16

Control of Gene Expression

Concept Outline

16.1 Gene expression is controlled by regulating transcription.

An Overview of Transcriptional Control. In bacteria transcription is regulated by controlling access of RNA polymerase to the promoter in a flexible and reversible way; eukaryotes by contrast regulate many of their genes by turning them on and off in a more permanent fashion.

16.2 Regulatory proteins read DNA without unwinding it.

How to Read a Helix without Unwinding It. Regulatory proteins slide special segments called DNA-binding motifs along the major groove of the DNA helix, reading the sides of the bases.

Four Important DNA-Binding Motifs. DNA-binding proteins contain structural motifs such as the helix-turn-helix which fit into the major groove of the DNA helix.

16.3 Bacteria limit transcription by blocking RNA polymerase.

Controlling Transcription Initiation. Repressor proteins inhibit RNA polymerase's access to the promoter, while activators facilitate its binding.

16.4 Transcriptional control in eukaryotes operates at a distance.

Designing a Complex Gene Control System. Eukaryotic genes use a complex collection of transcription factors and enhancers to aid the polymerase in transcription.

The Effect of Chromosome Structure on Gene Regulation. The tight packaging of eukaryotic DNA into nucleosomes does not interfere with gene expression.

Posttranscriptional Control in Eukaryotes. Gene expression can be controlled at a variety of levels after transcription.

FIGURE 16.1
Chromosome puffs. In this chromosome of the fly *Drosophila melanogaster*, individual active genes can be visualized as "puffs" on the chromosomes. The RNA being transcribed from the DNA template has been radioactively labeled, and the dark specks indicate its position on the chromosome.

In an orchestra, all of the instruments do not play all the time; if they did, all they would produce is noise. Instead, a musical score determines which instruments in the orchestra play when. Similarly, all of the genes in an organism are not expressed at the same time, each gene producing the protein it encodes full tilt. Instead, different genes are expressed at different times, with a genetic score written in regulatory regions of the DNA determining which genes are active when (figure 16.1).

16.1 Gene expression is controlled by regulating transcription.

An Overview of Transcriptional Control

Control of gene expression is essential to all organisms. In bacteria, it allows the cell to take advantage of changing environmental conditions. In multicellular organisms, it is critical for directing development and maintaining homeostasis.

Regulating Promoter Access

One way to control transcription is to regulate the initiation of transcription. In order for a gene to be transcribed, RNA polymerase must have access to the DNA helix and must be capable of binding to the gene's **promoter,** a specific sequence of nucleotides at one end of the gene that tells the polymerase where to begin transcribing. How is the initiation of transcription regulated? Protein-binding nucleotide sequences on the DNA regulate the initiation of transcription by modulating the ability of RNA polymerase to bind to the promoter. These protein-binding sites are usually only 10 to 15 nucleotides in length (even a large regulatory protein has a "footprint," or binding area, of only about 20 nucleotides). Hundreds of these regulatory sequences have been characterized, and each provides a binding site for a specific protein able to recognize the sequence. Binding the protein to the regulatory sequence either *blocks* transcription by getting in the way of RNA polymerase, or *stimulates* transcription by facilitating the binding of RNA polymerase to the promoter.

Transcriptional Control in Prokaryotes

Control of gene expression is accomplished very differently in bacteria than in the cells of complex multicellular organisms. Bacterial cells have been shaped by evolution to grow and divide as rapidly as possible, enabling them to exploit transient resources. In bacteria, the primary function of gene control is to adjust the cell's activities to its immediate environment. Changes in gene expression alter which enzymes are present in the cell in response to the quantity and type of available nutrients and the amount of oxygen present. Almost all of these changes are fully reversible, allowing the cell to adjust its enzyme levels up or down as the environment changes.

Transcriptional Control in Eukaryotes

The cells of multicellular organisms, on the other hand, have been shaped by evolution to be protected from transient changes in their immediate environment. Most of them experience fairly constant conditions. Indeed, **homeostasis—** the maintenance of a constant internal environment—is considered by many to be the hallmark of multicellular organisms. Although cells in such organisms still respond to signals in their immediate environment (such as growth factors and hormones) by altering gene expression, in doing so they participate in regulating the body as a whole. In multicellular organisms with relatively constant internal environments, the primary function of gene control in a cell is not to respond to that cell's immediate environment, but rather to participate in regulating the body as a whole.

Some of these changes in gene expression compensate for changes in the physiological condition of the body. Others mediate the decisions that *produce* the body, ensuring that the right genes are expressed in the right cells at the right time during development. The growth and development of multicellular organisms entail a long series of biochemical reactions, each catalyzed by a specific enzyme. Once a particular developmental change has occurred, these enzymes cease to be active, lest they disrupt the events that must follow. To produce these enzymes, genes are transcribed in a carefully prescribed order, each for a specified period of time. In fact, many genes are activated only once, producing irreversible effects. In many animals, for example, **stem cells** develop into differentiated tissues like skin cells or red blood cells, following a fixed genetic program that often leads to programmed cell death. The one-time expression of the genes that guide this program is fundamentally different from the reversible metabolic adjustments bacterial cells make to the environment. In all multicellular organisms, changes in gene expression within particular cells serve the needs of the whole organism, rather than the survival of individual cells.

Posttranscriptional Control

Gene expression can be regulated at many levels. By far the most common form of regulation in both bacteria and eukaryotes is **transcriptional control,** that is, control of the transcription of particular genes by RNA polymerase. Other less common forms of control occur after transcription, influencing the mRNA that is produced from the genes or the activity of the proteins encoded by the mRNA. These controls, collectively referred to as **posttranscriptional controls,** will be discussed briefly later in this chapter.

Gene expression is controlled at the transcriptional and posttranscriptional levels. Transcriptional control, more common, is effected by the binding of proteins to regulatory sequences within the DNA.

16.2 Regulatory proteins read DNA without unwinding it.

How to Read a Helix without Unwinding It

It is the ability of certain proteins to bind to *specific* DNA regulatory sequences that provides the basic tool of gene regulation, the key ability that makes transcriptional control possible. To understand how cells control gene expression, it is first necessary to gain a clear picture of this molecular recognition process.

Looking into the Major Groove

Molecular biologists used to think that the DNA helix had to unwind before proteins could distinguish one DNA sequence from another; only in this way, they reasoned, could regulatory proteins gain access to the hydrogen bonds between base-pairs. We now know it is unnecessary for the helix to unwind because proteins can bind to its outside surface, where the edges of the base-pairs are exposed. Careful inspection of a DNA molecule reveals two helical grooves winding round the molecule, one deeper than the other. Within the deeper groove, called the **major groove,** the nucleotides' hydrophobic methyl groups, hydrogen atoms, and hydrogen bond donors and acceptors protrude. The pattern created by these chemical groups is unique for each of the four possible base-pair arrangements, providing a ready way for a protein nestled in the groove to read the sequence of bases (figure 16.2).

DNA-Binding Motifs

Protein-DNA recognition is an area of active research; so far, the structures of over 30 regulatory proteins have been analyzed. Although each protein is unique in its fine details, the part of the protein that actually binds to the DNA is much less variable. Almost all of these proteins employ one of a small set of **structural,** or **DNA-binding, motifs,** particular bends of the protein chain that permit it to interlock with the major groove of the DNA helix.

Regulatory proteins identify specific sequences on the DNA double helix, without unwinding it, by inserting DNA-binding motifs into the major groove of the double helix where the edges of the bases protrude.

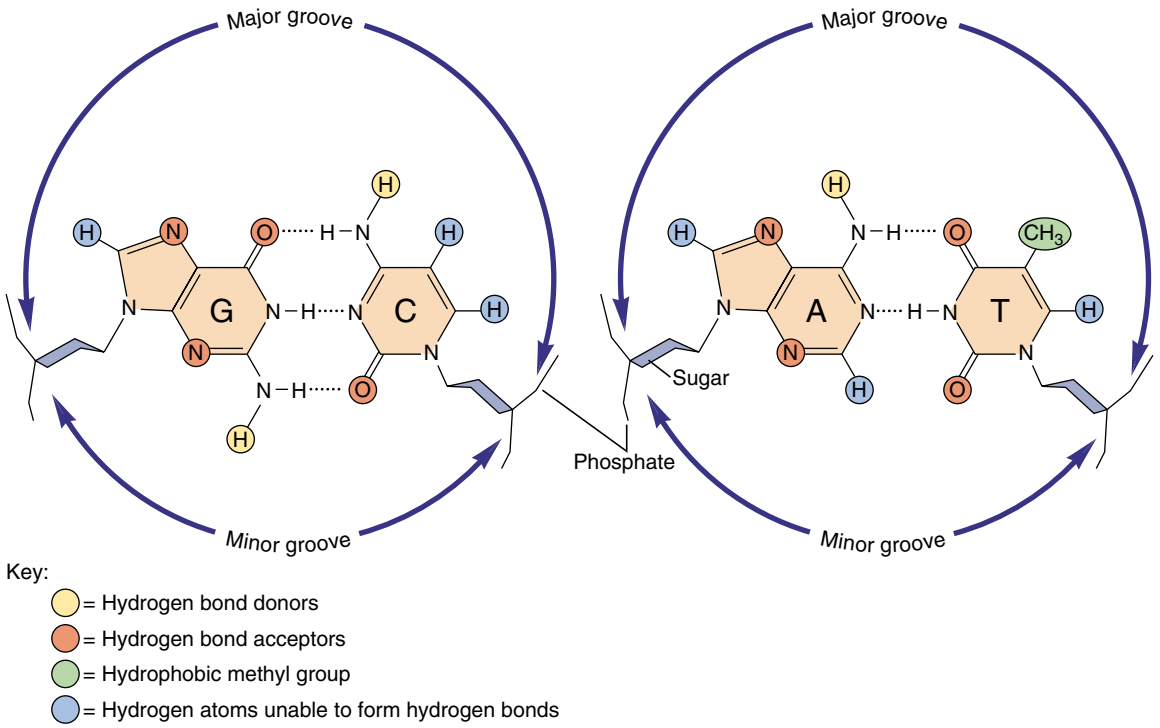


FIGURE 16.2
Reading the major groove of DNA. Looking down into the major groove of a DNA helix, we can see the edges of the bases protruding into the groove. Each of the four possible base-pair arrangements (two are shown here) extends a unique set of chemical groups into the groove, indicated in this diagram by differently colored balls. A regulatory protein can identify the base-pair arrangement by this characteristic signature.

Four Important DNA-Binding Motifs

The Helix-Turn-Helix Motif

The most common DNA-binding motif is the **helix-turn-helix,** constructed from two α-helical segments of the protein linked by a short nonhelical segment, the "turn" (figure 16.3). The first DNA-binding motif recognized, the helix-turn-helix motif has since been identified in hundreds of DNA-binding proteins.

A close look at the structure of a helix-turn-helix motif reveals how proteins containing such motifs are able to interact with the major groove of DNA. Interactions between the helical segments of the motif hold them at roughly right angles to each other. When this motif is pressed against DNA, one of the helical segments (called the recognition helix) fits snugly in the major groove of the DNA molecule, while the other butts up against the outside of the DNA molecule, helping to ensure the proper positioning of the recognition helix. Most DNA regulatory sequences recognized by helix-turn-helix motifs occur in symmetrical pairs. Such sequences are bound by proteins containing two helix-turn-helix motifs separated by 3.4 nm, the distance required for one turn of the DNA helix (figure 16.4). Having *two* protein/DNA-binding sites doubles the zone of contact between protein and DNA and so greatly strengthens the bond that forms between them.

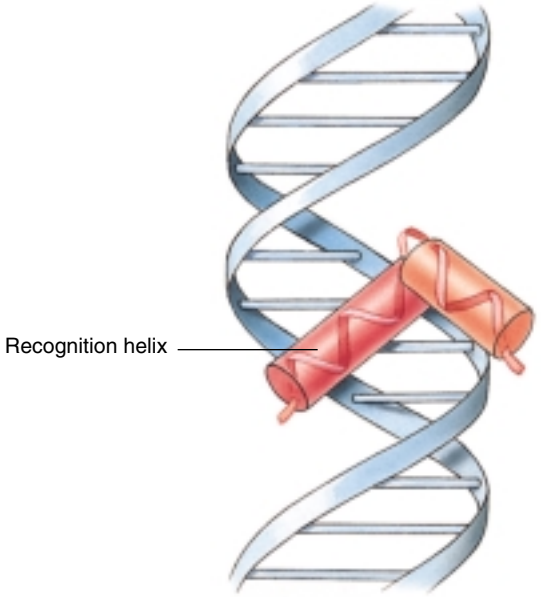


FIGURE 16.3
The helix-turn-helix motif. One helical region, called the recognition helix, actually fits into the major groove of DNA. There it contacts the edges of base-pairs, enabling it to recognize specific sequences of DNA bases.

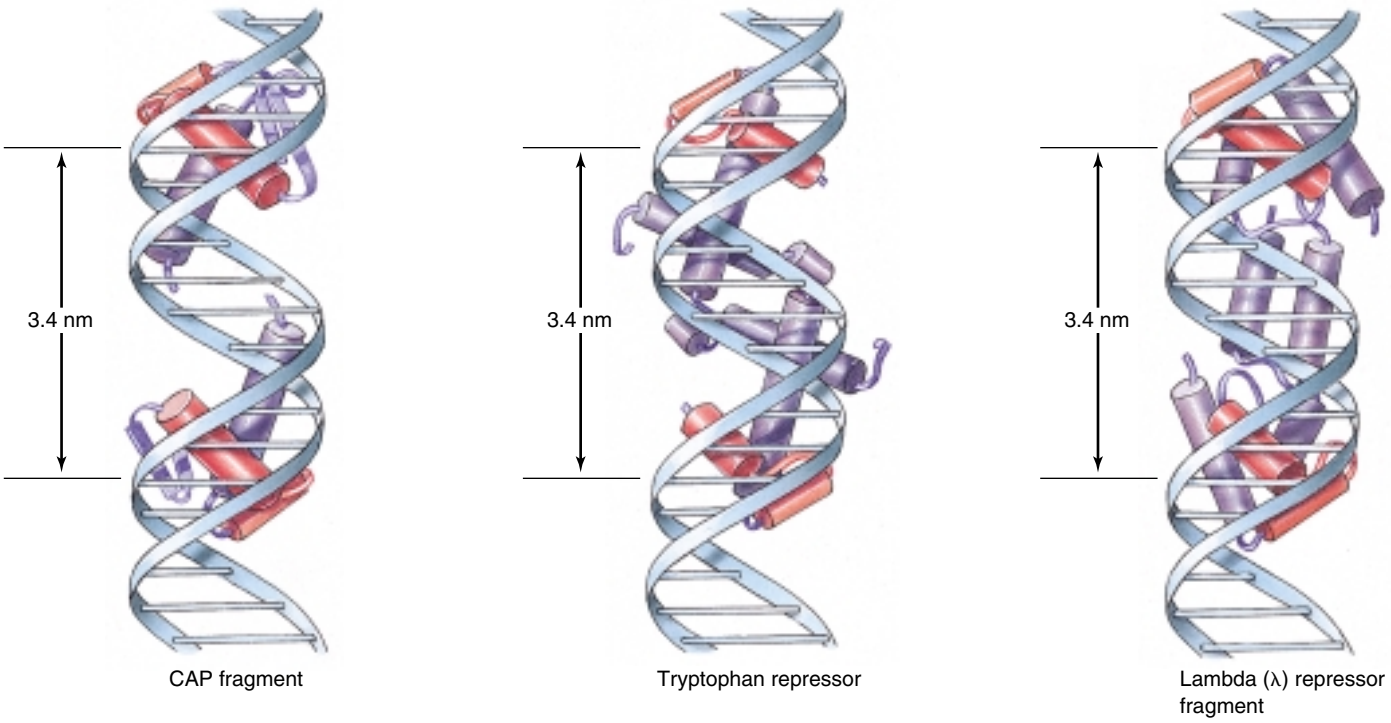


FIGURE 16.4
How the helix-turn-helix binding motif works. The three regulatory proteins illustrated here all bind to DNA using a pair of helix-turn-helix binding motifs. In each case, the two copies of the motif (*red*) are separated by 3.4 nm, precisely the spacing of one turn of the DNA helix. This allows the regulatory proteins to slip into two adjacent portions of the major groove in DNA, providing a strong attachment.

The Homeodomain Motif

A special class of helix-turn-helix motifs plays a critical role in development in a wide variety of eukaryotic organisms, including humans. These motifs were discovered when researchers began to characterize a set of homeotic mutations in *Drosophila* (mutations that alter how the parts of the body are assembled). They found that the mutant genes encoded regulatory proteins whose normal function was to initiate key stages of development by binding to developmental switch-point genes. More than 50 of these regulatory proteins have been analyzed, and they all contain a nearly identical sequence of 60 amino acids, the **homeodomain** (figure 16.5*b*). The center of the homeodomain is occupied by a helix-turn-helix motif that binds to the DNA. Surrounding this motif within the homeodomain is a region that always presents the motif to the DNA in the same way.

The Zinc Finger Motif

A different kind of DNA-binding motif uses one or more zinc atoms to coordinate its binding to DNA. Called **zinc fingers** (figure 16.5*c*), these motifs exist in several forms. In one form, a zinc atom links an α-helical segment to a β sheet segment so that the helical segment fits into the major groove of DNA. This sort of motif often occurs in clusters, the β sheets spacing the helical segments so that each helix contacts the major groove. The more zinc fingers in the cluster, the stronger the protein binds to the DNA. In other forms of the zinc finger motif, the β sheet's place is taken by another helical segment.

The Leucine Zipper Motif

In yet another DNA-binding motif, two different protein subunits cooperate to create a single DNA-binding site. This motif is created where a region on one of the subunits containing several hydrophobic amino acids (usually leucines) interacts with a similar region on the other subunit. This interaction holds the two subunits together at those regions, while the rest of the subunits are separated. Called a **leucine zipper,** this structure has the shape of a "Y," with the two arms of the Y being helical regions that fit into the major groove of DNA (figure 16.5*d*). Because the two subunits can contribute quite different helical regions to the motif, leucine zippers allow for great flexibility in controlling gene expression.

Regulatory proteins bind to the edges of base-pairs exposed in the major groove of DNA. Most contain structural motifs such as the helix-turn-helix, homeodomain, zinc finger, or leucine zipper.

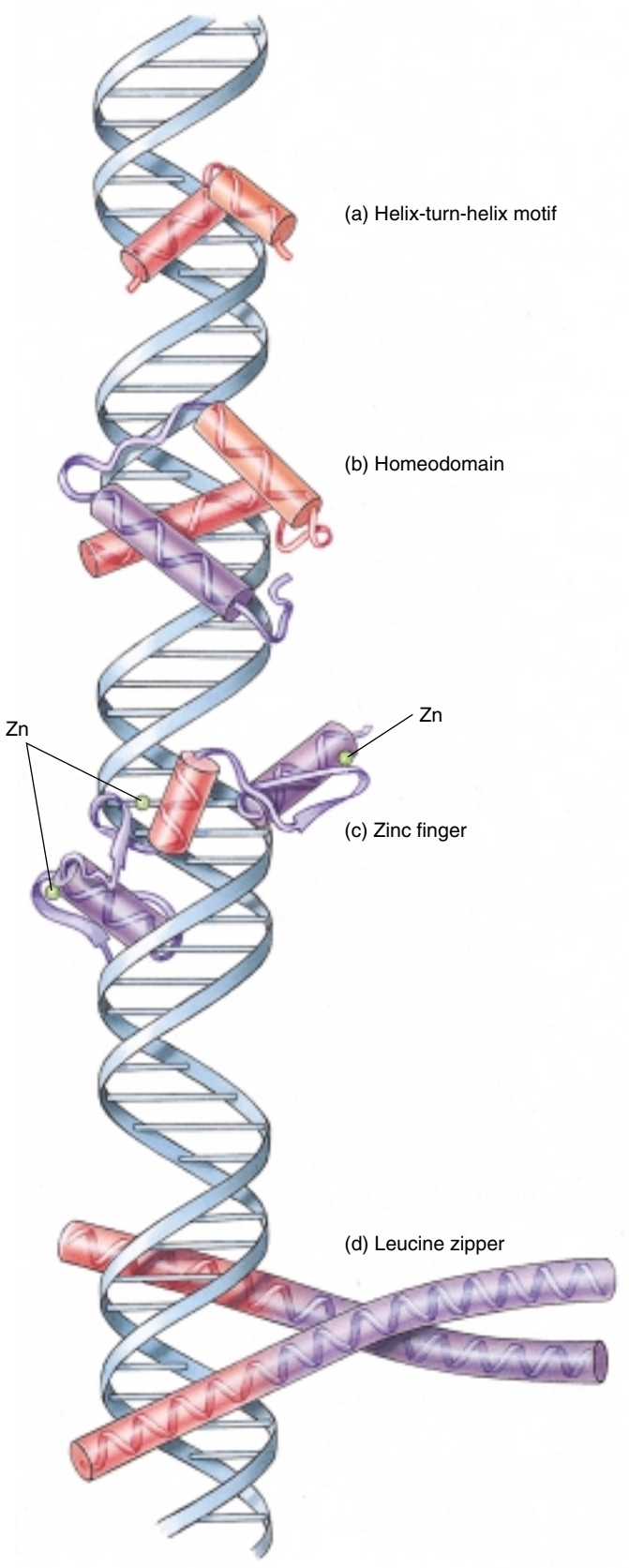


FIGURE 16.5
Major DNA-binding motifs.

16.3 Bacteria limit transcription by blocking RNA polymerase.

Controlling Transcription Initiation

How do organisms use regulatory DNA sequences and the proteins that bind them to control when genes are transcribed? The same basic controls are used in bacteria and eukaryotes, but eukaryotes employ several additional elements that reflect their more elaborate chromosomal structure. We will begin by discussing the relatively simple controls found in bacteria.

Repressors Are *OFF* Switches

A typical bacterium possesses genes encoding several thousand proteins, but only some are transcribed at any one time; the others are held in reserve until needed. When the cell encounters a potential food source, for example, it begins to manufacture the enzymes necessary to metabolize that food. Perhaps the best-understood example of this type of transcriptional control is the regulation of tryptophan-producing genes (*trp* genes), which was investigated in the pioneering work of Charles Yanofsky and his students at Stanford University.

Operons. The bacterium *Escherichia coli* uses proteins encoded by a cluster of five genes to manufacture the amino acid tryptophan. All five genes are transcribed together as a unit called an **operon,** producing a single, long piece of mRNA. RNA polymerase binds to a promoter located at the beginning of the first gene, and then proceeds down the DNA, transcribing the genes one after another. Regulatory proteins shut off transcription by binding to an operator site immediately in front of the promoter and often overlapping it.

When tryptophan is present in the medium surrounding the bacterium, the cell shuts off transcription of the *trp* genes by means of a tryptophan **repressor,** a helix-turn-helix regulatory protein that binds to the operator site located within the *trp* promoter (figure 16.6). Binding of the repressor to the operator prevents RNA polymerase from binding to the promoter. The key to the functioning of this control mechanism is that the tryptophan repressor cannot bind to DNA unless it has first bound to two molecules of tryptophan. The binding of tryptophan to the repressor alters the orientation of a pair of helix-turn-helix motifs in the repressor, causing their recognition helices to fit into adjacent major grooves of the DNA (figure 16.7).

Thus, the bacterial cell's synthesis of tryptophan depends upon the absence of tryptophan in the environment. When the environment lacks tryptophan, there is nothing to activate the repressor, so the repressor cannot prevent

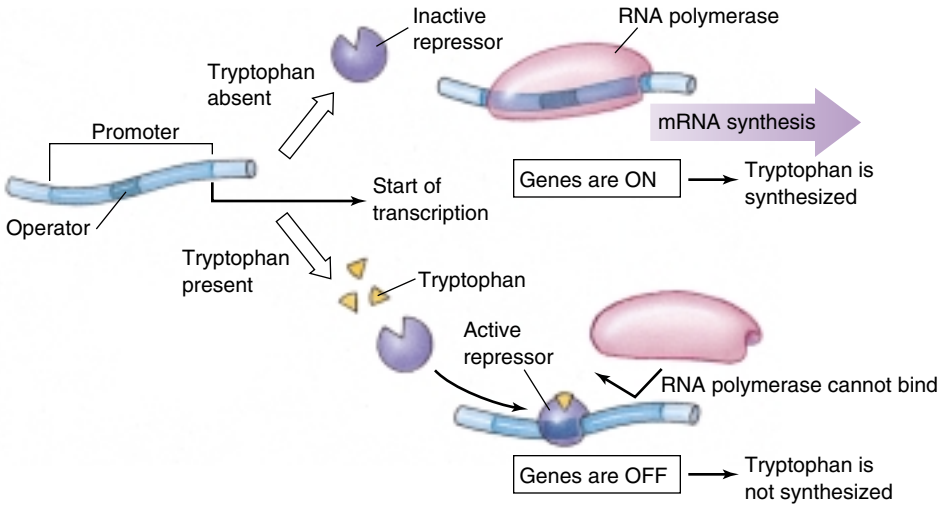


FIGURE 16.6
How the *trp* operon is controlled. The tryptophan repressor cannot bind the operator (which is located *within* the promoter) unless tryptophan first binds to the repressor. Therefore, in the absence of tryptophan, the promoter is free to function and RNA polymerase transcribes the operon. In the presence of tryptophan, the tryptophan-repressor complex binds tightly to the operator, preventing RNA polymerase from initiating transcription.

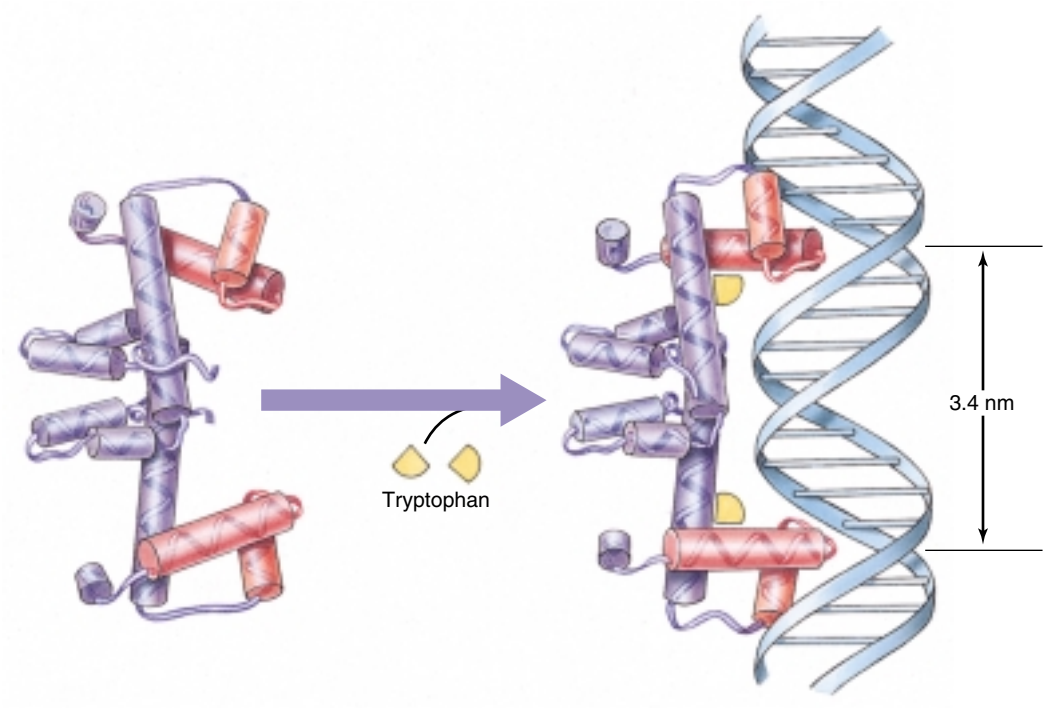


FIGURE 16.7
How the tryptophan repressor works. The binding of tryptophan to the repressor increases the distance between the two recognition helices in the repressor, allowing the repressor to fit snugly into two adjacent portions of the major groove in DNA.

RNA polymerase from binding to the *trp* promoter. The *trp* genes are transcribed, and the cell proceeds to manufacture tryptophan from other molecules. On the other hand, when tryptophan is present in the environment, it binds to the repressor, which is then able to bind to the *trp* promoter. This blocks transcription of the *trp* genes, and the cell's synthesis of tryptophan halts.

Activators Are *ON* Switches

Not all regulatory switches shut genes off—some turn them on. In these instances, bacterial promoters are deliberately constructed to be poor binding sites for RNA polymerase, and the genes these promoters govern are thus rarely transcribed—unless something happens to improve the promoter's ability to bind RNA polymerase. This can happen if a regulatory protein called a **transcriptional activator** binds to the DNA nearby. By contacting the polymerase protein itself, the activator protein helps hold the polymerase against the DNA promoter site so that transcription can begin.

A well-understood transcriptional activator is the catabolite activator protein (CAP) of *E. coli*, which initiates the transcription of genes that allow *E. coli* to use other molecules as food when glucose is not present. Falling levels of glucose lead to higher intracellular levels of the signaling molecule, cyclic AMP (cAMP), which binds to the CAP protein. When cAMP binds to it, the CAP protein changes shape, enabling its helix-turn-helix motif to bind to the DNA near any of several promoters. Consequently, those promoters are activated and their genes can be transcribed (figure 16.8).

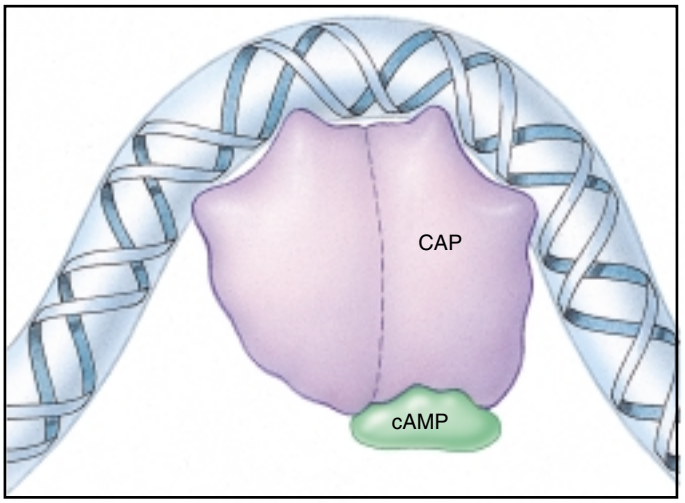


FIGURE 16.8
How CAP works. Binding of the catabolite activator protein (CAP) to DNA causes the DNA to bend around it. This increases the activity of RNA polymerase.

Combinations of Switches

By combining ON and OFF switches, bacteria can create sophisticated transcriptional control systems. A particularly well-studied example is the ***lac* operon** of *E. coli* (figure 16.9). This operon is responsible for producing three proteins that import the disaccharide lactose into the cell and break it down into two monosaccharides: glucose and galactose.

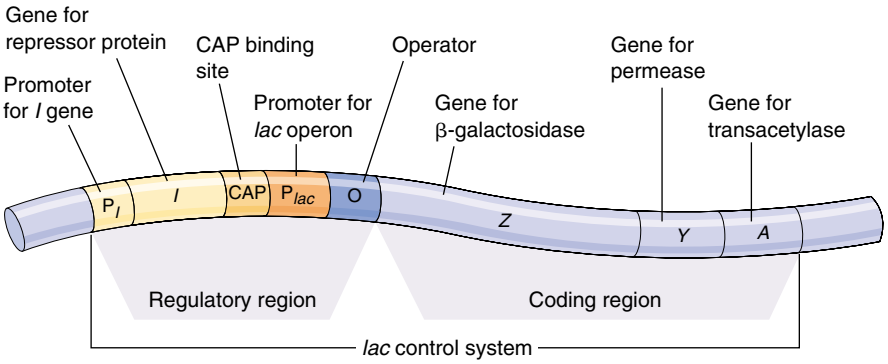


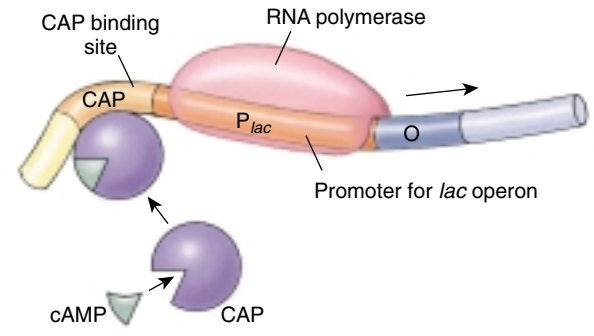
FIGURE 16.9
The *lac* region of the *Escherichia coli* chromosome. The *lac* operon consists of a promoter, an operator, and three genes that code for proteins required for the metabolism of lactose. In addition, there is a binding site for the catabolite activator protein (CAP), which affects whether or not RNA polymerase will bind to the promoter. Gene *I* codes for a repressor protein, which will bind to the operator and block transcription of the *lac* genes. The genes *Z*, *Y*, and *A* encode the two enzymes and the permease involved in the metabolism of lactose.

The Activator Switch. The *lac* operon possesses two regulatory sites. One is a CAP site located adjacent to the *lac* promoter. It ensures that the *lac* genes are not transcribed effectively when ample amounts of glucose are already present. In the absence of glucose, a high level of cAMP builds up in the cell. Consequently, cAMP is available to bind to CAP and allow it to change shape, bind to the DNA, and activate the *lac* promoter (figure 16.10). In the presence of glucose, cAMP levels are low, CAP is unable to bind to the DNA, and the *lac* promoter is not activated.

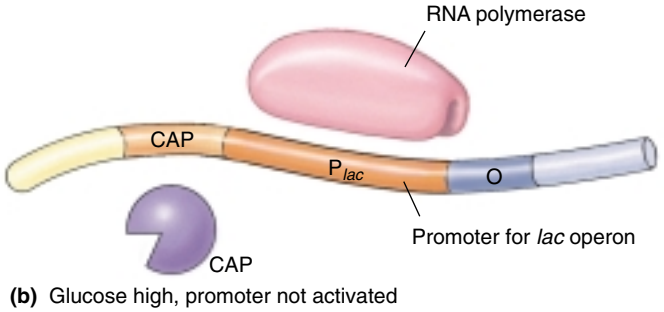
The Repressor Switch. Whether the *lac* genes are actually transcribed in the absence of glucose is determined by the second regulatory site, the **operator,** which is located adjacent to the promoter. A protein called the *lac* repressor is capable of binding to the operator, but only when lactose is absent. Because the operator and the promoter are close together, the repressor covers part of the promoter when it binds to the operator, preventing RNA polymerase from proceeding and so blocking transcription of the *lac* genes. These genes are then said to be "repressed" (figure 16.11). As a result, the cell does not transcribe genes whose products it has no use for. However, when lactose is present, a lactose isomer binds to the repressor, twisting its binding motif away from the major groove of the DNA. This prevents the repressor from binding to the operator and so allows RNA polymerase to bind to the promoter and transcribe the *lac* genes. Transcription of the *lac* operon is said to have been "induced" by lactose.

This two-switch control mechanism thus causes the cell to produce lactose-utilizing proteins whenever lactose is present but glucose is not, enabling it to make a metabolic decision to produce only what the cell needs, conserving its resources (figure 16.12).

Bacteria regulate gene expression transcriptionally through the use of repressor and activator "switches," such as the *trp* repressor and the CAP activator. The transcription of some clusters of genes, such as the *lac* operon, is regulated by both repressors and activators.



(a) Glucose low, promoter activated



(b) Glucose high, promoter not activated

FIGURE 16.10
How the CAP site works. The CAP molecule can attach to the CAP binding site only when the molecule is bound to cAMP. (*a*) When glucose levels are low, cAMP is abundant and binds to CAP. The cAMP-CAP complex binds to the CAP site, bends in the DNA, and gives RNA polymerase access to the promoter. (*b*) When glucose levels are high, cAMP is scarce, and CAP is unable to activate the promoter.

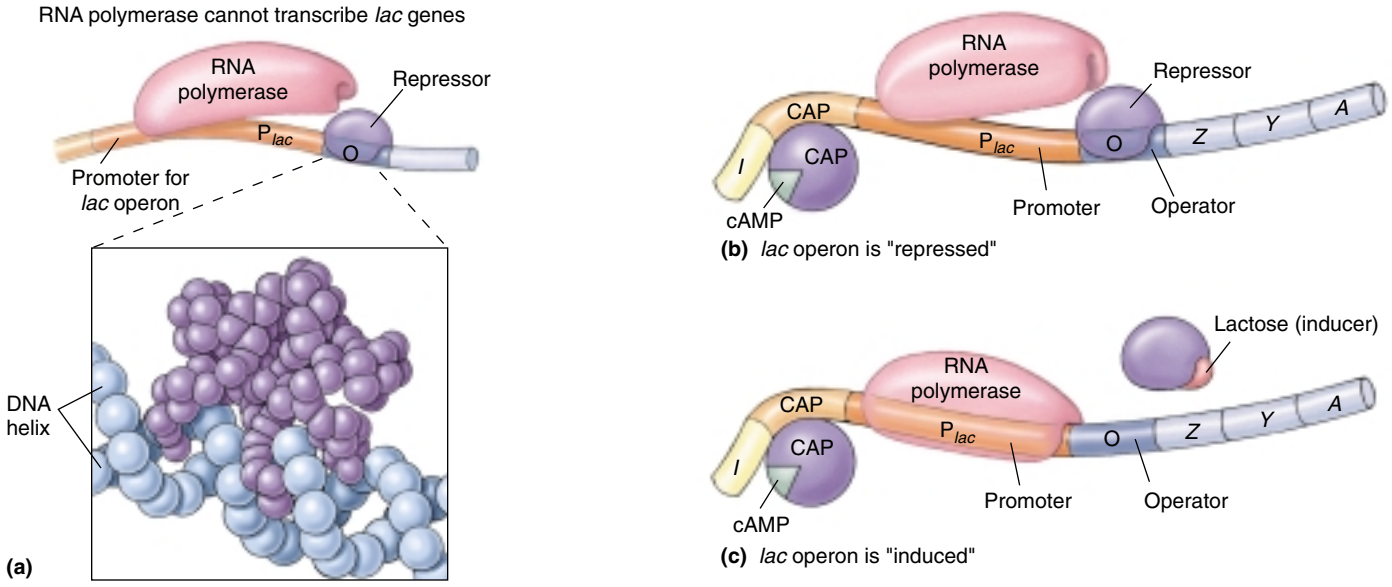


FIGURE 16.11
How the *lac* repressor works. (*a*) The *lac* repressor. Because the repressor fills the major groove of the DNA helix, RNA polymerase cannot fully attach to the promoter, and transcription is blocked. (*b*) The *lac* operon is shut down ("repressed") when the repressor protein is bound to the operator site. Because promoter and operator sites overlap, RNA polymerase and the repressor cannot functionally bind at the same time, any more than two people can sit in the same chair at once. (*c*) The *lac* operon is transcribed ("induced") when CAP is bound and when lactose binding to the repressor changes its shape so that it can no longer sit on the operator site and block RNA polymerase activity.

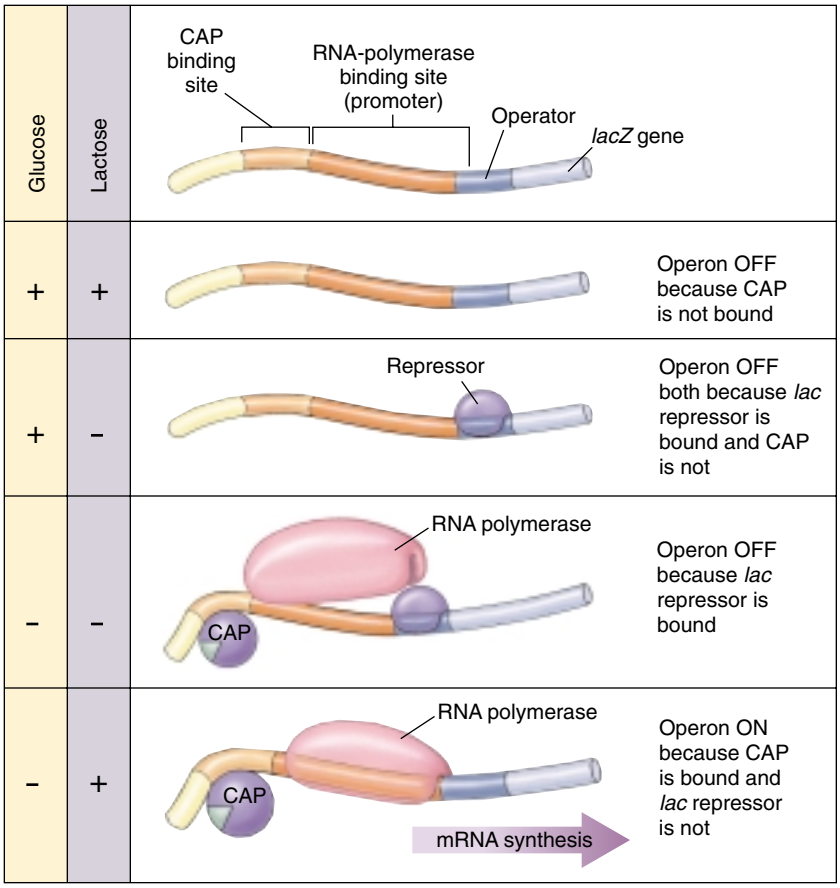


FIGURE 16.12
Two regulatory proteins control the *lac* operon. Together, the *lac* repressor and CAP provide a very sensitive response to the cell's need to utilize lactose-metabolizing enzymes.

16.4 Transcriptional control in eukaryotes operates at a distance.

Designing a Complex Gene Control System

As we have seen, combinations of ON and OFF control switches allow bacteria to regulate the transcription of particular genes in response to the immediate metabolic demands of their environment. All of these switches work by interacting directly with RNA polymerase, either blocking or enhancing its binding to specific promoters. There is a limit to the complexity of this sort of regulation, however, because only a small number of switches can be squeezed into and around one promoter. In a eukaryotic organism that undergoes a complex development, many genes must interact with one another, requiring many more interacting elements than can fit around a single promoter (table 16.1).

In eukaryotes, this physical limitation is overcome by having distant sites on the chromosome exert control over the transcription of a gene (figure 16.13). In this way, many regulatory sequences scattered around the chromosomes can influence a particular gene's transcription. This "control-at-a-distance" mechanism includes two features: a set of proteins that help bind RNA polymerase to the promoter, and modular regulatory proteins that bind to distant sites. These two features produce a truly flexible control system.

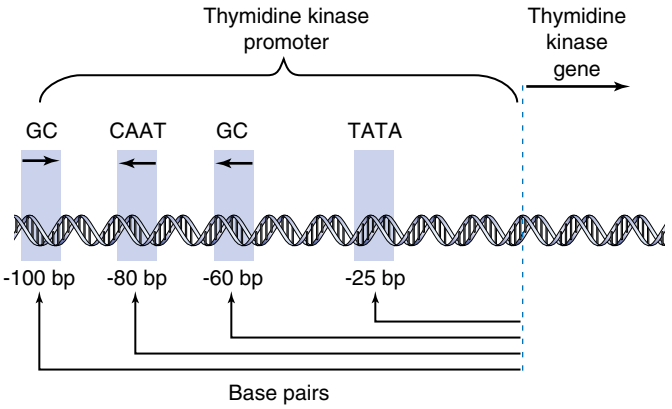


FIGURE 16.13
A eukaryotic promoter. This promoter for the gene encoding the enzyme thymidine kinase contains the TATA box that the initiation factor binds to, as well as three other DNA sequences that direct the binding of other elements of the transcription complex.

Table 16.1 Some Gene Regulatory Proteins and the DNA Sequences They Recognize

Regulatory Proteins of Species	DNA Sequence Recognized*	Regulatory Proteins of Species	DNA Sequence Recognized*
ESCHERICHIA COLI		***DROSOPHILA MELANOGASTER***	
lac repressor	AATTGTGAGCGGATAACAATT TTAACACTCGCCTATTGTTAA	Krüppel	AACGGGTTAA TTGCCCAATT
CAP	TGTGAGTTAGCTCACT ACACTCAATCGAGTGA	bicoid	GGGATTAGA CCCTAATCT
λ repressor	TATCACCGCCAGAGGTA ATAGTGGCGGTCTCCAT		
YEAST		**HUMAN**	
GAL4	CGGAGGACTGTCCTCCG GCCTCCTGACAGGAGGC	Spl	GGGCGG CCCGCC
MAT α2	CATGTAATT GTACATTAA	Oct-1	ATGCAAAT TACGTTTA
GCN4	ATGACTCAT TACTGAGTA	GATA-1	TGATAG ACTATC

*Each regulatory protein is able to recognize a family of closely related DNA sequences; only one member of each family is listed here.

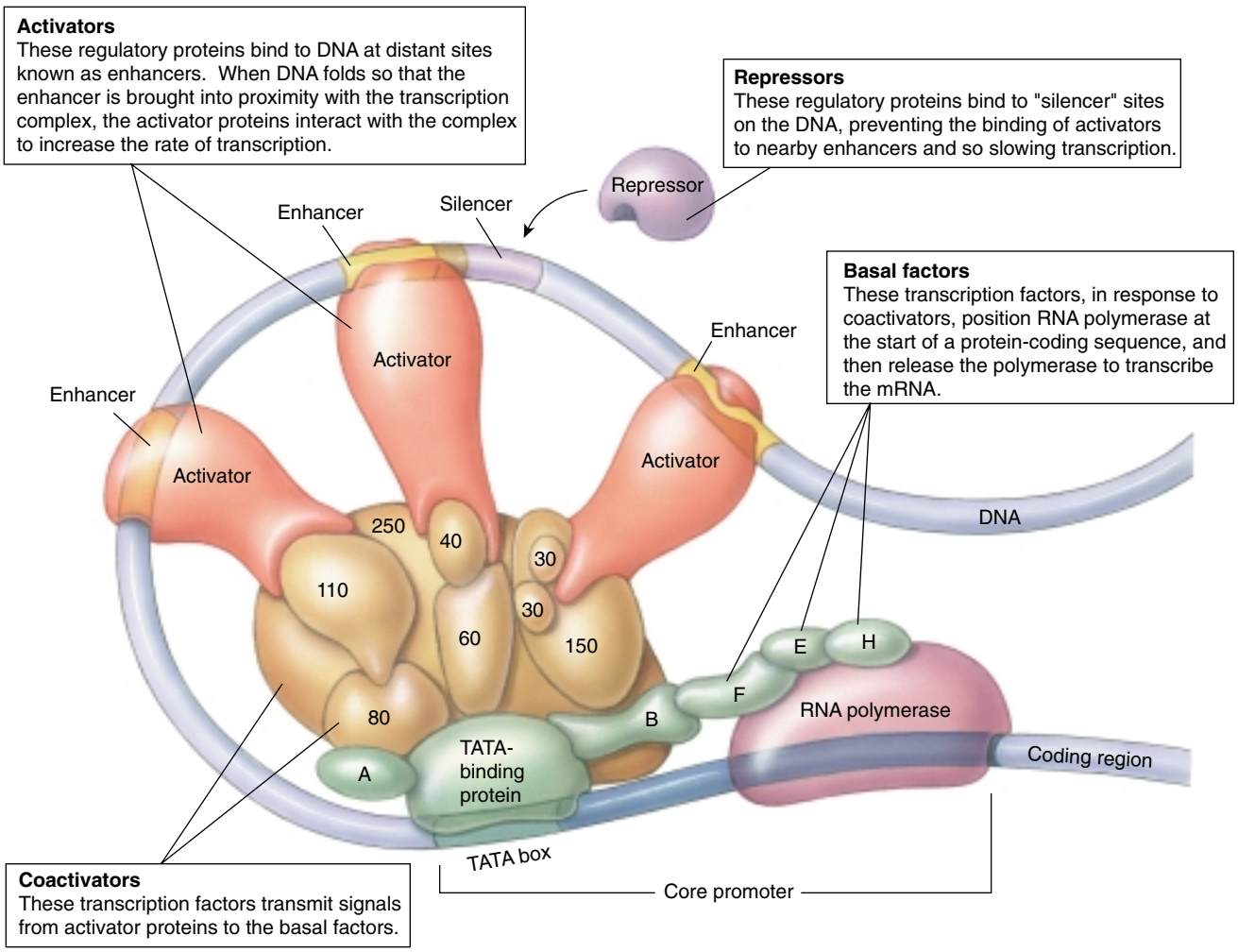


FIGURE 16.14
The structure of a human transcription complex. The transcription complex that positions RNA polymerase at the beginning of a human gene consists of four kinds of proteins. Basal factors (the green shapes at bottom of complex with letter names) are transcription factors that are essential for transcription but cannot by themselves increase or decrease its rate. They include the TATA-binding protein, the first of the basal factors to bind to the core promoter sequence. Coactivators (the tan shapes that form the bulk of the transcription complex, named according to their molecular weights) are transcription factors that link the basal factors with regulatory proteins called activators (the red shapes). The activators bind to enhancer sequences at other locations on the DNA. The interaction of individual basal factors with particular activator proteins is necessary for proper positioning of the polymerase, and the rate of transcription is regulated by the availability of these activators. When a second kind of regulatory protein called a repressor (the purple shape) binds to a so-called "silencer" sequence located adjacent to or overlapping an enhancer sequence, the corresponding activator that would normally have bound that enhancer is no longer able to do so. The activator is thus unavailable to interact with the transcription complex and initiate transcription.

Eukaryotic Transcription Factors

For RNA polymerase to successfully bind to a eukaryotic promoter and initiate transcription, a set of proteins called **transcription factors** must first assemble on the promoter, forming a complex that guides and stabilizes the binding of the polymerase (figure 16.14). The assembly process begins some 25 nucleotides upstream from the transcription start site, where a transcription factor composed of many subunits binds to a short TATA sequence (discussed in chapter 15). Other transcription factors then bind, eventually forming a full transcription factor complex able to capture RNA polymerase. In many instances, the transcription factor complex then phosphorylates the bound polymerase, disengaging it from the complex so that it is free to begin transcription.

The binding of several different transcription factors provides numerous points where control over transcription may be exerted. Anything that reduces the availability of a particular factor (for example, by regulating the promoter that governs the expression and synthesis of that factor) or limits its ease of assembly into the transcription factor complex will inhibit transcription.

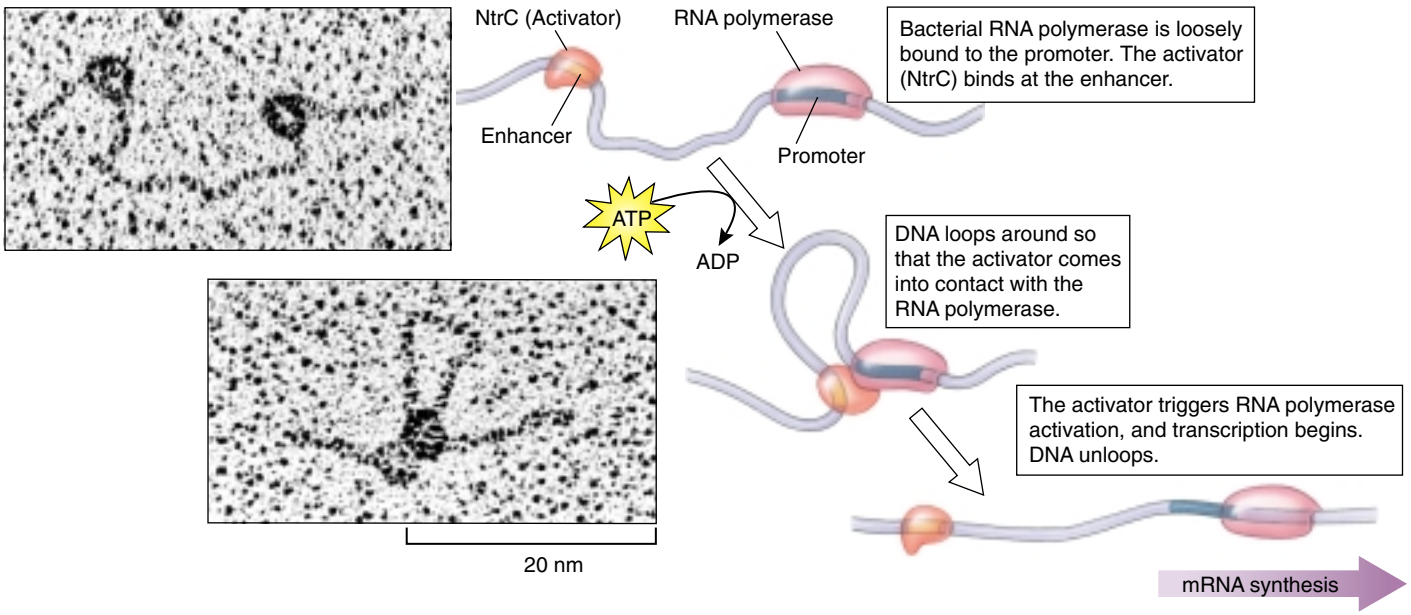


FIGURE 16.15
An enhancer in action. When the bacterial activator NtrC binds to an enhancer, it causes the DNA to loop over to a distant site where RNA polymerase is bound, activating transcription. While such enhancers are rare in bacteria, they are common in eukaryotes.

Enhancers

A key advance in the evolution of eukaryotic gene transcription was the advent of regulatory proteins composed of two distinct modules, or domains. The **DNA-binding domain** physically attaches the protein to the DNA at a specific site, using one of the structural motifs discussed earlier, while the **regulatory domain** interacts with other regulatory proteins.

The great advantage of this modular design is that it uncouples regulation from DNA binding, allowing a regulatory protein to bind to a specific DNA sequence at one site on a chromosome and exert its regulation over a promoter at another site, which may be thousands of nucleotides away. The distant sites where these regulatory proteins bind are called **enhancers.** Although enhancers also occur in exceptional instances in bacteria (figure 16.15), they are the rule rather than the exception in eukaryotes.

How can regulatory proteins affect a promoter when they bind to the DNA at enhancer sites located far from the promoter? Apparently the DNA loops around so that the enhancer is positioned near the promoter. This brings the regulatory domain of the protein attached to the enhancer into direct contact with the transcription factor complex attached to the promoter (figure 16.16).

The enhancer mode of transcriptional control that has evolved in eukaryotes adds a great deal of flexibility to the control process. The positioning of regulatory sites at a distance permits a large number of different regulatory sequences scattered about the DNA to influence a particular gene.

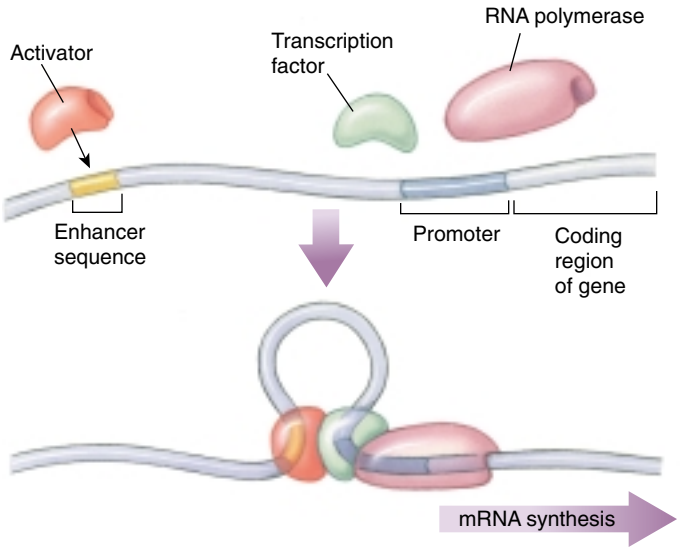


FIGURE 16.16
How enhancers work. The enhancer site is located far away from the gene being regulated. Binding of an activator (*red*) to the enhancer allows the activator to interact with the transcription factors (*green*) associated with RNA polymerase, activating transcription.

Transcription factors and enhancers confer great flexibility on the control of gene expression in eukaryotes.

The Effect of Chromosome Structure on Gene Regulation

The way DNA is packaged into chromosomes can have a profound effect on gene expression. As we saw in chapter 11, the DNA of eukaryotes is packaged in a highly compact form that enables it to fit into the cell nucleus. DNA is wrapped tightly around histone proteins to form nucleosomes (figure 16.17) and then the strand of nucleosomes is twisted into 30-nm filaments.

Promoter Blocking by Nucleosomes

Intensive study of eukaryotic chromosomes has shown that histones positioned over promoters block the assembly of transcription factor complexes. Therefore, transcription factors appear unable to bind to a promoter packaged in a nucleosome. In this way, nucleosomes may prevent continuous transcription initiation. On the other hand, nucleosomes do *not* inhibit activators and RNA polymerase. The regulatory domains of activators attached to enhancers apparently are able to displace the histones that block a promoter. In fact, this displacement of histones and the binding of activator to promoter are required for the assembly of the transcription factor complex. Once transcription has begun, RNA polymerase seems to push the histones aside as it traverses the nucleosome.

DNA Methylation

Chemical **methylation** of the DNA was once thought to play a major role in gene regulation in vertebrate cells. The addition of a methyl group to cytosine creates 5-methylcytosine but has no effect on base-pairing with guanine (figure 16.18), just as the addition of a methyl group to uracil produces thymine without affecting base-pairing with adenine. Many inactive mammalian genes are methylated, and it was tempting to conclude that methylation caused the inactivation. However, methylation is now viewed as having a less direct role, blocking accidental transcription of "turned-off" genes. Vertebrate cells apparently possess a protein that binds to clusters of 5-methylcytosine, preventing transcriptional activators from gaining access to the DNA. DNA methylation in vertebrates thus ensures that once a gene is turned off, it stays off.

Transcriptional control of gene expression occurs in eukaryotes despite the tight packaging of DNA into nucleosomes.

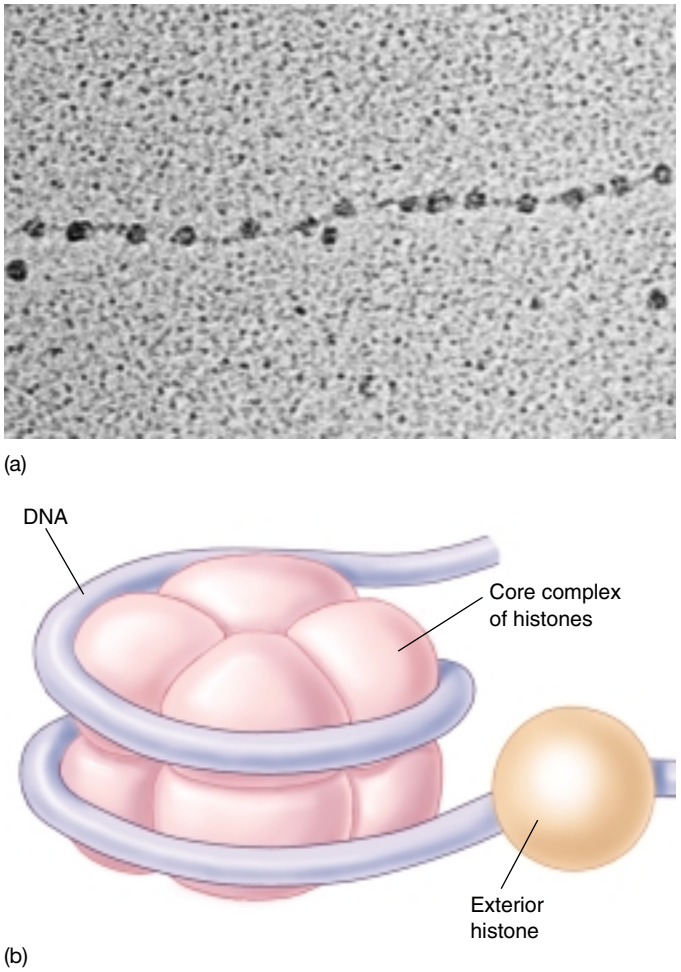


FIGURE 16.17
Nucleosomes. (*a*) In the electron micrograph, the individual nucleosomes have diameters of about 10 nm. (*b*) In the diagram of a nucleosome, the DNA double helix is wound around a core complex of eight histones; one additional histone binds to the outside of the nucleosome, exterior to the DNA.

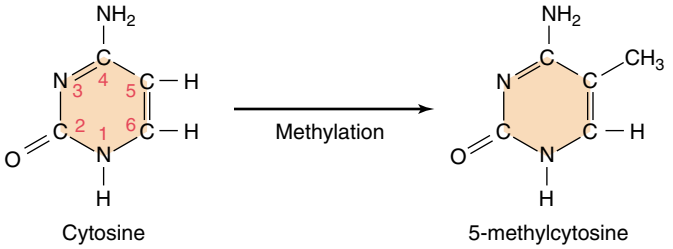


FIGURE 16.18
DNA methylation. Cytosine is methylated, creating 5-methylcytosine. Because the methyl group is positioned to the side, it does not interfere with the hydrogen bonds of a GC base-pair.

Posttranscriptional Control in Eukaryotes

Thus far we have discussed gene regulation entirely in terms of transcription initiation, that is, when and how often RNA polymerase starts "reading" a particular gene. Most gene regulation appears to occur at this point. However, there are many other points after transcription where gene expression could be regulated in principle, and all of them serve as control points for at least some eukaryotic genes. In general, these posttranscriptional control processes involve the recognition of specific sequences on the primary RNA transcript by regulatory proteins or other RNA molecules.

Processing of the Primary Transcript

As we learned in chapter 15, most eukaryotic genes have a patchwork structure, being composed of numerous short coding sequences (exons) embedded within long stretches of noncoding sequences (introns). The initial mRNA molecule copied from a gene by RNA polymerase, the **primary transcript,** is a faithful copy of the entire gene, including introns as well as exons. Before the primary transcript is translated, the introns, which comprise on average 90% of the transcript, are removed in a process called **RNA processing,** or **RNA splicing.** Particles called *small nuclear ribonucleoproteins*, or *snRNPs* (more informally, **snurps**), are thought to play a role in RNA splicing. These particles reside in the nucleus of a cell and are composed of proteins and a special type of RNA called *small nuclear RNA*, or *snRNA*. One kind of snRNP contains snRNA that can bind to the 5′ end of an intron by forming base-pairs with complementary sequences on the intron. When multiple snRNPs combine to form a larger complex called a **spliceosome,** the intron loops out and is excised (figure 16.19).

RNA splicing provides a potential point where the expression of a gene can be controlled, because exons can be

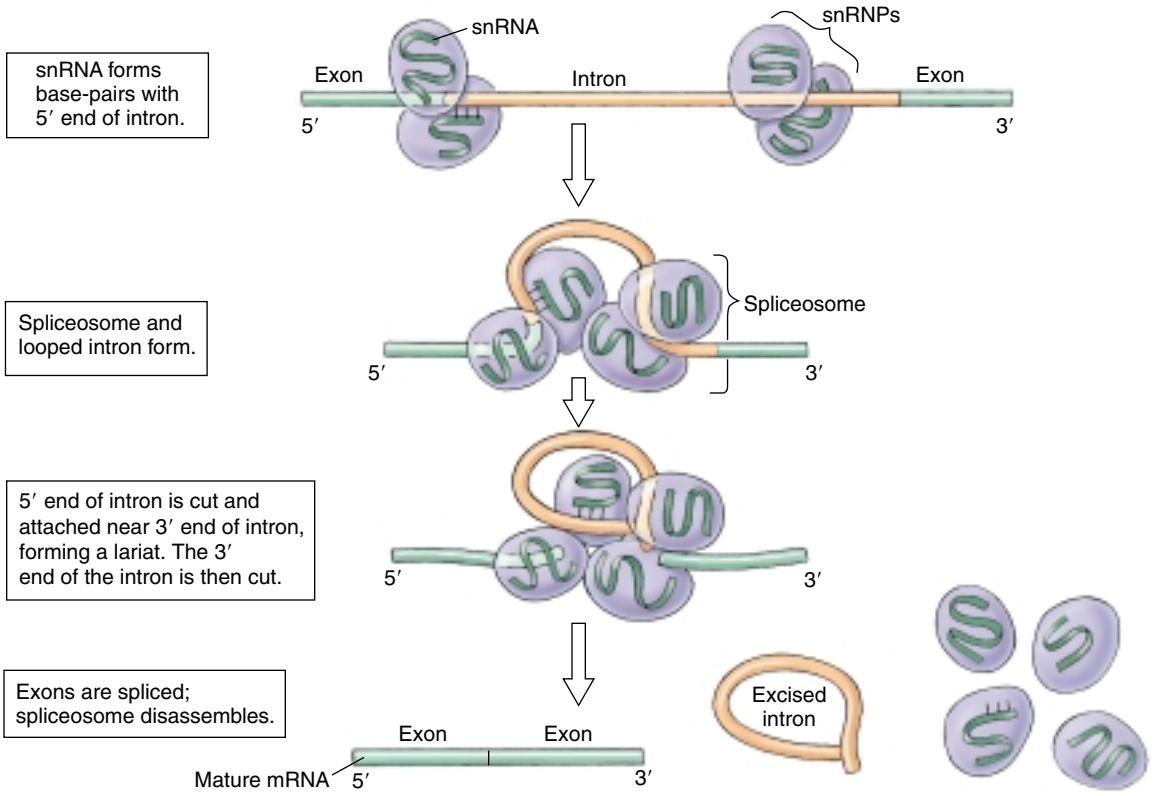


FIGURE 16.19
How spliceosomes process RNA. Particles called snRNPs contain snRNA that interacts with the 5′ end of an intron. Several snRNPs come together and form a spliceosome. As the intron forms a loop, the 5′ end is cut and linked to a site near the 3′ end of the intron. The intron forms a lariat that is excised, and the exons are spliced together. The spliceosome then disassembles and releases the mature mRNA.

spliced together in different ways, allowing a variety of different polypeptides to be assembled from the same gene! Alternative splicing is common in insects and vertebrates, with two or three different proteins produced from one gene. In many cases, gene expression is regulated by changing which splicing event occurs during different stages of development or in different tissues.

An excellent example of alternative splicing in action is found in two different human organs, the thyroid and the hypothalamus. The thyroid gland (see chapter 56) is responsible for producing hormones that control processes such as metabolic rate. The hypothalamus, located in the brain, collects information from the body (for example, salt balance) and releases hormones that in turn regulate the release of hormones from other glands, such as the pituitary gland (see chapter 56). The two organs produce two distinct hormones, calcitonin and CGRP (calcitonin gene-related peptide) as part of their function. Calcitonin is responsible for controlling the amount of calcium we take up from our food and the balance of calcium in tissues like bone and teeth. CGRP is involved in a number of neural and endocrine functions. Although these two hormones are used for very different physiological purposes, the hormones are made using the same transcript (figure 16.20). The appearance of one product versus another is determined by tissue-specific factors that regulate the processing of the primary transcript. This ability offers another powerful way to control the expression of gene products, ranging from proteins with subtle differences to totally unrelated proteins.

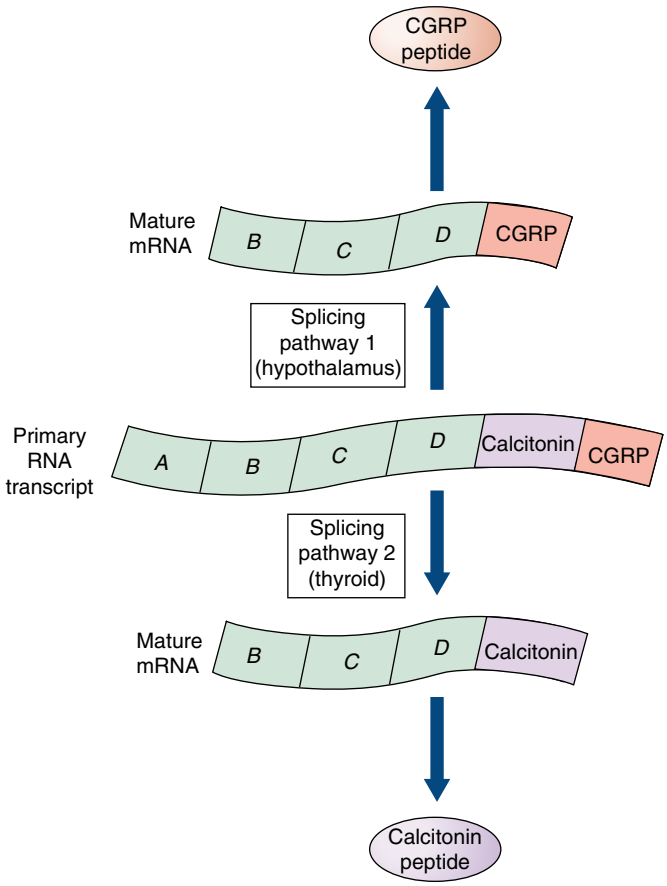


FIGURE 16.20
Alternative splicing products. The same transcript made from one gene can be spliced differently to give rise to two very distinct protein products, calcitonin and CGRP.

Transport of the Processed Transcript Out of the Nucleus

Processed mRNA transcripts exit the nucleus through the nuclear pores described in chapter 5. The passage of a transcript across the nuclear membrane is an active process that requires that the transcript be recognized by receptors lining the interior of the pores. Specific portions of the transcript, such as the poly-A tail, appear to play a role in this recognition. The transcript cannot move through a pore as long as any of the splicing enzymes remain associated with the transcript, ensuring that partially processed transcripts are not exported into the cytoplasm.

There is little hard evidence that gene expression is regulated at this point, although it could be. On average, about 10% of transcribed genes are exon sequences, but only about 5% of the total mRNA produced as primary transcript ever reaches the cytoplasm. This suggests that about half of the exon primary transcripts never leave the nucleus, but it is not clear whether the disappearance of this mRNA is selective.

Selecting Which mRNAs Are Translated

The translation of a processed mRNA transcript by the ribosomes in the cytoplasm involves a complex of proteins called translation factors. In at least some cases, gene expression is regulated by modification of one or more of these factors. In other instances, **translation repressor proteins** shut down translation by binding to the beginning of the transcript, so that it cannot attach to the ribosome. In humans, the production of ferritin (an iron-storing protein) is normally shut off by a translation repressor protein called aconitase. Aconitase binds to a 30-nucleotide sequence at the beginning of the ferritin mRNA, forming a stable loop to which ribosomes cannot bind. When the cell encounters iron, the binding of iron to aconitase causes the aconitase to dissociate from the ferritin mRNA, freeing the mRNA to be translated and increasing ferritin production 100-fold.

A Vocabulary of Gene Expression

activator A regulatory protein that promotes gene transcription by binding to DNA sequences upstream of a promoter. Activator binding stimulates RNA polymerase activity.

anticodon The three-nucleotide sequence on one end of a tRNA molecule that is complementary to and base-pairs with an amino acid–specifying codon in mRNA.

codon The basic unit of the genetic code; a sequence of three adjacent nucleotides in DNA or mRNA that codes for one amino acid or for polypeptide termination.

exon A segment of eukaryotic DNA that is both transcribed into mRNA and translated into protein. Exons are typically scattered within much longer stretches of nontranslated intron sequences.

intron A segment of eukaryotic DNA that is transcribed into mRNA but removed before translation.

nonsense codon A codon (UAA, UAG, or UGA) for which there is no tRNA with a complementary andicodon; a chain-terminating codon often called a "stop" codon.

operator A site of negative gene regulation; a sequence of nucleotides near or within the promoter that is recognized by a repressor. Binding of the repressor to the operator prevents the functional binding of RNA polymerase to the promoter and so blocks transcription.

operon A cluster of functionally related genes transcribed into a single mRNA molecule. A common mode of gene regulation in prokaryotes, it is rare in eukaryotes other than fungi.

promoter A site upstream from a gene to which RNA polymerase attaches to initiate transcription.

repressor A protein that regulates transcription by binding to the operator and so preventing RNA polymerase from initiating transcription from the promoter.

RNA polymerase The enzyme that transcribes DNA into RNA.

transcription The RNA polymerase-catalyzed assembly of an RNA molecule complementary to a strand of DNA.

translation The assembly of a polypeptide on the ribosomes, using mRNA to direct the sequence of amino acids.

Selectively Degrading mRNA Transcripts

Another aspect that affects gene expression is the stability of mRNA transcripts in the cell cytoplasm (figure 16.21). Unlike bacterial mRNA transcripts, which typically have a half-life of about 3 minutes, eukaryotic mRNA transcripts are very stable. For example, β-globin gene transcripts have a half-life of over 10 hours, an eternity in the fast-moving metabolic life of a cell. The transcripts encoding regulatory proteins and growth factors, however, are usually much less stable, with half-lives of less than 1 hour. What makes these particular transcripts so unstable? In many cases, they contain specific sequences near their 3′ ends that make them attractive targets for enzymes that degrade mRNA. A sequence of A and U nucleotides near the 3′ poly-A tail of a transcript promotes removal of the tail, which destabilizes the mRNA. Histone transcripts, for example, have a half-life of about 1 hour in cells that are actively synthesizing DNA; at other times during the cell cycle, the poly-A tail is lost and the transcripts are degraded within minutes. Other mRNA transcripts contain sequences near their 3′ ends that are recognition sites for endonucleases, which causes these transcripts to be digested quickly. The short half-lives of the mRNA transcripts of many regulatory genes are critical to the function of those genes, as they enable the levels of regulatory proteins in the cell to be altered rapidly.

An Example of a Complex Gene Control System

Sunlight is an important gene-controlling signal for plants, from germination to seed formation. Plants must regulate their genes according to the presence of sunlight, the quality of the light source, the time of day, and many other environmental signals. The combination of these responses culminate in the way the genes are regulated, such as the genes *cab* (a chlorophyll-binding photosynthetic protein) and *rbcS* (a subunit of a carbon-fixing enzyme). For instance, photosynthesis-related genes tend to express early in the day, to carry out photosynthesis, and begin to shut down later in the day. Expression levels may also be regulated according to lighting conditions, such as cloudy days versus sunny days. When darkness arrives, the transcripts must be degraded in preparation for the next day. This is an example of how complex a gene control system can be, and scientists are just beginning to understand parts of such a complicated system.

Although less common than transcriptional control, posttranscriptional control of gene expression occurs in eukaryotes via RNA splicing, translation repression, and selective degradation of mRNA transcripts.

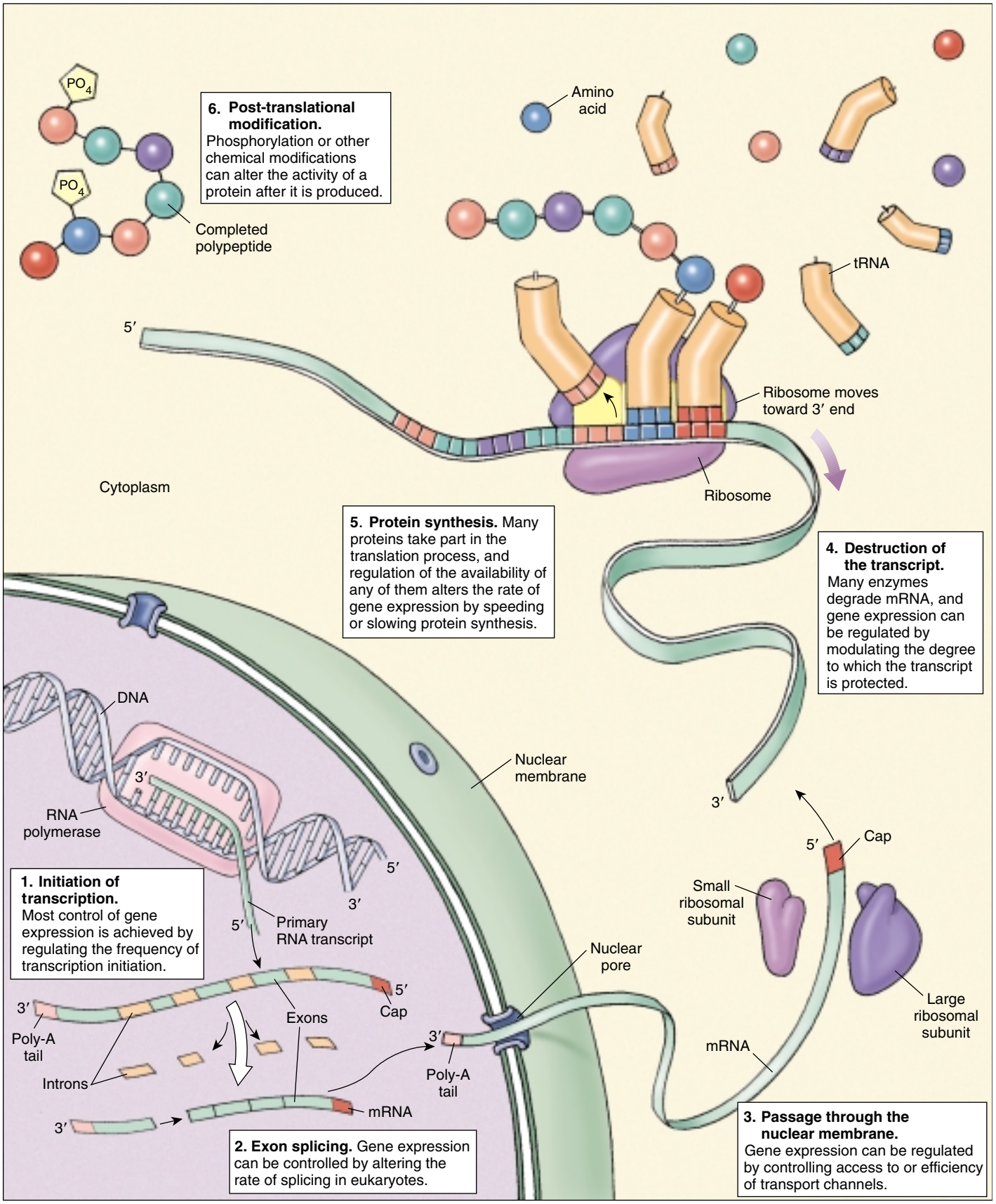


FIGURE 16.21
Six levels where gene expression can be controlled in eukaryotes.

Summary | Questions | Media Resources

16.1 Gene expression is controlled by regulating transcription.

- Regulatory sequences are short stretches of DNA that function in transcriptional control but are not transcribed themselves.
- Regulatory proteins recognize and bind to specific regulatory sequences on the DNA.

1. How do regulatory proteins identify specific nucleotide sequences without unwinding the DNA?

- Exploration: Gene regulation
- Student Research: Heat Shock Proteins

16.2 Regulatory proteins read DNA without unwinding it.

- Regulatory proteins possess structural motifs that allow them to fit snugly into the major groove of DNA, where the sides of the base-pairs are exposed.
- Common structural motifs include the helix-turn-helix, homeodomain, zinc finger, and leucine zipper.

2. What is a helix-turn-helix motif? What sort of developmental events are homeodomain motifs involved in?

- Exploration: Reading DNA

16.3 Bacteria limit transcription by blocking RNA polymerase.

- Many genes are transcriptionally regulated through repressors, proteins that bind to the DNA at or near the promoter and thereby inhibit transcription of the gene.
- Genes may also be transcriptionally regulated through activators, proteins that bind to the DNA and thereby stimulate the binding of RNA polymerase to the promoter.
- Transcription is often controlled by a *combination* of repressors and activators.

3. Describe the mechanism by which the transcription of *trp* genes is regulated in *Escherichia coli* when tryptophan is present in the environment.

4. Describe the mechanism by which the transcription of *lac* genes is regulated in *E. coli* when glucose is absent but lactose is present in the environment.

- Art Activity: The lac operon
- Regulation of *E.coli lac* operon
- Regulation of *E.coli trp* operon
- Gene Regulation

16.4 Transcriptional control in eukaryotes operates at a distance.

- In eukaryotes, RNA polymerase cannot bind to the promoter unless aided by a family of transcription factors.
- Anything that interferes with the activity of the transcription factors can block or alter gene expression.
- Eukaryotic DNA is packaged tightly in nucleosomes within chromosomes. This packaging appears to provide some inhibition of transcription, although regulatory proteins and RNA polymerase can still activate specific genes even when they are so packaged.
- Gene expression can also be regulated at the posttranscriptional level, through RNA splicing, translation repressor proteins, and the selective degradation of mRNA transcripts.

5. How do transcription factors promote transcription in eukaryotic cells? How do the enhancers of eukaryotic cells differ from most regulatory sites on bacterial DNA?

6. What role does the methylation of DNA likely play in transcriptional control?

7. How does the *primary* RNA transcript of a eukaryotic gene differ from the mRNA transcript of that gene as it is translated in the cytoplasm?

8. How can a eukaryotic cell control the translation of mRNA transcripts after they have been transported from the nucleus to the cytoplasm?