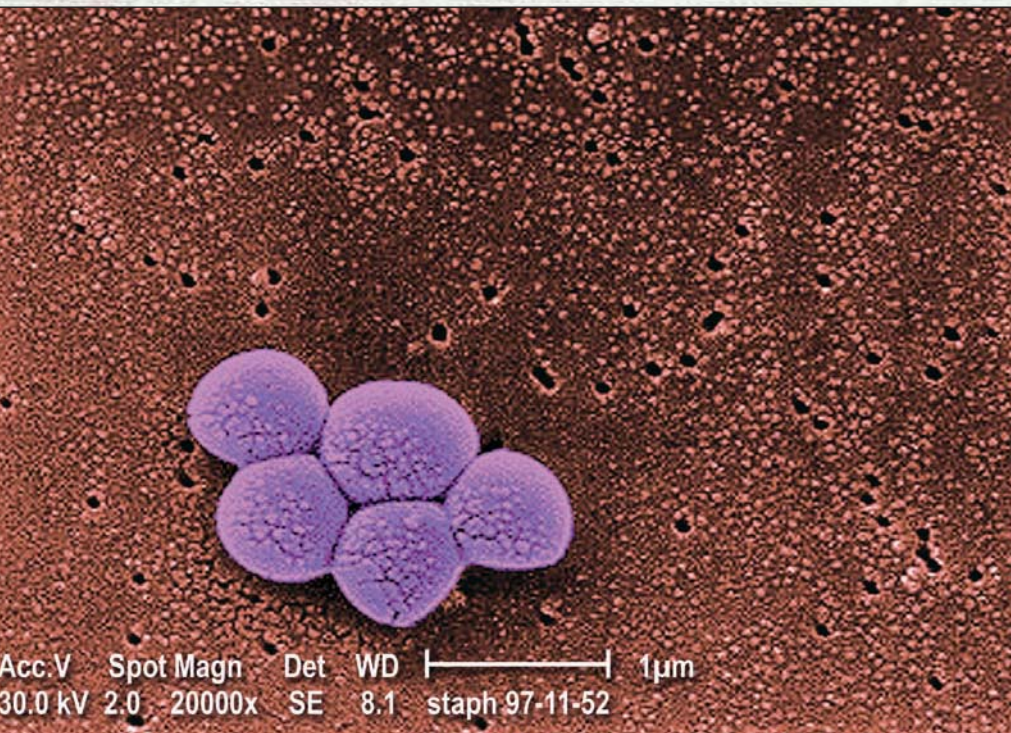


The Mechanism of Transcription in Bacteria



Colorized scanning electron micrograph of bacterial cells (*Staphylococcus aureus*). Centers for Disease Control and Prevention

In Chapter 3 we learned that transcription is the first step in gene expression. Indeed, transcription is a vital control point in the expression of many genes. Chapters 6–9 will examine in detail the mechanism of transcription and its control in bacteria. In Chapter 6 we will focus on the basic mechanism of transcription. We will begin with RNA polymerase, the enzyme that catalyzes transcription. We will also look at the interaction between RNA polymerase and DNA. This interaction begins when an RNA polymerase docks at a promoter (a specific polymerase binding site next to a gene), continues as the polymerase elongates the RNA chain, and ends when the polymerase reaches a terminator, or stopping point, and releases the finished transcript.

6.1 RNA Polymerase Structure

As early as 1960–1961, RNA polymerases were discovered in animals, plants, and bacteria. And, as you might anticipate, the bacterial enzyme was the first to be studied in great detail. By 1969, the polypeptides that make up the *E. coli* RNA polymerase had been identified by SDS polyacrylamide gel electrophoresis (SDS-PAGE) as described in Chapter 5.

Figure 6.1, lane 1, presents the results of an SDS-PAGE separation of the subunits of the *E. coli* RNA polymerase by Richard Burgess, Andrew Travers, and their colleagues. This enzyme preparation contained two very large subunits: beta (β) and beta-prime (β'), with molecular masses of 150 and 160 kD, respectively. These two subunits were not well separated in this experiment, but they were clearly distinguished in subsequent studies. The other RNA polymerase subunits visible on this gel are called **sigma** (σ) and alpha (α), with molecular masses of 70 and 40 kD, respectively. Another subunit, omega (ω), with a molecular mass of 10 kDa is not detectable here, but was clearly visible in urea gel electrophoresis experiments performed on this same enzyme preparation. In contrast to the other subunits, the ω -subunit is not required for cell viability, nor for enzyme activity in vitro. It seems to play a role, though not

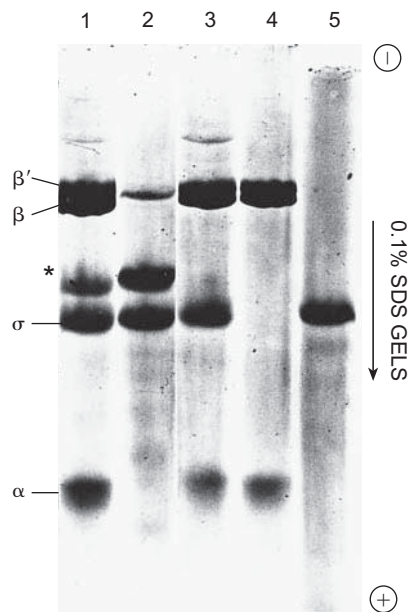


Figure 6.1 Separation of σ -factor from core *E. coli* RNA polymerase by phosphocellulose chromatography. Burgess, Travers, and colleagues subjected RNA polymerase holoenzyme to phosphocellulose chromatography, which yielded three peaks of protein: A, B, and C. Then they performed SDS-PAGE on the holoenzyme (lane 1), peaks A, B, and C (lanes 2–4, respectively), and purified σ (lane 5). Peak A contained σ , along with some contaminants (the most prominent of which is marked with an asterisk), B contained the holoenzyme, and C contained the functional core polymerase (subunits α , β , and β'). (Source: Burgess et al., “Factor Stimulating Transcription by RNA Polymerase.” *Nature* 221 (4 January 1969) p. 44, fig. 3. © Macmillan Magazines Ltd.

Table 6.1 Ability of Core and Holoenzyme to Transcribe DNAs

DNA Template	Relative Transcription Activity	
	Core	Holoenzyme
T4 (native, intact)	0.5	33.0
Calf thymus (native, nicked)	14.2	32.8

a vital one, in enzyme assembly. The polypeptide marked with an asterisk was a contaminant. Thus, the subunit content of an **RNA polymerase holoenzyme** is β' , β , σ , α_2 , ω ; in other words, two molecules of α and one of all the others are present.

When Burgess, Travers, and colleagues subjected the RNA polymerase holoenzyme to cation exchange chromatography (Chapter 5) using a phosphocellulose resin, they detected three peaks of protein, which they labeled A, B, and C. When they performed SDS-PAGE analysis of each of these peaks, they discovered that they had separated the σ -subunit from the remainder of the enzyme, called the **core polymerase**. Figure 6.1, lane 2 shows the composition of peak A, which contained the σ -subunit, along with a prominent contaminating polypeptide and perhaps a bit of β' . Lane 3 shows the polypeptides in peak B, which contained the holoenzyme. Lane 4 shows the composition of peak C, containing the core polymerase, which clearly lacks the σ -subunit. Further purification of the σ -subunit yielded the preparation in lane 5, which was free of most contamination.

Next, the investigators tested the RNA polymerase activities of the two separated components of the enzyme: the core polymerase and the σ -factor. Table 6.1 shows that this separation had caused a profound change in the enzyme’s activity. Whereas the holoenzyme could transcribe intact phage T4 DNA in vitro quite actively, the core enzyme had little ability to do this. On the other hand, core polymerase retained its basic RNA polymerizing function because it could still transcribe highly nicked templates (DNAs with single-stranded breaks) very well. (As we will see, transcription of nicked DNA is a laboratory artifact and has no biological significance.)

Sigma (σ) as a Specificity Factor

Adding σ back to the core reconstituted the enzyme’s ability to transcribe unnicked T4 DNA. Even more significantly, Ekkehard Bautz and colleagues showed that the holoenzyme transcribed only a certain class of T4 genes (called immediate early genes), but the core showed no such specificity.

Not only is the core enzyme indiscriminate about the T4 genes it transcribes, it also transcribes both DNA strands. Bautz and colleagues demonstrated this by hybridizing the

labeled product of the holoenzyme or the core enzyme to authentic T4 phage RNA and then checking for RNase resistance. That is, they attempted to get the two RNAs to base-pair together and form an RNase-resistant double-stranded RNA. Because authentic T4 RNA is made **asymmetrically** (only one DNA strand in any given region is copied), it should not hybridize to T4 RNA made properly in vitro because this RNA is also made asymmetrically and is therefore identical, not complementary, to the authentic RNA. Bautz and associates did indeed observe this behavior with RNA made in vitro by the holoenzyme. However, if the RNA is made symmetrically in vitro, up to half of it will be complementary to the in vivo RNA and will be able to hybridize to it and thereby become resistant to RNase. In fact, Bautz and associates found that about 30% of the labeled RNA made by the core polymerase in vitro became RNase-resistant after hybridization to authentic T4 RNA. Thus, the core enzyme acts in an unnatural way by transcribing both DNA strands.

Clearly, depriving the holoenzyme of its σ -subunit leaves a core enzyme with basic RNA synthesizing capability, but lacking specificity. Adding σ back restores specificity. In fact, σ was named only after this characteristic came to light, and the σ , or Greek letter s, was chosen to stand for “specificity.”

SUMMARY The key player in the transcription process is RNA polymerase. The *E. coli* enzyme is composed of a core, which contains the basic transcription machinery, and a σ -factor, which directs the core to transcribe specific genes.

6.2 Promoters

In the T4 DNA transcription experiments presented in Table 6.1, why was core RNA polymerase still capable of transcribing nicked DNA, but not intact DNA? Nicks and gaps in DNA provide ideal initiation sites for RNA polymerase, even core polymerase, but this kind of initiation is necessarily nonspecific. Few nicks or gaps occurred on the intact T4 DNA, so the core polymerase encountered only a few such artificial initiation sites and transcribed this DNA only weakly. On the other hand, when σ was present, the holoenzyme could recognize the authentic RNA polymerase binding sites on the T4 DNA and begin transcription there. These polymerase binding sites are called **promoters**. Transcription that begins at promoters in vitro is specific and mimics the initiation that would occur in vivo. Thus, σ operates by directing the polymerase to initiate at specific promoter sequences. In this section, we will examine the interaction of bacterial polymerase with promoters, and the structures of these promoters.

Binding of RNA Polymerase to Promoters

How does σ change the way the core polymerase behaves toward promoters? David Hinkle and Michael Chamberlin used nitrocellulose filter-binding studies (Chapter 5) to help answer this question. To measure how tightly holoenzyme and core enzyme bind to DNA, they isolated these enzymes from *E. coli* and bound them to ^3H -labeled T7 phage DNA, whose early promoters are recognized by the *E. coli* polymerase. Then they added a great excess of unlabeled T7 DNA, so that any polymerase that dissociated from a labeled DNA had a much higher chance of rebinding to an unlabeled DNA than to a labeled one. After varying lengths of time, they passed the mixture through nitrocellulose filters. The labeled DNA would bind to the filter only if it was still bound to polymerase. Thus, this assay measured the dissociation rate of the polymerase–DNA complex. As the last (and presumably tightest bound) polymerase dissociated from the labeled DNA, that DNA would no longer bind to the filter, so the filter would become less radioactive.

Figure 6.2 shows the results of this experiment. Obviously, the polymerase holoenzyme binds much more tightly to the T7 DNA than does the core enzyme. In fact, the holoenzyme dissociates with a half time ($t_{1/2}$) of 30–60 h, which lies far beyond the timescale of Figure 6.2. This means that after 30–60 h, only half of the complex had

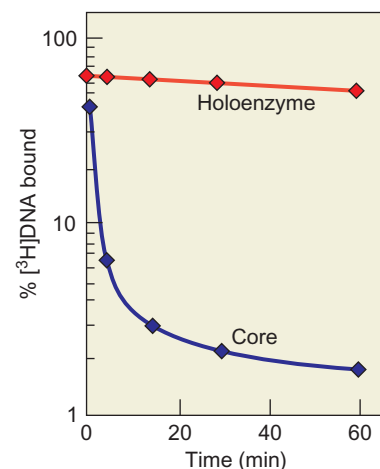


Figure 6.2 Sigma stimulates tight binding between RNA polymerase and promoter. Hinkle and Chamberlin allowed ^3H -labeled T7 DNA to bind to *E. coli* core polymerase (blue) or holoenzyme (red). Next, they added an excess of unlabeled T7 DNA, so that any polymerase that dissociated from the labeled DNA would be likely to rebind to unlabeled DNA. They filtered the mixtures through nitrocellulose at various times to monitor the dissociation of the labeled T7 DNA–polymerase complexes. (As the last polymerase dissociates from the labeled DNA, the DNA will no longer bind to the filter, which loses radioactivity.) The much slower dissociation rate of the holoenzyme (red) relative to the core polymerase (blue) shows much tighter binding between T7 DNA and holoenzyme. (Source: Adapted from Hinkle, D.C. and Chamberlin, M.J., “Studies of the Binding of *Escherichia coli* RNA Polymerase to DNA,” *Journal of Molecular Biology*, Vol. 70, 157–85, 1972.)

dissociated, which indicates very tight binding indeed. By contrast, the core polymerase dissociated with a $t_{1/2}$ of less than a minute, so it bound much less tightly than the holoenzyme did. Thus, the σ -factor can promote tight binding, at least to certain DNA sites.

In a separate experiment, Hinkle and Chamberlin switched the procedure around, first binding polymerase to unlabeled DNA, then adding excess labeled DNA, and finally filtering the mixture at various times through nitrocellulose. This procedure measured the dissociation of the first (and loosest bound) polymerase, because a newly dissociated polymerase would be available to bind to the free labeled DNA and thereby cause it to bind to the filter. This assay revealed that the holoenzyme, as well as the core, had loose binding sites on the DNA.

Thus, the holoenzyme finds two kinds of binding sites on T7 DNA: tight binding sites and loose ones. On the other hand, the core polymerase is capable of binding only loosely to the DNA. Because Bautz and coworkers had already shown that the holoenzyme, but not the core, can recognize promoters, it follows that the tight binding sites are probably promoters, and the loose binding sites represent the rest of the DNA. Chamberlin and colleagues also showed that the tight complexes between holoenzyme and T7 DNA could initiate transcription immediately on addition of nucleotides, which reinforces the conclusion that the tight binding sites are indeed promoters. If the polymerase had been tightly bound to sites remote from the promoters, a lag would have occurred while the polymerases searched for initiation sites. Furthermore, Chamberlin and coworkers titrated the tight binding sites on each molecule of T7 DNA and found only eight. This is not far from the number of early promoters on this DNA. By contrast, the number of loose binding sites for both holoenzyme and core enzyme is about 1300, which suggests that these loose sites are found virtually everywhere on the DNA and are therefore nonspecific. The inability of the core polymerase to bind to the tight (promoter) binding sites accounts for its inability to transcribe DNA specifically, which requires binding at promoters.

Hinkle and Chamberlin also tested the effect of temperature on binding of holoenzyme to T7 DNA and found a striking enhancement of tight binding at elevated temperature. Figure 6.3 shows a significantly higher dissociation rate at 25° than at 37°C, and a much higher dissociation rate at 15°C. Because high temperature promotes DNA melting (strand separation, Chapter 2) this finding is consistent with the notion that tight binding involves local melting of the DNA. We will see direct evidence for this hypothesis later in this chapter.

Hinkle and Chamberlin summarized these and other findings with the following hypothesis for polymerase–DNA interaction (Figure 6.4): RNA polymerase holoenzyme binds loosely to DNA at first. It either binds initially

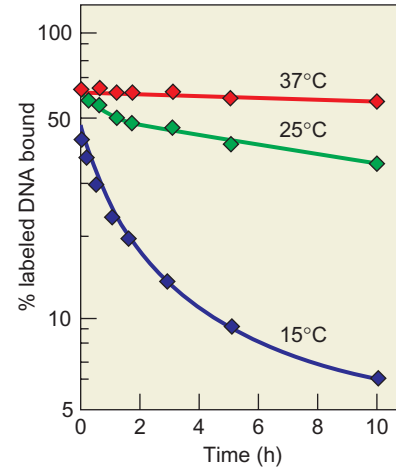
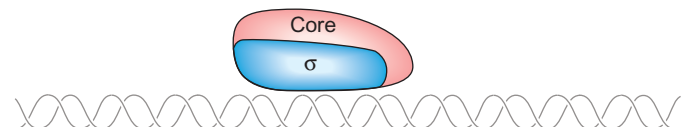
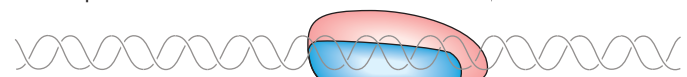


Figure 6.3 The effect of temperature on the dissociation of the polymerase–T7 DNA complex. Hinkle and Chamberlin formed complexes between *E. coli* RNA polymerase holoenzyme and ^3H -labeled T7 DNA at three different temperatures: 37°C (red), 25°C (green), and 15°C (blue). Then they added excess unlabeled T7 DNA to compete with any polymerase that dissociated; they removed samples at various times and passed them through a nitrocellulose filter to monitor dissociation of the complex. The complex formed at 37°C was more stable than that formed at 25°C, which was much more stable than that formed at 15°C. Thus, higher temperature favors tighter binding between RNA polymerase holoenzyme and T7 DNA. (Source: Adapted from Hinkle, D.C. and Chamberlin, M.J., "Studies of the Binding of *Escherichia coli* RNA Polymerase to DNA," *Journal of Molecular Biology*, Vol. 70, 157–85, 1972.)

(a) Promoter search



(b) Closed promoter complex formation



(c) Open promoter complex formation

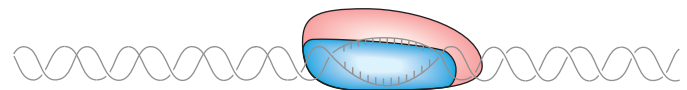


Figure 6.4 RNA polymerase/promoter binding. (a) The holoenzyme binds and rebinds loosely to the DNA, searching for a promoter. (b) The holoenzyme has found a promoter and has bound loosely, forming a closed promoter complex. (c) The holoenzyme has bound tightly, melting a local region of DNA and forming an open promoter complex.

at a promoter or scans along the DNA until it finds one. The complex with holoenzyme loosely bound at the promoter is called a **closed promoter complex** because the DNA remains in closed double-stranded form. Then the holoenzyme can melt a short region of the DNA at the promoter to form an **open promoter complex** in which the polymerase is bound tightly to the DNA. This is called an open promoter complex because the DNA has to open up to form it.

It is this conversion of a loosely bound polymerase in a closed promoter complex to the tightly bound polymerase in the open promoter complex that requires σ , and this is also what allows transcription to begin. We can now appreciate how σ fulfills its role in determining specificity of transcription: It selects the promoters to which RNA polymerase will bind tightly. The genes adjacent to these promoters will then be transcribed.

SUMMARY The σ -factor allows initiation of transcription by causing the RNA polymerase holoenzyme to bind tightly to a promoter. This tight binding depends on local melting of the DNA to form an open promoter complex and is stimulated by σ . The σ -factor can therefore select which genes will be transcribed.

Promoter Structure

What is the special nature of a bacterial promoter that attracts RNA polymerase? David Pribnow compared several *E. coli* and phage promoters and discerned a region they held in common: a sequence of 6 or 7 bp centered approximately 10 bp upstream of the start of transcription. This was originally dubbed the “Pribnow box,” but is now usually called the **–10 box**. Mark Ptashne and colleagues noticed another short sequence centered approximately 35 bp upstream of the transcription start site; it is known as the **–35 box**. Thousands of promoters have now been examined and a typical, or **consensus sequence** for each of these boxes has emerged (Figure 6.5).

These so-called consensus sequences represent probabilities. The capital letters in Figure 6.5 denote bases that have a high probability of being found in the given position. The lowercase letters correspond to bases that are usually found in the given position, but at a lower frequency than those

denoted by capital letters. The probabilities are such that one rarely finds –10 or –35 boxes that match the consensus sequences perfectly. However, when such perfect matches are found, they tend to occur in very strong promoters that initiate transcription unusually actively. In fact, mutations that destroy matches with the consensus sequences tend to be **down mutations**. That is, they make the promoter weaker, resulting in less transcription. Mutations that make the promoter sequences more like the consensus sequences usually make the promoters stronger; these are called **up mutations**. The spacing between promoter elements is also important, and deletions or insertions that move the –10 and –35 boxes unnaturally close together or far apart are deleterious. In Chapter 10 we will see that eukaryotic promoters have their own consensus sequences, one of which resembles the –10 box quite closely.

In addition to the –10 and –35 boxes, which we can call **core promoter elements**, some very strong promoters have an additional element farther upstream called an **UP element**. *E. coli* cells have seven genes (***rrn* genes**) that encode rRNAs. Under rapid growth conditions, when rRNAs are required in abundance, these seven genes by themselves account for the majority of the transcription occurring in the cell. Obviously, the promoters driving these genes are extraordinarily powerful, and their UP elements are part of the explanation. Figure 6.6 shows the structure of one of these promoters, the *rrnB* P1 promoter. Upstream of the core promoter (blue), there is an UP element (red) between positions –40 and –60. We know that the UP element is a true promoter element because it stimulates transcription of the *rrnB* P1 gene by a factor of 30 in the presence of RNA polymerase alone. Because it is recognized by the polymerase itself, we conclude that it is a promoter element.

This promoter is also associated with three so-called Fis sites between positions –60 and –150, which are binding sites for the transcription-activator protein Fis. The Fis sites, because they do not bind to RNA polymerase itself, are not classical promoter elements, but instead are members of another class of transcription-activating DNA elements called **enhancers**. We will discuss bacterial enhancers in greater detail in Chapter 9.

The *E. coli* *rrn* promoters are also regulated by a pair of small molecules: the initiating NTP (the iNTP) and an **alarmone**, guanosine 5'-diphosphate 3'-diphosphate (**ppGpp**). An abundance of iNTP indicates that the concentration of

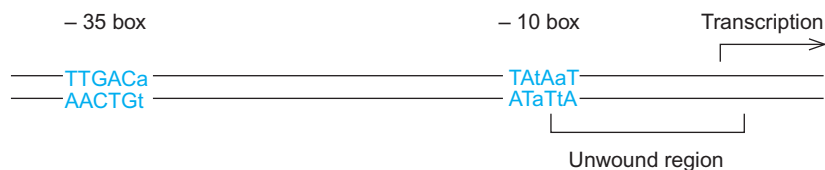


Figure 6.5 A bacterial promoter. The positions of –10 and –35 boxes and the unwound region are shown relative to the start of transcription for a typical *E. coli* promoter. Capital letters denote bases found in those positions in more than 50% of promoters examined; lower-case letters denote bases found in those positions in 50% or fewer of promoters examined.

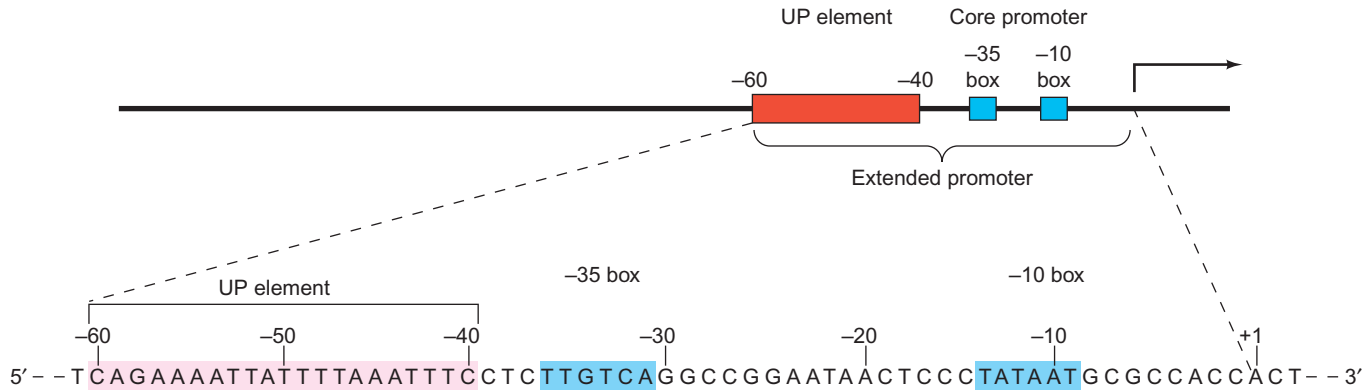


Figure 6.6 The *rrnB* P1 promoter. The core promoter elements (–10 and –35 boxes, blue) and the UP element (red) are shown schematically above, and with their complete base sequences (nontemplate strand) below, with the same color coding. (Source: Adapted from Ross et al., “A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase.” *Science* 262:1407, 1993.)

nucleotides is high, and therefore it is appropriate to synthesize plenty of rRNA. Accordingly, iNTP stabilizes the open promoter complex, stimulating transcription.

On the other hand, when cells are starved for amino acids, protein synthesis cannot occur readily and the need for ribosomes (and rRNA) decreases. Ribosomes sense the lack of amino acids when uncharged tRNAs bind to the ribosomal site where aminoacyl-tRNAs would normally bind. Under these conditions, a ribosome-associated protein called RelA receives the “alarm” and produces the “alarmone” ppGpp, which destabilizes open promoter complexes whose lifetimes are normally short, thus inhibiting transcription.

The protein DskA also plays an important role. It binds to RNA polymerase and reduces the lifetimes of the *rrn* open promoters to a level at which they are responsive to changes in iNTP and ppGpp concentrations. Thus, DskA is required for the regulation of *rrn* transcription by these two small molecules. Indeed, *rrn* transcription is insensitive to iNTP and ppGpp in mutants lacking DskA.

SUMMARY Bacterial promoters contain two regions centered approximately at –10 and –35 bp upstream of the transcription start site. In *E. coli*, these bear a greater or lesser resemblance to two consensus sequences: TATAAT and TTGACA, respectively. In general, the more closely regions within a promoter resemble these consensus sequences, the stronger that promoter will be. Some extraordinarily strong promoters contain an extra element (an UP element) upstream of the core promoter. This makes these promoters even more attractive to RNA polymerase. Transcription from the *rrn* promoters responds positively to increases in the concentration of iNTP, and negatively to the alarmone ppGpp.

6.3 Transcription Initiation

Until 1980, it was a common assumption that transcription initiation ended when RNA polymerase formed the first phosphodiester bond, joining the first two nucleotides in the growing RNA chain. Then, Agamemnon Carpousis and Jay Gralla reported that initiation is more complex than that. They incubated *E. coli* RNA polymerase with DNA bearing a mutant *E. coli lac* promoter known as the *lac* UV5 promoter. Along with the polymerase and DNA, they included heparin, a negatively charged polysaccharide that competes with DNA in binding tightly to free RNA polymerase. The heparin prevented any reassociation between DNA and polymerase released at the end of a cycle of transcription. These workers also included labeled ATP in their assay to label the RNA products. Then they subjected the products to gel electrophoresis to measure their sizes. They found several very small oligonucleotides, ranging in size from dimers to hexamers (2–6 nt long), as shown in Figure 6.7. The sequences of these oligonucleotides matched the sequence of the beginning of the expected transcript from the *lac* promoter. Moreover, when Carpousis and Gralla measured the amounts of these oligonucleotides and compared them to the number of RNA polymerases, they found many oligonucleotides per polymerase. Because the heparin in the assay prevented free polymerase from reassociating with the DNA, this result implied that the polymerase was making many small, **abortive transcripts** without ever leaving the promoter. Other investigators have since verified this result and have found abortive transcripts up to 9 or 10 nt in size.

Thus, we see that transcription initiation is more complex than first supposed. It is now commonly represented in four steps, as depicted in Figure 6.8: (1) formation of a closed promoter complex; (2) conversion of the closed promoter complex to an open promoter complex; (3) polymerizing the first few nucleotides (up to 10) while the polymerase remains at the promoter, in an **initial transcribing complex**;

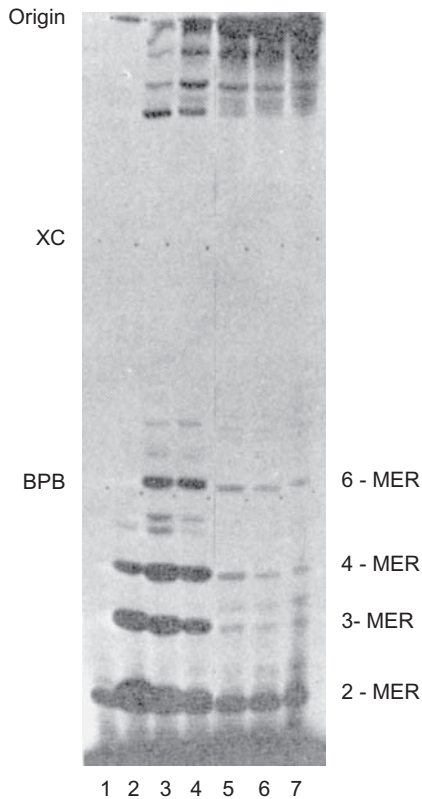


Figure 6.7 Synthesis of short oligonucleotides by RNA polymerase bound to a promoter. Carpousis and Gralla allowed *E. coli* RNA polymerase to synthesize ^{32}P -labeled RNA in vitro using a DNA containing the *lac* UV5 promoter, heparin to bind any free RNA polymerase, $[\text{}^{32}\text{P}]\text{ATP}$, and various concentrations of the other three nucleotides (CTP, GTP, and UTP). They electrophoresed the products on a polyacrylamide gel and visualized the oligonucleotides by autoradiography. Lane 1 is a control with no DNA; lane 2, ATP only; lanes 3–7; ATP with concentrations of CTP, GTP, and UTP increasing by twofold in each lane, from 25 μM in lane 3 to 400 μM in lane 7. The positions of 2-mers through 6-mers are indicated at right. The positions of two marker dyes (bromophenol blue [BPB] and xylene cyanol [XC]) are indicated at left. The apparent dimer in lane 1, with no DNA, is an artifact caused by a contaminant in the labeled ATP. The same artifact can be presumed to contribute to the bands in lanes 2–7. (Source: Carpousis A.J. and Gralla J.D. Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the *lac* UV5 promoter. *Biochemistry* 19 (8 Jul 1980) p. 3249, f. 2, © American Chemical Society.)

and (4) **promoter clearance**, in which the transcript becomes long enough to form a stable hybrid with the template strand. This helps to stabilize the transcription complex, and the polymerase changes to its elongation conformation and moves away from the promoter. In this section, we will examine the initiation process in more detail.

Sigma Stimulates Transcription Initiation

Because σ directs tight binding of RNA polymerase to promoters, it places the enzyme in a position to initiate transcription—at the beginning of a gene. Therefore, we

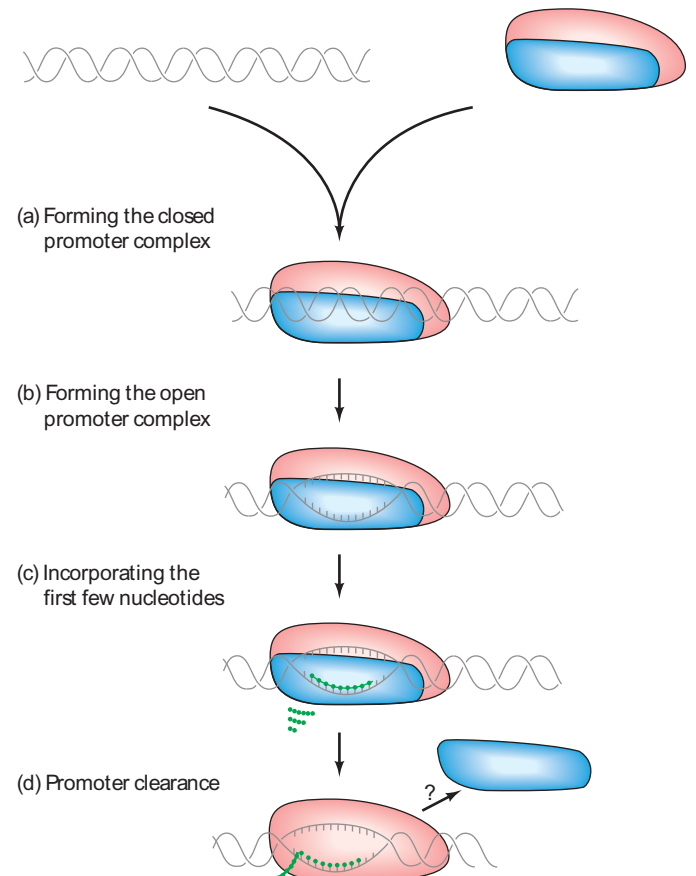


Figure 6.8 Stages of transcription initiation. (a) RNA polymerase binds to DNA in a closed promoter complex. (b) The σ -factor stimulates the polymerase to convert the closed promoter complex to an open promoter complex. (c) The polymerase incorporates the first 9 or 10 nt into the nascent RNA. Some abortive transcripts are pictured at left. (d) The polymerase clears the promoter and begins the elongation phase. The σ -factor may be lost at this point or later, during elongation.

would expect σ to stimulate initiation of transcription. To test this, Travers and Burgess took advantage of the fact that the first nucleotide incorporated into an RNA retains all three of its phosphates (α , β , and γ), whereas all other nucleotides retain only their α -phosphate (Chapter 3). These investigators incubated polymerase core in the presence of increasing amounts of σ in two separate sets of reactions. In some reactions, the labeled nucleotide was $[\text{}^{14}\text{C}]\text{ATP}$, which is incorporated throughout the RNA and therefore measures elongation, as well as initiation, of RNA chains. In the other reactions, the labeled nucleotide was $[\gamma\text{}^{32}\text{P}]\text{ATP}$ or $[\gamma\text{}^{32}\text{P}]\text{GTP}$, whose label should be incorporated only into the first position of the RNA, and therefore is a measure of transcription initiation. (They used ATP and GTP because transcription usually starts with a purine nucleotide—more often ATP than GTP.) The results in Figure 6.9 show that σ stimulated the incorporation of both ^{14}C - and $\gamma\text{}^{32}\text{P}$ -labeled nucleotides, which suggests that σ enhanced both initiation and elongation.

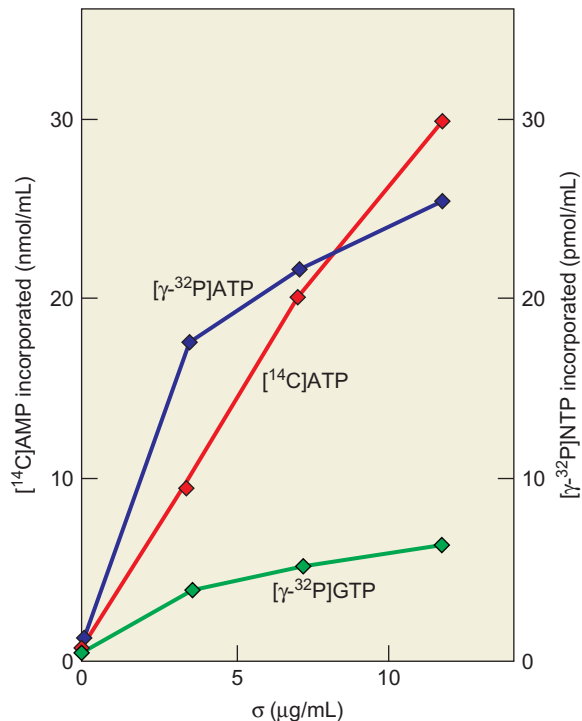


Figure 6.9 Sigma seems to stimulate both initiation and elongation. Travers and Burgess transcribed T4 DNA in vitro with *E. coli* RNA polymerase core plus increasing amounts of σ . In separate reactions, they included $[^{14}\text{C}]\text{ATP}$ (red), $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (blue), or $[\gamma\text{-}^{32}\text{P}]\text{GTP}$ (green) in the reaction mix. The incorporation of the $[^{14}\text{C}]\text{ATP}$ measured bulk RNA synthesis, or elongation; the incorporation of the $\gamma\text{-}^{32}\text{P}$ -labeled nucleotides measured initiation. Because all three curves rise with increasing σ concentration, this experiment makes it appear that σ stimulates both elongation and initiation. (Source: Adapted from Travers, A.A. and R.R. Burgess, "Cyclic re-use of the RNA polymerase sigma factor." *Nature* 222:537–40, 1969.)

However, initiation is the rate-limiting step in transcription (it takes longer to get a new RNA chain started than to extend one). Thus, σ could appear to stimulate elongation by stimulating initiation and thereby providing more initiated chains for core polymerase to elongate.

Travers and Burgess proved that is the case by demonstrating that σ really does not accelerate the rate of RNA chain growth. To do this, they held the number of RNA chains constant and showed that under those conditions σ did not affect the length of the RNA chains. They held the number of RNA chains constant by allowing a certain amount of initiation to occur, then blocking any further chain initiation with the antibiotic **rifampicin**, which blocks bacterial transcription initiation, but not elongation. Then they used ultracentrifugation to measure the length of RNAs made in the presence or absence of σ . They found that σ made no difference in the lengths of the RNAs. If it really had stimulated the rate of elongation, it would have made the RNAs longer. Therefore, σ does not stimulate elongation, and the apparent stimulation in the previous experiment was simply an indirect effect of enhanced initiation.

SUMMARY Sigma stimulates initiation, but not elongation, of transcription.

Reuse of σ

In the same 1969 paper, Travers and Burgess demonstrated that σ can be recycled. The key to this experiment was to run the transcription reaction at low ionic strength, which prevents RNA polymerase core from dissociating from the DNA template at the end of a gene. This caused transcription initiation (as measured by the incorporation of $\gamma\text{-}^{32}\text{P}$ -labeled purine nucleotides into RNA) to slow to a stop, as depicted in Figure 6.10 (red line). Then, when they added

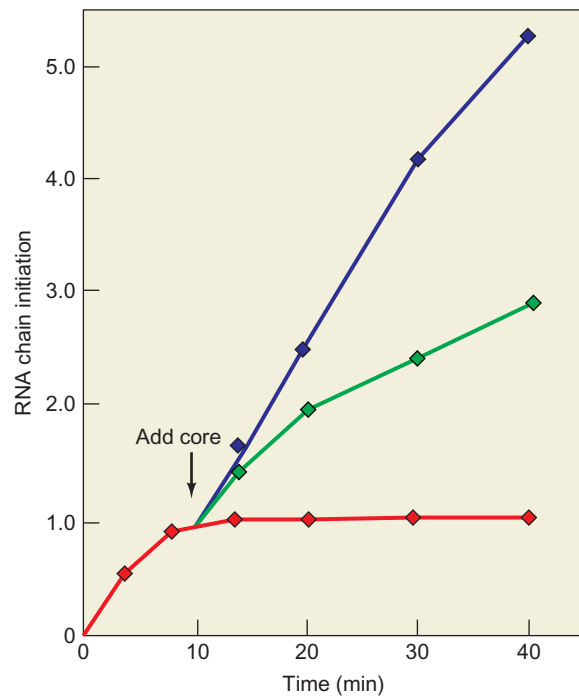


Figure 6.10 Sigma can be reused. Travers and Burgess allowed RNA polymerase holoenzyme to initiate and elongate RNA chains on a T4 DNA template at low ionic strength, so the polymerases could not dissociate from the template to start new RNA chains. The red curve shows the initiation of RNA chains, measured by $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and $[\gamma\text{-}^{32}\text{P}]\text{GTP}$ incorporation, under these conditions. After 10 min (arrow), when most chain initiation had ceased, the investigators added new, rifampicin-resistant core polymerase in the presence (green) or absence (blue) of rifampicin. The immediate rise of both curves showed that addition of core polymerase can restart RNA synthesis, which implied that the new core associated with σ that had been associated with the original core. In other words, the σ was recycled. The fact that transcription occurred even in the presence of rifampicin showed that the new core, which was from rifampicin-resistant cells, together with the old σ , which was from rifampicin-sensitive cells, could carry out rifampicin-resistant transcription. Thus, the core, not the σ , determines rifampicin resistance or sensitivity. (Source: Adapted from Travers, A.A. and R.R. Burgess, "Cyclic re-use of the RNA polymerase sigma factor." *Nature* 222:537–40, 1969.)

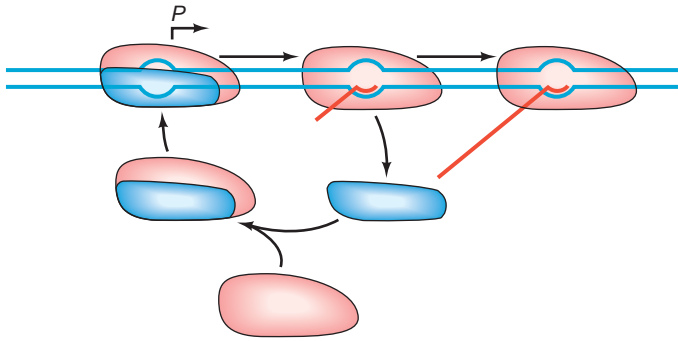


Figure 6.11 The σ cycle. RNA polymerase binds to the promoter at left, causing local melting of the DNA. As the polymerase moves to the right, elongating the RNA, the σ -factor dissociates and joins with a new core polymerase (lower left) to initiate another RNA chain.

new core polymerase, these investigators showed that transcription began anew (blue line). This meant that the new core was associating with σ that had been released from the original holoenzyme. In a separate experiment, they demonstrated that the new transcription could occur on a different kind of DNA added along with the new core polymerase. This supported the conclusion that σ had been released from the original core and was associating with a new core on a new DNA template. Accordingly, Travers and Burgess proposed that σ cycles from one core to another, as shown in Figure 6.11. They dubbed this the “ σ cycle.”

Figure 6.10 contains still another piece of valuable information. When Travers and Burgess added rifampicin, along with the core polymerase, which came from a rifampicin-resistant mutant, transcription still occurred (green line). Because the σ was from the original, rifampicin-sensitive polymerase, the rifampicin resistance in the renewed transcription must have been conferred by the newly added core. The fact that less initiation occurred in the presence of rifampicin probably means that the rifampicin-resistant core is still somewhat sensitive. We might have expected the σ -factor, not the core, to determine rifampicin sensitivity or resistance because rifampicin blocks initiation, and σ is the acknowledged initiation factor. Nevertheless, the core is the key to rifampicin sensitivity, and experiments to be presented later in this chapter will provide some clarification of why this is so.

SUMMARY At some point after σ has participated in initiation, it appears to dissociate from the core polymerase, leaving the core to carry out elongation. Furthermore, σ can be reused by different core polymerases, and the core, not σ , governs rifampicin sensitivity or resistance.

The Stochastic σ -Cycle Model

The σ -cycle model that arose from Travers and Burgess’s experiments called for the dissociation of σ from core as the polymerase undergoes promoter clearance and switches from initiation to elongation mode. This has come to be known as the **obligate release** version of the σ -cycle model. Although this model has held sway for over 30 years and has considerable experimental support, it does not fit all the data at hand. For example, Jeffrey Roberts and colleagues demonstrated in 1996 that σ is involved in pausing at position +16/+17 downstream of the late promoter ($P_{R'}$) in λ phage. This implies that σ is still attached to core polymerase at position +16/+17, well after promoter clearance has occurred.

Based on this and other evidence, an alternative view of the σ -cycle was proposed: the **stochastic release model**. (“Stochastic” means “random”; Greek: *stochos*, meaning guess.) This hypothesis holds that σ is indeed released from the core polymerase, but there is no discrete point during transcription at which this release is required; rather, it is released randomly. As we will see, the preponderance of evidence now favors the stochastic release model.

Richard Ebright and coworkers noted in 2001 that all of the evidence favoring the obligate release model relies on harsh separation techniques, such as electrophoresis or chromatography. These could strip σ off of core if σ is weakly bound to core during elongation and, thus, make it appear that σ had dissociated from core during promoter clearance. These workers also noted that previous work had generally failed to distinguish between active and inactive RNA polymerases. This is a real concern because a significant fraction of RNA polymerase molecules in any population is not competent to switch from initiation to elongation mode.

To test the obligate release hypothesis, Ebright and coworkers used a technique, **fluorescence resonance energy transfer (FRET)**, that allows the position of σ relative to a site on the DNA to be measured without using separation techniques that might themselves displace σ from core. The FRET technique relies on the fact that two fluorescent molecules close to each other will engage in transfer of resonance energy, and the efficiency of this energy transfer (FRET efficiency) will decrease rapidly as the two molecules move apart.

Ebright and coworkers measured FRET with fluorescent molecules (**fluorescence probes**) on both σ and DNA. The probe on σ serves as the fluorescence donor, and the probe on the DNA serves as the fluorescence acceptor. Sometimes the probe on the DNA was at the 5′, or upstream end (trailing-edge FRET), which allowed the investigators to observe the drop in FRET as the polymerase moved away from the promoter and the 5′-end of the DNA. In other experiments, the probe on the DNA was at the 3′-, or downstream end (leading-edge FRET), which allowed the investigators to observe the increase in FRET as the polymerase moved toward the downstream end. Figure 6.12 illustrates the strategies of trailing-edge and leading-edge FRET.

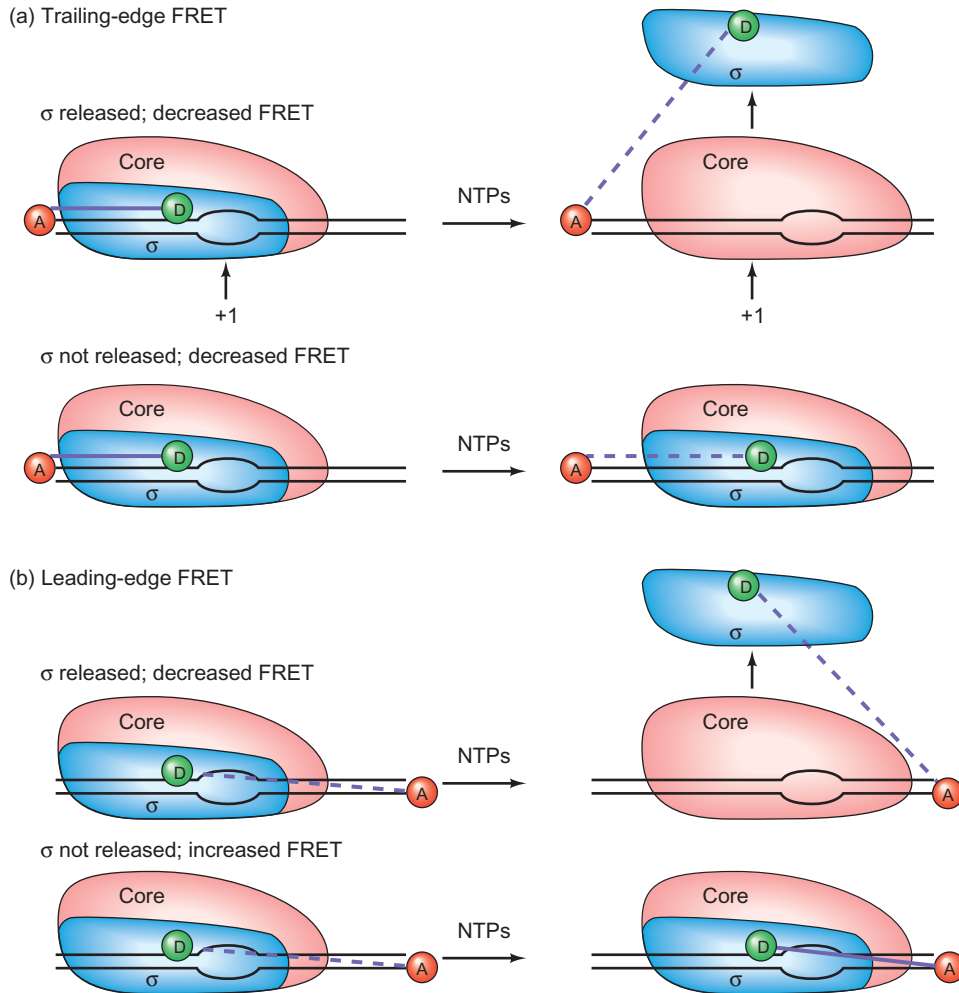


Figure 6.12 Rationale of FRET assay for σ movement relative to DNA. (a) Trailing-edge FRET. A fluorescence donor (D, green) is attached to the single cysteine residue in a σ^{70} mutant that had been engineered to eliminate all but one cysteine. A fluorescence acceptor (A, red) is attached to the 5'-end of the DNA. FRET efficiency is high (solid purple line) in the open promoter complex (RP_0) because the two probes are close together. On addition of 3 of the 4 nucleotides, the polymerase moves to a position downstream at which the fourth nucleotide (CTP) is required. This is at least position +11, so promoter clearance occurs. FRET efficiency decreases (dashed purple line) regardless of whether σ dissociates from the core, because the two probes grow farther apart in either case. If σ does not dissociate, it would travel with the core downstream during elongation, taking it farther from the probe at the 5'-end of the DNA.

If σ dissociates, it would be found at random positions in solution, but, on average, it would be much farther away from the core than it was in the open promoter complex before transcription began. (b) Leading-edge FRET. Again a fluorescence donor is attached to σ^{70} , but this time, the fluorescence acceptor is attached to the 3'-end of the DNA. FRET efficiency is low (dashed purple line) in the open promoter complex because the two probes are far apart. On the addition of nucleotides, the polymerase undergoes promoter clearance and elongates to a downstream position as in (a). Now FRET can distinguish between the two hypotheses. If σ dissociates from core, FRET should decrease (dashed purple line), as it did in panel (a). On the other hand, if σ remains bound to core, the two probes will grow closer together as the polymerase moves downstream, and FRET efficiency will increase (solid purple line).

The trailing-edge FRET strategy does not distinguish between one model in which σ dissociates from the core, and a second model in which σ does not dissociate, after promoter clearance. In both cases, the donor probe on σ gets farther away from the acceptor probe at the upstream end of the DNA after promoter clearance and the FRET efficiency therefore decreases. Indeed, Figure 6.13a shows that the FRET efficiency does decrease with time when the probe on the DNA is at the upstream end.

On the other hand, the leading-edge strategy can distinguish between the two models (Figure 6.12b). If σ dissociates from the core, then FRET efficiency should decrease, just as it did in the trailing-edge experiment. But if σ is not released from the core, it should move closer to the probe at the downstream end of the DNA with time, and FRET efficiency should increase. Figure 6.13b shows that FRET efficiency did indeed increase, which supports the hypothesis that σ remains with the core after promoter clearance. In fact, the

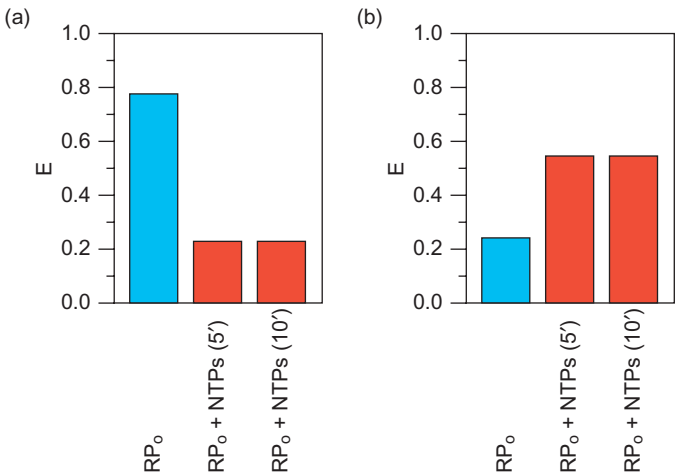


Figure 6.13 FRET analysis of σ -core association after promoter clearance. Ebright and coworkers performed FRET analysis as described in Figure 6.12. **(a)** Trailing-edge FRET results; **(b)** leading-edge FRET results. Blue bars, FRET efficiency (E) of open promoter complex (RP₀); red bars, FRET efficiency after 5 and 10 min, respectively, in the presence of the three nucleotides that allow the polymerase to move 11 bp downstream of the promoter.

magnitude of the FRET efficiency increase suggests that 100% of the complexes after promoter clearance still retained their σ -factor.

Ebright and coworkers performed the experiments in Figure 6.13a and b in a polyacrylamide gel as follows. They formed open promoter complexes in solution, then added heparin to bind to any uncomplexed polymerase. Then they subjected the complexes to nondenaturing electrophoresis in a polyacrylamide gel. They located the complexes in the gel, sliced the gel and removed the slice containing the complexes, placed that gel slice in a container called a cuvette that fits into the fluorescence-measuring instrument (a fluorometer), added transcription buffer, and measured FRET efficiency on RP₀. Then they added three nucleotides to allow the polymerase to move downstream, and measured FRET efficiency on the elongation complex. This in-gel assay has the advantage of measuring FRET efficiency only on active complexes, because gel electrophoresis removes inactive (closed promoter) complexes. To eliminate the possibility that electrophoresis introduced an artifact of some kind, Ebright and coworkers performed the same experiments in solution and obtained very similar results.

In 2001, Bar-Nahum and Nudler also presented evidence for retention of σ . They formed complexes between holoenzyme and a DNA containing one promoter, then added three out of four nucleotides to allow the polymerase to move to position +32. Then they purified this elongation complex (called EC32) rapidly and gently by annealing the upstream end of the elongating RNA to a complementary oligonucleotide attached to resin beads. This allowed the beads, along with the complexes, to be purified quickly by low-speed centrifugation. Only elongation

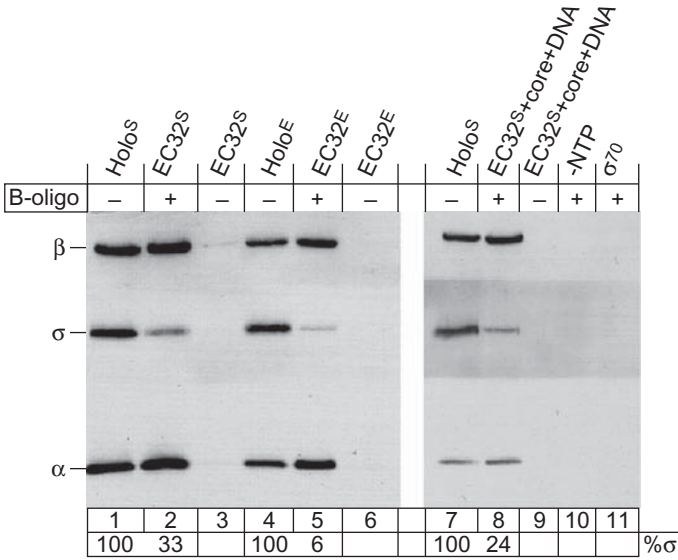


Figure 6.14 Measuring σ associated with transcription elongation complexes. Bar-Nahum and Nudler purified elongation complexes stalled at position +32 from stationary cells (EC32^S complexes) or from exponentially growing cells (EC32^E complexes), released the proteins from the nascent RNAs with nuclease, and subjected the proteins to SDS-PAGE, followed by immunoblotting. The nature of the complex and the presence or absence of an oligonucleotide on the beads used to purify the complexes is denoted at the top. Lanes 8 and 9 are controls in which excess amounts of core and DNA were added to EC32^S complexes prior to binding to the oligonucleotide beads. The purpose was to rule out σ attachment to beads due to nonspecific binding between σ and core or DNA. (Source: Reprinted from *Cell* v. 106, Bar-Nahum and Nudler, p. 444, © 2001, with permission from Elsevier Science.)

complexes are purified this way, because they are the only ones with a nascent RNA that can bind to the complementary oligonucleotide.

Finally, Bar-Nahum and Nudler released the complexes from the beads with nuclease, subjected the proteins to SDS-PAGE, and performed an immunoblot (Chapter 5) to identify the proteins associated with the complexes. Figure 6.14 shows that the purified EC32 complexes contained at least some σ . Quantification showed that complexes isolated from stationary phase cells contained $33 \pm 2\%$ of the full complement of σ per complex, and complexes isolated from exponential phase cells contained $6 \pm 1\%$ of the full complement of σ per complex. This is considerably less than the 100% observed by Ebright and coworkers and suggests relatively weak binding between σ and core in elongation complexes. Nevertheless, even these amounts of complexes that retain σ could aid considerably in reinitiation of transcription, because the association of core with σ is the rate-limiting step in transcription initiation.

Although the results of Bar-Nahum and Nudler, and those of Ebright and colleagues appear to rule out the obligate release model, and may seem to argue against the σ -cycle in general, they are actually consistent with the stochastic release version of the σ -cycle, which calls for σ

release at multiple points throughout transcription. Bar-Nahum and Nudler collected elongation complexes after only 32-nt of transcription, which could be too early in transcription to see complete σ release. And, while it is true that Ebright and colleagues did not observe significant σ dissociation after 50 nt of transcription in the experiments we have discussed, they were unwittingly using a DNA template (the *E. coli lacUV5* promoter) that contributed to this phenomenon. This promoter contains a second -10 -like box just downstream of the transcription start site. It has recently been learned that this sequence causes pausing that depends on σ , and indeed appears to aid in σ retention. When this second -10 -like box was mutated, the FRET signal decreased, and σ dissociation increased more than 4-fold. Furthermore, when they performed their original experiments with fluorescent labels on σ and core, rather than σ and DNA, Ebright and colleagues found that their FRET signal did decrease with increasing transcript length. All of these findings suggest that some σ was dissociating from core during the transcription process, and that the DNA sequence can influence the rate of such dissociation.

To probe further the validity of the σ -cycle hypothesis, Ebright and colleagues used leading and trailing edge single-molecule FRET analysis with alternating-laser excitation (single-molecule FRET ALEX). For leading edge FRET, they tagged the leading edge of σ with the donor fluorophore and a downstream DNA site with the acceptor. For trailing edge FRET, they tagged the trailing edge of σ with the donor and an upstream DNA site with the acceptor fluorophore. They measured both fluorescence efficiency and “stoichiometry,” or the presence of one or both of the fluorophores (donor and acceptor) in a small (femtoliter [10^{-15} L] scale) excitation volume, which should have at most one copy of the elongation complex at any given time. They switched rapidly between exciting the donor and acceptor fluorophore, such that each would be excited multiple times during the approximately 1 ms transit time through the excitation volume. Furthermore, they stalled the elongation complex at various points (nascent RNAs 11, 14, and 50 nt long) by coupling the *E. coli lacUV5* promoter to various G-less cassettes (Chapter 5) and leaving out CTP in the transcription reaction. By measuring both fluorescence efficiency and stoichiometry for the same elongation complex, they could tell: (1) how far transcription had progressed (by the fluorescence efficiency, which grows weaker in trailing edge FRET, and stronger in leading edge FRET, as transcription progresses); and (2) whether or not σ had dissociated from core (by the stoichiometry, which should be approximately 0.5 for holoenzyme, but nearer 0 for core alone and 1.0 for σ alone).

These studies confirmed that σ did indeed remain associated with the great majority (about 90%) of elongation complexes that had achieved promoter clearance (with transcripts 11 nt long). Again, this finding argued strongly against the obligate release model. But they also showed

that about half of halted elongation complexes with longer transcripts had lost their σ -factors, in accord with the stochastic release model. Finally, their results suggested that some elongation complexes may retain their σ -factors throughout the transcription process. If that is true, these elongation complexes are avoiding the σ cycle altogether.

SUMMARY The σ -factor appears to be released from the core polymerase, but not usually immediately upon promoter clearance. Rather, σ seems to exit from the elongation complex in a stochastic manner during the elongation process.

Local DNA Melting at the Promoter

Chamberlin’s studies on RNA polymerase–promoter interactions showed that such complexes were much more stable at elevated temperature. This suggested that local melting of DNA occurs on tight binding to polymerase, because high temperature would tend to stabilize melted DNA. Furthermore, such DNA melting is essential because it exposes bases of the template strand so they can base-pair with bases on incoming nucleotides.

Tao-shih Hsieh and James Wang provided more direct evidence for local DNA melting in 1978. They bound *E. coli* RNA polymerase to a restriction fragment containing three phage T7 early promoters and measured the hyperchromic shift (Chapter 2) caused by such binding. This increase in the DNA’s absorbance of 260-nm light is not only indicative of DNA strand separation, its magnitude is directly related to the number of base pairs that are opened. Knowing the number of RNA polymerase holoenzymes bound to their DNA, Hsieh and Wang calculated that each polymerase caused a separation of about 10 bp.

In 1979, Ulrich Siebenlist, identified the base pairs that RNA polymerase melted in a T7 phage early promoter. Figure 6.15 shows the strategy of his experiment. First he end-labeled the promoter DNA, then added RNA polymerase to form an open promoter complex. As we have seen, this involves local DNA melting, and when the strands separate, the N_1 of adenine—normally involved in hydrogen bonding to a T in the opposite strand—becomes susceptible to attack by certain chemical agents. In this case, Siebenlist methylated the exposed adenines with dimethyl sulfate (DMS). Then, when he removed the RNA polymerase and the melted region closed up again, the methyl groups prevented proper base-pairing between these N_1 -methyl-adenines and the thymines in the opposite strand and thus preserved at least some of the single-stranded character of the formerly melted region. Next, he treated the DNA with S1 nuclease, which specifically cuts single-stranded DNA. This enzyme should therefore cut wherever an adenine had been in a melted region of the

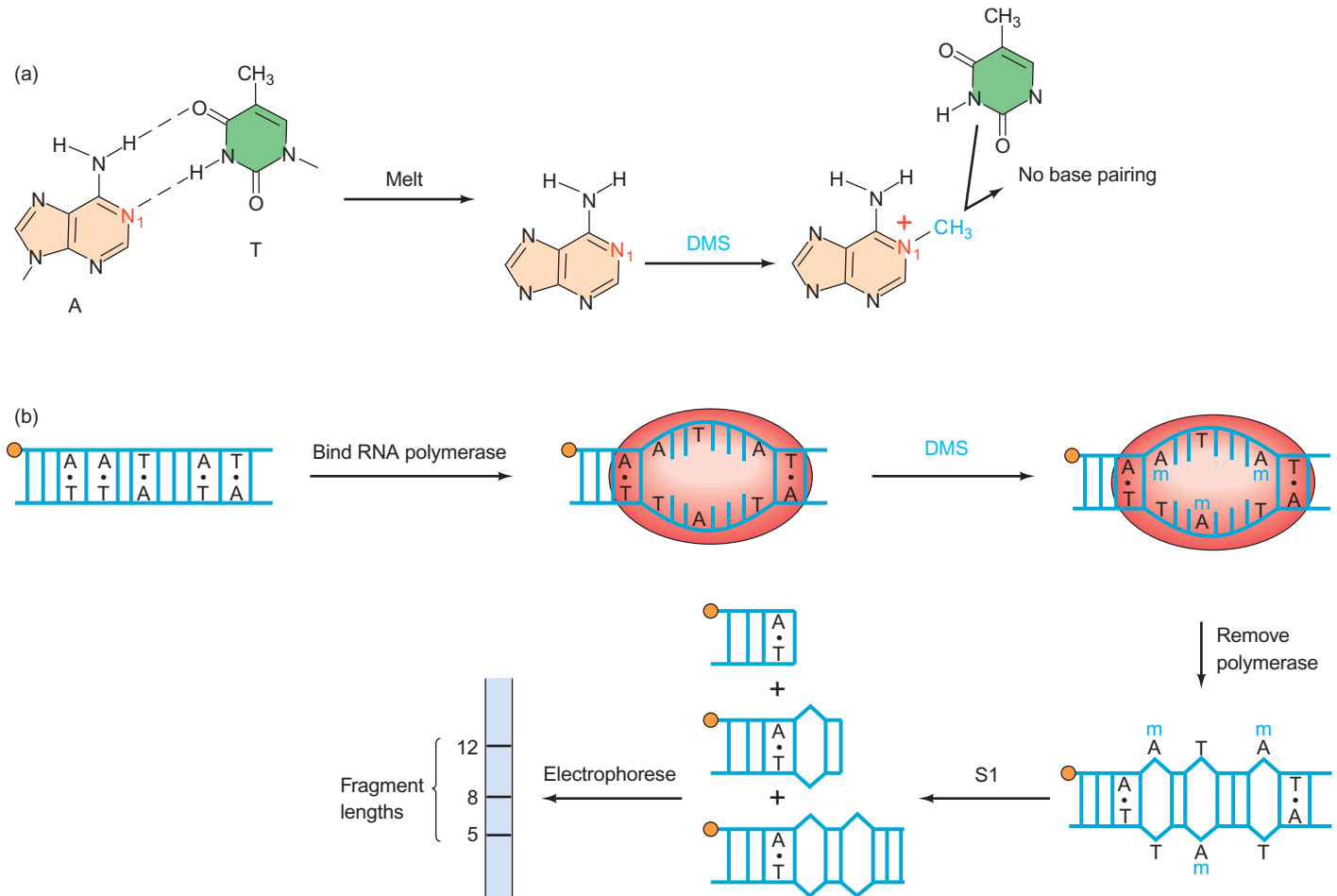


Figure 6.15 Locating the region of a T7 phage early promoter melted by RNA polymerase. (a) When adenine is base-paired with thymine (left) the N₁ nitrogen of adenine is hidden in the middle of the double helix and is therefore protected from methylation. On melting (right), the adenine and thymine separate; this opens the adenine up to attack by dimethyl sulfate (DMS, blue), and the N₁ nitrogen is methylated. Once this occurs, the methyl-adenine can no longer base-pair with its thymine partner. (b) A hypothetical promoter region containing five A-T base pairs is end-labeled (orange), then RNA polymerase (red) is bound, which causes local melting of the

promoter DNA. The three newly exposed adenines are methylated with dimethyl sulfate (DMS). Then, when the polymerase is removed, the A-T base pairs cannot reform because of the interfering methyl groups (m, blue). Now S1 nuclease can cut the DNA at each of the unmethylated base pairs because these are local single-stranded regions. Very mild cutting conditions are used so that only about one cut per molecule occurs. Otherwise, only the shortest product would be seen. The resulting fragments are denatured and electrophoresed to determine their sizes. These sizes tell how far the melted DNA region was from the labeled DNA end.

promoter and had become methylated. In principle, this should produce a series of end-labeled fragments, each one terminating at an adenine in the melted region. Finally, Siebenlist electrophoresed the labeled DNA fragments to determine their precise lengths. Then, knowing these lengths and the exact position of the labeled end, he could calculate accurately the position of the melted region.

Figure 6.16 shows the results. Instead of the expected neat set of fragments, we see a blur of several fragments extending from position +3 to -9. The reason for the blur seems to be that each of the multiple methylations in the melted region introduced a positive charge and therefore weakened base pairing so much that few strong base pairs could re-form; the whole melted region retained at least par-

tially single-stranded character and therefore remained open to cutting by S1 nuclease. The length of the melted region detected by this experiment is 12 bp, roughly in agreement with Hsieh and Wang's estimate, although this may be an underestimate because the next base pairs on either side are G-C pairs whose participation in the melted region would not have been detected. This is because neither guanines nor cytosines are readily methylated under the conditions used in this experiment. It is also satisfying that the melted region is just at the place where RNA polymerase begins transcribing.

The experiments of Hsieh and Wang, and of Siebenlist, as well as other early experiments, measured the DNA melting in a simple binary complex between polymerase and DNA. None of these experiments examined the size

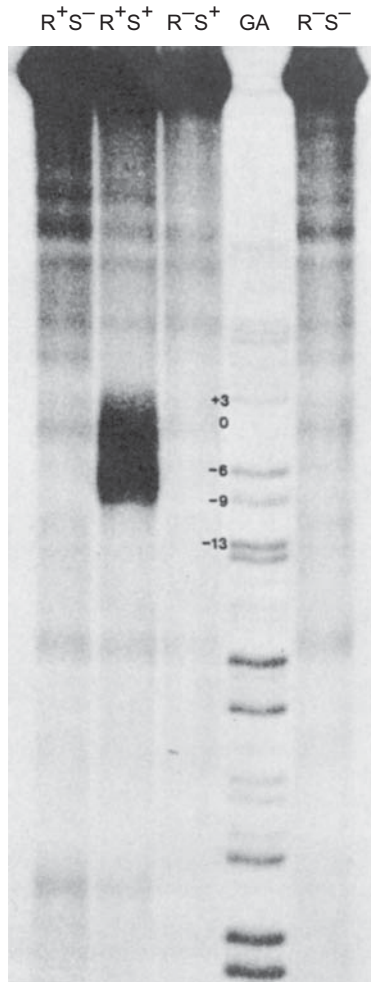


Figure 6.16 RNA polymerase melts the DNA in the -9 to $+3$ region of the T7 A3 promoter. Siebenlist performed a methylation-S1 assay as described in Figure 6.15. Lane R^+S^+ shows the results when both RNA polymerase (R) and S1 nuclease (S) were used. The other lanes were controls in which Siebenlist left out either RNA polymerase, or S1 nuclease, or both. The partial sequencing lane (GA) served as a set of markers and allowed him to locate the melted region approximately between positions -9 and $+3$. (Source: Siebenlist. RNA polymerase unwinds an 11-base pair segment of a phage T7 promoter. *Nature* 279 (14 June 1979) p. 652, f. 2, © Macmillan Magazines Ltd.)

of a DNA bubble in complexes in which initiation or elongation of RNA chains was actually taking place. Thus, in 1982, Howard Gamper and John Hearst set out to estimate the number of base pairs melted by polymerases, not only in binary complexes, but also in actively transcribing complexes that also contained RNA (ternary complexes). They used SV40 DNA, which happens to have one promoter site recognized by the *E. coli* RNA polymerase. They bound RNA polymerase to the SV40 DNA at either 5°C or 37°C in the absence of nucleotides to form binary complexes, or in the presence of nucleotides to form ternary complexes. Under the conditions of the experiment, each polymerase initiated only once, and no polymerase terminated

transcription, so all polymerases remained complexed to the DNA. This allowed an accurate assessment of the number of polymerases bound to the DNA.

After binding a known number of *E. coli* RNA polymerases to the DNA, Gamper and Hearst relaxed any supercoils that had formed with a crude extract from human cells, then removed the polymerases from the relaxed DNA (Figure 6.17a). The removal of the protein left melted regions of DNA, which meant that the whole DNA was underwound. Because the DNA was still a covalently closed circle, this underwinding introduced strain into the circle that was relieved by forming supercoils (Chapters 2 and 20). The higher the superhelical content, the greater the double helix unwinding that has been caused by the polymerase. The superhelical content of a DNA can be measured by gel electrophoresis because the more superhelical turns a DNA contains, the faster it will migrate in an electrophoretic gel.

Figure 6.17b is a plot of the change in the superhelicity as a function of the number of active polymerases per genome at 37°C . A linear relationship existed between these two variables, and one polymerase caused about 1.6 superhelical turns, which means that each polymerase unwound 1.6 turns of the DNA double helix. If a double helical turn contains 10.5 bp, then each polymerase melted about 17 bp ($1.6 \times 10.5 = 16.8$). A similar calculation of the data from the 5°C experiment yielded a value of 18 bp melted by one polymerase. From these data, Gamper and Hearst concluded that a polymerase binds at the promoter, melts 17 ± 1 bp of DNA to form a **transcription bubble**, and a bubble of this size moves with the polymerase as it transcribes the DNA. Subsequent experimental and theoretical work has suggested that the size of the transcription bubble actually increases and decreases within a range of approximately 11–16 nt, according to conditions, including the base sequence within the bubble. Larger bubbles can form, but their abundance decreases exponentially with size because of the energy required to melt more base pairs.

SUMMARY On binding to a promoter, RNA polymerase causes melting that has been estimated at 10–17 bp in the vicinity of the transcription start site. This transcription bubble moves with the polymerase, exposing the template strand so it can be transcribed.

Promoter Clearance

RNA polymerases cannot work if they do not recognize promoters, so they have evolved to recognize and bind strongly to them. But that poses a challenge when it comes time for promoter clearance: Somehow those strong bonds between polymerase and promoter must be broken in order for the polymerase to leave the promoter and enter the elongation phase. How can we explain that phenomenon?

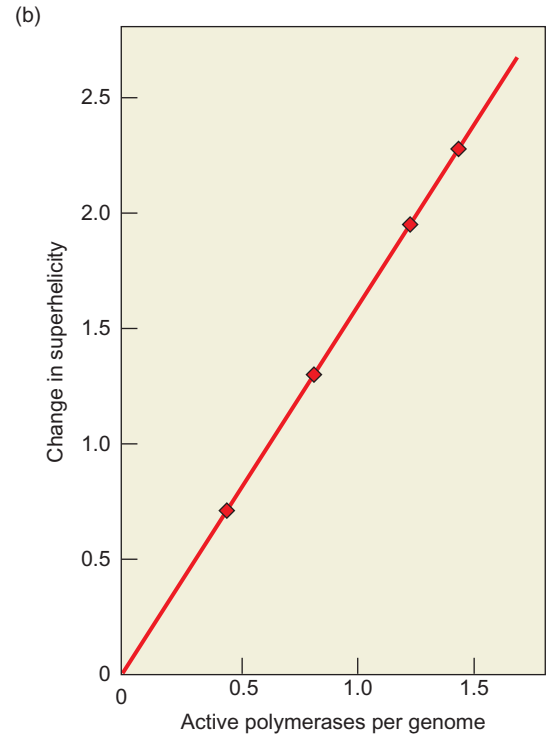
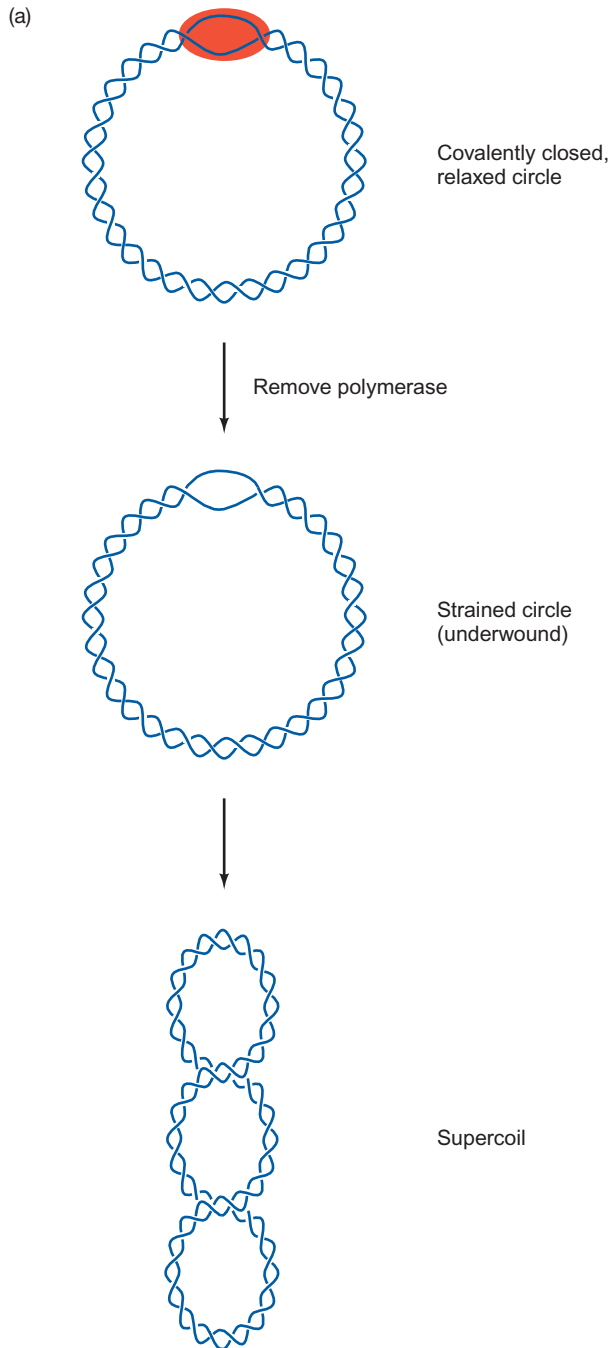


Figure 6.17 Measuring the melting of DNA by polymerase binding. (a) Principle of the experiment. Gamper and Hearst added *E. coli* RNA polymerase (red) to SV40 DNA, then relaxed any supercoils with a nicking-closing extract to produce the complexes shown at top. Then they removed the polymerase, leaving the DNAs strained (middle) because of the region that had been melted by the polymerase. This strain was quickly relieved by forming supercoils (bottom). The greater the superhelicity, the greater the unwinding caused by the polymerase. (b) Experimental results. Gamper and Hearst plotted the change in superhelicity of DNA as a function of the number of polymerases added. The plot was a straight line with a slope of 1.6 (1.6 superhelical turns introduced per polymerase).

Several hypotheses have been proposed, including the idea that the energy released by forming a short transcript (up to 10 nt long) is stored in a distorted polymerase or DNA, and the release of that energy in turn allows promoter clearance. However this process works, it is clearly not perfect, as it fails more often than not, giving rise to abortive transcripts.

The polymerase cannot move enough downstream to make a 10-nt transcript without doing one of three things: moving briefly downstream and then snapping back to the starting position (transient excursion); stretching itself by leaving its trailing edge in place while moving its leading

edge downstream (inchworming); or compressing the DNA without moving itself (**scrunching**). In 2006, Richard Ebright and colleagues applied two single-molecule strategies to show that scrunching appears to be the correct answer.

The first set of experiments used single-molecule FRET as described earlier in this chapter, but with a twist known as “FRET analysis with alternating-laser excitation” (**FRET-ALEX**). This adaptation can correct for the fact that the spectrum of a donor fluorophore depends on its exact protein environment, which can change during an experiment because proteins are dynamic molecules. This change in

spectrum can be perceived as a change in fluorescence energy, confusing the results. Ebright and colleagues examined both the leading and trailing edge of the *E. coli* RNA polymerase in complexes of polymerase attached to promoter DNA. For leading edge FRET, they tagged the leading edge of σ with the donor fluorophore and a downstream DNA site (position +20) with the acceptor. For trailing edge FRET, they tagged the trailing edge of σ with the donor and an upstream DNA site (position -39) with the acceptor fluorophore. They considered complexes only if they had a stoichiometry indicating the presence of both fluorophores.

They formed open promoter complexes (RP_o) by binding holoenzyme to a promoter DNA in the presence of the dinucleotide ApA (the first two nucleotides in the nascent transcript are A's). They formed initial transcribing complexes containing abortive transcripts up to 7 nt long ($RP_{itc \leq 7}$) by adding UTP and GTP in addition to ApA. This allowed the formation of the 7-mer AAUUGUG, but stopped because the next nucleotide called for was ATP, which was missing.

All three hypotheses predict the same result with leading edge FRET ALEX: All three should yield a decreased separation between the fluorophores, as illustrated in Figure 6.18a. Indeed, a comparison of RP_o and $RP_{itc \leq 7}$ showed an increase in FRET efficiency as the polymerase formed abortive transcripts up to 7 nt long, and therefore a decreased distance between fluorophores.

To begin to distinguish among the hypotheses, Ebright and colleagues performed trailing edge FRET ALEX (Figure 6.18b). Both the inchworming and scrunching models predict no change in the position of the trailing edge of the polymerase during abortive transcript production. But the transient excursion model predicts that the polymerase moves downstream in producing abortive transcripts and therefore $RP_{itc \leq 7}$ complexes should show a decrease in FRET efficiency relative to RP_o complexes. In fact, Ebright and colleagues observed no difference in FRET efficiency, ruling out the transient excursion model.

To distinguish between the inchworming and scrunching models, Ebright and colleagues placed the donor fluorophore on the leading edge of σ and the acceptor fluorophore on the DNA spacer between the -10 and -35 boxes of the promoter (Figure 6.18c). If the polymerase stretches, as the inchworming model predicts, the separation between fluorophores should increase, and the fluorescence efficiency should fall. On the other hand, the scrunching model predicts that downstream DNA is drawn into the enzyme, which should not change the separation between fluorophores. Indeed, the fluorescence efficiency did not change, supporting the scrunching model.

To check this result, Ebright and colleagues tested directly for the scrunching of DNA. They placed the donor fluorophore at DNA position -15, and the acceptor fluorophore in the downstream DNA, at position +15. If the

polymerase really does pull downstream DNA into itself, the distance between fluorophores on the DNA should decrease. Indeed, the fluorescence efficiency increased, supporting the scrunching hypothesis.

Thus, it may be the scrunched DNA that stores the energy expended in abortive transcript formation, rather like a spring, and enables the RNA polymerase finally to break away from the promoter and shift to the elongation phase. In another study, Ebright, Terence Strick, and colleagues used single-molecule DNA nanomanipulation to show that DNA scrunching indeed accompanies, and is probably required for, promoter clearance.

In this method, Ebright, Strick, and colleagues tethered a magnetic bead to one end of a piece of DNA, and a glass surface to the other (Figure 6.19). They made the DNA stick straight up from the glass surface by placing a pair of magnets above the magnetic bead. By rotating the magnets, they could rotate the DNA, introducing either positive or negative supercoils, depending on the direction of rotation. Then they added RNA polymerase, which bound to a promoter in the DNA. By adding different subsets of nucleotides, they could form either RP_o , $RP_{itc \leq 4}$, $RP_{itc \leq 8}$, or an elongation complex (RP_e). (With this promoter, addition of ATP and UTP leads to an abortive transcript up to 4 nt long, and addition of ATP, UTP, and CTP produces an abortive transcript up to 8 nt long.)

If scrunching occurs during abortive transcription, then the DNA will experience an extra unwinding, which causes a compensating loss of negative supercoiling, or gain of positive supercoiling. Every unwinding of one helical turn (about 10 bp) leads to loss of one negative, or gain of one positive, supercoil. The change in supercoiling can be measured as shown in Figure 6.19. Gain of one positive supercoil should decrease the apparent length (l) of the DNA (the distance between the bead and the glass surface) by 56 nm. Similarly, loss of one negative supercoil should increase l by 56 nm. Such changes in the position of the magnetic bead can be readily observed in real time by video-microscopy, yielding estimates of DNA unwinding with a resolution of about 1 bp.

Ebright, Strick, and colleagues observed the expected change in l upon converting RP_o to $RP_{itc \leq 4}$ and $RP_{itc \leq 8}$. Thus, unwinding of DNA accompanies formation of abortive transcripts, and the degree of unwinding depends on the length of the abortive transcript made. In particular, formation of abortive transcripts 4 and 8 nt long led to unwinding of 2 and 6 nt, respectively. This is consistent with the hypothesis that the active center of RNA polymerase can polymerize two nucleotides without moving relative to the DNA, but further RNA synthesis requires scrunching.

Does scrunching also accompany promoter clearance? To find out, Ebright, Strick, and colleagues looked at individual complexes over time: from the addition of polymerase and all four nucleotides until termination at a

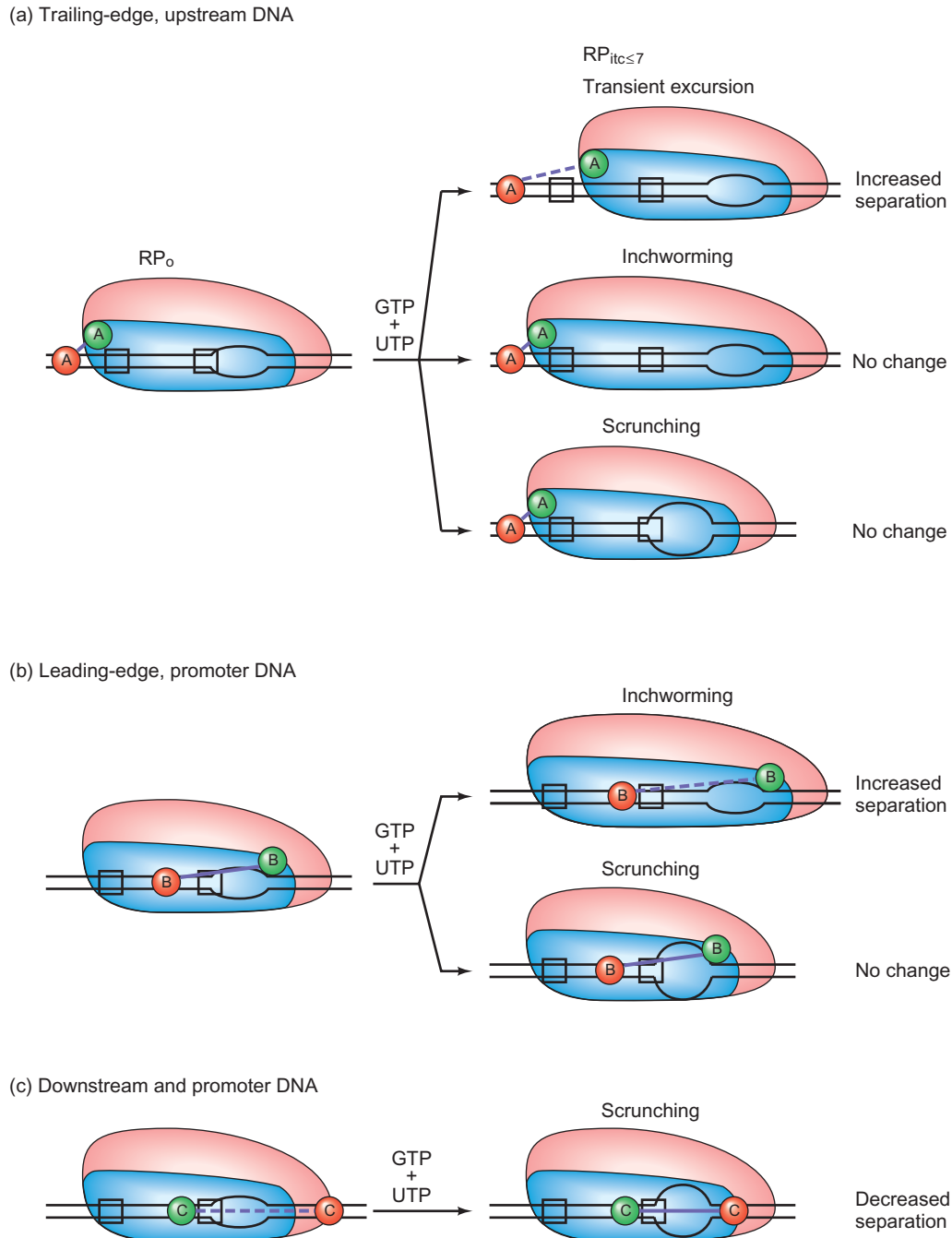


Figure 6.18 Evidence for DNA scrunching during abortive transcription. Ebright and colleagues used single-molecule FRET ALEX to distinguish among three hypotheses for the mechanism of abortive transcription: transient excursion, inchworming, and scrunching. They compared the average efficiency of single-molecule FRET of RP_0 and $RP_{itc \leq 7}$ complexes of *E. coli* RNA polymerase with promoter DNA. The latter complexes contained abortive transcripts up to 7 nt in length and were created by allowing transcription in the presence of the primer ApA

plus UTP and GTP. ATP is required in the eighth position, limiting the abortive transcripts to 7 nt. The position of the donor fluorophore is denoted in green, and the acceptor fluorophore in red, throughout. High-efficiency FRET, indicating short distance between fluorophores, is denoted by a solid purple line throughout. Lower-efficiency FRET, indicating a greater distance between fluorophores, is denoted by a dashed purple line throughout. The three experiments depicted in panels (a)–(c) are described in the text. The boxes represent the -10 and -35 boxes of the promoter.

terminator either 100 or 400 bp downstream of the promoter. In fact, since reinitiation could occur, the investigators could look at multiple rounds of transcription on each DNA. They found a four-phase pattern that repeated over

and over with each round. Considering a positively supercoiled DNA: First, the superhelicity increased, reflecting the DNA unwinding that occurs during RP_0 formation. Second, the superhelicity increased still further, reflecting the

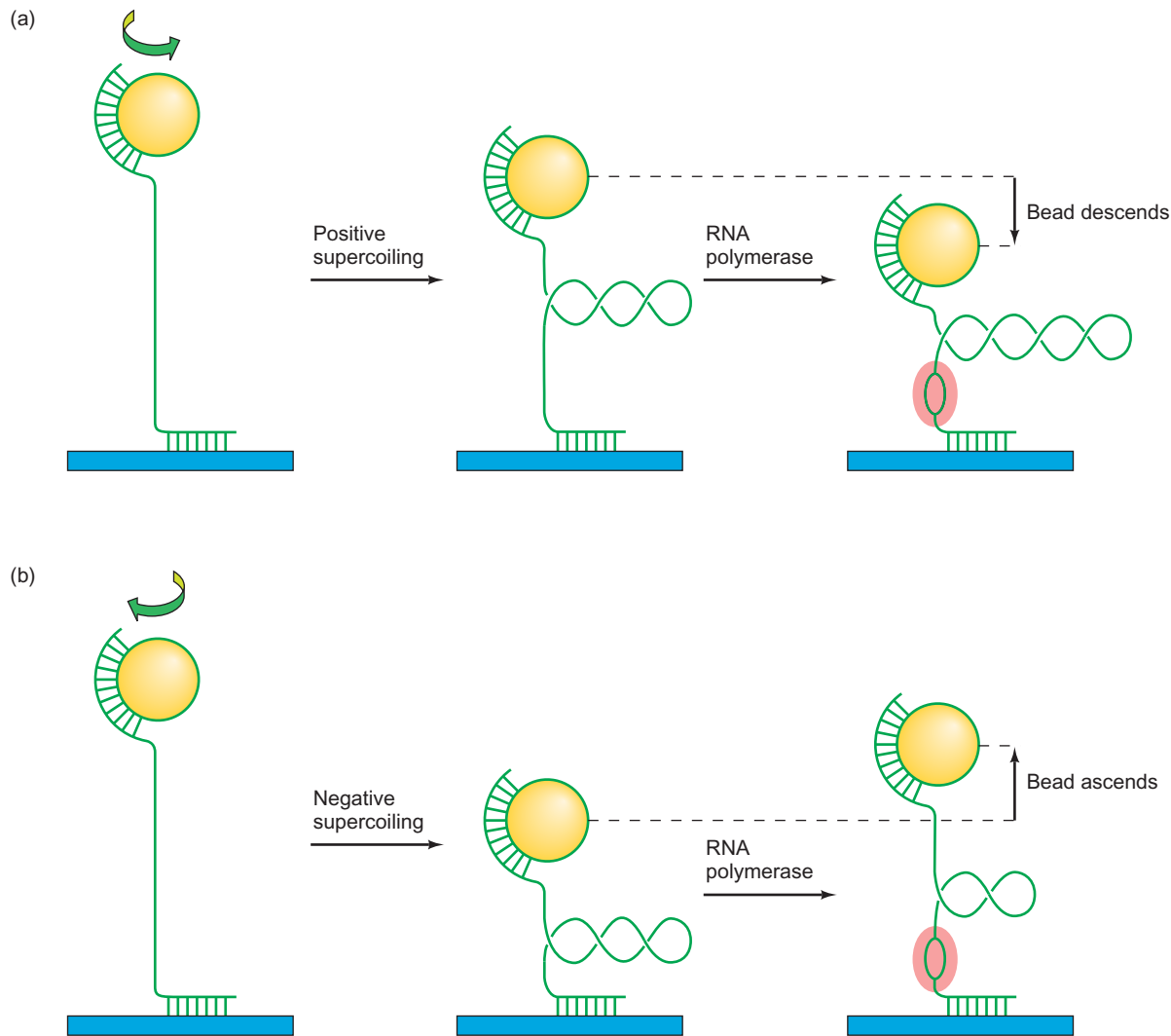


Figure 6.19 Basis of single-molecule nanomanipulation procedure. One end of a promoter-containing piece of DNA is tethered to a magnetic bead (yellow), and the other end is tethered to a glass surface (blue). A pair of magnets at the top extend the DNA vertically, and introduce a rightward (a) or leftward (b) twist to the bead, and therefore to the DNA. Every full turn of the bead introduces one superhelical turn into the DNA. The supercoiling is

positive in (a) and negative in (b). When RNA polymerase (pink) is added to the DNA, it binds to the promoter and unwinds about one double-helical turn of DNA, which adds one positive supercoil (a), which drags the magnetic bead down about 56 nm for every such supercoil. Similarly, unwinding of promoter DNA by the polymerase subtracts one negative supercoil (b). These changes in bead position are detected by videomicroscopy.

scrunching that occurs during RP_{itc} formation. Third, the superhelicity decreased, reflecting the reversal of scrunching during promoter clearance and RP_e formation. Finally, the superhelicity decreased back to the original level, reflecting the loss of RNA polymerase at termination. The amount of scrunching observed in these experiments was 9 ± 2 bp, which is within experimental error of the amount expected: Promoter clearance at this promoter was known to occur upon formation of an 11-nt transcript, 9 nt of which should require 9 bp of DNA scrunching, and 2 nt of which the polymerase can synthesize without scrunching.

Eighty percent of the transcription cycles studied had detectable scrunches. But 20% of the cycles were predicted

to have scrunches that lasted less than 1 s, and 1 s was the limit of resolution in these experiments. So this 20% of cycles probably also had scrunches. The authors concluded that approximately 100% of all the transcription cycles involve scrunching, which suggests that scrunching is required for promoter clearance.

E. coli RNA polymerase was used in all these studies, but the similarity among RNA polymerases, the strength of binding between polymerases and promoters, and the necessity to break that binding to start productive transcription, all suggest that scrunching could be a general phenomenon, and could be universally required for promoter clearance.

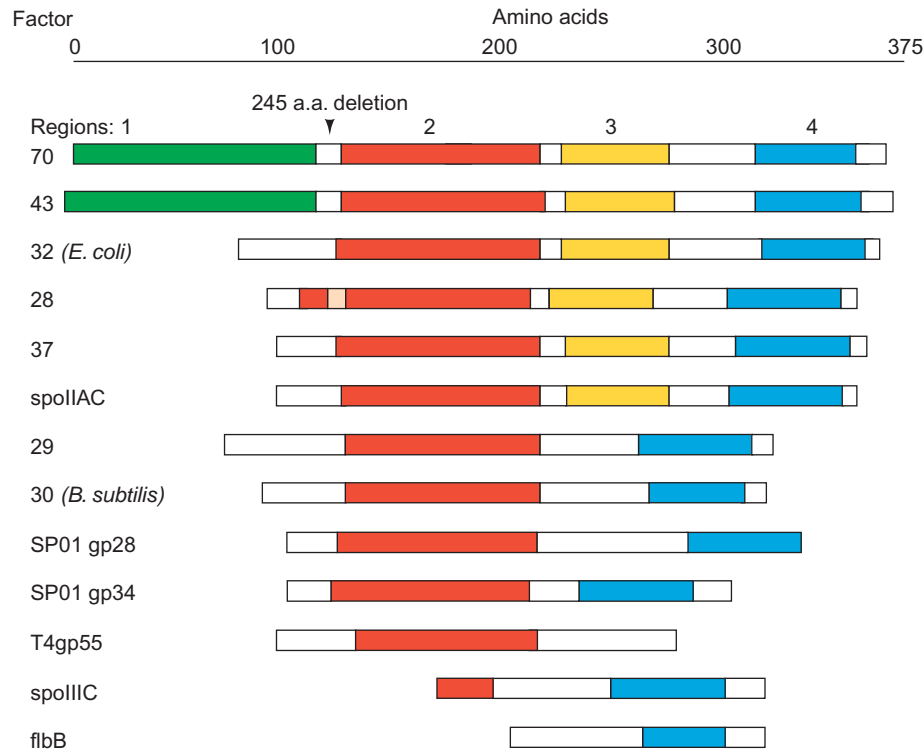


Figure 6.20 Homologous regions in various *E. coli* and *B. subtilis* σ -factors. The σ proteins are represented as horizontal bars, with homologous regions aligned vertically. Only the top two, the primary σ -factors of *E. coli* and *B. subtilis*, respectively, contain the first homologous region. Also, σ^{70} contains a sequence of 245 amino acids between regions 1 and 2 that is missing in σ^{43} . This is marked above the σ^{70} bar. Lighter shading denotes an area that is conserved only in some of the proteins.

SUMMARY The *E. coli* RNA polymerase achieves abortive transcription by scrunching: drawing downstream DNA into the polymerase without actually moving and losing its grip on promoter DNA. The scrunched DNA could store enough energy to allow the polymerase to break its bonds to the promoter and begin productive transcription.

Structure and Function of σ

By the late 1980s, the genes encoding a variety of σ -factors from various bacteria had been cloned and sequenced. As we will see in Chapter 8, each bacterium has a primary σ -factor that transcribes its vegetative genes—those required for everyday growth. For example, the primary σ in *E. coli* is called σ^{70} , and the primary σ in *B. subtilis* is σ^{43} . These proteins are named for their molecular masses, 70 and 43 kD, respectively, and they are also called σ^A because of their primary nature. In addition, bacteria have alternative σ -factors that transcribe specialized genes (heat shock genes, sporulation genes, and so forth). In 1988, Helmann and Chamberlin reviewed the literature on all these factors and analyzed the striking similarities in amino acid sequence among them, which are clustered

in four regions (regions 1–4, see Figure 6.20). The conservation of sequence in these regions suggests that they are important in the function of σ , and in fact they are all involved in binding to core and positively or negatively, in binding to DNA. Helmann and Chamberlin proposed the following functions for each region.

Region 1 This region is found only in the primary σ 's (σ^{70} and σ^{43}). Its role appears to be to prevent σ from binding by itself to DNA. We will see later in this chapter that a fragment of σ is capable of DNA binding, but region 1 prevents the whole polypeptide from doing that. This is important because free σ binding to promoters could inhibit holoenzyme binding and thereby inhibit transcription.

Region 2 This region is found in all σ -factors and is the most highly conserved σ region. It can be subdivided into four parts, 2.1–2.4 (Figure 6.21).

We have good evidence that region 2.4 is responsible for a crucial σ activity, recognition of the promoter's -10 box. First of all, if σ region 2.4 does recognize the -10 box, then σ 's with similar specificities should have similar regions 2.4. This is demonstrable; σ^{43} of *B. subtilis* and σ^{70} of *E. coli* recognize identical promoter sequences, including -10 boxes. Indeed, these two σ 's are interchangeable. And the regions 2.4 of these two σ 's are 95% identical.

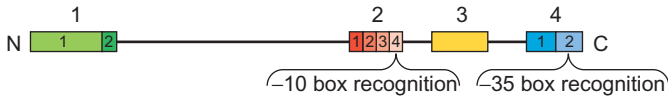


Figure 6.21 Summary of regions of primary structure in *E. coli* σ^{70} . The four conserved regions are indicated, with subregions delineated in regions 1, 2, and 4. (Source: Adapted from Dombroski, A.J., et al., "Polypeptides containing highly conserved regions of the transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA." *Cell* 70:501–12, 1992.)

Richard Losick and colleagues performed genetic experiments that also link region 2.4 with -10 box binding. Region 2.4 of the σ -factor contains an amino acid sequence that suggests it can form an α -helix. We will learn in Chapter 9 that an α -helix is a favorite DNA-binding motif, which is consistent with a role for this part of the σ in promoter binding. Losick and colleagues reasoned as follows: If this potential α -helix is really a -10 box-recognition element, then the following experiment should be possible. First, they could make a single base change in a promoter's -10 box, which destroys its ability to bind to RNA polymerase. Then, they could make a compensating mutation in one of the amino acids in region 2.4 of the σ -factor. If the σ -factor mutation can suppress the promoter mutation, restoring binding to the mutated promoter, it provides strong evidence that there really is a relationship between the -10 box and region 2.4 of the σ . So Losick and colleagues caused a G \rightarrow A transition in the -10 box of the *B. subtilis* *spoVG* promoter, which prevented binding between the promoter and RNA polymerase. Then they caused a Thr \rightarrow Ile mutation at amino acid 100 in region 2.4 of σ^H , which normally recognizes the *spoVG* promoter. This σ mutation restored the ability of the polymerase to recognize the mutant promoter.

Region 3 We will see later in this chapter that region 3 is involved in both core and DNA binding.

Region 4 Like region 2, region 4 can be subdivided into subregions. Also like region 2, region 4 seems to play a key role in promoter recognition. Subregion 4.2 contains a helix-turn-helix DNA-binding domain (Chapter 9), which

suggests that it plays a role in polymerase–DNA binding. In fact, subregion 4.2 appears to govern binding to the -35 box of the promoter. As with the σ region 2.4 and the -10 box, genetic and other evidence supports the relationship between the σ region 4.2 and the -35 box. Again, we see that σ 's that recognize promoters with similar -35 boxes have similar regions 4.2. And again, we observe suppression of mutations in the promoter (this time in the -35 box) by compensating mutations in region 4.2 of the σ -factor. For instance, Miriam Susskind and her colleagues showed that an Arg \rightarrow His mutation in position 588 of the *E. coli* σ^{70} suppresses G \rightarrow A or G \rightarrow C mutations in the -35 box of the *lac* promoter. Figure 6.22 summarizes this and other interactions between regions 2.4 and 4.2 of σ and the -10 and -35 boxes, respectively, of bacterial promoters.

These results all suggest the importance of σ regions 2.4 and 4.2 in binding to the -10 and -35 boxes, respectively, of the promoter. The σ -factor even has putative DNA-binding domains in strategic places. But we are left with the perplexing fact that σ by itself does not bind to promoters, or to any other region of DNA. Only when it is bound to the core can σ bind to promoters. How do we resolve this apparent paradox?

Carol Gross and her colleagues suggested that regions 2.4 and 4.2 of σ are capable of binding to promoter regions on their own, but other domains in σ interfere with this binding. In fact, we now know that region 1.1 prevents σ from binding to DNA in the absence of core. Gross and colleagues further suggested that when σ associates with core it changes conformation, unmasking its DNA-binding domains, so it can bind to promoters. To test this hypothesis, these workers made fusion proteins (Chapter 4) containing glutathione-S-transferase (GST) and fragments of the *E. coli* σ -factor (region 2.4, or 4.2, or both). (These fusion proteins are easy to purify because of the affinity of GST for glutathione.) Then they showed that a fusion protein containing region 2.4 could bind to a DNA fragment containing a -10 box, but not a -35 box. Furthermore, a fusion protein containing region 4.2 could bind to a DNA fragment containing a -35 box, but not a -10 box.

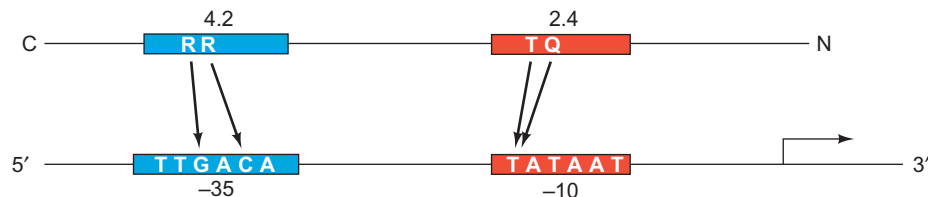


Figure 6.22 Specific interactions between σ regions and promoter regions. Arrows denote interactions revealed by mutation suppression experiments involving σ^{70} . The letters in the upper bar, representing the σ^{70} protein show the amino acid mutated and the arrows point to bases in the promoter that the respective amino acids in σ^{70} appear to contact. The two R's in σ^{70} region 4.2 represent arginines 584 and 588 (the 584th and 588th amino acids in the protein), and these amino acids contact a C and a G, respectively, in the -35 box of the

promoter. The Q and T in the σ^{70} 2.4 region represent glutamine 437 and threonine 440, respectively, both of which contact a T in the -10 box of the promoter. Notice that the linear structure of the σ -factor (top) is written with the C-terminus at left, to match the promoter written conventionally, 5' \rightarrow 3' left to right (bottom). (Source: Adapted from Dombroski, A.J., et al., "Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA." *Cell* 70:501–12, 1992.)

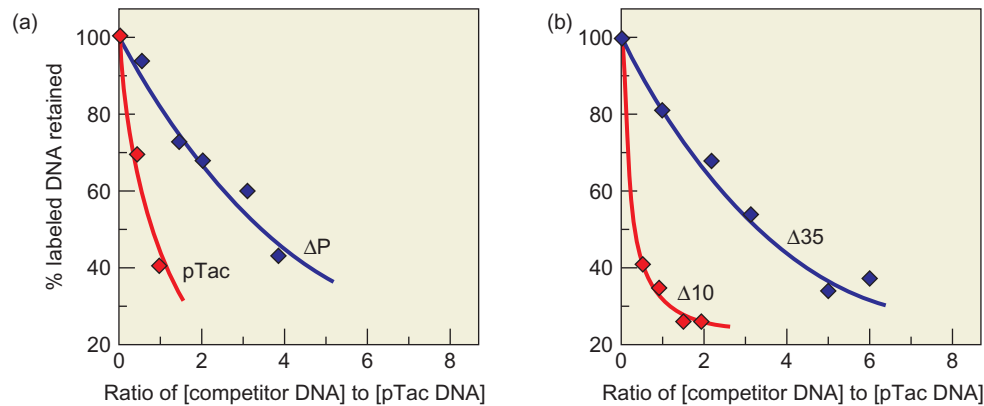


Figure 6.23 Analysis of binding between σ region 4.2 and the promoter –35 box. (a) Recognition of the promoter. Gross and colleagues measured binding between a σ fragment-GST fusion protein and a labeled DNA fragment (pTac) containing the *tac* promoter. The σ fragment in this experiment contained only the 108 amino acids at the C-terminus of the *E. coli* σ , which includes region 4, but not region 2. Gross and coworkers measured binding of the labeled DNA–protein complex to nitrocellulose filters in the presence of competitor DNA containing the *tac* promoter (pTac), or lacking the *tac* promoter (ΔP). Because pTac DNA competes much better than ΔP DNA, they concluded that the fusion protein with region 4 can bind to the *tac*

promoter. (b) Recognition of the –35 region. Gross and colleagues repeated the experiment but used two different competitor DNAs: One ($\Delta 10$) had a *tac* promoter with a 6-bp deletion in the –10 box; the other ($\Delta 35$) had a *tac* promoter with a 6-bp deletion in the –35 box. Because deleting the –35 box makes the competitor no better than a DNA with no *tac* promoter at all and removing the –10 box had no effect, it appears that the σ fragment with region 4 binds to the –35 box, but not to the –10 box. (Source: Adapted from Dombroski, A.J., et al., “Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA.” *Cell* 70:501–12, 1992.)

To measure the binding between fusion proteins and promoter elements, Gross and coworkers used a nitrocellulose filter-binding assay. They labeled the target DNA containing one or both promoter elements from the composite *tac* promoter. The *tac* promoter has the –10 box of the *lac* promoter and the –35 box of the *trp* promoter. Then they added a fusion protein to the labeled target DNA in the presence of excess unlabeled competitor DNA and measured the formation of a labeled DNA–protein complex by nitrocellulose binding.

Figure 6.23a shows the results of an experiment in which Gross and colleagues bound a labeled *tac* promoter to a GST– σ -region 4 fusion protein. Because σ -region 4 contains a putative –35 box-binding domain, we expect this fusion protein to bind to DNA containing the *tac* promoter more strongly than to DNA lacking the *tac* promoter. Figure 6.23a demonstrates this is just what happened. Unlabeled DNA containing the *tac* promoter was an excellent competitor, whereas unlabeled DNA missing the *tac* promoter competed relatively weakly. Thus, the GST– σ region 4 protein binds weakly to nonspecific DNA, but strongly to *tac* promoter-containing DNA, as we expect.

Figure 6.23b shows that the binding between the GST– σ region 4 proteins and the promoter involves the –35 box, but not the –10 box. As we can see, a competitor from which the –35 box was deleted competed no better than nonspecific DNA, but a competitor from which the –10 box was deleted competed very well because it still contained the –35 box. Thus, σ region 4 can bind specifically to the –35 box, but not to the –10 box. Similar experiments with a GST– σ region 2 fusion protein showed

that this protein can bind specifically to the –10 box, but not the –35 box.

We have seen that the polymerase holoenzyme can recognize promoters and form an open promoter complex by melting a short region of the DNA, approximately between positions –11 and +1. We suspect that σ plays a big role in this process, but we know that σ cannot form an open promoter complex on its own. One feature of open complex formation is binding of polymerase to the nontemplate strand in the –10 region of the promoter. Again, σ cannot do this on its own so, presumably, some part of the core enzyme is required to help σ with this task. Gross and colleagues have posed the question: What part of the core enzyme is required to unmask the part of σ that binds to the nontemplate strand in the –10 region of the promoter?

To answer this question, Gross and colleagues focused on the β' subunit, which had already been shown to collaborate with σ in binding to the nontemplate strand in the –10 region. They cloned different segments of the β' subunit, then tested these, together with σ , for ability to bind to radiolabeled single-stranded oligonucleotides corresponding to the template and nontemplate strands in the –10 region of a promoter. They incubated the β' segments, along with σ , with the labeled DNAs, then subjected the complexes to UV irradiation to cross-link σ to the DNA. Then they performed SDS-PAGE on the cross-linked complexes. If the β' fragment induced binding between σ and the DNA, then σ would be cross-linked to the labeled DNA and the SDS-PAGE band corresponding to σ would become labeled.

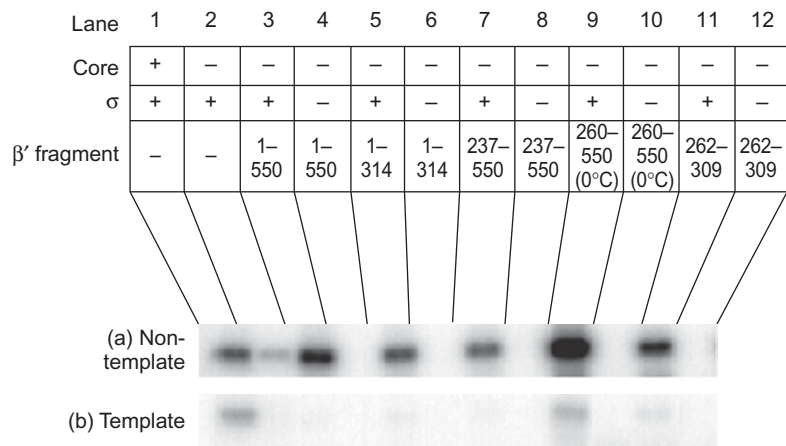


Figure 6.24 Induction of σ binding to the -10 region of a promoter. Gross and colleagues mixed σ plus various fragments of β' , as indicated at top, with labeled oligonucleotides representing either the nontemplate or template strand in the -10 region of the promoter. Then they UV-irradiated the complexes to cross-link any σ -subunit bound to the DNA, subjected the complexes to SDS-PAGE, and performed autoradiography to detect σ bound to labeled DNA. Lane 1 is a positive

control with whole core instead of a β' fragment; lane 2 is a control with no β' fragment; and all the other even-numbered lanes are negative controls with no protein. The experiments in lanes 9 and 10 were performed at 0°C; all other experiments were performed at room temperature. The autoradiography results are shown for experiments with (a) the nontemplate strand and (b) the template strand. (Source: Reprinted from *Cell* v. 105, Young et al., p. 940 © 2001, with permission from Elsevier Science.)

Figure 6.24 shows that the fragment of β' containing amino acids 1–550 caused binding between σ and the nontemplate strand DNA (but not the template strand), whereas σ by itself showed little binding. Next, Gross and colleagues used smaller fragments of the 1–550 region to pinpoint the part of β' that was inducing the binding. All of the fragments illustrated in Figure 6.24 could induce binding, although the 260–550 fragment would work only at low temperature. Strikingly, the very small 262–309 fragment, with only 48 amino acids, could stimulate binding very actively, even at room temperature. Mutations in three amino acids in this region (R275, E295, and A302) were already known to interfere with σ binding to promoters. Accordingly, Gross and colleagues tested these mutations for interference with σ binding to the nontemplate strand in the -10 region. In every case, these mutations caused highly significant interference.

The Role of the α -Subunit in UP Element Recognition

As we learned earlier in this chapter, RNA polymerase itself can recognize an upstream promoter element called an UP element. We know that the σ -factor recognizes the core promoter elements, but which polymerase subunit is responsible for recognizing the UP element? Based on the following evidence, it appears to be the α -subunit of the core polymerase.

Richard Gourse and colleagues made *E. coli* strains with mutations in the α -subunit and found that some of these were incapable of responding to the UP element—they gave no more transcription from promoters with UP elements than from those without UP elements. To measure transcription, they placed a wild-type form of the very strong *rrnB* P1 promoter, or a mutant form that was missing its UP element, about 170 bp upstream of an *rrnB* P1 transcription terminator in a cloning vector. They transcribed these constructs with three different RNA polymerases, all of which had been reconstituted from purified subunits: (1) wild-type polymerase with a normal α -subunit; (2) α -235, a polymerase whose α -subunit was missing 94 amino acids from its C-terminus; and (3) R265C, a polymerase whose α -subunit contained a cysteine (C) in place of the normal arginine (R) at position 265. They included a labeled nucleotide to label the RNA, then subjected this RNA to gel electrophoresis, and finally performed autoradiography to visualize the RNA products.

Figure 6.25a depicts the results with wild-type polymerase. The wild-type promoter (lanes 1 and 2) allowed a great deal more transcription than the same promoter with vector DNA substituted for its UP element (lanes 3 and 4), or having its UP element deleted (lanes 5 and 6). Figure 6.25b shows the same experiment with the polymerase with 94

SUMMARY Comparison of the sequences of different σ genes reveals four regions of similarity among a wide variety of σ -factors. Subregions 2.4 and 4.2 are involved in promoter -10 box and -35 box recognition, respectively. The σ -factor by itself cannot bind to DNA, but interaction with core unmasks a DNA-binding region of σ . In particular, the region between amino acids 262 and 309 of β' stimulates σ binding to the nontemplate strand in the -10 region of the promoter.

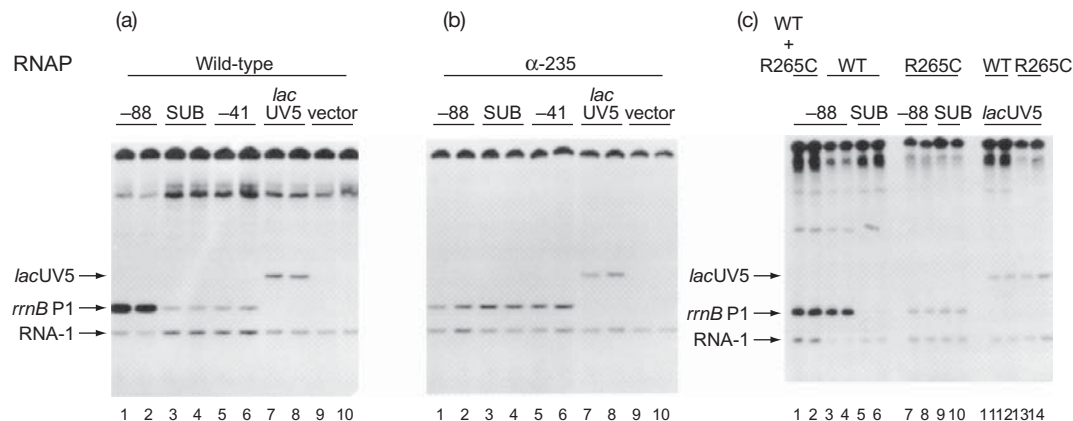


Figure 6.25 Importance of the α -subunit of RNA polymerase in UP element recognition. Gourse and colleagues performed *in vitro* transcription on plasmids containing the promoters indicated at top. They placed the promoters between 100 and 200 nt upstream of a transcription terminator to produce a transcript of defined size. After the reaction, they subjected the labeled transcripts to gel electrophoresis and detected them by autoradiography. The promoters were as follows: -88 contained wild-type sequence throughout the region between positions -88 and $+1$; SUB contained an irrelevant sequence instead of the UP element between positions -59 and -41 ; -41 lacked the UP element upstream of position -41 and had vector

sequence instead; *lacUV5* is a *lac* promoter without an UP element; vector indicates a plasmid with no promoter inserted. The positions of transcripts from the *rrnB* P1 and *lacUV5* promoters, as well as an RNA (RNA-1) transcribed from the plasmid's origin of replication, are indicated at left. RNAP at top indicates the RNA polymerase used, as follows: **(a)** Wild-type polymerase used throughout. **(b)** α -235 polymerase (missing 94 C-terminal amino acids of the α -subunit) used throughout. **(c)** Wild-type (WT) polymerase or R265C polymerase (with cysteine substituted for arginine 265) used, as indicated. (Source: Ross et al., A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* 262 (26 Nov 1993) f. 2, p. 1408. © AAAS.)

C-terminal amino acids missing from its α -subunit. We see that this polymerase is just as active as the wild-type polymerase in transcribing a gene with a core promoter (compare panels a and b, lanes 3–6). However, in contrast to the wild-type enzyme, this mutant polymerase did not distinguish between promoters with and without an UP element (compare lanes 1 and 2 with lanes 3–6). The UP element provided no benefit at all. Thus, it appears that the C-terminal portion of the α -subunit enables the polymerase to respond to an UP element.

Figure 6.25c demonstrates that the polymerase with a cysteine in place of an arginine at position 265 of the α -subunit (R265C) does not respond to the UP element (lanes 7–10 all show modest transcription). Thus, this single amino acid change appears to destroy the ability of the α -subunit to recognize the UP element. This phenomenon was not an artifact caused by an inhibitor in the R265C polymerase preparation because a mixture of R265C and the wild-type polymerase still responded to the UP element (lanes 1–4 all show strong transcription).

To test the hypothesis that the α -subunit actually contacts the UP element, Gourse and coworkers performed DNase footprinting experiments (Chapter 5) with DNA containing the *rrnB* P1 promoter and either wild-type or mutant RNA polymerase. They found that the wild-type polymerase made a footprint in the core promoter and the UP element, but that the mutant polymerase lacking the C-terminal domain of the α -subunit made a footprint in the core promoter only (data not shown). This indicates that the α -subunit C-terminal domain is required for interaction between polymerase and UP elements. Further evidence for this hypothesis came from an experiment in

which Gourse and coworkers used purified α -subunit dimers to footprint the UP element of the *rrnB* P1 promoter. Figure 6.26 shows the results—a clear footprint in the UP element caused by the α -subunit dimer all by itself.

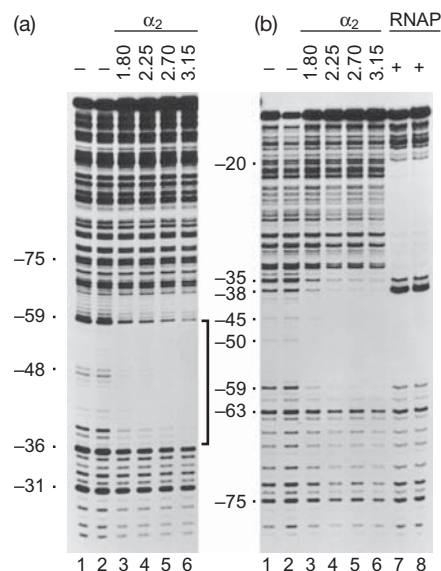


Figure 6.26 Footprinting the UP element with pure α -subunit.

Gourse and colleagues performed DNase footprinting with end-labeled template strand **(a)** or nontemplate strand **(b)** from the *rrnB* P1 promoter. They used the amounts listed at top (in micrograms) of purified α -dimers, or 10 nM RNA polymerase holoenzyme (RNAP). The bold brackets indicate the footprints in the UP element caused by the α -subunit, and the thin bracket indicates the footprint caused by the holoenzyme.

(Source: Ross et al., A third recognition element in bacterial promoter: DNA binding by the α -subunit of RNA polymerase. *Science* 262 (26 Nov 1993) f. 5, p. 1408. © AAAS.)

Richard Gourse, Richard Ebright, and their colleagues used **limited proteolysis** analysis to show that the α -subunit N-terminal and C-terminal domains (the α -NTD and α -CTD, respectively) fold independently to form two domains that are tethered together by a flexible linker. A protein **domain** is a part of a protein that folds independently to form a defined structure. Because of their folding, domains tend to resist proteolysis, so limited digestion with a proteolytic enzyme will attack unstructured elements between domains and leave the domains themselves alone. When Gourse and Ebright and collaborators performed limited proteolysis on the *E. coli* RNA polymerase α -subunit, they released a polypeptide of about 28 kD, and three polypeptides of about 8 kD. The sequences of the ends of these products showed that the 28-kD polypeptide contained amino acids 8–241, whereas the three small polypeptides contained amino acids 242–329, 245–329, and 249–329. This suggested that the α -subunit folds into two domains: a large N-terminal domain encompassing (approximately) amino acids 8–241, and a small C-terminal domain including (approximately) amino acids 249–329.

Furthermore, these two domains appear to be joined by an unstructured linker that can be cleaved in at least three places by the protease used in this experiment (Glu-C). This linker seems at first glance to include amino acids 242–248. Because Glu-C requires three unstructured amino acids on either side of the bond that is cleaved, however, the linker is longer than it appears at first. In fact, it must be at least 13 amino acids long (residues 239–251).

These experiments suggest a model such as the one presented in Figure 6.27. RNA polymerase binds to a core promoter via its σ -factor, with no help from the C-terminal domains of its α -subunits, but it binds to a promoter with an UP element using σ plus the α -subunit C-terminal domains. This allows very strong interaction between polymerase and promoter and therefore produces a high level of transcription.

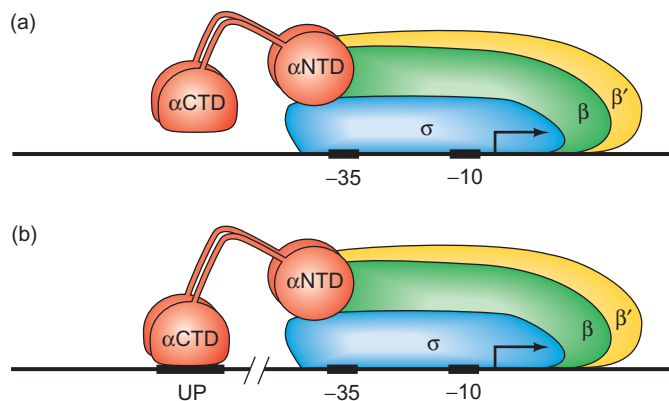


Figure 6.27 Model for the function of the C-terminal domain (CTD) of the polymerase α -subunit. (a) In a core promoter, the α -CTDs are not used, but (b) in a promoter with an UP element, the α -CTDs contact the UP element. Notice that two α -subunits are depicted: one behind the other.

SUMMARY The RNA polymerase α -subunit has an independently folded C-terminal domain that can recognize and bind to a promoter's UP element. This allows very tight binding between polymerase and promoter.

6.4 Elongation

After initiation of transcription is accomplished, the core continues to elongate the RNA, adding one nucleotide after another to the growing RNA chain. In this section we will explore this elongation process.

Core Polymerase Functions in Elongation

So far we have been focusing on the role of σ because of the importance of this factor in determining the specificity of initiation. However, the core polymerase contains the RNA synthesizing machinery, so the core is the central player in elongation. In this section we will see evidence that the β - and β' -subunits are involved in phosphodiester bond formation, that these subunits also participate in DNA binding, and that the α -subunit has several activities, including assembly of the core polymerase.

The Role of β in Phosphodiester Bond Formation Walter Zillig was the first to investigate the individual core subunits, in 1970. He began by separating the *E. coli* core polymerase into its three component polypeptides and then combining them again to reconstitute an active enzyme. The separation procedure worked as follows: Alfred Heil and Zillig electrophoresed the core enzyme on cellulose acetate in the presence of urea. Like SDS, urea is a denaturing agent that can separate the individual polypeptides in a complex protein. Unlike SDS, however, urea is a mild denaturant that is relatively easy to remove. Thus, it is easier to renature a urea-denatured polypeptide than an SDS-denatured one. After electrophoresis was complete, Heil and Zillig cut out the strips of cellulose acetate containing the polymerase subunits and spun them in a centrifuge to drive the buffer, along with the protein, out of the cellulose acetate. This gave them all three separated polypeptides, which they electrophoresed individually to demonstrate their purity (Figure 6.28).

Once they had separated the subunits, they recombined them to form active enzyme, a process that worked best in the presence of σ . Using this separation–reconstitution system, Heil and Zillig could mix and match the components from different sources to answer questions about their functions. For example, recall that the core polymerase determines sensitivity or resistance to the antibiotic rifampicin, and that rifampicin blocks transcription initiation.

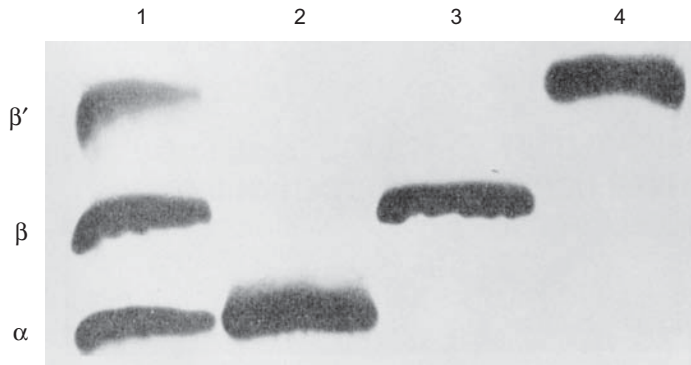


Figure 6.28 Purification of the individual subunits of *E. coli* RNA polymerase. Heil and Zillig subjected the *E. coli* core polymerase to urea gel electrophoresis on cellulose acetate, then collected the separated polypeptides. Lane 1, core polymerase after electrophoresis; lane 2, purified α ; lane 3, purified β ; lane 4, purified β' . (Source: Heil, A. and Zillig, W. Reconstitution of bacterial DNA-dependent RNA-polymerase from isolated subunits as a tool for the elucidation of the role of the subunits in transcription. *FEBS Letters* 11 (Dec 1970) p. 166, f. 1.)

Separation and reconstitution of the core allowed Heil and Zillig to ask which core subunit confers this antibiotic sensitivity or resistance. When they recombined the α -, β' -, and σ -subunits from a rifampicin-sensitive bacterium with the β -subunit from a rifampicin-resistant bacterium, the resulting polymerase was antibiotic-resistant (Figure 6.29). Conversely, when the β -subunit came from an antibiotic-sensitive bacterium, regardless of the origin of the other subunits. Thus, the β -subunit is obviously the determinant of rifampicin sensitivity or resistance.

Another antibiotic, known as *streptolydigin*, blocks RNA chain elongation. By the same separation and reconstitution strategy used for rifampicin, Heil and Zillig showed that the β -subunit also governed streptolydigin

resistance or sensitivity. At first this seems paradoxical. How can the same core subunit be involved in both initiation and elongation? The answer, which we will discuss in detail later in this chapter, is that rifampicin actually blocks *early* elongation, preventing the RNA from growing more than 2–3 nucleotides long. Thus, strictly speaking, it blocks initiation, because initiation is not complete until the RNA is up to 10 nucleotides long, but its effect is really on the elongation that is part of initiation.

In 1987, M. A. Grachev and colleagues provided more evidence for the notion that β plays a role in elongation, using a technique called **affinity labeling**. The idea behind this technique is to label an enzyme with a derivative of a normal substrate that can be cross-linked to protein. In this way, one can use the affinity reagent to seek out and then tag the active site of the enzyme. Finally, one can dissociate the enzyme to see which subunit the tag is attached to. Grachev and coworkers used 14 different affinity reagents, all ATP or GTP analogs. One of these, which was the first in the series, and therefore called I, has the structure shown in Figure 6.30a. When it was added to RNA polymerase, it went to the active site, as an ATP that is initiating transcription would normally do, and then formed a covalent bond with an amino group at the active site according to the reaction in Figure 6.30b.

In principle, these investigators could have labeled the affinity reagent itself and proceeded from there. However, they recognized a pitfall in that simple strategy: The affinity reagent could bind to other amino groups on the enzyme surface in addition to the one(s) in the active site. To circumvent this problem, they used an unlabeled affinity reagent, followed by a radioactive nucleotide ($[\alpha\text{-}^{32}\text{P}]\text{UTP}$ or CTP) that would form a phosphodiester bond with the affinity reagent in the active site and therefore label that site and no others on the enzyme. Finally, they dissociated the labeled enzyme and subjected the subunits to SDS-PAGE.

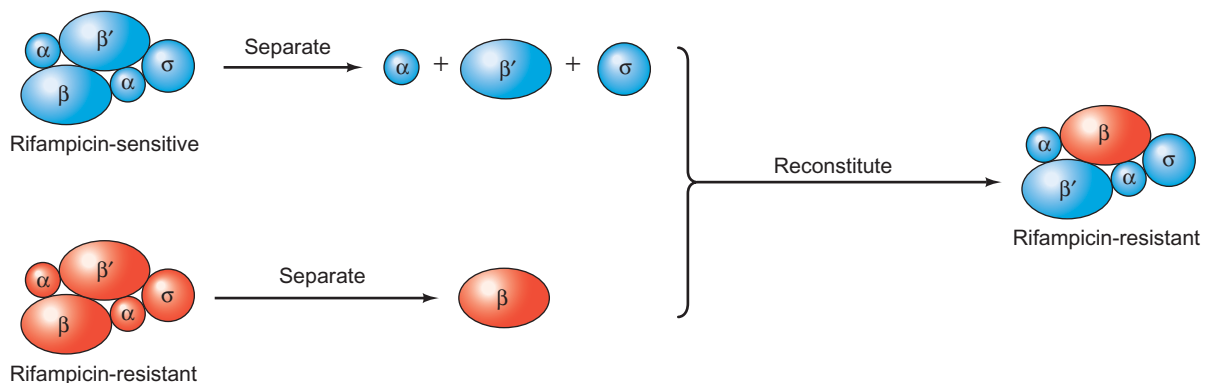


Figure 6.29 Separation and reconstitution of RNA polymerase to locate the determinant of antibiotic resistance. Start with RNA polymerases from rifampicin-sensitive and -resistant *E. coli* cells, separate them into their component polypeptides, and recombine them in various combinations to reconstitute the active enzyme. In this

case, the α -, β' -, and σ -subunits came from the rifampicin-sensitive polymerase (blue), and the β -subunit came from the antibiotic-resistant enzyme (red). The reconstituted polymerase is rifampicin-resistant, which shows that the β -subunit determines sensitivity or resistance to this antibiotic.

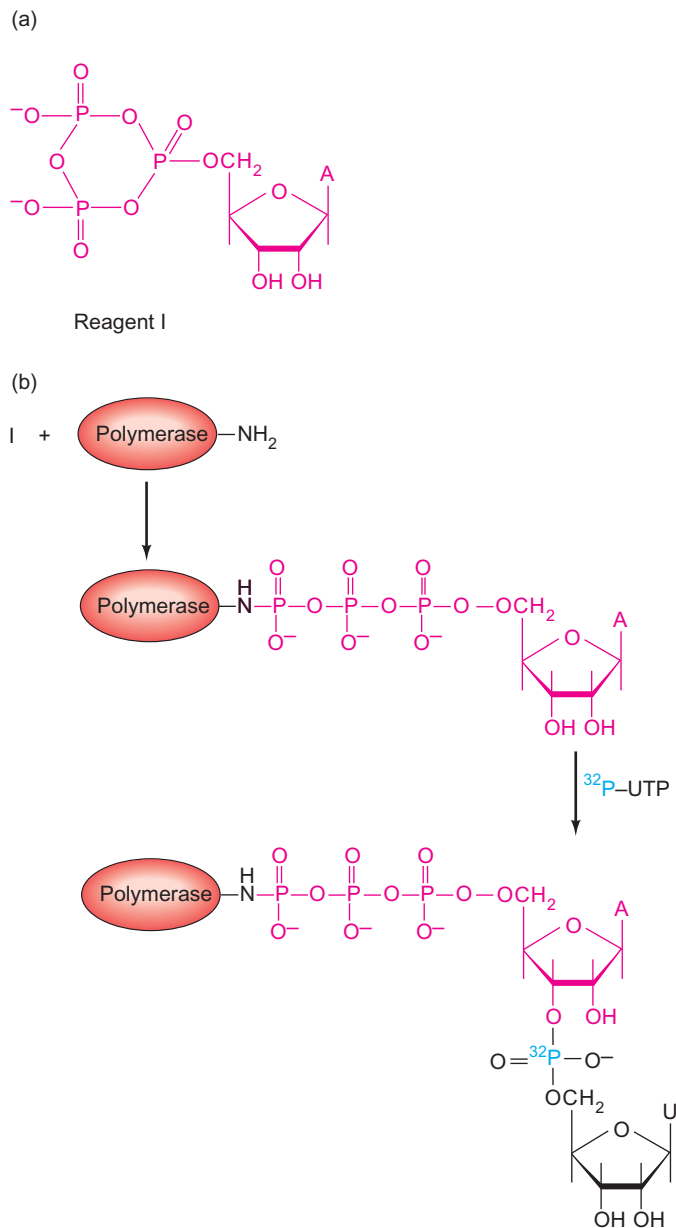


Figure 6.30 Affinity labeling RNA polymerase at its active site. (a) Structure of one of the affinity reagents (I), an ATP analog. (b) The affinity-labeling reactions. First, add reagent I to RNA polymerase. The reagent binds covalently to amino groups at the active site (and perhaps elsewhere). Next, add radioactive UTP, which forms a phosphodiester bond (blue) with the enzyme-bound reagent I. This reaction should occur only at the active site, so only that site becomes radioactively labeled.

The results are presented in Figure 6.31. Obviously, the β -subunit is the only core subunit labeled by any of the affinity reagents, suggesting that this subunit is at or very near the site where phosphodiester bond formation occurs. In some cases, we also see some labeling of σ , suggesting that it too may lie near the catalytic center.

SUMMARY The core subunit β lies near the active site of the RNA polymerase where phosphodiester bonds are formed. The σ -factor may also be near the nucleotide-binding site, at least during the initiation phase.

Structure of the Elongation Complex

Studies in the mid-1990s had suggested that the β and β' subunits are involved in DNA binding. In this section, we will see how well these predictions have been borne out by structural studies. We will also consider the topology of elongation: How does the polymerase deal with the problems of unwinding and rewinding its template, and of moving along its twisted (helical) template without twisting its RNA product around the template?

The RNA–DNA Hybrid Up to this point we have been assuming that the RNA product forms an RNA–DNA hybrid with the DNA template strand for a few bases before peeling off and exiting from the polymerase. But the length of this hybrid has been controversial, with estimates ranging from 3–12 bp, and some investigators even doubted whether it existed. But Nudler and Goldfarb and their colleagues applied a transcript walking technique, together with RNA–DNA cross-linking, to prove that an RNA–DNA hybrid really does occur within the elongation complex, and that this hybrid is 8–9 bp long.

The transcript walking technique works like this: Nudler and colleagues used gene cloning techniques described in Chapter 4 to engineer an RNA polymerase with six extra histidines at the C-terminus of the β -subunit. This string of histidines, because of its affinity for divalent metals such as nickel, allowed them to tether the polymerase to a nickel resin so they could change substrates rapidly by washing the resin, with the polymerase stably attached, and then adding fresh reagents. Accordingly, by adding a subset of nucleotides (e.g., ATP, CTP, and GTP, but no UTP), they could “walk” the polymerase to a particular position on the template (where the first UTP is required, in the present case). Then they could wash away the first set of nucleotides and add a second subset to walk the polymerase to a defined position further downstream.

These workers incorporated a UMP derivative (U^\bullet) at either position 21 or 45 with respect to the 5'-end of a ^{32}P -labeled nascent RNA. U^\bullet is normally unreactive, but in the presence of NaBH_4 it becomes capable of cross-linking to a base-paired base, as shown in Figure 6.32a. Actually, U^\bullet can reach to a purine adjacent to the base-paired A in the DNA strand, but this experiment was

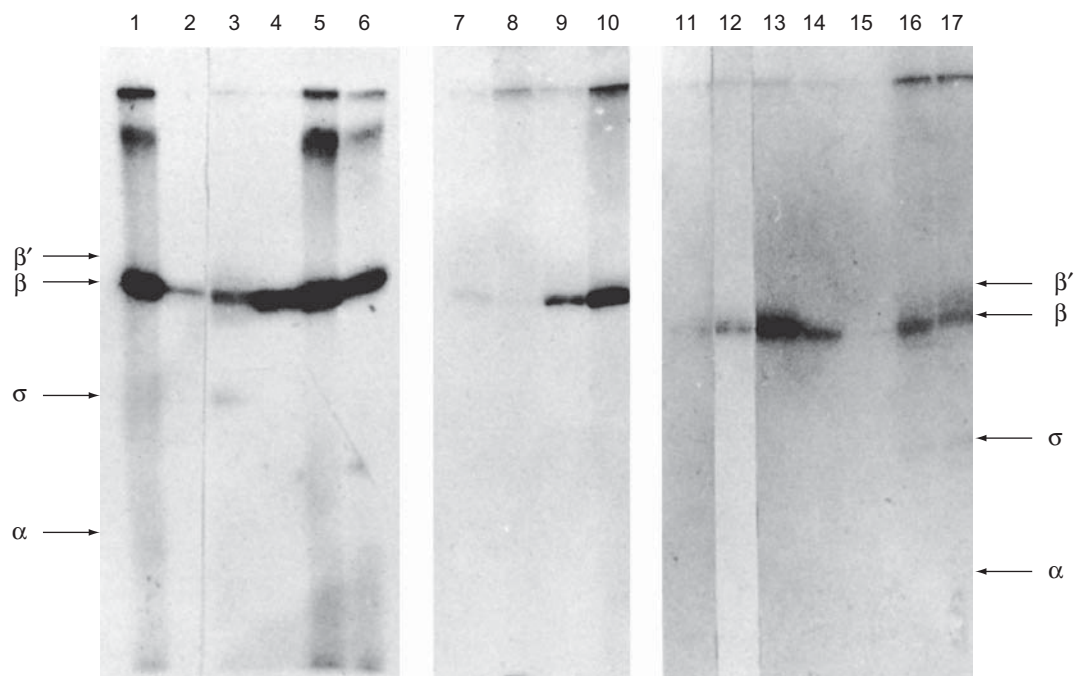


Figure 6.31 The β -subunit is at or near the active site where phosphodiester bonds are formed. Grachev and colleagues labeled the active site of *E. coli* RNA polymerase as described in Figure 6.30, then separated the polymerase subunits by electrophoresis to identify the subunits that compose the active site. Each lane represents labeling with a different nucleotide-affinity reagent plus radioactive UTP, except lanes 5 and 6, which resulted from using the same affinity

reagent, but either radioactive UTP (lane 5) or CTP (lane 6). The autoradiograph of the separated subunits demonstrates labeling of the β -subunit with most of the reagents. In a few cases, σ was also faintly labeled. Thus, the β -subunit appears to be at or near the phosphodiester bond-forming active site. (Source: Grachev et al., Studies on the functional topography of *Escherichia coli* RNA polymerase. *European Journal of Biochemistry* 163 (16 Dec 1987) p. 117, f. 2.)

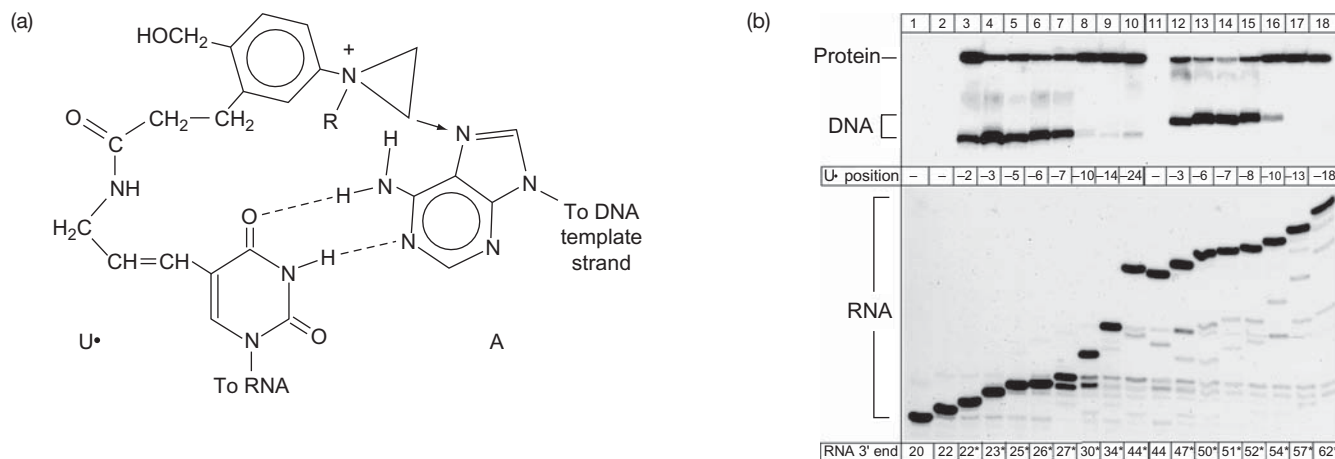


Figure 6.32 RNA-DNA and RNA-protein cross-linking in elongation complexes. (a) Structure of the cross-linking reagent U^\bullet base-paired with an A in the DNA template strand. The reagent is in position to form a covalent bond with the DNA as shown by the arrow. (b) Results of cross-linking. Nudler, Goldfarb, and colleagues incorporated U^\bullet at position 21 or 45 of a [^{32}P]nascent RNA in an elongation complex. Then they walked the U^\bullet to various positions between -2 and -24 with respect to the $3'$ -end (position -1) of the nascent RNA. Then they cross-linked the RNA to the DNA template (or the protein in the RNA polymerase). They then electrophoresed the DNA and protein in one gel (top) and the free RNA transcripts in another (bottom) and autoradiographed the gels. Lanes 1, 2, and 11 are negative controls in which the RNA contained no U^\bullet . Lanes 3–10 contained products from reactions in which the U^\bullet was in position 21; lanes 12–18 contained products from reactions in which the U^\bullet was in position 45 of the nascent RNA. Asterisks at bottom denote the presence of U^\bullet in the RNA. Cross-linking to DNA was prevalent only when U^\bullet was between positions -2 and -8 . (Sources: (a) Reprinted from *Cell* 89, Nudler, E. et al. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase fig.1, p. 34 © 1997 from Elsevier (b) Nudler, E. et al. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* 89 (1997) f. 1, p. 34. Reprinted by permission of Elsevier Science.)

designed to prevent that from happening. So cross-linking could occur only to an A in the DNA template strand that was base-paired to the U• base in the RNA product. If no base-pairing occurred, no cross-linking would be possible.

Nudler, Goldfarb, and their colleagues walked the U• base in the transcript to various positions with respect to the 3'-end of the RNA, beginning with position -2 (the nucleotide next to the 3'-end, which is numbered -1) and extending to position -44. Then they tried to cross-link the RNA to the DNA template strand. Finally, they electrophoresed both the DNA and protein in one gel, and just the RNA in another. Note that the RNA will always be labeled, but the DNA or protein will be labeled only if the RNA has been cross-linked to them.

Figure 6.32b shows the results. The DNA was strongly labeled if the U• base was in position -2 through position -8, but only weakly labeled when the U• base was in position -10 and beyond. Thus, the U• base was base-paired to its A partner in the DNA template strand only when it was in position -2 through -8, but base-pairing was much decreased when the reactive base was in position -10. So the RNA-DNA hybrid extends from position -1 to position -8, or perhaps -9, but no farther. (The nucleotide at the very 3'-end of the RNA, at position -1, must be base-paired to the template to be incorporated correctly.) This conclusion was reinforced by the protein labeling results. Protein in the RNA polymerase became more strongly labeled when the U• was not within the hybrid region (positions -1 through -8). This presumably reflects the fact that the reactive group was more accessible to the protein when it was not base-paired to the DNA template. More recent work on the T7 RNA polymerase has indicated a hybrid that is 8 bp long.

SUMMARY The RNA-DNA hybrid within the *E. coli* elongation complex extends from position -1 to position -8 or -9 with respect to the 3'-end of the nascent RNA. The T7 hybrid appears to be 8 bp long.

Structure of the Core Polymerase To get the clearest picture of the structure of the elongation complex, we need to know the structure of the core polymerase. X-ray crystallography would give the best resolution, but it requires three-dimensional crystals and, so far, no one has succeeded in preparing three-dimensional crystals of the *E. coli* polymerase. However, in 1999 Seth Darst and colleagues crystallized the core polymerase from another bacterium, *Thermus aquaticus*, and obtained a crystal structure to a resolution of 3.3 Å. This structure is very similar in overall shape to the lower-resolution structure of the *E. coli* core polymerase obtained by electron microscopy of

two-dimensional crystals, so the detailed structures are probably also similar. In other words, the crystal structure of the *T. aquaticus* polymerase is our best window right now on the structure of a bacterial polymerase. As we look at this and other crystal structures throughout this book, we need to remember a principle we will discuss more fully in Chapters 9 and 10: Proteins do not have just one static structure. Instead, they are dynamic molecules that can assume a wide range of conformations. The one we trap in a crystal may not be the one (or more than one) that the active form of the protein assumes in vivo.

Figure 6.33 depicts the overall shape of the enzyme in three different orientations. We notice first of all that it resembles an open crab claw. The four subunits (β , β' , and two α) are shown in different colors so we can distinguish them. This coloring reveals that half of the claw is composed primarily of the β -subunit, and the other half is composed primarily of the β' -subunit. The two α -subunits lie at the “hinge” of the claw, with one of them (α I, yellow) associated with the β -subunit, and the other (α II, green) associated with the β' -subunit. The small ω -subunit is at the bottom, wrapped around the C-terminus of β' .

Figure 6.34 shows the **catalytic center** of the core polymerase. We see that the enzyme contains a channel, about 27 Å wide, between the two parts of the claw, and the template DNA presumably lies in this channel. The catalytic center of the enzyme is marked by the Mg^{2+} ion, represented here by a pink sphere. Three pieces of evidence place the Mg^{2+} at the catalytic center. First, an invariant string of amino acids (NADFDGD) occurs in the β' -subunit from all bacteria examined so far, and it contains three aspartate residues (D) suspected of chelating a Mg^{2+} ion. Second, mutations in any of these Asp residues are lethal. They create an enzyme that can form an open-promoter complex at a promoter, but is devoid of catalytic activity. Thus, these Asp residues are essential for catalytic activity, but not for tight binding to DNA. Finally, as Figure 6.34 demonstrates, the crystal structure of the *T. aquaticus* core polymerase shows that the side chains of the three Asp residues (red) are indeed coordinated to a Mg^{2+} ion. Thus, the three Asp residues and a Mg^{2+} ion are at the catalytic center of the enzyme.

Figure 6.34 also identifies a rifampicin-binding site in the part of the β -subunit that forms the ceiling of the channel through the enzyme. The amino acids whose alterations cause rifampicin resistance are tagged with purple dots. Clearly, these amino acids are tightly clustered in the three-dimensional structure, presumably at the site of rifampicin binding. We also know that rifampicin allows RNA synthesis to begin, but blocks elongation of the RNA chain beyond just a few nucleotides. On the other hand, the antibiotic has no effect on elongation once promoter clearance has occurred.

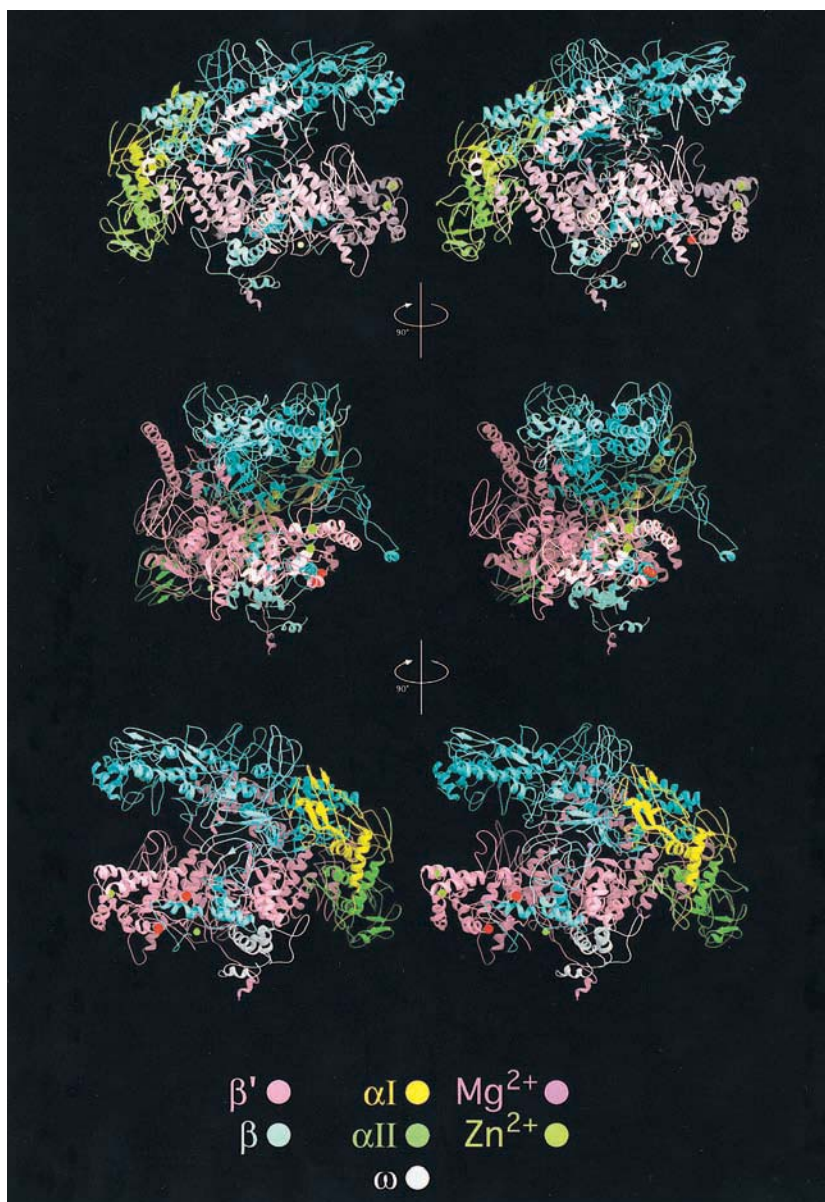


Figure 6.33 Crystal structure of the *Thermus aquaticus* RNA polymerase core enzyme. Three different stereo views are shown, differing by 90-degree rotations. The subunits and metal ions in the enzyme are color-coded as indicated at the bottom. The metal ions are depicted as small colored spheres. The larger red dots denote unstructured regions of the β - and β' -subunits that are missing from these diagrams.

(Source: Zhang, G. et al., Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98 (1999) 811–24. Reprinted by permission of Elsevier Science.)

How can we interpret the location of the rifampicin-binding site in terms of the antibiotic's activity? One hypothesis is that rifampicin bound in the channel blocks the exit through which the growing RNA should pass, and thus prevents growth of a short RNA. Once an RNA reaches a certain length, it might block access to the rifampicin-binding site, or at least prevent effective binding of the antibiotic.

Darst and colleagues validated this hypothesis by determining the crystal structure of the *T. aquaticus* polymerase core complexed with rifampicin. The antibiotic lies in the predicted site in such a way that it would block the exit of

the elongating transcript when the RNA reaches a length of 2 or 3 nt.

SUMMARY X-ray crystallography on the *Thermus aquaticus* RNA polymerase core has revealed an enzyme shaped like a crab claw designed to grasp DNA. A channel through the enzyme includes the catalytic center (a Mg^{2+} ion coordinated by three Asp residues), and the rifampicin-binding site.



Figure 6.34 Stereo view of the catalytic center of the core polymerase. The Mg^{2+} ion is shown as a pink sphere, coordinated by three aspartate side chains (red) in this stereo image. The amino acids involved in rifampicin resistance are denoted by purple spheres at the top of the channel, surrounding the presumed rifampicin-binding site, or Rif pocket, labeled Rif^r. The colors of the polymerase subunits are as in Figure 6.33 (β' , pink; β , turquoise; α 's yellow and green). Note that the two panels of this figure are the two halves of the stereo image. (Source: Zhang G. et al., "Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution." *Cell* 98 (1999) 811–24. Reprinted by permission of Elsevier and Green Science.)

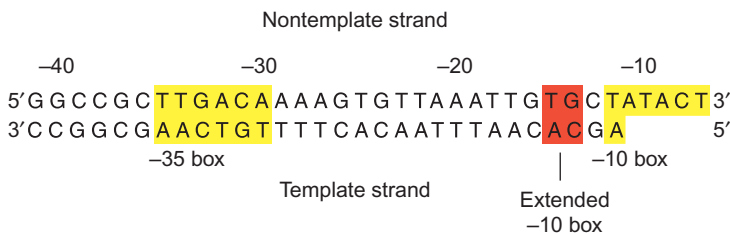


Figure 6.35 Structure of the DNA used to form the RF complex. The -10 and -35 boxes are shaded yellow, and an extended -10 element is shaded red. Bases -11 through -7 are in single-stranded form, as they would be in an open promoter complex.

Structure of the Holoenzyme–DNA Complex To generate a homogeneous holoenzyme–DNA complex, Darst and colleagues bound the *T. aquaticus* holoenzyme to the “fork-junction” DNA pictured in Figure 6.35. This DNA is mostly double-stranded, including the -35 box, but has a single-stranded projection on the nontemplate strand in the -10 box region, beginning at position -11 . This simulates the character of the promoter in the open promoter complex, and locks the complex into a form (RF, where F stands for “fork junction”) resembling RP_o .

Figure 6.36a shows an overall view of the holoenzyme–promoter complex. The first thing to notice is that the DNA stretches across the top of the polymerase in this view—where the σ -subunit is located. In fact, all of the specific DNA–protein interactions involve σ , not the core. Considering the importance of σ in initiation, that is not surprising.

Looking more closely (Figure 6.36b) we can see that the structure corroborates several features already inferred from biochemical and genetic experiments. First of all, as we saw earlier in this chapter, σ region 2.4 is implicated in recognizing the -10 box of the promoter. In particular, mutations in Gln 437 and Thr 440 of *E. coli* σ^{70} can suppress mutations in position -12 of the promoter, suggesting an interaction between these two amino acids and the base at position -12 (recall Figure 6.22). Gln 437 and Thr 440 in *E. coli* σ^{70} correspond to Gln 260 and Asn 263 of *T. aquaticus* σ^A , so we would expect these two amino acids to be close to the base at position -12 in the promoter. Figure 6.36b bears out part of this prediction. Gln 260 (Q260, green) is indeed close enough to contact base -12 . Asn 263 (N263, also colored green) is too far away to make contact in this structure, but a minor movement, which could easily occur in vivo, would bring it close enough.

Three highly conserved aromatic residues in *E. coli* σ^{70} (corresponding to Phe 248 (F248), Tyr 253 (Y253), and Trp 256 (W256) of *T. aquaticus* σ^A) have been implicated in promoter melting. These amino acids presumably bind the nontemplate strand in the -10 box in the open promoter complex. These amino acids (colored yellow-green in Figure 6.36b) are indeed in position to interact with the single-stranded nontemplate strand in the RF complex. In fact, Trp 256 is neatly positioned to stack with base pair 12, which is the last base pair before the melted region of the -10 box. In this way, Trp 256 would substitute for a base pair in position -11 and help melt that base pair.

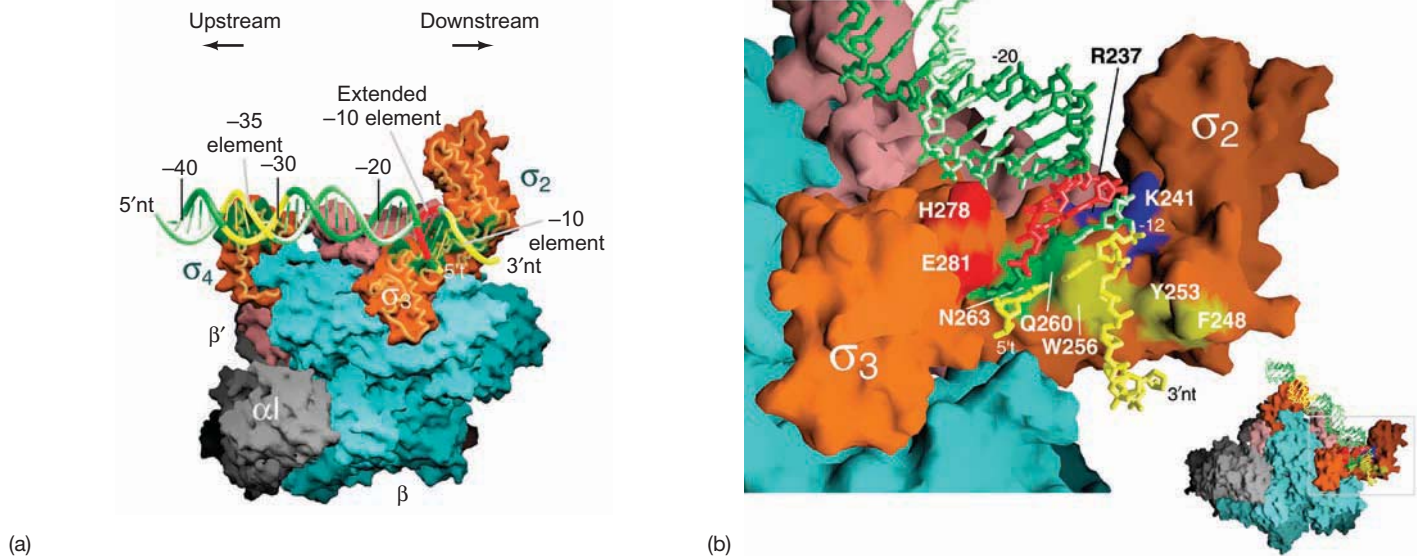


Figure 6.36 Structure of the RF complex. (a) The whole complex. The various subunits are color coded as follows: β , turquoise; β' , brown; α , gray; regions of σ (σ_2 – σ_4), tan and orange (σ_1 is not included in this crystal structure). The DNA is shown as a twisted ladder. The surface of σ is rendered partially transparent to reveal the path of the α -carbon backbone. (b) Contacts between the holoenzyme and downstream DNA. The σ_2 and σ_3 domains are colored as in (a), except for residues that have been implicated by genetic studies in downstream promoter binding. These are: extended -10 box recognition, red; -10 box recognition, green; -10 box melting and nontemplate strand binding, yellow-green; and invariant basic residues implicated in DNA binding, blue. The -10 box DNA is yellow and the extended -10 box DNA is red. The 3'-end of the nontemplate strand is denoted 3'nt. Specific amino acid side chains that are important in DNA binding are labeled. The box in the small structure at lower right shows the position of the magnified structure within the RF complex. (Source: Murakami et al., *Science* 296: (a), p. 1287; (b), p. 1288. Copyright 2002 by the AAAS.)

Two invariant basic residues in σ regions 2.2 and 2.3 (Arg 237 [R237] and Lys 241 [K241]) are known to participate in DNA binding. Figure 6.36b shows why: These two residues (colored blue in the figure) are well positioned to bind to the acidic DNA backbone by electrostatic interaction. These interactions are probably not sequence-specific.

Previous studies implicated region 3 of σ in DNA binding, in particular binding to the extended (upstream) -10 box. Specifically, Glu 281 (E281) was found to be important in recognizing the extended -10 box, while His 278 (H278) was implicated in more general DNA-binding in this region. The structure in Figure 6.36b is consistent with those findings: Both Glu 281 and His 278 (red shading on σ region 3) are exposed on an α -helix, and face the major groove of the extended -10 box (red DNA). Glu 281 is probably close enough to contact a thymine at position -13 , and His 278 is close enough to the extended -10 box that it could interact nonspecifically with the phosphodiester bond linking the nontemplate strand residues -17 and -18 .

We saw earlier in this chapter that specific residues in σ region 4.2 are instrumental in binding to the -35 box of the promoter. But, surprisingly, the RF structure does not confirm these findings. In particular, the -35 box seems about 6 Å out of position relative to $\sigma_{4.2}$, and the DNA is straight instead of bending to make the necessary interactions. Because the evidence for these -35 box– $\sigma_{4.2}$ interactions is so strong, Darst and colleagues needed to explain why their crystal structure does not allow them.

They concluded that the -35 box DNA in the RF structure is pushed out of its normal position relative to $\sigma_{4.2}$ by crystal packing forces—a reminder that the shape a molecule or a complex assumes in a crystal is not necessarily the same as its shape *in vivo*, and indeed that proteins are dynamic molecules that can change shape as they do their jobs.

The studies of Darst and colleagues, and others, have revealed only one Mg^{2+} ion at the active site. But all DNA and RNA polymerases are thought to use a mechanism that requires two Mg^{2+} ions. In accord with this mechanism, Dmitry Vassylyev and colleagues have determined the crystal structure of the *T. thermophilus* polymerase at 2.6 Å resolution. Their asymmetric crystals contained two polymerases, one with one Mg^{2+} ion, and one with two. The latter is probably the form of the enzyme that takes part in RNA synthesis. The two Mg^{2+} ions are held by the same three aspartate side chains that hold the single Mg^{2+} ion, in a network involving several nearby water molecules.

SUMMARY The crystal structure of a *Thermus aquaticus* holoenzyme–DNA complex mimicking an open promoter complex reveals several things. First, the DNA is bound mainly to the σ -subunit, which makes all the important interactions with the promoter DNA. Second, the predicted interactions between amino acids in region 2.4 of σ and the

–10 box of the promoter are really possible. Third, three highly conserved aromatic amino acids are predicted to participate in promoter melting, and they really are in a position to do so. Fourth, two invariant basic amino acids in σ are predicted to participate in DNA binding and they are in a position to do so. A higher resolution crystal structure reveals a form of the polymerase that has two Mg^{2+} ions, in accord with the probable mechanism of catalysis.

Structure of the Elongation Complex In 2007, Dmitry Vassylyev and colleagues presented the x-ray crystal structure of the *Thermus thermophilus* RNA polymerase elongation complex at 2.5Å resolution. This complex contained 14 bp of downstream double-stranded DNA that had yet to be melted by the polymerase, 9 bp of RNA–DNA hybrid, and 7 nt of RNA product in the RNA exit channel. Several important observations came from this work.

First, a valine residue in the β' subunit inserts into the minor groove of the downstream DNA. This could have two important consequences: It could prevent the DNA from slipping backward or forward in the enzyme; and it could induce the screw-like motion of the DNA through the enzyme, which we will examine later in this chapter. (Consider a screw being driven through a threaded hole in a piece of metal. The metal threads, because of their position between the threads of the screw, require the screw to turn in order to penetrate or withdraw.) There are analogous residues in the single-subunit phage T7 RNA polymerase (Chapter 8), and in the multi-subunit yeast enzyme (Chapter 10) that probably play the same role as the valine residue in the *T. thermophilus* β' subunit.

Second, as Figure 6.37a shows, the downstream DNA is double-stranded up to and including the +2 base pair, where +1 is the position at which the new nucleotide is added. This means that only one base pair (at position +1) is melted and available for base-pairing with an incoming nucleotide, so only one nucleotide at a time can bind specifically to the complex. Figure 6.37a also demonstrates that one amino acid in the β subunit is situated in a key position right at the site where nucleotides are added to the growing RNA chain. This is arginine 422 of the β fork 2 loop. It makes a hydrogen bond with the phosphate of the +1 template nucleotide, and van der Waals interactions with both bases of the +2 base pair. In the T7 polymerase elongation complex, phenylalanine 644 is in a similar position (Figure 6.37b). The proximity of these amino acids to the active site, and their interactions with key nucleotides there, suggests that they play a role in molding the active site for accurate substrate recognition. If this is so, then mutations in these amino acids should decrease the accuracy of transcription. Indeed, changing phenylalanine 644

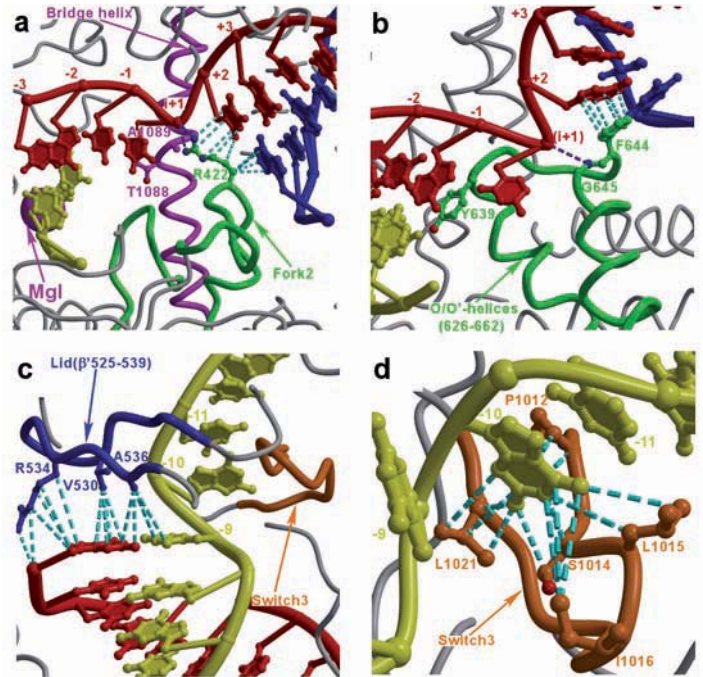


Figure 6.37 Strand separation in the DNA template and in the RNA–DNA hybrid. (a) Downstream DNA strand separation in the *T. thermophilus* polymerase. Note the interactions between R422 (green) and the template nucleotide phosphate and the +2 base pair. In all panels, polar interactions are in dark blue, and van der Waals interactions are in blue-green dashed lines. (b) Downstream DNA strand separation in the T7 enzyme. Note the interactions between F644 (green) and the template nucleotide phosphate and the +2 base pair. (c) RNA–DNA hybrid strand separation in the *T. thermophilus* enzyme. Note the stacking of three amino acids in the β' lid (blue) and the –9 base pair, and the interaction of the first displaced RNA base (–10, light green) with the pocket in the β switch 3 loop (orange). (d) Detail of interactions between the first displaced RNA base (–10) and five amino acids in the β switch 3 loop (orange). Source: Reprinted by permission from Macmillan Publishers Ltd: *Nature*, 448, 157–162, 20 June 2007. Vassylyev et al, Structural basis for transcription elongation by bacterial RNA polymerase. © 2007.

(or glycine 645) of the T7 polymerase to alanine does decrease fidelity. At the time this work appeared, the effect of mutations in arginine 422 of the bacterial enzyme had not been checked.

Third, in agreement with previous biochemical work, the enzyme can accommodate nine base pairs of RNA–DNA hybrid. Furthermore, at the end of this hybrid, a series of amino acids of the β' lid (valine 530, arginine 534, and alanine 536) stack on base pair –9, stabilizing it, and limiting any further base-pairing (Figure 6.37c). These interactions therefore appear to play a role in strand separation at the end of the RNA–DNA hybrid. A variety of experiments have shown the hybrid to vary between 8–10 bp in length, and the β' lid appears to be flexible enough to handle that kind of variability. But other forces are at work in limiting the length of the hybrid. One is the tendency of the two DNA strands to reanneal. Another is the trapping

of the first displaced RNA base (−10) in a hydrophobic pocket of a β loop known as switch 3 (Figure 6.37c). Five amino acids in this pocket make van der Waals interactions with the displaced RNA base (Figure 6.37d), stabilizing the displacement.

Fourth, the RNA product in the exit channel is twisted into the shape it would assume as one-half of an A-form double-stranded RNA. Thus, it is ready to form a hairpin that will cause pausing, or even termination of transcription (see later in this chapter and Chapter 8). Because RNA in hairpin form was not used in this structural study, we cannot see exactly how a hairpin would fit into the exit channel. However, Vassylyev and colleagues modeled the fit of an RNA hairpin in the exit channel, and showed that such a fit can be accomplished with only minor alterations of the protein structure. Indeed, the RNA hairpin could fit with the core enzyme in much the same way as the σ -factor fits with the core in the initiation complex.

In a separate study, Vassylyev and colleagues examined the structure of the elongation complex including an unhydrolyzable substrate analog, adenosine-5'-[(α , β)-methylene]-triphosphate (AMPcPP), which has a methylene (CH₂) group instead of an oxygen between the α - and β -phosphates of ATP. Since this is the bond that is normally broken when the substrate is added to the growing RNA chain, the substrate analog binds to the catalytic site and remains there unaltered. These investigators also looked at the elongation complex structure with AMPcPP and with and without the elongation inhibitor streptolydigin. This comparison yielded interesting information about how the substrate associates with the enzyme in a two-step process.

In the absence of streptolydigin, the so-called trigger loop (residues 1221–1266 of the β' subunit) is fully folded into two α -helices with a short loop in between. (Figure 6.38a). This brings the substrate into the active site in a productive way, with two metal ions (Mg^{2+} , in this case) close enough together to collaborate in forming the phosphodiester bond that will incorporate the new substrate into the growing RNA chain. Studies of many RNA and DNA polymerases (see Chapter 10) have shown that two metal ions participate in phosphodiester bond formation. One of these is permanently held in the active site, and the other shuttles in, bound to the β - and γ -phosphates of the NTP substrate. Once the substrate is added to the growing RNA, the second metal ion leaves, bound to the by-product, inorganic pyrophosphate (which comes from the β - and γ -phosphates of the substrate).

In the presence of streptolydigin, by contrast, the antibiotic forces a change in the trigger loop conformation: The two α -helices unwind somewhat to form a larger loop in between. This in turn forces a change in the way the substrate binds to the active site: The base and sugar of the substrate bind in much the same way, but the triphosphate part extends a bit farther away from the active site, taking

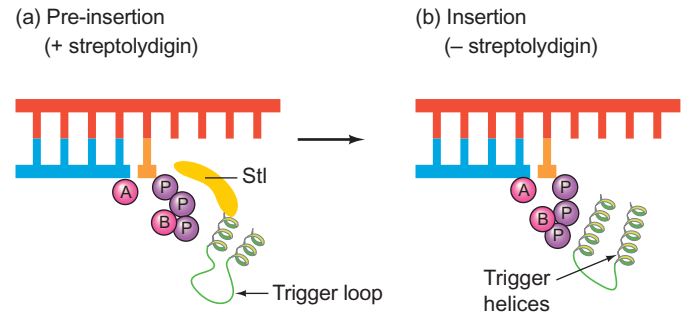


Figure 6.38 A two-step model for nucleotide insertion during RNA synthesis. (a) Pre-insertion state. This is presumably a natural first step in vivo, but it is stabilized by the antibiotic streptolydigin in vitro. Here, streptolydigin (yellow) is forcing the trigger loop out of its normal position close to the active site, which in turn allows the incoming nucleotide (orange with purple triphosphate) to extend its triphosphate moiety away from the active site (exaggerated in this illustration). Because the second metal (metal B) essential for catalysis is complexed to the β - and γ -phosphates of the incoming nucleotide, this places metal B too far away from metal A to participate in catalysis. (b) Insertion state. No streptolydigin is present, so the trigger loop can fold into trigger helices that lie closer to the active site, allowing the triphosphates of the incoming nucleotide, and their complexed metal B, to approach closer to metal A at the active site. This arrangement allows the two metal ions to collaborate in nucleotide insertion into the growing RNA chain.

with it one of the metal ions required for catalysis (Figure 6.38b). This makes catalysis impossible and explains how streptolydigin blocks transcription elongation.

Vassylyev and colleagues concluded that the two states of the elongation complex revealed by streptolydigin correspond to two natural states: a **preinsertion state** (seen in the presence of the antibiotic) and an **insertion state** (seen in the absence of the antibiotic). Presumably, the substrate normally binds first in the preinsertion state (Figure 6.38b), and this allows the enzyme to examine it for correct base-pairing and for the correct sugar (ribose vs. deoxyribose) before it switches to the insertion state (Figure 6.38a), where it can be examined again for correct base-pairing with the template base. Thus, the two-state model helps to explain the fidelity of transcription.

The great similarity in structure of the active site among RNA polymerases from all kingdoms of life suggests that all should use the same mechanism of substrate addition, including the two-state model described here. However, as we will see in Chapter 10, investigators of the yeast RNA polymerase have described a two-state model that includes an “entry state” that differs radically from the preinsertion state described here. The substrate in the “entry site” is essentially upside down with respect to the substrate in the insertion state. Clearly, in such a position, it cannot be checked for proper fit with the template base. Vassylyev and colleagues do not dispute the existence of the entry site, but postulate that, if it exists, it must precede the preinsertion state.

SUMMARY Structural studies of the elongation complex involving the *Thermus thermophilus* RNA polymerase have revealed the following features: A valine residue in the β' subunit inserts into the minor groove of the downstream DNA. In this position, it could prevent the DNA from slipping, and it could induce the screw-like motion of the DNA through the enzyme. Only one base-pair of DNA (at position +1) is melted and available for base-pairing with an incoming nucleotide, so only one nucleotide at a time can bind specifically to the complex. Several forces limit the length of the RNA–DNA hybrid. One of these is the length of the cavity in the enzyme that accommodates the hybrid. Another is a hydrophobic pocket in the enzyme at the end of the cavity that traps the first RNA base displaced from the hybrid. The RNA product in the exit channel assumes the shape of one-half of a double-stranded RNA. Thus, it can readily form a hairpin to cause pausing, or even termination of transcription. Structural studies of the enzyme with an inactive substrate analog and the antibiotic streptolydigin have identified a preinsertion state for the substrate that is catalytically inactive, but could provide for checking that the substrate is the correct one.

Topology of Elongation Does the core, moving along the DNA template, maintain the local melted region created during initiation? Common sense tells us that it does because this would help the RNA polymerase “read” the bases of the template strand and therefore insert the correct bases into the transcript. Experimental evidence also demonstrates that this is so. Jean-Marie Saucier and James Wang added nucleotides to an open promoter complex, allowing the polymerase to move down the DNA as it began elongating an RNA chain, and found that the same degree of melting persisted. Furthermore, the crystal structure of the polymerase–DNA complex shows clearly that the two DNA strands feed through separate channels in the holoenzyme, and we assume that this situation persists with the core polymerase during elongation.

The static nature of the transcription models presented in Chapter 6 is somewhat misleading. If we could see transcription as a dynamic process, we would observe the DNA double helix opening up in front of the moving “bubble” of melted DNA and closing up again behind. In theory, RNA polymerase could accomplish this process in two ways, and Figure 6.39 presents both of them. One way would be for the polymerase and the growing RNA to rotate around and around the DNA

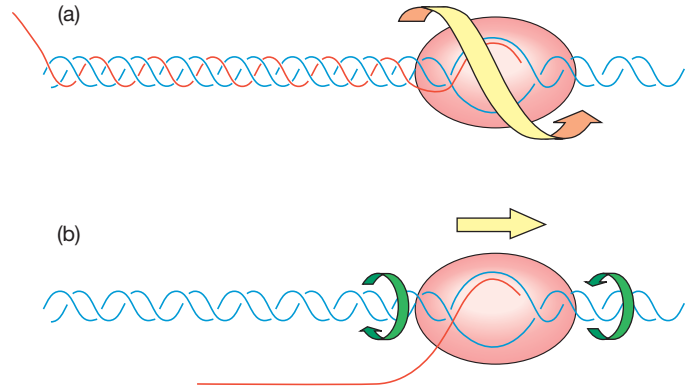


Figure 6.39 Two hypotheses of the topology of transcription of double-stranded DNA. (a) The RNA polymerase (pink) moves around and around the double helix, as indicated by the yellow arrow. This avoids straining the DNA, but it wraps the RNA product (red) around the DNA template. (b) The polymerase moves in a straight line, as indicated by the yellow arrow. This avoids twisting the RNA product (red) around the DNA, but it forces the DNA ahead of the moving polymerase to untwist and the DNA behind the polymerase to twist back up again. These two twists, represented by the green arrows, introduce strain into the DNA template that must be relieved by topoisomerases.

template, following the natural twist of the double-helical DNA, as transcription progressed (Figure 6.39a). This would not twist the DNA at all, but it would require considerable energy to make the polymerase gyrate that much, and it would leave the transcript hopelessly twisted around the DNA template, with no known enzyme to untwist it.

The other possibility is that the polymerase moves in a straight line, with the template DNA rotating in one direction ahead of it to unwind, and rotating in the opposite direction behind it to wind up again (Figure 6.39b). But this kind of rotating of the DNA introduces strain. To visualize this, think of unwinding a coiled telephone cord, or actually try it if you have one available. You can feel (or imagine) the resistance you encounter as the cord becomes more and more untwisted, and you can appreciate that you would also encounter resistance if you tried to wind the cord more tightly than its natural state. It is true that the rewinding of DNA at one end of the melted region creates an opposite and compensating twist for the unwinding at the other. But the polymerase in between keeps this compensation from reaching across the melted region, and the long span of DNA around the circular chromosome insulates the two ends of the melted region from each other the long way around.

So if this second mechanism of elongation is valid, we have to explain how the strain of unwinding the DNA is relaxed. As we will see in Chapter 20 when we discuss DNA replication, a class of enzymes called **topoisomerases** can introduce transient breaks into DNA strands and so relax

this kind of strain. We will see that strain due to twisting a double-helical DNA causes the helix to tangle up like a twisted rubber band. This process is called **supercoiling**, and the supercoiled DNA is called a **supercoil** or **superhelix**. Unwinding due to the advancing polymerase causes a compensating overwinding ahead of the unwound region. (Compensating overwinding is what makes it difficult to unwind a coiled telephone cord.) The supercoiling due to overwinding is by convention called positive. Thus, positive supercoils build up in front of the advancing polymerase. Conversely, negative supercoils form behind the polymerase. One line of evidence that directly supports this model of transcription comes from studies with topoisomerase mutants that cannot relax supercoils. If the mutant cannot relax positive supercoils, these build up in DNA that is being transcribed. On the other hand, negative supercoils accumulate during transcription in topoisomerase mutants that cannot relax that kind of superhelix.

SUMMARY Elongation of transcription involves the polymerization of nucleotides as the RNA polymerase travels along the template DNA. As it moves, the polymerase maintains a short melted region of template DNA. This requires that the DNA unwind ahead of the advancing polymerase and close up again behind it. This process introduces strain into the template DNA that is relaxed by topoisomerases.

Pausing and Proofreading The process of elongation is far from uniform. Instead, the polymerase repeatedly pauses, and in some cases backtracks, while elongating an RNA chain. Under *in vitro* conditions of 21°C and 1 mM NTPs, pauses in bacterial systems have been found to be very brief: generally only 1–6 sec. But repeated short pauses significantly slow the overall rate of transcription. Pausing is physiologically important for at least two reasons: First, it allows translation, an inherently slower process, to keep pace with transcription. This is important for phenomena such as attenuation (Chapter 7), and aborting transcription if translation fails. The second important aspect of pausing is that it is the first step in termination of transcription, as we will see later in this chapter.

Sometimes the polymerase even backtracks by reversing its direction and thereby extruding the 3'-end of the growing transcript out of the active site of the enzyme. This is more than just an exaggerated pause. For one thing, it tends to last much longer: 20 sec, up to irreversible arrest. For another, it occurs only under special conditions: when nucleotide concentrations are severely reduced, or when the polymerase has added the wrong nucleotide to the growing RNA chain. In the latter case, backtracking is part of a proofreading process in which auxiliary proteins known as **GreA** and

GreB stimulate an inherent RNase activity of the polymerase to cleave off the end of the growing RNA, removing the misincorporated nucleotide, and allowing transcription to resume. GreA produces only short RNA end fragments 2–3 nt long, and can prevent, but not reverse transcription arrest. GreB can produce RNA end fragments up to 18 nt long, and can reverse arrested transcription. We will discuss the analogous proofreading mechanism in eukaryotes in greater detail in Chapter 11.

One complication to this proofreading model is that the auxiliary proteins are dispensable *in vivo*. And yet one would predict that mRNA proofreading would be important for life. In 2006, Nicolay Zenkin and colleagues suggested a resolution to this apparent paradox: The nascent RNA itself appears to participate in its own proofreading.

Zenkin and colleagues simulated an elongation complex by mixing RNA polymerase with a piece of single-stranded DNA and an RNA that was either perfectly complementary to the DNA or had a mismatched base at its 3'-end. When they added Mg^{2+} , they observed that the mismatched RNA lost a dinucleotide from its 3'-end, including the mismatched nucleotide and the penultimate (next-to-last) nucleotide. This proofreading did not occur with the perfectly matched RNA. The fact that two nucleotides were lost suggests that the polymerase had backtracked one nucleotide in the mismatched complex. And this in turn suggested a chemical basis for the RNA-assisted proofreading: In the backtracked complex, the mismatched nucleotide, because it is not base-paired to the template DNA, is flexible enough to bend back and contact metal II, holding it at the active site of the enzyme. This would be expected to enhance phosphodiester bond cleavage, because metal II is presumably involved in the enzyme's RNase activity. In addition, the mismatched nucleotide can orient a water molecule to make it a better nucleophile in attacking the phosphodiester bond that links the terminal dinucleotide to the rest of the RNA. Both of these considerations help to explain why the mismatched RNA can stimulate its own cleavage, while a perfectly matched RNA cannot.

SUMMARY RNA polymerase frequently pauses, or even backtracks, during elongation. Pausing allows ribosomes to keep pace with the RNA polymerase, and it is also the first step in termination. Backtracking aids proofreading by extruding the 3'-end of the RNA out of the polymerase, where misincorporated nucleotides can be removed by an inherent nuclease activity of the polymerase, stimulated by auxiliary factors. Even without these factors, the polymerase can carry out proofreading: The mismatched nucleotide at the end of a nascent RNA plays a role in this process by contacting two key elements at the active site: metal II and a water molecule.

6.5 Termination of Transcription

When the polymerase reaches a **terminator** at the end of a gene it falls off the template, releasing the RNA. *E. coli* cells contain about equal numbers of two kinds of terminators. The first kind, known as **intrinsic terminators**, function with the RNA polymerase by itself without help from other proteins. The second kind depend on an auxiliary factor called **rho** (ρ). Naturally, these are called rho-dependent terminators. Let us consider the mechanisms of termination employed by these two systems, beginning with the simpler, intrinsic terminators.

Rho-Independent Termination

Rho-independent, or intrinsic, termination depends on terminators consisting of two elements: an inverted repeat followed immediately by a T-rich region in the non-template strand of the gene. The model of termination we will present later in this section depends on a “hairpin” structure in the RNA transcript of the inverted repeat. Before we get to the model, we should understand how an inverted repeat predisposes a transcript to form a hairpin.

Inverted Repeats and Hairpins Consider this inverted repeat:



Such a sequence is symmetrical around its center, indicated by the dot; it would read the same if rotated 180 degrees in the plane of the paper, and if we always read the strand that runs 5'→3' left to right. Now observe that a transcript of this sequence



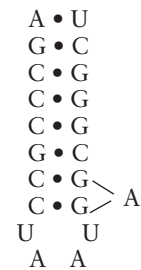
is self-complementary around its center (the underlined G). That means that the self-complementary bases can pair to form a hairpin as follows:



The A and the U at the apex of the hairpin cannot form a base pair because of the physical constraints of the turn in the RNA.

The Structure of an Intrinsic Terminator The *E. coli trp* operon (Chapter 7) contains a DNA sequence called an attenuator that causes premature termination of transcription. The *trp* attenuator contains the two elements (an inverted repeat and a string of T's in the nontemplate DNA strand) suspected to be vital parts of an intrinsic terminator, so Peggy Farnham and Terry Platt used attenuation as an experimental model for normal termination.

The inverted repeat in the *trp* attenuator is not perfect, but 8 bp are still possible, and 7 of these are strong G–C pairs, held together by three hydrogen bonds. The hairpin looks like this:



Notice that a small loop occurs at the end of this hairpin because of the U–U and A–A combinations that cannot base-pair. Furthermore, one A on the right side of the stem has to be “looped out” to allow 8 bp instead of just 7. Still, the hairpin should form and be relatively stable.

Farnham and Platt reasoned as follows: As the T-rich region of the attenuator is transcribed, eight A–U base pairs would form between the A's in the DNA template strand and the U's in the RNA product. They also knew that rU–dA base pairs are exceptionally weak; they have a melting temperature 20°C lower than even rU–rA or dT–rA pairs. This led the investigators to propose that the polymerase paused at the terminator, and then the weakness of the rU–dA base pairs allowed the RNA to dissociate from the template, terminating transcription.

What data support this model? If the hairpin and string of rU–dA base pairs in the *trp* attenuator are really important, we would predict that any alteration in the base sequence that would disrupt either one would be deleterious to attenuation. Farnham and Platt devised the following in vitro assay for attenuation (Figure 6.40): They started with a *HpaII* restriction fragment containing the *trp* attenuator and transcribed it in vitro. If attenuation works, and transcription terminates at the attenuator, a short (140-nt) transcript should be the result. On the other hand, if transcription fails to terminate at the attenuator, it will continue to the end of the fragment, yielding a run-off transcript 260 nt in length. These two transcripts are easily distinguished by electrophoresis.

When these investigators altered the string of eight T's in the nontemplate strand of the terminator to the sequence TTTTGCAA, creating the mutant they called *trp a1419*,

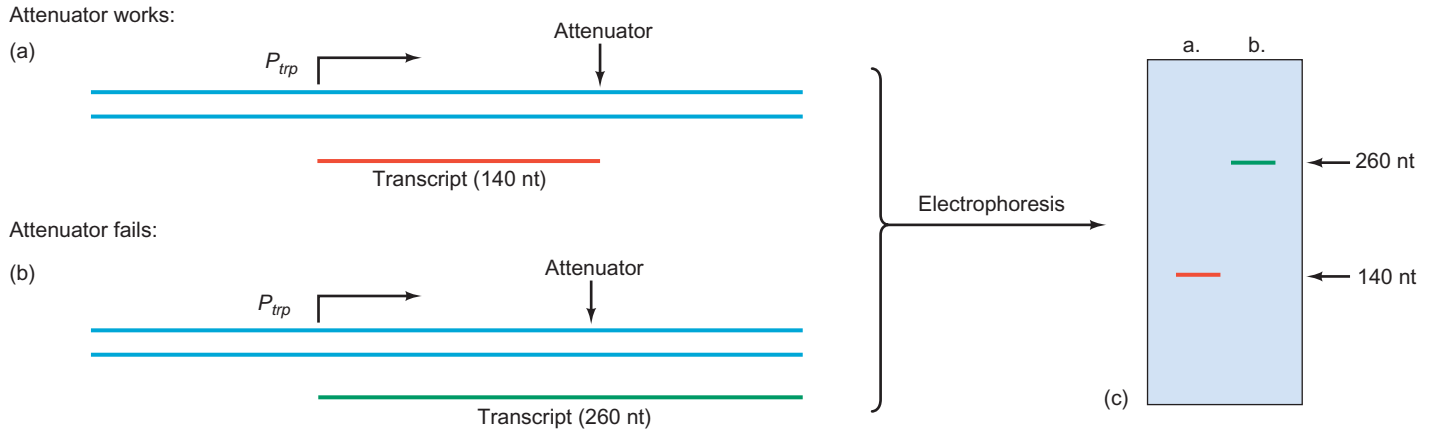


Figure 6.40 An assay for attenuation. (a) When the DNA fragment containing the *trp* promoter and attenuator is transcribed under conditions in which the attenuator works, transcription stops in the attenuator, and a 140-nt transcript (red) results. (b) When the same DNA fragment is transcribed under conditions that cause the

attenuator to fail, a run-off transcript of 260 nt (green) is the result. (c) The transcripts from the two different reactions can be distinguished easily by electrophoresis. Using this assay, one can tell whether the attenuator works under a variety of conditions.

attenuation was weakened. This is consistent with the hypothesis that the weak rU–dA pairs are important in termination, because half of them would be replaced by stronger base pairs in this mutant.

Moreover, this mutation could be overridden by substituting the nucleotide iodo-CTP (I-CTP) for normal CTP in the *in vitro* reaction. The most likely explanation is that base-pairing between G and iodo-C is stronger than between G and ordinary C. Thus, the GC-rich hairpin should be stabilized by I-CMP, and this effect counteracts the loss of weak base pairs in the region following the hairpin. On the other hand, IMP (inosine monophosphate, a GMP analog) should weaken base-pairing in the hairpin because I–C pairs, with only two hydrogen bonds holding them together, are weaker than G–C pairs with three. Sure enough, substituting ITP for GTP in the transcription reaction weakened termination at the attenuator. Thus, all of these effects are consistent with the hypothesis that the hairpin and string of U's in the transcript are important for termination. However, they do not identify the roles that these RNA elements play in pausing and termination.

SUMMARY Using the *trp* attenuator as a model terminator, Farnham and Platt showed that intrinsic terminators have two important features: (1) an inverted repeat that allows a hairpin to form at the end of the transcript; (2) a string of T's in the non-template strand that results in a string of weak rU–dA base pairs holding the transcript to the template strand.

A Model for Termination Several hypotheses have been proposed for the roles of the hairpin and string of rU–dA

base pairs in the mechanism of termination. Two important clues help narrow the field of hypotheses. First, hairpins are found to destabilize elongation complexes that are stalled artificially (not at strings of rU–dA pairs). Second, terminators in which half of the inverted repeat is missing still stall at the strings of rU–dA pairs, even though no hairpin can form. This leads to the following general hypothesis: The rU–dA pairs cause the polymerase to pause, allowing the hairpin to form and destabilize the already weak rU–dA pairs that are holding the DNA template and RNA product together. This destabilization results in dissociation of the RNA from its template, terminating transcription.

W. S. Yarnell and Jeffrey Roberts proposed a variation on this hypothesis in 1999, as illustrated in Figure 6.41. This model calls for the withdrawal of the RNA from the active site of the polymerase that has stalled at a terminator—either because the newly formed hairpin helps pull it out or because the polymerase moves downstream without elongating the RNA, thus leaving the RNA behind. To test their hypothesis, Yarnell and Roberts used a DNA template that contained two mutant terminators ($\Delta tR2$ and $\Delta t82$) downstream of a strong promoter. These terminators had a T-rich region in the nontemplate strand, but only half of an inverted repeat, so hairpins could not form. To compensate for the hairpin, these workers added an oligonucleotide that was complementary to the remaining half of the inverted repeat. They reasoned that the oligonucleotide would base-pair to the transcript and restore the function of the hairpin.

To test this concept, they attached magnetic beads to the template, so it could be easily removed from the mixture magnetically. Then they used *E. coli* RNA polymerase to synthesize labeled RNAs *in vitro* in the presence and absence of the appropriate oligonucleotides. Finally, they removed the template magnetically to form

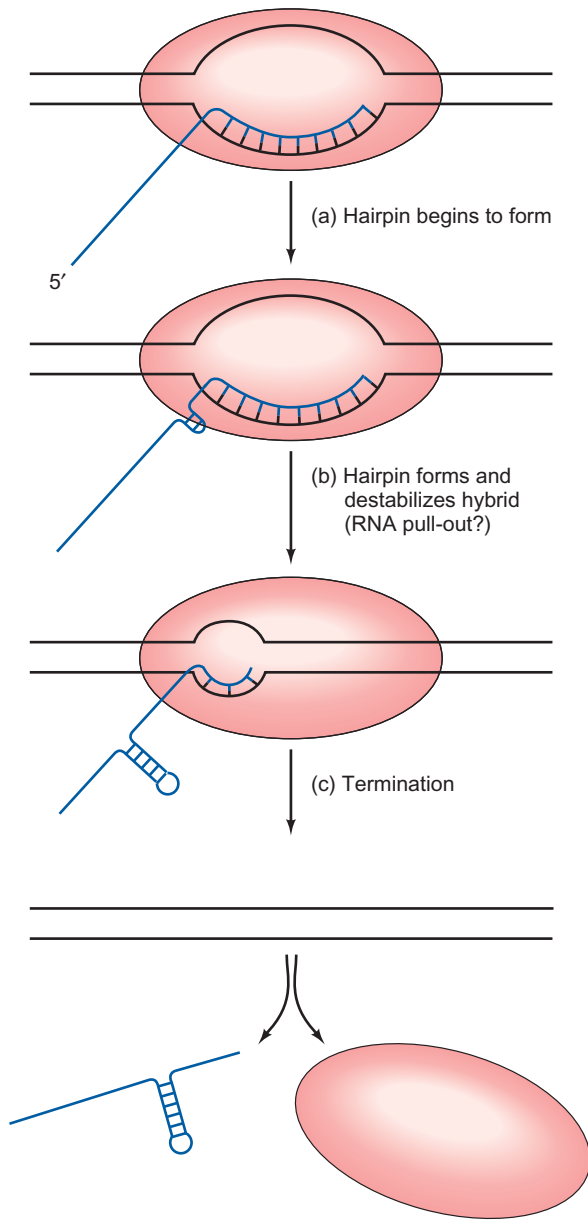


Figure 6.41 A model for rho-independent, or intrinsic termination.

(a) The polymerase has paused at a string of weak rU–dA base pairs, and a hairpin has started to form just upstream of these base pairs.
(b) As the hairpin forms, it further destabilizes the RNA–DNA hybrid. This destabilization could take several forms: The formation of the hairpin could physically pull the RNA out of the polymerase, allowing the transcription bubble to collapse; conversely, it could cause the transcription bubble to collapse, expelling the RNA from the hybrid.
(c) The RNA product and polymerase dissociate completely from the DNA template, terminating transcription.

a pellet and electrophoresed the material in the pellet and the supernatant and detected the RNA species by autoradiography.

Figure 6.42 shows the results. In lanes 1–6, no oligonucleotides were used, so little incomplete RNA was

released into the supernatant (see faint bands at $\Delta tR2$ and $\Delta t82$ markers in lanes 1, 3, and 5). However, pausing definitely did occur at both terminators, especially at short times (see stronger bands in lanes 2, 4, and 6). This was a clear indication that the hairpin is not required for pausing, though it is required for efficient release of the transcript. In lanes 7–9, Yarnell and Roberts included an oligonucleotide (t19) complementary to the remaining, downstream half of the inverted repeat in the $\Delta tR2$ terminator. Clearly, this oligonucleotide stimulated termination at the mutant terminator, as the autoradiograph shows a dark band corresponding to a labeled RNA released into the supernatant. This labeled RNA is exactly the same size as an RNA released by the wild-type terminator would be. Similar results, though less dramatic, were obtained with an oligonucleotide (t18) that is complementary to the downstream half of the inverted repeat in the $\Delta t82$ terminator.

To test further the importance of base-pairing between the oligonucleotide and the half-inverted repeat, these workers mutated one base in the t19 oligonucleotide to yield an oligonucleotide called t19H1. Lane 13 shows that this change caused a dramatic reduction in termination at $\Delta tR2$. Then they made a compensating mutation in $\Delta tR2$ and tested t19H1 again. Lane 14 shows that this restored strong termination at $\Delta tR2$. This template also contained the wild-type t82 terminator, so abundant termination also occurred there. Lanes 15 and 16 are negative controls in which no t19H1 oligonucleotide was present, and, as expected, very little termination occurred at the $\Delta tR2$ terminator.

Together, these results show that the hairpin itself is not required for termination. All that is needed is something to base-pair with the downstream half of the inverted repeat to destabilize the RNA–DNA hybrid. Furthermore, the T-rich region is not required if transcription can be slowed to a crawl artificially. Yarnell and Roberts advanced the polymerase to a site that had neither an inverted repeat nor a T-rich region and made sure it paused there by washing away the nucleotides. Then they added an oligonucleotide that hybridized upstream of the artificial pause site. Under these conditions, they observed release of the nascent RNA.

Termination is also stimulated by a protein called NusA, which appears to promote hairpin formation in the terminator. The essence of this model, presented in 2001 by Ivan Gusarov and Evgeny Nudler, is that the upstream half of the hairpin binds to part of the core polymerase called the **upstream binding site (UBS)**. This protein–RNA binding slows down hairpin formation and so makes termination less likely. But NusA loosens the association between the RNA and the UBS, thereby stimulating hairpin formation. This makes termination more likely. In Chapter 8, we will discuss NusA and its mode of action in more detail and see evidence for the model mentioned here.

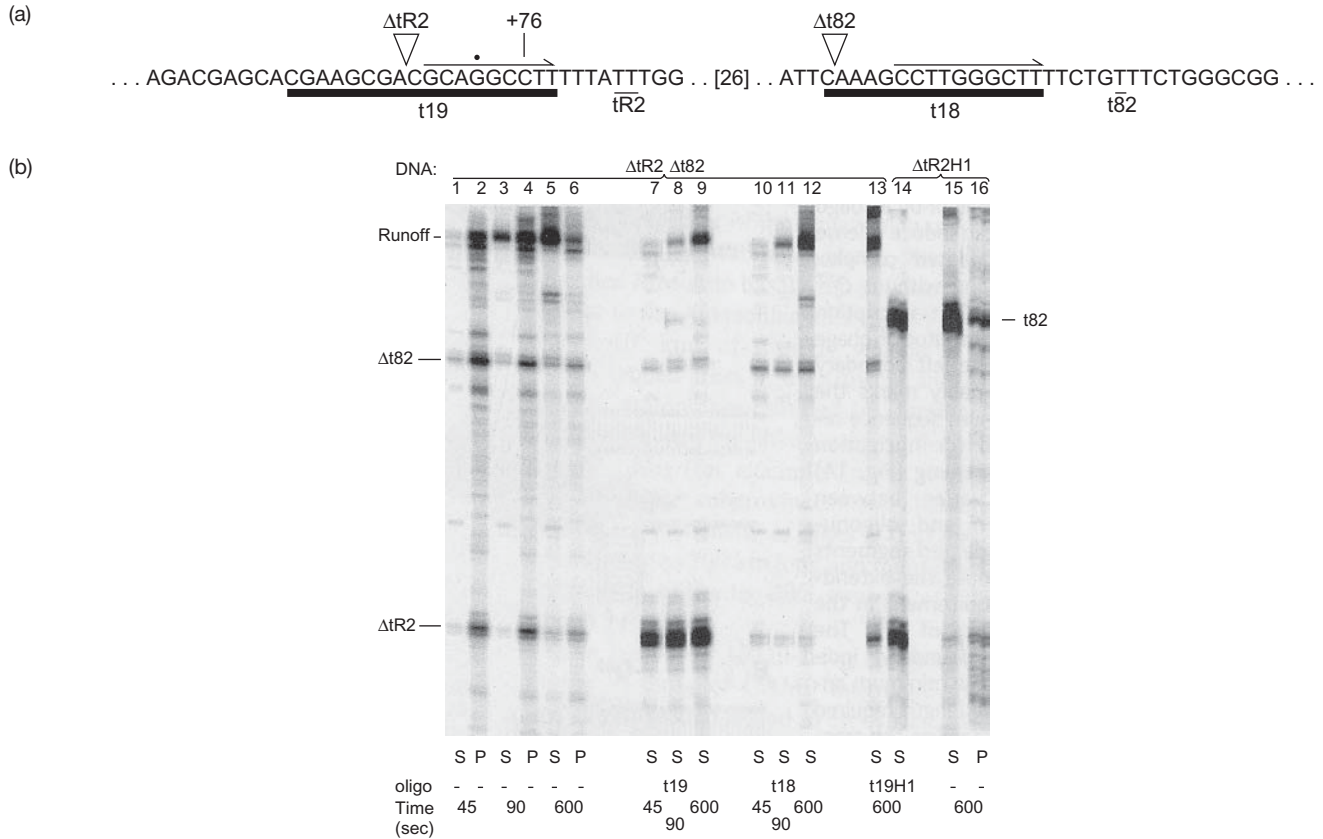


Figure 6.42 Release of transcripts from elongation complexes by oligonucleotides complementary to mutant terminators.

(a) Scheme of the template used in these experiments. The template contained two mutant terminators, $\Delta tR2$, and $\Delta t82$, situated as shown, downstream of a strong promoter. The normal termination sites for these two terminators are labeled with thin underlines. The black bars denote regions complementary to the oligonucleotides used (t19 and t18). The rightward arrows denote the half inverted repeats remaining in the mutant terminators. The dot indicates the site of a base altered in the t19HI oligonucleotide and of a compensating mutation in the DNA template in certain of the experiments. The template was attached to a magnetic bead so it could be removed from solution easily by

centrifugation. (b) Experimental results. Yarnell and Roberts synthesized labeled RNA in the presence of the template in panel (a) and; no oligonucleotide (lanes 1–6 and 15–16), the t19 oligonucleotide (lanes 7–9), the t18 oligonucleotide (lanes 10–12); and the t19HI oligonucleotide (lanes 13–14). They allowed transcription for the times given at bottom, then removed the template and any RNA attached to it by centrifugation. They electrophoresed the labeled RNA in the pellet (P) or supernatant (S), as indicated at bottom, and autoradiographed the gel. The positions of run-off transcripts, and of transcripts that terminated at the $\Delta tR2$ and $\Delta t82$ terminators, are indicated at left. (Source: (a–b) Yarnell, W.S. and Roberts, J.W. Mechanism of intrinsic transcription termination and antitermination. *Science* 284 (23 April 1999) 611–12. © AAAS.)

SUMMARY The essence of a bacterial terminator is twofold: (1) base-pairing of something to the transcript to destabilize the RNA–DNA hybrid; and (2) something that causes transcription to pause. A normal intrinsic terminator satisfies the first condition by causing a hairpin to form in the transcript, and the second by causing a string of U's to be incorporated just downstream of the hairpin.

Rho-Dependent Termination

Jeffrey Roberts discovered rho as a protein that caused an apparent depression of the ability of RNA polymerase to transcribe certain phage DNAs in vitro. This depression is

simply the result of termination. Whenever rho causes a termination event, the polymerase has to reinitiate to begin transcribing again. And, because initiation is a time-consuming event, less net transcription can occur. To establish that rho is really a termination factor, Roberts performed the following experiments.

Rho Affects Chain Elongation, But Not Initiation Just as Travers and Burgess used $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and $[^{14}\text{C}]\text{ATP}$ to measure transcription initiation and total RNA synthesis, respectively, Roberts used $[\gamma\text{-}^{32}\text{P}]\text{GTP}$ and $[^3\text{H}]\text{UTP}$ for the same purposes. He carried out in vitro transcription reactions with these two labeled nucleotides in the presence of increasing concentrations of rho. Figure 6.43 shows the results. We see that rho had little effect on initiation; if anything, the rate of initiation went up. But rho caused a

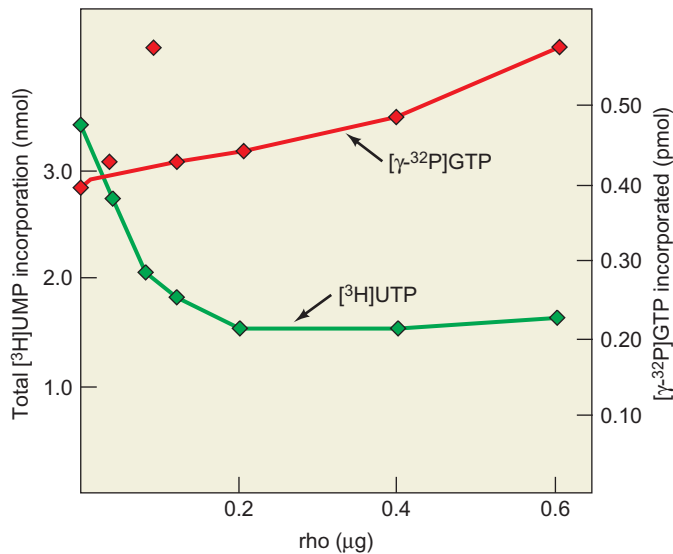


Figure 6.43 Rho decreases the net rate of RNA synthesis. Roberts allowed *E. coli* RNA polymerase to transcribe λ phage DNA in the presence of increasing concentrations of rho. He used $[\gamma\text{-}^{32}\text{P}]\text{GTP}$ to measure initiation (red) and $[\text{H}]\text{UTP}$ to measure elongation (green). Rho depressed the elongation rate, but not initiation. (Source: Adapted from Roberts, J.W. Termination factor for RNA synthesis, *Nature* 224:1168–74, 1969.)

significant decrease in total RNA synthesis. This is consistent with the notion that rho terminates transcription, thus forcing time-consuming reinitiation. This hypothesis predicts that rho would cause shorter transcripts to be made.

Rho Causes Production of Shorter Transcripts It is relatively easy to measure the size of RNA transcripts by gel electrophoresis or, in 1969, when Roberts performed his experiments, by ultracentrifugation. But just finding short transcripts would not have been enough to conclude that rho was causing termination. It could just as easily have been an RNase that chopped up longer transcripts into small pieces.

To exclude the possibility that rho was simply acting as a nuclease, Roberts first made ^3H -labeled λ RNA in the absence of rho, then added these relatively large pieces of RNA to new reactions carried out in the presence of rho, in which ^{14}C UTP was the labeled RNA precursor. Finally, he measured the sizes of the ^{14}C - and ^3H -labeled λ RNAs by ultracentrifugation. Figure 6.44 presents the results. The solid curves show no difference in the size of the preformed ^3H -labeled RNA even when it had been incubated with rho in the second reaction. Rho therefore shows no RNase activity. However, the ^{14}C -labeled RNA made in the presence of rho (red line in Figure 6.44b) is obviously much smaller than the preformed RNA made without rho. Thus, rho is causing the synthesis of much smaller RNAs. Again, this is consistent with the role of rho in terminating transcription. Without rho, the transcripts grew to abnormally large size.

Rho Releases Transcripts from the DNA Template Finally, Roberts used ultracentrifugation to analyze the sedimenta-

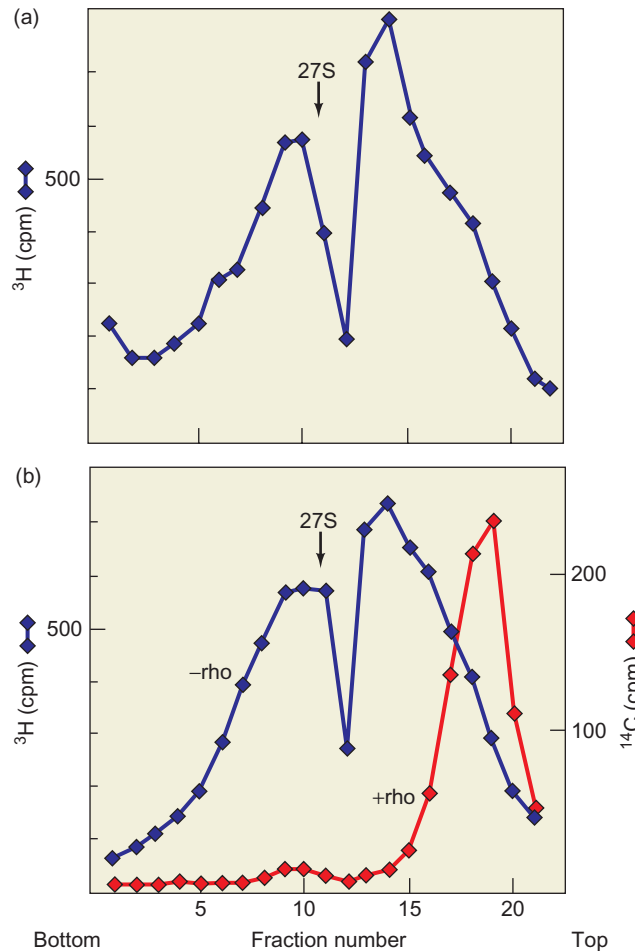


Figure 6.44 Rho reduces the size of the RNA product. (a) Roberts allowed *E. coli* RNA polymerase to transcribe λ DNA in the absence of rho. He included $[\text{H}]\text{UTP}$ in the reaction to label the RNA. Finally, he used ultracentrifugation to separate the transcripts by size. He collected fractions from the bottom of the centrifuge tube, so low-numbered fractions, at left, contained the largest RNAs. (b) Roberts used *E. coli* RNA polymerase to transcribe λ DNA in the presence of rho. He also included ^{14}C ATP to label the transcripts, plus the ^3H -labeled RNA from panel (a). Again, he ultracentrifuged the transcripts to separate them by size. The ^{14}C -labeled transcripts (red) made in the presence of rho were found near the top of the gradient (at right), indicating that they were relatively small. On the other hand, the ^3H -labeled transcripts (blue) from the reaction lacking rho were relatively large and the same size as they were originally. Thus, rho has no effect on the size of previously made transcripts, but it reduces the size of the transcripts made in its presence. (Source: Adapted from Roberts, J.W. Termination factor for RNA synthesis, *Nature* 224:1168–74, 1969.)

tion properties of the RNA products made in the presence and absence of rho. The transcripts made without rho (Figure 6.45a) cosedimented with the DNA template, indicating that they had not been released from their association with the DNA. By contrast, the transcripts made in the presence of rho (Figure 6.45b) sedimented at a much lower rate, independent of the DNA. Thus, rho seems to release RNA transcripts from the DNA template. In fact, rho (the Greek letter ρ) was chosen to stand for “release.”

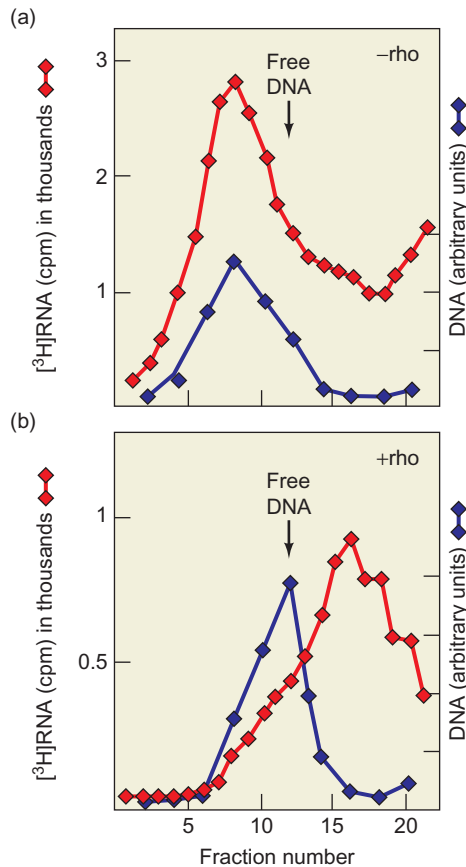


Figure 6.45 Rho releases the RNA product from the DNA template. Roberts transcribed λ DNA under the same conditions as in Figure 6.44, in the (a) absence or (b) presence of rho. Then he subjected the ^3H -labeled product (red) to ultracentrifugation to see whether the product was associated with the DNA template (blue). (a) The RNA made in the absence of rho sedimented together with the template in a complex that was larger than free DNA. (b) The RNA made in the presence of rho sedimented independently of DNA at a position corresponding to relatively small molecules. Thus, transcription with rho releases transcripts from the DNA template. (Source: Adapted from Roberts, J.W. Termination factor for RNA synthesis, *Nature* 224:1168–74, 1969.)

The Mechanism of Rho How does rho do its job? It has been known for some time that rho is able to bind to RNA at a so-called **rho loading site**, or **rho utilization (*rut*) site**, and has ATPase activity that can provide the energy to propel it along an RNA chain. Accordingly, a model has arisen that calls for rho to bind to a nascent RNA, and follow the polymerase by moving along the RNA chain in the 5'→3' direction. This chase continues until the polymerase stalls in the terminator region just after making the RNA hairpin. Then rho can catch up and release the transcript. In support of this hypothesis, Terry Platt and colleagues showed in 1987 that rho has RNA–DNA helicase activity that can unwind an RNA–DNA hybrid. Thus, when rho encounters the polymerase stalled at the terminator, it can unwind the RNA–DNA hybrid within the transcription bubble, releasing the RNA and terminating transcription.

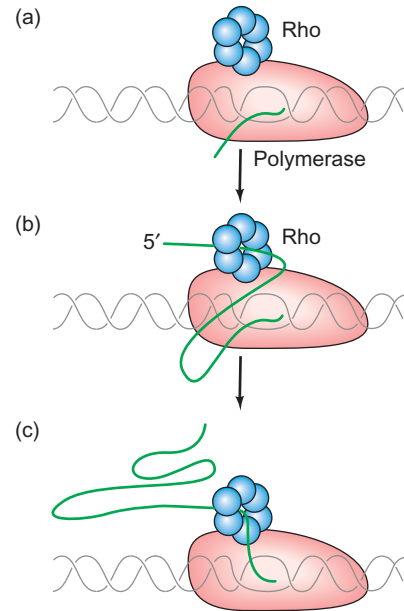


Figure 6.46 A model of rho-dependent termination. (a) Rho (blue) has joined the elongation complex by binding directly to RNA polymerase. The end of the nascent transcript (green) has just emerged from the polymerase. (b) The transcript has lengthened and has bound to rho via a rho loading site, forming an RNA loop. Rho can now feed the transcript through its central cavity. (c) The polymerase has paused at a terminator. By continuously feeding the transcript through itself, rho has tightened the RNA loop and irreversibly trapped the elongation complex. Rho has also begun to dissociate the RNA–DNA hybrid, which will lead to transcript release.

Evgeny Nudler and colleagues presented evidence in 2010 that this attractive hypothesis is probably wrong. These workers used their transcription walking method, as described earlier in this chapter, using His₆-tagged rho coupled to nickel beads. They found that elongation complexes (ECs) with RNA products only 11 nt long were retained by the beads. Because an 11-nt RNA is completely contained within RNA polymerase, this behavior means that the association between rho and the EC must involve the polymerase, not the RNA. Thus, if rho binds directly to the polymerase, it does not need to bind to the nascent RNA first and chase the polymerase until it catches up.

Furthermore, the EC tethered to the rho-nickel beads could be walked along the DNA template without dissociating, proving that the association between rho and the EC is stable. And the complex could terminate normally at rho-dependent terminators, showing that the rho that is bound to the polymerase is capable of sponsoring termination.

If rho is already bound to the polymerase at an early stage in transcription, how does its affinity for RNA come into play in termination? Nudler and colleagues proposed the model in Figure 6.46. First, rho binds to the polymerase when the transcript is still very short. When the transcript grows longer, and includes a rho loading site, the RNA binds to rho. X-ray crystallography studies have

shown that rho is a hexamer of identical subunits arranged in the shape of a lock washer—an open circle with slightly offset ends. This presumably allows the growing RNA to enter the hole in the middle of the hexamer, forming an RNA loop. As transcription progresses, rho continues to feed the RNA product through itself, progressively tightening the RNA loop. Ultimately, when the polymerase encounters a termination signal, it pauses, allowing the RNA loop to tighten so much that further transcription cannot occur. This creates a “trapped” elongation complex. Finally, rho could invade the RNA–DNA hybrid within the polymerase and cause termination in one of two ways: It could use its RNA–DNA helicase activity to unwind the hybrid, or it could unwind the hybrid by physically disrupting it.

SUMMARY Rho-dependent terminators consist of an inverted repeat, which can cause a hairpin to form in the transcript, but no string of T's. Rho binds to the RNA polymerase in an elongation complex. When the RNA transcript has grown long enough, rho binds to it via a rho loading site, forming an RNA loop between the polymerase and rho. Rho continues to feed the growing transcript through itself until the polymerase pauses at a terminator. This pause allows rho to tighten the RNA loop and trap the elongation complex. Rho then dissociates the RNA–DNA hybrid, terminating transcription.

SUMMARY

The catalytic agent in the transcription process is RNA polymerase. The *E. coli* enzyme is composed of a core, which contains the basic transcription machinery, and a σ -factor, which directs the core to transcribe specific genes. The σ -factor allows initiation of transcription by causing the RNA polymerase holoenzyme to bind tightly to a promoter. This σ -dependent tight binding requires local melting of 10–17 bp of the DNA in the vicinity of the transcription start site to form an open promoter complex. Thus, by directing the holoenzyme to bind only to certain promoters, a σ -factor can select which genes will be transcribed. The initiation process continues until 9 or 10 nt have been incorporated into the RNA, the core changes to an elongation-specific conformation, leaves the promoter, and carries on with the elongation process. The σ -factor appears to be released from the core polymerase, but not usually immediately upon promoter clearance. Rather, σ seems to exit from the elongation complex in a stochastic manner during the elongation process. The

σ -factor can be reused by different core polymerases. The core, not σ , governs rifampicin sensitivity or resistance. The *E. coli* RNA polymerase achieves abortive transcription by scrunching: drawing downstream DNA into the polymerase without actually moving and losing its grip on promoter DNA. The scrunched DNA could store enough energy to allow the polymerase to break its bonds to the promoter and begin productive transcription.

Prokaryotic promoters contain two regions centered at –10 and –35 bp upstream of the transcription start site. In *E. coli*, these have the consensus sequences TATAAT and TTGACA, respectively. In general, the more closely regions within a promoter resemble these consensus sequences, the stronger that promoter will be. Some extraordinarily strong promoters contain an extra element (an UP element) upstream of the core promoter. This makes these promoters better binding sites for RNA polymerase.

Four regions are similar among σ -factors, and subregions 2.4 and 4.2 are involved in promoter –10 box and –35 box recognition, respectively.

The core subunit β lies near the active site of the RNA polymerase where phosphodiester bonds are formed. The σ -factor is also nearby during the initiation phase. The α -subunit has independently folded N-terminal and C-terminal domains. The C-terminal domain can recognize and bind to a promoter's UP element. This allows very tight binding between polymerase and promoter.

Elongation of transcription involves the polymerization of nucleotides as the RNA polymerase core travels along the template DNA. As it moves, the polymerase maintains a short melted region of template DNA. This transcription bubble is 11–16 bases long and contains an RNA–DNA hybrid about 9 bp long. The movement of the transcription bubble requires that the DNA unwind ahead of the advancing polymerase and close up again behind it. This process introduces strain into the template DNA that is relaxed by topoisomerases.

The crystal structure of the *T. aquaticus* RNA polymerase core is shaped like a crab claw. The catalytic center, containing a Mg^{2+} ion coordinated by three Asp residues, lies in a channel that conducts DNA through the enzyme.

The crystal structure of a *T. aquaticus* holoenzyme–DNA complex mimicking an open promoter complex allows the following conclusions. (1) The DNA is bound mainly to the σ -subunit. (2) The predicted interactions between amino acids in region 2.4 of σ and the –10 box of the promoter are really possible. (3) Three highly conserved aromatic amino acids that are predicted to participate in promoter melting are really in a position to do so. (4) Two invariant basic amino acids in σ that are predicted to participate in DNA binding are in proper position to do so. A higher resolution crystal structure reveals a form of the polymerase that has two Mg^{2+} ions, in accord with the probable mechanism of catalysis.

Structural studies of the elongation complex involving the *Thermus thermophilus* RNA polymerase revealed that: A valine residue in the β' subunit inserts into the minor groove of the downstream DNA; thus, it could prevent the DNA from slipping, and it could induce the screw-like motion of the DNA through the enzyme. Only one base pair of DNA (at position +1) is melted and available for base-pairing with an incoming nucleotide, so only one nucleotide at a time can bind specifically to the complex. Several forces limit the length of the RNA–DNA hybrid, including the length of the cavity in the enzyme that accommodates the hybrid and a hydrophobic pocket in the enzyme at the end of the cavity that traps the first RNA base displaced from the hybrid. The RNA product in the exit channel assumes the shape of one-half of a double-stranded RNA. Thus, it can readily form a hairpin to cause pausing, or even termination of transcription. Structural studies of the enzyme with an inactive substrate analog and the antibiotic streptolydigin have identified a preinsertion state for the substrate that is catalytically inactive, but could provide for checking that the substrate is the correct one.

Intrinsic terminators have two important elements: (1) an inverted repeat that allows a hairpin to form at the end of the transcript to destabilize the RNA–DNA hybrid; (2) a string of T's in the nontemplate strand that results in a string of weak rU–dA base pairs holding the transcript to the template. Together, these elements cause the polymerase to pause and the transcript to be released. Rho-dependent terminators consist of an inverted repeat, which can cause a hairpin to form in the transcript, but no string of T's. Rho binds to the RNA polymerase in an elongation complex. When the RNA transcript has grown long enough, rho binds to it via a rho loading site, forming an RNA loop between the polymerase and rho. Rho continues to feed the growing transcript through itself until the polymerase pauses at a terminator. This pause allows rho to tighten the RNA loop and trap the elongation complex. Rho then dissociates the RNA–DNA hybrid, terminating transcription.

REVIEW QUESTIONS

1. Explain the following findings: (1) Core RNA polymerase transcribes intact T4 phage DNA only weakly, whereas holoenzyme transcribes this template very well; but (2) core polymerase can transcribe calf thymus DNA about as well as the holoenzyme can.
2. How did Bautz and colleagues show that the holoenzyme transcribes phage T4 DNA asymmetrically, but the core transcribes this DNA symmetrically?
3. Describe an experiment to measure the dissociation rate of the tightest complex between a protein and a DNA. Show sample results of weak and tight binding. How do these results relate to the binding of core polymerase and holoenzyme to DNA that contains promoters?
4. What effect does temperature have on the dissociation rate of polymerase–promoter complexes? What does this suggest about the nature of the complex?
5. Diagram the difference between a closed and an open promoter complex.
6. Diagram a typical prokaryotic promoter, and a promoter with an UP element. Exact sequences are not necessary.
7. Describe and give the results of an experiment that demonstrates the formation of abortive transcripts by *E. coli* RNA polymerase.
8. Diagram the four-step transcription initiation process in *E. coli*.
9. Describe and show the results of an experiment that measures the effects of σ on transcription initiation and elongation rates.
10. How can you show that σ does not really accelerate the rate of transcription elongation?
11. What final conclusion can you draw from the experiments in the previous two questions?
12. Describe and show the results of an experiment that demonstrates the reuse of σ . On the same graph, show the results of an experiment that shows that the core polymerase determines resistance to rifampicin.
13. Draw a diagram of the “ σ -cycle,” assuming σ dissociates from core during elongation.
14. Describe and show the results of a fluorescence resonance energy transfer (FRET) experiment that suggests that σ does not dissociate from the core polymerase during elongation.
15. In the σ -cycle, what is obligate release and what is stochastic release? Which is the favored hypothesis?
16. Propose three hypotheses for the mechanism of abortive transcription in *E. coli*. Describe and give the results of a FRET experiment that supports one of these hypotheses.
17. Describe and show the results of an experiment that shows which base pairs are melted when RNA polymerase binds to a promoter. Explain how this procedure works.
18. Describe and show the results of an experiment that gives an estimate of the number of base pairs melted during transcription by *E. coli* RNA polymerase.
19. What regions of the σ -factor are thought to be involved in recognizing (1) the -10 box of the promoter and (2) the -35 box of the promoter? Without naming specific residues, describe the genetic evidence for these conclusions.
20. Describe a binding assay that provides biochemical evidence for interaction between σ -region 4.2 and the -35 box of the promoter.
21. Cite evidence to support the hypothesis that the α -subunit of *E. coli* RNA polymerase is involved in recognizing a promoter UP element.

22. Describe how limited proteolysis can be used to define the domains of a protein such as the α -subunit of *E. coli* RNA polymerase.
23. Describe an experiment to determine which polymerase subunit is responsible for rifampicin and streptolydigin resistance or sensitivity.
24. Describe and give the results of an experiment that shows that the β -subunit of *E. coli* RNA polymerase is near the active site that forms phosphodiester bonds.
25. Describe an RNA–DNA cross-linking experiment that demonstrates the existence of an RNA–DNA hybrid at least 8 bp long within the transcription elongation complex.
26. Draw a rough sketch of the structure of a bacterial RNA polymerase core based on x-ray crystallography. Point out the positions of the subunits of the enzyme, the catalytic center, and the rifampicin-binding site. Based on this structure, propose a mechanism for inhibition of transcription by rifampicin.
27. Based on the crystal structure of the *E. coli* elongation complex, what factors limit the length of the RNA–DNA hybrid?
28. Based on the crystal structures of the *E. coli* elongation complex with and without the antibiotic streptolydigin, propose a mechanism for the antibiotic.
29. Draw a rough sketch of the crystal structure of the holoenzyme–DNA complex in the open promoter form. Focus on the interaction between the holoenzyme and DNA. What enzyme subunit plays the biggest role in DNA binding?
30. Sigma regions 2.4 and 4.2 are known to interact with the -10 and -35 boxes of the promoter, respectively. What parts of this model are confirmed by the crystal structure of the holoenzyme–DNA complex? Provide explanations for the parts that are not confirmed.
31. Present two models for the way the RNA polymerase can maintain the bubble of melted DNA as it moves along the DNA template. Which of these models is favored by the evidence? Cite the evidence in a sentence or two.
32. What are the two important elements of an intrinsic transcription terminator? How do we know they are important? (Cite evidence.)
33. Present evidence that a hairpin is not required for pausing at an intrinsic terminator.
34. Present evidence that base-pairing (of something) with the RNA upstream of a pause site is required for intrinsic termination.
35. What does a rho-dependent terminator look like? What role is rho thought to play in such a terminator?
36. How can you show that rho causes a decrease in net RNA synthesis, but no decrease in chain initiation? Describe and show the results of an experiment.
37. Describe and show the results of an experiment that demonstrates the production of shorter transcripts in the presence of rho. This experiment should also show that rho does not simply act as a nuclease.
38. Describe and show the results of an experiment that demonstrates that rho releases transcripts from the DNA template.

ANALYTICAL QUESTIONS

1. Draw the structure of an RNA hairpin with a 10-bp stem and a 5-nt loop. Make up a sequence that will form such a structure. Show the sequence in the linear as well as the hairpin form.
2. An *E. coli* promoter recognized by the RNA polymerase holoenzyme containing σ^{70} has a -10 box with the following sequence in the nontemplate strand: 5'-CATAGT-3'. (a) Would a C→T mutation in the first position likely be an up or a down mutation? (b) Would a T→A mutation in the last position likely be an up or down mutation? Explain your answers.
3. You are carrying out experiments to study transcription termination in an *E. coli* gene. You sequence the 3'-end of the gene and get the following results:
 5' – CGAAGCGCCGATTGCCGCGCTTTTTTTTTT –3'
 3' – GCTTCGCGGCTAACGCGCCGCGAAAAA –5'
 You then create mutant genes with this sequence changed to the following (top, or nontemplate strand, 5'→3'):
 Mutant A: CGAAACTAAGATTGCAGCAGTTTTTTTTT
 Mutant B: CGAAGCGCCGTAGCACGCGCTTTTTTTTTT
 Mutant C: CGAAGCGCCGATTGCCGCGCTTACGGCCC
 You put each of the mutant genes into an assay that measures termination and get the following results:

Mutant Gene Tested	Without Rho	With Rho
Wild-type gene	100% termination	100% termination
Mutant A	40% termination	40% termination
Mutant B	95% termination	95% termination
Mutant C	20% termination	80% termination

- a. Draw the structure of the RNA molecule that results from transcription of the wild-type sequence above.
- b. Explain these experimental results as completely as possible.
4. Examine the sequences below and determine the consensus sequence.
 TAGGACT – TCGCAGA – AAGCTTG – TACCAAG – TTCCTCG

SUGGESTED READINGS

General References and Reviews

- Busby, S. and R.H. Ebright. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* 79:743–46.
- Cramer, P. 2007. Extending the message. *Nature* 448:142–43.
- Epshtein, V., D. Dutta, J. Wade, and E. Nudler. 2010. An allosteric mechanism of Rho-dependent transcription termination. *Nature* 463:245–50.

- Geiduschek, E.P. 1997. Paths to activation of transcription. *Science* 275:1614–16.
- Helmann, J.D. and M.J. Chamberlin. 1988. Structure and function of bacterial sigma factors. *Annual Review of Biochemistry* 57:839–72.
- Landick, R. 1999. Shifting RNA polymerase into overdrive. *Science* 284:598–99.
- Landick, R. and J.W. Roberts. 1996. The shrewd grasp of RNA polymerase. *Science* 273:202–3.
- Mooney, R.A., S.A. Darst, and R. Landick. 2005. Sigma and RNA polymerase: An on-again, off-again relationship? *Molecular Cell* 20:335–46.
- Richardson, J.P. 1996. Structural organization of transcription termination factor rho. *Journal of Biological Chemistry* 271:1251–54.
- Roberts, J.W. 2006. RNA polymerase, a scrunching machine. *Science* 314:1097–98.
- Young, B.A., T.M. Gruber, and C.A. Gross. 2002. Views of transcription initiation. *Cell* 109:417–20.
- ## Research Articles
- Bar-Nahum, G. and E. Nudler. 2001. Isolation and characterization of σ^{70} -retaining transcription elongation complexes from *E. coli*. *Cell* 106:443–51.
- Bautz, E.K.F., F.A. Bautz, and J.J. Dunn. 1969. *E. coli* σ factor: A positive control element in phage T4 development. *Nature* 223:1022–24.
- Blatter, E.E., W. Ross, H. Tang, R.L. Gourse, and R.H. Ebricht. 1994. Domain organization of RNA polymerase α subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell* 78:889–96.
- Brennan, C.A., A.J. Dombroski, and T. Platt. 1987. Transcription termination factor rho is an RNA–DNA helicase. *Cell* 48:945–52.
- Burgess, R.R., A.A. Travers, J.J. Dunn, and E.K.F. Bautz. 1969. Factor stimulating transcription by RNA polymerase. *Nature* 221:43–46.
- Campbell, E.A., N. Korzheva, A. Mustaev, K. Murakami, S. Nair, A. Goldfarb, and S.A. Darst. 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* 104:901–12.
- Carpousis, A.J. and J.D. Gralla. 1980. Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the *lac* UV5 promoter. *Biochemistry* 19:3245–53.
- Dombroski, A.J., W.A. Walter, M.T. Record, Jr., D.A. Siegle, and C.A. Gross. (1992). Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA. *Cell* 70:501–12.
- Farnham, P.J. and T. Platt. 1980. A model for transcription termination suggested by studies on the *trp* attenuator in vitro using base analogs. *Cell* 20:739–48.
- Grachev, M.A., T.I. Kolocheva, E.A. Lukhtanov, and A.A. Mustaev. 1987. Studies on the functional topography of *Escherichia coli* RNA polymerase: Highly selective affinity labelling of initiating substrates. *European Journal of Biochemistry* 163:113–21.
- Hayward, R.S., K. Igarashi, and A. Ishihama. 1991. Functional specialization within the α -subunit of *Escherichia coli* RNA polymerase. *Journal of Molecular Biology* 221:23–29.
- Heil, A. and W. Zillig. 1970. Reconstitution of bacterial DNA-dependent RNA polymerase from isolated subunits as a tool for the elucidation of the role of the subunits in transcription. *FEBS Letters* 11:165–71.
- Hinkle, D.C. and M.J. Chamberlin. 1972. Studies on the binding of *Escherichia coli* RNA polymerase to DNA: I. The role of sigma subunit in site selection. *Journal of Molecular Biology* 70:157–85.
- Hsieh, T.-s. and J.C. Wang. 1978. Physicochemical studies on interactions between DNA and RNA polymerase: Ultraviolet absorbance measurements. *Nucleic Acids Research* 5:3337–45.
- Kapanidis, A.N., E. Margeat, S. O. Ho, E. Kortkhonja, S. Weiss, and R.H. Ebricht. 2006. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314:1144–47.
- Malhotra, A., E. Severinova, and S.A. Darst. 1996. Crystal structure of a σ^{70} subunit fragment from *E. coli* RNA polymerase. *Cell* 87:127–36.
- Mukhopadhyay, J., A.N. Kapanidis, V. Mekler, E. Kortkhonja, Y.W. Ebricht, and R.H. Ebricht. 2001. Translocation of σ^{70} with RNA polymerase during transcription: Fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* 106:453–63.
- Murakami, K.S., S. Masuda, E.A. Campbell, O. Muzzin, and S.A. Darst. 2002. Structural basis of transcription initiation: An RNA polymerase holoenzyme–DNA complex. *Science* 296:1285–90.
- Nudler, E., A. Mustaev, E. Lukhtanov, and A. Goldfarb. 1997. The RNA–DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* 89:33–41.
- Paul, B.J., M.M. Barker, W. Ross, D.A. Schneider, C. Webb, J.W. Foster, and R.L. Gourse. 2004. DskA, a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* 118:311–22.
- Revyakin, A., C. Liu, R.H. Ebricht, and T.R. Strick. 2006. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* 314:1139–43.
- Roberts, J.W. 1969. Termination factor for RNA synthesis. *Nature* 224:1168–74.
- Ross, W., K.K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov, and R.L. Gourse. 1993. A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase. *Science* 262:1407–13.
- Saucier, J.-M. and J.C. Wang. 1972. Angular alteration of the DNA helix by *E. coli* RNA polymerase. *Nature New Biology* 239:167–70.
- Sidorenkov, I., N. Komissarova, and M. Kashlev. 1998. Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Molecular Cell* 2:55–64.
- Siebenlist, U. 1979. RNA polymerase unwinds an 11-base pair segment of a phage T7 promoter. *Nature* 279:651–52.
- Toulkhonov, I., I. Artsimovitch, and R. Landick. 2001. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* 292:730–33.
- Travers, A.A. and R.R. Burgess. 1969. Cyclic re-use of the RNA polymerase sigma factor. *Nature* 222:537–40.

- Vassilyev, D.G., S.-i Sekine, O. Laptenko, J. Lee, M.N. Vassilyeva, S. Borukhov, and S. Yokoyama. 2002. Crystal structure of bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417:712–19.
- Vassilyev, D.G., M.N. Vassilyeva, A. Perederina, T.H. Tahirov, and I. Artsimovitch. 2007. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* 448:157–62.
- Vassilyev, D.G., M.N. Vassilyeva, J. Zhang, M. Palangat, and I. Artsimovitch. 2007. Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 448:163–68.
- Yarnell, W.S. and J.W. Roberts. 1999. Mechanism of intrinsic transcription termination and antitermination. *Science* 284:611–15.
- Young, B.A., L.C. Anthony, T.M. Gruber, T.M. Arthur, E. Heyduk, C.Z. Lu, M.M. Sharp, T. Heyduk, R.R. Burgess, and C.A. Gross. 2001. A coiled-coil from the RNA polymerase β' subunit allosterically induces selective nontemplate strand binding by σ^{70} . *Cell* 105:935–44.
- Zhang, G., E.A. Campbell, L. Minakhin, C. Richter, K. Severinov, and S.A. Darst. 1999. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98:811–24.
- Zhang, G. and S.A. Darst. 1998. Structure of the *Escherichia coli* RNA polymerase α subunit amino terminal domain. *Science* 281:262–66.