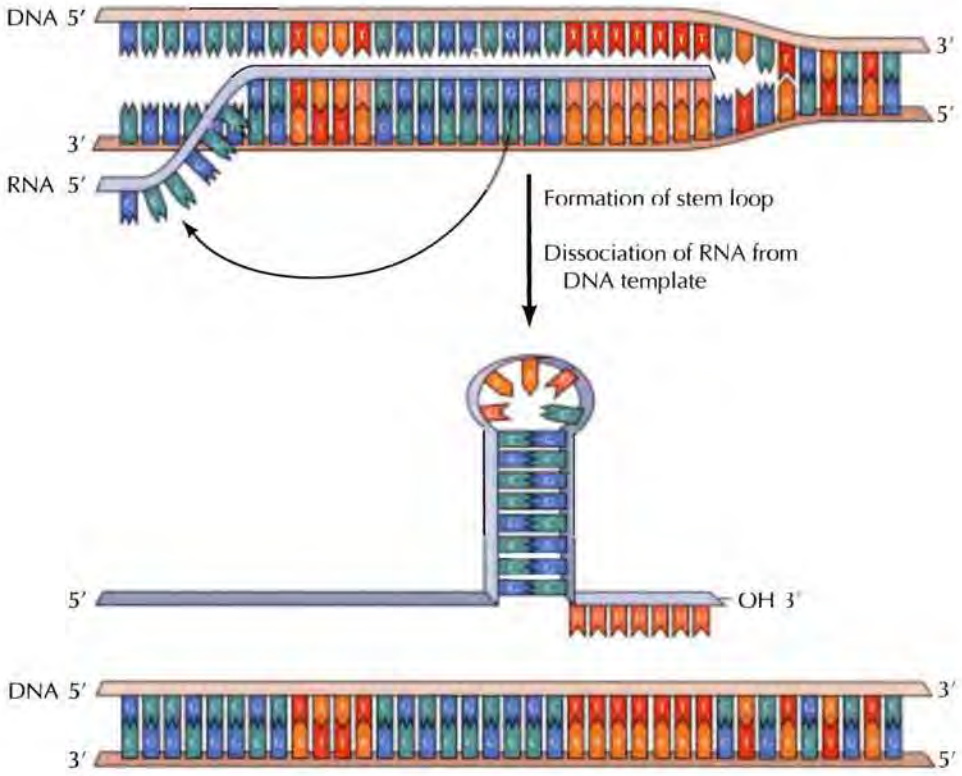

FIGURE 7.6 Transcription termination The termination of transcription is signaled by a GC-rich inverted repeat followed by seven A residues. The inverted repeat forms a stable stem-loop structure in the RNA, causing the RNA to dissociate from the DNA template.

RNA synthesis continues until the polymerase encounters a termination signal, at which point transcription stops, the RNA is released from the polymerase, and the enzyme dissociates from its DNA template. There are two alternative mechanisms for termination of transcription in *E. coli*. The simplest and most common type of termination signal consists of a symmetrical inverted repeat of a GC-rich sequence followed by approximately seven A residues (Figure 7.6). Transcription of the GC-rich inverted repeat results in the formation of a segment of RNA that can form a stable stem-loop structure by complementary base pairing. The formation of such a self-complementary structure in the RNA disrupts its association with the DNA template and terminates transcription. Because hydrogen bonding between A and U is weaker than that between G and C, the presence of A residues downstream of the inverted repeat sequences is thought to facilitate the dissociation of the RNA from its template. Alternatively, the transcription of some genes is terminated by a specific termination protein (called Rho), which binds extended segments (greater than 60 nucleotides) of single-stranded RNA. Since mRNAs in bacteria become associated with ribosomes and are translated while they are being transcribed, such extended regions of single-stranded RNA are exposed only at the end of an mRNA.

Repressors and Negative Control of Transcription

Transcription can be regulated at the stages of both initiation and elongation, but most transcriptional regulation in bacteria operates at the level of initiation. The pioneering studies of gene regulation in *E. coli* were carried out by François Jacob and Jacques Monod in the 1950s. These investigators and their colleagues analyzed the expression of enzymes involved in the metabolism of lactose, which can be used as a source of carbon and energy via cleavage to glucose and galactose (Figure 7.7). The enzyme that catalyzes the cleavage of lactose (β-galactosidase) and other enzymes involved in lac-

tose metabolism are expressed only when lactose is available for use by the bacteria. Otherwise, the cell is able to economize by not investing energy in the synthesis of unnecessary RNAs and proteins. Thus lactose induces the synthesis of enzymes involved in its own metabolism. In addition to requiring β-galactosidase, lactose metabolism involves the products of two other closely linked genes: lactose permease, which transports lactose into the cell, and a transacetylase, which is thought to inactivate toxic thiogalactosides that are transported into the cell along with lactose by the permease. On the basis of purely genetic experiments, Jacob and Monod deduced the mechanism by which the expression of these genes was regulated, thereby formulating a model that remains fundamental to our understanding of transcriptional regulation.

The starting point in this analysis was the isolation of mutants that were defective in regulation of the genes involved in lactose utilization. These mutants were of two types: constitutive mutants, which expressed all three genes even when lactose was not available, and noninducible mutants, which failed to express the genes even in the presence of lactose. Genetic mapping localized these regulatory mutants to two distinct loci, called *o* and *i*, with *o* located immediately upstream of the structural gene for β-galactosidase. Mutations affecting *o* resulted in constitutive expression; mutants of *i* were either constitutive or noninducible.

The function of these regulatory genes was probed by experiments in which two strains of bacteria were mated, resulting in diploid cells containing genes derived from both parents (Figure 7.8). Analysis of gene expression in such diploid bacteria provided critical insights by defining which alleles of these regulatory genes are dominant and which are recessive. For example, when bacteria containing a normal *i* gene (i^+) were mated with bacteria carrying an *i* gene mutation resulting in constitutive expression (an i^- mutation), the resulting diploid bacteria displayed normal inducibility; therefore

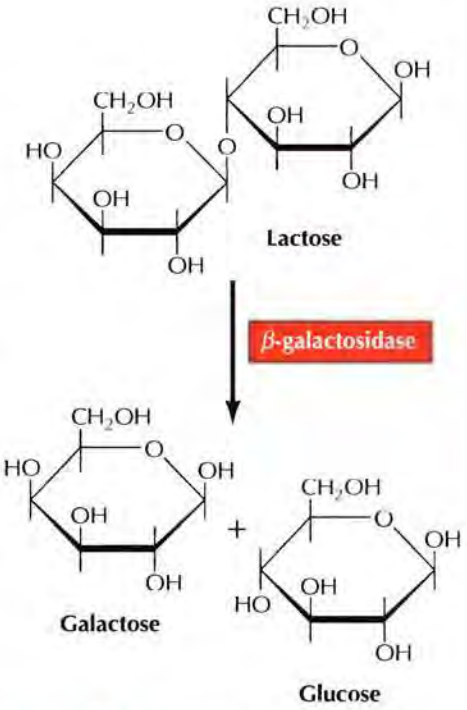

FIGURE 7.7 Metabolism of lactose β-galactosidase catalyzes the hydrolysis of lactose to glucose and galactose.

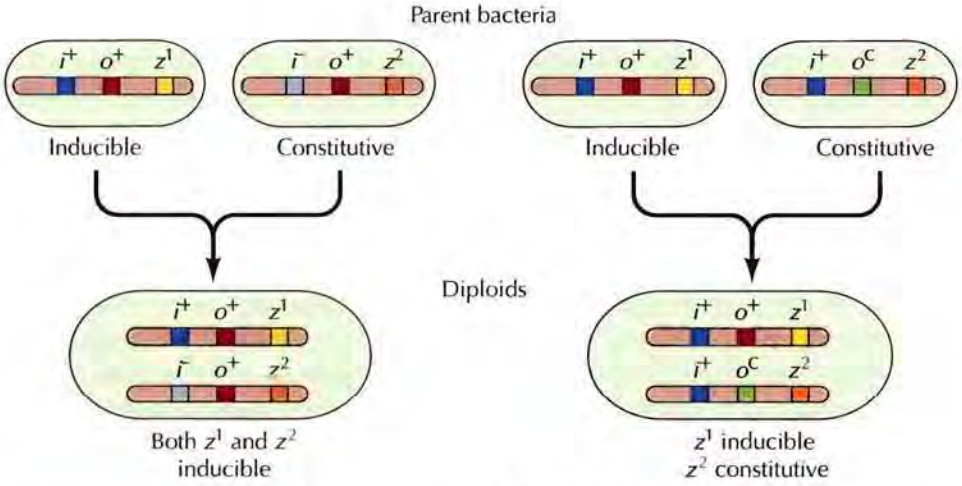

FIGURE 7.8 Regulation of β-galactosidase in diploid *E. coli* The mating of two bacterial strains results in diploid cells that contain genes from both parents. In these examples, it is assumed that the genes encoding β-galactosidase (the z genes) can be distinguished on the basis of structural gene mutations, designated z^1 and z^2. In an i^+/i^- diploid (left), both structural genes are inducible; therefore i^+ is dominant over i^- and affects expression of z genes on both chromosomes. In contrast, in an o^c/o^+ diploid (right), the z gene linked to o^c is constitutively expressed, whereas that linked to o^+ is inducible. Therefore *o* affects expression of only the adjacent z gene on the same chromosome.

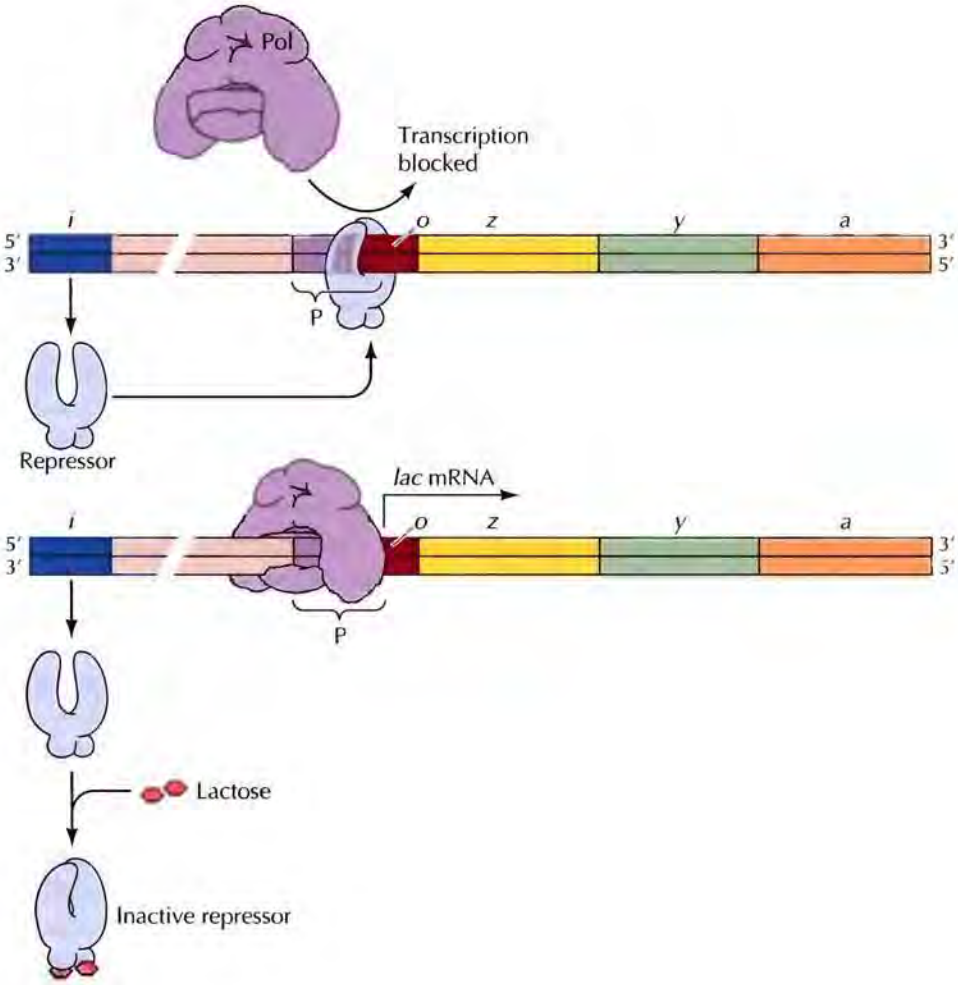

FIGURE 7.9 Negative control of the *lac* operon The *i* gene encodes a repressor which, in the absence of lactose (top), binds to the operator (*o*) and interferes with the binding of RNA polymerase to the promoter, blocking transcription of the three structural genes (*z*, β-galactosidase; *y*, permease; and *a*, transacetylase). Lactose induces expression of the operon by binding to the repressor (bottom), which prevents the repressor from binding to the operator. P = promoter; Pol = polymerase.

the normal i^+ gene was dominant over the i^- mutant. In contrast, matings between normal bacteria and bacteria with an o^c mutation (constitutive expression) yielded diploids with the constitutive expression phenotype, indicating that o^c is dominant over o^+. Additional experiments in which mutations in *o* and *i* were combined with different mutations in the structural genes showed that *o* affects the expression of only the genes to which it is physically linked, whereas *i* affects the expression of genes on both chromosome copies in diploid bacteria. Thus, in an o^c/o^+ cell, only the structural genes that are linked to o^c are constitutively expressed. In contrast, in an i^+/i^- cell, structural genes on both chromosomes are regulated normally. These results led to the conclusion that *o* represents a region of DNA that controls the transcription of adjacent genes, whereas the *i* gene encodes a regulatory factor (e.g., a protein) that can diffuse throughout the cell and control genes on both chromosomes.

The model of gene regulation developed on the basis of these experiments is illustrated in Figure 7.9. The genes encoding β-galactosidase, permease, and transacetylase are expressed as a single unit, called an **operon**. Transcription of the operon is controlled by *o* (the **operator**), which is adjacent to the transcription initiation site. The *i* gene encodes a protein that regulates transcription by binding to the operator. Since i^- mutants (which result in constitutive gene expression) are recessive, it was concluded that these mutants failed to make a functional gene product. This result implies that the normal *i* gene product is a **repressor**, which blocks transcription

when bound to *o*. The addition of lactose leads to induction of the operon because lactose binds to the repressor, thereby preventing it from binding to the operator DNA. In noninducible *i* mutants (which are dominant over i^+), the repressor fails to bind lactose, so expression of the operon cannot be induced.

The model neatly fits the results of the genetic experiments from which it was derived. In i^- cells, the repressor is not made, so the *lac* operon is constitutively expressed. Diploid i^+/i^- cells are normally inducible, since the functional repressor is encoded by the i^+ allele. Finally, in o^c mutants, a functional operator has been lost and the repressor cannot be bound. Consequently, o^c mutants are dominant but affect the expression only of linked structural genes.

Confirmation of this basic model has since come from a variety of experiments, including Walter Gilbert's isolation, in the 1960s, of the *lac* repressor and analysis of its binding to operator DNA. Molecular analysis has defined the operator as approximately 20 base pairs of DNA, starting a few bases before the transcription initiation site. Footprinting analysis has identified this region as the site to which the repressor binds, blocking transcription by interfering with the binding of RNA polymerase to the promoter. As predicted, lactose binds to the repressor, which then no longer binds to operator DNA. Also as predicted, o^c mutations alter sequences within the operator, thereby preventing repressor binding and resulting in constitutive gene expression.

The central principle of gene regulation exemplified by the lactose operon is that control of transcription is mediated by the interaction of regulatory proteins with specific DNA sequences. This general mode of regulation is broadly applicable to both prokaryotic and eukaryotic cells. Regulatory sequences like the operator are called ***cis*-acting control elements**, because they affect the expression of only linked genes on the same DNA molecule. On the other hand, proteins like the repressor are called ***trans*-acting factors** because they can affect the expression of genes located on other chromosomes within the cell. The *lac* operon is an example of negative control because binding of the repressor blocks transcription. This, however, is not always the case; many *trans*-acting factors are activators rather than inhibitors of transcription.

Positive Control of Transcription

The best-studied example of positive control in *E. coli* is the effect of glucose on the expression of genes that encode enzymes involved in the breakdown (catabolism) of other sugars (including lactose) that provide alternative sources of carbon and energy. Glucose is preferentially utilized; so as long as glucose is available, enzymes involved in catabolism of alternative energy sources are not expressed. For example, if *E. coli* are grown in medium containing both glucose and lactose, the *lac* operon is not induced and only glucose is used by the bacteria. Thus glucose represses the *lac* operon even in the presence of the normal inducer (lactose).

Glucose repression (generally called catabolite repression) is now known to be mediated by a positive control system, which is coupled to levels of cyclic AMP (cAMP) (Figure 7.10). In bacteria, the enzyme adenylyl cyclase, which converts ATP to cAMP, is regulated such that levels of cAMP increase when glucose levels drop. cAMP then binds to a transcriptional regulatory protein called catabolite activator protein (CAP). The binding of cAMP stimulates the binding of CAP to its target DNA sequences, which in the *lac* operon are located approximately 60 bases upstream of the transcription start

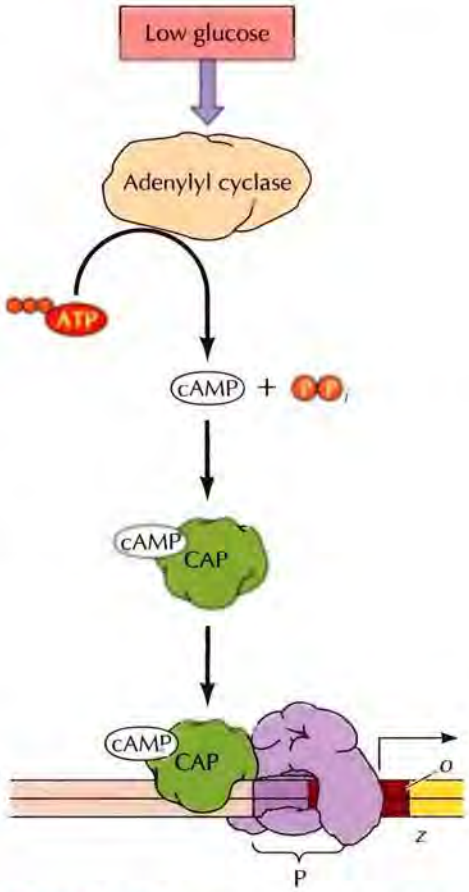

FIGURE 7.10 Positive control of the *lac* operon by glucose Low levels of glucose activate adenylyl cyclase, which converts ATP to cyclic AMP (cAMP). Cyclic AMP then binds to the catabolite activator protein (CAP) and stimulates its binding to regulatory sequences of various operons concerned with the metabolism of alternative sugars, such as lactose. CAP interacts with the α subunit of RNA polymerase to facilitate the binding of polymerase to the promoter.

TABLE 7.1 Classes of Genes Transcribed by Eukaryotic RNA Polymerases

Type of RNA synthesized	RNA polymerase
Nuclear genes	
mRNA	II
tRNA	III
rRNA	
5.8S, 18S, 28S	I
5S	III
snRNA and scRNA	II and III[a]
Mitochondrial genes	Mitochondrial[b]
Chloroplast genes	Chloroplast[b]

[a]Some small nuclear (sn) and small cytoplasmic (sc) RNAs are transcribed by polymerase II and others by polymerase III.

[b]The mitochondrial and chloroplast RNA polymerases are similar to bacterial enzymes.

site. CAP then interacts with the α subunit of RNA polymerase, facilitating the binding of polymerase to the promoter and activating transcription.

Eukaryotic RNA Polymerases and General Transcription Factors

Although transcription proceeds by the same fundamental mechanisms in all cells, it is considerably more complex in eukaryotic cells than in bacteria. This is reflected in three distinct differences between the prokaryotic and eukaryotic systems. First, whereas all genes are transcribed by a single core RNA polymerase in bacteria, eukaryotic cells contain multiple different RNA polymerases that transcribe distinct classes of genes. Second, eukaryotic RNA polymerases need to interact with a variety of additional proteins to specifically initiate and regulate transcription. Finally, transcription in eukaryotes takes place on chromatin rather than on free DNA, and regulation of chromatin structure is an important factor in the transcriptional activity of eukaryotic genes. This increased complexity of eukaryotic transcription presumably facilitates the sophisticated regulation of gene expression needed to direct the activities of the many different cell types of multicellular organisms.

Eukaryotic RNA Polymerases

Eukaryotic cells contain three distinct nuclear RNA polymerases that transcribe different classes of genes (Table 7.1). Protein-coding genes are transcribed by RNA polymerase II to yield mRNAs; ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) are transcribed by RNA polymerases I and III. RNA polymerase I is specifically devoted to transcription of the three largest species of rRNAs, which are designated 28S, 18S, and 5.8S according to their rates of sedimentation during velocity centrifugation. RNA polymerase III transcribes the genes for tRNAs and for the smallest species of ribosomal RNA (5S rRNA). Some of the small RNAs involved in splicing and protein transport (snRNAs and scRNAs) are also transcribed by RNA polymerase III, while others are polymerase II transcripts. RNA polymerase II also transcribes microRNAs, which are important regulators of both transcription and translation in eukaryotic cells. In addition, separate RNA polymerases (which are similar to bacterial RNA polymerases) are found in chloroplasts and mitochondria, where they specifically transcribe the DNAs of those organelles.

All three of the nuclear RNA polymerases are complex enzymes, consisting of 12 to 17 different subunits each. Although they recognize different promoters and transcribe distinct classes of genes, they share several features in common with each other as well as with bacterial RNA polymerase. In particular, all three eukaryotic RNA polymerases contain nine conserved subunits, five of which are related to the α, β, β', and ω subunits of bacterial RNA polymerase. The structure of yeast RNA polymerase II is strikingly similar to that of the bacterial enzyme (Figure 7.11), suggesting that all RNA polymerases utilize fundamentally conserved mechanisms to transcribe DNA.

General Transcription Factors and Initiation of Transcription by RNA Polymerase II

Because RNA polymerase II is responsible for the synthesis of mRNA from protein-coding genes, it has been the focus of most studies of transcription