For background material on the structure of RNA, you might review section 2.18.

REVIEW

1. How were Beadle and Tatum able to conclude that a gene encoded a specific enzyme?
2. Distinguish between the two-dimensional and three-dimensional structure of RNAs.

11.2 The Role of RNA Polymerases in Transcription

Transcription is a process in which a DNA strand provides the information for the synthesis of an RNA strand. The enzymes responsible for transcription in both prokaryotic and eukaryotic cells are called **DNA-dependent RNA polymerases**, or simply **RNA polymerases**. These enzymes are able to incorporate nucleotides, one at a time, into a strand of RNA whose sequence is complementary to one of the DNA strands, which serves as the *template*.

The first step in the synthesis of an RNA is the association of the polymerase with the DNA template. This brings up a matter of more general interest, namely, the specific interactions of two very different macromolecules, proteins and nucleic acids. Just as different proteins have evolved to bind different types of substrates and catalyze different types of reactions, so too have some of them evolved to recognize and bind to specific sequences of nucleotides in a strand of nucleic acid. The site on the DNA to which an RNA polymerase molecule binds prior to initiating transcription is called the **promoter**. Cellular RNA polymerases are not capable of recognizing promoters on their own but require the help of additional proteins called **transcription factors**. In addition to providing a binding site for the polymerase, the promoter contains the information that determines which of the two DNA strands is transcribed and the site at which transcription begins.

RNA polymerase moves along the template DNA strand toward its 5′ end (i.e., in a 3′→5′ direction). As the polymerase progresses, the DNA is temporarily unwound, and the polymerase assembles a complementary strand of RNA that grows starting from its 5′ terminus in a 3′ direction (**FIGURE 11.4a,b**). RNA polymerase catalyzes the highly favorable reaction

$$RNA_n + NTP \rightarrow RNA_{n+1} + PP_i$$

in which ribonucleoside triphosphate substrates (NTPs) are cleaved into nucleoside monophosphates as they are polymerized into a covalent chain (Figure 11.4*c*). Reactions leading to the synthesis of nucleic acids (and proteins) are inherently different from those of intermediary metabolism discussed in Chapter 3. Whereas some of the reactions leading to the formation of small molecules, such as amino acids, may be close enough to equilibrium that a considerable reverse reaction can be measured, those reactions leading to the synthesis of nucleic acids and proteins must occur under conditions in which there is virtually no reverse reaction. This condition is met during transcription with the aid of a second favorable reaction

$$PP_i \rightarrow 2P_i$$

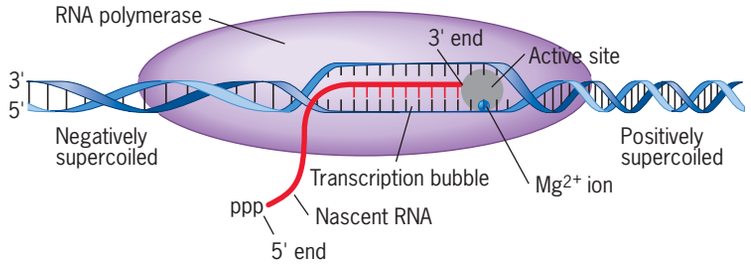
catalyzed by a different enzyme, a pyrophosphatase. In this case, the pyrophosphate (PP_i) produced in the first reaction is hydrolyzed to inorganic phosphate (P_i). The hydrolysis of pyrophosphate releases a large amount of free energy and makes the nucleotide incorporation reaction essentially irreversible.

As the polymerase moves along the DNA template, it incorporates complementary nucleotides into the growing RNA chain. A nucleotide is incorporated into the RNA strand if it is able to form a proper (Watson-Crick) base pair with the nucleotide in the DNA strand being transcribed. This can be seen in Figure 11.4*c* where the incoming adenosine 5'-triphosphate pairs with the thymine-containing nucleotide of the template. Once the polymerase has moved past a particular stretch of DNA, the DNA double helix re-forms (as in Figure 11.4*a*,*b*). Consequently, the RNA chain does not remain associated with its template as a DNA–RNA hybrid (except for about nine nucleotides just behind the site where the polymerase is operating). RNA polymerases are capable of incorporating from about 20 to 50 nucleotides into a growing RNA molecule per second, and many genes in a cell are transcribed simultaneously by a hundred or more polymerases. The frequency at which a gene is transcribed is tightly regulated and can vary dramatically depending on the given gene and the prevailing conditions. The electron micrograph of Figure 11.4*d* shows a molecule of phage DNA with a number of bound RNA polymerase molecules.

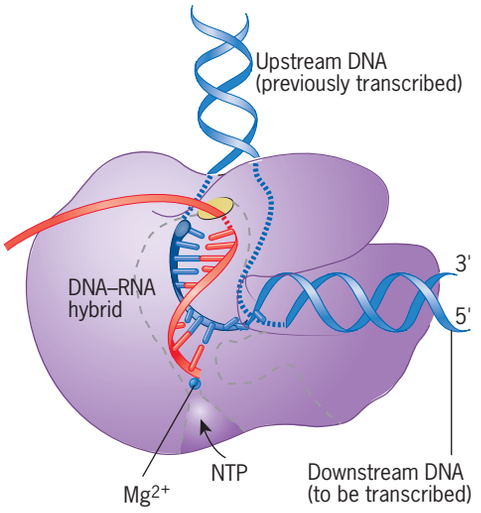
RNA polymerases are capable of forming prodigiously long RNAs. Consequently, the enzyme must remain attached to the DNA over long stretches of template (the enzyme is said to be *processive*). At the same time, the enzyme must be associated loosely enough that it can move from nucleotide to nucleotide of the template. It is difficult to study certain properties of RNA polymerases, such as processivity, using biochemical methodologies that tend to average out differences between individual protein molecules. Consequently, researchers have developed techniques to follow the activities of single RNA polymerase molecules similar to those used to study individual cytoskeletal motors. Two examples of such studies are depicted in **FIGURE 11.5**. In both of these examples, a single RNA polymerase is attached to the surface of a glass coverslip and allowed to transcribe a DNA molecule containing a fluorescent bead covalently linked to one of its ends. The movement of the fluorescent bead can be monitored under a fluorescence microscope.

In Figure 11.5*a*, the bead is free to move in the medium, and its range of movement is proportional to the length of the DNA between the polymerase and the bead. As the polymerase transcribes the template, the connecting DNA strand is elongated, and the movement of the bead is increased. This type of system allows investigators to study the rate of transcription of an individual polymerase and to determine if the polymerase transcribes the DNA in a steady or discontinuous movement. In Figure 11.5*b*, the bead at the end of the DNA molecule being transcribed is trapped by a focused laser beam (page 319). The minute force exerted by the laser trap can be varied, until it is just sufficient to stop the polymerase from continuing to transcribe the DNA. Measurements carried out on single RNA polymerase molecules in the act of transcription indicate that these enzymes can move over the template, one base (3.4 Å) at a time, with a force several times that of a myosin molecule (see Chapter 9 for discussion of myosin and other motor proteins).

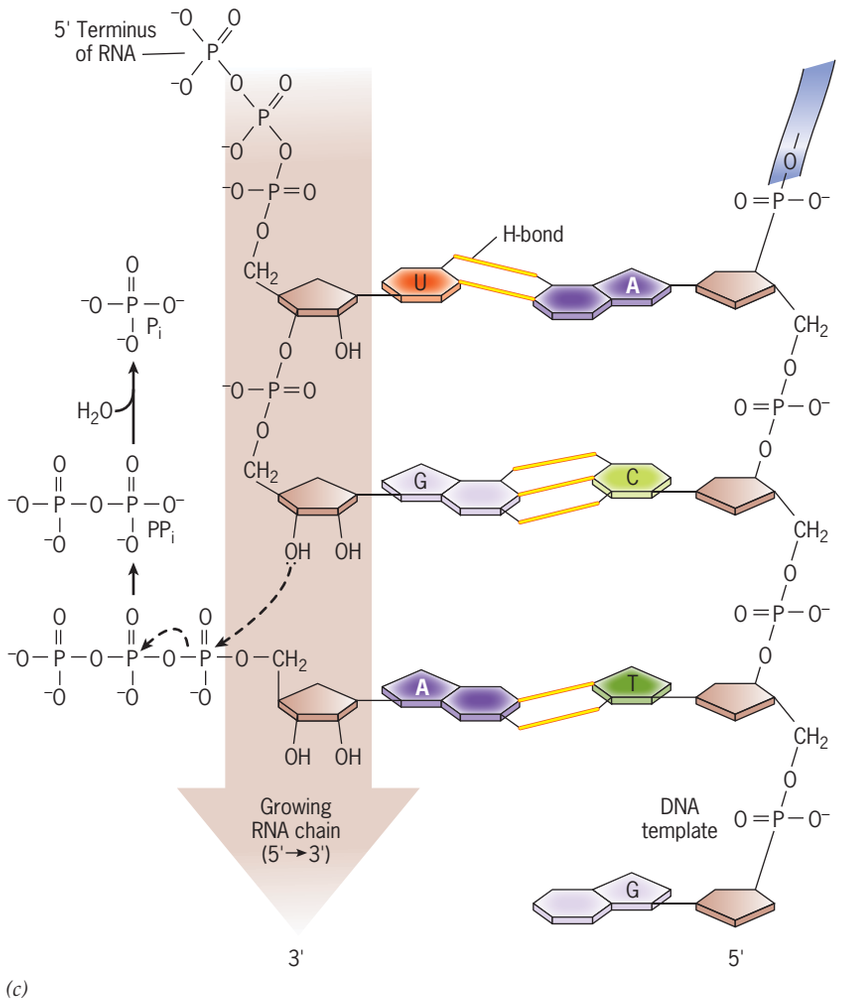
Even though polymerases are relatively powerful motors, these enzymes do not necessarily move in a steady, continuous fashion but may pause at certain locations along the template or even backtrack before resuming their forward progress. A number of elongation factors have been identified that enhance the enzyme's ability to traverse these various roadblocks. In some cases, as occurs following the rare incorporation of an incorrect nucleotide, a stalled polymerase



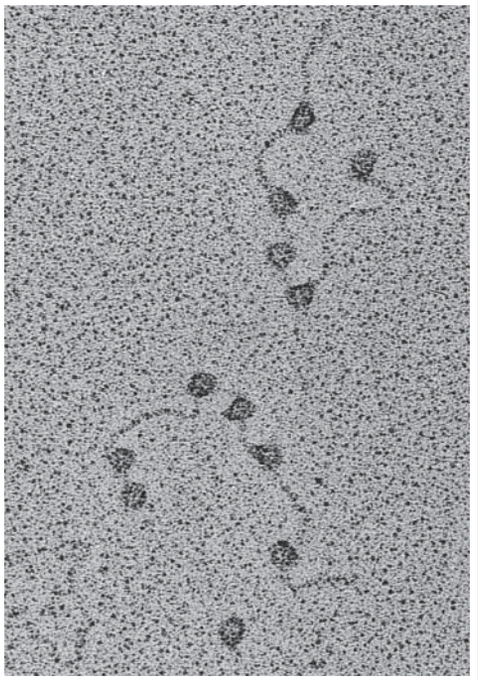
(a)



(b)



(c)



(d)

FIGURE 11.4 Chain elongation during transcription. (*a*) A schematic model of the elongation of a newly synthesized RNA molecule during transcription. The polymerase covers approximately 35 base pairs of DNA, the transcription bubble composed of single-stranded (melted) DNA contains about 15 base pairs, and the segment present in a DNA–RNA hybrid includes about nine base pairs. The enzyme generates overwound (positively supercoiled) DNA ahead of itself and underwound (negatively supercoiled) DNA behind itself (page 381). These conditions are relieved by topoisomerases (page 381). (*b*) Schematic drawing of an RNA polymerase in the act of transcription elongation. The downstream DNA lies in a groove within the polymerase, clamped by a pair of jaws formed by the two largest subunits of the enzyme. The DNA makes a sharp turn in the region of the active site, so that the upstream DNA extends upward in this drawing. The nascent RNA exits from the enzyme's active site through a separate channel. (*c*) Chain elongation results following an attack by the 3′ OH of the nucleotide at the end of the growing strand on the 5′ α-phosphate of the incoming nucleoside triphosphate. The pyrophosphate released is subsequently cleaved, which drives the reaction in the direction of polymerization. The geometry of base-pairing between the nucleotide of the template strand and the incoming nucleotide determines which of the four possible nucleoside triphosphates is incorporated into the growing RNA chain at each site. (*d*) Electron micrograph of several RNA polymerase molecules bound to a phage DNA template.

SOURCE: (*d*) Courtesy of Robley C. Williams.

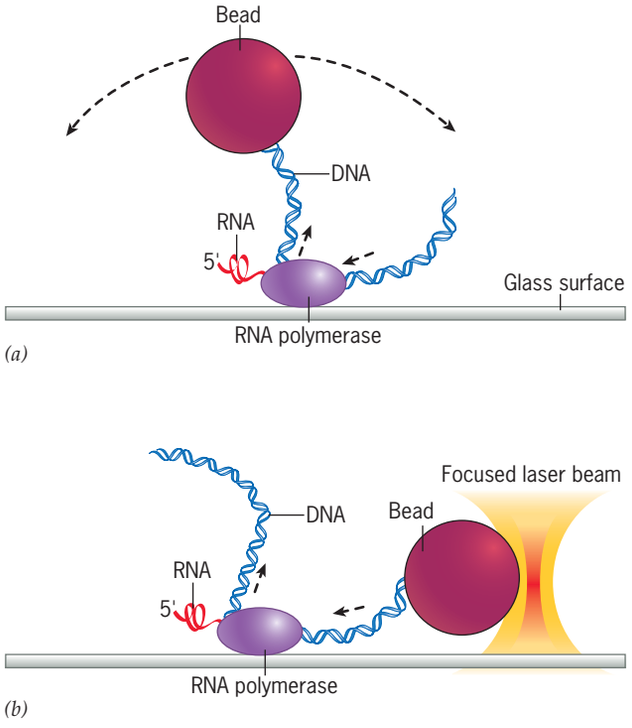


FIGURE 11.5 Examples of experimental techniques to follow the activities of single RNA polymerase molecules. (*a*) In this protocol, the RNA polymerase molecule is attached to the underlying coverslip and allowed to transcribe a DNA molecule containing a fluorescent bead at the upstream end. The arrows indicate the movement of the DNA through the polymerase. The rate of movement and progression of the polymerase can be followed by observing the position of the bead over time using a fluorescence microscope. (*b*) In this protocol, the attached polymerase is transcribing a DNA molecule with a bead at its downstream end. As in *a*, the arrows indicate direction of DNA movement. The bead is caught in an optical (laser) trap, which delivers a known force that can be varied by adjusting the laser beam. This type of apparatus can measure the forces generated by a transcribing polymerase.

SOURCE: Reprinted from J. Gelles and R. Landick, *Cell* 93:15, 1998, copyright 1998, with permission from Elsevier Science.)

must backtrack and then digest away the 3′ end of the newly synthesized transcript and resynthesize the missing portion before continuing its movement. The ability of RNA polymerase to recognize and remove a misincorporated nucleotide is referred to as **proofreading**. This corrective function is carried out by the same active site within the enzyme that is responsible for nucleotide incorporation.

REVIEW

1. What is the role of a promoter in gene expression? Where are the promoters for bacterial polymerases located?

11.3 An Overview of Transcription in Both Prokaryotic and Eukaryotic Cells

At this point, it is useful to examine the differences in the process of transcription between bacterial and eukaryotic cells.

Transcription in Bacteria

Bacteria, such as *E. coli*, contain a single type of RNA polymerase composed of five subunits that are tightly associated to form a *core enzyme*. If the core enzyme is purified from bacterial cells and added to a solution of bacterial DNA molecules and ribonucleoside triphosphates, the enzyme binds to the DNA and synthesizes RNA. The RNA molecules produced by a purified polymerase, however, are not the same as those found within cells because the core enzyme has attached to random sites in the DNA (**FIGURE 11.6a**), sites that it

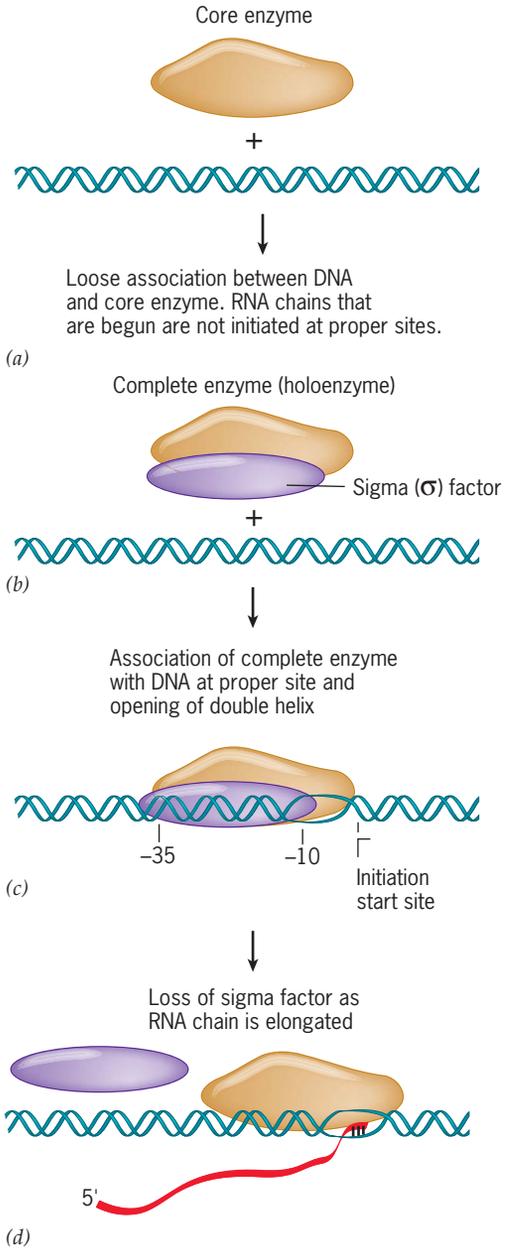


FIGURE 11.6 Schematic representation of the initiation of transcription in bacteria. (*a*) In the absence of the σ factor, the core enzyme does not interact with the DNA at specific initiation sites. (*b–d*) When the core enzyme is associated with the σ factor, the complete enzyme (or holoenzyme) is able to recognize and bind to the promoter regions of the DNA, separate the strands of the DNA double helix, and initiate transcription at the proper start sites (see Figure 11.7). In the traditional model shown here, the σ factor dissociates from the core enzyme, which is capable of transcription elongation. Several studies suggest that, in at least some cases, σ may remain with the polymerase.

would normally have ignored in vivo. If, however, a purified accessory polypeptide called *sigma factor* (σ) is added to the RNA polymerase before it attaches to DNA, transcription begins at selected locations (Figure 11.6*b-d*). Attachment of σ factor to the core enzyme increases the enzyme's affinity for promoter sites in DNA and decreases its affinity for DNA in general. As a result, the complete enzyme is thought to slide freely along the DNA until it recognizes and binds to a suitable promoter region.

X-ray crystallographic analysis of the bacterial RNA polymerase (see Figure 11.8) reveals a molecule shaped like a crab claw with a pair of mobile pincers (or jaws) enclosing a positively charged internal channel. As the σ factor interacts with the promoter, the jaws of the enzyme grip the downstream DNA duplex, which resides within the channel (as in Figure 11.4*b*). The enzyme then separates (or *melts*) the two DNA strands in the region surrounding the start site (Figure 11.6*c*). The complex of the polymerase, σ factor, and DNA with the strands separated is called the *Open Complex*. Strand separation makes the template strand accessible to the enzyme's active site, which resides at the back wall of the channel. Initiation of transcription appears to be a difficult undertaking because an RNA polymerase typically makes several unsuccessful attempts to assemble an RNA transcript. Once about 10 nucleotides have been successfully incorporated into a growing transcript, the enzyme undergoes a major change in conformation and is transformed into a *transcriptional elongation complex* that can move processively along the DNA. In the model shown in Figure 11.6*d*, the formation of an elongation complex is followed by release of the σ factor.

As noted above, promoters are the sites in DNA that bind RNA polymerase. Bacterial promoters are located in the region of a DNA strand just preceding the initiation site of RNA synthesis (**FIGURE 11.7**). The nucleotide at which transcription is initiated is denoted as +1 and the preceding nucleotide as −1. Those portions of the DNA preceding the initiation site (toward the 3′ end of the template strand) are said to be *upstream* from that site. Those portions of the DNA succeeding it (toward the 5′ end of the template strand) are said to be *downstream* from that site. Analysis of the DNA sequences just upstream from a large number of bacterial genes reveals that two short stretches of DNA are similar from one gene to another. One of these stretches is centered at approximately 35 bases upstream from the initiation site and typically occurs as the sequence TTGACA (Figure 11.7). This TTGACA sequence (known as the −35 element) is called a **consensus sequence**, which indicates that it is the most common version of a conserved sequence, but that some variation occurs from one gene to another. The second conserved sequence is found approximately 10 bases upstream from the initiation site and occurs at the consensus sequence TATAAT (Figure 11.7). This site in the promoter, named the *–10 element* after its position or the *Pribnow box* after its discoverer, is responsible for identifying the precise nucleotide at which transcription begins. As the sigma factor recognizes the Pribnow box, amino acid residues within the protein interact with each of the six nucleotides of the TATAAT sequence of the nontemplate strand. Two of these nucleotides are flipped out of the nucleotide stack and into the core of the protein. This action likely initiates melting of the adjoining region of promoter DNA and formation of the transcription bubble seen in Figures 11.6 and 11.7.

Bacterial cells possess a variety of different σ factors that recognize different versions of the promoter sequence. The σ^{70} is known as the "housekeeping" σ factor, because it initiates transcription of most genes. Alternative σ factors initiate transcription of a small number of specific genes that participate in a common response. For example, when *E. coli* cells are subjected to a sudden rise in temperature, a new σ factor is synthesized that recognizes a different promoter sequence and leads to the coordinated transcription of a battery of *heat-shock genes*. The products of these genes protect the proteins of the cell from thermal damage (See Experimental Pathways for Chapter 2).

Just as transcription is initiated at specific points in the chromosome, it also terminates when a specific nucleotide sequence is reached. In roughly half of the cases, a ring-shaped protein called *rho* is required for termination of bacterial transcription. Rho encircles the newly synthesized RNA and moves along the strand in a 5′→3′ direction to the polymerase, where it separates the RNA transcript from the DNA to which it is bound. In other cases, the polymerase stops transcription when it reaches a *terminator sequence*. Terminator sequences typically fold into a hairpin loop that causes the RNA polymerase to release the completed RNA chain without requiring additional factors.

Transcription and RNA Processing in Eukaryotic Cells

Eukaryotic cells have three distinct transcribing enzymes in their cell nuclei. Each of these enzymes is responsible for synthesizing a different group of RNAs (Table 11.1). Plants have two additional RNA polymerases that are not essential for life. No prokaryote has been found with multiple RNA polymerases, whereas the simplest eukaryotes (yeast) have the same three nuclear types that are present in mammalian cells. This difference in number of RNA polymerases is another sharp distinction between prokaryotic and eukaryotic cells.

FIGURE 11.8a shows the surface structures of RNA polymerases from each of the three domains of life: archaea, bacteria, and eukaryotes. Several features are apparent from a close examination of this illustration. It is evident that the archaeal and eukaryotic enzymes

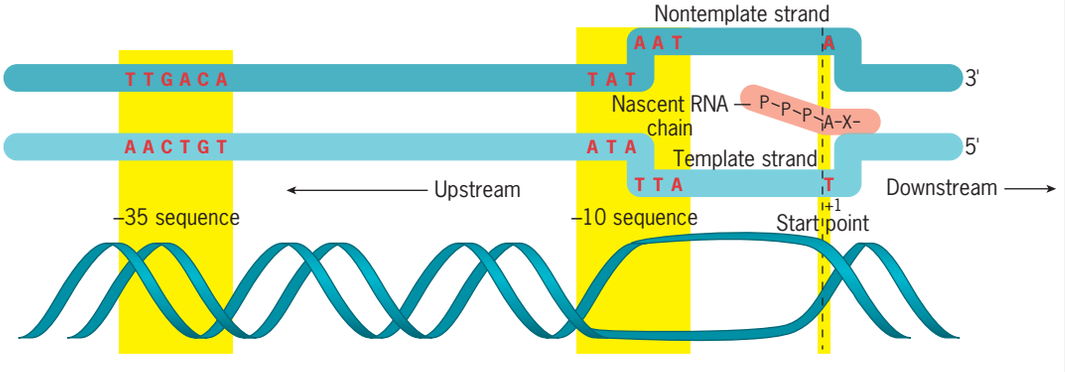


FIGURE 11.7 The basic elements of a promoter region in the DNA of the bacterium *E. coli*. The key regulatory sequences required for initiation of transcription are found in regions located at –35 and –10 base pairs from the site at which transcription is initiated. The initiation site marks the boundary between the + and – sides of the gene.

TABLE 11.1 Eukaryotic *Nuclear* RNA Polymerases

Enzyme	RNAs Synthesized
RNA polymerase I	larger rRNAs (28S, 18S, 5.8S)
RNA polymerase II	mRNAs, most small nuclear RNAs (snRNAs and snoRNAs), most microRNAs, and telomerase RNA
RNA polymerase III	other small RNAs, including tRNAs, 5S rRNA, and U6 snRNA
RNA Polymerase IV, V (Plants only)	siRNAs

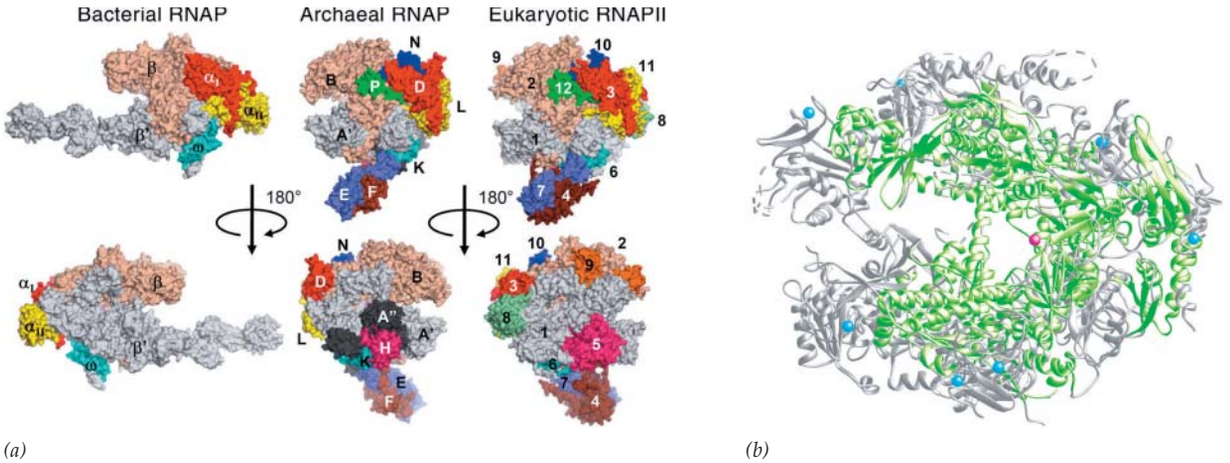
are more similar in structure to one another than are the bacterial and archaeal enzymes. This feature reflects the evolutionary relationship between archaea and eukaryotes that was discussed back in section 1.9. The subunits that make up each of the proteins depicted in Figure 11.8*a* are indicated by a different color, and it is immediately evident that RNA polymerases are multisubunit enzymes. Those subunits between the different enzymes that are homologous to one another (i.e., are derived from a common ancestral polypeptide) are shown in the same color. Thus, it is also evident from Figure 11.8*a* that the RNA polymerases from the three domains share a number of subunits. Although the yeast enzyme shown in Figure 11.8*a* has a total of 12 subunits, seven more than its bacterial counterpart, the fundamental core structure of the two enzymes is virtually identical. This evolutionary conservation in RNA polymerase structure is revealed in Figure 11.8*b*, which shows the homologous regions of the bacterial and eukaryotic enzymes at much higher resolution.

Our understanding of transcription in eukaryotes was greatly advanced with the 2001 publication of the X-ray crystallographic structure of yeast RNA polymerase II by Roger Kornberg and colleagues at Stanford University. As a result of these studies, and those of other laboratories in subsequent years, we now know a great deal about the mechanism of action of RNA polymerases as they move along the DNA, transcribing a complementary strand of RNA. A major distinction between transcription in prokaryotes and eukaryotes is the requirement in eukaryotes for a large variety of accessory proteins, or *transcription factors*. These proteins play a role in virtually every aspect of the transcription process, from the binding of the polymerase to the DNA template, to the initiation of transcription, to its elongation and termination. Although transcription factors are crucial for the operation of all three types of eukaryotic RNA polymerases, they will only be discussed in regard to the synthesis of mRNAs by RNA polymerase II (page 418).

All three major types of eukaryotic RNAs—mRNAs, rRNAs, and tRNAs—are derived from precursor RNA molecules that are considerably longer than the final RNA product. The initial precursor RNA is equivalent in length to the full length of the DNA transcribed and is called the **primary transcript**, or **pre-RNA**. The corresponding segment of DNA from which a primary transcript is transcribed is called a **transcription unit**. Primary transcripts do not exist within the cell as naked RNA but become associated with proteins even as they are synthesized. Primary transcripts typically have a fleeting existence, being *processed* into smaller, functional RNAs by a series of "cut-and-paste" reactions. RNA processing requires a variety of small RNAs (90 to 300 nucleotides long) and their associated proteins. In the following sections, we will examine the activities associated with the transcription and processing of each of the major eukaryotic RNAs.

REVIEW

1. Describe the steps during initiation of transcription in bacteria. What is the role of the σ factor? What is the nature of the reaction in which nucleotides are incorporated into a growing RNA strand? How is the specificity of nucleotide incorporation determined? What is the role of pyrophosphate hydrolysis?
2. How do the number of RNA polymerases distinguish prokaryotes and eukaryotes? What is the relationship between a pre-RNA and a mature RNA?



(*a*) (*b*)

FIGURE 11.8 A comparison of prokaryotic and eukaryotic RNA polymerase structure. (*a*) RNA polymerases from the three domains of life. Each subunit of an enzyme is denoted by a different color and labeled according to conventional nomenclature for that enzyme. Homologous subunits are depicted by the same color. It can be seen that the archaeal and eukaryotic polymerases are more similar in structure to one another than are the bacterial and eukaryotic enzymes. RNA polymerase II (shown here) is only one of three major eukaryotic nuclear RNA polymerases. (*b*) Ribbon diagram of the core structure of yeast RNA polymerase II. Regions of the bacterial polymerase that are structurally homologous to the yeast enzyme are shown in green. The large channel that grips the downstream DNA is evident. A divalent Mg^{2+} ion, situated at the end of the channel and within the active site, is seen as a red sphere.

SOURCE: (*a*) From Akira Hirata, Brianna J. Klein, and Katsuhiko S. Murakami, *Nature* 451:852, 2008, reprinted by permission from Macmillan Publishers Ltd.; (*b*) From Patrick Cramer, et al., *Science* 292:1874, 2001, Fig. 12b. © 2001, reprinted with permission from AAAS. Courtesy of Roger Kornberg, Stanford University School of Medicine.