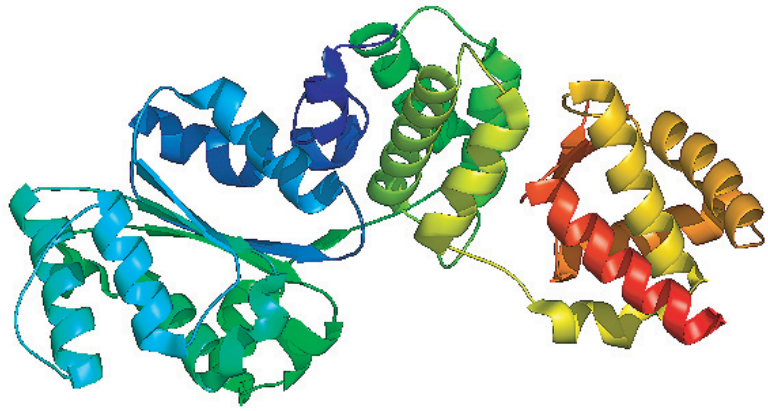
Archaeal DNA Replication: A Robust Model for Eukaryotes

Stephen D. Bell* | Isabelle Dionne | Si-houy Lao-Sirieix | Victoria L. Marsh | Adam McGeoch | Nicholas P. Robinson

Medical Research Council Cancer Cell Unit
Hutchison MRC Research Centre

Corresponding Author:
Medical Research Council Cancer Cell Unit
Hutchison MRC Research Centre
Hills Road, Cambridge, CB2 2XZ, UK

Phone: +44.1223.763311 **Fax:** +44.1223.763296 **E-mail:** sb419@hutchison-mrc.cam.ac.uk



ABSTRACT

The Yellowstone National Park environment plays host to a range of hyperthermophilic ecosystems. Hot sulfur springs are inhabited by a range of bacterial and archaeal microorganisms. Understanding how these organisms replicate their DNA under such apparently inimical conditions is of inherent interest. Furthermore, the archaeal replication system is proving to be a valuable model for the eukaryotic replication apparatus. With the publishing of the first archaeal genome sequences, it became apparent that archaea and eukaryotes have closely related machineries for replication of their genomic DNA. Many of the key players in the eukaryotic apparatus are present, usually in a simplified form in archaea. The relationship between the archaeal and eukaryotic machineries, coupled with the stripped-down nature of the archaeal complexes and the relative tractability of proteins derived from hyperthermophilic archaea, has led a number of laboratories around the world to focus on archaeal DNA replication as a potentially powerful model system to reveal the molecular basis of conserved events in replication. In this review, we will discuss recent advances in our understanding of both initiation and elongation phases of archaeal DNA replication.

Key Words

archaea
Cdc6/Orc1
DNA replication
MCM
origin of replication
PCNA

1.0 INTRODUCTION

The replicon hypothesis of Jacob and colleagues (1963) proposed that key initiator proteins interact with sites within a DNA molecule whereupon they lead to the initiation of DNA replication. Thus, replication initiates at defined locations in a given DNA molecule. There is a fundamental difference in the organisation of replication origins between bacteria and eukaryotes. Typically, bacterial chromosomes are replicated from a single site, usually termed *oriC*. In contrast, eukaryotic chromosomes contain many origins of replication, spaced between 10 kb and 330 kb apart. Very few archaeal origins of replication have been characterised but, until recently, the available data suggested that archaea, like bacteria, may only have a single origin per chromosome. The first archaeal origin to be identified, that of *Pyrococcus abyssi*, was described in an elegant series of studies by Myllykallio, Forterre, and colleagues (Myllykallio et al. 2000). Initial bioinformatics studies had suggested the location of a single origin in *P. abyssi*, and this prediction was confirmed by *in vivo* labelling studies. These initial results were confirmed by subsequent analyses using a two-dimensional gel electrophoresis methodology that resolves replication intermediates (Matsunaga et al. 2001). Finally, Matsunaga and colleagues used a high-resolution technique, termed RIP mapping, to map the position of replication initiation *in vivo* (Matsunaga et al. 2003). Intriguingly, as will be discussed in more detail below, the *P. abyssi* initiation site was found in a non-coding region immediately upstream of the gene encoding the single Orc1/Cdc6 homolog in this species. The fine mapping approach revealed that the initiation point was immediately adjacent to one of two inverted repeats at the origin (Matsunaga et al. 2003).

A more recent study on *Halobacterium* sp. NRC1 used a targeted genetic approach to screen selected genomic regions for DNA elements that could lead to maintenance of a selective marker on a plasmid otherwise incapable of autonomous replication (Berquist and DasSarma 2003). The selectable marker could be maintained either by conferring replication competence to the recipient plasmid, or by promoting high levels of reversible recombination with host chromosomal DNA. Using this approach, a candidate origin was proposed. As was the case with *Pyrococcus*, this was found in a region upstream of a homolog of Orc1/Cdc6. Intriguingly, a bioinformatics analysis had suggested that the main chromosome of *Halobacterium* might actually contain two origins of replication (Zhang and Zhang 2003). However, no second origin has been confirmed by experimental means to date. Thus, to this point, these studies indicated that archaea may, like bacteria, possess a single origin of replication per chromosome.

A recent study, however, has revealed the presence of at least two origins of replication, termed *oriC1* and *oriC2*, in the hyperthermophilic archaeon *Sulfolobus solfataricus* (Robinson et al. 2004). These were localised by a combination of 2D gel analysis and RIP mapping to regions upstream of two of the three Orc1/Cdc6 homologs in *S. solfataricus*. While it has not yet been demonstrated conclusively that both origins are used in every cell in every cell cycle, the RIP mapping result suggested that the majority of replicating chromosomes have leading strand synthesis starting at both origins. This study also revealed the presence of conserved elements within archaeal origins of replication, origin recognition boxes (ORB). Furthermore, ORB elements were found associated with the origins of replication identified in *P. abyssi* and *Halobacterium* (**Figure 1, next page**), organisms in a distinct phylogenetic kingdom from *S. solfataricus* (Robinson et al. 2004). As will be discussed below, these serve as recognition sites for key initiator proteins, archaeal homologs of the eukaryotic Orc1 and Cdc6 proteins.

2.0 ORIGIN RECOGNITION

Consensus sequences, known as DnaA boxes, within a bacterial origin are recognised by the DnaA protein (Messer et al. 2001). In eukaryotes there are many origins within a chromosome that are recognised by initiator proteins (Bell 2002). The eukaryal initiator is the origin recognition complex, ORC (Bell 2002). This is a heterohexameric assembly with three of the subunits capable of binding ATP. In some species, such as *Saccharomyces cerevisiae*, ORC binds in an ATP-dependent manner to readily recognised consensus sequences found within origins. In higher

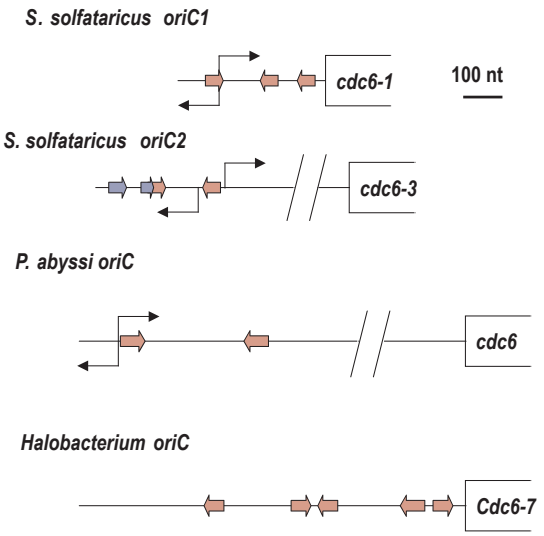


Figure 1. Diagram of the organisation of origins of replication in three archaeal species (Berquist and DasSarma 2003; Matsunaga et al. 2003; Robinson et al. 2004). ORB elements, binding sites for orthologs of *S. solfataricus* Cdc6-1 protein, are shown as pink arrows; binding sites for *S. solfataricus* Cdc6-3 are shown as blue arrows. Black arrows indicate transition points between leading and lagging strand synthesis identified by RIP mapping. The transition point for the *Halobacterium oriC* has not been mapped.

eukaryotes, no consensus sequences for origins have been described. However, ORC still binds specifically to origin-containing DNA (Bell 2002). Whether origin identity in higher eukaryotes is defined purely at the sequence level or by a combination of sequence and chromatin structure is not yet fully resolved (Schaarschmidt et al. 2003). Once origins are bound by ORC, the eukaryotic Cdc6 protein is recruited and is, along with other proteins, involved in recruiting and loading the presumptive replicative helicase, the MCM complex, onto origins (Bell and Dutta 2002).

Examination of archaeal genome sequences has failed to reveal archaeal homologs of bacterial DnaA. In contrast, with the notable exception of *Methanococcus jannaschii*, all archaeal genomes do encode at least one gene with homology to the eukaryotic Orc1 component of ORC (Kelman and Kelman 2003). As the archaeal proteins also show homology to another key eukaryotic initiator protein, Cdc6, we shall refer to the archaeal proteins as Orc1/Cdc6. The crystal structure of the single Orc1/Cdc6 homolog of *Pyrobaculum aerophilum* has been solved (**Figure 2**) and reveals the presence of an N-terminal AAA+, ATP-binding fold, and a C-terminal winged helix-turn-helix (wHTH) candidate DNA binding fold (Liu et al. 2000). Archaeal Orc1/Cdc6s have been proposed to play roles in origin recognition and have been shown to have sequence non-specific DNA-binding activity. Additionally, chromatin immunoprecipitation experiments have indicated that the *P. abyssi* Orc1/Cdc6 protein is associated with the origin *in vivo* (Matsunaga et al. 2001). Intriguingly, the archaeal proteins have been shown to possess a weak DNA-regulated auto-phosphorylation activity, although the physiological relevance of this is not yet clear (Grabowski and Kelman 2001). Another, initially puzzling feature of archaeal Orc1/Cdc6s is that many species encode multiple Orc1/Cdc6s, raising the possibility that these proteins may play roles in recognising distinct origin sequences, have distinct Orc1- or Cdc6-like roles, or have regulatory roles.

Recent work on the three Orc1/Cdc6 homologs of *S. solfataricus*, annotated by the genome sequencing project as Cdc6-1, Cdc6-2, and Cdc6-3 (She et al. 2001)—but homologous to both Orc1 and Cdc6—has proposed

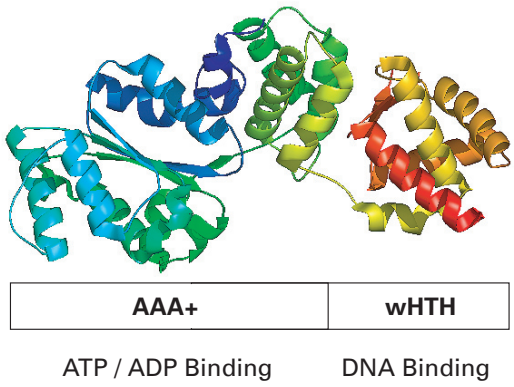


Figure 2. Domain organisation of archaeal Cdc6/Orc1. The structure of *Pyrobaculum aerophilum* Cdc6/Orc1 is shown (Liu et al. 2000) and ADP/ATP and DNA binding folds indicated. The figure was generated from PDB co-ordinates 1FNN using the prgram Pymol (available at www.pymol.org).

a molecular basis for the function of these proteins in origin recognition (Robinson et al. 2004). A phylogenetic analysis has indicated that three principal groupings of archaeal Orc1/Cdc6s exist (Berquist and DasSarma 2003). Intriguingly, *S. solfataricus* Cdc6-1 and Cdc6-3 fall into one clade, while Cdc6-2 falls in another. Furthermore, all archaea with more than one Orc1/Cdc6 have at least one member in both of these groupings. Cdc6-1 is the most highly conserved of the *S. solfataricus* Orc1/Cdc6s. This protein was found to bind in a sequence-specific manner to the ORB elements present in *S. solfataricus oriC1* and to related elements in *oriC2*. Cdc6-3 bound to elements adjacent to the Cdc6-1 binding sites at *oriC2*, but did not bind to *oriC1*. Finally, Cdc6-2 bound to sites overlapping the Cdc6-1 and Cdc6-3 sites at *oriC1* and *oriC2* respectively.

A possible explanation for this complex set of interactions came with the observation that the three Orc1/Cdc6s showed distinct patterns of temporal regulation during the cell cycle. Specifically, Cdc6-1 and Cdc6-3 were expressed in cells prior to and during replication of DNA while Cdc6-2 was expressed in post-replicative cells (Robinson et al. 2004). This suggests, therefore, that Cdc6-1 and Cdc6-3 may have roles promoting replication, whereas Cdc6-2 may act as an inhibitor, preventing inappropriate initiation of replication during the post-replicative phase of the cell cycle. Because archaea with multiple Orc1/Cdc6s generally appear to have at least one ortholog of Cdc6-1 and one ortholog of Cdc6-2, it is tempting to speculate that this situation may be widely applicable to the regulation of archaeal DNA replication.

Once the origin has been bound by the Orc1/Cdc6 proteins, it is presumed that the replicative helicase —most likely the archaeal MCM complex— is loaded onto origins (Kelman and Kelman 2003). In bacteria, DnaA does not directly load the helicase (DnaB) but requires a third protein, DnaC. Similarly, in eukaryotes the ORC complex loads the MCM complex in a reaction that requires Cdc6 and Cdt1 (Bell and Dutta 2002). To date no archaeal homolog of Cdt1 has been characterised. However, as *S. cerevisiae* and *Schizosaccharomyces pombe* Cdt1 proteins share only 10% identity (Tanaka and Diffley 2002), it is possible that highly diverged Cdt1 homologs may exist, as yet undetected, in archaeal genomes. Nevertheless, as discussed above, the archaeal Orc1/Cdc6 proteins do possess homology to both Orc1 and Cdc6. It is possible, therefore, that these proteins may play roles in loading the MCM complex. However, this has not yet been demonstrated experimentally. One piece of data that may support Cdc6-like roles for the archaeal Orc1/Cdc6s has come from studies on *Methanobacterium thermoautotrophicum* (Mth) Orc1/Cdc6s by Kelman and colleagues (Shin et al. 2003a). It has been well established that the bacterial helicase loader, DnaC, inhibits the helicase activity of DnaB. Similarly, Kelman and colleagues revealed that the Mth Orc1/Cdc6s both inhibited the helicase activity of Mth MCM. Intriguingly, the MCM from the crenarchaeon *S. solfataricus* was also inhibited by these euryarchaeal proteins, suggesting a general mechanism for archaea. The inhibition was found to be dependent on the wHTH motif in the Orc1/Cdc6s. Furthermore, a direct interaction between Mth Orc1/Cdc6s and MCM was detected by yeast two-hybrid analysis (Shin et al. 2003a).

3.0 MCM COMPLEX

The eukaryotic MCM complex is a large multi-protein assembly, containing six related subunits, that plays a pivotal role in licensing origins of replication (Bell and Dutta 2002). Indeed, in higher eukaryotes MCM is tightly regulated and expressed only in proliferating cells, leading to its recent exploitation as a clinical marker for cellular proliferation (Williams et al. 1998). The archaeal MCM is composed of multiple copies of a single subunit (Kelman and Kelman 2003). The majority of studies have focused on the *M. thermoautotrophicum* MCM (Chong et al. 2000; Kelman et al. 1999; Shechter et al. 2000). Initially characterised as a double hexamer by a combination of hydrodynamic analyses and electron microscopy (Chong et al. 2000), a recent study has suggested that the protein can, in fact, form a double heptamer (Yu et al. 2002). Biochemical assays have shown the protein has helicase activity and is able to melt double-stranded DNA. The helicase activity is quite processive, liberating single strands of over 500 nt in length in a reaction dependent upon the hydrolysis of ATP or dATP. Sequence analysis of the

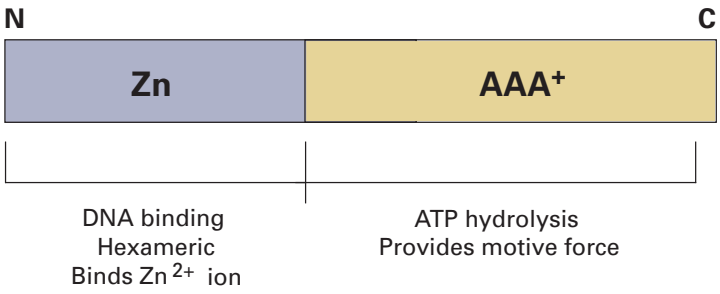


Figure 3. Domain organization of an archaeal MCM protein.

archaeal MCMs reveals a highly conserved AAA+ ATPase domain in the C-terminal two thirds of the protein (**Figure 3**), with a helix-turn-helix at the extreme C-terminus. Thus, it is within this region of the protein that the hydrolysis of ATP is catalysed and utilised to provide the motive force that manifests itself in the unwinding of DNA. The N-terminal region of the protein is less conserved, and the most obvious feature within this region is a zinc ribbon motif. Mutagenesis of this motif does not alter the multimeric status of the protein but does reduce ATPase and single-strand DNA binding, and abrogates helicase activity (Poplawski et al. 2001). The crystal structure of the N-terminal domain of Mth MCM has been solved (Fletcher et al. 2003). This revealed a double-hexamer arrangement with the two hexamers facing each other in a head-to-head alignment. Importantly, a large central cavity of at least 23 Å in diameter was apparent in the centre of the hexameric ring, and so could readily accommodate double- or single-stranded DNA. While the majority of studies have focussed on the ability of archaeal MCM to unwind DNA molecules, recent work has revealed that the complex also has the ability to translocate over double-stranded DNA without leading to melting of the double helix (Shin et al. 2003b). While the relevance of these two activities remains to be demonstrated *in vivo*, it is tempting to speculate that the MCM complex is initially loaded onto double-stranded DNA by Orc1/Cdc6 proteins. It is interesting to note that the studies of the *S. solfataricus* origins of replication have indicated that at both origins, replication initiates between inverted repeats that bind Cdc6-1 (Robinson et al. 2004). As these repeats are located 220-240 Å apart, it is possible that a double hexamer of MCM complex (roughly 200 Å in length) may be loaded onto DNA between two Cdc6-1 binding sites (**Figure 4**). If, as the crystal structure of Mth MCM indicates, the MCM hexamers face one another within the double hexamer, then they have the capacity to pump double-stranded DNA toward each other. Assuming an appropriate handedness of pumping, this could lead to under-winding and localised melting of DNA between the two hexamers. The single-stranded DNA thus extruded could then be recognised by the DNA primase molecule and strand synthesis initiated. This situation could be maintained and eventually the entire genome spooled through the MCMs. Alternatively, once the single-stranded region of DNA is generated between the two hexamers of MCM, the MCM could undergo a remodelling event and relocate onto the exposed single strands and proceed away from the initial site of melting. In light of the first model of double-strand DNA pumping, it is noteworthy that, in higher eukaryotes, the MCM complex appears to be localised distantly from the actual sites of DNA replication, leading Laskey and Madine to propose that in these organisms, the MCM may be acting as a double-strand pump (Laskey and Madine 2003).

In order to resolve these issues, it will be necessary to develop highly defined *in vitro* systems for origin recognition and MCM loading in order to dissect the key intermolecular transactions at the biochemical level. With the increasing knowledge of archaeal origins and their interactions with initiator proteins, it is anticipated that these systems will be developed in the near future.

4.0 THE ARCHAEAL DNA PRIMASE(S)

Once single-stranded DNA has been exposed at origins of replication it can be used as a template for the synthesis of daughter DNA strands. However, DNA polymerases (DNA pols) lack the capability to initiate *de novo* synthesis of DNA. Rather, they first require the action of a DNA primase to synthesise a short oligonucleotide primer that can then be extended by the DNA pol. The bacterial primase

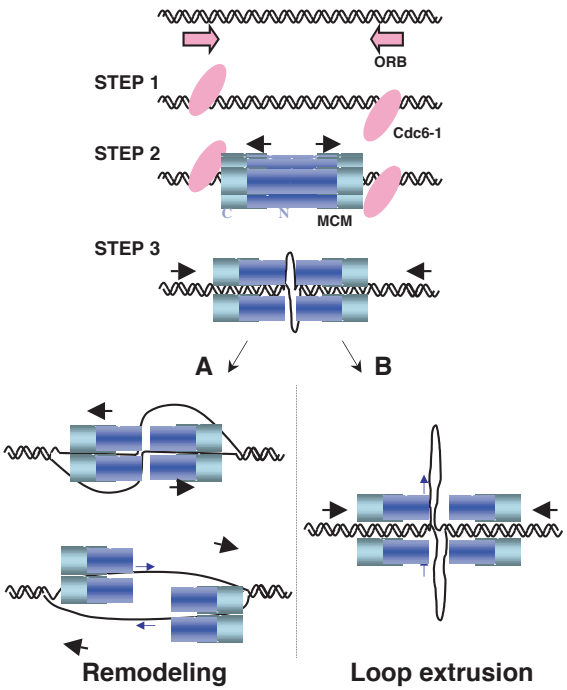


Figure 4. Model for the loading of MCM at a *S. solfataricus* origin of replication. ORB elements are indicated as pink arrows and Cdc6-1 shown as a pink oval. The MCM complex is shown in blue. As detailed in the text, our model suggests that two hexamers of MCM are loaded, facing each other between ORB elements and forming a double hexamer (note that for clarity we only show two MCM monomers per hexamer in this and later steps). **STEP 2.** When the hexamers attempt to translocate on DNA they are held together and this has the effect of spooling DNA into the centre of the double hexamer. **STEP 3.** At this point we envisage one of two scenarios occurring. In option A, the MCM hexamers reposition themselves on the spooled out single stranded DNA (ssDNA). Movement of the MCM along the ssDNA then sets up two conventional bi-directional replication forks. In scenario B, MCMs retain their double-stranded pumping mode, extruding ever larger "rabbit ears" of ssDNA from the centre of the MCM complex. This extruded DNA is the recognized by primase and strand synthesis initiated. This situation could occur until the entire genome has spooled through the MCM complex.

is the product of the *dnaG* gene (Frick and Richardson 2001). In light of the generally eukaryotic-like nature of the archaeal DNA replication proteins, it is somewhat surprising that many archaea encode a homolog of this protein. However, there has been no characterisation of the biochemical properties of the archaeal DnaG; moreover, a recent study has indicated that it co-purifies with the archaeal exosome—a complex involved in the degradation of RNA—suggesting a role in processes other than DNA replication (Evguenieva-Hackenberg et al. 2003).

The eukaryotic primase consists of a small subunit that possesses the catalytic activity. This forms a tight complex with a second, larger subunit, and, in turn, these interact with the B subunit and a DNA polymerase (DNA polα) to form a primase complex (Frick and Richardson 2001). Significantly, archaeal genomes encode homologs of two of the four subunits of the eukaryotic primase complex. Archaea do not have homologs of subunit B and polα but do have clear homologs of the large and small subunit, PriL and PriS. Initial studies focused on the isolated small subunit. A study of *M. jannaschii* PriS revealed the ability to synthesise RNA (Desogus et al. 1999). In contrast, studies of *Pyrococcus furiosus* PriS indicated that it had the capacity to generate extensive (over 1 kb) DNA molecules (Bocquier et al. 2001). Intriguingly, a subsequent study of the reconstituted *Pyrococcus* PriSL complex revealed that the presence of the large subunit reduced this novel DNA synthetic capability of PriS and conferred RNA synthetic capability upon the enzyme (Liu et al. 2001). It is not yet known if archaeal primases have specific sites at which they prefer to initiate synthesis, although it has been demonstrated that the *Pyrococcus* enzyme can initiate synthesis with an ATP molecule (Liu et al. 2001). The crystal structure of the *Pyrococcus* PriS subunit has been determined, revealing a rather flat molecule with a central groove, proposed to be the nucleic acid binding site, and a recess containing highly conserved aspartic acid residues, thought to be the catalytic centre of the enzyme (Augustin et al. 2001).

5.0 ARCHAEAL DNA POLYMERASES

All archaea possess at least one DNA pol of the B family. This is a broadly conserved family of DNA pols, with homologs identified in all three domains of life. Members of the euryarchaeal kingdom also possess a novel class of heterodimeric DNA pol, the D family (reviewed in Cann and Ishino 1999). The D family polymerases have two

subunits, DP1 and DP2 (Cann et al. 1998). The DP1 subunit contains a motif associated with pyrophosphatase activity that may increase the polymerisation rate of the enzyme. Additionally, DP1 contains a candidate-binding motif for the sliding clamp, PCNA (see section 6.0). DP2 is the catalytic subunit of the enzyme; however, isolated DP2 has low activity that is stimulated at least 50-fold by DP1. Like DP1, DP2 also contains a candidate PCNA interaction motif.

The precise roles of these two families of DNA pol *in vivo* remain to be determined. It is possible, for example, that they have compartmentalised roles in leading and lagging strand synthesis. Further, many crenarchaea possess multiple B family DNA pols; again, their functions *in vivo* remain to be determined (Cann et al. 1999a). Some archaea possess homologs of the "lesion bypass" or "error prone" family Y polymerases. In particular, Dpo4, the Y-family DNA pol of *S. solfataricus* P2, has been the subject of considerable study (Kulaeva et al. 1996). The lesion bypass polymerases have the capacity to synthesise DNA across from lesions (e.g., pyrimidine dimers) that would normally lead to the arrest of DNA polymerisation. An elegant series of structural studies has revealed the molecular basis for the low fidelity and template promiscuity of these enzymes (Ling et al. 2001, 2003). In particular, the co-crystal structure of a Dpo4/DNA/nucleotide complex revealed limited and non-specific interactions between Dpo4 and the replicating base pair. Additionally, unlike the case with other DNA polymerases, the active site is able to accommodate two template base pairs, suggesting a mechanism by which Dpo4 can bypass pyrimidine dimers.

6.0 ARCHAEAL SLIDING CLAMP AND ACCESSORY FACTORS

During leading strand synthesis, the DNA pol must be highly processive, potentially synthesising over a megabase of DNA without releasing the template. In contrast, on the lagging strand much shorter molecules, Okazaki fragments, are synthesised. Recent work has shown archaeal Okazaki fragments to be in the region of 100-200 nt (Matsunaga et al. 2003). The processivity of the DNA pol is not an innate property of the enzyme, but rather is conferred upon it by association with a so-called "sliding-clamp," a toroidal molecule that encircles double-stranded DNA behind the DNA pol, thereby holding it to the template (Warbrick 2000). In archaea and eukarya the sliding clamp is PCNA, the proliferating cell nuclear antigen. Crystal structures of *Pyrococcus* PCNA have been determined, revealing striking similarity to eukaryotic PCNA structures (Matsumiya et al. 2001). Indeed, this conservation is so high that the archaeal PCNA has been demonstrated to interact functionally with eukaryotic DNA pol (Ishino et al. 2001). As well as being important for conferring processivity to the leading strand DNA pol, PCNA also plays central roles in lagging strand synthesis. During lagging strand synthesis, adjacent Okazaki fragments must be joined. This involves a complex set of reactions in which the RNA primer of a downstream Okazaki fragment is displaced, generating a substrate for the flap endonuclease, Fen1. Fen1 cleavage results in exposure of a 5' phosphate that is then joined to the 3' hydroxyl of the upstream fragment by DNA ligase. Studies in eukaryotes have revealed that PCNA can interact with and stimulate the activity of both Fen1 and Ligase1 (Warbrick 2000). In eukaryotes and the euryarchaea, PCNA is a homotrimer, i.e., composed of three identical subunits (Cann et al. 1999b; Kelman and Hurwitz 2000). Intriguingly, in the crenarchaea, multiple PCNA homologs are encoded in many species. A study of the three *Aeropyrum pernix* PCNA homologs revealed that these subunits have the ability to both homo- and hetero-multimerise (Daimon et al. 2002). An even more extreme case was found in *S. solfataricus*, in which PCNA was found to be an obligate heterotrimer (Dionne et al. 2003). A series of functional and interaction studies revealed that distinct subunits of *S. solfataricus* PCNA had preferred interaction partners. Specifically, PCNA1 bound Fen1, PCNA2 interacted with DNA polB1 and the preferred partner for PCNA3 was Ligase 1. Furthermore, although individual PCNA subunits could interact with the partner protein, only the intact heterotrimer could stimulate Fen1, DNA polymerase, or ligase activities. This indicates that the likely mechanism by which PCNA stimulates the enzymatic activities is by acting as a passive DNA-binding tether and thereby facilitating recruitment of these enzymes to DNA. Finally, it was found that Fen1, DNA polB1, and Ligase 1 could bind simultaneously to the heterotrimeric PCNA ring

(Dionne et al. 2003). This indicates a possible mechanism to ensure tight coupling of Okazaki fragment synthesis and maturation (**Figure 5**). Although the asymmetry of the heterotrimeric *Sulfolobus* PCNA facilitated detection of this assembly, it is possible that this model of simultaneous occupancy of PCNA could be extended to homotrimeric PCNAs in other archaea and eukaryotes. In these cases, the specific geometry of such assemblies cannot be imposed at the level of PCNA-factor interaction, but may be mediated by an outer annulus of interactions between, for example, DNA pol, Fen1, and ligase. As well as roles in leading and lagging strand synthesis, PCNA has also been found to act as a docking platform for a number of DNA repair proteins. Examples in archaea include uracil DNA glycosylase (Yang et al. 2002), involved in the detection and removal of uracil in DNA, and the Xpf factor, a homolog of the eukaryotic Xpf nucleotide excision repair factor (Roberts et al. 2003). In *S. solfataricus*, Xpf is found in tight association with the heterotrimeric PCNA. Indeed, PCNA is an essential co-factor for the enzymatic activity of this factor (Roberts et al. 2003). Thus, it appears that, as in eukaryotes, archaeal PCNA acts as a general tether for a wide range of DNA replication and repair proteins.

6.1 Clamp Loader

The toroidal nature of the sliding clamp presents a potential problem in allowing DNA access to the central cavity. The PCNA ring must be opened and re-sealed to allow DNA access, or to remove the clamp from DNA. This activity is mediated by the "clamp-loader," replication factor C (RFC), in archaea and eukaryotes. This protein complex has five subunits; in eukaryotes, there is one large and four related small subunits. In archaea, the complex is somewhat simpler, consisting of one large and a homotetramer of small subunits. RFC homologs from a variety of archaea have now been characterised and, in general, these molecules have been shown to utilise ATP to load PCNA onto a circular DNA substrate (Cann et al. 2001; Kelman and Hurwitz 2000; Seybert et al. 2002). Both RFC subunits have been shown to interact with PCNA. A recent study on the heterotrimeric PCNA of *Sulfolobus* has indicated that two of the PCNA subunits interact with the RFC small subunit tetramer, and the

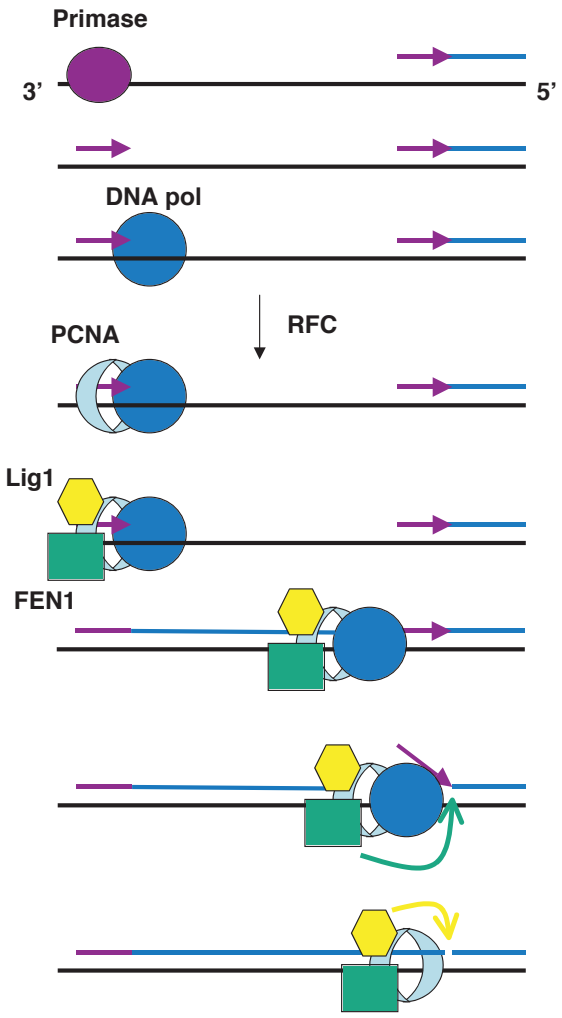


Figure 5. Model for co-ordinate synthesis and processing of Okazaki fragments. DNA primase initiates synthesis of a primer molecule (purple arrow). This is recognized by DNA polymerase and PCNA is loaded by the RFC factor. PCNA then acts as a scaffold for the assembly of flap endonuclease (FEN1) and DNA ligase (Lig1). DNA pol synthesizes new DNA (blue) and as it translocates is accompanied by PCNA, Lig1 and FEN1. On reaching the downstream Okazaki fragment, DNA pol displaces the primer creating a substrate for cleavage by FEN1 and the consequent nick is sealed by Lig1.

third PCNA subunit interacts with the large RFC subunit (Dionne et al. 2003). An ATP-induced conformational alteration in RFC could therefore be envisaged to prise open the PCNA ring and allow DNA access.

7.0 FUTURE DIRECTIONS

Considerable insight has been gained from the determination of the biochemical properties of a range of individual archaeal DNA replication-associated proteins. However, a key goal is the reconstitution of a defined *in vitro* replication system to address the roles of these proteins in the greater context. Of particular importance will be the elucidation of evolutionarily conserved interaction interfaces that modulate and integrate the various enzymatic activities.

Another aspect of the archaeal DNA replication field that will continue to provide insight into the functions of these important proteins is the ongoing attempts by a range of structural biologists to determine high-resolution structures of the replication factors. One eagerly anticipated achievement must be the determination of a high-resolution structure for an archaeal MCM. Additionally, although a structure has been obtained for the small subunit of the RFC complex (Oyama et al. 2001), it would be highly informative to determine the structure of RFC in complex with PCNA. This would allow determination of the molecular basis of the clamp loading reaction. As with the biochemical analyses, the reconstitution and structural determination of higher order assemblies of replication-associated proteins will be pivotal in understanding the coordinated actions of these key activities.

Beyond the fundamental replication machineries, very little is known about the regulation of archaeal DNA replication (Bernander 2003). In particular, species such as members of the *Sulfolobus* genus precisely regulate the replication of their genomic DNA, oscillating between 1N and 2N genomic contents. How this tight regulation of DNA replication and cell division is attained remains a mystery. It is anticipated that transcriptomic and proteomic analyses will begin to identify key regulatory molecules in this fundamental process.

ACKNOWLEDGMENTS

Work in SDB's lab is funded by the Medical Research Council.

REFERENCES

Augustin, M.A., R. Huber, and J.T. Kaiser. 2001. Crystal structure of a DNA-dependent RNA polymerase (DNA primase). *Nat Struct Biol* 8:57-61.

Bell, S.P. 2002. The origin recognition complex: from simple origins to complex functions. *Genes & Development* 16:659-72.

Bell, S.P., and A. Dutta. 2002. DNA replication in eukaryotic cells. *Annu Rev Biochem* 71:333-74.

Bernander, R. 2003. The archaeal cell cycle: current issues. *Mol Microbiol* 48:599-604.

Berquist, B.R., and S. DasSarma. 2003. An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, *Halobacterium* sp. strain NRC-1. *J Bacteriol* 185.

Bocquier, A.A., L.D. Liu, I.K.O. Cann, K. Komori, D. Kohda, and Y. Ishino. 2001. Archaeal primase: bridging the gap between RNA and DNA polymerases. *Curr Biol* 11:452-56.

Cann, I.K., S. Ishino, N. Nomura, Y. Sako, and Y. Ishino. 1999a. Two family B DNA polymerases from *Aeropyrum pernix*, an aerobic hyperthermophilic crenarchaeote. *J Bacteriol* 181:5984-92.

Cann, I.K., K. Komori, H. Toh, S. Kanai, and Y. Ishino. 1998. A heterodimeric DNA polymerase: evidence that members of *Euryarchaeota* possess a distinct DNA polymerase. *Proc Natl Acad Sci* 95:14250-5.

Cann, I.K.O., S. Ishino, I. Hayashi, K. Komori, H. Toh, K. Morikawa, and Y. Ishino. 1999b. Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. *J Bacteriol* 181:6591-9.

Cann, I.K.O., S. Ishino, M. Yuasa, H. Daiyasu, H. Toh, and Y. Ishino. 2001. Biochemical analysis of replication factor C from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol* 183:2614-23.

Cann, I.K.O., and Y. Ishino. 1999. Archaeal DNA replication: Identifying the pieces to solve a puzzle. *Genetics* 152:1249-67.

Chong, J.P.J., M.K. Hayashi, M.N. Simon, R.M. Xu, and B. Stillman. 2000. A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. *Proc Natl Acad Sci* 97:1530-5.

Daimon, K., Y. Kawarabayasi, H. Kikuchi, Y. Sako, and Y. Ishino. 2002. Three proliferating cell nuclear antigen-like proteins found in the hyperthermophilic archaeon *Aeropyrum pernix*: Interactions with the two DNA polymerases. *J Bacteriol* 184: 687-94.

Desogus, G., S. Onesti, P. Brick, M. Rossi, and F.M. Pisani. 1999. Identification and characterization of a DNA primase from the hyperthermophilic archaeon *Methanococcus jannaschii*. *Nucl Acids Res* 27:4444-50.

Dionne, I., R.K. Nookala, S.P. Jackson, A. Doherty, and S.D. Bell. 2003. A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Cell* 11:275-82.

Evguenieva-Hackenberg, E., P. Walter, E. Hochleitner, F. Lottspeich, and G. Klug. 2003. An exosome-like complex in *Sulfolobus solfataricus*. *EMBO Rep* 4:889-93.

Fletcher, R.J., B.E. Bishop, R.P. Leon, R.A. Sclafani, C.M. Ogata, and X.S. Chen. 2003. The structure and function of MCM from archaeal *M. thermoautotrophicum*. *Nat Struct Biol* 10: 160-7.

Frick, D.N., and C.C. Richardson. 2001. DNA primases. *Annu Rev Biochem* 70;:39-80.

Grabowski, B., and Z. Kelman. 2001. Autophosphorylation of archaeal Cdc6 homologues is regulated by DNA. *J Bacteriol* 183:5459-64.

Ishino, Y., T. Tsurimoto, S. Ishino, and I.K. Cann. 2001. Functional interactions of an archaeal sliding clamp with mammalian clamp loader and DNA polymerase delta. *Genes Cells* 6:699-706.

Jacob, F., S. Brenner, and F. Kuzin. 1963. On the regulation of DNA replication in bacteria. Cold Spring Harbor *Symp Quant Biol* 28:329-48.

Kelman, L.M., and Z. Kelman. 2003. Archaea: an archetype for replication initiation studies? *Mol Microbiol* 48:605-16.

Kelman, Z., and J. Hurwitz. 2000. A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon *Methanobacterium thermoautotrophicum* Delta H. *J Biol Chem* 275:7327-36.

Kelman, Z., J.K. Lee, and J. Hurwitz. 1999. The single minichromosome maintenance protein of *Methanobacterium thermoautotrophicum* Delta H contains DNA helicase activity. *Proc Natl Acad Sci* 96:14783-8.

Kulaeva, O.I., E.V. Koonin, J.P. McDonald, S.K. Randall, N. Rabinovich, J.F. Connaughton, A.S. Levine, and R. Woodgate. 1996. Identification of a DinB/UmuC homolog in the archeon *Sulfolobus solfataricus*. *Mutat Res* 357:245-53.

Laskey, R.A., and M.A. Madine. 2003. A rotary pumping model for helicase function of MCM proteins at a distance from replication forks. *EMBO Rep* 4:26-30.

Ling, H., F. Boudsocq, B.S. Plosky, R. Woodgate, and W. Yang. 2003. Replication of a cis-syn thymine dimer at atomic resolution. *Nature* 424:1083-7.

Ling, H., F. Boudsouq, R. Woodgate, and W. Yang. 2001. Crystal structure of a Y-family DNA polymerase in action: A mechanism for error-prone and lesion-bypass replication. *Cell* 107:91-102.

Liu, J.Y., C.L. Smith, D. DeRyckere, K. DeAngelis, G.S. Martin, and J.M. Berger. 2000. Structure and function of Cdc6/Cdc18: Implications for origin recognition and checkpoint control. *Mol Cell* 6:637-48.

Liu, L.D., K. Komori, S. Ishino, A.A. Bocquier, I.K.O. Cann, D. Kohda, and Y. Ishino. 2001. The archaeal DNA primase - Biochemical characterization of the p41-p46 complex from *Pyrococcus furiosus*. *J Biol Chem* 276:45484-90.

Matsumiya, S., Y. Ishino, and K. Morikawa. 2001. Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from *Pyrococcus furiosus*. *Protein Sci* 10:17-23.

Matsunaga, F., P. Forterre, Y. Ishino, and H. Myllykallio. 2001. *In vivo* interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc Natl Acad Sci* 98:11152-7.

Matsunaga, F., C. Norais, P. Forterre, and H. Myllykallio. 2003. Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep* 4: 154-8.

Messer, W., F. Blaesing, D. Jakimowicz, M. Krause, J. Majka, J. Nardmann, S, Schaper, H. Seitz, C. Speck, C. Weigel, et al. 2001. Bacterial replication initiator DnaA. Rules for DnaA binding and roles of DnaA in origin unwinding and helicase loading. *Biochimie* 83:5-12.

Myllykallio, H., P. Lopez, P. Lopez-Garcia, R. Heilig, W. Saurin, Y. Zivanovic, H. Philippe, and P. Forterre. 2000. Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* 288:2212-15.

Oyama, T., Y. Ishino, I.K.O. Cann, S. Ishino, and K. Morikawa. 2001. Atomic structure of the clamp loader small subunit from *Pyrococcus furiosus*. *Mol Cell* 8:455-63.

Poplawski, A., B. Grabowski, S.F. Long, and Z. Kelman. 2001. The zinc finger domain of the archaeal minichromosome maintenance protein is required for helicase activity. *J Biol Chem* 276: 49371-77.

Roberts, J.A., S.D. Bell, and M.F. White. 2003. An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. *Mol Microbiol* 48:361-72.

Robinson, N.P., I. Dionne, M. Lundgren, V.L. Marsh, R. Bernander, and S.D. Bell. 2004. Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* 116:25-38.

Schaarschmidt, D., J. Baltin, I.M. Stehle, H.J. Lipps, and R. Knippers. 2003. An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *EMBO J advance online publication*, December 11, 2003; doi:10.1038/sj.emboj.7600029.

Seybert, A., D.J. Scott, S. Scaife, M.R. Singleton, and D.B. Wigley. 2002. Biochemical characterisation of the clamp/clamp loader

proteins from the euryarchaeon *Archaeoglobus fulgidus*. *Nucl Acids Res* 30:4329-38.

She, Q., R.K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M.J. Awayez, C.C.Y. Chan-Weiher, I.G. Clausen, B.A. Curtis, A. De Moors, et al. 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci* 98: 7835-40.

Shechter, D.F., C.Y. Ying, and J. Gautier. 2000. The intrinsic DNA helicase activity of *Methanobacterium thermoautotrophicum* Delta H minichromosome maintenance protein. *J Biol Chem* 275:15049-59.

Shin, J.H., B. Grabowski, R. Kasiviswanathan, S.D. Bell, and Z. Kelman. 2003a. Regulation of minichromosome maintenance helicase activity by Cdc6. *J Biol Chem* 278:38059-67.

Shin, J.H., Y. Jiang, B. Grabowski, J. Hurwitz, and Z. Kelman. 2003b. Substrate requirements for duplex DNA translocation by the eukaryal and archaeal minichromosome maintenance helicases. *J Biol Chem* 278:49053-62.

Tanaka, S., and J.F.X. Diffley. 2002. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nat Cell Biol* 4:198-297.

Warbrick, E. 2000. The puzzle of PCNA's many partners. *Bioessays* 22:997-1006.

Williams, G.H., P. Romanowski, L. Morris, M. Madine, A.D. Mills, K. Stoeber, J. Marr, R.A. Laskey, and N. Coleman. 1998. Improved cervical smear assessment using antibodies against proteins that regulate DNA replication. *Proc Natl Acad Sci* 95:14932-7.

Yang, H., J.H. Chiang, S. Fitz-Gibbon, M. Lebel, A.A. Sartori, J. Jiricny, M.M. Slupska, and J.H. Miller. 2002. Direct interaction between uracil-DNA glycosylase and a proliferating cell nuclear antigen homolog in the crenarchaeon *Pyrobaculum aerophilum*. *J Biol Chem* 277:22271-8.

Yu, X., M.S. VanLoock, A. Poplawski, Z. Kelman, T. Xiang, B.K. Tye, and E.A. Egelman. 2002. The *Methanobacterium thermoautotrophicum* MCM protein can form heptameric rings. *EMBO Rep* 3:792-7.

Zhang, R., and C. Zhang. 2003. Multiple replication origins of the archaeon *Halobacterium* species NRC-1. *Biochem Biophys Res Commun* 302:728-34.