Identification of Two Origins of Replication in the Single Chromosome of the Archaeon Sulfolobus solfataricus

Nicholas P. Robinson,1 Isabelle Dionne,1,3 Magnus Lundgren,2,3 Victoria L. Marsh,1 Rolf Bernander,2 and Stephen D. Bell1,∗
1Medical Research Council Cancer Cell Unit Hutcheson MRC Research Centre Hills Road Cambridge CB2 2XZ United Kingdom
2Department of Molecular Evolution Evolutionary Biology Center Uppsala University Norbyvägen 18C, SE-752 36 Uppsala Sweden

Summary

Eukaryotic chromosomes possess multiple origins of replication, whereas bacterial chromosomes are replicated from a single origin. The archaeon Pyrococcus abyssi also appears to have a single origin, suggesting a common rule for prokaryotes. However, in the current work, we describe the identification of two active origins of replication in the single chromosome of the hyperthermophilic archaean Sulfolobus solfataricus. Further, we identify conserved sequence motifs within the origins that are recognized by a family of three Sulfolobus proteins that are homologous to the eukaryotic initiator proteins Orc1 and Cdc6. We demonstrate that the two origins are recognized by distinct subsets of these Orc1/Cdc6 homologs. These data, in conjunction with an analysis of the levels of the three Orc1/Cdc6 proteins in different growth phases and cell cycle stages, lead us to propose a model for the roles for these proteins in modulating origin activity.

Introduction

The Archaea constitute the third domain of life and, although prokaryotic, are as phylogenetically distinct from bacteria as they are from eukaryotes. Within the archaeal domain, deeply branching kingdoms have been proposed with organisms identified from the Nanoarchaeota, Crenarchaeota, and Euryarchaeota. The Crenarchaea are particularly abundant in the biosphere and are found in a wide variety of niches including the oceans. Indeed, a recent study has indicated that crenarchaea represent 39% of the picoplankton in the Pacific Ocean (Kamer et al., 2001).

In recent years, it has become apparent that the DNA replication apparatus of archaea and eukaryotes is derived from a common ancestor, such that the archaeal machinery represents a core version of that in eukaryotes (Bell and Dutta, 2002; Bernander, 2000; Kelman and Kelman, 2003). Therefore, archaea have the potential to be powerful model systems to understand the central mechanisms underpinning eukaryotic DNA replication.

While the elongation phase of archaeal replication has been the subject of substantial analysis, considerably less is known about the initiation stage. Indeed, only a single origin of replication has been conclusively identified in archaea to date, that of the euryarchaeon Pyrococcus abyssi (Matsunaga et al., 2001, 2003; Myllykallio et al., 2000). This origin is contained within a 800 nt region upstream of the single orcl/cdc6 gene. The sequence of this intergenic region is highly conserved in Pyrococcus species. Furthermore, the locus is linked to a variety of replication and recombination-associated genes (e.g., rfc, radA, and a helicase). Although a number of sequence repeats have been identified within this region, the relevance of these in protein binding and origin function remains to be determined (Matsunaga et al., 2003; Myllykallio et al., 2000).

Within the repertoire of archaeal homologs of eukaryotic replication proteins, the most likely candidates for initiator proteins that mediate origin recognition are the archaeal Orc1/Cdc6 homologs. The archaeal proteins are related to both the eukaryotic Orc1 subunit of the origin recognition complex (ORC) and Cdc6 (called Cdc18 in Schizosaccharomyces pombe). As discussed below, ORC binds to origins of replication; once bound, it then recruits Cdc6, which, in turn and in conjunction with additional factors, leads to the recruitment of the presumptive replicative helicase, the Mcm complex (Bell and Dutta, 2002). As the archaeal proteins are related to both Orc1 and Cdc6, they may combine origin recognition and Mcm loading activities. Three Orc1/Cdc6 homologs are encoded by Sulfolobus solfataricus, and, although they display similarity to both Orc1 and Cdc6, in accordance with the genome annotation we shall refer hereafter to these as Cdc6-1, Cdc6-2, and Cdc6-3 (She et al., 2001). Chromatin immunoprecipitation experiments have indicated that the single Pyrococcus Orc1/Cdc6 protein is associated with the origin region in that species (Matsunaga et al., 2001); however, the site(s) bound by the protein are not yet known. To date, all the sequenced archaeal genomes encode Orc1/Cdc6 homologs, with the exception of Methanococcus januschii. Intriguingly, the number of these homologs varies between species: members of the Pyrococcus genus and Pyrobaculum aerophilum have a single Orc1/Cdc6, Methanobacterium thermoautotrophicum, Aeropyrum pernix, and Archaeoglobus fulgidus have two homologs, Sulfolobus, Thermoplasma and Methanosarcina species have three Orc1/Cdc6s, and halophilic archaea have up to 10 homologs. However, the functions of these proteins remain unknown. It is possible that different Orc1/Cdc6s define distinct replication origins within a given species. Alternatively, it is conceivable that, although related to both eukaryotic Orc1 and Cdc6, individual archaeal Orc1/Cdc6s may play either Orc1- or Cdc6-like roles. Finally, it is possible that some of the Orc1/Cdc6 homologs act as regulators.

Insight into the mechanism of Orc1/Cdc6 function has come from the determination of the crystal structure of

∗Correspondence: sb419@cam.ac.uk
1These authors contributed equally to this work.
Orc1/Cdc6 from the crenarchaeon *P. aerophilum* (Liu et al., 2000). The structure reveals the presence of a winged helix turn helix (wHTH) in Orc1/Cdc6 that has been proposed to interact with DNA. DNA was observed to inhibit an autophosphorylation activity of archaeal Orc1/Cdc6s in a manner dependent upon the presence of the wHTH (Grabowski and Kelman, 2001). More recent work has demonstrated that one of the two Orc1/Cdc6s from *Archeoglobus fulgidus* possesses nonsequence-specific DNA binding activity (Grainge et al., 2003). To our knowledge, to date, there has been no report of sequence-specific binding by the archaeal Orc1/Cdc6 proteins.

In eukaryotes, the issue of sequence-specific origin recognition varies, dependent upon the species in question (reviewed in Bell, 2002). Studies of origins of replication in Saccharomyces cerevisiae have demonstrated the presence of conserved sequence elements. At least two elements, the ACS and B1 motifs, are bound by ORC in this species (Bell and Stillman, 1992). Similarly, in *Schizosaccharomyces pombe*, specific sequence elements within origins, although unrelated to those in *S. cerevisiae*, are bound by the *s. pombe* ORC complex (Chuang and Kelly, 1999; Kong and DePamphilis, 2001; Lee et al., 2001). Intriguingly, while DNA recognition by *S. cerevisiae* ORC is ATP dependent, recognition by *S. pombe* ORC is not. This may be in part due to the fact that *s. pombe* ORC recognizes DNA via multiple AT hooks in Orc4, a feature not found in other ORC complexes. The situation in metazoans is more complex and has been harder to define. Sequence-specific binding of *Drosophila melanogaster* ORC has been observed at the origins in the chorion amplification loci (Austin et al., 1999). DNA binding by *Drosophila* ORC is strongly stimulated by ATP binding (Chesnokov et al., 2001). Although ORC is absolutely required for initiation of replication in *Xenopus laevis* cell-free systems (Carpenter et al., 1996; Romanowski et al., 1996; Rowles et al., 1996), it has been demonstrated that there is little or no sequence specificity for initiation in these systems (Coverley and Laskey, 1994).

In bacteria, the initiator protein is DnaA, and this molecule is conserved across the bacterial domain of life (Messer et al., 2001). The architectural organization of DnaA is related to that of Orc1/Cdc6. Like Orc1/Cdc6, DnaA has an N-terminal AAA-ATPase fold and a C-terminal DNA binding fold. However, in the case of DnaA, this is a helix-turn-helix (Erzberger et al., 2002). There is no significant sequence homology between the bacterial and archaeal/eukaryotic proteins, apart from residues involved in nucleotide binding. DnaA binds a highly conserved sequence, the dnaA box, multiple copies of which are often associated with bacterial origins. Therefore, origins in bacteria are often bound by multiple DnaA molecules and, with the cooperation of bacterial nucleoid proteins such as HU and IHF, are thought to impose a higher order structure at the origin.

It is well established that eukaryotic chromosomes have multiple origins, spaced between 10 and 330 kb apart. In contrast, bacteria have single chromosomal origins. Although a bioinformatics analysis suggested that there may be two origins of replication in the main Halobacterium chromosome (Zhang and Zhang, 2003), a recent targeted genetic screen for Halobacterium autonomously replicating sequences only found a single origin in this species (Berquist and DasSarma, 2003). These data, in conjunction with the identification of a single origin in *P. abyssi*, suggest that, like bacteria, there may only be one origin per archaeal chromosome (Myllykallio et al., 2000). However, in the current work, we identify two origins of replication in the hyperthermophile *Sulfolobus solfataricus*, a crenarchaeon that possesses three Orc1/Cdc6 homologs (She et al., 2001). Remarkably, origin identity appears to be mediated by the binding of distinct subsets of the Orc1/Cdc6 proteins. Analysis of the levels of these proteins as a function of the cell cycle reveals that the three proteins show distinct abundance profiles and suggests that they may have different roles in controlling initiation of replication.

### Results

#### Identification of Two Origins of Replication in *Sulfolobus solfataricus* P2

We have initiated studies of the cell cycle and of the DNA replication apparatus of the aerobic hyperthermophilic crenarchaeon *S. solfataricus* P2 (Dionne et al., 2003). A key goal has been to identify an origin of replication in *Sulfolobus* and to determine how this is defined at the molecular level. In light of the *P. abyssi* findings described above (Myllykallio et al., 2000), we examined the genomic organization of the three orc1/cdc6 genes in *S. solfataricus* to determine if any of them was linked to a noncoding region that might contain an origin of replication. Of the three, only the cdc6-1 locus contains any significant noncoding sequences immediately upstream, extending ~450 nt from the gene. Intriguingly, as in *P. abyssi*, the cdc6-1 gene is linked to a radA homolog (Figure 1A). The cdc6-1 gene is also linked to that of the basal transcription factor, TFE (Bell et al., 2001). Additionally, Cdc6-1 is the most closely related of the three *S. solfataricus* Orc1/Cdc6 proteins to the single Orc1/Cdc6 of *P. abyssi*. The predicted protein sequences of *S. solfataricus* Cdc6-1, Cdc6-2, and Cdc6-3 possess 50%, 30%, and 37% sequence identity, respectively, to *P. abyssi* Orc1/Cdc6. The cdc6-2 gene is closely linked to genes encoding for RFC, MCM, and a RadA paralog. However, the upstream region contains two conserved short ORFs of unknown function (Figure 1B). While the cdc6-3 gene is not linked to any genes of known replication function, short noncoding regions are found a short distance from the gene (Figure 1C).

To test whether the loci associated with the cdc6 genes might contain origins of replication, we isolated DNA from asynchronously replicating cells, subjected the DNA to digestion with restriction enzymes, and used this in 2D neutral-neutral agarose gel analysis (Brewer and Fangman, 1987) to detect replication intermediates (Figures 1D–1P). This technique allows the resolution of distinct arcs on a gel corresponding to replication intermediates (Figure 1Q). This approach permits discrimination between Y-shaped molecules, arising from passage of a replication fork through a given region, and molecules containing a bubble intermediate, indicative of the presence of an origin of replication in that fragment. Typically, for a bubble to be detected, it must be contained within the central third of the restriction

---

**References**

Replication Origins in *Sulfolobus solfataricus*

Figure 1. Detection of Origins of Replication at the *cdc6-1* and *cdc6-3* Loci

(A–C) Diagram of the genomic organization of the loci encoding *cdc6-1*, 2, and 3. Boxes indicate the location of open reading frames. ORFs above the midline are transcribed left to right and those below the line transcribed right to left. Genes encoding DNA replication/repair-associated proteins are indicated, as is that of the basal transcription factor TFE. The identity of DNA species generated by restriction enzyme digest and detected by Southern blotting of neutral/neutral 2D gels in Figure 1D–1P are indicated.

(D–P) 2D gel analysis of DNA isolated from asynchronous replicating *S. solfataricus* cultures. DNA was digested with the enzymes indicated in Figures 1A–1C and, following electrophoresis, transferred to nylon membrane then hybridized with fragment-specific radiolabeled probes followed by detection by autoradiography. Panels E and N contain bubble arcs indicative of active origins of replication encompassed within the central third of these restriction fragments.

(Q) Cartoon of 2D gel result with positions of bubble and fork arcs illustrated. Bubble arcs are generated by molecules containing replication origins, and fork arcs contain “Y”-shaped molecules indicative of passage of replication forks through the restriction fragment.

Fragment being analyzed. Using an AccI restriction digest, it was possible to detect arcs corresponding to both bubble- and fork-shaped replication intermediates in a fragment centered on the region upstream of *cdc6-1*, indicating the presence of an origin of replication (Figure 1E). We shall refer to this as oriC1. Importantly, re-probing the filter with probes specific for the flanking AccI fragments only resulted in detection of fork arcs (Figures 1D and 1F).

Similar analyses were performed with the *cdc6-2* locus. However, while fork arcs were readily detected across this locus (Figures 1G–1L), we could not detect...
any bubble arcs, leading us to conclude that there are no functional origins in the immediate vicinity of the cdc6-2 gene. Strikingly, however, when we performed 2D gel analyses of the cdc6-3 locus, we detected the presence of an additional active origin of replication upstream of the cdc6-3 gene (Figure 1N). We shall refer to this as oriC2. Analysis of flanking regions reveals only the presence of fork arcs (Figures 1M, 1O, and 1P). An additional feature in the 2D gels is the presence of an intense spike rising to meet both bubble arcs (Figures 1E and 1N). This pattern is indicative of X-shaped molecules and may correspond to joint molecules, such as those observed in the eukaryote Physarum (Benard et al., 2001).

Thus, our studies reveal the presence of two active origins of replication in the S. solfataricus chromosome. It must be emphasized that we do not exclude the possibility of additional origins of replication in the S. solfataricus chromosome.

Replication Initiation Point Mapping

To characterize further the origins, we used the replication initiation point (RIP) mapping procedure developed by Gerbi and colleagues (Bielsky and Gerbi, 1998) to identify the replication initiation points in S. solfataricus. This approach exploits the short lengths of eukaryotic Okazaki fragments to map the transition point between leading and lagging strand synthesis using a primer extension methodology. Recent work has demonstrated that archaea also have short Okazaki fragments allowing this technique to be applied to map the initiation point of P. abyssi (Matsunaga et al., 2003). Our analysis at the S. solfataricus oriC1 indicates the presence of transition points between leading and lagging strand synthesis approximately 230 nt upstream of the cdc6-1 gene (Figures 2A and 2B). Because the exact size of the RNA primer synthesized by the S. solfataricus primase in vivo is not known, this technique does not allow the identification of the precise nucleotide at which replication initiates. However, by analogy with eukaryotes, the initiation position will be within 10–15 nt of the transition point that we have mapped.

We also performed RIP mapping at the cdc6-3 proximal origin, oriC2. Again, we could detect transition points with this technique, mapping at 630 nt and 678 nt upstream of the Cdc6-3 open reading frame for lower and upper strands, respectively (Figures 2C and 2D).

In addition to mapping the initiation point at these origins, the fact that we can detect transition points at both origins strongly indicates that both origins are being used in the majority of cells.

S. solfataricus Cdc6-1 Recognizes Conserved Elements in oriC1

Based on the sequence similarity between archaeal Orc1/Cdc6 proteins and the eukaryotic Orc1 subunit of the origin recognition complex, we hypothesized that the S. solfataricus Orc1/Cdc6 proteins may play roles in origin recognition. To test this, we purified recombinant Cdc6-1, -2, and -3 and used these in DNasel footprinting analyses. Five millimolar ATP was included in these binding reactions because the budding yeast ORC complex is known to require ATP for DNA binding (Bell, 2002), and, in addition, the structure of an archaeal Orc1/Cdc6 indicates the presence of an AAA+ ATP binding fold (Liu et al., 2000). Our initial experiments focused on the oriC1 region. We found three regions of the DNA protected on both strands by the Cdc6-1 protein (Figure 3A). We have termed these elements ORB (Origin Recognition Boxes) 1, 2, and 3, with ORB1 being closest to the cdc6-1 gene. Titration of Cdc6-1 in footprinting studies reveals that ORB1 is bound with the highest affinity, followed by ORB2 and ORB3 (Figure 3B). All three sites are bound by 33 nM Cdc6-1. ORB1, 2, and 3 are closely related to each other in sequence (Figure 3C) and contain a short dyad symmetric element at their core. We cannot detect closely related sequences to these full ORB elements elsewhere in the S. solfataricus genome. However, as discussed below, we can detect sequences related to the internal dyad symmetric element at the cdc6-3 proximal origin (oriC2). Significantly, we find that the initiation points identified by RIP mapping lie at the right hand edge of ORB3 (Figure 2B).

Specificity and Cofactor Requirements of Cdc6-1 Binding

Double-stranded oligonucleotides corresponding to ORB1 are recognized by Cdc6-1 (Figure 3D), with two protein-DNA complexes detected in electrophoretic mobility shift assays (EMSA). Mutation of the conserved core of the ORB1 element substantially reduces its binding by Cdc6-1 (Figure 3D, lanes 4–9). We then went on to introduce analogous mutations into ORB1, 2, or 3 in the context of the origin sequence. We find that mutation of an individual ORB element does not appear to influence recognition of the other ORB elements (Figure 3E), indicating independent binding.

The crystal structure of an archaeal Orc1/Cdc6 indicates that it contains a wHTH domain at the C terminus (Liu et al., 2000). We therefore introduced point mutations into the cdc6-1 gene, changing the codons for arginine 342 and aspartate 344 to encode alanine. R342A and D344A lie in the proposed DNA recognition helix of the wHTH domain. The mutant protein was purified and, when compared with wild-type Cdc6-1 in EMSA using oligonucleotides corresponding to ORB1, we find that the mutant protein has impaired DNA binding activity (Figure 3F). Significantly, mutations in the analogous positions of the homologous S. pombe cdc18 gene result in a late S phase arrest, indicative of inefficient origin firing (Liu et al., 2000).

The crystal structure of P. aerophilum Orc1/Cdc6 revealed the presence of ADP in the nucleotide binding site of the protein (Liu et al., 2000). To determine whether co-factors might modulate DNA binding by Cdc6-1, we performed EMSAs on ORB1 oligonucleotides in the presence or absence of ATP, ADP, and the nonhydrolyzable ATP analog, AMP-PNP. To our surprise, none of these molecules significantly stimulated DNA binding by Cdc6-1. Indeed, ATP caused a slight but highly reproducible reduction in protein-DNA complex yield (Figure 3G).

Conservation of ORB Elements in Diverse Archaea

Origins of replication in S. cerevisiae possess conserved sequence elements (Bell, 2002; Bell and Dutta, 2002);
Figure 2. Mapping of Replication Initiation Points at both Origins

(A) Identification of replication initiation points by RIP mapping performed on replication intermediates isolated from *S. solfataricus*. RIP mapping of the *cdc6-1* proximal origin is shown. For the lower strand start site, cycle primer extension was performed with oligonucleotide RIP2 (left panel) and for the upper strand (right panel), oligonucleotide RIP5 was used. Lanes contain di-deoxy sequencing reactions (GATC) and the product of cycle extensions on replication intermediates (R) or nonreplicating DNA (N). Arrowheads indicate the position of start sites specific to replicating DNA, and the apparent transition point between leading and lagging strand is marked (TP).

(B) Summary of RIP mapping results at oriC1. Sequence of the region upstream of the Cdc6-1 open reading frame is shown. The positions of large repeated motifs that we term ORB elements (see Figure 3) are shown in large open arrows. The medium-sized arrows at the right edge of ORB3 show the position of the transition points identified in the RIP mapping in Figure 2A. Small arrows indicate the approximate position of presumptive start sites of Okazaki fragments.

(C) RIP mapping at the *cdc6-3* proximal origin. For the lower strand start site, cycle primer extension was performed with oligonucleotide RIP8 (left panel) and for the upper strand (right panel), oligonucleotide RIP10 was used. Lanes contain di-deoxy sequencing reactions (GATC) and the product of cycle extensions on replication intermediates (R) or nonreplicating DNA (N). Arrowheads indicate the position of start sites specific to replicating DNA, and the apparent transition point between leading and lagging strand is marked (TP).

(D) Summary of RIP mapping results at oriC2. Sequence of the noncoding region 515 nucleotides upstream of the *cdc6-3* open reading frame is shown. The medium-sized arrows show the position of the transition points identified in the RIP mapping in Figure 2C. Small arrows indicate the approximate position of presumptive start sites of Okazaki fragments.

however, no consensus for higher eukaryotic origins has been detected. Bacterial replication origins typically contain *dnaA* boxes that are bound by DnaA (Messer et al., 2001). The *dnaA* boxes are recognizably conserved between species. We therefore speculated that the ORB elements might also show general conservation between archaeal species. Accordingly, we searched for candidate ORB elements in other complete archaeal genomes (Figure 4A). In many cases, we could find clusters of ORB sequences, and in several cases, ORB elements were found at the position of predicted origins (based on marker frequency analysis or oligonucleotide skew calculations). Our analysis reveals that two ORB-like repeats are found at the known origin of *P. abyssi*.

These ORB-like elements were previously identified as repeats of unknown function (Matsunaga et al., 2003). In addition, the mapped replication initiation point in *P. abyssi* is adjacent to one of these repeats, reflecting the proximity of ORB3 to the initiation point found in *S. solfataricus*. Thus, it appears that ORB elements may be a general feature of archaeal origins. Further, these findings demonstrate the surprising conservation of replication initiation recognition elements between these two unrelated organisms, which are found in distinct archaeal kingdoms.

DNaseI footprinting analysis reveals that *S. solfataricus Cdc6-1* (the ortholog of the single *P. abyssi cdc6* gene) can bind to the ORB sequences in the *Pyrococcus*
Figure 3. Cdc6-1 Binds to oriC1

(A) DNaseI footprint analysis of Cdc6-1 binding to oriC1. Reactions were performed in the presence or absence of 250 nM recombinant Cdc6-1 protein. Samples were electrophoresed alongside a Maxam-Gilbert A + G sequencing ladder. The positions of the three binding sites (ORB1, 2, and 3) are indicated and summarized in Figure 2B.

(B) Titration of Cdc6-1 on the origin. Lanes contain 0, 8, 16, 33, 63, 125, 250, 500 nM Cdc6-1.

(C) Sequence lineup of the ORB elements, the short inverted repeat in the elements is indicated by converging arrows. Mutations introduced into duplex oligonucleotides MUT1 and MUT2 (used in Figure 3D) are indicated.

(D) EMSA analysis of the binding of Cdc6-1 (0, 42, and 125 nM) to oligonucleotides corresponding to wild-type (ORB1) or mutant (MUT1 and MUT2) ORB1 elements. Positions of free DNA and Cdc6-1 DNA complex are indicated.

(E) DNaseI footprint analysis of 32, 63, and 125 nM Cdc6-1 on wild-type origin sequences or derivatives containing mutations in ORB1, 2, or 3.

(F) EMSA using ORB1 DNA as in Figure 3D with 0, 42, and 125 nM wild-type Cdc6-1 or a derivative with point mutations (R342A D344A).

(G) EMSA analysis of complex formation on ORB1 mediated by 0, 56, 167, 500 nM Cdc6-1 in the presence or absence of 5mM ATP, ADP, and AMP-PNP.

origin in vitro (Figures 4B and 4D). Recently, a candidate origin of replication upstream of the cdc6-7 gene has been described in the archaeon Halobacterium NRC1 (Berquist and DasSarma, 2003). We detected ORB-like elements in this region (Figure 4A) and again found that S. solfataricus Cdc6-1 could recognize these motifs in vitro (Figures 4C and 4E).

Together, these findings indicate that the ORB elements are functionally conserved between the Euryarchaeota and Crenarchaeota kingdoms.

Cdc6-2 also Binds to oriC1

Next, we examined whether Cdc6-2 or Cdc6-3 might also bind oriC1. Remarkably, Cdc6-2 binds to three sites, including a region that overlaps ORB2 (Figures 5A and 5D). In addition to the protected sites, Cdc6-2 also induces DNaseI-hypersensitive sites at ORB3. Cdc6-3 does not bind specifically to the origin (Figure 5B), although at concentrations of the protein greater than 100 nM and in the absence of competitor DNA, we have detected nonspecific binding by the protein (data not shown).

We were surprised by the observation that Cdc6-2 binds to ORB2 but not to the closely related ORB1 or 3. To test whether the binding of Cdc6-2 to ORB2 was dependent on the same sequences that are recognized by Cdc6-1, we introduced mutations into the dyad symmetric core element in ORB2. This prevented recognition by Cdc6-1 but not Cdc6-2, indicating that these proteins have distinct sequence preferences and explaining why ORB1 and 3 are not bound by Cdc6-2 (Figure 5C).

All Three Cdc6s Bind the Second Origin, oriC2

Next, we examined the binding of the three Cdc6s to oriC2. Although Cdc6-1 does bind to two regions within this origin (Figure 6A), it does so with a considerably reduced affinity compared to binding to oriC1. Whereas
Figure 4. Identification of ORB Elements across the Archaeal Domain

(A) Location and sequence of ORB elements in archaeal genomes. Species abbreviations and GenBank accession numbers for coordinates: AP, Aeropyrum pernix, NC_000854; SS, Sulfolobus solfataricus, NC_002754; ST, Sulfolobus tokodaii, NC_003106; AF, Archaeoglobus fulgidus, NC_000917; H, Halobacterium NRC-1, NC_002607; PA, Pyrococcus abyssi, NC_000868; PH, Pyrococcus horikoshii, NC_000961; PF, Pyrococcus furiosus, NC_003413. Spacing: distance to nearest upstream repeat. References: 1, (Zhang and Zhang, 2003); 2, (Maisnier-Patin et al., 2002); 3, (Berquist and DasSarma, 2003); 4, (Myllykallio et al., 2000).

(B) DNaseI footprinting of 0, 16, 31, 62, 125, 250, 500 nM Cdc6-1 on the Pyrococcus origin of replication. Black rectangles indicate the position of protected regions.

(C) Footprinting of Cdc6-1 on both strands of the candidate replication origin of Halobacterium. MG is a Maxam Gilbert A+T sequencing ladder. Black rectangles indicate the position of protected regions, and coordinates are given relative to the first nucleotide of the cdc6-7 open reading frame.

(D) Summary of footprinting on the Pyrococcus origin. The position of transition points mapped in the P. abyssi origin (Matsunaga et al., 2003) is indicated by arrows and labeled TP. The ORB elements bound by Cdc6-1 are shown as open arrows.

(E) Summary of footprinting on the Halobacterium origin. ORB elements are shown as open arrows.
Cdc6-1 had saturated the three binding sites at oriC1 at 33 nM, we do not detect binding to oriC2 until 500 nM. Inspection of the protected sequences, however, reveals a relationship with the ORB elements (Figure 6E). Specifically, both of the new binding sites show similarity to the inner core dyad of ORB elements. However, beyond this conserved core, they do not correspond to the extended ORB consensus defined above. We assume that the extra conserved sequences in ORB elements contribute to the greater binding affinity. It is interesting to note that the two “mini-ORB” (mORBa and mORBb) elements at oriC2 are inverted relative to each other and spaced 65 nt apart, a situation reminiscent of the 65 nt between the mini-ORBs at oriC1. At oriC2, a notable feature of the 65 nt between the mini-ORBs is that it contains a stretch of 32 nt with 79% AT composition. Such an organization is reminiscent of a duplex unwinding element, a feature found in many bacterial and eukaryotic origins of replication and thought to facilitate melting of DNA at the origin (Natale et al., 1992). In addition, one of the transition points determined by RIP mapping (Figures 2C and 2D) lies within this AT-rich element.

As with the oriC1, we detect binding of Cdc6-2 to this locus. Two footprints are detected that correspond to imperfect inverted repeats (Figure 6B). Interestingly, we cannot detect related sequences at the Cdc6-2 binding sites in oriC1, and we have thus far been unable to determine a consensus sequence for DNA recognition by this factor.

As described above, we found that Cdc6-3 binds nonspecifically at high concentrations (data not shown). However, in marked contrast to oriC1, at lower concentrations (below 62 nM) we could detect two specific footprints (Figure 6C), corresponding to imperfect direct repeats. Intriguingly, one of these is encompassed within one of the Cdc6-2 binding sites, and the second element overlaps a Cdc6-1 binding site (Figure 6D).

Levels and Binding of Cdc6 Proteins In Vivo

To address the significance of these in vitro data to the regulation of replication in vivo, we raised specific polyclonal antisera to the three Cdc6 proteins. First, we used these antibodies in chromatin immunoprecipitation (ChIP) experiments (Orlando and Paro, 1993). Exponentially growing cells were treated with formaldehyde and DNA was recovered by immunoprecipitation from crosslinked, sonicated extracts using the three specific antisera or their preimmune sera as controls. The DNA was subjected to PCR using three pairs of primers specific for oriC1, oriC2, and the lrs14 locus, a transcription regulator encoded over 700 kb from either origin (Bell and Jackson, 2000; Napoli et al., 1999). In good agreement with our in vitro footprinting data, we found that Cdc6-1 and Cdc6-2, but not Cdc6-3, could be crosslinked to oriC1 (Figure 7A, top panel). Additionally, all three anti-sera against the individual Cdc6s allowed the precipitation of oriC2 sequences (Figure 7A, middle panel). Importantly, the preimmune sera failed to precipitate significant amounts of DNA, and the lrs14 locus was not detected in any experiment (Figure 7A, bottom panel).

Next, we addressed the levels of the Cdc6s during cell growth. The antisera were used in Western blotting of whole-cell extracts of S. solfataricus prepared at various points throughout exponential growth and stationary phase (Figures 7B and 7C). Consistent with a role in actively replicating cells, we found that levels of all three Cdc6s are significantly reduced in nonreplicating
stationary phase cells (Figure 7C, lane 9). We next examined whether these effects were mediated at the transcriptional level. Northern blotting of RNA prepared from mid-logarithmic growth and early stationary phase reveals that while cdc6-2 mRNA is still detectable, cdc6-1 and cdc6-3 transcripts are absent in stationary phase cells (Figure 7D). Although regulation at the level of mRNA stability may play a role, it is tempting to speculate that there is control of expression of these genes at the level of transcriptional initiation. Indeed, the close linkage of the origins with the cdc6-1 and cdc6-3 genes may permit a dynamic mechanism for coupling origin activity with regulation of the levels of transcription of the initiators.

Next, we examined the levels of the proteins in a synchronised culture. As S. solfataricus cultures display a low efficiency of synchronization, this experiment was performed with the related S. acidocaldarius species. Synchronization was achieved by arresting cell growth with N'-guanyl-1,7-diaminoheptane (GC7), an inhibitor of translation. Treatment of cells with this compound induces G2 arrest in the cell population. Release from the arrest results in one synchronous round of cell division and subsequent chromosome replication, after

Figure 6. All Three Cdc6s Bind to oriC2
(A) DNaseI footprinting of 0, 250, 500, 1000 nM Cdc6-1 on oriC2. The positions of Cdc6-1-induced protection are indicated by open rectangles on the right of the panel and named mORBa and mORBb. The track marked MG contains a Maxam Gilbert A + G sequencing reaction. The coordinates limiting the extremities of the observed footprints are given relative to the first nucleotide of the cdc6-3 ORF.
(B) DNaseI footprinting of 0, 31, 62, 125 nM Cdc6-2 on oriC2. Positions of Cdc6-2-induced protection are indicated by gray rectangles at the right of the figure, coordinates are as in (A).
(C) DNaseI footprinting of 0, 16, 31, and 62 nM Cdc6-3 on oriC2. Cdc6-3 binding sites are indicated at the right of the figure as black rectangles.
(D) Summary of binding sites for the three Cdc6s at oriC2. The positions of transition points established in the RIP mapping in Figure 2 are indicated with fine arrows and TP. Open white arrows indicate the position and polarity of the mini-ORB elements (mORBa and mORBb) bound by Cdc6-1. Gray arrows indicate the positions of the regions (C2a and C2b) bound by Cdc6-2. Black filled arrows indicate the positions of the Cdc6-3 binding sites, C3a and C3b. The position of a candidate duplex unwinding element is shown as an open rectangle between the mini-ORB elements.
(E) Summary of sequences recognized by the Cdc6s. The mini-ORB elements are shown in a lineup with the ORB elements at oriC1.
Figure 7. Expression and In Vivo Binding of the Cdc6s

(A) Chromatin immunoprecipitation using either preimmune (P) or immune sera (I) raised against Cdc6-1, 2, or 3. Input samples were 1/10, 1/100, and 1/1000 of input samples generated as described in Experimental Procedures. DNA recovered from immunoprecipitates was amplified with primers specific for oriC1 (top panel), oriC2 (middle panel), or a distal control region lrs14 gene.

(B) Plot of log_{10} (optical density 600 nm) versus time. Aliquots of the S. solfataricus P2 culture were taken at the indicated times and whole-cell extract prepared.

(C) Extracts prepared from time points indicated in Figure 7B were analyzed by Western blotting using antisera specific for Cdc6-1, Cdc6-2, or Cdc6-3. To control for loading, the three blots were subsequently probed with anti-TBP antisera.

(D) Northern blot analysis of RNA isolated from logarithmic (OD_{600} = 0.35) or stationary phase cells (equivalent to time point 9 in Figure 7B). Blots were hybridized with probes specific for cdc6-1, 2, or 3. Blots were stripped and rehybridized with a probe specific for the single strand binding protein gene (Wadsworth and White, 2001).

(E) Flow cytometer profile of culture growth of S. acidocaldarius released from GC7 arrest. Bars indicate the positions of peaks corresponding to 1 and 2 genome equivalents. The peak to the left of the 1 genome region contains dead cells and debris. Samples were taken at the indicated time points; samples 1 and 2 are representative of G2 phase cells, samples 3 and 4 are primarily G1 and S phase cells, and sample 5 is asynchronous, whole-cell extract prepared and analyzed by Western blotting in (F).

(F) Western blot analysis of samples indicated in (E) using antisera specific for the three Cdc6s. Loading was controlled using anti-TBP antisera as above.
which synchrony is lost. Culture progression was monitored by flow cytometry (Figure 7E). We prepared whole-cell extracts from cultures harvested at selected time points and determined relative Cdc6 protein levels by Western blotting (antisera against the TATA box binding protein, TBP, was used as internal loading control). We found a dramatic difference in the expression profile of the cdc6 genes, with Cdc6-2 being present in G2 phase cells and Cdc6-1 and Cdc6-3 present in G1 and S phase cells (Figure 7F).

Discussion

In the current work, we have revealed the presence of two active origins of replication in the hyperthermophilic archaean S. solfataricus. This result has been supported by recent marker frequency analysis (M.L. and R.B., unpublished data). Additionally, we have demonstrated sequence-specific recognition of these origins by the S. solfataricus homologs of the eukaryotic Orc1 and Cdc6 proteins. Furthermore, we find that the highly conserved Cdc6-1 protein recognizes sequence motifs (ORB elements) associated with replication origins found in a phylogenetically diverse range of archaea. This suggests a broadly conserved mechanism for origin recognition in the archaean domain of life. However, we have been unable to detect clear ORB elements in the genomes of P. aerophilum and M. thermoautotrophicum. A recent phylogenetic analysis of the archaeal Orc1/Cdc6s has indicated that P. aerophilum lacks a Cdc6-1 ortholog (Berquist and DasSarma, 2003). Furthermore, the single Orc1/Cdc6 in this species is more closely related to S. solfataricus Cdc6-2 than to Cdc6-1, suggesting that P. aerophilum may possess a distinct mechanism of origin recognition. Within S. solfataricus, oriC2 does not contain full ORB elements, but rather has mini-ORB elements bound by Cdc6-1. These are related to the inner dyad element in full ORB elements. We speculate that at oriC2, interaction between the adjacent binding sites for Cdc6-3 and Cdc6-1 will facilitate the assembly of a cooperative complex at the origin. This situation may apply to other archaebacteria with multiple Orc1/Cdc6s.

We have revealed that initiation at the cdc6-1 proximal origin, oriC1, occurs in the vicinity of ORB3. It may be significant that ORB3 and ORB2 are inverted relative to each other. Furthermore, a similar situation occurs at oriC2. The mini-ORB elements are inverted relative to one another and flank one of the transition points detected in RIP mapping at oriC2. The inverted nature of these repeats may provide a mechanism to impart directionality to the resultant Cdc6-1/DNA complexes, and thereby direct recruitment of the replicative helicase (presumably the MCM complex) between the repeats. It is particularly noteworthy that the spacing between ORB2 and 3 at oriC1 is 70 nt and between the mini-ORBs at oriC2 is 85 nt. These spacings correspond to distances of approximately 240 and 220 Å, respectively, assuming B form DNA. Image reconstructions of electron micrographs of the MCM of the euarchaeon M. thermotogae indicate that a double hexamer, or double heptamer, of MCM is roughly 200 Å in length (Chong et al., 2000; Yu et al., 2002) and could therefore readily be accommodated between ORB2 and 3 and between mini-ORBs.

The temporal regulation of Cdc6 levels that we have observed leads us to speculate that Cdc6-1 and Cdc6-3, which are present in G1 and S phases, play roles in promoting replication and that Cdc6-2, which is present only in G2, may act as a negative regulator. In this regard, it may be significant that at both origins, Cdc6-2 binds to sites that overlap Cdc6-1 or Cdc6-3 binding sites. We propose, therefore, that Cdc6-2 acts as a repressor of replication, possibly by blocking access to the origins by residual Cdc6-1 and 3 during the postreplicative G2 phase of the cell cycle and perhaps during stationary phase. A recent phylogenetic analysis of the archaean Cdc6s has suggested that three principal groupings or clades of Cdc6s exist (Berquist and DasSarma, 2003); one contains orthologs of S. solfataricus Cdc6-1 (including the single Pyrococcus Cdc6 and Halobacterium Cdc6-7) and another clade contains Orc1/Cdc6s closely related to S. solfataricus Cdc6-2. Additionally, species with multiple Orc1/Cdc6s have at least one member in both Cdc6-1 and Cdc6-2 clades. In light of our data, we propose that many archaean species may have interplay between positively and negatively acting Orc1/Cdc6s as an integral component of their replication machineries. Intriguingly, a parallel may be drawn with bacteria. In E. coli, the most closely related protein to the E. coli DnaA replication initiation protein is Hda, a negative regulator of replication (Kato and Katayama, 2001).

Our characterization of two replication origins in S. solfataricus reveals a prokaryote with more than a single origin of replication and indicates that origin identity and regulation may be determined by selective recognition by distinct Orc1/Cdc6 homologs. Given that the majority of archaean genome sequences to date possess multiple Orc1/Cdc6s, we speculate that our findings will be generally applicable to the archaean domain of life. Furthermore, our ongoing studies of the function and coordinated regulation of origin firing in Sulfolobus will give insight into the evolution of modern eukaryotic chromosomess, with their multitude of coordinately regulated origins of replication.

Experimental Procedures

Neutral/Neutral 2D Agarose Gel Electrophoresis

Asynchronous S. solfataricus cultures were grown to early/mid log phase (optical density (A_{600}) of 0.3) and harvested by centrifugation at 7000 g for 10 min at 4°C. Cells were washed with chilled TEN solution (50 mM Tris, 50 mM EDTA, 100 mM NaCl, pH 8.0) and suspended in TEN to optical density (A_{600}) of 600. Genomic plugs were made by immobilizing the cells in an equal volume of 0.8% low melting point agarose (Biogene) and pouring into plug molds. The plugs were digested within the plugs overnight at 37°C with proteinase K (VWR International) and then transferred to fresh NDS solution (pH 8.0) supplemented with 1 mg.ml^{-1} proteinase K (WWR International Ltd) and then transferred to fresh NDS solution (pH 8.0 + 1 mg.ml^{-1} proteinase K) for a second 37°C overnight incubation. Following elution in 1 × restriction enzyme buffer (NEB), the unsheared genomic DNA was restriction digested within the plugs overnight at 37°C (NEB, 600 to 1000 Units). An entire digested agarose plug was run on the N/N 2D agarose gel in 1 × TBE (90 mM Tris-borate, 2 mM EDTA) at 4°C (first dimension 0.4% agarose [no ethidium bromide], 0.6 V.cm^{-1} for 48 hr; second dimension 1.0% agarose [0.3 μg.ml^{-1} ethidium bromide], 5.0 V.cm^{-1} for 7 hr). DNA was transferred to nylon membrane (Hybond-N or XL; Amersham Pharmacia) by Southern blot and bound by UV irradiation. Filters were probed using α-32P-radiolabeled DNA fragments (Rediprime II; Amersham Biosciences) specific to the locus under investigation (sequences supplied on
request) and washed to a final stringency of 0.5 × SSC (1 × SSC: 0.15 M NaCl, 0.015 M sodium citrate) in 0.1% sodium dodecyl sulfate at 60°C.

RIP Mapping
A 1 liter culture of S. solfataricus was grown in medium DSM88 to OD600 = 0.4. Replication intermediates were isolated as described (Matsumura et al., 2003). Points were detected using cycle primer extension with oligonucleotides RIP2 (CCGGTTATAGGGTTTATAGGGTTTATATGCTAGTGAATTAAC) using the following cycling procedure: with 30°C for 1 min, 94°C; 1 min, 64°C; 1 min, 75°C. Sequencing reactions were performed in parallel and all samples electrophoresed on a 6% acrylamide gel containing 8M urea and 1 × TBE.

Amplification and Cloning of Origins
OriC1 was amplified using primers ori7854not5/H11003 and ori333orf3/CCTACCTTCAAGAATTAAC) using the following cycling procedure: 2 min, 95°C; then 30 cycles of 1 min, 94°C; 1 min, 64°C; 1 min, 75°C. Sequencing reactions were performed in parallel and all samples electrophoresed on a 6% acrylamide gel containing 8M urea and 1 × TBE.

Site-Directed Mutagenesis of pOriC1
Site-directed mutagenesis on pOriC1 was performed using the QuikChange method (Stratagene) to introduce the point mutations in the ORB elements indicated below in bold and underlined.

ORB1:
Wild-type: TATTTACCTTAAGGTTTATATGCTAGTGAATTAAC
Mutant: TATTTACCTTAAGGTTTATATGCTAGTGAATTAAC

ORB2:
Wild-type: ATCATATATATAATTTTCCAAACGAAACAATGGGGTAGG
Mutant: ATCATATATATAATTTTCCAAACGAAACAATGGGGTAGG

ORB3:
Wild-type: GCTAGATTAAGTAATTTCCAGAGGAAATAGATGGGT
Mutant: GCTAGATTAAGTAATTTCCAGAGGAAATAGATGGGT

DNasel Footprinting
For DNasel footprinting assays, NolI-EcoRI restriction fragments were excised from the origin containing plasmids. For analysis of footprinting on the top strand, the Nol site was labeled using 50 μCi of 3000Ci/mmol 32P-ATP using T4 polynucleotide kinase. Formaldehyde was added to an exponentially growing culture of S. solfataricus and incubated at 37°C for 3 hr. A 25°C volume of 20 mM Tris acetate (pH 7.9 at 25°C), 10 mM magnesium acetate, 50 mM potassium acetate and 1 mM dithiothreitol, 5 mM ATP, 5% glycerol, 5 μg/ml poly dG.dC. Binding reactions were incubated at 48°C for 10 min before loading on an 8% polyacrylamide gel in 1 × TBE. Gels were run for 2 hr at 10 volts/cm before drying and autoradiography.

Maxam Gilbert A + G Sequencing Reaction
Ten milliters (~100 fmol) DNA probe was mixed with 4 μl of 2 μg/ml sonicated salmon sperm DNA and 3 μl 88% formic acid and incubated at 37°C for 7 min. One hundred and fifty milliters of 10% piperidine was added and the reaction incubated at 90°C for 30 min. 1.2 ml butanol was added, mixed vigorously, then centrifuged. The pellet was resuspended in 150 μl 1% SDS, 1.2 ml butanol added, mixed, and centrifuged. The pellet was dried, resuspended in 25 μl 10 mM Tris (pH 8.0), and passed through a Sephadex G50 microspin column (Amershams). Twenty-five milliters formamide loading buffer was added, the sample boiled, and 5 μl electrophoresed as above.

Synchronization of S. acidocaldarius
The synchronization method used was a modified version of the one described (Jansson et al., 2000). S. acidocaldarius was grown to early log phase (A600 = 0.1) at 78°C in Allen media supplemented with 0.2% tryptone (Grogan, 1989). The culture was divided in control and treatment culture. N'-guanylyl-7, 7-diaminoheptane (GC) was added to the treatment culture to a final concentration of 150 μM from a stock solution of 0.05 M GC; dissolved in 1 M acetic acid. The cultures were incubated for 4 hr and then pelleted by centrifugation at room temperature (3000 g 20 min) and resuspended in fresh preheated media. Eighty milliliter samples were taken from the treated culture 10, 70, 120, 150, and 220 min after resuspension in fresh media. The samples were cooled to room temperature in ice water and then pelleted by centrifugation (3000 g 15 min) at room temperature. The cells were then washed with 0.5 ml media with pH adjusted to 5.0 with NaOH and pelleted again. Samples were stored at ~ 80°C until protein extraction.

Flow cytometry using mitrhamycin A and ethidium bromide staining, as described (Bernander and Poplawski, 1997), were used to monitor initial culture, control, treated culture, and synchronized culture.

Chromatin Immunoprecipitation
Formaldehyde was added to an exponentially growing culture of S. solfataricus to a final concentration of 1% and cooled to room temperature for 20 min. Then the reaction was quenched by the addition of glycine to 125 mM. Cells were harvested and washed with phosphate-buffered saline, then resuspended and lysed in 4 ml TBST (20 mM Tris.Cl [pH 7.5], 150 mM NaCl, 0.1% Tween 20, 0.1% Triton X-100). The extract was sonicated, and the soluble fraction containing DNA primarily in the 500–1000 bp size range was prepared by centrifugation. Immunoprecipitations were performed by incubating 3 μl antiserum with 10 μg of extract in 100 μl TBST with shaking at 4°C for 3 hr. A 25 μl slurry of protein A sepharose was added, and incubation continued for a further hour. Immune complexes were collected by centrifugation and washed by five consecutive 5 min incubations with 1.2 ml TBST with vigorous shaking at room temperature. Beads were then washed once with...
TBSTT containing 500 mM NaCl and once with TBSTT containing Tween 20 and Triton X-100 at 0.5%. Finally, immune complexes were disrupted by resuspending the beads in 20 mM Tris (pH 7.8), 10 mM EDTA, and 0.5% SDS and heating to 65°C for 30 min. Beads were removed, and the DNA was recovered by treatment with 10 µg/ml proteinase K for 6 hr at 65°C then 10 hr at 37°C, followed by extraction with phenol/chloroform/isooamyl alcohol and back-extraction of the organic phase with TE, then chloroform extraction. Finally, DNA was precipitated with ethanol in the presence of 20 µg glycogen then resuspended in 50 µl TE. Input samples were treated as above without the addition of antiserum and beads.

Acknowledgments

We are indebted to Björn Canbäck and Björn Brindefalk for similarity searches for ORF elements. We thank Ian Holt and David Santamaria for suggestions on 2D gel protocols and Jessica Downs for comments on the manuscript. This work was funded by the Medical Research Council, the Swedish Research Council, and the Swedish Graduate Research School in Genomics and Bioinformatics.

Received: December 2, 2003
Revised: December 15, 2003
Accepted: December 15, 2003
Published online: December 24, 2003

References


