

Insights into DNA replication from the third domain of life

Bik K. Tye*

Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853-2703

The discovery of the Archaea (1), the third domain of life, has reinvigorated efforts to understand the phylogenetic relationships among all forms of life. It is therefore not surprising that the recent completion of four archaeal genomes (2–5) created a flurry of excitement among biologists (*Genetics*, Vol. 152). A comparison of genome sequences reveals that the prokaryotes, Archaea and Bacteria, have closer similarities in their metabolic processes, but Archaea and Eukarya have more similarities in their information processing machineries (DNA replication, transcription, and translation) (6, 7). If this initial observation were supported by further biochemical analyses, then it would suggest that Archaea and Bacteria diverged before the evolution of mechanisms that substantially increased the fidelity of information transfer. That pressure for accuracy may have increased as genomes became more and more complex. Furthermore, it would suggest Archaea and Eukarya share a common ancestor more recently than either shares with Bacteria (Fig. 1). This observation has drawn intense interests from biochemists who have long been battling with the complexities of eukaryal biology. If Archaea embody a more primitive form of the highly evolved Eukarya, then simplified models for complex systems of Eukarya may be found in Archaea. Indeed, a composite of the putative replication proteins identified from four archaeal genomes supports this view (Table 1) (7). This rationale provided the motivation for the biochemical characterization of the *Methanobacterium thermoautotrophicum* MCM protein.

MthMCM shows similarities to a family of six related minichromosome maintenance proteins, Mcm2-Mcm7, that are conserved from yeast to human and are known for their central role in the highly regulated replication initiation process (8). These six proteins have been postulated to form a hexameric ring complex (MCM complex) that serves as the DNA helicase to unwind DNA at replication origins as well as at replication forks. However, although inferences for bio-

chemical activities for the eukaryal MCMs are strong, evidence for such activities has been weak. In two recent issues of PNAS, two groups (9, 10) independently reported the exciting finding that purified recombinant *MthMCM* protein forms a double hexamer that acts as a 3'-5' DNA helicase as postulated for the eukaryal MCM proteins.

The genomes of eukaryotes and prokaryotes differ in size and organization. The generally larger genomes of eukaryotes are organized into multiple chromosomes, and initiation of DNA synthesis starts from not one but numerous sites on each chromosome. Compounding the complexity of this replication process is the necessity to coordinate these multiple initiation events such that on completion, the entire genome is duplicated exactly once every cell cycle. The strategy for this regulated process is beginning to emerge from the collective works of many laboratories. Key to this strategy is the periodic recruitment and discharge of the MCM complex at replication origins to forge a cycle of activity and inactivity at replication origins (8). Although this concept of the temporal separation of an active and inactive chromatin state at replication origins involving the MCM complex is relatively simple, the number of proteins involved to achieve this goal suggests a complex and intricate scheme. Chief among these proteins is the origin recognition complex (ORC), a complex of six nonidentical subunits, Orc1-Orc6, which binds an essential element at replication origins to act as a "bookmark" for these sites (11). During the G₁ phase, the MCM complex is delivered to replication origins with the aid of Cdc6 (12, 13), a short-lived protein that also has some sequence similarities to Orc1. Initiation of DNA synthesis occurs after a rapid succession of events involving additional proteins (including Cdc45) cued by the cell cycle-dependent protein kinases (Cdc7-Dbf4 and Cdc28-Clb) resulting in the melting of DNA and the recruitment of elongation machineries to replication origins (14, 15). As replication transitions from initiation to elongation, the MCM complex is

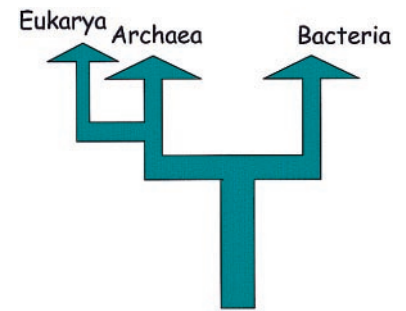


Fig. 1. The three domains of life.

believed to change its alliance with the initiation complex to the elongation complex, where it acts as the processive helicase of the growing fork (16), much like the large T antigen of simian virus SV40 (17). The MCM-vacated replication origins now assume an inactive state that no longer supports the initiation of DNA synthesis. This sequence of events is not to repeat again until the next G₁ phase.

For the purpose of this commentary, the strategy of the eukaryal cell used to restrict DNA replication to one round per cell cycle can be simply viewed as a concerted effort to regulate the many activities of the MCM complex. Although this concept has gained popularity as a working hypothesis for those who study the MCM proteins, much of the information that generated this picture was derived from genetic or *in vivo* experiments without the support of biochemical evidence. Because of the number of protein factors involved, the prospect to study the activities of the MCM proteins in the context of the many accessory proteins in an *in vitro* assay for replication initiation is daunting. The existence of a single MCM homolog in *M. thermoautotrophicum* in contrast to the six MCM proteins found in all eukaryotes is welcome news. The six eukaryal MCM proteins, ranging in size from 776 to 1,017 residues, contain three

See companion articles on page 14783 in issue 26 of volume 96 and page 1530 in issue 4 of volume 97.

*To whom reprint requests should be addressed. E-mail: bt16@cornell.edu.

Table 1. Replication proteins of Eukarya, Bacteria, and Archaea

Function	Archaeal	Eukaryal	Bacterial
Origin recognition	Orc1/Cdc6-like proteins	Origin recognition complex (ORC) six proteins	DnaA
Helicase	Dna2-like, MCM-like	Dna2, MCMs-six proteins	DnaB
Single-stranded DNA binding	RPA-like protein	Replication protein A (RPA, three subunits)	SSB
Primer synthesis	Eukaryotic-like primase	DNA Pol α	DnaG
Clamp loader	RFC-like proteins (small, large)	RFC	γ -complex
Clamp (elongation factor)	PCNA-like proteins	PCNA	Pol III β subunit
DNA strand synthesis	Family B DNA Pol Family D DNA Pol	DNA Pol α, δ, ϵ	Pol III
DNA strand ligation (on lagging strand)	DNA ligase (ATP-dependent)	DNA ligase (ATP-dependent)	DNA ligase (NAD-dependent)
Removal of primers	FEN1, RnaseH	FEN1, RnaseH	Pol I, Rnase H

Adapted from ref. 7.

highly conserved regions, including a NTP binding motif (Fig. 2). The *Mth*MCM protein, although slightly smaller (667 amino acids), is also conserved in those regions (Fig. 2). All six MCM proteins are known to be essential for the initiation of DNA synthesis and can be purified in a hexameric complex that has a globular structure (18). However, to date, biochemical activities associated with this hexameric complex or any one of the six proteins have yet to be reported. The only report of enzymatic activities associated with the MCM proteins is a complex that contains three of the six MCMs (Mcm4, Mcm6, and Mcm7) formed during the course of purification of the larger complex in HeLa cells (19, 20). Weak helicase, ATPase, and single-strand DNA binding activities were found associated with this complex. These results suggest that the MCM proteins have an intrinsic helicase activity that can only be activated when assembled in a specific conformation. The search for the physiologically active form of the MCM complex becomes the key to understanding the regulation of eukaryote DNA replication (8).

In the papers reported recently in PNAS, two groups purified and characterized the recombinant *Mth*MCM protein and obtained similar results (9, 10). In a purified preparation of *Mth*MCM protein, $\approx 80\%$ of the protein assembled in an oligomeric complex of about 850–950 kDa, corresponding to a dodecamer, and $\approx 20\%$ existed as a monomer of about 75 kDa. The dodecameric complex examined by scanning transmission electron microscopy appears to form a double hexamer with a ring structure (10) characteristic of proteins or enzymes that travel large distances along DNA such as the processivity factors (21) and helicases (22). When assayed for helicase and helicase associated activities, both forms exhibited single-strand DNA binding, ssDNA-stimulated ATPase, and ATP-dependent 3'-5' helicase activities. The double hexamer was further tested for processivity of its helicase activity and was shown to be able to unwind DNA of up to 500 base pairs (10). Mutational analysis indicated that the N-terminal 110 amino acids as well as the putative Walker A motif of the conserved NTP binding site were dispensable for the

ssDNA binding activity but essential for the ATPase as well as the helicase activities. The robust helicase activity is consistent with a role for the MCM complex in replication elongation as well as initiation as postulated for the eukaryal MCM complex.

Studying DNA replication in Archaea, the third domain of life, is important and interesting both in its own right and from an evolutionary standpoint. A major difference between Bacteria and Eukarya in replicating their genomes is the use of a single versus multiple replication origins. Where is Archaea in this evolutionary process? Is the evolution of a more complex eukaryal helicase containing six different MCMs a consequence of the expansion of genome sizes as well as the number of replication origins? The revelation that Archaea may have many of the basic components (Table 1) of the replication initiation machinery of Eukarya offers new opportunities for unraveling the complexities of eukaryal systems not offered by viral models such as the SV40 (17). The viral models provide their own initiator proteins that usurp the host rep-

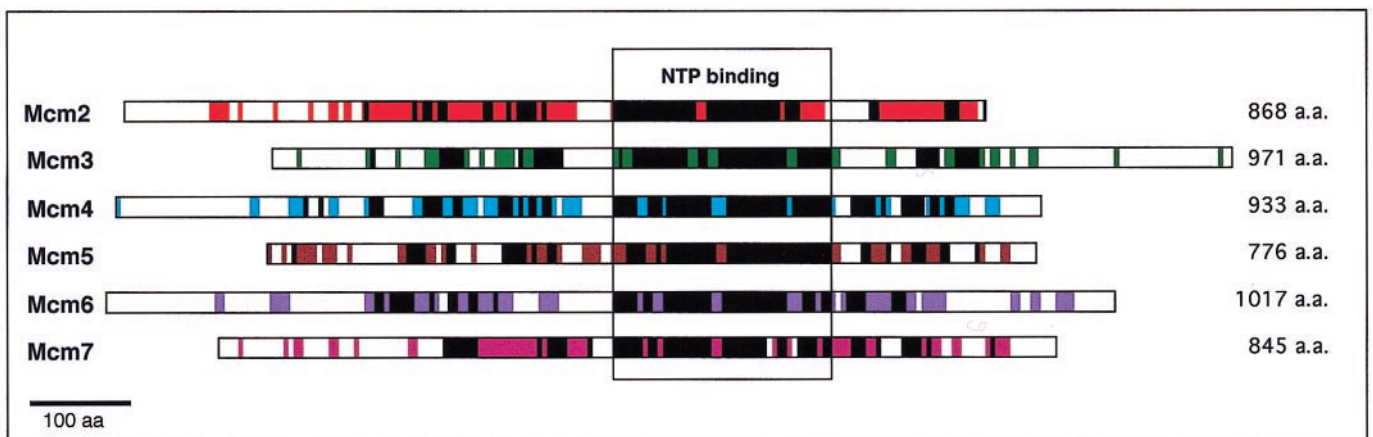


Fig. 2. Sequence conservation among *Saccharomyces cerevisiae* MCM proteins. Black bars represent regions conserved between *S. cerevisiae* MCMs and the single MCM protein of *M. thermoautotrophicum*, and colored bars represent regions conserved between yeast and mammalian MCMs of the same class. The largest conserved domain contains the nucleotide binding motif. [Reproduced with permission from Kearsley and Labib (27) (Copyright 1998, Elsevier).]

lication elongation machineries and therefore are useful for studying replication elongation as well as understanding the principles, although not the mechanistics, of eukaryal replication initiation.

The study of the *Mth*MCM suggests that Achaea as a model may provide useful information at two levels. The self-assembly of a single *Mth*MCM protein into an active double hexamer is a dramatic simplification of the Eukaryal MCM studies. It allows one to explore the complete repertoire of enzymatic activities associated with the MCM complex in isolation. Furthermore, setting aside the intricacies of assembling six MCMs into an active complex, it allows one to focus on the interaction of this complex with other components of the replication initiation complex, such as ORC, Cdc6, and RPA. Indeed, the existence of two Cdc6/Orc1-like proteins in *M. thermoautotrophicus*

instead of the six ORC subunits suggests significant simplification in the origin recognition function and bodes well for an *in vitro* system that is likely to have some resemblance to the eukaryal system.

There are also some important omissions in the *Mt* genome that suggest major differences between Eukarya and Archaea. Cdc45, an important factor in the transition of the initiation complex to elongation complex, and Cdc7-Dbf4 and Cdc28-Clb, the cell cycle-dependent protein kinases, all appear to be absent in the *Mt* genome. These omissions suggest a less elaborate regulatory mechanism for replication initiation in Archaea, although the existence of functional *Mt* homologs cannot be ruled out. Clearly, much work is needed for studies in both the archaeal and the eukaryal systems to begin to unravel the complexities of the regulation of DNA replication of eukaryotes that is

fundamental to our understanding of the control of cell proliferation. It is worth mentioning here that because expression of the MCMs is tightly coupled to the regulation of cell proliferation (23–25), this property of the MCMs is currently being exploited as a diagnostic marker for neoplasticity in precancerous cells (26). It is therefore not a stretch of the imagination to look for answers in the ancient microbe for cures of human diseases. The reports on the *Mth*MCM underscore the value of studying fundamental biological processes in diverse organisms and the prospect that these studies will reveal the underlying principles of life.

I thank Chip Aquadro, Jeff Roberts, and Sara Sawyer for critical reading of this commentary. I thank Tom Fox for discussions. Research from this laboratory is supported by a grant (GM34190) from the National Institutes of Health.

1. Woese, C. R. & Fox, G. E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5088–5090.
2. Bult, C. J., White, O., Olsen, G. J., Zhou, L. X., Fleischmann, R. D., Sutton, G. G., Blake, J. A., Fitzgerald, L. M., Clayton, R. A., Gocayne, J. D., et al. (1996) *Science* **273**, 1058–1073.
3. Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., et al. (1997) *Nature (London)* **390**, 364–370.
4. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H. M., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., et al. (1997) *J. Bacteriol.* **179**, 7135–7155.
5. Kawarabayasi, Y., Sawada, M., Horikawa, H., Kaikawa, Y., & Hino, Y. (1998) *DNA Res.* **5**, 55–76.
6. Whitman, W. B., Pfeifer, F., Blum, P. & Klein, A. (1999) *Genetics* **152**, 1245–1248.
7. Caan, I. K. O. & Ishino, Y. (1999) *Genetics* **152**, 1249–1267.
8. Tye, B. K. (1999) *Annu. Rev. Biochem.* **68**, 649–686.
9. Kelman, Z., Lee, J.-K. & Hurwitz, J. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 14783–14788.
10. Chong, J., Hayashi, M. K., Simon, M. N., Xu, R.-M. & Stillman, B. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 1530–1535.
11. Bell, S. & Stillman, B. (1992) *Nature (London)* **357**, 128–134.
12. Cocker, J. H., Piatti, S., Santocanale, C., Nasmyth, K. & Diffley, J. F. X. (1996) *Nature (London)* **379**, 180–182.
13. Coleman, T. R., Carpenter, P. B. & Dunphy, W. G. (1996) *Cell* **87**, 53–63.
14. Lei, M., Kawasaki, Y., Young, M. R., Kihara, M., Sugino, A. & Tye, B. K. (1997) *Genes Dev.* **11**, 3365–3374.
15. Zou, L. & Stillman, B. (1998) *Science* **280**, 593–596.
16. Aparicio, O. M., Weinstein, D. M. & Bell, S. P. (1997) *Cell* **91**, 59–69.
17. Li, J. J. & Kelly, T. J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6973–6977.
18. Adachi, Y., Usukura, J. & Yanagida, M. (1997) *Genes Cells* **2**, 467–479.
19. Ishimi, Y. (1997) *J. Biol. Chem.* **272**, 24508–24513.
20. You, Z., Komamura, Y. & Ishimi, Y. (1999) *Mol. Cell. Biol.* **19**, 8003–8015.
21. Kong, X.-P., Onrust, R., O'Donnell, M. & Kuriyan, J. (1992) *Cell* **69**, 425–437.
22. Yu, X. & Egelman, E. H. (1997) *Nat. Struct. Biol.* **4**, 101–104.
23. Springer, P. S., McCombe, W. R., Sundaresan, V. & Martienssen, R. A. (1995) *Science* **268**, 877–880.
24. Treisman, J. E., Follette, P. J., O'Farrell, P. H. & Rubin, G. M. (1995) *Genes Dev.* **9**, 1709–1715.
25. Young, M. & Tye, B. K. (1997) *Mol. Biol. Cell* **8**, 1587–1601.
26. Williams, G. H., Romanowski, P., Morris, L., Madine, M., Mills, A. D., Stoeber, K., Marr, J., Laskey, R. A. & Coleman, N. (1999) *Proc. Natl. Acad. Sci. USA* **95**, 14932–14937.
27. Kearsley, S. E. & Labib, K. (1998) *Biochim. Biophys. Acta* **1398**, 113–136.