



A conserved mechanism for replication origin recognition and binding in archaea

Alan I Majerník, James Pj Chong

► To cite this version:

Alan I Majerník, James Pj Chong. A conserved mechanism for replication origin recognition and binding in archaea. *Biochemical Journal*, 2007, 409 (2), pp.511-518. <10.1042/BJ20070213>. <hal-00478756>

HAL Id: hal-00478756

<https://hal.science/hal-00478756v1>

Submitted on 30 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A CONSERVED MECHANISM FOR REPLICATION ORIGIN RECOGNITION AND BINDING IN ARCHAEA

Alan I. Majerník* and James P.J. Chong†

*Institute of Animal Biochemistry and Genetics, Slovak Academy of Sciences, 900 28 Ivanka pri Dunaji, Slovak Republic

†Department of Biology, University of York, PO Box 373, York, YO10 5YW, UK

Corresponding author: James P.J. Chong
Department of Biology (Area 5)
University of York
PO Box 373, York
YO10 5YW, UK
Tel. +44 1904 328628
Fax. +44 1904 328505
Email jpjc1@york.ac.uk

Synopsis

To date, methanogens are the only group within the Archaea where firing DNA replication origins have not been demonstrated *in vivo*. Here we show that a previously identified cluster of Origin Recognition Box (ORB) sequences do indeed function as an origin of replication *in vivo* in the archaeon *Methanothermobacter thermautotrophicus*. Although the consensus sequence of ORBs in *M. thermautotrophicus* is somewhat conserved when compared to ORB sequences in other archaea, the Cdc6-1 protein from *M. thermautotrophicus* (*MthCdc6-1*) displays sequence-specific binding that is selective for the *MthORB* sequence and does not recognize ORBs from other archaeal species. Stabilization of *in vitro* *MthORB* DNA binding by *MthCdc6-1* requires additional conserved sequences 3' to those originally described for *M. thermautotrophicus*. By testing synthetic sequences bearing mutations in the *MthORB* consensus sequence, we show that Cdc6-ORB binding is critically dependent on the presence of an invariant guanine found in all archaeal ORB sequences. Mutation of a universally conserved arginine in the recognition helix of the winged helix domain of archaeal Cdc6-1 shows that specific origin sequence recognition is dependant on the interaction of this arginine with the invariant guanine. Recognition of a mutated origin sequence can be achieved by mutation of the conserved arginine to a lysine or glutamine. Thus, despite a number of differences in protein and DNA sequences between species, the mechanism of origin recognition and binding appears to be conserved throughout the archaea.

Short title

Replication Origin Recognition and Binding in Archaea

Keywords

DNA replication, origin of replication, archaea, methanogen, Cdc6

Introduction

DNA replication in archaea employs enzymes that are more similar to those utilised by eukaryotes than bacteria in the equivalent process. In general, the archaeal proteins are simplified versions of their eukaryotic counterparts [1, 2]. For example, in eukaryotes a six subunit Origin Recognition Complex (ORC) binds to origins of replication and acts as a “landing pad” for the assembly of pre-replication and pre-initiation proteins such as Cdc6 and MCM. The closest homologues of ORC proteins in archaea are the Orc/Cdc6-like proteins, which share homology with eukaryotic Cdc6 and the C-terminal portion of ORC1. Most archaea possess two Orc/Cdc6 genes, although there are a number of exceptions: *Methanopyrus kandleri*, *Methanocaldococcus jannashii* and *Methanococcus maripaludis* have only distantly related Cdc6 homologues [3, 4, 5,], species of the *Sulfolobus* genus possess three [6], whereas *Halobacterium* sp NRC-1 contains 10 Cdc6-like genes [7].

Structures of Cdc6 proteins from *Pyrobaculum aerophilum* [8] and *Aeropyrum pernix* [9] show that archaeal Cdc6 proteins consist of 3 domains. The C terminal domain forms a winged helix domain (WHD) and binds DNA in a sequence-specific manner via a helix-turn-helix motif (HTH) [6, 10, 11]. Comparison of archaeal HTH domain sequences reveals that “Class I”, but not “Class II” Orc/Cdc6 proteins, possess a conserved motif on the second helix consistent with a DNA recognition helix [9]. Classical WHD proteins recognize DNA sequences by inserting the recognition helix into the major groove of DNA, while the β -hairpin “wing” interacts with the phosphodiester backbone [12]. Compatible with this, mutation of Arg³³⁴ and Arg³³⁵ residues in the putative recognition helix of *MthCdc6-1* strongly reduced DNA binding whereas mutations in positively charged residues of the *MthCdc6-1* “wing” (Arg³⁵⁸, Lys³⁶⁰, Arg³⁶²) affected DNA binding only modestly [10]. Recent data has shown that multiple copies of the *A. pernix* Cdc6-1/ORC1 dimer recognize and bind four *A. pernix* ORB elements, causing a specific change in the topology of the DNA and partial unwinding of an AT-rich region, as judged by sensitivity to DNase I and P₁ endonuclease [13]. Class II (Orc2/Cdc6-2) proteins display only partial origin sequence specific DNA binding and do not possess conserved sequences within the recognition helix, but rather show conservation in the wing motif of the protein which is implicated in a different mechanism of DNA binding [9].

GC skew and Z-curve analyses have been used to predict origins of replication for a number of archaeal species, including *M. thermautotrophicus* and *Methanosarcina mazei* [14,

15], although these techniques have not been useful in other archaea such as *M. jannashii* or *Sulfolobus solfataricus* [15, 16]. Whether or not predictions exist, experimental data are still required to confirm that putative origins act as sites of replication initiation *in vivo*. Functional archaeal replication origins have been experimentally verified by a number of methods [6, 7, 17-20]. Unusually for prokaryotes with circular chromosomes, a number of archaea including both of the *Sulfolobus* species examined have been shown to possess multiple origins of replication [6, 18, 20, 21]. To date, methanogens are the only group within the Archaea where replication origins have not been experimentally demonstrated.

The presence of Origin Recognition Box (ORB) repeats is specific to archaeal replication origins. 36 bp ORBs were first described in *S. solfataricus* near an AT-rich region that might act as a site for initial duplex melting. Several ORB sequences are typically located upstream of Class I Orc/Cdc6 genes and ORBs were identified in most sequenced archaea apart from methanogens [6]. *In vitro* Cdc6 proteins specifically recognized *S. solfataricus* ORB sequences. This interaction is believed to be essential for DNA duplex melting and origin firing. Mutation of three consecutive bases in either of two positions in the SsORB consensus abolished binding of SsCdc6-1 protein to the mutated sequence [6].

A cluster of repeats was identified in the upstream region of the *MthCdc6-1* gene [14] and was shown to contain a “mini-ORB” sequence of 5'-TTACA(g/c)TTGAA(a/c)N-3' [10]. This sequence is a minimized version of the ORB sequences found in other archaea. *MthCdc6-1* shows 40-fold higher binding affinity for mini-ORB sequences than for a random DNA sequence of the same length and cooperative binding to a 181 bp duplex containing three mini-ORB sequences. The affinity of *MthCdc6-1* for the triple repeat is ~40-fold higher than that for a single site and 10-fold higher than expected for DNA containing three independent ORB sites [10].

Here we show that the region predicted to contain the origin of replication in *M. thermautotrophicus* does indeed function as an origin *in vivo*. By presenting mutated ORB sequences to *MthCdc6-1* in a previously described DNA filter binding assay [10] we show that a single guanine that is invariant among all archaeal origin sequences described to date is critical for Cdc6 binding. In addition, we show that a number of single point mutations within the mini-ORB consensus have little or no effect on *MthCdc6-1* binding *in vitro*, but that the DNA sequence required for tight binding of *MthCdc6-1* extends beyond the 13 bp mini-ORB

consensus previously reported. We have generated mutant Cdc6 proteins that confirm a R334A mutation results in loss of origin sequence binding. In addition, we have shown that a R334K Cdc6 variant shows promiscuous binding of origin sequences where the conserved guanine has been changed and that a R334Q variant displays altered origin sequence specificity. The R334Q protein no longer binds the wild-type origin sequence, but recognizes sequences where the guanine has been replaced by a cytosine or adenosine. Our data suggest that a conserved mechanism is employed by the Class I archaeal Cdc6/ORC proteins in specifically recognizing replication origin sequences.

Experimental

M. thermautotrophicus *Growth* – cultures of *M. thermautotrophicus* DSM1053 were grown in defined mineral-salt media [22] buffered with 50 mM MOPS (pH 7.4 at 20°C) at 58°C under strictly anaerobic conditions in a 3 L vessel (Applikon, Ltd) sparged with 240 ml/min H₂/CO₂ and mixed at 650 rpm. Approximately 10⁷ cells (equivalent to ~ 100 µl of a culture at OD 0.4) were aerobically fixed by the addition of 1 ml ice-cold 77% ethanol and analysed by flow cytometry as described previously [23]. Flow cytometry was performed with an A40 miniFCM (Apogee Flow Systems).

Replication Intermediate Analysis – a 4 kb HindIII fragment containing the putative origin of replication from *M. thermautotrophicus* was cloned into pUC118 to produce pJC064 (“b” in Fig. 1A). The adjacent 5.29 kb HindIII fragment (upstream of *MthCDC6-1*) was cloned into pUC118 to produce pJC070 (“a” in Fig. 1A). Two 2.8 kb fragments were PCR amplified from *M. thermautotrophicus* genomic DNA using the primer pairs: 5’-GATATCTCATCTATCCTTGGGAGTGG-3’ and 5’-CACCTAAACTTGACAGGCTAATGG-3’; 5’-GTTCCACGTTTATACAGGATGCAAGG-3’ and 5’-CTTAAGGAACTCCTGGAAAATGGAGG-3’). The purified products of these PCR reactions were used as a template to generate a control probe to the 5.7 kb HindIII fragment downstream of *MthCdc6-1* (“c” in Fig. 1A). *M. thermautotrophicus* cells were harvested aerobically into 133 mM EDTA / 33% glycerol, rapidly cooled in dry ice/ethanol and recovered by centrifugation (6,500 g, 10 min, 4°C). Cell pellets were lysed by three rounds of freezing in liquid nitrogen and grinding in a pestle and mortar. The lysate was diluted 4-fold (20 mM Tris-HCl, pH 8.0 / 5 mM EDTA / 10% sucrose) and incubated at 60°C for 30 min in 2% SDS.

Genomic DNA was recovered by isopropanol precipitation, RNase digestion and two rounds of phenol:chloroform extraction before being purified by cesium chloride gradient ultracentrifugation. Genomic DNA was dialysed against HindIII restriction buffer plus EDTA before being digested with HindIII and added MgCl₂. Replication intermediates were enriched, separated and visualised as previously described [24] using the probes described above.

Cloning, mutagenesis and isolation of MthCdc6-1 – *MthCdc6-1* (ORF1412) was PCR amplified from genomic DNA and cloned into pQE30 (Qiagen) to produce pJC062. pJC062 was used as a template to produce R334A, R334K and R334Q *MthCdc6-1* variants using QuikChange site-directed mutagenesis (Stratagene). Cultures were grown overnight in 2x LB without IPTG induction and *MthCdc6-1* protein was purified as previously described [10] with the following modifications: *MthCdc6-1* protein was immobilised on Talon Affinity Resin (BD Biosciences) pre-equilibrated with lysis buffer. The bound protein was washed with 10 column volumes of lysis buffer containing 75 mM imidazole before *MthCdc6-1* was eluted (40 mM Tris-HCl, pH 8.0 / 400 mM potassium acetate / 20% glycerol / 125 mM imidazole / 1 mM 2-mercaptoethanol / 1 mM PMSF / 1 mM pepstatin) and desalted using a NAP-25 column (GE Biosciences). Aliquots of *MthCdc6-1* were snap frozen in liquid nitrogen and stored at -80°C. Protein purity was checked by SDS-PAGE.

DNA binding assays – Protein-DNA binding was assessed using a modification of the method previously described [10]. Oligonucleotides labelled with [γ -³²P]ATP and T4 polynucleotide kinase were annealed to a complementary oligonucleotide. *MthCdc6-1* was desalted into binding buffer (40 mM bis-Tris propane, pH 7.0 / 70 mM, potassium acetate / 20% glycerol / 3 mM DTT / 2 mM MgCl₂ / 0.1 mg/ml BSA) using a NAP-10 column (GE Biosciences). 100 μ l reactions containing 0-1.2 mM *MthCdc6-1* protein, 0.8 nM substrate and binding buffer were assembled in microtitre plates and assayed as described [10]. Binding constants for the wild-type Cdc6 protein and Orb8 DNA sequence were calculated by curve fitting a simple independent binding sites model equation to experimental data in SigmaPlot [10, 25]. The resulting B_{max} value (65%) was then used to constrain fitting of the same equation to all other data. Where DNA binding was strongly disrupted we observed a poor fit of the equation and K_{d(app)} was not determined (n.d.).

Results

GC skew methods have predicted a single origin of replication from the genomic sequence of *M. thermautotrophicus* [14]. Consistent with this notion, a number of direct and inverted repeats have been identified in this region, although the number and exact sequence of the repeats varies between reports [10, 14]. We observed 11 repeats of the previously reported consensus 5'-TTACA(c/g)TTGAA(a/c)N-3' [10], where the repeats are regularly spaced some 49-53 nucleotides apart, possibly suggesting a phased binding of *MthCdc6-1* (Supp. Fig. 1). In addition to the mini-ORB sequence, we noticed that all 11 sequences contained an additional conserved region 3' to the mini-ORB consensus of 5'-NN(a/t)(t/c)(c/g)NCNC-3' (Supp. Fig. 1).

Analysis of the putative origin-containing region was performed by probing genomic DNA isolated from exponentially growing *M. thermautotrophicus* cells and separated by neutral/neutral (N/N) two-dimensional agarose gel electrophoresis [26]. A functional origin of replication was detected on hybridisation of the putative origin sequence using a 4 kb probe (Fig. 1), as indicated by the bubble arc structure (arrow, Fig. 1C) and diminished Y-arc characteristic of a firing origin of replication. The discontinuity between the top of the bubble arc and the Y-arc (the so-called "bubble-to-Y transition") is also highly characteristic of a firing origin of replication. In order to demonstrate that the bubble structure was specific to the fragment probed, the blot was stripped and hybridised to ~5 kb probes specific for adjacent fragments (Fig. 1B, D). Both probes yielded a complete Y-arc, and the absence of any bubble arc structures, indicating that these regions of DNA contained only fork structures and suggesting that DNA replication in *M. thermautotrophicus* is bidirectional and proceeds via the movement of a pair of replication forks (Fig. 1). The 2D analysis consistently yielded the same results when using DNA isolated from a number (n=5) of different exponentially growing cultures (e.g. Fig. 2). We took additional samples from an actively growing culture of *M. thermautotrophicus* to examine the initiation of replication as cells entered stationary phase (Fig. 2). Growth stage was confirmed by flow cytometry (Fig. 2 A, C, E) before replication intermediates were analysed. We observed a decrease in the frequency of replication initiation from stationary phase samples (Fig. 2 F), but replication did not appear to completely cease. These results are typical of an asynchronous microbial culture, where a small population of cells can continue to proliferate even under limiting growth conditions. Our data are also consistent with the growth characteristics of methanogens, which are not known to produce a dormant stage or sporulate, as they normally exist in hydrogen-limited environments which would result in slow but continuous ('cryptic')

growth. Although GC skew analysis predicted only a single origin of replication in the genome of *M. thermotrophicus* [14], we also probed a 11.5 kb DNA locus around the ORF encoding the *MthCdc6-2* gene (chromosomal position 1,460,708 – 1,472,211 bp). Our results showed no bubble structures in this DNA region (not shown). These results indicate that there is no replication origin near the *MthCdc6-2* gene.

The sequence requirements of *MthCdc6-1* DNA binding were examined using a filter binding assay [10]. 34 bp duplexes of ORBs 8, 9, 10 or 11 containing the previously described mini-ORB consensus sequence [10] plus flanking regions were incubated with purified *MthCdc6-1* (Fig. 3A). ORB8 DNA binding and apparent dissociation constants ($K_{d(app)}$) for *MthCdc6-1* (Table 1) were comparable to those published previously, even though the maximal protein concentration used was lower than in those experiments [10]. Interestingly, variations in the sequence of the regions flanking the ORB consensus had a small but measurable effect on DNA binding (Fig. 3A, Table 1). In contrast to DNase I footprinting experiments where *SsCdc6-1* was able to recognize ORB sequences from both *Pyrococcus* and *Halobacterium* [6], reactions containing *MthCdc6-1* and ORB sequences from *P. abyssi* or *S. solfataricus* resulted in reduced binding (Fig. 3A). A 34 bp duplex of random sequence was not bound by *MthCdc6-1*, but the addition of increasing amounts of this and other competitor DNA into reactions containing the origin sequence resulted in more specific origin binding (Fig. 3B, Table 1). These results were obtained using two different DNA duplexes of random sequence, indicating that the observation is not a sequence-based artefact, although it does not exclude the possibility of an effect of the filter-binding assay.

In order to more exactly determine the sequence requirements of *MthCdc6-1* for specific ORB binding, we challenged the wild-type Cdc6 protein with additional 34 bp duplexes containing point mutations in the consensus sequence. Interestingly, we found that most single mutations (SM) within the consensus sequence appeared to have little effect on DNA binding (Fig. 3C, Table 1). The exceptions to this were SM6 and SM7, where mutations at the same G (to C or A respectively) within the mini-ORB consensus sequence effectively resulted in a complete loss of binding of *MthCdc6-1* (Fig. 3C, Table 1). Two double mutations (DM1 and DM3) within the consensus sequence also resulted in abrogation of binding (Fig. 3D), one of which, DM3 overlapped with SM6/7. A third double mutation in the consensus sequence, DM2, resulted in a significant decrease in total DNA binding by *MthCdc6-1*. Consistent with the extended *MthORB*

consensus sequence we observed (above), a similar partial inhibition of *MthCdc6-1* binding was observed with duplexes where double mutations were made 3' to the consensus sequence (Fig. 3D, Table 1), implying that these 3' bases are also important for *MthCdc6-1* / ORB interactions. This was confirmed by replacing 13 bp of sequence 3' to the original consensus sequence with a random sequence, which resulted in an almost complete loss of *MthCdc6-1* binding (Fig. 3D, Table 1).

In order to test the hypothesis that the conserved guanine identified in our experiments interacts directly with the conserved Arg³³⁴ previously characterized [10], we generated and purified three *MthCdc-1* variants: R334A, R334K, R334Q, with the selection of amino acids being based on well characterized bidentate protein-DNA interactions [27] (Fig. 4). These proteins, along with the wild-type protein were subjected to *in vitro* DNA binding assays in the presence of Orb8, SM6 (G to C), SM7 (G to A) and random sequence duplexes (Fig. 5). As previously reported [10], the R334A mutation in *MthCdc6-1* disrupted the interaction of Cdc6 with Orb8. R334A also showed poor binding of SM6, SM7 and the random sequence (Fig. 5A). The R334K variant of Cdc6 showed a modest decrease in binding of Orb8 compared to wild-type, but was also able to bind the mutated origin sequences SM6 and SM7 (Fig. 5B). The R334Q variant was also able to bind SM6 and SM7, but in contrast to R334K or wild-type Cdc6, this protein did not bind the Orb8 sequence, thus showing an altered DNA sequence binding specificity compared to the wild-type *MthCdc6-1* (Fig. 5C).

Discussion

DNA replication in the archaea is interesting both for understanding this fundamental process in the third domain of life and as a way of probing the mechanisms that underlie eukaryotic replication. The replication initiation process is still only partially understood in archaea, and well-characterised model systems are required for mechanistic studies. Within the archaea, the replication proteins of *M. thermautotrophicus* are among some of the best characterized to date. *M. thermautotrophicus* possesses a strikingly organized array of ORB sequences. Evenly spaced repeats flanking both sides of an AT-rich region upstream of the *MthCDC6-1* gene (Fig 1A) were first reported to contain twelve 13mers with significant homology to repeats found by GC skew analysis in *Pyrococcus* species [14], although we find only eleven (Supp. Fig. 1). While the repeats of *P. furiosus* were identified as homologous to the

functional ORBs identified in *Sulfolobus* those of *M. thermautotrophicus* were not [6]. Limited homology between the mini-ORB sequences of *M. thermautotrophicus* and those of *Sulfolobus* and *Pyrococcus* has since been noted [10]. Our study supports the notion that the active origin of replication in *M. thermautotrophicus* contains eleven full-length ORB sequences that are homologous to those previously identified in other archaea. Our data provide the first experimental confirmation that the putative site of DNA replication initiation in *M. thermautotrophicus* is in fact a functional replication origin.

As previously reported [9, 10], we have shown that *MthCdc6-1* can bind to Orb8 sequences *in vitro*. We have additionally shown that this interaction is improved in the presence of competitor DNA. One possible explanation for this observation is that *MthCdc6-1* contains two separate DNA binding sites, and that binding of non-specific DNA to one site causes some change in the protein that stabilizes the binding of the specific DNA at the other site. We have determined that changes to the DNA sequences 3' to the mini-ORB consensus affect *MthCdc6-1* binding. Replacement of the entire 3' region of the duplex used for DNA binding resulted in a complete loss of DNA binding, suggesting that the originally reported mini-ORB consensus is not sufficient for effective Cdc6/ORB interactions. Mutations located outside the mini-ORB sequence in the conserved 3' residues (DM4, DM5) did not completely disrupt binding, although they did have a significant effect, supporting the notion that specific contacts between *MthCdc6-1* and the 3' sequence could be crucial for *MthORB* binding. The limited conservation in this region of the ORB suggests that the 3' sequence may play a role in enhancing Cdc6 / DNA binding, or stabilisation of the protein-DNA complex, rather than in sequence-specific recognition by *MthCdc6* protein. It should be noted that the sequences 3' to the mini-ORB identified here were present in the original *MthCdc6-1* DNA binding studies [10], and can also be identified in a synthetic triple repeat of the mini-ORB that was shown to be bound by *MthCdc6-1* [11] (Supp. Fig. 2). Thus our novel findings are completely consistent with other studies in this species and we are confident that the 3' sequences identified by our experiments form part of the essential *MthORB* sequence required for *MthCdc6-1* binding, although their contribution was not previously appreciated. ORB sequences from different archaeal species show weaker homology in this region and the *MthORB* diverges greatly from the consensus. We believe that the differences in this part of the ORB sequence are likely sufficient to explain the

differences between the *SsCdc6-1* and *MthCdc6-1* proteins with regards to their abilities to bind ORBs from other species.

All the double mutations we tested affected the total amount of DNA bound by *MthCdc6-1*. DM1 (CT to TG) showed a very strong reduction in binding that did not occur when either of these mutations was tested alone, suggesting that these bases may be involved in a cooperative interaction with Cdc6 protein residues. Two of the seven single point mutations that we tested in the consensus sequence had a significant effect on Cdc6-1 binding (SM6, G to C and SM7, G to A), and the location of this guanine residue corresponds directly to one of the “triple mutant” residues shown to disrupt *Sulfolobus* Cdc6-1 / ORB interactions [6]. Interestingly, a guanine corresponding to the residue mutated in SM6 is completely conserved in the ORB sequences of all archaeal species where an origin has been described to date (Fig. 6A). Furthermore, guanine but not cytosine can form two hydrogen bonds with arginine via the major groove of a double helix without disrupting base pairing [27, 28]. Given that the R334A mutation in *MthCdc6-1* has previously been shown to abolish ORB binding [10], the most obvious mechanism by which the SM6 mutation might effect changes in *MthCdc6-1* binding is if this base interacts with Arg³³⁴ via hydrogen bonding. Such an interaction could be key either in the initial positioning of *MthCdc6-1* on the DNA, or in stabilizing the protein-DNA interaction once the specific site has been recognized. Arg³³⁴ is also completely conserved in the Class I archaeal Cdc6 proteins known to bind to origin sequences, but not in the Class II Cdc6 variants that do not (Fig. 6B). Our mutagenesis of *MthCdc6-1* has confirmed that Arg³³⁴ is essential for the Cdc6 interaction with guanine. Mutating this amino acid to lysine reduced the specificity of binding and enabled Cdc6 to bind to a wider range of origin sequences. Most interestingly, we were able to show that a *MthCdc6-1* variant containing glutamine at position 334 displayed altered sequence specificity as it will no longer bind to the wild-type guanine-containing Orb8 sequence, but is now able to recognize cytosine or adenine in its place, albeit with a lower affinity than we observe for the wild-type interaction. Thus it appears that the potential to form hydrogen bonds through a major groove / recognition helix guanine-arginine interaction, as described here, is part of a generally conserved mechanism employed in replication origin binding by all archaeal ORC/Cdc6 proteins. Variation in the surrounding DNA sequences and changes to the amino acid sequence of the recognition helix in the WHD of Class I Cdc6 proteins could thus explain the sequence specificity observed in Cdc6-1 proteins and ORBs from different archaeal species.

Acknowledgements

We thank Alison Walters, Cyril Sanders and Marjan van der Woude for comments on the manuscript and Mrs Magdaléna Országhová for excellent technical assistance. This work was funded by a BBSRC David Phillips Research Fellowship (J.P.J.C.), grant APVT-51-024904 from the Science and Technology Assistance Agency (Slovak Republic) and a Royal Society International Short Visit / Fellowship (A.I.M.).

Additional Note

While this manuscript was under review, crystal structures of *S. solfataricus* cdc6-1 and cdc6-3 forming a heterodimer bound to *ori2* DNA (Dueber EL, Corn JE, Bell SD, Berger JM., 2007, *Science* **317**:1210-3); and that of *A. pernix* cdc6-1 bound to an origin sequence (Gaudier M, Schuwirth BS, Westcott SL, Wigley DB., *Science* **317**:1213-6) have been published. Both papers provide structural data fully consistent with our conclusions. In particular, Gaudier *et al* clearly show the essential interaction through the major groove between a guanine (Gd10) in the ORB and an arginine (R345) in Cdc6-1 of *A. pernix*, which correspond to those identified in our system (and predicted from our results). In addition the structures indicate that the less-conserved 3' ORB sequence is important for DNA binding and suggest that there are two independent DNA binding sites that interact with each other – clearly supporting our data concerning non-sequence-specific competitor. The fact that our data are completely consistent with the findings of these crystal structures serves to reinforce the conserved nature of these interactions across all archaea. We have created models where substitution of lysine or glutamine for R345 in the *A. pernix* structure was visualised (see Supp. Fig. 3; thanks to Dale Wigley for providing the crystal co-ordinates before release and Fred Anston for modelling assistance). This has allowed us to visualize the conserved hydrogen bonds of the lysine substitution, and the potential of glutamine to disrupt this interaction and contribute to the binding phenotypes observed for our mutant *MthCdc6-1* proteins.

References

- 1 Bernander, R. (2003) The archaeal cell cycle: current issues. *Mol Microbiol* **48**, 599-604
- 2 Edgell, D. R. and Doolittle, W. F. (1997) Archaea and the origin(s) of DNA replication proteins. *Cell* **89**, 995-998
- 3 Slesarev, A. I., Mezhevaya, K. V., Makarova, K. S., Polushin, N. N., Shcherbinina, O. V., Shakhova, V. V., Belova, G. I., Aravind, L., Natale, D. A., Rogozin, I. B., Tatusov, R. L., Wolf, Y. I., Stetter, K. O., Malykh, A. G., Koonin, E. V. and Kozyavkin, S. A. (2002) The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc Natl Acad Sci U S A* **99**, 4644-4649
- 4 Zhang, R. and Zhang, C. T. (2004) Identification of replication origins in the genome of the methanogenic archaeon, *Methanocaldococcus jannaschii*. *Extremophiles* **8**, 253-258
- 5 Zhang, R. and Zhang, C. T. (2005) Identification of replication origins in archaeal genomes based on the Z-curve method. *Archaea* **1**, 335-346
- 6 Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R. and Bell, S. D. (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* **116**, 25-38
- 7 Berquist, B. R. and DasSarma, S. (2003) An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, *Halobacterium* sp. strain NRC-1. *J Bacteriol* **185**, 5959-5966
- 8 Liu, J., Smith, C. L., DeRyckere, D., DeAngelis, K., Martin, G. S. and Berger, J. M. (2000) Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. *Mol Cell* **6**, 637-648
- 9 Singleton, M. R., Morales, R., Grainge, I., Cook, N., Isupov, M. N. and Wigley, D. B. (2004) Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. *J Mol Biol* **343**, 547-557
- 10 Capaldi, S. A. and Berger, J. M. (2004) Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon *Methanothermobacter thermoautotrophicus*. *Nucleic Acids Res* **32**, 4821-4832
- 11 Kasiviswanathan, R., Shin, J. H. and Kelman, Z. (2006) DNA binding by the *Methanothermobacter thermoautotrophicus* Cdc6 protein is inhibited by the minichromosome maintenance helicase. *J Bacteriol* **188**, 4577-4580

- 12 Gajiwala, K. S. and Burley, S. K. (2000) Winged helix proteins. *Curr Opin Struct Biol* **10**, 110-116
- 13 Grainge, I., Gaudier, M., Schuwirth, B. S., Westcott, S. L., Sandall, J., Atanassova, N. and Wigley, D. B. (2006) Biochemical analysis of a DNA replication origin in the archaeon *Aeropyrum pernix*. *J Mol Biol* **363**, 355-369
- 14 Lopez, P., Philippe, H., Myllykallio, H. and Forterre, P. (1999) Identification of putative chromosomal origins of replication in Archaea. *Mol Microbiol* **32**, 883-886
- 15 Zhang, R. and Zhang, C. T. (2002) Single replication origin of the archaeon *Methanosarcina mazei* revealed by the Z curve method. *Biochem Biophys Res Commun* **297**, 396-400
- 16 Zhang, R. and Zhang, C. T. (2003) Multiple replication origins of the archaeon *Halobacterium* species NRC-1. *Biochem Biophys Res Commun* **302**, 728-734
- 17 Maisnier-Patin, S., Malandrin, L., Birkeland, N. K. and Bernander, R. (2002) Chromosome replication patterns in the hyperthermophilic euryarchaea *Archaeoglobus fulgidus* and *Methanocaldococcus* (*Methanococcus*) *jannaschii*. *Mol Microbiol* **45**, 1443-1450
- 18 Lundgren, M., Andersson, A., Chen, L., Nilsson, P. and Bernander, R. (2004) Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A* **101**, 7046-7051
- 19 Matsunaga, F., Forterre, P., Ishino, Y. and Myllykallio, H. (2001) In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc Natl Acad Sci U S A* **98**, 11152-11157
- 20 Norais, C., Hawkins, M., Hartman, A. L., Eisen, J. A., Myllykallio, H. and Allers, T. (2007) Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet* **3**, e77
- 21 Robinson, N. P. and Bell, S. D. (2007) Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc Natl Acad Sci U S A* **104**, 5806-5811
- 22 Nölling, J., Frijlink, M. and de Vos, W. M. (1991) Isolation and characterisation of plasmids from different strains of *Methanobacterium thermoformicicum*. *J Gen Microbiol* **137**, 1981-1986

- 23 Majernik, A. I., Lundgren, M., McDermott, P., Bernander, R. and Chong, J. P. (2005) DNA content and nucleoid distribution in *Methanothermobacter thermautotrophicus*. *J Bacteriol* **187**, 1856-1858
- 24 Liang, C., Weinreich, M. and Stillman, B. (1995) ORC and Cdc6p interact and determine the frequency of initiation of DNA replication in the genome. *Cell* **81**, 667-676
- 25 Clarke, A. R. (1996) Enzymology. In *LabFax* (Engel, P. C., ed.), pp. 203-204, BIOS Scientific Publisher and Academic Press
- 26 Brewer, B. J. and Fangman, W. L. (1987) The localization of replication origins on ARS plasmids in *S. cerevisiae*. *Cell* **51**, 463-471
- 27 Luscombe, N. M., Laskowski, R. A. and Thornton, J. M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* **29**, 2860-2874
- 28 Bloomfield, V. A., Crothers, D. M. and Tinoco Jr, I. (2000) *Nucleic Acids: Structures, Properties, and Functions*. University Science Books, Sausalito, CA

Legends

Fig. 1. – The *M. thermautotrophicus* origin of replication is functional *in vivo*.

A, Cartoon of the genomic region of *M. thermautotrophicus* showing the presence of the fragments used to probe the gels in *B-D* (marked as a-c, respectively), location of nearby genes and the relative position and orientation of ORB sequences. An A/T-rich region that could be the site of initial unwinding is indicated. *B*, 2D gel probed with genomic fragment upstream of putative replication origin showing only a Y-arc consistent with the presence of a fork. *C*, 2D gel probed with genomic fragment containing the putative origin of replication. Bubble arc is indicated with an arrow. *D*, 2D gel probed with genomic fragment downstream of putative replication origin showing only a Y-arc consistent with the presence of a fork.

Fig. 2. – Replication initiation decreases but does not cease as *M. thermautotrophicus* cells enter stationary phase.

A, C, E, Flow cytometric analysis of cells taken at different stages of growth show a characteristic reduction in average genome content [23]. *B, D, E*, 2D gels of DNA samples taken from the corresponding growth stages show that initiation frequency decreases, but is not completely lost in stationary phase. Bubble arcs are indicated by arrows.

Fig. 3. – Specific DNA binding by *MthCdc6-1*.

A, Variations in the Orb sequences found in *M. thermautotrophicus* have a small effect on the extent of *MthCdc6-1* binding. Despite similarities in sequence, ORB sequences from other species are not bound by *MthCdc6-1*. *Pa*: *Pyrococcus abyssi*; *Ss*: *Sulfolobus solfataricus*. *B*, Addition of increasing quantities of competitor (random sequence) DNA results in more specific binding of *MthCdc6-1* to Orb8 duplex. *C*, Single point mutations within the Orb8 sequence has little effect on Cdc6-Orb binding other than the mutation of a conserved guanine to cytosine in SM6, which abrogates Orb binding. *D*, Mutation of pairs of nucleotides within the mini-ORB consensus affects the ability of *MthCdc6-1* to effectively bind DNA. Mutation of nucleotides 3' to the mini-ORB consensus also compromises the ability of *MthCdc6-1* to effectively bind DNA. Substrate sequences and $K_{d(app)}$ values (where appropriate) are listed in Table 1. All points are the mean of 3 independent reactions. Error bars indicate ± 1 SD.

Fig.4. – Variants in *MthCdc6-1* may affect Cdc6 / Orb interactions.

A, Coomassie stained SDS-PAGE gel of purified WT and variant *MthCdc6-1* proteins. Position and size of molecular weight markers are indicated. *B*, Likely non-disruptive interactions between amino acid side chains and DNA via the major groove (based on [27]). From top to bottom the interactions illustrated are: arginine – guanine, lysine – guanine and glutamine – adenosine.

Fig. 5. – Altered specificity of binding can be engineered into *MthCdc6-1*.

A, Confirmation that the introduction of the R334A mutation into *MthCdc6-1* disrupts Orb8 as well as all other DNA binding. *B*, An *MthCdc6-1* R334K variant shows slightly reduced affinity for Orb8, but promiscuous binding to the SM6 and SM7 substrates. *C*, The R334Q variant of *MthCdc6-1* cannot bind Orb8, but is able to bind DNA duplexes containing the SM6 and SM7 mutations, demonstrating an altered specificity for DNA sequences for this protein.

Fig. 6. – A conserved mechanism employed in replication origin binding by all archaeal ORC/Cdc6 proteins?

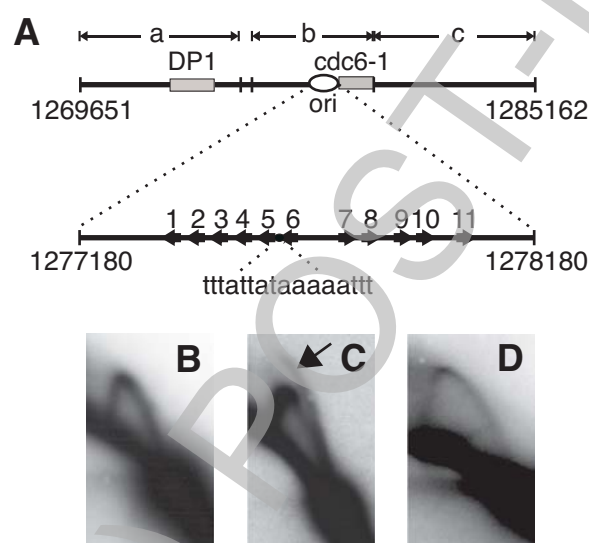
A, The SM6 guanine (highlighted) is conserved in all archaeal origins characterized to date, as shown by an alignment of the consensus sequences of mini-ORB regions from a number of species (based on [10]). *B*, ClustalW alignment of archaeal Cdc6 proteins in the region of Arg³³⁴ (highlighted) showing that this amino acid is completely conserved in all Class I archaeal Cdc6 proteins, but is not present in Class II Cdc6 proteins. Mth – *M. thermautotrophicus*, Sso – *S. solfataricus*, Afu – *A. fulgidus*, Ape – *A. pernix*, Pfu – *P. furiosus*.

Table 1 – kinetic data for the interaction between *MthCdc6-1* and different ORB sequences

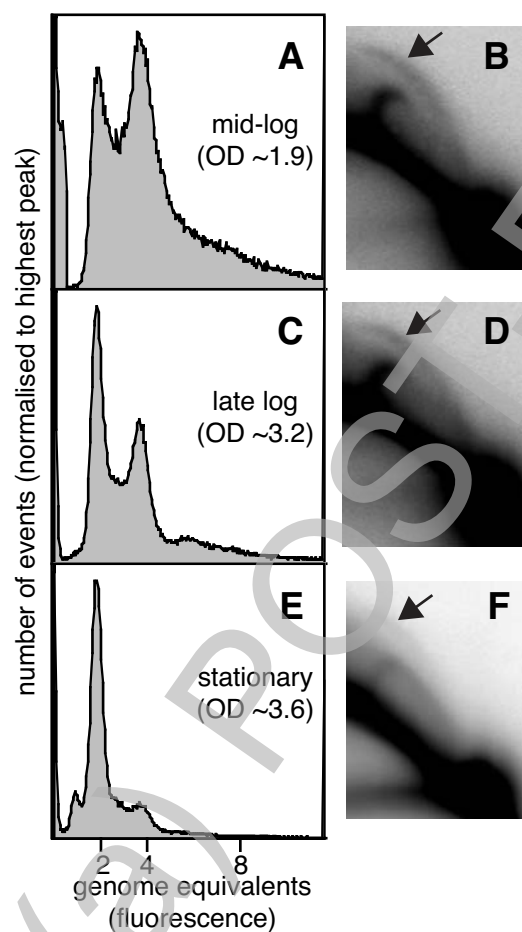
Name	Sequence	$K_{d(app)}$ (μ M)
ORB8	5'-catggtcagaTTACAcTTGAAatggatgtCtCcc-3'	0.35
ORB9	GCA••ATTTT•••••TC••CC••AT	0.43
ORB10	ATCCAGAT•G•••••A•••••C•A•	0.62
ORB11	ATAA•AG•AT•••••CT•CCA•A•••	0.68
ORB8	In the presence of 10-fold excess random competitor	0.30
ORB8	In the presence of 100-fold excess random competitor	0.23
ORB8	In the presence of 1000-fold excess random competitor	0.24
ORB8	In the presence of 1000-fold excess random competitor 2	0.21
SM1	•••••C•••••	0.34
SM2	•••••T•••••	0.51
SM3	•••••T•••••	0.48
SM4	•••••C•••••	0.64
SM5	•••••G•••••	0.43
SM6	•••••C•••••	n.d.
SM7	•••••A•••••	n.d.
DM1	•••••TG•••••	n.d.
DM2	•••••GG•••••	n.d.
DM3	•••••AG•••••	n.d.
DM4	•••••T•T••	n.d.
DM5	•••••G•A•••••	n.d.
3' end	•••••TAATGCCATGGAA	n.d.
<i>Pa</i> ORB	CAAC•C•G••C••G•G••••A•ACC•TGGGGG	7.41
<i>Ss</i> ORB	TTTACCTTAAGTTC•C••G•G•••C•AA•GGGGT	n.d.
Random	ACGTACTGACCAGTTGAGTTCTAATGCCATGGAA	n.d.

$K_{d(app)}$ values for the binding of *MthCdc6-1* to different Orb-sequence containing DNA duplexes. Values were derived by using a simple independent binding site model equation with B_{max} fixed to the value obtained for Orb8 (65%). ORB8: UPPER CASE bases indicate consensus sequence

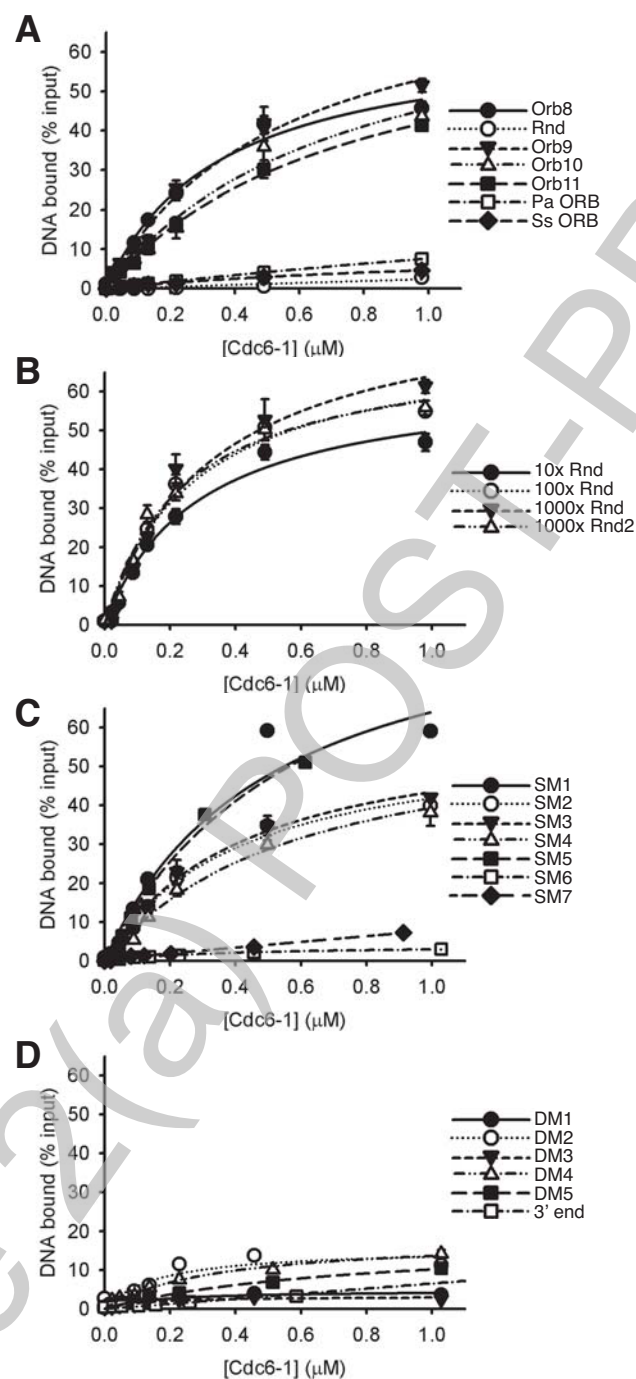
residues, lower case indicate the remaining (non-conserved) sequence. Other sequences: UPPER CASE indicates variation in sequence from ORB8, • indicates identity to ORB8. In all cases, $R^2 \geq 0.98$. n.d. indicates that the $K_{d(app)}$ for these samples was not determined.



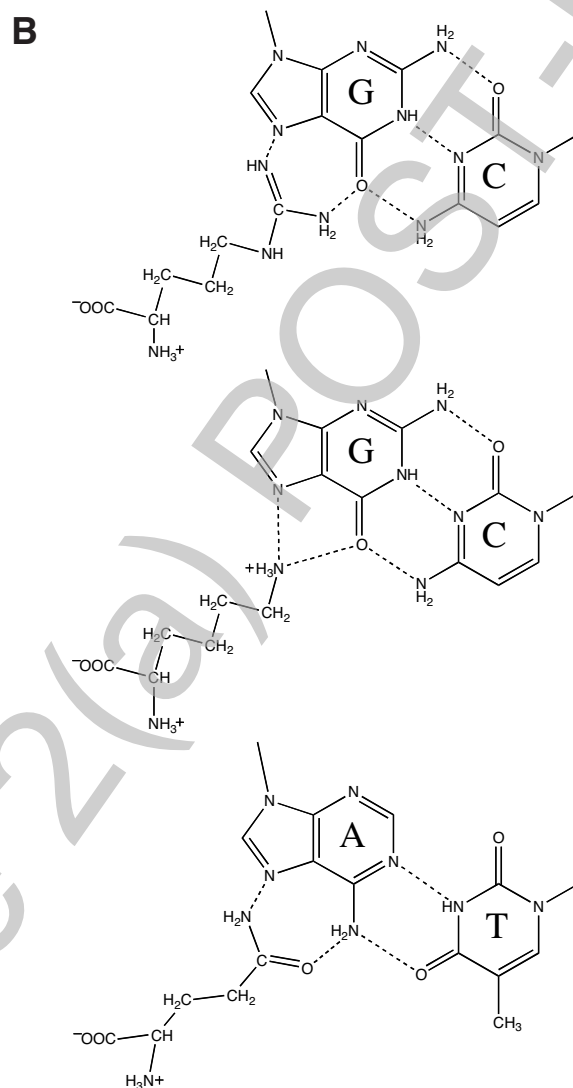
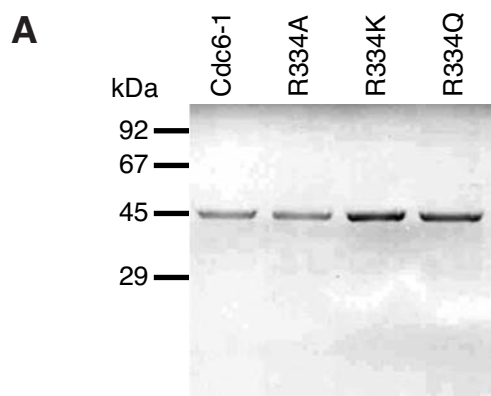
Majerník and Chong, Figure 1



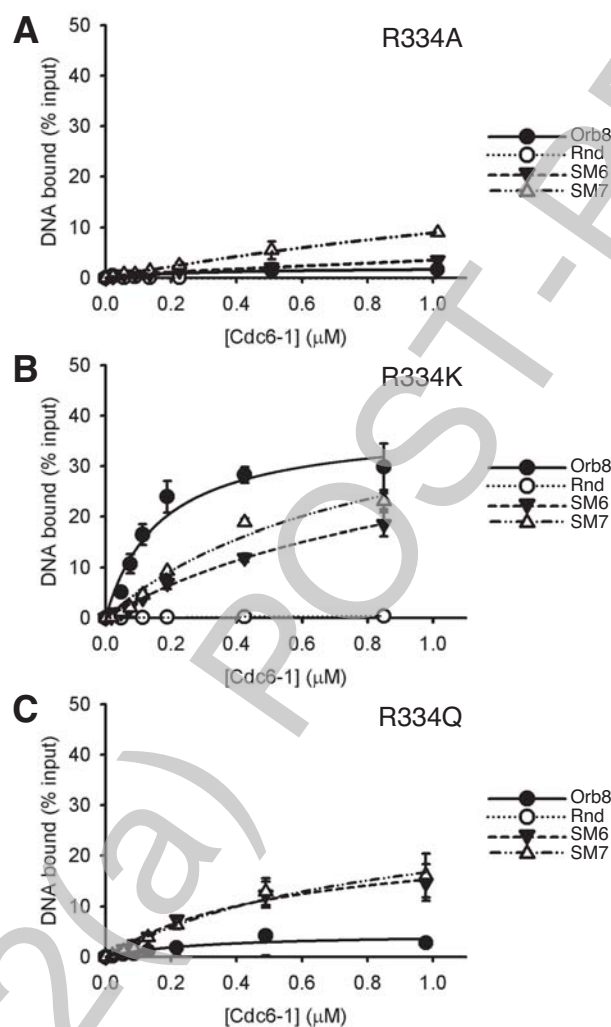
Majerník and Chong, Figure 2



Majerník and Chong, Figure 3



Majerník and Chong, Figure 4



Majernik and Chong, Figure 5

A

Mth	TTACA ^G TTGAA ^A _C
Sso	C ^G TCCA ^{GTG} AAC ^{GAAA}
Afu	TTCCACAGGAAA
Ape	C ^A TCCA ^{CA} GC ^{GAAA} _A
Pfu	T ^T C ^C TCCAGTGAAA
cons	T ^C _A CA GAA ^A _C

B

MthCdc6-1	GDSVSYRRIFDFINELEMLGLISTNTVSRGR
SsoCdc6-1	VEAVTQRRVSDIINELDMVGILTAKVVNRGR
SsoCdc6-3	QKPLSYRRFSDIISELDMFGIVKIRIINRGR
AfuCdc6-1	VDPLTQRRISDLISELDMGLILNSVVISKGR
ApeOrc1	LEHVTLRVSGIISELDMGLIVKSRVVSRRGR
PfuCdc6	LEPLTQRRISDLINELDMGLIINAKVVSRRGR
MthCdc6-2	---SSYSSFNRRIIEKLEFLRLIDTRLTGKGV
SsoCdc6-2	EEPRRHQTQVYEYLRKLKVIGIINTRQSGKGM
AfuCdc6-2	---MSYRKYYNILEKLERLKLIEISFGEKG-
ApeOrc2	VKPRGYTQYHIYLNKHLTSLGLVDAKPSGRGM

: . * . : : : *

Majernik and Chong, Figure 6