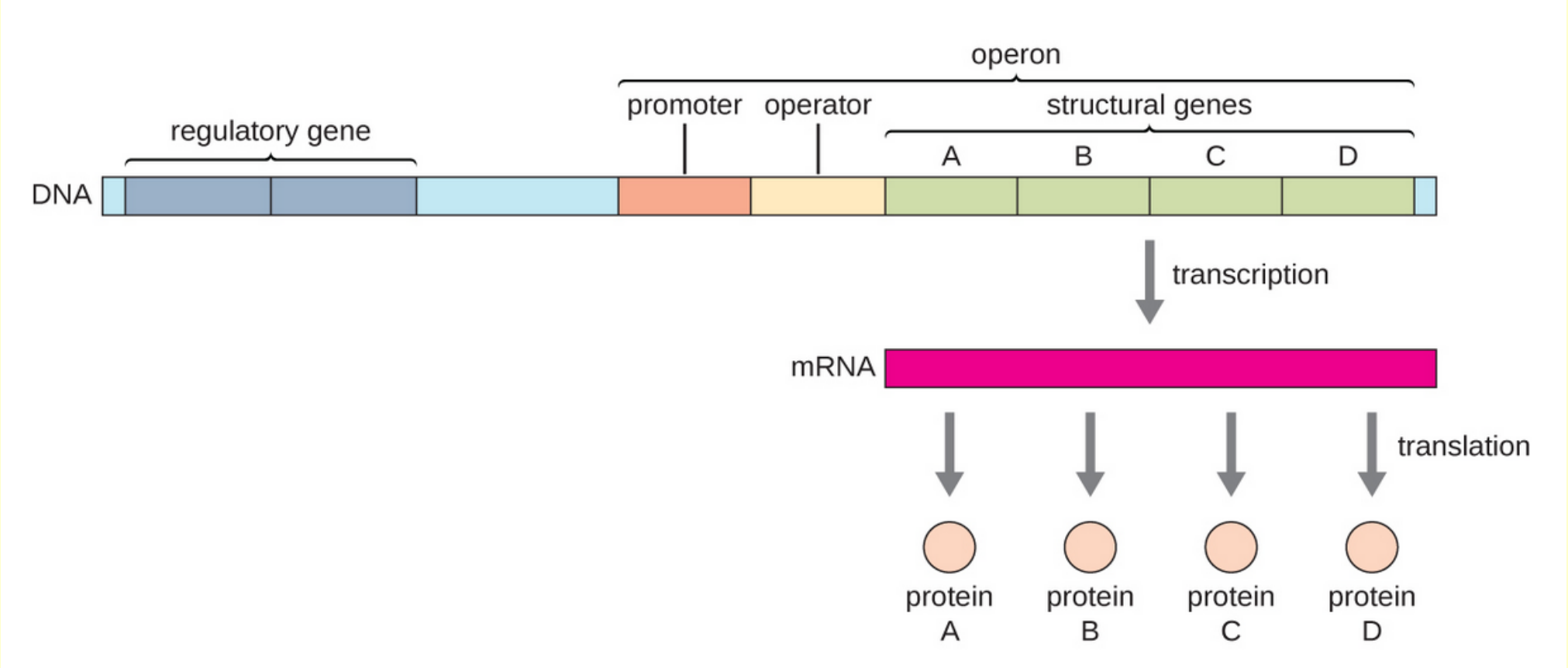
Gene Regulation in Prokaryotes

Introduction

In bacteria and archaea, clusters of genes under the control of a single promoter and are transcribed as a single large mRNA that contains multiple structural genes or cistrons called as an operon.

Each operon includes DNA sequences that influence its own transcription; these are located in a region called the regulatory region. The regulatory region includes the promoter and the region surrounding the promoter, to which **transcription factors**, proteins encoded by regulatory genes, can bind. Transcription factors influence the binding of **RNA polymerase** to the promoter and allow its progression to transcribe structural genes. A **repressor** is a transcription factor that suppresses transcription of a gene in response to an external stimulus by binding to a DNA sequence within the regulatory region called the **operator**, which is located between the RNA polymerase binding site of the promoter and the transcriptional start site of the first structural gene. Repressor binding physically blocks RNA polymerase from transcribing structural genes. Conversely, an **activator** is a transcription factor that increases the transcription of a gene in response to an external stimulus by facilitating RNA polymerase binding to the promoter. An **inducer**, a third type of regulatory molecule, is a small molecule that either activates or represses transcription by interacting with a repressor or an activator.



In prokaryotes, structural genes of related function are often organized together on the genome and transcribed together under the control of a single promoter. The operon's regulatory region includes both the promoter and the operator. If a repressor binds to the operator, then the structural genes will not be transcribed. Alternatively, activators may bind to the regulatory region, enhancing transcription.

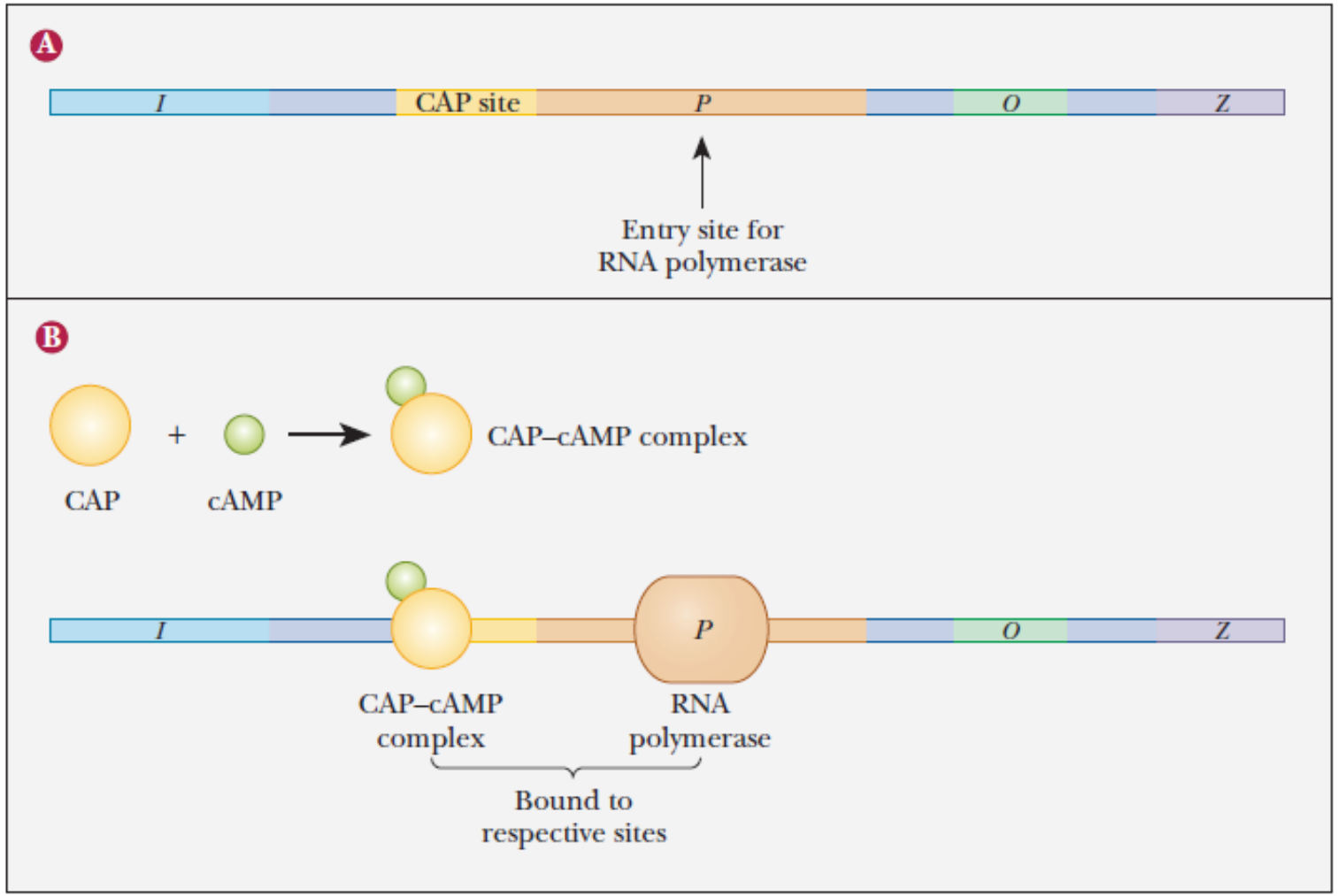
Lac Operon

French scientists François **Jacob** and Jacques **Monod** at the Pasteur Institute were the first to show the organization of bacterial genes into operons, through their studies on the ***lac* operon** of *E. coli*. They found that in *E. coli*, all of the structural genes that encode enzymes needed to use lactose as an energy source lie next to each other in the lactose (or *lac*) operon under the control of a single promoter, the *lac* promoter. For this work, they won the Nobel Prize in Physiology or Medicine in 1965. The lactose or *lac* operon is a cluster of three structural genes encoding proteins involved in lactose metabolism and the sites on the DNA involved in regulation of the operon. The three genes are: (1) *lacZ*, which encodes the enzyme β-galactosidase (which splits lactose into glucose and galactose); (2) *lacY*, which encodes lactose permease; and (3) *lacA*, which encodes a lactose transacetylase. Functional β-galactosidase and lactose permease are required for the utilization of lactose by this bacterium. These proteins are present in the cell in very low amounts when the organism is grown on carbon sources other than lactose. However, the presence of lactose and related compounds leads to the induction of the synthesis of these proteins. Interest in understanding the induction of β-galactosidase by its inducer, lactose.

Positive Gene Regulation

The *lac* operon is induced when *E. coli* has lactose, and no glucose, available to it as a carbon source. When both glucose and lactose are present, the cell does not make the *lac* proteins. The repression of the synthesis of the *lac* proteins by glucose is called **catabolite repression.** The mechanism by which *E. coli* recognizes the presence of glucose involves the promoter. The promoter has two regions. One is the binding site for RNA polymerase, and the other is the binding site for another regulatory protein, the **catabolite activator protein (CAP)** (Figure). The binding site for RNA polymerase also overlaps the binding site for the repressor in the operator region. The binding of CAP to the promoter depends on the presence or absence of 3,'5'-cyclic AMP (cAMP). When glucose is not present, cAMP is formed, serving as a “hunger signal” for the cell. CAP forms a complex with cAMP. The complex binds to the CAP site in the promoter region. When the complex is bound to the CAP site on the promoter, the RNA polymerase can bind at the binding site available to it and proceed with transcription (Figure). The *lac* promoter is particularly weak, and RNA polymerase binding is minimal in the absence of the CAP–cAMP complex bound to the CAP site. The CAP site is an example of an enhancer element, and the CAP–cAMP complex is a transcription factor. The modulation of transcription by CAP is a type of **positive regulation.**

Positive Gene Regulation

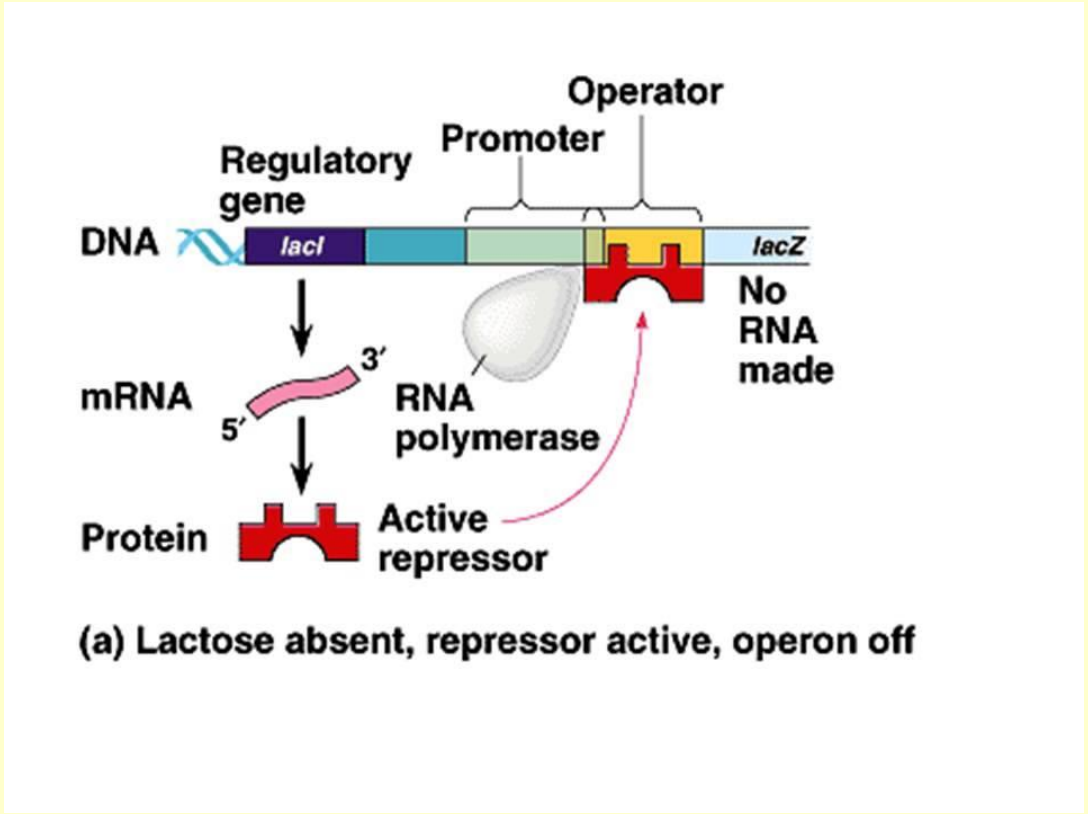


A The control sites of the *lac* operon. The CAP-cAMP complex, not CAP alone, binds to the CAP site of the *lac* promoter. When the CAP site on the promoter is not occupied, RNA polymerase does not bind. **B** In the absence of glucose, cAMP forms a complex with CAP. The complex binds to the CAP site, allowing RNA polymerase to bind to the entry site on the promoter and to transcribe the structural genes.

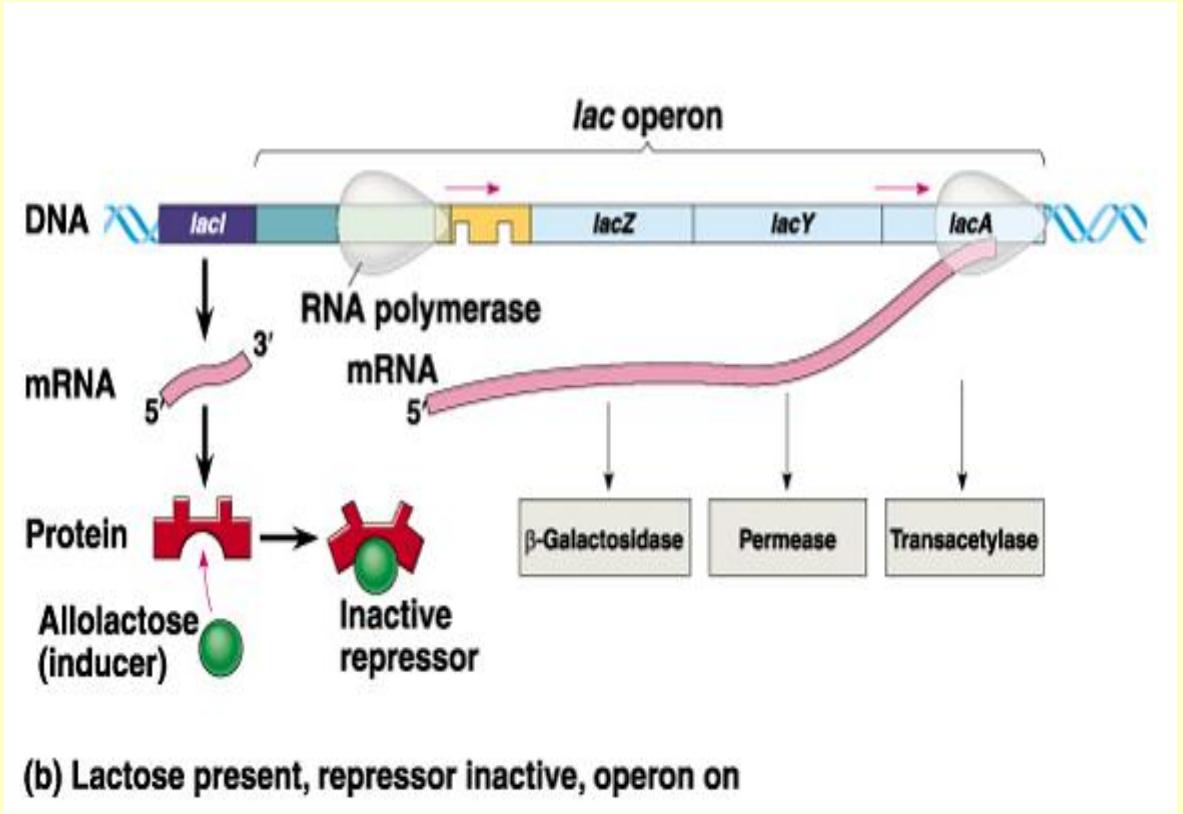
Negative Gene Regulation

- The *lac* operon in the bacterium *Escherichia coli* functions by a repression mechanism in which an inhibitor protein (lacI) binds to regulatory sites *(lacO)* in the promoter and turns off transcription (Fig.). On the addition of lactose, the lacI protein undergoes a conformational change, which changes its binding affinity for the *lacO* sequences. The lacI protein thereby comes off the *lacO* sites, and transcription can occur. *E. coli* uses this system to tightly control the genes required for the use of lactose, and it is completely reversible.
- The disaccharide *lactose* (a β-galactoside) is the substrate of β-galactosidase. The enzyme hydrolyzes the glycosidic linkage between galactose and glucose, the monosaccharides that are the component parts of lactose. *E. coli* can survive with lactose as its sole carbon source. To do so, the bacterium needs β-galactosidase to catalyze the first step in lactose degradation.
- The production of β-galactosidase takes place only in the presence of lactose, not in the presence of other carbon sources, such as glucose. A metabolite of lactose, allolactose, is the actual inducer, and β-galactosidase is an *inducible enzyme*.
- As the name indicates, the repressor inhibits the expression of the structural genes. In the presence of the inducer, this inhibition is removed. This is an example of **negative regulation** because the *lac* operon is turned on unless something is present to turn it off, which is the repressor in this case.

Negative Gene Regulation in *lac* operon



(a) Lactose absent, repressor active, operon off



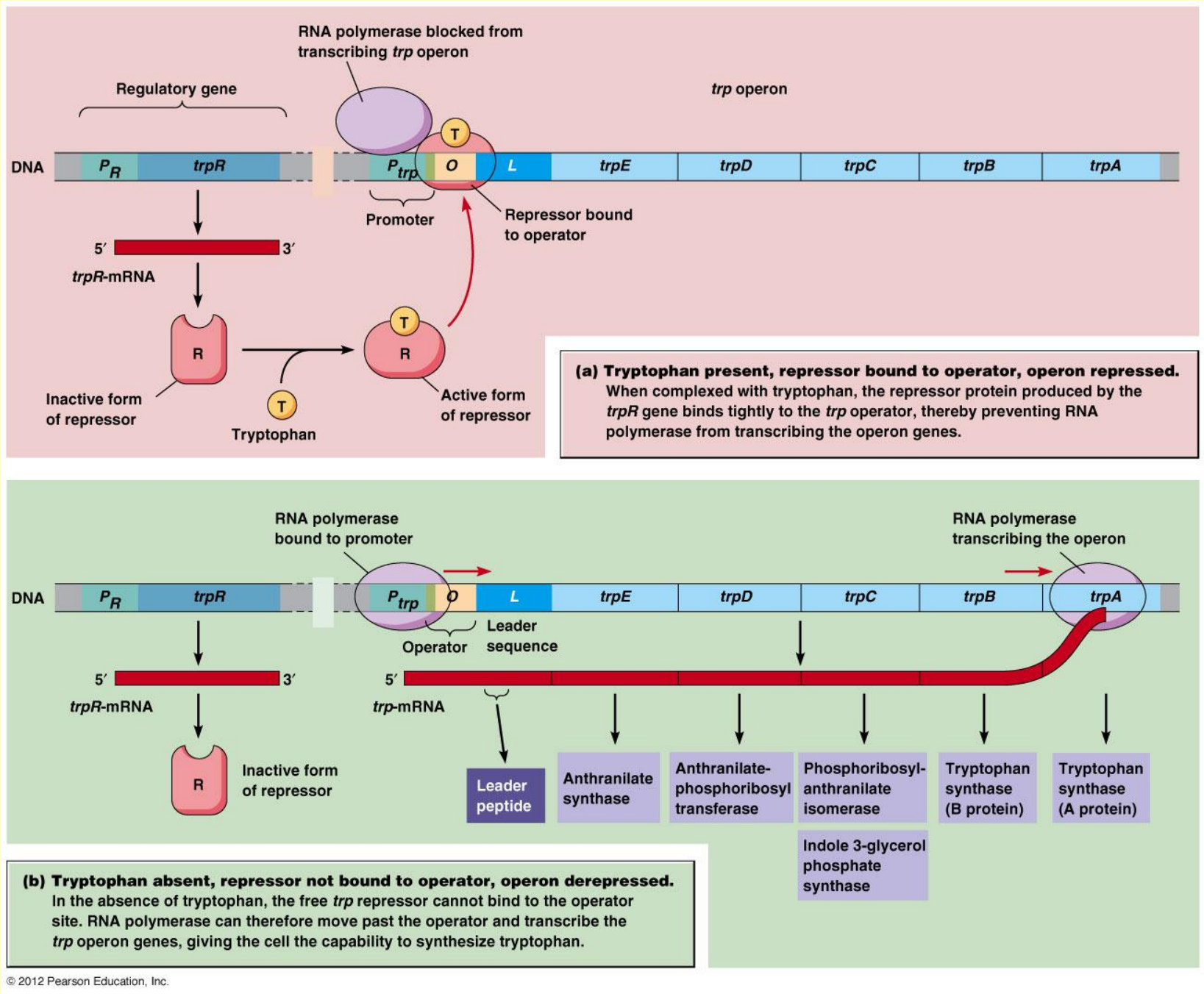
(b) Lactose present, repressor inactive, operon on

- When the cell has an adequate supply of glucose, the level of cAMP is low. CAP binds to the promoter only when it is complexed to cAMP. The combination of positive and negative regulation with the *lac* operon means that the presence of lactose is necessary, but not sufficient, for transcription of the operon structural genes. It takes the presence of lactose *and* the absence of glucose for the operon to be active.

The *trp* Operon: A Repressor Operon

- Bacteria such as *E. coli* need amino acids to survive. **Tryptophan** is one such amino acid that *E. coli* can ingest from the environment. *E. coli* can also synthesize tryptophan using enzymes that are encoded by five genes. These five genes are next to each other in what is called the **tryptophan (*trp*) operon** (Figure). If tryptophan is present in the environment, then *E. coli* does not need to synthesize it and the switch controlling the activation of the genes in the *trp* operon is switched off. However, when tryptophan availability is low, the switch controlling the operon is turned on, transcription is initiated, the genes are expressed, and tryptophan is synthesized.

- A DNA sequence that codes for proteins is referred to as the coding region. The five coding regions for the tryptophan biosynthesis enzymes are arranged sequentially on the chromosome in the operon. Just before the coding region is the **transcriptional start site**. This is the region of DNA to which RNA polymerase binds to initiate transcription. The promoter sequence is upstream of the transcriptional start site; each operon has a sequence within or near the promoter to which proteins (activators or repressors) can bind and regulate transcription.
- A DNA sequence called the operator sequence is encoded between the promoter region and the first *trp* coding gene. This **operator** contains the DNA code to which the repressor protein can bind. When tryptophan is present in the cell, two tryptophan molecules bind to the *trp* repressor, which changes shape to bind to the *trp* operator. Binding of the tryptophan–repressor complex at the operator physically prevents the RNA polymerase from binding, and transcribing the downstream genes.
- When tryptophan is not present in the cell, the repressor by itself does not bind to the operator; therefore, the operon is active and tryptophan is synthesized. Because the repressor protein actively binds to the operator to keep the genes turned off, the *trp* operon is negatively regulated and the proteins that bind to the operator to silence *trp* expression are **negative regulators**.



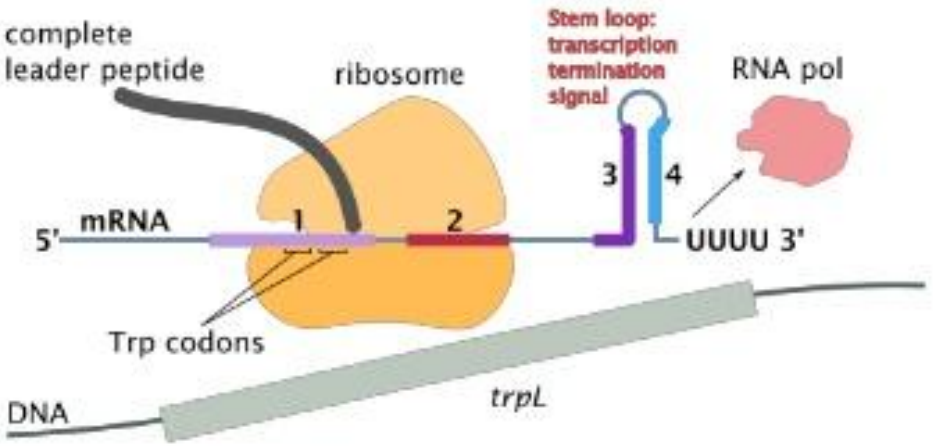
Transcription attenuation in *trp* operon

In addition to the standard negative regulation, the *trp* operon is regulated by another mechanism of control called transcription attenuation. This mechanism operates by causing premature termination of transcription of the operon when tryptophan is abundant. As shown in Figure, there are two loci, the *trp* leader and the *trp* attenuator, in between the operator and the gene *trpE*. Secondary structures formed in the mRNA of the leader sequence are responsible for this premature termination. The formation of such secondary structures comes from the transcription stop signals—an inverted repeat and a string of 8 A-T pairs in the attenuator. When tryptophan is scarce, the operon is translated normally. When it is plentiful, transcription is terminated prematurely after the leader sequences have been transcribed. Thus, attenuation imposes an extra level of control on an operon, over and above the repressor-operator system.

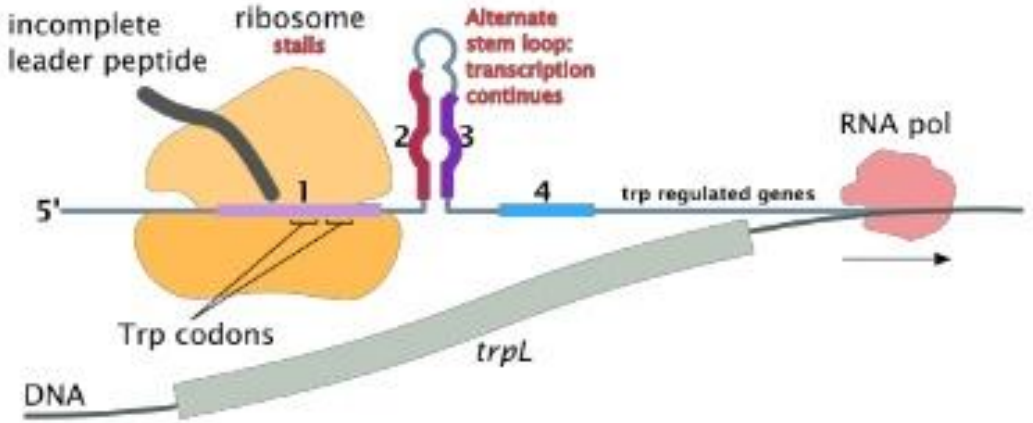
- The leader mRNA contains four regions capable of base pairing in various combinations to form hairpin structures. Attenuation depends upon the ability of **regions 1 and 2** and **regions 3 and 4** of the trp leader sequence to base pair and form hairpin secondary structures. A part of the leader mRNA containing regions 3 and 4 and a string of eight U's is called the ***attenuator***. The **region 3+4 hairpin** structure acts as a transcription termination signal; as soon as it forms, the RNA and the RNA polymerase are released from the DNA. During periods of tryptophan scarcity, a ribosome translating the coding sequence for the leader peptide may stall when it encounters the two tryptophan (trp) codons because of the shortage of tryptophan-carrying tRNA molecules. Because a stalled ribosome at this site blocks **region 1**, a **region 1+2** hairpin cannot form and an alternative, **region 2+3** hairpin is formed instead. The **region 2+3** base pairing prevents formation of the **region 3+4** transcription termination hairpin and therefore RNA polymerase can move on to transcribe the entire operon to produce enzymes that will synthesize tryptophan. When tryptophan is readily available, a ribosome can complete translation of the leader peptide without stalling. As it pauses at the stop codon, it blocks **region 2**, preventing it from base pairing. As a result, the **region 3+4** structure forms and terminates transcription near the end of the leader sequence and the structural genes of the operon are not transcribed (nor translated). This is example of a "riboswitch", a mechanism which can control transcription and translation through intereactions of molecules withe an mRNA.

ATTENUATION

High level of tryptophan



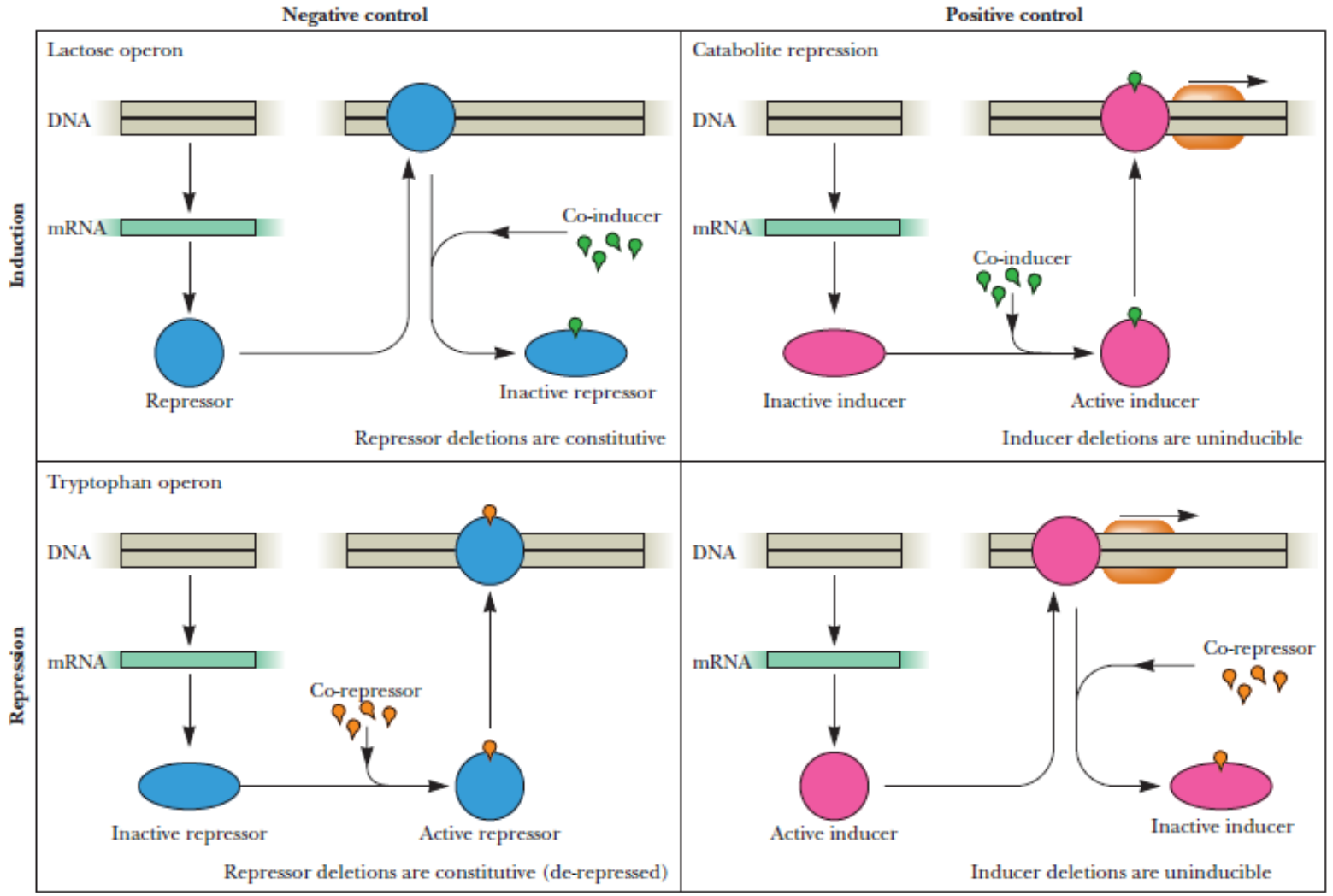
Low level of tryptophan



Positive and negative gene regulation

- Operons can be controlled by positive or negative regulation mechanisms. They are also classified as **inducible, repressible,** or both, depending on how they respond to the molecules that control their expression. There are four general possibilities(Figure). The top left figure shows a negative control system with induction. It is negative control because a repressor protein stops transcription when it binds to the promoter. It is an inducible system because the presence of the inducer or **co-inducer,** as it is often called, releases the repression, as we saw with the *lac* operon. Negative control systems can be identified by the fact that, if the gene for the repressor is mutated in some way that stops the expression of the repressor, the operon is always expressed. Genes that are always expressed are called **constitutive.** The top right figure shows a positively controlled inducible system. The controlling protein is an inducer that binds to the promoter, stimulating transcription, but it works only when bound to its co-inducer. This is what is seen with the catabolite activator protein with the *lac* operon. Such positively controlled systems can be identified by the fact that, if the gene for the inducer is mutated, it cannot be expressed—that is, it is **uninducible.** The bottom left figure shows a negatively controlled repressible system. A repressor stops transcription, but this repressor functions only in the presence of a **co-repressor.** The bottom right figure shows a positively controlled repressible system. An inducer protein binds to the promoter, stimulating transcription; but, in the presence of the co-repressor, the inducer is inactivated.

Positive and Negative Gene Regulation



Basic control mechanisms seen in the control of genes. They may be inducible or repressible, and they may be positively or negatively controlled.

Riboswitches

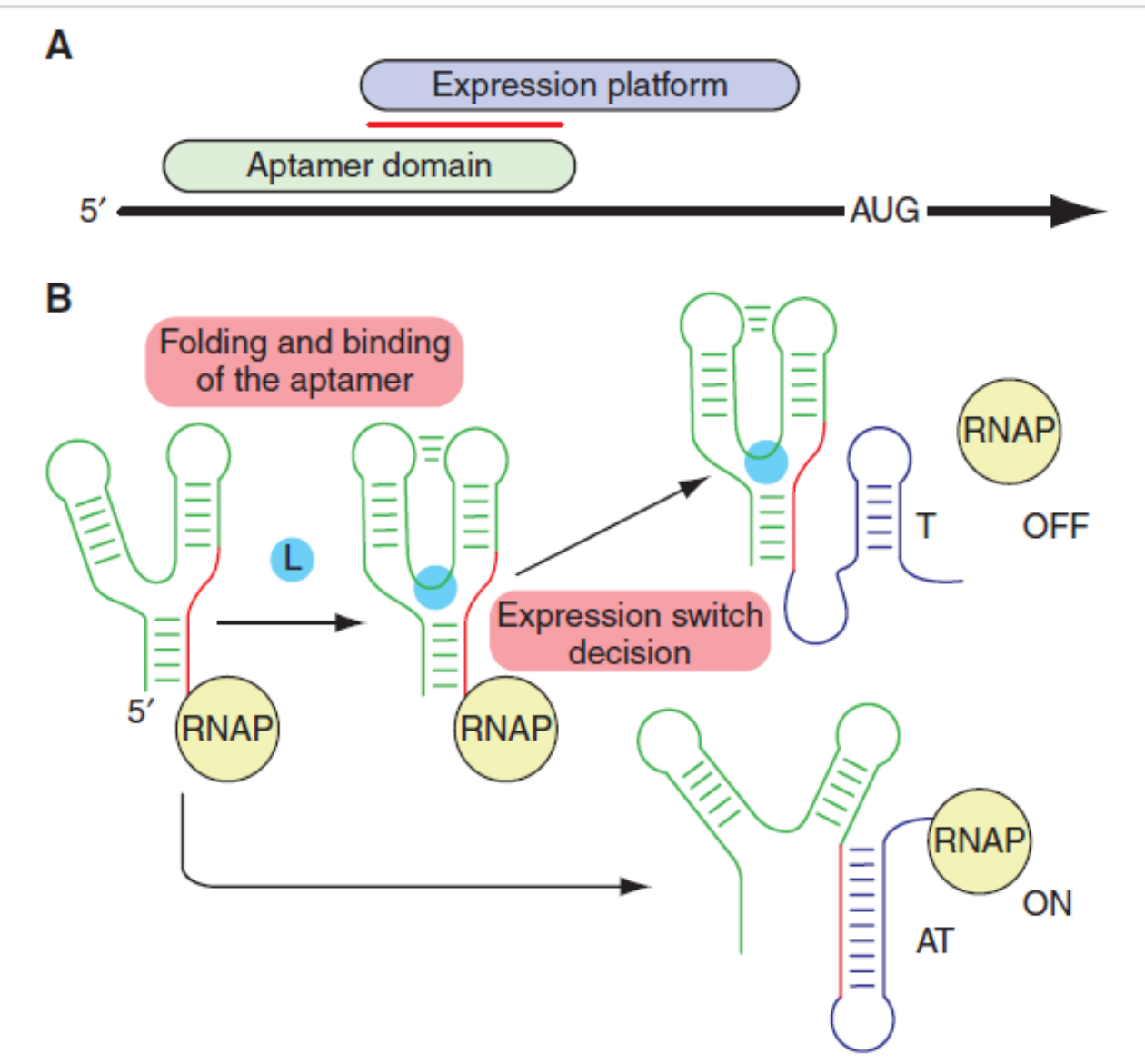
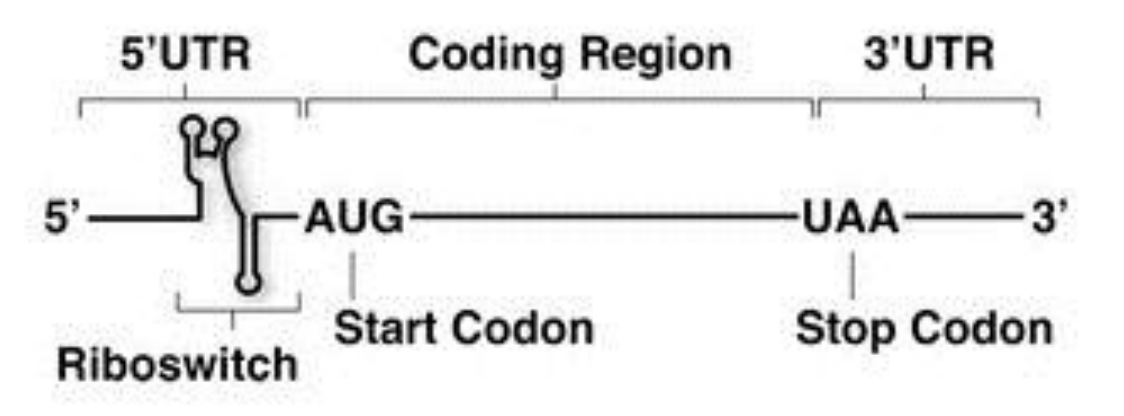
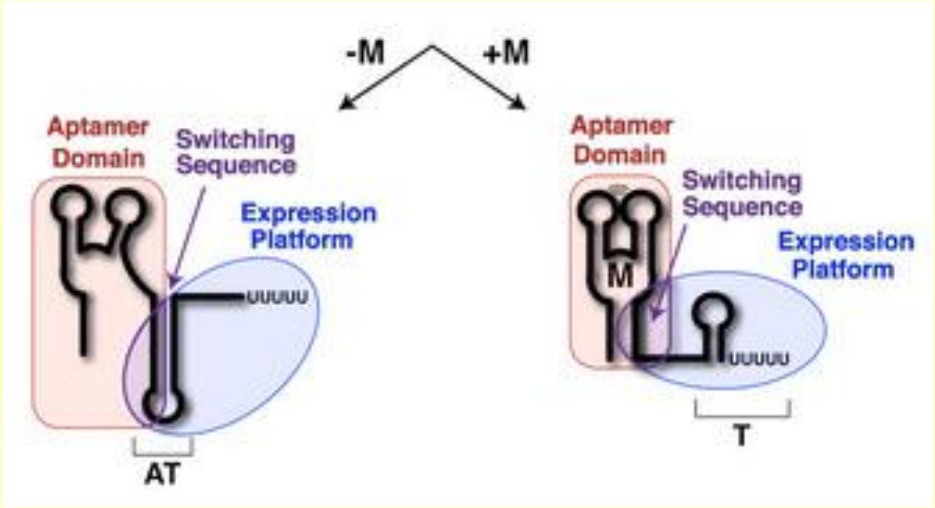


Figure1: Organization and mechanism of the typical riboswitch. (A) Arrangement of riboswitch structural elements in primary sequence of the mRNA transcript. The aptamer domain (green) and the expression platform (purple) overlap through a sequence that can base pair with either domain (red bar). (B) During transcription of the riboswitch, several events occur to elicit the appropriate regulatory outcome. Early events during transcription include the folding of the aptamer domain and potential binding of the effector (L,cyan). Depending on whether effector is bound, the RNA adopts one of two potential folds in the expression platform (an antiterminator, AT, or rho-independent terminator, T) that determines the regulatory response.

Riboswitches are elements commonly found in the 5′-untranslated region (UTR) of mRNAs that exert their regulatory control over the transcript in a cis-fashion by directly binding a small molecule ligand. The typical riboswitch contains two distinct functional domains (Fig.1A). The effector molecule is recognized by an aptamer domain, which adopts a compact three-dimensional fold to scaffold the ligand binding pocket. As with proteins, these RNA receptors must discriminate between chemically related metabolites with high selectivity to elicit the appropriate regulatory response. A second domain, the expression platform, contains a secondary structural switch that interfaces with the transcriptional or translational machinery. Regulation is achieved by virtue of a region of overlap between these two domains, known as the switching sequence, whose pairing directs folding of the RNA into one of two mutually exclusive structures in the expression platform that represent the on and off states of the mRNA (Fig.1B).



A typical bacterial mRNA transcript controlled by a riboregulatory element such as a riboswitch is composed of three sections: the 5' untranslated region (5' UTR), the protein-coding region beginning with the start codon (AUG) and ending with a stop codon (UAA), and the 3' untranslated region (3' UTR).



A riboswitch can adopt different secondary structures to effect gene regulation depending on whether ligand is bound. This schematic is an example of a riboswitch that controls transcription. When metabolite is not bound (-M), the expression platform incorporates the switching sequence into an antiterminator stem-loop (AT) and transcription proceeds through the coding region of the mRNA. When metabolite binds (+M), the switching sequence is incorporated into the aptamer domain, and the expression platform folds into a terminator stem-loop (T), causing transcription to abort. aptamer domain (red), switching sequence (purple), and expression platform (blue).

References

- Andrew D. Garst, Andrea L. Edwards, and Robert T. Batey, Riboswitches: Structures and Mechanisms, Cold Spring Harb Perspect Biol 2011;3:a003533
- https://www.nature.com/scitable/topicpage/riboswitches-a-common-rna-regulatory-element-14262702/#
- Biochem-Cambel-6th edition