



## CHAPTER EIGHT

# 8

### Control of Gene Expression

An organism's DNA encodes all of the RNA and protein molecules that are needed to make its cells. Yet a complete description of the DNA sequence of an organism—be it the few million nucleotides of a bacterium or the few billion nucleotides in each human cell—does not enable us to reconstruct the organism any more than a list of all the English words in a dictionary enables us to reconstruct a play by Shakespeare. We need to know how the elements in the DNA sequence or the words on a list work together to make the masterpiece.

In cell biology, the question comes down to *gene expression*. Even the simplest single-celled bacterium can use its genes selectively—for example, switching genes on and off to make the enzymes needed to digest whatever food sources are available. And, in multicellular plants and animals, gene expression is under even more elaborate control. Over the course of embryonic development, a fertilized egg cell gives rise to many cell types that differ dramatically in both structure and function. The differences between a mammalian neuron and a lymphocyte, for example, are so extreme that it is difficult to imagine that the two cells contain the same DNA (Figure 8-1). For this reason, and because cells in an adult organism rarely lose their distinctive characteristics, biologists originally suspected that genes might be selectively lost when a cell becomes specialized. We now know, however, that nearly all the cells of a multicellular organism contain the same genome. Cell differentiation is instead achieved by changes in gene expression.

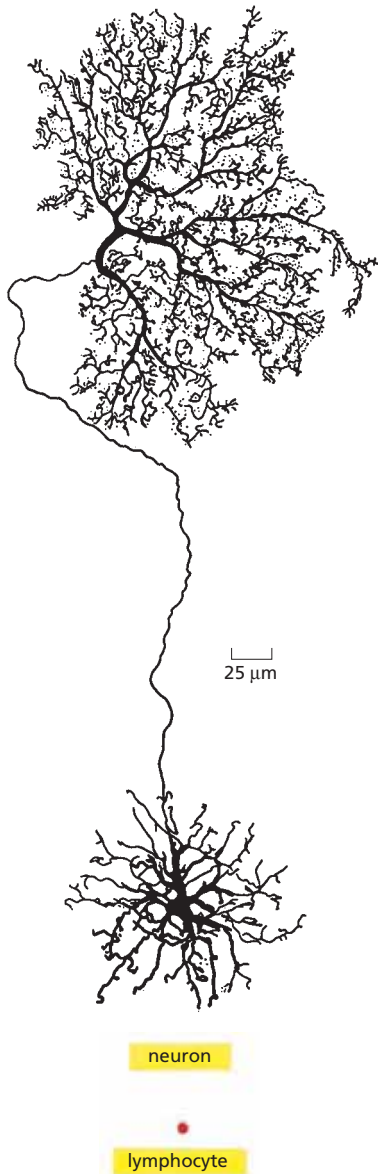
Hundreds of different cell types carry out a range of specialized functions that depend upon genes that are only switched on in that cell type: for

AN OVERVIEW OF GENE  
EXPRESSION

HOW TRANSCRIPTIONAL  
SWITCHES WORK

THE MOLECULAR  
MECHANISMS THAT CREATE  
SPECIALIZED CELL TYPES

POST-TRANSCRIPTIONAL  
CONTROLS



**Figure 8–1 A neuron and a lymphocyte share the same genome.** The long branches of this neuron from the retina enable it to receive electrical signals from many cells and carry them to many neighboring cells. The lymphocyte, a white blood cell involved in the immune response to infection (drawn to scale), moves freely through the body. Both of these mammalian cells contain the same genome, but they express different RNAs and proteins. (Neuron from B.B. Boycott in *Essays on the Nervous System* [R. Bellairs and E.G. Gray, eds.]. Oxford, U.K.: Clarendon Press, 1974. © Oxford University Press.)

example, the  $\beta$  cells of the pancreas make the protein hormone insulin, while the  $\alpha$  cells of the pancreas make the hormone glucagon; the lymphocytes of the immune system are the only cells in the body to make antibodies, while developing red blood cells are the only cells that make the oxygen-transport protein hemoglobin. The differences between a neuron, a lymphocyte, a liver cell, and a red blood cell depend upon the precise control of gene expression. In each case the cell is using only some of the genes in its total repertoire.

In this chapter, we shall discuss the main ways in which gene expression is controlled in bacterial and eucaryotic cells. Although some mechanisms of control apply to both sorts of cells, eucaryotic cells, through their more complex chromosomal structure, have ways of controlling gene expression that are not available to bacteria.

## AN OVERVIEW OF GENE EXPRESSION

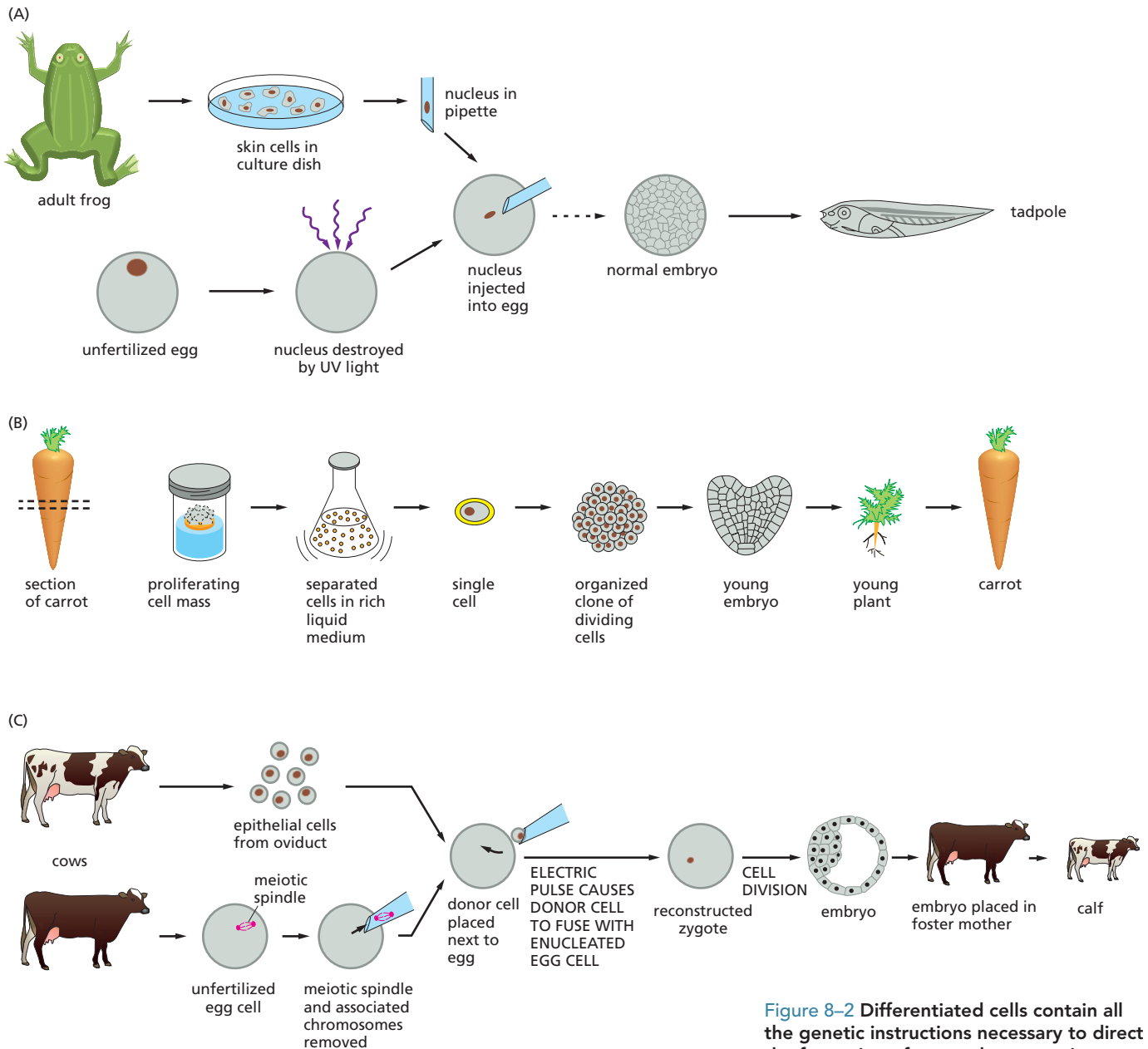
How does an individual cell specify which of its many thousands of genes to express? This decision is an especially important problem for multicellular organisms because, as the animal develops, cell types such as muscle, nerve, and blood cells become different from one another, eventually leading to the wide variety of cell types seen in the adult. Such **differentiation** arises because cells make and accumulate different sets of RNA and protein molecules: that is, they express different genes.

### The Different Cell Types of a Multicellular Organism Contain the Same DNA

As discussed above, cells have the ability to change which genes they express without altering the nucleotide sequence of their DNA. But how do we know this? If DNA were altered irreversibly during development, the chromosomes of a differentiated cell would be incapable of guiding the development of the whole organism. To test this idea, a nucleus from a skin cell of an adult frog was injected into a frog egg whose own nucleus had been removed. In at least some cases the egg developed normally into a tadpole, indicating that the transplanted skin cell nucleus cannot have lost any critical DNA sequences (**Figure 8–2**). Such nuclear transplantation experiments have also been carried out successfully using differentiated cells taken from adult mammals, including sheep, cows, pigs, goats, and mice. And in plants, individual cells removed from a carrot, for example, can be shown to regenerate an entire adult carrot plant. These experiments all show that the DNA in specialized cell types still contains the entire set of instructions needed to form a whole organism. The cells of an organism therefore differ not because they contain different genes, but because they express them differently.

### Different Cell Types Produce Different Sets of Proteins

The extent of the differences in gene expression between different cell types may be roughly gauged by comparing the protein composition of cells in liver, heart, brain, and so on using the technique of two-dimensional gel electrophoresis (see Panel 4–6, p. 167). Experiments of this kind reveal that many proteins are common to all the cells of a multicellular organism. These *housekeeping* proteins include the structural proteins of chromosomes, RNA polymerases, DNA repair enzymes, ribosomal proteins, enzymes involved in glycolysis and other basic metabolic processes, and many of the proteins that form the cytoskeleton. Each different cell type also produces specialized proteins that are responsible for the cell's distinctive properties. In mammals, for example, hemoglobin is



**Figure 8–2 Differentiated cells contain all the genetic instructions necessary to direct the formation of a complete organism.**

(A) The nucleus of a skin cell from an adult frog transplanted into an egg whose nucleus has been removed can give rise to an entire tadpole. The broken arrow indicates that to give the transplanted genome time to adjust to an embryonic environment, a further transfer step is required in which one of the nuclei is taken from the early embryo that begins to develop and is put back into a second enucleated egg. (B) In many types of plants, differentiated cells retain the ability to “dedifferentiate,” so that a single cell can form a clone of progeny cells that later give rise to an entire plant. (C) A differentiated cell from an adult cow introduced into an enucleated egg from a different cow can give rise to a calf. Different calves produced from the same differentiated cell donor are genetically identical and are therefore clones of one another. (A, modified from J.B. Gurdon, *Sci. Am.* 219(6):24–35, 1968. With permission from the Estate of Bunji Tagawa.)

made in reticulocytes, the cells that develop into red blood cells, but it cannot be detected in any other cell type.

Many proteins in a cell are produced in such small numbers that they cannot be detected by the technique of gel electrophoresis. A more sensitive technique, called mass spectrometry (see Figure 4–45) can be used to detect even rare proteins, and can also provide information about whether the proteins are covalently modified (for example by phosphorylation). Gene expression can also be studied by monitoring the mRNAs that encode proteins, rather than the proteins themselves. Estimates of the number of different mRNA sequences in human cells suggest that, at any one time, a typical differentiated human cell expresses perhaps 5000–15,000 genes from a repertoire of about 25,000. It is the expression of a different collection of genes in each cell type that causes the large variations seen in the size, shape, behavior, and function of differentiated cells.

## A Cell Can Change the Expression of Its Genes in Response to External Signals

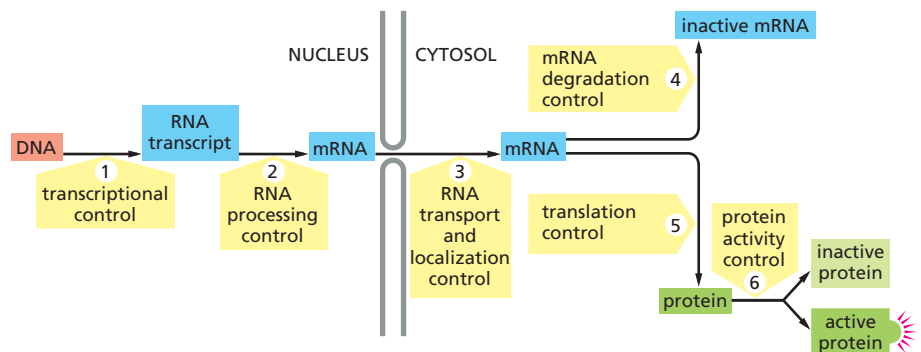
Most of the specialized cells in a multicellular organism are capable of altering their patterns of gene expression in response to extracellular cues. For example, if a liver cell is exposed to a glucocorticoid hormone (a type of steroid), the production of several specific proteins is dramatically increased. Released in the body during periods of starvation or intense exercise, glucocorticoids signal the liver to increase the production of glucose from amino acids and other small molecules. The set of proteins whose production is induced by glucocorticoids includes enzymes such as tyrosine aminotransferase, which helps to convert tyrosine to glucose. When the hormone is no longer present, the production of these proteins drops to its normal level.

Other cell types respond to glucocorticoids differently. In fat cells, for example, the production of tyrosine aminotransferase is reduced, while some other cell types do not respond to glucocorticoids at all. These examples illustrate a general feature of cell specialization: different cell types often respond in different ways to the same extracellular signal. Underlying such adjustments are features of the gene expression pattern that do not change and give each cell type its permanently distinctive character.

## Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein

If differences among the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? As we saw in the last chapter, there are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed, (2) controlling how an RNA transcript is spliced or otherwise processed, (3) selecting which mRNAs are exported from the nucleus to the cytosol, (4) selectively degrading certain mRNA molecules, (5) selecting which mRNAs are translated by ribosomes, or (6) selectively activating or inactivating proteins after they have been made (Figure 8–3).

Gene expression can be regulated at each of these steps, and throughout this chapter we will describe some of the key control points along their pathway from DNA to protein. For most genes, however, the control of transcription (step number 1 in Figure 8–3) is paramount. This makes sense because only transcriptional control can ensure that no unnecessary intermediates are synthesized. So it is the regulation of transcription—and the DNA and protein components that determine which genes a cell transcribes into RNA—that we address first.



**Figure 8–3 Eucaryotic gene expression can be controlled at several different steps.** Examples of regulation at each of the steps are known, although for most genes the main site of control is step 1: transcription of a DNA sequence into RNA.

## HOW TRANSCRIPTIONAL SWITCHES WORK

Only 50 years ago the idea that genes could be switched on and off was revolutionary. This concept was a major advance, and it came originally from the study of how *E. coli* bacteria adapt to changes in the composition of their growth medium. Many of the same principles apply to eucaryotic cells. However, the enormous complexity of gene regulation in higher organisms, combined with the packaging of their DNA into chromatin, creates special challenges and some novel opportunities for control—as we shall see. We begin with a discussion of the *transcription regulators*, proteins that control gene expression at the level of transcription.

### Transcription Is Controlled by Proteins Binding to Regulatory DNA Sequences

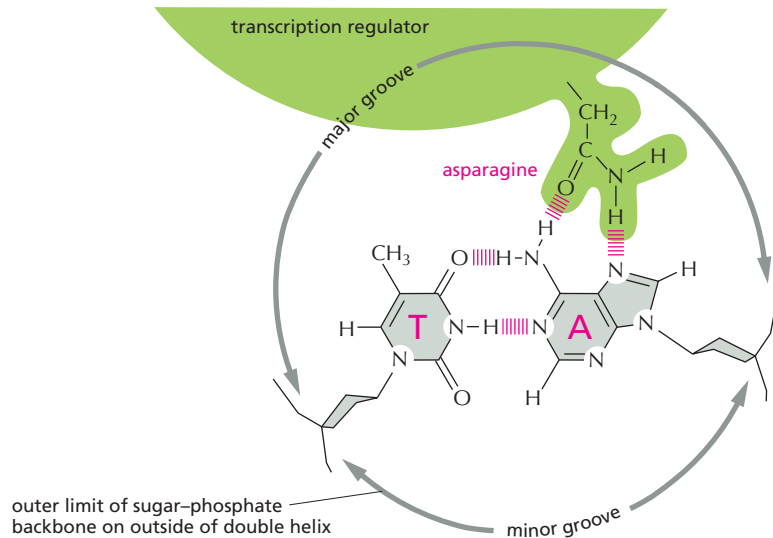
Control of transcription is usually exerted at the step at which the process is initiated. In Chapter 7, we saw that the promoter region of a gene attracts the enzyme RNA polymerase and correctly orients the enzyme to begin its task of making an RNA copy of the gene. The promoters of both bacterial and eucaryotic genes include an *initiation site*, where transcription actually begins, and a sequence of approximately 50 nucleotides that extends upstream from the initiation site (if one likens the direction of transcription to the flow of a river). This region contains sites that are required for the RNA polymerase to bind to the promoter. In addition to the promoter, nearly all genes, whether bacterial or eucaryotic, have **regulatory DNA sequences** that are used to switch the gene on or off.

Some regulatory DNA sequences are as short as 10 nucleotide pairs and act as simple gene switches that respond to a single signal. Such simple switches predominate in bacteria. Other regulatory DNA sequences, especially those in eucaryotes, are very long (sometimes more than 10,000 nucleotide pairs) and act as molecular microprocessors, integrating information from a variety of signals into a command that dictates how often transcription is initiated.

Regulatory DNA sequences do not work by themselves. To have any effect, these sequences must be recognized by proteins called **transcription regulators**, which bind to the DNA. It is the combination of a DNA sequence and its associated protein molecules that acts as the switch to control transcription. The simplest bacterium codes for several hundred transcription regulators, each of which recognizes a different DNA sequence and thereby regulates a distinct set of genes. Humans make many more—several thousand—signifying the importance and complexity of this form of gene regulation in producing a complex organism.

Proteins that recognize a specific DNA sequence do so because the surface of the protein fits tightly against the special surface features of the double helix in that region. These features will vary depending on the nucleotide sequence, and thus different proteins will recognize different nucleotide sequences. In most cases, the protein inserts into the major groove of the DNA helix (see Figure 5–7) and makes a series of molecular contacts with the base pairs. The protein forms hydrogen bonds, ionic bonds, and hydrophobic interactions with the edges of the bases, usually without disrupting the hydrogen bonds that hold the base pairs together (Figure 8–4). Although each individual contact is weak, the 20 or so contacts that are typically formed at the protein–DNA interface combine to ensure that the interaction is both highly specific and very strong; indeed, protein–DNA interactions are among the tightest and most specific molecular interactions known in biology.

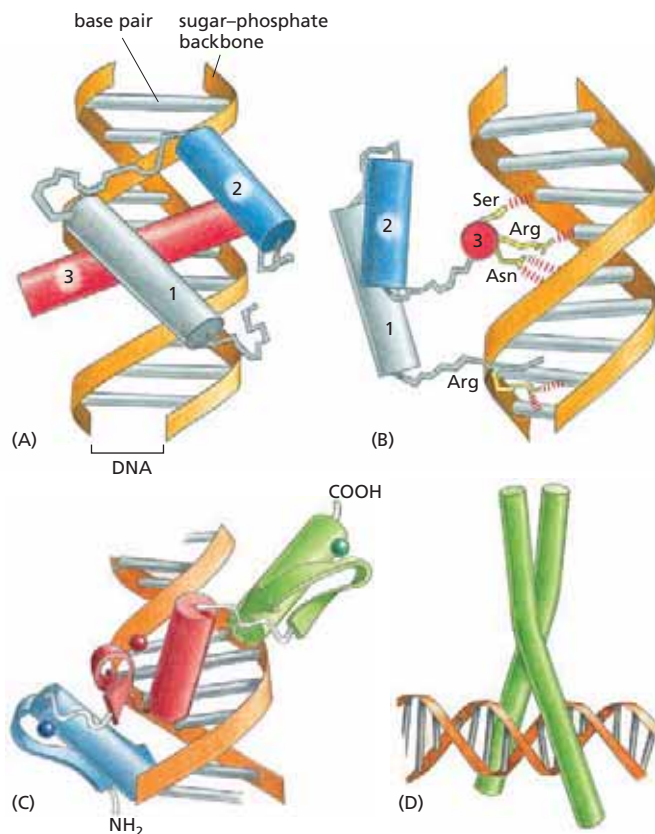
**Figure 8–4 A transcription regulator binds to the major groove of a DNA helix.** Only a single contact between the protein and one base pair in DNA is shown. Typically, the protein–DNA interface would consist of 10–20 such contacts, each involving a different amino acid and each contributing to the strength of the protein–DNA interaction.



Although each example of protein–DNA recognition is unique in detail, many of the proteins responsible for gene regulation recognize DNA through one of several structural motifs. These fit into the major groove of the DNA double helix and form tight associations with a short stretch of DNA base pairs. The *DNA-binding motifs* shown in **Figure 8–5**—the homeodomain, the zinc finger, and the leucine zipper—are found in transcription regulators that control the expression of thousands of different genes in virtually all eucaryotic organisms. Frequently, DNA-binding proteins bind in pairs (dimers) to the DNA helix. Dimerization roughly doubles the area of contact with the DNA, thereby greatly increasing the strength and specificity of the protein–DNA interaction. Because two different proteins can pair in different combinations, dimerization also makes it possible for many different DNA sequences to be recognized by a limited number of proteins.

**Figure 8–5 Transcription regulators contain a variety of DNA-binding motifs.** (A and B) Front and side views of the *homeodomain*—a structural motif found in many eucaryotic DNA-binding proteins (**Movie 8.1**). It consists of three consecutive  $\alpha$  helices, which are shown as *cylinders* in this figure. Most of the contacts with the DNA bases are made by helix 3 (which is seen end-on in B). The asparagine (Asn) in this helix contacts an adenine in the manner shown in **Figure 8–4**. (C) The *zinc finger* motif is built from an  $\alpha$  helix and a  $\beta$  sheet (the latter shown as a *twisted arrow*) held together by a molecule of zinc (indicated by the *colored spheres*). Zinc fingers are often found in clusters covalently joined together to allow the  $\alpha$  helix of each finger to contact the DNA bases in the major groove (**Movie 8.2**). The illustration here shows a cluster of three zinc fingers. (D) A *leucine zipper* motif. This DNA-binding motif is formed by two  $\alpha$  helices, each contributed by a different protein molecule. Leucine zipper proteins thus bind to DNA as dimers, gripping the double helix like a clothespin on a clothesline (**Movie 8.3**).

Each of these motifs makes many contacts with DNA. For simplicity, only the hydrogen-bond contacts are shown in (B), and none of the individual protein–DNA contacts are shown in (C) and (D).



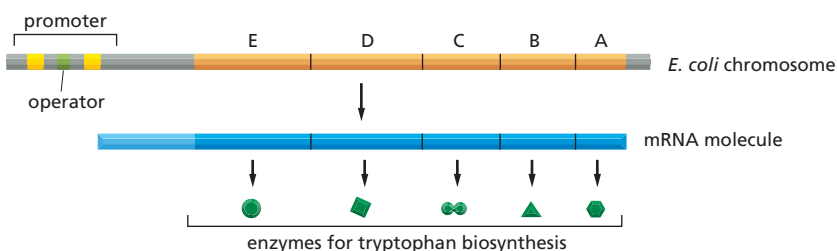
## Transcription Switches Allow Cells to Respond to Changes in the Environment

The simplest and most completely understood examples of gene regulation occur in bacteria and in the viruses that infect them. The genome of the bacterium *E. coli* consists of a single circular DNA molecule of about  $4.6 \times 10^6$  nucleotide pairs. This DNA encodes approximately 4300 proteins, although only a fraction of these are made at any one time. Bacteria regulate the expression of many of their genes according to the food sources that are available in the environment. For example, in *E. coli*, five genes code for enzymes that manufacture the amino acid tryptophan. These genes are arranged in a cluster on the chromosome and are transcribed from a single promoter as one long mRNA molecule from which the five proteins are translated (Figure 8–6). When tryptophan is present in the surroundings and enters the bacterial cell, these enzymes are no longer needed and their production is shut off. This situation arises, for example, when the bacterium is in the gut of a mammal that has just eaten a meal rich in protein. These five coordinately expressed genes are part of an *operon*—a set of genes that are transcribed into a single mRNA. Operons are common in bacteria but are not found in eucaryotes, where genes are transcribed and regulated individually (see Figure 7–36).

We now understand in considerable detail how the tryptophan operon functions. Within the promoter is a short DNA sequence (15 nucleotides in length) that is recognized by a transcription regulator. When this protein binds to this nucleotide sequence, termed the *operator*, it blocks access of RNA polymerase to the promoter; this prevents transcription of the operon and production of the tryptophan-producing enzymes. The transcription regulator is known as the *tryptophan repressor*, and it is controlled in an ingenious way: the repressor can bind to DNA only if it has also bound several molecules of the amino acid tryptophan (Figure 8–7).

The tryptophan repressor is an allosteric protein (see Figure 4–37): the binding of tryptophan causes a subtle change in its three-dimensional structure so that the protein can bind to the operator sequence. When the concentration of free tryptophan in the cell drops, the repressor no longer binds tryptophan and thus no longer binds to DNA, and the tryptophan operon is transcribed. The repressor is thus a simple device that switches production of a set of biosynthetic enzymes on and off according to the availability of the end product of the pathway that the enzymes catalyze.

The bacterium can respond very rapidly to the rise in tryptophan concentration because the tryptophan repressor protein itself is always present in the cell. The gene that encodes it is continuously transcribed at a low level, so that a small amount of the repressor protein is always being made. Such unregulated gene expression is known as *constitutive* gene expression.

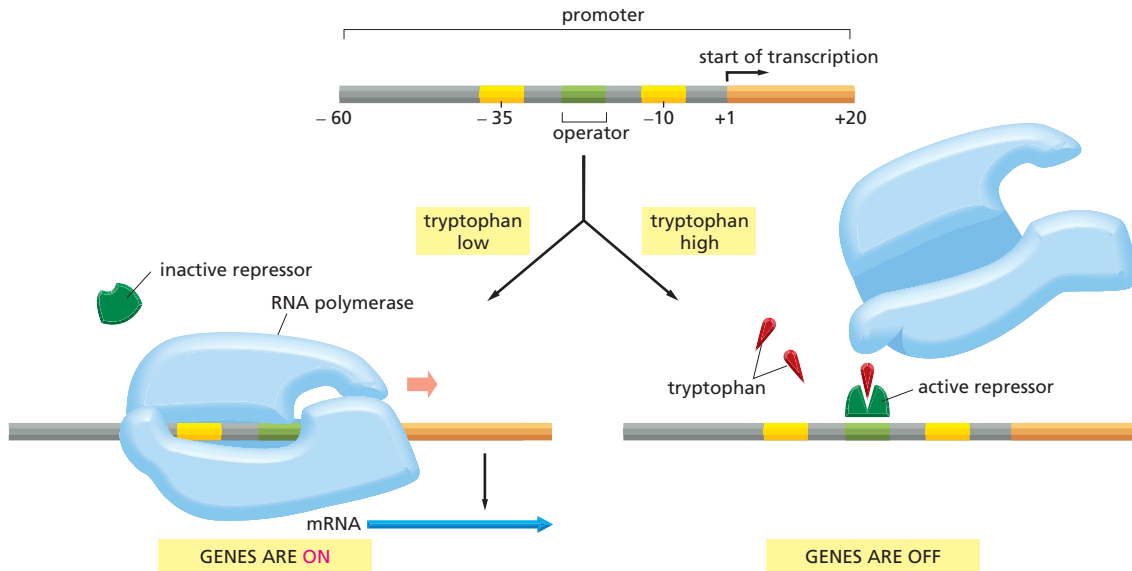


### QUESTION 8-1

Bacterial cells can take up the amino acid tryptophan (Trp) from their surroundings, or if there is an insufficient external supply they can synthesize tryptophan from other small molecules. The Trp repressor is a transcription regulator that shuts off the transcription of genes that code for the enzymes required for the synthesis of tryptophan (see Figure 8–7).

- A. What would happen to the regulation of the tryptophan operon in cells that express a mutant form of the tryptophan repressor that (1) cannot bind to DNA, (2) cannot bind tryptophan, or (3) binds to DNA even in the absence of tryptophan?
- B. What would happen in scenarios (1), (2), and (3) if the cells, in addition, produced normal tryptophan repressor protein from a second, normal gene?

**Figure 8–6 A cluster of bacterial genes can be transcribed from a single promoter.** Each of these five genes encodes a different enzyme; all of the enzymes are needed to synthesize the amino acid tryptophan. The genes are transcribed as a single mRNA molecule, a feature that allows their expression to be coordinated. Clusters of genes transcribed as a single mRNA molecule are common in bacteria. Each such cluster is called an *operon*; expression of the tryptophan operon shown here is controlled by a regulatory DNA sequence called the *operator*, situated within the promoter.



**Figure 8–7 Genes can be switched on and off with repressor proteins.** If the concentration of tryptophan inside the cell is low, RNA polymerase (*blue*) binds to the promoter and transcribes the five genes of the tryptophan operon (*left*). If the concentration of tryptophan is high, however, the repressor protein (*dark green*) becomes active and binds to the operator (*light green*), where it blocks the binding of RNA polymerase to the promoter (*right*). Whenever the concentration of intracellular tryptophan drops, the repressor releases its tryptophan and falls off the DNA, allowing the polymerase to again transcribe the operon. The promoter is marked by two key blocks of DNA sequence information, the  $-35$  and  $-10$  regions, highlighted in yellow (see Figure 7–10). The complete operon is shown in Figure 8–6.

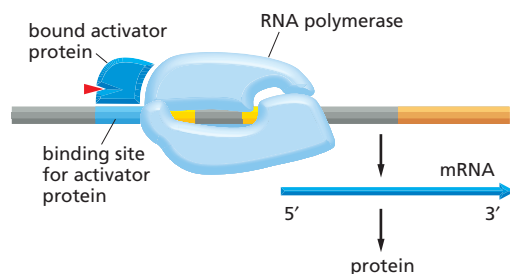
## Repressors Turn Genes Off, Activators Turn Them On

The tryptophan repressor, as its name suggests, is a **repressor** protein: in its active form, it switches genes off, or *represses* them. Some bacterial transcription regulators do the opposite: they switch genes on, or *activate* them. These **activator** proteins work on promoters that—in contrast to the promoter for the tryptophan operator—are, on their own, only marginally able to bind and position RNA polymerase; they may, for example, be recognized only poorly by the polymerase. However, these poorly functioning promoters can be made fully functional by activator proteins that bind to a nearby site on the DNA and contact the RNA polymerase to help it initiate transcription (Figure 8–8). In some cases, a bacterial transcription regulator can repress transcription at one promoter and activate transcription at another; whether the regulatory protein acts as an activator or repressor depends, in large part, on exactly where the regulatory sequences to which it binds are located with respect to the promoter.

Like the tryptophan repressor, activator proteins often have to interact with a second molecule to be able to bind DNA. For example, the bacterial activator protein CAP has to bind cyclic AMP (cAMP) before it can

### Figure 8–8 Gene expression can also be controlled with activator proteins.

An activator protein binds to a regulatory sequence on the DNA and then interacts with the RNA polymerase to help it initiate transcription. Without the activator, the promoter fails to initiate transcription efficiently. In bacteria, the binding of the activator to DNA is often controlled by the interaction of a metabolite or other small molecule (*red triangle*) with the activator protein. For example, the bacterial catabolite activator protein (CAP) must bind cyclic AMP (cAMP) before it can bind to DNA; thus CAP allows genes to be switched on in response to increases in intracellular cAMP concentration.

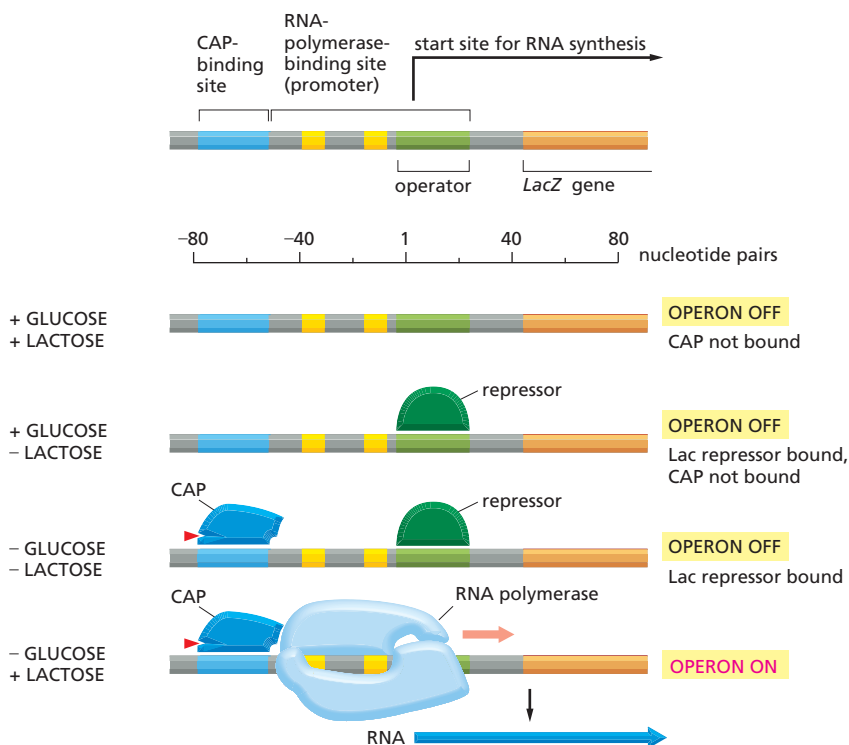


bind to DNA. Genes activated by CAP are switched on in response to an increase in intracellular cAMP concentration, which signals to the bacterium that glucose, its preferred carbon source, is no longer available; as a result, CAP drives the production of enzymes capable of degrading other sugars.

### An Activator and a Repressor Control the *Lac* Operon

In many instances, the activity of a single promoter is controlled by two different transcription regulators. The *Lac* operon in *E. coli*, for example, is controlled by both the *Lac* repressor and the activator protein CAP. The *Lac* operon encodes proteins required to import and digest the disaccharide lactose. In the absence of glucose, CAP switches on genes that allow the cell to utilize alternative sources of carbon—including lactose. It would be wasteful, however, for CAP to induce expression of the *Lac* operon when lactose is not present. Thus the *Lac* repressor shuts off the operon in the absence of lactose. This arrangement enables the control region of the *Lac* operon to integrate two different signals, so that the operon is highly expressed only when two conditions are met: lactose must be present and glucose must be absent (Figure 8–9). This genetic circuit thus behaves like a switch that carries out a logic operation in a computer. When lactose is present AND glucose is absent, the cell executes the appropriate program: in this case, transcription of the genes that permit the uptake and utilization of lactose.

The elegant logic of the *Lac* operon first attracted the attention of biologists more than 50 years ago. The molecular basis of the switch was uncovered by a combination of genetics and biochemistry, providing the first insight into how gene expression is controlled. In a eucaryotic cell, similar gene regulatory devices are combined to generate increasingly complex circuits. Indeed, the developmental program that takes a fertilized egg to adulthood can be viewed as an exceedingly complex circuit composed of simple components like those that control the *Lac* and tryptophan operons.



#### QUESTION 8-2

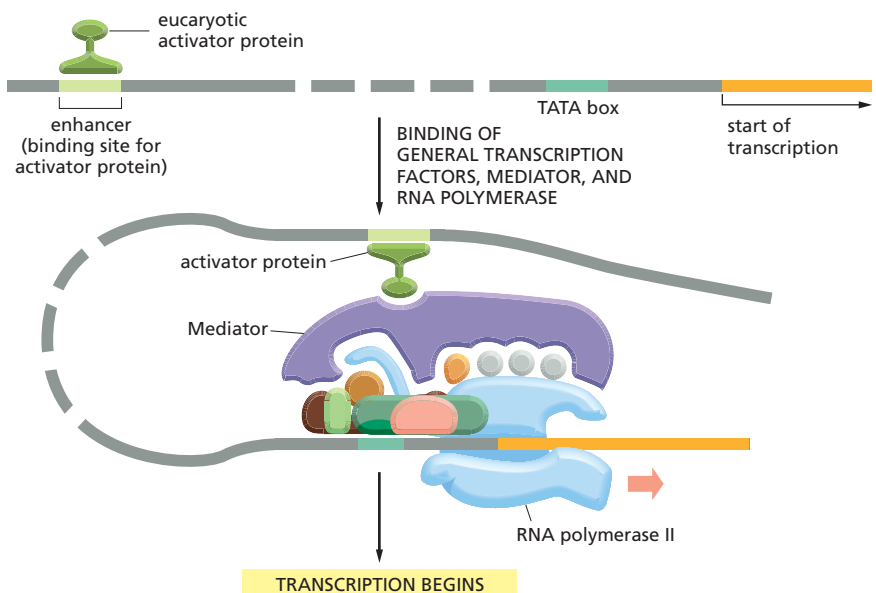
Explain how DNA-binding proteins can make sequence-specific contacts to a double-stranded DNA molecule without breaking the hydrogen bonds that hold the bases together. Indicate how, through such contacts, a protein can distinguish a T-A from a C-G pair. Give your answer in a form similar to Figure 8–4, and indicate what sorts of noncovalent bonds—hydrogen bonds, electrostatic attractions, or hydrophobic interactions (see Panel 2–7, pp. 76–77)—would be made. There is no need to specify any particular amino acid on the protein. The structures of all the base pairs in DNA are given in Figure 5–6.

## Eucaryotic Transcription Regulators Control Gene Expression from a Distance

Eucaryotes, too, use transcription regulators—both activators and repressors—to regulate the expression of their genes. The DNA sites to which eucaryotic gene activators bound were originally termed *enhancers*, because their presence dramatically enhanced, or increased, the rate of transcription. It was surprising to biologists when, in 1979, it was discovered that these activator proteins could enhance transcription even when they are bound thousands of nucleotide pairs away from a gene's promoter. They also work when bound either upstream or downstream from the gene. These observations raised several questions. How do enhancer sequences and the proteins bound to them function over such long distances? How do they communicate with the promoter?

Many models for this 'action at a distance' have been proposed, but the simplest of these seems to apply in most cases. The DNA between the enhancer and the promoter loops out to allow eucaryotic activator proteins to directly influence events that take place at the promoter (Figure 8–10). The DNA thus acts as a tether, causing a protein bound to an enhancer even thousands of nucleotide pairs away to interact with the proteins in the vicinity of the promoter—including RNA polymerase II and the *general transcription factors* (see Figure 7–12). Often, additional proteins serve to link the distantly bound transcription regulators to these proteins at the promoter; the most important is a large complex of proteins known as *Mediator* (see Figure 8–10). One of the ways in which eucaryotic activator proteins function is by aiding in the assembly of the general transcription factors and RNA polymerase at the promoter. Eucaryotic repressor proteins do the opposite: they decrease transcription by preventing or sabotaging the assembly of the same protein complex.

In addition to promoting—or repressing—the assembly of a transcription initiation complex, eucaryotic transcription regulators have an additional mechanism of action: they attract proteins that modulate chromatin structure and thereby affect the accessibility of the promoter to the general transcription factors and RNA polymerase, as we discuss next.



**Figure 8–10 In eucaryotes, gene activation occurs at a distance.** An activator protein bound to DNA attracts RNA polymerase and general transcription factors (see Figure 7–12) to the promoter. Looping of the DNA permits contact between the activator protein bound to the enhancer and the transcription complex bound to the promoter. In the case shown here, a large protein complex called Mediator serves as a go-between. The broken stretch of DNA signifies that the length of DNA between the enhancer and the start of transcription varies, sometimes reaching tens of thousands of nucleotide pairs in length.

## Packing of Promoter DNA into Nucleosomes Affects Initiation of Transcription

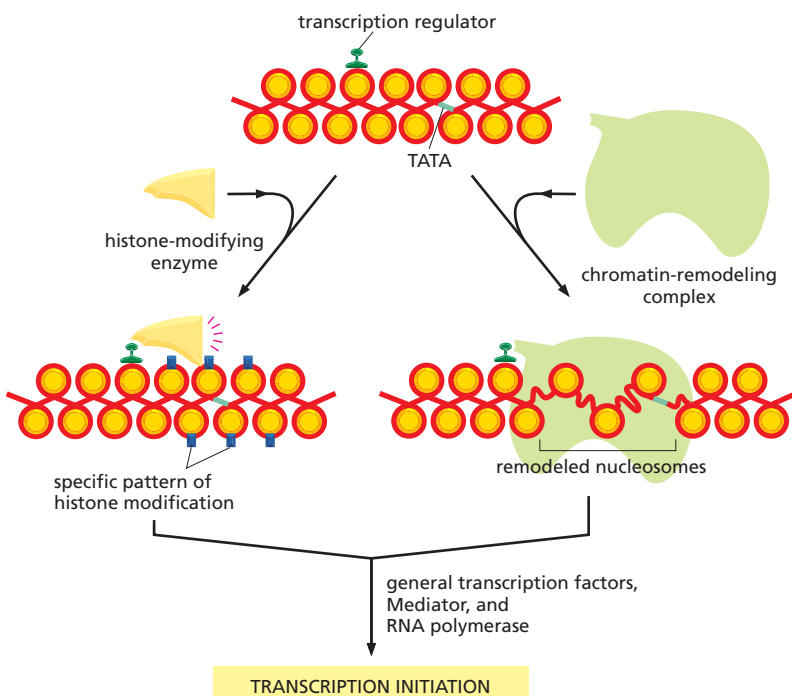
Initiation of transcription in eucaryotic cells must also take into account the packaging of DNA into chromosomes. As we saw in Chapter 5, the genetic material in eucaryotic cells is packed into nucleosomes, which, in turn, are folded into higher-order structures. How do transcription regulators, general transcription factors, and RNA polymerase gain access to such DNA? Nucleosomes can inhibit the initiation of transcription if they are positioned over a promoter, probably because they physically block the assembly of the general transcription factors or RNA polymerase on the promoter. In fact, such chromatin packaging may have evolved in part to prevent leaky gene expression—initiation of transcription in the absence of the proper activator proteins.

In eucaryotic cells, activator and repressor proteins exploit chromatin structure to help turn genes on and off. As we saw in Chapter 5, chromatin structure can be altered by chromatin-remodeling complexes and by enzymes that covalently modify the histone proteins that form the core of the nucleosome (see Figures 5–27 and 5–28). Many gene activators take advantage of these mechanisms by recruiting these proteins to promoters (Figure 8–11). For example, many transcription activators attract *histone acetylases*, which attach an acetyl group to selected lysines in the tail of histone proteins. This modification alters chromatin structure, probably allowing greater accessibility to the underlying DNA; moreover, the acetyl groups themselves are recognized by proteins that promote transcription, including some of the general transcription factors.

Likewise, gene repressor proteins can modify chromatin in ways that reduce the efficiency of transcription initiation. For example, many repressors attract *histone deacetylases*—enzymes that remove the acetyl groups from histone tails, thereby reversing the positive effects that acetylation has on transcription initiation. Although some eucaryotic repressor proteins work on a gene-by-gene basis, others can orchestrate the formation of large swaths of transcriptionally inactive chromatin containing many

### QUESTION 8–3

Some transcription regulators bind to DNA and cause the double helix to bend at a sharp angle. Such “bending proteins” can stimulate the initiation of transcription without contacting either the RNA polymerase, any of the general transcription factors, or any other transcription regulators. Can you devise a plausible explanation for how these proteins might work to modulate transcription? Draw a diagram that illustrates your explanation.



**Figure 8–11 Eucaryotic gene activator proteins can direct local alterations in chromatin structure.** Activator proteins can recruit histone-modifying enzymes and chromatin-remodeling complexes to the promoter region of a gene. The action of these proteins renders the DNA packaged in chromatin more accessible to other proteins in the cell, including those required for transcription initiation. In addition, the covalent histone modifications can serve as binding sites for proteins that stimulate transcription initiation.

genes. As discussed in Chapter 5, these transcription-resistant regions of DNA include the heterochromatin found in interphase chromosomes and the entire X chromosome in female mammals.

## THE MOLECULAR MECHANISMS THAT CREATE SPECIALIZED CELL TYPES

All cells must be able to switch genes on and off in response to signals in their environments. But the cells of multicellular organisms have evolved this capacity to an extreme degree and in highly specialized ways to form an organized array of differentiated cell types. In particular, once a cell in a multicellular organism becomes committed to differentiate into a specific cell type, the choice of fate is generally maintained through many subsequent cell generations. This means that the changes in gene expression, which are often triggered by a transient signal, must be remembered. This phenomenon of *cell memory* is a prerequisite for the creation of organized tissues and for the maintenance of stably differentiated cell types. In contrast, the simplest changes in gene expression in both eucaryotes and bacteria are often only transient; the tryptophan repressor, for example, switches off the tryptophan genes in bacteria only in the presence of tryptophan; as soon as the amino acid is removed from the medium, the genes are switched back on, and the descendants of the cell will have no memory that their ancestors had been exposed to tryptophan.

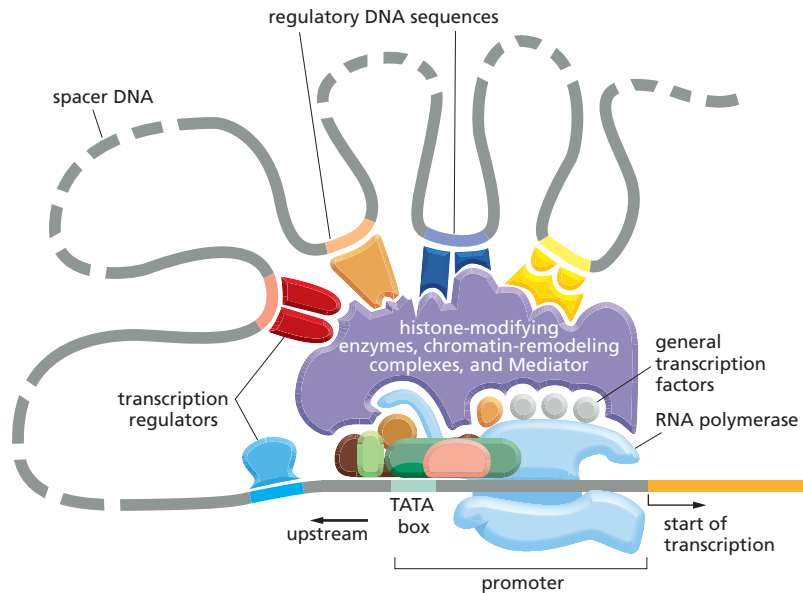
In this section, we discuss some of the special features of transcriptional regulation that are found in multicellular organisms. Our focus will be on how these mechanisms create and maintain the specialized cell types that give a worm, a fly, or a human its distinctive characteristics.

### Eucaryotic Genes Are Regulated by Combinations of Proteins

Because eucaryotic transcription regulators can control transcription initiation when bound to DNA many base pairs away from the promoter, the nucleotide sequences that control the expression of a gene can be spread over long stretches of DNA. In animals and plants it is not unusual to find the regulatory DNA sequences of a gene dotted over tens of thousands of nucleotide pairs, although much of this DNA serves as “spacer” sequence and is not directly recognized by the transcription regulators.

So far in this chapter we have treated transcription regulators as though each functions individually to turn a gene on or off. While this idea holds true for many simple bacterial activators and repressors, most eucaryotic transcription regulators work as part of a “committee” of regulatory proteins, all of which are necessary to express the gene in the right cell, in response to the right conditions, at the right time, and in the required amount.

The term **combinatorial control** refers to the way that groups of regulatory proteins work together to determine the expression of a single gene. We saw a simple example of such regulation by multiple signals when we discussed the bacterial *Lac* operon (see Figure 8–9). In eucaryotes, the regulatory inputs have been amplified, and a typical gene is controlled by dozens of transcription regulators (Figure 8–12). Often, some of these regulatory proteins are repressors and some are activators; the molecular mechanisms by which the effects of all of these proteins are added up to determine the final level of expression for a gene are only now beginning to be understood. An example of such a complex regulatory system—one that participates in the development of a fruit fly from a fertilized egg—is described in [How We Know](#), pp. 282–284.



**Figure 8–12 Transcription regulators work together as a “committee” to control the expression of a eucaryotic gene.** Whereas the general transcription factors that assemble at the promoter are the same for all genes transcribed by polymerase II, the transcription regulators and the locations of their binding sites relative to the promoters are different for different genes. The effects of multiple transcription regulators combine to determine the final rate of transcription initiation. It is not yet understood in detail how these multiple inputs are integrated.

## The Expression of Different Genes Can Be Coordinated by a Single Protein

In addition to being able to switch individual genes on and off, all organisms—whether prokaryote or eucaryote—need to coordinate the expression of different genes. When a eucaryotic cell receives a signal to divide, for example, a number of hitherto unexpressed genes are turned on together to set in motion the events that lead eventually to cell division (discussed in Chapter 18). One way in which bacteria coordinate the expression of a set of genes is by having them clustered together in an operon under the control of a single promoter (see Figure 8–6). This is not the case in eucaryotes, in which each gene is transcribed and regulated individually. So how do eucaryotes coordinate gene expression? In particular, given that a eucaryotic cell uses a committee of transcription regulators to control each of its genes, how can it rapidly and decisively switch whole groups of genes on or off? The answer is that even though control of gene expression is combinatorial, the effect of a single transcription regulator can still be decisive in switching any particular gene on or off, simply by completing the combination needed to activate or repress that gene. This is like dialing in the final number of a combination lock: the lock will spring open if the other numbers have been previously entered. Just as the same number can complete the combination for different locks, the same protein can complete the combination for several different genes. As long as different genes contain DNA sequences recognized by the same transcription regulator, they can be switched on or off together, as a unit.

An example of this style of regulation in humans is seen with the *glucocorticoid receptor protein*. In order to bind to regulatory sites in DNA this transcription regulator must first form a complex with a molecule of a glucocorticoid hormone (for example, cortisol; see Table 16–1, p. 535). In response to glucocorticoids, liver cells increase the expression of many different genes, one of which encodes the enzyme tyrosine aminotransferase, as discussed earlier. These genes are all regulated by the binding of the glucocorticoid hormone–receptor complex to a regulatory sequence in the DNA of each gene. When the body has recovered and the

# HOW WE KNOW:

## GENE REGULATION—THE STORY OF *EVE*

The ability to regulate gene expression is crucial to the proper development of a multicellular organism from a fertilized egg to a fertile adult. Beginning at the earliest moments in development, a succession of programs controls the differential expression of genes that allows an animal to form a proper body plan—helping to distinguish its back from its belly, and its head from its tail. These cues ultimately direct the correct placement of a wing or a leg, a mouth or an anus, a neuron or a sex cell.

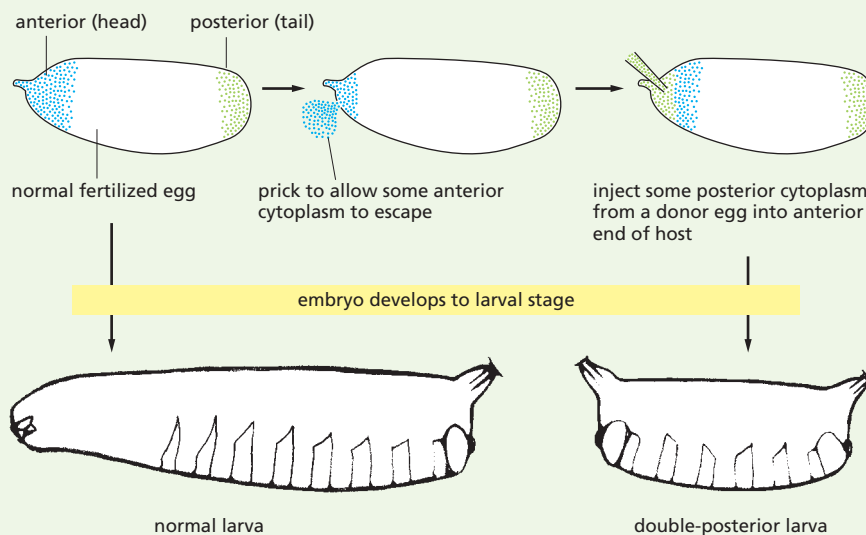
A central problem in development, then, is understanding how an organism generates these patterns of gene expression, which are laid down within hours of fertilization. A large part of the story rests on the action of transcription regulators. By interacting with different regulatory DNA sequences, these proteins instruct every cell in the embryo to switch on the genes that are appropriate for that cell at each time point during development. How can a protein binding to a piece of DNA help direct the development of a complex multicellular organism? To see how we can address that large question, we now review the story of *Eve*.

### In the Big Egg

*Even-skipped*—*Eve*, for short—is a gene whose expression plays an important role in the development of the *Drosophila* embryo. If this gene is inactivated by

mutation, many parts of the embryo fail to form and the fly larva dies early in development. At the stage of development when *Eve* is first switched on, the embryo developing within the egg is still a single, giant cell containing multiple nuclei afloat in a common cytoplasm. This embryo, which is some 400  $\mu\text{m}$  long and 160  $\mu\text{m}$  in diameter, is formed from the fertilized egg through a series of rapid nuclear divisions that occur without cell division. Eventually each nucleus will be enclosed in a plasma membrane and become a cell; however, the events that concern us happen before this cellularization.

The cytoplasm of this giant egg is far from uniform: the anterior (head) end of the embryo contains different proteins from those in the posterior (tail) end. The presence of these asymmetries in the fertilized egg and the early embryo was originally demonstrated by experiments in which *Drosophila* eggs were made to leak. If the front end of an egg is punctured carefully and a small amount of the anterior cytoplasm is allowed to ooze out, the embryo will fail to develop head segments. Further, if cytoplasm taken from the posterior end of another egg is then injected into this somewhat depleted anterior area, the animal will develop a second set of abdominal segments where its head parts should have been (Figure 8–13).



**Figure 8–13 Molecules localized at the ends of the *Drosophila* egg control its anterior–posterior polarity.** A small amount of cytoplasm is allowed to leak out of the anterior end of the egg and is replaced by an injection of posterior cytoplasm. The resulting double-tailed embryo (right) shows a duplication of the last three abdominal segments. A normal embryo (left) is shown for comparison. (Adapted from C. Nüsslein-Volhard, H.G. Frohnhöfer, and R. Lehmann, *Science* 238:1675–1681, 1987. With permission from AAAS.)

## Finding the Proteins

This egg-draining experiment shows that the normal head-to-tail pattern of development is controlled by substances located at each end of the embryo. And researchers were betting that these substances were proteins. To identify them, investigators subjected eggs to a treatment that would inactivate genes at random. They then searched for embryos whose head-to-tail body plan looked abnormal. In these mutant animals, the genes that were disrupted must encode proteins that are important for establishing proper anterior–posterior polarity.

Using this approach, researchers discovered many genes required for setting up anterior–posterior polarity, including genes encoding four key transcription regulators: Bicoid, Hunchback, Krüppel, and Giant. (*Drosophila* genes are often given colorful names that reflect the appearance of flies in which the gene is inactivated by mutation.) Once these proteins had been identified, researchers could prepare antibodies that would recognize each. These antibodies, coupled to fluorescent markers, were then used to determine where in the early embryo each protein is localized (see Panel 1–1, pp. 8–9).

The results of these antibody-staining experiments are striking. The cytoplasm of the early embryo, it turns out, contains a mixture of these transcription regulators, each distributed in a unique pattern along the length of the embryo (Figure 8–14). As a result, the nuclei inside this giant, multinucleate cell begin to express different genes depending on which transcription regulators they are exposed to, which in turn depends on the location of each nucleus along the embryo. Nuclei near the anterior end of the embryo, for example, encounter a set of transcription regulators that is distinct from the set that bathe nuclei at the posterior end. Thus the differing amounts of these proteins provide the many nuclei in the developing embryo with positional information along the anterior–posterior axis of the embryo.

This is where *Eve* comes in. The regulatory DNA sequences of the *Eve* gene can read the concentrations of the transcription regulators at each position along the length of the embryo. Based on this information, *Eve* is expressed in seven stripes, each at a precise location along the anterior–posterior axis of the embryo. To find out how these regulatory proteins control the expression of *Eve* with such precision, researchers next set their sights on the regulatory region of the *Eve* gene.

## Dissecting the DNA

As we have seen in this chapter, regulatory DNA sequences control which cells in an organism will express a particular gene, and at what point that gene will be turned on. One way to learn when and where a regulatory DNA sequence is active is to hook the sequence up to a **reporter gene**—a gene encoding a protein whose activity is easy to monitor experimentally. The regulatory DNA sequences will now drive the expression of the reporter gene. This artificial DNA construct is then reintroduced into a cell or an organism, and the activity of the reporter protein is measured.

By coupling various portions of the regulatory sequence of *Eve* to a reporter gene, researchers discovered that the *Eve* gene contains a series of seven regulatory modules, each of which is responsible for specifying a single stripe of *Eve* expression along the embryo. So, for example, researchers could remove the regulatory module that specifies stripe 2 from its normal setting upstream of *Eve*, place it in front of a reporter gene, and reintroduce this engineered DNA sequence into the *Drosophila* genome (Figure 8–15A). When embryos carrying this genetic construct are examined, the reporter gene is found to be expressed in precisely the position

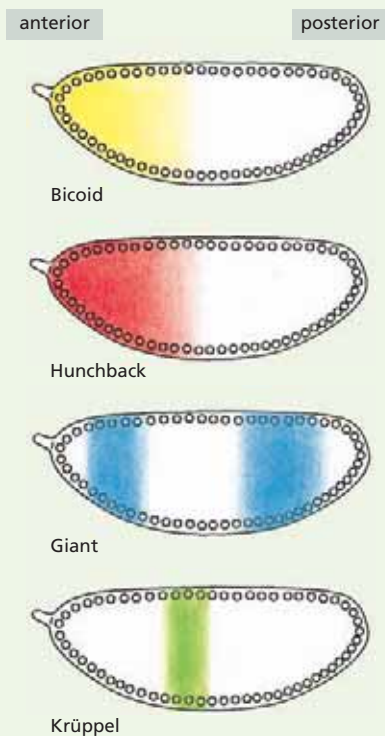
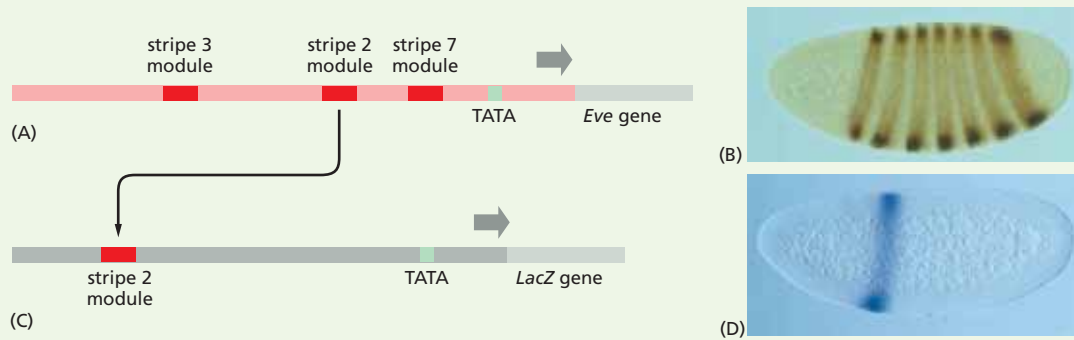


Figure 8–14 The early *Drosophila* embryo shows a nonuniform distribution of four transcription regulators.

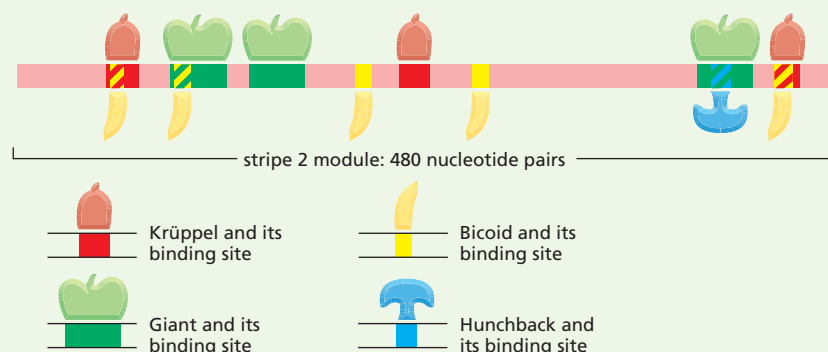


**Figure 8–15 A reporter gene reveals the modular construction of the *Eve* gene regulatory region.** (A) The *Eve* gene contains regulatory sequences that direct the production of *Eve* protein in stripes along the embryo. (B) Embryos stained with antibodies to the *Eve* protein show the seven characteristic *Eve* stripes. (C) In this experiment, a 480-nucleotide piece of the *Eve* regulatory region (the stripe 2 module from A) is removed and inserted upstream of the *E. coli* *LacZ* gene, which encodes the enzyme  $\beta$ -galactosidase (see Figure 8–9). (D) When the engineered DNA containing a single regulatory region is reintroduced into the genome of a *Drosophila* embryo, the resulting embryo expresses  $\beta$ -galactosidase precisely in the position of the second of the seven *Eve* stripes. Enzyme activity is assayed by the addition of X-gal, a modified sugar that when cleaved by  $\beta$ -galactosidase generates an insoluble blue product. (B and D, courtesy of Stephen Small and Michael Levine.)

of stripe 2 (Figure 8–15B). Similar experiments revealed the existence of other regulatory modules, one for each of the other six stripes.

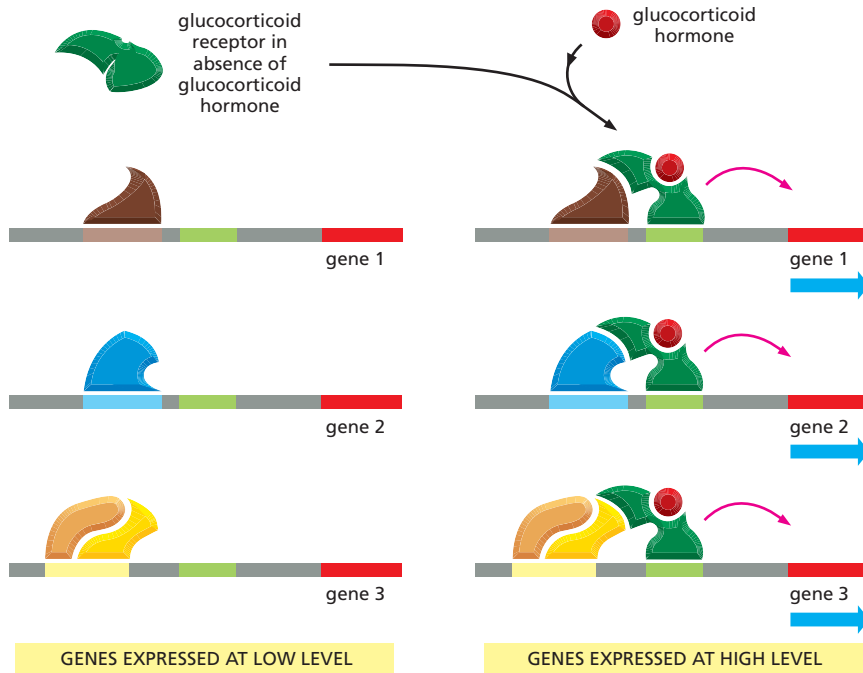
The question then becomes: how does each of these modules direct the formation of a single stripe in a specific position? The answer, researchers found, is that each module contains a unique combination of regulatory sequences that bind different combinations of the four transcription regulators that are present in gradients in the early embryo. The stripe 2 unit, for example, contains recognition sequences for all four regulators—Bicoid and Hunchback activate *Eve* transcription, while Krüppel and Giant repress it (Figure 8–16). The concentrations of these four proteins vary across the embryo (see Figure 8–14), and these patterns determine which of the proteins are bound to the *Eve* stripe 2 module at each position along the embryo. The combination of bound proteins then ‘tells’ the appropriate nuclei to express *Eve*, and stripe 2 is formed.

The other stripe regulatory modules are thought to function along similar lines; each module reads positional information provided by some unique combination of transcription regulators and expresses *Eve* on the basis of this information. The entire gene control region of *Eve* is strung out over 20,000 nucleotide pairs of DNA and binds more than 20 transcription regulators, including the four discussed here. A large and complex control region is thereby formed from a series of smaller modules, each of which consists of a unique arrangement of short DNA sequences recognized by specific transcription regulators. In this way, a single gene can respond to an enormous number of combinatorial inputs. *Eve* itself is a transcription regulator and it—in combination with many other regulatory proteins—controls key events later in the development of the fly. This organization begins to explain how the development of a complex organism can be orchestrated by repeated applications of a few basic principles.



**Figure 8–16 The regulatory module for *Eve* stripe 2 contains binding sites for four different transcription regulators.**

All four regulators are responsible for the proper expression of *Eve* in stripe 2. Flies that are deficient in the two activators, Bicoid and Hunchback, fail to form stripe 2 efficiently; in flies deficient in either of the two repressors, Giant or Krüppel, stripe 2 expands and covers an abnormally broad region of the embryo. As indicated in the top diagram, in some cases the binding sites for the transcription regulators overlap and the proteins compete for binding to the DNA. For example, the binding of Bicoid and Krüppel to the site at the far right is thought to be mutually exclusive.



**Figure 8–17 A single transcription regulator can coordinate the expression of many different genes.** The action of the glucocorticoid receptor is illustrated. On the left is shown a series of genes, each of which has various gene activator proteins bound to its regulatory region. However, these bound proteins are not sufficient on their own to activate transcription efficiently. On the right is shown the effect of adding an additional transcription regulator—the glucocorticoid receptor in a complex with glucocorticoid hormone—that can bind to the regulatory sequences of each gene. The glucocorticoid receptor completes the combination of transcription regulators required for efficient initiation of transcription, and the genes are now switched on as a set.

hormone is no longer present, the expression of all of these genes drops to its normal level. In this way a single transcription regulator can control the expression of many different genes (Figure 8–17).

### Combinatorial Control Can Create Different Cell Types

The ability to switch many different genes on or off using just one protein is not only useful in the day-to-day regulation of cell function. It is also one of the means by which eucaryotic cells differentiate into particular types of cells during embryonic development.

A striking example of the effect of a single transcription regulator on differentiation comes from studying the development of muscle cells. A mammalian skeletal muscle cell is a highly distinctive cell type. It is typically an extremely large cell that is formed by the fusion of many muscle precursor cells called *myoblasts*. The mature muscle cell is distinguished from other cells by the production of a large number of characteristic proteins, such as the actin and myosin that make up the contractile apparatus (discussed in Chapter 17) as well as the receptor proteins and ion channel proteins in the cell membranes that make the muscle cell sensitive to nerve stimulation. Genes encoding these muscle-specific proteins are all switched on coordinately as the myoblasts begin to fuse. Studies of muscle cells differentiating in culture have identified key transcription regulators, expressed only in potential muscle cells, that coordinate gene expression and thus are crucial for muscle-cell differentiation. These regulators activate the transcription of the genes that code for the muscle-specific proteins by binding to specific DNA sequences present in their regulatory regions.

These key transcription regulators can convert nonmuscle cells to myoblasts by activating the changes in gene expression typical of differentiating muscle cells. For example, when one of these regulators, MyoD, is artificially expressed in fibroblasts cultured from skin connective tissue, the fibroblasts start to behave like myoblasts and fuse to form musclelike cells. The dramatic effect of expressing the *MyoD* gene in fibroblasts is shown in Figure 8–18. It appears that the fibroblasts, which are derived from the same broad class of embryonic cells as muscle cells,



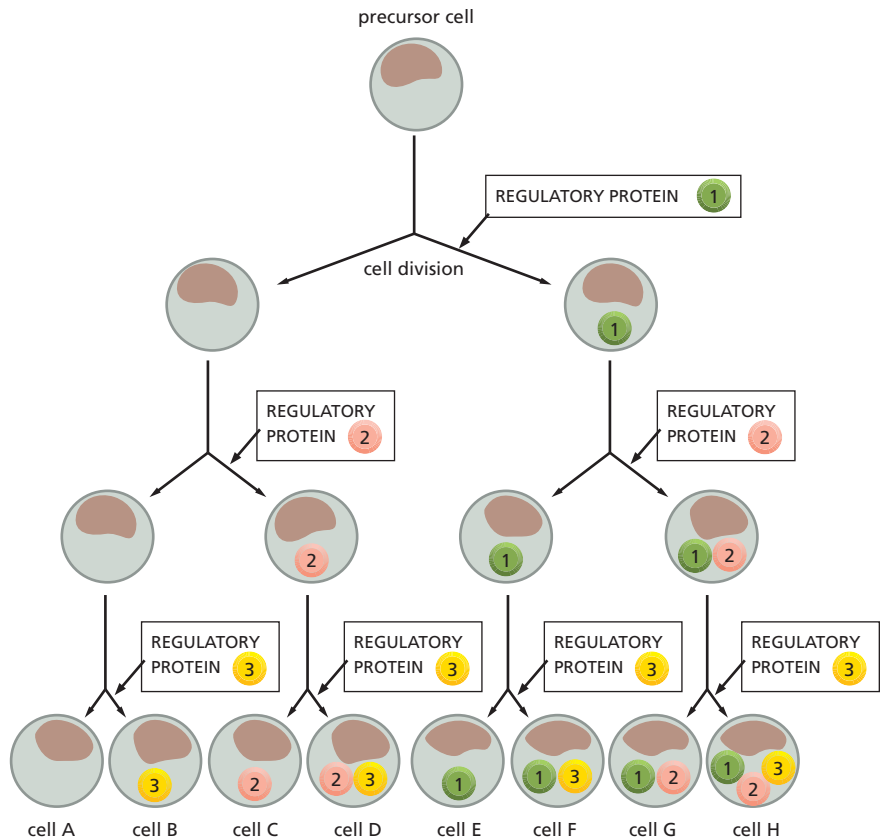
**Figure 8–18 Fibroblasts can be converted to muscle cells by a single transcription regulator.** As shown in this immunofluorescence micrograph, fibroblasts from the skin of a chick embryo have been converted to muscle cells by the experimentally induced expression of the *MyoD* gene. The fibroblasts that expressed the *MyoD* gene have fused to form elongated multinucleate musclelike cells, which are stained green with an antibody that detects a muscle-specific protein. Fibroblasts that do not express *MyoD* are barely visible in the background. (Courtesy of Stephen Tapscott and Harold Weintraub.)

have already accumulated many of the other necessary transcription regulators required for the combinatorial control of the muscle-specific genes, and that addition of *MyoD* completes the unique combination that directs the cells to become muscle. Some other cell types fail to be converted to muscle by the addition of *MyoD*; these cells presumably have not accumulated the other required transcription regulators during their developmental history.

How the accumulation of different transcription regulators can lead to the generation of different cell types is illustrated schematically in **Figure 8–19**. This figure also illustrates how, thanks to the possibilities of combinatorial control and shared regulatory DNA sequences, a limited set of transcription regulators can control the expression of a much larger number of genes.

The conversion of one cell type (fibroblast) to another (muscle) by a single transcription regulator emphasizes one of the most important principles discussed in this chapter: the dramatic differences between cell types—such as size, shape, and function—are produced by differences in gene expression.

**Figure 8–19 Combinations of a few transcription regulators can generate many different cell types during development.** In this simple scheme a “decision” to make a new regulator (shown as a numbered circle) is made after each cell division. Repetition of this simple rule enables eight cell types (A through H) to be created using only three different transcription regulators. Each of these hypothetical cell types would then express different genes, as dictated by the combination of transcription regulators that are present within it.



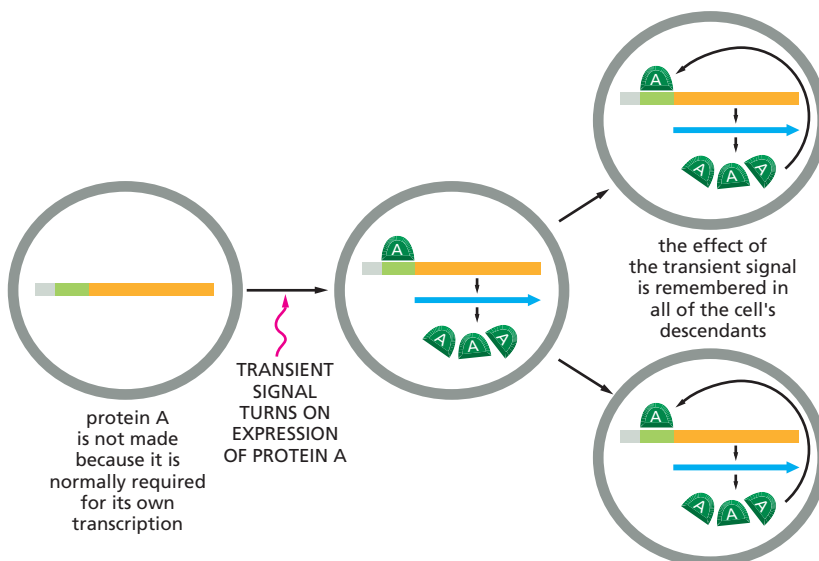
## Stable Patterns of Gene Expression Can Be Transmitted to Daughter Cells

As discussed earlier in this chapter, once a cell in a multicellular organism has become differentiated into a particular cell type, it will generally remain differentiated, and all its progeny cells will remain that same cell type. Some highly specialized cells never divide again once they have differentiated; for example, skeletal muscle cells and neurons. But many other differentiated cells, such as fibroblasts, smooth muscle cells, and liver cells (hepatocytes), will divide many times in the life of an individual. All of these cell types give rise only to cells like themselves when they divide: smooth muscle does not give rise to liver cells, nor liver cells to fibroblasts.

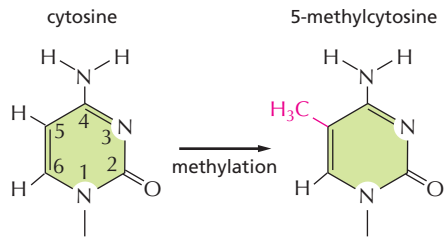
This preservation of cellular identity means that the changes in gene expression that give rise to a differentiated cell must be remembered and passed on to its daughter cells through all subsequent cell divisions. For example, in the cells illustrated in Figure 8–19, the production of each transcription regulator, once begun, has to be perpetuated in the daughter cells of each cell division. How might this be accomplished?

Cells have several ways of ensuring that their daughters “remember” what kind of cells they are supposed to be. One of the simplest is through a **positive feedback loop**, where a key transcription regulator activates transcription of its own gene in addition to that of other cell-type-specific genes (Figure 8–20). The MyoD protein discussed earlier functions in such a positive feedback loop. Another way of maintaining cell type is through the faithful propagation of a condensed chromatin structure from parent to daughter cell. We saw an example of this in Figure 5–30, where the same X chromosome is inactive through many cell generations.

A third way in which cells can transmit information about gene expression to their progeny is through **DNA methylation**. In vertebrate cells, DNA methylation occurs exclusively on cytosine bases (Figure 8–21). This covalent modification of cytosines generally turns off genes by attracting proteins that block gene expression. DNA methylation patterns are passed on to progeny cells by the action of an enzyme that copies the methylation pattern on the parent DNA strand to the daughter DNA strand immediately after replication (Figure 8–22). Because each of these mechanisms—positive feedback loops, certain forms of condensed chromatin,



**Figure 8–20 A positive feedback loop can create cell memory.** Protein A is a transcription regulator that activates its own transcription. All of the descendants of the original cell will therefore “remember” that the progenitor cell had experienced a transient signal that initiated the production of the protein.



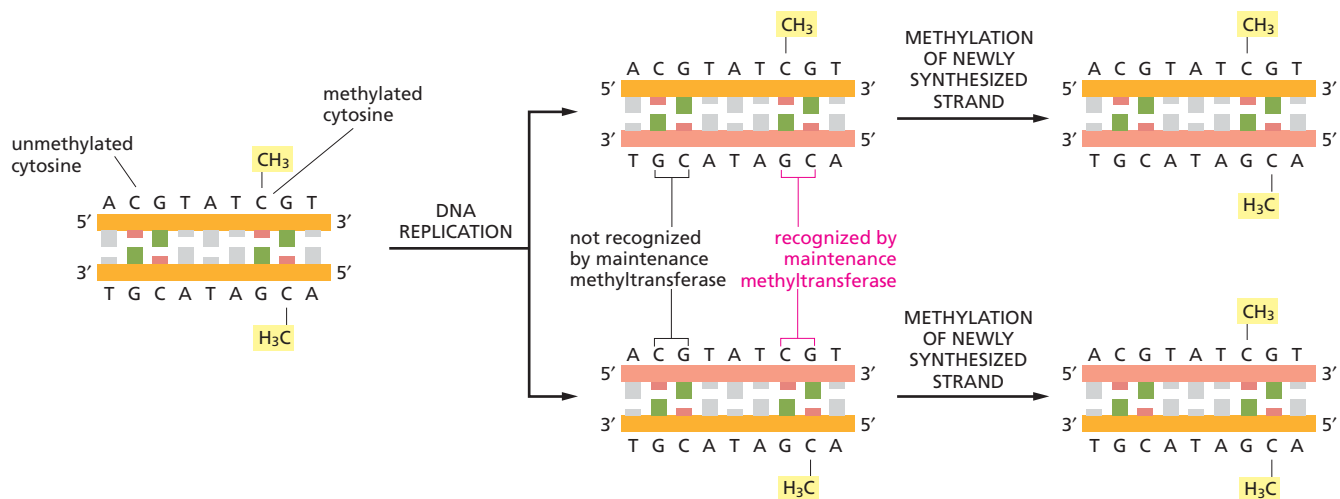
**Figure 8–21 Formation of 5-methylcytosine occurs by methylation of a cytosine base in the DNA double helix.** In vertebrates this event is confined to selected cytosine (C) nucleotides that fall next to a guanine (G).

and DNA methylation—transmits information from parent to daughter cell without altering the actual nucleotide sequence of the DNA, they are considered forms of *epigenetic inheritance* (see p. 192).

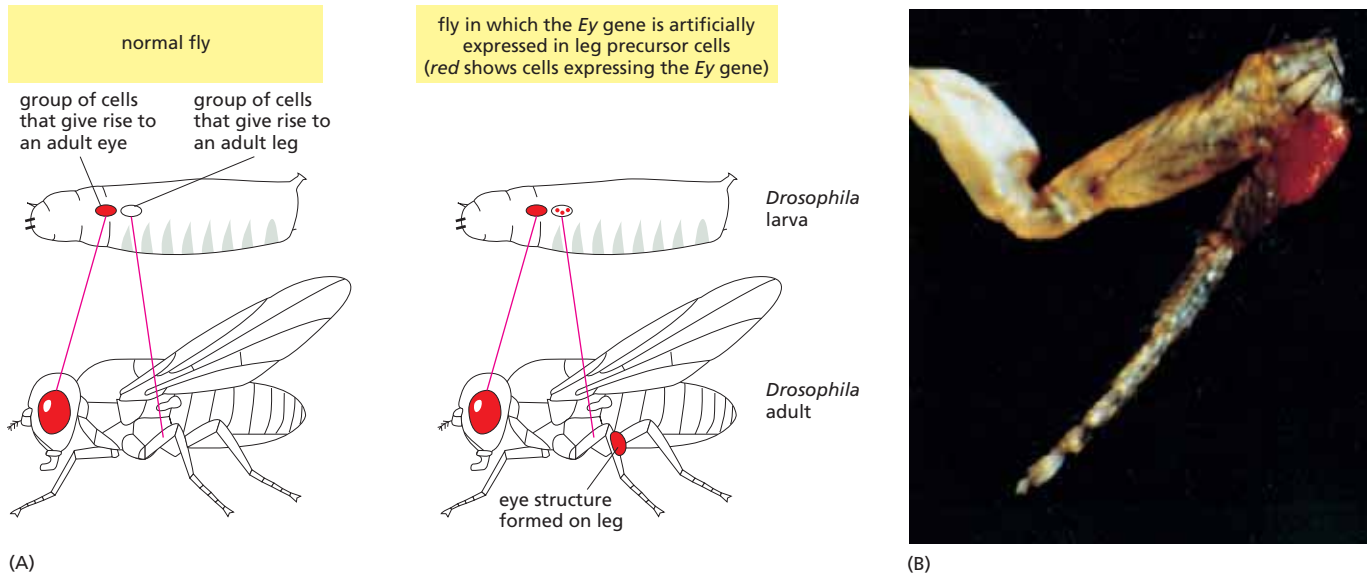
## The Formation of an Entire Organ Can Be Triggered by a Single Transcription Regulator

We have seen that even though combinatorial control is the norm for eucaryotic genes, a single transcription regulator can be decisive in switching a whole set of genes on or off, and can convert one cell type into another. A dramatic extension of this principle comes from studies of eye development in *Drosophila*, mice, and humans. Here, a transcription regulator, called *Ey* in flies and *Pax-6* in vertebrates, is crucial for eye development. When expressed in the proper type of cell, *Ey* can trigger the formation of not just a single cell type but a whole organ—the eye—composed of different types of cells all properly organized in three-dimensional space.

The best evidence for the action of *Ey* comes from experiments in fruit flies in which the *Ey* gene is artificially expressed early in development in cells that normally go on to form legs. This abnormal gene expression causes eyes to develop in the middle of the legs (Figure 8–23). The *Drosophila* eye is composed of thousands of cells, and how the *Ey* protein coordinates the specification of each cell in the eye is an actively studied topic in developmental biology. Here, we shall simply note that *Ey* directly controls the expression of many genes by binding to DNA sequences in their regulatory regions. Some of the genes controlled by *Ey* encode additional transcription regulators that, in turn, control the expression of other genes. Moreover, some of these regulators act back on *Ey* itself to create a positive feedback loop that ensures the continued production of the *Ey* protein. So the action of just one transcription regulator can produce a cascade of regulators whose combined actions lead to the formation of an organized group of many different types of cells. One can begin to imagine how, by repeated applications of this principle, a complex organism is built piece by piece.



**Figure 8–22 DNA methylation patterns can be faithfully inherited.** An enzyme called a maintenance methyltransferase guarantees that once a pattern of DNA methylation has been established, it is inherited by progeny DNA. Immediately after replication, each daughter helix will contain one methylated DNA strand—inherited from the parent helix—and one unmethylated, newly synthesized strand. The maintenance methyltransferase interacts with these hybrid helices, where it methylates only those CG sequences that are base-paired with a CG sequence that is already methylated. In vertebrate DNA, a large portion of the cytosines in CG sequences are methylated.



**Figure 8–23 Expression of the *Drosophila Ey* gene in the precursor cells of the leg triggers the development of an eye on the leg.** (A) Simplified diagrams showing the result when a fruit fly larva contains either the normally expressed *Ey* gene (left) or an *Ey* gene that is additionally expressed artificially in cells that will give rise to legs (right). (B) Photograph of an abnormal leg that contains a misplaced eye. (B, courtesy of Walter Gehring.)

## POST-TRANSCRIPTIONAL CONTROLS

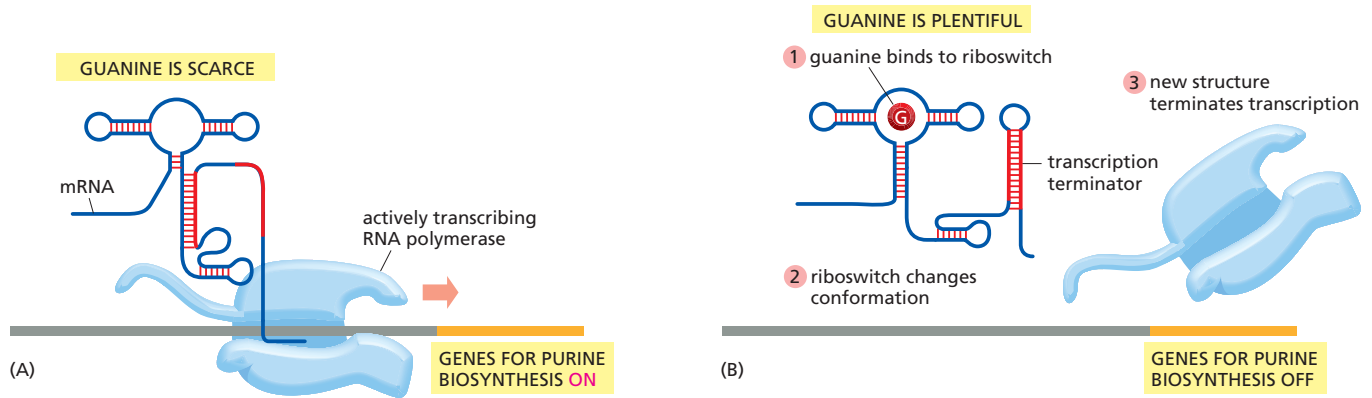
We have seen that transcription regulators control gene expression by switching on or off transcription initiation. The vast majority of genes in all organisms are regulated in this way. But additional points of control can come into play later in the pathway from DNA to protein, giving cells a further opportunity to manage the amount of gene product that is made. These **post-transcriptional controls**, which operate after RNA polymerase has bound to a gene's promoter and started to synthesize RNA, are crucial for the regulation of many genes.

In Chapter 7, we described one type of post-transcriptional control: alternative splicing, which allows different forms of a protein to be made in different tissues (Figure 7–21). Here we discuss a few more examples of the many ways in which cells can manipulate gene expression after transcription has begun.

### Riboswitches Provide An Economical Solution to Gene Regulation

The mechanisms for controlling gene expression we have described thus far all involve the participation of a regulatory protein. But scientists have recently discovered a number of mRNAs that can regulate their own transcription and translation. These self-regulating mRNAs contain **riboswitches**: short sequences of RNA that change their conformation when bound to small molecules such as metabolites. Many riboswitches have been discovered, and each recognizes a specific small molecule. The conformational change that is driven by the binding of that molecule can regulate gene expression (Figure 8–24). This mode of gene regulation is particularly common in bacteria, where riboswitches sense key small metabolites in the cell and adjust gene expression accordingly.

Riboswitches are perhaps the most economical examples of gene control devices, because they bypass the need for regulatory proteins altogether. The fact that short sequences of RNA can form such highly efficient gene



**Figure 8–24 A riboswitch controls purine biosynthesis genes in bacteria.** (A) When guanine is scarce, the riboswitch adopts a structure that allows the elongating RNA polymerase, which has already initiated transcription, to continue transcribing into the purine biosynthetic genes. The enzymes needed for guanine synthesis are thereby expressed. (B) When guanine is abundant, it binds to the riboswitch, causing it to undergo a conformational change. The new conformation includes a double-stranded structure (red) that forces the polymerase to terminate transcription before it reaches the purine biosynthetic genes. In the absence of guanine, formation of this double-stranded structure is blocked because one of the RNA strands that forms it pairs with a different region of the riboswitch (A). In this example, the riboswitch blocks completion of an mRNA. Other riboswitches control translation of mRNA molecules once they have been synthesized. (Adapted from M. Mandal and R.R. Breaker, *Nat. Rev. Mol. Cell Biol.* 5:451–63, 2004. With permission from Macmillan Publishers Ltd.)

control devices offers further evidence that, before modern cells arose, a world run by RNAs may have reached a high level of sophistication (see pp. 261–264).

### The Untranslated Regions of mRNAs Can Control Their Translation

Once an mRNA has been synthesized, one of the most common ways of regulating how much of its protein product is made is to control the initiation of translation. Although the details of translation initiation differ between eucaryotes and bacteria, both use the same basic strategies for regulating gene expression at this step.

Bacterial mRNAs contain a short ribosome-binding sequence located a few nucleotides upstream of the AUG codon where translation begins. This recognition sequence forms base pairs with the RNA in the small ribosomal subunit, correctly positioning the initiating AUG codon within the ribosome. Because this interaction is needed for efficient translation initiation, it provides an ideal target for translational control. By blocking—or exposing—the ribosome recognition sequence, the bacterium can either inhibit—or promote—the translation of an mRNA (Figure 8–25).

Eucaryotic mRNAs possess a 5' cap that helps guide the ribosome to the first AUG, the codon where translation will start (see Figure 7–35). In eucaryotic cells, repressors can inhibit translation initiation by binding to specific RNA sequences in the 5' untranslated region of the mRNA and keeping the ribosome from finding the first AUG. When conditions change, the cell can inactivate the repressor and thereby increase translation of the mRNA.

### Small Regulatory RNAs Control the Expression of Thousands of Animal and Plant Genes

As we saw in Chapter 7, RNAs perform many critical tasks in cells. In addition to acting as intermediate carriers of genetic information, they play key structural and catalytic roles, particularly in protein synthesis (see pp. 253–254). But a recent series of striking discoveries has revealed that noncoding RNAs—those that do not direct the production of a protein product—are far more prevalent than previously imagined and play unanticipated, widespread roles in regulating gene expression.

One particularly important type of noncoding RNA, found in plants and animals, is called **microRNA (miRNA)**. Humans, for example, produce more than 400 different miRNAs, which seem to regulate at least one-third of all protein-coding genes. These short, regulatory RNAs control gene expression by base-pairing with specific mRNAs and controlling their stability and their translation.

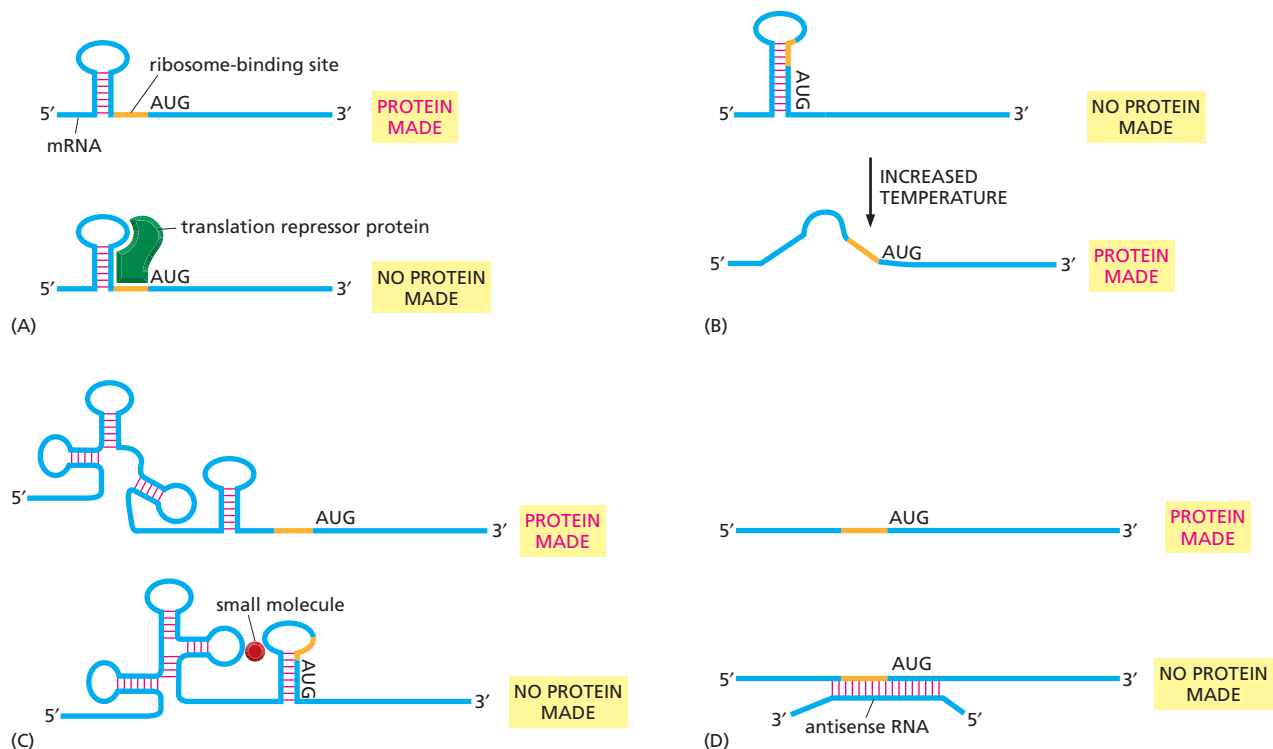
Like other noncoding RNAs, such as tRNA and rRNA, the precursor miRNA transcript undergoes a special type of processing to yield the mature miRNA. This miRNA is then assembled with specialized proteins to form an *RNA-induced silencing complex (RISC)*. The RISC patrols the cytoplasm, searching for mRNAs that are complementary to the miRNA it carries (Figure 8–26). Once a target mRNA forms base pairs with an miRNA, it is destroyed immediately by a nuclease present within the RISC or else its translation is blocked and it is delivered to a region of the cytoplasm where other nucleases will eventually degrade it. Once the RISC has taken care of an mRNA molecule, it is released and is free to seek out additional mRNA molecules. Thus a single miRNA—as part of a RISC—can eliminate one mRNA molecule after another, thereby efficiently blocking production of the protein that the mRNA encodes.

Two features of miRNAs make them especially useful regulators of gene expression. First, a single miRNA can regulate a whole set of different mRNAs so long as the mRNAs carry a common sequence; these sequences are often located in their 5' and 3' untranslated regions. In humans, some individual miRNAs control hundreds of different mRNAs in this manner. Second, a gene that encodes an miRNA occupies relatively little space in the genome compared with one that encodes a transcription regulator. Indeed, their small size is one reason that miRNAs were discovered only recently. Although we are only beginning to understand the full impact of miRNAs, it is clear that they represent a critical part of the cell's equipment for regulating the expression of its genes.

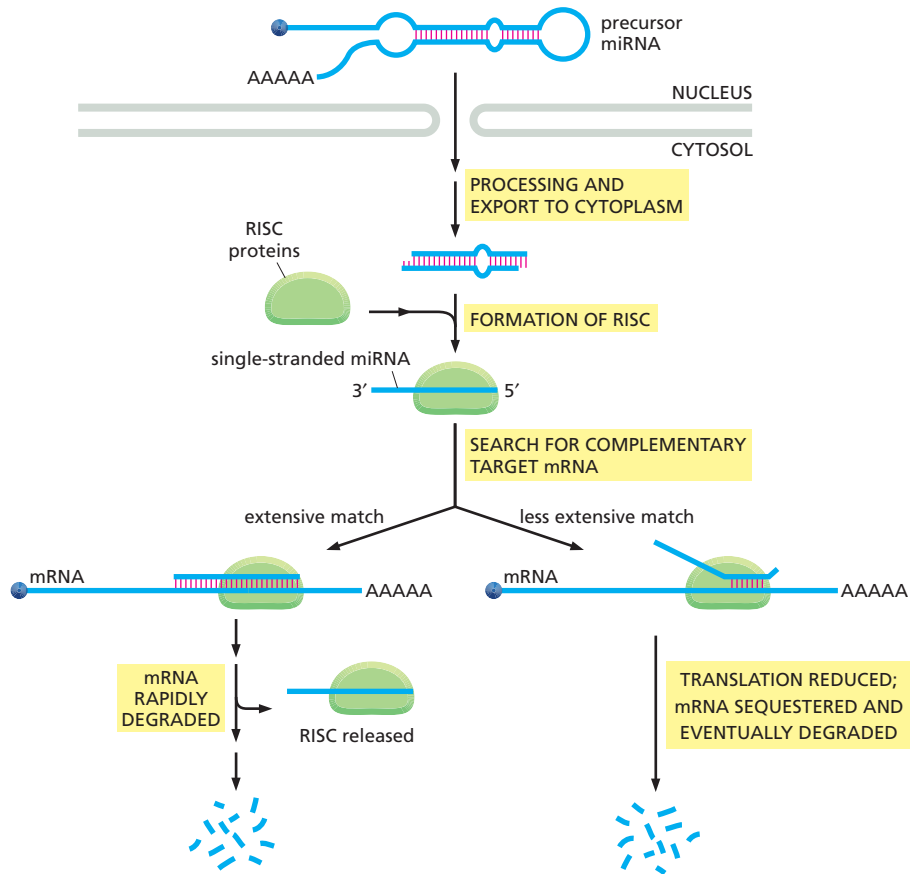
## RNA Interference Destroys Double-Stranded Foreign RNAs

Some of the proteins that process and package miRNAs also serve as a cell defense mechanism: they orchestrate the destruction of 'foreign' RNA molecules, specifically those that are double-stranded. Many viruses—and transposable genetic elements—produce double-stranded RNA

**Figure 8–25 Gene expression can be controlled by regulating translation initiation.** (A) Sequence-specific RNA-binding proteins can repress the translation of specific mRNAs by keeping the ribosome from binding to the ribosome-recognition sequence (orange) found at the start of a bacterial protein-coding gene. Some ribosomal proteins use this mechanism to inhibit the translation of their own mRNA. (B) An mRNA from the pathogen *Listeria monocytogenes* contains a 'thermosensor' RNA sequence that controls the translation of a set of virulence genes. At the warmer temperature that the bacterium encounters inside its human host, the thermosensor sequence is denatured and the virulence genes are expressed. (C) Binding of a small molecule to a riboswitch causes a structural rearrangement of the RNA, sequestering the ribosome-recognition sequence and blocking translation initiation. (D) A complementary, 'antisense' RNA produced by another gene base-pairs with a specific mRNA and blocks its translation. Although these examples of translational control are from bacteria, many of the same principles operate in eucaryotes.



**Figure 8–26 An miRNA targets a complementary mRNA transcript for destruction.** The precursor miRNA is processed to form a mature miRNA. It then assembles with a set of proteins into a complex called RISC. The miRNA then guides the RISC to mRNAs that have a complementary nucleotide sequence. Depending on how extensive the region of complementarity is, the target mRNA is either rapidly degraded by a nuclease within the RISC or transferred to an area of the cytoplasm where other cellular nucleases will destroy it.



some time in their life cycles. This targeted RNA degradation mechanism, called **RNA interference (RNAi)**, helps to keep these potentially dangerous invaders in check.

The presence of foreign, double-stranded RNA in the cell triggers RNAi by first attracting a protein complex containing a nuclease called Dicer. Dicer cleaves the double-stranded RNA into short fragments (approximately 23 nucleotide pairs in length) called **small interfering RNAs (siRNAs)**. These short, double-stranded RNAs are then incorporated into RISCs, the same complexes that can carry miRNAs. The RISC discards one strand of the siRNA duplex and uses the remaining single-stranded RNA to locate a complementary foreign RNA molecule (Figure 8–27). This target RNA molecule is then rapidly degraded, leaving the RISC free to search out more of the same foreign RNA molecules.

RNAi is found in a wide variety of organisms, including single-celled fungi, plants, and worms, indicating that it is evolutionarily ancient. In some organisms, including plants, the RNAi activity can spread from tissue to tissue by the movement of RNA between cells. This RNA transfer allows the entire plant to become resistant to a virus after only a few of its cells have been infected. In a broad sense, the RNAi response resembles certain aspects of the human immune system. In both cases, an infectious organism elicits the production of ‘attack’ molecules (either siRNAs or antibodies) that are custom designed to inactivate the invader and thereby protect the host.

### Scientists Can Use RNA Interference to Turn Off Genes

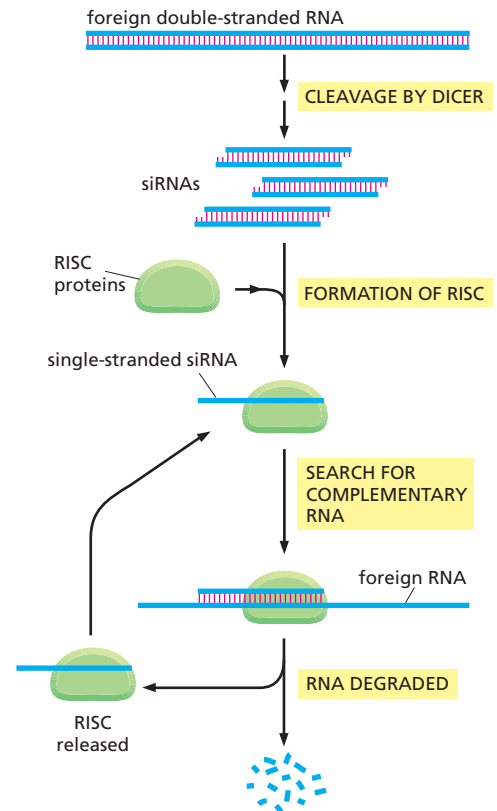
The discovery of miRNAs, siRNAs, and the mechanism of RNAi has been greeted with great enthusiasm. In a practical sense, RNAi has become

**Figure 8–27 siRNAs destroy foreign RNAs.** Double-stranded RNAs from a virus or transposable genetic element are first cleaved by a nuclease called Dicer. The resulting double-stranded fragments are incorporated into RISCs, which discard one strand of the duplex and use the other strand to locate and destroy complementary RNAs. This mechanism forms the basis for RNA interference (RNAi).

a powerful experimental tool that allows scientists to inactivate almost any gene in cultured cells or, in some cases, a whole plant or animal. We discuss how this method is being used to help determine the function of individual genes in Chapter 10.

At the same time, RNAi shows real potential as a powerful new approach for treating human disease. Because many human disorders result from the inappropriate expression of genes, the ability to turn these genes off by introducing complementary siRNA molecules holds great medical promise.

Finally, the discovery that RNAs play such a key role in controlling gene expression expands our understanding of the types of regulatory networks that cells have at their command. One of the great challenges of biology in this century will be to determine how these networks cooperate to specify the development of complex organisms, including ourselves.



## ESSENTIAL CONCEPTS

- A typical eucaryotic cell expresses only a fraction of its genes, and the distinct types of cells in multicellular organisms arise because different sets of genes are expressed as cells differentiate.
- Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of transcription is the most important point of control.
- The transcription of individual genes is switched on and off in cells by transcription regulators. These act by binding to short stretches of DNA called regulatory DNA sequences.
- Although each transcription regulator has unique features, most bind to DNA using one of a small number of structural motifs. The precise amino acid sequence that is folded into the DNA-binding motif determines the particular DNA sequence that is recognized.
- In bacteria, transcription regulators usually bind to regulatory DNA sequences close to where RNA polymerase binds. They can either activate or repress transcription of the gene. In eucaryotes, regulatory DNA sequences are often separated from the promoter by many thousands of nucleotide pairs.
- Eucaryotic transcription regulators act in two fundamental ways: (1) they can directly affect the assembly process of RNA polymerase and the general transcription factors at the promoter, and (2) they can locally modify the chromatin structure of promoter regions.
- In eucaryotes, the expression of a gene is generally controlled by a combination of transcription regulators.
- In multicellular plants and animals, the production of different transcription regulators in different cell types ensures the expression of only those genes appropriate to the particular type of cell.
- Cells in multicellular organisms have mechanisms that enable their progeny to 'remember' what type of cell they should be.
- A single transcription regulator, if expressed in the appropriate precursor cell, can trigger the formation of a specialized cell type or even an entire organ.

- Cells can also regulate gene expression by controlling events that occur after transcription has begun. Many of these mechanisms rely on RNA molecules that can influence their own transcription or translation.
- MicroRNAs (miRNAs) control gene expression by base-pairing with specific mRNAs and regulating their stability and their translation.
- Cells have a defense mechanism for destroying ‘foreign’ double-stranded RNAs, many of which are produced by viruses. Scientists can take advantage of this mechanism, called RNA interference, to inactivate genes of interest by simply injecting cells with double-stranded RNAs that target the mRNAs produced by those genes.

## KEY TERMS

activator  
combinatorial control  
differentiation  
DNA methylation  
epigenetic inheritance  
gene expression  
microRNA (miRNA)  
positive feedback loop

post-transcriptional control  
regulatory DNA sequence  
reporter gene  
repressor  
riboswitch  
RNA interference (RNAi)  
small interfering RNA (siRNA)  
transcription regulator

## QUESTIONS

## QUESTION 8-4

A virus that grows in bacteria (bacterial viruses are called bacteriophages) can replicate in one of two ways. In the prophage state, the viral DNA is inserted into the bacterial chromosome and is copied along with the bacterial genome each time the cell divides. In the lytic state, the viral DNA is released from the bacterial chromosome and replicates many times in the cell. This viral DNA then produces viral coat proteins that together with the replicated viral DNA form many new virus particles that burst out of the bacterial cell. These two forms of growth are controlled by two transcription regulators, called *cl* (“c one”) and *Cro*, that are encoded by the virus. In the prophage state, *cl* is expressed; in the lytic state, *Cro* is expressed. In addition to regulating the expression of other genes, *cl* represses the *Cro* gene, and *Cro* represses the *cl* gene (Figure Q8-4). When bacteria containing a phage in the prophage state are briefly irradiated with UV light, *cl* protein is degraded.

- What will happen next?
- Will the change in (A) be reversed when the UV light is switched off?
- Why might this response to UV light have evolved?

## QUESTION 8-5

(True/False) When the nucleus of a fully differentiated carrot cell is injected into a frog egg whose nucleus has been removed, the injected donor nucleus is capable of programming the recipient egg to produce a normal carrot. Explain your answer.

## QUESTION 8-6

Which of the following statements are correct? Explain your answers.

- In bacteria, but not in eucaryotes, most mRNAs encode more than one protein.
- Most DNA-binding proteins bind to the major groove of the double helix.

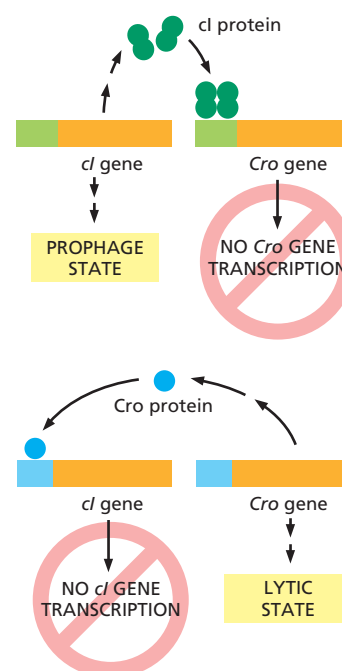


Figure Q8-4

C. Of the major control points in gene expression (transcription, RNA processing, RNA transport, translation, and control of a protein's activity), transcription initiation is one of the most common.

D. The zinc atoms in DNA-binding proteins that contain zinc finger domains contribute to the binding specificity through sequence-specific interactions that they form with the bases.

### QUESTION 8-7

Your task in the laboratory of Professor Quasimodo is to determine how far an enhancer (a binding site for an activator protein) could be moved from the promoter of the *straightspine* gene and still activate transcription. You systematically vary the number of nucleotide pairs between these two sites and then determine the amount of transcription by measuring the production of Straightspine mRNA. At first glance, your data look confusing (Figure Q8-7). What would you have expected for the results of this experiment? Can you save your reputation and explain these results to Professor Quasimodo?

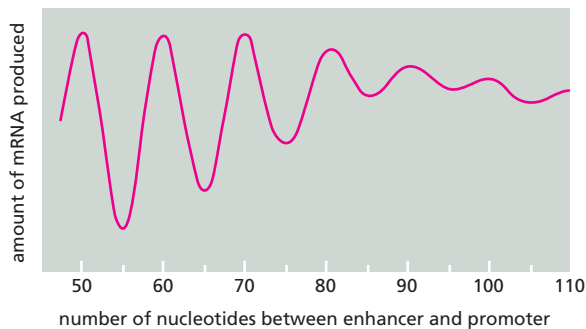


Figure Q8-7

### QUESTION 8-8

Many transcription regulators form dimers of identical subunits. Why is this advantageous? Describe three structural motifs that are often used to contact DNA. What are the particular features that suit them for this purpose?

### QUESTION 8-9

The  $\lambda$  repressor binds as a dimer to critical sites on the  $\lambda$  genome to repress the virus's lytic genes. This is necessary to maintain the prophage (integrated) state. Each molecule of the repressor consists of an N-terminal DNA-binding domain and a C-terminal dimerization domain (Figure Q8-9). Upon induction (for example, by irradiation

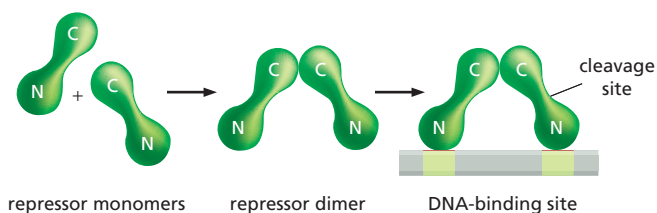


Figure Q8-9

with UV light), the genes for lytic growth are expressed,  $\lambda$  progeny are produced, and the bacterial cell is lysed (see Question 8-4). Induction is initiated by cleavage of the  $\lambda$  repressor at a site between the DNA-binding domain and the dimerization domain, which causes the repressor to dissociate from the DNA. In the absence of bound repressor, RNA polymerase binds and initiates lytic growth. Given that the number (concentration) of DNA-binding domains is unchanged by cleavage of the repressor, why do you suppose its cleavage results in its dissociation from the DNA?

### QUESTION 8-10

The enzymes for arginine biosynthesis are located at several positions around the genome of *E. coli*, and they are regulated coordinately by a transcription regulator encoded by the *ArgR* gene. The activity of ArgR is modulated by arginine. Upon binding arginine, ArgR alters its conformation, dramatically changing its affinity for the DNA sequences in the promoters of the genes for the arginine biosynthetic enzymes. Given that ArgR is a repressor protein, would you expect that ArgR would bind more tightly or less tightly to the DNA sequences when arginine is abundant? If ArgR functioned instead as an activator protein, would you expect the binding of arginine to increase or to decrease its affinity for its regulatory DNA sequences? Explain your answers.

### QUESTION 8-11

When enhancers were initially found to influence transcription many thousands of nucleotide pairs from the promoters they control, two principal models were invoked to explain this action at a distance. In the 'DNA looping' model, direct interactions between proteins bound at enhancers and promoters were proposed to stimulate transcription initiation. In the 'scanning' or 'entry-site' model, RNA polymerase (or another component of the transcription machinery) was proposed to bind at the enhancer and then scan along the DNA until it reached the promoter. These two models were tested using an enhancer on one piece of DNA and a  $\beta$ -globin gene and promoter on a separate piece of DNA (Figure Q8-11). The  $\beta$ -globin gene was not expressed from the mixture of pieces. However, when the two segments of DNA were joined via a protein linker, the  $\beta$ -globin gene was expressed.

Does this experiment distinguish between the DNA looping model and the scanning model? Explain your answer.

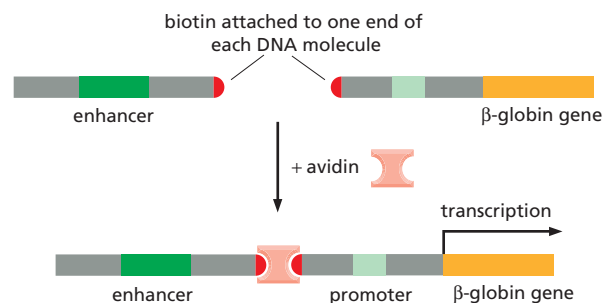


Figure Q8-11

## QUESTION 8-12

Differentiated cells of an organism contain the same genes. (Among the few exceptions to this rule are the cells of the mammalian immune system, in which the formation of specialized cells is based on limited rearrangements of the genome.) Describe an experiment that substantiates the first sentence of this question, and explain why it does.

## QUESTION 8-13

Figure 8-19 shows a simple scheme by which three transcription regulators are used during development to create eight different cell types. How many cell types could you create, using the same rules, with four different transcription regulators? As described in the text, MyoD is a transcription regulator that by itself is sufficient to induce muscle-specific gene expression in fibroblasts. How does this observation fit the scheme in Figure 8-19?

## QUESTION 8-14

Imagine the two situations shown in Figure Q8-14. In cell I, a transient signal induces the synthesis of protein A, which is a gene activator that turns on many genes including its own. In cell II, a transient signal induces the synthesis of protein R, which is a gene repressor that turns off many genes including its own. In which, if either, of these situations will the descendants of the original cell “remember” that the progenitor cell had experienced the transient signal? Explain your reasoning.

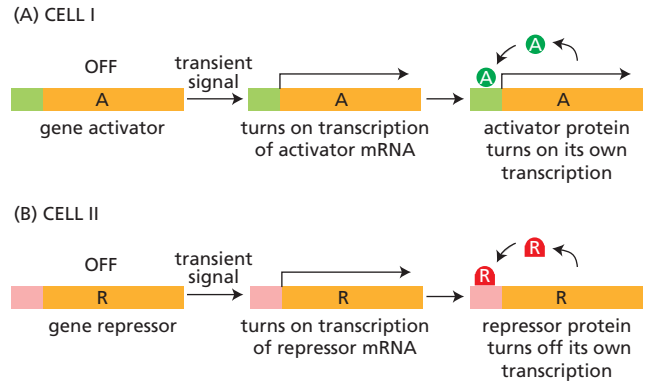


Figure Q8-14

## QUESTION 8-15

Discuss the following argument: “If the expression of every gene depends on a set of transcription regulators, then the expression of these regulators must also depend on the expression of other regulators, and their expression must depend on the expression of still other regulators, and so on. Cells would therefore need an infinite number of genes, most of which would code for transcription regulators.” How does the cell get by without having to achieve the impossible?

### ANSWER 8-1

A. Transcription of the tryptophan operon would no longer be regulated by the absence or presence of tryptophan; the enzymes would be permanently on in scenarios (1) and (2) and permanently off in scenario (3).

B. In scenarios (1) and (2), the normal tryptophan repressor molecules would completely restore the regulation of the tryptophan biosynthesis enzymes. In contrast, expression of the normal protein would have no effect in scenario 3, because the tryptophan operator would remain permanently occupied by the mutant protein.

### ANSWER 8-2

Contacts can form between the protein and the edges of the base pairs that are exposed in the major groove of the DNA (Figure A8-2). The bonds responsible for sequence-specific contacts are hydrogen bonds and hydrophobic interactions with the methyl group on thymine. Note that the arrangement of hydrogen-bond donors and hydrogen-bond acceptors of a T-A pair is different from that of a

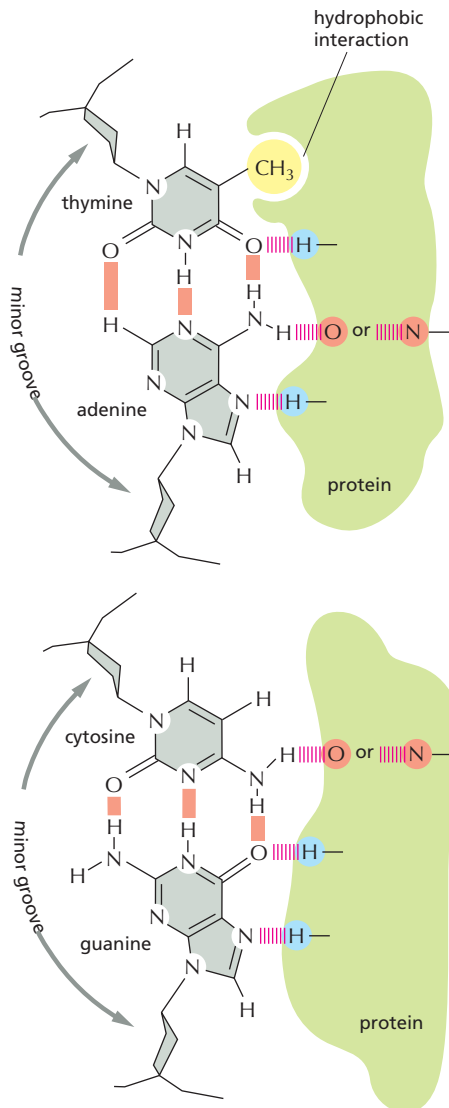


Figure A8-2

C-G pair. Similarly, the arrangement of hydrogen-bond donors and hydrogen-bond acceptors of A-T and G-C pairs would be different from one another and from the two pairs shown in the figure. In addition to the contacts shown in the figure, electrostatic attractions between the positively charged amino acid side chains of the protein and the negatively charged phosphate groups in the DNA backbone usually stabilize DNA-protein interactions.

### ANSWER 8-3

Bending proteins can help to bring distant DNA regions together that normally would contact each other only inefficiently (Figure A8-3). Such proteins are found in both prokaryotes and eukaryotes and are involved in many examples of transcriptional regulation.

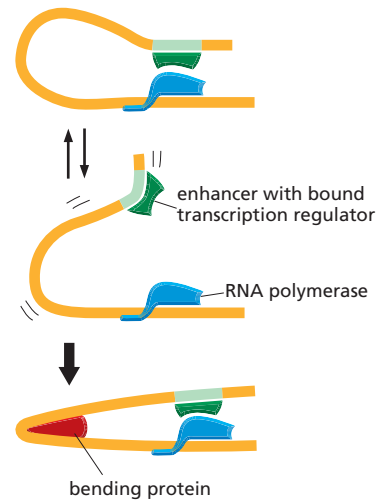


Figure A8-3

### ANSWER 8-4

A. UV light throws the switch from the prophage to the lytic state: when  $cI$  protein is destroyed, Cro is made and turns off the further production of  $cI$ . The virus produces coat proteins, and new virus particles are made.

B. When the UV light is switched off, the virus remains in the lytic state. Thus,  $cI$  and Cro form a gene regulatory switch that 'memorizes' its previous setting.

C. This switch makes sense in the viral life cycle: UV light tends to damage the bacterial DNA (see Figure 6-24), thereby rendering the bacterium an unreliable host for the virus. A prophage will therefore switch to the lytic state and leave the 'sinking ship' in search for new host cells to infect.

### ANSWER 8-5

False. Carrots can be grown from single carrot cells and tadpoles can be produced by injecting differentiated frog nuclei into frog eggs. But carrots cannot be produced from frog eggs, no matter what.

### ANSWER 8–6

A. True. Prokaryotic mRNAs are often transcripts of entire operons. Ribosomes can initiate translation at internal AUG start sites of these “polycistronic” mRNAs (see Figures 7–36 and 8–6).

B. True. The major groove of double-stranded DNA is sufficiently wide to allow a protein surface, such as one face of an  $\alpha$  helix, access to the base pairs.

C. True. It is advantageous to exert control at the earliest possible point in a pathway. This conserves metabolic energy because unnecessary products are not made in the first place.

D. False. The zinc atoms in zinc finger domains are required for the correct folding of the protein domain; they are internal to these domains and do not contact the DNA.

### ANSWER 8–7

From our knowledge of enhancers, one would expect their function to be relatively independent of their distance from the promoter—possibly weakening as this distance increases. The surprising feature of the data (which have been adapted from an actual experiment) is the periodicity: the enhancer is maximally active at certain distances from the promoter (50, 60, or 70 nucleotides), but almost inactive at intermediate distances (55 or 65 nucleotides). The periodicity of 10 suggests that the mystery can be explained by considering the structure of double-helical DNA, which has 10 base pairs per turn. Thus, placing an enhancer on the side of the DNA opposite to that of the promoter (Figure A8–7) would make it more difficult for the activator that binds to it to interact with the proteins bound at the promoter. At longer distances, there is more DNA to absorb the twist, and the effect is diminished.

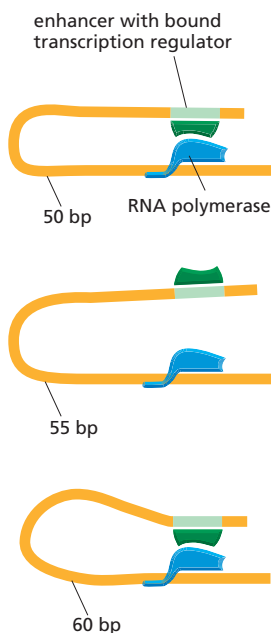


Figure A8–7

### ANSWER 8–8

Two advantages of dimeric DNA-binding proteins are (1) that dimer formation greatly increases the specificity and the strength of a protein–DNA interaction by doubling the number of protein–DNA contacts, and (2) that different subunits can be combined in different combinations to increase the number of DNA-binding specificities available to the cell. Three of the most common protein domains that bind DNA are leucine zippers, homeodomains, and zinc fingers. Each provides a particularly stable fold in the polypeptide chain that positions an  $\alpha$  helix so it can be inserted into the major groove of the DNA helix and contact the sides of the base pairs (see Figure 8–5).

### ANSWER 8–9

The affinity of the dimeric  $\lambda$  repressor for its binding site is the sum of the interactions made by each of the two DNA-binding domains. A single DNA-binding domain can make only half the contacts and provide just half the binding energy compared with the dimer. Thus, although the concentration of binding domains is unchanged, they are no longer coupled, and their individual affinities for DNA are sufficiently weak that they cannot remain bound. As a result, the genes for lytic growth are turned on.

### ANSWER 8–10

The function of these *Arg* genes is to synthesize arginine. When arginine is abundant, expression of the biosynthetic genes should be turned off. If ArgR acts as a gene repressor (which it does in reality), then binding of arginine should increase its affinity for its regulatory sites, allowing it to bind and shut off gene expression. If ArgR acted as a gene activator instead, then the binding of arginine would be predicted to reduce its affinity for its regulatory DNA, preventing its binding and thereby shutting off expression of the *Arg* genes.

### ANSWER 8–11

The results of this experiment favor DNA looping, which should not be affected by the protein bridge (so long as it allowed the DNA to bend, which it does). The scanning or entry site model, however, is predicted to be affected by the nature of the linkage between the enhancer and the promoter. If the proteins enter at the enhancer and scan to the promoter, they would have to traverse the protein linkage. If such proteins are geared to scan on DNA, they would likely have difficulty scanning across such a barrier.

### ANSWER 8–12

The most definitive result is one showing that a single differentiated cell taken from a specialized tissue can re-create a whole organism. This proves that the cell must contain all the information required to produce a whole organism, including all of its specialized cell types. Experiments of this type are summarized in Figure 8–2.

### ANSWER 8–13

You could create 16 different cell types with 4 different transcription regulators (all the 8 cell types shown in Figure 8–19, plus another 8 created by adding an additional transcription regulator). MyoD by itself is sufficient to

induce muscle-specific gene expression only in certain cell types, such as some kinds of fibroblasts. The action of MyoD is therefore consistent with the model shown in Figure 8–19: if muscle cells were specified, for example, by the combination of transcription regulators 1, 3, and MyoD, then the addition of MyoD would convert only two of the cell types of Figure 8–19 (cells F and H) to muscle.

#### ANSWER 8–14

The induction of a gene activator protein that stimulates its own synthesis can create a positive feedback loop that can produce cell memory. The continued self-stimulated synthesis of activator A can in principle last for many cell generations, serving as a constant reminder of an event that took place in the past. By contrast, the induction of a gene repressor that inhibits its own synthesis creates a negative feedback loop that guarantees only a transient response to the transient stimulus. Because repressor R shuts off its own synthesis, the cell will quickly return to the state that existed before the signal.

#### ANSWER 8–15

Many transcription regulators are continually made in the cell; that is, their expression is constitutive and the activity of the protein is controlled by signals from inside or outside the cell (e.g., the availability of nutrients, as for the tryptophan repressor, or by hormones, as for the glucocorticoid receptor), thereby adjusting the transcriptional program to the physiological needs of the cell. Moreover, a given transcription regulator usually controls the expression of many different genes. Transcription regulators are often used in various combinations and can affect each other's activity, thereby further increasing the possibilities for regulation with a limited set of proteins. Nevertheless, the cell devotes a large fraction of its genome to the control of transcription: an estimated 10% of all genes in eucaryotic cells code for transcription regulators.