UNIVERSITY OF BIRMINGHAM
University of Birmingham
Research at Birmingham

# Local and global regulation of transcription initiation in bacteria

Browning, Douglas F; Busby, Stephen J W

*DOI:*
10.1038/nrmicro.2016.103

*License:*
None: All rights reserved

*Document Version*
Peer reviewed version

*Citation for published version (Harvard):*
Browning, DF & Busby, SJW 2016, 'Local and global regulation of transcription initiation in bacteria', *Nature Reviews Microbiology*. https://doi.org/10.1038/nrmicro.2016.103

Link to publication on Research at Birmingham portal

**Publisher Rights Statement:**
Checked 11/08/2016

**General rights**
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

**Take down policy**
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 04. Jun. 2023

# Regulation of bacterial transcription initiation: local and global

Douglas F. Browning[1] and Stephen J. W. Busby[1].

[1] Institute of Microbiology and Infection, School of Biosciences, The University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.

For correspondence: Email: D.F.Browning@bham.ac.uk and S.J.W.Busby@bham.ac.uk

## Abstract

Bacterial gene expression relies on promoter recognition by the DNA-dependent RNA polymerase and subsequent initiation of transcription. Bacterial cells are able to tune their transcriptional programmes to changing environments through numerous mechanisms that regulate the activity of RNA polymerase, or change the set of promoters to which the RNA polymerase can bind. In this Review, we outline our current understanding of the different factors that direct the regulation of transcript initiation in bacteria, whether by interacting with promoters, with RNA polymerase, or with both, and we discuss the diverse molecular mechanisms used by these factors to control gene expression.

## Introduction

In any bacterium, at any moment, there is enormous variation in the expression of individual genes. Furthermore, expression of many genes is subject to variation in response to environmental change. Such regulation of gene expression is crucial for the overall fitness of the bacterial cell. For most genes in most bacteria, the key regulatory step modulating gene expression seems to be promoter recognition and transcript initiation by RNA polymerase[1]. RNA polymerase core enzyme consists of the large β and β' subunits, two α subunits and the small ω subunit (Figure 1a). Each α subunit consists of independently folded N-terminal and C-terminal domains that are joined by a flexible linker. The β and β' subunits are assembled by binding to the N-terminal domains of the α subunits, and form a cleft that contains the active site, whereas the ω subunit is primarily a chaperone for the β' subunit. However, although competent for DNA-dependent RNA synthesis, the core enzyme is unable to recognize promoters or to initiate transcription without the assistance of one of a set of additional proteins, the sigma factors. The complex formed by the binding of a sigma factor to core enzyme is known as RNA polymerase holo enzyme and is able to orchestrate transcript initiation from specific promoters[2]. Interactions with the promoter position the RNA polymerase holo enzyme such that it unwinds the double-stranded DNA in the region of the transcript start site. Positions +1 and +2 of the unwound template strand (corresponding to the 5' end of the RNA transcript) are then able to enter the active site to form the transcriptionally-competent open complex[3-7]. Subsequent to initiation, the transcription cycle proceeds with escape, elongation and termination steps (Figure 1a), which have been extensively documented and reviewed elsewhere[8-11].

Bacterial promoters contain several discrete sequence motifs, including the -35 element, the extended -10 element, the -10 element and the discriminator region, which are recognized by the sigma factor, and the UP element, which is recognized by the C-terminal domains of the α subunits (Figure 1b; Box 1)[1-3]. All bacteria contain one predominant essential sigma factor, known as the housekeeping sigma factor (such as $\sigma^{70}$ in *Escherichia coli*, also known as RpoD), which is responsible for recognizing most promoters[2]. These housekeeping sigma factors are composed of 4 structural domains that are connected to one another by flexible linkers. In the RNA polymerase holo enzyme, the sigma factor binds to the subunits of the core enzyme, such that each domain of the sigma factor is positioned to interact with a specific promoter element[2,3]. Contacts involving Domains 3 and 4 of the sigma factor seem to have the major role in the initial positioning of the RNA polymerase, whereas Domains 1 and 2 of the sigma factor drive the formation of the open complex (Figure 1a)[2-7,12]. A further function of housekeeping sigma factors, mediated by Domain 1, is to ensure that DNA

cannot enter the active site until the RNA polymerase is bound to a promoter, which triggers a conformational change that permits DNA access to the active site[13-15].

Transcription initiation can be regulated at the level of RNA polymerase holo enzyme formation, promoter recognition by RNA polymerase, or RNA polymerase activity, all of which can be mediated by a variety of factors. These regulatory mechanisms can either modulate the function of RNA polymerase itself, or can modulate the accessibility or affinity of promoters for RNA polymerase. In this Review, we discuss the different layers of regulation for each of these targets, focusing on the factors and mechanisms on which they rely. Although we mainly describe findings from *Escherichia coli* and other model bacterial species, we also consider regulation in non-model bacterial species, and we suggest that understanding the diversity of these mechanisms may shed light on the parallel evolution of other strategies for gene regulation.

## RNA polymerase-centred regulation

Many factors interact directly with bacterial RNA polymerase to influence its activity at different promoters. These include sigma factors and a host of other proteins and ligands that regulate either the formation of RNA polymerase holo enzyme or its activity or promoter preferences.

### *Sigma factors*

The role of sigma factors is to guide the positioning of RNA polymerase molecules at promoters and then to orchestrate formation of the open complex (Figure 1a). Thus, the promoter specificity of RNA polymerase depends on its sigma factor[2]. Although the housekeeping sigma factor recruits RNA polymerase to the majority of promoters, nearly all bacteria have one or more alternative sigma factors that guide the RNA polymerase to different promoters, providing a simple mechanism for reprogramming transcription to a different set of genes[2,16]. The available evidence suggests that the housekeeping sigma factor and alternative sigma factors bind to the same site on the surface of the RNA polymerase core enzyme but that, in most conditions, the housekeeping sigma factor is more abundant and thus able to outcompete alternative sigma factors[15,17-21]. However, when the abundance of an alternative sigma factor increases, it can then compete with and displace the housekeeping sigma factor to reprogramme a subset of RNA polymerase molecules[22].

Many alternative sigma factors have a simpler domain structure than housekeeping sigma factors[2]. These fall into a small number of classes, according to their domain structure, and one class has just two domains, corresponding to Domain 2 and Domain 4 in housekeeping sigma factors[23]. Another difference between alternative sigma factors and housekeeping sigma factors is that housekeeping sigma factors recognise a larger range of target sequences, and can be regulated by diverse transcription factors, which results in a range of transcriptional outputs, whereas most alternative sigma factors bind to a smaller set of promoters, with more stringent sequence specificity, associated with fewer transcription factors[24-26]. A possible explanation for this is that many alternative sigma factors evolved to switch on the concerted expression of sets of genes that have stress-responsive functions, for which fine-tuning is not required (though there are exceptions[27]).

The activity of alternative sigma factors can be controlled by several mechanisms, including covalent modification, sequestration by cognate anti-sigma factors, their subcellular localization, and their rate of synthesis and proteoloytic turnover[16,28]. In some cases, regulation by anti-sigma factors is countered by anti-anti-sigma factors that sequester anti-sigma factors in response to a particular trigger, in a process termed 'partner switching'[23,28,29]. Some anti-anti-sigma factors share structural similarities with the cognate sigma factor, which suggests that they have evolved to be mimics[30,31]. All of these regulatory mechanisms combine to form complex regulatory circuits that adjust the number of RNA polymerase molecules programmed by a particular alternative sigma factor according to the environmental conditions.

Despite our knowledge of this circuitry, many details of the mechanisms that mediate sigma factor exchange are unclear. For example, we are still ignorant of how many alternative sigma factors, with weak binding affinities for the RNA polymerase core enzyme, are able to recruit sufficient RNA polymerase molecules to have an impact on the transcriptional programme[22]. One well studied case of sigma factor exchange occurs between $\sigma^{38}$ (also known RpoS) and the housekeeping sigma factor, $\sigma^{70}$, in *E. coli*, in response to certain stresses, or at entry into stationary phase[20, 21]. Displacement of $\sigma^{70}$ following an increase in the abundance of $\sigma^{38}$ is facilitated by Crl[32-34], which interacts with $\sigma^{38}$, and by Rsd[35-37], an anti-sigma factor for $\sigma^{70}$, which reduces the level of functional $\sigma^{70}$, thereby biasing the competition between the two sigma factors in favour of $\sigma^{38}$. In addition, the small guanine nucleotide ppGpp, which undergoes a sharp rise in abundance in response to certain metabolic responses, may have a role in the exchange of some sigma factors [38, 39]

Nearly all alternative sigma factors are evolutionarily related to housekeeping sigma factors, consisting of 2, 3 or 4 domains that retain common functions in their different contexts[2]. However, most bacteria contain one additional alternative sigma factor that is 'in a class of its own' and belongs to the family of $\sigma^{54}$ sigma factors, which is evolutionarily unrelated to other sigma factors[40-42]. These sigma factors recognize different elements at target promoters to other sigma factors, with specificity determined by elements at positions -24 and -12, rather than -35 and -10. Crucially, unlike the complex formed by $\sigma^{70}$, RNA polymerase holo enzyme that contains $\sigma^{54}$ is unable to proceed to the open complex and requires activation by a special class of ATP-dependent activators (see below)[42]. Recent structural data has shown that the site at which the template strand needs to be located for open complex formation is blocked by parts of $\sigma^{54}$ [19]. Interestingly, different activators are required at different target promoters and, at least in *E. coli*, levels of $\sigma^{54}$ appear to be unregulated[21]. Together, these features of $\sigma^{54}$ argue that, in keeping with being in a class of its own, $\sigma^{54}$ uses a different regulatory strategy to other sigma factors, in that it demits regulation of transcript initiation to an activator, instead of regulating transcription initiation by its own abundance.

*Other regulators of holo enzyme activity*

The simplicity and ubiquity of regulation by sigma factors has detracted attention from numerous other regulatory factors that interact with RNA polymerase (Figure 2). Although many of these regulatory factors act at the level of transcript elongation and termination (such as Nus factors), and hence are beyond the scope of this Review[9], a small number are, like sigma factors, regulators of transcript initiation. Some of these regulators are restricted to particular groups of bacteria, such as RbpA and CarD, which bind to and stabilize open complexes in Actinomycetes[43] (Figure 2a). Recent studies show that RbpA binds directly to

Domain 2 of the housekeeping sigma factor and is positioned to make direct stabilizing contacts with the upstream end of unwound DNA in open complexes[44]. CarD has a C-terminal domain that makes stabilizing contacts that are complementary to those made by RbpA, which suggests that RbpA and CarD function coordinately, whilst the N-terminal domain of CarD interacts with the RNA polymerase β subunit[45,46].

Although *E. coli* and related Enterobacteriaceae lack RbpA and CarD, these bacteria use other factors to modulate the stability of open complexes. For example, the extended coiled-coil protein motif in DksA inserts into a narrow channel (known as the secondary channel) that leads from the surface of the RNA polymerase to the active site. The insertion of DksA, in cooperation with ppGpp, selectively stabilizes open complexes at some promoters, but destabilizes open complexes at others, by a mechanism that awaits elucidation, but which likely involves a lowering of the energy barrier between the closed and open RNA polymerase-promoter complexes[47-52]. The structure of DksA and its mode of binding to the polymerase resemble those of the GreA and GreB transcription elongation factors[47, 53], and recent studies suggest that DksA, GreA and GreB all function as 'inspectors' that continually probe the enzyme active site by rapid transient 'visits' to the secondary channel[54-56]. Increased levels of ppGpp result from certain metabolic stresses, and the *E. coli* ppGpp–DksA mechanism provides a sigma-free means of responding to these stresses by reprogramming the transcriptome[52].

Other factors reduce the number of RNA polymerase molecules available for transcription by sequestering the holo enzyme. One example is 6S RNA, a ~180-nt non-coding RNA, which is synthesized in response to slow growth, and forms a 1:1 complex with the RNA polymerase holo enzyme (Figure 2b)[57]. In *E. coli*, 6S RNA is a mimetic for the DNA of promoters that are targets for the housekeeping RNA polymerase holo enzyme (that contains $\sigma^{70}$). As a result, the proportion of functional (that is, non-sequestered) RNA polymerase holo enzyme that contains $\sigma^{38}$ increases, but the total transcription potential of the cell is reduced as the number of 'free' RNA polymerase molecules is reduced[57]. Hence, 6S RNA appears to work cooperatively with the Rsd anti-sigma factor (see above) to couple gene expression to growth, since levels of both increase as cell growth slows[37].

Some bacteriophage encode factors that inhibit the activity of bacterial RNA polymerases to favour the activity of their own bespoke RNA polymerases. For example, during infection of *E. coli* cells, the Gp2 protein of 'phage T7 induces a conformational change in part of Domain 1 of $\sigma^{70}$ that blocks access of template DNA to the active site (Figure 2c)[14]. By contrast, the P7 protein of the related phage Xp10, which infects *Xanthomonas oryzae*, inhibits the activity of RNA polymerase by displacing the sigma factor from the holo enzyme; the association of P7 with RNA polymerase is stable, which prevents recruitment of a replacement sigma factor to reform the holo enzyme (Figure 2d)[58].

Some regulators of RNA polymerase holo enzyme remodel parts of the polymerase, and are known as 'appropriators'. Similar to alternative sigma factors, their function is to alter the promoter preference of the RNA polymerase (Figure 3). The most well studied example of a bacteriophage appropriator is the AsiA protein of 'phage T4, which infects *E. coli*. As T4 does not encode its own RNA polymerase, its invasion strategy is to appropriate, rather than to silence, the host RNA polymerase. AsiA is encoded by a gene expressed during the early stage of the T4 life cycle, and remodels and repositions Domain 4 of $\sigma^{70}$ so that it is unable to recognise -35 elements[59, 60] and, instead, becomes a target for the T4 MotA transcription activator. MotA, which is encoded by early-stage genes, is the essential activator for

transcription of genes expressed during the middle stage of the T4 life cycle (Figure 3a)[61]. In this way, AsiA and MotA redirect the housekeeping RNA polymerase holo enzyme from the transcription of host genes to the transcription of phage genes. Interestingly, other T4 proteins, Alt and ModA, also alter the promoter preferences of the host RNA polymerase, by ADP-ribosylating residue R265 in the C-terminus of either one or both RNA polymerase α subunits. This prevents the RNA polymerase recognizing UP elements and so helps divert the polymerase from transcribing host genes to transcribing T4 genes (Figure 3b)[61].

The two examples of host-encoded appropriators that have been studied in most detail are SoxS in *E. coli* and Spx in *Bacillus subtilis*. SoxS targets RNA polymerase holo enzyme to promoters carrying upstream Sox-box sequences, and is essential for the induction of dozens of genes in response to the sensing of oxidative stress by the SoxR repressor, which triggers an increase in the abundance of SoxS. Evidence suggests SoxS guides the RNA polymerase holo enzyme to target promoters using a 'pre-recruitment' mechanism; that is, SoxS contacts the DNA-recognition surfaces of the C-terminal domain of each RNA polymerase α subunit prior to binding of the SoxS–RNA polymerase holo enzyme complex to the DNA[62, 63]. The SoxS–RNA polymerase holo enzyme complex is unable to recognise UP elements but instead recognizes Sox-boxes (Figure 3c)[63]. Similarly, Spx, in response to oxidative stress, binds to the C-terminal domain of each RNA polymerase holo enzyme α subunit, and this alters the promoter preference such that certain promoters are favoured and others are disfavoured (Figure 3d) [64]. Structures of Spx in complex with the C-terminal domain of the α subunit suggests that complex formation does not produce any major structural change in either Spx or the α subunit[65,66].

Finally, in addition to regulation by proteins, RNA polymerase holo enzyme activity can be regulated by fluctuations in the levels of its four nucleoside triphosphate (NTP) substrates. Note that the Michaelis constant for the initiating nucleotide is higher than that for subsequently added NTPs[67]; that is, for a given rate of transcription, the required concentration of the initiating NTP is higher than that for subsequent NTPs, which means that the concentration of the initiating NTP is most crucial to the activity of the RNA polymerase. Since, in *E. coli*, the initiating NTP for rRNA transcripts is ATP, initiation of these transcripts is expected to be sensitive to the cellular concentration of ATP, which increases as cells leave stationary phase with a concomitant burst of rRNA synthesis[68, 69]. In contrast, in *Bacilus subtilis*, the initiating NTP for rRNA is GTP, which decreases in abundance when it is consumed by ppGpp synthesis. Hence, ppGpp levels in *B. subtilis* couple metabolism to RNA polymerase activity indirectly through corresponding changes to GTP levels[70, 71]. This contrasts with *E. coli*, where the cooperative action of ppGpp and DksA directly regulates the stability of RNA polymerase holo enzyme open complexes (see above)[51].

More complex effects of NTP levels are observed during transcript initiation at *E. coli* promoters that control genes concerned with pyrimidine biosynthesis. At many of these promoters, fluctuations in UTP or CTP levels affect transcript stability and translation efficiency. This can occur either by changing the location of the transcript start site, or by non-templated insertion of extra bases into the nascent transcript[72]. Since UTP and CTP are both pyrimidine-based nucleotides, this form of regulation produces a feedback mechanism in which pyrimidine biosynthesis can be adjusted in response to changes in their levels[72].

**Promoter-centred regulation**

Regulation by factors that bind directly to the RNA polymerase is complemented by factors that directly target the promoter DNA. Such regulation can be mediated by many different mechanisms ranging from supercoiling to transcription factors that simply block access to the promoter (Figure 4). In this scenario, first imagined by François Jacob and collaborators (Box 2), transcription factors contain structural motifs that bind to promoters containing cognate 'operators', which are specific sequences of base pairs. The operators for most bacterial DNA-binding structural motifs, such as the helix-turn-helix motif, have 4–5 base pair targets. Since any given 5-base pair sequence will arise, on average, once every $4^5$ base pairs, further specificity is required to target transcription factors only to the desired promoters. Such specificity can be achieved using various mechanisms, including homodimerization (or higher order multimerization) of the transcription factor, association of the transcription factor with another DNA-binding factor, or the incorporation of several DNA-binding structural motifs into a single transcription factor. Of these mechanisms, dimerization or further multimerization is the most common, and therefore most operators contain direct or inverted repeats of a 4–5 base pair sequence.

The key role of transcription factors is to couple promoter activity to environmental cues, so their activity has to be regulated in response to these cues. In many transcription factors, regulatory modules mediate DNA binding in response to environmental cues sensed by the binding of a small ligand or protein, or covalent modification. The activities of other transcription factors depend on their abundance and availability in the cell, which can be regulated by synthesis, turnover or sequestration. Although some transcription factors regulate only a single promoter, most transcription factors regulate many promoters. In addition, at least in *E. coli*, most promoters are regulated by more than one factor. Furthermore, many transcription factors are expressed from promoters that are themselves regulated by other transcription factors. All this generates a complex regulatory network, in which the concerted activities of specific, global and master regulators orchestrate the distribution of RNA polymerase to the various transcription units present in the genome[73-76].

It is important to appreciate that transcription initiation occurs in the context of the bacterial nucleoid, and the compaction of the bacterial chromosome that occurs during the formation of the nucleoid is thought to have an overall negative effect on promoter activity. Therefore, modulation of the local structure of the nucleoid to change the level of compaction may result in changes in transcription. Compaction is primarily due to supercoiling and to the contortions induced in the DNA by nucleoid associated proteins (NAPs), which 'sculpt' the bacterial chromosome by bending, wrapping, looping and twisting DNA such that it fits inside the bacterial cell[77] (Box 3). Many NAPs are abundant and bind to a large number of targets sites using a relaxed sequence specificity[78-80]. By contrast, some NAPs, such as Fis and IHF, clearly do have sequence specificity, which enables them to mediate promoter-specific effects on the local DNA structure, such that they can be regarded as transcription factors[81]. For example, Fis blocks binding of the housekeeping RNA polymerase holo enzyme to the *dps* promoter in *E. coli*, silencing the promoter so that it is active only in stress conditions[82]. In parallel, global alterations in supercoiling, which are probably mediated by fluctuations in ATP levels, lead to up- and down-regulation at scores of promoters; in some cases, this results in local superhelical density changes that modulate the transcription activity at specific target promoters[83-88].

Recent observations that the activities of certain promoters vary according to their precise location in the folded bacterial chromosome may be best explained by the suggestion that the availability of free RNA polymerase holo enzyme, or at least the degree to which the enzyme can access promoter regions, is not the same at every location[89]. Thus, the longstanding assumption that promoter activity in bacteria is determined only by the correspondence of sequences of promoter elements to consensus sequences and regulation by local transcription factors (Box 1) is an oversimplification.

*Repressing transcription factors*

Several mechanisms have been described for the repression of transcription initiation by repressive transcription factors (known as 'repressors'). At many promoters, repression of transcript initiation occurs by simple steric hindrance, in which repressors bind to operators that overlap the -10 or -35 elements of the promoter, and thus block access of RNA polymerase to the promoter (Figure 4a)[1]. Many promoters contain arrays of operators that are organized adjacently to each other. For these promoters, a greater number of bound repressors usually translates to more efficient repression. At other promoters, the simultaneous binding of a repressor to operators that are located distally to one another gives rise to a loop in the local DNA structure that prevents binding by RNA polymerase (Figure 4a)[90]. The best documented case of repression by DNA looping is mediated by the GalR repressor of the galactose operon in *E. coli*. The operators that are recognized by GalR are located upstream and downstream of the different *gal* operon promoter elements but do not overlap with these elements [90,91]. At some promoters, two or more different repressor act together, and, mostly, when this happens, the different repressors bind independently to their targets. This can lead to very tight repression, such as at promoters that control colicin expression[92,93].

At some promoters, repression is more complicated and is mediated by 'anti-activator' repressors (Figure 4a), which counter the function of activating transcription factors ('activators', see below). For example, the CytR repressor is as an anti-activator that simultaneously interacts with its operator and an adjacent activator, the cyclic AMP receptor protein (CRP; also known as catabolite activator protein (CAP)). At some promoters, CytR binding requires a combination of CytR–CRP and CytR–DNA interactions to prevent binding of RNA polymerase[94].

Another complex scenario is found at promoters where RNA polymerase binds to the promoter DNA but the repressor interacts directly with bound RNA polymerase to mediate repression. For example, the p4 protein encoded by phage ϕ29, which infects *B. subtilis*, simultaneously binds to the C-terminal domain of the α subunit of RNA polymerase and to the DNA upstream of the polymerase, preventing promoter clearance[95].

*Activating transcription factors*

Activating transcription factors (known as 'activators') increase the activity of promoters, often from low basal levels, by one of three mechanisms: Class I activation, Class II activation or activation by a conformational change (Figure 4b) [96]. In Class I activation, the activator binds to an operator located upstream of the promoter elements and then recruits RNA polymerase to the promoter through interactions formed between a small surface-exposed patch on the activator (known as an Activating Region) and the C terminal domain of the RNA polymerase α subunit[96]. Structural analysis of activation by CRP shows that this interaction occurs without any structural change in either partner, and, as such, is often

referred to as 'velcro'[97]. Furthermore, Class I activation occurs at promoters that require activators to recruit RNA polymerase because one or more of the promoter elements has a sequence that is sub-optimal for RNA polymerase binding. Thus, for those promoters dependent on Class I activation, RNA polymerase is recruited by protein–protein interactions rather than protein–DNA interactions. Optimal Class I activation occurs when the activator binds to the same face of the DNA helix as the RNA polymerase α subunit[98], and this facilitates the interaction between the two proteins. Hence, activation is reduced if the operator is displaced by 1–5 base pairs (corresponding to up to half a turn of the helix) but restored if the displacement is ~10 base pairs (corresponding to a full turn of the helix). In some cases, decreased activation due to suboptimal placement of an activator can be reversed if the bases sequence of one or more of the promoter elements is modified to enhance RNA polymerase binding[99].

In Class II activation, the activator binds to an operator that overlaps the -35 element of the promoter. Once bound to the operator, the activator recruits the RNA polymerase by forming direct interactions with Domain 4 of the sigma factor, or the N-terminal domain of the α subunit, or with other components of the RNA polymerase (Figure 4b)[96]. Here, discrete activating regions in the activator recruit RNA polymerase to the promoter but, at some promoters, the interactions between the activator and the RNA polymerase are more favourable in the open complex, and this facilitates the transition from the closed to the open complex[100-103]. One of the consequences of Class II activation is that the C terminal domains of the RNA polymerase α subunits are unable to bind to their preferred binding site immediately upstream of the -35 promoter element and, instead, bind immediately upstream of the activator[96]. This permits Class II activators to function together with Class I activators, and this combination is responsible for synergy at many bacterial promoters where activity is co-dependent on two signals[1].

A third mechanism by which a single activator can drive transcription initiation at a promoter involves binding of the activator to an operator located between the -35 and -10 elements of the promoter (Figure 4b). These promoters are usually poorly active due to non-optimal spacing between the two elements, and the current view is that activation involves distortion of the DNA to so that the -35 and -10 elements of the promoter are better juxtaposed for RNA polymerase binding[104, 105]. The available structural evidence argues that this distortion is not evenly spread across the DNA between the -35 and -10 elements[105,106].

In the above examples (Figure 4b), which can be viewed as 'direct' activation, the activator alters the target promoter to make it more 'attractive' to the RNA polymerase. However, activation can also occur indirectly, usually by removing a repressor, and in these cases the activator can be thought of as an anti-repressor[1]. For example, in enterohemorrhagic *E. coli*, expression of one of the principal virulence determinants is repressed by the NAP, H-NS, but expression of Ler, an H-NS homologue, disrupts the repression and triggers virulence[107, 108]. A more complex example is found at the *E. coli nir* promoter that controls the expression of a nitrite reductase. Here, Class II activation, repression by NAPs, and indirect activation by anti-repressors, all compete to regulate promoter activity. Class II activation of *nir* is mediated in response to oxygen deprivation by an activator known as FNR, but activation is suppressed by the NAPs, IHF and Fis, which bind to specific target sites located just upstream of the DNA site for FNR[109,110]. However, this suppression is disrupted by either NarL or NarP, acting as anti-repressors, whose activities are regulated by levels of

extracellular nitrate and nitrite ions[111]. Thus, expression from FNR-regulated promoters integrates two different environmental signals, oxygen and nitrate/nitrite.

The common feature of many activation mechanisms is that they work primarily by promoting recruitment of RNA polymerase, rather than promoting the transition of the RNA polymerase closed complex to the open complex, which is a process that does not require an activator for most RNA polymerase holo enzymes. However RNA polymerase holo enzymes containing sigma factors related to $\sigma^{54}$ are unable to make this transition and hence require special activators from the class of activators known as enhancer binding proteins (EBPs). These EBPs use ATP hydrolysis to change the conformation of a surface-exposed loop that interacts directly with $\sigma^{54}$ to induce a conformational change that drives open complex formation and enables transcriptionally competent complexes to be formed[112]. As with activators that mediate Class I activation, EBPs bind upstream of promoter regions. However, since they need to interact with the $\sigma^{54}$-related sigma factor, rather than with the adjacent α subunits of the RNA polymerase holoenzyme, looping of upstream DNA has to occur to 'deliver' the EBP to the sigma factor, and this is often assisted by DNA-bending proteins such as IHF [112, 113].

*Promoter modifications*

Other mechanisms by which the activity of bacterial promoters can be regulated involve modification of specific bases, or changes to the base sequence, which alter the affinity of an operator for a transcription factor, or affect RNA polymerase binding or its activity. The most common chemical modification is DNA methylation, which can result in changes in the binding affinities of specific transcription factors to operators that can potentially lead to global changes in the transcriptional programme[114-118]. For example, methylation of the operators of two *E. coli* genes, *pap* and *agn43*, by the Dam methylase results in the loss of repressor binding, which leads to the induction of gene expression. Repression is re-established in daughter cells, as the DNA synthesized during replication is unmethylated (Figure 5a).

The most extreme examples of base sequence changes at regulatory regions are inversions of DNA segments that reverse the orientation of promoters in response to a signal. This creates promoters that are either 'on' or 'off', depending on whether they are oriented towards or away from the genes that they control. This contrasts with promoters that are directly controlled by transcription factors, for which output can be proportional to the input signal[117, 118]. An example of regulation by DNA inversion is the promoter that controls the expression of the *fim* operon, which encodes Type 1 fimbriae in both commensal and pathogenic *E. coli* (Figure 5b). As might be expected, the FimB and FimE enzymes that catalyse DNA inversion at this operon are themselves tightly regulated.

Local sequence variation at promoters provides a more sophisticated alternative strategy to DNA inversion. This is usually driven by tracts of variable length that repeat a single nucleotide (or dinucleotide), often in the vicinity of the -35 element of the promoter (Figure 5c). For example, the promoters of *hifA* and *hifB* in *Haemophilus influenza* have variable numbers of TA repeats that correspond to changes in the level of gene expression Individual bacterial cells in a population have a different number of repeats in the variable region of the promoter, with each tract length conferring a different level of transcriptional activity. Thus, at any moment, a subset of cells will have the optimal level of transcriptional activity at the promoter for the given conditions [117-120]. This regulation 'by lottery', driven by

repetitive sequences that differ from one generation to another, contrasts sharply with regulation by transcription factors, which drive specific responses to specific environmental cues. Hence transcription factors may not always be essential for bacterial survival, and their deployment is not the only strategy available for bacteria to adapt their transcriptional programme to changing environments.

**Evolutionary origins**

François Jacob stated that 'evolution was a tinkerer' and nowhere is this famous statement more evident than in the regulation of transcription initiation in bacteria[121]. Thus, it appears that bacterial regulatory circuits evolved by the sequential addition of layers of complexity, and this evolution continues today. This view accounts for the great diversity of strategies adopted by different bacteria for regulating transcription initiation, and underscores the danger of making generalisations based solely on studies using *E. coli* as a model. Hence, for example, different bacteria deploy sigma factors and transcription factors to different extents, in order to harness the evolutionary advantage of making defined transcripts at specific levels in a regulated way.

From the perspective of evolutionary history, we can assume that transcription was present in the first cells and it is easy to imagine that this transcription might have been much less specific and much less regulated than what we observe now. Indeed, it has been suggested that the widely observed, unregulated, pervasive transcription, seen in many bacteria, is an evolutionary relic from the time before control was exerted[122]. Furthermore, the RNA polymerase core enzyme, which is common to all three domains of life, can make transcripts but lacks the ability to select specific start sites or efficiently initiate transcription[123]. In bacteria, these functions were acquired by the evolution of sigma factors, whereas different solutions evolved in the other domains of life[123].

Following the acquisition of sigma factors, the bacterial RNA polymerase had to overcome the compacted DNA structure of the nucleoid and access many potential promoters that were likely blocked by NAPs (Box 3). It has been argued that transcriptional regulation evolved by tinkering with NAPs to create regulated repression or regulated activation of specific genes[124]. Hence transcription activation by anti-repression (for example by Ler, NarL or NarP, see above) may well reflect the oldest form of activation, and, perhaps later, further tinkering of NAPs led to the evolution of activating regions. Evidence of a common origin for transcription factors and NAPs comes from findings that many transcription factors bind to some targets without any measurable effect on transcription[125,126], and that many NAPs function like transcription factors at some promoters, for example, by deploying activating regions to help recruit RNA polymerase[127, 128].

**Outlook**

Since 2004, when we last reviewed this topic[1], enormous progress has been made in understanding the molecular details of transcription initiation, together with mechanisms that regulate its activation and repression. At the forefront have been advances in understanding the structure and dynamics of the bacterial RNA polymerase, and these have opened the way to detailed mechanistic understanding of the actions of the many interacting factors. In parallel, genome sequencing, together with the -omics revolution, has

alerted us to the diversity of factors in different bacteria, and to the breadth of the regulatory repertoire. However, despite this, we are still unable to predict transcriptomes and regulatory patterns from first principles. Therefore, an important aspiration for the years ahead, in our view, is to develop our understanding of regulatory mechanisms such that the many layers of regulation can be modelled, in order to accurately predict transcriptional outcomes from a bacterial genome.

**Acknowledgements.**

The authors were supported by BBSRC grant BB/J006076/1 and by the Industrial Biotechnology Catalyst (Innovate UK, BBSRC, EPSRC) to support the translation, development and commercialisation of innovative Industrial Biotechnology processes.

**Competing interests statement:** The authors declare no competing interests.

**Author Biographies**

Doug Browning is a Senior Research Fellow at the University of Birmingham, UK, where he moved in 1997, after completing his Ph.D. at the University of Warwick. His research has mainly focused on two areas: how bacteria regulate gene expression in response to environmental and metabolic signals and how these signals are integrated, and the mechanisms by which the outer membranes of Gram-negative bacteria are assembled and can be disrupted.

Steve Busby received his doctorate from Oxford University and then moved to the Pasteur Institute in Paris, where he became interested in bacteria and how they organize and manage their genes. Since returning to the UK in 1983, Steve and his colleagues have contributed to our understanding of transcription initiation and gene activation in bacteria, and have also developed new methodologies to study regulation on a global scale.

**References.**

1. Browning, D.F. & Busby, S.J. The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2**, 57-65 (2004).
2. Feklistov, A., Sharon, B.D., Darst, S.A. & Gross, C.A. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol* **68**, 357-76 (2014).
3. Murakami, K.S. & Darst, S.A. Bacterial RNA polymerases: the wholo story. *Curr Opin Struct Biol* **13**, 31-9 (2003).
4. Feklistov, A. & Darst, S.A. Structural basis for promoter -10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell* **147**, 1257-69 (2011).
5. Zhang, Y. et al. Structural basis of transcription initiation. *Science* **338**, 1076-80 (2012).

**This work complements previous structural work from the Darst laboratory (ref. 3) by revealing the interactions of bacterial holoenzyme containing $\sigma^{70}$ with the downstream end of the initiation 'bubble'. The results introduce us to unstacking of certain bases and their insertion into pockets in the sigma factor, and to specific interactions between side chains in the core enzyme and certain bases round the transcript start (the CRE, core recognition element).**

6. Zuo, Y. & Steitz, T.A. Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Mol Cell* **58**, 534-40 (2015).
7. Bae, B., Feklistov, A., Lass-Napiorkowska, A., Landick, R. & Darst, S.A. Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *Elife* **4** doi:10.7554/eLife.08504 (2015).
8. Ruff, E.F., Record, M.T., Jr. & Artsimovitch, I. Initial events in bacterial transcription initiation. *Biomolecules* **5**, 1035-62 (2015).
9. Washburn, R.S. & Gottesman, M.E. Regulation of transcription elongation and termination. *Biomolecules* **5**, 1063-78 (2015).
10. Roberts, J.W., Shankar, S. & Filter, J.J. RNA polymerase elongation factors. *Annu Rev Microbiol* **62**, 211-33 (2008).
11. Zhang, J. & Landick, R. A two-way street: regulatory interplay between RNA polymerase and nascent RNA structure. *Trends Biochem Sci* **41**, 293-310 (2016).
12. Campbell, E.A. et al. Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell* **9**, 527-39 (2002).
13. Mekler, V. et al. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**, 599-614 (2002).
14. Bae, B. et al. Phage T7 Gp2 inhibition of *Escherichia coli* RNA polymerase involves misappropriation of sigma70 domain 1.1. *Proc Natl Acad Sci U S A* **110**, 19772-7 (2013).
15. Murakami, K.S. X-ray crystal structure of *Escherichia coli* RNA polymerase sigma70 holoenzyme. *J Biol Chem* **288**, 9126-34 (2013).
16. Gruber, T.M. & Gross, C.A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* **57**, 441-66 (2003).
17. Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O. & Darst, S.A. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**, 1285-90 (2002).
18. Vassylyev, D.G. et al. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature* **417**, 712-9 (2002).
19. Yang, Y. et al. Structures of the RNA polymerase-sigma54 reveal new and conserved regulatory strategies. *Science* **349**, 882-5 (2015).

**This paper presents the long awaited structure of $\sigma^{54}$ and its interactions with RNA polymerase. The structure shows how $\sigma^{54}$ really is different to $\sigma^{70}$, and why it is incompetent for transcript initiation, and suggests how its activators might work.**

20. Jishage, M. & Ishihama, A. Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of sigma 70 and sigma 38. *J Bacteriol* **177**, 6832-5 (1995).
21. Jishage, M., Iwata, A., Ueda, S. & Ishihama, A. Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions. *J Bacteriol* **178**, 5447-51 (1996).
22. Maeda, H., Fujita, N. & Ishihama, A. Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res* **28**, 3497-503 (2000).
23. Campbell, E.A., Westblade, L.F. & Darst, S.A. Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. *Curr Opin Microbiol* **11**, 121-7 (2008).
24. Rhodius, V.A. et al. Design of orthogonal genetic switches based on a crosstalk map of sigmas, anti-sigmas, and promoters. *Mol Syst Biol* **9**, 702 (2013).
25. Campagne, S., Marsh, M.E., Capitani, G., Vorholt, J.A. & Allain, F.H. Structural basis for -10 promoter element melting by environmentally induced sigma factors. *Nat Struct Mol Biol* **21**, 269-76 (2014).
26. Koo, B.M., Rhodius, V.A., Nonaka, G., deHaseth, P.L. & Gross, C.A. Reduced capacity of alternative sigmas to melt promoters ensures stringent promoter recognition. *Genes Dev* **23**, 2426-36 (2009).
27. Hollands, K., Lee, D.J., Lloyd, G.S. & Busby, S.J. Activation of sigma 28-dependent transcription in *Escherichia coli* by the cyclic AMP receptor protein requires an unusual promoter organization. *Mol Microbiol* **75**, 1098-111 (2010).
28. Paget, M.S. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules* **5**, 1245-65 (2015).
29. Osterberg, S., del Peso-Santos, T. & Shingler, V. Regulation of alternative sigma factor use. *Annu Rev Microbiol* **65**, 37-55 (2011).
30. Francez-Charlot, A. et al. Sigma factor mimicry involved in regulation of general stress response. *Proc Natl Acad Sci U S A* **106**, 3467-72 (2009).

**Following Campbell et al's structures of sigma-anti-sigma complexes (ref. 23) this paper reports that an anti-anti-sigma factor can function by mimicking a sigma factor. Although predictable that this would be the case, it needed to be shown, so this is a landmark paper.**

31. Herrou, J., Rotskoff, G., Luo, Y., Roux, B. & Crosson, S. Structural basis of a protein partner switch that regulates the general stress response of alpha-proteobacteria. *Proc Natl Acad Sci U S A* **109**, E1415-23 (2012).
32. Typas, A., Barembruch, C., Possling, A. & Hengge, R. Stationary phase reorganisation of the *Escherichia coli* transcription machinery by Crl protein, a fine-tuner of sigmas activity and levels. *EMBO J* **26**, 1569-78 (2007).
33. Banta, A.B. et al. Key features of sigmaS required for specific recognition by Crl, a transcription factor promoting assembly of RNA polymerase holoenzyme. *Proc Natl Acad Sci U S A* **110**, 15955-60 (2013).
34. Banta, A.B. et al. Structure of the RNA polymerase assembly factor Crl and identification of its interaction surface with sigma S. *J Bacteriol* **196**, 3279-88 (2014).
35. Yuan, A.H. et al. Rsd family proteins make simultaneous interactions with regions 2 and 4 of the primary sigma factor. *Mol Microbiol* **70**, 1136-51 (2008).
36. Piper, S.E., Mitchell, J.E., Lee, D.J. & Busby, S.J. A global view of *Escherichia coli* Rsd protein and its interactions. *Mol Biosyst* **5**, 1943-7 (2009).
37. Sharma, U.K. & Chatterji, D. Transcriptional switching in *Escherichia coli* during stress and starvation by modulation of $\sigma^{70}$ activity. *FEMS Microbiol Rev* **34**, 646-57 (2010)
38. Jishage, M., Kvint, K., Shingler, V. & Nystrom, T. Regulation of sigma factor competition by the alarmone ppGpp. *Genes Dev* **16**, 1260-70 (2002).

39. Costanzo, A. et al. ppGpp and DksA likely regulate the activity of the extracytoplasmic stress factor sigmaE in *Escherichia coli* by both direct and indirect mechanisms. *Mol Microbiol* **67**, 619-32 (2008).
40. Merrick, M.J. In a class of its own--the RNA polymerase sigma factor sigma 54 (sigma N). *Mol Microbiol* **10**, 903-9 (1993).
41. Studholme, D.J. & Buck, M. The biology of enhancer-dependent transcriptional regulation in bacteria: insights from genome sequences. *FEMS Microbiol Lett* **186**, 1-9 (2000).
42. Wigneshweraraj, S. et al. Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor. *Mol Microbiol* **68**, 538-46 (2008).
43. Flentie, K., Garner, A.L. & Stallings, C.L. The *Mycobacterium tuberculosis* transcription machinery: ready to respond to host attacks. *J Bacteriol* **198**, 1360-73(2016).
44. Hubin, E.A. et al. Structural, functional, and genetic analyses of the actinobacterial transcription factor RbpA. *Proc Natl Acad Sci U S A* **112**, 7171-6 (2015).
45. Srivastava, D.B. et al. Structure and function of CarD, an essential mycobacterial transcription factor. *Proc Natl Acad Sci U S A* **110**, 12619-24 (2013).
46. Bae, B. et al. CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. *Elife* **4** doi:10.7554/eLife.08505 (2015).
47. Perederina, A. et al. Regulation through the secondary channel--structural framework for ppGpp-DksA synergism during transcription. *Cell* **118**, 297-309 (2004).
48. Paul, B.J. et al. DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* **118**, 311-22 (2004).
49. Paul, B.J., Berkmen, M.B. & Gourse, R.L. DksA potentiates direct activation of amino acid promoters by ppGpp. *Proc Natl Acad Sci U S A* **102**, 7823-8 (2005).

**Here, an *in vitro* assay is exploited to provide the proof that ppGpp, together with DksA, really can directly stimulate transcript initiation at a promoter. This paper provides a wonderful lesson in how the limitations of genetics can be atoned by amazing biochemistry.**

50. Lennon, C.W. et al. Direct interactions between the coiled-coil tip of DksA and the trigger loop of RNA polymerase mediate transcriptional regulation. *Genes Dev* **26**, 2634-46 (2012).
51. Ross, W., Vrentas, C.E., Sanchez-Vazquez, P., Gaal, T. & Gourse, R.L. The magic spot: a ppGpp binding site on *E. coli* RNA polymerase responsible for regulation of transcription initiation. *Mol Cell* **50**, 420-9 (2013).
52. Hauryliuk, V., Atkinson, G.C., Murakami, K.S., Tenson, T. & Gerdes, K. Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Rev Microbiol* **13**, 298-309 (2015).
53. Zenkin, N. & Yuzenkova, Y. New insights into the functions of transcription factors that bind the RNA polymerase secondary channel. *Biomolecules* **5**, 1195-209 (2015).
54. Yuzenkova, Y., Roghanian, M. & Zenkin, N. Multiple active centers of multi-subunit RNA polymerases. *Transcription* **3**, 115-8 (2012).
55. Friedman, L.J. & Gelles, J. Multi-wavelength single-molecule fluorescence analysis of transcription mechanisms. *Methods* **86**, 27-36 (2015).
56. Zhang, Y. et al. DksA guards elongating RNA polymerase against ribosome-stalling-induced arrest. *Mol Cell* **53**, 766-78 (2014).
57. Cavanagh, A.T. & Wassarman, K.M. 6S RNA, a global regulator of transcription in *Escherichia coli*, *Bacillus subtilis*, and beyond. *Annu Rev Microbiol* **68**, 45-60 (2014).
58. Liu, B. et al. A bacteriophage transcription regulator inhibits bacterial transcription initiation by sigma-factor displacement. *Nucleic Acids Res* **42**, 4294-305 (2014).
59. Lambert, L.J., Wei, Y., Schirf, V., Demeler, B. & Werner, M.H. T4 AsiA blocks DNA recognition by remodeling sigma70 region 4. *EMBO J* **23**, 2952-62 (2004).

60. Gregory, B.D. et al. A regulator that inhibits transcription by targeting an intersubunit interaction of the RNA polymerase holoenzyme. *Proc Natl Acad Sci U S A* **101**, 4554-9 (2004).
61. Hinton, D.M. Transcriptional control in the prereplicative phase of T4 development. *Virol J* **7**, 289 (2010).
62. Griffith, K.L., Shah, I.M., Myers, T.E., O'Neill, M.C. & Wolf, R.E., Jr. Evidence for "pre-recruitment" as a new mechanism of transcription activation in *Escherichia coli*: the large excess of SoxS binding sites per cell relative to the number of SoxS molecules per cell. *Biochem Biophys Res Commun* **291**, 979-86 (2002).
63. Shah, I.M. & Wolf, R.E., Jr. Novel protein--protein interaction between *Escherichia coli* SoxS and the DNA binding determinant of the RNA polymerase alpha subunit: SoxS functions as a co-sigma factor and redeploys RNA polymerase from UP-element-containing promoters to SoxS-dependent promoters during oxidative stress. *J Mol Biol* **343**, 513-32 (2004).
64. Zuber, P. Management of oxidative stress in Bacillus. *Annu Rev Microbiol* **63**, 575-97 (2009).
65. Newberry, K.J., Nakano, S., Zuber, P. & Brennan, R.G. Crystal structure of the *Bacillus subtilis* anti-alpha, global transcriptional regulator, Spx, in complex with the alpha C-terminal domain of RNA polymerase. *Proc Natl Acad Sci U S A* **102**, 15839-44 (2005).
66. Lamour, V., Westblade, L.F., Campbell, E.A. & Darst, S.A. Crystal structure of the *in vivo*-assembled *Bacillus subtilis* Spx/RNA polymerase alpha subunit C-terminal domain complex. *J Struct Biol* **168**, 352-6 (2009).
67. Mangel, W.F. & Chamberlin, M.J. Studies of ribonucleic acid chain initiation by *Escherichia coli* ribonucleic acid polymerase bound to T7 deoxyribonucleic acid. I. An assay for the rate and extent of ribonucleic acid chain initiation. *J Biol Chem* **249**, 2995-3001 (1974).
68. Murray, H.D., Schneider, D.A. & Gourse, R.L. Control of rRNA expression by small molecules is dynamic and nonredundant. *Mol Cell* **12**, 125-34 (2003).
69. Schneider, D.A., Gaal, T. & Gourse, R.L. NTP-sensing by rRNA promoters in *Escherichia coli* is direct. *Proc Natl Acad Sci U S A* **99**, 8602-7 (2002).
70. Krasny, L. & Gourse, R.L. An alternative strategy for bacterial ribosome synthesis: *Bacillus subtilis* rRNA transcription regulation. *EMBO J* **23**, 4473-83 (2004).
71. Liu, K., Bittner, A.N. & Wang, J.D. Diversity in (p)ppGpp metabolism and effectors. *Curr Opin Microbiol* **24**, 72-9 (2015).

**A very clear account of how ppGpp (and pppGpp) plays different roles in different bacteria. A warning that what is true for *E. coli* may not be true for other bacteria, let alone elephants.**

72. Turnbough, C.L., Jr. & Switzer, R.L. Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. *Microbiol Mol Biol Rev* **72**, 266-300 (2008).
73. Martinez-Antonio, A. & Collado-Vides, J. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol* **6**, 482-9 (2003).
74. Ishihama, A. Prokaryotic genome regulation: a revolutionary paradigm. *Proc Jpn Acad Ser B Phys Biol Sci* **88**, 485-508 (2012).
75. Cho, B.K., Palsson, B. & Zengler, K. Deciphering the regulatory codes in bacterial genomes. *Biotechnol J* **6**, 1052-63 (2011).
76. Salgado, H. et al. Extracting regulatory networks of *Escherichia coli* from RegulonDB. *Methods Mol Biol* **804**, 179-95 (2012).
77. Dillon, S.C. & Dorman, C.J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol* **8**, 185-95 (2010).
78. Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. & Ishihama, A. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J Bacteriol* **181**, 6361-70 (1999).
79. Azam, T.A. & Ishihama, A. Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J Biol Chem* **274**, 33105-13 (1999).

80. Luijsterburg, M.S., Noom, M.C., Wuite, G.J. & Dame, R.T. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J Struct Biol* **156**, 262-72 (2006).
81. Browning, D.F., Grainger, D.C. & Busby, S.J. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr Opin Microbiol* **13**, 773-80 (2010).
82. Grainger, D.C., Goldberg, M.D., Lee, D.J. & Busby, S.J. Selective repression by Fis and H-NS at the *Escherichia coli dps* promoter. *Mol Microbiol* **68**, 1366-77 (2008).
83. Sobetzko, P., Glinkowska, M., Travers, A. & Muskhelishvili, G. DNA thermodynamic stability and supercoil dynamics determine the gene expression program during the bacterial growth cycle. *Mol Biosyst* **9**, 1643-51 (2013).
84. Sobetzko, P., Travers, A. & Muskhelishvili, G. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc Natl Acad Sci U S A* **109**, E42-50 (2012).
85. Dorman, C.J. Co-operative roles for DNA supercoiling and nucleoid-associated proteins in the regulation of bacterial transcription. *Biochem Soc Trans* **41**, 542-7 (2013).
86. Zhang, W. & Baseman, J.B. Transcriptional regulation of MG_149, an osmoinducible lipoprotein gene from *Mycoplasma genitalium*. *Mol Microbiol* **81**, 327-39 (2011).
87. Neumann, S. & Quinones, A. Discoordinate gene expression of *gyrA* and *gyrB* in response to DNA gyrase inhibition in *Escherichia coli*. *J Basic Microbiol* **37**, 53-69 (1997).
88. Lal, A. et al. Genome scale patterns of supercoiling in a bacterial chromosome. *Nat Commun* doi: 10.1038/ncomms1105 (2016).
89. Bryant, J.A., Sellars, L.E., Busby, S.J. & Lee, D.J. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res* **42**, 11383-92 (2014).
90. Swint-Kruse, L. & Matthews, K.S. Allostery in the LacI/GalR family: variations on a theme. *Curr Opin Microbiol* **12**, 129-37 (2009).
91. Semsey, S., Tolstorukov, M.Y., Virnik, K., Zhurkin, V.B. & Adhya, S. DNA trajectory in the Gal repressosome. *Genes Dev* **18**, 1898-907 (2004).
92. Butala, M. et al. Double locking of an *Escherichia coli* promoter by two repressors prevents premature colicin expression and cell lysis. *Mol Microbiol* **86**, 129-39 (2012).
93. Kamensek, S. et al. Silencing of DNase colicin E8 gene expression by a complex nucleoprotein assembly ensures timely colicin induction. *PLoS Genet* **11**, e1005354 (2015).
94. Valentin-Hansen, P., Sogaard-Andersen, L. & Pedersen, H. A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. *Mol Microbiol* **20**, 461-6 (1996).
95. Monsalve, M., Mencia, M., Salas, M. & Rojo, F. Protein p4 represses phage phi 29 A2c promoter by interacting with the alpha subunit of *Bacillus subtilis* RNA polymerase. *Proc Natl Acad Sci U S A* **93**, 8913-8 (1996).
96. Lee, D.J., Minchin, S.D. & Busby, S.J. Activating transcription in bacteria. *Annu Rev Microbiol* **66**, 125-52 (2012).
97. Benoff, B. et al. Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. *Science* **297**, 1562-6 (2002).
98. Gaston, K., Bell, A., Kolb, A., Buc, H. & Busby, S. Stringent spacing requirements for transcription activation by CRP. *Cell* **62**, 733-43 (1990).
99. Zhou, Y., Kolb, A., Busby, S.J. & Wang, Y.P. Spacing requirements for Class I transcription activation in bacteria are set by promoter elements. *Nucleic Acids Res* **42**, 9209-16 (2014).
100. Dove, S.L., Huang, F.W. & Hochschild, A. Mechanism for a transcriptional activator that works at the isomerization step. *Proc Natl Acad Sci U S A* **97**, 13215-20 (2000).
101. Jain, D., Nickels, B.E., Sun, L., Hochschild, A. & Darst, S.A. Structure of a ternary transcription activation complex. *Mol Cell* **13**, 45-53 (2004).

102. Niu, W., Kim, Y., Tau, G., Heyduk, T. & Ebright, R.H. Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell* **87**, 1123-34 (1996).
103. Feng, Y., Zhang, Y. & Ebright, R.H. Structural basis of transcription activation. *Science* **352**, 1330-3 (2016).

**This paper describes the first full molecular picture of Class II transcription activation at a target promoter, and contrasts with previous reports that have described components of the process.**

104. Brown, N.L., Stoyanov, J.V., Kidd, S.P. & Hobman, J.L. The MerR family of transcriptional regulators. *FEMS Microbiol Rev* **27**, 145-63 (2003).
105. Philips, S.J. et al. Allosteric transcriptional regulation via changes in the overall topology of the core promoter. *Science* **349**, 877-81 (2015).

**Working with CueR, which senses copper and silver ions, this paper presents high resolution structures of its complexes at a target sequence when either in activator or repressor mode. The results complement Dick Brennan's previous work (ref 106) showing that DNA distortion by such factors is uneven**.

106. Heldwein, E.E. & Brennan, R.G. Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature* **409**, 378-82 (2001).
107. Bustamante, V.H., Santana, F.J., Calva, E. & Puente, J.L. Transcriptional regulation of type III secretion genes in enteropathogenic *Escherichia coli*: Ler antagonizes H-NS-dependent repression. *Mol Microbiol* **39**, 664-78 (2001).
108. Sperandio, V. et al. Activation of enteropathogenic *Escherichia coli* (EPEC) LEE2 and LEE3 operons by Ler. *Mol Microbiol* **38**, 781-93 (2000).
109. Browning, D.F., Cole, J.A. & Busby, S.J. Transcription activation by remodelling of a nucleoprotein assembly: the role of NarL at the FNR-dependent *Escherichia coli nir* promoter. *Mol Microbiol* **53**, 203-15 (2004).
110. Browning, D.F., Cole, J.A. & Busby, S.J. Regulation by nucleoid-associated proteins at the *Escherichia coli nir* operon promoter. *J Bacteriol* **190**, 7258-67 (2008).
111. Tyson, K.L., Cole, J.A. & Busby, S.J. Nitrite and nitrate regulation at the promoters of two *Escherichia coli* operons encoding nitrite reductase: identification of common target heptamers for both NarP- and NarL-dependent regulation. *Mol Microbiol* **13**, 1045-55 (1994).
112. Bush, M. & Dixon, R. The role of bacterial enhancer binding proteins as specialized activators of sigma54-dependent transcription. *Microbiol Mol Biol Rev* **76**, 497-529 (2012).
113. Buck, M. et al. A second paradigm for gene activation in bacteria. *Biochem Soc Trans* **34**, 1067-71 (2006).
114. Casadesus, J. & Low, D. Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev* **70**, 830-56 (2006).
115. van der Woude, M.W. & Henderson, I.R. Regulation and function of Ag43 (flu). *Annu Rev Microbiol* **62**, 153-69 (2008).
116. Sanchez-Romero, M.A., Cota, I. & Casadesus, J. DNA methylation in bacteria: from the methyl group to the methylome. *Curr Opin Microbiol* **25**, 9-16 (2015).
117. van der Woude, M.W. & Baumler, A.J. Phase and antigenic variation in bacteria. *Clin Microbiol Rev* **17**, 581-611 (2004).
118. van der Woude, M.W. Phase variation: how to create and coordinate population diversity. *Curr Opin Microbiol* **14**, 205-11 (2011).
119. Cerdeno-Tarraga, A.M. et al. Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* **307**, 1463-5 (2005).
120. Moxon, R., Bayliss, C. & Hood, D. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet* **40**, 307-33 (2006).

121. Jacob, F. Evolution and tinkering. *Science* **196**, 1161-6 (1977).
122. Wade, J.T. & Grainger, D.C. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat Rev Microbiol* **12**, 647-53 (2014).
123. Grohmann, D. & Werner, F. Recent advances in the understanding of archaeal transcription. *Curr Opin Microbiol* **14**, 328-34 (2011).
124. Visweswariah, S.S. & Busby, S.J. Evolution of bacterial transcription factors: how proteins take on new tasks, but do not always stop doing the old ones. *Trends Microbiol* **23**, 463-7 (2015).
125. Grainger, D.C., Hurd, D., Harrison, M., Holdstock, J. & Busby, S.J. Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc Natl Acad Sci U S A* **102**, 17693-8 (2005).
126. Shimada, T., Ishihama, A., Busby, S.J. & Grainger, D.C. The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Res* **36**, 3950-5 (2008).
127. Aiyar, S.E. et al. Architecture of Fis-activated transcription complexes at the *Escherichia coli rrnB* P1 and *rrnE* P1 promoters. *J Mol Biol* **316**, 501-16 (2002).
128. Rossiter, A.E. et al. Expression of different bacterial cytotoxins is controlled by two global transcription factors, CRP and Fis, that co-operate in a shared-recruitment mechanism. *Biochem J* **466**, 323-35 (2015).
129. Lisser, S. & Margalit, H. Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Res* **21**, 1507-16 (1993).
130. Hook-Barnard, I.G. & Hinton, D.M. Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regul Syst Bio* **1**, 275-93 (2007).
131. Ross, W., Ernst, A. & Gourse, R.L. Fine structure of *E. coli* RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev* **15**, 491-506 (2001).
132. Sclavi, B. et al. Real-time characterization of intermediates in the pathway to open complex formation by *Escherichia coli* RNA polymerase at the T7A1 promoter. *Proc Natl Acad Sci U S A* **102**, 4706-11 (2005).
133. Davis, C.A., Bingman, C.A., Landick, R., Record, M.T., Jr. & Saecker, R.M. Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by *Escherichia coli* RNA polymerase. *Proc Natl Acad Sci U S A* **104**, 7833-8 (2007).
134. Buckle, M., Pemberton, I.K., Jacquet, M.A. & Buc, H. The kinetics of sigma subunit directed promoter recognition by *E. coli* RNA polymerase. *J Mol Biol* **285**, 955-64 (1999).
135. Stracy, M. et al. Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proc Natl Acad Sci U S A* **112**, E4390-9 (2015).

**Live microscopy can be used to observe individual molecules of RNA polymerase in *E. coli*. This is a prerequisite for future direct observation of responses to regulatory triggers.**

136. Patrick, M., Dennis, P.P., Ehrenberg, M. & Bremer, H. Free RNA polymerase in *E. coli*. *Biochimie* **119**, 80-91 (2015).
137. Hsu, L.M. Promoter Escape by *Escherichia coli* RNA Polymerase. *Ecosal Plus* **3** (2008).
138. Skancke, J., Bar, N., Kuiper, M. & Hsu, L.M. Sequence-dependent promoter escape efficiency is strongly influenced by bias for the pretranslocated state during initial transcription. *Biochemistry* **54**, 4267-75 (2015).
139. Bauer, D.L.V., Duchi, D. & Kapanidis, A.N. *E. coli* RNA polymerase pauses during initial transcription. *Biophysical Journal* **110**, 21a (2016).
140. Reppas, N.B., Wade, J.T., Church, G.M. & Struhl, K. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Mol Cell* **24**, 747-57 (2006).

141. Sendy, B., Lee, D.J., Busby, S.J. & Bryant J.A. RNA polymerase supply and flux through the *lac* operon in *Escherichia coli*. *Phil Trans R Soc B in press* (2016)
142. Jacob, F. La statue intérieure (Éditions Odile Jacob, Paris, 1987).
143. Müller-Hill, B. The *lac* Operon. A Short History of a Genetic Paradigm. (Walter de Gruyter, Berlin New York, 1996).
144. Schwartz, M. in Origins of Molecular Biology. A Tribute to Jacques Monod. (eds. Lwoff, A. & Ullmann, A.) 207-216 (Academic Press, Inc., New York, 1979).
145. Cases, I. & de Lorenzo, V. The black cat/white cat principle of signal integration in bacterial promoters. *EMBO J* **20**, 1-11 (2001).
146. Pul, Ü. & Wagner, R. in Bacterial Chromatin (eds. Dame, R.T. & Dorman, C.J.) 149-173 (Springer Science + Business Media, London, 2010).
147. Dorman, C.J. H-NS: a universal regulator for a dynamic genome. *Nat Rev Microbiol* **2**, 391-400 (2004).
148. Stoebel, D.M., Free, A. & Dorman, C.J. Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology* **154**, 2533-45 (2008).
149. Ohniwa, R.L. et al. Dynamic state of DNA topology is essential for genome condensation in bacteria. *EMBO J* **25**, 5591-602 (2006).
150. Micka, B. & Marahiel, M.A. The DNA-binding protein HBsu is essential for normal growth and development in *Bacillus subtilis*. *Biochimie* **74**, 641-50 (1992).
151. Grove, A. Functional evolution of bacterial histone-like HU proteins. *Curr Issues Mol Biol* **13**, 1-12 (2011).
152. Ball, C.A., Osuna, R., Ferguson, K.C. & Johnson, R.C. Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. *J Bacteriol* **174**, 8043-56 (1992).
153. Dame, R.T., Noom, M.C. & Wuite, G.J. Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature* **444**, 387-90 (2006).
154. Dame, R.T., Wyman, C. & Goosen, N. H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res* **28**, 3504-10 (2000).
155. Maurer, S., Fritz, J. & Muskhelishvili, G. A systematic *in vitro* study of nucleoprotein complexes formed by bacterial nucleoid-associated proteins revealing novel types of DNA organization. *J Mol Biol* **387**, 1261-76 (2009).
156. Grainger, D.C., Hurd, D., Goldberg, M.D. & Busby, S.J. Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res* **34**, 4642-52 (2006).
157. Kahramanoglou, C. et al. Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res* **39**, 2073-91 (2011).
158. Cho, B.K., Knight, E.M., Barrett, C.L. & Palsson, B.O. Genome-wide analysis of Fis binding in *Escherichia coli* indicates a causative role for A-/AT-tracts. *Genome Res* **18**, 900-10 (2008).
159. Singh, S.S. et al. Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev* **28**, 214-9 (2014).
160. Vora, T., Hottes, A.K. & Tavazoie, S. Protein occupancy landscape of a bacterial genome. *Mol Cell* **35**, 247-53 (2009).

**A novel application of genomics to monitor the protein landscape in different parts of a bacterial chromosome, with direct observation of tracts of the genome where transcription is silenced.**

161. Haugen, S.P., Ross, W. & Gourse, R.L. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat Rev Microbiol* **6**, 507-19 (2008).

**Box 1: Recognition of promoters by RNA polymerase.**

***Optimizing promoter strength.*** Bacterial promoter sequences contain motifs known as the -35 element, the extended -10 element, the -10 element, the discriminator region, the UP element, and the core recognition element. The activity of a promoter is set by these elements, with stronger promoters often having elements with sequences that better resemble the consensus sequences (Figure 1b). Bacteria have evolved a 'mix and match' approach to setting promoter strength, in which the strength of the individual promoter elements combine to determine the overall activity of the promoter. This modular organization probably has the benefit of enabling promoters to evolve easily to produce different output levels. Most naturally occurring promoters possess sequences for each element that are suboptimal for maximum activity[129], although the strength of each element is often not uniform across the promoter, such that different combinations of elements can generate promoters of similar strength[130]. An additional feature that contributes to promoter strength is the initiating nucleotide. For example, this is important for transcription initiation at ribosomal RNA operon promoters, where cellular concentrations of initiating nucleotides (either ATP or GTP) influence the formation and stability of the open complex[68, 69]**.**

***Double-stranded or single-stranded?*** For many promoter elements, recognition by RNA polymerase is specific for double-stranded DNA (dsDNA). For example, the C-terminal domains of the α subunits recognize the UP element by docking into adjacent minor grooves of the dsDNA[131], the -35 element is recognized specifically as dsDNA by a helix-turn-helix motif in Domain 4 of the sigma factor, and the dsDNA of the extended -10 element is bound by a long helix in Domain 3 of the sigma factor[2,3]. However, recent structural studies have shown that the -10 element and the discriminator region are both recognized as single–stranded DNA (ssDNA); that is, once the DNA has been unwound during the formation of the transcription bubble[4-7]. To bind to the RNA polymerase, specific bases from the non-template strand of the -10 element (the A and T bases at positions -11 and -7) and the discriminator region (the G at position -6) are flipped out and inserted into pockets in Domains 2 and 1, respectively, of the sigma factor. The core recognition element, which is formed by bases around the transcription start site, is recognized by residues in the RNAP β and β' subunits of the RNA polymerase[5-7]. The order in which each element is recognized by RNA polymerase during transcription initiation has been examined using rapid kinetic methods to study the A1 promoter of the T7 phage, which showed that the UP element and the -35 element are bound first, followed by unwinding of the DNA and recognition of the -10 element[132]. Other studies of the *lacUV5* and $\lambda P_R$ promoters showed that RNA polymerase seems to bind to dsDNA at the -10 element and downstream DNA, before the DNA is unwound to produce ssDNA[133, 134].

***Promoters as "bottlenecks".*** By visualizing fluorescently labelled RNA polymerase in live cells[135] or by measuring transcript levels as a proxy for RNA polymerase activity[136], studies in

*Escherichia coli* have shown that a pool of RNA polymerase holo enzyme that scans for promoters seems to be present on bacterial DNA in most conditions. However, as a promoter can only be occupied by a single RNA polymerase at any time, transcript formation at many promoters is subject to a "bottleneck" effect, in which formation of the open complex and transition to the elongation step of the transcription cycle become rate-limiting [137-139]. Accordingly, experiments using chromatin immunoprecipitation to measure promoter occupancy have shown that RNA polymerase is stalled at many promoters, which results in decreased flux of RNA polymerase through specific transcription units[140,141].

**Box 2: 50 years of studying gene regulatory mechanisms.**

In 1965, François Jacob, André Lwoff and Jacques Monod received a Nobel Prize for 'discoveries concerning the genetic regulation of enzyme and virus synthesis'. Together, they had discovered the mechanism by which lambda phage is silenced, as a lysogen in *Escherichia coli*, until exposure to UV light, and the mechanism by which β-galactosidase activity is silenced, unless *E. coli* is grown in a medium containing lactose. In both cases, the mechanism centred on a transcriptional repressor. Since the fundamentals of molecular biology had only just been established, transcription regulation had been a non-starter as a possible mechanism when the work began, and, indeed, the idea that the solution was a transcriptional repressor was resisted by Monod[142] until even he was persuaded by the clarity of genetic experiments[143]. The discovery that these functions are regulated at the level of transcription gave birth to a research area covering promoters, transcript start sites and operators, together with all the paraphernalia that mediates induction and repression of gene expression. From these studies emerged the concept that what matters is not the genes one has, but how one expresses them, and it is no exaggeration to state that all work in this area has its origins in the discovery of transcriptional repressors. And yet, despite these brilliant fundamental contributions, two shortcomings are notable in the writings of Monod. First, he assumed that all regulation would be explained by the activity of repressors and, unlike Jacob, was never persuaded that transcriptional regulation might also rely on activators[144]. Second, the notion popularized by Monod that 'anything found to be true of *E. coli* must also be true of elephants' gave undue prominence to *E. coli* and its *lac* operon as a model for molecular biology, the limitations of which were appreciated by Jacob[121] and, more recently, articulated by Victor de Lorenzo as the black cat-white cat principle[145].

**Box 3: Nucleoid-associated proteins.**

Although not enclosed by a membrane in a nucleus, bacterial chromosomes are compacted and highly folded, forming a structure known as the nucleoid. This compaction enables large

bacterial chromosomes to fit inside cells and is achieved by the combined effects of DNA supercoiling, molecular crowding, and the presence of RNA and nucleoid-associated proteins (NAPs). Initially, it was thought that NAPs would have similar structures to eukaryotic histones, but it is now clear that NAPs are, instead, a diverse group of proteins that recognize DNA target sites using many different structural motifs[77, 146]. *Escherichia coli* has at least 12 different NAPs, including Fis (Factor for Inversion Stimulation), IHF (Integration Host Factor), HU (Histone protein from strain U9), H-NS (Histone-like Nucleoid-structuring protein) and LRP (Leucine Responsive Protein), all of which have been studied extensively[77, 79]. The most studied example is probably H-NS, which forms filaments along AT-rich segments of DNA; these filaments then form bridges between different segments to repress the activity of any associated promoter[147]. How the activity of H-NS is modulated is not fully understood, but one mechanism used to relieve repression is the expression of H-NS homologues, such as Ler, that are incorporated into the filaments but are unable to form the repressive bridges[148].

Much of what we understand about NAPs comes from studies with *E. coli*; however, the set of NAPs present in any bacterium varies according to the species. For example, Fis is restricted to Gammaproteobacteria, whereas HU is found in almost all bacteria and is the principal NAP in *Bacillus subtilis*, in which it is an essential protein[77, 149, 150]. Some species of *Actinomycetes* have a NAP, Hlp, that has an N-terminal domain that is similar to HU, but a C-terminal domain that resembles eukaryotic histone H1[151]. Adding further complexity, the levels of some NAPs change in response to growth conditions and growth phase, which means that the composition of a cell's nucleoid can change over time. For example, rapidly growing *E. coli* cells at early exponential phase contain over 50,000 molecules of Fis, whereas the number of Fis molecules in cells at stationary phase is close to zero[78, 152].

Characterization of the binding preferences of individual NAPs has shown that some, such as IHF and Fis, are sequence-specific DNA-binding proteins, whereas others, such as H-NS and HU, bind to DNA non-specifically, often to distorted DNA structures[79]. Biochemical, biophysical and single-molecule experiments have shown that NAPs package DNA by several different mechanisms. For example, IHF and Fis package DNA by bending the target site, and, as mentioned above, H-NS forms bridges between DNA molecules[77, 146, 153-155]. Chromatin immunoprecipitation studies have determined the genome-wide binding profiles of some NAPs, which have shown that many NAPs bind to hundreds of target sites [156-158]. Many of these target sites are located in intergenic regulatory regions, where several different NAPs may bind[156]. Thus, it is not surprising that many examples have been described where NAPs regulate gene expression by modulating directly promoter activity[81]. For example, the formation of extended filaments of H-NS on DNA silences many promoters[147], and even prevents the unwanted expression of horizontally acquired A/T-rich genes, which often encode virulence factors[159]. By contrast, Fis is able to repress transcription initiation by a variety of mechanisms and can also activate transcription directly by interacting with holo RNA polymerase using an activating region[127, 128]. Other

NAPs can prevent transcription over a much larger segment of a bacterial chromosome and a genome-wide survey of the *E. coli* chromosome identified over 100 loci at which transcription seemed to be silenced[160]. The ability of NAPs to regulate transcription suggests that evolution has blurred the distinction between NAPs and transcription factors so as to ensure both nucleoid integrity and coordinated gene regulation[124].

## Text Figure Legends

Figure 1 | **Transcription of bacterial genes.**

**a |** The bacterial transcription cycle. Holo RNA polymerase, which comprises the RNA polymerase core enzyme and a sigma factor, interacts with promoter DNA to form the closed complex. The closed complex transitions to the open complex by unwinding the duplex DNA in the region of the transcription start site. The addition of nucleoside triphosphates enables a further transition to the initiating complex, which synthesizes the RNA transcript. Initially, the template DNA is pulled into the initiating complex, which is known as 'scrunching'. The scrunched complex can be held at the promoter, which results in cycles of abortive initiation that only produce small RNA fragments. Alternatively, the RNA polymerase can escape the promoter to enter the elongation phase, leading to the release of the sigma factor and elongation of the RNA transcript using NTPs and elongation factors (not shown). Transcription proceeds until the RNA polymerase encounters a transcriptional terminator, after which the RNA transcript is released and the polymerase dissociates from the DNA template to reengage with sigma to repeat the cycle.

**b |** The principal DNA elements that are recognized by RNA polymerase at bacterial promoters include the UP element (positions -37 to -58, if the transcriptional start site is denoted +1), the -35 element (positions -35 to -30), the extended -10 element (Ext) (position -17 to -14), the -10 element (positions -12 to -7) and the discriminator element (Dis) (-6 to -4). Note that the exact positions of each element can vary according to the particular promoter. The regions of the sigma factor (Domains 1 to 4), and the C-terminal domain of the α subunit of RNA polymerase, that contact these promoter elements are indicated schematically. The figure is adapted from reference[161].

Figure 2 | **Modulation of RNA polymerase activity.**

**a |** Polymerase centric transcription factors. Both RbpA and CarD are global transcription factors that bind directly to RNA polymerase to regulate transcription at promoters. RbpA contacts Domain 2 of the housekeeping sigma factor, whereas CarD binds to the β subunit of RNA polymerase. Both are positioned to stabilize the open complex by interacting with promoter DNA in addition to the RNA polymerase. Adapted from reference[44].

**b |** Regulation by a promoter mimic. The structure of the 6S RNA mimics that of a promoter in an open complex, which enables the 6S RNA to sequester RNA polymerase holo enzyme containing the housekeeping sigma factor, and thus down-regulate global transcription .

**c |** Sequestration of a sigma factor domain. On infection of *E. coli* by 'phage T7*,* the 'phage Gp2 protein binds to two targets of the *E. coli* RNA polymerase holo enzyme: the channel that contains the active site and Domain 1 of the sigma factor. This inhibits transcript initiation by preventing DNA from entering the active site, and by stopping the movement of Domain 1 of the sigma factor. Adapted from reference[14].

**d |** Sigma factor displacement. On infection of *Xanthomonas oryzae* by 'phage Xp10, the 'phage P7 protein binds to the β and β' subunits of the bacterial RNA polymerase, which inhibits transcription by displacing Domain 4 of the sigma factor from the holo enzyme when the RNA polymerase is associated with promoter DNA. Adapted from reference[58].

Figure 3 | **Appropriation of RNA polymerase for specific transcriptional programmes.**

**a |** During infection of *Escherichia coli*, the AsiA protein of 'phage T4 remodels the structure of Domain 4 of the host housekeeping sigma factor so that it is able to interact directly with

the ‘phage MotA transcription activator protein, bound at ‘phage target promoters, thereby activating expression of the middle stage genes of the T4 infection cycle.
**b |** Early during infection, the Alt ‘phage T4 protein ADP-ribosylates amino acid residue R265 on the C-terminal domain of the α subunit of *E. coli* RNA polymerase. This reduces recognition of UP element sequences at bacterial promoters, thereby increasing transcription at T4 promoters controlling the genes of the early stage of the infection cycle.
**c |** In *E. coli*, SoxS is expressed in response to oxidative stress. By binding to the C-terminal domains of the α subunits of RNA polymerase, SoxS directs RNA polymerase to promoters that contain ‘Soxbox’ SoxS-recognition sites in upstream regions, which leads to the expression of the genes controlled by these promoters.
**d |** In *Bacillus subtilis*, the Spx protein is activated in response to thiol-oxidative stress. Activated Spx is able to associate with the C-terminal domains of the α subunits of RNA polymerase, which modulates the binding preferences of RNA polymerase, such that expression of *trxB* and other genes that combat thiol-oxidative stress is induced.

Figure 4 | **Repression or activation at promoters by transcription factors.**
The principal mechanisms by which transcription factors directly repress or activate transcription initiation by holo RNA polymerase carrying *Escherichia coli* $\sigma^{70}$, or a related sigma factor, at promoters.
**a** | In repression by steric hindrance, the repressor binds to a site overlapping the core elements of the promoter recognized by RNA polymerase, and thus blocks binding of RNA polymerase to the promoter. In repression by looping, protein–protein interactions are formed between repressors that bind to sites upstream and downstream of the promoter, inducing looping of the DNA between the two repressor binding sites. The looping prevents recognition of promoter elements by RNA polymerase and transcription is repressed. For some genes that require activators for transcription, repressors can repress transcription by modulation of these activators so that RNA polymerase is no longer recruited by them. These repressors target activators, rather than directly masking promoter elements from RNA polymerase, but the ultimate effect remains to prevent recognition of the promoter and thus repress transcription.
**b** | In Class I activation, the activator binds to a site upstream of the promoter and recruits RNA polymerase to the promoter by contacting the C-terminal domain of the α subunit. In Class II activation, the activator binds to a site in the promoter adjacent to or overlapping the -35 element, where it recruits RNA polymerase through direct interactions with Domain 4 of the sigma factor. Rather than directly recruiting RNA polymerase, some activators activate transcription by inducing a conformation change in the promoter DNA. These activators bind at, or near to, the core RNA polymerase recognition elements of the promoter, and often realign the -10 and -35 elements so that they can be recognized by RNA polymerase, thereby enabling the recruitment of RNA polymerase to the promoter and activation of transcription.

Figure 5 | **Regulation by promoter DNA modification.**
**a** | Regulation by DNA methylation. To repress transcription of *agn43*, the OxyR transcription factor binds to sites that include unmethylated GATC sequences in the promoter region of the *agn43* locus. This results in transcriptional repression because OxyR blocks access of RNA polymerase to the promotor. After DNA replication, the Dam methylase can methylate the adenines of these GATC sequences on one or both strands,

which prevents OxyR binding. Once the GATC sites are no longer occupied by OxyR, RNA polymerase is able to bind to the promoter and initiate transcription.

**b** | Regulation by DNA inversion. The *fim* operon contains *fimA* and other genes. The expression of *fimA* is regulated by inversion of a DNA element in the promoter region of the gene. The inversion is mediated by FimB and FimE recombinases and switches the orientation of the promoter between off and on states. In the 'off' state, the promoter is oriented away from *fimA*, which results in the production of non-coding transcripts, whereas, in the 'on' state, the promoter orientation enables the production of *fimA* transripts. The *fim* locus is not drawn to scale.

**c** | Regulation by local sequence variation. In *Haemophilus influenza*, *hifA* and *hifB* encode fimbrial components. The promoter of these genes has a variable number of TA dinucleotide repeats, which alters the spacing between the -10 and -35 elements. As the spacing between these elements is a determinant of RNA polymerase holo enzyme binding, and thus promoter strength, variation of the number of repeats produces changes in gene expression, such that expression can be switched off entirely or adjusted to different levels.

## Glossary

**Actinomycetes:** *A class of soil bacteria with a particular morphology*

**Chromatin immunoprecipitation (ChIP):** *A method whereby antibodies are used to isolate DNA fragments that have been crosslinked to a specific protein*

**Coiled-coil:** *An extended motif found in proteins*

**RNA polymerase core enzyme:** *The form of bacterial DNA-dependent RNA polymerase that lacks a sigma factor*

**Closed complex:** *The complex between RNA polymerase and a promoter before DNA duplex unwinding has occurred*

**Enterohemorrhagic *E. coli*:** *A virulent strain of E. coli that causes bloody diarrhoea*

**Housekeeping sigma factor:** *The sigma factor in a bacterium that is responsible for recognition of promoters that control the transcription of most genes*

**Initiating nucleotide:** *The 5' nucleotide of a transcript.*

**Michaelis constant:** *This is a key character of any enzyme and refers to the concentration of its substrate at which the catalysed reaction proceeds at half of its maximum spreed.*

**Nucleoid:** *The structure that forms after a bacterial chromosome is compacted inside a bacterium*

**Open complex:** *The complex between RNA polymerase and a promoter after DNA duplex unwinding has occurred and the RNA polymerase is ready to start transcription*

**Pervasive transcription:** *The synthesis of transcripts that appear not to correspond to any functional genetic unit*

**ppGpp(p):** *3' diphospho guanosine 5' di(tri)phosphate. A small molecule synthesised in response to certain stresses.*

**Stationary phase:** *The period when bacteria have stopped growing*

**Superhelical density:** *The measure of the degree to which the winding of one DNA strand around the other differs from the periodicity of the Watson-Crick structure*

**Template strand:** *The strand of the DNA duplex that acts as a template for RNA synthesis*

**ONLINE ONLY**

## KEY POINTS:

- **Transcript initiation involves the interaction of DNA-dependent RNA polymerase with promoters**
- **In bacteria, transcription initiation is highly regulated**
- **Many regulators interact directly with the bacterial DNA-dependent RNA polymerase**
- **Other regulators interact directly with promoters**
- **Regulation occurs in the context of folding and compaction of bacterial chromosomes**
- **A very wide range of different strategies to regulate transcription initiation are used and these differ from one species to another**