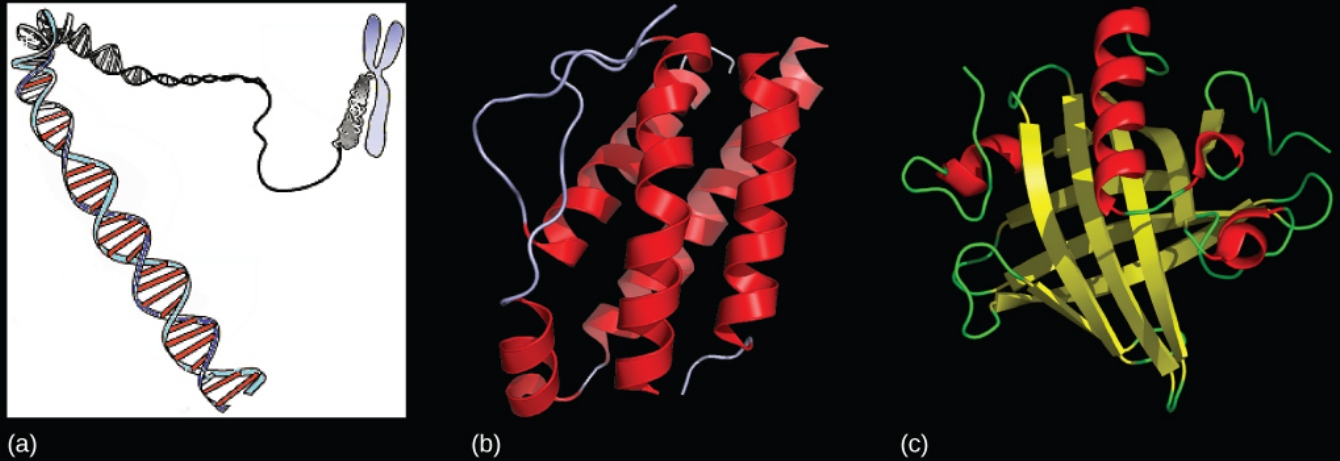# CHAPTER 15
Genes and Proteins

Figure 15.1 Genes, which are carried on (a) chromosomes, are linearly organized instructions for making the RNA and protein molecules that are necessary for all of the processes of life. The (b) interleukin-2 protein and (c) alpha-2u-globulin protein are just two examples of the array of different molecular structures that are encoded by genes. (credit "chromosome: National Human Genome Research Institute; credit "interleukin-2": Ramin Herati/Created from PDB 1M47 and rendered with Pymol; credit "alpha-2u-globulin": Darren Logan/rendered with AISMIG)

INTRODUCTION Since the rediscovery of Mendel's work in 1900, the definition of the gene has progressed from an abstract unit of heredity to a tangible molecular entity capable of replication, expression, and mutation (Figure 15.1). Genes are composed of DNA and are linearly arranged on chromosomes. Genes specify the sequences of amino acids, which are the building blocks of proteins. In turn, proteins are responsible for orchestrating nearly every function of the cell. Both genes and the proteins they encode are absolutely essential to life as we know it.

15.1 The Genetic Code

By the end of this section, you will be able to do the following:

- Explain the "central dogma" of DNA-protein synthesis
- Describe the genetic code and how the nucleotide sequence prescribes the amino acid and the protein sequence

The cellular process of transcription generates messenger RNA (mRNA), a mobile molecular copy of one or

more genes with an alphabet of A, C, G, and uracil (U). Translation of the mRNA template on ribosomes converts nucleotide-based genetic information into a protein product. That is the central dogma of DNA-protein synthesis. Protein sequences consist of 20 commonly occurring amino acids; therefore, it can be said that the protein alphabet consists of 20 "letters" (Figure 15.2). Different amino acids have different chemistries (such as acidic versus basic, or polar and nonpolar) and different structural constraints. Variation in amino acid sequence is responsible for the enormous variation in protein structure and function.
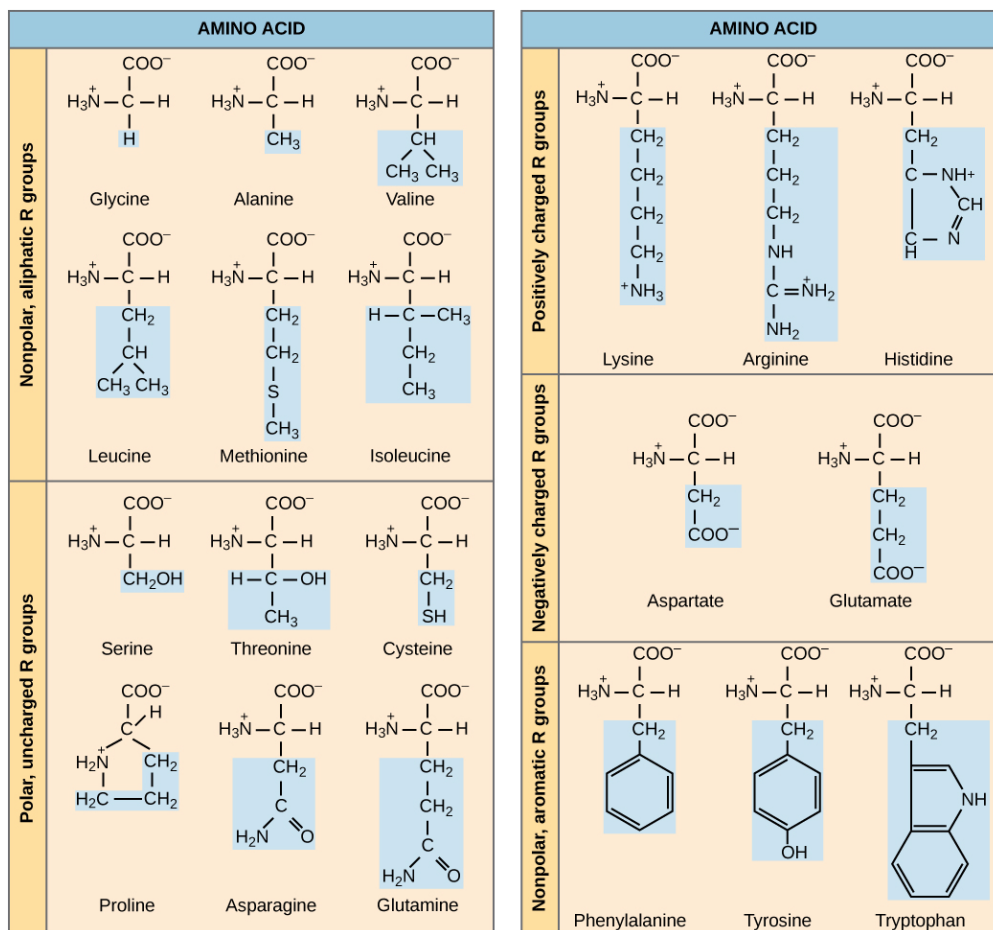
Figure 15.2 Structures of the 20 amino acids found in proteins are shown. Each amino acid is composed of an amino group (NH_3^+), a carboxyl group (COO⁻), and a side chain (blue). The side chain may be nonpolar, polar, or charged, as well as large or small. It is the variety of amino acid side chains that gives rise to the incredible variation of protein structure and function.

The Central Dogma: DNA Encodes RNA; RNA Encodes Protein

The flow of genetic information in cells from DNA to mRNA to protein is described by the **central dogma** (Figure 15.3), which states that genes specify the sequence of mRNAs, which in turn specify the sequence of amino acids making up all proteins. The decoding of one molecule to another is performed by specific proteins and RNAs. Because the information stored in DNA is so central to cellular function, it makes intuitive sense that the cell would make mRNA copies of this information for protein synthesis, while keeping the DNA itself intact and protected. The copying of DNA to RNA is relatively straightforward, with one nucleotide being added to the mRNA strand for every nucleotide read in the DNA strand. The translation to protein is a bit more complex because three mRNA nucleotides correspond to one amino acid in the polypeptide sequence. However, the translation to protein is still systematic and **colinear**, such that nucleotides 1 to 3 correspond to amino acid 1, nucleotides 4 to 6 correspond to amino acid 2, and so on.

Transcription

RNA processing

Translation
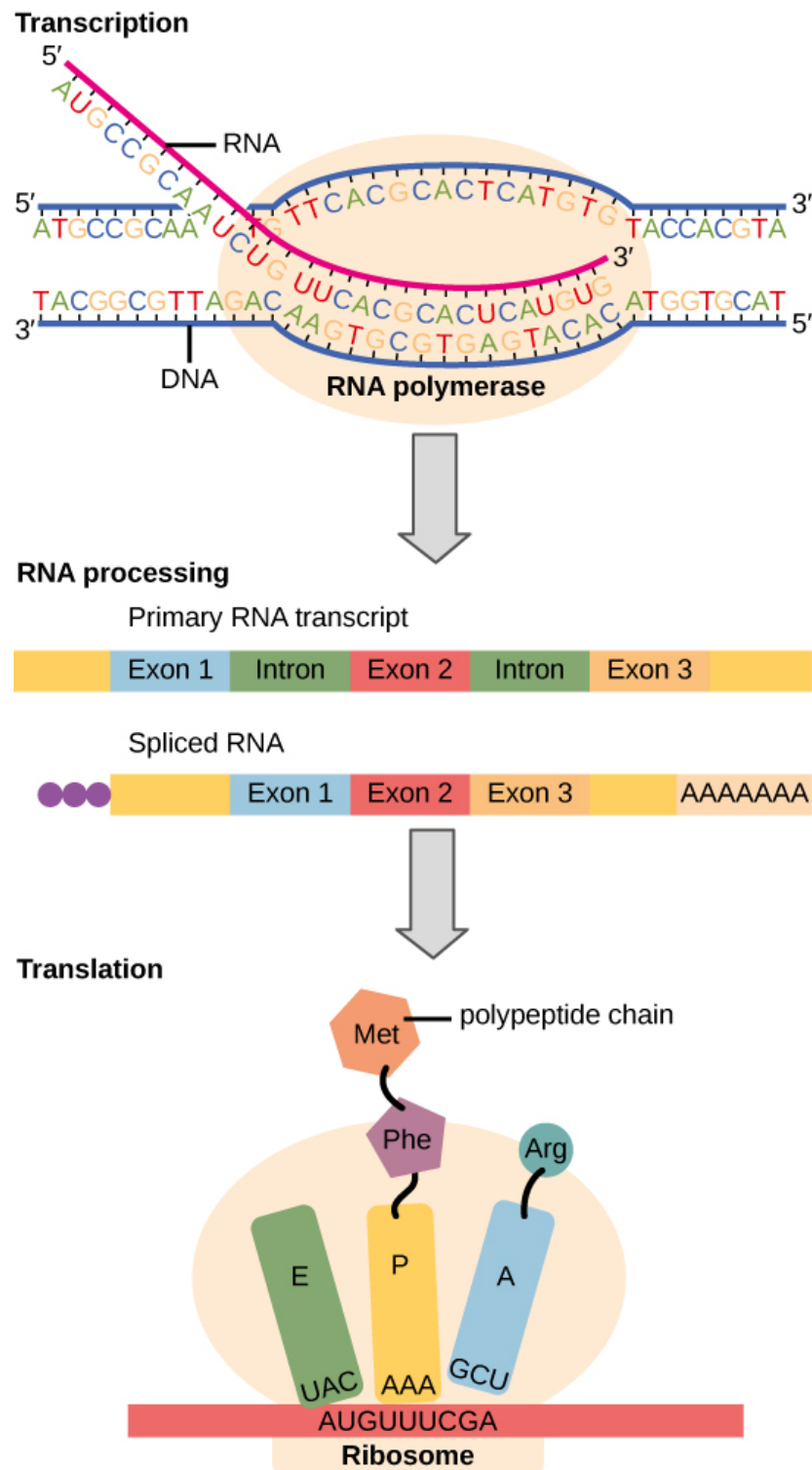
Figure 15.3 Instructions on DNA are transcribed onto messenger RNA. Ribosomes are able to read the genetic information inscribed on a strand of messenger RNA and use this information to string amino acids together into a protein.

The Genetic Code Is Degenerate and Universal

Each amino acid is defined by a three-nucleotide sequence called the triplet codon. Given the different numbers of "letters" in the mRNA and protein "alphabets," scientists theorized that single amino acids must be represented by combinations of nucleotides. Nucleotide doublets would not be sufficient to specify every amino acid because there are only 16 possible two-nucleotide combinations (4^2). In contrast, there are 64 possible nucleotide triplets (4^3), which is far more than the number of

amino acids. Scientists theorized that amino acids were encoded by nucleotide triplets and that the genetic code was **"degenerate."** In other words, a given amino acid could be encoded by more than one nucleotide triplet. This was later confirmed experimentally: Francis Crick and Sydney Brenner used the chemical mutagen proflavin to insert one, two, or three nucleotides into the gene of a virus. When one or two nucleotides were inserted, the normal proteins were not produced. When three nucleotides were inserted, the protein was synthesized and functional. This demonstrated that the amino acids must be specified by groups of three nucleotides. These nucleotide triplets are called **codons**. The insertion of one or two nucleotides completely changed the triplet **reading frame**, thereby altering the message for every subsequent amino acid (Figure 15.5). Though insertion of three nucleotides caused an extra amino acid to be inserted during translation, the integrity of the rest of the protein was maintained.

Scientists painstakingly solved the genetic code by translating synthetic mRNAs in vitro and sequencing the proteins they specified (Figure 15.4).

Figure 15.4 This figure shows the genetic code for translating each nucleotide triplet in mRNA into an amino acid or a termination signal in a protein. (credit: modification of work by NIH)

In addition to codons that instruct the addition of a specific amino acid to a polypeptide chain, three of the 64 codons terminate protein synthesis and release the polypeptide from the translation machinery. These triplets are called **nonsense codons**, or *stop codons*. Another codon, AUG, also has a special function. In addition to specifying the amino acid methionine, it also serves as the start codon to initiate translation. The **reading frame** for translation is set by the AUG start codon near the 5' end of the mRNA. Following the start codon, the mRNA is read in groups of three until a stop codon is encountered.

The arrangement of the coding table reveals the structure of the code. There are sixteen "blocks" of codons, each specified by the first and second nucleotides of the codons within the block, e.g., the "AC*" block that corresponds to the amino acid threonine (Thr). Some blocks are divided into a pyrimidine half, in which the codon ends with U or C, and a purine half, in which the codon ends with A or G. Some amino acids get a whole block of four codons, like alanine (Ala), threonine (Thr) and proline (Pro). Some get the pyrimidine half of their block, like histidine (His) and asparagine (Asn). Others get the purine half of their block, like glutamate (Glu) and lysine (Lys). Note that some amino acids get a block and a half-block for a total of six codons.

The specification of a single amino acid by multiple similar codons is called "degeneracy." Degeneracy is believed to be a cellular mechanism to reduce the negative impact of random mutations. Codons that specify the same amino acid typically only differ by one nucleotide. In addition, amino acids with chemically similar side chains are encoded by similar codons. For example, aspartate (Asp) and glutamate (Glu), which occupy the GA* block, are both negatively charged. This nuance of the genetic code ensures that a single-nucleotide substitution mutation might specify the same amino acid but have no effect or specify a similar amino acid, preventing the protein from being rendered completely nonfunctional.

The genetic code is nearly universal. With a few minor exceptions, virtually all species use the same genetic code for protein synthesis. Conservation of codons means that a purified mRNA encoding the globin protein in horses could be transferred to a tulip cell, and the tulip would synthesize horse globin. That there is only one genetic code is powerful evidence that all of life on Earth shares a common origin, especially considering that there are about 10^{84} possible combinations of 20 amino acids and 64 triplet codons.

🔗 LINK TO LEARNING

Transcribe a gene and translate it to protein using complementary pairing and the genetic code at this site (http://openstax.org/l/create_protein) .
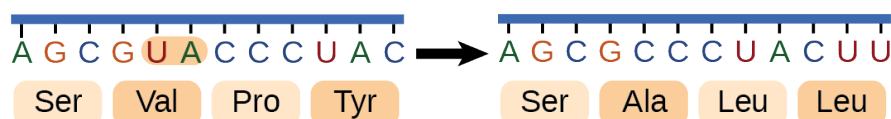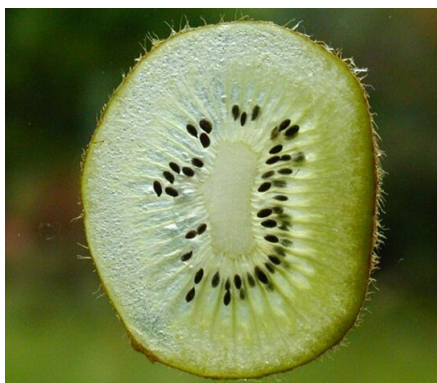

Figure 15.5 The deletion of two nucleotides shifts the reading frame of an mRNA and changes the entire protein message, creating a nonfunctional protein or terminating protein synthesis altogether.

📋 SCIENTIFIC METHOD CONNECTION

Which Has More DNA: A Kiwi or a Strawberry?

Figure 15.6 Do you think that a kiwi or a strawberry has more DNA per fruit? (credit "kiwi": "Kelbv"/Flickr; credit: "strawberry": Alisdair McDiarmid)

Question: Would a kiwi and strawberry that are approximately the same size (Figure 15.6) also have approximately the same amount of DNA?

Background: Genes are carried on chromosomes and are made of DNA. All mammals are diploid, meaning they have two copies of each chromosome. However, not all plants are diploid. The common strawberry is octoploid ($8n$) and the cultivated kiwi is hexaploid ($6n$). Research the total number of chromosomes in the cells of each of these fruits and think about how this might correspond to the amount of DNA in these fruits' cell nuclei. What other factors might contribute to the total amount of DNA in a single fruit? Read about the technique of DNA isolation to understand how each step in the isolation protocol helps liberate and precipitate DNA.

Hypothesis: Hypothesize whether you would be able to detect a difference in DNA quantity from similarly sized strawberries and kiwis. Which fruit do you think would yield more DNA?

Test your hypothesis: Isolate the DNA from a strawberry and a kiwi that are similarly sized. Perform the experiment in at least triplicate for each fruit

1.   Prepare a bottle of DNA extraction buffer from 900 mL water, 50 mL dish detergent, and two teaspoons of table salt. Mix by

inversion (cap it and turn it upside down a few times).

2. Grind a strawberry and a kiwi by hand in a plastic bag, or using a mortar and pestle, or with a metal bowl and the end of a blunt instrument. Grind for at least two minutes per fruit.

3. Add 10 mL of the DNA extraction buffer to each fruit, and mix well for at least one minute.

4. Remove cellular debris by filtering each fruit mixture through cheesecloth or porous cloth and into a funnel placed in a test tube or an appropriate container.

5. Pour ice-cold ethanol or isopropanol (rubbing alcohol) into the test tube. You should observe white, precipitated DNA.

6. Gather the DNA from each fruit by winding it around separate glass rods.

Record your observations: Because you are not quantitatively measuring DNA volume, you can record for each trial whether the two fruits produced the same or different amounts of DNA as observed by eye. If one or the other fruit produced noticeably more DNA, record this as well. Determine whether your observations are consistent with several pieces of each fruit.

Analyze your data: Did you notice an obvious difference in the amount of DNA produced by each fruit? Were your results reproducible?

Draw a conclusion: Given what you know about the number of chromosomes in each fruit, can you conclude that chromosome number necessarily correlates to DNA amount? Can you identify any drawbacks to this procedure? If you had access to a laboratory, how could you standardize your comparison and make it more quantitative?

15.2 Prokaryotic Transcription

By the end of this section, you will be able to do the following:

- List the different steps in prokaryotic transcription
- Discuss the role of promoters in prokaryotic transcription
- Describe how and when transcription is terminated

The prokaryotes, which include Bacteria and Archaea, are mostly single-celled organisms that, by definition, lack membrane-bound nuclei and other organelles. A bacterial chromosome is a closed circle that, unlike eukaryotic chromosomes, is not organized around histone proteins. The central region of the cell in which prokaryotic DNA resides is called the nucleoid region. In addition, prokaryotes often have abundant **plasmids**, which are shorter, circular DNA molecules that may only contain one or a few genes. Plasmids can be transferred independently of the bacterial chromosome during cell division and often carry traits such as those involved with antibiotic resistance.

Transcription in prokaryotes (and in eukaryotes) requires the DNA double helix to partially unwind in the region of mRNA synthesis. The region of unwinding is called a **transcription bubble.** Transcription always proceeds from the same DNA strand for each gene, which is called the **template strand**. The mRNA product is complementary to the template strand and is almost identical to the other DNA strand, called the **nontemplate strand**, or the coding strand. The only nucleotide difference is that in mRNA, all of the T nucleotides are replaced with U nucleotides (Figure 15.7). In an RNA double helix, A can bind U via two hydrogen bonds, just as in A–T pairing in a DNA double helix.
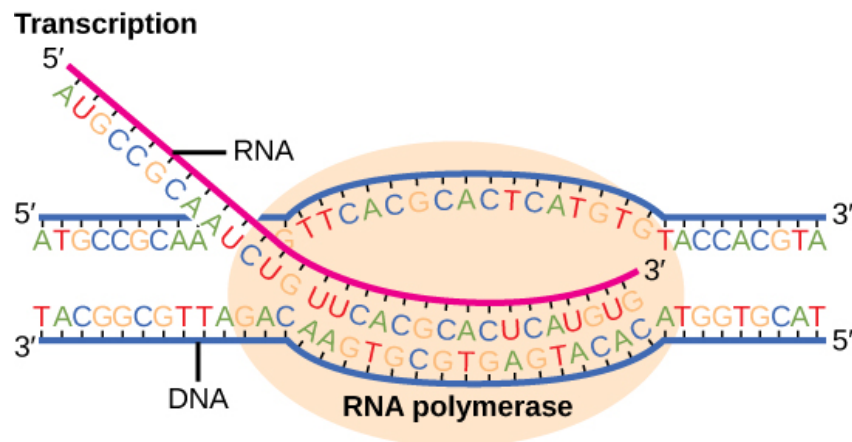
Figure 15.7 Messenger RNA is a copy of protein-coding information in the coding strand of DNA, with the substitution of U in the RNA for T

in the coding sequence. However, new RNA nucleotides base pair with the nucleotides of the template strand. RNA is synthesized in its 5'-3' direction, using the enzyme RNA polymerase. As the template is read, the DNA unwinds ahead of the polymerase and then rewinds behind it.

The nucleotide pair in the DNA double helix that corresponds to the site from which the first 5' mRNA nucleotide is transcribed is called the +1 site, or the **initiation site**. Nucleotides preceding the initiation site are denoted with a "-" and are designated *upstream nucleotides*. Conversely, nucleotides following the initiation site are denoted with "+" numbering and are called *downstream nucleotides*.

Initiation of Transcription in Prokaryotes

Prokaryotes do not have membrane-enclosed nuclei. Therefore, the processes of transcription, translation, and mRNA degradation can all occur simultaneously. The intracellular level of a bacterial protein can quickly be amplified by multiple transcription and translation events that occur concurrently on the same DNA template. Prokaryotic genomes are very compact, and prokaryotic transcripts often cover more than one gene or cistron (a coding sequence for a single protein). Polycistronic mRNAs are then translated to produce more than one kind of protein.

Our discussion here will exemplify transcription by describing this process in *Escherichia coli*, a well-studied eubacterial species. Although some differences exist between transcription in *E. coli* and transcription in archaea, an understanding of *E. coli* transcription can be applied to virtually all bacterial species.

Prokaryotic RNA Polymerase

Prokaryotes use the same RNA polymerase to transcribe all of their genes. In *E. coli*, the polymerase is composed of five polypeptide subunits, two of which are identical. Four of these subunits, denoted α, α, β, and β', comprise the polymerase **core enzyme**. These subunits assemble every time a gene is transcribed, and they disassemble once transcription is complete. Each subunit has a unique role; the two α-subunits are necessary to assemble the polymerase on the DNA; the β-subunit binds to the ribonucleoside triphosphate that will become part of the nascent mRNA molecule; and the β' subunit binds the DNA template strand. The fifth subunit, σ, is involved only in transcription initiation. It confers transcriptional specificity such that the polymerase begins to synthesize mRNA from an appropriate initiation site. Without σ, the core enzyme would transcribe from random sites and would produce mRNA molecules that specified protein gibberish. The polymerase comprised of all five subunits is called the **holoenzyme**.

Prokaryotic Promoters

A **promoter** is a DNA sequence onto which the transcription machinery, including RNA polymerase, binds and initiates transcription. In most cases, promoters exist upstream of the genes they regulate. The specific sequence of a promoter is very important because it determines whether the corresponding gene is transcribed all the time, some of the time, or infrequently. Although promoters vary among prokaryotic genomes, a few elements are evolutionarily conserved in many species. At the -10 and -35 regions upstream of the initiation site, there are two *promoter consensus sequences*, or regions that are similar across all promoters and across various bacterial species (Figure 15.8). The -10 sequence, called the -10 region, has the consensus sequence TATAAT. The -35 sequence has the consensus sequence TTGACA. These consensus sequences are recognized and bound by σ. Once this interaction is made, the subunits of the core enzyme bind to the site. The A–T-rich -10 region facilitates unwinding of the DNA template, and several phosphodiester bonds are made. The transcription initiation phase ends with the production of abortive transcripts, which are polymers of approximately 10 nucleotides that are made and released.
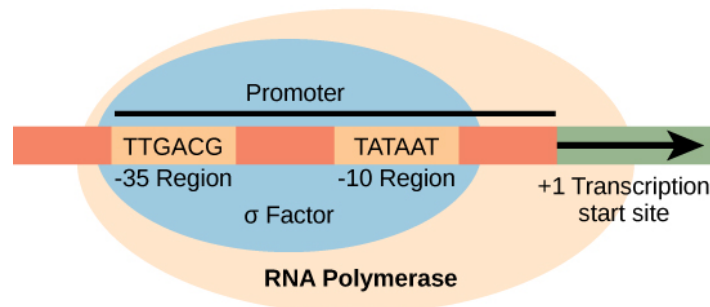
Figure 15.8 The σ subunit of prokaryotic RNA polymerase recognizes consensus sequences found in the promoter region upstream of the transcription start site. The σ subunit dissociates from the polymerase after transcription has been initiated.

View this MolecularMovies animation (http://openstax.org/l/transcription) to see the first part of transcription and the base sequence repetition of the TATA box.

## Elongation and Termination in Prokaryotes

The transcription elongation phase begins with the release of the σ subunit from the polymerase. The dissociation of σ allows the core enzyme to proceed along the DNA template, synthesizing mRNA in the 5' to 3' direction at a rate of approximately 40 nucleotides per second. As elongation proceeds, the DNA is continuously unwound ahead of the core enzyme and rewound behind it. The base pairing between DNA and RNA is not stable enough to maintain the stability of the mRNA synthesis components. Instead, the RNA polymerase acts as a stable linker between the DNA template and the nascent RNA strands to ensure that elongation is not interrupted prematurely.

Prokaryotic Termination Signals

Once a gene is transcribed, the prokaryotic polymerase needs to be instructed to dissociate from the DNA template and liberate the newly made mRNA. Depending on the gene being transcribed, there are two kinds of termination signals. One is protein-based and the other is RNA-based. **Rho-dependent termination** is controlled by the rho protein, which tracks along behind the polymerase on the growing mRNA chain. Near the end of the gene, the polymerase encounters a run of G nucleotides on the DNA template and it stalls. As a result, the rho protein collides with the polymerase. The interaction with rho releases the mRNA from the transcription bubble.

Rho-independent termination is controlled by specific sequences in the DNA template strand. As the polymerase nears the end of the gene being transcribed, it encounters a region rich in C–G nucleotides. The mRNA folds back on itself, and the complementary C–G nucleotides bind together. The result is a **stable hairpin** that causes the polymerase to stall as soon as it begins to transcribe a region rich in A–T nucleotides. The complementary U–A region of the mRNA transcript forms only a weak interaction with the template DNA. This, coupled with the stalled polymerase, induces enough instability for the core enzyme to break away and liberate the new mRNA transcript.

Upon termination, the process of transcription is complete. By the time termination occurs, the prokaryotic transcript would already have been used to begin synthesis of numerous copies of the encoded protein because these processes can occur concurrently. The unification of transcription, translation, and even mRNA degradation is possible because all of these processes occur in the same 5' to 3' direction, and because there is no membranous compartmentalization in the prokaryotic cell (Figure 15.9). In contrast, the presence of a nucleus in eukaryotic cells precludes simultaneous transcription and translation.
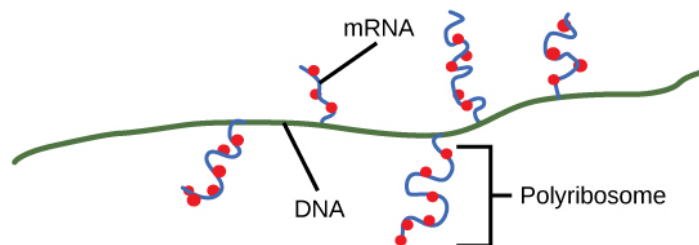
Figure 15.9 Multiple polymerases can transcribe a single bacterial gene while numerous ribosomes concurrently translate the mRNA transcripts into polypeptides. In this way, a specific protein can rapidly reach a high concentration in the bacterial cell.

Visit this BioStudio animation (http://openstax.org/l/transcription2) to see the process of prokaryotic transcription.

15.3 Eukaryotic Transcription

By the end of this section, you will be able to do the following:

- List the steps in eukaryotic transcription
- Discuss the role of RNA polymerases in transcription
- Compare and contrast the three RNA polymerases
- Explain the significance of transcription factors

Prokaryotes and eukaryotes perform fundamentally the same process of transcription, with a few key differences. The most important difference between prokaryote and eukaryote transcription is due to the latter's membrane-bound nucleus and organelles. With the genes bound in a nucleus, the eukaryotic cell must be able to transport its mRNA to the cytoplasm and must protect its mRNA from degrading before it is translated. Eukaryotes also employ three different polymerases that each transcribe a different subset of genes. Eukaryotic mRNAs are usually *monogenic*, meaning that they specify a single protein.

Initiation of Transcription in Eukaryotes

Unlike the prokaryotic polymerase that can bind to a DNA template on its own, eukaryotes require several other proteins, called transcription factors, to first bind to the promoter region and then to help recruit the appropriate polymerase.

The Three Eukaryotic RNA Polymerases

The features of eukaryotic mRNA synthesis are markedly more complex than those of prokaryotes. Instead of a single polymerase comprising five subunits, the eukaryotes have three polymerases that are each made up of 10 subunits or more. Each eukaryotic polymerase also requires a distinct set of transcription factors to bring it to the DNA template.

RNA polymerase I is located in the nucleolus, a specialized nuclear substructure in which ribosomal RNA (rRNA) is transcribed, processed, and assembled into ribosomes (Table 15.1). The rRNA molecules are considered structural RNAs because they have a cellular role but are not translated into protein. The rRNAs are components of the ribosome and are essential to the process of translation. RNA polymerase I synthesizes all of the rRNAs from the tandemly duplicated set of 18S, 5.8S, and 28S ribosomal genes. (Note that the "S" designation applies to "Svedberg" units, a nonadditive value that characterizes the speed at which a particle sediments during centrifugation.)

Locations, Products, and Sensitivities of the Three Eukaryotic RNA Polymerases

RNA Polymerase	Cellular Compartment	Product of Transcription	α-Amanitin Sensitivity
I	Nucleolus	All rRNAs except 5S rRNA	Insensitive
II	Nucleus	All protein-coding nuclear pre-mRNAs	Extremely sensitive
III	Nucleus	5S rRNA, tRNAs, and small nuclear RNAs	Moderately sensitive

Table 15.1

RNA polymerase II is located in the nucleus and synthesizes all protein-coding nuclear pre-mRNAs. Eukaryotic pre-mRNAs undergo extensive processing after transcription but before translation. For clarity, this module's discussion of transcription and translation in eukaryotes will use the term "mRNAs" to describe only the mature, processed molecules that are ready to be translated. RNA polymerase II is responsible for transcribing the overwhelming majority of eukaryotic genes.

RNA polymerase III is also located in the nucleus. This polymerase transcribes a variety of structural RNAs that includes the 5S pre-rRNA, transfer pre-RNAs (pre-tRNAs), and **small nuclear** pre-**RNAs**. The tRNAs have a critical role in translation; they serve as the "adaptor molecules" between the mRNA template and the growing polypeptide chain. Small nuclear RNAs have a variety of functions, including "splicing" pre-mRNAs and regulating transcription factors.

A scientist characterizing a new gene can determine which polymerase transcribes it by testing whether the gene is expressed in the presence of α-amanitin, an oligopeptide toxin produced by the fly agaric toadstool mushroom and other species of *Amanita*. Interestingly, the α-amanitin affects the three polymerases very differently (Table 15.1). RNA polymerase I is completely

insensitive to α-amanitin, meaning that the polymerase can transcribe DNA in vitro in the presence of this poison. RNA polymerase III is moderately sensitive to the toxin. In contrast, RNA polymerase II is extremely sensitive to α-amanitin. The toxin prevents the enzyme from progressing down the DNA, and thus inhibits transcription. Knowing the transcribing polymerase can provide clues as to the general function of the gene being studied. Because RNA polymerase II transcribes the vast majority of genes, we will focus on this polymerase in our subsequent discussions about eukaryotic transcription factors and promoters.
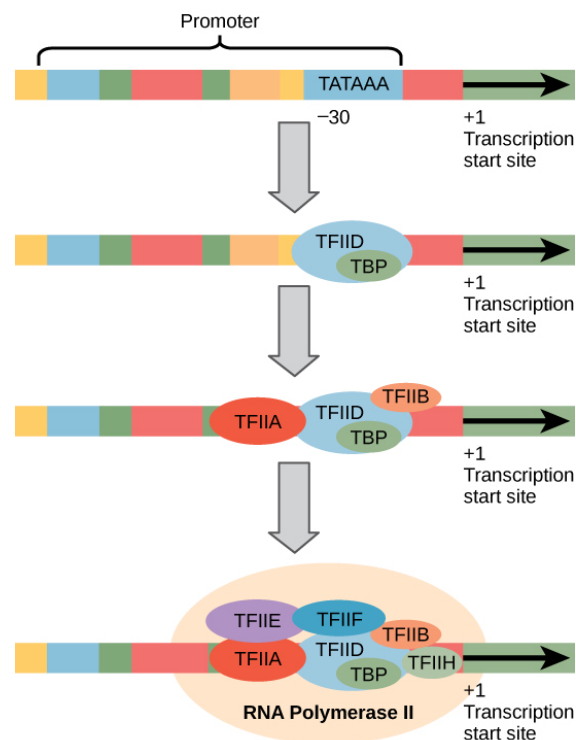
RNA Polymerase II Promoters and Transcription Factors

Eukaryotic promoters are much larger and more intricate than prokaryotic promoters. However, both have a sequence similar to the -10 sequence of prokaryotes. In eukaryotes, this sequence is called the TATA box, and has the consensus sequence TATAAA on the coding strand. It is located at -25 to -35 bases relative to the initiation (+1) site (Figure 15.10). This sequence is not identical to the *E. coli* -10 box, but it conserves the A–T rich element. The thermostability of A–T bonds is low and this helps the DNA template to locally unwind in preparation for transcription.

Instead of the simple σ factor that helps bind the prokaryotic RNA polymerase to its promoter, eukaryotes assemble a complex of transcription factors required to recruit RNA polymerase II to a protein coding gene. Transcription factors that bind to the promoter are called *basal transcription factors*. These basal factors are all called TFII (for Transcription Factor/polymerase II) plus an additional letter (A-J). The core complex is TFIID, which includes a TATA-binding protein (TBP). The other transcription factors systematically fall into place on the DNA template, with each one further stabilizing the pre-initiation complex and contributing to the recruitment of RNA polymerase II.

👁 VISUAL CONNECTION

Figure 15.10 A generalized promoter of a gene transcribed by RNA polymerase II is shown. Transcription factors recognize the promoter. RNA polymerase II then binds and forms the transcription initiation complex.

A scientist splices a eukaryotic promoter in front of a bacterial gene and inserts the gene in a bacterial chromosome. Would you expect the bacteria to transcribe the gene?

Some eukaryotic promoters also have a conserved **CAAT box** (GGCCAATCT) at approximately -80. Further upstream of the TATA box, eukaryotic promoters may also contain one or more **GC-rich boxes** (GGCG) or **octamer boxes** (ATTTGCAT). These elements

bind cellular factors that increase the efficiency of transcription initiation and are often identified in more "active" genes that are constantly being expressed by the cell.

Basal transcription factors are crucial in the formation of a preinitiation complex on the DNA template that subsequently recruits RNA polymerase II for transcription initiation. The complexity of eukaryotic transcription does not end with the polymerases and promoters. An army of other transcription factors, which bind to upstream enhancers and silencers, also help to regulate the frequency with which pre-mRNA is synthesized from a gene. Enhancers and silencers affect the efficiency of transcription but are not necessary for transcription to proceed.

Promoter Structures for RNA Polymerases I and III

The processes of bringing RNA polymerases I and III to the DNA template involve slightly less complex collections of transcription factors, but the general theme is the same.

The conserved promoter elements for genes transcribed by polymerases I and III differ from those transcribed by RNA polymerase II. RNA polymerase I transcribes genes that have two GC-rich promoter sequences in the -45 to +20 region. These sequences alone are sufficient for transcription initiation to occur, but promoters with additional sequences in the region from -180 to -105 upstream of the initiation site will further enhance initiation. Genes that are transcribed by RNA polymerase III have upstream promoters or promoters that occur within the genes themselves.

Eukaryotic transcription is a tightly regulated process that requires a variety of proteins to interact with each other and with the DNA strand. Although the process of transcription in eukaryotes involves a greater metabolic investment than in prokaryotes, it ensures that the cell transcribes precisely the pre-mRNAs that it needs for protein synthesis.

EVOLUTION CONNECTION

The Evolution of Promoters

The evolution of genes may be a familiar concept. Mutations can occur in genes during DNA replication, and the result may or may not be beneficial to the cell. By altering an enzyme, structural protein, or some other factor, the process of mutation can transform functions or physical features. However, eukaryotic promoters and other gene regulatory sequences may evolve as well. For instance, consider a gene that, over many generations, becomes more valuable to the cell. Maybe the gene encodes a structural protein that the cell needs to synthesize in abundance for a certain function. If this is the case, it would be beneficial to the cell for that gene's promoter to recruit transcription factors more efficiently and increase gene expression.

Scientists examining the evolution of promoter sequences have reported varying results. In part, this is because it is difficult to infer exactly where a eukaryotic promoter begins and ends. Some promoters occur within genes; others are located very far upstream, or even downstream, of the genes they are regulating. However, when researchers limited their examination to human core promoter sequences that were defined experimentally as sequences that bind the preinitiation complex, they found that promoters evolve even faster than protein-coding genes.

It is still unclear how promoter evolution might correspond to the evolution of humans or other complex organisms. However, the evolution of a promoter to effectively make more or less of a given gene product is an intriguing alternative to the evolution of the genes themselves.[1]

Eukaryotic Elongation and Termination

Following the formation of the preinitiation complex, the polymerase is released from the other transcription factors, and elongation is allowed to proceed as it does in prokaryotes with the polymerase synthesizing pre-mRNA in the 5' to 3' direction. As discussed previously, RNA polymerase II transcribes the major share of eukaryotic genes, so in this section we will focus on how this polymerase accomplishes elongation and termination.

Although the enzymatic process of elongation is essentially the same in eukaryotes and prokaryotes, the DNA template is considerably more complex. When eukaryotic cells are not dividing, their genes exist as a diffuse mass of DNA and proteins called chromatin. The DNA is tightly packaged around charged histone proteins at repeated intervals. These *DNA–histone complexes*, collectively called nucleosomes, are regularly spaced and include 146 nucleotides of DNA wound around eight

1H Liang et al., "Fast evolution of core promoters in primate genomes," *Molecular Biology and Evolution* 25 (2008): 1239–44.

histones like thread around a spool.

For polynucleotide synthesis to occur, the transcription machinery needs to move histones out of the way every time it encounters a nucleosome. This is accomplished by a special protein complex called **FACT**, which stands for "*facilitates chromatin transcription.*" This complex pulls histones away from the DNA template as the polymerase moves along it. Once the pre-mRNA is synthesized, the FACT complex replaces the histones to recreate the nucleosomes.

The termination of transcription is different for the different polymerases. Unlike in prokaryotes, elongation by RNA polymerase II in eukaryotes takes place 1,000 to 2,000 nucleotides beyond the end of the gene being transcribed. This pre-mRNA tail is subsequently removed by cleavage during mRNA processing. On the other hand, RNA polymerases I and III require termination signals. Genes transcribed by RNA polymerase I contain a specific 18-nucleotide sequence that is recognized by a termination protein. The process of termination in RNA polymerase III involves an mRNA hairpin similar to rho-independent termination of transcription in prokaryotes.

15.4 RNA Processing in Eukaryotes

*By the end of this section, you will be able to do the following:*
- Describe the different steps in RNA processing
- Understand the significance of exons, introns, and splicing for mRNAs
- Explain how tRNAs and rRNAs are processed

After transcription, eukaryotic pre-mRNAs must undergo several processing steps before they can be translated. Eukaryotic (and prokaryotic) tRNAs and rRNAs also undergo processing before they can function as components in the protein-synthesis machinery.

mRNA Processing

The eukaryotic pre-mRNA undergoes extensive processing before it is ready to be translated. Eukaryotic protein-coding sequences are not continuous, as they are in prokaryotes. The coding sequences (exons) are interrupted by noncoding introns, which must be removed to make a translatable mRNA. The additional steps involved in eukaryotic mRNA maturation also create a molecule with a much longer half-life than a prokaryotic mRNA. Eukaryotic mRNAs last for several hours, whereas the typical *E. coli* mRNA lasts no more than five seconds.

Pre-mRNAs are first coated in RNA-stabilizing proteins; these protect the pre-mRNA from degradation while it is processed and exported out of the nucleus. The three most important steps of pre-mRNA processing are the addition of stabilizing and signaling factors at the 5' and 3' ends of the molecule, and the removal of the introns (Figure 15.11). In rare cases, the mRNA transcript can be "edited" after it is transcribed.
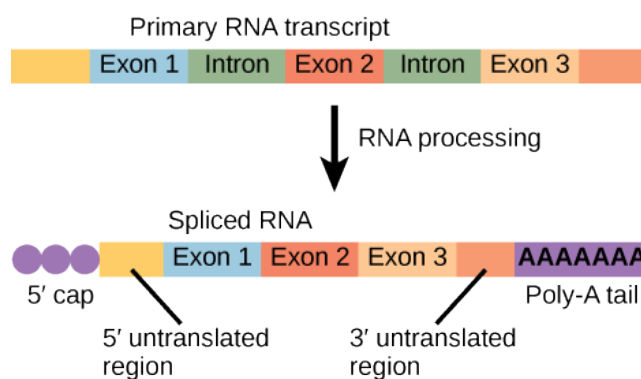
Figure 15.11 Eukaryotic mRNA contains introns that must be spliced out. A 5' cap and 3' poly-A tail are also added.

EVOLUTION CONNECTION

RNA Editing in Trypanosomes
The trypanosomes are a group of protozoa that include the pathogen *Trypanosoma brucei*, which causes nagana in cattle and sleeping sickness in humans throughout great areas of Africa (Figure 15.12). The trypanosome is carried by biting flies in the

genus *Glossina* (commonly called tsetse flies). Trypanosomes, and virtually all other eukaryotes, have organelles called mitochondria that supply the cell with chemical energy. Mitochondria are organelles that express their own DNA and are believed to be the remnants of a symbiotic relationship between a eukaryote and an engulfed prokaryote. The mitochondrial DNA of trypanosomes exhibit an interesting exception to the central dogma: their pre-mRNAs do not have the correct information to specify a functional protein. Usually, this is because the mRNA is missing several U nucleotides. The cell performs an additional RNA processing step called **RNA editing** to remedy this.

Figure 15.12 *Trypanosoma brucei* is the causative agent of sleeping sickness in humans. The mRNAs of this pathogen must be modified by the addition of nucleotides before protein synthesis can occur. (credit: modification of work by Torsten Ochsenreiter)

Other genes in the mitochondrial genome encode 40- to 80-nucleotide guide RNAs. One or more of these molecules interacts by complementary base pairing with some of the nucleotides in the pre-mRNA transcript. However, the *guide RNA* has more A nucleotides than the pre-mRNA has U nucleotides with which to bind. In these regions, the guide RNA loops out. The 3' ends of guide RNAs have a long poly-U tail, and these U bases are inserted in regions of the pre-mRNA transcript at which the guide RNAs are looped. This process is entirely mediated by RNA molecules. That is, guide RNAs—rather than proteins—serve as the catalysts in RNA editing.

RNA editing is not just a phenomenon of trypanosomes. In the mitochondria of some plants, almost all pre-mRNAs are edited. RNA editing has also been identified in mammals such as rats, rabbits, and even humans. What could be the evolutionary reason for this additional step in pre-mRNA processing? One possibility is that the mitochondria, being remnants of ancient prokaryotes, have an equally ancient RNA-based method for regulating gene expression. In support of this hypothesis, edits made to pre-mRNAs differ depending on cellular conditions. Although speculative, the process of RNA editing may be a holdover from a primordial time when RNA molecules, instead of proteins, were responsible for catalyzing reactions.

5' Capping

While the pre-mRNA is still being synthesized, a **7-methylguanosine cap** is added to the 5' end of the growing transcript by a phosphate linkage. This functional group protects the nascent mRNA from degradation. In addition, factors involved in protein synthesis recognize the cap to help initiate translation by ribosomes.

3' Poly-A Tail

Once elongation is complete, the pre-mRNA is cleaved by an endonuclease between an AAUAAA consensus sequence and a GU-rich sequence, leaving the AAUAAA sequence on the pre-mRNA. An enzyme called poly-A polymerase then adds a string of approximately 200 A residues, called the **poly-A tail**. This modification further protects the pre-mRNA from degradation and is also the binding site for a protein necessary for exporting the processed mRNA to the cytoplasm.

Pre-mRNA Splicing

Eukaryotic genes are composed of **exons**, which correspond to protein-coding sequences (*ex*-on signifies that they are *ex*pressed), and *int*ervening sequences called **introns** (*int*-ron denotes their *int*ervening role), which may be involved in gene regulation but are removed from the pre-mRNA during processing. Intron sequences in mRNA do not encode functional
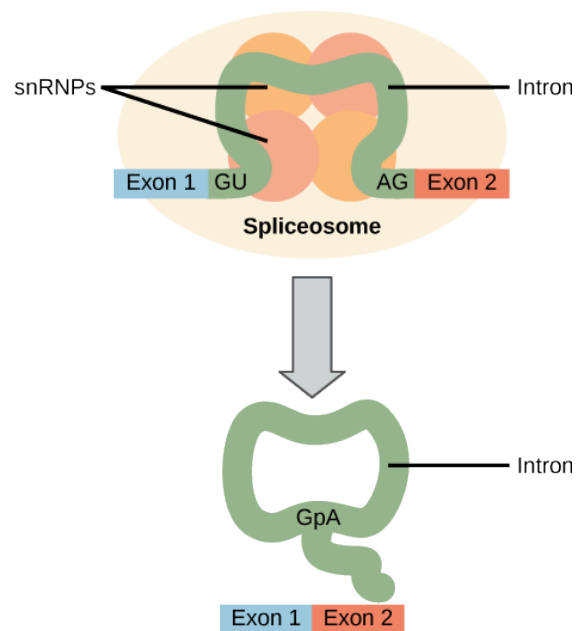
proteins.

The discovery of introns came as a surprise to researchers in the 1970s who expected that pre-mRNAs would specify protein sequences without further processing, as they had observed in prokaryotes. The genes of higher eukaryotes very often contain one or more introns. These regions may correspond to regulatory sequences; however, the biological significance of having many introns or having very long introns in a gene is unclear. It is possible that introns slow down gene expression because it takes longer to transcribe pre-mRNAs with lots of introns. Alternatively, introns may be nonfunctional sequence remnants left over from the fusion of ancient genes throughout the course of evolution. This is supported by the fact that separate exons often encode separate protein subunits or domains. For the most part, the sequences of introns can be mutated without ultimately affecting the protein product.

All of a pre-mRNA's introns must be completely and precisely removed before protein synthesis. If the process errs by even a single nucleotide, the reading frame of the rejoined exons would shift, and the resulting protein would be dysfunctional. The process of removing introns and reconnecting exons is called **splicing** (Figure 15.13). Introns are removed and degraded while the pre-mRNA is still in the nucleus. Splicing occurs by a sequence-specific mechanism that ensures introns will be removed and exons rejoined with the accuracy and precision of a single nucleotide. Although the intron itself is noncoding, the beginning and end of each intron is marked with specific nucleotides: GU at the 5' end and AG at the 3' end of the intron. The splicing of pre-mRNAs is conducted by complexes of proteins and RNA molecules called spliceosomes.

👁 VISUAL CONNECTION

Figure 15.13 Pre-mRNA splicing involves the precise removal of introns from the primary RNA transcript. The splicing process is catalyzed by protein complexes called spliceosomes that are composed of proteins and RNA molecules called small nuclear RNAs (snRNAs). Spliceosomes recognize sequences at the 5' and 3' end of the intron.

Errors in splicing are implicated in cancers and other human diseases. What kinds of mutations might lead to splicing errors? Think of different possible outcomes if splicing errors occur.

Note that more than 70 individual introns can be present, and each has to undergo the process of splicing—in addition to 5' capping and the addition of a poly-A tail—just to generate a single, translatable mRNA molecule.

🔗 LINK TO LEARNING

See how introns are removed during RNA splicing at this website (http://openstax.org/l/RNA_splicing) .

Processing of tRNAs and rRNAs

The tRNAs and rRNAs are structural molecules that have roles in protein synthesis; however, these RNAs are not themselves translated. Pre-rRNAs are transcribed, processed, and assembled into ribosomes in the nucleolus. Pre-tRNAs are transcribed and processed in the nucleus and then released into the cytoplasm where they are linked to free amino acids for protein synthesis.

Most of the tRNAs and rRNAs in eukaryotes and prokaryotes are first transcribed as a long precursor molecule that spans multiple rRNAs or tRNAs. Enzymes then cleave the precursors into subunits corresponding to each structural RNA. Some of the bases of pre-rRNAs are *methylated*; that is, a $-CH_3$ methyl functional group is added for stability. Pre-tRNA molecules also undergo methylation. As with pre-mRNAs, subunit excision occurs in eukaryotic pre-RNAs destined to become tRNAs or rRNAs.

Mature rRNAs make up approximately 50 percent of each ribosome. Some of a ribosome's RNA molecules are purely structural, whereas others have catalytic or binding activities. Mature tRNAs take on a three-dimensional structure through local regions of base pairing stabilized by intramolecular hydrogen bonding. The tRNA folds to position the amino acid binding site at one end and the **anticodon** at the other end (Figure 15.14). The anticodon is a three-nucleotide sequence in a tRNA that interacts with an mRNA codon through complementary base pairing.
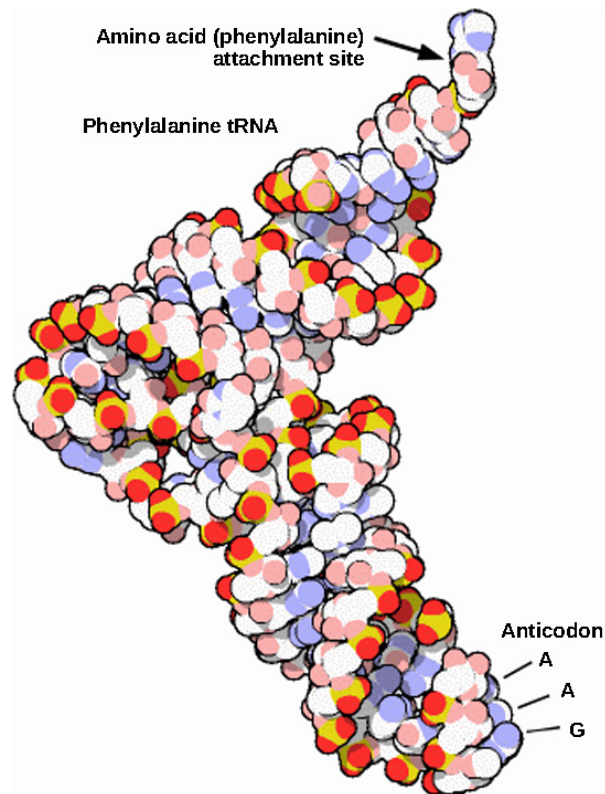
Figure 15.14 This is a space-filling model of a tRNA molecule that adds the amino acid phenylalanine to a growing polypeptide chain. The anticodon AAG binds the Codon UUC on the mRNA. The amino acid phenylalanine is attached to the other end of the tRNA.

15.5 Ribosomes and Protein Synthesis

By the end of this section, you will be able to do the following:
- Describe the different steps in protein synthesis
- Discuss the role of ribosomes in protein synthesis

The synthesis of proteins consumes more of a cell's energy than any other metabolic process. In turn, proteins account for more mass than any other component of living organisms (with the exception of water), and proteins perform virtually every function of a cell. The process of translation, or protein synthesis, involves the decoding of an mRNA message into a polypeptide product. Amino acids are covalently strung together by interlinking peptide bonds in lengths ranging from approximately 50 to more

than 1000 amino acid residues. Each individual amino acid has an amino group (NH_2) and a carboxyl (COOH) group. Polypeptides are formed when the amino group of one amino acid forms an amide (i.e., peptide) bond with the carboxyl group of another amino acid (Figure 15.15). This reaction is catalyzed by ribosomes and generates one water molecule.
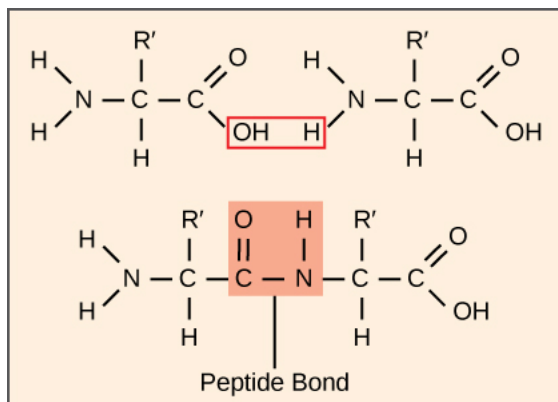
Figure 15.15 A peptide bond links the carboxyl end of one amino acid with the amino end of another, producing one water molecule during the process. For simplicity in this image, only the functional groups involved in the peptide bond are shown. The R and R' designations refer to the rest of each amino acid structure.

The Protein Synthesis Machinery

In addition to the mRNA template, many molecules and macromolecules contribute to the process of translation. The composition of each component may vary across species; for example, ribosomes may consist of different numbers of rRNAs and polypeptides depending on the organism. However, the general structures and functions of the protein synthesis machinery are comparable from bacteria to human cells. Translation requires the input of an mRNA template, ribosomes, tRNAs, and various enzymatic factors. (Note: A ribosome can be thought of as an enzyme whose amino acid binding sites are specified by mRNA.)

🔗 LINK TO LEARNING

Click through the steps of this PBS interactive (http://openstax.org/l/prokary_protein) to see protein synthesis in action.

Ribosomes

Even before an mRNA is translated, a cell must invest energy to build each of its ribosomes. In *E. coli*, there are between 10,000 and 70,000 ribosomes present in each cell at any given time. A **ribosome** is a complex macromolecule composed of structural and catalytic rRNAs, and many distinct polypeptides. In eukaryotes, the nucleolus is completely specialized for the synthesis and assembly of rRNAs.

Ribosomes exist in the cytoplasm of prokaryotes and in the cytoplasm and rough endoplasmic reticulum of eukaryotes. Mitochondria and chloroplasts also have their own ribosomes in the matrix and stroma, which look more similar to prokaryotic ribosomes (and have similar drug sensitivities) than the ribosomes just outside their outer membranes in the cytoplasm. Ribosomes dissociate into large and small subunits when they are not synthesizing proteins and reassociate during the initiation of translation. *In E. coli, the small subunit is described as 30S, and the large subunit is 50S, for a total of 70S (recall that Svedberg units are not additive).* Mammalian ribosomes have a small 40S subunit and a large 60S subunit, for a total of 80S. The small subunit is responsible for binding the mRNA template, whereas the large subunit sequentially binds tRNAs. Each mRNA molecule is simultaneously translated by many ribosomes, all synthesizing protein in the same direction: reading the mRNA from 5' to 3' and synthesizing the polypeptide from the N terminus to the C terminus. The complete mRNA/poly-ribosome structure is called a **polysome**.

tRNAs

The tRNAs are structural RNA molecules that were transcribed from genes by RNA polymerase III. Depending on the species, 40 to 60 types of tRNAs exist in the cytoplasm. Transfer RNAs serve as adaptor molecules. Each tRNA carries a specific amino acid and recognizes one or more of the mRNA codons that define the order of amino acids in a protein. Aminoacyl-tRNAs bind to the ribosome and add the corresponding amino acid to the polypeptide chain. Therefore, tRNAs are the molecules that actually

"translate" the language of RNA into the language of proteins.

Of the 64 possible mRNA codons—or triplet combinations of A, U, G, and C—three specify the termination of protein synthesis and 61 specify the addition of amino acids to the polypeptide chain. Of these 61, one codon (AUG) also encodes the initiation of translation. Each tRNA anticodon can base pair with one or more of the mRNA codons for its amino acid. For instance, if the sequence CUA occurred on an mRNA template in the proper reading frame, it would bind a leucine tRNA expressing the complementary sequence, GAU. The ability of some tRNAs to match more than one codon is what gives the genetic code its blocky structure.

As the adaptor molecules of translation, it is surprising that tRNAs can fit so much specificity into such a small package. Consider that tRNAs need to interact with three factors: 1) they must be recognized by the correct aminoacyl synthetase (see below); 2) they must be recognized by ribosomes; and 3) they must bind to the correct sequence in mRNA.

Aminoacyl tRNA Synthetases

The process of pre-tRNA synthesis by RNA polymerase III only creates the RNA portion of the adaptor molecule. The corresponding amino acid must be added later, once the tRNA is processed and exported to the cytoplasm. Through the process of tRNA "charging," each tRNA molecule is linked to its correct amino acid by one of a group of enzymes called **aminoacyl tRNA synthetases**. At least one type of aminoacyl tRNA synthetase exists for each of the 20 amino acids; the exact number of aminoacyl tRNA synthetases varies by species. These enzymes first bind and hydrolyze ATP to catalyze a high-energy bond between an amino acid and adenosine monophosphate (AMP); a pyrophosphate molecule is expelled in this reaction. The activated amino acid is then transferred to the tRNA, and AMP is released. The term "charging" is appropriate, since the high-energy bond that attaches an amino acid to its tRNA is later used to drive the formation of the peptide bond. Each tRNA is named for its amino acid.

The Mechanism of Protein Synthesis

As with mRNA synthesis, protein synthesis can be divided into three phases: *initiation, elongation, and termination*. The process of translation is similar in prokaryotes and eukaryotes. Here we'll explore how translation occurs in *E. coli*, a representative prokaryote, and specify any differences between prokaryotic and eukaryotic translation.

Initiation of Translation

Protein synthesis begins with the formation of an **initiation complex**. In *E. coli*, this complex involves the small 30S ribosome, the mRNA template, three initiation factors (IFs; IF-1, IF-2, and IF-3), and a special **initiator tRNA**, called tRNA$^{\text{Metf}}$.

In *E. coli* mRNA, a sequence upstream of the first AUG codon, called the Shine-Dalgarno sequence (AGGAGG), interacts with the rRNA molecules that compose the ribosome. This interaction anchors the 30S ribosomal subunit at the correct location on the mRNA template. Guanosine triphosphate (GTP), which is a purine nucleotide triphosphate, acts as an energy source during translation—both at the start of elongation and during the ribosome's translocation. Binding of the mRNA to the 30S ribosome also requires IF-III.

The initiator tRNA then interacts with the **start codon** AUG (or rarely, GUG). This tRNA carries the amino acid methionine, which is formylated after its attachment to the tRNA. The formylation creates a "faux" peptide bond between the formyl carboxyl group and the amino group of the methionine. Binding of the fMet-tRNA$^{\text{Metf}}$ is mediated by the initiation factor IF-2. The fMet begins every polypeptide chain synthesized by *E. coli*, but it is usually removed after translation is complete. When an in-frame AUG is encountered during translation elongation, a non-formylated methionine is inserted by a regular Met-tRNA$^{\text{Met}}$. After the formation of the initiation complex, the 30S ribosomal subunit is joined by the 50S subunit to form the translation complex. In eukaryotes, a similar initiation complex forms, comprising mRNA, the 40S small ribosomal subunit, eukaryotic IFs, and nucleoside triphosphates (GTP and ATP). The methionine on the charged initiator tRNA, called Met-tRNA$_i$, is not formylated. However, Met-tRNA$_i$ is distinct from other Met-tRNAs in that it can bind IFs.

Instead of depositing at the Shine-Dalgarno sequence, the eukaryotic initiation complex recognizes the 7-methylguanosine cap at the 5' end of the mRNA. A cap-binding protein (CBP) and several other IFs assist the movement of the ribosome to the 5' cap. Once at the cap, the initiation complex tracks along the mRNA in the 5' to 3' direction, searching for the AUG start codon. Many eukaryotic mRNAs are translated from the first AUG, but this is not always the case. According to **Kozak's rules**, the nucleotides around the AUG indicate whether it is the correct start codon. Kozak's rules state that the following consensus sequence must appear around the AUG of vertebrate genes: 5'-gccRccAUGG-3'. The R (for purine) indicates a site that can be either A or G, but cannot be C or U. Essentially, the closer the sequence is to this consensus, the higher the efficiency of translation.

Once the appropriate AUG is identified, the other proteins and CBP dissociate, and the 60S subunit binds to the complex of Met-tRNA$_i$, mRNA, and the 40S subunit. This step completes the initiation of translation in eukaryotes.

## Translation, Elongation, and Termination

In prokaryotes and eukaryotes, the basics of elongation are the same, so we will review elongation from the perspective of *E. coli*. When the translation complex is formed, the tRNA binding region of the ribosome consists of three compartments. The A (aminoacyl) site binds incoming charged aminoacyl tRNAs. The P (peptidyl) site binds charged tRNAs carrying amino acids that have formed peptide bonds with the growing polypeptide chain but have not yet dissociated from their corresponding tRNA. The E (exit) site releases dissociated tRNAs so that they can be recharged with free amino acids. The initiating methionyl-tRNA, however, occupies the P site at the beginning of the elongation phase of translation in both prokaryotes and eukaryotes.

During translation elongation, the mRNA template provides tRNA binding specificity. As the ribosome moves along the mRNA, each mRNA codon comes into register, and specific binding with the corresponding charged tRNA anticodon is ensured. If mRNA were not present in the elongation complex, the ribosome would bind tRNAs nonspecifically and randomly (?).

Elongation proceeds with charged tRNAs sequentially entering and leaving the ribosome as each new amino acid is added to the polypeptide chain. Movement of a tRNA from A to P to E site is induced by conformational changes that advance the ribosome by three bases in the 3' direction. The energy for each step along the ribosome is donated by elongation factors that hydrolyze GTP. GTP energy is required both for the binding of a new aminoacyl-tRNA to the A site and for its translocation to the P site after formation of the peptide bond. Peptide bonds form between the amino group of the amino acid attached to the A-site tRNA and the carboxyl group of the amino acid attached to the P-site tRNA. The formation of each peptide bond is catalyzed by **peptidyl transferase**, an RNA-based enzyme that is integrated into the 50S ribosomal subunit. The energy for each peptide bond formation is derived from the high-energy bond linking each amino acid to its tRNA. After peptide bond formation, the A-site tRNA that now holds the growing peptide chain moves to the P site, and the P-site tRNA that is now empty moves to the E site and is expelled from the ribosome (Figure 15.16). Amazingly, the *E. coli* translation apparatus takes only 0.05 seconds to add each amino acid, meaning that a 200-amino-acid protein can be translated in just 10 seconds.

VISUAL CONNECTION
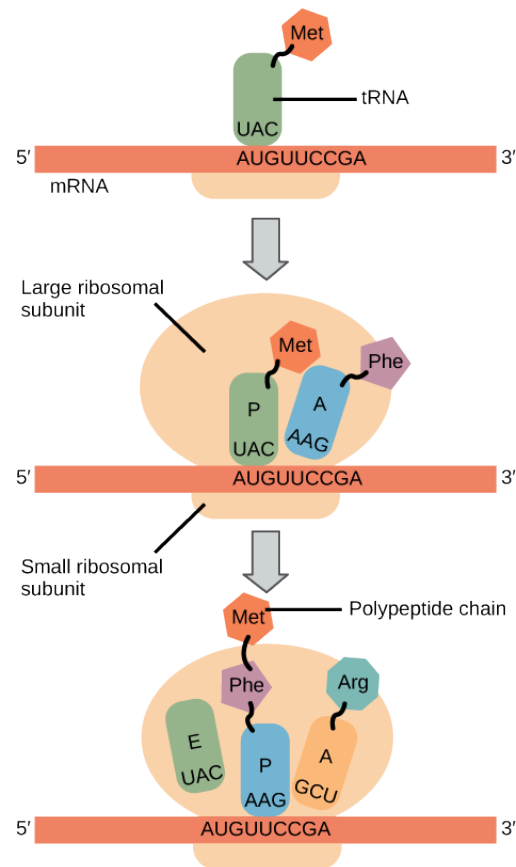
Figure 15.16 Translation begins when an initiator tRNA anticodon recognizes a start codon on mRNA bound to a small ribosomal subunit. The large ribosomal subunit joins the small subunit, and a second tRNA is recruited. As the mRNA moves relative to the ribosome, successive tRNAs move through the ribosome and the polypeptide chain is formed. Entry of a release factor into the A site terminates translation and the components dissociate.

Many antibiotics inhibit bacterial protein synthesis. For example, tetracycline blocks the A site on the bacterial ribosome, and chloramphenicol blocks peptidyl transfer. What specific effect would you expect each of these antibiotics to have on protein synthesis?

Tetracycline would directly affect:

a. tRNA binding to the ribosome
b. ribosome assembly
c. growth of the protein chain

Chloramphenicol would directly affect:

a. tRNA binding to the ribosome
b. ribosome assembly
c. growth of the protein chain

Termination of translation occurs when a nonsense codon (UAA, UAG, or UGA) is encountered. Upon aligning with the A site, these nonsense codons are recognized by *protein release factors* that resemble tRNAs. The releasing factors in both prokaryotes and eukaryotes instruct peptidyl transferase to add a water molecule to the carboxyl end of the P-site amino acid. This reaction forces the P-site amino acid to detach from its tRNA, and the newly made protein is released. The small and large ribosomal subunits dissociate from the mRNA and from each other; they are recruited almost immediately into another translation

initiation complex. After many ribosomes have completed translation, the mRNA is degraded so the nucleotides can be reused in another transcription reaction.

Protein Folding, Modification, and Targeting

During and after translation, individual amino acids may be chemically modified, signal sequences appended, and the new protein "folded" into a distinct three-dimensional structure as a result of intramolecular interactions. A **signal sequence** is a short sequence at the amino end of a protein that directs it to a specific cellular compartment. These sequences can be thought of as the protein's "train ticket" to its ultimate destination, and are recognized by signal-recognition proteins that act as conductors. For instance, a specific signal sequence terminus will direct a protein to the mitochondria or chloroplasts (in plants). Once the protein reaches its cellular destination, the signal sequence is usually clipped off.

Many proteins fold spontaneously, but some proteins require helper molecules, called *chaperones*, to prevent them from aggregating during the complicated process of folding. Even if a protein is properly specified by its corresponding mRNA, it could take on a completely dysfunctional shape if abnormal temperature or pH conditions prevent it from folding correctly.

KEY TERMS

7-methylguanosine cap   modification added to the 5' end of pre-mRNAs to protect mRNA from degradation and assist translation

aminoacyl tRNA synthetase   enzyme that "charges" tRNA molecules by catalyzing a bond between the tRNA and a corresponding amino acid

anticodon   three-nucleotide sequence in a tRNA molecule that corresponds to an mRNA codon

CAAT box   (GGCCAATCT) essential eukaryotic promoter sequence involved in binding transcription factors

central dogma   states that genes specify the sequence of mRNAs, which in turn specify the sequence of proteins

codon   three consecutive nucleotides in mRNA that specify the insertion of an amino acid or the release of a polypeptide chain during translation

colinear   in terms of RNA and protein, three "units" of RNA (nucleotides) specify one "unit" of protein (amino acid) in a consecutive fashion

consensus   DNA sequence that is used by many species to perform the same or similar functions

core enzyme   prokaryotic RNA polymerase consisting of α, α, β, and β' but missing σ; this complex performs elongation

degeneracy   (of the genetic code) describes that a given amino acid can be encoded by more than one nucleotide triplet; the code is degenerate, but not ambiguous

downstream   nucleotides following the initiation site in the direction of mRNA transcription; in general, sequences that are toward the 3' end relative to a site on the mRNA

exon   sequence present in protein-coding mRNA after completion of pre-mRNA splicing

FACT   complex that "facilitates chromatin transcription" by disassembling nucleosomes ahead of a transcribing RNA polymerase II and reassembling them after the polymerase passes by

GC-rich box   (GGCG) nonessential eukaryotic promoter sequence that binds cellular factors to increase the efficiency of transcription; may be present several times in a promoter

hairpin   structure of RNA when it folds back on itself and forms intramolecular hydrogen bonds between complementary nucleotides

holoenzyme   prokaryotic RNA polymerase consisting of α, α, β, β', and σ; this complex is responsible for transcription initiation

initiation site   nucleotide from which mRNA synthesis proceeds in the 5' to 3' direction; denoted with a "+1"

initiator tRNA   in prokaryotes, called $tRNA_f^{Met}$; in eukaryotes, called $tRNA_i$; a tRNA that interacts with a start codon, binds directly to the ribosome P site, and links to a special methionine to begin a polypeptide chain

intron   non–protein-coding intervening sequences that are spliced from mRNA during processing

Kozak's rules   determines the correct initiation AUG in a eukaryotic mRNA; the following consensus sequence must appear around the AUG: 5'-GCC(**purine**)CC<u>AUG</u>**G**-3'; the bolded bases are most important

nonsense codon   one of the three mRNA codons that specifies termination of translation

nontemplate strand   strand of DNA that is not used to transcribe mRNA; this strand is identical to the mRNA except that T nucleotides in the DNA are replaced by U nucleotides in the mRNA

Octamer box   (ATTTGCAT) nonessential eukaryotic promoter sequence that binds cellular factors to increase the efficiency of transcription; may be present several times in a promoter

peptidyl transferase   RNA-based enzyme that is integrated into the 50S ribosomal subunit and catalyzes the formation of peptide bonds

plasmid   extrachromosomal, covalently closed, circular DNA molecule that may only contain one or a few genes; common in prokaryotes

poly-A tail   modification added to the 3' end of pre-mRNAs to protect mRNA from degradation and assist mRNA export from the nucleus

polysome   mRNA molecule simultaneously being translated by many ribosomes all going in the same direction

preinitiation complex   cluster of transcription factors and other proteins that recruit RNA polymerase II for transcription of a DNA template

promoter   DNA sequence to which RNA polymerase and associated factors bind and initiate transcription

reading frame   sequence of triplet codons in mRNA that specify a particular protein; a ribosome shift of one or two nucleotides in either direction completely abolishes synthesis of that protein

rho-dependent termination   in prokaryotes, termination of transcription by an interaction between RNA polymerase and the rho protein at a run of G nucleotides on the DNA template

rho-independent   termination sequence-dependent termination of prokaryotic mRNA synthesis; caused by hairpin formation in the mRNA that stalls the polymerase

RNA editing   direct alteration of one or more nucleotides in an mRNA that has already been synthesized

Shine-Dalgarno sequence   (AGGAGG); initiates prokaryotic translation by interacting with rRNA molecules comprising the 30S ribosome

signal sequence   short tail of amino acids that directs a protein to a specific cellular compartment

small nuclear RNA   molecules synthesized by RNA

polymerase III that have a variety of functions, including splicing pre-mRNAs and regulating transcription factors

splicing   process of removing introns and reconnecting exons in a pre-mRNA

start codon   AUG (or rarely, GUG) on an mRNA from which translation begins; always specifies methionine

TATA box   conserved promoter sequence in eukaryotes and prokaryotes that helps to establish the initiation site for

transcription

template strand   strand of DNA that specifies the complementary mRNA molecule

transcription bubble   region of locally unwound DNA that allows for transcription of mRNA

upstream   nucleotides preceding the initiation site; in general, sequences toward the 5' end relative to a site on the mRNA

CHAPTER SUMMARY

15.1 The Genetic Code

The genetic code refers to the DNA alphabet (A, T, C, G), the RNA alphabet (A, U, C, G), and the polypeptide alphabet (20 amino acids). The central dogma describes the flow of genetic information in the cell from genes to mRNA to proteins. Genes are used to make mRNA by the process of transcription; mRNA is used to synthesize proteins by the process of translation. The genetic code is degenerate because 64 triplet codons in mRNA specify only 20 amino acids and three nonsense codons. Most amino acids have several similar codons. Almost every species on the planet uses the same genetic code.

15.2 Prokaryotic Transcription

In prokaryotes, mRNA synthesis is initiated at a promoter sequence on the DNA template comprising two consensus sequences that recruit RNA polymerase. The prokaryotic polymerase consists of a core enzyme of four protein subunits and a σ protein that assists only with initiation. Elongation synthesizes mRNA in the 5' to 3' direction at a rate of 40 nucleotides per second. Termination liberates the mRNA and occurs either by rho protein interaction or by the formation of an mRNA hairpin.

15.3 Eukaryotic Transcription

Transcription in eukaryotes involves one of three types of polymerases, depending on the gene being transcribed. RNA polymerase II transcribes all of the protein-coding genes, whereas RNA polymerase I transcribes the tandemly duplicated rRNA genes, and RNA polymerase III transcribes various small RNAs, like the 5S rRNA, tRNA, and small nuclear RNA genes. The initiation of transcription in eukaryotes involves the binding of several transcription factors to complex promoter sequences that are usually located upstream of the gene being copied. The mRNA is synthesized in the 5' to 3' direction, and the FACT complex moves and reassembles nucleosomes as the polymerase

passes by. Whereas RNA polymerases I and III terminate transcription by protein- or RNA hairpin-dependent methods, RNA polymerase II transcribes for 1,000 or more nucleotides beyond the gene template and cleaves the excess during pre-mRNA processing.

15.4 RNA Processing in Eukaryotes

Eukaryotic pre-mRNAs are modified with a 5' methylguanosine cap and a poly-A tail. These structures protect the mature mRNA from degradation and help export it from the nucleus. Pre-mRNAs also undergo splicing, in which introns are removed and exons are reconnected with single-nucleotide accuracy. Only finished mRNAs that have undergone 5' capping, 3' polyadenylation, and intron splicing are exported from the nucleus to the cytoplasm. Pre-rRNAs and pre-tRNAs may be processed by intramolecular cleavage, splicing, methylation, and chemical conversion of nucleotides. Rarely, RNA editing is also performed to insert missing bases after an mRNA has been synthesized.

15.5 Ribosomes and Protein Synthesis

The players in translation include the mRNA template, ribosomes, tRNAs, and various enzymatic factors. The small ribosomal subunit binds to the mRNA template either at the Shine-Dalgarno sequence (prokaryotes) or the 5' cap (eukaryotes). Translation begins at the initiating AUG on the mRNA, specifying methionine. The formation of peptide bonds occurs between sequential amino acids matched to the mRNA template by their tRNAs according to the genetic code. Charged tRNAs enter the ribosomal A site, and their amino acid bonds with the amino acid at the P site. The entire mRNA is translated in three-nucleotide "steps" of the ribosome. When a nonsense codon is encountered, a release factor binds and dissociates the components and frees the new protein. Folding of the protein occurs during and after translation.

VISUAL CONNECTION QUESTIONS

1. Figure 15.11 A scientist splices a eukaryotic promoter in front of a bacterial gene and inserts the gene in a bacterial chromosome. Would you expect the bacteria to transcribe the gene?

2. Figure 15.13 Errors in splicing are implicated in cancers and other human diseases. What kinds of mutations might lead to splicing errors? Think of different possible outcomes if splicing errors occur.

3. Figure 15.16 Many antibiotics inhibit bacterial protein synthesis. For example, tetracycline blocks the A site on the bacterial ribosome, and chloramphenicol blocks peptidyl transfer. What specific effect would you expect each of these antibiotics to have on protein synthesis? Tetracycline would directly affect:
    a. tRNA binding to the ribosome
    b. ribosome assembly
    c. growth of the protein chain

    Chloramphenicol would directly affect
    a. tRNA binding to the ribosome
    b. ribosome assembly
    c. growth of the protein chain

REVIEW QUESTIONS

4. The AUC and AUA codons in mRNA both specify isoleucine. What feature of the genetic code explains this?
    a. complementarity
    b. nonsense codons
    c. universality
    d. degeneracy

5. How many nucleotides are in 12 mRNA codons?
    a. 12
    b. 24
    c. 36
    d. 48

6. Which event contradicts the central dogma of molecular biology?
    a. Poly-A polymerase enzymes process mRNA in the nucleus.
    b. Endonuclease enzymes splice out and repair damaged DNA.
    c. Scientists use reverse transcriptase enzymes to make DNA from RNA.
    d. Codons specifying amino acids are degenerate and universal.

7. Which subunit of the *E. coli* polymerase confers specificity to transcription?
    a. α
    b. β
    c. β'
    d. σ

8. The -10 and -35 regions of prokaryotic promoters are called consensus sequences because _____.
    a. they are identical in all bacterial species
    b. they are similar in all bacterial species
    c. they exist in all organisms
    d. they have the same function in all organisms

9. Three different bacteria species have the following consensus sequences upstream of a conserved gene.

	Species A	Species B	Species C
-10	TAATAAT	TTTAAT	TATATT
-35	TTGACA	TTGGCC	TTGAAA

Table 15.2

Order the bacteria from most to least efficient initiation of gene transcription.
    a. A > B > C
    b. B > C > A
    c. C > B > A
    d. A > C > B

10. Which feature of promoters can be found in both prokaryotes and eukaryotes?
    a. GC box
    b. TATA box
    c. octamer box
    d. -10 and -35 sequences

11. What transcripts will be most affected by low levels of α-amanitin?
    a. 18S and 28S rRNAs
    b. pre-mRNAs
    c. 5S rRNAs and tRNAs
    d. other small nuclear RNAs

12. How do enhancers and promoters differ?
    a. Enhancers bind transcription factors to silence gene expression, while promoters activate transcription.
    b. Enhancers increase the efficiency of gene expression, but are not essential for transcription. Promoter recognition is essential to transcription initiation.
    c. Promoters bind transcription factors to increase the efficiency of transcription. Enhancers bind RNA polymerases to initiate transcription.
    d. There is no difference. Both are transcription factor-binding sequences in DNA.

13. Which pre-mRNA processing step is important for initiating translation?
    a. poly-A tail
    b. RNA editing
    c. splicing
    d. 7-methylguanosine cap

14. What processing step enhances the stability of pre-tRNAs and pre-rRNAs?
    a. methylation
    b. nucleotide modification
    c. cleavage
    d. splicing

15. A scientist identifies a pre-mRNA with the following structure.

    What is the predicted size of the corresponding mature mRNA in base pairs (bp), excluding the 5' cap and 3' poly-A tail?
    a. 220bp
    b. 295bp
    c. 140bp
    d. 435bp

16. The RNA components of ribosomes are synthesized in the _____.
    a. cytoplasm
    b. nucleus
    c. nucleolus
    d. endoplasmic reticulum

17. In any given species, there are at least how many types of aminoacyl tRNA synthetases?
    a. 20
    b. 40
    c. 100
    d. 200

18. A scientist introduces a mutation that makes the 60S ribosomal subunit nonfunctional in a human cell line. What would be the predicted effect on translation?
    a. Translation stalls after the initiation AUG codon is identified.
    b. The ribosome cannot catalyze the formation of peptide bonds between the tRNAs in the A and P sites.
    c. The ribosome cannot interact with mRNAs.
    d. tRNAs cannot exit the E site of the ribosome.

CRITICAL THINKING QUESTIONS

19. Imagine if there were 200 commonly occurring amino acids instead of 20. Given what you know about the genetic code, what would be the shortest possible codon length? Explain.

20. Discuss how degeneracy of the genetic code makes cells more robust to mutations.

21. A scientist sequencing mRNA identifies the following strand: CUAUGUGUCGUAACAGCCGAUGACCCG What is the sequence of the amino acid chain this mRNA makes when it is translated?

22. If mRNA is complementary to the DNA template strand and the DNA template strand is complementary to the DNA nontemplate strand, then why are base sequences of mRNA and the DNA nontemplate strand not identical? Could they ever be?

23. In your own words, describe the difference between rho-dependent and rho-independent termination of transcription in prokaryotes.

24. A fragment of bacterial DNA reads: 3' –TACCTATAATCTCAATTGATAGAAGCACTCTAC– 5' Assuming that this fragment is the template strand, what is the sequence of mRNA that would be transcribed? (Hint: Be sure to identify the initiation site.)

25. A scientist observes that a cell has an RNA polymerase deficiency that prevents it from making proteins. Describe three additional observations that would together support the conclusion that a defect in RNA polymerase I activity, and not problems with the other polymerases, causes the defect.

26. Chronic lymphocytic leukemia patients often harbor nonsense mutations in their spliceosome machinery. Describe how this mutation of the spliceosome would change the final location and sequence of a pre-mRNA.

27. Transcribe and translate the following DNA sequence (nontemplate strand): 5'-ATGGCCGGTTATTAAGCA-3'

28. Explain how single nucleotide changes can have vastly different effects on protein function.

29. A normal mRNA that reads 5' – UGCCAUGGUAAUAACACAUGAGGCCUGAAC– 3' has an insertion mutation that changes the sequence to 5' -UGCCAUGG**U**UAAUAACACAUGAGGCCUGAAC– 3'. Translate the original mRNA and the mutated mRNA, and explain how insertion mutations can have dramatic effects on proteins. (Hint: Be sure to find the initiation site.)