

# Chapter 5

## Initiation of DNA Replication in the Archaea

Stephen D. Bell

**Abstract** Organisms within the archaeal domain of life possess a simplified version of the eukaryotic DNA replication machinery. While some archaea possess a bacterial-like mode of DNA replication with single origins of replication per chromosome, the majority of species characterized to date possess chromosomes with multiple replication origins. Genetic, structural, and biochemical studies have revealed the nature of archaeal origin specification. Recent work has begun to shed light on the mechanisms of replication initiation in these organisms.

**Keywords** DNA replication • Initiator protein • Helicase • Replication origin • Archaea • *Sulfolobus*

The archaea are a diverse range of microorganisms that share more recent evolutionary history with eukaryotes than do the bacteria (Woese and Fox 1977). The precise timing of the divergence of the archaeal and eukaryotic lineages is the subject of considerable debate, with some studies even suggesting that eukaryotes arose from within the archaeal domain of life (Williams et al. 2013; Rivera and Lake 2004; Forterre 2015). A number of phyla have been identified within the Archaea; again controversy exists regarding the precise nature of the taxonomic divisions between archaeal phyla. With increased sampling, particularly at the metagenomic level, some degree of consensus is being established. It is generally accepted that there is a broad divide between the phylum of the Euryarchaea and those of the Thaumarchaea, Aigarchaea, Crenarchaea, and Korarchaea. The latter four taxonomic groupings appear more closely related to one another and have been termed the “TACK superphylum” (Guy and Ettema 2011; Brochier-Armanet et al. 2008; Forterre 2015). At the morphological level, archaea are prokaryotes; most species have a single cell membrane and are devoid of any organelle structures.

---

S.D. Bell (✉)

Department of Molecular and Cellular Biochemistry, Department of Biology, Indiana University, Bloomington, IN, USA

e-mail: [stedbell@indiana.edu](mailto:stedbell@indiana.edu)

**Table 5.1** Taxonomic distribution of archaeal species described in the text

Phylum	Species	Origins in main chromosome	Copy number
Crenarchaea	<i>Sulfolobus islandicus</i>	3	1C-2C
	<i>Sulfolobus solfataricus</i>	3	1C-2C
	<i>Aeropyrum pernix</i>	2	1C-2C
	<i>Pyrobaculum calidifontis</i>	4	1C-2C
Euryarchaea	<i>Pyrococcus abyssi</i>	1	Polyploid
	<i>Haloferax volcanii</i>	3 (or 4)	Polyploid
	<i>Haloferax mediterranei</i>	3	Polyploid
	<i>Methanothermobacter thermautotrophicum</i>	one mapped thus far	Polyploid
	<i>Thermococcus kodakarensis</i>	ND	Polyploid

The ambiguity in origin number in *Haloferax volcanii* lies in the description of the integration of an extrachromosomal element into the 3-origin main chromosome in a lab-adapted strain (Hawkins et al. 2013). While one origin has been mapped in *M. thermautotrophicum* (Majernik and Chong 2008), no genome-wide replication profiling has been performed on this organism. It is possible, therefore, that additional origins exist in this species

ND not determined

Thus far, all archaea characterized have circular chromosomes; however, the chromosome copy number shows considerable variation across taxonomic divides. To a broad approximation, euryarchaea appear to be generally oligoploid or polyploid, while the members of TACK that have been studied have cell cycles that oscillate between one and two copies of their chromosome (Table 5.1) (Samson and Bell 2014; Breuert et al. 2006; Hildenbrand et al. 2011). Flow cytometry studies have revealed that the TACK superphylum organisms, such as members of the *Sulfolobus* genus of hyperthermophilic acidophiles, have cell cycles that contain defined gap phases separating DNA replication and cell division (Lundgren et al. 2008; Pelve et al. 2013). These observations have led to the adoption of the G1, S, G2, and M phase nomenclatures established in studies of the eukaryotic cell cycle to describe the analogous stages of archaeal cell cycle progression. It must be emphasized, however, that there is no evidence that archaeal chromosome segregation is in any way related to eukaryotic mitosis. Interestingly, in marked contrast to the orchestrated cell cycles of crenarchaea, the euryarchaea that have been studied appear to lack obvious gap phases, perhaps hinting that cell division can occur during ongoing rounds of replication of the multiple copies of the chromosome, in a manner somewhat reminiscent of fast-growing *E. coli* (Sherratt 2003).

## 5.1 The Replication Machinery of Archaea

With the availability of whole genome sequences of archaeal species in the 1990s, it became apparent that archaea possess clear orthologs of eukaryotic DNA replication-associated proteins (Edgell and Doolittle 1997). In general, and in keeping with the organizational simplicity of the organisms, the archaeal replication proteins are

simplified versions of their eukaryotic counterparts (Barry and Bell 2006; Kelman and Kelman 2014). For example, the eukaryotic MCM(2-7) replicative helicase has six distinct subunits. However, all six subunits are related to each other in sequence, suggesting derivation from a common ancestor. Indeed, the majority of present-day archaea encode a single *mcm* gene the product of which homo-multimerizes to form a homohexameric assembly (Costa and Onesti 2009; Bochman and Schwacha 2009). Similarly, almost all archaea encode a protein that is related to both Cdc6 and Orc1 component of eukaryotic origin recognition complex, ORC (Bell 2012). Interestingly, early branching eukaryotes, such as trypanosomes, also encode an archaeal-like Orc1/Cdc6 protein, suggesting that the gene duplication and sequence diversification leading to “higher” eukaryotic Orc1 and Cdc6 occurred within the eukaryotic lineage (Samson and Bell 2016; Tiengwe et al. 2012). Importantly, the bacterial replication machinery, although, ultimately, performing the same function, is largely non-orthologous to the shared archaeal/eukaryotic replication apparatus. The key exceptions lie in the clamp/loader and sliding clamp that facilitate DNA polymerization, leading to the proposal that the elongation machinery is fundamentally conserved and thus ancestral, even though the rest of the replisome components are not conserved between bacteria and archaea/eukarya (Yao and O’Donnell 2016).

## 5.2 Archaeal Replication Initiation

The first archaeon in which the replication mode was experimentally determined, a euryarchaeon *Pyrococcus abyssi*, revealed a single origin of replication. The origin, *oriC*, is located in a gene environment that contained genes for several replication-associated proteins, including the candidate initiator protein *orc1/cdc6* gene (Myllykallio et al. 2000; Bell 2012). The *orc1/cdc6* nomenclature is cumbersome, and orthologs in archaeal genomes have been variously annotated as *orc1* or *cdc6* on an apparently random basis. In this chapter, for simplicity’s sake, I will refer to these genes as *orc1*. Many archaea encode multiple Orc1 paralogs, and I will refer to these as Orc1-1, Orc1-2, etc.

Interestingly, the single-origin paradigm in *Pyrococcus* species actually appears to be atypical among the archaea, and it is now known that many archaea from both euryarchaea and TACK species have multiple replication origins per chromosome (Robinson and Bell 2007; Robinson et al. 2004; Robinson et al. 2007; Lundgren et al. 2004; Norais et al. 2007; Hawkins et al. 2013; Yang et al. 2015; Pelve et al. 2012). The highest number of origins reported is four per chromosome for lab strains of the euryarchaeon *Haloferax volcanii* and also the crenarchaeon *Pyrobaculum calidifontis* (Pelve et al. 2012; Hawkins et al. 2013). For most species, while origin number and location have been established, the extent to which each origin is used remains poorly resolved. The exception to this lies in *Sulfolobus* species where three origins have been mapped, and these have been experimentally determined to fire once per cell cycle (Duggin et al. 2008). Studies with synchronized cell populations have revealed that two of the origins, *oriC1* and *oriC3*, fire

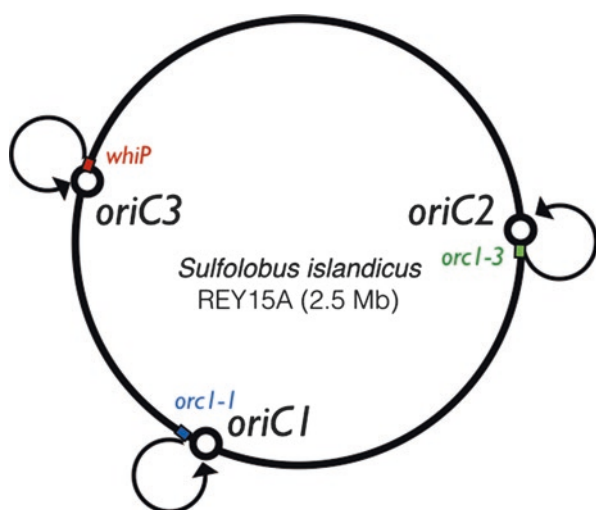
synchronously, thereby defining the start of S phase. Notably, *oriC2* fires a few minutes later. As will be discussed below, this temporal delay is likely linked to the expression of the initiator protein that defines this origin.

Many archaea encode multiple Orc1 paralogs. In the case of *Sulfolobus*, three such proteins, Orc1-1, Orc1-2, and Orc1-3, are encoded in the 2.2–3 megabase-pair genome. *Sulfolobus* also encodes a further candidate initiator protein, WhiP, that is a distant homolog of another eukaryotic replication initiation protein, the helicase co-loader, Cdt1 (Robinson and Bell 2007).

### 5.3 Origin Specification

Genetic studies in *Sulfolobus islandicus* have revealed a simple one-to-one relationship between the location of initiator protein genes (Fig. 5.1) and the origins that they specify (Samson et al. 2013). More specifically, Orc1-1 is encoded adjacent to, and specifies origin function at, *oriC1*; Orc1-3 is adjacent to *oriC2* and is required for function at that origin, and finally the gene for WhiP is beside *oriC3*, and the WhiP gene product is necessary for *oriC3* function. Furthermore, the initiator protein encoded adjacent to each origin is both necessary and sufficient for its cognate origin function. What then of Orc1-2? The *orc1-2* gene is not encoded immediately adjacent to any of the three origins, and deletion of *orc1-2* does not affect firing at any of the three origins. A range of biochemical and transcriptomic analyses have implicated Orc1-2 as a negative regulator of replication (Robinson et al. 2004; Maaty et al. 2009; Frols et al. 2007; Gotz et al. 2007). However, its role in this regard remains to be firmly established. Thus, the *Sulfolobus islandicus* chromosome is a mosaic of three distinct replicons, each origin having its own specific

**Fig. 5.1** Diagram of the organization of the 2.5 Mb chromosome of *Sulfolobus islandicus*. The relative positions of the three origins are indicated along with their cognate initiator proteins (Samson et al. 2013). Genetic dependence of the origin upon initiators is indicated by the circular arrows

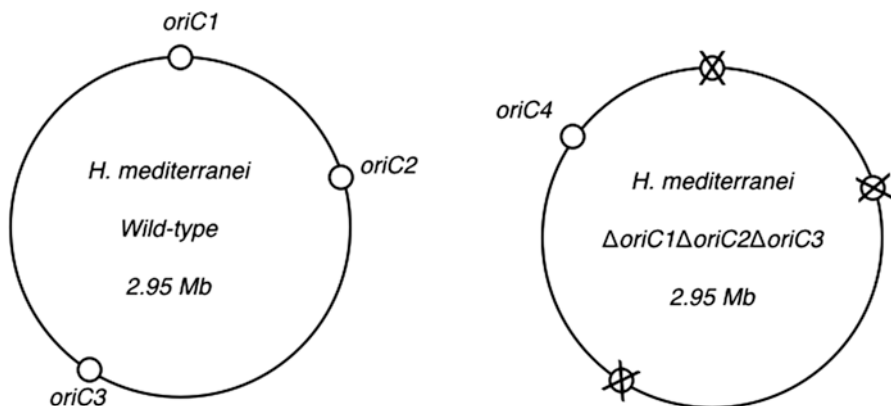


initiator. Analyses of the phyletic distribution of the initiator proteins reveal that Orc1-1 is highly conserved across a broad range of archaeal species. For example, the single *orc1* gene encoded by *Pyrococcus* is most closely related to *Sulfolobus* Orc1-1. Indeed, it was demonstrated that *Sulfolobus solfataricus* Orc1-1 can bind specifically to conserved sequence elements, termed ORB (origin recognition box), in the *Pyrococcus oriC* in vitro (Robinson et al. 2004). ORB elements are conserved across the archaeal domain of life and possess a dyad symmetric element flanked uniquely on one side by a G-rich element. Interestingly, all characterized *oriC1* origins in archaea possess at least two ORB elements in inverted orientation and separated by an AT-rich candidate duplex unwinding element (see Samson and Bell 2016 for a review). The nature of Orc1-1 interaction with ORB elements is discussed below.

In contrast to the near universality of Orc1-1, Orc1-3 appears to be restricted to the Sulfolobales, and WhiP is found in both Sulfolobales and Desulfurococcales. This patchy distribution of the initiators suggests that the *oriC2* and *oriC3* origin/initiator cassettes are relatively recent acquisitions, and it has been proposed that they have been acquired by incorporation of extrachromosomal elements into an ancestral *oriC1*/Orc1-1 containing chromosome (McGeoch and Bell 2008; Robinson and Bell 2007; Samson and Bell 2014).

Direct evidence for functional incorporation of extrachromosomal origins has been documented in the halophilic euryarchaeon *Haloferax volcanii* where a lab strain differs from the parental strain by incorporation of a large plasmid, pHV4, into the host chromosome (Hawkins et al. 2013). Importantly, the origin on the plasmid remains functional in its new integrated location. The malleability of the replicon architecture of *H. volcanii* main chromosome is underscored by the remarkable observation that its replication can be maintained even in the apparent absence of active replication origins. More specifically, experiments to delete all four origins in the lab strain of *H. volcanii*'s main chromosome were successful, and, very strangely, the resultant "zero origin" strain actually outcompeted the wild type in coculture experiments. The zero origin strain was highly dependent on the RAD51/RecA ortholog, RadA, suggesting a recombination-based mechanism was able to drive genome duplication (Hawkins et al. 2013). How universal this remarkable observation is is not yet clear (Michel and Bernander 2014). When similar experiments were performed in the closely related *H. mediterranei*, the main chromosome of which normally has three active origins (Fig. 5.2), deletion of the three origins led to activation of a cryptic novel origin of replication (Yang et al. 2015). It is possible that the high ploidy, sexual promiscuity (as manifested by high levels of intraspecies and even interspecies genetic exchange mediated by this organism), and natural competence, i.e., ability to uptake DNA from the media, may be contributory to *H. volcanii*'s remarkable genomic plasticity (Zerulla et al. 2014; Zerulla and Soppa 2014; Naor et al. 2012).

Genetic studies in *Sulfolobus islandicus* (Sis) reveal that at least one replication origin is essential for viability and that each origin has a unique initiator protein. Intriguingly, this simple binary relationship of origin and initiator is not conserved across the *Sulfolobus* genus. Studies in *Sulfolobus solfataricus* (Sso) have revealed

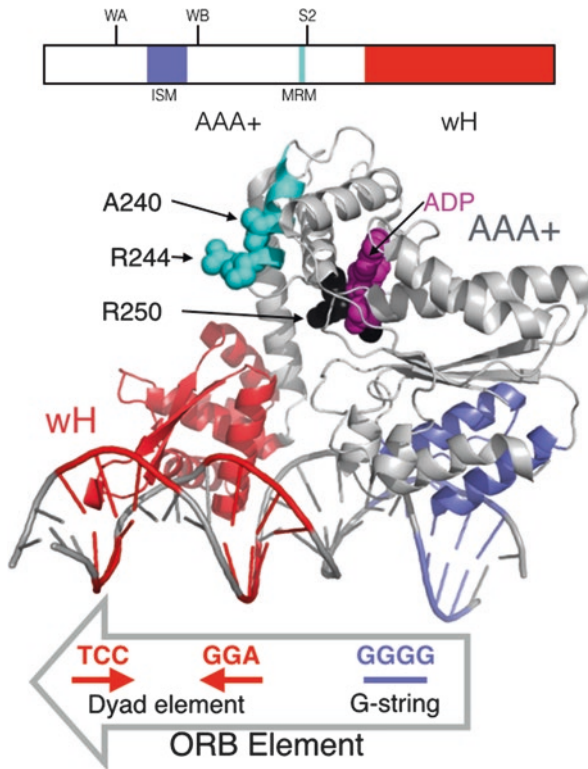


**Fig. 5.2** Diagram of the organization of the 2.95 Mb main chromosome of *Haloferax mediterranei*. The locations of the three active origins in wild-type cells are shown in the left-hand panel. The right-hand panel indicated that, upon deletion of *oriC1*, *oriC2*, and *oriC3*, cell viability is maintained by activation of a novel cryptic origin, *oriC4*. For details see Yang et al. (2015)

that *oriC2* in that species is bound by both Orc1-1 and Orc1-3. While the genetic dependence of this origin on both initiators has not been tested, a range of chromatin immunoprecipitation and biochemical and structural studies have demonstrated that this origin is bound by both Orc1-1 and Orc1-3 (Robinson et al. 2004; Dueber et al. 2007; Dueber et al. 2011). The Orc1-1 and Orc1-3 binding sites at this origin are immediately adjacent, and the two proteins have a 360 Å<sup>2</sup> protein-protein interface (Dueber et al. 2007). A biochemical comparison of Orc1-1 and *oriC2* between *S. islandicus* and *S. solfataricus* revealed that both origin sequence and protein sequence have evolved to allow the binding of *S. solfataricus* Orc1-1 to *oriC2* in that species (Samson et al. 2013). This enhanced complexity of origin specification may give insight into the evolutionary transitions that drove the evolution of the multi-subunit present-day ORC complex found in eukaryotes.

## 5.4 Orc1 Protein Structure and Function

The structural studies of *Sulfolobus* Orc1-1 and Orc1-3 bound to *oriC2*, in conjunction with the work from the Wigley lab on *Aeropyrum pernix* Orc1-1 bound to its cognate *oriC1*, revealed some general principles of Orc1 protein/DNA interactions (Dueber et al. 2007; Gaudier et al. 2007). The archaeal Orc1 proteins are approximately 43 kDa in size and possess an N-terminal AAA+ domain and a C-terminal winged-helix (wH) DNA-binding domain (Fig. 5.3). While mutational studies had demonstrated the importance of the wH domain in DNA binding, the structural studies revealed that the AAA+ domain also contacted the DNA (Gaudier et al. 2007; Robinson et al. 2004; Dueber et al. 2007; Dueber et al. 2011). The contact between the AAA+ domain and DNA is mediated by a signature embellishment to



**Fig. 5.3** Structure of the Orc1 proteins. The upper panel is a linear representation of the protein. The N-terminal two-thirds are a AAA+ domain, and the positions of the Walker A (WA), Walker B (WB), and Sensor 2 (S2) motifs are indicated. The ISM is the signature initiator-specific motif embellishment to the AAA+ fold found in the initiator clade of AAA+ proteins. MRM indicates the location of the MCM recruitment motif. The C-terminal third of the protein forms a winged-helix (wH) fold. The crystal structure shown below is of Orc1-1 bound to an ORB element (PDB Accession Number 2V1U). The ORB element is shown by a large gray arrow with internal dyad element and G-string element indicated. The orientation of the arrow is the same as that in Fig. 5.5. The wH domain in red interacts with the dyad symmetric element of the DNA. The ISM, in blue, mediates contacts a G-rich element, the so-called G-string. ADP is present in the active site of the AAA+ domain and is shown in magenta. The residues highlighted in cyan have been demonstrated to be essential for recruitment of MCM by Orc1-1 (residue numbering based on the *S. islandicus* Orc1-1 protein)

the classical AAA+ fold found in the initiator clade of AAA+ proteins, termed the initiator-specific motif (ISM). Thus, the *orc1* proteins make extended bipartite interactions with the origin DNA (Fig. 5.3). It had been demonstrated that Orc1-1 bound to conserved sequence elements, termed ORBs, at *oriC1* (Robinson et al. 2004). ORB elements contain a dyad symmetric element flanked on one side only by a string of G-C base pairs. The wH domain recognizes the dyad element, and the G-string interacts with the ISM (Fig. 5.3). Despite the presence of the dyad element, only a single Orc1-1 molecule binds per ORB element. The structural studies revealed that binding of Orc1-1 substantially distorts and underwinds the DNA to

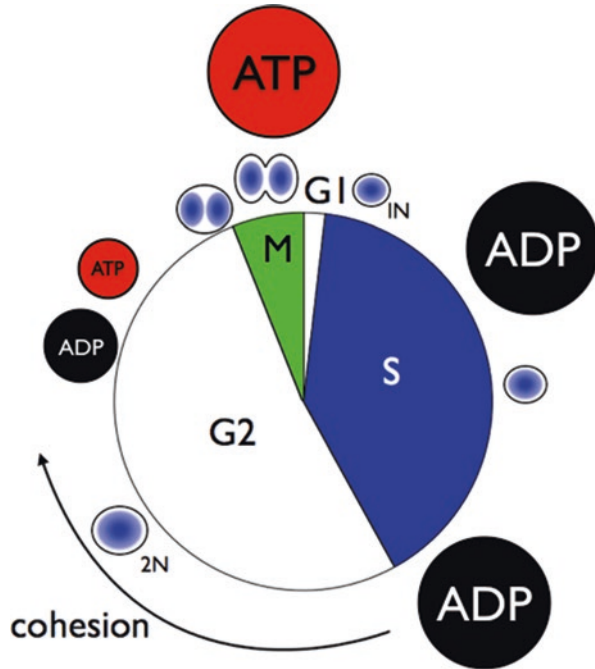
the extent that a second Orc1-1 molecule is unable to recognize the symmetry-related binding site (Gaudier et al. 2007). The preferred orientation of Orc1-1 on an ORB element is presumably defined by the unique ISM-G-string interaction.

Thus, at *oriC1*, Orc1-1 binds to ORB elements as a monomer. Another key feature in the structural studies was that the active site of the AAA+ ATPase domain was occupied by ADP. As no nucleotide was supplemented during purification and crystallization, this presumably reflects ATP bound during expression in *E. coli* and hydrolyzed during the expression and purification processes. Biochemical studies have confirmed that ADP is extremely stably bound to Orc1 proteins. Indeed, protein denaturation and extensive and subsequent re-folding are required to obtain nucleotide-free protein with which to perform ATPase studies (Grainge et al. 2006; Samson et al. 2013). Such studies have revealed that Orc1-1 undergoes a single-turnover ATP hydrolysis event leaving ADP stably bound in the active site. While bacterial DnaA is also active in its ATP-bound state, this activation is manifested in a fundamentally distinct manner from that of Orc1-1. ATP facilitates multimerization of DnaA, ultimately resulting in direct remodeling and melting of the origin DNA (see Bleichert et al. 2017 for a review). In contrast, Orc1-1 remains monomeric when ATP bound and undergoes a subtle conformational change (Samson et al. 2013) that facilitates interaction with MCM, as described below.

Studies using mutated versions of Orc1-1 in vivo and in vitro have revealed that stabilization of the ATP-bound form of the protein by substitution of the so-called Walker B glutamic acid residue by alanine results in a highly active form of the protein (Samson et al. 2013; Samson et al. 2016). In contrast, the ADP-bound form of Orc1-1 is inactive in MCM loading in vitro. On the biochemical level, ATP binding did not alter either the affinity or stoichiometry of Orc1-1 binding to DNA. Rather, ATP binding simply induced a modest conformational change in the protein, as detected by analytical ultracentrifugation and protease sensitivity assays. Despite these modest changes, the constitutively ATP-bound form of the protein was far more active in vitro than the ADP form (Samson et al. 2013). Thus, it appears that ATP binding and not hydrolysis is required for Orc1-1 function. Importantly, expression of the Walker B mutant form of the protein in vivo resulted in an overreplication phenotype, suggesting that ATP hydrolysis serves as an off switch. In this regard, it is significant that the *orc1-1* gene shows cell cycle-dependent regulation of its expression with transcript levels highest in cells about to enter G1 (Samson et al. 2013). Thus, the cell cycle dependence of *orc1-1* expression, coupled with the single-turnover ATP hydrolysis activity, indicates that Orc1-1 is acting as a molecular switch, permitting MCM recruitment in the Orc1-1•ATP state and inhibiting it in the Orc1-1•ADP state. Such a binary switch behavior is likely important for ensuring once-per-cell cycle regulation of origin activity (Fig. 5.4). The timing of expression of the initiator protein gene thus helps define a permissive window for initiator function. As mentioned above, *oriC2* fires a few minutes later in the cell cycle than does *oriC1*. This is reflected in the later peak of transcription of the Orc1-3 mRNA, relative to that for Orc1-1 (Samson et al. 2013). How the ADP-bound form of the initiator is removed from the origin at the end of the cell cycle is currently unknown. Possible



**Fig. 5.4** Cartoon of the *Sulfolobus* cell cycle. The red and black spheres labeled ATP and ADP represent the nucleotide status of Orc1 proteins. A pulse of transcription of *orc1-1* at the time of cell division will produce Orc1-1 associated with ATP; subsequently ATP will be hydrolyzed to ADP in a single-turnover event. ADP will thus remain stably associated with Orc1-1 for the rest of the cell cycle



explanations include an ATP exchange factor or, perhaps more likely given the new wave of *orc1-1* transcription, targeted destruction of Orc1-1•ADP at cell division.

## 5.5 MCM Recruitment to Archaeal Replication Origins

As alluded to above, Orc1-1 is able to recruit MCM to *oriC1* in a defined reaction using recombinant proteins purified from *E. coli* (Samson et al. 2016). These experiments reveal that, in addition to Orc1-1 sharing sequence homology with Orc1 and Cdc6 of eukaryotes, Orc1-1 also shares Orc1 and Cdc6's respective functions of origin binding and helicase recruitment. Orc1-1•ATP was shown to contact MCM's C-terminal wH domain via a conserved motif in the lid domain of the AAA+ domain (the MRM – MCM recruitment motif; see Fig. 5.3). The basis of the ATP dependence of Orc1-1's functionality was ascribed to the Sensor 2 motif. This conserved arginine residue has the capacity to coordinate the gamma phosphate of ATP and in doing so modulate the relative positions of the two subdomains of the AAA+ module. Importantly, mutation of the Sensor 2 residue led to a protein that bound ATP but had substantially reduced ATPase activity. However, unlike the Walker B mutant that has similar ATPase-null behavior, the Sensor 2 mutant Orc1-1 was unable to recruit MCM to the origin *in vitro* and did not support origin firing *in vivo*. Thus, the

Sensor 2 residue may act to transduce the information of the nucleotide status of Orc1-1 to the conformation of the MCM recruitment site (Samson et al. 2016).

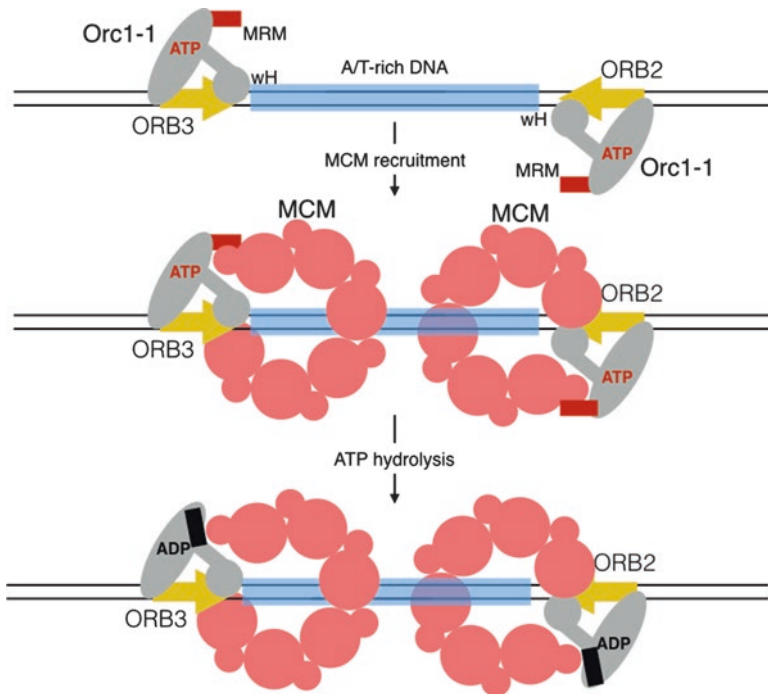
## 5.6 Active Loading or Passive Recruitment of MCM?

Classical views of the MCM helicase portray it as a ring-shaped hexamer (Costa and Onesti 2009). However, structural studies of both eukaryotic MCM2-7 and archaeal MCM have revealed a range of conformations. With regard to the archaeal MCMs, single and double hexamers and heptamers have been described, as have open-ring and even filamentous forms of the protein (Chen et al. 2005; Pape et al. 2003; Slaymaker et al. 2013; Samson and Bell 2016; Samson et al. 2016). There has been considerable debate about how the MCM ring might be opened to allow loading onto DNA (Yardimci and Walter 2014; Sakakibara et al. 2009). With regard to the archaeal protein, a notable electron microscopy study demonstrated that simply heating the MCM of *Methanothermobacter thermautotrophicum* to its normal physiological growth temperature resulted in greater than half of the particles adopting an open-ring conformation (Chen et al. 2005). Similarly, heat treatment of *Sulfolobus* MCM resulted in substantial elevation of recruitment of MCM by Orc1-1 to *oriC1* in vitro (Samson et al. 2016). Thus, based on Orc1-1's monomeric behavior, single-turnover ATP hydrolysis, activity when ATP bound, switch-off upon ATP hydrolysis, and the thermodynamically favored opening of MCM, we have proposed that Orc1-1 is acting as a conditional platform for MCM recruitment to replication origins. Importantly, *oriC1* possesses ORB elements aligned in inverted orientation flanking a ~90 bp AT-rich region. Replication initiation has been mapped at the boundary of this candidate duplex unwinding element, and so it is believed that two hexamers of MCM are loaded into this region by Orc1-1 bound to the flanking ORB elements (Fig. 5.5).

## 5.7 The Archaeal CMG Complex

The molecular basis of how initial DNA unwinding at replication origins is effected remains unknown at this time in both archaeal and eukaryotic systems. In eukaryotes, it is well established that the ultimate activation of the MCM helicase is tightly regulated and involves the facilitated recruitment of Cdc45 and GINS to form an active helicase assembly, termed CMG, that is capable of driving replication fork progression (Bell and Labib 2016).

Eukaryotic GINS is composed of four distinct subunits, Psf1, Psf2, Psf3, and Sld5 (Labib and Gambus 2007; MacNeill 2010). The subunits fall into two classes, related to each other by circular permutation. Psf2 and Psf3 have a domain order BA with a beta-strand domain followed by an alpha-helical domain. In Psf1 and Sld5, the order of the domains is switched to AB. The archaeal orthologs were initially



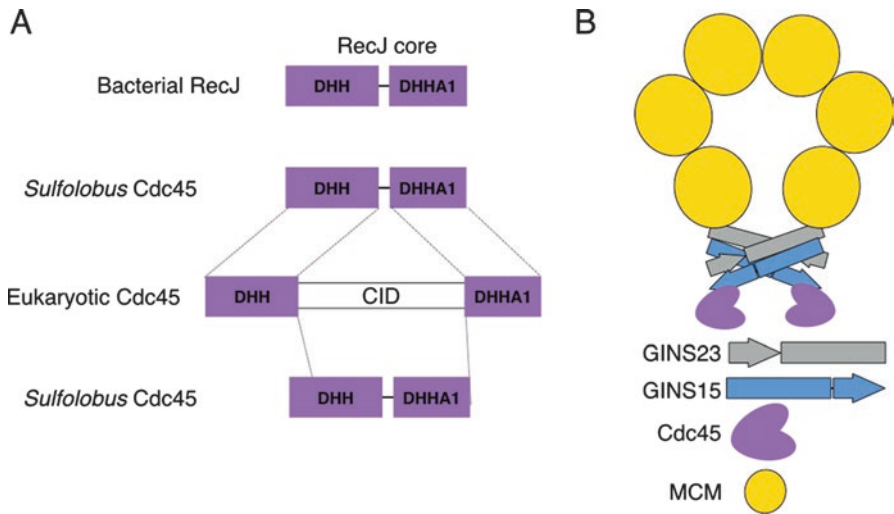
**Fig. 5.5** Model of the ATP-dependent recruitment of MCM by Orc1 proteins. ATP-bound Orc1-1 associates with inverted ORB elements at *oriC1*. As illustrated in Fig. 5.3, Orc1-1 binds to ORB elements as a monomer with a defined polarity – the AAA+ module contacting a G-rich element and the WH domain binding a short inverted repeat. The region between the inverted ORB elements, colored in blue, is highly AT rich. The MRM is positioned such that it can interact with MCM, leading to MCM’s recruitment to the origin with both hexamers encircling double-stranded DNA. Subsequent hydrolysis of ATP to ADP repositions the MRM (shown in black in the “off” state), preventing further rounds of MCM recruitment

identified by virtue of their ability to interact with the N-terminal domains of MCM in a yeast two-hybrid screen. The first archaeal GINS ortholog identified was shown to be related to both Psf2 and Psf3 and was thus named Gins23 (Marinsek et al. 2006). Interestingly, the *gins23* gene is encoded within a bi-cistronic operon with *mcm*. Biochemical studies revealed that Gins23 co-purified with another small protein that was revealed to be related to Psf1 and Sld5 and thus named Gins15. The archaeal GINS assembly was shown to be a tetramer, containing two copies each of Gins15 and Gins23 (Marinsek et al. 2006). While the initial work was performed in *Sulfolobus*, the archaeal GINS complex is now known to be conserved across the archaeal domain of life (MacNeill 2010; Oyama et al. 2011; Yoshimochi et al. 2008; Oyama et al. 2016). During the biochemical isolation of *Sulfolobus* GINS, a further polypeptide co-purified over eight steps and was identified as being related to the DNA-binding fold of the RecJ superfamily of proteins, leading to its initial name of RecJdbh (Marinsek et al. 2006). Subsequent work has revealed an unambiguous

relationship between RecJ and eukaryotic Cdc45, and so RecJdbh has been renamed as Cdc45 (Sanchez-Pulido and Ponting 2011; Xu et al. 2016). Interestingly, Cdc45-related proteins have been identified across the archaeal domain of life but appear phylogenetically diverse (Makarova et al. 2012). One such protein, termed GAN, has been shown to be capable of association with GINS in the organism *Thermococcus kodakarensis* and, intriguingly, appears to be active as a nuclease (Li et al. 2011; Oyama et al. 2016). Recent structural studies have confirmed the GAN•GINS interaction and revealed the basis of the interaction between the GAN and the C-terminal domain of Gins15 (Oyama et al. 2016). Notably, in eukaryotes, an analogous interaction is observed between Psf1's CTD and Cdc45 (Costa et al. 2011).

In *Sulfolobus*, Cdc45 appears to be very tightly associated with GINS as evidenced by their co-purification over multiple steps (Marinsek et al. 2006). Furthermore, experiments with recombinant GINS and Cdc45 have revealed that the Cdc45•GINS complex (termed CG) is resistant to up to 8 M urea (Xu et al. 2016). Chromatin immunoprecipitation experiments have demonstrated that Cdc45 (and by inference, GINS) associates with MCM at replication origins and proceeds with the helicase during DNA synthesis. At the biochemical level, association of CG with MCM leads to a robust stimulation of helicase activity. Importantly, neither Cdc45 nor GINS when individually added to MCM results in detectable stimulation of helicase activity (Xu et al. 2016). While this latter observation agrees with initial reports that *Sulfolobus* GINS did not stimulate MCM's helicase activity (Marinsek et al. 2006), a report from the Huang laboratory has suggested that *Sulfolobus* GINS alone could stimulate MCM (Lang and Huang 2015).

One important difference between the archaeal and eukaryotic Cdc45 and GINS association lies in the composition of the assembly. While both eukaryotic Psf1 and Sld5 possess the AB domain organization, only Psf1 interacts with Cdc45 (Costa et al. 2011). This enforces a stoichiometry of one Cdc45 per GINS complex. In contrast, in the archaeal GINS, two identical copies of Gins15 are present, thus conferring the potential to interact with two Cdc45 molecules per GINS complex. Native electrospray ionization mass spectrometry experiments on the reconstituted *Sulfolobus* CG complex revealed that this was indeed the case, revealing a mass compatible with two copies each of Cdc45, Gins15, and Gins23 (Xu et al. 2016). While it has not been directly determined, it seems likely that this organization will also apply to the euryarchaeal *Thermococcus* GINS•GAN assembly (Oyama et al. 2016). Although this observation suggests a distinct difference between archaeal and eukaryotic CMG, hidden Markov modeling of the predicted structure of *Sulfolobus* Cdc45 revealed a hitherto undocumented similarity with an unanticipated region of eukaryotic Cdc45 (Xu et al. 2016). More specifically, the RecJ fold of eukaryotic Cdc45 is interrupted by a so-called CID domain (Simon et al. 2016). Surprisingly, *Sulfolobus* Cdc45 was predicted to form a structure related to this CID domain. As it had already been documented that *Sulfolobus* Cdc45 has similarities to the RecJ fold, this observation suggests that eukaryotic Cdc45 may have arisen via a gene duplication and internal fusion event, yielding a Russian doll-like organization (Fig. 5.6a). Thus, eukaryotic Cdc45 can be viewed as a pseudodimer when compared to its archaeal antecedents.



**Fig. 5.6** The archaeal CMG complex. **(a)** Relationship between bacterial RecJ and archaeal and eukaryotic Cdc45. The *Sulfolobus* Cdc45 corresponds to the core fold of RecJ – comprised of the DHH and DHHA1 domains. Eukaryotic Cdc45 has these two domains separated by the “CID” domain. Hidden Markov modeling revealed that the CID may have evolved from a partial copy of a core RecJ fold. See Xu et al. (2016) for details. **(b)** Speculative model for the architecture of the archaeal CMG complex. Gins23 and Gins15 are shown in gray and blue, respectively. Their beta-strand-rich domains are shown as arrows and their alpha-helical domains as rectangles. Gins15 and Gins23 form a 2:2 complex. Further, Gins15 interacts with Cdc45, and Gins 23 interacts with MCM. An open-ring form of MCM, such as that loaded on the replication origins, is depicted

Electron microscopy studies of the eukaryotic CMG complex reveal that GINS and Cdc45 interact over the interface between MCM2 and MCM5 subunits (Costa et al. 2011). This interface serves as a gate in the MCM ring, and elegant cross-linking studies have revealed that the ability of this gate to open is key to loading eukaryotic MCM(2-7) at replication origins (Samel et al. 2014). The innate asymmetry of the eukaryotic heterohexameric MCM(2-7) makes it easy to understand how the location and stoichiometry of Cdc45 and GINS association are imposed. This contrasts with the situation in archaea where the MCM is composed of six identical subunits. However, the available data indicate that MCM is recruited to origins in an open-ring form (Samson et al. 2016). It is possible that the nature of the opening between MCM subunits is such that it favors association of CG with that locus on the MCM complex (Fig. 5.6b). It may be significant that CG interacts with MCM’s N-terminal domains via the Gins23 subunit (Marinsek et al. 2006). It is conceivable that the presence of two identical MCM-interaction interfaces on archaeal CG favors interactions between MCM N-terminal domains juxtaposed across the opening in the MCM ring.

In eukaryotes, the sequential and regulated associations of first Cdc45 and then GINS with loaded MCM are pivotal events in the control of the initiation of DNA replication (Siddiqui et al. 2013; Tanaka and Araki 2013; Bell and Labib 2016).

Interestingly, the so-called firing factors that facilitate this process (e.g., Sld2, Sld3, Sld7, Dpb11) are eukaryotic innovations with no discernable homologs in the archaea. Furthermore, the CDK and DDK kinases that in turn govern the behavior of the firing factors are also absent from archaea. The tight association of Cdc45 and GINS in archaeal cell extracts might imply that these factors interact en bloc with origin-associated MCM, leading to activation of MCM's helicase activity. Whether this step in archaeal DNA replication initiation is subject to regulatory control is currently unknown. However, in species such as *Sulfolobus* where multiple replication origins are coordinately regulated to trigger a single initiation event per cell cycle, it is very tempting to speculate that MCM activation by CG could be a key and committing step in regulating replication initiation.

**Acknowledgments** I would like to thank Rachel Samson for helpful discussions of this material.

## References

- Barry ER, Bell SD (2006) DNA replication in the archaea. *Microbiol Mol Biol Rev* 70(4):876–887. <https://doi.org/10.1128/membr.00029-06>
- Bell SD (2012) Archaeal *orc1/cdc6* proteins. *Subcell Biochem* 62:59–69. [https://doi.org/10.1007/978-94-007-4572-8\\_4](https://doi.org/10.1007/978-94-007-4572-8_4)
- Bell SP, Labib K (2016) Chromosome duplication in *Saccharomyces cerevisiae*. *Genetics* 203(3):1027–1067. <https://doi.org/10.1534/genetics.115.186452>
- Bleichert F, Botchan MR, Berger JM (2017) Mechanisms for initiating cellular DNA replication. *Science* 355:eaah6317. <https://doi.org/10.1126/science.aah6317>
- Bochman ML, Schwacha A (2009) The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiol Mol Biol Rev* 73(4):652–683. <https://doi.org/10.1128/membr.00019-09>
- Breuert S, Allers T, Spohn G, Soppa J (2006) Regulated polyploidy in halophilic archaea. *PLoS One* 1(1):e92. <https://doi.org/10.1371/journal.pone.0000092>
- Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P (2008) Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* 6(3):245–252. <https://doi.org/10.1038/nrmicro1852>
- Chen YJ, Yu XO, Kasiviswanathan R, Shin JH, Kelman Z, Egelman EH (2005) Structural polymorphism of *Methanothermobacter thermoautotrophicus* MCM. *J Mol Biol* 346(2):389–394. <https://doi.org/10.1016/j.jmb.2004.11.076>
- Costa A, Onesti S (2009) Structural biology of MCM helicases. *Crit Rev Biochem Mol Biol* 44(5):326–342. <https://doi.org/10.1080/10409230903186012>
- Costa A, Ilves I, Tamberg N, Petojevic T, Nogales E, Botchan MR, Berger JM (2011) The structural basis for MCM2-7 helicase activation by GINS and Cdc45. *Nat Struct Mol Biol* 18(4):471–U110. <https://doi.org/10.1038/nsmb.2004>
- Dueber ELC, Corn JE, Bell SD, Berger JM (2007) Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science* 317(5842):1210–1213. <https://doi.org/10.1126/science.1143690>
- Dueber EC, Costa A, Corn JE, Bell SD, Berger JM (2011) Molecular determinants of origin discrimination by Orc1 initiators in archaea. *Nucleic Acids Res* 39(9):3621–3631. <https://doi.org/10.1093/nar/gkq1308>

- Duggin IG, McCallum SA, Bell SD (2008) Chromosome replication dynamics in the archaeon *Sulfolobus acidocaldarius*. *Proc Natl Acad Sci U S A* 105(43):16737–16742. <https://doi.org/10.1073/pnas.0806414105>
- Edgell DR, Doolittle WF (1997) Archaea and the origin(s) of DNA replication proteins. *Cell* 89(7):995–998
- Forster P (2015) The universal tree of life: an update. *Front Microbiol* 6:717. <https://doi.org/10.3389/fmicb.2015.00717>
- Frols S, Gordon PM, Panlilio MA, Duggin IG, Bell SD, Sensen CW, Schleper C (2007) Response of the hyperthermophilic archaeon *Sulfolobus solfataricus* to UV damage. *J Bacteriol* 189(23):8708–8718. <https://doi.org/10.1128/JB.01016-07>
- Gaudier M, Schuwirth BS, Westcott SL, Wigley DB (2007) Structural basis of DNA replication origin recognition by an ORC protein. *Science* 317(5842):1213–1216. <https://doi.org/10.1126/science.1143664>
- Gotz D, Paytubi S, Munro S, Lundgren M, Bernander R, White MF (2007) Responses of hyperthermophilic crenarchaea to UV irradiation. *Genome Biol* 8(10):R220. <https://doi.org/10.1186/gb-2007-8-10-r220>
- Grainge I, Gaudier M, Schuwirth BS, Westcott SL, Sandall J, Atanassova N, Wigley DB (2006) Biochemical analysis of a DNA replication origin in the archaeon *Aeropyrum pernix*. *J Mol Biol* 363(2):355–369. <https://doi.org/10.1016/j.jmb.2006.07.076>
- Guy L, Ettema TJ (2011) The archaeal ‘TACK’ superphylum and the origin of eukaryotes. *Trends Microbiol* 19(12):580–587. <https://doi.org/10.1016/j.tim.2011.09.002>
- Hawkins M, Malla S, Blythe MJ, Nieduszynski CA, Allers T (2013) Accelerated growth in the absence of DNA replication origins. *Nature* 503(7477):544–547. <https://doi.org/10.1038/nature12650>
- Hildenbrand C, Stock T, Lange C, Rother M, Soppa J (2011) Genome copy numbers and gene conversion in methanogenic archaea. *J Bacteriol* 193:734–743
- Kelman LM, Kelman Z (2014) Archaeal DNA replication. *Annu Rev Genet* 48:71–97. <https://doi.org/10.1146/annurev-genet-120213-092148>
- Labib K, Gambus A (2007) A key role for the GINS complex at DNA replication forks. *Trends Cell Biol* 17(6):271–278. <https://doi.org/10.1016/j.tcb.2007.04.002>
- Lang S, Huang L (2015) The *Sulfolobus solfataricus* GINS complex stimulates DNA binding and processive DNA unwinding by minichromosome maintenance helicase. *J Bacteriol* 197(21):3409–3420. <https://doi.org/10.1128/JB.00496-15>
- Li Z, Pan M, Santangelo TJ, Chemnitz W, Yuan W, Edwards JL, Hurwitz J, Reeve JN, Kelman Z (2011) A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res* 39(14):6114–6123. <https://doi.org/10.1093/nar/gkr181>
- Lundgren M, Andersson A, Chen LM, Nilsson P, Bernander R (2004) Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A* 101(18):7046–7051. <https://doi.org/10.1073/pnas.0400656101>
- Lundgren M, Malandrini L, Eriksson S, Huber H, Bernander R (2008) Cell cycle characteristics of Crenarchaeota: unity among diversity. *J Bacteriol* 190(15):5362–5367. <https://doi.org/10.1128/jb.00330-08>
- Maaty WS, Wiedenheft B, Tarlykov P, Schaff N, Heinemann J, Robison-Cox J, Valenzuela J, Dougherty A, Blum P, Lawrence CM, Douglas T, Young MJ, Bothner B (2009) Something old, something new, something borrowed; how the thermoacidophilic archaeon *Sulfolobus solfataricus* responds to oxidative stress. *PLoS One* 4(9):e6964. <https://doi.org/10.1371/journal.pone.0006964>
- MacNeill SA (2010) Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochem J* 425(3):489–500. <https://doi.org/10.1042/BJ20091531>
- 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. <https://doi.org/10.1186/1745-6150-7-7>

- Majernik AI, Chong JP (2008) A conserved mechanism for replication origin recognition and binding in archaea. *Biochem J* 409(2):511–518
- Marinsek N, Barry ER, Makarova KS, Dionne I, Koonin EV, Bell SD (2006) GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep* 7(5):539–545. <https://doi.org/10.1038/sj.embor.7400649>
- McGeoch AT, Bell SD (2008) Extra-chromosomal elements and the evolution of cellular DNA replication machineries. *Nat Rev Mol Cell Biol* 9(7):569–574. <https://doi.org/10.1038/nrm2426>
- Michel B, Bernander R (2014) Chromosome replication origins: do we really need them? *BioEssays* 36(6):585–590. <https://doi.org/10.1002/bies.201400003>
- Myllykallio H, Lopez P, Lopez-Garcia P, Heilig R, Saurin W, Zivanovic Y, Philippe H, Forterre P (2000) Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* 288(5474):2212–2215
- Naor A, Lapiere P, Mevarech M, Papke RT, Gophna U (2012) Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr Biol* 22(15):1444–1448. <https://doi.org/10.1016/j.cub.2012.05.056>
- Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, Allers T (2007) Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet* 3:729–743. <https://doi.org/10.1371/journal.pgen.0030077>
- Oyama T, Ishino S, Fujino S, Ogino H, Shirai T, Mayanagi K, Saito M, Nagasawa N, Ishino Y, Morikawa K (2011) Architectures of archaeal GINS complexes, essential DNA replication initiation factors. *BMC Biol* 9:28. <https://doi.org/10.1186/1741-7007-9-28>
- Oyama T, Ishino S, Shirai T, Yamagami T, Nagata M, Ogino H, Kusunoki M, Ishino Y (2016) Atomic structure of an archaeal GAN suggests its dual roles as an exonuclease in DNA repair and a CMG component in DNA replication. *Nucleic Acids Res* 44(19):9505–9517. <https://doi.org/10.1093/nar/gkw789>
- Pape T, Meka H, Chen SX, Vicentini G, van Heel M, Onesti S (2003) Hexameric ring structure of the full-length archaeal MCM protein complex. *EMBO Rep* 4(11):1079–1083. <https://doi.org/10.1038/sj.embor.7400010>
- Pelve EA, Lindas AC, Knoppel A, Mira A, Bernander R (2012) Four chromosome replication origins in the archaeon *Pyrobaculum calidifontis*. *Mol Microbiol* 85(5):986–995. <https://doi.org/10.1111/j.1365-2958.2012.08155.x>
- Pelve EA, Martens-Habbena W, Stahl DA, Bernander R (2013) Mapping of active replication origins in vivo in thaum- and euryarchaeal replicons. *Mol Microbiol* 90(3):538–550. <https://doi.org/10.1111/mmi.12382>
- Rivera MC, Lake JA (2004) The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* 431(7005):152–155. <https://doi.org/10.1038/nature02848>
- Robinson NP, Bell SD (2007) Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc Natl Acad Sci U S A* 104(14):5806–5811. <https://doi.org/10.1073/pnas.0700206104>
- Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, Bell SD (2004) Identification of two origins of replication in the single chromosome of the Archaeon *Sulfolobus solfataricus*. *Cell* 116(1):25–38
- Robinson NP, Blood KA, McCallum SA, Edwards PAW, Bell SD (2007) Sister chromatid junctions in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *EMBO J* 26(3):816–824. <https://doi.org/10.1038/sj.emboj.7601529>
- Sakakibara N, Kelman LM, Kelman Z (2009) How is the archaeal MCM helicase assembled at the origin? Possible mechanisms. *Biochem Soc Trans* 37(Pt 1):7–11. <https://doi.org/10.1042/BST0370007>
- Samel SA, Fernandez-Cid A, Sun J, Riera A, Tognetti S, Herrera MC, Li H, Speck C (2014) A unique DNA entry gate serves for regulated loading of the eukaryotic replicative helicase MCM2-7 onto DNA. *Genes Dev* 28(15):1653–1666. <https://doi.org/10.1101/gad.242404.114>
- Samson RY, Bell SD (2014) Archaeal chromosome biology. *J Mol Microbiol Biotechnol* 24(5–6):420–427. <https://doi.org/10.1159/000368854>



- Samson RY, Bell SD (2016) Archaeal DNA replication origins and recruitment of the MCM replicative helicase. *Enzyme* 39:169–190. <https://doi.org/10.1016/bs.enz.2016.03.002>
- Samson RY, Xu Y, Gadelha C, Stone TA, Faqiri JN, Li D, Qin N, Pu F, Liang YX, She Q, Bell SD (2013) Specificity and function of archaeal DNA replication initiator proteins. *Cell Rep* 3(2):485–496. <https://doi.org/10.1016/j.celrep.2013.01.002>
- Samson RY, Abeyrathne PD, Bell SD (2016) Mechanism of archaeal MCM helicase recruitment to DNA replication origins. *Mol Cell* 61(2):287–296. <https://doi.org/10.1016/j.molcel.2015.12.005>
- Sanchez-Pulido L, Ponting CP (2011) Cdc45: the missing RecJ ortholog in eukaryotes? *Bioinformatics* 27(14):1885–1888. <https://doi.org/10.1093/bioinformatics/btr332>
- Sherratt DJ (2003) Bacterial chromosome dynamics. *Science* 301(5634):780–785
- Siddiqui K, On KF, Diffley JF (2013) Regulating DNA replication in eukarya. *Cold Spring Harb Perspect Biol* 5(9):a010371. <https://doi.org/10.1101/cshperspect.a012930>
- Simon AC, Sannino V, Costanzo V, Pellegrini L (2016) Structure of human Cdc45 and implications for CMG helicase function. *Nat Commun* 7:11638. <https://doi.org/10.1038/ncomms11638>
- Slaymaker IM, Fu Y, Toso DB, Ratanunga N, Brewster A, Forsburg SL, Zhou ZH, Chen XS (2013) Mini-chromosome maintenance complexes form a filament to remodel DNA structure and topology. *Nucleic Acids Res* 41(5):3446–3456. <https://doi.org/10.1093/nar/gkt022>
- Tanaka S, Araki H (2013) Helicase activation and establishment of replication forks at chromosomal origins of replication. *Cold Spring Harb Perspect Biol* 5(12):a010371. <https://doi.org/10.1101/cshperspect.a010371>
- Tiengwe C, Marcello L, Farr H, Gadelha C, Burchmore R, Barry JD, Bell SD, McCulloch R (2012) Identification of ORC1/CDC6-interacting factors in *Trypanosoma brucei* reveals critical features of origin recognition complex architecture. *PLoS One* 7(3):e32674. <https://doi.org/10.1371/journal.pone.0032674>
- Williams TA, Foster PG, Cox CJ, Embley TM (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504(7479):231–236. <https://doi.org/10.1038/nature12779>
- Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74(11):5088–5090
- 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(47):13390–13395. <https://doi.org/10.1073/pnas.1613825113>
- Yang H, Wu Z, Liu J, Liu X, Wang L, Cai S, Xiang H (2015) Activation of a dormant replication origin is essential for *Haloferax mediterranei* lacking the primary origins. *Nat Commun* 6:8321. <https://doi.org/10.1038/ncomms9321>
- Yao NY, O'Donnell ME (2016) Evolution of replication machines. *Crit Rev Biochem Mol Biol* 51(3):135–149. <https://doi.org/10.3109/10409238.2015.1125845>
- Yardimci H, Walter JC (2014) Prereplication-complex formation: a molecular double take? *Nat Struct Mol Biol* 21(1):20–25. <https://doi.org/10.1038/nsmb.2738>
- Yoshimochi T, Fujikane R, Kawanami M, Matsunaga F, Ishino Y (2008) The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. *J Biol Chem* 283(3):1601–1609. <https://doi.org/10.1074/jbc.M707654200>
- Zerulla K, Soppa J (2014) Polyploidy in haloarchaea: advantages for growth and survival. *Front Microbiol* 5:274. <https://doi.org/10.3389/fmicb.2014.00274>
- Zerulla K, Chimileski S, Nather D, Gophna U, Papke RT, Soppa J (2014) DNA as a phosphate storage polymer and the alternative advantages of polyploidy for growth or survival. *PLoS One* 9(4):e94819. <https://doi.org/10.1371/journal.pone.0094819>