activation and repression of transcription initiation in bacteria

georgina lloyd, paolo landini\(^1\) and steve busby\(^2\)

*school of biosciences, the university of birmingham, edgbaston, birmingham b15 2tt, u.k.*

**introduction**

a remarkable feature of bacteria is their ability to respond to environmental stimuli and to adapt to changing environmental conditions. rapid adaptation is achieved by switching on and off the expression of specific genes. bacteria are very efficient at increasing the production of certain proteins and enzymes when they are needed, only to shut off their production when it becomes metabolically and energetically wasteful. tight control of bacterial gene expression can be achieved at many levels, including regulation of transcription initiation, transcript elongation, translation, messenger rna stability or availability and protein turnover. these mechanisms are not mutually exclusive, and expression of some genes is subject to all the above levels of regulation. in this chapter we focus on regulation at the level of transcription initiation simply because, in most cases, it is the dominant form of regulation. most of this chapter concerns regulation in the simple enteric bacterium *escherichia coli*, which is rightly regarded as a paradigm that can instruct our

\(^1\)present address: department of microbiology, swiss federal institute for environmental science and technology, 133 überlandstrasse, ch-8600 dubendorf, switzerland.

\(^2\)to whom correspondence should be addressed (e-mail: s.j.w.busby@bham.ac.uk).
understanding of other bacterial systems. Additionally, *E. coli* was the first organism in which the molecular biology of gene expression was investigated in depth. The logic of these early investigations has been repeatedly applied to other more complex systems. An excellent account of these early studies and their relevance today can be found in [1].

**Transcription regulation**

The complete genome sequence of *E. coli* has revealed that the number of potential genes is 4290 [2]. However, the protein responsible for all transcription, DNA-dependent RNA polymerase holoenzyme (RNAP), is present at only 2000 molecules/cell (see [3] for an account of the quantitative aspects of transcription). Thus, because of the relative scarcity of the transcription apparatus, *E. coli* cells need to be very efficient in directing transcription in response to stimuli and stresses and, because of this, gene regulation in bacteria is principally mediated at the level of transcription initiation. This regulation is an elaborate process, involving many proteins. The most important of these proteins are the RNAP σ subunits, which confer different promoter specificity and therefore direct RNAP towards different sets of genes. Next, there is a set of ~150 transcription regulators that up-regulate and down-regulate expression from specific sets of promoters in response to particular environmental signals. Control of bacterial gene expression at the level of transcription initiation has been described for genes involved in almost all cellular processes, from energy metabolism to DNA repair and replication, and to responses to environmental stresses. At some promoters, more than one transcription factor can interact and co-regulate transcription. Thus, we first give a brief outline of the bacterial RNAP, placing particular emphasis on the σ subunit. Next, we describe how transcription factors interact with RNAP to up-regulate and down-regulate expression from promoters. Finally, we consider how promoters are designed to respond to multiple signals transduced via different transcription regulators.

**Promoter recognition by RNAP — role of the σ subunit**

*E. coli* RNAP consists of five subunits: two identical α subunits and one copy each of the β, β′ and σ subunits [4]. RNAP (α₂ββ′σ) can be separated into two components: the core enzyme (α₂ββ′) and the σ subunit. The core enzyme can bind to DNA in a non-specific fashion and retains the catalytic activity of RNAP, i.e. the ability to synthesize RNA from a DNA template. However, the core enzyme cannot initiate transcription at promoters; the σ subunit is necessary for promoter recognition and regulated transcription initiation [5]. Assembly of a σ subunit into RNAP introduces a major change in the RNAP–DNA interaction, reducing non-specific binding to DNA and conferring the ability to recognize particular promoters. This σ-subunit-dependent RNAP–promoter interaction involves several steps. Initial binding
to the promoter results in the so-called ‘closed complex’, which is then converted into the ‘open complex’ by local denaturation (‘melting’), of the DNA sequence immediately surrounding the transcription start point. On formation of the open complex, the synthesis of the transcript is initiated. RNAP then moves away from the promoter, and after about 20 nt have been incorporated into the nascent RNA chain, the σ subunit is released.

If the core enzyme is considered as the ‘constant’ part of RNAP, the σ subunit is the ‘variable’ part. The *E. coli* genome encodes seven different σ subunits; incorporation of each different σ subunit into RNAP results in recognition of different sets of promoters [6]. Most genes expressed in rapidly growing cells are transcribed from promoters dependent on the σ70 subunit, which is considered the main σ subunit in *E. coli*. The σ70 protein contains 613 amino acids and is the product of the *rpoD* gene. Despite being responsible for the specific binding of RNAP to DNA, σ70 does not bind to DNA as a free subunit, but only when assembled into RNAP. This is due to the inhibitory activity of the σ70 N-terminal region (region 1). Deletion of this region allows σ70 to bind to DNA as a free subunit. A typical σ70-dependent promoter contains two conserved hexamer sequences located near positions −10 and −35 relative to the transcription start point. The consensus sequences for these two promoter elements are TATAAT and TTGACA, respectively. These elements are recognized by distinct regions of the σ70 subunit: region 2 for the −10 sequence and region 4 for the −35 sequence (Figure 1).

Alternative σ subunits are involved in the induction of specific sets of genes that are required for adaptation and cell survival during environmental stress [5–7]. For example, σ32 is required when *E. coli* is grown at temperatures higher than the optimal 37°C, and triggers the expression of the heat-shock

---

**Figure 1. Interactions between RNAP and promoter elements at an activator-independent promoter**

The σ subunit of RNAP is responsible for contacting the −10 and −35 hexamers and is shown divided into four distinct regions (shaded). Region 2 and region 4 are involved in promoter recognition; they contact the −10 sequence and −35 sequence, respectively. The α subunits of RNAP consist of two domains, αNTD and αCTD, that are joined by a flexible linker. αCTD provides additional sequence-specific protein–DNA interactions by binding the upstream element (UPE), usually located immediately upstream of the −35 hexamer, thus strengthening RNAP binding. An arrow indicates the transcription start point and the direction of transcription.
genes. A further σ subunit, σE, is responsible for transcription of the so-called extreme heat-shock genes. Another example concerns the transition from exponential growth to stationary phase, when a set of genes essential for survival is expressed. The σ subunit that controls expression of these genes is σS. In addition to controlling many stationary-phase-specific genes, σS is involved in several stress responses. It controls the expression of some genes involved in oxidative stress and in the adaptive response to methylation damage even during exponential bacterial growth.

**Subunits of the RNAP core**

The structure at 3.3 Å resolution of core RNAP from *Thermus aquaticus* has been reported [8]. Sequence analysis suggests that this structure, although not identical with that of the *E. coli* core RNAP, is very similar. The structure shows a large cleft, which accommodates the template DNA. One face of the cleft is lined by the β subunit and the opposite face is lined by the β′ subunit. Side chains from the β and β′ subunits form the enzyme active site and the binding sites for template DNA and product RNA.

The β subunit (151 kDa; encoded by the *rpoB* gene) can be considered as the main catalytic subunit of RNAP. It binds ribonucleoside triphosphates and promotes polymerization of the RNA chain. The β subunit is the target of the transcription inhibitors, rifampicin and streptolydigin. It also carries the target for the alarmone molecules, ppGpp and pppGpp, that are synthesized as part of the stringent response triggered by starvation for amino acids. Alarmone binding to the β subunit shuts down transcription of ribosomal RNA by preventing transcription initiation.

The function of the β′ subunit (155 kDa; encoded by the *rpoC* gene) is not fully understood. The β′ subunit is rich in positively charged amino acids, and is thought to be the main contributor to non-specific binding by core enzyme to the DNA template. Indeed, the β′ subunit binds avidly to DNA and to heparin, a polyanion whose structure mimics DNA. The β′ subunit has been found to be the target of the N4 bacteriophage single-stranded DNA-binding protein during transcriptional activation of the N4 late promoters. This suggests that β′ also plays a role in transcription initiation, either directly or via interaction with the σ subunit.

The RNAP α subunit (37 kDa; encoded by the *rpoA* gene) is the only subunit that is present as a dimer. It performs three critical functions [9] — it is the initiator of RNAP assembly, it contributes to promoter recognition by sequence-specific interaction with DNA and it interacts with many transcription factors, both in transcription initiation and in anti-termination. Structural and biochemical studies indicate that α consists of two independently folded domains with distinct functions: the N-terminal domain (αNTD) of roughly 230 amino acids and the C-terminal domain (αCTD) of approximately 80 amino acids, linked by a flexible linker region (Figure 1). This α linker region
is highly sensitive to protease cleavage. αNTD carries the determinants for α dimerization, the first step in RNAP assembly, which follows the pathway $2\alpha_2\beta\epsilon\alpha_2\beta\epsilon\alpha_2\beta\epsilon\alpha_2\beta\epsilon\alpha_2$. αCTD carries the determinants for recognition of a 20 bp promoter element known as the upstream element (UPE), first identified in the promoters of the genes encoding ribosomal RNA [10]. At these promoters, the UPE, which is located just upstream of the $-35$ and $-10$ elements, is a major contributor to promoter activity, stimulating transcription initiation by at least 30-fold. The binding of αCTD to the UPE strengthens RNAP binding (Figure 1). In addition, αCTD carries the contact sites for a large number of transcription activators, including the well-characterized cAMP receptor protein (CRP) [11]. Different transcription activators target different regions of αCTD, but in most cases, the interaction appears to stimulate the binding of αCTD, and hence the rest of RNAP, to the promoter DNA (Figure 2a). Interestingly, at some promoters, CRP also interacts with the αNTD, showing that αNTD can also function as a target of transcription activators [12]. Note that some of the determinants in αCTD that are important for interaction with the UPE and with transcription activators are also involved in the interaction with termination and anti-termination proteins, and possibly bind to the nascent RNA chain. This suggests parallels between transcription initiation and the resumption of transcription elongation during anti-termination.

**Negative and positive control of transcription initiation**

The expression of many bacterial genes is regulated at the initiation of transcription by transcription factors that interact at or near the corresponding promoter. The majority of transcription factors are DNA-binding proteins, and their promoter specificity is determined by preferential binding to particular base sequences. Most transcription factors function as dimers, or as higher multimers. In many cases, each subunit of the dimer (or multimer) contributes to specific DNA binding. Regulation results from the interaction of such transcription factors with promoter DNA and often also involves interactions with RNAP. These transcription factors can increase the rate of transcription initiation (activators) or prevent RNAP from initiating transcription (repressors). As mentioned above, during transcription initiation RNAP progresses through the ‘closed complex’ into the ‘open complex’, in which the promoter DNA is unwound at the transcription start. At this stage RNAP initiates RNA synthesis and must clear the promoter before entering the elongation phase of transcription. Activators and repressors can, at least in theory, affect all these steps: initial binding of RNAP to the promoter, open-complex formation, or promoter clearance [13]. A surprising discovery was that some transcription factors can function as both activators and repressors. Examples of such proteins are the bacteriophage λ repressor, which inhibits transcription from the $P_R$ and $P_L$ promoters but activates transcription at $P_{RM}$.
and the CRP protein, which activates transcription of many *E. coli* genes (the *lac* operon is the best known example) but represses others (e.g. the proP1 promoter). It appears that the effect of a transcription factor is determined principally by the location of its binding site in the promoter region. Thus, for example, CRP activates transcription when its binding site is centred near positions $-61.5$ or $-41.5$ upstream of the transcription start, but it interferes with RNAP binding, and therefore represses transcription, when the binding site is centred near positions $-51.5$ or $-34.5$. 

**Figure 2. Activation of transcription by activator–RNAP interactions**

In each diagram, the activator is shown as a dimer (e.g. as for CRP, bacteriophage λ ci protein, etc.). (a) The activator (A) makes a specific protein–protein interaction (●) with αCTD, which recruits αCTD to the DNA and increases the overall affinity of RNAP for the promoter. (b) The activator (A) makes a specific protein–protein interaction (■) with region 4 of σ, to either recruit RNAP to the promoter or enhance a subsequent step in transcription initiation. αCTD is shown displaced and making no specific contacts with the promoter DNA. (c) The activator (A) contacts both αCTD and region 4 of σ (● and ■, respectively). Each interaction contributes to transcription initiation.
Transcription activators

The majority of transcription activators bind to DNA sites located either upstream of, or overlapping, the target promoter $-35$ element [14]. Activators can function by either of two alternative modes of action: direct protein–protein interaction with RNAP, or introduction of a conformational change in the promoter DNA that makes it accessible to RNAP [15]. Both biochemical and genetic data suggest that most activators make direct contact with RNAP. In particular, the identification of mutant activator proteins that result in activation defects without affecting DNA binding (positive control mutants) has supported the idea that direct protein–protein interaction is the principal mechanism for transcription activation. The locations of the substitutions in these mutants define the region in the activator protein involved in contact with RNAP. In the same way, contact surfaces on RNAP have been identified from the location of substitutions that interfere with transcription activation without affecting factor-independent transcription initiation. Since the targets for activators can be located on different subunits of RNAP, it is possible to classify activators according to their target in RNAP.

Activators that contact $\alpha$CTD

Many activators contact the $\alpha$CTD [9]. The best characterized of these activators is CRP, but FNR, IHF, Fis, OxyR, OmpR, and CysB are also thought to interact with $\alpha$CTD. The location for the binding site of these activators can vary considerably, from near position $-91$ to near position $-41$; however, it is always located upstream of the $-35$ sequence. This variability in the location of activators that contact $\alpha$CTD is due to the presence of the flexible linker between the $\alpha$NTD and $\alpha$CTD [16]. Activation by these factors also requires DNA binding by the RNAP $\alpha$ subunits; the activator appears to facilitate docking of RNAP to the promoter via an $\alpha$CTD–DNA interaction (Figure 2a). The consequence of this is that these activators primarily accelerate formation of the closed complex. Support for this comes from experiments showing that artificial protein–protein contacts can replace the CRP–$\alpha$CTD interaction and activate transcription at the lac promoter [17]. Thus conformational changes in RNAP relayed by activator–$\alpha$CTD contacts do not appear to play a major role.

Activators that contact the RNAP $\sigma$ subunit

Together with $\alpha$CTD, the $\sigma^{70}$ subunit of RNAP is the major target of transcription activators. Significantly, these two RNAP segments both make specific contacts with the promoter DNA ($-35$ and $-10$ sequences for $\sigma^{70}$ and the UP element for the $\alpha$ subunit) [16]. Activators known to interact with the $\sigma^{70}$ subunit are PhoB, FNR and several bacteriophage proteins ($\lambda$ cI protein, T4 MotA protein, and Mu Mor protein) [5]. Activators that contact the RNAP
σ70 subunit bind to DNA sites that overlap the target promoter −35 element (Figure 2b). Substitutions in σ70 that specifically affect activator-dependent transcription map between amino acids 570 and 580 and amino acids 591 and 613, located in region 4 close to the recognition motif for the −35 hexamer. Interestingly, a consensus-like −35 promoter element is usually absent. Thus protein–protein interaction between the activator and σ70 appears to substitute for the usual protein–DNA contact found at activator-independent promoters. This interaction can either recruit RNAP to the promoter region, similar to the role played by activators that contact αCTD, or enhance a subsequent step in transcription initiation. In the case of PhoB, initial recruitment of RNAP to the promoter region is likely to be the step affected by the activator, whereas the bacteriophage λ cI protein increases open-complex formation. It is noteworthy that several activators that contact σ70, such as FNR and the Mu Mor protein, can interact simultaneously with αCTD (Figure 2c). It is likely that the two different interactions contribute to different steps in transcription initiation. The activator–αCTD interaction enhances recruitment of RNAP to the promoter, whereas the activator–σ interaction facilitates isomerization of the closed complex to the open complex [15].

Although most studies have focused on σ70, the alternative σ subunits can also be targets for transcription activators. However, the extent of this kind of regulation is poorly understood at present. For example, activators such as Ada and CRP have been shown to contact RNAP containing σ5, but the physiological significance of this is not known. It is likely that some activators are able to interact with similar contact sites in many different σ subunits.

Activators that contact other surfaces in RNAP
In principle, an activator could recruit RNAP to a promoter by contacting any surface, and a number of activators are known to contact RNAP surfaces other than αCTD and region 4 of σ70 [3]. Thus the DnaA protein activates the bacteriophage λ P_R promoter via interaction with the β subunit of RNAP. At CRP-dependent promoters, where the CRP binding site overlaps the promoter −35 sequence, CRP contacts αNTD, in addition to interacting with αCTD [12]. CRP contains two distinct surface-exposed regions that interact with αCTD and αNTD. Interestingly, at these promoters, the contact with αCTD is made by the upstream subunit of the CRP dimer and the contact with αNTD is made by the downstream subunit. The two contacts make distinct contributions to transcription activation. The CRP–αCTD interaction recruits RNAP to the promoter, whereas the CRP–αNTD interaction accelerates open-complex formation [12].

Activators that alter promoter DNA conformation
Several transcription activators appear to function primarily by altering the conformation of the target promoter DNA, without making direct contact with any of the RNAP subunits [15]. For example, some activators bind to
targets upstream from the promoter elements and introduce a loop in the DNA that allows RNAP to contact distal DNA sequences, stabilizing its interaction with the promoter (Figure 3a). Another group of activators that function by altering DNA conformation are the members of the MerR family. For example, MerR, when triggered by mercuric ions, activates transcription of the merP gene (part of the mercury resistance locus of the Tn501 transposon). Unlike most activator-dependent promoters, the merP promoter carries −10 and −35 hexamer sequences that resemble the consensus. However, the spacing between these two elements (19 bp) is greater than the optimal spacing (17 bp) found in activator-independent promoters. Thus levels of transcription from the merP promoter in the absence of the activator are extremely low. However, when triggered by mercuric ions, MerR binds between the −10 and −35 hexamer elements at the merP promoter and induces a local twisting of the promoter DNA (Figure 3b). This results in repositioning of the −10 and −35 promoter elements so that efficient transcription initiation can take place.

Figure 3. Activation of transcription by activator–promoter DNA interactions
(a) An upstream-bound activator (A) introduces a loop in the promoter DNA. This allows RNAP to contact distal promoter sequences, which stabilize its interaction with the promoter. (b) The −10 and −35 hexamers are repositioned due to local twisting of the promoter DNA induced by the binding of the activator (indicated by the arrows). RNAP can now bind and initiate transcription.
**Transcription repressors**

Many repressors prevent transcription by binding DNA at positions that interfere directly with RNAP binding [14]. Thus at many promoters subject to repression, operator sequences for a repressor are found to overlap or be immediately adjacent to the transcription start site (Figure 4a). A well studied example is the *E. coli lac* repressor, the first transcription regulator to be identified, which binds to a target site centred at position +11 with respect to the *lac* operon transcription start point. It has been found that the *lac* repressor, and many other repressors, function simply by interfering with the binding of RNAP to the target promoter. For factors that can act both as an activator or a repressor, it is the location of the binding site that determines their effect on transcription initiation.

---

**Figure 4. Mechanisms of transcriptional repression**

In each diagram, the repressor is drawn as a dimer. (a) Repressor (octagon) blocks access of RNAP by binding to DNA and interfering directly with RNAP binding. (b) Distally bound repressors form a repression loop that hinders access of RNAP to the promoter. (c) Anti-activation mechanism: a repressor (R) binds to a tandemly bound activator (A) and prevents activation.
In many cases the mechanism of repression, although apparently simple, is unexpectedly complex. For example, for efficient repression of lac transcription, one of two other binding sites for the lac repressor is required in addition to the binding site at position +11 [18]. One of these supplementary sites is located at position −82 and the other is at +412. The presence of additional binding sites (operators) is a common theme in negative control of transcription initiation in bacteria – binding of more than one repressor molecule increases the local concentration of the repressor. Additionally, multiple binding sites can result in co-operative interactions that improve the affinity of the repressor for DNA, or in the formation of DNA micro-loops that hinder access of RNAP to the promoter (Figure 4b). Similar complexities are found during repression of the gal operon promoter by GalR, a homologue of the lac repressor. GalR binds two target sites centred at positions −60.5 and +53.5 and, together with the histone-like protein HU, creates a micro-loop around the gal promoter region [19]. Recent work has proposed that GalR does not prevent RNAP binding to the gal promoter, but that direct interaction between GalR and αCTD of RNAP maintains RNAP in an inactive state [20]. Thus, for full repression of gal transcription, GalR protein interferes both with the initial binding of RNAP to the promoter (via micro-loop formation) and with a subsequent step (via direct protein–protein interaction).

**Transcription regulation by a repressor and an activator**

Many naturally occurring bacterial promoters are extremely complex and most are regulated by more than one transcription factor. This permits the synthesis of a particular gene product to be controlled by more than one environmental stimulus; in effect, the promoter integrates information from different signals. For simplicity, we consider here only promoters that are regulated by two factors. Clearly there are three possible scenarios: regulation by two repressors, regulation by a repressor and an activator, and regulation by two activators. Although there are very few instances of promoters that are controlled by two repressors, there are many cases of regulation by a repressor and an activator. In most of these the repressor and the activator apparently function independently. For example, the lac promoter is active only when the lac repressor is removed from its target at position +11 (in response to lactose in the environment) and when the activator, CRP, binds to its target at position −61.5 (in response to low levels of glucose in the environment). However, there are cases where the functions of the repressor and the activator are not independent (Figure 4c). For example, promoters that control gene products needed for pyrimidine salvage are dependent on CRP for activity, but are repressed by CytR (a repressor whose activity is controlled by cytidine). At these promoters CytR binds directly to CRP and prevents CRP from activating transcription [21]. Thus CytR can be considered as an anti-activator rather than a simple repressor. Most of these promoters carry tandem
DNA sites for CRP separated by 52 base pairs. Interestingly, this arrangement is the optimal target for CytR, and hence CytR is targeted only to certain CRP-dependent promoters.

**Transcription regulation by two activators**

Many naturally occurring bacterial promoters are co-dependent on the action of two activators that make promoter activity contingent on two different physiological signals. Five different situations can account for co-dependence on two activators [22].

(i) The binding of one activator may be dependent on the binding of the other, and vice-versa (Figure 5a).
(ii) The binding of one activator may trigger the repositioning of the other, moving it from a location where it is unable to activate transcription to a location where it can activate transcription (Figure 5b).
(iii) The binding of one activator may alter the conformation of promoter DNA so that a second activator is able to make contact with RNAP (Figure 5c).
(iv) The two different activators may each make independent contacts with RNAP (Figure 5d).
(v) The second activator may counteract the action of a repressor that is interfering with the function of the first activator (Figure 5e).

Although examples of all five mechanisms have been found (see Figure 5), the independent contact mechanism (Figure 5d) appears to be the most common as it can be effected without interactions between the two different activators. In principle, this mechanism could accommodate as many activators as there are contact sites on the RNAP.

**Perspectives**

Over 5% of the genes of *E. coli* encode products concerned with transcriptional regulation and it is clear that a large variety of mechanisms are exploited. However, we can now propose a number of simple models to explain the majority of cases. These models hinge on our understanding of the structure of RNAP and of the role of the different RNAP subunits. Also, the realization that most bacterial transcription factors belong to a small number of families has greatly facilitated our understanding of different promoters. Immediate priorities for the future include the description, in structural terms, of the process of transcription initiation, elucidation of the precise mechanisms by which different activators and repressors function, and the structures at various activator–RNAP interfaces. Methods to address these priorities are now in place and progress should be rapid.
Summary

• Transcription initiation is the principal step at which bacterial gene expression is regulated. Bacterial transcription is due to a single multi-subunit RNA polymerase.
• The potential transcription initiation rate of any promoter is set by the efficiency with which RNA polymerase recognizes the different promoter sequence elements. The \( \sigma \) subunit plays the major role in the process of promoter recognition.
• Different RNA polymerase $\sigma$ subunits can guide RNA polymerase to different promoters. The E. coli genome encodes seven different $\sigma$ subunits, each of which allows the cell to respond to different environmental stimuli.

• A large number of transcription factors up-regulate and down-regulate expression from different promoters in response to environmental signals.

• Many transcription activators function by making a direct interaction with RNA polymerase. Some activators function by altering the conformation of promoter DNA.

• Most transcription repressors function by blocking access of RNA polymerase to their target promoter. In some cases, optimal repression depends on multiply bound repressor molecules that interact in complex ways.

• Many promoters are regulated by more than one transcription factor. A variety of mechanisms whereby a promoter can be regulated by a repressor and an activator, or by two activators, is known.

References


