



chapter 28

REGULATION OF GENE EXPRESSION

28.1 Principles of Gene Regulation 1082

28.2 Regulation of Gene Expression in Prokaryotes 1092

28.3 Regulation of Gene Expression in Eukaryotes 1102

The fundamental problem of chemical physiology and of embryology is to understand why tissue cells do not all express, all the time, all the potentialities inherent in their genome.

—François Jacob and Jacques Monod,
article in *Journal of Molecular Biology*, 1961

Of the 4,000 or so genes in the typical bacterial genome, or the perhaps 35,000 genes in the human genome, only a fraction are expressed in a cell at any given time. Some gene products are present in very large amounts: the elongation factors required for protein synthesis, for example, are among the most abundant proteins in bacteria, and ribulose 1,5-bisphosphate carboxylase/oxygenase (rubisco) of plants and photosynthetic bacteria is, as far as we know, the most abundant enzyme in the biosphere. Other gene products occur in much smaller amounts; for instance, a cell may contain only a few molecules of the enzymes that repair rare DNA lesions. Requirements for some gene products change over time. The need for enzymes in certain metabolic pathways may wax and wane as food sources change or are depleted. During development of a multicellular organism, some proteins that influence cellular differentiation are present for just a brief time in only a few cells. Specialization of cellular function can dramatically affect the need for various gene products; an example is the uniquely high concentration of a single

protein—hemoglobin—in erythrocytes. Given the high cost of protein synthesis, regulation of gene expression is essential to making optimal use of available energy.

The cellular concentration of a protein is determined by a delicate balance of at least seven processes, each having several potential points of regulation:

1. Synthesis of the primary RNA transcript (transcription)
2. Posttranscriptional modification of mRNA
3. Messenger RNA degradation
4. Protein synthesis (translation)
5. Posttranslational modification of proteins
6. Protein targeting and transport
7. Protein degradation

These processes are summarized in Figure 28–1. We have examined several of these mechanisms in previous chapters. Posttranscriptional modification of mRNA, by processes such as alternative splicing patterns (see Fig. 26–19b) or RNA editing (see Box 27–1), can affect which proteins are produced from an mRNA transcript and in what amounts. A variety of nucleotide sequences in an mRNA can affect the rate of its degradation (p. 1020). Many factors affect the rate at which an mRNA is translated into a protein, as well as the posttranslational modification, targeting, and eventual degradation of that protein (Chapter 27).

This chapter focuses primarily on the regulation of transcription initiation, although aspects of posttranscriptional and translational regulation are also described. Of the regulatory processes illustrated in Figure 28–1, those operating at the level of transcription initiation are the best documented and probably the most

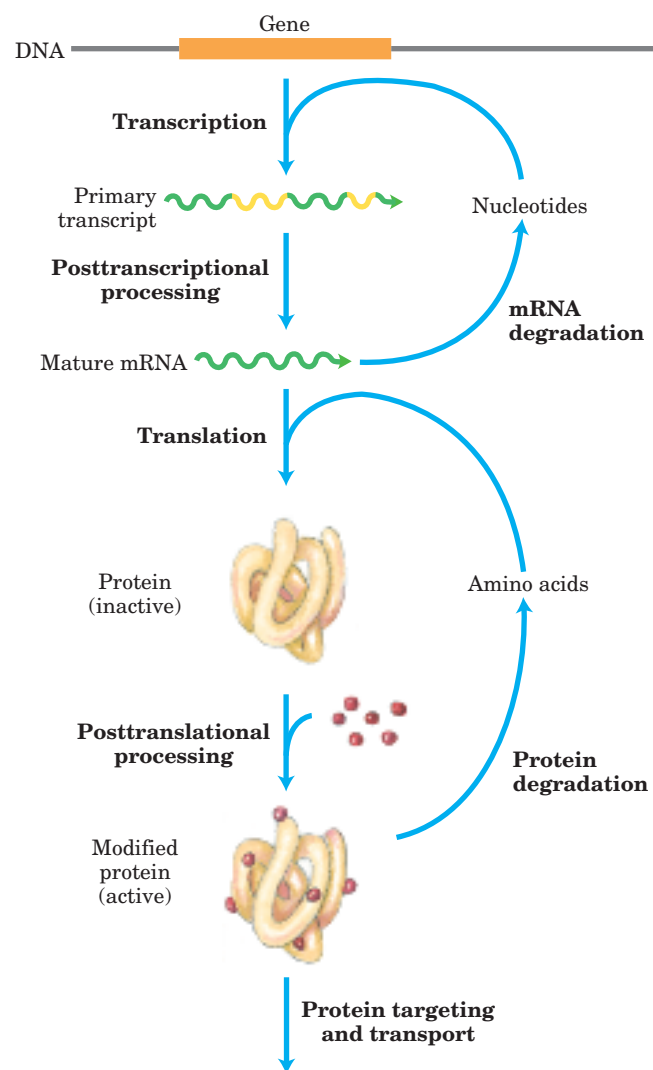


FIGURE 28-1 Seven processes that affect the steady-state concentration of a protein. Each process has several potential points of regulation.

common. As in all biochemical processes, an efficient place for regulation is at the beginning of the pathway. Because synthesis of informational macromolecules is so extraordinarily expensive in terms of energy, elaborate mechanisms have evolved to regulate the process. Researchers continue to discover complex and sometimes surprising regulatory mechanisms. Increasingly, posttranscriptional and translational regulation are proving to be among the more important of these processes, especially in eukaryotes. In fact, the regulatory processes themselves can involve a considerable investment of chemical energy.

Control of transcription initiation permits the synchronized regulation of multiple genes encoding products with interdependent activities. For example, when their DNA is heavily damaged, bacterial cells require a coordinated increase in the levels of the many DNA repair enzymes. And perhaps the most sophisticated form

of coordination occurs in the complex regulatory circuits that guide the development of multicellular eukaryotes, which can involve many types of regulatory mechanisms.

We begin by examining the interactions between proteins and DNA that are the key to transcriptional regulation. We next discuss the specific proteins that influence the expression of specific genes, first in prokaryotic and then in eukaryotic cells. Information about posttranscriptional and translational regulation is included in the discussion, where relevant, to provide a more complete overview of the rich complexity of regulatory mechanisms.

28.1 Principles of Gene Regulation

Genes for products that are required at all times, such as those for the enzymes of central metabolic pathways, are expressed at a more or less constant level in virtually every cell of a species or organism. Such genes are often referred to as **housekeeping genes**. Unvarying expression of a gene is called **constitutive gene expression**.

For other gene products, cellular levels rise and fall in response to molecular signals; this is **regulated gene expression**. Gene products that increase in concentration under particular molecular circumstances are referred to as **inducible**; the process of increasing their expression is **induction**. The expression of many of the genes encoding DNA repair enzymes, for example, is induced by high levels of DNA damage. Conversely, gene products that decrease in concentration in response to a molecular signal are referred to as **repressible**, and the process is called **repression**. For example, in bacteria, ample supplies of tryptophan lead to repression of the genes for the enzymes that catalyze tryptophan biosynthesis.

Transcription is mediated and regulated by protein-DNA interactions, especially those involving the protein components of RNA polymerase (Chapter 26). We first consider how the activity of RNA polymerase is regulated, and proceed to a general description of the proteins participating in this process. We then examine the molecular basis for the recognition of specific DNA sequences by DNA-binding proteins.

RNA Polymerase Binds to DNA at Promoters

RNA polymerases bind to DNA and initiate transcription at promoters (see Fig. 26-5), sites generally found near points at which RNA synthesis begins on the DNA template. The regulation of transcription initiation often entails changes in how RNA polymerase interacts with a promoter.

The nucleotide sequences of promoters vary considerably, affecting the binding affinity of RNA polymerases and thus the frequency of transcription initiation. Some

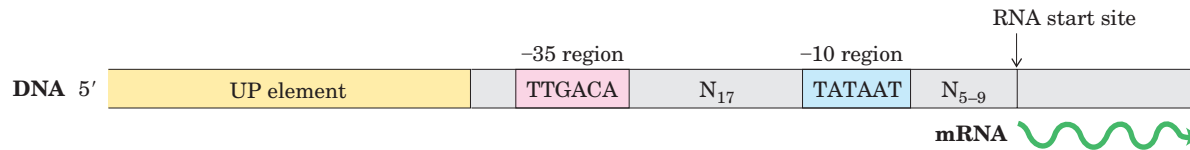


FIGURE 28-2 Consensus sequence for many *E. coli* promoters. Most base substitutions in the -10 and -35 regions have a negative effect on promoter function. Some promoters also include the UP (upstream promoter) element (see Fig. 26-5). By convention, DNA sequences

are shown as they exist in the nontemplate strand, with the 5' terminus on the left. Nucleotides are numbered from the transcription start site, with positive numbers to the right (in the direction of transcription) and negative numbers to the left. N indicates any nucleotide.

Escherichia coli genes are transcribed once per second, others less than once per cell generation. Much of this variation is due to differences in promoter sequence. In the absence of regulatory proteins, differences in promoter sequences may affect the frequency of transcription initiation by a factor of 1,000 or more. Most *E. coli* promoters have a sequence close to a consensus (Fig. 28-2). Mutations that result in a shift away from the consensus sequence usually decrease promoter function; conversely, mutations toward consensus usually enhance promoter function.

Although housekeeping genes are expressed constitutively, the cellular concentrations of the proteins they encode vary widely. For these genes, the RNA polymerase–promoter interaction strongly influences the rate of transcription initiation; differences in promoter sequence allow the cell to synthesize the appropriate level of each housekeeping gene product.

The basal rate of transcription initiation at the promoters of nonhousekeeping genes is also determined by the promoter sequence, but expression of these genes is further modulated by regulatory proteins. Many of these proteins work by enhancing or interfering with the interaction between RNA polymerase and the promoter.

The sequences of eukaryotic promoters are more variable than their prokaryotic counterparts (see Fig. 26-8). The three eukaryotic RNA polymerases usually require an array of general transcription factors in order to bind to a promoter. Yet, as with prokaryotic gene expression, the basal level of transcription is determined by the effect of promoter sequences on the function of RNA polymerase and its associated transcription factors.

Transcription Initiation Is Regulated by Proteins That Bind to or near Promoters

At least three types of proteins regulate transcription initiation by RNA polymerase: **specificity factors** alter the specificity of RNA polymerase for a given promoter or set of promoters; **repressors** impede access of RNA polymerase to the promoter; and **activators** enhance the RNA polymerase–promoter interaction.

We introduced prokaryotic specificity factors in Chapter 26 (see Fig. 26-5), although we did not refer to them by that name. The σ subunit of the *E. coli* RNA polymerase holoenzyme is a specificity factor that mediates promoter recognition and binding. Most *E. coli* promoters are recognized by a single σ subunit (M_r 70,000), σ^{70} . Under some conditions, some of the σ^{70} subunits are replaced by another specificity factor. One notable case arises when the bacteria are subjected to heat stress, leading to the replacement of σ^{70} by σ^{32} (M_r 32,000). When bound to σ^{32} , RNA polymerase is directed to a specialized set of promoters with a different consensus sequence (Fig. 28-3). These promoters control the expression of a set of genes that encode the heat-shock response proteins. Thus, through changes in the binding affinity of the polymerase that direct it to different promoters, a set of genes involved in related processes is coordinately regulated. In eukaryotic cells, some of the general transcription factors, in particular the TATA-binding protein (TBP; see Fig. 26-8), may be considered specificity factors.

Repressors bind to specific sites on the DNA. In prokaryotic cells, such binding sites, called **operators**, are generally near a promoter. RNA polymerase binding,

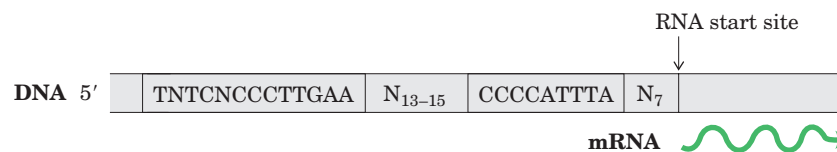


FIGURE 28-3 Consensus sequence for promoters that regulate expression of the *E. coli* heat-shock genes. This system responds to temperature increases as well as some other environmental stresses, resulting in the induction of a set of proteins. Binding of RNA polymerase to heat-shock promoters is mediated by a specialized σ subunit of the polymerase, σ^{32} , which replaces σ^{70} in the RNA polymerase initiation complex.

or its movement along the DNA after binding, is blocked when the repressor is present. Regulation by means of a repressor protein that blocks transcription is referred to as **negative regulation**. Repressor binding to DNA is regulated by a molecular signal (or **effector**), usually a small molecule or a protein, that binds to the repressor and causes a conformational change. The interaction between repressor and signal molecule either increases or decreases transcription. In some cases, the conformational change results in dissociation of a DNA-bound repressor from the operator (Fig. 28–4a). Transcription initiation can then proceed unhindered. In other cases, interaction between an inactive repressor and the signal molecule causes the repressor to bind to the operator (Fig. 28–4b). In eukaryotic cells, the binding site for a repressor may be some distance from the promoter; binding has the same effect as in bacterial cells: inhibit-

ing the assembly or activity of a transcription complex at the promoter.

Activators provide a molecular counterpoint to repressors; they bind to DNA and *enhance* the activity of RNA polymerase at a promoter; this is **positive regulation**. Activator binding sites are often adjacent to promoters that are bound weakly or not at all by RNA polymerase alone, such that little transcription occurs in the absence of the activator. Some eukaryotic activators bind to DNA sites, called **enhancers**, that are quite distant from the promoter, affecting the rate of transcription at a promoter that may be located thousands of base pairs away. Some activators are normally bound to DNA, enhancing transcription until dissociation of the activator is triggered by the binding of a signal molecule (Fig. 28–4c). In other cases the activator binds to DNA only after interaction with a signal molecule

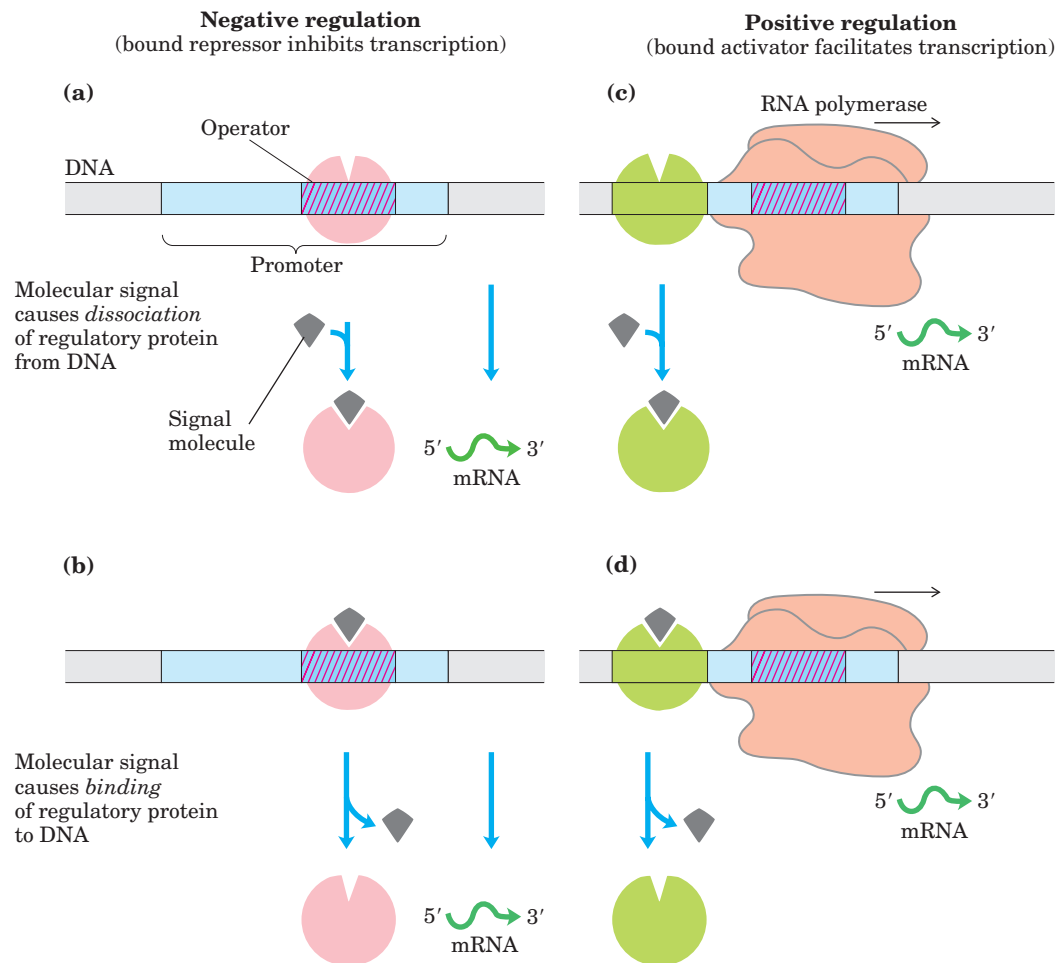


FIGURE 28–4 Common patterns of regulation of transcription initiation. Two types of negative regulation are illustrated. (a) Repressor (pink) binds to the operator in the absence of the molecular signal; the external signal causes dissociation of the repressor to permit transcription. (b) Repressor binds in the presence of the signal; the repressor dissociates and transcription ensues when the signal is removed. Positive regulation is mediated by gene activators. Again, two types are shown. (c) Activator (green) binds in the absence of the mo-

lecular signal and transcription proceeds; when the signal is added, the activator dissociates and transcription is inhibited. (d) Activator binds in the presence of the signal; it dissociates only when the signal is removed. Note that “positive” and “negative” regulation refer to the type of regulatory protein involved: the bound protein either facilitates or inhibits transcription. In either case, addition of the molecular signal may increase or decrease transcription, depending on its effect on the regulatory protein.

(Fig. 28–4d). Signal molecules can therefore increase or decrease transcription, depending on how they affect the activator. Positive regulation is particularly common in eukaryotes, as we shall see.

Many Prokaryotic Genes Are Clustered and Regulated in Operons

Bacteria have a simple general mechanism for coordinating the regulation of genes encoding products that participate in a set of related processes: these genes are clustered on the chromosome and are transcribed together. Many prokaryotic mRNAs are polycistronic—multiple genes on a single transcript—and the single promoter that initiates transcription of the cluster is the site of regulation for expression of all the genes in the cluster. The gene cluster and promoter, plus additional sequences that function together in regulation, are called an **operon** (Fig. 28–5). Operons that include two to six genes transcribed as a unit are common; some operons contain 20 or more genes.

Many of the principles of prokaryotic gene expression were first defined by studies of lactose metabolism in *E. coli*, which can use lactose as its sole carbon source. In 1960, François Jacob and Jacques Monod published a short paper in the *Proceedings* of the French Academy of Sciences that described how two adjacent genes involved in lactose metabolism were coordinately regulated by a genetic element located at one end of the gene cluster. The genes were those for β -galactosidase, which cleaves lactose to galactose and glucose, and galactoside permease, which transports lactose into the cell (Fig. 28–6). The terms “operon” and “operator” were first introduced in this paper. With the operon model, gene regulation could, for the first time, be considered in molecular terms.

The *lac* Operon Is Subject to Negative Regulation

The lactose (*lac*) operon (Fig. 28–7a) includes the genes for β -galactosidase (*Z*), galactoside permease (*Y*), and thiogalactoside transacetylase (*A*). The last of these enzymes appears to modify toxic galactosides to facilitate their removal from the cell. Each of the three genes is preceded by a ribosome binding site (not shown in Fig. 28–7) that independently directs the translation

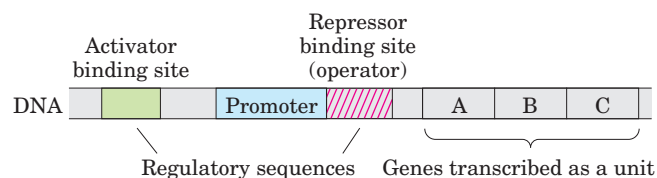


FIGURE 28–5 Representative prokaryotic operon. Genes A, B, and C are transcribed on one polycistronic mRNA. Typical regulatory sequences include binding sites for proteins that either activate or repress transcription from the promoter.

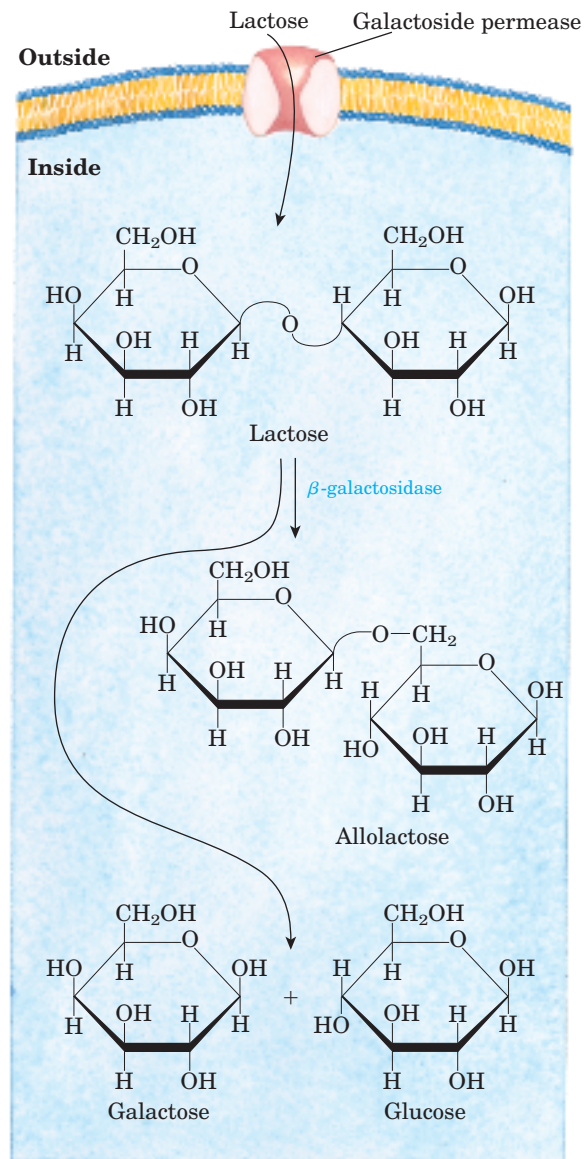


FIGURE 28–6 Lactose metabolism in *E. coli*. Uptake and metabolism of lactose require the activities of galactoside permease and β -galactosidase. Conversion of lactose to allolactose by transglycosylation is a minor reaction also catalyzed by β -galactosidase.



François Jacob



Jacques Monod, 1910–1976

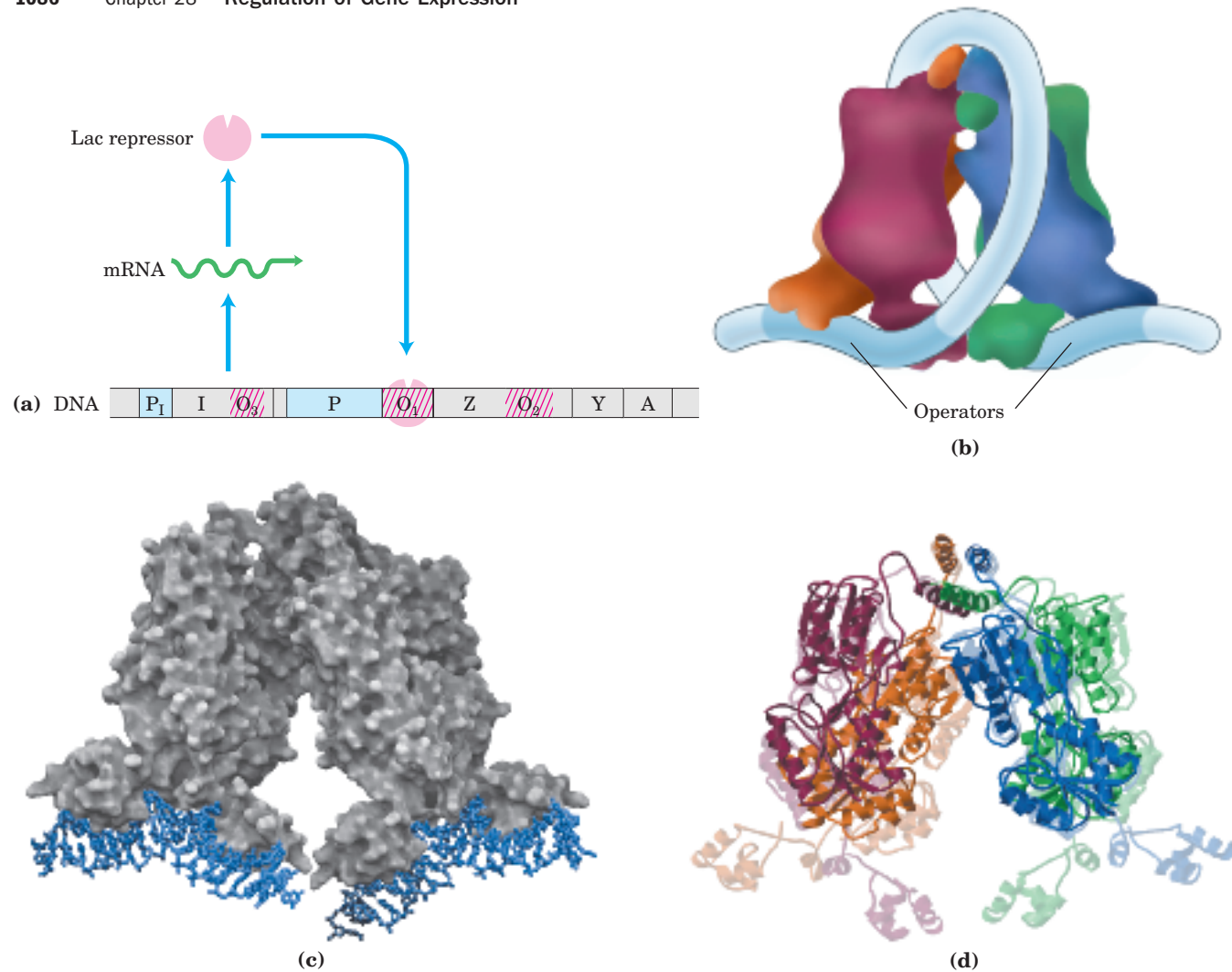


FIGURE 28-7 The *lac* operon. (a) The *lac* operon in the repressed state. The *I* gene encodes the Lac repressor. The *lac* Z, Y, and A genes encode β-galactosidase, galactoside permease, and thiogalactoside transacetylase, respectively. P is the promoter for the *lac* genes, and P_I is the promoter for the *I* gene. O₁ is the main operator for the *lac* operon; O₂ and O₃ are secondary operator sites of lesser affinity for the Lac repressor. (b) The Lac repressor binds to the main operator and O₂ or O₃, apparently forming a loop in the DNA that might wrap around the repressor as shown. (c) Lac repressor bound to DNA (derived from PDB ID 1LBG). This shows the protein (gray) bound to short,

discontinuous segments of DNA (blue). (d) Conformational change in the Lac repressor caused by binding of the artificial inducer isopropylthiogalactoside, IPTG (derived from PDB ID 1LBH and 1LBG). The structure of the tetrameric repressor is shown without IPTG bound (transparent image) and with IPTG bound (overlaid solid image; IPTG not shown). The DNA bound when IPTG is absent (transparent structure) is not shown. When IPTG is bound and DNA is not bound, the repressor's DNA-binding domains are too disordered to be defined in the crystal structure.

of that gene (Chapter 27). Regulation of the *lac* operon by the *lac* repressor protein (Lac) follows the pattern outlined in Figure 28-4a.

The study of *lac* operon mutants has revealed some details of the workings of the operon's regulatory system. In the absence of lactose, the *lac* operon genes are repressed. Mutations in the operator or in another gene, the *I* gene, result in constitutive synthesis of the gene products. When the *I* gene is defective, repression can be restored by introducing a functional *I* gene into the cell on another DNA molecule, demonstrating that the *I* gene encodes a diffusible molecule that causes gene repression. This molecule proved to be a protein, now

called the Lac repressor, a tetramer of identical monomers. The operator to which it binds most tightly (O₁) abuts the transcription start site (Fig. 28-7a). The *I* gene is transcribed from its own promoter (P_I) independent of the *lac* operon genes. The *lac* operon has two secondary binding sites for the Lac repressor. One (O₂) is centered near position +410, within the gene encoding β-galactosidase (Z); the other (O₃) is near position -90, within the *I* gene. To repress the operon, the Lac repressor appears to bind to both the main operator and one of the two secondary sites, with the intervening DNA looped out (Fig. 28-7b, c). Either binding arrangement blocks transcription initiation.

Despite this elaborate binding complex, repression is not absolute. Binding of the Lac repressor reduces the rate of transcription initiation by a factor of 10^3 . If the O_2 and O_3 sites are eliminated by deletion or mutation, the binding of repressor to O_1 alone reduces transcription by a factor of about 10^2 . Even in the repressed state, each cell has a few molecules of β -galactosidase and galactoside permease, presumably synthesized on the rare occasions when the repressor transiently dissociates from the operators. This basal level of transcription is essential to operon regulation.

When cells are provided with lactose, the *lac* operon is induced. An inducer (signal) molecule binds to a specific site on the Lac repressor, causing a conformational change (Fig. 28-7d) that results in dissociation of the repressor from the operator. The inducer in the *lac* operon system is not lactose itself but allolactose, an isomer of lactose (Fig. 28-6). After entry into the *E. coli* cell (via the few existing molecules of permease), lactose is converted to allolactose by one of the few existing β -galactosidase molecules. Release of the operator by Lac repressor, triggered as the repressor binds to allolactose, allows expression of the *lac* operon genes and leads to a 10^3 -fold increase in the concentration of β -galactosidase.

Several β -galactosides structurally related to allolactose are inducers of the *lac* operon but are not substrates for β -galactosidase; others are substrates but not inducers. One particularly effective and nonmetabolizable inducer of the *lac* operon that is often used experimentally is isopropylthiogalactoside (IPTG):

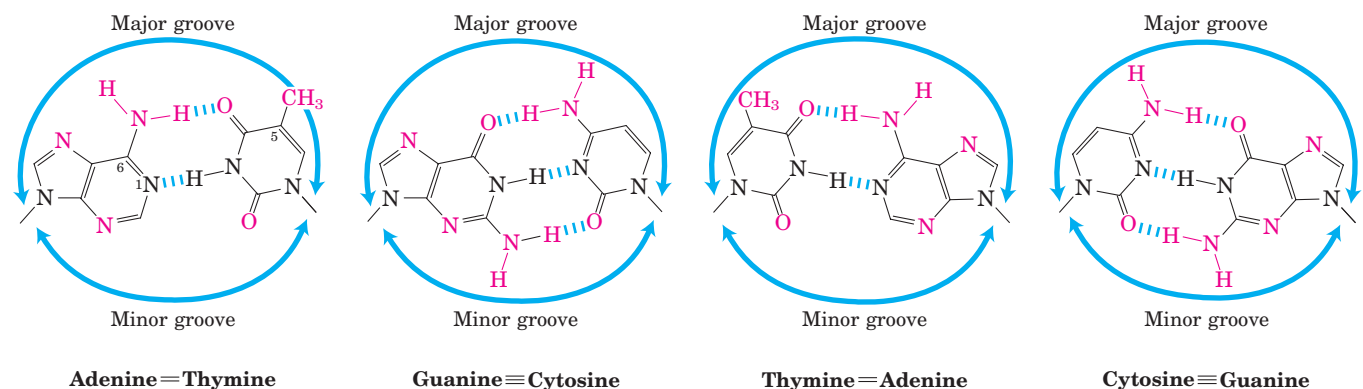
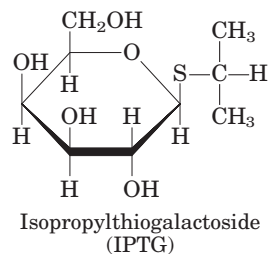


FIGURE 28-8 Groups in DNA available for protein binding. Shown here are functional groups on all four base pairs that are displayed in

An inducer that cannot be metabolized allows researchers to explore the physiological function of lactose as a carbon source for growth, separate from its function in the regulation of gene expression.

In addition to the multitude of operons now known in bacteria, a few polycistronic operons have been found in the cells of lower eukaryotes. In the cells of higher eukaryotes, however, almost all protein-encoding genes are transcribed separately.

The mechanisms by which operons are regulated can vary significantly from the simple model presented in Figure 28-7. Even the *lac* operon is more complex than indicated here, with an activator also contributing to the overall scheme, as we shall see in Section 28.2. Before any further discussion of the layers of regulation of gene expression, however, we examine the critical molecular interactions between DNA-binding proteins (such as repressors and activators) and the DNA sequences to which they bind.

Regulatory Proteins Have Discrete DNA-Binding Domains

Regulatory proteins generally bind to specific DNA sequences. Their affinity for these target sequences is roughly 10^4 to 10^6 times higher than their affinity for any other DNA sequences. Most regulatory proteins have discrete DNA-binding domains containing substructures that interact closely and specifically with the DNA. These binding domains usually include one or more of a relatively small group of recognizable and characteristic structural motifs.

To bind specifically to DNA sequences, regulatory proteins must recognize surface features on the DNA. Most of the chemical groups that differ among the four bases and thus permit discrimination between base pairs are hydrogen-bond donor and acceptor groups exposed in the major groove of DNA (Fig. 28-8), and most of the protein-DNA contacts that impart specificity are hydrogen bonds. A notable exception is the nonpolar surface

the major and minor grooves of DNA. Groups that can be used for base-pair recognition by proteins are shown in red.

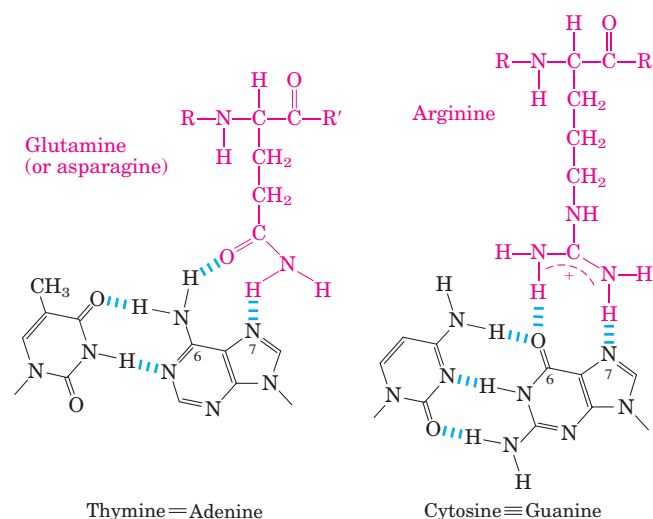


FIGURE 28-9 Two examples of specific amino acid–base pair interactions that have been observed in DNA–protein binding.

near C-5 of pyrimidines, where thymine is readily distinguished from cytosine by its protruding methyl group. Protein–DNA contacts are also possible in the minor groove of the DNA, but the hydrogen-bonding patterns here generally do not allow ready discrimination between base pairs.

Within regulatory proteins, the amino acid side chains most often hydrogen-bonding to bases in the DNA are those of Asn, Gln, Glu, Lys, and Arg residues. Is there a simple recognition code in which a particular amino acid always pairs with a particular base? The two hydrogen bonds that can form between Gln or Asn and the N^6 and N^7 positions of adenine cannot form with any other base. And an Arg residue can form two hydrogen bonds with N^7 and O^6 of guanine (Fig. 28-9). Examination of the structures of many DNA-binding proteins, however, has shown that a protein can recognize each base pair in more than one way, leading to the conclusion that there is no simple amino acid–base code. For some proteins, the Gln–adenine interaction can specify A=T base pairs, but in others a van der Waals pocket for the methyl group of thymine can recognize A=T base pairs. Researchers cannot yet examine the structure of a DNA-binding protein and infer the DNA sequence to which it binds.

To interact with bases in the major groove of DNA, a protein requires a relatively small structure that can stably protrude from the protein surface. The DNA-binding domains of regulatory proteins tend to be small (60 to 90 amino acid residues), and the structural motifs within these domains that are actually in contact with the DNA are smaller still. Many small proteins are unstable because of their limited capacity to form layers of structure to bury hydrophobic groups (p. 118). The DNA-binding motifs provide either a very compact stable structure or a way of allowing a segment of protein to protrude from the protein surface.

The DNA-binding sites for regulatory proteins are often inverted repeats of a short DNA sequence (a palindrome) at which multiple (usually two) subunits of a regulatory protein bind cooperatively. The Lac repressor is unusual in that it functions as a tetramer, with two dimers tethered together at the end distant from the DNA-binding sites (Fig. 28-7b). An *E. coli* cell normally contains about 20 tetramers of the Lac repressor. Each of the tethered dimers separately binds to a palindromic operator sequence, in contact with 17 bp of a 22 bp region in the *lac* operon (Fig. 28-10). And each of the tethered dimers can independently bind to an operator sequence, with one generally binding to O_1 and the other to O_2 or O_3 (as in Fig. 28-7b). The symmetry of the O_1 operator sequence corresponds to the twofold axis of symmetry of two paired Lac repressor subunits. The tetrameric Lac repressor binds to its operator sequences in vivo with an estimated dissociation constant of about 10^{-10} M. The repressor discriminates between the operators and other sequences by a factor of about 10^6 , so binding to these few base pairs among the 4.6 million or so of the *E. coli* chromosome is highly specific.

Several DNA-binding motifs have been described, but here we focus on two that play prominent roles in the binding of DNA by regulatory proteins: the **helix-turn-helix** and the **zinc finger**. We also consider a type of DNA-binding domain—the homeodomain—found in some eukaryotic proteins.

Helix-Turn-Helix This DNA-binding motif is crucial to the interaction of many prokaryotic regulatory proteins with DNA, and similar motifs occur in some eukaryotic regulatory proteins. The helix-turn-helix motif comprises about 20 amino acids in two short α -helical segments,

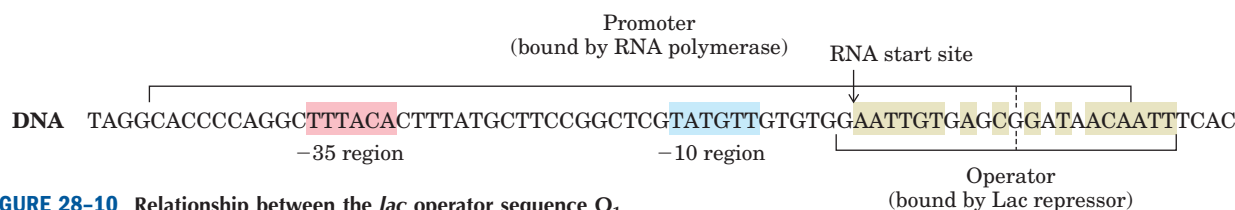


FIGURE 28-10 Relationship between the *lac* operator sequence O_1 and the *lac* promoter. The bases shaded beige exhibit twofold (palindromic) symmetry about the axis indicated by the dashed vertical line.

each seven to nine amino acid residues long, separated by a β turn (Fig. 28–11). This structure generally is not stable by itself; it is simply the reactive portion of a somewhat larger DNA-binding domain. One of the two α -helical segments is called the recognition helix, because it usually contains many of the amino acids that

interact with the DNA in a sequence-specific way. This α helix is stacked on other segments of the protein structure so that it protrudes from the protein surface. When bound to DNA, the recognition helix is positioned in or nearly in the major groove. The Lac repressor has this DNA-binding motif (Fig. 28–11).

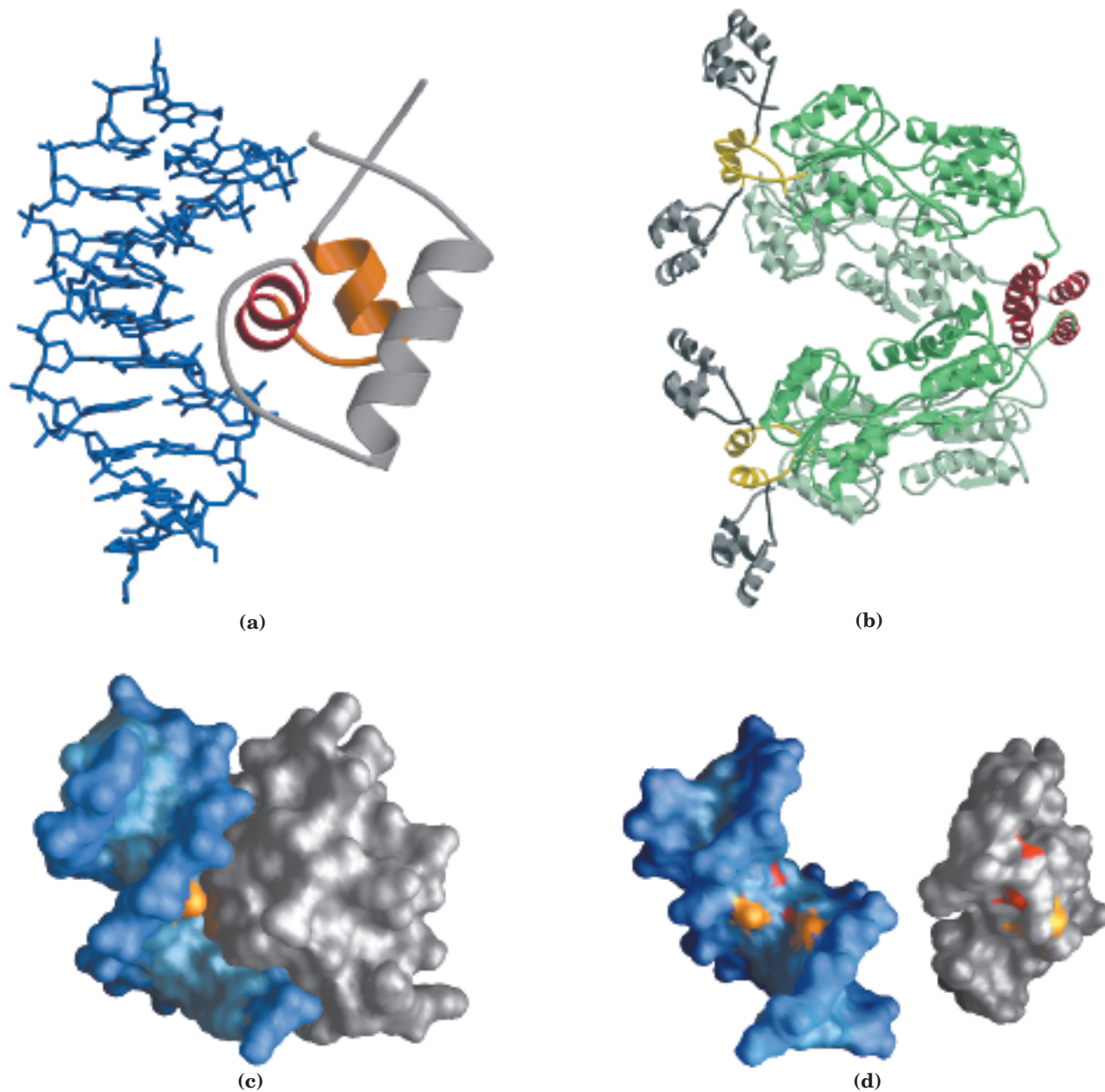


FIGURE 28–11 Helix-turn-helix. (a) DNA-binding domain of the Lac repressor (PDB ID 1LCC). The helix-turn-helix motif is shown in red and orange; the DNA recognition helix is red. (b) Entire Lac repressor (derived from PDB ID 1LBG). The DNA-binding domains are gray, and the α helices involved in tetramerization are red. The remainder of the protein (shades of green) has the binding sites for allolactose. The allolactose-binding domains are linked to the DNA-binding domains through linker helices (yellow). (c) Surface rendering of the

DNA-binding domain of the Lac repressor (gray) bound to DNA (blue). (d) The same DNA-binding domain as in (c), but separated from the DNA, with the binding interaction surfaces shown. Some groups on the protein and DNA that interact through hydrogen-bonding are shown in red; some groups that interact through hydrophobic interactions are in orange. This model shows only a few of the groups involved in sequence recognition. The complementary nature of the two surfaces is evident.

Zinc Finger In a zinc finger, about 30 amino acid residues form an elongated loop held together at the base by a single Zn^{2+} ion, which is coordinated to four of the residues (four Cys, or two Cys and two His). The zinc does not itself interact with DNA; rather, the coordination of zinc with the amino acid residues stabilizes this small structural motif. Several hydrophobic side chains in the core of the structure also lend stability. Figure 28–12 shows the interaction between DNA and three zinc fingers of a single polypeptide from the mouse regulatory protein Zif268.

Many eukaryotic DNA-binding proteins contain zinc fingers. The interaction of a single zinc finger with DNA is typically weak, and many DNA-binding proteins, like Zif268, have multiple zinc fingers that substantially enhance binding by interacting simultaneously with the DNA. One DNA-binding protein of the frog *Xenopus* has 37 zinc fingers. There are few known examples of the zinc finger motif in prokaryotic proteins.

The precise manner in which proteins with zinc fingers bind to DNA differs from one protein to the next. Some zinc fingers contain the amino acid residues that are important in sequence discrimination, whereas others appear to bind DNA nonspecifically (the amino acids required for specificity are located elsewhere in the protein). Zinc fingers can also function as RNA-binding motifs—for example, in certain proteins that bind eukaryotic mRNAs and act as translational repressors. We discuss this role later (Section 28.3).

Homeodomain Another type of DNA-binding domain has been identified in a number of proteins that function as transcriptional regulators, especially during eukaryotic

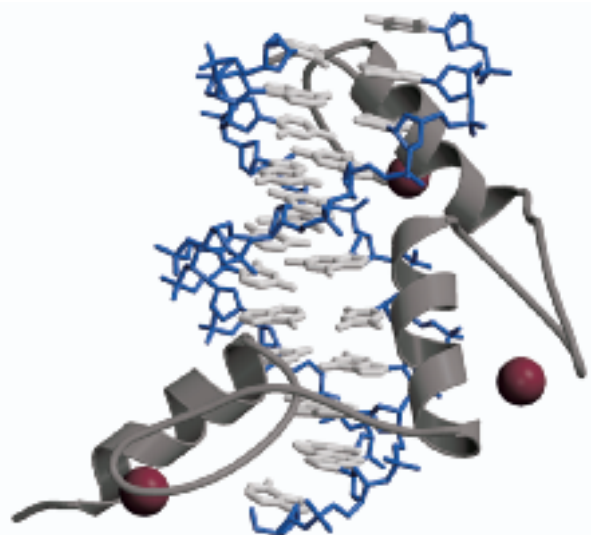


FIGURE 28–12 Zinc fingers. Three zinc fingers (gray) of the regulatory protein Zif268, complexed with DNA (blue and white) (PDB ID 1A1L). Each Zn^{2+} (maroon) coordinates with two His and two Cys residues (not shown).

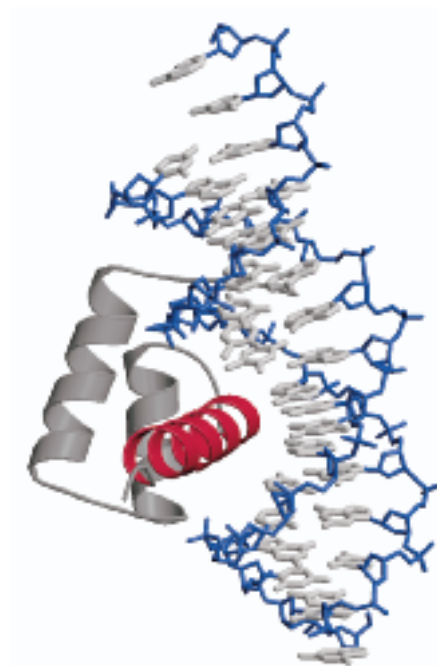


FIGURE 28–13 Homeodomain. Shown here is a homeodomain bound to DNA; one of the α helices (red), stacked on two others, can be seen protruding into the major groove (PDB ID 1B8I). This is only a small part of the much larger protein Ultrabithorax (Ubx), active in the regulation of development in fruit flies.

development. This domain of 60 amino acids—called the **homeodomain**, because it was discovered in homeotic genes (genes that regulate the development of body patterns)—is highly conserved and has now been identified in proteins from a wide variety of organisms, including humans (Fig. 28–13). The DNA-binding segment of the domain is related to the helix-turn-helix motif. The DNA sequence that encodes this domain is known as the **homeobox**.

Regulatory Proteins Also Have Protein-Protein Interaction Domains

Regulatory proteins contain domains not only for DNA binding but also for protein-protein interactions—with RNA polymerase, other regulatory proteins, or other subunits of the same regulatory protein. Examples include many eukaryotic transcription factors that function as gene activators, which often bind as dimers to the DNA, using DNA-binding domains that contain zinc fingers. Some structural domains are devoted to the interactions required for dimer formation, which is generally a prerequisite for DNA binding. Like DNA-binding motifs, the structural motifs that mediate protein-protein interactions tend to fall within one of a few common categories. Two important examples are the **leucine zipper** and the **basic helix-loop-helix**. Structural motifs such as

these are the basis for classifying some regulatory proteins into structural families.

Leucine Zipper This motif is an amphipathic α helix with a series of hydrophobic amino acid residues concentrated on one side (Fig. 28–14), with the hydrophobic surface forming the area of contact between the two polypeptides of a dimer. A striking feature of these α helices is the occurrence of Leu residues at every seventh position, forming a straight line along the hydrophobic surface. Although researchers initially thought the Leu residues interdigitated (hence the name “zipper”), we now know that they line up side by side as the interacting α helices coil around each other (forming a coiled coil; Fig. 28–14b). Regulatory proteins with leucine zippers often have a separate DNA-binding domain with a high concentration of basic (Lys or Arg) residues that can interact with the negatively charged phosphates of the DNA backbone. Leucine zippers have been found in many eukaryotic and a few prokaryotic proteins.

Basic Helix-Loop-Helix Another common structural motif occurs in some eukaryotic regulatory proteins implicated

in the control of gene expression during the development of multicellular organisms. These proteins share a conserved region of about 50 amino acid residues important in both DNA binding and protein dimerization. This region can form two short amphipathic α helices linked by a loop of variable length, the helix-loop-helix (distinct from the helix-turn-helix motif associated with DNA binding). The helix-loop-helix motifs of two polypeptides interact to form dimers (Fig. 28–15). In these proteins, DNA binding is mediated by an adjacent short amino acid sequence rich in basic residues, similar to the separate DNA-binding region in proteins containing leucine zippers.

Subunit Mixing in Eukaryotic Regulatory Proteins Several families of eukaryotic transcription factors have been defined based on close structural similarities. Within each family, dimers can sometimes form between two identical proteins (a homodimer) or between two different members of the family (a heterodimer). A hypothetical family of four different leucine-zipper proteins could thus form up to ten different dimeric species. In many cases, the different combinations appear to have distinct regulatory and functional properties.

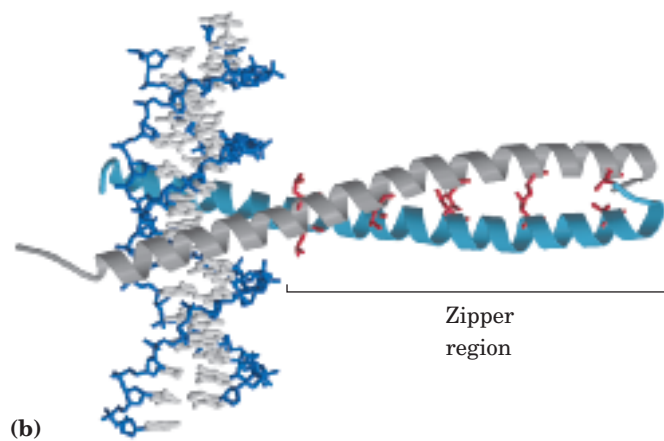
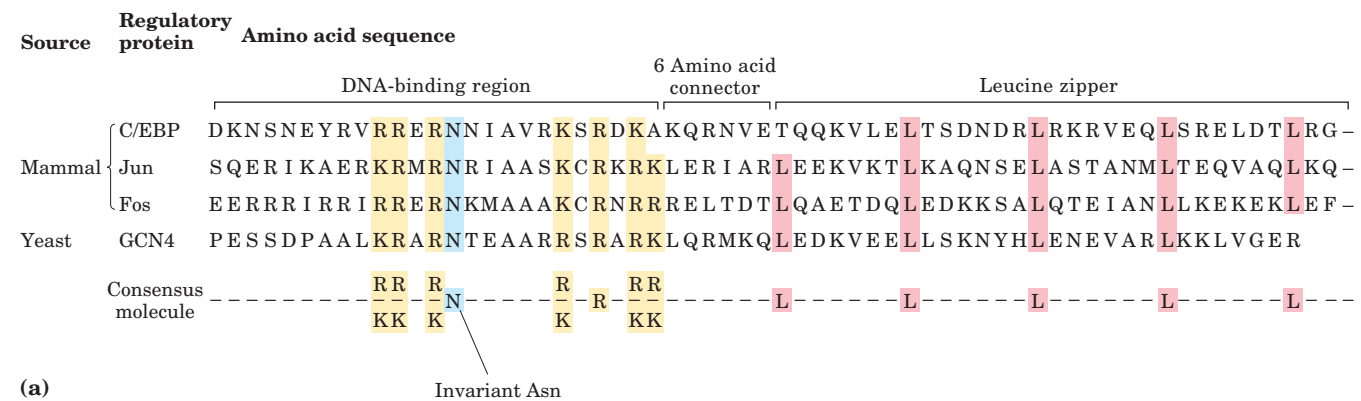


FIGURE 28–14 Leucine zippers. (a) Comparison of amino acid sequences of several leucine zipper proteins. Note the Leu (L) residues at every seventh position in the zipper region, and the number of Lys (K) and Arg (R) residues in the DNA-binding region. (b) Leucine zipper from the yeast activator protein GCN4 (PDB ID 1YSA). Only the “zippered” α helices (gray and light blue), derived from different subunits of the dimeric protein, are shown. The two helices wrap around each other in a gently coiled coil. The interacting Leu residues are shown in red.

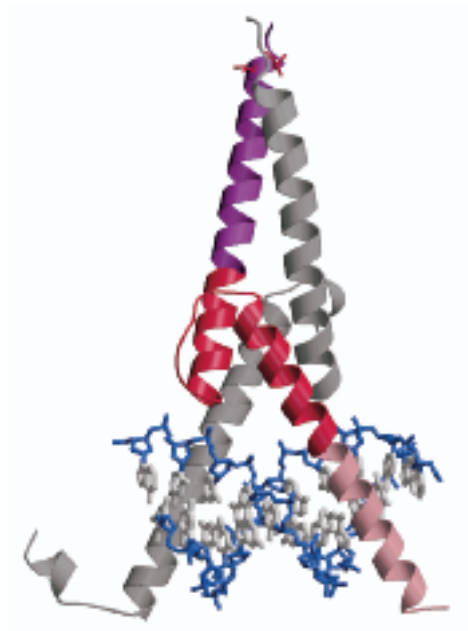


FIGURE 28-15 Helix-loop-helix. The human transcription factor Max, bound to its DNA target site (PDB ID 1HLO). The protein is dimeric; one subunit is colored. The DNA-binding segment (pink) merges with the first helix of the helix-loop-helix (red). The second helix merges with the carboxyl-terminal end of the subunit (purple). Interaction of the carboxyl-terminal helices of the two subunits describes a coiled coil very similar to that of a leucine zipper (see Fig. 28-14b), but with only one pair of interacting Leu residues (red side chains near the top) in this particular example. The overall structure is sometimes called a helix-loop-helix/leucine zipper motif.

In addition to structural domains devoted to DNA binding and dimerization (or oligomerization), many regulatory proteins must interact with RNA polymerase, with unrelated regulatory proteins, or with both. At least three different types of additional domains for protein-protein interaction have been characterized (primarily in eukaryotes): glutamine-rich, proline-rich, and acidic domains, the names reflecting the amino acid residues that are especially abundant.

Protein-DNA binding interactions are the basis of the intricate regulatory circuits fundamental to gene function. We now turn to a closer examination of these gene regulatory schemes, first in prokaryotic, then in eukaryotic systems.

SUMMARY 28.1 Principles of Gene Regulation

- The expression of genes is regulated by processes that affect the rates at which gene products are synthesized and degraded. Much of this regulation occurs at the level of transcription initiation, mediated by regulatory proteins that either repress transcription

(negative regulation) or activate transcription (positive regulation) at specific promoters.

- In bacteria, genes that encode products with interdependent functions are often clustered in an operon, a single transcriptional unit. Transcription of the genes is generally blocked by binding of a specific repressor protein at a DNA site called an operator. Dissociation of the repressor from the operator is mediated by a specific small molecule, an inducer. These principles were first elucidated in studies of the lactose (*lac*) operon. The Lac repressor dissociates from the *lac* operator when the repressor binds to its inducer, allolactose.
- Regulatory proteins are DNA-binding proteins that recognize specific DNA sequences; most have distinct DNA-binding domains. Within these domains, common structural motifs that bind DNA are the helix-turn-helix, zinc finger, and homeodomain.
- Regulatory proteins also contain domains for protein-protein interactions, including the leucine zipper and helix-loop-helix, which are involved in dimerization, and other motifs involved in activation of transcription.

28.2 Regulation of Gene Expression in Prokaryotes

As in many other areas of biochemical investigation, the study of the regulation of gene expression advanced earlier and faster in bacteria than in other experimental organisms. The examples of bacterial gene regulation presented here are chosen from among scores of well-studied systems, partly for their historical significance, but primarily because they provide a good overview of the range of regulatory mechanisms employed in prokaryotes. Many of the principles of prokaryotic gene regulation are also relevant to understanding gene expression in eukaryotic cells.

We begin by examining the lactose and tryptophan operons; each system has regulatory proteins, but the overall mechanisms of regulation are very different. This is followed by a short discussion of the SOS response in *E. coli*, illustrating how genes scattered throughout the genome can be coordinately regulated. We then describe two prokaryotic systems of quite different types, illustrating the diversity of gene regulatory mechanisms: regulation of ribosomal protein synthesis at the level of translation, with many of the regulatory proteins binding to RNA (rather than DNA), and regulation of a process called phase variation in *Salmonella*, which results from genetic recombination. First, we return to the *lac* operon to examine its features in greater detail.

The *lac* Operon Undergoes Positive Regulation

The operator-repressor-inducer interactions described earlier for the *lac* operon (Fig. 28–7) provide an intuitively satisfying model for an on/off switch in the regulation of gene expression. In truth, operon regulation is rarely so simple. A bacterium's environment is too complex for its genes to be controlled by one signal. Other factors besides lactose affect the expression of the *lac* genes, such as the availability of glucose. Glucose, metabolized directly by glycolysis, is *E. coli*'s preferred energy source. Other sugars can serve as the main or sole nutrient, but extra steps are required to prepare them for entry into glycolysis, necessitating the synthesis of additional enzymes. Clearly, expressing the genes for proteins that metabolize sugars such as lactose or arabinose is wasteful when glucose is abundant.

What happens to the expression of the *lac* operon when both glucose and lactose are present? A regulatory mechanism known as **catabolite repression** restricts expression of the genes required for catabolism of lactose, arabinose, and other sugars in the presence of glucose, even when these secondary sugars are also present. The effect of glucose is mediated by cAMP, as a coactivator, and an activator protein known as **cAMP receptor protein**, or **CRP** (the protein is sometimes called CAP, for catabolite gene activator protein). CRP is a homodimer (subunit M_r 22,000) with binding sites for DNA and cAMP. Binding is mediated by a helix-turn-helix motif within the protein's DNA-binding domain (Fig. 28–16). When glucose is absent, CRP-cAMP binds to a site near the *lac* promoter (Fig. 28–17a) and stimulates RNA transcription 50-fold. CRP-cAMP is therefore a positive regulatory element responsive to glucose levels, whereas the Lac repressor is a negative regulatory element responsive to lactose. The two act in con-

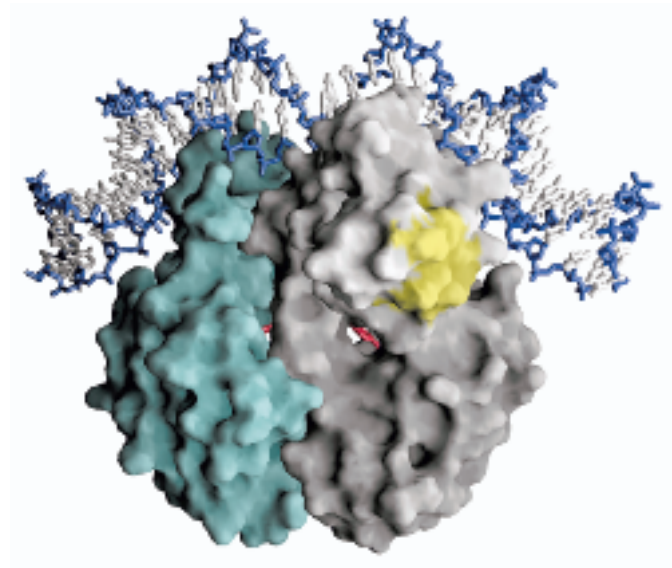


FIGURE 28–16 CRP homodimer. (PDB ID 1RUN) Bound molecules of cAMP are shown in red. Note the bending of the DNA around the protein. The region that interacts with RNA polymerase is shaded yellow.

cert. CRP-cAMP has little effect on the *lac* operon when the Lac repressor is blocking transcription, and dissociation of the repressor from the *lac* operator has little effect on transcription of the *lac* operon unless CRP-cAMP is present to facilitate transcription; when CRP is not bound, the wild-type *lac* promoter is a relatively weak promoter (Fig. 28–17b). The open complex of RNA polymerase and the promoter (see Fig. 26–6) does not form readily unless CRP-cAMP is present. CRP interacts directly with RNA polymerase (at the region shown in Fig. 28–16) through the polymerase's α subunit.

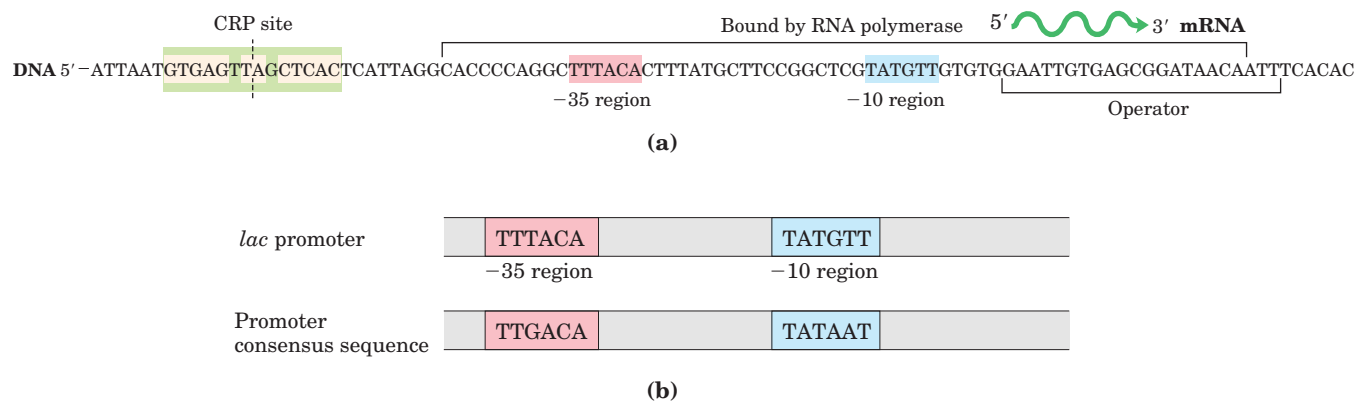


FIGURE 28–17 Activation of transcription of the *lac* operon by CRP.

(a) The binding site for CRP-cAMP is near the promoter. As in the case of the *lac* operator, the CRP site has twofold symmetry (bases shaded beige) about the axis indicated by the dashed line. (b) Sequence of

the *lac* promoter compared with the promoter consensus sequence. The differences mean that RNA polymerase binds relatively weakly to the *lac* promoter until the polymerase is activated by CRP-cAMP.

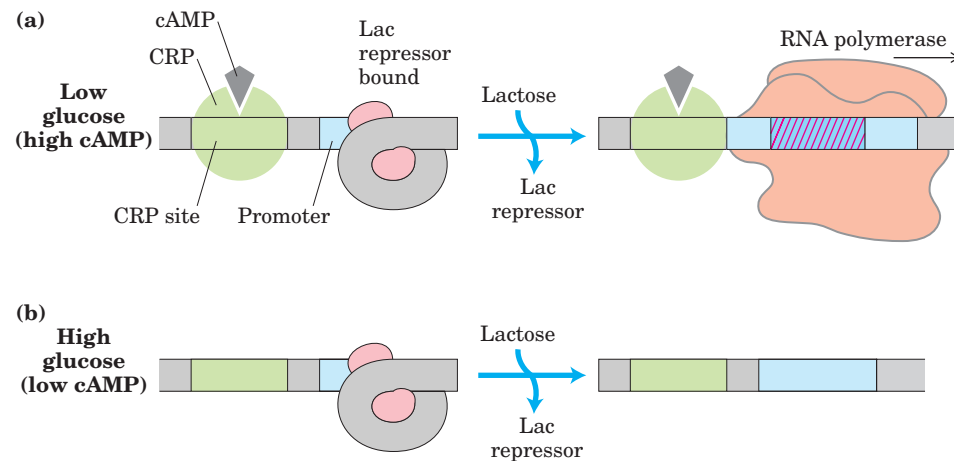


FIGURE 28-18 Combined effects of glucose and lactose on expression of the *lac* operon. (a) High levels of transcription take place only when glucose concentrations are low (so cAMP levels are high and CRP-cAMP is bound) and lactose concentrations are high (so the Lac repressor is not bound). (b) Without bound activator (CRP-cAMP), the *lac* promoter is poorly transcribed even when lactose concentrations are high and the Lac repressor is not bound.

The effect of glucose on CRP is mediated by the cAMP interaction (Fig. 28-18). CRP binds to DNA most avidly when cAMP concentrations are high. In the presence of glucose, the synthesis of cAMP is inhibited and efflux of cAMP from the cell is stimulated. As [cAMP] declines, CRP binding to DNA declines, thereby decreasing the expression of the *lac* operon. Strong induction of the *lac* operon therefore requires both lactose (to inactivate the *lac* repressor) and a lowered concentration of glucose (to trigger an increase in [cAMP] and increased binding of cAMP to CRP).

CRP and cAMP are involved in the coordinated regulation of many operons, primarily those that encode enzymes for the metabolism of secondary sugars such as lactose and arabinose. A network of operons with a common regulator is called a **regulon**. This arrangement, which allows for coordinated shifts in cellular functions that can require the action of hundreds of genes, is a major theme in the regulated expression of dispersed networks of genes in eukaryotes. Other bacterial regulons include the heat-shock gene system that responds to changes in temperature (p. 1083) and the genes induced in *E. coli* as part of the SOS response to DNA damage, described later.

Many Genes for Amino Acid Biosynthetic Enzymes Are Regulated by Transcription Attenuation

The 20 common amino acids are required in large amounts for protein synthesis, and *E. coli* can synthesize all of them. The genes for the enzymes needed to synthesize a given amino acid are generally clustered in an operon and are expressed whenever existing supplies of that amino acid are inadequate for cellular requirements. When the amino acid is abundant, the biosyn-

thetic enzymes are not needed and the operon is repressed.

The *E. coli* tryptophan (*trp*) operon (Fig. 28-19) includes five genes for the enzymes required to convert chorismate to tryptophan. Note that two of the enzymes catalyze more than one step in the pathway. The mRNA from the *trp* operon has a half-life of only about 3 min, allowing the cell to respond rapidly to changing needs for this amino acid. The Trp repressor is a homodimer, each subunit containing 107 amino acid residues (Fig. 28-20). When tryptophan is abundant it binds to the Trp repressor, causing a conformational change that permits the repressor to bind to the *trp* operator and inhibit expression of the *trp* operon. The *trp* operator site overlaps the promoter, so binding of the repressor blocks binding of RNA polymerase.

Once again, this simple on/off circuit mediated by a repressor is not the entire regulatory story. Different cellular concentrations of tryptophan can vary the rate of synthesis of the biosynthetic enzymes over a 700-fold range. Once repression is lifted and transcription begins, the rate of transcription is fine-tuned by a second regulatory process, called **transcription attenuation**, in which transcription is initiated normally but is abruptly halted *before* the operon genes are transcribed. The frequency with which transcription is attenuated is regulated by the availability of tryptophan and relies on the very close coupling of transcription and translation in bacteria.

The *trp* operon attenuation mechanism uses signals encoded in four sequences within a 162 nucleotide **leader** region at the 5' end of the mRNA, preceding the initiation codon of the first gene (Fig. 28-21a). Within the leader lies a region known as the **attenuator**, made up of sequences 3 and 4. These sequences base-pair to

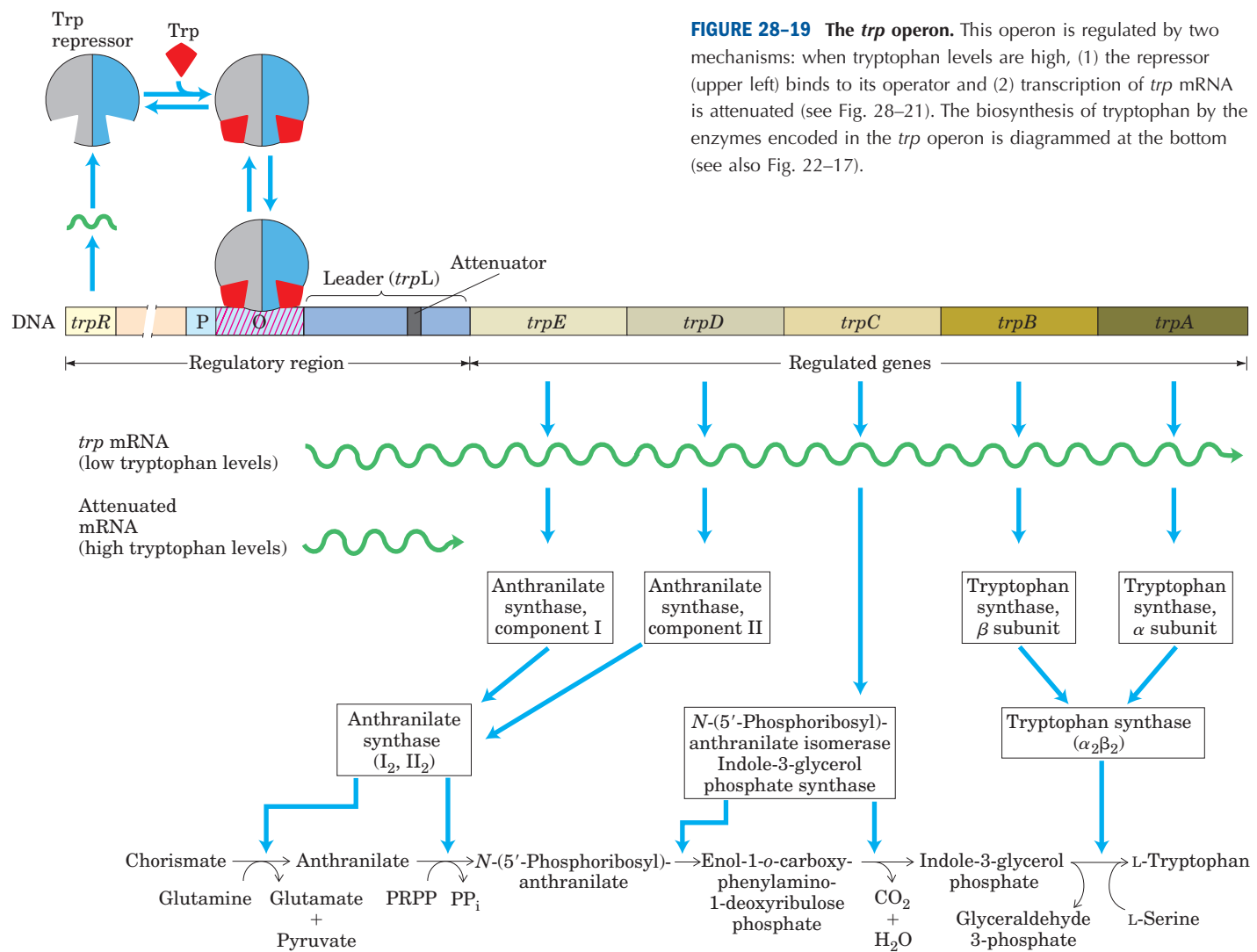


FIGURE 28-19 The *trp* operon. This operon is regulated by two mechanisms: when tryptophan levels are high, (1) the repressor (upper left) binds to its operator and (2) transcription of *trp* mRNA is attenuated (see Fig. 28-21). The biosynthesis of tryptophan by the enzymes encoded in the *trp* operon is diagrammed at the bottom (see also Fig. 22-17).

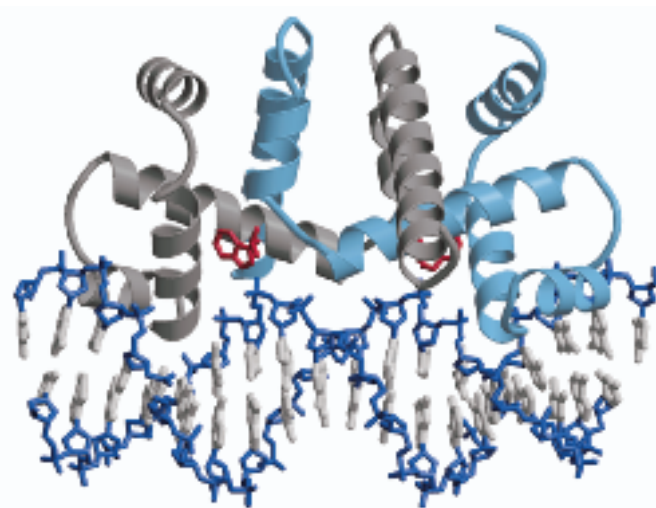


FIGURE 28-20 Trp repressor. The repressor is a dimer, with both subunits (gray and light blue) binding the DNA at helix-turn-helix motifs (PDB ID 1TRO). Bound molecules of tryptophan are in red.

form a G≡C-rich stem-and-loop structure closely followed by a series of U residues. The attenuator structure acts as a transcription terminator (Fig. 28-21b). Sequence 2 is an alternative complement for sequence 3 (Fig. 28-21c). If sequences 2 and 3 base-pair, the attenuator structure cannot form and transcription continues into the *trp* biosynthetic genes; the loop formed by the pairing of sequences 2 and 3 does not obstruct transcription.

Regulatory sequence 1 is crucial for a tryptophan-sensitive mechanism that determines whether sequence 3 pairs with sequence 2 (allowing transcription to continue) or with sequence 4 (attenuating transcription). Formation of the attenuator stem-and-loop structure depends on events that occur during *translation* of regulatory sequence 1, which encodes a leader peptide (so called because it is encoded by the leader region of the mRNA) of 14 amino acids, two of which are Trp residues. The leader peptide has no other known cellular function; its synthesis is simply an operon regulatory device.

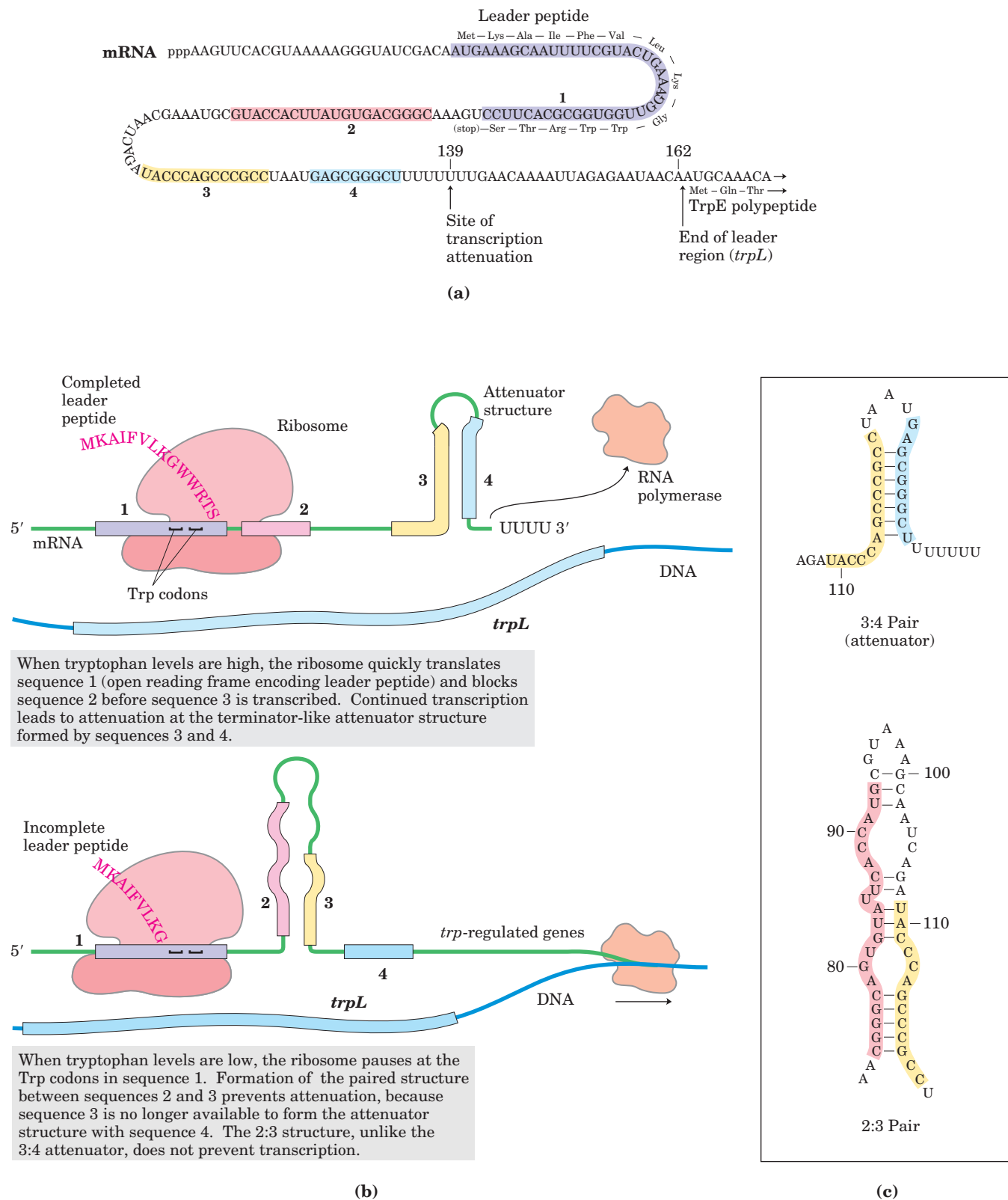


FIGURE 28-21 Transcriptional attenuation in the *trp* operon. Transcription is initiated at the beginning of the 162 nucleotide mRNA leader encoded by a DNA region called *trpL* (see Fig. 28-19). A regulatory mechanism determines whether transcription is attenuated at the end of the leader or continues into the structural genes. **(a)** The *trp* mRNA leader (*trpL*). The attenuation mechanism in the *trp* operon involves sequences 1 to 4 (highlighted). **(b)** Sequence 1 encodes a small peptide, the leader peptide, containing two Trp residues (W); it is translated immediately after transcription begins. Sequences 2 and

3 are complementary, as are sequences 3 and 4. The attenuator structure forms by the pairing of sequences 3 and 4 (top). Its structure and function are similar to those of a transcription terminator (see Fig. 26-7). Pairing of sequences 2 and 3 (bottom) prevents the attenuator structure from forming. Note that the leader peptide has no other cellular function. Translation of its open reading frame has a purely regulatory role that determines which complementary sequences (2 and 3 or 3 and 4) are paired. **(c)** Base-pairing schemes for the complementary regions of the *trp* mRNA leader.

This peptide is translated immediately after it is transcribed, by a ribosome that follows closely behind RNA polymerase as transcription proceeds.

When tryptophan concentrations are high, concentrations of charged tryptophan tRNA (Trp-tRNA^{Trp}) are also high. This allows translation to proceed rapidly past the two Trp codons of sequence 1 and into sequence 2, before sequence 3 is synthesized by RNA polymerase. In this situation, sequence 2 is covered by the ribosome and unavailable for pairing to sequence 3 when sequence 3 is synthesized; the attenuator structure (sequences 3 and 4) forms and transcription halts (Fig. 28–21b, top). When tryptophan concentrations are low, however, the ribosome stalls at the two Trp codons in sequence 1, because charged tRNA^{Trp} is less available. Sequence 2 remains free while sequence 3 is synthesized, allowing these two sequences to base-pair and permitting transcription to proceed (Fig. 28–21b, bottom). In this way, the proportion of transcripts that are attenuated declines as tryptophan concentration declines.

Many other amino acid biosynthetic operons use a similar attenuation strategy to fine-tune biosynthetic enzymes to meet the prevailing cellular requirements. The

15 amino acid leader peptide produced by the *phe* operon contains seven Phe residues. The *leu* operon leader peptide has four contiguous Leu residues. The leader peptide for the *his* operon contains seven contiguous His residues. In fact, in the *his* operon and a number of others, attenuation is sufficiently sensitive to be the *only* regulatory mechanism.

Induction of the SOS Response Requires Destruction of Repressor Proteins

Extensive DNA damage in the bacterial chromosome triggers the induction of many distantly located genes. This response, called the SOS response (p. 976), provides another good example of coordinated gene regulation. Many of the induced genes are involved in DNA repair (see Table 25–6). The key regulatory proteins are the RecA protein and the LexA repressor.

The LexA repressor (M_r 22,700) inhibits transcription of all the SOS genes (Fig. 28–22), and induction of the SOS response requires removal of LexA. This is not a simple dissociation from DNA in response to binding of a small molecule, as in the regulation of the *lac* operon described above. Instead, the LexA repressor is

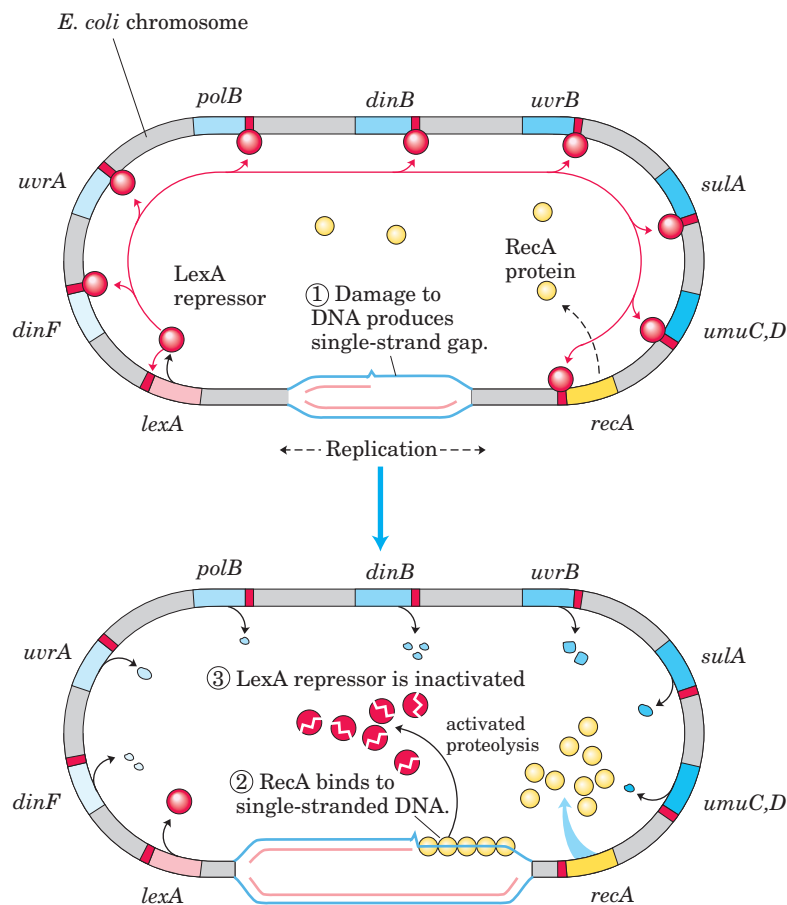


FIGURE 28–22 SOS response in *E. coli*. See Table 25–6 for the functions of many of these proteins. The LexA protein is the repressor in this system, which has an operator site (red) near each gene. Because the *recA* gene is not entirely repressed by the LexA repressor, the normal cell contains about 1,000 RecA monomers. ① When DNA is extensively damaged (e.g., by UV light), DNA replication is halted and the number of single-strand gaps in the DNA increases. ② RecA protein binds to this damaged, single-stranded DNA, activating the protein's coprotease activity. ③ While bound to DNA, the RecA protein facilitates cleavage and inactivation of the LexA repressor. When the repressor is inactivated, the SOS genes, including *recA*, are induced; RecA levels increase 50- to 100-fold.

inactivated when it catalyzes its own cleavage at a specific Ala–Gly peptide bond, producing two roughly equal protein fragments. At physiological pH, this autocleavage reaction requires the RecA protein. RecA is not a protease in the classical sense, but its interaction with LexA facilitates the repressor's self-cleavage reaction. This function of RecA is sometimes called a co-protease activity.

The RecA protein provides the functional link between the biological signal (DNA damage) and induction of the SOS genes. Heavy DNA damage leads to numerous single-strand gaps in the DNA, and only RecA that is bound to single-stranded DNA can facilitate cleavage of the LexA repressor (Fig. 28–22, bottom). Binding of RecA at the gaps eventually activates its co-protease activity, leading to cleavage of the LexA repressor and SOS induction.

During induction of the SOS response in a severely damaged cell, RecA also cleaves and thus inactivates the repressors that otherwise allow propagation of certain viruses in a dormant lysogenic state within the bacterial host. This provides a remarkable illustration of evolutionary adaptation. These repressors, like LexA, also undergo self-cleavage at a specific Ala–Gly peptide bond, so induction of the SOS response permits replication of the virus and lysis of the cell, releasing new viral particles. Thus the bacteriophage can make a hasty exit from a compromised bacterial host cell.

Synthesis of Ribosomal Proteins Is Coordinated with rRNA Synthesis

In bacteria, an increased cellular demand for protein synthesis is met by increasing the number of ribosomes rather than altering the activity of individual ribosomes. In general, the number of ribosomes increases as the cellular growth rate increases. At high growth rates, ribosomes make up approximately 45% of the cell's dry weight. The proportion of cellular resources devoted to making ribosomes is so large, and the function of ribosomes so important, that cells must coordinate the synthesis of the ribosomal components: the ribosomal proteins (r-proteins) and RNAs (rRNAs). This regulation is distinct from the mechanisms described so far, because it occurs largely at the level of *translation*.

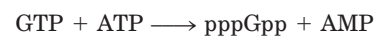
The 52 genes that encode the r-proteins occur in at least 20 operons, each with 1 to 11 genes. Some of these operons also contain the genes for the subunits of DNA primase (see Fig. 25–13), RNA polymerase (see Fig. 26–4), and protein synthesis elongation factors (see Fig. 27–23)—revealing the close coupling of replication, transcription, and protein synthesis during cell growth.

The r-protein operons are regulated primarily through a translational feedback mechanism. One r-protein encoded by each operon also functions as a **translational repressor**, which binds to the mRNA

transcribed from that operon and blocks translation of all the genes the messenger encodes (Fig. 28–23). In general, the r-protein that plays the role of repressor also binds directly to an rRNA. Each translational repressor r-protein binds with higher affinity to the appropriate rRNA than to its mRNA, so the mRNA is bound and translation repressed only when the level of the r-protein exceeds that of the rRNA. This ensures that translation of the mRNAs encoding r-proteins is repressed only when synthesis of these r-proteins exceeds that needed to make functional ribosomes. In this way, the rate of r-protein synthesis is kept in balance with rRNA availability.

The mRNA binding site for the translational repressor is near the translational start site of one of the genes in the operon, usually the first gene (Fig. 28–23). In other operons this would affect only that one gene, because in bacterial polycistronic mRNAs most genes have independent translation signals. In the r-protein operons, however, the translation of one gene depends on the translation of all the others. The mechanism of this translational coupling is not yet understood in detail. However, in some cases the translation of multiple genes appears to be blocked by folding of the mRNA into an elaborate three-dimensional structure that is stabilized both by internal base-pairing (as in Fig. 8–26) and by binding of the translational repressor protein. When the translational repressor is absent, ribosome binding and translation of one or more of the genes disrupts the folded structure of the mRNA and allows all the genes to be translated.

Because the synthesis of r-proteins is coordinated with the available rRNA, the regulation of ribosome production reflects the regulation of rRNA synthesis. In *E. coli*, rRNA synthesis from the seven rRNA operons responds to cellular growth rate and to changes in the availability of crucial nutrients, particularly amino acids. The regulation coordinated with amino acid concentrations is known as the **stringent response** (Fig. 28–24). When amino acid concentrations are low, rRNA synthesis is halted. Amino acid starvation leads to the binding of uncharged tRNAs to the ribosomal A site; this triggers a sequence of events that begins with the binding of an enzyme called **stringent factor** (RelA protein) to the ribosome. When bound to the ribosome, stringent factor catalyzes formation of the unusual nucleotide guanosine tetraphosphate (ppGpp; see Fig. 8–42); it adds pyrophosphate to the 3' position of GTP, in the reaction



then a phosphohydrolase cleaves off one phosphate to form ppGpp. The abrupt rise in ppGpp level in response to amino acid starvation results in a great reduction in rRNA synthesis, mediated at least in part by the binding of ppGpp to RNA polymerase.

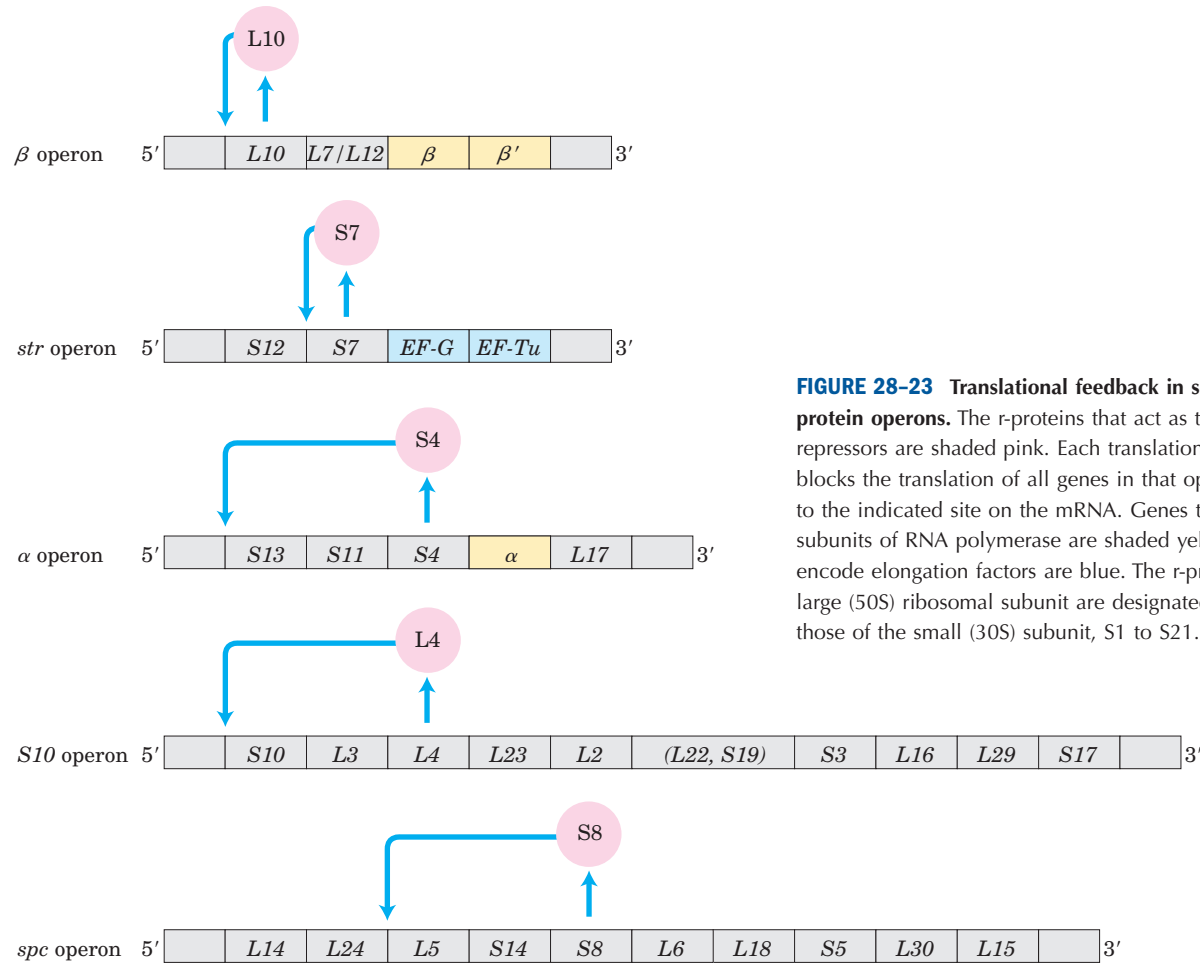
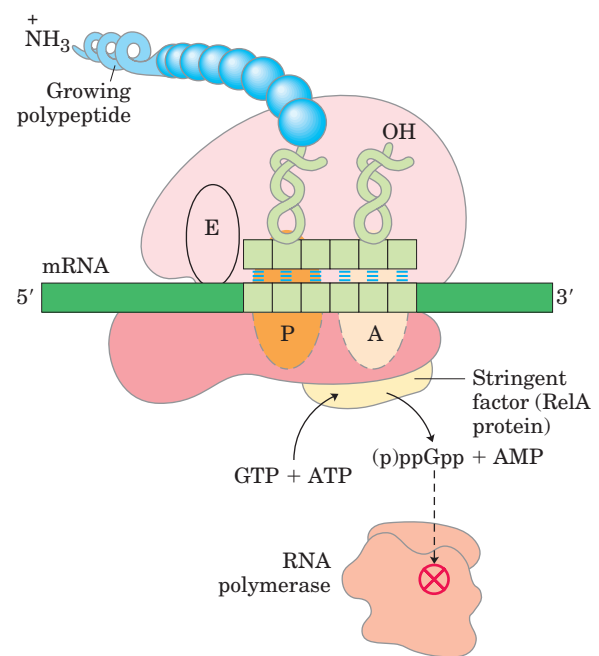


FIGURE 28-23 Translational feedback in some ribosomal protein operons. The r-proteins that act as translational repressors are shaded pink. Each translational repressor blocks the translation of all genes in that operon by binding to the indicated site on the mRNA. Genes that encode subunits of RNA polymerase are shaded yellow; genes that encode elongation factors are blue. The r-proteins of the large (50S) ribosomal subunit are designated L1 to L34; those of the small (30S) subunit, S1 to S21.

FIGURE 28-24 Stringent response in *E. coli*. This response to amino acid starvation is triggered by binding of an uncharged tRNA in the ribosomal A site. A protein called stringent factor binds to the ribosome and catalyzes the synthesis of pppGpp, which is converted by a phosphohydrolase to ppGpp. The signal ppGpp reduces transcription of some genes and increases that of others, in part by binding to the β subunit of RNA polymerase and altering the enzyme's promoter specificity. Synthesis of rRNA is reduced when ppGpp levels increase.



The nucleotide ppGpp, along with cAMP, belongs to a class of modified nucleotides that act as cellular second messengers (p. 302). In *E. coli*, these two nucleotides serve as starvation signals; they cause large changes in cellular metabolism by increasing or decreasing the transcription of hundreds of genes. In eukaryotic cells, similar nucleotide second messengers also have multiple regulatory functions. The coordination of cellular metabolism with cell growth is highly complex, and further regulatory mechanisms undoubtedly remain to be discovered.

Some Genes Are Regulated by Genetic Recombination

Salmonella typhimurium, which inhabits the mammalian intestine, moves by rotating the flagella on its cell surface (Fig. 28–25). The many copies of the protein flagellin (M_r 53,000) that make up the flagella are prominent targets of mammalian immune systems. But *Salmonella* cells have a mechanism that evades the immune response: they switch between two distinct flagellin proteins (FljB and FliC) roughly once every 1,000 generations, using a process called **phase variation**.

The switch is accomplished by periodic inversion of a segment of DNA containing the promoter for a flagellin gene. The inversion is a site-specific recombination reaction (see Fig. 25–39) mediated by the Hin recombinase at specific 14 bp sequences (*hix* sequences)



FIGURE 28–25 *Salmonella typhimurium*, with flagella evident.

at either end of the DNA segment. When the DNA segment is in one orientation, the gene for FljB flagellin and the gene encoding a repressor (FljA) are expressed (Fig. 28–26a); the repressor shuts down expression of the gene for FliC flagellin. When the DNA segment is inverted (Fig. 28–26b), the *fljA* and *fljB* genes are no longer transcribed, and the *fliC* gene is induced as the repressor becomes depleted. The Hin recombinase, encoded by the *hin* gene in the DNA segment that undergoes inversion, is expressed when the DNA segment is in either orientation, so the cell can always switch from one state to the other.

This type of regulatory mechanism has the advantage of being absolute: gene expression is impossible

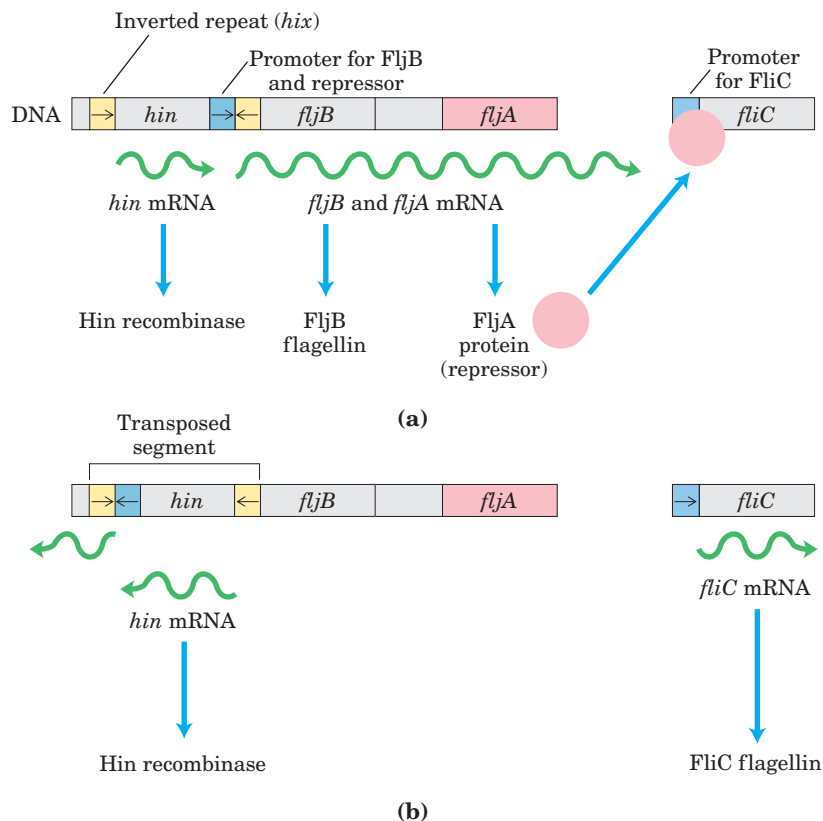


FIGURE 28–26 Regulation of flagellin genes in *Salmonella*: phase variation. The products of genes *fliC* and *fljB* are different flagellins. The *hin* gene encodes the recombinase that catalyzes inversion of the DNA segment containing the *fljB* promoter and the *hin* gene. The recombination sites (inverted repeats) are called *hix* (yellow). (a) In one orientation, *fljB* is expressed along with a repressor protein (product of the *fljA* gene) that represses transcription of the *fliC* gene. (b) In the opposite orientation only the *fliC* gene is expressed; the *fljA* and *fljB* genes cannot be transcribed. The interconversion between these two states, known as phase variation, also requires two other nonspecific DNA-binding proteins (not shown), HU (histonelike protein from *U13*, a strain of *E. coli*) and FIS (factor for inversion stimulation).

TABLE 28–1 Examples of Gene Regulation by Recombination

| System | Recombinase/ recombination site | Type of recombination | Function |
|---|--|-----------------------------------|--|
| Phase variation (<i>Salmonella</i>) | Hin/hix | Site-specific | Alternative expression of two flagellin genes allows evasion of host immune response. |
| Host range (bacteriophage μ) | Gin/gix | Site-specific | Alternative expression of two sets of tail fiber genes affects host range. |
| Mating-type switch (yeast) | HO endonuclease, RAD52 protein, other proteins/MAT | Nonreciprocal gene conversion* | Alternative expression of two mating types of yeast, a and α , creates cells of different mating types that can mate and undergo meiosis. |
| Antigenic variation (trypanosomes) [†] | Varies | Nonreciprocal gene conversion* | Successive expression of different genes encoding the variable surface glycoproteins (VSGs) allows evasion of host immune response. |

*In nonreciprocal gene conversion (a class of recombination events not discussed in Chapter 25), genetic information is moved from one part of the genome (where it is silent) to another (where it is expressed). The reaction is similar to replicative transposition (see Fig. 25–43).

[†]Trypanosomes cause African sleeping sickness and other diseases (see Box 22–2). The outer surface of a trypanosome is made up of multiple copies of a single VSG, the major surface antigen. A cell can change surface antigens to more than 100 different forms, precluding an effective defense by the host immune system.

when the gene is physically separated from its promoter (note the position of the *fljB* promoter in Fig. 28–26b). An absolute on/off switch may be important in this system (even though it affects only one of the two flagellin genes), because a flagellum with just one copy of the wrong flagellin might be vulnerable to host antibodies against that protein. The *Salmonella* system is by no means unique. Similar regulatory systems occur in a number of other bacteria and in some bacteriophages, and recombination systems with similar functions have been found in eukaryotes (Table 28–1). Gene regulation by DNA rearrangements that move genes and/or promoters is particularly common in pathogens that benefit by changing their host range or by changing their surface proteins, thereby staying ahead of host immune systems.

SUMMARY 28.2 Regulation of Gene Expression in Prokaryotes

- In addition to repression by the Lac repressor, the *E. coli lac* operon undergoes positive regulation by the cAMP receptor protein (CRP). When [glucose] is low, [cAMP] is high and CRP-cAMP binds to a specific site on the DNA, stimulating transcription of the *lac* operon and production of lactose-metabolizing enzymes. The presence of glucose depresses [cAMP], decreasing expression of *lac* and other

genes involved in metabolism of secondary sugars. A group of coordinately regulated operons is referred to as a regulon.

- Operons that produce the enzymes of amino acid synthesis have a regulatory circuit called attenuation, which uses a transcription termination site (the attenuator) in the mRNA. Formation of the attenuator is modulated by a mechanism that couples transcription and translation while responding to small changes in amino acid concentration.
- In the SOS system, multiple unlinked genes repressed by a single repressor are induced simultaneously when DNA damage triggers RecA protein-facilitated autocatalytic proteolysis of the repressor.
- In the synthesis of ribosomal proteins, one protein in each r-protein operon acts as a translational repressor. The mRNA is bound by the repressor, and translation is blocked only when the r-protein is present in excess of available rRNA. Some genes are regulated by genetic recombination processes that move promoters relative to the genes being regulated. Regulation can also take place at the level of translation. These diverse mechanisms permit very sensitive cellular responses to environmental change.

28.3 Regulation of Gene Expression in Eukaryotes

Initiation of transcription is a crucial regulation point for both prokaryotic and eukaryotic gene expression. Although some of the same regulatory mechanisms are used in both systems, there is a fundamental difference in the regulation of transcription in eukaryotes and bacteria.

We can define a transcriptional ground state as the inherent activity of promoters and transcriptional machinery *in vivo* in the absence of regulatory sequences. In bacteria, RNA polymerase generally has access to every promoter and can bind and initiate transcription at some level of efficiency in the absence of activators or repressors; the transcriptional ground state is therefore nonrestrictive. In eukaryotes, however, strong promoters are generally inactive *in vivo* in the absence of regulatory proteins; that is, the transcriptional ground state is restrictive. This fundamental difference gives rise to at least four important features that distinguish the regulation of gene expression in eukaryotes from that in bacteria.

First, access to eukaryotic promoters is restricted by the structure of chromatin, and activation of transcription is associated with many changes in chromatin structure in the transcribed region. Second, although eukaryotic cells have both positive and negative regulatory mechanisms, positive mechanisms predominate in all systems characterized so far. Thus, given that the transcriptional ground state is restrictive, virtually every eukaryotic gene requires activation to be transcribed. Third, eukaryotic cells have larger, more complex multimeric regulatory proteins than do bacteria. Finally, transcription in the eukaryotic nucleus is separated from translation in the cytoplasm in both space and time.

The complexity of regulatory circuits in eukaryotic cells is extraordinary, as the following discussion shows. We conclude the section with an illustrated description of one of the most elaborate circuits: the regulatory cascade that controls development in fruit flies.

Transcriptionally Active Chromatin Is Structurally Distinct from Inactive Chromatin

The effects of chromosome structure on gene regulation in eukaryotes have no clear parallel in prokaryotes. In the eukaryotic cell cycle, interphase chromosomes appear, at first viewing, to be dispersed and amorphous (see Figs 12–41, 24–25). Nevertheless, several forms of chromatin can be found along these chromosomes. About 10% of the chromatin in a typical eukaryotic cell is in a more condensed form than the rest of the chromatin. This form, **heterochromatin**, is transcriptionally inactive. Heterochromatin is generally associated

with particular chromosome structures—the centromeres, for example. The remaining, less condensed chromatin is called **euchromatin**.

Transcription of a eukaryotic gene is strongly repressed when its DNA is condensed within heterochromatin. Some, but not all, of the euchromatin is transcriptionally active. Transcriptionally active chromosomal regions can be detected based on their increased sensitivity to nuclease-mediated degradation. Nucleases such as DNase I tend to cleave the DNA of carefully isolated chromatin into fragments of multiples of about 200 bp, reflecting the regular repeating structure of the nucleosome (see Fig. 24–26). In actively transcribed regions, the fragments produced by nuclease activity are smaller and more heterogeneous in size. These regions contain **hypersensitive sites**, sequences especially sensitive to DNase I, which consist of about 100 to 200 bp within the 1,000 bp flanking the 5' ends of transcribed genes. In some genes, hypersensitive sites are found farther from the 5' end, near the 3' end, or even within the gene itself.

Many hypersensitive sites correspond to binding sites for known regulatory proteins, and the relative absence of nucleosomes in these regions may allow the binding of these proteins. Nucleosomes are entirely absent in some regions that are very active in transcription, such as the rRNA genes. Transcriptionally active chromatin tends to be deficient in histone H1, which binds to the linker DNA between nucleosome particles.

Histones within transcriptionally active chromatin and heterochromatin also differ in their patterns of covalent modification. The core histones of nucleosome particles (H2A, H2B, H3, H4; see Fig. 24–27) are modified by irreversible methylation of Lys residues, phosphorylation of Ser or Thr residues, acetylation (see below), or attachment of ubiquitin (see Fig. 27–41). Each of the core histones has two distinct structural domains. A central domain is involved in histone-histone interaction and the wrapping of DNA around the nucleosome. A second, lysine-rich amino-terminal domain is generally positioned near the exterior of the assembled nucleosome particle; the covalent modifications occur at specific residues concentrated in this amino-terminal domain. The patterns of modification have led some researchers to propose the existence of a histone code, in which modification patterns are recognized by enzymes that alter the structure of chromatin. Modifications associated with transcriptional activation would be recognized by enzymes that make the chromatin more accessible to the transcription machinery.

5-Methylation of cytosine residues of CpG sequences is common in eukaryotic DNA (p. 296), but DNA in transcriptionally active chromatin tends to be undermethylated. Furthermore, CpG sites in particular genes are more often undermethylated in cells from tissues where the genes are expressed than in those where

the genes are not expressed. The overall pattern suggests that active chromatin is prepared for transcription by the removal of potential structural barriers.

Chromatin Is Remodeled by Acetylation and Nucleosomal Displacements

The detailed mechanisms for transcription-associated structural changes in chromatin, called **chromatin remodeling**, are now coming to light, including identification of a variety of enzymes directly implicated in the process. These include enzymes that covalently modify the core histones of the nucleosome and others that use the chemical energy of ATP to remodel nucleosomes on the DNA (Table 28–2).

The acetylation and deacetylation of histones figure prominently in the processes that activate chromatin for transcription. As noted above, the amino-terminal domains of the core histones are generally rich in Lys residues. Particular Lys residues are acetylated by histone acetyltransferases (HATs). Cytosolic (type B) HATs acetylate newly synthesized histones before the histones are imported into the nucleus. The subsequent assembly of the histones into chromatin is facilitated by additional proteins: CAF1 for H3 and H4, and NAP1 for H2A and H2B. (See Table 28–2 for an explanation of some of these abbreviated names.)

Where chromatin is being activated for transcription, the nucleosomal histones are further acetylated by nuclear (type A) HATs. The acetylation of multiple Lys residues in the amino-terminal domains of histones H3 and H4 can reduce the affinity of the entire nucleosome for DNA. Acetylation may also prevent or promote interactions with other proteins involved in transcription or its regulation. When transcription of a gene is no

longer required, the acetylation of nucleosomes in that vicinity is reduced by the activity of histone deacetylases, as part of a general gene-silencing process that restores the chromatin to a transcriptionally inactive state. In addition to the removal of certain acetyl groups, new covalent modification of histones marks chromatin as transcriptionally inactive. As an example, the Lys residue at position 9 in histone H3 is often methylated in heterochromatin.

Chromatin remodeling also requires protein complexes that actively move or displace nucleosomes, hydrolyzing ATP in the process (Table 28–2). The enzyme complex SWI/SNF found in all eukaryotic cells, contains 11 polypeptides (total M_r 2×10^6) that together create hypersensitive sites in the chromatin and stimulate the binding of transcription factors. SWI/SNF is not required for the transcription of every gene. NURF is another ATP-dependent enzyme complex that remodels chromatin in ways that complement and overlap the activity of SWI/SNF. These enzyme complexes play an important role in preparing a region of chromatin for active transcription.

Many Eukaryotic Promoters Are Positively Regulated

As already noted, eukaryotic RNA polymerases have little or no intrinsic affinity for their promoters; initiation of transcription is almost always dependent on the action of multiple activator proteins. One important reason for the apparent predominance of positive regulation seems obvious: the storage of DNA within chromatin effectively renders most promoters inaccessible, so genes are normally silent in the absence of other regulation. The structure of chromatin affects access to some promoters more than others, but repressors that

TABLE 28–2 Some Enzyme Complexes Catalyzing Chromatin Structural Changes Associated with Transcription

| Enzyme complex* | Oligomeric structure (number of polypeptides) | Source | Activities |
|-----------------|--|----------------------------------|---|
| GCN5-ADA2-ADA3 | 3 | Yeast | GCN5 has type A HAT activity |
| SAGA/PCAF | >20 | Eukaryotes | Includes GCN5-ADA2-ADA3 |
| SWI/SNF | 11; total M_r 2×10^6 | Eukaryotes | ATP-dependent nucleosome remodeling |
| NURF | 4; total M_r 500,000 | <i>Drosophila</i> | ATP-dependent nucleosome remodeling |
| CAF1 | >2 | Humans; <i>Drosophila</i> | Responsible for binding histones H3 and H4 to DNA |
| NAP1 | 1; M_r 125,000 | Widely distributed in eukaryotes | Responsible for binding histones H2A and H2B to DNA |

*The abbreviations for eukaryotic genes and proteins are often more confusing or obscure than those used for bacteria. The complex of GCN5 (general control *n*onderepressible) and ADA (alteration/deficiency activation) proteins was discovered during investigation of the regulation of nitrogen metabolism genes in yeast. These proteins can be part of the larger SAGA complex (SPF, ADA2,3, GCN5, acetyltransferase) in yeasts. The equivalent of SAGA in humans is PCAF (p300/CBP-associated factor). SWI (switching) was discovered as a protein required for expression of certain genes involved in mating-type switching in yeast, and SNF (sucrose nonfermenting) as a factor for expression of the yeast gene for sucrase. Subsequent studies revealed multiple SWI and SNF proteins that acted in a complex. The SWI/SNF complex has a role in the expression of a wide range of genes and has been found in many eukaryotes, including humans. NURF is nuclear remodeling factor; CAF1, chromatin assembly factor; and NAP1, nucleosome assembly protein.

bind to DNA so as to preclude access of RNA polymerase (negative regulation) would often be simply redundant. Other factors are at play in the use of positive regulation, and speculation generally centers around two: the large size of eukaryotic genomes and the greater efficiency of positive regulation.

First, nonspecific DNA binding of regulatory proteins becomes a more important problem in the much larger genomes of higher eukaryotes. And the chance that a single specific binding sequence will occur randomly at an inappropriate site also increases with genome size. Specificity for transcriptional activation can be improved if each of several positive-regulatory proteins must bind specific DNA sequences and then form a complex in order to become active. The average number of regulatory sites for a gene in a multicellular organism is probably at least five. The requirement for binding of several positive-regulatory proteins to specific DNA sequences vastly reduces the probability of the random occurrence of a functional juxtaposition of all the necessary binding sites. In principle, a similar strategy could be used by multiple negative-regulatory elements, but this brings us to the second reason for the use of positive regulation: it is simply more efficient. If the 30,000 to 35,000 genes in the human genome were negatively regulated, each cell would have to synthesize, at all times, this same number of different repressors (or many times this number if multiple regulatory elements were used at each promoter) in concentrations sufficient to permit specific binding to each “unwanted” gene. In positive regulation, most of the genes are normally inactive (that is, RNA polymerases do not bind to the promoters) and the cell synthesizes only the activator proteins needed to promote transcription of the subset of genes required in the cell at that time. These arguments notwithstanding, there are examples of negative regulation in eukaryotes, from yeast to humans, as we shall see.

DNA-Binding Transactivators and Coactivators Facilitate Assembly of the General Transcription Factors

To continue our exploration of the regulation of gene expression in eukaryotes, we return to the interactions between promoters and RNA polymerase II (Pol II), the enzyme responsible for the synthesis of eukaryotic mRNAs. Although most (but not all) Pol II promoters include the TATA box and Inr (initiator) sequences, with their standard spacing (see Fig. 26–8), they vary greatly in both the number and the location of additional sequences required for the regulation of transcription. These additional regulatory sequences are usually called **enhancers** in higher eukaryotes and **upstream activator sequences (UASs)** in yeast. A typical enhancer may be found hundreds or even thousands of base pairs

upstream from the transcription start site, or may even be downstream, within the gene itself. When bound by the appropriate regulatory proteins, an enhancer increases transcription at nearby promoters regardless of its orientation in the DNA. The UASs of yeast function in a similar way, although generally they must be positioned upstream and within a few hundred base pairs of the transcription start site. An average Pol II promoter may be affected by a half-dozen regulatory sequences of this type, and even more complex promoters are quite common.

Successful binding of active RNA polymerase II holoenzyme at one of its promoters usually requires the action of other proteins (Fig. 28–27), of three types: (1) **basal transcription factors** (see Fig. 26–9, Table 26–1), required at every Pol II promoter; (2) **DNA-binding transactivators**, which bind to enhancers or UASs and facilitate transcription; and (3) **coactivators**. The latter group act indirectly—not by binding to the DNA—and are required for essential communication between the DNA-binding transactivators and the complex composed of Pol II and the general transcription factors. Furthermore, a variety of repressor proteins can interfere with communication between the RNA polymerase and the DNA-binding transactivators, resulting in repression of transcription (Fig. 28–27b). Here we focus on the protein complexes shown in Figure 28–27 and on how they interact to activate transcription.

TATA-Binding Protein The first component to bind in the assembly of a preinitiation complex at the TATA box of a typical Pol II promoter is the **TATA-binding protein (TBP)**. The complete complex includes the basal (or general) transcription factors TFIIB, TFIIE, TFIIIF, TFIIH; Pol II; and perhaps TFIIA (not all of the factors are shown in Fig. 28–27). This minimal preinitiation complex, however, is often insufficient for the initiation of transcription and generally does not form at all if the promoter is obscured within chromatin. Positive regulation leading to transcription is imposed by the transactivators and coactivators.

DNA-Binding Transactivators The requirements for transactivators vary greatly from one promoter to another. A few transactivators are known to facilitate transcription at hundreds of promoters, whereas others are specific for a few promoters. Many transactivators are sensitive to the binding of signal molecules, providing the capacity to activate or deactivate transcription in response to a changing cellular environment. Some enhancers bound by DNA-binding transactivators are quite distant from the promoter’s TATA box. How do the transactivators function at a distance? The answer in most cases seems to be that, as indicated earlier, the intervening DNA is looped so that the various protein complexes can interact directly. The looping is promoted by certain non-

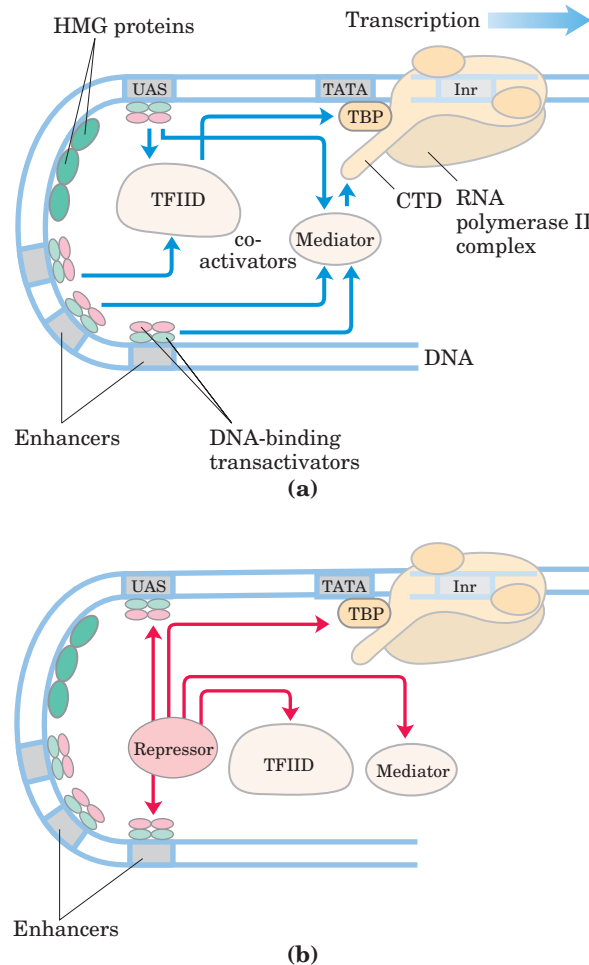


FIGURE 28-27 Eukaryotic promoters and regulatory proteins. RNA polymerase II and its associated general transcription factors form a preinitiation complex at the TATA box and Inr site of the cognate promoters, a process facilitated by DNA-binding transactivators, acting through TFIID and/or mediator. **(a)** A composite promoter with typical sequence elements and protein complexes found in both yeast and higher eukaryotes. The carboxyl-terminal domain (CTD) of Pol II (see Fig. 26-9) is an important point of interaction with mediator and other protein complexes. Not shown are the protein complexes required for histone acetylation and chromatin remodeling. For the DNA-binding transactivators, DNA-binding domains are shown in green, activation domains in pink. The interactions symbolized by blue arrows are discussed in the text. **(b)** A wide variety of eukaryotic transcriptional repressors function by a range of mechanisms. Some bind directly to DNA, displacing a protein complex required for activation; others interact with various parts of the transcription or activation complexes to prevent activation. Possible points of interaction are indicated with red arrows.

histone proteins that are abundant in chromatin and bind nonspecifically to DNA. These high mobility group (HMG) proteins (Fig. 28-27; “high mobility” refers to their electrophoretic mobility in polyacrylamide gels) play an important structural role in chromatin remodeling and transcriptional activation.

Coactivator Protein Complexes Most transcription requires the presence of additional protein complexes. Some major regulatory protein complexes that interact with Pol II have been defined both genetically and biochemically. These coactivator complexes act as intermediaries between the DNA-binding transactivators and the Pol II complex.

The best-characterized coactivator is the transcription factor TFIID (Fig. 28-27). In eukaryotes, TFIID is a large complex that includes TBP and ten or more TBP-associated factors (TAFs). Some TAFs resemble histones and may play a role in displacing nucleosomes during the activation of transcription. Many DNA-binding transactivators aid in transcription initiation by interacting with one or more TAFs. The requirement for TAFs to initiate transcription can vary greatly from one gene to another. Some promoters require TFIID, some do not, and some require only subsets of the TFIID TAF subunits.

Another important coactivator consists of 20 or more polypeptides in a protein complex called **mediator** (Fig. 28-27); the 20 core polypeptides are highly conserved from fungi to humans. Mediator binds tightly to the carboxyl-terminal domain (CTD) of the largest subunit of Pol II. The mediator complex is required for both basal and regulated transcription at promoters used by Pol II, and it also stimulates the phosphorylation of the CTD by TFIIH. Both mediator and TFIID are required at some promoters. As with TFIID, some DNA-binding transactivators interact with one or more components of the mediator complex. Coactivator complexes function at or near the promoter’s TATA box.

Choreography of Transcriptional Activation We can now begin to piece together the sequence of transcriptional activation events at a typical Pol II promoter. First, crucial remodeling of the chromatin takes place in stages. Some DNA-binding transactivators have significant affinity for their binding sites even when the sites are within condensed chromatin. Binding of one transactivator may facilitate the binding of others, gradually displacing some nucleosomes.

The bound transactivators can then interact directly with HATs or enzyme complexes such as SWI/SNF (or both), accelerating the remodeling of the surrounding chromatin. In this way a bound transactivator can draw in other components necessary for further chromatin remodeling to permit transcription of specific genes. The bound transactivators, generally acting through complexes such as TFIID or mediator (or both), stabilize the binding of Pol II and its associated transcription factors and greatly facilitate formation of the preinitiation transcription complex. Complexity in these regulatory circuits is the rule rather than the exception, with multiple DNA-bound transactivators promoting transcription.

The script can change from one promoter to another, but most promoters seem to require a precisely ordered assembly of components to initiate transcription. The assembly process is not always fast. At some genes it may take minutes; at certain genes in higher eukaryotes the process can take days.

Reversible Transcriptional Activation Although rarer, some eukaryotic regulatory proteins that bind to Pol II promoters can act as repressors, inhibiting the formation of active preinitiation complexes (Fig. 28–27b). Some transactivators can adopt different conformations, enabling them to serve as transcriptional activators or repressors. For example, some steroid hormone receptors (described later) function in the nucleus as DNA-binding transactivators, stimulating the transcription of certain genes when a particular steroid hormone signal is present. When the hormone is absent, the receptor proteins revert to a repressor conformation, *preventing* the formation of preinitiation complexes. In some cases this repression involves interaction with histone deacetylases and other proteins that help restore the surrounding chromatin to its transcriptionally inactive state.

The Genes of Galactose Metabolism in Yeast Are Subject to Both Positive and Negative Regulation

Some of the general principles described above can be illustrated by one well-studied eukaryotic regulatory circuit (Fig. 28–28). The enzymes required for the importation and metabolism of galactose in yeast are encoded by genes scattered over several chromosomes (Table 28–3). Each of the *GAL* genes is transcribed separately, and yeast cells have no operons like those in bacteria. However, all the *GAL* genes have similar promoters and are regulated coordinately by a common set of proteins. The promoters for the *GAL* genes consist of the TATA box and Inr sequences, as well as an upstream activator sequence (UAS_G) recognized by a DNA-binding transcriptional activator known as Gal4p protein (Gal4p). Regulation of gene expression by galactose entails an interplay between Gal4p and two other proteins, Gal80p and Gal3p (Fig. 28–28). Gal80p forms a complex with Gal4p, preventing Gal4p from functioning as an activator of the *GAL* promoters. When galactose is present, it binds Gal3p, which then interacts with Gal80p, allowing Gal4p to function as an activator at the various *GAL* promoters.

Other protein complexes also have a role in activating transcription of the *GAL* genes. These may include the SAGA complex for histone acetylation, the SWI/SNF complex for nucleosome remodeling, and the mediator complex. Figure 28–29 provides an idea of the complexity of protein interactions in the overall process of transcriptional activation in eukaryotic cells.

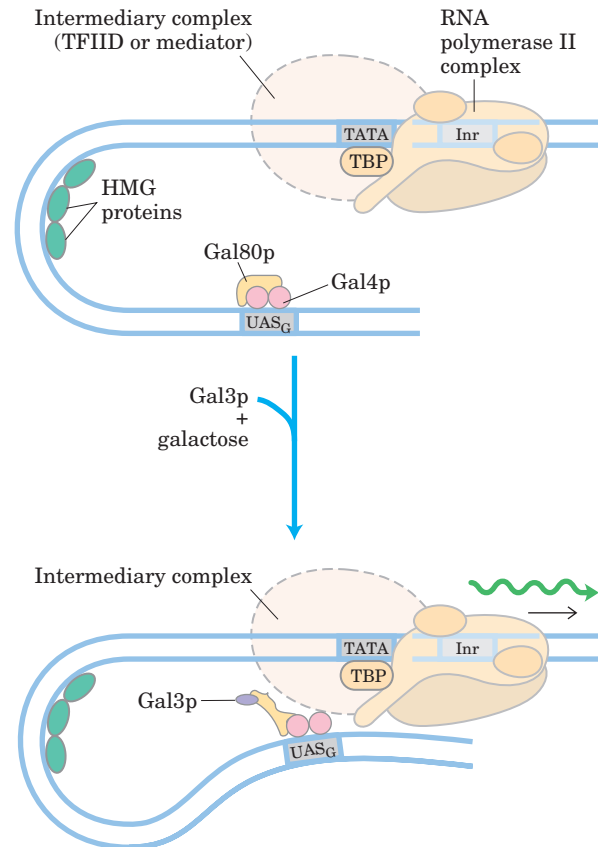


FIGURE 28–28 Regulation of transcription at genes of galactose metabolism in yeast. Galactose is imported into the cell and converted to galactose 6-phosphate by a pathway involving six enzymes whose genes are scattered over three chromosomes (see Table 28–3). Transcription of these genes is regulated by the combined actions of the proteins Gal4p, Gal80p, and Gal3p, with Gal4p playing the central role of DNA-binding transactivator. The Gal4p-Gal80p complex is inactive in gene activation. Binding of galactose to Gal3p and its interaction with Gal80p produce a conformational change in Gal80p that allows Gal4p to function in transcription activation.

Glucose is the preferred carbon source for yeast, as it is for bacteria. When glucose is present, most of the *GAL* genes are repressed—whether galactose is present or not. The *GAL* regulatory system described above is effectively overridden by a complex catabolite repression system that includes several proteins (not depicted in Fig. 28–29).

DNA-Binding Transactivators Have a Modular Structure

DNA-binding transactivators typically have a distinct structural domain for specific DNA binding and one or more additional domains for transcriptional activation or for interaction with other regulatory proteins. Interaction of two regulatory proteins is often mediated by domains containing leucine zippers (Fig. 28–14) or helix-loop-helix motifs (Fig. 28–15). We consider here three

TABLE 28-3 Genes of Galactose Metabolism in Yeast

| | Protein function | Chromosomal location | Protein size (number of residues) | Relative protein expression in different carbon sources | | |
|-------------------------|--|----------------------|--------------------------------------|--|----------|-----------|
| | | | | Glucose | Glycerol | Galactose |
| Regulated genes | | | | | | |
| <i>GAL1</i> | Galactokinase | II | 528 | — | — | +++ |
| <i>GAL2</i> | Galactose permease | XII | 574 | — | — | +++ |
| <i>PGM2</i> | Phosphoglucomutase | XIII | 569 | + | + | ++ |
| <i>GAL7</i> | Galactose 1-phosphate uridylyltransferase | II | 365 | — | — | +++ |
| <i>GAL10</i> | UDP-glucose 4-epimerase | II | 699 | — | — | +++ |
| <i>MEL1</i> | α -Galactosidase | II | 453 | — | + | ++ |
| Regulatory genes | | | | | | |
| <i>GAL3</i> | Inducer | IV | 520 | — | + | ++ |
| <i>GAL4</i> | Transcriptional activator | XVI | 881 | + / — | + | + |
| <i>GAL80</i> | Transcriptional inhibitor | XIII | 435 | + | + | ++ |

Source: Adapted from Reece, R. & Platt, A. (1997) Signaling activation and repression of RNA polymerase II transcription in yeast. *Bioessays* 19, 1001–1010.

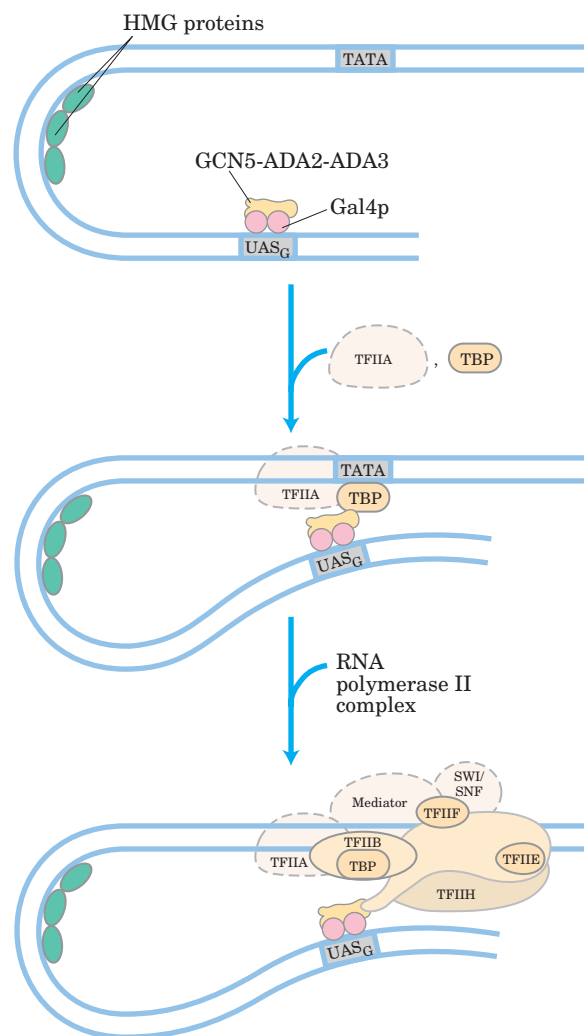


FIGURE 28-29 Protein complexes involved in transcription activation of a group of related eukaryotic genes. The *GAL* system illustrates the complexity of this process, but not all these protein complexes are yet known to affect *GAL* gene transcription. Note that many of the complexes (such as SWI/SNF, GCN5-ADA2-ADA3, and mediator) affect the transcription of many genes. The complexes assemble stepwise. First the DNA-binding transactivators bind, then the additional protein complexes needed to remodel the chromatin and allow transcription to begin.

distinct types of structural domains used in activation by DNA-binding transactivators (Fig. 28–30a): Gal4p, Sp1, and CTF1.

Gal4p contains a zinc fingerlike structure in its DNA-binding domain, near the amino terminus; this domain has six Cys residues that coordinate two Zn^{2+} . The protein functions as a homodimer (with dimerization mediated by interactions between two coiled coils) and binds to UAS_G, a palindromic DNA sequence about 17 bp long. Gal4p has a separate activation domain with many acidic amino acid residues. Experiments that substitute a variety of different peptide sequences for the **acidic activation domain** of Gal4p suggest that the acidic nature of this domain is critical to its function, although its precise amino acid sequence can vary considerably.

Sp1 (M_r 80,000) is a DNA-binding transactivator for a large number of genes in higher eukaryotes. Its DNA binding site, the GC box (consensus sequence

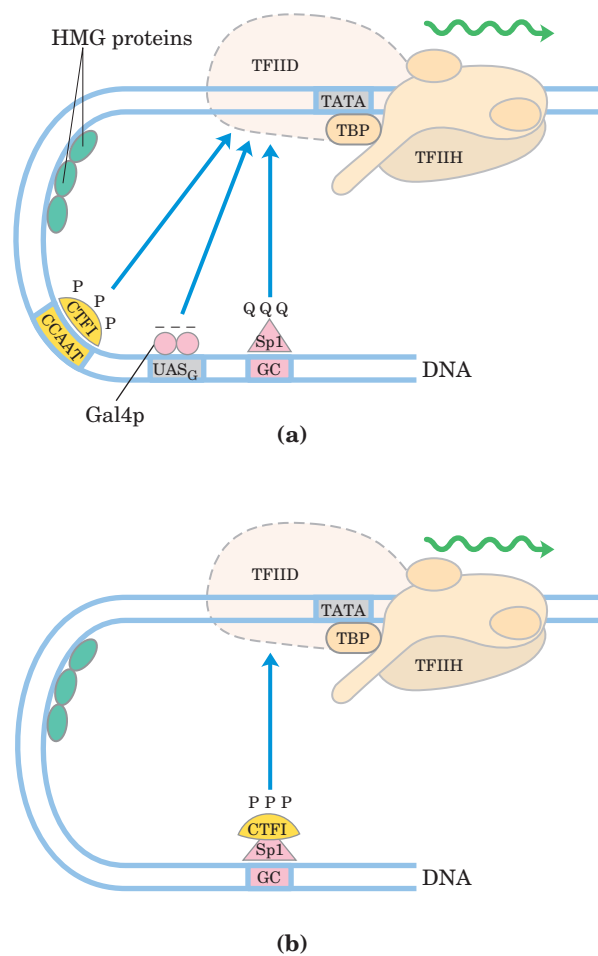


FIGURE 28-30 DNA-binding transactivators. (a) Typical DNA-binding transactivators such as CTF1, Gal4p, and Sp1 have a DNA-binding domain and an activation domain. The nature of the activation domain is indicated by symbols: — — —, acidic; Q Q Q, glutamine-rich; P P P, proline-rich. Some or all of these proteins may activate transcription by interacting with intermediary complexes such as TFIID or mediator. Note that the binding sites illustrated here are not generally found together near a single gene. (b) A chimeric protein containing the DNA-binding domain of Sp1 and the activation domain of CTF1 activates transcription if a GC box is present.

GGGCGG), is usually quite near the TATA box. The DNA-binding domain of the Sp1 protein is near its carboxyl terminus and contains three zinc fingers. Two other domains in Sp1 function in activation, and are notable in that 25% of their amino acid residues are Gln. A wide variety of other activator proteins also have these **glutamine-rich domains**.

CCAAT-binding transcription factor 1 (CTF1) belongs to a family of DNA-binding transactivators that bind a sequence called the CCAAT site (its consensus sequence is TGGN₆GCCAA, where N is any nucleotide). The DNA-binding domain of CTF1 contains many basic amino acid residues, and the binding region is probably arranged as an α helix. This protein has neither a helix-

turn-helix nor a zinc finger motif; its DNA-binding mechanism is not yet clear. CTF1 has a **proline-rich activation domain**, with Pro accounting for more than 20% of the amino acid residues.

The discrete activation and DNA-binding domains of regulatory proteins often act completely independently, as has been demonstrated in “domain-swapping” experiments. Genetic engineering techniques (Chapter 9) can join the proline-rich activation domain of CTF1 to the DNA-binding domain of Sp1 to create a protein that, like normal Sp1, binds to GC boxes on the DNA and activates transcription at a nearby promoter (as in Fig. 28–30b). The DNA-binding domain of Gal4p has similarly been replaced experimentally with the DNA-binding domain of the prokaryotic LexA repressor (of the SOS response; Fig. 28–22). This chimeric protein neither binds at UAS_G nor activates the yeast *GAL* genes (as would normal Gal4p) unless the UAS_G sequence in the DNA is replaced by the LexA recognition site.

Eukaryotic Gene Expression Can Be Regulated by Intercellular and Intracellular Signals

The effects of steroid hormones (and of thyroid and retinoid hormones, which have the same mode of action) provide additional well-studied examples of the modulation of eukaryotic regulatory proteins by direct interaction with molecular signals (see Fig. 12–40). Unlike other types of hormones, steroid hormones do not have to bind to plasma membrane receptors. Instead, they can interact with intracellular receptors that are themselves transcriptional transactivators. Steroid hormones too hydrophobic to dissolve readily in the blood (estrogen, progesterone, and cortisol, for example) travel on specific carrier proteins from their point of release to their target tissues. In the target tissue, the hormone passes through the plasma membrane by simple diffusion and binds to its specific receptor protein in the nucleus. The hormone-receptor complex acts by binding to highly specific DNA sequences called **hormone response elements (HREs)**, thereby altering gene expression. Hormone binding triggers changes in the conformation of the receptor proteins so that they become capable of interacting with additional transcription factors. The bound hormone-receptor complex can either enhance or suppress the expression of adjacent genes.

The DNA sequences (HREs) to which hormone-receptor complexes bind are similar in length and arrangement, but differ in sequence, for the various steroid hormones. Each receptor has a consensus HRE sequence (Table 28–4) to which the hormone-receptor complex binds well, with each consensus consisting of two six-nucleotide sequences, either contiguous or separated by three nucleotides, in tandem or in a palindromic arrangement. The hormone receptors have a highly conserved DNA-binding domain with two zinc fingers

TABLE 28–4 Hormone Response Elements (HREs) Bound by Steroid-Type Hormone Receptors

| Receptor | Consensus sequence bound* |
|----------------------|---------------------------------|
| Androgen | GG(A/T)ACAN ₂ TGTTCT |
| Glucocorticoid | GGTACAN ₃ TGTTCT |
| Retinoic acid (some) | AGGTCAN ₅ AGGTCA |
| Vitamin D | AGGTCAN ₃ AGGTCA |
| Thyroid hormone | AGGTCAN ₃ AGGTCA |
| RX [†] | AGGTCANAGGTCANAGGTCANAGGTCA |

*N represents any nucleotide.

[†]Forms a dimer with the retinoic acid receptor or vitamin D receptor.

(Fig. 28–31). The hormone-receptor complex binds to the DNA as a dimer, with the zinc finger domains of each monomer recognizing one of the six-nucleotide sequences. The ability of a given hormone to act through the hormone-receptor complex to alter the expression of a specific gene depends on the exact sequence of the HRE, its position relative to the gene, and the number of HREs associated with the gene.

Unlike the DNA-binding domain, the ligand-binding region of the receptor protein—always at the carboxyl terminus—is quite specific to the particular receptor. In the ligand-binding region, the glucocorticoid receptor is only 30% similar to the estrogen receptor and 17% similar to the thyroid hormone receptor. The size of the ligand-binding region varies dramatically; in the vitamin D receptor it has only 25 amino acid residues, whereas in the mineralocorticoid receptor it has 603 residues. Mutations that change one amino acid in these regions can result in loss of responsiveness to a specific hormone.

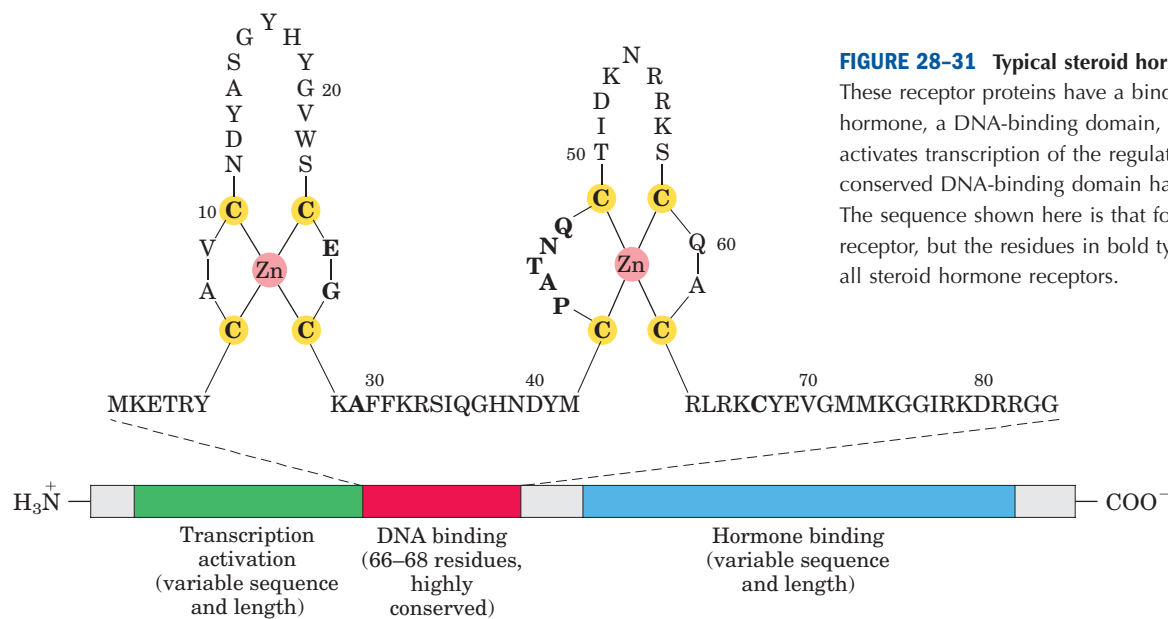
Some humans unable to respond to cortisol, testosterone, vitamin D, or thyroxine have mutations of this type.

Regulation Can Result from Phosphorylation of Nuclear Transcription Factors

We noted in Chapter 12 that the effects of insulin on gene expression are mediated by a series of steps leading ultimately to the activation of a protein kinase in the nucleus that phosphorylates specific DNA-binding proteins and thereby alters their ability to act as transcription factors (see Fig. 12–6). This general mechanism mediates the effects of many nonsteroid hormones. For example, the β -adrenergic pathway that leads to elevated levels of cytosolic cAMP, which acts as a second messenger in eukaryotes as well as in prokaryotes (see Figs 12–12, 28–18), also affects the transcription of a set of genes, each of which is located near a specific DNA sequence called a cAMP response element (CRE). The catalytic subunit of protein kinase A, released when cAMP levels rise (see Fig. 12–15), enters the nucleus and phosphorylates a nuclear protein, the CRE-binding protein (CREB). When phosphorylated, CREB binds to CREs near certain genes and acts as a transcription factor, turning on the expression of these genes.

Many Eukaryotic mRNAs Are Subject to Translational Repression

Regulation at the level of translation assumes a much more prominent role in eukaryotes than in bacteria and is observed in a range of cellular situations. In contrast to the tight coupling of transcription and translation in bacteria, the transcripts generated in a eukaryotic nucleus

**FIGURE 28–31** Typical steroid hormone receptors.

These receptor proteins have a binding site for the hormone, a DNA-binding domain, and a region that activates transcription of the regulated gene. The highly conserved DNA-binding domain has two zinc fingers. The sequence shown here is that for the estrogen receptor, but the residues in bold type are common to all steroid hormone receptors.

must be processed and transported to the cytoplasm before translation. This can impose a significant delay on the appearance of a protein. When a rapid increase in protein production is needed, a translationally repressed mRNA already in the cytoplasm can be activated for translation without delay. Translational regulation may play an especially important role in regulating certain very long eukaryotic genes (a few are measured in the millions of base pairs), for which transcription and mRNA processing can require many hours. Some genes are regulated at both the transcriptional and translational stages, with the latter playing a role in the fine-tuning of cellular protein levels. In some anucleate cells, such as reticulocytes (immature erythrocytes), transcriptional control is entirely unavailable and translational control of stored mRNAs becomes essential. As described below, translational controls can also have spatial significance during development, when the regulated translation of prepositioned mRNAs creates a local gradient of the protein product.

Eukaryotes have at least three main mechanisms of translational regulation.

1. Initiation factors are subject to phosphorylation by a number of protein kinases. The phosphorylated forms are often less active and cause a general depression of translation in the cell.
2. Some proteins bind directly to mRNA and act as translational repressors, many of them binding at specific sites in the 3' untranslated region (3'UTR). So positioned, these proteins interact with other translation initiation factors bound to the mRNA or with the 40S ribosomal subunit to prevent translation initiation (Fig. 28–32; compare this with Fig. 27–22).
3. Binding proteins, present in eukaryotes from yeast to mammals, disrupt the interaction between eIF4E and eIF4G (see Fig. 27–22). The mammalian versions are known as 4E-BPs (eIF4E binding proteins). When cell growth is slow, these proteins limit translation by binding to the site on eIF4E that normally interacts with eIF4G. When cell growth resumes or increases in response to growth factors or other stimuli, the binding proteins are inactivated by protein kinase-dependent phosphorylation.

The variety of translational regulation mechanisms provides flexibility, allowing focused repression of a few mRNAs or global regulation of all cellular translation.

Translational regulation has been particularly well studied in reticulocytes. One such mechanism in these cells involves eIF2, the initiation factor that binds to the initiator tRNA and conveys it to the ribosome; when Met-tRNA has bound to the P site, the factor eIF2B

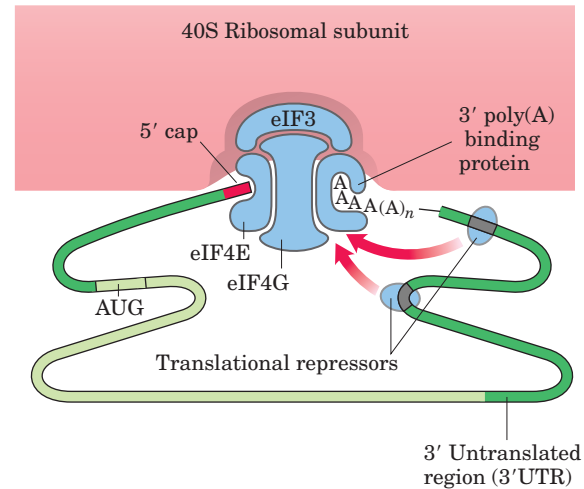


FIGURE 28–32 Translational regulation of eukaryotic mRNA. One of the most important mechanisms for translational regulation in eukaryotes involves the binding of translational repressors (RNA-binding proteins) to specific sites in the 3' untranslated region (3'UTR) of the mRNA. These proteins interact with eukaryotic initiation factors or with the ribosome (see Fig. 27–22) to prevent or slow translation.

binds to eIF2, recycling it with the aid of GTP binding and hydrolysis. The maturation of reticulocytes includes destruction of the cell nucleus, leaving behind a plasma membrane packed with hemoglobin. Messenger RNAs deposited in the cytoplasm before the loss of the nucleus allow for the replacement of hemoglobin. When reticulocytes become deficient in iron or heme, the translation of globin mRNAs is repressed. A protein kinase called HCR (*hemin-controlled repressor*) is activated, catalyzing the phosphorylation of eIF2. In its phosphorylated form, eIF2 forms a stable complex with eIF2B that sequesters the eIF2, making it unavailable for participation in translation. In this way, the reticulocyte coordinates the synthesis of globin with the availability of heme.

Many additional examples of translational regulation have been found in studies of the development of multicellular organisms, as discussed in more detail below.

Posttranscriptional Gene Silencing Is Mediated by RNA Interference

In higher eukaryotes, including nematodes, fruit flies, plants, and mammals, a class of small RNAs has been discovered that mediates the silencing of particular genes. The RNAs function by interacting with mRNAs, often in the 3'UTR, resulting in either mRNA degradation or translation inhibition. In either case, the mRNA, and thus the gene that produces it, is silenced. This form of gene regulation controls developmental timing in at least some organisms. It is also used as a mechanism to protect against invading RNA viruses (particularly

important in plants, which lack an immune system) and to control the activity of transposons. In addition, small RNA molecules may play a critical (but still undefined) role in the formation of heterochromatin.

The small RNAs are sometimes called micro-RNAs (miRNAs). Many are present only transiently during development, and these are sometimes referred to as small temporal RNAs (stRNAs). Hundreds of different miRNAs have been identified in higher eukaryotes. They are transcribed as precursor RNAs about 70 nucleotides long, with internally complementary sequences that form hairpinlike structures (Fig. 28–33). The precursors are cleaved by endonucleases to form short duplexes about 20 to 25 nucleotides long. The best-characterized nuclease goes by the delightfully suggestive name Dicer; endonucleases in the Dicer family are widely distributed in higher eukaryotes. One strand of the processed miRNA is transferred to the target mRNA (or to a viral or transposon RNA), leading to inhibition of translation or degradation of the RNA (Fig. 28–33a).

This gene regulation mechanism has an interesting and very useful practical side. If an investigator introduces into an organism a duplex RNA molecule corresponding in sequence to virtually any mRNA, the Dicer endonuclease cleaves the duplex into short segments,

called small interfering RNAs (siRNAs). These bind to the mRNA and silence it (Fig. 28–33b). The process is known as **RNA interference (RNAi)**. In plants, virtually any gene can be effectively shut down in this way. In nematodes, simply introducing the duplex RNA into the worm's diet produces very effective suppression of the target gene. The technique has rapidly become an important tool in the ongoing efforts to study gene function, because it can disrupt gene function without creating a mutant organism. The procedure can be applied to humans as well. Laboratory-produced siRNAs have already been used to block HIV and poliovirus infections in cultured human cells for a week or so at a time. Although this work is in its infancy, the rapid progress makes RNA interference a field to watch for future medical advances.

Development Is Controlled by Cascades of Regulatory Proteins

For sheer complexity and intricacy of coordination, the patterns of gene regulation that bring about development of a zygote into a multicellular animal or plant have no peer. Development requires transitions in morphology and protein composition that depend on tightly coordinated changes in expression of the genome. More genes are expressed during early development than in any other part of the life cycle. For example, in the sea urchin, an oocyte has about 18,500 *different* mRNAs, compared with about 6,000 different mRNAs in the cells of a typical differentiated tissue. The mRNAs in the oocyte give rise to a cascade of events that regulate the expression of many genes across both space and time.

Several animals have emerged as important model systems for the study of development, because they are easy to maintain in a laboratory and have relatively short generation times. These include nematodes, fruit flies, zebra fish, mice, and the plant *Arabidopsis*. This discussion focuses on the development of *Drosophila melanogaster* is particularly well advanced and can be used to illustrate patterns and principles of general significance.

The life cycle of the fruit fly includes complete metamorphosis during its progression from an embryo to an adult (Fig. 28–34). Among the most important characteristics of the embryo are its **polarity** (the anterior and posterior parts of the animal are readily distinguished, as are its dorsal and ventral parts) and its **metamerism** (the embryo body is made up of serially repeating segments, each with characteristic features). During development, these segments become organized into a head, thorax, and abdomen. Each segment of the adult thorax has a different set of appendages. Development of this complex pattern is under genetic control, and a variety of pattern-regulating genes have been

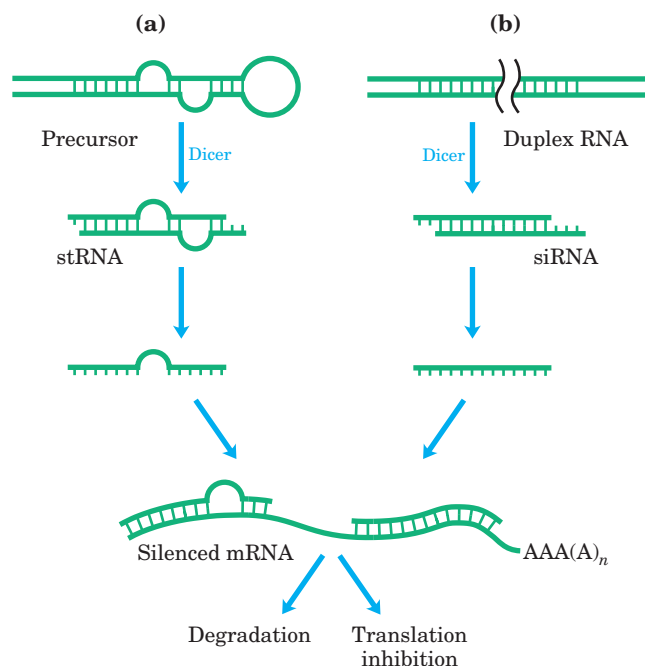


FIGURE 28–33 Gene silencing by RNA interference. (a) Small temporal RNAs (stRNAs) are generated by Dicer-mediated cleavage of longer precursors that fold to create duplex regions. The stRNAs then bind to mRNAs, leading to degradation of mRNA or inhibition of translation. (b) Double-stranded RNAs can be constructed and introduced into a cell. Dicer processes the duplex RNAs into small interfering RNAs (siRNAs), which interact with the target mRNA. Again, the mRNA is either degraded or its translation is inhibited.

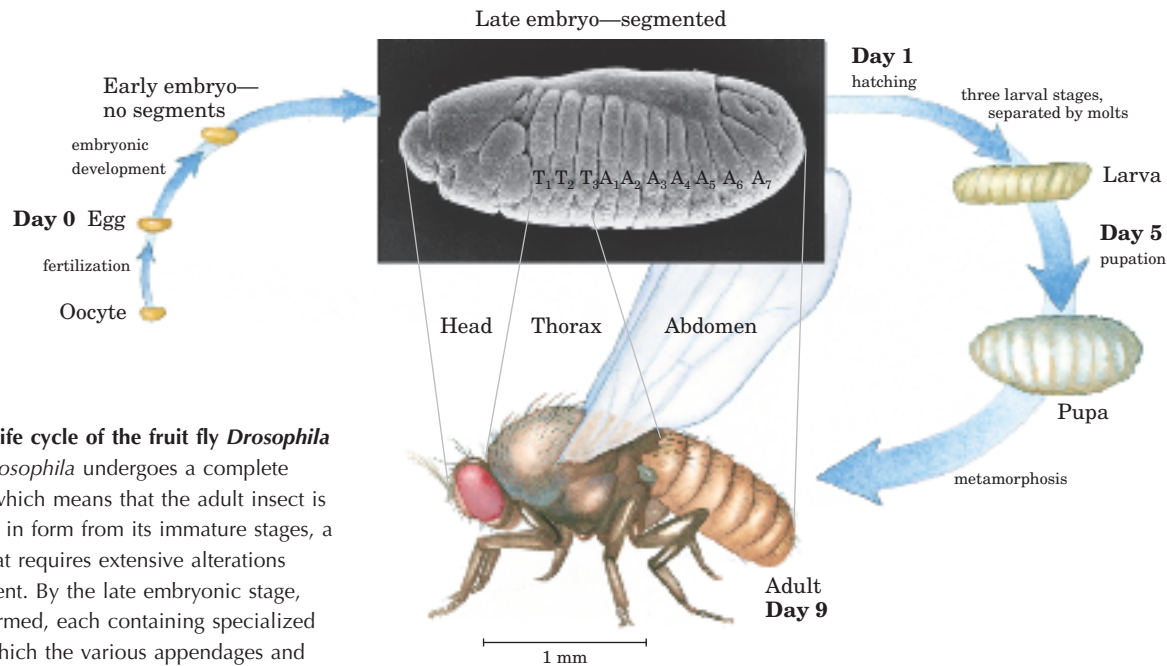


FIGURE 28-34 Life cycle of the fruit fly *Drosophila melanogaster*. *Drosophila* undergoes a complete metamorphosis, which means that the adult insect is radically different in form from its immature stages, a transformation that requires extensive alterations during development. By the late embryonic stage, segments have formed, each containing specialized structures from which the various appendages and other features of the adult fly will develop.

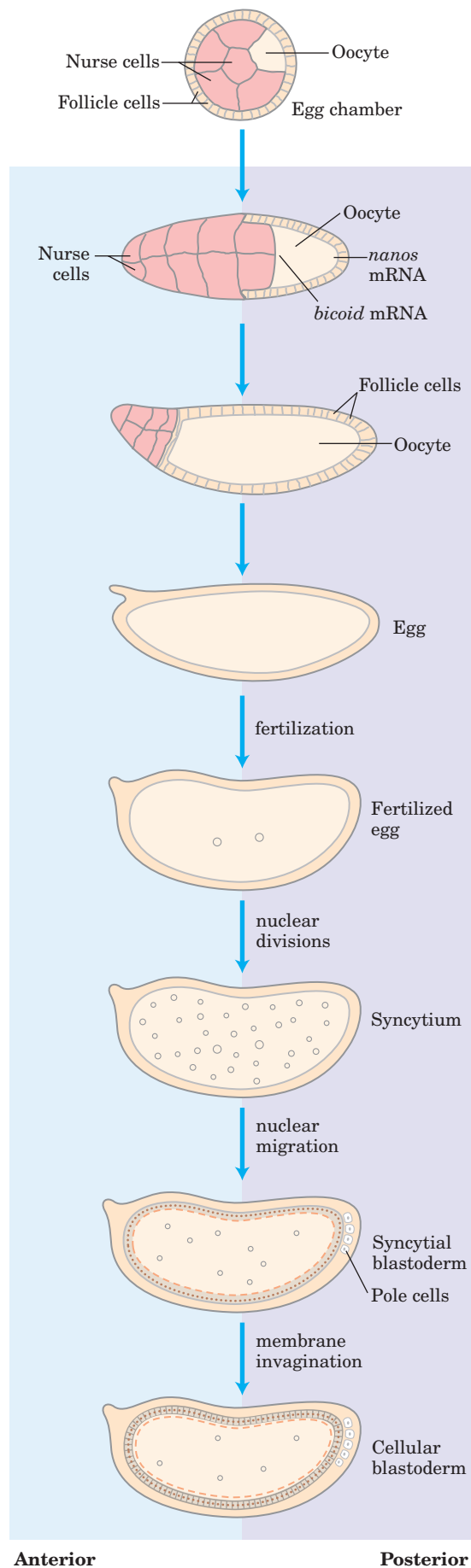
discovered that dramatically affect the organization of the body.

The *Drosophila* egg, along with 15 nurse cells, is surrounded by a layer of follicle cells (Fig. 28–35). As the egg cell forms (before fertilization), mRNAs and proteins originating in the nurse and follicle cells are deposited in the egg cell, where some play a critical role in development. Once a fertilized egg is laid, its nucleus divides and the nuclear descendants continue to divide in synchrony every 6 to 10 min. Plasma membranes are not formed around the nuclei, which are distributed within the egg cytoplasm (or syncytium). Between the eighth and eleventh rounds of nuclear division, the nuclei migrate to the outer layer of the egg, forming a monolayer of nuclei surrounding the common yolk-rich cytoplasm; this is the syncytial blastoderm. After a few additional divisions, membrane invaginations surround the nuclei to create a layer of cells that form the cellular blastoderm. At this stage, the mitotic cycles in the various cells lose their synchrony. The developmental fate of the cells is determined by the mRNAs and proteins originally deposited in the egg by the nurse and follicle cells.

Proteins that, through changes in local concentration or activity, cause the surrounding tissue to take up a particular shape or structure are sometimes referred to as **morphogens**; they are the products of pattern-regulating genes. As defined by Christiane Nüsslein-Volhard, Edward B. Lewis, and Eric F. Wieschaus, three major classes of pattern-regulating genes—maternal, segmentation, and homeotic genes—function in successive stages of development to specify the basic fea-

tures of the *Drosophila* embryo's body. **Maternal genes** are expressed in the unfertilized egg, and the resulting **maternal mRNAs** remain dormant until fertilization. These provide most of the proteins needed in very early development, until the cellular blastoderm is formed. Some of the proteins encoded by maternal mRNAs direct the spatial organization of the developing embryo at early stages, establishing its polarity. **Segmentation genes**, transcribed after fertilization, direct the formation of the proper number of body segments. At least three subclasses of segmentation genes act at successive stages: **gap genes** divide the developing embryo into several broad regions, and **pair-rule genes** together with **segment polarity genes** define 14 stripes that become the 14 segments of a normal embryo. **Homeotic genes** are expressed still later; they specify which organs and appendages will develop in particular body segments.

The many regulatory genes in these three classes direct the development of an adult fly, with a head, thorax, and abdomen, with the proper number of segments, and with the correct appendages on each segment. Although embryogenesis takes about a day to complete, all these genes are activated during the first four hours. Some mRNAs and proteins are present for only a few minutes at specific points during this period. Some of the genes code for transcription factors that affect the expression of other genes in a kind of developmental cascade. Regulation at the level of translation also occurs, and many of the regulatory genes encode translational repressors, most of which bind to the 3'UTR of the mRNA (Fig. 28–32). Because many mRNAs are



deposited in the egg long before their translation is required, translational repression provides an especially important avenue for regulation in developmental pathways.

Maternal Genes Some maternal genes are expressed within the nurse and follicle cells, and some in the egg itself. Within the unfertilized *Drosophila* egg, the maternal gene products establish two axes—anterior-posterior and dorsal-ventral—and thus define which regions of the radially symmetric egg will develop into the head and abdomen and the top and bottom of the adult fly. A key event in very early development is establishment of mRNA and protein gradients along the body axes. Some maternal mRNAs have protein products that diffuse through the cytoplasm to create an asymmetric distribution in the egg. Different cells in the cellular blastoderm therefore inherit different amounts of these proteins, setting the cells on different developmental paths. The products of the maternal mRNAs include transcriptional activators or repressors as well as translational repressors, all regulating the expression of other pattern-regulating genes. The resulting specific patterns and sequences of gene expression therefore differ between cell lineages, ultimately orchestrating the development of each adult structure.

The anterior-posterior axis in *Drosophila* is defined at least in part by the products of the *bicoid* and *nanos* genes. The *bicoid* gene product is a major anterior morphogen, and the *nanos* gene product is a major posterior morphogen. The mRNA from the *bicoid* gene is synthesized by nurse cells and deposited in the unfertilized egg near its anterior pole. Nüsslein-Volhard found that this mRNA is translated soon after fertilization, and the Bicoid protein diffuses through



Christiane Nüsslein-Volhard

FIGURE 28-35 Early development in *Drosophila*. During development of the egg, maternal mRNAs (including the *bicoid* and *nanos* gene transcripts, discussed in the text) and proteins are deposited in the developing oocyte (unfertilized egg cell) by nurse cells and follicle cells. After fertilization, the two nuclei of the fertilized egg divide in synchrony within the common cytoplasm (syncytium), then migrate to the periphery. Membrane invaginations surround the nuclei to create a monolayer of cells at the periphery; this is the cellular blastoderm stage. During the early nuclear divisions, several nuclei at the far posterior become pole cells, which later become the germ-line cells.

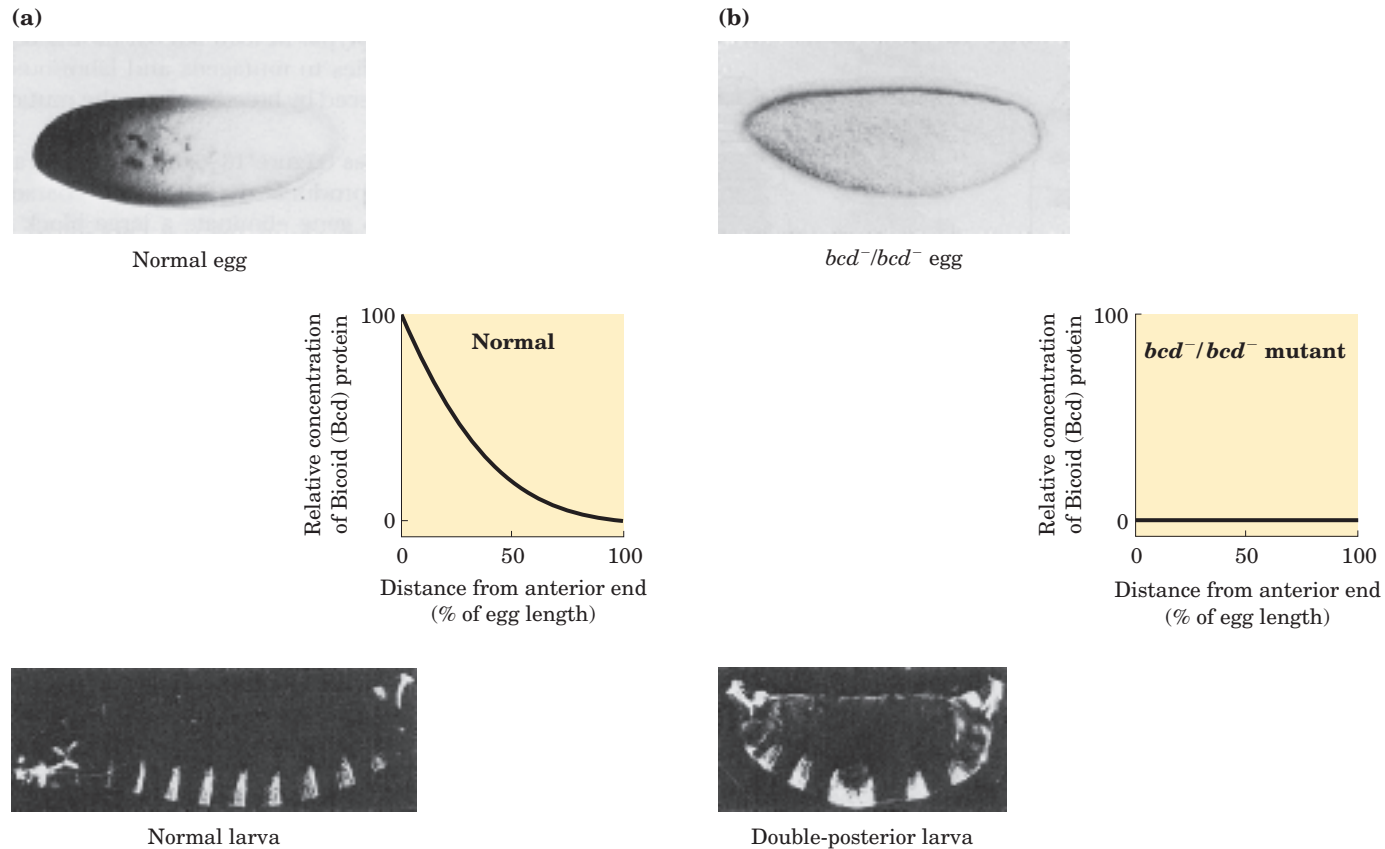


FIGURE 28-36 Distribution of a maternal gene product in a *Drosophila* egg. (a) Micrograph of an immunologically stained egg, showing distribution of the *bicoid* (*bcd*) gene product. The graph measures stain intensity. This distribution is essential for normal develop-

ment of the anterior structures of the animal. (b) If the *bcd* gene is not expressed by the mother (bcd^-/bcd^- mutant) and thus no *bicoid* mRNA is deposited in the egg, the resulting embryo has two posteriors (and soon dies).

the cell to create, by the seventh nuclear division, a concentration gradient radiating out from the anterior pole (Fig. 28-36a). The Bicoid protein is a transcription factor that activates the expression of a number of segmentation genes; the protein contains a homeodomain (p. 1090). Bicoid is also a translational repressor that inactivates certain mRNAs. The amounts of Bicoid protein in various parts of the embryo affect the subsequent expression of a number of other genes in a threshold-dependent manner. Genes are transcriptionally activated or translationally repressed only where the Bicoid protein concentration exceeds the threshold. Changes in the shape of the Bicoid concentration gradient have dramatic effects on the body pattern. Lack of Bicoid protein results in development of an embryo with two abdomens but neither head nor thorax (Fig. 28-36b); however, embryos without Bicoid will develop normally if an adequate amount of *bicoid* mRNA is injected into the egg at the appropriate end. The *nanos* gene has an analogous role, but its mRNA is deposited at the posterior end of the egg and the anterior-posterior protein gradient peaks at the posterior pole. The Nanos protein is a translational repressor.

A broader look at the effects of maternal genes reveals the outline of a developmental circuit. In addition to the *bicoid* and *nanos* mRNAs, which are deposited in the egg asymmetrically, a number of other maternal mRNAs are deposited uniformly throughout the egg cytoplasm. Three of these mRNAs encode the Pumilio, Hunchback, and Caudal proteins, all affected by *nanos* and *bicoid* (Fig. 28-37). Caudal and Pumilio are involved in development of the posterior end of the fly. Caudal is a transcriptional activator with a homeodomain; Pumilio is a translational repressor. Hunchback protein plays an important role in the development of the anterior end and is also a transcriptional regulator of a variety of genes, in some cases a positive regulator, in other cases negative. Bicoid suppresses translation of *caudal* in the anterior and also acts as a transcriptional activator of *hunchback* in the cellular blastoderm. Because *hunchback* is expressed both from maternal mRNAs and from genes in the developing egg, it is considered both a maternal and a segmentation gene. The result of the activities of Bicoid is an increased concentration of Hunchback at the anterior end of the

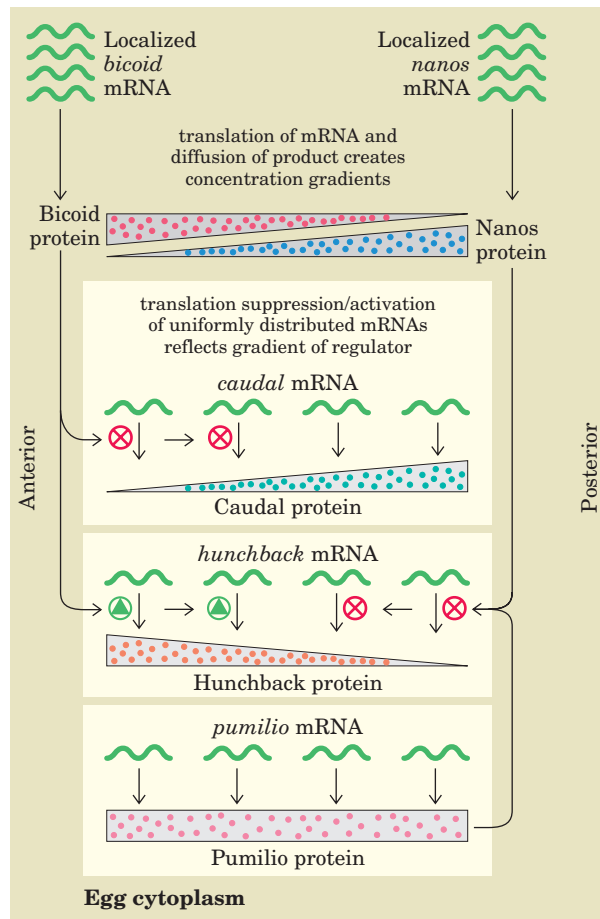


FIGURE 28-37 Regulatory circuits of the anterior-posterior axis in a *Drosophila* egg. The *bicoid* and *nanos* mRNAs are localized near the anterior and posterior poles, respectively. The *caudal*, *hunchback*, and *pumilio* mRNAs are distributed throughout the egg cytoplasm. The gradients of Bicoid (Bcd) and Nanos proteins lead to accumulation of Hunchback protein in the anterior and Caudal protein in the posterior of the egg. Because Pumilio protein requires Nanos protein for its activity as a translational repressor of *hunchback*, it functions only at the posterior end.

egg. The Nanos and Pumilio proteins act as translational repressors of *hunchback*, suppressing synthesis of its protein near the posterior end of the egg. Pumilio does not function in the absence of the Nanos protein, and the gradient of Nanos expression confines the activity of both proteins to the posterior region. Translational repression of the *hunchback* gene leads to degradation of *hunchback* mRNA near the posterior end. However, lack of Bicoid protein in the posterior leads to expression of *caudal*. In this way, the Hunchback and Caudal proteins become asymmetrically distributed in the egg.

Segmentation Genes Gap genes, pair-rule genes, and segment polarity genes, three subclasses of segmentation genes in *Drosophila*, are activated at successive

stages of embryonic development. Expression of the gap genes is generally regulated by the products of one or more maternal genes. At least some of the gap genes encode transcription factors that affect the expression of other segmentation or (later) homeotic genes.

One well-characterized segmentation gene is *fushi tarazu* (*ftz*), of the pair-rule subclass. When *ftz* is deleted, the embryo develops 7 segments instead of the normal 14, each segment twice the normal width. The Fushi-tarazu protein (Ftz) is a transcriptional activator with a homeodomain. The mRNAs and proteins derived from the normal *ftz* gene accumulate in a striking pattern of seven stripes that encircle the posterior two-thirds of the embryo (Fig. 28-38). The stripes demarcate the positions of segments that develop later; these segments are eliminated if *ftz* function is lost. The Ftz protein and a few similar regulatory proteins directly or indirectly regulate the expression of vast numbers of genes in the continuing developmental cascade.

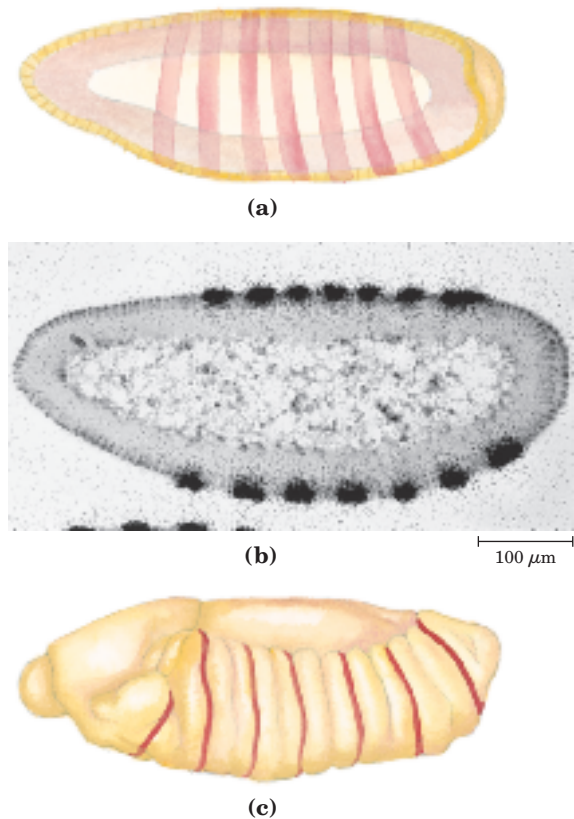


FIGURE 28-38 Distribution of the *fushi tarazu* (*ftz*) gene product in early *Drosophila* embryos. (a) In the normal embryo, the gene product can be detected in seven bands around the circumference of the embryo (shown schematically). These bands (b) appear as dark spots (generated by a radioactive label) in a cross-sectional autoradiograph and (c) demarcate the anterior margins of the segments in the late embryo (marked in red).

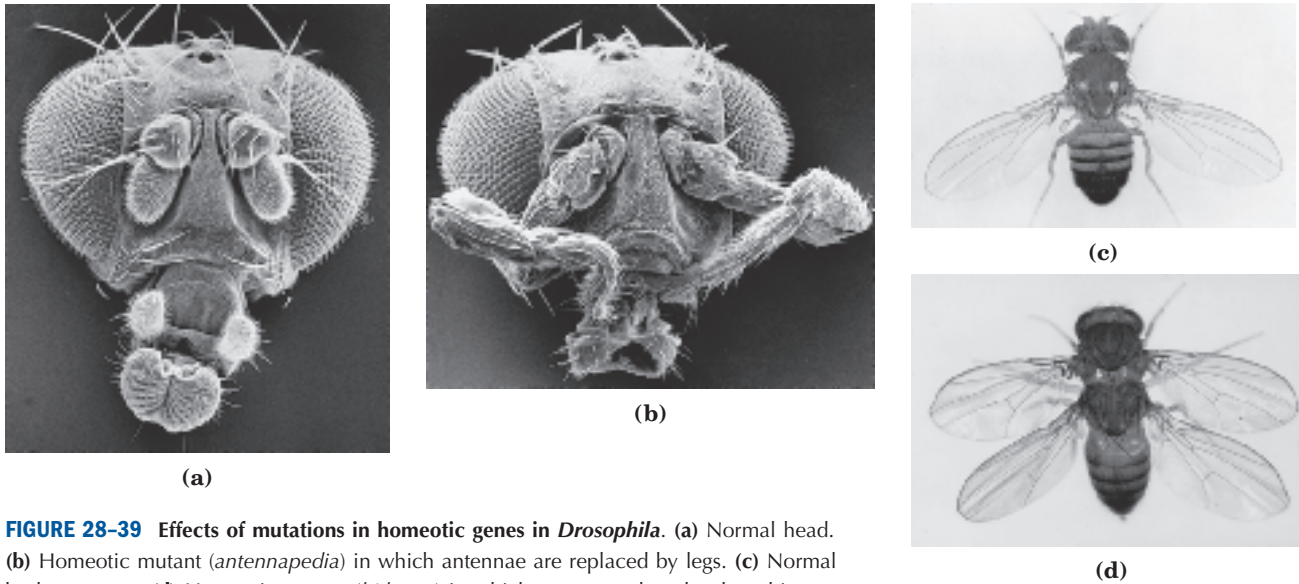


FIGURE 28-39 Effects of mutations in homeotic genes in *Drosophila*. (a) Normal head. (b) Homeotic mutant (*antennapedia*) in which antennae are replaced by legs. (c) Normal body structure. (d) Homeotic mutant (*bithorax*) in which a segment has developed incorrectly to produce an extra set of wings.

Homeotic Genes Loss of homeotic genes by mutation or deletion causes the appearance of a normal appendage or body structure at an inappropriate body position. An important example is the *ultrabithorax* (*ubx*) gene. When Ubx function is lost, the first abdominal segment develops incorrectly, having the structure of the third thoracic segment. Other known homeotic mutations cause the formation of an extra set of wings, or two legs at the position in the head where the antennae are normally found (Fig. 28–39).

The homeotic genes often span long regions of DNA. The *ubx* gene, for example, is 77,000 bp long. More than 73,000 bp of this gene are in introns, one of which is more than 50,000 bp long. Transcription of the *ubx* gene takes nearly an hour. The delay this imposes on *ubx* gene expression is believed to be a timing mechanism involved in the temporal regulation of subsequent steps in development. The Ubx protein is yet another transcriptional activator with a homeodomain (Fig. 28–13).

Many of the principles of development outlined above apply to eukaryotes from nematodes to humans. Some of the regulatory proteins themselves are conserved. For example, the products of the homeobox-containing genes *HOX 1.1* in mouse and *antennapedia* in fruit fly differ in only one amino acid residue. Of course, although the molecular regulatory mechanisms may be similar, many of the ultimate developmental events are not conserved (humans do not have wings or antennae). The discovery of structural determinants with identifiable molecular functions is the first step in understanding the molecular events underlying development. As more genes and their protein products are discovered, the biochemical side of this vast puzzle will be elucidated in increasingly rich detail.

SUMMARY 28.3 Regulation of Gene Expression in Eukaryotes

- In eukaryotes, positive regulation is more common than negative regulation, and transcription is accompanied by large changes in chromatin structure. Promoters for Pol II typically have a TATA box and Inr sequence, as well as multiple binding sites for DNA-binding transactivators. The latter sites, sometimes located hundreds or thousands of base pairs away from the TATA box, are called upstream activator sequences in yeast and enhancers in higher eukaryotes.
- Large complexes of proteins are generally required to regulate transcriptional activity. The effects of DNA-binding transactivators on Pol II are mediated by coactivator protein complexes such as TFIID or mediator. The modular structures of the transactivators have distinct activation and DNA-binding domains. Other protein complexes, including histone acetyltransferases such as GCN5-ADA2-ADA3 and ATP-dependent complexes such as SWI/SNF and NURF, reversibly remodel chromatin structure.
- Hormones affect the regulation of gene expression in one of two ways. Steroid hormones interact directly with intracellular receptors that are DNA-binding regulatory proteins; binding of the hormone has either positive or negative effects on the transcription of genes targeted by the hormone. Nonsteroid

hormones bind to cell-surface receptors, triggering a signaling pathway that can lead to phosphorylation of a regulatory protein, affecting its activity.

- Development of a multicellular organism presents the most complex regulatory challenge. The fate of cells in the early embryo is determined by establishment of anterior-posterior and dorsal-ventral gradients

of proteins that act as transcriptional transactivators or translational repressors, regulating the genes required for the development of structures appropriate to a particular part of the organism. Sets of regulatory genes operate in temporal and spatial succession, transforming given areas of an egg cell into predictable structures in the adult organism.

Key Terms

Terms in bold are defined in the glossary.

| | | | |
|--------------------------|---|--|------------------------------|
| housekeeping genes 1082 | leucine zipper 1090 | chromatin remodeling 1103 | ments (HREs) 1108 |
| induction 1082 | basic helix-loop-helix 1090 | enhancers 1104 | RNA interference (RNAi) 1111 |
| repression 1082 | catabolite repression 1093 | upstream activator sequences (UASs) 1104 | polarity 1111 |
| specificity factor 1083 | cAMP receptor protein (CRP) 1093 | basal transcription factors 1104 | metamerism 1111 |
| repressor 1083 | regulon 1094 | DNA-binding transactivators 1104 | morphogens 1112 |
| activator 1083 | transcription attenuation 1094 | coactivators 1104 | maternal genes 1112 |
| operator 1083 | translational repressor 1098 | TATA-binding protein (TBP) 1104 | maternal mRNAs 1112 |
| negative regulation 1084 | stringent response 1098 | mediator 1105 | segmentation genes 1112 |
| positive regulation 1084 | phase variation 1100 | hormone response elements (HREs) 1108 | gap genes 1112 |
| operon 1085 | hypersensitive sites 1102 | | pair-rule genes 1112 |
| helix-turn-helix 1088 | | | segment polarity genes 1112 |
| zinc finger 1088 | | | homeotic genes 1112 |
| homeodomain 1090 | | | |
| homeobox 1090 | | | |

Further Reading

General

Hershey, J.W.B., Mathews, M.B., & Sonenberg, N. (1996) *Translational Control*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Many detailed reviews cover all aspects of this topic.

Müller-Hill, B. (1996) *The lac Operon: A Short History of a Genetic Paradigm*, Walter de Gruyter, New York.

An excellent detailed account of the investigation of this important system.

Neidhardt, F.C. (ed.) (1996) *Escherichia coli and Salmonella typhimurium*, 2nd edn, Vol. 1: *Cellular and Molecular Biology* (Curtiss, R., Ingraham, J.L., Lin, E.C.C., Magasanik, B., Low, K.B., Reznikoff, W.S., Riley, M., Schaechter, M., & Umberger, H.E., vol. eds), American Society for Microbiology, Washington, DC.

An excellent source for reviews of many bacterial operons. The Web-based version, *EcoSal*, is updated regularly.

Pabo, C.O. & Sauer, R.T. (1992) Transcription factors: structural factors and principles of DNA recognition. *Annu. Rev. Biochem.* **61**, 1053–1095.

Schleif, R. (1993) *Genetics and Molecular Biology*, 2nd edn, The Johns Hopkins University Press, Baltimore.

Provides an excellent account of the experimental basis of important concepts of prokaryotic gene regulation.

Regulation of Gene Expression in Prokaryotes

Condon, C., Squires, C., & Squires, C.L. (1995) Control of rRNA transcription in *Escherichia coli*. *Microbiol. Rev.* **59**, 623–645.

Gourse, R.L., Gaal, T., Bartlett, M.S., Appleman, J.A., & Ross, W. (1996) rRNA transcription and growth rate-dependent regulation of ribosome synthesis in *Escherichia coli*. *Annu. Rev. Microbiol.* **50**, 645–677.

Jacob, F. & Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356.

The operon model and the concept of messenger RNA, first proposed in the *Proceedings* of the French Academy of Sciences in 1960, are presented in this historic paper.

Johnson, R.C. (1991) Mechanism of site-specific DNA inversion in bacteria. *Curr. Opin. Genet. Dev.* **1**, 404–411.

Kolb, A., Busby, S., Buc, H., Garges, S., & Adhya, S. (1993) Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.* **62**, 749–795.

Romby, P. & Springer, M. (2003) Bacterial translational control at atomic resolution. *Trends Genet.* **19**, 155–161.

Yanofsky, C., Konan, K.V., & Sarsero, J.P. (1996) Some novel transcription attenuation mechanisms used by bacteria. *Biochimie* **78**, 1017–1024.

Regulation of Gene Expression in Eukaryotes

Agami, R. (2002) RNAi and related mechanisms and their potential use for therapy. *Curr. Opin. Chem. Biol.* **6**, 829–834.

Bashirullah, A., Cooperstock, R.L., & Lipshitz, H.D. (1998) RNA localization in development. *Annu. Rev. Biochem.* **67**, 335–394.

Becker, P.B. & Horz W. (2002) ATP-dependent nucleosome remodeling. *Annu. Rev. Biochem.* **71**, 247–273.

Boube, M., Joulia, L., Cribbs, D.L., & Bourbon, H.M. (2002) Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell* **110**, 143–151.

Cerutti, H. (2003) RNA interference: traveling in the cell and gaining functions? *Trends Genet.* **19**, 9–46.

Conaway, R.C., Brower, C.S., & Conaway, J.W. (2002) Gene expression—emerging roles of ubiquitin in transcription regulation. *Science* **296**, 1254–1258.

Cosma, M.P. (2002) Ordered recruitment: gene-specific mechanism of transcription activation. *Mol. Cell* **10**, 227–236.

Dean, K.A., Aggarwal, A.K., & Wharton, R.P. (2002) Translational repressors in *Drosophila*. *Trends Genet.* **18**, 572–577.

DeRobertis, E.M., Oliver, G., & Wright, C.V.E. (1990) Homeobox genes and the vertebrate body plan. *Sci. Am.* **263** (July), 46–52.

Edmondson, D.G. & Roth, S.Y. (1996) Chromatin and transcription. *FASEB J.* **10**, 1173–1182.

Gingras, A.-C., Raught, B., & Sonenberg, N. (1999) eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* **68**, 913–963.

Gray, N.K. & Wickens, M. (1998) Control of translation initiation in animals. *Annu. Rev. Cell Dev. Biol.* **14**, 399–458.

Hannon, G.J. (2002) RNA interference. *Nature* **418**, 244–251.

Luger, K. (2003) Structure and dynamic behavior of nucleosomes. *Curr. Opin. Genet. Dev.* **13**, 127–135.

Mannervik, M., Nibu, Y., Zhang, H., & Levine, M. (1999) Transcriptional coregulators in development. *Science* **284**, 606–609.

Martens, J.A. & Winston, F. (2003) Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Curr. Opin. Genet. Dev.* **13**, 136–142.

McKnight, S.L. (1991) Molecular zippers in gene regulation. *Sci. Am.* **264** (April), 54–64.

A good description of leucine zippers.

Melton, D.A. (1991) Pattern formation during animal development. *Science* **252**, 234–241.

Muller, W.A. (1997) *Developmental Biology*, Springer, New York. A good elementary text.

Myers, L.C. & Kornberg, R.D. (2000) Mediator of transcriptional regulation. *Annu. Rev. Biochem.* **69**, 729–749.

Reese, J.C. (2003) Basal transcription factors. *Curr. Opin. Genet. Dev.* **13**, 114–118.

Rivera-Pomar, R. & Jackle, H. (1996) From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trends Genet.* **12**, 478–483.

Struhl, K. (1999) Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98**, 1–4.

Waterhouse, P.M. & Helliwell, C.A. (2003) Exploring plant genomes by RNA-induced gene silencing. *Nat. Rev. Genet.* **4**, 29–38.

Problems

1. Effect of mRNA and Protein Stability on Regulation *E. coli* cells are growing in a medium with glucose as the sole carbon source. Tryptophan is suddenly added. The cells continue to grow, and divide every 30 min. Describe (qualitatively) how the amount of tryptophan synthase activity in the cells changes with time under the following conditions:

- (a) The *trp* mRNA is stable (degraded slowly over many hours).
- (b) The *trp* mRNA is degraded rapidly, but tryptophan synthase is stable.
- (c) The *trp* mRNA and tryptophan synthase are both degraded rapidly.

2. Negative Regulation Describe the probable effects on gene expression in the *lac* operon of a mutation in (a) the *lac* operator that deletes most of O₁; (b) the *lacI* gene that inactivates the repressor; and (c) the promoter that alters the region around position –10.

3. Specific DNA Binding by Regulatory Proteins A typical prokaryotic repressor protein discriminates between its specific DNA binding site (operator) and nonspecific DNA by a factor of 10⁴ to 10⁶. About 10 molecules of repressor per cell are sufficient to ensure a high level of repression. Assume that a very similar repressor existed in a human cell, with a similar specificity for its binding site. How many copies of the repressor would be required to elicit a level of repression similar to that in the prokaryotic cell? (Hint: The *E. coli* genome contains about 4.6 million bp; the human haploid genome has about 3.2 billion bp.)

4. Repressor Concentration in *E. coli* The dissociation constant for a particular repressor-operator complex is very low, about 10^{–13} M. An *E. coli* cell (volume 2 × 10^{–12} mL) contains 10 copies of the repressor. Calculate the cellular concentration of the repressor protein. How does this value compare with the dissociation constant of the repressor-operator complex? What is the significance of this result?

5. Catabolite Repression *E. coli* cells are growing in a medium containing lactose but no glucose. Indicate whether each of the following changes or conditions would increase, decrease, or not change the expression of the *lac* operon. It may be helpful to draw a model depicting what is happening in each situation.

- (a) Addition of a high concentration of glucose
- (b) A mutation that prevents dissociation of the Lac repressor from the operator
- (c) A mutation that completely inactivates β -galactosidase
- (d) A mutation that completely inactivates galactoside permease
- (e) A mutation that prevents binding of CRP to its binding site near the *lac* promoter

6. Transcription Attenuation How would transcription of the *E. coli trp* operon be affected by the following manipulations of the leader region of the *trp* mRNA?

- (a) Increasing the distance (number of bases) between the leader peptide gene and sequence 2
- (b) Increasing the distance between sequences 2 and 3
- (c) Removing sequence 4
- (d) Changing the two Trp codons in the leader peptide gene to His codons
- (e) Eliminating the ribosome-binding site for the gene that encodes the leader peptide
- (f) Changing several nucleotides in sequence 3 so that it can base-pair with sequence 4 but not with sequence 2

7. Repressors and Repression How would the SOS response in *E. coli* be affected by a mutation in the *lexA* gene that prevented autocatalytic cleavage of the LexA protein?

8. Regulation by Recombination In the phase variation system of *Salmonella*, what would happen to the cell if the Hin recombinase became more active and promoted recombination (DNA inversion) several times in each cell generation?

9. Initiation of Transcription in Eukaryotes A new RNA polymerase activity is discovered in crude extracts of cells derived from an exotic fungus. The RNA polymerase initiates transcription only from a single, highly specialized promoter. As the polymerase is purified its activity declines, and the purified enzyme is completely inactive unless crude extract is added to the reaction mixture. Suggest an explanation for these observations.

10. Functional Domains in Regulatory Proteins A biochemist replaces the DNA-binding domain of the yeast Gal4 protein with the DNA-binding domain from the Lac repressor, and finds that the engineered protein no longer regulates transcription of the *GAL* genes in yeast. Draw a diagram of the different functional domains you would expect to find in the Gal4 protein and in the engineered protein. Why does the engineered protein no longer regulate transcription of the *GAL* genes? What might be done to the DNA-binding site recognized by this chimeric protein to make it functional in activating transcription of *GAL* genes?

11. Inheritance Mechanisms in Development A *Drosophila* egg that is *bcd*⁻/*bcd*⁻ may develop normally but as an adult will not be able to produce viable offspring. Explain.

Biochemistry on the Internet

12. TATA Binding Protein and the TATA Box To examine the interactions between transcription factors and DNA, go to the Protein Data Bank (www.rcsb.org/pdb) and download the PDB file 1TGH. This file models the interactions between a human TATA-binding protein and a segment of double-stranded DNA. Use the Noncovalent Bond Finder at the Chime Resources website (www.umass.edu/microbio/chime) to examine the roles of hydrogen bonds and hydrophobic interactions involved in the binding of this transcription factor to the TATA box.

Within the Noncovalent Bond Finder program, load the PDB file and display the protein in Spacefill mode and the DNA in Wireframe mode.

(a) Which of the base pairs in the DNA form hydrogen bonds with the protein? Which of these contribute to the specific recognition of the TATA box by this protein? (Hydrogen-bond length between hydrogen donor and hydrogen acceptor ranges from 2.5 to 3.3 Å.)

(b) Which amino acid residues in the protein interact with these base pairs? On what basis did you make this determination? Do these observations agree with the information presented in the text?

(c) What is the sequence of the DNA in this model and which portions of the sequence are recognized by the TATA-binding protein?

(d) Can you identify any hydrophobic interactions in this complex? (Hydrophobic interactions usually occur with interatomic distances of 3.3 to 4.0 Å.)



W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010
Houndmills, Basingstoke RG21 6XS, England
www.whfreeman.com

COVER DESIGN: VICKI TOMASELLI
COVER ILLUSTRATION: ADAM H. STEINBERG

ISBN-13: 978-0-7167-7108-1
ISBN-10: 0-7167-7108-X



EAN

9

780716 771081

