Лекция 8.
Регуляция экспрессии генов.
Система передачи сигнала

Ion Channels Are Gated by Ligands and by Membrane Potential

In a fourth class of signal transducers, receptors are coupled directly or indirectly to ion channels in the plasma membrane. The best-understood example of such a receptor is the **nicotinic acetylcholine receptor,** which responds to the neurotransmitter acetylcholine. It is found in the postsynaptic cells in certain nerve synapses (Fig. 22–34) and in the junction between a muscle fiber and the neuron that controls it. The acetylcholine receptor complex (M_r 250,000) is composed of four different polypeptide chains, one of which is present in two copies. The transmembrane arrangement of these five chains provides a hydrophilic channel through which ions can traverse the lipid bilayer. When acetylcholine released from the presynaptic nerve ending binds

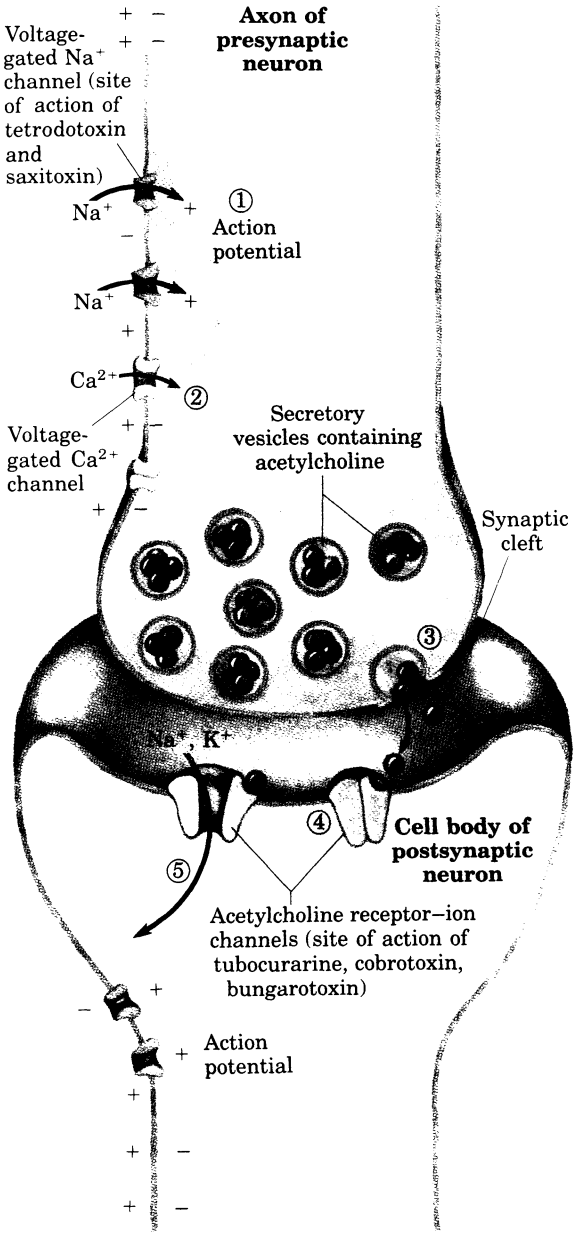
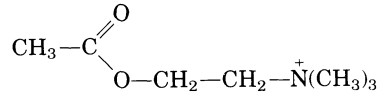


Figure 22–34 Role of voltage-gated and ligand-gated ion channels in passage of an electrical signal between two neurons. Initially, the plasma membrane of the presynaptic neuron is polarized, with the inside negative; this results from the action of the electrogenic Na^+K^+ ATPase, which pumps three Na^+ outward for every two K^+ pumped into the neuron (see Fig. 10–22). ① A stimulus to this neuron causes an action potential to move downward along its axon (white arrow). The opening of one voltage-gated Na^+ channel allows Na^+ entry, and the resulting local depolarization causes the adjacent Na^+ channel to open, and so on. The directionality of movement of the action potential is ensured by the brief refractory period that follows the opening of each voltage-gated Na^+ channel. ② When this wave of depolarization reaches the axon tip, voltage-gated Ca^{2+} channels open, allowing Ca^{2+} entry into the presynaptic neuron. ③ The resulting increase in internal $[Ca^{2+}]$ triggers exocytosis of the neurotransmitter acetylcholine into the space between the neurons (synaptic cleft). ④ Acetylcholine binds to its specific receptor in the plasma membrane of the cell body of the postsynaptic neuron, causing the ligand-gated ion channel that is part of the receptor to open. ⑤ Extracellular Na^+ and K^+ enter through this channel, depolarizing the postsynaptic cell. The electrical signal has thus passed to the postsynaptic cell, and will move along its axon to a third neuron by this same sequence of events. The effects of the toxins shown in parentheses are discussed on p. 774.



Acetylcholine

to its receptor in the postsynaptic cell (Fig. 22–34), the receptor–ion channel opens, allowing transmembrane passage of Na^+ and K^+ ions (pp. 292–293). The receptor is therefore referred to as a **ligand-gated ion channel.** The resulting depolarization of the postsynaptic membrane triggers muscle contraction or initiates an action potential in the postsynaptic neuron.

The action potential is a wave of transient depolarization that sweeps the neuron from the site of the initial stimulus (in the cell body of the neuron), along the long, thin cytoplasmic extension (axon), to the next synapse. Essential to this signaling mechanism are several types of "voltage-gated" ion channels in the plasma membrane of the neuron. These channels, formed by transmembrane proteins, open and close in response to changes in the transmembrane electrical potential. Along the entire length of the axon are **voltage-gated Na^+ channels** (Fig. 22–34), which are closed when the membrane is polarized, but open briefly when the membrane potential is reduced (i.e., during depolarization). After each opening of a Na^+ channel there follows a brief refractory period during which the channel cannot open again, and thus a unidirectional wave of depolarization sweeps from the nerve cell body toward the end of the axon.

At the distal tip of the neuron are **voltage-gated Ca^{2+} channels.** When the wave of depolarization reaches these channels they open, letting Ca^{2+} enter from the extracellular space and triggering acetylcholine release into the synaptic cleft (Fig. 22–34). Acetylcholine diffuses to the postsynaptic cell, where it binds to acetylcholine receptors; thus the message is passed to the next cell in the circuit.

Toxins, Oncogenes, and Tumor Promoters Interfere with Signal Transductions

Biochemical studies of signal transductions have led to an improved understanding of the pathological effects of toxins produced by the bacteria that cause cholera and pertussis (whooping cough). Both toxins are enzymes that interfere with normal signal transductions in the host animal. **Cholera toxin,** secreted by *Vibrio cholerae* found in contaminated drinking water, catalyzes the transfer of ADP-ribose from NAD^+ to the α subunit of G_s, blocking its GTPase activity (Fig. 22–26) and thereby rendering it permanently activated (Fig. 22–35). This results in continuous activation of the adenylate cyclase of intestinal epithelial cells, and the resultant high concentration of cAMP triggers continual secretion of Cl^-, HCO_3^-, and water into the intestinal lumen. The resulting dehydration and electrolyte loss are the major pathologies in cholera. The **pertussis toxin** produced by *Bordetella pertussis* catalyzes ADP-ribosylation of G_i, preventing GDP displacement by GTP and blocking inhibition of adenylate cyclase by G_i; this defect produces the symptoms of whooping cough, including hypersensitivity to histamines and lowered blood glucose.

The critical importance of ligand- and voltage-gated ion channels in nerve signal conduction as described above is clear from the effects of several naturally occurring toxins. **Tubocurarine,** the active component of curare (used as an arrow poison in the Amazon), and toxins from snake venoms (**cobrotoxin** and **bungarotoxin**), block the acetylcholine receptor or prevent the opening of its ion channel (Fig. 22–34). By blocking signals from nerves to muscles, these toxins cause paralysis and death. **Tetrodotoxin** (from the internal organs of puffer fish) and **saxitoxin** (produced by the marine dinoflagellate that occa-

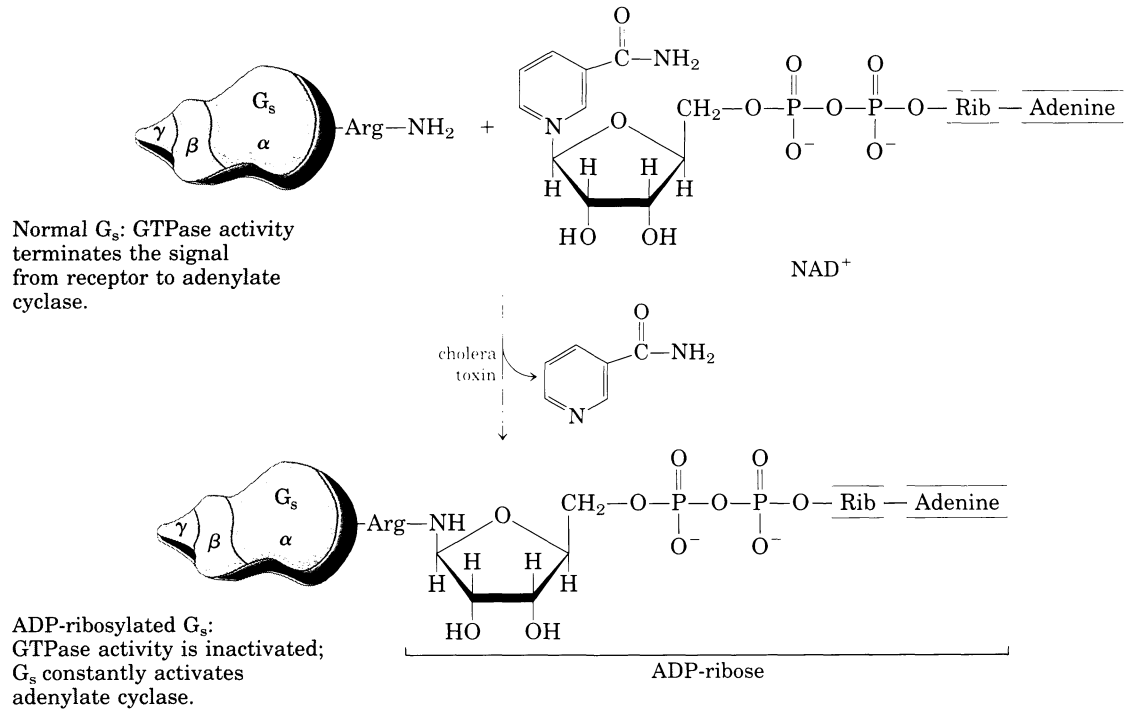


Figure 22–35 The toxins produced by the bacteria that cause cholera and whooping cough (pertussis) are enzymes that catalyze transfer of the ADP-ribose moiety of NAD^+ to an Arg residue of G proteins: G_s in the case of cholera (as shown here) and G_i in whooping cough. The G proteins thus modified fail to respond to normal hormonal stimuli. The pathology of both diseases results from defective regulation of adenylate cyclase and overproduction of cAMP.

sionally causes "red tides") are also deadly poisons, which block neurotransmission by preventing the opening of Na^+ channels.

Tumors and cancer are the result of uncontrolled cell division. Normally, cell division is highly regulated by a family of **growth factors,** proteins that cause resting cells to undergo cell division and, in some cases, differentiation. Some growth factors are cell type–specific, stimulating division of only those cells with appropriate receptors; other growth factors are more general in their effects. Among the well-studied growth factors are epidermal growth factor (EGF), nerve growth factor (NGF), fibroblast growth factor (FGF), platelet-derived growth factor (PDGF), erythropoietin, and a family of proteins called lymphokines, which includes interleukins (IL-1, IL-2, etc.) and interferon γ. There are also extracellular factors that antagonize the effects of growth factors, slowing or preventing cell division; transforming growth factor β (TGFβ) and tumor necrosis factor (TNF) are such factors.

These extracellular signals act through cell-surface receptors very similar to those for hormones, and by similar mechanisms: the production of intracellular second messengers, protein phosphorylation, and ultimately, alteration of gene expression.

It is becoming clear that many types of cancer are the result of abnormal signal-transducing proteins, which lead to continual production of the signal for cell division. The mutated genes that encode these defective signaling proteins are **oncogenes.** (Oncogenes, and gene function in general, are discussed in Chapter 25.) Oncogenes were originally discovered in tumor-causing viruses, then later found to be closely similar to or derived from genes present in the animal host cells. Most likely, these viral genes originated from normal host genes (proto-oncogenes) that encode growth-regulating proteins. During certain types of viral infections, these DNA sequences can be copied by the virus and incorporated into its genome (Fig. 22–36). At some point during the cycle of viral infection, the gene can become defective as a

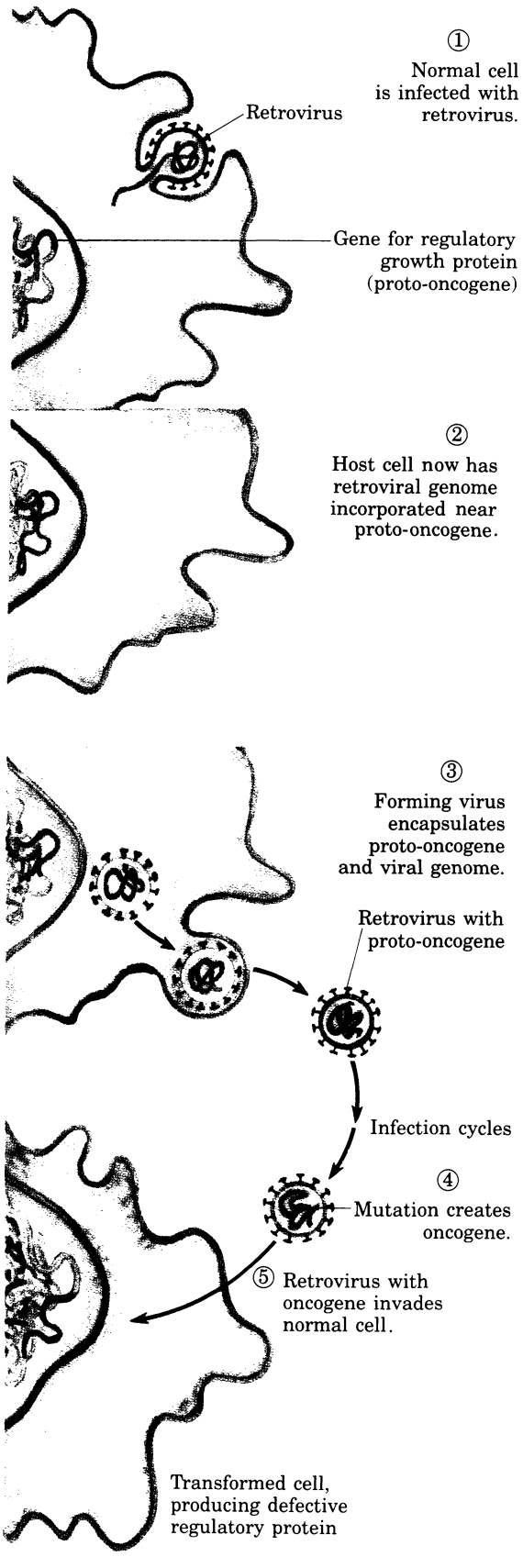


Figure 22–36 Conversion of a normal regulatory gene into a viral oncogene. ① A normal cell is infected by a retrovirus, which ② inserts its own genome into the chromosome of the host cell, near the gene for a regulatory protein (the proto-oncogene). ③ Virus particles released from the infected cell infrequently "capture" a host gene, in this case the proto-oncogene that encodes a regulatory protein. ④ During several cycles of infection, a mutation occurs in the viral proto-oncogene, converting it into an oncogene. ⑤ When the virus subsequently infects a normal cell, it introduces the oncogene into the host-cell DNA. Transcription of the oncogene leads to the production of a defective regulatory protein that continuously gives the signal for host-cell division, overriding normal mechanisms for limiting cell division. Host cells infected with oncogene-carrying viruses therefore undergo unregulated cell division—they form tumors. Proto-oncogenes can also undergo mutation to oncogenes without the intervention of a retrovirus; these cellular oncogenes also confer unregulated growth on the cells in which they occur.

result of truncation or some other mutation. During a subsequent infection, when this viral oncogene is expressed in its host cell, the abnormal protein product interferes with normal regulation of cell growth, and the unregulated growth can result in a tumor. Oncogenes can also arise from proto-oncogenes without viral involvement. Chromosomal rearrangements, chemical agents, radiation, or other factors can cause mutations in the genes that encode signal-transducing proteins. The resulting oncogenes express defective proteins and defective signaling, once again leading to tumor growth.

Many viral oncogenes encode unregulated tyrosine kinase activities, and in some cases the oncogene product is nearly identical to a normal animal-cell receptor, but with the normal signal-binding site defective or missing. For example, the *erb*B oncogene product, a protein called ErbB, is essentially identical to the normal receptor for epidermal growth factor, except that ErbB lacks the domain that normally binds EGF (Fig. 22–37, p. 777). The *erb*B2 oncogene is commonly associated with adenocarcinomas (cancers) of the breast, stomach, and ovary.

Other signal-transducing proteins with oncogene analogs are the GTP-binding (G) proteins. One well-characterized oncogene, *ras,* encodes a protein with normal GTP binding but no GTPase activity. When the Ras protein (p. 682) is produced in an animal cell, it remains always in the activated form, regardless of the signals coming through normal receptors. Again, the result is unregulated growth—cancer. Mutations in *ras* are associated with 30 to 50% of lung and colon carcinomas and over 90% of pancreatic carcinomas.

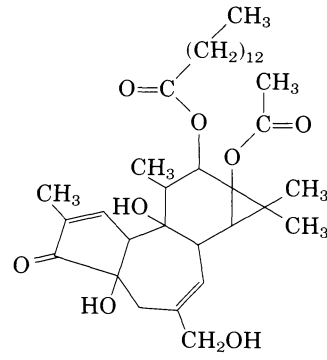
The action of a group of compounds known as **tumor promoters** can also be understood in the light of what we know of signal transduction. The best understood of these compounds, phorbol esters, are

chemically synthesized compounds that are potent activators of protein kinase C. They apparently mimic cellular diacylglycerol as second messengers (Fig. 22–32), but unlike naturally occurring diacylglycerols they are not rapidly metabolized. By permanently activating protein kinase C, these synthetic tumor promoters interfere with the normal regulation of cell growth and division.

Protein Phosphorylation and Dephosphorylation Are Central to Cellular Control

One common denominator in signal transductions—whether they involve adenylate cyclase, a transmembrane receptor–tyrosine kinase, phospholipase C, or an ion channel—is the eventual regulation of the activity of a protein kinase. We have seen examples of kinases activated by cAMP, insulin, Ca^{2+}/calmodulin, Ca^{2+}/diacylglycerol, and by phosphorylation catalyzed by another protein kinase. The number of known protein kinases has grown remarkably since their discovery by Edwin G. Krebs and Edmond H. Fischer in 1959. Hundreds of different protein kinases, each with its own specific activator and its own specific protein target(s), may be present in eukaryotic cells. Although many other types of covalent modifications are known to occur on proteins, it is clear that phosphorylations make up the vast majority of known regulatory modifications of proteins.

The addition of a phosphate group to a Ser, Thr, or Tyr residue introduces a bulky, highly charged group into a region that was only moderately polar. When the modified side chain is located in a region of the protein critical to its three-dimensional structure, phosphorylation can be expected to have dramatic effects on protein conformation and thus on the catalytic activity of the protein. As a result of evolution, the kinase-phosphorylated Ser, Thr, and/or Tyr residues of regulated proteins occur within common structural motifs (consensus sequences) that are recognized by their specific protein kinases (Table 22–9).



Myristoylphorbol acetate
(a phorbol ester)

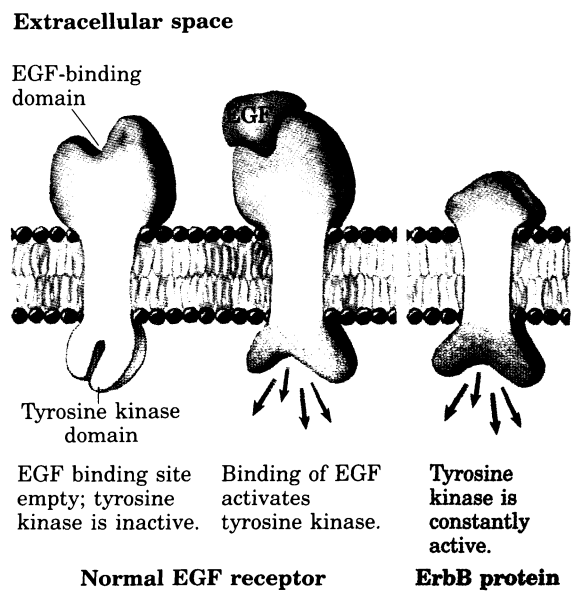


Figure 22–37 The product of the *erb*B oncogene (the ErbB protein) is a truncated version of the normal receptor for epidermal growth factor (EGF). Its intracellular domain has the structure normally induced by EGF binding, but the protein lacks the extracellular binding site for EGF. Unregulated by EGF, ErbB continuously signals cell division.

Table 22–9 Consensus sequences for protein kinases

Protein kinase	Consensus sequence*
Protein kinase A	–X–R–(R/K)–X–(S/T)–X–
Protein kinase G	–X–$(R/K)_{2-3}$–X–(S/T)–X–
Protein kinase C	–X–(R/K_{1-3}, X_{0-2})–(S/T)–(X_{0-2}, R/K_{1-3})–X–
Ca^{2+}/calmodulin kinase II	–X–R–X–X–(S/T)–X–
Phosphorylase *b* kinase	–K–R–K–Q–I–(S/T)–V–R–
Insulin receptor kinase	–T–R–D–I–Y–E–T–D–Y–Y–R–K–
EGF receptor kinase	–T–A–E–N–A–E–Y–L–R–V–A–P–

Source: Data from Kemp, B.E. & Pearson, R.B. (1990) Protein kinase recognition sequence motifs. *Trends Biochem. Sci.* **15,** 342–346; and Kennelly, P.J. & Krebs, E.G. (1991) Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J. Biol. Chem.* **266,** 15555–15558.

* (S/T) and Y are the Ser (or Thr) and Tyr residues that are phosphorylated. X is a less essential residue; any of several amino acids may be at this position. Essential residues are indicated by their one-letter abbreviations (see Table 5–1). The notation –(R/K_{1-3}, X_{0-2})– means that at this position there are from one to three amino acids, which can be R (Arg) or K (Lys), as well as zero to two of any amino acids, in any sequence (the comma indicates that no sequence is implied).

Figure 22–38 The enzyme glycogen synthase contains at least nine separate sites in five designated regions susceptible to phosphorylation by one of the cellular protein kinases. The activity of this enzyme is therefore capable of modulation in response to a variety of second messengers produced in response to different extracellular signals. Thus regulation is a matter not of binary (on/off) switching but of finely tuned modulation of the activity over a wide range.

Kinase	Glycogen synthase sites phosphorylated	Degree of synthase inactivation
cAMP-dependent protein kinase	1A, 1B, 2, 4	+
cGMP-dependent protein kinase	1A, 1B, 2	+
Phosphorylase *b* kinase	2	+
Ca^{2+}/calmodulin-dependent kinase	1B, 2	+
Glycogen synthase kinase 3	3A, 3B, 3C	+ + +
Glycogen synthase kinase 4	2	+
Casein kinase II	5	0
Casein kinase I	At least 9 sites	+ + + +
Protein kinase C	1A	+

Not all cases of regulation by phosphorylation are as simple as those we have described. Some proteins have consensus sequences recognized by several different protein kinases, each of which can phosphorylate the protein and alter its enzymatic activity. For example, glycogen synthase is inactivated by cAMP-dependent phosphorylation of specific Ser residues, and is also modulated by at least four other protein kinases that phosphorylate four other sites in the protein (Fig. 22–38). Some of the phosphorylations inhibit the enzyme more than others, and some combinations of phosphorylation are cumulative. The result of all of these regulations is the potential for extremely subtle modulation of the activity of glycogen synthase, allowing very finely tuned responses to varying metabolic circumstances.

The end effect of epinephrine's interaction with the β-adrenergic receptor is the phosphorylation of several cellular enzymes, including glycogen synthase and glycogen phosphorylase. To serve as an effective regulatory mechanism, this phosphorylation must be reversible, allowing the regulated enzymes to return to their prestimulus level when the hormonal signal stops. In muscle, for example, the enzyme phosphoprotein phosphatase-1 dephosphorylates glycogen phosphorylase, phosphorylase *b* kinase, and glycogen synthase (see Figs. 14–17, 19–15), reversing the effects of cAMP on the activities of these enzymes. This enzyme (sometimes called phosphorylase *a* phosphatase, synthase phosphatase, or kinase phosphatase to indicate its substrate specificity) is regulated by another protein, **phosphoprotein phosphatase inhibitor.** This inhibitor, when phosphorylated by protein kinase A, inhibits phosphoprotein phosphatase-1. A rise in the concentration of cAMP therefore stimulates phosphorylation of certain regulated proteins such as glycogen phosphorylase and also slows dephosphorylation of these proteins, prolonging the effect of phosphorylation.

Cells contain a family of phosphoprotein phosphatases that hydrolyze specific phosphoserine, phosphothreonine, and phosphotyrosine

esters, releasing P_i. Although this class of enzymes is not yet as thoroughly studied as the protein kinases, it is very likely that these phosphatases will turn out to be just as important as the protein kinases in regulating cellular processes and metabolism. The known phosphoprotein phosphatases show substrate specificity, acting on only a subset of phosphoproteins, and they are in some cases regulated by a second messenger or an extracellular signal. Some protein phosphatases are transmembrane proteins of the plasma membrane, with extracellular receptorlike domains and intracellular phosphatase domains; they may well prove to be regulated by extracellular signals in a fashion similar to regulation of the tyrosine kinase of the insulin receptor. The complexity and the subtlety of the regulatory mechanisms achieved by evolution strain the imagination, and the experimental challenges of discovering the full range of regulatory mechanisms remain to be met.

Steroid and Thyroid Hormones Act in the Nucleus to Change Gene Expression

The mechanism by which steroid and thyroid hormones exert their effects is fundamentally different from that for the other types of hormones. Steroid hormones (estrogen, progesterone, and cortisol, for example), too hydrophobic to dissolve readily in the blood, are carried on specific carrier proteins from the point of their release to their target tissues. In the target tissue, these hormones pass through the plasma membrane by simple diffusion and bind to specific receptor proteins in the nucleus (Fig. 22–39). The hormone–receptor complexes act by

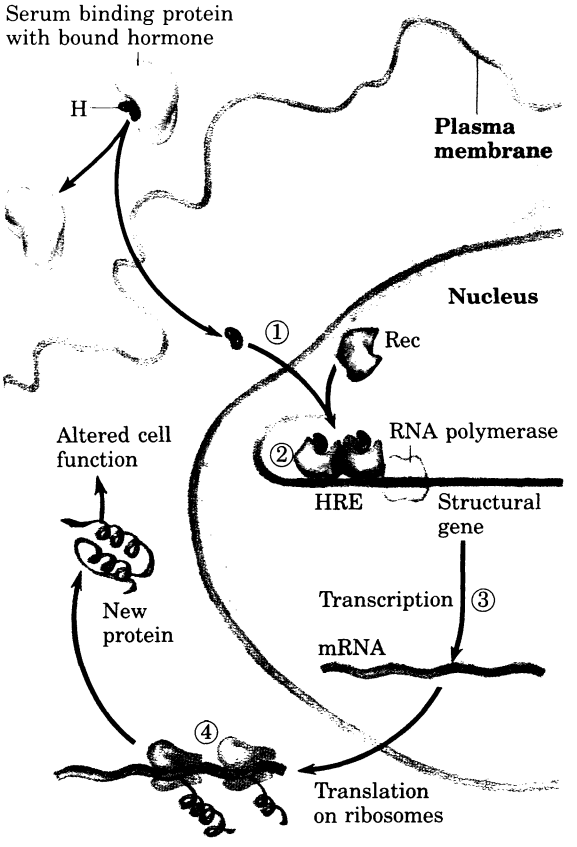


Figure 22–39 The general mechanism by which steroid and thyroid hormones, retinoids, and vitamin D act to regulate gene expression. ① Hormone (H) carried to the target tissue on serum binding proteins diffuses across the plasma membrane and binds to its specific receptor protein (Rec) in the nucleus. ② Hormone binding changes the conformation of the receptor, allowing it to form dimers in the nucleus with other hormone–receptor complexes of the same type and to bind to specific regulatory regions, hormone response elements (HREs), in the DNA adjacent to specific genes. ③ This binding somehow facilitates transcription of the adjacent gene(s) by RNA polymerase (Chapter 25), increasing the rate of messenger RNA formation and ④ bringing about new synthesis of the hormone-regulated gene product. The changed level of the newly synthesized protein produces the cellular response to the hormone. The details of protein synthesis are discussed in Chapter 26.

binding to highly specific DNA sequences called **hormone response elements** (HREs) (Fig. 22–39) and altering gene expression. Hormone binding triggers changes in the conformation of the receptor proteins so that they become capable of interacting with specific transcription factors (Chapter 27). The bound hormone–receptor complex can either enhance or suppress the expression (transcription into messenger RNA; Chapter 25) of specific genes adjacent to HREs, and thus the synthesis of the genes' protein products (Chapter 26).

The DNA sequences (HREs) to which hormone–receptor complexes bind are similar in length and arrangement, but different in sequence, for the various steroid hormones. The HRE sequences recognized by a given receptor are very similar but not identical; for each receptor there is a "consensus sequence" (Table 22–10), which the hormone–receptor complex binds at least as well as it binds the natural HREs. Each HRE consensus sequence consists of two six-nucleotide sequences, either contiguous or separated by three nucleotides. The two hexameric sequences occur either in tandem or in a palindromic arrangement (Fig. 12–20). The hormone–receptor complex binds to the DNA as a dimer, with each monomer recognizing one of the six-nucleotide sequences. The ability of a given hormone to alter the expression of a specific gene depends upon the HRE element's exact sequence and on its position relative to the gene and the number of HREs associated with the gene.

Table 22–10 Consensus sequences of some hormone response elements

Hormone	Sequence of DNA (both strands)*	
Glucocorticoid	(5′) AGAACAXXXTGTTCT (3′)	(strand 1)
	(3′) TCTTGTXXXACAAGA (5′)	(strand 2)
Estrogen	(5′) AGGTCAXXXTGACCT (3′)	(strand 1)
	(3′) TCCAGTXXXACTGGA (5′)	(strand 2)
Thyroid	(5′) AGGTCATGACCT (3′)	(strand 1)
	(3′) TCCAGTACTGGA (5′)	(strand 2)

Source: Data from Schwabe, J.W.R. & Rhodes, D. (1991) Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition. *Trends Biochem. Sci.* **16,** 291–296; and Fuller, P.J. (1991) The steroid receptor superfamily: mechanisms of diversity. *FASEB J.* **5,** 3092–3099.

* X represents any nucleotide.

Comparison of the amino acid sequences of receptors for several steroid hormones as well as receptors for thyroid hormone, vitamin D, and retinoids has revealed several highly conserved sequences and some regions in which the sequences differ considerably with receptor type (Fig. 22–40). (Retinoids are compounds related to retinoate, the carboxylate form of vitamin A_1 (see Fig. 9–18), which have hormonelike actions on some cell types.) A centrally located sequence of 66 to 68 residues is very similar in all of the receptors; this is the DNA-binding region, which resembles regions of other proteins known to bind DNA. All of these DNA-binding regions share the "zinc finger" structure (see Fig. 27–12), a sequence containing eight Cys residues that provide binding sites for two Zn^{2+} ions, which stabilize the DNA-binding domain.

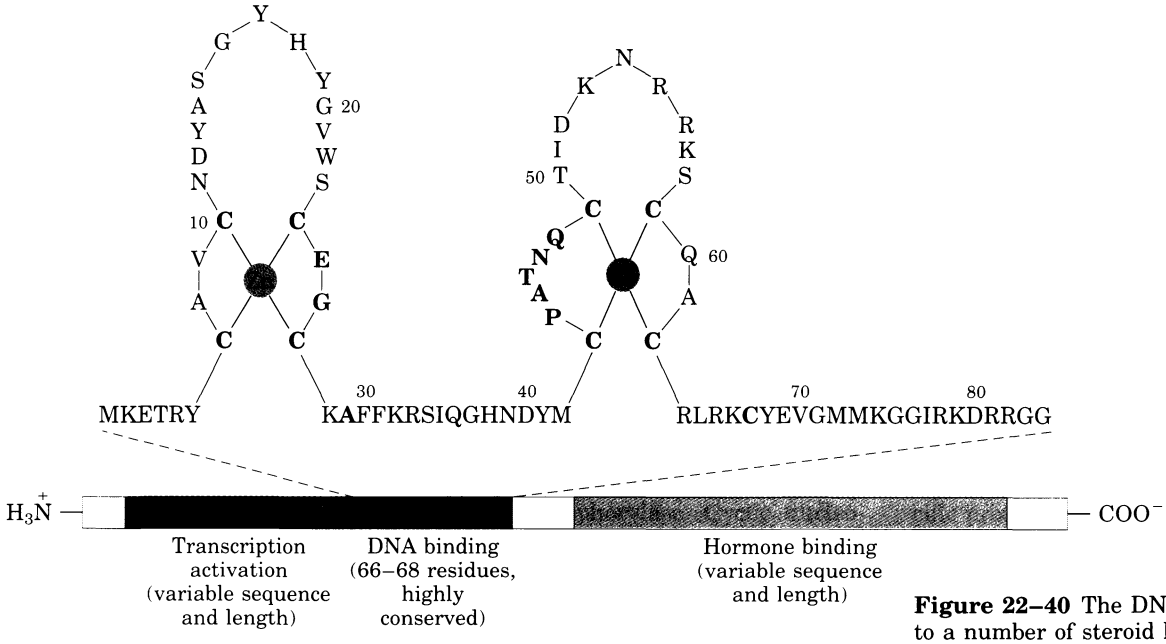


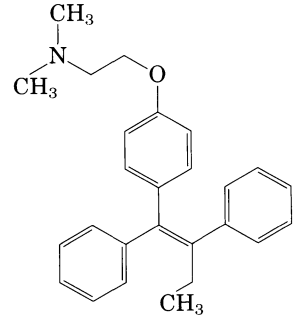
Figure 22–40 The DNA-binding domain common to a number of steroid hormone receptor proteins. These proteins have a binding site for the hormone, a DNA-binding domain, and a region that activates the transcription of the regulated gene. The DNA-binding region is highly conserved. The sequence shown here (see Table 5–1 for amino acid abbreviations) is that for the estrogen receptor, but the residues in bold type are common to all such receptors. Eight critical Cys residues bind to two Zn^{2+} ions that stabilize the "zinc finger" structure shared with many other DNA-binding proteins (see Fig. 27–12). The regulation of gene expression is described in more detail in Chapter 27.

The region of the hormone receptor responsible for hormone binding (the ligand-binding region, always at the carboxyl terminus) is quite different in different members of the hormone receptor family. The glucocorticoid receptor is only 30% homologous with the estrogen receptor and 17% homologous with the thyroid hormone receptor. In the vitamin D receptor, the ligand-binding region consists of only 25 residues, whereas it has 603 residues in the mineralocorticoid receptor. The different sequences are reflected in different specificities for hormone binding. Mutations that change one amino acid residue in this region result in loss of responsiveness to a specific hormone; some humans unable to respond to cortisol, testosterone, vitamin D, or thyroxine have been shown to have such mutations in the corresponding hormone receptor.

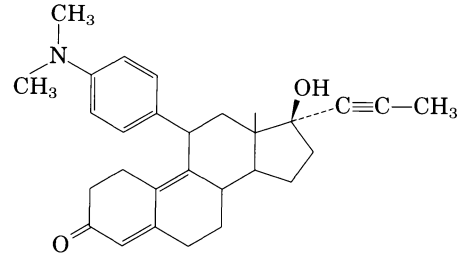
The specificity of the ligand-binding site is exploited in the use of a drug, **tamoxifen,** in the treatment of breast cancer in humans. In some types of breast cancer, division of the cancerous cells depends on the continued presence of the hormone estrogen. Tamoxifen competes with estrogen in binding to the estrogen receptor, but the tamoxifen–receptor complex is inactive in gene regulation. Consequently, tamoxifen administration after surgery or chemotherapy for this type of breast cancer slows or stops the growth of remaining cancerous cells, prolonging the life of the patient.

Another steroid analog, the drug **RU486,** is used in the very early termination of pregnancy. An antagonist of the hormone progesterone, RU486 binds to the progesterone receptor and blocks hormone actions essential to the implantation of the fertilized ovum in the uterus. As of 1992, RU486 had not been approved for use in the United States.

The ability of a given steroid or thyroid hormone to act on a specific cell type depends not only on whether the receptor for that hormone is synthesized by the cell, but also on whether the cell contains enzymes that metabolize the hormone. Some hormones (testosterone, thyroxine, vitamin D) are enzymatically converted into more active derivatives within the target cell; others, such as cortisol, are converted to an inactive form in some cells, making these cells resistant to that hormone.



Tamoxifen



RU486
(mifepristone)

In addition to the DNA-binding and ligand-binding regions, steroid receptors also have two domains that interact (in a way not fully understood) with elements of the transcriptional (RNA-synthesizing) machinery in the nucleus. The combination of DNA binding and this interaction with the transcriptional apparatus allows the steroid hormone–receptor complex to modulate the rate at which proteins are produced from a specific gene. The relatively slow action of steroid hormones (hours or days are required for their full effect) is a consequence of their mode of action; time is required for RNA synthesis in the nucleus and for the subsequent protein synthesis.

Summary

In mammals there is a division of metabolic labor among specialized tissues and organs. Coordination of the body's diverse metabolic activities is accomplished by hormonal signals that circulate in the blood. The liver is the central distributing and processing organ for nutrients. Sugars and amino acids produced in digestion cross the intestinal epithelium and enter the blood, which carries them to the liver. Some triacylglycerols derived from ingested lipids also make their way to the liver, where the constituent fatty acids are used in a variety of processes. Glucose-6-phosphate is the key intermediate in carbohydrate metabolism. It may be polymerized into glycogen, dephosphorylated to blood glucose, or converted to fatty acids via acetyl-CoA. It may undergo degradation by glycolysis and the citric acid cycle to yield ATP energy or by the pentose phosphate pathway to yield pentoses and NADPH. Amino acids are used to synthesize liver and plasma proteins, or their carbon skeletons may be converted into glucose and glycogen by gluconeogenesis; the ammonia formed by their deamination is converted into urea. Fatty acids may be converted by the liver into other triacylglycerols, cholesterol, or plasma lipoproteins for transport to and storage in adipose tissue. They may also be oxidized to yield ATP, and to form ketone bodies to be circulated to other tissues.

Skeletal muscle is specialized to produce ATP for mechanical work. During strenuous muscular activity, glycogen is the ultimate fuel and is fermented into lactate, supplying ATP. During recovery the lactate is reconverted (through gluconeogenesis) to glycogen and glucose in the liver. Phosphocreatine is an immediate source of ATP during active contraction. Heart muscle obtains all of its ATP from oxidative phosphorylation. The brain uses only glucose and β-hydroxybutyrate as fuels, the latter being important during fasting or starvation. The brain uses most of its ATP energy for the active transport of Na^+ and K^+ and the maintenance of the electrical potential of neuronal membranes. The blood links all of the organs, carrying nutrients, waste products, and hormonal signals between them.

Hormones are chemical messengers (peptides, amines, or steroids) secreted by certain tissues into the blood, serving to regulate the activity of other tissues. They act in a hierarchy of functions. Nerve impulses stimulate the hypothalamus to send specific hormones to the pituitary gland, stimulating (or inhibiting) the release of tropic hormones. The anterior pituitary hormones in turn stimulate other endocrine glands (thyroid, adrenals, pancreas) to secrete their characteristic hormones, which in turn stimulate specific target tissues.

The concentration of glucose in the blood is hormonally regulated. Fluctuations in blood glucose (which is normally about 80 mg/100 mL or 4.5 mM) due to dietary uptake or vigorous exercise are counterbalanced by a variety of hormonally triggered changes in the metabolism of several organs. Epinephrine prepares the body for increased activity by mobilizing blood glucose from glycogen and other precursors. Low blood glucose results in the release of glucagon, which stimulates glucose release from liver glycogen and shifts the fuel metabolism in liver and muscle to fatty acids, sparing glucose for use by the brain. In prolonged fasting, triacylglycerols become the principal fuels; the liver converts the fatty acids to ketone bodies for export to other tissues, including the brain. High blood glucose elicits the release of insulin, which speeds the uptake of glucose by tissues and favors the storage of fuels as glycogen and triacylglycerols. In untreated diabetes, insulin is either not produced or is not recognized by the tissues, and the utilization of blood glucose is compromised. When blood glucose levels are high, glucose is excreted intact into the urine. Tissues then depend upon fatty acids for fuel (producing ketone bodies) and degrade cellular proteins to make glucose from

their glucogenic amino acids. Untreated diabetes is characterized by high glucose levels in the blood and urine and the production and excretion of ketone bodies.

Hormones act through a small number of fundamentally similar mechanisms. Epinephrine binds to specific β-adrenergic receptors on the outer face of hepatocytes and myocytes. A stimulatory GTP-binding protein (G_s) mediates between the adrenergic receptor and adenylate cyclase on the inner face of the plasma membrane. When the adrenergic receptor is occupied, adenylate cyclase is activated and converts ATP to cAMP (the second messenger), which then activates the cAMP-dependent protein kinase. This protein kinase phosphorylates and activates inactive phosphorylase *b* kinase, which in a subsequent step phosphorylates and activates glycogen phosphorylase. Cyclic nucleotide phosphodiesterase terminates the signal by converting cAMP to AMP. The cAMP-dependent protein kinase also phosphorylates and regulates a number of other enzymes present in target tissues. (Glucagon acts by an essentially similar mechanism except that the tissue distribution of glucagon receptors is different; this hormone acts primarily on the liver.) This cascade of events, in which a single molecule of hormone activates a catalyst that in turn activates another catalyst and so on, results in large signal amplification; this is characteristic of all hormone-activated systems. Cyclic GMP acts as the second messenger for other hormones, by a similar mechanism.

Protein phosphorylation is a universal mechanism for rapid and reversible enzyme regulation. To reverse the effects of signal-stimulated protein kinases, cells contain a variety of phosphatases. These enzymes, too, are subject to regulation by extracellular and intracellular signals.

The insulin receptor represents a second signal-transducing mechanism. The receptor is an integral protein of the plasma membrane. Binding of insulin to its extracellular domain activates a tyrosine-specific protein kinase in the receptor's cytosolic domain. This kinase activates several protein kinases by phosphorylating specific Tyr residues. The phosphorylated protein kinases bring about changes in metabolism by phosphorylating additional key enzymes, altering their enzymatic activities.

A third general class of hormone mechanisms involves the coupling of hormone receptors, via another group of GTP-binding proteins, to a phospholipase C of the plasma membrane. Hormone binding activates this enzyme, which hydrolyzes inositol-containing phospholipids in the plasma membrane. This generates two second messengers: diacylglycerol, which activates protein kinase C, and inositol-1,4,5-trisphosphate (IP_3), which causes the release of Ca^{2+} sequestered in the endoplasmic reticulum. Ca^{2+} is a common second messenger in hormone-sensitive cells and in neural signaling; it alters the enzymatic activities of specific protein kinases. Calmodulin is a small Ca^{2+}-binding subunit of a number of Ca^{2+}-dependent enzymes.

The fourth general transduction mechanism triggered by hormones is the opening of hormone-sensitive ion channels. The nicotinic acetylcholine receptor is a ligand-gated ion channel, which, when occupied by acetylcholine, allows transmembrane passage of Na^+ and K^+ ions and consequent depolarization of the target cell. A wave of depolarization sweeps along nerves through the action of voltage-gated Na^+ and Ca^{2+} ion channels, triggering neurotransmitter release.

A variety of pathological conditions are associated with defects in signal-transduction mechanisms. Some bacterial toxins interfere with signal transductions. Oncogenes in a cell's DNA permit uncontrolled cell division, possibly through formation of defective signal-transducing proteins that are insensitive to modulation by growth factors or hormonal signals. Tumor promoters also interfere with cell regulation and growth.

Steroid hormones enter cells and bind to specific receptor proteins. The hormone–receptor complex binds specific regions of nuclear DNA called hormone response elements and regulates the expression of nearby genes. Tamoxifen and RU486 are drugs that act as steroid hormone antagonists.

Further Reading

General Background and History

Molecular Biology of Signal Transduction. (1988) *Cold Spring Harb. Symp. Quant. Biol.* **53.**

This entire volume is filled with short research and review papers on a wide variety of signal-transducing systems, from bacteria to humans.

Nishizuka, Y., Tanaka, C., & Endo, M. (eds) (1990) *The Biology and Medicine of Signal Transduction,* Adv. Second Messenger Phosphoprotein Res., **24.**

A collection of papers on receptor–transducer systems and the medical effects of defective signal transducers.

Sutherland, E.W. (1972) Studies on the mechanisms of hormone action. *Science* **177,** 401–408.

The author's Nobel lecture, describing the classic experiments on cAMP.

Wilson, J.D. & Foster, D.W. (eds) (1992) *Williams Textbook of Endocrinology,* 8th edn, W.B. Saunders Company, Philadelphia.

Especially relevant are Chapter 1, an introduction to hormonal regulation; Chapter 3, on the mechanism of action of steroid hormones; and Chapter 4, on the mechanisms of hormones that act at the cell surface.

Yalow, R.S. (1978) Radioimmunoassay: a probe for the fine structure of biologic systems. *Science* **200,** 1236–1245.

A history of the development of radioimmunoassays; the author's Nobel lecture.

Tissue-Specific Metabolism: Division of Labor

Arias, I.M., Jakoby, W.B., Popper, H., Schachter, D., & Shafritz, D.A. (eds) (1988) *The Liver: Biology and Pathobiology,* 2nd edn, Raven Press, New York.

An advanced-level text; includes chapters on the metabolism of carbohydrates, fats, and proteins in the liver.

Hormones: Communication among Cells and Tissues

Crapo, L. (1985) *Hormones: The Messengers of Life,* W.H. Freeman and Company, New York.

A short, entertaining account of the history and recent state of hormone research.

Snyder, S.H. (1985) The molecular basis of communication between cells. *Sci. Am.* **253** (October), 132–141.

An introductory-level discussion of the human endocrine system.

Hormonal Regulation of Fuel Metabolism

Harris, R.A. & Crabb, D.W. (1992) Metabolic interrelationships. In *Textbook of Biochemistry with Clinical Correlations,* 3rd edn (Devlin, T.M., ed), pp. 576–606, John Wiley & Sons, Inc., New York.

A description of the metabolic interplay among human tissues during normal metabolism, and the effect on tissue-specific energy metabolism of the stresses of exercise, lactation, diabetes, and renal disease.

Pilkis, S.J. & Claus, T.H. (1991) Hepatic gluconeogenesis/glycolysis: regulation and structure/function relationships of substrate cycle enzymes. *Annu. Rev. Nutr.* **11,** 465–515.

A review at the advanced level.

Roach, P.J. (1990) Control of glycogen synthase by hierarchal protein phosphorylation. *FASEB J.* **4,** 2961–2968.

Phosphorylation of one enzyme at several positions by several different protein kinases can produce finely graded changes in enzyme activity.

Molecular Mechanisms of Signal Transduction

Aaronson, S.A. (1991) Growth factors and cancer. *Science* **254,** 1146–1153.

A clear description of defects in the signal-transducing mechanisms that regulate cell division, which result from mutations in the genes for growth-factor receptors.

Becker, A.B. & Roth, R.A. (1990) Insulin receptor structure and function in normal and pathological conditions. *Annu. Rev. Med.* **41,** 99–115.

A brief description of the structure of the receptor and its gene, and a discussion of the clinical syndromes associated with receptor defects.

Berridge, M.J. (1985) The molecular basis of communication within the cell. *Sci. Am.* **253** (October), 142–152.

An introduction to the transductions mediated by adenylate cyclase, guanylate cyclase, and phospholipase C.

Berridge, M.J. & Irvine, R.F. (1989) Inositol phosphates and cell signalling. *Nature* **341,** 197–205.

Not the latest, but one of the best descriptions of the role of inositol phospholipids in signal transduction.

Brent, G.A., Moore, D.D., & Larsen, P.R. (1991) Thyroid hormone regulation of gene expression. *Annu. Rev. Physiol.* **53,** 17–36.

An advanced discussion.

Collins, S., Lohse, M.J., O'Dowd, B., Caron, M.G., & Lefkowitz, R.J. (1991) Structure and regulation of G protein-coupled receptors: the β_2-adrenergic receptor as a model. *Vitam. Horm.* **46,** 1–39.

An advanced discussion.

Fisher, S.K., Heacock, A.M., & Agranoff, B.W. (1992) Inositol lipids and signal transduction in the nervous system: an update. *J. Neurochem.* **58,** 18–38.

A review of inositol phospholipids in signaling, including a good description of the various phosphorylated derivatives of inositol and their functions as second messengers; advanced level.

Gilman, A.G. (1989) G proteins and regulation of adenylyl cyclase. *JAMA* **262,** 1819–1825.

Hille, B. (1991) *Ionic Channels of Excitable Membranes,* 2nd edn, Sinauer Associates, Sunderland, MA.

Very broad coverage, at an intermediate level.

Hollenberg, M.D. (1991) Structure-activity relationships for transmembrane signaling: the receptor's turn. *FASEB J.* **5,** 178–186.

A description of how information about the amino acid sequences of receptors, derived from cloning receptor genes, can be used to discover structural bases for receptor interactions with ligands, G proteins, and other elements of a transducing system.

Kennelly, P.J. & Krebs, E.G. (1991) Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J. Biol. Chem.* **266,** 15555–15558.

A concise summary of the sequence specificity of protein kinases.

Krebs, E.G. (1989) Role of the cyclic AMP–dependent protein kinase in signal transduction. *JAMA* **262,** 1815–1818.

A clear account of the research on protein kinase A and its history.

Linder, M.E. & Gilman, A.G. (1992) G proteins. *Sci. Am.* **267** (July), 56–65.

An introductory level description of the discovery and functions of GTP-binding proteins.

O'Malley, B.W., Tsai, S.Y., Bagchi, M., Weigel, N.L., Schrader, W.T., & Tsai, M.-J. (1991) Molecular mechanism of action of a steroid hormone receptor. *Recent Prog. Horm. Res.* **47,** 1–26.

A brief history of the discovery of steroid hormone receptors and their genes, and a review of the effects of the hormone–receptor complex on mRNA and protein synthesis in vitro.

Rasmussen, H. (1989) The cycling of calcium as an intracellular messenger. *Sci. Am.* **261** (October), 66–73.

An introduction to the role of Ca^{2+} as a second messenger.

Snyder, S.H. & Bredt, D.S. (1992) Biological roles of nitric oxide. *Sci. Am.* **266** (May), 68–77.

An intermediate-level review of the role of NO as a second messenger.

Taylor, S.S., Buechler, J.A., & Yonemoto, W. (1990) cAMP-dependent protein kinase: framework for a diverse family of regulatory enzymes. *Annu. Rev. Biochem.* **59,** 971–1005.

An advanced review of the structure and function of protein kinase A and a comparison of its activation mechanism and catalytic mechanism with those of other protein kinases.

Ulmann, A., Teutsch, G., & Philibert, D. (1990) RU 486. *Sci. Am.* **262** (June), 42–48.

The effects of this steroid antagonist, the "morning-after pill," on the female reproductive system; an introduction.

Ullrich, A. & Schlessinger, J. (1990) Signal transduction by receptors with tyrosine kinase activity. *Cell* **61,** 203–212.

A review of the common structural and functional features of receptors in the insulin receptor family.

Problems

1. *ATP and Phosphocreatine as Sources of Energy for Muscle* In contracting skeletal muscle, the concentration of phosphocreatine drops while the concentration of ATP remains fairly constant. Explain how this happens.

In a classic experiment, Robert Davies found that if the muscle is first treated with 1-fluoro-2,4-dinitrobenzene (see Fig. 5–14), the concentration of ATP in the muscle declines rapidly, whereas the concentration of phosphocreatine remains unchanged during a series of contractions. Suggest an explanation.

2. *Metabolism of Glutamate in the Brain* Glutamate in the blood flowing into the brain is transformed into glutamine, which appears in the blood leaving the brain. What is accomplished by this metabolic conversion? How does it take place? Actually, the brain can generate more glutamine than can be made from the glutamate entering in the blood. How does this extra glutamine arise? (Hint: You may want to review amino acid catabolism in Chapter 17. Recall that NH_3 is very toxic to the brain.)

3. *Absence of Glycerol Kinase in Adipose Tissue* Glycerol-3-phosphate is a key intermediate in the biosynthesis of triacylglycerols. Adipocytes, which are specialized for the synthesis and degradation of triacylglycerols, cannot directly use glycerol because they lack glycerol kinase, which catalyzes the reaction

$$\text{Glycerol} + \text{ATP} \longrightarrow \text{glycerol-3-phosphate} + \text{ADP}$$

How does adipose tissue obtain the glycerol-3-phosphate necessary for triacylglycerol synthesis? Explain.

4. *Hyperglycemia in Patients with Acute Pancreatitis* Patients with acute pancreatitis are treated by withholding protein from the diet and by intravenous administration of glucose–saline solution. What is the biochemical basis for these measures? Patients undergoing this treatment commonly experience hyperglycemia. Why?

5. *Oxygen Consumption during Exercise* A sedentary adult consumes about 0.05 L of O_2 during a 10 s period. A sprinter, running a 100 m race, consumes about 1 L of O_2 during the same time period. After finishing the race, the sprinter will continue to breathe at an elevated but declining rate for some minutes, consuming an extra 4 L of O_2 above the amount consumed by the sedentary individual.

(a) Why do the O_2 needs increase dramatically during the sprint?

(b) Why do the O_2 demands remain high after the sprint is completed?

6. *Thiamin Deficiency and Brain Function* Individuals with thiamin deficiency display a number of characteristic neurological signs: loss of reflexes, anxiety, and mental confusion. Suggest a reason why thiamin deficiency is manifested by changes in brain function.

7. *Significance of Hormone Concentration* Under normal conditions, the human adrenal medulla secretes epinephrine ($C_9H_{13}NO_3$) at a rate sufficient to maintain a concentration of 10^{-10} M in the circulating blood. To appreciate what that concentration means, calculate the diameter of a round swimming pool, with a water depth of 2 m, that would be needed to dissolve 1 g (about 1 teaspoon) of epinephrine to a concentration equal to that in blood.

8. *Regulation of Hormone Levels in the Blood* The half-life of most hormones in the blood is relatively short. For example, if radioactively labeled insulin is injected into an animal, one can determine that within 30 min half the hormone has disappeared from the blood.

(a) What is the importance of the relatively rapid inactivation of circulating hormones?

(b) In view of this rapid inactivation, how can the circulating hormone level be kept constant under normal conditions?

(c) In what ways can the organism make possible rapid changes in the level of circulating hormones?

9. *Water-Soluble versus Lipid-Soluble Hormones* On the basis of their physical properties, hormones fall into one of two categories: those that are very soluble in water but relatively insoluble in lipids (e.g., epinephrine) and those that are relatively insoluble in water but highly soluble in lipids (e.g., steroid hormones). In their role as regulators of cellular activity, most water-soluble hormones do not penetrate into the interior of their target cells. The lipid-soluble hormones, by contrast, do penetrate into their target cells and ultimately act in the nucleus. What is the correlation between solubility, the location of receptors, and the mode of action of the two classes of hormones?

10. *Hormone Experiments in Cell-Free Systems* In the 1950s, Earl Sutherland and his colleagues carried out pioneering experiments to elucidate the mechanism of action of epinephrine and glucagon. In the light of our current understanding of hormone action as described in this chapter, interpret each of the experiments described below. Identify the components and indicate the significance of the results.

(a) The addition of epinephrine to a homogenate or broken-cell preparation of normal liver resulted in an increase in the activity of glycogen phosphorylase. However, if the homogenate was first centrifuged at a high speed and epinephrine or glucagon was added to the clear supernatant fraction containing phosphorylase, no increase in phosphorylase activity was observed.

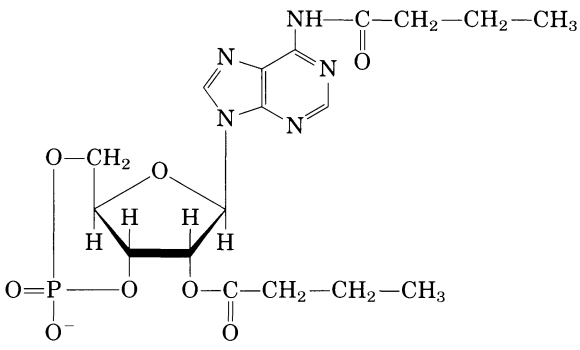
(b) When the particulate fraction sedimented from a liver homogenate by centrifugation was separated and treated with epinephrine, a new substance was produced. This substance was isolated and purified. Unlike epinephrine, this substance activated glycogen phosphorylase when added to the clear supernatant fraction of the homogenate.

(c) The substance obtained from the particulate fraction was heat-stable; that is, heat treatment did not prevent its capacity to activate phosphorylase. (Hint: Would this be the case if the substance were a protein?) The substance appeared nearly identical to a compound obtained when pure ATP was treated with barium hydroxide. (Figure 12–6 will be helpful.)

11. *Effect of Dibutyryl-cAMP versus cAMP on Intact Cells* The physiological effects of the hormone epinephrine should in principle be mimicked by the addition of cAMP to the target cells. In practice, the addition of cAMP to intact target cells elicits only a minimal physiological response. Why?

When the structurally related derivative dibutyryl-cAMP (shown below) is added to intact cells, the expected physiological responses can readily be seen. Explain the basis for the difference in cellu-

lar response to these two substances. Dibutyryl cAMP is a widely used derivative in studies of cAMP function.



Dibutyryl-cAMP

12. *Effect of Cholera Toxin on Adenylate Cyclase* The gram-negative bacterium *Vibrio cholerae* produces a protein, cholera toxin (M_r 90,000), responsible for the characteristic symptoms of cholera: extensive loss of body water and Na^+ through continuous, debilitating diarrhea. If body fluids and Na^+ are not replaced, severe dehydration will occur; untreated, the disease is often fatal. When the cholera toxin gains access to the human intestinal tract it binds tightly to specific sites in the plasma membrane of the epithelial cells lining the small intestine, causing adenylate cyclase to undergo activation that persists for hours or days.

(a) What is the effect of cholera toxin on the level of cAMP in the intestinal cells?

(b) Based on the information above, can you suggest how cAMP normally functions in intestinal epithelial cells?

(c) Suggest a possible treatment for cholera.

13. *Metabolic Differences in Muscle and Liver in a "Fight or Flight" Situation* During a "fight or flight" situation, the release of epinephrine promotes glycogen breakdown in the liver, heart, and skeletal muscle. The end product of glycogen breakdown in the liver is glucose. In contrast, the end product in skeletal muscle is pyruvate.

(a) Why are different products of glycogen breakdown observed in the two tissues?

(b) What is the advantage to the organism during a "fight or flight" condition of having these specific glycogen breakdown routes?

14. *Excessive Amounts of Insulin Secretion: Hyperinsulinism* Certain malignant tumors of the pancreas cause excessive production of insulin by the β cells. Affected individuals exhibit shaking and trembling, weakness and fatigue, sweating, and hunger. If this condition is prolonged, brain damage occurs.

(a) What is the effect of hyperinsulinism on the metabolism of carbohydrate, amino acids, and lipids by the liver?

(b) What are the causes of the observed symptoms? Suggest why this condition, if prolonged, leads to brain damage.

15. *Thermogenesis Caused by Thyroid Hormones* Thyroid hormones are intimately involved in regulating the basal metabolic rate. Liver tissue of animals given excess thyroxine shows an increased rate of O_2 consumption and increased heat output (thermogenesis), but the ATP concentration in the tissue is normal. Different explanations have been offered for the thermogenic effect of thyroxine. One is that excess thyroid hormone causes uncoupling of oxidative phosphorylation in mitochondria. How could such an effect account for the observations? Another explanation suggests that the thermogenesis is due to an increased rate of ATP utilization by the thyroid-stimulated tissue. Is this a reasonable explanation? Why?

16. *Function of Prohormones* What are the possible advantages in the synthesis of hormones as prohormones or preprohormones?

17. *Action of Aminophylline* Aminophylline, a purine derivative resembling theophylline of tea, is often administered together with epinephrine to individuals with acute asthma. What is the purpose and biochemical basis for this treatment?

CHAPTER

Regulation of Gene Expression

Of the 4,000 genes in the typical bacterial genome or the estimated 100,000 genes in the human genome, only a fraction are expressed at any given time. Some gene products have functions that mandate their presence in very large amounts. The elongation factors required for protein synthesis, for example, are among the most abundant proteins in bacteria. Other gene products are needed in much smaller amounts; for instance, a cell may contain only a few molecules of the enzymes that repair rare DNA lesions. Requirements for a given gene product may also change with time. The need for enzymes in certain metabolic pathways may wax or wane as food sources change or are depleted. During development in a multicellular eukaryote, some proteins that influence cellular differentiation are present for only a brief time in a small subset of an organism's cells. The specialization of some cells for particular functions can also dramatically affect the need for various gene products, one example being the uniquely high concentration of hemoglobin in erythrocytes.

The regulation of gene expression is a critical component in regulating cellular metabolism and in orchestrating and maintaining the structural and functional differences that exist in cells during development. Given the high energetic cost of protein synthesis, regulation of gene expression is essential if the cell is to make optimal use of available energy.

Regulating the concentration of a cellular protein involves a delicate balance of many processes. There are at least six potential points at which the amount of protein can be regulated (Fig. 27–1): synthesis of the primary RNA transcript, posttranscriptional processing of mRNA, mRNA degradation, protein synthesis (translation), posttranslational modification of proteins, and protein degradation. The concentration of a given protein is controlled by regulatory mechanisms at any or all of these points. Some of these mechanisms have been examined in previous chapters. Posttranscriptional modification of mRNAs by processes such as differential splicing (p. 873) or RNA editing (see Box 26–1) can affect which proteins are produced from an mRNA transcript and in what amounts. A variety of sequences can affect the rate at which an mRNA is degraded (p. 880). Many factors that affect the rate at which an mRNA is translated into a protein, as well as the posttranslational modification and eventual degradation of that protein, were described in Chapter 26.

Our primary focus in this chapter is the regulation of transcription initiation (although some aspects of the regulation of translation will

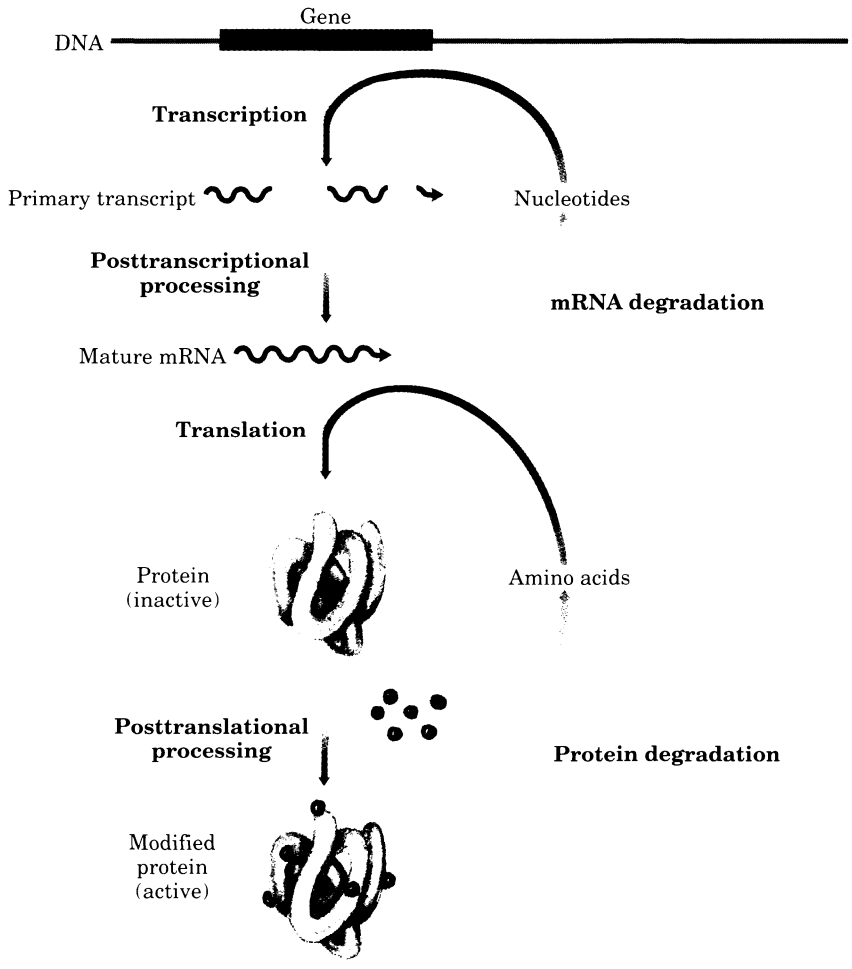


Figure 27–1 Six processes that affect the steady-state concentration of a protein. Each of these processes is a potential point of regulation.

be described). Of all the processes illustrated in Figure 27–1, regulation at the level of transcription initiation is the best documented and may be the most common. At least one important reason is clear: as for all biosynthetic pathways, the most efficient place for regulation is the first reaction in the pathway. In this way, unnecessary biosynthesis can be halted before energy is invested. Transcription initiation also is an excellent point at which to coordinate the regulation of multiple genes whose products have interdependent activities. For example, when DNA is heavily damaged, bacterial cells require a coordinated increase in the levels of many enzymes involved in DNA repair. Perhaps the most sophisticated form of coordination occurs in the complex regulatory circuits that guide the development of multicellular eukaryotes.

In this chapter, we first describe the interactions between proteins and DNA that are the key to transcriptional regulation. Specific proteins that regulate the expression of specific genes will then be discussed, first for prokaryotes and then for eukaryotes. In the course of this discussion we will examine several different mechanisms by which cells regulate gene expression and coordinate the expression of multiple genes.

Gene Regulation: Principles and Proteins

Just as the cellular requirements for different proteins vary, the mechanisms by which their respective genes are regulated also vary. The degree and type of regulation naturally reflect the function of the protein product of the gene. Some gene products are required all the time and their genes are expressed at a more or less constant level in virtually all the cells of a species or organism. Many of the genes for enzymes that catalyze steps in central metabolic pathways such as the citric acid cycle fall into this category. These genes are often referred to as **housekeeping genes.** Constant, seemingly unregulated expression of a gene is called **constitutive** gene expression. The amounts of other gene products rise and fall in response to molecular signals. Gene products that increase in concentration under prescribed molecular circumstances are referred to as inducible, and the process of increasing the expression of the gene is called **induction.** The expression of many genes encoding DNA repair enzymes, for example, is induced in response to 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 decrease in gene expression is called **repression.** For example, the presence of ample supplies of the amino acid tryptophan leads to repression of the genes for the enzymes catalyzing tryptophan biosynthesis in bacteria.

Transcription is mediated and regulated by protein–DNA interactions. The central component is RNA polymerase, an enzyme described in some detail in Chapter 25. We begin here with a further description of RNA polymerase from the standpoint of regulation, then proceed to a general description of the proteins that modulate the activity of RNA polymerase. Finally we discuss the molecular basis for the recognition of specific DNA sequences by DNA-binding proteins.

The Activity of RNA Polymerase Is Regulated

RNA polymerases bind to DNA and initiate transcription at specific sites in the DNA called promoters (Chapter 25). Promoters generally are found very near the position where RNA synthesis begins on the DNA template. The regulation of transcription initiation is, in effect, regulation of the interaction of RNA polymerase with its promoter.

Promoters vary considerably in their nucleotide sequence, and this affects the binding affinity of RNA polymerases. The binding affinity in turn affects the frequency of transcription initiation. In *E. coli,* some genes are transcribed once each second whereas others are transcribed less than once per cell generation. Much of this variation is accounted for simply by differences in promoter sequences. In the absence of regulatory proteins, differences in the sequences of two promoters may affect the frequency of transcription initiation by factors of 1,000 or more. Recall (see Fig. 25–5) that *E. coli* promoters have a consensus sequence (Fig. 27–2). Promoters that exactly match the consensus se-

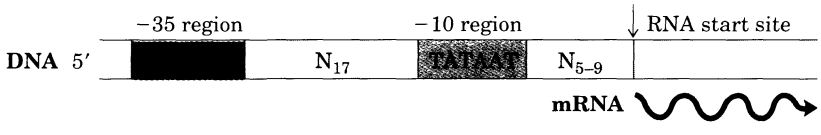


Figure 27–2 Consensus sequence for many *E. coli* promoters. N indicates any nucleotide. Most base substitutions in the −10 and −35 regions have a negative effect on promoter function. (Recall from Chapter 25 that by convention, DNA sequences are shown as they occur on the coding (nontemplate) strand.)

quence generally have the highest affinity for RNA polymerase and the highest frequency of transcription initiation. Mutations that change a consensus base pair to a nonconsensus pair generally decrease promoter function: mutations that change a nonconsensus base pair to a consensus pair usually enhance promoter function.

Although housekeeping genes are expressed constitutively, the proteins they encode are present in widely varying amounts. For these genes the RNA polymerase–promoter interaction is the only factor affecting transcription initiation, and differences in promoter sequences allow the cell to maintain the required level of each housekeeping protein.

Transcription initiation at the promoters of many genes that do not fall in the housekeeping category is further regulated in response to molecular signals. These promoters have a basal rate of transcription initiation (determined by the promoter sequence), superimposed on which is regulation mediated by several types of regulatory proteins. These proteins affect the interaction between RNA polymerase and the promoters.

Transcription Initiation Is Regulated by Proteins Binding to or near Promoters

At least three types of proteins regulate transcription initiation by RNA polymerase: (1) **specificity factors** alter the specificity of RNA polymerase for a given promoter or set of promoters; (2) **repressors** bind to a promoter, blocking access of RNA polymerase to the promoter; (3) **activators** bind near a promoter, enhancing the RNA–promoter interaction.

We encountered prokaryotic specificity factors in Chapter 25, although they were not given that name. The σ subunit (M_r 70,000) called σ^{70} of the *E. coli* RNA polymerase holoenzyme is a prototypical specificity factor that mediates specific promoter recognition and binding. Under some conditions, notably when the bacteria are subjected to heat stress, σ^{70} is replaced with another specificity factor (M_r 32,000) called σ^{32} (p. 863). When bound to σ^{32}, RNA polymerase does not bind to the standard *E. coli* promoters (Fig. 27–2), but instead is directed to a specialized set of promoters with the sequence structure shown in Figure 27–3. The promoters control the expression of a set of genes that make up the heat-shock response. Altering the polymerase to direct it to different promoters is one mechanism by which a set of related genes can be coordinately regulated. Other mechanisms will be encountered throughout this chapter.

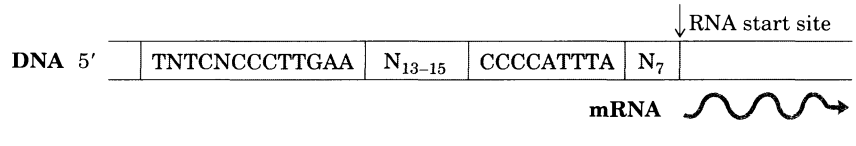
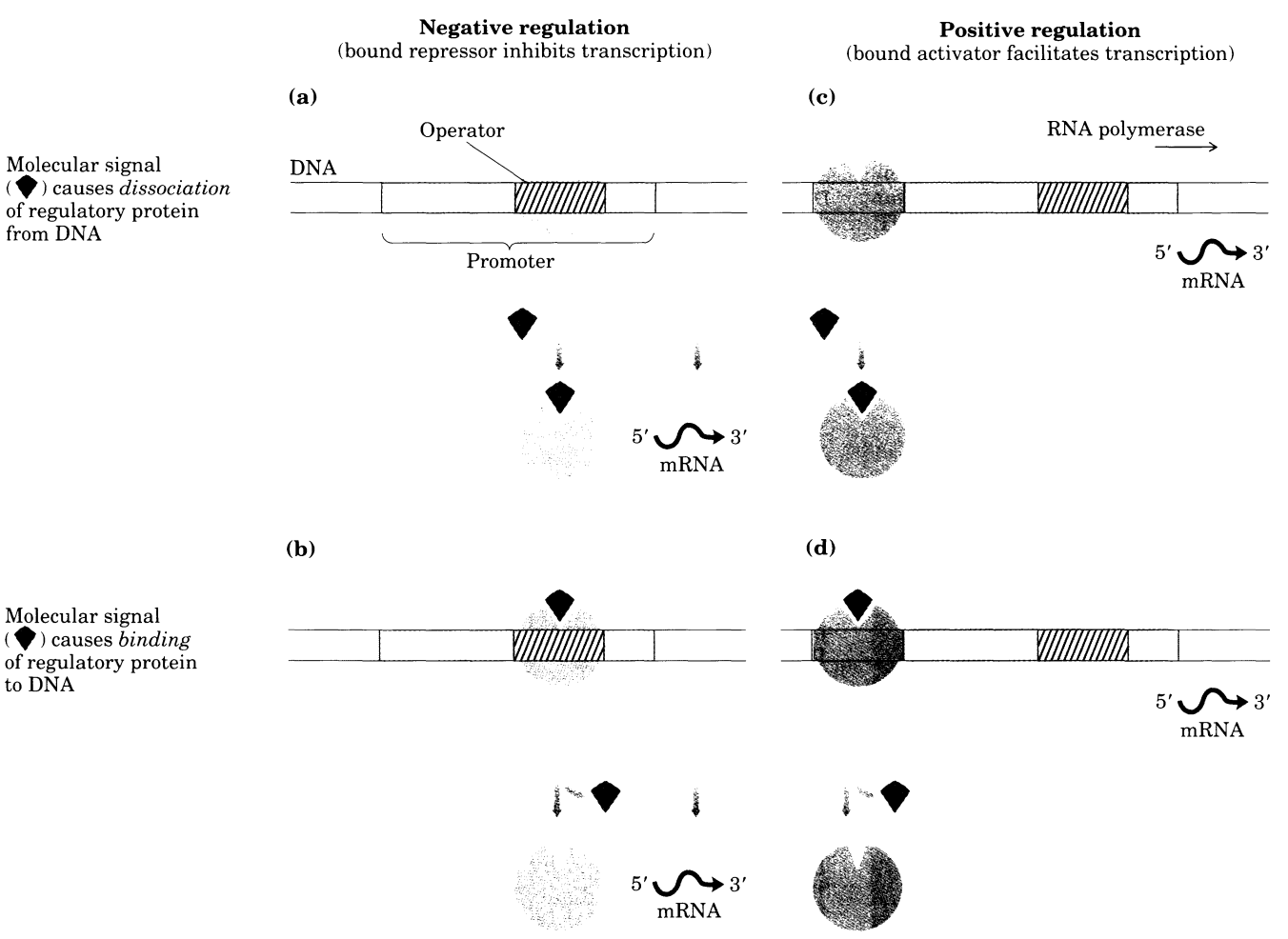


Figure 27–3 Consensus sequence for promoters that regulate the expression of genes involved in the heat-shock response in *E. coli*. This system responds to temperature increases as well as some other environmental stresses, and it involves the induction of a set of proteins. Binding of RNA polymerase to heat-shock promoters is mediated by a specialized σ subunit of the enzyme called σ^{32}, which replaces σ^{70}.

Repressors bind to specific sites in the DNA. In prokaryotes, the binding sites for repressors are called **operators.** Operator sites are generally near and often overlap the promoter so that RNA polymerase binding, or its movement along the DNA after binding, is blocked whenever the repressor is present. Regulation by means of a repressor



protein that binds to DNA and blocks transcription is referred to as **negative regulation.** Repressor binding is regulated by a molecular signal, usually a specific small molecule that binds to and induces a conformational change in the repressor. The interaction between repressor and signal molecule may lead to either an increase or a decrease in transcription. In some cases the conformational change results in dissociation of a DNA-bound repressor from the operator (Fig. 27–4a). Transcription initiation can then proceed unhindered. In other cases the interaction between an inactive repressor and the signal molecule causes the repressor to bind to the operator (Fig. 27–4b).

Activators provide a molecular counterpoint to repressors. Regulation mediated by an activator is called **positive regulation.** Activators bind to sites adjacent to a promoter and enhance the binding and activity of RNA polymerase at that promoter. The binding sites for activators are often found adjacent to promoters that are normally bound weakly or not at all by RNA polymerase. Transcription at these genes is therefore often negligible in the absence of activator. Sometimes the activator is normally bound to DNA and dissociates when it binds to the signal molecule, often a specific small molecule or another protein (Fig. 27–4c). When bound to the DNA, the activator protein facilitates RNA polymerase binding and increases the rate of transcription initiation. In other cases the activator is not bound to the DNA until it also binds to a molecular signal (Fig. 27–4d). Positive regulation is particularly common in eukaryotes, as we shall see. We now turn to a fundamental unit of gene expression, the study of which gave rise to much of our current understanding of the regulation of gene expression.

Figure 27–4 Common patterns of regulation of transcription initiation. Two types of negative regulation are illustrated. **(a)** The repressor (red) is bound to the operator in the absence of the molecular signal; the signal causes dissociation of the repressor to permit transcription. **(b)** The repressor is bound in the presence of the signal; the repressor dissociates and transcription ensues when the signal is removed. Positive regulation is mediated by gene activators. **(c)** The activator (green) binds in the absence of the molecular signal and transcription proceeds; the activator dissociates and transcription is inhibited when the signal is added. **(d)** The activator binds in the presence of the signal; it dissociates only when the signal is removed. Note that "positive" and "negative" regulation are defined by the type of regulatory protein involved. In either case the addition of the molecular signal may increase or decrease transcription, depending on the effect of the signal on the regulatory protein.

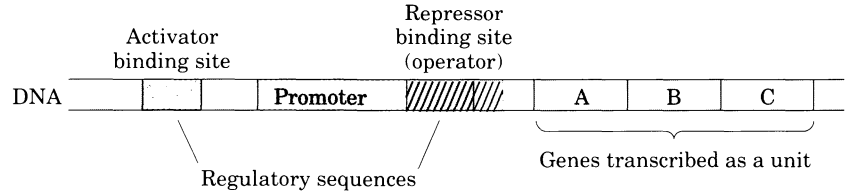


Figure 27–5 An operon. Genes A, B, and C are transcribed on one polycistronic mRNA. Typical regulatory sequences include binding sites for proteins that either activate or repress transcription from the promoter.

Many Prokaryotic Genes Are Regulated in Units Called Operons

Bacteria have a simple general mechanism for coordinating the regulation of genes whose products are involved in related processes: the genes are clustered on the chromosome and transcribed together. Most prokaryotic mRNAs are polycistronic. The single promoter required to initiate transcription of the cluster is the point where expression of all of the genes is regulated. The gene cluster, the promoter, and additional sequences that function in regulation are together called an **operon** (Fig. 27–5). Operons that include 2 to 6 genes transcribed as a unit are common; some operons contain 20 or more genes.

Many of the principles guiding the regulation of gene expression in bacteria were defined by studies of the regulation of lactose metabolism in *E. coli*. The disaccharide lactose can be used as the sole carbon source for the growth of *E. coli*. In 1960, François Jacob and Jacques Monod published a short paper in the *Proceedings of the French Academy of Sciences* demonstrating that two genes involved in lactose metabolism were coordinately regulated by a genetic element located adjacent to them. The genes were those for β-galactosidase, which cleaves lactose to galactose and glucose, and galactoside permease, which transports lactose into the cell (Fig. 27–6). The terms operon and operator were first introduced in this paper. The operon model that evolved from this and subsequent studies permitted biochemists to think about gene regulation in molecular terms for the first time.

The *lac* Operon Is Subject to Negative Regulation

The model for regulation of the lactose *(lac)* operon deduced from these studies is shown in Figure 27–7; it follows the pattern outlined in Figure 27–4a. In addition to the genes for β-galactosidase (Z) and galactoside permease (Y), the operon includes a gene for thiogalactoside transacetylase (A), whose physiological function is unknown. Each of the three genes is preceded by translational signals (not shown in Fig. 27–7) to guide ribosome binding and protein synthesis (Chapter 26). In

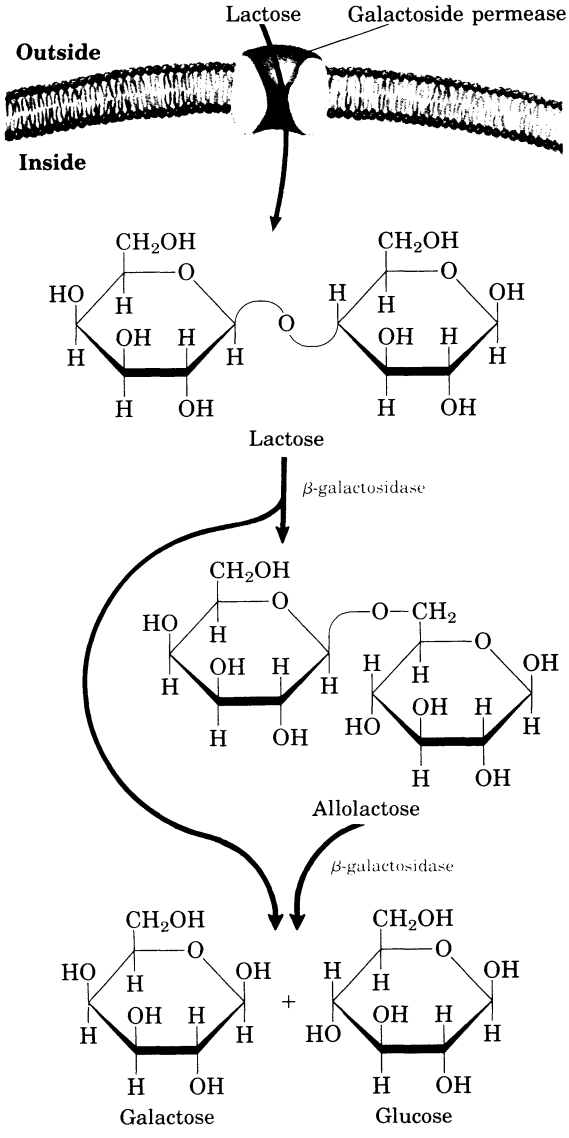


Figure 27–6 The activities of galactoside permease and β-galactosidase in lactose metabolism in *E. coli*. The conversion of lactose to allolactose by transglycosylation is a minor reaction catalyzed by β-galactosidase.

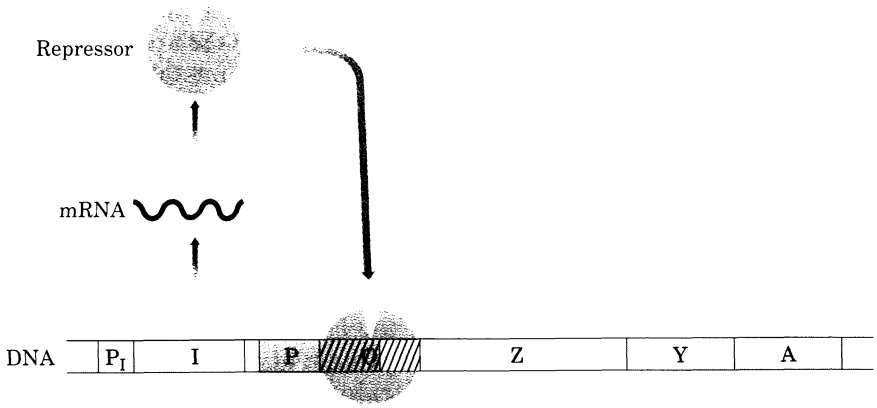


Figure 27–7 The *lac* operon in the repressed state. The I gene encodes the Lac repressor. The *lac* Z, Y, and A genes encode β-galactosidase, galactoside permease, and transacetylase, respectively. The P and O sites are the promoter and operator for the *lac* genes, respectively. The P_I site is the promoter for the I gene.

the absence of the substrate lactose, the *lac* operon genes are repressed, and β-galactosidase is present in only a few copies (a few molecules) per cell. Jacob and Monod found that mutations in the operator or in another gene called I led to constitutive synthesis of the *lac* operon gene products. When the I gene was defective, repression could be restored by introducing a functional I gene to the cell on another DNA molecule. This showed that the I gene encoded a diffusible molecule that caused gene repression; the molecule was later shown to be a protein, now called the Lac repressor. Repression is not absolute. Even in the repressed state each cell has a few copies of β-galactosidase and galactoside permease, presumably synthesized on the rare occasions when the repressor briefly dissociates from its DNA binding site (the operator).

When cells are provided with lactose, the *lac* operon is induced. An inducer molecule binds to a specific site on the repressor causing a conformational change in the repressor that results in its dissociation from the operator (Fig. 27–8). The inducer in this system is not lactose itself but an isomer of lactose called allolactose (Fig. 27–6). Lactose entering the *E. coli* cell is converted to allolactose in a reaction catalyzed by the few copies of β-galactosidase in the cell. Allolactose then binds to the Lac repressor. After the repressor dissociates, the *lac* operon genes are expressed and the concentration of β-galactosidase increases by a factor of 1,000.

Jacques Monod

François Jacob

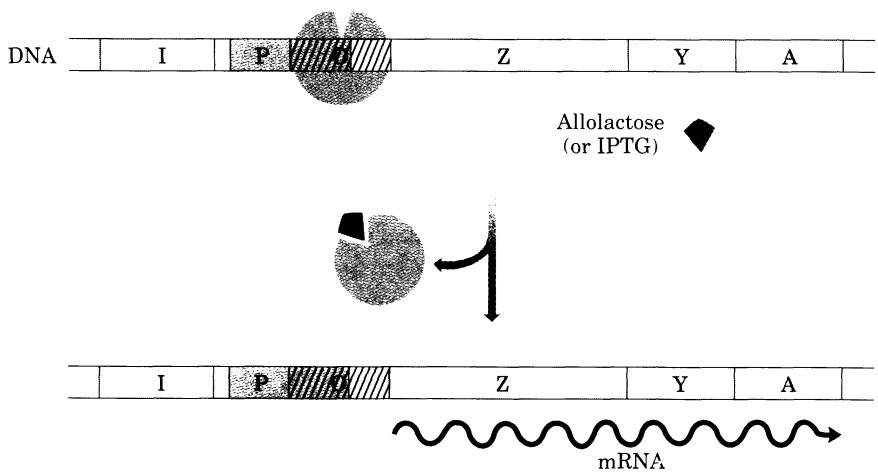
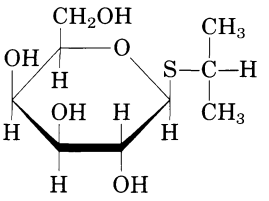


Figure 27–8 Induction of the *lac* operon in response to a molecular signal. Binding of allolactose to the Lac repressor causes a conformational change. The repressor dissociates from the operator, allowing transcription to proceed. Other β-galactosides, such as isopropylthiogalactoside (IPTG), can also act as inducers.



Isopropylthiogalactoside (IPTG)

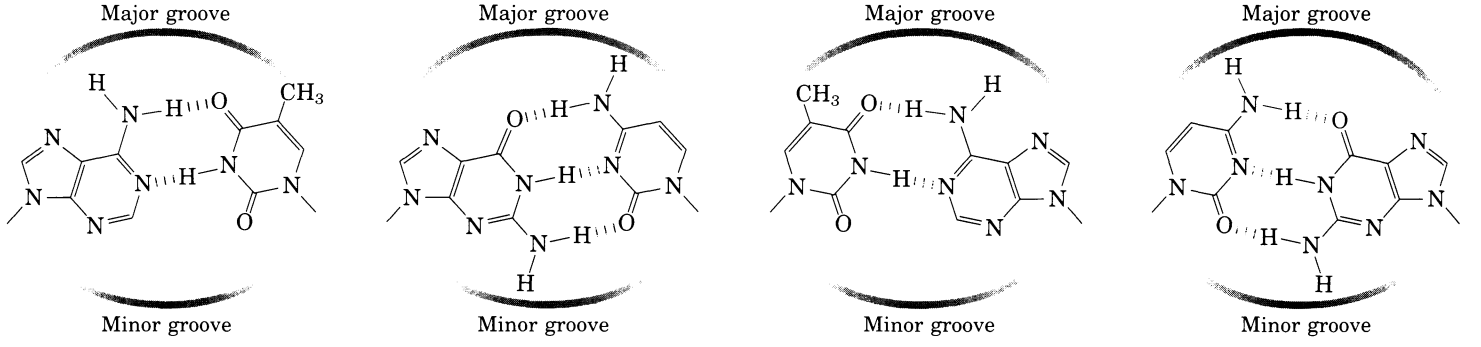
Several β-galactosides structurally related to allolactose are inducers of, but not substrates for, β-galactosidase, and some are substrates but not inducers. One particularly effective and nonmetabolizable inducer of the *lac* operon often used experimentally is isopropylthiogalactoside (IPTG). Such nonmetabolized inducers permit the separation of the physiological function of lactose as a carbon source for growth from its function in the regulation of gene expression.

Many operons are now known in bacteria and a few have been found in lower eukaryotes. The mechanisms by which they are regulated can vary significantly from the simple model presented in Figure 27–7. Research has shown that even the *lac* operon is more complex than indicated here, with an activator protein also contributing to the overall scheme. The regulation of several well-studied bacterial operons, including *lac,* is described in more detail later in this chapter. We now consider the critical molecular interactions between DNA-binding proteins (e.g., repressors and activators) and the specific DNA sequences to which they bind.

Regulatory Proteins Have Discrete DNA-Binding Domains

Regulatory proteins generally bind to specific DNA sequences. They also bind to nonspecific DNA, but their affinity for their target sequences is generally 10^5 to 10^7 times higher. The molecular basis for this discrimination has been the subject of intensive investigation. A general conclusion is that regulatory proteins usually have discrete DNA-binding domains. In addition, the substructures within these domains that actually come in contact with the DNA fall into one of a rather small group of recognizable and characteristic structural motifs.

Before examining these protein structures, it is useful to consider the recognition surfaces on the DNA with which regulatory proteins must interact. Most of the groups that differ from one base to another and can therefore permit discrimination between base pairs are hydrogen-bond donor and acceptor groups exposed in the major DNA groove (Fig. 27–9). Most of the protein–DNA contacts that impart specificity are therefore hydrogen bonds. One notable exception is a nonpolar surface near C-5 of pyrimidines, where thymine is readily distinguished from cytosine by virtue of thymine's protruding methyl group (Fig. 27–9). 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 different base pairs.



Adenine: Thymine Guanine: Cytosine Thymine: Adenine Cytosine: Guanine

Figure 27–9 Functional groups on DNA base pairs in the major groove of DNA. The groups that can be used for base-pair recognition are shown in red for all four base pairs.

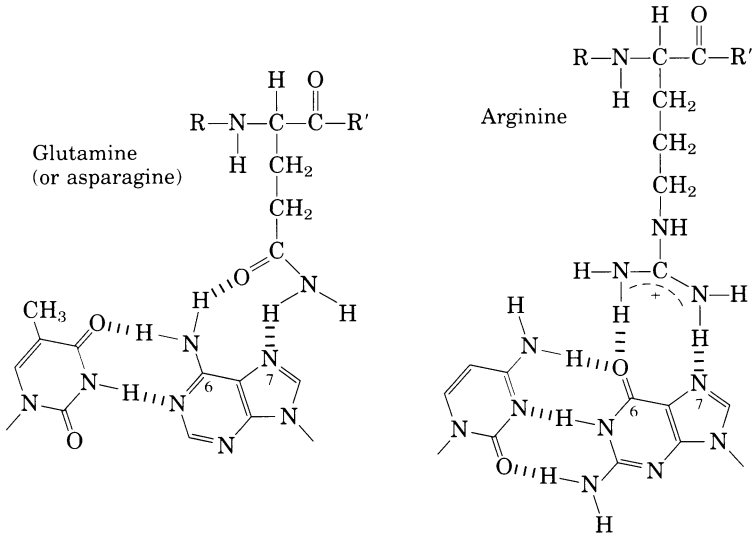
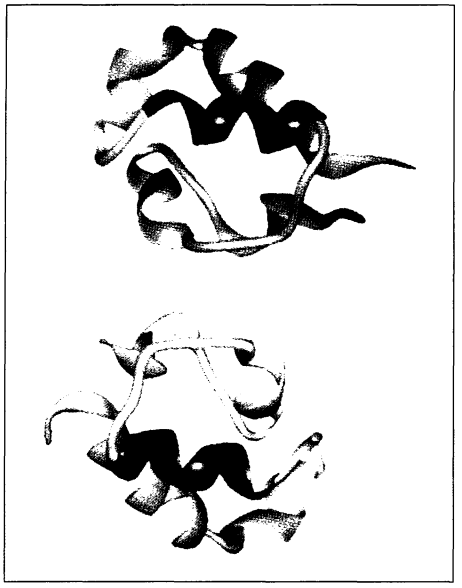


Figure 27–10 Two examples of specific amino acid–base pair interactions that have been observed in the structures of DNA-bound regulatory proteins.

As for the regulatory proteins themselves, the amino acid residues whose side chains are most often found hydrogen-bonded to bases in the DNA include Asn, Gln, Glu, Lys, and Arg. Is there a simple "recognition code" in which an amino acid is always paired with a certain base? The two hydrogen bonds that can form between Gln or Asn and the N^6 and N-7 positions of adenine (Fig. 27–10) constitute a pattern that cannot form with any other base. An Arg residue can similarly form two hydrogen bonds to both N-7 and O^6 of guanine (Fig. 27–10). However, examination of the structures of many DNA-binding proteins has shown that there are multiple ways for a protein to recognize each base pair, and no simple code exists. The Gln–adenine interaction specifies A=T base pairs in some cases, whereas a van der Waals pocket for the methyl group of thymine is the mechanism used to recognize A=T base pairs in other proteins. It is not yet possible to examine the structure of a DNA-binding protein and infer the sequence of the DNA to which it binds.

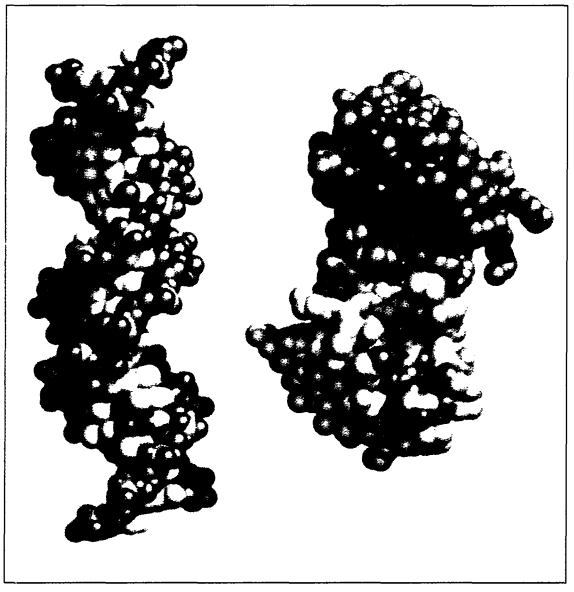
The DNA-binding domains of regulatory proteins tend to be small (60 to 90 amino acid residues). Only a small subset of the amino acids within these domains actually contact the DNA, and the structure of the protein in the region where these amino acids occur is not random. Two structural motifs that play a major role in DNA binding have been found in numerous regulatory proteins: the **helix-turn-helix** motif and the **zinc finger.** Other DNA-binding motifs exist in some proteins, but the discussion here focuses on these well-studied examples.

The Helix-Turn-Helix This DNA-binding motif was the first to be studied in detail. It is the physical basis for protein–DNA interactions for many prokaryotic regulatory proteins. Closely related DNA-binding motifs also occur in some eukaryotic regulatory proteins. The helix-turn-helix motif consists of two short α-helical segments 7 to 9 amino acid residues long, separated by a β turn (about 20 amino acids total). This structure generally is not stable by itself, but it represents the reactive portion of the larger DNA-binding domain. One of the two α helices is referred to as the recognition helix, because it usually contains many of the amino acids that interact with the DNA; this helix is positioned in the major groove. The Cro repressor protein from bacteriophage 434 (a close relative of bacteriophage λ) provides a good example (Fig. 27–11).

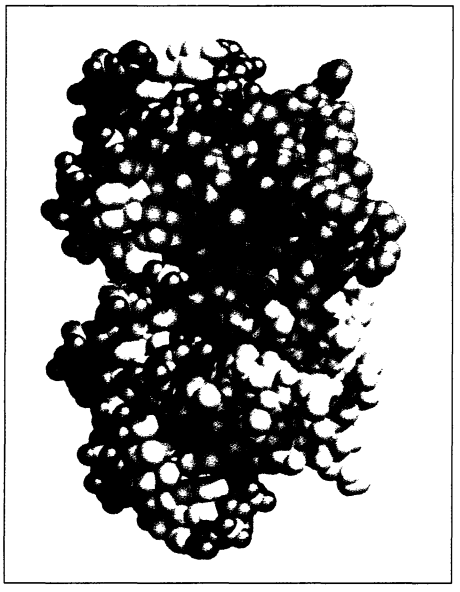


(a)

(b)

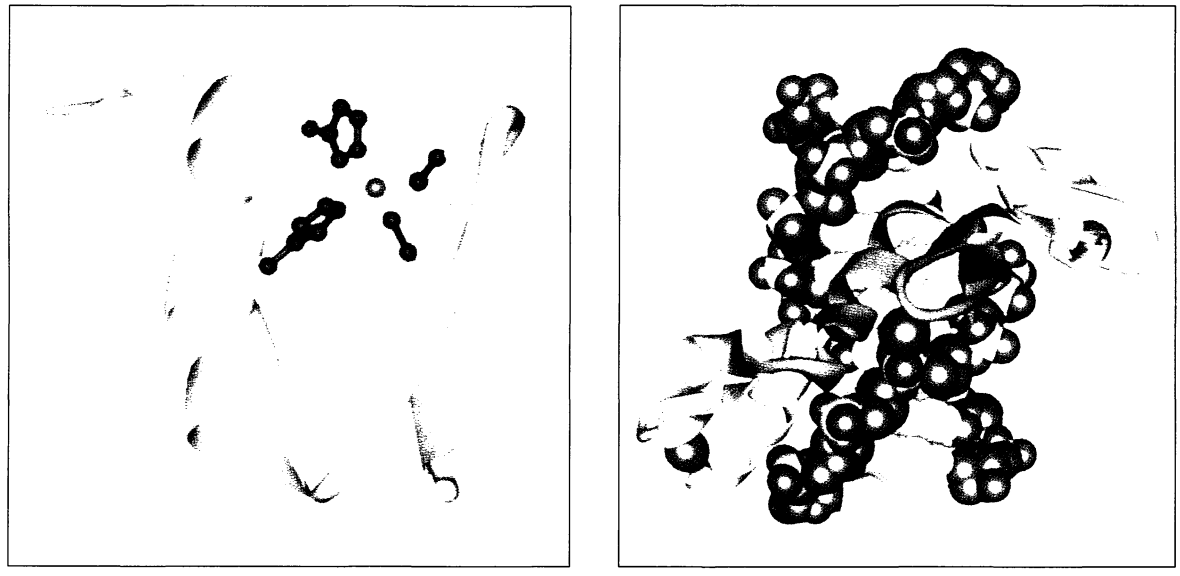


(c)



(d)

Figure 27–11 The Cro repressor of bacteriophage 434 and its interaction with DNA. Each subunit of this dimeric protein contains 71 amino acids. It is presented as a ribbon in **(a)** and **(b),** alone and complexed with its specific DNA binding site. The two subunits are shown in gray and light blue, except for the helix-turn-helix motif in each which is shown in red and yellow. The red helices are the recognition helices, which are positioned in adjacent major grooves of the DNA as seen in **(b).** The interactions between protein and DNA that allow this repressor to discriminate between its specific DNA binding site (shown here) and other DNA sequences are illustrated in **(c)** and **(d).** The protein subunits are again shown in gray and light blue; chemical groups on both the DNA and protein that interact through hydrogen bonds or van der Waals (hydrophobic) interactions are highlighted in red and orange, respectively. Discrimination is mediated by interactions between each protein subunit and four bases (the DNA binding site is a palindrome, and the interactions are the same for both subunits). One set of hydrogen bonds is formed between a Gln residue and the N^6 and N-7 of an adenine (see Fig. 27–10); another hydrogen bond is formed between the O^6 of a guanine and another Gln. In addition, van der Waals pockets on each subunit bind to the C-5 methyl groups of two adjacent thymines. The complementary interacting groups are evident in **(c),** and the complex is shown in **(d).** Many nonspecific contacts (not shown) also exist between protein and DNA in this complex. These do not contribute to discrimination between DNA sequences, but do contribute to the overall DNA-binding affinity. An interesting feature of this structure is that the DNA is bent slightly when it is bound. This occurs in the binding of many proteins to DNA (see Fig. 27–16.)



(a) **(b)**

Figure 27–12 Zinc fingers. **(a)** A ribbon representation of a single zinc finger derived from the regulatory protein Zif 268. The zinc atom is in orange and the amino acid residues that coordinate it (two His and two Cys) are shown in red. **(b)** Three zinc fingers (light blue and gray) from Zif 268 are shown complexed with DNA. The zinc atoms are again shown in orange.

The Zinc Finger Zinc fingers consist of about 30 amino acid residues; four of the residues, either four Cys or two Cys and two His, coordinate a single Zn^{2+} atom (Fig. 27–12). This structural motif is found in many eukaryotic DNA-binding proteins, with several often present in a single protein. There are few, if any, known examples among prokaryotic proteins. Bacteriophage T4 has a protein, the gene 32 protein, that binds single-stranded DNA. It binds a single zinc atom within a structure that may be similar to a zinc finger. An apparent record is held by a DNA-binding protein derived from the frog *Xenopus,* which has 37 zinc fingers. The precise manner in which proteins containing zinc fingers bind to DNA may vary from one protein to the next. In some cases these structures contain the amino acid residues that are involved in sequence discrimination; in other cases the zinc fingers appear to bind DNA nonspecifically, and the amino acids required for specificity are found elsewhere in the protein. The interaction of three zinc fingers (derived from a mouse regulatory protein called Zif 268) with DNA is shown in Figure 27–12b. It should be noted that some regulatory proteins contain zinc bound within structures that are distinct from the zinc finger.

Regulatory Proteins Also Interact with Other Proteins

Regulatory proteins generally contain additional domains that are involved in interactions with RNA polymerase, other regulatory proteins, or additional copies of the same regulatory protein (Fig. 27–13). The DNA binding sites for regulatory proteins are generally inverted repeats of a short DNA sequence (a palindrome) at which two or four copies of a regulatory protein bind cooperatively, as in Figures 27–11 and 27–13.

The Lac repressor is a tetramer of identical subunits (M_r 37,000). A wild-type *E. coli* cell generally contains about ten copies of Lac repressor. The *i* gene is transcribed from its own promoter independently of the *lac* operon genes (Fig. 27–7). The repressor binds to a palindromic operator sequence that spans 22 base pairs within the larger regula-

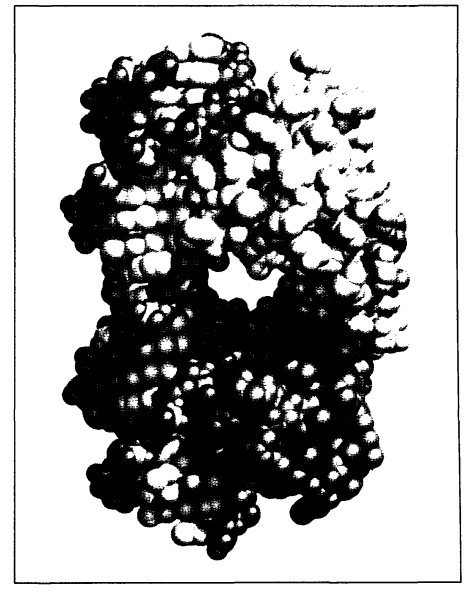


Figure 27–13 The bacteriophage λ repressor bound to DNA. The two identical subunits of the dimeric protein are shown in gray and light blue.

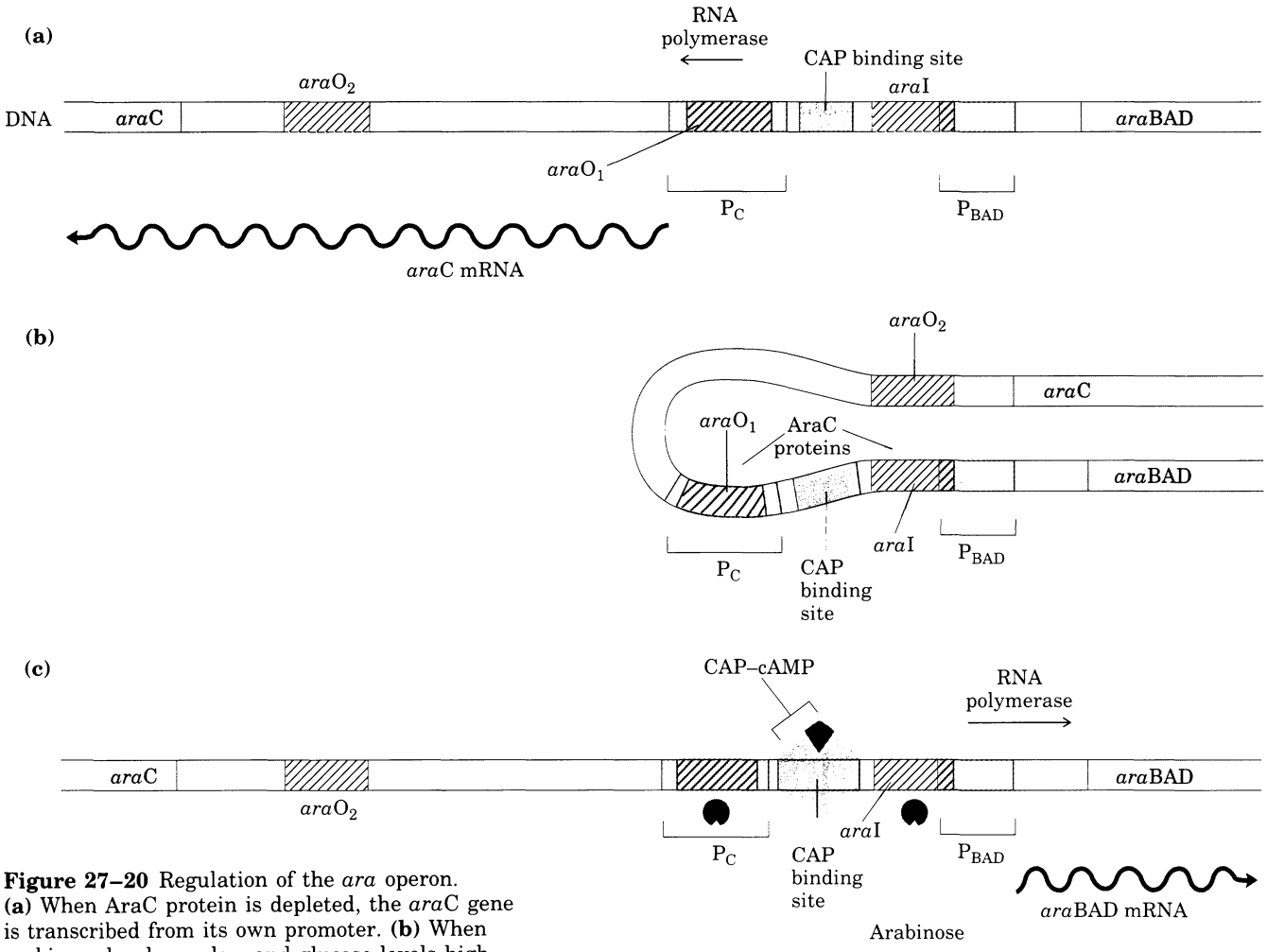


Figure 27–20 Regulation of the *ara* operon. **(a)** When AraC protein is depleted, the *ara*C gene is transcribed from its own promoter. **(b)** When arabinose levels are low and glucose levels high, AraC protein binds to both *ara*I and $araO_2$ and brings these sites together to form a DNA loop. The operon is repressed in this state. AraC protein also binds to $araO_1$, repressing further synthesis of AraC. **(c)** When arabinose is present and glucose concentration is low, AraC protein binds arabinose and changes conformation to become an activator. The DNA loop is opened, and the AraC protein acts in concert with CAP–cAMP to facilitate transcription.

Under these conditions, the AraC protein bound to $araO_2$ and that bound to *ara*I bind to each other, forming a DNA loop of about 210 base pairs. In this configuration the system represses transcription from the promoter for the *ara*BAD genes (Fig. 27–20b). (2) Glucose is not present (or is at low levels) but arabinose is available. Under these conditions, CAP–cAMP becomes abundant and binds to its site adjacent to *ara*I. Arabinose also binds to the AraC protein, altering its conformation. The DNA loop is opened, and the AraC protein bound at *ara*I now becomes an activator, acting in concert with CAP–cAMP to induce transcription of the *ara*BAD genes (Fig. 27–20c). (3) Arabinose and glucose are both abundant. (4) Arabinose and glucose are both absent. For both (3) and (4), the status of the system is not entirely clear, but it remains repressed in both cases. The *ara* operon is a complex regulatory system that provides rapid and reversible responses to changes in environmental conditions.

Genes for Amino Acid Biosynthesis Are Regulated by Transcription Attenuation

Amino acids are required in large amounts for protein synthesis, and *E. coli* has enzymes for synthesizing all of them. Not surprisingly, the genes for the enzymes needed to synthesize a given amino acid are generally clustered in an operon. These enzymes are needed, and hence the operon corresponding to an amino acid is expressed, whenever existing supplies of the amino acid are inadequate for cellular requirements. When the amino acid is in abundant supply, the biosynthetic enzymes are no longer needed and the operon is repressed.

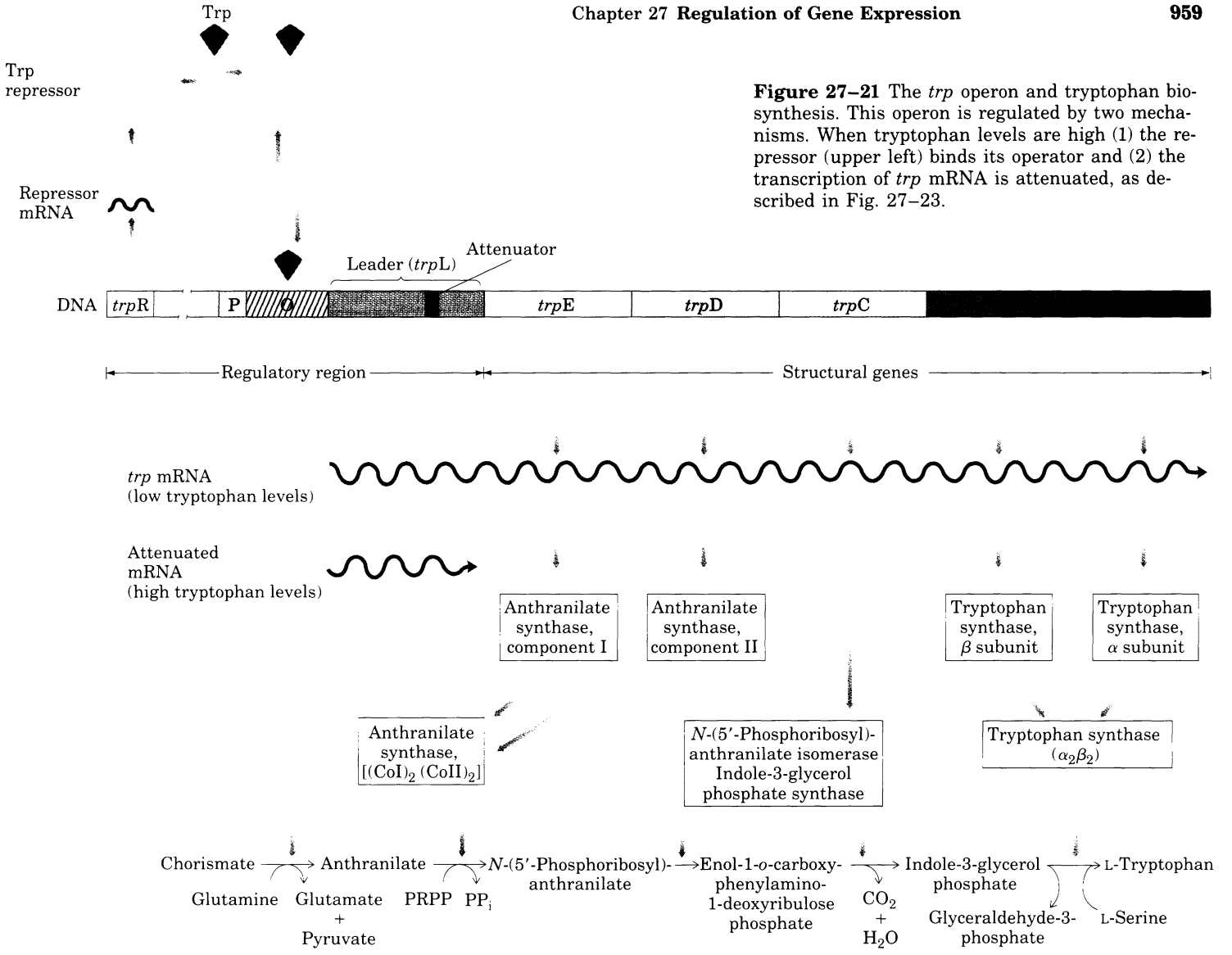


Figure 27–21 The *trp* operon and tryptophan biosynthesis. This operon is regulated by two mechanisms. When tryptophan levels are high (1) the repressor (upper left) binds its operator and (2) the transcription of *trp* mRNA is attenuated, as described in Fig. 27–23.

A well-defined example is the *E. coli* tryptophan *(trp)* operon, which includes five genes for the enzymes required to convert chorismate into tryptophan (Fig. 27–21). 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, with each subunit containing 107 amino acid residues (Fig. 27–22). When tryptophan is abundant, it binds to the Trp repressor, causing a conformational change that permits the repressor to bind its operator. The *trp* operator site overlaps the promoter, and binding of the repressor blocks binding of RNA polymerase.

Here, as elsewhere, this simple "on/off" circuit mediated by a repressor is not the entire regulatory story. This system responds to different tryptophan concentrations by varying 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.

Transcription attenuation describes a process in which transcription is initiated normally but is abruptly halted *before* the operon genes are transcribed. The frequency with which transcription is attenuated depends on the available concentration of tryptophan. The basis for the mechanism, as worked out by Charles Yanofsky, is the very close coupling between transcription and translation in bacteria.

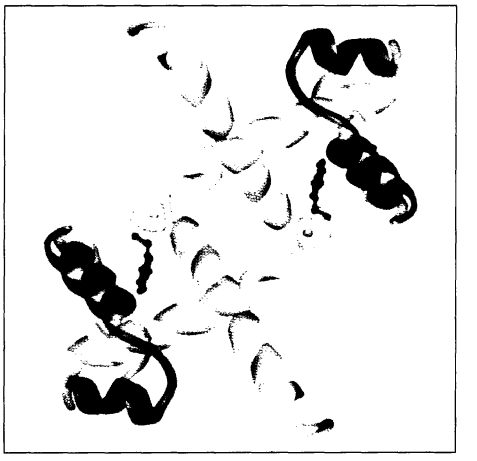


Figure 27–22 Structure of the Trp repressor. The dimeric protein is shown with the helix-turn-helix DNA-binding motifs in red and bound molecules of tryptophan in blue.

activators Sp1 and CTF1 may act through additional bridging proteins called coactivators. The complexity of these interactions, the number of proteins involved, and the central role of these regulatory processes in the life of every eukaryote ensure that this will continue to be an area of vigorous inquiry.

Development Is Controlled by a Cascade of Regulatory Proteins

The transitions in morphology and protein composition observed in the development of a zygote into a multicellular animal or plant with many distinctly different tissues and cell types involve tightly coordinated changes in the expression of the organism's genome. More genes are expressed during early development than in any other part of the life cycle. For example, there are about 18,500 *different* mRNAs in the sea urchin oocyte, but only about 6,000 different mRNAs in the cells of typical differentiated tissues. The mRNAs present in the oocyte give rise to a cascade of events that not only regulate the expression of many genes but also determine where and when the gene products will appear in the developing organism.

Several organisms have emerged as important model systems for the study of development. These include yeasts, nematodes, fruit flies, sea urchins, frogs, chickens, and mice. Our discussion will focus on the development of fruit flies. The emerging picture of the molecular events that occur in development is particularly well advanced in fruit flies and can be used to illustrate patterns and principles of general significance.

The fruit fly, *Drosophila melanogaster,* has a complex life cycle that includes complete metamorphosis in its progression from an embryo to an adult (Fig. 27–35). Among the most important characteristics of the embryo are its **polarity** (the anterior and posterior, dorsal and ventral parts of the animal are readily distinguished) and its **me-**

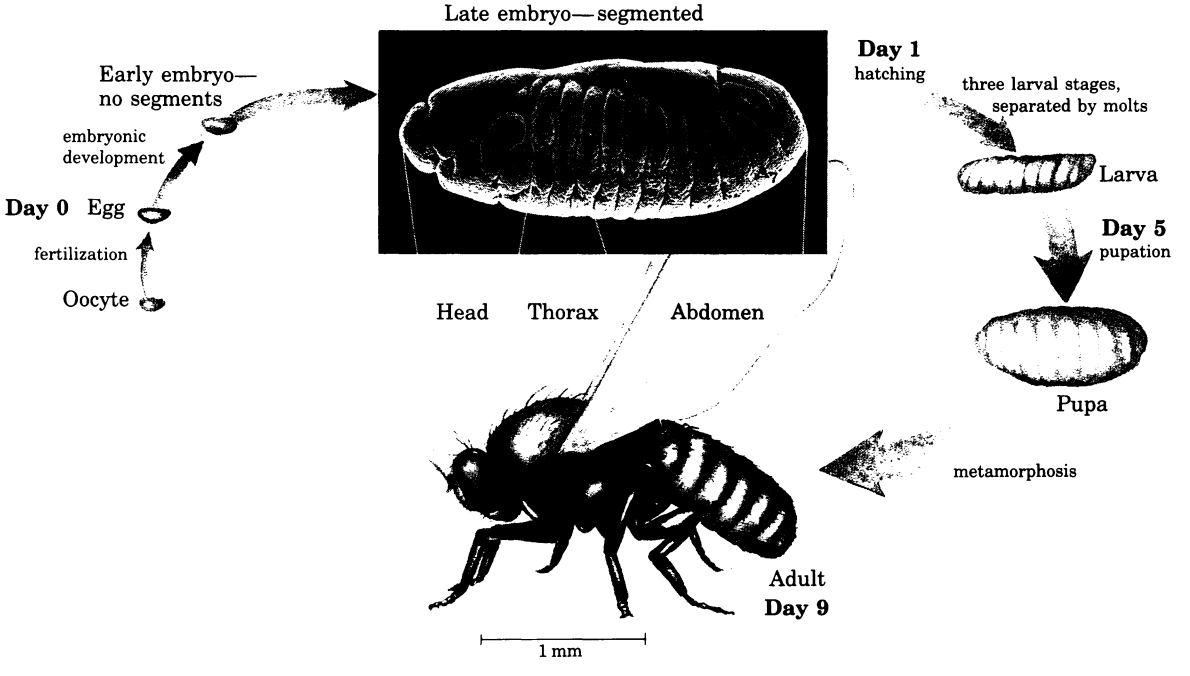


Figure 27–35 The life cycle of the fruit fly *Drosophila melanogaster.* In complete metamorphosis, the adult insect is radically different in form from its immature stages; this process requires extensive "remodeling" during development. By the late embryonic stage, segments have formed from which the various structures in the adult fly will develop.

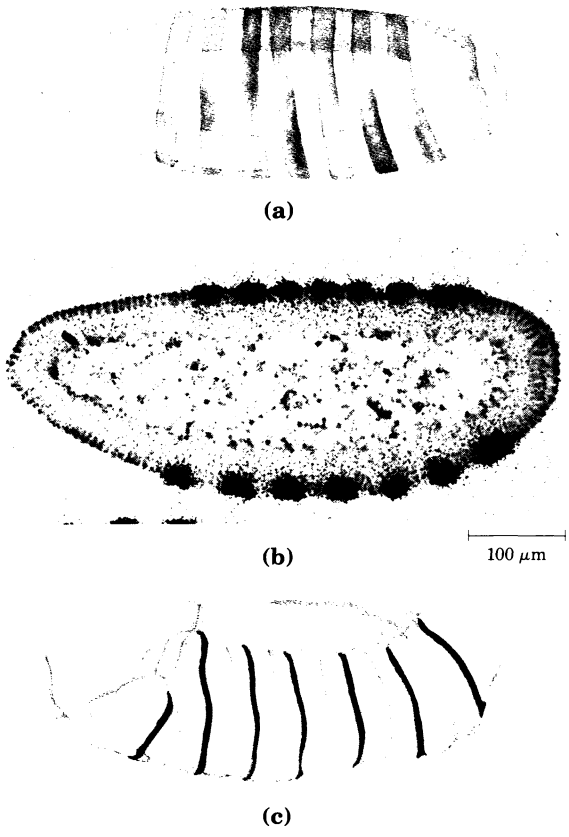


Figure 27–38 Distribution of the *fushi tarazu (ftz)* gene product in early embryos. In the normal embryo, the gene product can be detected in seven bands around the circumference of the embryo, as shown schematically in **(a).** These bands are seen as dark spots (generated by a radioactive label) in a cross-sectional autoradiograph **(b),** and give rise to the segments shown here in red in the late embryo **(c).**

homeotic genes. One well-characterized segmentation gene is *fushi tarazu (ftz),* which belongs to the "pair-rule" subclass. When this gene is lost, the embryo develops seven double-wide segments instead of the normal 14. 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. 27–38). The stripes correspond to the positions of segments that develop later, and which are eliminated if *ftz* function is lost. The expression of pattern-regulating genes such as *ftz* (and *bcd,* expressed earlier in development) establishes a kind of chemical blueprint for the body plan that precedes the actual formation of a body structure.

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. 27–39).

The homeotic genes span long regions of DNA. The *Ubx* gene, for example, is 77,000 base pairs in length and contains introns that are as long as 50,000 base pairs. Transcription of this gene takes nearly one 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 precise nature of many of the events directed by these proteins, and in many cases the biochemical function of the proteins themselves, are unknown. A likely DNA-binding domain has been identified

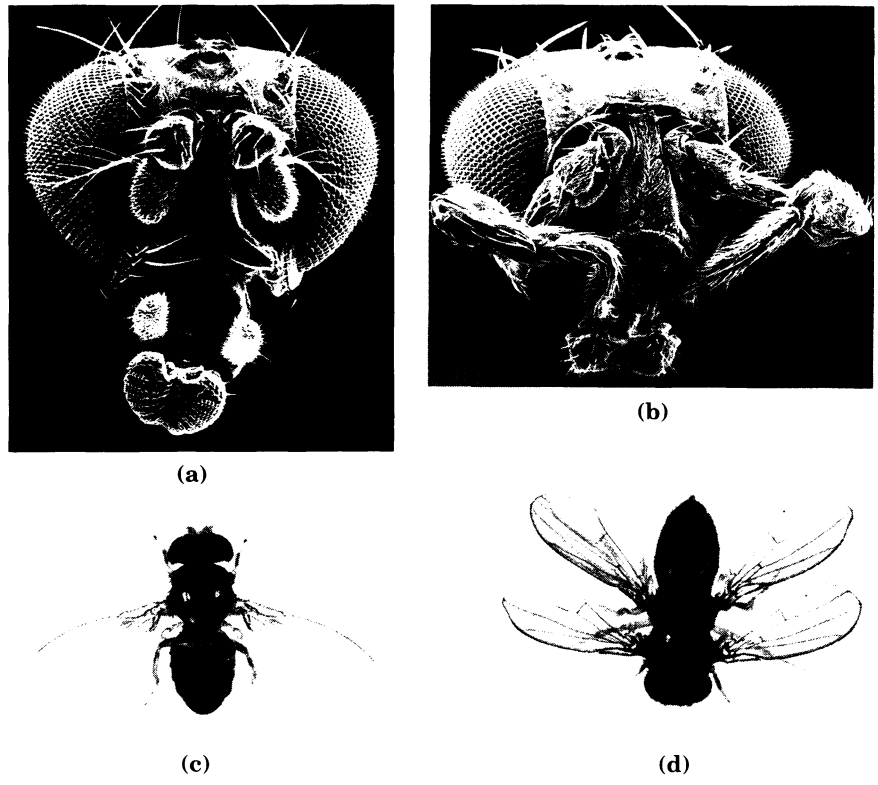


Figure 27–39 The effects of mutations in homeotic genes. **(a)** Normal *Drosophila* head. **(b)** *Drosophila* homeotic mutant *(Antennapaedia)* in which antennae are replaced by legs. **(c)** Normal *Drosophila* body structure. **(d)** Homeotic mutant *(Bithorax)* in which a segment has developed incorrectly to produce an extra set of wings.

in a number of these proteins, however, which suggests that they are regulatory proteins. This domain contains 60 amino acids and is called the **homeodomain** because it was found first in homeotic genes. The DNA sequence encoding this domain is called the **homeobox.** It is highly conserved and has been identified in proteins from a wide variety of organisms. The DNA-binding segment of the domain is related to the helix-turn-helix motif.

The identification 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 slowly come together.

Summary

The expression of genes is regulated by a number of processes that affect the rates at which gene products are synthesized and degraded. Much of this regulation occurs at the level of the initiation of transcription and is mediated by regulatory proteins that either repress or activate transcription from specific promoters. Regulation by repressors and activators is called negative and positive regulation, respectively.

In prokaryotes, genes with interdependent functions are often clustered as a single transcriptional unit called an operon. The transcription of operon genes is generally blocked by the 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, called 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 the biological inducer, allolactose.

Regulatory proteins are DNA-binding proteins that recognize specific sequences in the DNA. Most of these proteins have distinct DNA-binding domains. Within these domains, common structural motifs involved in DNA binding are the helix-turn-helix and zinc finger motifs. Regulatory proteins also contain domains for protein–protein interactions, including leucine zipper and helix-loop-helix motifs involved in dimerization and several classes of domains involved in the activation of transcription.

The lactose operon of *E. coli* also exhibits positive regulation by the catabolite gene activator protein (CAP). When cAMP concentrations are high (glucose concentrations are low), CAP 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 concentrations, restricting expression of *lac* (and other) genes and suppressing the use of secondary sugars. Several operons that are coordinately regulated, as with CAP and cAMP, are referred to as a regulon.

Other mechanisms of regulation are also observed in prokaryotes. In the arabinose *(ara)* operon, the AraC protein acts as both activator and repressor. Some repressors, as in the *ara* operon and the bacteriophage λ system, regulate their own synthesis (autoregulation). Some regulatory proteins in the *ara* system bind sites many base pairs distant from each other and interact by DNA looping mechanisms. Amino acid biosynthetic operons have a regulatory circuit called attenuation that uses a transcription termination site (the attenuator), modulating its formation in the mRNA by a mechanism that couples transcription and translation and responds to small changes in amino acid concentration. In the SOS system, multiple unlinked genes are repressed by a single type of repressor protein, and all of the genes are induced simultaneously when DNA damage triggers RecA protein–mediated proteolysis of the repressor. The bacteriophage λ has a complex regulatory circuit that oversees the choice between lysis and lysogeny. Two λ proteins, N and Q, act as antiterminators, modifying the host RNA polymerase so that it can bypass transcription termination sites. Finally, some prokaryotic genes are regulated by genetic recombination processes that physically move promoters relative to the genes being regulated. These diverse mechanisms permit very sensitive cellular responses to changes in environmental conditions.

Some regulation also occurs at the level of translation. The synthesis of ribosomal proteins in bacteria is mediated by a strategy in which one protein in each ribosomal protein operon acts as a

translational repressor. The mRNA is bound by the repressor and translation is blocked only when the ribosomal protein is present in excess relative to available rRNA.

Eukaryotes employ many of the same regulatory schemes, although positive regulation appears to be more common and transcription is also accompanied by large changes in chromatin structure. Eukaryotic transcriptional activator proteins are generally required for RNA polymerase binding and activity. Some transcription factors have general functions; the TFII factors associated with RNA polymerase II, for example, are required at almost all RNA polymerase II promoters. Other transcriptional activators, unique to one gene or set of genes, have distinct domains for DNA binding and activation, and their DNA binding sites are often found hundreds of base pairs from the site where RNA synthesis begins.

Perhaps the most complex regulatory problem is the development of a multicellular animal. Here, sets of regulating genes operate in temporal and spatial succession, turning a given area of an egg cell into a predictable structure in the adult animal. Research continues into the molecular basis for this highly coordinated process.

Further Reading

General

Ingraham, J.L., Magasanik, B., Low, K.B., Schaechter, M., & Umbarger, H.E. (eds) (1987) Escherichia coli *and* Salmonella typhimurium, *Cellular and Molecular Biology,* Vol. 2, American Society for Microbiology, Washington, DC.

An excellent reference source for reviews of many bacterial operons.

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

Schleif, R. (1986) *Genetics and Molecular Biology,* Addison-Wesley Publishing Co., Inc., Reading, MA.

Chapters 12, 13, and 14 provide an excellent account of the experimental basis of major concepts of gene regulation in prokaryotes.

Schleif, R. (1992) DNA looping. *Annu. Rev. Biochem.* **61,** 199–223.

Struhl, K. (1989) Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. *Trends Biochem. Sci.* **14,** 137–140.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A., & Weiner, A.M. (1987) *Molecular Biology of the Gene,* 4th edn, The Benjamin/Cummings Publishing Company, Menlo Park, CA.

Regulation of Gene Expression in Prokaryotes

Gottesman, S. (1984) Bacterial regulation: global regulatory networks. *Annu. Rev. Genet.* **18,** 415–441.

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 were proposed in this historic paper.

Nomura, M., Gourse, R., & Baughman, G. (1984) Regulation of the synthesis of ribosomes and ribosomal components. *Annu. Rev. Biochem.* **53,** 75–117.

Ptashne, M., Johnson, A.D., & Pabo, C.O. (1982) A genetic switch in a bacterial virus. *Sci. Am.* **247** (November), 128–140.

Stephens, J.C., Artz, S.W., & Ames, B.N. (1975) Guanosine 5′-diphosphate 3′-diphosphate (ppGpp): positive effector for histidine operon transcription and general signal for amino acid deficiency. *Proc. Natl. Acad. Sci. USA* **72,** 4389–4393.

Yanofsky, C. (1981) Attenuation in the control of expression of bacterial operons. *Nature* **289,** 751–758.

Zieg, J., Silverman, M., Hilmen, M., & Simon, M. (1977) Recombinational switch for gene expression. *Science* **196,** 170–172.

Regulation of Gene Expression in Eukaryotes

Beardsley, T. (1991) Smart genes. *Sci. Am.* **265** (August), 86–95.

A good overview of gene regulation during development.

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

Guarente, L. (1988) UASs and enhancers: common mechanism of transcriptional activation in yeast and mammals. *Cell* **52,** 303–305.

Kornberg, R.D. & Lorch, Y. (1991) Irresistible force meets immovable object: transcription and the nucleosome. *Cell* **67,** 833–836.

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.

Ptashne, M. (1989) How gene activators work. *Sci. Am.* **260** (January), 40–47.

Pugh, B.F. & Tjian, R. (1992) Diverse transcriptional functions of the multisubunit eukaryotic TFIID complex. *J. Biol. Chem.* **267,** 679–682.

Struhl, K. (1987) Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. *Cell* **49,** 295–297.

Thummel, C.S. (1992) Mechanisms of transcriptional timing in *Drosophila. Science* **255,** 39–40.

Zlatanova, J. (1990) Histone H1 and the regulation of transcription of eukaryotic genes. *Trends Biochem. Sci.* **15,** 273–276.

Problems

1. *Negative Regulation* In the *lac* operon, describe the probable effect on gene expression of:
(a) Mutations in the *lac* operator
(b) Mutations in the *lac*I gene
(c) Mutations in the promoter

2. *Effect of mRNA and Protein Stability on Regulation* An *E. coli* cell is growing in a solution 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 cell changes if:
(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.

3. *Functional Domains in Regulatory Proteins* A biochemist replaces the DNA-binding domain of the yeast GAL4 protein with the DNA-binding domain from the λ repressor (CI) and finds that the engineered protein no longer functions as a transcriptional activator (it no longer regulates transcription of the *gal* operon in yeast). What might be done to the GAL4 DNA-binding site to make the engineered protein functional in activating *gal* operon transcription?

4. *Bacteriophage λ* Bacteria that become lysogenic for bacteriophage λ are immune to subsequent λ lytic infections. Why?

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

6. *Transcription Attenuation* In the leader region of the *trp* mRNA, what would be the effect of:
(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?

7. *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 10^5 to 10^6. About ten molecules of the repressor per cell are sufficient to ensure a high level of repression. Assume that a very similar repressor existed in a human cell and had a similar specificity for its binding site. How many copies of the repressor would be required per cell to elicit a level of repression similar to that seen in the prokaryotic cell? (Hint: The *E. coli* genome contains about 4.7 million base pairs and the human genome contains about 2.4 billion base pairs.)

8. *Positive Regulation* 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 is observed to decline. The purified enzyme is completely inactive unless crude extract is added to the reaction mixture. Suggest an explanation for these observations.