6.4: Replication landscape in E. coli

Replication landscape in E. coli: Initiation at oriC, elongation and termination at ters

The origin of replication on the circular chromosome of *E. coli* illustrates to interactions of specific DNA sequences and proteins in the tightly regulated process of initiating replication. Replication in *E. coli* begins at a specific sequence called *oriC*. This is the single origin of replication on this chromosome, and DNA synthesis proceeds in both directions from it (Figure 6.7). The sequence *oriC* was identified by its ability to confer the capacity for autonomous replication on a DNA molecule. In this experiment, the origin of replication of a plasmid containing a drug-resistance marker gene was inactivated by mutation, hence making it impossible to replicate in bacteria. Random fragments of *E. coli* DNA were ligated into the mutated plasmid, and these recombinants were transformed into *E. coli*, screening for the ability of the bacterial DNA fragments to provide the ability to replicate, thereby producing a drug-resistant strain. Note that this genetic assay reveals a replicator, i.e. the DNA fragment required in cis for a DNA molecule to replicate. Further biochemical analyses showed that DNA synthesis also initiates within *oriC*, hence it is also an origin of replication. Although replicators and origins often map close to each other (and may be the same for some replication units), that is not a requirement. In some replicators, the origin is a broad zone that encompasses a more precisely defined replicator, such as the origin of replication for bacteriophage λ.
Figure 6.7. Sites for initiation and termination of replication in *E. coli*.

A. Annotated sequence of *oriC*.

The sequence is from GenBank locus ECOORI, accession J01657. The probable left and right ends of *oriC* are 128 and 377. Binding sites for DnaA (from GenBank annotation) are doubly underlined, and the 9 bp repeat within them is red. The consensus for the 9 bp repeat is TTATMCAMA (M=C or A) or its reverse complement TKTGKATAA (K=G or T). A minor DnaA binding site is underlined with a dotted line. The *Bgl*II cleavage sites are underlined. They are contained within two of the three 13 bp repeats, which are colored blue. GATC motifs are underlined with a wavy line; note that the *Bgl*II cleavage sites contain GATC.

B.
B. Aligned sequences of oriC from *E. coli* and homologs from several enteric bacteria. The alignment is from the *Menteric* server at [http://bio.cse.psu.edu](http://bio.cse.psu.edu). The 13 bp repeats are colored blue. Conserved sequences identified at the default parameters of the *Menteric* server are boxed with a black outline. Two of these are binding sites for DnaA, and these are colored red. Two other DnaA binding sites are slightly less conserved; these have a lighter shade of red and no black outline. Note that the functional DnaA binding sites are conserved in this range of bacteria. Also, some highly conserved sequences have not yet been identified as a specific binding site for a protein; these would be interesting for further study.

The minimal fragment of *E. coli* DNA active in the above assay is referred to as oriC, for the *origin of the E. coli* chromosome. It is approximately 245 bp long, and contains 3 copies of a 13-mer repeat, 4 copies of a 9-mer repeat, and 11 GATC motifs (Figure 6.8A). These features, and others, are highly conserved in the enteric bacteria such as *E. coli* and its relatives such as the *Salmonella* species, *Klebsiella* and *Vibrio cholera*. (Figure 6.8B).

The DnaA protein binds specifically to the 9-mer repeats (Figure 6.9). Temperature-sensitive mutations in the dnaA gene cause a slow-stop phenotype at the restrictive temperature, showing that the primary role for the DnaA protein is in initiation. It is the only protein known to be used only in initiation of replication, not other stages. Once the 4 copies of the 9-mer sequence are occupied by DnaA protein, many more molecules of DnaA bind cooperatively to those on the DNA, eventually leading to binding of 20-40 protein monomers in a large core. This large protein-DNA complex causes the DNA to melt at the three 13-mer repeats.
Question 6.6. The diagram for Figure 6.9 contains information that can be used to design two assays for melting of DNA at the origin. Nuclease P1 cleaves single stranded DNA, whereas the restriction endonuclease BglII cuts only duplex DNA. After DnaA has bound to DNA at oriC in the presence of ATP, how can you distinguish between the initial complex and the open complex?

The DnaB hexamer can now bind to the melted DNA at the origin. This is the same DnaB that we encountered in the primosome, and like those reactions, it is brought to DNA in a complex with 6 monomers of the DnaC protein. After the DnaB helicase is loaded on the melted DNA, it can carry out its DNA unwinding activity, using the energy of ATP hydrolysis to break apart base pairs in the DNA. Action of DnaB melts the DNA beyond the 13-mer repeats and displaces the DnaA protein complex. In the absence of polymerase, a long segment of DNA can be unwound (about 1000 bp) but in the presence of replicating polymerases, the region unwound is only about 60 bp.

This unwinding of about 60 bp allows other proteins to bind to establish the two replication forks at this bidirectional origin. SSB coats the single strands formed by melting and unwinding. DnaB and the single stranded DNA activate the Dna G primase to form the primers for the replication forks. DNA polymerase III holoenzymecan bind a begin replication from the primers. Movement of the replication forks proceeds as described in the previous chapter.

Question 6.7. (a) Assume that gyrase activity maintains a constant superhelical density while DNA is being replicated. How often will it have to act? (b) If one cycle of gyrase action requires the hydrolysis of one ATP molecule, how many ATP molecules are consumed by the unwinding and writhing for that one cycle of gyrase action?

The two replication forks launched from oriC proceed in opposite directions around the circular chromosome, synthesizing DNA at a rate of approximately 50,000 bp per min. Note that this means the DNA is untwisting at about 5,000 revolutions per min ahead of each fork. Not only are the helicases working efficiently and consuming large amounts of ATP, but gyrase is highly active, providing a critical swivel point for the replication machinery, allowing the rapid rotation required for the unwinding.

The two replication forks effectively divide the E. coli chromosome into two replicores, each containing about half the chromosome (Figure 6.10). The replicore is the chromosomal DNA synthesized by a particular replication fork. Replicore
1 is synthesized by the replication fork moving in a clockwise direction on the conventional genetic map of *E. coli*. For this replicore, the leading strand is the one running 5’ to 3’ in the same direction as the genetic map (increasing from 0 to 100 min). Replicore 2 is synthesized by the replication fork moving in a counterclockwise direction, and of course the opposite strand will be the leading strand. For both replicores, the leading strand has more G than C. Also, the trinucleotide CTG occurs more frequently on the leading than lagging strand. The leading strand is the template for lagging strand synthesis, and a CTG on the leading strand serves as a primase binding site and a primer initiation site. Hence the oligonucleotide bias on leading versus lagging strand fits with the needs for multiple priming events during discontinuous replication. The recombination hot-spot Chi is more frequent along the leading strands. Finally, most genes are transcribed in the same direction as the replication fork moves in these replicores. The full significance of some of these observations is still not clear, but they point to an overall organization of the genome with respect to replication. It will be of considerable interest to see whether these patterns are found in replicores in other organisms.

**Figure 6.10. Replicores on the *E. coli* chromosome.** The term “top” strand refers to DNA strand running 5’ to 3’ in the same direction as the genetic map (increasing from 0 to 100 min. This is the same strand as is listed in the standard *E. coli*K12 sequence in the databases.

The two replication forks meet on the side of the chromosome opposite oriC. **Termination** occurs in a zone where the forks meet (Figure 6.7). It is restricted to this zone by the action of the Tus protein at ter sequences. The tersequences block further progression of the replication fork, with a clear polarity. The sequences terD and terA block the progress of the counter-clockwise fork (Fork 1 in Figure 6.7) but allow clockwise replication (Fork 2) to proceed through. In contrast, terC and terB block the progress of the clockwise fork (Fork 2 in Figure 6.7) but allow counter-clockwise replication (Fork 1) to proceed through. The tersequences are 23 bp and are binding sites for the Tus protein, the product of the tus gene (“ter utilization substance”), which is required for termination. It prevents further helicase action from the replication fork.

**Resolution** of the replicated chromosomes occurs when the two replication forks meet. Since these are moving in opposite directions, the distribution of ter sites roughly opposite to the ori insures that the two replication forks will meet in the zone between the oppositely oriented tersites.

One scenario is illustrated in Figure 6.11. Let Fork 1, moving in a counter-clockwise direction, proceed as far as it can,
i.e. to the terD, terA sites. Fork 2, moving in a clockwise direction, can proceed past these ter sites, and will it will meet Fork 1. The two sets of products from each replication fork are then joined. The leading strand synthesized from Fork 2 joins the lagging strand synthesized from Fork 1. Likewise, the lagging strand from Fork 2 joins the leading strand from Fork 1.

**Figure 6.11. Resolution of replication forks in the termination zone.** The abbreviations are cw=clockwise and ccw=counter-clockwise, lead=leading strand, lag=lagging strand.

**Question 6.8.** Use the model for lagging strand synthesis to explain how the leading strand is joined to the lagging strand when the replication forks meet and resolve.

**Contributors**

- Ross C. Hardison, T. Ming Chu Professor of Biochemistry and Molecular Biology (The Pennsylvania State University)