The archaeal DNA replication machinery: past, present and future

Sonoko Ishino¹, Lori M. Kelman², Zvi Kelman³ and Yoshizumi Ishino¹*

¹Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, and Faculty of Agriculture, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan
²Program in Biotechnology, Montgomery College, 20200 Observation Drive, Germantown, MD 20876, USA
³National Institute of Standards and Technology and the Institute for Bioscience and Biotechnology Research, 9600 Gudelsky Drive, Rockville, MD 20850, USA

(Received 24 January 2014, accepted 4 March 2014)

Living organisms are divided into three domains: Bacteria, Archaea, and Eukarya. Whereas Bacteria and Archaea are both prokaryotes, proteins involved in information processes; replication, transcription, and translation, are more similar in Archaea and Eukarya. Here the history of the research on archaeal DNA replication is summarized and the future of the field is discussed.

Key words: DNA polymerase, extremophiles, molecular mechanism, three domains of life

INTRODUCTION

Archaea is one of the three domains of life. Archaeal organisms have been found in all habitats, including extreme environments such as strict anaerobic conditions, high temperature, and high salinity. The archaeal domain can be divided into five Phyla: Aigarchaeota, Crenarchaeota, Euryarchaeota, Korarchaeota, and Thaumarchaeota.

DNA replication plays an essential role in all life forms. The study of archaeal DNA replication initiated shortly after the recognition of Archaea as the third domain of life. However, at that time, studies focused mainly on the isolation and characterization of thermostable DNA polymerases. Major advances in the field initiated about 15 years ago when the first complete genome of an archaeon was determined (Bult et al., 1996). As it became apparent that the archaeal replication machinery is similar to that in Eukarya, a concentrated effort toward elucidating the archaeal replication machinery began in earnest. Early studies focused on the biochemical and structural characterization of the enzymes and factors needed for the initiation and elongation phases of DNA replication. Genomic studies identified homologs of bacterial and eukaryal replication factors. In the past few years, genetic tools have enabled functional analysis of the replication proteins in vivo and the identification of new replication enzymes. Here a perspective on the field of archaeal DNA replication, past, present and future, is described. Due to space limitations only key papers and reviews are included.

THE BIOCHEMISTRY ERA

The study of archaeal DNA replication started in the 1980s (Fig. 1). Early on it became apparent that archaeal replication is different from that in Bacteria because aphidicolin, a specific inhibitor of eukaryotic DNA replication, also inhibits DNA synthesis and cell growth of haloarchaea and methanogenic archaea (Forterre et al., 1984; Zabel et al., 1985). These early studies suggested that DNA polymerases in Archaea may be similar to the eukaryotic replicative polymerases [DNA polymerase (Pol) α, δ, and ε]. After these initial studies, however, the focus shifted to the isolation and characterization of DNA polymerases from hyperthermophilic archaea for their commercial value in polymerase chain reaction (PCR). These early studies utilized polymerase activity as a tool to isolate the enzymes from cell extracts followed by cloning, expression and purification from Escherichia coli cells (Perler et al., 1996). These enzymes all belonged to family B DNA polymerases (PolB), as do the eukaryotic Pol α, δ, and ε. It was therefore suggested that PolB homologs are the replicative enzymes in Archaea. This is true in some archaeal species but not in others (discussed below). The studies of the replication machinery based on the knowledge of the biochemical properties of the enzymes was limited to DNA polymerase and a few other enzymes [e.g. topoisomerase (Forterre and Elie, 1993)] with well-defined activities that could be easily followed during purification.
THE GENOMICS ERA

The field of archaeal DNA replication experienced an expansion following the completion of the genomic sequence of *Methanocaldococcus jannaschii* (Bult et al., 1996). The analysis of this and subsequent genomes suggested that the archaeal replication proteins are similar to those found in Eukarya, but with lower complexity (Edgell and Doolittle, 1997) (Fig. 2). These *in silico* observations led to an influx of biochemical and structural studies on the replication machinery of archaea (Grabowski and Kelman, 2003). The genomic sequences, together with biochemical studies, also enabled the identification of differences in the replication apparatuses among the species and kingdoms. The current knowledge is summarized in a recent review article (Ishino and Ishino, 2012).

The genome sequences were also instrumental in identifying the archaeal origins of replication (*oriC*) (Fig. 2). The first report on the archaeal *oriC* came from a skew analysis of genome sequences (Lopez et al., 1999). The *in silico* analyses were later followed by *in vivo* studies that demonstrated that archaea contain a single origin on a circular genome, as in Bacteria. It was later shown that *Sulfolobus* has three origins, and multiple origins have also been reported in other species (Kelman and Kelman, 2004). Studies on the termination of DNA replication are also in progress. Most archaeal organisms have a single homolog of bacterial XerC/D proteins that is the site-specific recombinase involved in resolution of replicated chromosomes. The *dif*-like sequences are also predicted to be the replication termination region of *Thermococcal* DNA (Cortez et al., 2010), but outside of the fork fusion zone (ffz) of the multiple origins of the *Sulfolobus* DNA (Duggin et al., 2011). However, the mechanism of fork fusion and chromosomal resolution in archaea is not yet known.

One of the first proteins identified and studied based on the genome sequences was the ssDNA binding protein, replication protein A (RPA), from *M. jannaschii* (Kelly et al., 1998). Subsequent studies focused on the polymerase and its accessory factors, proliferating cell nuclear antigen (PCNA), replication factor C (RFC), and proteins involved in lagging strand processing such as DNA ligase and Fen1. All of these factors exhibited biochemical properties similar to their eukaryotic counterparts (Ishino and Ishino, 2012).

Some replication proteins were not readily identified by similarity to bacterial or eukaryotic counterparts. For example, archaeal DNA primase was not identified by strong sequence similarity. However, a candidate protein with limited similarity to the catalytic subunit of the eukaryotic primase was identified (Desogus et al., 1999). In some archaea, the gene encoding the putative catalytic subunit of primase is in an operon with another gene, and careful examination of the protein encoded by that gene revealed weak similarity to the non-catalytic subunit of the eukaryotic primase (Desogus et al., 1999). In some archaea, the gene encoding the putative catalytic subunit of primase is in an operon with another gene, and careful examination of the protein encoded by that gene revealed weak similarity to the non-catalytic subunit of the eukaryotic primase. Subsequent studies demonstrated that the complex of the two subunits constitutes the archaeal primase (Liu et al., 2001). Although similar to the eukaryotic enzyme, the archaeal primase exhibits unique features not found in other DNA primases. *In vitro* the enzyme is capable of incorporating both ribo- and deoxyribonucleotides. This observation underscores the idea that although archaeal DNA replication is similar to that of Eukarya, it is not merely a simpler version of the eukaryotic replication machinery.
Although many of the DNA replication proteins were identified and characterized biochemically during the genomic era based on sequence similarity to proteins in Eukarya and Bacteria, biochemical approaches were still important in identifying new replication factors. For example, to identify new DNA polymerases in *Pyrococcus furiosus*, a genomic library was constructed and expressed in *E. coli*. The expressed proteins were screened for polymerase activity at high temperature, and a new dimeric DNA polymerase, designated PolD, was identified. PolD is archaeal-specific and therefore could not be identified using sequence similarity to other DNA polymerases.
Following the characterization of the enzymes and factors involved in the elongation phase, the focus in the field shifted to the study of the proteins needed for the initiation of DNA replication. The first protein to be characterized was the minichromosome maintenance (MCM) helicase (Sakakibara et al., 2009), followed by studies on the initiation protein Cdc6, also referred to as Cdc6/Orc1 (Costa et al., 2013).

Also during this period, structures of most known archaeal replication proteins were determined either at high resolution using X-ray crystallography or at low resolution by electron microscopy (EM), small-angle neutron scattering (SANS), or small-angle X-ray scattering (SAXS). The structures of several protein complexes and protein-DNA complexes were also determined (Brewster and Chen, 2010; Costa et al., 2013; Ishino and Ishino, 2013). Two important structures that have not been solved are those of PolD and the full-length MCM protein.

**THE GENETICS ERA**

Until recently, genetic tools were not readily available in Archaea, but manipulation in several species is now possible (Leigh et al., 2011). These tools opened new avenues in archaeal research. For example, bioinformatic analysis identified two putative primases in Archaea, a two-subunit enzyme with homology to the eukaryotic primase (discussed above) and an enzyme with limited amino acid sequence similarity to the bacterial DnaG protein. Genetic deletion studies showed that only the eukaryotic-like primase is essential for cell viability (Le Breton et al., 2007). Taken together with the biochemical studies, this observation strongly suggests that the eukaryotic-like dimeric enzyme is the archaeal DNA primase. Another example is the cellular role of PolB and PolD. All archaeal species except of those belonging to the Crenarchaeal phylum contain both PolB and PolD. The biochemical properties of both enzymes are consistent with the activities expected of a replicative polymerase including stimulation by PCNA and 3′-5′ exonuclease proofreading activity (Ishino and Ishino, 2012). Thus, it was not clear if both are required for chromosomal replication. Based on in vitro biochemical characterizations including primer usage and strand displacement activity, it has been suggested that PolB may replicate the leading strand while PolD replicates the lagging strand. This would make archaeal organisms similar to eukaryotic organisms, in which PolB replicates the leading strand while Polα replicates the lagging strand. Using gene knockouts, however, it was found that, at least in some species, PolB is not required for cell viability (Cubonova et al., 2013; Sarmiento et al., 2013). These results suggest that only PolD is the replicative enzyme in those species.

Genetic tools were also used to expand the pool of proteins known to participate in the archaeal replication process. While many replication factors have been identified using in silico approaches, it is clear that some factors and enzymes involved in archaeal replication have not yet been identified. In vivo tagging of known replication proteins and the isolation of complexes that co-purified with them resulted in the identification of several putative new replication proteins (Li et al., 2010). For example, this strategy identified an archaeal protein that interacts with the GINS complex (Li et al., 2011). The protein, called GAN (GINS-associated nuclease) shares amino acid sequence and biochemical properties with the bacterial RecJ nuclease. Sequence analysis revealed that GAN, RecJ, and the eukaryotic Cdc45 protein all contain a DHH phosphodiesterase domain, suggesting they are homologs (Sanchez-Pulido and Ponting, 2011; Makarova et al., 2012). It had been long hypothesized that archaea have a functional Cdc45 homolog, and genetic tools enabled its isolation and characterization.

**THE FUTURE**

Many of the proteins needed for archaeal DNA replication have been identified and characterized in the last 15 years. However, there are a few goals in the field that have not been achieved. One is the reconstitution of an in vitro DNA replication system for Archaea. An in vitro DNA replication system has not been established in Archaea despite the efforts of several labs. When developed, such a system will provide a new tool for detailed mechanistic studies on the replication machinery. It will enable the elucidation of the roles played by the different proteins and complexes needed for the initiation and elongation phases, the regulation of the process, the coupling of leading and lagging strand synthesis, etc. It is noteworthy that a partial reconstitution of a rolling circle replication using synthetic DNA and proteins from *T. kodakarensis* was published (Chenmitz Galal et al., 2012). It will be helpful to investigate the elongation process of the archaeal replication in more detail. Furthermore, progress will be made in the structural determination of high-order complexes including proteins and DNA to provide a better understanding of the initiation and elongation phases of DNA replication.

Although progress has been made, we are still far from completely understanding the initiation and termination of DNA replication in archaea and these are two of the main subjects for future research. It will also be important to study the initiation process in species with single and multiple origins and the coordination between initiations at different origins. Recently it was shown that *Haloferax volcanii* can survive without a canonical origin (Hawkins et al., 2013). Further work is needed to show how initiation occurs in those cells, and to determine if
this is halophile-specific or a more widespread phenomenon.

Further accumulated information on DNA replication in different archaeal species will be used to understand the evolutionary relationships among different archaeal phyla and between Archaea and the other domains of life.

This perspective stemmed from discussions during Dr. Z. Kelman’s visit to Dr. Y. Ishino’s laboratory thanks to the generous support of the Japan Society for the Promotion of Science (JSPS).

REFERENCES

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.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., et al. (1996) Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. Science 273, 1056–1073.

Cheminetz Galal, W., Pan, M., Kelman, Z., and Hurwitz, J. (2012) Characterization of DNA primase complex isolated from the archaeon, Thermococcus kodakarense. J. Biol. Chem. 287, 16209–16219.

Cortez, D., Quevillon-Cheruel, S., Gribaldo, S., Desnoues, N., Sezonov, G., Forterre, P., and Serre, M. C. (2010) Evidence for a XerCid system for chromosome resolution in archaea. PLoS Genet. 6, e1001166.

Costa, A., Hood, I. V., and Berger, J. M. (2013) Mechanisms for initiating cellular DNA replication. Annu. Rev. Biochem. 82, 25–54.

Cubonova, L., Richardson, T., Burkhart, B. W., Kelman, Z., Connolly, B. A., Reeve, J. N., and Santangelo, T. J. (2013) Archaeal DNA polymerase D but not DNA polymerase B is required for genome replication in Thermococcus kodakarense. J. Bacteriol. 195, 2322–2328.

Desogus, G., Onesti, S., Brick, P., Rossi, M., and Pisani, F. M. (1999) Identification and characterization of a DNA primase from the hyperthermophilic archaeon Methanococcus jannaschii. Nucleic Acids Res. 27, 4444–4450.

Diggins, I. G., Dubarry, N., and Bell, S. D. (2011) Replication termination and chromosome dimer resolution in the archaeon Sulfolobus solfataricus. EMBO J. 30, 145–153.

Edgell, D. R., and Doolittle, W. F. (1997) Archaea and the origins of DNA replication proteins. Cell 89, 995–998.

Forterre, P., and Elie, C. (1993) Chromosome structure, DNA topoisomerases, and DNA polymerases in archaeabacteria (archaea). In: New Comprehensive Biochemistry, vol. 26. (eds.: M. Kates, D. J. Kushner, and A. T. Matheson), pp. 323–365. Elsevier, Amsterdam.

Forterre, P., Elie, C., and Kohiyama, M. (1984) Aphidicolin inhibits growth and DNA synthesis in halophile archaeabacteria. J. Bacteriol. 159, 800–802.

Grabowski, B., and Kelman, Z. (2003) Archaeal DNA replication: eukaryal proteins in a bacterial context. Annu. Rev. Microbiol. 57, 487–516.

Hawkins, M., Malla, S., Blythe, M. J., Nieduszynski, C. A., and Allers, T. (2013) Accelerated growth in the absence of DNA replication origins. Nature 503, 544–547.

Ishino, S., and Ishino, Y. (2013) DNA polymerases and DNA ligases. In: Thermophilic Microbes in Environmental and Industrial Biotechnology. (eds.: T. Satyanarayana, J. Littlechild, and Y. Kawarabayasi), pp. 429–457. Springer, Heidelberg.

Ishino, Y., and Ishino, S. (2012) Rapid progress of DNA replication studies in Archaea, the third domain of life. Sci. China Life Sci. 55, 386–403.

Kelly, T. J., Simancek, P., and Brush, G. S. (1998) Identification and characterization of a single-stranded DNA-binding protein from the archaeon Methanococcus jannaschii. Proc. Natl. Acad. Sci. USA 95, 14634–14639.

Kelman, L. M., and Kelman, Z. (2004) Multiple origins of replication in archaea. Trends Microbiol. 12, 399–401.

Le Breton, M., Henneke, G., Norais, C., Flament, D., Myllykallio, H., Querellou, J., and Raffin, J. P. (2007) The heterodimeric primase from the euryarchaeon Pyrococcus abyssii: a multifunctional enzyme for initiation and repair? J. Mol. Biol. 374, 1172–1185.

Leigh, J. A., Albers, S. V., Atomi, H., and Allers, T. (2011) Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfohalobiales. FEMS Microbiol. Rev. 35, 577–608.

Li, Z., Santangelo, T. J., Cubonova, L., Reeve, J. N., and Kelman, Z. (2010) Affinity purification of an archaeal DNA replication protein network. MBio 1, e00221-10.

Li, Z., Pan, M., Santangelo, T. J., Cheminetz, W., Yuan, W., Edwards, J. L., Hurwitz, J., Reeve, J. N., and Kelman, Z. (2011) A novel DNA nuclease is stimulated by association with the GINS complex. Nucleic Acids Res. 39, 6114–6123.

Liu, L., Romori, K., Ishino, S., Becquier, A. A., Cann, I. K. O., Kohda, D., and Ishino, Y. (2001) The archaeal DNA primase: biochemical characterization of the p41–p46 complex from Pyrococcus furiosus. J. Biol. Chem. 276, 45484–45490.

Lopez, P., Philippe, H., Myllykallio, H., and Forterre, P. (1999) Identification of putative chromosomal origins of replication in Archaea. Mol. Microbiol. 32, 883–886.

Makarova, K. S., Koonin, E. V., and Kelman, Z. (2012) The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. Biol. Direct 7, 7.

Perler, F. B., Kumar, S., and Kong, H. (1996) Thermostable DNA polymerases. Adv. Protein Chem. 48, 377–435.

Sakakibara, N., Kelman, L. M., and Kelman, Z. (2009) Unwinding the structure and function of the archaeal MCM helicase. Mol. Microbiol. 72, 286–296.

Sanchez-Pulido, L., and Ponting, C. P. (2011) Cdc45: the missing RecJ ortholog in eukaryotes? Bioinformatics 27, 1885–1888.

Sarmiento, F., Mrazek, J., and Whitman, W. B. (2013) Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon Methanococcus maripaludis. Proc. Natl. Acad. Sci. USA 110, 4726–4731.

Silvert, C., Yano, H., Kato, I., Doi, H., and Ishino, Y. (1997) A novel DNA polymerase in the hyperthermophilic archaeon, Pyrococcus furiosus: gene cloning, expression, and characterization. Genes Cells 2, 499–512.

Zabel, H. P., Fischer, H., Holler, E., and Winter, J. (1985) In vivo and in vitro evidence for Eucaryotic t-type DNA-polymerases in methanogens. Purification of the DNA-polymerase of Methanococcus vannielli. Syst. Appl. Microbiol. 6, 111–118.

Perspective on archaeal DNA replication 319