Regulation of Gene Expression in Prokaryotes

(Photo above): Colonies of *lac*$^+$ *E. coli* cells (blue) and *lac*$^-$ *E. coli* cells (white) growing on X-gal medium. (Photo by Kristine Kirkeby, Biological Sciences Art Services, University of Minnesota.)

BACTERIA such as *Escheriscia coli* are exposed to a wide variety of environmental conditions. For example, *E. coli* cells may encounter rapidly changing growth conditions as they pass from mammalian intestinal tracts to sewer systems to polluted rivers, lakes, ponds, and so on. Each of these ecological niches will provide different organic molecules for use as energy sources. Thus, it seems quite reasonable to assume that natural selection will have preserved those organisms that have evolved ways of adapting to the wide range of environmental conditions encountered during their evolution. Indeed, the available information indicates that most prokaryotes such as *E. coli* exhibit remarkable capacities to adapt to diverse environmental conditions.

To a considerable degree, the adaptability of bacteria and other prokaryotes depends on their ability to "turn on" and "turn off" the expression of specific sets of genes in response to the specific demands of the environmental milieu. Stated differently, these organisms exhibit a striking ability to *regulate the expression of specific genes in response to environmental signals*. The expression of particular genes is "turned on" when the products of these genes are needed for growth in a given environment. Their expression is "turned off" when their products are no longer needed for growth in the existing milieu. Clearly, the ability of an organism to regulate gene expression in this way will increase its overall "fitness" (its ability to grow and leave progeny under a variety of environmental conditions). The synthesis of gene transcripts and translation products requires the expenditure of considerable energy. By "turning off" the expression of genes when their products are not needed, an organism can avoid wasting energy and can utilize the conserved energy to synthesize products that maximize the growth rate in the existing milieu. What, then, are the mechanisms by which these organisms regulate gene expression in response to changes in the environment? Is there a single mechanism by which the expression of different genes or sets of genes are regulated? Or are different genes controlled by different mechanisms?

Certain genes, for example, the genes specifying ribosomal RNAs, ribosomal proteins, and transfer RNAs, are undoubtedly expressed at some time in virtually all cells regardless of the environmental conditions. The products of these genes are required for growth of all cells in all environments. However, the products of many other genes are required for growth only in certain environments, and the expression of these genes is regulated such that the products are synthesized only when they are needed. As a result, the expression of these genes is continually being "turned on" and "turned off" in response to changes in the environment.

As it turns out, *gene expression can be (and is) regulated at several different levels*—for example, transcription, mRNA processing, mRNA turnover, translation, and enzyme function. However, extensive data indicate that the *regulation of transcription is the most important mode of the control of gene expression*, at least in prokaryotes. That is not to say that regulation does not occur at other levels. Regulatory fine-tuning at translational levels is clearly important in the overall control of metabolic processes in living organisms. The regulatory mechanisms with the largest effects on phenotype, however, have been shown to act at the level of transcription.

Based on what is presently known about the regulation of transcription in both prokaryotes and eukaryotes, the various regulatory mechanisms seem to fit into two general categories. The first, and best understood, category includes mechanisms involved in the *rapid turn-on and turn-off of gene expression in response to environmental changes*. Regulatory mechanisms of this type are very important in microorganisms because of the frequent exposure of these organisms to sudden changes in environment. They provide microorganisms with a great deal of "plasticity," an ability to rapidly adjust their metabolic processes in order to achieve maximal growth and reproduction under highly variable environmental conditions. These quick responding on–off switches seem to be less important in higher eukaryotes. This might be expected since the circulatory systems of higher eukaryotes buffer their cells against many sudden environmental changes.

The second major category of regulatory mechanisms includes what might be called *preprogrammed circuits of gene expression*. In these cases, some event (such as infection by a virus) triggers the expression of one set of genes. The product (or products) of one (or more) of these genes functions by turning off the transcription of the first set of genes and/or turning on the transcription of a second set of genes. In turn, one or more of the products of the second set acts by turning on a third set, and so on. In these cases, the sequential expression of genes is genetically preprogrammed, and the genes usually *cannot* be turned on out of sequence. Such preprogrammed sequences of gene expression in viral infections are well documented. In most of these preprogrammed sequences, it seems the circuitry is cyclical. For example, in viral infections some event associated with the packaging of the viral DNA or RNA inside the protein coat somehow

seems to reset the program so that the first set of genes will again be expressed when a progeny virus subsequently infects another host cell.

INDUCTION AND REPRESSION IN PROKARYOTES

Certain gene-products, such as tRNA molecules, rRNA molecules, ribosomal proteins, RNA polymerase components (polypeptides), and other enzymes catalyzing metabolic processes that are frequently referred to as cellular "housekeeping" functions, are essential components of almost all living cells. Genes that specify products of this type are *continually being expressed* in most cells. Such genes are said to be expressed *constitutively* and are frequently referred to as *constitutive genes*.

Other gene-products are needed for cell growth only under certain environmental conditions. Constitutive synthesis of such gene-products would clearly be wasteful, using energy that could otherwise be utilized for more rapid growth and reproduction under the existing environmental conditions. The evolution of regulatory mechanisms that would provide for the synthesis of such gene-products only when and where they were needed would clearly provide organisms possessing these regulatory mechanisms with a selective advantage over organisms lacking these mechanisms. This undoubtedly explains why presently existing organisms, including the "primitive" bacteria and viruses, exhibit highly developed and very efficient mechanisms for the control of gene expression.

Escherichia coli and most other bacteria are capable of growth using any one of several carbohydrates (e.g., glucose, sucrose, galactose, arabinose, lactose) as an energy source. If glucose is present in the environment, it will be preferentially metabolized by *E. coli* cells. In the absence of glucose, however, *E. coli* cells can grow very well on other carbohydrates. Cells growing in medium containing the sugar lactose, for example, as the sole carbon source synthesize two enzymes, β-galactosidase and β-galactoside permease, that are uniquely required for the catabolism of lactose. (A third enzyme, β-galactoside transacetylase, is also synthesized. It has no *known* metabolic function, however.) β-galactosidase cleaves lactose into glucose and galactose, and β-galactoside permease pumps β-galactosides into the cell. Neither of these enzymes is of any use to *E. coli* cells when present in environments not containing lactose. The synthesis of these two enzymes, of course, requires the utilization of considerable energy (in the form of ATP and GTP; see Chapter 10). Thus, *E. coli* cells have evolved a regulatory mechanism by which the synthesis of these lactose-catabolizing enzymes is turned on in the presence of lactose and turned off in its absence.

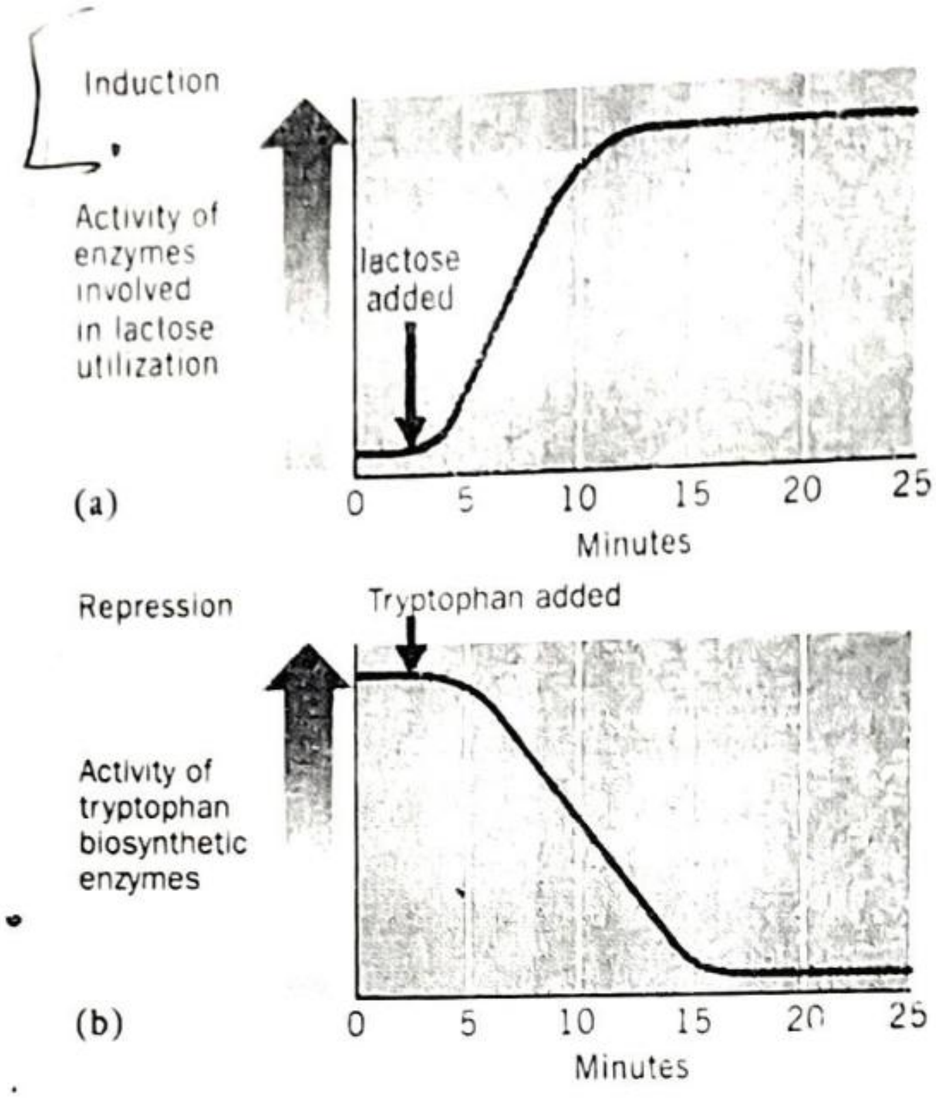


Figure 14.1 Induction (a) and ***repression*** (b) of enzyme synthesis in bacteria. Induction is characteristic of catabolic (degradative) pathways; repression is characteristic of anabolic (biosynthetic) pathways. (a) The induction of the synthesis of enzymes required for the utilization of the sugar lactose as an energy source in *E. coli* is illustrated. In the absence of lactose in the environment, *E. coli* cells synthesize only very minute amounts of the lactose-utilizing enzymes. When such cells are transferred to an environment containing lactose as the sole carbon source (occurring at the time indicated by the arrow labeled ***lactose added***), the synthesis of the enzymes required for lactose catabolism is rapidly ***induced*** (turned on). (b) The repression of the synthesis of the enzymes required for the biosynthesis of tryptophan in *E. coli* is illustrated. When tryptophan is not present in the environment, the *E. coli* cells synthesize the enzymes required for tryptophan biosynthesis. If tryptophan is added to the environment of such cells (e.g., at the time indicated by the arrow labeled ***tryptophan added***), the synthesis of the tryptophan biosynthetic enzymes is rapidly ***repressed*** (turned off). The kinetics shown are only approximate.

In natural environments (intestinal tracts and sewers), *E. coli* cells probably encounter an absence of glucose and the presence of lactose relatively infrequently. Most of the time, therefore, the *E. coli* genes coding for the enzymes involved in lactose utilization are not being expressed. If cells growing on a carbohydrate other than lactose are transferred to medium containing lactose as the only carbon source, they rapidly begin synthesizing the enzymes required for lactose utilization (Fig. 14.1*a*). This process, by which the expression of genes is turned on in response to a

substance in the environment, is called *induction*. Genes whose expression are so regulated are called *inducible genes*; their products, if enzymes, are called *inducible enzymes*. The substances or molecules responsible for induction are known as *inducers*.

Enzymes that are involved in *catabolic (degradative) pathways*, such as in lactose, galactose, or arabinose utilization, are characteristically inducible. As will become apparent in the following sections of this chapter, *induction occurs at the level of transcription*. It alters the rate of synthesis of enzymes, not the activity of existing enzyme molecules. Induction should not be confused with *enzyme activation*, in which the binding of a small molecule to an enzyme increases the *activity* of the enzyme (but does not affect its rate of synthesis).

Bacteria possess the metabolic capacity to synthesize most of the organic molecules (such as amino acids, purines, and vitamins) required for their growth. For example, *E. coli* has five genes coding for enzymes that are required in the synthesis of tryptophan. These five genes must be expressed in *E. coli* cells growing in an environment devoid of tryptophan in order to provide adequate amounts of this amino acid for ongoing protein synthesis.

When *E. coli* cells are present in an environment containing concentrations of tryptophan sufficient to support optimal growth, the continued synthesis of the tryptophan biosynthetic enzymes would be a waste of energy, because these bacteria have the capacity to take in external tryptophan. Thus, a regulatory mechanism has evolved in *E. coli* by which the synthesis of the tryptophan biosynthetic enzymes is turned off when tryptophan is present in the external milieu (Fig. 14.1*b*). This process of "turning off" the expression of sets of genes is called *repression*. A gene whose expression has been turned off in this way is said to be *repressed*; when its expression is turned on, a gene of this type is said to be *derepressed*.

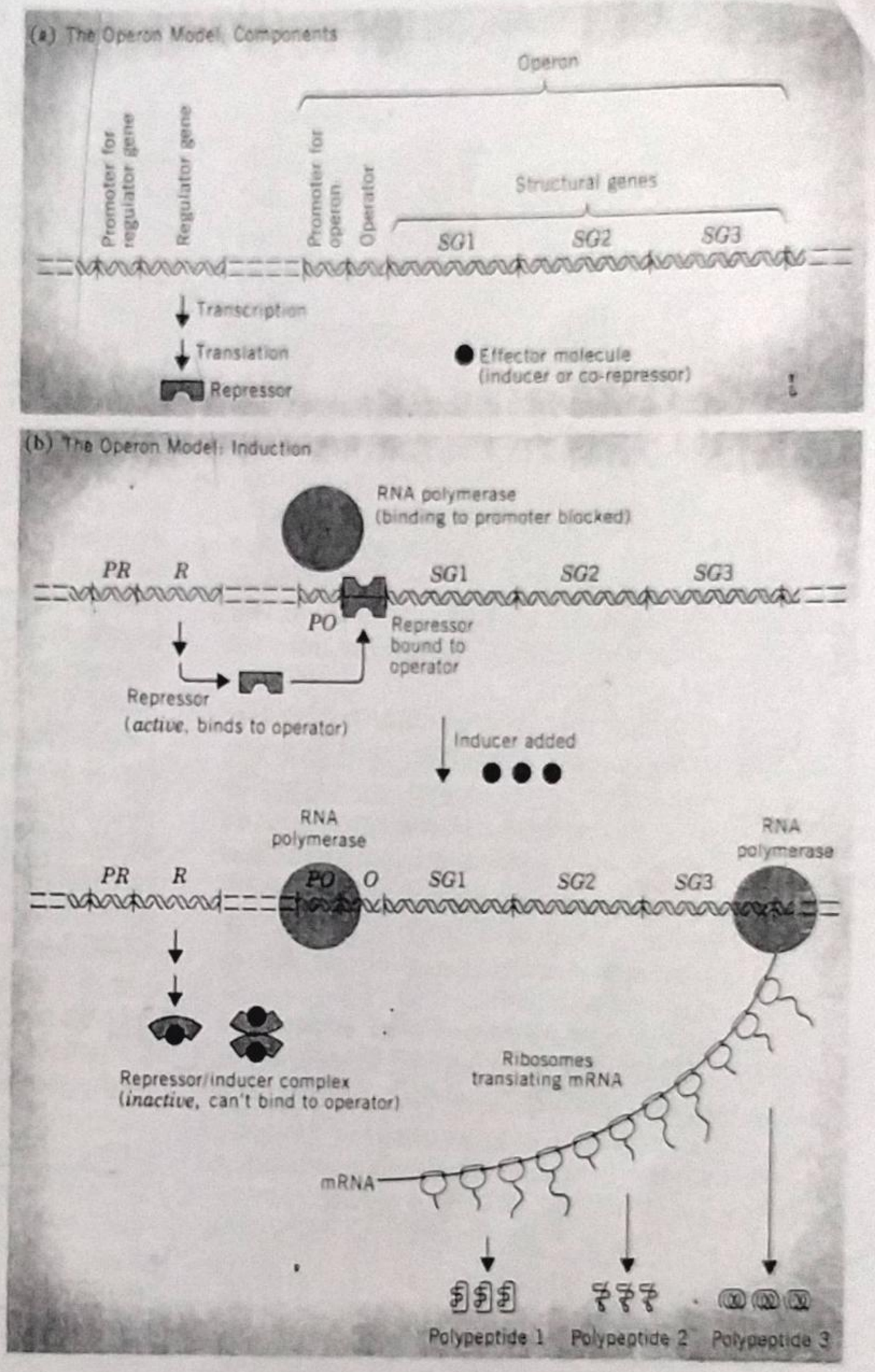
Enzymes that are components of *anabolic (biosynthetic) pathways* are frequently subject to *repression* (are *repressible*). *Repression*, like induction, *occurs at the level of transcription*. Repression should not be confused with *feedback inhibition*, in which the binding of an end product to the first enzyme in a biosynthetic pathway *inhibits* the *activity* of the enzyme (but does not affect its synthesis).

The Operon Model

Induction and repression of gene expression can be accomplished by essentially the same mechanism. This mechanism was first accurately described in 1961 when F. Jacob and J. Monod, both 1965 Nobel Prize recipients, proposed the *operon model* to explain the regulation of genes encoding the enzymes required for lactose utilization in *E. coli*. Jacob and Monod proposed that the transcription of one or a set of contiguous structural genes (genes coding for polypeptides) is regulated by two controlling elements (Fig. 14.2*a*). One of these elements, called the *regulator gene* (or *repressor gene*), codes for a protein called the *repressor*; under the appropriate conditions, the repressor binds to the second element, the *operator* (or *operator sequence*). The operator is always located contiguous to the structural gene or genes whose expression it regulates. *When the repressor is bound to the operator, transcription of the structural genes cannot occur.* We now know that this results because the binding of the repressor to the operator sterically prevents RNA polymerase from binding at the *promoter site* (the RNA polymerase binding site; see Chapter 10), which is always located contiguous with (or even overlapping) the operator sequence. The operator is usually located between the promoter and the structural genes (Fig. 14.2*a*). (The promoter was not recognized at the time of Jacob and Monod's proposal, but has since been shown to be an essential component of an operon.) The complete contiguous unit, including the structural gene or genes, the operator, and the promoter, is called an *operon*.

Whether the repressor will bind to the operator and turn off the transcription of the structural genes in an operon is determined by the presence or absence of *effector molecules* (small molecules such as amino acids and sugars) in the environment. In the case of inducible operons, these effector molecules are called *inducers*. Those active on repressible operons are called *co-repressors*. These effector molecules (inducers and co-repressors) act by binding to (or forming a complex with) the repressors.

The only essential difference between inducible operons and repressible operons is whether the naked repressor or the repressor–effector molecule complex is active in binding to the operator. (1) In the case of an inducible operon, the *free repressor binds to the operator*, turning off transcription (Fig. 14.2*b*). When the effector molecule (the inducer) is present, it binds to the repressor, releasing the repressor from the operator, that is, the *repressor–inducer complex cannot bind to the operator*. Thus, the addition of inducer turns on or induces the transcription of the structural genes in the operon (Fig. 14.2*b*). (2) In the case of a repressible operon, the situation is just reversed. *The free repressor cannot bind to the operator. Only the repressor–effector molecule (co-repressor) complex is active in binding to the operator* (Fig. 14.2*c*). Thus, transcription of the structural genes in a repressible operon is turned on in the absence of and turned off in the presence of the effector molecule (co-repressor).

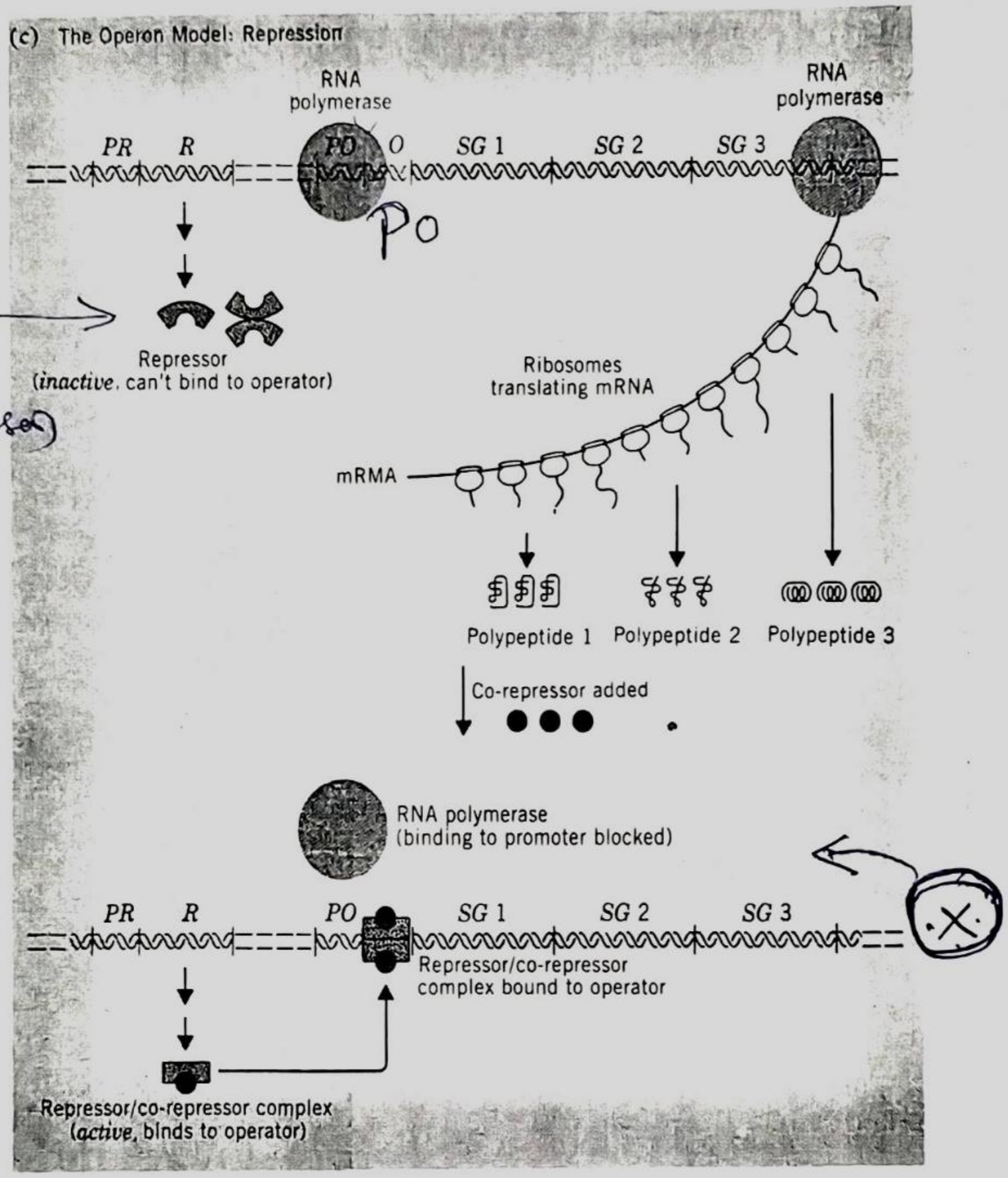


Except for this difference in the operator-binding behavior of the repressor, inducible and repressible operons are comparable.

A single mRNA transcript carries the coding information of an entire operon. Thus, the mRNAs of operons consisting of more than one structural gene are polygenic. For example, the tryptophan operon mRNA of *E. coli* is a huge macromolecule carrying the coding sequences that specify five different polypeptides (see Fig. 14.5). Because of their cotranscription, all the structural genes in an operon are *coordinately expressed*.

Because the product of the regulator gene, the repressor, acts by shutting off the transcription of structural genes, the operon model, as originally proposed by Jacob and Monod, is referred to as a *negative control* system. In *positive control* systems, the products of regulator genes are required to turn on tran-

Figure 14.2 The operon model for regulation of gene expression. (a) Diagram showing the essential components of regulation specified by the operon model. The *operon* consists of one or more ***structural genes*** (three—*SG1*, *SG2*, and *SG3*—are arbitrarily shown) and the adjoining *operator* and ***promoter*** sequences. The promoter for the operon (*PO*) is the site at which RNA polymerase must bind to initiate transcription of the structural genes. The operator (*O*) is the site at which the protein ***repressor***—the product of the ***regulator gene*** (or ***repressor gene***)—binds. The regulator gene need not be closely linked to the operon; in fact, it can be located at any position in the genome. The transcription of the regulator gene is initiated by RNA polymerase, which binds to its promoter (labeled *PR*, for promoter for regulator gene). When the repressor is bound to the operator, it sterically prevents RNA polymerase from binding to the adjoining promoter (*PO*) and from initiating transcription of the structural genes. Whether the repressor binds to the operator or not depends on the presence or absence of a metabolite called an ***effector molecule***. (b) Mode of regulation of gene expression for an ***inducible operon***. The product of the regulator gene (*R*), the repressor, in the ***absence*** of



the effector molecule (called an ***inducer,*** for inducible operons), binds to the operator, preventing RNA polymerase from binding to the promoter for the operon (*PO*). Thus, transcription of the structural genes ***cannot*** occur. When inducer is added, it binds to the repressor, causing it to be released from the operator (*O*). This, in turn, allows RNA polymerase to bind to the promoter (*PO*) and initiate transcription of the structural genes. The resulting multigenic mRNA is rapidly translated by ribosomes, producing the three polypeptide products of the structural genes. (c) Mode of regulation of gene expression for a ***repressible operon.*** In this case, the repressor can ***only bind to the operator*** in the ***presence*** of the ***effector molecule*** (called a ***co-repressor,*** for repressible operons). In its absence, the operator is free, permitting RNA polymerase to bind at the adjoining promoter (*PO*) and to initiate transcription of the structural genes. When co-repressor is added, it forms a complex with the repressor. This repressor–co-repressor complex then binds to the operator (*O*). This, in turn, prevents RNA polymerase from binding at *PO* and transcribing the three structural genes.

scription. We will discuss examples of positive control mechanisms later in this chapter.

lac, an Inducible Operon

Jacob and Monod proposed the operon model largely as a result of their studies of the *lac* operon of *E. coli*. More is known about the *lac* operon than any other operon. The *lac* operon contains a promoter, an operator, and three structural genes, *z*, *y*, and *a*, coding for the enzymes β-galactosidase, β-galactoside permease, and β-galactoside transacetylase, respectively (Fig. 14.3). β-galactoside permease "pumps" lactose into the cell, where β-galactosidase cleaves it into glucose and galactose (Fig. 14.4). The function of the transacetylase is still not clear.

The *lac* regulator gene, designated the *i* gene, codes for a repressor that is 360 amino acids in length.

Encoding for Enz. Biosynthesis

trp R – str. Gene (5 sets)

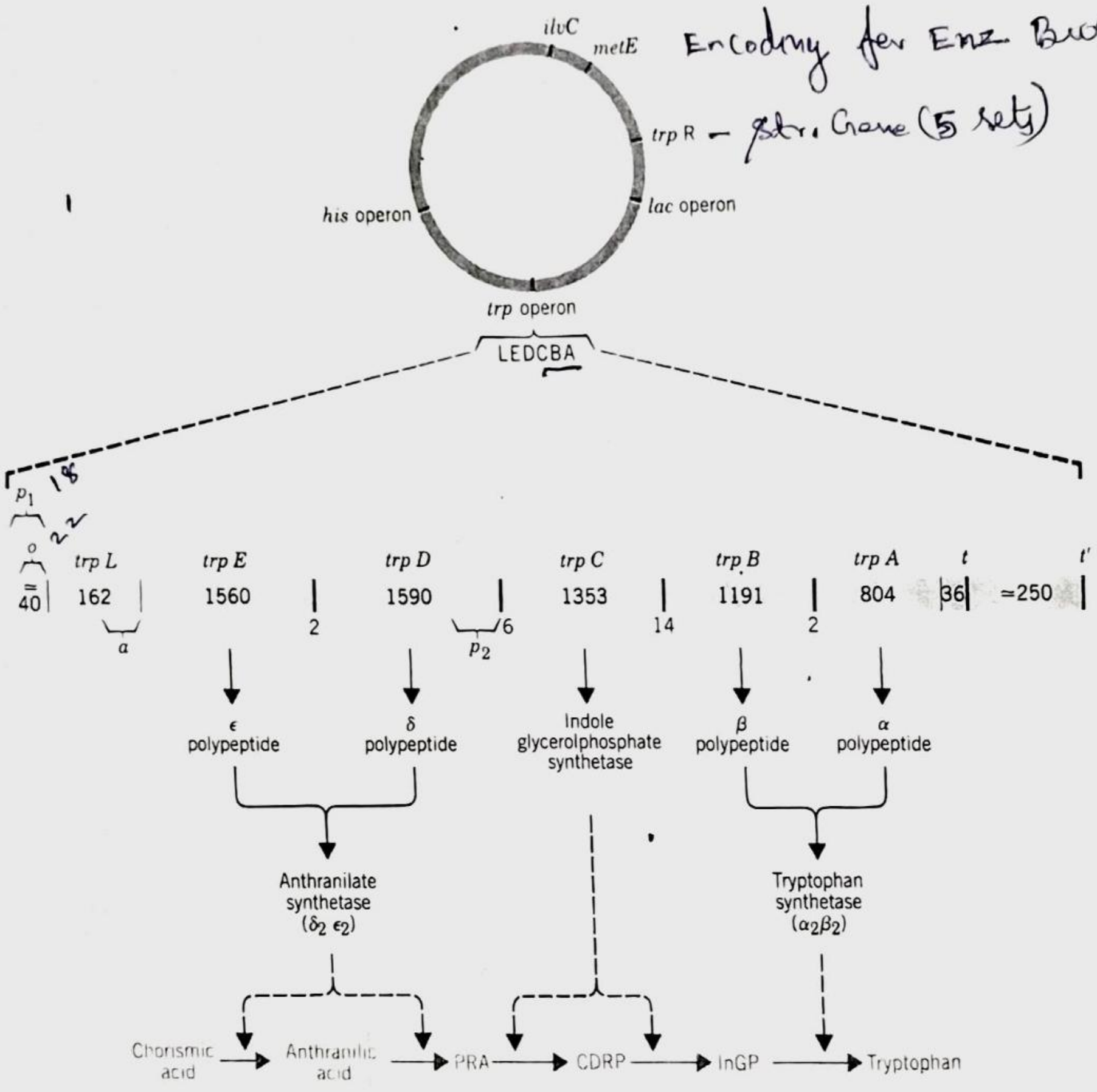


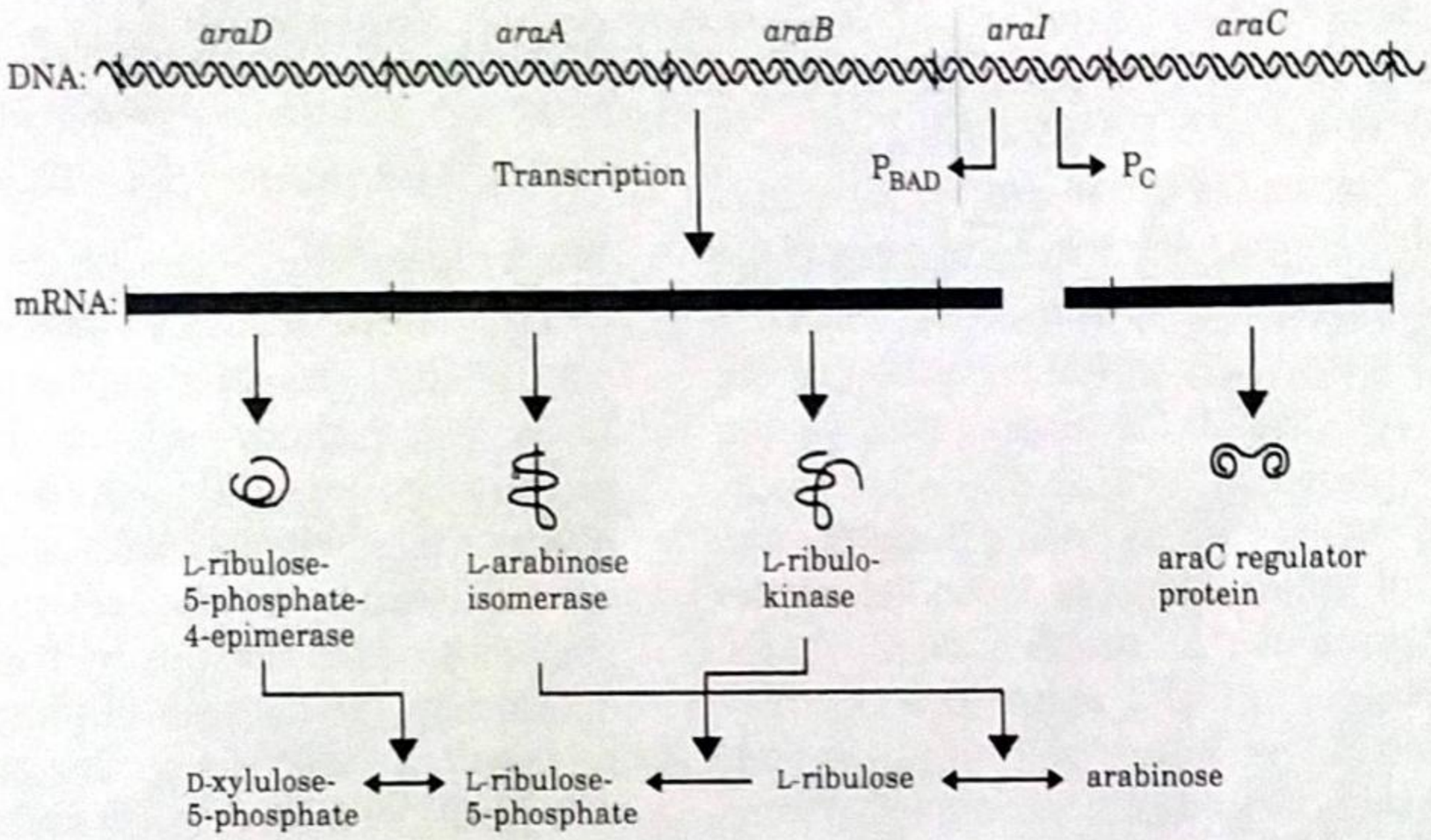
Figure 14.5 Organization of the tryptophan (*trp*) operon of *E. coli*. (The *Salmonella typhimurium trp* operon is essentially identical.) The operon contains five structural genes that encode the enzymes involved in the biosynthesis of tryptophan as shown at the bottom. The *trpR* gene, which encodes the *trp* repressor, is not closely linked to the *trp* operon (top). The operator (*o*) region of the *trp* operon lies entirely within the primary promoter (p_1) region. There is also a weak promoter (p_2), at the operator–distal end of the *trpD* gene, that results in a somewhat increased basal level of constitutive transcription of the *trpC*, *B* and *A* genes. There are two transcription termination sequences (*t* and *t'*) downstream from *trpA*. The *trpL* region specifies a 162-nucleotide-long mRNA leader sequence; it contains the attenuator (*a*) region that provides a second level of control of *trp* operon expression (see pp. 403–406). The p_1 promoter actually extends about 18 nucleotide-pairs into *trpL*. The length of each gene or region is given in nucleotide-pairs; the intergenic distances are given in nucleotide-pairs below the gene sequence. The abbreviations used are: PRA = phosphoribosyl anthranilate; CDRP = carboxyphenylamino-deoxyribulose phosphate; InGP = indole-glycerol phosphate. (Based on the data summarized by C. Yanofsky, *Nature* 289: 751–758, 1981, and by T. Platt, *Cell* 24: 10–23, 1981.)

tryptophan-mediated termination of transcription in the *trpL* (mRNA leader) region of the operon (see pp. 403–406).

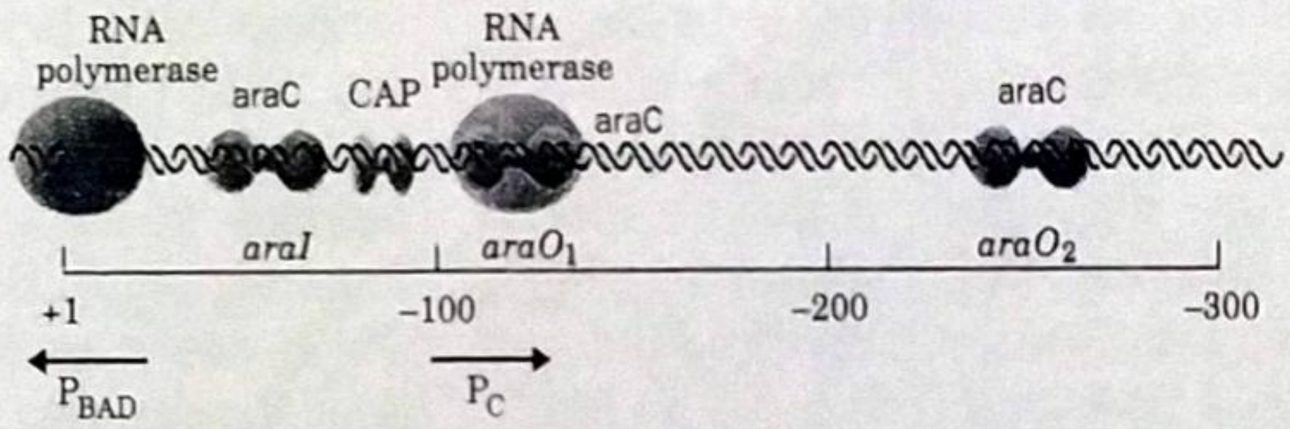
Positive Control of the *lac* Operon by CAP and Cyclic AMP

The operon model was proposed by Jacob and Monod to explain the induction of the biosynthesis of the enzymes involved in lactose utilization when this sugar is added to the medium in which *E. coli* cells are growing. The presence of glucose, however, has long been known to prevent the induction of the *lac* operon, as well as other operons controlling enzymes involved in carbohydrate catabolism (e.g., the arabinose and galactose operons). This phenomenon, called *catabolite repression* (or the *glucose effect*), has apparently evolved to assure that glucose is metabolized when present, in preference to other, less efficient, energy sources.

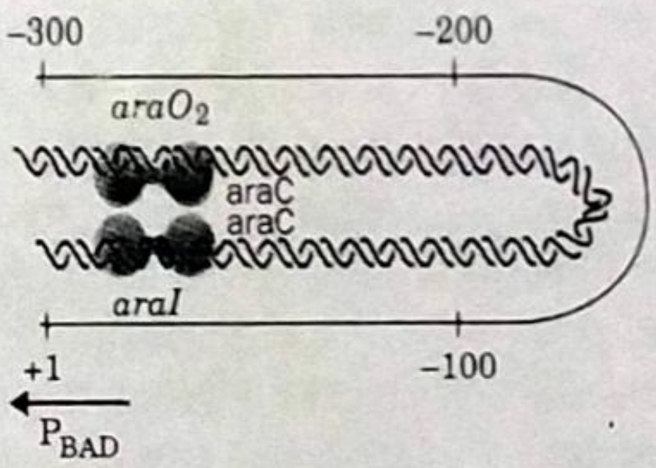
(a) *ara* operon: components



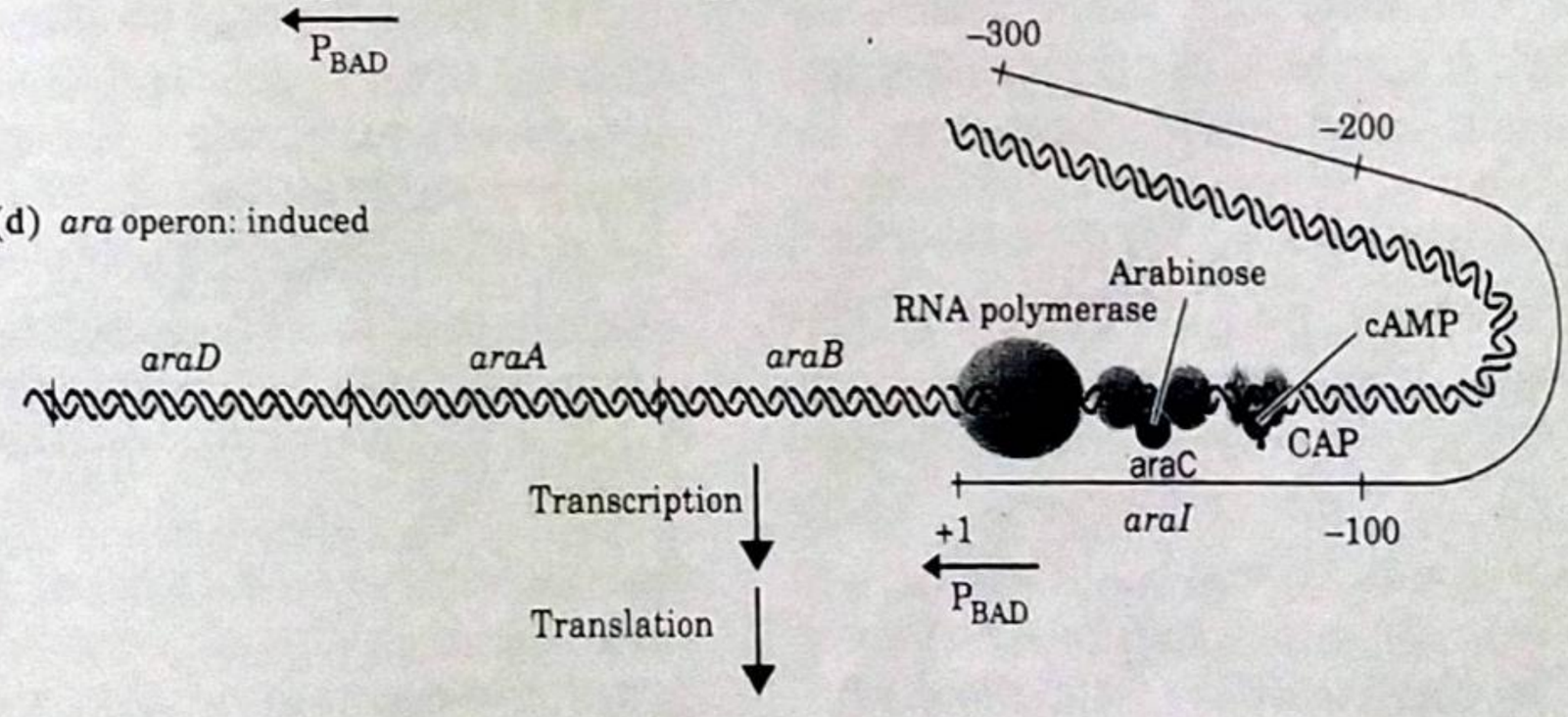
(b) *ara* operon: protein binding sites in regulatory region



(c) *ara* operon: repressed



(d) *ara* operon: induced



initiated at a promoter called P_C. Transcript extension occurs in opposite directions from the P_{BAD} and P_C promoters. The *araI* region contains the binding sites for the *araC* regulator protein and the cAMP–CAP complex. (b) Enlargement of the *araI* region showing the positions of the binding sites for *araC* protein, cAMP–CAP, and RNA polymerase, and the location of the *araC* protein binding site at $araO_2$. (c) Model for the repression of the *ara* operon. *AraC* protein dimers bind at the *araI* and $araO_2$ sites, and then bind to each other to form a DNA loop. RNA polymerase cannot bind at P_{BAD} to initiate transcription when the DNA is in this looped conformation. (d) Induction of the *ara* operon occurs when the arabinose–*araC* protein and cAMP–CAP complexes both bind at their *araI* sites. When these complexes are bound at their *araI* sites, the DNA loop structure cannot form, and RNA polymerase is able to bind at P_{BAD} and initiate transcription of the *araB*, *araA*, and *araD* structural genes. Additional details are given in the text. (After D.-H. Lee and R. F. Schleif, "*In Vivo* DNA Loops in *araCBAD*: Size limits and Helical Repeat," *Proc. Natl. Acad. Sci. U.S.A.* 86: 476–480, 1989.)

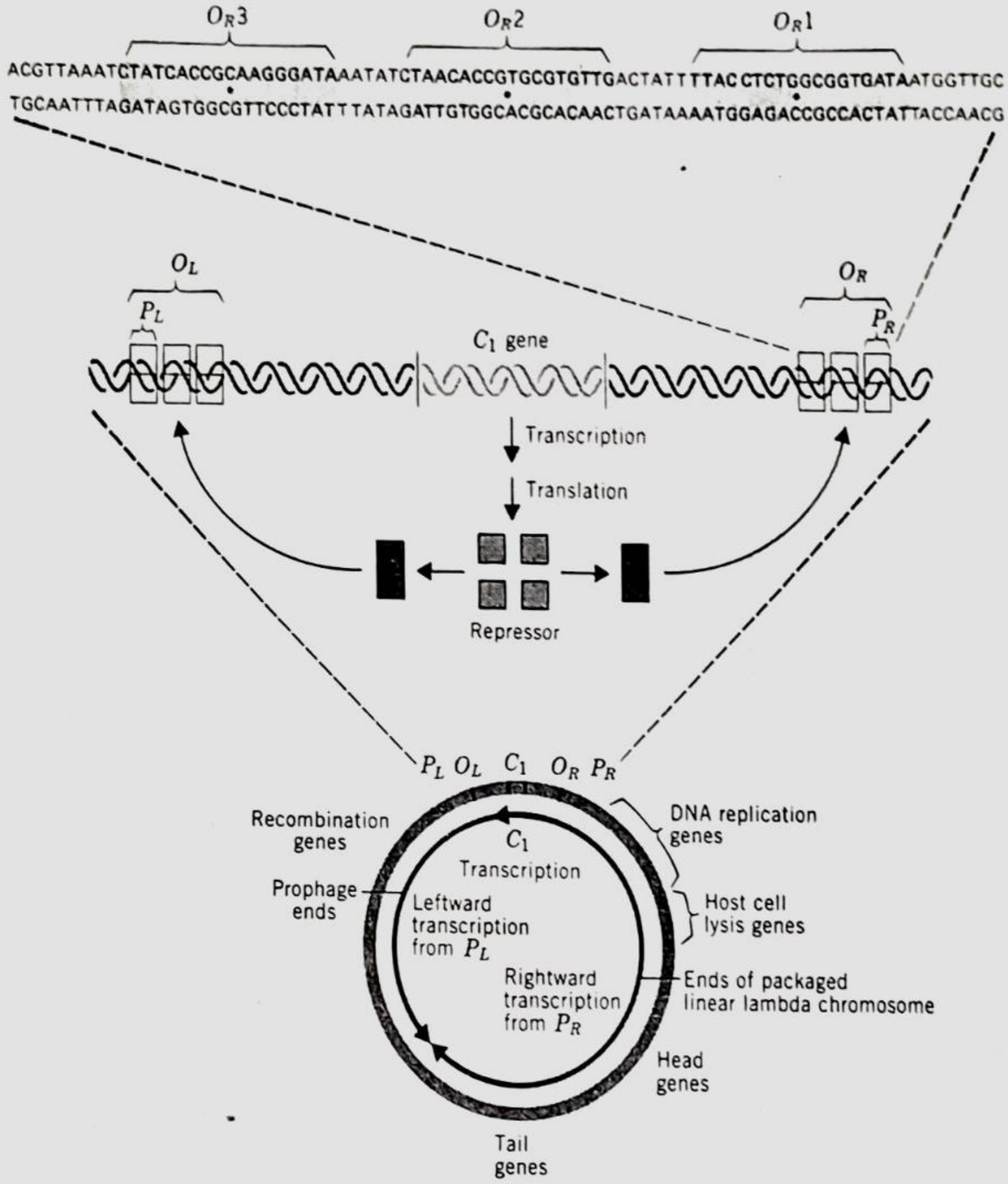
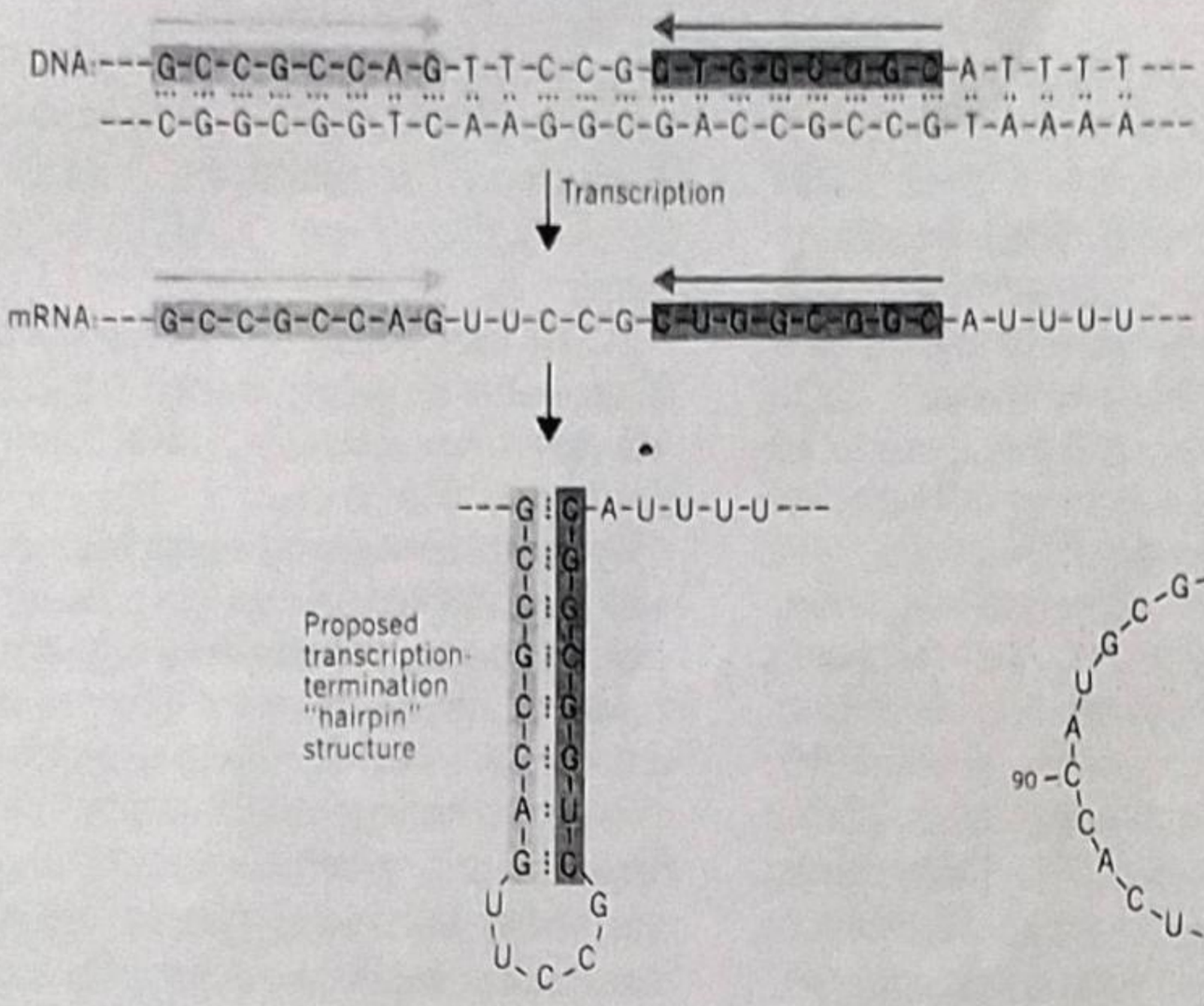


Figure 14.9 Repression of the phage lambda genes that control lytic development by the binding of the C_1 gene-product (repressor) to the two operator sequences (O_L and O_R) that control leftward and rightward transcription of the lambda chromosome. Transcription of the C_1 gene itself is in a leftward direction (solid black arrow), beginning at the C_1 promoter (not shown), which is located between the C_1 gene and O_R. The circular intracellular form of the lambda chromosome (see Chapter 5, pp. 110–113) is shown at the bottom, with the approximate location of the C_1 gene, P_LO_L (promoter leftward, operator leftward), O_RP_R (operator rightward, promoter rightward), the ends of the linear forms of the lambda chromosome (prophage and mature forms), and clusters of some of the genes controlling lytic development. The orange arrows indicate the transcription of genes from P_L and P_R, respectively. The repression of the genes controlling lytic development by the binding of repressor molecules to O_L and O_R is illustrated in the center enlargement. Note that each operator sequence has three repressor binding sites and that the operator and promoter (RNA polymerase binding site) overlap. The nucleotide-pair sequence of the O_R region is shown at the top, with the 17-nucleotide-pair sequence of each repressor binding site in brackets. The orange dot between the two strands of DNA within each repressor binding site indicates the axis of partial symmetry. (The nucleotide-pair sequence data are from M. Ptashne, K. Backman, M. Z. Humayun, A. Jeffrey, R. Maurer, B. Meyer, and R. T. Sauer, *Science* 194: 156, 1976.)

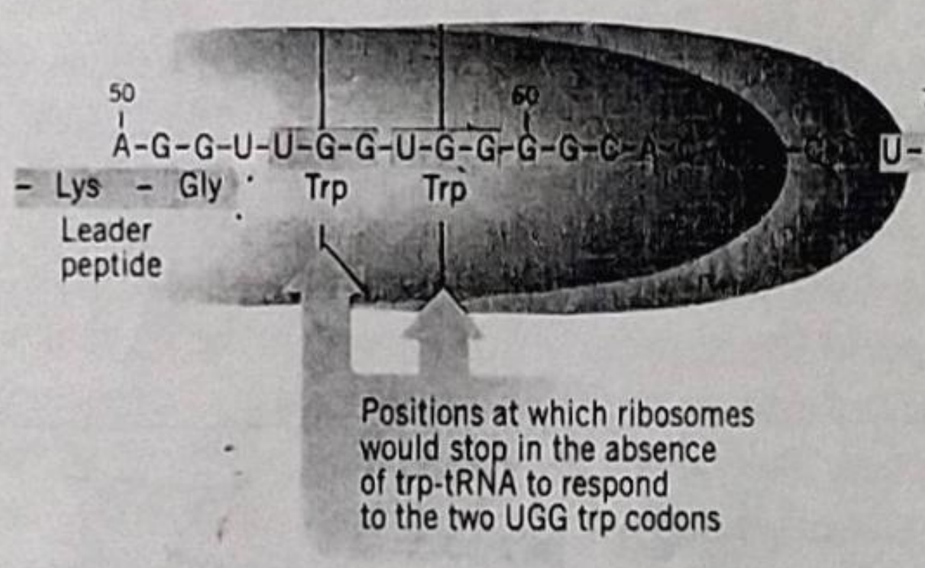
CONTROL OF THE *trp* OPERON BY ATTENUATION

Repression and derepression can change the level of expression of the structural genes of the *trp* operon by about 70-fold. There is a second level of regulation of *trp* operon expression, however. In *trpR* mutants that cannot make repressor, the addition of tryptophan to a culture of cells growing in the absence of tryptophan will cause an 8- to 10-fold decrease in the rates of synthesis of the tryptophan biosynthetic enzymes.

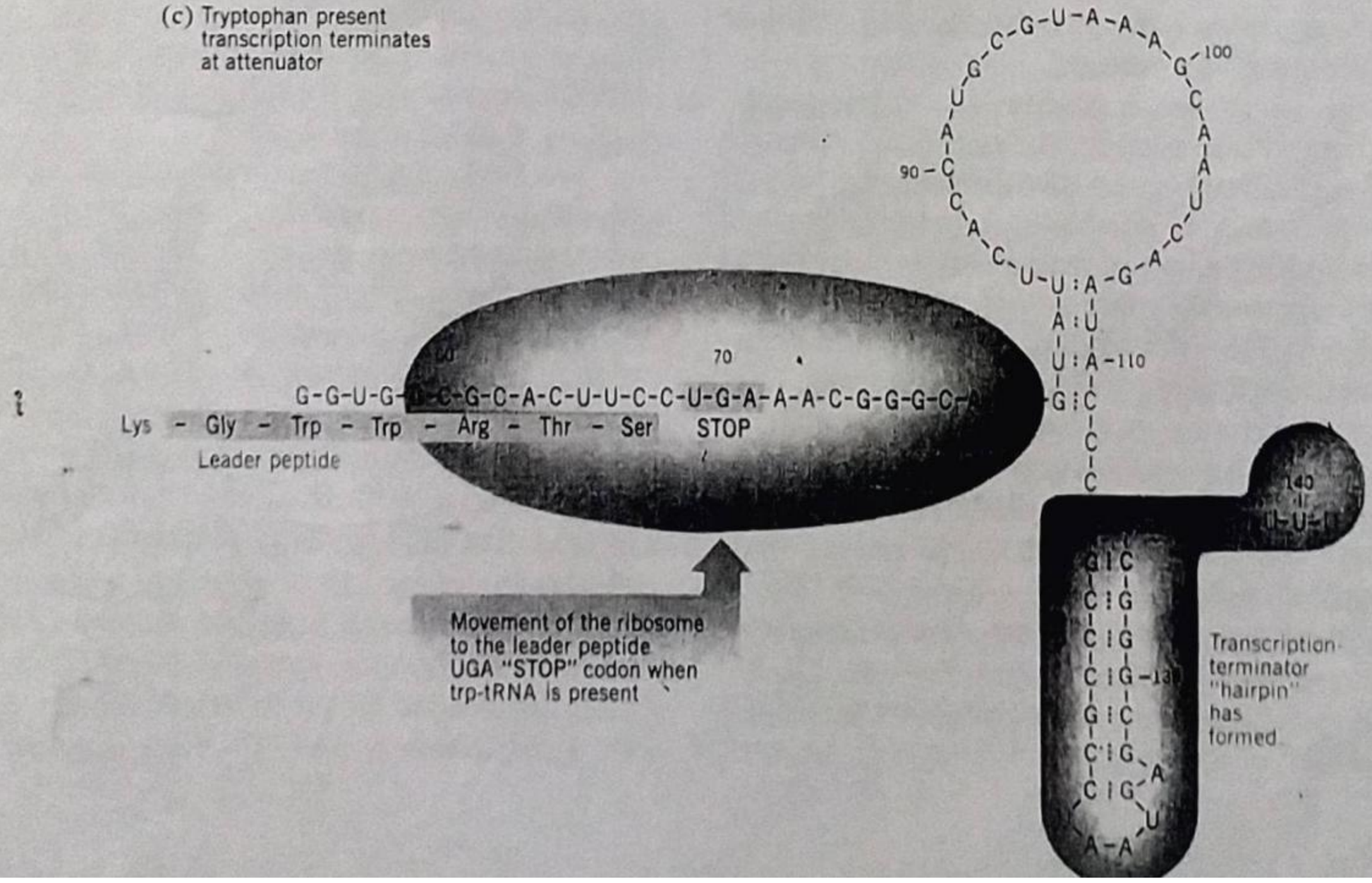
(a) Structure of trp operon transcription-termination sequence t



(b) Tryptophan absent transcription proceeds past attenuator



(c) Tryptophan present transcription terminates at attenuator



FEEDBACK INHIBITION AND ALLOSTERIC ENZYMES

Earlier in this chapter, we described the mechanism by which the transcription of bacterial genes coding for enzymes in a biosynthetic pathway is repressed when the end product of the pathway is present in the medium in which the cells are growing. A second, and more rapid, regulatory fine-tuning of metabolism often occurs at the level of enzyme activity. The presence of sufficient concentrations of an end product (such as histidine or tryptophan) of a biosynthetic pathway will frequently result in the inhibition of the first enzyme in the pathway. This phenomenon is called *feedback inhibition* or *end product inhibition*; it should *not be confused with repression (inhibition of enzyme synthesis)*. Feedback inhibition results in an almost instantaneous arrest of the synthesis of an end product when it is added to the medium.

Feedback inhibition-sensitive enzymes have been shown to have an end product binding site (or sites) in addition to the substrate binding site (or sites). In the case of some multimeric enzymes, the *end product or regulatory binding site* is on a different subunit (polypeptide) than the substrate site. Upon binding the end product, such enzymes are believed to undergo changes in conformation, called *allosteric transitions*, that reduce their affinity for their substrates. Proteins that undergo such conformational changes are usually referred to as *allosteric proteins*. Many examples are known, including numerous feedback inhibition-sensitive enzymes and the repressor molecules discussed in the preceding sections.