

Initiation of Replication

This section will begin with a brief discussion of chromosomal DNA sequences that sponsor the initiation of replication.

Replicons and origins of replication

The unit of DNA replication is called the replicon. It is at the level of the replicon that DNA replication is regulated. Consider that DNA replication must be coordinated with the cell cycle, which demands a strict alteration of DNA replication and mitosis. Thus each replicon must be replicated at the appropriate time, and must be replicated only once per cell cycle (i.e. once per S phase).

Replicons have a unique origin of replication (ori), where replication initiates. (Ori's are sometimes called replicators, but we will use the term ori, or origin.) Origin sequences bind initiator proteins that promote the formation of replication forks. A replicon may also have a discrete terminus, where replication stops. Replicons may be unidirectional or bidirectional, depending on whether one or two replication forks advance from the origin. In the case of cellular chromosomes (prokaryotic and eukaryotic), replication occurs bidirectionally.

Simple origins (those from bacterial chromosomes, viral chromosomes, and chromosomes of lower eukaryotes such as budding yeast) can be identified genetically as cis-acting sequences that sponsor the replication (extrachromosomal maintenance) of DNA molecules in which they are inserted. These simple origins are modular in nature, and consist of a series of short sequence blocks spaced over a region of 100-200 bp.

In budding yeast, for example, replication origins have been identified as autonomously replicating sequences (ARS). ARS's were discovered by fragmenting the yeast genome by restriction endonuclease digestion, and inserting the fragments into plasmids containing a selectable marker (e.g. nutritional marker or antibiotic resistance) (Fig. 59). These plasmids were then used to transform yeast, and cells capable of forming colonies contained plasmids that were capable of replicating in yeast cells; i.e. the DNA fragment inserted into the plasmid contained an ARS. Some ARS sequences have been confirmed as bona fide origins of replication.

In contrast, complex oris (those found in the chromosomes of multicellular eukaryotes) usually cannot be identified in this way. That is, they do not behave as ARS's, probably because their function requires chromatin structure. Thus, relatively few oris from higher organisms have been

identified, and relatively little is known of their structure. Most encompass several kilobases of DNA and it has been difficult to identify critical sequences within them. That is, multiple sequence elements over a large stretch of DNA contribute to origin function, and these sequences seem to be redundant.

On the other hand, in some early embryos (*Drosophila* and *Xenopus*) origins appear to require little or no sequence specificity, perhaps to allow for a very rapid S phase.

Key to understanding how cells regulate DNA synthesis is understanding the nature of DNA replication origins. Replication origins play two important roles in DNA replication:

- They provide discrete starting points where replication begins. This ensures that each time the DNA must be replicated, all the necessary components of the replication machinery are brought to the correct location for assembly.
- They provide a means to regulate when initiation events occur. This ensures that the DNA is replicated only once per cell division; that is, a chromosomal replication origin is used once and only once per cell cycle.

Consider the *E. coli* chromosome. The genome exists as a single large circular molecule ($1N = 4.2 \times 10^6$ bp). It constitutes a single replicon, as it contains only one origin of replication (*oriC*). Replication initiates at the origin and proceeds bidirectionally (theta mode replication). Two replication forks leave the single origin in opposite directions and proceed at about the same rate until they meet. Replication of the *E. coli* chromosome is relatively rapid: 4,200,000 bp/100,000 bp per min = 42 minutes needed to replicate the entire genome.

In contrast, eukaryotic genomes are very much larger and fork movement is considerably slower (500-4,000 bp/min vs. 50,000 bp/min in *E. coli*). To compensate, eukaryotic genomes are divided into several (linear) chromosomes, each of which is divided into a large number of relatively small replicons.

In eukaryotes, one origin of replication occurs every 10-330 kb (compare with the 4,200 kb *E. coli* genome) (Fig. 60). Replication proceeds bidirectionally from each origin, until it encounters a replication fork from an adjacent origin (or reaches the end of the DNA molecule).

Consider the mouse genome ($1N = 2 \times 10^9$ bp; $2N = 4 \times 10^9$ bp). Even though it is about a 1,000 times larger than the *E. coli* genome, the fact that it is divided into 25,000 replicons of average

size 150 kb allows for rapid duplication. Just over an hour would be required if all replicons fired at the same time.

But in fact, eukaryotic genomes typically require 6-8 hours to be replicated, because not all origins fire at once. Replicons containing active genes tend to be replicated early, while others are replicated later. (This implies additional regulation, which is not yet understood.)

Obviously, events which occur at the origin of replication are of central importance for replication and its control. There are many questions: How is replication initiated? How is initiation coordinated with the cell cycle? How is initiation restricted to once per S phase? How is the replication of thousands of origins coordinated? How do viruses escape these controls? At present, we have only partial answers to these questions, and of course we know more about prokaryotic systems than we do about eukaryotic systems. But again, despite the differences between pro- and eukaryotes, there are elements of mechanism that are common to both. Here we will focus on origin structure and how replication is initiated.

Characteristics of simple origins

What do replication origins look like? Origins of DNA replication can be divided into two general groups: simple origins such as *E. coli* oriC and those found in simple eukaryotic genomes (viruses and yeast), and those found in complex genomes of plants and metazoa (flies, frogs and mammals).

Simple origins have a modular anatomy composed of unique DNA sequence motifs, at least one of which interacts with soluble proteins (Fig. 61). They usually consist of at least four motifs/elements. The organization of these elements relative to each other is similar in most simple origins (*DePamphilis, M. (1996) Origins of DNA replication. In "DNA Replication in Eukaryotic Cells". Cold Spring Harbor Laboratory Press. pp. 45-86.*)

1. The origin recognition element (ORE).

An ORE is a DNA binding site for one or more origin recognition proteins (ORPs) that are required for initiation of DNA replication. Binding is sequence specific in simple oris, although other factors (such as chromatin structure) can influence binding. Origin recognition proteins serve at least two functions:

- They initiate DNA unwinding using either their own helicase activity (some viral proteins), by recruiting a helicase, or by distorting the helix.

- They bind other replication proteins and recruit them to the origin. Therefore, binding of the ORE by ORPs is the key step that determines where replication will begin.

2. The DNA unwinding element (DUE).

DNA unwinding begins at an easily unwound DNA sequence called the DUE. A DUE is defined by cis-acting mutations that both increase the stability of the double helix and reduce DNA replication efficiency.

DUE's are usually A/T rich but lack any other sequence specificity. Any easily unwound DNA sequences can substitute for a bona fide DUE. Therefore, it is unlikely that a specific replication protein interacts with the DUE. Rather, binding of the ORE by ORP results in unwinding at the DUE. In support of this idea, spacing between the ORE and DUE is critical for unwinding.

The DUE corresponds to the origin of bidirectional replication (OBR). This site is defined by the transition between continuous and discontinuous DNA synthesis on each template. Thus, the DUE is the entry site for the replication machinery.

3. The A/T element

This element is critical for origin function, although what it does is not entirely clear. It should not be confused with the DUE, which is usually also A/T rich. A/T elements characteristically have a "bent" structure (bent DNA), which is favored by the appearance of an unbroken stretch of A-T base pairs. DNA bending may contort the origin and facilitate protein binding to the ORE and unwinding of the DUE (or both). The ORE, the DUE and the A/T rich element are essential and comprise the origin core.

4. Auxiliary elements

Auxiliary elements are not essential, but stimulate replication 2 to 1000-fold. Aux elements typically bind transcription factors.

How do transcription factors stimulate origin activity? This is not entirely clear. But it is likely that the transcription factors recruit proteins capable of remodeling chromatin structure; thus they may help to "open" chromatin and allow replication factors (such as an ORP) to access the DNA.

Initiation- Assembly of replication forks

Replication forks are assembled at origins of replication. The components needed to build the replication fork include proteins that recognize the origin (origin recognition proteins), proteins required for assembly but not for subsequent DNA synthesis, and components of the fork replication machinery.

Initiation at oriC in E. coli.

Using plasmids containing oriC, an *in vitro* system utilizing purified proteins has been established. This has led to significant advances in our understanding of initiation.

The goal of the first step in initiation is to open up the duplex, and to recruit replication proteins (in particular, helicase). The product is a prepriming complex. The proteins required for formation of the prepriming complex include DnaA, DnaB, DnaC, and SSB. The reaction is favored by template supercoiling, which facilitates strand separation.

OriC contains 3 x 13 bp repeats (the DUE) and 4 x 9 bp repeats (R1-4; the ORE). DnaA acts as an origin recognition protein (initiator protein). It recognizes its cognate origin (oriC) in a sequence specific manner. This is the foundation for all subsequent events (Fig. 62).

DnaA is both a sequence-specific dsDNA binding protein and an ssDNA binding protein. It is also an ATPase. In the presence of ATP, DnaA-ATP first acts as a dsDNA binding protein and binds cooperatively to the ORE until a complex consisting of 10-20 DnaA monomers is formed. OriC (150-250 bp of DNA) is wrapped around the DnaA complex. This distorts the helix, which promotes melting at the DUE. This step is facilitated by supercoiling. DnaA-ATP then interacts with ssDNA at the DUE to stabilize the open conformation.

In the next step, the hexameric helicase (DnaB) is recruited and loaded onto the ssDNA. Recruiting is accomplished by DnaA which interacts with a helicase loading factor, DnaC. DnaC also interacts with DnaB in the presence of ATP. Six molecules of DnaC-ATP bind six molecules of DnaB. Two such DnaB:DnaC complexes are recruited to the origin by the ability of DnaC to bind DnaA.

The presence of ssDNA and/or interaction with DnaA triggers ATP hydrolysis by DnaC. The free DnaC then leaves DnaB loaded onto the lagging strand template. Again, two DnaB helicases are

loaded at oriC, one for each fork. At this point, the prepriming complex is formed. It consists of stable replication protein complexes assembled on the ori.

The next step in initiation is priming. The helicase, with the aid of SSB, begins to expand the replication bubble. The DnaG primase is recruited next, it forms a mobile complex with DnaB (the primosome) (Fig. 63).

After the primosome is formed, primase recognizes the template and synthesizes an RNA primer 10-20 NT long. This provides the 3'-OH terminus that is recognized by Pol III holoenzyme. The interaction between DnaB and the β subunit no doubt assists in recruiting Pol III. β clamps are loaded on the primer termini by the β complex clamp loader.

Polymerase loading is the final step in the assembly of the replisome at the replication fork. But replicative synthesis cannot proceed until the complex is further remodeled. Recent evidence indicates that this final remodeling step is ATP hydrolysis by DnaA-ATP. ATP hydrolysis is triggered by the presence of the β clamp at the primer terminus. ATP hydrolysis inactivates and releases DnaA from the origin, which allows elongation to proceed. This provides a mechanism to ensure that replication does not occur until all components are assembled. In addition, since DnaA cannot bind oriC unless it is complexed with ATP, this mechanism also prevents re-replication.

References:

Baker, TA, and Wickner, S.H. (1992) Genetics and enzymology of DNA replication in Escherichia coli. Annual Review of Genetics 26: 447-478.

Katayama, T., Kubota, T., Kurokawa, K., Crooke, E., and Sekimizu, K. (1998) The initiator function of DnaA protein is negatively regulated by the sliding clamp of the E. coli chromosomal replicase. Cell 94: 61-71.

Kurokawa, K., Nishida, S., Emoto, A., Sekimizu, K., and Katayama, T. (1999) The EMBO Journal 18: 6642-6652.

Lee, D.G., and Bell, S.P. (2000) ATPase switches controlling DNA replication initiation. Current Opinion in Cell Biology 12: 280-285.

Initiation at eukaryotic chromosomal origins

Much less is known about initiation at eukaryotic chromosomal origins. However, a general picture is beginning to emerge. Most of the recent information has come from genetic and biochemical studies using budding yeast (*Saccharomyces cerevisiae*), fission yeast

(*Schizosaccharomyces pombe*), *Drosophila melanogaster*, the African toad (*Xenopus laevis*), mouse and human cells.

In budding yeast, ARS sequences contain elements which function similarly to those of other simple origins. All ARS's contain an A element, which is characterized by a conserved, 11-bp ARS consensus sequence (an ORE). The A element binds the origin recognition complex (ORC; see below). Three B elements have also been defined. B1, which is adjacent to A, facilitates ORC binding. B2 appears to be a DUE. In the ARS1 ori, B3 is a binding site for transcription factor Abf1, and thus it probably functions as an auxiliary element.

The organization of complex metazoan oris is not clear. It is known that they bind ORC, but exactly how is not known, as they do not contain an ARS consensus sequence, or any other obviously conserved sequence. Chromatin structure also plays an important role in origin recognition.

Nevertheless, the sequence of events appears to be very similar in yeast and higher eukaryotes. We can say this because the proteins involved are highly conserved and in many cases are interchangeable between organisms.

The initiation process is tied to the cell cycle through the action of cyclin dependent kinases (CDKs) which either inhibit or stimulate complex formation by phosphorylating specific proteins, depending on cell cycle timing. The CDKs are a family of related protein kinases that require association with an activating cyclin protein for catalytic function. Cyclin expression/accumulation is regulated in a cell cycle-dependent manner. For example, G1/S-cyclins are active at the end of G1; S-cyclins are active during S phase, M-cyclins promote mitosis, and G1-cyclins allow passage through G1. The exact targets of the CDKs within the initiation complex are not yet known.

Origins are recognized by a group of proteins known as the origin recognition complex (ORC). ORC is a complex of six different proteins (ORC 1-6), which binds the ORE in the presence of ATP (Fig. 64). Several of the ORC proteins are ATPases, and the complex has both dsDNA and ssDNA binding activity. ORC-ATP binds dsDNA at the origin. Interestingly, ssDNA appears to stimulate ORC-ATPase activity in a length-dependent manner. Thus unwinding at the origin may result in the loss of ORC dsDNA binding activity. ORC appears to be associated with the origin through through G2-M-G1, and may be released in S.

In G1, Cdc6 enters the complex. Cdc6 binding to the origin requires ORC. Cdc6 is also binds ATP and is active as a Cdc6-ATP complex. Cdt1 (not shown on the Fig. 64 diagram) forms a complex with Cdc6. Current data suggests that the Cdc6:Cdt1 complex is the helicase-loader.

The Mcm proteins (Mcm 2-7) enter next. Mcm proteins interact with Cdc6 and are not bound to origins until Cdc6 is present. A subset of the Mcm complex, Mcm 4, 6 and 7 (two copies of each), is a weak hexameric helicase *in vitro*. The other Mcm proteins, particularly Mcm 2, may inhibit helicase activity. Later, at the time of initiation, the helicase could be activated by removal of Mcm 2, 3, and 5. Presumably the active Mcm helicase functions in local unwinding at the origin, and also acts to separate parental strands during replication fork progression. A complex with Mcm (helicase) at the nascent fork constitutes the pre-initiation complex.

The transition to replication occurs at the G1/S boundary and involves removal of Cdc6 and Cdt1. CDK phosphorylation of Cdc6 results in its degradation (yeast) or export from the nucleus (mammalian cells). A novel protein called geminin (found in metazoans; not shown) inhibits and probably removes Cdt1.

Once Cdc6 and Cdt1 are removed, Cdc45 protein associates with the origin. Recruitment of Cdc45 requires ORC and MCM proteins. Cdc45 also interacts with RPA (SSB), DNA Pol α :primase, and DNA Pol δ / ϵ . Cdc45 appears to be part of the replisome. So it may not only recruit the polymerase:primase, it may also tether it to the helicase (Mcm:Cdc45:Pol α : primase; a mobile primosome?).

With the arrival of the primase and polymerases, all components needed for replication are present. But additional proteins that were necessary for initiation but inhibitory to further growth of the fork must first be released. This remodeling step may be mediated in part by Cdk7-Dbf4 kinase (also known as DDK). This kinase is known to phosphorylate Mcm proteins. It is possible that this releases Mcm 2, 3, 5, which could activate the Mcm 4,6,7 helicase.

Additional unwinding generates additional ssDNA, which stimulates the ORC ATPase. ATP hydrolysis inactivates the ORP function of ORC, and allow its release from the origin.

Note that there are several steps that prevent re-replication (re-initiation):

CDK-mediated prevention of re-replication. CDK phosphorylation inactivates Cdc6 (clamp loader). This allows complexes established in G1 to transition toward replication, but prevents

formation of new pre-priming complexes through S, G2, and M. CDKs also phosphorylate ORC and MCM components, although the functional consequences are less clear.

Geminin-mediated prevention of re-replication. Geminin appears after cells enter S-phase and levels fall during mitosis. This protein inhibits Cdt1 (part of the clamp loader) and the consequences of this are similar to CDK phosphorylation of Cdc6.

ORC-mediated prevention of re-initiation. ORC binds the origin as a nucleotide complex (ORC-ATP). Unwinding generates ssDNA, which stimulates of ORC ATPase, which inactivates ORC as an ORP.

Although there are differences in detail, what is striking about this mechanism is its general similarity to initiation at *E. coli* oriC: both mechanisms employ a series of discrete steps involving rounds of recruitment and remodeling to produce an active replication fork on which primases and DNA polymerases can be loaded .

Most notably, ATP is used as a switch to separate and regulate recruitment and remodeling. (Fig. 65). In each mechanism, protein + ATP is required for an early step, and later ATP hydrolysis results in a switch in protein activity (e.g. DnaA, ORC, DnaC, Cdc6). Previously, we saw a similar sequence of events used during clamp loading by the complex and RFC.

References:

Baker, T.A., and Bell, S.P. (1998) Polymerases and the replisome: Machines within machines. Cell 92: 295-305.

Kelly, T.J., and Brown, G.W. (2000) Regulation of chromosomal replication. Annual Review of Biochemistry 69: 829-880.

Lee, D.G., and Bell, S.P. (2000) ATPase switches controlling DNA replication initiation. Current Opinion in Cell Biology 12: 280-285.

Bell, S.P. and Dutta, A. (2002) DNA replication in eukaryotic cells (2002) Annual Review of Biochemistry 71: 333-374.