

1

The Bacterial Genome – Where the Genes Are

1.1 Genome Philosophy

The genome of a bacterium consists of its entire collective of genes, and these can be located in a chromosome (or chromosomes) and on extra-chromosomal autonomous replicons such as plasmids. Chromosomes and plasmids replicate, copying the genes that they carry, with the replicon copies being segregated into the daughter cells at cell division. This process drives the vertical transmission of genetic information from one generation to the next and its fidelity determines the stability of the genetic information in the genome. If non-lethal errors occur during the replication of the genome the resulting mutations will be transmitted to the next generation.

The vertical transmission of genetic copying errors is one of the driving forces of evolution in all types of organisms. Bacterial cells are also prone to the evolutionary influence that is horizontal gene transfer (HGT). Here, foreign genes are transmitted to the bacterium, adding to its genetic complement. Evolution through HGT is much less subtle than evolution through the vertical transmission of copying errors which often involves the gradual accumulation of single nucleotide changes to the genome over many generations. Other types of mutation that are transmitted vertically are inversions, insertions, and deletions of genomic DNA. In HGT a bacterium can acquire entire clusters of genes *en bloc*, resulting in the acquisition of novel capabilities in a single generation. Examples include the arrival of genes that make the bacterium resistant to an antibiotic or to a heavy metal that previously could kill it, or genes that allow the organism to colonise a niche in the environment from which it had previously been excluded.

HGT played a key role in the early research work that led to the mapping of the bacterial genome and to our understanding of the locations of its genes. Among the autonomously replicating plasmids found in bacteria are elements that can promote their own transfer from cell to cell. The fertility, or 'F', factor of *Escherichia coli* was among the first to be studied. F encodes proteins that can build a connecting bridge between the F⁺ (or male) cell and one that is F⁻ (female). F transfers one of its DNA strands to the F⁻ cell in a process called conjugation. This is the bacterial equivalent of sex. It resembles sex in higher organisms in that the participants are male (F⁺) and female (F⁻) but it differs from conventional sex because the process converts the female into an F⁺ male. The F plasmid has segments of DNA that are identical in sequence, or almost identical, to DNA segments found in the chromosome. Usually these are mobile genetic elements called insertion

sequences (IS). The homologous recombination machinery of the cell can recombine the F-associated region of DNA sequence identity with a chromosomal counterpart, causing F to become fused with that part of the chromosome (Hadley and Deonier 1980). Where this happens is determined by the location of the IS element, and these mobile elements can be found at sites distributed around the chromosome.

Once F has become one with the chromosome, it is replicated as a part of that molecule. It can still engage in conjugation, however. When this happens, the DNA that is transferred to the F⁻ female bacterium consists of chromosomal DNA with F DNA in the vanguard. Strains that can act as DNA donors in these matings are called ‘Hfr’ (high frequency of recombination) (Reeves 1960) and homologous recombination between the incoming, horizontally transferred DNA and the resident chromosome allows the order of the genes on the chromosome to be determined. Experiments, in which chromosomal gene transfer mediated by the F plasmid was monitored as a function of time, allowed a rudimentary genetic map of the *E. coli* chromosome to be assembled (Bachmann 1983; Brooks Low 1991). Because the mating experiments were allowed to proceed for fixed periods of time before the deliberate breakage of the conjugation bridges by mechanical shearing, these early genetic maps were calibrated in ‘minutes’. It was discovered that it took 100 minutes to transfer the entire *E. coli* chromosome from one cell to another by conjugation (Bachmann 1983; Brooks Low 1991). Similar experiments were performed for other bacterial species, including the pathogen *Salmonella*, giving rough approximations of the physical scale of bacterial genomes (Sanderson and Roth 1988). Hfr strains could also mediate gene transfer between *E. coli* and *Salmonella* (Schneider et al. 1961). When the F plasmid is excised from the chromosome, genes that had been adjacent to the plasmid can be removed too, becoming part of the autonomously replicating episome. The plasmids are called F-prime (F') and have proved to be very useful in genetic analysis. The chromosomal gene ‘cargo’ can be transferred to F-minus strains by conjugation and this phenomenon can be exploited in genetic complementation experiments. Work of this type provided useful information about gene order and the position and nature of genetic mutations. F-primes have been used to investigate plasmid stability, incompatibility, and DNA replication: for example, the F'-*lac* episome was used extensively to study plasmid replication in *E. coli* (Davis and Helmstetter 1973; Dubnau and Maas 1968). Experiments with *E. coli* mutants deficient in Hfr recombination led to the discovery of important genes involved in homologous recombination: for example, *recA* (Clark and Margulies 1965), *recB*, and *recC* (Barbour and Clark 1970; Willetts et al. 1969; Youngs and Bernstein 1973).

HGT also provided a means for more refined mapping of genomes. Bacteriophages (often abbreviated to ‘phages’) are viruses that replicate in bacterial cells. Some phages package bacterial DNA in their viral heads as they exit the bacterial host and this DNA is transferred to the next bacterium that they manage to infect in a process known as transduction. The length of the DNA segment that a phage head can accommodate is finite and known in the cases of the viruses most commonly used for generalised transduction in *E. coli* (P1, 100–115 kb) and *Salmonella* (P22, 42 kb) (Sternberg and Maurer 1991). Therefore, genes that are co-transduced must be within a distance of one another that is compatible with being co-packaged by the phage. Very sophisticated experiments with transducing phage allowed not only gene-to-gene distance relationships to be determined but also the

measurement of the physical relationships between features of individual genes, such as their regulatory elements.

Thus, HGT has driven bacterial evolution and microbial geneticists have exploited it to assemble the first genetic maps of bacterial genomes. Genetic engineers have also used HGT to build novel variants of bacterial genomes in the lab. Cloning experiments using vectors based on natural or engineered plasmids rely on the HGT process known as transformation to move new DNA sequences into bacterial cells. A bacterium that is susceptible to transformation is said to be 'competent' and competence can be induced chemically or by electric shock (Hanahan et al. 1991). In addition, many bacterial species are naturally competent and therefore open to the uptake of foreign DNA from the environment. Knowledge and application of the HGT processes of transformation, transduction, and conjugation have revolutionised our understanding of bacterial genomes in a matter of decades. Genome sequencing has extended and deepened this knowledge.

Foreign DNA entering bacterial cells may undergo surveillance. Much attention has been focused on clustered regularly interspaced short palindromic repeat (CRISPR) systems both in their natural roles as systems that identify and destroy 'non-self' DNA, and as a result of their promise as agents of genome editing (Barrangou et al. 2007; Brouns et al. 2008; Garneau et al. 2010). Restriction endonucleases and their associated DNA methylases represent another mechanism for defending the bacterial cell from foreign DNA. Here, the methylases chemically modify the newly synthesised DNA of the bacterial genome so that it matches the 'approved' pattern: incoming DNA that lacks this methylation pattern is cut into pieces by the restriction endonucleases. These DNA surveillance mechanisms help to control the access of foreign DNA to established genomes.

The concepts of 'foreign DNA' and 'established genomes' can also be expressed by the terms 'accessory genome' and 'core genome', respectively. 'Accessory' implies that portion of the genome is not essential for the life of the bacterium, and that may be true in the artificial environment of the laboratory. The name also suggests that some form of value is added to the life of the organism, but that this is conditional. In fact, the same can be said of any gene or portion of the genome. In some cases, the essential nature of a genome component is made obvious because the bacterium dies if this component is eliminated. However, this can apply to a portion of the accessory genome just as much as to one of the core genome, depending on the circumstances of the bacterium. A facile example concerns the presence in a bacterium of a gene encoding resistance to penicillin. The gene is not a part of the core genome (it may even be located on a plasmid and not on the chromosome) and it is not essential unless penicillin-class antibiotics appear in the neighbourhood. In the absence of this gene during periods of cell wall peptidoglycan synthesis, the bacterium dies if penicillin-class antibiotics are introduced; despite being a part of the accessory genome, the resistance gene is now an essential gene.

The accessory genome is distinguished from the core genome in being of more recent arrival in the cell. It is not a monolithic entity but a mosaic of imported genetic components that have arrived through HGT, possibly over a very long period of time. In this context, it is very important to realise that genome evolution is not only a process of gene acquisition: gene loss is equally important. A gene may be lost safely if another member of the genome can supply its function, if the cell can acquire the lost product from an exogenous source, or if the selective pressure to retain the gene has been removed.

The core genome consists of those genes that are essential for the life and reproduction of the cell and that are widely shared by other organisms, even those that are only distantly related. Thus, the analysis of genome composition using computers to compare and contrast the genes possessed by thousands of bacteria has helped to identify those genes that are truly almost universally present in bacteria. These designations of essentiality have been supported in some cases by experiments in which the genes have been removed and the impact of their loss on the survival of the bacterium has been measured (Baba et al. 2006; Gerdes et al. 2003; Goodall et al. 2018; Rousset et al. 2018). Essential genes include those whose products replicate the genome, transcribe the genes, translate the messages, and operate the principal metabolic pathways of the cell. In many cases, redundancies are revealed where more than one gene can contribute to an essential process. For example, *E. coli* has seven operons that contribute to the building of ribosomes, so losing one is not life threatening. On the other hand, the loss of even an apparently redundant gene might impose a fitness cost when a bacterium that has lost this gene competes with one that has not (Condon et al. 1995a).

An essential gene may cease to be essential if another microbe can supply the missing function. This phenomenon is easily illustrated in the laboratory by cross feeding of the mutant by a strain lacking the mutation, but it is not confined to metabolic functions. For example, the absence of an apparently essential virulence gene in one pathogen during infection can be compensated by a function encoded by a second, co-infecting pathogen (Ibberson et al. 2017).

Loss of competitive fitness arises when a change to the genome (a mutation) renders the bacterium unable to compete with an otherwise genetically identical counterpart. While this can result from the loss of a gene it can also be caused by gene acquisition. Indeed, the negative effect even may arise simply due to the process of expressing the new gene, and not to the effect on the cell of the new gene product (Stoebel et al. 2008a). This illustrates the subtle nature of the causes of competitive fitness differences and their relationships to genome composition and structure.

We will begin by considering genome composition and structure in the model bacterium *E. coli* and some others where useful data are available. This survey will provide information about any discernable rules governing these important aspects of microbial cell biology.

1.2 The Bacterial Chromosome

E. coli K-12 has played a central role in the history of bacterial genetics and bacterial physiology. The original K-12 isolate came from a stool sample from a human patient suffering from diphtheria and was cultured in Palo Alto, California, USA, in 1922 (Bachmann 1996). This isolate was the ancestor of W1485 from the Joshua Lederberg laboratory, the isolate that was named MG1655 by Mark Guyer (hence ‘MG’). The first *E. coli* chromosome to be sequenced came from this intensively studied MG1655 strain (Blattner et al. 1997). However, this was not the first bacterial chromosome to have its complete nucleotide sequence determined: that honour belongs to *Haemophilus influenza* (Fleischmann et al. 1995).

The Blattner lab chose MG1655 because it has undergone relatively little genetic manipulation and is considered a good representative of wild-type *E. coli*. It has been cured of bacteriophage lambda and of the F plasmid and has few genetic lesions. An *ilvG*

mutation deprives it of acetohydroxy acid synthase II, making it prone to valine-dependent isoleucine starvation (Lawther et al. 1981, 1982) and there is an IS5 insertion in the *rfb* locus that interferes with O-antigen synthesis (Liu and Reeves 1994). If this mutation is repaired, the bacterium has its lipopolysaccharide expression reinstated and it becomes pathogenic in an infection model based on the worm *Caenorhabditis elegans* (Browning et al. 2013). Strain MG1655 displays mild starvation for pyrimidine arising from poor expression of its *pyrE* gene: the cause is a frameshift mutation at the end of the *rph* locus (*rph-1*) (Jensen 1993). Interestingly, genome sequence analysis shows that MG1655 is closely related to NCTC 86, the bacterium originally named *Bacillus coli* by Theodor Escherich in 1885, isolated before the antibiotic era (Dunne et al. 2017).

The *E. coli* K-12 chromosome is a single, covalently closed, circular, double-stranded DNA molecule of 4 639 221 bp (Blattner et al. 1997). Although chromosome circularity is the norm in *E. coli*, cells in which the chromosome is artificially linearised (with the ends closed by hairpin turns) are viable, show few alterations in gene expression, have normal nucleoid structure, and do not display growth defects (Cui et al. 2007). Thus, the circular nature of the chromosome is not essential for its functionality or for its ability to be replicated and to be segregated at cell division.

The *E. coli* chromosome was visualised originally in the early 1960s by autoradiography of cells fed with tritiated thymidine in a classic experiment that also revealed the existence of the moving replication fork (Cairns 1963a,b). The chromosome undergoes bi-directional replication from its *oriC* locus (Kaguni 2011), creating two replichores: Left and Right (Figure 1.1) (Lesterlin et al. 2005; Wang, X., et al. 2006). Through a process of

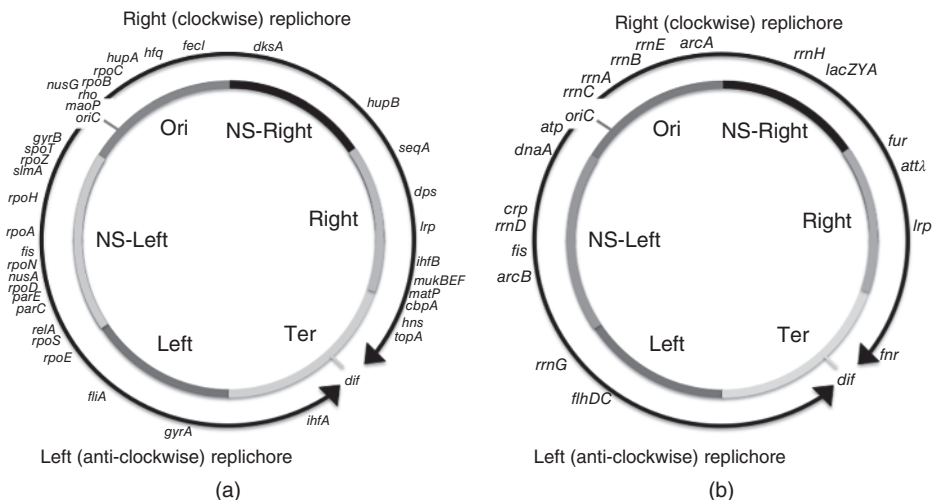


Figure 1.1 The macrodomain structure of the *E. coli* chromosome. Shaded segments represent the Ori, Right, Ter, and Left macrodomains, and the Left and Right non-structured regions. The curved arrows outside the circular chromosome represent the Left (anticlockwise) and Right (clockwise) replichores. (a) The positions of genes that encode NAPs, chromosome organisation factors, topoisomerases, proteins involved in the process of transcription, the Hfq RNA chaperone are indicated around the periphery of the chromosome. (b) The positions of the seven *rrn* operons and genes encoding transcription regulators that are discussed in the text are shown. The positions of the *lac* operon and the bacteriophage lambda attachment site (*attλ*) are also indicated. (See colour plate section for colour representation of this figure)

semi-conservative DNA replication, the bacterium acquires a second copy of its chromosome prior to cell division. In rapidly growing bacteria, one or more additional rounds of chromosome replication are initiated before the first one is completed, creating multiple copies of those chromosomal sequences that lie closest to *oriC* (Figure 1.1) (Cooper and Helmstetter 1968). Genes in the *oriC*-proximal zones of the *E. coli* chromosome will be present in higher copy numbers than genes in *Ter*, the region of the chromosome where replication terminates. In slower-growing bacterial populations, gene copy numbers are more in balance around the chromosome with only a twofold difference in copy number between genes close to *oriC* and those near *Ter*.

Most of our knowledge about chromosome replication and segregation comes from studying a handful of model organisms: *E. coli*, *Caulobacter crescentus*, *Vibrio cholerae*, and *Bacillus subtilis*. The focus in this chapter will be on *E. coli*, with comparisons to other organisms where this is useful.

1.3 Chromosome Replication: Initiation

Chromosome replication, segregation, and cell division are complex processes that must be coordinated to ensure the successful replication of the cell (Reyes-Lamothe et al. 2012). The nutritional status of the cell and its metabolic flux are very influential in achieving this coordination and they have a direct bearing on the growth rate of the culture (Wang and Levin 2009).

Replication of the *E. coli* chromosome begins at a specific site, *oriC*, which has a number of important DNA sequence elements called DnaA boxes that make up the DnaA Oligomerisation Region, DOR (Figure 1.2) (Fuller et al. 1984; Jameson and Wilkinson 2017; Katayama et al. 2017). These boxes are bound by DnaA, an adenosine triphosphate (ATP)-dependent initiator protein (Schaper and Messer 1995; Sutton and Kaguni 1997), which then forms a right-handed helical protein oligomer along the DNA that unwinds the duplex at an A+T-rich element known as the DNA Unwinding Element, DUE (Bramhill and Kornberg 1988a; Kowalski and Eddy 1989) (Figure 1.2). The DnaA oligomerisation process is assisted by another protein called DiaA (Ishida et al. 2004). The DUE has an A-rich and a T-rich DNA strand; once it is unwound, the T-rich strand binds to the DnaA oligomers at the DOR. A helicase loader known as DnaC then loads the DnaB helicase onto the single-stranded DNA (Koboris and Kornberg 1982). This helicase then recruits in turn the DnaG primase and DnaN, the DNA polymerase beta-clamp (Fang et al. 1999). When fully assembled, this complex is known as the replisome (Figures 1.3 and 1.4).

In *oriC* of *E. coli*, the DnaA boxes are of variable affinity for the DnaA protein (Blaesing et al. 2000) (Figure 1.2). Boxes with high affinity bind DnaA that is in a complex with either ATP or ADP, whereas weak boxes bind only DnaA that has bound ATP (Grimwade et al. 2007). Binding of DnaA to *oriC* is cooperative, with DnaA-ATP that has bound to strong boxes facilitating the subsequent binding of DnaA-ATP to the weaker sites, promoting the formation of the DnaA oligomer at the origin of replication (Miller et al. 2009; Kaur et al. 2014). The activity of DnaA may also be controlled by reversible acetylation at lysine residues: of the 13 lysine amino acids in DnaA, acetylation of residues K178 and K243 seems to be especially important in promoting the initiation of chromosome replication (Li et al. 2017; Zhang, Q., et al. 2016).

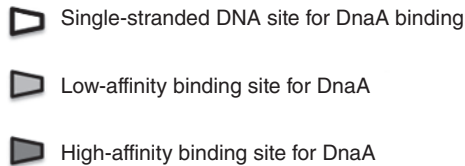


Figure 1.2 Structure of *oriC* on the *E. coli* chromosome. The ATP-dependent DnaA protein binds to sites throughout *oriC* and oligomerises in the DnaA Oligomerisation Region (DOR), driving DNA unwinding at the A+T-rich DNA Unwinding Element (DUE). Single-stranded T-rich DNA in DUE binds to the DnaA oligomers at DOR. High-affinity sites bind DnaA-ATP or DnaA-ADP; low affinity sites bind just DnaA-ATP. Binding sites for the NAPs FIS and IHF are also shown: FIS and IHF modulate the process of replication initiation negatively and positively, respectively. The Dam methylase methylates *oriC* at several 5'-GATC-3' sites (indicated by vertical arrows): hemimethylated sites bind SeqA, excluding DnaA and preventing untimely re-initiation of chromosome replication.

The availability of DnaA-ATP is a rate-limiting factor for the initiation of chromosome replication. A protein called Hda converts active DnaA-ATP into inactive DnaA-ADP through ATP hydrolysis (Kato and Katayama 2001). This conversion also requires DnaN, the DNA polymerase beta-clamp, linking ATP hydrolysis to the elongation phase of DNA synthesis (Takata et al. 2004).

The nucleoid-associated proteins (NAPs) Integration Host Factor (IHF) and the Factor for Inversion Stimulation (FIS) are DNA-binding and -bending proteins that are thought to play important architectural roles at the origin of replication (Figure 1.2) (Kasho et al. 2014; Ryan et al. 2004). IHF has a positive role at *oriC* where it binds to a specific DNA sequence, introducing a DNA bend that encourages DnaA binding and oligomer formation; it can also redistribute DnaA on supercoiled DNA (Grimwade et al. 2000). While some work has not found a major role for FIS in regulating events at *oriC* (Weigel et al. 2001) data from other investigations show that, in contrast to IHF, the role of FIS is inhibitory to DNA replication: when it binds to *oriC* it interferes with the binding of IHF and DnaA, blocking unwinding of the DUE sequence (Ryan et al. 2004).

The many 5'-GATC-3' sites found throughout *oriC* (Figure 1.2) are hemimethylated in the period immediately following the initiation of chromosome replication (Lu et al. 1994).

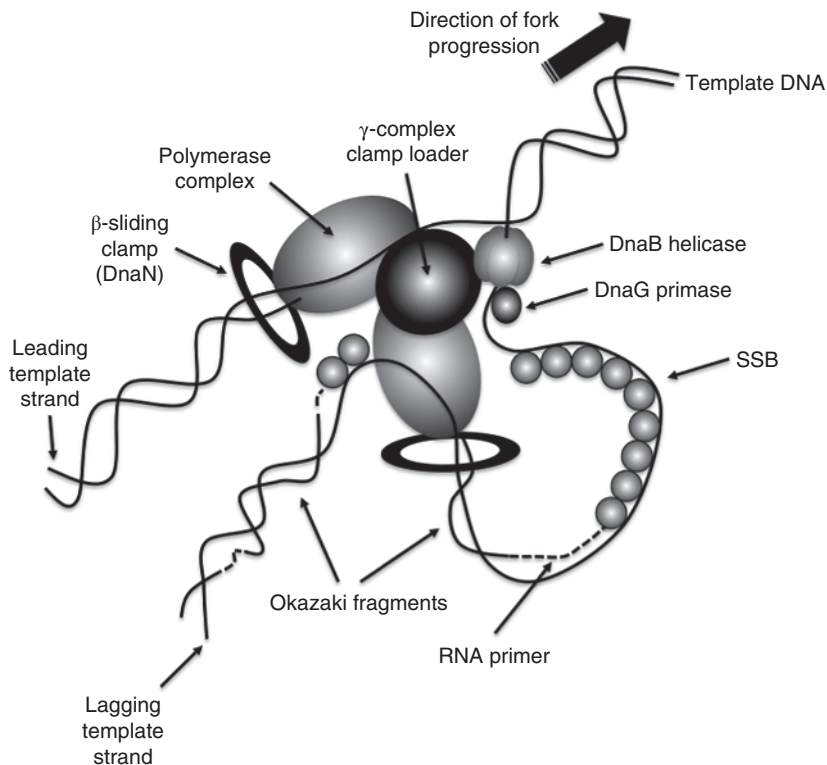


Figure 1.3 Structure of the *E. coli* replisome in chromosome replication. The replisome is made up of the two cores of DNA Polymerase III, a gamma (γ) complex (or clamp loader) and the beta clamp together with a hexameric helicase, the DnaG primase, and the single-stranded binding protein, SSB. The DnaB helicase uses energy from ATP hydrolysis to translocate along the lagging strand, unwinding the DNA duplex. The two Polymerase III cores, linked by the tau subunits (Figure 1.4), are each dedicated to coordinated and simultaneous replication of the leading and lagging template strands of the replication fork. The ring-like beta (β) clamp (DnaN), or processivity factor, encircles DNA and is attached to the replisome via the alpha (α) subunit. The β clamp stabilises the moving replication machine on its template, allowing it to operate with a high degree of processivity. A single-stranded DNA bubble is formed by the unwinding action of the replisome and SSB protein coats the ssDNA. The DnaG primase interacts with the helicase to generate RNA primers that are used to prime Okazaki fragment synthesis.

The SeqA protein binds to these hemimethylated sites, preventing immediate and untimely re-initiation of chromosome replication by DnaA: SeqA also downregulates the expression of the negatively autoregulated *dnaA* gene (Campbell and Kleckner 1990; Waldminghaus and Skarstad 2009). Dam-mediated methylation of the 5'-GATC-3' sites is inhibitory to SeqA binding and re-admits DnaA to *oriC* (Lu et al. 1994).

E. coli uses clusters of DnaA binding sites that are located outside *oriC* to modulate the initiation of chromosome replication (Figure 1.5). One of these is the 183-bp *datA* site, located next to the *vjeV* gene on the *E. coli* chromosome. The *datA* site is made up of five high-affinity DnaA binding sites (Kitagawa et al. 1996); *datA* also binds IHF (Nozaki et al. 2009). The interaction of IHF with *datA* occurs immediately after the initiation of

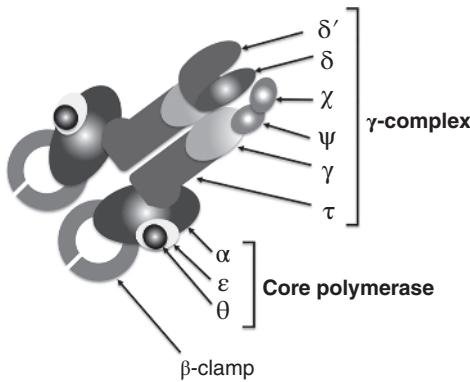


Figure 1.4 The structure of the DNA polymerase III subassembly (Pol III*). The core is made up of the alpha (α), epsilon (ϵ), and theta (θ) subunits and the holoenzyme contains two cores. The tau (τ) subunit (two copies) links the cores together, ensuring simultaneous replication of the leading and lagging strands (Figure 1.3). The function of the cores is DNA synthesis on the leading and lagging strands (Figure 1.3). The clamp loader (or gamma complex) is made up of the chi (χ), delta (δ), delta-prime (δ'), gamma (γ), and psi (ψ) subunits. The gamma subunit loads the beta clamp onto the DNA that is primed for de novo DNA synthesis. The arrival of the beta (β) clamp (processivity factor) converts the Pol III* subassembly into the Pol III holoenzyme (Figure 1.3). (See colour plate section for colour representation of this figure)

chromosome replication and this facilitates the binding of DnaA-ATP to *dataA* (Nozaki et al. 2009). DnaA-ATP bound to *dataA* undergoes ATP hydrolysis, reducing the size of the pool of DnaA-ATP that is available for binding to *oriC* (Ogawa et al. 2002). This IHF-dependent process has a negative influence on the frequency with which chromosome replication is initiated at *oriC* (Kasho and Katayama 2012).

Conversion of DnaA-ADP to DnaA-ATP has been associated with two so-called DnaA Reactivation Sites, DARS1 and DARS2 (Fujimitsu et al. 2009) (Figure 1.5). DARS1 is 103 bp in length, has three DnaA binding sites, and is located upstream of *uvrB* in *E. coli*. The DARS2 site is more sophisticated. It is 455 bp in length and is located upstream of the *mutH* gene in *E. coli*. DARS2 binds IHF and FIS in addition to DnaA. Binding of these NAPs to DARS2 stimulates the conversion of DnaA-ADP to DnaA-ATP. IHF binding is cell cycle determined while FIS binding is growth phase determined: FIS binds in rapidly growing cells and this is consistent with the observation that FIS stimulates DNA replication in rapidly growing *E. coli* (Flåtten and Skarstad 2013; Kasho et al. 2014). The chromosomal locations of *dataA* and the DARS elements seem to be important for their function: if they are repositioned, the chromosome replication control is disrupted (Frimodt-Møller et al. 2016).

The *oriC* locus is found between two highly conserved genes, *mioC* and *gidA* (Figure 1.6). The *mioC* gene is transcribed towards *oriC* while *gidA* is transcribed away from it. The two genes exhibit reciprocal transcription patterns that are functions of the cell cycle: *mioC* transcription is maximal midway through chromosome replication while *gidA* transcription is minimal at that point; maximal expression of *gidA* coincides with the onset of septation and cell division (Lies et al. 2015). MraZ, a protein possibly involved in cell division control, binds and represses the *mioC* promoter (Eraso et al. 2014) and this promoter is also stringently regulated, linking *mioC* transcription to growth rate

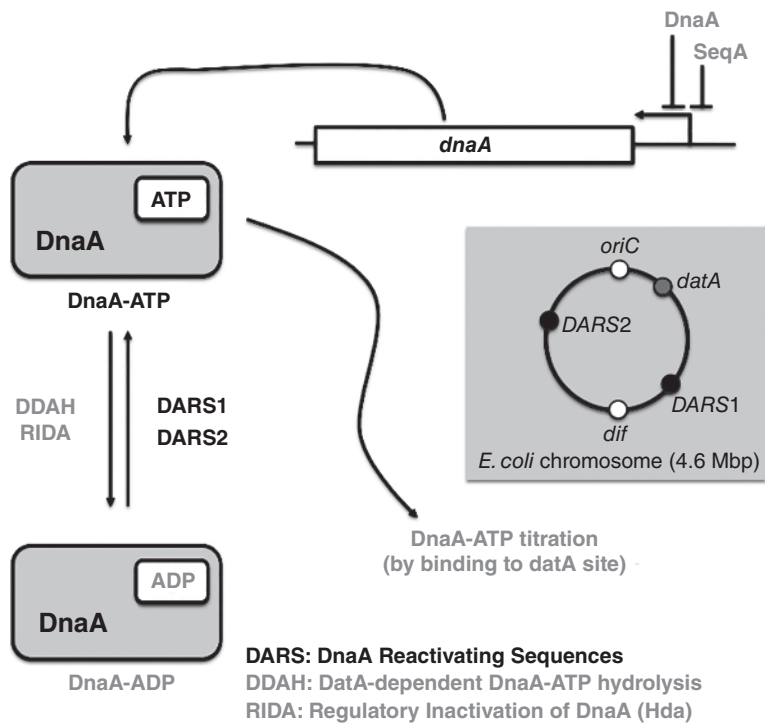


Figure 1.5 The control of DnaA production and activity. The SeqA and DnaA proteins regulate expression of the *dnaA* gene negatively. DnaA-ATP is generated at the DnaA Reactivating Sequences DARS1 and DARS2, and is converted to DnaA-ADP by ATP hydrolysis (i) at the *datA* site stimulated by binding of IHF in a process called *datA*-dependent DnaA-ATP Hydrolysis (DDAH) and (ii) by Regulatory Inactivation of DnaA (RIDA) in which the DnaA inhibitor protein Hda catalyses the hydrolysis of DnaA-bound ATP to ADP, yielding DnaA-ADP. Hda activation in RIDA follows interaction with the DNA polymerase clamp on newly synthesised DNA. The relative locations of *datA* (4.39 Mb), *DARS1* (0.81 Mb) and *DARS2* (2.97 Mb) with respect to the *oriC* and *dif* sites on the 4.6 Mb *E. coli* chromosome are shown (inset). Black lettering: generation of DnaA-ATP, grey lettering: conversion of DnaA-ATP to DnaA-ADP.

(Chiaromello and Zyskind 1989). The biological function of MioC is not firmly established, although it has been reported to be involved in biotin metabolism (Birch et al. 2001). The GidA protein contributes to tRNA modification, working in association with MnmE (GidA is also known as MnmG) (Bregon et al. 2001). Neither protein is thought to have DNA-binding activity. Transcription of *mioC* is repressed by DnaA acting at a DnaA box in the promoter. The initiation of chromosome replication displaces DnaA and de-represses *mioC*, with the return of DnaA being delayed as the protein is recruited by the new DnaA binding sites generated by replication (Bogan and Helmstetter 1996). Transcription of *gidA* is repressed by SeqA when this protein binds to the 5'-GATC-3' sites at the promoter that become hemimethylated following DNA replication (Bogan and Helmstetter 1997). The process of transcribing *gidA* and *mioC* is important for the initiation of chromosome replication at *oriC* (Bramhill and Kornberg 1988b; Theisen et al. 1993), at least under some circumstances (Asai et al. 1998; Bates et al. 1997; Lies et al. 2015).

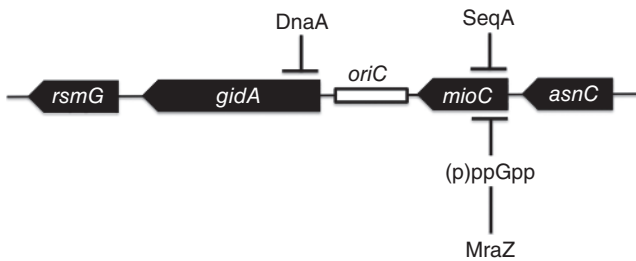


Figure 1.6 The genetic neighbourhood of *oriC* in *E. coli*. Filled arrows represent the genes and an open rectangle indicates the position of *oriC*. DnaA represses the *gidA* gene transcriptionally through DnaA boxes that overlap the *gidA* promoter. The *mioC* gene is repressed by SeqA binding to hemimethylated versions of 5'-GATC-3' sites at the promoter that are generated by DNA replication. The *mioC* promoter is also subject to stringent control via the (p)ppGpp alarmone and it is repressed by MraZ, a protein that has been linked to the control of cell division. The *rsmG* gene encodes a methyltransferase for the modification of 16S rRNA (see Benítez-Páez et al. 2012). The *asnC* gene encodes a HTH-motif-containing transcription regulator that is related to LRP and controls genes involved in asparagine metabolism (see Kölling and Lother 1985; Willins et al. 1991). Termination of transcription extending from *asnC* to *mioC* is dependent on a DnaA-DNA complex at the *asnC* terminator, as described by Schaefer and Messer (1988).

1.4 Chromosome Replication: Elongation

Once replication has been initiated, the replisome is responsible for progressive DNA synthesis during the elongation phase of chromosome replication. This large complex is composed of a pentameric clamp loader, the DNA polymerase clamp (DnaN), the three-subunit DNA primase (DnaG), and the hexameric helicase DnaB (Bailey et al. 2007; Reyes-Lamothe et al. 2010) (Figure 1.4). The helicase uses ATP hydrolysis to unwind the DNA duplex, moving along the lagging strand of the DNA as it does so. Single-stranded DNA-binding protein (SSB) coats the separated ssDNA strands, thus preventing reformation of the duplex by religation and attack by nucleases (Beattie and Reyes-Lamothe 2015).

The primase, DnaG, possesses a central RNA polymerase domain where the RNA primers used in DNA synthesis are manufactured (Corn et al. 2008). The primer emerges from the DnaG-DnaB complex and is transferred to DNA polymerase and SSB (Corn et al. 2008). DNA Polymerase III works with the beta-clamp protein (DnaN) to extend the primer, creating a new DNA strand at a rate of 1000 bases per second (Beattie and Reyes-Lamothe 2015). It is advantageous to have DnaN as a component of the replisome because a beta-clamp must be reloaded for the synthesis of each lagging strand Okazaki fragment (Beattie and Reyes-Lamothe 2015). If the replication fork stalls or breaks, replication can be restarted through a DnaA-independent mechanism. Here, the PriA helicase, in association with accessory proteins such as PriB, PriC, and DnaT, binds to the gapped replication fork and loads DnaBC. In some cases, the restart may be associated with a strong transcription promoter that generates an R-loop where PriA can introduce DnaBC on the displaced DNA strand (Heller and Marians 2006; Kogoma 1997). Of the approximately 300 copies of DNA gyrase that are bound to the *E. coli* chromosome at any one time, about 12 accompany each moving replication fork to manage the DNA topological disturbance that is associated with fork migration (Stracy et al. 2019).

1.5 Chromosome Replication: Termination

Termination of DNA synthesis occurs within *Ter*, located at a point that is diametrically opposite *oriC* on the chromosome (Hill et al. 1987) (Figure 1.7). The *Ter* region has five copies of a 23-bp DNA element on each flank and the 36-kDa Tus protein binds to these sequences (Neylon et al. 2005). The Tus binding sites are asymmetric and have a permissive and a non-permissive orientation (Figure 1.7). Replication forks can pass the Tus-*Ter* nucleoprotein complexes when the DNA sequences are in the permissive orientation, but fork movement becomes arrested when the sequences are oriented in the non-permissive direction. The mechanism of replication fork passage at *Ter* sites that are in the permissive orientation involves displacement of Tus by the DnaB helicase; when in

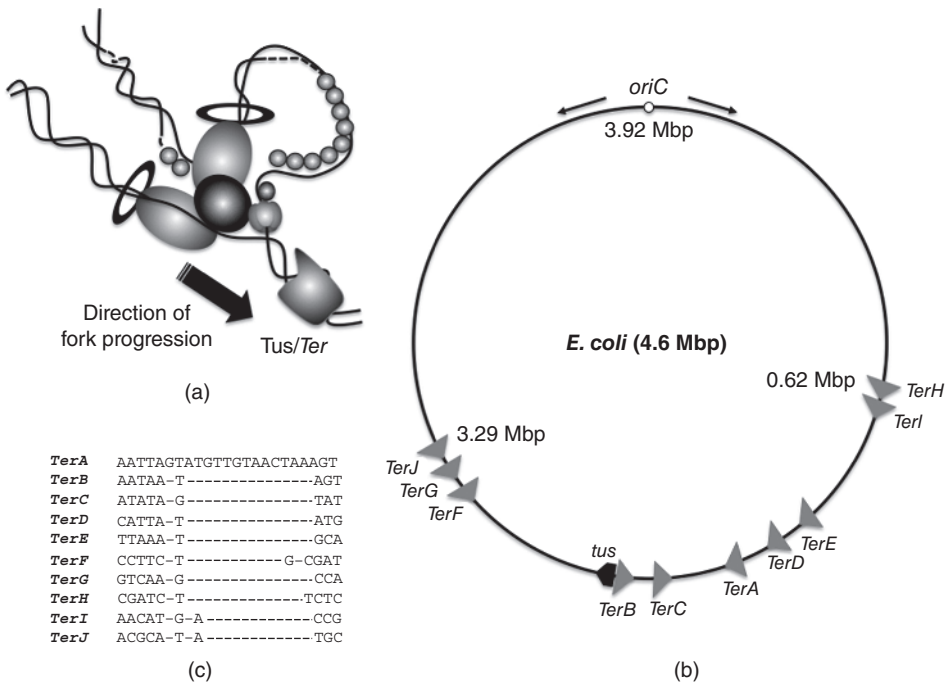


Figure 1.7 Termination of chromosome replication in *E. coli*. (a) The moving replisome encounters an appropriately oriented Tus/*Ter* nucleoprotein complex and the interaction between Tus and the DnaB helicase halts replisome movement, leading to the termination of chromosome replication. (b) The 4.6 Mb chromosome of *E. coli* is shown, indicating the relative positions and orientations of the *Ter* sequences (grey arrowheads) with respect to one another and *oriC* and the *tus* gene. The black arrows on either side of *oriC* indicate the bidirectional nature of *E. coli* chromosome replication. *Ter* sites aligned with the direction of replication are in the permissive orientation and will allow the replication fork to pass; those oriented against the direction of fork movement are in the non-permissive configuration and will halt fork movement if bound by Tus. The promoter of the *tus* gene overlaps the *TerB* sequence, resulting in negative autoregulation of *tus* transcription by the Tus protein. (c) An alignment of the DNA sequences of the *Ter* elements, showing the high degree of sequence conservation among the sites and the lack of dyad symmetry within each site. The latter feature ensures that the sites operate to stop forks moving in one direction only.

the non-permissive orientation, Tus prevents DnaB, and the replication fork, from translocating past that point (Bastia et al. 2008; Berghuis et al. 2015; Mulcair et al. 2006). Single-molecule experiments performed *in vitro* have shown that the DNA also plays a critical role: in the non-permissive orientation, the unwinding of the DNA by the approaching replication fork creates a powerful lock at the Tus-Ter site that is an effective roadblock to further translocation by the fork; in the permissive orientation the lock does not operate and the fork can proceed (Berghuis et al. 2015).

The newly synthesised DNA strand is unmethylated and forms one part of a hemimethylated duplex. For this reason, the products of chromosome replication are chemically distinct from the template duplex until a full methylation of the newly synthesised strand has taken place. DNA adenine methyltransferase, Dam, methylates DNA at 5'-GATC-3' sites and there are 11 of these sites in *oriC* (Figure 1.2). The SeqA protein binds to these sites while they are still in their hemimethylated form, sequestering the origin and excluding DnaA (Han et al. 2003; Slater et al. 1995; von Freiesleben et al. 1994). The sequestered state persists for about one third of the cell cycle when it is relieved by dissociation of SeqA and methylation of the 5'-GATC-3' sites by Dam (Kang, S. et al. 1999; Lu et al. 1994). SeqA also interferes with expression of the *dnaA* gene, reducing the levels of the DnaA protein available for binding to *oriC* (Campbell and Kleckner 1990). In addition, SeqA contributes to processes that ensure proper segregation of the chromosome copies at cell division (Helgesen et al. 2015; Waldminghaus and Skarstad 2009). It is interesting to note that both hemimethylated *oriC* and SeqA have been shown to associate with the cell envelope (Ogden et al. 1988; Slater et al. 1995), perhaps indicating a role for the complex in the positioning of *oriC* in the cell.

1.6 Replication Produces Physically Connected Products

The converging replication forks moving along the chromosome will create a topological problem as they approach one another in the Ter region. As chromosome replication comes to an end, the products that it generates will emerge as intertwined DNA duplexes. This physical linkage must be resolved if it is not to impede chromosome segregation. Furthermore, if homologous recombination occurs between the sister chromosomes it may create a chromosome dimer. This process is made more likely by oxidative damage to DNA, as occurs in mutants defective in the Fur iron regulatory protein (Steiner and Kuempel 1998). The dimers arise from RecBCD- and RecFOR-mediated events with roughly equal frequency (Barre et al. 2001). Once formed, this dimer must be resolved at or before cell division or an anucleate cell will be created (see Section 1.8).

1.7 Decatenating the Sister Chromosomes

Bidirectional replication produces two interlinked copies of the chromosome and these must be decatenated before they can be segregated at cell division. Decatenation of fully intact duplexes is an intermolecular event that is catalysed by type II topoisomerases. Topoisomerase IV is the most efficient decatenase in such cases (Adams et al. 1992; Espéli et al. 2003; Zechiedrich and Cozzarelli 1995), although DNA gyrase can provide this function too

(Steck and Drlica 1984). Type I topoisomerases can also supply a decatenase function, but in this case one of the DNA strands in at least one of the DNA duplexes must be nicked, with the topoisomerase providing the break in the intact strand of the same duplex to permit passage of the intact duplex through the gap (DiGate and Mariani 1988).

1.8 Resolving Chromosome Dimers

The creation of chromosome dimers by homologous recombination between sister chromosomes threatens to interfere with chromosome segregation at cell division. The dimers are resolved by site-specific recombination at the *dif* sites within the terminus regions of the cointegrated chromosomes. These sites are arranged as directly repeated sequences in the dimer, albeit 4.6 million base pairs apart, and *dif* synapsis, strand exchange, and resolution separate the chromosome copies as monomeric molecules (Figure 1.8). The XerC and XerD tyrosine integrases act sequentially at *dif* to catalyse the reaction (Lesterlin et al. 2004; Sherratt et al. 1995); XerC creates the Holliday junction but this will collapse to substrate unless XerD completes the reaction (Barre et al. 2000; Hallet et al. 1999; Recchia and Sherratt 1999). The FtsK cell division protein triggers XerC/D-mediated dimer resolution at *dif* (Steiner et al. 1999). FtsK is located at the cell division septum at the mid-cell where it uses ATP hydrolysis to activate chromosome dimer resolution. Its location at the septum is

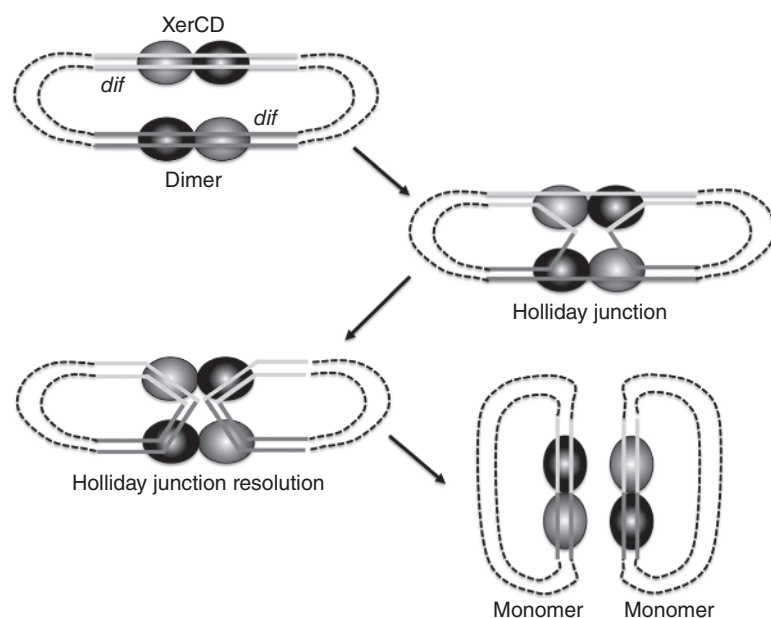


Figure 1.8 Resolution of a chromosome dimer by XerCD-mediated recombination at *dif*. Chromosome dimerisation can occur as a result of recombination-mediated DNA repair during replication. The dimers are resolved by site-specific recombination at two directly repeated *dif* sites located in the Ter region. The reaction is catalysed by the XerCD tyrosine integrases, generating monomeric chromosomes that can be segregated into the two daughter cells at cell division.

dependent on its interaction with the FtsZ septum ring protein (Dorazi and Dewar 2000a; Yu et al. 1998; Wang and Lutkenhaus 1998). FtsK moves DNA within the cell, assists sister chromosome synapsis and reduces DNA entanglement/catenaion of sister chromosomes (Sherratt et al. 2004). FtsK Orienting Polar Sequences (KOPS) located near the *dif* site in the Ter region of the chromosome contribute to the directional loading of the FtsK translocase, allowing it to segregate the daughter chromosome by driving a copy into each daughter cell (Bigot et al. 2006, 2007; Sivanathan et al. 2009; Stouf et al. 2013). The *dif* site and the wider Ter region of the chromosome remain in close association with the septum/nascent septum throughout the cell cycle (Niki et al. 2000).

It is not in the interest of the bacterium to have chromosome dimers arising frequently because this poses a risk that either anucleate cells will arise or that the chromosome may be damaged by the division septum as the cell attempts to segregate a dimeric chromosome. RecA-dependent homologous recombination events, whether they arise via the RecBCD or RecFOR pathways (Section 1.6), generate Holliday junctions that must be resolved by the Ruv resolvosome-based system. There is a bias in this process in favour of Ruv-mediated resolution that does not involve crossover, and therefore the creation of chromosome dimers (Cromie and Leach 2000; van Gool et al. 1999). This bias reduces the frequency at which chromosome dimers, and the associated threat to the wellbeing of the daughter cell genomes, arise.

The XerCD system is versatile and is used in site-specific recombination reactions with sequences related, but not identical, to *dif*, and accompanied by co-factor proteins, to resolve dimers in autonomously replicating plasmids (Clerget 1991; Colloms 2013; Colloms et al. 1990, 1998; Summers 1989). In the human pathogen *V. cholerae*, the CTX ϕ bacteriophage that carries the cholera toxin *ctxAB* operon integrates into chromosome I at its *dif* site using the XerC recombinase from the *V. cholerae* XerCD recombinases to catalyse the reaction (it can enter the corresponding location on chromosome II at a lower frequency) (Das et al. 2013; Huber and Waldor 2002; McLeod and Waldor 2004). CTX ϕ is a filamentous phage and only the plus strand of its genome integrates. The plus strand of the CTX ϕ genome folds to create a double-stranded region that encompasses the XerC and XerD binding sites of the phage flanking a mismatched and bulging phage *dif* site. Only a single stranded exchange occurs, mediated by XerC alone, and this creates a Holliday junction that is resolved by DNA replication (Val et al. 2005). The minus strand of the CTX ϕ , phage fails to generate a *dif* site with enough homology to recombine with the chromosomal counterpart, so this strand of the phage genome does not integrate. The integrated phage does not excise because its *dif* sites also lack sufficient homology with the chromosomal site to promote site-specific recombination (Val et al. 2005).

1.9 Segregating the Products of Chromosome Replication

The daughter chromosomes have to be moved to locations in the mother cell that correspond to the emerging daughter cells. At the end of the movement period, it should be possible to close the cell division septum between the nascent daughter cells without damaging the chromosome copies by guillotining them. The chromosome segregation

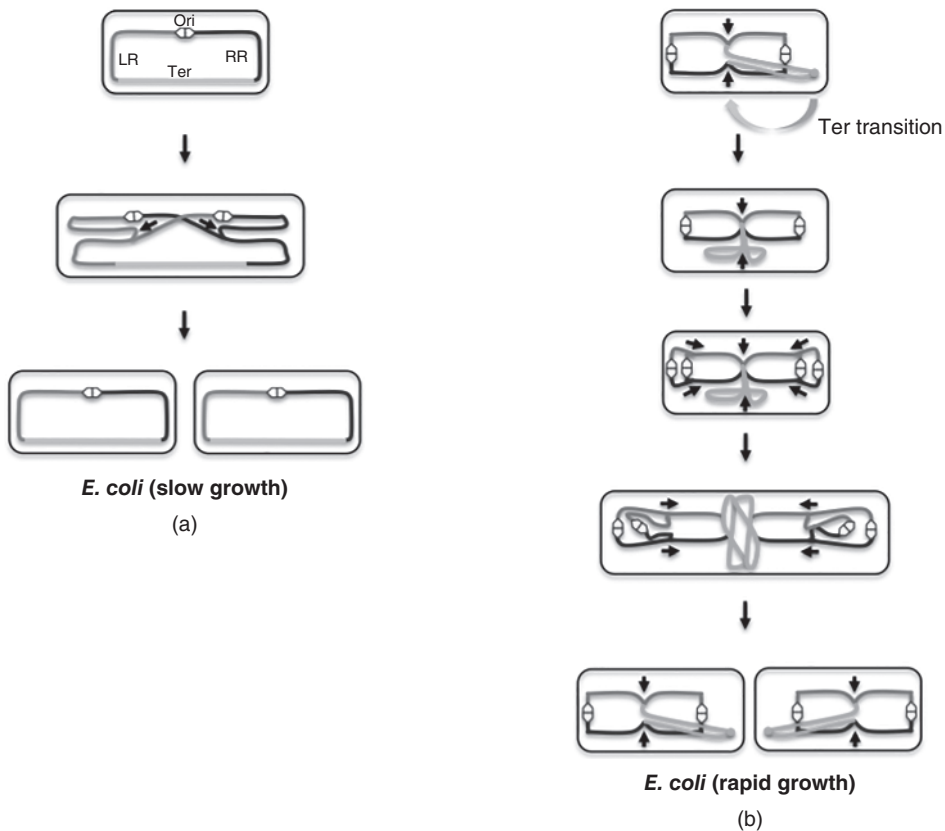


Figure 1.9 The choreography of chromosome movement in *E. coli* during the cell cycle. (a) In slow-growing *E. coli* cells, chromosome replication is initiated at the origin (Ori, green/yellow lozenge) located at mid-cell, and proceeds bidirectionally, copying the left (LR, light blue) and right (RR, dark blue) replichores and ending in the terminus (red), where the products of replication are decatenated prior to segregation into the daughter cells. Arrows drawn alongside the chromosome are used to indicate the direction of replication fork movement. (b) Rapidly growing *E. coli* cells have multiple rounds of chromosome replication underway simultaneously. Instead of the mid-cell position seen in slow-growing *E. coli* cells, the Ori is at the cell pole in the rapidly growing bacteria. The newly born cell has its Ter region displaced towards one pole of the cell and this undergoes a transition to the mid-cell. A second round of chromosome replication starts before the previous one is complete and multiple replication forks can be observed. The final separation of the daughter chromosomes is thought to exert a force at the terminus that moves this part of the chromosome to an eccentric position that is maintained in the daughter cell immediately after its birth (Youngren et al. 2014). (See colour plate section for colour representation of this figure)

process follows the Origin-to-Terminus axis, just as the replication process does (Bouet et al. 2014) (Figure 1.9).

The immediate products of replication behind the moving fork and replisome are catenated, positively supercoiled, interwound, double-stranded DNA molecules. Topo IV will attempt to decatenate these (Deibler et al. 2001; Espéli et al. 2003; Khodursky et al. 2000; Lopez et al. 2012). However, in a 300- to 400-kb sliding window immediately in the wake of

fork passage, the DNA is only hemimethylated. Any hemimethylated SeqA sites can bind the SeqA protein, which has the potential both to bridge DNA molecules and to exclude Topo IV, preventing decatenation (Joshi et al. 2013). The result is cohesion of the chromosome copies (Joshi et al. 2011; Nielsen et al. 2006a).

Two factors are thought to contribute to cohesion: exclusion of Topo IV by SeqA and bridging of daughter chromosomes by SeqA bound to hemimethylated copies of its binding sites. In support of the model, a positive correlation has been described between the time delay in ending cohesion and the density of SeqA-binding sites at *oriC* and at two so-called 'snap' loci (in the Right Replichore in *E. coli*) at which particularly strong physical tethers seem to fail suddenly as the replication-and-segregation process proceeds (Joshi et al. 2011). Chromosome cohesion must be overcome if segregation is to proceed to completion and this involves efficient decatenation of interwound chromosome copies and the breaking of any inter-chromosome bridges. The resulting segregation process appears to include a series of successive jerks as each tether breaks in turn (Fisher et al. 2013). Dam-mediated methylation of the newly synthesised DNA strands seems to be a key step in preventing SeqA binding and using its capacity to bridge chromosome copies and to interfere with Topo IV access to catenated DNA substrates behind the fork (Joshi et al. 2011). Cohesion does not require the Structural Maintenance of Chromosome (SMC) proteins, such as MukBEF in *E. coli*, despite their potential to encircle DNA duplexes (Adachi et al. 2008; Danilova et al. 2007; Joshi et al. 2013) (Figure 1.10).

The presence or absence of a ParAB-*parS* system is a major determinant of the pattern of chromosome segregation seen in a bacterium. *E. coli* does not possess such a system and forces of mutual repulsion acting on the chromosome copies as they emerge in the confined space of the rod-shaped cell may drive them to segregate, perhaps aided by the imprinted structural and super-structural features of the chromosomes (Jun and Mulder 2006; Jun and Wright 2010; Junier et al. 2014; Pelletier et al. 2012; Wiggins et al. 2010).

ParAB-*parS* systems may be useful rather than essential in bacteria that have just one chromosome, unless the bacterium is sporulating (Ireton et al. 1994) or going through a growth phase transition (Godfrin-Estevenon et al. 2002). If the microbe has more than one chromosome, then the partitioning system is essential if the segregation of its different chromosomes is to be properly coordinated, as, for example, in the case of *V. cholerae* (Yamaichi et al. 2007) or members of the Burkholderias (Passot et al. 2012). The roles of chromosome-encoded ParAB-*parS* systems as functioning partitioning machines was confirmed in early work where it was demonstrated that they could replace the native plasmid *par* systems on single copy episomes (Godfrin-Estevenon et al. 2002; Lin and Grossman 1998; Yamaichi and Niki 2000).

The *cis*-acting *parS* centromere-like sequences, and the genes that encode the ParA and ParB proteins, typically are found close to the *oriC* region of those bacterial chromosomes that harbour these systems (Livny et al. 2007; Reyes-Lamothé et al. 2012). This centromere positioning ensures that the first part of the chromosome to be duplicated is going to be delivered to the appropriate cellular site for the orientation of the segregation of the rest of the chromosome. Positioning varies and can be at mid-cell or quarter-cell (Fogel and Waldor 2005; Webb et al. 1997); in the case of chromosomes with *oriC* tethering at the poles of rod-shaped cells, it will be at the cell pole (Bowman et al. 2008; Fogel and Waldor 2006; Harms et al. 2013) (Figure 1.11). Dimeric ParB is in excess compared with its multiple 16-bp

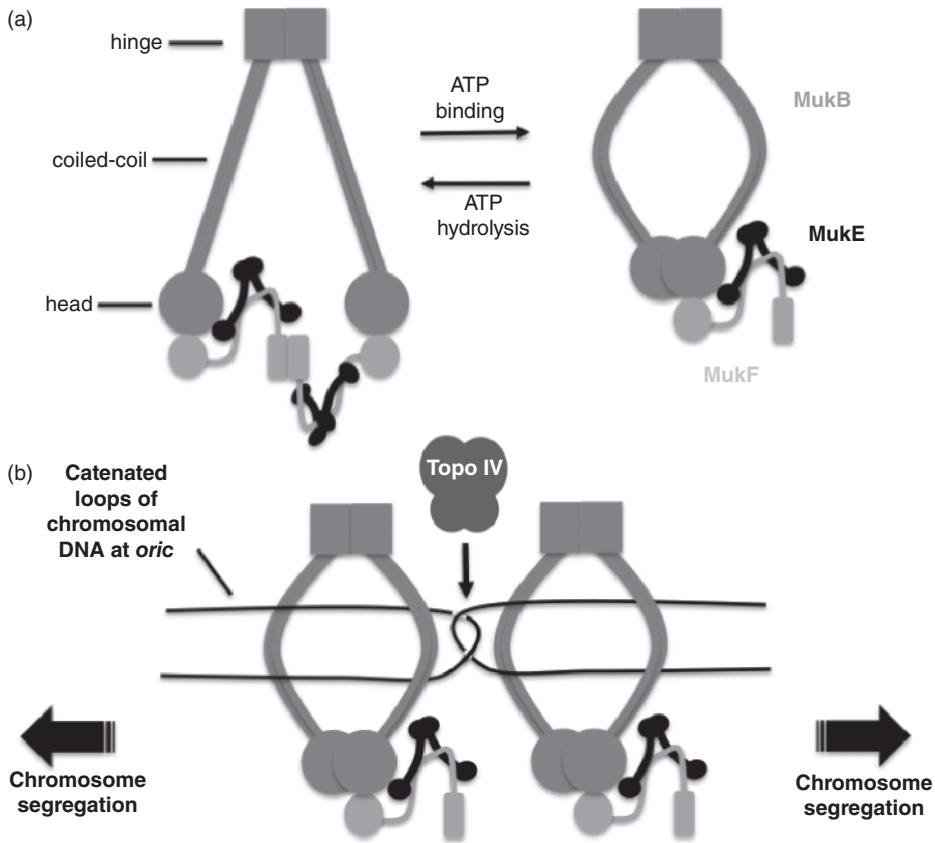


Figure 1.10 MukBEF structure and function. (a) The Dimeric MukB protein (intermediate grey) consists of a head domain with ATPase activity that is attached to a hinge domain by a coiled-coil. A ‘cap’ region in the head domain interacts with carboxyl terminus of MukF (light grey) while the central segment of MukF interacts with a MukE dimer (black). When ATP is absent, the complex has two copies of MukB, MukF and four of MukE. One MukF protein is displaced following ATP binding by MukB. This form of the MukBEF complex may in turn dimerise (not shown) (Nolivos and Sherratt 2014). (b) Clusters of MukBEF complexes, here represented by just two, have the potential to organise and segregate the *ori* regions of sister chromosomes following replication. The catenated *ori* regions are decatenated by Topo IV (dark grey), working in combination with the DNA segregation activity of the MukBEF motor.

parS binding sites and it seems that ParB can spread beyond *parS*, perhaps by a bridging mechanism that has the effect of folding the centromere region into a tightly organised complex (Graham et al. 2014; Sanchez et al. 2013).

The ParA protein drives bidirectional segregation of the *parS*-ParB complexes in an ATP-dependent manner. It can form filaments, and these have been proposed to be a factor in chromosome segregation (Bouet et al. 2007; Hui et al. 2010; Ptacin et al. 2010). However, it is also possible that ParA-driven chromosome segregation works by a diffusion-ratchet-type mechanism that has been described for its plasmid-encoded counterparts (Hwang et al. 2013; Vecchiarelli et al. 2014) or a *trans*-nucleoid relay mechanism (Lim et al. 2014).

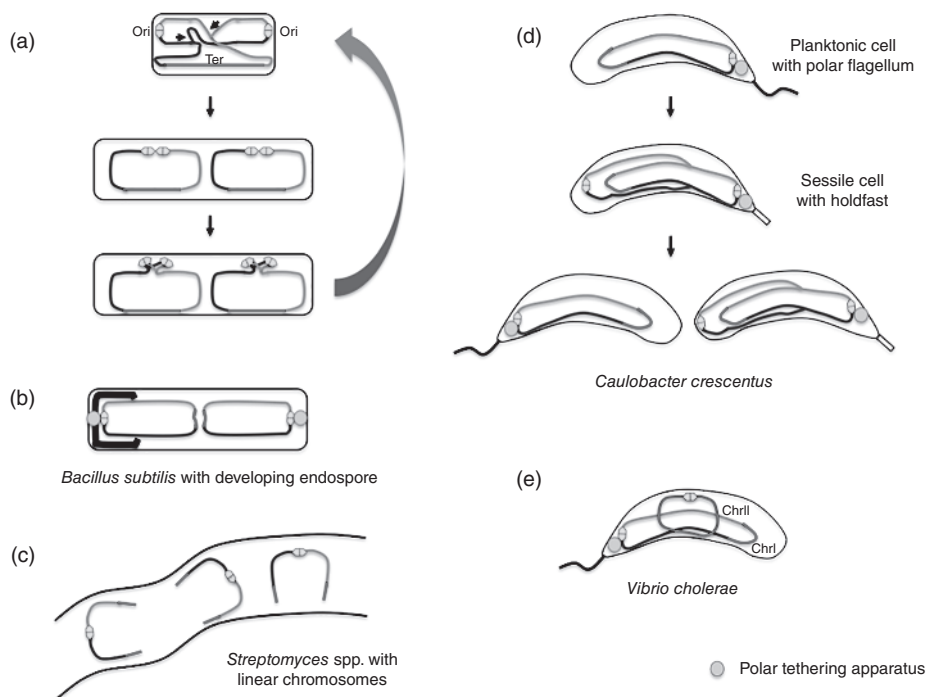


Figure 1.11 Spatiotemporal dispositions of chromosomes in model bacteria other than *E. coli*. (a) In vegetative cells of *Bacillus subtilis*, the origins of chromosome replication are at mid cell in newborn cells before moving to the poles. The origins are at quarter cell positions at the onset of cell division. (b) In sporulating cells, the origins are tethered at the poles as one chromosome copy (left) is segregated into the developing fore-spore. (c) *Streptomyces* species have linear chromosomes with the origin of replication at the midsection. (d) *Caulobacter crescentus* has two cell types: planktonic cells with a polar flagellum and sessile cells that adhere to the substratum via a polar holdfast. The origin of chromosome replication is tethered to the pole where the flagellum/holdfast is located. (e) *Vibrio cholerae* has two chromosomes. ChrI is the primary chromosome and its origin of replication is tethered to the cell pole where the single flagellum is located. The second chromosome, or chromid (ChrII), is plasmid-like and synchronises its replication with that of ChrI. (See colour plate section for colour representation of this figure)

Ter and associated genetic loci replicate at mid-cell and cohere for an extended period as a result of the *matS*-MatP nucleoprotein complex (Dupaigne et al. 2012; Reyes-Lamothe et al. 2008; Mercier et al. 2008; Stouf et al. 2013; Wu et al. 2019). The C-terminal domain of MatP interacts with the ZapB protein in the cell division apparatus and this interaction contributes to the extended cohesion of the Ter domain copies (Dupaigne et al. 2012; Espéli et al. 2012). MatP is displaced from *matS* by the action of the FtsK motor (or SpoIIIE in *B. subtilis*) as it drives the Ter domain copies into the nascent daughter cells (Deghorain et al. 2011; Graham et al. 2010; Marquis et al. 2008; Massey et al. 2006; Sherratt et al. 2010). FtsK uses its *dif*-oriented KOPS repeats to bind and to guide this process; the XerCD recombination *dif* site is the final component of the chromosome to be segregated (Stouf et al. 2013). The formation of FtsK hexamers, triggered by the onset of cell division, is a critical step in FtsK's own activation (Bisicchia et al. 2013) and its activation of the XerCD recombination

apparatus (Zawadzki et al. 2013), illustrating the exquisite integration of the Ter segregation and the chromosome dimer resolution systems.

1.10 Polar Tethering of Chromosome Origins

Anchoring the origin of replication to one pole of the cell is likely to assist in reinforcement of the *ori-Ter* orientation of the chromosome seen along the long axis of rod shaped cells and in ensuring that daughter cells receive an entire chromosomal copy at cell division (Badrinarayanan et al. 2015) (Figure 1.11). The PopZ protein fulfils this role in *C. crescentus* by forming a matrix at the pole and interacting with the ParB-*parS* complex at *oriC* (Bowman et al. 2008; Ebersbach et al. 2008). Displacement of *parS* to a different chromosome site interferes with this arrangement: while *parS* continues to be located at the pole *oriC*, from which *parS* is now disconnected, lies elsewhere in the cell (Umbarger et al. 2011).

The cytoplasmic protein HubP connects the origin of replication of ChrI to the cell pole in *V. cholerae*. The connection is made between HubP and the ParAI-ParBI-*parS* complex. In addition to its membrane location, the HubP protein is connected to the cell wall through a peptidoglycan-binding LysM motif, a feature that is required for its polar localisation (Yamaichi et al. 2012).

Polar attachment of the chromosome occurs in *B. subtilis* at the onset of sporulation. The RacA protein interacts with the DivIVA membrane protein that is located at the cell pole (Ben-Yehuda et al. 2003; Lenarcic et al. 2009; Oliva et al. 2010; Ramamurthi and Losick 2009; Wu and Errington 2003). RacA also binds to *ram* (RacA binding motifs) that are found in 25 copies at *oriC* (Ben-Yehuda et al. 2005). In the absence of RacA or DivIVA, sporulating bacteria fail to position the chromosome correctly and have the *oriC* at mid-cell. This misplacement leads to the production of prespore compartments without chromosomes (Ben-Yehuda et al. 2003). *B. subtilis* cells do not have their chromosomes attached to the cell pole during vegetative growth, although their origins occupy positions that alternate between pole-proximal and at quarter-cell, arrangements that require the cytoplasmic SMC complex (Wang, X., et al. 2014), just as the MukBEF equivalent in *E. coli* is required for that organism's chromosome to exhibit its customary *ori-Ter* orientation during rapid growth (Danilova et al. 2007).

1.11 Some Bacterial Chromosomes Are Linear

Most of the literature on bacterial chromosomes describes work with covalently closed, circular molecules. On the face of it, chromosome circularity is not essential for survival: work with *E. coli* has shown that linearisation of its circular chromosome through a phage-mediated process that leaves the ends closed by DNA hairpins does not interfere significantly with the life of the bacterium (Cui et al. 2007). Going in the other direction, the linear chromosome of *Streptomyces lividans* can be circularised without killing the microbe, although its genetic instability increases (Voff et al. 1997).

Some organisms have linear chromosomes naturally. For example, *Borrelia burgdorferi*, the spirochete and causative agent of Lyme disease, has a complex genome consisting of

a linear chromosome and 23 plasmids, some of which are circular while others are linear (Chaconas and Kobryn 2010). Essential metabolic functions are encoded by the plasmids, so these are parts of the core genome and not simply ancillary components. The ends of the linear DNA molecules are closed covalently by hairpin telomere-like structures (Barbour and Garon 1987). Such structures are not found widely in bacteria; other examples have been reported in the plant pathogen *Agrobacterium tumefaciens* and in some phage (Chaconas and Kobryn 2010; Slater et al. 2013). Telomere resolvases, enzymes that are related to the integrase family of tyrosine site-specific recombinases, promote fusions between the linear replicons at their telomeres, driving genome evolution (Huang et al. 2017). Replication of linear replicons in *Borrelia* spp. is thought to occur bidirectionally from a central origin, producing a double-stranded dimeric circle that is resolved by the telomere resolvase (ResT in *B. burgdorferi*) to produce two linear molecules with closed telomeres at their ends. Positive DNA supercoiling, probably arising from the local overwinding of the DNA during replication, assists telomere resolution (Bankhead et al. 2006). Although the role of DNA supercoiling, positive or negative, in linear replicons has not been studied comprehensively, there is some evidence that it is a factor in setting the level of transcription of promoters found on linear plasmids when those replicons are artificially circularised. This has led to the proposal that linear replicons may avoid instability caused by topological changes in circular molecules (Beaurepaire and Chaconas 2007). *Streptomyces* spp. also have linear chromosomes and linear plasmids, and intra-replicon interactions mediated by ‘terminal proteins’ that are covalently bound to the telomeres allow the creation of negatively supercoiled DNA circles from the linear replicons (Tsai et al. 2011). These negative supercoils are relaxed by DNA topoisomerase I, which is a component of the telomere complex in *Streptomyces*. It has been proposed that negative supercoiling is likely to be important for both DNA replication and for transcription, especially of genes located close to the telomeres (Tsai et al. 2011).

1.12 Some Bacteria Have More than One Chromosome

Among bacteria that have more than one chromosome are the well-studied organisms *A. tumefaciens* (Allardet-Servent et al. 1993), *Brucella* spp. (Jumas-Bilak et al. 1998), *Rhodobacter sphaeroides* spp. (Choudhary et al. 2007; Suwanto and Kaplan 1989), and *Vibrio* spp. (Val et al. 2014). Of the organisms listed here, *A. tumefaciens*, has one circular and one linear chromosome; the others have two circular chromosomes. *Paracoccus denitrificans* is a bacterium that has three chromosomes (Winterstein and Ludwig 1998).

Are all of the chromosomes in a multi-chromosome genome ‘equal’? It appears that one chromosome is usually the primary replicon, with the other being relegated to secondary chromosome status. For example, the two chromosomes of the pathogen *V. cholerae* are designated chromosomes, ChrI and ChrII, with ChrI having the majority of the metabolically important and virulence-associated genes. ChrII does harbour essential genes, so its designation as a second chromosome is justified, despite its having plasmid-like features. For example, the origin of replication of ChrII shows structural features that are similar to those of plasmids, which is consistent with the secondary chromosome having evolved from an ancestral plasmid (Orlova et al. 2017). The plasmid-like nature of ChrII is also emphasised

by its encoding plasmid RK2-like toxin systems that ensure post-segregation killing of those *V. cholerae* cells that lose ChrII (Yuan et al. 2011). The replication initiation system of ChrI resembles that of *E. coli*: initiation of replication of each chromosome is independent of, but coordinated with, that of the other (Duigou et al. 2006). ChrII begins replicating later in the cell cycle than the larger ChrI, but both finish together. This has been interpreted as a mechanism that compensates for the differences in size of the two molecules and the need to end replication simultaneously so that the chromosomes can be segregated together and that any dimers can be resolved simultaneously (Rasmussen et al. 2007). Further investigation has revealed the generality of coordinated termination of replication in members of the *Vibrionaceae* with two chromosomes (Kemter et al. 2018). ChrI and ChrII each possess their own ParAB-*parS* systems and use these for efficient segregation of chromosome copies at cell division (Fogel and Waldor 2005; Yamaichi et al. 2007, 2011).

1.13 Plasmids

In many bacteria, autonomously replicating and segregating genetic elements called plasmids accompany the chromosome in the cell. Like most bacterial chromosomes, plasmids are usually covalently closed, circular DNA molecules, but this is not always the case: some are linear. Certain plasmids are categorised as additional chromosomes (or ‘chromids’) due to their size, their carriage of genes normally found on bona fide chromosomes, their unitary copy number, and/or the coordination of their replication and segregation with the main chromosome (Barloy-Hubler and Jebbar 2008; Fournes et al. 2018). Other very big plasmids are called ‘mega-plasmids’ and can encode functions required for symbiosis or virulence (Schwartz 2008). In general, plasmids carry genes that are useful rather than essential, so their loss is not usually fatal to the cell; in contrast, loss of the chromosome is fatal.

Plasmids came to attention due to their involvement in bacterial sex (the Fertility, or F factor) and when it was discovered that they carried genes for resistance to antimicrobial agents, including antibiotics (R factors). Investigations of these phenomena led to the discovery of plasmid conjugation and the existence of other mobile genetic elements such as transposons and integrons. Plasmid studies revealed a wealth of information about plasmid replication processes, segregation systems, and copy number control mechanisms. This field also provided cloning vectors to support the emergence of the recombinant DNA technology that, in part, underpins biotechnology. Plasmid research has provided important insights into gene regulation mechanisms, including the provision of early examples of the regulatory roles of small RNA molecules.

1.14 Plasmid Replication

The term plasmid was introduced in the mid-twentieth century to describe self-replicating extrachromosomal DNA elements (Lederberg 1952). Self-transmissible plasmids are important contributors to HGT; indeed, some plasmids can even participate in DNA transfer between different domains of life (Suzuki et al. 2008; Zambryski et al. 1989). Limits to the

host range of a plasmid include barriers to mating bridge establishment, to successful DNA transfer, and to successful plasmid replication (Jain and Srivastava 2013). Plasmid size is a poor predictor of host range: while many broad-host-range plasmids are large, many others are quite small. For example, pBC1 can function in both Gram-negative and Gram-positive bacteria, yet it is only 1.6 kb in size (De Rossi et al. 1992). At the other end of the scale, the intensively studied RK2/RP4 plasmid group is in the 60-kb-size range but is confined to Gram-negative hosts, albeit a wide selection of these (Thomas et al. 1982). It is an advantage to have several origins of plasmid replication as this improves the chances of being able to replicate in a given host. However, the presence of multiple origins is not in itself a reliable indicator of broad host range: the F plasmid has a narrow range yet it has three origins of replication. Instead, it is the structure of the origin(s) that seems to be important in determining host range. Plasmids with a minimum dependence on host-encoded factors for replication are likely to have a broad host range; for example, RSF1010 from incompatibility group Q (IncQ) uses a strand-displacement mode of replication that relies on no host-encoded factors for the initiation of DNA synthesis (Loftie-Eaton and Rawlings 2012).

The RK2 plasmid (IncP) has a broad host range and can replicate in *E. coli* or *Pseudomonas aeruginosa* (Shah et al. 1995). It can replicate in *E. coli* using just a minimal origin, *oriV* ('vegetative' origin), containing five iterons (directly repeated sequences) and four binding sites for DnaA (Doran et al. 1999) (Figure 1.12). In *P. aeruginosa*, RK2 needs an additional three iterons but can dispense with the DnaA boxes (Schmidhauser et al. 1983). RK2 is also capable of replicating without DnaA when in *C. crescentus* (Wegrzyn et al. 2013).

Iterons are the binding sites for plasmid-encoded Rep (replication) proteins. The Rep proteins are needed to initiate plasmid DNA replication, but they can also inhibit this process. In addition, Rep proteins have roles as transcription regulators, acting as auto-repressors. They are also subject to turnover by host-encoded proteases (Konieczny et al. 2014). The positions, numbers, orientations, lengths, DNA helical phasing, and spacer lengths of iterons vary from plasmid to plasmid and making alterations to any of these details within a particular plasmid typically has a negative effect on replication (Konieczny et al. 2014). In RK2, the Rep protein (called TrfA) and DnaA both target the origin, which has an A+T-rich DUE adjacent to the iteron arrays. Some iteron plasmids (but not RK2, which requires HU) have a requirement for IHF binding to the origin for efficient initiation of replication (Shah et al. 1995). Sites for Dam methylation (5'-GATC-3') and SeqA binding are also features of some iteron-dependent origins (Brendler et al. 1995). SeqA binding blocks replication initiation by excluding the replication proteins. A GC-rich sequence motif is located adjacent to the DUE in RK2, but its significance is unclear (Figure 1.12). Other iteron origins have requirements for the FIS NAP and the IciA protein, an inhibitor of DNA unwinding at the DUE.

Many of the *cis* or *trans*-acting components of iteron origins, and their architectures, are reminiscent of *oriC* on the chromosome. The TrfA replication protein of RK2 interacts with, and recruits, the DnaB helicase. The ability of a plasmid replication protein to recruit a host helicase may be a determining factor limiting plasmid host range (Zhong et al. 2005). TrfA also acts with Hda, the inhibitor of DnaA activity, to prevent over-initiation of RK2 replication (Kim et al. 2003). It has been suggested that TrfA has a motif that is suitable for interaction with the β -clamp of *E. coli* DNA Pol III (Kongsuwan et al. 2006). In

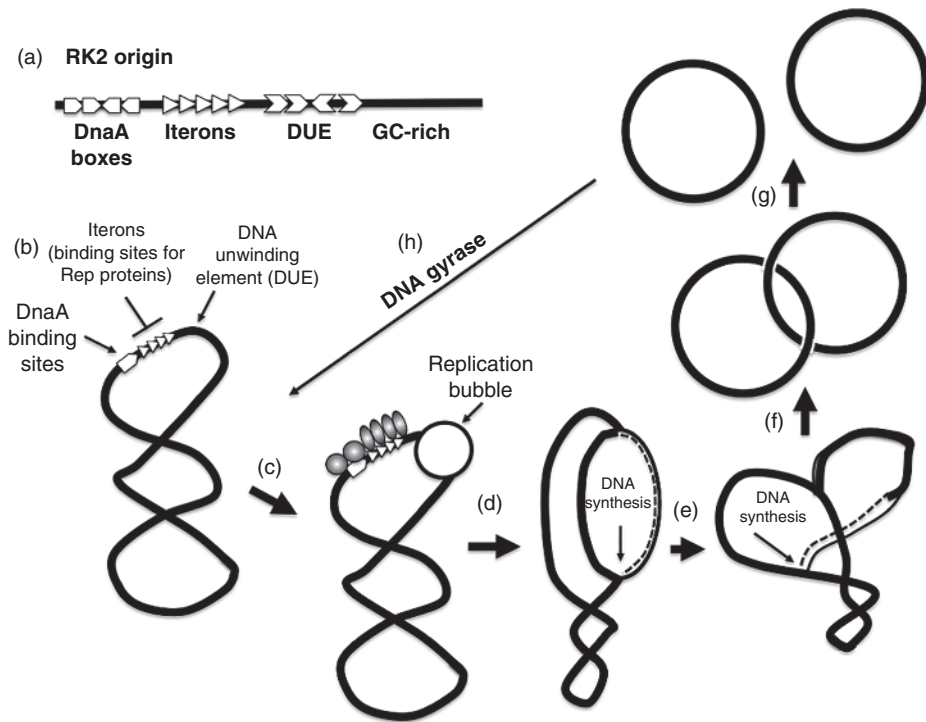


Figure 1.12 Theta model of plasmid replication. (a) The structure of the origin of replication in the broad-host-range, single-copy, IncP plasmid RK2, showing the relative positions and numbers of the DnaA boxes, the iterons (binding sites for the replication protein, TrfA), the AT-rich DNA unwinding element (DUE), and the adjacent GC-rich sequence. (b) A replication cycle is shown for an idealised plasmid using theta replication. Replication begins with the binding of the replication protein to the iterons and the recruitment of the host-encoded DnaA to the DnaA boxes. (c) The DNA in the DUE becomes single-stranded, creating a replication bubble to which host-encoded replication proteins are recruited. (d, e) Depending on the plasmid, DNA synthesis proceeds either uni- (e.g. ColE1) or bidirectionally (e.g. RK2). (f) The products of plasmid replication are catenated, double-stranded circles and these are unlinked by Topo IV. (g, h) The unlinked plasmids are negatively supercoiled by DNA gyrase, recreating the substrate for another round of replication.

addition to DnaB, iteron-based initiation also requires DnaC (in *E. coli*), the DnaG primase, DNA gyrase, the Pol III holoenzyme, and the SSB, as is also seen at *oriC* (Section 1.3). Unlike initiation of chromosome replication at *oriC*, initiation of plasmid DNA replication at iteron origins is ATP-independent; there is no requirement for the DnaA-ATP form of DnaA (Konieczny et al. 2014).

An important mechanism of copy number control in iteron plasmids is ‘handcuffing’, where the monomeric Rep proteins bound to iterons on two plasmids dimerise, bridging the two replicons (Das and Chattoraj 2004). Handcuffing may be counteracted by the DnaJ-DnaK-GrpE protein chaperone triad, which can convert the Rep dimers to monomers (Toukdarian and Helinski 1998). The Rep proteins downregulate replication initiation through the auto-repression of their own genes (Kelley and Bastia 1985). The level of active Rep proteins in the cell is controlled by proteolysis and protein chaperones: monomers

are active and dimers are inactive in promoting replication (Konieczny et al. 2014). Active Rep proteins are a proxy for the number of plasmid copies in the cell and allow feedback through dimer formation to downregulate Rep activity, and hence to inhibit the synthesis of new plasmid copies. Antisense RNAs also play important roles in modulating plasmid replication through their ability to interfere with the expression of plasmid-encoded replication factors (Brantl 2014).

RK2 and its plasmid relatives use a theta model of DNA replication (Figure 1.12). Plasmid ColE1, the backbone for many cloning vectors used in recombinant DNA technology, also uses theta replication, but differs from RK2-like plasmids in relying on host factors to open the double-stranded origin and to prime synthesis (Lilly and Camps 2015; Wang et al. 2004). DNA duplex unwinding is driven by transcription of a stable RNA pre-primer that forms an R-loop in the *ori* region of ColE1. This process is driven by negative supercoiling of the plasmid DNA. RNase H then processes the bound RNA to generate the primer RNA that is then extended by Pol I. This marks the beginning of leading strand DNA synthesis. As the newly synthesised DNA strand makes progress through the plasmid DNA duplex, it base pairs with the template to create a D-loop that recruits PriA. Pol III takes over leading strand synthesis and initiates the synthesis of the lagging strand; the converging replisomes continue moving until they are at or near the termination site *terH* (Nakasu and Tomizawa 1992). Gaps between the strands are then filled in by Pol I (Troll et al. 2014).

Rolling circle replication allows plasmids to replicate independently of chromosomal DNA (Khan 2005). The process relies on a nick made by a plasmid-encoded initiator protein in one plasmid DNA strand, providing a primer for leading strand initiation and a lagging strand origin (Figure 1.13). No RNA primer is required. Rolling circle replication is chiefly found in plasmids from Gram-positive bacteria, although it does occur in replicons from Gram-negatives and archaea (del Solar et al. 1993; Ruiz-Masó et al. 2015). The process replicates the leading strand and the lagging strand in two separate steps. Leading strand replication begins with the nicking of the double-strand origin (*dso*) by a plasmid-encoded replication protein, Rep. This is a member of the HuH superfamily of DNA endonucleases (Chandler et al. 2013) and it has a binding site located adjacent to the *dso* that positions it appropriately to cut the DNA. The DNA to be cleaved is presented to Rep in a single-stranded form within a stem-loop structure that extrudes from the negatively supercoiled plasmid. This extrusion event is Rep-binding-dependent (Ruiz-Masó et al. 2007). Rep forms a covalent bond with the cleaved DNA through an active site tyrosine (Noirot-Gros et al. 1994; Thomas et al. 1990). Host DNA polymerases use the intact template strand to guide DNA synthesis while simultaneously displacing the non-template strand. The displaced strand is coated with SSB and is ejected as a covalently closed, single-stranded circle at the end of leading strand synthesis. This single-stranded circle is then used as the template for lagging strand synthesis, a process that involves only host-encoded proteins (especially RNA polymerase and DNA polymerase I) and initiates at a structured region in the circle known as the single-strand origin, *ssu* (del Solar et al. 1987; Gruss et al. 1987; Kramer et al. 1997). Control of rolling circle replication is achieved principally through the control of Rep protein production. For this reason, the expression of the *rep* gene is strictly regulated, typically via mechanisms that employ an antisense RNA or an antisense RNA working with a DNA-binding regulatory protein. The first type operates through transcriptional attenuation while the second involves protein-mediated transcriptional repression

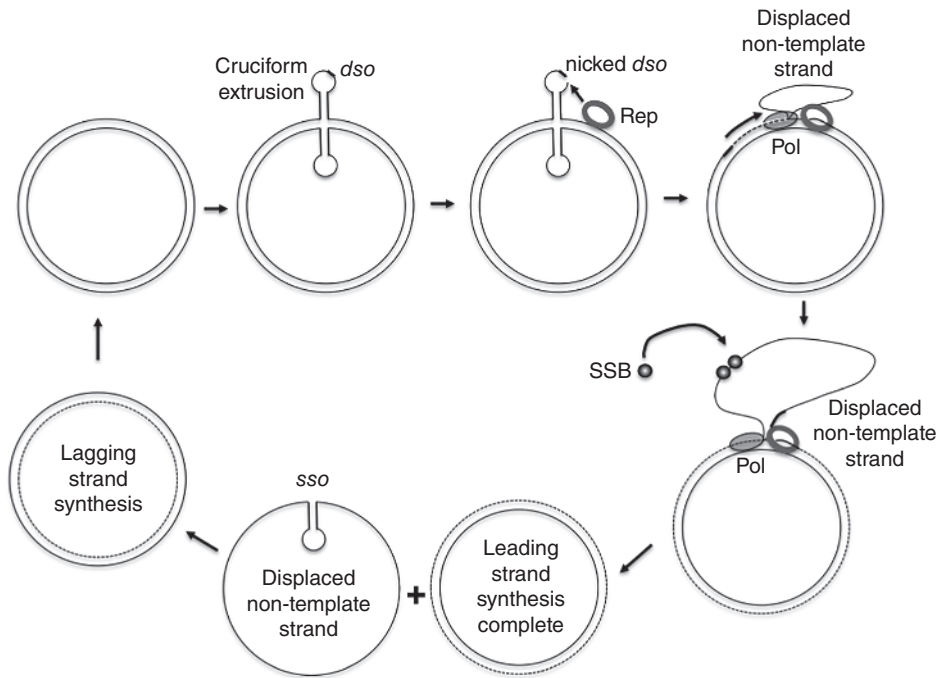


Figure 1.13 Rolling circle plasmid replication. A circular plasmid using the rolling circle mechanism of replication is shown at top left. The double-stranded circle is shown in a topologically relaxed state, but it would be negatively supercoiled in the bacterium, a state that encourages extrusion of a cruciform that contains the double-strand origin (*dso*), the site of replication initiation. The *dso* is represented by a slightly thicker line in the drawing. Extrusion of the cruciform presents the *dso* in single-stranded form to the plasmid-encoded replication protein, Rep. The Rep protein is positioned appropriately by binding to a recognition site on the plasmid adjacent to the *dso*. The bacterial DNA polymerase use the 3'-OH at the nick to prime DNA synthesis; no RNA primer is required. A dotted line represents the newly synthesised DNA and an arrow next to this line shows the direction of DNA synthesis. The plasmid duplex unwinds as DNA synthesis proceeds, displacing the non-template DNA strand, which is then coated by the single-stranded DNA-binding protein, SSB. A full round of replication displaces the non-template strand completely, producing a double-stranded plasmid (with one newly synthesised strand) and a single-stranded circle. This circle is used as the template for the synthesis of the lagging strand. Host proteins exclusively conduct lagging strand synthesis (especially RNA polymerase and DNA polymerase I), a process that begins with priming by RNA polymerase via RNA synthesis at the single-strand origin, *sso*.

backed up by translational inhibition using a *trans*-acting RNA (Brantl 2014; del Solar and Espinosa 2000; Novick et al. 1989).

1.15 Plasmid Segregation

Plasmids employ two strategies to ensure segregation of their copies at cell division: active partitioning mechanisms (low copy number plasmids) and reliance on dispersal through the cytosol of the mother cell to ensure that some copies end up in each daughter

cell (high copy number plasmids). A third strategy acts post-segregationally. It is based on toxin–antitoxin systems and eliminates those bacterial cells that do not acquire a plasmid copy at cell division (Hayes 2003) (see also Sections 2.30 and 2.35). The plasmid encodes both a stable toxin and an unstable antitoxin in an operon known as an addiction module: maintenance of the supply of the antitoxin requires the continued presence of the plasmid and the antitoxin-encoding gene. Bacteria that become liberated from the burden of plasmid-carriage may outgrow their plasmid-carrying counterparts. Eliminating the plasmidless bacteria helps to prevent extinction of the plasmid carriers by the fitter, plasmid-free, segregants. The CcdA/B antitoxin/toxin pair produced by the F plasmid provides an example of this post-segregational killing strategy. CcdB inhibits DNA gyrase by trapping it in a cleavage complex with DNA, and CcdB activity is neutralised when it binds the unstable CcdA antidote. A bacterium that loses the F plasmid will retain the toxic CcdB molecule after the unstable CcdA molecule has been broken down by the ATP-dependent Lon protease (van Melder et al. 1994). The resulting poisoning of DNA gyrase by CcdB kills the plasmidless cell.

High copy number plasmids, i.e. those with 10 or more copies per cell, lack genes that are capable of encoding active partitioning machinery (Million-Weaver and Camps 2014). Random distribution of the multicopy plasmid through the cytosol seems to account for the faithful inheritance of plasmids such as ColE1 (Durkacz and Sherratt 1983). The formation of plasmid multimers poses a risk because this process reduces the copy number of independently segregating units, but multimer resolution systems such as *cer*/XerCD in ColE1 provide a potent antidote (Summers and Sherratt 1984). This resolution mechanism is a close relative of the chromosomal *dif*/XerCD system, albeit with additional co-factors (Section 1.8). Plasmid distribution in the cytosol is likely to be influenced by the presence of other molecules and structures, not least the nucleoid, and nucleoid exclusion does seem to be a factor in confining plasmids to the space just inside the cytoplasmic membrane (Reyes-Lamothe et al. 2014; Wang et al. 2016; Yao et al. 2007). The plasmids occur in clusters and frequently these clusters are seen at the poles of the cell; clusters are dynamic, they can divide, with some sub-clusters relocating to the mid-cell (Yao et al. 2007). The introduction of another type of plasmid produces an even more complex clustering pattern (Diaz et al. 2015; Yao et al. 2007). It appears that multicopy plasmids move between existing in clusters and being alone, and that these forms diffuse randomly within the confines imposed by the nucleoid and other cell structures (Wang 2017).

Low-copy number plasmids cannot rely on strategies based on random spatial distribution to ensure their segregation to the daughter cells at division. These plasmids have active partitioning systems, systems that have counterparts in the chromosomes of many bacteria (but not *E. coli*). These partitioning (Par) systems consist of two proteins, ParA and ParB, and a centromere-like DNA site called *parS* (Baxter and Funnell 2014; Gerdes et al. 2010). The ParB protein binds to *parS* and ParA interacts with ParB, hydrolysing ATP or GTP to provide the energy needed to drive the partitioning process.

Plasmid Par systems, such as those in the single-copy F plasmid or the P1 prophage plasmid, whose ParA protein has a Walker-type ATPase motif, use the surface of the nucleoid as a scaffold over which plasmids are actively moved. The mechanism is termed a diffusion-ratchet, with ParA diffusing over the nucleoid and ParB binding to the *parS* sequence on the plasmid to form the partition complex (Vecchiarelli et al. 2013, 2014).

ParA-ParB interaction triggers ATP hydrolysis by ParA, denuding the nucleoid surface in the vicinity of the plasmid *parS*-ParB complex of active ParA. This depletion effect creates a ParA gradient across the nucleoid surface, moving the *parS*-ParB complex (and the plasmid) along the gradient. With two daughter plasmids in play, the effect of ParA depletion and the associated gradients is to move the two plasmids away from each other, segregating them into the two daughter cells. This diffusion ratchet mechanism has replaced earlier hypothetical models of ParA-ParB-*parS* segregation systems that were based on ParA assembly into cytoskeletal filaments (Brooks and Hwang 2017).

The R1 drug-resistance single-copy plasmid has a ParMRC partitioning system that consists of a centromere-like *parC* site, an adaptor protein ParR that binds to *parC* and an actin-like ATPase, ParM. ParM forms filaments that grow bidirectionally, with a ParR-*parC* complex on either end. As the filament grows in length, the plasmid copies are separated. ParM searches the cell for ParR-*parC* complexes, the complexes stabilise ParM filaments whose dynamic instability requires ATP hydrolysis; the stabilised filaments grow, pushing *parC*-containing plasmids to opposite ends of the cell (Garner et al. 2004, 2007). The TubZ-TubR-*tubZ* partitioning system found in many plasmids in *Bacillus* spp. (e.g. *B. thuringiensis*) differs from ParMRC in that the TubZ filament grows unidirectionally by recycling TubZ subunits from the leading edge to extend the trailing edge ('treadmilling') and uses GTP hydrolysis to form the filament (Fink and Löwe 2015; Larsen et al. 2007).

1.16 The Nucleoid

A defining characteristic of prokaryotes is that they do not possess a membrane-bound nucleus. Instead, prokaryotes have a nucleoid, a body within the cytoplasm that contains the genetic material but lacks a surrounding membrane (Piekarski 1937). The nucleoid is composed of the chromosome and associated molecules including RNA polymerase, DNA polymerase, DNA-binding proteins, and RNA molecules (Dorman 2014b; Macavín and Adhya 2012). In electron micrographs of thinly sectioned bacteria, the nucleoid can be seen as an amoeboid shape surrounded by the electron-dense ribosomes within the cytoplasm (Kellenberger 1952; Robinow and Kellenberger 1994). Staining of the DNA with 4', 6-diamidino-2-phenylindole (DAPI) has confirmed the presence of a zone around the nucleoid in *E. coli* and *B. subtilis* where translation can take place (Mascarenhas et al. 2001). While even more sophisticated imaging has improved our knowledge of the structure of the nucleoid, it has taken a multi-pronged approach using a variety of techniques over several decades to bring about our current (but still incomplete) understanding of the bacterial nucleoid.

The chromosome is packaged within the bacterial cell in a conformation that permits gene expression and DNA replication to proceed. The 4.6 Mb circular chromosome of *E. coli* strain MG1655 has a circumference of 1.5 mm and, if opened out fully, a diameter of approximately 0.5 mm. In contrast, the bacterial cell is approximately 2 µm in length, 1 µm in diameter and has a volume of 1 fl, or 1×10^{-15} l (Dorman 2013; Kubitschek and Friske 1986). Understanding how the need to package the DNA efficiently is reconciled with the requirements of DNA replication, gene transcription, DNA recombination, and DNA repair is a major goal of research into the structure of the nucleoid.

1.17 The Chromosome Has Looped Domains

Examination of electron micrographs of thinly sectioned *E. coli* cells gives few clues as to the fine detail of chromosome organisation in the nucleoid. The images of chromosomes extruded from gently lysed *E. coli* obtained by Cairns (1963a,b) using autoradiography hint at a subdivision of the chromosome into looped, supercoiled domains. The nature of the domain boundaries was obscure but seemed to be associated with RNA (Kavenoff and Bowen 1976). Analysis with electron microscopy suggested that the chromosome was subdivided into between 12 and 80 supercoiled loops (Delius and Worcel 1974).

Sinden and Pettijohn (1981) used photobinding of trimethylpsoralen to intracellular DNA to estimate the number of independently looped domains. This agent binds to duplex DNA at a rate that is proportional to the superhelical tension in the DNA (Sinden et al. 1980). By estimating the number of gamma-radiation-induced nicks that were required to release most of the superhelical tension, an estimate of the number of topologically independent domains was obtained. For *E. coli* growing with a generation time of 30 minutes, it was estimated that the chromosome is divided into between 33 and 53 independent domains (or between 90 and 150 domains per nucleoid) (Sinden and Pettijohn 1981). The existence of independent domains is an important concept in nucleoid architecture: genes in one domain may be isolated from DNA topological changes taking place in other domains, making gene location along the chromosome significant for reasons other than differences in copy number arising as a result of gene distance from *oriC*.

More refined measurements of domain number in the *E. coli* chromosome or that of its close relative, *Salmonella enterica* serovar Typhimurium, have been made by exploiting site-specific recombination reactions that can take place within but not between domains, by counting the number of looped domains in multiple images of extruded chromosomes, and by releasing superhelical tension from domains using restriction enzymes and observing the distance over which an effect on the transcription of a supercoiling-sensitive gene can be exerted (Postow et al. 2004; Stein et al. 2005). The results from these experimental approaches indicate that the chromosome is subdivided into about 400 independent domains in *E. coli* cells during exponential growth, with each being approximately 12–14 kb in length. The boundaries between the domains do not seem to be fixed and fewer of them are found in bacteria that have entered stationary phase (Staczek and Higgins 1998).

1.18 The Macrodomain Structure of the Chromosome

The bacteriophage lambda integrase-mediated site-specific recombination system has been exploited in studies of nucleoid organisation in *E. coli* and *Salmonella* (Garcia-Russell et al. 2004; Valens et al. 2004). Recombination between copies of the lambda attachment site requires physical contact between the sites and these can be created by random collision (Crisona et al. 1999; Dorman and Bogue 2016). Sites placed at different distances from one another around the chromosome can be assessed for interaction frequency, providing an estimate of the frequency of contact between different parts of the chromosome. At the same time, regions of the chromosome that rarely interact have also been discovered. This analysis led to the proposal that the chromosome is divided into a small number of

large territories called macrodomains (Valens et al. 2004) (Figure 1.1). *E. coli* and its close relatives have four macrodomains (Ori, Left, Ter, and Right) and two non-structured (NS) regions: NS-Left and NS-Right (Cameron et al. 2017; Jiang et al. 2015; Thiel et al. 2012). The NS domains are determined by their proximity to the Ori macrodomain: any region that is placed next to Ori acquires the features of an NS domain (Duigou and Boccard 2017).

1.19 The Chromosome Displays Spatial Arrangement Within the Cell

The bacterial chromosome is oriented in the cytoplasm with reference to the poles and the midpoint of the cell (Figure 1.9). This positioning is important for the successful segregation of the daughter chromosomes at cell division and it assigns domains to specific regions of the cytoplasm. Experiments using fluorescence *in situ* hybridisation (FISH) in *E. coli* show that in newborn cells, Ori and Ter are located at the cell poles and that genetic loci in between follow approximately the gene order along the chromosome (Niki et al. 2000). Before the initiation of chromosome replication, Ori and Ter move to the mid-cell position. Chromosomes with gross rearrangements of the macrodomains misplace Ori and Ter, leading to the appearance of anucleate daughter cells at cell division (Niki et al. 2000).

FISH experiments involve the permeabilisation and fixing of the bacterial cells; these treatments may produce artefacts, making less intrusive approaches more desirable. Experiments in which arrays of binding sites for fluorescently labelled DNA-binding proteins are placed at specific chromosomal locations have allowed chromosome dynamics to be studied in live bacteria (Robinett et al. 1996). This approach has revealed that in slow-growing *E. coli* cells, the origin and terminus are found at mid-cell and the repli-chores are located in each cell half (Nielson et al. 2006b; Wang et al. 2006) (Figure 1.9). In fast-growing *E. coli* cells, the Ori copies are found at the cell poles while the Ter region is recovered from the pole to the mid-cell position as the replication forks close in (Youngren et al. 2014) (Figure 1.9). This situation is reminiscent of that described in *B. subtilis* for the Ori and Ter regions during chromosome replication (Teleman et al. 1998; Webb et al. 1997). *C. crescentus* has also served as an important model organism in nucleoid structure studies and its chromosome has been observed as adopting an arrangement in which Ori is located at one cell pole and Ter at the other (Viollier et al. 2004) (Figure 1.11). Analysis of *C. crescentus* using ‘carbon copy chromosome conformation capture’ (5C) has shown that the repli-chores lying in between the cell-pole-located Ori and Ter regions are intertwined (Umbarger et al. 2011).

Methods based on different forms of chromosome conformation capture offer the possibility of generating high-resolution data of the co-location of different parts of the chromosome at different points in time. This offers the possibility of assembling a four-dimensional map of the bacterial genome. In contrast, investigations that rely on arrays of binding sites for fluorescently labelled DNA-binding proteins are limited to information about just those few positions where the arrays have been placed. There may also be technical difficulties that arise if the binding site copies in the array recombine; shortening array length and so reducing the strength of the fluorescent signal as fewer labelled proteins bind. The nucleo-protein complexes at the arrays may impede replication fork passage or interfere with local

transcription or the diffusion of DNA supercoils (Le and Laub 2014). It is also possible that the bound proteins may interact with one another, bridging together sister chromosomes prior to chromosome segregation (Mirkin et al. 2014).

The movement of the Ori and Ter domains within the *E. coli* cytosol is thought to reflect the separation of sister chromosomes that are connected along their length (Bates and Kleckner 2005). The molecular nature of the inter-sister connections is unknown (Kleckner et al. 2014). The final phase of the replication process copies the Ter region, setting the stage for chromosome segregation (Figure 1.9). In fast-growing bacteria, another round of chromosome replication will have started by the time this stage is reached (Youngren et al. 2014). The *E. coli* pattern of chromosome domain positioning with Ori and Ter at mid-cell is not universal: *C. crescentus* has its Ori and Ter regions in opposite poles of the cell while *B. subtilis* oscillates between the two modes (Figure 1.11). The differences are thought to reflect different chromosome segregation mechanisms (Wang et al. 2014).

Decatenation of the interlinked chromosome copies by topoisomerase IV and the resolution of any chromosome multimers by XerCD are needed prior to sister chromosome segregation (Hiraga 1993). In the absence of Topo IV, XerCD-*dif*-FtsK can achieve the same outcome by a process of local reconnection involving multiple rounds of site-specific recombination (Grainge et al. 2007). Finally, the segregation process will move one chromosome, together with any associated live replication forks, into one of the daughter cells. In *E. coli* and its close relatives, this occurs without the aid of a dedicated protein-based active partitioning system equivalent to the ParAB proteins and the *parS* *cis*-acting partitioning DNA sequence that are found in most other bacteria (Badrinarayanan et al. 2015; Bignell and Thomas 2001). Radial confinement of the two sister chromosomes has been proposed as playing a role in segregation in the case of *E. coli*. Here the two chromosome polymers repel one another through an entropic exclusion mechanism that drives the copies into separate compartments before the closure of the cell division septum (Jun and Wright 2010; Junier et al. 2014).

The specificity of chromosome orientation within the cytoplasm has led to the interesting proposal that the chromosome provides the prokaryotic cell with an internal frame of reference, something that has been lost in eukaryotes because the genetic material there is in a membrane-enclosed nucleus (Theriot 2013). This reference frame is useful in providing each molecule in the cell with a set of spatial coordinates. Developing the point further, Theriot has proposed that eukaryotes rely on their cytoskeleton to provide a reference frame (Theriot 2013). We will return to the issue of spatial and temporal positioning of molecules (Chapter 8).

1.20 SeqA and Nucleoid Organisation

DNA-binding proteins play important roles in organising the structure of the folded chromosome, with some of these proteins having a chromosome-domain-specific binding pattern (Dame et al. 2011). The SeqA protein was introduced during the description of factors involved in the control of the initiation of chromosome replication (Section 1.3). SeqA accompanies the moving replication fork (Brendler et al. 2000; Onogi et al. 1999), resulting in a dynamic pattern of binding around the chromosome (Sánchez-Romero et al.

2010; Waldminghaus and Skarstad 2010). This protein seems to be excluded from binding within the Ter macrodomain, possibly reflecting the absence from Ter of matches to the consensus sequence for SeqA high-affinity binding sites (Sánchez-Romero et al. 2010; Waldminghaus and Skarstad 2010). The SeqA protein can interact with the cell envelope as well as hemimethylated DNA (D’Alençon et al. 1999; Mika et al. 2015; Shakibai et al. 1998; Slater et al. 1995) so may it play a role in the positioning of Ori during cell division (Dame et al. 2011).

1.21 MukB, a Condensin-Like Protein

The bacterial chromosome is maintained in an orderly superstructure to facilitate replication, transcription, and other DNA-based transactions. The family of SMC proteins play an important role in achieving this organisation (Uhlmann 2016). These large polypeptides have a DNA-binding head domain and long coiled-coil domains that bring the head-domain-DNA complexes together in a condensed nucleoprotein complex (Figure 1.10). The head domains have ATPase activity and a DNA-binding hinge region in the coiled-coil domain promotes dimer formation (Chen, N., et al. 2008; Kumar et al. 2017a). SMC activity is found in eukaryotes and in prokaryotes, with one of the best-studied examples of an SMC-like protein in bacteria being the MukB protein from *E. coli* (Niki et al. 1991; Rybenkov et al. 2014).

MukB forms topologically stable loops in the chromosomal DNA and protects the supercoils in these protected loops (Kumar et al. 2017a). It forms a complex with the MukE and MukF proteins, with these seeming to play a role in the turnover of the MukB complex on DNA in combination with the ATPase activity of MukB itself (Kumar et al. 2017a). MukF performs a bridging role between the ATPase heads of the two MukB in the complex (Figure 1.10). Proteins performing this task in SMC complexes are called kleisins. The equivalent system in *B. subtilis* and *C. crescentus* consists of the proteins Smc (MukB), ScpA (MukF, kleisin), and ScpB (MukE); the ‘Scp’ designation indicates that the protein is a ‘segregation and condensation protein’ (Britton et al. 1998; Burmann et al. 2013; Jensen and Shapiro 1999, 2003; Mascarenhas et al. 2005).

The MukBEF complex has important architectural and segregational roles in the nucleoid, operating mainly outside the Ter macrodomain of the chromosome. Its principal site of action seems to be at *ori* and MukB requires MukE, MukF, and ATP hydrolysis to gain and maintain this association; MukBEF/SMC complexes do not seem to track moving replisomes (Badrinarayanan et al. 2012a,b; Danilova et al. 2007; Gruber and Errington 2009; Sullivan et al. 2009). MukBEF is responsible for guiding the newly replicated *ori* regions into the two halves of the cell, driving bipolar segregation; if MukBEF is removed, the *ori* shifts from mid-cell to the pole, disturbing normal chromosome orientation and segregation patterns (Danilova et al. 2007). Inside the Ter macrodomain, the MatP protein prevents MukBEF from playing a structural role by displacing it and so making it available for *ori* binding (Lioy et al. 2018; Nolivos and Sherratt 2014; Nolivos et al. 2016). The MukBEF complex is not required for sister chromosome cohesion because *muk* mutants have a higher degree of cohesion of sister chromosomes than wild-type cells (Danilova et al. 2007).

Despite being able to stimulate Topo IV activity by direct interaction (Hayama and Mariani 2010; Hayama et al. 2013; Li et al. 2010; Vos et al. 2013), MukBEF forms a complex with Topo IV that seems to stabilise MukBEF on the DNA and drive chromosome condensation independently of the catalytic activities of the topoisomerase (Kumar et al. 2017b). This observation is indicative of an architectural role for the MukBEF-Topo-IV complex. Data from chromosome conformation capture experiments show that, together with the HU NAP, MukBEF promotes long-range DNA contacts in the megabase range (Lioy et al. 2018), adding further detail to our picture of the nucleoid-structuring contributions of MukBEF. However, the MukBEF-Topo-IV complex also has functions in chromosome segregation that do depend on the catalytic properties of Topo IV. The complex is responsible for the timely decatenation of newly replicated *ori* copies and their segregation (Figure 1.10), while also contributing to the management of DNA supercoiling set points (Zawadzki et al. 2015). The primary site of action of the MukBEF-Topo-IV complex has been regarded as *ori*, with MukBEF recruiting the topoisomerase to that region of the chromosome (Nicolas et al. 2014). It is becoming clear that MukBEF-Topo-IV has a much more dynamic relationship with the chromosome, both spatially and temporally, and that its relationship with the MatP protein in *ter* has an important management role in its choreography. The final stages of chromosome replication and the associated need to effect decatenation in a timely manner are consistent with a need to grant MukBEF-Topo-IV access to *ter* to perform the necessary cohesion/decatenation steps, with MatP then evicting the complex, making it available for re-association with *ori* (Nolivos et al. 2016).

1.22 MatP, the *matS* Site and Ter Organisation

The 17-kDa MatP protein binds to 23 high-affinity *matS* sites found exclusively within the Ter macrodomain of the *E. coli* chromosome (Mercier et al. 2008; Thiel et al. 2012). The Ter domains of both chromosome I and chromosome II of *V. cholerae* also have *matS* sites that are bound by MatP and these affect the spatiotemporal coordination of the replication and segregation of both chromosomes (Demarre et al. 2014). Counterparts of *matS* are present in the Ter regions of the chromosomes of *Erwinia carotovora*, *S. enterica* serovar Typhimurium LT2, and *Yersinia pestis* (Mercier et al. 2008). MatP forms bridged tetramers that link distant *matS* sites to condense the Ter DNA (Dupaigne et al. 2012). MatP interacts with ZapB, a cell division-associated protein, to position the Ter macrodomain at the cell midpoint and to ensure its segregation (Espéli et al. 2012). With proteins ZapA, C, and D, ZapB functions to organise the Z ring that acts as a scaffold for the assembly of the complexes responsible for cell division (Buss et al. 2013). ZapB is dependent on ZapA for interaction with the tubulin-like FtsZ protein (Galli and Gerdes 2010). MatP also interacts with the MukBEF-Topo-IV complex at the Ter macrodomain, displacing MukBEF from Ter and facilitating its interaction with the Ori macrodomain. This is part of a step-wise process in which MukBEF manoeuvres the chromosome such that Ori and Ter adopt their assigned positions at the pole and mid-cell, respectively, by the time cell division takes place (Lioy et al. 2018; Nolivos et al. 2016). Together with MukBEF and SeqA, MatP is part of a group of proteins found exclusively in bacteria that express the Dam methyltransferase (Brezellec et al. 2006; Mercier et al. 2008).

1.23 MaoP and the *maoS* Site

The MaoP protein is important for the organisation of the Ori macrodomain (Duigou and Bocard 2017; Valens et al. 2016). It binds to a 17-bp sequence that lies adjacent to the *maoP* gene. Both MaoP and *maoS* are required for accurate positioning of Ori within the cell during the division cycle and for restricting the ability of the Ori macrodomain to interact with other territories of the chromosome, especially the Right Macrodomain and the NS-Right region (Valens et al. 2016). MaoP has been described as belonging to a class of DNA-binding proteins that has co-evolved with Dam, including MatP (see below), MukBEF, and SeqA: all of these proteins are involved in chromosome replication/segregation (Brezellec et al. 2006; Valens et al. 2016).

1.24 SlmA and Nucleoid Occlusion

The nucleoid is protected from bisection by the Z ring through a process known as nucleoid occlusion, or NO (Woldringh et al. 1991). If NO fails to occur the result is guillotining of the chromosome, fragmentation of the DNA, and cell death. In *E. coli*, the DNA-binding SlmA protein plays a key role in NO through an interaction with FtsZ (Bernhardt and de Boer 2005). SlmA works by interfering with the polymerisation activity of FtsZ, inhibiting Z-ring formation (Cho, H., et al. 2011). SlmA binds to DNA and to FtsZ simultaneously. Its DNA-binding sites are distributed around the chromosome, but it does not bind in the Ter macrodomain. The inhibitory activity of SlmA on FtsZ polymerisation outside of Ter is thought to delay septum formation until after the Ter macrodomain has been replicated (Tonthat et al. 2011).

1.25 The Min System and Z Ring Localisation

The cell division septum must be placed centrally if rod-shaped bacteria like *E. coli* are to divide into daughter cells of equal size. This placement is achieved through a gradient of FtsZ-inhibitor Min proteins. The gradients extend from regions of maximum Min density at the cell poles to a region of minimum density (and therefore minimum FtsZ inhibition) at mid-cell (Bramkamp and van Baarle 2009; Monahan and Harry 2012; Rowlett and Margolin 2015). While nucleoid occlusion operates to prevent guillotining of the nucleoid by the closing division septum, the Min system works independently of the nucleoid and is concerned with the correct localisation of the septum. Indeed, bacteria that are rendered chromosomeless still tend to form the cell division septum at the mid cell (Sun et al. 1998).

The term ‘Min’ is derived from ‘minicell’, a phenotype in *E. coli min* mutants (and other rod-shaped bacteria) where eccentric placement of the division septum produces two daughter cells of uneven length, one of which is too small to accommodate a chromosome (Adler et al. 1967; Reeve et al. 1973), although it can house plasmids (Roozen et al. 1971).

E. coli uses the MinC protein to inhibit FtsZ polymerisation (Hu et al. 1999; Hu and Lutkenhaus 2000) but, unlike *B. subtilis*, it lacks a cell-pole-anchoring protein that can be used to recruit MinC and other Min complex components to that part of the envelope. It relies instead on a MinC protein gradient extending from each pole to the midcell, with MinC forming a complex with the membrane-binding MinD ParA-like ATPase protein (de Boer et al. 1989, 1991; Hu and Lutkenhaus 2003). A third protein, MinE, is used to target the MinCD complex to the cell poles, with MinE (and phospholipid) stimulating the ATPase activity of MinD (Hu and Lutkenhaus 2001). MinE binds to the membrane at the pole, targeting MinCD complexes, displacing both MinC and MinD and stimulating ATP hydrolysis by MinD (Loose et al. 2011; Park et al. 2011). MinE and MinD set up a high-speed oscillating system in which MinC is trafficked from pole to pole, on average spending a minimum of time at mid-cell and most of the time at the poles (Raskin and de Boer 1997; Hu and Lutkenhaus 1999; Hu et al. 2002). It is the relative paucity of MinC at mid-cell that diminishes the inhibitory influence on FtsZ polymerisation and Z-ring formation (Hu and Lutkenhaus 1999; Raskin and de Boer 1999a,b). In addition to inhibiting FtsZ polymerisation by protein-protein interaction, the oscillation of MinC populations from pole to pole has an impact on the distribution of other FtsZ-interactors. Together with FtsZ itself, the ZapA, ZapB, and ZipA proteins oscillate oppositely to MinC and with a similar dynamic pattern. ZapB does not bind FtsZ directly but through ZapA, which does bind FtsZ. ZipA, with FtsA, connects FtsZ to the cytoplasmic membrane (Pichoff and Lutkenhaus 2005) while ZapA-ZapB stimulates Z-ring formation and stabilises it (Buss et al. 2013; Galli and Gerdes 2010; Gueiros-Filho and Losick 2002). Therefore, the oscillatory movements of MinC proteins probably trigger periodic assembly and disassembly of the Z ring complexes (Bisicchia et al. 2013; Thanedar and Margolin 2004).

B. subtilis possesses the cell-pole-targeting protein DivIVA, which is involved both in chromosome attachment at the pole in sporulating cells (Section 1.10) and in directing the cellular localisation of MinC (Cha and Stewart 1997; Edwards and Errington 1997). The utility of DivIVA as a general pole-targeting protein arises from its ability to sense cell membrane curvature, which is maximal at the poles (Edwards et al. 2000; Lenarcic et al. 2009). The MinC protein is bound by MinD and an adaptor protein, MinJ, connects this complex to DivIVA (Bramkamp et al. 2008; Patrick and Kearns 2008). As the division septum develops, invagination of the membrane, and the associated membrane curvature, recruit DivIVA from the pole to the mid-cell (which is the soon-to-be pole of the new daughter cell) (Gregory et al. 2008; Rodriguez and Harry 2012; van Baarle and Bramkamp 2010).

Genetic elimination of the Min system and of nucleoid occlusion is deleterious for cell growth in rich medium, but the mutants can grow and divide in minimal medium (Bernhardt and de Boer 2005; Yu and Margolin 1999). This suggests that additional systems exist to ensure chromosome segregation and cell division (Bailey et al. 2014; Cambridge et al. 2014). A link between the Ter-matS-binding MatP protein and ZapB connects the Ter macrodomain of the chromosome to the divisome's ZapB-ZapA-FtsZ complex (Espéli et al. 2012) and this may afford the nucleoid itself a role in determining Z ring placement (Rowlett and Margolin 2015; Yu and Margolin 1999).

The Min system operates on Z ring placement through an inhibitory mechanism. This strategy is not used universally among bacteria. For example, positive placement is used

to direct Z ring positioning in *Streptomyces* spp. Here, the SsgA protein takes up position at mid-cell and recruits a partner, SsgB, which is thought to promote FtsZ polymerisation and to link the Z rings to the membrane. These rings form the basis of sporulation septa in these actinomycetes (Traag and van Wezel 2008; Willemse et al. 2011). Sporulation takes place in the aerial hyphae where up to one hundred division septa are laid down, dividing the hyphae into compartments that each contain one genome copy (Zhang, L. et al. 2016). The SepG membrane protein recruits SsgB to future septum sites; it is also required for nucleoid compaction indicating that plays a role in coordinating chromosome organisation and placement of the division septum (Zhang et al. 2016). The PomZ protein in *Myxococcus* spp. is also a positive regulator of Z-ring positioning; PomZ shares with MinD the property of being a ParA-like ATPase (Treuner-Lange et al. 2013). It forms a complex with PomX and PomY that moves over the surface of the nucleoid in a biased, randomised motion that becomes constrained at the mid-cell. Once at the mid-cell location, PomXYZ recruits FtsZ (Schumacher et al. 2017).

Macrodomains play important roles in determining the choreography of the daughter chromosomes, as these segregate prior to cell division (Espéli et al. 2008). They also correlate with global gene expression patterns, suggesting that the overall gene expression programme of the cell is written into the architecture of the nucleoid (Cameron et al. 2017; Sobetzko 2016; Sobetzko et al. 2012). To appreciate the significance of the connections between nucleoid structure and gene expression, it will be necessary to consider the contributions made to both by variable DNA structure and NAPs.

1.26 DNA in the Bacterial Nucleoid

DNA in bacterial cells is maintained in an underwound state and this affects the shape that the DNA duplex adopts as it seeks to adopt a minimal energy conformation. The underwinding arises due to a deficit in helical turns, i.e. the number of times the two DNA strands twist around the duplex axis. The twist deficiency imparts torsional stress to the duplex, which is relieved by allowing the duplex to adopt a writhed confirmation in which the helical axis coils around itself. This coiling of the already coiled DNA duplex creates supercoiling and has the effect of making the DNA molecule more compact. In the context of the nucleoid, such compaction assists with solving the problem of packaging the genetic material within the cell. The most supercoiled parts of the chromosome form branches, facilitating further compaction.

1.27 DNA Topology

The topological state of a DNA molecule is described by three parameters: the linking number (the number of times one DNA strand crosses the other in the duplex); the twist (the number of complete turns made by the strands around one another along the length of the duplex); and the writhe (the number of writhing turns made by the duplex axis around itself) (Bauer et al. 1980; Boles et al. 1990; Vinograd et al. 1965). Changes to the linking number can be made by breaking one or both strands of the duplex, twisting the DNA with

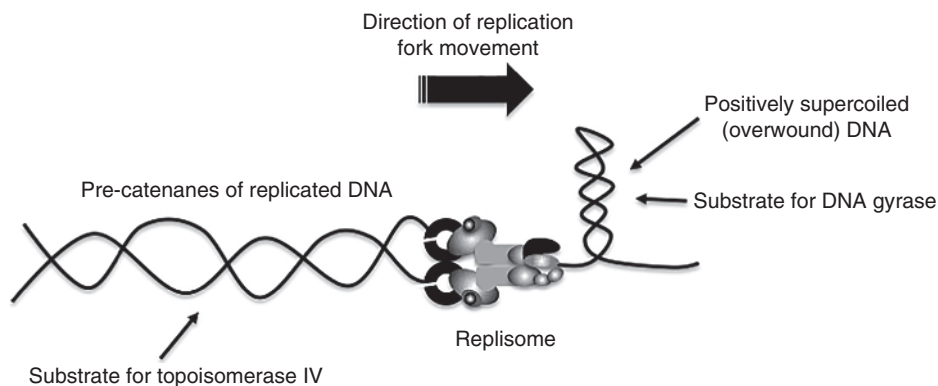


Figure 1.14 The topological consequences of replisome activity. The replication fork, moving from left to right, overwinds the DNA ahead to create a domain of positive supercoiling. Unless this is relaxed, the replisome will be unable to advance further. Relaxation of the positively supercoiled domain is carried out by DNA gyrase: it neutralises the positive supercoils by introducing negative ones. In *C. crescentus*, the GapR protein binds to positively supercoiled DNA and stimulates its relaxation by type II topoisomerases (see Guo et al. 2018). The newly synthesised DNA behind the replisome becomes catenated. These interwound, double-stranded DNA molecules are decatenated by DNA topoisomerase IV, allowing the daughter chromosomes to be segregated at cell division. For further information, see Lopez et al. (2012) and Postow et al. (2001).

or against the sense of the double helix, and then resealing the strand breaks (Higgins and Vologodskii 2015). Twisting the DNA in keeping with the sense of the double helix tightens the duplex and imparts positive writhe, resulting in positive supercoiling. Twisting the DNA against the sense of the double helix underwinds the molecule, imposing negative writhe and therefore negative supercoiling (Sinden 1994). In nature, negative supercoiling is the norm although positive supercoils do arise naturally, especially during the movement of the DNA replication fork (Figure 1.14) and during transcription (Liu and Wang 1987) (Figure 1.15). In contrast to mesophilic bacteria like *E. coli*, hyperthermophilic archaea that live in environments that are characterised by very high temperatures possess a reverse gyrase activity and maintain their DNA in a positively supercoiled state (Couturier et al. 2014; Lipscomb et al. 2017). Reverse gyrase, first described in *Sulfolobus acidocaldarius* in 1984, uses ATP to introduce positive supercoils into DNA (Kikuchi and Asai 1984; Ogawa et al. 2015).

A covalently closed, circular duplex DNA molecule that is neither overwound nor underwound is said to be topologically relaxed. If this circular duplex undergoes an increase in its linking number (overwinding) or a decrease (underwinding) it retains an identical nucleotide sequence compared with the relaxed form but differs from this form topologically (Sinden 1994). For this reason, the relaxed, overwound, and underwound isomeric forms of the circular duplex are referred to as topoisomers of the same DNA molecule. Enzymes that produce topological changes in DNA by altering the linking number are called topoisomerases and *E. coli* has four: topoisomerase I (topo I), DNA gyrase (a topo II family member), topoisomerase III (topo III), and topoisomerase IV (topo IV) (Bates and Maxwell 2005; Wang 2002) (Table 1.1).

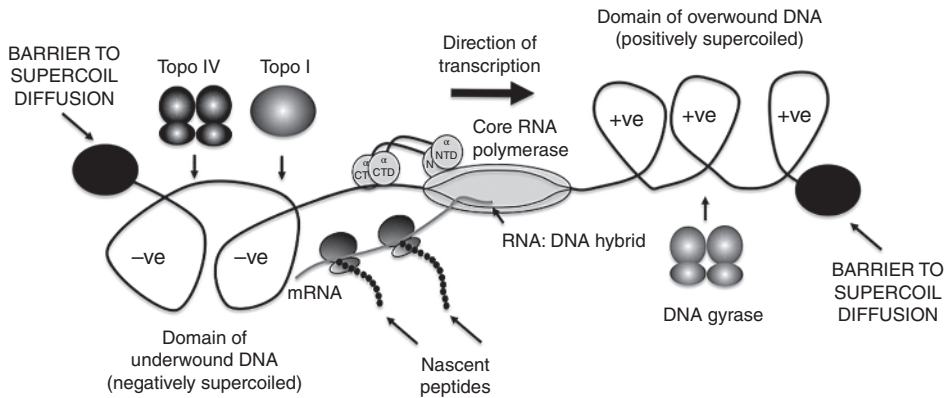


Figure 1.15 The topological consequences of transcription. This is the twin supercoil domain model, as proposed by Liu and Wang (1987) and supported by numerous independent experiments. Core RNA polymerase is engaged in transcript elongation and the mRNA is simultaneously being translated by polyribosomes to produce nascent peptides. As the coupled transcription-translation complex moves from left to right, the DNA template ahead becomes overwound, or positively supercoiled, while the DNA behind becomes underwound, or negatively supercoiled. This situation will bring transcription to a halt as the machinery jams. This is because: (a) the domains of supercoiled DNA cannot be resolved by supercoil lateral diffusion because the transcribed region is bounded by topological barriers (solid black discs) and (b) the bulky transcription-translation complex cannot rotate around the DNA to relieve the torsional tension in the DNA duplex. The solution to the problem is provided by the DNA-relaxing activities of topoisomerases: DNA gyrase removes the positive supercoils by introducing negative supercoils, while the negatively supercoiled domain is relaxed by DNA topoisomerase I and IV. Interference with these relaxation processes can result in undesirable outcomes, such as R-loop formation (Figure 1.16). Topological barriers can be created by head-to-head transcription collisions and by collisions between converging replisomes and RNA polymerases; they can also arise from the presence of nucleoprotein complexes and distortions (e.g. sharp bends) in the DNA duplex.

1.28 DNA Topoisomerases: DNA Gyrase

Topoisomerases are classed as type I if they change the linking number of DNA in steps of one, and as type II if the linking number changes in steps of two (Champoux 2001; Wang 2002) (Table 1.1). DNA gyrase is a type II enzyme and it has the property, unique to prokaryotes, of being able to introduce negative supercoils into DNA (Gellert et al. 1976a; Higgins et al. 1978; Nöllmann et al. 2007). This negative supercoiling activity is ATP dependent and there is an ATP-binding site in the B subunit of gyrase (Gellert et al. 1979; Mizuuchi et al. 1978). Gyrase is an A_2B_2 hetero-tetramer and it is essential; knockout mutations in either of the genes that encode its A (*gyrA*) or B (*gyrB*) subunits are lethal (Bates and Maxwell 2005). Its essentiality has made gyrase a very attractive target for antimicrobial therapy and a number of drugs are available that target its subunits (Collin et al. 2011; Maxwell 1999). The coumarin class of antimicrobials have been particularly useful as research tools because they compete with ATP for access to the B subunit ATPase and inhibit gyrase activity without inducing the SOS response (DeMarini and Lawrence 1992; Gellert et al. 1976b; Gormley et al. 1996; Pugsley 1981; Sugino and Cozzarelli 1980; Sugino et al. 1978).

Table 1.1 The topoisomerases of *E. coli*.

Enzyme name (type ^a)	Molecular mass (kDa)	Gene(s)	Comments
(Type I)			
Topoisomerase I	97	<i>topA</i>	DNA swivelase that makes a transient cut in one strand of the DNA duplex, binds to the cut site via a 5'-phosphotyrosine bond; relaxes negatively supercoiled DNA; requires Mg ²⁺
Topoisomerase III	73.2	<i>topB</i>	Relaxes negatively supercoiled DNA; decatenase; has catenase activity in association with RecQ; requires Mg ²⁺
(Type II)			
DNA gyrase (Topoisomerase II)	105 (A subunit)	<i>gyrA</i> (A subunit)	ATP-dependent negative supercoiling activity; relaxes negative supercoils in an ATP-independent manner; relaxes positively supercoiled DNA; binds DNA transiently via a 5'-phosphotyrosine bind; requires Mg ²⁺
	95 (B subunit)	<i>gyrB</i> (B subunit)	
Topoisomerase IV	75 (ParC)	<i>parC</i> (ParC; GyrA-like)	Decatenase activity; interacts with MukBEF; relaxes negative supercoils; requires ATP, Mg ²⁺
	70 (ParE)	<i>parE</i> (ParE; GyrB-like)	

- a) Type I enzymes change the linking number of the duplex DNA substrate in steps of 1 ($\Delta Lk = 1$) while type II enzymes change the linking number in steps of 2 ($\Delta Lk = 2$).

In contrast, those drugs (e.g. quinolones) that inhibit the A subunit during DNA cleavage and religation cause DNA damage that results in induction of the SOS response (Gellert et al. 1977), something that can complicate experimental design and data interpretation. Gyrase also has an ATP-independent DNA relaxing activity that is unmasked only in the absence of ATP (Gellert et al. 1977; Higgins et al. 1978; Williams and Maxwell 1999). The ATP-dependent mechanism by which gyrase introduces negative supercoils into DNA is also capable of relaxing positive supercoils (Ashley et al. 2017). This activity is especially important when gyrase processes the positively supercoiled DNA that is a by-product of transcription and DNA replication (Koster et al. 2010) (Figures 1.14 and 1.15). DNA gyrase in living cells responds to the [ATP]/[ADP] ratio, linking the management of DNA topology to the metabolic activity of the bacterium (Hsieh et al. 1991a,b; Snoep et al. 2002; van Workum et al. 1996). Gyrase activity is also tuned in living bacteria by stresses such as the acidification of the bacterial cytosol that accompanies adaptation to acid stress (Colgan et al. 2018).

1.29 DNA Topoisomerases: DNA Topoisomerase IV

Topo IV was discovered in *E. coli* 14 years after gyrase, its type II topoisomerase companion (Kato et al. 1990) (Table 1.1). It is encoded by two genes, *parC* and *parE*, whose names hint at a defect in chromosome partitioning that is associated with mutants with a topo IV deficiency (Kato et al. 1990). Topo IV is an ATP-dependent topoisomerase but, unlike gyrase, which it closely resembles in amino acid sequence and subunit structure, it cannot introduce negative supercoils into DNA. Instead, Topo IV relaxes both positively and negatively supercoiled DNA and is an important DNA decatenase (Bates and Maxwell 2007; Crisona and Cozzarelli 2006; Kato et al. 1992; Peng and Mariani 1993; Zawadzki et al. 2015). Its relationship with the MukBEF SMC-like complex is emerging as one of Topo IV's most physiologically significant functions, one that is important for the correct spatiotemporal management of chromosome replication, architecture, and segregation during the cell cycle (Section 1.21).

1.30 DNA Topoisomerases: DNA Topoisomerase I

The principal source of relaxation activity for negatively supercoiled DNA is Topo I, a monomeric, ATP-independent enzyme that is encoded by the *topA* gene in the Ter macrodomain (Margolin et al. 1985) (Table 1.1). This enzyme also has a catenase/decatenase activity on double-stranded circular DNA with single-stranded regions (Terekhova et al. 2012; Tse and Wang 1980) and prevents over-replication of the chromosome originating at *oriC* (Usongo and Drolet 2014). Topo I is a type I topoisomerase that has a 'swivelase' activity. It cuts one of the two DNA strands in a negatively supercoiled molecule, forming a covalent link to the cut strand, and allows the torsional strain in the DNA to drive the rotation of the intact strand through the gap (Kirkegaard and Wang 1978). The result is an increase of 1 in the linking number of the DNA (Bates and Maxwell 2005). The *topA* gene is not essential, although knockout mutants grow slowly (Margolin et al. 1985; Sternglanz et al. 1981). Bacteria that lose Topo I through *topA* knockout mutations can compensate in different ways, restoring a growth rate that is close to that of wild-type strains (Raji et al. 1985; Richardson et al. 1984). One option is to acquire non-lethal mutations in *gyrA* or *gyrB* that result in the expression of a gyrase with a reduced negative supercoiling activity (DiNardo et al. 1982; Pruss and Drlica 1985; Pruss et al. 1982, 1986; Richardson et al. 1984, 1988). Another possibility exploits the amplification of the copy number of the *parC* and *parE* genes, resulting in increased expression of Topo IV. In these strains the elevated expression of Topo IV with its DNA-relaxing activity can compensate for the missing Topo I, restoring the growth rate of the mutant to one that is similar to wild type (Dorman et al. 1989; Free and Dorman 1994; McNairn et al. 1995). Bacteria that are not exposed to stressful growth conditions such as elevated temperature, low pH, or raised osmotic stress or lack of oxygen do not require compensatory mutations in order to display normal rates of growth (Ní Bhriain and Dorman 1993). This suggests that a link exists between environmental stress and the management of DNA supercoiling in bacteria (Dorman and Dorman 2016).

1.31 DNA Topoisomerases: DNA Topoisomerase III

Topo III is a second type I topoisomerase that, like Topo I, is an ATP-independent monomeric enzyme (Table 1.1). It is encoded by the *topB* gene and is not essential for the survival of the bacterium (Usongo et al. 2013). However, mutants deficient in both Topo I and in Topo III do not survive (Stupina and Wang 2005). Topo III has weak DNA-relaxing activity and functions principally as a decatenase (Nurse et al. 2003; Perez-Cheeks et al. 2012). Its apparent weakness as a DNA-relaxing enzyme compared with Topo I arises from a difference in the mechanisms used by the two topoisomerases: Topo I operates in a processive manner with short pauses between processive runs, whereas Topo III takes long pauses, leading to a relaxing process with an overall rate that seems slower (Terekhova et al. 2012). While Topo I plays an important role in controlling the frequency of chromosome replication initiation at *oriC*, Topo III contributes to the management of replication fork collision in the Ter macrodomain (Suski and Marians 2008). In fact, all four topoisomerases are important components of the chromosome replication machinery and display both a division of labour and an interesting degree of redundancy that allows the cell to continue to function even if one of the enzymes experiences interference.

1.32 DNA Replication and Transcription Alter Local DNA Topology

The linking number of DNA is changed at a local level by the processes of transcription and DNA replication. In 1987, Liu and Wang proposed, in a landmark theoretical paper, that the process of transcription would induce overwinding of the DNA template ahead of RNA polymerase and underwinding behind (Liu and Wang 1987) (Figure 1.15). Experimental studies provided support for the proposal, leading to the realisation that topoisomerases play important roles in transcription by relieving the torsional stress that the process creates (Ahmed et al. 2017; Chong et al. 2014; Higgins 2014; Rahmouni and Wells 1992; Rani and Nagaraja 2019; Wu et al. 1988). The role of local DNA supercoiling in the modulation of transcription and in gene-to-gene communication will be addressed in Section 8.2. Here we will consider the impact of transcription on nucleoid architecture and on DNA replication.

1.33 Transcription and Nucleoid Structure

Several investigations have made links between patterns of transcription and the superstructure of the bacterial nucleoid. At a practical level, replication fork movement must be reconciled with the needs of transcription (initiation, elongation, and termination), so aligning replisome movement with the direction of gene transcription avoids significant conflicts between DNA and RNA polymerases. Collisions between the replisome and RNA polymerase are known to cause severe inhibition of replisome progression (Mirkin and Mirkin 2005). Transcriptional promoters that oppose the direction of replisome movement serve to pause the replication fork, while transcription terminators that are aligned with the direction of replisome movement also act as replication fork pause sites (Mirkin et al.

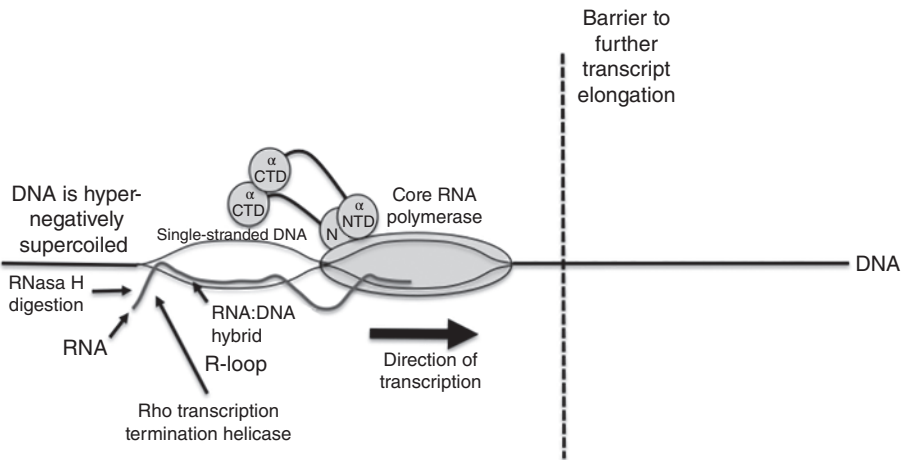


Figure 1.16 R-loop formation. When RNA polymerase reads a G+C-rich template, stalls and backtracks, it leaves a domain of hypernegatively supercoiled in its wake and the associated transcription stalling may allow the RNA transcript to base pair with its DNA template strand, leaving the non-transcribed strand as a single-stranded bubble. Other impediments to RNA polymerase progression include head-on collisions with other transcription units or with replisomes (the barrier is represented by the vertical gapped line). Loss of topoisomerase I activity is known to promote R-loop formation because it encourages the accumulation of hyper negative superhelicity in DNA that is being transcribed (or replicated). Failure to process and remove RNA loops can lead to DNA damage, including double-stranded breaks and hyper-recombination. The Rho transcription-terminating helicase (Figure 3.4) helps to suppress R-loop formation by preventing backtracking by RNA polymerase while RNase H eliminates R-loops by removing the RNA component of the RNA:DNA hybrid in R-loops.

2006). Conflicts between the replisome and RNA polymerase can generate R-loops, stalling replication, and transcription in the affected region until RNase H removes the R-loop (Kuziminov 2018). Unresolved R-loops also result in hyper-recombination and genome instability, so avoiding replication–transcription conflicts is very desirable (Figure 1.16).

A correlation has also been reported between gene *essentiality* and alignment with the direction of replication fork movement (Rocha and Danchin 2003). Overall, one finds more genes on the leading strand than on the lagging strand of the chromosome and this may reflect the outcome of evolutionary pressure to minimise collisions (Rocha 2008). The question of replication and transcription alignment/collision also has a DNA topological dimension, in that converging polymerases will create, and trap between them, a domain of positively supercoiled DNA that must be resolved by type II topoisomerases: gyrase or topo IV (Crisona et al. 2000; Kato et al. 1992). The mechanisms of action of these enzymes bring an increased risk of double-stranded breaks occurring in the chromosome with potentially lethal consequences for the cell (Hiasa et al. 1996; Lockshon and Morris 1983). These factors may impose limits on the options for gene orientation on the chromosome, influencing its evolution.

Experiments using chromosome conformation capture (3C) methods (Dekker et al. 2002) have suggested a link between DNA replication, transcription, and the organisation of the bacterial nucleoid (Cagliero et al. 2013; Le et al. 2013; Le and Laub 2016; Liroy et al.

2018; Meyer et al. 2018). Work that exploited site-specific recombination as a measure of interaction frequency within the bacterial genome suggested that transcription activity and the associated changes in local DNA topology have a structural function in the nucleoid (Booker et al. 2010; Higgins 2014). Furthermore, the global level of DNA supercoiling in the nucleoid can be tuned by treating the bacterium with rifampicin, an inhibitor of RNA polymerase (Rovinskiy et al. 2012). Experiments with trimethylpsoralen crosslinking have provided evidence that a gradient of DNA supercoiling exists extending from the origin of chromosome replication to the Ter macrodomain (Lal et al. 2016). Bioinformatic and experimental studies show that the distribution of binding sites around the chromosome for DNA gyrase is non-uniform, with more sites being detected close to the origin of replication (Jeong et al. 2004; Sobetzko et al. 2012; Sutormin et al. 2019). This could indicate that the Ori-proximal part of the chromosome is the most underwound, contradicting the data from the psoralen-binding-and-crosslinking studies (Lal et al. 2016). On the other hand, it may indicate a need to compensate for the paucity of negative supercoiling in the Ori-proximal part of the chromosome that is predicted by those studies. Further evidence that the chromosome is not uniformly supercoiled has come from investigations in which supercoiling-sensitive genes have been placed at different locations in the genome (Bryant et al. 2014). A comprehensive survey of the effects of gene position on transcription in *E. coli* showed that the propensity for transcription varies with chromosomal location: horizontally acquired genetic elements are associated with quiescent regions while ribosomal and other metabolic genes are in highly active zones (Scholz et al. 2019). It should be noted, however, that several studies which have explored the influence of gene position (including the possible contribution of differences in local DNA supercoiling) did not detect clear changes in the level of expression of the test gene(s) at different sites around the chromosome (Block et al. 2012; Brambilla and Sclavi 2015; Chandler and Pritchard 1975; Miller and Simons 1993; Pavitt and Higgins 1993; Schmid and Roth 1987; Sousa et al. 1997; Thompson and Gasson 2001; Ying et al. 2014).

1.34 Nucleoid-associated Proteins (NAPs) and Nucleoid Structure

Sequence-dependent DNA-binding proteins such as MaoP, MatP, SeqA, and SlmA play important roles in the organisation of the chromosome in the nucleoid. Yet the term ‘nucleoid-associated protein’, or NAP, is usually reserved for another group of DNA-binding proteins, not all of which are sequence-dependent for DNA binding. Most NAPs were discovered in roles other than organising the nucleoid. Several were first encountered as contributors to the efficient operation of site-specific recombination systems in bacteria or their phage (reviewed in Dorman and Bogue 2016). Later, their more general contributions to cell physiology became appreciated. We have already encountered two, FIS and IHF, as components of the systems that govern the initiation of chromosome replication (Section 1.3). Originally, FIS was identified as an enhancer-binding protein that improved the efficiency of the DNA inversion events responsible for phase-variable expression of tail fibre genes in bacteriophage Mu, and for flagellar phase variation in *Salmonella* (Koch and Kahmann 1986; Johnson et al. 1986). IHF was detected as an

essential factor for the site-specific entry of the bacteriophage lambda genome into the *att λ* site on the chromosome of *E. coli* (Nash and Robertson 1981) (Section 1.35). These and other NAPs have emerged as important regulators of a multitude of bacterial genes (Dillon and Dorman 2010). They have mechanisms of action that typically involve the making of adjustments to local DNA architecture. In many cases, these adjustments affect gene expression at the level of transcription (or beyond) and include direct or indirect effects on nucleoid structure. A summary of the key features of some of the most important (i.e. best-studied) NAPs is given in the following sections.

1.35 DNA Bending Protein Integration Host Factor (IHF)

IHF is essential both for the integration and the excision of bacteriophage lambda into/from the *E. coli* chromosome (Bushman et al. 1984; Seah et al. 2014) (Figure 1.17). These site-specific recombination reactions are catalysed by the tyrosine integrase Int (Craig and Nash 1983; Han et al. 1994; Hoess et al. 1980; Kikuchi and Nash 1979; Tong et al. 2014). IHF is usually expressed as a heterodimer composed of an alpha and a beta subunit encoded by the *ihfA* and *ihfB* genes, respectively, which are located at different places on the chromosome (Haluzi et al. 1991; Mendelson et al. 1991; Miller and Friedman 1980; Nash et al. 1987). IHF has strict DNA sequence requirements for binding and its binding sites typically are located in A+T-rich DNA (Miller and Friedman 1980). The protein inserts a looped beta strand with a proline amino acid at its apex into the minor groove of the DNA at the target site, producing a distortion that causes the DNA duplex to bend (Engelhorn and Geiselman 1998; Engelhorn et al. 1995; Rice et al. 1996; Sun et al. 1996; Vivas et al. 2012) (Figure 1.18). The combined effects of the bends introduced by each subunit is to create a turn of up to 180° in the pathway taken by the DNA helix (Rice et al. 1996; Sugimura and Crothers 2006). This allows IHF to play important roles in nucleoid architecture, chromosome replication, site-specific recombination, transposition, plasmid maintenance, and transcription regulation (Biek and Cohen 1989; Crellin et al. 2004; Dorman and Bogue 2016; Prieto et al. 2012; Ryan et al. 2004; Saha et al. 2013; Swinger and Rice 2004). In terms of amino acid sequence, protein structure, and subunit composition, IHF is a close relative of the HU NAP. Both proteins have an $\alpha\beta$ heterodimeric structure and all four proteins have similar amino acid sequences (Dey et al. 2017). IHF appears to be a specialist member of the HU superfamily of DNA-binding proteins, a group that includes relatives encoded by the human genome that contribute to chromosome partitioning (Burroughs et al. 2017). Although IHF is usually considered as acting in an $\alpha\beta$ heterodimeric form, transcriptomic studies carried out in *Salmonella* indicate that bacteria expressing just the α , just the β , and both α and β subunits, are altered in expression of distinct-yet-overlapping groups of genes, implying that IHF can form homodimers that are biologically active (Mangan et al. 2006). Indeed, previous investigations have found that IHF homodimers have DNA-binding activity (Huszczynska-Sawicka and Kur 1997; Werner et al. 1994; Zablewska and Kur 1995; Zulianello et al. 1994). When reading the older IHF literature, the reader should be aware that in 1996 the names of the genes were changed from *himA* and *himD* (where ‘*him*’ referred to H.I. Miller) to *ihfA* and *ihfB*, respectively (Weisberg et al. 1996).

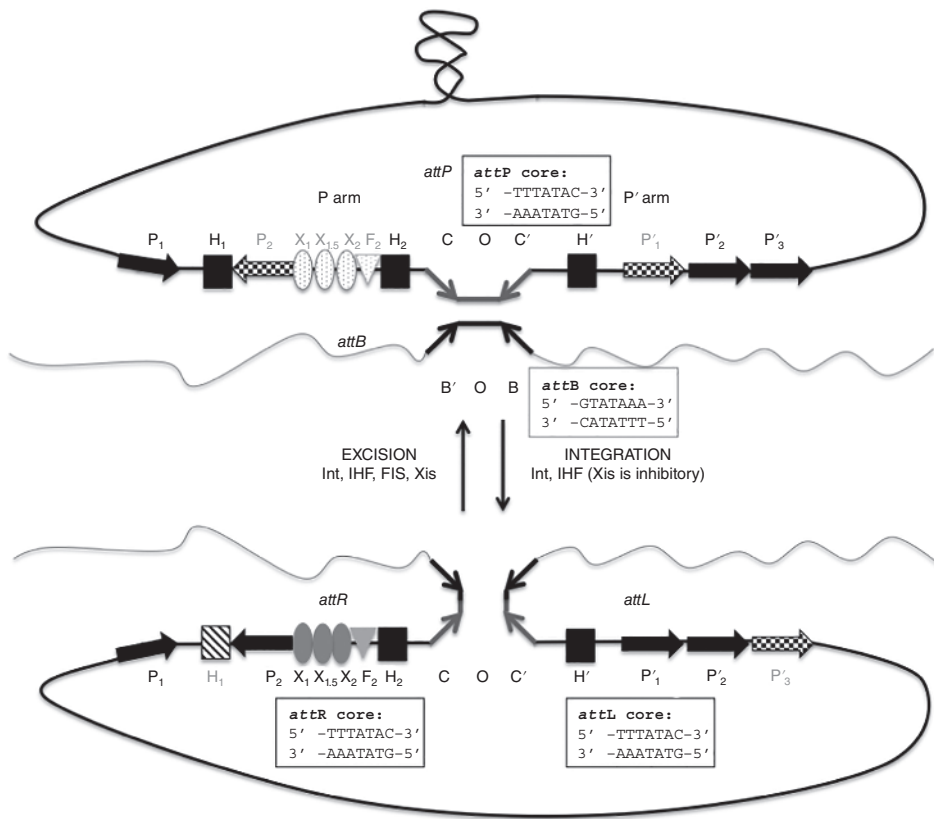


Figure 1.17 Integration of bacteriophage lambda at the lambda attachment site on the *E. coli* chromosome and excision of the prophage. Integration and excision are catalysed by the phage-encoded Int tyrosine integrase. In the integration step, the recombining sites at *attP* (on the phage) and *attB* (on the bacterial chromosome) undergo intermolecular site-specific recombination to generate the lambda prophage. Excision recreates *attB* and *attP* from *attL* and *attR*, the direct repeats that form the boundaries between the prophage and chromosomal DNA. The nucleotide sequences of each of these four elements are shown in the figure. In addition to the Int recombinase, the integration step requires the architectural protein, integration host factor, IHF. Excision also requires IHF together with the phage-encoded directionality factor, Xis, and the host-encoded FIS protein. Xis is inhibitory to integration and FIS has a stimulatory effect on excision (see Ball and Johnson 1991; Seah et al. 2014). Binding sites for these proteins are represented by labelled arrows (P, Int), squares (H, IHF), ovals (X, Xis), and a triangle for a FIS (F) binding site. Binding sites in the P' arm of the free circular phage genome have primes added to their designations and these are retained in the *attL* segment of the prophage. Not every protein-binding site is occupied in each reaction. Occupied sites have a solid filling while unoccupied sites have a speckled or hatched filling. During integration, Int-binding sites P₁, P'₂, and P'₃ are occupied while P₂ and P'₁ are not; all three IHF-binding sites are occupied but the XIS and FIS sites are vacant. During excision, Int sites P₁, P₂, P'₁, and P'₂ are occupied, as are IHF sites H₂ and H' and all of the Xis and FIS sites; IHF site H₁ and Int site P₃ are vacant.

1.36 HU, a NAP with General DNA-binding Activity

The HU protein from *E. coli* is the founder member of a superfamily of related NAPs found throughout the prokaryotic world and beyond (Burroughs et al. 2017). HU interacts with DNA in the minor groove and this encourages the bound DNA to follow a looped path (Figure 1.18). This property helps HU to overcome the resistance of the DNA to loop formation by overcoming DNA's intrinsic stiffness (Johnson et al. 1986). HU-assisted loop formation contributes to the formation of nucleoprotein complexes involved in the control of transcription and site-specific recombination (Haykinson and Johnson 1993; Semsey et al. 2004). It also has RNA-binding activity, enabling it to influence translation (Balandina et al. 2001).

Each HU subunit inserts a beta sheet with an apical proline amino acid into the minor groove of the DNA at the binding site, inducing the DNA to bend (Figure 1.18). The bend angle is typically in the range of 105° – 140° and bends are not coplanar, having a dihedral angle that is consistent with the path taken in negatively supercoiled DNA (Swinger et al. 2003). The flexibility in the bend angle, coupled with the absence of a strict nucleotide sequence for DNA binding, may allow HU to participate as an architectural component in a wide variety of DNA-based transactions.

The α and β subunits of HU are encoded, respectively, by the *hupA* and *hupB* genes, located at distinct positions on the chromosome: *hupA hupB* double mutants that fail to express the HU protein display a filamentous cell phenotype because of disruption of the

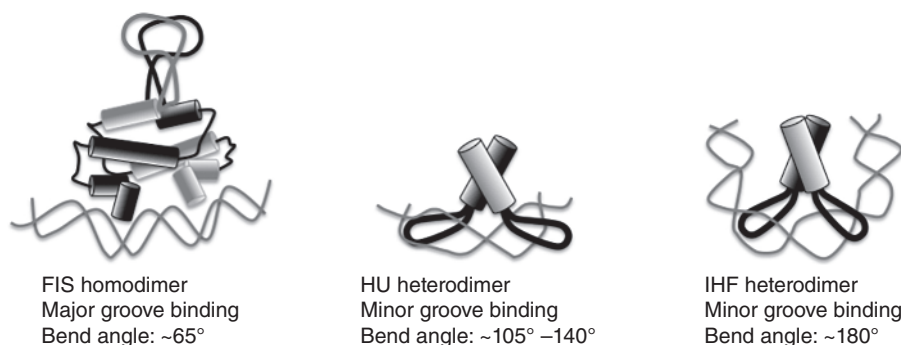


Figure 1.18 The interactions of the nucleoid-associated proteins FIS, HU, and IHF with DNA. The FIS protein is homodimeric, while HU and IHF are heterodimeric. HU and IHF are closely related at the level of amino acid sequence and the alpha and beta subunits of each protein are similar in sequence and secondary structure. The tertiary structures of HU and IHF are also quite similar, as their modes of binding to DNA. Each inserts a beta sheet from each subunit into the minor groove of its DNA target. IHF differs from HU in having a strict nucleotide sequence requirement for binding and in making more contacts with the DNA at its binding site. IHF also induces a much greater bend angle in the DNA. The bends on either side of the HU-DNA complex are non-coplanar. FIS binds in the major groove of DNA using an alpha helix, one of four alpha helices found in each FIS monomer. The protein uses an induced fit binding mechanism that compresses the minor groove lying between the two sites of insertion of the alpha helices in the major groove. FIS binds to a variety of sites with differing binding affinities; a Logo has been assembled that summarises the chief sequence characteristics of the highest affinity sites (see Stella et al. 2010).

cell cycle due to the arrest of DNA replication (Dri et al. 1991). HU interacts with the DnaA protein at *oriC* where it stimulates formation of the initiation complex in chromosome replication. Like IHF, HU is usually heterodimeric and is composed of an alpha and a beta subunit. The alpha subunit seems to have the primary responsibility for interacting with DnaA. This preference for the alpha subunit may facilitate enhanced HU–DnaA interaction at early stages of growth when an HU α_2 homodimer predominates rather than the $\alpha\beta$ heterodimer (Chodavarapu et al. 2008). HU can also influence chromosome replication initiation indirectly by repressing the expression of the gene that encodes SeqA (Lee, H., et al. 2001), the protein that sequesters *oriC* and excludes DnaA (Han et al. 2003; Slater et al. 1995; von Freiesleben et al. 1994).

The HU protein can form nucleosome-like structures in *E. coli* that are dependent on the local HU-to-DNA ratio (Sagi et al. 2004). It has been described as insulating transcription units on the chromosome by preventing changes in DNA supercoiling caused by transcription in one unit from influencing an adjacent one (Berger et al. 2016). HU may be particularly important for the maintenance of DNA supercoiling levels in the Ter macrodomain as the bacterium enters the stationary phase of the growth cycle (Lal et al. 2016). It has been reported to induce, together with FIS, weak and transient domain boundaries around the *E. coli* chromosome (Wu et al. 2019).

In laboratory-grown cultures, the subunit composition of the HU protein changes as a function of growth phase: In lag phase, as the bacterium adapts to its new environment, the α_2 form of HU occurs; in exponential growth the $\alpha\beta$ form predominates and the β_2 form is detected as the culture enters stationary phase (Claret and Rouvière-Yaniv 1997). The changing subunit composition of HU and the different DNA interaction properties of the distinct HU forms may contribute to processes that differentially compact the chromosome in the nucleoid and affect gene expression patterns (Hammel et al. 2016). Transcriptomic studies in *Salmonella* have shown that each form of the HU protein seems to govern a distinct group of genes, with overlaps between the three sub-regulons (Mangan et al. 2011).

1.37 The Very Versatile FIS Protein

FIS is the Factor for Inversion Stimulation, so called because it was discovered originally as an important architectural element in the DNA inversion mechanisms responsible for the phase-variable expression of flagella in *Salmonella* (Johnson et al. 1986) and of tail fibre proteins in bacteriophage Mu (Koch and Kahmann 1986). FIS is now known to contribute to a wide range of molecular events in bacteria, including DNA replication (Cassler et al. 1995; Filutowisz et al. 1992; Gille et al. 1991), site-specific recombination (Dhar et al. 2009; McLean et al. 2013), transposition (Weinreich and Reznikoff 1992), transcription regulation (Grainger et al. 2008; Hirvonen et al. 2001; Kelly et al. 2004; Pemberton et al. 2002), bacteriophage life cycles (Betermier et al. 1993; van Drunen et al. 1993; Papagiannis et al. 2007; Seah et al. 2014), illegitimate recombination (Shanado et al. 1997), and chromosome domain boundary formation (Hardy and Cozzarelli 2005; Wu et al. 2019).

1.38 FIS and the Early Exponential Phase of Growth

FIS is a homodimeric NAP that is encoded by the second gene in the *dusB-fis* operon and shows strong homology to the DNA-binding domain of the NtrC transcription factor (Bishop et al. 2002; Morett and Bork 1998). Transcription of the *fis* gene is maximal in the early stages of exponential growth and FIS plays an important role in boosting the expression of genes that encode components of the translational machinery of the cell (Appleman et al. 1998; Ball and Johnson 1991; Hirvonen et al. 2001; Osuna et al. 1995). FIS binds to the major groove of the DNA using a helix-turn-helix (HTH) motif that interacts with A+T-rich sites that match a weak consensus sequence (Hancock et al. 2016). The protein uses an induced fit binding mechanism that compresses the minor groove between those parts of the major groove that accommodate the HTH motifs of the two subunits (Figure 1.18) (Hancock et al. 2016; Stella et al. 2010). This creates a bend in the DNA of 65° according to FIS-DNA co-crystal structure data (Stella et al. 2010) with bends of up to 90° also being reported (Kostrewa et al. 1992; Pan et al. 1996).

Transcription factors that introduce bends into DNA can facilitate additional contacts between DNA (including proteins bound to that DNA) located upstream of the promoter and bound RNA polymerase, increasing the efficiency of transcription initiation (Huo et al. 2006; Rivetti et al. 1999; Verbeek et al. 1991). FIS acts as a ‘conventional’ transcription factor at some promoters, making protein–protein contacts with RNA polymerase (Bokal et al. 1997) and its DNA-bending activity has the potential to enhance the efficiency of the early stages of the transcription process. FIS can also influence promoter function without the need to contact RNA polymerase. The *leuV* operon consists of three genes that encode three of the four tRNA_{Leu} isoreceptors and its promoter is under the positive control of FIS. The single binding site for the FIS protein upstream of the *leuV* promoter is located in a DNA segment that is prone to becoming single-stranded under the torsional stress imposed by negative supercoiling. This phenomenon is known as supercoiling-induced DNA duplex destabilisation, SIDD (Benham 1992, 1993). Binding of the FIS protein to its site within the SIDD element displaces the tendency towards duplex destabilisation to the nearest susceptible site, in this case, the *leuV* promoter – assisting in the formation there of an open transcription complex (Opel et al. 2004). This mechanism is not peculiar to FIS and has been demonstrated for the IHF NAP too (Sheridan et al. 1998). It is likely to be used at many other promoters and represents an under-researched aspect of the link between NAP binding, DNA topology, and promoter activation.

FIS has also been shown to create a nucleoprotein complex at promoters with a series of FIS-binding sites that stabilise the topological state of the DNA in ways that favour transcription initiation (Rochman et al. 2004). Many of these promoters express genes that encode components of the translational apparatus, such as ribosomal proteins, tRNA, and rRNA (Champagne and Lapointe 1998; Newlands et al. 1992; Nilsson et al. 1990). Increased translation capacity is necessary to support rapid bacterial growth, so the stimulatory role of FIS during the lag-to-log phase of the growth cycle is important. Consistent with this is the observation that while mutants that lack the FIS protein remain viable, they display reduced competitive fitness when grown in co-culture with their otherwise isogenic wild-type parent (Schneider et al. 1997).

1.39 FIS and the Stringent Response

Stable RNA (tRNA and rRNA) genes that are stimulated by FIS are subject to control by the stringent response (Condon et al. 1995b; Potrykus and Cashel 2008). Here, an intracellular signal known as an ‘alarmone’ interferes with the ability of RNA polymerase to transcribe a subset of genes, including the stable RNA genes. The alarmone is guanosine tetraphosphate (ppGpp) or pentaphosphate (pppGpp) and it is synthesised in response to a build-up of uncharged tRNA molecules and the interaction of the RelA protein with stalled ribosomes (Brown et al. 2016; Hauryliuk et al. 2015; Richter 1976) (see Section 6.18 for a more complete description of the stringent response). This accumulation results from a shortage of amino acids to charge the tRNAs and is an indication that the translational capacity of the cell exceeds demand. Hence the feedback loop that shuts down the transcription of genes involved in the production of ribosomes and other parts of the translational apparatus. The stringent response also affects DNA synthesis and mRNA translation both negatively and directly, while stimulating the transcription of genes outside the stringently regulated group (Ferullo and Lovett 2008; Haugen et al. 2006; Paul et al. 2005). What distinguishes the members of the two groups? One important factor is the possession by stringently regulated promoters of a discriminator sequence consisting of a G+C-rich DNA between its –10 and +1 elements (Figure 1.19) (Lamond and Travers 1985; Mizushima-Sugano and Kaziro 1985; Travers 1980; Travers et al. 1986; Zacharias et al. 1989). The discriminator is an effective barrier to open complex formation, possibly due to the extra hydrogen bonding between DNA strands consisting of G+C-rich sequences. Negative supercoiling of the DNA has a stimulatory effect on the promoters of stable RNA genes and this may assist with the melting of the recalcitrant discriminators when negative supercoiling is available (Schneider et al. 2000). However, in bacteria experiencing low metabolic flux (e.g. those in lag phase or stationary phase) this stimulatory influence is absent and the resulting relaxation of the DNA template, combined with the negative influences of the (p)ppGpp alarmone and the DksA protein, cooperate to repress transcription of stable RNA genes (Potrykus and Cashel 2008; Schneider et al. 2000). Genes subject to stimulation by (p)ppGpp and DksA also possess a discriminator, but in these cases this element is an A+T-rich DNA sequence (Figure 1.19) (Gummesson et al. 2013).

1.40 FIS and DNA Topology

The involvement of FIS in SIDD-based regulatory mechanisms has already been described. The promoter of the *dusB-fis* operon is subject to transcriptional stimulation by DNA negative supercoiling (Schneider et al. 2000) in addition to being auto-repressed by FIS and controlled negatively by the stringent response (Ninnemann et al. 1992). At a global level, the FIS protein is intimately associated with the general management of DNA topology in the bacterial cell. It represses the transcription of the *gyrA* and *gyrB* genes in *E. coli* (Schneider et al. 1999) and *Salmonella* (Keane and Dorman 2003) and has a complicated relationship with the promoters of the *topA* gene, where its influence is conditional on factors such as oxidative stress (Weinstein-Fischer and Altuvia 2007). Although *E. coli* and *Salmonella* have distinct DNA supercoiling set points, with *Salmonella* DNA being more

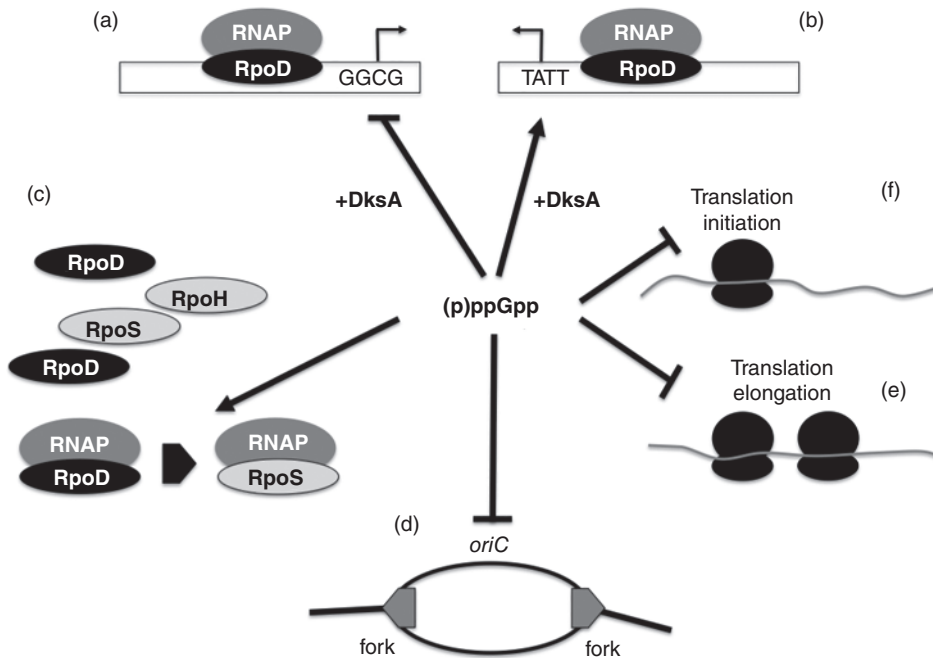


Figure 1.19 The multifaceted stringent response. A summary is shown of the processes that are inhibited or enabled by the alarmone (p)ppGpp. (a) (p)ppGpp and DksA affect a stringently regulated promoter that contains a G+C-rich discriminator sequence negatively. Genes encoding rRNA or tRNA are in this category. (b) In contrast, (p)ppGpp and DksA affect a promoter with an A+T-rich discriminator positively. Genes involved in amino acids biosynthesis are in this category. (c) The (p)ppGpp alarmone biases the selection of sigma factors by RNA polymerase away the RpoD housekeeping sigma factor and towards sigma factors that are required for various stress responses. (d) The initiation of chromosome replication is inhibited by (p)ppGpp. (e) Translation initiation and translation elongation are affected negatively because (p)ppGpp has an inhibitory influence on Initiation Factor 2 (IF2) and on the translation elongation factor, EF, respectively.

relaxed than in *E. coli* (Champion and Higgins 2007), this distinction is dependent on the presence of FIS (Cameron et al. 2011). Thus, the pattern of expression of the topoisomerases responsible for negative supercoiling (DNA gyrase) and relaxation (Topo I) of DNA is modulated by FIS. The activities of these topoisomerases is also affected by FIS because the protein influences their access to DNA by binding to it: since FIS prefers to bind to DNA with intermediate levels of negative supercoiling it acts to preserve this topological form (Schneider et al. 1997; Cameron and Dorman 2012).

In order to exert its influence on DNA topology, FIS must be present in the cell. This restricts its influence to the early stages of exponential growth when it is most abundant (Schneider et al. 1997). An exception has been discovered in bacteria growing under micro-aerobic conditions: here FIS levels are sustained into the stationary phase of growth (Cameron et al. 2013; O Cróinín and Dorman 2007). This may be of special significance in environments such as the mammalian gut epithelial surface where FIS-dependent gene expression is required for colonisation and invasion (Falconi et al. 2001; Kelly et al. 2004; Prosseda et al. 2004; Rossiter et al. 2015).

1.41 Ferritin-Like Dps and the Curved-DNA-binding Protein CbpA

While FIS is associated with the early stages of rapid exponential growth, the Dps (DNA-binding protein from starved cells) and CbpA proteins exhibit the polar opposite expression pattern and are seen predominantly in stationary phase (Ali Azam and Ishihama 1999). Both are NAPs and Dps has been studied in the most detail. CbpA expression is prevented during exponential growth by the FIS protein. FIS, which is abundant in this period of the growth cycle, binds and represses the activity of an RpoD-dependent promoter that is located in a gene (*yccE*) adjacent to *cbpA* that is partly responsible for *cbpA* transcription in stationary phase. A second promoter immediately upstream of *cbpA* depends on RpoS, a sigma factor that is only available in stationary phase or in stressed cells (Figure 1.20) (Chintakayala et al. 2013). CbpA forms dimers in solution and aggregates following binding to DNA, forming nucleoprotein complexes similar to those produced by the Dps NAP (Cosgriff et al. 2010). The preferred DNA targets of CbpA are A+T-rich and intrinsically curved; this protein has a marked preference for binding within the Ter macrodomain of the chromosome, a zone of high DNA curvature and with a high A+T content (Chintakayala et al. 2013). CbpA binds at the minor groove of DNA and *cbpA* mutants display aberrant DNA topology, observations that are consistent with a role in organising the DNA in the Ter macrodomain during stationary phase (Chintakayala et al. 2013, 2015).

Dps is dodecameric in *E. coli* and has ferritin-like properties (Grant et al. 1998). This protein accumulates in stationary phase bacteria and was found initially to protect the genomic DNA from oxidative damage (Almirón et al. 1992; Martinez and Kolter 1997). It does not impede transcription, despite being an abundant DNA-binding protein (Janissen et al. 2018). Dps was subsequently discovered also to afford protection against gamma radiation, ultraviolet light, copper and iron toxicity, heat stress, and pH shock (Algu et al. 2007; Jeong et al. 2008; Nair and Finkel 2004). It also protects DNA from cleavage by restriction enzymes (Janissen et al. 2018).

Although Dps is usually grouped with the NAPs (Ali Azam and Ishihama 1999), its relationship with DNA has been difficult to determine with precision. This protein can form a co-crystal with DNA, perhaps accounting for its ability to protect the chromosome from damage in stressed cells (Wolf et al. 1999). Through the application of SELEX (systematic evolution of ligands by exponential enrichment) to *E. coli*, a DNA sequence has been identified that seems to contain the elements of a Dps-binding site (Ishihama et al. 2016). A closely related motif has been detected in *E. coli* by chromatin immunoprecipitation on chip (Antipov et al. 2017). These Dps-binding sites overlap with those of other NAPs, leading to speculation that Dps supplies the genome architectural functions of those proteins (such as FIS) that are no longer expressed as the bacterium enters stationary phase (Antipov et al. 2017). Alternatively, Dps binding and the binding of other NAPs, such as IHF, may alternate depending on the environmental conditions that accompany entry of the bacterium into stationary phase (Lee et al. 2015).

The FIS protein plays a key role in controlling the expression of the *dps* gene during the growth cycle (Grainger et al. 2008). This gene is transcribed from a single promoter that is recognised by both the RpoD- and the RpoS-containing forms of RNA polymerase (Altuvia et al. 1994). Three NAPs are involved in the regulation of *dps* transcription: FIS, H-NS,

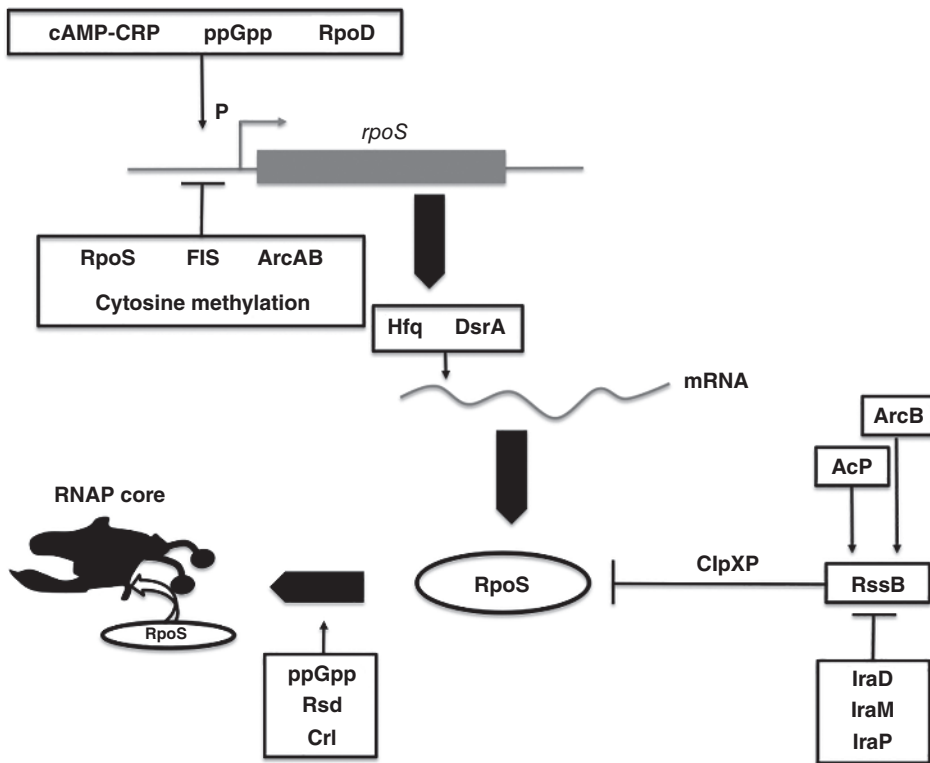


Figure 1.20 The stress and stationary phase sigma factor, RpoS. The *rpoS* gene is influenced at the transcriptional level by several factors. The glucose-sensitive cAMP-CRP complex and the stringent-response signal, ppGpp, act at the RpoD-dependent promoter to enhance *rpoS* transcription; production of RpoS is enhanced in mutants deficient in cytosine methylation (see Kahramanoglou et al. 2012). It is important to note that *rpoS* is expressed under all growth conditions and that the principal regulatory effects are imposed at the level of RpoS protein stability; rapidly growing bacteria have few copies of RpoS and non-growing bacteria have many. The expansion of the population of RpoS proteins occurs in bacteria when growth is slowed or stopped due to stress. The stress can be physical or chemical in nature. Once transcribed, RpoS mRNA is translated poorly due to the formation of secondary structures that sequester the translation initiation signals. These stem-loops are eliminated by the DksA sRNA that binds to the 5' end of the mRNA in the presence of the Hfq RNA chaperone protein. DksA also controls the translation of the *hns* transcript, albeit negatively due to sequestration of the translation initiation signals. The RpoS protein is degraded by the ClpXP protease. Proteolytic cleavage of RpoS is enhanced by the adaptor protein, RssB. RssB activity is in turn modulated negatively by the sRNAs IraD, IraM, and IraP in response to stresses that impede the growth of the bacterium. In *Salmonella*, IraM is called RssC. RssB activity is controlled in response to changes to oxygen supply (ArcB) and carbon levels (Acp). The sigma factor must compete with RpoD and other sigma factors for access to the core RNA polymerase and it is assisted in doing so by ppGpp, Crl, and the anti-sigma factor, Rsd.

and IHF. IHF and the OxyR transcription factor activate the *dps* promoter in association with RpoD in bacteria experiencing oxidative stress; in stationary phase the same promoter is utilised by RNA polymerase containing RpoS (Altuvia et al. 1994). During exponential growth, FIS holds RNA polymerase containing RpoD at the promoter while H-NS excludes the RpoD form of the RNA polymerase holoenzyme from the same promoter. The transcriptionally inert FIS-RpoD-RNA-polymerase complex prevents entry of the RpoS-containing RNA polymerase holoenzyme. Since FIS levels decline to negligible values as the bacterium enters stationary phase, this negative control is no longer exerted, and as H-NS is unable to impede the activity of the RpoS-containing form of RNA polymerase, *dps* transcription can begin (Grainger et al. 2008).

1.42 The H-NS Protein: A Silencer of Transcription

Throughout the 1980s, the gene that encodes the H-NS NAP was discovered and rediscovered by investigators working independently of each other because this protein is involved in controlling the expression of so many different components of the bacterium. One of the consequences of the broad influence of H-NS is that the gene that encodes it has been given many names, such as *bglY*, *osmZ*, *pilG*, *virR*, and *dxrR*, among others. In each case, the name linked a mutation in the gene, now referred to universally as *hns*, to a specific H-NS-dependent system in the cell such as beta-glucoside uptake and utilisation (*bglY*), pilus expression (*pilG*), the osmotic stress response (*osmZ*), expression of a virulent phenotype in the pathogen *Shigella* (*virR*), or thermo-regulated adhesin expression in *E. coli* (*drdX*) (Defez and de Felice 1981; Higgins et al. 1988; Göransson et al. 1990; Maurelli and Sansonetti 1988; Spears et al. 1986). The *hns* gene is located in the Ter macrodomain of the chromosome, close to the *topA* gene, and early experiments revealed a connection between some *hns* alleles and alterations in the DNA topology of reporter plasmids in those strains (Dorman et al. 1990; Higgins et al. 1988). The finding that H-NS binding to DNA impedes access to that DNA by DNA gyrase may explain these observations (Sutormin et al. 2019). H-NS can influence transposition and site-specific recombination as well as transcription (Corcoran and Dorman 2009; Liu et al. 2011; Whitfield et al. 2009) and it has effects at the level of mRNA translation too (Park et al. 2010).

The H-NS protein is a small, abundant NAP that is produced at all stages of the growth cycle (Ali Azam and Ishihama 1999; Dorman 2013; Free and Dorman 1995). It binds to A+T-rich DNA and has been described as having a preference for DNA with intrinsic curvature (Yamada et al. 1991). These features are commonly associated with transcriptional promoters, allowing H-NS to target large numbers of genes for repression, even if the functions of the gene products are not related in any obvious way. The protein is almost always a repressor, allowing it to silence transcription of a large subset of the genes in the genome (Figure 1.21).

The very low base sequence requirement, as opposed to base content requirement, of H-NS for its binding sites allows this protein to interact with DNA on the basis of DNA shape rather than nucleotide sequence. Its reliance on indirect readout for DNA-binding-site recognition underlies the contribution made by H-NS to silencing the transcription of genes that have been acquired by HGT in Gram-negative organisms (Dorman 2007, 2014a). This xenogeneic silencing is hypothesised to allow potentially

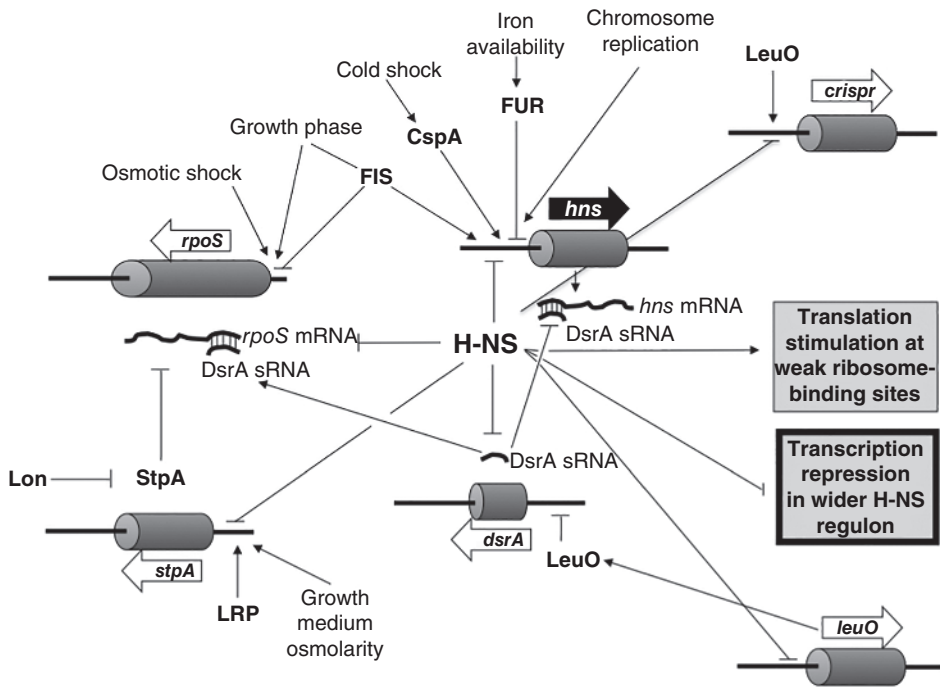


Figure 1.21 The vast H-NS regulon. The H-NS protein controls the expression of hundreds of genes in the pan genome, the accessory genome, and the core genome. Its own gene is also subject to complex control at the transcriptional and posttranscriptional levels. Among the factors influencing *hns* transcription positively are chromosome replication, the growth-phase-dependent FIS protein, and the cold-shock regulatory protein CspA; it is auto-repressed by its own gene product and repressed by the iron-binding Fur protein, and translation of H-NS's own mRNA is inhibited by the DsrA sRNA in an Hfq-dependent manner. H-NS, in turn, inhibits DsrA expression, resulting in a positive effect throughout the DsrA regulon, including production of the RpoS sigma factor. H-NS silences *crispr* transcription, an effect that is antagonised by the LeuO transcription factor, with the *leuO* gene in turn being H-NS-repressed. Like StpA (whose gene H-NS represses) the H-NS protein can act at the level of RNA, for example improving the translation of the maltose regulatory gene by repositioning ribosomes so that translation can proceed. The H-NS paralogue, StpA, is a substrate for Lon-mediated proteolysis but dimerisation with H-NS protects StpA from this fate (H-NS is not degraded by Lon). StpA can act as a backup for H-NS and its expression is governed by an independent set of cues, such as the LRP protein and bacterial growth phase (H-NS is present at all stages of growth). The H-NS regulon includes genes involved in bacterial virulence and whose expression is controlled by pH, osmotic stress, temperature, and other environmental influences. Its (usually negative) influence is overcome by an impressive array of mechanisms that link the expression of the H-NS target genes to information relevant to the infection process (see Stoebe et al. 2008b). Arrows represent positive regulatory inputs and negative ones are indicated by 'T' symbols.

harmful genes to be imported into the genome without compromising the competitive fitness of the bacterium (Dorman 2004, 2007; Lucchini et al. 2006; Navarre et al. 2006; Oshima et al. 2006). It should be pointed out that although H-NS displays low nucleotide sequence specificity at its binding sites, a preferred sequence has been identified at the *proU* operon (Bouffartigues et al. 2007; Lang et al. 2007). This operon encodes a transport

system for glycine-betaine, a modified amino acid that protects macromolecules in bacteria from the deleterious effects of water loss following osmotic shock (Cairney et al. 1985). The *proU* promoter is induced by osmotic up-shock in the presence of potassium ions and is subject to repression by the H-NS protein (Gowrishankar and Manna 1996; Higgins et al. 1988; Sutherland et al. 1986). It contains two matches to an A+T-rich consensus sequence that contains the very flexible TpA di-nucleotide step, possibly important in specifying local DNA curvature (Bouffartigues et al. 2007; Lang et al. 2007). This motif may represent an evolutionary step towards a more specific binding site for the H-NS protein.

H-NS forms dimers and these can link together to form higher-order oligomers (Arold et al. 2010; Shahul Hameed et al. 2018) (Figure 1.22). The monomer consists of an amino-terminal dimerisation domain that is connected by a linker region to a carboxyl-terminal nucleic-acid-binding domain (Dorman et al. 1999). The dimer can bridge different parts of the same DNA molecule or link separate DNA molecules together, activities that have been confirmed *in vitro* using single molecule methods and atomic force microscopy (Dame et al. 2000, 2005, 2006). The protein displays two DNA-binding modes, one involving bridging of two DNA duplexes (or two sections of the same DNA duplex) and the other involving polymerisation of H-NS along a single DNA duplex (Figure 1.23) (Corcoran and Dorman 2009; O’Gara and Dorman 2000; Lim et al. 2012). Polymerisation, with or without bridging, occurs following initial binding of H-NS to a

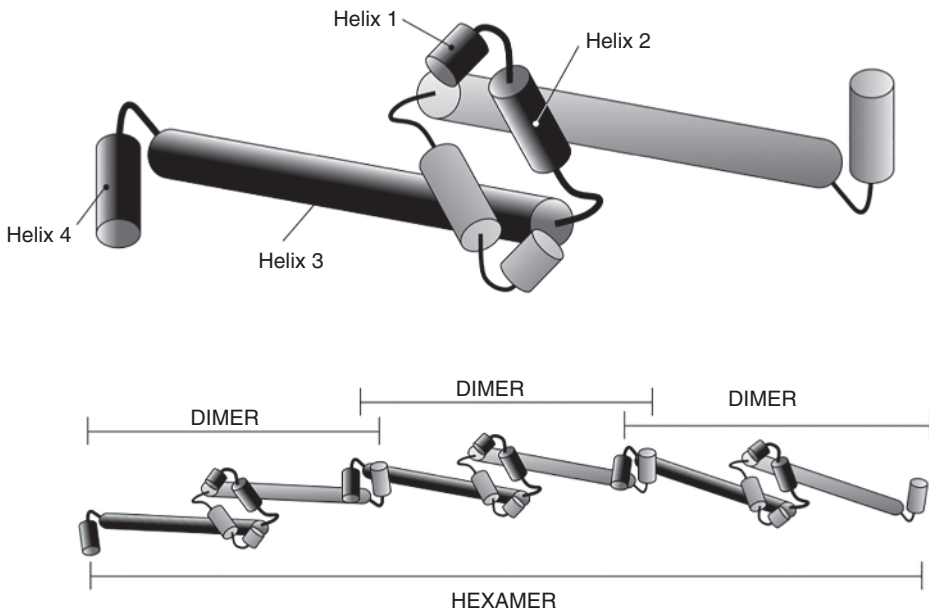


Figure 1.22 The dimerisation and oligomerisation of H-NS. The amino-terminal domain of H-NS is shown, with its four alpha helices represented by linked cylinders. The carboxyl terminal domain with its nucleotide binding activity and the linker region that connects this domain to the amino-terminus are omitted. Interlinkage of alpha helices 1 and 2 of one monomer with the corresponding parts of another allows dimer formation. Linking alpha helices 4 allows two dimers to form a tetramer, and this process can continue indefinitely to produce an H-NS polymer. For further information, see Arold et al. (2010).

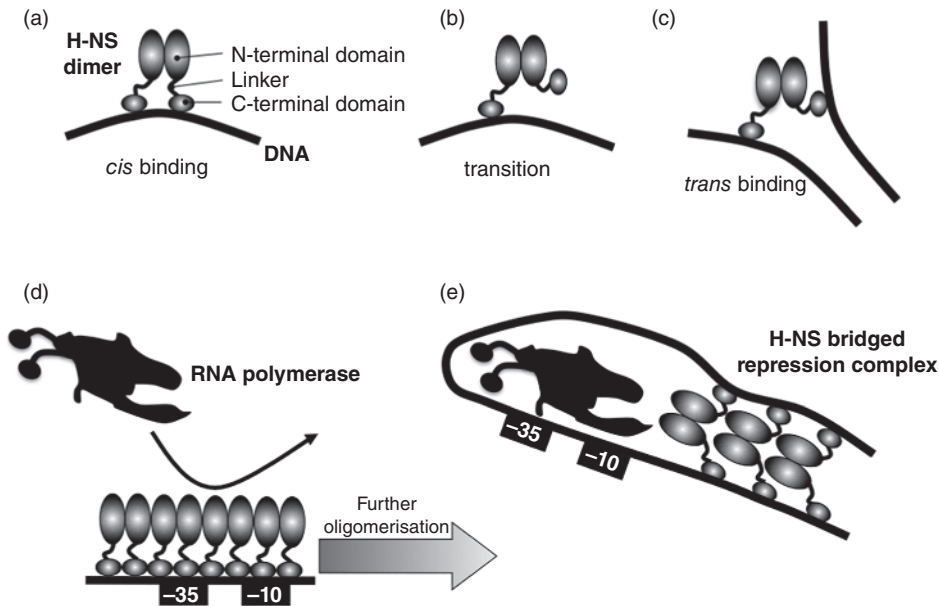


Figure 1.23 The DNA-binding modes of H-NS. The H-NS monomer is shown as two linked ovals with one oval representing the amino-terminal dimerisation/oligomerisation domain and another the carboxyl-terminal domain with the nucleic acid binding activity. The functional domains are connected by a flexible linker. (a) An individual dimer is shown in a *cis*-binding mode, with both DNA-binding modules on the same DNA molecule. (b) It then undergoes a transition to a *trans*-binding mode in which it attaches simultaneously to two DNA molecules or to two separate sections of the same DNA molecule. (c) This *trans*-binding mode is the basis for H-NS-mediated DNA bridging and the transition between the two modes is toggled by Mg^{2+} concentrations. Either mode can silence transcription. (d) In one case polymerisation of H-NS along DNA excludes RNA polymerase from a promoter. (e) In the other case, H-NS bridges DNA from upstream and downstream of the transcription start site, blockading the promoter so that RNA polymerase is either excluded or becomes trapped in the initiation stages of transcription.

preferred nucleation site (Badaut et al. 2002). *In vitro* data indicate that changes in the concentration of magnesium ions can toggle H-NS between these two binding modes (Liu et al. 2010), though the significance of this effect in the bacterial cell is unknown. In addition to the influence of divalent cations, the toggling of H-NS between these two binding modes is modulated by co-regulator proteins such as Hha, YdgT, and YmoA, which resemble the dimerisation domain of H-NS (van der Valk et al. 2017). By coating the DNA surface, H-NS can exclude other proteins or protein complexes, including RNA polymerase (Figure 1.23). However, the negative influence of H-NS on transcription can be overcome in a variety of ways (Dorman and Kane 2009; Stoebel et al. 2008b).

H-NS displacement can be assisted by properties of the DNA to which it binds. In the genetic switch that controls the expression of the major virulence genes of *Shigella flexneri*, a region of curved DNA undergoes a thermally induced reorganisation that displaces H-NS, leading to derepression of the *virF* regulatory gene (Prosseda et al. 2004). H-NS may also use an internal conformational change in response to rising temperature to disrupt its transcription-silencing complexes on DNA (Shahul Hameed et al. 2018). However, most

of the anti-H-NS mechanisms that have been described involve roles for a second protein. Also in *S. flexneri*, the VirB DNA-binding protein derepresses the promoters of the principal virulence gene operons by binding and wrapping DNA, making H-NS binding and bridging untenable (Turner and Dorman 2007). This form of antagonism can be transferred to other H-NS-repressed promoters simply by appropriate placement there of the VirB-binding site (Kane and Dorman 2011). The similarity of VirB to proteins involved in the partitioning of plasmids with protein-dependent active partitioning systems (Section 1.15) shows that DNA-binding molecules can be co-opted from other purposes to oppose the transcription silencing imposed by H-NS (Adler et al. 1989; Kane and Dorman 2012; Turner and Dorman 2007). This suggests that simple genetic switches based on transcription silencing and anti-silencing can arise quickly in the evolution of gene regulatory circuits.

The DNA-binding and bridging activity of H-NS may lend itself to nucleoid organisation as well as to transcription control (Dorman 2013; Japaridze et al. 2017). The distribution pattern of H-NS binding sites around the chromosome appears to be periodic and genetic experiments have identified the *hns* gene as being important for the formation of chromosomal domain boundaries (Hardy and Cozzarelli 2005). Experiments with super resolution imaging and chromosome conformation capture have produced some data that indicate a central role for H-NS in nucleoid architecture (Wang et al. 2011) while other 3C experiments seem to rule out a role for this protein in the organisation of long range interactions in the nucleoid (Cagliero et al. 2013). Other HiC experiments show that H-NS promotes DNA interactions in the chromosome at short ranges, but not the long-range interactions between H-NS-regulated genes and operons that were detected by Wang et al. (2011) (Lioy et al. 2018). Work with HiC in *Caulobacter* has implicated the high-frequency transcription of long genes as playing a role in boundary formation and maintenance, independent of any effect of translation (Le and Laub 2016). This work suggests that the act of transcribing a long gene inhibits interactions between the DNA regions that flank that gene, causing this long gene to act as a boundary element between domains, a process that may become modulated by H-NS-mediated transcription silencing.

1.43 **StpA: A Parologue of H-NS**

The ratio of H-NS to its genomic binding sites appears to be important for the competitive fitness of the bacterial cell. Bacteria that gain extra copies of the *hns* gene through the introduction of multicopy recombinant plasmids that encode this NAP gain in their capacity to replicate themselves (C.J. Dorman, unpublished). The model organisms *E. coli* and *Salmonella typhimurium* encode the StpA protein, a second H-NS-like molecule. It shares many features with H-NS and has been described as an RNA chaperone (Doetsch et al. 2010). StpA can substitute for H-NS and it can form heterodimers with its parologue (Johansson et al. 2001; Leonard et al. 2009; Sonden and Uhlin 1996; Zhang et al. 1996 Refs). The *stpA* gene is expressed maximally during exponential growth, perhaps to provide an auxiliary supply of H-NS-like protein at a point in the growth cycle where the number of H-NS binding sites is most numerous (Deighan et al. 2003; Free and Dorman 1997). In this context, it is interesting to note that the transcription of the *hns* gene is linked positively to chromosome replication (Free and Dorman 1995).

1.44 H-NS Orthologues Encoded by Plasmids and Phage

Copies of genes encoding H-NS-like proteins occur naturally on plasmids, including large, self-transmissible plasmids (Shintani et al. 2015). These molecules have A+T-rich DNA and may impose a competitive fitness cost on bacteria that receive them in conjugation by diverting H-NS from A+T-rich DNA sites on the chromosome, leading to disruption of the gene expression pattern in the bacterium. By providing its own supply of H-NS activity, the plasmid can avoid this regulatory disturbance and its associated impact on competitive fitness (Doyle et al. 2007). The R27 plasmid, originally detected in *Salmonella*, and more recently in *S. flexneri* 2a 2457T, encodes the H-NS orthologue Sfh (Beloin et al. 2003a; Deighan et al. 2003). The two proteins exhibit considerable overlaps in their binding sites on the bacterial chromosome. When H-NS is present, Sfh is restricted to a subset of the sites that it can occupy when H-NS is removed by inactivation of the *hns* gene (Dillon et al. 2012). These observations are consistent with Sfh acting as an auxiliary to H-NS, a role that it shares with StpA. The expression patterns of the three proteins are instructive in this regard: H-NS is present at a constant level per chromosome throughout the growth cycle, StpA is expressed when the cells are in exponential phase, and Sfh appears at the beginning of stationary phase (Deighan et al. 2003).

H-NS-like paralogues encoded by self-transmissible plasmids (Shintani et al. 2015) and bacteriophage (Skenner et al. 2011) can downregulate the expression of CRISPR-*cas* loci, allowing the mobile genetic element to evade the host immune system (Dillon et al. 2012; Lin et al. 2016; Medina-Aparicio et al. 2011; Pul et al. 2010). The LysR-like transcription factor LeuO overcomes H-NS-mediated repression of the CRISPR-*cas* locus, but the *leuO* gene is itself silenced by H-NS, and its paralogues (Dillon et al. 2012; Medina-Aparicio et al. 2011; Pul et al. 2010). Stochastic upregulation of *leuO* transcription may provide a mechanism for overcoming silencing of the immunity function in some cells in the population that encounter plasmid or bacteriophage invaders.

1.45 H-NSB/Hfp and H-NS2: H-NS Homologues of HGT Origin

Genes encoding proteins related to H-NS are found in pathogenicity islands that have been acquired by HGT. Hfp/H-NSB is an H-NS-like protein that is expressed by a gene in the *serU* pathogenicity island in the chromosome of uropathogenic *E. coli* (Müller et al. 2010; Williams and Free 2005). It can form heterodimers with H-NS and may modulate its activity in helping UPEC to adapt to environmental conditions encountered during the infection process (Dorman 2010; Müller et al. 2010). The H-NSB/Hfp protein is also encoded by a chromosomal island in *E. carotovora* but is absent from an island in enteropathogenic *E. coli* (EPEC) that is closely related to the *serU* island of UPEC (Williams and Free 2005). This association of a gene encoding an H-NS-like protein with a former mobile genetic element is reminiscent of similar associations with elements that are currently mobile such as self-transmissible plasmids and bacteriophage (Section 1.44). The absence of the *hnsB* gene from the version of the *serU* island that is present in EPEC is intriguing, given that *hnsB* is both present and expressed in the corresponding island in UPEC. Perhaps the gene had performed its role once the EPEC island was established in the chromosome and it

was subsequently lost in the absence of a selective pressure to keep it? If so, the different environmental circumstances experienced by EPEC and UPEC seem to have selected for retention of *hnsB* by the latter organism.

Enteroaggregative *E. coli* (EAEC) strains express, in addition to H-NS and StpA, an H-NS2 protein that is closely related to H-NS. H-NS2 behaves somewhat like H-NS when the latter is in a complex with Hha: it targets A+T-rich genes that have been acquired by HGT and silences them transcriptionally (Prieto et al. 2018). The amino acid sequence of H-NS2 is similar to those of H-NSB and Hfp, but differs from them in a number of respects. It does not exhibit the sensitivity to proteolytic turnover that is a characteristic of these H-NS homologues and StpA (Prieto et al. 2018). It is possible, and plausible, that H-NS2 and other ‘third homologues’ could form heteromeric complexes with H-NS or StpA that have distinct activities from those of the homodimers. Certainly, H-NS heterodimers with StpA have properties that are distinct from those of the homodimers (Johansson et al. 2001; Leonard et al. 2009), so expanding the number of interacting partners may represent a way of modulating NAP function (Beloin et al. 2003a; Sonden and Uhlin 1996; Zhang et al. 1996).

1.46 A Truncated H-NS-Like Protein

The *serU* island in UPEC that encodes H-NSB/Hfp also encodes a protein that resembles a truncated H-NS. This is H-NST and its gene is tightly linked to the *hnsB* gene in the chromosomal island. H-NST consists of the first 80 amino acids of H-NS and the corresponding island in EPEC encodes a closely related protein; EAEC also encodes a relative of H-NST (Williams and Free 2005). H-NST from UPEC can form a heterodimer with H-NS and it can antagonise its activity as a transcription silencer. The corresponding protein from EPEC is much attenuated in its ability to interact with H-NS and to attenuate its biological activity: a key substitution at residue 16 of the amino acid sequence seems to be responsible for this difference between the UPEC and EPEC H-NSTs (Williams and Free 2005).

The action of H-NST recalls that of the gene 5.5 protein that is encoded by bacteriophage T7. Like H-NST, the gene 5.5 protein co-purifies with H-NS and antagonises the transcription silencing activity of H-NS, presumably to the benefit of the phage (Liu and Richardson 1993). H-NST and the gene 5.5 protein resemble one another in size and mode of action structure but not in amino acid sequence (Williams and Free 2005). The ability of H-NST from EPEC to inhibit H-NS activity has been exploited to explore the H-NS⁻ phenotype of *Yersinia enterocolitica*, a bacterium where H-NS is essential (Baños et al. 2008). The essential nature of H-NS in *Y. enterocolitica* probably reflects the absence of a paralogous protein such as StpA that can offset the severe phenotype associated with the loss of H-NS. Expressing EPEC H-NST ectopically in *Y. enterocolitica* titrates the transcription silencing of H-NS, revealing that it has similar effects on global gene expression patterns to those seen in other Gram-negative bacteria such as *E. coli* (Baños et al. 2008).

1.47 Hha-like Proteins

H-NST’s close similarity to the oligomerisation domain of H-NS and its ability to form heteromeric complexes with H-NS is superficially similar to the relationship of H-NS to

the Hha family of proteins. Found only in the *Enterobacteriaceae*, these too mimic the oligomerisation domain of H-NS in their structure but unlike H-NST, their interaction does not interfere with H-NS-mediated transcription silencing. Instead it channels the negative influence of H-NS to certain target promoters (Baños et al. 2009; Madrid et al. 2007).

Y. enterocolitica possesses just one housekeeping Hha-like protein (called Hha), in contrast to other model organisms like *E. coli* and *Salmonella* that have both Hha and a closely related paralogue, YdgT. However, pathogenic strains of *Yersinia* express YmoA (*Yersinia* modulator) from a chromosomal locus. YmoA is a founding member of the Hha protein family that regulates virulence genes negatively in *Yersinia* spp. (Cornelis et al. 1991; de la Cruz et al. 1992). It does this by forming a complex with H-NS in which H-NS provides the DNA-binding activity (Ellison and Miller 2006b). YmoA potentiates the transcription repression activity of H-NS, targeting virulence gene promoters in *Yersinia* (Ellison and Miller 2006b). It shares this property with Hha itself and with the Hha paralogue YdgT (Nieto et al. 2002; Starke and Fuchs 2014). YdgT is a paralogue of Hha and shares with Hha an ability to form heteromeric complexes with H-NS and StpA (Paytubi et al. 2004).

YmoA is structurally closely related to Hha, it mimics the oligomerisation domain of H-NS (McFeeters et al. 2007), and it is turned over by Lon- and ClpXP-mediated proteolysis (Jackson et al. 2004). YmoA and Hha each interact with an H-NS dimer, stabilising the transcription-silencing complex at target promoters (Cordeiro et al. 2015).

In the case of *Salmonella*, Hha-like proteins target H-NS to the major virulence genes in the SPI1 and SPI2 pathogenicity islands and on the *Salmonella* virulence plasmid, pSLT (Silphaduang et al. 2007; Vivero et al. 2008). They can also influence the DNA-binding mode of H-NS and whether this protein forms polymers along DNA or creates bridges between different segments of DNA (van der Valk et al. 2017). Hha and YdgT direct H-NS towards horizontally acquired genes, causing them to be silenced preferentially (Aznar et al. 2013). In *E. coli*, Hha/YdgT also targets horizontally acquired genes via H-NS/StpA binding, together with genes involved in the osmotic and carbon starvation stress responses (Ueda et al. 2013). Genes encoding Hha-like proteins also occur on self-transmissible plasmids and their products can interact with H-NS. For example, the virulence plasmid in the pathogenic *E. coli* strain O157:H7 encodes a form of Hha that directs H-NS to just a subset of the targets that are bound by H-NS when in a complex with the chromosomally encoded Hha protein (Paytubi et al. 2013). Plasmid-encoded H-NS proteins, such as the one expressed by the R27 self-transmissible plasmid in *Salmonella*, target genes of HGT origin in the chromosome, a task that requires the Hha helper protein when it is performed by the chromosomally encoded H-NS protein (Baños et al. 2009).

1.48 Other H-NS Homologues: The Ler Protein from EPEC

Ler is a paralogue of H-NS that controls virulence gene expression in EPEC (Bustamante et al. 2001) and Enterohaemorrhagic *E. coli*, EHEC (Hansen and Kaper 2009). The principal Ler-dependent virulence genes are located in the locus of enterocyte effacement, LEE, an A+T-rich pathogenicity island that has been acquired by HGT (Clarke et al. 2003; Nataro and Kaper 1998) (Figure 1.24). LEE encodes a type 3 secretion apparatus and its effector proteins and consists of five operons: *LEE1* to *LEE5*, with the *ler* gene being

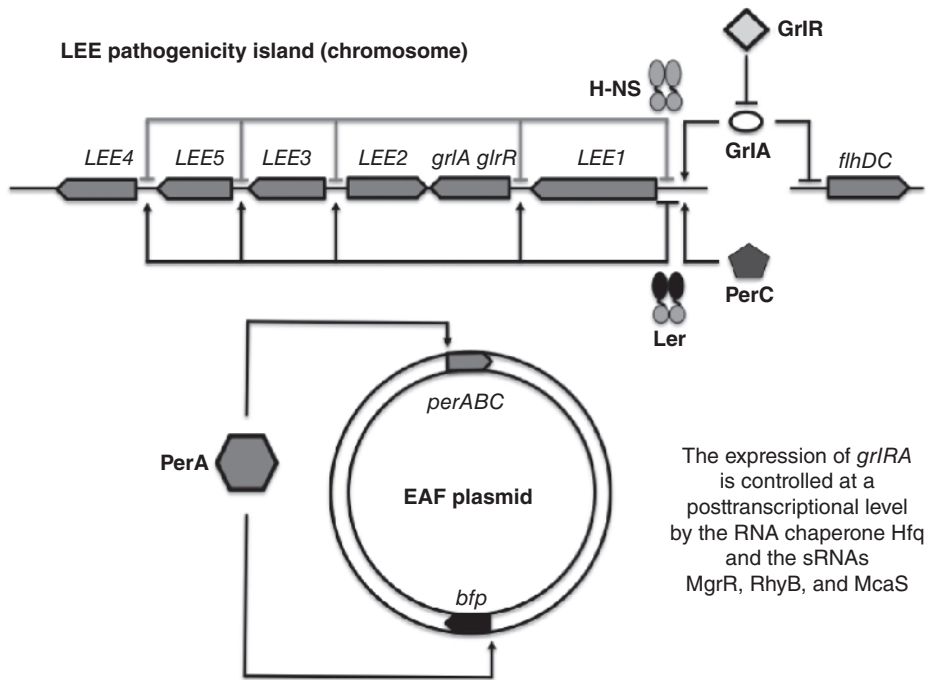


Figure 1.24 The genetic switch controlling LEE virulence gene expression in EPEC. The LEE pathogenicity island in the EPEC chromosome encodes a T3SS and its associated effector proteins for mammalian epithelial cell invasion. Transcription of the LEE operons is silenced by the H-NS protein and this is relieved by the antagonistic action of the Ler DNA-binding protein, an H-NS homologue. The *ler* gene is the first in the *LEE1* operon and its transcription is negatively autoregulated by Ler and positively controlled by GrlA (encoded by the LEE island) and the PerC protein that is expressed from the EAF virulence plasmid. Production of PerC depends on the PerA regulatory protein, which also activates the transcription of the *bfp* operon on the EAF plasmid (this operon expresses the bundle-forming pilus that attaches EPEC to its target host cells). GrlA is a negative regulator of the *flhDC* operon, the master switch for bacterial motility: when LEE is activated, the bacteria cease to be motile. GrlA activity is inhibited by the LEE-encoded GrlR protein. Ler and H-NS share a common nucleation site and a competition between these proteins determines whether or not LEE genes are expressed. The concentration of the Ler protein reflects the activities of the GrlA, PerC, Ler, and H-NS proteins at the *ler* promoter. In enterohaemorrhagic *E. coli* strains, the Hfq RNA chaperone negatively regulates *grlRA* post-transcriptionally and exerts a GrlA-independent negative effect on LEE gene expression in stationary phase (Hansen and Kaper 2009). In EPEC, Hfq exerts its effects at *grlRA* mRNA through the sRNAs MgrR, RhyB, and McaS (see Bhatt et al. 2017).

located in *LEE1* (Elliott et al. 1998). H-NS silences transcription throughout LEE (Leh et al. 2017; Shin et al. 2012). Ler, the LEE-encoded regulator, controls LEE operon expression positively by relieving the transcriptional silencing that is imposed by the H-NS NAP in LEE (Umanski et al. 2002) and at some other loci (Bingle et al. 2014; Elliott et al. 2000).

The LEE system is under complex control that involves plasmid-encoded regulators in addition to those encoded by genes on the chromosome (Figure 1.24). The EPEC Adherence Factor (EAF) plasmid encodes Bfp type IV (bundle-forming) pili for adherence to host cells

and a *perABC* operon that is autoregulated through the PerA protein and cross-regulates the *ler* gene in *LEE1* via the PerC protein. PerC is redundant with the island-encoded GrlA regulatory protein, another positive regulator of *ler* transcription. In addition, GrlA represses expression of the flagella master regulator locus, *flhDC* (Kitagawa et al. 2011; Padavannil et al. 2013). The *grlRA* operon is the subject of posttranscriptional regulation by the Hfq RNA chaperone (Bhatt et al. 2017; Hansen and Kaper 2009). GrlR is an inhibitor of GrlA activity and it achieves this effect by direct protein–protein interaction at the GrlA HTH motif (Padavannil et al. 2013).

Upregulation of the *LEE* operons occurs in minimal medium at 37°C and transcription is repressed in EPEC cells growing in LB. Two switches operate in the system to create physiological variety. In one, the PerA protein activates the *bfp* adherence operon on the EAF plasmid directly and the *LEE* genes indirectly through PerC (Figure 1.24). Positive auto-control of *perABC* transcription by PerA is stochastic, creating sub-populations of bacteria that maintain a hypervirulent (LEE^+) phenotype even if inducing conditions are removed and a second population of non-virulent (LEE^-) cells. Allowing the culture to reach stationary phase resets the system (Ronin et al. 2017).

The second switch involves a competition between Ler and H-NS for access to a nucleation site upstream of the *LEE5* promoter (Leh et al. 2017) (Figure 1.24). Although Ler and H-NS are paralogues, they create distinct nucleoprotein complexes when they bind to DNA, with the Ler complex favouring transcription and the H-NS one causing transcriptional silencing (Leh et al. 2017). Stochastic expression of the *perABC* operon, with downstream effects via PerC on *ler* transcription, may be expected to tip the balance in the Ler/H-NS competition back and forth, leading to LEE^+ and LEE^- phenotypes among members of the EPEC population.

Ler is not a general antagonist of H-NS because it binds only to a small subset of H-NS targets in the genome, mostly those associated with the *LEE* pathogenicity island. The two proteins are dissimilar in amino acid sequence at their N-termini but share similar C-terminal domains, including the nucleic acid-binding domain. However, a key arginine residue, found in Ler but not H-NS, seems to underlie the more restricted range of Ler binding in DNA. Both proteins rely on an indirect readout mechanism for binding site recognition: in the case of Ler, the introduction of its arginine residue into the minor groove of DNA is permitted at only a subset of H-NS binding sites (Cordeiro et al. 2011). This represents an interesting example of specialisation within the large family of H-NS-like proteins.

1.49 H-NS Functional Homologues

Proteins performing a foreign-gene-silencing function analogous to that associated with H-NS seem to be restricted to bacteria and to fall into four classes: H-NS itself, Rok, MvaT, and Lsr2. In contrast, other types of NAP are widely distributed among prokaryotes. When a bacterium possesses one type of xenogeneic silencer, it typically will not also have an example of a different type, indicating specialisation between each protein type and its genome (Perez-Rueda and Ibarra 2015).

1.50 H-NS Functional Homologues: Rok from *Bacillus* spp.

The Rok protein was discovered in *B. subtilis* during an investigation of gene regulation in the competence system: Rok emerged as a transcription silencer of *comK*, the autoregulated master controller of competence (Hoa et al. 2002). Rok controls the expression of an extensive regulon of genes (Albano et al. 2005) and at some of its gene targets its activity is amplified by co-binding of the DnaA protein (Seid et al. 2017). Rok binds to A+T-rich DNA targets (Smits and Grossman 2010) and, like H-NS, it has been implicated in the silencing of genes that have been acquired by HGT (Duan et al. 2018). Rok exhibits a higher preference for specific DNA sequences than other xenogenic silencer proteins (e.g. H-NS) and these targets are relatively rare in the *B. subtilis* core genome, allowing Rok to focus on imported genes (Duan et al. 2018). Rok binds only in the DNA minor groove and uses a winged helix fold to do this. It avoids rigid poly-A tracts with their very narrow minor grooves (Rohs et al. 2009), preferring 5'-AACTA-3' and 5'-TACTA-3' (both underrepresented in the core genome) and sequences that contain the flexible TpA step (Duan et al. 2018; Travers 2005).

1.51 H-NS Functional Homologues: Lsr2 from Actinomycetes

The 12-kDa Lsr2 NAP has been described as a functional analogue of H-NS in actinomycetes, including *Mycobacterium* spp. (Datta et al. 2019a; Kriel et al. 2018). It targets genes that have high A+T content that are thought to have been acquired by HGT (Gordon et al. 2010). Like H-NS, it can form DNA–protein–DNA bridges (Chen, J.M., et al. 2008), has a similar domain structure, and it can substitute functionally for H-NS (Gordon et al. 2008). Lsr2 and H-NS also bind DNA in the minor groove through a similar mechanism: using a so-called AT-hook-like grip (Gordon et al. 2011). The similarities between Lsr2 and H-NS are primarily functional and probably arose by convergent evolution: their amino acid sequences and their DNA-binding domains have distinct tertiary structures (Gordon et al. 2010, 2011; Shindo et al. 1995). Like H-NS, Lsr2 can polymerise along DNA to form stiff nucleoprotein structures from which other DNA-binding proteins are excluded (Qu et al. 2013).

Rv3852 is a 13.8-kDa protein that is highly conserved among *Mycobacterium* spp. and has been annotated as H-NS because its N-terminus resembles histone 1 from humans (Cole et al. 1998). Rv3852 is not an essential protein and a careful study of its properties rules out a role for it in controlling the virulence phenotype of *Mycobacterium tuberculosis* or in compacting the bacterial nucleoid (Odermatt et al. 2017).

1.52 H-NS Functional Homologues: MvaT from *Pseudomonas* spp.

Identified originally as a transcription regulator of *mvaAB*, an operon involved in mevalonate metabolism in *Pseudomonas mevalonii* (Rosenthal and Rodwell 1998), MvaT is now recognised as a NAP with properties analogous to those of H-NS (Castang and Dove 2010; Tendeng et al. 2003; Winardhi et al. 2012). MvaT binds to AT-rich DNA in genes that have

been acquired by HGT but it uses a binding mechanism that is distinct from other xenogeneic silencers: MvaT prefers binding sites that contain a series of flexible TpA steps and is tolerant of GC interruptions to the target sequence (Ding et al. 2015). MvaT has a paralogue, MvaU, with which it can form heteromeric complexes (Castang et al. 2008). Like MvaT, MvaU can bridge DNA and form filaments along the DNA that exclude other DNA-binding proteins, enabling it to silence transcription (Winardhi et al. 2014). Mutants deficient in these proteins have altered phenotypes affecting prophage activation, pyocyanin expression, biofilm production, and the elaboration of surface fimbriae (Li et al. 2009; Vallet et al. 2004; Vallet-Gely et al. 2005).

Genes encoding MvaT-like proteins are found on self-transmissible plasmids and these proteins influence the transcriptome of the host cell in cooperation with their chromosomally encoded counterparts (Yun et al. 2015). Some bacteria express multiple members of the MvaT family; for example, *Pseudomonas putida* KT2440 encodes five MvaT orthologues: TurA, TurB, TurC, TurD, and TurE (Renzi et al. 2010). TurC, TurD, and TurE have species-specific properties while TurA and TurB are similar to MvaT proteins found in all members of the *Pseudomonadaceae*. TurB is reported not to act generally as a repressor and to affect a smaller group of genes than TurA. These findings illustrate the versatile nature of MvaT-like proteins and their capacity to acquire new functions through evolution (Renzi et al. 2010).

1.53 The Leucine-responsive Regulatory Protein, LRP

The leucine-responsive regulatory protein (LRP) DNA-binding protein affects the expression of about 10% of the protein-encoding genes in *E. coli*, many of which are involved in determining the structure of the bacterial surface, in transport, in metabolism, and in adaptation to stationary phase (Cho, B.K., et al. 2008, 2011; Engstrom and Mobley 2016; Tani et al. 2002). More recent data, based on ChIP-seq and RNA-seq analyses, have led to a revision of the estimate of LRP's influence to up to 38% of the *E. coli* genome (Kroner et al. 2019). In many cases, LRP interacts with target promoters in a poised mode, not influencing promoter activity until it operates in combination with other regulatory proteins; it also shifts between more- and less-sequence specific DNA-binding modes in response to nutrient signals (Kroner et al. 2019).

LRP contributes to the genetic switches that govern the phase-variable expression of Pap and type 1 fimbriae in *E. coli* (and fimbriae in *Salmonella*), linking LRP to bacterial virulence and to biofilm formation (Aviv et al. 2017; Hernday et al. 2004; Kelly et al. 2009; Lahooti et al. 2005; McFarland et al. 2008). LRP is also a regulator of the *stpA* gene, encoding the H-NS paralogue StpA that is both a DNA- and an RNA-binding protein (Free and Dorman 1997; Sonden and Uhlin 1996). Together with StpA (and with H-NS and FIS) LRP controls the transcription of *rsd*, the gene encoding the Rsd anti-sigma factor that targets RpoD and, to a lesser extent, RpoS (Hofmann et al. 2011). These links confer on LRP the potential to influence transcription patterns throughout the genome.

The 18.8-kDa LRP monomer forms octamers and hexadecamers and has the ability to wrap, bend, and bridge DNA (Chen et al. 2001); the *B. subtilis* homologue, LrpC,

forms structures with DNA that are reminiscent of a eukaryotic histone core (Beloin et al. 2003b). LRP expression peaks at the transition from the exponential phase to the stationary phase of the growth cycle in rapidly growing *E. coli* (Landgraf et al. 1996). In keeping with its name, the interactions of LRP with its target genes can be potentiated, inhibited, or unaffected by leucine and other branched-chain amino acids (Calvo and Matthews 1994; Lahooti et al. 2005; Peterson and Reich 2010). Leucine can also influence the oligomeric state of LRP, perhaps helping to explain the distinct effects (including no effects) that branched chain amino acids can have on different LRP regulated systems.

The LRP protein competes with the Dam methylase for access to two 5'-GATC-3' sites in the *pap* and *pef* fimbrial operons of uropathogenic *E. coli* and *Salmonella*, respectively. This outcome of competition decides if the Dam sites will be methylated or not and this, in turn, determines if the fimbrial structural genes will be transcribed or not. The result is a stochastic switch that is reset by the synthesis of hemimethylated DNA during chromosome (*pap*) or virulence plasmid (*pef*) replication (Hernday et al. 2002; Nicholson and Low 2000). In contrast, LRP acts as a directionality determinant at the invertible *fimS* genetic switch that governs the phase-variable expression of type 1 fimbriae in *E. coli*, an inversion event that is catalysed by two tyrosine integrases, FimB and FimE (Corcoran and Dorman 2009; Kelly et al. 2006). In the case of *pap/pef*, LRP is acting as a DNA-binding protein in competition with a DNA modifying enzyme; in the case of *fimS*, LRP's ability to shape DNA is likely to be influencing the directionality of the On/Off invertible genetic switch. It also plays a role in virulence in *M. tuberculosis* by modulating the innate immune response of macrophage (Liu and Cai 2018). These examples illustrate the versatility of the LRP protein and its ability to influence diverse systems by different molecular mechanisms.

1.54 Small, Acid-soluble Spore Proteins, SASPs

Aerobic and anaerobic spore-forming bacteria rely on small, acid-soluble spore proteins (SASPs) to protect their DNA from damage during the long (or very long) periods that may elapse between sporulation and spore germination. Most research on SASPs has been concerned with those produced by *Bacillus* spp. and *Clostridium* spp., aerobic and anaerobic organisms, respectively (Setkow 2007). The SASPs fall into two broad groups, the α/β type and the γ type. Their genes are transcribed with the G sigma factor (Nicholson et al. 1989), protect the genomic DNA in the spore, and after germination they are cannibalised as a source of amino acids by the emerging bacterial cell (Hackett and Setlow 1988; Setlow 1988). SASPs are specialists in that they are expressed specifically to accompany the genome during its period of storage in the spore and are degraded during germination; they do not have physiological roles in vegetative cells, yet they exhibit properties that are shared with NAPs. For example, they stiffen DNA, and eliminate DNA bends, they increase the persistence length of DNA and introduce supertwists into relaxed or nicked circular DNA (Griffith et al. 1994; Nicholson et al. 1990), including plasmids from spores (Nicholson and Setlow 1990). When cloned *ssp* genes expressing SASPs are

introduced to *E. coli*, they induce nucleoid condensation; however, they also induce mutagenesis and cell killing (Setlow et al. 1991, 1992). The mutations accompanied expression of the *B. subtilis* SspC SASP and required RecA and Pol V, suggesting that the effects followed the arrest of replication forks in growing *E. coli*; a derivative of SspC that was deficient in DNA binding failed to elicit these deleterious effects when expressed in *E. coli* (Setlow et al. 1992). Overall, the effect of SASP expression in *E. coli* was to cause the Gram-negative bacterium to assume some of the characteristics of a sporulating organism (Setlow et al. 1991).