the FEBS Journal

REVIEW ARTICLE

Origins of DNA replication in the three domains of life

Nicholas P. Robinson and Stephen D. Bell

MRC Cancer Cell Unit, Hutchison MRC Research Centre, Hills Road, Cambridge, UK

Keywords
Cdc6; DnaA; DNA Replication; MCM; ORC

Correspondence
S. D. Bell, MRC Cancer Cell Unit, Hutchison MRC Research Centre, Hills Road, Cambridge, CB2 2XZ, UK
Fax: +44 1223 763296
Tel: +44 1223 763311
E-mail: sb419@hutchison-mrc.cam.ac.uk

(Received 8 April 2005, revised 11 May 2005, accepted 13 May 2005)

doi:10.1111/j.1742-4658.2005.04768.x

Replication of DNA is essential for the propagation of life. It is somewhat surprising then that, despite the vital nature of this process, cellular organisms show a great deal of variety in the mechanisms that they employ to ensure appropriate genome duplication. This diversity is manifested along classical evolutionary lines, with distinct combinations of replicon architecture and replication proteins being found in the three domains of life: the Bacteria, the Eukarya and the Archaea. Furthermore, although there are mechanistic parallels, even within a given domain of life, the way origins of replication are defined shows remarkable variation.

The origin of origins

In a now classic 1963 paper, Jacob, Brenner & Cuzin proposed that, in a manner analogous to the interaction of *trans*-acting regulators with *cis*-acting operators in control of gene expression, an initiator factor would act at a replicator sequence in the chromosome to control and facilitate DNA replication [1]. However, in contrast to the then prevalent models for negative regulation of gene expression, it was proposed that the replication initiator factor would act positively to promote replication at the replicator, or as it is now named, origin of replication. In the following 40 years much has been learnt about the nature of initiators and origins of replication, particularly in simple model systems. However, many of the molecular details of the basis of origin selection remain poorly understood, particularly in higher eukaryotes.

Bacteria

In bacteria the origin of replication is termed *oriC*, and typically a single origin exists per bacterial chromosome [2]. In *Escherichia coli*, *oriC* is located between the *gldA* and *mioC* genes. The ≈ 250 bp *oriC* region contains multiple repeated sequences containing a nine base pair consensus element termed the DnaA box [3]. Other bacteria also possess single origins of replication with multiple DnaA boxes although both the precise number and distribution of these boxes vary between species [4]. Interestingly, in many bacteria the origin of replication is found adjacent to the gene for DnaA itself, suggesting a mechanism for the coordinate control of origin activity and levels of initiator proteins [4]. An individual consensus DnaA box is bound by a monomer of the DnaA protein and this interaction induces a sharp bend in the binding site [5]. However, in natural bacterial origins there are multiple DnaA boxes and these orchestrate complex cooperative binding events to DnaA boxes with varying degrees of conformity to the consensus sequence. A particularly interesting ramification of this is that a DnaA box with poor conservation to the consensus may not be able to bind DnaA on its own. However, binding to this ‘weak’ site can be facilitated by binding of DnaA to an adjacent high affinity consensus site [4].

Abbreviations
ACS, ARS-consensus sequence; ARS, autonomously replicating sequences; DBD, DNA binding domain; MCM, minichromosomal maintenance; ORB, origin recognition box; ORC, origin recognition complex; pre-RC, prereplicative complex.

Bacterial origins of replication also possess a second conserved element, a highly AT rich region. The unwinding of this intrinsically meltable DNA is a key step in replication initiation at origins. Under highly defined *in vitro* conditions, DnaA is capable of mediating partial unwinding of this region on its own (Fig. 1). It appears therefore, that the combination of DNA bending induced by DnaA and the cooperative interactions between DnaA monomers on DNA result in local topological tension that manifests itself by unwinding of this intrinsically less stable region of duplex [6].

A further level of complexity arises from the fact that DnaA is a member of the AAA$^+$ family of ATPases. This class of protein possess a nucleotide-binding domain that can bind ATP and catalyse its hydrolysis. The conformation of the AAA$^+$ domain alters depending on the phosphorylation status of the bound nucleotide. Furthermore, AAA$^+$ proteins often exist as multimers and neighbouring subunits communicate by extending so-called arginine fingers into the ATP binding subunit of a neighbour [7]. Thus, there is the capacity to transduce the effects of ATP to ADP hydrolysis in one subunit through a network of interacting proteins. It has been found that although ADP–DnaA and ATP–DnaA have similar affinities for consensus DnaA boxes [8], the ATP bound form is able to recognize an additional six base pair element, providing a consensus 'strong' DnaA box is present in the vicinity. In addition, single stranded versions of these 'ATP–DnaA box' hexameric sequences can also be recognized by ATP–DnaA. Six of these ATP–DnaA boxes are found in the AT rich region of *E. coli oriC*. Thus, once the topological tension induced by DnaA binding has melted this region, the exposed single stranded ATP–DnaA boxes can be bound by ATP–DnaA stabilizing the DNA in the melted form [4,8].

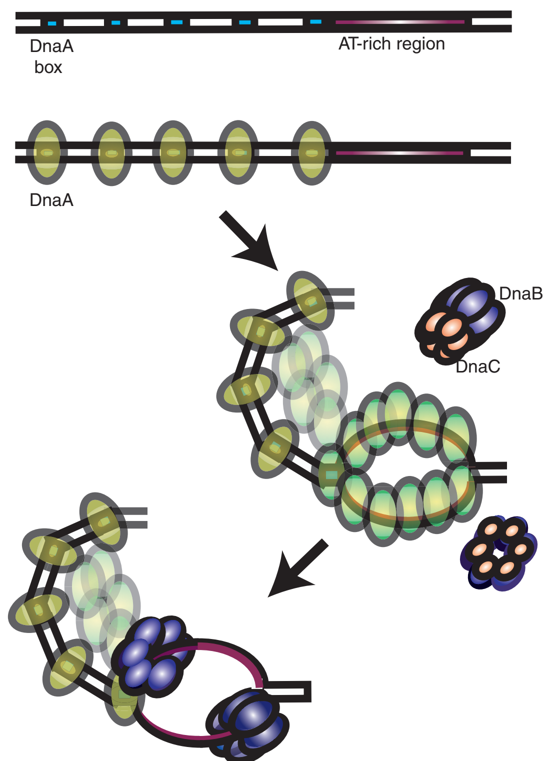

Fig. 1. Cartoon of the assembly of the DNA replication machinery on the *E. coli oriC* region. Binding of DnaA (green ovals) to the DnaA boxes (blue boxes) is shown. This leads to local DNA distortion and facilitates binding of DnaA to melted DNA in an AT-rich region (purple region). DnaB (blue), in a complex with DnaC (peach) is recruited to the melted region, followed by disassociation of DnaC as detailed in the text.

It is apparent therefore, that the combination of multiple DNA recognition sites, distortion of duplex DNA, and cooperative interactions between DNA bound initiator proteins at bacterial origins leads to a complex nucleoprotein architecture, the precise stoichiometry of which remains unclear, that both mediates the initial melting and stabilizes the resultant single stranded DNA. It therefore comes as no surprise that bacterial architectural chromatin proteins such as HU and IHF play important roles in facilitating the assembly of this complex [9].

Once the melted origin–DnaA complex has formed, the replicative helicase DnaB can be loaded. Although DnaA interacts physically with DnaB [4], this reaction requires the action of another protein, DnaC [10–13]. Interestingly DnaC, like DnaA, is a member of the AAA$^+$ family of ATPases, however, the role of ATP appears to be rather more subtle than simply being required as an energy source to facilitate loading of the ring shaped DnaB helicase. Indeed, ATP hydrolysis by DnaC is not required for DnaB loading as this reaction can be performed by ATP–DnaC, ADP–DnaC or even nucleotide free forms of DnaC [14]. Rather the role of ATP in the reaction is to serve as a switch that controls the activity of the helicase. The ATP bound form of DnaC severely inhibits the helicase activity of DnaB and also increases the affinity of DnaC for single stranded DNA. In contrast ADP–DnaC does not inhibit DnaB helicase and has lower DNA binding affinity. Thus, it has been proposed that ATP–DnaC interacts with DnaB and facilitates loading of the helicase onto the single stranded region of the melted origin. However, the high affinity of ATP–DnaC for DNA effectively glues DnaB to the origin, preventing its translocation and therefore suppressing its helicase activity. Subsequently, the hydrolysis of

ATP to ADP by DnaC releases DnaB allowing it to act as the replicative helicase [14].

Eukaryotic origins

The identification of initiation sites in eukaryotic organisms has been an arduous task. In contrast to the single, clearly defined sites of bacterial replication, eukaryotic DNA synthesis commences from hundreds or even thousands of origins, which rarely contain obvious sequence motifs, and are often difficult to characterize (reviewed in [15–19]). This complexity is compounded by the fact that eukaryotic origin activation is asynchronous. In addition, initiation site usage displays considerable flexibility under varying growth conditions, or throughout different stages of development. In more recent years, it has become increasingly apparent that epigenetic factors govern the regulation of eukaryotic origin activity (reviewed in [15,16]). These modulations provide the elasticity necessary for coordinated initiation from multiple sites.

Lessons from budding yeast

Although eukaryotic replication initiation is inevitably more complicated than the bacterial process, some parallels can be drawn between the two systems. These similarities are perhaps most obvious in the budding yeast *Saccharomyces cerevisiae*, where conserved sequence motifs have been identified at the origins. Budding yeast initiation sites, or autonomously replicating sequences (ARS), are noncoding regions of DNA, approximately 100–200 bp in length. These sites encompass the short, highly conserved and essential ARS-consensus sequence (ACS or A element), and more divergent motifs known as B elements [20,21]. It is important to note, however, that *S. cerevisiae* and its close relatives appear to be the only eukaryotic organisms that utilize specific sequence elements within its origins. Fortuitously, these conserved elements were instrumental to the isolation of the origin recognition complex (ORC) [22]. This complex, constituted by the interaction of six closely associated proteins (Orc1–6) (reviewed in [21]), has been identified as the eukaryotic replication initiator, performing an analogous function to bacteria DnaA. Before DNA synthesis commences, ORC recruits a number of additional proteins to the origin to form the prereplicative complex (pre-RC), licensing the site for initiation (Fig. 2; reviewed in [23–25]). As seen in bacteria, a key step in origin function is the recruitment of the replicative helicase. In eukaryotes, the hexameric minichromosomal maintenance (MCM) complex, composed of the six related proteins Mcm2–7, is the most probable candidate for this role. As in bacteria, this recruitment requires supplementary proteins, and the factors Cdc6 and Cdt1 have been shown to be critical for the loading process [20,23–25]. Interestingly, Cdc6 displays homology to the Orc1 subunit and thus Cdc6 and Orc1 are presumably derived from a common ancestor (see below). In addition, it has recently been demonstrated that ORC itself is also actively involved in the MCM assembly in budding yeast [26]. The formation and activation of the pre-RC is crucial to the regulation of replication, ensuring that any potential origin can only fire once per cell cycle [23,24].

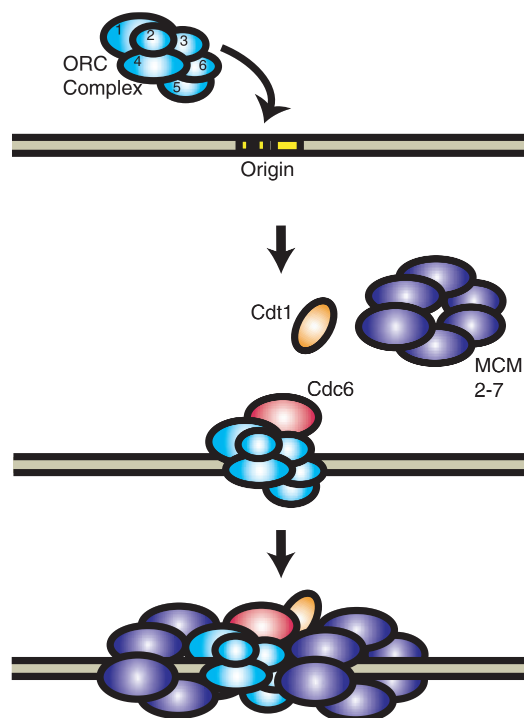

Fig. 2. Model for recruitment of the eukaryotic minichromosomal maintenance (MCM) complex to an origin of replication. The origin is first bound by the heterohexameric ORC (light blue). Cdc6 (red) is then recruited and, in conjunction with Cdt1 (orange), recruits the MCM complex (dark blue) to the origin.

Again, as in bacteria, multiple components of the eukaryotic pre-RC possess ATPase domains. More specifically, ORC subunits Orc1, 4 and 5 have AAA^{+} domains, as does Cdc6, and genetic studies have revealed that mutation of the ATP binding sites in either ORC or Cdc6 impairs the loading of the MCM helicase onto origins. Indeed, with the exception of *Schizosaccharomyces pombe* discussed below, all characterized eukaryotic ORCs require ATP to bind DNA [21]. Although at this level the loading of MCM may

appear superficially similar to the bacterial system, there are a number of fundamental differences between the bacterial and eukaryotic systems. First, in the bacterial system, once DnaA has formed the appropriate open form of the origin, DnaB is loaded and replication initiates. In contrast, there is no evidence that ORC melts the DNA [27] and, additionally, ORC remains bound at origins throughout the cell cycle [21]. A second difference between the systems lies in the observation that, in bacteria, a single pair of DnaB helicases is recruited [13]. In contrast, many MCM molecules are loaded per origin (reviewed in [28]).

Interestingly, this iterative loading of MCM has recently been shown to be dependent on the ATPase activity of ORC [26]. More specifically, mutational analysis of the arginine finger of Orc4 resulted in a mutant protein complex that could still bind both ATP and DNA, but had impaired ATPase activity. This mutant complex supported a single round of MCM loading but was unable to mediate the iterative loading. Thus, it appears that the eukaryotic initiator complex, ORC, plays an active role in the helicase loading process [26].

Finally, it has become apparent that the status of chromatin at a given origin in budding yeast can have important consequences for origin activity. For example, the positioning of nucleosomes at an origin can be influenced by the binding of ORC [29]. Additionally, the acetylation status of chromatin can influence timing of origin firing in budding yeast. More specifically, it has been demonstrated that deletion of the histone deacetylase, RPD3, results in earlier firing of origins in *S. cerevisiae* [30]. Thus, it appears that epigenetic factors have the capacity to regulate origin activity in this model eukaryote.

Origins in other eukaryotes

Although homologues of the subunits of the ORC initiator have been identified in every eukaryotic organism analysed thus far, origin sequences have proven to be considerably more elusive. Indeed, even in budding yeast, origin definition is not as simple as is often portrayed. For example, it is problematical to predict yeast origins by sequence alone, because a large number of candidate ACS elements within the genome do not coincide with initiation zones, and, additionally, some ACS elements deviate from the consensus [31]. Furthermore, intricate compound origins, which contain multiple ACS elements, have also been described [32]. Recently, the use of global, microarray-based techniques have circumvented the difficulties associated with sequence based searches, and successfully mapped the distribution of replication origins throughout the budding yeast genome [33,34]. In addition to identifying novel replication origins, these analyses have also revealed valuable information regarding the duplication of the genome.

Fission yeast

Origins of replication have been identified in fission yeast, *S. pombe*. Remarkably, these show little similarity to those of budding yeast and lack detectable consensus sequences. The principal feature common to *S. pombe* origin regions is that they are rich in A and T bases. Intriguingly, ORC from *S. pombe* recognizes origins solely via a unique feature, an AT-hook DNA binding domain on the Orc4 subunit [35–37], and additionally, *S. pombe* ORC does not require ATP to bind to origins. Recent analyses have suggested that *S. pombe* may have quite relaxed constraints for what constitutes an origin of replication. First, A + T rich regions of the genome were identified bioinformatically. Twenty of these AT rich islands were chosen at random and tested in 2D gel analysis to look for possible origin activity. Eighteen of these regions showed clear evidence of replication intermediates indicative of origin activity [38]. More recently, with reference to the *S. pombe* genome sequence it was noted that the features shared by characterized origins of replication, namely AT-richness and asymmetric strand composition, were common to many intergenic regions in this organism's genome. Using a genetic screen for origin activity, it was found that four of 26 intergenic regions tested had the ability to support maintenance of an episome [39]. Furthermore, dimerization of the intergenic regions led to the discovery that an additional 10 of the intergenic regions could function as origins in this context [37]. In light of these data, it has been proposed that *S. pombe* uses a mode of replication distinct from the original replicon hypothesis [39]. Thus, instead of depending on a highly selective system as in bacteria or even budding yeast, *S. pombe* appears to have little sequence dependence in selection of origins; rather it makes use of a relatively promiscuous DNA binding motif to direct binding of ORC to common features in the genome. Consequently, origin selection in *S. pombe* may be a rather stochastic phenomenon. Furthermore, as the AT-rich regions map to intergenic regions it is possible that origin selectivity may be in part governed by epigenetic phenomena such as the state of chromatin in these intergenic regions. In this light, it is tempting to speculate that the ability of these AT-rich intergenic regions to function as origins of replication in *S. pombe* may correlate with the status of promoters for the encompas-

sing genes. This could manifest itself both at the level of the immediate chromatin environment of the intergenic region, and also at the level of topological status of the DNA as a result of transcription of the adjacent genes.

Origins in higher eukaryotes

Although ORC is conserved in higher eukaryotes, and is clearly essential for replication, the molecular basis of origin identity and function remains poorly understood. Indeed, early studies revealed that a strikingly diverse range of molecules, even from completely heterologous sources, could be replicated in *Xenopus* cell-free systems (reviewed in [40]). One of the best characterized higher eukaryotic origins lies in the chorion amplification locus in *Drosophila melanogaster*. This region undergoes a dynamic localized amplification by multiple rounds of re-replication during oocyte development. Analyses have revealed that the *ACE3* and *ori-β* elements, important for the amplification, are bound by the *Drosophila* ORC [41]. In addition, a complex containing the *Drosophila* homologue of the Myb transcription factor also binds both these elements [42]. Also, immunoprecipitation experiments suggest direct interactions between Myb and the ORC, and cells mutant in Myb showed drastically reduced levels of DNA replication [40]. These data, in conjunction with the observation that Myb is required for S-phase progression in many (although not all) *Drosophila* cell types, suggest a broad role for Myb in DNA replication [42]. Furthermore, interplay between transcription factors and the replication machinery may be instrumental in exerting developmental control of DNA replication in tissue- and temporal-specific manners. More recently, a study has revealed that the chromatin status of the chorion amplification locus has an important role in governing origin activity. The chorion amplification loci were found to co-locate with hyperacetylated histone H4 [43]. More generally, either genetic or chemical reduction of histone deacetylase levels resulted in elevated replication throughout the genome, suggesting a causal link between histone acetylation and replication. Significantly, tethering a histone deacetylase to the chorion amplification locus resulted in a local repression of replication and conversely, tethering a histone acetylase resulted in local stimulation of replication. Thus, it appears that the local epigenetic or structural status of the chromatin in the vicinity of an origin can influence the activity of this region [43]. It is possible therefore, that the stimulatory effect that *Drosophila* Myb has on replication may in part be due to its recruitment of chromatin modifying activities. The interplay between the transcription and replication machineries has been further underscored by a microarray-based analysis of replication and transcription profiles of the left arm of *Drosophila* chromosome 2 [44]. This work revealed that early replicating regions correlated with transcriptionally active locations. Furthermore, these early replicating regions also correlated with ORC binding sites. These sites showed a preponderance of AT-rich regions and generally fell within intergenic regions. Interestingly, there was also significant overlap between ORC and RNA polymerase II binding sites [44]. This latter finding further emphasizes the connection between transcription and replication apparatuses and, as discussed above, suggests that gene specific transcription factors could facilitate ORC recruitment, either via direct protein–protein interaction or by generating a chromatin environment favourable to ORC binding.

This interplay between transcription and replication machineries has also been observed in *Xenopus* cell-free systems. Plasmid DNA introduced into *Xenopus* egg extracts forms chromatin and replication initiates at random positions around the plasmid. However, when a plasmid containing a strong promoter is introduced under conditions where that promoter is active, the plasmid shows preferential replication initiation in the vicinity of the promoter [45]. Interestingly, transcription was not required for the localization of origin activity, indeed the potent activator, GAL4-VP16, alone, is capable of specifying initiation location. It is likely therefore that GAL4-VP16 is acting to facilitate an open chromatin structure conducive to pre-RC assembly. Consistent with this possibility, it was found that there was increased histone H3 acetylation in the vicinity of the localized replication initiation. Interestingly, this study found that while ORC was associated with plasmid DNA it did not show any preferential localization, even in the presence of the GAL4-VP16, suggesting that it may be bound randomly but activated in a locus specific manner [45].

Another study has also found a close relationship between promoter activity and origin function, in this case in the context of a mammalian episome. The plasmid pEPI-1 replicates stably in a once per cell cycle manner in a range of mammalian cell lines. Recent work has shown that stable replication is dependent on the presence of the strong CMV promoter in the plasmid [46]. However, attempts to map replication initiation sites on the plasmid revealed that initiation occurred at apparently random positions around the episome. Similarly, no distinct or preferred localization of the ORC was detected [47]. Given the dependence of replication on the presence of the CMV promoter it is again tempting to speculate that chromatin remodel-

ling activities recruited by *trans*-activators bound to the promoter facilitate the generation of a permissive chromatin structure in the episome. In addition, it is possible that the circular nature and small size (< 7000 bp) of the episome may have topological consequences that also promote binding of ORC. Indeed, it has recently been demonstrated that purified *Drosophila* ORC has little or no sequence specificity in binding site selection but does show a considerable (roughly 30-fold) preference for negatively supercoiled DNA [48].

Thus, eukaryotes appear to use a striking diversity of mechanisms to define origins of replication, ranging from high affinity sequence specific binding to apparently sequence nonspecific but topology-dependent binding. Additionally, epigenetic phenomena clearly play an important role in governing the selectivity of origin usage. Finally, the observation that Myb may directly interact with ORC opens the possibility of facilitated recruitment of ORC to developmentally regulated sites within the chromosome.

Archaea

In contrast to the wealth of molecular, genetic and biochemical detail that is now known about origins of replication and their interaction with initiators in bacteria and eukaryotes, very little is known about the molecular basis of replication initiation in the third domain of life, the archaea.

It is well established that archaea possess an intriguing blend of bacterial and eukaryotic features as well as aspects that are unique to this domain of life. Archaeal chromosomes resemble those of most bacteria, being small, circular and having polycistronic transcription units. In addition, archaea are likely to have coupled transcription and translation. However, it has become apparent that the core information processing machineries of the archaea are fundamentally related to those of eukaryotes. Thus, the transcription and DNA replication machineries of archaea are closely related to, but significantly simpler than, their eukaryotic counterparts and distinct from those of bacteria [49,50]. Therefore, archaea present themselves as a potentially simple model system to understand the conserved events in DNA replication. A number of studies have described the biochemical properties of archaeal DNA replication proteins (reviewed in [50]). It is also of considerable interest to understand how the simple bacterial-like chromosomes of the archaea are replicated by a eukaryotic-type replication apparatus, to elucidate the nature of the archaeal replicon organization, and to establish the mechanisms by which archaeal replication origins are defined.

Initial attempts to identify archaeal origins of replication were bioinformatic in nature, exploiting the observation that leading and lagging strands often have differential nucleotide composition. Such analyses led to the prediction of the existence of single origins of replication in *Methanobacterium thermoautotrophicum* (now called *Methanothermobacter thermoautotrophicus*) and *Pyrococcus horikoshii* [51]. Subsequent work confirmed the position of the origin of replication in *Pyrococcus*, providing the first experimental proof of a localized origin of replication in the archaea [52]. Interestingly, in a situation reminiscent of that in several bacteria where their origin is adjacent to the gene for the initiator, DnaA, the single *Pyrococcus* origin, termed *oriC*, lies immediately upstream of the gene for the candidate replication initiator protein, a homologue of Orc1 and Cdc6 [52]. As mentioned above, eukaryotic Orc1 and Cdc6 proteins show sequence similarity and are presumably derived from a common ancestor. Archaeal genomes encode proteins that are approximately equally related to both Orc1 and Cdc6, and although individual genome projects variously refer to these as Orc or Cdc6 in this review we shall describe these proteins as Orc1/Cdc6. Fine mapping of the *Pyrococcus* replication origin *in vivo* revealed that the start site of leading strand synthesis was adjacent to a repeat motif of unknown function present in two inverted copies in the *Pyrococcus oriC* [53]. Additionally chromatin immunoprecipitation studies indicated that, *in vivo*, the product of the *orc1/cdc6* gene was associated specifically with the origin of replication [54]. Thus, it appears that in *Pyrococcus*, there is a bacterial-like replicon architecture with a single origin of replication that is recognized (and presumably defined) by a homolog of components of the eukaryotic pre-RC.

A genetic study in a second archaeal species, *Halobacterium* NRC-1 provided evidence for an origin of replication adjacent to the *orc7* gene that encodes the orthologue of the *Pyrococcus* Orc1/Cdc6 protein [55]. Interestingly, *Halobacterium* encodes a total of 10 Orc1/Cdc6 homologues and has three distinct replicons; a main chromosome and two large plasmids. The large chromosome encodes four Orc1/Cdc6 homologues and the remaining homologues are encoded on the plasmids. However, only the *orc7* gene on the main chromosome appears to be associated with an origin of replication. Whether additional origins exist elsewhere on the *Halobacterium* main chromosome remains unknown. Intriguingly, a bioinformatics study has suggested that a second origin may exist in

Halobacterium [56], but attempts to identify this candidate origin experimentally have been unsuccessful [55]. Thus, the available evidence points to both *Pyrococcus* and *Halobacterium* main chromosomes having a bacterial-like situation of a single origin of replication.

A very different situation has been shown to exist in the hyperthermophilic archaeon *Sulfolobus solfataricus*. This organism belongs to the Crenarchaea, a distinct Kingdom from *Halobacterium* and *Pyrococcus* (both Euryarchaea). *S. solfataricus* encodes three Orc1/Cdc6 homologues and, in a systematic 2D gel mapping approach [57], it was demonstrated that origins of replication, termed *oriC1* and *oriC2*, are closely linked to two of these genes (*cdc6-1* and *cdc6-3*). The initiation points of replication were mapped at both origins and found to lie in an AT-rich region (Fig. 3). This region was flanked by various repeat motifs and these were found to be binding sites for the Orc1/Cdc6 proteins. The *oriC1* origin is located upstream of the *cdc6-1* gene, encoding the *Sulfolobus* ortholog of the *Pyrococcus* Orc/Cdc6 and *Halobacterium* Orc7 proteins. Moreover, the sequence elements bound by Cdc6–1 at *Sulfolobus oriC1* are related to sequence repeats at both *Halobacterium* and *Pyroccocus* origins. Indeed, these conserved motifs, termed origin recognition box (ORB) elements, in both *Pyrococcus* and *Halobacterium*, can be recognized by purified *Sulfolobus* Cdc6-1 protein [57]. Thus it appears that these ORB elements, like DnaA boxes in bacteria, are conserved features of a number of archaeal origins of replication and this has allowed the prediction of the localization of replication origins in a diverse range of archaea. Interestingly, the second *Sulfolobus* origin has sequence repeats that are related to a core inverted repeat present in the full ORB elements. These shorter elements, termed mini-ORBs, were also capable of binding Cdc6-1 but did so with at least 10-fold lower affinity than ORB elements [57]. Mini-ORBs also appear broadly conserved and have recently been identified in the predicted origin in *M. thermoautotrophicus* [58].

The presence of broadly conserved Orc1/Cdc6 binding sites in archaea is reminiscent of DnaA boxes in bacteria. The parallel with the bacterial system can be further extended with the elucidation of the crystal structures of DnaA [59] and Orc1/Cdc6 proteins [60,61]. As can be seen in Fig. 4, both proteins possess N-terminal AAA$^+$ domains and C-terminal DNA binding domains (DBDs). In DnaA, the DBD contains a helix-turn-helix; in the archaeal proteins, the DBD has a winged helix domain (reviewed in [62]). Interestingly, the relative position of the AAA$^+$ domain and the winged helix domain of *Aeropyrum pernix* Orc1/Cdc6 homolog was influenced by the nature of the nucleotide bound by the protein, suggesting that binding and hydrolysis of ATP might modulate the nature of the protein–DNA interaction [61]. Intriguingly, however, biochemical studies with the *S. solfataricus* Cdc6-1 protein did not detect any significant effect of the presence or absence of ATP or ADP on the ability of this protein to bind to ORB elements [57].

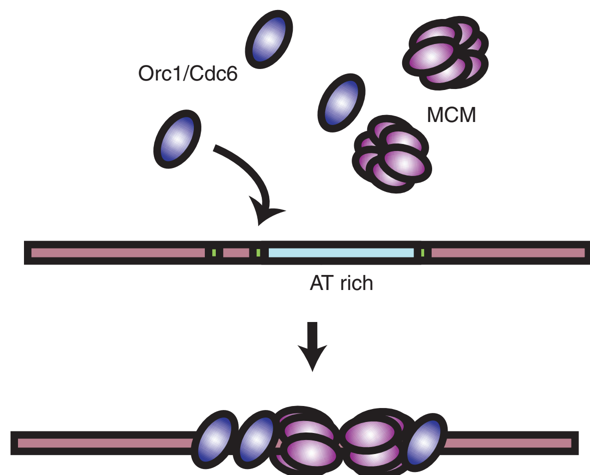

Fig. 3. Model for the recognition of an archaeal origin of replication (based on *S. solfataricus oriC1* [55]). Green boxes depict ORB elements that are recognized by the dark blue Orc1/Cdc6 protein (encoded by the *cdc6-1* gene). This event is presumed to lead to the recruitment of the MCM complex (purple), however, it is currently unknown whether additional factors are required for this process.

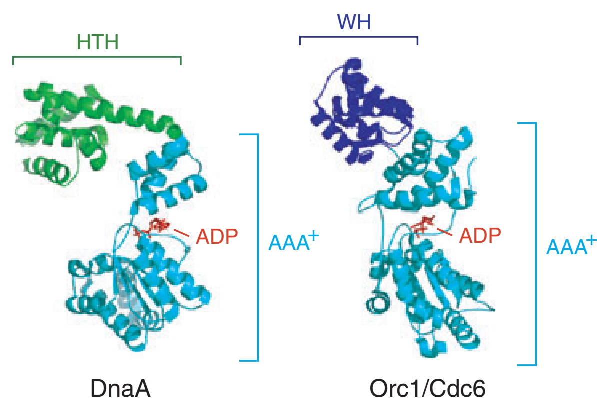

Fig. 4. Structures of bacterial DnaA and archaeal Orc1/Cdc6. The figure was generated using the PYMOL software package (http://pymol.sourceforge.net) and coordinates from PDB files 1FNN (Orc1/Cdc6) and 1L8Q (DnaA). The AAA$^+$ domains of both proteins are shown in cyan with ADP indicated in red. The helix-turn-helix (HTH)-containing DNA binding domain of DnaA is in green and the winged helix (WH)-containing domain of Orc1/Cdc6 is in dark blue.

While ORB/mini-ORB elements appear to be broadly conserved and perhaps play a role analogous to DnaA boxes, they are clearly not the only sequences bound by Orc1/Cdc6 homologues in archaea. *Sulfolobus oriC1* and *oriC2* are also recognized by the Cdc6-2 protein and *oriC2* is additionally bound by Cdc6-3 [57]. However, it has not yet been possible to establish consensus sequences for DNA recognition by these two proteins. It is possible that the Cdc6-2 protein may play a regulatory role in origin activity as it was found to be at highest levels in postreplicative cells, and preliminary data suggest that Cdc6-3 may act to facilitate mini-ORB recognition by Cdc6-1 (NP Robinson & SD Bell, unpublished data). Thus, the differential expression of these proteins may play a key role in regulating origin activity in *Sulfolobus* [57]. How conserved this potential mechanism is amongst the archaea is currently unclear, but it is enticing to note that many archaea encode more than one Orc1/Cdc6 homologue [50].

The *Sulfolobus oriC1* and *oriC2* were identified using a candidate locus approach in a 2D gel electrophoresis analysis to identify replication intermediates associated with replication initiation. However, bioinformatics had suggested that a third origin may exist in the *Sulfolobus* genome [56]. This proposal was confirmed by a whole genome microarray-based marker frequency analysis that, in addition to confirming the identity of the two previously characterized *Sulfolobus* origins, presented compelling evidence for a third origin, *oriC3* [63]. This origin has now been fine mapped and has been shown to bind all three Orc1/Cdc6 homologues (NP Robinson & SD Bell, unpublished data). The marker frequency analysis also revealed that all three origins appear to fire synchronously, however, how this is controlled remains unknown [63].

Thus, although much remains unknown about both the mechanisms, and particularly the control, of archaeal DNA replication initiation, these initial studies suggest that there is an intriguing level of complexity to the archaeal system. The combination of multiple replication origins in some species, together with multiple initiator proteins, some of which appear to be cell cycle regulated, suggests that comparatively sophisticated regulatory networks will be regulating origin activity in these organisms.

Finally, in bacteria it has been demonstrated that nucleoid proteins play key roles in assembly of the appropriate geometry of the DnaA–*oriC* complex. Additionally, as discussed above, the local chromatin architecture may play important roles in modulating, and even possibly facilitating, recruitment of the eukaryotic ORC. In this light it is likely that archaeal chromatin proteins may play roles in assisting pre-RC assembly on origins, Furthermore, the discovery that in *Sulfolobus* the chromatin protein Alba is regulated by reversible acetylation [64] presents the exciting possibility of epigenetic control of origin activity in the archaea.

Acknowledgements

Work in SDB's laboratory is funded by the Medical Research Council. We thank members of the Bell lab and Jessica Downs for helpful discussions.

References

1 Jacob F, Brenner S & Kuzin F (1963) On the regulation of DNA replication in bacteria. Cold Spring Harbor Symposium. *Quant Biol* **28**, 329–348.

2 Kornberg A & Baker TA (1992) *DNA Replication*, 2nd edn. Freeman, New York.

3 Fuller RS, Funnell BE & Kornberg A (1984) The dnaA Protein Complex with the *E. coli* Chromosomal Replication Origin (Oric) and Other DNA Sites. *Cell* **38**, 889–900.

4 Messer W (2002) The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Micro Rev* **26**, 355–374.

5 Schaper S & Messer W (1995) Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J Biol Chem* **270**, 17622–17626.

6 Speck C & Messer W (2001) Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *EMBO J* **20**, 1469–1476.

7 Davey MJ, Jeruzalmi D, Kuriyan J & O'Donnell M (2002) Motors and switches: AAA+ machines within the replisome. *Nat Rev Mol Cell Biol* **3**, 826–835.

8 Speck C, Weigel C & Messer W (1999) ATP- and ADP-DnaA protein, a molecular switch in gene regulation. *EMBO J* **18**, 6169–6176.

9 Hwang DS & Kornberg A (1992) Opening of the replication origin of *Escherichia coli* by DnaA protein with protein Hu or Ihf. *J Biol Chem* **267**, 23083–23086.

10 Bramhill D & Kornberg A (1988) A model for initiation at origins of DNA replication. *Cell* **54**, 915–918.

11 Funnell BE, Baker TA & Kornberg A (1987) *In vitro* assembly of a prepriming complex at the origin of the *Escherichia coli* chromosome. *J Biol Chem* **262**, 10327–10334.

12 Baker TA, Funnell BE & Kornberg A (1987) Helicase action of DnaB protein during replication from the *Escherichia coli* chromosomal origin *in vitro*. *J Biol Chem* **262**, 6877–6885.

13 Fang LH, Davey MJ & O'Donnell M (1999) Replisome assembly at oriC, the replication origin of *E. coli*,

reveals an explanation for initiation sites outside an origin. *Mol Cell* **4**, 541–553.
14 Davey MJ, Fang LH, McInerney P, Georgescu RE & O'Donnell M (2002) The DnaC helicase loader is a dual ATP/ADP switch protein. *EMBO J* **21**, 3148–3159.
15 Schwob E (2004) Flexibility and governance in eukaryotic DNA replication. *Curr Opin Microbiol* **7**, 680–690.
16 Antequera F (2004) Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. *EMBO J* **23**, 4365–4370.
17 Gilbert DM (2004) In search of the holy replicator. *Nat Rev Mol Cell Biol* **5**, 848–855.
18 Aladjem MI & Fanning E (2004) The replicon revisited: an old model learns new tricks in metazoan chromosomes. *EMBO Report* **5**, 686–691.
19 Mechali M (2001) DNA replication origins: from sequence specificity to epigenetics. *Nat Rev Genet* **2**, 640–645.
20 Bell SP & Dutta A (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem* **71**, 333–374.
21 Bell SP (2002) The origin recognition complex: from simple origins to complex functions. *Genes Dev* **16**, 659–672.
22 Bell SP & Stillman B (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**, 128–134.
23 Stillman B (2005) Origin recognition and the chromosome cycle. *FEBS Lett* **579**, 877–884.
24 Diffley JF (2004) Regulation of early events in chromosome replication. *Curr Biol* **14**, R778–R786.
25 Mendez J & Stillman B (2003) Perpetuating the double helix: molecular machines at eukaryotic DNA replication origins. *Bioessays* **25**, 1158–1167.
26 Bowers JL, Randell JCW, Chen SY & Bell SP (2004) ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. *Mol Cell* **16**, 967–978.
27 Geraghty DS, Ding M, Heintz NH & Pederson DS (2000) Premature structural changes at replication origins in a yeast minichromosome maintenance (MCM) mutant. *J Biol Chem* **275**, 18011–18021.
28 Laskey RA & Madine MA (2003) A rotary pumping model for helicase function of MCM proteins at a distance from replication forks. *EMBO Report* **4**, 26–30.
29 Lipford JR & Bell SP (2001) Nucleosomes Positioned by ORC Facilitate the Initiation of DNA Replication. *Mol Cell* **7**, 21–30.
30 Vogelauer M, Rubbi L, Lucas I, Brewer BJ & Grunstein M (2002) Histone Acetylation Regulates the Time of Replication Origin Firing. *Mol Cell* **10**, 1223–1233.
31 Theis JF, Yang C, Schaefer CB & Newlon CS (1999) DNA sequence and functional analysis of homologous ARS elements of *Saccharomyces cerevisiae* and *S. carlsbergensis*. *Genetics* **152**, 943–952.
32 Theis JF & Newlon CS (2001) Two compound replication origins in *Saccharomyces cerevisiae* contain redundant origin recognition complex binding sites. *Mol Cell Biol* **21**, 2790–2801.
33 Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP & Aparicio OM (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: High-resolution mapping of replication origins. *Science* **294**, 2357–2360.
34 Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ & Fangman WL (2001) Replication dynamics of the yeast genome. *Science* **294**, 115–121.
35 Chuang RY & Kelly TJ (1999) The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc Natl Acad Sci USA* **96**, 2656–2661.
36 Lee JK, Moon KY, Jiang Y & Hurwitz J (2001) The *Schizosaccharomyces pombe* origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. *Proc Natl Acad Sci USA* **98**, 13589–13594.
37 Kong D & DePamphilis ML (2001) Site-specific DNA binding of the *Schizosaccharomyces pombe* origin recognition complex is determined by the Orc4 subunit. *Mol Cell Biol* **21**, 8095–8103.
38 Segurado M, de Luis A & Antequera F (2003) Genome-wide distribution of DNA replication origins at A+T-rich islands in *Schizosaccharomyces pombe*. *EMBO Rep* **4**, 1048–1053.
39 Dai JL, Chuang RY & Kelly TJ (2005) DNA replication origins in the *Schizosaccharomyces pombe* genome. *Proc Natl Acad Sci USA* **102**, 337–342.
40 Coverley D & Laskey RA (1994) Regulation of eukaryotic DNA replication. *Annu Rev Biochem* **63**, 745–776.
41 Austin RJ, Orr-Weaver TL & Bell SP (1999) *Drosophila* ORC specifically binds to ACE3, an origin of DNA replication control element. *Genes Dev* **13**, 2639–2649.
42 Beall EL, Manak JR, Zhou S, Bell M, Lipsick JS & Botchan MR (2002) Role for a *Drosophila* Myb-containing protein complex in site-specific DNA replication. *Nature* **420**, 833–837.
43 Aggarwal BD & Calvi BR (2004) Chromatin regulates origin activity in *Drosophila* follicle cells. *Nature* **430**, 372–376.
44 MacAlpine DM, Rodriguez HK & Bell SP (2004) Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev* **18**, 3094–3105.
45 Danis E, Brodolin K, Menut S, Maiorano D, Girard-Reydet C & Mechali M (2004) Specification of a DNA replication origin by a transcription complex. *Nat Cell Biol* **6**, 721–730.
46 Stehle IM, Scinteie MF, Baiker A, Jenke ACW & Lipps HJ (2003) Exploiting a minimal system to study the

epigenetic control of DNA replication: the interplay between transcription and replication. *Chromosome Res* **11**, 413–421.
47 Schaarschmidt D, Baltin J, Stehle IM, Lipps HJ & Knippers R (2004) An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *EMBO J* **23**, 191–201.
48 Remus D, Beall EL & Botchan MR (2004) DNA topology, not DNA sequence, is a critical determinant for *Drosophila* ORC-DNA binding. *EMBO J* **23**, 897–907.
49 Bell SD & Jackson SP (2001) Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol* **4**, 208–213.
50 Kelman LM & Kelman Z (2003) Archaea: an archetype for replication initiation studies? *Mol Microbiol* **48**, 605–616.
51 Lopez P, Philippe H, Myllykallio H & Forterre P (1999) Identification of putative chromosomal origins of replication in Archaea. *Mol Micro* **32**, 883–886.
52 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**, 2212–2215.
53 Matsunaga F, Norais C, Forterre P & Myllykallio H (2003) Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep* **4**, 154–158.
54 Matsunaga F, Forterre P, Ishino Y & Myllykallio H (2001) In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc Natl Acad Sci USA* **98**, 11152–11157.
55 Berquist BR & DasSarma S (2003) An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, *Halobacterium* sp. strain NRC-1. *J Bacteriol* **185**, 5959–5966.
56 Zhang R & Zhang C (2003) Multiple replication origins of the archaeon *Halobacterium* species NRC-1. *Biochem Biophys Res Commun* **302**, 728–734.
57 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**, 25–38.
58 Capaldi SA & Berger JM (2004) Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon *Methanothermobacter thermoautotrophicus*. *Nucleic Acids Res* **32**, 4821–4832.
59 Erzberger JP, Pirruccello MM & Berger JM (2002) The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *EMBO J* **21**, 4763–4773.
60 Liu JY, Smith CL, DeRyckere D, DeAngelis K, Martin GS & Berger JM (2000) Structure and function of Cdc6/Cdc18: Implications for origin recognition and checkpoint control. *Mol Cell* **6**, 637–648.
61 Singleton MR, Morales R, Grainge I, Cook N, Isupov MN & Wigley DB (2004) Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. *J Mol Biol* **343**, 547–557.
62 Cunningham EL & Berger JM (2005) Unraveling the early steps of prokaryotic replication. *Curr Opin Struct Biol* **15**, 68–76.
63 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 USA* **101**, 7046–7051.
64 Bell SD, Botting CH, Wardleworth BN, Jackson SP & White MF (2002) The interaction of Alba, a conserved archaeal, chromatin protein, with Sir2 and its regulation by acetylation. *Science* **296**, 148–151.