Unwinding twenty years of the archaeal minichromosome maintenance helicase

Lori M. Kelman[1], William B. O'Dell[2] and Zvi Kelman[2]

1. Program in Biotechnology, Montgomery College, 20200 Observation Drive, Germantown, MD 20876, USA

2. Biomolecular Labeling Laboratory, National Institute of Standards and Technology and Institute for Bioscience and Biotechnology Research, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850, USA

Corresponding author:
Zvi Kelman
NIST/IBBR
9600 Gudelsky Drive
Rockville, MD 20850
USA

Phone: 240-314-6294
Fax: 240-314-6255
E-mail: zkelman@umd.edu

Running Title: The archaeal MCM helicase

Keywords: archaea, DNA replication, helicase, MCM, minichromosome maintenance, three-dimensional structure

Abstract

Replicative DNA helicases are essential cellular enzymes that unwind duplex DNA in front of the replication fork during chromosomal DNA replication. Replicative helicases were discovered, beginning in the 1970s, in bacteria, bacteriophages, viruses, eukarya, and, in the mid-1990s, in archaea. This year marks the 20th anniversary of the first report on the archaeal replicative helicase, the minichromosome maintenance (MCM) protein. This minireview summarizes two decades of work on the archaeal MCM.

Introduction

In 1996, the complete genome of the first archaeon, *Methanocaldococcus jannaschii* (named *Methanococcus jannaschii* at the time) was published (1). Since then many aspects of archaeal biology and physiology have been studied. Because many archaeal species are extremophiles, some of these studies focused on the biotechnological applications of archaea and archaeal enzymes (e.g. PCR, molecular cloning, environmental remediation), while others concentrated on exploring the similarities and differences between archaea and the other two domains, bacteria and eukarya, with respect to physiology and cellular processes. Figure 1 summarizes the timeline of research on the archaeal MCM helicase.

Many of these studies focused on the archaeal DNA replication machinery both as a source for biotechnology reagents (e.g. thermostable DNA polymerases for PCR) and as a group of microorganisms with a unique replication process. When the complete genomes of several archaeal species were determined, bioinformatics studies suggested that although archaea are prokaryotes with a circular chromosome, like bacteria, their replication machinery is more similar to that of eukarya (Table 1) (the reader is referred to several reviews on the archaeal replication machinery for details (2-4)). In the following years, biochemical, structural, and genetic studies demonstrated the relationship between the archaeal and eukaryal DNA replication machineries. These studies also revealed that, although, in general, the archaeal replication process is more similar to that of eukarya, some aspects are more bacterial-like, and others are

archaeal-specific (Table 1). For example, the replicative helicase in archaea, the MCM (minichromosome maintenance) protein, is a homologue of the eukaryotic MCM and not the bacterial DnaB protein, and it translocates on DNA in the 3'-5' direction as does the eukaryotic helicase. The bacterial DnaB translocates in the 5'-3' direction (Table 2). Another example is the DNA sliding clamp. While the bacterial protein, the β-subunit of DNA polymerase III, forms homo-dimers (5), the eukaryotic and archaeal proteins, proliferating cell nuclear antigen (PCNA), form homo-trimers (5, 6). Worth noting, however, all three clamps have similar three-dimensional structures and all have a pseudo six-fold symmetry (7). However, some features of the replication machinery are archaeal-specific, such as the archaeal-specific DNA Polymerase D, found in some species as the only essential DNA polymerase (8) (Table 1).

The replicative helicase of bacteria and eukarya

In bacteria, the replicative helicase is the DnaB protein, which forms a homo-hexameric ring with helicase activity and is essential for DNA replication and cell viability ((9) and references therein). In eukarya, the MCM protein is a family of six related proteins, MCM2-7, that are essential for chromosomal DNA replication (10-12). All six proteins belong to the AAA+ family of ATPases (ATPases associated with diverse cellular activities) and contain all the hallmarks of other members of the family (13, 14). Based on amino acid sequence analysis, the largest conserved portion of the six proteins is a region of about 300 amino acids that contains the domains involved in ATPase activity. A region of about 250 residues, N-terminal to the catalytic part, is also conserved among the six eukaryotic MCM proteins. Outside of these regions the eukaryotic MCM proteins show no similarity with each other and each contains long, diverse N- and C-terminal regions (15).

Although the eukaryotic MCM2-7 proteins contain all the elements of a DNA helicase, *in vivo*, the MCM2-7 complex is tightly associated with two additional factors, the Cdc45 protein and the hetero-tetrameric GINS complex. Together, these form the CMG (Cdc45, MCM, GINS) complex that functions as the replicative helicase in eukarya (10,

11). Each of the components of the CMG complex are essential for cell viability (Table 1).

All archaeal genomes encode for MCM homologues

When the genome sequences of several archaeal species were analyzed, some proteins were annotated as putative helicases. Edgell and Doolittle were the first to recognize the presence of MCM homologues in the archaeal genomes (Fig. 1) (16). Subsequent studies showed that all archaeal species contain at least one homologue of a MCM protein (17), and this was suggested to function as the replicative helicase. The archaeal MCM proteins, however, are shorter than the eukaryotic enzymes. Most are about 650 amino acids in length, and include a 250-residue N-terminal portion and an approximately 300-amino acid catalytic region (Fig. 2). Both of these regions are similar to the eukaryotic MCM2-7 proteins. The enzymes also contain ~100 amino acid C-terminal regions suggested to fold into a helix-turn-helix (HTH) motif (17, 18) (Fig. 2). The C-terminal region is thought to play a regulatory function (19, 20). In several archaeal species with multiple MCM homologues, some are longer than 650 amino acids. However, in the few cases where the enzymes were studied, it was found that only the MCM proteins that are similar to all other archaeal MCMs are essential for cell viability (21, 22).

The biochemical properties of the archaeal MCM proteins

The first report on the biochemical properties of the archaeal MCM was a talk given by James Chong, then a post-doc in Bruce Stillman's laboratory, at the 1999 Cold Spring Harbor meeting on "Eukaryotic DNA Replication". This presentation, and subsequent publications from three groups, focused on the initial characterization of the MCM protein from *Methanothermobacter thermautotrophicus* (then called *Methanobacterium thermoautotrophicum* ΔH) (23-25). These early studies showed that the protein is a 3'-to-5' ATP-dependent DNA helicase, binds to single stranded (ss) and double stranded

(ds) DNA, has a processivity of several hundred bases, and forms a homo-dodecameric structure in solution (Table 2).

Research on the biochemical properties of the archaeal MCM proteins was expanded to enzymes from other species and kingdoms. These studies illuminated the diverse activities of the helicase, the role of specific residues and domains in MCM function, and factors involved in the regulation of helicase activity. The similarities and differences between MCM homologues from different species were also examined. These studies explored the processivity of the enzymes (26), and regions involved in DNA binding including the Zn-finger motif (27) and the N-terminal portion (28). The studies also demonstrated the ability of the helicase to translocate along ss- and dsDNA (29), the ability to displace proteins from DNA during translocation (30), and to displace RNA from DNA–RNA hybrid duplexes while translocating on the DNA strand (31) (Table 2). Many of these activities are consistent with MCM serving as the archaeal replicative helicase, as they are shared by the eukaryotic MCM and the bacterial replicative helicase DnaB (32).

In eukarya, under most experimental conditions the MCM helicase is not active on its own. Only the CMG complex possesses helicase activity, and the CMG complex is the active helicase *in vivo* (33, 34). The situation in archaea, however, is more complex. While most of the archaeal MCM proteins studied are active on their own (e.g. (23)), some require additional factors for appreciable helicase activity (e.g. (35)). And in some cases, opposite effects can be observed with the proteins from different species. For example, while the initiator protein Cdc6 (also referred to as Orc1) stimulates the *in vitro* helicase activity of MCM from some species (for example *Thermoplasma acidophilum* (35)), it inhibits the activity of others (for example *M. thermautotrophicus* (36)). Another example of MCM-interacting enzymes that affect helicase activity is the MCM association with the archaeal GINS and GAN proteins (also referred to as the archaeal Cdc45 protein or RecJ). In some species the GMG (GAN, MCM, GINS) complex (also referred to as the archaeal CMG) has no effect on helicase activity *in vitro*, although all

three components are present in all archaeal species (37). In other species, however, the complex stimulates helicase activity (38, 39).

Single molecule analysis studies were also employed to determine the properties of the helicase. Single-molecule FRET (fluorescence resonance energy transfer) studies identified the interactions between the MCM protein and the DNA substrate and show that the helicase interacts better with a fork substrate than with a substrate with only a 3'-overhanging ssDNA region (40). The processivity of the helicase was also determined using a high temperature single-molecule bead tether assay to study the speed and processivity of several archaeal enzymes. These studies revealed that, *in vitro*, archaeal MCMs from some species possess a processivity of several thousand bases without the need for accessory factors (Table 2) (26).

MCM structure

The three-dimensional structures of the MCM proteins were determined using different techniques. The first observation on the structure of the MCM complex came from low resolution size-exclusion chromatography studies reported in the first few publications on the *M. thermautotrophicus* protein (23, 24). These studies suggested that, in solution, the helicase forms a double-hexameric ring structure. This was exciting, as it strongly suggested that the MCM protein is the replicative helicase. This stemmed from knowledge that the bacterial replicative helicase, DnaB, and the large tumor antigen (T-Ag) of simian virus 40 (SV40) are single polypeptides that form dodecameric rings that encircle DNA ((9) and references therein).

These observations were followed by electron microscopy (EM) studies of the full-length protein from *M. thermautotrophicus*. These studies showed that the protein can adopt different oligomeric structures depending on protein concentration and buffer conditions. These structures include hexamers, heptamers, octamers, dodecamers, open rings, and filaments (41, 42). Although the enzyme can form multiple structures, it was suggested that, at least *in vitro*, only the hexamers possess helicase activity (43). EM

studies also showed that when provided with long dsDNA the DNA wraps around the hexameric ring (44). This wrapping was suggested to play a role during the initiation of replication.

The first high-resolution structures of the MCM were an X-ray structure of the N-terminal portion of the *M. thermautotrophicus* protein (45, 46) followed by the structure of the N-terminal part of the protein from other species (47, 48) (Fig. 3). The structures revealed a hexameric arrangement, with each monomer folded into two distinct domains: domain A and domain B/C. The structures opened the door for detailed biochemical, functional, and structure–function studies of the different domains, regions, and residues of the N-terminal region. These studies elucidated the role of the N-terminal portion in MCM multimerization, ss- and dsDNA binding, and ATPase activity (28). The structures also revealed a loop, not identified by sequence analysis, that is highly conserved among archaeal and eukaryal MCM proteins. This loop was shown to play an important role in communication between the N-terminal DNA binding region and the ATPase activity of the catalytic portion (49).

In addition, the solution structure of the N-terminal part of the protein was also determined using small-angle neutron scattering (SANS) and demonstrated a large movement of domain A with respect to the other domain (50).

The structures of the N-terminal portion were followed by an X-ray structure of the near-full-length MCM protein from *Sulfolobus solfataricus* (51). This structure, although it does not include the entire protein and was of low resolution, was instrumental in advancing the research on the MCM proteins (52). As had been predicted by amino acid sequence analysis, the structure confirmed the presence of all conserved motifs found in other AAA+ proteins. However, several motifs not identified by sequence analysis were also observed . The structure revealed four β-hairpins per monomer, three located within the main channel and one on the exterior of the hexamer. Mutational analysis of the latter elucidated its role in DNA binding and helicase activity (52, 53). The structure of the full-length protein in the presence of ssDNA was also

determined (54) (Fig. 4). The structure suggested that, like DnaB, the helicase moves with a step of two nucleotides per MCM subunit. A structure of a chimeric MCM protein that included the N-terminal portion of the *S. solfataricus* protein and the catalytic domains of *Pyrococcus furiosus* was also determined using X-ray crystallography (55).

The solution structure of the full-length protein from *M. thermautotrophicus* was also determined using SANS (56) and suggested that all twelve AAA+ domains lie at approximately the same distance from the axis. The results also indicated that domain A of the N-terminal portion of each monomer is next to the AAA+ region for all twelve monomers.

Genetic studies

Two decades ago, the ability to study archaeal proteins *in vivo* was very limited due to the lack of robust genetic tools. This changed, however, and in the past decade genetic methods were developed for several archaeal species (57-60). Genetic studies show that all archaeal species depend on a single MCM protein for chromosomal replication. Here, archaea are similar to bacteria, where a single protein, DnaB, is multimerized to assemble the active helicase (Table 1). However, the archaeal helicase is biochemically and structurally similar to eukarya (Table 1).

Genetic tools were also used to identify proteins that interact with MCM. For example, the *Thermococcus kodakarensis* MCM proteins were tagged *in vivo*, and interacting proteins were identified by protein complex purification followed by mass-spectrometric analysis (61). Some of the proteins identified were known to be involved in DNA replication (e.g. DNA polymerase), while others are of unknown function and only future studies will determine their role, if any, in DNA replication or other cellular processes and the roles of their interactions with MCM.

Future directions

One of the outstanding questions regarding the MCM is how the hexameric ring is loaded onto DNA at the origin of replication. Although the initiator protein, Cdc6, was implicated in the assembly process (62, 63) the mechanism is not known, and several different processes were suggested (64). The newly developed single molecule approaches may help in addressing this essential question in MCM function.

In the past several years, a large number of new archaeal species, lineages, groups, and supergroups have been identified (for examples see (65, 66)). Unfortunately, many of the newly identified organisms cannot be cultured, and the classification is based largely on metagenomics of environmental samples. Therefore, the organisms cannot be studied directly, but their DNA sequences can be used to express recombinant MCM homologues for *in vitro* analysis. It will be interesting to elucidate the structures and functions of these proteins and to determine their similarities and differences to enzymes from other species.

To date, most of the studies on the archaeal MCM were *in vitro* or *in vivo* genetic studies involving tagged proteins and attempts to delete the gene(s) encoding for MCM from the genome. Few other types of *in vivo* studies have been reported. In the future, *in vivo* imaging studies of proteins in live cells could determine cellular location and kinetics (for examples see (67)). The development of tools for *in vivo* protein labeling for mesophilic and thermophilic organisms may enable the study of helicase activity and localization within the cell during the different stages of the cell cycle (68). Such tools may also help to determine if the MCM protein is needed only for DNA replication or for other cellular processes.

The studies on the replicative helicases of archaea, bacteria, and eukarya illustrate the similarities and differences between the enzymes in the three domains (Table 2). However, while the DnaB proteins in bacteria and the MCM and CMG complexes in eukarya are quite similar between species, it was shown that archaeal MCM proteins are more diverse. This includes the requirement of additional factors for activity and the mechanisms by which helicase activity is regulated. In addition, to date, most archaeal

MCM proteins studied are from thermophilic organisms. It will be of interest to determine if MCM proteins from organisms growing in other extreme environments, such as psychrophiles, are similar to those from thermophiles. Although a great deal has been learned in the last two decades, much remains to be discovered about the archaeal replicative helicase.

Acknowledgments

We thank the dozens of scientists who contributed to the study of the archaeal MCM in the last twenty years. Unfortunately, due to space limitations, we could not cite all of the primary literature.

Dedication

Lori Kelman and Zvi Kelman would like to dedicate the paper to the memory of Jerard “Jerry" Hurwitz, a mentor, colleague, and friend.

References

1. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Venter JC. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii* [see comments]. Science 273:1058-1073.

2. Kelman LM, Kelman Z. 2003. Archaea: an archetype for replication initiation studies? Mol Microbiol 48:605-615.

3. Grabowski B, Kelman Z. 2003. Archaeal DNA replication: eukaryal proteins in a bacterial context. Annu Rev Microbiol 57:487-516.

4. Kelman LM, Kelman Z. 2014. Archaeal DNA replication. Annu Rev Genet 48:71-97.

5. Kelman Z, O'Donnell M. 1995. Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps. Nucleic Acids Res 23:3613-3620.
6. Pan M, Kelman LM, Kelman Z. 2011. The archaeal PCNA proteins. Biochem Soc Trans 39:20-24.
7. Kelman Z, Finkelstein J, O'Donnell M. 1995. Why have six-fold symmetry? Curr Biol 5:1239-1242.
8. Čuboňová L, Richardson T, Burkhart BW, Kelman Z, Reeve JN, Connolly BA, Santangelo TJ. 2013. Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in *Thermococcus kodakarensis*. J Bacteriol 195:2322-2328.
9. Kornberg A, Baker TA. 1992. DNA replication, 2nd / ed. W.H. Freeman, New York.
10. O'Donnell ME, Li H. 2018. The ring-shaped hexameric helicases that function at DNA replication forks. Nat Struct Mol Biol 25:122-130.
11. Onesti S, MacNeill SA. 2013. Structure and evolutionary origins of the CMG complex. Chromosoma 122:47-53.
12. Bell SD, Botchan MR. 2013. The minichromosome maintenance replicative helicase. Cold Spring Harb Perspect Biol 5.
13. Duderstadt KE, Berger JM. 2008. AAA+ ATPases in the initiation of DNA replication. Crit Rev Biochem Mol Biol 43:163-187.
14. Erzberger JP, Berger JM. 2006. Evolutionary relationships and structural mechanisms of AAA+ proteins. Annu Rev Biophys Biomol Struct 35:93-114.
15. Tye BK, Sawyer SL. 2000. The hexameric eukaryotic MCM helicase: Building symmetry from nonidentical parts. J Biol Chem 275:34833-34836.
16. Edgell DR, Doolittle WF. 1997. Archaea and the origin(s) of DNA replication proteins. Cell 89:995-998.
17. Sakakibara N, Kelman LM, Kelman Z. 2009. Unwinding the structure and function of the archaeal MCM helicase. Mol Microbiol 72:286-296.
18. Brewster AS, Chen XS. 2010. Insights into the MCM functional mechanism: lessons learned from the archaeal MCM complex. Crit Rev Biochem Mol Biol 45:243-256.

19. Jenkinson ER, Chong JP. 2006. Minichromosome maintenance helicase activity is controlled by N- and C-terminal motifs and requires the ATPase domain helix-2 insert. Proc Natl Acad Sci U S A 103:7613-7618.
20. Barry ER, McGeoch AT, Kelman Z, Bell SD. 2007. Archaeal MCM has separable processivity, substrate choice and helicase domains. Nucleic Acids Res 35:988-998.
21. Ishino S, Fujino S, Tomita H, Ogino H, Takao K, Daiyasu H, Kanai T, Atomi H, Ishino Y. 2011. Biochemical and genetical analyses of the three MCM genes from the hyperthermophilic archaeon, *Thermococcus kodakarensis*. Genes Cells 16:1176-1189.
22. Pan M, Santangelo TJ, Li Z, Reeve JN, Kelman Z. 2011. *Thermococcus kodakarensis* encodes three MCM homologs but only one is essential. Nucleic Acids Res 39:9671-9680.
23. Kelman Z, Lee JK, 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.
24. Chong JP, Hayashi MK, Simon MN, Xu RM, 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.
25. Shechter DF, Ying CY, Gautier J. 2000. The intrinsic DNA helicase activity of *Methanobacterium thermoautotrophicum* ΔH minichromosome maintenance protein. J Biol Chem 275:15049-15059.
26. Schermerhorn KM, Tanner N, Kelman Z, Gardner AF. 2016. High-temperature single-molecule kinetic analysis of thermophilic archaeal MCM helicases. Nucleic Acids Res 44:8764-8771.
27. Poplawski A, Grabowski B, Long SE, Kelman Z. 2001. The zinc finger domain of the archaeal minichromosome maintenance protein is required for helicase activity. J Biol Chem 276:49371-49377.
28. Kasiviswanathan R, Shin J-H, Melamud E, Kelman Z. 2004. Biochemical characterization of the *Methanothermobacter thermautotrophicus* minichromosome maintenance (MCM) helicase N-terminal domains. J Biol Chem 279:28358-28366.

29. Shin J-H, Jiang Y, Grabowski B, Hurwitz J, Kelman Z. 2003. Substrate requirements for duplex DNA translocation by the eukaryal and archaeal minichromosome maintenance helicases. J Biol Chem 278:49053-49062.
30. Shin J-H, Santangelo TJ, Xie Y, Reeve JN, Kelman Z. 2007. Archaeal minichromosome maintenance (MCM) helicase can unwind DNA bound by archaeal histones and transcription factors. J Biol Chem 282:4908-4915.
31. Shin J-H, Kelman Z. 2006. The replicative helicases of bacteria, archaea and eukarya can unwind RNA-DNA hybrid substrates. J Biol Chem 281:26914-26921.
32. Li Y, Araki H. 2013. Loading and activation of DNA replicative helicases: the key step of initiation of DNA replication. Genes Cells 18:266-277.
33. Li H, O'Donnell ME. 2018. The eukaryotic CMG helicase at the replication fork: emerging architecture reveals an unexpected mechanism. Bioessays 40;1700208.
34. MacNeill SA. 2010. Structure and function of the GINS complex, a key component of the eukaryotic replisome. Biochem J 425:489-500.
35. Haugland GT, Shin J-H, Birkeland NK, Kelman Z. 2006. Stimulation of MCM helicase activity by a Cdc6 protein in the archaeon *Thermoplasma acidophilum*. Nucleic Acids Res 34:6337-6344.
36. Shin J-H, Grabowski B, Kasiviswanathan R, Bell SD, Kelman Z. 2003. Regulation of minichromosome maintenance helicase activity by Cdc6. J Biol Chem 278:38059-38067.
37. Makarova KS, Koonin EV, 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.
38. Xu Y, Gristwood T, Hodgson B, Trinidad JC, Albers SV, Bell SD. 2016. Archaeal orthologs of Cdc45 and GINS form a stable complex that stimulates the helicase activity of MCM. Proc Natl Acad Sci U S A 113:13390-13395.
39. Nagata M, Ishino S, Yamagami T, Ogino H, Simons JR, Kanai T, Atomi H, Ishino Y. 2017. The Cdc45/RecJ-like protein forms a complex with GINS and MCM, and is important for DNA replication in *Thermococcus kodakarensis*. Nucleic Acids Res 45:10693-10705.

399 40. Rothenberg E, Trakselis MA, Bell SD, Ha T. 2007. MCM fork substrate specificity
400 involves dynamic interaction with the 5' tail. J Biol Chem 282:34229-34234.
401 41. Chen YJ, Yu X, Kasiviswanathan R, Shin J-H, Kelman Z, Egelman EH. 2005.
402 Structural Polymorphism of *Methanothermobacter thermautotrophicus* MCM. J Mol Biol
403 346:389-394.
404 42. Yu X, VanLoock MS, Poplawski A, Kelman Z, Xiang T, Tye BK, Egelman EH.
405 2002. The *Methanobacterium thermoautotrophicum* MCM protein can form heptameric
406 rings. EMBO Rep 3:792-797.
407 43. Shin J-H, Heo G-Y, Kelman Z. 2009. The *Methanothermobacter*
408 *thermautotrophicus* MCM helicase is active as a hexameric ring. J Biol Chem 284:540-
409 546.
410 44. Costa A, van Duinen G, Medagli B, Chong J, Sakakibara N, Kelman Z, Nair SK,
411 Patwardhan A, Onesti S. 2008. Cryo-electron microscopy reveals a novel DNA-binding
412 site on the MCM helicase. Embo J 27:2250-2258.
413 45. Fletcher RJ, Bishop BE, Leon RP, Sclafani RA, Ogata CM, Chen XS. 2003. The
414 structure and function of MCM from archaeal *M. thermoautotrophicum*. Nat Struct Biol
415 10:160-167.
416 46. Kelman Z, Hurwitz J. 2003. Structural lessons in DNA replication from the third
417 domain of life. Nat Struct Biol 10:148-150.
418 47. Liu W, Pucci B, Rossi M, Pisani FM, Ladenstein R. 2008. Structural analysis of
419 the *Sulfolobus solfataricus* MCM protein N-terminal domain. Nucleic Acids Res 36:3235-
420 3243.
421 48. Meagher M, Enemark EJ. 2016. Structure of a double hexamer of the
422 Pyrococcus furiosus minichromosome maintenance protein N-terminal domain. Acta
423 Crystallogr F Struct Biol Commun 72:545-551.
424 49. Sakakibara N, Kasiviswanathan R, Melamud E, Han M, Schwarz FP, Kelman Z.
425 2008. Coupling of DNA binding and helicase activity is mediated by a conserved loop in
426 the MCM protein. Nucleic Acids Res 36:1309-1320.
427 50. Krueger S, Shin JH, Raghunandan S, Curtis JE, Kelman Z. 2011. Atomistic
428 ensemble modeling and small-angle neutron scattering of intrinsically disordered protein

complexes: applied to minichromosome maintenance protein. Biophys J 101:2999-3007.
51. Brewster AS, Wang G, Yu X, Greenleaf WB, Carazo JM, Tjajadi M, Klein MG, Chen SX. 2008. Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase. Proc Natl Acad Sci U S A 105:20191-20196.
52. Slaymaker IM, Chen XS. 2012. MCM structure and mechanics: what we have learned from archaeal MCM. Subcell Biochem 62:89-111.
53. Brewster AS, Slaymaker IM, Afif SA, Chen XS. 2010. Mutational analysis of an archaeal minichromosome maintenance protein exterior hairpin reveals critical residues for helicase activity and DNA binding. BMC Mol Biol 11:62.
54. Meagher M, Epling LB, Enemark EJ. 2019. DNA translocation mechanism of the MCM complex and implications for replication initiation. Nat Commun 10:3117.
55. Miller JM, Arachea BT, Epling LB, Enemark EJ. 2014. Analysis of the crystal structure of an active MCM hexamer. Elife 3:e03433.
56. Krueger S, Shin JH, Curtis JE, Rubinson KA, Kelman Z. 2014. The solution structure of full-length dodecameric MCM by SANS and molecular modeling. Proteins 82:2364-2374.
57. Kohler PR, Metcalf WW. 2012. Genetic manipulation of *Methanosarcina* spp. Front Microbiol 3:259.
58. Atomi H, Imanaka T, Fukui T. 2012. Overview of the genetic tools in the Archaea. Front Microbiol 3:337.
59. Farkas JA, Picking JW, Santangelo TJ. 2013. Genetic techniques for the archaea. Annu Rev Genet 47:539-561.
60. Zatopek KM, Gardner AF, Kelman Z. 2018. Archaeal DNA replication and repair: new genetic, biophysical and molecular tools for discovering and characterizing enzymes, pathways and mechanisms. FEMS Microbiol Rev 42:477-488.
61. Li Z, Santangelo TJ, Čuboňová L, Reeve JN, Kelman Z. 2010. Affinity purification of an archaeal DNA replication protein network. MBio 1:e00221-10.
62. Shin J-H, Heo GY, Kelman Z. 2008. The *Methanothermobacter thermautotrophicus* Cdc6-2 protein, the putative helicase loader, dissociates the minichromosome maintenance helicase. J Bacteriol 190:4091-4094.

63. Samson RY, Abeyrathne PD, Bell SD. 2016. Mechanism of archaeal MCM helicase recruitment to DNA replication origins. Mol Cell 61:287-296.
64. Sakakibara N, Kelman LM, Kelman Z. 2009. How is the archaeal MCM helicase assembled at the origin? Possible mechanisms. Biochem Soc Trans 37:7-11.
65. Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. 2017. Archaea and the origin of eukaryotes. Nat Rev Microbiol 15:711-723.
66. Da Cunha V, Gaia M, Nasir A, Forterre P. 2018. Asgard archaea do not close the debate about the universal tree of life topology. PLoS Genet 14:e1007215.
67. Bisson-Filho AW, Zheng J, Garner E. 2018. Archaeal imaging: leading the hunt for new discoveries. Mol Biol Cell 29:1675-1681.
68. Visone V, Han W, Perugino G, Del Monaco G, She Q, Rossi M, Valenti A, Ciaramella M. 2017. *In vivo* and i*n vitro* protein imaging in thermophilic archaea by exploiting a novel protein tag. PLoS One 12:e0185791.

Author Biographies (NOTE: no biography for Zvi Kelman; he does not want one printed)

Lori M. Kelman is a Professor of Biotechnology at Montgomery College, Germantown, Maryland. She received an A.B. in biochemistry from Mount Holyoke College, a M.S. in biology from St. John's University, a MBA in management from Iona College, and a Ph.D. in molecular biology from Cornell University. Prior to coming to Montgomery College, she was on the faculty of Iona College in New Rochelle, NY. She has performed research at the Rockefeller University, Memorial Sloan-Kettering Cancer Center, the National Institutes of Health, the National Institute of Standards and Technology, and the Institute for Bioscience and Biotechnology Research. She is Editor of *BIOS: a quarterly journal of biology*, a journal of undergraduate research and the journal of the Beta Beta Beta Biological Society.

William Brad O'Dell was born in Newport, Tennessee. He received a B.A. in College Scholars Honors (concentration: structural chemistry) from the University of Tennessee, Knoxville in 2009. He completed a Ph.D. in Biochemistry (with Prof. Flora

Meilleur) at NC State University in 2017 while conducting research in neutron protein crystallography in residence at the Neutron Sciences Directorate, Oak Ridge National Laboratory. In 2017, he was awarded a National Research Council Postdoctoral Associateship (with Zvi Kelman) to join the Biomolecular Structure and Function Group, Biomolecular Measurement Division within the National Institute of Standards and Technology, Materials Measurement Laboratory where he works today as a biologist. He pursues his research interests in protein structure determination using neutron scattering methods and in biological consequences of deuterium isotopic labeling through affiliation with the Biomolecular Labeling Laboratory (BL^2) of the Institute for Bioscience and Biotechnology Research.

Figure legends

Figure 1. Milestones of archaeal MCM helicase research. Blue, genetic studies; black, bioinformatics analysis; red, biochemical studies; green, structural studies.

Figure 2. Schematic representation of the archaeal MCM protein. The N-terminal region is responsible for DNA binding and protein multimerization, the AAA+ region is the catalytic portion, and the C-terminal region is unique to the archaeal MCM and is a predicted Helix-Turn-Helix motif. The three major regions of the protein are shown at the top, and some of the structural motifs are shown at the bottom.

Figure 3. Structures of the archaeal MCM proteins N-terminal regions. A) Ribbon diagrams of (left to right) *M. thermautotrophicus* (PDB ID 1LTL), *S. solfataricus* (PDB ID 2VL6), *T. acidophilum* (PDB ID 4ME3) and *Pyrococcus furiosus* (PDB ID 4POF) viewed from the N-terminal face. For *M. thermautotrophicus* and *S. solfataricus*, crystallographic symmetry was applied to reconstruct the hexamer, while for *T. acidophilum* the hexamer was constructed by superposition with the crystallized *P. furiosus* hexamer. B) The same viewed from right of the N-terminal face. C) Calculated solvent-accessible surfaces colored by electrostatic potential.

Figure 4: The structure of the full-length *S. solfataricus* MCM protein in the presence of ssDNA. A) Ribbon diagram (PDB ID 6MII) viewed from the N-terminal face. The ssDNA molecule is shown in gray. B) Calculated protein solvent-accessible surface colored by electrostatic potential viewed from the right of the N-terminal face. Two monomers are omitted to show the internal surface of the helicase channel. C) Enlargement of the ssDNA (gray) within the helicase channel.

Table 1. A comparison of the common features of chromosomal DNA replication in *E. coli*, yeast/human, and euryarchaeota, with bacterial or bacterial-like features shown in green, eukaryotic or eukaryotic-like features in blue, and archaeal-specific factors in red.[a]

	E. coli	Yeast/human	Euryarchaea
Chromosome	Circular	Linear	Circular
Replication origin	Single	Multiple	Single or Multiple
Pre-replication complex			
Origin recognition	DnaA (1)	ORC (6)	Cdc6 (Orc1)[b] (≥1)
Helicase	DnaB[c] (1)	MCM (6)	MCM (1)
Helicase loader	DnaC[c] (1)	ORC (6) and Cdc6 (1)	Cdc6 (Orc1)[b] (≥1)
Pre-initiation complex			
Cdc45	-	Cdc45 (1)	GAN (Cdc45, RecJ) (1)
GINS	-	GINS (4)	GINS (1-2)
CMG/GMG complex[d]	-	+	+
Single-stranded DNA binding protein	SSB (1)	RPA (3)	RPA (1-3)
Replisome assembly			
Primase	DnaG (1)	Polα/Primase[e,f] (4)	Primase (2)
Sliding clamp	β-clamp (1)	PCNA (1)	PCNA (1)
Clamp loader	τ-complex (5)	RFC (5)	RFC (2)
DNA polymerase			
Leading strand	PolC (3)	Polε[f] (4)	PolB[g] (1) and/or PolD (2)
Lagging strand	PolC (3)	Polδ[f] (4)	PolB[g] (1) and/or PolD (2)
Okazaki fragment maturation			
Primer removal	PolI (1)	Fen1 (1) and Dna2 (1)	Fen1 (1)
Gap filling	PolI (1)	Polδ (4)	PolB/PolD (1 / 2)
Ligation	NAD^+-dependent (1)	ATP-dependent (1)	ATP-dependent[h] (1)

a. The number of different proteins forming the active unit are shown in parentheses. The comparison includes the Euryarchaea as representative archaea. There are many lineages and kingdoms, each with a slightly different set of replication proteins.

b. The genomes of species belonging to Methanococcales and Methanopyrales do not contain genes encoding for Cdc6 (Orc1) homologues.

c. In bacteria the helicase and helicase loader are not considered to be part of the pre-RC but rather the pre-IC. As this paper is about archaea, these proteins were included under pre-RC.

d. The archaeal CMG complex is also called GMG (GAN, MCM, GINS).

e. Polα/Primase is a complex of four subunits that includes polymerase and primase activity.

f. All three replicative polymerases in eukarya (Polα, Polε and Polδ) belong to family B.

g. In some archaeal species PolB is not essential for cell viability.

h. Most archaeal ligases use ATP, but some use NAD^+ as a co-factor.

Table 2. Comparison of the replicative helicases from the three domains of life.

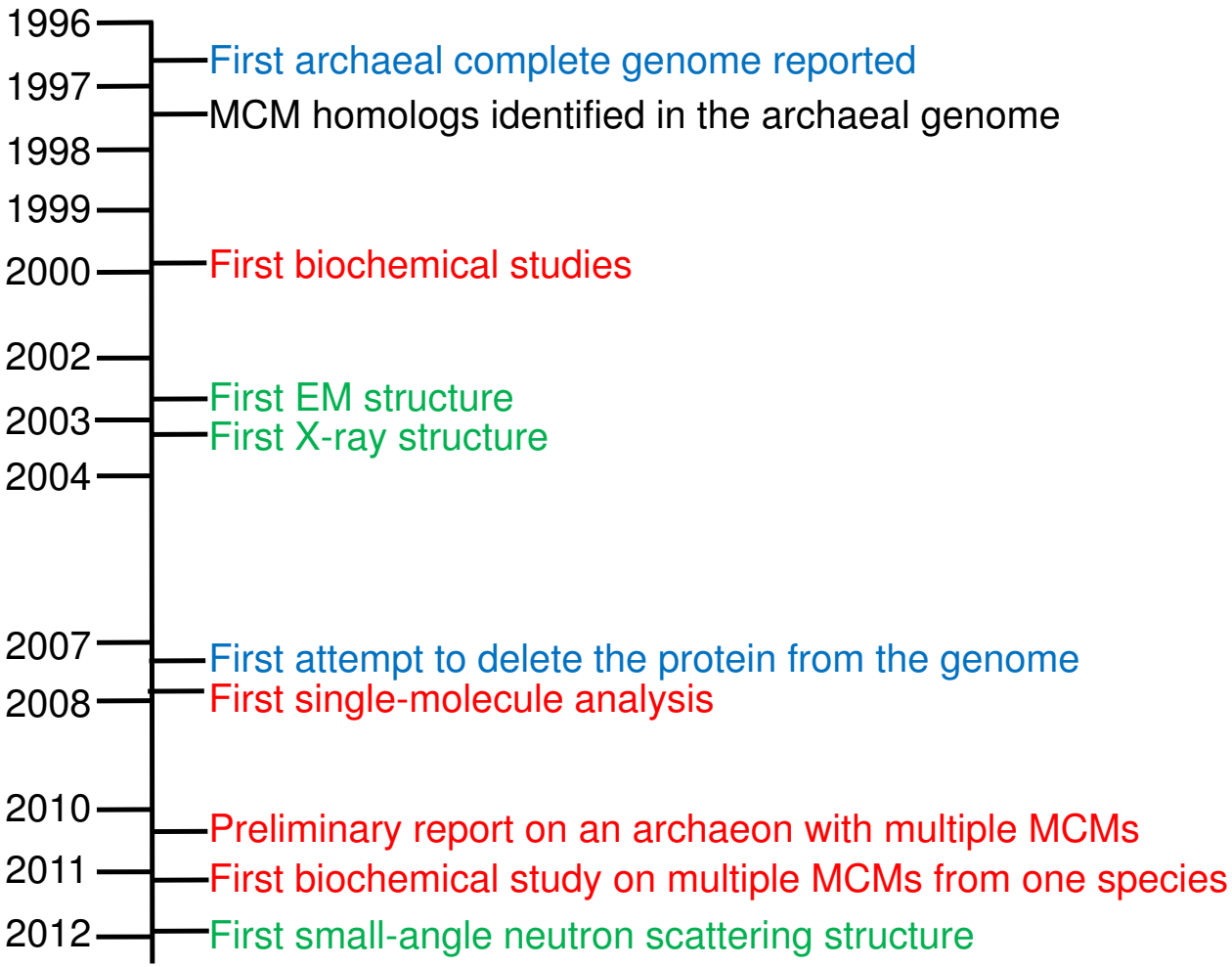
	Bacteria	Eukarya	Archaea
Protein(s)	DnaB	MCM2-7	MCM
Essential for viability?	Yes	Yes	Yes
Oligomeric structure	Homo-hexamer	Hetero-hexamer	Homo-dodecamer
Direction of translocation on ssDNA	5'-to-3'	3'-to-5'	3'-to-5'
Additional factors required for activity *in vitro*	None	Cdc45 and GINS[a]	None[b]
In vitro processivity (bp)			
Alone	400	0[c]	4,500
Replication complex	86,000	500	nd[d]
Bind to ssDNA and dsDNA?	Yes	Yes	Yes
Translocate on ssDNA and dsDNA?	Yes	Yes	Yes
Unwind DNA-RNA hybrid?	Yes	Yes	Yes

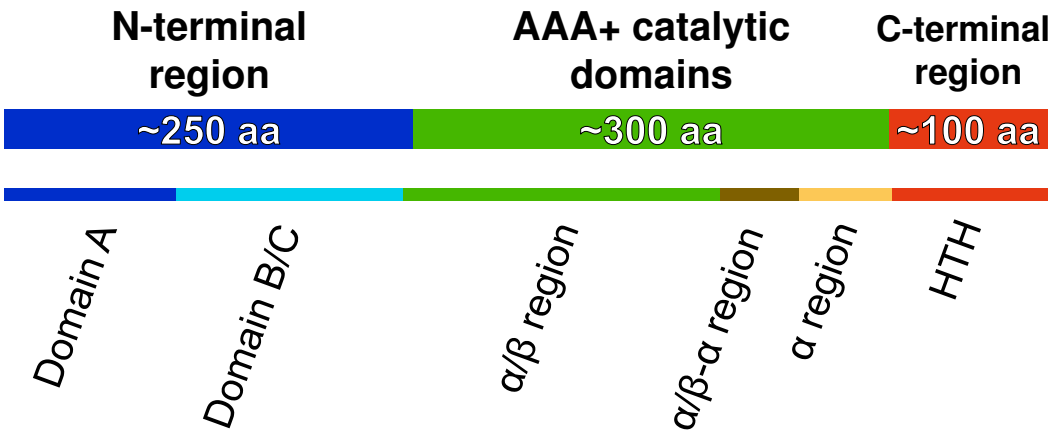
a. Under some conditions the eukaryotic MCM possess *in vitro* activity on its own.

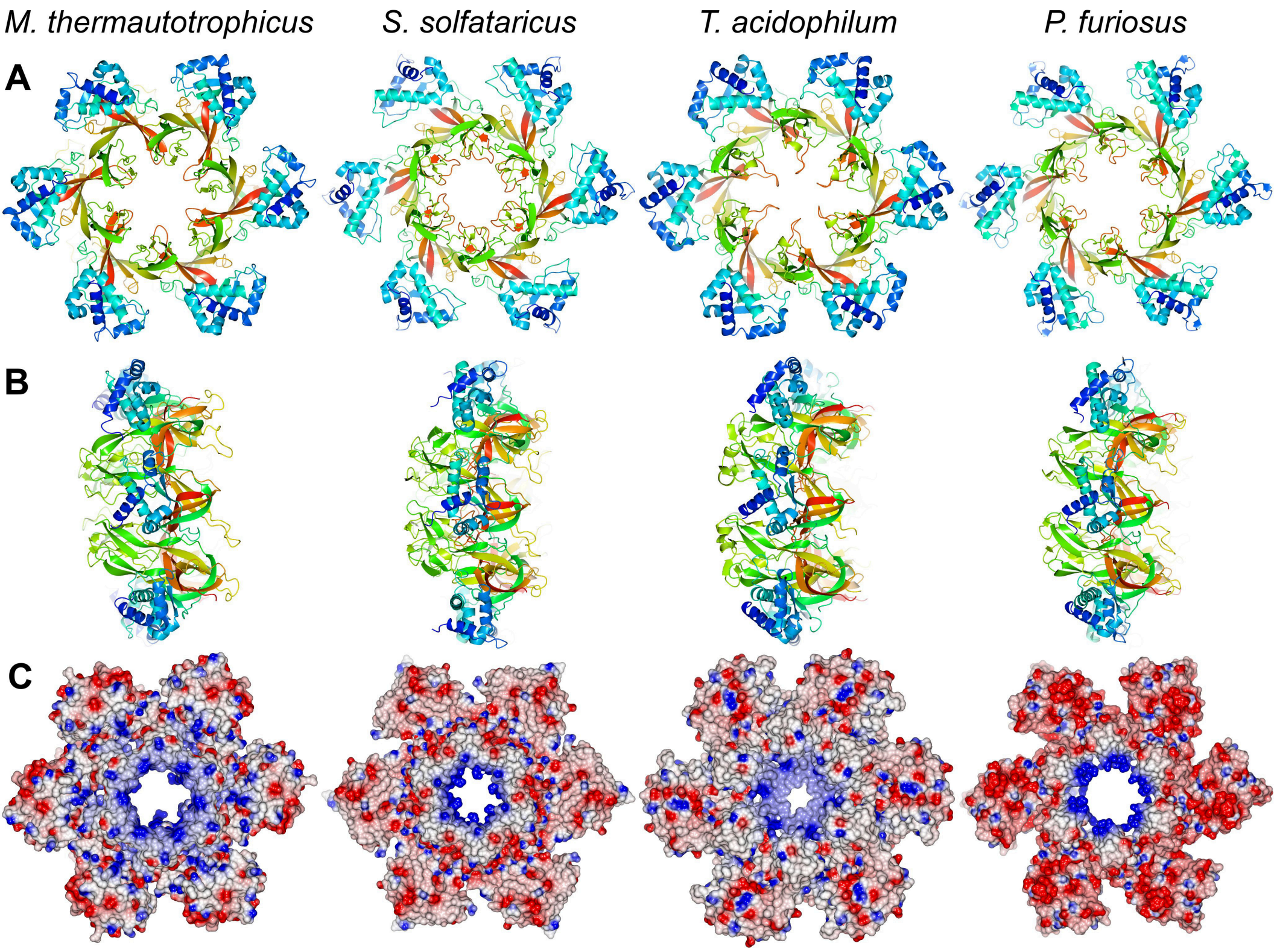
b. In most species.

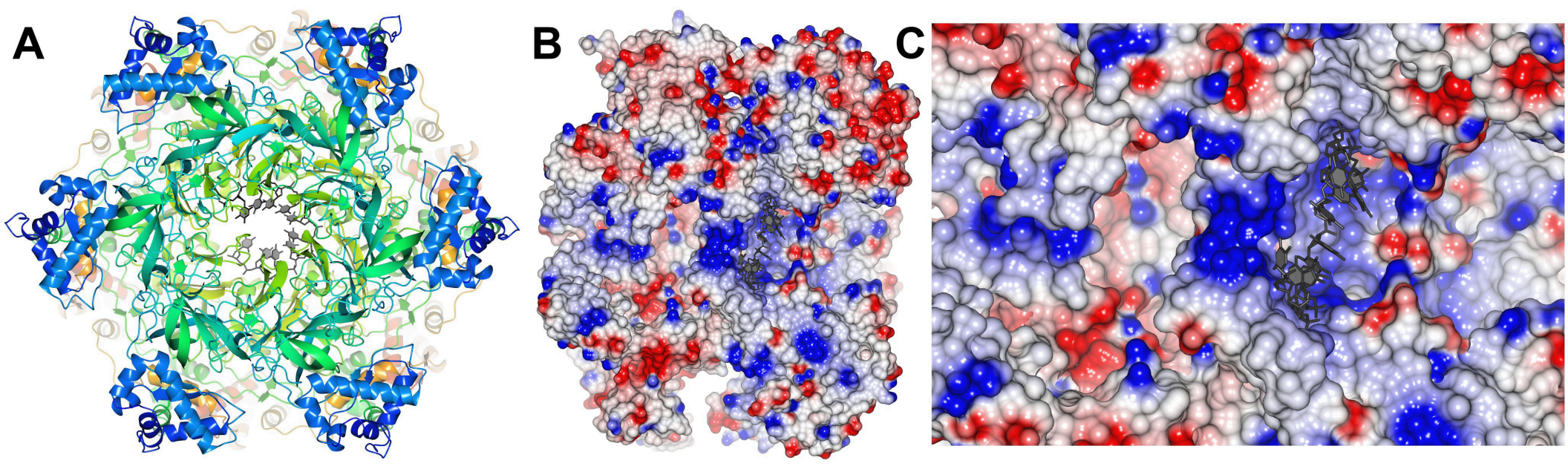
c. For the MCM2-7 complex.

d. Not determined.


1996
First archaeal complete genome reported
1997
MCM homologs identified in the archaeal genome
1998
1999
2000
First biochemical studies
2002
First EM structure
2003
First X-ray structure
2004
2007
First attempt to delete the protein from the genome
2008
First single-molecule analysis
2010
Preliminary report on an archaeon with multiple MCMs
2011
First biochemical study on multiple MCMs from one species
2012
First small-angle neutron scattering structure


N-terminal region
AAA+ catalytic domains
C-terminal region
~250 aa
~300 aa
~100 aa
Domain A
Domain B/C
α/β region
α/β-α region
α region
HTH


M. thermautotrophicus
S. solfataricus
T. acidophilum
P. furiosus
A
B
C


A
B
C