

Regulation of Gene Expression



▲ **Figure 18.1** What regulates the precise pattern of gene expression in the developing wing of a fly embryo?

KEY CONCEPTS

- 18.1** Bacteria often respond to environmental change by regulating transcription
- 18.2** Eukaryotic gene expression is regulated at many stages
- 18.3** Noncoding RNAs play multiple roles in controlling gene expression
- 18.4** A program of differential gene expression leads to the different cell types in a multicellular organism
- 18.5** Cancer results from genetic changes that affect cell cycle control

OVERVIEW

Conducting the Genetic Orchestra

It's almost concert time! Dissonance reigns as the orchestra members individually tune their instruments. Then, after a brief hush, the conductor's baton rises, pauses, and begins a series of elaborate movements, directing specific instruments to

join in and others to raise or lower their volume at defined moments. Properly balanced and timed, discordant sounds are thus transformed into a beautiful symphony that enraptures the audience.

In a similar way, cells intricately and precisely regulate their gene expression. Both prokaryotes and eukaryotes must alter their patterns of gene expression in response to changes in environmental conditions. Multicellular eukaryotes must also develop and maintain multiple cell types. Each cell type contains the same genome but expresses a different subset of genes, a significant challenge in gene regulation.

An adult fruit fly, for example, develops from a single fertilized egg, passing through a wormlike stage called a larva. At every stage, gene expression is carefully regulated, ensuring that the right genes are expressed only at the correct time and place. In the larva, the adult wing forms in a disk-shaped pocket of several thousand cells, shown in **Figure 18.1**. The tissue in this image has been treated to reveal the mRNA for three genes—labeled red, blue, and green—using techniques covered in Chapter 20. (Red and green together appear yellow.) The intricate pattern of expression for each gene is the same from larva to larva at this stage, and it provides a graphic display of the precision of gene regulation. But what is the molecular basis for this pattern? Why is one particular gene expressed only in the few hundred cells that appear blue in this image and not in the other cells?

In this chapter, we first explore how bacteria regulate expression of their genes in response to different environmental conditions. We then examine how eukaryotes regulate gene expression to maintain different cell types. Gene expression in eukaryotes, as in bacteria, is often regulated at the stage of transcription, but control at other stages is also important. In recent years, researchers have been surprised to discover the many roles played by RNA molecules in regulating eukaryotic gene expression, a topic we cover next. We then consider what happens when a complex program of gene regulation works properly during embryonic development: A single cell—the fertilized egg—becomes a fully functioning organism made up of many different cell types. Finally, we investigate how cancer can result when gene regulation goes awry. Orchestrating proper gene expression by all cells is crucial to the functions of life.

CONCEPT 18.1

Bacteria often respond to environmental change by regulating transcription

Bacterial cells that can conserve resources and energy have a selective advantage over cells that are unable to do so. Thus, natural selection has favored bacteria that express only the genes whose products are needed by the cell.

Consider, for instance, an individual *E. coli* cell living in the erratic environment of a human colon, dependent for its nutrients on the whimsical eating habits of its host. If the environment is lacking in the amino acid tryptophan, which the bacterium needs to survive, the cell responds by activating a metabolic pathway that makes tryptophan from another compound. Later, if the human host eats a tryptophan-rich meal, the bacterial cell stops producing tryptophan, thus saving itself from squandering its resources to produce a substance that is available from the surrounding solution in prefabricated form. This is just one example of how bacteria tune their metabolism to changing environments.

Metabolic control occurs on two levels, as shown for the synthesis of tryptophan in **Figure 18.2**. First, cells can adjust the activity of enzymes already present. This is a fairly fast response, which relies on the sensitivity of many enzymes to chemical cues that increase or decrease their catalytic activity (see Chapter 8). The activity of the first enzyme in the tryptophan synthesis pathway is inhibited by the pathway's end product (**Figure 18.2a**). Thus, if tryptophan accumulates in a cell, it shuts down the synthesis of more tryptophan by inhibiting enzyme activity. Such *feedback inhibition*, typical of

anabolic (biosynthetic) pathways, allows a cell to adapt to short-term fluctuations in the supply of a substance it needs.

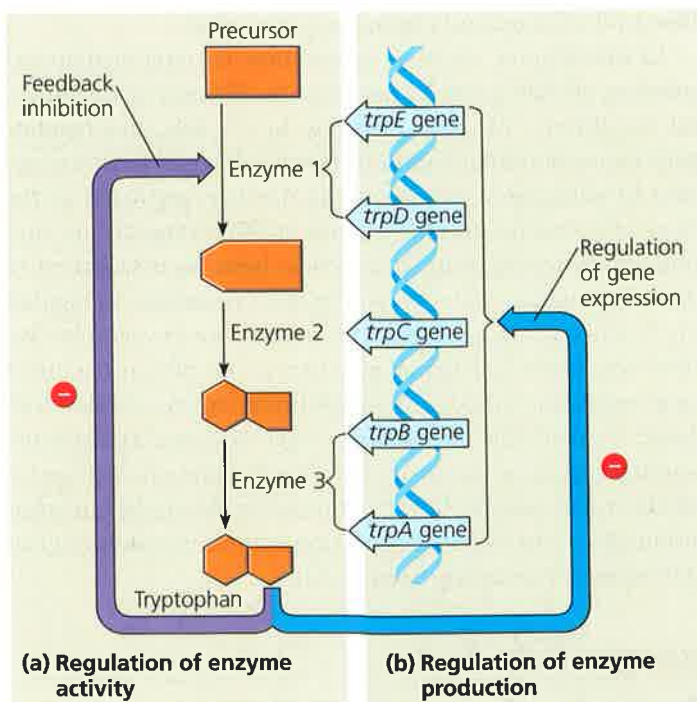
Second, cells can adjust the production level of certain enzymes; that is, they can regulate the expression of the genes encoding the enzymes. If, in our example, the environment provides all the tryptophan the cell needs, the cell stops making the enzymes that catalyze the synthesis of tryptophan (**Figure 18.2b**). In this case, the control of enzyme production occurs at the level of transcription, the synthesis of messenger RNA coding for these enzymes. More generally, many genes of the bacterial genome are switched on or off by changes in the metabolic status of the cell. One basic mechanism for this control of gene expression in bacteria, described as the *operon model*, was discovered in 1961 by François Jacob and Jacques Monod at the Pasteur Institute in Paris. Let's see what an operon is and how it works, using the control of tryptophan synthesis as our first example.

Operons: The Basic Concept

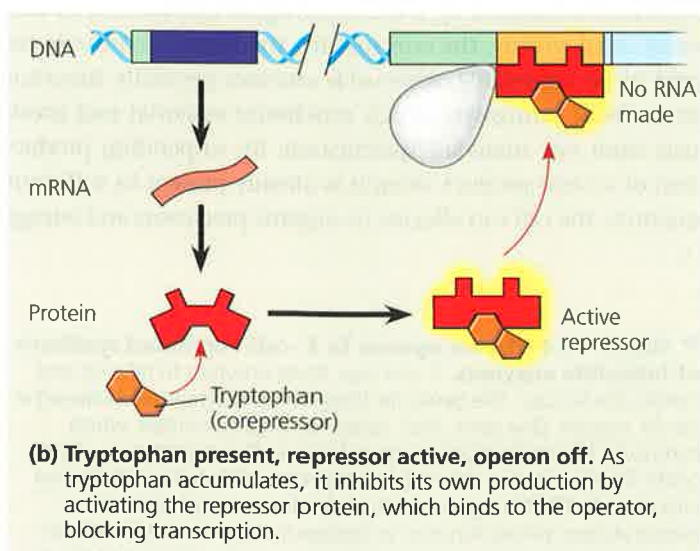
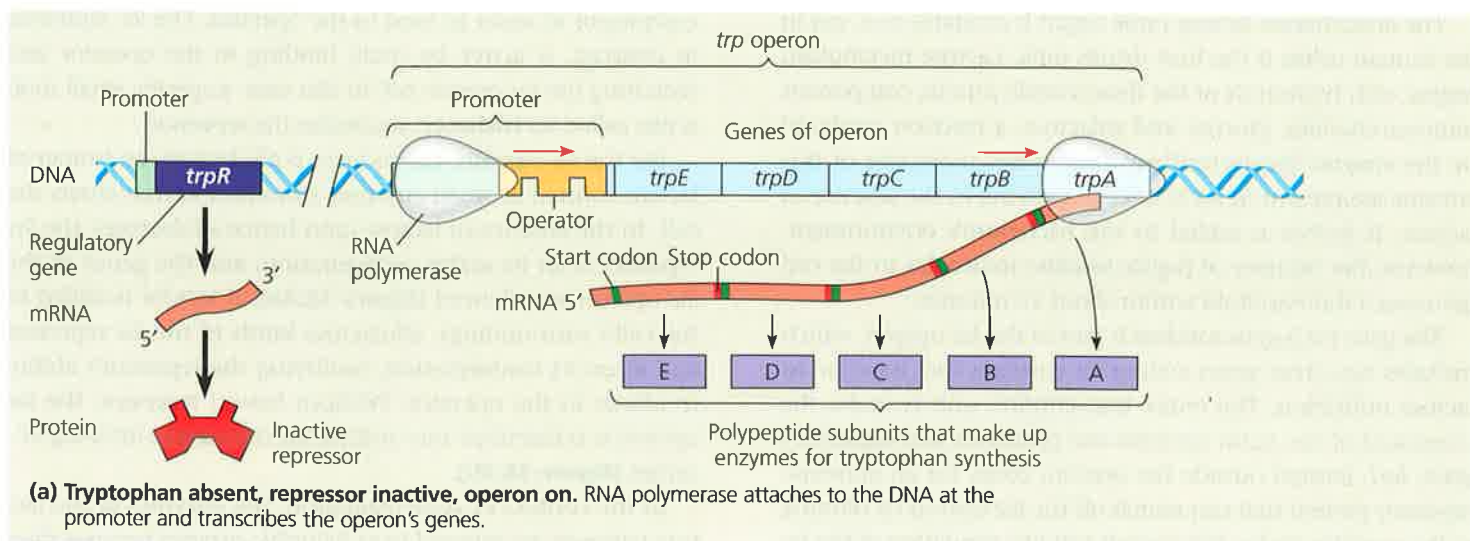
E. coli synthesizes the amino acid tryptophan from a precursor molecule in the multistep pathway shown in Figure 18.2. Each reaction in the pathway is catalyzed by a specific enzyme, and the five genes that code for the subunits of these enzymes are clustered together on the bacterial chromosome. A single promoter serves all five genes, which together constitute a transcription unit. (Recall from Chapter 17 that a promoter is a site where RNA polymerase can bind to DNA and begin transcription.) Thus, transcription gives rise to one long mRNA molecule that codes for the five polypeptides making up the enzymes in the tryptophan pathway. The cell can translate this one mRNA into five separate polypeptides because the mRNA is punctuated with start and stop codons that signal where the coding sequence for each polypeptide begins and ends.

A key advantage of grouping genes of related function into one transcription unit is that a single “on-off switch” can control the whole cluster of functionally related genes; in other words, these genes are *coordinately controlled*. When an *E. coli* cell must make tryptophan for itself because the nutrient medium lacks this amino acid, all the enzymes for the metabolic pathway are synthesized at one time. The switch is a segment of DNA called an **operator**. Both its location and name suit its function: Positioned within the promoter or, in some cases, between the promoter and the enzyme-coding genes, the operator controls the access of RNA polymerase to the genes. All together, the operator, the promoter, and the genes they control—the entire stretch of DNA required for enzyme production for the tryptophan pathway—constitute an **operon**. The *trp* operon (*trp* for tryptophan) is one of many operons in the *E. coli* genome (**Figure 18.3**).

If the operator is the operon's switch for controlling transcription, how does this switch work? By itself, the *trp* operon is turned on; that is, RNA polymerase can bind to the promoter and transcribe the genes of the operon. The operon



▲ Figure 18.2 Regulation of a metabolic pathway. In the pathway for tryptophan synthesis, an abundance of tryptophan can both (a) inhibit the activity of the first enzyme in the pathway (feedback inhibition), a rapid response, and (b) repress expression of the genes encoding all subunits of the enzymes in the pathway, a longer-term response. Genes *trpE* and *trpD* encode the two subunits of enzyme 1, and genes *trpB* and *trpA* encode the two subunits of enzyme 3. (The genes were named before the order in which they functioned in the pathway was determined.) The \ominus symbol stands for inhibition.



▲ **Figure 18.3 The *trp* operon in *E. coli*: regulated synthesis of repressible enzymes.** Tryptophan is an amino acid produced by an anabolic pathway catalyzed by repressible enzymes. **(a)** The five genes encoding the polypeptide subunits of the enzymes in this pathway (see Figure 18.2) are grouped, along with a promoter, into the *trp* operon. The *trp* operator (the repressor binding site) is located within the *trp* promoter (the RNA polymerase binding site). **(b)** Accumulation of tryptophan, the end product of the pathway, represses transcription of the *trp* operon, thus blocking synthesis of all the enzymes in the pathway and shutting down tryptophan production.

? Describe what happens to the *trp* operon as the cell uses up its store of tryptophan.

can be switched off by a protein called the *trp* repressor. The repressor binds to the operator and blocks attachment of RNA polymerase to the promoter, preventing transcription of the genes. A repressor protein is specific for the operator of a particular operon. For example, the repressor that switches off the *trp* operon by binding to the *trp* operator has no effect on other operons in the *E. coli* genome.

The *trp* repressor is the protein product of a **regulatory gene** called *trpR*, which is located some distance from the *trp* operon and has its own promoter. Regulatory genes are expressed continuously, although at a low rate, and a few *trp* repressor molecules are always present in *E. coli* cells. Why, then, is the *trp* operon not switched off permanently? First, the binding of repressors to operators is reversible. An operator vacillates between two states: one without the repressor bound and one with the repressor bound. The relative duration of each state depends on the number of active repressor molecules around. Second, the *trp* repressor, like most regulatory proteins, is an allosteric protein, with two alternative

shapes, active and inactive (see Figure 8.20). The *trp* repressor is synthesized in an inactive form with little affinity for the *trp* operator. Only if tryptophan binds to the *trp* repressor at an allosteric site does the repressor protein change to the active form that can attach to the operator, turning the operon off.

Tryptophan functions in this system as a **corepressor**, a small molecule that cooperates with a repressor protein to switch an operon off. As tryptophan accumulates, more tryptophan molecules associate with *trp* repressor molecules, which can then bind to the *trp* operator and shut down production of the tryptophan pathway enzymes. If the cell's tryptophan level drops, transcription of the operon's genes resumes. The *trp* operon is one example of how gene expression can respond to changes in the cell's internal and external environment.

Repressible and Inducible Operons: Two Types of Negative Gene Regulation

The *trp* operon is said to be a *repressible operon* because its transcription is usually on but can be inhibited (repressed) when a specific small molecule (in this case, tryptophan) binds allosterically to a regulatory protein. In contrast, an *inducible operon* is usually off but can be stimulated (induced) when a specific small molecule interacts with a regulatory protein. The classic example of an inducible operon is the *lac* operon (*lac* for lactose), which was the subject of Jacob and Monod's pioneering research.

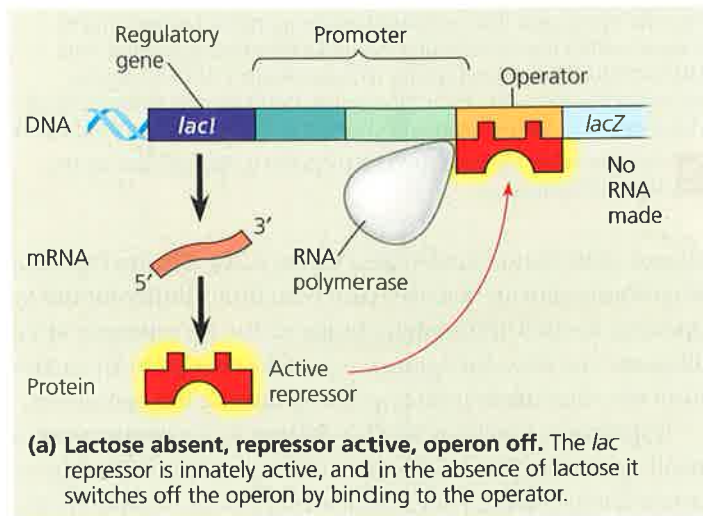
The disaccharide lactose (milk sugar) is available to *E. coli* in the human colon if the host drinks milk. Lactose metabolism begins with hydrolysis of the disaccharide into its component monosaccharides, glucose and galactose, a reaction catalyzed by the enzyme β -galactosidase. Only a few molecules of this enzyme are present in an *E. coli* cell growing in the absence of lactose. If lactose is added to the bacterium's environment, however, the number of β -galactosidase molecules in the cell increases a thousandfold within about 15 minutes.

The gene for β -galactosidase is part of the *lac* operon, which includes two other genes coding for enzymes that function in lactose utilization. The entire transcription unit is under the command of one main operator and promoter. The regulatory gene, *lacI*, located outside the operon, codes for an allosteric repressor protein that can switch off the *lac* operon by binding to the operator. So far, this sounds just like regulation of the *trp* operon, but there is one important difference. Recall that the *trp* repressor is inactive by itself and requires tryptophan as a

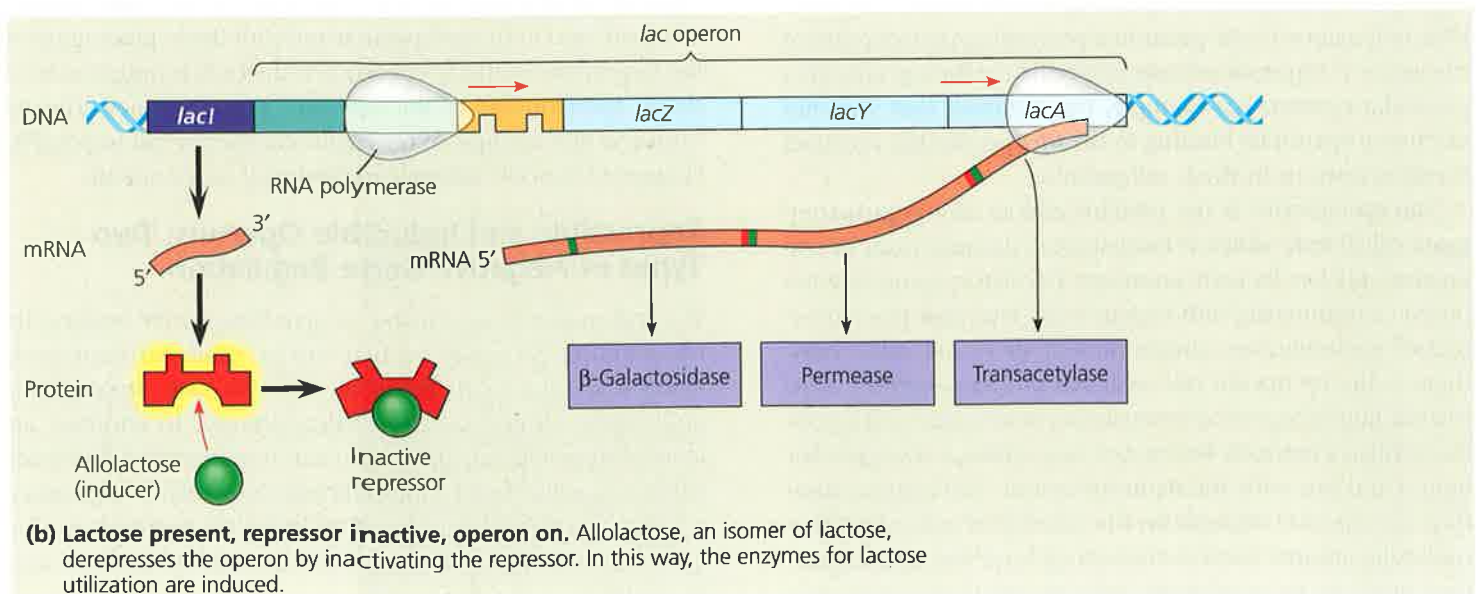
corepressor in order to bind to the operator. The *lac* repressor, in contrast, is active by itself, binding to the operator and switching the *lac* operon off. In this case, a specific small molecule, called an **inducer**, *inactivates* the repressor.

For the *lac* operon, the inducer is allolactose, an isomer of lactose formed in small amounts from lactose that enters the cell. In the absence of lactose (and hence allolactose), the *lac* repressor is in its active configuration, and the genes of the *lac* operon are silenced (**Figure 18.4a**). If lactose is added to the cell's surroundings, allolactose binds to the *lac* repressor and alters its conformation, nullifying the repressor's ability to attach to the operator. Without bound repressor, the *lac* operon is transcribed into mRNA for the lactose-utilizing enzymes (**Figure 18.4b**).

In the context of gene regulation, the enzymes of the lactose pathway are referred to as *inducible enzymes* because their synthesis is induced by a chemical signal (allolactose, in this case). Analogously, the enzymes for tryptophan synthesis are said to be repressible. *Repressible enzymes* generally function in anabolic pathways, which synthesize essential end products from raw materials (precursors). By suspending production of an end product when it is already present in sufficient quantity, the cell can allocate its organic precursors and energy



▼ **Figure 18.4 The *lac* operon in *E. coli*: regulated synthesis of inducible enzymes.** *E. coli* uses three enzymes to take up and metabolize lactose. The genes for these three enzymes are clustered in the *lac* operon. One gene, *lacZ*, codes for β -galactosidase, which hydrolyzes lactose to glucose and galactose. The second gene, *lacY*, codes for a permease, the membrane protein that transports lactose into the cell. The third gene, *lacA*, codes for an enzyme called transacetylase, whose function in lactose metabolism is still unclear. The gene for the *lac* repressor, *lacI*, happens to be adjacent to the *lac* operon, an unusual situation. The function of the teal region at the upstream end of the promoter (the left end in these diagrams) will be revealed in Figure 18.5.



for other uses. In contrast, inducible enzymes usually function in catabolic pathways, which break down a nutrient to simpler molecules. By producing the appropriate enzymes only when the nutrient is available, the cell avoids wasting energy and precursors making proteins that are not needed.

Regulation of both the *trp* and *lac* operons involves the *negative* control of genes, because the operons are switched off by the active form of the repressor protein. It may be easier to see this for the *trp* operon, but it is also true for the *lac* operon. Allolactose induces enzyme synthesis not by acting directly on the genome, but by freeing the *lac* operon from the negative effect of the repressor. Gene regulation is said to be *positive* only when a regulatory protein interacts directly with the genome to switch transcription on. Let's look at an example of the positive control of genes, again involving the *lac* operon.

Positive Gene Regulation

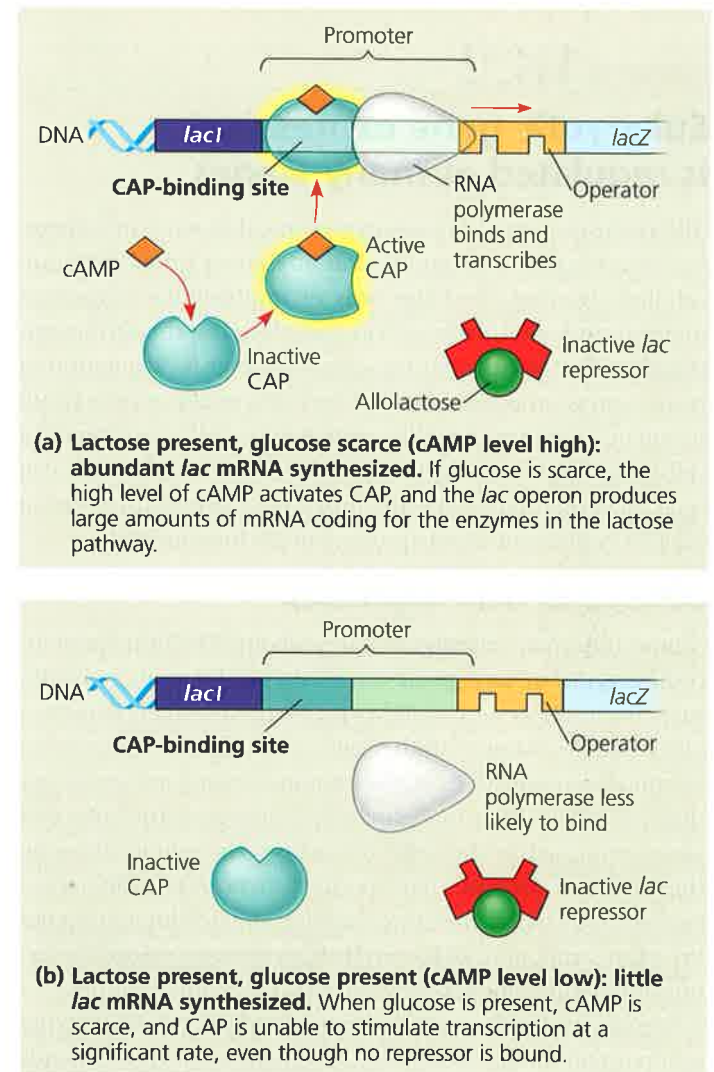
When glucose and lactose are both present in its environment, *E. coli* preferentially uses glucose. The enzymes for glucose breakdown in glycolysis (see Figure 9.9) are continually present. Only when lactose is present *and* glucose is in short supply does *E. coli* use lactose as an energy source, and only then does it synthesize appreciable quantities of the enzymes for lactose breakdown.

How does the *E. coli* cell sense the glucose concentration and relay this information to the genome? Again, the mechanism depends on the interaction of an allosteric regulatory protein with a small organic molecule, in this case **cyclic AMP (cAMP)**, which accumulates when glucose is scarce (see Figure 11.11 for the structure of cAMP). The regulatory protein, called *catabolite activator protein* (CAP), is an **activator**, a protein that binds to DNA and stimulates transcription of a gene. When cAMP binds to this regulatory protein, CAP assumes its active shape and can attach to a specific site at the upstream end of the *lac* promoter (**Figure 18.5a**). This attachment increases the affinity of RNA polymerase for the promoter, which is actually rather low even when no repressor is bound to the operator. By facilitating the binding of RNA polymerase to the promoter and thereby increasing the rate of transcription, the attachment of CAP to the promoter directly stimulates gene expression. Therefore, this mechanism qualifies as positive regulation.

If the amount of glucose in the cell increases, the cAMP concentration falls, and without cAMP, CAP detaches from the operon. Because CAP is inactive, RNA polymerase binds less efficiently to the promoter, and transcription of the *lac* operon proceeds at only a low level, even in the presence of lactose (**Figure 18.5b**). Thus, the *lac* operon is under dual control: negative control by the *lac* repressor and positive control by CAP. The state of the *lac* repressor (with or without bound allolactose) determines whether or not transcription of the *lac* operon's genes occurs at all; the state of CAP (with or without bound cAMP) controls the *rate* of transcription if

the operon is repressor-free. It is as though the operon has both an on-off switch and a volume control.

In addition to regulating the *lac* operon, CAP helps regulate other operons that encode enzymes used in catabolic pathways. All told, it may affect the expression of more than 100 genes in *E. coli*. When glucose is plentiful and CAP is inactive, the synthesis of enzymes that catabolize compounds other than glucose generally slows down. The ability to catabolize other compounds, such as lactose, enables a cell deprived of glucose to survive. The compounds present in the cell at the moment determine which operons are switched on—the result of simple interactions of activator and repressor proteins with the promoters of the genes in question.



▲ Figure 18.5 Positive control of the *lac* operon by catabolite activator protein (CAP). RNA polymerase has high affinity for the *lac* promoter only when catabolite activator protein (CAP) is bound to a DNA site at the upstream end of the promoter. CAP attaches to its DNA site only when associated with cyclic AMP (cAMP), whose concentration in the cell rises when the glucose concentration falls. Thus, when glucose is present, even if lactose also is available, the cell preferentially catabolizes glucose and makes very little of the lactose-utilizing enzymes.

CONCEPT CHECK 18.1

1. How does binding of the *trp* corepressor and the *lac* inducer to their respective repressor proteins alter repressor function and transcription in each case?
2. Describe the binding of RNA polymerase, repressors, and activators to the *lac* operon when both lactose and glucose are scarce. What is the effect of these scarcities on transcription of the *lac* operon?
3. **WHAT IF?** A certain mutation in *E. coli* changes the *lac* operator so that the active repressor cannot bind. How would this affect the cell's production of β -galactosidase?

For suggested answers, see Appendix A.

CONCEPT 18.2

Eukaryotic gene expression is regulated at many stages

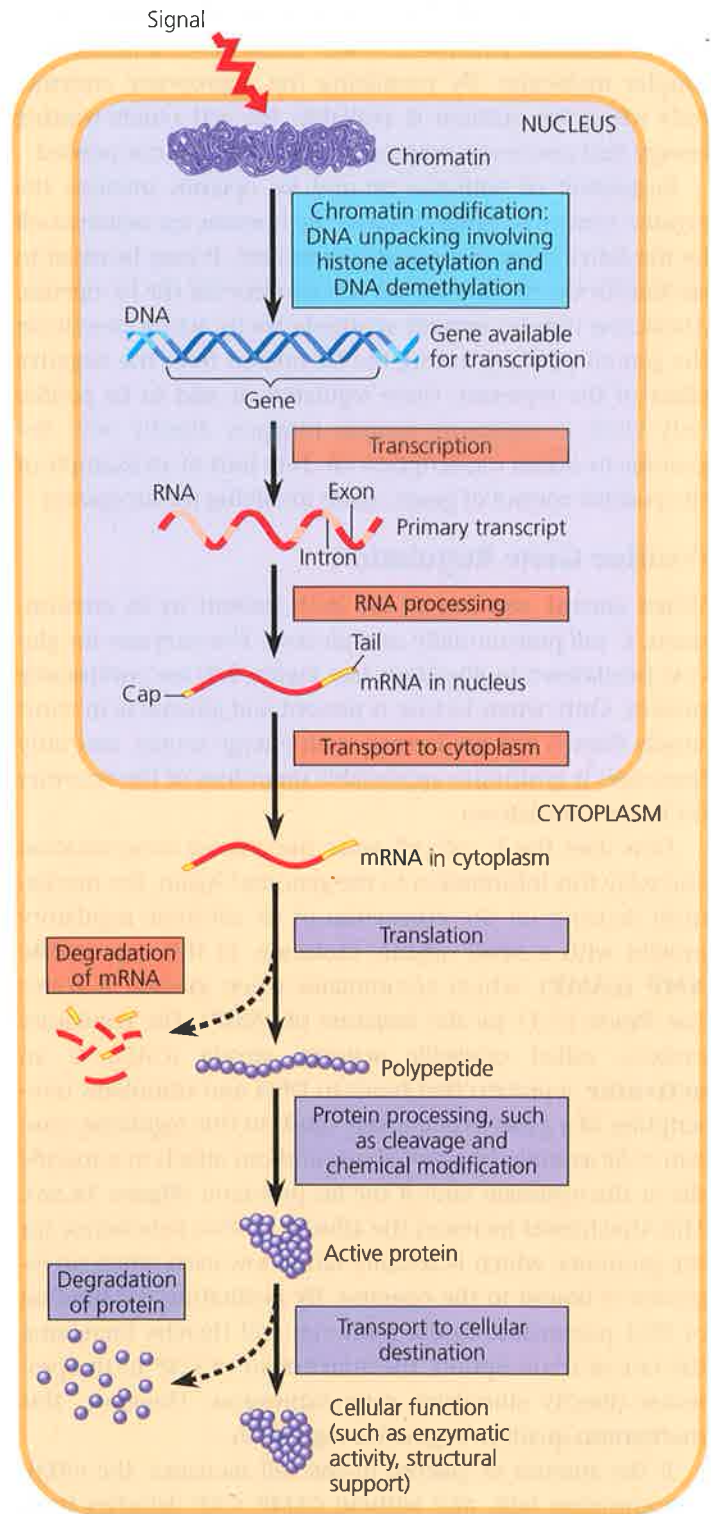
All organisms, whether prokaryotes or eukaryotes, must regulate which genes are expressed at any given time. Both unicellular organisms and the cells of multicellular organisms must continually turn genes on and off in response to signals from their external and internal environments. Regulation of gene expression is also essential for cell specialization in multicellular organisms, which are made up of different types of cells, each with a distinct role. To perform its role, each cell type must maintain a specific program of gene expression in which certain genes are expressed and others are not.

Differential Gene Expression

A typical human cell might express about 20% of its protein-coding genes at any given time. Highly differentiated cells, such as muscle or nerve cells, express an even smaller fraction of their genes. Almost all the cells in an organism contain an identical genome. (Cells of the immune system are one exception, as you will see in Chapter 43.) However, the subset of genes expressed in the cells of each type is unique, allowing these cells to carry out their specific function. The differences between cell types, therefore, are due not to different genes being present, but to **differential gene expression**, the expression of different genes by cells with the same genome.

The function of any cell, whether a single-celled eukaryote or a particular cell type in a multicellular organism, depends on the appropriate set of genes being expressed. The transcription factors of a cell must locate the right genes at the right time, a task on a par with finding a needle in a haystack. When gene expression proceeds abnormally, serious imbalances and diseases, including cancer, can arise.

Figure 18.6 summarizes the entire process of gene expression in a eukaryotic cell, highlighting key stages in the expression of a protein-coding gene. Each stage depicted in



▲ **Figure 18.6 Stages in gene expression that can be regulated in eukaryotic cells.** In this diagram, the colored boxes indicate the processes most often regulated; each color indicates the type of molecule that is affected (blue = DNA, orange = RNA, purple = protein). The nuclear envelope separating transcription from translation in eukaryotic cells offers an opportunity for post-transcriptional control in the form of RNA processing that is absent in prokaryotes. In addition, eukaryotes have a greater variety of control mechanisms operating before transcription and after translation. The expression of any given gene, however, does not necessarily involve every stage shown; for example, not every polypeptide is cleaved.

Figure 18.6 is a potential control point at which gene expression can be turned on or off, accelerated, or slowed down.

Only 50 years ago, an understanding of the mechanisms that control gene expression in eukaryotes seemed almost hopelessly out of reach. Since then, new research methods, notably advances in DNA technology (see Chapter 20), have enabled molecular biologists to uncover many of the details of eukaryotic gene regulation. In all organisms, a common control point for gene expression is at transcription; regulation at this stage often occurs in response to signals coming from outside the cell, such as hormones or other signaling molecules. For this reason, the term *gene expression* is often equated with transcription for both bacteria and eukaryotes. While this is most often the case for bacteria, the greater complexity of eukaryotic cell structure and function provides opportunities for regulating gene expression at many additional stages (see Figure 18.6). In the remainder of this section, we'll examine some of the important control points of eukaryotic gene expression more closely.

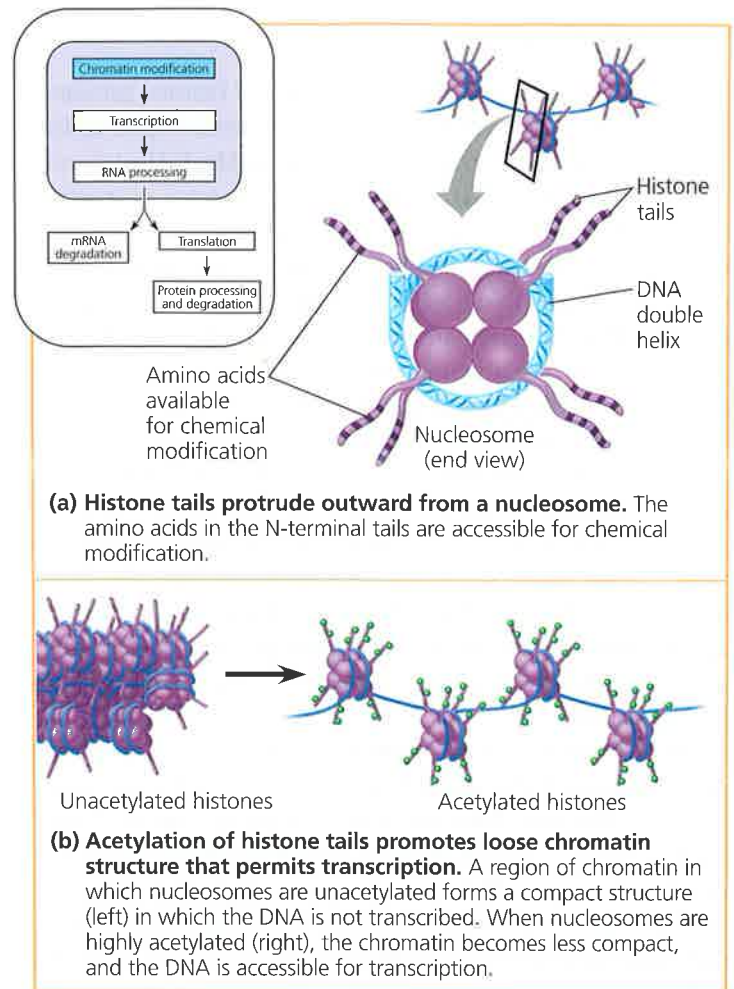
Regulation of Chromatin Structure

Recall that the DNA of eukaryotic cells is packaged with proteins in an elaborate complex known as chromatin, the basic unit of which is the nucleosome (see Figure 16.22). The structural organization of chromatin not only packs a cell's DNA into a compact form that fits inside the nucleus, but also helps regulate gene expression in several ways. The location of a gene's promoter relative to nucleosomes and to the sites where the DNA attaches to the chromosome scaffold or nuclear lamina can affect whether the gene is transcribed. In addition, genes within heterochromatin, which is highly condensed, are usually not expressed. Lastly, certain chemical modifications to the histone proteins and to the DNA of chromatin can influence both chromatin structure and gene expression. Here we examine the effects of these modifications, which are catalyzed by specific enzymes.

Histone Modifications

There is abundant evidence that chemical modifications to histones, the proteins around which the DNA is wrapped in nucleosomes, play a direct role in the regulation of gene transcription. The N-terminus of each histone molecule in a nucleosome protrudes outward from the nucleosome (Figure 18.7a). These histone tails are accessible to various modifying enzymes that catalyze the addition or removal of specific chemical groups.

In **histone acetylation**, acetyl groups ($-\text{COCH}_3$) are attached to lysines in histone tails; deacetylation is the removal of acetyl groups. When the lysines are acetylated, their positive charges are neutralized and the histone tails no longer bind to neighboring nucleosomes (Figure 18.7b). Such binding promotes the folding of chromatin into a more compact structure; when this binding does not occur, chromatin has a looser structure. As a result, transcription proteins have easier



▲ Figure 18.7 A simple model of histone tails and the effect of histone acetylation. In addition to acetylation, histones can undergo several other types of modifications that also help determine the chromatin configuration in a region.

access to genes in an acetylated region. Researchers have shown that some enzymes that acetylate or deacetylate histones are closely associated with or even components of the transcription factors that bind to promoters (see Figure 17.8). These observations suggest that histone acetylation enzymes may promote the initiation of transcription not only by remodeling chromatin structure, but also by binding to and thus “recruiting” components of the transcription machinery.

Other chemical groups, such as methyl and phosphate groups, can be reversibly attached to amino acids in histone tails. Addition of methyl groups ($-\text{CH}_3$) to histone tails (histone methylation) can promote condensation of the chromatin, while addition of a phosphate group (phosphorylation) to an amino acid next to a methylated amino acid can have the opposite effect. The recent discovery that modifications to histone tails can affect chromatin structure and gene expression has led to the *histone code hypothesis*. This hypothesis proposes that specific combinations of modifications, as well as the order in which they have occurred, help determine the chromatin configuration, which in turn influences transcription.

DNA Methylation

While some enzymes methylate the tails of histone proteins, a different set of enzymes can methylate certain bases in the DNA itself, usually cytosine. Such **DNA methylation** occurs in most plants, animals, and fungi. Long stretches of inactive DNA, such as that of inactivated mammalian X chromosomes (see Figure 15.8), are generally more methylated than regions of actively transcribed DNA, although there are exceptions. On a smaller scale, individual genes are usually more heavily methylated in cells in which they are not expressed. Removal of the extra methyl groups can turn on some of these genes.

At least in some species, DNA methylation seems to be essential for the long-term inactivation of genes that occurs during normal cell differentiation in the embryo. For instance, experiments have shown that deficient DNA methylation (due to lack of a methylating enzyme) leads to abnormal embryonic development in organisms as different as mice and *Arabidopsis* (a mustard plant). Once methylated, genes usually stay that way through successive cell divisions in a given individual. At DNA sites where one strand is already methylated, enzymes methylate the correct daughter strand after each round of DNA replication. Methylation patterns are thus passed on, and cells forming specialized tissues keep a chemical record of what occurred during embryonic development. A methylation pattern maintained in this way also accounts for *genomic imprinting* in mammals, where methylation permanently regulates expression of either the maternal or paternal allele of particular genes at the start of development (see Figure 15.17).

Epigenetic Inheritance

The chromatin modifications that we have just discussed do not entail a change in the DNA sequence, yet they may be passed along to future generations of cells. Inheritance of traits transmitted by mechanisms not directly involving the nucleotide sequence is called **epigenetic inheritance**. Whereas mutations in the DNA are permanent changes, modifications to the chromatin can be reversed, by processes that are not yet fully understood. The molecular systems for chromatin modification may well interact with each other in a regulated way. In *Drosophila*, for example, experiments have suggested that a particular histone-modifying enzyme recruits a DNA methylation enzyme to one region and that the two enzymes collaborate to silence a particular set of genes. Working in the opposite order, proteins have also been found that first bind to methylated DNA and then recruit histone deacetylation enzymes. Thus, a dual mechanism, involving both DNA methylation and histone deacetylation, can repress transcription.

Researchers are amassing more and more evidence for the importance of epigenetic information in the regulation of gene expression. Epigenetic variations might help explain why one identical twin acquires a genetically based disease,

such as schizophrenia, but the other does not, despite their identical genomes. Alterations in normal patterns of DNA methylation are seen in some cancers, where they are associated with inappropriate gene expression. Evidently, enzymes that modify chromatin structure are integral parts of the eukaryotic cell's machinery for regulating transcription.

Regulation of Transcription Initiation

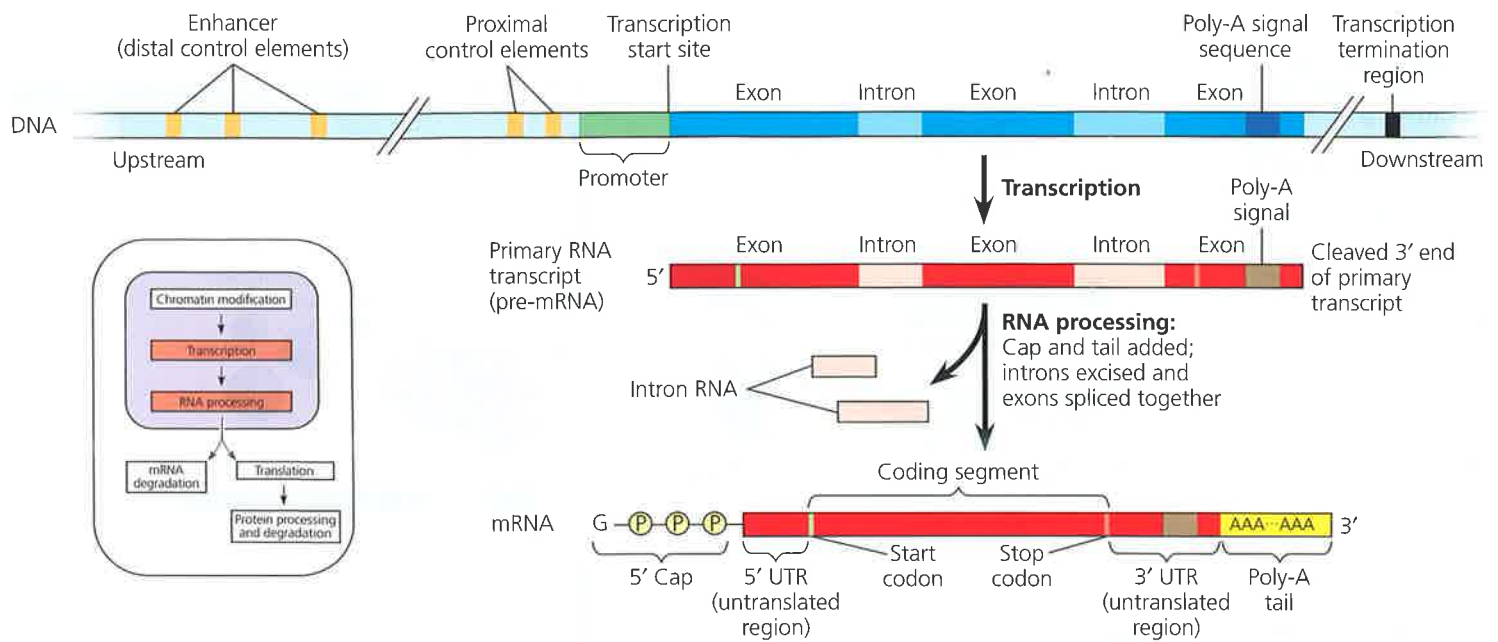
Chromatin-modifying enzymes provide initial control of gene expression by making a region of DNA either more or less able to bind the transcription machinery. Once the chromatin of a gene is optimally modified for expression, the initiation of transcription is the next major step at which gene expression is regulated. As in bacteria, the regulation of transcription initiation in eukaryotes involves proteins that bind to DNA and either facilitate or inhibit binding of RNA polymerase. The process is more complicated in eukaryotes, however. Before looking at how eukaryotic cells control their transcription, let's review the structure of a typical eukaryotic gene and its transcript.

Organization of a Typical Eukaryotic Gene

A eukaryotic gene and the DNA elements (segments) that control it are typically organized as shown in **Figure 18.8**, which extends what you learned about eukaryotic genes in Chapter 17. Recall that a cluster of proteins called a *transcription initiation complex* assembles on the promoter sequence at the "upstream" end of the gene. One of these proteins, RNA polymerase II, then proceeds to transcribe the gene, synthesizing a primary RNA transcript (pre-mRNA). RNA processing includes enzymatic addition of a 5' cap and a poly-A tail, as well as splicing out of introns, to yield a mature mRNA. Associated with most eukaryotic genes are multiple **control elements**, segments of noncoding DNA that serve as binding sites for the proteins called transcription factors, which in turn regulate transcription. Control elements and the transcription factors they bind are critical to the precise regulation of gene expression seen in different cell types.

The Roles of Transcription Factors

To initiate transcription, eukaryotic RNA polymerase requires the assistance of transcription factors. Some transcription factors, such as those illustrated in Figure 17.8, are essential for the transcription of *all* protein-coding genes; therefore, they are often called *general transcription factors*. Only a few general transcription factors independently bind a DNA sequence, such as the TATA box within the promoter; the others primarily bind proteins, including each other and RNA polymerase II. Protein-protein interactions are crucial to the initiation of eukaryotic transcription. Only when the complete initiation complex has assembled can the polymerase begin to move along the DNA template strand, producing a complementary strand of RNA.



▲ **Figure 18.8 A eukaryotic gene and its transcript.** Each eukaryotic gene has a promoter, a DNA sequence where RNA polymerase binds and starts transcription, proceeding “downstream.” A number of control elements (gold) are involved in regulating the initiation of transcription; these are DNA sequences located near (proximal to) or far from

(distal to) the promoter. Distal control elements can be grouped together as enhancers, one of which is shown for this gene. A polyadenylation (poly-A) signal sequence in the last exon of the gene is transcribed into an RNA sequence that signals where the transcript is cleaved and the poly-A tail added. Transcription may continue for hundreds of nucleotides beyond the poly-A

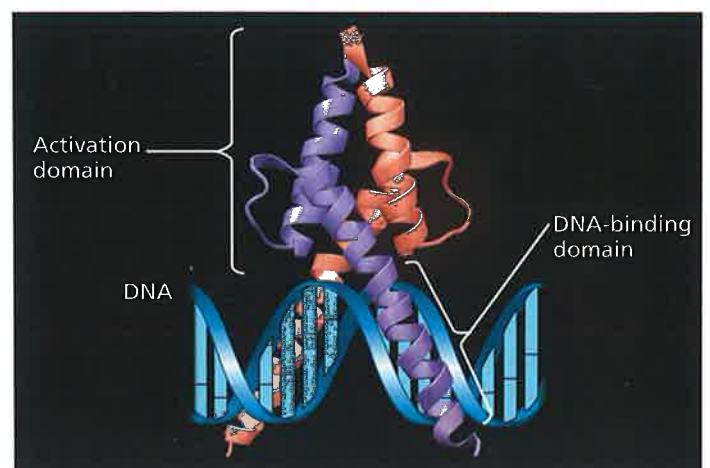
signal before terminating. RNA processing of the primary transcript into a functional mRNA involves three steps: addition of the 5' cap, addition of the poly-A tail, and splicing. In the cell, the 5' cap is added soon after transcription is initiated; splicing and poly-A tail addition may also occur while transcription is still under way (see Figure 17.10).

The interaction of general transcription factors and RNA polymerase II with a promoter usually leads to only a low rate of initiation and production of few RNA transcripts. In eukaryotes, high levels of transcription of particular genes at the appropriate time and place depend on the interaction of control elements with another set of proteins, which can be thought of as *specific transcription factors*.

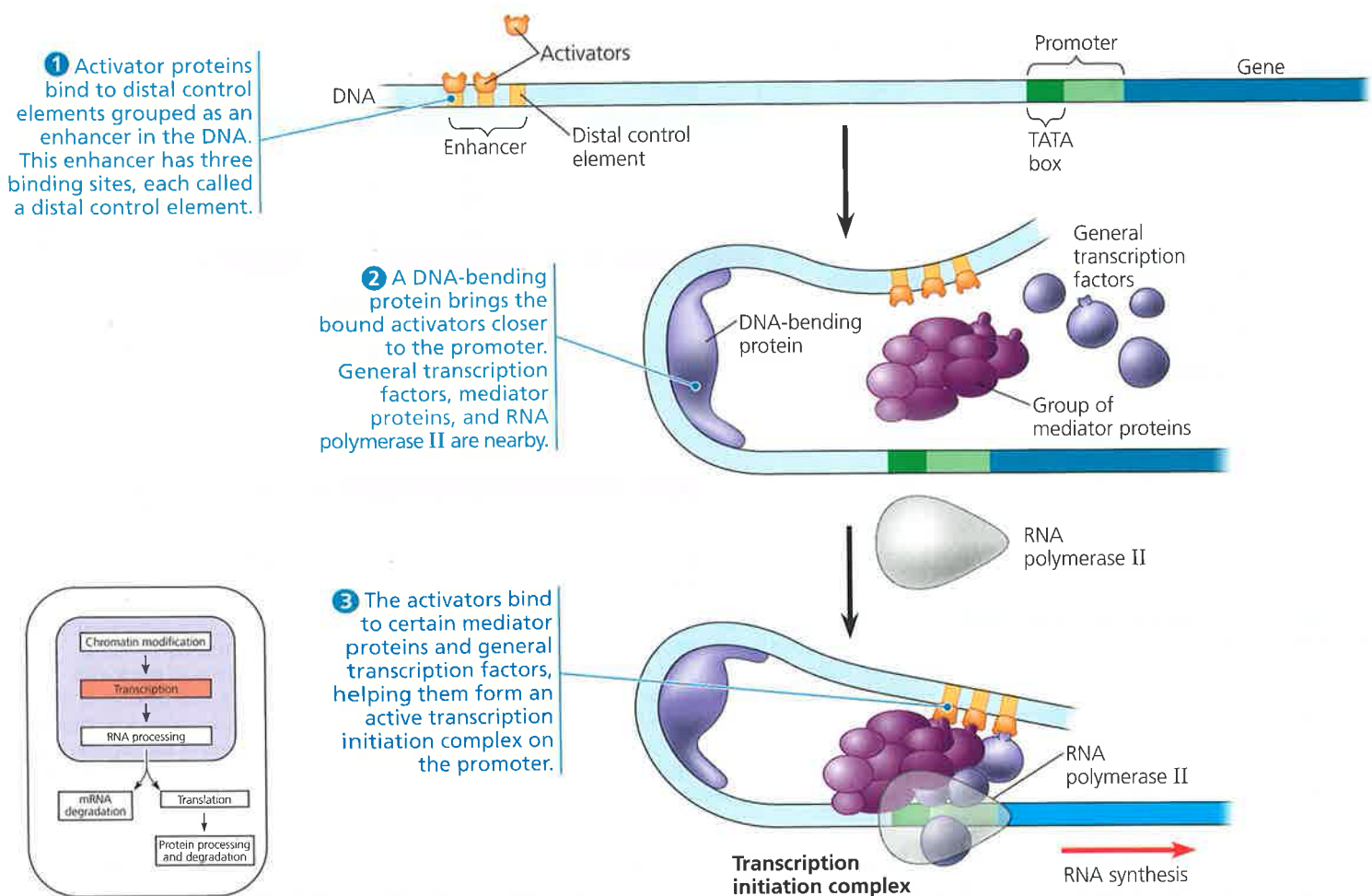
Enhancers and Specific Transcription Factors As you can see in Figure 18.8, some control elements, named *proximal control elements*, are located close to the promoter. (Although some biologists consider proximal control elements part of the promoter, in this book we do not.) The more distant *distal control elements*, groupings of which are called **enhancers**, may be thousands of nucleotides upstream or downstream of a gene or even within an intron. A given gene may have multiple enhancers, each active at a different time or in a different cell type or location in the organism. Each enhancer, however, is generally associated with only that gene and no other.

In eukaryotes, the rate of gene expression can be strongly increased or decreased by the binding of specific transcription factors, either activators or repressors, to the control elements of enhancers. Hundreds of transcription activators have been discovered in eukaryotes; the structure of one example is shown in **Figure 18.9**. Researchers have identified two common structural elements in a large number of activator proteins: a DNA-binding domain—a part of the protein’s

three-dimensional structure that binds to DNA—and one or more activation domains. Activation domains bind other regulatory proteins or components of the transcription machinery, facilitating a series of protein-protein interactions that result in transcription of a given gene.



▲ **Figure 18.9 The structure of MyoD, a specific transcription factor that acts as an activator.** The MyoD protein is made up of two subunits (purple and salmon) with extensive regions of α helix. Each subunit has a DNA-binding domain and an activation domain (indicated by brackets for the purple subunit). The activation domain includes binding sites for the other subunit as well as other proteins. MyoD is involved in muscle development in vertebrate embryos and will be discussed further in Concept 18.4.



▲ Figure 18.10 A model for the action of enhancers and transcription activators. Bending of the DNA by a protein enables enhancers to influence a promoter hundreds or even thousands of nucleotides away. Specific transcription factors called

activators bind to the enhancer DNA sequences and then to a group of mediator proteins, which in turn bind to general transcription factors, assembling the transcription initiation complex. These protein-protein interactions facilitate the correct positioning of the complex

on the promoter and the initiation of RNA synthesis. Only one enhancer (with three orange control elements) is shown here, but a gene may have several enhancers that act at different times or in different cell types.

Figure 18.10 shows a current model for how binding of activators to an enhancer located far from the promoter can influence transcription. Protein-mediated bending of the DNA is thought to bring the bound activators into contact with a group of *mediator proteins*, which in turn interact with proteins at the promoter. These multiple protein-protein interactions help assemble and position the initiation complex on the promoter. Support for this model includes a study showing that the proteins regulating a mouse globin gene contact both the gene's promoter and an enhancer located about 50,000 nucleotides upstream. Evidently, these two regions in the DNA must come together in a very specific fashion for this interaction to occur.

Specific transcription factors that function as repressors can inhibit gene expression in several different ways. Some repressors bind directly to control element DNA (in enhancers or elsewhere), blocking activator binding or, in some cases, turning off transcription even when activators are bound. Other repressors block the binding of activators to proteins that allow the activators to bind to DNA.

In addition to influencing transcription directly, some activators and repressors act indirectly by affecting chromatin structure. Studies using yeast and mammalian cells show that some activators recruit proteins that acetylate histones near the promoters of specific genes, thus promoting transcription (see Figure 18.7). Similarly, some repressors recruit proteins that deacetylate histones, leading to reduced transcription, a phenomenon referred to as *silencing*. Indeed, recruitment of chromatin-modifying proteins seems to be the most common mechanism of repression in eukaryotes.

Combinatorial Control of Gene Activation In eukaryotes, the precise control of transcription depends largely on the binding of activators to DNA control elements. Considering the great number of genes that must be regulated in a typical animal or plant cell, the number of completely different nucleotide sequences found in control elements is surprisingly small. A dozen or so short nucleotide sequences appear again and again in the control elements for different genes. On average, each

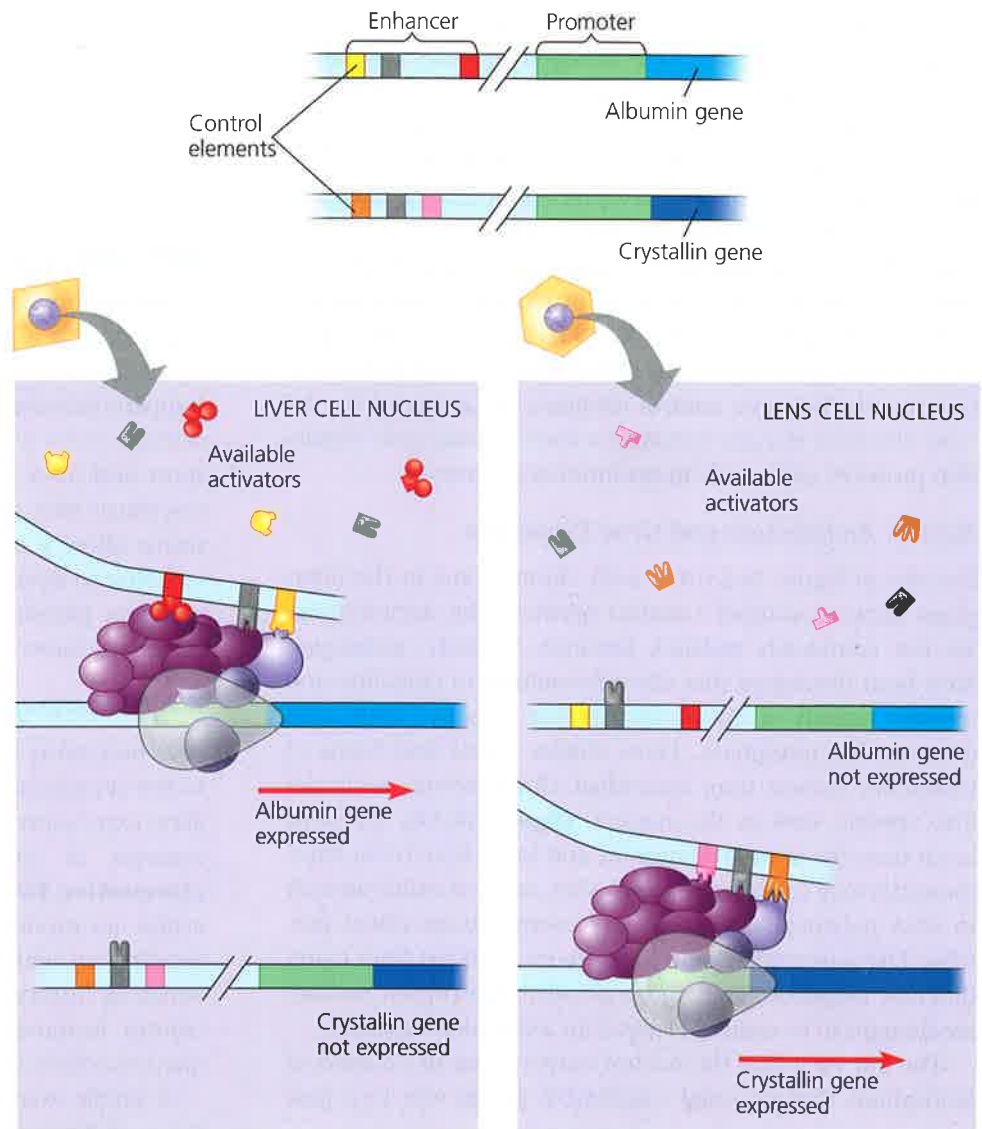
enhancer is composed of about ten control elements, each of which can bind only one or two specific transcription factors. It is the particular *combination* of control elements in an enhancer associated with a gene, rather than the presence of a single unique control element, that is important in regulating transcription of the gene.

Even with only a dozen control element sequences available, a very large number of combinations are possible. A particular combination of control elements will be able to activate transcription only when the appropriate activator proteins are present, which may occur at a precise time during development or in a particular cell type. **Figure 18.11** illustrates how the use of different combinations of just a few control elements can allow differential regulation of transcription in two cell types. This can occur because each cell type contains a different group of activator proteins. How these groups came to differ will be explored in Concept 18.4.

Coordinately Controlled Genes in Eukaryotes

How does the eukaryotic cell deal with genes of related function that need to be turned on or off at the same time? Earlier in this chapter, you learned that in bacteria, such coordinately controlled genes are often clustered into an operon, which is regulated by a single promoter and transcribed into a single mRNA molecule. Thus, the genes are expressed together, and the encoded proteins are produced concurrently. With a few minor exceptions, operons that work in this way have *not* been found in eukaryotic cells.

Co-expressed eukaryotic genes, such as genes coding for the enzymes of a metabolic pathway, are typically scattered over different chromosomes. In these cases, coordinate gene expression depends on the association of a specific combination of control elements with every gene of a dispersed group. The presence of these elements can be compared to the raised flags on a few mailboxes out of many, signaling to the mail carrier to check those boxes. Copies of the activators that recognize the control elements bind to them, promoting simultaneous transcription of the genes, no matter where they are in the genome.



(a) Liver cell. The albumin gene is expressed, and the crystallin gene is not.

(b) Lens cell. The crystallin gene is expressed, and the albumin gene is not.

▲ Figure 18.11 Cell type-specific transcription. Both liver cells and lens cells have the genes for making the proteins albumin and crystallin, but only liver cells make albumin (a blood protein) and only lens cells make crystallin (the main protein of the lens of the eye). The specific transcription factors made in a cell determine which genes are expressed. In this example, the genes for albumin and crystallin are shown at the top, each with an enhancer made up of three different control elements. Although the enhancers for the two genes share one control element (gray), each enhancer has a unique combination of elements. All the activators required for high-level expression of the albumin gene are present only in liver cells (a), whereas the activators needed for expression of the crystallin gene are present only in lens cells (b). For simplicity, we consider only the role of activators here, although the presence or absence of repressors may also influence transcription in certain cell types.

? Describe the enhancer for the albumin gene in each cell. How would the nucleotide sequence of this enhancer in the liver cell compare with that in the lens cell?

Coordinate control of dispersed genes in a eukaryotic cell often occurs in response to chemical signals from outside the cell. A steroid hormone, for example, enters a cell and binds to a specific intracellular receptor protein, forming a hormone-receptor complex that serves as a transcription activator (see Figure 11.9). Every gene whose transcription is stimulated by a particular steroid hormone, regardless of its chromosomal

location, has a control element recognized by that hormone-receptor complex. This is how estrogen activates a group of genes that stimulate cell division in uterine cells, preparing the uterus for pregnancy.

Many signaling molecules, such as nonsteroid hormones and growth factors, bind to receptors on a cell's surface and never actually enter the cell. Such molecules can control gene expression indirectly by triggering signal transduction pathways that lead to activation of particular transcription activators or repressors (see Figure 11.15). Coordinate regulation in such pathways is the same as for steroid hormones: Genes with the same control elements are activated by the same chemical signals. Systems for coordinating gene regulation probably arose early in evolutionary history.

Nuclear Architecture and Gene Expression

You saw in Figure 16.23 that each chromosome in the interphase nucleus occupies a distinct territory. The chromosomes are not completely isolated, however. Recently, techniques have been developed that allow researchers to cross-link and identify regions of chromosomes that associate with each other during interphase. These studies reveal that loops of chromatin extend from individual chromosomal territories into specific sites in the nucleus (Figure 18.12). Different loops from the same chromosome and loops from other chromosomes may congregate in such sites, some of which are rich in RNA polymerases and other transcription-associated proteins. Like a recreation center that draws members from many different neighborhoods, these so-called *transcription factories* are thought to be areas specialized for a common function.

The old view that the nuclear contents are like a bowl of amorphous chromosomal spaghetti is giving way to a new

model of a nucleus with a defined architecture and regulated movements of chromatin. Relocation of particular genes from their chromosomal territories to transcription factories may be part of the process of readying genes for transcription. This is an exciting area of current research that raises many fascinating questions for future study.

Mechanisms of Post-Transcriptional Regulation

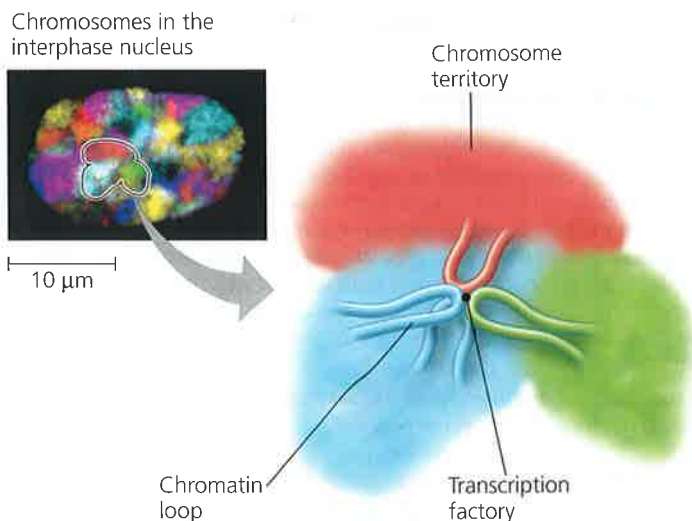
Transcription alone does not constitute gene expression. The expression of a protein-coding gene is ultimately measured by the amount of functional protein a cell makes, and much happens between the synthesis of the RNA transcript and the activity of the protein in the cell. Researchers are discovering more and more regulatory mechanisms that operate at various stages after transcription (see Figure 18.6). These mechanisms allow a cell to fine-tune gene expression rapidly in response to environmental changes without altering its transcription patterns. Here we discuss how cells can regulate gene expression once a gene has been transcribed.

RNA Processing

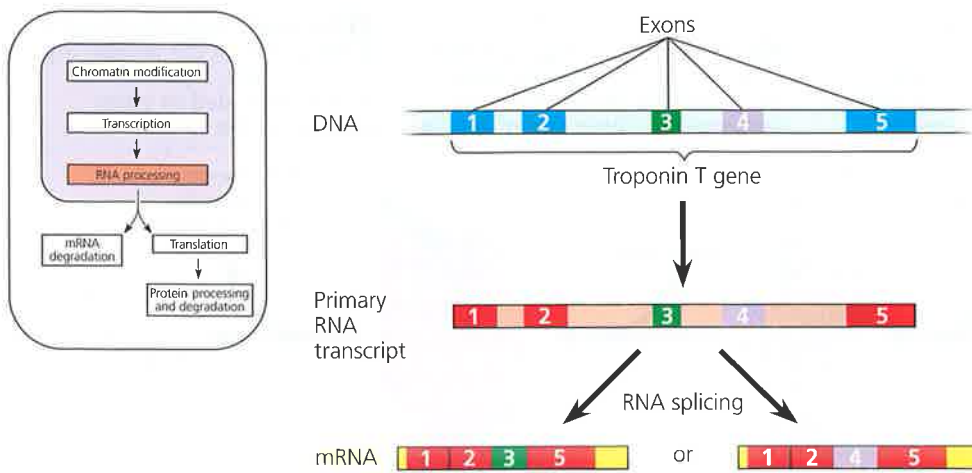
RNA processing in the nucleus and the export of mature RNA to the cytoplasm provide several opportunities for regulating gene expression that are not available in prokaryotes. One example of regulation at the RNA-processing level is **alternative RNA splicing**, in which different mRNA molecules are produced from the same primary transcript, depending on which RNA segments are treated as exons and which as introns. Regulatory proteins specific to a cell type control intron-exon choices by binding to regulatory sequences within the primary transcript.

A simple example of alternative RNA splicing is shown in Figure 18.13 for the troponin T gene, which encodes two different (though related) proteins. Other genes offer possibilities for far greater numbers of products. For instance, researchers have found a *Drosophila* gene with enough alternatively spliced exons to generate about 19,000 membrane proteins that have different extracellular domains. At least 17,500 (94%) of the alternative mRNAs are actually synthesized. Each developing nerve cell in the fly appears to synthesize a unique form of the protein, which acts as an identification badge on the cell surface.

It is clear that alternative RNA splicing can significantly expand the repertoire of a eukaryotic genome. In fact, alternative splicing was proposed as one explanation for the surprisingly low number of human genes counted when the human genome was sequenced about ten years ago. The number of human genes was found to be similar to that of a soil worm (nematode), a mustard plant, or a sea anemone. This discovery prompted questions about what, if not the number of genes, accounts for the more complex morphology (external form) of humans. It turns out that 75–100% of human genes that have multiple exons probably undergo alternative splicing. Thus, the extent of alternative splicing



▲ **Figure 18.12 Chromosomal interactions in the interphase nucleus.** Although each chromosome has its own territory (see Figure 16.23), loops of chromatin may extend into other sites in the nucleus. Some of these sites are transcription factories that are occupied by multiple chromatin loops from the same chromosome (blue loops) or other chromosomes (red and green loops).



◀ **Figure 18.13 Alternative RNA splicing of the troponin T gene.** The primary transcript of this gene can be spliced in more than one way, generating different mRNA molecules. Notice that one mRNA molecule has ended up with exon 3 (green) and the other with exon 4 (purple). These two mRNAs are translated into different but related muscle proteins.

greatly multiplies the number of possible human proteins, which may be better correlated with complexity of form.

mRNA Degradation

The life span of mRNA molecules in the cytoplasm is important in determining the pattern of protein synthesis in a cell. Bacterial mRNA molecules typically are degraded by enzymes within a few minutes of their synthesis. This short life span of mRNAs is one reason bacteria can change their patterns of protein synthesis so quickly in response to environmental changes. In contrast, mRNAs in multicellular eukaryotes typically survive for hours, days, or even weeks. For instance, the mRNAs for the hemoglobin polypeptides (α -globin and β -globin) in developing red blood cells are unusually stable, and these long-lived mRNAs are translated repeatedly in these cells.

Nucleotide sequences that affect how long an mRNA remains intact are often found in the untranslated region (UTR) at the 3' end of the molecule (see Figure 18.8). In one experiment, researchers transferred such a sequence from the short-lived mRNA for a growth factor to the 3' end of a normally stable globin mRNA. The globin mRNA was quickly degraded.

During the past few years, other mechanisms that degrade or block expression of mRNA molecules have come to light. These mechanisms involve an important group of newly discovered RNA molecules that regulate gene expression at several levels, and we will discuss them later in this chapter.

Initiation of Translation

Translation presents another opportunity for regulating gene expression; such regulation occurs most commonly at the initiation stage (see Figure 17.18). For some mRNAs, the initiation of translation can be blocked by regulatory proteins that bind to specific sequences or structures within the untranslated region at the 5' or 3' end (5' or 3' UTR), preventing the attachment of ribosomes. (Recall from Chapter 17 that both the 5' cap and the poly-A tail of an mRNA molecule are important for ribosome binding.) A different mechanism for blocking translation is seen in a variety of mRNAs present in

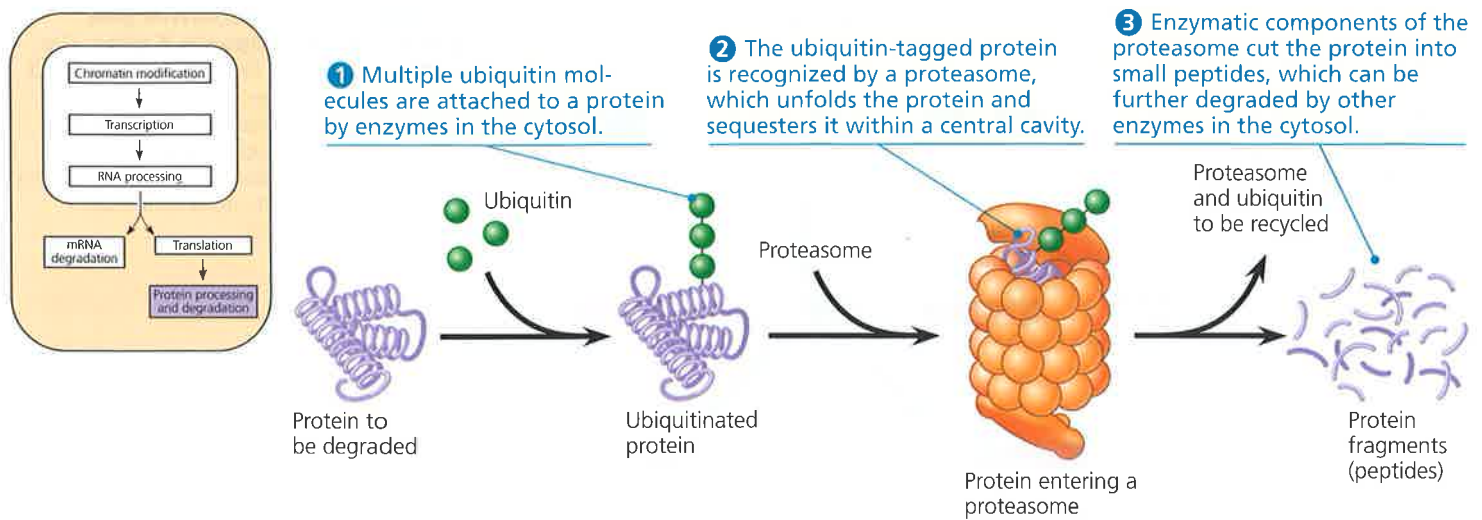
the eggs of many organisms: Initially, these stored mRNAs lack poly-A tails of sufficient length to allow translation initiation. At the appropriate time during embryonic development, however, a cytoplasmic enzyme adds more adenine (A) nucleotides, prompting translation to begin.

Alternatively, translation of *all* the mRNAs in a cell may be regulated simultaneously. In a eukaryotic cell, such "global" control usually involves the activation or inactivation of one or more of the protein factors required to initiate translation. This mechanism plays a role in starting translation of mRNAs that are stored in eggs. Just after fertilization, translation is triggered by the sudden activation of translation initiation factors. The response is a burst of synthesis of the proteins encoded by the stored mRNAs. Some plants and algae store mRNAs during periods of darkness; light then triggers the reactivation of the translational apparatus.

Protein Processing and Degradation

The final opportunities for controlling gene expression occur after translation. Often, eukaryotic polypeptides must be processed to yield functional protein molecules. For instance, cleavage of the initial insulin polypeptide (pro-insulin) forms the active hormone. In addition, many proteins undergo chemical modifications that make them functional. Regulatory proteins are commonly activated or inactivated by the reversible addition of phosphate groups, and proteins destined for the surface of animal cells acquire sugars. Cell-surface proteins and many others must also be transported to target destinations in the cell in order to function. Regulation might occur at any of the steps involved in modifying or transporting a protein.

Finally, the length of time each protein functions in the cell is strictly regulated by means of selective degradation. Many proteins, such as the cyclins involved in regulating the cell cycle, must be relatively short-lived if the cell is to function appropriately (see Figure 12.17). To mark a particular protein for destruction, the cell commonly attaches molecules of a small protein called ubiquitin to the protein. Giant protein complexes called **proteasomes** then recognize the



▲ **Figure 18.14 Degradation of a protein by a proteasome.** A proteasome, an enormous protein complex shaped like a trash can, chops up unneeded proteins in the cell. In most cases, the proteins attacked by a

proteasome have been tagged with short chains of ubiquitin, a small protein. Steps 1 and 3 require ATP. Eukaryotic proteasomes are as massive as ribosomal subunits and are

distributed throughout the cell. Their shape somewhat resembles that of chaperone proteins, which protect protein structure rather than destroy it (see Figure 5.23).

ubiquitin-tagged proteins and degrade them (**Figure 18.14**). The importance of proteasomes is underscored by the finding that mutations making specific cell cycle proteins impervious to proteasome degradation can lead to cancer. The 2004 Nobel Prize in Chemistry was awarded to three scientists—two from Israel and one from the United States—who worked out the regulated process of protein degradation.

CONCEPT CHECK 18.2

1. In general, what is the effect of histone acetylation and DNA methylation on gene expression?
2. Compare the roles of general and specific transcription factors in regulating gene expression.
3. Suppose you compared the nucleotide sequences of the distal control elements in the enhancers of three genes that are expressed only in muscle tissue. What would you expect to find? Why?
4. Once mRNA encoding a particular protein reaches the cytoplasm, what are four mechanisms that can regulate the amount of the protein that is active in the cell?
5. **WHAT IF?** Examine Figure 18.11 and suggest a mechanism by which the yellow activator protein comes to be present in the liver cell but not in the lens cell.

For suggested answers, see Appendix A.

CONCEPT 18.3

Noncoding RNAs play multiple roles in controlling gene expression

Genome sequencing has revealed that protein-coding DNA accounts for only 1.5% of the human genome and a similarly

small percentage of the genomes of many other multicellular eukaryotes. A very small fraction of the non-protein-coding DNA consists of genes for RNAs such as ribosomal RNA and transfer RNA. Until recently, most of the remaining DNA was assumed to be untranscribed. The idea was that since it didn't specify proteins or the few known types of RNA, such DNA didn't contain meaningful genetic information. However, a flood of recent data has contradicted this idea. For example, an in-depth study of a region comprising 1% of the human genome showed that more than 90% of that region was transcribed. Introns accounted for only a fraction of this transcribed, nontranslated RNA. These and other results suggest that a significant amount of the genome may be transcribed into non-protein-coding RNAs (also called *noncoding RNAs*, or *ncRNAs*), including a variety of small RNAs. While many questions about the functions of these RNAs remain unanswered, researchers are uncovering more evidence of their biological roles every day.

Biologists are excited about these recent discoveries, which hint at a large, diverse population of RNA molecules in the cell that play crucial roles in regulating gene expression—and have gone largely unnoticed until now. Clearly, we must revise our long-standing view that because mRNAs code for proteins, they are the most important RNAs functioning in the cell. This represents a major shift in the thinking of biologists, one that you are witnessing as students entering this field of study. It's as if our exclusive focus on a famous rock star has blinded us to the many backup musicians and songwriters working behind the scenes.

Regulation by both small and large ncRNAs is known to occur at several points in the pathway of gene expression, including mRNA translation and chromatin modification. We will focus mainly on two types of small ncRNAs that have been extensively studied in the past few years; the importance

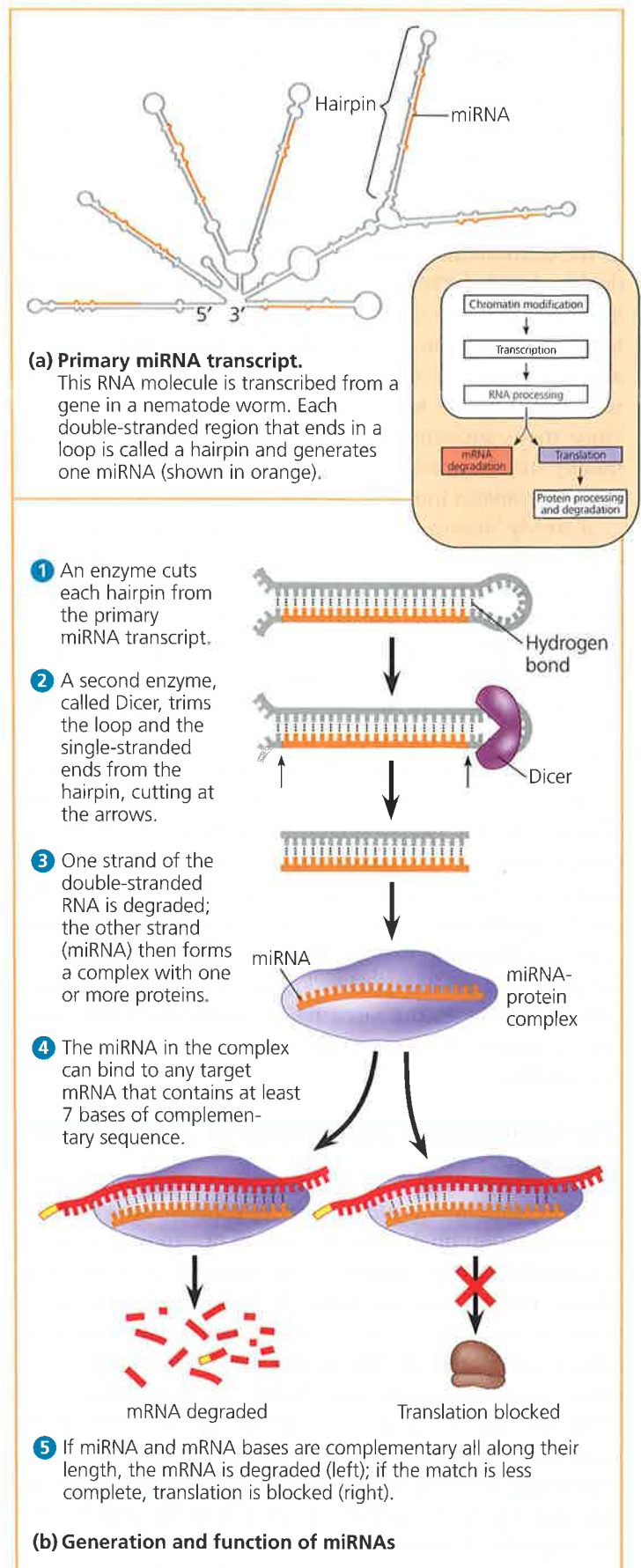
of these RNAs was acknowledged when they were the focus of the 2006 Nobel Prize in Physiology or Medicine.

Effects on mRNAs by MicroRNAs and Small Interfering RNAs

Since 1993, a number of research studies have uncovered small single-stranded RNA molecules, called **microRNAs (miRNAs)**, that are capable of binding to complementary sequences in mRNA molecules. The miRNAs are made from longer RNA precursors that fold back on themselves, forming one or more short double-stranded hairpin structures, each held together by hydrogen bonds (**Figure 18.15**). After each hairpin is cut away from the precursor, it is trimmed by an enzyme (fittingly called Dicer) into a short double-stranded fragment of about 22 nucleotide pairs. One of the two strands is degraded, while the other strand, which is the miRNA, forms a complex with one or more proteins; the miRNA allows the complex to bind to any mRNA molecule with 7–8 nucleotides of complementary sequence. The miRNA-protein complex then either degrades the target mRNA or blocks its translation. It has been estimated that expression of at least one-half of all human genes may be regulated by miRNAs, a remarkable figure given that the existence of miRNAs was unknown a mere two decades ago.

A growing understanding of the miRNA pathway provided an explanation for a perplexing observation: Researchers had found that injecting double-stranded RNA molecules into a cell somehow turned off expression of a gene with the same sequence as the RNA. They called this experimental phenomenon **RNA interference (RNAi)**. It was later shown to be due to **small interfering RNAs (siRNAs)**, which are similar in size and function to miRNAs. In fact, subsequent research showed that the same cellular machinery generates miRNAs and siRNAs and that both can associate with the same proteins, producing similar results. The distinction between miRNAs and siRNAs is based on the nature of the precursor molecule for each. While an miRNA is usually formed from a single hairpin in a precursor RNA (see Figure 18.15), multiple siRNAs are formed from a much longer, linear, double-stranded RNA molecule.

We mentioned that laboratory investigators had injected double-stranded RNAs into cells, and you may wonder whether such molecules are ever found naturally. As you will learn in Chapter 19, some viruses have double-stranded RNA genomes. Because the cellular RNAi pathway can lead to the destruction of RNAs with sequences complementary to those found in double-stranded RNAs, this pathway may have evolved as a natural defense against infection by such viruses. However, the fact that RNAi can also affect the expression of nonviral cellular genes may reflect a different evolutionary origin for the RNAi pathway. Moreover, many species, including mammals, apparently produce their own long, double-stranded RNA precursors to small RNAs such as siRNAs. Once produced, these RNAs can interfere with gene expression at stages other than translation, as we'll discuss next.



▲ **Figure 18.15** Regulation of gene expression by miRNAs.

Chromatin Remodeling and Effects on Transcription by ncRNAs

In addition to affecting mRNAs, small RNAs can cause remodeling of chromatin structure. In some yeasts, siRNAs produced by the yeast cells themselves are required for the formation of heterochromatin at the centromeres of chromosomes. According to one model, an RNA transcript produced from DNA in the centromeric region of the chromosome is copied into double-stranded RNA by a yeast enzyme and then processed into siRNAs. These siRNAs associate with a complex of proteins (different from the one shown in Figure 18.15) and act as a homing device, targeting the complex back to RNA transcripts being made from the centromeric sequences of DNA. Once there, proteins in the complex recruit enzymes that modify the chromatin, turning it into the highly condensed heterochromatin found at the centromere.

A newly discovered class of small ncRNAs called *piwi-associated RNAs* (*piRNAs*) also induce formation of heterochromatin, blocking expression of some parasitic DNA elements in the genome known as transposons. (Transposons are discussed in Chapter 21.) Usually 24–31 nucleotides in length, piRNAs are probably processed from single-stranded RNA precursors. They play an indispensable role in the germ cells of many animal species, where they appear to help re-establish appropriate methylation patterns in the genome during gamete formation.

The cases we have just described involve chromatin remodeling that blocks expression of large regions of the chromosome. Several recent experiments have shown that related RNA-based mechanisms may also block the transcription of specific genes. For instance, some plant miRNAs have sequences that bind to gene promoters and can repress transcription, and piRNAs can block expression of specific genes. And in a twist on the same theme, some cases have even been reported of *activation* of gene expression by miRNAs and piRNAs.

The Evolutionary Significance of Small ncRNAs

EVOLUTION Small ncRNAs can regulate gene expression at multiple steps and in many ways. In general, extra levels of gene regulation might allow evolution of a higher degree of complexity of form. Therefore, the versatility of miRNA regulation has led some biologists to hypothesize that an increase in the number of miRNAs specified by the genome of a given species has allowed morphological complexity to increase over evolutionary time. While this hypothesis is still being debated, it is logical to expand the discussion to include all small ncRNAs. Exciting new techniques for rapidly sequencing genomes are beginning to allow biologists to ask how many genes for ncRNAs are present in the genome of a given species. A survey of different species supports the notion that siRNAs evolved first, followed by miRNAs and later piRNAs, which are found only in animals. And while there

are hundreds of types of miRNAs, there appear to be many thousands of types of piRNAs, allowing the potential for very sophisticated gene regulation by piRNAs.

Given the extensive functions of ncRNAs, it is not surprising that many of the ncRNAs characterized thus far play important roles in embryonic development—the topic we turn to in the next section. Embryonic development is perhaps the ultimate example of precisely regulated gene expression.

CONCEPT CHECK 18.3

1. Compare and contrast miRNAs and siRNAs.
2. **WHAT IF?** If the mRNA being degraded in Figure 18.15 coded for a protein that promotes cell division in a multicellular organism, what would happen if a mutation disabled the gene encoding the miRNA that triggers this degradation?
3. **MAKE CONNECTIONS** In Concept 15.2 (pp. 337–338), you learned about inactivation of one of the X chromosomes in female mammals. Reread those pages, and suggest a model for how the *XIST* noncoding RNA functions to cause Barr body formation.

For suggested answers, see Appendix A.

CONCEPT 18.4

A program of differential gene expression leads to the different cell types in a multicellular organism

In the embryonic development of multicellular organisms, a fertilized egg (a zygote) gives rise to cells of many different types, each with a different structure and corresponding function. Typically, cells are organized into tissues, tissues into organs, organs into organ systems, and organ systems into the whole organism. Thus, any developmental program must produce cells of different types that form higher-level structures arranged in a particular way in three dimensions. The processes that occur during development in plants and animals are detailed in Chapters 35 and 47, respectively. In this chapter, we focus instead on the program of regulation of gene expression that orchestrates development, using a few animal species as examples.

A Genetic Program for Embryonic Development

The photos in **Figure 18.16** illustrate the dramatic difference between a zygote and the organism it becomes. This remarkable transformation results from three interrelated processes: cell division, cell differentiation, and morphogenesis. Through a succession of mitotic cell divisions, the zygote gives rise to a



(a) Fertilized eggs of a frog

(b) Newly hatched tadpole

▲ **Figure 18.16 From fertilized egg to animal: What a difference four days makes.** It takes just four days for cell division, differentiation, and morphogenesis to transform each of the fertilized frog eggs shown in (a) into a tadpole like the one in (b).

large number of cells. Cell division alone, however, would merely produce a great ball of identical cells, nothing like a tadpole. During embryonic development, cells not only increase in number, but also undergo cell **differentiation**, the process by which cells become specialized in structure and function. Moreover, the different kinds of cells are not randomly distributed but are organized into tissues and organs in a particular three-dimensional arrangement. The physical processes that give an organism its shape constitute **morphogenesis**, meaning “creation of form.”

All three processes have their basis in cellular behavior. Even morphogenesis, the shaping of the organism, can be traced back to changes in the shape, motility, and other characteristics of the cells that make up various regions of the embryo. As you have seen, the activities of a cell depend on the genes it expresses and the proteins it produces. Almost all cells in an organism have the same genome; therefore, differential gene expression results from the genes being regulated differently in each cell type.

In Figure 18.11, you saw a simplified view of how differential gene expression occurs in two cell types, a liver cell and a lens cell. Each of these fully differentiated cells has a particular mix of specific activators that turn on the collection of genes whose products are required in the cell. The fact that both cells arose through a series of mitoses from a common fertilized egg inevitably leads to a question: How do different sets of activators come to be present in the two cells?

It turns out that materials placed into the egg by the mother set up a sequential program of gene regulation that is carried out as cells divide, and this program makes the cells become different from each other in a coordinated fashion. To understand how this works, we will consider two basic developmental processes: First, we’ll explore how cells that arise from early embryonic mitoses develop the differences that start each cell along its own differentiation pathway. Second, we’ll see how cellular differentiation leads to one particular cell type, using muscle development as an example.

Cytoplasmic Determinants and Inductive Signals

What generates the first differences among cells in an early embryo? And what controls the differentiation of all the various cell types as development proceeds? By this point in the chapter, you can probably deduce the answer: The specific genes expressed in any particular cell of a developing organism determine its path. Two sources of information, used to varying extents in different species, “tell” a cell which genes to express at any given time during embryonic development.

One important source of information early in development is the egg’s cytoplasm, which contains both RNA and proteins encoded by the mother’s DNA. The cytoplasm of an unfertilized egg is not homogeneous. Messenger RNA, proteins, other substances, and organelles are distributed unevenly in the unfertilized egg, and this unevenness has a profound impact on the development of the future embryo in many species. Maternal substances in the egg that influence the course of early development are called **cytoplasmic determinants** (Figure 18.17a, on the next page). After fertilization, early mitotic divisions distribute the zygote’s cytoplasm into separate cells. The nuclei of these cells may thus be exposed to different cytoplasmic determinants, depending on which portions of the zygotic cytoplasm a cell received. The combination of cytoplasmic determinants in a cell helps determine its developmental fate by regulating expression of the cell’s genes during the course of cell differentiation.

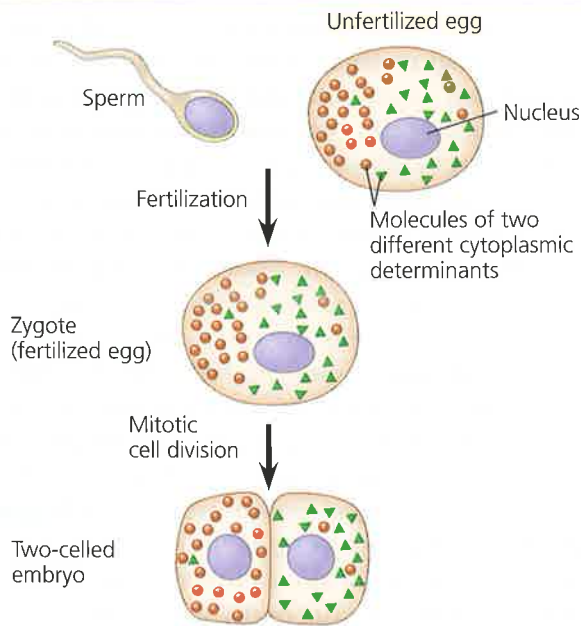
The other major source of developmental information, which becomes increasingly important as the number of embryonic cells increases, is the environment around a particular cell. Most influential are the signals impinging on an embryonic cell from other embryonic cells in the vicinity, including contact with cell-surface molecules on neighboring cells and the binding of growth factors secreted by neighboring cells. Such signals cause changes in the target cells, a process called **induction** (Figure 18.17b). The molecules conveying these signals within the target cell are cell-surface receptors and other proteins expressed by the embryo’s own genes. In general, the signaling molecules send a cell down a specific developmental path by causing changes in its gene expression that eventually result in observable cellular changes. Thus, interactions between embryonic cells help induce differentiation of the many specialized cell types making up a new organism.

Sequential Regulation of Gene Expression During Cellular Differentiation

As the tissues and organs of an embryo develop and their cells differentiate, the cells become noticeably different in structure and function. These observable changes are actually the outcome of a cell’s developmental history beginning at the first mitotic division of the zygote, as we have just seen. The

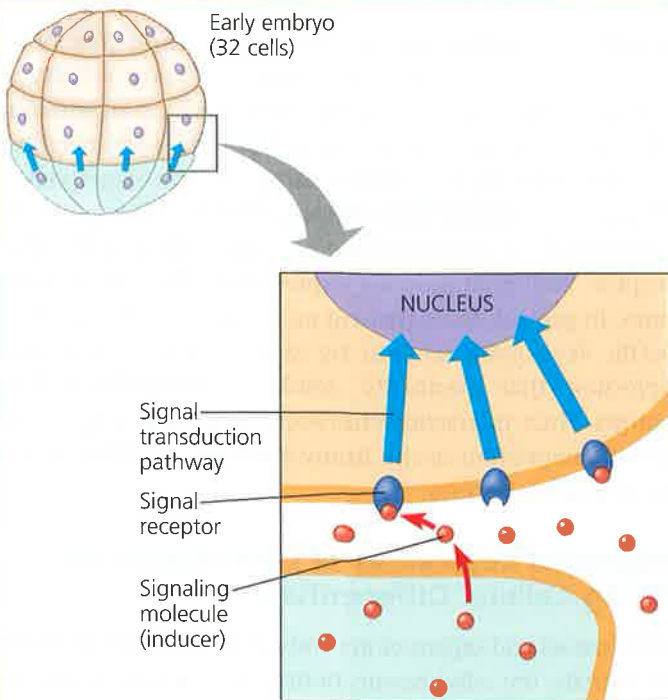
▼ Figure 18.17 Sources of developmental information for the early embryo.

(a) Cytoplasmic determinants in the egg



The unfertilized egg has molecules in its cytoplasm, encoded by the mother's genes, that influence development. Many of these cytoplasmic determinants, like the two shown here, are unevenly distributed in the egg. After fertilization and mitotic division, the cell nuclei of the embryo are exposed to different sets of cytoplasmic determinants and, as a result, express different genes.

(b) Induction by nearby cells



The cells at the bottom of the early embryo depicted here are releasing chemicals that signal nearby cells to change their gene expression.

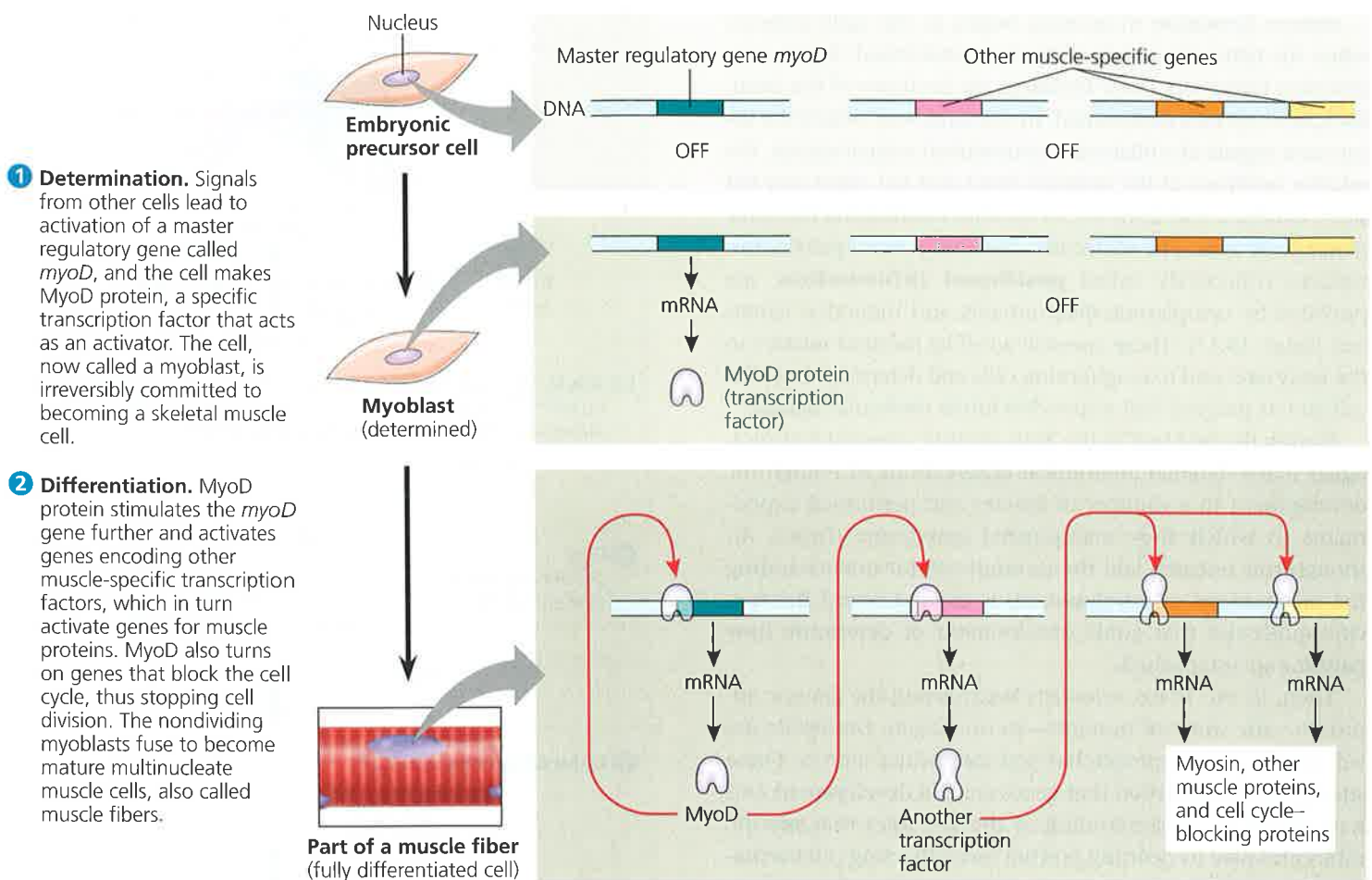
earliest changes that set a cell on a path to specialization are subtle ones, showing up only at the molecular level. Before biologists knew much about the molecular changes occurring in embryos, they coined the term **determination** to refer to the events that lead to the observable differentiation of a cell. Once it has undergone determination, an embryonic cell is irreversibly committed to its final fate. If a committed cell is experimentally placed in another location in the embryo, it will still differentiate into the cell type that is its normal fate.

Today we understand determination in terms of molecular changes. The outcome of determination, observable cell differentiation, is marked by the expression of genes for *tissue-specific proteins*. These proteins are found only in a specific cell type and give the cell its characteristic structure and function. The first evidence of differentiation is the appearance of mRNAs for these proteins. Eventually, differentiation is observable with a microscope as changes in cellular structure. On the molecular level, different sets of genes are sequentially expressed in a regulated manner as new cells arise from division of their precursors. A number of the steps in gene expression may be regulated during differentiation, with transcription among the most important. In the fully differentiated cell, transcription remains the principal regulatory point for maintaining appropriate gene expression.

Differentiated cells are specialists at making tissue-specific proteins. For example, as a result of transcriptional regulation, liver cells specialize in making albumin, and lens cells specialize in making crystallin (see Figure 18.11). Skeletal muscle cells in vertebrates are another instructive example. Each of these cells is a long fiber containing many nuclei within a single plasma membrane. Skeletal muscle cells have high concentrations of muscle-specific versions of the contractile proteins myosin and actin, as well as membrane receptor proteins that detect signals from nerve cells.

Muscle cells develop from embryonic precursor cells that have the potential to develop into a number of cell types, including cartilage cells and fat cells, but particular conditions commit them to becoming muscle cells. Although the committed cells appear unchanged under the microscope, determination has occurred, and they are now *myoblasts*. Eventually, myoblasts start to churn out large amounts of muscle-specific proteins and fuse to form mature, elongated, multinucleate skeletal muscle cells (**Figure 18.18**).

Researchers have worked out what happens at the molecular level during muscle cell determination by growing myoblasts in culture and analyzing them using molecular biological techniques you will learn about in Chapter 20. In a series of experiments, they isolated different genes, caused each to be expressed in a separate embryonic precursor cell, and then looked for differentiation into myoblasts and muscle cells. In this way, they identified several so-called "master regulatory genes" whose protein products commit the cells to becoming skeletal muscle. Thus, in the case of muscle cells,



▲ **Figure 18.18 Determination and differentiation of muscle cells.** Skeletal muscle cells arise from embryonic cells as a result of changes in gene expression. (In this depiction, the process of gene activation is greatly simplified.)

WHAT IF? What would happen if a mutation in the *myoD* gene resulted in a MyoD protein that could not activate the *myoD* gene?

the molecular basis of determination is the expression of one or more of these master regulatory genes.

To understand more about how commitment occurs in muscle cell differentiation, let's focus on the master regulatory gene called *myoD* (see Figure 18.18). This gene encodes MyoD protein, a transcription factor that binds to specific control elements in the enhancers of various target genes and stimulates their expression (see Figure 18.9). Some target genes for MyoD encode still other muscle-specific transcription factors. MyoD also stimulates expression of the *myoD* gene itself, thus perpetuating its effect in maintaining the cell's differentiated state. Presumably, all the genes activated by MyoD have enhancer control elements recognized by MyoD and are thus coordinately controlled. Finally, the secondary transcription factors activate the genes for proteins such as myosin and actin that confer the unique properties of skeletal muscle cells.

The MyoD protein deserves its designation as a master regulatory gene. Researchers have shown that it is even capable of changing some kinds of fully differentiated nonmuscle cells, such as fat cells and liver cells, into muscle cells. Why

doesn't it work on *all* kinds of cells? One likely explanation is that activation of the muscle-specific genes is not solely dependent on MyoD but requires a particular *combination* of regulatory proteins, some of which are lacking in cells that do not respond to MyoD. The determination and differentiation of other kinds of tissues may play out in a similar fashion.

We have now seen how different programs of gene expression that are activated in the fertilized egg can result in differentiated cells and tissues. But for the tissues to function effectively in the organism as a whole, the organism's *body plan*—its overall three-dimensional arrangement—must be established and superimposed on the differentiation process. Next we'll investigate the molecular basis for the establishment of the body plan, using the well-studied *Drosophila* as an example.

Pattern Formation: Setting Up the Body Plan

Cytoplasmic determinants and inductive signals both contribute to the development of a spatial organization in which the tissues and organs of an organism are all in their characteristic places. This process is called **pattern formation**.

Pattern formation in animals begins in the early embryo, when the major axes of an animal are established. Before construction begins on a new building, the locations of the front, back, and sides are determined. In the same way, before the tissues and organs of a bilaterally symmetrical animal appear, the relative positions of the animal's head and tail, right and left sides, and back and front are set up, thus establishing the three major body axes. The molecular cues that control pattern formation, collectively called **positional information**, are provided by cytoplasmic determinants and inductive signals (see Figure 18.17). These cues tell a cell its location relative to the body axes and to neighboring cells and determine how the cell and its progeny will respond to future molecular signals.

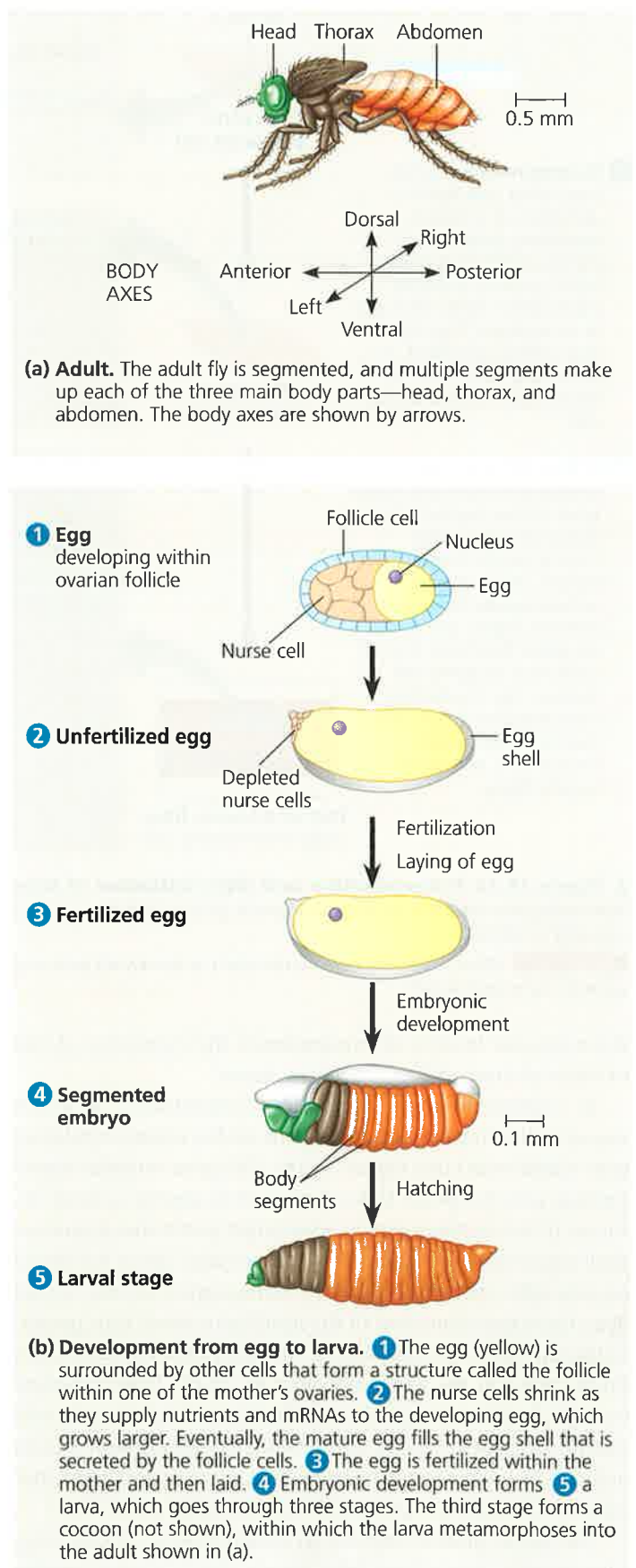
During the first half of the 20th century, classical embryologists made detailed anatomical observations of embryonic development in a number of species and performed experiments in which they manipulated embryonic tissues. Although this research laid the groundwork for understanding the mechanisms of development, it did not reveal the specific molecules that guide development or determine how patterns are established.

Then, in the 1940s, scientists began using the genetic approach—the study of mutants—to investigate *Drosophila* development. That approach has had spectacular success. These studies have established that genes control development and have led to an understanding of the key roles that specific molecules play in defining position and directing differentiation. By combining anatomical, genetic, and biochemical approaches to the study of *Drosophila* development, researchers have discovered developmental principles common to many other species, including humans.

The Life Cycle of *Drosophila*

Fruit flies and other arthropods have a modular construction, an ordered series of segments. These segments make up the body's three major parts: the head, the thorax (the midbody, from which the wings and legs extend), and the abdomen (Figure 18.19a). Like other bilaterally symmetrical animals, *Drosophila* has an anterior-posterior (head-to-tail) axis, a dorsal-ventral (back-to-belly) axis, and a right-left axis. In *Drosophila*, cytoplasmic determinants that are localized in the unfertilized egg provide positional information for the placement of anterior-posterior and dorsal-ventral axes even before fertilization. We'll focus here on the molecules involved in establishing the anterior-posterior axis.

The *Drosophila* egg develops in the female's ovary, surrounded by ovarian cells called nurse cells and follicle cells (Figure 18.19b, top). These support cells supply the egg with nutrients, mRNAs, and other substances needed for development and make the egg shell. After fertilization and laying of the egg, embryonic development results in the formation of a segmented larva, which goes through three larval stages. Then, in a process much like that by which a caterpillar



▲ **Figure 18.19** Key developmental events in the life cycle of *Drosophila*.

becomes a butterfly, the fly larva forms a cocoon in which it metamorphoses into the adult fly pictured in Figure 18.19a.

Genetic Analysis of Early Development: Scientific Inquiry

Edward B. Lewis was a visionary American biologist who, in the 1940s, first showed the value of the genetic approach to studying embryonic development in *Drosophila*. Lewis studied bizarre mutant flies with developmental defects that led to extra wings or legs in the wrong place (Figure 18.20). He located the mutations on the fly's genetic map, thus connecting the developmental abnormalities to specific genes. This research supplied the first concrete evidence that genes somehow direct the developmental processes studied by embryologists. The genes Lewis discovered, called **homeotic genes**, control pattern formation in the late embryo, larva, and adult.

Insight into pattern formation during early embryonic development did not come for another 30 years, when two researchers in Germany, Christiane Nüsslein-Volhard and Eric Wieschaus, set out to identify *all* the genes that affect segment formation in *Drosophila*. The project was daunting for three reasons. The first was the sheer number of *Drosophila* genes, now known to total about 13,700. The genes affecting segmentation might be just a few needles in a haystack or might be so numerous and varied that the scientists would be unable to make sense of them. Second, mutations affecting a

process as fundamental as segmentation would surely be **embryonic lethals**, mutations with phenotypes causing death at the embryonic or larval stage. Because organisms with embryonic lethal mutations never reproduce, they cannot be bred for study. The researchers dealt with this problem by looking for recessive mutations, which can be propagated in heterozygous flies that act as genetic carriers. Third, cytoplasmic determinants in the egg were known to play a role in axis formation, so the researchers knew they would have to study the mother's genes as well as those of the embryo. It is the mother's genes that we will discuss further as we focus on how the anterior-posterior body axis is set up in the developing egg.

Nüsslein-Volhard and Wieschaus began their search for segmentation genes by exposing flies to a mutagenic chemical that affected the flies' gametes. They mated the mutagenized flies and then scanned their descendants for dead embryos or larvae with abnormal segmentation or other defects. For example, to find genes that might set up the anterior-posterior axis, they looked for embryos or larvae with abnormal ends, such as two heads or two tails, predicting that such abnormalities would arise from mutations in maternal genes required for correctly setting up the offspring's head or tail end.

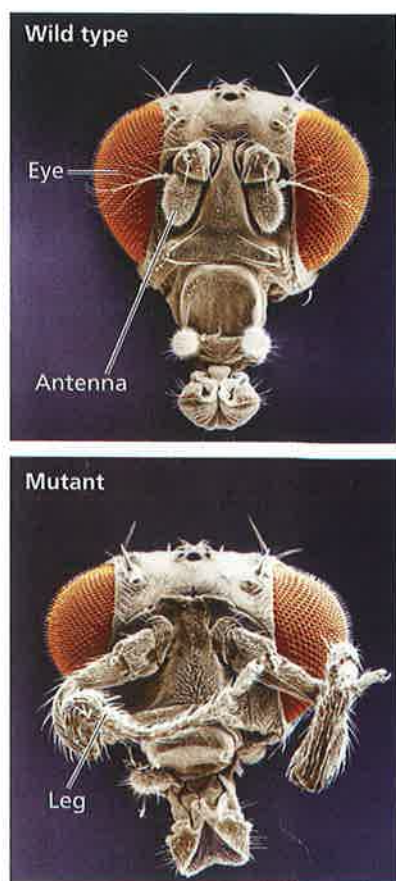
Using this approach, Nüsslein-Volhard and Wieschaus eventually identified about 1,200 genes essential for pattern formation during embryonic development. Of these, about 120 were essential for normal segmentation. Over several years, the researchers were able to group these segmentation genes by general function, to map them, and to clone many of them for further study in the lab. The result was a detailed molecular understanding of the early steps in pattern formation in *Drosophila*.

When the results of Nüsslein-Volhard and Wieschaus were combined with Lewis's earlier work, a coherent picture of *Drosophila* development emerged. In recognition of their discoveries, the three researchers were awarded a Nobel Prize in 1995.

Let's consider further the genes that Nüsslein-Volhard, Wieschaus, and co-workers found for cytoplasmic determinants deposited in the egg by the mother. These genes set up the initial pattern of the embryo by regulating gene expression in broad regions of the early embryo.

Axis Establishment

As we mentioned earlier, cytoplasmic determinants in the egg are the substances that initially establish the axes of the *Drosophila* body. These substances are encoded by genes of the mother, fittingly called maternal effect genes. A **maternal effect gene** is a gene that, when mutant in the mother, results in a mutant phenotype in the offspring, regardless of the offspring's own genotype. In fruit fly development, the mRNA or protein products of maternal effect genes are placed in the egg while it is still in the mother's ovary. When the mother



◀ **Figure 18.20**
Abnormal pattern formation in *Drosophila*. Mutations in certain regulatory genes, called homeotic genes, cause a misplacement of structures in an animal. These scanning electron micrographs contrast the head of a wild-type fly, bearing a pair of small antennae, with that of a homeotic mutant (a fly with a mutation in a single gene), bearing a pair of legs in place of antennae.

has a mutation in such a gene, she makes a defective gene product (or none at all), and her eggs are defective; when these eggs are fertilized, they fail to develop properly.

Because they control the orientation (polarity) of the egg and consequently of the fly, maternal effect genes are also called **egg-polarity genes**. One group of these genes sets up the anterior-posterior axis of the embryo, while a second group establishes the dorsal-ventral axis. Like mutations in segmentation genes, mutations in maternal effect genes are generally embryonic lethals.

Bicoid: A Morphogen Determining Head Structures To see how maternal effect genes determine the body axes of the offspring, we will focus on one such gene, called ***bicoid***, a term meaning “two-tailed.” An embryo whose mother has two mutant alleles of the *bicoid* gene lacks the front half of its body and has posterior structures at both ends (Figure 18.21). This phenotype suggested to Nüsslein-Volhard and her colleagues that the product of the mother’s *bicoid* gene is essential for setting up the anterior end of the fly and might be concentrated at the future anterior end of the embryo. This hypothesis is an example of the *morphogen gradient hypothesis* first proposed by embryologists a century ago; in this hypothesis, gradients of substances called **morphogens** establish an embryo’s axes and other features of its form.

DNA technology and other modern biochemical methods enabled the researchers to test whether the *bicoid* product, a protein called Bicoid, is in fact a morphogen that determines the anterior end of the fly. The first question they asked was whether the mRNA and protein products of these genes are located in the egg in a position consistent with the hypothesis.

They found that *bicoid* mRNA is highly concentrated at the extreme anterior end of the mature egg, as predicted by the hypothesis (Figure 18.22). After the egg is fertilized, the mRNA is translated into protein. The Bicoid protein then diffuses from the anterior end toward the posterior, resulting in a gradient of protein within the early embryo, with the highest concentration at the anterior end. These results are consistent with the hypothesis that Bicoid protein specifies the fly’s anterior end. To test the hypothesis more specifically, scientists injected pure *bicoid* mRNA into various regions of early embryos. The protein that resulted from its translation caused anterior structures to form at the injection sites.

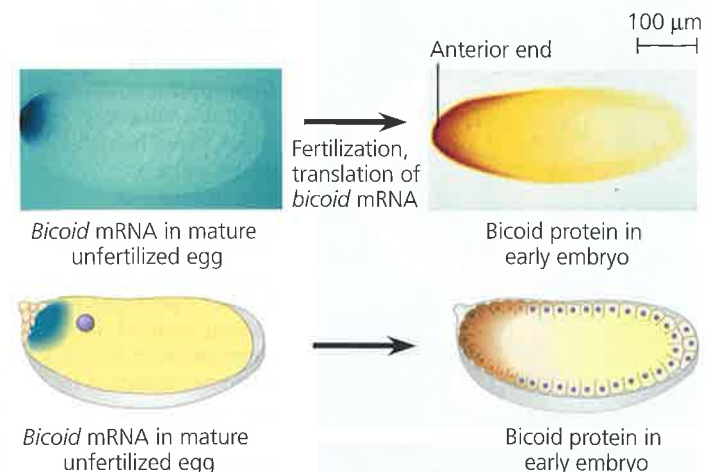
▼ Figure 18.22

INQUIRY

Is Bicoid a morphogen that determines the anterior end of a fruit fly?

EXPERIMENT Using a genetic approach to study *Drosophila* development, Christiane Nüsslein-Volhard and colleagues at the European Molecular Biology Laboratory in Heidelberg, Germany, analyzed expression of the *bicoid* gene. The researchers hypothesized that *bicoid* normally codes for a morphogen that specifies the head (anterior) end of the embryo. To test this hypothesis, they used molecular techniques to determine where the mRNA and protein encoded by this gene were found in the fertilized egg and early embryo of wild-type flies.

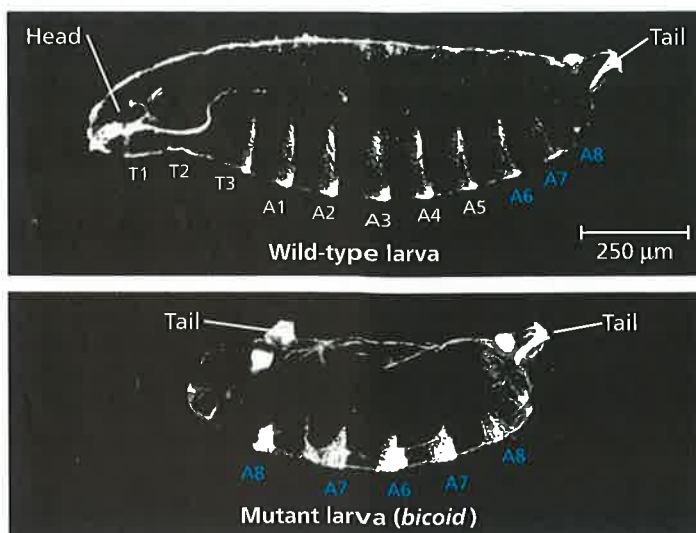
RESULTS *Bicoid* mRNA (dark blue) was confined to the anterior end of the unfertilized egg. Later in development, Bicoid protein (dark orange) was seen to be concentrated in cells at the anterior end of the embryo.



CONCLUSION The location of *bicoid* mRNA and the diffuse gradient of Bicoid protein seen later support the hypothesis that Bicoid protein is a morphogen specifying formation of head-specific structures.

SOURCE: C. Nüsslein-Volhard et al., Determination of anteroposterior polarity in *Drosophila*, *Science* 238:1675–1681 (1987); W. Driever and C. Nüsslein-Volhard, A gradient of *bicoid* protein in *Drosophila* embryos, *Cell* 54:83–93 (1988); T. Berleth et al., The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo, *EMBO Journal* 7:1749–1756 (1988).

WHAT IF? If the hypothesis is correct, predict what would happen if you injected *bicoid* mRNA into the anterior end of an egg from a female with a mutation disabling the *bicoid* gene.



▲ **Figure 18.21 Effect of the *bicoid* gene on *Drosophila* development.** A wild-type fruit fly larva has a head, three thoracic (T) segments, eight abdominal (A) segments, and a tail. A larva whose mother has two mutant alleles of the *bicoid* gene has two tails and lacks all anterior structures (LMs).

The *bicoid* research was groundbreaking for several reasons. First, it led to the identification of a specific protein required for some of the earliest steps in pattern formation. It thus helped us understand how different regions of the egg can give rise to cells that go down different developmental pathways. Second, it increased our understanding of the mother's critical role in the initial phases of embryonic development. Finally, the principle that a gradient of morphogens can determine polarity and position has proved to be a key developmental concept for a number of species, just as early embryologists had thought.

Maternal mRNAs are crucial during development of many species. In *Drosophila*, gradients of specific proteins encoded by maternal mRNAs determine the posterior and anterior ends and establish the dorsal-ventral axis. As the fly embryo grows, it reaches a point when the embryonic program of gene expression takes over, and the maternal mRNAs must be destroyed. (This process involves miRNAs in *Drosophila* and other species.) Later, positional information encoded by the embryo's genes, operating on an ever finer scale, establishes a specific number of correctly oriented segments and triggers the formation of each segment's characteristic structures. When the genes operating in this final step are abnormal, the pattern of the adult is abnormal, as you saw in Figure 18.20.

In this section, we have seen how a carefully orchestrated program of sequential gene regulation controls the transformation of a fertilized egg into a multicellular organism. The program is carefully balanced between turning on the genes for differentiation in the right place and turning off other genes. Even when an organism is fully developed, gene expression is regulated in a similarly fine-tuned manner. In the final section of the chapter, we'll consider how fine this tuning is by looking at how specific changes in expression of one or a few genes can lead to the development of cancer.

CONCEPT CHECK 18.4

1. As you learned in Chapter 12, mitosis gives rise to two daughter cells that are genetically identical to the parent cell. Yet you, the product of many mitotic divisions, are not composed of identical cells. Why?
2. **MAKE CONNECTIONS** Explain how the signaling molecules released by an embryonic cell can induce changes in a neighboring cell without entering the cell. (See Figures 11.15 and 11.16, pp. 265 and 266.)
3. Why are fruit fly maternal effect genes also called egg-polarity genes?
4. **WHAT IF?** In the blowup box in Figure 18.17b, the lower cell is synthesizing signaling molecules, whereas the upper cell is expressing receptors for these molecules. In terms of gene regulation, explain how these cells came to synthesize different molecules.

For suggested answers, see Appendix A.

CONCEPT 18.5

Cancer results from genetic changes that affect cell cycle control

In Chapter 12, we considered cancer as a set of diseases in which cells escape from the control mechanisms that normally limit their growth. Now that we have discussed the molecular basis of gene expression and its regulation, we are ready to look at cancer more closely. The gene regulation systems that go wrong during cancer turn out to be the very same systems that play important roles in embryonic development, the immune response, and many other biological processes. Thus, research into the molecular basis of cancer has both benefited from and informed many other fields of biology.

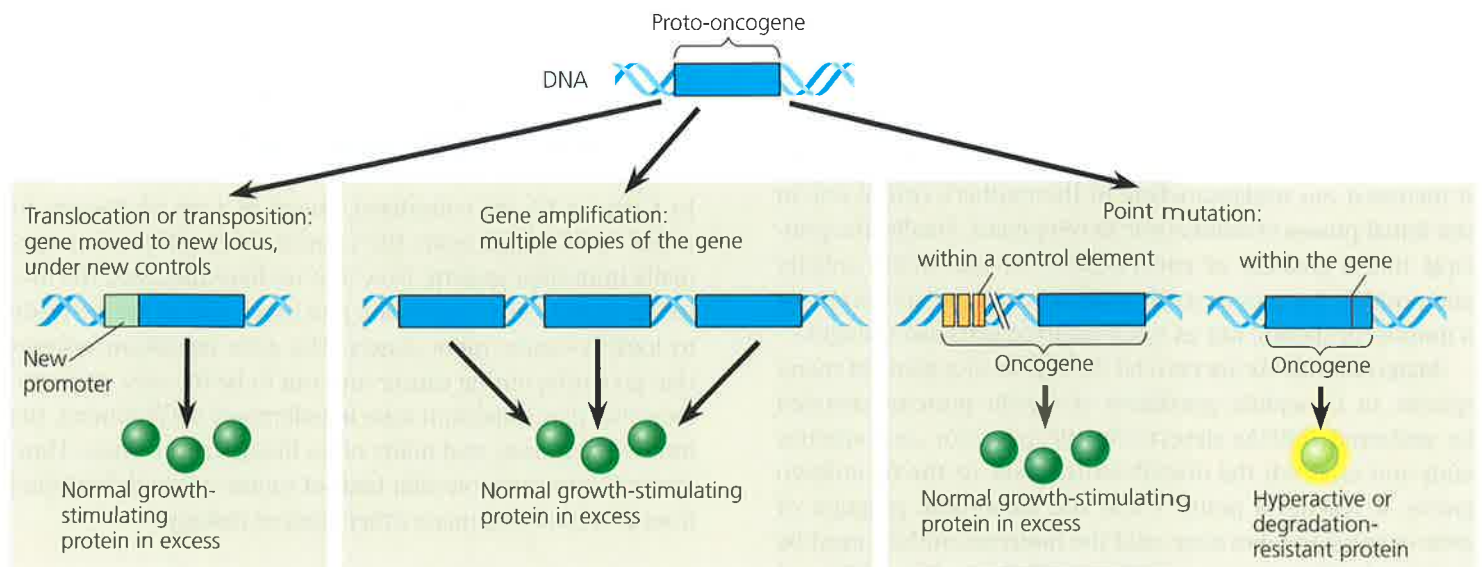
Types of Genes Associated with Cancer

The genes that normally regulate cell growth and division during the cell cycle include genes for growth factors, their receptors, and the intracellular molecules of signaling pathways. (To review the cell cycle, see Chapter 12.) Mutations that alter any of these genes in somatic cells can lead to cancer. The agent of such change can be random spontaneous mutation. However, it is likely that many cancer-causing mutations result from environmental influences, such as chemical carcinogens, X-rays and other high-energy radiation, and some viruses.

Cancer research led to the discovery of cancer-causing genes called **oncogenes** (from the Greek *onco*, tumor) in certain types of viruses (see Chapter 19). Subsequently, close counterparts of viral oncogenes were found in the genomes of humans and other animals. The normal versions of the cellular genes, called **proto-oncogenes**, code for proteins that stimulate normal cell growth and division.

How might a proto-oncogene—a gene that has an essential function in normal cells—become an oncogene, a cancer-causing gene? In general, an oncogene arises from a genetic change that leads to an increase either in the amount of the proto-oncogene's protein product or in the intrinsic activity of each protein molecule. The genetic changes that convert proto-oncogenes to oncogenes fall into three main categories: movement of DNA within the genome, amplification of a proto-oncogene, and point mutations in a control element or in the proto-oncogene itself (**Figure 18.23**, on the next page).

Cancer cells are frequently found to contain chromosomes that have broken and rejoined incorrectly, translocating fragments from one chromosome to another (see Figure 15.14). Now that you have learned how gene expression is regulated, you can understand the possible consequences of such translocations. If a translocated proto-oncogene ends up near an especially active promoter (or other control element), its transcription may increase, making it an oncogene. The second main type of genetic change, amplification, increases the



▲ **Figure 18.23** Genetic changes that can turn proto-oncogenes into oncogenes.

number of copies of the proto-oncogene in the cell through repeated gene duplication (discussed in Chapter 21). The third possibility is a point mutation either (1) in the promoter or an enhancer that controls a proto-oncogene, causing an increase in its expression, or (2) in the coding sequence, changing the gene's product to a protein that is more active or more resistant to degradation than the normal protein. All these mechanisms can lead to abnormal stimulation of the cell cycle and put the cell on the path to malignancy.

Tumor-Suppressor Genes

In addition to genes whose products normally promote cell division, cells contain genes whose normal products *inhibit* cell division. Such genes are called **tumor-suppressor genes** because the proteins they encode help prevent uncontrolled cell growth. Any mutation that decreases the normal activity of a tumor-suppressor protein may contribute to the onset of cancer, in effect stimulating growth through the absence of suppression.

The protein products of tumor-suppressor genes have various functions. Some tumor-suppressor proteins repair damaged DNA, a function that prevents the cell from accumulating cancer-causing mutations. Other tumor-suppressor proteins control the adhesion of cells to each other or to the extracellular matrix; proper cell anchorage is crucial in normal tissues—and is often absent in cancers. Still other tumor-suppressor proteins are components of cell-signaling pathways that inhibit the cell cycle.

Interference with Normal Cell-Signaling Pathways

The proteins encoded by many proto-oncogenes and tumor-suppressor genes are components of cell-signaling pathways. Let's take a closer look at how such proteins function in normal

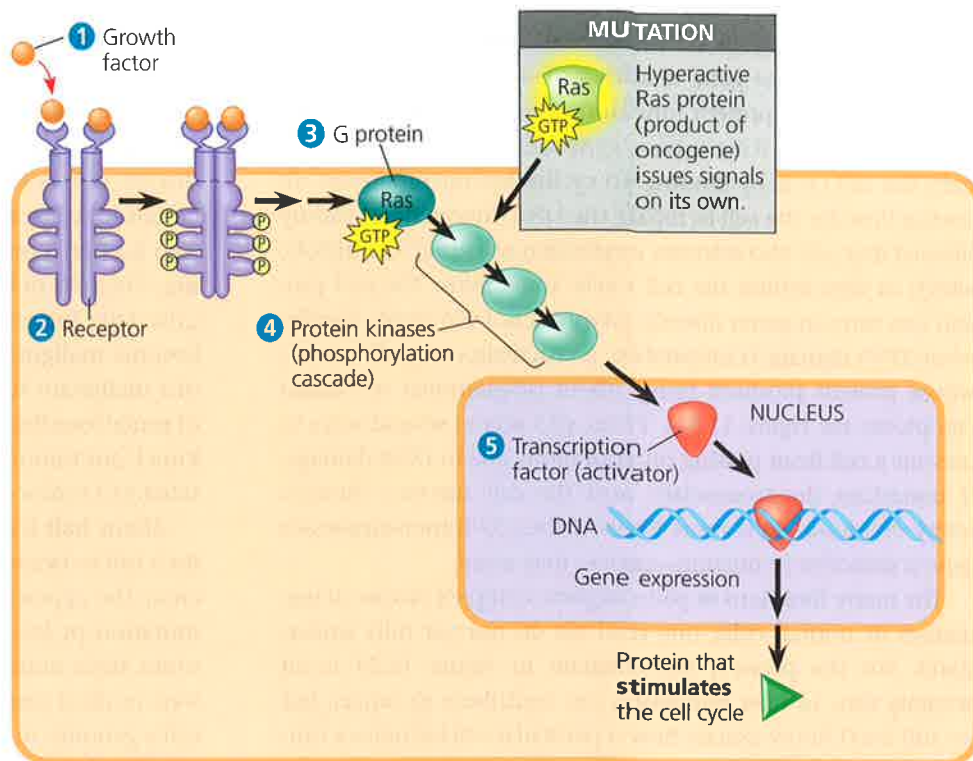
cells and what goes wrong with their function in cancer cells. We will focus on the products of two key genes, the *ras* proto-oncogene and the *p53* tumor-suppressor gene. Mutations in *ras* occur in about 30% of human cancers, and mutations in *p53* in more than 50%.

The Ras protein, encoded by the ***ras* gene** (named for *rat* sarcoma, a connective tissue cancer), is a G protein that relays a signal from a growth factor receptor on the plasma membrane to a cascade of protein kinases (see Figure 11.7). The cellular response at the end of the pathway is the synthesis of a protein that stimulates the cell cycle (**Figure 18.24a**). Normally, such a pathway will not operate unless triggered by the appropriate growth factor. But certain mutations in the *ras* gene can lead to production of a hyperactive Ras protein that triggers the kinase cascade even in the absence of growth factor, resulting in increased cell division. In fact, hyperactive versions or excess amounts of any of the pathway's components can have the same outcome: excessive cell division.

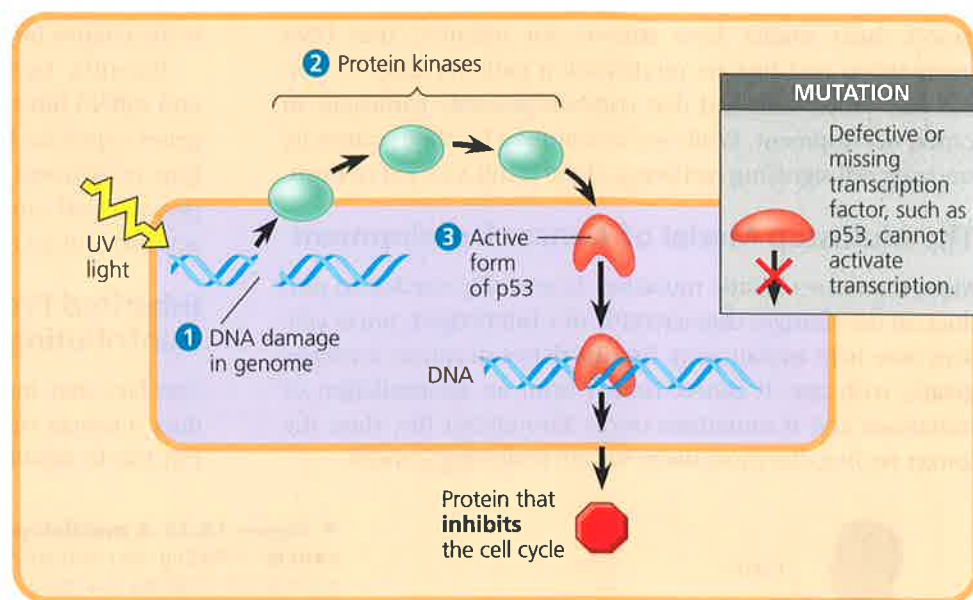
Figure 18.24b shows a pathway in which a signal leads to the synthesis of a protein that suppresses the cell cycle. In this case, the signal is damage to the cell's DNA, perhaps as the result of exposure to ultraviolet light. Operation of this signaling pathway blocks the cell cycle until the damage has been repaired. Otherwise, the damage might contribute to tumor formation by causing mutations or chromosomal abnormalities. Thus, the genes for the components of the pathway act as tumor-suppressor genes. The ***p53* gene**, named for the 53,000-dalton molecular weight of its protein product, is a tumor-suppressor gene. The protein it encodes is a specific transcription factor that promotes the synthesis of cell cycle-inhibiting proteins. That is why a mutation that knocks out the *p53* gene, like a mutation that leads to a hyperactive Ras protein, can lead to excessive cell growth and cancer (**Figure 18.24c**).

(a) Cell cycle–stimulating pathway.

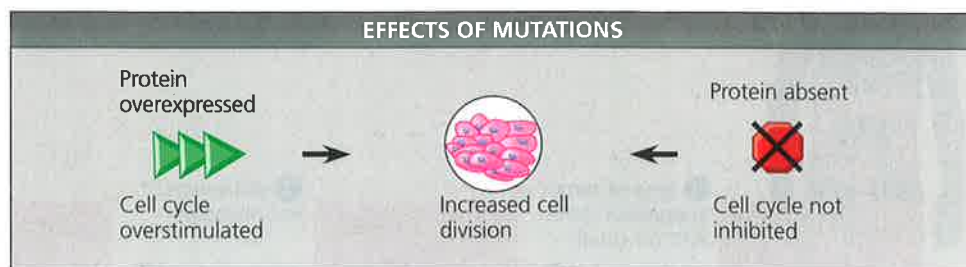
This pathway is triggered by **1** a growth factor that binds to **2** its receptor in the plasma membrane. The signal is relayed to a G protein called Ras. Like all G proteins, Ras is active when GTP is bound to it. Ras passes the signal to **4** a series of protein kinases. The last kinase activates **5** a transcription activator that turns on one or more genes for proteins that stimulate the cell cycle. If a mutation makes Ras or any other pathway component abnormally active, excessive cell division and cancer may result.



(b) Cell cycle–inhibiting pathway. In this pathway, **1** DNA damage is an intracellular signal that is passed via **2** protein kinases and leads to activation of **3** p53. Activated p53 promotes transcription of the gene for a protein that inhibits the cell cycle. The resulting suppression of cell division ensures that the damaged DNA is not replicated. If the DNA damage is irreparable, the p53 signal leads to programmed cell death (apoptosis). Mutations causing deficiencies in any pathway component can contribute to the development of cancer.



(c) Effects of mutations. Increased cell division, possibly leading to cancer, can result if the cell cycle is overstimulated, as in (a), or not inhibited when it normally would be, as in (b).



▲ Figure 18.24 Signaling pathways that regulate cell division. Both stimulatory and inhibitory pathways regulate the cell cycle, commonly by influencing transcription. Cancer can result from aberrations in such pathways, which may be caused by mutations, either spontaneous or environmentally triggered.

? Looking at the pathway in (b), explain whether a cancer-causing mutation in a tumor-suppressor gene, such as p53, is more likely to be a recessive or a dominant mutation.

The *p53* gene has been called the “guardian angel of the genome.” Once the gene is activated—for example, by DNA damage—the *p53* protein functions as an activator for several other genes. Often it activates a gene called *p21*, whose product halts the cell cycle by binding to cyclin-dependent kinases, allowing time for the cell to repair the DNA. Researchers recently showed that *p53* also activates expression of a group of miRNAs, which in turn inhibit the cell cycle. In addition, the *p53* protein can turn on genes directly involved in DNA repair. Finally, when DNA damage is irreparable, *p53* activates “suicide” genes, whose protein products bring about programmed cell death (apoptosis; see Figure 11.21). Thus, *p53* acts in several ways to prevent a cell from passing on mutations due to DNA damage. If mutations do accumulate and the cell survives through many divisions—as is more likely if the *p53* tumor-suppressor gene is defective or missing—cancer may ensue.

The many functions of *p53* suggest a complex picture of regulation in normal cells, one that we do not yet fully understand. For the present, the diagram in Figure 18.24 is an accurate view of how mutations can contribute to cancer, but we still don’t know exactly how a particular cell becomes a cancer cell. As we discover previously unknown aspects of gene regulation, it is informative to study their role in the onset of cancer. Such studies have shown, for instance, that DNA methylation and histone modification patterns differ in normal and cancer cells and that miRNAs probably participate in cancer development. While we’ve learned a lot about cancer by studying cell-signaling pathways, there is still a lot left to learn.

The Multistep Model of Cancer Development

More than one somatic mutation is generally needed to produce all the changes characteristic of a full-fledged cancer cell. This may help explain why the incidence of cancer increases greatly with age. If cancer results from an accumulation of mutations and if mutations occur throughout life, then the longer we live, the more likely we are to develop cancer.

The model of a multistep path to cancer is well supported by studies of one of the best-understood types of human cancer, colorectal cancer. About 135,000 new cases of colorectal cancer are diagnosed each year in the United States, and the disease causes 60,000 deaths each year. Like most cancers, colorectal cancer develops gradually (Figure 18.25). The first sign is often a polyp, a small, benign growth in the colon lining. The cells of the polyp look normal, although they divide unusually frequently. The tumor grows and may eventually become malignant, invading other tissues. The development of a malignant tumor is paralleled by a gradual accumulation of mutations that convert proto-oncogenes to oncogenes and knock out tumor-suppressor genes. A *ras* oncogene and a mutated *p53* tumor-suppressor gene are often involved.

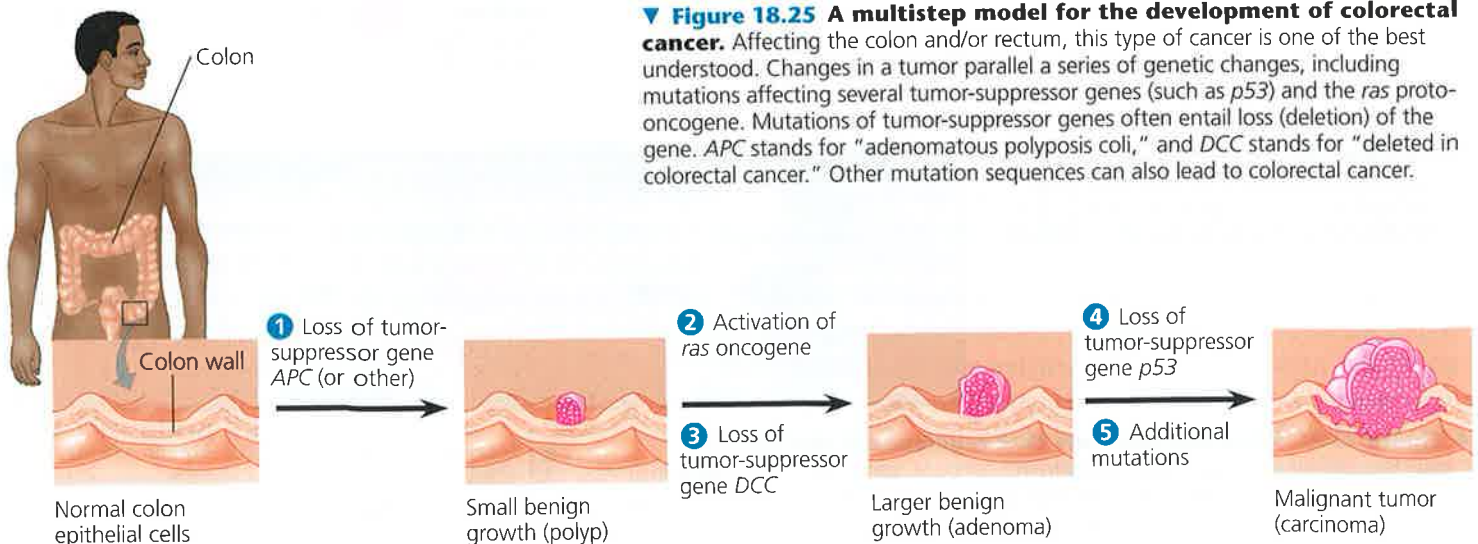
About half a dozen changes must occur at the DNA level for a cell to become fully cancerous. These changes usually include the appearance of at least one active oncogene and the mutation or loss of several tumor-suppressor genes. Furthermore, since mutant tumor-suppressor alleles are usually recessive, in most cases mutations must knock out *both* alleles in a cell’s genome to block tumor suppression. (Most oncogenes, on the other hand, behave as dominant alleles.) The order in which these changes must occur is still under investigation, as is the relative importance of different mutations.

Recently, technical advances in the sequencing of DNA and mRNA have allowed medical researchers to compare the genes expressed by different types of tumors and by the same type in different individuals. These comparisons have led to personalized cancer treatments based on the molecular characteristics of an individual’s tumor (see Figure 12.21.)

Inherited Predisposition and Other Factors Contributing to Cancer

The fact that multiple genetic changes are required to produce a cancer cell helps explain the observation that cancers can run in families. An individual inheriting an oncogene or

▼ Figure 18.25 A multistep model for the development of colorectal cancer. Affecting the colon and/or rectum, this type of cancer is one of the best understood. Changes in a tumor parallel a series of genetic changes, including mutations affecting several tumor-suppressor genes (such as *p53*) and the *ras* proto-oncogene. Mutations of tumor-suppressor genes often entail loss (deletion) of the gene. *APC* stands for “adenomatous polyposis coli,” and *DCC* stands for “deleted in colorectal cancer.” Other mutation sequences can also lead to colorectal cancer.



a mutant allele of a tumor-suppressor gene is one step closer to accumulating the necessary mutations for cancer to develop than is an individual without any such mutations.

Geneticists are devoting much effort to identifying inherited cancer alleles so that predisposition to certain cancers can be detected early in life. About 15% of colorectal cancers, for example, involve inherited mutations. Many of these affect the tumor-suppressor gene called *adenomatous polyposis coli*, or *APC* (see Figure 18.25). This gene has multiple functions in the cell, including regulation of cell migration and adhesion. Even in patients with no family history of the disease, the *APC* gene is mutated in 60% of colorectal cancers. In these individuals, new mutations must occur in both *APC* alleles before the gene's function is lost. Since only 15% of colorectal cancers are associated with known inherited mutations, researchers continue in their efforts to identify “markers” that could predict the risk of developing this type of cancer.

There is evidence of a strong inherited predisposition in 5–10% of patients with breast cancer. This is the second most common type of cancer in the United States, striking over 180,000 women (and some men) annually and killing 40,000 each year. In 1990, after 16 years of research, geneticist Mary-Claire King convincingly demonstrated that mutations in one gene—*BRCA1*—were associated with increased susceptibility to breast cancer, a finding that flew in the face of medical opinion at the time. (*BRCA* stands for *breast cancer*.) Mutations in that gene or the related *BRCA2* gene are found in at least half of inherited breast cancers, and tests using DNA sequencing can detect these mutations (Figure 18.26). A woman who inherits one mutant *BRCA1* allele has a 60% probability of developing breast cancer before the age of 50, compared with only a 2% probability for an individual homozygous for the normal allele. Both *BRCA1* and *BRCA2* are considered tumor-suppressor genes because their wild-type alleles



▲ **Figure 18.26 Testing for mutations in *BRCA1* and *BRCA2*.** Genetic testing for mutations that increase the risk of breast cancer is available for individuals with a family history of breast cancer. New “high-throughput” sequencing techniques can sequence many DNA samples at once, as shown here.

protect against breast cancer and their mutant alleles are recessive. Apparently, the *BRCA1* and *BRCA2* proteins both function in the cell's DNA damage repair pathway. More is known about *BRCA2*, which, in association with another protein, helps repair breaks that occur in both strands of DNA; it is crucial for maintaining undamaged DNA in a cell's nucleus.

Because DNA breakage can contribute to cancer, it makes sense that the risk of cancer can be lowered by minimizing exposure to DNA-damaging agents, such as the ultraviolet radiation in sunlight and chemicals found in cigarette smoke. Novel methods for early diagnosis and treatment of specific cancers are being developed that rely on new techniques for analyzing, and perhaps interfering with, gene expression in tumors. Ultimately, such approaches may lower the death rate from cancer.

The study of genes associated with cancer, inherited or not, increases our basic understanding of how disruption of normal gene regulation can result in this disease. In addition to the mutations and other genetic alterations described in this section, a number of *tumor viruses* can cause cancer in various animals, including humans. In fact, one of the earliest breakthroughs in understanding cancer came in 1911, when Peyton Rous, an American pathologist, discovered a virus that causes cancer in chickens. The Epstein-Barr virus, which causes infectious mononucleosis, has been linked to several types of cancer in humans, notably Burkitt's lymphoma. Papillomaviruses are associated with cancer of the cervix, and a virus called HTLV-1 causes a type of adult leukemia. Worldwide, viruses seem to play a role in about 15% of the cases of human cancer.

Viruses may at first seem very different from mutations as a cause of cancer. However, we now know that viruses can interfere with gene regulation in several ways if they integrate their genetic material into the DNA of a cell. Viral integration may donate an oncogene to the cell, disrupt a tumor-suppressor gene, or convert a proto-oncogene to an oncogene. In addition, some viruses produce proteins that inactivate p53 and other tumor-suppressor proteins, making the cell more prone to becoming cancerous. Viruses are powerful biological agents, and you'll learn more about their function in Chapter 19.

CONCEPT CHECK 18.5

1. **MAKE CONNECTIONS** The p53 protein can activate genes involved in apoptosis, or programmed cell death. Review Concept 11.5 (pp. 269–271) and discuss how mutations in genes coding for proteins that function in apoptosis could contribute to cancer.
2. Under what circumstances is cancer considered to have a hereditary component?
3. **WHAT IF?** Explain how the types of mutations that lead to cancer are different for a proto-oncogene and a tumor-suppressor gene, in terms of the effect of the mutation on the activity of the gene product.

For suggested answers, see Appendix A.

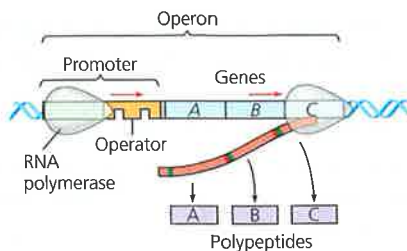
18 CHAPTER REVIEW

SUMMARY OF KEY CONCEPTS

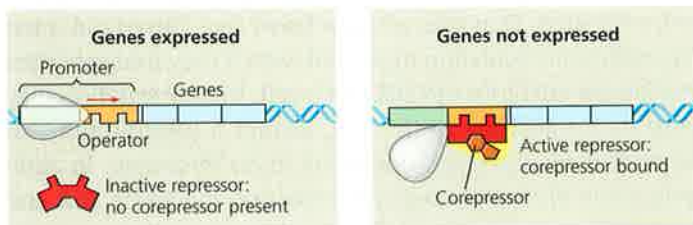
CONCEPT 18.1

Bacteria often respond to environmental change by regulating transcription (pp. 397–402)

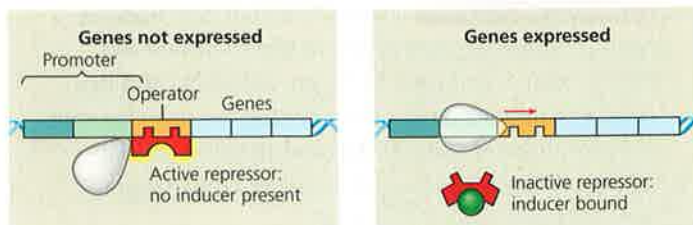
- Cells control metabolism by regulating enzyme activity or the expression of genes coding for enzymes. In bacteria, genes are often clustered into **operons**, with one promoter serving several adjacent genes. An **operator** site on the DNA switches the operon on or off, resulting in coordinate regulation of the genes.



- Both repressible and inducible operons are examples of negative gene regulation. In either type of operon, binding of a specific **repressor** protein to the operator shuts off transcription. (The repressor is encoded by a separate **regulatory gene**.) In a repressible operon, the repressor is active when bound to a **corepressor**, usually the end product of an anabolic pathway.



In an inducible operon, binding of an **inducer** to an innately active repressor inactivates the repressor and turns on transcription. Inducible enzymes usually function in catabolic pathways.

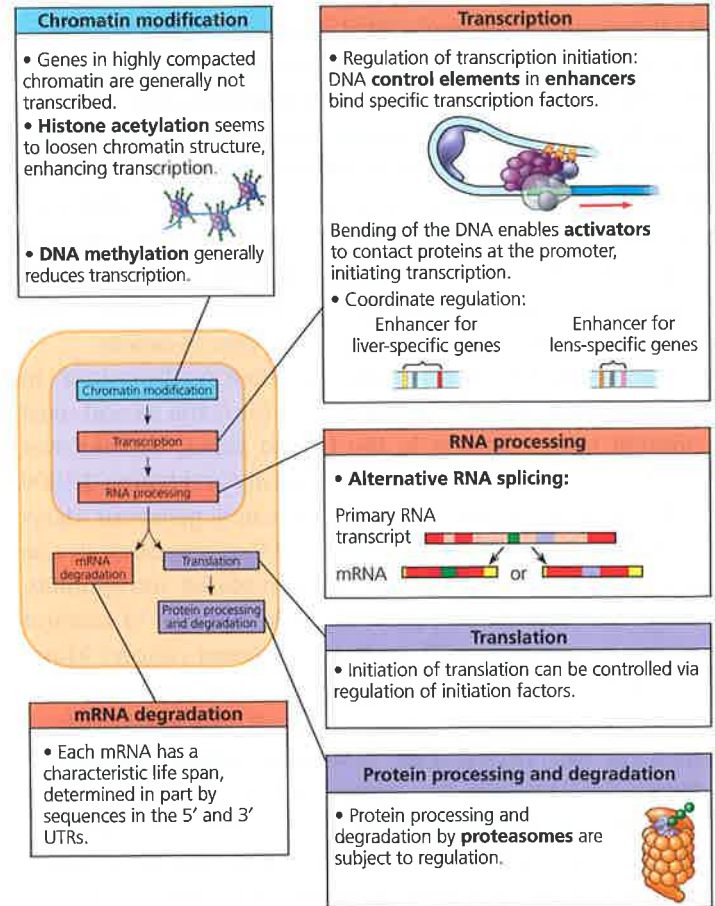


- Some operons are also subject to positive gene regulation via a stimulatory **activator** protein, such as catabolite activator protein (CAP), which, when activated by **cyclic AMP**, binds to a site within the promoter and stimulates transcription.

? Compare and contrast the roles of the corepressor and the inducer in negative regulation of an operon.

CONCEPT 18.2

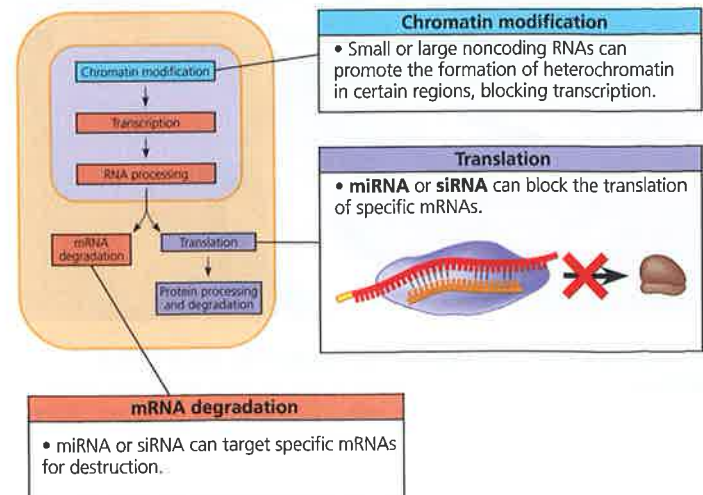
Eukaryotic gene expression is regulated at many stages (pp. 402–410)



? Describe what must happen for a cell-type-specific gene to be transcribed in a cell of that type.

CONCEPT 18.3

Noncoding RNAs play multiple roles in controlling gene expression (pp. 410–412)



? Why are miRNAs called noncoding RNAs? Explain how they participate in gene regulation.

CONCEPT 18.4

A program of differential gene expression leads to the different cell types in a multicellular organism (pp. 412–419)

- Embryonic cells undergo **differentiation**, becoming specialized in structure and function. **Morphogenesis** encompasses the processes that give shape to the organism and its various parts. Cells differ in structure and function not because they contain different genes but because they express different portions of a common genome.
- Cytoplasmic determinants** in the unfertilized egg regulate the expression of genes in the zygote that affect the developmental fate of embryonic cells. In the process called **induction**, signaling molecules from embryonic cells cause transcriptional changes in nearby target cells.
- Differentiation is heralded by the appearance of tissue-specific proteins, which enable differentiated cells to carry out their specialized roles.
- In animals, **pattern formation**, the development of a spatial organization of tissues and organs, begins in the early embryo. **Positional information**, the molecular cues that control pattern formation, tells a cell its location relative to the body's axes and to other cells. In *Drosophila*, gradients of **morphogens** encoded by **maternal effect genes** determine the body axes. For example, the gradient of **Bicoid** protein determines the anterior-posterior axis.

? Describe the two main processes that cause embryonic cells to head down different pathways to their final fates.

CONCEPT 18.5

Cancer results from genetic changes that affect cell cycle control (pp. 419–423)

- The products of **proto-oncogenes** and **tumor-suppressor genes** control cell division. A DNA change that makes a proto-oncogene excessively active converts it to an **oncogene**, which may promote excessive cell division and cancer. A tumor-suppressor gene encodes a protein that inhibits abnormal cell division. A mutation in such a gene that reduces the activity of its protein product may also lead to excessive cell division and possibly to cancer.
- Many proto-oncogenes and tumor-suppressor genes encode components of growth-stimulating and growth-inhibiting signaling pathways, respectively, and mutations in these genes can interfere with normal cell-signaling pathways. A hyperactive version of a protein in a stimulatory pathway, such as **Ras** (a G protein), functions as an oncogene protein. A defective version of a protein in an inhibitory pathway, such as **p53** (a transcription activator), fails to function as a tumor suppressor.
- In the multistep model of cancer development, normal cells are converted to cancer cells by the accumulation of mutations affecting proto-oncogenes and tumor-suppressor genes. Technical advances in DNA and mRNA sequencing are enabling cancer treatments that are more individually based.
- Individuals who inherit a mutant oncogene or tumor-suppressor allele have a predisposition to develop a particular cancer. Certain viruses promote cancer by integration of viral DNA into a cell's genome.

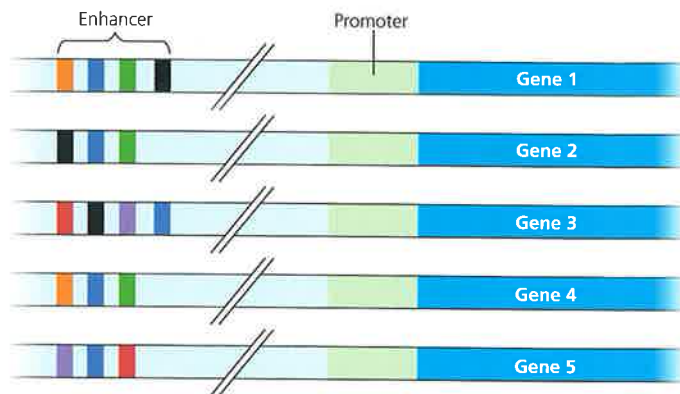
? Compare the usual functions of proteins encoded by proto-oncogenes with the functions of proteins encoded by tumor-suppressor genes.

TEST YOUR UNDERSTANDING



Multiple-choice Self-Quiz questions #1–10 can be found in the Study Area at www.masteringbiology.com.

11. **DRAW IT** The diagram below shows five genes, including their enhancers, from the genome of a certain species. Imagine that orange, blue, green, black, red, and purple activator proteins exist that can bind to the appropriately color-coded control elements in the enhancers of these genes.



- Draw an X above enhancer elements (of all the genes) that would have activators bound in a cell in which only gene 5 is transcribed. Which colored activators would be present?
 - Draw a dot above all enhancer elements that would have activators bound in a cell in which the green, blue, and orange activators are present. Which gene(s) would be transcribed?
 - Imagine that genes 1, 2, and 4 code for nerve-specific proteins, and genes 3 and 5 are skin specific. Which activators would have to be present in each cell type to ensure transcription of the appropriate genes?
12. **EVOLUTION CONNECTION**
DNA sequences can act as “tape measures of evolution” (see Chapter 5). Scientists analyzing the human genome sequence were surprised to find that some of the regions of the human genome that are most highly conserved (similar to comparable regions in other species) don’t code for proteins. Propose a possible explanation for this observation.
13. **SCIENTIFIC INQUIRY**
Prostate cells usually require testosterone and other androgens to survive. But some prostate cancer cells thrive despite treatments that eliminate androgens. One hypothesis is that estrogen, often considered a female hormone, may be activating genes normally controlled by an androgen in these cancer cells. Describe one or more experiments to test this hypothesis. (See Figure 11.9, p. 260, to review the action of these steroid hormones.)
14. **SCIENCE, TECHNOLOGY, AND SOCIETY**
Trace amounts of dioxin were present in Agent Orange, a defoliant sprayed on vegetation during the Vietnam War. Animal tests suggest that dioxin can cause birth defects, cancer, liver and thymus damage, and immune system suppression, sometimes leading to death. But the animal tests are equivocal; a hamster is not affected by a dose that can kill a guinea pig. Dioxin acts somewhat like a steroid hormone, entering a cell and binding to a receptor protein that then attaches to the cell's DNA. How might this mechanism help explain the variety of dioxin's effects on different body systems and in different animals? How might you determine whether a type

of illness is related to dioxin exposure? How might you determine whether a particular individual became ill as a result of exposure to dioxin? Which would be more difficult to demonstrate? Why?

15. **WRITE ABOUT A THEME**

Feedback Regulation In a short essay (100–150 words), discuss how the processes shown in Figure 18.24a and b are examples of feedback mechanisms regulating biological systems.

For selected answers, see Appendix A.



www.masteringbiology.com

1. MasteringBiology® Assignments

Tutorials Regulation of Gene Expression in Bacteria • Regulation of Gene Expression in Eukaryotes • Pattern Formation

Activities The *lac* Operon • The *lac* Operon in *E. coli* • Transcription Initiation in Eukaryotes • Overview: Control of Gene Expression • Control of Transcription • Review: Control of Gene Expression • Early Pattern Formation in *Drosophila* • Role of *bicoid* Gene in *Drosophila* Development

Questions Student Misconceptions • Reading Quiz • Multiple Choice • End-of-Chapter

2. eText

Read your book online, search, take notes, highlight text, and more.

3. The Study Area

Practice Tests • Cumulative Test • **BioFlix** 3-D Animations • MP3 Tutor Sessions • Videos • Activities • Investigations • Lab Media • Audio Glossary • Word Study Tools • Art