CHAPTER 7

RNA Synthesis and Processing

CHAPTERS 5 AND 6 DISCUSSED THE ORGANIZATION and maintenance of genomic DNA, which can be viewed as the set of genetic instructions governing all cellular activities. These instructions are implemented via the synthesis of RNAs and proteins. Importantly, the behavior of a cell is determined not only by what genes it inherits but also by which of those genes are expressed at any given time. Regulation of gene expression allows cells to adapt to changes in their environments and is responsible for the distinct activities of the multiple differentiated cell types that make up complex plants and animals. Muscle cells and liver cells, for example, contain the same genes; the functions of these cells are determined not by differences in their genomes, but by regulated patterns of gene expression that govern development and differentiation.

The first step in expression of a gene, the transcription of DNA into RNA, is the initial level at which gene expression is regulated in both prokaryotic and eukaryotic cells. RNAs in eukaryotic cells are then modified in various ways—for example, introns are removed by splicing—to convert the primary transcript into its functional form. Different types of RNA play distinct roles in cells: Messenger RNAs (mRNAs) serve as templates for protein synthesis; ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) function in mRNA translation. Still other small RNAs function in gene regulation, mRNA splicing, rRNA processing, and protein sorting in eukaryotes. In fact, some of the most exciting advances in recent years have pertained to the roles of noncoding RNAs (microRNAs) as regulators of both transcription and translation in eukaryotic cells. Transcription and RNA processing are discussed in this chapter. The final step in gene expression, the translation of mRNA to protein, is the subject of Chapter 8.

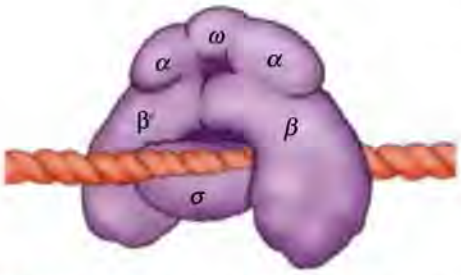

FIGURE 7.1 *E. coli* RNA polymerase The complete enzyme consists of six subunits: two α, one β, one β', one ω and one σ. The σ subunit is relatively weakly bound and can be dissociated from the other five subunits, which constitute the core polymerase.

Transcription in Prokaryotes

As in most areas of molecular biology, studies of *E. coli* have provided the model for subsequent investigations of transcription in eukaryotic cells. As reviewed in Chapter 4, mRNA was discovered first in *E. coli*. *E. coli* was also the first organism from which RNA polymerase was purified and studied. The basic mechanisms by which transcription is regulated were likewise elucidated by pioneering experiments in *E. coli* in which regulated gene expression allows the cell to respond to variations in the environment, such as changes in the availability of nutrients. An understanding of transcription in *E. coli* has thus provided the foundation for studies of the far more complex mechanisms that regulate gene expression in eukaryotic cells.

RNA Polymerase and Transcription

The principal enzyme responsible for RNA synthesis is **RNA polymerase**, which catalyzes the polymerization of ribonucleoside 5′-triphosphates (NTPs) as directed by a DNA template. The synthesis of RNA is similar to that of DNA, and like DNA polymerase, RNA polymerase catalyzes the growth of RNA chains always in the 5′ to 3′ direction. Unlike DNA polymerase, however, RNA polymerase does not require a preformed primer to initiate the synthesis of RNA. Instead, transcription initiates *de novo* at specific sites at the beginning of genes. The initiation process is particularly important because this is a major step at which transcription is regulated.

RNA polymerase, like DNA polymerase, is a complex enzyme made up of multiple polypeptide chains. The intact bacterial enzyme consists of five different types of subunits, called α, β, β', ω and σ (Figure 7.1). The σ subunit is relatively weakly bound and can be separated from the other subunits, yielding a core polymerase consisting of two α, one β, one β' and one ω subunits. The core polymerase is fully capable of catalyzing the polymerization of NTPs into RNA, indicating that σ is not required for the basic catalytic activity of the enzyme. However, the core polymerase does not bind specifically to the DNA sequences that signal the normal initiation of transcription; therefore the σ subunit is required to identify the correct sites for transcription initiation. The selection of these sites is a critical element of transcription because synthesis of a functional RNA must start at the beginning of a gene.

The DNA sequence to which RNA polymerase binds to initiate transcription of a gene is called the **promoter**. The DNA sequences involved in promoter function were first identified by comparisons of the nucleotide sequences of a series of different genes isolated from *E. coli*. These comparisons revealed that the region upstream of the transcription initiation site contains two sets of sequences that are similar in a variety of genes. These common sequences encompass six nucleotides each and are located approximately 10 and 35 base pairs upstream of the transcription start site (Figure 7.2). They are called the –10 and –35 elements, denoting their position relative to the transcription initiation site, which is defined as the +1 position. The sequences at the –10 and –35 positions in different promoters are not

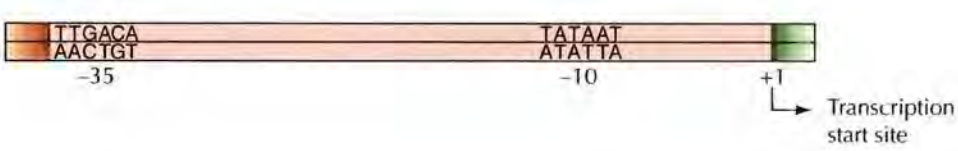

FIGURE 7.2 Sequences of *E. coli* promoters *E. coli* promoters are characterized by two sets of sequences located 10 and 35 base pairs upstream of the transcription start site (+1). The consensus sequences shown correspond to the bases most frequently found in different promoters.

identical, but they are all similar enough to establish consensus sequences—the bases most frequently found at each position.

Several types of experimental evidence support the functional importance of the –10 and –35 promoter elements. First, genes with promoters that differ from the consensus sequences are transcribed less efficiently than genes whose promoters match the consensus sequences more closely. Second, mutations introduced in either the –35 or –10 consensus sequences have strong effects on promoter function. Third, the sites at which RNA polymerase binds to promoters have been directly identified by **footprinting** experiments, which are widely used to determine the sites at which proteins bind to DNA (Figure 7.3). In experiments of this type, a DNA frag-

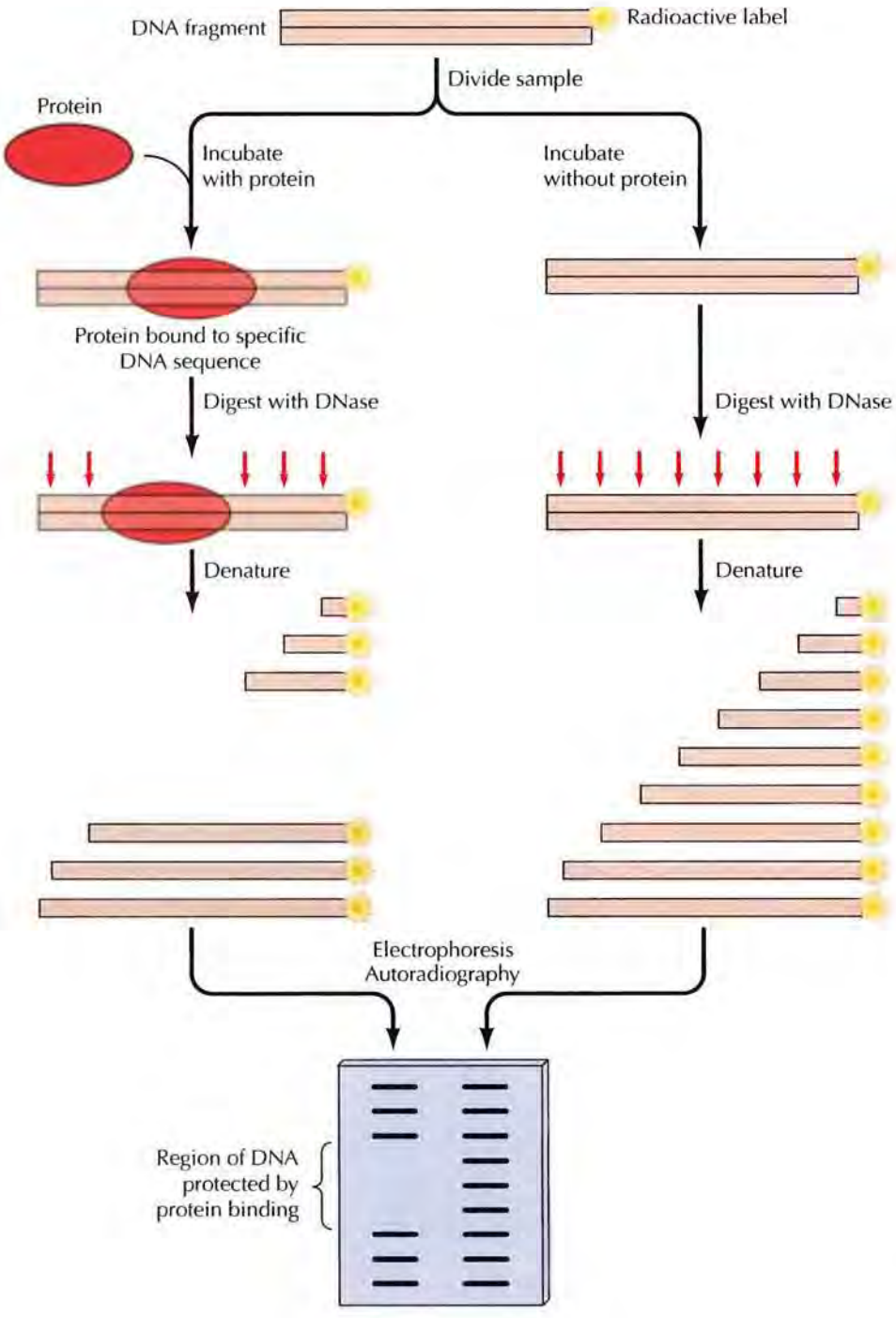

FIGURE 7.3 DNA footprinting A sample containing fragments of DNA radiolabeled at one end is divided in two, and one half of the sample is incubated with a protein that binds to a specific DNA sequence within the fragment. Both samples are then digested with DNase, under conditions such that the DNase introduces an average of one cut per molecule. The region of DNA bound to the protein is protected from DNase digestion. The DNA-protein complexes are then denatured, and the sizes of the radiolabeled DNA fragments produced by DNase digestion are analyzed by electrophoresis. Fragments of DNA resulting from DNase cleavage within the region protected by protein binding are missing from the sample of DNA that was incubated with protein.

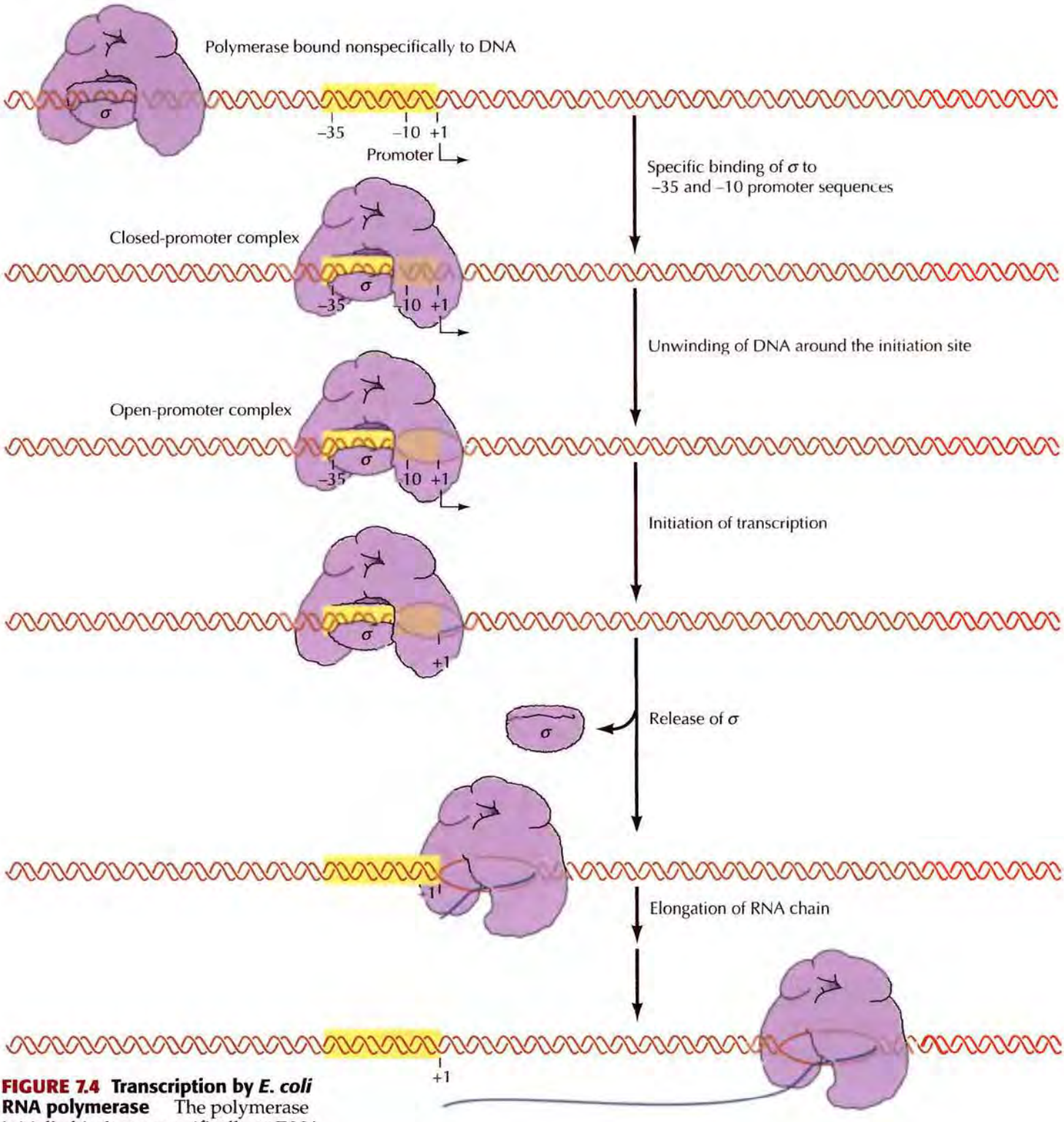

FIGURE 7.4 Transcription by *E. coli* RNA polymerase The polymerase initially binds nonspecifically to DNA and migrates along the molecule until the σ subunit binds to the –35 and –10 promoter elements, forming a closed-promoter complex. The polymerase then unwinds DNA around the initiation site, and transcription is initiated by the polymerization of free NTPs. The σ subunit then dissociates from the core polymerase, which migrates along the DNA and elongates the growing RNA chain.

ment is radiolabeled at one end. The labeled DNA is incubated with the protein of interest (e.g., RNA polymerase) and then subjected to partial digestion with DNase. The principle of the method is that the regions of DNA to which the protein binds are protected from DNase digestion. These regions can therefore be identified by comparison of the digestion products of the protein-bound DNA with those resulting from identical DNase treatment of a parallel sample of DNA that was not incubated with protein. Vari-

ations of this basic method, which employ chemical reagents to modify and cleave DNA at particular nucleotides, can be used to identify the specific DNA bases that are in contact with protein. Such footprinting analysis has shown that RNA polymerase generally binds to promoters over approximately a 60-base-pair region, extending from –40 to +20 (i.e., from 40 nucleotides upstream to 20 nucleotides downstream of the transcription start site). The σ subunit binds specifically to sequences in both the –35 and –10 promoter regions, substantiating the importance of these sequences in promoter function. In addition, some *E. coli* promoters have a third sequence, located upstream of the –35 region, that serves as a specific binding site for the RNA polymerase α subunit.

In the absence of σ, RNA polymerase binds nonspecifically to DNA with low affinity. The role of σ is to direct the polymerase to promoters by binding specifically to both the –35 and –10 sequences, leading to the initiation of transcription at the beginning of a gene (Figure 7.4). The initial binding between the polymerase and a promoter is referred to as a closed-promoter complex because the DNA is not unwound. The polymerase then unwinds 12–14 bases of DNA, from about –12 to +2, to form an open-promoter complex in which single-stranded DNA is available as a template for transcription. Transcription is initiated by the joining of two free NTPs. After addition of about the first 10 nucleotides, σ is released from the polymerase, which then leaves the promoter and moves along the template DNA to continue elongation of the growing RNA chain.

During elongation, the polymerase remains associated with its template while it continues synthesis of mRNAs. As it travels, the polymerase unwinds the template DNA ahead of it and rewinds the DNA behind it, maintaining an unwound region of about 15 base pairs in the region of transcription. Within this unwound portion of DNA, 8–9 bases of the growing RNA chain are bound to the complementary template DNA strand. High resolution structural analysis of bacterial RNA polymerase indicates that the β and β' subunits form a crab claw-like structure that grips the DNA template (Figure 7.5). An internal channel between the β and β' subunits accommodates approximately 20 base pairs of DNA and contains the polymerase active site.

7.1 WEBSITE ANIMATION

Transcription

Transcription is DNA-directed RNA synthesis, catalyzed by the enzyme RNA polymerase.

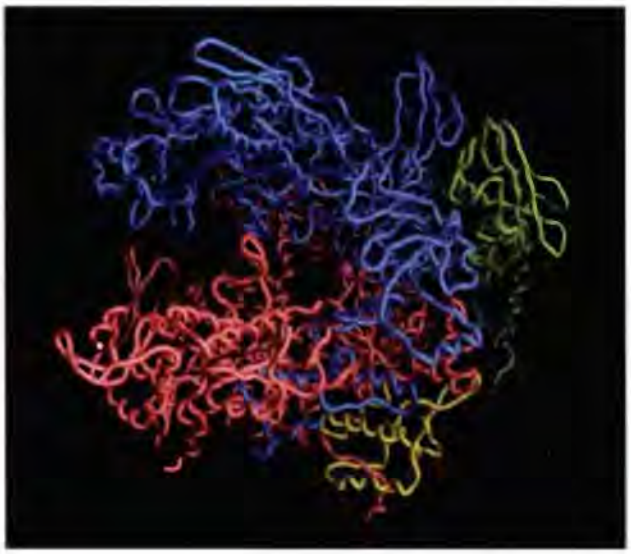

FIGURE 7.5 Structure of bacterial RNA polymerase The α subunits of the polymerase are colored dark green and light green, β blue, β' pink, and ω yellow. (Courtesy of Seth Darst, Rockefeller University.)

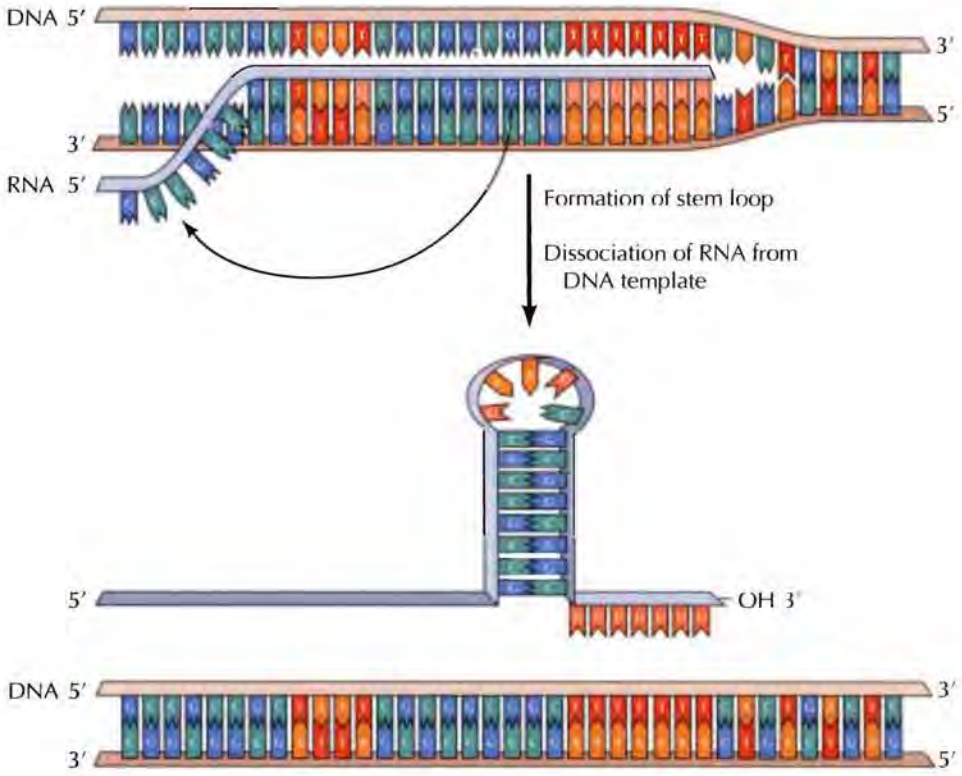

FIGURE 7.6 Transcription termination The termination of transcription is signaled by a GC-rich inverted repeat followed by seven A residues. The inverted repeat forms a stable stem-loop structure in the RNA, causing the RNA to dissociate from the DNA template.

RNA synthesis continues until the polymerase encounters a termination signal, at which point transcription stops, the RNA is released from the polymerase, and the enzyme dissociates from its DNA template. There are two alternative mechanisms for termination of transcription in *E. coli*. The simplest and most common type of termination signal consists of a symmetrical inverted repeat of a GC-rich sequence followed by approximately seven A residues (Figure 7.6). Transcription of the GC-rich inverted repeat results in the formation of a segment of RNA that can form a stable stem-loop structure by complementary base pairing. The formation of such a self-complementary structure in the RNA disrupts its association with the DNA template and terminates transcription. Because hydrogen bonding between A and U is weaker than that between G and C, the presence of A residues downstream of the inverted repeat sequences is thought to facilitate the dissociation of the RNA from its template. Alternatively, the transcription of some genes is terminated by a specific termination protein (called Rho), which binds extended segments (greater than 60 nucleotides) of single-stranded RNA. Since mRNAs in bacteria become associated with ribosomes and are translated while they are being transcribed, such extended regions of single-stranded RNA are exposed only at the end of an mRNA.