TRANSCRIPTION

# 3.9.1 Introduction

## Key Terms

The **coding strand (Sense strand)** of DNA has the same sequence as the mRNA and is related by the genetic code to the protein sequence that it represents.

The **antisense strand (Template strand)** of DNA is complementary to the sense strand, and is the one that acts as the template for synthesis of mRNA.

**RNA polymerases** are enzymes that synthesize RNA using a DNA template (formally described as DNA-dependent RNA polymerases).

A **promoter** is a region of DNA where RNA polymerase binds to initiate transcription.

**Startpoint (startsite) (Startsite)** refers to the position on DNA corresponding to the first base incorporated into RNA.

A **terminator** is a sequence of DNA that causes RNA polymerase to terminate transcription.

A **transcription unit** is the distance between sites of initiation and termination by RNA polymerase; may include more than one gene.

**Upstream** identifies sequences proceeding in the opposite direction from expression; for example, the bacterial promoter is upstream of the transcription unit, the initiation codon is upstream of the coding region.

**Downstream** identifies sequences proceeding farther in the direction of expression; for example, the coding region is downstream of the initiation codon.

A **primary transcript** is the original unmodified RNA product corresponding to a transcription unit.

Transcription involves synthesis of an RNA chain representing one strand of a DNA duplex. By "representing" we mean that the RNA is *identical in sequence* with one strand of the DNA, which is called the **coding strand**. It is *complementary* to the other strand, which provides the **template strand** for its synthesis. **Figure 9.1** recapitulates the relationship between double-stranded DNA and its single-stranded RNA transcript.

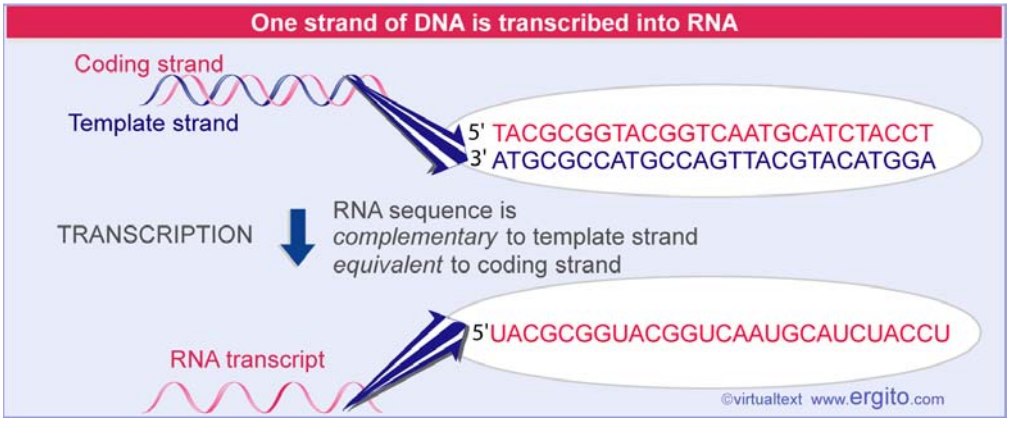


**Figure 9.1** The function of RNA polymerase is to copy one strand of duplex DNA into RNA.

RNA synthesis is catalyzed by the enzyme **RNA polymerase**. Transcription starts when RNA polymerase binds to a special region, the **promoter**, at the start of the gene. The promoter surrounds the first base pair that is transcribed into RNA, the **startpoint**. From this point, RNA polymerase moves along the template, synthesizing RNA, until it reaches a **terminator** sequence. This action defines a **transcription unit** that extends from the promoter to the terminator. The critical feature of the transcription unit, depicted in **Figure 9.2**, is that it constitutes a stretch of DNA *expressed via the production of a single RNA molecule.* A transcription unit may include more than one gene.

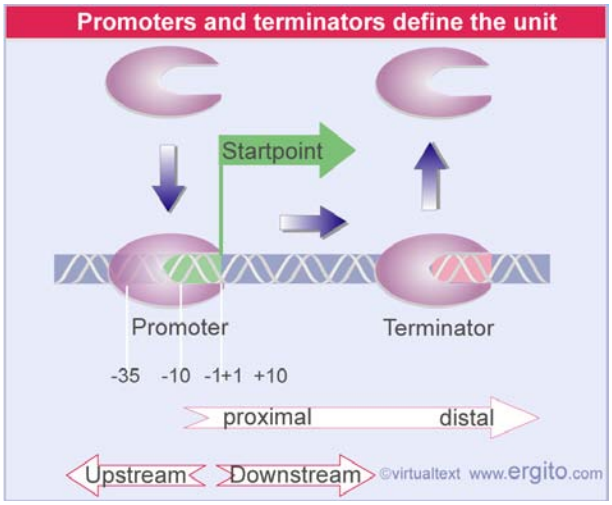


**Figure 9.2** A transcription unit is a sequence of DNA transcribed into a single RNA, starting at the promoter and ending at the terminator.

Sequences prior to the startpoint are described as **upstream** of it; those after the startpoint (within the transcribed sequence) are **downstream** of it. Sequences are conventionally written so that transcription proceeds from left (upstream) to right (downstream). This corresponds to writing the mRNA in the usual 5′ → 3′ direction.

Often the DNA sequence is written to show only the coding strand, which has the

same sequence as the RNA. Base positions are numbered in both directions away from the startpoint, which is assigned the value +1; numbers are increased going downstream. The base before the startpoint is numbered –1, and the negative numbers increase going upstream. (There is no base assigned the number 0.)

The immediate product of transcription is called the **primary transcript**. It would consist of an RNA extending from the promoter to the terminator, possessing the original 5′ and 3′ ends. However, the primary transcript is almost always unstable. In prokaryotes, it is rapidly degraded (mRNA) or cleaved to give mature products (rRNA and tRNA). In eukaryotes, it is modified at the ends (mRNA) and/or cleaved to give mature products (all RNA).

Transcription is the first stage in gene expression, and the principal step at which it is controlled. Regulatory proteins determine whether a particular gene is available to be transcribed by RNA polymerase. The initial (and often the only) step in regulation is the decision on whether or not to transcribe a gene. Most regulatory events occur at the initiation of transcription, although subsequent stages in transcription (or other stages of gene expression) are sometimes regulated.

Within this context, there are two basic questions in gene expression:

- How does RNA polymerase find promoters on DNA? This is a particular example of a more general question: how do proteins distinguish their specific binding sites in DNA from other sequences?
- How do regulatory proteins interact with RNA polymerase (and with one another) to activate or to repress specific steps in the initiation, elongation, or termination of transcription?

In this chapter, we analyze the interactions of bacterial RNA polymerase with DNA, from its initial contact with a gene, through the act of transcription, culminating in its release when the transcript has been completed. *Molecular Biology 3.10 The operon* discusses the various means by which regulatory proteins can assist or prevent bacterial RNA polymerase from recognizing a particular gene for transcription. *Molecular Biology 3.11 Regulatory circuits* discusses other means of regulation, including the use of small RNAs, and considers how these interactions can be connected into larger regulatory networks. In *Molecular Biology 3.12 Phage strategies* we consider how individual regulatory interactions can be connected into more complex networks. In *Molecular Biology 5.21 Promoters and enhancers* and *Molecular Biology 5.22 Activating transcription*, we consider the analogous reactions between eukaryotic RNA polymerases and their templates.

TRANSCRIPTION

# 3.9.2 Transcription occurs by base pairing in a "bubble" of unpaired DNA

## Key Concepts

- RNA polymerase separates the two strands of DNA in a transient "bubble" and uses one strand as a template to direct synthesis of a complementary sequence of RNA.
- The length of the bubble is ~12-14 bp, and the length of RNA-DNA hybrid within it is ~8-9 bp.

Transcription takes place by the usual process of complementary base pairing. **Figure 9.3** illustrates the general principle of transcription. RNA synthesis takes place within a "transcription bubble," in which DNA is transiently separated into its single strands, and the template strand is used to direct synthesis of the RNA strand.

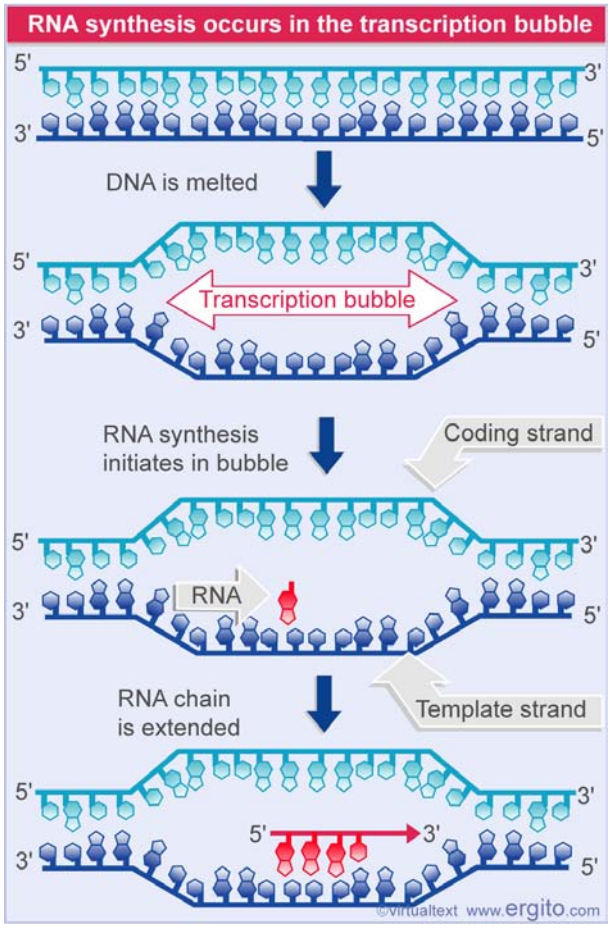


**Figure 9.3** DNA strands separate to form a transcription bubble. RNA is synthesized by complementary base pairing with one of the DNA strands.

The RNA chain is synthesized from the 5′ end toward the 3′ end. The 3′–OH group of the last nucleotide added to the chain reacts with an incoming nucleoside 5′ triphosphate. The incoming nucleotide loses its terminal two phosphate groups ( γ and β); its α group is used in the phosphodiester bond linking it to the chain. The overall reaction rate is ~40 nucleotides/second at 37°C (for the bacterial RNA polymerase); this is about the same as the rate of translation (15 amino acids/sec), but much slower than the rate of DNA replication (800 bp/sec).

RNA polymerase creates the transcription bubble when it binds to a promoter. **Figure 9.4** shows that as RNA polymerase moves along the DNA, the bubble moves with it, and the RNA chain grows longer. The process of base pairing and base addition within the bubble is catalyzed and scrutinized by the enzyme.

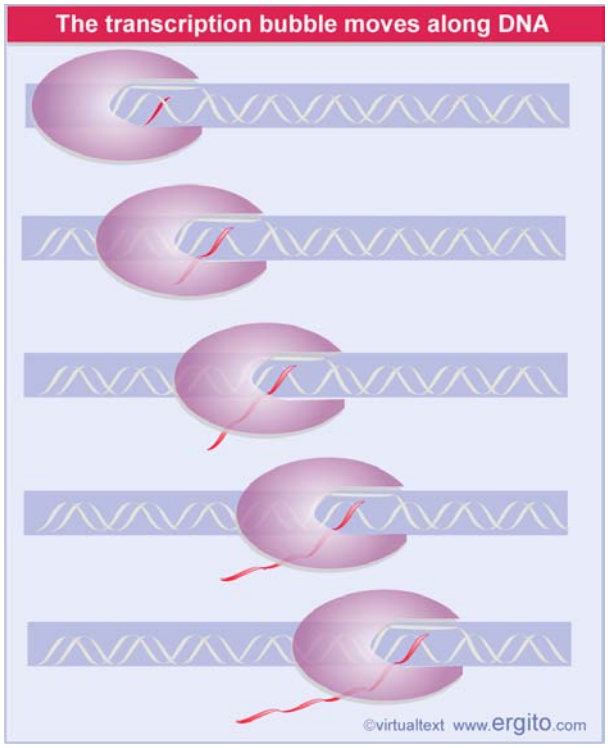


**Figure 9.4** Transcription takes place in a bubble, in which RNA is synthesized by base pairing with one strand of DNA in the transiently unwound region. As the bubble progresses, the DNA duplex reforms behind it, displacing the RNA in the form of a single polynucleotide chain.

The structure of the bubble within RNA polymerase is shown in the expanded view of **Figure 9.5**. As RNA polymerase moves along the DNA template, it unwinds the duplex at the front of the bubble (the unwinding point), and rewinds the DNA at the back (the rewinding point). The length of the transcription bubble is ~12-14 bp, but the length of the RNA-DNA hybrid region within it is shorter (1903; for review see 68).

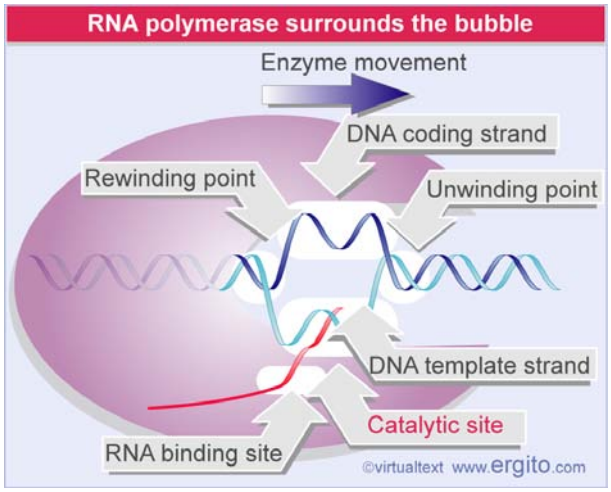


**Figure 9.5** During transcription, the bubble is maintained within bacterial RNA polymerase, which unwinds and rewinds DNA, maintains the conditions of the partner and template DNA strands, and synthesizes RNA.

There is a major change in the topology of DNA extending over ~1 turn, but it is not clear how much of this region is actually base paired with RNA at any given moment. Certainly the RNA-DNA hybrid is short and transient. As the enzyme moves on, the DNA duplex reforms, and the RNA is displaced as a free polynucleotide chain. About the last 25 ribonucleotides added to a growing chain are complexed with DNA and/or enzyme at any moment.

*Last updated on November 17, 2003*

## Reviews

68. Losick, R. and Chamberlin, M. (1976). *RNA Polymerase*. Cold Spring Harbor Symp. Quant. Biol..

## References

1903. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. A. (2000). *A structural model of transcription elongation.* Science 289, 619-625.

TRANSCRIPTION

# 3.9.3 The transcription reaction has three stages

## Key Terms

**Initiation** describes the stages of transcription up to synthesis of the first bond in RNA. This includes binding of RNA polymerase to the promoter and melting a short region of DNA into single strands.

**Elongation** is the stage in a macromolecular synthesis reaction (replication, transcription, or translation) when the nucleotide or polypeptide chain is being extended by the addition of individual subunits.

**Termination** is a separate reaction that ends a macromolecular synthesis reaction (replication, transcription, or translation), by stopping the addition of subunits, and (typically) causing disassembly of the synthetic apparatus.

## Key Concepts

- RNA polymerase initiates transcription after binding to a promoter site on DNA.
- During elongation the transcription bubble moves along DNA and the RNA chain is extended in the 5′–3′ direction.
- Transcription stops, the DNA duplex reforms and RNA polymerase dissociates at a terminator site.

The transcription reaction can be divided into the stages illustrated in **Figure 9.6**, in which a bubble is created, RNA synthesis begins, the bubble moves along the DNA, and finally is terminated:

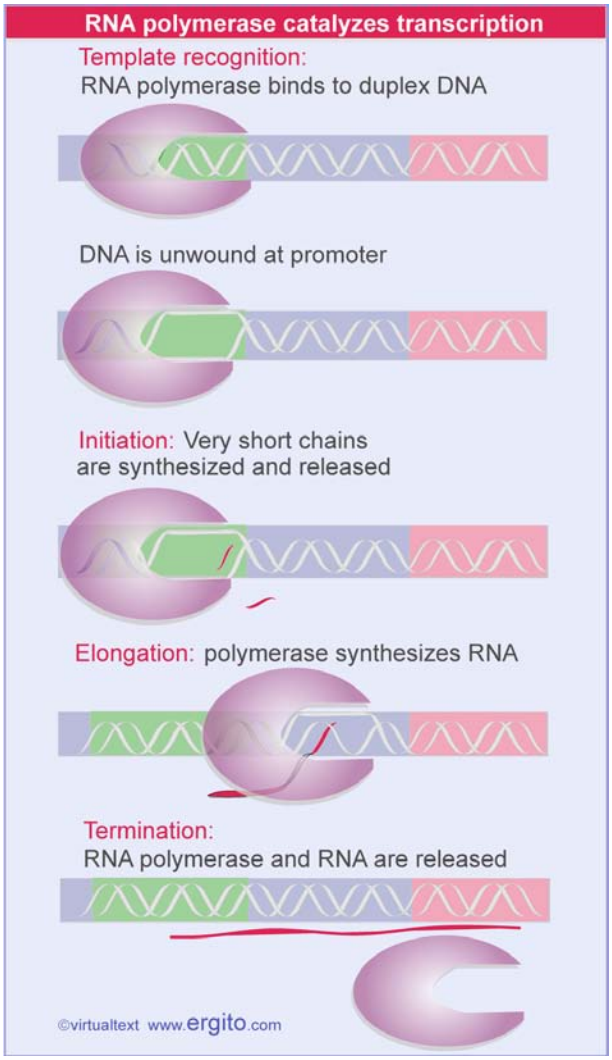


**Figure 9.6** Transcription has four stages, which involve different types of interaction between RNA polymerase and DNA. The enzyme binds to the promoter and melts DNA, remains stationary during initiation, moves along the template during elongation, and dissociates at termination.

- *Template recognition* begins with the binding of RNA polymerase to the double-stranded DNA at a promoter to form a "closed complex". Then the strands of DNA are separated to form the "open complex" that makes the template strand available for base pairing with ribonucleotides. The transcription bubble is created by a local unwinding that begins at the site bound by RNA polymerase.

- **Initiation** describes the synthesis of the first nucleotide bonds in RNA. The enzyme remains at the promoter while it synthesizes the first ~9 nucleotide bonds. The initiation phase is protracted by the occurrence of abortive events, in which the enzyme makes short transcripts, releases them, and then starts synthesis of RNA again. The initiation phase ends when the enzyme succeeds in extending the chain and clears the promoter. *The sequence of DNA needed for*

*RNA polymerase to bind to the template and accomplish the initiation reaction defines the promoter*. Abortive initiation probably involves synthesizing an RNA chain that fills the active site. If the RNA is released, the initiation is aborted and must start again. Initiation is accomplished if and when the enzyme manages to move along the template to move the next region of the DNA into the active site.

- During **elongation** the enzyme moves along the DNA and extends the growing RNA chain. As the enzyme moves, it unwinds the DNA helix to expose a new segment of the template in single-stranded condition. Nucleotides are covalently added to the 3′ end of the growing RNA chain, forming an RNA-DNA hybrid in the unwound region. Behind the unwound region, the DNA template strand pairs with its original partner to reform the double helix. The RNA emerges as a free single strand. *Elongation involves the movement of the transcription bubble by a disruption of DNA structure, in which the template strand of the transiently unwound region is paired with the nascent RNA at the growing point*.

- **Termination** involves recognition of the point at which no further bases should be added to the chain. To terminate transcription, the formation of phosphodiester bonds must cease, and the transcription complex must come apart. When the last base is added to the RNA chain, the transcription bubble collapses as the RNA-DNA hybrid is disrupted, the DNA reforms in duplex state, and the enzyme and RNA are both released. *The sequence of DNA required for these reactions defines the terminator*.

The traditional view of elongation has been that it is a monotonic process, in which the enzyme moves forward 1 bp along DNA for every nucleotide added to the RNA chain. Changes in this pattern occur in certain circumstances, in particular when RNA polymerase pauses. One type of pattern is for the "front end" of the enzyme to remain stationary while the "back end" continues to move, thus compressing the footprint on DNA. After movement of several base pairs, the "front end" is released, restoring a footprint of full length. This gave rise to the "inchworm" model of transcription, in which the enzyme proceeds discontinuously, alternatively compressing and releasing the footprint on DNA. However, it may be the case that these events describe an aberrant situation rather than normal transcription (502; 504).

## References

502. Rice, G. A., Kane, C. M., and Chamberlin, M. (1991). *Footprinting analysis of mammalian RNA polymerase II along its transcript: an alternative view of transcription elongation*. Proc. Natl. Acad. Sci. USA 88, 4245-281.

504. Wang, D. et al. (1995). *Discontinuous movements of DNA and RNA in RNA polymerase accompany formation of a paused transcription complex*. Cell 81, 341-350.

TRANSCRIPTION

# 3.9.4 Phage T7 RNA polymerase is a useful model system

## Key Concepts

- T3 and T7 phage RNA polymerases are single polypeptides with minimal activities in recognizing a small number of phage promoters.
- Crystal structures of T7 RNA polymerase with DNA identify the DNA-binding region and the active site.

The existence of very small RNA polymerases, comprising single polypeptide chains coded by certain phages, gives some idea of the "minimum" apparatus necessary for transcription. These RNA polymerases recognize just a few promoters on the phage DNA; and they have no ability to change the set of promoters to which they respond. They provide simple model systems for characterizing the binding of RNA polymerase to DNA and the initiation reaction.

The RNA polymerases coded by the related phages T3 and T7 are single polypeptide chains of <100 kD each. They synthesize RNA at rates of ~200 nucleotides/second at 37°C, more rapidly than bacterial RNA polymerase.

The T7 RNA polymerase is homologous to DNA polymerases, and has a similar structure, in which DNA lies in a "palm" surrounded by "fingers" and a "thumb" (see **Figure 14.7**). We now have a direct view of the active site from a crystal structure of a phage T7 RNA polymerase engaged in transcription (925).

The T7 RNA polymerase recognizes its target sequence in DNA by binding to bases in the major groove at a position upstream from the startpoint, as shown in **Figure 9.7**. The enzyme uses a *specificity loop* that is formed by a β ribbon (1565). This feature is unique to the RNA polymerase (it is not found in DNA polymerases). The common point with all RNA polymerases is that the enzyme recognizes specific bases in DNA that are upstream of the sequence that is transcribed.

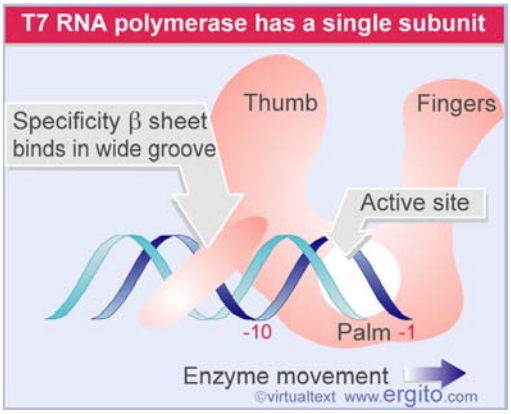


**Figure 9.7** T7 RNA polymerase has a specificity loop that binds positions -7 to -11 of the promoter while positions -1 to -4 enter the active site.

When transcription initiates, the conformation of the enzyme remains essentially the same while several nucleotides are added, and the transcribed template strand is "scrunched" in the active site. The active site can hold a transcript of 6-9 nucleotides. The transition from initiation to elongation is defined as the point when the enzyme begins to move along DNA. This occurs when the nascent transcript extends beyond the active site and interacts with the specificity loop (1566). The RNA emerges to the surface of the enzyme when 12-14 nucleotides have been synthesized. These features are similar to those displayed by bacterial RNA polymerase.

*Last updated on 3-7-2001*

## References

925. Cheetham, G. M. T. and Steitz, T. A. (1999). *Structure of a transcribing T7 RNA polymerase initiation complex*. Science 286, 2305-2309.

1565. Cheetham, G. M., Jeruzalmi, D., and Steitz, T. A. (1999). *Structural basis for initiation of transcription from an RNA polymerase-promoter complex.* Nature 399, 80-83.

1566. Temiakov, D., Mentesana, D., Temiakov, D., Ma, K., Mustaev, A., Borukhov, S., and McAllister, W. T. (2000). *The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance.* Proc. Natl. Acad. Sci. USA 97, 14109-14114.

TRANSCRIPTION

# 3.9.5 A model for enzyme movement is suggested by the crystal structure

## Key Concepts

- DNA moves through a groove in yeast RNA polymerase that makes a sharp turn at the active site.
- A protein bridge changes conformation to control the entry of nucleotides to the active site.

We now have much information about the structure and function of RNA polymerase as the result of the crystal structures of the bacterial and yeast enzymes. Bacterial RNA polymerase has overall dimensions of ~90 × 95 × 160 Å. Eukaryotic RNA polymerase is larger but less elongated. Structural analysis shows that they share a common type of structure, in which there is a "channel" or groove on the surface ~25 Å wide that could be the path for DNA. This is illustrated in **Figure 9.8** for the example of bacterial RNA polymerase. The length of the groove could hold 16 bp in the bacterial enzyme, and ~25 bp in the eukaryotic enzyme, but this represents only part of the total length of DNA bound during transcription. The enzyme surface is largely negatively charged, but the groove is lined with positive charges, enabling it to interact with the negatively charged phosphate groups of DNA.

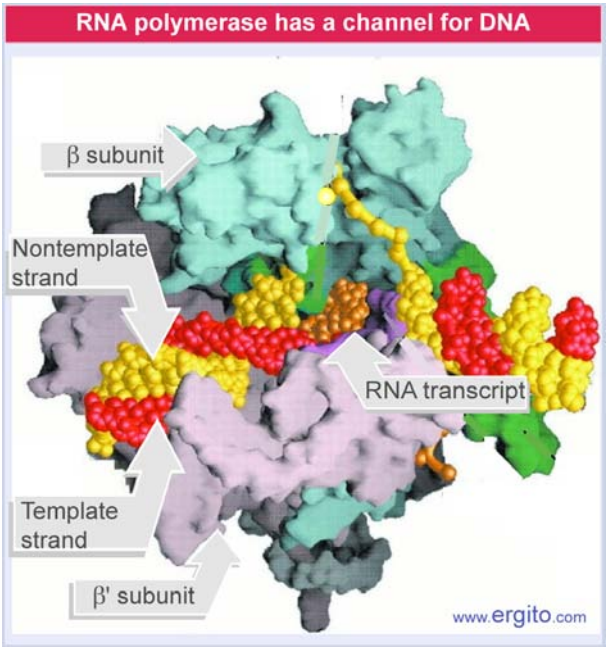


**Figure 9.8** The β (cyan) and β ′ subunit (pink) of RNA polymerase have a channel for the DNA template. Synthesis of an RNA transcript (copper) has just begun; the DNA template (red) and coding (yellow) strands are separated in a transcription bubble. Photograph kindly provided by Seth Darst.

The yeast enzyme is a large structure with 12 subunits (see *Molecular Biology 5.21.2 Eukaryotic RNA polymerases consist of many subunits*). Ten subunits of the yeast RNA polymerase II have been located on the crystal structure, as shown in **Figure 9.9**. The catalytic site is formed by a cleft between the two large subunits (#1 and #2), which grasp DNA downstream in "jaws" as it enters the RNA polymerase. Subunits 4 and 7 are missing from this structure; they form a subcomplex that dissociates from the complete enzyme. The structure is generally similar to that of bacterial RNA polymerase (1714; 1913; 1914; for review see 4528). This can be seen more clearly in the crystal structure of **Figure 9.10**. RNA polymerase surrounds the DNA, as seen in the view of **Figure 9.11**. A catalytic $Mg^{2+}$ ion is found at the active site. The DNA is clamped in position at the active site by subunits 1, 2, and 6. **Figure 9.12** shows that DNA is forced to take a turn at the entrance to the site, because of an adjacent wall of protein. The length of the RNA hybrid is limited by another protein obstruction, called the rudder. Nucleotides probably enter the active site from below, via pores through the structure.

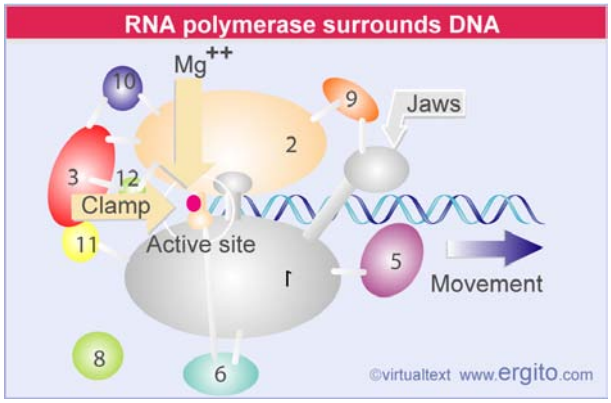


**Figure 9.9** Ten subunits of RNA polymerase are placed in position from the crystal structure. The colors of the subunits are the same as in the crystal structures of the following figures.

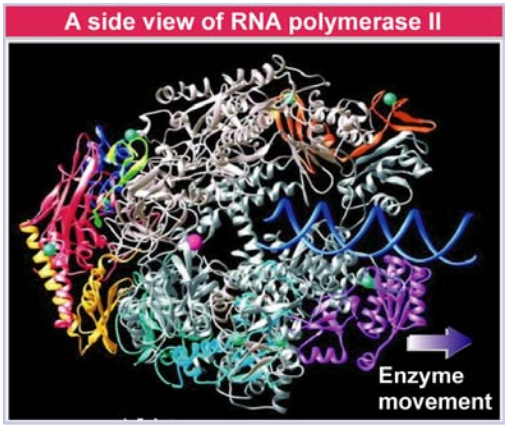


**Figure 9.10** The side view of the crystal structure of RNA polymerase II from yeast shows that DNA is held downstream by a pair of jaws and is clamped in position in the active site, which contains an $Mg^{++}$ ion. Photograph kindly provided by Roger Kornberg (see 1714).

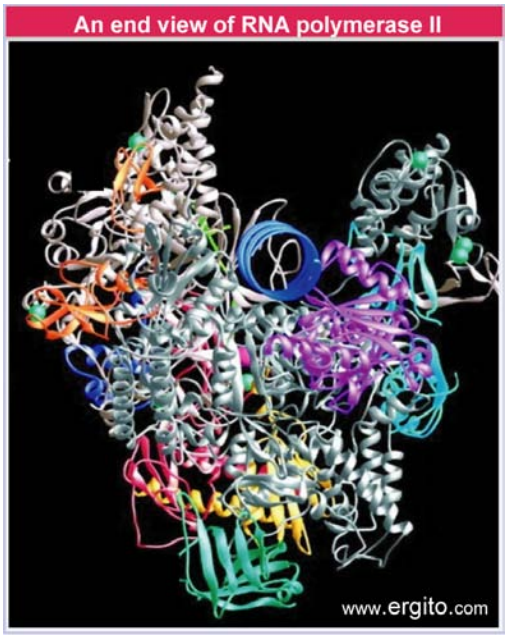


**Figure 9.11** The end view of the crystal structure of RNA polymerase II from yeast shows that DNA is surrounded by ~270° of protein. Photograph kindly provided by Roger Kornberg (see 1714).

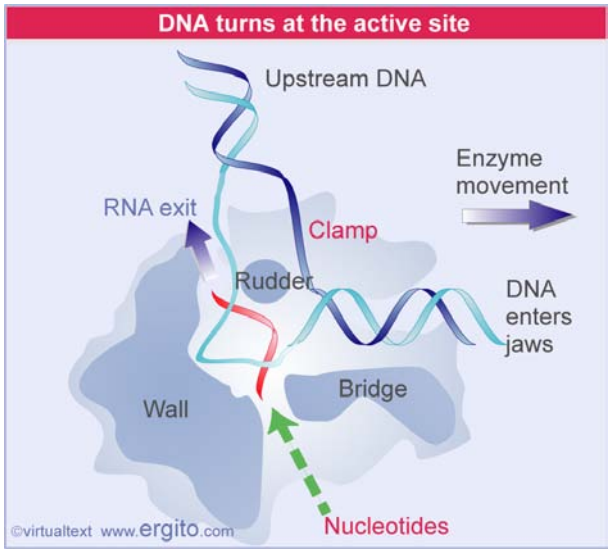


**Figure 9.12** DNA is forced to make a turn at the active site by a wall of protein. Nucleotides may enter the active site through a pore in the protein (see 1913).

The expanded view of the active site in **Figure 9.13** shows that the transcription bubble includes 9 bp of DNA-RNA hybrid. Where the DNA takes its turn, the bases downstream are flipped out of the DNA helix. As the enzyme moves along DNA, the base in the template strand at the start of the turn will be flipped to face the nucleotide entry site. The ′ end of the RNA is forced to leave the DNA when it hits the protein rudder (see **Figure 9.12**).

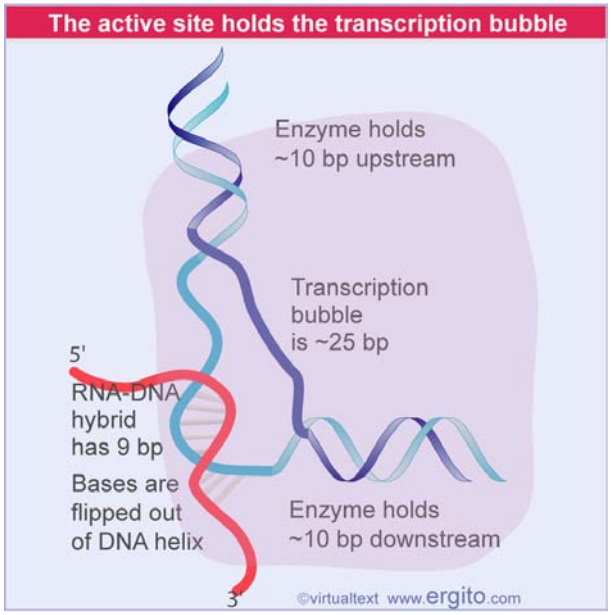


**Figure 9.13** An expanded view of the active site shows the sharp turn in the path of DNA.

Once DNA has been melted, the individual strands have a flexible structure in the transcription bubble. This enables DNA to take its turn in the active site. But before transcription starts, the DNA double helix is a relatively rigid straight structure. How does this structure enter the polymerase without being blocked by the wall? The answer is that a large conformational shift must occur in the enzyme. Adjacent to the wall is a clamp. In the free form of RNA polymerase, this clamp swings away from the wall to allow DNA to follow a straight path through the enzyme. After DNA has been melted to create the transcription bubble, the clamp must swing back into position against the wall.

One of the dilemmas of any nucleic acid polymerase is that the enzyme must make tight contacts with the nucleic acid substrate and product, but must break these contacts and remake them with each cycle of nucleotide addition. Consider the situation illustrated in **Figure 9.14**. A polymerase makes a series of specific contacts with the bases at particular positions. For example, contact "1" is made with the base at the end of the growing chain, and contact "2" is made with the base in the template strand that is complementary to the next base to be added. But the bases that occupy these locations in the nucleic acid chains change every time a nucleotide is added!

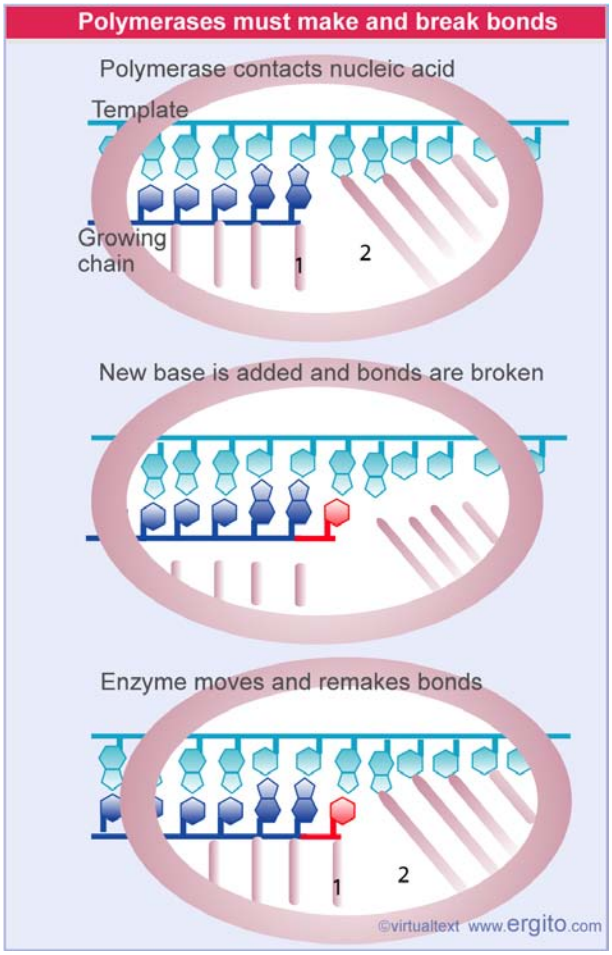


**Figure 9.14** Movement of a nucleic acid polymerase requires breaking and remaking bonds to the nucleotides that occupy fixed positions relative to the enzyme structure. The nucleotides in these positions change each time the enzyme moves a base along the template.

The top and bottom panels of the figure show the same situation: a base is about to be added to the growing chain. The difference is that the growing chain has been extended by one base in the bottom panel. The geometry of both complexes is exactly the same, but contacts "1" and "2" in the bottom panel are made to bases in the nucleic acid chains that are located one position farther along the chain. The middle panel shows that this must mean that, after the base is added, and before the enzyme moves relative to the nucleic acid, the contacts made to specific positions must be broken so that they can be remade to bases that occupy those positions after the movement.

The RNA polymerase structure suggests an insight into how the enzyme retains contact with its substrate while breaking and remaking bonds (1914). A structure in the protein called the bridge is adjacent to the active site (see **Figure 9.12**). This feature is found in both the bacterial and yeast enzymes, but it has different shapes in the different crystal structures. In one it is bent, and in the other it is straight. **Figure 9.15** suggests that the change in conformation of the bridge structure is closely related to translocation of the enzyme along the nucleic acid.

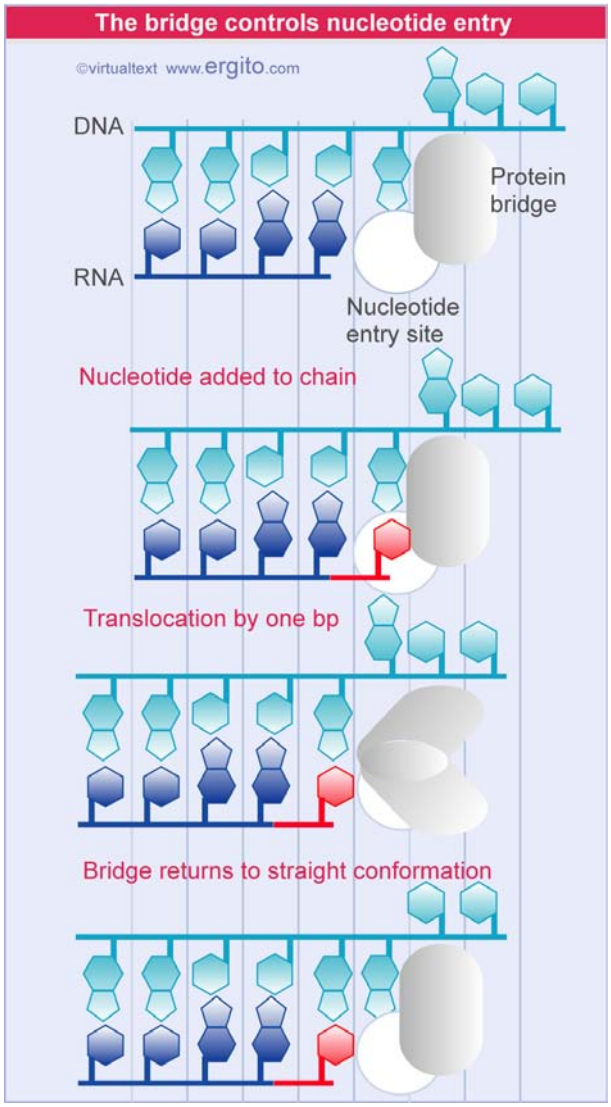


**Figure 9.15** The RNA polymerase elongation cycle starts with a straight bridge adjacent to the nucleotide entry site. After nucleotide addition, the enzyme moves one base pair and bridge bends as it retains contact with the newly added nucleotide. When the bridge is released, the cycle can start again (see 1914).

At the start of the cycle of translocation, the bridge has a straight conformation adjacent to the nucleotide entry site. This allows the next nucleotide to bind at the nucleotide entry site. The bridge is in contact with the newly added nucleotide. Then the protein moves one base pair along the substrate. The bridge changes its conformation, bending to keep contact with the newly added nucleotide. In this conformation, the bridge obscures the nucleotide entry site. To end the cycle, the bridge returns to its straight conformation, allowing access again to the nucleotide entry site. The bridge acts as a ratchet that releases the DNA and RNA strands for translocation while holding on to the end of the growing chain.

*Last updated on 10-3-2003*

## Reviews

4528. Shilatifard, A., Conaway, R. C., and Conaway, J. W. (2003). *The RNA polymerase II elongation complex.* Annu. Rev. Biochem. 72, 693-715.

## References

1714. Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R., and Kornberg, R. D. (2000). *Architecture of RNA polymerase II and implications for the transcription mechanism.* Science 288, 640-649.

1913. Cramer, P., Bushnell, P., and Kornberg, R. D. (2001). *Structural basis of transcription: RNA polymerase II at 2.8* Å *resolution.* Science 292, 1863-1876.

1914. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. (2001). *Structural basis of transcription: an RNA polymerase II elongation complex at 3.3* Å *resolution.* Science 292, 1876-1882.

TRANSCRIPTION

# 3.9.6 Bacterial RNA polymerase consists of multiple subunits

## Key Terms

The **holoenzyme (complete enzyme)** is the complex of five subunits including core enzyme ( $\alpha_2 \beta \beta'$ and $\sigma$) factor that is competent to initiate bacterial transcription.

## Key Concepts

- Bacterial RNA core polymerases are ~500 kD multisubunit complexes with the general structure $\alpha_2 \beta \beta'$.
- DNA is bound in a channel and is contacted by both the $\beta$ and $\beta'$ subunits.

The best characterized RNA polymerases are those of eubacteria, for which *E. coli* is a typical case. *A single type of RNA polymerase appears to be responsible for almost all synthesis of mRNA, and all rRNA and tRNA, in a eubacterium.* About 7000 RNA polymerase molecules are present in an *E. coli* cell. Many of them are engaged in transcription; probably 2000-5000 enzymes are synthesizing RNA at any one time, the number depending on the growth conditions.

The **complete enzyme** or **holoenzyme** in *E. coli* has a molecular weight of ~465 kD. Its subunit composition is summarized in **Figure 9.16**.

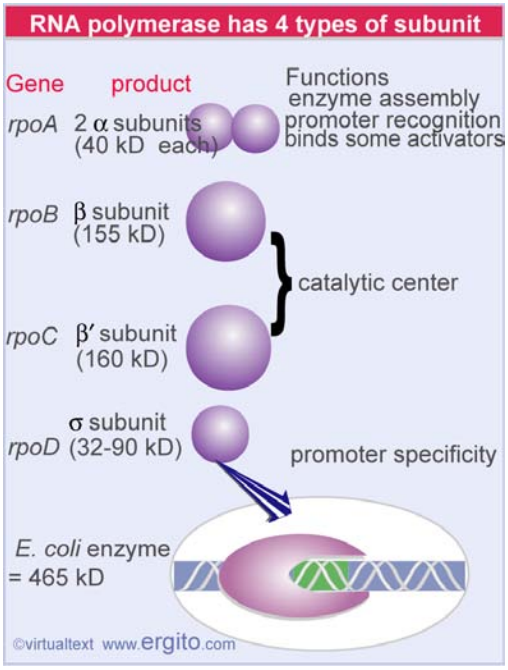


**Figure 9.16** Eubacterial RNA polymerases have four types of subunit; α , β , and β have rather constant sizes in different bacterial species, but σ varies more widely.

The β and β ′ subunits together make up the catalytic center. Their sequences are related to those of the largest subunits of eukaryotic RNA polymerases (see *Molecular Biology 5.21.2 Eukaryotic RNA polymerases consist of many subunits*), suggesting that there are common features to the actions of all RNA polymerases. The β subunit can be crosslinked to the template DNA, the product RNA, and the substrate ribonucleotides; mutations in *rpoB* affect all stages of transcription. Mutations in *rpoC* show that β ′ also is involved at all stages.

The α subunit is required for assembly of the core enzyme. When phage T4 infects *E. coli*, the α subunit is modified by ADP-ribosylation of an arginine. The modification is associated with a reduced affinity for the promoters formerly recognized by the holoenzyme, suggesting that the α subunit plays a role in promoter recognition. The α subunit also plays a role in the interaction of RNA polymerase with some regulatory factors.

The σ subunit is concerned specifically with promoter recognition, and we have more information about its functions than on any other subunit (see *Molecular Biology 3.9.7 RNA polymerase consists of the core enzyme and sigma factor*).

The crystal structure of the bacterial enzyme (**Figure 9.8**) shows that the channel for DNA lies at the interface of the β and β ′ subunits (1880). (The α subunits are not visible in this view.) The DNA is unwound at the active site, where an RNA chain is being synthesized. Crosslinking experiments identify the points at which the RNA polymerase subunits contact DNA (1903). These are summarized in **Figure 9.17**. The β and β ′ subunits contact DNA at many points downstream of the active site. They make several contacts with the coding strand in the region of the transcription

bubble, thus stabilizing the separated single strands. The RNA is contacted largely in the region of the transcription bubble.

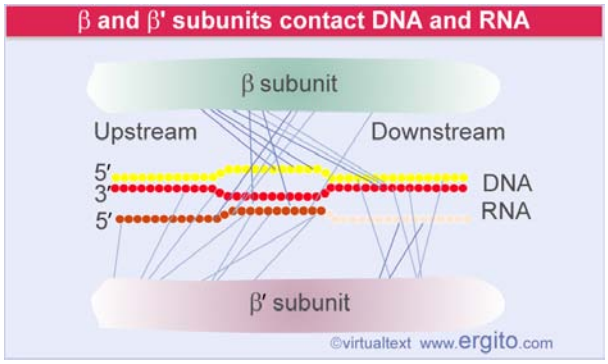


**Figure 9.17** Both the template and coding strands of DNA are contacted by the β and β ′ subunits largely in the region of the transcription bubble and downstream. The RNA is contacted mostly in the transcription bubble. (Usually there is no downstream RNA, and contacts with RNA downstream occur only in the special case when the enzyme backtracks.) (Based on data of 1903.)

The drug rifampicin (a member of the rifamycin antibiotic family) blocks transcription by bacterial RNA polymerase. It is a major drug used against tuberculosis. The crystal structure of RNA polymerase bound to rifampicin explains its action: it binds in a pocket of the β subunit, >12Å away from the active site, but in a position where it blocks the path of the elongating RNA (1904). By preventing the RNA chain from extending beyond 2-3 nucleotides, this blocks transcription.

Originally defined simply by its ability to incorporate nucleotides into RNA under the direction of a DNA template, the enzyme RNA polymerase now is seen as part of a more complex apparatus involved in transcription. *The ability to catalyze RNA synthesis defines the minimum component that can be described as RNA polymerase.* It supervises the base pairing of the substrate ribonucleotides with DNA and catalyzes the formation of phosphodiester bonds between them.

All of the subunits of the basic polymerase that participate in elongation are necessary for initiation and termination. But transcription units differ in their dependence on additional polypeptides at the initiation and termination stages. Some of these additional polypeptides are needed at all genes, but others may be needed specifically for initiation or termination at particular genes. The analogy with the division of labors between the ribosome and the protein synthesis factors is obvious.

*E. coli* RNA polymerase can transcribe any one of many (>1000) transcription units. The enzyme therefore requires the ability to interact with a variety of host and phage functions that modify its intrinsic transcriptional activities. The complexity of the enzyme therefore at least in part reflects its need to interact with regulatory factors, rather than any demand inherent in its catalytic activity (for review see 76).

*Last updated on 6-1-2001*

## Reviews

76. Helmann, J. D. and Chamberlin, M. (1988). *Structure and function of bacterial sigma factors.* Annu. Rev. Biochem. 57, 839-872.

## References

1880. Zhang, G., Campbell, E. A., Zhang, E. A., Minakhin, L., Richter, C., Severinov, K., and Darst, S. A. (1999). *Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 Å resolution.* Cell 98, 811-824.

1903. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S. A. (2000). *A structural model of transcription elongation.* Science 289, 619-625.

1904. Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S. A. (2001). *Structural mechanism for rifampicin inhibition of bacterial RNA polymerase.* Cell 104, 901-912.

TRANSCRIPTION

# 3.9.7 RNA polymerase consists of the core enzyme and sigma factor

## Key Terms

The **core enzyme** is the complex of RNA polymerase subunits that undertakes elongation. It does not include additional subunits or factors that may needed for initiation or termination.

**Sigma factor** is the subunit of bacterial RNA polymerase needed for initiation; it is the major influence on selection of promoters.

A **loose binding site** is any random sequence of DNA that is bound by the core RNA polymerase when it is not engaged in transcription.

## Key Concepts

- Bacterial RNA polymerase can be divided into the $\alpha_2\ \beta\ \beta'$ core enzyme that catalyzes transcription and the sigma subunit that is required only for initiation.
- Sigma factor changes the DNA-binding properties of RNA polymerase so that its affinity for general DNA is reduced and its affinity for promoters is increased.
- Binding constants of RNA polymerase for different promoters vary over 6 orders of magnitude, corresponding to the frequency with which transcription is initiated at each promoter.

The holoenzyme ( $\alpha_2\ \beta\ \beta'\ \sigma$) can be separated into two components, the **core enzyme** ( $\alpha_2\ \beta\ \beta'$ ) and the **sigma factor** (the $\sigma$ polypeptide). *Only the holoenzyme can initiate transcription. Sigma factor ensures that bacterial RNA polymerase binds in a stable manner to DNA only at promoters.* The sigma "factor" is usually released when the RNA chain reaches 8-9 bases, leaving the core enzyme to undertake elongation. *Core enzyme has the ability to synthesize RNA on a DNA template, but cannot initiate transcription at the proper sites* (496).

The core enzyme has a general affinity for DNA, in which electrostatic attraction between the basic protein and the acidic nucleic acid plays a major role. Any (random) sequence of DNA that is bound by core polymerase in this general binding reaction is described as a **loose binding site**. No change occurs in the DNA, which remains duplex. The complex at such a site is stable, with a half-life for dissociation of the enzyme from DNA ~60 minutes. *Core enzyme does not distinguish between promoters and other sequences of DNA.*

**Figure 9.18** shows that sigma factor introduces a major change in the affinity of RNA polymerase for DNA. *The holoenzyme has a drastically reduced ability to recognize loose binding sites* – that is, to bind to any general sequence of DNA. The association constant for the reaction is reduced by a factor of $\sim 10^4$, and the half-life of the complex is <1 second. So sigma factor destabilizes the general binding ability very considerably.

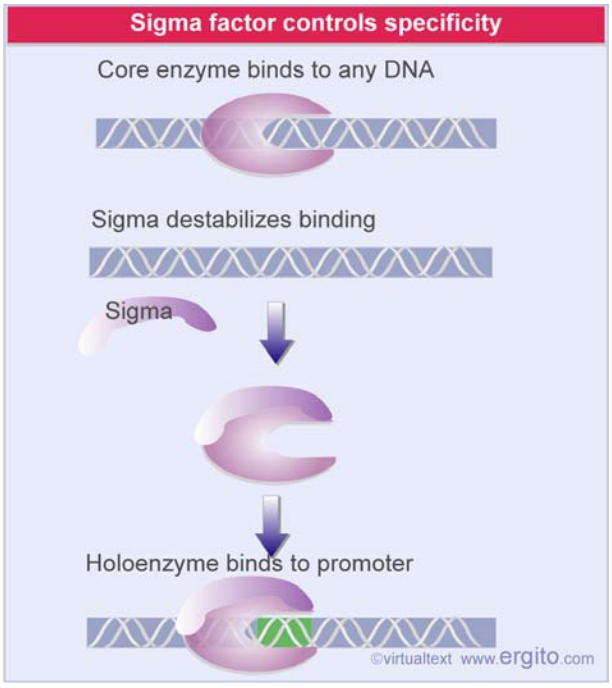


**Figure 9.18** Core enzyme binds indiscriminately to any DNA. Sigma factor reduces the affinity for sequence-independent binding, and confers specificity for promoters.

But sigma factor also *confers the ability to recognize specific binding sites.* The holoenzyme binds to promoters very tightly, with an association constant increased from that of core enzyme by (on average) 1000 times and with a half-life of several hours.

The specificity of holoenzyme for promoters compared to other sequences is ~$10^7$, but this is only an average, because there is wide variation in the rate at which the holoenzyme binds to different promoter sequences. This is an important parameter in determining the efficiency of an individual promoter in initiating transcription. The binding constants range from ~$10^{12}$ to ~$10^6$. Other factors also affect the frequency of initiation, which varies from ~1/sec (rRNA genes) to ~1/30 min (the *lacI* promoter).

## References

496. Travers, A. A. and Burgess, R. R. (1969). *Cyclic reuse of the RNA polymerase sigma factor*. Nature 222, 537-540.

TRANSCRIPTION

# 3.9.8 The association with sigma factor changes at initiation

## Key Terms

An **open complex** describes the stage of initiation of transcription when RNA polymerase causes the two strands of DNA to separate to form the "transcription bubble".

**Tight binding** of RNA polymerase to DNA describes the formation of an open complex (when the strands of DNA have separated).

The **ternary complex** in initiation of transcription consists of RNA polymerase and DNA and a dinucleotide that represents the first two bases in the RNA product.

**Abortive initiation** describes a process in which RNA polymerase starts transcription but terminates before it has left the promoter. It then reinitiates. Several cycles may occur before the elongation stage begins.

## Key Concepts

- When RNA polymerase binds to a promoter, it separates the DNA strands to form a transcription bubble and incorporates up to 9 nucleotides into RNA.
- There may be a cycle of abortive initiations before the enzyme moves to the next phase.
- Sigma factor may be released from RNA polymerase when the nascent RNA chain reaches 8-9 bases in length.

We can now describe the stages of transcription in terms of the interactions between different forms of RNA polymerase and the DNA template. The initiation reaction can be described by the parameters that are summarized in **Figure 9.19**:

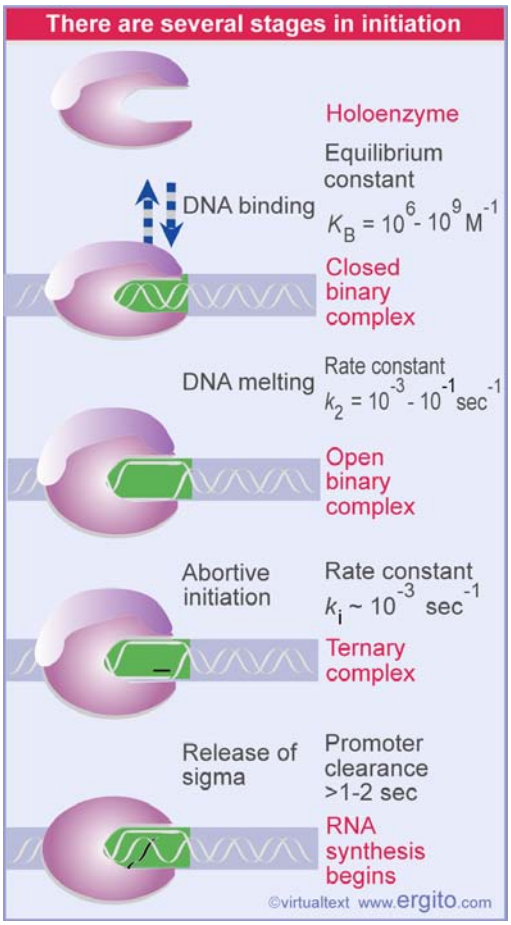


**Figure 9.19** RNA polymerase passes through several steps prior to elongation. A closed binary complex is converted to an open form and then into a ternary complex.

- The holoenzyme·promoter reaction starts by forming a closed binary complex. "Closed" means that the DNA remains duplex. Because the formation of the closed binary complex is reversible, it is usually described by an equilibrium constant ($K_B$). There is a wide range in values of the equilibrium constant for forming the closed complex.

- The closed complex is converted into an **open complex** by "melting" of a short region of DNA within the sequence bound by the enzyme. The series of events leading to formation of an open complex is called **tight binding**. For strong promoters, conversion into an open binary complex is irreversible, so this reaction is described by a rate constant ($k_2$). This reaction is fast. Sigma factor is involved in the melting reaction (see *Molecular Biology 3.9.16 Substitution of sigma factors may control initiation* ).

- The next step is to incorporate the first two nucleotides; then a phosphodiester bond forms between them. This generates a **ternary complex** that contains RNA as well as DNA and enzyme. Formation of the ternary complex is described by the rate constant $k_i$; this is even faster than the rate constant $k_2$. Further

nucleotides can be added without any enzyme movement to generate an RNA chain of up to 9 bases. After each base is added, there is a certain probability that the enzyme will release the chain. This comprises an **abortive initiation**, after which the enzyme begins again with the first base. A cycle of abortive initiations usually occurs to generate a series of very short oligonucleotides.

- When initiation succeeds, sigma is no longer necessary, and the enzyme makes the transition to the elongation ternary complex of core polymerase·DNA·nascent RNA. The critical parameter here is *how long it takes for the polymerase to leave the promoter so another polymerase can initiate.* This parameter is the promoter clearance time; its minimum value of 1-2 sec establishes the maximum frequency of initiation as <1 event per second. The enzyme then moves along the template, and the RNA chain extends beyond 10 bases.

When RNA polymerase binds to DNA, the elongated dimension of the protein extends along the DNA, but some interesting changes in shape occur during transcription. Transitions in shape and size identify three forms of the complex, as illustrated in **Figure 9.20** (501):

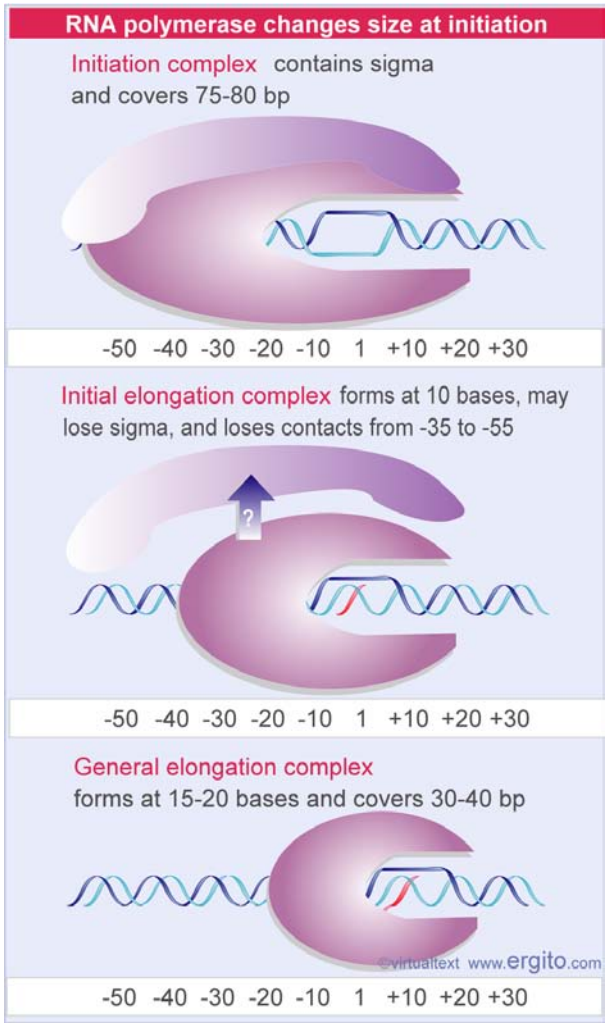


**Figure 9.20** RNA polymerase initially contacts the region from -55 to +20. When sigma dissociates, the core enzyme contracts to -30; when the enzyme moves a few base pairs, it becomes more compactly organized into the general elongation complex.

- When RNA polymerase holoenzyme initially binds to DNA, it covers some 75-80 bp, extending from –55 to +20. (The long dimension of RNA polymerase (160 Å) could cover ~50 bp of DNA in extended form, which implies that binding of a longer stretch of DNA must involve some bending of the nucleic acid.)

- The shape of the RNA polymerase changes at the transition from initiation to elongation. This is associated with the loss of contacts in the –55 to –35 region, leaving only ~60 bp of DNA covered by the enzyme. This corresponds with the concept that the more upstream part of the promoter is involved in initial recognition by RNA polymerase, but is not required for the later stages of initiation (*Molecular Biology 3.9.13 Promoter efficiencies can be increased or decreased by mutation*).

- When the RNA chain extends to 15-20 bases, the enzyme makes a further transition, to form the complex that undertakes elongation; now it covers 30-40 bp (depending on the stage in the elongation cycle).

It has been a tenet of transcription since soon after the discovery of sigma factor that it is released after initiation. However, this may not be strictly true. Direct measurements of elongating RNA polymerase complexes show that ~70% of them retain sigma factor (2003; 2004). Since a third of elongating polymerases lack sigma, the original conclusion is certainly correct that it is not necessary for elongation. In those cases where it remains associated with core enzyme, the nature of the association has almost certainly changed (see *Molecular Biology 3.9.11 Sigma factor controls binding to DNA*).

*Last updated on 9-18-2001*

## References

501. Krummel, B. and Chamberlin, M. J. (1989). *RNA chain initiation by E. coli RNA polymerase. Structural transitions of the enzyme in early ternary complexes.* Biochemistry 28, 7829-7842.

2003. Bar-Nahum, G. and Nudler, E. (2001). *Isolation and characterization of sigma(70)-retaining transcription elongation complexes from E. coli.* Cell 106, 443-451.

2004. Mukhopadhyay, J., Kapanidis, A. N., Mekler, V., Kortkhonjia, E., Ebright, Y. W., and Ebright, R. H. (2001). *Translocation of sigma(70) with RNA Polymerase during Transcription. Fluorescence Resonance Energy Transfer Assay for Movement Relative to DNA.* Cell 106, 453-463.

TRANSCRIPTION

# 3.9.9 A stalled RNA polymerase can restart

## Key Concepts

- An arrested RNA polymerase can restart transcription by cleaving the RNA transcript to generate a new 3′ end.

RNA polymerase must be able to handle situations when transcription is blocked. This can happen, for example, when DNA is damaged. A model system for such situations is provided by arresting elongation *in vitro* by omitting one of the necessary precursor nucleotides. When the missing nucleotide is restored, the enzyme can overcome the block by cleaving the 3′ end of the RNA, to create a new 3′ terminus for chain elongation. The cleavage involves accessory factors in addition to the enzyme itself. In the case of *E. coli* RNA polymerase, the proteins GreA and GreB release the RNA polymerase from elongation arrest. In eukaryotic cells, RNA polymerase II requires an accessory factor ($TF_{II}S$), which enables the polymerase to cleave a few ribonucleotides from the 3′ terminus of the RNA product.

The catalytic site of RNA polymerase undertakes the actual cleavage in each case. The roles of GreB and $TF_{II}S$ are to convert the enzyme's catalytic site into a ribonucleolytic site. Although there is no sequence homology between the factors, crystal structures of their complexes with the respective RNA polymerases suggest that they function in a similar way (4187; 4188). Each of the factors inserts a narrow protein domain (in one case a zinc ribbon, in the other a coiled coil) deep into RNA polymerase, where it terminates within the catalytic site. The inserted domain positions two acidic amino acids close to the primary catalytic magnesium ion of the active site; this allows the introduction of a second magnesium ion, which converts the catalytic site to a ribonucleolytic site.

The reason for this reaction may be that stalling causes the template to be mispositioned, so that the 3′ terminus is no longer located in the active site. Cleavage and backtracking is necessary to place the terminus in the right location for addition of further bases.

We see therefore that RNA polymerase has the facility to unwind and rewind DNA, to hold the separated strands of DNA and the RNA product, to catalyze the addition of ribonucleotides to the growing RNA chain, and to adjust to difficulties in progressing by cleaving the RNA product and restarting RNA synthesis (with the assistance of some accessory factors).

*Last updated on 9-26-2003*

## References

4187. Opalka, N., Chlenov, M., Chacon, P., Rice, W. J., Wriggers, W., and Darst, S. A. (2003). *Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase.* Cell 114, 335-345.

4188. Kettenberger, H., Armache, K. J., and Cramer, P. (2003). *Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage.* Cell 114, 347-357.

TRANSCRIPTION

# 3.9.10 How does RNA polymerase find promoter sequences?

## Key Concepts

- The rate at which RNA polymerase binds to promoters is too fast to be accounted for by random diffusion.
- RNA polymerase probably binds to random sites on DNA and exchanges them with other sequences very rapidly until a promoter is found.

How is RNA polymerase distributed in the cell? A (somewhat speculative) picture of the enzyme's situation is depicted in **Figure 9.21**:

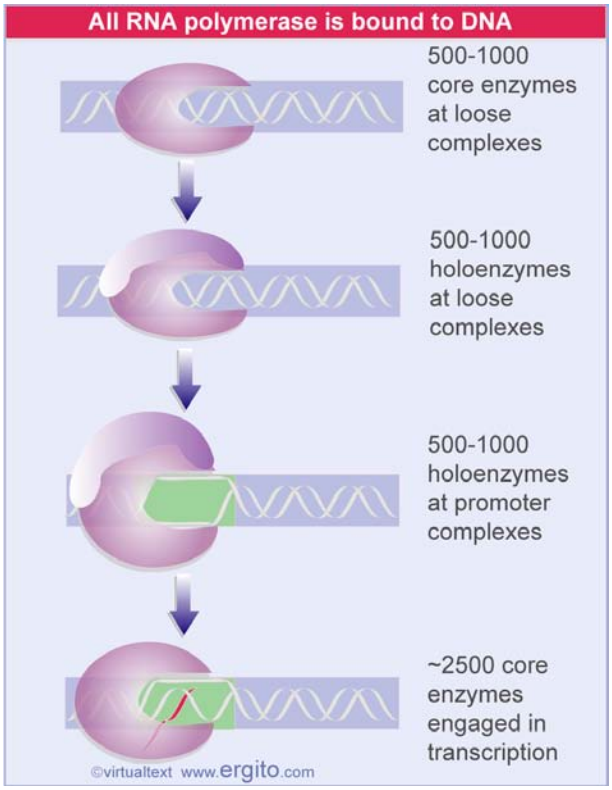


**Figure 9.21** Core enzyme and holoenzyme are distributed on DNA, and very little RNA polymerase is free.

- Excess core enzyme exists largely as closed loose complexes, because the enzyme enters into them rapidly and leaves them slowly. There is very little, if any, free core enzyme.

- There is enough sigma factor for about one third of the polymerases to exist as holoenzymes, and they are distributed between loose complexes at nonspecific sites and binary complexes (mostly closed) at promoters.
- About half of the RNA polymerases consist of core enzymes engaged in transcription.
- How much holoenzyme is free? We do not know, but we suspect that the amount is very small.

RNA polymerase must find promoters within the context of the genome. Suppose that a promoter is a stretch of ~60 bp; how is it distinguished from the $4 \times 10^6$ bp that comprise the *E. coli* genome? The next three figures illustrate the principle of some possible models.

**Figure 9.22** shows the simplest model for promoter binding, in which RNA polymerase moves by random diffusion. Holoenzyme very rapidly associates with, and dissociates from, loose binding sites. So it could continue to make and break a series of closed complexes until (by chance) it encounters a promoter. Then its recognition of the specific sequence would allow tight binding to occur by formation of an open complex.

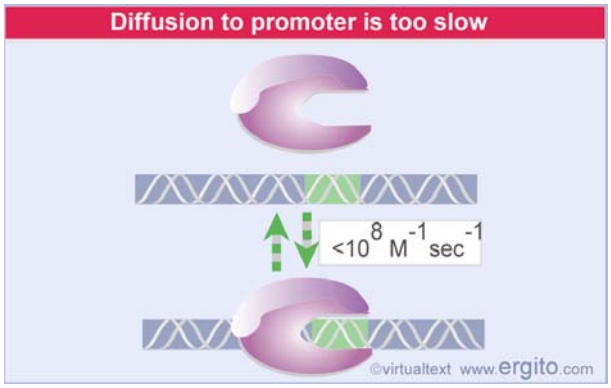


**Figure 9.22** The forward rate constant for RNA polymerase binding to promoters is faster than random diffusion.

For RNA polymerase to move from one binding site on DNA to another, it must dissociate from the first site, find the second site, and then associate with it. Movement from one site to another is limited by the speed of diffusion through the medium. Diffusion sets an upper limit for the rate constant for associating with a 60 bp target of $<10^8$ $M^{-1}$ $sec^{-1}$. But the actual forward rate constant for some promoters *in vitro* appears to be ~$10^8$ $M^{-1}$ $sec^{-1}$, at or above the diffusion limit. If this value applies *in vivo*, the time required for random cycles of successive association and dissociation at loose binding sites is too great to account for the way RNA polymerase finds its promoter.

RNA polymerase must therefore use some other means to seek its binding sites. **Figure 9.23** shows that the process could be speeded up if the initial target for RNA polymerase is the whole genome, not just a specific promoter sequence. By increasing the target size, the rate constant for diffusion to DNA is correspondingly

increased, and is no longer limiting.

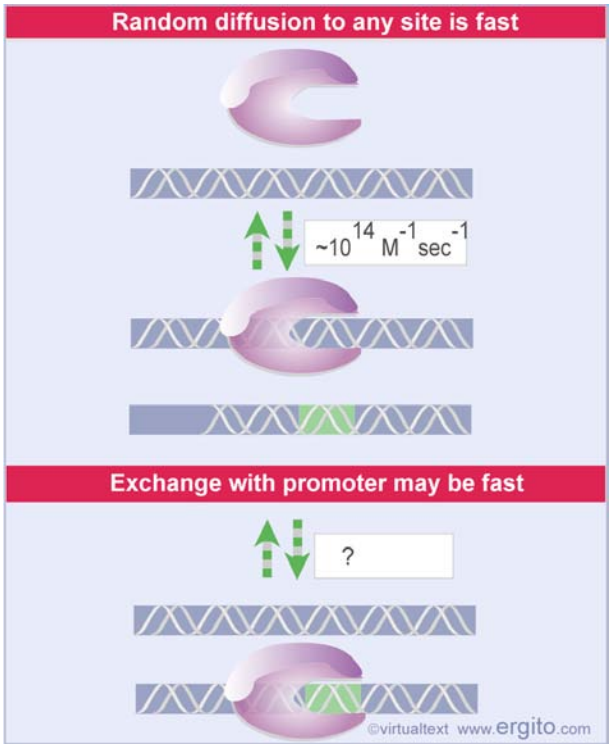


**Figure 9.23** RNA polymerase binds very rapidly to random DNA sequences and could find a promoter by direct displacement of the bound DNA sequence.

If this idea is correct, a free RNA polymerase binds DNA and then remains in contact with it. How does the enzyme move from a random (loose) binding site on DNA to a promoter? The most likely model is to suppose that the bound sequence is directly displaced by another sequence. Having taken hold of DNA, the enzyme exchanges this sequence with another sequence very rapidly, and continues to exchange sequences until a promoter is found. Then the enzyme forms a stable, open complex, after which initiation occurs. The search process becomes much faster because association and dissociation are virtually simultaneous, and time is not spent commuting between sites. Direct displacement can give a "directed walk," in which the enzyme moves preferentially from a weak site to a stronger site.

Another idea supposes that the enzyme slides along the DNA by a one-dimensional random walk, as shown in **Figure 9.24**, being halted only when it encounters a promoter. However, there is no evidence that RNA polymerase (or other DNA-binding proteins) can function in this manner.

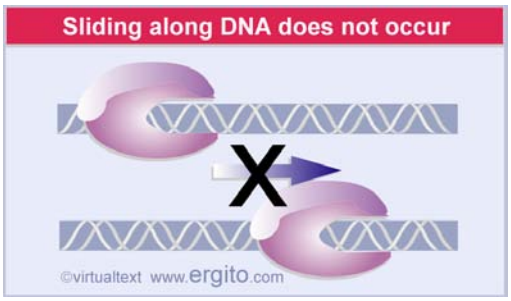


**Figure 9.24** RNA polymerase does not slide along DNA.

*This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.3.9.10*

TRANSCRIPTION

# 3.9.11 Sigma factor controls binding to DNA

## Key Concepts

- A change in association between sigma and holoenzyme changes binding affinity for DNA so that core enzyme can move along DNA.

RNA polymerase encounters a dilemma in reconciling its needs for initiation with those for elongation. Initiation requires tight binding *only* to particular sequences (promoters), while elongation requires close association with *all* sequences that the enzyme encounters during transcription. **Figure 9.25** illustrates how the dilemma is solved by the reversible association between sigma factor and core enzyme. As mentioned previously (see *Molecular Biology 3.9.8 The association with sigma factor changes at initiation*), sigma factor is either released following initiation or changes its association with core enzyme so that it no longer participates in DNA binding (2003; 2004). Because there are fewer molecules of sigma than of core enzyme, the utilization of core enzyme requires that sigma recycles. This occurs immediately after initiation (as shown in the figure) in about one third of cases; presumably sigma and core dissociate at some later point in the other cases.

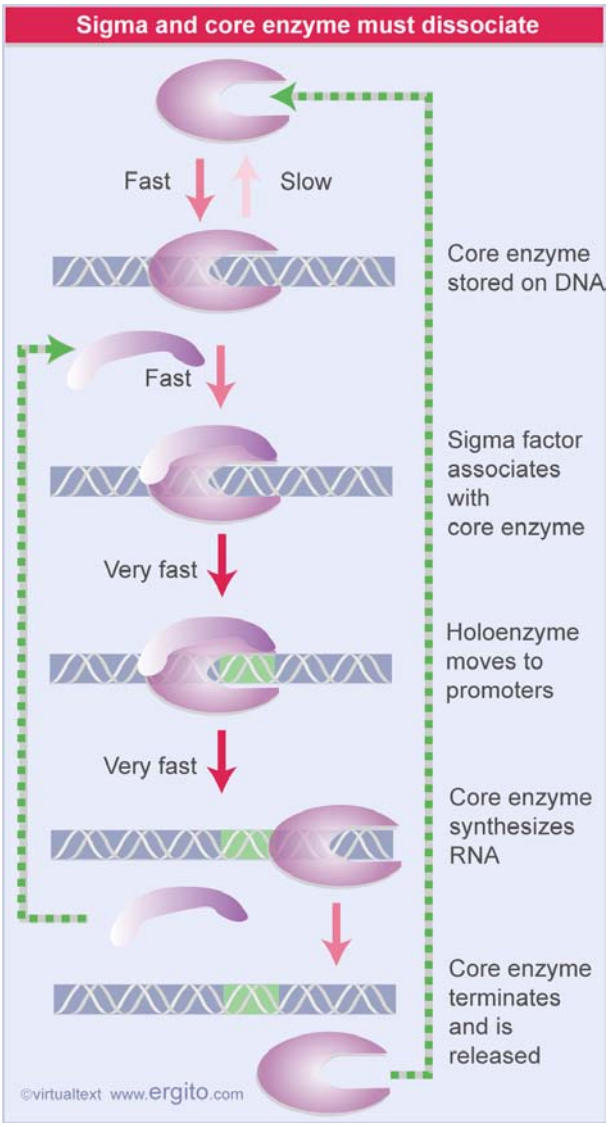


**Figure 9.25** Sigma factor and core enzyme recycle at different points in transcription.

Irrespective of the exact timing of its release from core enzyme, sigma factor is involved only in initiation. It becomes unnecessary when abortive initiation is concluded and RNA synthesis has been successfully initiated. We do not know whether the state of polymerase changes as a consequence of overcoming abortive initiation, or whether instead it is the change in state that ends abortive initiation and allows elongation to commence.

When sigma factor is released from core enzyme, it becomes immediately available for use by another core enzyme. Whether sigma is released or remains more loosely associated with core enzyme, the core enzyme in the ternary complex is bound very tightly to DNA. It is essentially "locked in" until elongation has been completed. When transcription terminates, the core enzyme is released. It is then "stored" by binding to a loose site on DNA. If it has lost its sigma factor, it must find another sigma factor in order to undertake a further cycle of transcription.

Core enzyme has a high intrinsic affinity for DNA, which is increased by the presence of nascent RNA. But its affinity for loose binding sites is too high to allow the enzyme to distinguish promoters efficiently from other sequences. By reducing the stability of the loose complexes, sigma allows the process to occur much more rapidly; and by stabilizing the association at tight binding sites, the factor drives the reaction irreversibly into the formation of open complexes. When the enzyme releases sigma (or changes its association with it), it reverts to a general affinity for all DNA, irrespective of sequence, that suits it to continue transcription.

What is responsible for the ability of holoenzyme to bind specifically to promoters? Sigma factor has domains that recognize the promoter DNA. As an independent polypeptide, sigma does not bind to DNA, but when holoenzyme forms a tight binding complex, σ contacts the DNA in the region upstream of the startpoint. This difference is due to a change in the conformation of sigma factor when it binds to core enzyme. The N-terminal region of free sigma factor suppresses the activity of the DNA-binding region; when sigma binds to core, this inhibition is released, and it becomes able to bind specifically to promoter sequences (see also **Figure 9.35** later). The inability of free sigma factor to recognize promoter sequences may be important: if σ could freely bind to promoters, it might block holoenzyme from initiating transcription.

*Last updated on 9-18-2001*

## References

2003. Bar-Nahum, G. and Nudler, E. (2001). *Isolation and characterization of sigma(70)-retaining transcription elongation complexes from E. coli.* Cell 106, 443-451.

2004. Mukhopadhyay, J., Kapanidis, A. N., Mekler, V., Kortkhonjia, E., Ebright, Y. W., and Ebright, R. H. (2001). *Translocation of sigma(70) with RNA Polymerase during Transcription. Fluorescence Resonance Energy Transfer Assay for Movement Relative to DNA.* Cell 106, 453-463.

TRANSCRIPTION

# 3.9.12 Promoter recognition depends on consensus sequences

## Key Terms

**Conserved** positions are defined when many examples of a particular nucleic acid or protein are compared and the same individual bases or amino acids are always found at particular locations.

A **consensus sequence** is an idealized sequence in which each position represents the base most often found when many actual sequences are compared.

The **-10 sequence** is the consensus sequence centered about 10 bp before the startpoint of a bacterial gene. It is involved in melting DNA during the initiation reaction.

The **-35 sequence** is the consensus sequence centered about 35 bp before the startpoint of a bacterial gene. It is involved in initial recognition by RNA polymerase.

## Key Concepts

- A promoter is defined by the presence of short consensus sequences at specific locations.
- The promoter consensus sequences consist of a purine at the startpoint, the hexamer TATAAT centered at –10, and another hexamer centered at –35.
- Individual promoters usually differ from the consensus at one or more positions.

As a sequence of DNA whose function is to be *recognized by proteins*, a promoter differs from sequences whose role is to be transcribed or translated. The information for promoter function is provided directly by the DNA sequence: its structure is the signal. This is a classic example of a *cis*-acting site, as defined previously in **Figure 1.40** and **Figure 1.41**. By contrast, expressed regions gain their meaning only after the information is transferred into the form of some other nucleic acid or protein.

A key question in examining the interaction between an RNA polymerase and its promoter is how the protein recognizes a specific promoter sequence. Does the enzyme have an active site that distinguishes the chemical structure of a particular sequence of bases in the DNA double helix? How specific are its requirements?

One way to design a promoter would be for a particular sequence of DNA to be recognized by RNA polymerase. Every promoter would consist of, or at least include, this sequence. In the bacterial genome, the minimum length that could provide an adequate signal is 12 bp. (Any shorter sequence is likely to occur – just by chance – a sufficient number of additional times to provide false signals. The minimum length required for unique recognition increases with the size of genome.) The 12 bp sequence need not be contiguous. If a specific number of base pairs

separates two constant shorter sequences, their combined length could be less than 12 bp, since the *distance* of separation itself provides a part of the signal (even if the intermediate *sequence* is itself irrelevant).

Attempts to identify the features in DNA that are necessary for RNA polymerase binding started by comparing the sequences of different promoters. Any essential nucleotide sequence should be present in all the promoters. Such a sequence is said to be **conserved**. However, a conserved sequence need not necessarily be conserved at every single position; some variation is permitted. How do we analyze a sequence of DNA to determine whether it is sufficiently conserved to constitute a recognizable signal?

Putative DNA recognition sites can be defined in terms of an idealized sequence that represents the base most often present at each position. A **consensus sequence** is defined by aligning all known examples so as to maximize their homology. For a sequence to be accepted as a consensus, each particular base must be reasonably predominant at its position, and most of the actual examples must be related to the consensus by rather few substitutions, say, no more than 1-2.

The striking feature in the sequence of promoters in *E. coli* is the *lack of any extensive conservation of sequence* over the 60 bp associated with RNA polymerase. The sequence of much of the binding site is irrelevant. But some short stretches within the promoter are conserved, and they are critical for its function. *Conservation of only very short consensus sequences is a typical feature of regulatory sites (such as promoters) in both prokaryotic and eukaryotic genomes.*

There are four (perhaps five) conserved features in a bacterial promoter: the startpoint; the –10 sequence; the –35 sequence; the separation between the –10 and –35 sequences; and (sometimes) the UP element (for review see 2560):

- The startpoint is usually (>90% of the time) a purine. It is common for the startpoint to be the central base in the sequence CAT, but the conservation of this triplet is not great enough to regard it as an obligatory signal.
- Just upstream of the startpoint, a 6 bp region is recognizable in almost all promoters. The center of the hexamer generally is close to 10 bp upstream of the startpoint; the distance varies in known promoters from position -18 to -9. Named for its location, the hexamer is often called the **-10 sequence**. Its consensus is *TATAAT*, and can be summarized in the form

$$T_{80}\ A_{95}\ T_{45}\ A_{60}\ A_{50}\ T_{96}$$

where the subscript denotes the percent occurrence of the most frequently found base, varying from 45-96%. (A position at which there is no discernible preference for any base would be indicated by N.) If the frequency of occurrence indicates likely importance in binding RNA polymerase, we would expect the initial highly conserved TA and the final almost completely conserved T in the –10 sequence to be the most important bases.

- Another conserved hexamer is centered ~35 bp upstream of the startpoint. This is called the **-35 sequence**. The consensus is *TTGACA*; in more detailed form, the conservation is

$T_{82}\ T_{84}\ G_{78}\ A_{65}\ C_{54}\ A_{45}$

- The distance separating the –35 and –10 sites is between 16-18 bp in 90% of promoters; in the exceptions, it is as little as 15 or as great as 20 bp. *Although the actual sequence in the intervening region is unimportant, the distance is critical in holding the two sites at the appropriate separation for the geometry of RNA polymerase.*

- Some promoters have an A-T-rich sequence located farther upstream. This is called the UP element (2559). It interacts with the α subunit of the RNA polymerase. It is typically found in promoters that are highly expressed, such as the promoters for rRNA genes.

The optimal promoter is a sequence consisting of the –35 hexamer, separated by 17 bp from the –10 hexamer, lying 7 bp upstream of the startpoint. The structure of a promoter, showing the permitted range of variation from this optimum, is illustrated in **Figure 9.26**.


**Figure 9.26** A typical promoter has three components, consisting of consensus sequences at -35 and -10, and the startpoint.

*Last updated on 7-20-2002*

## Reviews

2560. McClure, W. R. (1985). *Mechanism and control of transcription initiation in prokaryotes.* Annu. Rev. Biochem. 54, 171-204.

## References

2559. Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R. L. (1993). *A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase.* Science 262, 1407-1413.

TRANSCRIPTION

# 3.9.13 Promoter efficiencies can be increased or decreased by mutation

## Key Terms

A **down mutation** in a promoter decreases the rate of transcription.

An **up mutation** in a promoter increases the rate of transcription.

## Key Concepts

- Down mutations to decrease promoter efficiency usually decrease conformance to the consensus sequences, whereas up mutations have the opposite effect.
- Mutations in the –35 sequence usually affect initial binding of RNA polymerase.
- Mutations in the –10 sequence usually affect the melting reaction that converts a closed to an open complex.

Mutations are a major source of information about promoter function. Mutations in promoters affect the level of expression of the gene(s) they control, without altering the gene products themselves. Most are identified as bacterial mutants that have lost, or have very much reduced, transcription of the adjacent genes. They are known as **down mutations**. Less often, mutants are found in which there is increased transcription from the promoter. They have **up mutations**.

It is important to remember that "up" and "down" mutations are defined relative to the *usual* efficiency with which a particular promoter functions. This varies widely. So a change that is recognized as a down mutation in one promoter might never have been isolated in another (which in its wild-type state could be even less efficient than the mutant form of the first promoter). Information gained from studies *in vivo* simply identifies the overall direction of the change caused by mutation.

Is the most effective promoter one that has the actual consensus sequences? This expectation is borne out by the simple rule that up mutations usually increase homology with one of the consensus sequences or bring the distance between them closer to 17 bp. Down mutations usually decrease the resemblance of either site with the consensus or make the distance between them more distant from 17 bp. Down mutations tend to be concentrated in the most highly conserved positions, which confirms their particular importance as the main determinant of promoter efficiency (for review see2560). However, there are occasional exceptions to these rules.

To determine the absolute effects of promoter mutations, we must measure the affinity of RNA polymerase for wild-type and mutant promoters *in vitro*. There is ~100-fold variation in the rate at which RNA polymerase binds to different promoters *in vitro*, which correlates well with the frequencies of transcription when their genes are expressed *in vivo*. Taking this analysis further, we can investigate the stage at which a mutation influences the capacity of the promoter. Does it change the

affinity of the promoter for binding RNA polymerase? Does it leave the enzyme able to bind but unable to initiate? Is the influence of an ancillary factor altered?

By measuring the kinetic constants for formation of a closed complex and its conversion to an open complex, as defined in **Figure 9.19**, we can dissect the two stages of the initiation reaction:

- Down mutations in the –35 sequence reduce the rate of closed complex formation (they reduce $K_B$), but do not inhibit the conversion to an open complex.
- Down mutations in the –10 sequence do not affect the initial formation of a closed complex, but they slow its conversion to the open form (they reduce $k_2$).

These results suggest the model shown in **Figure 9.27**. The function of the –35 sequence is to provide the signal for recognition by RNA polymerase, while the –10 sequence allows the complex to convert from closed to open form. We might view the –35 sequence as comprising a "recognition domain," while the –10 sequence comprises an "unwinding domain" of the promoter.

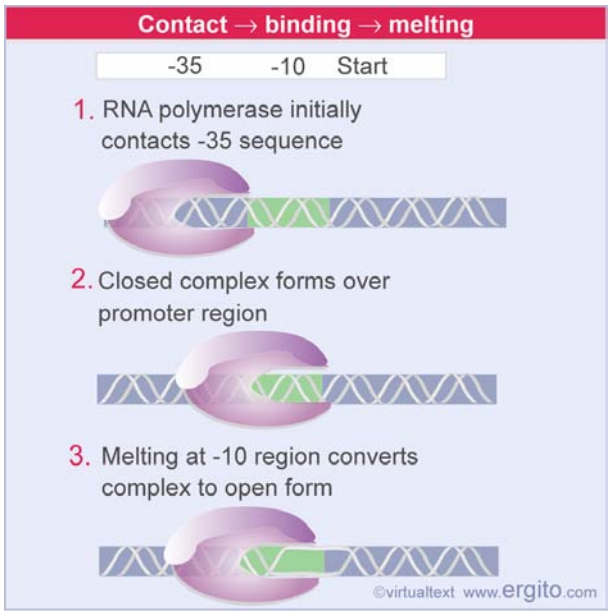


**Figure 9.27** The –35 sequence is used for initial recognition, and the –10 sequence is used for the melting reaction that converts a closed complex to an open complex.

The consensus sequence of the –10 site consists exclusively of A·T base pairs, which assists the initial melting of DNA into single strands. The lower energy needed to disrupt A·T pairs compared with G·C pairs means that a stretch of A·T pairs demands the minimum amount of energy for strand separation.

The sequence immediately around the startpoint influences the initiation event. And the initial transcribed region (from +1 to +30) influences the rate at which RNA

polymerase clears the promoter, and therefore has an effect upon promoter strength. So the overall strength of a promoter cannot be predicted entirely from its –35 and –10 consensus sequences.

A "typical" promoter relies upon its –35 and –10 sequences to be recognized by RNA polymerase, but one or the other of these sequences can be absent from some (exceptional) promoters. In at least some of these cases, the promoter cannot be recognized by RNA polymerase alone, and the reaction requires ancillary proteins, which overcome the deficiency in intrinsic interaction between RNA polymerase and the promoter.

## Reviews

2560. McClure, W. R. (1985). *Mechanism and control of transcription initiation in prokaryotes.* Annu. Rev. Biochem. 54, 171-204.

TRANSCRIPTION

# 3.9.14 RNA polymerase binds to one face of DNA

## Key Terms

**Footprinting** is a technique for identifying the site on DNA bound by some protein by virtue of the protection of bonds in this region against attack by nucleases.

## Key Concepts

- The consensus sequences at –35 and –10 provide most of the contact points for RNA polymerase in the promoter.
- The points of contact lie on one face of the DNA.

The ability of RNA polymerase (or indeed any protein) to recognize DNA can be characterized by **footprinting**. A sequence of DNA bound to the protein is *partially* digested with an endonuclease to attack individual phosphodiester bonds within the nucleic acid. Under appropriate conditions, any particular phosphodiester bond is broken in some, but not in all, DNA molecules. The positions that are cleaved are recognized by using DNA labeled on one strand at one end only. The principle is the same as that involved in DNA sequencing; partial cleavage of an end-labeled molecule at a susceptible site creates a fragment of unique length.

As **Figure 9.28** shows, following the nuclease treatment, the broken DNA fragments are recovered and electrophoresed on a gel that separates them according to length. Each fragment that retains a labeled end produces a radioactive band. The position of the band corresponds to the number of bases in the fragment. The shortest fragments move the fastest, so distance from the labeled end is counted up from the bottom of the gel.

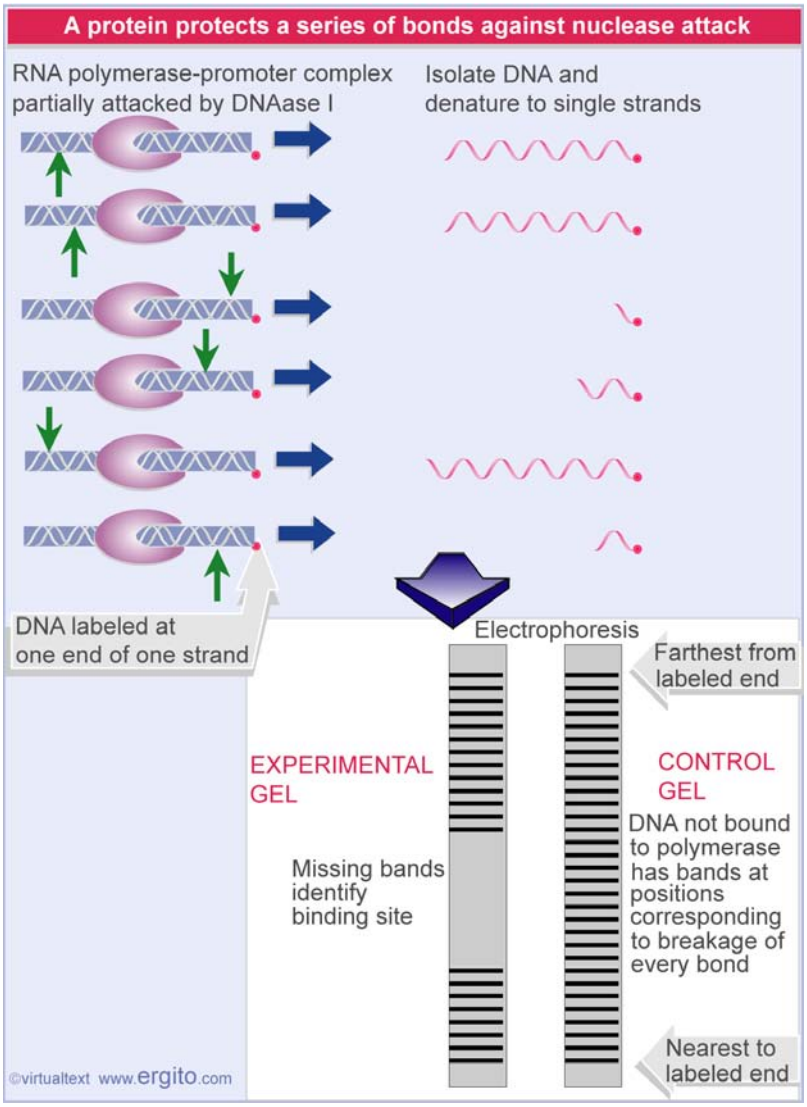


**Figure 9.28** Footprinting identifies DNA-binding sites for proteins by their protection against nicking.

In a free DNA, *every* susceptible bond position is broken in one or another molecule. But when the DNA is complexed with a protein, the region covered by the DNA-binding protein is protected in every molecule. So two reactions are run in parallel: a control of DNA alone; and an experimental mixture containing molecules of DNA bound to the protein. *When a bound protein blocks access of the nuclease to DNA, the bonds in the bound sequence fail to be broken in the experimental mixture.*

In the control, every bond is broken, generating a series of bands, one representing each base. There are 31 bands in the figure. In the protected fragment, bonds cannot be broken in the region bound by the protein, so bands representing fragments of the corresponding sizes are not generated. The absence of bands 9-18 in the figure identifies a protein-binding site covering the region located 9-18 bases from the labeled end of the DNA. By comparing the control and experimental lanes with a sequencing reaction that is run in parallel it becomes possible to "read off" the

corresponding sequence directly, thus identifying the nucleotide sequence of the binding site.

As described previously (see **Figure 9.20**), RNA polymerase initially binds the region from –50 to +20. The points at which RNA polymerase actually contacts the promoter can be identified by modifying the footprinting technique to treat RNA polymerase·promoter complexes with reagents that modify particular bases. We can perform the experiment in two ways:

- The DNA can be modified before it is bound to RNA polymerase. If the modification prevents RNA polymerase from binding, we have identified a base position where contact is essential.
- The RNA polymerase·DNA complex can be modified. Then we compare the pattern of protected bands with that of free DNA and the unmodified complex. Some bands disappear, identifying sites at which the enzyme has protected the promoter against modification. Other bands increase in intensity, identifying sites at which the DNA must be held in a conformation in which it is more exposed.

These changes in sensitivity reveal the geometry of the complex, as summarized in **Figure 9.29** for a typical promoter. The regions at –35 and –10 contain most of the contact points for the enzyme. Within these regions, the same sets of positions tend both to prevent binding if previously modified, and to show increased or decreased susceptibility to modification after binding. Although the points of contact do not coincide completely with sites of mutation, they occur in the same limited region.

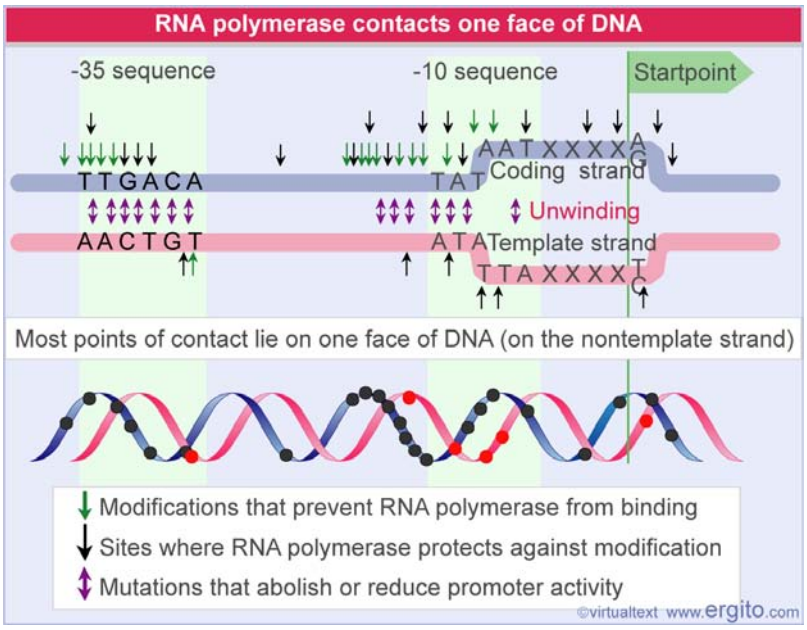


**Figure 9.29** One face of the promoter contains the contact points for RNA.

It is noteworthy that the same *positions* in different promoters provide the contact points, even though a different base is present. This indicates that there is a common

mechanism for RNA polymerase binding, although the reaction does not depend on the presence of particular bases at some of the points of contact. This model explains why some of the points of contact are not sites of mutation. Also, not every mutation lies in a point of contact; they may influence the neighborhood without actually being touched by the enzyme (500).

It is especially significant that the experiments with prior modification identify *only* sites in the same region that is protected by the enzyme against subsequent modification. These two experiments measure different things. Prior modification identifies all those sites that the enzyme must recognize in order to bind to DNA. Protection experiments recognize all those sites that actually make contact in the binary complex. The protected sites include all the recognition sites and also some additional positions, which suggests that the enzyme first recognizes a set of bases necessary for it to "touch down," and then extends its points of contact to additional bases.

The region of DNA that is unwound in the binary complex can be identified directly by chemical changes in its availability. When the strands of DNA are separated, the unpaired bases become susceptible to reagents that cannot reach them in the double helix. Such experiments implicate positions between –9 and +3 in the initial melting reaction. The region unwound during initiation therefore includes the right end of the –10 sequence and extends just past the startpoint.

Viewed in three dimensions, the points of contact upstream of the –10 sequence all lie on one face of DNA. This can be seen in the lower drawing in **Figure 9.29**, in which the contact points are marked on a double helix viewed from one side. Most lie on the coding strand. These bases are probably recognized in the initial formation of a closed binary complex. This would make it possible for RNA polymerase to approach DNA from one side and recognize that face of the DNA. As DNA unwinding commences, further sites that originally lay on the other face of DNA can be recognized and bound.

## Reviews

500. Siebenlist, U., Simpson, R. B., and Gilbert, W. (1980). *E. coli RNA polymerase interacts homologously with two different promoters*. Cell 20, 269-281.

TRANSCRIPTION

# 3.9.15 Supercoiling is an important feature of transcription

## Key Concepts

- Negative supercoiling increases the efficiency of some promoters by assisting the melting reaction.
- Transcription generates positive supercoils ahead of the enzyme and negative supercoils behind it, and these must be removed by gyrase and topoisomerase.

The importance of strand separation in the initiation reaction is emphasized by the effects of supercoiling. Both prokaryotic and eukaryotic RNA polymerases can initiate transcription more efficiently *in vitro* when the template is supercoiled, presumably because the supercoiled structure requires less free energy for the initial melting of DNA in the initiation complex.

The efficiency of some promoters is influenced by the degree of supercoiling. The most common relationship is for transcription to be aided by negative supercoiling. We understand in principle how this assists the initiation reaction. But why should some promoters be influenced by the extent of supercoiling while others are not? One possibility is that the dependence of a promoter on supercoiling is determined by its sequence. This would predict that some promoters have sequences that are easier to melt (and are therefore less dependent on supercoiling), while others have more difficult sequences (and have a greater need to be supercoiled). An alternative is that the location of the promoter might be important if different regions of the bacterial chromosome have different degrees of supercoiling.

Supercoiling also has a continuing involvement with transcription. As RNA polymerase transcribes DNA, unwinding and rewinding occurs, as illustrated in **Figure 9.4**. This requires that either the entire transcription complex rotates about the DNA or the DNA itself must rotate about its helical axis. The consequences of the rotation of DNA are illustrated in **Figure 9.30** in the *twin domain* model for transcription. As RNA polymerase pushes forward along the double helix, it generates positive supercoils (more tightly wound DNA) ahead and leaves negative supercoils (partially unwound DNA) behind. For each helical turn traversed by RNA polymerase, +1 turn is generated ahead and –1 turn behind (506).

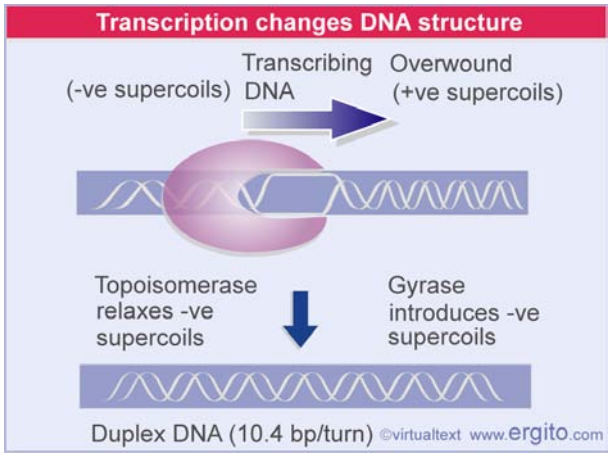


**Figure 9.30** Transcription may generate more tightly wound (positively supercoiled) DNA ahead of RNA polymerase, while the DNA behind becomes less tightly wound (negatively supercoiled).

Transcription therefore has a significant effect on the (local) structure of DNA. As a result, the enzymes gyrase (introduces negative supercoils) and topoisomerase I (removes negative supercoils) are required to rectify the situation in front of and behind the polymerase, respectively. Blocking the activities of gyrase and topoisomerase causes major changes in the supercoiling of DNA. For example, in yeast lacking an enzyme that relaxes negative supercoils, the density of negative supercoiling doubles in a transcribed region. A possible implication of these results is that transcription is responsible for generating a significant proportion of the supercoiling that occurs in the cell.

A similar situation occurs in replication, when DNA must be unwound at a moving replication fork, so that the individual single strands can be used as templates to synthesize daughter strands. Solutions for the topological constraints associated with such reactions are indicated later in **Figure 15.20**.

## References

506. Wu, H.-Y. et al. (1988). *Transcription generates positively and negatively supercoiled domains in the template*. Cell 53, 433-440.

TRANSCRIPTION

# 3.9.16 Substitution of sigma factors may control initiation

## Key Terms

**Heat shock genes** are a set of loci that are activated in response to an increase in temperature (and other abuses to the cell). They occur in all organisms. They usually include chaperones that act on denatured proteins.

## Key Concepts

- *E. coli* has several sigma factors, each of which causes RNA polymerase to initiate at a set of promoters defined by specific –35 and –10 sequences.
- $\sigma^{70}$ is used for general transcription, and the other sigma factors are activated by special conditions.

The division of labors between a core enzyme that undertakes chain elongation and a sigma factor involved in site selection immediately raises the question of whether there is more than one type of sigma, each specific for a different class of promoters. **Figure 9.31** shows the principle of a system in which a substitution of the sigma factor changes the choice of promoter.

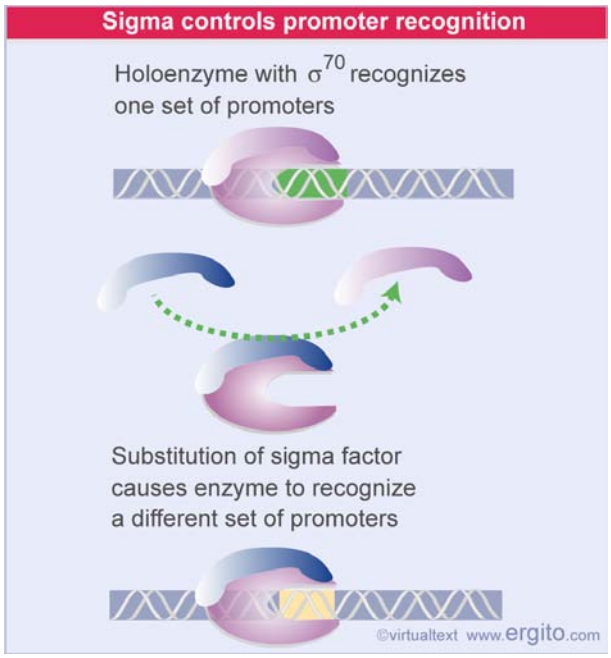


**Figure 9.31** The sigma factor associated with core enzyme determines the set of promoters where transcription is initiated.

*E. coli* uses alternative sigma factors to respond to general environmental changes.

They are listed in **Figure 9.32**. (They are named either by molecular weight of the product or for the gene.) The general factor, responsible for transcription of most genes under normal conditions, is $\sigma^{70}$. The alternative sigma factors $\sigma^{S}$, $\sigma^{32}$, $\sigma^{E}$, and $\sigma^{54}$ are activated in response to environmental changes; $\sigma^{28}$ is used for expression of flagellar genes during normal growth, but its level of expression responds to changes in the environment. All the sigma factors except $\sigma^{54}$ belong to the same protein family and function in the same general manner.

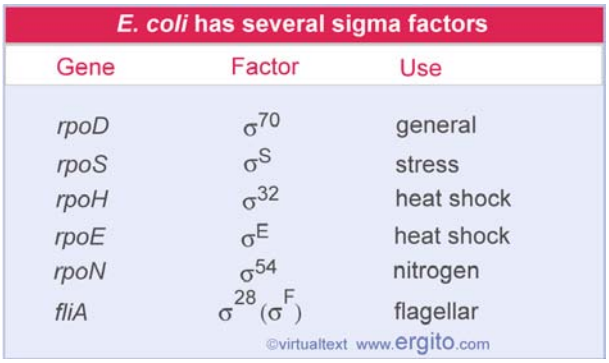


**Figure 9.32** In addition to $\sigma^{70}$, *E. coli* has several sigma factors that are induced by particular environmental conditions. (A number in the name of a factor indicates its mass.)

Temperature fluctuation is a common type of environmental challenge. Many organisms, both prokaryotic and eukaryotic, respond in a similar way. Upon an increase in temperature, synthesis of the proteins currently being made is turned off or down, and a new set of proteins is synthesized. The new proteins are the products of the **heat shock genes**. They play a role in protecting the cell against environmental stress, and are synthesized in response to other conditions as well as heat shock. Several of the heat shock proteins are chaperones. In *E. coli*, the expression of 17 heat shock proteins is triggered by changes at transcription. The gene *rpoH* is a regulator needed to switch on the heat shock response. Its product is $\sigma_{32}$, which functions as an alternative sigma factor that causes transcription of the heat shock genes (499).

The heat shock response is accomplished by increasing the amount of $\sigma^{32}$ when the temperature increases, and decreasing its activity when the temperature change is reversed. The basic signal that induces production of $\sigma^{32}$ is the accumulation of unfolded (partially denatured) proteins that results from increase in temperature. The $\sigma^{32}$ protein is unstable, which is important in allowing its quantity to be increased or decreased rapidly. $\sigma^{70}$ and $\sigma^{32}$ can compete for the available core enzyme, so that the set of genes transcribed during heat shock depends on the balance between them.

Changing sigma factors is a serious matter that has widespread implications for gene expression in the bacterium. It is not surprising, therefore, that the production of new sigma factors can be the target of many regulatory circuits. The factor $\sigma^{S}$ is induced when bacteria make the transition from growth phase to stationary phase, and also in other stress conditions (for review see 3070). It is controlled at two levels. Translation of the *rpoS* mRNA is increased by low temperature or high osmolarity. Proteolysis of the protein product is inhibited by carbon starvation (the typical signal of stationary phase) and by high temperature.

Another group of heat-regulated genes is controlled by the factor $\sigma^E$. It responds to more extreme temperature shifts than $\sigma^{32}$, and is induced by accumulation of unfolded proteins in the periplasmic space or outer membrane. It is controlled by the intricate circuit summarized in **Figure 9.33**. $\sigma^E$ binds to a protein (RseA) that is located in the inner membrane. As a result, it cannot activate transcription. The accumulation of unfolded proteins activates a protease (DegS) in the periplasmic space, which cleaves off the C-terminal end of the RseA protein. This cleavage activates another protein in the inner membrane (YaeL) which cleaves the N-terminal region of RseA (3014; 3010). When this happens, the $\sigma^E$ factor is released, and can then activate transcription. The net result is that the accumulation of unfolded proteins at the periphery of the bacterium is responsible for activating the set of genes controlled by the sigma factor.

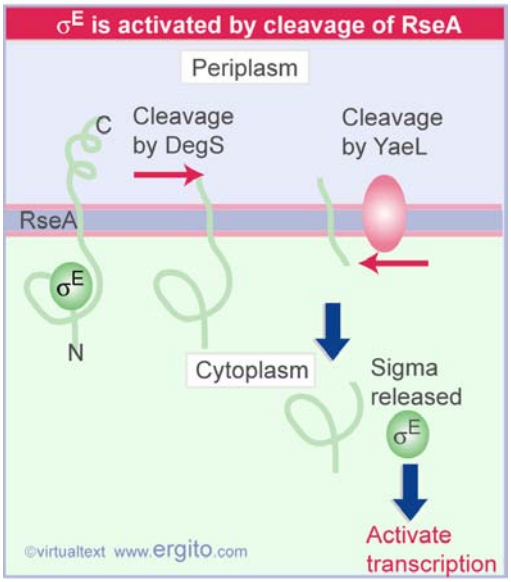


**Figure 9.33** RseA is synthesized as a protein in the inner membrane. Its cytoplasmic domain binds the $\sigma^E$ factor. RseA is cleaved sequentially in the periplasmic space and then in the cytoplasm. The cytoplasmic cleavage releases $\sigma^E$.

This circuit has two interesting parallels with other regulatory circuits. The response to unfolded proteins in eukaryotic cells also uses a pathway in which an unfolded protein (within the endoplasmic reticulum) activates a membrane protein. In this case, the membrane protein is an endonuclease that cleaves an RNA, leading ultimately to a change in splicing that causes the production of a transcription factor (see *Molecular Biology 5.24.17 The unfolded protein response is related to tRNA splicing*). And a more direct parallel is with the first case to be discovered in which cleavage of a membrane protein activates a transcription factor. In this case, the transcription factor itself is synthesized as a membrane protein, and the level of sterols in the membrane controls the activation of proteases that release the transcription factor from the cytosolic domain of the protein (3015).

Another sigma factor is used under conditions of nitrogen starvation. *E. coli* cells contain a small amount of $\sigma^{54}$, which is activated when ammonia is absent from the medium. In these conditions, genes are turned on to allow utilization of alternative

nitrogen sources. Counterparts to this sigma factor have been found in a wide range of bacteria, so it represents a response mechanism that has been conserved in evolution.

Another case of evolutionary conservation of sigma factors is presented by the factor $\sigma^F$, which is present in small amounts and causes RNA polymerase to transcribe genes involved in chemotaxis and flagellar structure. Its counterpart in *B. subtilis* is $\sigma_D$, which controls flagellar and motility genes; factors with the same promoter specificity are present in many species of bacteria.

Each sigma factor causes RNA polymerase to initiate at a particular set of promoters. By analyzing the sequences of these promoters, we can show that each set is identified by unique sequence elements. Indeed, the sequence of each type of promoter ensures that it is recognized only by RNA polymerase directed by the appropriate sigma factor. We can deduce the general rules for promoter recognition from the identification of the genes responding to the sigma factors found in *E. coli* and those involved in sporulation in *B. subtilis* (see *Molecular Biology 3.9.19 Sporulation is controlled by sigma factors*).

A significant feature of the promoters for each enzyme is that *they have the same size and location relative to the startpoint, and they show conserved sequences only around the usual centers of –35 and –10.* ( $\sigma^{54}$ is an exception for which the consensus sequences are closer together, and are positioned at –24 and –12; see *Molecular Biology 3.9.17 Sigma factors directly contact DNA*.) As summarized in **Figure 9.34**, the consensus sequences for each set of promoters are different from one another at either or both of the –35 and –10 positions. This means that an enzyme containing a particular sigma factor can recognize only its own set of promoters, so that transcription of the different groups is mutually exclusive. Substitution of one sigma factor by another therefore turns off transcription of the old set of genes as well as turning on transcription of a new set of genes. (Some genes are expressed by RNA polymerases with different sigma factors because they have more than one promoter, each with a different set of consensus sequences.)

**Sigma factors recognize promoters by consensus sequences**

| Gene | Factor | -35 Sequence | Separation | -10 Sequence |
|---|---|---|---|---|
| *rpoD* | $\sigma^{70}$ | TTGACA | 16-18 bp | TATAAT |
| *rpoH* | $\sigma^{32}$ | CCCTTGAA | 13-15 bp | CCCGATNT |
| *rpoN* | $\sigma^{54}$ | CTGGNA | 6 bp | TTGCA |
| *fliA* | $\sigma^{28}$ ($\sigma^F$) | CTAAA | 15 bp | GCCGATAA |
| *sigH* | $\sigma^H$ | AGGANPuPu | 11-12 bp | GCTGAATCA |



**Figure 9.34** *E. coli* sigma factors recognize promoters with different consensus sequences.

*Last updated on 11-10-2002*

## Reviews

3070. Hengge-Aronis, R. (2002). *Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase.* Microbiol. Mol. Biol. Rev. 66, 373-393.

## References

499. Grossman, A. D., Erickson, J. W., and Gross, C. A. (1984). *The htpR gene product of E. coli is a sigma factor for heat-shock promoters*. Cell 38, 383-390.

3010. Alba, B. M., Onufryk, C., Lu, C. Z., and Gross, C. A. (2002). *DegS and YaeL participate sequentially in the cleavage of RseA to activate the sigma(E)-dependent extracytoplasmic stress response.* Genes Dev. 16, 2156-2168.

3014. Kanehara, K., Ito, K., and Akiyama, Y. (2002). *YaeL (EcfE) activates the sigma(E) pathway of stress response through a site-2 cleavage of anti-sigma(E), RseA.* Genes Dev. 16, 2147-2155.

3015. Sakai, J., Duncan, E. A., Rawson, R. B., Hua, X., Brown, M. S., and Goldstein, J. L. (1996). *Sterol-regulated release of SREBP-2 from cell membranes requires two sequential cleavages, one within a transmembrane segment.* Cell 85, 1037-1046.

TRANSCRIPTION

# 3.9.17 Sigma factors directly contact DNA

## Key Concepts

- $\sigma^{70}$ changes its structure to release its DNA-binding regions when it associates with core enzyme.
- $\sigma^{70}$ binds both the –35 and –10 sequences.

The definition of a series of different consensus sequences recognized at –35 and –10 by holoenzymes containing different sigma factors (see **Figure 9.34**) carries the immediate implication that the sigma subunit must itself contact DNA in these regions. This suggests the general principle that there is a common type of relationship between sigma and core enzyme, in which the sigma factor is positioned in such a way as to make critical contacts with the promoter sequences in the vicinity of –35 and –10.

Direct evidence that sigma contacts the promoter directly at both the –35 and –10 consensus sequences is provided by mutations in sigma that suppress mutations in the consensus sequences. When a mutation at a particular position in the promoter prevents recognition by RNA polymerase, and a compensating mutation in sigma factor allows the polymerase to use the mutant promoter, the most likely explanation is that the relevant base pair in DNA is contacted by the amino acid that has been substituted.

Comparisons of the sequences of several bacterial sigma factors identify regions that have been conserved. Their locations in *E. coli* $\sigma^{70}$ are summarized in **Figure 9.35** (503). The crystal structure of a sigma factor fragment from the bacterium *Thermus aquaticus* shows that these regions fold into three independent domains in the protein: domain $\sigma_2$ contains 1.2-2.4, $\sigma_3$ contains 3.0-1.3, and $\sigma_4$ contains 4.1-4.2 (3071).

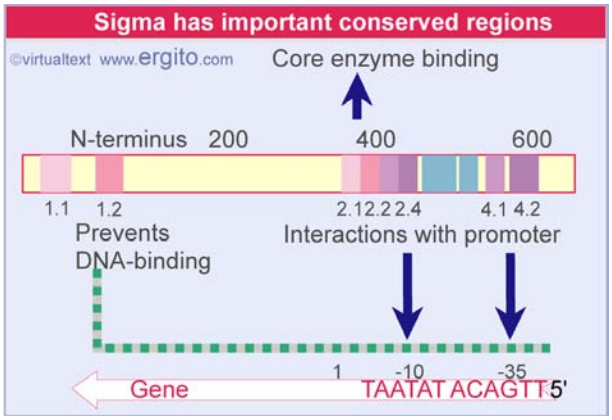


**Figure 9.35** A map of the *E. coli* $\sigma^{70}$ factor identifies conserved regions. Regions 2.1 and 2.2 contact core polymerase, 2.3 is required for melting, and 2.4 and 4.2 contact the –10 and –35 promoter elements. The N-terminal region prevents 2.4 and 4.2 from binding to DNA in the absence of core enzyme.

**Figure 9.35** shows that two short parts of regions 2 and 4 (named 2.4 and 4.2) are involved in contacting bases in the –10 and –35 elements, respectively. Both of these regions form short stretches of α-helix in the protein. Experiments with heteroduplexes show that $\sigma^{70}$ makes contacts with bases principally on the coding strand, and it continues to hold these contacts after the DNA has been unwound in this region. This suggests that sigma factor could be important in the melting reaction.

The use of α-helical motifs in proteins to recognize duplex DNA sequences is common (see *Molecular Biology 3.12.12 Repressor uses a helix-turn-helix motif to bind DNA*). Amino acids separated by 3-4 positions lie on the same face of an α-helix and are therefore in a position to contact adjacent base pairs. **Figure 9.36** shows that amino acids lying along one face of the 2.4 region α-helix contact the bases at positions –12 to –10 of the –10 promoter sequence.

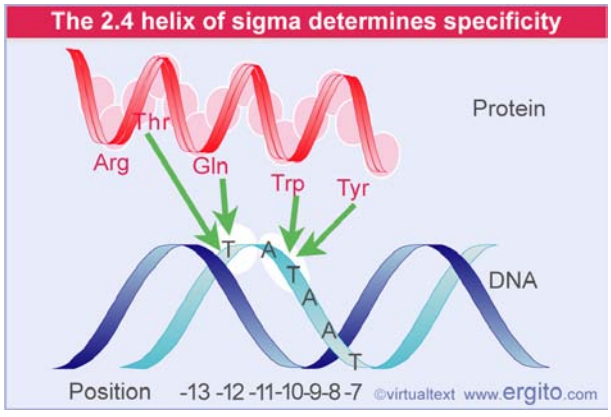


**Figure 9.36** Amino acids in the 2.4 α -helix of $\sigma^{70}$ contact specific bases in the coding strand of the –10 promoter sequence.

Region 2.3 resembles proteins that bind single-stranded nucleic acids, and is involved in the melting reaction. Regions 2.1 and 2.2 (which is the most highly conserved part of sigma) are involved in the interaction with core enzyme. It is assumed that all sigma factors bind the same regions of the core polymerase (ensuring that the reactions are competitive).

The N-terminal region of $\sigma^{70}$ has important regulatory functions. If it is removed, the shortened protein becomes able to bind specifically to promoter sequences. This suggests that the N-terminal region behaves as an autoinhibition domain. It occludes the DNA-binding domains when $\sigma^{70}$ is free. Association with core enzyme changes the conformation of sigma so that the inhibition is released, and the DNA-binding domains can contact DNA.

**Figure 9.37** schematizes the conformational change in sigma at open complex formation. When sigma binds to the core polymerase, the N-terminal domain swings ~20 Å away from the DNA-binding domains, and the DNA-binding domains separate from one another by ~15 Å, presumably to acquire a more elongated conformation appropriate for contacting DNA. Mutations in either the –10 or –35 sequences prevent an (N-terminal-deleted) $\sigma^{70}$ from binding to DNA, which suggests that $\sigma^{70}$ contacts both sequences simultaneously. This implies that the sigma factor must have a rather elongated structure, extending over the ~68 Å of two turns of DNA.

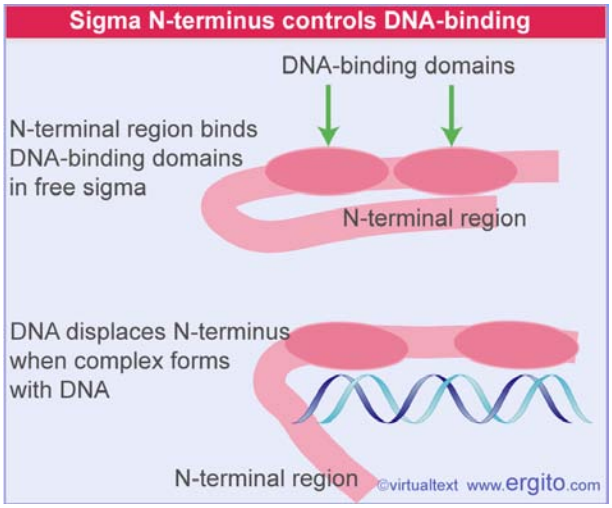


**Figure 9.37** The N-terminus of sigma blocks the DNA-binding regions from binding to DNA. When an open complex forms, the N-terminus swings 20 Å away, and the two DNA-binding regions separate by 15 Å.

In the free holoenzyme, the N-terminal domain is located in the active site of the core enzyme components, essentially mimicking the location that DNA will occupy when a transcription complex is formed (2405). When the holoenzyme forms an open complex on DNA, the N-terminal sigma domain is displaced from the active site. Its relationship with the rest of the protein is therefore very flexible, and changes when sigma binds to core enzyme, and again when the holoenzyme binds to DNA.

Comparisons of the crystal structures of the core enzyme and holoenzyme show that

sigma factor lies largely on the surface of the core enzyme (2509). **Figure 9.38** shows that it has an elongated structure that extends past the DNA-binding site. This places it in a position to contact DNA during the initial binding. The DNA helix has to move some 16Å from the initial position in order to enter the active site. **Figure 9.39** illustrates this movement, looking in cross-section down the helical axis of the DNA.

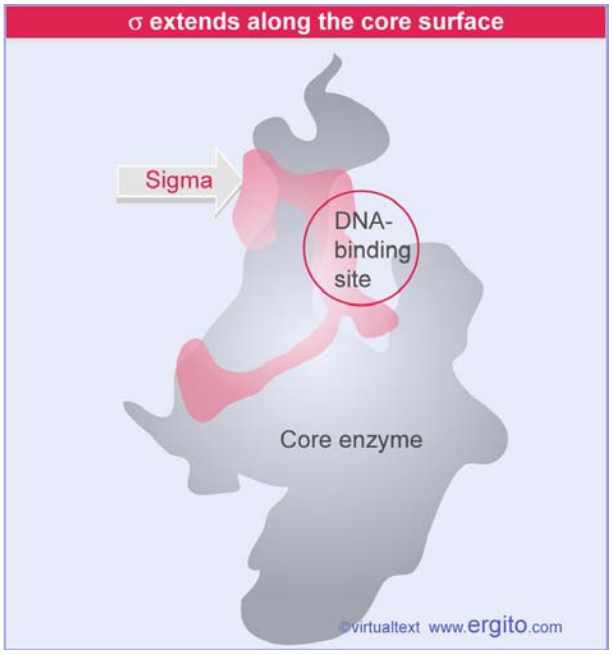


**Figure 9.38** Sigma factor has an elongated structure that extends along the surface of the core subunits when the holoenzyme is formed.

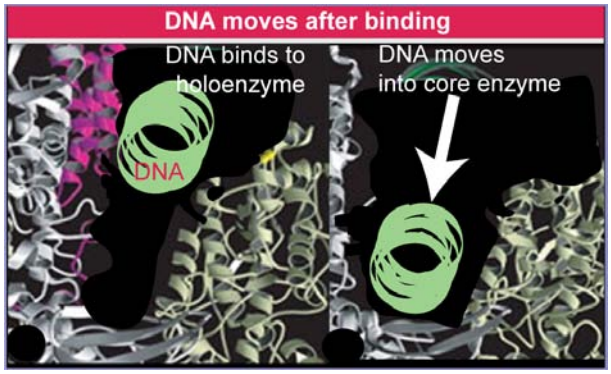


**Figure 9.39** DNA initially contacts sigma factor (pink) and core enzyme (gray). It moves deeper into the core enzyme to make contacts at the –10 sequence. When sigma is released, the width of the passage containing DNA increases.

An interesting difference in behavior is found with the $\sigma^{54}$ factor. This causes RNA polymerase to recognize promoters that have a distinct consensus sequence, with a conserved element at –12 and another close by at –24 (given in the "–35" column of **Figure 9.32**). So the geometry of the polymerase-promoter complex is different

under the direction of this sigma factor. Another difference in the mechanism of regulation is that high level transcription directed by $\sigma^{54}$ requires other activators to bind to sites that are quite distant from the promoter. This contrasts with the other types of bacterial promoter, where the regulator sites are always in close proximity to the promoter. The behavior of $\sigma^{54}$ itself is different from other sigma factors, most notably in its ability to bind to DNA independently of core polymerase. In these regards, $\sigma^{54}$ is more like the eukaryotic regulators we discuss in *Molecular Biology 5.21 Promoters and enhancers* than the typical prokaryotic regulators discussed in *Molecular Biology 3.10 The operon*.

*Last updated on 11-10-2002*

## References

503. Dombrowski, A. J. et al. (1992). *Polypeptides containing highly conserved regions of transcription initiation factor $\sigma^{70}$ exhibit specificity of binding to promoter DNA.* Cell 70, 501-512.

2405. Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A. N., Niu, W., Ebright, Y. W., Levy, R., and Ebright, R. H. (2002). *Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex.* Cell 108, 599-614.

2509. Vassylyev, D. G, Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S., Yokoyama S. (2002). *Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6* Å *resolution.* Nature 417, 712-719.

3071. Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L., and Darst, S. A. (2002). *Structure of the bacterial RNA polymerase promoter specificity sigma subunit.* Mol. Cell 9, 527-539.

TRANSCRIPTION

# 3.9.18 Sigma factors may be organized into cascades

## Key Terms

**Early genes** are transcribed before the replication of phage DNA. They code for regulators and other proteins needed for later stages of infection.

**Middle genes** are phage genes that are regulated by the proteins coded by early genes. Some proteins coded by middle genes catalyze replication of the phage DNA; others regulate the expression of a later set of genes.

**Late genes** are transcribed when phage DNA is being replicated. They code for components of the phage particle.

A **cascade** is a sequence of events, each of which is stimulated by the previous one. In transcriptional regulation, as seen in sporulation and phage lytic development, it means that regulation is divided into stages, and at each stage, one of the genes that are expressed codes for a regulator needed to express the genes of the next stage.

## Key Concepts

- A cascade of sigma factors is created when one sigma factor is required to transcribe the gene coding for the next sigma factor.
- The early genes of phage SPO1 are transcribed by host RNA polymerase.
- One of the early genes codes for a sigma factor that causes RNA polymerase to transcribe the middle genes.
- Two of the middle genes code for subunits of a sigma factor that causes RNA polymerase to transcribe the late genes.

Sigma factors are used extensively to control initiation of transcription in the bacterium *B. subtilis*, where ~10 different σ factors are known. Some are present in vegetative cells; other are produced only in the special circumstances of phage infection or the change from vegetative growth to sporulation.

The major RNA polymerase found in *B. subtilis* cells engaged in normal vegetative growth has the same structure as that of *E. coli*, $\alpha_2$ β β′ σ. Its sigma factor (described as $\sigma^{43}$ or $\sigma^{A}$) recognizes promoters with the same consensus sequences used by the *E. coli* enzyme under direction from $\sigma^{70}$. Several variants of the RNA polymerase that contain other sigma factors are found in much smaller amounts. The variant enzymes recognize different promoters on the basis of consensus sequences at –35 and –10.

Transitions from expression of one set of genes to expression of another set are a common feature of bacteriophage infection. In all but the very simplest cases, the

development of the phage involves shifts in the pattern of transcription during the infective cycle. These shifts may be accomplished by the synthesis of a phage-encoded RNA polymerase or by the efforts of phage-encoded ancillary factors that control the bacterial RNA polymerase. A well characterized example of control via the production of new sigma factors occurs during infection of *B. subtilis* by phage SPO1.

The infective cycle of SPO1 passes through three stages of gene expression. Immediately on infection, the **early genes** of the phage are transcribed. After 4-5 minutes, the early genes cease transcription and the **middle genes** are transcribed. Then at 8-12 minutes, middle gene transcription is replaced by transcription of **late genes**.

The early genes are transcribed by the holoenzyme of the host bacterium. They are essentially indistinguishable from host genes whose promoters have the intrinsic ability to be recognized by the RNA polymerase $\alpha_2 \beta \beta' \sigma^{43}$.

Expression of phage genes is required for the transitions to middle and late gene transcription. Three regulatory genes, named *28*, *33*, and *34*, control the course of transcription. Their functions are summarized in **Figure 9.40**. The pattern of regulation creates a **cascade**, in which the host enzyme transcribes an early gene whose product is needed to transcribe the middle genes; and then two of the middle genes code for products that are needed to transcribe the late genes.

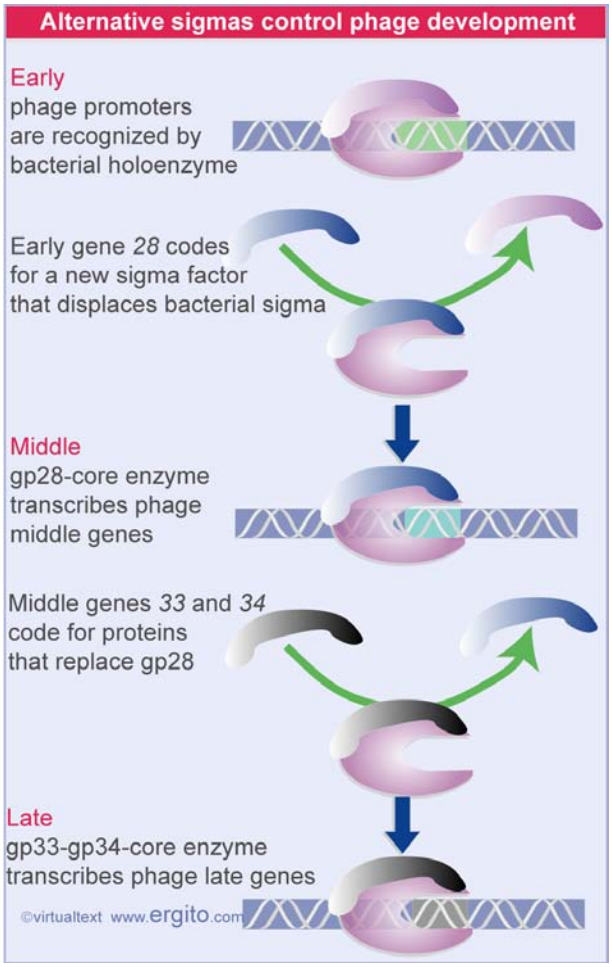


**Figure 9.40** Transcription of phage SPO1 genes is controlled by two successive substitutions of the sigma factor that change the initiation specificity.

Mutants in the early gene *28* cannot transcribe the middle genes. The product of gene *28* (called gp28) is a protein of 26 kD that replaces the host sigma factor on the core enzyme. *This substitution is the sole event required to make the transition from early to middle gene expression.* It creates a holoenzyme that can no longer transcribe the host genes, but instead specifically transcribes the middle genes. We do not know how gp28 displaces $\sigma^{43}$, or what happens to the host sigma polypeptide.

Two of the middle genes are involved in the next transition. Mutations in either gene *33* or *34* prevent transcription of the late genes. The products of these genes form a dimer that replaces gp28 on the core polymerase. Again, we do not know how gp33 and gp34 exclude gp28 (or any residual host $\sigma^{43}$), *but once they have bound to the core enzyme, it is able to initiate transcription only at the promoters for late genes.*

The successive replacements of sigma factor have dual consequences. Each time the subunit is changed, the RNA polymerase becomes able to recognize a new class of genes, *and* it no longer recognizes the previous class. These switches therefore constitute global changes in the activity of RNA polymerase. Probably all or virtually all of the core enzyme becomes associated with the sigma factor of the moment; and

the change is irreversible.

*This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.3.9.18*

TRANSCRIPTION

# 3.9.19 Sporulation is controlled by sigma factors

## Key Terms

**Sporulation** is the generation of a spore by a bacterium (by morphological conversion) or by a yeast (as the product of meiosis).

The **vegetative phase** describes the period of normal growth and division of a bacterium. For a bacterium that can sporulate, this contrasts with the sporulation phase, when spores are being formed.

A **phosphorelay** describes a pathway in which a phosphate group is passed along a series of proteins.

## Key Concepts

- Sporulation divides a bacterium into a mother cell that is lysed and a spore that is released.
- Each compartment advances to the next stage of development by synthesizing a new sigma factor that displaces the previous sigma factor.
- Communication between the two compartments coordinates the timing of sigma factor substitutions.

Perhaps the most extensive example of switches in sigma factors is provided by **sporulation**, an alternative lifestyle available to some bacteria. At the end of the **vegetative phase** in a bacterial culture, logarithmic growth ceases because nutrients in the medium become depleted. This triggers sporulation, as illustrated in **Figure 9.41**. DNA is replicated, a genome is segregated at one end of the cell, and eventually it is surrounded by the tough spore coat. When the septum forms, it generates two independent compartments, the mother cell and the forespore. At the start of the process, one chromosome is attached to each pole of the cell. The growing septum traps part of one chromosome in the forespore, and then a translocase (SpoIIIE) pumps the rest of the chromosome into the forespore.

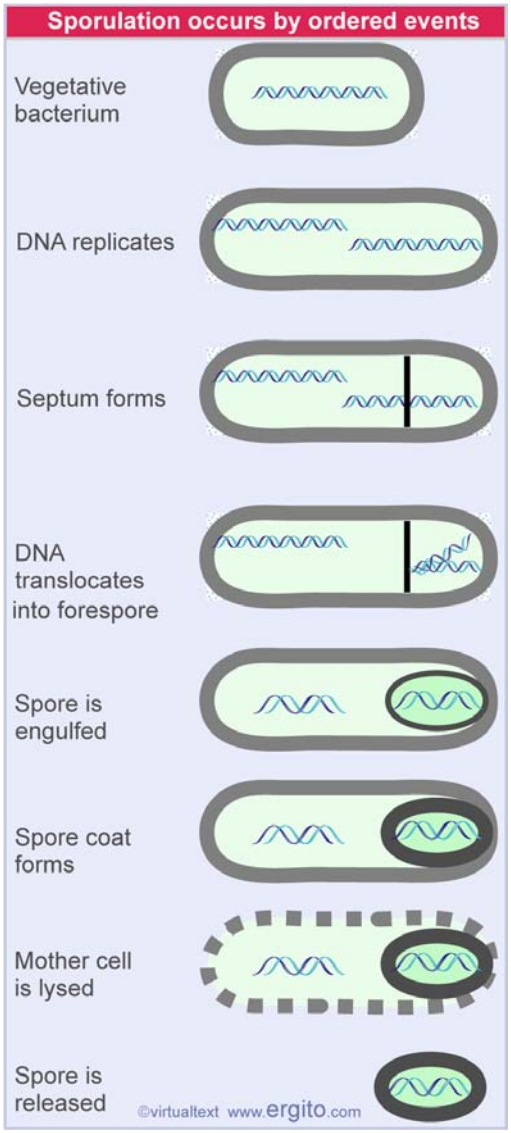


**Figure 9.41** Sporulation involves the differentiation of a vegetative bacterium into a mother cell that is lysed and a spore that is released.

Sporulation takes ~8 hours. It can be viewed as a primitive sort of differentiation, in which a parent cell (the vegetative bacterium) gives rise to two different daughter cells with distinct fates: the mother cell is eventually lysed, while the spore that is released has an entirely different structure from the original bacterium.

Sporulation involves a drastic change in the biosynthetic activities of the bacterium, in which many genes are involved. The basic level of control lies at transcription. Some of the genes that functioned in the vegetative phase are turned off during sporulation, but most continue to be expressed. In addition, the genes specific for sporulation are expressed only during this period. At the end of sporulation, ~40% of the bacterial mRNA is sporulation-specific.

New forms of the RNA polymerase become active in sporulating cells; they contain the same core enzyme as vegetative cells, but have different proteins in place of the vegetative $\sigma^{43}$. The changes in transcriptional specificity are summarized in **Figure 9.42**. The principle is that in each compartment the existing sigma factor is successively displaced by a new factor that causes transcription of a different set of genes. Communication between the compartments occurs in order to coordinate the timing of the changes in the forespore and mother cell (497; 498; for review see 72; 80).

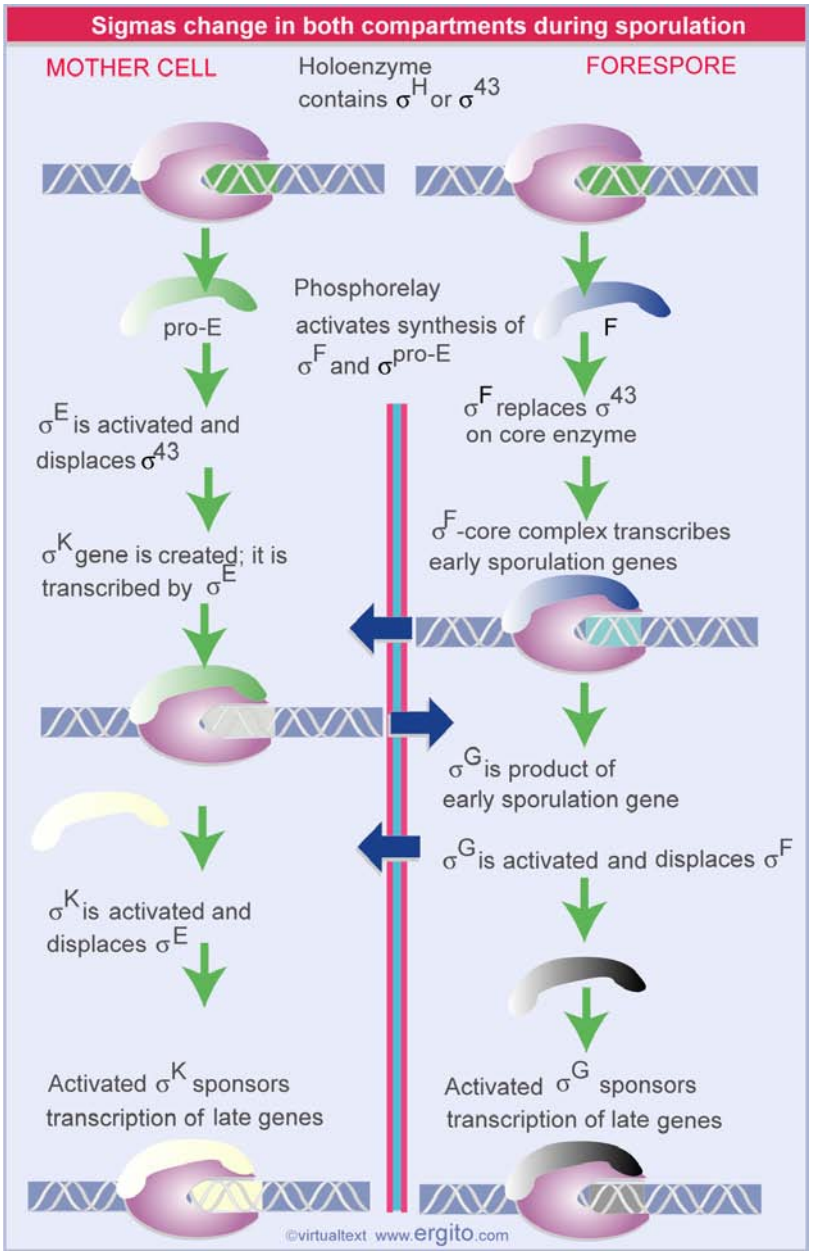


**Figure 9.42** Sporulation involves successive changes in the sigma factors that control the initiation specificity of RNA polymerase. The cascades in the forespore (left) and the mother cell (right) are related by signals passed across the septum (indicated by horizontal arrows).

The sporulation cascade is initiated when environmental conditions trigger a **phosphorelay**, in which a phosphate group is passed along a series of proteins until it reaches SpoOA (see *Molecular Biology Supplement 32.13 Two-component signal transduction*). (Several gene products are involved in this process, whose complexity may reflect the need to avoid mistakes in triggering sporulation unnecessarily.) SpoOA is a transcriptional regulator whose activity is affected by phosphorylation. In the phosphorylated form, it activates transcription of two operons, each of which is transcribed by a different form of the host RNA polymerase. Under the direction of phosphorylated SpoOA, host enzyme utilizing the general $\sigma^{43}$ transcribes the gene coding for the factor $\sigma^{F}$; and host enzyme under the direction of a minor factor, $\sigma^{H}$, transcribes the gene coding for the factor pro- $\sigma^{E}$. Both of these new sigma factors are produced before septum formation, but become active later.

$\sigma^{F}$ is the first factor to become active in the forespore compartment. It is inhibited by an anti-sigma factor that binds to it; in the forespore, an anti-anti-sigma factor removes the inhibitor. This reaction is controlled by a series of phosphorylation/dephosphorylation events. The initial determinant is a phosphatase (SpoIIE) that is an integral membrane protein, and which accumulates at the pole, with the result that its phosphatase domain becomes more concentrated in the forespore. It dephosphorylates, and thereby activates, SpoIIAA, which in turn displaces the anti-sigma factor SpoIIAB from the complex of SpoIIAB- $\sigma^{F}$. Release of $\sigma^{F}$ activates it.

Activation of $\sigma^{F}$ is the start of sporulation. Under the direction of $\sigma^{F}$, RNA polymerase transcribes the first set of sporulation genes instead of the vegetative genes it was previously transcribing. The replacement reaction probably affects only part of the RNA polymerase population, since $\sigma^{F}$ is produced only in small amounts. Some vegetative enzyme remains present during sporulation. The displaced $\sigma^{43}$ is not destroyed, but can be recovered from extracts of sporulating cells.

Two regulatory events follow from the activity of $\sigma^{F}$, as detailed in **Figure 9.43**. In the forespore itself, another factor, $\sigma^{G}$, is the product of one of the early sporulation genes. $\sigma^{G}$ is the factor that causes RNA polymerase to transcribe the late sporulation genes in the forespore. Another early sporulation gene product is responsible for communicating with the mother cell compartment. $\sigma^{F}$ activates SpoIIR, which is secreted from the forespore. It then activates the membrane-bound protein SpoIIGA to cleave the inactive precursor pro- $\sigma^{E}$ into the active factor $\sigma^{E}$ in the mother cell. (Any $\sigma^{E}$ that is produced in the forespore is degraded by forespore-specific functions.)

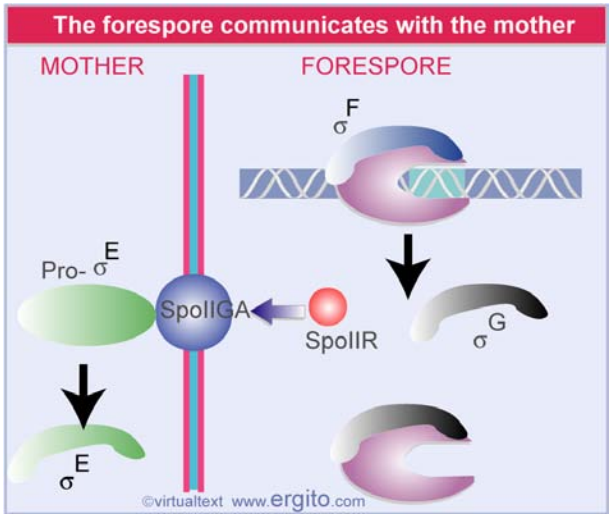


**Figure 9.43** $\sigma^F$ triggers synthesis of the next sigma factor in the forespore ( $\sigma^G$) and turns on SpoIIR which causes SpoIIGA to cleave pro- $\sigma^E$.

The cascade continues when $\sigma^E$ in turn is replaced by $\sigma^K$. (Actually the production of $\sigma^K$ is quite complex, because first its gene must be created by a recombination event!) This factor also is synthesized as an inactive precursor (pro- $\sigma^K$) that is activated by a protease. Once $\sigma^K$ has been activated, it displaces $\sigma^E$ and causes transcription of the late genes in the mother cell. The timing of these events in the two compartments are coordinated by further signals. The activity of $\sigma^E$ in the mother cell is necessary for activation of $\sigma^G$ in the forespore; and in turn the activity of $\sigma^G$ is required to generate a signal that is transmitted across the septum to activate $\sigma^K$.

Sporulation is thus controlled by two cascades, in which sigma factors in each compartment are successively activated, each directing the synthesis of a particular set of genes. **Figure 9.44** outlines how the two cascades are connected by the transmission of signals from one compartment to the other. As new sigma factors become active, old sigma factors are displaced, so that transitions in sigma factors turn genes off as well as on. The incorporation of each factor into RNA polymerase dictates when its set of target genes is expressed; and the amount of factor available influences the level of gene expression. More than one sigma factor may be active at any time, and the specificities of some of the sigma factors overlap. We do not know what is responsible for the ability of each sigma factor to replace its predecessor (for review see 77; 78; 81).

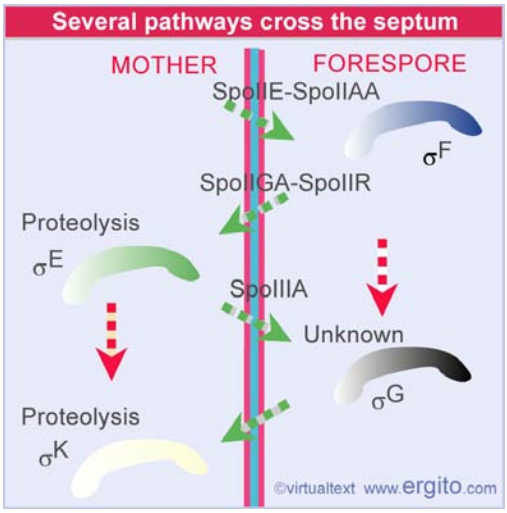


**Figure 9.44** The crisscross regulation of sporulation coordinates timing of events in the mother cell and forespore.

## Reviews

72. Losick, R. et al. (1986). *Genetics of endospore formation in B. subtilis*. Annu. Rev. Genet. 20, 625-669.

77. Losick, R. and Stragier, P. (1992). *Crisscross regulation of cell-type specific gene expression during development in B. subtilis.* Nature 355, 601-604.

78. Errington, J. (1993). *B. subtilis sporulation: regulation of gene expression and control of morphogenesis.* Microbiol. Rev. 57, 1-33.

80. Haldenwang, W. G. (1995). *The sigma factors of B. subtilis.* Microbiol. Rev. 59, 1-30.

81. Stragier, P and Losick, R. (1996). *Molecular genetics of sporulation in B. subtilis.* Annu. Rev. Genet. 30, 297-341.

## References

497. Haldenwang, W. G. and Losick, R. (1980). *A novel RNA polymerase sigma factor from B. subtilis*. Proc. Natl. Acad. Sci. USA 77, 7000-7004.

498. Haldenwang, W. G., Lang, N., and Losick, R. (1981). *A sporulation-induced sigma-like regulatory protein from B. subtilis*. Cell 23, 615-624.

TRANSCRIPTION

# 3.9.20 Bacterial RNA polymerase terminates at discrete sites

## Key Terms

A **terminator** is a sequence of DNA that causes RNA polymerase to terminate transcription.

**Antitermination** is a mechanism of transcriptional control in which termination is prevented at a specific terminator site, allowing RNA polymerase to read into the genes beyond it.

**Readthrough** at transcription or translation occurs when RNA polymerase or the ribosome, respectively, ignores a termination signal because of a mutation of the template or the behavior of an accessory factor.

## Key Concepts

- Termination may require both recognition of the terminator sequence in DNA and the formation of a hairpin structure in the RNA product.

Once RNA polymerase has started transcription, the enzyme moves along the template, synthesizing RNA, until it meets a **terminator** (*t*) sequence. At this point, the enzyme stops adding nucleotides to the growing RNA chain, releases the completed product, and dissociates from the DNA template. Termination requires that all hydrogen bonds holding the RNA-DNA hybrid together must be broken, after which the DNA duplex reforms.

It is difficult to define the termination point of an RNA molecule that has been synthesized in the living cell. It is always possible that the 3′ end of the molecule has been generated by *cleavage* of the primary transcript, and therefore does not represent the actual site at which RNA polymerase terminated.

The best identification of termination sites is provided by systems in which RNA polymerase terminates *in vitro*. Because the ability of the enzyme to terminate is strongly influenced by parameters such as the ionic strength, its termination at a particular point *in vitro* does not prove that this same point is a natural terminator. But we can identify authentic 3′ ends when the same end is generated *in vitro* and *in vivo*.

**Figure 9.45** summarizes the two types of feature found in bacterial terminators.

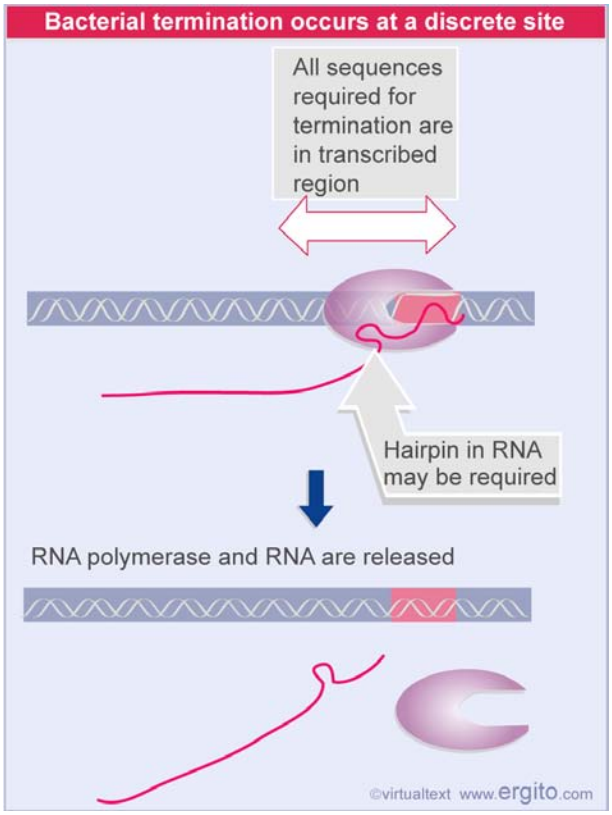


**Figure 9.45** The DNA sequences required for termination are located prior to the terminator sequence. Formation of a hairpin in the RNA may be necessary.

- Terminators in bacteria and their phages have been identified as sequences that are needed for the termination reaction (*in vitro* or *in vivo*). The sequences at prokaryotic terminators show no similarities beyond the point at which the last base is added to the RNA. The responsibility for termination lies with the *sequences already transcribed* by RNA polymerase. So termination relies on scrutiny of the template or product that the polymerase is currently transcribing (for review see 69; 73; 74).

- Many terminators require a hairpin to form in the secondary structure of the RNA being transcribed. *This indicates that termination depends on the RNA product and is not determined simply by scrutiny of the DNA sequence during transcription.*

Terminators vary widely in their efficiencies of termination. At some terminators, the termination event can be *prevented* by specific ancillary factors that interact with RNA polymerase. **Antitermination** causes the enzyme to continue transcription past the terminator sequence, an event called **readthrough** (the same term used to describe a ribosome's suppression of termination codons).

In approaching the termination event, we must regard it not simply as a mechanism for generating the 3′ end of the RNA molecule, but as an opportunity to control gene expression. So the stages when RNA polymerase associates with DNA (initiation) or

dissociates from it (termination) both are subject to specific control. There are interesting parallels between the systems employed in initiation and termination. Both require breaking of hydrogen bonds (initial melting of DNA at initiation, RNA-DNA dissociation at termination); and both require additional proteins to interact with the core enzyme. In fact, they are accomplished by alternative forms of the polymerase. However, whereas initiation relies solely upon the interaction between RNA polymerase and duplex DNA, the termination event involves recognition of signals in the transcript by RNA polymerase or by ancillary factors as well as the recognition of sequences in DNA.

## Reviews

69. Adhya, S. and Gottesman, M. (1978). *Control of transcription termination.* Annu. Rev. Biochem. 47, 967-996.

73. Platt, T. (1986). *Transcription termination and the regulation of gene expression.* Annu. Rev. Biochem. 55, 339-372.

74. Friedman, D. I., Imperiale, M. J., and Adhya, S. L. (1987). *RNA 3′ end formation in the control of gene expression.* Annu. Rev. Genet. 21, 453-488.

TRANSCRIPTION

# 3.9.21 There are two types of terminators in *E. coli*

## Key Terms

**Intrinsic terminators** are able to terminate transcription by bacterial RNA polymerase in the absence of any additional factors.

**Rho-dependent** terminators are sequences that terminate transcription by bacterial RNA polymerase in the presence of the rho factor.

**Rho factor** is a protein involved in assisting *E. coli* RNA polymerase to terminate transcription at certain terminators (called rho-dependent terminators).

## Key Concepts

- Intrinsic terminators consist of a G·C-rich hairpin in the RNA product followed by a U-rich region in which termination occurs.

Terminators are distinguished in *E. coli* according to whether RNA polymerase requires any additional factors to terminate *in vitro:*

- Core enzyme can terminate *in vitro* at certain sites in the absence of any other factor. These sites are called **intrinsic terminators**.
- **Rho-dependent** terminators are defined by the need for addition of **rho factor** ( ρ) *in vitro;* and mutations show that the factor is involved in termination *in vivo.*

Intrinsic terminators have the two structural features evident in **Figure 9.46**: a hairpin in the secondary structure; and a region that is rich in U residues at the very end of the unit. Both features are needed for termination. The hairpin usually contains a G·C-rich region near the base of the stem. The typical distance between the hairpin and the U-rich region is 7-9 bases. There are ~1100 sequences in the *E. coli* genome that fit these criteria, suggesting that about half of the genes have intrinsic terminators (2830).

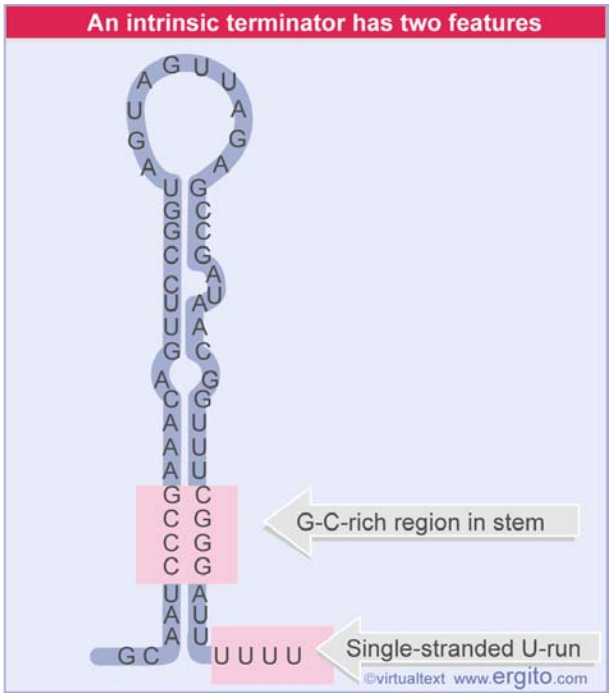


**Figure 9.46** Intrinsic terminators include palindromic regions that form hairpins varying in length from 7-20 bp. The stem-loop structure includes a G-C-rich region and is followed by a run of U residues.

Point mutations that prevent termination occur within the stem region of the hairpin. What is the effect of a hairpin on transcription? Probably all hairpins that form in the RNA product cause the polymerase to slow (and perhaps to pause) in RNA synthesis.

Pausing creates an opportunity for termination to occur. Pausing occurs at sites that resemble terminators but have an increased separation (typically 10-11 bases) between the hairpin and the U-run. But if the pause site does not correspond to a terminator, usually the enzyme moves on again to continue transcription. The length of the pause varies, but at a typical terminator lasts ~60 seconds.

A downstream U-rich region destabilizes the RNA-DNA hybrid when RNA polymerase pauses at the hairpin (for review see 2831). The rU·dA RNA-DNA hybrid has an unusually weak base-paired structure; it requires the least energy of any RNA-DNA hybrid to break the association between the two strands. When the polymerase pauses, the RNA-DNA hybrid unravels from the weakly bonded rU·dA terminal region. Often the actual termination event takes place at any one of several positions toward or at the end of the U-rich region, as though the enzyme "stutters" during termination. The U-rich region in RNA corresponds to an A·T-rich region in DNA, so we see that A·T-rich regions are important in intrinsic termination as well as initiation.

Both the sequence of the hairpin and the length of the U-run influence the efficiency of termination. However, termination efficiency *in vitro* varies from 2-90%, and does not correlate in any simple way with the constitution of the hairpin or the number of U residues in the U-rich region (2833). The hairpin and U-region are therefore

necessary, but not sufficient, and additional parameters influence the interaction with RNA polymerase. In particular, the sequences both upstream and downstream of the intrinsic terminator influence its efficiency (2832).

Less is known about the signals and ancillary factors involved in termination for eukaryotic polymerases. Each class of polymerase uses a different mechanism (see *Molecular Biology 5.24 RNA splicing and processing*).

*Last updated on 8-8-2002*

## Reviews

2831. von Hippel, P. H. (1998). *An integrated model of the transcription complex in elongation, termination, and editing.* Science 281, 660-665.

## References

2830. Lesnik, E. A., Sampath, R., Levene, H. B., Henderson, T. J., McNeil,J.A., and Ecker, D. J. (2001). *Prediction of rho-independent transcriptional terminators in E. coli,*. Nucleic Acids Res. 29, 3583-3594.

2832. Lee, D. N., Phung, L., Stewart, J., and Landick, R. (1990). *Transcription pausing by E. coli RNA polymerase is modulated by downstream DNA sequences.* J. Biol. Chem. 265, 15145-15153.

2833. Reynolds, R., Bermadez-Cruz, R. M., and Chamberlin, M. J. (1992). *Parameters affecting transcription termination by E. coli RNA polymerase. I. Analysis of 13 rho-independent terminators.* J. Mol. Biol. 224, 31-51.

TRANSCRIPTION

# 3.9.22 How does rho factor work?

## Key Terms

**Polarity** refers to the effect of a mutation in one gene in influencing the expression (at transcription or translation) of subsequent genes in the same transcription unit.

## Key Concepts

- Rho factor is a terminator protein that binds to a *rut* site on nascent RNA and translocates along the RNA to release it from the RNA-DNA hybrid structure at the RNA polymerase.

Rho factor is an essential protein in *E. coli*. It functions solely at the stage of termination. It is a ~275 kD hexamer of identical subunits (for review see 2836). The subunit has an N-terminal RNA-binding domain and a C-terminal ATP hydrolysis domain. Rho is a member of the family of hexameric ATP-dependent helicases that function by passing nucleic acid through the hole in the middle of the hexamer formed from the RNA-binding domains of the subunits. Rho functions as an ancillary factor for RNA polymerase; typically its maximum activity *in vitro* is displayed when it is present at ~10% of the concentration of the RNA polymerase (507).

Rho-dependent terminators account for about half of *E. coli* terminators. They were discovered in phage genomes, where they have been most fully characterized. The sequences required for rho-dependent termination (sometimes called **rut** sites) are 50-90 bases long and lie upstream of the actual termination site. Their common feature is that the RNA is rich in C residues and poor in G residues. An example is given in **Figure 9.47**; C is by far the most common base (41%) and G is the least common base (14%). As a general rule the efficiency of a *rut* site increases with the length of the C-rich/G-poor region.


**Figure 9.47** A *rut* site has a sequence rich in C and poor in G preceding the actual site(s) of termination. The sequence correspodns to the 3′ end of the RNA.

**Figure 9.48** shows that the structure of rho gives some indications of how it functions. It winds RNA from the 3&#; end around the exterior of the N-terminal domains, and pushes the 5&#; end of the bound region into the interior, where it is bound by a secondary RNA-binding domain in the C-terminal domains (4189). The

initial form of rho is a gapped ring, but binding of the RNA converts it to a closed ring.

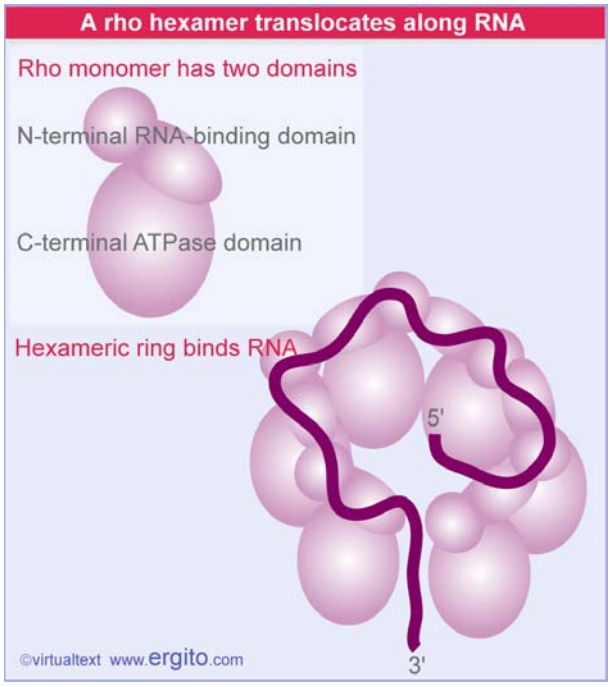


**Figure 9.48** Rho has an N-terminal RNA-binding domain and a C-terminal ATPase domain. A hexamer in the form of a gapped ring binds RNA along the exterior of the N-terminal domains. The 5&#; end of the RNA is bound by a secondary binding site in the interior of the hexamer.

**Figure 9.49** shows that after binding to the *rut* site, it uses its helicase activity, driven by ATP hydrolysis, to translocate along RNA until it reaches the RNA-DNA hybrid stretch in RNA polymerase. Then it unwinds the duplex structure.(2835, 2834; for review see 2831). We do not know whether this action is sufficient to release the transcript or whether rho also interacts with RNA polymerase to help release RNA (for review see 79).

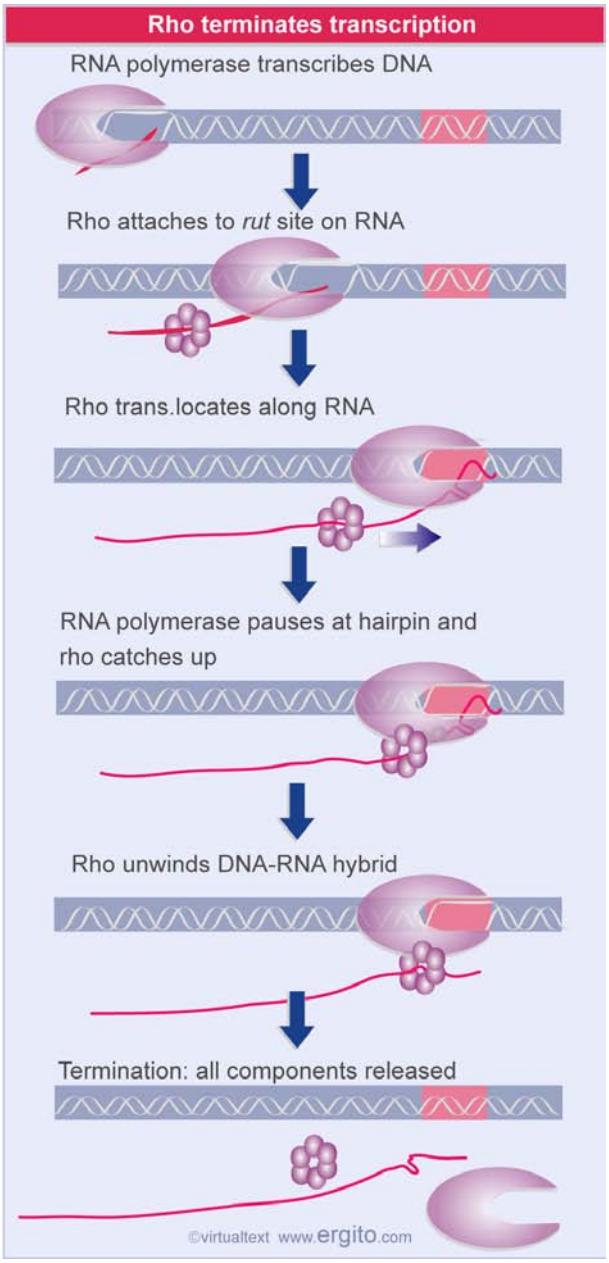


**Figure 9.49** Rho factor binds to RNA at a *rut* site and translocates along RNA until it reaches the RNA-DNA hybrid in RNA polymerase, where it releases the RNA from the DNA.

Rho needs to translocate along RNA from the *rut* site to the actual point of termination. This requires the factor to move faster than RNA polymerase. The enzyme pauses when it reaches a terminator, and termination occurs if rho catches it there. Pausing is therefore important in rho-dependent termination, just as in intrinsic termination, because it gives time for the other necessary events to occur.

The idea that rho moves along RNA leads to an important prediction about the relationship between transcription and translation. Rho must first have access to a binding sequence on RNA, and then must be able to move along the RNA. Either or

both of these conditions may be prevented if ribosomes are translating an RNA. So the ability of rho factor to reach RNA polymerase at a terminator depends on what is happening in translation.

This model explains a puzzling phenomenon. In some cases, a nonsense mutation in one gene of a transcription unit prevents the expression of subsequent genes in the unit. This effect is called **polarity**. A common cause is the absence of the mRNA corresponding to the subsequent (distal) parts of the unit.

Suppose that there are rho-dependent terminators *within* the transcription unit, that is, before the terminator that *usually* is used. The consequences are illustrated in **Figure 9.50**. Normally these earlier terminators are not used, because the ribosomes prevent rho from reaching RNA polymerase. But a nonsense mutation releases the ribosomes, so that rho is free to attach to and/or move along the mRNA, enabling it to act on RNA polymerase at the terminator. As a result, the enzyme is released, and the distal regions of the transcription unit are never expressed. (Why should there be internal terminators? Perhaps they are simply sequences that by coincidence mimic the usual rho-dependent terminator.) Some stable RNAs that have extensive secondary structure are preserved from polar effects, presumably because the structure impedes rho attachment or movement.

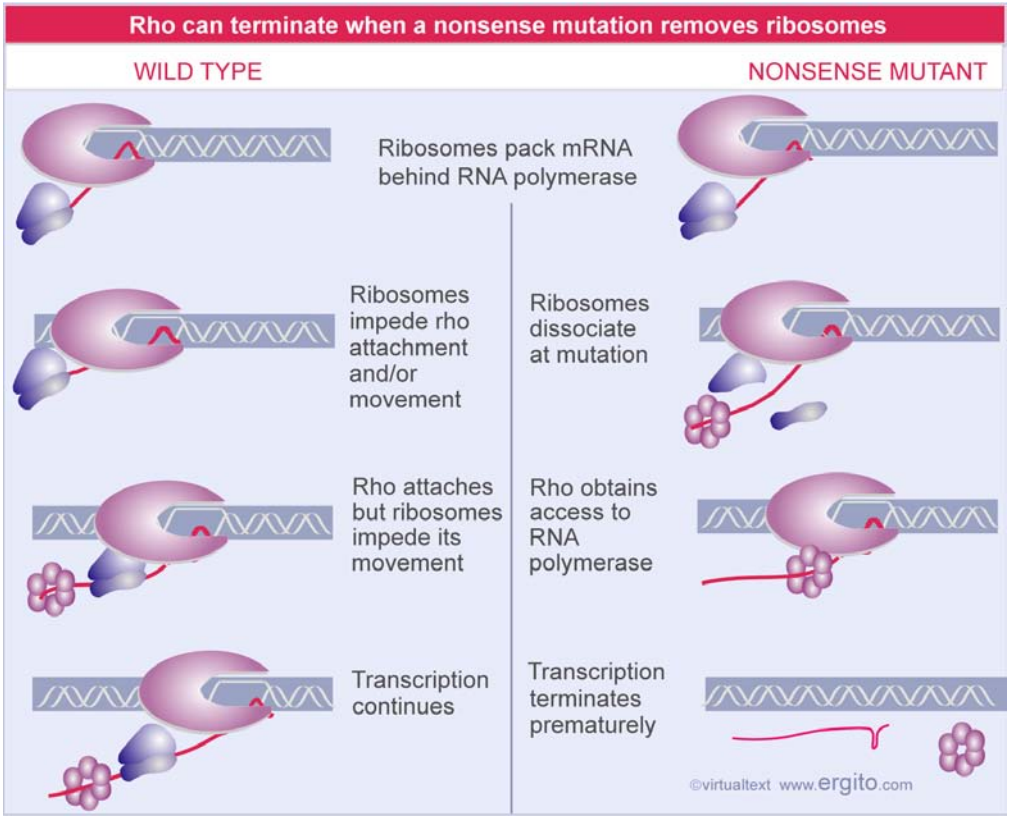


**Figure 9.50** The action of rho factor may create a link between transcription and translation when a rho-dependent terminator lies soon after a nonsense mutation.

*rho* mutations show wide variations in their influence on termination. The basic nature of the effect is a failure to terminate. But the magnitude of the failure, as seen

in the percent of readthrough *in vivo*, depends on the particular target locus. Similarly, the need for rho factor *in vitro* is variable. Some (rho-dependent) terminators require relatively high concentrations of rho, while others function just as well at lower levels. This suggests that different terminators require different levels of rho factor for termination, and therefore respond differently to the residual levels of rho factor in the mutants (*rho* mutants are usually leaky).

Some *rho* mutations can be suppressed by mutations in other genes. This approach provides an excellent way to identify proteins that interact with rho. The β subunit of RNA polymerase is implicated by two types of mutation. First, mutations in the *rpoB* gene can reduce termination at a rho-dependent site. Second, mutations in *rpoB* can restore the ability to terminate transcription at rho-dependent sites in *rho* mutant bacteria. However, we do not know what function the interaction plays.

*Last updated on January 20, 2004*

## Reviews

79. Das, A. (1993). *Control of transcription termination by RNA-binding proteins.* Annu. Rev. Biochem. 62, 893-930.

2831. von Hippel, P. H. (1998). *An integrated model of the transcription complex in elongation, termination, and editing.* Science 281, 660-665.

2836. Richardson, J. P. (1996). *Structural organization of transcription termination factor Rho.* J. Biol. Chem. 271, 1251-1254.

## References

507. Roberts, J. W. (1969). *Termination factor for RNA synthesis*. Nature 224, 1168-1174.

2834. Geiselmann, J., Wang, Y., Seifried, S. E., and von Hippel, P. H. (1993). *A physical model for the translocation and helicase activities of E. coli transcription termination protein Rho.* Proc. Natl. Acad. Sci. USA 90, 7754-7758.

2835. Brennan, C. A., Dombroski, A. J., and Platt, T. (1987). *Transcription termination factor rho is an RNA-DNA helicase.* Cell 48, 945-952.

4189. Skordalakes, E. and Berger, J. M. (2003). *Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading.* Cell 114, 135-146.

TRANSCRIPTION

# 3.9.23 Antitermination is a regulatory event

## Key Terms

**Antitermination** is a mechanism of transcriptional control in which termination is prevented at a specific terminator site, allowing RNA polymerase to read into the genes beyond it.

**Antitermination proteins** allow RNA polymerase to transcribe through certain terminator sites.

**Immediate early** phage genes in phage lambda are equivalent to the early class of other phages. They are transcribed immediately upon infection by the host RNA polymerase.

**Delayed early** genes in phage lambda are equivalent to the middle genes of other phages. They cannot be transcribed until regulator protein(s) coded by the immediate early genes have been synthesized.

## Key Concepts

- Termination is prevented when antitermination proteins act on RNA polymerase to cause it to readthrough a specific terminator or terminators.
- Phage lambda has two antitermination proteins, pN and pQ, that act on different transcription units.

**Antitermination** is used as a control mechanism in both phage regulatory circuits and bacterial operons. **Figure 9.51** shows that antitermination controls the ability of the enzyme to read past a terminator into genes lying beyond. In the example shown in the figure, the default pathway is for RNA polymerase to terminate at the end of region 1. But antitermination allows it to continue transcription through region 2. Because the promoter does not change, both situations produce an RNA with the same 5′ sequences; the difference is that after antitermination the RNA is extended to include new sequences at the 3′ end.

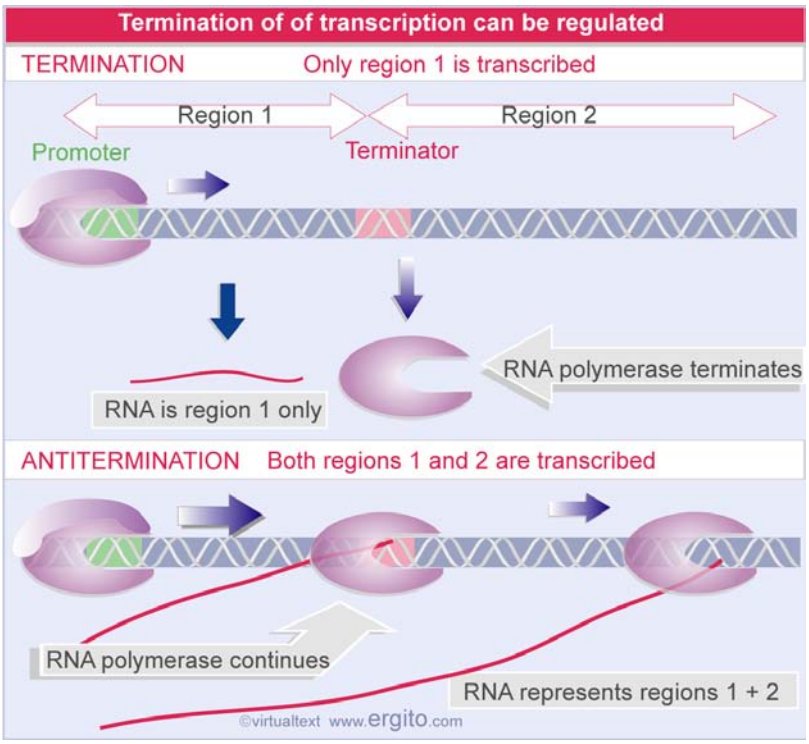


**Figure 9.51** Antitermination can be used to control transcription by determining whether RNA polymerase terminates or reads through a particular terminator into the following region.

Antitermination was discovered in bacteriophage infections. A common feature in the control of phage infection is that very few of the phage genes (the "early" genes) can be transcribed by the bacterial host RNA polymerase. Among these genes, however, are regulator(s) whose product(s) allow the next set of phage genes to be expressed (see *Molecular Biology 3.12.4 Two types of regulatory event control the lytic cascade*). One of these types of regulator is an **antitermination protein**. **Figure 9.52** shows that it enables RNA polymerase to read through a terminator, extending the RNA transcript. In the absence of the antitermination protein, RNA polymerase terminates at the terminator (top panel). When the antitermination protein is present, it continues past the terminator (middle panel).

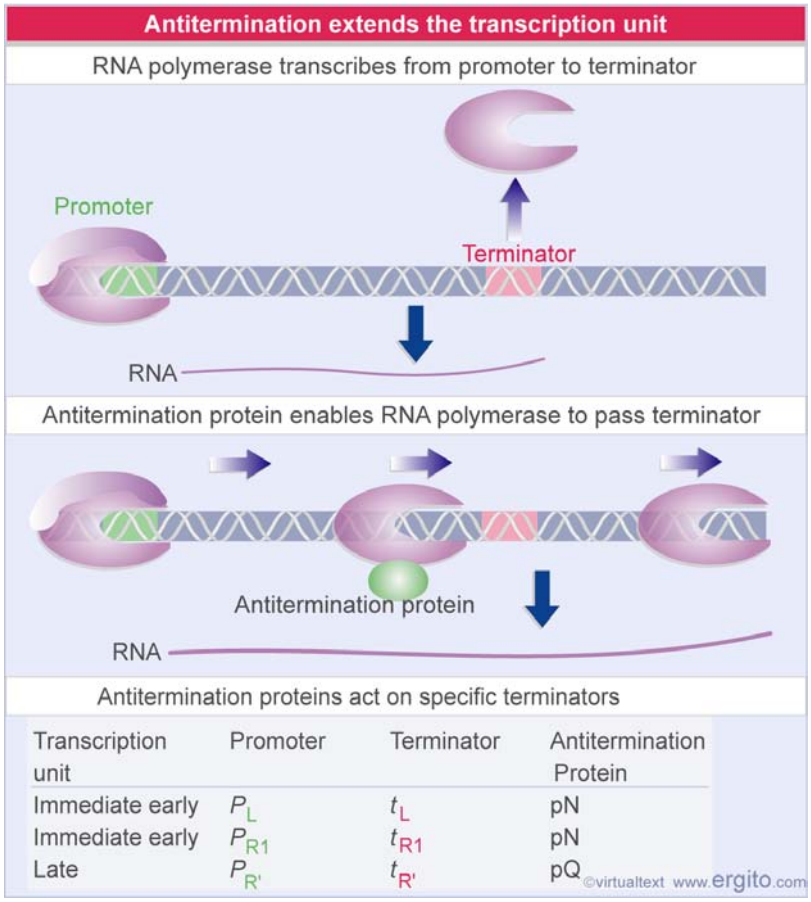


**Figure 9.52** An antitermination protein can act on RNA polymerase to enable it to readthrough a specific terminator.

The best characterized example of antitermination is provided by phage lambda, with which the phenomenon was discovered. It is used at two stages of phage expression. The antitermination protein produced at each stage is specific for the particular transcription units that are expressed at that stage, as summarized in the bottom panel of **Figure 9.52**.

The host RNA polymerase initially transcribes two genes, which are called the **immediate early** genes. The transition to the next stage of expression is controlled by preventing termination at the ends of the immediate early genes, with the result that the **delayed early** genes are expressed. (We discuss the overall regulation of lambda development in *Molecular Biology 3.12 Phage strategies*.)

The regulator gene that controls the switch from immediate early to delayed early expression is identified by mutations in lambda gene *N* that can transcribe *only* the immediate early genes; they proceed no further into the infective cycle. There are two transcription units of immediate early genes (transcribed from the promoters $P_L$ and $P_R$). Transcription by *E. coli* RNA polymerase itself stops at the terminators at the ends of these transcription units ($t_{L1}$ and $t_{R1}$, respectively.) Both terminators depend on rho; in fact, these were the terminators with which rho was originally identified. The situation is changed by expression of the *N* gene. The product pN is an antitermination protein that acts on both of the immediate early transcription units,

and allows RNA polymerase to read through the terminators into the delayed early genes beyond them.

Like other phages, still another control is needed to express the late genes that code for the components of the phage particle. This switch is regulated by gene *Q*, itself one of the delayed early genes. Its product, pQ, is another antitermination protein, one that specifically allows RNA polymerase initiating at another site, the late promoter $P_{R'}$, to read through a terminator that lies between it and the late genes.

The different specificities of pN and pQ establish an important general principle: *RNA polymerase interacts with transcription units in such a way that an ancillary factor can sponsor antitermination specifically for some transcripts.* Termination can be controlled with the same sort of precision as initiation.

TRANSCRIPTION

# 3.9.24 Antitermination requires sites that are independent of the terminators

## Key Concepts

- The site where an antiterminator protein acts is upstream of the terminator site in the transcription unit.
- The location of the antiterminator site varies in different cases, and can be in the promoter or within the transcription unit.

What sites are involved in controlling the specificity of antitermination? The antitermination activity of pN is highly specific, but *the antitermination event is not determined by the terminators* $t_{L1}$ *and* $t_{R1}$*; the recognition site needed for antitermination lies upstream in the transcription unit, that is, at a different place from the terminator site at which the action eventually is accomplished*. This conclusion establishes a general principle. When we know the site on DNA at which some protein exercises its effect, we cannot assume that this coincides with the DNA sequence that it initially recognizes. They may be separate. **Figure 9.53** shows the locations of the sites required for antitermination in phage lambda.

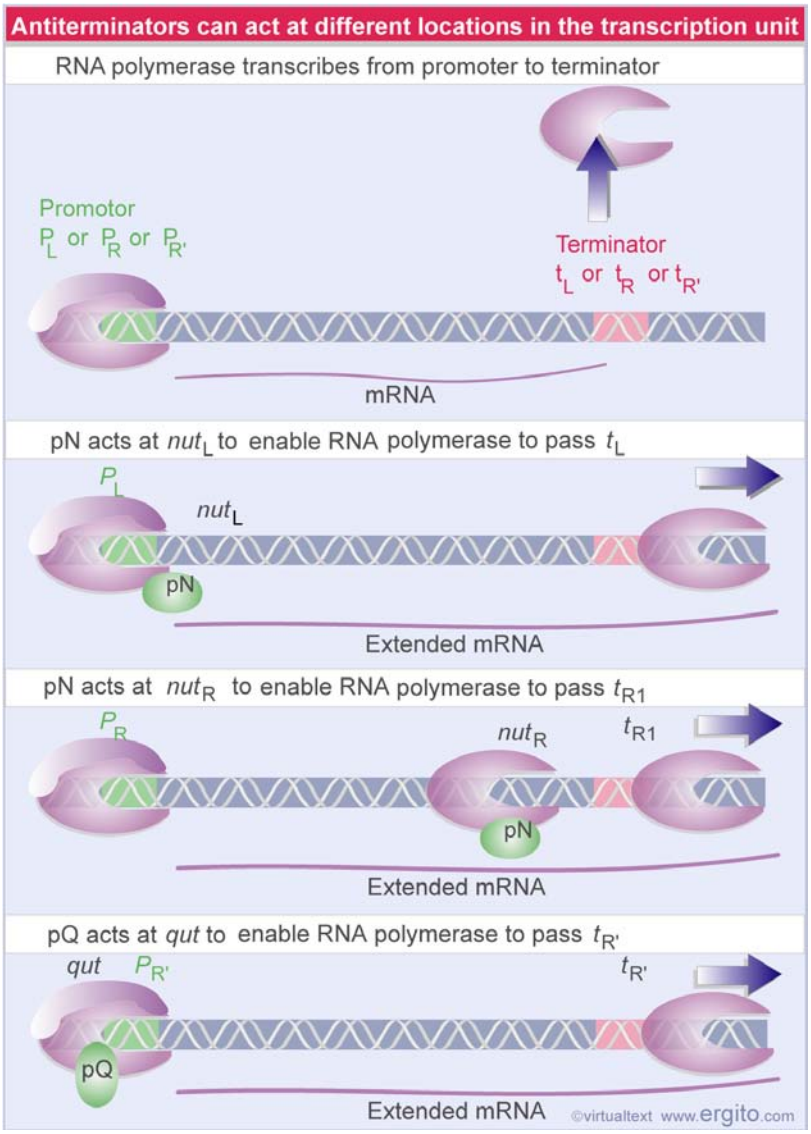


**Figure 9.53** Host RNA polymerase transcribes lambda genes and terminates at *t* sites. pN allows it to read through terminators in the L and R1 units; pQ allows it to read through the R′ terminator. The sites at which pN acts (*nut*) and at which pQ acts (*qut*) are located at different relative positions in the transcription units.

The recognition sites required for pN action are called *nut* (for *N ut*ilization). The sites responsible for determining leftward and rightward antitermination are described as *nutL* and *nutR*, respectively. Mapping of *nut* mutations locates *nutL* between the startpoint of $P_L$ and the beginning of the *N* coding region. By contrast, *nutR* lies between the end of the *cro* gene and $t_{L1}$. This means that the two *nut* sites lie in different positions relative to the organization of their transcription units. Whereas *nutL* is near the promoter, *nutR* is near to the terminator. (*qut* is different yet again, and lies within the promoter.)

How does antitermination occur? When pN recognizes the *nut* site, it must act on RNA polymerase to ensure that the enzyme can no longer respond to the terminator. The variable locations of the *nut* sites indicate that this event is linked neither to

initiation nor to termination, but can occur to RNA polymerase as it elongates the RNA chain past the *nut* site. As illustrated in **Figure 9.54**, the polymerase then becomes a juggernaut that continues past the terminator, heedless of its signal. (This reaction involves antitermination at rho-dependent terminators, but pN also suppresses termination at intrinsic terminators.)

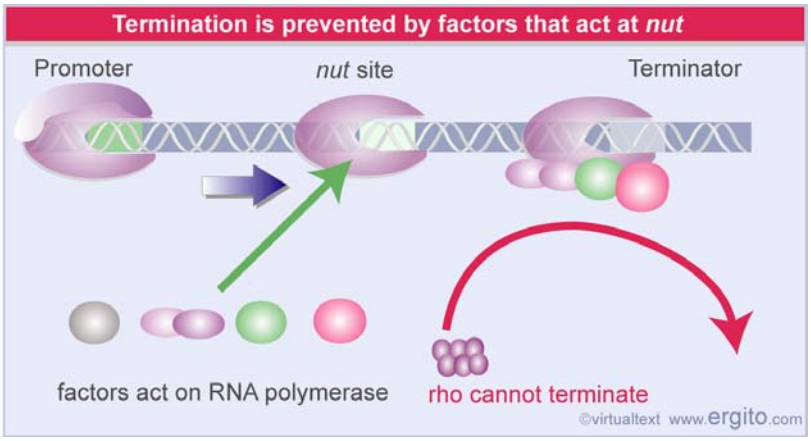


**Figure 9.54** Ancillary factors bind to RNA polymerase as it passes the *nut* site. They prevent rho from causing termination when the polymerase reaches the terminator.

Is the ability of pN to recognize a short sequence within the transcription unit an example of a more widely used mechanism for antitermination? Other phages, related to lambda, have different *N* genes and different antitermination specificities. The region of the phage genome in which the *nut* sites lie has a different sequence in each of these phages, and each phage must therefore have characteristic *nut* sites recognized specifically by its own pN. Each of these pN products must have the same general ability to interact with the transcription apparatus in an antitermination capacity, but has a different specificity for the sequence of DNA that activates the mechanism.

*This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.3.9.24*

TRANSCRIPTION

# 3.9.25 Termination and anti-termination factors interact with RNA polymerase

## Key Concepts

- Several bacterial proteins are required for lambda pN to interact with RNA polymerase.
- These proteins are also involved in antitermination in the *rrn* operons of the host bacterium.
- The lambda antiterminator pQ has a different mode of interaction that involves binding to DNA at the promoter.

Termination and antitermination are closely connected, and involve bacterial and phage proteins that interact with RNA polymerase. Several proteins concerned with termination have been identified by isolating mutants of *E. coli* in which pN is ineffective. Several of these mutations lie in the *rpoB* gene. This argues that pN (like rho factor) interacts with the β subunit of the core enzyme. Other *E. coli* mutations that prevent pN function identify the *nus* loci: *nusA*, *nusB*, *nusE*, and *nusG*. (The term "*nus*" is an acronym for *N u*tilization *s*ubstance.)

A lambda *nut* site consists of two sequence elements, called *boxA* and *boxB* (1776). Sequence elements related to *boxA* are also found in bacterial operons. *boxA* is required for binding bacterial proteins that are necessary for antitermination in both phage and bacterial operons. *boxB* is specific to the phage genome, and mutations in *boxB* abolish the ability of pN to cause antitermination.

The *nus* loci code for proteins that form part of the transcription apparatus, but that are not isolated with the RNA polymerase enzyme. The *nusA*, *nusB*, and *nusG* functions are concerned solely with the termination of transcription. *nusE* codes for ribosomal protein S10; the relationship between its location in the 30S subunit and its function in termination is not clear. The Nus proteins bind to RNA polymerase at the *nut* site, as summarized in **Figure 9.55**.

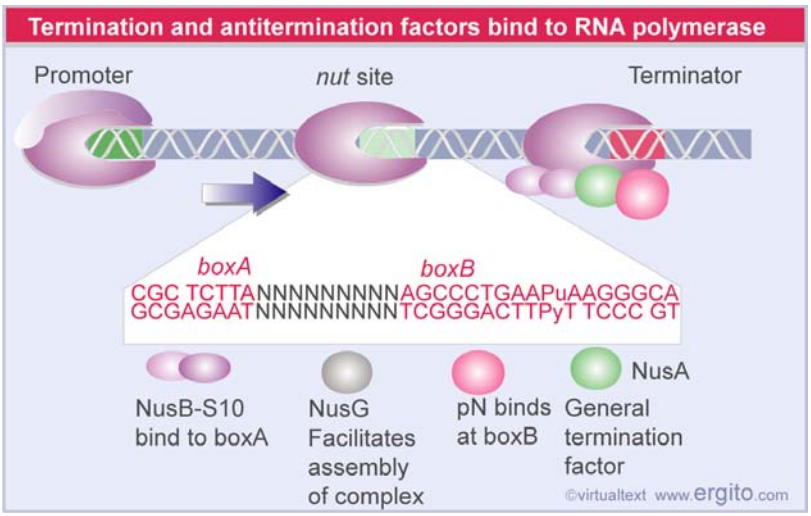


**Figure 9.55** Ancillary factors bind to RNA polymerase as it passes certain sites. The *nut* site consists of two sequences. NusB-S10 join core enzyme as it passes *boxA*. Then NusA and pN protein bind as polymerase passes *boxB*. The presence of pN allows the enzyme to read through the terminator, producing a joint mRNA that contains immediate early sequences joined to delayed early sequences.

NusA is a general transcription factor that increases the efficiency of termination, probably by enhancing RNA polymerase's tendency to pause at terminators (and indeed at other regions of secondary structure; see below). NusB and S10 form a dimer that binds specifically to RNA containing a *boxA* sequence. NusG may be concerned with the general assembly of all the Nus factors into a complex with RNA polymerase.

Intrinsic and rho-dependent terminators have different requirements for the Nus factors. NusA is required for termination at intrinsic terminators, and the reaction can be prevented by pN. At rho-dependent terminators, all 4 Nus proteins are required, and again pN alone can inhibit the reaction (for review see 102). The common feature of pN at both types of terminator is to prevent the role of NusA in termination (1779). Binding of pN to NusA inhibits the ability of NusA to bind RNA, which is necessary for termination.

Antitermination occurs in the *rrn* (rRNA) operons of *E. coli*, and involves the same *nus* functions. The leader regions of the *rrn* operons contain *boxA* sequences; NusB-S10 dimers recognize these sequences and bind to RNA polymerase as it elongates past *boxA*. This changes the properties of RNA polymerase in such a way that it can now read through rho-dependent terminators that are present within the transcription unit.

The *boxA* sequence of lambda RNA does not bind NusB-S10, and is probably enabled to do so by the presence of NusA and pN; the *boxB* sequence could be required to stabilize the reaction. So variations in *boxA* sequences may determine which particular set of factors is required for antitermination. The consequences are the same: when RNA polymerase passes the *nut* site, it is modified by addition of appropriate factors, and fails to terminate when it subsequently encounters the terminator sites.

Antitermination in lambda requires pN to bind to RNA polymerase in a manner that depends on the sequence of the transcription unit. Does pN recognize the *boxB* site in DNA or in the RNA transcript? It does not bind independently to either type of sequence, but does bind to a transcription complex when core enzyme passes the *boxB* site. pN has separate domains that recognize the *boxB* RNA sequence and the NusA protein (1777; 1778). After joining the transcription complex, pN remains associated with the core enzyme, in effect becoming an additional subunit whose presence changes recognition of terminators. It is possible that pN in fact continues to bind to both the *boxB* RNA sequence and to RNA polymerase, maintaining a loop in the RNA; thus the role of *boxB* RNA would partly be to tether pN in the vicinity, effectively increasing its local concentration.

pQ, which prevents termination later in phage infection by acting at *qut*, has a different mode of action. The *qut* sequence lies at the start of the late transcription unit. The upstream part of *qut* lies within the promoter, while the downstream part lies at the beginning of the transcribed region. This implies that pQ action involves recognition of DNA, and implies that its mechanism of action, at least concerning the initial binding to the complex, must be different from that of pN. pQ interacts with the holoenzyme during the initiation phase. In fact, $\sigma^{70}$ is required for the interaction with pQ. This reinforces the view of RNA polymerase as an interactive structure in which conformational changes induced at one phase may affect its activity at a later phase.

The basic action of pQ is to interfere with pausing; and once pQ has acted upon RNA polymerase, the enzyme shows much reduced pausing at all sites, including rho-dependent and intrinsic terminators. So pQ does not act directly on termination *per se*, but instead allows the enzyme to pass the terminator more quickly, thus depriving the core polymerase and/or accessory factor of the opportunity to cause termination.

The general principle is that RNA polymerase may exist in forms that are competent to undertake particular stages of transcription, and its activities at these stages can be changed only by modifying the appropriate form. So substitutions of sigma factors may change one initiation-competent form into another; and additions of Nus factors may change the properties of termination-competent forms.

Termination seems to be closely connected with the mode of elongation. In its basic transcription mode, core polymerase is subject to many pauses during elongation; and pausing at a terminator site is the prerequisite for termination to occur. Under the influence of factors such as NusA, pausing becomes extended, increasing the efficiency of termination; while under the influence of pN or pQ, pausing is abbreviated, decreasing the efficiency of termination. Because recognition sites for these factors are found only in certain transcription units, pausing and consequently termination are altered only in those units.

*Last updated on 8-8-2002*

## Reviews

102. Greenblatt, J., Nodwell, J. R., and Mason, S. W. (1993). *Transcriptional antitermination*. Nature 364, 401-406.

## References

1776. Olson, E. R., Flamm, E. L., and Friedman, D. I. (1982). *Analysis of nutR: a region of phage lambda required for antitermination of transcription.* Cell 31, 61-70.

1777. Legault, P., Li, J., Mogridge, J., Kay, L. E., and Greenblatt, J. (1998). *NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif.* Cell 93, 289-299.

1778. Mogridge, J., Mah, J., and Greenblatt, J. (1995). *A protein-RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the lambda N protein.* Genes Dev. 9, 2831-2845.

1779. Mah, T. F., Kuznedelov, K., Mushegian, A., Severinov, K., and Greenblatt, J. (2000). *The alpha subunit ofE. coli RNA polymerase activates RNA binding by NusA.* Genes Dev. 14, 2664-2675.

## TRANSCRIPTION

# 3.9.26 Summary

A transcription unit comprises the DNA between a promoter, where transcription initiates, and a terminator, where it ends. One strand of the DNA in this region serves as a template for synthesis of a complementary strand of RNA. The RNA-DNA hybrid region is short and transient, as the transcription "bubble" moves along DNA. The RNA polymerase holoenzyme that synthesizes bacterial RNA can be separated into two components. Core enzyme is a multimer of structure $\alpha_2 \beta \beta'$ that is responsible for elongating the RNA chain. Sigma factor ( $\sigma$ ) is a single subunit that is required at the stage of initiation for recognizing the promoter.

Core enzyme has a general affinity for DNA. The addition of sigma factor reduces the affinity of the enzyme for nonspecific binding to DNA, but increases its affinity for promoters. The rate at which RNA polymerase finds its promoters is too great to be accounted for by diffusion and random contacts with DNA; direct exchange of DNA sequences held by the enzyme may be involved.

Bacterial promoters are identified by two short conserved sequences centered at –35 and –10 relative to the startpoint. Most promoters have sequences that are well related to the consensus sequences at these sites. The distance separating the consensus sequences is 16-18 bp. RNA polymerase initially "touches down" at the –35 sequence and then extends its contacts over the –10 region. The enzyme covers ~77 bp of DNA. The initial "closed" binary complex is converted to an "open" binary complex by melting of a sequence of ~12 bp that extends from the –10 region to the startpoint. The A·T-rich base pair composition of the –10 sequence may be important for the melting reaction.

The binary complex is converted to a ternary complex by the incorporation of ribonucleotide precursors. There are multiple cycles of abortive initiation, during which RNA polymerase synthesizes and releases very short RNA chains without moving from the promoter. At the end of this stage, there is a change in structure, and the core enzyme contracts to cover ~50 bp. Sigma factor is either released (30% of cases) or changes its form of association with the core enzyme. Then core enzyme moves along DNA, synthesizing RNA. A locally unwound region of DNA moves with the enzyme. The enzyme contracts further in size to cover only 30-40 bp when the nascent chain has reached 15-20 nucleotides; then it continues to the end of the transcription unit.

The "strength" of a promoter describes the frequency at which RNA polymerase initiates transcription; it is related to the closeness with which its –35 and –10 sequences conform to the ideal consensus sequences, but is influenced also by the sequences immediately downstream of the startpoint. Negative supercoiling increases the strength of certain promoters. Transcription generates positive supercoils ahead of RNA polymerase and leaves negative supercoils behind the enzyme.

The core enzyme can be directed to recognize promoters with different consensus sequences by alternative sigma factors. In *E. coli*, these sigma factors are activated by adverse conditions, such as heat shock or nitrogen starvation. *B. subtilis* contains

a single major sigma factor with the same specificity as the *E. coli* sigma factor, and also contains a variety of minor sigma factors. Another series of factors is activated when sporulation is initiated; sporulation is regulated by two cascades in which sigma factor replacements occur in the forespore and mother cell. A cascade for regulating transcription by substitution of sigma factors is also used by phage SPO1.

The geometry of RNA polymerase-promoter recognition is similar for holoenzymes containing all sigma factors (except $\sigma^{54}$). Each sigma factor causes RNA polymerase to initiate transcription at a promoter that conforms to a particular consensus at –35 and –10. Direct contacts between sigma and DNA at these sites have been demonstrated for *E. coli* $\sigma^{70}$. The $\sigma^{70}$ factor of *E. coli* has an N-terminal autoinhibitory domain that prevents the DNA-binding regions from recognizing DNA. The autoinhibitory region is displaced by DNA when the holoenzyme forms an open complex.

Bacterial RNA polymerase terminates transcription at two types of sites. Intrinsic terminators contain a G·C-rich hairpin followed by a U-rich region. They are recognized *in vitro* by core enzyme alone. Rho-dependent terminators require rho factor both *in vitro* and *in vivo*; rho binds to *rut* sites that are rich in C and poor in G residues and that precede the actual site of termination. Rho is a hexameric ATP-dependent helicase activity that translocates along the RNA until it reaches the RNA-DNA hybrid region in the transcription bubble of RNA polymerase, where it dissociates the RNA from DNA. In both types of termination, pausing by RNA polymerase is important in order to allow time for the actual termination event to occur.

The Nus factors are required for termination. NusA is required for intrinsic terminators, and in addition NusB-S10 is required for rho-dependent terminators. The NusB-S10 dimer recognizes the *boxA* sequence of a *nut* site in the elongating RNA; NusA joins subsequently.

Antitermination is used by some phages to regulate progression from one stage of gene expression to the next. The lambda gene *N* codes for an antitermination protein (pN) that is necessary to allow RNA polymerase to read through the terminators located at the ends of the immediate early genes. Another antitermination protein, pQ, is required later in phage infection. pN and pQ act on RNA polymerase as it passes specific sites (*nut* and *qut*, respectively). These sites are located at different relative positions in their respective transcription units. pN recognizes RNA polymerase carrying NusA when the enzyme passes the sequence *boxB*. pN then binds to the complex and prevents termination by antagonizing the action of NusA when the polymerase reaches the rho-dependent terminator.

*Last updated on 9-29-2003*