Chapter 4. DNA replication in detail

The (re)production of a complete chromosome is one of the most complex processes taking place in a living bacterial cell, in which many proteins participate. The activities of DNA polymerases I and III have been briefly introduced in the previous chapter, in which it was already explained that DNA replication involves more than just knitting nucleotides together to form a double strand of DNA. This chapter describes the multiple proteins involved in the process. DNA replication has been studied in *E. coli* for more than a half-century, and because much is known about this system, the nomenclature from *E. coli* is used here. When other bacterial species significantly differ in some of these steps this will be briefly mentioned, without going into details.

4.1. Initiation of replication

The first step: opening the origin of replication

Replication of chromosomes is initiated at the origin of replication, called oriC in *E. coli*. This region can be between 180 and 3000 bp long, depending on the species; in *E. coli* the minimally required oriC region is 245 bp long, and devoid of genes. The first protein required to melt the double-stranded chromosome at oriC is DnaA, a member of the ATPase family of proteins called AAA+. The gene *dnaA* is usually located close to ori (but not always, for instance in *Coxiella burnetti* or *Caulobacter crescentus* it is not) and is conserved in almost all bacterial genomes (with the exception of some endosymbiontic bacteria containing extremely reduced genomes). The protein DnaA forms complexes with itself and also with DNA: it binds to specific short repeated AT-rich sequences, called DnaA-boxes, which are found within the oriC region. Multiple copies of DnaA bind to these repeats in an orderly manner: the 5 DnaA-binding repeats found in *E. coli* are slight variations of the consensus sequence TTATNCACA (the exact DNA sequence can also vary in different bacterial species), and hence have a different affinity for the DnaA protein to bind. The sites located on the edges of the assembly region have the highest affinity, so that the order of loading is defined (the strongest binding sites are always occupied, so that oriC in the cell is always bound to some DnaA). The protein molecules also interact with each other, resulting in a spiraling multimer of proteins around which the DNA is wrapped, as shown in Figure 4.1.

The general conservation of oriC sequences within all eubacteria, and the conservation of DnaA suggest that this mechanism of replication initiation is generally conserved in the bacterial world (archaea have their own type of DnaA-like protein, called Orc1, which acts in tandem with a protein called Cdc6 and in this respect more resemble eukaryotes). DnaA needs to be loaded with ATP before it can bind to the DnaA-boxes. The protein can also bind ADP, but the ADP-DnaA complex has a lower affinity for the DnaA-boxes, so the weaker binding sites are less likely to be occupied. This means that replication can only initiate if sufficient energy is available in the form of ATP, so that the ATP-DnaA/ADP-DnaA ratio is high enough. Abundant histone-like DNA-binding proteins, such as IHF and Fis, also play a role in regulating initiation of replication. The complex of specific proteins binding to the oriC region that initiates replication is called the orisome.

Wrapping the DNA around the spiraling DnaA multimer destabilizes the helix. As already discussed in Chapter 2, introducing a wrap, or loop in the DNA helix can result in local torsional stress (negative supercoiling) which forces the two strands to separate: the torsional stress can provide sufficient energy to overcome the stabilization of the helix due to stacking and H-bonding, and a local area becomes single-strand (an "open bubble" is formed). Apart from the DnaA binding sites, the oriC of *E. coli* contains a second repeat region called the duplex unwinding element (DUE), where three copies of another AT-rich repeat (consensus
sequence GATCTNTTNTTTT) are located, and once these become unwound, DnaA-ATP specifically binds to the T-rich single-strand.

![Figure 4.1. Opening of the origin of replication by DnaA.](image)

**Figure 4.1. Opening of the origin of replication by DnaA.** At the top, the ori region is schematically drawn with three 13-mer repeats of the duplex unwinding element (DUE) and five DnaA-boxes (marked '9' as the consensus is 9 nt). The protein DnaA has highest affinity to the two outermost boxes. The binding sites for IHF and Fis are indicated. The asterisks show the presence of Dam methylation sites. DnaA protein can oligomerize due to protein-protein interaction of an arginine finger. Binding of DnaA multimers induces a curl in the DNA, which melts the DUE. This structure is stabilized by binding of a DnaB hexamer to the single stranded DNA, for which DnaC is required.

The opening of DUE is again dependent on high intracellular ATP levels. A helicase protein, DnaB, quickly stabilizes the other strand. DnaB is a hexamer (formed by six identical protein subunits) producing a ring that encircles single-stranded DNA (the lagging strand, specifically), but in order to do so it needs six molecules of DnaC, to load the ring onto the DNA. Once the DnaB hexameric rings are in place (one for each lagging strand), DnaC is released in an ATP-dependent manner. Together, DnaB and DnaA prevent the two strands from re-hybridizing, and allow primase to produce the first RNA primers (see below). The loading of DnaB onto the open DUE sequences can only take place when DnaA is bound to its recognition sites, and direct interaction between DnaB and DnaA has been demonstrated.

A further protein is needed to stabilize the unfavorable open helix: single strand binding protein, SSB. The protein consists of four subunits (it is a tetramer) and binds to single-strand DNA, as its name suggests. Whereas only one copy of the DnaB hexamer will be located at each replication fork, multiple copies of SSB remain dispersed over the single-strand DNA region, to help keep it open, and to protect it from degradation. This is shown in **Figure 4.2.** When the replisome slides along, DnaB (helicase) and the chi subunit of Pol III (see below) will push SSB aside.

![Figure 4.2. SSB protein stabilizes single-strand DNA.](image)

**Figure 4.2. SSB protein stabilizes single-strand DNA.** The two DNA strands that are separated by helicase (DnaB) are stabilized by tetramers of single strand DNA binding protein (SSB).
Initiation of replication is tightly regulated

Once a first replication round has started, initiation of replication must be repressed, to prevent too many replication cycles occurring simultaneously. In Enterobacteriaceae, both DNA methylation and regulatory inactivity of DnaA (abbreviated as RIDA) are key features in this repression. The newly-produced DNA strands (one for each replichore) are not methylated, but the older template strand is, so the double-strand replichores consist of one methylated and one non-methylated DNA, called hemimethylated DNA. This hemimethylated DNA is particularly prominent at oriC, which contains many methylation sites, such as GATC Dam sites, in E. coli (marked in Figure 4.1). Hemimethylated DNA is recognized and bound by protein SeqA (so called because it sequesters oriC), and this binding prevents DnaA from starting a new replication cycle. In addition to hemimethylated DNA, other mechanisms are in place to limit replication initiation. Free DnaA molecules are out-titrated by DNA-boxes located elsewhere on the chromosome (as part of the datA locus, shown for E. coli and Streptococcus coelicolor), thus lowering the concentration of free ATP-DnaA, and preventing a new initiation round. Finally, as already stated, the ratio between ATP-DnaA and ADP-DnaA further determines whether a new replication cycle can start. RIDA (regulatory inhibition of DnaA activity) regulates this ratio. The sliding clamp of the replisome (see below) that acts at both replication forks actively reduces this ratio as it hydrolyses DnaA-bound ATP to DnaA-bound ADP. As a result, as long as the replication forks are still close to the origin, there is negative feedback to prevent novel initiation. Moreover, DnaA of all bacteria known to date contain an ATPase domain, so that excess ATP-DnaA can be auto-hydrolyzed to ADP-DnaA. Finally, a protein called Hda (it is homologous to DnaA) specifically hydrolyzes DnaA-bound ATP. RIDA may be the main regulatory mechanism to down-regulate replication initiation for organisms lacking DNA methylation.

Multiple initiation rounds can exist simultaneously

The maximum speed of DNA pol III would only allow a 4.5 Mb chromosome to be produced every 75 min or so, whereas many bacteria, including E. coli, multiply faster than that. E. coli growing exponentially in rich media can double every 20 minutes. E. coli multiplies at a moderate speed. The fastest doubling time for bacteria is less than 7 minutes. (As an aside, bacterial growth can also be extremely slow, and depends both on the species and on the growth conditions applied). How can a bacterium double faster than the time it takes to make a copy of its chromosome? This is possible because in fast-growing cells the next round of replication is initiated before termination is completed. It is estimated that up to six cycles of replication can co-exist in a single E. coli cell growing under optimal conditions; faster growing species may allow even more rounds of replication initiation simultaneously. As a result, a dividing cell receives more than a complete copy of a chromosome, as the region around the origin has already been duplicated by the time the DNA is redistributed over the two dividing cells. The protein DiaA, a recently discovered E. coli DNA binding protein, is thought to coordinate synchronous initiation in not-yet-finished replichores. DiaA forms tetramers that each can bind multiple ATP-DnaA (or ADP-DnaA) copies. It is believed that by bringing ATP-DnaA molecules together, the overall affinity for the DNA binding sites is increased by DiaA, thus stimulating replication once sufficient ATP-DnaA is present in the cell. DiaA homologs are found in other bacteria.

The picture is still not complete, as more proteins are involved in regulation of the orisome formation. Once replication has initiated, it will continue and cannot be halted, unless the cell divides or dies. Thus, initiation is carefully regulated, and numerous proteins are involved in this, such as HU, Dpi, IciA, Cnu, Hha, Rob, and ArcA, but their specific roles will not be
Box 4.1 summarizes the most important proteins involved in initiation of replication.

**Information box 4.1. The main proteins involved in initiation of replication**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DnaA</td>
<td>Binds to DnaA-boxes within oriC, initiating strand separation. Conserved in most bacteria (not found in some endosymbionts)</td>
</tr>
<tr>
<td>DnaB</td>
<td>Helicase, stabilizes melted, ssDNA at the DUE site of oriC, hexamer.</td>
</tr>
<tr>
<td>DnaC</td>
<td>Loads DnaB onto the singlestranded DNA</td>
</tr>
<tr>
<td>seqA</td>
<td>Binds to hemimethylated Dam methylation sites in oriC to prevent re-initiation</td>
</tr>
<tr>
<td>DiaA</td>
<td>Binds DnaA to assist new initiation even when growing replichores are not yet complete</td>
</tr>
<tr>
<td>IHF, FIS</td>
<td>Histone-like proteins that participate in regulation of initiation</td>
</tr>
<tr>
<td>SSB</td>
<td>Single-strand binding protein protects any single-strand DNA from degradation, tetramer.</td>
</tr>
<tr>
<td>CtrA (in <em>C. crescentis</em>)</td>
<td>Blocks replication in swarmer cells</td>
</tr>
<tr>
<td>Spo0A (in <em>B. subtilis</em>)</td>
<td>Involved in blockage of replication in spores</td>
</tr>
</tbody>
</table>

**Initiation of replication in various bacterial species**

Initiation of replication has only been studied in a limited number of organisms. The origin of replication can now accurately be predicted from a genome sequence, mainly based on the skew of short sequences (up to 8 nt long) biased towards only one of the two replichores. These predictions have been confirmed by experiments for a number of organisms, so that they are now considered reliable without further experimental confirmation. Nevertheless, it is of interest to study initiation of replication in organisms that differ from *E. coli*, such as those that grow more rapidly. One fast multiplying species is *Bacillus subtilis*, which can divide in seven minutes under optimal growth conditions. Its DnaA-boxes have recently been characterized. It lacks sequence homologs of IHF, Hha, SeqA or Dam, but alternative protein candidates (functional homologs) have been identified that may play similar roles in regulating replication initiation.

In some instances, replication must be switched off on a long-term basis. *Caulobacter crescentis* divides into two distinct cells at each cell division: a stalked cell and a swarmer cell. The latter cannot divide, and won't replicate their chromosome, until they have differentiated into stalked cells. Replication is inhibited in swarmer cells by the action of CtrA, which binds to specific sequences within Cori (as its ori is known), blocking DnaA binding. A protease (ClpXP) degrades CtrA during the transition to stalker cells, after which replication can start. The activity of CtrA is tightly regulated in which the response regulator DivK plays a key role. Likewise, spores of *Bacillus subtilis* have to inhibit their replication, and Spo0A, the key regulatory protein to initiate sporulation, most likely binds to '0A boxes' within oriC, to block initiation of replication in spores.

It has not yet been determined how endosymbiotic bacteria that lack DnaA, such as *Wigglesworthia glossinidia* or *Blochmannia* species (endosymbionts to insects) initiate replication, but they might employ recA or recBCD genes, which in *E. coli* is used to re-establish replication at damaged sites.

**4.2. Priming and strand elongation**

*RNA Primer production and the start of DNA synthesis*
As was mentioned the previous chapter, DNA polymerase cannot start producing DNA unless it has a primer which it can extend. An RNA primer, usually approximately 10 nucleotides long, is produced in both directions at ori by DnaG primase (an RNA polymerase). This primase is inhibited by ppGpp (guanine tetraphosphate), that serves as a stress regulator and senses the energy available in the cell; ppGpp can halt replication via primase inhibition during conditions of low nutrition. Bacterial primases are single-unit proteins, though in eukaryotes and archaea most primases are dimers. The enzyme contains a single-strand DNA binding site, and binds at least two nucleotides at a time, which it will connect to form a short piece of RNA.

For the leading strand, only one primer will be produced after which DNA polymerase takes over, but for the lagging strand, primers are needed every 1000 to 2000 nucleotides for the production of Okazaki fragments. At the origin of replication, DnaG primase starts its first primer production once DnaB helicase has encircled the single-stranded DNA. Binding of primase to DnaB drastically increases (>1000 times) its enzymatic activity. This makes for a clever strategy: the enzyme that can start producing primers all by itself requires DnaB to properly function, and DnaB can only be loaded on a single-stranded DNA at the origin of replication. As we have seen, the opening of the oriC region is tightly regulated, so that the activity of primase is kept in check at the oriC, while it can efficiently start for every needed Okazaki fragment once DnaB comes along, which moves on with the replication fork.

Primases don't start at random, but prefer a combination of three nucleotides, which still is a rather frequent recognition signal. *E. coli* primase recognizes 3'-GTC, and with a GC content of 50%, one can expect approximately 70,000 copies of the recognition triplet on a chromosome by chance. However, DNA sequences are not random, and the actual frequency of the primase recognition triplet is less. The primase of *S. aureus* recognizes 3'-ATT and 3'-ATC. Most primases require a pyrimidine at the first position (primase of *Aquilifex aeolicus* is an exception to this rule), and produce a primer starting with ppp-Pu, shorthand for a primer with a purine at the 5'-triphosphate end.

Interestingly, the primase enzyme can count, in that it won't produce primers longer than a single turn of the DNA helix, or 10 nucleotides, on average. How does the enzyme do this? Helicase (DnaB), to which primase binds, is a hexamer and binds three molecules of primase. It is quite likely that only one of these produces the primer, while the other two spatially hinder a production longer than 10 nucleotides by their sheer presence. A Zn²⁺-binding domain of primase is essential for this interaction; deleting that part of the molecule destroys its counting ability. On the other hand, archaeal primases don't count, and can continue to produce primers of over 1000 nucleotides. (Eukaryotic primases count via an alternative mechanism.) Viruses infecting bacteria (bacteriophages) frequently encode their own primases, and these can produce shorter primers. The ability to count also allows the protein to hand over the primer to DNA polymerase. For this step, the β subunit of Pol III is essential, which is discussed in the next section.

**DNA polymerases: more than one protein**

Once a primer is in place, DNA can then be synthesized, by extending it with dNTPs to produce complimentary DNA (rather than RNA that primase has so far produced), for which DNA polymerase is responsible. A bacterial cell contains various DNA polymerase enzymes. Besides Pol I and Pol III, introduced in section 1.6, *E. coli* contains three more types of DNA polymerase. Even Pol I and Pol III are not single enzymes, but act as heterologous multimers of proteins or subunits. The most complex structure is Pol III, also known as the replisome, and since it is responsible for the majority of DNA production in a dividing cell, this will be treated first, though it was not the first DNA polymerase that was discovered (that honor goes to DNA polymerase I, the proteins were numbered in order of their discovery). Pol III belongs
to the C family of DNA polymerases. It is the largest of all DNA polymerases and it is also the most rapidly working enzyme: it can add up to 1000 nucleotides per second to a growing DNA strand. This enzyme is quite abundant in rapidly growing bacterial cells, and produces a prominent protein band of 150 kDa on SDS page gels of *E. coli* cell lysates.

DNA polymerase III makes the leading and lagging strand of the chromosome simultaneously, but the production of the lagging strand is more complex than that of the leading strand, because it requires re-initiation of replication for every Okazaki fragment to be formed (this was introduced in section 1.6). *In vitro* studies using *E. coli* protein extracts have suggested that the replisome consists of two identical Pol III cores (working on both strands simultaneously) plus additional proteins (some of which are called subunits) that are non-covalently attached to form one big replisome protein complex. The two identical cores are again built up of individual subunits: an alpha (α), epsilon (ε) and theta (θ) subunit each per core. The α subunit is the protein that performs the polymerization step to add nucleotides, one at a time, at the 3'-OH of the growing strand (which can be the RNA primer or growing DNA). The subunit has no sequence or structural similarity to polymerases of the other families, though the protein is strongly conserved between bacteria. The subunit has the shape of a cupped right hand, as was deduced from its crystal structure. It is believed that its 'fingers' check the smooth minor groove of the produced double-strand helix, to detect any aberrations due to wrong base-pairing. In case a wrong base-pair was produced, the ε subunit, which has exonuclease activity, removes the faulty nucleotides in the 3' to 5' direction. The θ subunit may regulate the activity of ε, but its function has not been determined with certainty. Together, θ and ε have an editing, or proofreading function: when α has made a mistake, and put in the wrong nucleotide, that error will be removed and α will replace it with the right one.

Each core (α, ε and θ) is attached to a tau (τ) subunit, and two τ subunits together form a bridge to connect the two cores: τ is responsible for the dimerization of Pol III (but see below). The two τ subunits connect the two cores also with a single gamma (γ) complex, the part of the replisome that uses ATP in order to function. Two cores and a complete gamma complex together constitutes a DNA polymerase III holoenzyme. The γ complex is built of three γ subunits, which bind and hydrolyze ATP, a delta (δ) and a δ' subunit, complemented with a chi (χ) and a psi (ψ) subunit. The γ complex, also called clamp loader, functions like a pair of tongs: it can open and close a circular protein called the β clamp. The δ subunit is the 'wrench' that wedges open a clamp ring, see below, while δ' may provide stability to the clamp loader. The χ subunit removes SSB from the single-stranded DNA when it is no longer needed and assists in the handover of the primer from primase to the α subunit. The protein complex is schematically shown in **Figure 4.3**.

**Figure 4.3. DNA polymerase.** Two polymerase core complexes are connected by a tau subunit to a single gamma complex, also called the clamp loader. Two DNA strands are shown, each with a tetrameric SSB bound to it. The various subunits of the DNA pol core are not specified. The β clamp is a dimer ring put around the ssDNA by the clamp loader, which itself consists of different subunits that are not distinguished for simplicity.
The clamp loader plays a key role in DNA production

The DNA pol holoenzyme is still not the complete structure responsible for DNA polymerization. The holoenzyme would easily dissociate from its DNA template, unless it is kept in place by the β subunit, which consists of a dimer that forms a ring around the newly formed DNA that functions as a clamp. The γ complex, by means of δ', loads the β ring onto the double-strand DNA (for which it uses ATP), which is why it is also known as the clamp loader. The β clamp then slides along the DNA, holding the replisome in place like a sliding tether. Without the β clamp, the DNA pol holoenzyme is slow, connecting less than 20 nucleotides per second and connecting at most 10 nucleotides before falling off its template; with the β clamp, the speed increases to 500 - 1000 nucleotides per second, continuing for a complete leading strand (provided the template is damage-free). The rate is actually slowed by the lagging strand, which was observed when in vitro systems were designed to study leading strand synthesis in absence of a lagging strand: the replisome could move 25% faster in that case.

The stoichiometry of the replisome may be different in vivo than these in vitro findings suggest. Recently, it was found that an active replisome inside the cell combines three instead of two Pol III cores, connected by three τ subunits. While one Pol III core continuously produces the leading strand, two cores are responsible for synthesis of the lagging strand, which would make the switch from one Okasaki fragment to the next more efficient. A schematic drawing of the complete process is presented in Figure 4.4.

Figure 4.4. The replisome is shown in four panels, with increasing complexity. Panel A shows how the lagging strand loops out, so that both strands can be produced from the replication fork, despite their production being performed in opposite directions. In panel B, the components of only half a replisome are shown, with helicase separating the two strands
of the template by encircling the lagging strand, and SSB stabilizing the ssDNA. In panel C, primase is putting an RNA primer (red, numbered '2') in place for replication of the lagging strand. In panel D, the second DNA Pol core is shown that is responsible for production of the lagging strand. This continues until it hits the previous Okazaki fragment. Notice the direction of production is from a newly produced primer (2) towards an already extended primer (1). A new start will then be provided by a novel primer produced by primase (that primer is not shown).

An obvious question is how the clamp loader can load a circular β protein on a DNA molecule. It doesn't act on a complete single-stranded molecule, and it requires a 5'-OH (of the last nucleotide of the primer). The clamp loader binds to and opens a β dimer (the ring-shaped sliding clamp) and selectively puts it around the DNA at a primer-template junction; the χ subunit interacts with primase and removes any SSB, ensuring that the β ring is put in the correct position. The loaded β ring then slides along the DNA-RNA hybrid of the primed DNA (it would fall off a short linear molecule). Now the α subunit of the replisome finds the primer end and starts Its job.

Separation of the two template strands is continuously carried out by DnaB, the helicase that slides along the lagging strand and forms the 'front end' of the replisome, after which SSB will keep the strands separated until removed by χ. DnaB, together with a protein called Rep, pushes other proteins aside that may be bound to the DNA, such as the histone-like proteins, or DNA-binding proteins that function in gene expression. The α-subunit that creates the leading strand can start once it has been handed over the primer that primase (DnaG) had produced at the origin of replication.

Formation of the lagging strand

The most complex part of DNA replication is to understand how the individual Okazaki fragments of the lagging strand are produced. While the leading strand of a chromosome just slides along, the lagging strand is produced in the opposite direction, which produces a growing loop of DNA. The loop consists of single strand template (stabilized by SSB) upstream of the DNA pol core, that is responsible for the lagging strand, and of newly produced double-strand DNA downstream of it, as long as the next primer has not yet been reached. This loop maximally spans the length of a complete Okazaki fragment, which is typically between 1.5 and 2 kb. The DNA pol core will let go of the loop when it has reached the next primer, at which stage the clamp loader will put a new beta clamp on, and the whole process repeats itself, with a loop growing again. (The old beta clamp presumably remains on the produced DNA and may be removed by separate clamp loaders, which can both load and unload beta clamps).

Production of the lagging strand is completed by two more proteins: DNA polymerase I replaces the RNA primer with DNA nucleotides, and ligase connects the ribose-phosphate backbone of the fragments. Pol I (also called PolA) was the first DNA polymerase to be discovered. It is a protein with two domains: the Klenow domain that combines 5'-to-3' polymerase activity with 3'-to-5' exonuclease activity (eating away the RNA primers of the Okazaki fragments and filling in the gaps). A smaller domain performs 5' to 3' exonuclease activity, as Pol I is also involved in DNA repair. Finally, the new DNA sequences are attached to the rest of the DNA with the ligase enzyme. Bacterial ligase is special in that it receives energy from NAD+ instead of ATP; ligases in eukaryotes all depend on ATP. The main proteins active in a replication fork are summarized in Box 4.2.

Information box 4.2. The main proteins and subunits acting at the replication fork

The replisome: DNA polymerase III complex
α subunit  Responsible for DNA polymerization
β subunit  Sliding clamp (also called DnaN)
γ subunit  Part of the clamp loading gamma complex
δ, δ’  Parts of the clamp loading gamma complex
ε subunit  (DnaQ) has 5’ to 3’ exonuclease activity
θ subunit  Probably regulates epsilon activity, together responsible for proofreading
τ subunit  Dimerization (or trimerization) of two (or three) pol III cores
γ and ψ  Part of the clamp loading gamma complex

Proteins assisting at replication:
- DnaG  Primase, an RNA polymerase that produces the primers for Okazaki fragments
- DnaB  Helicase, binds to the lagging strand and separates the DNA duplex
- SSB  Single strand binding protein
- DNA Pol I  Removes the RNA primers and fills the gaps in between Okazaki fragments
- Ligase  Connects Okazaki fragments
topoisomerase, gyrase  Correct negative and positive supercoiling ahead and following the replication fork, untangle catenonic DNA

4.3. Termination of replication

Termination of chromosomal replication is far easier than initiation. In some bacteria, the two replisomes just meet somewhere and stop; in *E. coli*, specific sequences restrict this event to the terC region. A replisome shooting through the terminus region and continuing in the wrong direction will be stopped by a polar (direction-dependent) fork arrest. The sequence motifs responsible for this polar block are called ter, to which a protein called Tus specifically binds. Such a ter-Tus complex blocks passage of a replisome in one direction, but allows it to pass from the other direction; as a result Tus prevents a replisome from shooting through the terminal region. A ter-Tus complex has two sides: when a replisome approaches from the permissive site, Tus will give way of the DNA and the replisome continues. A replisome approaching the complex from the non-permissive site, however, results in Tus binding more tightly to its binding site. The increased affinity of Tus in such a case is actually due to the local melting of the replication fork, as it can be mimicked by DNA manipulation, when single-strand DNA is produced at the non-permissive flank of ter due to introduced mismatches. One particular (strictly conserved) cytosine base within the ter sequence will flip out of the helix and bind to Tus when single-stranded, and as a result the protein is fixed to the DNA, until a replisome from the permissive direction comes along; this has a destabilizing effect on the fixed complex (again, as a result of DNA melting, but this time from the permissive flank) and Tus will let go; thus replication continues in the correct direction.

Supercoiling and catenation hamper strand separation

After replication of the terminus region and all gaps of both strands have been filled in, the two molecules are likely to be intermingled into catenanes (like two segments of a chain), and have to be actively separated. As was presented in Chapter 2, the topoisomerase enzymes are responsible for this.

How important are topoisomerases in the replication process? Consider a double-strand template of a chromosome of 4.5 Mb; this contains approximately 400,000 crosses of one strand over another, once for every 10.5 basepairs (the number varies slightly due to supercoiling). Obviously, these crosses have to be removed completely, or the two sister chromosomes will remain tangled. The opened replication fork causes local torsional stress.
(positive supercoiling) ahead of the replisome, which is released by gyrase. This greatly unwinds the two strands, and thus the two sister molecules-in-the-making. Although topoisomerases can't 'choose' a direction and can both link and unlink two DNA strands, their activity is driven by mechanical, torsional stress, and by releasing this stress they actually separate the two sister chromosomes by and large. Moreover, supercoils formed in the newly produced DNA (by the action of gyrase) actually protect it from becoming intermingled again. The chance that two DNA segments cross each other is far greater for long stretches of relaxed DNA than it is for tightly compacted supercoiled structures, so that supercoiling reduces the chance that topoisomerase (accidentally) links the two molecules. Experimental evidence has indeed confirmed that supercoiling protects against the formation of catenated structures.

While the replicons near completion, the DNA translocator FtcC pumps one of the two sister chromosomes through the closing septum. The protein consists of several domains and forms a hexameric ring (not unlike helicase does) with one of its domains; another domain is attached to the membrane. The protein pumps DNA out in one direction only, due to sequences on the DNA called dif (for 'defined sites') in the terminus region.

The complete untangling of two sister chromosomes to the very last link, remains somewhat enigmatic, though it is known that topoisomerase IV is mostly responsible. Active topoisomerase IV is a tetramer built of two different subunits, called ParC and ParE. ParC is co-localized with the replisome but ParE is normally dispersed in the cell. At the final stage of replication, ParC approaches the translocator FtsK and the two proteins interact with each other. This interaction possibly brings ParC and ParE together, so that active Topoisomerase IV is mostly found in the final stage of replication, where it can separate the two replicons completely.

Margin box: replication proteins come in varieties
As explained in the main text, replication of a bacterial genome requires many different proteins, each of which has a specific task. Nevertheless, few of these proteins are found in every bacterial genome, which means that they can be missed or replaced by alternative proteins in individual species. Although replication is one of the most fundamental processes in a living cell, few of the components required are truly conserved in the bacterial world. Replication of viral DNA has been shown to depend on a simplified replisome. For instance, bacteriophage T4, which infects Proteobacteria, uses a replisome of only 10 proteins. The so-called replicase is composed of DNA polymerase, a clamp, a clamp loader (made of two proteins), and ssDNA binding protein. In addition helicase, primase, and helicase loading protein are needed. An RnaseH removes the RNA primers and T4 DNA ligase repairs the Okazaki fragments. These 10 proteins are enough to replicate the viral DNA that contains 288 genes in total.

Alternative DNA polymerases
Not even the key protein of DNA replication, DNA polymerase III, is strictly conserved in all species. In Gram-positive Firmicutes, the leading and lagging strands are produced by two slightly different DNA polymerases: the leading strand is produced by a subunit called PolC (coded by polC). In this polymerase, the proofreading domain is integrated in a single subunit, so it lacks the θ and ε subunits. The lagging strand depends on PolC as well as on a DnaE homolog: the RNA primers of the lagging strand are first extended with dNTPs by DnaE before PolC takes over. Other bacterial phyla also employ both DnaE and PolC. For instance, *Thermotoga maritima* (of the phylum Thermotogae), *Fusobacterium nucleatum* (Fusobacteria) and *Aquifex aeolicus* (Aquificae) were shown to contain both DnaE and PolC. The presence of these two enzymes happens to coincide with the direction of the AT-skew. Bacteria with both DnaE and PolC contain chromosomes with A's being over-represented on the leading strand, while species lacking PolC (the Actinobacteria, Proteobacteria,
Chlamydiae, Bacteroidetes and Spirochetes) tend to have T's on the leading strand. All these bacteria also have a GC-skew with a conserved direction (G's on the leading strand). This explains the differences in GC-skew presented in the previous chapter (Figure 3.3). In the case of the Firmicute *Thermoanaerobacter tengcongensis*, which contains *polC*, both the G's and A's are on the same (leading) strand, whilst in the *Vibrio fisheri* genome (Proteobacteria, and no PolC proofreading), the G's and A's are on opposite strands. The GC-skew is usually stronger than AT-skew, but sometimes both are very weak or absent, as was shown in the previous chapter, in Figure 3.3.

### Margin box. DNA polymerase nomenclature

The five known bacterial DNA polymerases were numbered according to their order of discovery. DNA polymerase I (also called PolA) was the first discovered polymerase and is responsible for replacing the RNA primers of a lagging strand with DNA. The enzyme is also known as Klenow fragment, and has been used in many genetic engineering experiments. DNA pol II (alternatively called PolB and belonging to the B family of DNA polymerases) is responsible for replication-restart following UV DNA damage; it can spring into action within minutes following UV irradiation. DNA pol III is the main enzyme responsible for DNA production during replication. It produces a leading and a lagging strand simultaneously. In *E. coli*, this is called ‘polC’, but this is different than the PolC proof-reading subunit found in Gram positives, mentioned in the text. Two more polymerases are produced to repair DNA damage and these belong to the Y family of polymerases. Pol IV (also called DinB) is a translation (TLS) DNA polymerase that specifically repairs N<sup>2</sup>-deoxyguanosine, a nucleotide derivative that results from mutagens such as nitrofurazone. Finally, PolV (coded by the *umuDC* operon) is specialized to repair abasic sites and thymine-thymine dimers that frequently result from UV radiation. Its production is induced by DNA damage and thus it can only start working after it has been produced, some 50 min after irradiation.

### 4.4. When things go wrong

Replication is a complicated task, and can easily go wrong. A cell with a halted replisome would die, unless the machinery can be made to continue; a number of alternatives are available. The lagging strand causes fewest problems: the repeated initiation by primase and reloading of β rings ensures that the process can restart when a problem has occurred. A halt in production of the leading strand is serious, however, because this is ideally supposed to finish its job in one go (in reality, its can usually process about 85,000 bp on average per binding event, though it may continue for 300 kb when all goes well). Stalled replication forks are mostly caused by lesions in the leading strand. Several mechanisms have evolved to facilitate the restart of inactivated replication forks: reactivation downstream from a leading strand block or break, translesion synthesis, and recombinational repair.

**Restoring stalled replication forks by downstream reactivation**

The first mechanism to release a stalled replication fork discussed here is reactivation downstream of a blockage or break. It makes use of the ability of DNA pol to dissociate from its β clamp and re-associate to another beta molecule, which it has to do repeatedly for production of the lagging strand. It turns out that DNA pol frequently dissociates from its β-ring during leading strand production as well, when an obstacle is in its way. A strand break in the template, for instance, stops pol III, but it is recognized by primase (DnaG), which will produce an RNA primer downstream of the break. A second β-ring clamp can now be attached by the clamp-loader downstream of the obstacle, to which DNA pol will 'hop' in order to pass the problematic site. A single-strand gap remains in the newly-synthesized
strand around the strand break, to be repaired by alternative DNA polymerases and ligase. Another blockage for the replisome is a gene that is transcribed from the template strand at the replication fork; in that case RNA polymerase is in the way of DNA polymerase. This is easy to resolve: the RNA that is in the making is recognized by the clamp loader as a primer, so it will attach a beta clamp behind it, after which DNA polymerase can continue.

Primase frequently makes mistakes and builds in a wrong nucleotide, which is troublesome when this remains the last nucleotide of a primer. In that case, pol III can't continue, and an alternative DNA polymerase enzyme takes over: pol II (also called PolB, a member of the family B polymerases). It is a high fidelity enzyme with 3’ to 5’ exonuclease activity so it will remove the faulty RNA nucleotide. This DNA polymerase is much slower than pol III, so that it is only used for problematic sites, and not for the lump production of DNA.

Whereas polA and polC (the genes coding for Pol I and Pol III, respectively) are constitutively expressed, the genes for Pol II (polB), Pol IV (dinB) and Pol V (umuDC) are regulated by the SOS-regulon and are produced in response to DNA damage. These so-called trans-lesion synthesis DNA polymerases (TLS polymerases) are specialized enzymes to overcome lesions in the template DNA that would permanently halt DNA pol III. They have a lower fidelity and frequently introduce mutations at the site of the lesion, but that is a small price to pay for continuation of replication. It may even have a beneficial selective advantage: DNA is most frequently under attack from environmental stresses, and in such conditions TLS polymerases are most active, while the mutations they introduce may help the bacterial population to adapt to these damaging conditions. Pol IV, coded by dinB, and its role in stress-related mutagenesis has been extensively studied. Nevertheless, its primary role is most likely to keep the replisome going. Both Pol IV and Pol V belong to the Y-family of DNA polymerases and lack proofreading 3’-to-5’ exonuclease activity.

Overcoming strand breaks

Strand breaks are not the only difficulty for Pol III. Other lesions of the DNA duplex that frequently occur are pyrimidine dimers, deaminated bases, or abasic sites. These were introduced in Chapter 2, which also briefly mentioned the repair mechanisms to overcome these obstacles of replication. Deamination, the removal of the amine group (-NH2) from a base, produces an alternative base that can no longer participate in proper base pairing. A nucleotide in the template strand lacking its base obviously is problematic during replication. DNA polymerases of the Y-family can handle any of these obstacles. Their active site is more spacious, which allows faulty structures to be processed, but may also contribute to their low fidelity. Moreover, the ‘finger’ domain of the molecule, that typically checks the minor groove of the produced DNA helix, is more rigid and thus less sensitive for aberrations. What signals determine that TLS polymerase should take over from Pol III? Apparently, the β clamp can bind two polymerases at the same time, one of which can be a TLS Polymerase, which would mean one is always at hand in a replisome. A stalled replisome seems to be the trigger for this polymerase to take over, but once the replisome moves again, Pol III continues and is preferred over the alternative enzyme, by an unknown mechanism. Different TLS polymerases have different specificities, so the type of lesion determines which one can solve the problem best. Maybe the binding of these alternative enzymes to the beta clamp is stochastic, and once a replisome is stalled, sooner or later the correct 'repair enzyme' will come along.

Recombination repair is a third mechanism by which stalled replication forks can be resolved, as was presented in Chapter 3. It involves the enzyme RecA. The recombinational machinery of the cell has probably evolved to resolve stalled replication forks, though the same enzymes can also introduce permanent changes in the genome, as will be discussed in the next chapter.
4.5. Concluding remarks

The complexity of chromosomal DNA replication is amazing, although from a biological perspective, such complexity is necessary for this important process. DNA synthesis needs to be robust and reliable, as it is vital to the cell. This is accomplished by built-in redundancy and back-ups, which in combination can handle many problems. There are many different DNA polymerase systems in bacteria, and although the system of E. coli is the best studied, it is not the only system in the bacterial world.

The speed of chromosomal DNA replication is also amazing. A growing chromosome increases in length at a speed of 1000 nt per second, or about 1 Mbp every 17 minutes. DNA replication is also amazingly accurate, when all proof-reading systems are active. The error rate in bacteria grown in culture (where they are well-fed and DNA-damaging conditions are absent, in contrast to many bacteria growing in Nature) can be in the range of 1 in 10 million. As a consequence, an entire bacterial chromosome can be replicated without a single error! From this perspective, the complexity yields a great deal of accuracy. However, under conditions of stress and low energy levels, the mutational frequency is much higher, which can allow the bacteria to perhaps mutate in ways that might be beneficial.

Recommended reading:


