Thermococcus kodakarensis encodes three MCM homologs but only one is essential

Miao Pan[1], Thomas J. Santangelo[2,3], Zhuo Li[1], John N. Reeve[2] and Zvi Kelman[1,*]

[1]Institute for Bioscience and Biotechnology Research, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850, [2]Department of Microbiology and [3]Center for RNA Biology, Ohio State University, Columbus, OH 43210, USA

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 *Thermococcus kodakarensis* genome encodes three archaeal MCM homologs, 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 (MCM3) have been unsuccessful arguing that that MCM3 is essential and is likely the replicative helicase in *T. kodakarensis*. The origins and possible function(s) of the three MCM proteins are discussed.

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 homologs (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 (1–4)]. The heterohexameric complex, and also trimeric complexes formed by MCM4, MCM6 and MCM7 have $3' \rightarrow 5'$ helicase activity *in vitro* although the replicative helicase *in vivo* seems to be a larger assembly of all six MCM homologs, Cdc45 and the GINS complex, together designated the CMG complex [Cdc45, MCM and GINS; (5–7)]. The replicative helicase in archaea is also thought to contain an MCM complex [reviewed in (8–12)], but most archaea have only one MCM homolog. A central domain that embodies the AAA+ catalytic ATPase is conserved in both the archaeal and eukaryotic MCMs, but the eukaryotic proteins have 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 $3' \rightarrow 5'$ helicase activity and can bind and translocate along single-stranded (ss) and double-stranded (ds) DNA, displace proteins bound to DNA and unwind DNA–RNA hybrids [reviewed in (9,11,12)].

With many archaeal genome sequences now available, a few species with more than one MCM homolog have been identified with gene duplication and lateral gene transfer posited as explanations (13–17). The genome of *Thermococcus kodakarensis* encodes three MCM homologs (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.

*To whom correspondence should be addressed. Tel: +240 314 6294; Fax: +240 314 6255; Email: zkelman@umd.edu

MATERIALS AND METHODS

Media and growth conditions

Thermococcus 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 l (ASW-YT medium) or with a mixture of 20 amino acids (ASW-AA) (18). 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 (19).

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 (20). The sequences of all PCR primers used in this study are available on request. The plasmids generated (Table 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 (Table 2) confirmed that the (TK0254 + TK0664) cassette was integrated into the *T. kodakarensis* genome, adjacent to the target gene and flanked by a direct duplication of genomic DNA. Expression of TK0664 resulted in these transformants being sensitive to 6-methyl purine (6 MP). Mutants spontaneously resistant to 6MP were selected as clones that grew on ASW-AA 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 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 *T. kodakarensis* TS601 (Δ*pyrF*; Δ*trpE*::*pyrF*; ΔTK0664; ΔTK0096) as the recipient strain, generated *T. kodakarensis* TS604 (Δ*pyrF*; Δ*trpE*::*pyrF*; ΔTK0664; ΔTK0096; ΔTK1361). All the *T. kodakarensis* strains used and generated in this study are listed in Table 3.

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 TK1620 (MCM3) includes sequences that encode two inteins that were removed by using PCR as previously described (21). The resulting open reading frame (here designated TK1620) and TK1361 (MCM2) were cloned into pET-21a (Novagen), with six histidine codons added in-frame to their 3′-termini, resulting in plasmids designated pET-TK1620 and pET-TK1361 (Table 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 generate 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

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

Primer set	Sequence	Chromosomal location
I	GGCCAACGCCACTCGACCC GGGACC	TK1194083-1194107
	GAGGATTTGAGTGGTGGT GAGCCG	TK1194581-1194558
II	CAACCCAATCTGCCGTAACG	TK1191965-1191984
	TCTTCCTCTTCGGCTTCATG	TK1195741-1195722
III	GAGGAGAAGCTGGTCAGGG CTTTC	TK81123-81146
	CTCATACCCACACTTAGGA CACAC	TK81572-81549
IV	TATGTACCTGTTCTCTGCGC	TK84557-84538
	GTGATCACTTCTTCACAATG	TK80127-80146

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

Plasmid name	Used to/for
pZLE031	Delete TK0096 *in vivo*
pZLE029	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 3. *Thermococcus kodakarensis* strains used in this study

Strain designation	Relevant genotype	Origin
TS517	Δ*pyrF*; Δ*trpE*::*pyrF*; ΔTK0664	(20)
TS601	Δ*pyrF*; Δ*trpE*::*pyrF*; ΔTK0664; ΔTK0096	This study
TS602	Δ*pyrF*; Δ*trpE*::*pyrF*; ΔTK0664; ΔTK1361	This study
TS604	Δ*pyrF*; Δ*trpE*::*pyrF*; ΔTK0664; ΔTK0096; ΔTK1361	This study

PCR-based site-directed mutagenesis as previously described (21).

Synthesis 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 Luria-Bertani containing 100 μg/ml ampicillin and 25 μg/ml chloramphenicol at an OD_{600} of ∼0.6. Incubation was continued for 16 h at 16°C. Cells were collected by centrifugation, resuspended and incubated in 50 mM Tris–HCl (pH 8), 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 on to 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), 1 M NaCl, 10% glycerol, 50 mM imidazole and then with 50 mM Tris–HCl (pH 8), 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), 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 ethylenediaminetetraacetic acid (EDTA), 2 mM dithiothreitol (DTT) and 10% glycerol.

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 1 h 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% sodium dodecyl sulfate (SDS) and visualized by staining with Coomassie brilliant blue (R250).

Helicase assays

Oligonucleotides MD007 and MD015 (Table 4) were [^{32}P]-end labeled by incubation with [γ-^{32}P]-ATP (3 kCi/mmol; Perkin Elmer) and T4 polynucleotide kinase (Fermentas). The 25 and 96 bp double-stranded (ds) [^{32}P]-labeled DNA substrates used for helicase assays were generated by hybridization of [^{32}P]-MD007 with MD008, and [^{32}P]-MD014 with MD015, respectively (Table 4), and purified as previously described (22).

DNA helicase activity was assayed in reaction mixtures (15 μl) that contained 20 mM Tris–HCl (pH 8), 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 1 and 5. The reaction mixtures were incubated at 70°C for 1 h and the reaction was then stopped by addition of 5 μl loading buffer (0.1% xylene cyanol, 0.1% bromophenol blue, 1% SDS, 50% glycerol and 100 mM EDTA), and placing the tube 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 0.5× TBE (45 mM Tris, 45 mM boric acid, 0.5 mM EDTA) for 40 min at 180 V. The helicase assays were repeated at least three times, and the averages 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.

ATPase assays

ATPase activity was assayed in reaction mixtures (15 μl) that contained 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 (3 kCi/mmol), plus or minus 10 pmol of the 49-mer oligonucleotide MD008 (Table 4), plus the MCM protein, as noted in the Figure 3 legend. 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

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

Name	Sequence
MD007	5′-TTTGTTTGTTTGTTTGTTTGTTTGCCGACGTGCCAGGCCGACGCGTCCC
MD008	5′-GGGACGCGTCGGCCTGGCACGTCGGTTTGTTTGTTTGTTTGTTTGTTTG
MD012	5′-TTTGTTTGTTTGTTTGTTTGTTTGCGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCG
MD013	5′-CGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCGTTTGTTTGTTTGTTTGTTTGTTTG
MD014	5′-TTTGTTTGTTTGTTTGTTTGCGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCG
	CGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCGCGACGGCCAGTGCCAAGCTTGCAT
MD015	5′-ATGCAAGCTTGGCACTGGCCGTCGCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTC
	GCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTGTTTGTTTGTTTGTTTGTTTG
MD016	5′-CGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCG
A1	5′-(Cy5)-GGGGCGAGTCCAGGTCAGGACCTTGCGGGG

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 the rates of ATP hydrolysis, reaction mixtures (45 µl) that contained 25 mM Tris–HCl (pH 8), 5 mM MgCl$_2$, 1 mM DTT, 4.5 µg BSA, 4.5 nmol [γ-^{32}P]-ATP (3 kCi/mmol) and 4 pmol MCM1 or MCM3 monomer, or 0.2 pmol of MCM2 monomer were incubated at 75°C, 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 MCM3) or 0, 2, 5, 10, 15, 30 and 45 min (MCM2) incubation, mixed with 1 µl of 0.5 M EDTA, and the extent of ATP hydrolysis was determined as described above.

Measurement of DNA binding by fluorescence polarization anisotropy (FPA)

A 30-mer oligonucleotide (A1, Table 4), synthesized with Cy5 at the 5′-terminus, 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 287 900/M/cm) for DNA and A$_{646}$ (extinction coefficient 250 000/M/cm) 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 25°C, FPA measurements were taken at 25°C using a Fluoromax-3 spectrofluorimeter equipped with an autopolarizer. The cuvette (3 mm 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 s 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, ΔA_T is the total anisotropy change, E_T is the enzyme concentration at each titration point, D_T is the total concentration of DNA (assumed to be 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.
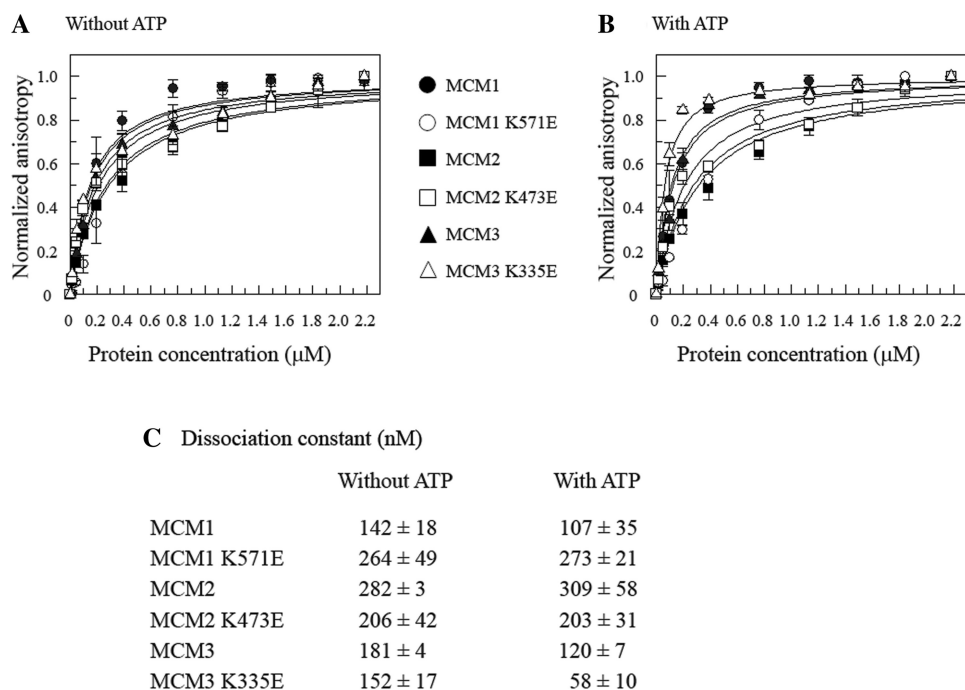
RESULTS

The *T. kodakarensis* genome encodes three MCM homologs

Sequencing the *T. kodakarensis* genome (13) revealed the presence of three genes (TK0096, TK1361, TK1620) that were predicted to encode MCM homologs, here

designated MCM1, MCM2 and MCM3, respectively. An alignment of their amino acid sequence with those of established archaeal (Supplementary Figure S1) and eukaryotic MCMs confirms the presence of all motifs
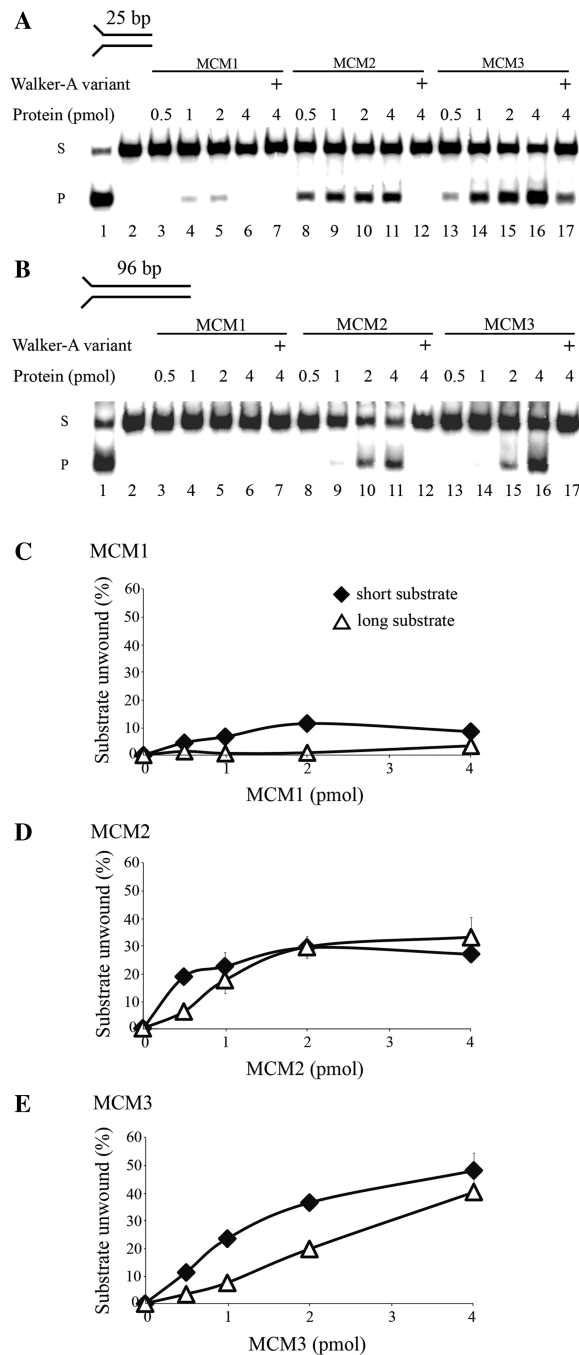
Figure 1. Helicase assays of MCM1, MCM2 and MCM3. Electrophoretic separation of the products of helicase assays with 10 fmol of the [^{32}P]-labeled (**A**) 25 bp and (**B**) 96 bp duplex substrates generated with increasing concentrations of MCM1 (lanes 3–6), MCM2 (lanes 8–11) and MCM3 (lanes 13–16). 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 Without ATP

B With ATP

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. DNA binding by MCM1, MCM2, MCM3 and the Walker-A K→E variants. DNA binding to 10 nM Cy5-labeled oligonucleotide A1 measured by FPA 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**).

needed for helicase activity, and demonstrates that MCM1 and MCM2 have unique 205 and 136 residue N-terminal extensions, respectively (Supplementary Figure S1). As encoded in the *T. kodakarensis* genome, MCM3 contains two inteins, one between the Walker-A and Walker-B motifs of the AAA+ domain and the second is C-terminal to the Walker-B motif. An intein is present at the same location as the first MCM3 intein in the single MCM encoded in related *Pyrococcal* genomes, and there is an intein in a similar location to the second MCM3 intein in the MCMs of *Methanoculleus marisnigri* and *Staphylothermus marinus* (23).

MCM2 and MCM3 have helicase activity

MCM2 and MCM3 had robust helicase activity (Figure 1), and efficiently unwound both the 25 and 96 bp substrate, consistent with processive enzymes and with the activities reported for other archaeal MCMs [as examples see Refs (24–26)]. The helicase activities of MCM2 and MCM3 required the presence of ATP or dATP (Supplementary Figure S2) and a lysine residue conserved in the Walker-A motif (K473 in MCM2 and K335 in MCM3), was required for both ATPase (data not shown) and helicase (Figure 1 and Supplementary Figure S2) activities. In contrast, MCM1 had much lower helicase activity *in vitro*. MCM1 exhibited low but detectable helicase activity with the 25-bp substrate at low enzyme concentrations (Figure 1A, lane 5) but, for reasons that remain unclear, this activity decreased at higher enzyme concentrations (Figure 1A, lane 6; Figure 1C).

All three MCMs bind DNA and hydrolyze ATP

Helicase activity requires DNA binding followed by ATP-dependent translocation along the DNA substrate. Although MCM1 exhibited only low helicase activity, MCM1 bound DNA in the absence (Figure 2A) and presence (Figure 2B) of ATP with affinities similar to that of MCM2, MCM3 and the DNA-binding affinities reported for other archaeal MCMs (27–29). MCM1 and MCM3 had relatively low but readily measurable ATPase activities (Figure 3A and C), with rates of ATP hydrolysis that were stimulated only ~2-fold by the presence of DNA (Figure 3D), close to the activities reported for other archaeal MCMs (28,30,31). MCM2, in contrast, exhibited robust ATPase activity (Figure 3B) and the rate of ATP hydrolysis was stimulated ~7-fold by the presence of DNA (Figure 3D). Given that MCM1 and MCM3 have similar affinities for DNA and ATPase activities, it seems unlikely that a deficiency in DNA binding or ATPase activity explains the minimal ability of MCM1 to unwind DNA *in vitro* (Figure 1).

MCM2 and MCM3 form hexameric complexes

For helicase activity, six MCM subunits assemble to form a hexameric ring-shaped complex (4,11,32,33). Size exclusion chromatography revealed that MCM1 monomers (104 kDa) formed complexes in solution with an estimated molecular mass of ~345 kDa (Figure 4A) and therefore that contained either three (312 kDa) or four (416 kDa) monomers. There was no evidence for assembly MCM1 into hexamers (636 kDa) *in vitro* providing one
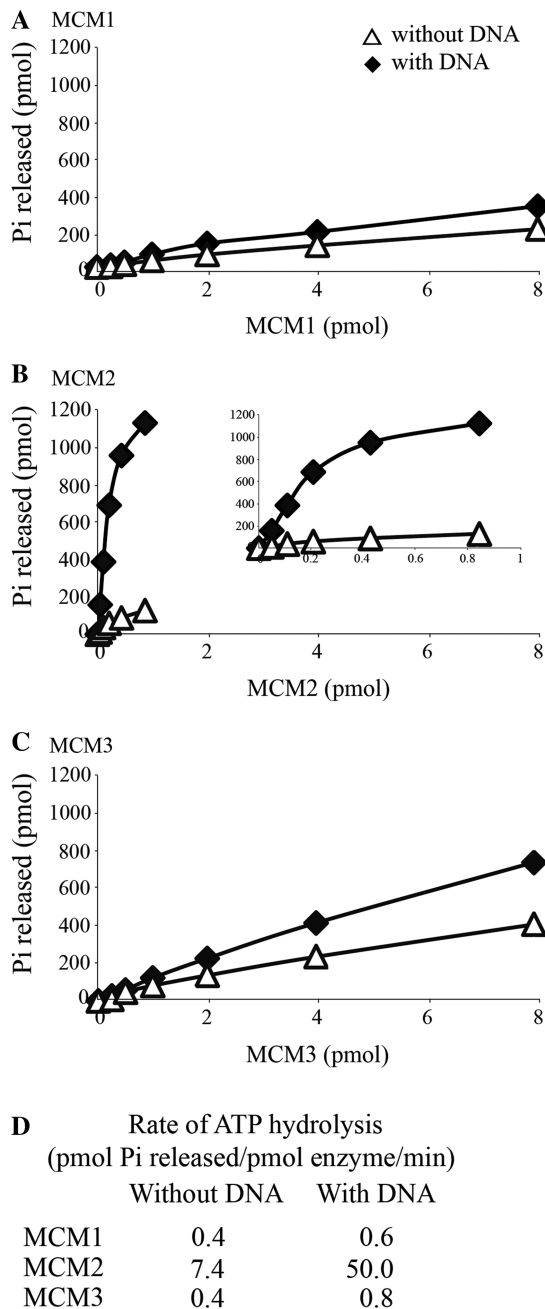
Figure 3. ATPase assays of MCM1, MCM2 and MCM3. Assays were carried in reaction mixtures that contained increasing amounts of (**A**) MCM1, (**B**) MCM2 and (**C**) MCM3 in the presence or absence of the oligonucleotide MD008. The rates of ATP hydrolysis were calculated and (**D**) shows the averages of the results from three independent experiments. The standard deviations of all experiments were <10%.

explanation for the lack of helicase activity under the experimental conditions investigated. It remains possible that MCM1 could assemble into hexamers at higher protein concentrations, as might exist *in vivo*. 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) (Figure 4B). MCM3 (77.4 kDa) eluted as a single protein peak (Figure 4C), 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 (data not shown).

Removal of the N-terminal extension increases MCM2 helicase activity

As illustrated in Supplementary Figure S1, MCM1 and MCM2 have N-terminal extensions, formed by 205 and 136 amino acid residues, respectively. MCM1-ΔN and MCM2-ΔN variants that lacked the N-terminal extensions were generated (Figure 5A), purified and their helicase activities determined (Figure 5B and 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 (Figure 5B and C) although the N-terminal deletion did not similarly increase the protein's ATPase activity (Supplementary Figure S3A and B) or DNA binding ability (Supplementary Figure S3C). Apparently, the N-terminal structure of MCM2 modulates its helicase activity by a mechanism independent of ATPase activity and DNA binding.

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, respectively. *T. kodakarensis* TS604 (ΔTK0096; ΔTK1361), a strain lacking both MCM1 and MCM2, was then readily generated from *T. kodakarensis* TS601 consistent with the loss of MCM1 not significantly reducing homologous recombination. Figure 6A illustrates the strategies used to confirm the genome organizations in *T. kodakarensis* TS601, TS602 and TS604, and examples of the diagnostic PCR and Southern blots results are shown in Figure 6B and C, respectively. Despite repeated attempts, we have been unable to generate a strain with TK1620 (MCM3) deleted arguing that MCM3 is likely essential for *T. kodakarensis* viability. As *T. kodakarensis* TS604 (ΔTK0096; ΔTK1361) exhibits no detectable growth defects, the presence of MCM3 is apparently sufficient for genome replication and, as in most archaea, the *T. kodakarensis* replisome can function with homohexamer assembly of one MCM homolog.

DISCUSSION

MCM3 is the replicative helicase in *T. kodakarensis*

Recombinant His$_6$-tagged versions of the three MCM proteins predicted by bioinformatics to exist in *T. kodakarensis* have been purified and all three bind DNA and have ATPase activity *in vitro* that is dependent on an intact Walker-A motif (data not shown). However, only the MCM2 and MCM3 homologs spontaneously assembled into hexamers and exhibit robust helicase activity *in vitro*. In contrast, recombinant MCM1 did

A MCM1

B MCM2

C MCM3

D
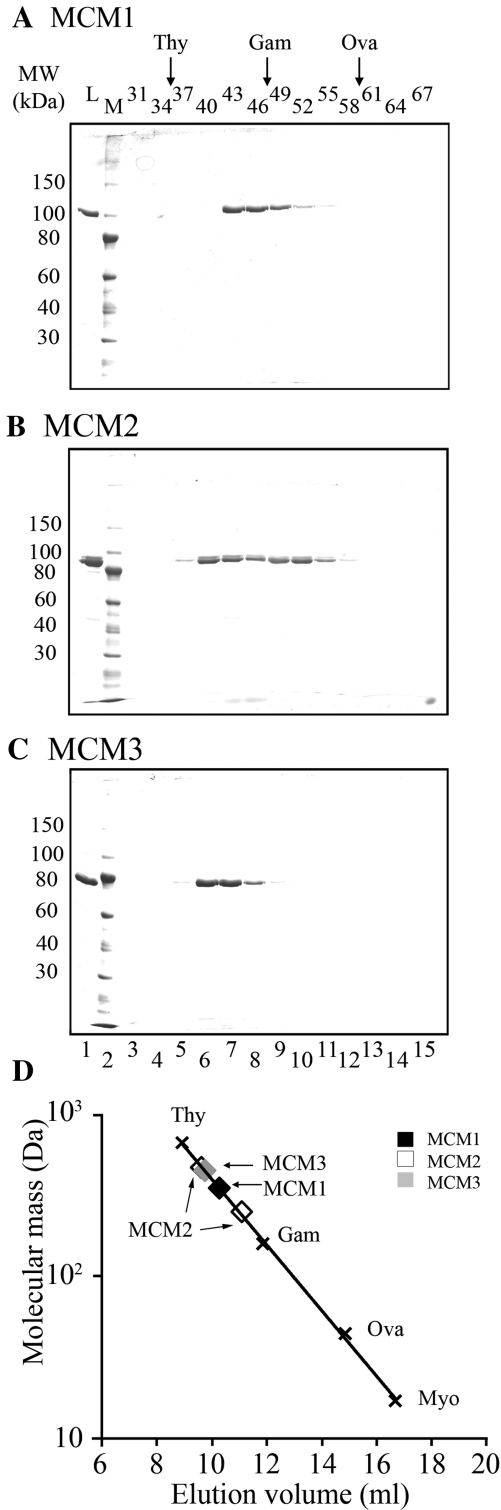
Figure 4. Sephadex-200 size exclusion chromatography of MCM1, MCM2 and MCM3. The proteins present in fractions from (**A**) MCM1, (**B**) MCM2 and (**C**) MCM3, separated by Superdex-200 gel filtration chromatography, Coomassie blue stained after resolution by SDS–PAGE. The positions at which thyroglobulin (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). (**D**) Shows the elution positions of the MCM1, MCM2 and MCM3 complexes relative to those of the mass standards.

not form stable hexamers and exhibited barely detectable helicase activity *in vitro*. Deletion of the genes encoding MCM1 (TK0096) and MCM2 (TK1361) had no 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 for this conclusion, MCM3 is most similar in size to other archaeal MCMs and is encoded in an operon that also encodes the GINS23 (TK1619) subunit of the replisome (34,35). In contrast, MCM1 and MCM2 have atypical structures, with unique N-terminal extensions (Figure 5A, Supplementary Figure S1) and their encoding genes are not closely linked in the *T. kodakarensis* genome to genes that encode known replication proteins. The initial genome annotation and a subsequent in-depth bioinformatics analysis have both concluded that TK0096 (MCM1) and TK1361 (MCM2) are located in regions of

A

B

C
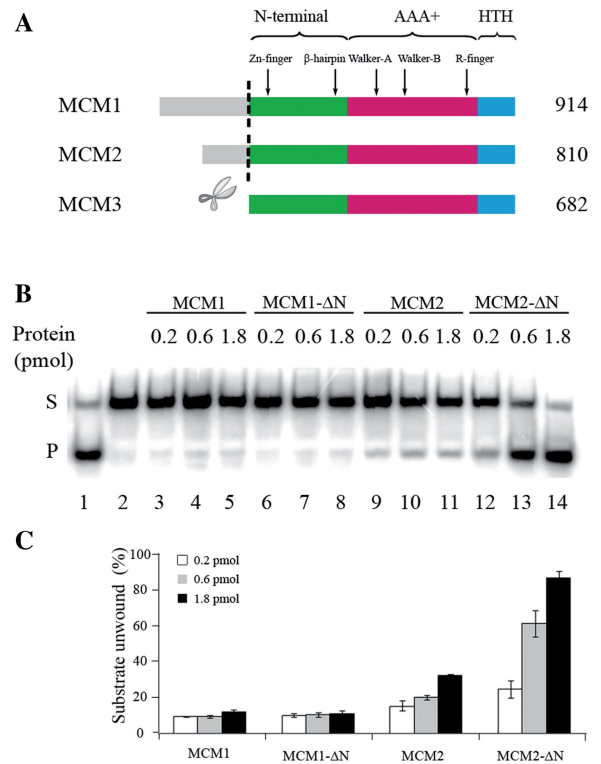
Figure 5. Helicase assays of MCM1-ΔN and MCM2-ΔN. (**A**) Illustration of MCM1, MCM2 and MCM3 with the N-terminal, AAA+ catalytic region and the helix-turn-helix (HTH) regions identified. Arrows indicate the locations of motifs required for helicase activity, and the number of amino acids present is listed to right of each protein. The sites at which MCM1 and MCM2 were truncated to generate MCM1-ΔN and MCM2-ΔN are noted and identified precisely in Supplementary Figure S1. (**B**) Helicase activities of the indicated MCM proteins in reaction mixtures that contained 10 fmol of 25 bp duplex DNA substrate 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.
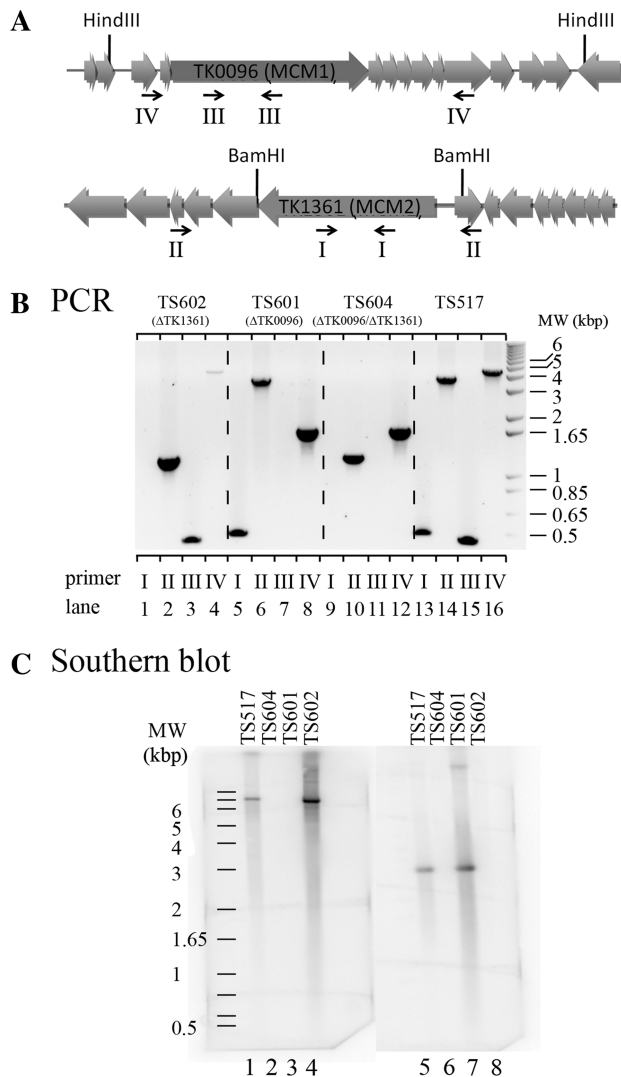
Figure 6. Genome organizations, PCR and Southern blot confirmation of the *T. kodakarensis* ΔTK0096 and ΔTK1361 deletions. (**A**) Genome organizations surrounding TK0096 (MCM1) and TK1361 (MCM2). The positions at which the PCR primers (Roman numeral primer pairs I through IV, Table 2) hybridized and 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 (ΔpyrF, ΔtrpE::pyrF, ΔTK0664), TS601 (ΔpyrF, ΔtrpE::pyrF, ΔTK0664, ΔTK0096), TS602 (ΔpyrF, ΔtrpE::pyrF, ΔTK0664, ΔTK1361) and TS604 (ΔpyrF, ΔtrpE::pyrF, ΔTK0664, ΔTK0096, ΔTK1361) with the positions of DNA size standards indicated. As shown, primers internal to TK1361 [primer pair I (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 ∼3.8 kbp amplicons, which contain the TK1361 sequence (2.4 kbp) from *T. kodakarensis* TS517 and TS601, but amplicons that were only ∼1.4 kbp from *T. kodakarensis* TS602 and TS604 genomic DNAs consistent with the loss of TK1361. Primers specific to TK0096 (primer pair III) amplified ∼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 ∼4.4 kbp amplicon, which included the TK0096 sequence (2.7 kbp) from *T. kodakarensis* TS517 and TS602, but amplicons that were only ∼1.6 kbp amplicon from *T. kodakarensis* TS601 and TS604 genomic DNAs consistent with the loss of TK0096. (**C**) Southern blot analyses of genomic DNA. Genomic DNA (10 μg)

the *T. kodakarensis* genome that are remnants of past viral infections (13,16). It seems a reasonable hypothesis therefore 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 (14–16,36). As proposed (14), all the MCMs present may still be functional but, as the *T. kodakarensis* results suggest, one MCM, most likely the endogenous enzyme, may predominate in the archaeal genome replisome.

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 regulation and/or replication of an infecting viral or plasmid DNA or in the activation of a prophage. This would resemble the role of the simian virus 40 (SV40) large T antigen (37,38) that, together with host proteins, contributes to both SV40 origin recognition and functions as the viral replicative helicase. Possibly, MCM1 and/or MCM2 had similar functions in archaeal viral replications in ancestors of the *T. kodakarensis* lineage with these activities regulated by their atypical N-terminal extensions.

Although their abundances remain to be determined, the presence of MCM1 and MCM2 *in vivo* has been established (39) arguing that they do likely have functions in *T. kodakarensis*. MCM1-His$_6$ and MCM2-His$_6$ isolated from *T. kodakarensis* cell lysates by binding to a Ni^{2+}-charged matrix were present in complexes that also contained the archaeal DNA polymerases B and D and the processivity factor PCNA1 (39). The presence of MCM1 and MCM2 in such complexes argues for their participation in DNA metabolic events, possibly in recombination and/or DNA repair and, consistent with this notion, MCM2 also co-purified from *T. kodakarensis* cell lysates with a MutS homolog, an established DNA repair enzyme (39). 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

Figure 6. Continued
from *T. kodakarensis* TS517 (ΔpyrF, ΔtrpE::pyrF, ΔTK0664); TS601 (ΔpyrF, ΔtrpE::pyrF, ΔTK0664, ΔTK0096); TS602 (ΔpyrF, ΔtrpE::pyrF, ΔTK0664, ΔTK1361) and TS604 (ΔpyrF, ΔtrpE::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 (19). The membranes were incubated with a [^{32}P]-labeled oligonucleotide that hybridized to a sequence internal to TK0096 (left panel) or TK1361 (right panel). TK0096 is located within ∼7 kbp HindIII fragment that was present only in *T. kodakarensis* TS517 and TS602 genomic DNAs. TK1361 is located within ∼3 kbp BamHI fragment (right panel) that was present only in *T. kodakarensis* TS517 and TS602 genomic DNAs.

archaeal PCNAs (40), and only PCNA1 is essential for *T. kodakarensis* viability (TJS, MP, ZL, JNR and ZK, unpublished data). Based on their co-isolation in complexes, both MCM1 and MCM2 interact with PCNA1 but not PCNA2 (encoded by TK0582). Studies are now underway to determine if *T. kodakarensis* strains lacking MCM1, MCM2 and/or PCNA2 exhibit defects in DNA repair and/or recombination.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Tye,B.K. (1999) MCM proteins in DNA replication. *Annu. Rev. Biochem.*, **68**, 649–686.
2. Tye,B.K. and Sawyer,S.L. (2000) The hexameric eukaryotic MCM helicase: building symmetry from nonidentical parts. *J. Biol. Chem.*, **275**, 34833–34836.
3. Forsburg,S.L. (2004) Eukaryotic MCM proteins: beyond replication initiation. *Microbiol. Mol. Biol. Rev.*, **68**, 109–131.
4. Bochman,M.L. and Schwacha,A. (2009) The mcm complex: unwinding the mechanism of a replicative helicase. *Microbiol. Mol. Biol. Rev.*, **73**, 652–683.
5. 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. USA*, **103**, 10236–10241.
6. 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.
7. 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.
8. Kelman,L.M. and Kelman,Z. (2003) Archaea: an archetype for replication initiation studies? *Mol. Microbiol.*, **48**, 605–615.
9. Sakakibara,N., Kelman,L.M. and Kelman,Z. (2009) Unwinding the structure and function of the archaeal MCM helicase. *Mol. Microbiol.*, **72**, 286–296.
10. Duggin,I.G. and Bell,S.D. (2006) The chromosome replication machinery of the archaeon *Sulfolobus solfataricus*. *J. Biol. Chem.*, **281**, 15029–15032.
11. Costa,A. and Onesti,S. (2009) Structural biology of MCM helicases. *Crit. Rev. Biochem. Mol. Biol.*, **44**, 326–342.
12. 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.
13. 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.
14. Walters,A.D. and Chong,J.P. (2009) *Methanococcus maripaludis*: an archaeon with multiple functional MCM proteins? *Biochem. Soc. Trans.*, **37**, 1–6.
15. Walters,A.D. and Chong,J.P. (2010) An archaeal order with multiple minichromosome maintenance genes. *Microbiology*, **156**, 1405–1414.
16. 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 hitch-hiking of mobile genetic elements. *Mol. Biol. Evol.*, **27**, 2716–2732.
17. Chia,N., Cann,I. and Olsen,G.J. (2010) Evolution of DNA replication protein complexes in eukaryotes and Archaea. *PLoS One*, **5**, e10866.
18. Santangelo,T.J. and Reeve,J.N. (2010) In Horikoshi,K. (ed.), *Extremophiles Handbook*. Springer, Japan.
19. 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.
20. 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.
21. 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.
22. 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.
23. Perler,F.B. (2002) InBase: the Intein Database. *Nucleic Acids Res.*, **30**, 383–384.
24. 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. USA*, **96**, 14783–14788.
25. 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. USA*, **97**, 1530–1535.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.

31. 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.

32. Patel,S.S. and Picha,K.M. (2000) Structure and function of hexameric helicases. *Annu. Rev. Biochem.*, **69**, 651–697.

33. Enemark,E.J. and Joshua-Tor,L. (2008) On helicases and other motor proteins. *Curr. Opin. Struct. Biol.*, **18**, 243–257.

34. 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.

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

36. 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.

37. Fanning,E. and Knippers,R. (1992) Structure and function of simian virus 40 large tumor antigen. *Annu. Rev. Biochem.*, **61**, 55–85.

38. Challberg,M.D. and Kelly,T.J. (1989) Animal virus DNA replication. *Annu. Rev. Biochem.*, **58**, 671–717.

39. Li,Z., Santangelo,T.J., Čuboňová,L., Reeve,J.N. and Kelman,Z. (2010) Affinity purification of an archaeal DNA replication protein network. *MBio*, **1**, e00221–e00210.

40. 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. USA*, **108**, 2711–2716.