ABSTRACT

Title of Document: *THERMOCOCCUS KODAKARENSIS* DNA REPLICATION MACHINERY

Miao Pan, Ph.D., 2012

Directed By: Professor Zvi Kelman,
Institute for Bioscience and Biotechnology Research

DNA replication is the basis for the propagation and evolution of living organisms. It requires the combined efforts of numerous proteins. DNA replication in archaea has been shown to be more similar to eukarya than bacteria. Therefore, we use archaea as a model to study DNA replication.

Euryarchaeon is one of the five main branches of archaeon. In this thesis, the replication machinery of the thermophilic euryarchaeon *Thermococcus kodakarensis* was investigated. In particular, this work focuses on two essential DNA replication proteins, the minichromosome maintenance (MCM) helicase and the processivity factor, proliferating cell nuclear antigen (PCNA).

The MCM complex is thought to function as the replicative helicase in

archaea and eukaryotes. In most archaea, one MCM homolog assembles to form the active homohexameric complex. Atypically, the genome of *T. kodakarensis* encodes three MCM homologs, here designated MCM1-3. Although all three MCM exhibit helicase activity, DNA binding and ATPase activities, only MCM3 appears to be essential for cell viability. Taken together with bioinformatics analysis, the results suggest that MCM3 is the replicative helicase in *T. kodakarensis*.

PCNA is a ring shaped protein that encircles duplex DNA and, upon binding to the polymerase and other proteins, tethers them to the DNA. All euryarchaeal genomes, except *T. kodakarensis*, encode for a single PCNA protein. *T. kodakarensis* is unique because it contains two genes encoding for PCNA1 and PCNA2. It is shown here that both PCNA proteins stimulate DNA polymerase activity. It was found that PCNA1 is expressed *in vivo* at high levels in comparison to PCNA2. Furthermore, it was determined that PCNA2 is dispensable for cell viability.

Taking together, the data presented herein suggest that *T. kodakarensis* is similar to other archaeal species studied, requiring only one MCM and one PCNA protein for viability. The results obtained from this work provide essential knowledge about the replication machinery in eukarya.

# *THERMOCOCCUS KODAKARENSIS* DNA REPLICATION MACHINERY

By

Miao Pan.

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor Zvi Kelman, Chair
Professor James Culver
Professor Jeffery Destefano
Professor Douglas Julin
Professor John Moult
Professor Brian Bequette, Dean's representative

# Dedication

I dedicate the accomplishments of my learning to my parents and my wife.

# Acknowledgment

I would like to express my profound gratitude to my advisor Dr. Zvi Kelman, whose guidance and encouragement have been a great source of inspiration through my Ph.D. study. His dedication and attitude to science have taught me very important lesson not only in my study, but also in my life. I feel very fortunate to have an opportunity to work together with him. His patience and enthusiasm always guide me to work optimistically towards my future endeavors.

I am especially grateful to my committee members who have exhibited much kindness and have given valuable suggestions in the course of my Ph.D. study. I thank Dr. Douglas Julin for his "Nucleic acids" course inspiring me with DNA replication. I also appreciate Drs. Jeffrey DeStefano, James Culver, John Moult and Brian Bequette who spend the time talking about experiment design and career development. I learned a great deal from their expertise and insights.

I would also like to thank the members of the Kelman lab – current and past – Nozomi, Roxane and Zhou gave me a lot of help physically and spiritually. Also, I would like to thank our helpful undergraduate and high school students, Carol, Diane, Tobias, Harlett and Charles for their help. And, I would like to thank Dr. Lori Kelman who took a lot of time to proofread the manuscripts and provide valuable suggestions on my presentations.

I would also like to thank Drs. John Marino, Jane Ladner and Daniel Nelson.

I also made many friends during my Ph.D. study at Maryland. Thanks for

all the help from the members from John Marino group, John Moult group and “the third floor meeting group”. Also thanks for my friends I made when I was in college park, although many of them have already moved to other places. Their friendship always made me feel supported.

Lastly, I would like take this opportunity to thank my parents and my wife. Without their encouragement, I could not finish this work.

## TABLE OF CONTENTS

# List of tables

# List of figures

# List of abbreviations

| | |
|---|---|
| 6MP | 6-methyl purine |
| AAA+ | ATPases associated with various cellular activities |
| ASW | Artificial sea water |
| bp | Base pair |
| BSA | Bovine serum albumin |
| CDC6 | Cell division cycle 6 |
| CDC45 | Cell division cycle 45 |
| Da | Dalton |
| dNTP | Deoxynucleotide triphosphate |
| dsDNA | Double-stranded DNA |
| EDTA | Ethylenediaminetetraacetic acid |
| Fen1 | Flap endonuclease 1 |
| FPA | Fluorescence polarization anisotropy |
| IR | Inverted repeat |
| IPTG | Isopropyl-β-D-thiogalactopyranoside |
| MCM | Minichromosome maintenance |
| NTP | Nucleotide triphosphate |
| ORB | Origin recognition box |
| ORC | Origin recognition complex |
| PCNA | Proliferating cell nuclear antigen |
| PIP motif | PCNA-interacting protein motif |

| | |
|---|---|
| PolB | DNA polymerase B |
| PolD | DNA polymerase D |
| RFC | Replication factor C |
| RPA | Replication protein A |
| SDS | Sodium dodecyl sulfate |
| SDS-PAGE | SDS-Polyacrylamide gel electrophoresis |
| SSB | Single strand DNA binding protein |
| ssDNA | Single-stranded DNA |

# Chapter 1 Introduction

## 1.1 Overview

DNA replication is the basis for the propagation and evolution of living organisms. The mechanism of DNA replication guarantees the duplication and transfer of genetic information during cell division. The process needs to be precise during the cell cycle. Our studies will contribute to the understanding of the mechanisms controlling DNA replication and its regulation in archaea.

DNA replication starts at a specific sequence known as the origin of replication. It has an AT-rich region and contains specific inverted repeats (IRs) that facilitate binding of origin binding proteins (OBPs). OBPs bind to the origin and partially melt the duplex DNA to form a replication bubble. After that, the helicase loader loads the helicase in an ATP-dependent manner. Following the dissociation of the helicase loader from the helicase-DNA complex, the helicase unwinds the duplex DNA into single strand DNA (ssDNA). Single strand binding protein (SSB) covers the unwound ssDNA to protect it from degradation by nuclease. DNA polymerase will use the ssDNA as template to synthesize new DNA. However, polymerase cannot start DNA synthesis *de novo* but needs a primer. Primase binds to the unwound ssDNA and uses it as a template to synthesize a short RNA primer, which is used by the polymerase to synthesize the complementary DNA strand. Sliding clamp is loaded by the clamp loader onto the primer-template junction, and then the clamp will encircle the duplex and tether polymerase onto the DNA to increase the polymerase processivity. This

process occurs once in the leading strand and multiple times in the lagging strand. It has been shown that in bacteria, when the polymerase encounters the 5'-terminus of the previous Okazaki fragment in the lagging strand, the affinity of the polymerase to the sliding clamp is attenuated, which frees the polymerase to dissociate from the Okazaki fragment and allows it to cycle to the new primer-template junction to begin synthesis of the next Okazaki fragment. This process is still not clear in archaea and eukarya, though it is proposed that the eukaryotic and archaeal polymerase can displace the primer of the previous Okazaki fragment. The sliding clamp remaining on the DNA molecule interacts with other proteins needed for Okazaki fragment maturation and regulates their actions in removing RNA primers, synthesizing DNA to the template, and then covalently joining the completed Okazaki fragments into a continuous DNA molecule.

## 1.2 Archaea

The archaea is one of the three domains of life, along with bacteria and eukarya [(1,2), Fig. 1-1]. Archaea often inhabit extreme environments and are classified into five major kingdoms; Euryarchaeota, Crenarchaeota, Korearchaeota, Nanoarchaeota, and Thaumarchaeota (3,4). Since the genome of the first archaea was completed attention has been given to the study of these unique organisms. Phylogenetic analyses of archaeal genomes have revealed that they exhibit a mosaic of features from the other two domains, as well as domain specific features. For example, proliferation processes in archaea are an interesting mix: proteins involved in the biochemical property of the DNA

replication process are simplified versions of those found in eukarya [(5), Table 1], whereas the cell division process appears to be quite similar to the system found in bacteria (6-8).

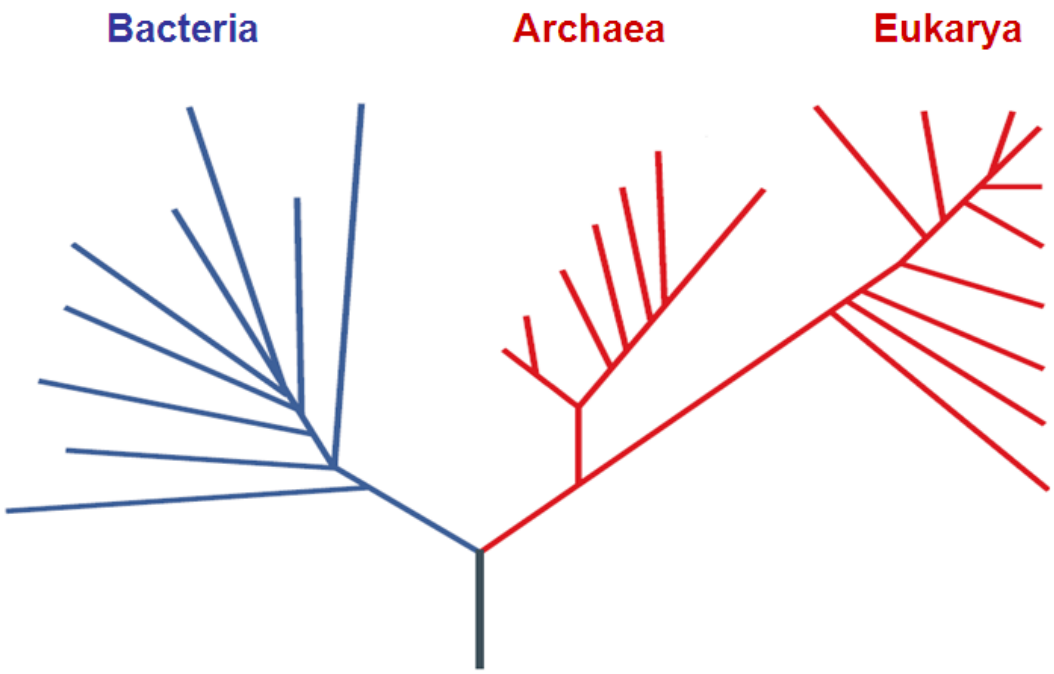



Figure 1-1. Phylogenetic trees of the three domains of life based on 16s rDNA sequences. Adapted from (9)

## 1.3 *Thermococcus kodakarensis*

Euryarchaeaota contains a diverse archaeal group of members, including halophiles, hyperthermophiles, methanogens and thermophilic methanogens. In this study, we used the euryarchaeon *Thermococcus kodakarensis* KOD1 as a model organism to study DNA replication machinery. *T. kodakarensis* was isolated from a solfatara off the shore of Kodakara Island, Kagoshima, Japan (10). It is a hyperthermophilic microorganism with an optimal growth temperature of

Table 1-1. Key components of the DNA replication machinery.

| **Function** | *E. coli* Protein | Subunit Organization | Archaeal Protein | Subunit Organization | Eukaryotic Protein | Subunit Complex |
|---|---|---|---|---|---|---|
| **Origin reorganization** | DnaA | 8-12 mer | ORC/CDC6 | Unknown | ORC1-6 | Heterohexamer |
| **Helicase loading** | DnaC | Homohexamer | Unknown | Unknown | CDC6 | Unknown |
| **Helicase** | DnaB | Homohexamer | MCM | Homohexamer | MCM2-7 | Heterohexamer |
| **Primase** | DnaG | Monomer | DnaG-like Primase | Unknown Heterodimer | Polα/Primase | Heterotetramer |
| **DNA polymerase** | Pol III core | Heterotrimer | Pol B | Monomer | Pol ε | Heterotetramer |
| | | | Pol D | Heterodimer | Pol δ | Heterotetramer |
| **Processivity factor** | β | Homodimer | PCNA | Homotrimer/Heterotrimer[1] | PCNA | Homotrimer |
| **Processivity complex** | γ complex | $\gamma_3\delta\delta'$ | RFC | Heteropentamer | RFC | Heteropentamer |
| **Single strand DNA binding** | SSB | Homotetramer | RPA | Heterotrimer/Monomer | RPA | Heterotrimer |
| **Replisome assembly** | τ | homodimer | GINS[2] | Heterotetramer/Homotetramer[3] | GINS | Heterotetramer |

[1] Euryarchaea contain one PCNA homolg which forms homotrimers; crenarchaea contain three PCNA homologs which forms heterotrimers.

[2] Proposed to function in replisome assemblys.

[3] Some archaeal species contain two GINS homologs which form a heterotetramer, while other contains only one GINS homolog which forms a homotetramer.

85°C and generation time of about 2 hours. *T. kodakarensis* is a strictly anaerobic obligate heterotrophy that grows on complex proteinaceous substrates, and whose growth is strongly associated with the reduction of elemental sulfur (11). *T. kodakarensis* has a single genome of 2.1M bp with 2,306 open reading frames (12). It was previously reported as *Pyrococcus* sp. KOD1. Later detailed phylogenetic analysis indicated that it is a member of the genus *Thermococcus* [(11), Fig. 1-2]. The GC content of *T. kodakarensis* genome is 52.0%, higher than its closest euryarchaeal order *Pyrococcus* genomes (40-50%). Four potential viral integration regions (TKV1-TKV4) have been identified in its genome (12). It is interesting that several DNA replication factors are located in these regions (more detail will be discussed below). Genetic tools generated in *T. kodakarensis* make it an attractive and useful archaeal model organism in the research on hyperthermophiles (13-16).

## 1.4 Archaeal DNA replication

### 1.4.1 The replication origin

DNA replication origin is a particular sequence in the chromosomal DNA where replication is initiated. Archaea was thought to contain a single origin of replication because its chromosome structure is similar to bacteria, which usually contain a single replication origin. Past studies showed that *Pyrococcus abyssi*, *Pyrococcus furiosus, Pyrococcus horikoshii* and *Archeoglobus fulgidus* have only one origin (17,18). However, multiple origins have also been identified in several archaeal species. (19,20). Interestingly, in most cases the identified origins are

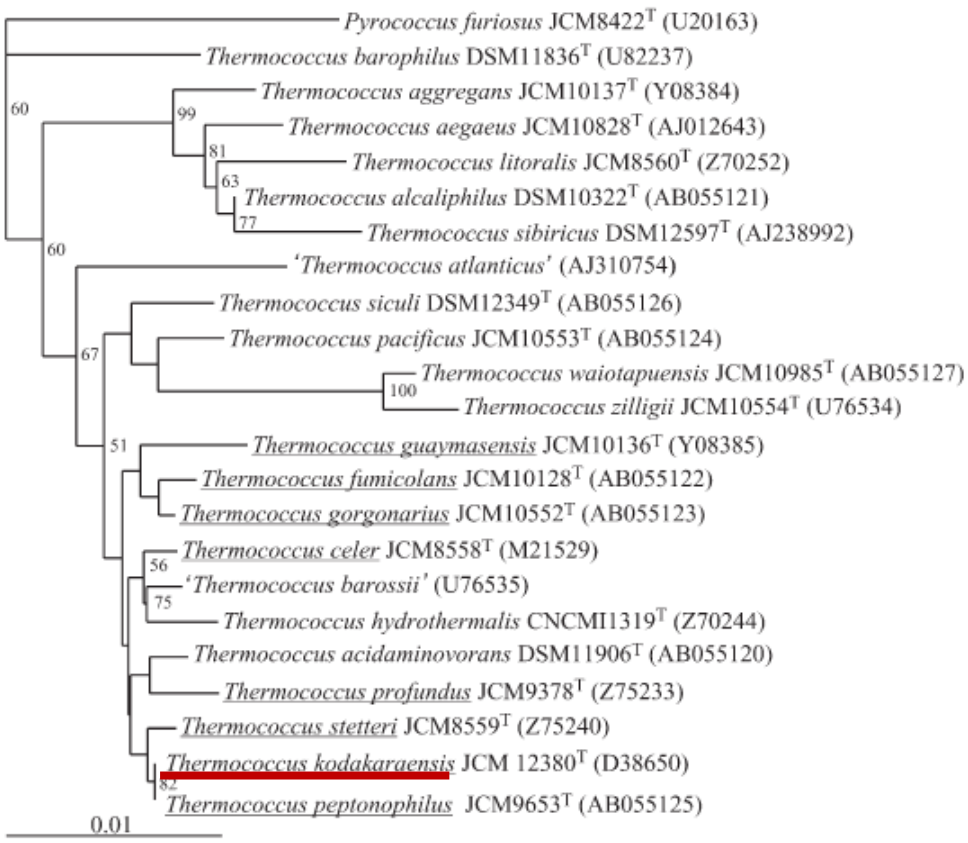



Figure 1-2. Phylogenetic tree showing the KOD1 strain is positioned more close to other *Thermococcus* species than *Pyrococcus* species. *T. kodakarensis* KOD1 is underlined in red. Adapted from (11)

located upstream of the gene encoding for the initiator protein CDC6. Archaeal origins contain a number of origin recognition boxes (ORB). ORB is the functional homolog of the bacterial DnaA boxes, which serve as binding sites for the origin recognition. Structural and biochemical studies indicated that archaeal initiator proteins CDC6 interact with ORB specifically and tightly. Although archaeal origin sequences are highly diverse, conserved inverted reverse (IR) sequence elements are identified in archaeal ORB. This makes archaeal origins

similar to bacterial and eukaryotic origins which are defined by specific sequence elements. Archaeal mini-ORB is defined as TnCAnnnGAAA (where n can be substituted by any nucleotide). The flanking sequences may vary between IRs.

Like most other *Thermococcus* and *Pyrococcus*, *T. kodakarensis* is predicted to have a single origin in its genome and it is located upstream of the operon encoding for CDC6 and Polymerase D (PolD). Although it has not been confirmed experimentally, based on the DNA sequence similarity between *T. kodakarensis* genome sequence and other archaeal sequences, the *T. kodakarensis* origin is predicted as TTCCAgTggAAACggAA (Fig. 1-3).

### 1.4.2 Initiation

Initiator protein CDC6

The function of the initiator protein is to recognize the origin of replication and partially melt the origin duplex region. In bacteria, DnaA are recruited onto the DNA replication origin in the initiation process. In eukaryotes, origin recognition complex (ORC) binds at origins, and then recruits many other proteins, such as CDC6 and MCM onto the origin (Table 1-1). So far, most sequenced archaeal species contain at least one homolog of CDC6. Archaeal CDC6 has a similar amino acid sequence to the eukaryotic CDC6 protein and the subunits of ORC. However, it is still unclear whether the archaeal CDC6 combine the function of ORC and CDC6 in eukaryotic cells. Most archaea contain only one CDC6 gene, some contain multiple copies and several, (e.g. methanogens) do not contain clear CDC6 homolog. The different copy number


(A)
CDC6
PolD-S
PolD-L
TK1900
TK1901
TK1902
TK1903
(B)
M
CATGCAAGGGGTTAAAAACGTTGGGACGTGGTTAAGTGAACTTTAATGAA
+/-
CATAAGAATACCACTAGGCACAAGTTTTCCAGTGGAAACGGAACTCTTTTT
+/-
CGATACTCATCGAAAAGACCCCCCAGAGTTTTATTTCCACTGGAACCAACT
TTCGGAGAATGGATGGATTTGATGGCTTTAAACAACAGAAAGGGTATTTTT
AAAGCGTCTGTGTCAATGTAAATATGATTTTGTTCATTTAAACTTTGTTCTTT
CAGTATTTTTTCAATCTTACATCTTAATACCCGCTAATTATTGAACATTGTTTA
+
TGGAAAATTATAAATACATGATGGGAAAGAAAAAGCGGGAGTGTTTGCAC
+/-
CAAAATTCGATATTGAAAATGGGTGTAGCAACGAAAGCTGCCTCCACTGGA
AACAGAACTCTGGGGGGACTACAAAATTGTTTATTTTTTGAGATTGCGAAAA
CAAAGATGTTTTCTAAAAATGATCCCTTTTGTCCATCATTGTTCGTGTTTTCA
-
+
TGTATATCCCTGTTCATTCCTCTTTCAGTGGAAATAGGCACACCTACCGATG
-
TTTTGTTTGTCTAATGCAAACTTGTATGTGCATCCCTGATCATTGCTTGCTTC
GCGTTTCCTCTGGAGAGGTCTTTTTGCACAGGTTAGAACTGGAACTTCGAG
TGGAATTCTTTGACAGCTCCATGTTCAACAATCCATATAAACCAGCGGCCTA
CACATAGGTATGCCTTCCAGTGTATCCGCTGGTTCCTGTGGAGTCCCTCGTG
AATGTTCAATATTTCCCCATATTTCCTATCAAAATGTTGAAATAAGGATATGT
ACTGGTCATTTCGATGGAAAGCGCACTATCATATTGACGGAACGACTTCCA
M
TCGAGTATGATTTCAAGTGGAGAGTGGGACGATG

Figure 1-3. Putative *T. kodakarensis* origin of replication. (A) Schematic representation of the region flanking the origin (oriC). (B) The sequence of the putative oriC in *T. kodakarensis* is shown. The initiation codons for CDC6 and TK1900 are shown in red. Putative ORB sequences are shown with an arrow (+ or - refer to the orientation of the IRs).

in the archaeal CDC6 copy may indicate gene duplication, gene transfer or heterocomplex formation similar to eukaryotic ORC counterpart. It is also possible that archaeal CDC6 has been evolved into different sequence, which cannot be identified by homology. The fact that the CDC6 gene is usually expressed next to the predicted origin, suggests that the two may be co-regulated.

As a member of the AAA+ protein family, CDC6 contains several conserved motifs for ATP binding and hydrolysis. Crystal structure of a CDC6 homolog from *Pyrobaculum aerophilum* indicated that it is featured with a novel loop and α-helix, which is not found in other AAA+ family proteins (21). Another crystal structure of ORC from *A. pernix* suggested that ATP binding and hydrolysis may substantially influence its structure (22). To date, the biochemical properties of archaeal ORC/CDC6 are not yet clear. It has been shown that CDC6 protein undergoes autophosphorylation on Ser residue (23). It is interesting to note that eukaryotic CDC6 can also be autophosphorylated. However, the function of this activity *in vivo* is unclear. It has also been reported that CDC6 can stimulate MCM helicase activity in *Thermoplasma acidophilum*, which suggests that it possesses a regulatory function during DNA replication (24).

Although the mechanism of helicase assembly at the origin is not yet clear, several possible mechanisms have been proposed. Studies showed that the CDC6 protein can dissociate the MCM complex, suggesting a potential role in the origin assembly process (25). Another study also indicated that CDC6 may be involved in the recognition of DNA replication origin in *Pyrococcus furiosus* (26) .

Attempt to knockout the CDC6 gene in *T. kodakarensis* has failed, suggesting that CDC6 is essential and probably plays an important role in *T. kodakarensis* DNA replication (Santangelo, TS *et al* personal communication). Similar to other *Thermococcus* and *Pyrococcus*, *T. kodakarensis* contains only one CDC6 homolog. The gene encoding for CDC6 is adjacent to the putative DNA replication origin (Fig. 1-3), consistent with many other species. The function of CDC6 is not clear yet. However, it has been shown that CDC6 can form a complex with PCNA and PolD *in vivo* (see below), which indicates its function in DNA replication (27). Future work needs to elucidate the role of CDC6 in *T. kodakarensis*.

MCM helicase

After the initiator protein melts the origin, DNA helicase is recruited onto the origin. During chromosomal DNA replication, the replicative helicase unwinds the duplex DNA to provide the ssDNA substrate for DNA primase and polymerase. In bacteria, the helicase is named DnaB and it is comprised of six identical polypeptides to form homohexamer rings. In eukarya and archaea, the replicative helicase is the minichromosome maintenance (MCM) complex. In

eukaryotes the MCM complex is a family of six essential polypeptides (Mcm2–7), which is a stable hexamer in solution. Archaea usually contain one MCM homologue in their genomes, although some archaea contain up to 8 MCM homologues (Table1-2). MCM expression level in archaeal cells has been reported to be 300-1500 molecules of MCM per cell (17,28). Recently, several archaeal species with multiple MCM homologues have been identified that are thought to have resulted from gene duplication or lateral gene transfer from viruses or other archaeal species.

The archaeal MCM usually contains 650-680 amino acids and can be divided into three major regions (Fig. 1-4): the N-terminal containing around 250 residues which is responsible for DNA binding and oligomerization; a catalytic region of about 300 residues consisting of several conserved motifs for nucleotide binding and hydrolysis; the C-terminal region containing a helix-turn-helix domain, which can only be identified in archaeal MCM, though the function is not yet clear. In solution, archaeal MCM has been reported to adopt different oligomeric states, although a hexamer is suggested to be the active form of MCM (29). Salt and protein concentration greatly influence MCM structure. To further investigate its catalytic mechanism, much effort has been focused on defining the structure of archaeal MCM by X-ray and electron microscopy. Unfortunately, there is still no high resolution of the full length sequence of this family of enzymes, so the information provided is limited.

In addition to structural features, the archaeal and eukaryotic MCM proteins share similar biochemical properties, some of which are different from

Table 1-2. Archaeal species with multiple MCMs

| Species Name | Numbers of MCM homologs |
|---|---|
| *Sulfolobus neozealandicus* | 2 |
| *Natronomonas pharaonis* DSM 2160 | 2 |
| *Haloarcula marismortui* ATCC 43049 | 3 |
| *Halorubrum lacusprofundi* ATCC 49239 | 2 |
| *Methanosarcina acetivorans* C2A | 2 |
| *Thermococcus kodakarensis* KOD1 | 3 |
| *Methanopyrus kandleri* AV19 | 2 |
| *Methanocaldococcus jannaschii* DSM 2661 | 4 |
| *Methanococcus aeolicus* Nankai-3 | 3 |
| *Methanococcus vannielii* SB | 3 |
| *Methanococcus maripaludis* C5 | 4 |
| *Methanococcus maripaludis* C6 | 8 |
| *Methanococcus maripaludis* C7 | 4 |
| *Methanococcus maripaludis* S2 | 4 |
| *Methanococcus vannielii* A3 | 2 |

the two other extensively studied replicative helicases, *Escherichia coli* DnaB and phage T7. Both the archaeal and eukaryotic MCM translocate on DNA in the 3'-5' direction (DnaB has a 5'-3' directionality), both can assemble around closed circular DNA (while DnaB cannot) and both have similar processivity *in vitro*. In addition to structure features, the enzymes can bind and translocate along dsDNA, and unwind DNA-RNA hybrids when moving along the DNA strand. The archaeal MCM protein was also shown to displace proteins, such as histones, from DNA during DNA unwinding and it was suggested that the eukaryotic

enzyme may be able to displace proteins from DNA as well.

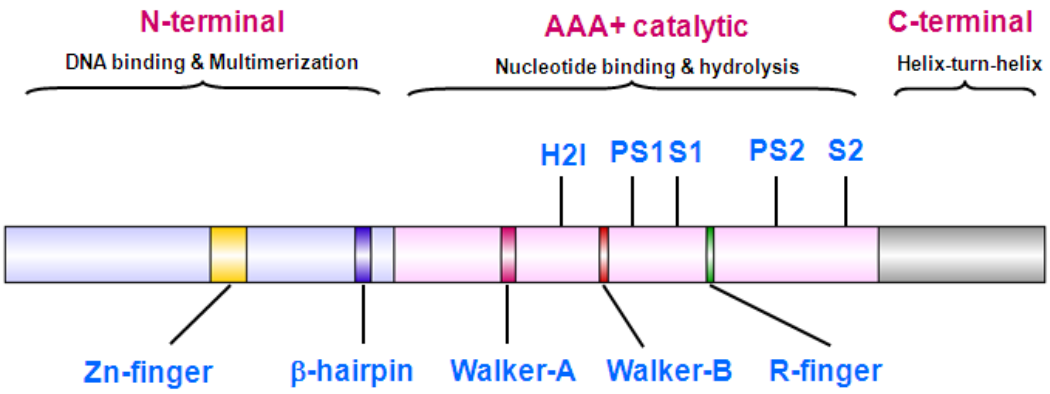



Figure 1-4. The schematic structure of archaeal MCM. N-terminal (in blue) is for DNA binding and multimerization; AAA+ catalytic domain has conserved motifs for ATP binding and hydrolysis; C-terminal has a predicted helix-turn-helix domain. Several motifs required for MCM helicase activity is labeled in blue. Zn-finger, Zinc finger; β-hairpin; Walker-A motif; Walker-B motif; R-finger, Arginine finger; H2I, Helix 2 insertion; PS1, Presensor 1; PS2, Presensor 2; S1, Sensor 1 and S2, Sensor 2.

Unlike many other *Thermococcus*, *Pyrococcus* and most other archaeal species, the *T. kodakarensis* genome encodes three MCM homologs (Table1-2 and also see Chapter 2). Bioinformatics analysis suggested that all three MCMs contain the essential motifs for helicase activity and these three MCMs may originate from the same origin. Our biochemical study has shown that although all three MCMs exhibited helicase activity, only MCM3 is essential for *T. kodakarensis* growth, while the other two (MCM1 and 2) are deletable (30). The essential MCM gene is located in the same operon with GINS23, which is also

found in other archaeal species. Another interesting finding is that the two non-essential MCM genes are located in different viral integration regions, suggesting that the two genes are acquired via lateral gene transfer. When the cell grows in the mid-log phase, MCM3 expression level is highest at 2000 molecules per cell, while the other two are expressed at levels much lower than MCM3. Therefore, a tempting hypothesis is that *T. kodakarensis* controls the different expression levels of the three MCMs for its replication. Although both MCM2 and 3 exhibit robust helicase activity *in vitro*, only MCM3 is indispensable for *T. kodakarensis* due to its high concentration *in vivo*. Future experiment will be performed to replace the genomic MCM3 gene with MCM2 to establish that the cell remains viable.

### 1.4.3 Elongation

Primase

The ssDNA unwound by DNA helicase is covered by the ssDNA binding protein RPA from nuclease degradation, and becomes the template for DNA polymerase to elongate. However, DNA polymerases require a 3'-OH primed template in order to elongate DNA chains and these primers are synthesized by DNA primases. In bacteria, DNA primase consists of a single subunit DnaG to synthesize short RNA primers. In eukaryotes, DNA primase is a heterodimer containing a catalytic subunit (Pri-S) that associates with a regulatory subunit Pri-L. This heterodimer is found to synthesize short oligoribonucleotides that are elongated by the Polα complex to generate RNA-DNA oligonucleotides (31). These short chains are recognized by the clamp loader RFC which loads PCNA

onto a primer-template junction that tethers the DNA polymerase which catalyzes the synthesis of lagging and leading strands (32).

Although archaea have been shown to contain both the DnaG and eukaryotic-type DNA primases, the DnaG-like primase was reported to play a role in RNA degradation rather than DNA replication (33). Genetic studies suggest that the genes encoding for Pri-L and Pri-S are essential whereas DnaG is dispensable for cell growth (34). The archaeal primase contains a two-subunit structure with significant homology to the eukaryotic Pri-L and Pri-S subunits. However, the counterpart to the eukaryotic Polα subunit has not been detected in archaeal genomes (35). Similar to the eukaryotic primase heterodimer, the archaeal Pri-S contains DNA primase activity and the Pri-L subunit may regulate its primase activity. It has been shown that both archaeal and eukaryotic Pri-L contain a [4Fe4S] cluster which is essential for primer synthesis, although the mechanism of the metal cluster in this process is not yet clear (36,37). In some archaeal species, RNA primed intermediates have been found, similar to those observed in eukaryotic Okazaki fragments (19,38). However, studies with the heterodimeric DNA primase complex isolated from *P. furiosus* suggested that the Pri-S alone preferentially utilized dNTPs and formed DNA chains *de novo* that were extended up to several kilobases (39,40). Furthermore, *in vitro* replication by the *P. furiosus* primase complex was stimulated by ATP and ATP was reported to start DNA chains. Similar findings have been reported for the primase complex isolated from *P. abyssi* (34).

The *T. kodakarensis* primase complex can synthesize both RNA and DNA,

but preferentially utilizes dNTP rather than rNTP (41). This finding suggests that archaeal DNA replication may be initiated by a DNA primer. In addition, the catalytic subunit alone is found to form the dNMP-Tris or dNMP-glycerol through a 5' phosphorus linkage of dAMP (41,42). The significance of the dNMP adduct formation is being explored.

GINS complex

The GINS complex has been identified as a component required for the establishment and maintenance of replication forks in eukarya (43). Archaeal GINS homolog was originally identified by bioinformatics study and all sequenced archaeal species contain GINS homolog (44). Most archaea contain two GINS homologs, GINS15 and GINS23 and there are some species that lack GINS23 (45). It is possible that this protein is present in archaeal cells but that it has diverged in sequence to the point where it cannot be detected. It has also been proposed that the GINS23 protein was present in the last common ancestor but then lost during the evolutionary process.

Archaeal GINS structure has been elucidated. It has similar overall structure to human GINS structure. The major difference between these is the location of the C-terminal small domain of GINS15 (46). In addition, archaeal GINS structure also showed different subunit contacts from those in human GINS. The overall similarity in the architectures between the archaeal and eukaryotic GINS complexes suggests that the GINS function is conserved (47).

To date, four archaeal GINS complexes have been characterized. In

*Sulfolobus solfataricus*, GINS was found to interact with MCM and both subunits of primase (48). It was also reported that RecJ homolog RecJdbh protein can be co-purified with the GINS complex. However, no obvious functional effect was observed from those interactions (48). In *P. furiosus*, GINS23 and GINS15 form a tetramer in a 2:2 ratio. The GINS23 protein can interact with MCM helicase and GINS15 can interact with CDC6 homolog (49). Surprisingly, it is claimed that GINS complex can stimulate MCM helicase activity through an increase in ATP hydrolysis, which is not seen in *S. solfataricus* or *T. kodakarensis* (see Chapter 2). In *T. acidophilum* and other euryarchaea, only GINS15 is found. It is shown that GINS15 can form a homotetramer and physically interact with MCM. It has also been claimed that GINS stimulates MCM helicase activity under high concentration of GINS (50). In *T. kodakarensis*, biochemical studies indicate that GINS can form a stable complex with a RecJ homolog nuclease, and stimulate its exo-nuclease activity (51). However, no interaction between GINS and MCM has been identified and there is no obvious influence on MCM helicase activity by the GINS complex (Chapter 2). It has also been found that GINS can interact with Polymerase D (27), though the significance of this interaction has not been elucidated. The fact that GINS is reported to interact with various proteins, indicates that it has multiple functions during DNA replication.

DNA polymerase

DNA polymerase requires ssDNA as a template and an RNA primer generated by primase to synthesize the complementary DNA. Similar to bacteria

and eukaryote, archaea possess multiple polymerases. Studies have suggested that both DNA polymerase B (PolB) and PolD are involved in archaeal chromosomal DNA replication (52).

Polymerase B (PolB) is a ubiquitous family that exists in all archaeal organisms. All archaeal species contain at least one PolB homolog. Because all three replicative polymerases in eukarya (Polα, δ, and ε) belong to PolB family, it was proposed that the members of this family also participate in chromosomal replication in archaea. PolB possess a 3'-5' exonuclease proofreading activity. The *T. kodakarensis* PolB crystal structure has been elucidated (53), and this has provided useful information to clarify the mechanisms for the rapid and accurate elongation reaction. Interestingly, many members of PolB contain inteins. The intein location is conserved among different species. It is also found that the catalytic subunit of PolB from *M. thermautotrophicus* is split into two parts and the association of both polypeptides is required to support the activity (54).

PolD can only be found in euryarchaea (55-57). It is a heterodimer of PolD-L and PolD-S. In *Pyrococcus* and *Thermococcus*, the genes encoding for PolD-L and PolD-S are in an operon and located next to the putative DNA replication origin. PolD has also been proposed to function at the replication fork as its small subunit shares amino acid sequence similarity with several of the small non-catalytic subunits of the eukaryotic Polε and Polδ . In addition, it has been shown that the large subunit of PolD shares sequence similarity with the catalytic subunit of the eukaryotic Polε (58) and may be the functional homologue of Polε in archaea. However, the biochemical properties of PolD and

its function in DNA replication are still poorly understood. To date, few biochemical studies on PolD have been reported. PolD has low processivity, 3'-5' exonuclease activity and strand displacement activity (56,57). Furthermore, it has been reported that PolD can be stimulated by RFC and PCNA (52).

Genome sequence analysis suggested that both PolB and PolD can be found in *T. kodakarensis*. It was hypothezed that both PolB and PolD function in the replication fork and work independently. In eukaryotes, Polδ is mainly responsible for copying the lagging strand template; while Polε primarily participates in the leading strand (59). Like the eukaryotes, PolB and PolD may participate in different strand synthesis. To support this idea, a bioinformatics study indicated that PolD evolved together with Polδ; while PolB is more similar to Polε (58). However, a genetic study by Santangelo group indicated that PolB is dispensable for *T. kodakarensis* cell growth (Santangelo, TS *et al* unpublished result). It implies that *T. kodakarensis* genome replication can be performed exclusively by only one DNA polymerase. Nevertheless, studies are still needed to explain this observation.

PCNA and RFC

The processivity of a polymerase can be defined by the probability that a polymerase will incorporate another nucleotide or will dissociate from the template. The replicative polymerase on its own has very low processivity. To synthesize over a million bases in the leading strand, the required processivity is conferred by the ring shaped sliding clamp, which can encircle the duplex DNA

and tether a polymerase to the DNA template. In bacteria, the sliding clamp is a homodimer, β dimer (60). In eukaryotes and archaea, the sliding clamp is a trimer, named as proliferating cell nuclear antigen (PCNA). Studies have shown that PCNA is an essential factor required for chromosomal DNA replication in archaea and eukarya. Crenarchaea usually encodes three distinct PCNA genes that form heterotrimers. In contrast, most members of euryarchaea and thaumarchaea encode a single PCNA gene which results in formation of a homotrimer ring as the active form. Several archaeal PCNA structures have been elucidated [(61), reference therein, also see Chapters 3 and 4]. In all organisms studied PCNA forms ring-shaped trimers.

In addition to their function as the processivity factor to the replicative polymerases, PCNA proteins from archaea and eukarya have been shown to associate with a large number of other cellular proteins required for nucleic acid metabolic processes (61). The interaction between PCNA and its client protein is through the PCNA interacting protein (PIP) motif. The PIP motif is usually located in the C-/N-terminal of the client proteins. Structural study indicated that one PIP peptide can bind each PCNA subunits. Since the PCNA ring contains three subunits, it is possible that PCNA can interact with multiple partner proteins simultaneously as a platform. However, more biochemical and structural studies are needed to support this idea. In addition to the PIP motif, there are also other interacting sites in the client proteins, which are required for their activity.

The PCNA trimer is a circular molecule that encircles the DNA. However, PCNA itself cannot spontaneously assemble onto DNA and requires a specific

clamp loader, replication factor C (RFC) to facilitate clamp loading onto the primed DNA template. This process is usually ATP- and DNA-dependent (62,63). Biochemical and structure studies have provide much of the information on the mechanism of this process. Like MCM and CDC6, RFC subunits are members of the AAA+ protein family. In most archaea two homologues of RFC have been identified, RFC-S and RFC-L, and these form a pentameric complex containing 4 subunits of RFC-S associated with one subunit of RFC-L. Four ATP bind to RFC and then the intact RFC complex interacts with PCNA. This interaction can open the PCNA ring and facilitate its binding to DNA. This loading process depends on ATP binding to RFC. Although it is not yet fully confirmed, it is believed that ATP hydrolysis releases the clamp loader from the loaded PCNA.

To date, among all sequenced archaea, there are only two archaeal species that contain two PCNA homologs: *P. aerophiluma* and *T. kodakarensis.* In this thesis, the biochemical properties of the two PCNA proteins from *T. kodakarensis* were determined and compared: although both PCNA can form a homotrimer, they have different stabilities. Both PCNAs function with RFC and DNA polymerase. Genetic studies also showed that only one PCNA homolog is required for cell survival (Chapter 3).

### 1.4.4 Okazaki fragment maturation

Several additional proteins required for the elongation phase and the maturation of Okazaki fragments have been isolated and studied in several archaea. These include the topoisomerase, ligases, Fen-1 and RNase HII. Our

group identified a novel DNA nuclease which can be stimulated by GINS complex in *T. kodakarensis*, and named as GAN (GINS associated nuclease). Structural analysis indicated that GAN has a similar structure to eukaryotic CDC45, although the primary sequence has no obvious homology. It has been reported that CDC45, GINS and MCM form a functional complex in eukarya (43,64). However, the archaeal counterpart of CDC45 is absent and has not been discovered based on amino acid sequence alignment (65). It is possible that the archaeal CDC45 has evolved to some new function and its sequence is different from that of eukaryotes. Therefore, no obvious homolog can be readily identified by simple primary sequence. It may also apply to other DNA replication factors which appear in eukarya but are absent in archaea. In addition, it has also been reported that PCNA can interact with Fen1, ligase and RNase HII in archaea. In *T. kodakarensis*, PCNA can stimulate Fen1 endonuclease activity but has no major effect on ligase activity. This mechanism needs to be elucidated in the future.

# Chapter 2 Biochemical characterization of multiple MCMs in *T. kodakarensis*

## 2.1 Abstract

The minichromosome maintenance (MCM) complex is thought to function as the replicative helicase in Archaea and Eukaryotes. In Eukaryotes, this complex is an assembly of six different but related polypeptides (MCM2-7) but, in most Archaea, one MCM protein assembles to form a homohexameric complex. Atypically, the *T. kodakarensis* genome encodes three archaeal MCM homologues, here designated MCM1-3, although MCM1 and MCM2 are unusual in having long and unique N-terminal extensions. The results reported establish that MCM2 and MCM3 assemble into homohexamers and exhibit DNA binding, helicase and ATPase activities *in vitro* typical of archaeal MCMs. In contrast, MCM1 does not form homohexamers and although MCM1 binds DNA and has ATPase activity, it has only minimal helicase activity *in vitro*. Removal of the N-terminal extension had no detectable effects on MCM1 but increased the helicase activity of MCM2. A *T. kodakarensis* strain with the genes TK0096 (MCM1) and TK1361 (MCM2) deleted has been constructed that exhibits no detectable defects in growth or viability, but all attempts to delete TK1620 that encodes MCM3 have been unsuccessful. This suggests that MCM3 is the replicative helicase in *T. kodakarensis*. The likely origins and possible function(s) of the three MCM proteins are discussed.

## 2.2 Introduction

The minichromosome maintenance (MCM) complex is thought to function as the replicative helicase in Archaea and Eukaryotes, the role played by DnaB in bacteria. In all eukaryotes, the replicative MCM complex is a heterohexamer formed by the assembly of six different homologues (MCM2 through MCM7) all of which are essential for viability. The MCM complex participates in both the initiation and elongation phases of DNA replication [reviewed in: (66-69). The heterohexameric complex, and also trimeric complexes formed by MCM4, MCM6 and MCM7 have 3'→5' helicase activity *in vitro* although the replicative helicase *in vivo* seems to be a larger assembly of all six MCM homologues, Cdc45 and the GINS complex, together designated the CMG complex (Cdc45, MCM and GINS) (43,64,70). In archaea, the replicative helicase is also thought to be the MCM [reviewed in: (71-75)] but most Archaea have only one MCM homologue. A central domain that embodies the AAA+ catalytic ATPase is conserved in both the archaeal and eukaryotic MCMs but the eukaryotic homologues have longer sequence, with N- and C-extensions that are not present in the archaeal MCM proteins. In contrast, archaeal MCMs have a C-terminal domain containing a helix-turn-helix motif that is not conserved in eukaryotic MCMs. As predicted, archaeal MCM complexes also have ATP-dependent 3′→5′ helicase activity and can bind and translocate along ss and ds DNA, displace proteins bound to DNA, and unwind DNA–RNA hybrids [reviewed in: (71,73,74)].

With many archaeal genome sequences now available, a few species with

more than one MCM homologue have been identified with gene duplication and lateral gene transfer as explanations (12,76-79). The genome of *T. kodakarensis* encodes three MCM homologues (here designated MCM1, MCM2 and MCM3) although MCM1 and MCM2 are unusual in having long and unique N-terminal extensions. With facile genetic technologies now established for *T. kodakarensis*, this species has become a model system for archaeal molecular biology research. Here we report the results of a combination of biochemical and genetic approaches that establish that MCM2 and MCM3 have the activities expected for an archaeal MCM helicase but that only MCM3 appears essential for replication and viability.

## 2.3 Materials and methods

### 2.3.1 Multiple amino acid sequence alignment of MCM proteins

The amino acid sequences of the three *T. kodakarensis* MCM proteins were compared to those of *S. solfataricus* and *M. thermautotrophicus*. The alignment was generated using ClustalW. Amino acids with 100% identity in all species are highlighted in red, 80% identity in green and 60% identity in yellow. For clarity, the two inteins found in MCM3 are not included although their location is marked by arrow heads. The location of the truncation of the N-terminal regions of MCM1 and 2 is marked by an arrow. The location of the Walker-A sites on MCM1-3 (Lys571, Lys473, and Lys335, respectively) is marked by an asterisk. The location of the putative Walker-A box motif lysine of MCM3 that was replaced by a glutamate to generate MCM3 (K335E/K355E) is

marked by a triangle. The accession numbers used are: MCM1, YP_182509.1; MCM2, YP_183774.1; MCM3. YP_184033.1; *M. thermautotrophicus*, NP_276876; *S. solfataricus*, NP_342281.1.

### 2.3.2 Media and growth conditions

*T. kodakarensis* cultures were grown anaerobically at 85˚C in artificial sea water (ASW) containing trace minerals and vitamins supplemented with 5 g yeast extract and 5 g tryptone per liter (ASW-YT medium) or with a mixture of 20 amino acids (ASW-AA) (80). Sulfur (2 g/l) and/or sodium pyruvate (5 g/l) were also added to ASW-YT or ASW-AA where indicated. Gelrite (1% w/v) was added to solidify these media for plating. Cells competent for DNA uptake were prepared as described (81).

### 2.3.3 Construction of *T. kodakarensis* deletion strains

Sequences that flank TK0096, TK1361 and TK1620 were PCR amplified from *T. kodakarensis* KW128 genomic DNA and were cloned into plasmid pTS535, adjacent to the [TK0254 (*trpE*) +TK0664] expression cassette, essentially as previously described (16). The sequences of all PCR primers used in this study are list in appendix. The plasmids generated (Table 2-1) were used to transform *T. kodakarensis* TS517 (Δ*pyrF*; Δ*trpE*::*pyrF*; ΔTK0664) with transformants selected by growth in the absence of tryptophan. Diagnostic PCR (primers used are list in Table 2-2) confirmed that the (TK0254+TK0664) cassette was integrated into the *T. kodakarensis* genome, adjacent to the target gene and

flanked a direct duplication of genomic DNA. Expression of TK0664 resulted in these transformants being sensitive to 6-methyl purine (6MP). Mutants, spontaneously resistant to 6MP were selected as clones that grew on ASW-YT plates containing 100 μM 6MP. PCR and sequencing of genomic DNA isolated from representative clones, designated *T. kodakarensis* TS601 and TS602 confirmed that recombination between the duplicated genomic regions had precisely deleted the (TK0254+TK0664) cassette and TK0096 or TK1361 respectively (Table 2-3). In contrast, although the $6MP^R$ clones isolated following transformation with plasmid DNA containing TK1620 had lost the (TK0254+TK0664) cassette they all retained TK1620. Repetition of the transformation, selection and counter-selection steps, starting with the *T. kodakarensis* TS601 (Δ*pyrF*; Δ*trpE*::*pyrF*; ΔTK0664; ΔTK0096) as recipient strain, generated *T. kodakarensis* TS604 (Δ*pyrF*; Δ*trpE*::*pyrF*; ΔTK0664; ΔTK0096; ΔTK1361). The *T. kodakarensis* strains generated and used in this study are summarized in Table 2-3.

### 2.3.4 Construction of MCM expression plasmids

Standard molecular biology procedures were used to construct plasmids, transform and select *Escherichia coli* DH5α transformants and isolate plasmid DNA from *E. coli* and *T. kodakarensis*. The genomic copy of TK1361 (MCM2) includes sequences that encode two inteins that were removed by using PCR as described previously (82). The resulting ORF (here designated TK1361) and TK1620 (MCM3) were cloned into pET-21a (Novagen), with six histidine codons

added in-frame to their 3'-termini, resulting in plasmids designated pET-TK1361 and pET-TK1620 (Table 2-1). The gene (TK0096) encoding MCM1 was synthesized with six histidine codons added at the 3'-terminus and cloned into pET-21a by GeneArt, resulting in plasmid pET-TK0096. Derivatives of these plasmids were generated by site-directed mutagenesis in which the MCM-encoding sequence was changed to generated plasmids that encode Walker-A-box variants of MCM1 (K571E), MCM2 (K473E) and MCM3 (K335E). Derivatives that encode variants of MCM1 and MCM2 that lack the N-terminal extension (designated MCM1-ΔN and MCM2-ΔN), were generated by PCR-based site-directed mutagenesis as previously described (82).

### 2.3.5 Expression and purification of recombinant $His_6$-tagged MCM1, MCM2 and MCM3

Plasmids pET-TK0096, pET-TK1361 and pET-TK1620 were transformed into *E. coli* BL21 DE3 Rosetta (Invitrogen) and expression of the MCM encoding gene was induced by addition of 0.5 mM IPTG to cultures growing in LB containing 100 μg/ml ampicillin and 25 μg/ml chloramphenicol at an $OD_{600}$ of ~0.6. Incubation was continued for 16 hr at 16°C. Cells were collected by centrifugation, resuspended and incubated in 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 mM imidazole and 10% glycerol (lysis buffer) at 55°C for 30 min, and then lysed by sonication. The lysate was clarified by centrifugation and loaded onto a $Ni^{2+}$-charged column (Chelating Sepharose Fast Flow, GE Healthcare) pre-equilibrated with lysis buffer. The column was washed with 50 mM Tris-HCl (pH

8.0), 1 M NaCl, 10% glycerol, 50 mM imidazole and then with 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 50 mM imidazole and 10% glycerol. The $His_6$-tagged MCM protein was eluted from the column by washing with 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 250 mM imidazole and 10% glycerol, dialyzed and stored in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 0.5 mM EDTA, 2 mM DTT and 10% glycerol.

### 2.3.6 Size exclusion chromatography

An aliquot (200 μg) of each MCM protein, dissolved in 200 μl 25 mM Tris–HCl (pH 7.5), 500 mM NaCl and 10% (v/v) glycerol, was incubated for 1h at 22°C and then subjected to chromatography, at 22°C, by passage through a Superdex-200 gel-filtration column (HR10/30; GE Healthcare) pre-equilibrated with 25 mM Tris–HCl (pH 7.5), 500 mM NaCl and 10% glycerol (v/v). Fractions (250 μl) were collected and the proteins present in a sample of each fraction were separated by electrophoresis through a 10% (w/v) polyacrylamide gel containing 0.1% SDS and visualized by staining with Coomassie brilliant blue (R250).

### 2.3.7 Light scattering of MCMs

The molecular mass of the MCM complexes were determined using 100 μg of proteins in 20 μl buffer with 25 mM Tris-HCl pH 7.5, 50 mM NaCl and 10% glycerol. A 1200 series HPLC system (Agilent Technologies) with a Shodex KW-802.5 or a Shodex KW-804 column (Showa Denko K.K.) was used. The flow rate was 0.5 ml/min of a solution containing 25 mM Tris–HCl (pH 7.5), 500

mM NaCl and 10% glycerol (v/v). Light scattering was measured on a miniDawn Treos (Wyatt Technology) and the protein concentration was measured with an OptilabrEX differential refractometer (Wyatt Technology) (83).

### 2.3.8 Helicase assays

Oligonucleotides MD007 and MD015 (Table 2-4) were $^{32}$P-end-labeled by incubation with [$\gamma$-$^{32}$P]ATP (3000 Ci/mmol; Perkin Elmer) and T4 polynucleotide kinase (Fermentas). The 25 bp and 96 bp double stranded $^{32}$P-labeled DNA substrates used for helicase assays were generated by hybridization of the $^{32}$P-MD007 with MD008, and $^{32}$P-MD014 with MD015, respectively (Table 2-4), and purified as previously described (84).

DNA helicase activity was assayed in reaction mixtures (15 µl) that contained 20 mM Tris-HCl (pH 8.0), 2 mM DTT, 10 mM $MgCl_2$, 1.5 µg BSA, 2 mM ATP, 10 fmol of $^{32}$P-labeled substrate and the MCM protein, as noted in the legends to Figures 2-2, 2-3, 2-4, 2-5 and 2-6. The reaction mixtures were incubated at 70°C for 1 h and the reaction stopped by addition of 5 µl loading buffer (0.1% xylene cyanol, 0.1% bromophenol blue, 1% SDS, 50% glycerol and 100 mM EDTA), after which the sample was placed on ice. Aliquots (10 µl) were loaded onto an 8% (w/v) polyacrylamide gel and the $^{32}$P-labeled nucleic acids present were visualized and quantified by phosphor-imaging after separation by electrophoresis in 0.5X TBE for 40 min at 180 V. The helicase assays were repeated at least three times, and the average of the results obtained, with standard deviations, are reported.

To determine the nucleotide requirements for helicase activity, reaction mixtures (15 μl) that contained the 25 bp DNA substrate, 1 pmol of the MCM protein and 2 mM ATP, dATP, ADP, or [γ-S]ATP, or 1 mM CTP, dCTP, GTP, dGTP, UTP or dTTP were incubated at 70°C for 1 h. The reactions were stopped and the $^{32}$P-labeled products were separated, visualized and quantified as described above.

### 2.3.9 ATPase assays

ATPase activity was assayed in reaction mixtures (15 μl) that contained 25 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 1 mM DTT, 100 μg/ml BSA, 1.5 nmol of [γ-$^{32}$P]ATP (3000 Ci/mmol; Perkin Elmer), plus or minus 10 pmol of the 49-mer oligonucleotide MD008 (Table 2-4), plus the MCM protein, as noted in the figure legends. After incubation at 75°C for 1 h, an aliquot (1 μl) of the reaction mixture was spotted on a polyethyleneimine cellulose thin layer plate. ATP and $P_i$ were separated by chromatography in 1 M formic acid containing 0.5 M LiCl, and the extent of ATP hydrolysis was calculated based on phosphorimage quantification. The ATPase assays were repeated at least three times, and the averages of the results obtained with standard deviations, are reported.

To establish rates of ATP hydrolysis, reaction mixtures (45 μl) that contained 25 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 1 mM DTT, 100 μg/ml BSA, 4.5 nmol [γ-$^{32}$P]ATP (3000 Ci/mmol; Perkin Elmer) and 4 pmol MCM1 or MCM3 monomer, or 0.2 pmol of MCM2 monomer were incubated at 75ºC, with or without 10 pmol of the 49-mer oligonucleotide MD008. Aliquots (3 μl) of the

reaction mixture were removed after 0, 30, 45, 60, 90, 120, 150 and 175 min (MCM1 and 3) or 0, 2, 5, 10, 15, 30 and 45 min (MCM2) incubation, mixed with 1 μl 0.5 M EDTA, and the extent of ATP hydrolysis determined as described above.

### 2.3.10 Measurement of DNA binding by fluorescence polarization anisotropy (FPA)

A 30-mer oligonucleotide (A1, Table 2-4) was synthesized with a Cy5 at the 5'-terminus and the dye-labeled product was purified by chromatography through a 15 % acrylamide gel. The concentration of the DNA solution was determined by measurements of $A_{260}$ (extinction coefficient of 287900 $M^{-1}$ $cm^{-1}$) for DNA and $A_{646}$ (extinction coefficient 250,000 $M^{-1}$ $cm^{-1}$) for Cy5. The MCM protein was added to reaction mixtures that contained 25 mM HEPES-NaOH (pH 7.5), 2 mM DTT, 5 mM $MgCl_2$, 10 nM DNA, plus or minus 1 mM ATP. After 5 min incubation at 25ºC, FPA measurements were taken at 25°C using a Fluoromax-3 spectrofluorimeter equipped with an autopolarizer. The cuvette (3mm path length) contained a starting volume of 150 μl, the reaction mixtures were excited at 645 nm and emission measured at 670 nm. Three measurements were taken, averaged over 5 sec integration periods. The anisotropy values were directly tabulated with measured G factor and dark corrections acquired for each blank for each experiment. Binding constants ($K_d$) were calculated by using Grafit version 5.0.1, based on the following equation for fluorescent polarization anisotropy measurements: $\Delta A = \frac{\Delta A_T}{2D_T}\left\{(E_T + D_T + K_d) - \sqrt{(E_T + D_T + K_d)^2 - 4E_T D_T}\right\}$,

where ΔA is the change in anisotropy, $\Delta A_T$ is the total anisotropy change, $E_T$ is the enzyme concentration at each titration point, $D_T$ is the total concentration of DNA (assuming it is constant at 10 nM) and $K_d$ is the dissociation constant for the binding isotherm. All experiments were repeated and the average values obtained, with standard deviations, are reported.

Table 2-1. Plasmids used to generate the knockout strains and for protein expression

| Plasmid name | Used for |
|---|---|
| pZLE031 | Delete TK0096 *in vivo* |
| pZLE29 | Delete TK1361 *in vivo* |
| pET-TK0096 | MCM1 expression |
| pET-TK1361 | MCM2 expression |
| pET-TK1620 | MCM3 expression |
| pET-TK0096WA | MCM1 K571E expression |
| pET-TK1361WA | MCM2 K473E expression |
| pET-TK1620WA | MCM3 K335E expression |
| pET-TK0096-ΔN | MCM1-ΔN expression |
| pET-TK1361-ΔN | MCM2-ΔN expression |

Table 2-2. Oligonucleotides used for diagnostic PCR of the deleted strains

| Primer set | Sequence | Chromosomal location |
|---|---|---|
| I | GGCCAACGCCACTCGACCCGGGACC | TK1194083-1194107 |
| | GAGGATTTGAGTGGTGGTGAGCCG | TK1194581-1194558 |
| II | CAACCCAATCTGCCGTAACG | TK1191965-1191984 |
| | TCTTCCTCTTCGGCTTCATG | TK1195741-1195722 |
| III | GAGGAGAAGCTGGTCAGGGCTTTC | TK81123-81146 |
| | CTCATACCCACACTTAGGACACAC | TK81572-81549 |
| IV | TATGTACCTGTTCTCTGCGC | TK84557-84538 |
| | GTGATCACTTCTTCACAATG | TK80127-80146 |

Table 2-3. *T. kodakarensis* strains used in this study

| Strain designation | Relevant Genotype | Origin |
|---|---|---|
| TS517 | ΔpyrF; ΔtrpE::pyrF; ΔTK0664 | (85) |
| TS601 | ΔpyrF; ΔtrpE::pyrF; ΔTK0664; ΔTK0096 | This study |
| TS602 | ΔpyrF; ΔtrpE::pyrF; ΔTK0664; ΔTK1361 | This study |
| TS604 | ΔpyrF; ΔtrpE::pyrF; ΔTK0664; ΔTK0096; ΔTK1361 | This study |

Table 2-4. Oligonucleotides used for DNA binding and to generate the helicase substrates

| Oligo names | Sequence | Formed substrate names |
|---|---|---|
| MD007 | 5'-TTTGTTTGTTTGTTTGTTTGTTTGCCGACGTGC CAGGCCGACGCGTCCC | 25F |
| MD008 | 5’-GGGACGCGTCGGCCTGGCACGTCGGTTTGTTT GTTTGTTTGTTTGTTTG | 25F |
| MD012 | 5'-TTTGTTTGTTTGTTTGTTTGTTTGCGACGGCCA GTGCCAAGCTTGCATGCCTGCAGGTCG | 36B |
| MD013 | 5'-CGACGGCCAGTGCCAAGCTTGCATGCCTGCAG GTCGTTTGTTTGTTTGTTTGTTTGTTTG | 36C |
| MD014 | 5’-TTTGTTTGTTTGTTTGTTTGTTTGCGACGGCCA GTGCCAAGCTTGCATGCCTGCAGGTCGCGACGGC CAGTGCCAAGCTTGCATGCCTGCAGGTCGCGACG GCCAGTGCCAAGCTTGCAT | 96F |
| MD015 | 5’-ATGCAAGCTTGGCACTGGCCGTCGCGACCTGC AGGCATGCAAGCTTGGCACTGGCCGTCGCGACCT GCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTG TTTGTTTGTTTGTTTGTTTG | 96F |
| MD016 | 5'-CGACGGCCAGTGCCAAGCTTGCATGCCTGCAG GTCG | 36A |
| A1 | 5’-(Cy5)-GGGGCGAGTCCAGGTCAGGACCTTGCG GGG | |

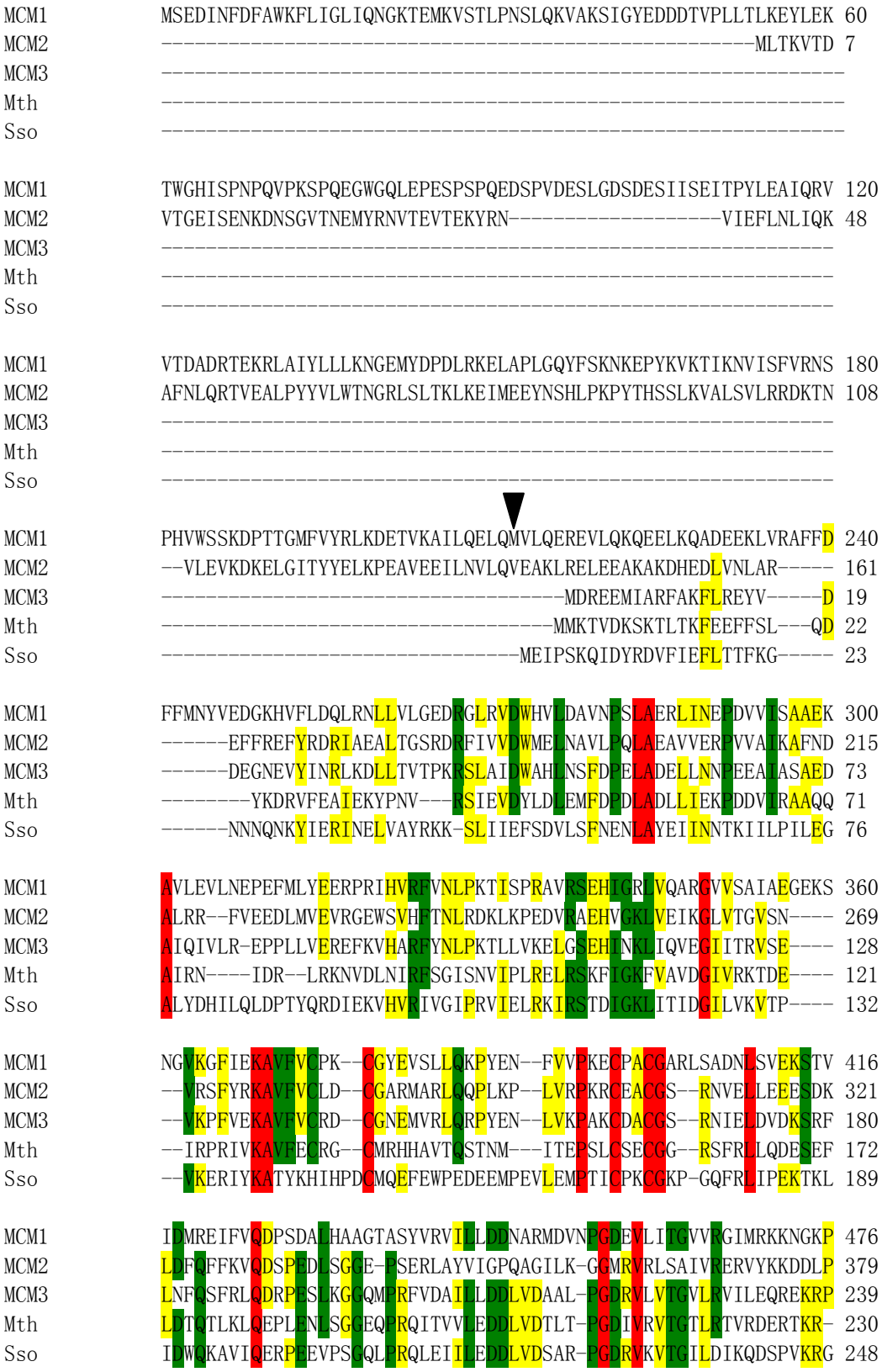
## 2.4 Results

### 2.4.1 The *T. kodakarensis* genome encodes three MCM homologues

The genome sequence of *T. kodakarensis* revealed the presence of three genes (TK0096, TK1361 and TK1620) that were predicted to encode MCM homologues (12), here designated MCM1, MCM2 and MCM3, respectively. An alignment of their amino acid sequence with those of other archaeal MCM proteins (Fig. 2-1) and eukaryotic MCM proteins confirms the presence of all known motifs needed for helicase activity. The analysis also suggested that MCM1 and MCM2 have unique 205 and 136 residue N-terminal extensions, respectively (Fig. 2-1). The *T. kodakarensis* genome, MCM3 protein contains two inteins, one between the Walker-A and Walker-B motifs and the second is downstream of the Walker-B motif. The first intein is present in the same location in the single MCM encoded in related *Pyrococcus* genomes and the second intein is present in a similar location in the MCMs of *Methanoculleus marisnigri* and *Staphylothermus marinus* (86).

### 2.4.2 MCM1-3 proteins posses helicase activity

The MCM proteins from archaea and eukarya are the presumptive replicative helicases. However, all other archaeal MCM studied are from organisms with single MCM homologue. Thus, the ability of MCM1-3 to unwind duplex DNA was evaluated. It has been reported that both archaeal and eukaryotic MCM can unwind primed M13 substrates, although they prefer fork-like substrates. First, three different M13 based helicase substrates were used for


MCM1 MSEDINFDFAWKFLIGLIQNGKTEMKVSTLPNSLQKVAKSIGYEDDDTVPLLTLKEYLEK 60
MCM2 ----------MLTKVTD 7
MCM3
Mth
Sso
MCM1 TWGHISPNPQVPKSPQEGWGQLEPESPSPQEDSPVDESLGDSDESIISEITPYLEAIQRV 120
MCM2 VTGEISENKDNSGVTNEMYRNVTEVTEKYRN----------VIEFLNLIQK 48
MCM3
Mth
Sso
MCM1 VTDADRTEKRLAIYLLLKNGEMYDPDLRKELAPLGQYFSKNKEPYKVKTIKNVISFVRNS 180
MCM2 AFNLQRTVEALPYYVLWTNGRLSLTKLKEIMEEYNSHLPKPYTHSSLKVALSVLRRDKTN 108
MCM3
Mth
Sso
MCM1 PHVWSSKDPTTGMFVYRLKDETVKAILQELQMVLQEREVLQKQEELKQADEEKLVRAFFD 240
MCM2 --VLEVKDKELGITYYELKPEAVEEILNVLQVEAKLRELEEAKAKDHEDLVNLAR----- 161
MCM3 ----------MDREEMIARFAKFLREYV-----D 19
Mth ----------MMKTVDKSKTLTKFEEFFSL---QD 22
Sso ----------MEIPSKQIDYRDVFIEFLTTFKG----- 23
MCM1 FFMNYVEDGKHVFLDQLRNLLVLGEDRGLRVDWHVLDAVNPSLAERLINEPDVVTSAAEK 300
MCM2 ------EFFREFYRDRIAEALTGSRDRFIVVDWMELNAVLPQLAEAVVERPVVAIKAFND 215
MCM3 ------DEGNEVYINRLKDLLTVTPKRSLAIDWAHLNSFDPELADELLNNPEEAIASAED 73
Mth ---------YKDRVFEAIEKYPNV---RSIEVDYLDLEMFDPDLADLLIEKPDDVIRAAQQ 71
Sso ------NNNQNKYIERINELVAYRKK-SLIIEFSDVLSFNENLAYEIINNTKIILPILEG 76
MCM1 AVLEVLNEPEFMLYEERPRIHVRFVNLPKTISPRAVRSEHIGRLVQARGVVSAIAEGEKS 360
MCM2 ALRR--FVEEDLMVEVRGEWSVHFTNLRDKLKPEDVRAEHVGKLVEIKGLVTGVSN----- 269
MCM3 AIQIVLR-EPPLLVEREFKVHARFYNLPKTLLVKELGSEHINKLIQVEGIITRVSE----- 128
Mth AIRN----IDR--LRKNVDLNIRFSGISNVIPLRELRSKFIGKFVAVDGIVRKTDE----- 121
Sso ALYDHILQLDPTYQRDIEKVHVRIVGIPRVIELRKIRSTDIGKLITIDGILVKVTP----- 132
MCM1 NGVKGFIEKAVFVCPK--CGYEVSLLQKPYEN--FVVPKECPACGARLSADNLSVEKSTV 416
MCM2 --VRSFYRKAVFVCLD--CGARMARLQQPLKP--LVRPKRCEACGS--RNVELLEEESDK 321
MCM3 --VKPFVEKAVFVCRD--CGNEMVRLQRPYEN--LVKPAKCDACGS--RNIELDVDKSRF 180
Mth --IRPRIVKAVFECRG--CMRHHAVTQSTNM---ITEPSLCSECGG--RSFRLLQDESEF 172
Sso --VKERIYKATYKHIHPDCMQEFEWPEDEEMPEVLEMPTICPKCGKP-GQFRLIPEKTKL 189
MCM1 IDMREIFVQDPSDALHAAGTASYVRVILLDDNARMDVNPGDEVLITGVVRGIMRKKNGKP 476
MCM2 LDFQFFKVQDSPEDLSGGE-PSERLAYVIGPQAGILK-GGMRVRLSAIVRERVYKKDDLP 379
MCM3 LNFQSFRLQDRPESLKGGQMPRFVDAILLDDLVDAAL-PGDRVLVTGVLRVILEQREKRP 239
Mth LDTQTLKLQEPLENLSGGEQPRQITVVLEDDLVDTLT-PGDIVRVTGTLRTVRDERTKR- 230
Sso IDWQKAVIQERPEEVPSGQLPRQLEIILEDDLVDSAR-PGDRVKVTGILDIKQDSPVKRG 248

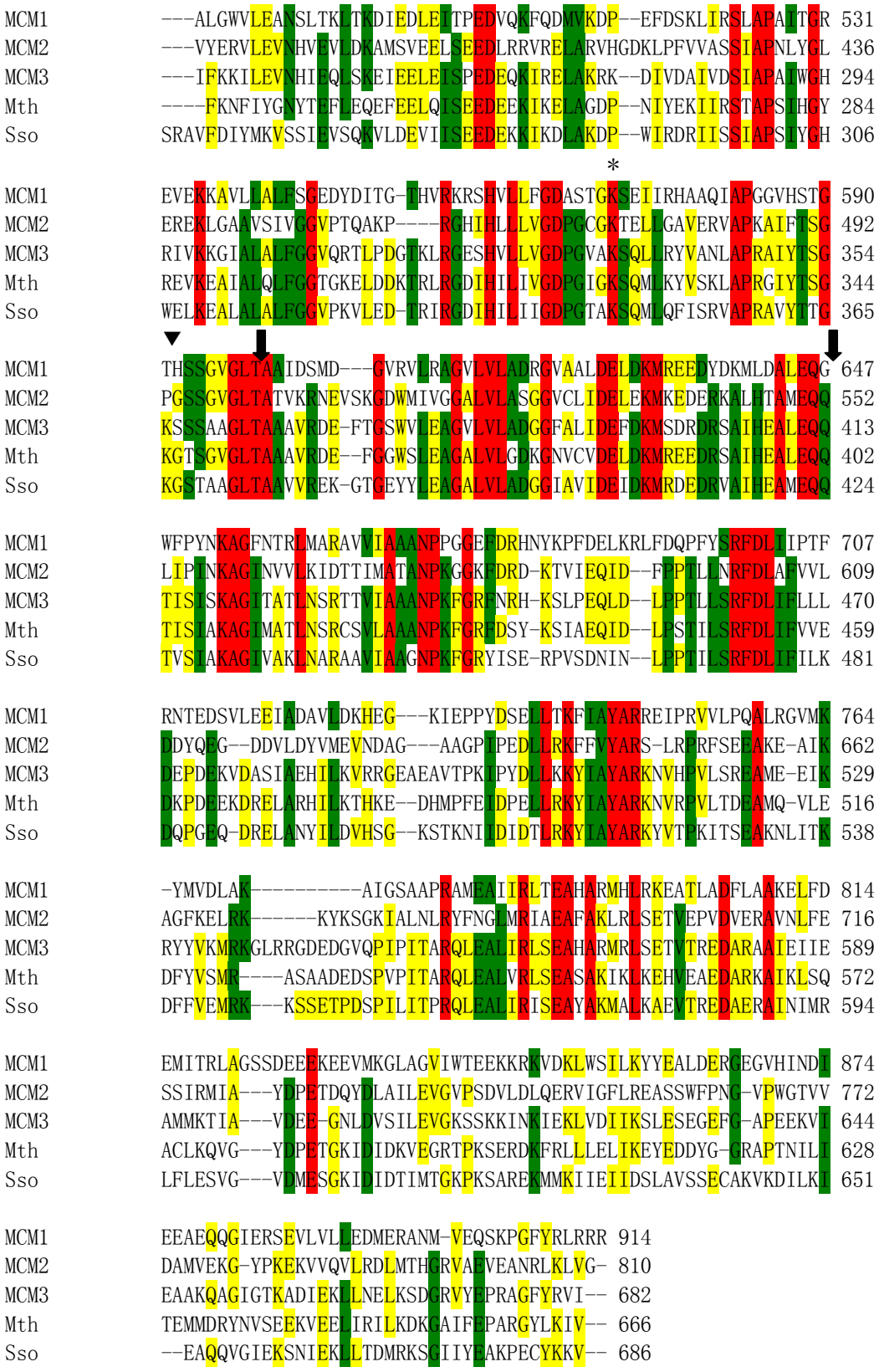

MCM1 ---ALGWVLEANSLTKLTKDIEDLEITPEDVQKFQDMVKDP--EFDSKLIRSLAPAITGR 531
MCM2 ---VYERVLEVNHVHVLDKAMSVEELSEEDLRRVRELARVHGDKLPFVVASSIAPNLYGL 436
MCM3 ---IFKKILEVNHIEQLSKEIEELEISPEDEQKIRELAKRK--DIVDAIVDSIAPAIWGH 294
Mth ----FKNFIYGNYTEFLEQEFEELQISEEDEEKIKELAGDP--NIYEKIIRSTAPSIHGY 284
Sso SRAVFDIYMKVSSIEVSQKVLDEVIISEEDEKKIKDLAKDP--WIRDRIISSIAPSIYGH 306
*
MCM1 EVEKKAVLLALFSGEDYDITG-THVRKRSHVLLFGDASTGKSEIIRHAAQIAPGGVHSTG 590
MCM2 EREKLGAAVSIVGGVPTQAKP----RGHIHLLLVGDPGCGKTELLGAVERVAPKAIFTSG 492
MCM3 RIVKKGIALALFGGVQRTLPDGTKLRGESHVLLVGDPGVAKSQLLRYVANLAPRAIYTSG 354
Mth REVKEAIALQLFGGTGKELDDKTRLRGDIHILIVGDPGIGKSQMLKYVSKLAPRGIYTSG 344
Sso WELKEALALALFGGVPKVLED-TRIRGDIHILIIGDPGTAKSQMLQFISRVAPRAVYTTG 365
MCM1 THSSGVGLTAAIDSMD---GVRVLRAGVLVLADRGVAALDELDKMREEDYDKMLDALEQG 647
MCM2 PGSSGVGLTATVKRNEVSKGDWMIVGGALVLASGGVCLIDELEKMKEEDEKKALHTAMEQQ 552
MCM3 KSSSAAGLTAAAVRDE-FTGSWVLEAGVLVLADGGFALIDEFDKMSDRDRSAIHEALEQQ 413
Mth KGTSGVGLTAAAVRDE--FGGWSLEAGALVLGDKGNVCVDELDKMREEDRSAIHEALEQQ 402
Sso KGSTAAGLTAAVVREK-GTGEYYLEAGALVLADGGIAVIDEIDKMRDEDRVAIHEAMEQQ 424
MCM1 WFPYNKAGFNTRLMARAVVIAAANPPGGEFDRHNYKPFDELKRLFDQPFYSRFDLIIPTF 707
MCM2 LIPINKAGINVVLKIDTTIMATANPKGGKFDRD-KTVIEQID--FPPTLLNRFDLAFVVL 609
MCM3 TISISKAGITATLNSRTTVIAAANPKFGRFNRH-KSLPEQLD--LPPTLLSRFDLIFLLL 470
Mth TISIAKAGIMATLNSRCSVLAAANPKFGRFDSY-KSIAEQID--LPSTILSRFDLIFVVE 459
Sso TVSIAKAGIVAKLNARAAVIAAGNPKFGRYISE-RPVSDNIN--LPPTILSRFDLIFILK 481
MCM1 RNTEDSVLEEIADAVLDKHEG---KIEPPYDSELLTKFIAYARREIPRVVLPQALRGVMK 764
MCM2 DDYQEG--DDVLDYVMEVNDAG---AAGPIPEDLLRKFFVYARS-LRPRFSEEAKE-AIK 662
MCM3 DEPDEKVDASIAEHILKVRRGEAEAVTPKIPYDLLKKYIAYARKNVHPVLSREAME-EIK 529
Mth DKPDEEKDRELARHILKTHKE--DHMPFEIDPELLRKYIAYARKNVRPVLTDEAMQ-VLE 516
Sso DQPGEQ-DRELANYILDVHSG--KSTKNIIDIDTLRKYIAYARKYVTPKITSEAKNLITK 538
MCM1 -YMVDLAK----------AIGSAAPRAMEAIIRLTEAHARMHLRKEATLADFLAAKELFD 814
MCM2 AGFKELRK------KYKSGKIALNLRYFNGEMRIAEAFAKLRLSETVEPVDVERAVNLFE 716
MCM3 RYYVKMRKGLRRGDEDGVQPIPITARQLEALIRLSEAHARMRLSETVTREDARAAIEIIE 589
Mth DFYVSMR----ASAADEDSPVPITARQLEALVRLSEASAKIKLKEHVEAEDARKAIKLSQ 572
Sso DFFVEMRK---KSSETPDSPILITPRQLEALIRISEAYAKMALKAEVTREDAERAINIMR 594
MCM1 EMITRLAGSSDEEEKEEVMKGLAGVIWTEEKKRKVDKLWSILKYYEALDERGEGVHINDI 874
MCM2 SSIRMIA---YDPETDQYDLAILEVGVPSDVLDLQERVIGFLREASSWFPNG-VPWGTVV 772
MCM3 AMMKTIA---VDEE-GNLDVSILEVGKSSKKINKIEKLVDIIKSLESEGEFG-APEEKVI 644
Mth ACLKQVG---YDPETGKIDIDKVEGRTPKSERDKFRLLLELIKEYEDDYG-GRAPTNILI 628
Sso LFLESVG---VDMESGKIDIDTIMTGKPKSAREKMMKIIEIIDSLAVSSECAKVKDILKI 651
MCM1 EEAEQQGIERSEVLVLLEDMERANM-VEQSKPGFYRLRRR 914
MCM2 DAMVEKG-YPKEKVVQVLRDLMTHGRVAEVEANRLKLVG- 810
MCM3 EAAKQAGIGTKADIEKLLNELKSDGRVYEPRAGFYRVI-- 682
Mth TEMMDRYNVSEEKVEELIRILKDKGAIFEPARGYLKIV-- 666
Sso --EAQQVGIEKSNIEKLLTDMRKSGIIYEAKPECYKKV-- 686

Figure 2-1. Alignment of MCM amino acid sequences. The sequences of MCM1, MCM2 and MCM3 are shown aligned by ClustalW with the sequences of archaeal MCMs from *S. solfataricus* and *M. thermautotrophicus.* Residues conserved 100%, 80% and 60% are highlighted in red, green and yellow, respectively. The locations of the two inteins in MCM3 are indicated by arrows. MCM1 and MCM2 were truncated at the position marked by an arrowhead to generate MCM1-ΔN and MCM2-ΔN. The locations of the lysine in the Walker-A box motif that were replaced by a glutamate to generate the MCM1 (K571E), MCM2 (K473E) and MCM3 (K335E) variants are marked by an asterisk. The location of the putative Walker-A box motif lysine of MCM3 that was replaced by a glutamate to generate MCM3 (K335E/K355E) is marked by a triangle. The GenBank accession numbers of the sequences shown are YP_182509.1 (MCM1); YP_183774.1 (MCM2); YP_184033.1 (MCM3); NP_276876 (*M. thermautotrophicus*) and NP_342281.1 (*S. solfataricus*).

this test. As shown in Figure 2-2, both MCM2 and 3 exhibited good helicase activity on primed M13 substrates with 5'-overhang (Fig. 2-2 B, lane 8-10 and lane 13-15 and Fig. 2-2 F) and with 3'-overhang (Fig. 2-2 C, lane 8-10 and lane 13-15 and Fig. 2-2 G); while MCM1 only show very poor activity even under highest concentration (Fig. 2-2 B and C, lane 3-5 and Fig. 2-2 F and G). When primed M13 without overhang substrate was used, the helicase activity of all three MCMs decreased (Fig. 2-2 A and E), which is consistent with the previous results that MCM prefers the forked substrate. To further confirm this result, an oligo

based fork-like substrate was used for the helicase assay (Fig. 2-2 D and H). MCM2 and MCM3 still exhibit good helicase activity (Fig. 2-2 D lane 8-10 and lane 13-15), while MCM1 shows low activity (Fig. 2-2 D lane 3-5). This result is similar to primed M13 substrates mentioned above. Collectively, MCM2 and 3 process better helicase activity than MCM1.

All MCM proteins required the energy from ATP hydrolysis to translocate along one strand of the DNA duplex and displace the complementary strand. Therefore, the omission of ATP can be served as a control. As expected, when ATP was omitted from the assays no DNA unwinding activity could be detected (Fig. 2-2 A, B, C and D lanes 6, 11 and 16) and only ATP and dATP could support helicase activity (Figure 2-3) as was reported for other member of the MCM family [for example: (87)]. MCM is a member of the AAA+ family of ATPases and thus contain a conserved Lys at the Walker-A motif required for ATP binding and hydrolysis. Therefore, as an additional control, the Walker- A Lys residues in the three MCM proteins were replace by Glu, MCM1 (K571E), MCM2 (K473E) and MCM3 (K335E). When the variant enzymes of MCM1 and 2 were used in lieu of the wild-type protein no helicase activities could be observed (Fig. 2-2 A, B, C and D lanes 7, 12 and 17, see also Fig. 2-3).

The data presented in Figures 2-2 and 2-3 demonstrate that MCM2 and 3 robust helicase activities on short duplex DNA, while MCM1 activity is low. Next, the activity of the enzyme on longer duplex was evaluated. It was found that both MCM2 and MCM3 can unwind the short and long duplex to similar extent (Fig. 2-4) suggesting that both enzymes are processive as was reported for

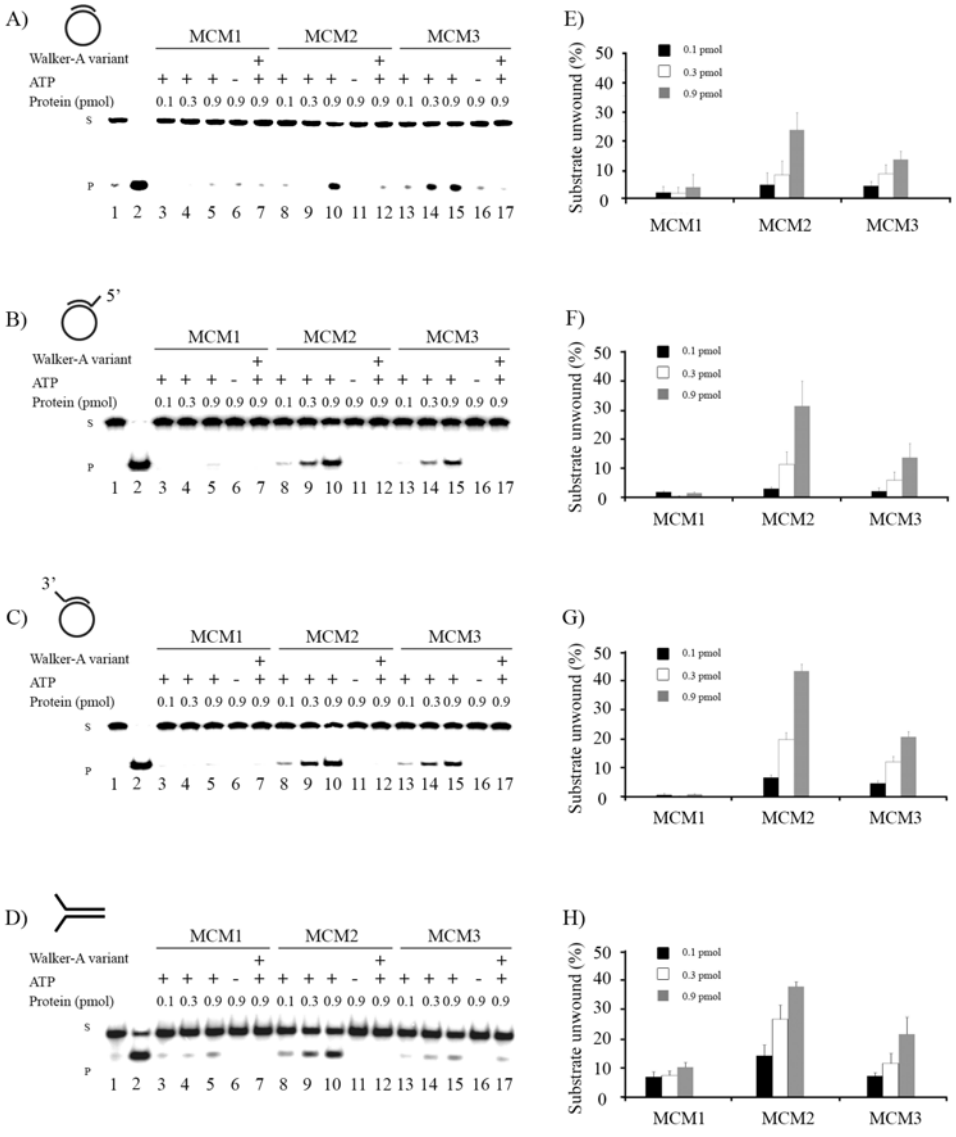



Figure 2-2. MCM2 and MCM3 exhibit better helicase activities than MCM1. Electrophoresis separation of the products of helicase assays with 10 fmol of the $^{32}$P-labeled substrate 36A (A), 36B (B), 36C (C), and 25F (D) (Table 2-4) performed with increasing concentrations of MCM1 (lanes 3-6), MCM2 (lanes 8-11) and MCM3 (lanes 13-16). Control lanes contained aliquots of the substrate (S), product (P) and the products of reaction mixtures incubated with 0.9 pmol of the Walker-A K→E variants (lanes 7, 12 and 17). E, F, G and H provide the average

values, with standard deviations, resulting from quantification of three independent repetitions of the experiments exemplified in (A), (B), (C) and (D).

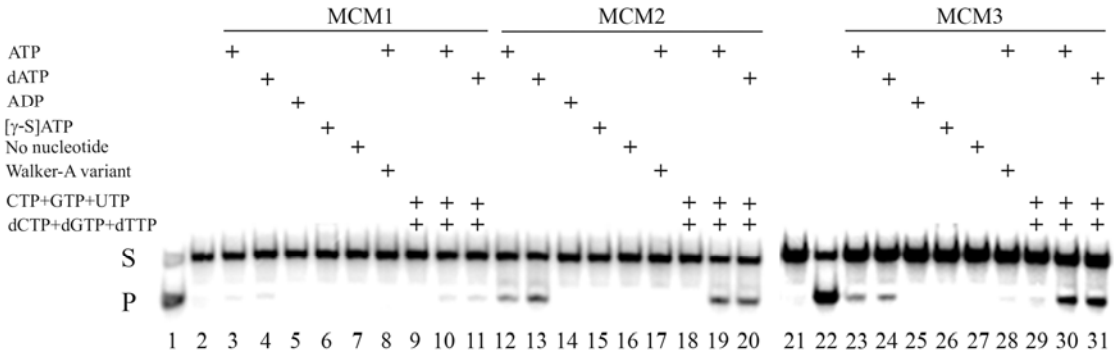



Figure 2-3. Helicase activities of MCM1, MCM2 and MCM3 are dependent on ATP and dATP.

Reaction mixtures contained 10 fmol of the $^{32}$P-labeled substrate 25F and 1 pmol of the indicated MCM protein or Walker-A variant plus the nucleotide(s) listed above each track. Control lanes contained substrate (S) and product (P).

other archaeal MCM helicases [for examples see: (87-89)]. MCM1 still exhibited little activity, which suggests it has poor processivity.

It is also noticed that the Walker-A mutant of MCM3 K335E still processes some substrate unwinding activity (see Fig. 2-2 D, lane 17 and Fig. 2-3A, lane 17) when 25 bp duplex forked substrate was used. It may be due to the different mutation in Walker-A may influence its helicase activity to different level. Sequence analysis of MCM3 indicated a "cryptic" Walker-A motif may exist between the canonical Walker-A and Walker-B motifs. This cryptic Walker-

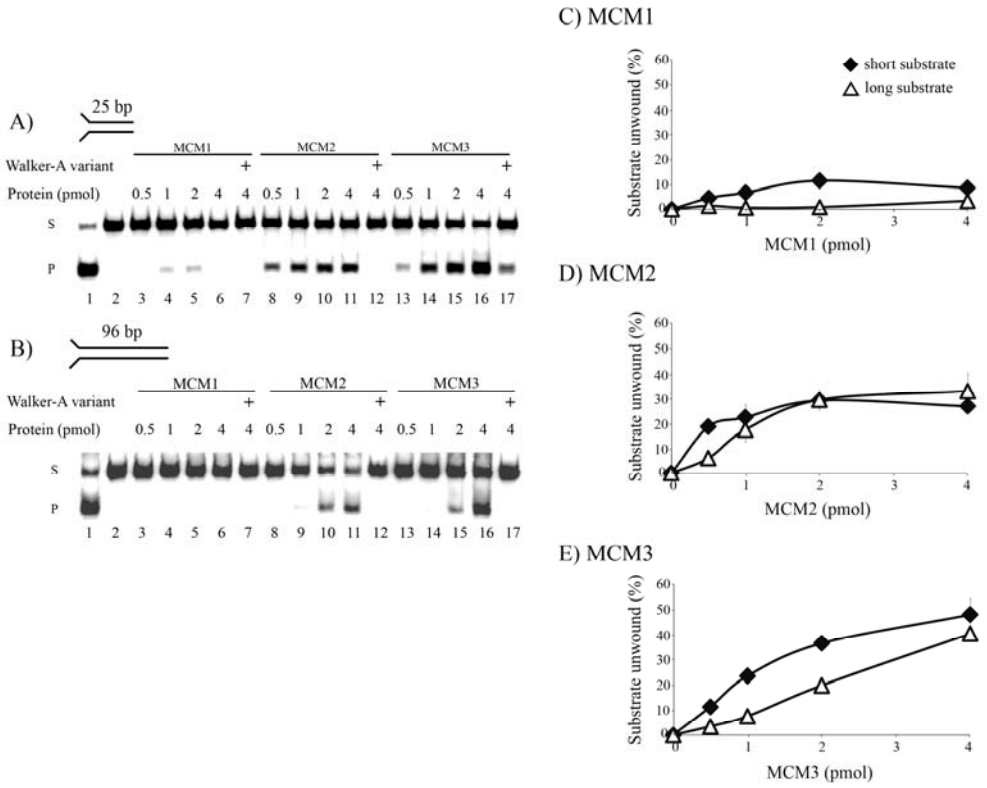



Figure 2-4. Both MCM2 and MCM3 are processive helicases.

Electrophoresis separation of the products of helicase assays with 10 fmol of the $^{32}$P-labeled 25F (A) and 96F (B) forked substrates (Table 2-4) performed with increasing concentrations of MCM1 (lanes 3-6), MCM2 (lanes 8-11) and MCM3 (lanes 13-16). Control lanes contained aliquots of the substrate (S), product (P) and the products of reaction mixtures incubated with 4 pmol of the Walker-A K→E variants (lanes 7, 12 and 17). (C), (D) and (E) provide the average values, with standard deviations, resulting from quantification of three independent repetitions of the experiments exemplified in (A) and (B).

A motif may compensate some function to the canonical one (the mutated lysine residue is marked by a triangle in Fig. 2-1). To test this hypothesis, we

constructed double mutant in which both known Walker-A and the cryptic Walker-A sites were point mutated (K335E/K355E). To further test the influence of the mutations on the protein activity, helicase assay were performed (Fig. 2-5). It is found that MCM3 K335E/K355E still possess limited substrate unwinding ability, which is similar to the MCM3 K335E (Fig. 2-5 A lane 4 and 5, Fig. 2-5 B and also see above helicase assay result). However, in the absence of ATP, MCM3 has no substrate unwinding ability, because its helicase activity is dependent on ATP binding and hydrolysis (Fig. 2-5 C lane 3 and 5); while MCM3 K335E still exhibited substrate unwinding ability, which suggest that this activity is not dependent on ATP (Fig. 2-5 C lane 4 and 6). Therefore, the duplex unwinding by MCM3 K335E is not due to helicase activity. The reason remains to be elucidated.

### 2.4.3 GINS has no major effect on three MCM helicase activity

In eukaryotic cell, MCM2-7 need to form stable complex with CDC45 and the GINS complex for helicase activity (64,70). In archaea, some studies have been performed to test the function of GINS on MCM activity, while the results are different. In *S. solfataricus*, it was reported that GINS could interact with MCM but did not influence its helicase activity (48). In *P. furiosus*, it was shown that GINS can stimulate MCM helicase activity through increasing its ATP hydrolysis ability, although the amount of GINS input into the reaction is substantially over MCM (49). *Thermococcus* is the closest species to *Pyrococcus*. To test if GINS has the similar stimulatory effect in *T. kodakarensis*, we add

different amount of GINS into the helicase assay. Our result indicated that there is no major difference on helicase activity when GINS is present in the MCM1 and MCM2 (Fig. 2-6 A, lane 4-6 and 7-9). However, it is found that GINS slightly increase MCM3 helicase activity (Fig. 2-6 A lane 10-12).

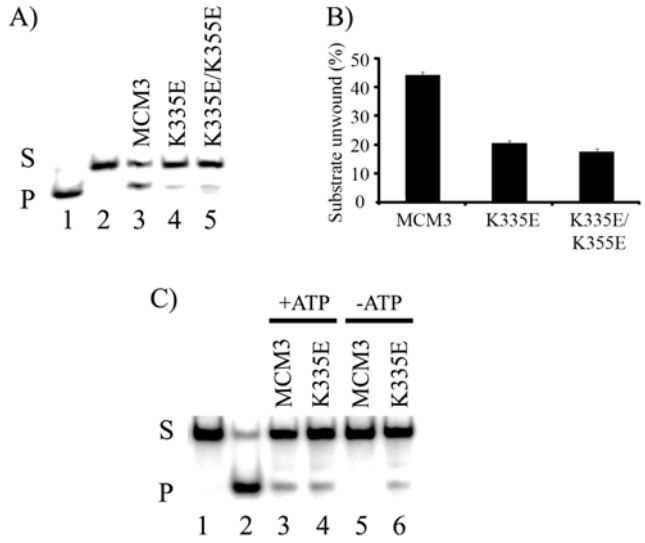



Figure 2-5. The duplex DNA unwinding ability exhibited by MCM3 Walker-A mutant K335E is ATP-independent.

Electrophoresis separation of the products of helicase assays with 10 fmol of the $^{32}$P-labeled substrate 25F (A) performed with 1pmol of MCM3, K335E, K335E/K355E. The averages of three repeats were reported in (B). (C) MCM3 K335E mutant strand unwinding ability is not dependent on ATP. One pmol of MCM3 and K335E were used for helicase assay in the presence (lane 3 and 4) or absence (lane 5 and 6) of 2mM ATP.

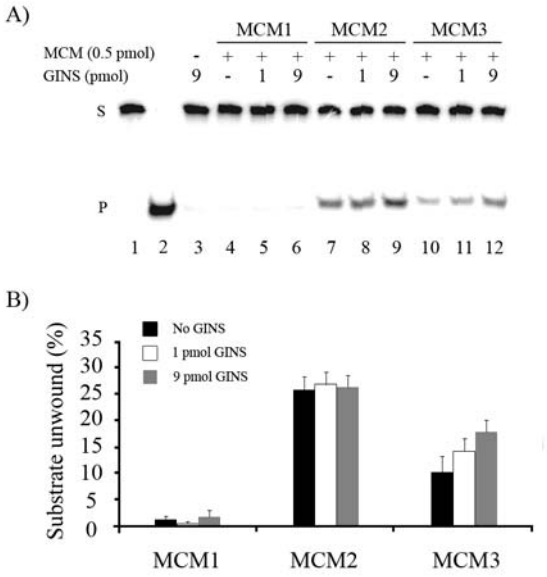
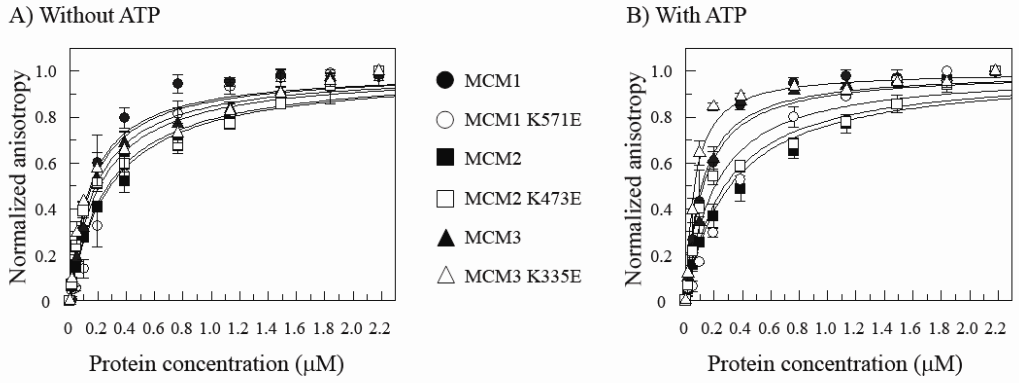



Figure 2-6. GINS has no major effect on MCM1-3 helicase activity.

Electrophoresis separation of the products of helicase assays with 10 fmol of the $^{32}$P-labeled substrate 36B (A). 1 or 9 pmol GINS (as a tetramer) was added into the standard helicase assay with 0.5 pmol of each MCM. (B) is the average results with standard deviation from three repeats.

### 2.4.4 All three MCMs bind DNA and hydrolyze ATP

Helicase activity requires DNA binding followed by ATP-dependent translocation along the DNA substrate. To determine all three MCMs DNA binding ability, fluorescence polarization anisotropy (FPA) was used. Although MCM1 had only minimal helicase activity, it bound DNA in the absence (Fig. 2-7 A) and presence (Fig. 2-7 B) of ATP with affinities similar to that of MCM2, MCM3 and DNA-binding affinities reported for other archaeal MCMs (28,90,91).Then, ATPase assays were performed to evaluate the ability of MCMs




C) Dissociation constant (nM)

| | Without ATP | With ATP |
|---|---|---|
| MCM1 | 142 ± 18 | 107 ± 35 |
| MCM1 K571E | 264 ± 49 | 273 ± 21 |
| MCM2 | 282 ± 3 | 309 ± 58 |
| MCM2 K473E | 206 ± 42 | 203 ± 31 |
| MCM3 | 181 ± 4 | 120 ± 7 |
| MCM3 K335E | 152 ± 17 | 58 ± 10 |

Figure 2-7. MCM1-3 bind DNA with similar affinities.

DNA binding to 10 mM Cy5-labeled oligonucleotide A1 measured by fluorescence polarization anisotropy in the absence (A) and presence (B) of 1 mM ATP. The changes in anisotropy were measured as the MCM protein identified was titrated into the reaction mixture. The apparent dissociation constants calculated are listed in (C).

to hydrolyze ATP. MCM1 and MCM3 had relatively low but readily measurable ATPase activities (Fig. 2-8 A and C), with rates of ATP hydrolysis that were stimulated only ~2-fold by the presence of DNA (Fig. 2-8 D), close to the activities reported for other archaeal MCMs (24,90,92). MCM2, in contrast, exhibited robust ATPase activity (Fig. 2-8 B) and the rate of ATP hydrolysis was

stimulated ~7-fold by the presence of DNA (Fig. 2-8 D). Given that MCM1 and MCM3 have similar affinities for DNA and comparable ATPase activities, it seems unlikely that a deficiency in DNA binding or ATPase activity is the explanation for the inability of MCM1 to unwind DNA *in vitro* (Fig. 2-2 and 2-3).

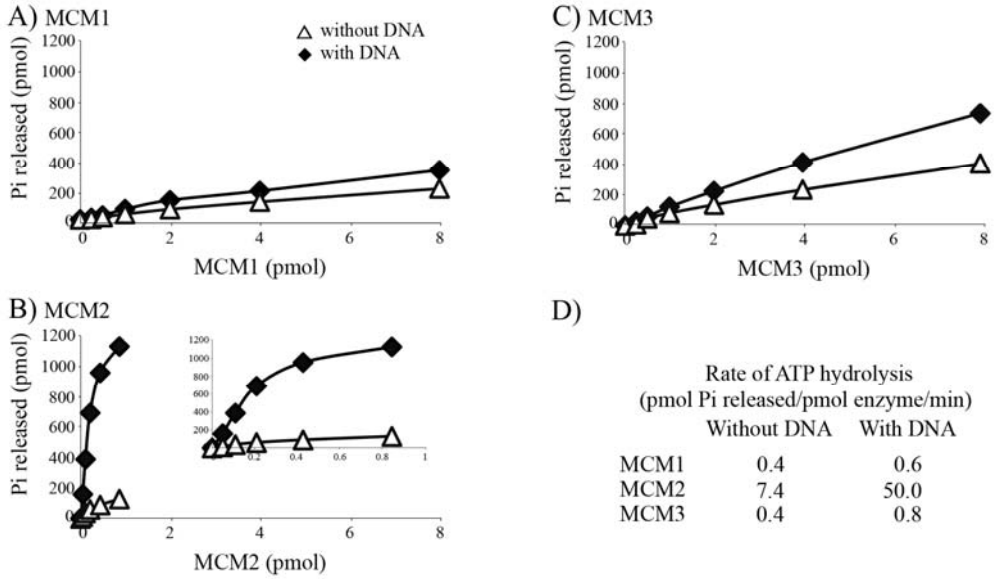



Figure 2-8. MCM2 has higher ATP hydrolysis rate than MCM1 and 3.

Assays were carried in reaction mixtures that contained increasing amounts of (A) MCM1, (B) MCM2 and (C) MCM3 in the presence or absence of the oligonucleotide MD008. The rates of ATP hydrolysis were calculated and (D) shows the averages of the results from three independent experiments. The standard deviations of all experiments were below 10%.

### 2.4.5 Simultaneous measurement of helicase and ATPase activity

To further determine the ATP consumption by MCM during DNA unwinding, helicase and ATPase coupled assay were performed. This experiment

allowed us to monitor the ATPase activity and helicase activity simultaneously. As shown in Figure 2-9, it is clear that MCM1 hydrolyzed low amount ATP corresponding to its exhibited low DNA duplex unwinding activity (Fig. 2-9 A lane 3-7; Fig. 2-9 B lane 3-7 and C); while MCM3 hydrolyzed the lowest amount of ATP among the three MCMs (Fig. 2-9 B lane 13-17 and E), however, it can reach the best unwinding substrate activity when the enzyme concentration increases up to 4 pmol (Fig. 2-9 A lane 13-17 and E). MCM2 is the most complicated to explain. Under low protein concentration, the high level ATP hydrolysis level is corresponding to the high helicase activity (Fig. 2-9 A lane 8-12 and D); while after reach to 1 pmol per reaction, the helicase activity reaches ~40% unwinding saturation (also see the above helicase assay result, Figs. 2-2 and 2-4), but the ATP hydrolysis is still increasing (Fig. 2-9 B lane 8-12 and D). In other words, the helicase activity of MCM2 is not coupled with its own ATPase activity under high enzyme concentration, because at this point, the substrate for ATPase activity is not only the duplex DNA, but those unwound ssDNA also become the substrate to stimulate MCM2 ATPase activity (also see the above ATP hydrolysis assay in Fig. 2-8 B and D). Therefore, it is suggested that only ATPase activity was increased without unwinding activity coupled. Collectively, this assay indicated that the three MCMs exhibited different ATP hydrolysis ability during their duplex unwinding process, and this result was consistent with the ATPase assay and helicase assay shown in other experiments.

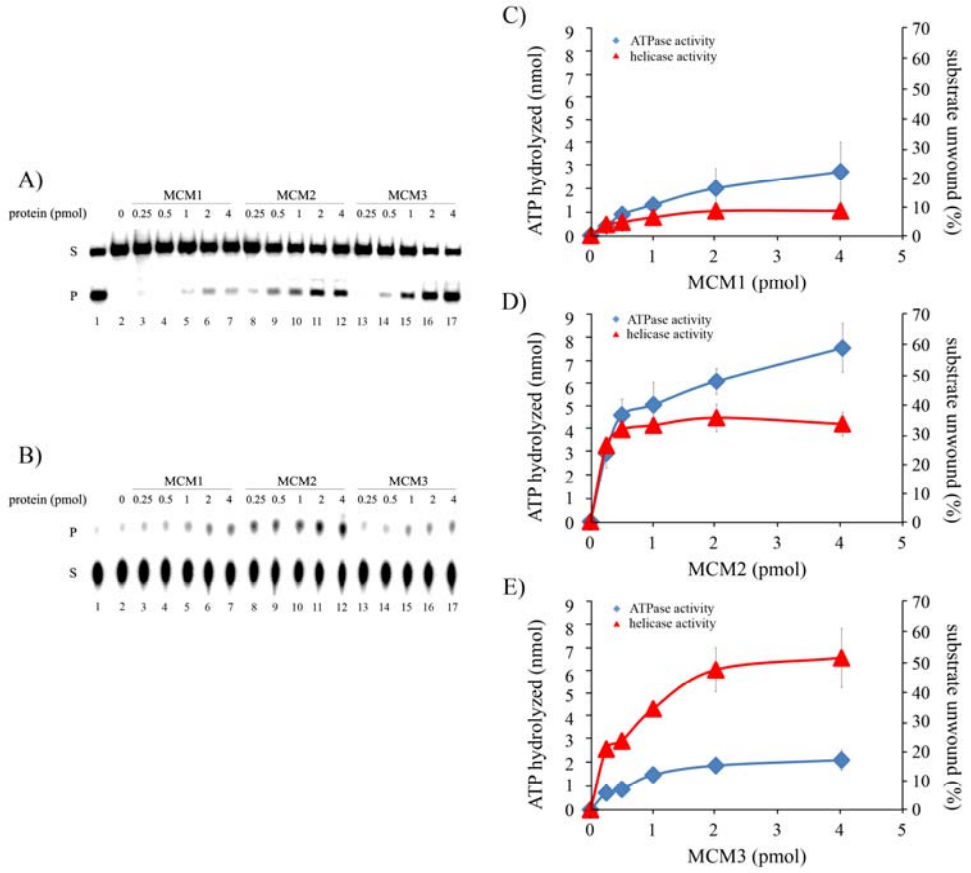



Figure 2-9. Simultaneous measurement of helicase and ATPase activities of MCM1-3.

Standard helicase assays were performed as described in Materials and Methods, with the exception of adding 9000 pmol of [γ-$^{32}$P]ATP into each reaction. Electrophoresis separation of the products of helicase assays with 10 fmol of the $^{32}$P-labeled substrate 25F (A) and thin layer chromatography separation of ATP and Pi (B) with increasing concentrations of MCM1 (lanes 3-7), MCM2 (lanes 8-12) and MCM3 (lanes 13-17). Control lanes contained aliquots of the substrate (S), product (P). (C), (D) and (E) provide the average values, with standard deviations, resulting from quantification of three independent repetitions of the experiments exemplified in (A) and (B). Blue diamond stands for ATPase activity

and red triangle for helicase activity.

### 2.4.6 MCM2 and MCM3 form hexameric complexes

For helicase activity, six MCM subunits assemble to form a hexameric ring-shaped complex (69,73,93,94). Size exclusion chromatography revealed that MCM1 monomers (104 kDa) formed complexes in solution with an estimated molecular mass of ~345 kDa (Fig. 2-10 A) and therefore contained either three (312 kDa) or four (416 kDa) monomers. There was no evidence for assembly into MCM1 hexamers (636 kDa) providing one explanation for the lack of helicase activity. In contrast, MCM2 (94 kDa) eluted from the Superdex-200 column as two protein peaks, the first consistent with a mixture of monomers and dimers (estimated molecular mass of ~140 kDa), and the second with hexameric complexes (estimated molecular mass of ~500 kDa) (Fig. 2-10 B). MCM3 (77.4 kDa) eluted as a single protein peak (Fig. 2-10 C), with an estimated molecular mass of ~440 kDa, consistent with a hexameric complex. Similar estimates for the sizes of the complexes formed by MCM1, MCM2 and MCM3 in solution were obtained by static light scattering (Table 2-5).

Table 2-5. The oligomeric states of MCM1-3 complexes determined by light scattering.

| Protein | Monomeric size (kDa) | Light scattering (kDa) | Oligomeric state |
|---|---|---|---|
| MCM1 | 104.0 | 298.4 | trimer |
| MCM2 | 94.0 | 285.3 and 451.5 | trimer and hexamer |
| MCM3 | 77.4 | 460.8 | hexamer |

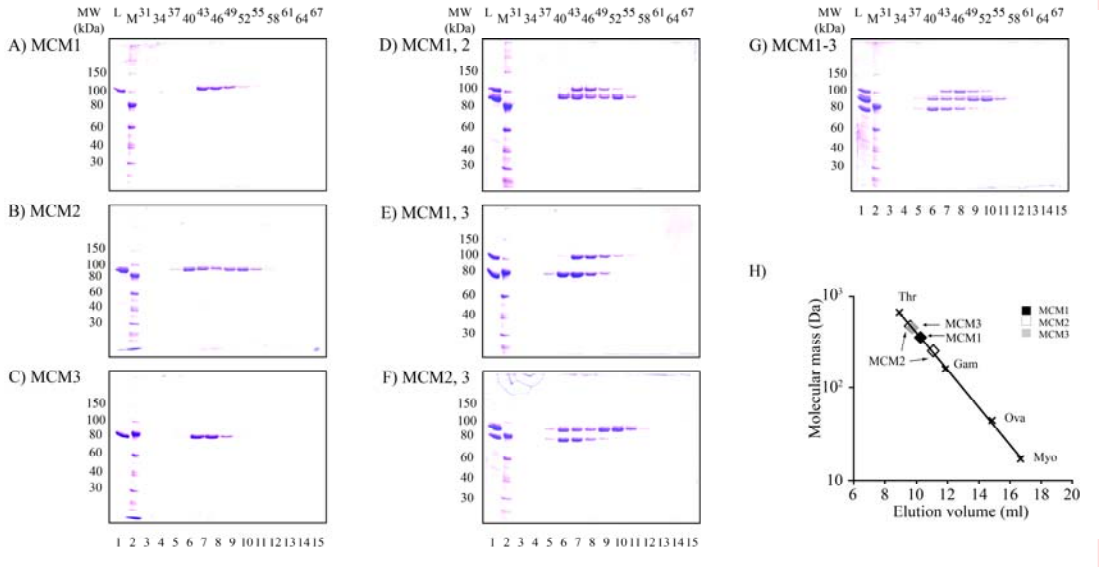



Figure 2-10. Sephadex-200 size exclusion chromatography of MCM1, MCM2 and MCM3 and the mixture of the three MCMs.

The proteins present in fractions from (A) MCM1, (B) MCM2, (C) MCM3, (D) MCM1 plus MCM2, (E) MCM1 plus MCM3, (F) MCM2 plus MCM3 and (G) MCM1 plus MCM2 and 3 separated by Superdex-200 gel filtration chromatography, Coomassie blue stained after resolution by 10% SDS-PAGE. (H) Shows the elution positions of the MCM1, MCM2 and MCM3 complexes relative to those of the mass standards. The positions at which thyroglobulin (Thy, 669 kDa), gamma globulin (Gam, 158 kDa) and ovalbumin (Ova, 44 kDa) eluted from the Superdex-200 column are noted at the top of the figure. Control gel lanes contained aliquots of the material loaded onto the column (L) and size standards (M).

### 2.4.7 MCM1-3 heterocomplex formation test

Based on the results shown above, it has been clear that the three MCMs

have different properties of duplex substrate unwinding and ATP hydrolysis but not DNA binding. To further test if the three MCM could form heterocomplexes, like the eukaryotic MCM2-7, to work together, several assays has been performed. First, the three MCMs were mixed together in the same amount (100 μg) and subjected to Superdex 200 gel filtration column after incubation under 25°C for 1h (Fig. 2-10 D-G). If either two or all of the MCMs interact, they should form a new complex which may change the protein elution peak position. Unfortunately, it is not found that new elution position which is different from the one by each MCM alone (Compare Fig. 2-10 A, B and C with D, E, F and G). This result indicated that there is no heterocomplex formation under our test condition. However, no new elution profile may be due to the resolution limitation of our gel filtration column or the MCM heterocomplex is not stable under the high salt test condition (500 mM NaCl) we used. Therefore, to get a better understanding of the heterocomplex formation, we tried to see if the biochemical property is changed when three MCMs are mixed together.

To test if MCM1-3 have functional interaction, ATPase assays were performed with each MCM alone and the mixture of the three MCMs. In this assay, we tried to mix two or three MCMs in a 1:1 or 1:1:1 molar ratio to see if there is any stimulation or inhibition effect. Under our test condition, only MCM2 exhibit obvious ATPase activity, while MCM1 and 3 activities are very weak, which are negligible compared with MCM2 (Fig. 2-11). When MCM2 was mixed with the same molar ratio with MCM1, or MCM3 or MCM1 and 3 mixtures, the ATPase activity exhibited is similar to the level by MCM2 alone.

The fact that the ATP hydrolysis amount is not changed by the mixture of MCM1 or/and MCM2, indicates MCM1-3 did not have functional interactions, indicating that there is no heterocomplex formation among the three MCMs.

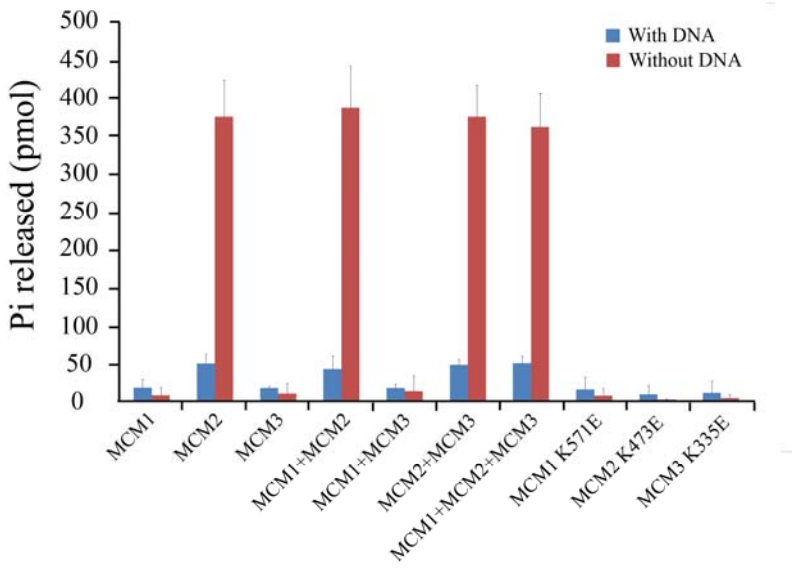



Figure 2-11. ATP hydrolysis by MCM1-3 mixtures indicated that no heterocomplex being formed.

0.1 pmol of MCM1, MCM2 and MCM3 (as monomer) and their Walker-A mutants were used to incubate with (blue bars) or without (red bars) 0.5 pmol ssDNA oligo (Miao 008, Table 2-4) to detect ATP hydrolysis level either by mixture of the two or three MCMs together or each MCM alone. The average result from three independent experiments with standard deviation was shown.

### 2.4.8 Removal of the N-terminal extension increases MCM2 helicase activity

As illustrated in Figures 2-1 and 2-12A, MCM1 and MCM2 have N-terminal extensions, consist of 205 and 136 residues, respectively. These

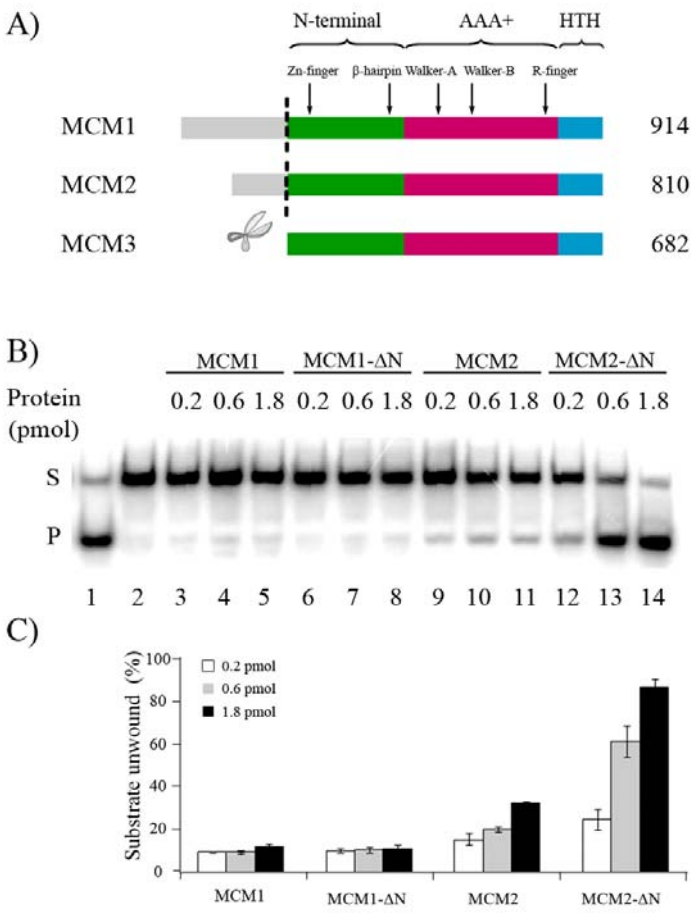



Figure 2-12. N-terminal extension of MCM2 may have regulatory function on its helicase activity.

(A) Illustration of MCM1, MCM2 and MCM3 with the sites at which MCM1 and MCM2 were truncated to generate MCM1-ΔN and MCM2-ΔN (see also Fig. 2-1). (B) Helicase activities of the indicated MCM proteins in reaction mixtures that contained 10 fmol of the $^{32}$P-labeled substrate 25F and increasing concentrations of the MCM protein. Control lanes contained an aliquot of the substrate (S) and product (P). (C) The average values, with standard deviation, calculated from three independent experiments.

extension sequences appear to be unique since no homolog could be identified in database. It is possible that the N-terminal extensions may function as a regulator to inhibit MCM1 helicase activity. To determine the role of the N-terminal extensions on helicase activity, MCM1-ΔN and MCM2-ΔN variants that lacked the N-terminal extensions were generated (Fig. 2-12 A) and purified and their helicase activities determined (Figs. 2-12 B and 2-12 C). As observed for MCM1, MCM1-ΔN had only minimal helicase activity *in vitro*. Surprisingly, in contrast, MCM2-ΔN had higher helicase activity than MCM2 (Figs. 2-12 B and 2-12 C), revealing that the N-terminal extension of MCM2 modulates its helicase activity.

To determine the reason that influences MCM2-ΔN has better helicase activity, ATPase activity and DNA binding ability of MCM1-ΔN and MCM2-ΔN have been investigated and compared with the full length MCM1 and MCM2. It was shown that there is no obvious difference between full length MCMs and truncated MCMs regarding to DNA binding (Fig. 2-13 C) and ATPase activity (Fig 2-13 A and B). Therefore, they are not the reasons to explain the better helicase activity by MCM2-ΔN. It is also shown that MCM2-ΔN is like other MCMs whose helicase activity is ATP-dependant (Fig. 2-14, lane 3 and 5). However, there may be another possibility that the MCM2-ΔN has different oligomeric state from the full length MCM2. Archaeal MCM has been shown to be active as a hexamer (29). When MCM2 was analyzed by gel filtration and light scattering, two populations were found (hexamer and monomer, see Fig. 2-10 and Table. 2-5). It is possible that MCM2-ΔN become more stable as a hexamer, therefore leads to better helicase activity. Future experiment will need

to confirm this hypothesis.

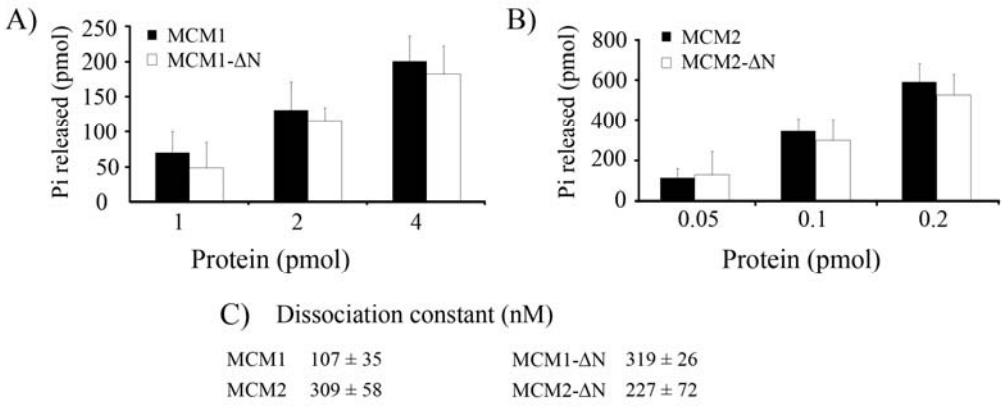



Figure 2-13. MCM1-ΔN and MCM2-ΔN have similar ATPase and DNA binding activities to the full length MCM1 and MCM2. (A and B) ATPase assays were carried out in reaction mixtures that contained increasing amounts of (A) MCM1 and MCM1-ΔN and (B) MCM2 and MCM2-ΔN in the presence of 10 pmol of the oligonucleotide MD008. The amount of ATP hydrolysis was calculated and the average values, with standard deviation, calculated from three independent experiments are shown. (C) DNA binding to 10 nM Cy5-labeled oligonucleotide A1 measured by fluorescence polarization anisotropy in the presence of 1 mM ATP. The changes in anisotropy were measured as the MCM protein identified was titrated into the reaction mixture. The apparent dissociation constants calculated are listed.

### 2.4.9 Only MCM3 is essential for *T. kodakarensis* viability

*T. kodakarensis* TS601 (ΔTK0096) and TS602 (ΔTK1361) were constructed without difficulty generating strains lacking MCM1 and MCM2,

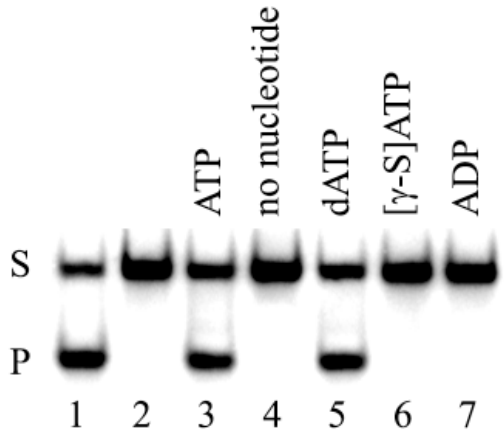



Figure 2-14. MCM2-ΔN helicase activity is dependent on ATP and dATP. Helicase assays of MCM2-ΔN with and without ATP. Reaction mixtures contained 10 fmol of the 25 bp $^{32}$P-labeled substrate and 1 pmol of the MCM2-ΔN protein plus the nucleotide(s) listed above each track. Control lanes contained substrate (S) and product (P).

respectively. *T. kodakarensis* TS604 (ΔTK0096; ΔTK1361) was then generated from *T. kodakarensis* TS601 with both TK0096 and TK1361 deleted. The strategies and results of the diagnostic PCRs (Fig. 2-15 B) and Southern blots (Fig. 2-15 C) that confirmed the genome organizations in *T. kodakarensis* TS601, TS602 and TS604 are shown in Figure 2-15 A. Despite repeated attempts, we have been unable to generate a strain with TK1620 (MCM3) deleted arguing that MCM3 is likely required for *T. kodakarensis* viability. As *T. kodakarensis* TS604 (TK0096; ΔTK1361) exhibits no detectable growth defects, the presence of MCM3 is apparently sufficient for genome replication and therefore *T. kodakarensis*, in common with most *Archaea*, requires only one MCM for chromosomal replication.

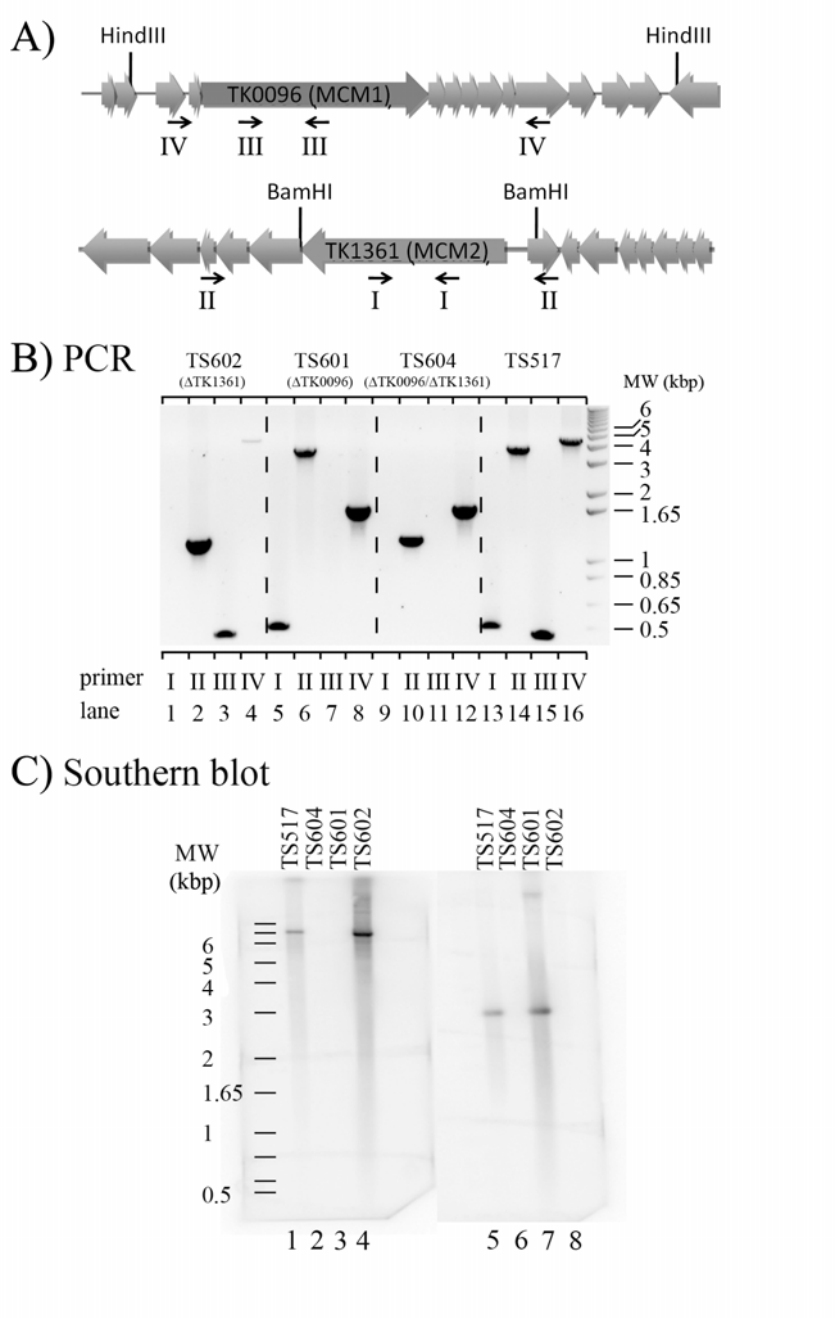



Figure 2-15. Genome organizations, PCR and Southern blot confirmation of the *T. kodakarensis* ΔTK0096 and ΔTK1361 deletions.

(A). Genome organizations surrounding TK0096 (MCM1) and TK1361 (MCM2). The positions at which the PCR primers (Roman numeral primer pairs I through IV, Table 2) hybridized and the locations of the HindIII and BamHI sites used in

the Southern blot analyses are shown.

(B). Agarose gel electrophoretic separation of PCR amplicons from genomic DNA of *T. kodakarensis* TS517 (Δ*pyrF*, Δ*trpE*::*pyrF*, ΔTK0664), TS601 (Δ*pyrF*, Δ*trpE*::*pyrF*, ΔTK0664, ΔTK0096), TS602 (Δ*pyrF*, Δ*trpE*::*pyrF*, ΔTK0664, ΔTK1361), and TS604 (Δ*pyrF*, Δ*trpE*::*pyrF*, ΔTK0664, ΔTK0096, ΔTK1361) with the positions of DNA size standards indicated. As shown, primers internal to TK1361 (primer pair I, panel A) amplified a ~600 bp molecule from *T. kodakarensis* TS517 and TS601 but failed to generate an amplicon from *T. kodakarensis* TS602 and TS604 genomic DNAs. Primers hybridizing to sequences that flank TK1361 (primer pair II) generated ~3,800 bp amplicons, which contain the TK1361 sequence (2,433 bp) from *T. kodakarensis* TS517 and TS601 but amplicons that were only ~1,400 kbp from *T. kodakarensis* TS602 and TS604 genomic DNAs consistent with the loss of TK1361. Primers specific to TK0096 (primer pair III) amplified an ~450 bp amplicon from *T. kodakarensis* TS517 and TS602, but failed to generate an amplicon from *T. kodakarensis* TS601 and TS604 genomic DNAs. Primers that hybridized to sequences flanking TK0096 (primer pair IV) generated an ~4,400 bp amplicon, which included the TK0096 sequence (2,745 bp) from *T. kodakarensis* TS517 and TS602 but amplicons that were only ~1,600 bp amplicon from *T. kodakarensis* TS601 and TS604 genomic DNAs consistent with the loss of TK0096.

(C). Southern blot analyses of genomic DNA. Genomic DNA (10μg) from *T. kodakarensis* TS517 (Δ*pyrF*, Δ*trpE*::*pyrF*, ΔTK0664); TS601 (Δ*pyrF*, Δ*trpE*::*pyrF*, ΔTK0664, ΔTK0096); TS602 (Δ*pyrF*, Δ*trpE*::*pyrF*, ΔTK0664,

ΔTK1361), and TS604 (Δ*pyrF*, Δ*trpE*::*pyrF*, ΔTK0664, ΔTK0096, ΔTK1361) was digested with HindIII (left panel) and BamHI (right panel). The products were separated by electrophoresis through 0.8% agarose gels, denatured and transferred to a Zeta-probe membrane as previously described (81). The membranes were incubated with a $^{32}$P-labeled oligonucleotide that hybridized to a sequence internal toTK0096 (left panel) or TK1361 (right panel). TK0096 is located within an ~7,000 bp HindIII fragment that was present only in *T. kodakarensis* TS517 and TS602 genomic DNAs. TK1361 is located within an ~3,000 bp BamHI fragment (right panel) that was present only in *T. kodakarensis* TS517 and TS602 genomic DNAs. The data was generated by Dr. Thomas Santangelo, Ohio State University.

## 2.5 Discussion

### 2.5.1 MCM3 is the replicative helicase in *T. kodakarensis*

All three MCM proteins predicted by bioinformatics to exist in *T. kodakarensis* have been purified and all three bind DNA and have ATPase activity, dependent on an intact Walker-A motif (Fig. 2-11). However, only MCM2 and MCM3 spontaneously assembled into hexameric complexes and exhibited robust helicase activity *in vitro*. In contrast, recombinant MCM1 does not form stable hexamers and exhibited barely detectable helicase activity *in vitro*. Deletion of the genes encoding MCM1 (TK0096) and MCM2 (TK1361) had detectable effects on growth or viability arguing that these MCM homologs are not essential for replication provided MCM3 is present. In contrast, our inability to delete

TK1620 strongly suggests, although it does not categorically prove, that MCM3 is essential for viability in *T. kodakarensis* TS517 and is likely the predominant and possibly the only MCM catalyzing *T. kodakarensis* genome replication. Providing further support to this conclusion, MCM3 is the most similar in size to other archaeal MCMs, and TK1620 is apparently part of an operon that also encodes the GINS23 (TK1619) subunit of the replisome (45,95). MCM1 and MCM2 have atypical N-terminal extensions (Figs. 2-1 and 2-15A) and are encoded by genes that are not closely linked in the *T. kodakarensis* genome to genes that encode known replication proteins. Both the initial genome annotation and a subsequent in-depth bioinformatics analysis predict that TK0096 (MCM1) and TK1361 (MCM2) are present in regions of the *T. kodakarensis* genome that originated from past viral infections (12,78). Thus, it seems a reasonable hypothesis that MCM1 and MCM2 are vestigial viral helicases, and that MCM3 is the endogenous archaeal MCM in *T. kodakarensis*. The same explanation, namely that one MCM homolog is the endogenous enzyme and any additional MCMs present were acquired through replicon infection, may also apply to the other archaea now found to have multiple MCMs (96).

### 2.5.2 Do MCM1 and MCM2 have non-essential functions in *T. kodakarensis*?

As MCM1 and MCM2 can be deleted, they do not have essential functions in *T. kodakarensis* TS517 synthesizing MCM3. In the past, MCM1 and/or MCM2, most likely in collaboration with host proteins, may have participated in the replication of infecting viruses or plasmid DNA or in the activation of a prophage.

Their N-terminal extensions perhaps play roles regulating the viral replication. This would resemble the role of the simian virus 40 (SV40) large T antigen (97,98) that, together with host proteins, contributes to both SV40 origin recognition and functions as the viral replicative helicase. As TK0096 and TK1361 are still expressed *in vivo* (27), it seems also likely that MCM1 and MCM2 now also do have similar functions in archaeal viral replications in ancestors of the *T. kodakarensis* lineage with these activities regulated by their atypical N-terminal extensions.

When MCM1-$His_6$ and MCM2-$His_6$ were isolated from *T. kodakarensis* cell lysates by binding to a $Ni^{2+}$-charged matrix, they were isolated in complexes that also contained the archaeal DNA polymerases B and D, and the processivity factor PCNA1 (27). Their presence in such complexes *in vivo* argues for MCM1 and MCM2 participating in DNA metabolic processes, possibly in recombination and/or DNA repair in *T. kodakarensis*. In this regard, it is noteworthy that MCM2 also co-purified from *T. kodakarensis* cell lysates with a MutS homologue, an established DNA repair enzyme (27). The *T. kodakarensis* genome encodes two PCNA homologs, although biochemical and structural studies argue that only PCNA1 (encoded by TK0535) has properties in common with all other archaeal PCNAs, and only PCNA1 is essential for *T. kodakarensis* viability (Chapter 3). Studies are now underway to determine if *T. kodakarensis* strains lacking MCM1, MCM2 and/or PCNA2 exhibits defects in DNA repair and/or recombination.

# Chapter 3 Biochemical and genetical characterization of multiple PCNAs in *T. kodakarensis*

## 3.1 Abstract

PCNA forms ring-shaped molecules that encircle duplex DNA and provide essential contacts for, and stability and processivity to, many DNA metabolic processes. All eukaryotic and most euryarchaea genomes contain a single gene encoding for PCNA, while *T. kodakarensis* is unique by containing two genes, TK0535 and TK0582, that encode for PCNA1 and PCNA2, respectively. It is shown here that these two PCNA proteins stimulate DNA synthesis by PolB, and ATP hydrolysis by the RFC complex. Examination of their roles in *vivo* revealed that PCNA1 is very highly expressed compared to PCNA2 in rapidly dividing cells, and that only PCNA1 was essential for viability. The implications of these observations for PCNA function, and the origin of the two PCNA-encoding genes, are discussed.

## 3.2 Introduction

PCNA is a ring-shaped trimeric complex that encircles duplex DNA and plays an essential role in many DNA metabolic processes in Archaea and Eukarya (99,100). PCNA cannot autonomously assemble around DNA, but must be loaded onto DNA by the clamp loader, RFC complex. RFC recognizes the 3' end of the single-strand/duplex (primer-template) junction and utilizes ATP hydrolysis

to assemble the PCNA ring around the primer (101).

The first reported function of PCNA was as a processivity factor for the replicative DNA polymerases. PCNA encircles DNA and upon binding to a replicative polymerase, tethers it to the DNA template for processive DNA synthesis. Subsequent studies showed that PCNA proteins from Archaea and Eukarya also associate and modulate the activities of a large number of other proteins involved in nucleic acid metabolic transactions and the regulation of the cell cycle (102).

The Archaea can be divided into three main branches: crenarchaeota, euryarchaeota and thaumarchaeota. While most Archaea belonging to the euryarchaeota and thaumarchaeota kingdom encode a single PCNA homologue that forms homotrimers, members of the crenarchaeota domain contain three distinct genes that result in the formation of heterotrimers as the active form (61,103). Most archaeal species also contain two homologues of RFC, referred to as RFC large (RFC-L) and short (RFC-S), that form a pentameric complex of one copy of RFC-L with four subunits of RFC-S.

*T. kodakarensis*, a hyperthermophilic euryarchaea that grows optimally at 85°C, was isolated from a solfatara on the shore of Kodakara Island, Japan (11). *T. kodakarensis* is unique among euryarchaeal species by containing two genes encoding for PCNA homologues, TK0535 and TK0582, referred to here as PCNA1 and PCNA2, respectively. It was previously shown that the two genes are expressed *in vivo* (27) and form homotrimeric rings [Fig. 3-1, (83)]. In this study the biochemical properties of the two PCNA proteins were determined as

well as their *in vivo* roles. It was found that although both proteins have similar biochemical properties, only PCNA1 is essential for cell viability.

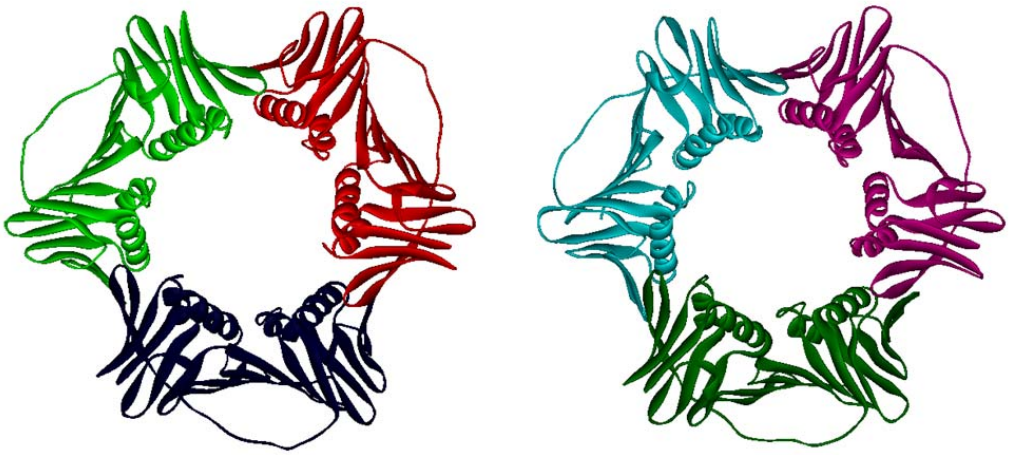



Figure 3-1. The crystal structures of *T. kodakarensis* PCNA1 and PCNA2. Both PNCA is composed of three subunits and form homotrimer. PDB ID codes 3LX1 and 3LX2 for PCNA1 and PCNA2 respectively (83).

## 3.3 Materials and methods

### 3.3.1 Protein expression and purification

The two *T. kodakarensis* encoded-PCNA proteins, the *T. kodakarensis* RFC complex and *T. kodakarensis* PolB were purified as previously described (41,83). The genes encoding for the three RPA genes (TK1959, TK1960 and TK1961) were amplified using PCR from *T. kodakarensis* genomic DNA and cloned into pET-21a (Novagen) to generate pET21a-TK1959, pET21a-TK1960 and pET21a-TK1961. All three constructs contain an in-frame six histidine tag

(His6) at the C-terminus. The three RPA proteins were expressed in *E. coli* BL21 DE3 Rosetta cells and purified on a $Ni^{2+}$ column as previously described for the expression and purification of RPA from *M. thermautotrophicus* (54). The RPA complex was formed by mixing the three proteins in a 1:1:1 molar ratio.

### 3.3.2 Gel filtration analysis

An aliquot (200 μg) of each PCNA protein and RFC complex in 200 μl buffer containing 25 mM Tris–HCl (pH 7.5), 500 mM NaCl and 10% glycerol (v/v) was fractionated through a Superdex-200 gel-filtration column (HR10/30; GE Healthcare) pre-equilibrated with 25 mM Tris–HCl (pH 7.5), 500 mM NaCl and 10% glycerol (v/v) at 22°C. Fractions (15 μl) were separated though a 10% SDS-PAGE and visualized by staining with Coomassie brilliant blue (R250).

### 3.3.3 Static light scattering of PCNA and RFC

The solution molecular mass of each PCNA protein and RFC were determined using size exclusion chromatography within-line multiangle light scattering. A 1200 series HPLC system (Agilent Technologies) with a Shodex KW-802.5 or a Shodex KW-804 column (Showa Denko K.K.) was used for this purpose. For each PCNA, different amount of purified proteins as shown in Figure 3-2 was detected. The flow rate was 0.5 ml/min, and the mobile phase was 0.01 M phosphate buffer (pH 6.5), 0.1 M ammonium sulfate, 0.1 mM EDTA, and 0.01% sodium azide. For RFC complex, the molecular mass was determined using 100 μg of protein in 20 μl 25 mM Tris-HCl (pH 7.5), 50 mM NaCl and 10%

glycerol. The flow rate was 0.5 ml/min, in a solution containing 25 mM Tris–HCl (pH 7.5), 500 mM NaCl and 10% glycerol (v/v) (Chapter 2). Light scattering was measured with a miniDawn Treos (Wyatt Technology); the protein concentration was measured with an Optilab rEX differential refractometer (Wyatt Technology).

### 3.3.4 ATPase assays

ATPase activity of RFC was assayed in reaction mixtures (15 μl) containing 25 mM Tris-HCl (pH 8), 5 mM $MgCl_2$, 1 mM DTT, 1.5 μg BSA, 1.5 nmol of [γ-$^{32}$P]ATP (3000 Ci/mmol), 0.5 pmol of RFC and 0.01, 0.05, 0.1, 0.25 and 0.5 pmol of PCNA1 or PCNA2 proteins (as trimers) as indicated in the figure legend in the presence or absence of 50 pmol of primed substrate formed by annealing oligonucleotides MD001 (5'-GCGGCGAGTCCAGCTCAGGAGCTC GCGCCG) and MD002 (5'-TTTGTTTGTTTGTTTGTTTGTTTGTTTGTTTGTT TGCGGCGCGAGCTCCTGAGCTGGACTCGCCGC). After incubation at 70°C for 1 h, an aliquot (1 μl) of the reaction mixture was spotted on a polyethyleneimine cellulose thin layer plate. ATP and Pi were separated by chromatography in 1 M formic acid containing 0.5 M LiCl, and the extent of ATP hydrolysis was calculated based on phosphorimaging quantification. The ATPase assays were repeated three times, and the averages of the results obtained with standard deviations are shown in the Figure 3-5.

To establish the rate of ATP hydrolysis by RFC, reaction mixtures (45 μl) containing 25 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 1 mM DTT, 100 μg/ml BSA, 4.5 nmol [γ-$^{32}$P]ATP (3000 Ci/mmol), 0.5 pmol RFC and 0.05 pmol PCNA were

incubated at 70ºC, with or without 50 pmol of primed substrate as described above. Aliquots (3 μl) of the reaction mixture were removed after 0, 5, 10, 15, 30 and 45 min, mixed with 1 μl of 0.5 M EDTA, and the extent of ATP hydrolysis determined using polyethyleneimine cellulose thin layer chromatography as described above.

### 3.3.5 PCNA protein expression

Fifty ml cultures of *T. kodakarensis* cells (KW128) were grown as previously reported to $OD_{600}$ of 0.5-0.6 resulting in 2-5×$10^8$ cells/ml. The cells were harvested by centrifugation at 3000 rpm for 20 min at 23°C. The cell pellet was resuspended in 500 μl 25 mM Tris-HCl (pH 8.0), 500 mM NaCl and 10% glycerol followed by sonication. The mixture was clarified by centrifugation at 13,000 rpm for 10 min at 4˚C.

Known amounts of recombinantly-produced *T. kodakarensis* PCNA1 and 2 were resolved alongside specific volumes of cell extract, representing known numbers of cells, on 12% SDS-PAGE followed by electrobloting onto a nitrocellulose membrane. Western analysis was performed using a combination of guinea pig polyclonal antibodies (Cocalico Biologicals Inc) generated against the recombinant PCNA proteins and rabbit anti-guinea pig antibodies coupled to horse-radish peroxidase (Sigma-Aldrich); blots were developed using enhanced chemiluminescence (ECL) (GE Healthcare). The intensity of each band was quantified using ImageQuant TL (GE Healthcare). The experiment was repeated three times. The average results of three experiments were used to calculate the *in*

*vivo* levels of the PCNA proteins.

### 3.3.6 DNA replication assay

Elongation assays of singly primed M13 DNA templates were performed as reported previously (83). A reaction mixture (20 μl) containing 40 mM Tris-HCl (pH8.0), 250 mM NaCl, 1.5 mM DTT, 0.01% BSA, 10 mM Mg-acetate, 2 mM ATP, 100 μM each of dTTP, dCTP and dGTP and 20 μM [$\alpha$-$^{32}$P]dATP and 10 fmol singly primed M13 mp18. The proteins used in the reaction were indicated as in Figure 4-4. Following incubation at 70ºC for 20 min the reaction was stopped by the addition of EDTA to final concentration of 10 mM. The products were separated by electrophoresis through 1.1% alkaline agarose gel. Total nucleotide incorporation was quantified by liquid scintillation (GE Healthcare) from 4μl aliquots of each reaction.

### 3.3.7 Constructs and techniques to generate *T. kodakarensis* knock out strains

TS032 and TS033 target the genomic loci of TK0535 and TK0582 respectively, and were constructed using the same principles employed to design similar vectors as described [Chapter 2 and (30)]. The selective and counter-selective pressures employed to generate intermediate and markerless deletion strains were as described [Chapter 2 and (30)]. Briefly, TS032 and TS033 were independently transformed into *T. kodakarensis* strain TS517 with selection on artificial seawater (ASW) based solid medium containing amino acids (AA) but lacking tryptophan. Colonies arising from double crossover homologous

integrations at the desired locus were identified by diagnostic PCR, and then spread on ASW-AA medium containing 6-methyl purine (6MP). Intermediate strain construction was possible for each construct, whereas only TS033, targeting TK0582 (PCNA2) allowed viable excision of the markers to generate a markerless deletion of TK0582. Figure 3-6 outlines the diagnostic PCRs confirming loss of TK0582 from the genome.

## 3.4 Results

### 3.4.1 The oligomeric states of PCNA1, PCNA2 and RFC

*T. kodakarensis* PCNA1 and PCNA2 form very similar trimeric ring-shaped molecules that differ significantly only at the subunit boundaries, excluding stable heteromeric complexes of the two PCNA proteins (83). The two PCNA trimers have different stability (Figs. 3-2 and 3-3). At the concentration used for crystallization, 13 mg/ml, PCNA2 gave a molar mass average of 87.0 kDa which is close to the calculated molar mass of 87.8 kDa for a trimer. The experimental molar mass average for PCNA1 at the concentration used for crystallization, 6.25 mg/ml, was 67.6 kDa which is significantly low for a trimer (87.2 kDa) and significantly high for a dimer (58.1 kDa). When lower concentrations were used, the light scattering indicated that PCNA2, under the solution conditions used, remained a stable trimer. Under the same conditions, the structure of PCNA1 was most likely rapidly alternating between a dimer and a trimer at high concentrations and between dimer and monomer at lower concentrations. The different stabilities of the two PCNAs are also reflected in

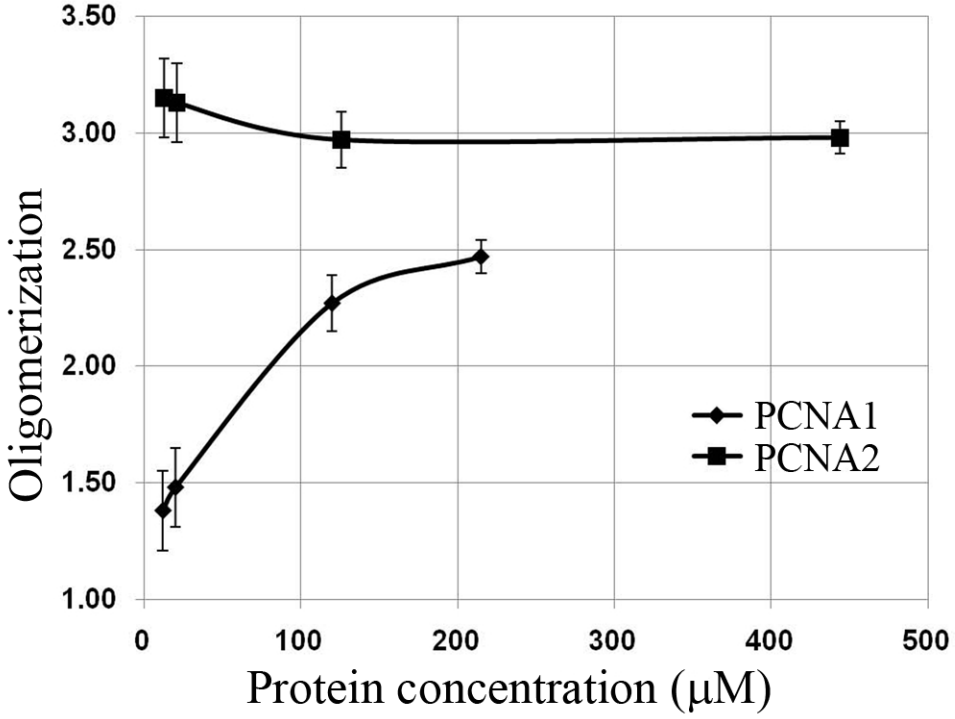



Figure 3-2. PCNA2 is a stable trimer in solution. Static light scattering of dilution series of the two PCNA proteins demonstrate the stability of the trimer of PCNA2 and the instability of the trimer of PCNA1 at low protein concentrations.

their ability to support DNA polymerase activity.

To further confirm the above result, we used size exclusion chromatography to determine the oligomeric state of archaeal PCNA proteins. Similar to the light scattering results, PCNA1 is not a stable trimer and eluted from the column as a protein with an approximate molecular mass of 62.0 kDa (Fig. 3-3 A and B). The results suggest that PCNA1 is in equilibrium between trimers (87.2 kDa) and dimers (58.1 kDa). PCNA2, in contrast, eluted in two major peaks, (Fig. 3-3 C and D) suggesting a mixture of complexes with molecular masses of 86.4 kDa and 177.9 kDa, consistent with trimeric (87.8 kDa)

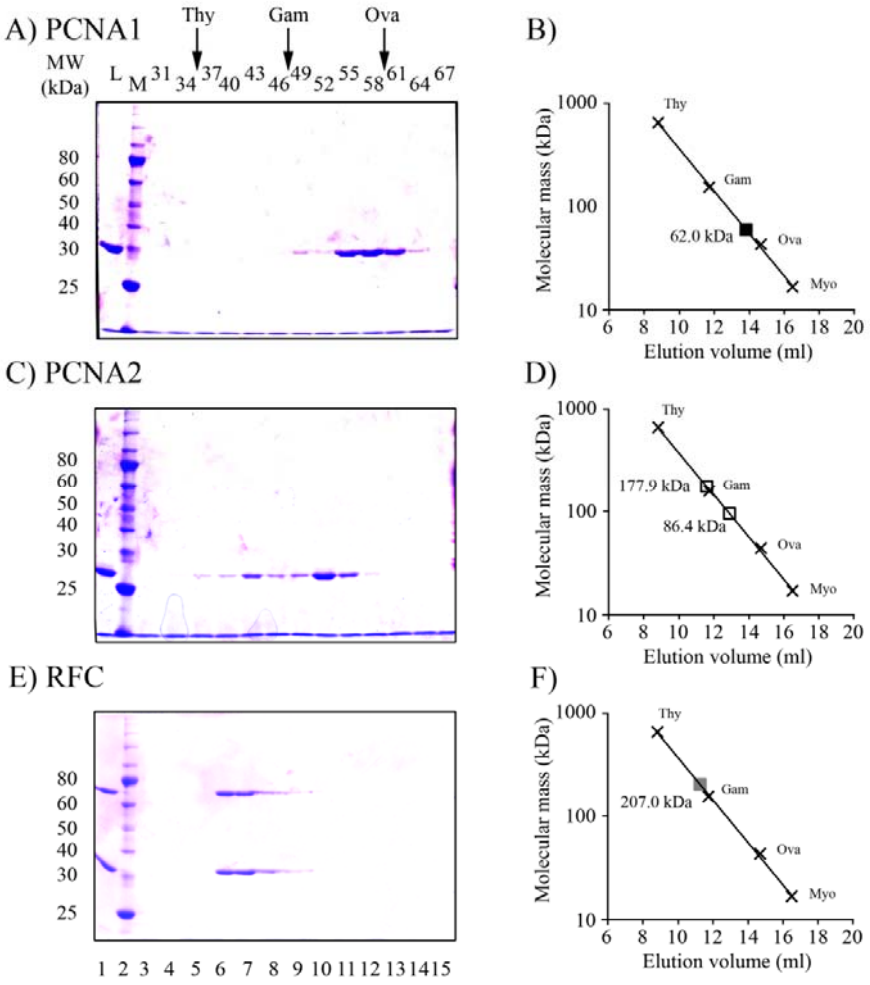



Figure 3-3. Size exclusion chromatography of PCNA1, PCNA2 and the RFC complex. PCNA1, PCNA2 and the RFC complex were analyzed using Superdex-200 gel filtration column as described under "Experimental Procedures". Aliquots (15 µl) of each fraction were subjected to 10% SDS-PAGE analysis followed by Coomassie Blue staining. A. PCNA1; C, PCNA2 and E, RFC. The peak positions of thyroglobulin (Thy, 669 kDa), gamma globulin (Gam, 158 kDa) and ovalbumin (Ova, 44 kDa) are indicated at the top of the figure. L, load on; M, molecular mass standard. The peak elution in respect to standards are shown in panels (B), PCNA1; (D) PCNA2 and (F) RFC.

and hexameric (175.6 kDa) structures, respectively. The human PCNA may also form hexameric rings (104).

When the RFC complex was analyzed on gel-filtration (Fig. 3-3 E and F) it elute in a single peak corresponding to a protein with molecular mass of 207.6 kDa. This suggest, similar to all other archaeal RFC studied, that the *T. kodakarensis* RFC is a complex in which the large subunit (58.1 kDa) is associated with four small subunits (37.2 kDa) resulting in a complex pentameric complex of 206.9 kDa. Light scattering experiments revealed a molecular mass of 203.0 kDa for the RFC complex also consistent with a pentameric complex.

### 3.4.2 Both PCNA1 and PCNA2 stimulate the activity of PolB.

The best understood function for PCNA is as the processivity factor for the replicative DNA polymerases (99). To determine if both *T. kodakarensis* PCNA proteins can stimulate *T. kodakarensis* DNA polymerase, replication assays using *T. kodakarensis* PolB were performed in the presence and absence of PCNA1 and PCNA2. The elongation of a singly primed M13 template by PolB was detected only in the presence of both RFC and PCNA1 or PCNA2, and the rate of DNA synthesis was depended on the level of PCNA added (Fig. 3-4, lanes 9-11 and 6-8). Thus, both PCNA complexes retain function, although at low PCNA1 concentrations the size of DNA products formed were reduced in comparison to PCNA2 (Fig. 3-4, compare lanes 10 and 11 to lanes 7 and 8). This effect is likely due to the instability of the PCNA1 trimer at low concentrations (Figs. 3-2 and 3-3 A and B).

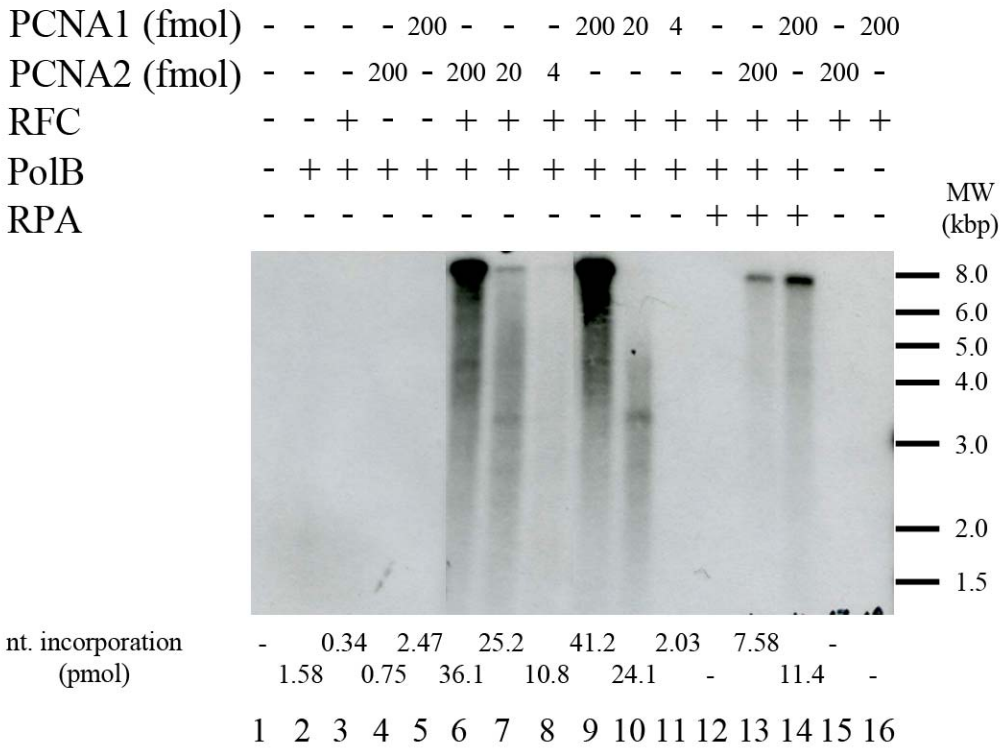



Figure 3-4. Both PCNA1 and PCNA2 stimulate polymerase activity. DNA replication assays were performed as described under "Experimental Procedures" using 440 fmol of PolB, 430 fmol of RFC (as pentamers), 6000 fmol of RPA (as trimmers) and different amount of PCNA1 or PCNA2 as indicated. Reactions were incubated for 20 min at 70°C. An aliquot (4 µl) was used to measure DNA synthesis, and the remaining mixture was subjected to 1.1% alkaline-agarose gel electrophoresis. After drying, gels were autoradiographed for 1 hour at -80°C and then developed. The data was generated by Dr. Jerard Hurwitz, Memorial Sloan-Kettering Institute.

In Bacteria and Eukarya, the replicative polymerase cannot replicate long single stranded DNA template in the absence of SSB. Therefore, the effect of *the T. kodakarensis* SSB - referred to as RPA in Archaea and Eukarya - on *T.*

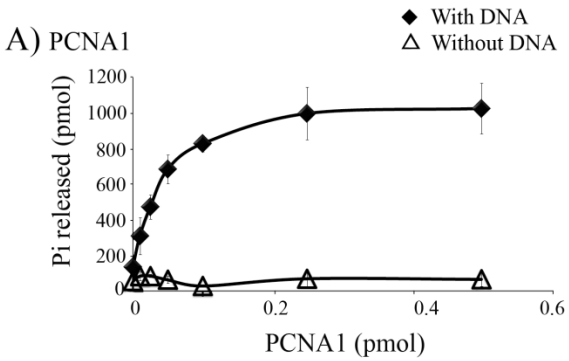
*kodakarensis* PolB activity was evaluated. As shown in Figure 2 RPA addition inhibits rather than stimulates the activity of PolB in the presence of either PCNA1 (compare Fig. 3-4 lane 14 to lane 9) or PCNA2 (compare lane 13 to lane 6). This observation is similar to those made with other archaeal members of the PolB family in which RPA inhibit polymerase activity (54), and suggests that PCNA1 nor PCNA2 interaction with PolB are insufficient to overcome such inhibition.

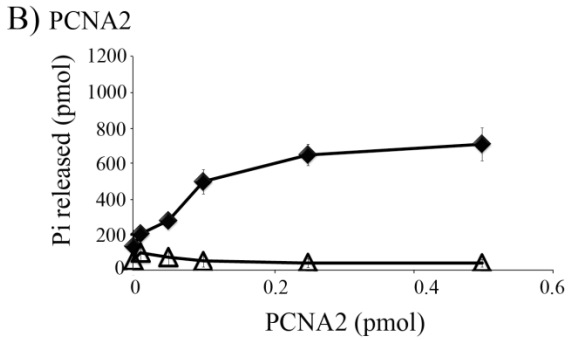
### 3.4.3 Both PCNA1 and 2 stimulate the ATPase activity of RFC

It is well established that the ATPase activity of the archaeal and eukaryotic RFC is stimulated in the presence of PCNA and primed substrate DNA (105,106). However, it remained unclear if the two *T. kodakarensis* PCNAs would behave similarly in this regard, and thus the effect of each PCNA complex on RFC ATPase activity were evaluated. In the presence of a primed DNA substrate both PCNA1 (Fig. 3-5 A) and PCNA2 (Fig. 3-5 B) stimulate ATP hydrolysis by RFC in a concentration dependent manner. However, the rate of ATP hydrolysis is ~ 3 fold greater in the presence of PCNA1 than PCNA 2 (Fig. 3-5 C). This is consistent with the observation that PCNA1 stimulates PolB activity slightly better than PCNA2 [Fig. 3-4, (83)].

### 3.4.4 Only PCNA1 is essential for *T. kodakarensis* viability

Both PCNA proteins stimulate PolB activity (Fig. 3-4), RFC ATPase activity (Fig. 3-5) and Fen1 activity (Li *et al* unpublished data). Thus, it is




C)

| | Rate of ATP hydrolysis (pmol Pi released/pmol enzyme/min) | |
|---|---|---|
| | Without DNA | With DNA |
| RFC | 2.69 | 4.81 |
| RFC+PCNA1 | 3.06 | 28.42 |
| RFC+PCNA2 | 2.82 | 9.32 |

Figure 3-5. Both PCNA1 and PCNA2 stimulate ATP hydrolysis by RFC. A and B, ATPase assays were carried as described under "Experimental Procedures" in reaction mixtures that contained increasing amounts of (A) PCNA1 or (B) PCNA2 in the presence or absence of the primed DNA substrate. The rates of ATP hydrolysis by RFC were determined as described under "Experimental Procedures" in the presence of PCNA1 or PCNA2 and primed DNA substrate. The average results of three experiments are shown in panel C.

possible that they are redundant and *T. kodakarensis* cells could survive with only one protein. To address this possibility we attempted to delete the genes encoding PCNA1 (TK0535 and PCNA2 (TK0582). *T. kodakarensis* TS033 (ΔTK0582) was constructed without difficulty generating a strain lacking PCNA2. Figure 3-6 A illustrates the strategy used to confirm the genome organizations in *T. kodakarensis* TS033 and examples of the diagnostic PCR results are shown in Figure 3-6 B. Despite repeated attempts, we were unable to generate a strain with TK0535 (PCNA1) deleted suggesting that PCNA1 is essential for *T. kodakarensis* viability. As *T. kodakarensis* TS033 exhibits no detectable growth defects, the presence of PCNA1 is apparently sufficient for genome replication and, as in most euryarchaea, the *T. kodakarensis* replisome can function with one PCNA homologue.

### 3.4.5 PCNA1 is abundant in *T. kodakarensis* cells

Although PCNA1 is essential for cell viability (Fig. 3-6) it is not a stable trimer at low protein concentrations (Figs. 3-2 and 3-3 A and B). Therefore, the *in vivo* expression levels of PCNA1 and 2 were determined using quantitative Western analysis (Fig. 3-7). It was found that while PCNA1 is expressed and accumulates to high levels (5,000-10,000 molecules per cell), the cells contain less that 60 molecules of PCNA2 (the capacities of the antibodies limits further definition). The high concentration of PCNA1 may explain how the relatively unstable trimer functions as the replicative PCNA *in vivo*.

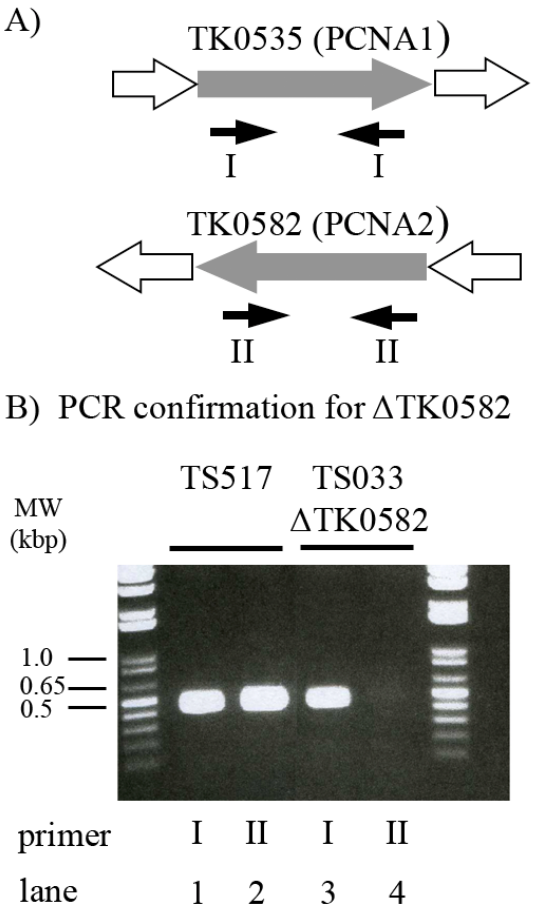



Figure 3-6. Genome organizations and PCR confirmation of the *T. kodakarensis* ΔTK0582 deletion.

(A) Genome organizations surrounding TK0535 (PCNA1) and TK0582 (PCNA2). The positions at which the PCR primers (Roman numeral primer pairs I and II) hybridized are shown.

(B) Agarose gel electrophoretic separation of PCR amplicons from genomic DNA of *T. kodakarensis* TS517 (parental strain, lane 1 and 2), TS033 (ΔTK0582 strain, lane 3 and 4) with the positions of DNA size standards indicated. As shown in lane 1 and 3, primers internal to TK0535 (primer pair I, panel A) amplified a ~600 bp molecule from *T. kodakarensis* TS517 and TS033. Primers specific to TK0582

(primer pair II, panel A) amplified an ~600 bp amplicon from *T. kodakarensis* TS517, but failed to generate an amplicon from *T. kodakarensis* TS033 genomic DNAs (compare lane 2 and 4). The data was generated by Dr. Thomas Santangelo, Ohio State University.

It is currently unknown if *T. kodakarensis* requires a high number of PCNA molecules per cell for its general physiology or if PCNA is produced at high levels to promote formation of active trimers. Although no information is available on PCNA concentrations in other archaeal species, one can speculate that *T. kodakarensis* does not require grossly greater number of clamps than *E. coli*; both have similarly sized circular chromosomes and both are likely to replicate their chromosome from a single origin of replication. The β-subunit is the bacterial functional homologue of PCNA and retains a similar overall structure and shared biochemical properties (99,107). The main difference is that the β-subunit is a dimer while PCNA is a trimer (107). *E. coli* contains about 600 copies of the β-subunit per cell resulting in 300 active molecules (dimers). Thus, *T. kodakarensis* needs about 900 molecules of PCNA to achieve a similar number of trimeric rings. Assuming the same numbers of rings are needed for both organisms replisome, the results presented in Fig. 3-7 suggest that the high number of PCNA molecule per cell may not be required for function but rather to make the concentration sufficiently high to form trimmers, or that PCNA1 has more diverse roles in *T. kodakarensis* than currently known.

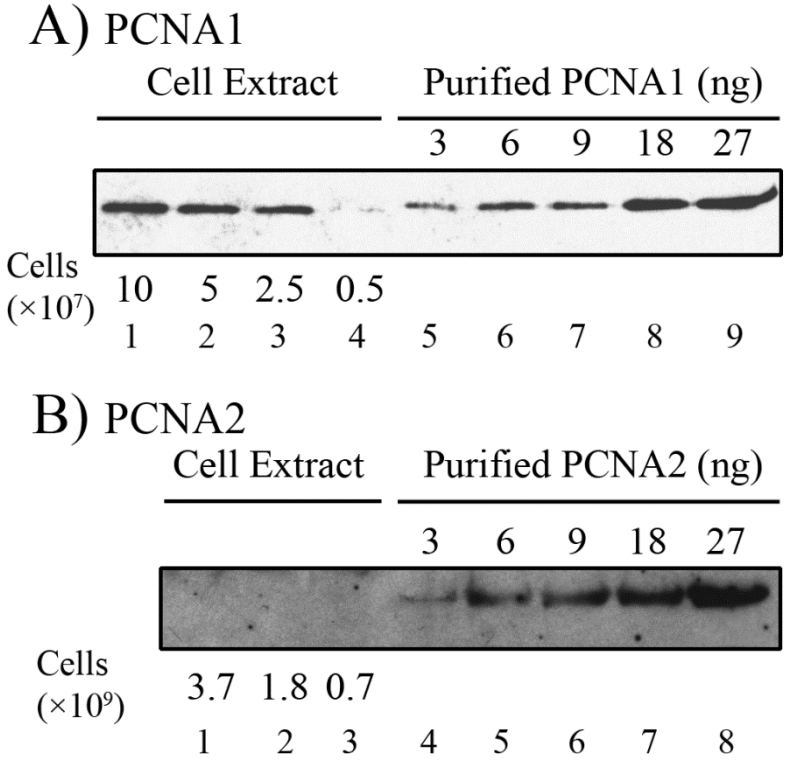



Figure 3-7. PCNA1 is more abundant than PCNA2 in *T. kodakarensis* cells. Quantitative Western was performed as describe in "Experimental Procedures" with PCNA1 (A) or PCNA2 (B).

## 3.5 Discussion

*T. kodakarensis* is the only known euryarchaeote that encodes multiple PCNA homologues, however, the data presented here suggest that, like all other Euryarchaea, *T. kodakarensis* requires only one PCNA homologue for viability. Despite similar biochemical activities, and thus the potential for redundant roles *in vivo*, we demonstrate that the gene encoding PCNA2 is not essential and can be readily deleted from the chromosome without an obvious phenotype. We could not, however, delete the gene encoding PCNA1. Although the inability to delete a

gene is not, by itself, a proof of its essentiality, there are a number of supporting observations suggesting that PCNA1 is the replicative PCNA in *T. kodakarensis*.

Three additional lines of evidence support the notion that PCNA1 is the PCNA normally involved in replication and replisome activity. First, in many archaeal species the gene encoding PCNA is in an operon with a gene encoding a subunit of GINS, another replication factor (45,95). The gene encoding PCNA1 (TK0535) is in an operon with the gene encoding GINS15 protein (TK0536) while PCNA2 is not. Furthermore, the gene encoding PCNA2 is located in a part of the chromosome suggested to be of a viral origin (12) suggesting that PCNA2 protein is of viral origin. A second line of evidence comes from the atomic structures of PCNA1 and PCNA2. Although the overall structures of PCNA1 and 2 are very similar the dimer interfaces are not (83). The trimer interfaces of PCNA1 are more similar to those found in other archaeal PCNA proteins in comparison to PCNA2 (83). Third, the expression of PCNA1 dominates *in vivo*, whereas PCNA2 expression is minimal. It is unlikely, based on the *E. coli* model and β-subunit concentrations in *E. coli*, that PCNA2 expression would be sufficient to support replication in the absence of PCNA1.

Taken together our data strongly suggest that PCNA1 is the functional replication protein in *T. kodakarensis* cells while PCNA2 is likely from a viral origin. This scenario is reminiscent of another replication protein found in multiple copies in *T. kodakarensis* (Chapter 2). The minichromosome maintenance (MCM) is the replicative helicase responsible for separating the duplex DNA at the replication fork. While most archaeal species contain a single

MCM homologue, *T. kodakarensis* contains three. It was shown, however, that only one is essential for viability (30) while the other two are likely have a viral origin.

# Chapter 4 The archaeal PCNA proteins

## 4.1 Abstract

PCNA is a ring-shaped protein that encircles duplex DNA and plays an essential role in many DNA metabolic processes. The PCNA protein interacts with a large number of cellular factors and modulates their enzymatic activities. Here the structures, functions and interactions of the archaeal PCNA proteins are summarized.

## 4.2 Introduction

PCNA was originally identified in humans as an antigen in patients with systemic lupus erythematosis and as a protein that was synthesized during the S phase of the cell cycle. The presence of the protein in the nucleus of dividing cells led to the name PCNA. Subsequent studies demonstrated that genes that encode for PCNA are present in all eukarya and that the proteins are essential for cell viability. It was also found that the protein plays diverse roles in many aspects of DNA metabolic processes by interacting and modulating the activities of a large number of enzymes (108,109). When genome sequences became available, it was found that all archaeal species sequenced contain at least one gene encoding a homologue of the eukaryotic PCNA.

All PCNA proteins studied were shown to form ring-shaped structures that encircle duplex DNA and slide bi-directionally along it. This ability to slide on DNA leads to the term "DNA sliding clamp" commonly used to describe the

protein. The first described, and to date best understood, function for PCNA is its role in chromosomal DNA replication as the processivity factor for the replicative DNA polymerases. Following primer synthesis by primase, PCNA is assembled around the primer and upon binding to the polymerase tethers it to the template, resulting in processive DNA synthesis.

To date, all activities described for the PCNA proteins require them to encircle the duplex; no biochemical function for PCNA off DNA has been reported. However, the PCNA proteins form stable rings in solution that cannot assemble independently around the duplex. RFC complex was shown to function as the clamp loader in archaea and eukarya. RFC is capable of assembling PCNA around the DNA [reviewed in: (110,111)]. Following primer synthesis by primase, RFC recognizes the primer terminus and uses the energy from ATP hydrolysis to open the PCNA ring and assemble it around the primer. Although the mechanism of clamp opening by RFC is not yet understood, molecular dynamics simulation suggested a lateral opening and right handed spiral of the ring which matches the right handed spiral of RFC resulting in clamp opening (112). However, as PCNA participates in other cellular functions besides replication (see below), it is possible that other proteins or complexes, not yet identified, are also capable of loading PCNA onto DNA. The number of Okazaki fragments synthesized during the replication of the chromosome is larger than the number of PCNA molecules within the cell. Therefore, PCNA has to be unloaded from the DNA so that it can be used on subsequent Okazaki fragments. It was shown that the eukaryotic RFC could also unload the PCNA from duplex DNA (113). It is likely that the archaeal

RFC has a similar function.Here a summary of the current knowledge on the archaeal PCNA will be described. For studies on the eukaryotic PCNA the reader is referred to several reviews on the subject (114).

## 4.3 The archaeal PCNA proteins

All sequenced archaeal genomes contain at least one homologue of the eukaryotic PCNA protein. However, the number of homologues differs between members of the different archaeal kingdoms. While the genomes of organisms belonging to the euryarchaeota and thaumarchaeota kingdoms contain a single homologue of PCNA (115-117), those belonging to the crenarchaeota branch contain three homologues each (118,119). While the single PCNA in euryarchaeota forms a homotrimer similar to the eukaryotic PCNA, the crenarchaeota proteins are active as heterotrimers (118,119). To date, the genomes of only two archaeal species contain two genes encoding for PCNA. One is the crenarchaeota *P. aerophilum* while the other is the euryarchaeota *T. kodakarensis*. It is possible that each participates in a subset of PCNA activities within the cell. *In vitro* studies with the two proteins from *T. kodakarensis* show that both form homotrimers and can stimulate DNA polymerase activity in the presence of RFC (83). It was suggested that one of the *T. kodakarensis* genes was acquired via a recent lateral gene transfer (12).

## 4.4 PCNA structure

In all organisms studied PCNA forms ring-shape trimers. To date, the

three-dimensional structures of archaeal PCNA proteins have been determined for *P. furiosus* (Fig. 4-1 A) (120), *A. fulgidus* (Fig. 4-1 B) (121), *H. volcanii* (Fig. 4-1 C) (122,123), the two PCNA homologues in *T. kodakarensis* (Fig. 4-1 D) (83), and *S. solfataricus* (Fig. 4-1 E) (124). While PCNA from euryarchaeota form homotrimers, those from crenarchaeota form heterotrimers. Nevertheless, the overall structures of the rings are very similar to each other as well as to the structure of the eukaryotic PCNA (125). Each monomer is composed of two structurally similar domains, although they share very little amino acid sequence similarity. The two domains are connected by a long loop that was shown to be the major interaction with PIP-box containing proteins (discussed below).

To form the trimeric rings three monomers interact in a head-to-tail manner resulting in a pseudo six-fold symmetry for the trimeric complex. The interface includes antiparallel β-strands forming an extended β-sheet, a core of hydrophobic residues and ionic interactions. The interfaces between PCNA proteins from different organisms vary mainly in the extent of hydrogen bonding and ion pairing and the length of the β-strand interaction.

The inner diameter of the PCNA rings from all organisms studied is about 35 Å; this is sufficiently large to accommodate duplex DNA. The ring has two different faces. One interacts with proteins that associate with PCNA while, to date, the other face has not been implicated in protein-protein interactions. Although no structure of a thaumarchaeota PCNA has been determined, it is likely to possess a structure similar to those of the euryarchaeota proteins.

All PCNA proteins identified to date are acidic with low pI (114). The

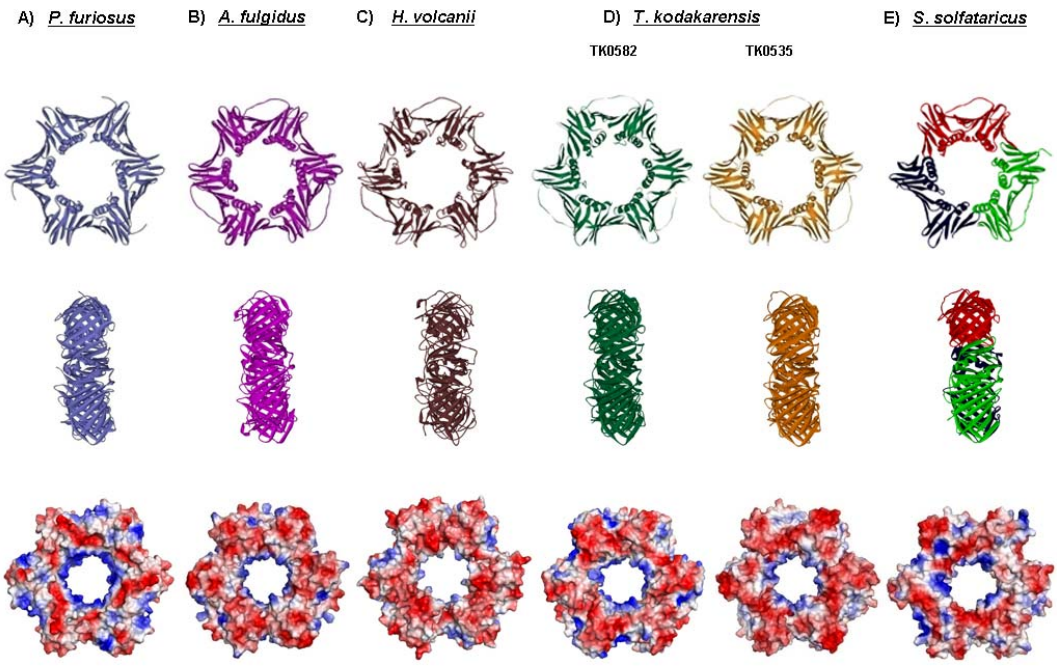



Figure 4-1. Three dimensional structures of archaeal PCNA proteins.

Top and bottom, front view; middle, side view. Top and middle images are ribbon representations. The bottom panel shows the charge distribution. Negatively charged residues are shown in red; positively charged residues are blue. Panel A, *P. furiosus*; Panel B, *A. fulgidus*; Panel C, *H. volcanii*; Panels D, *T. kodakarensis*; Panel E, *S. solfataricus*.

charge distribution on the ring surface, however, is not symmetric. While the outer surface is highly negatively charged, there is a net positive electrostatic potential in the central cavity (Fig. 4-1). It was proposed that the negatively charged surface might prevent non-specific interactions with DNA. Using the eukaryotic PCNA it was shown that the positively charged residues in the central

cavity are required for PCNA interactions with DNA (126). Furthermore, mutating four conserved Arg and Lys residues in the central cavity resulted in reduced DNA binding (126).

In contrast to the PCNA proteins from the thermophilic archaeons *P. furiosus*, *A. fulgidus*, *S. solfataricus*, *T. kodakarensis*, and the eukaryotic PCNA, the protein from the halophilic archaeon *H. volcanii* has very few positively charged residues within the central cavity. Interestingly, out of the four positively charged residues in the eukaryotic PCNA required for DNA binding (126) only one is present in the *H. volcanii* protein (122,123). These observations may suggest a different mechanism of interaction with DNA by the halophilic PCNAs. It was suggested that the increased number of negatively charged residues in the central cavity may serve as cation-binding sites and thus DNA binding may be mediated by cation interactions with the phosphate backbone of the DNA (122).

As mentioned above, *T. kodakarensis* contains two PCNA homologues, PCNA1 and PCNA2 encoded by TK0535 and TK0582 respectively. It was suggested that one of the genes encoding for PCNA was acquired via lateral gene transfer and is not a result of gene duplication (Chapter 3). The three-dimensional structure supports this hypothesis (83). While the antiparallel β-strand interactions found in the interface and the core hydrophobic interactions are conserved between PCNA1 and PCNA2, the side chains participating in hydrogen bonding and ionic interactions are not (83).

## 4.5 PCNA functions

The PCNA protein has no known enzymatic activity. The only known function is its ability to move along duplex DNA. The protein, however, interacts with a large number of cellular factors and regulates their activities [summarized in (102,108,109)]. It was proposed that the ability of PCNA to slide along duplex DNA might function as a moving platform for enzymes that participate in DNA metabolic processes but have low sequence specificity (127).

Many of the proteins that interact with PCNA do so, at least in part, via a PCNA-interacting protein (PIP) motif (128,129). PIP- motif is 8 amino acid sequence defined as Qxxhxxaa where "h" is a moderately hydrophobic amino acid (I, L or M) and "a" is an aromatic residue followed by a non-conserved sequence containing basic amino acids. The PIP-motif interacts with the loop that connects the two domains in each PCNA monomer. However, not all PCNA interacting proteins contain a PIP-motif and it was found that other, less common PCNA binding motifs also present in some proteins. Furthermore, some proteins contain multiple PIP-boxes or several motifs needed for PCNA interactions.

The archaeal PCNA interacts with a number of enzymes needed for lagging strand replication and Okazaki fragment maturation. PCNA was shown to interact and stimulate the activity of the replicative DNA polymerases of euryarchaeota PolB and PolD, and crenarchaeota, PolB. When the synthesis of an Okazaki fragment is completed the PCNA clamps left on DNA interact and regulate the activity of several enzymes needed for Okazaki fragment maturation. These interactions include the binding to Fen1 and RNase H, which participate in

removing the RNA primers, and DNA ligase which seals the nick between adjacent Okazaki fragments (Table 4-1).

As mentioned above, crenarchaeal PCNA is a heterotrimer. Studies with the heterotrimeric PCNA from *S. solfataricus* show that different subunits interact withPolB, Fen1 and DNA ligase (118). This observation may suggest that during chromosomal replication there is a preformed complex that contains all the components needed for lagging strand replication and Okazaki fragment maturation. In euryarchaeota PCNA is a homotrimer and it is not yet clear if PCNA can simultaneously bind to different enzymes. It is thus possible that euryarchaea and crenarchaea utilize PCNA differently.

In addition to their roles in chromosomal replication, the archaeal PCNA proteins also participate in DNA repair and recombination. The role of PCNA in these processes is less well understood. PCNA binds and regulates the activity of several repair and recombination enzyme enzymes (Table 4-1). PCNA was shown to interact and regulate the activity of the archaeal members of the translesion DNA polymerases (PolY), which are involved in DNA repair by replicating through damaged DNA. PCNA also interacts with uracil-DNA glycosylase (UDG), which eliminates uracil from DNA damaged by cytosine deamination and with apurinic/apyrimidinic (AP) endonuclease to generate a gap for efficient DNA repair synthesis. PCNA was also shown to interact with XPF (xeroderma pigmentosum complementation group F), Hjc (Holliday junction endonuclease) and Hjm (Holliday junction migration helicase) proteins that participate in the resolution of Holiday junctions either during recombination mediated repair

Table 4-1. PCNA interacting proteins

| Protein | Effect | Organisms in which interactions been demonstrated |
|---|---|---|
| PolB | Increase polymerase processivity | *A. pernix; A. fulgidus; M. thermoautotrophicus; M. acetivorans P. abyssi; P. furiosus; S. solfataricus* |
| PolD | Increase polymerase processivity | *A. fulgidus; P. abyssi; P. furiosus; P. horikoshii* |
| PolY | Increase polymerase processivity | *M. acetivorans; S. solfataricus* |
| RFC | Loading PCNA onto DNA | *A. pernix; A. fulgidus; M. thermoautotrophicus; P. abyssi; P. furiosus; S. solfataricus* |
| DNA ligase | Stimulate ligase activity | *A. pernix; P. furiosus; S. solfataricus* |
| Fen1 | Stimulate nuclease activity | *A. pernix; P. aerophilum; S. solfataricus* |
| RNase HII | | *A. fulgidus; P. abyssi* |
| Rad2 | | *A. fulgidus* |
| UDG | Stimulate activity * | *P. aerophilum; P. furiosus; S. solfataricus* |
| XPF | Stimulate nuclease activity | *S. solfataricus* |
| AP endonuclease | Stimulate nuclease activity | *P. furiosus* |
| NucS | Increase nuclease cleavage specificity | *P. abyssi* |
| Hjc | Stimulate nuclease activity | *S. solfataricus* |
| Hjm | Stimulate helicase activity | *P. furiosus* |

*Stimulatory effect on *P. furiosus* UDG but no effect was observed on *P. aerophilum* and *S. solfataricus* UDG.

and/or in DNA replication restart of stalled or broken replication forks. PCNA also interacts with other enzymes that may be involved in DNA repair [e.g. (130)].

## 4.6 Regulating PCNA activity

In eukarya, post-translational modification of the PCNA proteins regulates its stability and the binding between PCNA and its interacting enzymes. These modifications include phosphorylation, acetylation, ubiquitination and sumoylation. To date, neither ubiquitination nor sumoylation have been identified in archaea. However, a small molecule modifier of archaeal proteins has been identified (131). It is thus possible that post-translational modification of PCNA also plays a role in the function of the enzyme in archaea. It was found that PCNA in *H. volcanii* is enriched in the phosphoproteome fractions, thus suggesting that the protein may be phosphorylated *in vivo* (132). In eukarya, it was shown that phosphorylation of several enzymes that interact with PCNA modulates their interactions with PCNA [e.g. Fen1 (133)]. As protein phosphorylation is commonly found in archaea (134), a similar mechanism may also exist in this domain. Future studies, however, are needed to determine if post-translational modification participates in regulating the archaeal PCNA.

## 4.7 Discussion

To date, most of the proteins that interact with the archaeal PCNA participate in either DNA replication or repair (Table 4-1). In eukarya, on the other hand, PCNA also interacts with enzymes involved in post replication

process (e.g. methyltransferase and chromatin remodeling factors), cell cycle regulatory proteins (e.g. p21/cip1), etc. Thus, it is possible that PCNA interacts with other, not yet identified cellular proteins and modulates their activities. The recently developed genetic tools for a number of archaeal species may enable new experimental approaches to identify those proteins.

# Chapter 5 Concluding remarks

In the past decade, a lot of progress has been made on the study of the mechanism of DNA replication in archaea. However, the main drawback of the field was the lack of genetic tools. This has changed in the past few years. *T. kodakarensis* is found to be naturally competent for DNA uptake and to efficiently incorporate donor DNA into its genome. Benefiting from this, many shuttle and integrating vectors have been developed, making *T. kodakarensis* an ideal archaeal model system. In addition, this organism has several unique features regarding its replication proteins, not found in other archaeal species, which make it an interesting organism to study DNA replication.

MCM helicase is an essential enzyme in both the initiation and elongation stages of DNA replication. All archaeal and eukaryotic species contains MCM genes. In eukarya, the replicative MCM complex is a heterohexamer formed by the assembly of six different homologs (MCM2-7). The replicative helicase in archaea is also thought to contain an MCM complex, but most archaeal species possess a single MCM homolog. Prior to this work, all archaeal MCM studies were from organisms with a single MCM homologue. However, there are several archaeal species, including *T. kodakarensis*, containing multiple MCM homologs (Chapter 1 and 2), which may infer that these organisms may be the evolution intermediates between archaea and eukarya. Therefore, it would be important to find whether these paralogues/orthologues form hetero multimeric complexes, like their eukaryotic counterparts; or they can be functional individually. It would

be also interesting to see if all of the MCMs in the cell are essential for the cell viability. To answer these questions, several experiments were performed (Chapter 2). It was found that all three MCMs contain all essential motifs required for archaeal MCM helicase activity reported previously. Based on the biochemical experiments, all three MCMs exhibited ATP hydrolysis ability, DNA binding ability and duplex DNA unwinding activity but with different levels. However, the three MCMs have different oligomeric states: only MCM3 forms a stable hexamer in solution, which is similar to other reported archaeal and eukayotic MCMs. This result may explain why MCM1 is such a poor helicase, since it is not a stable hexamer in solution. Future work will be needed to determine how the helicase activity of MCM2 is influenced by its different oligomeric states. To determine whether all three MCMs are required for cell growth, we tried to knock out each gene. The result indicated that the genes encoding for MCM1 and MCM2 can be knocked out individually and simultaneously. The fact that we could not knockout MCM3 gene suggests it is essential for *T. kodakarensis*. Collectively, the biochemical and genetic results indicated that only MCM3 is involved in DNA replication in *T. kodakarensis*.

PCNA is the processivity factor which can tether the DNA polymerase onto the template to prevent it from falling off the DNA. Both archaeal and eukaryotic PCNA are composed of three subunits to form trimers (Chapter 3). To date, the structures of six archaeal PCNA have been elucidated, including *T. kodakarensis* (Chapter 3 and 4). Uniquely, *T. kodakarensis* is the only euryarchaea containing two PCNA homologs. The two PCNAs from *T.*

*kodakarensis* share similar overall structure to each other as well as to other previously characterized archaeal PCNA structures (Chapter 1 and 3). However, it is also found that the PCNAs have different trimer stability. Further tests were performed to check the difference in biochemical properties of the two PCNAs. It is shown that both PCNAs can stimulate RFC ATPase activity and both can stimulate PolB activity. These biochemical assays confirmed that both PCNAs are functional *in vitro*. Next, to detect if both PCNAs are needed for cell growth *in vivo*, knock out for each PCNA gene was attempted. Only the gene encoding PCNA2 can be deleted, but not PCNA1, suggesting that PCNA1 is essential for cell survival and probably involved in DNA replication.

Another interesting finding is that the three deletable genes encoding for MCM1, MCM2 and PCNA2 are located in three different putative viral integration regions (TKV1, TKV4 and TKV3, respectively). This analysis suggests that these three potential DNA replication factors may originate from viral sources. They may be acquired through lateral gene transfer, but not via gene duplication into *T. kodakarensis* genome. In fact, it is rare to find that replication factors are located in viral integration regions. To support it, there are totally 151 open reading frames (ORFs) in *T. kodakarensis* viral integration regions. More than 95% ORFs are annotated as hypothetical proteins, without any evident functional homolog. Based on our study, MCM1, MCM2 and PCNA2 are not essential for cell growth. However, all of them are expressed *in vivo* but with different level. It is possible that although these proteins are not directly involved in DNA replication, they may participate in other DNA

metabolism pathways, like DNA repair, recombination or other cellular process. To further determine if the knockout strains for MCM or PCNA have abnormal cell growth due to DNA replication and repair, *T. kodakarensis* cell will need to be subject to UV radiation and other DNA damage. Furthermore, *T. kodakarensis* can grow under a wide range of condition (temperature from 60-100°C, pH from 5-9). Therefore, it is also possible that the cell may utilize different MCMs or PCNAs to adapt its changing living environment. In the future, the thermostablity and pH/salt sensitivity of each MCM and PCNA can be analyzed to approve this hypothesis. Interestingly, both MCM1 and2 have unique N-terminal extensions, through which the proteins may interact with other protein factors *in vivo* to regulate the MCM helicase activity. In fact, i*n vivo* pull down assay also suggest that MCM2 may form a complex with MutS homolog, a DNA repair enzyme (Chapter 2). Alternatively, the N-terminal extensions of MCM1 and 2 may also be responsible for other enzyme functions required for archaeal viral replication and infection regulation. To support it, it was found that in polyomavirus SV40, the large tumor antigen contains an origin DNA binding domain in addition to its helicase domain (Chapter 2). In bacteriophage T7, gene 4 encodes a polypeptide composed of both primase and helicase domain. Another possibility is that the host cell may add this N-terminal extension sequence to modify the viral replication protein and control its activity. In other words, these N-terminal extensions may reflect the interaction between the host and virus. Future experiment can be designed to characterize the function of the N-terminal elongation region. This study also raised another possibility that the whole viral

integration regions in *T. kodakarensis* (TKV1-TKV4) are not essential for the cell growth. These regions may be formed and developed during the evolution process as an "immune system" to defend archaeal viral infection.

In addition to *T. kodakarensis*, there are several other archaeal species that have been sequenced and shown to contain multiple MCM homologs. Based on our biochemical and genetic results, *T. kodakarensis* requires only one MCM and one PCNA for its DNA replication, like most other archaeal species. However, it is still not clear how multiple MCMs and PCNAs formed in *T. kodakarensis* and other archaeal organisms. Bioinformatics study suggest that MCM1, MCM2 and MCM3 are grouped in the same cluster in archaeal MCM phylogenetic tree, which indicated these three MCMs share the same origin (78). Given that MCM2 and MCM3 are located in putative viral integration regions, it is possible that MCM1 and 2 is generated from MCM3 via viral infections (Fig. 5-1). It is understood that virus could utilize host replication machinery to amplify new virus. During this process, archaeal virus may "hi-jack" the endogenous MCM3 for its replication and mutated to new MCMs due to low viral replication fidelity. Then the virus containing the newly mutated MCM (MCM1 or 2) could invade the archaeal cell again. At this point, *T. kodakarensis* cell uptake the foreign DNA material to integrate to it genome to form the multiple MCM. This model may also be applicable to the evolution of multiple PCNAs in *T. kodakarensis*. Nevertheless, more information and study will be needed for archaeal viruses and archaeal evolution to confirm this speculation. This study will shed light on the evolution of multiple MCMs in other archaeal organisms. Future work will need

to address the relationship between these viral proteins and DNA repair and recombination. *Thermococcus* and *Pyrococcus* are the largest orders among euryarchaea. Interestingly, *T. kodakarensis* is the only organism has been found to have viral integration region in both *Thermococcus* and *Pyrococcus*. It will be important to understand the reason *T. kodakarensis* is more easily to be infected by virus and acquire foreign materials.

DNA replication is the basis for the propagation and evolution of living organisms. The process needs to be precisely controlled and regulated. This work focuses on the study of two essential DNA replication proteins, the MCM helicase and the processivity factor PCNA. Using biochemical, structural and genetic approaches we studied the three MCM and two PCNA homologs in *T. kodakarensis*. We showed that although *T. kodakarensis* contains multiple homologues of each protein, with similar structures and biochemical properties, only one of each is essential for cell viability. These observations make *T. kodakarensis* similar to most other archaeal species studied which contain a single MCM and a single PCNA homologues.

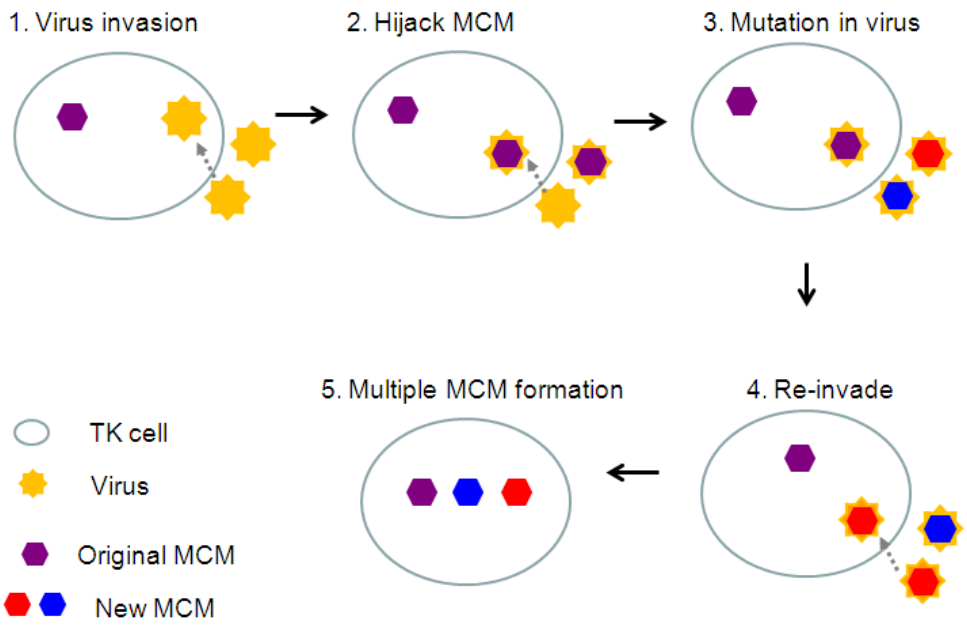



Figure 5-1. Putative model for *T. kodakarensis* multiple MCM evolution.

Dashed arrow stands for the viral invasion process. Hexagons represent MCMs; oval represents *T. kodakarensis* cell; yellow star represents archaeal virus which can infect the *T. kodakarensis* cells; black arrow represents evolution process; dashed arrow stands for archaeal virus infection process. In this process, *T. kodakarensis* is originated from single MCM containing cell to multiple MCM harboring cell due to integrated viral genetical materials into its own genome.

# Appendix

## Oligonucleotides used for this study

| Primer | Sequence |
|---|---|
| MD001 | 5’-GCGGCGAGTCCAGCTCAGGAGCTCGCGCCG |
| MD002 | 5'-TTTGTTTGTTTGTTTGTTTGTTTGTTTGTTTGTTTGCGGCGCGAGCTCCTGAGCTGGACTCGCCGC |
| MD003 | 5'-CGGCGCGAGCTCCTGAGCTGGACTCGCCGC |
| MD004 | 5'-GCGGCGAGTCCAGCTCAGGAGCTCGCGCCGGTTTGTTTGTTTTGTTTGTTTGTTTGTTTGTTTGTTT |
| MD005 | 5'-CGGCGCGAGCTCCTGAGCTGGACTCGCCGC |
| MD006 | 5'-CCGACGTGCCAGGCCGACGCGTCCC |
| MD007 | 5'-TTTGTTTGTTTGTTTGTTTGTTTGCCGACGTGCCAGGCCGACGCGTCCC |
| MD008 | 5’-GGGACGCGTCGGCCTGGCACGTCGGTTTGTTTGTTTGTTTGTTTGTTTG |
| MD009 | 5’-TTTGTTTGTTTGTTTGTTTGTTTG CAGTGCCAAGCTTGCATGCCTGCAG |
| MD010 | 5’-CAGTGCCAAGCTTGCATGCCTGCAGTTTGTTTGTTTGTTTGTTTGTTTG |
| MD011 | 5'-CGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTGTTTGTTTGTTTGTTTGTTTG |
| MD012 | 5'-TTTGTTTGTTTGTTTGTTTGTTTGCGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCG |
| MD013 | 5'-CGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCGTTTGTTTGTTTGTTTGTTTGTTTG |
| MD014 | 5’-TTTGTTTGTTTGTTTGTTTGTTTGCGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCGCGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCGCGACGGCCAGTGCCAAGCTTGCAT |

| | |
|---|---|
| MD015 | 5'-ATGCAAGCTTGGCACTGGCCGTCGCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTGTTTGTTTGTTTGTTTGTTTG |
| MD016 | 5'-CGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCG |
| MD017 | 5'-TTTGTTTGTTTGTTTGTTTGTTTGCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTGTTTGTTTGTTTGTTTGTTTG |
| MD018 | 5'-TTTGTTTGTTTGTTTGTTTGTTTGCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTGTTTGTTTGTTTGTTTGTTTG |
| Miao001 | 5'-ccgCATATGGAAGTTCTGACAAAGGACGAG |
| Miao002 | 5'-ccgGGATCCTCAGTGATGGTGATGGTGATGCTCAAGGCCGTACTCCTTCAG |
| Miao003 | 5'-ccgCATATGAAGAAGCGCCTACCAGCGAGC |
| Miao004 | 5’-ccgGGATCCTCAGTGATGGTGATGGTGATGGAGCGGCTCGTAGAAGCCTATC |
| Miao005 | 5’-ccgCATATGGAGGAAACGAGGTTTAGAAGG |
| Miao006 | 5’-ccgGGATCCTCAGTGATGGTGATGGTGATGTTCCTCACCACCTCCGAAGAACTCG |
| Miao007 | 5’-ccgCATATGGACGATGACTACCTCAGTTC |
| Miao008 | 5’-ccgGGATCCTCAGTGATGGTGATGGTGATGGAACAACCTCCTCTGGCGAGAGAGGC |
| Miao009 | 5’-ccgCATATGTTAACAAAAGTTACAGACGTTACAG |
| Miao010 | 5’-ccgGGATCCTCAGTGATGGTGATGGTGATGCCCCACCAACTTCAATCTATTAG |
| Miao011 | 5’-ccgCATATGAGTGAGGACATCAATTTTGACTTCG |
| Miao012 | 5’-ccgGAATTCTCAGTGATGGTGATGGTGATGCCTCCTTCTCAAGCGGTAAAAG |
| Miao013 | 5’-ccgCATATGGACAGGGAAGAGATGATAGC |

| | |
|---|---|
| | 5’- |
| Miao014 | ccgGGATCCTCAGTGATGGTGATGGTGATGGATGACGCGGTA GAATCCAGC |
| Miao015 | 5’-ccgCATATGACTTTTGAGATTGTGTTTGATTC |
| | 5’- |
| Miao016 | ccgGTCGACTCAGTGATGGTGATGGTGATGTGAGCGACCCT CCTCGACTC |
| Miao017 | 5’-ccgCATATGCCGTTCGAAGTTGTTTTTGACG |
| | 5’- |
| Miao018 | ccgGGATCCTCAGTGATGGTGATGGTGATGCTCCTCAACGCG CGGAGCG |
| Miao019 | 5’-ccgCATATGCTCGACCCCTTTGGAAAAAGAGC |
| | 5’- |
| Miao020 | ccgGGATCCTTAGTGATGGTGATGGTGATGCTCACTGGTCTC ACCTTCCTTC |
| Miao021 | 5’-ccgCCATGGTGAGTAAGCTGTTGAGGGAAGTAAC |
| | 5’- |
| Miao022 | ccgGGATCCTCAGTGATGGTGATGGTGATGTCCCTCATGTAG CTCCTTCCATTC |
| Miao023 | 5’-ccgCATATGTCCGAGGAAGTGAAGGAAG |
| | 5’- |
| Miao024 | ccgGGATCCTCAGTGATGGTGATGGTGATGCTTACCCATAAT CGTGAACTGG |
| Miao025 | 5’-ccgCATATGACGGAAGTCCCATGGGTTG |
| | 5’- |
| Miao026 | ccgGGATCCTCAGTGATGGTGATGGTGATGCTTCTTGAGGAA GTCGAACAG |
| Miao027 | 5’-ccgCATATGATCCTCGACACTGACTACATAAC |
| | 5’- |
| Miao028 | ccgGGATCCTCAGTGATGGTGATGGTGATGAGTTCCCTTCGG CTTCAGCC |
| Miao029 | 5’-ccgCATATGGATATAGTGAAGCTCAGGGAAC |
| | 5’- |
| Miao030 | ccgGGATCCTTAGTGATGGTGATGGTGATGGAGGAATATCCT TACTCTTCGTGC |
| Miao031 | 5’-ccgCATATGTTCACGGGTAAAGCCCTC |
| | 5’ |
| Miao032 | ccgGGATCCTCAGTGATGGTGATGGTGATGGGCATCACCGAG CCACTCGTTTC |
| Miao033 | 5’-ccgCATATGCTGGTGGAGGACCTCCTTAAGAAC |

| | |
|---|---|
| Miao034 | 5’-ccgGGATCCTCAGTGATGGTGATGGTGATGAACCCCCTCGCAAAACTGGTCAAAC |
| Miao035 | 5’-ccgCATATGAGCGAGGAGATTTACTCGCC |
| Miao036 | 5’-ccgGAATTCTTAGTGATGGTGATGGTGATGCGAGCCGAAGAACTCGTCGAGGC |
| Miao037 | 5'-ccgCATATGATCCTCGACACTGACTACATAACCGAGGATGGAAAGCC |
| Miao038 | 5'-ccgGGATCCTCAGTGATGGTGATGGTGATGAGTTCCCTTCGGCTTCAGCCAAGCACTCAAACC |
| Miao039 | 5'-ccgCATATGAGCGAGGAGATTTACTCGCCTGAAATGAAGGC |
| Miao049 | 5'-ATCTTCTCCAGCTCATCAAT |
| Miao050 | 5'-GAAGTCAAAGAAAGCCCTGACC |
| Miao051 | 5'-CAGGCTTCCCGTTCTTCTTCC |
| Miao052 | 5'-GAAACGGCTGTAAAACGGC |
| Miao053 | 5'-GAATATGGGCCTCTTCTCCC |
| Miao054 | 5'-TTGGAGACCTTGACTGGGCC |
| Miao055 | 5'-CTTGATCTTCGAGACCACTCC |
| Miao056 | 5'-CAGAGTTGCGGTAATTCCGGCC |
| Miao057 | 5'-TGAATCTTCGGCTCGCTTCCATCCC |
| Miao058 | 5'-CAACGGGAAGCTTATGATTGG |
| Miao059 | 5'-GATGAATCCAATGAGTGTTTGG |
| Miao060 | 5'-GGTTATCTTTATTCTCCTCCC |
| Miao061 | 5'-GGCGAGTCTCTCCATATCGTCC |
| Miao062 | 5'-CCTTGCAGTACCAGCGCGCCC |
| Miao063 | 5'-GGTGGTTCGTCTCAACACCGGGG |
| Miao064 | 5'-CTAGCTCCTGAACCCACCACTCC |
| Miao065 | 5'-CAACCCTGGCCGTTTTTCCTTCC |
| Miao066 | 5'-TATCCTCCCAATGATTCCAGCC |
| Miao067 | 5'-TTATCGCGCTTAAATCCAAGCC |
| Miao068 | 5'-GCCTGACTTTCACCAGCCAGCC |
| Miao069 | 5'-GACGGTGAGGACTATCTTTCC |
| Miao071 | 5’-GCTAAAGCCAGTGCAGCGGTTGTCTTCCCGACGCCGGGTGGCCCTGCG |

| | |
|---|---|
| | 5’- |
| Miao072 | CGCAGGGCCACCCGGCGTCGGGAAGACAACCGCTGCACT GGCTTTAGC |
| | 5’- |
| Miao073 | GCGTCTTCGTCGCCGTCGCAGTTCCTCCTCTTCGCCGCATG GTAATAC |
| | 5’- |
| Miao074 | GTATTACCATGCGGCGAAGAGGAGGAACTGCGACGGCGAC GAAGACGC |
| | 5’- |
| Miao075 | CGCGGACGGCTGCTGCAGTTAGGCCCGCCGCCGATGAACT CTTCCCG |
| | 5’- |
| Miao076 | CGGGAAGAGTTCATCGGCGGCGGGCCTAACTGCAGCAGCC GTCCGCG |
| | 5’- |
| Miao077 | GGCCTTGGAGATGCTGATTGTCTGCTGCTCCAGCGCTTCGT GTATCG |
| | 5’- |
| Miao078 | CGATACACGAAGCGCTGGAGCAGCAGACAATCAGCATCTC CAAGGCC |
| | 5’- |
| Miao079 | GATGATTGAGGGGTACAGGGATCTAAAATCTAGGTACACTA TGTTCTCC |
| | 5’- |
| Miao080 | GGAGAACATAGTGTACCTAGATTTTAGATCCCTGTACCCCT CAATCATC |
| | 5’- |
| Miao081 | GCCGTAGTAACCGTAGTAGCTGTTTGCCAGGATCTTGATGG CCCTC |
| | 5’- |
| Miao082 | GAGGGCCATCAAGATCCTGGCAAACAGCTACTACGGTTACT ACGGC |
| Miao083 | 5’-GCCGTCGCAGTTCCTCCTCTTCGCCGCATGGTAATAC |
| Miao084 | 5’-GTATTACCATGCGGCGAAGAGGAGGAACTGCGACGGC |
| | 5’- |
| Miao085 | CGGCTGCTGCAGTTAGGCCCGCCGCCGATGAACTCTTCCC G |

| | |
|---|---|
| Miao086 | 5’-CGGGAAGAGTTCATCGGCGGCGGGCCTAACTGCAGCAGCCG |
| Miao087 | 5’-GGAGATGCTGATTGTCTGCTGCTCCAGCGCTTCGTGTATC |
| Miao088 | 5’-GATACACGAAGCGCTGGAGCAGCAGACAATCAGCATCTCC |
| Miao089 | 5’-GGGGTACAGGGATCTAAAATCTAGGTACACTATGTTCTCC |
| Miao090 | 5’-GGAGAACATAGTGTACCTAGATTTTAGATCCCTGTACCCC |
| Miao091 | 5’-GTAACCGTAGTAGCTGTTTGCCAGGATCTTGATGGCCCTC |
| Miao092 | 5-GAGGGCCATCAAGATCCTGGCAAACAGCTACTACGGTTAC |
| Miao093 | 5’-ACCTTGTGGGTCACCTCGTTATAGGTCTCGCTCCACACACCTCGGCTGGAATCATTGGGAGGATAATCGGCTTTTCAGACGTTCTCGTTGGCTACGCACACCCGTATTACCATGCGGC |
| Miao094 | 5’-CGAGGCAGGGAGGTACCTCCTCAAGGTGGCGCGCTTCATCGATGACCTCCTTGAGAAGTTCTACGGATTACCCAGGTTCTACAATGCCGAGAAGATGGAAGACCTTGTGGGTCACC |
| Miao095 | 5’-GTATTACCATGCGGCGAAGAGGAGGAACTGCGACGG |
| Miao096 | 5’-ccgGAATTCTTAGTGATGGTGATGGTGATGCGAGCCGAAGAACTCGTCGAGG |
| Miao097 | 5’-ccgGAATTCTTACGAGCCGAAGAACTCGTCGAGG |
| Miao098 | 5'-TTAACGAAAGTATTGTAACACACC |
| Miao099 | 5'-GATCGAACACACGAAAAGGAAACC |
| Miao100 | 5’-AAGAGATGATAGCACGCTTTGCC |
| Miao101 | 5’-CCCTGTCCATATGTATATCTCC |
| Miao102 | 5’-ccgCATATGGACAGGGAAGAGATGATAGC |
| Miao103 | 5’-CTAAGTGTGGGTATGAGGTTTCTCTCC |
| Miao104 | 5’-GACACACAAAAACAGCCTTCTCGATGAAGCC |
| Miao105 | 5’-ATACATATGGACAGGGAAGAGATGATAGCACG |
| Miao106 | 5’-CGTGCTATCATCTCTTCCCTGTCCATATGTAT |
| Miao107 | 5'-cctGTCGACAGCGATATATTTATATAGGGATATAGTAATAGATAATATCACAGGTGGTATGAATGAGTGAGGACATCAATTTTGACTTCG |

| | |
|---|---|
| | 5’- |
| Miao108 | CCGAAGCGGCCGCTCAGTGATGGTGATGGTGATGGTGATG GTGATGCCTCCTTCTCAAGCGGTAAAA |
| Miao109 | 5'-ccgGGATCCTCACTTACCCATAATCGTGAACTGG |
| Miao110 | 5'-5'-catgaggggaagattgaaccgcc |
| Miao111 | 5'-ccgGGATCCTTACTCACTGGTCTCACCTTCCTTC |
| | 5'- |
| Miao112 | cctGTCGACAGCGATATATTTATATAGGGATATAGTAATAGATA ATATCACAGGTGGTATGAATGGACGATGACTACCTCAGTTC |
| | 5’- |
| Miao113 | CCGAAGCGGCCGCTCAGTGATGGTGATGGTGATGGTGATG GTGATGGAACAACCTCCTCTGGCGAGAGAGGC |
| Miao114 | 5’-ccgCATATGTACCCCGAAAACTATAAAAGG |
| | 5’- |
| Miao115 | ccgGGATCCTCAGTGATGGTGATGGTGATGTTGGATTCCCTC CAGCAGG |
| Miao116 | 5’-ccgCATATGATTCCTTTTCTAATTTTCTTC |
| | 5’- |
| Miao117 | ccgGGATCCCTAGTGATGGTGATGGTGATGGTGTTCCCCGTA TTTAGG |
| Miao118 | 5'-CCATCAAGGCCTTCAACGATGCC |
| Miao119 | 5'-GGACAAGGCCATGTCGGTTGAGG |
| Miao120 | 5'-CCGTGGAGAGGGTTGCGCCTAAGGCG |
| Miao121 | 5'-TGAGACTGAGCTGTTGGGGG |
| Miao125 | 5'-ccgGGATCCTCAAACCCCCTCGCAAAACTGGTCAAACC |
| Miao126 | 5'-ATGAGTCCCTTCCAGTGGAAAACGGG |
| Miao127 | 5’-GGGCAAAGGAGATTCCCGACGAGTGGGAGC |
| Miao128 | 5’-TGTAGTGGAAGCCCTCCTTTGGGGC |
| Miao129 | 5'-ccgCATATGCTGGTGGAGGACCTCCTTAAGAAC |
| Miao130 | 5’-ccgCATATGACTTTTGAGATTGTGTTTGATTC |
| Miao131 | 5'-ccgGTCGACctgcagTCATGAGCGACCCTCCTCGACTC |
| Miao132 | 5'-ccgGGATCCctgcagTCACTCCTCAACGCGCGGAGCG |
| Miao133 | 5’-TCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGG |
| Miao134 | 5’-TATACATATGGTACTGCAAGAACGTGAAGTTCTGCAG |
| Miao135 | 5’-TATACATATGGAGGCTAAGCTCAGGGAGCTGGAGGAGG |
| Miao136 | 5’-GAGAGTTCATCGGCGGCGGGCCTAACTGCAGCAGCCG |
| Miao137 | 5’-CCCGCTCGTGTAAATAGCCCTTGGAGCCAGATTGGCG |
| Miao138 | 5’-CCGCATATGAAGATAGCGGGCATTGACGAGG |

| | |
|---|---|
| Miao139 | 5’-CCGGGATCCTCAGTGATGGTGATGGTGATGGACCTTCCGAAAGTAGCGG |
| Miao140 | 5’-TGAGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCC |
| Miao141 | 5’-GCGACGACGCAGACGATAAAAACCCGGTTTGCTCTGTTCAACC |
| Miao142 | 5’-CCCCACCAACTTCAATCTATTAGCCTCAACCTCCGCAACTCTCC |
| Miao143 | 5’-GATGACGCGGTAGAATCCAGCTCTCGGTTCATAGACCCTGCC |
| Miao144 | 5'-CACGGGATCCTCAACCAACAGAAGCTCTTCGAAGGTGATGGTGATGGTGATGCTCCTC |
| Miao145 | 5'-CACGGTCGACTCAACCAACAGAAGCTCTTCGAAGGTGATGGTGATGGTGATGTGAGCGAC |
| Miao146 | 5'-AAGCTTGCGGCCGCACTCGAGCACCACC |

# Bibliography


1. Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A*, **74**, 5088-5090.
2. Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A*, **87**, 4576-4579.
3. Pester, M., Schleper, C. and Wagner, M. (2011) The *Thaumarchaeota*: an emerging view of their phylogeny and ecophysiology. *Curr Opin Microbiol*, **14**, 300-306.
4. Brochier-Armanet, C., Forterre, P. and Gribaldo, S. (2011) Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr Opin Microbiol*, **14**, 274-281.
5. Grabowski, B. and Kelman, Z. (2003) Archeal DNA replication: eukaryal proteins in a bacterial context. *Annu Rev Microbiol*, **57**, 487-516.
6. Zivanovic, Y., Lopez, P., Philippe, H. and Forterre, P. (2002) *Pyrococcus* genome comparison evidences chromosome shuffling-driven evolution. *Nucleic Acids Res*, **30**, 1902-1910.
7. Lundgren, M. and Bernander, R. (2005) Archaeal cell cycle progress. *Curr Opin Microbiol*, **8**, 662-668.
8. Lundgren, M. and Bernander, R. (2007) Genome-wide transcription map of an archaeal cell cycle. *Proc Natl Acad Sci U S A*, **104**, 2939-2944.
9. McGeoch, A.T. and Bell, S.D. (2008) Extra-chromosomal elements and the evolution of cellular DNA replication machineries. *Nat Rev Mol Cell Biol*, **9**, 569-574.
10. Morikawa, M., Izawa, Y., Rashid, N., Hoaki, T. and Imanaka, T. (1994) Purification and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic *Pyrococcus* sp. *Appl Environ Microbiol*, **60**, 4559-4566.
11. Atomi, H., Fukui, T., Kanai, T., Morikawa, M. and Imanaka, T. (2004) Description of *Thermococcus kodakaraensis* sp. nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp. KOD1. *Archaea*, **1**, 263-267.
12. Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S. and Imanaka, T. (2005) Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res*, **15**, 352-363.
13. Sato, T., Fukui, T., Atomi, H. and Imanaka, T. (2003) Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J Bacteriol*, **185**, 210-220.
14. Sato, T., Fukui, T., Atomi, H. and Imanaka, T. (2005) Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *Appl Environ Microbiol*, **71**, 3889-3899.
15. Santangelo, T.J., Cubonova, L. and Reeve, J.N. (2008) Shuttle vector

expression in *Thermococcus kodakaraensis*: contributions of cis elements to protein synthesis in a hyperthermophilic archaeon. *Appl Environ Microbiol*, **74**, 3099-3104.
16. Santangelo, T.J., Cubonova, L. and Reeve, J.N. (2010) *Thermococcus kodakarensis* genetics: TK1827-encoded beta-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl Environ Microbiol*, **76**, 1044-1052.
17. Matsunaga, F., Forterre, P., Ishino, Y. and Myllykallio, H. (2001) *In vivo* interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc Natl Acad Sci U S A*, **98**, 11152-11157.
18. Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H. and Forterre, P. (2000) Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science*, **288**, 2212-2215.
19. Robinson, N.P., Dionne, I., Lundgren, M., Marsh, V.L., Bernander, R. and Bell, S.D. (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell*, **116**, 25-38.
20. Norais, C., Hawkins, M., Hartman, A.L., Eisen, J.A., Myllykallio, H. and Allers, T. (2007) Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet*, **3**, e77.
21. Liu, J., Smith, C.L., DeRyckere, D., DeAngelis, K., Martin, G.S. and Berger, J.M. (2000) Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. *Mol Cell*, **6**, 637-648.
22. Singleton, M.R., Morales, R., Grainge, I., Cook, N., Isupov, M.N. and Wigley, D.B. (2004) Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. *J Mol Biol*, **343**, 547-557.
23. Grabowski, B. and Kelman, Z. (2001) Autophosphorylation of archaeal Cdc6 homologues is regulated by DNA. *J Bacteriol*, **183**, 5459-5464.
24. Haugland, G.T., Sakakibara, N., Pey, A.L., Rollor, C.R., Birkeland, N.K. and Kelman, Z. (2008) *Thermoplasma acidophilum* Cdc6 protein stimulates MCM helicase activity by regulating its ATPase activity. *Nucleic Acids Res*, **36**, 5602-5609.
25. Shin, J.H., Heo, G.Y. and Kelman, Z. (2008) The *Methanothermobacter thermautotrophicus* Cdc6-2 protein, the putative helicase loader, dissociates the minichromosome maintenance helicase. *J Bacteriol*, **190**, 4091-4094.
26. Akita, M., Adachi, A., Takemura, K., Yamagami, T., Matsunaga, F. and Ishino, Y. (2011) Cdc6/Orc1 from *Pyrococcus furiosus* may act as the origin recognition protein and Mcm helicase recruiter. *Genes Cells*, **15**, 537-552.
27. Li, Z., Santangelo, T.J., Cubonova, L., Reeve, J.N. and Kelman, Z. (2010) Affinity purification of an archaeal DNA replication protein network. *MBio*, **1**.
28. Haugland, G.T., Rollor, C.R., Birkeland, N.K. and Kelman, Z. (2009) Biochemical characterization of the minichromosome maintenance protein

from the archaeon *Thermoplasma acidophilum*. *Extremophiles*, **13**, 81-88.
29. Shin, J.H., Heo, G.Y. and Kelman, Z. (2009) The *Methanothermobacter thermautotrophicus* MCM helicase is active as a hexameric ring. *J Biol Chem*, **284**, 540-546.
30. Pan, M., Santangelo, T.J., Li, Z., Reeve, J.N. and Kelman, Z. (2011) *Thermococcus kodakarensis* encodes three MCM homologs but only one is essential. *Nucleic Acids Res*, **39**, 9671-9680.
31. Waga, S. and Stillman, B. (1998) The DNA replication fork in eukaryotic cells. *Annu Rev Biochem*, **67**, 721-751.
32. Nick McElhinny, S.A., Gordenin, D.A., Stith, C.M., Burgers, P.M. and Kunkel, T.A. (2008) Division of labor at the eukaryotic replication fork. *Mol Cell*, **30**, 137-144.
33. Walter, P., Klein, F., Lorentzen, E., Ilchmann, A., Klug, G. and Evguenieva-Hackenberg, E. (2006) Characterization of native and reconstituted exosome complexes from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Microbiol*, **62**, 1076-1089.
34. Le Breton, M., Henneke, G., Norais, C., Flament, D., Myllykallio, H., Querellou, J. and Raffin, J.P. (2007) The heterodimeric primase from the euryarchaeon *Pyrococcus abyssi*: a multifunctional enzyme for initiation and repair? *J Mol Biol*, **374**, 1172-1185.
35. Lao-Sirieix, S.H., Pellegrini, L. and Bell, S.D. (2005) The promiscuous primase. *Trends Genet*, **21**, 568-572.
36. Klinge, S., Hirst, J., Maman, J.D., Krude, T. and Pellegrini, L. (2007) An iron-sulfur domain of the eukaryotic primase is essential for RNA primer synthesis. *Nat Struct Mol Biol*, **14**, 875-877.
37. Vaithiyalingam, S., Warren, E.M., Eichman, B.F. and Chazin, W.J. (2011) Insights into eukaryotic DNA priming from the structure and functional interactions of the 4Fe-4S cluster domain of human DNA primase. *Proc Natl Acad Sci U S A*, **107**, 13684-13689.
38. Matsunaga, F., Norais, C., Forterre, P. and Myllykallio, H. (2003) Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep*, **4**, 154-158.
39. Liu, L., Komori, K., Ishino, S., Bocquier, A.A., Cann, I.K., Kohda, D. and Ishino, Y. (2001) The archaeal DNA primase: biochemical characterization of the p41-p46 complex from *Pyrococcus furiosus*. *J Biol Chem*, **276**, 45484-45490.
40. Bocquier, A.A., Liu, L., Cann, I.K., Komori, K., Kohda, D. and Ishino, Y. (2001) Archaeal primase: bridging the gap between RNA and DNA polymerases. *Curr Biol*, **11**, 452-456.
41. Chemnitz Galal, W., Pan, M., Kelman, Z. and Hurwitz, J. (2012) Characterization of the DNA primase complex isolated from the archaeon, *Thermococcus kodakaraensis*. *J Biol Chem*, **287**, 16209-16219.
42. Chemnitz Galal, W., Pan, M., Giulian, G., Yuan, W., Li, S., Edwards, J.L., Marino, J.P., Kelman, Z. and Hurwitz, J. (2012) Formation of dAMP-glycerol and dAMP-Tris derivatives by the *Thermococcus kodakaraensis* DNA primase. *J Biol Chem*, **287**, 16220-16229.

43. Moyer, S.E., Lewis, P.W. and Botchan, M.R. (2006) Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci U S A*, **103**, 10236-10241.
44. Makarova, K.S., Wolf, Y.I., Mekhedov, S.L., Mirkin, B.G. and Koonin, E.V. (2005) Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res*, **33**, 4626-4638.
45. MacNeill, S.A. (2010) Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochem J*, **425**, 489-500.
46. Oyama, T., Ishino, S., Fujino, S., Ogino, H., Shirai, T., Mayanagi, K., Saito, M., Nagasawa, N., Ishino, Y. and Morikawa, K. (2011) Architectures of archaeal GINS complexes, essential DNA replication initiation factors. *BMC Biol*, **9**, 28.
47. Bell, S.D. (2011) DNA replication: archaeal oriGINS. *BMC Biol*, **9**, 36-37.
48. Marinsek, N., Barry, E.R., Makarova, K.S., Dionne, I., Koonin, E.V. and Bell, S.D. (2006) GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep*, **7**, 539-545.
49. Yoshimochi, T., Fujikane, R., Kawanami, M., Matsunaga, F. and Ishino, Y. (2008) The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. *J Biol Chem*, **283**, 1601-1609.
50. Ogino, H., Ishino, S., Mayanagi, K., Haugland, G.T., Birkeland, N.K., Yamagishi, A. and Ishino, Y. (2011) The GINS complex from the thermophilic archaeon, *Thermoplasma acidophilum* may function as a homotetramer in DNA replication. *Extremophiles*, **15**, 529-539.
51. Li, Z., Pan, M., Santangelo, T.J., Chemnitz, W., Yuan, W., Edwards, J.L., Hurwitz, J., Reeve, J.N. and Kelman, Z. (2011) A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res*, **39**, 6114-6123.
52. Henneke, G., Flament, D., Hubscher, U., Querellou, J. and Raffin, J.P. (2005) The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. *J Mol Biol*, **350**, 53-64.
53. Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T. and Kai, Y. (2001) Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *J Mol Biol*, **306**, 469-477.
54. Kelman, Z., Pietrokovski, S. and Hurwitz, J. (1999) Isolation and characterization of a split B-type DNA polymerase from the archaeon *Methanobacterium thermoautotrophicum* ΔH. *J Biol Chem*, **274**, 28751-28761.
55. Cann, I.K., Komori, K., Toh, H., Kanai, S. and Ishino, Y. (1998) A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proc Natl Acad Sci U S A*, **95**, 14250-14255.
56. Gueguen, Y., Rolland, J.L., Lecompte, O., Azam, P., Le Romancer, G., Flament, D., Raffin, J.P. and Dietrich, J. (2001) Characterization of two

DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*. *Eur J Biochem*, **268**, 5961-5969.
57. Shen, Y., Tang, X.F., Matsui, E. and Matsui, I. (2004) Subunit interaction and regulation of activity through terminal domains of the family D DNA polymerase from *Pyrococcus horikoshii*. *Biochem Soc Trans*, **32**, 245-249.
58. Tahirov, T.H., Makarova, K.S., Rogozin, I.B., Pavlov, Y.I. and Koonin, E.V. (2009) Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. *Biol Direct*, **4**, 11.
59. Garg, P. and Burgers, P.M. (2005) DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit Rev Biochem Mol Biol*, **40**, 115-128.
60. Stukenberg, P.T., Studwell-Vaughan, P.S. and O'Donnell, M. (1991) Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J Biol Chem*, **266**, 11328-11334.
61. Pan, M., Kelman, L.M. and Kelman, Z. (2011) The archaeal PCNA proteins. *Biochem Soc Trans*, **39**, 20-24.
62. Seybert, A., Singleton, M.R., Cook, N., Hall, D.R. and Wigley, D.B. (2006) Communication between subunits within an archaeal clamp-loader complex. *EMBO J*, **25**, 2209-2218.
63. Seybert, A. and Wigley, D.B. (2004) Distinct roles for ATP binding and hydrolysis at individual subunits of an archaeal clamp loader. *EMBO J*, **23**, 1360-1371.
64. Ilves, I., Petojevic, T., Pesavento, J.J. and Botchan, M.R. (2010) Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Mol Cell*, **37**, 247-258.
65. Makarova, K.S., Koonin, E.V. and Kelman, Z. (2012) The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol Direct*, **7**, 7.
66. Tye, B.K. (1999) Minichromosome maintenance as a genetic assay for defects in DNA replication. *Methods*, **18**, 329-334.
67. Tye, B.K. and Sawyer, S. (2000) The hexameric eukaryotic MCM helicase: building symmetry from nonidentical parts. *J Biol Chem*, **275**, 34833-34836.
68. Forsburg, S.L. (2004) Eukaryotic MCM proteins: beyond replication initiation. *Microbiol Mol Biol Rev*, **68**, 109-131.
69. Bochman, M.L. and Schwacha, A. (2009) The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiol Mol Biol Rev*, **73**, 652-683.
70. Gambus, A., Jones, R.C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R.D. and Labib, K. (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol*, **8**, 358-366.
71. Sakakibara, N., Kelman, L.M. and Kelman, Z. (2009) Unwinding the structure and function of the archaeal MCM helicase. *Mol Microbiol*, **72**, 286-296.
72. Duggin, I.G. and Bell, S.D. (2006) The chromosome replication

machinery of the archaeon *Sulfolobus solfataricus*. *J Biol Chem*, **281**, 15029-15032.
73. Costa, A. and Onesti, S. (2009) Structural biology of MCM helicases. *Crit Rev Biochem Mol Biol*, **44**, 326-342.
74. Brewster, A.S. and Chen, X.S. (2010) Insights into the MCM functional mechanism: lessons learned from the archaeal MCM complex. *Crit Rev Biochem Mol Biol*, **45**, 243-256.
75. Kelman, L.M. and Kelman, Z. (2003) Archaea: an archetype for replication initiation studies? *Mol Microbiol*, **48**, 605-615.
76. Walters, A.D. and Chong, J.P. (2009) *Methanococcus maripaludis*: an archaeon with multiple functional MCM proteins? *Biochem Soc Trans*, **37**, 1-6.
77. Walters, A.D. and Chong, J.P. (2010) An archaeal order with multiple minichromosome maintenance genes. *Microbiology*, **156**, 1405-1414.
78. Krupovic, M., Gribaldo, S., Bamford, D.H. and Forterre, P. (2010) The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. *Mol Biol Evol*, **27**, 2716-2732.
79. Chia, N., Cann, I. and Olsen, G.J. (2010) Evolution of DNA replication protein complexes in eukaryotes and Archaea. *PLoS One*, **5**, e10866.
80. Santangelo, T.J. and Reeve, J.N. (2010) In Horikoshi, K. (ed.), *Extremophiles Handbook* Springer Japan.
81. Santangelo, T.J., Cubonova, L., James, C.L. and Reeve, J.N. (2007) TFB1 or TFB2 is sufficient for *Thermococcus kodakaraensis* viability and for basal transcription *in vitro*. *J Mol Biol*, **367**, 344-357.
82. Kasiviswanathan, R., Shin, J.H., Melamud, E. and Kelman, Z. (2004) Biochemical Characterization of the *Methanothermobacter thermautotrophicus* Minichromosome Maintenance (MCM) Helicase N-terminal Domains. *J Biol Chem*, **279**, 28358-28366.
83. Ladner, J.E., Pan, M., Hurwitz, J. and Kelman, Z. (2011) Crystal structures of two active proliferating cell nuclear antigens (PCNAs) encoded by *Thermococcus kodakaraensis*. *Proc Natl Acad Sci U S A*, **108**, 2711-2716.
84. Shin, J.H., Jiang, Y., Grabowski, B., Hurwitz, J. and Kelman, Z. (2003) Substrate requirements for duplex DNA translocation by the eukaryal and archaeal minichromosome maintenance helicases. *J Biol Chem*, **278**, 49053-49062.
85. Santangelo, T.J., Cubonova, L. and Reeve, J.N. (2010) *Thermococcus kodakarensis* genetics: TK1827-encoded beta-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl Environ Microbiol*, **76**, 1044-1152.
86. Perler, F.B. (2002) InBase: the Intein Database. *Nucleic Acids Res*, **30**, 383-384.
87. Kelman, Z., Lee, J.K. and Hurwitz, J. (1999) The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* ΔH contains DNA helicase activity. *Proc Natl Acad Sci U S A*, **96**, 14783-14788.

88. Chong, J.P., Hayashi, M.K., Simon, M.N., Xu, R.M. and Stillman, B. (2000) A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. *Proc Natl Acad Sci U S A*, **97**, 1530-1535.
89. Shechter, D.F., Ying, C.Y. and Gautier, J. (2000) The intrinsic DNA helicase activity of *Methanobacterium thermoautotrophicum* ΔH minichromosome maintenance protein. *J Biol Chem*, **275**, 15049-15059.
90. Sakakibara, N., Kasiviswanathan, R., Melamud, E., Han, M., Schwarz, F.P. and Kelman, Z. (2008) Coupling of DNA binding and helicase activity is mediated by a conserved loop in the MCM protein. *Nucleic Acids Res*, **36**, 1309-1320.
91. Atanassova, N. and Grainge, I. (2008) Biochemical characterization of the minichromosome maintenance (MCM) protein of the crenarchaeote *Aeropyrum pernix* and its interactions with the origin recognition complex (ORC) proteins. *Biochemistry*, **47**, 13362-13370.
92. Liew, L.P. and Bell, S.D. (2011) The interplay of DNA binding, ATP hydrolysis and helicase activities of the archaeal MCM helicase. *Biochem J*, **436**, 409-414.
93. Patel, S.S. and Picha, K.M. (2000) Structure and function of hexameric helicases. *Annu Rev Biochem*, **69**, 651-697.
94. Enemark, E.J. and Joshua-Tor, L. (2008) On helicases and other motor proteins. *Curr Opin Struct Biol*, **18**, 243-257.
95. Berthon, J., Cortez, D. and Forterre, P. (2008) Genomic context analysis in Archaea suggests previously unrecognized links between DNA replication and translation. *Genome Biol*, **9**, R71.
96. Soler, N., Gaudin, M., Marguet, E. and Forterre, P. (2011) Plasmids, viruses and virus-like membrane vesicles from *Thermococcales*. *Biochem Soc Trans*, **39**, 36-44.
97. Fanning, E. and Knippers, R. (1992) Structure and function of simian virus 40 large tumor antigen. *Annu Rev Biochem*, **61**, 55-85.
98. Challberg, M.D. and Kelly, T.J. (1989) Animal virus DNA replication. *Annu Rev Biochem*, **58**, 671-717.
99. Jeruzalmi, D., O'Donnell, M. and Kuriyan, J. (2002) Clamp loaders and sliding clamps. *Curr Opin Struct Biol*, **12**, 217-224.
100. Indiani, C. and O'Donnell, M. (2006) The replication clamp-loading machine at work in the three domains of life. *Nat Rev Mol Cell Biol*, **7**, 751-761.
101. Jeruzalmi, D. (2005) The opened processivity clamp slides into view. *Proc Natl Acad Sci U S A*, **102**, 14939-14940.
102. Vivona, J.B. and Kelman, Z. (2003) The diverse spectrum of sliding clamp interacting proteins. *FEBS Lett*, **546**, 167-172.
103. Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., Stahl, D.A., Wagner, M. and Schleper, C. Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum *Thaumarchaeota*. *Trends Microbiol*, **18**, 331-340.
104. Naryzhny, S.N., Zhao, H. and Lee, H. (2005) Proliferating cell nuclear

antigen (PCNA) may function as a double homotrimer complex in the mammalian cell. *J Biol Chem*, **280**, 13888-13894.
105. Kelman, Z. and Hurwitz, J. (2000) A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon M*ethanobacterium thermoautotrophicum* ΔH. *J Biol Chem*, **275**, 7327-7336.
106. Cann, I.K., Ishino, S., Yuasa, M., Daiyasu, H., Toh, H. and Ishino, Y. (2001) Biochemical analysis of replication factor C from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol*, **183**, 2614-2623.
107. Kelman, Z. and O'Donnell, M. (1995) Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps. *Nucleic Acids Res*, **23**, 3613-3620.
108. Stoimenov, I. and Helleday, T. (2009) PCNA on the crossroad of cancer. *Biochem Soc Trans*, **37**, 605-613.
109. Moldovan, G.L., Pfander, B. and Jentsch, S. (2007) PCNA, the maestro of the replication fork. *Cell*, **129**, 665-679.
110. Indiani, C. and O'Donnell, M. (2006) The replication clamp-loading machine at work in the three domains of life. *Nat Rev Mol Cell Biol*, **7**, 751-761.
111. Dionne, I., Brown, N.J., Woodgate, R. and Bell, S.D. (2008) On the mechanism of loading the PCNA sliding clamp by RFC. *Mol Microbiol*, **68**, 216-222.
112. Kazmirski, S.L., Zhao, Y., Bowman, G.D., O'Donnell, M. and Kuriyan, J. (2005) Out-of-plane motions in open sliding clamps: molecular dynamics simulations of eukaryotic and archaeal proliferating cell nuclear antigen. *Proc Natl Acad Sci U S A*, **102**, 13801-13806.
113. Yao, N., Turner, J., Kelman, Z., Stukenberg, P.T., Dean, F., Shechter, D., Pan, Z.Q., Hurwitz, J. and O'Donnell, M. (1996) Clamp loading, unloading and intrinsic stability of the PCNA, β and gp45 sliding clamps of human, *E. coli* and T4 replicases. *Genes Cells*, **1**, 101-113.
114. Kelman, Z. (1997) PCNA: structure, functions and interactions. *Oncogene*, **14**, 629-640.
115. Cann, I.K., Ishino, S., Hayashi, I., Komori, K., Toh, H., Morikawa, K. and Ishino, Y. (1999) Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in *Archaea*. *J Bacteriol*, **181**, 6591-6599.
116. Kelman, Z. and Hurwitz, J. (2000) A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon *Methanobacterium thermoautotrophicum* ΔH. *J Biol Chem*, **275**, 7327-7336.
117. Seybert, A., Scott, D.J., Scaife, S., Singleton, M.R. and Wigley, D.B. (2002) Biochemical characterisation of the clamp/clamp loader proteins from the euryarchaeon *Archaeoglobus fulgidus*. *Nucleic Acids Res*, **30**, 4329-4338.
118. Dionne, I., Nookala, R., Jackson, S.P., Doherty, A.J. and Bell, S.D. (2003)

A heterotrimeric PCNA in the hyperthermophilic archaeaon *Sulfolobus solfataricus*. *Mol Cell*, **11**, 275-282.
119. Daimon, K., Kawarabayasi, Y., Kikuchi, H., Sako, Y. and Ishino, Y. (2002) Three proliferating cell nuclear antigen-like proteins found in the hyperthermophilic archaeon *Aeropyrum pernix*: interactions with the two DNA polymerases. *J Bacteriol*, **184**, 687-694.
120. Matsumiya, S., Ishino, Y. and Morikawa, K. (2001) Crystal structure of an archaeal DNA sliding clamp: Proliferating cell nuclear antigen from *Pyrococcus furiosus*. *Protein Sci*, **10**, 17-23.
121. Chapados, B.R., Hosfield, D.J., Han, S., Qiu, J., Yelent, B., Shen, B. and Tainer, J.A. (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell*, **116**, 39-50.
122. Morgunova, E., Gray, F.C., Macneill, S.A. and Ladenstein, R. (2009) Structural insights into the adaptation of proliferating cell nuclear antigen (PCNA) from *Haloferax volcanii* to a high-salt environment. *Acta Crystallogr D Biol Crystallogr*, **65**, 1081-1088.
123. Winter, J.A., Christofi, P., Morroll, S. and Bunting, K.A. (2009) The crystal structure of *Haloferax volcanii* proliferating cell nuclear antigen reveals unique surface charge characteristics due to halophilic adaptation. *BMC Struct Biol*, **9**, 55.
124. Williams, G.J., Johnson, K., Rudolf, J., McMahon, S.A., Carter, L., Oke, M., Liu, H., Taylor, G.L., White, M.F. and Naismith, J.H. (2006) Structure of the heterotrimeric PCNA from *Sulfolobus solfataricus*. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **62**, 944-948.
125. Krishna, T.S., Kong, X.P., Gary, S., Burgers, P.M. and Kuriyan, J. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell*, **79**, 1233-1243.
126. McNally, R., Bowman, G.D., Goedken, E.R., O'Donnell, M. and Kuriyan, J. (2010) Analysis of the role of PCNA-DNA contacts during clamp loading. *BMC Struct Biol*, **10**, 3.
127. Kelman, Z. and Hurwitz, J. (1998) Protein-PCNA interactions: a DNA-scanning mechanism? *Trends Biochem Sci*, **23**, 236-238.
128. Warbrick, E. (1998) PCNA binding through a conserved motif. *Bioessays*, **20**, 195-199.
129. Warbrick, E., Heatherington, W., Lane, D.P. and Glover, D.M. (1998) PCNA binding proteins in *Drosophila melanogaster*: the analysis of a conserved PCNA binding domain. *Nucleic Acids Res*, **26**, 3925-3932.
130. Ren, B., Kuhn, J., Meslet-Cladiere, L., Briffotaux, J., Norais, C., Lavigne, R., Flament, D., Ladenstein, R. and Myllykallio, H. (2009) Structure and function of a novel endonuclease acting on branched DNA substrates. *Embo J*, **28**, 2479-2489.
131. Humbard, M.A., Miranda, H.V., Lim, J.M., Krause, D.J., Pritz, J.R., Zhou, G., Chen, S., Wells, L. and Maupin-Furlow, J.A. (2010) Ubiquitin-like small archaeal modifier proteins (SAMPs) in *Haloferax volcanii*. *Nature*, **463**, 54-60.
132. Kirkland, P.A. and Maupin-Furlow, J.A. (2009) Stabilization of an

archaeal DNA-sliding clamp protein, PCNA, by proteasome-activating nucleotidase gene knockout in *Haloferax volcanii*. *FEMS Microbiol Lett*, **294**, 32-36.
133. Henneke, G., Koundrioukoff, S. and Hubscher, U. (2003) Phosphorylation of human Fen1 by cyclin-dependent kinase modulates its role in replication fork regulation. *Oncogene*, **22**, 4301-4313.
134. Kennelly, P.J. (2003) Archaeal protein kinases and protein phosphatases: insights from genomics and biochemistry. *Biochem J*, **370**, 373-389.