

✓CHECK YOUR UNDERSTANDING

Eukaryotic DNA is wrapped tightly around histones, forming nucleosomes, which are then coiled into dense structures called 30-nm fibers. Before transcription can begin, the chromatin (DNA-protein complex) must be relaxed by chromatin-remodeling complexes and HATs so that RNA polymerase can contact the promoter. You should be able to (1) draw and label the structure of a nucleosome and a 30-nm fiber and explain their function, and (2) compare and contrast the structure of DNA in bacteria versus that in eukaryotes.

18.3 Regulatory Sequences and Regulatory Proteins

Chapter 16 introduced the **promoter**—the site in DNA where RNA polymerase binds to initiate transcription. Eukaryotic promoters are located about 30 nucleotides upstream of the point where RNA polymerase begins transcription. Many contain a specific base sequence, known as a **TATA box**, where a sigma-like protein binds and allows RNA polymerase to contact DNA. Based on their position and function, eukaryotic promoters are much like bacterial promoters. Recall from Chapter 16 that bacterial promoters may vary in sequence and bind different types of sigma proteins. Eukaryotic genes that are transcribed by RNA polymerase II also have promoters that vary in sequence, but all eukaryotic promoters are bound by the same protein: the **TATA-binding protein (TBP)**.

If eukaryotic genes have promoters that interact with the same promoter-binding protein, how can transcription be controlled? The answer lies in interactions between regulatory proteins and regulatory sequences. Eukaryotic regulatory proteins are analogous to *E. coli*'s CAP and the repressor protein analyzed in Chapter 17. **Regulatory sequences** are sections of DNA that are involved in controlling the activity of genes, similar to the CAP site and operators introduced in Chapter 17.

As it turns out, some regulatory sequences found in eukaryotes are very similar to those observed in bacteria. Others are radically different. Let's take a closer look.

Promoter-Proximal Elements

The first regulatory sequences in eukaryotic DNA were discovered in the late 1970s, when Yasuji Oshima and co-workers developed an important model system to study control of gene expression in eukaryotes. Much as François Jacob and Jacques Monod had focused on lactose metabolism in *E. coli*, Oshima's team set out to understand how yeast cells control the metabolism of the sugar galactose. When galactose is absent, yeast cells produce tiny quantities of the enzymes required to metabolize it. But when galactose is present, transcription of the genes encoding these enzymes increases by a factor of 1000.

Oshima and co-workers focused their early work on mutants that were unable to use galactose. The team's first major result was the discovery of mutant cells that failed to produce any of the five enzymes required for galactose metabolism, even

if galactose was present. To interpret this observation, they hypothesized that the cells had a knock-out (loss-of-function) mutation that completely disabled a regulatory protein. Like CAP, this hypothesized protein was thought to exert positive control over the five genes. The team called the gene *GAL4* and the protein product GAL4. (In *S. cerevisiae*, the names of genes are written in full caps and italicized. As in other species, gene names are italicized but protein names are not.)

This hypothesis was supported when other researchers isolated the regulatory protein and found that it has a DNA-binding domain, analogous to the helix-turn-helix motif introduced in Chapter 17. The domain binds to a 20-base-pair stretch of DNA located just upstream from the promoter for the five genes that GAL4 regulates. The location and structure of this regulatory sequence are comparable to those of the CAP binding site in the *lac* operon of *E. coli*. Similar regulatory sequences have now been found in a wide array of eukaryotic genes and species. Because such sequences are located close to the promoter and bind regulatory proteins, they are termed **promoter-proximal elements**.

As Figure 18.5 shows, promoter-proximal elements are just upstream from the promoter and the gene's start site; the exons and introns that make up the actual gene are farther downstream. Unlike the promoter itself, promoter-proximal elements have sequences that are unique to specific genes. In this way, they furnish a mechanism for eukaryotic cells to exert precise control over transcription.

The discovery of positive control and promoter-proximal elements provided a satisfying parallel between gene regulation in bacteria and in eukaryotes. This picture changed, however, when researchers discovered a new class of eukaryotic DNA regulatory sequences—sequences unlike anything in bacteria.

Enhancers

Susumu Tonegawa and colleagues made a discovery that may rank as the most startling in the history of research on gene expression. Tonegawa's group was exploring how human immune system cells regulate the genes involved in the production of antibodies. Antibodies are proteins that bind to specific sites on other molecules. In your immune system, antibodies bind to viruses and bacteria and mark them for destruction by other

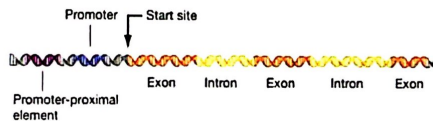


FIGURE 18.5 Promoter-Proximal Elements Regulate the Expression of Some Eukaryotic Genes

All eukaryotic genes have promoters, the sites where RNA polymerase initially is in contact with the gene. Exons and introns are not drawn to scale throughout this chapter. They are typically very large compared with promoters and promoter-proximal (regulatory) elements.

cells. The antibody gene that the researchers were working with is broken into many introns and exons. Introns are DNA sequences spliced out of the primary mRNA transcript; exons are regions of eukaryotic genes that are included in the mature RNA once splicing is complete. The biologists used techniques that will be introduced in Chapter 19 to place copies of an intron in new locations and had found that, when they placed the intron close to a gene, the gene's transcription rate increased. Based on this observation, Tonegawa and co-workers hypothesized that the intron contained some sort of regulatory sequence.

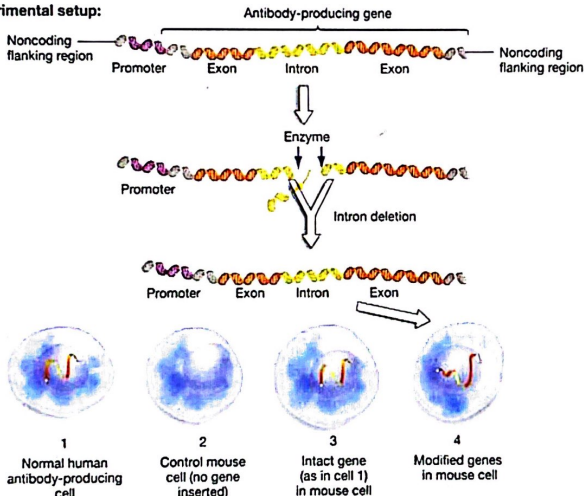
To test this hypothesis, the biologists performed what is now considered a classic experiment (Figure 18.6). The protocol was simple in concept: Starting with a human antibody-producing gene that included an intron flanked by two exons, as shown in step 1 of Figure 18.6, the team removed different portions of the intron and observed whether transcription of the gene still took place. The researchers used enzymes to cut out specific pieces of the intron in multiple copies of the gene (step 2) and to ligate the resulting sections back together (step 3). Each of the modified genes was missing a different section of the intron. Then the team

Question: Can a regulatory sequence in DNA be located far from the promoter?

Hypothesis: A regulatory sequence exists far from the promoter of an antibody-producing gene—in an intron.

Alternative hypothesis: Regulatory sequences are located close to the promoter—not in introns or in other distant sequences.

Experimental setup:



1. An antibody-producing gene in humans has a large intron.

2. Use enzymes to cut out specific sections of the intron. (Several sections were cut out—only one example is shown.)

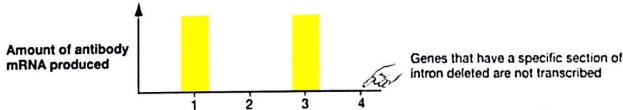
3. Use an enzyme to put DNA fragments back together.

4. Introduce intact or modified genes into mouse cells, which do not normally produce this protein. Different types of modified genes are missing different portions of the intron.

Prediction: Transcription will be reduced when part of the intron is deleted.

Prediction of alternative hypothesis: Transcription will not be affected by deleting part of the intron.

Results:



Conclusion: The deleted part of the intron must contain a regulatory sequence that is required for transcription. Thus, regulatory sequences in DNA may be located far from the promoter—in this case, in an intron.

FIGURE 18.6 Evidence That Enhancers Are Required for Transcription

QUESTION Why did the researchers assess mRNA production in human cells that did not receive a modified or intact gene?

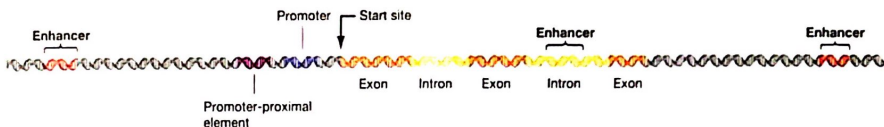


FIGURE 18.7 Enhancers Are Far from the Genes They Regulate

Eukaryotic genes often have more than one enhancer. Enhancers can be located either 5' or 3' to the gene they regulate or in introns and can be located tens of thousands of base pairs away from the promoter.

inserted different versions of the gene into mouse cells, which do not normally contain the antibody-producing gene. Some mouse cells received normal copies of the gene; some received copies of the gene missing different portions of the intron; some received no gene (step 4). Finally, the researchers analyzed the mRNAs that were produced and compared them with mRNA from a normal human cell and from mouse cells that received the normal gene. If their hypothesis was correct and a regulatory sequence was located inside the intron, then some of the modified genes should lack that sequence. Transcription should be reduced in genes lacking the regulatory sequence.

As the "Results" of Figure 18.6 shows, some of the modified copies of the gene were not transcribed at all. The message is clear: The gene is not transcribed if certain parts of the intron are missing. Tonegawa and co-workers proposed that the intron contains a regulatory sequence that is required for transcription to be activated.

This result was remarkable for two reasons: (1) the regulatory sequence was thousands of bases away from the promoter, and (2) it was downstream of the promoter instead of upstream. Regulatory elements that are far from the promoter are termed **enhancers**. Follow-up work has shown that enhancers occur in all eukaryotes and that they have several key characteristics:

- Enhancers can be more than 100,000 bases away from the promoter. They can be located in introns or in untranscribed 5' or 3' sequences flanking the gene (see Figure 18.7). Researchers have yet to find enhancers located in exons.
- Like promoter-proximal elements, many types of enhancers exist. Different genes are associated with different enhancers.
- Enhancers can work even if their normal 5' → 3' orientation is flipped.
- Enhancers can work even if they are moved to a new location in the vicinity of the gene, on the same chromosome.

Enhancers are regulatory sequences that are unique to eukaryotes. In addition, eukaryotic genomes contain regulatory sequences that are similar in structure to enhancers but opposite in function. These sequences are **silencers**. When silencers are active, they shut down transcription.

Once enhancers and silencers had been characterized, it was clear that they represented a type of regulatory sequence that worked very differently than promoter-proximal elements did.

For example, there was good evidence that GAL4 and other types of regulatory proteins bind to promoter-proximal elements and interact directly with TBP or RNA polymerase. These interactions stabilize binding and promote transcription. But how could a regulatory sequence that is distant from the promoter help initiate transcription?

How Do Enhancers Work?

An experiment that followed up on Tonegawa's work with antibody-producing genes supported the hypothesis that enhancers are binding sites for proteins that regulate transcription. As Figure 18.8 shows, researchers cut the enhancer sequence from an antibody-producing gene and spliced it into copies of the gene that codes for the protein β -globin. When this modified gene was inserted into antibody-producing cells, which normally do not express the gene for β -globin, the β -globin mRNA was produced. This result suggests that antibody-producing cells contain some factor that interacts specifically with the antibody-producing gene's enhancer to induce transcription. Follow-up work showed that the factor is a regulatory protein. Enhancers are regulatory sequences where regulatory proteins bind.

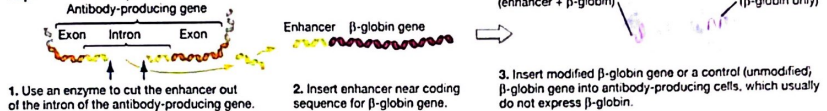
By analyzing mutant yeast, fruit flies, and roundworms that have defects in the expression of particular genes, biologists have identified a large number of regulatory proteins that bind to enhancers and silencers. This research program has supported one of the most general conclusions about gene regulation in eukaryotes: In multicellular species, different types of cells express different genes because they contain different regulatory proteins. These regulatory proteins, in turn, are produced in response to signals that arrive from other cells early in embryonic development. For example, a signaling molecule might arrive at a cell early in development and trigger the production of regulatory proteins that are specific to muscle cells. Because the regulatory proteins bind to specific enhancers and silencers and promoter-proximal elements, they trigger the production of muscle-specific proteins (Figure 18.9). If no "become a muscle cell" signal arrives, then no muscle-specific regulatory proteins are produced and no muscle-specific gene expression takes place.

To summarize, differential gene expression is based on the production of specific regulatory proteins. Eukaryotic genes are turned on when specific regulatory proteins bind to enhancers and promoter-proximal elements; the genes are turned off

Question: What is the role of enhancers in gene expression?

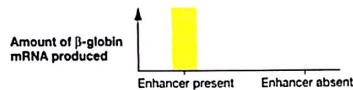
Hypothesis: Enhancers are binding sites for regulatory proteins that activate transcription.

Null hypothesis: Enhancers are not binding sites for regulatory proteins that activate transcription.

Experimental setup:

Prediction: β -Globin mRNAs will be produced in the cells with enhancer present but not in the control.

Prediction of null hypothesis: β -Globin mRNAs will not be produced in either group of cells.

Results:

Conclusion: Antibody-producing cells must contain a factor that interacts specifically with the enhancer for the antibody-producing gene and results in transcription.

FIGURE 18.8 Evidence That Enhancers Are Involved in Tissue-Specific Expression of Genes

Researchers attached the enhancer sequence from an antibody-producing gene to the coding sequence for the β -globin gene. The altered gene was expressed in antibody-producing cells, which normally do not produce β -globin.

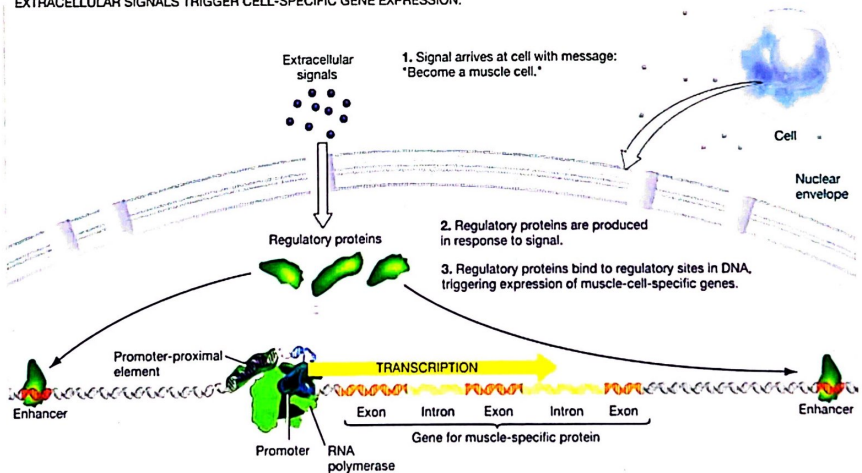
EXTRACELLULAR SIGNALS TRIGGER CELL-SPECIFIC GENE EXPRESSION.

FIGURE 18.9 Why Are Certain Proteins Produced Only in Certain Types of Cells?

Signaling molecules from outside the cell trigger the production or activation of cell-specific regulatory proteins, which then influence transcription by binding to enhancers, silencers, or promoter-proximal elements in DNA.

when regulatory proteins bind to silencers. Distinctive regulatory proteins are what make a muscle cell a muscle cell and a bone cell a bone cell.

How do the interactions between regulatory proteins and DNA sequences control transcription? As we'll see in the next section, the initiation of transcription is a marvelously complex process in eukaryotes. It requires the coordinated activity of many types of proteins interacting with each other and a variety of DNA sequences.

✓CHECK YOUR UNDERSTANDING

Eukaryotic genes have regulatory sequences called promoter-proximal elements close to their promoters and regulatory sequences called enhancers or silencers far from their promoters. You should be able to (1) explain the logic of the experiment that led to the discovery of an enhancer in an intron, and (2) compare and contrast regulatory sequences in the DNA of bacteria versus those in eukaryotes.

18.4 Transcription Initiation

After several decades of research, many questions remain about how transcription is initiated in eukaryotes. It is clear, however, that two broad classes of proteins interact with regulatory sequences at the start of transcription: (1) regulatory transcription factors and (2) basal transcription factors. **Regulatory transcription factors** are proteins that bind to enhancers and promoter-proximal elements. These transcription factors are responsible for the expression of particular genes in particular cell types and at particular stages of development. **Basal transcription factors**, in contrast, interact with the promoter. Basal transcription factors must be present for transcription to occur, but they do not provide much in the way of regulation. TBP is a basal transcription factor that is common to all genes. Other basal transcription factors are specific to promoters recognized by RNA polymerase I, II, or III.

In addition, proteins called **coactivators** are involved in starting transcription. Coactivators do not bind to DNA. Instead they link the proteins involved in initiating transcription—regulatory transcription factors and basal transcription factors.

Figure 18.10 summarizes the current model for how transcription is initiated in eukaryotes. The first step is the binding of regulatory transcription factors to DNA; these factors recruit chromatin-remodeling complexes and histone acetyl transferases (HATs). The result is chromatin remodeling, a loosening of the chromatin structure. Once the chromatin-remodeling complexes and HATs are in place, they open a broad swath of chromatin that includes the promoter region (step 2).

Once chromatin's grip is relaxed, other regulatory transcription factors can bind enhancers and promoter-proximal

elements. When these regulatory transcription factors are bound to DNA, they interact with basal transcription factors. Because enhancers are located far from the promoter, the regulatory transcription factors are normally far from the basal transcription factors. For the two sets of proteins to make contact, DNA has to loop out and away from the promoter, as shown in Figure 18.10 (step 3). When all of the basal transcription factors have assembled at the promoter in response to interactions with regulatory transcription factors and co-activators, they form a multi-protein machine called the **basal transcription complex**. The basal transcription complex then "recruits" RNA polymerase II so that transcription can begin (step 4).

Figure 18.11 provides a closer look at the basal transcription complex as it is assembled at the promoter. Construction of the complex begins when TBP binds to the TATA box. In a multistep process, as many as 60 other proteins then assemble around the DNA-bound TBP. The assembly process depends on interactions with regulatory transcription factors that are bound to enhancers, silencers, and promoter-proximal elements; the result is the formation of a basal transcription complex that can position RNA polymerase II in a way that initiates transcription. Compared with the situation in bacteria, where just 3 to 5 proteins may interact at the promoter to initiate transcription, the state of affairs in eukaryotes is remarkably complicated.

Currently, biologists are focused on understanding exactly how regulatory and basal transcription factors interact to control the formation of the basal transcription complex. Given the number of different proteins involved, progress may be slow. Progress is important, however, for transcription initiation lies at the heart of gene expression. Careful regulation of transcription is critical not only to the development of embryos but also to the daily life of eukaryotes. Right now, cells throughout your body are starting and stopping the transcription of specific genes in response to signals from nearby and distant cells. As the environment inside and outside your body continually changes, your cells continually change which genes are being transcribed.

✓CHECK YOUR UNDERSTANDING

Transcription initiation is a multistep process that begins when regulatory transcription factors bind to DNA and recruit proteins that open chromatin. Interactions between regulatory transcription factors and basal transcription factors result in the formation of the basal transcription complex and the arrival of RNA polymerase at the gene's start site. You should be able to compare and contrast the sigma proteins and regulatory proteins observed in bacteria with the basal and regulatory transcription factors found in eukaryotes.

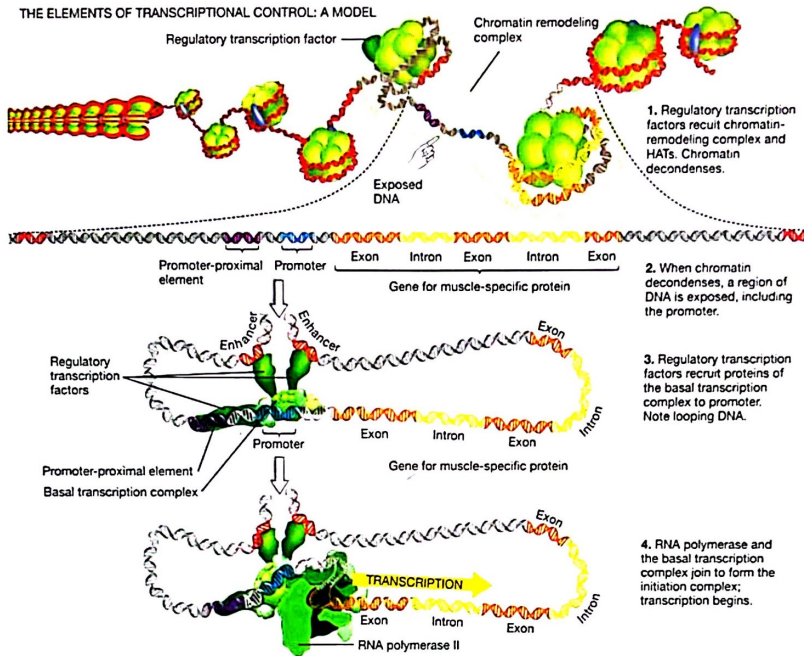


FIGURE 18.10 The Elements of Transcriptional Control: A Model
According to the currently accepted model, transcription is initiated through a series of steps.

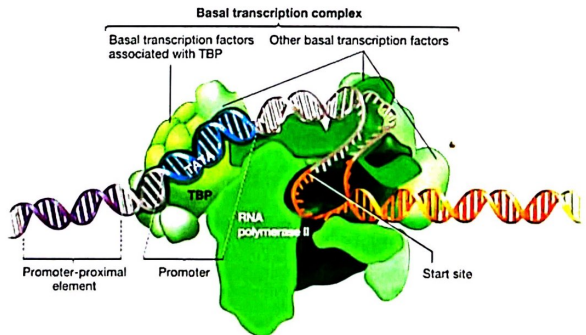


FIGURE 18.11 Many Proteins Interact at the Start Site for Transcription
Basal transcription factors associate with TBP when it binds to the promoter. The basal transcription complex also includes other basal transcription factors and RNA polymerase II.

18.5 Post-Transcriptional Control

In the process of regulating gene expression, chromatin re-modeling and transcription are just the start of the story. Once an mRNA is made, a series of events have to occur if the final product is going to affect the cell. Each of these events offers an opportunity to regulate gene expression, and each is used in some cells at least some of the time. Control points include splicing mRNAs in various ways, altering the rate at which translation is initiated, and modifying the life span of mRNAs and proteins after translation has occurred. Let's consider each in turn.

Alternative Splicing of mRNAs

Introns are spliced out of primary RNA transcripts while the message is still inside the nucleus. The mRNA that results from splicing consists of sequences that are encoded by exons and are protected by a cap on the 5' end and a long poly (A) tail on the 3' end. Recall from Chapter 16 that splicing is accomplished by the molecular machines called *spliceosomes*. What that chapter did not mention, however, is that splicing provides an opportunity for the regulation of gene expression.

During splicing, changes in gene expression are possible because selected exons, as well as introns, may be removed. As a result, the same primary RNA transcript can yield mature, processed mRNAs with several different combinations of transcribed exons. This is important, because, if the sequence of ribonucleotides in the mature mRNAs can vary, then the polypeptides translated from those mature mRNAs can also vary. When the same primary RNA transcript is spliced in dif-

ferent ways to produce different mature mRNAs and thus different proteins, **alternative splicing** is said to occur.

To see how alternative splicing works, consider the muscle-cell protein tropomyosin. The tropomyosin gene is expressed in both skeletal muscle and smooth muscle, which are distinct muscle types. (Skeletal muscle is responsible for moving your bones; smooth muscle lines many parts of your gut and certain blood vessels. The two types of muscle are composed of distinct types of muscle cells.) As Figure 18.12a shows, the primary transcript from the tropomyosin gene contains 14 exons. In skeletal muscle and smooth muscle cells, different subsets of the 14 exons are spliced together to produce two different messages for translation (Figure 18.12b). Each mature mRNA contains information from a different combination of exons. As a result of alternative splicing, the tropomyosin proteins found in these two cell types are distinct. One of the reasons that skeletal muscle and smooth muscle are different is that they contain different types of tropomyosin.

Alternative splicing is controlled by regulatory proteins that bind to mRNAs in the nucleus and interact with spliceosomes. When cells that are destined to become skeletal muscle or smooth muscle are developing, they receive signals leading to the production of specific proteins that are active in the regulation of splicing. Instead of transcribing different versions of the tropomyosin gene, the cells splice the same primary RNA transcript but in different ways.

Although alternative splicing of RNA was viewed as a curiosity until about a decade ago, attitudes have changed. Alternative splicing is now considered to be a major mechanism in the control of gene expression in multicellular eukaryotes. In

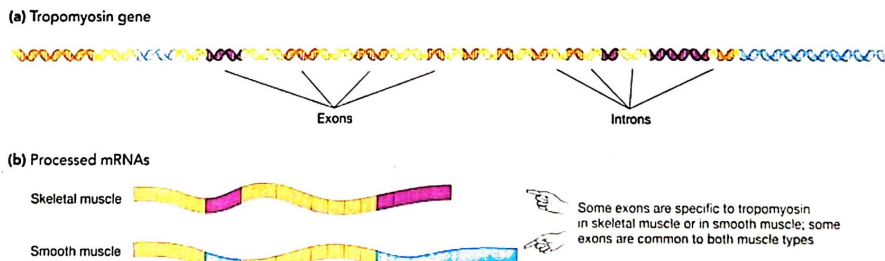


FIGURE 18.12 Alternative Splicing Produces More than One Mature mRNA from the Same Gene
(a) The tropomyosin gene has a large series of exons and introns. The purple exons are found in skeletal muscle tropomyosin, and the blue exons are found in smooth muscle tropomyosin. The orange exons are found in the tropomyosin of both types of muscle cells. (b) In skeletal and smooth muscle cells, alternative splicing leads to the production of distinct mature mRNAs for the tropomyosin protein.

fruit flies, for example, the gene that determines gender is transcribed into the same primary RNA transcript in both males and females. A cascade of regulated splicing in females ultimately results in the production of a protein called *doublesex*, which represses the transcription of male-specific genes. In males, an alternative series of splicing events results in a version of the *doublesex* protein that represses the transcription of female-specific genes.

Alternative splicing makes it possible to produce more than one protein from a single gene. Before the importance of alternative splicing was recognized, a gene was considered to be a nucleotide sequence that encodes a single protein or RNA. Now genes have to be thought of as DNA sequences that are capable of directing the production of one or more related polypeptides or RNAs. As a result, the number of proteins that an organism can produce far exceeds the number of genes it has. For example, at least 35 percent of human genes undergo alternative splicing. Although the human genome contains an estimated 40,000 genes, the number of different proteins that your cells can produce is believed to be between 100,000 and 1 million.

The current record holder for the number of mRNAs derived from one gene is the *Dscam* gene in the fruit fly *Drosophila melanogaster*. The products of this gene are involved in guiding growing nerve cells within the embryo. Because the primary transcript is spliced into about 38,000 distinct forms of mRNA, the *Dscam* gene can produce about 38,000 different products. Researchers are testing the hypothesis that particular combinations of *Dscam* proteins allow particular nerve cells to connect to each other in the fly's brain.

Current research on alternative splicing is focused on two topics: (1) understanding how regulatory proteins control the process and (2) documenting the extent to which alternative splicing enriches the genetic information stored in a genome. Stay tuned.

Translational Control

Once splicing is complete and processed mRNAs are exported to the cytoplasm, several new regulatory mechanisms come into play prior to translation. The following examples illustrate the array of mechanisms that influence the timing, or extent of translation:

- The life span of an mRNA in the cell can vary. For example, the mRNA for casein—the major protein in milk—is produced in the mammary gland tissue of female mammals. Normally this mRNA has a half-life of just over 1 hour, and little casein protein is produced. But when a female mouse is lactating, regulatory molecules increase the mRNA's half-life to 28.5 hours—leading to a huge increase in the production of casein protein.

- In many species of animals, eggs are loaded with mRNAs that are not translated until fertilization occurs. The proteins produced from these messages are involved in directing the early development of the embryo, so they are not needed until fertilization is complete. Translation of the mRNAs in the egg can be prevented when a regulatory protein binds to them, or when the message's cap or tail is modified.
- The overall rate of translation in a cell may slow or stop in response to a sudden increase in temperature or infection by a virus. The slowdown occurs because regulatory proteins add a phosphate group to a protein that is part of the ribosome. As we have pointed out in earlier chapters, phosphorylation frequently leads to changes in the shape and chemical reactivity of proteins. In the case of the phosphorylated ribosomal protein, the shape change slows or prevents translation. For the cell, this dramatic change in gene expression can mean the difference between life and death. If the danger is due to a sudden increase in ambient temperature, shutting down translation prevents the production of improperly folded polypeptides; if the insult is a virus, the cell avoids manufacturing viral proteins.

Post-Translational Control

Control of gene expression may continue even after translation occurs and a protein product is complete. Recall from Chapter 17 that, in bacteria, mechanisms of post-translational regulation are important because they allow the cell to respond to new conditions rapidly. The same is true for eukaryotes. Instead of waiting for transcription and translation to occur, the cell can respond to altered conditions by quickly activating or inactivating existing proteins. Regulatory mechanisms occurring late in the flow of information from DNA to RNA to protein involve a trade-off between speed and resource use, however, because transcription and translation require energy and materials.

A group of regulatory transcription factors that is found in mammals is called **signal transducers and activators of transcription (STATs)** and provides a good example of post-translational control. STATs participate in several signal transduction pathways. Recall from Chapter 8 that signal transduction pathways are interactions that convert a signal at the cell surface into a response by the cell. In white blood cells, STAT activity results in increased cell growth and cell division.

In white blood cells, STATs normally reside in the cytoplasm as single polypeptide chains. In this form, the STATs are inactive. But a chain of events is triggered when a signaling molecule binds to and activates a receptor on the cell

surface (Figure 18.13a). The activated receptor adds a phosphate to a STAT polypeptide chain, triggering the formation of a dimer (a unit made of two parts). The activated STAT dimer then moves into the nucleus, binds to an enhancer, and activates transcription (Figure 18.13b). The STAT remains active until the phosphate is removed by another regulatory protein. At this point, the two polypeptide chains come apart and are transported back to the cytoplasm.

Phosphorylation is a very common mechanism of post-translational control over gene expression, particularly in signal transduction pathways. It is only one of several post-translational control mechanisms that have been documented, however. The activity of a protein may also be modified by folding or by enzymes that cleave off a portion of the molecule. But no matter what the mechanism, post-translational control is correlated with particularly rapid changes in gene expression.

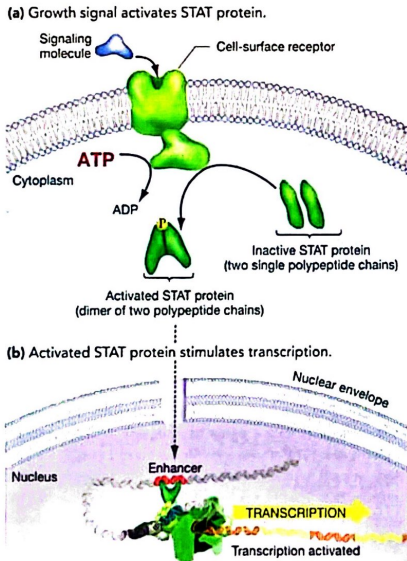


FIGURE 18.13 STATs Activate Transcription after They Are Phosphorylated

(a) A series of events, beginning with the binding of a signaling molecule to a cell-surface receptor, leads to the phosphorylation of STAT proteins. This causes two STAT polypeptide chains to join, forming an activated dimer. (b) Activated STAT proteins move to the nucleus, bind to DNA, and activate the transcription of selected target genes.

A 50-Year Retrospective: How Does Gene Expression in Bacteria and in Eukaryotes Compare?

For about five decades, biologists have been studying how gene expression is controlled. Almost as soon as they knew that information in DNA is transcribed into RNA and then translated into proteins, researchers began asking questions about how that flow of information is regulated. Early studies on bacteria established the existence of promoters, the structure and function of regulatory proteins and regulatory sequences, and the principles of positive control and negative control. More recent work has highlighted the importance of chromatin structure and alternative splicing in eukaryotes.

What are the fundamental differences in how gene expression is regulated in bacteria and eukaryotes? Biologists point to four primary differences, two of which involve levels of control that exist in eukaryotes but not in bacteria:

1. **Packaging** The chromatin structure of eukaryotic DNA must be opened in order for TBP, the basal transcription complex, and RNA polymerase to gain access to genes and initiate transcription. A key insight here is that, because eukaryotic DNA is packaged so tightly, the default state of transcription in eukaryotes is off. In contrast, the default state of transcription in bacteria, with their supercoiled DNA and freely accessible promoters, is on.
2. **Alternative splicing** Prior to translation, primary transcripts in eukaryotes must be spliced—an event that does not occur in bacteria. The fundamental consequence is that the one-to-one correspondence between the number of genes and the number of gene products observed in bacteria is not seen in eukaryotes. Instead, each eukaryotic gene may code for one to thousands of distinct products.
3. **Complexity** Transcriptional control is much more complex in eukaryotes than in bacteria. The function of sigma proteins in bacteria is analogous to the role of the basal transcription complex in eukaryotes. Likewise, the function of CAP, the repressor, and other regulatory proteins is analogous to the role of regulatory transcription factors in eukaryotes. But the sheer number of eukaryotic proteins involved in regulating transcription—and the complexity of their interactions—dwarfs those in bacteria.
4. **Coordinated expression** In bacteria, genes that are involved in the same cellular response are organized into operons controlled by a single promoter. Because their polycistronic mRNAs are translated together, several proteins are produced in a coordinated fashion. In contrast, operons are rare in eukaryotes. In these organisms, genes that are physically scattered can be expressed at the same time because a single set of regulatory transcription factors can trigger the transcription of several genes. For example, muscle-specific genes found on several different chromosomes can be transcribed in response to the same muscle-specific regulatory transcript.

tion factor. In this way, eukaryotes coordinate the expression of functionally related genes.

To explain why these differences exist, biologists point to a fundamental contrast between unicellular and multicellular individuals. In both types of organisms, cells have to respond to environmental changes in an appropriate way. Bacteria and yeast cells change gene expression to respond to changes in the availability of sugars and other nutrients; your muscle cells change gene expression in response to changes in temperature, the availability of oxygen and sugars, and general activity. But in addition, the cells of multicellular organisms have to differentiate as an individual develops. Changes in gene expression are responsible for the differentiation of muscle cells, bone cells, leaf cells, and flower cells in response to signals from other cells. The need for each cell type to have a unique pattern of gene expression is thought to be the reason why control of gene expression is so much more complex in eukaryotes than in bacteria.

One of the two great frontiers in research on gene expression is to understand how developmental signals produce cell-specific gene expression in multicellular organisms. The other major frontier is to understand how certain defects in gene regulation result in uncontrolled cell growth and the suite of diseases called cancer.

18.6 Linking Cancer with Defects in Gene Regulation

Normal regulation of gene expression results in the orderly development of an embryo and appropriate responses to environmental changes. Abnormal regulation of gene expression, in contrast, can lead to developmental abnormalities and diseases such as cancer. Hundreds of distinct cancers exist. These diseases are enormously variable regarding the tissues they affect, their rate of progression, and their outcome. Because the underlying defects, symptoms, and consequences are so diverse, cancer is not a single disease but a family of related diseases. Cancers are related because they all result from uncontrolled cell growth.

From this perspective, the fundamental question in cancer biology is a simple one: What causes uncontrolled cell growth? The short answer is that each type of cancer is caused by a different set of defects that lead to uncontrolled cell growth. The longer and more accurate answer is based on two key observations. First, recall from Chapter 11 that cancer results from defects in the proteins that control the cell cycle. Second, cancer is associated with mutations that knock out key genes. For example, Chapter 14 linked the cancer-causing disorder xeroderma pigmentosum and hereditary nonpolyposis colorectal cancer with defects in DNA repair. This association supports the hypothesis that increased mutation rates are involved in triggering cancer. The hypothesis is also supported by the observation that people

who have been exposed to **mutagens**—that is, radiation or chemicals that induce mutation—have an increased risk of developing cancer. For example, cancer rates were extremely high in the tens of thousands of people who were exposed to massive doses of radiation from the atomic bombs dropped on Hiroshima and Nagasaki, Japan, at the end of World War II. The link between mutation and cancer is so strong that suspected **chemical carcinogens** (cancer-causing compounds) are first tested by their ability to cause mutations in bacteria.

The fundamental question of cancer biology comes into sharper focus once the role of cell-cycle regulators and mutation is acknowledged. Now we can ask, Which genes, when mutated, disrupt the cell cycle and trigger uncontrolled cell growth?

Intensive research over the past two decades has shown that many cancers are associated with mutations in regulatory transcription factors or regulatory sequences in DNA. These mutations lead to cancer when they affect one of two classes of genes: (1) genes that stop or slow the cell cycle, and (2) genes that trigger cell growth and division by initiating specific phases in the cell cycle.

Genes that stop or slow the cell cycle are called **tumor suppressor genes**. Their products prevent the cell cycle from progressing unless specific signals indicate that conditions are right for moving forward with mitosis and cell division. If a mutation disrupts normal function of a tumor suppressor gene, then a key “brake” on the cell cycle is eliminated.

Genes that encourage cell growth by triggering specific phases in the cell cycle are called **proto-oncogenes** (“first-cancer-genes”). In normal cells, proto-oncogenes are required to initiate each phase in the cell cycle. They are active only when conditions are appropriate for growth, however. In cancerous cells, defects in the regulation of proto-oncogenes cause these genes to stimulate growth at all times. In cases such as this, a mutation has converted the proto-oncogene into an **oncogene**—an allele that promotes cancer development.

To gain a deeper understanding of how defects in gene expression can lead to cancer, let us consider two examples from recent research. The first, the *p53* gene, illustrates how problems with regulating tumor suppressor genes can lead to uncontrolled cell growth. The second, *STAT* mutations, demonstrates how defects in regulatory transcription factors can lead to the constant activation of proto-oncogenes, constant cell growth, and tumor formation.

p53—Guardian of the Genome

The gene that is most often defective in human cancers codes for a regulatory transcription factor. This gene is called *p53*, because the protein it codes for has a molecular weight of approximately 53 kilodaltons (53,000 daltons, or 53,000 amu—see Chapter 2). DNA sequencing studies have revealed that mutant, nonfunctional forms of *p53* are found in over half of all human cancers.

What is the link between a loss of p53 protein activity and cancer? A key observation suggested an answer to this question: When researchers exposed normal, noncancerous human cells to UV radiation, levels of p53 protein increased markedly. Recall from Chapter 14 that UV radiation damages DNA. Follow-up studies confirmed that there is a close correlation between DNA damage and the amount of p53 in a cell. In addition, analyses of the protein's structure showed that it contains a DNA-binding domain.

These observations inspired the hypothesis that p53 is a transcription factor that serves as the master brake on the cell cycle (Figure 18.14a). In this model, p53 is activated after DNA damage occurs. The hypothesis contends that the activated protein binds to the enhancers of genes that arrest the cell cycle. Once the cell cycle is arrested, the cell has time to repair its DNA before continuing to grow and divide.

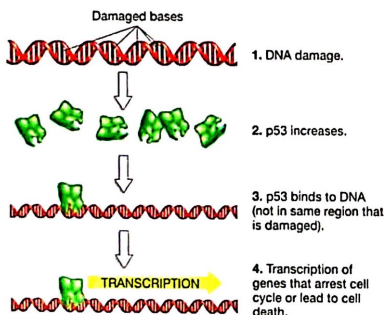
Recent research has shown that this model of p53 function is correct in almost every detail. For example, consider the results of studies based on X-ray crystallography. The three-dimensional models generated by this technique confirmed that p53 binds directly to DNA (Figure 18.14b). In addition, researchers who mapped the location of mutations that make p53 defective found that virtually all of the cancer-causing mutations were located in p53's DNA-binding site (see the gray regions of Figure 18.14b). This observation supports the hypothesis that defective forms of the protein can't bind to en-

hancers. Further, investigators have documented that one of the genes induced by p53 encodes a protein that binds to the cyclin-Cdk complexes introduced in Chapter 11. Binding by this protein prevents the cyclin-Cdk complexes from triggering M (mitosis) phase in the cell cycle.

Recent research has also shown that when a cell's DNA is extensively damaged and cannot be repaired, p53 activates the transcription of genes that cause the cell to take its own life via apoptosis (see Chapter 11). But if mutations in the p53 gene make the protein product inactive, then damaged cells are not shut down or killed. They continue to move through the cell cycle, except now they are likely to contain many mutations because of the DNA damage they have left unrepaired. If these mutations create oncogenes or inactivate genes for tumor suppressors, the cells have taken a key step on the road to cancer.

To summarize, p53 functions as a tumor suppressor gene. It prevents cancer onset by stopping the cell cycle when DNA is damaged. When p53 is functioning normally, mutations that produce oncogenes are repaired or eliminated before uncontrolled cell growth can begin. The role of p53 in preventing cancer is so fundamental that biologists call this gene "the guardian of the genome." Currently, research on the p53 protein is forging ahead on two fronts: Biologists are striving to identify more of the genes that are regulated by this protein and to find molecules that could act as anticancer drugs by mimicking p53's shape and activity.

(a) p53 MODE OF ACTION



(b) DNA-binding domain of p53

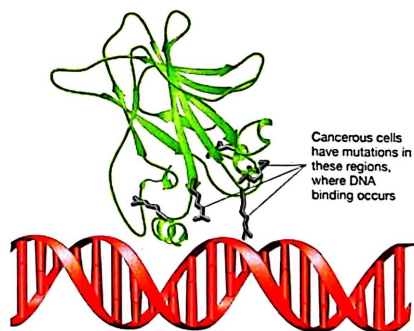


FIGURE 18.14 p53 Is a DNA-Binding Protein That Acts as a Tumor Suppressor

(a) Summary of the current model for p53's mode of action. (b) Ribbon model showing the DNA-binding domain of p53. In cancer patients, the amino acids in the gray portions of the molecule are often different than the normal versions. **QUESTION** Why would substituting one amino acid for another at one of the locations highlighted in part (b) affect a protein's ability to bind to DNA? If the model in part (a) is correct, what happens when p53 cannot bind to DNA efficiently?

STAT Mutations in Cancer

The family of transcription activators called STATs is another type of protein implicated in cancer. Recall from Section 18.5 that, in several types of cells, STATs activate genes that promote cell growth in response to signals from outside the cell. In normal cells, STATs must be phosphorylated before they can bind to regulatory sequences in DNA and activate transcription. But in some human cancers, STATs bind to DNA all the time, or constitutively—not just when they are phosphorylated. When this happens, STATs constantly stimulate cell growth and division. These observations suggest that mutations can convert STATs into oncogenes.

What type of mutations could turn STATs into oncogenes? To answer this question, recall that the activated form of a STAT is a dimer—a combination of two polypeptides. For a dimer to form, an amino acid residue at the point of contact between the two STAT proteins must be phosphorylated (Figure 18.15a).

These observations inspired a hypothesis that STATs turn into oncogenes if mutations cause them to form dimers without

being phosphorylated—that is, without being stimulated by a normal growth signal. More specifically, researchers suggested that if mutations caused cysteine residues to be substituted in the portion of the protein where dimer formation takes place, then disulfide bonds might form between STAT3 monomers, creating constitutively activated dimers (Figure 18.15b).

To test this hypothesis, researchers created mutant versions of the STAT3 protein with cysteines in the appropriate location. The researchers expressed the protein in normal (control) cells and analyzed the amount of protein produced by a gene regulated by STAT3. Compared with cells with normal STAT3 proteins, the cysteine-containing mutant cells produced 10 times as much of the target gene product (Figure 18.15c). This observation was consistent with the biologists' prediction that cysteine mutants form dimers and bind to DNA constitutively.

To confirm their result, the biologists injected mice either with cells that expressed cysteine-mutant STAT3 or with cells that expressed normal STAT3. After just four weeks, all of the

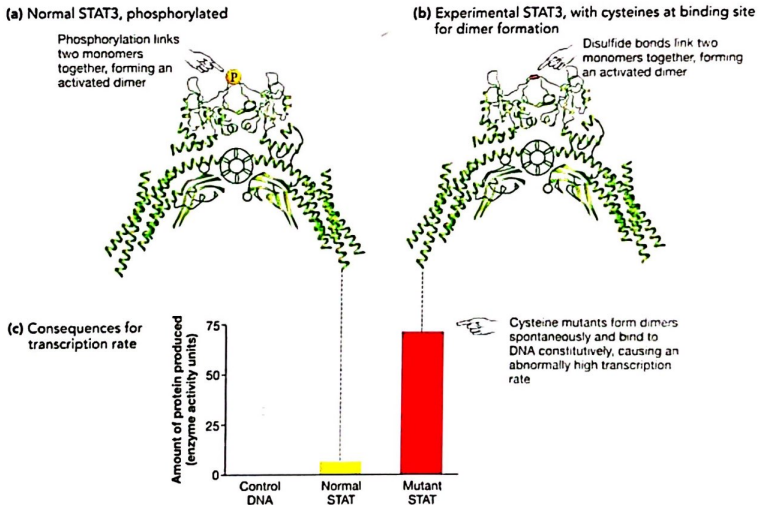


FIGURE 18.15 Generating a Cancer-Causing Mutation Experimentally

(a) Phosphorylated sections of STAT3 proteins bind to one another, forming a dimer. (b) Substitution of cysteines at the locations shown in red result in the formation of disulfide bonds between STAT3 proteins. As a result, the STAT proteins are activated at all times. (c) Histogram showing the transcriptional activity of a gene that is normally activated by STAT3. The data come from cells that received either a control DNA (a gene that does not encode a transcription activator), a normal STAT3 gene, or a copy of the cysteine-mutant version of the STAT3 gene.

mice that had been injected with mutant cells had developed tumors at the site of injection. In contrast, none of the mice that had been injected with normal cells showed evidence of tumors, even after eight weeks.

These experiments have an important message: Any mutation that constitutively forms dimers of STAT proteins causes constitutive activation of STAT target genes. As a result, these mutations set the stage for cancer. An even more global message is that constant activation of a regulatory transcription factor can lead to cancer. Cancers originate with mutations that disrupt the normal regulation of cell-cycle control genes.

✓CHECK YOUR UNDERSTANDING

Cancer is associated with mutations and with loss of control over the cell cycle. These two observations are connected, because mutations can cause cancer when they occur in genes that are involved in regulating cell-cycle control genes. For example, when a protein that activates the cell cycle is activated constitutively because of a mutation in a regulatory gene, then uncontrolled cell growth may occur. You should be able to explain why knock-out (loss-of-function) mutations in p53 occur in so many cancers.

ESSAY Fly Eyes and Other Strange Tales of Gene Regulation

Imagine a 1950s-vintage science-fiction movie. As thunder cracks and lightning flashes, a mad-eyed scientist working in the bowels of a sinister-looking laboratory inserts a mouse gene into a fruit fly's genome. The next day, the scientist cackles with delight as she discovers eyes sprouting in bizarre locations on the fly. But then the fly escapes. It mates with a normal fly and sires offspring with hundreds of eyes. Thanks to their remarkable 360° vision, no predator or flyswatter can stop the little monsters. Clouds of them soar off, bent on global domination.

Preposterous? A team of not-so-mad-eyed scientists has inserted a mouse gene into fruit flies and observed the formation of eyes on the adult fly's legs, antennae, and torso (Figure 18.16). The gene in question is a regulatory transcription factor. When the gene product binds to enhancers, it sets off a cascade of gene expression events that are required for eye development.

Walter Gehring and colleagues were first to isolate and characterize the gene involved. It was called *eyeless*, because it was

originally discovered as a mutant allele that causes flies to develop without eyes. When Gehring's team created *Drosophila* embryos that expressed the normal form of the gene in locations that naturally give rise to appendages, eyes developed on the legs

When eyeless is activated in leg cells, the leg cells give rise to eyes.

and antennae. This result underscores an important conclusion: During development, the fate of a cell depends on the activity of regulatory transcription factors. When *eyeless* is activated in leg cells, the leg cells give rise to eyes.

When the researchers looked for genes in other organisms that were similar to *eyeless* in DNA sequence, the team found them in a wide array of species. Consider a mouse gene called *Small Eye* or *pax6*, which is similar in sequence to *eyeless*. The name *Small Eye* was inspired by the observation that a mutant allele of this gene leads to reduced eye development. Because *eyeless* and *Small Eye* are similar in structure and function, Gehring hypothesized that they are homologous—that the genes are descended from a similar gene in a common ancestor of flies and mice, which, according to the fossil record, lived some 550 million years ago. To test this hypothesis, Gehring inserted the mouse gene into *Drosophila* embryos. *Small Eye* induced the development of fly eyes essentially anywhere in the embryo that it was expressed. He got the same result when he inserted a squid gene that was similar to *eyeless*.

What does this tell us about biology? First, it is reasonable to conclude that *eyeless*, *Small Eye*, and like genes in other species are similar in structure and function because they are homologous. Second, these regulatory transcription factors and their recognition sequences in DNA have been so conserved that the proteins are nearly interchangeable among organisms separated by 550 million years of evolution. Third, one key regulatory transcription factor can unleash a cascade of responses that results in a structure as complex as an eye. Fourth, the same starting point—expression of the *eyeless* gene—can lead to very different ends, depending on the context. In flies, the mouse gene *Small Eye* leads to a fly eye, not a mouse eye.

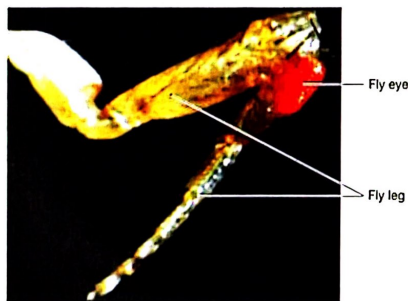


FIGURE 18.16 *Eyeless* Encodes a Transcription Factor That Induces Eye Development

When *Drosophila* embryos are made to express the *eyeless* gene in cells that gave rise to the legs, adults develop with eyes on their legs.

CHAPTER REVIEW

Summary of Key Concepts

- Changes in gene expression allow eukaryotic cells to respond to changes in the environment and cause distinct cell types to develop.

In a multicellular eukaryote, different types of cells are radically different in size, shape, and function—even though they have the same DNA. Cells are different not because they have different genes but because they express different genes.

In embryos, cells begin expressing specific genes in response to signals from other cells. Because different cells receive different signals, one cell might begin expressing muscle-specific genes while a nearby cell expresses bone-specific genes. Once cells mature, they continue to receive signals that are released by other cells and that carry information about changes in the environment. In response, changes in gene expression may occur.

In both embryonic cells and mature cells, several types of changes in gene expression may occur. The transcription of specific genes may be initiated or repressed, mRNAs may be spliced in different ways to produce a different product, the life span of specific mRNAs may be extended or shortened, and the life span or activity of particular proteins may be altered.

- In eukaryotes, DNA is packaged with proteins into complex structures that must be opened before transcription can occur. Eukaryotic DNA is wrapped around histone proteins to form a bead-like nucleosome, which is then coiled into 30-nm fibers and higher-order chromatin structures. Transcription cannot be initiated until the interaction between DNA and histones in chromatin is relaxed. These changes depend on the acetylation of histones and the action of chromatin-remodeling complexes.
- In eukaryotes, transcription can be initiated only when specific proteins bind to the promoter and to regulatory sequences that may be close to the promoter or far from it.

Regulatory transcription factors are proteins that bind to regulatory sequences called enhancers and silencers, which

are often located at a distance from the gene in question, or to promoter-proximal sequences near the start of the coding sequence. The first regulatory transcription factors that bind to DNA recruit proteins that loosen the histones' grip on the gene, making the promoter accessible to basal transcription factors. Interactions between regulatory and basal transcription factors lead to the formation of the basal transcription complex. Once this large, multi-protein complex is intact, RNA polymerase is recruited to the site and transcription begins.

Web Tutorial 18.1 Transcription Initiation in Eukaryotes

- Alternative splicing allows a single gene to code for several different products.

Once a message has been transcribed, several other regulatory events come into play. Alternative splicing allows a single locus to produce more than one mRNA and more than one protein. Mechanisms of post-translational control, such as the activation of transcription factors by phosphorylation, trigger rapid responses to signals from outside the cell.

- Cancer can develop when mutations disable genes that regulate cell-cycle control genes.

The cell cycle is controlled by specific genes. But if mutations alter the regulatory sequences of cell-cycle control genes or the genes for their regulatory transcription factors, then uncontrolled cell growth and tumor formation may result. For example, the cell-cycle regulator *p53* is responsible for stopping the cell cycle when DNA is damaged. If *p53* is mutated in a way that prevents its protein product from binding to DNA, then the cell cycle is not arrested and the damaged DNA is not repaired—leading to mutations. If the mutations alter genes that control the cell cycle, such as *STAT* genes, cancer may occur. Mutant *STAT* proteins that promote cancer bind to DNA at all times and trigger the expression of genes that activate the cell cycle, leading to uncontrolled cell growth.

Questions

Content Review

1. What is chromatin?
 - a. the protein core of the nucleosome, which consists of histones
 - b. the 30-nm fiber
 - c. the DNA-protein complex found in eukaryotes
 - d. the histone and non-histone proteins in eukaryotic nuclei
2. What is a tumor suppressor?
 - a. a gene associated with tumor formation when its product does not function
 - b. a gene associated with tumor formation when its product functions normally
 - c. a gene that accelerates the cell cycle and leads to uncontrolled cell growth
 - d. a gene that codes for a transcription factor involved in tumor formation
3. Which of the following statements about enhancers is correct?
 - a. They contain a unique base sequence called a TATA box.
 - b. They are located only in 5'-flanking regions.
 - c. They are located only in introns.
 - d. They are found in a variety of locations and are functional in any orientation.
4. In eukaryotes, why are certain genes expressed only in certain types of cells?
 - a. Regulatory transcription factors vary from cell to cell.
 - b. The promoter sequence varies from cell to cell.
 - c. The location of enhancers varies from cell to cell.
 - d. The type of RNA polymerase varies from cell to cell.

400 Unit 3 Gene Structure and Expression

5. What is alternative splicing?
 - a. the phosphorylation events that lead to different types of post-translational regulation
 - b. mRNA processing events that lead to different combinations of exons being spliced together
 - c. folding events that lead to proteins with alternative conformations
 - d. action by regulatory proteins that leads to changes in the life span of an mRNA
6. What types of proteins bind to promoter-proximal elements?
 - a. the basal transcription complex
 - b. the basal transcription complex plus RNA polymerase
 - c. basal transcription factors
 - d. regulatory transcription factors

Conceptual Review

1. Compare and contrast (a) enhancers and the CAP site; (b) promoter-proximal elements and the *Lac* operon operator; (c) basal transcription factors and sigma; and (d) eukaryotic promoters and bacterial promoters.
2. Why does chromatin need to be "remodeled" for transcription to occur? Why was it logical to observe that DNA-protein interactions can be loosened by the addition of acetyl groups to histones?
3. Compare and contrast (a) enhancers and silencers; (b) promoter-proximal elements and enhancers; (c) transcription factors and coactivators; and (d) regulatory transcription factors and basal transcription factors.
4. Explain the logic behind the experiment by Tonegawa and colleagues that was responsible for the discovery of enhancers. Which treatments in the experiment served as controls, and what did they control for?
5. When the human genome was first sequenced, researchers estimated that only about 30,000 genes were present. Previously, most investigators expected that the human genome held at least 100,000 genes. Explain how alternative splicing can explain the discrepancy between the observed and predicted numbers of genes.
6. Explain why mutations in regulatory proteins and regulatory sequences can lead to loss of control over the cell cycle and the development of cancer. Include a specific example.

Group Discussion Problems

1. Histone proteins have been extremely highly conserved during evolution. The histones found in fruit flies and humans, for example, are nearly identical in amino acid sequence. Offer an explanation for this observation. (Hint: What are the consequences of a mutation in a histone?)
2. Cancers are most common in tissues where cell division is common, such as blood cells and cells in the lining of the lungs or gut. Why is this observation logical?
3. Levels of p53 protein in the cytoplasm increase after DNA damage. Design an experiment to determine whether this increase is due to increased transcription of the *p53* gene or to activation of preexisting p53 proteins via a post-translational mechanism such as phosphorylation.
4. The frequency of cancer was elevated in individuals who were exposed to the nuclear bombs that were dropped on Japan in World War II. How are these data related to the observation that individuals who experience severe sunburn or deep tanning are more likely to develop skin cancer?

Answers to Multiple-Choice Questions 1. c; 2. a; 3. d; 4. a; 5. b; 6. d

www.prenhall.com/tryeman is your resource for the following: Web Tutorials, Online Quizzes and other Online Study Guide materials; Answers to Conceptual Review Questions; Solutions to Group Discussion Problems; Answers to Figure Caption Questions and Exercises; and Additional Readings and Research.