majority of genes carry information to build protein molecules, and it is the RNA copies of such *protein-coding genes* that constitute the mRNA molecules of cells. The DNA molecules of small viruses contain only a few genes, whereas the single DNA molecule in each of the chromosomes of higher animals and plants may contain several thousand genes.

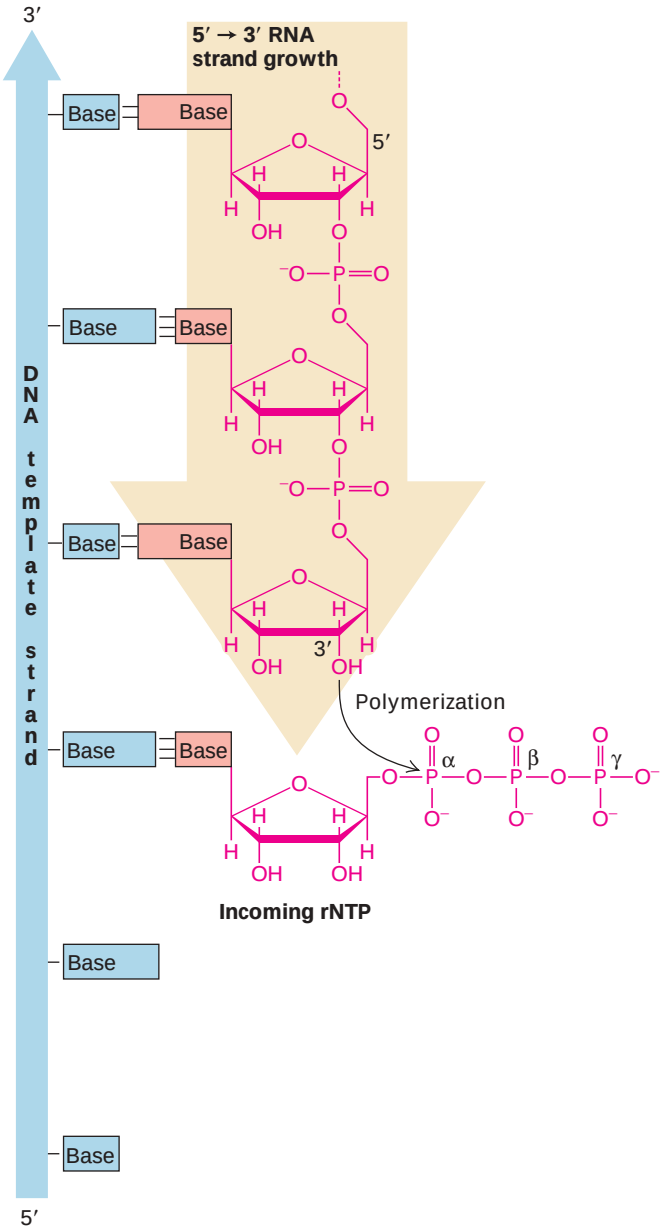
During synthesis of RNA, the four-base language of DNA containing A, G, C, and T is simply copied, or *transcribed*, into the four-base language of RNA, which is identical except that U replaces T. In contrast, during protein synthesis the four-base language of DNA and RNA is *translated* into the 20–amino acid language of proteins. In this section we focus on formation of functional mRNAs from protein-coding genes (see Figure 4-1, step 1). A similar process yields the precursors of rRNAs and tRNAs encoded by rRNA and tRNA genes; these precursors are then further modified to yield functional rRNAs and tRNAs (Chapter 12).

A Template DNA Strand Is Transcribed into a Complementary RNA Chain by RNA Polymerase

During transcription of DNA, one DNA strand acts as a *template*, determining the order in which ribonucleoside triphosphate (rNTP) monomers are polymerized to form a complementary RNA chain. Bases in the template DNA strand base-pair with complementary incoming rNTPs, which then are joined in a polymerization reaction catalyzed by **RNA polymerase.** Polymerization involves a nucleophilic attack by the 3′ oxygen in the growing RNA chain on the α phosphate of the next nucleotide precursor to be added, resulting in formation of a phosphodiester bond and release of pyrophosphate (PP_i). As a consequence of this mechanism, RNA molecules are always synthesized in the 5′→3′ direction (Figure 4-9).

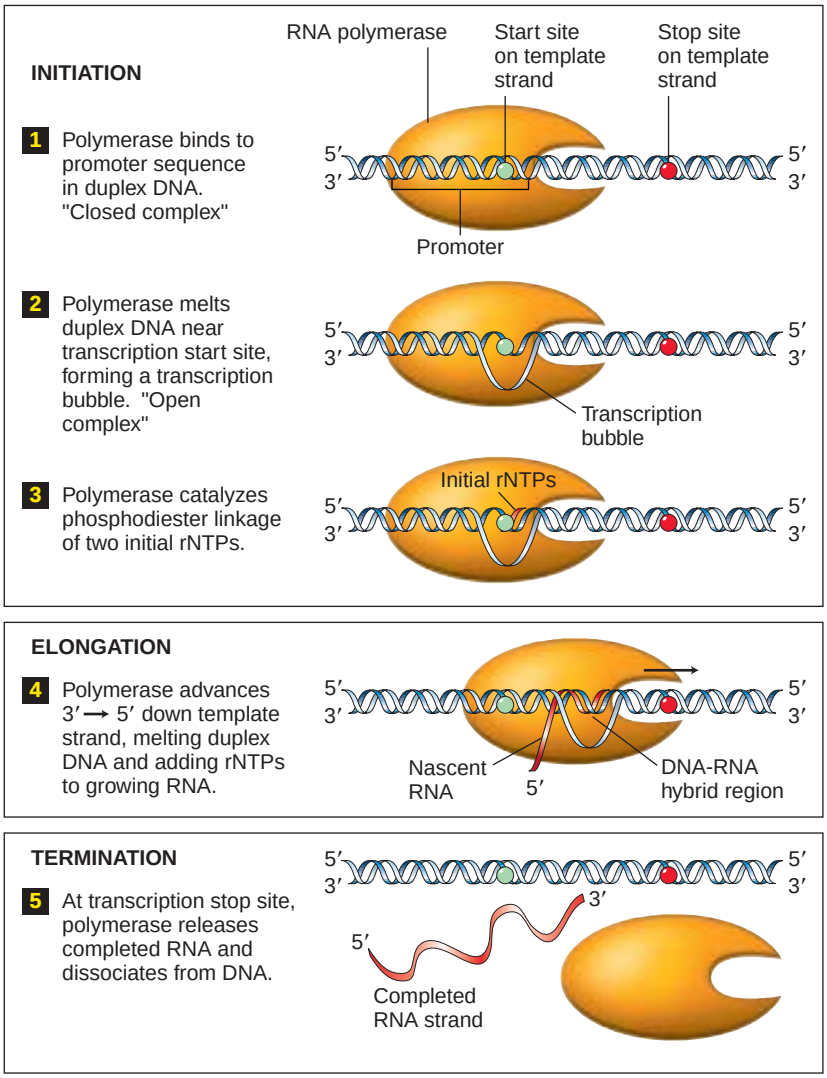
The energetics of the polymerization reaction strongly favors addition of ribonucleotides to the growing RNA chain because the high-energy bond between the α and β phosphate of rNTP monomers is replaced by the lower-energy phosphodiester bond between nucleotides. The equilibrium for the reaction is driven further toward chain elongation by pyrophosphatase, an enzyme that catalyzes cleavage of the released PP_i into two molecules of inorganic phosphate. Like the two strands in DNA, the template DNA strand and the growing RNA strand that is base-paired to it have opposite 5′→3′ directionality.

By convention, the site at which RNA polymerase begins transcription is numbered +1. **Downstream** denotes the direction in which a template DNA strand is transcribed (or mRNA translated); thus a downstream sequence is toward the 3′ end relative to the start site, considering the DNA strand with the same polarity as the transcribed RNA. **Upstream** denotes the opposite direction. Nucleotide positions in the DNA sequence downstream from a start site are indicated by a positive (+) sign; those upstream, by a negative (−) sign.



▲ FIGURE 4-9 **Polymerization of ribonucleotides by RNA polymerase during transcription.** The ribonucleotide to be added at the 3′ end of a growing RNA strand is specified by base pairing between the next base in the template DNA strand and the complementary incoming ribonucleoside triphosphate (rNTP). A phosphodiester bond is formed when RNA polymerase catalyzes a reaction between the 3′ O of the growing strand and the α phosphate of a correctly base-paired rNTP. RNA strands always are synthesized in the 5′→3′ direction and are opposite in polarity to their template DNA strands.

Stages in Transcription To carry out transcription, RNA polymerase performs several distinct functions, as depicted in Figure 4-10. During transcription *initiation*, RNA polymerase recognizes and binds to a specific site, called a **promoter,** in double-stranded DNA (step 1). Nuclear RNA



◀ FIGURE 4-10 **Three stages in transcription.** During initiation of transcription, RNA polymerase forms a transcription bubble and begins polymerization of ribonucleotides (rNTPs) at the start site, which is located within the promoter region. Once a DNA region has been transcribed, the separated strands reassociate into a double helix, displacing the nascent RNA except at its 3′ end. The 5′ end of the RNA strand exits the RNA polymerase through a channel in the enzyme. Termination occurs when the polymerase encounters a specific termination sequence (stop site). See the text for details.

MEDIA CONNECTIONS
Focus Animation: Basic Transcriptional Mechanism

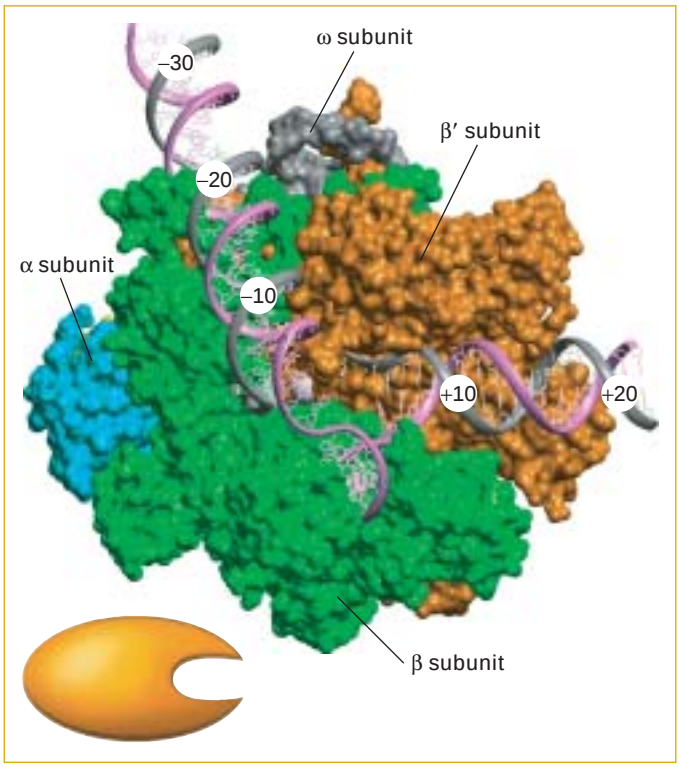
polymerases require various protein factors, called general **transcription factors,** to help them locate promoters and initiate transcription. After binding to a promoter, RNA polymerase melts the DNA strands in order to make the bases in the template strand available for base pairing with the bases of the ribonucleoside triphosphates that it will polymerize together. Cellular RNA polymerases melt approximately 14 base pairs of DNA around the transcription start site, which is located on the template strand within the promoter region (step 2). Transcription initiation is considered complete when the first two ribonucleotides of an RNA chain are linked by a phosphodiester bond (step 3).

After several ribonucleotides have been polymerized, RNA polymerase dissociates from the promoter DNA and general transcription factors. During the stage of *strand elongation*, RNA polymerase moves along the template DNA one base at a time, opening the double-stranded DNA in front of its direction of movement and hybridizing the strands behind it (Figure 4-10, step 4). One ribonucleotide at a time is added to the 3′ end of the growing (*nascent*) RNA chain during strand elongation by the polymerase. The enzyme maintains a melted region of approximately 14 base pairs, called the *transcription bubble*. Approximately eight nucleotides at the 3′ end of the growing RNA strand remain base-paired to the template DNA strand in the transcription bubble. The elongation complex, comprising RNA polymerase, template DNA, and the growing (nascent) RNA strand, is extraordinarily stable. For example, RNA polymerase transcribes the longest known mammalian genes, containing $\approx 2 \times 10^6$ base pairs, without dissociating from the DNA template or releasing the nascent RNA. Since RNA synthesis occurs at a rate of about 1000 nucleotides per minute at 37 °C, the elongation complex must remain intact for more than 24 hours to assure continuous RNA synthesis.

During transcription *termination*, the final stage in RNA synthesis, the completed RNA molecule, or **primary transcript,**

is released from the RNA polymerase and the polymerase dissociates from the template DNA (Figure 4-10, step 5). Specific sequences in the template DNA signal the bound RNA polymerase to terminate transcription. Once released, an RNA polymerase is free to transcribe the same gene again or another gene.

Structure of RNA Polymerases The RNA polymerases of bacteria, archaea, and eukaryotic cells are fundamentally similar in structure and function. Bacterial RNA polymerases are composed of two related large subunits (β′ and β), two copies of a smaller subunit (α), and one copy of a fifth subunit (ω) that is not essential for transcription or cell viability but stabilizes the enzyme and assists in the assembly of its subunits. Archaeal and eukaryotic RNA polymerases have several additional small subunits associated with this core complex, which we describe in Chapter 11. Schematic diagrams of the transcription process generally show RNA polymerase bound to an unbent DNA molecule, as in Figure 4-10. However, according to a current model of the interaction between bacterial RNA polymerase and promoter DNA, the DNA bends sharply following its entry into the enzyme (Figure 4-11).



▲ FIGURE 4-11 **Current model of bacterial RNA polymerase bound to a promoter.** This structure corresponds to the polymerase molecule as schematically shown in step 2 of Figure 4-10. The β′ subunit is in orange; β is in green. Part of one of the two α subunits can be seen in light blue; the ω subunit is in gray. The DNA template and nontemplate strands are shown, respectively, as gray and pink ribbons. A Mg^{2+} ion at the active center is shown as a gray sphere. Numbers indicate positions in the DNA sequence relative to the transcription start site, with positive (+) numbers in the direction of transcription and negative (−) numbers in the opposite direction. [Courtesy of R. H. Ebright, Waksman Institute.]

Organization of Genes Differs in Prokaryotic and Eukaryotic DNA

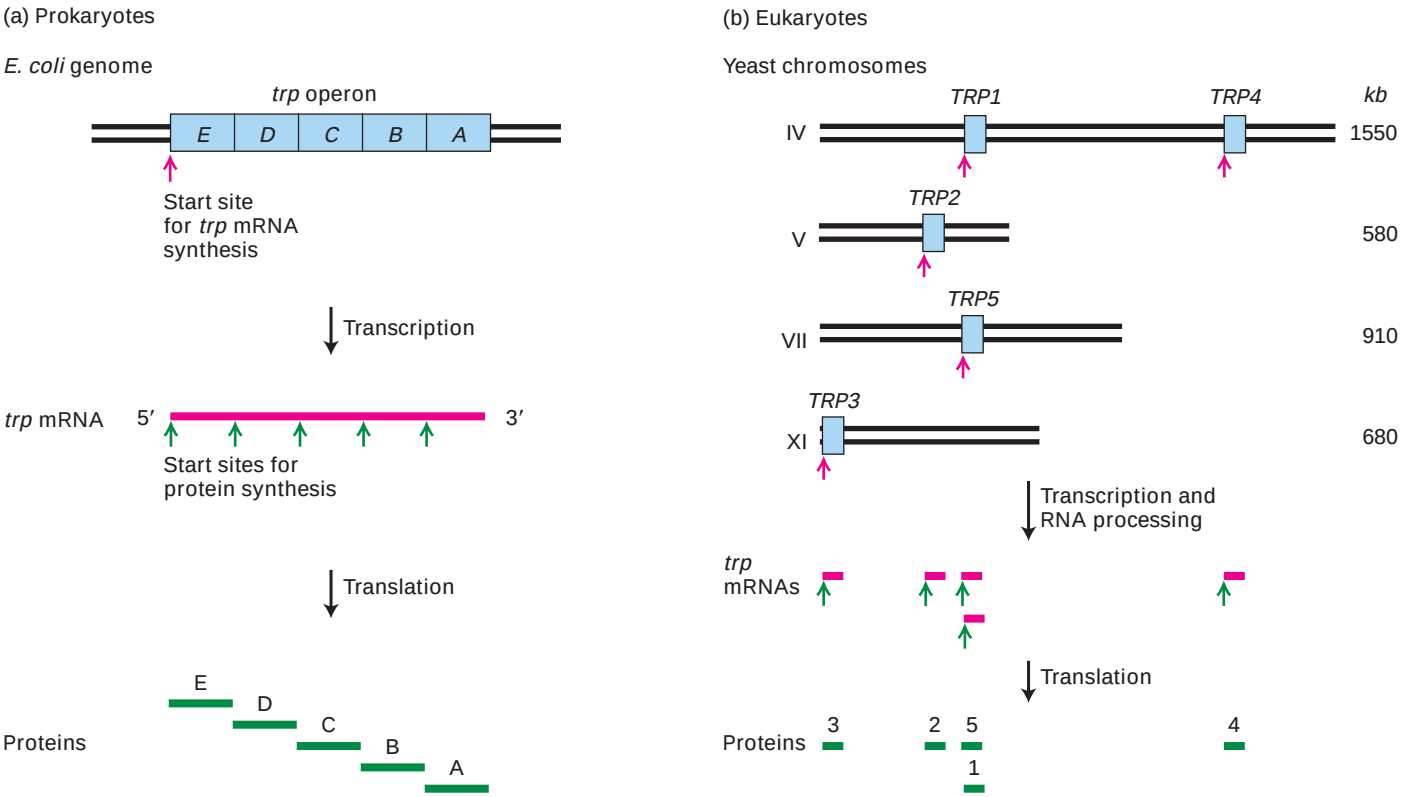
Having outlined the process of transcription, we now briefly consider the large-scale arrangement of information in DNA and how this arrangement dictates the requirements for RNA synthesis so that information transfer goes smoothly. In recent years, sequencing of the entire **genomes** from several organisms has revealed not only large variations in the number of protein-coding genes but also differences in their organization in prokaryotes and eukaryotes.

The most common arrangement of protein-coding genes in all prokaryotes has a powerful and appealing logic: genes devoted to a single metabolic goal, say, the synthesis of the amino acid tryptophan, are most often found in a contiguous array in the DNA. Such an arrangement of genes in a functional group is called an **operon,** because it operates as a unit from a single promoter. Transcription of an operon produces a continuous strand of mRNA that carries the message for a related series of proteins (Figure 4-12a). Each section of the mRNA represents the unit (or gene) that encodes one of the proteins in the series. In prokaryotic DNA the genes are closely packed with very few noncoding gaps, and the DNA is transcribed directly into colinear mRNA, which then is translated into protein.

This economic clustering of genes devoted to a single metabolic function does not occur in eukaryotes, even simple ones like yeasts, which can be metabolically similar to bacteria. Rather, eukaryotic genes devoted to a single pathway are most often physically separated in the DNA; indeed such genes usually are located on different chromosomes. Each gene is transcribed from its own promoter, producing one mRNA, which generally is translated to yield a single polypeptide (Figure 4-12b).

When researchers first compared the nucleotide sequences of eukaryotic mRNAs from multicellular organisms with the DNA sequences encoding them, they were surprised to find that the uninterrupted protein-coding sequence of a given mRNA was broken up (discontinuous) in its corresponding section of DNA. They concluded that the eukaryotic gene existed in pieces of coding sequence, the **exons,** separated by non-protein-coding segments, the **introns.** This astonishing finding implied that the long initial primary transcript—the RNA copy of the entire transcribed DNA sequence—had to be clipped apart to remove the introns and then carefully stitched back together to produce many eukaryotic mRNAs.

Although introns are common in multicellular eukaryotes, they are extremely rare in bacteria and archaea and



▲ FIGURE 4-12 **Comparison of gene organization, transcription, and translation in prokaryotes and eukaryotes.** (a) The tryptophan (*trp*) operon is a continuous segment of the *E. coli* chromosome, containing five genes (blue) that encode the enzymes necessary for the stepwise synthesis of tryptophan. The entire operon is transcribed from one promoter into one long continuous *trp* mRNA (red). Translation of this mRNA begins at five different start sites, yielding five proteins (green). The order of the genes in the bacterial genome parallels the sequential function of the encoded proteins in the tryptophan pathway. (b) The five genes encoding the enzymes required for tryptophan synthesis in yeast (*Saccharomyces cerevisiae*) are carried on four different chromosomes. Each gene is transcribed from its own promoter to yield a primary transcript that is processed into a functional mRNA encoding a single protein. The lengths of the yeast chromosomes are given in kilobases (10^3 bases).

uncommon in many unicellular eukaryotes such as baker's yeast. However, introns are present in the DNA of viruses that infect eukaryotic cells. Indeed, the presence of introns was first discovered in such viruses, whose DNA is transcribed by host-cell enzymes.

Eukaryotic Precursor mRNAs Are Processed to Form Functional mRNAs

In prokaryotic cells, which have no nuclei, translation of an mRNA into protein can begin from the 5′ end of the mRNA even while the 3′ end is still being synthesized by RNA polymerase. In other words, transcription and translation can occur concurrently in prokaryotes. In eukaryotic cells, however, not only is the nucleus separated from the cytoplasm where translation occurs, but also the primary transcripts of protein-coding genes are precursor mRNAs (**pre-mRNAs**) that must undergo several modifications, collectively termed *RNA processing*, to yield a functional mRNA (see Figure 4-1, step 2). This mRNA then must be exported to the cytoplasm before it can be translated into protein. Thus transcription and translation cannot occur concurrently in eukaryotic cells.

All eukaryotic pre-mRNAs initially are modified at the two ends, and these modifications are retained in mRNAs. As the 5′ end of a nascent RNA chain emerges from the surface of RNA polymerase II, it is immediately acted on by several enzymes that together synthesize the *5′ cap*, a 7-methylguanylate that is connected to the terminal nucleotide of the RNA by an unusual 5′,5′ triphosphate linkage (Figure 4-13). The cap protects an mRNA from enzymatic degradation and assists in its export to the cytoplasm. The cap also is bound by a protein factor required to begin translation in the cytoplasm.

Processing at the 3′ end of a pre-mRNA involves cleavage by an endonuclease to yield a free 3′-hydroxyl group to which a string of adenylic acid residues is added one at a time by an enzyme called *poly(A) polymerase*. The resulting *poly(A) tail* contains 100–250 bases, being shorter in yeasts and invertebrates than in vertebrates. Poly(A) polymerase is