

15

Genes and How They Work

Concept Outline

15.1 The Central Dogma traces the flow of gene-encoded information.

Cells Use RNA to Make Protein. The information in genes is expressed in two steps, first being transcribed into RNA, and the RNA then being translated into protein.

15.2 Genes encode information in three-nucleotide code words.

The Genetic Code. The sequence of amino acids in a protein is encoded in the sequence of nucleotides in DNA, three nucleotides encoding an amino acid.

15.3 Genes are first transcribed, then translated.

Transcription. The enzyme RNA polymerase unwinds the DNA helix and synthesizes an RNA copy of one strand.

Translation. mRNA is translated by activating enzymes that select tRNAs to match amino acids. Proteins are synthesized on ribosomes, which provide a framework for the interaction of tRNA and mRNA.

15.4 Eukaryotic gene transcripts are spliced.

The Discovery of Introns. Eukaryotic genes contain extensive material that is not translated.

Differences between Bacterial and Eukaryotic Gene Expression. Gene expression is broadly similar in bacteria and eukaryotes, although it differs in some respects.

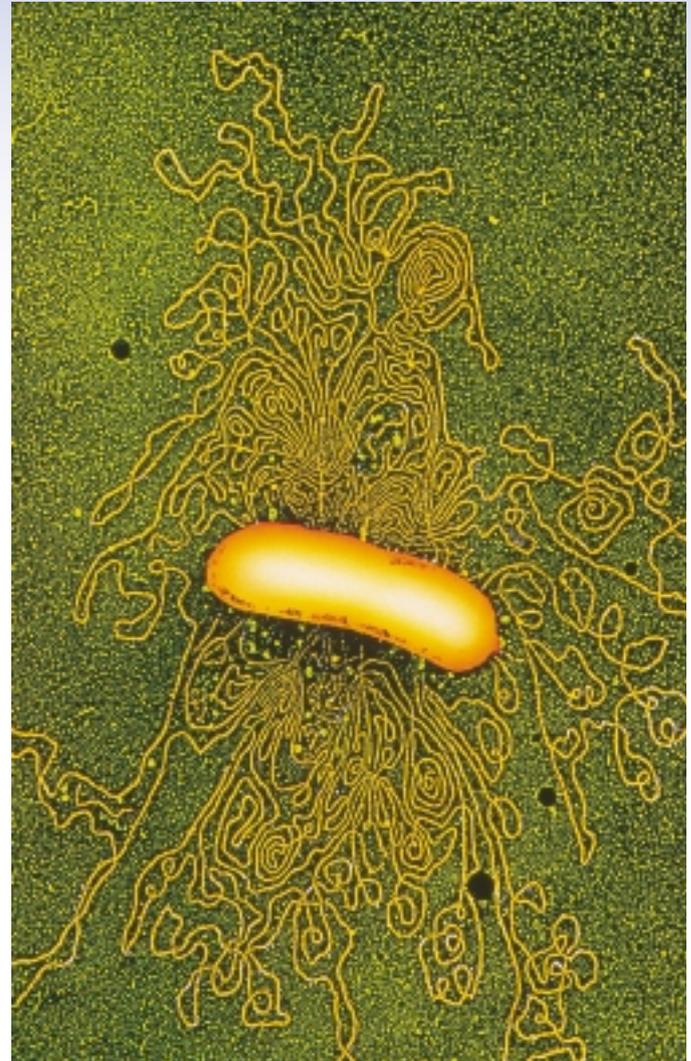


FIGURE 15.1

The unraveled chromosome of an *E. coli* bacterium. This complex tangle of DNA represents the full set of assembly instructions for the living organism *E. coli*.

Every cell in your body contains the hereditary instructions specifying that you will have arms rather than fins, hair rather than feathers, and two eyes rather than one. The color of your eyes, the texture of your fingernails, and all of the other traits you receive from your parents are recorded in the cells of your body. As we have seen, this information is contained in long molecules of DNA (figure 15.1). The essence of heredity is the ability of cells to use the information in their DNA to produce particular proteins, thereby affecting what the cells will be like. In that sense, proteins are the tools of heredity. In this chapter, we will examine how proteins are synthesized from the information in DNA, using both prokaryotes and eukaryotes as models.

15.1 The Central Dogma traces the flow of gene-encoded information.

Cells Use RNA to Make Protein

To find out how a eukaryotic cell uses its DNA to direct the production of particular proteins, you must first ask where in the cell the proteins are made. We can answer this question by placing cells in a medium containing radioactively labeled amino acids for a short time. The cells will take up the labeled amino acids and incorporate them into proteins. If we then look to see where in the cells radioactive proteins first appear, we will find that it is not in the nucleus, where the DNA is, but rather in the cytoplasm, on large RNA-protein aggregates called **ribosomes** (figure 15.2). These polypeptide-making factories are very complex, composed of several RNA molecules and over 50 different proteins (figure 15.3). Protein synthesis involves three different sites on the ribosome surface, called the P, A, and E sites, discussed later in this chapter.

Kinds of RNA

The class of RNA found in ribosomes is called **ribosomal RNA (rRNA)**. During polypeptide synthesis, rRNA provides the site where polypeptides are assembled. In addition to rRNA, there are two other major classes of RNA in cells. **Transfer RNA (tRNA)** molecules both transport the amino acids to the ribosome for use in building the polypeptides and position each amino acid at the correct place on the elongating polypeptide chain (figure 15.4). Human cells contain about 45 different kinds of tRNA molecules. **Messenger RNA (mRNA)** molecules are long strands of RNA that are transcribed from DNA and that travel to the ribosomes to direct precisely *which* amino acids are assembled into polypeptides.

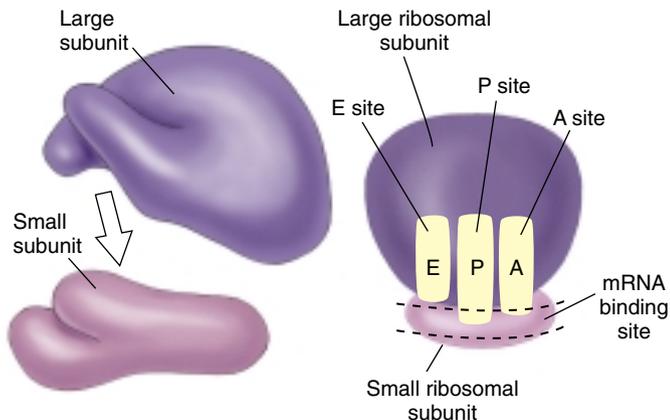


FIGURE 15.2

A ribosome is composed of two subunits. The smaller subunit fits into a depression on the surface of the larger one. The A, P, and E sites on the ribosome, discussed later in this chapter, play key roles in protein synthesis.

These RNA molecules, together with ribosomal proteins and certain enzymes, constitute a system that reads the genetic messages encoded by nucleotide sequences in the DNA and produces the polypeptides that those sequences specify. As we will see, biologists have also learned to read these messages. In so doing, they have learned a great deal about what genes are and how they are able to dictate what a protein will be like and when it will be made.

The Central Dogma

All organisms, from the simplest bacteria to ourselves, use the same basic mechanism of reading and expressing genes, so fundamental to life as we know it that it is often referred to as the “Central Dogma”: Information passes from the genes (DNA) to an RNA copy of the gene, and the RNA copy directs the sequential assembly of a chain of amino acids (figure 15.5). Said briefly,

DNA → RNA → protein

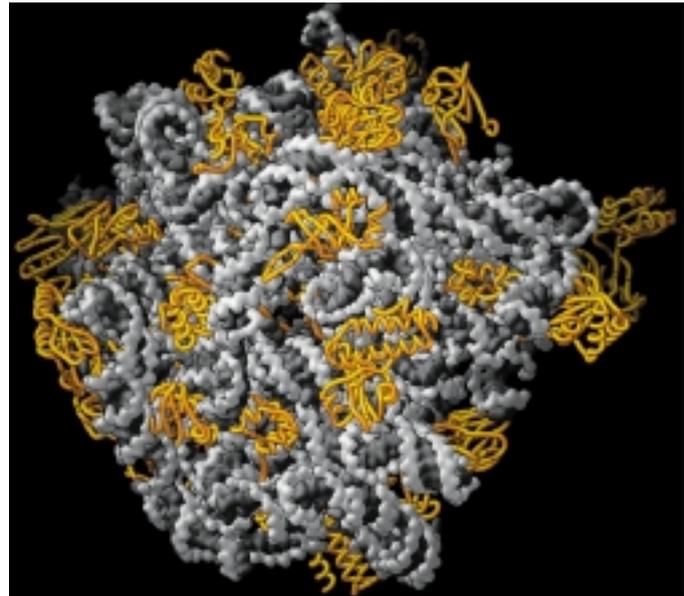


FIGURE 15.3

Ribosomes are very complex machines. The complete atomic structure of a bacterial large ribosomal subunit has been determined at 2.4 Å resolution. The RNA of the subunit is shown in gray and the proteins in gold. The subunit’s RNA is twisted into irregular shapes that fit together like a three-dimensional jigsaw puzzle. Proteins are abundant everywhere on its surface except where peptide bonds form and where it contacts the small subunit. The proteins stabilize the structure by interacting with adjacent RNA strands, often with folded extensions that reach into the subunit’s interior.

Transcription: An Overview

The first step of the Central Dogma is the transfer of information from DNA to RNA, which occurs when an mRNA copy of the gene is produced. Like all classes of RNA, mRNA is formed on a DNA template. Because the DNA sequence in the gene is transcribed into an RNA sequence, this stage is called **transcription**. Transcription is initiated when the enzyme **RNA polymerase** binds to a particular binding site called a **promoter** located at the beginning of a gene. Starting there, the RNA polymerase moves along the strand into the gene. As it encounters each DNA nucleotide, it adds the corresponding complementary RNA nucleotide to a growing mRNA strand. Thus, guanine (G), cytosine (C), thymine (T), and adenine (A) in the DNA would signal the addition of C, G, A, and uracil (U), respectively, to the mRNA.

When the RNA polymerase arrives at a transcriptional “stop” signal at the opposite end of the gene, it disengages from the DNA and releases the newly assembled RNA chain. This chain is a complementary transcript of the gene from which it was copied.

Translation: An Overview

The second step of the Central Dogma is the transfer of information from RNA to protein, which occurs when the information contained in the mRNA transcript is used to direct the sequence of amino acids during the synthesis of polypeptides by ribosomes. This process is called **translation** because the nucleotide sequence of the mRNA transcript is translated into an amino acid sequence in the polypeptide. Translation begins when an rRNA molecule within the ribosome recognizes and binds to a “start” sequence on the mRNA. The ribosome then moves along the mRNA molecule, three nucleotides at a time. Each group of three nucleotides is a codeword that specifies which amino acid will be added to the growing polypeptide chain. The ribosome continues in this fashion until it encounters a translational “stop” signal; then it disengages from the mRNA and releases the completed polypeptide.

The two steps of the Central Dogma, taken together, are a concise summary of the events involved in the expression of an active gene. Biologists refer to this process as **gene expression**.

The information encoded in genes is expressed in two phases: transcription, in which an RNA polymerase enzyme assembles an mRNA molecule whose nucleotide sequence is complementary to the DNA nucleotide sequence of the gene; and translation, in which a ribosome assembles a polypeptide, whose amino acid sequence is specified by the nucleotide sequence in the mRNA.

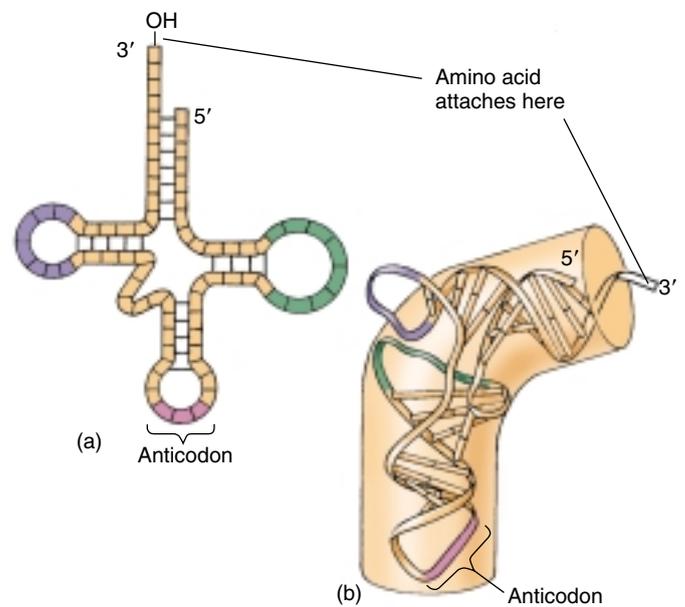


FIGURE 15.4

The structure of tRNA. (a) In the two-dimensional schematic, the three loops of tRNA are unfolded. Two of the loops bind to the ribosome during polypeptide synthesis, and the third loop contains the anticodon sequence, which is complementary to a three-base sequence on messenger RNA. Amino acids attach to the free, single-stranded —OH end. (b) In the three-dimensional structure, the loops of tRNA are folded.

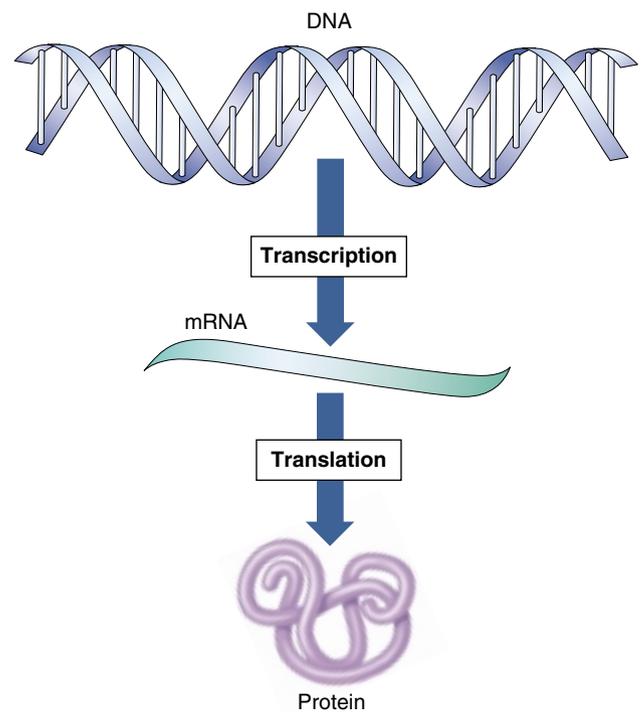


FIGURE 15.5

The Central Dogma of gene expression. DNA is transcribed to make mRNA, which is translated to make a protein.

15.2 Genes encode information in three-nucleotide code words.

The Genetic Code

The essential question of gene expression is, “How does the *order* of nucleotides in a DNA molecule encode the information that specifies the order of amino acids in a polypeptide?” The answer came in 1961, through an experiment led by Francis Crick. That experiment was so elegant and the result so critical to understanding the genetic code that we will describe it in detail.

Proving Code Words Have Only Three Letters

Crick and his colleagues reasoned that the genetic code most likely consisted of a series of blocks of information called **codons**, each corresponding to an amino acid in the encoded protein. They further hypothesized that the information within one codon was probably a sequence of three nucleotides specifying a particular amino acid. They arrived at the number three, because a two-nucleotide codon would not yield enough combinations to code for the 20 different amino acids that commonly occur in proteins. With four DNA nucleotides (G, C, T, and A), only 4², or 16, different pairs of nucleotides could be formed. However, these same nucleotides can be arranged in 4³, or 64, different combinations of three, more than enough to code for the 20 amino acids.

In theory, the codons in a gene could lie immediately adjacent to each other, forming a continuous sequence of transcribed nucleotides. Alternatively, the sequence could be punctuated with untranscribed nucleotides between the codons, like the spaces that separate the words in this sentence. It was important to determine which method cells employ because these two ways of transcribing DNA imply different translating processes.

To choose between these alternative mechanisms, Crick and his colleagues used a chemical to delete one, two, or three nucleotides from a viral DNA molecule and then asked whether a gene downstream of the deletions was transcribed correctly. When they made a single deletion or two deletions near each other, the **reading frame** of the genetic message shifted, and the downstream gene was transcribed as nonsense. However, when they made three deletions, the correct reading frame was restored, and the sequences downstream were transcribed correctly. They obtained the same results when they made additions to the DNA consisting of one, two, or three nucleotides. As shown in figure 15.6, these results could not have been obtained if the codons were punctuated by untranscribed nucleotides. Thus, Crick and his colleagues concluded that the genetic code is read in increments consisting of three nucleotides (in other words, it is a **triplet code**) and that reading occurs continuously without punctuation between the three-nucleotide units.

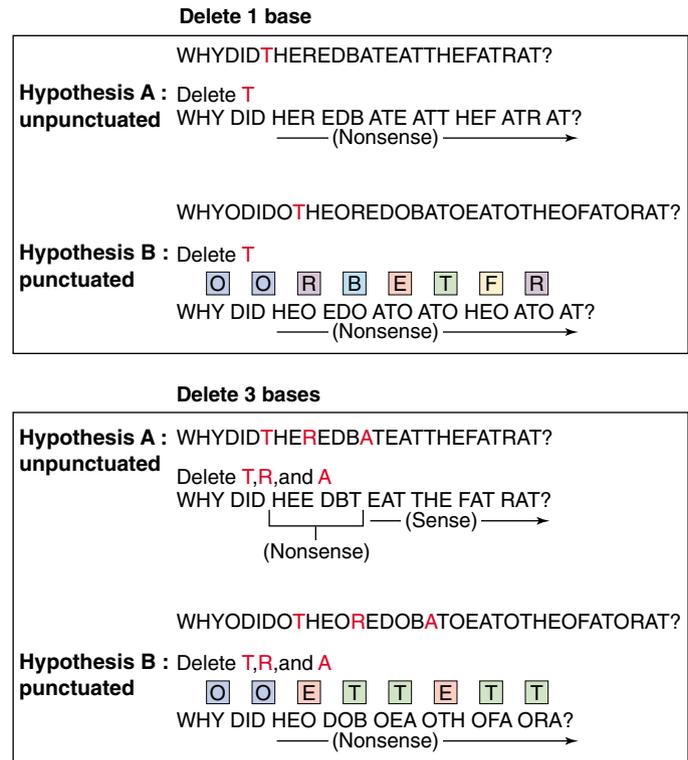


FIGURE 15.6

Using frame-shift alterations of DNA to determine if the genetic code is punctuated. The hypothetical genetic message presented here is “Why did the red bat eat the fat rat?” Under hypothesis B, which proposes that the message is punctuated, the three-letter words are separated by nucleotides that are not read (indicated by the letter “O”).

Breaking the Genetic Code

Within a year of Crick’s experiment, other researchers succeeded in determining the amino acids specified by particular three-nucleotide units. Marshall Nirenberg discovered in 1961 that adding the synthetic mRNA molecule polyU (an RNA molecule consisting of a string of uracil nucleotides) to cell-free systems resulted in the production of the polypeptide polyphenylalanine (a string of phenylalanine amino acids). Therefore, one of the three-nucleotide sequences specifying phenylalanine is UUU. In 1964, Nirenberg and Philip Leder developed a powerful **triplet binding assay** in which a specific triplet was tested to see which radioactive amino acid (complexed to tRNA) it would bind. Some 47 of the 64 possible triplets gave unambiguous results. Har Gobind Khorana decoded the remaining 17 triplets by constructing artificial mRNA molecules of defined sequence and examining what polypeptides they directed. In these ways, all 64 possible three-nucleotide sequences were tested, and the full genetic code was determined (table 15.1).

Table 15.1 The Genetic Code

First Letter	Second Letter								Third Letter
	U		C		A		G		
U	UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	Leucine	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG	Stop	UGG	Tryptophan	G
C	CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	CGA	A		
	CUG		CCG		CAG	CGG	G		
A	AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine	U
	AUC		ACC		AAC		AGC		C
	AUA	Methionine; Start	ACA		AAA	Lysine	AGA	Arginine	A
	AUG		ACG		AAG		AGG		G
G	GUU	Valine	GCU	Alanine	GAU	Aspartate	GGU	Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glutamate	GGA		A
	GUG		GCG		GAG		GGG		G

A codon consists of three nucleotides read in the sequence shown. For example, ACU codes for threonine. The first letter, A, is in the First Letter column; the second letter, C, is in the Second Letter column; and the third letter, U, is in the Third Letter column. Each of the mRNA codons is recognized by a corresponding anticodon sequence on a tRNA molecule. Some tRNA molecules recognize more than one codon in mRNA, but they always code for the same amino acid. In fact, most amino acids are specified by more than one codon. For example, threonine is specified by four codons, which differ only in the third nucleotide (ACU, ACC, ACA, and ACG).

The Code Is Practically Universal

The genetic code is the same in almost all organisms. For example, the codon AGA specifies the amino acid arginine in bacteria, in humans, and in all other organisms whose genetic code has been studied. The universality of the genetic code is among the strongest evidence that all living things share a common evolutionary heritage. Because the code is universal, genes transcribed from one organism can be translated in another; the mRNA is fully able to dictate a functionally active protein. Similarly, genes can be transferred from one organism to another and be successfully transcribed and translated in their new host. This universality of gene expression is central to many of the advances of genetic engineering. Many commercial products such as the insulin used to treat diabetes are now manufactured by placing human genes into bacteria, which then serve as tiny factories to turn out prodigious quantities of insulin.

But Not Quite

In 1979, investigators began to determine the complete nucleotide sequences of the mitochondrial genomes in

humans, cattle, and mice. It came as something of a shock when these investigators learned that the genetic code used by these mammalian mitochondria was not quite the same as the “universal code” that has become so familiar to biologists. In the mitochondrial genomes, what should have been a “stop” codon, UGA, was instead read as the amino acid tryptophan; AUA was read as methionine rather than isoleucine; and AGA and AGG were read as “stop” rather than arginine. Furthermore, minor differences from the universal code have also been found in the genomes of chloroplasts and ciliates (certain types of protists).

Thus, it appears that the genetic code is not quite universal. Some time ago, presumably after they began their endosymbiotic existence, mitochondria and chloroplasts began to read the code differently, particularly the portion of the code associated with “stop” signals.

Within genes that encode proteins, the nucleotide sequence of DNA is read in blocks of three consecutive nucleotides, without punctuation between the blocks. Each block, or codon, codes for one amino acid.

15.3 Genes are first transcribed, then translated.

Transcription

The first step in gene expression is the production of an RNA copy of the DNA sequence encoding the gene, a process called **transcription**. To understand the mechanism behind the transcription process, it is useful to focus first on RNA polymerase, the remarkable enzyme responsible for carrying it out (figure 15.7).

RNA Polymerase

RNA polymerase is best understood in bacteria. Bacterial RNA polymerase is very large and complex, consisting of five subunits: two α subunits bind regulatory proteins, a β' subunit binds the DNA template, a β subunit binds RNA nucleoside subunits, and a σ subunit recognizes the promoter and initiates synthesis. Only one of the two strands of DNA, called the **template strand**, is transcribed. The RNA transcript's sequence is complementary to the template strand. The strand of DNA that is not transcribed is called the **coding strand**. It has the same sequence as the RNA transcript, except T takes the place of U. The coding strand is also known as the sense (+) strand, and the template strand as the antisense (-) strand.

In both bacteria and eukaryotes, the polymerase adds ribonucleotides to the growing 3' end of an RNA chain. No primer is needed, and synthesis proceeds in the 5' \rightarrow 3' direction. Bacteria contain only one RNA polymerase enzyme, while eukaryotes have three different RNA polymerases: RNA polymerase I synthesizes rRNA in the nucleolus; RNA polymerase II synthesizes mRNA; and RNA polymerase III synthesizes tRNA.

Promoter

Transcription starts at RNA polymerase binding sites called **promoters** on the DNA template strand. A promoter is a short sequence that is not itself transcribed by the polymerase that binds to it. Striking similarities are evident in the sequences of different promoters. For example, two six-base sequences are common to many bacterial promoters, a TTGACA sequence called the **-35 sequence**, located 35 nucleotides upstream of the position where transcription actually starts, and a TATAAT sequence called the **-10 sequence**, located 10 nucleotides upstream of the start site. In eukaryotic DNA, the sequence TATAAA, called the **TATA box**, is located at -25 and is very similar to the prokaryotic -10 sequence but is farther from the start site.

Promoters differ widely in efficiency. Strong promoters cause frequent initiations of transcription, as often as every 2 seconds in some bacteria. Weak promoters may tran-

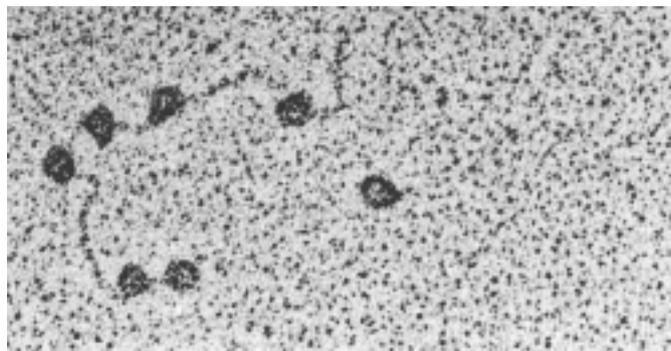


FIGURE 15.7
RNA polymerase. In this electron micrograph, the dark circles are RNA polymerase molecules bound to several promoter sites on bacterial virus DNA.

scribe only once every 10 minutes. Most strong promoters have unaltered -35 and -10 sequences, while weak promoters often have substitutions within these sites.

Initiation

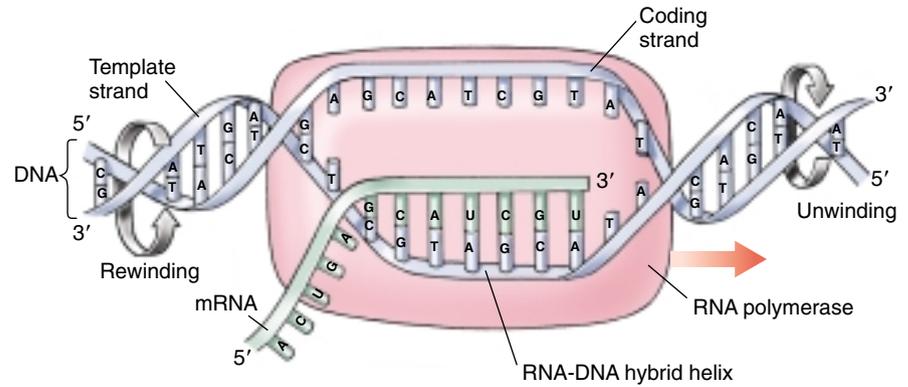
The binding of RNA polymerase to the promoter is the first step in gene transcription. In bacteria, a subunit of RNA polymerase called σ (**sigma**) recognizes the -10 sequence in the promoter and binds RNA polymerase there. Importantly, this subunit can detect the -10 sequence without unwinding the DNA double helix. In eukaryotes, the -25 sequence plays a similar role in initiating transcription, as it is the binding site for a key protein factor. Other eukaryotic factors then bind one after another, assembling a large and complicated **transcription complex**. The eukaryotic transcription complex is described in detail in the following chapter.

Once bound to the promoter, the RNA polymerase begins to unwind the DNA helix. Measurements indicate that bacterial RNA polymerase unwinds a segment approximately 17 base-pairs long, nearly two turns of the DNA double helix. This sets the stage for the assembly of the RNA chain.

Elongation

The transcription of the RNA chain usually starts with ATP or GTP. One of these forms the 5' end of the chain, which grows in the 5' \rightarrow 3' direction as ribonucleotides are added. Unlike DNA synthesis, a primer is not required. The region containing the RNA polymerase, DNA, and growing RNA transcript is called the **transcription bubble** because it contains a locally unwound "bubble" of DNA (figure 15.8). Within the bubble, the first 12 bases of the

FIGURE 15.8
Model of a transcription bubble. The DNA duplex unwinds as it enters the RNA polymerase complex and rewinds as it leaves. One of the strands of DNA functions as a template, and nucleotide building blocks are assembled into RNA from this template.



newly synthesized RNA strand temporarily form a helix with the template DNA strand. Corresponding to not quite one turn of the helix, this stabilizes the positioning of the 3' end of the RNA so it can interact with an incoming ribonucleotide. The RNA-DNA hybrid helix rotates each time a nucleotide is added so that the 3' end of the RNA stays at the catalytic site.

The transcription bubble moves down the DNA at a constant rate, about 50 nucleotides per second, leaving the growing RNA strand protruding from the bubble. After the transcription bubble passes, the now transcribed DNA is rewound as it leaves the bubble.

Unlike DNA polymerase, RNA polymerase has no proofreading capability. Transcription thus produces many more copying errors than replication. These mistakes, however, are not transmitted to progeny. Most genes are transcribed many times, so a few faulty copies are not harmful.

Termination

At the end of a gene are “stop” sequences that cause the formation of phosphodiester bonds to cease, the RNA-DNA hybrid within the transcription bubble to dissociate, the RNA polymerase to release the DNA, and the DNA within the transcription bubble to rewind. The simplest stop signal is a series of GC base-pairs followed by a series of AT base-pairs. The RNA transcript of this stop region forms a GC hairpin (figure 15.9), followed by four or more U ribonucleotides. How does this structure terminate transcription? The hairpin causes the RNA polymerase to pause immediately after the polymerase has synthesized it, placing the polymerase directly over the run of four uracils. The pairing of U with DNA's A is the weakest of the four hybrid base-pairs and is not strong enough to hold the hybrid strands together during the long pause. Instead, the RNA strand dissociates from the DNA within the transcription bubble, and transcription stops. A variety of protein factors aid hairpin loops in terminating transcription of particular genes.

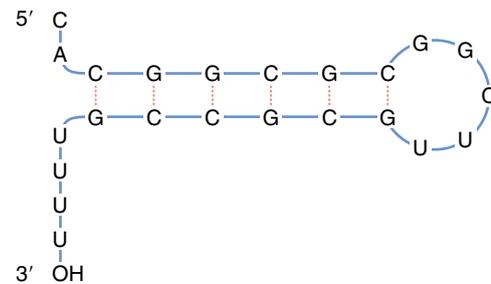


FIGURE 15.9
A GC hairpin. This structure stops gene transcription.

Posttranscriptional Modifications

In eukaryotes, every mRNA transcript must travel a long journey out from the nucleus into the cytoplasm before it can be translated. Eukaryotic mRNA transcripts are modified in several ways to aid this journey:

5' caps. Transcripts usually begin with A or G, and, in eukaryotes, the terminal phosphate of the 5' A or G is removed, and then a very unusual 5'-5' linkage forms with GTP. Called a **5' cap**, this structure protects the 5' end of the RNA template from nucleases and phosphatases during its long journey through the cytoplasm. Without these caps, RNA transcripts are rapidly degraded.

3' poly-A tails. The 3' end of eukaryotic transcript is cleaved off at a specific site, often containing the sequence AAUAAA. A special poly-A polymerase enzyme then adds about 250 A ribonucleotides to the 3' end of the transcript. Called a **3' poly-A tail**, this long string of As protects the transcript from degradation by nucleases. It also appears to make the transcript a better template for protein synthesis.

Transcription is carried out by the enzyme RNA polymerase, aided in eukaryotes by many other proteins.

Translation

In prokaryotes, translation begins when the initial portion of an mRNA molecule binds to an rRNA molecule in a ribosome. The mRNA lies on the ribosome in such a way that only one of its codons is exposed at the polypeptide-making site at any time. A tRNA molecule possessing the complementary three-nucleotide sequence, or anticodon, binds to the exposed codon on the mRNA.

Because this tRNA molecule carries a particular amino acid, that amino acid and no other is added to the polypeptide in that position. As the mRNA molecule moves through the ribosome, successive codons on the mRNA are exposed, and a series of tRNA molecules bind one after another to the exposed codons. Each of these tRNA molecules carries an attached amino acid, which it adds to the end of the growing polypeptide chain (figure 15.10).

There are about 45 different kinds of tRNA molecules. Why are there 45 and not 64 tRNAs (one for each codon)? Because the third base-pair of a tRNA anticodon allows some “wobble,” some tRNAs recognize more than one codon.

How do particular amino acids become associated with particular tRNA molecules? The key translation step, which pairs the three-nucleotide sequences with appropriate amino acids, is carried out by a remarkable set of enzymes called activating enzymes.

Activating Enzymes

Particular tRNA molecules become attached to specific amino acids through the action of activating enzymes called **aminoacyl-tRNA synthetases**, one of which exists for each of the 20 common amino acids (figure 15.11). Therefore, these enzymes must correspond to specific anticodon sequences on a tRNA molecule as well as particular amino acids. Some activating enzymes correspond to only one anticodon and thus only one tRNA molecule. Others recognize two, three, four, or six different tRNA molecules, each with a different anticodon but coding for the same amino acid (see table 15.1). If one considers the nucleotide sequence of mRNA a coded message, then the 20 activating enzymes are responsible for decoding that message.

“Start” and “Stop” Signals

There is no tRNA with an anticodon complementary to three of the 64 codons: UAA, UAG, and UGA. These codons, called **nonsense codons**, serve as “stop” signals in the mRNA message, marking the end of a polypeptide. The “start” signal that marks the beginning of a polypeptide within an mRNA message is the codon AUG, which also encodes the amino acid methionine. The ribosome will usually use the first AUG that it encounters in the mRNA to signal the start of translation.

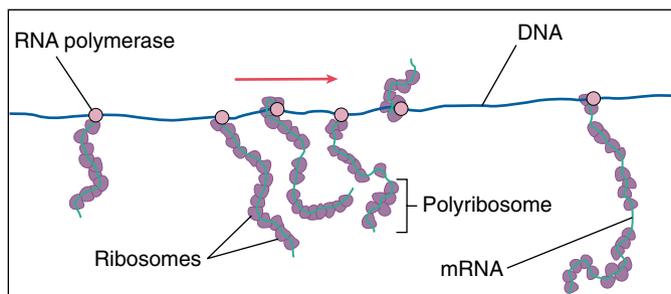
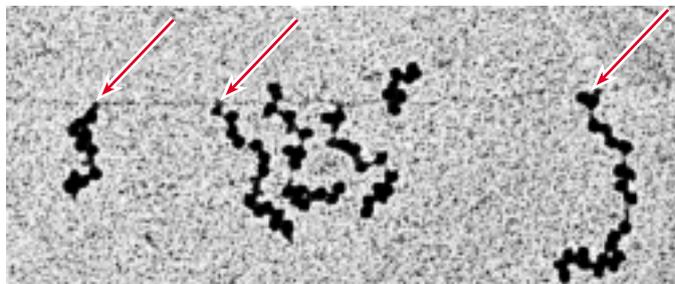


FIGURE 15.10

Translation in action. Bacteria have no nucleus and hence no membrane barrier between the DNA and the cytoplasm. In this electron micrograph of genes being transcribed in the bacterium *Escherichia coli*, you can see every stage of the process. The arrows point to RNA polymerase enzymes. From each mRNA molecule dangling from the DNA, a series of ribosomes is assembling polypeptides. These clumps of ribosomes are sometimes called “polyribosomes.”

Initiation

In prokaryotes, polypeptide synthesis begins with the formation of an **initiation complex**. First, a tRNA molecule carrying a chemically modified methionine called *N*-formylmethionine ($\text{tRNA}^{\text{fMet}}$) binds to the small ribosomal subunit. Proteins called **initiation factors** position the $\text{tRNA}^{\text{fMet}}$ on the ribosomal surface at the *P site* (for peptidyl), where peptide bonds will form. Nearby, two other sites will form: the *A site* (for aminoacyl), where successive amino acid-bearing tRNAs will bind, and the *E site* (for exit), where empty tRNAs will exit the ribosome (figure 15.12). This initiation complex, guided by another initiation factor, then binds to the anticodon AUG on the mRNA. Proper positioning of the mRNA is critical because it determines the reading frame—that is, which groups of three nucleotides will be read as codons. Moreover, the complex must bind to the beginning of the mRNA molecule, so that all of the transcribed gene will be translated. In bacteria, the beginning of each mRNA molecule is marked by a *leader sequence* complementary to one of the rRNA molecules on the ribosome. This complementarity ensures that the mRNA is read from the beginning. Bacteria often include several genes within a single mRNA transcript (polycistronic mRNA), while each eukaryotic gene is transcribed on a separate mRNA (monocistronic mRNA).

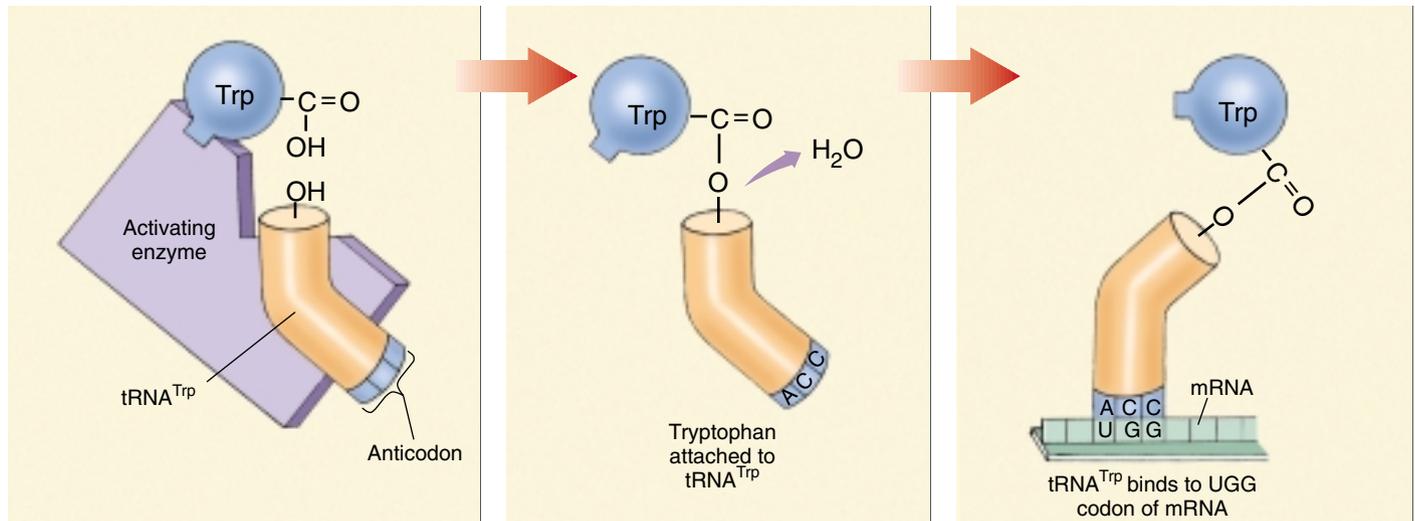


FIGURE 15.11

Activating enzymes “read” the genetic code. Each kind of activating enzyme recognizes and binds to a specific amino acid, such as tryptophan; it also recognizes and binds to the tRNA molecules with anticodons specifying that amino acid, such as ACC for tryptophan. In this way, activating enzymes link the tRNA molecules to specific amino acids.

Initiation in eukaryotes is similar, although it differs in two important ways. First, in eukaryotes, the initiating amino acid is methionine rather than *N*-formylmethionine. Second, the initiation complex is far more complicated than in bacteria, containing nine or more protein factors, many consisting of several subunits. Eukaryotic initiation complexes are discussed in detail in the following chapter.

Elongation

After the initiation complex has formed, the large ribosome subunit binds, exposing the mRNA codon adjacent to the

initiating AUG codon, and so positioning it for interaction with another amino acid-bearing tRNA molecule. When a tRNA molecule with the appropriate anticodon appears, proteins called elongation factors assist in binding it to the exposed mRNA codon at the A site. When the second tRNA binds to the ribosome, it places its amino acid directly adjacent to the initial methionine, which is still attached to its tRNA molecule, which in turn is still bound to the ribosome. The two amino acids undergo a chemical reaction, catalyzed by *peptidyl transferase*, which releases the initial methionine from its tRNA and attaches it instead by a peptide bond to the second amino acid.

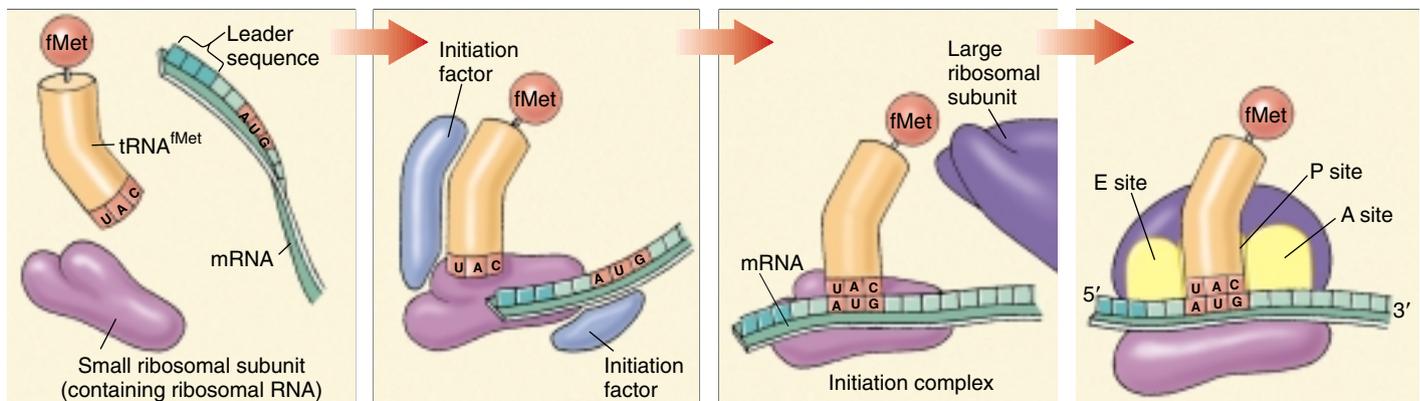


FIGURE 15.12

Formation of the initiation complex. In prokaryotes, proteins called initiation factors play key roles in positioning the small ribosomal subunit and the *N*-formylmethionine, or tRNA^{fMet}, molecule at the beginning of the mRNA. When the tRNA^{fMet} is positioned over the first AUG codon of the mRNA, the large ribosomal subunit binds, forming the P, A, and E sites where successive tRNA molecules bind to the ribosomes, and polypeptide synthesis begins.

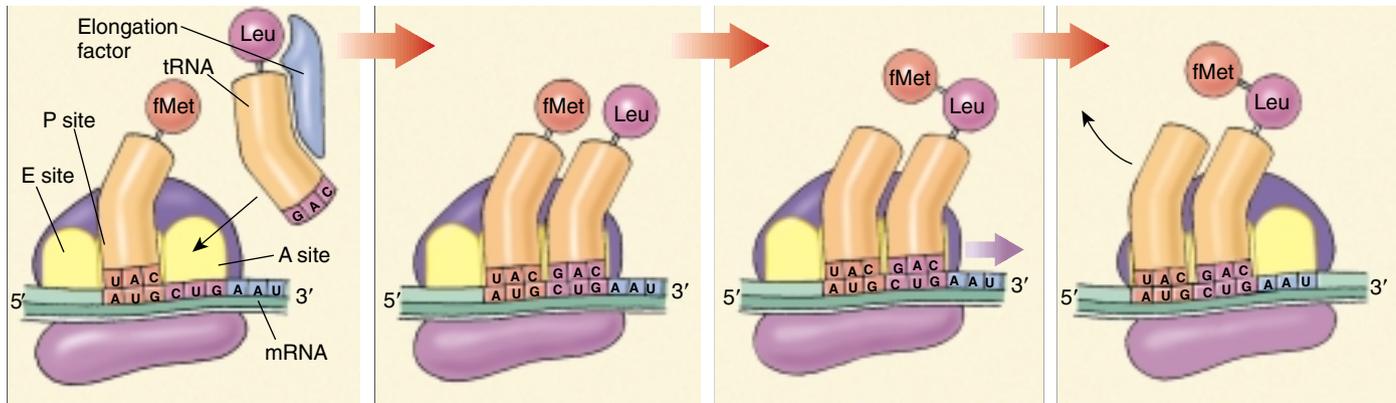


FIGURE 15.13

Translocation. The initiating tRNA^{fMet} in prokaryotes (tRNA^{fMet} in eukaryotes) occupies the P site, and a tRNA molecule with an anticodon complementary to the exposed mRNA codon binds at the A site. fMet is transferred to the incoming amino acid (Leu), as the ribosome moves three nucleotides to the right along the mRNA. The empty tRNA^{fMet} moves to the E site to exit the ribosome, the growing polypeptide chain moves to the P site, and the A site is again exposed and ready to bind the next amino acid–laden tRNA.

Translocation

In a process called **translocation** (figure 15.13), the ribosome now moves (translocates) three more nucleotides along the mRNA molecule in the 5′ → 3′ direction, guided by other elongation factors. This movement relocates the initial tRNA to the E site and ejects it from the ribosome, repositions the growing polypeptide chain (at this point containing two amino acids) to the P site, and exposes the next codon on the mRNA at the A site. When a tRNA molecule recognizing that codon appears, it binds to the codon at the A site, placing its amino acid adjacent to the growing chain. The chain then transfers to the new amino acid, and the entire process is repeated.

Termination

Elongation continues in this fashion until a chain-terminating nonsense codon is exposed (for example, UAA in figure 15.14). Nonsense codons do not bind to tRNA, but they are recognized by **release factors**, proteins that release the newly made polypeptide from the ribosome.

The first step in protein synthesis is the formation of an initiation complex. Each step of the ribosome’s progress exposes a codon, to which a tRNA molecule with the complementary anticodon binds. The amino acid carried by each tRNA molecule is added to the end of the growing polypeptide chain.

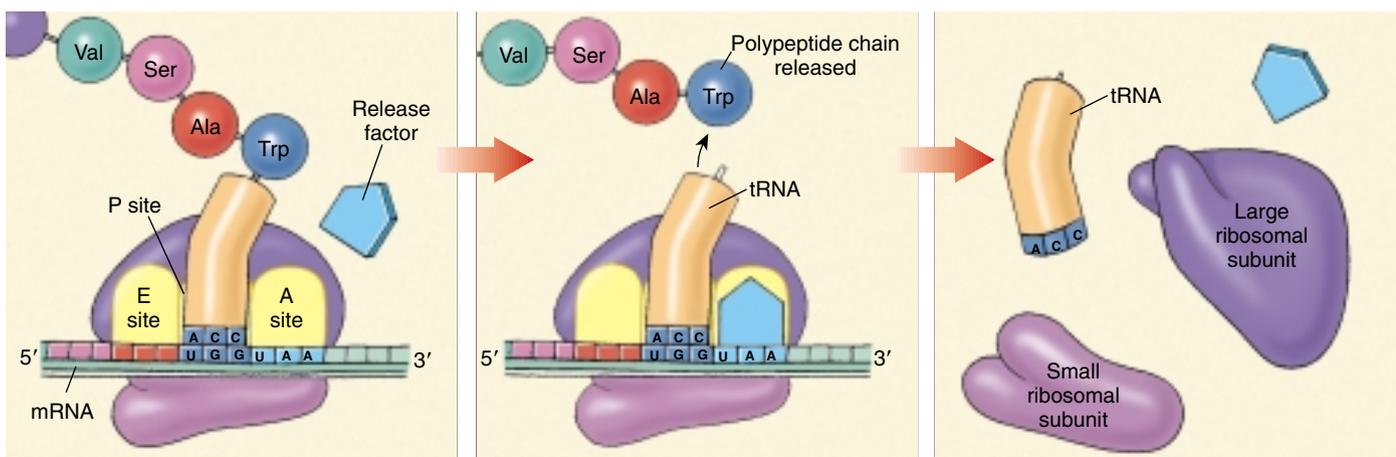


FIGURE 15.14

Termination of protein synthesis. There is no tRNA with an anticodon complementary to any of the three termination signal codons, such as the UAA nonsense codon illustrated here. When a ribosome encounters a termination codon, it therefore stops translocating. A specific release factor facilitates the release of the polypeptide chain by breaking the covalent bond that links the polypeptide to the P-site tRNA.

15.4 Eukaryotic gene transcripts are spliced.

The Discovery of Introns

While the mechanisms of protein synthesis are similar in bacteria and eukaryotes, they are not identical. One difference is of particular importance. Unlike bacterial genes, most eukaryotic genes are larger than they need to be to produce the polypeptides they code for. Such genes contain long sequences of nucleotides, known as **introns**, that do not code for any portion of the polypeptide specified by the gene. Introns are inserted between **exons**, much shorter sequences in the gene that do code for portions of the polypeptide.

In bacteria, virtually every nucleotide within a bacterial gene transcript is part of an amino acid-specifying codon. Scientists assumed for many years that this was true of all organisms. In the late 1970s, however, biologists were amazed to discover that many of the characteristics of prokaryotic gene expression did not apply to eukaryotes. In particular, they found that eukaryotic proteins are encoded by RNA segments that are excised from several locations along what is called the **primary RNA transcript** (or **primary transcript**) and then spliced together to form the mRNA that is eventually translated in the cytoplasm. The experiment that revealed this unexpected mode of gene expression consisted of several steps:

1. The mRNA transcribed from a particular gene was isolated and purified. For example, ovalbumin mRNA could be obtained fairly easily from unfertilized eggs.

2. Molecules of DNA complementary to the isolated mRNA were synthesized with the enzyme **reverse transcriptase**. These DNA molecules, which are called “copy” DNA (cDNA), had the same nucleotide sequence as the template strand of the gene that produced the mRNA.
3. With genetic engineering techniques (chapter 19), the portion of the nuclear DNA containing the gene that produced the mRNA was isolated. This procedure is referred to as *cloning* the gene in question.
4. Single-stranded forms of the cDNA and the nuclear DNA were mixed and allowed to pair with each other (to *hybridize*).

When the researchers examined the resulting hybrid DNA molecules with an electron microscope, they found that the DNA did not appear as a single duplex. Instead, they observed unpaired loops. In the case of the ovalbumin gene, they discovered seven loops, corresponding to sites where the nuclear DNA contained long nucleotide sequences not present in the cDNA. The conclusion was inescapable: nucleotide sequences must have been removed from the gene transcript before it appeared as cytoplasmic mRNA. These removed sequences are introns, and the remaining sequences are exons (figure 15.15). Because introns are excised from the RNA transcript before it is translated into protein, they do not affect the structure of the protein encoded by the gene in which they occur.

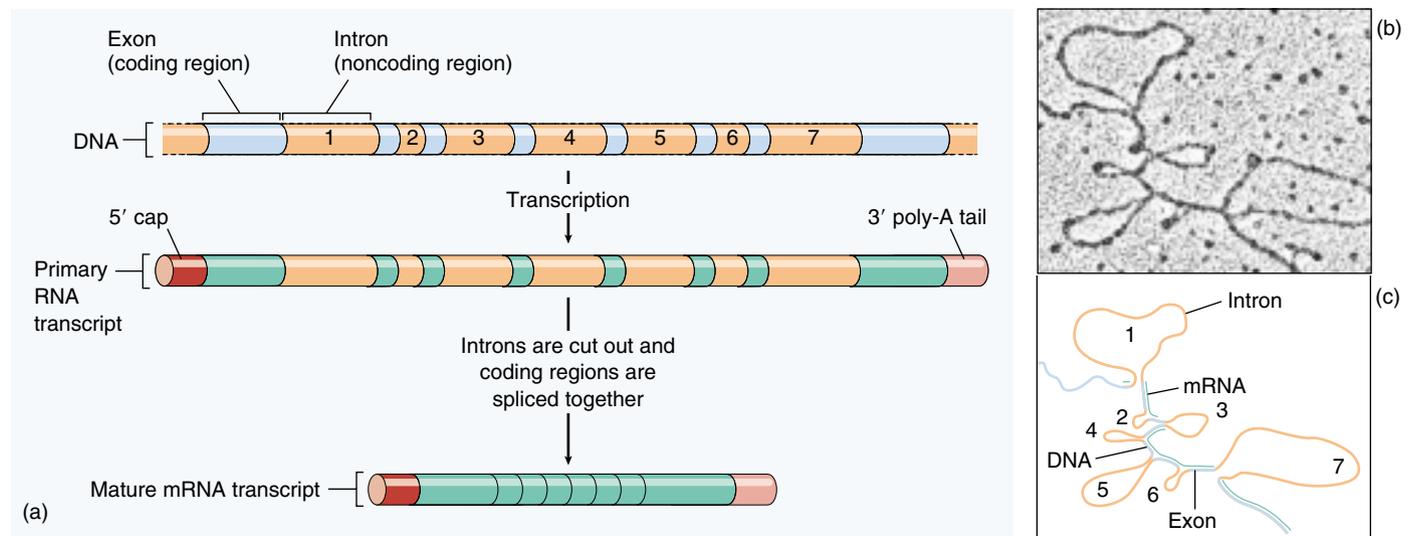


FIGURE 15.15

The eukaryotic gene that codes for ovalbumin in eggs contains introns. (a) The ovalbumin gene and its primary RNA transcript contain seven segments not present in the mRNA the ribosomes use to direct protein synthesis. Enzymes cut these segments (introns) out and splice together the remaining segments (exons). (b) The seven loops are the seven introns represented in the schematic drawing (c) of the mature mRNA transcript hybridized to DNA.

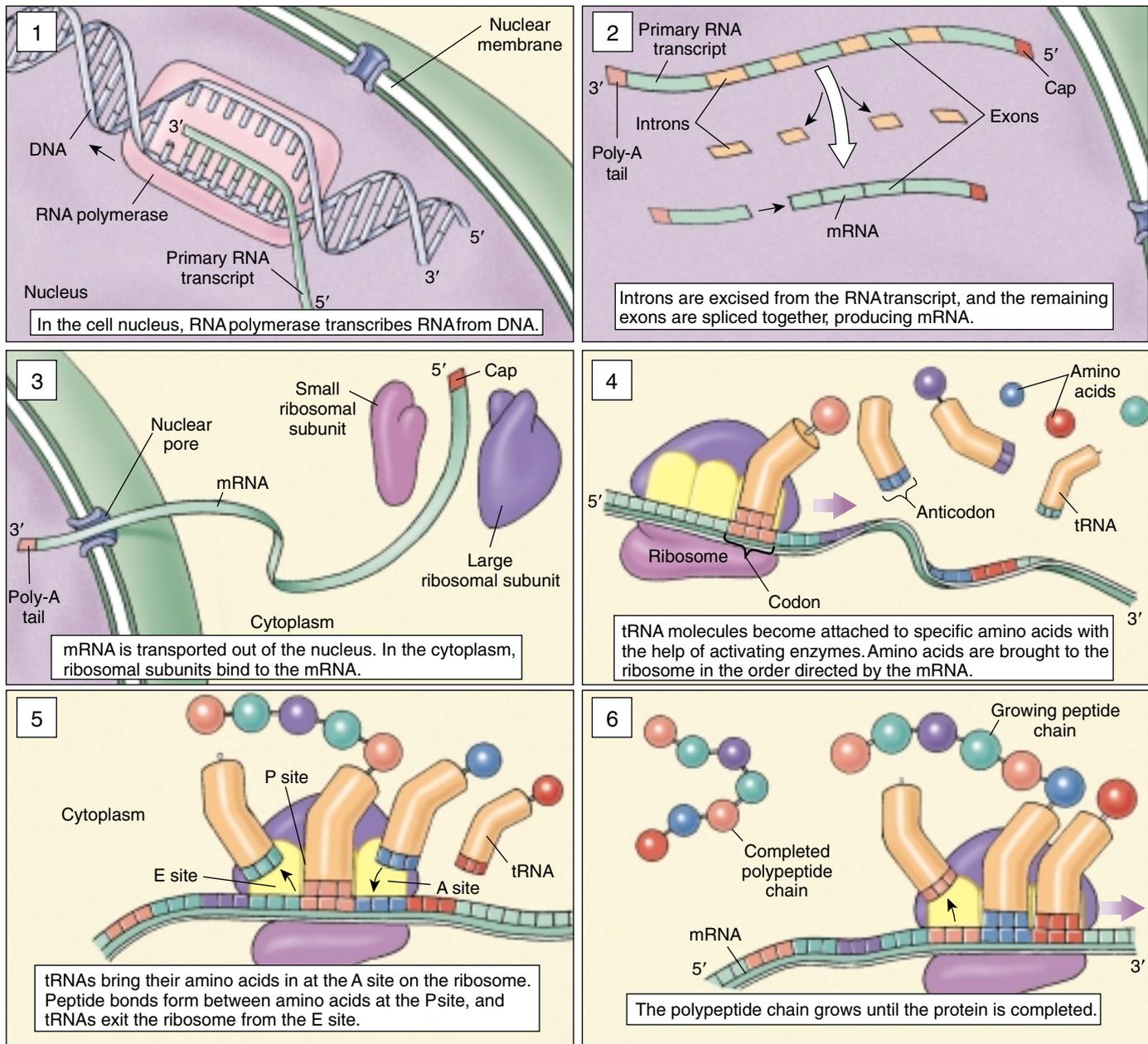


FIGURE 15.16
An overview of gene expression in eukaryotes.

RNA Splicing

When a gene is transcribed, the primary RNA transcript (that is, the gene copy as it is made by RNA polymerase, before any modification occurs) contains sequences complementary to the entire gene, including introns as well as exons. However, in a process called **RNA processing**, or **splicing**, the intron sequences are cut out of the primary transcript before it is used in polypeptide synthesis; therefore, those sequences are not translated. The remaining sequences, which correspond to the exons, are spliced together to form the final, “processed” mRNA molecule that

is translated. In a typical human gene, the introns can be 10 to 30 times larger than the exons. For example, even though only 432 nucleotides are required to encode the 144 amino acids of hemoglobin, there are actually 1356 nucleotides in the primary mRNA transcript of the hemoglobin gene. Figure 15.16 summarizes eukaryotic protein synthesis.

Much of a eukaryotic gene is not translated. Noncoding segments scattered throughout the gene are removed from the primary transcript before the mRNA is translated.

Differences between Bacterial and Eukaryotic Gene Expression

1. Most eukaryotic genes possess introns. With the exception of a few genes in the Archaeobacteria, prokaryotic genes lack introns (figure 15.17).
2. Individual bacterial mRNA molecules often contain transcripts of several genes. By placing genes with related functions on the same mRNA, bacteria coordinate the regulation of those functions. Eukaryotic mRNA molecules rarely contain transcripts of more than one gene. Regulation of eukaryotic gene expression is achieved in other ways.
3. Because eukaryotes possess a nucleus, their mRNA molecules must be completely formed and must pass across the nuclear membrane before they are translated. Bacteria, which lack nuclei, often begin translation of an mRNA molecule before its transcription is completed.

4. In bacteria, translation begins at an AUG codon preceded by a special nucleotide sequence. In eukaryotic cells, mRNA molecules are modified at the 5' leading end after transcription, adding a 5' cap, a methylated guanosine triphosphate. The cap initiates translation by binding the mRNA, usually at the first AUG, to the small ribosomal subunit.
5. Eukaryotic mRNA molecules are modified before they are translated: introns are cut out, and the remaining exons are spliced together; a 5' cap is added; and a 3' poly-A tail consisting of some 200 adenine (A) nucleotides is added. These modifications can delay the destruction of the mRNA by cellular enzymes.
6. The ribosomes of eukaryotes are a little larger than those of bacteria.

Gene expression is broadly similar in bacteria and eukaryotes, although it differs in some details.

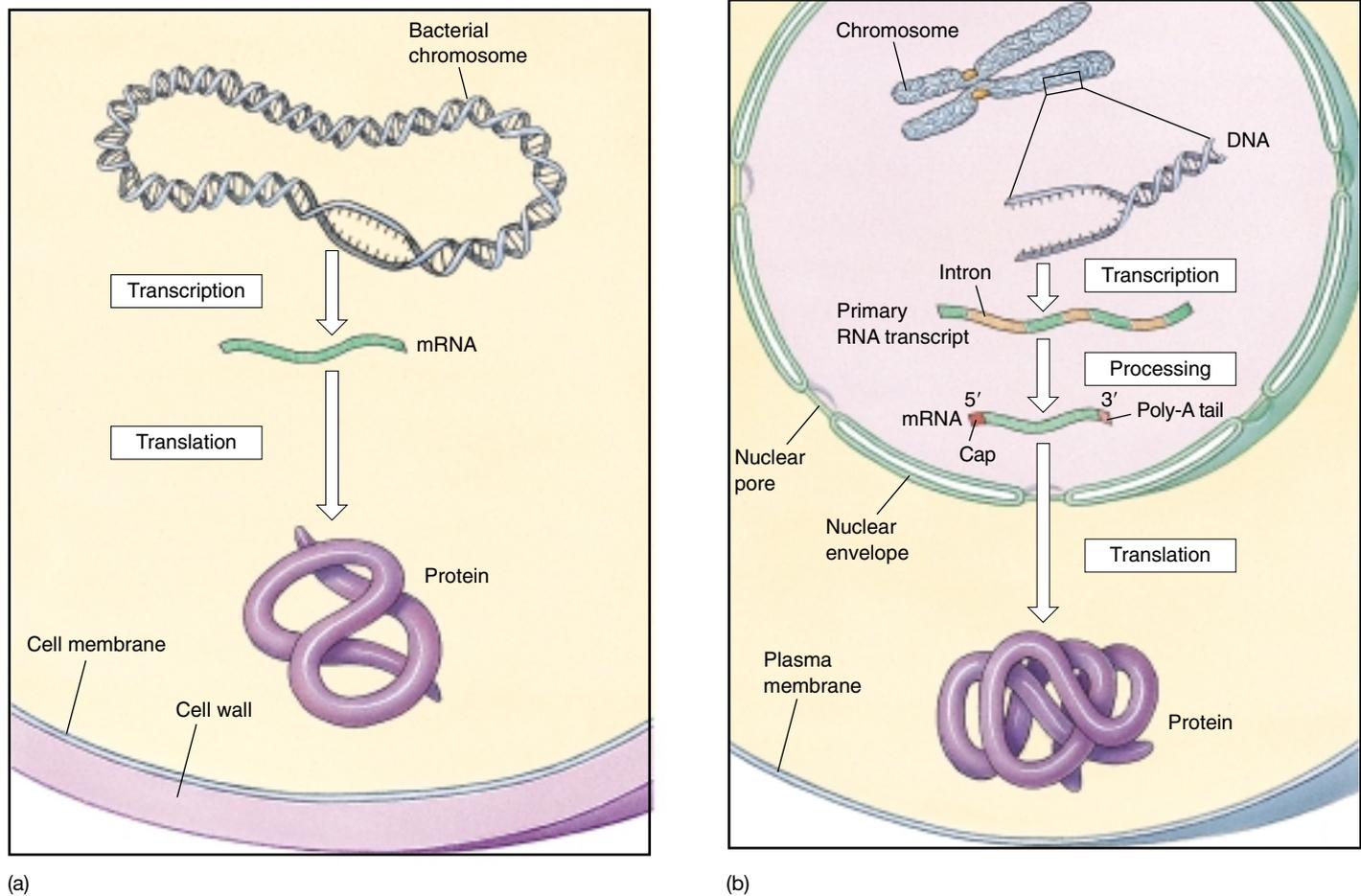


FIGURE 15.17

Gene information is processed differently in prokaryotes and eukaryotes. (a) Bacterial genes are transcribed into mRNA, which is translated immediately. Hence, the sequence of DNA nucleotides corresponds exactly to the sequence of amino acids in the encoded polypeptide. (b) Eukaryotic genes are typically different, containing long stretches of nucleotides called introns that do not correspond to amino acids within the encoded polypeptide. Introns are removed from the primary RNA transcript of the gene and a 5' cap and 3' poly-A tail are added before the mRNA directs the synthesis of the polypeptide.

**Summary****Questions****Media Resources****15.1 The Central Dogma traces the flow of gene-encoded information.**

- There are three principal kinds of RNA: messenger RNA (mRNA), transcripts of genes used to direct the assembly of amino acids into proteins; ribosomal RNA (rRNA), which combines with proteins to make up the ribosomes that carry out the assembly process; and transfer RNA (tRNA), molecules that transport the amino acids to the ribosome for assembly into proteins.

1. What are the three major classes of RNA? What is the function of each type?
2. What is the function of RNA polymerase in transcription? What determines where RNA polymerase begins and ends its function?



- Experiment: Jacob/Meselson/Brenner-Discovery of Messenger RNA (mRNA)



- Gene Activity

15.2 Genes encode information in three-nucleotide code words.

- The sequence of nucleotides in DNA encodes the sequence of amino acids in proteins. The mRNA transcribed from the DNA is read by ribosomes in increments of three nucleotides called codons.

3. How did Crick and his colleagues determine how many nucleotides are used to specify each amino acid? What is an anticodon?



- Experiment: Nirenberg/Khorana-Breaking the Genetic Code
- Experiment: The Genetic Code is Read in Three Bases at a Time

15.3 Genes are first transcribed, then translated.

- During transcription, the enzyme RNA polymerase manufactures mRNA molecules with nucleotide sequences complementary to particular segments of the DNA.
- During translation, the mRNA sequences direct the assembly of amino acids into proteins on cytoplasmic ribosomes.
- The information in a gene and in an mRNA molecule is read in three-nucleotide blocks called codons.
- On the ribosome, the mRNA molecule is positioned so that only one of its codons is exposed at any time.
- This exposure permits a tRNA molecule with the complementary base sequence (anticodon) to bind to it.
- Attached to the other end of the tRNA is an amino acid, which is added to the end of the growing polypeptide chain.

4. During protein synthesis, what mechanism ensures that only one amino acid is added to the growing polypeptide at a time? What mechanism ensures the correct amino acid is added at each position in the polypeptide?



- Transcription
- Translation
- Polyribosomes



- Transcription
- Translation



- Experiment: Chapeville-Proving the tRNA Hypothesis

5. How does an mRNA molecule specify where the polypeptide it encodes should begin? How does it specify where the polypeptide should end?

6. What roles do elongation factors play in translation?

15.4 Eukaryotic gene transcripts are spliced.

- Most eukaryotic genes contain noncoding sequences (introns) interspersed randomly between coding sequences (exons).
- The portions of an mRNA molecule corresponding to the introns are removed from the primary RNA transcript before the remainder is translated.

7. What is an intron? What is an exon? How is each involved in the mRNA molecule that is ultimately translated?



- Experiment: Chambon-Discovery of Introns