Structural Basis of DNA Replication Origin Recognition by an ORC Protein

Martin Gaudier, Barbara S. Schwuricht, Sarah L. Westcott, Dale B. Wigley*

DNA replication in archaea and in eukaryotes share many similarities. We report the structure of an archaeal origin recognition complex protein, ORC1, bound to an origin recognition box, a DNA sequence that is found in multiple copies at replication origins. DNA binding is mediated principally by a C-terminal winged helix domain that inserts deeply into the major and minor grooves, widening them both. However, additional DNA contacts are made with the N-terminal AAA+ domain, which inserts into the minor groove at a characteristic G-rich sequence, inducing a 35° bend in the duplex and providing directionality to the binding site. Both contact regions also induce substantial unwinding of the DNA. The structure provides insight into the initial step in assembly of a replication origin and recruitment of minichromosome maintenance (MCM) helicase to that origin.

Archaal DNA replication and repair processes share closer similarity to those in eukaryotes than to those in eubacteria (1), albeit with fewer proteins and hence complexity. Furthermore, whereas some archaea have a single replication origin (2), others have multiple origins, more like the situation in eukaryotes (3, 4). Archael replication origins are recognized by proteins with homology to eukaryotic origin recognition complex (ORC) and Cdc6 proteins (often annotated as ORC/Cdc6 but referred to here as ORC). The number of ORC proteins varies between archaea but is usually one or two proteins, although it can be as many as 14 (4).

Crystal structures of two archaeal ORC proteins have been determined (5, 6). The proteins comprise two domains: an AAA+ domain (7) and a C-terminal domain of the winged helix (WH) family, a structural motif commonly used to bind to specific nucleic acid sequences (8). The isolated AAA+ domain retains adenosine triphosphatase (ATPase) activity, and the WH domain binds to DNA (6, 9).

Analysis of archaeal genome sequences revealed a series of short (13 bp) conserved repeats located close to ORC genes that were proposed to be a signature for probearchal replication origins (10), which was confirmed experimentally in Pyrocococcus abyssi (11, 12). The later study revealed that two longer repeats were located on either side of an AT-rich region named a duplex unwinding element (DUE). Similar repeats flank a DUE in a region of chromosomal DNA from Halobacterium that conveys autonomous replication to plasmids in that species (13). These extended repeat sequences were the origin recognition box (ORB) elements later identified at a replication origin in Sulfolobus solfataricus (14). The ORC1 protein of S. solfataricus footprints at these ORB elements in both P. abyssi and Halobacterium, demonstrating their conservation across species. Curiously, not all of the origins in Sulfolobus contain full-length ORB elements but instead have shorter sequences called mini-ORBs (14), also found at a proposed origin in Methanobacterium thermoautotrophicum (15). The situation in Sulfolobus is complicated further because the three origins are quite different from one another and they all contain binding sites for multiple ORC proteins (14, 15).

Using the sequence of the conserved ORB elements, Robinson et al. proposed the location of a replication origin for A. pernix (14), which was

Cancer Research UK Clare Hall Laboratories, London Research Institute, Blanche Lane, South Mimms, Potters Bar, Herts EN6 3LD, UK.

*To whom correspondence should be addressed. E-mail: Dale.Wigley@cancer.org.uk

References and Notes

17. See the nomenclature section in the supporting online material. Materials and methods are available as supporting material on Science Online.
27. We thank D. Wigley for sharing unpublished data, J. Erzberger for initiator phylogenetic profiles, and the Advanced Light Source beamline 8.3.1 staff for assistance with data collection. E.L.C.D. and J.M.B. acknowledge support from the Miller Institute and NIH (grant GM071747), respectively. S.D.B. is supported by the Medical Research Council of the UK. Coordinates and structure factors have been deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (accession number 2QBY).

Supporting Online Material

www.sciencemag.org/cgi/content/full/117/5842/1210/D1C Materials and Methods

SOM Text

Figs. S1 to S10

Tables S1 and S2

References

Sequence Alignment Files

11 April 2007; accepted 26 June 2007 10.1126/science.1134690

www.sciencemag.org SCIENCE VOL 317 31 AUGUST 2007 1213

REPORTS
Fig. 1. (A) Organization of the Ori1 replication origin in A. pernix. The four ORB sequences are located on either side of the DUE. The sequence of the 5′-to-3′ ("top") strand of ORB4 is shown. (B) Overall structure of the ORC1-DNA complex. (C) Contacts between the AAA+ domain and DNA. Thr122 (T122) (21) contacts the Gd18-Cd5 base pair, L124 main chain oxygen contacts Gd19, and E128 contacts Gd20 via a water molecule. Residues T103, R106, G123, and R132 make direct interactions with the phosphodiester backbone on either side of the minor groove. (D) Insertion of the wing of the WH domain into the DNA minor groove widens it by 5 Å. Residues S366, G368, G371, and K372 interact directly with bases Td17, Gd18, and Gd19 of the complementary strand. G368, G371, K372, and T373 also interact with the DNA backbone on both sides of the minor groove. (E) Insertion of the recognition helix of the WH domain into the major groove. R345 makes a base-specific interaction with Gd10. Residues T343, R346, S348, and R374 contact the DNA phosphate backbone. The R346 side chain is stabilized in a noncanonical conformation by a salt bridge interaction with E353, which enables it to bind the DNA phosphate backbone. In the same way, interaction between S352 and R374 brings the arginine side chain close to the DNA backbone.

Fig. 2. Deformation of the ORB element compared with regular B-form DNA. (A) A comparison between B-form DNA (top) and the ORB4 DNA in the structure (below). The helical axis of the DNA is shown as a blue (B-form DNA) or red (ORB4 DNA) bar running down the center of the duplex. Overall, the DNA is underwound as shown by the twist, which averages 33° per base pair, giving 11 bp per helical turn. DNA parameters were calculated with the program 3DNA (22). (B) Comparison of the base pair roll for each base pair. The roll is greater for all but one base pair in the ORB4 DNA. (C) Comparison of the major and minor groove widths of the ORB element with those for B-form DNA. The minor groove is wider at every position but one along the ORB4 DNA, and the major groove is also wider in all but two places.
later confirmed (16). At this origin (Ori1), there are four ORB elements arranged in pairs on either side of a DUE (Fig. 1A). Only one of the two ORC proteins in *A. pernix* (ORC1) binds at the origin. Consequently, *A. pernix* Ori1 can be considered an archetype for archaeal replication origins such as the origin in *Pyrococcus* (12), the oriC1 origin of *Sulfolobus* (14), at least one of the origins in *Halobacterium* (13), and the five origins identified in *Halofex volcanii* (4), all of which contain full-length ORB elements. Similarities with the eukaryotic system also give insight into conserved elements of eukaryotic origin recognition.

To gain an understanding of replication origin assembly, we determined the crystal structure of the *A. pernix* ORC1 protein complexed with a 22–base pair (bp) canonical ORB element (Fig. 1B). The structure of ORC1 is similar to that of other archaeal ORC proteins (5, 6) comprising AAA+ and WH domains. In common with both previous archaeal ORC protein structures, adenosine diphosphate (ADP) copurifies with the protein (5, 6). The main difference between ORC1 and ORC2 is the region connecting the AAA+ and WH domains. In ORC2, the flexibility of this region allows the protein to adopt several conformations (6). By contrast, the linker in ORC1 is more rigid, a structural difference that is a key element in the interaction between ORC1 and DNA. The principal contact between ORC1 and its DNA target is with the WH domain, as predicted in Fig. 3.

Fig. 3. The stoichiometry of ORC1 binding at the ORB4 element. (A) Isothermal calorimetry data collected by using purified WH domain protein and a 40-mer DNA containing an ORB4 element. (B) Figure mapping the footprint data (C) onto the crystal structure. (C) Deoxyribonuclease I footprints across the ORB4 region. DNA sequences are indicated at the side for reference, with the ORB4 sequence boxed.

Fig. 4. ORC1 proteins bound to ORB elements flanking a DUE would be oriented to place the WH domains, which interact with the MCM helicase, facing the DUE and hence the replication start site.
by crystal structures (5, 6) and biochemical data (6, 9, 16–18). The WH domain interacts with the ORB sequence in a canonical mode (Fig. 1), with the a helix inserted into the major groove and the wing reaching across the adjacent minor groove, similar to structures seen in other WH domain/ DNA complexes (8). ORB elements contain a conserved symmetric dyad with a consensus sequence TCCxxGGA (where x is any base), and mutations in this sequence compromise the ability of ORC1 to recognize ORB elements (14). The four ORB elements of A. pernix Ori1 contain this inverted repeat, which is recognized by the WH domain. Insertion of the recognition helix widens the major groove by over 2 Å, and the central G of the GGA sequence is contacted by an arginine residue (Fig. 1). This arginine is crucial for DNA binding by ORC1 proteins (14, 18, 19). Several residues contact the phosphodiester backbones across the major groove.

In addition to the recognition helix, the wing plays an important part in the interaction between ORC1 and DNA. The wing is longer than that seen in most WH domains, resulting in an extended contact region that spans five base pairs, including the TCC motif (Fig. 1). The wing inserts deeply into the minor groove, causing it to widen by over 5 Å, and direct contacts are made with the DNA bases. This feature is unusual because the wing in other WH domains merely contacts the phosphodiester backbone over the minor groove. A consequence of this widening of the minor groove is that a protein induces substantial unwinding of the duplex at the binding site (Fig. 2).

The symmetry of WH domain binding-site sequences commonly permits binding of two proteins (8). Mutations in the ORB dyad sequence abolish binding of Sulfolobus ORC1 to ORB elements (14). However, the inverted repeat is located at one end of the conserved ORB binding rather than at its center, and there is an additional G-rich sequence flanking it. The crystal structure explains this puzzling feature.

Unexpectedly, there is also a substantial contact between the AAA’ domain and the DNA. A short loop at the end of an a helix inserts into the minor groove in the region of the conserved guanine residues at the 3’ end of the ORB element (Fig. 1). This region is an insertion into the structure of a canonical AAA’ domain. This is characteristic of the Clade II AAA’ proteins that are associated with initiation of DNA replication (20). There is only one direct sequence-specific contact and one water-mediated interaction between the AAA’ domain and the conserved G-rich sequence that it contacts (Fig. 1C). The protein, however, grips one of the DNA phosphodiester backbones through a number of residues. The interaction between the AAA’ domain and the DNA has two effects. The first is a widening of the minor groove (Fig. 2). The second is that the DNA unwinds as this distortion takes place and is bent by 35°. The extended helix connecting the AAA’ and WH domains is a key component in this interaction. This rigid connection pushes the AAA’ domain against the DNA duplex, acting as a brace against which distortion of the DNA can be forced. The net effect of these interactions is extensive unwinding of the DNA. The mean twist per base pair across the DNA is reduced by 3°, resulting in 11 rather than 10 bp per turn and an overall untwisting of over 60° across the ORB element. DNA unwinding is a key aspect in the process of origin assembly.

Although our crystallization conditions contained a ratio of ORC1:DNA of 2:1, our DNA substrate only had a single ORC1 molecule bound. ORC1 forms dimers at higher protein concentrations, and dimerization requires the AAA’ domain (16). Although full-length ORC1 protein is poorly soluble when not bound to DNA, we used the more soluble WH domain to evaluate the stoichiometry of binding at ORB elements using isothermal calorimetry. These data revealed that a single WH domain binds to an ORB element (Fig. 3A), which can be explained by the crystal structure. In canonical WH domain dimers, the wing does not insert into the minor groove. However, in ORC1, insertion of the wing into the minor groove causes it to widen by 5 Å and the associated major groove to narrow by up to 2 Å. Similarly, binding of the helix widens the adjacent major groove by 2 Å. Hence, binding of a symmetric pair at the site requires the major groove to be 4 Å wider than observed, and the recognition helix of a second WH domain cannot be accommodated at the site. Similarly, the associated minor groove would need to widen by over 3 Å to accommodate the wing.

Given the paucity of sequence-specific contacts, particularly those that are not palindromic within the ORB4 sequence, it is unclear how the protein distinguishes sufficiently between the two possible modes of binding. However, of the two possible binding orientations, one is favored by the interaction of the AAA’ domain with the G-rich sequence, providing a directionality of ORC1 binding at an ORB site. This is important because the WH domain of ORC1 interacts with the minichromosome maintenance (MCM) helicase (19). Most archaean origins characterized to date have an ORB element. This is the location of the replication start site (Fig. 4), facing the DUE (4, 12–14, 16), which in A. pernix is the location of the replication start site (16). Consequently, the G-rich sequence in a full-length ORB element controls the arrangement of ORC1 proteins in an orientation appropriate to interact with MCM helicase during replication initiation (Fig. 4).

To evaluate the contacts observed in our crystal structure, we used DNA footprinting for ORC1 protein and the WH domain alone (Fig. 3). At the 5’ end of the top strand, contacts with the WH domain near the dyad repeat produce footprints that are identical in both cases. The 3’ end of the ORC1 footprint extends four bases further than that of the WH domain alone because of contacts with the AAA’ domain. On the bottom strand, there is a small region of sensitivity at the center of the footprint. The structure shows that this strand is protected except for the two unprotected bases that are located between the two contact regions made by the WH domain (Fig. 3). The 3’ ends of both footprints are identical but differ at the 5’ end because of contacts made by the AAA’ domain, extending the footprint by two bases in agreement with the structure. Consequently, the footprints are consistent with the contacts that we observe in the structure, with a single ORC1 molecule binding at an ORB element.

Although we only saw a single ORC1 molecule binding at an ORB, once the initial binding events have taken place then a higher order assembly process begins, the culmination of which is unwinding of the DUE (16). The nature of these later events remains unclear but could involve binding of additional ORC1 molecules combined with structural changes in the origin DNA itself. Whatever these changes may be, they likely involve communication between ORC1 monomers, and this may be manifested as the dimer that we saw for the unbound protein at high protein concentrations (16). Consequently, although the structure we present here enhances our understanding of the initial stage in this complex process, further work will be required to uncover the structural changes that take place as an origin assembles.

References and Notes
7. A. F. Neuwald et al., Genome Res. 9, 27 (1999).
21. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
23. We thank J. Berger for generously sharing data prior to publication and J. Diffley and M. Singleton for helpful discussions. This work was supported by Cancer Research UK and by European Molecular Biology Organization (EMBO) fellowships to M.G. and B.S.S. Coordinates and structure factors have been deposited in the Protein Data Bank with identification code 2V1U.

Supporting Online Material
www.sciencemag.org/cgi/content/full/317/5842/1213/DC1
Fig. S1
Table S1
11 April 2007; accepted 26 June 2007
10.1126/science.1134664

1216 31 AUGUST 2007 VOL 317 SCIENCE www.sciencemag.org
Structural Basis of DNA Replication Origin Recognition by an ORC Protein
Martin Gaudier, Barbara S. Schuwirth, Sarah L. Westcott and Dale B. Wigley

Science 317 (5842), 1213-1216.
DOI: 10.1126/science.1143664

http://science.sciencemag.org/content/317/5842/1213

http://science.sciencemag.org/content/suppl/2007/09/04/317.5842.1213.DC1

http://science.sciencemag.org/content/sci/317/5842/1210.full
http://science.sciencemag.org/content/sci/317/5842/1181.full

This article cites 20 articles, 9 of which you can access for free
http://science.sciencemag.org/content/317/5842/1213#BIBL

http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service