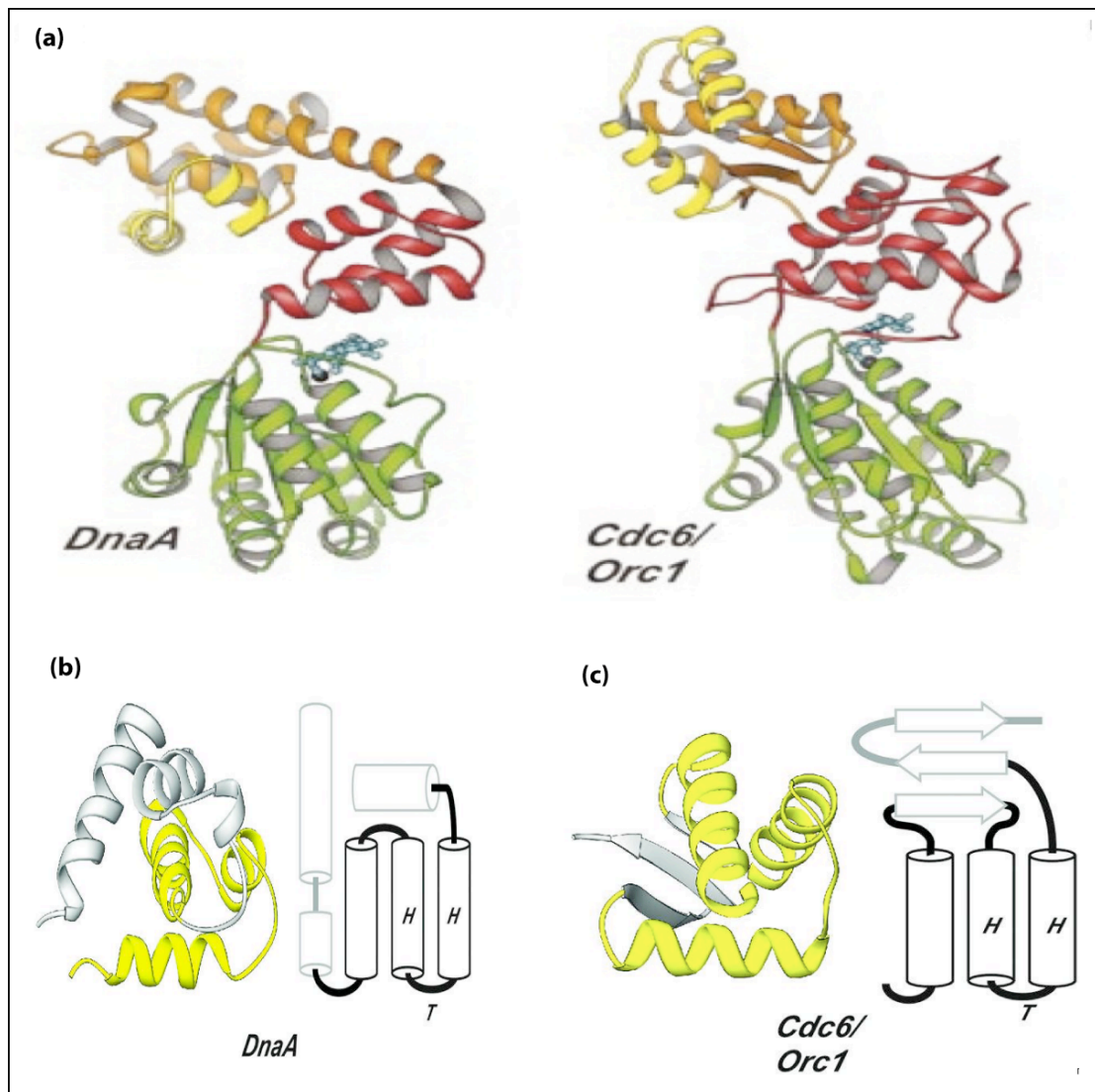
Figure 1.9. DnaA (Bacteria) and Cdc6/Orc1 (Archaea) show structural homology despite having different DNA binding domains (from Erzberger et al., 2002). (a) DnaA possesses a similar structure to archaeal Cdc6/Orc, with AAA+ domains on N-terminal (green) and DNA binding domain on the C- terminal (red). (b) DnaA possesses helix-turn-helix motif in its DNA binding domain, (c) Cdc6/Orc1 possesses winged helix motif. Although the proteins do not share any homology they are structurally similar and also perform similar functions at the origin of replication.

(responsible for the transition from initiation to extension of DNA replication) in eukaryotes (Zhang *et al.,* 2009). There have been several reports of Cdc6 /ORC proteins interactions with MCM proteins (De Felice, 2003, Kasiviswanathan *et al.,* 2005, Haugland *et al.,* 2006). A protein complex recruitment assay showed that MCM proteins are also recruited onto *oriC* by Cdc6/ORC1 but in ATP independent manner (Akita *et al.,* 2010). It is proposed that Cdc6/Orc1 plays a major role in the assembly of MCM at *oriC*, as a helicase loader. Shin *et al.,* (2008) suggested that MthCdc6-2 protein in *M. thermautotrophicus* might act as helicase loader (Kashiviswanathan *et al.,* 2005, Shin *et al.,* 2008). The MthCdc6-2 protein has been shown to inhibit MthMCM helicase activity, suggesting an involvement in loading the DNA helicase at the origin of replication, similar to the DnaB loading carried out by DnaC in bacteria (Shin *et al.,* 2003, 2008, Fang *et al.,* 1999).

Biochemical studies with the MCM proteins from *M. thermautotrophicus, S. solfataricus, Archaeoglobus fulgidus* and *Aeropyrum pernix* revealed that the helicase possesses 3'-5' unwinding activity like the eukaryotic MCM heterohexamer complex and shows single-stranded (ss) and double-stranded (ds) DNA-binding activity, ssDNA and dsDNA translocation and a DNA-dependent ATPase activity (Chong *et al.,* 2000, Kelman *et al.,* 1999). The archaeal MCM proteins can be divided into N- and C- terminal regions. N- terminal region contains three small sub-domains A, B and C (Fletcher *et al.,* 2003), domain A is an α-helical domain, domain B has a Zn-finger motif, and domain C consists of β-strands and forms a β-barrel like structure (Fletcher *et al.,* 2003, Liu *et al.,* 2008). The C- terminal region has a central AAA^{+} ATPase domain and is responsible for DNA helicase activity (Fletcher *et al.,* 2003). The N- terminal is poorly conserved between different species and contributes towards DNA binding, multimerization and regulation, on the other hand, the AAA^{+} domain is responsible for catalytic activity and is well conserved (Chong *et al.,* 2000). The archaeal MCMs have been shown to form various multimers i.e. double hexameric, hexameric, heptameric and filamentous forms, although the double hexamer is the most common (Kelman *et al.,* 1999, Fletcher *et al.,* 2003, Chong *et al.,* 2000, Carpentieri *et al.,* 2002, Barry and Bell, 2005). Recent report by Shin *et al.,* (2009) showed that the oligomerization by MthMCM is

dependent on the concentration of salt and protein in the absence of nucleotide and DNA (Shin *et al.,* 2009). It was observed that with the increasing protein concentration the degree of multimerization increased i.e. from hexamers to dodecamers (Shin *et al.,* 2009).

DNA unwinding by MCMs has been explained by several models such as steric exclusion (Patel and Picha, 2000, Kaplan *et al.,* 2003), rotary pump (Laskey and Madine, 2003, Takahashi *et al.,* 2005, Sakakibara *et al.,* 2009), strand extrusion and ploughshare (Takahashi *et al.,* 2005, Singleton *et al.,* 2004) and looping models (Li *et al.,* 2003, Gai *et al.,* 2004, Brewster and Chen, 2010). In the most recent paper by Fu *et al.* (2011), it is proposed that MCM hexamer is loaded on dsDNA in G1 phase and in S phase, the hexamer excircles ssDNA with no duplex DNA remaining inside the central channel of the helicase leading to further DNA unwinding by steric exclusion (Fu *et al.,* 2011). The model proposed explains that DDK phosphorylation leads to the split of MCM2-7 double hexamer at the Mcm2-Mcm5 interactions (also known as MCM2-5 gate) (Bochman and Schwacha, 2008). The ring split leads to extrution of one strand of the DNA duplex from the central channel, perhaps by the binding of Mcm10 or Sld2 to ssDNA (Kanter and Kaplan, 2010; Warren *et al.,* 2008). The MCM2-5 gate is reclosed and the helicase motor is jump-started by Cdc45 and GINS (Ilves *et al.,* 2010).

After the helicase is loaded onto the DNA, single strand DNA binding proteins (SSB in Crenarchaea and RPA in Euryarchaea) coat the exposed ssDNA (Kerr *et al.,* 2003, Kelly *et al.,* 1998). Primase, DNA polymerase and rest of replication machinery is assembled on the SSB/RPA-ssDNA complex and DNA replication is initiated (Kelman, 2000). Although many of the archaeal proteins participating in the elongation phase show clear similarity to eukaryotic proteins, some are more closely related to bacterial proteins (e.g. the crenarchaeaota SSB) and others are archaeal specific (e.g. PolD) (Kelman and Kelman, 2003). The presence of archaeal-specific elongation proteins may suggest that archaeal-specific initiation factors also exist and are still to be discovered.

1.3.3.1   Archaea as model organisms for eukaryotic DNA replication

Although the process of DNA replication is similar in all three domains of life, archaeal proteins show strong sequence conservation with eukaryotes (Table 1.2) (Brown, 2001, Robinson and Bell, 2005, Yutin *et al.,* 2008, Cox *et al.,* 2008, Brown *et al.,* 1997). It is generally observed that the archaeal homologues of bacterial proteins are involved in metabolism and the eukaryotic homologues are mainly involved in information processing task such as DNA replication, transcription and translation (Edgell and Doolittle, 1997). Archaeal chromosomes are circular, small and have polycistronic transcription units like bacteria. However, the core information processing machinery is related to eukarya.

Table 1.2. A comparison of chromosomal DNA replication in three domains of life, The bacteria – like features and proteins are shown in red, archaea-like are denoted in blue and eukarya-like are shown in black (Kelman and Kelman, 2004).

Features	Bacteria	Archaea	Eukarya
Chromosome	Linear or circular	Circular (Edgell and Doolittle, 1997)	Linear
Replication origin(s)	Single	Single (Lopez *et al.,* 1999) or multiple (Robinson *et al.,* 2004)	Multiple
Origin recognition	DnaA	Cdc6/ORC (Uemori *et al.,* 1997, Myllikalio and Forterre, 2000)	ORC
Helicase	DnaB	MCM (Kelman *et al.,* 1999; Chong *et al.,* 2000; Shechter *et al.,* 2000; Poplawski *et al.,* 2001)	MCM
Helicase loader	DnaA and DnaC	Cdc6/ORC (suggested) (Shin *et al.,* 2009, Akita *et al.,* 2010)	Cdc6 and Cdt1
ssDNA binding protein	SSB	SSB (Wadsworth and White, 2001; Haseltine and Kowalczykowski, 2002) or RPA (Chedin et al .,1998; Komori and Ishino, 2001)	RPA
Primase	DnaG	Primase (Desogus *et al.,* 1999)	Polα/Primase
Sliding clamp	β-subunit	PCNA (Cann *et al.,* 1999)	PCNA
Clamp loader	γ-complex	RFC (Ellison and Stillman, 2001)	RFC
Polymerase	PolC	PolD (Imamura *et al.,* 1995) and/or PolB (Uemori *et al.,* 1995)	PolB

In general, the archaeal proteins are simplified versions of their eukaryotic counterparts (Table 1.2). Archaea possess the Orc/Cdc6 like proteins, which share homology with eukaryotic Cdc6 and C- terminal portion of ORC1 (Majernik and Chong, 2008). The *Methanothermobacter thermautotrophicus* MCM homohexamer is a dramatic simplification of its eukaryotic heterohexamer homologue (Tye, 2000). Other replication proteins like GINS and primase have been characterized in archaea (Bocquier *et al.,* 2001, Liu *et al.,* 2001, Yoshimochi *et al.,* 2008). Therefore, archaea are potentially simple model systems to understand the conserved events in DNA replication.

Despite similarities there are some important replication proteins like Cdc45 (factor required for transition from initiation complex to elongation complex), Cdc7-Dbf4 and Cdc28-Clb (cell cycle dependent kinases) in eukaryotes that are not found in archaea so far. On one hand, their absence postulates a less complicated regulatory mechanism in archaea, alternatively it is also possible that archaea contain novel proteins, which are specific to this domain (e.g. PolD DNA polymerase) and have not been identified so far.

1.4   DNA replication in M. thermautotrophicus

1.4.1   Model organism

The organism used in this study is the euryarchaeon *Methanothermobacter thermautotrophicus*. It is a thermophillic lithoautotrophic archaeon, growing at an optimum temperature of 65°C (Zeikus and Wolfe, 1972). The organism has a doubling time of 3h and can be grown to high cell concentrations (optical density OD_{600} 4). *M. thermautotrophicus* has become a model organism to study DNA replication in archaea, as it possesses homologues of eukaryotic proteins involved in DNA replication, many of which have been characterized (structure and function). The archaeal MCM helicase activity and structure were first identified in *M. thermautotrophicus* (Chong *et al.,* 2000, Kelman *et al.,* 1999, Pape *et al.,* 2003). Homologues of the eukaryotic proteins ORC and Cdc6 are also found in this organism and play an important role in the initiation of DNA replication (Capaldi and Berger, 2004, Matsunaga *et al.,* 2010, Akita *et al.,* 2010, Grabowski and Kelman, 2001). Other replication proteins like RPA, DNA polymerase, RFC, PCNA, GINS, Primase have

been characterized in *M. thermautotrophicus* (Bocquier *et al.,* 2001, Liu *et al.,* 2001, Yoshimochi *et al.,* 2008).

1.4.2 Initiation of DNA replication

The origin of replication for *M. thermautotrophicus* contains multiple 13 bp A/T rich (ORB) sites, which serve as binding site for MthCdc6-1 (Lopez *et al.* 1999, Capaldi and berger, 2004, Majerick and Chong, 2008). These repeats are present upstream of the MthCdc6-1 gene (shown in Figure 1.10) and constitute mini-origin recognition box (ORB) elements (Capaldi and Berger, 2004). MthCdc6-1 has been shown to bind these mini-ORBs *in vitro* (Capaldi and Berger, 2004, Majerick and Chong, 2008).

1.4.3 Proteins involved in initiation of DNA replication

1.4.3.1 MthMCM

M. thermautotrophicus has only one homologue of eukaryotic MCM. The protein has 666 amino acids and a molecular weight of 75 kDa. Structural analysis of *M. thermautotrophicus* MCM using low resolution EM has revealed the formation of hexameric and double hexameric ring like structures (Chen *et al.,* 2005, Chong *et al.,* 2000). Another independent study has shown formation of heptameric rings by MthMCM (Yu *et al.,* 2002). It has been suggested by Yu (2002), that probably, under different conditions and stages of cell growth the MthMCM forms different structures. The N- terminal domain has been crystallized and the structure of MthMCM is well characterized (Fletcher *et al.,* 2003). The N- terminal domain of the protein has three domains: A (N- terminal highly conserved domain containing 4 α-helices), B (contains 3 anti-paralllel β strands and a Zn motif) and C (present between A and B and contain β strands that form a β barrel structure) (Fletcher *et al.,* 2003). The hexameric structure is a result of interactions between domains B and C of the N- terminal MthMCM protein, and the dodecamerization of the protein is mediated by the Zn motif of domain B (Fletcher *et al.,* 2003). The positively charged residues inside the central cavity of the hexamer and some surface residues are responsible for ss and ds DNA binding by the complex (Chong *et al.,* 2000, Fletcher *et al.,* 2003, 2008, Chen *et al.,* 2005, Rothernburg *et al.,* 2007, Costa *et al.,* 2008). The N- and C-
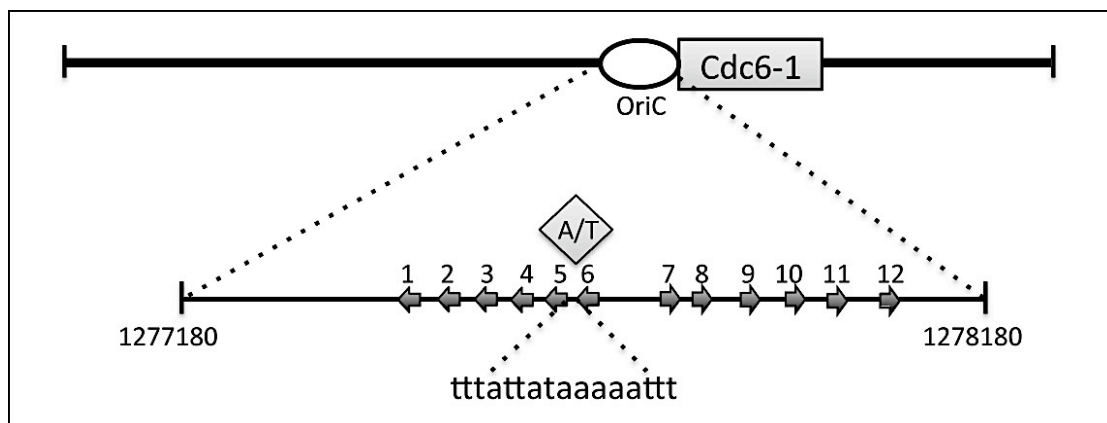
Figure 1.10. A schematic representation of M. thermautotrophicus origin of DNA replication (adapted from Majernik and Chong, 2008, Capaldi and Berger, 2004). The conserved mini-origin recognition sequences (mini-ORBs) present at the origin of replication. The *oriC* is present upstream of MthCdc6-1 (homologue of eukaryotic replication initiation genes ORC and Cdc6, shown in grey box). The A/T rich consensus sequence is also shown.

terminal domains are connected through a conserved loop called as allosteric control loop (ACL). This loop is thought to be responsible for the regulation of interactions between the two domains in response to ATP hydrolysis (Brewster *et al.,* 2008, Sakakibara *et al.,* 2008, Barry *et al.,* 2007). The C- terminal consists of an AAA$^+$ ATPase core and a small winged helix domain, and is considered to be the site of ATP hydrolysis and DNA unwinding (Tye and Sawyer, 2000).

1.4.3.2   MthCdc6-1 and MthCdc6-2

M. thermautotrophicus contains 2 Cdc6/Orc1 homologues: MthCdc6-1 and MthCdc6-2. MthCdc6-1 binds specifically to the 13 bp mini-ORB sequences near the origin of replication and conserved sequences present 3' to mini-ORB sequences (Capaldi and Berger, 2004, Majernik and Chong, 2008). This binding is via a winged helix domain (Kashiviswanathan *et al.,* 2006). It is thought that MthCdc6-1 may bind cooperatively and oligomerize at mini-ORB elements like DnaA at *oriC* in bacteria, in contrast, MthCdc6-2 does not bind specifically to origin sequences (Kashiviswanathan *et al.,* 2006).

MthCdc6-1 and MthCdc6-2 perform different functions in initiation of replication in the cell. It has been shown that while MthCdc6-1 binds only dsDNA with preference for origin-derived sequences; MthCdc6-2 binds both ssDNA and dsDNA with no preferential binding to origin sequences. It was also shown that MthMCM binding to MthCdc6-1 and -2 inhibits their DNA binding (Kashiviswanathan *et al.,* 2006). It was hypothesized by Kashiviswanathan (2005) that MthCdc6-1 may be a functional homologue of eukaryotic ORC and bacterial DnaA, whereas MthCdcd6-2 is a functional homologue of the helicase loader DnaC in bacteria and Cdc6 in eukaryotes since MthCdc6-2 inhibits MthMCM helicase activity more than MthCdc6-1. Recently it has been proposed that MthCdc6-2 acts as helicase loader and loads MthMCM by ring dissociation (Shin *et al.,* 2009).

1.4.3.3   MthCdc6-1 and MthMCM interactions

MthCdc6-1 and MthMCM interactions have been demonstrated using a yeast two-hybrid system (Kashiviswanathan *et al.,* 2005), whereas no interactions between MthCdc6-1 and MthCdc6-2 are reported in the same study. The interaction is mediated by the winged helix domain of MthCdc6-1 and Domain C of the N- terminal

portion of MthMCM. In contrast a full length MthCdc6-2 is required to show appreciable MthMCM binding (Kashiviswanathan *et al.,* 2005).

The helicase and DNA translocation activity of MthMCM is inhibited when bound to Cdc6 proteins. This is similar to the inhibition of DnaB helicase activity when bound to DnaC (helicase loader) (Shin *et al.,* 2003). The presence of MthMCM reduces the binding of MthCdc6-1 and MthCdc6-2 to DNA (Kashiviswanathan *et al.,* 2006), suggesting that MthMCM and DNA may compete for MthCdc6 binding.

Homologues of the eukaryotic initiation proteins ORC, MCM and Cdc6 have been identified in *M. thermautotrophicus*. However, any other proteins helping in the formation of the pre-replication initiation complex and regulation remain unknown. Therefore, the aim of this project was to investigate the presence of proteins performing such functions by interacting with MthMCM, MthCdc6-1 and the origin of replication.

1.4.3.4   Mth203

In a study performed to identify a regulator and initiator for DNA replication in *M. thermautotrophicus*, a yeast two-hybrid screen was carried out using MthCdc6-1 as bait (Figure 1.11 A). Thirteen putative proteins that interact with MthCdc6-1 were identified (Dr. Richard Parker thesis, PhD thesis, 2005). The ORF code and genome annotation for these genes are summarized in table 1.3.

The presence of MthCdc6-1 and MthMCM was expected as it has been previously reported that the MthCdc6-1 interacts with itself to form dimers and probably multimerizes on the replication origin, and interacts with MthMCM during pre-replication complex formation (Shin *et al.,* 2003).

It was very likely that out of the 13 other proteins shown to interact with MthCdc6-1, some might be false positives. However, as very little is known about the archaeal cell cycle and its control, none of the interactions could be immediately ruled out. Out of the interacting proteins, 3 were ATP-dependent RNA helicases belonging to Superfamily 2 (SF2) class of proteins. SF2 proteins are associated with processes involving rearrangements of DNA and RNA (Singleton and Wigley, 2004).

Table 1.3. Positive clones from yeast two-hybrid assay specific for MthCdc6-1. The ORF code and the putative function of the proteins is shown below (Smith *et al.,* 1995).

Mth ORF Code	*M. thermautotrophicus* Genome annotation
203	ATP-dependent RNA helicase, elF-4A family
1408	Cobalamine biosynthesis protein G
446	Sensory transduction regulatory protein
1907	Conserved hypothetical protein
151	Methyl coenzyme M reductase system
492	ATP-dependent RNA helicase system
1412	MthCdc6-1
1770	MthMCM
1624	DNA topoisomerase-1
1458	Hypothetical protein
1215	Fibrillarin like pre-rRNA processing protein
1715	Phycocyanin alpha phycocyanobillin lyase CocE related
656	ATP-dependent RNA helicase protein
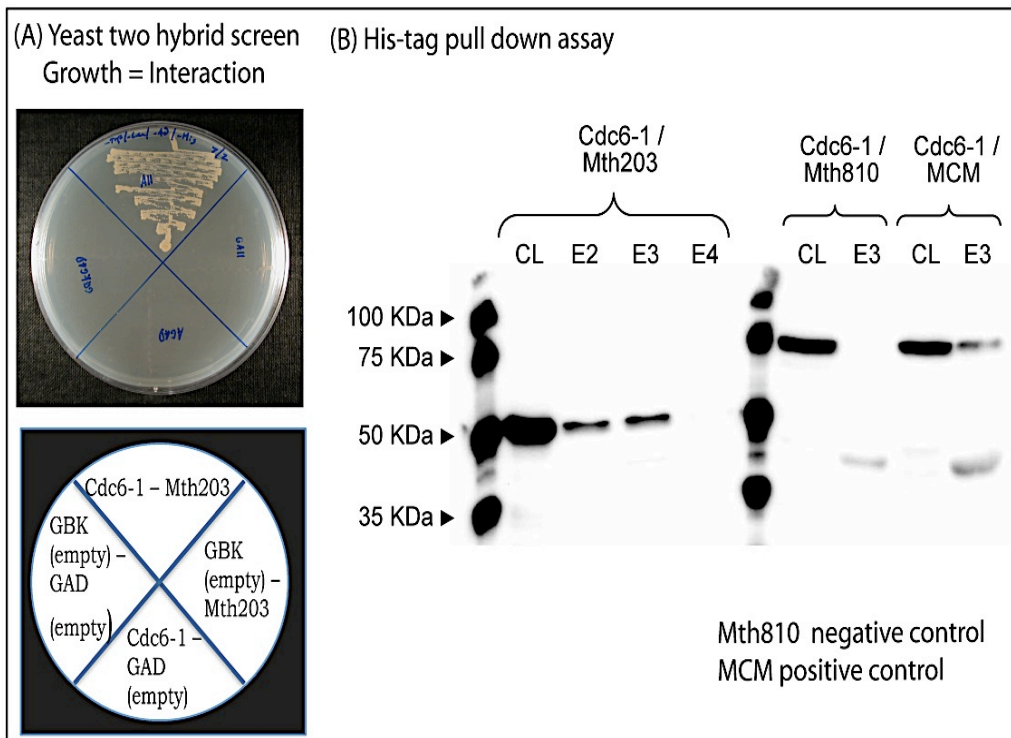
Figure 1.11. Mth203 interacts with MthCdc6-1 in yeast two-hybrid assay. (A) Interaction between MthCdc6-1 and Mth203, SD/-Trp/-Leu/-His/-Ade selection plates were streaked with *S. cerevisiae* AH109 containing the following: pGBKT7/MthCdc6-1 & pGADT7/Mth203 (top quadrant), pGBKT7/empty & pGADT7/empty (left quadrant), pGBKT7/MthCdc6-1 & pGADT7/empty (bottom quadrant) and pGBKT7/empty & pGADT7/Mth203 (right quadrant). (B) S-protein western blot depicting the results of protein pull-down assay using His-MthCdc6-1 as a bait. CL, denotes clarified lysate, E, denotes elution fraction. Mth203 and MthMCM were present in clarified lysate and elution fractions, on the other hand Mth810 was present only in clarified lysate suggesting Mth203 was interacting with MthCdc6-1 and pulled down in the elution fraction.

The initiation and elongation phase of DNA replication involves significant rearrangements of DNA structure such as DNA melting at origin, DNA unwinding at replication forks, priming DNA synthesis, chromatin remodeling, etc. Thus, Mth203, Mth492 and Mth656 were very interesting candidates for study and hence were selected to further investigate MthCdc6-1 binding. His-tag full-length protein pull downs were performed using His-tagged MthCdc6-1 (bait) bound to the column and only Mth203 was successfully pulled down, confirming a definite and strong binding of Mth203 with MthCdc6-1 (Figure 1.11 b). In addition, a 53 amino acid (aa) long peptide from Mth203 C- terminal was shown to bind independently with MthCdc6-1, suggesting that this could be a potential binding site for Mth203 binding with MthCdc6-1.

Mth203 belongs to the Superfamily 2 (SF2) class of DEAD-box proteins (Smith *et al.,* 1997). SF2 helicases perform a wide array of functions in DNA/RNA processing unwinding of polynucleotides and translocation along DNA and often are a part of a multi-protein complex. This may suggest that the interactions between Mth203 and MthCdc6-1 could be a hint of a function performed by Mth203 in replication or transcription (Dr. Richard Parker, PhD thesis, 2005).

A crystal structure of a homologue of Mth203 (MjDEAD) in *M. jannaschii* has been solved but its function is currently unknown (Story *et al.,* 2001) (Figure 1.12). The protein exists as a dimer in crystal. Each monomer has two α/β domains, containing Rec-A like folds found in all known monomeric helicases (Story and Steitz, 1992). From similar structures, it appears that the N- terminal domain contains Walker motifs associated with ATP binding and hydrolysis (Walker *et al.,* 1984). MjDEAD shows 36% sequence similarity with yeast DEAD-box protein eIF4A (SF2 helicase involved in translation initiation in yeast), which is a prototype of SF2 DEAD-box RNA helicase family (Caruthers *et al.,* 2000). No additional domain insertions were observed in the core domain structure of MjDEAD, suggesting MjDEAD and eIF4A proteins are very similar and the conserved features responsible for the helicase activity are possibly present in domains of both the structures. As eIF4A possesses helicase activity *in vitro* (Rogers *et al.,* 1999), MjDEAD crystal structure can serve as a model for minimal helicases; however, no biochemical studies have been performed on the protein to ascertain its helicase activity.
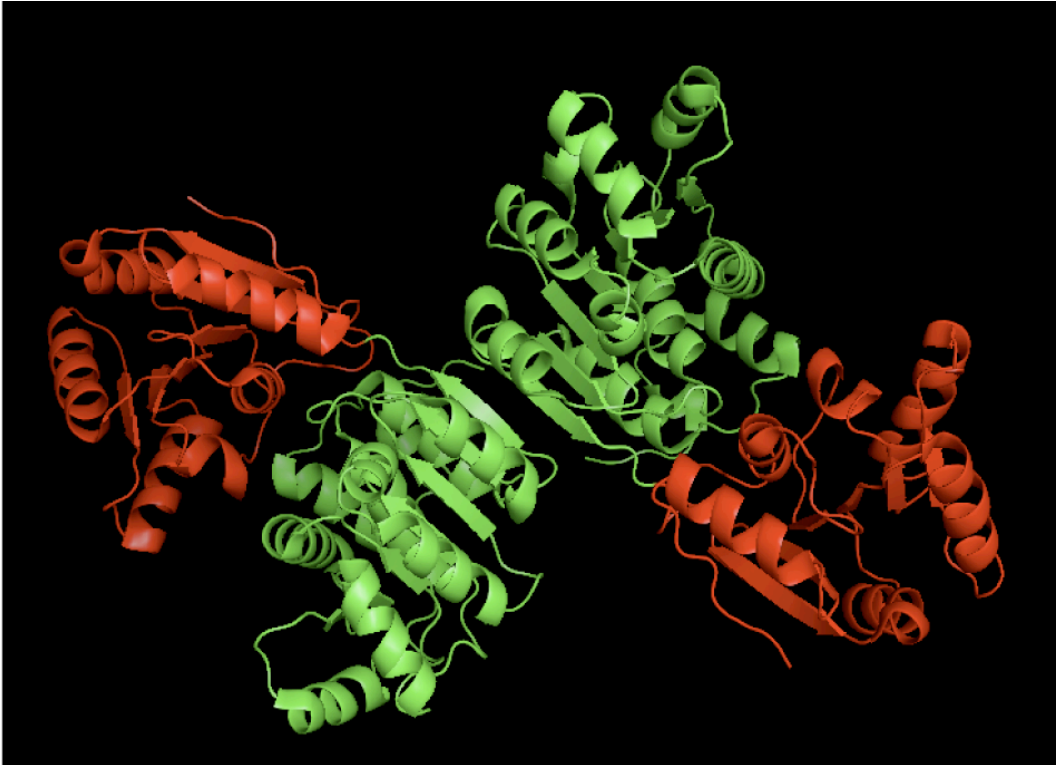
Figure 1.12. Crystal structure of Mth203 homologues MjDEAD (M. jannaschii). The MjDEAD dimer (PDB is 1hv8) has N- terminal (green, residues 1-210) and C- terminal (red, residues 211-365) domains attached by a linker, the two monomers are held together by hydrogen bonding between the β strands of the N terminal domain and hydrophobic interactions between the end β strands on the N-termini.

MjDEAD differs from other helicase structures as the N and C- terminal domains are arranged in an "open form", however this can also be an artifact caused by crystal packing or is a result of the absence of interacting proteins. The two molecules in the dimer interact by hydrogen bonding between the β strands of the N terminal domain and hydrophobic interactions between Phe and Tyr residues on the N- terminal strand (Figure 1.12). This interface is quite similar to the insulin dimer (Adams *et al.,* 1969). The authors also proposed a plausible model for the function of this protein where RNA duplex binds in between the two domains of each molecule (Story *et al.,* 2001). No biochemical activities or interactions with other proteins have been reported for this protein so far. The homologue of Mth203 in yeast, eIF-4A exists as a monomer in its crystal structure and acts as a part of a multiprotein complex along with eIF4B and eIF4H. Together these proteins are necessary to unwind 5' untranslated regions of mRNA preparing it for translation (Caruthers *et al.,* 2000).

1.5   *DEAD-box family of RNA helicases*

Helicases are key modulators in the regulation of different cellular processes (Abdel-Monem *et al.,* 1976). The helicases are generally divided between Superfamily L (SF1) comprising of DNA helicases possessing the classic Walker motif A (G-X-X-G-X-G-K-T) and Superfamily II (SF2) comprising RNA helicases with variation of Walker A domain (A-X-X-G-X-G-K-T) (Luking *et al.,* 1998). RNA helicases are required to activate inactive RNA secondary and tertiary structures, by inducing conformational changes (Tanner and Linder, 2001). The proteins belonging to the DEAD-box RNA helicases form the largest group in the SF2 helicase family, and possess a conserved DEAD domain, first characterised in 1993 (Gorbalenya and Koonin, 1993). These helicases were discovered as special helicases which are able to unwind RNA duplexes and RNA:DNA hybrids (Luking *et al.,* 1998). The conserved domains (I, Ia, Ib II, III, IV, V, VI, DEAD and Q) and functions of DEAD-box helicases are described in Table 1.4 and Figure 1.13 (Tanner *et al.,* 2003, Tanner and Linder, 2001, Rocak and Linder, 2004). The core region in all DEAD-box proteins shows approximately 40% sequence identity and the amino- and carboxyl- termini possess a high degree of sequence and length variability, indicating a role for these regions in individual protein functionality and complex formation with other factors (Luking *et al.,* 1998).

Table 1.4 Functions of conserved motifs of DEAD-box proteins (modified from Tanner and Linder, 2001, Rocak and Linder, 2004). Standard single letter code of conserved motifs: x = any; o = S, T; a = F, W, Y (Cordin *et al.,* 2006).

Motif	Sequence	Function
F	F	Substrate specificty
I	AxxGxGKT	P-loop; Walker A NTP-binding motif; binds phosphates of NTP
Ia	PTRELA	Binds substrate through sugar-phosphate backbone
GG	GG	Substrate specificty
Ib	TPGR	Substrate binding; not highly conserved
II	DEAD	Walker B NTP binding motif; binds β and γ phosphate through Mg^{+2}; coordinates hydrolysis of NTP with water molecule
III	SAT	Binds γ phosphate; links NTP hydrolysis with unwinding activity
IV	LIF	Substrate binding
QxxR	QxxR	Substrate specificty
V	ARGID	Binds substrate through sugar-phosphate backbone; may interact with NTP
VI	HRxGRxGR	Binds γ phosphate; converts NTP binding/hydrolysis with domain 1 and 2 movement
Q	GaxxPoxxQ	ATP binding and hydrolysis

Figure 1.13. A cartoon showing the location of nine conserved motifs of DEAD-box proteins (from Rocak and Linder, 2004). The Q-motif and motifs I and II (Walker motif A and B, respectively) bind ATP and are required for its hydrolysis. Motif I (AxxGxGKT) forms a loop structure (P loop) that accommodates the α- and β-phosphates of ATP. Motif II (or the DEAD motif) forms interactions with the β- and γ-phosphates through Mg^{2+} and is required for ATP hydrolysis. Motif III links the ATP binding and hydrolysis to conformational changes that are required for helicase activity. Motif VI is believed to participate in ATP binding, and mutations therein affect ATP hydrolysis. The structures of RNA helicases indicate that the remaining motifs (Ia, Ib, IV and V) are probably involved in RNA binding, although biochemical data are still lacking.
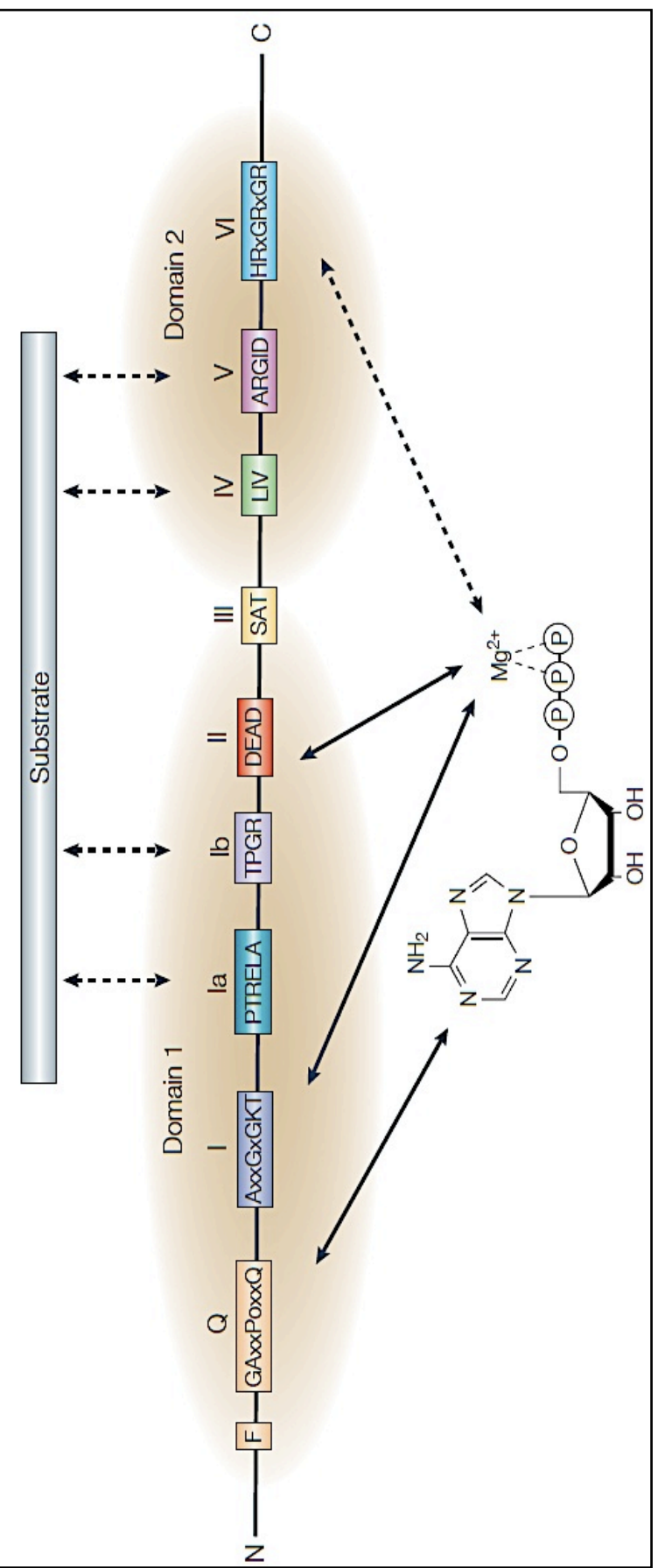
In Hepatitis C virus (HCV), the carboxyl- terminal sequence of NS3 (a DEXH-box SF2 helicase involved in DNA replication) enhances affinity for the substrate and the amino terminal sequence possess a serine protease activity (Kuang *et al.,* 2004). The structural information for DEXD/H-box proteins has been obtained from crystal structures of yeast translation initiation factor eIF4 domains, full-length *Methanocaldococcus jannaschii* DEAD-box protein, HCV NS3, UvrB and human splicing factor UAP56 (Shie *et al.,* 2004, Johnson and Mckay, 1999, Kuang *et al.,* 2004, Cordin *et al.,* 2005, Caruthers *et al.,* 2000). The enzymatic cores of DEAD-box helicases have a typical structure consisting of two discrete domains connected by a linker region forming a cleft to bind NTP (Tanner and Linder, 2001). The amino-terminal domain contains the ATP binding motifs (Q, I, II), the ATP hydrolysing motif (III) and RNA binding motifs (Ia and Ib). And the carboxyl- terminal domain contains RNA binding motif (IV and V), ATPase and unwinding coordinating motif (VI) (Figure 1.13).

In general, RNA helicases are not substrate specific as they interact with polynucleotides by base stacking and sugar phosphate backbone interactions (Kim *et al.,* 1998, Lin and Kim, 1999). It has also been observed that while the core provides substrate-binding affinity, the interactions of flanking sequences with other proteins provide specificity e.g. large complexes found in DNA replication, translation, ribosome precursors and spliceosomes (Singleton and Wigley, 2002). *In vitro* studies show that not all proteins have ATPase and helicase activity, suggesting improper assay conditions or that *in vivo* activity is influenced by substrate and interacting partner proteins, which contribute to the spatial and temporal control of these enzymes (Cordin *et al.,* 2006). The yeast protein eIF4A requires eIF4B or eIF4F for its helicase activity and is a non-processive helicase *in vitro* (Grifo *et al.,* 1984, Bi *et al.,* 2000, Rozen *et al.,* 1989). Dbp5 (DEAD-box helicase in yeast) is inactive in helicase assay *in vitro*, but shows activity when immunoprecipitated with cell extracts (Tseng *et al.,* 1998). Also, most of the DEAD-box proteins have been shown to possess RNA dependent ATPase activity, suggesting their function in context of RNA (summarised in Cordin *et al.,* 2006).

Numerous models are proposed to explain the mechanism of monomeric helicase activity, but two models, the active rolling model and inchworm model, have gained prominence over the years (Soulatanas and Wigley 2000, Tanner and Linder, 2001). The active rolling model requires a dimer in two conformational states and varying substrate affinity (ss or ds polynucleotide) (Figure 1.14). With binding and hydrolysis of NTP the helicase dimer moves along the RNA/DNA in a 'Hand over hand' motion by rearranging the duplex. In the Inchworm model, the monomer changes conformation (domain contraction and release) in conjunction with binding and hydrolysis of NTP (Figure 1.14).

DEAD-box proteins are involved in many processes and may have an important role in RNA metabolism such as RNA splicing (Prp5, Ded1, Prp28), ribosome biogenesis (SrmB, CsdA, DbpA), RNA degradation (eIF4AIII, RhlB), nuclear mRNA transport (Dbp5, Sub2), mitochondrial RNA processing, transcription (Ddx20/DP103, p68, p72) and translation (eIF4A, Ded1) (Rocak and Linder, 2004, Caruthers *et al.,* 2000).

RNA metabolism is present in all organisms and therefore DEAD-box proteins are also present ubiquitously in nature. A sequence database search indicates eukaryotes (*Homo sapiens* (38), *S. cerevisiae* (25), *Arabidopsis thaliana* (55)) possess a higher more number of DEAD-box proteins than archaea (*Methanococcoides burtonii* (1), *Methanothermobacter thermautotrophicus* (3), *Sulfolobus solfataricus* (1)) and bacteria (*E. coli* (5), *Bacillus subtilis* (5)), however not all of them were found to be essential for growth (Rocak and Linder, 2004). In yeast, only 17 out of 25 DEAD-box proteins are essential. However, DEAD-box helicases are completely absent in *Chlamydia sp., Halobacterium salinarum, Pyrococcus furiosus, Methanopyrus kandleri, Aeropyrum pernix* (Rocak and Linder, 2004, Shimada *et al.,* 2009). There are reports of DEAD-box protein synthesis in response to stress conditions like cold-shock (*Methanococcoides burtoni, Methanocaldococcus jannaschii, Anabena variabilis, Bacillus subtilis, E.coli*) (Chamot *et al.,* 1999, Cartier *et al.,* 2010, Hunger *et al.,* 2006, Charollais *et al.,* 2004).

Recently, a few DEAD-box proteins have been reported to bind and unwind DNA:RNA hybrids. Ded1 protein, an essential protein in translation initiation in *S. cerevisiae*, binds DNA independent of ATP concentrations, but is unable to unwind
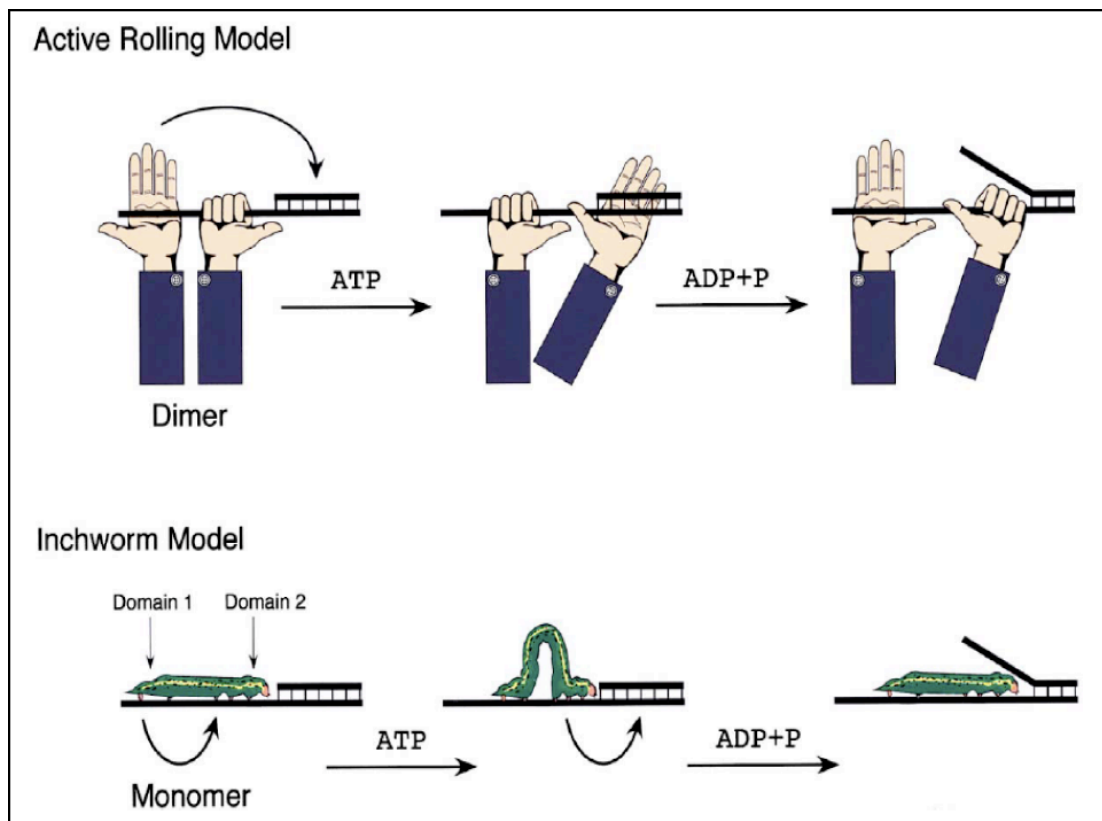
Figure 1.14. Proposed models to explain the monomeric helicase activity (from Tanner and Linder, 2001). The active rolling model depicts the motion of a helicase dimer in 'hand-over-hand' mechanism involving substrate specificity changes and conformational changes of each monomer synchronized with ATP binding and hydrolysis. In the Inchworm model, the monomer moves along the polynucleotide by conformational changes induced by NTP binding and hydrolysis.

duplex DNA (Yang and Jankowsky, 2006). Another *S. cerevisiae* protein Dbp9, involved in 27S ribosomal RNA processing, has also demonstrated ATP dependent DNA:DNA and DNA:RNA helicase activity (Kikuma *et al.,* 2004). There are 35 different DEAD-box RNA helicases in *Homo sapiens* and a few show binding to DNA and are important co-factors in gene expression and regulation. DDX1 protein in *H. sapiens* is hypothesised to play an important role in DNA repair activity of transcriptionally active regions of the genome (Gustafson and Wessel, 2010). DDX17 RNA helicase showed DNA binding at specific Androgen responsive elements (ARE sites) along with androgen receptor (AR) and is possibly involved in DNA remodelling and removal of histone to initiate gene expression (Wong *et al.,* 2008). p72/82 and p68 RNA helicases (gene products of DDX5 and DDX11, respectively) in *H. sapiens* have also shown DNA binding activity (Cordin *et al.,* 2006). These proteins are involved in transcription, cell proliferation, survival, and development of embryos (Janknecht, 2010). Also, these proteins have been characterized as tumour markers and >90% up regulation of these proteins has been observed in breast tumours, colorectal tumours, hyperplastic polyps, adenomas and adenocarcinomas (Janknecht, 2010). Several other DEAD-box proteins are over-expressed in cancer cells (DDX1, DDX43, DDX53, DDX6 proteins in *Homo sapiens*), or required for propagation of viruses (HCV NS3 helicase). Notably, they are also related to other diseases like obesity, Down's syndrome and hepatic fibrosis (Janknecht, 2010). These characteristics make DEAD-box helicases effective drug targets, further suggesting the importance of understanding the mechanism and function of DEAD-box RNA helicases.

Altogether, the emerging picture of DEAD-box proteins depicts a large family of proteins involved in many RNA metabolic processes in all three domains of life. It is likely that in these processes, the DEAD-box proteins are important placeholders or checkpoint proteins, allowing processes to proceed efficiently in one direction and connected with the steps in the RNA metabolism machinery. Despite the advancement of our knowledge regarding the genetic, biochemical, structural and bioinformatics insights into the enzymatic activity, little is known about the specific function of RNA helicases. Furthermore, little information is available regarding the cofactors and substrates of these enzymes. Finally, understanding the mechanism of action will probably be of great importance in understanding RNA metabolism and

devising new drug targets against viruses, and diseases that are the result of RNA helicase malfunction.

1.6 Hypothesis of study

Mth203 was found to bind DNA replication initiation protein MthCdc6-1 in a yeast two-hybrid screen and His-tagged full-length protein pull downs. A 53 aa long C-terminal domain of Mth203 was sufficient for Mth203-MthCdc6-1 interactions (Dr. Richard Parker, PhD thesis, 2006). Recently, Mmp0457, the Mth203 homologue in *M. maripaludis* was co-eluted with MmpMCMA (MCM) by affinity co-purification (Dr. Alison Walters, PhD thesis, 2010).

The discovery of a RNA helicase Mth203 binding to the initiation protein Cdc6 suggests that this protein may be involved in DNA replication initiation, nonetheless, it is also possible that Mth203 has no DNA replication function and is involved in some other related process. Mth203 has been annotated as a putative RNA helicase (Smith *et al.,* 1997), but its biological function and role in any pathway remains unknown. To date, there are no reports of DEAD-box RNA helicases that bind with DNA replication proteins and regulate DNA replication.

The hypotheses of the present study are:

a) Mth203 interacts with MthCdc6-1 and other replication proteins to form a pre-initiation complex by binding to origin

b) Deletion of *mth203* may exert physiological effects by prolonging cell division time and/or affecting the cell structure

1.7   Aim of the project

The present study was carried out to investigate the role of Mth203 by examining the following key questions:

1) What is the distribution of Mth203 homologues in archaea and the nearest homologues in all three domains of life?

2) Does Mth203 interact with DNA/nucleic acids under physiological conditions? Does Mth203 modulate MthCdc6-1–DNA interactions? And is this specific for the MthCdc6-1 homologue? Does Mth203 possess any biochemical activity?

3) What are the proteins interacting with Mth203 in the crude cell extract of *M. thermautotrophicus*?

4) Is *mth203* expressed under normal growth conditions and is it essential for the growth of *M. thermautotrophicus*?

In this study bioinformatics techniques were used to study the homologues of Mth203 in various methanogens, analyse their putative function and elucidate Mth203 protein structure. Mth203 function and protein expression *in vivo* was investigated using various biochemical approaches and genetic manipulation techniques were used to study the effect of the overexpression of Mth203 homologue, Mmp0457 in *M. maripaludis*.

2   Materials and methods

2.1   *M. thermautotrophicus cell culture*

2.1.1   Growth media

M. thermautotrophicus cells were cultured in a fermenter containing NM3 liquid media (Nolling *et al.,* 1991). The components are listed in Table 2.1. Recipe for stock solutions are listed below the table.

Table 2.1. The components of liquid NM3 medium

Components	Amount added for 2.5 L medium
NaHCO$_3$	10 g
100× NM$_3$ salts	25 ml
100× NM$_3$ trace elements	25 ml
1000× Resazurin	2.5 ml
Autoclave	
300× Cysteine-HCl	8.3 ml
300× Sodium thiosulphate	8.3 ml

100× NM$_3$ salts solution: for 500 ml 15 g KH$_2$PO$_4$, 50 g NH$_4$Cl, 30 g NaCl, 5 g MgCl$_2$.6H$_2$O, 3 g CaCl$_2$.2H$_2$O

100× NM$_3$ trace elements: for 1 litre 12.8 g Nitrilotriacetic acid, 1.35 g FeCl$_3$.6H$_2$O, 0.1 g MnCl$_2$.4H$_2$O, 0.024 g CoCl$_2$.6H$_2$O, 0.1 g CaCl$_2$.2H$_2$O, 0.1 g ZnCl$_2$, 0.025 g CuCl$_2$.2H$_2$O, 0.01 g H$_3$BO$_3$,).024 g Na$_2$MoO$_4$.2H$_2$O, 1 g NaCl, 0.12 g NiCl$_2$.6H$_2$O, 0.026 g Na$_2$SeO$_4$.6H$_2$O, 0.05 g AlCl$_3$.6H$_2$O

1000× Resazurin solution: 0.5 g/L Resazurin

300× Cysteine solution: 150 g/L cysteine-HCl

300× Sodium thiosulphate solution: 298 g/L sodium thiosulphate

      Reducers and inoculum were added to the anaerobic tubes using syringes made anaerobic by pre-gassing under a stream of N$_2$ and small gauge needles in order to maintain the anaerobic conditions. Medium components were added to a

2.5 L fermenter (Applikon) and autoclaved. After autoclaving, the water jacket was attached to a recirculating water bath (Heto OBN 8) set to 60°C. The fermenter was connected to a gassing manifold and the medium was sparged with 200 ml/min H_2, 50 ml/min CO_2. The stirrer (Applikon P100) was fixed on the top of the fermenter and the motor controller (Applikon motor controller ADI 1012) was set to an initial speed of 200 rpm. The condenser was attached to a cold water supply and the gas outlet was connected to an exhaust line. While the medium was still hot, 8.3 ml of anaerobic 300× Cysteine and 8.3 ml of anaerobic 300× sodium thiosulphate were added using a needle and syringe through a butyl septum. The medium was left to reduce and cool for one hour. Once the medium was fully reduced (resazurin is a dye which changes colour from blue to colourless when reduced) the stirrer speed was increased to 650 rpm and the fermenter was inoculated with the starter culture.

2.1.2   Inoculation method and growth conditions

2.5 L NM_3 medium in the fermenter were inoculated with 20 ml of *M. thermautotrophicus* culture (OD_{600nm} 4) stored in anaerobic serum bottles at room temperature. The cells were added to the fermenter anaerobically using a pre-gassed needle and syringe through the butyl stopper and into the liquid medium. After inoculation the optical density (OD) was measured at 600 nm every 3-4 hours. Samples for OD measurements were withdrawn using a long needle with a luer adapter placed through the butyl stopper and into the liquid medium. A syringe was attached to the long needle and used to obtain the required volume of medium to measure the OD at each time point. The OD was measured using 1 ml cuvettes in a Bio Mate 3 spectrophotometer blanked against NM_3 medium. Before the measurement of OD a small amount of sodium dithionite powder was added to reduce the sample to ensure that the sample was colourless. Samples with OD above 0.6 were diluted 1:10 in NM_3 medium before taking the measurements. Growth rate of the fermenter grown cultures was calculated as follows:

$$T_d = \ln /(\ln (A/A_0)t)$$

Where T_d is doubling time, ln is natural logarithm, A is OD at the time of sampling, A_0 is starting OD and t is time in minutes.

2.1.3 Synchronisation of *M. thermautotrophicus* cell culture

Once the fermenter OD reached 0.7 the H_2 gas was replaced with N_2 gas (200 ml/min). After 26 hours, the N_2 gas was turned off and the H_2 flow was reinstated. Once the hydrogen flow was resumed, samples were taken for OD examination, ethanol fixed and stored after regular intervals (as mentioned in section 2.1.4).

2.1.4 Cell harvesting and storage

The cells were harvested both aerobically and anaerobically. For aerobic harvesting, once the required OD was reached, the fermenter was dismantled and the medium containing cells was decanted into 1L centrifuge bottles. The cells were centrifuged at 3200 g for 20 min at 4°C using SLC6000 rotor in High speed Sorvall Evolution centrifuge. Samples for protein expression studies were harvested aerobically using a syringe, 1 ml of the diluted sample (OD 0.2) was centrifuged at 16,000 g for 5 minutes (Eppendorf centrifuge, 5810 R) and the cell pellet was stored at -20°C.

Cells were harvested anaerobically for long-term storage and used as inoculum for future fermenter runs. For harvesting, the gas outlet was blocked, causing a build-up of pressure inside the fermenter. Once the pressure had been raised an anaerobic evacuated serum bottle stoppered with butyl rubber was attached to the long needle fitted with a customised connecter. The high pressure in the fermenter forced the media into the serum bottle. Once the bottle was half full it was removed and pressurised with H_2/CO_2 (4:1) to 10 psi and stored at room temperature. These cultures were pressurised with fresh gas at least once a month to keep cells viable for inoculation.

2.1.5 Isolation of *M. thermautotrophicus* genomic DNA

Cells were harvested aerobically and pelleted at 4000 g for 20 minutes at 4°C (Eppendorf centrifuge, 5810 R). 2 g cell pellet was added to a mortar and frozen using liquid nitrogen. Frozen cells were ground using mortar and pestle into a fine powder. Liquid nitrogen was added whenever the cell pellet started forming a paste. 8 ml of lysis buffer (20 mM $Tris_{8.0}$, 5 mM EDTA, 10% sucrose) was added to the cell powder and incubated for 5 minutes at room temperature. After incubation 1.6 ml of 10% SDS was added to the cell lysate and further incubated for 30 minutes at 60°C. 2 ml 5M NaCl and 3 ml MilliQ were added to the solution and incubated on ice

for 40 minutes. The mix was transferred to fresh autoclaved oakridge tubes and centrifuged at 10000 g at 4°C for 15 minutes using the SS34 rotor in High speed Sorvall centrifuge. The supernatant was collected in a fresh oakridge tube, precipitated by adding 14 ml isopropanol and mixed by inversion. The mix was again centrifuged at 10000 g at 4°C for 15 minutes. The supernatant was discarded and the pellet was air-dried. The pellet was then resuspended in 4 ml of 20 mM $TE_{7.8}$ and 20 μl of 10 mg/ml RNase and incubated at 37°C for 30 minutes. An equal volume (4 ml) of phenol:chloroform (1:1) was added to the suspension and mixed by inversion and the top layer containing DNA was extracted carefully using a 1 ml pipette. The extraction was repeated three times to ensure high purity of DNA.

10 g of CsCl were added to the extracted DNA and volume was made to 10 ml using 20 mM $TE_{7.8}$. 300 μl of 5 mg/ml Hoechst H33258 were added to stain the DNA. The entire sample was loaded onto a 12.5 ml heat-seal tube through a 16 guage needle and 5 ml syringe. The tube was topped off with 20 mM $TE_{7.8}$ and sealed. The tube was centrifuged for 24 h in 65K Ti75 rotor at 14°C. After centrifugation, the tube was removed gently and DNA band was visualised under long wave UV light. This DNA band was extracted from the sealed tube using a 16 gauge needle and 1 ml syringe. The tube was carefully pierced using the needle and bright blue DNA band was slowly taken into the syringe. Once the entire band was collected the DNA was dialysed overnight in 2 L dialysis buffer (200 mM NaCl, 10 mM $Tris_{8.0}$, 1 mM EDTA) and 10,000 MWCO dialysis membrane (Sigma). After dialysis the DNA was ethanol precipitated and quantified using the nanodrop (as described in section 2.4.3 and 2.4.4). The extracted DNA was run on 1% agarose gel by gel electrophoresis (as described in section 2.4.12) to check the quality and stored at -20°C until further use.

2.2   *M. maripaludis cell culture*

2.2.1   Growth media

Liquid McCas medium (Moore, 2005) was used for the growth of *M. maripaludis* cultures in tubes and fermenter. The components of the medium are listed in the table below (Table 2.2). Recipes for stock solutions are listed below in the table 2.2.

Table 2.2. The components of liquid McCas medium for the small-scale tube cultures

Components	Amount added for 200 ml medium
General salts solution	100 ml
H_2O	100 ml
$NaHCO_3$	1 g
NaCl	4.4 g
K_2HPO_4 solution	2 ml
$FeSO_4$ solution	1 ml
1000× Trace minerals	0.2 ml
100× vitamins solution	2 ml
Resazurin solution	0.2 ml
Sodium acetate	0.28 g
Casamino acids	0.4 g

General salts solution: for 1 litre 0.67 g KCl, 5.5 g $MgCl_2.2H_2O$, 6.9 g $MgSO_4.7H_2O$, 0.28 g $CaCl_2.2H_2O$ and 1 g NH_4Cl

K_2HPO_4 solution: 14 g/L K_2HPO_4

$FeSO_4$ solution: 1.9 g $FeSO_4$ in 1 litre 10 mM HCl

1000× trace minerals: for 100 ml, 21 g sodium citrate (adjust pH to 6.5), 0.5 g $MnSO_4.7H_2O$, 0.1 g $CoSO_4.6H_2O$, 0.1 g $ZnSO_4.7H_2O$, 0.01 g $CuSO_4.5H_2O$, 0.01 g $AlK(SO_4)_2$, 0.01 g H_3BO_4, 0.1 g $Na_2MoO_4.2H_2O$, 0.025 g $NiCl_2.6H_2O$, 0.2 g Na_2SeO_3 and 0.01 g V(III)Cl

100× vitamin solution: for 1 litre 2 mg biotin, 2 mg folic acid, 10 mg pyridoxine HCl, 5 mg thiamine HCl, 5 mg riboflavin, 5 mg nicotinic acid, 5 mg DL-calcium pantothenate, 0.1 mg vitamin B_{12}, 5 mg *p*-aminobenzoic acid and 5 mg lipoic acid

Reaszurin solution: 1 g/L Resazurin

Sulphide solution: 2.5% (w/v) sodium sulphide. The salt crystals were rinsed with water, dried by blotting and weighed. The crystals were transferred in an anaerobic chamber and dissolved in anaerobic water.

The medium components were added to a stoppered round-bottomed flask and heated under a stream of N_2/CO_2 (4:1), until the medium started to boil. 0.05 g of cysteine HCl was added to the medium immediately stopped and taped. The medium was taken into anaerobic chamber and 5 ml medium was dispensed into glass Balch tubes (Bellco Glass Co.). The tubes were sealed with 20 mm butyl-septum stoppers (Bellco Glass Co.), removed from the anaerobic chamber and crimped with aluminium stoppers (Wheaton). The gas in the tubes was exchanged three times to H_2/CO_2 (4:1) to a pressure of 10 psi using a gassing manifold and then the tubes were autoclaved.

For solid McCas medium, all the medium components were added using the same recipe except 0.4 g $NaHCO_3$ and 3 g Difco Noble Agar were added per 200 ml of medium. Medium components were added to 1 litre flask, 0.1 g dithiothreitol was added per 200 ml medium, the flask was stoppered with rubber butyl stopper and autoclaved. After autoclaving, the flask was immediately transferred to a water bath at 55°C and 0.1 g cysteine HCl was added to reduce the medium. The flask was incubated with a continuous stream of gas N_2/CO_2 (4:1) for 1 hour in the water bath. A colour change was observed from pink to colourless suggesting the medium was reduced. The flask was removed from the water bath and left at room temperature for 5 min. Once cooled, if required neomycin or 8-azahypoxanthine was added (section 2.2.3) to the medium and the medium was taken into the anaerobic chamber and poured into petri dishes. Plates were left to set and dry for 1 h in the chamber and then conditioned using a sealed vessel containing 2.5% Na_2S (w/v), pressurized with N_2 to 10 psi and left at room temperature overnight.

2.2.2 Inoculation method and growth conditions

M. maripaludis was grown on a small-scale liquid media in Balch tubes (Belco Glass Co.). Reducers and inoculum were added to the anaerobic tubes using syringes made anaerobic by pre-gassing under a stream of N_2 and 16 gauge needles in order to maintain the anaerobic conditions. 0.1 ml 2.5% Na_2S (w/v) was added to each tube and then required antibiotics and base analogues were added. 0.1 ml cell culture was added as cell inoculum. After inoculation each tube was pressurised to 30 psi using H_2/CO_2 (4:1) and then incubated at 37°C in a water bath shaking at 110 rpm.

For growth on solid media, 100 µl cells were spread onto pre-conditioned plates in the anaerobic chamber using sterile plating beads (Novagen). Plates were allowed to dry and then placed in a sealed pressure vessel containing 2.5% Na_2S (w/v) in a beaker and pressurised to 25 psi using H_2/CO_2 (4:1). The sealed vessel was then placed in an incubator at 37°C for 3-5 days.

After colonies were obtained, they were picked into 5 ml Balch tubes. The sealed vessel was vented with N_2/CO_2 (4:1) for 15 minutes before opened in the anaerobic chamber, a colony was picked from the plate, resuspended into 100 µl McCas liquid media and added to a prepared McCas media tube containing required antibiotic or 8-aza hypoxanthine using a needle and syringe. The tube was pressurised to 30 psi using H_2/CO_2 (4:1) and then incubated at 37°C in a water bath shaking at 110 rpm.

2.2.3   Antibiotics and base analogues for selection

Antibiotics and base analogue (8-aza hypoxanthine) were filter sterilised and made anaerobic before adding to the growth medium (as described in section 2.6). In the solid medium, the antibiotics were added to a final concentration of Neomycin (1 mg/ml), Puromycin (2.5 µg/ml) and the base analogue, 8 aza hypoxanthine was added to a final concentration of 0.25 mg/ml. In liquid medium, Neomycin was added to a final concentration of 0.5 mg/ml, Puromycin was added at 2.5 µg/ml and 8 aza-hypoxanthine was added at 0.25 mg/ml. 8 aza-hypoxanthine was dissolved in 100 mM NaOH and heated to 60°C before each use to ensure it was in solution.

2.2.4   Large-scale growth of *M. maripaludis*

Large-scale culture of *M. maripaludis* was grown in 2.5 L fermenter. 2 L modified McCas medium was prepared and autoclaved. The medium components are described in Table 2.3. Recipes for stock solutions are listed in section 2.2.1 except for general salts solution II and 1000× cysteine HCl.

Table 2.3. The components of liquid McCas medium for 2 L fermenter

Components	Amount added for 2 litre medium
General salts solution II	1000 ml
H_2O	1000 ml
$NaHCO_3$	10 g
NaCl	44 g
$FeSO_4$ solution	10 ml
1000× trace minerals	2 ml
100× vitamins solution	20 ml
Resazurin	2 ml
Sodium acetate	2.8 g
Casamino acids	4 g

General salts solution II: for 1 litre, 0.67 g KCl and 1 g NH_4Cl

1000× Cysteine HCl: 500 g/L cysteine

After autoclaving, the water jacket was attached to a water bath (Heto OBN 8) set at 37°C. The fermenter was connected to gassing system and the medium was sparged with 200 ml/min H_2, 80 ml/min CO_2 and 200 ml/min N_2. The gas outlet was connected to the exhaust line via a foam trap and the condenser was attached to a cold water supply. The impeller (Applikon P100) was fixed on the top of fermenter and the motor was set to run at the speed of 200 rpm. 2 ml anaerobic 1000× cysteine HCl was added using needle and syringe while the media was still hot. The medium was left to cool and reduce (from blue to colourless) for 1 hour. Once the media was reduced H_2S (1% H_2S/ 99% N_2) was sparged at 15 ml/min and the stirrer speed was increased to 500 rpm. 20 ml sterile anaerobic K_2HPO_4 solution and 100 ml sterile anaerobic 20 × divalent cations solution were added to the medium through butyl rubber septum using needle and syringe. The medium was inoculated using 4 × 5 ml *M. maripaludis* tube cultures that had reached a OD_{600} of 0.7-1.0 in a spectrophotometer (Bio Mate 3). The cells were added to the fermenter by pressurising the tubes at 20 psi and then transferring the culture into the fermenter by placing one end of the double ended vacutainer needle (BD) on the butyl septum

of fermenter and the other end on the tube stopper. After inoculation the growth in the fermenter was monitored by OD measurements every 3-4 hours. Samples were withdrawn using a long needle with a luer adapter placed through the butyl stopper and into the liquid medium. A syringe was attached to the needle and the required volume of cell culture was withdrawn from the fermenter and the OD was measured using 1 ml cuvettes in a spectrophotometer (Bio Mate 3). The spectrophotometer was blanked using McCas liquid medium. Once the culture reached an OD of 0.6 for the accurate measurements of growth the culture was diluted 1:10 using McCas medium. Sodium dithionite was added to each sample to ensure that the solution is colourless at the time of OD measurements.

2.2.5   Cell harvesting and storage

Cells were harvested aerobically by dismantling the fermenter. The cells were transferred to 2× 50 ml falcon tubes and centrifuged at 4000 g for 15 minutes at 4°C. The supernatant was discarded and the cell pellet was stored at -80°C until further use.

2.2.6   Isolation of *M. maripaludis* genomic DNA

Genomic DNA of *M. maripaludis* was extracted from 2 ml of cell culture at OD of 0.7-1.0. The cell culture was centrifuged at 4000 g for 10 minutes at 4°C. the supernatant was discarded and the cell pellet was resuspended in 200 μl of $TE_{8.0}$ and stored at -20°C. The cell suspension was thawed at 37°C and 25 μl 10% SDS and 2.5 μl of 20 mg/ml proteinase K were added and mixed by inversion. The suspension was incubated in a water bath at 37°C for 1 hour. Furthermore, 112 μl 4M NaCl and 75 μl CTAB/NaCl solution (10% cetyltrimethylammonium bromide, 0.7 M NaCl) were added and the suspension was mixed gently by inversion and then incubated at 65°C for 20 minutes. 500 μl Chloroform:IsoAmyl alcohol (24:1) was added, mixed by inversion and centrifuged at 6000 g for 10 minutes at room temperature. The supernatant was decanted to a fresh tube using a wide bore pipette to avoid shearing of DNA. 300 μl of isopropanol was added to the supernatant, mixed by inversion until the DNA precipitated and centrifuged at 10,000 g for 10 minutes at room temperature (Eppendorf centrifuge 5415 D). The supernatant was discarded

and the DNA pellet was air dried and then resuspended in 50 μl TE$_{8.0}$ and stored at -20°C. The DNA concentration was measured using the nanodrop (as described in section 2.4.4) and run on 1% agarose gel by electrophoresis (as described in section 2.4.12) to check the DNA quality.

2.3 *E. coli cell culture*

2.3.1 Growth media

E. coli cells were cultured in Luria-Bertani (LB) liquid medium (Bertani, 1951) unless otherwise stated. For 1 litre liquid medium 10 g Tryptone, 5 g yeast extract and 10 g NaCl were added. For solid medium 15 g/L agar was also added along with the other components.

For transformation, the cells were resuspended in SOC medium containing 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 2.5 ml 1M KCL and 20 ml 1M Glucose.

For MthCdc6-1 protein expression and purification, the expression strain was grown in 2× YT medium containing 16 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl.

For Mth203 protein purification, the cells were grown in ZYP-5052 auto induction medium (Studier, 2005). The recipe is described in Table 2.4. The recipes of the stock solutions are listed below in the table 2.4.

Table 2.4. The components of ZYP-5052 auto-induction medium

Components	Amount required for 1 litre medium
ZY	928 ml
MgSO$_4$ solution	1 ml
50× 5052	20 ml
20× NPS	50 ml
Antibiotics	As needed

ZY medium: for 1 litre, 10 g bacto tryptone and 5 g yeast extract

MgSO$_4$ solution: 24.65 g MgSO$_4$.7H$_2$O in 100 ml MilliQ water

50× 5052: for 100 ml, 25 g glycerol, 2.5 g glucose and 10 g α-lactose

20× NPS: 0.5 M $(NH_4)_2SO_4$, 1 M KH_2PO_4 and 1 M Na_2HPO_4

2.3.2   Antibiotics selection

All antibiotics stocks were filter sterilised before adding to the growth media. Antibiotics were added to liquid and solid media to the final concentrations as follows: Ampicillin 100 μg/ml, Chloramphenicol 34 μg/ml and Kanamycin 30 μg/ml.

2.3.3   Cell harvesting and storage

Small cell cultures (1-200 ml) were harvested by centrifugation at 4000 g for 12 minutes at 4°C in Eppendorf 5810 R centrifuge. Larger cultures were transferred to 1L centrifuge bottles and centrifuged at 4000 g for 20 minutes at 4°C in Sorvall Evolution centrifuge. The supernatant was discarded and the cell pellets were stored at -80°C.

2.4   Cloning and plasmid construction

2.4.1   PCR

PCR was carried out using a Biometra T personal thermocycler (Whatman). Analytical reactions were carried out in a volume of 10 μl. The reaction master mix is described in table 2.5.

Table 2.5. The components of master mix of PCR reaction

Components	Final concentration
10 × Reaction mix	1 ×
Primers	1 μM
dNTPs	0.2 mM
DNA polymerase	1-2 units
gDNA/plasmid DNA	40-10 ng

GoTaq DNA polymerase (Promega) was used for analytical reactions. Preparative PCR was carried out using Fidelitaq DNA polymerase (USB) in 50 μl reaction volume. PCR cloning from *M. maripaludis* genomic DNA was carried out using Fidelitaq DNA polymerase. All the primer sequences used for cloning are listed in appendix A.

2.4.2 PCR purification

PCR products were purified using the Qiagen PCR purification kit (following manufacturer's instructions).

2.4.3 Ethanol Precipitation of DNA

Purified DNA was ethanol precipitated by adding 1/10th of the sample volume of 3 M sodium acetate (pH 5.2) and 3 times sample volume of 100% ethanol. The sample was mixed by inversion and stored at -20°C for 1-24 h to precipitate DNA. The sample was centrifuged at 16000 g for 30 minutes at room temperature in Eppendorf microcentrifuge 5415 D. The supernatant was discarded and the pellet was washed twice with 70% ethanol and centrifuged at 16000 g for 5 minutes. The supernatant was removed and the DNA pellet was air-dried and resuspended in 10 mM $Tris_{8.0}$. The DNA concentration was measured using nanodrop (as described in section (2.4.4)).

2.4.4 Measurement of DNA concentration

DNA concentration was measured on a Nanodrop ND-1000 spectrophotometer (Thermo Scientific). 1 µl of DNA sample was loaded on the nanodrop, which measures the A260/280 ratio of the sample. 10 mM $Tris_{8.0}$ was used to blank the instrument.

2.4.5 Restriction digestion

Restriction digestion was carried out in a reaction volume of 10-50 µl, containing 1 µg DNA substrate, 1× restriction digestion buffer (NEB) corresponding to the restriction enzyme and 10 units of restriction enzyme (NEB). The reaction was incubated at 37°C for 30 minutes in a water bath. The digested product was analysed by gel electrophoresis (as described in section 2.4.12). The DNA of required length was extracted from the gel by cutting the DNA band and purified using Qiagen gel extraction kit (following manufacturer's instructions). The extracted DNA was quantified using nanodrop (as described in section 2.4.4).

2.4.6 DNA ligation

DNA ligation was carried out in a final volume of 10 µl. The reaction mixture contained 2× ligation buffer (Promega), 50 ng purified vector DNA, 1 unit DNA ligase

and insert DNA to a molar ratio of 3:1 or 1:1 (insert:vector). The ligation reaction was incubated at 37°C for 1 hour or 4°C overnight.

2.4.7 DNA sequencing

100-150 ng/μl of DNA in 10 mM Tris$_{8.0}$ was sent for sequencing at University of York, Technology Facility. The resulting sequence was compared to the expected sequence using Seqman Pro DNA analysis software (Lasergene).

2.4.8 Glycerol stocks

E. coli cells containing plasmids of desired sequence were stored at -80°C as glycerol stocks. For 1 ml glycerol stock, 550 μl *E. coli* cell culture OD 1.0 was mixed with 450 μl of 50% glycerol (autoclaved). The mixture was vortexed and stored in -80°C.

2.4.9 Transformation of *E.coli*

Competent *E. coli* Novablue cells (Novagen) were transformed according to manufacturer's manual provided by Novagen. 200 μl glycerol stock was thawed on ice. 1 μl plasmid was added to 20 μl cells and incubated on ice for 5 minutes. The tube was immersed in a 42°C water bath for 45 sec then briefly cooled on ice. 800 μl of SOC medium was added and the cells were incubated for 1 hour at 37°C with shaking (200 rpm) to allow expression of the plasmid encoded antibiotic resistance marker. 100 μl of cells were spread onto a LB agar plate (containing appropriate antibiotic) and incubated overnight at 37°C.

2.4.10 Purification of Plasmid DNA

5 ml of medium containing appropriate antibiotics in a sterile 50 ml falcon tube was inoculated with a single colony of the *E. coli* strain harbouring the plasmid of interest. The culture was incubated overnight at 37°C with shaking (200 rpm). The culture was centrifuged at 4000 g for 5 minutes at room temperature (Eppendorf centrifuge 5810 R), the supernatant was discarded and the cell pellet was then processed using a QIAprep Spin Miniprep Kit (Qiagen) (according to the manufacturer's instructions). Plasmids were eluted in 30 μl of 10 mM Tris$_{8.0}$.

2.4.11 Preparing expression constructs

pET28a expression vector was used to prepare expression constructs for Mth203ΔC53. Mth203ΔC53 was amplified by PCR (57°C / 1.5 minutes Fidelitaq) using Mth203startNhe1 and Mth203C53stopHind3 primers. Mth203ΔC53 was first cloned in the cloning vector pGEMT (Promega) and transformed in *E. coli* Novablue cells. The gene was extracted from cloning vector by RE digestion (Nhe I and Hind III) and ligated into cloning vector pET28a and transformed into cloning Novablue strain. The plasmid was recovered and the insert was checked by DNA sequencing. The plasmid harbouring desired sequence was transformed into *E. coli* expression strain Rosetta pLysS (Merck).

2.4.12 DNA electrophoresis

DNA samples were electrophoresed using 0.8-1% agarose gels at 100 V for 45 minutes. Q-step IV molecular weight marker (York Bioscience) was run in each gel. Samples to be loaded on the gel were prepared by addition of 6 × loading dye (30% glycerol, 0.1% bromophenol blue) to a final concentration of 1 ×.

2.5 Transformation in M. maripaludis

2.5.1 Transformation buffers

Transformation buffer (50 mM Tris, 0.35 M sucrose, 0.38 M NaCl, 1 mM $MgCl_2$, 0.1 ng/ml resazurin, pH 7.5) components were mixed and transferred to a 100 ml serum bottle. The buffer was taken into anaerobic chamber and left overnight to become anaerobic. 1 ml of anaerobic reducing agent was added to every 49 ml transformation buffer, causing the colour change from light pink to colourless. Anaerobic reducing agent was prepared by dissolving 1% cysteine HCl and 50 mM DTT in anaerobic water and the pH was adjusted to 7.5 anaerobically using 1 M Tris base.

The PEG solution (40% PEG 8000) was weighed, taken into the anaerobic chamber, dissolved in anaerobic transformation buffer and left overnight to become anaerobic.

2.5.2 Transformation method

M. maripaludis transformations were carried out using MM900 and S0001 strains (see appendix B for strain details). The plasmids used for transformation reactions were resuspended in TE buffer$_{8.0}$ and made anaerobic overnight before starting the transformation. TE buffer$_{8.0}$ was used as negative control for transformation. 10 ml cultures were grown overnight at 37°C to an OD of 0.7-1.0. After overnight growth the cultures were pressurized to 30 psi using H_2/CO_2 (4:1) and then centrifuged at 1500 g for 15 minutes at room temperature in Eppendorf 5810 R centrifuge. The supernatant was removed by inverting the tube and inserting a needle through the butyl septum to allow the pressure to push the entire spent medium out of the tube. 5 ml transformation buffer was added using an anaerobic syringe and needle, the pellet was resuspended in the buffer and the tube was pressurised to 30 psi as before. The tubes were centrifuged at 1500 g for 15 minutes at room temperature and the supernatant was removed as mentioned above. 0.375 ml of transformation buffer was added to the cell pellets anaerobically through an ananerobic syringe and needle, the pellet was resuspended in buffer by gentle shaking and the tubes were taken into the anaerobic chamber. The metal crimps and butyl stopper were removed and plasmid DNA was added carefully to the cell suspension. In the negative control, TE$_{8.0}$ was added instead of plasmid. The tubes were stoppered and taken out of the anaerobic chamber and crimped with an aluminium cap. 0.225 ml of anaerobic PEG solution was added using an anaerobic needle and syringe. The cells were mixed thoroughly and pressurised with 100% N_2 to 30 psi and incubated at 37°C for 1 hour without shaking.

5 ml McCas medium in an autoclaved tube was prepared for inoculation by adding 0.1 ml of anaerobic 0.25% Na_2S solution and pressurised at 30 psi with N_2. After 1 hour incubation, 5 ml of McCas medium was transferred into the transformation tube by placing one end of the double ended vacutainer needle (Becton Dickinson) through the butyl stopper of transformation tube and just as the pressure was released, the fresh medium tube was inverted and placed on the other end of vacutainer needle. The cell pellet was resuspended in the medium, pressurised at 30 psi using H_2/CO_2 (4:1) and centrifuged at 1500 g for 25 minutes at room temperature (Eppendorf centrifuge 5810 R). The supernatant was removed as

described earlier and 5 ml fresh McCas medium was transferred to the tubes as before. The cell pellet was resuspended and the tube was flushed with H_2/CO_2 (4:1) for 1 minute and then pressurised to 30 psi. OD of the cells was measured and then the cells were incubated at 37°C in a water bath with continuous shaking at 110 rpm. After overnight growth the OD of the cells was measured again, if the transformation was successful it was found to have increased than the previous night. The cells were centrifuged at 1500 g for 15 minutes at room temperature (Eppendorf centrifuge 5810 R) and supernatant was removed as described above. The cell pellets were taken into the anaerobic chamber, resuspended in fresh McCas medium containing 0.1 ml of anaerobic 0.25% Na_2S solution as reducer and plated onto the plates containing desired selection medium (supplemented with antibiotics or base analogues) using plating beads. The plates were then put into sealed vessels pressurised with H_2/CO_2 (4:1) to 25 psi, with 2 ml 25% Na_2S in a beaker and incubated for 3-5 days at 37°C.

2.6   *Preparing anaerobic solutions*

All the solutions were made anaerobic by placing them in glass serum bottles that could be sealed using 20 mm butyl septa. A needle was placed through the butyl septum and a short length of thin-walled PTFE tubing (WZ-06417-31, Cole Palmer) was placed on the other end of the needle. The stopper was placed on the bottle such that the PTFE tubing reached the bottom of the tube. The needle was connected to the gassing manifold and N_2 gas was passed through the solution at low pressure (10 psi). A second open-ended needle was also placed on the stopper to release the pressure developed into the serum bottles. Small volumes of solutions (up to 100 ml) were made anaerobic by passing gas for 2 hours and larger volumes were gassed overnight. After incubation, both the needles were removed, the bottles were crimped using aluminium cap and stored at room temperature or 4°C.

2.7   *Protein purification*

2.7.1   MthCdc6-1 overproduction and affinity purification

MthCdc6-1 was purified according to the method described by (Capaldi and Berger, 2004).

Sample preparation

The glycerol stocks were revived on LB plates containing chloramphenicol (34 µg /ml) and ampicillin (100 µg/ml) and incubated overnight at 37°C. A single colony was incubated in 10 ml YT medium (chloramphenicol (20 µg/ml) and ampicillin (100 µg /ml)) and incubated overnight at 200 rpm at 37°C. 1 L conical flask containing 750 ml of YT medium (chloramphenicol (20 µg/ml) and ampicillin (100 µg /ml)) was inoculated with 2 ml of starter culture and incubated overnight at 200 rpm at 37°C. The cells were harvested by centrifugation for 10 minutes at 4000 g (Eppendorf centrifuge 5810 R). The wet weight of cells was noted and the cell pellet was stored at -80°C until required.

Protein purification

Cells were resuspended and lysed in 5 ml of lysis buffer (40 mM Tris-HCl, pH 8, 500 mM NaCl, 20% glycerol, 1 mM 2-mercaptoethanol and 10 mM imidazole) per gram of cells and 1 µg/ml Pepstatin A, 1 µg/ml Aproteinin, 1 µg/ml Leupeptin, 1 µg/ml Lysozyme, 10 µg/ml DNase I, 5 µg/ml RNase A. The cells were lysed via sonication at 50% power for 1 minute at 4°C using sonicator (Bandelin). The lysate was centrifuged at 50,000 g at 4°C for 30 minutes (Sorvall Evolution centrifuge, SS34) and the supernatant was loaded on 1 ml Ni column in AKTA FPLC system (GE Healthcare). The column was pre-equilibrated with lysis buffer. The column was washed with 40 column volumes of wash buffer (lysis buffer and 75 mM imidazole and eluted by 10 ml elution buffer (40 mM Tris-HCl, 400 mM KOAc, 20 % (v/v) glycerol, 400 mM imidazole plus all the protease inhibitors) in 0.5 ml fractions. On the basis of A280 readings the samples with maximum protein concentrations were collected and pooled. This sample was loaded onto a 1 ml Hi-Trap desalting column AKTA FPLC system (GE healthcare). The protein was eluted with elution buffer 2 (40 mM Tris-HCl, 400 mM KOAc, 20 % (v/v) glycerol). Pure MthCdc6-1 protein was collected in the flow through (0.5 ml fractions), flash frozen in liquid nitrogen and stored in 100 µl aliquots at -80°C until required.

2.7.2 MthCdc6-1 R334A mutant overproduction

R334A is a mutant of Cdc6 in its winged helix (WH) domain required to bind to DNA. This mutation has been shown to prevent the binding of MthCdc6-1 to mini-ORB sequence ORB8 (Majernik and Chong, 2008, Capaldi and Berger 2004). The gene with this mutation was cloned in plasmid pQE30 and transformed in *E.coli*. Frozen glycerol stock of this strain was revived on LB plates with ampicillin (100 μg/ml). Plasmid was purified and transformed in competent Rosetta 2 strain of *E. coli*. The protein was overproduced and purified like wild type MthCdc6-1.

2.7.3 Mth203 overproduction and purification

Sample preparation

The open reading frame of *mth203* was cloned into expression vector pET28a to express Mth203 protein with an N- terminal His tag (Dr. Richard Parker, PhD thesis, 2005). The glycerol stock (R31) was revived on LB plates containing chloramphenicol (34 μg/ml) and kanamycin (30 μg /ml) and incubated overnight at 37°C. A single colony was incubated in 10 ml ZYP 0.8G medium (chloramphenicol (34 μg /ml) and kanamycin (30 μg /ml)) and incubated overnight at 200 rpm in 37°C. 1 L conical flask containing 750 ml of ZY autoinduction medium (Studier 2005) (chloramphenicol (34 μg/ml) and kanamycin (30 μg/ml)) was inoculated with 2 ml of starter culture and incubated overnight at 200 rpm at 37°C. When the final O.D. reached 2.9 the cells were harvested by centrifugation for 10 minutes at 10,000 g. The wet weight of cells was noted and cells were stored at -80°C for further use.

Protein purification

1 g of cells were resuspended in 15 ml lysis buffer (20 mM Tris-HCl, pH 7.9, 1 M NaCl, 10% [v/v] glycerol, 10 mM imidazole, 0.1 mM PMSF, 5 mM β-mercaptoethanol, 1 μg/ml Pepstatin A, 1 μg/ml Aproteinin, 1 μg/ml Leupeptin, 1 μg/ml Lysozyme, 10 μg/ml DNase I, 5 μg/ml RNase A). The resuspended cells were lysed by sonication at 50% power for 1 minute using a sonicator (Bandelin) at 4°C. The cell debris was pelleted by centrifugation at 48,000 g for 30 minutes at 4°C. A sample of clarified lysate was prepared for SDS-PAGE and stored at -20°C. The supernatant was added to His-binding Talon beads prewashed with lysis buffer at room temperature. The

beads were allowed to bind the His-tagged protein for 10 minutes with stirring and centrifuged at 4000 g for 5 minutes. The beads were resuspended in 10 ml of lysis buffer incubated on rocker for 5 min and centrifuged at 1500 g for 5 minutes. The beads were resuspended in wash buffer (25 mM Tris-HCl, pH 8, 1 M NaCl, 10% (v/v) glycerol, 30 mM imidazole, 1 µg/ml Pepstatin A, 1 µg/ml Aproteinin, 1 µg/ml Leupeptin, 1 µg lysozyme, 10 µg/ml Dnase I, 5 µg/ml RNase A) and added to a bench top 1 ml gravity flow column (GE Healthcare). The beads were washed three times with 10 ml wash buffer. Bound protein was eluted by addition of 4 ml elution buffer (same composition as wash buffer with 250 mM imidazole) and collected in 0.5 ml fractions. The samples were run on 12.5% SDS-PAGE. Third and fourth fraction (with highest concentration of protein) were pooled for the next step of purification.

Gel filtration

Mth203 was purified by size exclusion chromatography using a Sephadex 200 16/26 column connected to AKTA FPLC system (GE Healthcare). The column was pre-equilibrated by gel filtration buffer (25 mM Tris-HCl, pH 8, 1 M NaCl). The pooled protein sample was passed through the column at 1 ml/min and purified protein was collected in 1 ml fractions. The most concentrated fractions were selected on the basis of the A280 trace. The protein was run on 12.5% SDS-PAGE to check the purity and quantified by Bradford assay (section 2.7.5). The samples were aliquoted in 100 µl aliquots, cooled by freezing in liquid nitrogen and stored at -80°C freezer for further use.

2.7.4   Mth203ΔC53 overproduction and purification

Mth203ΔC53 is a mutant of wild-type Mth203 where 53 amino acids from the C-terminal are deleted. Mth203ΔC53 (Mth203 gene without 159 bp corresponding to the 53 aa at C- terminal) was amplified by PCR (57°C / 1.5 minutes Fidelitaq) using Mth203startNhe1 and Mth203C53stopHind3 primers. Mth203ΔC53 was first cloned in the cloning vector pGEMT (Promega) and transformed in *E. coli* Novablue cells. The gene was excised from cloning vector by RE digestion (Nhe I and Hind III), ligated into cloning vector pET28a and transformed into cloning Novablue strain. The plasmid was recovered and the insert was checked by DNA sequencing. The plasmid

harbouring the desired sequence was transformed into *E. coli* expression strain BL21 (DE3) Star Rosetta2 (Merck). The protein was expressed like wild type Mth203 and purified (as described in section 2.7.3).

2.7.5   Mmp0457 overproduction and purification for Western blot analysis

2 g cells were harvested in mid-log phase (OD 1.0) and pelleted by centrifugation and resuspended in lysis buffer (1×PBS pH 7.0). The cells were lysed by sonication at 50% power for 1 minute using a sonicator (Bandelin) at 4°C and centrifuged. 100 µl Ni-NTA resin were added to the soluble fraction and incubated at room temperature for 1 h at 4°C. The resin was pelleted by centrifugation and washed twice with 5 ml lysis buffer before bound protein was eluted with 1 ml elution buffer (1×PBS and 1.5M NaCl). The protein was 50-fold concentrated by acetone precipitation (see section 2.12) and resuspended in 20 µl lysis buffer. The concentrated protein sample was run on 12.5% SDS-PAGE and transferred on a PVDF membrane (see section 2.14).

2.7.6   Measurement of protein concentration

Proteins were quantified using Bradford's assay reagent (Bio-Rad) (according to manufacturer's instructions). A standard curve of Bovine serum albumin (BSA) was generated every time the assay was performed.

2.8   SDS-PAGE

SDS-PAGE was performed using a minigel system (CBS Scientific). All protein gels contained 12.5% polyacrylamide. The recipe for 12.5% resolving gel is shown in table 2.6 and 5% stacking gel is shown in table 2.7.

Table 2.6. The components of 12.5% resolving gel

Components	Volume added in 5 ml gel
Protogel premix (30% solution, National diagnostics)	2000 μl
1 M Tris$_{8.7}$	1875 μl
10% SDS	50 μl
MilliQ water	1080 μl
10% APS	16 μl
TEMED	5 μl

Table 2.7. The components of 5% stacking gel

Components	Volume added in 2 ml gel
Protogel premix (30% solution)	333 μl
1 M Tris$_{6.9}$	100 μl
10% SDS	20 μl
MilliQ water	1547 μl
10% APS	16 μl
TEMED	5 μl

The gel assembly was prepared using glass plates and 4 ml of resolving gel was poured between the plates and allowed to set for 30 minutes. Once set, un-polymerized acrylamide was washed off using MilliQ water. 1 ml of stacking gel was poured over resolving gel, 10 well-comb was inserted and the gel was allowed to set for 5 minutes. Samples to load on the gel were prepared by adding 6× Laemmli buffer (30% 2-mercaptoethanol, 12% SDS, 10% glycerol, 0.1% bromophenol blue, 440 mM Tris$_{6.8}$) to a final concentration of 1×, then boiling at 95°C for 5 min. Gels were run at 200 V for 90 minutes in 1× SDS buffer (25 mM Tris, 192 mM glycine and 0.1% SDS). 10 μl unstained precision plus protein standards (Bio-Rad) was run along with samples in each gel.

Coomassie Blue staining

Gels were stained with Coomasie blue R250 stain (40% Methanol, 10% acetic acid, 0.1% Coomassie blue R250 (Fisher)) for 10 minutes and destained with destain solution (40% methanol, 10% acetic acid) until a colourless gels with bright blue bands appeared.

2.9 *Size Exclusion Chromatography – Multi Angle Laser Light Scattering (SEC-MALLS)*

A Shimadzu HPLC system (LC-20AD pump, SIL-20A autosampler and SPD20A UV/Vis detector) was used for size exclusion chromatography and all the samples were separated using a Superose 6 HR10/30 column (GE Healthcare). Column temperature was maintained at 20°C. The samples and running buffer (25 mM $Tris_{8.0}$, 1 M NaCl) were filtered using 0.2 µ filter (Millipore) to remove any particulate matter. 120 µl sample of protein (1 mg/ml) was aliquoted in small vials used to load the sample using an autosampler. MALLS data was gathered using a Wyatt Dawn HELEOS-II light scattering detector, Wyatt Optilab rEX refractive index monitor. The data analysis was carried out using Astra software (Wyatt Technology corporation).

2.10 *CD-spectra analysis*

The integrity of the overproduced proteins Mth203 and Mth203 ΔC53 was measured using a Jasco J810 CD spectrophotometer at 20°C. For analysis, samples were dialyzed into CD buffer (25 mM Tris $_{8.0,}$ 1 M Na_2SO_4) and protein concentration was measured using Bradford assay. 400 µl of sample (0.2 mg/ml) was aliquoted in 1 mm quartz cuvettes and spectrum was obtained in UV range between 195 – 260 nm. Each spectrum was an average of five repeated scans. The molar ellipticity (θ) was calculated using the following equation (Kelly *et al.,* 2005):

$$(\theta)= \Psi M_w/10^4 Lcn$$

Where, Ψ(degrees) is the observable signal, M_w is the molecular weight, L is the path length (cm), c is the protein concentration (g/ml) and n is the number of amino acids. (θ) was then plotted against the wavelength for Mth203 and Mth203ΔC53.

The spectra obtained from the spectrophotometer showed smaller amplitude for Mth203 spectra as compared to Mth203ΔC53. Difference in concentrations of the proteins was the main cause of error in a spectral reading. Comparison of protein concentrations of Mth203 and Mth203ΔC53 revealed a difference of 1.44 units. The Mth203 spectral readings at various wavelengths were adjusted for this factor (multiplication with 1.44) to yield the final spectra.

2.11 Protein pull-down assay

Proteins interacting with Mth203 *in vivo* in *M. thermautotrophicus* cells were identified by pull-down assay.

Column preparation

Mth203 protein was used as bait for the pull-down assay. The protein was covalently attached to the 1 ml NHS column (GE healthcare) using manufacturer's protocol.

Preparation of cell extract

A 10 g freshly harvested *M. thermautotrophicus* cell pellet was resuspended in 20 ml 1× PBS (pH 7.0). The cells were lysed by sonication (Bandelin) 3 times at 50% power for 45 seconds. The cell lysate was centrifuged at 50,000 g for 30 minutes at 4°C. The supernatant was filtered with 0.2 μm filter, collected in fresh falcon tube and kept at 4°C all the time.

Pull-down assay

The prepared affinity column was washed with 10 ml of lysis buffer (1× PBS, pH 7.0) and 5 ml elution buffer (1× PBS, 1.5 M NaCl, pH 7.0), then equilibrated with 15 ml lysis buffer. 20 ml cell extract was passed through the column at flow rate of 1 ml /min. The column was washed with lysis buffer until the protein leaching out of column stopped. The protein bound to the column was eluted in 1 ml fractions using 5 ml elution buffer. The elution fractions containing high concentration of proteins were pooled. The eluted proteins were visualised by SDS-PAGE. The presence of replication proteins in the elution fractions was analysed by western blot (See section 2.14) and full analysis of proteins was performed using MALDI-TOF.

2.12   Acetone precipitation of protein

Proteins were precipitated by adding 8 volumes of ice-cold acetone, mixed and incubated at -20°C for 24 hours. The precipitated protein was pelleted by centrifugation at 16,000 g for 30 minutes at 4°C. The supernatant was removed and pellet was air-dried before resuspending in 25 mM $HEPES_{8.0}$.

2.13   MALDI-TOF

2 ml elution fraction from the pull-down assay was precipitated using acetone precipitation (as described in section 2.12). The sample was resuspended in 100 μl of 25 mM $HEPES_{8.0}$. 15 μl of sample was run on 12% polyacrylamide gel and the bands of interest were excised and sent for MALDI-TOF-TOF analysis. The analysis was carried out using an Applied Biosystems 4700 Proteomics analyser.

2.14   Western blotting

Samples were run on 12.5% polyacrylamide gels and transferred to a PVDF membrane using the Transblot-SD semi-dry transfer cell (Bio-Rad). Gels for western blot analysis were loaded with 10 μl pre-stained precision plus protein standards (Bio-Rad) so that efficient blot transfer could be confirmed. Before starting the assembly of western blot PVDF membrane was activated by soaking in methanol for 15 sec, washed with MilliQ water for 2 minutes and incubated at room temperature in anode buffer II (25 mM $Tris_{10.4}$, 10% methanol). The gel was pre-soaked in cathode buffer for 15 minutes. The blot was assembled by semi-dry transfer method using Transblot 3D electrodes unit (Bio-Rad). Two sheets of 3 mm filter paper (Whatman) soaked in anode buffer I (0.3 M $Tris_{10.4}$, 10% methanol) were placed on the anode plate, followed by one piece of filter paper soaked in anode buffer II (25 mM $Tris_{10.4}$, 10% methanol) and then a piece of activated PVDF membrane (GE Healthcare) was added to the stack. The gel (soaked in cathode buffer) was placed on top of the membrane, followed by three filter papers soaked in cathode buffer (25 mM $Tris_{9.4}$, 40 mM glycine, 10% methanol). The cathode electrode was placed on the top of the stack and transfer was carried out using 70 mA per gel. After transfer, the membrane was washed 3 times for 5 minutes in 20 ml PBS-T (1× PBS with 0.02% Tween-20). Membrane was blocked in 20 ml blocking buffer (3% Marvel dried milk, 1× PBS-T).

Blocking was carried out for 1-12 hours at room temperature with continuous shaking.

After blocking, the membrane was washed 6 times for 5 minutes incubations in 20 ml PBS-T at room temperature. Primary Antibody stocks were prepared as 1:1000 dilution in PBS-T supplemented with 3% BSA (Sigma). Rabbit polyclonal antibodies were used for MthCdc6-1 (CS 1090) and MthMCM. Mouse monoclonal antibodies were used for His-tag (Qiagen) and mouse polyclonal antibodies were raised using full length Mth203 (GenScript). Primary antibody incubations were carried out for 1 hour at room temperature in 20 ml antibody solution. After washing the membrane for 6 times for 5 minute incubation in 20 ml PBST-T, HRP-conjugated secondary antibody solution (1:10,000 dilution in PBS-T) was added to the membrane and incubated for 1 hour. Mouse secondary antibody (Amersham) was used for anti-His-tag and anti-Mth203 primary antibodies. Rabbit secondary antibody (Sigma) was used for anti-MthCdc6-1 and anti-MthMCM primary antibodies. Thereafter, the membrane was washed 6 times for 5 minutes incubation in 20 ml PBST.

Cross-reactive bands were detected using Supersignal ECL kit (Peirce) and imaged using photographic film (Thermo Scientific) developed in an XO graph compact X9 analyzer (Packard).

2.15   Southern blotting

Southern blotting was carried out using the DIG-labelling and detection starter kit and the DIG wash and block buffer set (Roche).

2.15.1   Probe labelling

1 µg target DNA was gel purified (as described in section 2.4.2), ethanol precipitated (as described in section 2.4.3) and resuspended in 16 µl of 10 mM $Tris_{8.5}$. The DNA was denatured at 100°C for 10 minutes and cooled immediately by placing on ice. The DNA was labelled by 4 µl DIG-High Prime 5× labelling mix (Roche) and incubated for 20 hours at 37°C for efficient labelling. The reaction was stopped by adding 2 µl 0.2 M $EDTA_{8.0}$ and the labelled probe was heated at 65°C for 10 minutes. The probe was stored at -20°C until required. Probe yield was estimated by applying a series of

1 μl spots of approximately 1 ng/μl diluted probe to a piece of nylon membrane (Bio Trans). In parallel, 1 ng/μl solution of the DIG-labelled control DNA in 10 mM $Tris_{8.5}$ was prepared and a series of dilutions were also applied in 1 μl spots on the nylon membrane. The DNA was cross-linked on the membrane using UV-light (Bio-Rad) for 1 minute at 254 nm. The membrane was washed for 20 minutes in wash buffer (0.1 M maleic acid, 0.15 M NaCl, 0.3% v/v Tween 20, pH 7.5) with continuous shaking. The wash buffer was discarded and the membrane was washed for 30 minutes at room temperature in blocking buffer (Roche). The antibody solution (1:10,000 antibody in blocking buffer) was added to the membrane and incubated for 30 minutes at room temperature, washed twice for 15 minute at room temperature in washing buffer. The membrane was equilibrated in detection buffer (0.1 M $Tris_{9.5}$, 0.1 M NaCl) for 2-5 minutes and CSPD detection reagent (Roche) was applied over the membrane and incubated for 5 minutes. Excess reagent was removed, the membrane was sealed in a plastic envelope and incubated at 37°C for 10 minutes before exposure to chemiluminescence scanner for 5-20 minutes using an exposure cassette (Fuji). The yield of labelled probe was estimated by comparison of the signal from control and test-probe spots.

2.15.2   Gel electrophoresis and blotting

1 μg of genomic DNA was digested using 10 U PsiI at 37°C, overnight and run on a 0.8% agarose gel in 1x TBE running buffer (90 mM $Tris_{8.0}$, 90 mM boric acid, 2 mM EDTA) to separate DNA bands by gel electrophoresis (as described in section 2.4.12). DIG labelled markers (Roche) were also loaded on the gel. The gel was stained in 0.5 μg/ml EtBr in 1× TBE for 20 minutes, rinsed with MilliQ water briefly and then photographed to visualize that DNA has been digested. If the target DNA was larger than 5 kb, the DNA was depurinated by washing with 250 mM HCl for no longer than 10 minutes and rinsed with MilliQ water. The gel was denatured by washing twice for 15 minutes in gel wash buffer (0.5 M NaOH, 1.5 M NaCl) at room temperature. The gel was rinsed with MilliQ water and neutralised by washing twice for 15 minutes in neutralisation solution (0.5 M $Tris_{7.5}$, 1.5 M NaCl). The pH of the gel was measured by pressing pH paper on the gel to ensure the pH was below 9. The gel was equilibrated in 20 × SSC buffer (3 M NaCl, 0.3 M sodium citrate, pH 7.0) for 10

minutes and the DNA was transferred to a positively charged nylon membrane (BioTrans) using capillary transfer method in 20× SSC buffer overnight.

After transfer, the membrane was placed on 3 MM filter paper (Whatman) pre-soaked in 20 × SSC buffer and the DNA was fixed using UV cross-linking for 1 minute at 254 nm. The membrane was rinsed with MilliQ water and dried by placing on a dry piece of 3 MM filter paper at room temperature. The membrane was placed in a hybridisation bottle and incubated in hybridisation buffer (Roche) for 30 minutes at 39°C with slow rotation. 25 ng of labelled probe / ml in hybridisation buffer was added to 50 µl of water and boiled at 100°C for 5 minutes and cooled immediately by placing on ice and then added to pre-warmed hybridisation buffer. The first aliquot of hybridisation buffer was poured out of the hybridisation bottle and fresh buffer containing probe was added. The membrane was incubated with the probe at 39°C overnight. The membrane was washed twice for 5 minutes in 2 × SSC/0.1% SDS at room temperature and then twice for 5 minutes with 0.5× SSC/0.1% SDS at 65°C. The membrane was rinsed for 2 minutes in wash buffer (0.1 M Maleic acid, 0.15 M NaCl, 0.3% v/v Tween 20, pH 7.5), incubated in 1 × blocking solution (Roche) for 1 hour and 1:10,000 antibody solution (Roche) for 30 minutes. The membrane was washed twice for 15 minutes in wash buffer and equilibrated in detection buffer for 3 minutes at room temperature. The membrane was developed as described in section 2.14.1.

2.16   *Fluorescence Anisotropy*

Fluorescence anisotropy was used to study DNA-protein and protein-protein interactions.

2.16.1   Anisotropy substrates

ORB8 sequence was used as single ORB specific substrate and scrambled ORB sequence was used as non-specific substrate. Sense and antisense oligos were ordered from (MWG operon) (sequences of single ORB specific and non-specific substrates are in the appendix C).

4ORB oligos were amplified from *M. thermautotrophicus* genome using specific primers (Appendix A). For specific substrate containing origin recognition

box 7-10 (Capaldi and Berger, 2004) 4ORBspF and 4ORBspR primers were used. Non-specific substrates were amplified using the primers, 4ORBnspF, 4ORBnspR and $ORBnsp2F, 4ORBnsp2R primers respectively (Appendix A).

2.16.2   Oligo annealing

0.1 mM Oregon green (λex 495 nm, λem 515 nm) labelled sense and unlabelled anti sense oligos (sequences of single ORB specific and non-specific substrates are in the appendix C) were annealed by heating at 95°C for 5 minutes. The mix was cooled slowly at room temperature and then diluted 10 × in annealing buffer (100 M Tris-HCl, pH 7.5, 500 mM NaCl, 10 mM EDTA). The annealed oligos were run on 12% polyacrylamide gel (as described in section 2.4.12) and visualised on phosphorimager (Bio-Rad).

2.16.3   Protein labelling

MthCdc6-1 was N- terminally labelled with amine reactive Oregon green succinimidyl ester (Invitrogen). The protein (1 mg/ml) was dialysed in carbonate buffer (0.2 M sodium bicarbonate, pH 8.3) and labelled according to manufacturer's instructions. After labelling MthCdc6-1 was dialysed overnight in elution buffer (40 mM Tris-HCl, 400 mM KOAc, 20 % (v/v) glycerol), the concentration was measured by Bradford assay (section 2.7.5) and stored at -80°C.

2.16.4   Anisotropy assay

DNA:Protein binding

10 nM of labelled annealed oligo duplex was added to anisotropy buffer (40 mM Bis-Tris propane, pH 7.0, 70 mM KOAc, 2 mM MgCl2, 3 mM DTT, 20% glycerol, 0.1 mg/ml BSA). Anisotropy and change in polarization of the polarized light (λex 495 nm, λem 515 nm) were measured with buffer as blank, with oligos and later change in anisotropy was measured with increasing concentrations of the protein of interest. The molecular binding and K_m values were calculated using one-site saturation equation in Sigma plot 12 software (Systat Software Inc). Average change in anisotropy was plotted against protein concentration and fitted to a ligand binding equation to obtain apparent dissociation constant K_d (app).

$$y = B_{max} * abs\ (x) / (k_d + abs(x))$$

where, (y)=specific binding, x=concentration of free ligand, K_d=concentration of ligand to reach half of maximum binding and B_{max}=maximum number of binding sites.

Protein:protein binding

100 mM Oregon-green labelled *M. thermautotrophicus* MthCdc6-1 was added to anisotropy buffer. Anisotropy and change in polarization of the polarized light (λex 495 nm, λem 515 nm) were measured with buffer as blank, with OG-MthCdc6-1 and later change in anisotropy was measured with increasing concentrations of Mth203. The change in anisotropy was plotted against protein concentration to visualize molecular binding. Sigma plot 12 software was used to calculate the K_m values as mentioned above.

2.17 Auto-phosphorylation assay

A modified autophosphorylation assay was used (Grabowski and Kelman, 2001). MthCdc6-1 (250 ng) was incubated for 30 minutes at 50°C in a reaction mixture containing 0.03 mM [γ-32P] ATP, 25 mM HEPES (pH 7.5), 5 mM $MgCl_2$, 70 mM NaCl, 5% glycerol and 2 mM dithiothreitol (DTT) in the absence or presence of 1 μg of single- or double-stranded ϕ×174 DNA (Sigma) and the presence or absence of DNA binding proteins Mth203 (250 ng) and RPA (kindly received from Dr. Ed Bolt, University of Nottingham) (250 ng). The reaction was carried out in 10 μl reaction volume. Following incubation, the proteins were separated on 12.5% SDS–PAGE, visualized by Coomassie blue staining and photographed (as described in section 2.8). The gel was dried before exposure to a chemiluminescence screen for 24-48 hours using an exposure cassette (Fuji) and visualised by phosphorimager (Bio-Rad).

2.18 DNA Helicase assay

2.18.1 Preparation of substrate

The helicase substrate was prepared by labelling HS2 oligo with γ-^{32}P-ATP and annealing to HS1 oligo (see appendix A). Oligo HS2 was labelled in a 12 μl reaction. The master mix composition of the labelling reaction is described in table 2.8.

Table 2.8. The components of helicase substrate labelling reaction master mix

Components	Final concentration
HS2 oligo	5 μM
Polynucleotide kinase (Promega)	1 unit
PNK Buffer	1 ×
γ-^{32}P-ATP (3000 Ci/mmol)	3.6 μl

The reaction was incubated at 37°C for 1 hour and the enzyme was denatured at 90°C for 10 minutes. 0.5 μl of the labelling reaction was added to 24.5 μl of stop buffer (120 mM EDTA, 0.6% SDS, 60% glycerol, 0.1% bromophenol blue) to be used for substrate quantification as 100 nM control sample. The labelled HS2 oligo was annealed to complimentary unlabelled HS1 oligo in 40 μl reaction as described in table 2.9. The stocks composition is mentioned below the table.

Table 2.9. The components of helicase substrate annealing reaction

Components	Final concentration
Annealing buffer	1 ×
Labelled HS2 oligo	0.625 μM
HS1 oligo	0.625 M

1× annealing buffer: 200 mM HEPES pH 7.5, 250 mM NaCl, 5 mM EDTA

The reaction mix was heated at 95°C for 5 minutes and slowly allowed to cool at room temperature.

The annealed substrate was run on 12% acrylamide gel (1× TBE, 12 % acrylamide bis-acrylamide from a 19:1 20% stock, 0.07% APS, 0.1% TEMED). The substrate was run on the gel for 60 minutes at 100 V in 1× TBE buffer. The gel was wrapped in cling film and exposed to photographic film (Fuji) for 5 minutes. The band containing annealed substrate was cut from the gel transferred to a fresh microfuge tube, crushed and weighed. The substrate was eluted from the gel by adding 3 μl of PAGE elution buffer (0.5 M Na-acetate, 10 mM Mg acetate, 1 mM

EDTA, 0.1% SDS) for every 1 g of gel and incubating for 2 hours at 37°C. The gel and buffer were centrifuged at 16,000 g for 2 minutes and carefully 50% of the buffer was removed into a fresh tube without disturbing the gel pellet. An equal amount of fresh PAGE elution buffer was added into the tube containing the gel pieces and incubated at 4°C overnight. Buffer was removed as stated previously, 1 μl of 20 mg/ml glycogen was added to the eluted substrate and the DNA was ethanol precipitated (as described in section 2.4.3) and resuspended in 50 μl TE pH 8.0.

The substrate was quantified by spotting 3×1 μl spots on DE81 paper (Whatman) along with 100 nM control substrate spots (3 ×). The paper was washed 3 times for 5 minutes with 0.5 M phosphate buffer, pH 7.0, then with 70% and 100% ethanol for 10 minutes each. The paper was dried and the spots were imaged using a phosphorimaging screen in a phosphorimager (Bio-Rad). The substrate was quantified using Quantity One software (Bio-Rad) and stored at 4°C.

2.18.2   Helicase assay

The helicase reactions were carried out in 10 μl final volume and the reaction mix was prepared on ice as described in table 2.10.

Table 2.10. The components of DNA helicase assay master mix

Components	Final Concentration
HEPES$_{7.5}$	20 mM
MgCl$_2$	10 mM
DTT	2 mM
BSA	1 mg/ml
ATP	5 mM
Labelled substrate	2 nM
Proteins	Varying concentrations

The reaction was incubated at 50°C for 1 hour and stopped by addition of 4 μl proteinase K (20 mg/ml) and 3 μl of STOP buffer (200 mM EDTA, 1% SDS, 20% glycerol). The reaction was loaded on 12% 1× TBE polyacrylamide gel and run for 90

minutes at 130 V. The gel was fixed by washing in 7 % acetic acid for 5 minutes and then dried. The dried gel was exposed to a phosphorimager screen for 24-48 hours using an exposure cassette (Fuji) and visualised by phosphorimager (Bio-Rad). The unwinding activity was quantified using Quantity One software (Bio-Rad).

2.19   RNA helicase assay

2.19.1   Preparation of substrate

RNA helicase substrate was prepared by *in vitro* transcription of DNA A and DNA B substrates (see appendix A). 100 mM DNA templates and 100 mM T7 promoter template were annealed in a 50 µl reaction mixture containing DNA and 10 mM $TE_{8.0}$. The mixture was heated at 95°C and slowly cooled to room temperature. The DNA concentration was measured using nanodrop (as described in section 2.4.4). *In vitro* transcription was carried out following manufacturer's instructions of Megascript T7 high yield transcription kit (Ambion) (table 2.11).

Table 2.11. The components of in vitro transcription master mix for 50 µl reaction

Components	Final Concentration
10 × Transcription buffer	1 ×
DNA template	1 µg
ATP	0.5 mM
UTP	0.5 mM
CTP	3 µM
GTP	0.05 mM
Cap Analog (m^{7}G(5')ppp(5')G)	1 A_{254} U
CTP- α^{32}P	825 nM
Enzyme	40 U

The mix was prepared and incubated at 37°C for 2 hours for maximum yield. For unlabeled RNA 0.5 mM CTP was used instead of the radiolabeled CTP- α^{32}P. After incubation the reaction mix was run on 12.5 % polyacrymide 1 × TBE gel for 90 minutes at 150 V. The gel was wrapped in cling film and exposed to photographic

film (Fuji) for 5 minutes. The band containing substrate was cut from the gel and transferred in a fresh microfuge tube, crushed and weighed. The substrate was eluted from the gel as described in section 2.18.1. The RNA was resuspended in 50 μl TE pH 8.0 and stored at -80°C in 1 μl aliquots.

The amount of RNA synthesised was calculated using the counts per minute (cpm) values and the formula provided in the manufacturer's protocol. 1 μl of RNA transcription product was mixed with 2 ml scintillation fluid (National Diagnostics) and cpm values were calculated using P^{32} quick count method on a liquid scintillation counter (TriCarb Liquid Scintillation Counter). 0.1 μl of fresh CTP- $\alpha^{32}P$ in 2 ml scintillation fluid (National Diagnostics) was used as a control. Unlabelled RNA was quantified using nanodrop.

2.19.2   Helicase assay

RNA helicase assay was modified from the protocol in (Rozen *et al.,* 1990) and the reaction mix was prepared on ice as described in table 2.12.

Table 2.12. The components of RNA helicase assay master mix for 10 μl reaction

Components	Final concentration
HEPES$_{7.5}$	20 mM
KCl	70 mM
DTT	2 mM
ATP	1 mM
MnCl$_2$	0.5 mM
GTP	0.1 mM
RNAsin (Promega)	40 U
tRNA	1 μg
Labelled RNA substrate	900 ng
Protein	Varying concentration

Tubes containing increasing amount of protein (0-5 μg) were prepared and the reaction mix was added to make up the final volume 10 μl. The reaction was mixed

and incubated at 37°C for 30 minutes. The reaction was stopped by adding 1 μl Proteinase K (20 mg/ml) and 4 μl of RNA-STOP buffer (0.5 M EDTA, 10 % SDS) and incubated at 37°C for 10 minutes. 2 μl DNA loading dye was added to the reaction and the reaction was run on 12.5 % polyacrymide 1 × TBE gel for 90 minutes at 150 V. The gel was fixed by washing with 7 % acetic acid for 5 minutes and dried at 60°C for 20 minutes. The dried gel was exposed to chemiluminescence screen for 24-48 hours using a exposure cassette (Fuji) and visualised by phosphorimager (Bio-Rad). The unwinding activity was quantified using Quantity One software (Bio-Rad).

2.20   NTPase assay

NTPase activity of Mth203 was calculated using a colorimetric assay based on molybdate/malachite green reaction from a previous study (Rocak *et al.,* 2005). The reaction was carried out in a 50 μl reaction volume (Table 2.13).

Table 2.13. The components of NTPase assay reaction mix

Components	Final concentration
$HEPES_{8.0}$	20 mM
$MgCl_2$	2 mM
KOAc	20 mM
DTT	2 mM
BSA	0.1 mg/ml
rATP	1 mM
rGTP	1 mM
RNA	900 ng
Protein	25 μM

The reaction was incubated at 37°C for 30 minutes. After incubation, malachite green reagent (0.03% malachite green oxaolate, 8.3 mM sodium molybdate, 0.7 M HCl and 0.05% Triton X-100) was diluted with 0.5 volume MilliQ water and 150 μl was added to the reaction. The mix was incubated at room temperature for 20 minutes and transferred to a 96-well plate. The phosphate concentration was

quantified by measuring A_{630}. The standard reaction was carried out by measuring change in colour intensity of phosphate buffer (KH_2PO_4, pH 7) standards at increasing concentrations (0 - 100 nM).

2.20.1   Standard curve

The malachite green assay was used to estimate the NTPase activity of Mth203 and Mth203ΔC53. The protocol was first standardised by measuring the samples containing known concentration of potassium phosphate at various concentrations (0-100nM). 50 μl samples prepared in Mth203 elution buffer (25 mM Tris, 1 M NaCl) were mixed with 2:1 dilute solution of malachite green incubated for 20 minutes and absorbance was measured at 630 nm using a plate reader. A regression line was generated and the equation was used to convert the quantities of inorganic Pi released in NTPase assays to nM concentration.

2.21   Flow cytometry

For flow cytometry analysis, the cell samples were harvested depending on the type of cell culture. For *M. thermautotrophicus* cell culture the staining method was modified from a previous study (Bernander *et al.,* 1998), 100 μl of cell culture (OD 0.3) was added to 900 μl of ice-cold ethanol, mixed thoroughly and stored at 4°C. The fixed cells were centrifuged at 16,000 g for 5 minutes at room temperature and resuspended in 1 ml buffer A (10 mM $Tris_{8,0}$, 10 mM $MgCl_2$). The cells were resuspended and centrifuged again as mentioned above and resuspended in 50 μl of buffer A and 50 μl of 2 × Mithramycin dye (40 μg/ml EtBr and 200 μg/ml Mithramycin in buffer A).

For *M. maripaludis*, 100 μl of cells (OD 0.6) were centrifuged at 16,000 g for 5 minutes. The supernatant was discarded and the pellet was resuspended in 100 μl of TSE buffer (10 mM Tris pH 7.5, 10 mM EDTA, 380 mM NaCl, 200 mM KCl). Cells were fixed by adding 900 μl of ice-cold fixing buffer (600 mM LiCl, 77 % Ethanol) and stored at 4°C. Fixed cells were centrifuged at 16,000 g for 5 minutes at room temperature, the supernatant was discarded and the cell pellet was resuspended in 1 ml buffer A (10 mM $Tris_{8,0}$, 10 mM $MgCl_2$). The cells were centrifuged again and

then resuspended in 75 μl Buffer A and 75 μl 2× Mithramycin A dye. Stained cells were analysed using an Apogee A40 MiniFCM flow cytometer with 50 mW 405 nm laser and the data was analysed using FlowJo software version 8.

2.22 Bioinformatics tools

BLAST analysis of the protein sequences was carried out using tools from http://www.ncbi.nlm.nih.gov and multiple sequence alignments were generated using ClustalX (Thompson *et al.,* 1997). Poorly aligned regions were edited using Gblocks programme (Castresana, 2000). The alignments were used to generate a maximum likelihood phylogenetic tree using Lasergene 9 software and Njplot (Perrière and Gouy, 1996). The bootstrap values for the tree were calculated as follows (Efron, 1982): If there are m sequences of n nucleotides (or codons or amino acids, a phylogenetic tree was constructed on the basis of sequence similarity. Now, from each sequence, n nucleotides were randomly chosen with replacements, giving rise to m rows of n columns each. These now constitute a new set of sequences and tree was then reconstructed with these new sequences using the same tree building method as before. Next the topology of this tree was compared to that of the original tree. Each interior branch of the original tree that was different from the bootstrap tree the sequence it partitions was given a score of 0; all other interior branches are given the value 1. This procedure of resampling the sites and the subsequent tree reconstruction was repeated hundred times, and the percentage of times each interior branch was given a value of 1 is noted and each treenode is labeled with the sum of these values. A branch topology is generally considered correct if the the bootstrap value is 95% or more (Efron, 1982).

Genome context analysis was performed using tools from http://www.ncbi.nlm.nih.gov and Seqbuilder (Lasergene 9) software.

Mth203 protein secondary structure predictions were carried out using the modelling program ModWeb version SVN.r1278 (https://modbase.compbio.ucsf.edu/scgi/modweb.cgi). Mth203 protein structure was threaded on crystal structure of Mj0069 using the program Pymol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). XtalPred was used

to   determine   the   crystallizability   classification   of   Mth203 (http://ffas.burnham.org/XtalPred-cgi/xtal.pl, Slabinski *et al.,* 2007).

2.23   Protein crystallization

Protein crystallization experiments were carried out on Mth203 and Mth203ΔC53 in YSBL, University of York.

2.23.1   Thrombin digestion of His-Tag

In order to decrease the disordered region at the N- terminal end, His-tag was cleaved by thrombin digestion. 2 mg/ml protein was digested with 1 U/μl thrombin (GE Life Sciences). 50 mg protein was incubated at room temperature overnight with dialysis in Gel filtration buffer (25 mM $Tris_{8.0}$, 1 M NaCl). The protein was concentrated using 6 ml spin concentrators (10,000 MWCO, Sartorius) and used for setting up protein crystallization trays (see section 2.23.4).

2.23.2   Reductive methylation of Mth203

The Mth203 protein sequence has 16 lysine residues. In order to carry out methylation of Mth203 lysine residues, reductive methylation was carried out using JBScreen methylation kit and manufacturer's protocol (Jena Bioscience). 20 μl of 1 M di-methyl amine borane complex (ABC) was added to 1 ml of protein in suitable buffer (25 mM NaOAc, 1 M $MgCl_2$). 40 μl of 37 % formaldehyde was added to the reaction, mixed and incubated at 4°C for 2 hour. After incubation, 20 μl of ABC and 40 μl of 37 % formaldehyde were added to the mix and incubated for 2 hours at 4°C. Lastly, 10 μl of ABC was added, mixed gently and tube was incubated for 24 hours at 4°C. 20 μl of 1 M $Tris_{8.0}$ was added and incubated at 4°C for 1 hour. Buffer exchange in gel filtration buffer (25 mM $Tris_{8.0}$, 1 M NaCl) was carried out after the reaction by size exclusion chromatography. The methylation of lysine residues was tested using mass spectrometry.

2.23.3   Dynamic light scattering analysis

Protein solubility was analysed using a sparse matrix approach (Lindwall *et al.,* 2000). Suitable buffers for Mth203 crystallization were tested for particularity by dynamic light scattering (Protein solution, DynaPro). 20 μl of 1 mg/ml protein in a specific

buffer was aliquoted in a quartz cuvette and dynamic light scattering was measured against a buffer control. Polydispersity index was calculated from the data obtained and a suitable buffer for crystallization was selected. For a soluble protein, the polydispersity index should be below 20%.

2.23.4 Setting-up a crystal tray

Mth203 and Mth203ΔC53 proteins were concentrated to 25-50 mg/ml by centrifugation at 8000 g at room temperature for 2-4 hours using spin concentrators (Sartorius). Index, PACT, Hampton 1,2 and CSS 1,2/PEG/Ion/25% PEG Bis-Tris propane pH 6.5 plates (Hampton) were set for the proteins Mth203 and Mth203ΔC53. Crystal trays were set up using automatic robot (Mosquito© TTP Labtech) and stored at 4°C and 37°C. Development of crystals was monitored daily and weekly by observing each well under a light microscope (Leica M2).

3 Bioinformatics and structural analysis of Mth203

3.1 Introduction

The DEAD-box family of RNA helicases is the largest family of SF2 helicases, containing a high degree of conservation in their signature motifs, which can be used to detect and predict new helicases in the sequenced genomes (Rocak and Linder, 2004). In addition to DEAD-box proteins and closely related family of RNA helicases DEAH-box, DEVH-box helicases are together classified as DEXD/H proteins. These proteins are involved in RNA metabolism and share eight conserved motifs with DEAD-box proteins (Walker A and Walker motifs are present) but do not possess the conserved DEAD sequence in motif II or the phenylalanine upstream of the Q motif (Tanner and Linder, 2001, Caruthers and McKay 2002, Rocak and Linder, 2004, Cordin *et al.,* 2006).

The DEAD-box protein structure is composed of two RecA-like globular domains connected by a variable length linker forming a groove for substrate binding (Story and Steitz, 1992). Proteins belonging to the DEAD-box family possess a conserved core domain of 400 amino acids as well as variable amino- and carboxyl-terminal extensions, providing each protein with unique signature sites for the binding of accessory co-factors or proteins involved in various biological process (Korolev *et al.,* 1998). All the SF2 DEAD-box helicase structures solved to date (UvrB, eIF4a, MjDEAD, NS3, UAP56, BstDEAD) are close to minimal size constituting of the core conserved domains but lack the long amino- and carboxyl- terminal extensions (Benz *et al.,* 1999, Kim *et al.,* 1998, Theis *et al.,* 1999, Caruthers *et al.,* 2000, Story *et al.,* 2001, Shi *et al.,* 2004, Carmel *et al.,* 2004, Cordin *et al.,* 2006).

The C- terminus domain of NS3 (a DEXH RNA helicase in HCV) was co-crystallized with a RNA substrate and was shown to be composed of α-helices. The C- terminus interacts with the core domains to provide substrate 3'- single stranded region recognition and couple NTP hydrolysis in the helicase catalytic cycle of the protein (Yao *et al.,* 1997). The N- terminus flanking sequence of NS3 possesses serine protease activity and the C- terminus flanking sequence enhances affinity towards the substrate (Kim *et al.,* 1998). Nevertheless, there is no data available for functional characteristics of the DEAD-box proteins with N- and C- terminus flanking

regions. Since the flanking sequences appear to modify and regulate enzyme activity, the determination of the structure of a DEAD-box protein with its flanking domains and an RNA substrate would clearly help to clarify the complex network of interactions established between the different elements.

In the domain archaea, three DEAD-box RNA helicases have been characterized, MjDEAD (*Methanocaldococcus jannascchii*) (Story *et al.,* 2001), DeaD (*Methanococcoides burtonii*) (Lim *et al.,* 2000) and Tk-DeaD (*Thermococcus kodokaraensis*) (Shimada *et al.,* 2009). These proteins have been suggested as cold-inducible RNA helicases produced during cold-shock response (Story *et al.,* 2001, Lim *et al.,* 2000, Shimada *et al.,* 2009). Out of these proteins, MjDEAD is the only DEAD-box helicase crystallized as a dimer where the two monomers interact by hydrogen bonding of the asymmetric N- terminal (Story *et al.,* 2001). The MjDEAD monomer consists of two α/β domains with Rec-A like topology (Story and Steitz, 1992). The amino terminus contains Walker-motifs responsible for NTP binding and hydrolysis (Walker *et al.,* 1982). The protein sequence lacks long amino- or carboxyl- terminal domains observed in other proteins belonging to the DEAD-box family and thus represents only the common structural core (Korolev *et al.* 1998). MjDEAD shares 36% sequence similarity with eIF4A (yeast DEAD-box RNA helicase and a prototype for DEAD-box family) and the amino- terminal structure of both proteins show good superimposition (rmsd of 1.15 Å), suggesting that the proteins might have very similar structures (Story *et al.,* 2001). eIF4A possesses RNA helicase activity *in vitro* (Rogers *et al.,* 1999), but the full-length protein has not been crystallized yet. Thus, the protein structure of MjDEAD can serve as a model for minimal helicase (Story *et al.,* 2001).

MjDEAD amino- and carboxyl- terminal domains are crystallised in open conformation, suggesting the requirement of large motion to achieve ATP/RNA bound closed state. This conformation is different from the closed-conformation observed in other crystallized DEAD-box helicases (Korolev *et al.,* 1997, Kim *et al.,* 1998, Velankar *et al.,* 1999). The unusual orientation could be caused by crystal packing anomaly or it is possible that the open-complex conformation of protein is stable. It is also possible that the protein requires the presence of other cofactors or substrates to achieve stable helicase closed-conformation.

Mth203 (*M. thermautotrophicus*) is a putative RNA helicase (Smith *et al.,* 1997) and has been shown to interact with MthCdc6-1 (Dr. Richard Parker, PhD thesis, 2006). A *M. maripaludis* homologue, Mmp0457 has also been observed to interact with MmpMCMA (Dr. Alison Walters, PhD thesis, 2010). Ded1 and Dbp9, DEAD-box RNA helicases in *S. cerevisiae* have been shown to interact with DNA (Yang and Jankowsky, 2006) and Dbp9 (*S. cerevisiae*) also demonstrated ATP dependent DNA:DNA and DNA:RNA helicase activity (Kikuma *et al.,* 2004). However, there are no reports of interaction of DEAD-box helicases with DNA replication proteins.

A range of bioinformatics tools has been used to investigate whether Mth203 is an exception of SF2 helicases which interacts with replication proteins, and to gain some insight into how the carboxyl- and amino- terminals provide specificity for certain characteristic function.

3.2 Results

3.2.1 Mth203 homologues are ubiquitous in the three domains of life

M. thermautotrophicus complete genome has been sequenced, and the genome has three putative ATP-dependent DEXD/H RNA helicases, Mth203, Mth656 and Mth492 (Smith *et al.,* 1997). However, the sequence analysis and ClustalX alignment of their homologous proteins revealed that only Mth203 is a DEAD-box helicase with conserved DEAD-box domain in motif 'II', Mth656 possesses a DEIH-box and Mth492 is a DEAH-box RNA helicase (Figure 3.1). In addition, Mth656 and Mth492 primary sequences are longer (roughly 800 amino acids) as compared to that of Mth203 (425 amino acids).

The non-redundant NCBI protein database was searched (BLAST) to identify the most structurally related homologues of Mth203 (Altschul *et al.,* 1990). In order to generate a tree, the phylogenetic analysis was restricted to close homologues in archaea (e value less than 1 e^{-20}) and four bacterial and eukaryotic relatives were selected to root the maximum likelihood tree. 100 bootstrap values were calculated by maximum likelihood method and are represented at the nodes of each branch (section 2.22). A bootstrap value less that 95 is not considered correct suggesting the branch may have a different topology than as represented in the tree.

BLASTp results showed that Mth203 homologues were found in all the three domains of life (Figure 3.2) and one or two DEAD-box protein homologues were present in each organism (for building a decipherable tree only one protein per organisms is shown in (Figure 3.2). The archaeal and bacterial homologues appear to form close clusters compared to their eukaryotic counterparts (Figure 3.2).

The first ten hits from a BLASTp search (Altschul *et al.,* 1990) of the non-redundant sequences using Mth203 primary sequence as a query were predicted SF2 superfamily helicases in archaea *(M. marburgensis, M. smithii, M. stadmanae)* and bacteria (*Halanaerobium sp.*, *C. botulinum, H. orenii, E. halii, A. caccae*, *A. arabaticum*) (Table 3.1). However, no published structural or functional information is available for any of these proteins.

Another BLASTp was performed to investigate the nearest homologues of Mth203 in the domain archaea and maximum likelihood tree was generated using

Njplot (Figure 3.3). No apparent clustering arrangement was noticed in the maximum likelihood phylogenetic tree. In addition, when the two homologous proteins from same organism (*M. mazei* Go1) were examined, no phylogenetic

```
Mth203   ------------------------------------------------------------
Mth656   MRDIMIVLNRRKRSVDFIPAGNPKKILNTRRKPAYWGRLKIRSTEAGPRISRFTVEKGERE
Mth492   ------------------------------------------------------------

Mth203   ------------------------------------------------------------
Mth656   TLRKPSEALKILKKQAVILTGRDPEIEDLLSSYGISYRYARVCQHCLHEGYLTVVSSRSST
Mth492   ------------------------------------------------------------

Mth203   ------------------------------------------------------------
Mth656   VHAGQIICSRCVDELIKRELKFAGMDGSTFRNFRRLIRRGVSLDKILEMMSPRFDPLENHE
Mth492   ------------------------------------------MVKRVLERFRRDWR

                       F                            Q
Mth203   -------------MKGLEFSEFDISGDINRALDDMGFESTTPIQALTLPVTLDG-MDVVGE
Mth656   LTRYDTVTSESERTPRVPLDKLAVPEKFKRMLKREGNTVLRPVQVLAVDAGLLEGEDLMVV
Mth492   FRDAIEHVETIPGREARYSDSTDIPDNIRDYLDAGGIRLYR--HQREALDAIREGMNILIT

           I                         Ia
Mth203   AQTGTGKTAAFAIPVLENLEAER-VPQALIICPTRELCLQVSEEIKRIGKYMKVKVLA-VY
Mth656   SATASGKTLIAELAGIPRALGGE---KFIYLTPLVALANQKYRDFRRRYSPLKLKTAI-KV
Mth492   TPTASGKTLAFNLPIMETLTLNP-DATALYIYPVKALSRDQLGVLEELESELGLKLNPSTY

         GG           Ib                         II
Mth203   GGQSIGN----QIAQLRRGVHVIVATPGRLIDHIERG---TVDLGGISTVVLDEADEMLNM
Mth656   GMSRIRARDELRIPETDVSKADIVVGTYEGMDYILRAG-RSGILGDVGVVVIDEIHTLEDE
Mth492   DGDTPRE----RRPWIRENSRVILTNPHQLHRILPWHHQWRRFYSNLRYVVIDEAHQYRGV

                                          III
Mth203   G------FIDDIERILSHVPERRQTMLFSATVSKPILRIARKYMRNPQVMRVEKKHS---P
Mth656   ERG---SRLKGMIKRIRRLFPDAQIIALSATVKNNLEVASEFGLRLVEYDRRPVP------
Mth492   FGSNVALLLRRLRRICRYYGSEPQFVLASATLANPAEFSRKLTGLDFRVISDDTSPSGPKH

                                       IV
Mth203   KIDEFYFKTREEDKVELLDWILSSNNIRMG-------LIFCNTKRRVQRLRRQLNRMG--
Mth656   LERHLVFSRGEEDKKNLILRLAEREFSTESEKGFRGQTIVFTNSRRKTRLIADYLTRRG--
Mth492   FILYNPYRKRGDPSAHLETSAILSYLVLRG-----VQTLCFTISRKMAELVTRWTREQLDR

                    QxxR                     V
Mth203   ------YSADEIHGDLSQSKRERVMERFRRGDFSLLVATDVAARGIHVPDVEAVVNYDLPF
Mth656   ------VRAAAYHAGLSYRERQRIERAFASQELAAIVTTAALAAGVDFPASQVVFETLLMG
Mth492   ENRKLVDRVTSYRAGYLAEQRRRIEADLKEGRLLGVTATNALELGIDIGSLDAVIISGYPG

                    VI
Mth203   E----NEYYVHRIGRTGRAGSSGKSFTLVVGSEVHRLRRIQSFTGKRIKQSNMPSPEEIRR
Mth656   NRWLMPNEFAQMLGRAGRPSYHDRGVVYVLAEVGMEFDGESEEAMALKLLESGPDPVDVNY
Mth492   T----MISTWQQAGRAGRNLKDSMVILVAFENPLDQYIMKNPSFIFERPHENAIIDLQNRL

Mth203   GYEMDLREILRRNLESKTYSDSEILESLAGEGYSFRDISHALLDVLESSK-----------
Mth656   TEEDVLENILADITSGAIKSESEISPDPSWPLDPEGALDILESHGMIVRDGGLRATGYGVA
Mth492   ILRNQTACAASELPLILDDFLEYDFSVGVAGEMLADGELELAPDGLTCSGDPQFRYGLDDV

Mth203   ------------------------------------------------------------
Mth656   VSKSFIGVTDAEYIRGHLKGASRRPLDIALELEPFEGAYLSGRLHRALSRAAGARFSMNLM
Mth492   SQDTFTVICEGRVLETMSRSQAYREAHEGAVLMNHGNTYTVRDFDLEGRRIIVDRRDVEYH

Mth203   ------------------------------------------------------------
Mth656   ADSTLDILSDGNLVKLDSKLQEAVLNLQMDFLSCECRDRPFCGCIQRRLSEHIIAERISGR
Mth492   TSVLRDTELRIIRKLRQRRLGDLEVNFGEVEVTEHYTGYRVMSYSRVLGKRELDLPPLRFR

Mth203   ------------------------------------------------------------
Mth656   DPVEISGLLSRYQIQAYPGDIFSWLDTTVRALESIGRIASAFKRRRYLKECSGIVRAIEKG
Mth492   RALWFTIPDSLRDEVEEMYGGDAFHGGIHGAEDALISLFPLHVLCDRSDIGGISTPMHPDT

Mth203   ------------------------------------------------------------
Mth656   RGA---------------------------------------------------------
Mth492   QEATIFIYDGFEGGIGLAEKGVEVFGDLLASTLELVSSCSCTEGCPSCIYSPKCGNENSPL

Mth203   ---------------------------------------
Mth656   ---------------------------------------
Mth492   HRDATVKILEYLMEQTSGEVREASEVDGDVLAGVEELCSEGNLLDAKLRLHEVLERDPGNP

Mth203   ---------------------------
Mth656   ---------------------------
Mth492   GACYLMGEILRRQGEVEMAAYFHERAMNGS
```

Figure 3.1. M. thermautotrophicus has three putative DEAD/H box RNA helicases Mth203, Mth656 and Mth492 (Smith *et al.,* 1997). The alignment was constructed using Clustal-X2 program (Thompson et. al., 1997). Mth203 is a DEAD-box protein as DEAD sequence is present in motif II (green), however Mth492 has DEAH sequence

(turquoise) and Mth656 has DEIH sequence (yellow) in motif II. Mth203 also contains all the conserved motifs characteristic of DEAD-box family: F, Q, I, Ia, GG, Ib, II, III, IV, QxxR, V, VI (green).

Figure 3.2. Mth203 homologues in the three domains of life. The maximum likelihood phylogenetic tree was drawn using full-length Mth203 protein sequence alignment with protein homologues in the three domains of life. 100 bootstraps were performed for each tree, and the percentage likelihood values are shown at each tree node. The eukaryotic (green) and bacterial (red) out-groups are highlighted.

Table 3.1: BLAST results for the closest homologues of Mth203

Locus tag	Organism	Annotation	E value
MTBMA_c06520	*Methanothermobacter marburgensis st. Marburg*	ATP dependent RNA helicase	0.0
Halsa_1215	*Halanaerobium* sp. *sapolanicus*	DEAD/DEAH box helicase domain protein	2e-121
Msm_1498	*Methanobrevibacter smithii* ATCC 35061	Helicase	8e-121
ADO77237	*Halanaerobium praevalens* DSM 2228	DEAD/DEAH box helicase domain protein	2e-117
Hore_05480	*Halothermothrix orenii* H 168	DEAD/DEAH box helicase domain protein	3e-113
EUBHAL_03251	*Eubacterium hallii* DSM 3353	Hypothetical protein EUBAL_03251	3e-113
Msp_1228	*Methanosphaera stadmanae* DSM 3091	Helicase	3e-113
Mmah_1100	*Methanohalophils mahii* DSM 1219	DEAD/DEAH box helicase domain protein	9e-112
ANACAC_03142	*Anaerostipes caccae* DSM 14662	Hypothetical protein	1e-111
CLG_B2065	*Clostridium botulinum* D str.1873	Cold-shock DEAD-box protein a	2e-111
Acear_2000	*Acetohalobium arabaticum* DSM 5501	DEAD/DEAH box helicase domain protein	3e-111
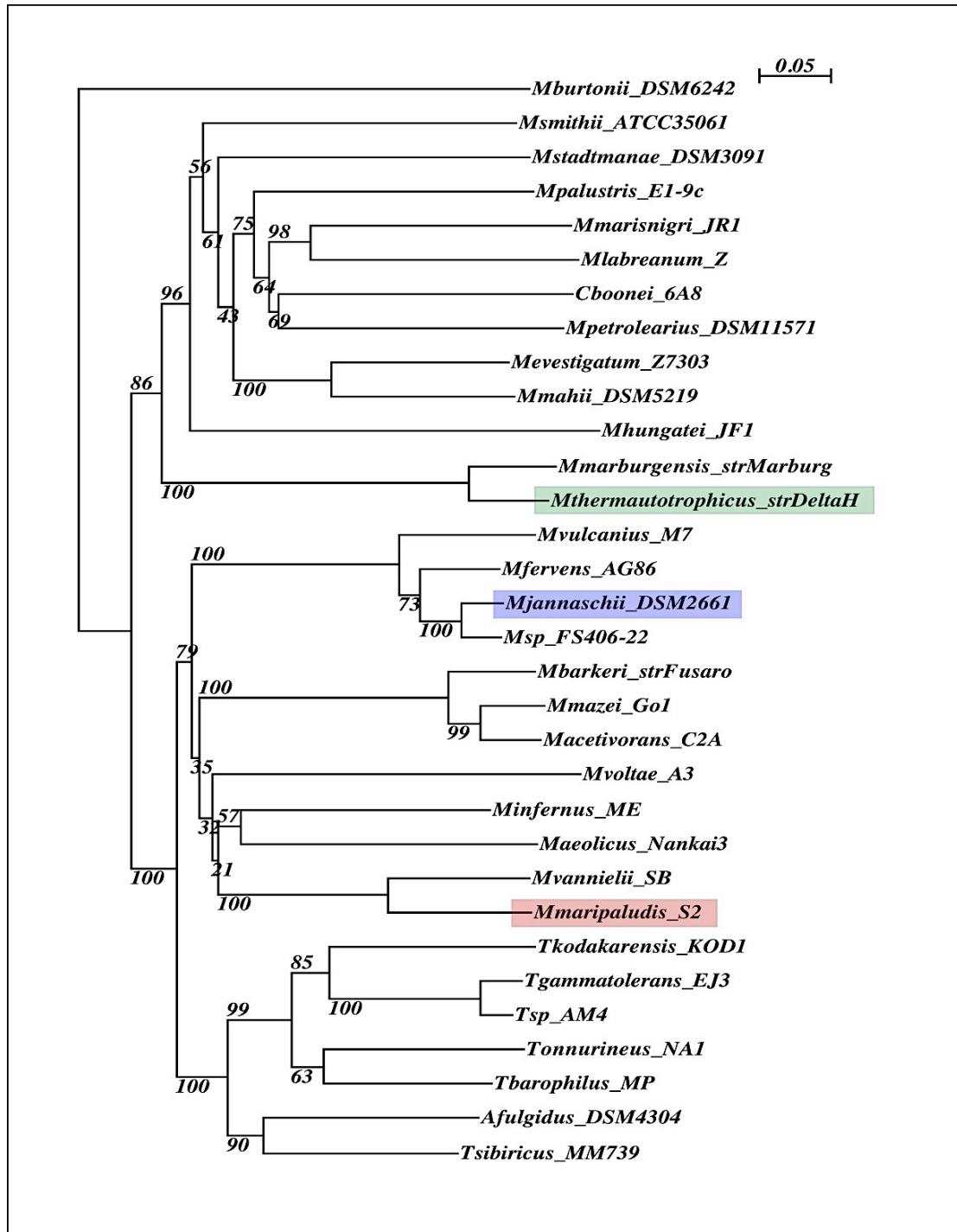
Figure 3.3. Mth203 homologues in the domain archaea. The maximum likelihood phylogenetic tree was constructed using full-length Mth203 protein sequence alignment with homologues in all the sequenced archaea genomes. 100 bootstraps were performed for each tree, and the percentage values are shown at each tree node. Mth203 (green) protein structure was threaded on the crystal structure of Mj0069 (blue) and Mmp0457 (pink), the homologue in *M. maripaludis*, is used for knock-out studies.

clustering was observed (not shown). DEAD-box RNA helicase protein homologues were present in all the archaea except, some methanogens, *Aeropyrum pernix* and *Pyrococcus* sp. (summarized in Table 3.2). With the exception of *M. ruminantium* all other archaea lacking DEAD-box helicases are thermophiles. It has been suggested that archaea harbouring DEAD-box RNA helicases generally possess lower growth temperature limit than those that do not harbour DEAD-box RNA helicase, with the exception of *Sulfolobus tokodaii* (Shomada *et al.,* 2009).

A search for the presence of homologous protein sequences in these organisms revealed that they possess DEXH-box RNA helicases, which are closely related to DEAD-box helicases (Table 3.2).

Table 3.2. DEXD/H-box proteins in archaea lacking in DEAD-box proteins

Organism	Locus tag	DEXD/H motif
Aeropyrum pernix	NP_148447.2	DEIH
	NP_148413.2	DEIH
Methanosaeta thermophila	YP_843209.1	DEIH
	YP_003423853.1	DEVH
Methanopyrus kandleri AV19	NP_614118.1	DEVH
Methanobrevibacter ruminantium M1	YP_003423853.1	DEIH
	YP_003423343.1	DEAH
Methanothermus fervidus DSM 2088	YP_00403595.1	DEIH
Pyrococcus furiosus DSM 3688	NP_577782.1	DEIH
Pyrococcus abyssi	NP_126656.1	DELH
Pyrococcus horikoshi	NP_143216.1	DELH

3.2.2 Sequence alignment of Mth203 with Mmp0457 and MjDEAD

A multiple sequence alignment of putative RNA helicases Mth203, Mmp0457 (Mth203 homologue in *M. maripaludis*) and MjDEAD (Mth203 homologue in *M. jannaschii*) was carried out using ClustalX (Figure 3.4) (Thompson *et al.,* 1997). The proteins possess all twelve motifs characteristic of the DEAD-box family of SF2 helicases (Tanner and Linder, 2001, Rocak and Linder, 2004, Hilbert *et al.,* 2009). The MjDEAD primary sequence consists of only the core conserved region containing all the conserved motifs, whereas Mth203 and Mmp0457 possess additional flanking sequences at the carboxyl- terminal end. Mth203 has a 53 amino acid long carboxyl-terminus flanking sequence and Mmp0457 has 182 amino acids (Figure 3.4).

Concomitantly, there were a few notable substitutions in the conserved motifs of Mth203 and MjDEAD primary sequences (summarised in Table 3.3). In Mth203, the 'Ia' motif non-polar alanine was substituted with a polar uncharged cysteine, however this is a conservative change as both alanine and cysteine are small amino acids. In Motif 'V', the aspartate (negatively charged residue) is replaced by histidine residue (positively charged residue). Since, the Ia and V motifs are involved in substrate binding through the sugar phosphate backbone of nucleic acids (Cordin *et al.,* 2006, Tanner and Linder, 2001), the presence of polar and charged amino acids may contribute towards a change in substrate specificity and binding characteristics.

MjDEAD sequence shows replacement of an isoleucine residue in motif 'IV' with a valine, and alanine in motif 'V' is replaced by serine. Both these substitutions may not have any effect on the over all structure of the domains as they are small residues.

On the other hand, no substitutions in the consensus sequences of the conserved motifs were found in Mmp0457 (Table 3.3).

Table 3.3. Conserved motifs present in Mth203, MjDEAD and Mmp0457. Standard single letter code of conserved motifs: x = any; o = S, T; a = F, W, Y (Cordin *et al.,* 2006, Hilbert *et al.,* 2009). The substitutions in the conserved motifs are highlighted in bold and underlined.

Motifs	Consensus	MjDEAD	Mth203	Mmp0457
F	F	F	F	F
I	AxxGxGKT	ARTGSGKT	AQTGTGKT	AQTGTGKT
Ia	PTRELA	PTRELA	PTREL**C**	PTRELA
GG	GG	GG	GG	GG
Ib	TPGR	TPGR	TPGR	TPGR
II	DEAD	DEAD	DEAD	DEAD
III	SAT	SAT	SAT	SAT
IV	LIF	L**V**F	LIF	LIF
QxxR	QxxR	QSQR	QSKR	QAQR
V	ARGID	**S**RGID	ARGI**H**	ARGID
VI	HRxGRxGR	HRIGRTGR	HRIGRTGR	HRIGRTGR
Q	GaxxPoxxQ	GFEKPTDIQ	GFESTTPIQ	GFTNPTPIQ

```
                   F                          Q                              I
Mmp0457    ----MESFKNLGLSDEILEALEKKGFTNPTPIQEQAIPILIEGKRDIVGQAQTGTGKTAA
Mj0669     MEVEYMNFNELNLSDNILNAIRNKGFEKPTDIQMKVIPLFLNDEYNIVAQARTGSGKTAS
Mth203     --MKGLEFSEFDISGDINRALDDMGFESTTPIQALTLPVTLDG-MDVVGEAQTGTGKTAA

                              Ia                         GG
Mmp0457    FGIPILETIDEHSRNTQALILAPTRELAIQVAEEIDSIKGSKRLNVFPVYGGQSIDRQIR
Mj0669     FAIPLIELVNENN-GIEAIILTPTRELAIQVADEIESLKGNKNLKIAKIYGGKAIYPQIK
Mth203     FAIPVLENL-EAERVPQALIICPTRELCLQVSEEIKRIGKYMKVKVLAVYGGQSIGNQIA

                      Ib                          II
Mmp0457    ELRRGVQIVVGTPGRILDHISRRTIKLENVSYVVLDEADEMLNMGFIDDVEEILKSVSTD
Mj0669     ALKN-ANIVVGTPGRILDHINRGTLNLKNVKYFILDEADEMLNMGFIKDVEKILNACNKD
Mth203     QLRRGVHVIVATPGRLIDHIERGTVDLGGISTVVLDEADEMLNMGFIDDIERILSHVPER

                  III
Mmp0457    KRMLLFSATLPDSIMKLAKNYMREYDIIKVKRQQLTTTLTDQSFYEIHSRDKFELLSRII
Mj0669     KRILLFSATMPREILNLAKKYMGDYSFIKAKIN----ANIEQSYVEVNENERFEALCRLL
Mth203     RQTMLFSATVSKPILRIARKYMRNPQVMRVEKKHS--PKIDEFYFKTREEDKVELLDWIL

                  IV                         QxxR
Mmp0457    DTEKEFYGLIFCKTKADVDEVANRLNEKGYAAEGLHGDMTQAQREKTLDKFKGKKINVLV
Mj0669     K-NKEFYGLVFCKTKRDTKELASMLRDIGFKAGAIHGDLSQSQREKVIRLFKQKKIRILI
Mth203     SSNNIRMGLIFCNTKRRVQRLRRQLNRMGYSADEIHGDLSQSKRERVMERFRRGDFSLLV

                V                         VI
Mmp0457    ATDVAARGIDINDLTHVVNFDIPQNPESYVHRIGRTGRAGKQGYAITFVEPSEFRKFKYI
Mj0669     ATDVMSRGIDVNDLNCVINYHLPQNPESYMHRIGRTGRAGKKGKAISIINRREYKKLRYI
Mth203     ATDVAARGIHVPDVEAVVNYDLPFENEYYVHRIGRTGRAGSSGKSFTLVVGSEVHRLRRI

Mmp0457    QKIAKTEIKREEVPDVKDIISAKKIKIISGIKEVLESGKYADCEKMASDLLEDADPQEVL
Mj0669     ERAMKLKIKKLKFG----------------------------------------------
Mth203     QSFTGKRIKQSNMPSPEEIRRGYEMDLREILRRNLESKTYSDSEILESLAGEGYSFRDIS

Mmp0457    SAVLKYALKDELSESNYKKIGRGSQRSERSGRNDSRGRRSFAPGENVRLFVALGKLDKMN
Mj0669     ------------------------------------------------------------
Mth203     HALLDVLESSK-------------------------------------------------

Mmp0457    PKKLVDHISRKSDVKGRDIDDVKVFEKFSFVTVSSSDAEIILDSFKNERRGRRSIIEVAS
Mj0669     ------------------------------------------------------------
Mth203     ------------------------------------------------------------

Mmp0457    GN--------
Mj0669     ----------
Mth203     ----------
```

Figure 3.4. Mth203, MjDEAD (Mj0669) and Mmp0457 possess all the conserved domains of DEAD-box helicases and Mth203 and Mmp0457 have longer C-termini as compared to MjDEAD. The alignment was constructed in Clustal-X2 (Thompson et. al., 1997). The proteins contain all the conserved motifs (green) of DEAD-box family proteins (F, Q, I, Ia, GG, Ib, II, III, IV, QxxR, V, VI), however Mmp0457 (turquoise) and Mth203 (yellow) protein sequences have additional flanking sequences at the C-terminus.

3.2.3   C- terminus variability region of DEAD-box RNA helicases in *Methanomicrobiales*

Mth203 homologues in the order *Methanomicrobiales* show the presence of variable lengths of amino- and carboxyl- terminal ends as revealed by multiple sequence alignment using ClustalX. It was observed that all the proteins contain a conserved core sequence of roughly 350 amino acids but the length of flanking sequences at the carboxyl- terminus ranges from 8 (*Methanocaldococcus infernus*) to 304 amino acids (*Methanocorpusculum labreanum Z*) (Figure 3.5). The difference in the length of carboxyl- termini is perhaps an indication of a role of this domain in functional specialization of individual proteins to interact with different protein machinery in different species (Barry and Bell, 2006).

Amino- terminal flanking sequences were shorter ranging between 2-30 amino acids with exception of the MbDeaD protein of *Methanococcoides burtonii* DAM 6242, which is a psychrophilic methanogen (Lim *et al.,* 2000). MbDeaD also formed a distinct outgroup in a maximum likelihood phylogenetic tree of homologues in archaea (Figure 3.2).

Organism	N-terminal	Helicase core	C-terminal
Methanothermobacter thermautotrophicus str. Delta H	5	Core 349 residues	71
Methanothermobacter marburgensis str. marburg	5	Core 349 residues	71
Methanosphaera stadtmanae DSM 3091	5	Core 346 residues	232
Methanobrevibacter smithii ATCC 35061	5	Core 315 residues	119
Methanospirillum hungatei JF-1	5	Core 353 residues	173
Methanosphaerula palustris E1-9c	28	Core 353 residues	169
Candidatus Methanoregula boonei 6A8	6	Core 353 residues	173
Methanoculleus marisnigri JR1	6	Core 352 residues	169
Methanocorpusculum labreanum Z	6	Core 346 residues	304 (polar amino acid rich)
Methanoplanus petrolearius DSM 11571	6	Core 353 residues	**177**
Methanocaldococcus infernus ME	2	Core 347 residues	8
Methanocaldococcus jannaschii DSM 2661	7	Core 408 residues	14
Methanocaldococcus fervens AG86	2	Core 346 residues	14
Methanocaldococcus sp. FS406-22	2	Core 346 residues	**14**
Methanocaldococcus vulcanius M7	7	Core 346 residues	51
Methanococcus aeolicus Nankai-3	5	Core 351 residues	173
Methanococcus voltae A3	2	Core 361 residues	221
Methanococcus vannielii SB	3	Core 353 residues	182
Methanococcus maripaludis S2	3	Core 353 residues	182
Methanothermococcus okinawensis IH1	5	Core 350 residues	**177**
Methanosarcina mazei Go1	6	Core 354 residues	228
Methanosarcina mazei Go1	30	Core 349 residues	50
Methanosarcina barkeri str. Fusaro	6	Core 350 residues	220
Methanosarcina acetivorans C2A	6	Core 352 residues	197
Methanohalophilus mahii DSM 5219	5	Core 336 residues	**90**
Methanohalobium evestigatum Z-7303	5	Core 311 residues	**115**
Methanococcoides burtonii DSM 6242	460	Core 349 residues	170 (Serine rich)

Figure 3.5. Mth203 homologues in methanogens have varying lengths of C-termini.

The figure shows diagrammatic representation of helicase core and N and C-terminus extensions of Mth203 homologues in methanogens. The helicase core contains conserved motifs characteristic of SF2 helicases. Mth203 contains 71 amino acids (aa) in its flanking sequences at the C-terminus. The 53 aa long C- terminus fragment of Mth203 was shown to be sufficient for binding to MthCdc6-1.

3.2.4   Genomic context of Mth203

The availability of *M. thermautotrophicus* (Smith *et al.,* 1997), *M. jannaschii* (Bult *et al.,* 1996) and *M. maripaludis* S2 (Hendrickson *et al.,* 2004) complete annotated genome sequence has facilitated the comparison of the genomic context of the gene encoding *mth203, mj0669* and *mmp0457*. *M. thermautotrophicus* belongs to the order *Methanobacteriales* whereas *M. jannaschii* and *M. maripaludis* S2 are in the order *Methanococcales*.

A genome context analysis was carried out using tools at http://www.ncbi.nlm.nih.gov/gene. *mth203, mmp0457* and *mj0669* do not appear to be part of any operon and the genes in their vicinity do not seem to have related functions (Figure 3.6). The gene encoding S8e, a 30S ribosomal protein is present upstream of *mth203* and *mj0669*, perhaps suggesting involvement in ribosomal biogenesis by these proteins. However, the gene encoding S8e protein was found to be present in different orientations in the genomes as compared to the respective DEAD-box helicases. The gene encoding another protein HypE (Hydrogenase expression protein) was also found in the close vicinity of both *mth203* and *mj0669*. The significance of this is unknown and could be coincidental. *mmp0457* does not share any of the genomic features with *mth203* and *mj0669*. Thus the genomic context of the gene encoding *mth203* does not provide any suggestion regarding the possible function of *mth203* and its homologues (*mmp0457* and *mj0669*).
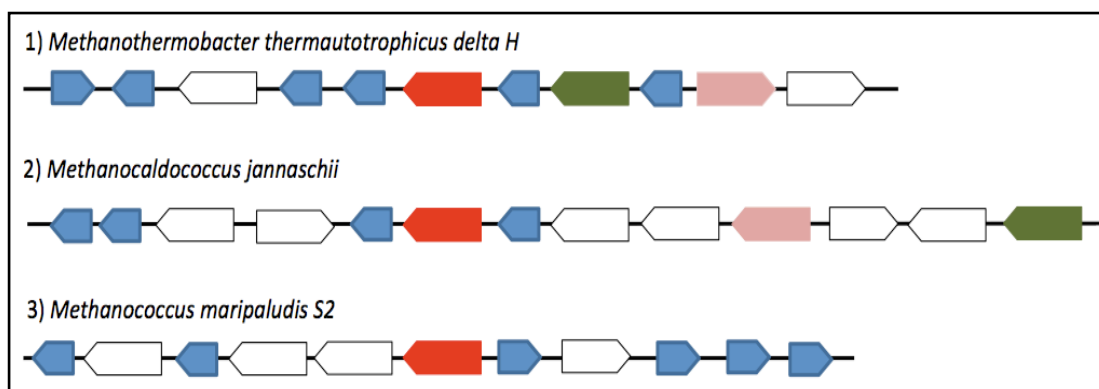
Figure 3.6. Genome context analysis of mth203 (M. thermautotrophicus), mj0669 (M. jannaschii) and mmp0457 (M. maripaludis) (shown in red) suggests absence of conserved sequences of gene clusters. The 30S ribosomal protein s8e (pink) and Hydrogenase expression protein, HypE (green) encoding genes were found in the close vicinities of *mth203* and *mj0669*. Homologous ORFs within each organism are colour coded. Colourless ORFs show no homology to others in the group and hypothetical proteins are depicted as blue arrows.

3.2.5 Expression and purification of Mth203 and Mth203ΔC53

In order to characterize Mth203 and investigate the effect of the carboxyl- terminus on the full-length protein, a deletion mutant, Mth203ΔC53 lacking the carboxyl-terminal sequences was generated. The sequences encoding for Mth203 and Mth203ΔC53 were cloned and expressed in *E.coli* for *in vitro* analysis of the protein structure and function. The proteins were His-tagged and had a predicted molecular weight of 48.9 kDa (Mth203) (Figure 3.7) and 42.7 kDa (Mth203ΔC53) (Figure 3.8). Mth203 and Mth203ΔC53 cultures were harvested and processed (as described in 2.7.3 and 2.7.4). The concentration of purified protein was 25.3 µM (Mth203) and 14 µM (Mth203ΔC53) measured using a Bradford assay with BSA as the protein standard.
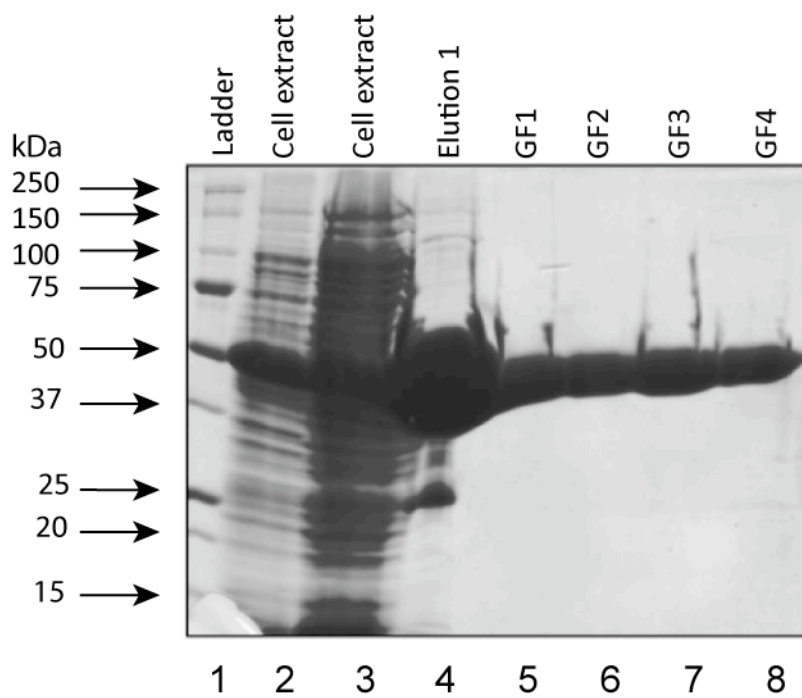
Figure 3.7. Recombinant Mth203 was purified and has a predicted molecular weight of 48.9 kDa. The figure shows 12% SDS-PAGE polyacrylamide gel of Mth203 purification, lane 1 contains protein marker, lane 2 and 3 show total protein in crude cell extracts in different concentrations, lane 4 contains Talon elution fraction, lanes 5-8 correspond to Gel filtration elution fractions (GF 1-4). The elution fractions were pooled and the total protein concentration was measured using a Bradford assay (25.3 µM). Mth203 appears to have predicted molecular weight of 48.9 kDa.
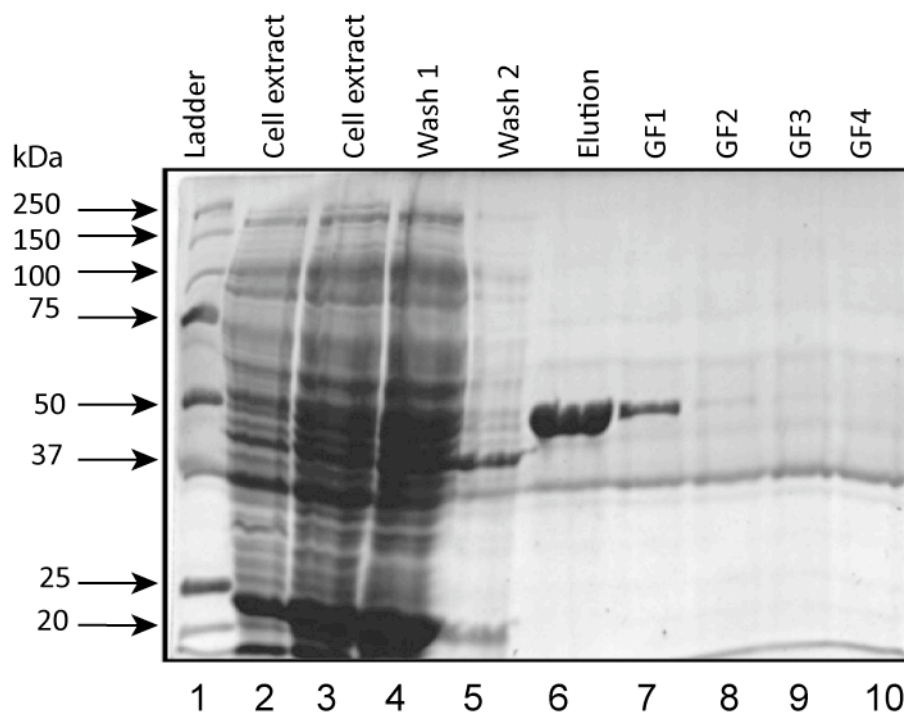
Figure 3.8. Recombinant Mth203 Δ C53 was purified and has a predicted molecular weight of 42.7 kDa. The figure shows 12% SDS-PAGE polyacrylamide gel of Mth203 Δ C53 purification, lane 1 contains protein marker, lanes 2 and 3 show total protein in crude cell extracts, lane 4, 5 show protein washed from the column, lane 6 contains Talon elution fraction, lanes 7-10 correspond to gel filtration fractions (GF 1-4). The elution fraction was selected and the total protein concentration was measured using a Bradford assay (14 μM)..

3.2.6   Secondary structure of Mth203 and Mth203ΔC53

Recombinant His-tagged Mth203 and Mth203ΔC53 were successfully purified from *E.coli*. The secondary structure was predicted by using the program Xtalpred (Slabinski *et al.,* 2007). The program predicted a secondary structure of α-helices (red) connected with short strands (blue) and loops (black) (Figure 3.9a). The sequence also possesses a highly disordered region at the carboxyl- terminal flanking sequence, which was removed to create deletion mutant Mth203ΔC53.

However, in order to perform functional analysis of the respective proteins it was necessary to elucidate if the proteins are folded when overproduced in the bacterial expression system. In addition, to find out whether the deletion of carboxyl- terminal flanking sequences has resulted in any structural changes that might alter the properties of Mth203ΔC53, the secondary structure analysis of Mth203 and Mth203ΔC53 was carried out using CD spectroscopy. The far UV spectra of Mth203 (blue) and Mth203ΔC53 (red) are shown in Figure 3.9b..

The spectral pattern for Mth203ΔC53 was found to be the same as wild type protein Mth203. The minimal value at 208 nm for Mth203 and Mth203ΔC53 is typical of the pattern of α-helical protein structures (Kelly *et al.,* 2005).

a)

```
1...*...10....*...20....*...30....*...40....*...50....*...60....*...70....*...80....*...90....*..100
MKGLEFSEFDISGDINRALDDMGFESTTPIQALTLPVTLDGMDVVGEAQTGTGKTAAFAIPVLENLEAERVPQALIICPTRELCLQVSEEIKRIGKYMKV

....*..110....*..120....*..130....*..140....*..150....*..160....*..170....*..180....*..190....*..200
KVLAVYGGQSIGNQIAQLRRGVHVIVATPGRLIDHIERGTVDLGGISTVVLDEADEMLNMGFIDDIERILSHVPERRQTMLFSATVSKPILRIARKYMRN

....*..210....*..220....*..230....*..240....*..250....*..260....*..270....*..280....*..290....*..300
PQVMRVEKKHSPKIDEFYFKTREEDKVELLDWILSSNNIRMGLIFCNTKRRVQRLRRQLNRMGYSADEIHGDLSQSKRERVMERFRRGDFSLLVATDVAA

....*..310....*..320....*..330....*..340....*..350....*..360....*..370....*..380....*..390....*..400
RGIHVPDVEAVVNYDLPFENEYYVHRIGRTGRAGSSGKSFTLVVGSEVHRLRRIQSFTGKRIKQSNMPSPEEIRRGYEMDLREILRRNLESKTYSDSEIL

....*..410....*..420....*
ESLAGEGYSFRDISHALLDVLESSK
```

b)
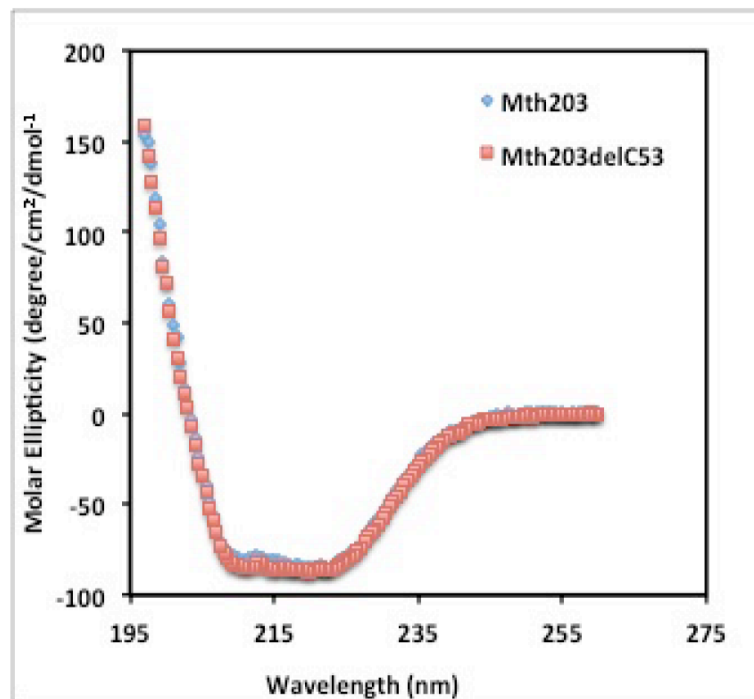
Figure 3.9. Recombinant Mth203 and Mth203ΔC53 show similar folding when expressed in E. coli and the predicted protein structure shows a series of α-helices.
(a) XtalPred program analysis (Slabinski *et al.,* 2007), the protein sequence shows a series of alpha helical secondary structures (red) flanked with strands (blue) and loops (black). The highly structurally disordered regions are underlined. One of the biggest disordered regions is in the carboxyl- terminal flanking sequence of the protein. (b) CD spectra analysis of Mth203 (blue) and Mth203ΔC53 (red), both the proteins show similar CD spectral signatures suggesting that the recombinant proteins are expressed and folded correctly in *E.coli* and have similar secondary structure mainly composed of helices.

3.2.7   Oligomerization of Mth203 and Mth203ΔC53

SEC-MALLS was used to study the oligomerization status of Mth203. The observed molecular weight in the elution fraction was 110 kDa for Mth203 (red) and 45 kDa for Mth203ΔC53 (blue) (Figure 3.10). The predicted molecular weight of Mth203 is 48.9 kDa, this observation suggests that the protein is a dimer in solution. On the other hand, the SEC-MALLS predicted molecular weight and the theoretical weight of Mth203ΔC53 are very similar, suggesting the protein is a monomer. This is a very important observation, which shows for the first time that the removal of the carboxyl- terminal sequences has a marked impact on the structural organization of the protein.

Interestingly, MjDEAD (*M. jannaschii*) also exists as a dimer, in contrast to other closely related DEXD/H box helicases that were crystallized as monomers (Story *et al.,* 2001). The two monomers are held together by hydrogen bonding of the last β-strand of the amino- terminal region forming an interface region. Additionally, there are tightly packed hydrophobic interactions due to the presence of a YSF motif located downstream of motif 'III' (Story *et al.,* 2001). Similar dimerization is observed in the insulin dimer which is held together by two-fold interaction (hydrogen bonding between β sheets and hydrophobic interaction of aromatic residues) of its α/β structures (Adams *et al.,* 1969). A database search for aromatic-X-aromatic residue present downstream of motif 'III' showed such motifs in many DEAD-box helicases such as Ded1 (yeast), An3 (Xenopus), protein 3(Human and mouse), Vasa (Drosophila) (Story *et al.,* 2001).

Mth203 shares homology with MjDEAD in the last β-strand of the amino- terminal region, however, an aromatic-X-aromatic motif is absent in the Mth203 primary sequence. It is possible that in addition to hydrogen bonding of N- terminal domains, Mth203 employs a separate set of interactions unique to the protein for dimerization. Deletion of the carboxyl- terminal leads to monomerization of the protein suggesting that the non-homologous carboxyl- terminal sequences are involved in dimerization of Mth203.
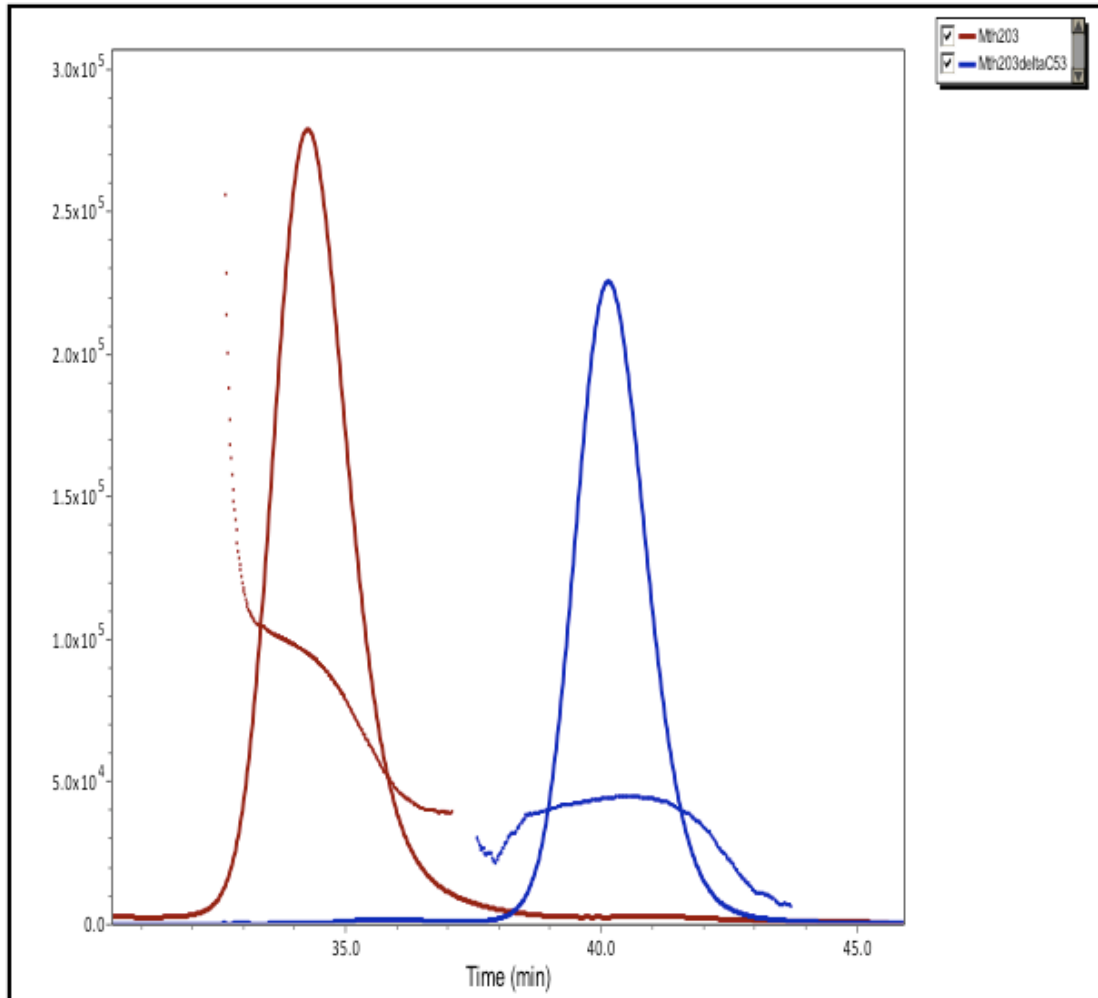
Figure 3.10. Mth203 (red) is a dimer whereas Mth203ΔC53 (blue) is a monomer. The Y-axis on the graph represents molar mass (g/mol) and X-axis shows elution time (min). The peaks demonstrate the protein elution fractions (Mth203 (red), Mth203 Δ C53 (blue)) and the horizontal lines demonstrate the average molecular weight in each sample. Mth203 elutes at size of 110 kDa, suggesting that the protein is a dimer (monomeric molecular weight of 55kDa), whereas Mth203ΔC53 appears to elute at 45 kDa, suggesting a monomeric status of the protein.

3.2.8 Threading of Mth203 protein sequence on MjDEAD crystal structure

Mth203 three-dimensional structure was modelled using the program Modweb (ModWeb version SVN.r1278) and eighteen different structures were predicted on the basis of the presence of conserved motifs in the protein sequence (summarized in Table 3.5). The suggested proteins included DNA gyrase, DNA remodelling protein, ATP-dependent RNA helicases involved in RNA splicing, RNA binding, and proteins involved in DNA recombination (UVRB) and repair (Table 3.5).

The Mth203 primary sequence was threaded onto the nearest archaeal homologue MjDEAD (Mj0669) crystal structure using Pymol software (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC) (Figure 3.11). The proteins have 42% sequence identity and the Mth203 primary sequence was modelled over 360 out of 367 amino acids of MjDEAD. Mth203 has additional 53 amino acids on its carboxyl- terminal end, which could not be modelled, as they are absent in MjDEAD. Thus, a crystal structure of Mth203 may provide information about the location and arrangement of carboxyl- terminal fragment on the Mth203 structure.
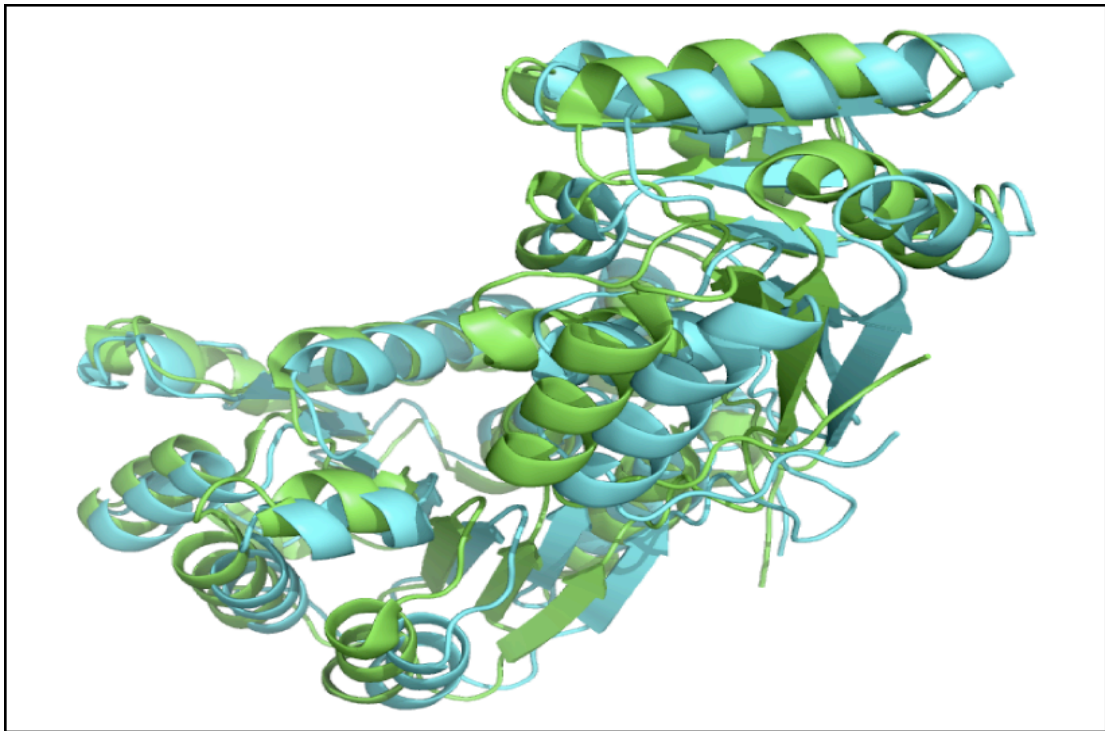
Figure 3.11. Mth203 (turquoise) protein sequence (3-364 aa out of 450 aa) can be threaded on MjDEAD protein crystal structure (green) to visualize the protein structure. The threading shows that the proteins may have a very similar structure. However, as the MjDEAD protein is smaller than Mth203, this structure does not show the structure or position of the 53 amino acids at the C- terminus of Mth203.

Table 3.4. Structural homologues of Mth203

Sr. No.	Protein	PDB id	E Value	Seq ident	Score	Protein length	Target region
1	N- terminal domain of RNA helicase of *Thermus thermophillus* (RNA dependent helicase) (Rudolph *et al.,* 2006)	2gxqA	0	48%	1.00	207	4-208
2	DEAD-box protein from *Methanocaldococcus jannaschii (*ATP-dependent RNA helicase) (Story *et al.,* 2001)	1hv8aA	0	42%	1.00	367	4-364
3	DNA helicase hjm apostatin form 1 of *Pyrococcus furiosus* (holiday junction migration activity) (Oyama *et al.,* 2009)	2zj2A	5e-05	23%	1.00	720	4-409
4	Human DEAD-box RNA helicase ddx3x (ATP dependent RNA helicase) (Hogbom *et al.,* 2007)	2j4lA	0	41%	1.00	417	6-340
5	DEAD-box protein in *Drosophila vasa* (RNA helicase) (Sengoku *et al.,* 2006)	2db3A	0	37%	1.00	434	6-374
6	Human uap56 (ATP dependent mRNA splicing) (Shi *et al.,* 2004)	1xtiA	0	35%	1.00	391	6-343
7	ATP dependent RNA helicase from *Sulpholobus tukodaii* (RNA binding protein) (Nakagawa *et al.,* (in press))	2z0mA	0	42%	1.00	337	15-344
8	SWI1/SNF2 chromatin remodelling domains of Rad54 from *Danio rerio* (DNA remodelling)(Thoma *et al.,* 2005)	1z3iX	0.2	24%	0.95	644	26-397
9	Reverse gyrase from *Archaeoglobus fulgidis* (Rodriguez *et al.,* 2002)	1gkuB	0.34	24%	0.59	1054	48-307

10	*Archaeoglobus fulgidis* xpb (core DNA unwinding protein) (Fan *et al.,* 2006)	2fwrB	0.003	26%	0.98		50-342
11	*Pyrococcus furiosus* hef helicase domain (branched DNA processing) (Nishino *et al.,* 2005)	1wp9A	9.9 e-08	27%	1.00	494	126-332
12	Second domain of *Bacillus subtilis* protein yxin (RNA helicase) (Caruthers *et al.,* 2006)	2hjvA	0	45%	1.00	163	214-364
13	DDX3 human RNA helicase domain (RNA processing) (Rodamilans *et al.,* 2007)	2jgnA	9e-11	42%	1.00	185	223-337
14	*Thermus thermophillus* DEAD-box helicase Hera (RNA metabolism) (Klostermeier *et al.,* 2009)	3eaqA	0	40%	1.00	212	226-405
15	Human DEAD-box RNA helicase DDX41 (ATP dependent RNA helicase) (Schutz *et al.,* 2010)	2p6nA	0	47%	1.00	191	243-334
16	UVRB of *Bacillus subtilis* (DNA dependent ATPase) (Eryilmaz *et al.,* 2006)	2d7sA	0.044	22%	1.00	661	243-372
17	DNA repair helicase Hel308 of *Sulpholobus solfataricus* (DNA recombination) (Richards *et al.,* 2008)	2va8A	0.039	29%	0.86	715	270-420
18	Hel308 of *Archaeoglobus fulgidis* (DNA binding protein) (Buttner *et al.,* 2007)	2pr6rA	0.039	29%	0.60	702	270-382

3.2.9 Crystallization of Mth203 and Mth203ΔC53

XtalPRED and UniProt predictions were performed using the protein sequence of Mth203 and the analysis ranks the protein among the proteins highly unlikely to crystallize, due to the presence of disordered regions in the carboxyl terminus of the protein. Thus, in order to maximise the chances of obtaining protein crystals, all the crystallization trials were set up using both full-length Mth203 protein and carboxyl-terminal deletion mutant Mth203ΔC53.

INDEX (Hampton) and PACT (Newman *et al.,* 2005) crystallization plates were set up to try and obtain crystals of Mth203 and Mth203ΔC53 recombinant proteins with penta-his-tag on N- terminus but no crystals were obtained. As the His-tag might have contributed to the flexibility of the protein and might hinder the crystallization, thrombin digestion of the His-tag was performed. INDEX, PACT, PGA plates were set-up to try and obtain the crystals of Mth203 and Mth203ΔC53 recombinant proteins after removal of penta-his-tag, although it was not possible to obtain crystals.

Assuming that the protein buffer is affecting crystallization, the proteins were tested for the stability in various buffers using a sparse matrix approach (Lindwall *et al.,* 2000). Five buffers providing high solubility of Mth203 were selected and the stability of the proteins was tested in these buffers. Two buffers were selected on the basis of low poly dispersity index (below 20%). The selected buffers had the following composition:

Buffer 1: 100 mM potassium phosphate pH7, 750 mM NaCl

Buffer 2: 100 mM sodium acetate pH 5.5, 1 M $MgSO_4$

Protein crystallization plates were set up for the protein in these buffers, some crystals were obtained but further analysis showed they were salt crystals.

Reductive methylation of lysine residues is known to reduce the disorder of the proteins in solution by reducing flexibility of the secondary structure and thus help crystallization (Walter *et al.,* 2006, Kim *et al.,* 2008). To decrease the disorder or flexibility of the proteins, reductive methylation of the lysine residues of Mth203 and Mth203ΔC53 was carried out. An increase in molecular weight of the proteins was used as a marker for successful methylation of proteins. Further, Index, PACT,

Hampton 1,2 (Hampton) and CSS1, 2/PEG/Ion/25% PEG Bis-Tris propane pH 6.5 plates (Hampton) were set for the methylated Mth203 and Mth203ΔC53.

Several microcrystals were observed in many wells but the crystals were too small to act as nucleation points for crystallization (Figure 3.12). Spherolites (amorphous crystals) were observed in only one of the conditions (30% PEG 4000, 0.1 M Tris HCl pH 8.5, 0.2 M $MgCl_2$). This condition was then selected to optimize crystallization. Various buffers with the sequential variation of buffer concentrations and pH were prepared and an optimization plate was set up using the methylated proteins. Crystals could not be obtained from any of the conditions.
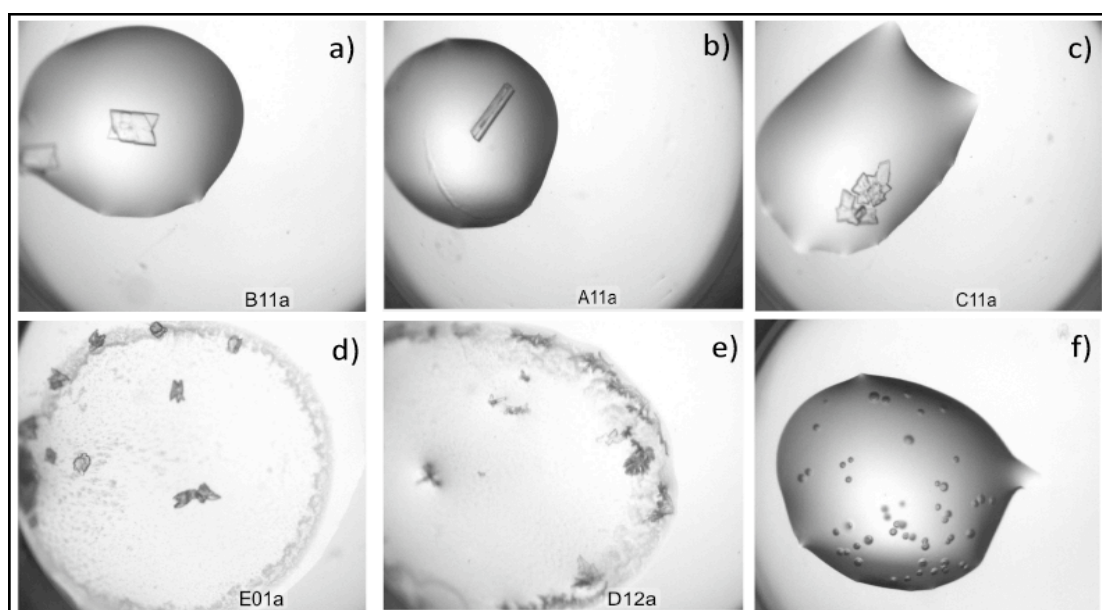
Figure 3.12. Micro-crystals were obtained in INDEX, PACT, Hampton and CSS crystallization plates. (a)-(e) show few examples of salt crystals obtained in the crystallization plates. The sharp edges and two-dimensional nature are typical features of a salt crystal, (f) shows a crystal screen with amorphous crystals called spherolites (30% PEG 4000, 0.1 M Tris pH 8.5, 0.2 M $MgCl_2$).

3.3   Discussion

BLAST searches showed that Mth203 homologues are present in all three domains of life. Strong structural conservation has been observed between the DEXD/H-box proteins along with functional overlap in carrying out RNA metabolic activities (Tanner and Linder, 2001). Thus, perhaps in those archaea, which do not possess DEAD-box proteins, RNA metabolism is carried out by DEXD/H proteins. A multiple sequence alignment of Mth203 with MjDEAD and Mmp0457 shows that the three proteins possess twelve conserved motifs believed to be the signature motifs of DEAD-box helicases. Close analysis of the Mth203 sequence revealed that there are amino acid substitutions in the Ia and V motifs (Table 3.3). Since, Ia and V motifs are involved in substrate binding through sugar phosphate backbone of nucleic acids (Cordin *et al.,* 2006, Tanner and Linder, 2001), the presence of polar and charged amino acids may contribute towards a change in substrate specificity and binding characteristics.

MjDEAD is one of the few DEAD-box helicases containing a short C- terminal region (Story *et al.,* 2001). Mth203 has a carboxyl- terminus flanking sequence, which is 53 amino acids longer than the MjDEAD sequence. It is possible that these sequences are responsible for the interaction with MthCdc6-1. This region also interacts with MthCdc6-1 without the full-length protein (Dr. Richard Parker, thesis, 2006). The DEAD-box helicases have variable amino- and carboxyl- terminus sequences but it is not been possible to determine the structure of these sequences to date. All the DEAD-box proteins crystallized so far possess the minimal core domain structure (Cordin *et al.,* 2006). The flanking sequences appear to confer specificity and affect the activity of these proteins. The Mth203 flanking sequences were found to interact with MthCdc6-1 independent of the rest of the core sequences and required for dimerization (Dr. Richard Parker, thesis, 2006).

The Mth203 protein was found to exist as a dimer in solution, no other DEAD-box helicases have been reported as dimers with an exception of MjDEAD, which was crystalized as a dimer (Story *et al.,* 2001). It was hypothesised that the YSF domain in the β-strand of the amino terminus was responsible for the dimerization of this protein. However, this motif is not present in Mth203, in addition, the

removal of the C- terminal end from the wild-type Mth203 leads to monomerization of the protein. This suggests that even though the two proteins are close homologues, they may have a different structural organization. The structure of the protein could not be determined because it was not possible to crystallize the Mth203 protein.

The putative interaction of the C- terminus with the replication initiation protein MthCdc6-1 and the absence of a homologue with known functional characteristics raise many interesting questions regarding the functional characteristics of Mth203. What is the nature of the MthCdc6-1-Mth203 interactions? Is Mth203 part of a multi-protein complex? And does Mth203 have any regulatory role in DNA replication?

4   Characterization of Mth203 protein

4.1   *Introduction*

MthCdc6-1 is a replication initiation protein in *M. thermautotrophicus* and demonstrates specific DNA binding activity at the origin recognition box (ORB) sequences present at the origin of replication (Capaldi and Berger, 2004, Majernik and Chong, 2008). Concomitantly, Mth203 was shown to interact with full length MthCdc6-1 in a yeast two-hybrid assay and by His-tag affinity co-purification (Dr. Richard Parker, PhD thesis, 2006). Therefore, an important question is whether MthCdc6-1 and Mth203 interact at the origin of replication. If so, then one of the proteins might be involved in recruiting the other in order to carry out some function at the origin. Under such circumstances Mth203 would be expected to interact with DNA not RNA. Thus, elucidation of the interactions between MthCdc6-1, Mth203 and DNA substrates is an important step in understanding the mechanism by which these two proteins interact with each other and perhaps with the origin of DNA replication. So far, there are no reports of any DEAD-box helicases binding with DNA replication proteins. Hence, knowledge regarding the properties and mechanism of Mth203 binding to DNA and MthCdc6-1 is needed to understand the molecular mechanism underlying Mth203 function and perhaps its role in DNA replication. A florescence anisotropy assay (Heyduk at al., 1996) was used to study the protein:protein and DNA:protein interactions using Oregon green labelled protein and DNA substrates.

SF2 helicases are found to work as part of multi-protein complexes (Rocak and Linder, 2004) and their function is influenced by the presence of these interacting proteins. For example, eIF4a, the archetype for DEAD-box helicases, demonstrates an increase in RNA helicase activity in the presence of eIF4B and eIF4F (Jagus *et al.,* 1981, Jaramillo *et al.,* 1991). Hence, in order to characterize the function of Mth203 it will be useful to identify the proteins interacting *in vivo*. Several techniques are available for the identification of protein:protein interactions like phage-based expression cloning, yeast-two hybrid assay, co-immunoprecipitation, chemical cross-linking, affinity co-purification and far western analysis (Phizicky and Fields, 1995, Ausubel, 1987, Sambrook and Russell, 2001). In this study, NHS column mediated

affinity purification and MALDI-TOF were used to identify the proteins interacting with Mth203 *in vivo*.

4.2   *Results*

4.2.1   Anisotropy assay

Fluorescence anisotropy is a technique used to study DNA:protein and protein:protein interactions. The main principle of this technique is that when polarized light of excitation wavelength falls on a fluorochrome labelled species the polarization of emitted light changes according to molecular size (Figure 4.1). A high molecular weight species normally rotates in solution more slowly than a lower molecular weight species. Therefore, on excitation with polarized light, a high molecular weight fluorochrome-labelled species will normally emit light with a greater extent of polarization than a lower-molecular weight fluorochrome-labelled species. This dependence of emitted-light polarization on the size of fluorochrome-labelled macromolecules and macromolecule complexes is used to measure macromolecular interactions (Heyduk *et al.* 1996).

If the fluorophore is immobile within the time scale of the fluorochrome excited-state lifetime (1-100 nsec) then the emission light will exhibit the same polarization. In contrast, a mobile fluorochrome will have partially or fully randomized within the time scale of the fluorescence excited-state lifetime. Thus, in this case the emitted light will display a lesser extent of polarization (Heyduk *et al.,* 1996).
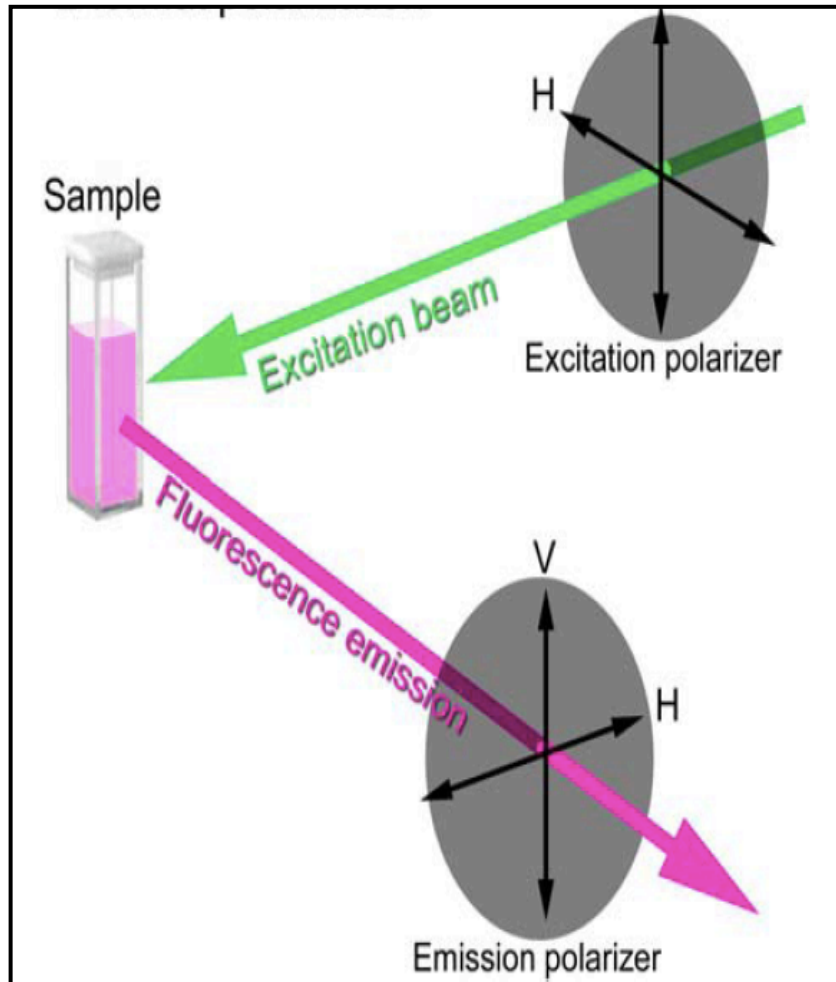
Figure 4.1. Principle of anisotropy assay (Heyduk *et al.,* 1996). When polarised light excites a fluorophore, it changes the polarization of the light depending on the size of the molecule. Protein:protein, DNA:protein interactions can be detected when one of the interacting partners is fused to a fluorophore: upon binding of the partner molecule a larger, more stable complex is formed which will tumble more slowly (thus, increasing the polarization of the emitted light).

4.2.2   DNA substrates used in the anisotropy assay

Steady state fluorescence anisotropy was used to study the DNA:protein interactions between Mth203 and Oregon green labelled DNA substrates. ORB8, a single-ORB sequence from the origin of replication of *M. thermautotrophicus* was used to test the DNA binding activity of Mth203 (Appendix C) (Majernik and Chong, 2008, Capaldi and Berger, 2004). The origin sequence was termed "specific" DNA sequence and a random sequence (with same %GC) from genomic DNA of *M. thermautotrophicus* was used as a non-specific DNA for the DNA binding studies (Majernik and Chong, 2008). The sense and nonsense complementary strands were annealed (section 2.16.1) and 100 nM of 5' Oregon-green labelled DNA substrate was used to carry out the DNA binding assays.

4.2.3   Mth203 has DNA binding activity

The binding studies of Mth203 to the single-ORB sequences in the presence and absence of ATP showed that Mth203 binds to all nucleic acid substrates suggesting inherent non-specific DNA binding activity (Figure 4.2). The K_d (app) for specific and non-specific single-ORB sequences was relatively similar (K_d (app) 0.4-0.48 μM) (Table 4.1).

MthCdc6-1 shows a marked increase in DNA binding (40%) in the presence of longer DNA sequence from the origin of replication (Capaldi and Berger, 2004). In order to test the effect of longer specific and non-specific DNA substrate on Mth203 DNA binding activity, a 205 bp sequence containing ORB 7-10 sequences was selected from *M. thermautotrophicus* origin sequences (Appendix C) (Capaldi and Berger, 2004). Mth203 showed tighter binding to the origin sequence (4ORB_specific) with a K_d (app) of 0.56 μM as compared to a random sequence from the genome of *M. thermautotrophicus* (K_d (app) 1.1 μM). Thus, suggesting that Mth203 protein binds indiscriminately to shorter DNA sequences but shows a slightly higher degree of specificity to longer DNA origin sequences as compared with non-specific DNA.

Table 4.1. K_d (app) values of for Mth203 binding to various DNA substrates

DNA substrate	Mth203 (K_d (app))
Single-ORB specific (ORB8) ds DNA	0.4 μM
Single-ORB specific (ORB8) ds DNA + ATP	0.42 μM
Single-ORB specific (ORB8) ss DNA	0.43 μM
Single-ORB specific (ORB8) ss DNA + ATP	0.42 μM
Single-ORB non-specific (scrambled sequence) ds DNA	0.4 μM
Single-ORB non-specific (scrambled sequence) ds DNA + ATP	0.48 μM
Single-ORB non-specific (scrambled sequence) ss DNA	0.43 μM
Single-ORB non-specific (scrambled sequence) ss DNA + ATP	0.46 μM
4ORB_specific (ORB 7-10)	0.56 μM
4ORB_non-specific (random sequence)	1.1 μM
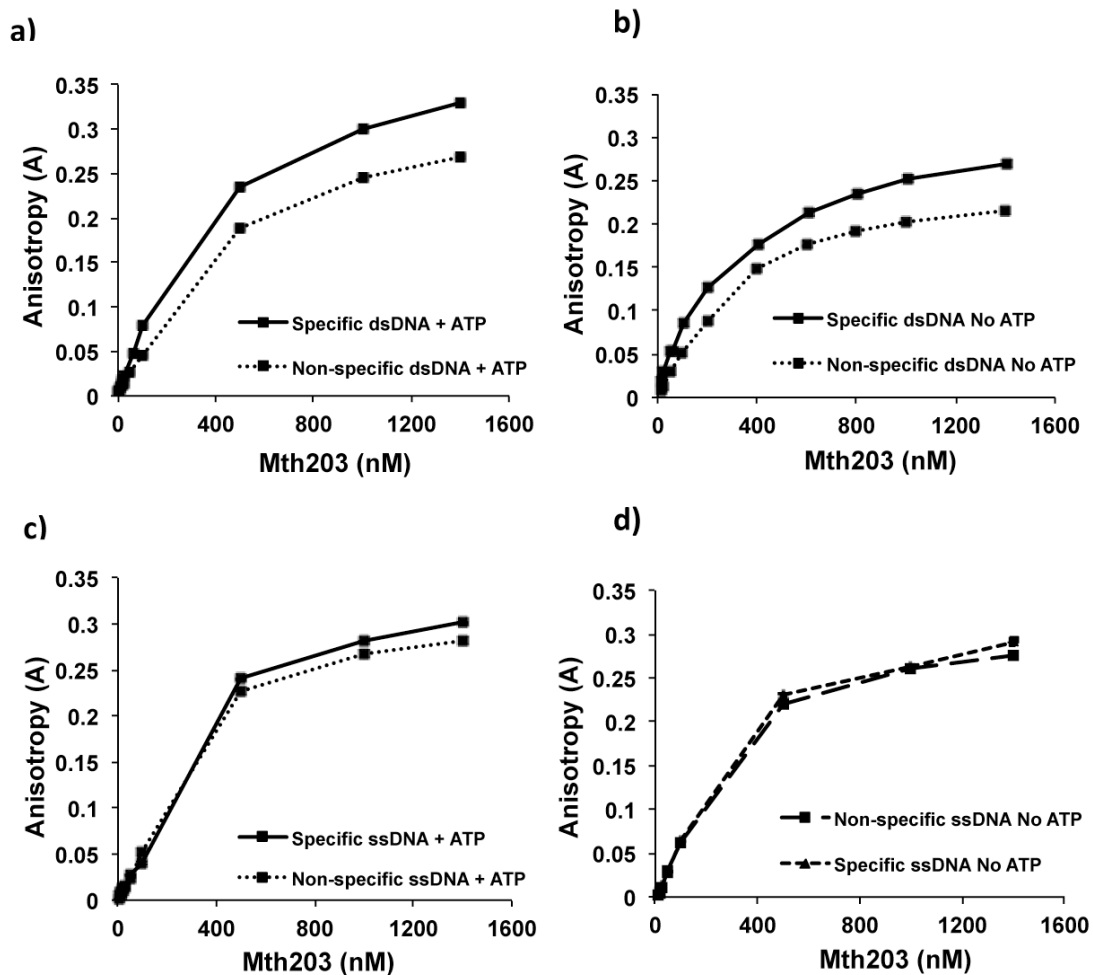
Figure 4.2. Mth203 binds non-specifically to dsDNA and ssDNA short substrates (34 bp). The specific DNA substrate used in the experiment was ORB8 containing origin recognition box (ORB8) present at the *oriC* of *M. thermautotrophicus*. Non-specific DNA substrate was a scrambled ORB8 sequence. The figure shows increase in anisotropy as Mth203 binds to (a) dsDNA substrates (specific and non-specific) in presence of ATP, b) dsDNA substrate (specific and non-specific) in absence of ATP, c) ssDNA substrate (specific and non-specific) in presence of ATP and d) ssDNA substrate (specific and non-specific) in absence of ATP. The standard error is less than 0.001 units from five replicates.
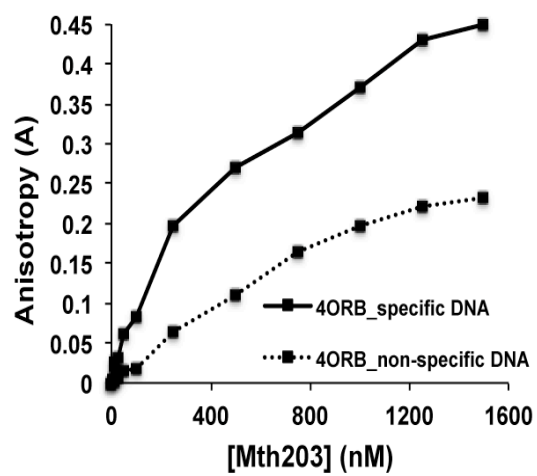
Figure 4.3. Mth203 shows more affinity to the origin sequences (4ORB specific substrate) as compared to non-specific sequences. Mth203 binds strongly to the specific ORB (K_d (app) 0.56) as compared to non-specific sequence (K_d (app) 1.11). The standard error is less than 0.001 units from five replicates.

4.2.4 Regulation of DNA binding activity by the C- terminal peptide

The C- terminus of Mth203 is involved in binding to MthCdc6-1 (Dr. Richard Parker, PhD thesis, 2006) and removal of the C- terminus 53 amino acids changes the oligomerization status of Mth203 (Mth203ΔC53 is a monomer in solution, Chapter 3). Thus, the C- terminus of Mth203 is perhaps involved in regulating the molecular interactions of the protein. In order to further investigate the Mth203 origin-binding activity it was important to investigate the role of the C- terminal sequences in DNA binding activity of Mth203.

The DNA binding anisotropy assay was repeated with Mth203 and Mth203ΔC53 with both single-ORB (ORB8) and multiple-ORB (ORB 7-10) DNA substrates. Analysis of DNA binding showed that Mth203 and Mth203ΔC53 have similar binding affinity for 4ORB DNA (K_d (app) 0.56, 0.716 respectively) (Figure 4.4, Table 4.2). Interestingly, when binding affinities were compared between specific and non-specific DNA, Mth203ΔC53 showed higher affinity for random non-specific DNA (K_d (app) 0.37 μM) whereas Mth203 DNA binding affinity decreased to almost half (K_d (app) 1.11-1.12 μM).

However, for single-ORB binding no significant variation in dissociation constants was observed between Mth203 (K_d (app) 0.4 (specific), 0.38 (non-specific)) and Mth203ΔC53 (K_d (app) 0.42 (specific), 0.37 (non-specific)) (Figure 4.5).

The difference in binding to longer DNA sequences could be explained if the protein core sequences provide the DNA binding activity and the C- terminal sequence provides DNA sequence specificity to the protein.

Table 4.2. K_d (app) values of for Mth203 and Mth203ΔC53 binding to various DNA substrates

DNA substrate	Mth203 (K_d (app))	Mth203ΔC53 (K_d (app))
4ORB_specific (ORB 7-10)	0.567 μM	0.716 μM
4ORB_non-specific (random sequence)	1.11 μM	0.37 μM
4ORB_non-specific2 (random sequence)	1.12	0.35
Single-ORB specific (ORB8)	0.4 μM	0.42 μM
Single-ORB non-specific (jumbled sequence)	0.4 μM	0.37 μM
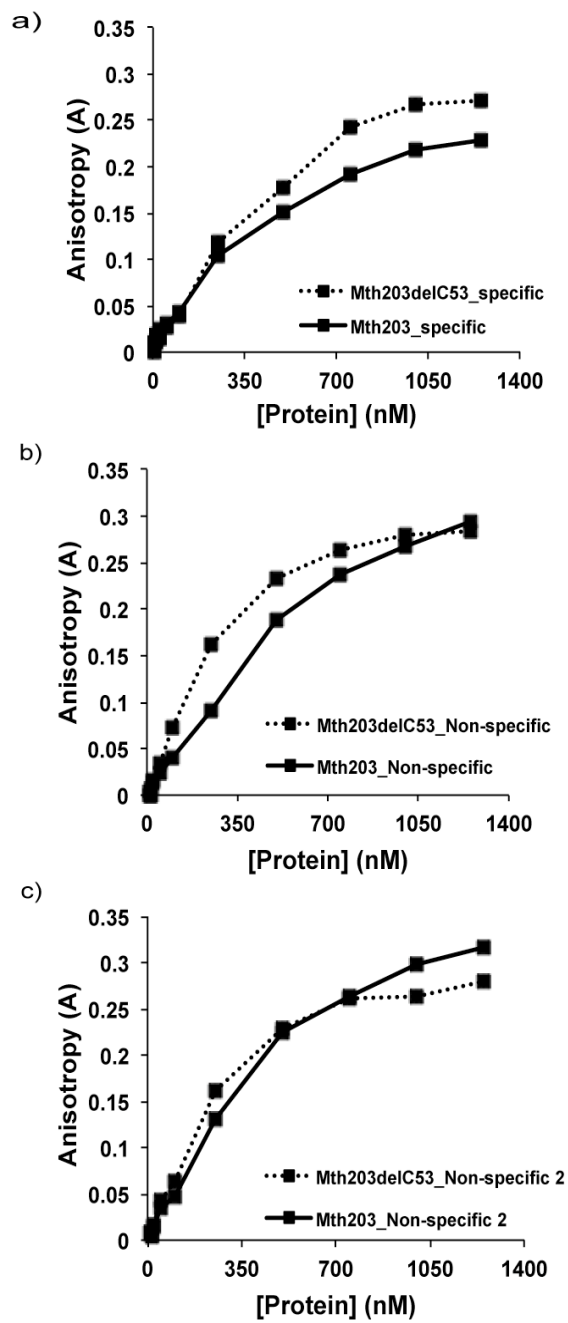
Figure 4.4. Mth203 and Mth203ΔC53 bind to dsDNA specific substrate (205 bp) with similar affinity. a, b and c show binding of Mth203 and Mth203ΔC53 to specific (ORB 7-10), non-specific substrate 1 and 2 respectively. The standard error is less than 0.001 units from five replicates.
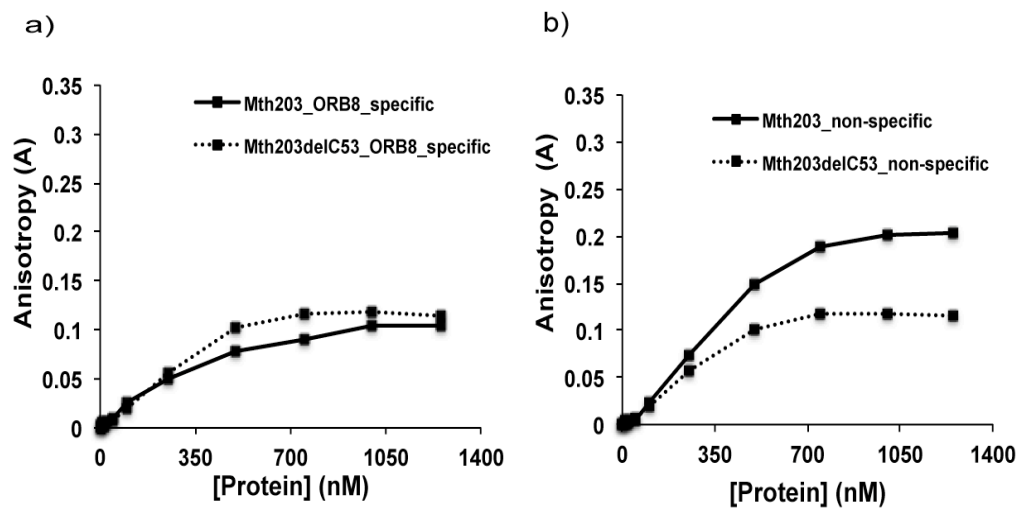
Figure 4.5. Mth203 and Mth203ΔC53 bind to single-ORB specific and non-specific substrates (34 bp) with similar affinity. a and b show binding to specific and non-specific DNA substrate respectively. The standard error is less than 0.001 units from five replicates.

4.2.5   Mth203 and Mth203ΔC53 bind MthCdc6-1 *in vitro*

Previously, Mth203 binding to MthCdc6-1 was confirmed by yeast two-hybrid assay and His-tag protein pull down (Dr. Richard Parker, PhD thesis, 2006). However, the binding in solution was not measured and the apparent dissociation constant K_d (app) of the binding reaction was unknown. For protein:protein binding assays, fluorescence anisotropy was carried out using Oregon green labelled MthCdc6-1.

In agreement with previous results, the binding of MthCdc6-1 to Mth203 and Mth203ΔC53 was observed in solution (Figure 4.6). The change in anisotropy showed similar binding curves with both Mth203 and Mth203ΔC53. As the two proteins differ in the oligomerization status and might also differ in their Cdc6-1 binding mechanism, K_d apparent values were not calculated to compare the binding affinity. However, a comparison of the change in anisotropy curves suggests that the deletion of C- terminus markedly decreased the binding of Mth203 to the replication initiation protein, MthCdc6-1 (Figure 4.6).

MthCdc6-1 is a DNA binding protein and DNA binding experiments on Mth203 also show affinity towards specific origin binding sequences, thus, it is possible that either one of the protein recruits the other onto the DNA or both the proteins compete to bind to origin sequences.
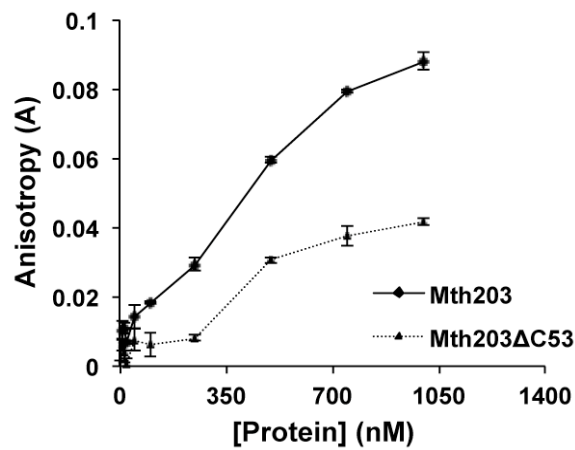
Figure 4.6. Deletion of C- terminal 53 amino acids from Mth203 caused decreased binding of the protein to MthCdc6-1. The graph shows Mth203 and Mth203 Δ C53 binding to Oregon green labelled MthCdc6-1***.*** a change in anisotropy was measured by addition of the proteins in increasing concentrations to 100 nM Oregon green labelled MthCdc6-1. The standard error is less than 0.001 units from five replicates.

4.2.6   Mth203 prevents the DNA mediated inhibition of autophosphorylation of MthCdc6-1

In DNA replication and cell cycle regulation, protein phosphorylation serves as an important regulatory function by activation/deactivation of proteins involved in cell cycle. Cdc6 protein is important in the formation of the pre-initiation complex in eukaryotes and also interacts with other replication proteins like ORC and MCM (Dutta and Bell, 1997, Donovan *et al.,* 1997, Tanaka *et al.,* 1997, Weinreich *et al.,* 1999, Kelly and Brown, 2000). In order to prevent re-initiation of DNA replication MthCdc6-1 is phosphorylated at the initiation of S-phase and then degraded by cyclin-dependent kinases (Calzada *et al.,* 2000). In addition, it was shown that the binding of ATP to Cdc6 causes a conformational change and increases its binding with MCM (Liu *et al.,* 2000, Mizushima *et al.,* 2000). However, the role of ATP in the regulation of the biochemical properties of Cdc6 is unknown and significant ATPase activity has not been detected in any Cdc6 proteins (Weinreich *et al.,* 1999). A study involving autophosphorylation of archaeal and eukaryotic Cdc6 proteins has shown that *M. thermautotrophicus* (MthCdc6-1 and MthCdc6-2), *Pyrococcus aerophilum* (paCdc6) and *Schizosaccharomyces pombe* (spCdc6) show autophosphorylation in the presence of $\gamma\text{-}^{32}\text{P-ATP}$ (Grabowski and Kelman, 2001). Also, the autophosphorylation activity was inhibited in the presence of ssDNA or dsDNA (Grabowski and Kelman, 2001). The authors suggested that perhaps Cdc6 activity is regulated by autophosphorylation for the up-regulation or down-regulation of DNA replication and the reversal of inhibition of the autophosphorylation activity may require the presence of other signal molecules or co-factors (Grabowski and Kelman, 2001).

Mth203 binding to MthCdc6-1 may have a regulatory role. Thus, the effect of Mth203 on the autophosphorylation activity of MthCdc6-1 was measured to investigate if the protein has a regulatory role in MthCdc6-1 activity (see section 2. 17). RPA, a ssDNA binding protein in *M. thermautotrophicus* (Kelman *et al.,* 1999) was used as a DNA binding protein control. The effect of Mth203 on the autophosphorylation of MthCdc6-1 in the presence of both ssDNA and dsDNA was examined (Figure 4.7 and 4.8, respectively). MthCdc6-1 was autophosphorylated in the absence of dsDNA and ssDNA (Lane 3 in Figure 4.7 and 4.8), and the presence of

Mth203 and RPA had no effect on the autophosphorylation activity of the protein. Neither, Mth203 nor RPA had any autophosphorylation activity (Figure 4.8 and 4.9).

In the presence of dsDNA (Figure 4.8), the MthCdc6-1 autophosphorylation was inhibited (Figure 4.8b, lane 9), but the presence of Mth203 in the reaction prevents dsDNA mediated inhibition of MthCdc6-1 autophosphorylation (Figure 4.7b, lane 12). The presence of RPA does not have any effect on the dsDNA mediated inhibition of MthCdc6-1 autophosphorylation. Similarly, in the presence of ssDNA, the autophosphorylation activity of MthCdc6-1 was inhibited (Figure 4.8b, lane 9). However, in the presence of ssDNA binding proteins Mth203 and RPA this inhibition was rescued and autophosphorylation was observed even in the presence of ssDNA (Figure 4.8b, lane 12 and 13). This suggests that the prevention of DNA inhibition of MthCdc6-1 autophosphorylation shown by Mth203 and RPA is probably due to DNA binding activity rather than a direct interaction with MthCdc6-1. Many SF-2 helicases are known to work as part of multi-protein complexes thus, in order to understand the Mth203 function perhaps it will be useful to look into other proteins interacting with the protein *in vivo*.
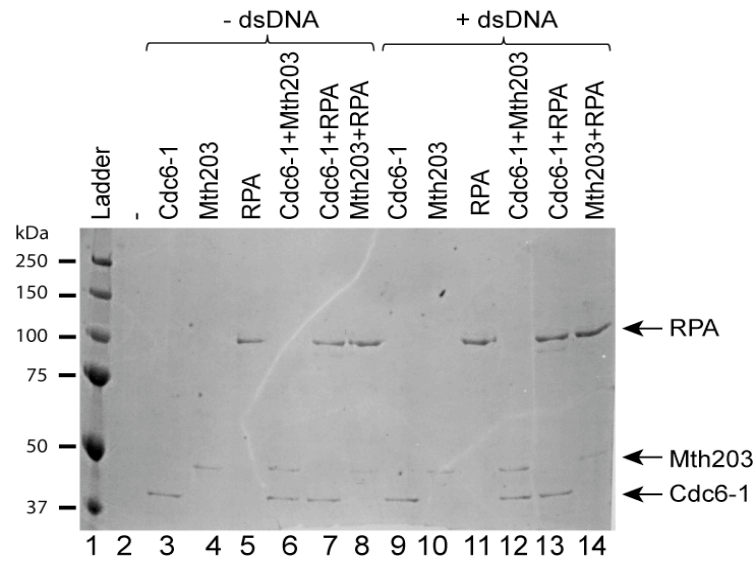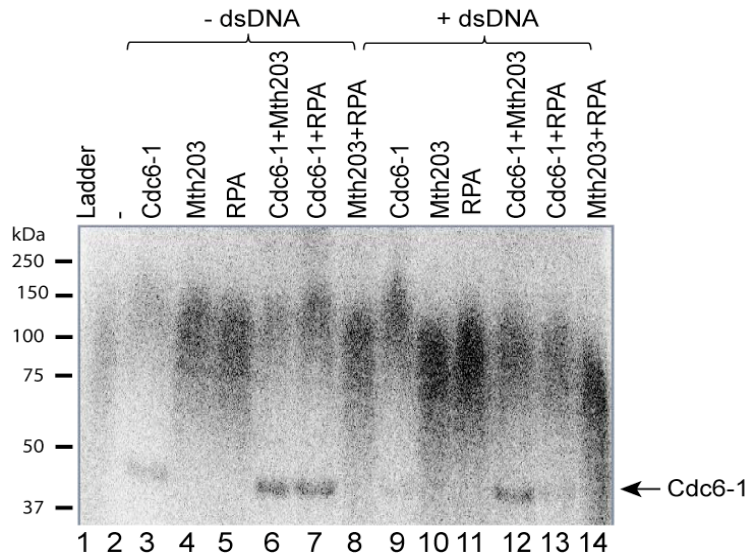
Figure 4.7. Mth203 prevents dsDNA inhibition of MthCdc6-1 autophosphorylation.
(a) Coomassie blue stained protein gel (b) the bands in the figure show autophosphorylation of MthCdc6-1. Lane 3 shows autophosphorylation of MthCdc6-1 in absence of dsDNA, lanes 6, 7 show autophosphorylation of MthCdc6-1 in the presence of Mth203 and RPA, respectively. Lanes 9, 13 show inhibition of the autophosphorylation activity of MthCdc6-1 in the presence of dsDNA and dsDNA+RPA. Lane 12 shows MthCdc6-1 autophosphorylation in the presence of dsDNA and Mth203. A significant amount of radioactive smearing observed in the gel, which decreased when a fresh batch of BSA was used (as observed in Figure 4.9).
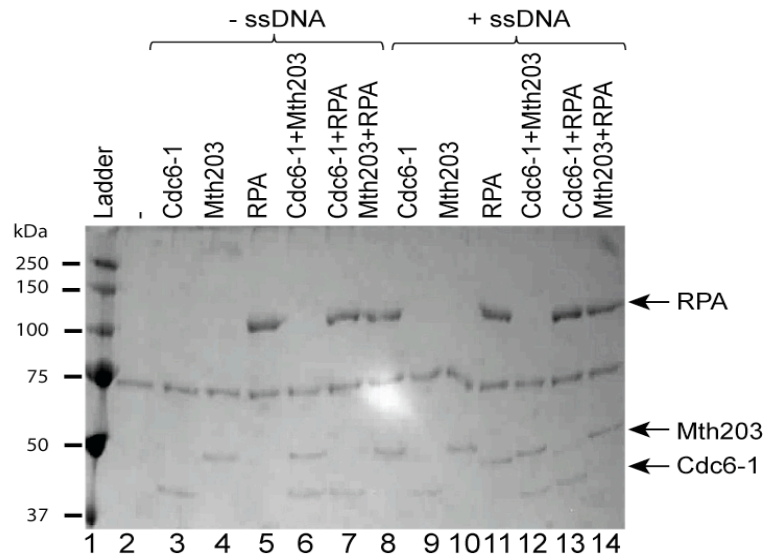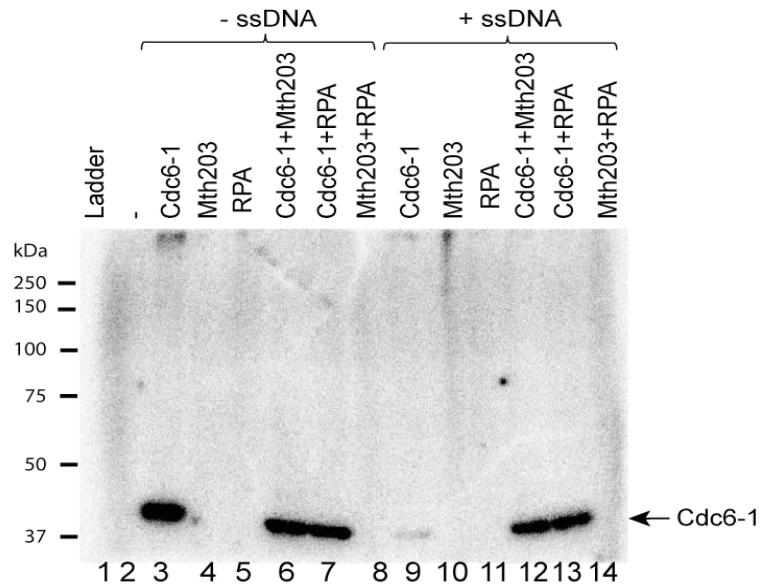
Figure 4.8. Mth203 and RPA prevent ssDNA inhibition of MthCdc6-1 autophosphorylation. (a) Coomassie blue stained protein gel, (b) the bands in the figure show autophosphorylation of MthCdc6-1. Lane 3 shows autophosphorylation of MthCdc6-1 in absence of dsDNA, lanes 6, 7 show autophosphorylation of MthCdc6-1 in the presence of Mth203 and RPA, respectively. Lane 9 shows inhibition of the autophosphorylation activity of MthCdc6-1 in the presence of ssDNA. Lanes 12, 13 show MthCdc6-1 autophosphorylation in the presence of ssDNA+ Mth203 and ssDNA + RPA. Prevention of ssDNA inhibition of MthCdc6-1 autophosphorylation is perhaps due to the ssDNA binding activity of Mth203 and RPA.

154

4.2.7   Identification of proteins interacting with Mth203 by NHS-column mediated pull-down assay

To elucidate the function of Mth203 in *M. thermautotrophicus* and to identify proteins interacting with Mth203 *in vivo*, a pull-down assay was carried out, by covalently linking His-tagged Mth203 to the NHS-column.

4.2.7.1   Cell extract is highly susceptible to auto-degradation

The stability of *M. thermautotrophicus* cell extracts was checked by SDS-PAGE analysis of the cell lysate after 0, 6 h and 24 h of cell harvesting (Figure 4.9). Cdc6-1 was selected as a marker protein and the protein stability was checked by probing the western blots of the cell extracts at 0, 6 and 24 h of harvesting with α-Cdc6-1. At 0 h, two bands (50 kDa and 37 kDa) showed cross-reaction with α-Cdc6-1 suggesting the wild-type protein starts degrading immediately after harvesting. At 6, 24 h, only one band was observed at 37 kDa suggesting no full-length protein was present in the cell extract.

Hence, the best conditions for the pull-down assay were determined to be immediately after cell harvesting, as the proteins in the cell extract start to auto-degrade even when stored at -80°C (as seen as 37kDa degraded product in the western blot, Figure 4.9).
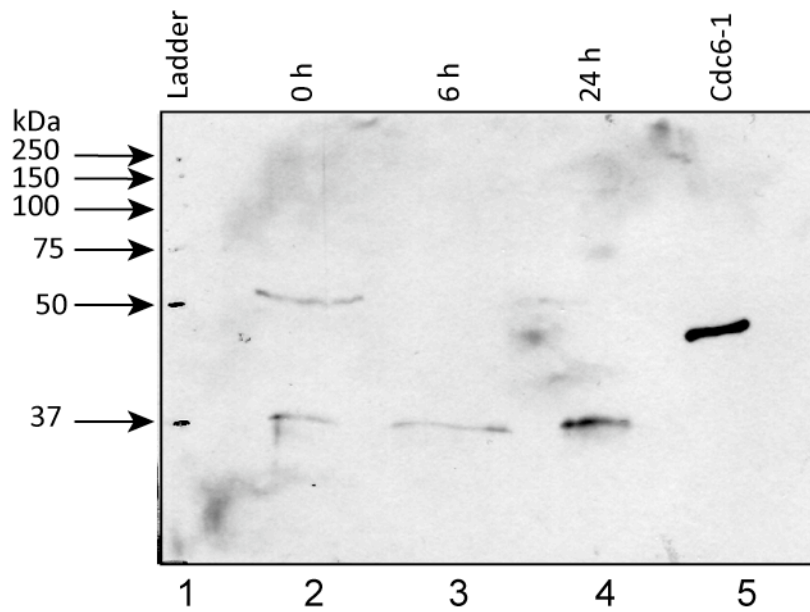
Figure 4.9. The proteins in M. thermautotrophicus cell extract show auto-degradation after 6 h of cell harvesting. Lane 1 contains molecular weight marker, lane 2 contains cell extract after cell lysis at 0 h, lane 3 contains cell extract after cell lysis at 6 h, lane 4 contains cell extract after cell lysis at 24 h, lane 5 contains purified MthCdc6-1 as a standard. Western blot analysis of Cdc6-1 using α-Cdc6-1 shows the protein in the cell extract

4.2.7.2   Preparation of column and pull-down assay

A pull-down assay was carried out on NHS-column covalently bound to Mth203 (see section 2.11 and Figure 4.10).

Schematic representation of pull-down assay

Covalently link Mth203 to NHS column

M. thermautotrophicus whole cell extract was passed over the column to identify Mth203 interacting proteins

Bound proteins were eluted using 1.5 M NaCl and replication initiation proteins were detected by western blot and MALDI-TOF.
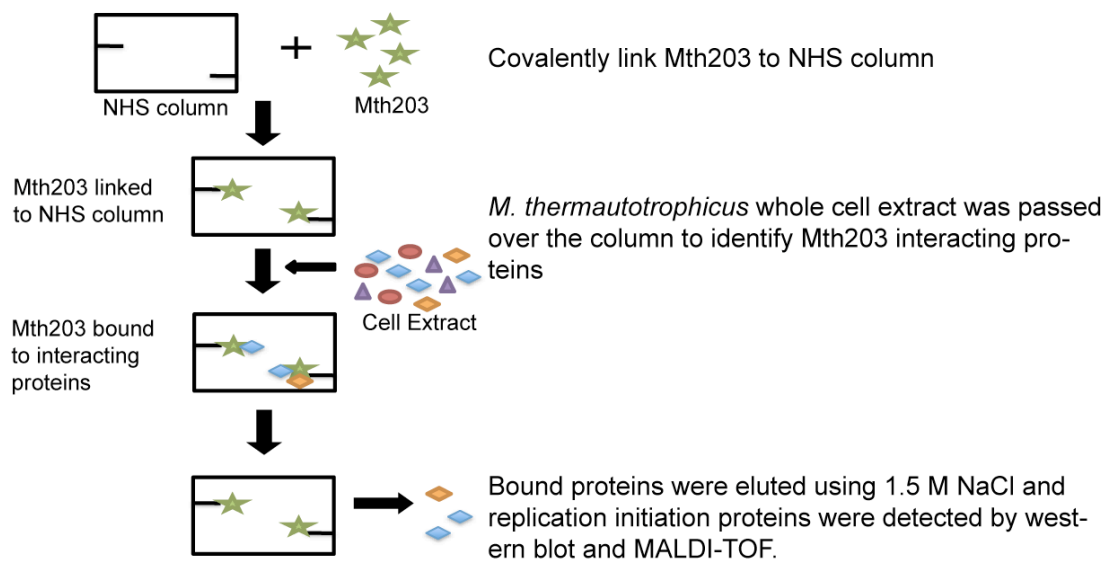
Figure 4.10. Schematic representation of pull-down assay to purify proteins interacting with Mth203 from whole cell extract of M. thermautotrophicus.

4.2.7.3   Western blot and SDS PAGE analysis of elution fractions from pull-down
*   assay*

SDS-PAGE was run for the elution fractions from the pull-down assay (Figure 4.11a). The presence of replication proteins and Mth203 in the elution fractions was tested by western blot using α-MthCdc6-1, α-Mth203 and α-MthMCM antibodies. α-His antibodies were used to detect any protein leaching from the column as His-tagged Mth203 was covalently linked to the column. The western blot showed that MthCdc6-1 was present in the elution fractions further suggesting that Mth203 and MthCdc6-1 interact in vivo (Figure 4.11b). In addition, a large amount of Mth203 was also present in the elution fractions, indicating that the protein is expressed under normal conditions and interacts with other Mth203 *in vivo*. In order to test that this protein was present *in vivo* and not leaching from the column, the blot was probed with α-His antibodies and only the control band was illuminated suggesting the protein eluted was native protein from the cell extract rather them the His-tagged bait (Figure 4.11b). However, another replication protein MthMCM was not found in the elution fractions.

SDS-PAGE analysis showed the presence of other proteins in the elution fraction, which are probably interacting with Mth203. The identification of these proteins might further suggest the role of Mth203 *in vivo*.
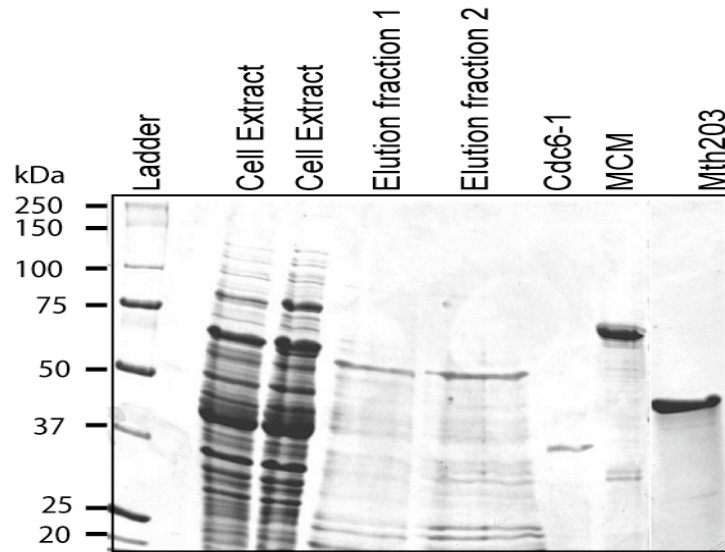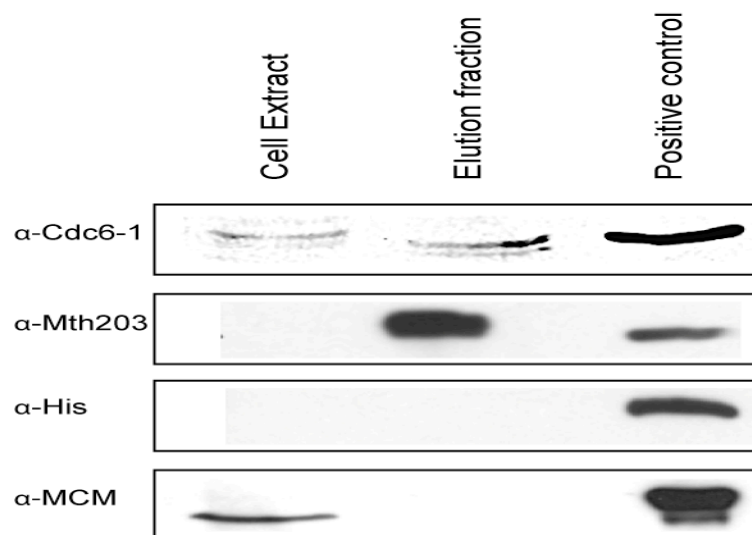
Figure 4.11. Mth203 interacts with MthCdc6-1 and Mth203 in pull-down assays. (a) Coomassie blue stained gel showing the protein profile of the cell extract and elution fractions from the pull-down assay along with positive controls (MthCdc6-1, MthMCM and Mth203), (b) Western blots of pull down assay elution fraction, lane 1 contains cell extract after cell lysis, lane 2 contains elution fraction, lane 3 contains purified recombinant proteins as positive control. The blot shows presence of MthCdc6-1 and Mth203 in the elution fraction. The gel probed with α-His antibody does not show any band in the elution fraction, suggesting no covalently bound protein was washed off the column. MthMCM protein was also absent in the elution fraction.

4.2.7.4   *Identification of other proteins interacting with Mth203 in vivo by*
            MALDI-TOF

The elution fraction was acetone precipitated, and separated by SDS-PAGE (Figure
4.12). Nine specific bands on the gel were excised and processed for identification by
MALDI-TOF-TOF of specific proteins (as described in section 2.13). The analysis
provided matches for the proteins present in the fractions as described in Table 4.3.

Table 4.3. MALDI-TOF identification of the proteins interacting with Mth203 in pull-down assay

Band No.	Protein id
1	No significant match
2	Phosphopyruvate hydratase
3	Mixture of isopentenyl pyrophosphate isomerase and 50S ribosomal protein L3P
4	3-chlorobenzoate-3,4-dioxygenase dihydrogenase related protein
5	50S ribosomal protein L4P
6	30S ribosomal protein S4e
7	Mixture of 30S ribosomal protein S3Ae and 30S ribosomal protein S7P
8	50S ribosomal protein L18P
9	30S ribosomal protein S13P

The presence of replication proteins could not be confirmed by MALDI-TOF. Large
and small subunit ribosomal proteins were detected in the MALDI-TOF-TOF analysis
of elution fractions.
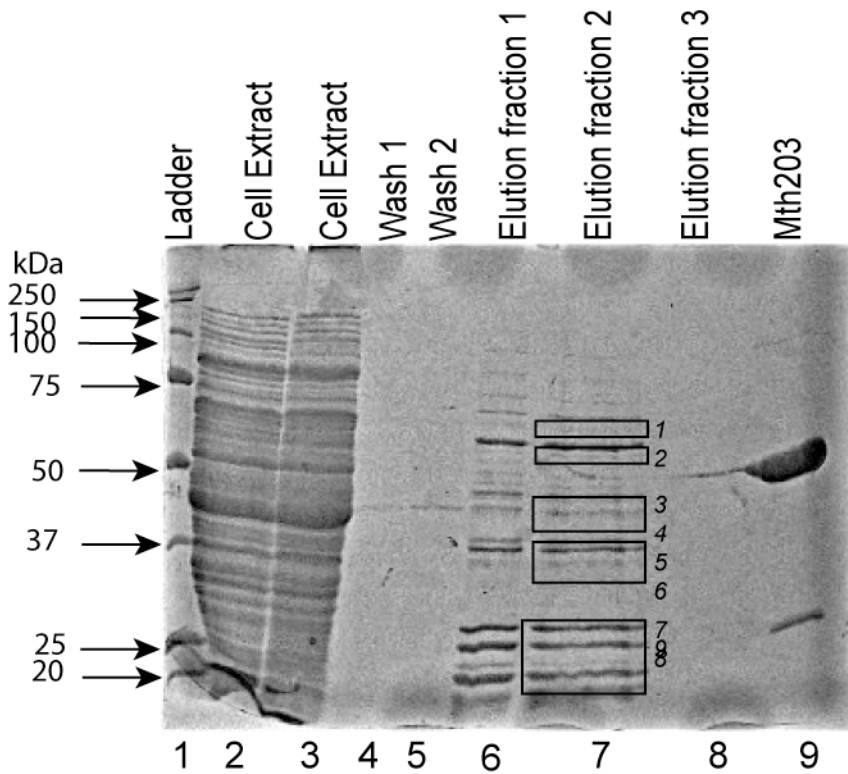
Figure 4.12. The protein bands obtained by NHS column pull down assay of Mth203 were sent for MALDI-TOF-TOF analysis. Lane 1 contains molecular weight marker, lane 2 contains cell extract after cell lysis, lane 3 contains the flow through of cell extract when passed through the column, lanes 4 and 5 contains wash fraction 2 and 3, lanes 6-8 contains elution fractions and lane 9 contains purified Mth203 as a standard. The bands labelled 1-9 were sent for MALDI-TOF-TOF identification.

4.3   Discussion

DNA binding assays of the protein have demonstrated that Mth203 binds DNA with no sequence preference and the binding is not affected by the presence or absence of ATP. This is not unusual as a DEAD-box protein Ded1 in yeast shows DNA binding, which cannot be modulated by ATP (Yang *et al.,* 2006). Deletion of the C-terminal domain caused no change in DNA binding to specific single-ORB sequences. Both Mth203 and Mth203ΔC53 showed a high DNA binding affinity to the specific longer origin sequences but deletion of C- terminal (Mth203ΔC53) lead to an increased affinity for non-specific substrates, thus suggesting a role for the C-terminus sequences of Mth203 in DNA binding and regulation.

The C- terminal domain of Mth203 was shown to bind MthCdc6-1 independent of the full-length protein (Dr. Richard Parker, PhD thesis, 2006). In protein:protein interaction assays, deletion of the C- terminal domain of Mth203 decreased the binding affinity of the mutant (Mth203ΔC53) to MthCdc6-1 compared to the full-length protein (Mth203), suggesting the two proteins may actively interact by C- terminal. If both MthCdc6-1 and Mth203 are involved in DNA replication initiation, the C- terminal domain of Mth203 can be involved in substrate specificity (by interactions with Cdc6-1 and origin DNA) and perhaps be involved in the DNA recognition and binding activity of Mth203.

Thus, a role of Mth203 in origin recognition can be envisaged where the full-length protein will recognise and bind specifically to the origin of replication and will be involved in formation of replication initiation complex. Once the replication is initiated, cleavage of the C- terminal will make the protein non-specific for origin recognition and thus will prevent re-initiation.

The DNA binding activity of Mth203 is consistent with the DNA binding properties of an origin binding protein that binds long origin sequences and then multimerize causing DNA bending and hence forming an open complex at the origin of replication in archaea (cdc6), bacteria (ORC) and eukaryotes (Cdc6 and OBP) and once the initiation is achieved the protein is rendered inactive by substrate competition or specific inactivation mechanisms like RIDA (bacteria) (Messer, 2002, Bell and Dutta, 2002, Mizushima *et al.,* 2011).

Although there are no DEAD-box helicase crystal structures containing the C-terminal domains, it is hypothesised that the helicases that unwind nucleic acids with defined polarity possess a C- terminal domain on top of the nucleotide binding site, thus providing substrate specificity and polarity for dsDNA substrate unwinding (Fairman-Williams *et al.,* 2010). This would explain the increased specificity for dsDNA in the presence of the C- terminal of Mth203.

It is also possible that Mth203 is a DNA binding protein and has a regulatory effect on MthCdc6-1. The autophosphorylation of Cdc6 in eukaryotes is reported to be one of the mechanisms that inactivate the protein and cause its degradation once DNA replication initiation is initiated (Calzada *et al.,* 2000). MthCdc6-1 autophosphorylation is inhibited in the presence of DNA (Grabowski and Kelman, 2001), but when DNA binding proteins Mth203 or RPA are added the inhibition is prevented. RPA is known to inhibit DNA polymerase B in DNA replication of *M. thermautotrophicus* (Kelman *et al.,* 1999). Similarly, MthCdc6-1 phosphorylation may play a positive or negative role in the activation of DNA replication and the autophosphorylation assay suggests DNA binding proteins (Mth203 and RPA) may regulate MthCdc6-1 activity.

MthCdc6-1 and Mth203 binding was observed *in vitro* and western blots probed with α-MthCdc6-1 revealed the presence of MthCdc6-1 in the pull-down assay, which further strengthened the notion that MthCdc6-1 binds with Mth203 *in vivo*. In addition, MALDI-TOF-TOF analysis identified an abundance of ribosomal proteins in the elution fraction, suggesting that the protein may perform more than one function in the cell.

5   Identification of a biochemical function of Mth203

5.1   Introduction

RNA helicase activity is fundamental to numerous cellular processes in unwinding/rearranging RNA secondary structures or disrupting RNA-protein interactions. DEAD-box RNA helicases are known to use ATP to remodel macromolecular interactions between RNA and proteins. Nevertheless, specific substrates and functions are not known for the majority of these proteins (Rocak and Linder, 2004). All the purified DEAD-box helicases have shown NTPase activity, but the RNA helicase activity has been characterized in only few proteins (Iost *et al.,* 1999). Out of 39 DEXD/H-box proteins in *S. cerevisiae* only four (eIF4A, Dbp5p, involved in polyA$^+$ RNA export; Prp16p and Prp22p, RNA splicing proteins) have been shown to possess RNA helicase activity (Iost *et al.,* 1999).

The most extensively studied DEAD-box proteins are Vasa protein, in *Drosophila melanogaster* and eIF4A, eukaryotic translation initiation factor 4A in yeast (Linder and Lasko, 2006, Rozen *et al.,* 1990). Vasa is involved in targeted mRNA translation and unwinds RNA duplexes whereas eIF4A is a non-processive helicase unwinding RNA, DNA/RNA duplexes and requires other co-factors for its activity (Yao *et al.,* 1997, Jaramillo *et al.,* 2010, Rogers *et al.,* 2001). A close homologue NS3, DEXH RNA helicase in HCV virus is a processive helicase unwinding RNA and DNA duplexes (Tai *et al.,* 1996).

Mth203 shows non-specific DNA binding (Chapter 4) and interacts with DNA replication initiation protein MthCdc6-1 in pull-down assay (Chapter 4) and a yeast two hybrid assay (Dr. Richard Parker, Ph.D. thesis, 2006). It is important to characterize its biochemical activities to address the following questions: is Mth203 able to unwind RNA/DNA duplexes in an ATP dependent manner? Is Mth203 an RNA dependent ATPase *in vitro*? What is the expression profile of Mth203 in the cell cycle of *M. thermautotrophicus*?

## 5.2	Results

### 5.2.1	DNA helicase assays to test the DNA unwinding activity of Mth203 and Mth203ΔC53

*5.2.1.1	Mth203 and Mth203ΔC53 do not possess DNA helicase activity*

MthMCM in *M. thermautotrophicus* forms a homo-hexamer and has been shown to possess ATP-dependent 3'-5' helicase activity on a forked substrate (Kelman *et al.,* 1999, Chong *et al.,* 2000). DNA helicase assays were carried out to test the DNA helicase activity of Mth203 and Mth203ΔC53, MthMCM was used as a positive control for substrate unwinding. A forked substrate (Figure 5.1) was used in the strand displacement assay, generated by annealing a ^{32}P labelled 57 bp oligonucleotide (HS2) and a non-labelled 74 bp oligonucleotide (HS1) (section 2.18.1). The annealed substrate contained 25 bp ds-DNA and the non-complimentary regions on HS1 and HS2 formed the fork.

The helicase reaction was carried out in a reaction buffer containing 20 mM HEPES$_{7.5}$, 10 mM MgCl$_2$, 2 mM DTT, 1 mg.ml BSA, 5 mM ATP, 2 nM of labelled substrate and 50 or 100 nM of protein. Helicase activity was measured as the amount of ^{32}P labelled oligonucleotide displaced in 1 hour at 50°C. Reactions were stopped using STOP buffer (section 2.18.2). The products were treated with proteinase K to prevent the presence of secondary high molecular weight bands due to protein:DNA binding. A positive control with 100% unwinding was generated by boiling the labelled substrate at 100°C for 3 minutes and then cooling immediately on ice. Results were analysed by separating the annealed and displaced substrates on a 12% polyacrylamide (1× TBE) gel (as described in section 2.18.2). The band intensities of unwound substrate were detected by phosphorimaging (Quantity One) and column graphs were plotted to further visualize the degree of unwinding.

Neither Mth203 (Figure 5.2) nor Mth203ΔC53 (Figure 5.3) displayed DNA unwinding of the forked 3'-5' DNA helicase substrate, although unwinding of the substrate was observed in the presence of MthMCM as described in previous studies
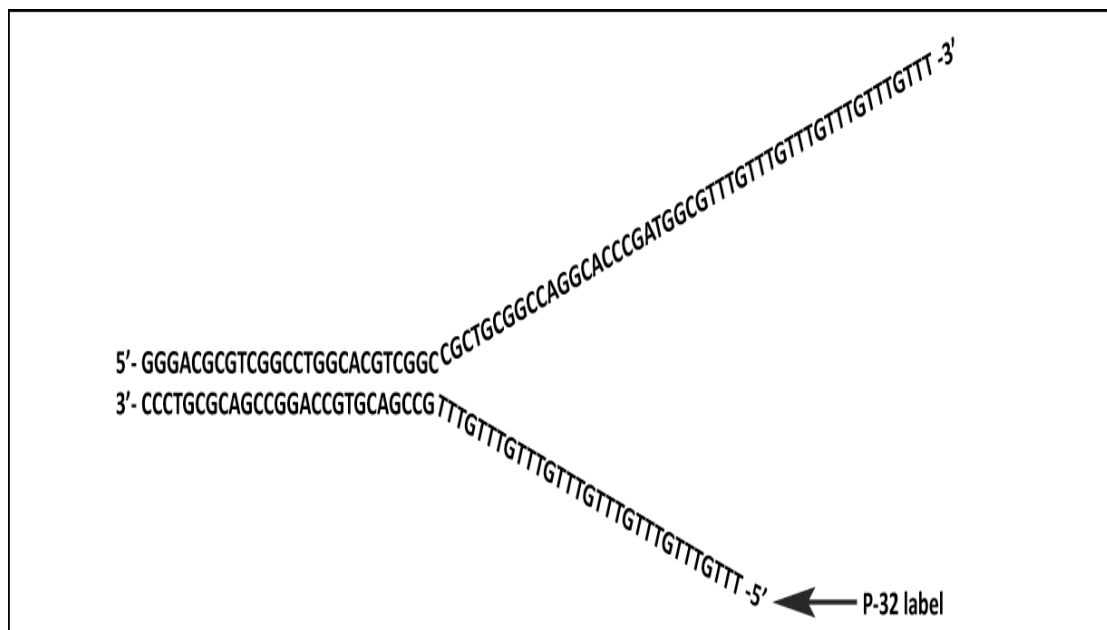
Figure 5.1. The forked substrate used for DNA helicase assay. The substrate has 25 bp ds DNA region, 49 bp long 3' overhang and 32 bp long 5' overhang which was ^{32}P labelled using T4 phosphonucleotide kinase.

a)

MCM (nM)	-			-	50	100	-	-
Mth203 (nM)	-			-	-	-	50	100

ds substrate

ss product

1   2   3   4   5   6
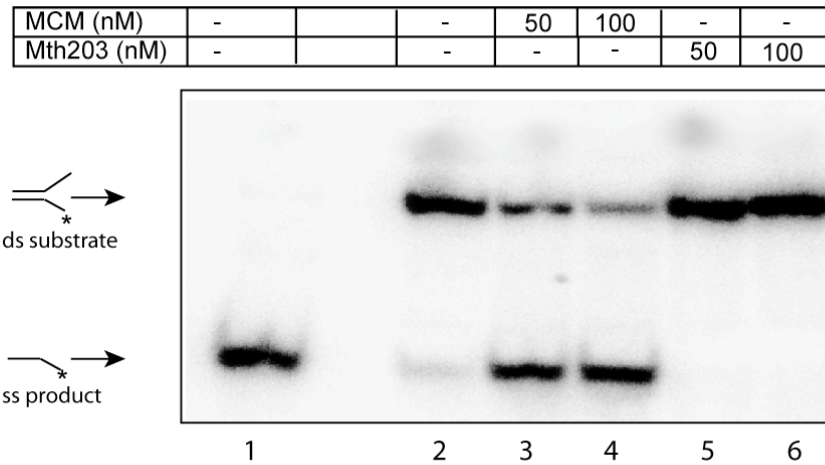
b)
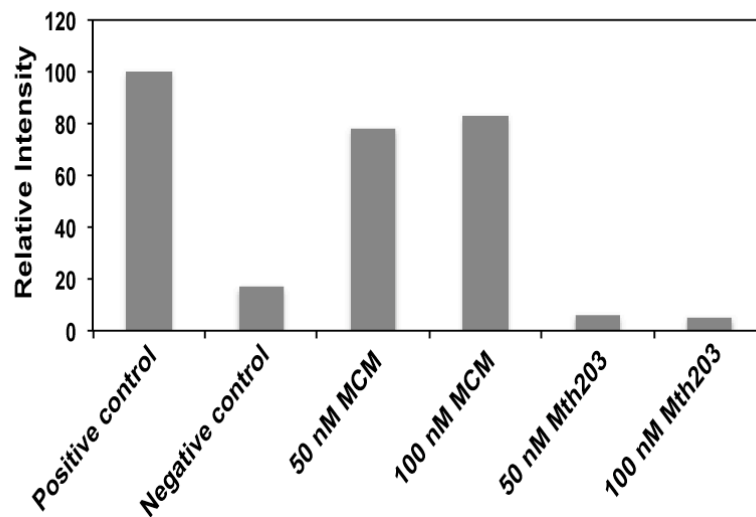
Figure 5.2. Mth203 does not unwind a forked substrate. (a) 11% acrylamide gel showing results of helicase assay of MthMCM and Mth203. Lane 1 contains boiled substrate as positive control, lane 2 is no protein negative control, lane 3,4 contains 50 and 100 nM MthMCM, and lane 5,6 contain 50 and 100 nM Mth203 protein. (b) Histogram showing the band intensities of unwound single stranded substrate under different conditions. DNA unwinding was observed in the presence of MthMCM and no unwinding was seen in the presence of Mth203.

a)

MCM (nM)	-	-	50	-	50	-	-
Mth203 (nM)	-	-	-	50	50	-	-
Mth203ΔC53 (nM)	-	-	-	-	-	50	100
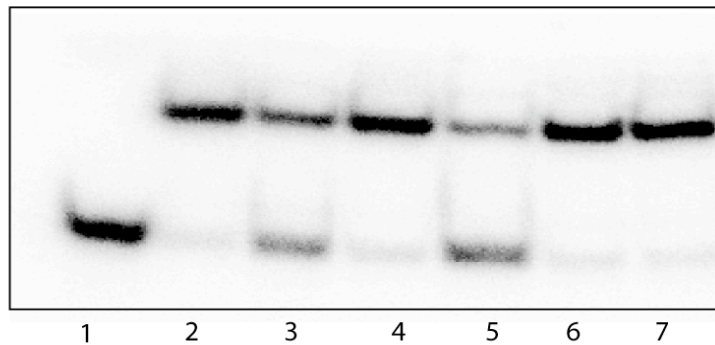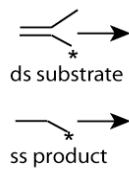
b)
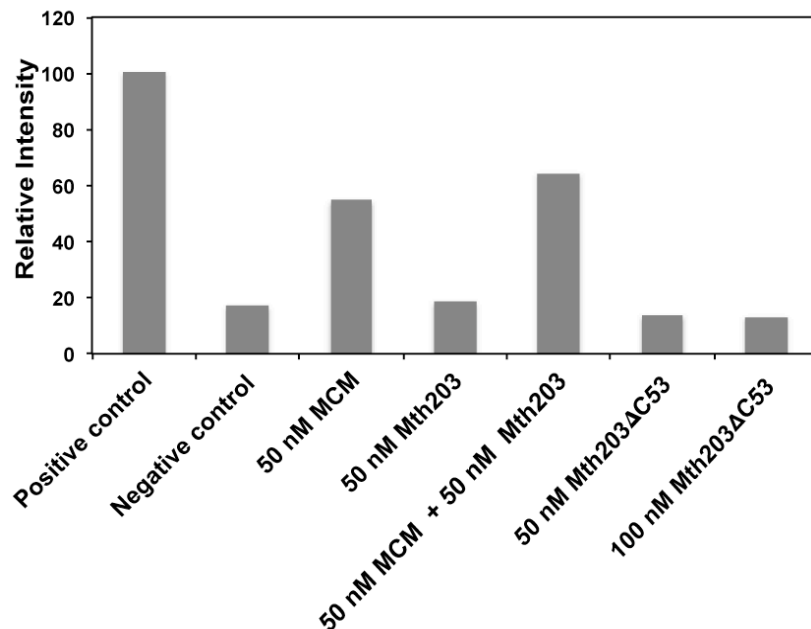
Figure 5.3. Mth203Δ53 does not unwind dsDNA. (a) 11% acrylamide gel showing results of helicase assay of MthMCM and Mth203ΔC53. Lane 1 contains boiled substrate as positive control, lane 2 is no protein negative control, lane 3 contains 50 nM MthMCM, lane 4 contains 50 nM Mth203, lane 5 contains 50 nM MthMCM and 50 nM Mth203 and lanes 6,7 contain 50 nM and 100 nM Mth203ΔC53 protein. (b) Histogram showing the band intensities of unwound single stranded substrate under different conditions. DNA unwinding was observed in the presence of MthMCM and no unwinding was seen in the presence of Mth203ΔC53. A small increase in MthMCM unwinding was observed in the presence of Mth203 (lane 4).

(Kelman *et al.,* 1999, Chong *et al.,* 2000). The band intensity of unwound substrate was compared to the unwinding of negative control, suggesting the presence of these proteins does not have any apparent effect on DNA unwinding.

M. thermautotrophicus MCM has been shown to unwind DNA bound to biotinylated oligonucleotides (Shin *et al.,* 2003). Also, MthMCM can unwind DNA bound to DNA binding proteins like histones, transcription regulator (TrpY) and transcription pre-initiation complex (Shin *et al.,* 2007). Mth203 and Mth203ΔC53 possess DNA binding activity *in vitro*, thus experiments were undertaken to determine whether MthMCM could also unwind the DNA bound by Mth203 and Mth203ΔC53. The forked substrate unwinding in the presence of Mth203 was tested in the presence of 50 nM MthMCM and 50 nM Mth203 and a slight increase in the helicase activity of MthMCM was observed in the presence of 50 nM Mth203 (Figure 5.3).

5.2.1.2   Mth203 and Mth203ΔC53 inhibit MthMCM DNA helicase assay activity

HMt2B (*M. thermautotrophicus* Histone ORF mth2454) (Smith *et al.,* 1997) forms homo-dimers and nucleosome like structures (NLS) on DNA longer than 52 bp (Bailey *et al.,* 1999, Tabassum *et al.,* 1992, Sandman *et al.,* 1990, 1998)). HMt2B was used as a positive control for DNA binding protein displaced by MthMCM.

Strand displacement assays were carried out to test the DNA helicase activity of MthMCM in the presence of Mth203, Mth203ΔC53 and HMt2B. The reaction was carried out in helicase reaction buffer (section 5.2.1.1). 50 or 100 nM MthMCM was added to each reaction and helicase activity was tested in the presence of 50 and 100 nM Mth203, Mth203ΔC53 and HMt2B. Therefore, the ratio between MthMCM (hexamer): Mth203 (dimer) or Mth203ΔC53 (monomer) or HMt2B (dimer) in the helicase reaction was 1:1 and 1:2. Helicase activity was measured by the amount of ^{32}P-labelled oligonucleotide displaced in 1 hour at 50°C. Reactions were terminated by addition of STOP buffer (section 2.18.2). The band intensities of unwound substrate were measured (Quantity One) and graphs were plotted to further visualize the degree of unwinding.

The presence of full-length Mth203 protein did not appear to have any effect on the unwinding activity of MthMCM (Figure 5.4a, lane 5) when compared with 50 nM MthMCM only control (Figure 5.4a, lane 3). However, DNA unwinding was inhibited in the presence of MthMCM: Mth203 concentration of 1:2 (Figure 5.4a, lane 5, 6). The C- terminal deletion mutant Mth203ΔC53, on the other hand seems to inhibit the unwinding activity of MthMCM even at 1:1 concentration (Figure 5.4a, lane 7,8). A decrease in the degree of unwinding by MthMCM was also observed in the presence of HMt2B as the concentration of histone increased from 50 to 100 nM but to a lesser extent than Mth203 (Figure 5.4a, lane 9,10).

a)

MCM	-	-	50	100	50	50	50	50	50	50
Mth203	-	-	-	-	50	100	-	-	-	-
Mth203ΔC53	-	-	-	-	-	-	50	100	-	-
Histone	-	-	-	-	-	-	-	-	50	100
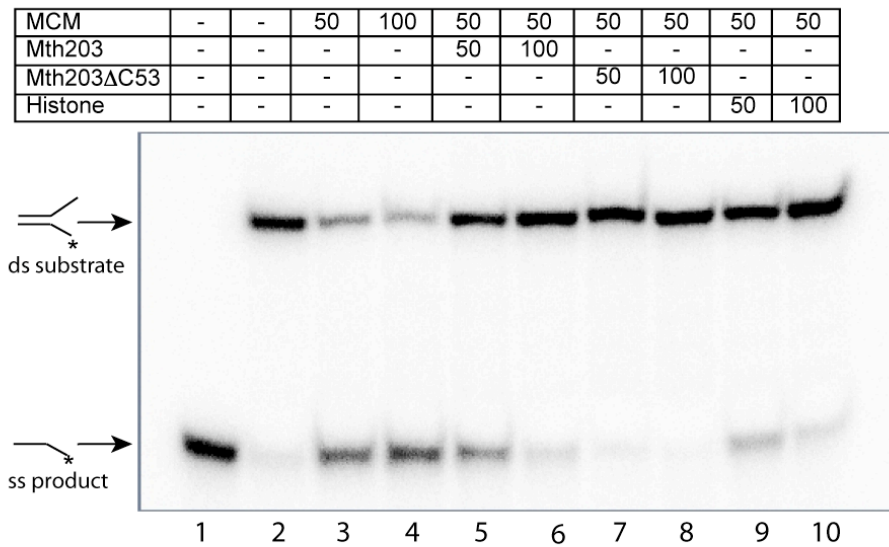
b)
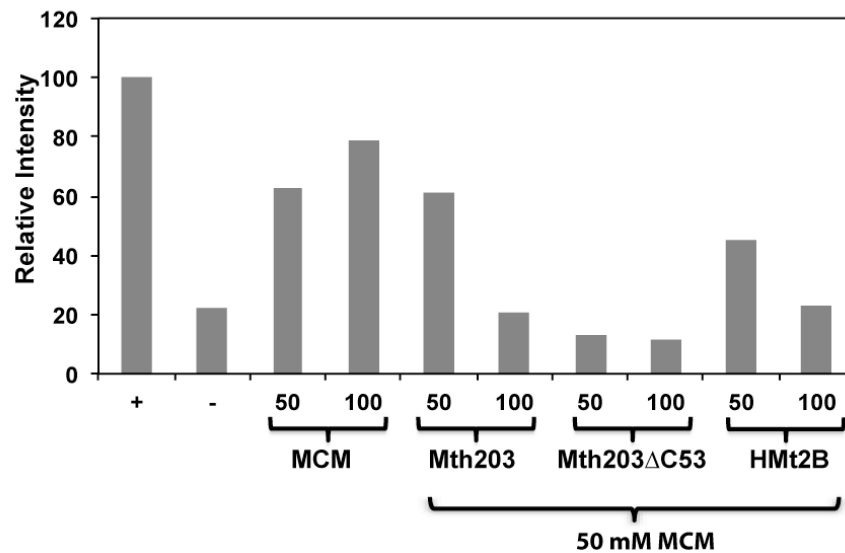
Figure 5.4. DNA helicase activity of MthMCM on a forked substrate is inhibited in the presence of Mth203, Mth203ΔC53 and HMt2B. (a) 11% acrylamide gel showing results of helicase assay of MthMCM, Mth203, Mth203ΔC53 and HMt2B. Lane 1 contains boiled substrate, lane 2 is negative control, lane 3-10 contain various concentrations of MthMCM, Mth203, Mth203ΔC53 and HMt2B as described in the table above. (b) Histogram showing the band intensities of unwound single stranded substrate under different conditions. The DNA unwinding was observed in the presence of MthMCM and inhibited in the presence of Mth203, Mth203ΔC53 and HMt2B.

171

5.2.2 RNA helicase assays to test the RNA unwinding activity of Mth203 and Mth203ΔC53

5.2.2.1 Mth203 has bi-directional RNA helicase activity

Mth203 has been characterized as a putative RNA helicase, thus the RNA helicase activity of the protein was quantified using RNA substrates designed with 10 bp duplex containing alternating C and G nucleotides flanked by single stranded regions of 30 nucleotides (Figure 5.5) (Rozen *et al.,* 1990). Two substrates were designed carrying 3' overhang (RNA A) and 5' overhang (RNA B) to determine the directionality of the RNA unwinding activity of Mth203. The sequence of nucleotides in both the substrates is similar (when reading 3' to 5' direction), the 10 bp duplex contains the same sequence and the overhang sequences are the same but inverted.

(A) DNA A substrate

T7 promoter

5'- TAATACGACTCACTATAG -3'
3'- ATTATGCTGAGTGATATCGCGCGCGCGTTGTTGTAATGTTAAATGTTAGGTGGTACC -3'

in vitro transcription

RNA A substrate

5'- GpppG=C AACAACAUUACAAUUUACAAUCCACCAUGG -3'
C=G
G=C
C=G
G=C
C=G
C=G
G=C
C=GpppG -5'
3'- GGUACCACCUAACAUUUAACAUUACAACAA
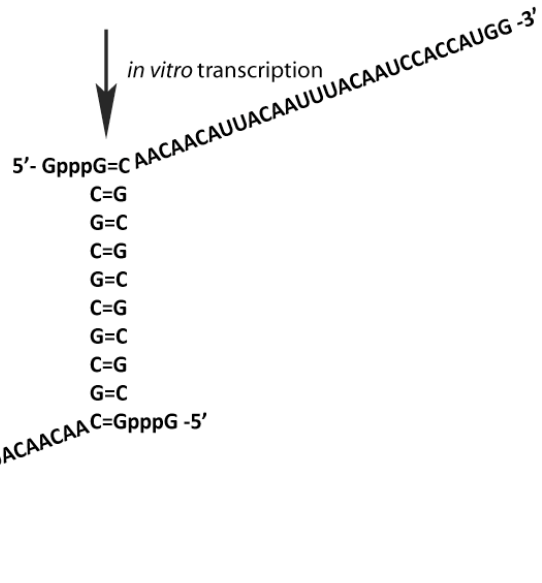
(B) DNA B substrate

T7 promoter

5'- TAATACGACTCACTATAG -3'
3'- ATTATGCTGAGTGATATCCTGGTACCTTGTTGTAATGTTAAATGTTAGCGCGCGCGC-3'

in vitro transcription

RNA B substrate

UAACAUUUAACAUUACAACAAGGUACCAGGpppG -5'
3'- G=C
C=G
G=C
C=G
G=C
C=G
G=C
C=G
G=C
C=G -3'
5'- GpppGGACCAUGGAACAACAUUACAAUUUACAAUC
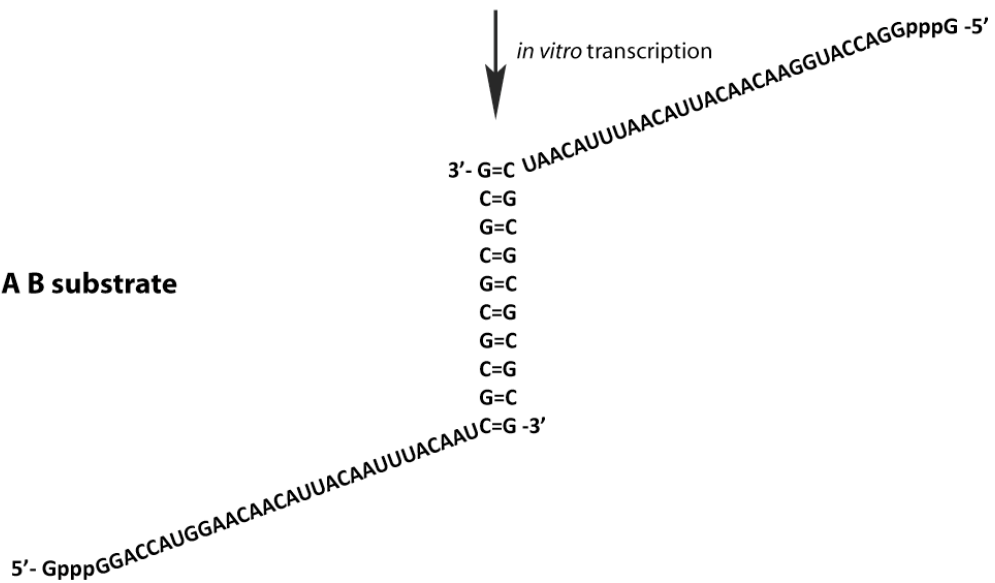
Figure 5.5. The RNA A and RNA B substrates used for RNA helicase assay (Rozen et al., 1990). The substrates have 10 bp dsRNA region, 30 bp long 3' (RNA A) and 5' (RNA B) overhangs. in vitro transcription was carried out using ^{32}P labelled rCTP to label the substrate.

The RNA unwinding activity was observed as the amount of single stranded product formed by unwinding of double stranded substrate and the increased band intensity was plotted as a graph (Figures 5.6 and 5.7). Addition of increasing concentrations of Mth203 from 0 – 5 μg demonstrated bi-directional RNA unwinding activity as shown by an increase in unwinding of both RNA A (3'-5' unwinding) and B (5'-3' unwinding) substrates (Figure 5.6 and 5.7, respectively). Mth203 showed increased unwinding of RNA B substrate (5' overhang) observed as increasing band intensity of single stranded product as the concentration of the protein increased from 0 to 5 μg (Figure 5.6b). Increased unwinding with increasing protein concentration was also observed with RNA A (3'-5' unwinding), however, due to a lot of background in RNA A substrate single stranded product band, the degree of unwinding was plotted as the decrease in the double stranded substrate as observed in Figure 5.7b. As the protein demonstrated unwinding of both the substrates but the results are better visualized in RNA B, due to time constraints further characterization of the RNA helicase activity was carried out using the RNA B substrate.

a)
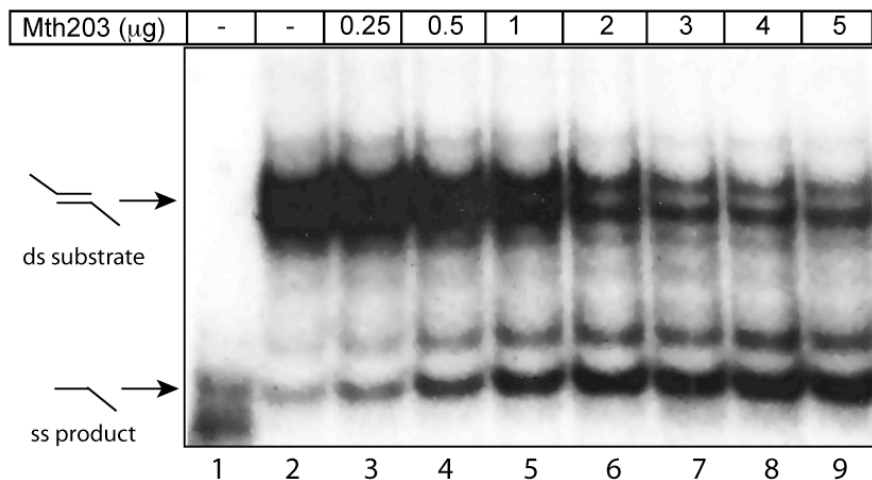
b)
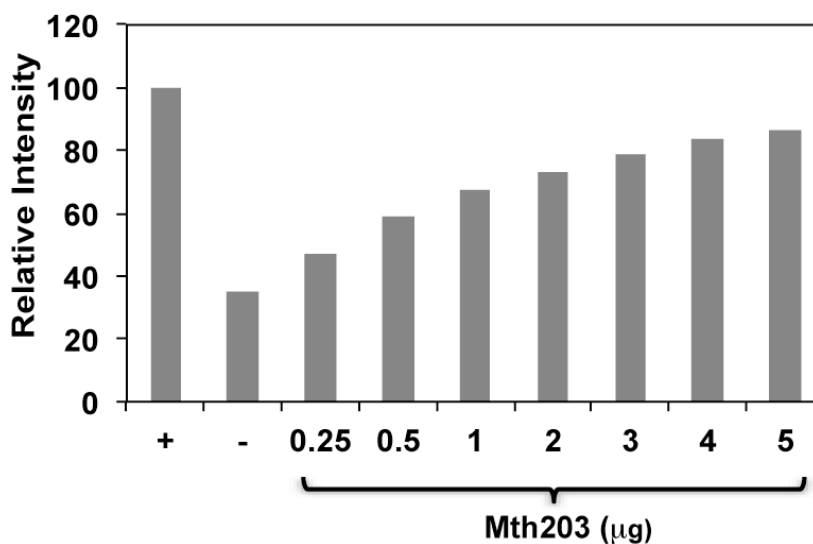
Figure 5.6. Mth203 unwinds RNA B substrate. (a) 12% acrylamide gel showing the degree of unwinding of RNA B substrate in the presence of increasing concentrations of Mth203. Lane 1 contains boiled substrate, lane 2 is negative control, lane 3-9 contain various concentrations of Mth203 concentrations as described in the table above. (b) Histogram showing the band intensities of unwound single stranded substrate under different conditions. An increase in the single stranded RNA product was observed with the increase in Mth203 concentration.

a)
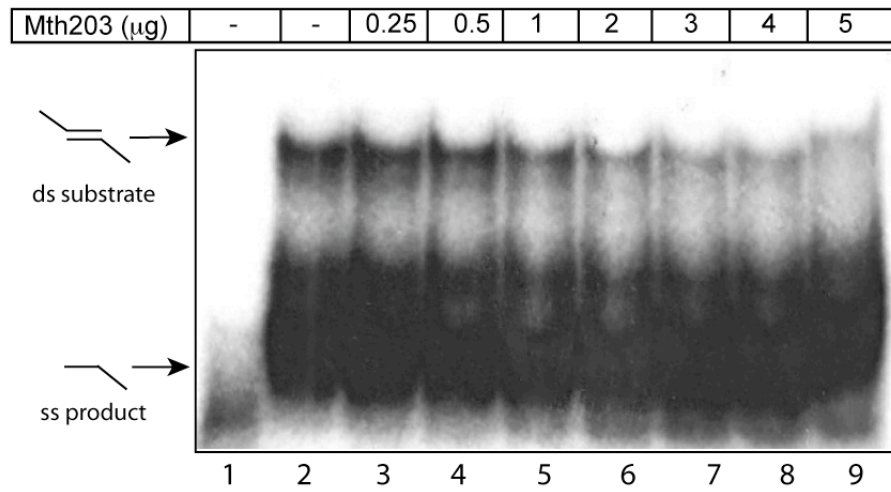
b)
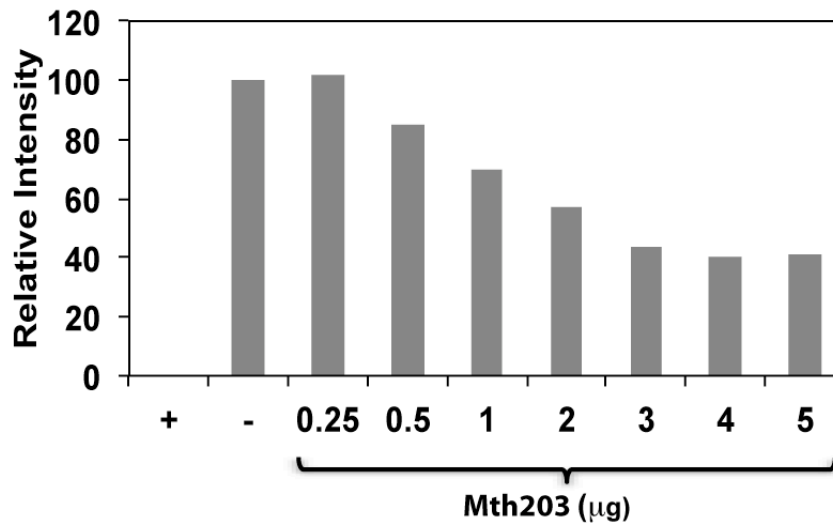
Figure 5.7. Mth203 unwinds RNA A substrate. (a) 12% acrylamide gel showing the degree of unwinding of RNA A substrate in the presence of increasing concentrations of Mth203. Lane 1 contains boiled substrate, lane 2 is negative control, lane 3-9 contain various concentrations of Mth203 concentrations as described in the table above. (b) Histogram showing the decreasing band intensities of double stranded substrate with the increase in Mth203 concentration. A decrease in the unwound RNA substrate was observed with the increase in Mth203 concentration.

5.2.2.2 Mth203 RNA helicase activity is NTP-dependent

DNA and RNA helicase activity is characterized as oligonucleotide unwinding in the presence of NTP (Geider *et al.,* 1981). *In vitro* RNA-dependent ATPase activity has been reported for many RNA helicases, whereas, in vitro NTP-dependent RNA helicase activity has been reported for very few RNA helicases (Svitkin *et al.,* 2001, Rocak and Linder, 2004). The requirement for NTP hydrolysis for Mth203 RNA unwinding activity was tested in the presence of ATP and GTP. Mth203 was able to unwind RNA in the presence of 0.5 mM ATP (Figure 5.8a, lane 4) and similar unwinding was also observed in the presence of 0.1 mM GTP (Figure 5.8A, lane 8). No unwinding was observed in the absence of NTPs (Figure 5.8a, lane 3, 6). Under the present experimental conditions, 1 mM ATP and 0.1 mM GTP are required for an optimum unwinding (as seen in Figure 5.8a, lane 10). Almost 100 % substrate unwinding was observed in the presence of Mth203ΔC53 (Figure 5.8a, lane 9), this was very unusual and it seems that there was most probably a contamination in the sample and the experiment should be repeated to observe the RNA helicase activity of the C- terminal deletion mutant Mth203ΔC53.

a)

	1	2	3	4	5	6	7	8	9	10
Mth203ΔC53(μg)	-	-	-	-	-	-	-	-	2	-
Mth203 (μg)	-	-	2	2	2	2	2	2	-	2
ATP (mM)	1	1	0	0.5	1	-	-	-	1	1
GTP (mM)	0.1	0.1	-	-	-	0	0.05	0.1	0.1	0.1
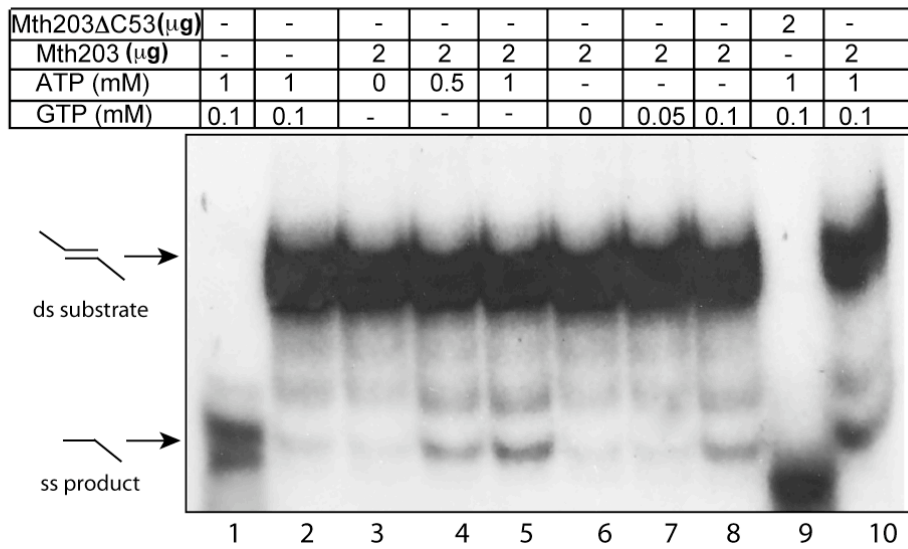
b)
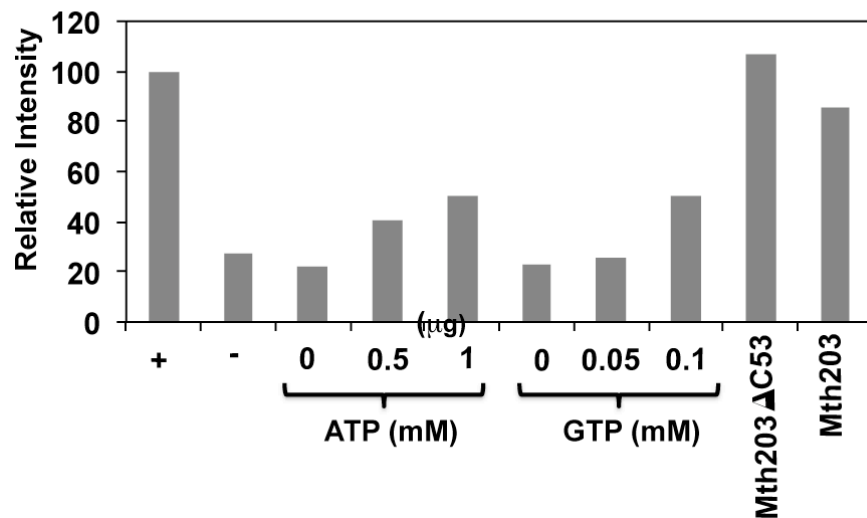
Figure 5.8. Mth203 has NTP-dependent RNA helicase activity. (a) 12% acrylamide gel showing the degree of unwinding of RNA B substrate in the presence of different ATP/GTP concentrations. Lane 1 contains boiled substrate, lane 2 is a negative control, lanes 3-5 contain increasing concentrations of ATP, lane 6-8 contain increasing concentrations of GTP, lane 9 contains Mth203ΔC53 and lane 10 contains Mth203 as described in the table above. (b) Histogram showing the band intensities of unwound single stranded substrate under different conditions. The RNA helicase activity does not require presence of specific NTPs. Same amount of RNA substrate unwinding was observed in the presence of 1 mM ATP and 0.1 mM GTP, showing highest degree of unwinding when both are present in the assay.

5.2.2.3   Mth203 prefers Mn^{2+} divalent ions for unwinding RNA substrates

In order to optimize the activity, the effect of divalent cations on the RNA helicase activity of Mth203 was investigated in the presence of varying concentrations (0 - 0.5 mM) of Mg^{+2} and Mn^{+2} divalent ions (Figure 5.9). Mth203 shows RNA helicase activity dependent on the presence of divalent cations. Mth203 shows unwinding in the presence of both Mg^{+2} and Mn^{+2}, but more unwinding was observed in the presence of Mn^{+2} (Figure 5.9a, lane 3-5). No unwinding was observed in the absence of cations (Figure 5.9a, lane 3, 6).

a)

Mth203	-	-	2	2	2	2	2	2
Mn 2+ (mM)	0.5	0.5	0	0.25	0.5	-	-	-
Mg 2+ (mM)	0.5	0.5	-	-	-	0	0.25	0.5

ds substrate

ss product

1   2   3   4   5   6   7   8
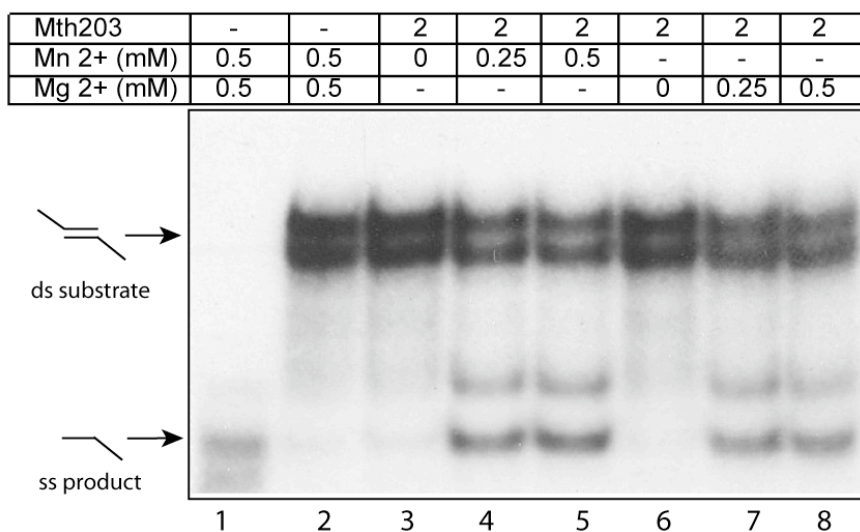
b)
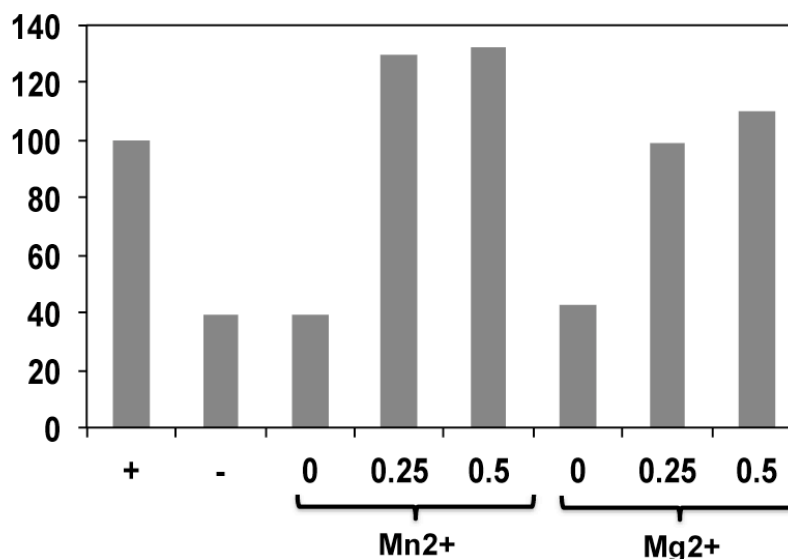
Figure 5.9. Mth203 shows higher degree of RNA B substrate unwinding in the presence of Mn^{+2} compared to Mg^{+2}. (a) 12% acrylamide gel showing the degree of unwinding of RNA B substrate in the presence of different Mg^{2+}/Mn^{2+} concentrations. Lane 1 contains boiled substrate, lane 2 is a negative control, lanes 3-5 contain increasing concentrations of Mn^{2+} and lanes 6-8 contain increasing concentrations of Mg^{2+} described in the table above. (b) Histogram showing the band intensities of unwound single stranded substrate under different conditions. The RNA helicase activity is dependent on the presence of divalent cations as no unwinding is observed in lane 3 and 6. In addition, Mth203 prefers the presence of Mn^{+2} as compared to Mg^{+2} for the unwinding activity.

5.2.2.4   RNA helicase activity of Mth203 over time

A time course of the unwinding reaction was performed on Mth203 (5 μg). The reactions were started at the same time by addition of Mth203 and the reactions were stopped at designated time periods by addition of STOP buffer and stored at 4°C until loaded on a gel. The time course results show that the reaction is fast in first 10 minutes and then slowly peaks to maximum (Figure 5.10 a and b). More than 50% unwinding was observed in the first 10 minutes (64%) after the addition of Mth203, 91% unwinding was observed at 25 minutes (Figure 5.10 b) and maximum unwinding was observed in 30 minutes (Figure 5.10 b).

a)

Mth203 (μg)	-	-	2	2	2	2	2	2	2	2
Time (min)	0'	0'	0'	2.5'	5'	10'	15'	25'	30'	45'

ds substrate →

ss product →

    1   2   3   4   5   6   7   8   9   10
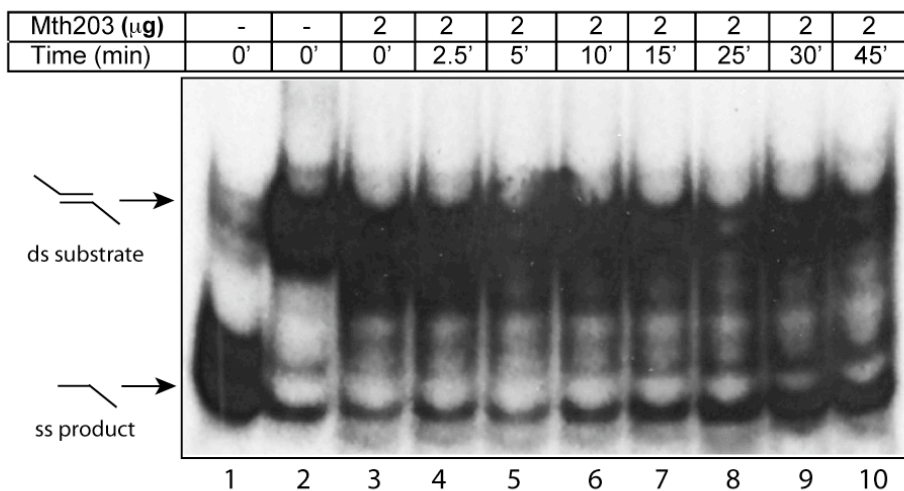
b)
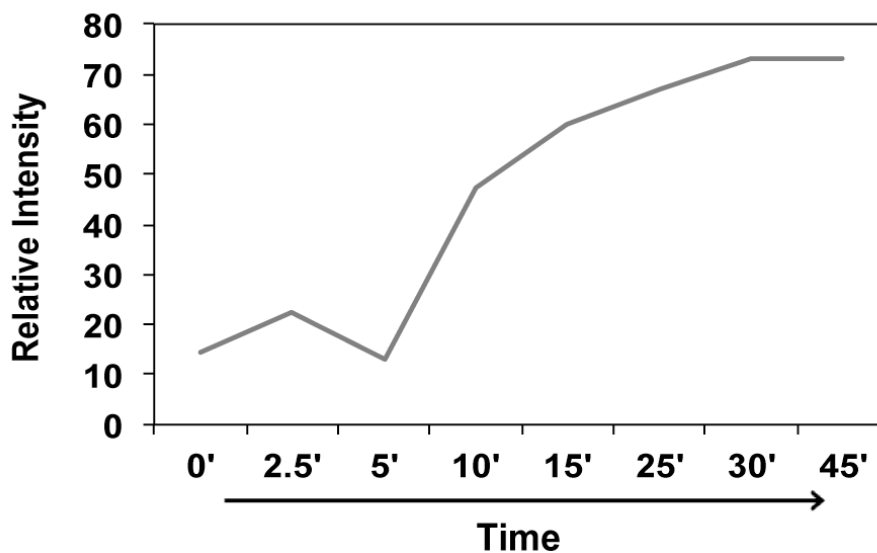
Figure 5.10. Mth203 shows increase in the RNA helicase activity over time. (a) 12% acrylamide gel showing the increase in RNA B substrate unwinding by Mth203 over time. Lane 1 contains boiled substrate, lane 2 contains negative control, lane 3-10 shows unwinding of Mth203 over time 0-45 min (b) Graph showing the band intensities of unwound single stranded substrate over time. 91% substrate unwinding was observed at 25 minutes of assay, reaching maximum at 30 minutes.

5.2.3 Analysis of NTP hydrolysing activity of Mth203 and Mth203ΔC53

5.2.3.1 Mth203 has RNA-independent NTPase activity

The presence of RNA does not have any effect on the NTPase activity of Mth203, whereas the Walker A (AAA$^+$ domain) mutant shows complete loss of NTPase activity (Figure 5.11b). Mth203 was able to hydrolyse both GTP and ATP, however, ATP hydrolysis is almost twenty times that of GTP. There is no apparent increase in the NTPase activity when both ATP and GTP are present, the hydrolysis obtained appears to be a sum of individual ATP/GTP hydrolysis (Figure 5.11 B). In addition, the removal of the C- terminal domain does not have any effect on the NTPase activity of Mth203 (Figure 5.11b) as very little decrease in the NTPase activity in the presence and absence of RNA was observed, suggesting that the NTPase activity requires an intact Walker A domain but not a full-length protein.
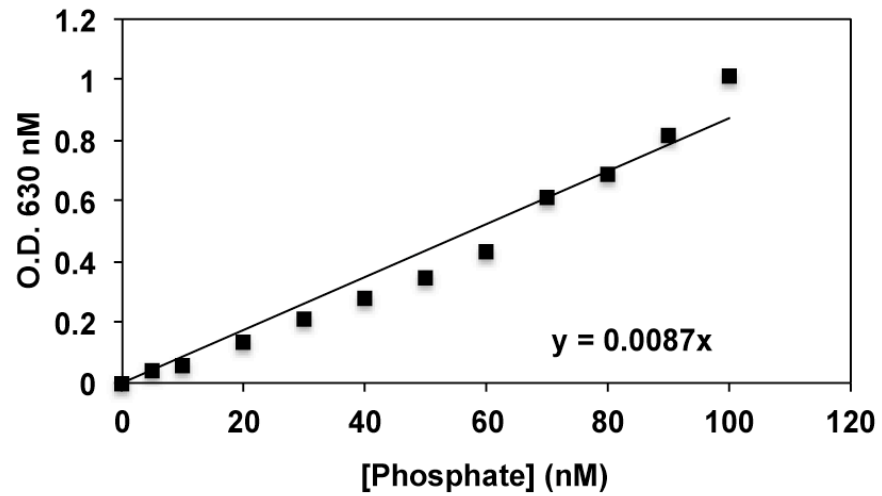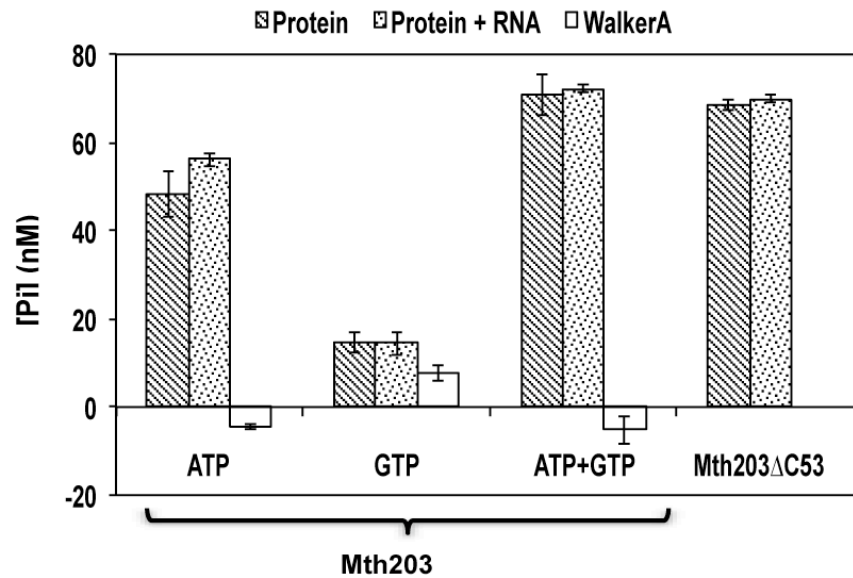
Figure 5.11. Mth203 and Mth203ΔC53 possess RNA independent NTPase activity.
(a) Calibration curve of malachite green assay with a range of phosphate concentrations (0-100 nM). Five replicates of each concentration were assayed with Malachite green. Mean OD_{630} was plotted against concentration of phosphate. (b) NTPase assay of Mth203, Mth203ΔC53 and Walker A mutant in the presence and absence of RNA, ATP and GTP. The reactions were set-up containing ATP, GTP or both and assayed for 30 min. The NTPase activity was highest in presence of ATP and GTP, however, RNA does not have any effect on the NTPase activity.

5.2.4 Mth203 expression in the cell cycle of *M. thermautotrophicus*

The RNA helicase and NTPase activities suggest a possible physiological role of Mth203 in RNA metabolism of *M. thermautotrophicus*, however, the exact function of the protein in the organism still remains unknown. Studying the expression profile of the profile might help in understanding and visualising the role of Mth203, if any, in the cell cycle.

5.2.4.1 Synchronization of cells in early log phase

A nutritional starvation method to synchronize cells leads to the arrest of rapidly growing cells right before cell division (Ron *et al.,* 1975). In order to investigate the expression profile of Mth203 in the cells under growth conditions *M. thermautotrophicus* cells were synchronized using modified method of nutritional starvation (Morgan *et al.,* 1997, Dr. Paul McDermott, PhD thesis, 2009). The cells were synchronized by N_2 treatment, so that all the growing cells reach the same phase of the cell cycle and did not progress further until the treatment was removed (see section 2.1.3) (Dr. Paul McDermott, PhD thesis, 2009). Once the treatment was removed all the cells progressed through the cell cycle at the same speed for a couple of divisions before becoming asynchronous again.

M. thermautotrophicus cells were grown in 2 L culture in a fermenter (section 2.1). After the cells reached $O.D._{600}$ of 0.6-0.7, they were synchronised by replacing H_2 gas with N_2 and decreasing the impeller speed for 26 hours, causing all the cells to reach at the same phase of cell cycle (before cell division) and did not progress further until H_2 was provided (shown by arrow in Figure 5.12a) (section 2.1.3, 2.11). The samples were taken aerobically every 10 minutes, O.D. measurements were taken and cells were stored for flow cytometry (section 2.21) and western blots (section 2.14).

The growth curve of the *M. thermautotrophicus* culture shows an initial lag phase and exponential growth of cells until they reach OD_{600} 0.7 and then the growth in inhibited due H_2 limitation (seen as a plateau in the graph for 26 hours). When the N_2 is again replaced by H_2 the cells start growing in 10 minutes and exponential growth is resumed. The cells reached OD_{600} 0.7 at 1755 minutes and
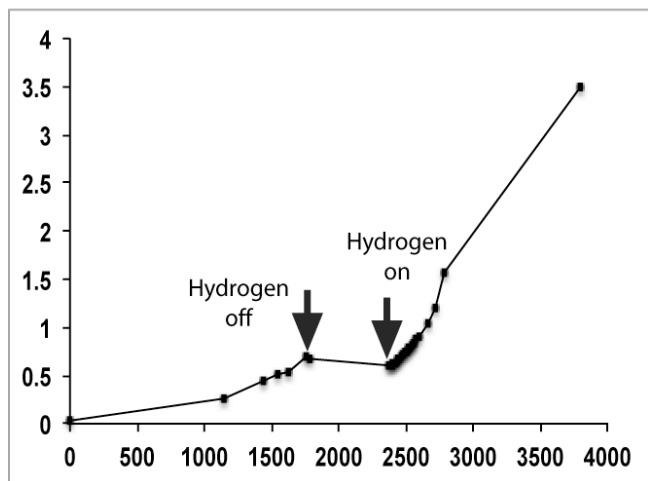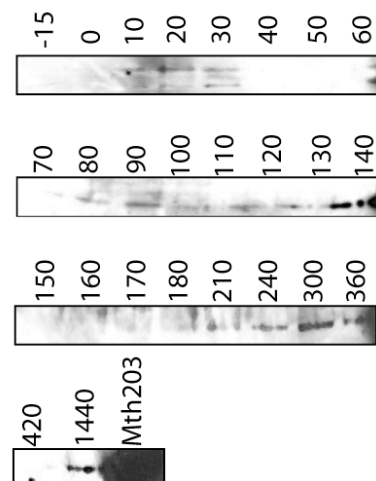
Figure 5.12. Cyclic expression of Mth203 in the cell cycle of M. thermautotrophicus.
(a) Growth curve of synchronized *M. thermautotrophicus* cell culture, the cells were synchronized by switching-off the hydrogen gas for 26 hours in mid-log phase. (b) Western blot of the cells harvested at regular intervals after re-starting the Hydrogen and probed with anti-Mth203 antibodies. The time at which the samples were harvested is shown in minutes above the blots.

then the H_2 was replaced by N_2, and the growth rate stopped (Figure 5.12a). At 2370 minutes, N_2 was stopped and H_2 was turned on at the OD_{600} of 0.616. Once the H_2 was restarted the cells grew exponentially to an OD_{600} of 3.5 at 3805 minutes (Figure 5.12 a).

5.2.4.2 *Western blot to visualize the expression of Mth203 in the cell cycle*

The nucleoid distribution and DNA content of *M. thermautotrophicus* could not be studied due to problems encountered with the flow cytometer. The western blots for the expression of Mth203 in the whole cell extract of *M. thermautotrophicus* show a possible cyclic expression of Mth203 during the cell cycle. It was observed that no Mth203 was expressed before the addition of H_2, and at zero minutes (resting phase). After 10 minutes of the addition of H_2, the protein expression was observed for a period of approximately 20 minutes (Figure 5.12b). Then after 40 minutes, Mth203 expression was observed again (at 80 minutes after addition of H_2 Figure 5.12b) for a period of 60 minutes and disappeared for 40 minutes. Thereafter, the cell samples were taken every 30 minutes and 1 hour, and a strong band of Mth203 indicating the expression of Mth203 was observed at 240, 300, 360 and 1440 minutes respectively (Figure 5.12b). A complete replication run out is not observed in M. thermautotrophicus, which would explain the presence of Mth203 even at very high OD_{600} (Majernik *et al.,* 2005). As the latter samples were far apart, no conclusions can be derived regarding the presence of bands after 180 minutes of the addition of H_2, thus, making it difficult to deduce the periodicity of Mth203 expression in the *M. thermautotrophicus* cell cycle.

5.3   Discussion

Mth203 contains a conserved DEAD-box domain and was assumed to hydrolyse ATP and unwind RNA, but a direct biochemical demonstration of helicase activity was essential. In the present study the biochemical activity of Mth203 and Mth203ΔC53 was extensively characterised for the first time. We have shown that although the protein possesses DNA binding activity neither Mth203 nor Mth203ΔC53 demonstrate DNA (3'-5') unwinding activity. Inhibition of MthMCM DNA helicase activity was observed in the presence of Mth203 and Mth203ΔC53. In addition, it is demonstrated for the first time that Mth203 possesses an ATP-dependent bidirectional RNA helicase activity and RNA-independent ATPase activity.

Except Dbp9 (Kikuma *et al.,* 2004), all DEAD-box proteins show bi-directional RNA helicase activity and require at least one RNA strand for unwinding activity (Cordin *et al.,* 2006, Rocak and Linder, 2004). Also, some DEAD-box proteins like eIF4A demonstrate RNA helicase activity in the presence of cofactors or other proteins (Rozen *et al.,* 1990). Mth203 displays an NTP-dependent bidirectional RNA unwinding activity. The RNA helicase activity of Mth203 requires the presence of NTPs and divalent cations, the standardization assays have suggested that maximum activity was observed in the presence of ATP and Mn^{+2}.

The NTPase activity of Mth203 and Mth203ΔC53 did not require the presence of RNA substrate suggesting the Walker A motif (ATPase domain) might function independently of the helicase domain. Vasa (DEAD-box protein in *D. melanogaster*) (Liang *et al.,* 1994), Cap-Rf (DEAD-box protein in HCV) are the only known RNA helicases possessing RNA-independent ATPase activity (You *et al.,* 1999). Other DEAD-box RNA helicases possess ATPase activity that is dependent on or stimulated by RNA substrates (summarised in Cordin *et al.,* 2006). In addition, the ATPases assays performed show that the ATPase activities did not require any specific RNA substrates with an exception of DbpA (DEAD-box protein in *E.coli*), which is the only known DEAD-box protein possessing RNA substrate specificity for ATPase assay (Fuller-Pace *et al.,* 1992, Tsu *et al.,* 1998).

Using western blotting experiments it was observed that the Mth203 protein levels varied during the cell cycle of *M. thermautotrophicus*. It was proposed earlier

that *M. thermautotrophicus* has a cell cycle of 135 minutes (Dr. Paul McDermott, PhD thesis, 2009) using a synchronized cell population it was observed that the protein appeared to be expressed in a periodic manner most probably after G1 phase. The RNA helicase and ATPase activity of Mth203 and its implication on Mth203 function in the cell cycle of *M. thermautotrophicus* is discussed in further detail in Chapter 7.

A thorough understanding of the effect of Mth203 would require genetic manipulation studies to create mutants of *M. thermautotrophicus* containing variants of the Mth203 protein. However, basic tools for genetic manipulations e.g. self-replicating high-copy number plasmids, selectable markers, transformation methods, etc., are absent in *M. thermautotrophicus*.

6   **Genetic manipulation studies of the Mth203 homologue in *M. maripaludis* (Mmp0457)**

6.1   *Introduction*

M. maripaludis is a well-studied methanogen that can be used to carry out genetic manipulation including gene-knock out studies and protein expression (Blank *et al.,* 1995, Argyle *et al.,* 1996, Gardner *et al.,* 1999). In an independent study to identify proteins involved in DNA replication in *M. maripaludis*, the N- terminus of MmpMCMA (MCM), one of four MCMs in *M. maripaludis*, was Histidine-tagged and affinity purified from cell extract leading to co-purification of Mmp0457, a Mth203 homologue (Dr. Alison Walters, PhD thesis, 2010). The interaction of Mmp0457 with MCMA, further strengthens the hypothesis that Mth203 and its homologues may play a role in archaeal DNA replication. To elucidate whether Mmp0457 is essential, a markerless mutagenesis strategy (Moore and Leigh, 2005) was used to delete the gene. To identify the role of the protein in the cell cycle, N- terminal Histidine-tagged Mmp0457 was expressed *in vivo* and physiological changes in the cell size and shape were visualized using flow cytometry.

6.2 Results

6.2.1 Deletion of *mmp0457* in *M. maripaludis* using markerless mutagenesis

The markerless mutagenesis strategy (Moore and Leigh, 2005) was used to delete *mmp0457* from the genome of *M. maripaludis* (Figure 6.1), if the deletion mutant was viable, the protein is not essential, however, if a deletion mutant could not be generated, it would suggest the protein is essential.

6.2.1.1 Construction of complete and partial deletion plasmids of mmp0457

1617 bp of *mmp0457* along with 500 bp of up- and downstream flanking DNA was amplified from *M. maripaludis* gDNA by PCR using primers MMP457A and MMP457B containing *NotI* sites (see appendix A for primer sequences). The PCR products were gel purified and cloned into pGEM-T (Promega) to generate a plasmid pCB01 (Figure 6.1). In order to generate a plasmid for the deletion of the gene, a second PCR was carried out on pCB01 using primers MMP457C and MMP457D that bound between the coding and flanking sequences. The plasmid was amplified outwards resulting in a product possessing self ligating ends with 10 amino acids of in-frame *mmp0457* coding sequence including the start and stop codons, 500 bp up- and down- stream flanking sequence and pGEM-T backbone. The product was self-ligated to generate pCB02 (Figure 6.1).

For partial deletion of *mmp0457* (deletion of 269 amino acids from the C-terminal)*,* PCR was carried out using primers MMP0457A and MMP0457E such that the product contains 269 amino acids of in-frame *mmp0457* coding sequence (N-terminal half of the original protein length) along with start and stop codons, 500 bp up- and down- stream flanking sequence and pGEM-T backbone, resulting in pCB04 (Figure 6.1). The 500 bp flanking sequence and intervening coding sequence from pCB02 and pCB04 were excised using NotI and sub-cloned into the *NotI* site of the multiple cloning site of pCRPrtNeo to generate *Δmmp0457* plasmid, pCB03 (Figure 6.2) and *partΔmmp0457* plasmid, pCB05 (Figure 6.3). pCRPrtNeo cannot be maintained independently in the cell and must integrate into genomic DNA by homologous recombination (Moore and Leigh, 2005). pCRPrtNeo backbone contains ampicillin and kanamycin resistance genes for selection in *E. coli* and neomycin resistance cassette and an *hpt* gene for selection in *M. maripaludis*. The *hpt* gene

confers resistance to the base analogue 8-azahypoxanthine when expressed in *M. maripaludis* and allows negative selection by forced removal of the vector by a second recombination event, resulting in either wild-type or deletion mutant genotypes (Figure 6.1) (Moore and Leigh, 2005).

A → mmp0457

M. maripaludis mmp0457 from genomic DNA along with 500bp up- and down-stream regions was amplified using PCR.

C/E
0457
D/F
pCB01

The product was cloned into pGEM-T. Then amplified outwards from the coding sequence by PCR to create deletion of *mmp0457*.

NotI   NotI
pCB02
or
pCB04

The ends of PCR product were ligated to reform circular plasmid containing flanking regions and a small fragment of Mmp0457 coding sequence.

NotI   NotI
pCB03
or
pCB05
neo   hpt

The flanking regions were subcloned into pCRPrtNeo using NotI. This plasmid will be used to transform into *M. maripaludis* cells.

genomic DNA
mmp0457

NotI   NotI
pCB03
or
pCB05
neo   hpt

Recombination of transformed plasmid an Mmp0457 on genomic DNA. Recombination should occur between *mmp0457* flanking regions in the vector and genomic DNA.

The integrants were selected by growth on neomycin.

hpt   neo   mmp0457

The transformants were grown without selection to allow looping out and then add 8-azahypoxanthine.

50%   50%
mmp0457

8-azahypoxanthine selected cells were identified which have WT or Δ *mmp0457* genotype by PCR and Southern blot. 100% of selected cells having WT *mmp0457* would suggest *mmp0457* being essential.

PCR and Southern blots

(modified from Ally Walters, PhD thesis)

Figure 6.1. Schematic representation of marker-less mutagenesis of Mmp0457 in MM900 strain of M. maripaludis using plasmids pCB03 and pCB05.

193
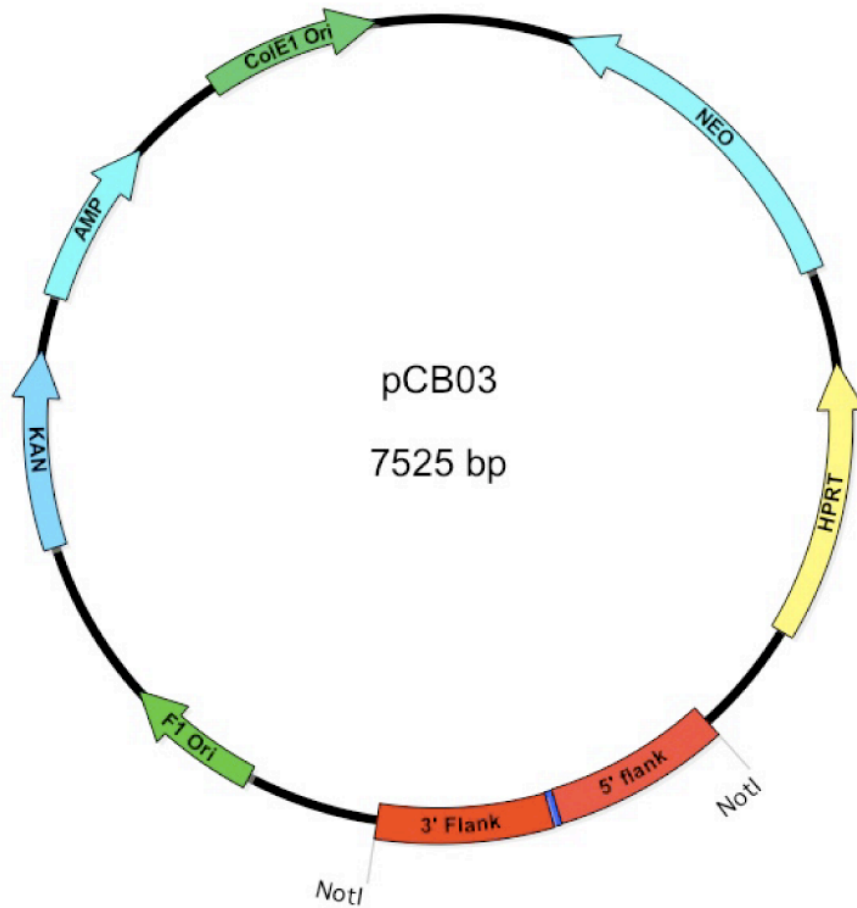
Figure 6.2. Map of the mmp0457 deletion plasmid, pCB03. *mmp0457* 500 bp flanking regions (red) with 10 amino acids of *mmp0457* coding sequence (blue) were cloned into the *NotI* site of pCRPrtNeo (Moore and Leigh, 2005). This plasmid can replicate in *E. coli* but has no functional replication origin for *M. maripaludis*.
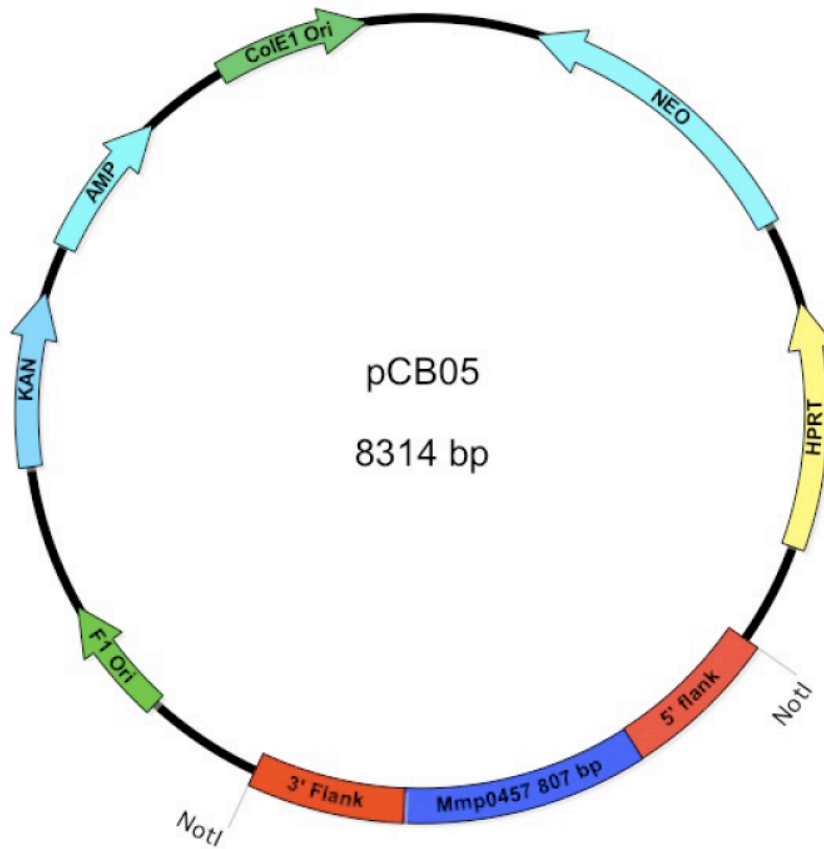
Figure 6.3. Map of the mmp0457 partial deletion plasmid, pCB05. *mmp0457* 500 bp flanking regions (red) with 268 amino acids of *mmp0457* coding sequence (blue) were cloned into the *NotI* site of pCRPrtNeo (Moore and Leigh, 2005). This plasmid can replicate in *E. coli* but has no functional replication origin for *M. maripaludis*.

6.2.1.2   *Transformation of deletion plasmids*

An outline of the markerless-mutagenesis method used for selection and counter-selection of transformants is shown in figure 6.1. The markerless-mutagenesis experiments were carried out in the *Δhpt* strain of *M. maripaludis*, MM900 (Moore and Leigh, 2005). 5 μg of each plasmid DNA was used for transformation using the PEG-based transformation method (see materials and methods section 2.5.2). 30 colonies were grown in liquid McCas medium with no selection for each transformation reaction. Genomic DNA was isolated from each culture and tested for gene deletion by PCR and Southern blot analysis (see section 2.2.6, 2.15). If the gene was not essential and there was no bias in the looping-out event, a 1:1 ratio of wildtype:deletion genotype was expected in the results (Figure 6.1).

6.2.1.3   *Isolation of genomic DNA from transformants*

Out of 30 colonies, 27 cultures from the complete deletion and 28 cultures from the partial deletion showed growth. Genomic DNA was isolated from all these 8-aza hypoxanthine selection cultures. DNA was also extracted from the wild-type MM900 control and a neomycin-selected culture from pCB03 (03 Neo) and pCB05 (05 Neo) as controls for the intermediate stage of the deletion process, where the plasmid DNA was integrated into the genome.

6.2.1.4   *Attempted deletion of full-length mmp0457*

6.2.1.4.1   PCR and Southern blot analysis of transformation with pCB03

Genomic DNA from the pCB03 transformation was screened by PCR using primers MMP457F and MMP0457G, which bound to flanking regions of *mmp0457*. The wild-type genotype should result in 2617 bp long PCR product and the deletion mutant *Δmmp0457* should result in a 1030 bp long product. MM900 genomic DNA and plasmid pCB03 were used as wild-type and deletion controls respectively. The PCR results were resolved on a 1% agarose gel and southern blotted (Figure 6.4). Products were not obtained for all the transformants suggesting the PCR may require optimization. However, in all reactions where product was obtained the bands coincided with a wild-type genotype product (2019 bp) shown by MM900

gDNA PCR. Smaller bands (1030 bp) suggesting a deletion product were absent in the transformants.

In order to confirm the genotypes of pCB03 transformants, a southern blot was carried out (as described in materials and methods section 2.15). Genomic DNA from 27 pCB03 transformants, MM900 (WT) and Neomycin intermediate (03 Neo) was digested using *PsiI*. The expected size of the bands was 2617 bp for wild type and 1030 bp for deletion genotype. The southern blot showed that all 27 transformants produced wild-type size bands indicating that the plasmid DNA has been lost in all the transformants (Figure 6.5). The banding pattern displayed by the neomycin intermediate (03 neo) showed plasmid integration on the chromosome and formed a unique restriction site pattern not observed in any of the transformants (Figure 6.5). The complete process of transformation, selection and analysis of transformants was repeated using pCB03 in order to validate the results obtained in first transformation. The results again demonstrated 100% wild-type size bands.
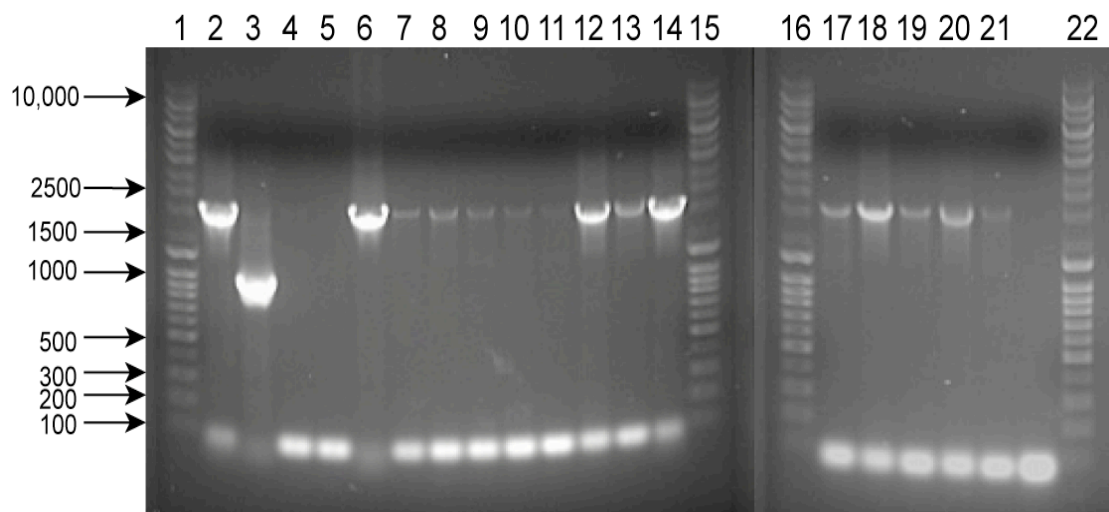
Figure 6.4. Colony PCR demonstrates all the pCB03 transformants are 100% wild type. Lane 1, 15, 16, 22 contains DNA molecular weight marker (bp), lane 2 contains *mmp0457* gene amplification from gDNA, lane 3 shows *mmp0457* deletion with flanking regions in pCB03 plasmid, lane 4-14, 17-21 contain pcr products from the gDNA of selected colonies. 100% colonies are wild type suggesting *mmp0457* may be essential.
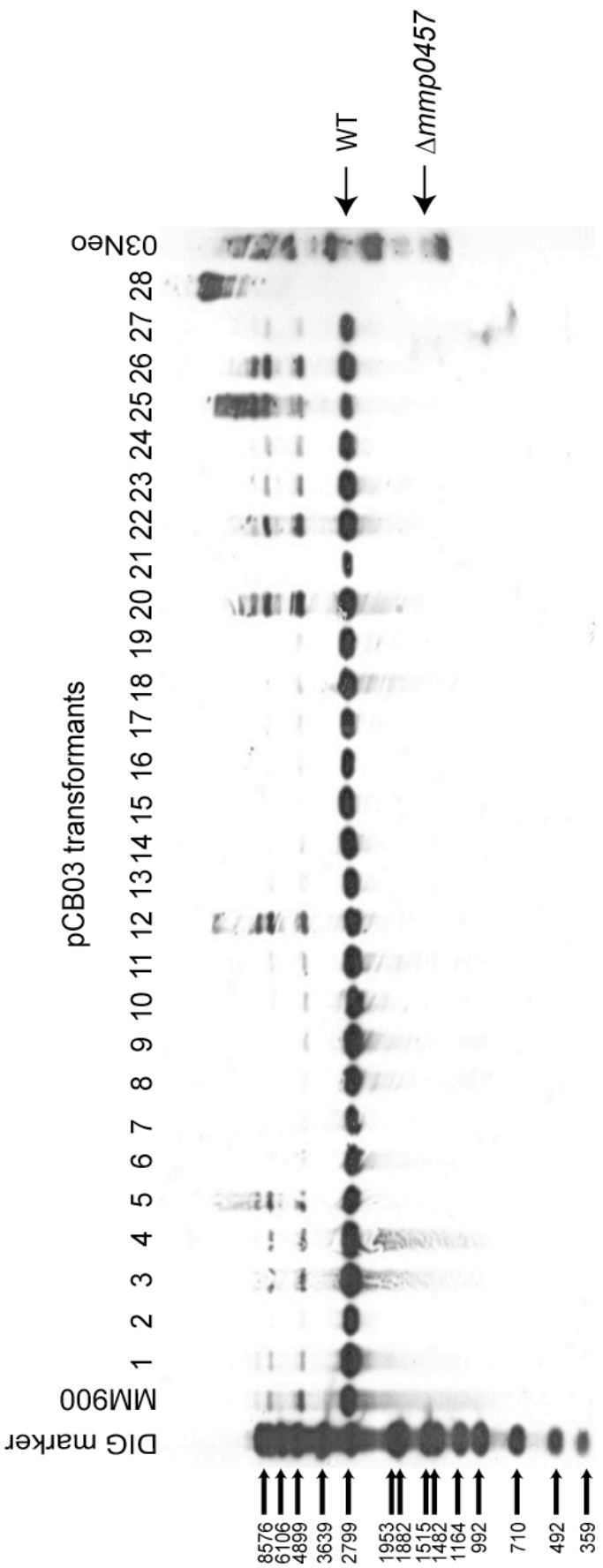
Figure 6.5. All the pCB03 transformants possess WT genotype. Genomic DNA from 28 transformants was digested, blotted and then hybridized with a DIG-labelled probe specific for *mmp0457* flanking regions. Expected fragment sizes were 2617 bp (WT) and 1030 bp for Δ*mmp0457* respectively. All 28 transformants gave a WT sized fragment. Genomic DNA from the neomycin intermediate (03 neo) was also digested and shows a distinct pattern.

pCB03 transformants

WT

Δ*mmp0457*

03Neo

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

MM900

DIG marker

8576
6106
4899
3639
2799
1953
1882
1515
1482
1164
992
710
492
359

6.2.1.5   *Attempted partial deletion of mmp0457*

6.2.1.5.1   Southern blot analysis of transformation with pCB05

Genomic DNA was extracted from 28 pCB05 transformants and digested with *NotI*. The resulting fragments were separated by gel electrophoresis and transferred on positively charged nylon membrane. The membrane was hybridised with a DIG-labelled probe (see section 2.15). All the transformants displayed the wild-type band at 2617 bp. The expected band size of 1530 bp for the deletion mutant was not observed in any of the transformants (Figure 6.6). The banding pattern displayed by the neomycin intermediate (05 neo) was similar to that of neomycin intermediate of pCB03 (03 neo), suggesting that the plasmid integrates into the chromosome using the same homologous regions, generating the same patterns of restriction sites.

6.2.1.6   *Conclusion*

As it was not possible to create a null mutant or partial deletion mutant for *mmp0457*, it can be concluded that the gene is probably essential for the normal growth of *M. maripaludis*. However, it is also possible that Δ*mmp0457* mutant is viable but deletion occurs at very low frequency. In order to further substantiate that *mmp0457* is essential, plasmid-based complementation experiments are required (Zhang *et al.,* 2002, Sandbeck and Leigh, 1991). The process involves the cells to be transformed with a replicative plasmid containing gene of interest (*mmp0457)* and genomic copy of the gene is deleted, then plasmid loss is forced by counter selection. If *mmp0457* is essential, loss of plasmid would be lethal.

Figure 6.6. All the pCB05 transformants possess WT genotype. Genomic DNA from 27 transformants was digested, blotted and then hybridized with a DIG-labelled probe specific for *mmp0457* flanking regions. Expected fragment sizes were 2617 bp and 1530 bp for WT and Δ*mmp0457* respectively. All 28 transformants gave a WT sized fragment. Genomic DNA from the neomycin intermediate (05 neo) was also digested and shows a similar pattern of plasmid insertion in the genome as the complete deletion intermediate 03 Neo, as the plasmid integrates into the chromosome using the same homologous regions, generating same patterns of restriction sites.
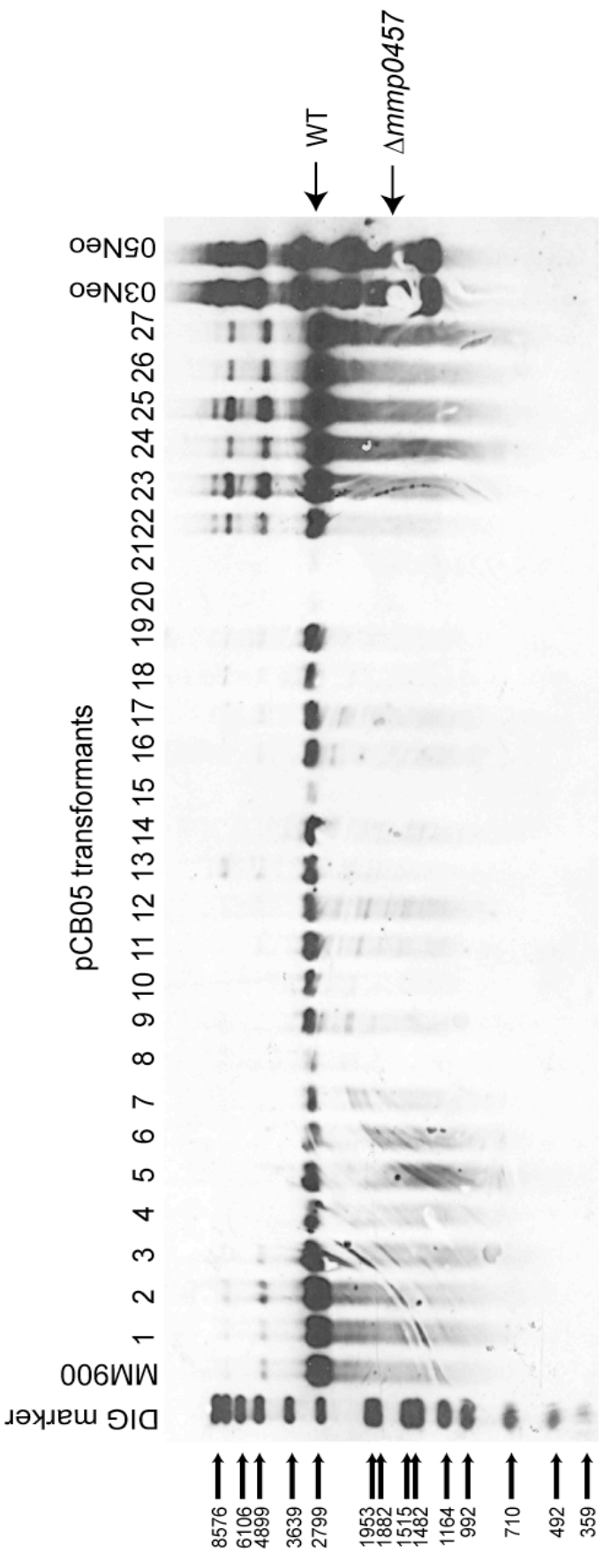
6.2.2 Expression of *mmp0457* in *M. maripaludis* using expression vector pAW42

6.2.2.1 *Construction of expression vector*

An outline of the cloning and transformation method for *in vivo* expression of *mmp0457* is shown in Figure 6.7. 6-Histidine-tagged protein was over-expressed from plasmid (pCB07) (Figure 6.7) to observe the physiological effect on cell size and DNA content in *M. maripaludis*. The *mmp0457* gene was amplified from *M. maripaludis* S2 genomic DNA by PCR using primers Mmp0457startAsc1 and Mmp0457stopAsc1 containing *AscI* restriction sites (see appendix A for primer sequence). The PCR product was cloned into cloning vector pGEM-T, restriction digested using *AscI* and gel purified. The purified products were cloned into the *AscI* site of the multiple cloning site pAW42 (Walters *et al.,* 2011) to generate an over-expression plasmid, pCB07 (Figure 6.8). pAW42 contains the puromycin resistance gene for selection of transformants in *M. maripaludis* S2.

P1 →
Mmp0457
← P2

M. maripaludis mmp0457 from genomic DNA was amplified using PCR.

0457
pCB06

The product was cloned into pGEM-T. Then digested using restriction enzyme *AscI*. The digested product was gel purified.

AscI   Asc1
0457
pCB07
amp   pur

The ends of the digested product were ligated to expression vector pAW42. This plasmid will be used to transform into *M. maripaludis* cells and the transformants will be seleted for growth on Puromycin. The protein expression will be tested by western blot.

Protein purification and Western blots

Figure 6.7. Schematic representation of generation of expression plasmid pCB07 for over-expression of mmp0457 in S0001 strain of M. maripaludis.
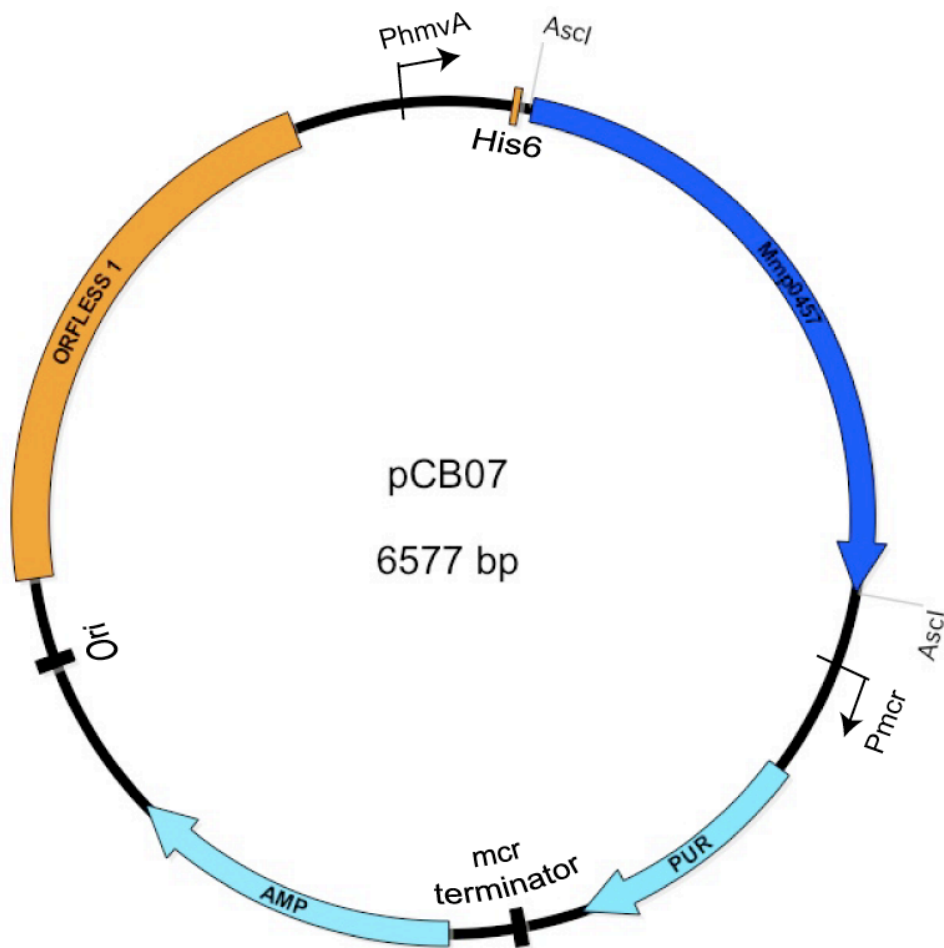
Figure 6.8. Map of the mmp0457 expression plasmid, pCB07. *mmp0457* was cloned into the *AscI* site of pAW42 (Walters *et al.,* 2011). This plasmid can replicate in the ORF1-containing S0001 strain of *M. maripaludis* and expressed His-tagged Mmp0457.

6.2.2.2 Transformation of expression vector

pAW42 requires the presence of the ORF1 sequence for maintenance in the cell, thus, pCB07 was transformed into *M. maripaludis*, S0001 strain which has ORF1 integrated into the *upt* gene (Walters *et al.,* 2011). 5 μg of plasmid DNA was transformed using a PEG-based transformation method (Materials and methods section 2.5.2). Transformants were resuspended into fresh medium and allowed to recover overnight without selection. The transformants were selected by growing the cells on puromycin plates and then sub-cultured in fresh liquid media with puromycin. The transformed cells were maintained on puromycin to prevent plasmid loss.

6.2.2.3 Expression of mmp0457 in M. maripaludis

6.2.2.3.1 Large-scale growth of expression strain (S0001 + pCB07) and wild-type (S0001) cells in fermenter

In order to compare the growth rate of the strain overexpressing mmp0457 (S0001+pCB07) with wild type cells (S0001) and to utilize flow cytometry analysis at various stages of growth, a large volume of cells was required. To achieve this, the strains were grown in a 2.5 L anaerobic fermenter (see section 2.2.4). 25 ml of log-phase cell culture was used to inoculate the fermenter medium and OD_{600} was measured every 2 h. The resulting growth curves for wild-type (S0001) and overexpression strain (S0001+pCB07) are shown in Figure 6.8a. A comparison of the growth curves shows a significant difference in the growth rate and the length of lag phase between the two cell types. S0001 reached OD_{600}=0.17 in 25 h whereas S0001+pCB07 cells took 50 h to reach a similar $OD_{600.}$ In addition, S0001 reached stationary phase at OD_{600}=1.3, where the strain overexpressing mmp0457 (S0001+pCB07) reached a higher OD_{600}=1.86 before entering stationary phase. A decrease in OD_{600} was observed for both the strains after reaching stationary phase suggesting a decrease in cell numbers due to lysis. The doubling time of the two strains was calculated by plotting logarithmic growth curve (Figure 6.8b). The S0001+pCB07 strain grew slowly, with a doubling time of 6.25 h compared to 2.5 h doubling time of wild-type (S0001) strain.

In addition to OD_{600} measurements the cells were harvested anaerobically from the fermenter, fixed and stored for flow cytometry analysis at 4°C (Materials and methods section 2.2.5).

a)

b)
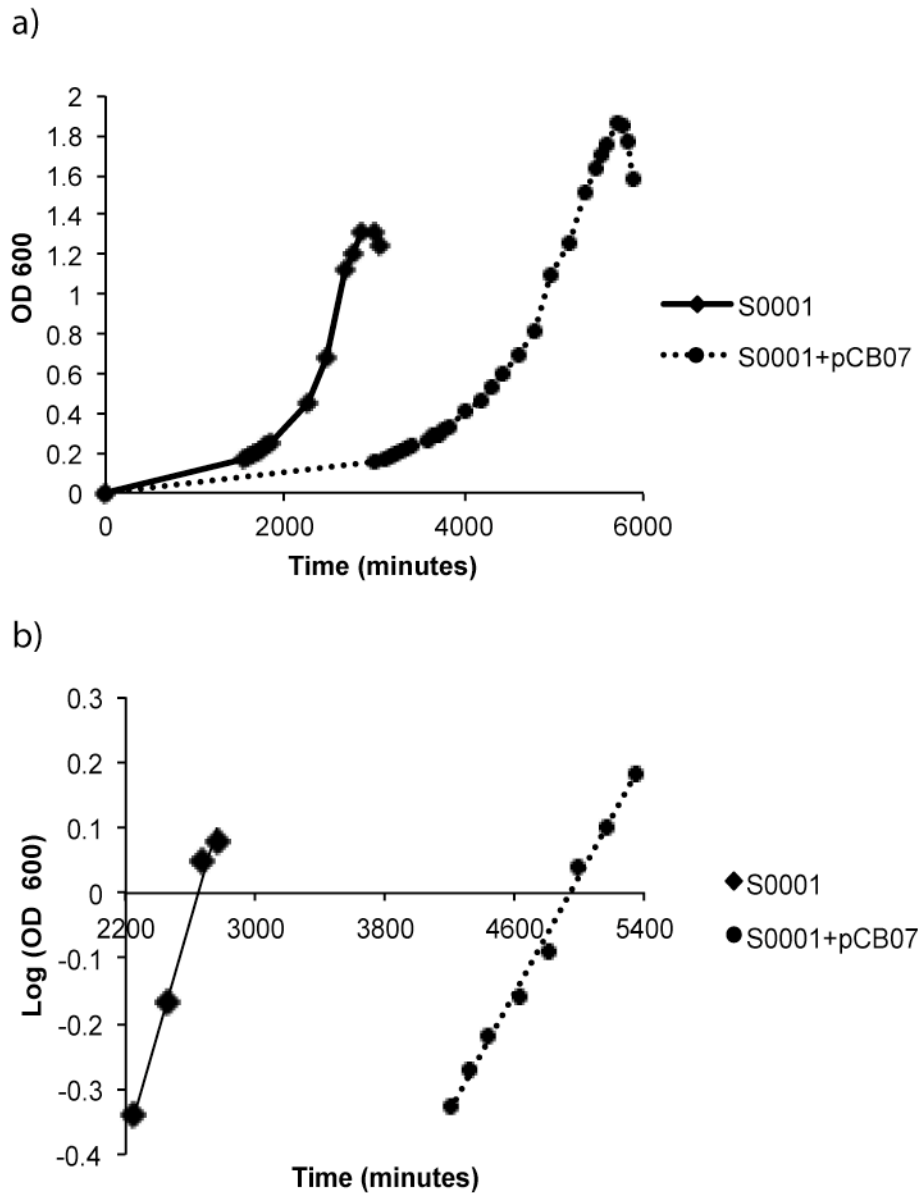
Figure 6.9. The overexpression strain (S0001 + pCB07) displays a longer lag phase (50 h compared to 25 h for wild-type) but reached stationary phase at much higher O.D. (a) Growth curve of wild type (S0001) and *mmp0457* overexpression strain (S0001 + pCB07) (b) The logarithmic growth curve showing doubling time of wild type cells is 2.5 h as compared to 6.5 h of overexpression strain.

6.2.2.3.2   Restriction digestion analysis of transformants

To confirm that pCB07 was present in the tube cultures and fermentor, plasmid purification was carried out from S0001+pCB07 tube culture cells and S0001+pCB07 fermenter culture cells. Plasmid purification was also carried out on S0001 tube culture cells as a negative control. The plasmids were digested with *AscI* and the digests were separated by gel electrophoresis. Stock pCB07 plasmid was digested and run on the gel as a positive control for the plasmid and insert size. The gel showed that the S0001 + pCB07 tube and fermenter culture cells possess copies of pCB07 as the plasmid 4952 bp and the *mmp0457* insert at the size of 1617 bp (Figure 6.9a). The plasmid or insert was absent in S0001 culture cells.

6.2.2.3.3   Western blot analysis for expression of *mmp0457* in S0001 + pCB07

To confirm that the his-tagged Mmp0457 was being expressed, western blots were carried out on the Mmp0457 elution fraction from S0001+pCB07 fermenter culture cells. Blots were also carried out on S0001 fermenter cells as negative control. The blot was probed with anti-His antibodies and developed (see section 2.14).

The expected full-length protein band was at size 61 kDa in the elution fraction, but the western blot analysis showed the presence of an intense band at 39 kDa in the elution fraction of overexpression strain (Figure 6.9b). This band was not visible in the S0001 elution fraction, consistent with the expression of His-tagged protein from plasmid pCB07 (Figure 6.9b). The presence of a low molecular weight protein band might be an indication of a posttranslational modification or that the protein is unstable in cell extracts. The experiment needs to be repeated with western blot analysis of whole cell extract and purified Mmp0457 over a time-course to study the protein degradation, if any. Also, a mass spectroscopy analysis of the protein is required to confirm that the protein thus purified is Mmp0457.

There was also a band between 15 and 20 kDa in both wild-type and overexpression strain (Figure 6.9 b). This protein has been previously identified as CbiX, a histidine rich protein involved in vitamin B12 synthesis (Hendrickson *et al.*, 2004, Xia *et al.*, 2009,). The protein contains 2 strings of 5-6 Histidine residues at its C- terminus and therefore interacts with Ni-NTA resins and anti-His antibodies (Protein id: NP_987284.1, http://www.ncbi.nlm.nih.gov/protein).
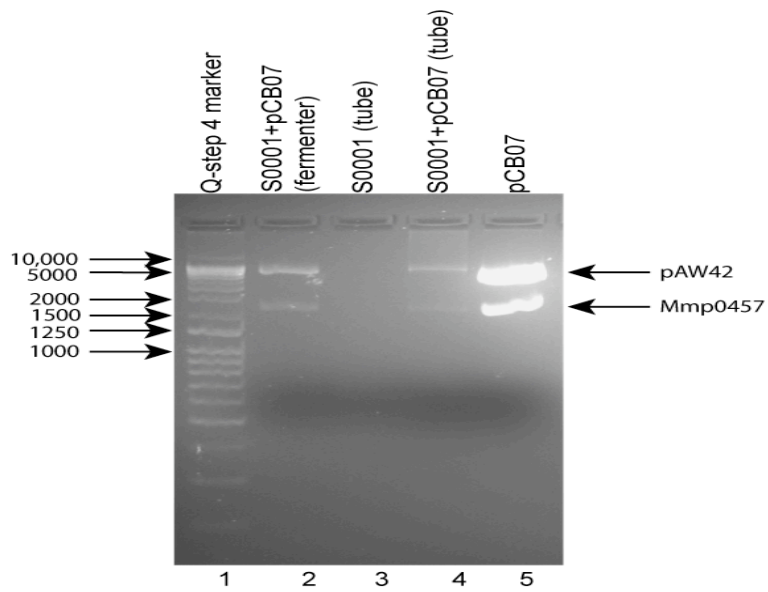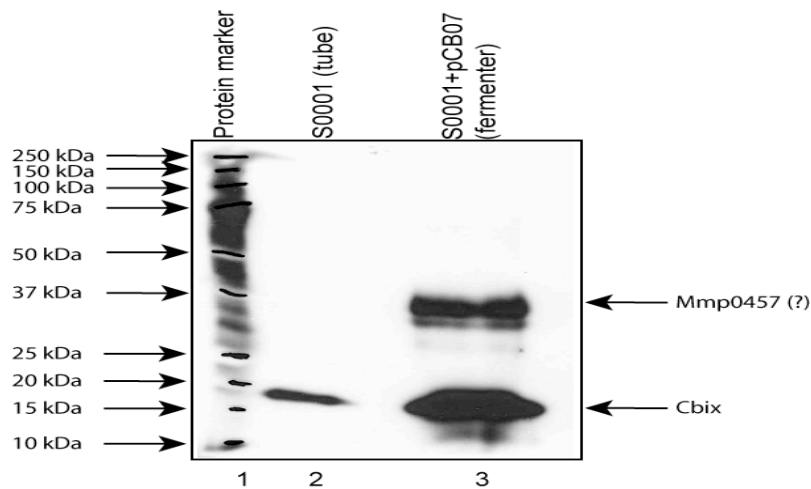
Figure 6.9. The overexpression strain (S0001+pCB07) expresses mmp0457. (a) Restriction digestion (AscI) of plasmids extracted from S0001 (tube culture), S0001+pCB07 (tube and fermenter cultures), lane 1 contains DNA ladder Q step 4, lane 2-4 contains digested plasmid from S0001+pCB07 fermenter cells, S0001 (negative control) and S0001+pCB07 tube culture respectively and Lane 5 contains digest of stock pCB07 as a positive control. The plasmid pCB07 was present in both the fermenter and tube cultures of S0001+pCB07. (b) Western blot showing expression of mmp0457 in overexpression strain (S0001+pCB07). Lane 1 contains protein molecular weight marker, lane 2 contains protein extracted from wild type and lane 3 contains protein isolated from the overexpression strain.

6.2.2.3.4   Phenotypic analysis of S0001 + pCB07 strain

Flow cytometry was carried out on S0001+pCB07 cells to identify whether the overexpression of *mmp0457* has altered the cells as compared to the wild type, S0001. The cells were harvested at different stages of growth and fixed (see section 2.2.5). The cells were stained with mithramycin/ethidium bromide stain just before flow cytometry analysis. Cell size was analysed for 100,000-500,000 cells per sample. Comparisons of the dot plot profiles of DNA content (FL3) and cell size (LS2) for the two strains were carried out. The cell size distribution in lag phase and early log phase (OD$_{600}$ 0.28) of the growth curve of S0001+pCB07 was very similar to that observed in wild-type cells. A small distinct population of cells (10%) with larger cell size and low DNA content (shown by arrow Figure 6.10) was distinctly visible from mid log phase (OD$_{600}$ 0.47) and was observed throughout the growth curve thereafter. The appearance of a distinct population suggests that the overexpression of Mmp0457 possibly affects S-phase in cells, which leads to a lower average DNA content in the cell population. Nevertheless, no distinct genome peaks were observed in the graphs showing the log of DNA content against the cell numbers (Figure 6.11). The cell cycle DNA distribution pattern was similar to studies carried out on closely related genus *M. jannaschii* (Maisnier-Patin *et al.,* 2002, Malandrin *et al.,* 1990) and MM900 strain of *M. maripaludis* S2 in deletion studies of MCM proteins (Dr. Alison Walters, PhD thesis).

The absence of distinct genome peaks also suggest multiple rounds of DNA replication occurring without cell division, as observed in certain bacteria (Cooper and Helmstetter, 1968). Studies have shown the presence of as many as 50 genome copies per cell in *M. maripaludis* (Hildenbrand *et al.,* 2011) and 3-15 genome copies in *M. jannaschii* (Malandrin *et al.,* 1999) suggesting asynchronous replication initiation of the different genome copies, which may result in the DNA distribution as observed in Figure 6.11. However, some genome peaks were observed in *M. jannaschii* when the cells stopped dividing in the stationary phase, and not observed in the flow analysis of *mmpMCM* deletion studies in *M. maripaludis* cells (Dr. Alison Walters, PhD thesis, 2010) and in the present study, which might be a characteristic of *M. maripaludis*.

One possible method to target this problem could be the synchronization of cell culture using nutrient (H_2) starvation (Morgan *et al.,* 1997). The synchronized cells will start the cell cycle in the same phase and will possibly form separate peaks showing DNA distribution and thus, a comparison of wild-type and overexpression strains may show an increase or decrease of certain distinct DNA peaks. In addition, a further microscopic analysis of the cells overexpressing the protein might help in understanding the type of defect caused.
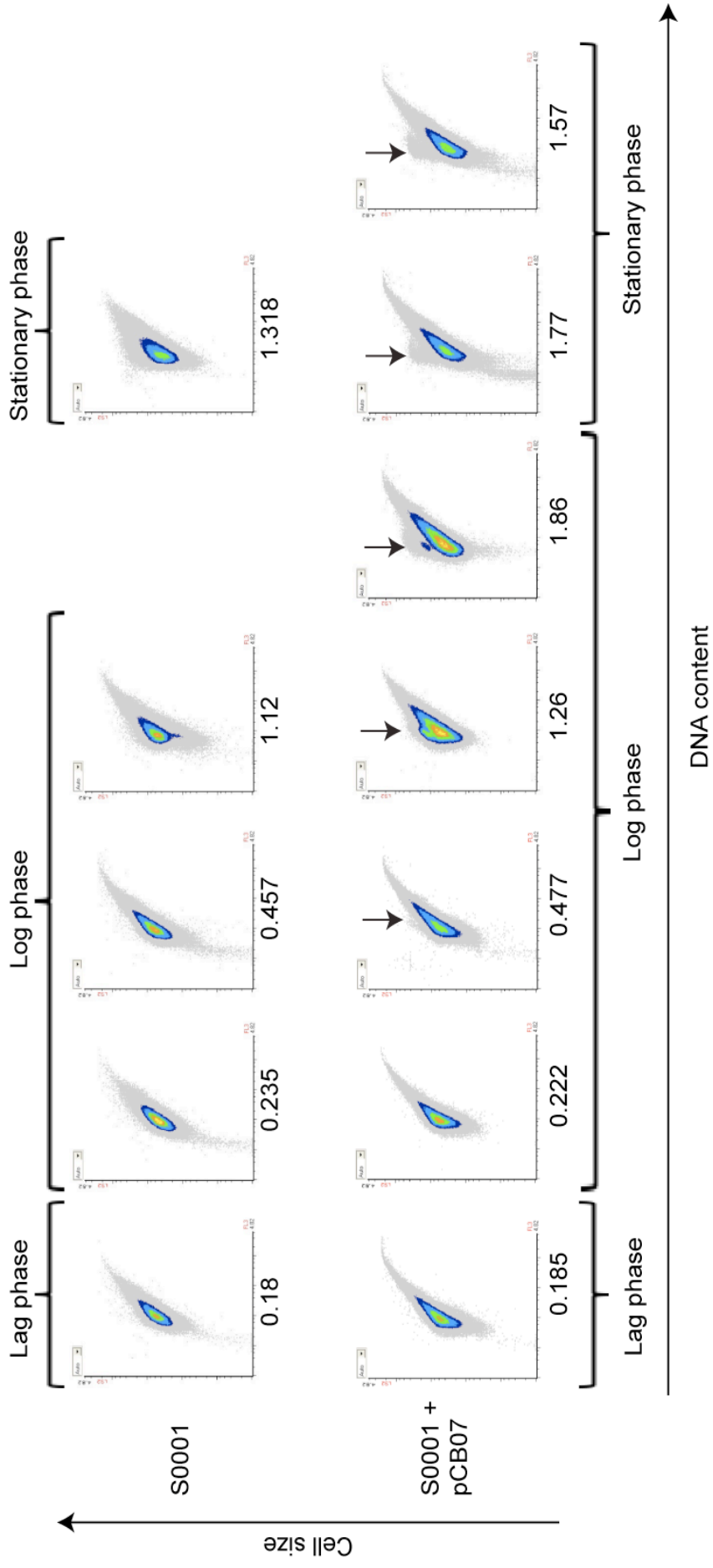
Figure 6.10. *The overexpression strain shows a small population (10%) of cells with larger cell size and relatively low DNA content.* The graphs show cell size (LS2) on Y-axis and DNA content on X-axis. The cell size comparisons were carried out between the two strains at similar O.D. 600. Optical densities for each sample are shown on the right of each pair of graphs (S0001 above S0001+pCB07). Optical densities for each sample are shown below the graphs. Both the strains show similar cell size and DNA distribution pattern, but from early log phase the overexpression strain showed an additional small population (10%) of cells with larger cell size and relatively low DNA content (shown by arrow). The appearance of a distinct population suggests that the overexpression of Mmp0457 possibly affects S-phase in cells, which leads to a lower average DNA content in the cell population.

212
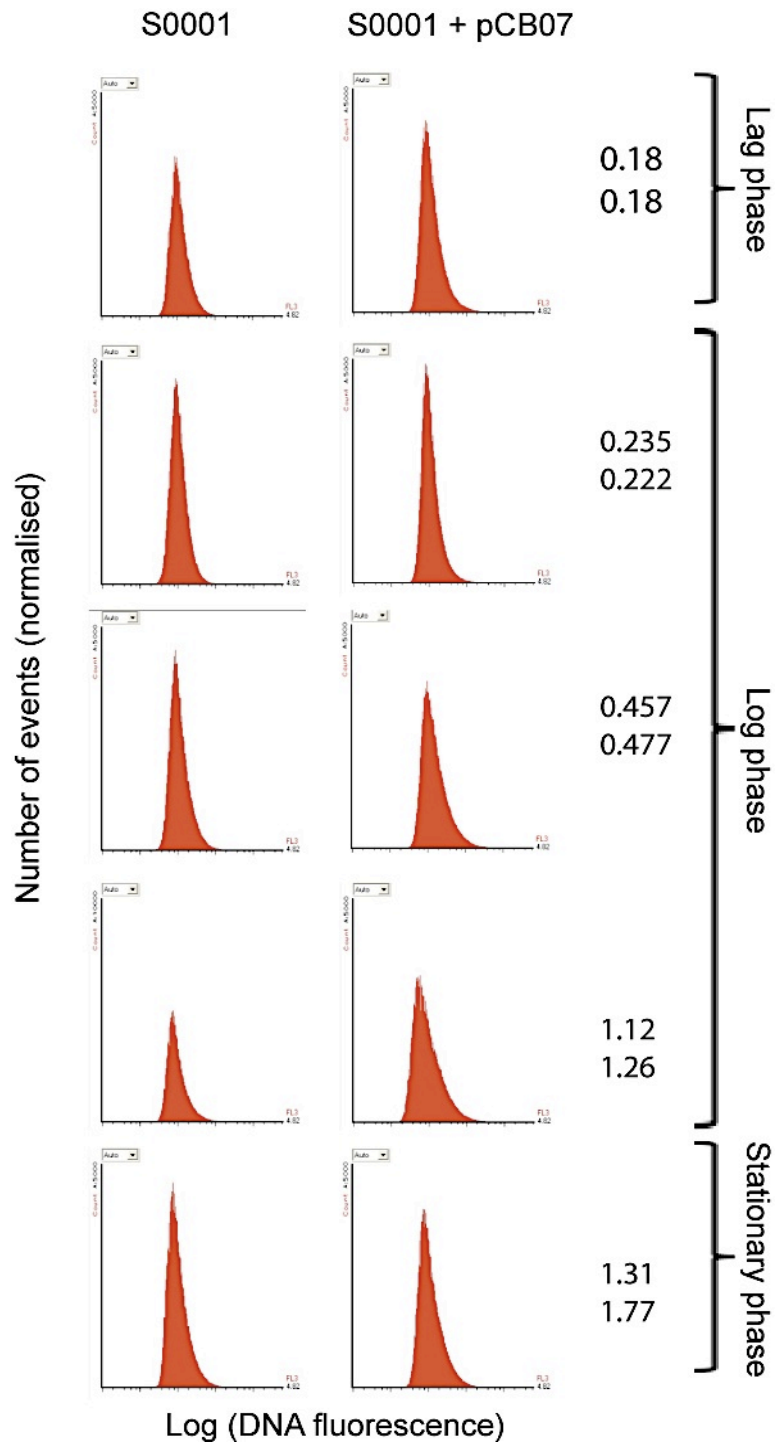
Figure 6.11. Flow cytometry analysis of S0001 and S0001+pCB07. The histograms have log DNA fluorescence on X-axis and cell numbers on Y-axis. Optical densities for each sample are shown on the right of each pair of graphs (S0001 above S0001+pCB07).

213

6.3   Discussion

Studies on model organisms have shown that a minimum number of DEAD-box proteins are required in eukaryotes however there is no such requirement in bacteria and archaea (Rocak and Linder, 2004). 17 out of 25 DEAD-box proteins were found to be essential in *S. cerevisiae* (de la Cruz *et al.,* 1999, Linder *et al.,* 2000), whereas none of the five DEAD-box proteins are essential in *E.coli* as individual deletion mutants and double deletions can be produced for all the proteins (Iost *et al.,* 2006). Gene deletion studies on RNA helicases have not been carried out in archaea. *M. maripaludis* contains two DEXD/H box proteins Mmp0457 (DEAD) and Mmp1284 (DEVH) (Hendrickson *et al.,* 2004). In this study, we show for the first time that like eukaryotes the presence of the DEAD-box RNA helicase is essential in at least one euryarchaeon *M. maripaludis*. In addition, although *M. maripaludis* and *M. thermautotrophicus* are very different organisms, both possess only one copy of the homologous DEAD-box helicase (Chapter 3), and it might be possible to extrapolate the results of the Mmp0457 homologue Mth203 in *M. thermautotrophicus*, suggesting that Mth203 may also be an essential DEAD-box protein. Further, flow cytometry analysis was carried out to examine the morphological changes and it was observed that the cells were found to be bigger in size and contained lesser DNA. It is most likely that this physiological consequence might be due to overexpression, discussed in further detail in Chapter 7, section 7.2.3.

7   Discussion

The discovery of archaea by Carl Woese established these organisms as the third domain of life (Woese and Fox, 1977). In the 1990s, the availability of sequenced genomes provided a new approach to study and investigate the similarities and differences between the three domains of life (Fleischmann *et al.,* 1995, Brown and Doolittle, 1997). It was observed that archaea and eukaryotes share striking similarity in the protein machinery that regulates cellular mechanisms such as DNA replication (Keeling and Doolittle, 1995, Doolittle and Logsden, 1998, Cann and Ishino, 1999, Kelman and Kelman, 2003, Grabowski and Kelman, 2003). In addition, the archaeal machinery is much simplified and thus poses a biochemical advantage to study complex processes. DNA replication in bacteria and eukaryotes has been extensively studied, although much of the detail is still unknown. On the other hand, very little information is available regarding archaeal DNA replication. For instance, bacteria-like single origins of replication (Myllikalio *et al.,* 2000, Berquist and Dassarma, 2003) and eukaryote-like multiple origins of replication (Robinson *et al.,* 2004, Lundergen *et al.,* 2004) have been found in archaea. Homologues of eukaryotic DNA replication proteins like MCM, Cdc6, PCNA, RPA and Polα have been identified in archaea (Grabowski and Kelman, 2003). However, other components of DNA replication such as the helicase loader and other associated proteins have not been identified in archaea. A study to identify the proteins that interact with replication proteins such as MthMCM and MthCdc6-1 in *M. thermautotriophicus* revealed that Mth203 forms stable interactions with MthCdc6-1 in a yeast-two-hybrid assay and protein pull-down assay (Dr. Richard Parker, PhD thesis, 2006). However, the function of Mth203 and its role in DNA replication, if any, is unknown.

Therefore, the aim of this study was the functional characterization of Mth203 and its role in the DNA replication of *M. thermautotrophicus*.

7.1   Mth203 is a DEAD-box RNA helicase

There are three putative Superfamily II DEXD/H-box RNA helicases encoded by the *M. thermautotrophicus* genome (Smith *et al.,* 1997). Sequence analysis classified

Mth203 as a DEAD-box RNA helicase (Fairman-Williams *et al.,* 2010). This work demonstrates that Mth203 possesses RNA-independent ATPase activity, bi-directional non-specific RNA unwinding activity that requires the presence of ATP and divalent cations (Mg^{2+}/Mn^{2+}) (Chapter 5). This observation is in conjunction with the RNA helicase activity observed for many DEAD-box proteins (Gorbalanya and Koonin, 1993, Rocak and Linder, 2004, Cordin *et al.,* 2006).

Both sequence and structural analysis (Chapter 3) revealed that Mth203 shares homology with MjDEAD, a DEAD-box protein and putative RNA helicase in *M. jannaschii* and eIF4A, an important RNA helicase involved in the initiation of protein translation in eukaryotic cells (Story *et al.,* 2001, Benz *et al.,* 1999). The protein contains all the 12 conserved motifs characteristic of DEAD-box proteins (Cordin *et al.,* 2006, Hilbert *et al.,* 2009). MjDEAD and eIF4A proteins are exceptions as their structures consist of an isolated helicase core, whereas in other helicase proteins the helicase core provides basic DEAD-box protein function and the N- and C- terminal extensions modulate the activity. It is proposed that the DEAD-box proteins functionally interact with polynucleotides (RNA) via the helicase core and the flanking domains can contribute towards the affinity/specificity of the binding (Hilbert *et al.,* 2009). The C- terminal domains of the splicing helicases CYT-19 and MSS116 mediate interactions with structured RNA (Grohman *et al.,* 2007, Mohr *et al.,* 2008). The C- terminal domains of DbpA and YxiN interact with the hairpin in ribosomal RNA and assists unwinding (Diges and Uhlenbeck, 2001, Wang et al, 2006, Karginov *et al.,* 2005, Tsu *et al.,* 2001, Kossen *et al.,* 2002). The C- terminal domain of the Hera protein provides high affinity for ribosomal RNA and RNase P RNA (Morlang *et al.,* 1999, Linden *et al.,* 2008).

A comparison of the full-length homologues of MjDEAD and Mth203 in methanogens has revealed that all the RNA helicases including Mth203 have a longer C- terminal, which might serve a modulatory function in the activity of Mth203. The presence and absence of the C- terminal of Mth203 has a profound effect on the oligomerization status of the protein. The structural analysis presented here has demonstrated that the full-length protein is a dimer and exists as a monomer in the absence of the C- terminal domain (Chapter 3). It was not possible to identify any functional advantages conferred by the C- terminal in the present study. The

absence of C- terminal did not have any effect in the DNA helicase assays, however, a difference in the DNA binding affinity was observed between the two proteins (Mth203 and Mth203ΔC53).

There have not been any DNA binding studies on DEAD-box RNA helicases. Although DNA unwinding activity has been observed in several helicases belonging to this group. For example Dbp9, an RNA helicase in *S. cerevisiae* unwinds RNA:DNA hybrid and DNA duplexes, DP103 and ATDRH1 show dsRNA unwinding and DNA-RNA helicase activities (Kikuma *et al.,* 2004, Yan *et al.,* 2003, Okanami *et al.,* 1998). In the present study we have characterised for the first time that Mth203 and Mth203ΔC53 have higher binding affinities for origin sequences (ORB) when compared with random DNA sequences (Chapter 4). Such preferential binding is usually observed in proteins involved in initiation of DNA replication (such as ORC, Cdc6 in eukaryotes, DnaA in bacteria), or regulation of DNA replication (such as SeqA in bacteria) (Hwang *et al.,* 1990, Wiegel *et al.,* 1997). Although the dissociation constant of Mth203-DNA binding (K_d 560 nM) is 100-fold higher than MthCdc6-1-DNA binding (K_d 3 nM), it is possible that Mth203 interacts with other co-factors, which might increase its affinity towards DNA. A similar effect is observed in the case of eIF4A whose ATP and RNA binding affinity is increased in the presence of co-factors eIF4B, eIF4G and eIF4H (Bi *et al.,* 2000, Rogers *et al.,* 2001). Hence, it cannot be ruled out that Mth203 might be involved in origin recognition and binding.

7.2 What is the function of Mth203?

Protein pull-down assays and western blots were carried out to identify proteins interacting with Mth203 *in vitro*. Non-specific interactions were prevented with the use of a high-salt wash buffer before eluting the proteins. Results have shown that Mth203 interacts with MthCdc6-1 (replication initiation protein), Mth203, pyruvate kinase, phosphopyruvate hydratase, isopentenyl pyrophosphate isomerase, 3-chlorobenzoate-3,4-dioxygenase dihydrogenase related protein, 50S ribosomal proteins (L3P, L4P, L18P) and 30S ribosomal proteins (S4e, S34e, S7P, S13P). However, it is highly unlikely that all the interactions observed are biologically meaningful. The enzymes like pyruvate kinase, phosphopyruvate hydratase, isopentenyl pyrophosphate isomerase, 3-chlorobenzoate-3,4-dioxygenase

dihydrogenase related protein are involved in vital biosynthetic pathways and may interact with Mth203 to couple some cellular function with energy production or more likely these proteins are pulled down in the elution fractions as they are abundant in cytoplasm. The co-elution of DNA replication protein and ribosomal proteins, suggests that perhaps Mth203 has more than one function in *M. thermautotrophicus*: 1) RNA helicase and 2) DNA replication/ regulation. There are a few proteins in eukaryotes, which are involved in DNA replication and transcription. For example, CTF/NF-I proteins in eukaryotes can serve both as a transcription selectivity factor for RNA polymerase II and as an initiation factor for adenovirus DNA replication (Jones *et al.,* 1987). The yeast origin recognition complex functions in transcription silencing and DNA replication (Bell *et al.,* 1993). Yph1P protein (an essential protein with BRCT domain) is involved in ORC binding, cell proliferation and ribosomal biogenesis (Du and Stillman, 2002, Oeffinger, 2002).

Perhaps Mth203 is part of a transcription initiation complex or translation machinery and interacts with MthCdc6-1 and MthMCM. In one study the MCM complex was observed to interact with RNA polymerase (Holland *et al.,* 2002) and MCM can unwind DNA ahead of RNA polymerase (Snyder *et al.,* 2005).

7.2.1 DNA replication/DNA regulation

Interactions between Mth203 and MthCdc6-1 were observed in yeast-two hybrid assay, affinity co-purification (Dr. Richard Parker, PhD thesis, 2006), western blots and fluorescence anisotropy assays (Chapter 4). All these observations seem to suggest that there is a specific and stable interaction between the replication initiation protein MthCdc6-1 and Mth203. Hence, an important question arises concerning whether Mth203 interacts with MthCdc6-1 at the origin or away from it?

MthCdc6-1 is a DNA initiation protein and thus specifically interacts with origin sequences (Majernik and Chong, 2008, Capaldi and Berger, 2004). Mth203 binds non-specifically to short DNA sequences but shows a preferential binding to long sections of DNA containing ORB sequences (Chapter 4). It is possible that MthCdc6-1 recruits Mth203 to the origin to carry out a specific function. The open-complex for initiation of DNA replication is formed by cooperative binding of initiation proteins at the origin leading to spatial unwinding of DNA at the origin

(DnaA in bacteria, ORC/Cdc6 in eukaryotes). In *M. thermautotrophicus*, MthCdc6-1 binds to the origin of replication but the process that leads to the formation of the open-complex is yet unknown (Capaldi and Berger, 2004, Majernik *et al.,* 2008). It is thus possible that Mth203 acts as a cofactor that assists MthCdc6-1 in the formation of the open-complex. A few DEAD-box helicases like Dbp9p in *S. cerevisiae* and Vasa protein in *D. melanogaster* have shown DNA binding activity (Liang *et al.,* 1994, Kikuma *et al.,* 2004). Dbp9p also shows dsDNA and RNA:DNA hybrid unwinding and is considered as an exception in the DEAD-box protein family (Kikuma *et al.,* 2004). PDH45 and PDH47 proteins in *Pisum sativum* possess ATP binding, and ATP-dependent DNA and RNA unwinding activities (Pham *et al.,* 2000). However, thorough studies of these proteins have shown that the protein is involved in translation and immuno-depletion of PDH45 and PDH47 leads to inhibition of protein synthesis (Sanan-Mishra *et al.,* 2005, Vashist *et al.,* 2005 a,b).

Some viral proteins belonging to the closely related DEXD/H family are involved in initiation of DNA replication and origin recognition. One such example of DNA replication initiation is observed in Herpes simplex virus 1 (HSV-1) where replication initiation is carried out by the UL9 protein, a DEAX/H box helicase (Malik *et al.,* 1992, Martinez *et al.,* 1992, Marintcheva and Weller, 2003). UL9 binds and unwinds the replication origins of the virus and is the limiting component of DNA replication in HSV-1 (Malik *et al.,* 1992, Martinez *et al.,* 1992). The NS3 protein of Hepatitis C virus, is another DEXH/D box helicase involved in the replication of the RNA virus (Iwai *et al.,* 2011). The protein displays both RNA and DNA unwinding activities (Pang *et al.,* 2002). The role of the DNA unwinding activity displayed by NS3 is still unknown.

It is possible that Mth203 is involved in DNA replication regulation by a) sequestering MthCdc6-1 or b) sequestering of DNA origin sequences and hence prohibiting MthCdc6-1 access to the origin of replication. Geminin protein in eukaryotes sequesters Cdt1, a replication initiation protein, as a mechanism to prevent re-initiation of DNA replication (Saxena and Dutta, 2005).

In *E.coli* Hda protein causes regulatory inactivation of DnaA (RIDA) where interaction of ADP-Hda with ATP-DnaA is mediated by the DNA polymerase-loaded clamp and causes hydrolysis of ATP-DnaA to ADP-DnaA catalysed by Hda (Figure 1.5)

(Su'etsugu *et al.,* 2008). In many archaea and eukaryotes, MthCdc6-1 possesses autophosphorylation activity, which is inhibited in the presence of DNA (Grabowski and Kelman, 2001). This could be one of the regulatory mechanisms for activation of MthCdc6-1 for initiation of DNA replication. In the present study, when Mth203 was added to the reaction, the MthCdc6-1 autophosphorylation activity was re-activated, suggesting perhaps Mth203 competes with MthCdc6-1 for interaction with DNA and hence might be involved in DNA replication by sequestering of origin sequences. Such a role might be reminiscent of SeqA protein in bacteria, which competes with DnaA for *oriC* (Kitagawa *et al.,* 1998, Messer, 2002).

Helicases are nucleic acid–dependent ATPase motor proteins capable of unwinding DNA or RNA duplex substrates. The SF2 putative helicases contain a series of conserved motifs, which are characteristic of proteins that are able to move directionally along nucleic acid strands, known as translocases (Rocak and Linder, 2004). These "putative helicases" do not unwind nucleic acids but instead couple NTP hydrolysis to directional motion (translocation) along nucleic acids. And the protein motors may or may not harness the translocation to drive the unwinding of duplex into single strands (Singleton *et al.,* 2007, Erzberger and Berger, 2007). For double-strand translocases, the antiparallel phosphodiester backbones give the duplex an intrinsic symmetry, resulting in DNA binding but no unwinding activity (Singleton *et al.,* 2007). Mth203 shows DNA binding activity but strand separation activity was not observed in the DNA helicase assay (Chapter 5). This probably suggests that Mth203 may translocate along dsDNA instead of unwinding dsDNA. There are many SF2 helicases, which possess translocase activity and are involved in DNA repair activity (Tuteja and Tuteja, 2004). It is possible that instead of DNA replication, Mth203 could be involved in DNA repair. UvrB, a SF2 helicase is a part of multienzyme nucleotide excision repair (NER) complex UvrABCD. UvrA and B form a complex and translocate on DNA to detect the site of DNA damage, UvrB complexes with UvrC and cleaves the nucleotides on either side of damage and finally UvrD removes the excised segment by localised unwinding (Koo *et al.,* 1991). Other DNA repair SF2 helicase enzymes include Rad3 (Sung *et al.,* 1987), Rad15 (Berneburg *et al.,* 2000), human helicase ERCC2 (Weber *et al.,* 1990), *E.coli* DinG (Voloshin and Otero, 2007), Rad54 (Gorbalenya and Koonin, 1992, Tuteja and Tuteja, 2004).

Perhaps Mth203 has both DNA translocation and unwinding activities activated in the presence of another cofactor. This is observed in other SF2 helicases. It has been shown that RuvB, a SF2 helicase, acts as an ATP-dependent helicase in the presence of RuvA, where a substrate consists of a short single-stranded DNA fragment annealed to its complementary sequence in a long single stranded DNA molecule (Tsaneva *et al.,* 1993). RuvB shows very little or no DNA binding in the absence of RuvA (Muller *et al.,* 1993). Eukaryotic translation initiation factor eIF4A requires eIF4B for RNA unwinding activity (Rozen *et al.,* 1990), *E. coli* UvrB requires UvrA for translocating along DNA and excising damaged DNA (Oh *et al.,* 1989), herpes simplex virus UL5 requires UL52 (Dodson *et al.,* 1991), and herpes simplex virus UL9 requires ICP8 for recognition of origin sequences and DNA unwinding (Boehmer *et al.,* 1993). In another study, RecG and PriA (both SF2 helicases) were shown to restart DNA replication by loading DnaB (replicative helicase) at a stalled forked (McGlynn and Lloyd, 2000). RecG protein is involved in the formation of a Holliday junction from a stalled replication fork by regression of the fork so that DNA damage repair systems such as RuvABC/RecBCD may act (Gregg at al., 2002).

In an alternative scenario, Mth203 might be involved in chromatin remodelling by unwinding secondary structures such as hairpins, often formed under supercoiled structures. Such a function is displayed by SWI/SNF proteins belonging to the Snf-2 class of the SF2 family of RNA helicases (Becker and Horz, 2002). These proteins are involved in ATP-dependent nucleosome remodelling as a means of transcription regulation (Vignali *et al.,* 2000).

The inhibition of DNA unwinding activity conferred by Mth203 is similar to an observation made with the *E.coli* helicase loader DnaC, which inhibits DnaB helicase activity on an artificial substrate (Wahle *et al.,* 1989). In a previous study on *M. thermautotrophicus* proteins, MthCdc6-2 protein was suggested to function as a helicase loader (Shin *et al.,* 2008). The protein was shown to dissociate the MthMCM multimer and proposed to work as a ring dissociater in order for the loading of the hexamer at the origin (Shin *et al.,* 2008). However, the process by which the hexamer re-associates at the origin was not understood and hence, the presence of other cofactors in the process was proposed. Mth203 inhibits MthMCM helicase

activity, interacts with MthCdc6-1 and origin sequences suggesting Mth203 may play an important role in helicase assembly.

7.2.2   Is Mth203 a Cold-shock protein?

All RNA helicases are known to facilitate translational and ribosomal assembly (Tanner and Linder, 2001). One interesting observation in the BLASTp analysis of Mth203 was the presence of protein CLG_B2065 (*Clostridium botulinum*), which is annotated as a cold-shock protein. The transcriptional profiling of the hyperthermophilic methanorchaeon *Methanocaldooccus jannaschii* in response to non-lethal cold shock revealed that RNA helicase MjDEAD (homologue of Mth203) was the most up-regulated gene (Boonyaratanakornkit *et al.,* 2005). A homologue of Mth203 called DeaD is up-regulated in response to cold-shock in *Methanococcides burtonii* (Lim *et al.,* 2000). Interestingly, over-expression of a DEAD-box helicase is also observed in the psychrophile *Pseudoalteromonas haloplanktis* (Medigue *et al.,* 2005).

Genes encoding for certain small and large ribosomal subunits were also up regulated. Together this may result in the formation of a functional ribosome under cold-shock conditions. In *E.coli*, over-expression of DeaD and srmB DEAD-box proteins rescues cold-sensitive phenotypes by assembling 50S ribosomes (Iggo *et al.,* 1990). In yeast, inactivation of DEAD-box protein Dbp2p results in a cold-sensitive phenotype (Iggo *et al.,* 1991, Barta and Iggo, 1995). In cyanobacteria, crhC, a DEAD-box RNA helicase of unknown function is over expressed under cold-shock conditions (Chamot *et al.,* 1999).

The DEAD-box helicases are induced not only in cold-stress but in other abiotic stresses such as high salinity, heat, drought, radiation, nutrient loss, oxidative stress and pH. Since RNA is more prone to form stable secondary structures, for the proper functioning of cells, the presence of RNA chaperones, like DEAD-box helicases is essential (Vashist and Tuteja, 2006). DEAD-box helicases use ATP to unwind or remodel such secondary structures (Jones *et al.,* 1996). Some stress up-regulated DEAD-box helicases are summarised in the Table 7.1 (Vashist and Tuteja, 2006). Very few stress-induced DEAD-box helicases have been biochemically characterised so far. All the proteins have shown ATPase activity and RNA helix destabilization activity

(Vashisht and Tuteja, 2006). Crh protein in *E.coli* has also shown RNA helicase activity and is part of a multi-subunit complex most probably involved in translation (Yu *et al.,* 2000). In addition proteins like PDH45, PDH47 (*P. sativum*), are among the few DEAD-box proteins demonstrating DNA and RNA helicase activity, which is very unusual for DEAD-box proteins (Sanan-Mishra *et al.,* 2005, Vashisht *et al.,* 2005).

When *S. cerevisiae* is exposed to various abiotic stress conditions, the increased expression of eIF4A was observed (Owttrim, 2006). eIF4A is part of multi-subunit complex eIF4F and is responsible for removing secondary structure in the 5'UTR of mRNA (Owttrim, 2006). Also, PDH45 in *P. sativum* induced during cold-stress conditions interacts with DNA modifying enzymes such as topoisomerases (Vashisht *et al.,* 2005). Thus, DEAD-box helicases may help in cold-shock response by (1) transcription and translation to enhance protein synthesis, and/or (2) DNA associated multi-subunit complexes to alter gene expression (Vashisht and Tuteja, 2006).

Kato *et al.,* (2008) performed a comparative transcriptome analysis of *Methanothermobacter thermautotrophicus* in various environmental conditions. Although the protein was expressed under normal growth condition, the Mth203 transcript was observed to show highest increase in the cold stress conditions. Also, the proteins L4P, S46, L3P, L18P and S13P were increased in cold stress conditions. Further, in the pull-down assays both 50S and 30S ribosomal proteins and DNA replication protein MthCdc6-1 were co-eluted with Mth203 (Chapter 4) implying, a possible role of Mth203 in rescuing cells from cold-shock conditions by removing secondary structures of RNA or increased gene expression by interaction with DNA modifying enzymes.

Since Mth203 is the only DEAD-box protein in *M. thermautotrophicus* (Smith *et al.,* 1997), it might also be involved in other biochemical functions associated with DEAD-box RNA helicases such as translation initiation, gene expression, degradation of RNA molecules, ribonucleoprotein complex (RNP) remodelling, pre-mRNA splicing and ribosome biogenesis (Tanner and Linder, 2001, Rocak and Linder, 2004, Cordin *et al.,* 2006).

Table 7.1 Table summarizing stress up-regulated DEAD-box helicases and their functions

Gene	Organism	Type of Stress	Possible function
CsdA	*E.coli*	Cold-stress (37-15°C)	Helix destabilization protein (Jones *et al.,* 1996)
DeaD	*E.coli*	Cold-stress (37-15°C)	Ribosomal assembly (Iggo *et al.,* 1990)
Srm	*E. coli*	Cold-stress (37-15°C)	Ribosomal assembly (Schidt and Linder, 1991
CrhC	*Anabena sp. Strain PCC7120*	Cold-stress (30-20°C)	Translation (Chamot *et al.,* 1999, 2000)
13055	*Clostridium perfringes*	Oxidative stress	Overcoming oxidative stress (Briolat *et al.,* 2002)
CrhR	*Syncheocystis sp. PCC 6803*	Redox regulated	Transcription and translation (Kujat *et al.,* 2000)
DeaD	*Methanococcoides burtonii*	Cold-stress (23-4°C)	Unknown (Lim *et al.,* 2000)
Hera	*Thermus thermophiles*	Heat	Ribosome assembly, RNA processing (Morlang *et al.,* 1999)
TIF2	*S. cerevisiae*	Salt stress (lithium)	Translation initiation (Montero-Lomeli *et al.,* 2002)
eIF4A	*S. cerevisiae*	Various stresses	Transcription initiation (Sanan-Mishra *et al.,* 2005)
FL25A4	*Arabidopsis thaliana*	Cold-stress (4°C)	Unknown (Seki *et al.,* 2001)
Los4-1, Los4-2	*A. thaliana*	Cold-stress (22-4°C)	mRNA export (Gong *et al.,* 2002, 2005)
HVD1	*Hordem vulgare*	Salt and Cold-stress	Transcription regulation (Nakamura *et al.,* 2004)
PDH45	*Pisum sativum*	Salt stress	Translation and regulation of DNA/RNA metabolism (Pham *et al.,* 2000, Sanan-Mishra *et al.,* 2005)
PDH47	*P. sativum*	Salt and Cold-stress	Regulation of RNA metabolism (Vashist *et al.,* 2005)

7.2.3 *mth203* expression in the cell cycle

Very little is known about the expression pattern of DEAD-box helicases in the cell cycle of prokaryotes and eukaryotes. The cell division cycle is thus a complex self-regulating program, such that many genes involved in aspects of the cell cycle are also controlled by it. Progression through the cell cycle is controlled by expression of periodic and phase-specific defined set of proteins. Many cell cycle-regulated genes are involved in processes that occur only once per cell cycle like DNA synthesis, budding, and cytokinesis. Additionally, many of these genes are involved in controlling the cell cycle itself, although in most cases it is unclear whether their regulated transcription is absolutely required (Spellman *et al.,* 1998).

A cyclic expression of Mth203 was observed during *M. thermautotrophicus* growth. Rad54, a SF2 helicase in *S. solfataricus,* is expressed in a cyclic pattern through out the cell cycle, being predominantly expressed in G1 to repair chromosomal DNA in preparation of S-phase (Spellman *et al.,* 1998). p68 protein in *S. cerevisiae* is present in the nucleoplasm in interphase and is transiently associated with the nucleoli during late telophase, at the time when pre-nucleolar bodies are condensing after mitosis (Iggo *et al.,* 1991). p68 is involved in rRNA processing and probably required for nucleolar assembly. *mth203* cyclic expression suggests a role for Mth203 in the maintenance of cellular activities in the cell cycle. Perhaps the protein is involved in targeted gene expression by processes like transcription and translation.

Genetic manipulation of *M. thermautotrophicus* is not possible due to lack of adequate molecular biology tools. The physiological effects of Mth203 were studied by gene-deletion and overexpression studies of a Mth203 homologue (Mmp0457) in *M. maripaludis* (Chapter 6). *mmp0457* full-length gene deletion and C- terminal deletion were lethal for the cells, hinting that the proteins perform some crucial function for cell viability. Due to time constraints it was not possible to generate variation of mutations in the protein sequence, in order to study the protein function. Generation of mutations in p68/72 DEAD-box proteins in *D. melanogaster* resulted in a multitude of phenotypes that range from sterility to lethality (Buszczak *et al.,* 2006). Studies on p68/72 homologues (DDX5/17) in higher eukaryotes (Chick,

frog, mouse, rat, human) have also shown these proteins are important for cell growth, division and survival (Stevenson *et al.,* 1998, Seufert *et al.,* 2000, Ip *et al.,* 2000, Fukuda *et al.,* 2007). These proteins are associated with functions like RNA unwinding, RNA splicing, processing of microRNAs, pre-mRNA and pre-rRNA processing, transcriptional co-activators (Janknceht, 2010).

mmp0457 was overexpressed in *M. maripaludis* cells (Chapter 6) and resulted in a small population of large cells with low-DNA content. Such a phenotype may arise due to a number of problems associated with cell division, DNA separation, or regulation.

7.3   Summary

With the discovery of MthCdc6-1 and Mth203 interactions, Mth203 was an interesting protein with a potential new function in DNA replication. The broad aim of this project was to elucidate the biochemical functions of Mth203 and thus define its function in *M. thermautotrophicus*. Recombinant Mth203 was expressed and purified. Various activities assays like polynucleotide-unwinding assays, ATPase assays, have demonstrated that Mth203 is an ATP-dependent non-specific RNA helicase and also possesses DNA binding activity. In addition, experiments conducted with full-length protein and Mth203ΔC53 suggests that the carboxyl-terminal of the protein might play an important role in regulating the biochemical function of the protein. Genetic manipulation studies with mmp0457 display that the protein is essential for cell viability and a change in the cellular concentration can affect the cell morphology and DNA content. Mth203 interaction with ribosomal proteins suggests that the protein is involved in cellular processes like ribosomal assembly, transcription, RNP remodelling or translation. Furthermore, binding with MthCdc6-1 and inhibition of MthMCM helicase activity suggests Mth203 involvement in DNA replication/regulation.

8   Future Work

8.1   *Protein structure*

The DEAD-box helicases have variable amino- and carboxyl- terminal sequences but it has not been possible to find out the structure of these sequences. The flanking sequences potentially confer the specificity and affect the activity of the protein (Cordin *et al.,* 2006). In the present study, Mth203 was found to exist as a dimer in solution, and Mth203 flanking sequences were found to interact with MthCdc6-1 independent of the rest of the sequence (Dr. Richard Parker, PhD thesis, 2006). The structure of the protein and localization of the flanking sequences on the protein structure could not be determined because it was not possible to crystallize the Mth203 protein. However, spherolites (amorphous crystals) were obtained in some conditions. Further experiments could be carried out to crystallize Mth203 using modifications of the conditions (change in pH, salt concentration, temperature) where spherolites were obtained (Mcpherson, 2004). An alternative to crystallization would be the use of NMR to determine the 3D structure of C-terminus in solution and then overlapping it with the structure of protein-core threaded on MjDEAD structure (Grishaev and Llinas, 2002).

8.2   *Mth203-MthCdc6-1 interactions*

Anisotropy assays have revealed that Mth203 interacts with MthCdc6-1 and DNA. However, due to time constraints it was not possible to further characterize the interactions, leaving lots of unanswered questions such as what is the oligomeric status of the MthCdc6-1 and Mth203 complex? Does the interaction take place on DNA? Oligomerization of MthCdc6-1 and Mth203 and its assembly on DNA could be further studied by analysing the protein complexes through SEC-MALLS. Similar studies of the assembly of DNA binding complexes in solution have been carried out using SEC-MALLS (Ryan *et al.* 2008, Moullintraffort *et al.,* 2010, Newman, 2011). The complex formation might provide an insight into the molecular interactions taking place and determine whether Mth203 or MthCdc6-1 sequesters DNA or both MthCdc6-1 and Mth203 bind DNA in a complex.

Mth203ΔC53 shows decreased binding to MthCdc6-1 (Chapter 4) and the C-terminal of Mth203 was also found to independently interact with MthCdc6-1 (Dr. Richard Parker, PhD thesis, 2006). These interactions could be further characterized by using site directed mutagenesis of individual amino acids in the C- terminal domain and studying for protein:protein interactions using anisotropy assays.

8.3 RNA helicase activity

In the present study, Mth203 was found to be an ATP-dependent bidirectional RNA helicase with RNA-independent ATPase activity. The RNA helicase activity was studied using forked RNA substrates. DEAD-box helicases have also demonstrated unwinding of blunt end, uncapped RNA duplexes (Rozen *et al.* 1990), RNA hairpins (Digens and Uhlenbeck, 2001), RNA:DNA hybrids (Kikuma *et al.,* 2004), RNP complexes (Jankowsky *et al.,* 2001). Evaluation of RNA unwinding activity using different substrates might help further understand the role of Mth203 in the cell. In addition, study of the RNA unwinding activity of Mth203ΔC53 might help in understanding the effect of C- terminal on the activity of Mth203.

Mth203 was found to show unwinding in the presence of ATP and GTP. However, the unwinding was not studied in the presence of other nucleotides (CTP and UTP). Thus, helicase assays in the presence of CTP and UTP and non-hydrolysable ATP (ATPγP or AMP-PNP) would help in further characterizing the helicase activity.

8.4 mth203 expression in vivo

Transcriptome analysis has revealed that Mth203 transcripts are increased under cold-shock (Kato *et al.,* 2008), but protein expression analysis has not been performed on Mth203. Mth203 expression profile under cold-stress conditions and protein pull-down assays to study the proteins interacting with Mth203 under cold-shock conditions will further help in understanding Mth203 function. Also, further study of Mth203 cell cycle expression, would shed light on Mth203 function in the cell cycle and if the protein is involved in regulation of any cellular process.

As DEAD-box proteins are involved in a variety of processes involving RNA metabolism (Schmidt *et al.,* 1992, 2002), in order to deduce Mth203 function we

need to apply a bottom-up approach by performing biochemical studies for all the possible functions and eliminating the ones which test negative. As genetic studies are not possible in *M. thermautotrophicus*, these studies could be carried out on *mmp0457* in *M. maripaludis*. The S0001+pCB07 strain overexpresses *mmp0457* and this strain can be used to perform further biochemistry to assign the role of Mmp0457 in RNA metabolism. Mmp0457 might be involved in translation, which could be tested by polysome profile analysis (Foiani *et al.,* 1991) where a change in polysome content will imply that the protein is involved in protein synthesis. Mmp0457 involvement in ribosomal assembly and rRNA processing can be estimated by measuring relative amounts of 50S and 30S subunits in a sucrose gradient (Jamieson *et al.,* 1992). Further analysis of the overexpression of *mmp0457*, DNA content quantification and cell size will potentially help in understanding the role of Mth203 and its homologue in archaea.

Bibliography

- Abdel Monem M., Hoffmann Berling H. (1976) Enzymic Unwinding of DNA. European Journal of Biochemistry 65:431-440.

- Adams M.J., Blundell T.L., Dodson E.J., Dodson G.G., Vijayan M., Baker E.N., Harding M.M., Hodgkin D.C., Rimmer B., Sheat S. (1969) Structure of rhombohedral 2 zinc insulin crystals. Nature 224:491-495.

- Akita M., Adachi A., Takemura K., Yamagami T., Matsunaga F., Ishino Y. (2010) Cdc6/Orc1 from *Pyrococcus furiosus* may act as the origin recognition protein and Mcm helicase recruiter. Genes to Cells 15:537-552.

- Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990) Basic local alignment search tool. Journal of Molecular Biology 215:403-410.

- Argyle J.L., Tumbula D.L., Leigh J.A. (1996) Neomycin resistance as a selectable marker in *Methanococcus maripaludis*. Applied and Environmental Microbiology 62:4233.

- Arias E.E., Walter J.C. (2005) Replication-dependent destruction of Cdt1 limits DNA replication to a single round per cell cycle in *Xenopus* egg extracts. Genes & Development 19:114.

- Arias E.E., Walter J.C. (2007) Strength in numbers: preventing re-replication via multiple mechanisms in eukaryotic cells. Genes & Development 21:497.

- Ausubel F.M., Brent R., Kingston R.E., Moore D.D., Smith J.A., Seidman J.G., Struhl K.e. (1987) Protocols in Molecular Biology John Wiley and Sons, New York, NY.

- Barry E.R., Bell S.D. (2006) DNA replication in the archaea. Microbiology and Molecular Biology Reviews 70:876.

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

- Barta I., Iggo R. (1995) Autoregulation of expression of the yeast Dbp2p 'DEAD-box' protein is mediated by sequences in the conserved DBP2 intron. The EMBO journal 14:3800.

- Becker P.B., Hˆrz W. (2002) ATP-dependent nucleosome remodeling. Annual

Review of Biochemistry 71:247-273.

- Bell S.P. (2002) The origin recognition complex: from simple origins to complex functions. Genes & Development 16:659.

- Bell S.P., Dutta A. (2002) DNA replication in eukaryotic cells. Annual Review of Biochemistry 71:333.

- Bell S.P., Kobayashi R., Stillman B. (1993) Yeast origin recognition complex functions in transcription silencing and DNA replication. Science 262:1844.

- Benz J., Trachsel H., Baumann U. (1999) Crystal structure of the ATPase domain of translation initiation factor 4A from *Saccharomyces cerevisiae*-the prototype of the DEAD-box protein family. Structure 7:671-679.

- Bernander R. (2007) The cell cycle of *Sulfolobus*. Molecular Microbiology 66:557-562.

- Bernander R., Stokke T., Boye E. (1998) Flow cytometry of bacterial cells: Comparison between different flow cytometers and different DNA stains. Cytometry 31:29-36.

- Berneburg M., Lowe J.E., Nardo T., Arajo S., Fousteri M.I., Green M.H.L., Krutmann J., Wood R.D., Stefanini M., Lehmann A.R. (2000) UV damage causes uncontrolled DNA breakage in cells from patients with combined features of XP-D and Cockayne syndrome. The EMBO journal 19:1157-1166.

- Berquist B.R., DasSarma S. (2003) An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, Halobacterium sp. strain NRC-1. Journal of Bacteriology 185:5959.

- Bertani G. (1951) Studies on lysogenesis I., The mode of phage liberation by lysogenic *Escherichia coli*. Journal of Bacteriology 62:293.

- Bi X., Ren J., Goss D.J. (2000) Wheat germ translation initiation factor eIF4B affects eIF4A and eIFiso4F helicase activity by increasing the ATP binding affinity of eIF4A. Biochemistry 39:5758-5765.

- Bianchi M.E. (1994) Prokaryotic HU and eukaryotic HMG1: a kinked relationship. Molecular Microbiology 14:1-5.

- Blank C.E., Kessler P.S., Leigh J.A. (1995) Genetics in methanogens: transposon insertion mutagenesis of a *Methanococcus maripaludis* nifH gene. Journal of

Bacteriology 177:5773.

- Blow J.J., Hodgson B. (2002) Replication licensing--Origin licensing: defining the proliferative state? Trends in Cell Biology 12:72-78.

- Blow J.J., Laskey R.A. (1988) A role for the nuclear envelope in controlling DNA replication within the cell cycle. Nature 332:546-548.

- Bochar D.A., Stauffacher C.V., Rodwell V.W. (1999) Sequence Comparisons Reveal Two Classes of 3-Hydroxy-3-methylglutaryl Coenzyme A Reductase* 1. Molecular Genetics and Metabolism 66:122-127.

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

- Bocquier A.A., Liu L., Cann I.K.O., Komori K., Kohda D., Ishino Y. (2001) Archaeal primase: bridging the gap between RNA and DNA polymerases. Current Biology 11:452-456.

- Boehmer P.E., Lehman I. (1993) Physical interaction between the herpes simplex virus 1 origin-binding protein and single-stranded DNA-binding protein ICP8. Proceedings of the National Academy of Sciences 90:8444.

- Boonyaratanakornkit B.B., Simpson A.J., Whitehead T.A., Fraser C.M., El-Sayed N., Clark D.S. (2005) Transcriptional profiling of the hyperthermophilic methanarchaeon *Methanococcus jannaschii* in response to lethal heat and non-lethal cold shock. Environmental Microbiology 7:789.

- Bork P., Koonin E.V. (1993) An expanding family of helicases within the ìDEAD/Hî superfamily. Nucleic Acids Research 21:751.

- Bramhill D., Kornberg A. (1988) A model for initiation at origins of DNA replication. Cell 54:915-918.

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

- Brewster A.S., Chen X.S. (2010) Insights into the MCM functional mechanism: lessons learned from the archaeal MCM complex. Critical reviews in biochemistry and molecular biology 45:243-256.

- Brewster A.S., Wang G., Yu X., Greenleaf W.B., Carazo J.M., Tjajadi M., Klein M.G., Chen X.S. (2008) Crystal structure of a near-full-length archaeal MCM: functional

insights for an AAA+ hexameric helicase. Proceedings of the National Academy of Sciences 105:20191.

- Briolat V., Reysset G. (2002) Identification of the *Clostridium perfringens* genes involved in the adaptive response to oxidative stress. Journal of Bacteriology 184:2333.

- Brochier-Armanet C., Boussau B., Gribaldo S., Forterre P. (2008) Mesophilic *Crenarchaeota*: proposal for a third archaeal phylum, the *Thaumarchaeota*. Nature Reviews Microbiology 6:245-252.

- Broek D., Bartlett R., Crawford K., Nurse P. (1991) Involvement of p34cdc2 in establishing the dependency of S phase on mitosis. Nature 349:388-393.

- Brown G.W., Jallepalli P.V., Huneycutt B.J., Kelly T.J. (1997) Interaction of the S phase regulator cdc18 with cyclin-dependent kinase in fission yeast. Proceedings of the National Academy of Sciences 94:6142.

- Brown J.R. (2001) Genomic and phylogenetic perspectives on the evolution of prokaryotes. Systematic Biology 50:497.

- Brown J.R., Doolittle W.F. (1997) Archaea and the prokaryote-to-eukaryote transition. Microbiology and Molecular Biology Reviews 61:456.

- Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D., Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273:1058.

- Buszczak M., Spradling A.C. (2006) The Drosophila P68 RNA helicase regulates transcriptional deactivation by promoting RNA release from chromatin. Genes & Development 20:977.

- Buttner K., Wenig K., Hopfner K.P. (2005) Structural framework for the mechanism of archaeal exosomes in RNA processing. Molecular Cell 20:461-471.

- Calzada A., S·nchez M., S·nchez E., Bueno A. (2000) The Stability of the Cdc6 Protein Is Regulated by Cyclin-dependent Kinase/Cyclin B Complexes inSaccharomyces cerevisiae. Journal of Biological Chemistry 275:9734.

- Cann I.K.O., Ishino S., Hayashi I., Komori K., Toh H., Morikawa K., Ishino Y. (1999) Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. Journal of Bacteriology 181:6591.

- Capaldi S.A., Berger J.M. (2004) Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon *Methanothermobacter thermoautotrophicus*. Nucleic Acids Research 32:4821.

- Carmel A.B., Matthews B.W. (2004) Crystal structure of the BstDEAD N- terminal domain: A novel DEAD protein from *Bacillus stearothermophilus*. Rna 10:66.

- Carpentieri F., De Felice M., De Falco M., Rossi M., Pisani F.M. (2002) Physical and functional interaction between the MCM-like DNA helicase and the single-stranded DNA binding protein from the crenarchaeon *Sulfolobus solfataricus*. Journal of Biological Chemistry.

- Cartier G., Lorieux F., Allemand F., Dreyfus M., Bizebard T. (2010) Cold adaptation in DEAD-box proteins. Biochemistry 49:2636-2646.

- Caruthers J.M., Hu Y., McKay D.B. (2006) Structure of the second domain of the *Bacillus subtilis* DEAD-box RNA helicase YxiN. Acta Crystallographica Section F: Structural Biology and Crystallization Communications 62:1191-1195.

- Caruthers J.M., Johnson E.R., McKay D.B. (2000) Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase. Proceedings of the National Academy of Sciences 97:13080.

- Caruthers J.M., McKay D.B. (2002) Helicase structure and mechanism. Current Opinion in Structural Biology 12:123-133.

- Castresana J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular biology and evolution 17:540.

- Chamot D., Magee W.C., Yu E., Owttrim G.W. (1999) A cold shock-induced cyanobacterial RNA helicase. Journal of Bacteriology 181:1728.

- Charollais J., Dreyfus M., Iost I. (2004) CsdA, a cold shock RNA helicase from *Escherichia coli*, is involved in the biogenesis of 50S ribosomal subunit. Nucleic Acids Research 32:2751.

- Chedin F., Seitz E.M., Kowalczykowski S.C. (1998) Novel homologs of replication protein A in archaea: implications for the evolution of ssDNA-binding proteins. Trends in Biochemical Sciences 23:273-277.

- Chen L., Brugger K., Skovgaard M., Redder P., She Q., Torarinsson E., Greve B., Awayez M., Zibat A., Klenk H.P. (2005) The genome of *Sulfolobus acidocaldarius*, a

model organism of the Crenarchaeota. Journal of Bacteriology 187:4992.

- Chesnokov I., Gossen M., Remus D., Botchan M. (1999) Assembly of functionally active *Drosophila* origin recognition complex from recombinant proteins. Genes & Development 13:1289.

- Chong J.P.J., Hayashi M.K., Simon M.N., Xu R.M., Stillman B. (2000) A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. Proceedings of the National Academy of Sciences 97:1530.

- Chong J.P.J., Mahbubani H.M., Khoo C.Y., Blow J.J. (1995) Purification of an MCM-containing complex as a component of the DNA replication licensing system. Nature 375:418-421.

- Chuang R.Y., ChrÈtien L., Dai J., Kelly T.J. (2002) Purification and Characterization of the *Schizosaccharomyces pombe* Origin Recognition Complex. Journal of Biological Chemistry 277:16920.

- Chuang R.Y., Kelly T.J. (1999) The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. Proceedings of the National Academy of Sciences 96:2656.

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

- Cooper S., Helmstetter C.E. (1968) Chromosome replication and the division cycle of *Escherichia coli* B/r* 1. Journal of Molecular Biology 31:519-540.

- Cordin O., Banroques J., Tanner N.K., Linder P. (2006) The DEAD-box protein family of RNA helicases. Gene 367:17-37.

- 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. The EMBO journal 27:2250-2258.

- Cox C.J., Foster P.G., Hirt R.P., Harris S.R., Embley T.M. (2008) The archaebacterial origin of eukaryotes. Proceedings of the National Academy of Sciences 105:20356.

- De Felice M., Esposito L., Pucci B., Carpentieri F., De Falco M., Rossi M., Pisani F.M. (2003) Biochemical characterization of a CDC6-like protein from the crenarchaeon *Sulfolobus solfataricus*. Journal of Biological Chemistry 278:46424.

- Desogus G., Onesti S., Brick P., Rossi M., Pisani F.M. (1999) Identification and characterization of a DNA primase from the hyperthermophilic archaeon *Methanococcus jannaschii*. Nucleic Acids Research 27:4444.

- Devault A., Vallen E.A., Yuan T., Green S., Bensimon A., Schwob E. (2002) Identification of Tah11/Sid2 as the ortholog of the replication licensing factor Cdt1 in *Saccharomyces cerevisiae*. Current Biology 12:689-694.

- Diffley J.F.X. (2004) Regulation of early events in chromosome replication. Current Biology 14:R778-R786.

- Diffley J.F.X., Cocker J.H., Dowell S.J., Rowley A. (1994) Two steps in the assembly of complexes at yeast replication origins in vivo. Cell 78:303-316.

- Diges C.M., Uhlenbeck O.C. (2001) *Escherichia coli* DbpA is an RNA helicase that requires hairpin 92 of 23S rRNA. The EMBO journal 20:5503-5512.

- Dijkwe P.A., Vaughn J.P., Hamlin J.L. (1994) Replication initiation sites are distributed widely in the amplified CHO dihydrofolate reductase domain. Nucleic Acids Research 22:4989.

- DiMarco A.A., Bobik T.A., Wolfe R.S. (1990) Unusual coenzymes of methanogenesis. Annual Review of Biochemistry 59:355-394.

- Dodson M.S., Lehman I. (1991) Association of DNA helicase and primase activities with a subassembly of the herpes simplex virus 1 helicase-primase composed of the UL5 and UL52 gene products. Proceedings of the National Academy of Sciences 88:1105.

- Donovan S., Harwood J., Drury L.S., Diffley J.F.X. (1997) Cdc6-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. Proceedings of the National Academy of Sciences 94:5611-5616.

- Doolittle W.F., Brown J.R. (1994) Tempo, mode, the progenote, and the universal root. Proceedings of the National Academy of Sciences 91:6721.

- Doolittle W.F., Logsdon Jr J.M. (1998) Archaeal genomics: Do archaea have a mixed heritage? Current Biology 8:R209-R211.

- Drury L.S., Perkins G., Diffley J.F.X. (1997) The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. The EMBO journal 16:5966-5976.

- Du Y.C.N., Stillman B. (2002) Yph1p, an ORC-Interacting Protein: Potential Links

between Cell Proliferation Control, DNA Replication, and Ribosome Biogenesis. Cell 109:835-848.

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

- Duggin I.G., Bell S.D. (2006) The chromosome replication machinery of the archaeon *Sulfolobus solfataricus*. Journal of Biological Chemistry 281:15029.

- Dutta A., Bell S.P. (1997) Initiation of DNA replication in eukaryotic cells. Annual review of Cell and Developmental Biology 13:293-332.

- Edgell D.R., Doolittle W.F. (1997) Archaea and the origin (s) of DNA replication proteins. Cell 89:995-998.

- Efron B (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*. CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38, SIAM, Philadelphia.

- Ellison V., Stillman B. (2001) Opening of the Clamp: An Intimate View of an ATP-Driven Biological Machine. Cell 106:655-660.

- Eryilmaz J., Ceschini S., Ryan J., Geddes S., Waters T.R., Barrett T.E. (2006) Structural insights into the cryptic DNA-dependent ATPase activity of UvrB. Journal of Molecular Biology 357:62-72.

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

- Erzberger J.P., Pirruccello M.M., Berger J.M. (2002) The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. The EMBO Journal 21:4763-4773.

- Fairman-Williams M.E., Guenther U.P., Jankowsky E. (2010) SF1 and SF2 helicases: family matters. Current Opinion in Structural Biology 20:313-324.

- Fan L., Arvai A.S., Cooper P.K., Iwai S., Hanaoka F., Tainer J.A. (2006) Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. Molecular Cell 22:27-37.

- Fang L., Davey M.J., O'Donnell M. (1999) Replisome assembly at *oriC*, the replication origin of E. coli, reveals an explanation for initiation sites outside an origin. Molecular Cell 4:541-553.

- Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496.

- Fletcher R.J., Bishop B.E., Leon R.P., Sclafani R.A., Ogata C.M., Chen X.S. (2003) The structure and function of MCM from archaeal *Methanothermobacter thermoautotrophicum*. Nature Structural & Molecular Biology 10:160-167.

- Fletcher R.J., Shen J., Holden L.G., Chen X.S. (2008) Identification of amino acids important for the biochemical activity of *Methanothermobacter thermautotrophicus* MCM. Biochemistry 47:9981-9986.

- Foiani M., Cigan A., Paddon C., Harashima S., Hinnebusch A. (1991) GCD2, a translational repressor of the GCN4 gene, has a general function in the initiation of protein synthesis in *Saccharomyces cerevisiae*. Molecular and Cellular Biology 11:3203.

- Fu Y.V., Yardimci H., Long D.T., Guainazzi A., Bermudez V.P., Hurwitz J., van Oijen A., Scherer O.D., Walter J.C. (2011) Selective Bypass of a Lagging Strand Roadblock by the Eukaryotic Replicative DNA Helicase. Cell 146:931-941.

- Fujimitsu K., Senriuchi T., Katayama T. (2009) Specific genomic sequences of *E. coli* promote replicational initiation by directly reactivating ADP-DnaA. Genes & Development 23:1221.

- Fukuda T., Yamagata K., Fujiyama S., Matsumoto T., Koshida I., Yoshimura K., Mihara M., Naitou M., Endoh H., Nakamura T. (2007) DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. Nature Cell Biology 9:604-611.

- Fukui T., Atomi H., Kanai T., Matsumi R., Fujiwara S., Imanaka T. (2005) Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with Pyrococcus genomes. Genome Research 15:352.

- Fuller R.S., Funnell B.E., Kornberg A. (1984) The dnaA protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. Cell 38:889-900.

- Fuller-Pace F., Nicol S., Reid A., Lane D. (1993) DbpA: a DEAD-box protein specifically activated by 23s rRNA. The EMBO Journal 12:3619.

- Gai D., Zhao R., Li D., Finkielstein C.V., Chen X.S. (2004) Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. Cell 119:47-60.

- Gardner W.L., Whitman W.B. (1999) Expression vectors for *Methanococcus maripaludis*: overexpression of acetohydroxyacid synthase and fl-galactosidase. Genetics 152:1439.

- Garrity G.M., Holt J.G. (2001) The road map to manual. In Bergey's Manual of Systematic Bacteriology, pp. 119-166, 2nd edn. Edited by D. R. Boone, R. W. Castenholz & G. M. Garrity. New York: Springer.

- Gaudier M., Schuwirth B.S., Westcott S.L., Wigley D.B. (2007) Structural basis of DNA replication origin recognition by an ORC protein. Science 317:1213.

- Geider K., Hoffmann-Berling H. (1981) Proteins controlling the helical structure of DNA. Annual review of biochemistry 50:233-260.

- Gong Q., Li P., Ma S., Indu Rupassara S., Bohnert H.J. (2005) Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative Arabidopsis thaliana. The Plant Journal 44:826-839.

- Gong Z., Morales-Ruiz T., Ariza R.R., Rold·n-Arjona T., David L., Zhu J.K. (2002) ROS1, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. Cell 111:803-814.

- Gorbalenya A.E., Koonin E.V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. Current Opinion in Structural Biology 3:419-429.

- Gorbalenya A.E., Koonin E.V., Donchenko A.P., Blinov V.M. (1989) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Research 17:4713.

- Grabowski B., Kelman Z. (2003) Archaeal DNA replication: eukaryal proteins in a bacterial context. Annual Reviews in Microbiology 57:487-516.

- Grainge I., Gaudier M., Schuwirth B.S., Westcott S.L., Sandall J., Atanassova N., Wigley D.B. (2006) Biochemical analysis of a DNA replication origin in the archaeon *Aeropyrum pernix*. Journal of Molecular Biology 363:355-369.

- Grayling R.A., Sandman K., Reeve J.N. (1994) Archaeal DNA binding proteins and

chromosome structure. Systematic and Applied Microbiology 16:582-590.

- Gregg A.V., McGlynn P., Jaktaji R.P., Lloyd R.G. (2002) Direct rescue of stalled DNA replication forks via the combined action of PriA and RecG helicase activities. Molecular Cell 9:241-251.

- Grifo J.A., Abramson R.D., Satler C.A., Merrick W.C. (1984) RNA-stimulated ATPase activity of eukaryotic initiation factors. Journal of Biological Chemistry 259:8648.

- Grigoriev A. (1998) Analyzing genomes with cumulative skew diagrams. Nucleic Acids Research 26:2286.

- Grishaev A., Llin·s M. (2002) Protein structure elucidation from NMR proton densities. Proceedings of the National Academy of Sciences 99:6713.

- Grohman J.K., Del Campo M., Bhaskaran H., Tijerina P., Lambowitz A.M., Russell R. (2007) Probing the mechanisms of DEAD-box proteins as general RNA chaperones: the C- terminal domain of CYT-19 mediates general recognition of RNA. Biochemistry 46:3013-3022.

- Gupta R.S., Shami A. (2011) Molecular signatures for the *Crenarchaeota* and the *Thaumarchaeota*. Antonie van Leeuwenhoek:1-25.

- Gustafson E.A., Wessel G.M. (2010) DEAD-box helicases: Posttranslational regulation and function. Biochemical and Biophysical Research Communications 395:1.

- Haseltine C.A., Kowalczykowski S.C. (2002) A distinctive single stranded DNA binding protein from the Archaeon *Sulfolobus solfataricus*. Molecular Microbiology 43:1505-1515.

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

- Hayles J., Fisher D., Woollard A., Nurse P. (1994) Temporal order of S phase and mitosis in fission yeast is determined by the state of the p34cdc2-mitotic B cyclin complex. Cell 78:813-822.

- Helmstetter C.E., Leonard A.C. (1987) Coordinate initiation of chromosome and minichromosome replication in *Escherichia coli*. Journal of Bacteriology 169:3489.

- Hendrickson E., Kaul R., Zhou Y., Bovee D., Chapman P., Chung J., Conway de

Macario E., Dodsworth J., Gillett W., Graham D. (2004) Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*. Journal of Bacteriology 186:6956.

- Heyduk T., Ma Y., Tang H., Ebright R.H. (1996) Fluorescence anisotropy: rapid, quantitative assay for protein-DNA and protein-protein interaction. Methods in Enzymology 274:492-503.

- Hilbert M., Karow A.R., Klostermeier D. (2009) The mechanism of ATP-dependent RNA unwinding by DEAD-box proteins. Biological Chemistry 390:1237-1250.

- Hildenbrand C., Stock T., Lange C., Rother M., Soppa J. (2011) Genome copy numbers and gene conversion in methanogenic archaea. Journal of Bacteriology 193:734.

- Hjort K., Bernander R. (1999) Changes in cell size and DNA content in *Sulfolobus* cultures during dilution and temperature shift experiments. Journal of Bacteriology 181:5669.

- Hogbom M., Collins R., van den Berg S., Jenvert R.M., Karlberg T., Kotenyova T., Flores A., Hedestam G.B.K., Schiavone L.H. (2007) Crystal structure of conserved domains 1 and 2 of the human DEAD-box helicase DDX3X in complex with the mononucleotide AMP. Journal of Molecular Biology 372:150-159.

- Holland L., Downey M., Song X., Gauthier L., Bell Rogers P., Yankulov K. (2002) Distinct parts of minichromosome maintenance protein 2 associate with histone H3/H4 and RNA polymerase II holoenzyme. European Journal of Biochemistry 269:5192-5202.

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

- Huber W., Von Heydebreck A., S¸ltmann H., Poustka A., Vingron M. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 18:S96.

- Hunger K., Beckering C.L., Wiegeshoff F., Graumann P.L., Marahiel M.A. (2006) Cold-induced putative DEAD-box RNA helicases CshA and CshB are essential for cold adaptation and interact with cold shock protein B in *Bacillus subtilis*. Journal

of Bacteriology 188:240.

- Hwang D.S., Crooke E., Kornberg A. (1990) Aggregated dnaA protein is dissociated and activated for DNA replication by phospholipase or dnaK protein. Journal of Biological Chemistry 265:19244.

- Iggo R., Jamieson D., MacNeill S., Southgate J., McPheat J., Lane D. (1991) p68 RNA helicase: identification of a nucleolar form and cloning of related genes containing a conserved intron in yeasts. Molecular and Cellular Biology 11:1326.

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

- Imamura M., Uemori T., Kato I., Ishino Y. (1995) A non-alpha-like DNA polymerase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Biological & Pharmaceutical Bulletin 18:1647.

- Iost I., Dreyfus M., Linder P. (1999) Ded1p, a DEAD-box protein required for translation initiation in *Saccharomyces cerevisiae*, is an RNA helicase. Journal of Biological Chemistry 274:17677.

- Ip F.C.F., Chung S.S.K., Fu W.Y., Ip N.Y. (2000) Developmental and tissue-specific expression of DEAD-box protein p72. Neuroreport 11:457.

- Ishida T., Akimitsu N., Kashioka T., Hatano M., Kubota T., Ogata Y., Sekimizu K., Katayama T. (2004) DiaA, a novel DnaA-binding protein, ensures the timely initiation of *Escherichia coli* chromosome replication. Journal of Biological Chemistry 279:45546.

- Iwai A., Takegami T., Shiozaki T., Miyazaki T. (2011) Hepatitis C Virus NS3 Protein Can Activate the Notch-Signaling Pathway through Binding to a Transcription Factor, SRCAP.

- Jacob F., Brenner S., Cuzin F. (1963) On the regulation of DNA replication in bacteria, Cold Spring Harbor Laboratory Press. pp. 329.

- Jagus R., Anderson W., Safer B. (1981) Initiation of mammalian protein biosynthesis. Progress in Nucleic Acid Research and Molecular Biology 25:127-185.

- Jamieson D.J., Rahe B., Pringle J., Beggs J.D. (1991) A suppressor of a yeast splicing mutation (prp8-1) encodes a putative ATP-dependent RNA helicase. Nature

349:715-717.

- Janknecht R. (2010) Multi-talented DEAD-box proteins and potential tumor promoters: p68 RNA helicase (DDX5) and its paralog, p72 RNA helicase (DDX17). American Journal of Translational Research 2:223.

- Jaramillo M., Dever T., Merrick W., Sonenberg N. (1991) RNA unwinding in translation: assembly of helicase complex intermediates comprising eukaryotic initiation factors eIF-4F and eIF-4B. Molecular and Cellular Biology 11:5992.

- Jarrell K.F., Walters A.D., Bochiwal C., Borgia J.M., Dickinson T., Chong J.P.J. (2011) Major players on the microbial stage: why archaea are important. Microbiology 157:919-936.

- Johnson E.R., McKay D.B. (1999) Crystallographic structure of the amino terminal domain of yeast initiation factor 4A, a representative DEAD-box RNA helicase. Rna 5:1526.

- Johnson R.T., Rao P.N. (1970) Mammalian cell fusion: induction of premature chromosome condensation in interphase nuclei.

- Jones K.A., Kadonaga J.T., Rosenfeld P.J., Kelly T.J., Tjian R. (1987) A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell 48:79-89.

- Jones P.G., Mitta M., Kim Y., Jiang W., Inouye M. (1996) Cold shock induces a major ribosomal-associated protein that unwinds double-stranded RNA in *Escherichia coli*. Proceedings of the National Academy of Sciences 93:76.

- Kamimura Y., Tak Y.S., Sugino A., Araki H. (2001) Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in *Saccharomyces cerevisiae*. The EMBO Journal 20:2097-2107.

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

- Kanter D.M., Kaplan D.L. (2011) Sld2 binds to origin single-stranded DNA and stimulates DNA annealing. Nucleic Acids Research 39:2580.

- Kaplan D.L., Davey M.J., O'Donnell M. (2003) Mcm4, 6, 7 uses a 'pump in ring' mechanism to unwind DNA by steric exclusion and actively translocate along a

duplex. Journal of Biological Chemistry 278:49171.

- Karginov F.V., Caruthers J.M., Hu Y.X., McKay D.B., Uhlenbeck O.C. (2005) YxiN is a modular protein combining a DExD/H core and a specific RNA-binding domain. Journal of Biological Chemistry 280:35499.

- Kasiviswanathan R., Shin J.H., Kelman Z. (2005) Interactions between the archaeal Cdc6 and MCM proteins modulate their biochemical properties. Nucleic Acids Research 33:4940.

- Kasiviswanathan R., Shin J.H., Kelman Z. (2006) DNA binding by the *Methanothermobacter thermautotrophicus* Cdc6 protein is inhibited by the minichromosome maintenance helicase. Journal of Bacteriology 188:4577.

- Katayama T., Kubota T., Kurokawa K., Crooke E., Sekimizu K. (1998) The initiator function of DnaA protein is negatively regulated by the sliding clamp of the *E. coli* chromosomal replicase. Cell 94:61-71.

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

- Kato S., Kosaka T., Watanabe K. (2008) Comparative transcriptome analysis of responses of *Methanothermobacter thermautotrophicus* to different environmental stimuli. Environmental Microbiology 10:893-905.

- Kawakami H.K.H., Katayama T.K.T. (2010) DnaA, ORC, and Cdc6: similarity beyond the domains of life and diversity This paper is one of a selection of papers published in this special issue entitled 8th International Conference on AAA+ Proteins and has undergone the Journal's usual peer review process. Biochemistry and Cell Biology 88:49-62.

- Keeling P.J., Charlebois R.L., Ford Doolittle W. (1994) Archaebacterial genomes: eubacterial form and eukaryotic content. Current Opinion in Genetics & Development 4:816-822.

- Keeling P.J., Doolittle W.F. (1995) Archaea: narrowing the gap between prokaryotes and eukaryotes. Proceedings of the National Academy of Sciences of the United States of America 92:5761.

- Kelly T.J., Brown G.W. (2000) Regulation of chromosome replication. Annual Review of Biochemistry 69:829-880.

- Kelly T.J., Martin G.S., Forsburg S.L., Stephen R.J., Russo A., Nurse P. (1993) The fission yeast cdc18+ gene product couples S phase to START and mitosis. Cell 74:371-382.

- Kelly T.J., Simancek P., Brush G.S. (1998) Identification and characterization of a single-stranded DNA-binding protein from the archaeon *Methanococcus jannaschii.* Proceedings of the National Academy of Sciences 95:14634.

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

- Kelman L.M., Kelman Z. (2004) Multiple origins of replication in archaea. Trends in Microbiology 12:399-401.

- Kelman Z. (2000) DNA replication in the third domain (of life). Current Protein and Peptide Science 1:139-154.

- Kelman Z., Lee J.K., Hurwitz J. (1999) The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* H contains DNA helicase activity. Proceedings of the National Academy of Sciences 96:14783.

- Kerr I.D., Wadsworth R.I.M., Cubeddu L., Blankenfeldt W., Naismith J.H., White M.F. (2003) Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein. The EMBO Journal 22:2561-2570.

- Keyamura K., Fujikawa N., Ishida T., Ozaki S., Suíetsugu M., Fujimitsu K., Kagawa W., Yokoyama S., Kurumizaka H., Katayama T. (2007) The interaction of DiaA and DnaA regulates the replication cycle in E. coli by directly promoting ATP-DnaA-specific initiation complexes. Genes & Development 21:2083.

- Kikuma T., Ohtsu M., Utsugi T., Koga S., Okuhara K., Eki T., Fujimori F., Murakami Y. (2004) Dbp9p, a member of the DEAD-box protein family, exhibits DNA helicase activity. Journal of Biological Chemistry 279:20692.

- Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A., Murcko M.A., Lin C., Caron P.R. (1998) Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. Structure 6:89-100.

- Kim Y., Quartey P., Li H., Volkart L., Hatzos C., Chang C., Nocek B., Cuff M., Osipiuk J., Tan K. (2008) Large-scale evaluation of protein reductive methylation for

improving protein crystallization. Nature Methods 5:853.

- Kitagawa R., Ozaki T., Moriya S., Ogawa T. (1998) Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for *Escherichia coli* DnaA protein. Genes & Development 12:3032.

- Klemm R.D., Austin R.J., Bell S.P. (1997) Coordinate binding of ATP and origin DNA regulates the ATPase activity of the origin recognition complex. Cell 88:493-502.

- Klostermeier D., Rudolph M.G. (2009) A novel dimerization motif in the C-terminal domain of the *Thermus thermophilus* DEAD-box helicase Hera confers substantial flexibility. Nucleic Acids Research 37:421.

- Komori K., Ishino Y. (2001) Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. Journal of Biological Chemistry 276:25654.

- Koo H.S., Claassen L., Grossman L., Liu L.F. (1991) ATP-dependent partitioning of the DNA template into supercoiled domains by Escherichia coli UvrAB. Proceedings of the National Academy of Sciences 88:1212.

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

- Kornberg A., Baker T.A. (1992) DNA replication WH Freeman New York:.

- Korolev S., Lohman T.M., Waksman G., Yao N., Weber P.C. (1998) Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 super families of helicases. Protein Science 7:605-610.

- Kossen K., Karginov F.V., Uhlenbeck O.C. (2002) The Carboxy- terminal Domain of the DEx Protein YxiN is Sufficient to Confer Specificity for 23 S rRNA. Journal of Molecular Biology 324:625-636.

- Kuang W.F., Lin Y.C., Jean F., Huang Y.W., Tai C.L., Chen D.S., Chen P.J., Hwang L.H. (2004) Hepatitis C virus NS3 RNA helicase activity is modulated by the two domains of NS3 and NS4A. Biochemical and Biophysical Research Communications 317:211-217.

- Kujat S.L., Owttrim G.W. (2000) Redox-regulated RNA helicase expression. Plant Physiology 124:703.

- Kunau W.H., Beyer A., Franken T., Gˆtte K., Marzioch M., Saidowsky J., Skaletz-

Rorowski A., Wiebel F.F. (1993) Two complementary approaches to study peroxisome biogenesis in *Saccharomyces cerevisiae*: forward and reversed genetics. Biochimie 75:209-224.

- Labib K., Diffley J.F.X., Kearsey S.E. (1999) G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus. Nature Cell Biology 1:415-422.

- Lacks S., Greenberg B. (1977) Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation* 1. Journal of Molecular Biology 114:153-168.

- Lam W.L., Doolittle W.F. (1992) Mevinolin-resistant mutations identify a promoter and the gene for a eukaryote-like 3-hydroxy-3-methylglutaryl-coenzyme A reductase in the archaebacterium *Haloferax volcanii*. Journal of Biological Chemistry 267:5829.

- Langer D., Hain J., Thuriaux P., Zillig W. (1995) Transcription in archaea: similarity to that in eucarya. Proceedings of the National Academy of Sciences 92:5768.

- Laskey R.A., Madine M.A. (2003) A rotary pumping model for helicase function of MCM proteins at a distance from replication forks. EMBO reports 4:26-30.

- Lee D.G., Makhov A.M., Klemm R.D., Griffith J.D., Bell S.P. (2000) Regulation of origin recognition complex conformation and ATPase activity: differential effects of single-stranded and double-stranded DNA binding. The EMBO Journal 19:4774-4782.

- Lee J.K., Moon K.Y., Jiang Y., Hurwitz J. (2001) The *Schizosaccharomyces pombe* origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. Proceedings of the National Academy of Sciences 98:13589.

- Lei M., Tye B.K. (2001) Initiating DNA synthesis: from recruiting to activating the MCM complex. Journal of Cell Science 114:1447.

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

- Liang C., Stillman B. (1997) Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in cdc6 mutants. Genes & Development

11:3375.

- Liang L., Diehl-Jones W., Lasko P. (1994) Localization of vasa protein to the *Drosophila* pole plasm is independent of its RNA-binding and helicase activities. Development 120:1201.

- Liku M.E., Nguyen V.Q., Rosales A.W., Irie K., Li J.J. (2005) CDK phosphorylation of a novel NLS-NES module distributed between two subunits of the Mcm2-7 complex prevents chromosomal rereplication. Molecular Biology of the Cell 16:5026.

- Lim J., Thomas T., Cavicchioli R. (2000) Low temperature regulated DEAD-box RNA helicase from the antarctic archaeon, *Methanococcoides burtonii1*. Journal of Molecular Biology 297:553-567.

- Lin C., Kim J.L. (1999) Structure-based mutagenesis study of hepatitis C virus NS3 helicase. Journal of Virology 73:8798.

- Linden M.H., Hartmann R.K., Klostermeier D. (2008) The putative RNase P motif in the DEAD-box helicase Hera is dispensable for efficient interaction with RNA and helicase activity. Nucleic Acids Research 36:5800.

- Linder P., Lasko P. (2006) Bent out of shape: RNA unwinding by the DEAD-box helicase Vasa. Cell 125:219-221.

- Lindwall G., Chau M.F., Gardner S., Kohlstaedt L. (2000) A sparse matrix approach to the solubilization of overexpressed proteins. Protein Engineering 13:67.

- Liu J., Smith C.L., DeRyckere D., DeAngelis K., Martin G.S., Berger J.M. (2000) Structure and Function of Cdc6/Cdc18:: Implications for Origin Recognition and Checkpoint Control. Molecular Cell 6:637-648.

- Liu L., Komori K., Ishino S., Bocquier A.A., Cann I.K.O., Kohda D., Ishino Y. (2001) The Archaeal DNA Primase. Journal of Biological Chemistry 276:45484.

- 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 Research 36:3235.

- Lopez P., Philippe H., Myllykallio H., Forterre P. (1999) Identification of putative chromosomal origins of replication in Archaea. Molecular Microbiology 32:883-886.

- Lu M., Campbell J.L., Boye E., Kieckner N. (1994) SeqA: a negative modulator of DNA replication initiation in *E. coli*. Cell 77:413-426.

- Luking A., Stahl U., Schmidt U. (1998) The protein family of RNA helicases. Critical Reviews in Biochemistry and Molecular Biology 33:259-296.

- Lundgren M., Andersson A., Chen L., Nilsson P., Bernander R. (2004) Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. Proceedings of the National Academy of Sciences of the United States of America 101:7046.

- Maiorano D., Moreau J. (2000) XCDT1 is required for the assembly of pre-replicative complexes in Xenopus laevis. Nature 404:560-1.

- Maisnier Patin S., Malandrin L., Birkeland N.K., Bernander R. (2002) Chromosome replication patterns in the hyperthermophilic euryarchaea *Archaeoglobus fulgidus* and *Methanocaldococcus (Methanococcus) jannaschii*. Molecular Microbiology 45:1443-1450.

- Majernik A., Chong J. (2008) A conserved mechanism for replication origin recognition and binding in archaea. Biochemistry Journal 409:511-518.

- Majernik A.I., Lundgren M., McDermott P., Bernander R., Chong J.P.J. (2005) DNA content and nucleoid distribution in *Methanothermobacter thermautotrophicus*. Journal of Bacteriology 187:1856.

- Malandrin L., Huber H., Bernander R. (1999) Nucleoid structure and partition in *Methanococcus jannaschii*: an archaeon with multiple copies of the chromosome. Genetics 152:1315.

- Malik A.K., Martinez R., Muncy L., Carmichael E.P., Weller S.K. (1992) Genetic analysis of the herpes simplex virus type 1 UL9 gene: isolation of a LacZ insertion mutant and expression in eukaryotic cells. Virology 190:702-715.

- Malorano D., Moreau J., Mechall M. (2000) XCDT1 is required for the assembly of pre-replicative complexes in *Xenopus laevis*. Nature 404:622-625.

- Marintcheva B., Weller S.K. (2003) Existence of transdominant and potentiating mutants of UL9, the herpes simplex virus type 1 origin-binding protein, suggests that levels of UL9 protein may be regulated during infection. Journal of Virology 77:9639.

- Martinez R., Shao L., Weller S.K. (1992) The conserved helicase motifs of the herpes simplex virus type 1 origin-binding protein UL9 are important for function. Journal of Virology 66:6735.

- Matsunaga F., Forterre P., Ishino Y., Myllykallio H. (2001) In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. Proceedings of the National Academy of Sciences 98:11152.

- Matsunaga F., Norais C., Forterre P., Myllykallio H. (2003) Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. EMBO reports 4:154-158.

- Matsunaga F., Takemura K., Akita M., Adachi A., Yamagami T., Ishino Y. (2010) Localized melting of duplex DNA by Cdc6/Orc1 at the DNA replication origin in the hyperthermophilic archaeon *Pyrococcus furiosus*. Extremophiles 14:21-31.

- McDermott P. (2009) Proliferation mechanisms in *Methanothermobacter thermautotrophicus*. PhD thesis, Biology, University of York.

- McGlynn P., Lloyd R.G. (2000) Modulation of RNA polymerase by (p) ppGpp reveals a RecG-dependent mechanism for replication fork progression. Cell 101:35-45.

- McPherson A. (2004) Introduction to protein crystallization. Methods 34:254-265.

- Mechali M., Kearsey S. (1984) Lack of specific sequence requirement for DNA replication in *Xenopus* eggs compared with high sequence specificity in yeast. Cell 38:55-64.

- MÈdigue C., Krin E., Pascal G., Barbe V., Bernsel A., Bertin P.N., Cheung F., Cruveiller S., D'Amico S., Duilio A. (2005) Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. Genome Research 15:1325.

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

- Messer W. (2002) The bacterial replication initiator DnaA. DnaA and *oriC*, the bacterial mode to initiate DNA replication. FEMS microbiology reviews 26:355-374.

- Messer W., Bellekes U., Lother H. (1985) Effect of dam methylation on the activity of the *E. coli* replication origin, *oriC*. The EMBO Journal 4:1327.

- Mimura S., Seki T., Tanaka S., Diffley J.F.X. (2004) Phosphorylation-dependent binding of mitotic cyclins to Cdc6 contributes to DNA replication control. Nature 431:1118-1123.

- Mizushima T., Takahashi N., Stillman B. (2000) Cdc6p modulates the structure and DNA binding activity of the origin recognition complex in vitro. Genes & Development 14:1631.

- Mohr G., Del Campo M., Mohr S., Yang Q., Jia H., Jankowsky E., Lambowitz A.M. (2008) Function of the C- terminal domain of the DEAD-box protein Mss116p analyzed in vivo and in vitro. Journal of Molecular Biology 375:1344-1364.

- Moissl C., Rachel R., Briegel A., Engelhardt H., Huber R. (2005) The unique structure of archaeal *ëhamií*, highly complex cell appendages with nano grappling hooks. Molecular Microbiology 56:361-370.

- Moll T., Tebb G., Surana U., Robitsch H., Nasmyth K. (1991) The role of phosphorylation and the CDC28 protein kinase in cell cycle-regulated nuclear import of the *S. cerevisiae* transcription factor SW15. Cell 66:743-758.

- Montero-Lomel M., Morais B.L.B., Figueiredo D.L., Neto D., Martins J.R.P., Masuda C.A. (2002) The initiation factor eIF4A is involved in the response to lithium stress in *Saccharomyces cerevisiae*. Journal of Biological Chemistry 277:21542.

- Moon K.Y., Kong D., Lee J.K., Raychaudhuri S., Hurwitz J. (1999) Identification and reconstitution of the origin recognition complex from *Schizosaccharomyces pombe.* Proceedings of the National Academy of Sciences 96:12367.

- Moore B.C., Leigh J.A. (2005) Markerless mutagenesis in *Methanococcus maripaludis* demonstrates roles for alanine dehydrogenase, alanine racemase, and alanine permease. Journal of Bacteriology 187:972.

- Moreno S., Nurse P. (1994) Regulation of progression through the G1 phase of the cell cycle by the rum1+ gene. Nature 367:236.

- Morgan R.M., Pihl T.D., Nolling J., Reeve J.N. (1997) Hydrogen regulation of growth, growth yields, and methane gene transcription in *Methanobacterium thermoautotrophicum deltaH*. Journal of Bacteriology 179:889.

- Morlang S., Weglˆhner W., Franceschi F. (1999) Hera from *Thermus thermophilus*: the first thermostable DEAD-box helicase with an RNase P protein motif1. Journal of Molecular Biology 294:795-805.

- Moullintraffort L., Bruneaux M., Nazabal A., Allegro D., Giudice E., Zal F., Peyrot V., Barbier P., Thomas D., Garnier C. (2010) Biochemical and biophysical characterization of the Mg2+-induced 90-kDa heat shock protein oligomers. Journal of Biological Chemistry 285:15100.

- Muller B., Tsaneva I.R., West S.C. (1993) Branch migration of Holliday junctions promoted by the Escherichia coli RuvA and RuvB proteins. I. Comparison of RuvAB-and RuvB-mediated reactions. Journal of Biological Chemistry 268:17179.

- Myllykallio H., Forterre P. (2000) Mapping of a chromosome replication origin in an archaeon: response. Trends in Microbiology 8:537-39.

- Nakamura T., Muramoto Y., Yokota S., Ueda A., Takabe T. (2004) Structural and transcriptional characterization of a salt-responsive gene encoding putative ATP-dependent RNA helicase in barley. Plant Science 167:63-70.

- Newman E.R. (2011) Biophysical analysis of the Nucleosome Assembly Protein and its interaction with histones. PhD thesis, University of Portsmouth.

- Newman J., Egan D., Walter T.S., Meged R., Berry I., Ben Jelloul M., Sussman J.L., Stuart D.I., Perrakis A. (2005) Towards rationalization of crystallization screening for small-to medium-sized academic laboratories: the PACT/JCSG+ strategy. Acta Crystallographica Section D: Biological Crystallography 61:1426-1431.

- Nguyen V.Q. (2001) Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature 411:1068-1073.

- Nguyen V.Q., Co C., Irie K., Li J.J. (2000) Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2-7. Current Biology 10:195-205.

- Nickell S., Hegerl R., Baumeister W., Rachel R. (2003) *Pyrodictium cannulae* enter the periplasmic space but do not enter the cytoplasm, as revealed by cryo-electron tomography. Journal of Structural Biology 141:34-42.

- Nishino T., Komori K., Tsuchiya D., Ishino Y., Morikawa K. (2005) Crystal structure and functional implications of *Pyrococcus furiosus* hef helicase domain involved in branched DNA processing. Structure 13:143-153.

- Nishitani H., Sugimoto N., Roukos V., Nakanishi Y., Saijo M., Obuse C., Tsurimoto T., Nakayama K.I., Nakayama K., Fujita M. (2006) Two E3 ubiquitin ligases, SCF-Skp2 and DDB1-Cul4, target human Cdt1 for proteolysis. The EMBO Journal 25:1126-1136.

- Nolling J., Frijlink M., de Vos W.M. (1991) Isolation and characterization of plasmids from different strains of *Methanobacterium thermoformicicum*. Journal of General Microbiology 137:1981.

- Oeffinger M., Leung A., Lamond A., Tollervey D., Lueng A. (2002) Yeast Pescadillo is required for multiple activities during 60S ribosomal subunit synthesis. Rna 8:626.

- Ogawa T., Yamada Y., Kuroda T., Kishi T., Moriya S. (2002) The datA locus predominantly contributes to the initiator titration mechanism in the control of replication initiation in Escherichia coli. Molecular Microbiology 44:1367-1375.

- Ogden G.B., Pratt M.J., Schaechter M. (1988) The replicative origin of the E. coli chromosome binds to cell membranes only when hemimethylated. Cell 54:127-135.

- Ogura T., Wilkinson A.J. (2001) AAA+ superfamily ATPases: common structure and diverse function. Genes to Cells 6:575-597.

- Oh E.Y., Claassen L., Thiagalingam S., Mazur S., Grossman L. (1989) ATPase activity of the UvrA and UvrAB protein complexes of the Escherichia coli UvrABC endonuclease. Nucleic Acids Research 17:4145.

- Okanami M., Meshi T., Iwabuchi M. (1998) Characterization of a DEAD-box ATPase/RNA helicase protein of *Arabidopsis thaliana*. Nucleic Acids Research 26:2638.

- Owttrim G.W. (2006) RNA helicases and abiotic stress. Nucleic Acids Research 34:3220.

- Oyama T., Oka H., Mayanagi K., Shirai T., Matoba K., Fujikane R., Ishino Y., Morikawa K. (2009) Atomic structures and functional implications of the archaeal RecQ-like helicase Hjm. BMC Structural Biology 9:2.

- Pang P.S., Jankowsky E., Planet P.J., Pyle A.M. (2002) The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding. The

EMBO Journal 21:1168-1176.

- Pape T., Meka H., Chen S., Vicentini G., Van Heel M., Onesti S. (2003) Hexameric ring structure of the full-length archaeal MCM protein complex. EMBO reports 4:1079-1083.

- Parker R. (2006) Identifying protein interactions involved in archaeal DNA replication licensing, PhD thesis, University of Bath.

- Patel S.S., Picha K.M. (2000) Structure and function of hexameric helicases 1. Annual Review of Biochemistry 69:651-697.

- Perriere G., Gouy M. (1996) WWW-query: an on-line retrieval system for biological sequence banks. Biochimie 78:364-369.

- Pham X.H., Reddy M.K., Ehtesham N.Z., Matta B., Tuteja N. (2000) A DNA helicase from *Pisum sativum* is homologous to translation initiation factor and stimulates topoisomerase I activity. The Plant Journal 24:219-229.

- Phizicky E.M., Fields S. (1995) Protein-protein interactions: methods for detection and analysis. Microbiology and Molecular Biology Reviews 59:94.

- Poplawski A., Grabowski B., Long S.E., Kelman Z. (2001) The zinc finger domain of the archaeal minichromosome maintenance protein is required for helicase activity. Journal of Biological Chemistry 276:49371.

- Puhler G., Leffers H., Gropp F., Palm P., Klenk H.P., Lottspeich F., Garrett R.A., Zillig W. (1989) Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. Proceedings of the National Academy of Sciences 86:4569.

- Ramirez C., Kopke A.K.E., Yang C., Boeckh T., Matheson A.T. (1993) The structure, function and evolution of archaeal genomes., in: M. Kates, *et al.* (Eds.), The biochemistry of archaea (archaeabacteria). Elsevier Science Publishers B.V., Amsterdam. pp. 439-466.

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

- Reeve J.N., Sandman K., Daniels C.J. (1997) Archaeal histones, nucleosomes, and transcription initiation. Cell 89:999-1002.

- Robinson N.P., Bell S.D. (2005) Origins of DNA replication in the three domains of life. FEBS Journal 272:3757-3766.

- Robinson N.P., Dionne I., Lundgren M., Marsh V.L., Bernander R., Bell S.D. (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. Cell 116:25-38.

- Rocak S., Emery B., Tanner N.K., Linder P. (2005) Characterization of the ATPase and unwinding activities of the yeast DEAD-box protein Has1p and the analysis of the roles of the conserved motifs. Nucleic Acids Research 33:999.

- Rocak S., Linder P. (2004) DEAD-box proteins: the driving forces behind RNA metabolism. Nature Reviews Molecular Cell Biology 5:232-241.

- Rodamilans B., Montoya G. (2007) Expression, purification, crystallization and preliminary X-ray diffraction analysis of the DDX3 RNA helicase domain. Acta Crystallographica Section F: Structural Biology and Crystallization Communications 63:283-286.

- Rodriguez A.C. (2002) Studies of a positive supercoiling machine: nucleotide hydrolysis and a multifunctional 'latch in' the mechanism of reverse gyrase. Journal of Biological Chemistry 277: 29865–29873.

- Rogers G.W. (2002) eIF4A: the godfather of the DEAD-box helicases. Progress in Nucleic Acid Research and Molecular Biology 72:307-331.

- Ron E., Rozenhak S., Grossman N. (1975) Synchronization of cell division in Escherichia coli by amino acid starvation: strain specificity. Journal of Bacteriology 123:374.

- Rothenberg E., Trakselis M.A., Bell S.D., Ha T. (2007) MCM forked substrate specificity involves dynamic interaction with the 5 -tail. Journal of Biological Chemistry 282:34229.

- Rowlands T., Baumann P., Jackson S.P. (1994) The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. Science 264:1326.

- Rozen F., Edery I., Meerovitch K., Dever T., Merrick W., Sonenberg N. (1990) Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. Molecular and Cellular Biology 10:1134.

- Rozen F., Pelletier J., Trachsel H., Sonenberg N. (1989) A lysine substitution in the

ATP-binding site of eucaryotic initiation factor 4A abrogates nucleotide-binding activity. Molecular and Cellular Biology 9:4061.

- Rudolph M.G., Heissmann R., Wittmann J.G., Klostermeier D. (2006) Crystal structure and nucleotide binding of the *Thermus thermophilus* RNA helicase Hera N- terminal domain. Journal of Molecular Biology 361:731-743.

- Ryan D.P., Duncan J.L., Lee C., Kuchel P.W., Matthews J.M. (2008) Assembly of the oncogenic DNA binding complex LMO2 Ldb1 TAL1 E12. Proteins: Structure, Function, and Bioinformatics 70:1461-1474.

- Sakakibara N., Kelman L.M., Kelman Z. (2009) Unwinding the structure and function of the archaeal MCM helicase. Molecular Microbiology 72:286-296.

- Salzberg S.L., Salzberg A.J., Kerlavage A.R., Tomb J.F. (1998) Skewed oligomers and origins of replication. Gene 217:57-67.

- Sambrook J., Russell D.W. (2001) Molecular cloning: a laboratory manual. CSHL press.

- Sanan-Mishra N., Pham X.H., Sopory S.K., Tuteja N. (2005) Pea DNA helicase 45 overexpression in tobacco confers high salinity tolerance without affecting yield. Proceedings of the National Academy of Sciences of the United States of America 102:509.

- Sandbeck K.A., Leigh J.A. (1991) Recovery of an integration shuttle vector from tandem repeats in *Methanococcus maripaludis*. Applied and Environmental Microbiology 57:2762.

- Sandman K., Krzycki J.A., Dobrinski B., Lurz R., Reeve J.N. (1990) HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. Proceedings of the National Academy of Sciences 87:5788.

- Saraste M., Sibbald P.R., Wittinghofer A. (1990) The P-loop-a common motif in ATP-and GTP-binding proteins. Trends in Biochemical Sciences 15:430-434.

- Saxena S., Dutta A. (2005) Geminin-Cdt1 balance is critical for genetic stability. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 569:111-121.

- Schaper S., Messer W. (1995) Interaction of the initiator protein DnaA of

Escherichia coli with its DNA target. Journal of Biological Chemistry 270:17622.

- Schmid S., Linder P. (1992) D E A D protein family of putative RNA helicases. Molecular Microbiology 6:283-292.

- Schmidt U., Lehmann K., Stahl U. (2002) A novel mitochondrial DEAD-box protein (Mrh4) required for maintenance of mtDNA in *Saccharomyces cerevisiae*. FEMS yeast research 2:267-276.

- Schutz P., Karlberg T., van den Berg S., Collins R., Lehti L., Hogbom M., Holmberg-Schiavone L., Tempel W., Park H.W., Hammarstrum M. (2010) Comparative structural analysis of human DEAD-box RNA helicases. PloS one 5:e12791.

- Seki M., Narusaka M., Abe H., Kasuga M., Yamaguchi-Shinozaki K., Carninci P., Hayashizaki Y., Shinozaki K. (2001) Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. The Plant Cell Online 13:61.

- Sekimizu K., Bramhill D., Kornberg A. (1987) ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the E. coli chromosome. Cell 50:259-265.

- Sengoku T., Nureki O., Nakamura A., Kobayashi S., Yokoyama S. (2006) Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. Cell 125:287-300.

- Seufert D.W., Kos R., Erickson C.A., Swalla B.J. (2000) p68, a DEAD-box RNA helicase, is expressed in chordate embryo neural and mesodermal tissues. Journal of Experimental Zoology 288:193-204.

- Shechter D.F., Ying C.Y., Gautier J. (2000) The intrinsic DNA helicase activity of *Methanobacterium thermoautotrophicum deltaH* minichromosome maintenance protein. Journal of Biological Chemistry 275:15049.

- Sheu Y.J., Stillman B. (2006) Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. Molecular Cell 24:101-113.

- Shi H., Cordin O., Minder C.M., Linder P., Xu R.M. (2004) Crystal structure of the human ATP-dependent splicing and export factor UAP56. Proceedings of the National Academy of Sciences of the United States of America 101:17628.

- Shimada Y., Fukuda W., Akada Y., Ishida M., Nakayama J., Imanaka T., Fujiwara S. (2009) Property of cold inducible DEAD-box RNA helicase in hyperthermophilic archaea. Biochemical and Biophysical Research Communications 389:622-627.

- Shin J.H., Grabowski B., Kasiviswanathan R., Bell S.D., Kelman Z. (2003) Regulation of minichromosome maintenance helicase activity by Cdc6. Journal of Biological Chemistry 278:38059.

- Shin J.H., Heo G.Y., Kelman Z. (2008) The *Methanothermobacter thermautotrophicus* Cdc6-2 protein, the putative helicase loader, dissociates the minichromosome maintenance helicase. Journal of Bacteriology 190:4091.

- Shin J.H., Heo G.Y., Kelman Z. (2009) The *Methanothermobacter thermautotrophicus* MCM helicase is active as a hexameric ring. Journal of Biological Chemistry 284:540.

- Shin J.H., Santangelo T.J., Xie Y., Reeve J.N., Kelman Z. (2007) Archaeal minichromosome maintenance (MCM) helicase can unwind DNA bound by archaeal histones and transcription factors. Journal of Biological Chemistry 282:4908.

- Siddiqui K., Stillman B. (2007) ATP-dependent assembly of the human origin recognition complex. Journal of Biological Chemistry 282:32370.

- Siebers B., Schonheit P. (2005) Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. Current Opinion in Microbiology 8:695-705.

- Singleton M.R., Morales R., Grainge I., Cook N., Isupov M.N., Wigley D.B. (2004) Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. Journal of Molecular Biology 343:547-557.

- Singleton M.R., Dillingham M.S., Wigley D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. Annu. Rev. Biochem. 76:23-50.

- Singleton M.R., Wigley D.B. (2002) Modularity and specialization in superfamily 1 and 2 helicases. Journal of Bacteriology 184:1819.

- Skarstad K., Boye E., Steen H.B. (1986) Timing of initiation of chromosome replication in individual *Escherichia coli* cells. The EMBO Journal 5:1711.

- Slabinski L., Jaroszewski L., Rychlewski L., Wilson I.A., Lesley S.A., Godzik A. (2007) XtalPred: a web server for prediction of protein crystallizability. Bioinformatics

23:3403.

- Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H., Dubois J., Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K. (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. Journal of Bacteriology 179:7135.

- Smith D.W., Garland A.M., Herman G., Enns R.E., Baker T.A., Zyskind J.W. (1985) Importance of state of methylation of *oriC* GATC sites in initiation of DNA replication in Escherichia coli. The EMBO Journal 4:1319.

- Snyder M., He W., Zhang J.J. (2005) The DNA replication factor MCM5 is essential for Stat1-mediated transcriptional activation. Proceedings of the National Academy of Sciences of the United States of America 102:14539.

- Soultanas P., Wigley D.B. (2000) DNA helicases: inching forward. Current Opinion in Structural Biology 10:124-128.

- Soultanas P., Wigley D.B. (2001) Unwinding the Gordian knot of helicase action. Trends in biochemical sciences 26:47-54.

- Speck C., Chen Z., Li H., Stillman B. (2005) ATPase-dependent, cooperative binding of ORC and Cdc6p to origin DNA. Nature Structural & Molecular Biology 12:965.

- Speck C., Messer W. (2001) Mechanism of origin unwinding: sequential binding of DnaA to double-and single-stranded DNA. The EMBO Journal 20:1469-1476.

- Spellman P.T., Sherlock G., Zhang M.Q., Iyer V.R., Anders K., Eisen M.B., Brown P.O., Botstein D., Futcher B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Molecular Biology of the Cell 9:3273.

- Starich M.R., Sandman K., Reeve J.N., Summers M.F. (1996) NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. Journal of Molecular Biology 255:187-203.

- Stevenson R.J., Hamilton S.J., MacCallum D.E., Hall P.A., Fuller Pace F.V. (1998) Expression of the DEAD-box RNA helicase p68 is developmentally and growth regulated and correlates with organ differentiation/maturation in the fetus. The Journal of Pathology 184:351-359.

- Stillman B. (1994) Smart machines at the DNA replication fork. Cell 78:725-728.

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

- Story R.M., Li H., Abelson J.N. (2001) Crystal structure of a DEAD-box protein from the hyperthermophile *Methanococcus jannaschii*. Proceedings of the National Academy of Sciences 98:1465.

- Story R.M., Steitz T.A. (1992) Structure of the recA proteinñADP complex. Nature 355:374-376.

- Studier F.W. (2005) Protein production by auto-induction in high-density shaking cultures. Protein expression and purification 41:207-234.

- Suíetsugu M., Takata M., Kubota T., Matsuda Y., Katayama T. (2004) Molecular mechanism of DNA replication coupled inactivation of the initiator protein in *Escherichia coli*: interaction of DnaA with the sliding clamp loaded DNA and the sliding clamp Hda complex. Genes to Cells 9:509-522.

- Sung P., Prakash L., Matson S.W., Prakash S. (1987) RAD3 protein of *Saccharomyces cerevisiae* is a DNA helicase. Proceedings of the National Academy of Sciences 84:8951.

- Svitkin Y.V., Pause A., Haghighat A., Pyronnet S., Witherell G., Belsham G.J., Sonenberg N. (2001) The requirement for eukaryotic initiation factor 4A (elF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. Rna 7:382.

- Tai C.L., Chi W.K., Chen D.S., Hwang L.H. (1996) The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). Journal of Virology 70:8477.

- Takahashi T.S., Wigley D.B., Walter J.C. (2005) Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. Trends in Biochemical Sciences 30:437-444.

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

- Tanaka S., Diffley J.F.X. (2002) Deregulated G1-cyclin expression induces genomic instability by preventing efficient pre-RC formation. Genes & Development 16:2639.

- Tanaka S., Umemori T., Hirai K., Muramatsu S., Kamimura Y., Araki H. (2006) CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. Nature 445:328-332.

- Tanaka T., Knapp D., Nasmyth K. (1997) Loading of an Mcm protein onto DNA replication origins is regulated by Cdc6p and CDKs. Cell 90:649-660.

- Tanner N.K., Cordin O., Banroques J., Doere M., Linder P. (2003) The Q Motif:: A Newly Identified Motif in DEAD-box Helicases May Regulate ATP Binding and Hydrolysis. Molecular Cell 11:127-138.

- Tanner N.K., Linder P. (2001) DExD/H Box RNA Helicases: From Generic Motors to Specific Dissociation Functions. Molecular Cell 8:251-262.

- Thauer R.K., Kaster A.K., Seedorf H., Buckel W., Hedderich R. (2008) Methanogenic archaea: ecologically relevant differences in energy conservation. Nature Reviews Microbiology 6:579-591.

- Theis K., Chen P.J., Skorvaga M., Van Houten B., Kisker C. (1999) Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. The EMBO journal 18:6899-6907.

- Thom N.H., Czyzewski B.K., Alexeev A.A., Mazin A.V., Kowalczykowski S.C., Pavletich N.P. (2005) Structure of the SWI 2/SNF 2 chromatin-remodeling domain of eukaryotic Rad 54. Nature Structural & Molecular Biology 12:350-356.

- Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25:4876.

- Tsaneva I.R., M¸ller B., West S.C. (1993) RuvA and RuvB proteins of Escherichia coli exhibit DNA helicase activity in vitro. Proceedings of the National Academy of Sciences 90:1315.

- Tseng S.S., Weaver P.L., Liu Y., Hitomi M., Tartakoff A.M., Chang T.H. (1998) Dbp5p, a cytosolic RNA helicase, is required for poly (A)+ RNA export. The EMBO Journal 17:2651.

- Tsu C.A., Kossen K., Uhlenbeck O.C. (2001) The *Escherichia coli* DEAD protein DbpA recognizes a small RNA hairpin in 23S rRNA. RNA 7:702.

- Tuteja N., Tuteja R. (2004) Prokaryotic and eukaryotic DNA helicases. European

Journal of Biochemistry 271:1835-1848.

- Tye B.K., Sawyer S. (2000) The hexameric eukaryotic MCM helicase: building symmetry from non identical parts. Journal of Biological Chemistry 275:34833.

- Uemori T., Sato Y., Kato I., Doi H., Ishino Y. (1997) A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. Genes to Cells 2:499-512.

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

- Vashisht A.A., Pradhan A., Tuteja R., Tuteja N. (2005) Cold and salinity stress induced bipolar pea DNA helicase 47 is involved in protein synthesis and stimulated by phosphorylation with protein kinase C. The Plant Journal 44:76-87.

- Vashisht A.A., Tuteja N. (2005) Cold stress-induced pea DNA helicase 47 is homologous to eIF4A and inhibited by DNA-interacting ligands. Archives of Biochemistry and Biophysics 440:79-90.

- Vashisht A.A., Tuteja N. (2006) Stress responsive DEAD-box helicases: a new pathway to engineer plant stress tolerance. Journal of Photochemistry and Photobiology B: Biology 84:150-160.

- Velankar S.S., Soultanas P., Dillingham M.S., Subramanya H.S., Wigley D.B. (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell 97:75-84.

- Verhees C.H., Kengen S.W.M., Tuininga J.E., Schut G.J., Adams M.W.W., De Vos W.M., Van Der Oost J. (2003) The unique features of glycolytic pathways in Archaea. Biochemical Journal 375:231.

- Vignali M., Hassan A.H., Neely K.E., Workman J.L. (2000) ATP-dependent chromatin-remodeling complexes. Molecular and Cellular Biology 20:1899.

- Voloshin O.N., Camerini-Otero R.D. (2007) The DinG protein from Escherichia coli is a structure-specific helicase. Journal of Biological Chemistry 282:18437.

- Von Freiesleben U., Rasmussen K.V., Schaechter M. (1994) SeqA limits DnaA activity in replication from *oriC* in *Escherichia coli*. Molecular Microbiology 14:763.

- Wadsworth R.I.M., White M.F. (2001) Identification and properties of the

crenarchaeal single-stranded DNA binding protein from *Sulfolobus solfataricus*. Nucleic Acids Research 29:914.

- Wahle E., Lasken R., Kornberg A. (1989) The dnaB-dnaC replication protein complex of *Escherichia coli*. II. Role of the complex in mobilizing dnaB functions. Journal of Biological Chemistry 264:2469.

- Walker J.E., Saraste M., Runswick M.J., Gay N.J. (1982) Distantly related sequences in the alpha-and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. The EMBO Journal 1:945.

- Walter J., Newport J. (2000) Initiation of Eukaryotic DNA Replication: Origin Unwinding and Sequential Chromatin Association of Cdc45, RPA, and DNA Polymerase [alpha]. Molecular Cell 5:617-627.

- Walter T.S., Meier C., Assenberg R., Au K.F., Ren J., Verma A. (2006) Lysine methylation as a routine rescue strategy for protein crystallization. Structure 14:1617-1622.

- Walters A. (2009) The role of multiple minichromosome maintenance proteins in *Methanococcus maripaludis*, PhD thesis, Biology, University of York.

- Walters A.D., Smith S.E., Chong J.P.J. (2011) A shuttle vector system with improved transformation efficiency in *Methanococcus maripaludis*. Applied and Environmental Microbiology. 02919-10.

- Wang S., Hu Y., Overgaard M.T., Karginov F.V., Uhlenbeck O.C., McKay D.B. (2006) The domain of the Bacillus subtilis DEAD-box helicase YxiN that is responsible for specific binding of 23S rRNA has an RNA recognition motif fold. Rna 12:959.

- Warren, E. M., Vaithiyalingam, S., Haworth, J., Greer, B., Bielinsky, K., Chazin, W., Eichman, B. (2008) Structural Basis for DNA Binding by Replication Initiator Mcm10. Struture 16(12):1892-1901.

- Weber C.A., Salazar E.P., Stewart S., Thompson L. (1990) ERCC2: cDNA cloning and molecular characterization of a human nucleotide excision repair gene with high homology to yeast RAD3. The EMBO Journal 9:1437.

- Weigel C., Schmidt A., R¸ckert B., Lurz R., Messer W. (1997) DnaA protein binding to individual DnaA boxes in the *Escherichia coli* replication origin, *oriC*. The EMBO

Journal 16:6574-6583.

- Weinberg D.H., Collins K.L., Simancek P., Russo A., Wold M.S., Virshup D.M., Kelly T.J. (1990) Reconstitution of simian virus 40 DNA replication with purified proteins. Proceedings of the National Academy of Sciences of the United States of America 87:8692.

- Weinreich M., Stillman B. (1999) Cdc7pñDbf4p kinase binds to chromatin during S phase and is regulated by both the APC and the RAD53 checkpoint pathway. The EMBO Journal 18:5334-5346.

- Weiss D.S., Thauer R.K. (1993) Methanogenesis and the unity of biochemistry. Cell 72:819-822.

- Werner F. (2007) Structure and function of archaeal RNA polymerases. Molecular Microbiology 65:1395-1404.

- Werner F., Weinzierl R.O.J. (2002) A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. Molecular Cell 10:635-646.

- Westergaard O., Brutlag D., Kornberg A. (1973) Initiation of Deoxyribonucleic Acid Synthesis. Journal of Biological Chemistry 248:1361.

- Wettach J., Gohl H.P., Tschochner H., Thomm M. (1995) Functional interaction of yeast and human TATA-binding proteins with an archaeal RNA polymerase and promoter. Proceedings of the National Academy of Sciences 92:472.

- Wilmes G.M., Archambault V., Austin R.J., Jacobson M.D., Bell S.P., Cross F.R. (2004) Interaction of the S-phase cyclin Clb5 with an 'RXL' docking sequence in the initiator protein Orc6 provides an origin-localized replication control switch. Genes & Development 18:981.

- Woese C.R. (2004) A new biology for a new century. Microbiology and Molecular Biology Reviews 68:173.

- Woese C.R., Fox G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proceedings of the National Academy of Sciences 74:5088.

- Woese C.R., Kandler O., Wheelis M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences 87:4576.

- Woese C.R., Magrum L.J., Fox G.E. (1978) Archaebacteria. Journal of Molecular

Evolution 11:245-252.

- Woese C.R., Magrum L.J., Gupta R., Siegel R.B., Stahl D.A., Kop J., Crawford N., Brosius R., Gutell R., Hogan J.J. (1980) Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. Nucleic Acids Research 8:2275.

- Wong H.Y., Demmers J.A.A., Bezstarosti K., Grootegoed J.A., Brinkmann A.O. (2009) DNA dependent recruitment of DDX17 and other interacting proteins by the human androgen receptor. Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics 1794:193-198.

- Xia Q., Wang T., Hendrickson E., Lie T., Hackett M., Leigh J. (2009) Quantitative proteomics of nutrient limitation in the hydrogenotrophic methanogen *Methanococcus maripaludis.* BMC Microbiology 9:149.

- Yan X., Mouillet J.F., Ou Q., Sadovsky Y. (2003) A novel domain within the DEAD-box protein DP103 is essential for transcriptional repression and helicase activity. Molecular and Cellular Biology 23:414.

- Yang Q., Jankowsky E. (2006) The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. Nature Structural & Molecular Biology 13:981-986.

- Yao N., Hesson T., Cable M., Hong Z., Kwong A., Le H., Weber P.C. (1997) Structure of the hepatitis C virus RNA helicase domain. Nature Structural & Molecular Biology 4:463-467.

- Yoshimochi T., Fujikane R., Kawanami M., Matsunaga F., Ishino Y. (2008) The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. Journal of Biological Chemistry 283:1601.

- Yu E., Owttrim G.W. (2000) Characterization of the cold stress-induced cyanobacterial DEAD-box protein CrhC as an RNA helicase. Nucleic Acids Research 28:3926.

- Yu X., VanLoock M.S., Poplawski A., Kelman Z., Xiang T., Tye B.K., Egelman E.H. (2002) The *Methanobacterium thermoautotrophicum* MCM protein can form heptameric rings. EMBO reports 3:792-797.

- Yutin N., Makarova K.S., Mekhedov S.L., Wolf Y.I., Koonin E.V. (2008) The deep

archaeal roots of eukaryotes. Molecular Biology and Evolution 25:1619.

- Zegerman P., Diffley J.F.X. (2006) Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. Nature 445:281-285.

- Zeikus J.G., Wolee R.S. (1972) Methanobacterium thermoautotrophicus sp. n., an anaerobic, autotrophic, extreme thermophile. Journal of Bacteriology 109:707.

- Zhang J.K., White A.K., Kuettner H.C., Boccazzi P., Metcalf W.W. (2002) Directed mutagenesis and plasmid-based complementation in the methanogenic archaeon Methanosarcina acetivorans C2A demonstrated by genetic analysis of proline biosynthesis. Journal of Bacteriology 184:1449.

- Zhang L., Liu Y., Yang S., Gao C., Gong H., Feng Y., He Z.G. (2009) Archaeal eukaryote-like Orc1/Cdc6 initiators physically interact with DNA polymerase B1 and regulate its functions. Proceedings of the National Academy of Sciences 106:7792.

- Zhang R., Zhang C.T. (2004) Identification of replication origins in the genome of the methanogenic archaeon, *Methanocaldococcus jannaschii*. Extremophiles 8:253-258.

- Zhou C., Huang S.H., Jong A.Y. (1989) Molecular cloning of *Saccharomyces cerevisiae* CDC6 gene. Isolation, identification, and sequence analysis. Journal of Biological Chemistry 264:9022.

- Zillig W. (1991) Comparative biochemistry of Archaea and Bacteria. Current Opinion in Genetics & Development 1:544-551.

- Zillig W., Stetter K.O., Tobien M. (1978) DNA Dependent RNA Polymerase from *Halobacterium halobium*. European Journal of Biochemistry 91:193-199.

- Zuckerkandl E., Pauling L. (1965) Evolutionary divergence and convergence in proteins, Academic Press, New York. pp. 166.

- Zyskind J.W., Smith D.W. (1992) DNA replication, the bacterial cell cycle, and cell growth. Cell 69:5.

Appendix A: Primers

Oligo name	Sequence (5'-3')
HS1	GGGACGCGTCGGCCTGGCACGTCGGCCGCTGCGGCCAG GCACCCGATGGCGTTTGTTTGTTTGTTTGTTTGTTT
HS2	TTTGTTTGTTTGTTTGTTTGTTTGTTTGGCCGACGTGCCAG GCCGACGCGTC
RNA A	CCATGGTGGATTGTAAATTGTAATGTTGTTGCGCGCGCGC TATAGTGAGTCGTATTA
RNA B	CGCGCGCGCGATTGTAAATTGTAATGTTGTTCCATGGTCC TATAGTGAGTCGTATTA
T7 promoter	TAATACGACTCACTATAG
Mth203startNhe1	GGAAGCTAGCATGAAAGGATTAGAATTTAGTGAG
Mth203C53stopHind3	CTCAAAAGCTTATTACTCCTCAGGGGAGGG
Mmp0457startAsc1	CTCAAGGCGCGCCTTAGTTGCCGCTTG
Mmp0457stopAsc1	GGAAGCCGGCATGGAAAGTTTTAAAAATTTAGG
Mmp0457-A	ACGTTATGTAACAACATTACAAAATGAGTAACTT
Mmp0457-B	TTAGTTGCCGCTTGCACTTCGGAT
Mmp0457-C	TTCCTGATGTTAAAGATATAATCAGCGCTAAAA
Mmp0457-D	AAAACTTTCCATAGTATATCCTCTTTTTTGAATTTTAC
Mmp0457-E	GAGTCTTACGTTTTCACCAGGT
Mmp0457F	GCCCTATAGTGAGTCGTA
Mmp0457G	CCAGTTTGGAACAAGAGTCC
4ORBspF	TAATACTAACTTACACTTGAAATG
4ORBspR	TGTGAACTTTATCTACAGGGTC
4ORBnspF	GGTGATTTAATGGCAGCAATAG
4ORBnspR	GAGTTCAGTCTACTATAACGCC
4ORBnsp2F	TTTTCTTATGCAGAAGACATTTATATA
4ORBnsp2R	GACCACAGCCACTCTTGTC

Appendix B: Strains

1. *E. coli*
- Novablue (Novagen)
  - Genotype: endA1 hsdR17 (r_{K12} $^-$$m_{K12}$ $^+$) supE44 thi-1 recA1 gyrA96 relA1 lacF'[proA$^+$ B$^+$ lacI^{q}ZΔM15::Tn10] (Tet^R)
- BL21 (DE3) pLysS (Merck)
  - Genotype: F$^-$ ompT gal dcm lon hsdS$_B$(r_B^- m_B^-) λ(DE3) pLysS(cm^R)
- BL21 (DE3) Star Rosetta 2 (Merck)
  - Genotype: F$^-$ ompT gal dcm lon hsdS$_B$(r_B^- m_B^-) rne131 λ(DE3) pRARE2(cm^R)

2. *M. maripaludis*
- MM900 (Moore and Leigh, 2002)
  - Genotype: *M. maripaludis* S2 Δ*hpt*
- S0001 (Walters *et al.,* 2011)
  - Genotype: MM900 with pAW42 ORF1 sequence integrated in *upt* gene

Appendix C: Oligonucleotides for anisotropy

The sequence of oligonucleotides used for anisotropy assays are described below. The origin recognition boxes (ORB) are highlighted in grey. For the assay, the oligos were labelled with Oregon green at 5'.

a) 4ORB specific substrate (containing ORB 7-10) 205 bp
5'- TAA TAC TAA C TT ACA CTT GAA AT G AAT GTC TCC CTT ACA GGT CAT CAG AAC CAT GGT CAG ATT ACA CTT GAA ATG GAT GTC TCC CAC ATC TAG CCA TGA ATC AGA GAA CTG GAT AAG GAA CAG CAG GAT TTT TTA CAC TTG AAA TTC ATC CCT CAT GAA TTC CCA TCG AGG ATC CAG ATG GTT ACA CTT GAA ATA GAT GTC CCA C -3'

b) 4ORB non-specific substrate 205 bp
5'- GGT GAT TTA ATG GCA GCA ATA GAA GTT GGA AGA GTA TGT GTA AAA ACC GCA GGA AGA GAA GCC GGT GAA AAT GGC GTG ATA CTC GAT ATC ATC GAC AAA AAC TTC GTT GAA GTT GTG GGT GTT AAC GTT AAA AAT AGA AGG TGC AAC GTG AGC CAC CTC GAA CCC ACT GAG AAT AAG ATA GAA CTC AAG TCA GAT GAT ATT GAG G -3'

c) 4ORB non-specific substrate-2 205 bp
5'- TTT TCT TAT GCA GAA GAC ATT TAT ATA TTT ATT CGC CGT CAT CTA TTA AAA AGC TGA ATC TTA TAA TAC TCA TGA CCA TAT ATT AAT CTT GCT GCC TGC GCC GTA ACT CTC TAA GAC CTG ATC ATG GAA TAT GGA GAT AAG CCT GTG AAA TGT ATG GAA GTG AGG GTG GAG CCA TGA GGA TAG TGG CTG GTG TCG GTG AGA ACA G -3'

d) Single ORB specific substrate (ORB8) 34 bp
5'- CAT GGT CAG ATT ACA CTT GAA ATG GAT GTC TCC C -3'

e) Single ORB non-specific substrate 34 bp
5'- ACG TAC TGA CCA GTT GAG TTC TAA TGC CAT GGA A -3'