Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

A structural view of the initiators for chromosome replication

Kin Fan On[1,2,3], Matt Jaremko[1,2,3], Bruce Stillman[3] and Leemor Joshua-Tor[1,2,3]

Addresses
[1] W.M. Keck Structural Biology Laboratory, United States
[2] Howard Hughes Medical Institute, United States
[3] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, United States

Corresponding authors: Stillman, Bruce (stillman@cshl.edu), Joshua-Tor, Leemor (leemor@cshl.edu)

Current Opinion in Structural Biology 2018, **53**:131–139

This review comes from a themed issue on **Protein–nucleic acid interactions**

Edited by **Eric Westhof** and **Dinshaw Patel**

https://doi.org/10.1016/j.sbi.2018.08.003

Introduction

Propagation of genetic material is central to the survival and fitness of multicellular organisms. To efficiently duplicate eukaryotic genomes of sizes between $\sim 10^7$ bp (yeast) to $\sim 10^{11}$ bp (plants), replication initiates from multiple start sites called 'replication origins' distributed on multiple chromosomes [1]. This strategy imposes a need for cellular controls to ensure complete DNA replication while also preventing re-replication, which can cause genomic instability, leading to developmental diseases (*e.g.* Meier–Gorlin Syndrome) and cancer [2,3]. Once-per-cell-cycle replication requires the oscillation of cyclin-dependent-kinases (CDKs) and ubiquitin-mediated proteolysis activities [4], establishing two discrete time periods during which origin licensing (assembly of replicative helicases onto origins) and origin firing (activation of helicases) occur in a mutually exclusive manner. Recent progress in acquiring structural information on origin licensing has advanced our mechanistic understanding and regulation of this process. This review discusses current knowledge of origin licensing and compares the origin recognition machinery with other multi-subunit, ATP-driven cellular motors that are required for later stages of DNA replication.

The function of initiators at replication origins

Replicators and initiators

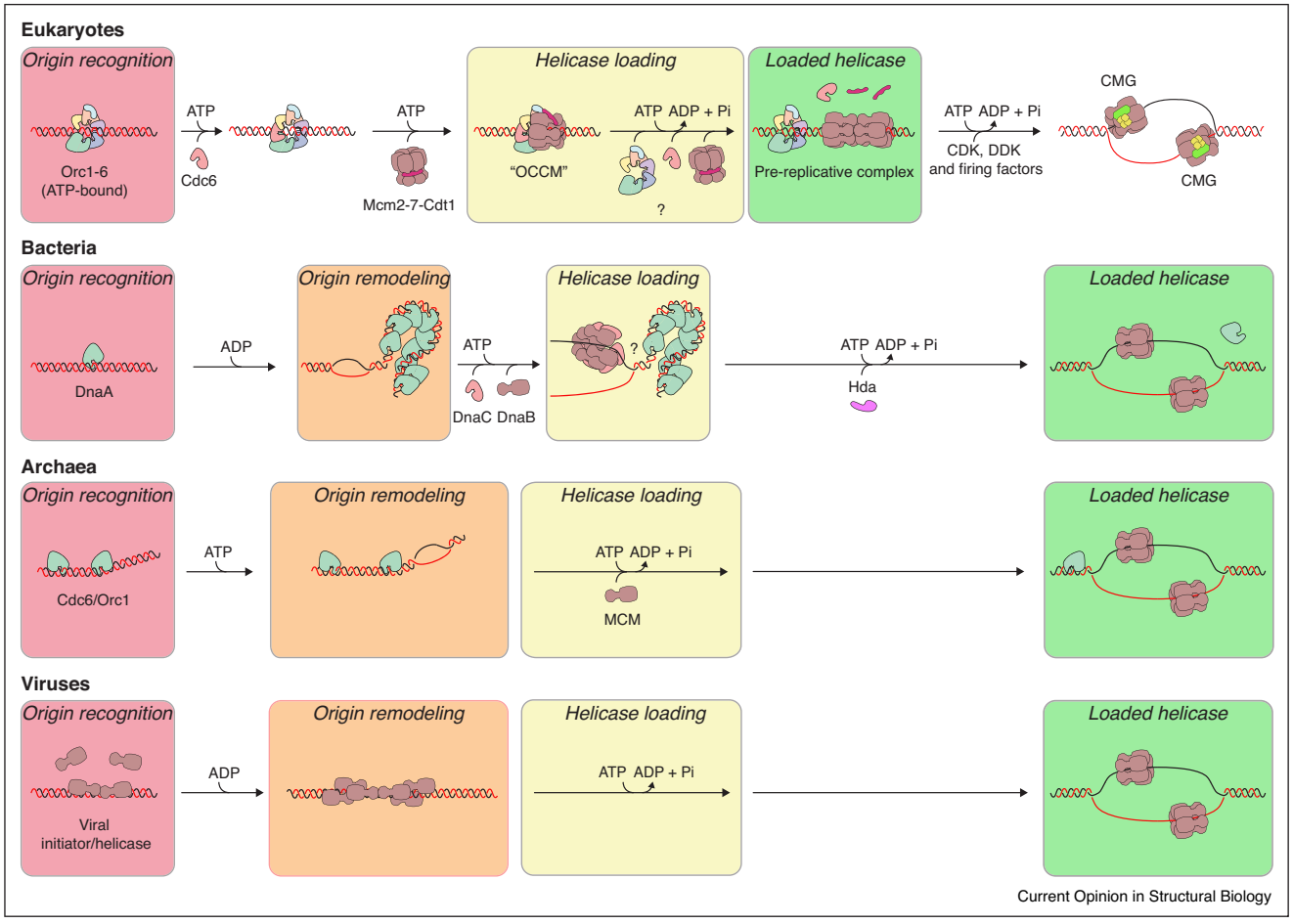
In 1963, a dual-element replication initiation model was proposed to describe the *Escherichia coli* replicon. In this model, a *cis*-acting replicator acts as the genetic element (DNA sequence) that is specifically recognized by a *trans*-acting initiator [5]. This model was confirmed by the identification of the *E. coli* replicator (*oriC*) and initiator (DnaA) respectively [6,7]. Bacteria tend to have a single origin of replication per chromosome, but in eukaryotes, the necessity of multiple origins on each chromosome creates a more complicated mechanism for origin licensing and regulation. The eukaryotic initiator is a complex of the six-subunit Origin Recognition Complex (ORC) and the related Cdc6 ATPase [8••,9,10]. The nature of the eukaryotic *cis*-acting origins, however, is variable and depends on the species studied. Initiator proteins play multiple roles: they recognize or define the start sites within the DNA/chromatin for the initiation of DNA replication, they promote the recruitment of the helicase that unwinds the double helix and they ensure that the process is highly regulated (Figure 1). In viral replication systems, the helicase itself often functions as the initiator.

Origin recognition in different domains of life

Bacterial origin recognition

Genome replication in bacteria initiates from a single origin called *oriC*, which contains a highly AT-rich sequence required for initial melting of the DNA-unwinding element (DUE) [11]. In *E. coli*, the best characterized prokaryotic replication system, replication initiation involves the recognition of a 9-bp consensus sequence within the *oriC* by the initiator protein DnaA, a sequence-specific DNA binding protein that is also an AAA+ family ATPase. The consensus sequence [5′-TTAT(C/A)CA(C/A)A-3′] essential for initiation is called the DnaA-box [12]. Surrounded by weaker binding sites, the R1, R2 and R4 DnaA-boxes are recognized by multiple ATP-bound DnaA subunits with high-binding affinity [13]. The crystal structure of *E. coli* DnaA domain IV bound to the 9-bp DnaA-box showed the insertion of the highly conserved recognition helix from the helix-turn-helix (HTH) motif within DnaA into the major groove of the DnaA-box [14••]. After the initial interaction of DnaA with high-affinity binding sites on *oriC*, multiple DnaA are then recruited to weak sites in an ATP-dependent manner. Oligomerization of multiple DnaA monomers leads to the formation of a high-order nucleoprotein assembly [13,15]. Subsequently, this ATP-bound DnaA assembly facilitates melting at the DUE region and allows for recruitment of the bacterial replicative helicase DnaB

Figure 1



Models for origin recognition and helicase loading in the three domains of life and in viruses. In eukaryotes, bacteria and archaea, the origin is first recognized by the origin recognition machinery, with concomitant DNA remodeling occurring in bacteria and archaea. The replicative helicase is then recruited and directly interacts with the origin recognition complex. ATP binding but not hydrolysis is required for this interaction (see text). In eukaryotes, multiple origin recognition proteins may be required for loading a double hexameric form of the helicase. After helicase loading, ATP hydrolysis allows for reiterate assembly of helicases on other locations on the DNA. In viruses (human papillomavirus is shown as an example in this case), the viral helicase often serves the function of both origin recognition and self-assembly onto DNA into the mature helicase. In papillomavirus, a double-trimeric assembly of the helicase complex, linked to origin melting, is a prerequisite for the fully functional helicase complexes to be formed in subsequent steps in an ATP hydrolysis-dependent manner (for simplicity, other auxiliary factors are omitted in the diagram).

onto single-stranded DNA [16]. The ATPase activity of DnaA is activated by its interaction with a complex consisting of the Hda protein and the β-clamp of DNA polymerase that are loaded as DNA synthesis begins at *oriC*, limiting initiation of replication to once per cell cycle (Figure 1) [17,18].

Archaeal origin recognition

Replication initiation in archaea combines features from bacteria and eukaryotes (Figure 1). Most archaeal genomes contain one origin, while others have several [1]. In the archaeal *oriC*, DNA sequences called the 'origin recognition boxes' (ORBs) are the target sites recognized and bound by an initiator called the Cdc6/Orc1 protein, a AAA+ protein with a primary structure similar to the Orc1 and Cdc6 proteins of the eukaryotic initiator [19]. ORBs, which are clustered within the origin, can be categorized into two main types: long ORB motifs (22–35 bp) [20•,21•,22], and short ORB motifs (also called miniORBs) (12–13 bp) [1].

A crystal structure of Cdc6/Orc1 in complex with an ORB showed the main DNA contacts occur through the insertion of a helix from the C-terminal winged-helix domain (WHD) into the major groove, while the wing forms an unusually deep insertion into the adjacent minor groove [20,21•]. The AAA+ domain makes another contact with an adjacent G-rich sequence [20,21•]. These three binding events result in significant widening of both major and minor grooves, resulting in a bend of the DNA of up to 35°. This deformation may induce local unwinding of the