Chapter 12

Gene expression and regulation

Bacterial genomes usually contain several thousand different genes. Some of the gene products are required by the cell under all growth conditions and are called **housekeeping genes**. These include the genes that encode such proteins as DNA polymerase, RNA polymerase, and DNA gyrase. Many other gene products are required under specific growth conditions. These include enzymes that synthesize amino acids, break down specific sugars, or respond to a specific environmental condition such as DNA damage.

Housekeeping genes must be expressed at some level all of the time. Frequently, as the cell grows faster, more of the housekeeping gene products are needed. Even under very slow growth, some of each housekeeping gene product is made. The gene products required for specific growth conditions are not needed all of the time. These genes are frequently expressed at extremely low levels, or not expressed at all when they are not needed and yet made when they are needed. This chapter will examine **gene regulation** or how bacteria regulate the expression of their genes so that the genes that are being expressed meet the needs of the cell for a specific growth condition.

Gene regulation can occur at three possible places in the production of an active gene product. First, the transcription of the gene can be regulated. This is known as **transcriptional regulation**. When the gene is transcribed and how much it is transcribed influences the amount of gene product that is made. Second, if the gene encodes a protein, it can be regulated at the translational level. This is known as **translational regulation**. How often the mRNA is translated influences the amount of gene product that is made. Third, gene products can be regulated after they are completely synthesized by either **post-transcriptional** or **post-translational regulation** mechanisms. Both RNA and protein can be regulated by degradation to control how much active gene product is present. Both can also be subjected to modifications such as the methylation of nucleosides in rRNA, the extensive modifications made to tRNAs (over 80 modified nucleosides have been described), or the phosphorylation of response-regulator proteins (see below). These modifications can play a major role in the function of the gene product.

In general, every step that is required to make an active gene product can be the focus of a regulatory event. In practice, most bacterial regulation occurs at the transcriptional level. Transcriptional regulation is thought to be more frequent because it would be a waste to make the RNA if neither the RNA nor its encoded protein is needed. The first three systems discussed—repression of the *lac* operon by Lac

repressor, activation of the *lac* operon by cAMP-CAP, and control of the *trp* operon by attenuation—describe classic transcriptional regulation mechanisms. The last three systems—regulation of the heat shock genes, regulation of the SOS response, and regulation of capsule synthesis—rely on different combinations of transcriptional, post-transcriptional, and post-translational regulatory mechanisms and will be used to demonstrate how several regulatory mechanisms can be integrated to fine-tune gene expression.

The players in the regulation game

Ribonucleic acid, or RNA, exists as messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). In most cases, RNA is a single-stranded, rather than a double-stranded, molecule. RNA participates in both genetic and functional activities. As mRNA, it allows the genetic information stored in the DNA molecule to be transmitted into proteins. As tRNA, it transfers amino acids during translation. As rRNA, it maintains the structure of the ribosome and helps carry out translation.

RNA is chemically synthesized by the action of RNA polymerase in a process called **transcription**. There are four identifiable steps during transcription: **promoter recognition**; **chain initiation**; **chain elongation**; and **chain termination** (Fig. 12.1). RNA polymerase catalyzes the formation of phosphodiester bonds between ribonucleotides using DNA as a template. Unlike DNA polymerase, RNA polymerase does not require a primer to begin synthesis of the RNA molecule (see Chapter 2 for details of DNA synthesis). Growth of the RNA chain, like the growth of a DNA chain, is in the 5′ to 3′ direction because RNA polymerase can only add a new nucleotide to a free 3′ OH group. The order in which the different ribonucleotides are added to a free 3′ OH is determined by a double-stranded DNA molecule in which one strand acts as the template (Fig. 12.2). The **template strand** is complementary to the RNA. The non-template or **coding strand** contains the same sequence as the RNA except for the substitution of uracil for thymine.

Core RNA polymerase is a complex composed of the proteins β, β′ and two subunits of α (Fig. 12.3a). Core RNA polymerase is responsible for the synthesis of RNA, but it imparts no specificity as to where the start of the RNA occurs. Core RNA polymerase is converted to **holoenzyme** when one additional protein, **sigma factor** (σ) is associated with it (Fig.

Fig. 12.1 The four major steps of transcription. (a) RNA polymerase recognizes the promoter. (b) RNA polymerase moves to the start site and begins polymerizing RNA. (c) RNA polymerase moves along the DNA template, elongating the RNA. (d) RNA polymerase stops RNA synthesis. The newly synthesized RNA disassociates from RNA polymerase and RNA polymerase disassociates from the DNA. RNA polymerase transcribes mRNA, rRNA, and tRNA.
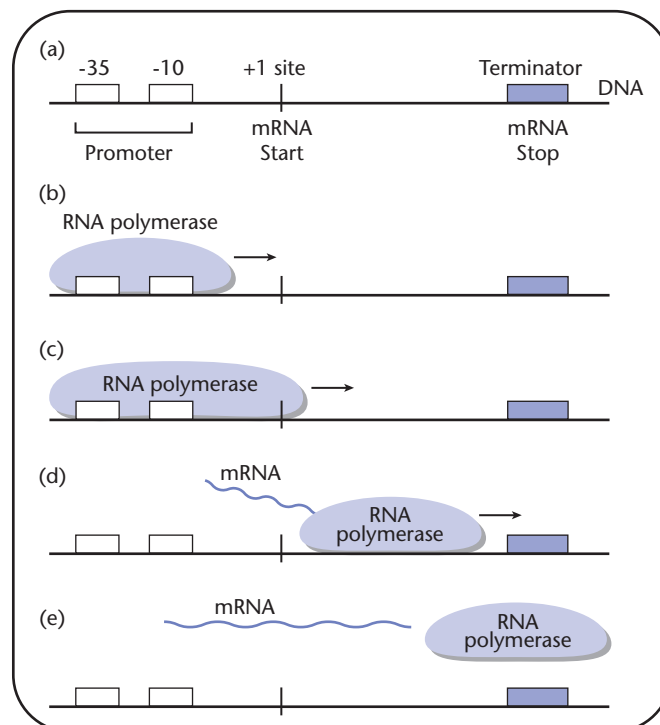
12.3a). σ factor directs RNA polymerase to specific sequences in the DNA called **promoters** so that transcription initiates at the proper place.

Bacteria can contain more than one σ factor. In *E. coli*, the major σ factor present under normal growth conditions is called sigma 70 or σ^{70}. σ^{70} recognizes promoters that have a specific DNA sequence and directs the RNA polymerase molecule that it is part of to begin transcription near these specific sequences. σ^{70} was named for its molecular weight, which is 70 kilodaltons. Different σ factors recognize different sequences as promoters. For example, when cells are exposed to an increase in temperature or heat shock a group of genes is induced to cope with this stress (see below). The expression of the heat-shock genes is controlled by an alternative σ factor and all of the heat-shock genes have a common promoter sequence that is recognized by the alternative σ factor. The heat-shock promoter sequence is different from the promoter sequence recognized by σ^{70}.

Promoters contain two distinct sequence motifs that reside ~10 bases and ~35 bases upstream of the transcriptional start site or first base of the RNA. The transcriptional start site is known as the **+1 site**. All of the bases following the +1 site are transcribed into RNA and are numbered consecutively with positive numbers (Fig. 12.3b). The bases prior to the +1 site are numbered consecutively with negative numbers. The motif at ~10 bases upstream of the +1 site is called the **–10 region** and the motif at ~35 bases upstream of the +1 is called the **–35 region**. σ^{70} recognizes promoters with a **consensus sequence** consisting of TAATAT at the –10 region and TTGACA at the –
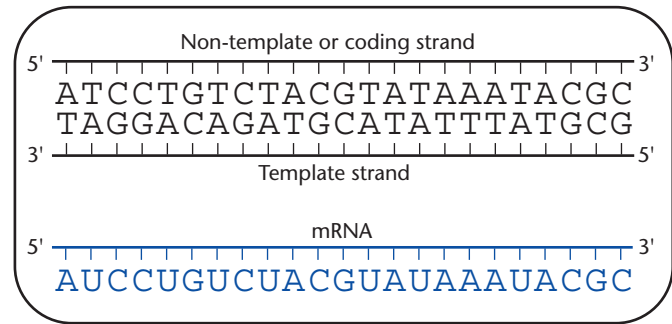

Fig. 12.2 The mRNA is synthesized using one strand of the DNA as a template. This makes the mRNA complementary to the template strand. When the DNA sequence encoding a gene is shown, by convention the DNA strand that is the same sequence as the mRNA (except for T to U) is usually shown.
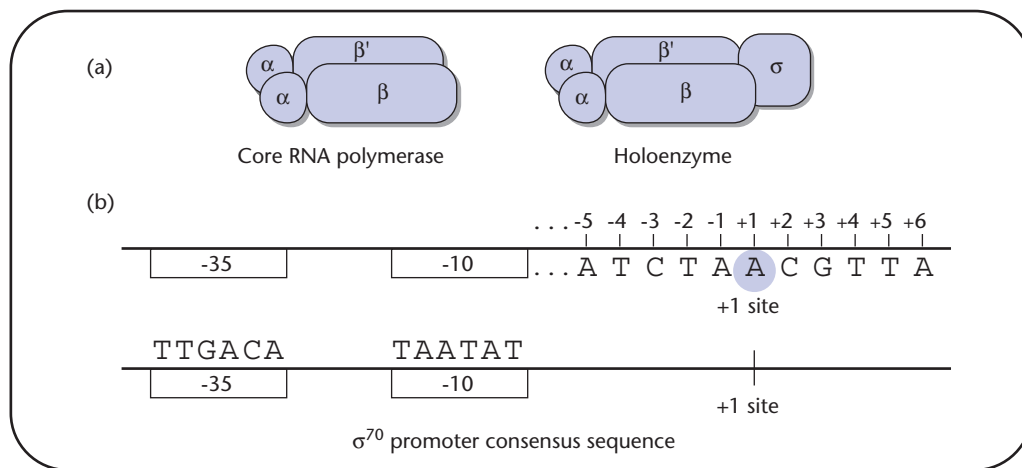

Fig. 12.3 The structure of RNA polymerase and the regions in the promoter it binds to. (a) RNA polymerase has two forms. Core RNA polymerase synthesizes RNA in vitro but starts at many places on the DNA. Holoenzyme also synthesizes RNA in vitro but it starts only at promoters. (b) The general organization of the signals used by RNA polymerase. RNA polymerase recognizes promoters, which consist of a –10 region and a –35 region. RNA polymerase begins synthesis of RNA at the +1 base. RNA polymerase containing σ^{70} recognizes specific sequences at the –10 and –35 regions. The closer a sequence is to the consensus, the better σ^{70} is at recognizing it as a promoter and the more times it will be transcribed into RNA. If any of the bases are changed, σ^{70} may still recognize the sequence as a promoter but it will do so with a lower affinity.

35 region (Fig. 12.3b). Promoters are defined according to their strength. This means that the stronger the promoter, the stronger the interaction between that promoter sequence and RNA polymerase. A general rule of thumb is that the closer the −10 and −35 sequences of a promoter are to the consensus sequence, the stronger the promoter.

RNA polymerase holoenzyme binds to promoter sequences and covers approximately 75 bases of the DNA from −55 to +20. Once bound, RNA polymerase initiates transcription by causing the double-stranded DNA template to open, effectively melting the hydrogen bonds that hold the two DNA strands together in the promoter region (Fig. 12.4). As RNA polymerase starts transcribing at the +1 site, it continues to open the double-stranded DNA molecule, creating a short region of single-stranded DNA. After RNA polymerase has passed through the opened DNA, this region will re-form hydrogen bonds to give a closed DNA molecule. Approximately two turns of the double helix or 17 base pairs are unwound at any given time during the elongation phase of transcription.

Although σ factor is needed for RNA polymerase to bind to the right promoter sequences, it is not needed during the elongation phase of transcription. After a short RNA transcript is synthesized, σ factor dissociates from RNA polymerase and core RNA polymerase continues to elongate the RNA transcript. Transcription continues until core RNA polymerase encounters a transcription termination signal or
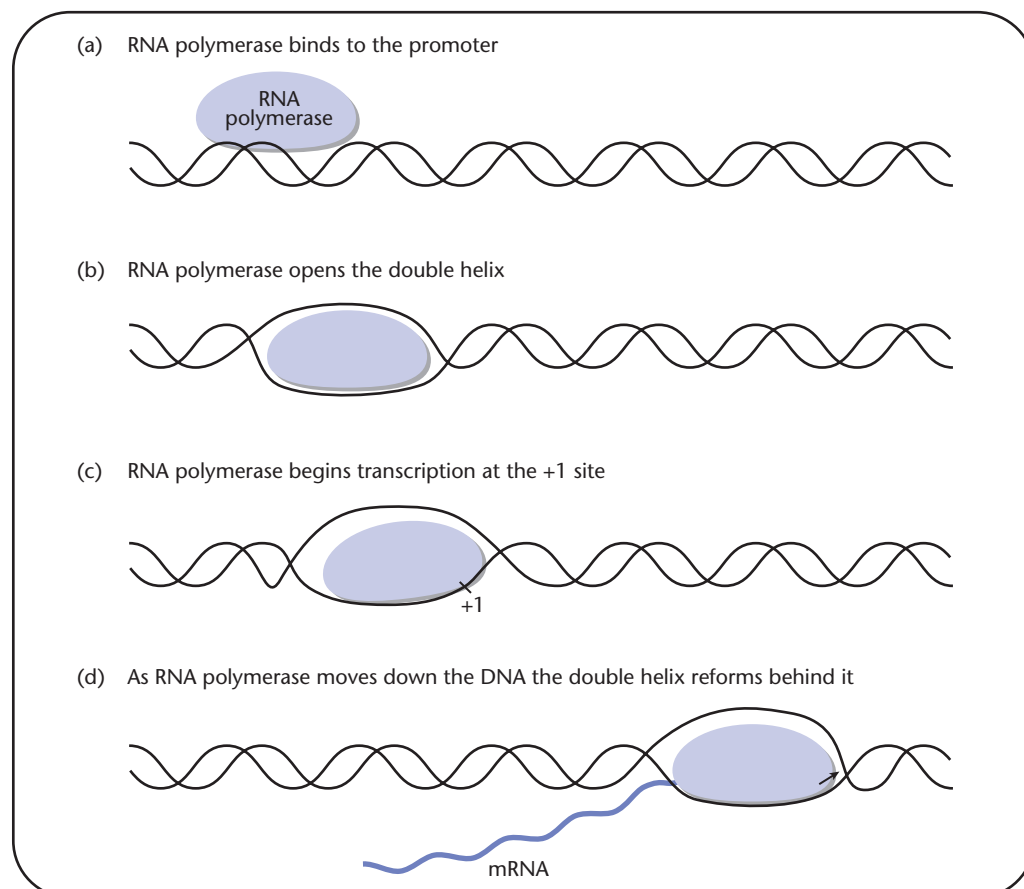

(a) RNA polymerase binds to the promoter

(b) RNA polymerase opens the double helix

(c) RNA polymerase begins transcription at the +1 site

(d) As RNA polymerase moves down the DNA the double helix reforms behind it

mRNA

Fig. 12.4 The beginning steps of transcription. RNA polymerase opens a 17 base pair bubble in the DNA to gain access to the DNA template.

terminator where RNA polymerase and the newly synthesized RNA dissociate from the DNA to end transcription.

Inverted repeat sequences are often found as part of a terminator along with a stretch of uracils found at the end of the RNA transcript (Fig. 12.5). The inverted repeat sequences are able to hydrogen bond with each other and create a stretch of double-stranded RNA. When this secondary structure forms in the RNA, it takes on the appearance of a **stem loop structure**. Terminating transcription can either involve a protein called Rho or be Rho independent, depending on the sequence comprising the terminator. Terminators that do not involve Rho rely on the stem loop structure that forms at the end of the RNA to halt RNA polymerase's forward motion along the DNA. In contrast, Rho-dependent terminators rely on both Rho and the formation of the stem loop structure at the end of the RNA to halt transcription.

Initiation of a new round of transcription does not require that the previous round be terminated. This means that there can be several RNA polymerase complexes transcribing the same template DNA at the same time (Fig. 12.6). Multiple RNA polymerases, because of the size of RNA polymerase, must be spread at least 75 nucleotides apart. In bacteria, translation of an mRNA into its corresponding polypeptide does not require that the entire mRNA be synthesized or transcription be terminated before translation is initiated. Ribosomes are able to bind the **ribosome binding site** (RBS) or **Shine–Dalgarno sequence** in the mRNA and initiate translation at the
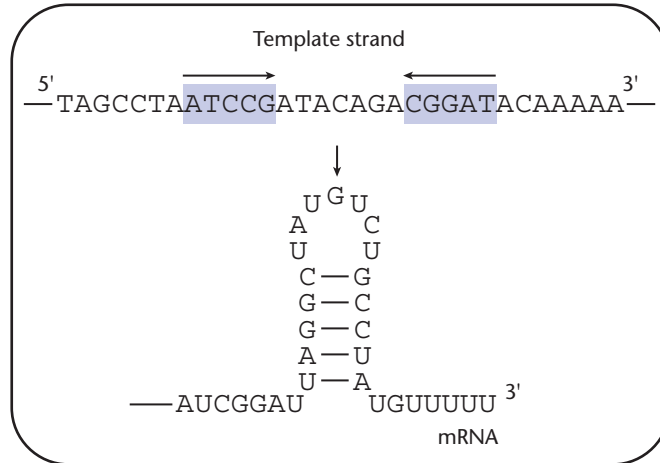

Template strand

```
5'                                                              3'
 —TAGCCTAATCCGATACAGACGGATACAAAAA—
```

```
              U G U
            A       C
            U       U
            C —— G
            G —— C
            G —— C
            A —— U
            U —— A
  ——AUCGGAU        UGUUUUU 3'
```

mRNA

Fig. 12.5 Termination of transcription takes place at specific sites. Frequently, a hairpin structure is part of the termination site. Inverted repeats form the base of the hairpin. The hairpin is frequently followed by a stretch of uracils.
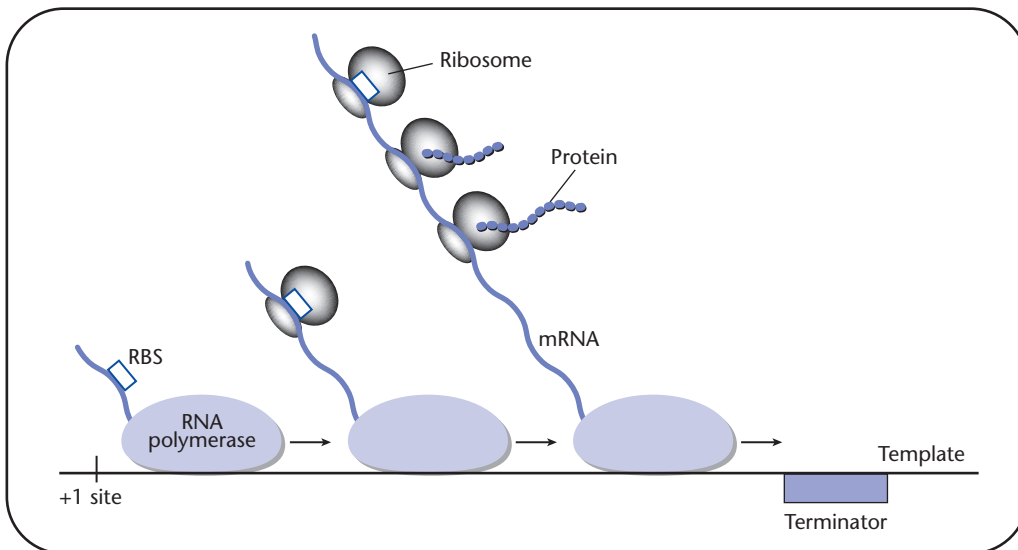

Ribosome

Protein

mRNA

RBS

RNA polymerase

+1 site

Template

Terminator

Fig. 12.6 Multiple rounds of transcription and translation take place at the same time. As RNA polymerase moves along the DNA template, the mRNA becomes accessible to ribosomes. The ribosomes bind to the RBS and initiate translation at the AUG start codon.

starting methionine codon (AUG) even if RNA polymerase is still transcribing the mRNA.

In transcriptional regulation, DNA sequences called **control regions** or **operators** are found adjacent to or overlapping the −35 and −10 regions of the promoter. Specific DNA binding proteins recognize these control regions and exert an effect on holoenzyme's ability to initiate transcription. Regulatory proteins can bind to their control region and prevent transcription. This is known as **repression**. Regulatory proteins can also bind to their control regions and promote RNA polymerase binding to the promoter. This effect is known as **activation**. The regulatory mechanisms controlling gene expression are typically discovered by mutational analysis. Mutations that eliminate the regulation or make the regulation even better usually localize to one of two places (Fig. 12.7). The mutations can be in the genes encoding the repressors or activators. The mutations can also be in the nucleotide sequences that these regulatory proteins bind to.

Operons and regulons

An adaptive response, metabolic pathway, or developmental program usually requires more than one gene. Frequently, the genes for a given process are regulated in a similar manner. Genes that need to be coordinately regulated can be located right next to each other or spread around the chromosome. An **operon** is a group of genes physically linked on the chromosome and under the control of the same promoter(s) (Fig. 12.8a). In an operon, the linked genes give rise to a single mRNA that is trans-
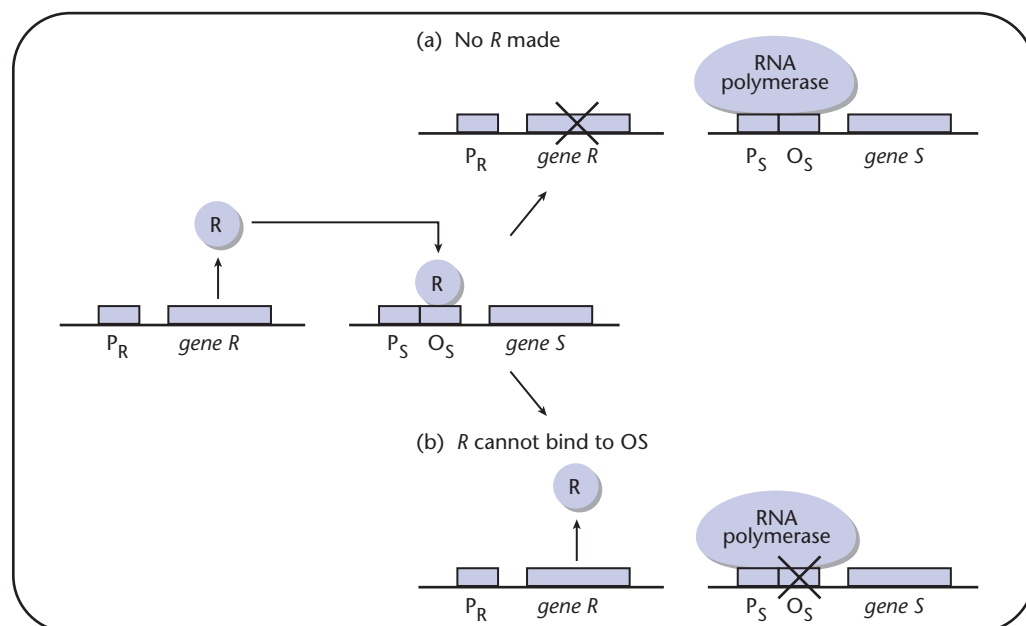

Fig. 12.7 Regulatory mechanisms are frequently studied by mutational analysis. Two classes of mutations are possible. (a) Mutations in the *R gene* or *R gene* promoter result in no R protein being produced. (b) Mutations in the operator (O) DNA sequence can result in the R protein not being able to bind to the DNA and carry out regulation.
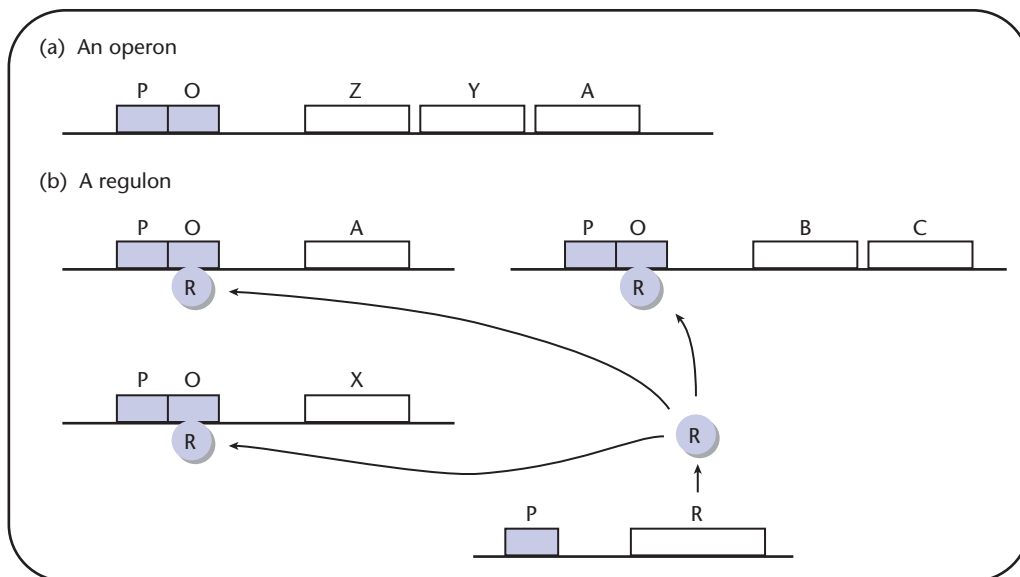
Fig. 12.8 There are two widely used organizational schemes in bacteria for multiple genes that are coordinately regulated. (a) In an operon, all of the genes are physically linked and under the control of one promoter. (b) In a regulon, the genes are in different places on the chromosome but they all have promoters that respond to the same regulators.

lated into the different gene products. This type of mRNA is called a **polycistronic mRNA** (Fig. 12.9). A **regulon** is a group of genes all needed for the same process but physically located in different parts of the chromosome and containing their own promoter(s) (Fig. 12.8b). In a regulon, the promoters are all regulated in the same fashion and allow for coordinate expression of the necessary genes.
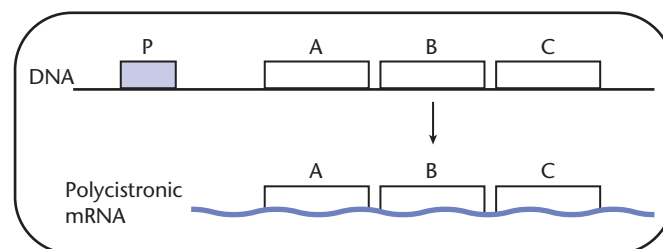

Fig. 12.9 Polycistronic mRNAs are translated into more than one protein. In this example, *gene A*, *gene B*, and *gene C* are encoded on one mRNA and will all be translated from this one mRNA.

Repression of the *lac* operon

The metabolism of the sugar, lactose, requires at least two proteins in *E. coli*. The lactose permease, encoded by the *lacY* gene, transports lactose into the cell. The enzyme β-galactosidase, called β-gal for short and encoded by the *lacZ* gene, catalyzes the cleavage of the β1,4-glycosidic linkage in lactose resulting in the monosaccharides, galactose and glucose (Fig. 12.10a). When a compound is broken down, the process is called **catabolic**. The *lac* operon expresses genes whose products catabolize lactose. Analogs are available that mimic lactose's ability to either be a substrate for β-gal (X-gal, ONPG) or converted to an **inducer** of the *lac* operon (IPTG) (Fig. 12.10b). This has made the study of the *lac* operon and its regulation amenable to examination in the laboratory. These analogs are frequently used in DNA cloning experiments as described in Chapters 13 and 14.

   lacZ, *lacY*, and one other gene, *lacA*, are linked on the *E. coli* chromosome (Fig. 12.11). *lacA* encodes a detoxifying enzyme called β-galactoside transacetylase. The intracellular concentrations of β-gal and Lac permease are very low in *E. coli* cells grown
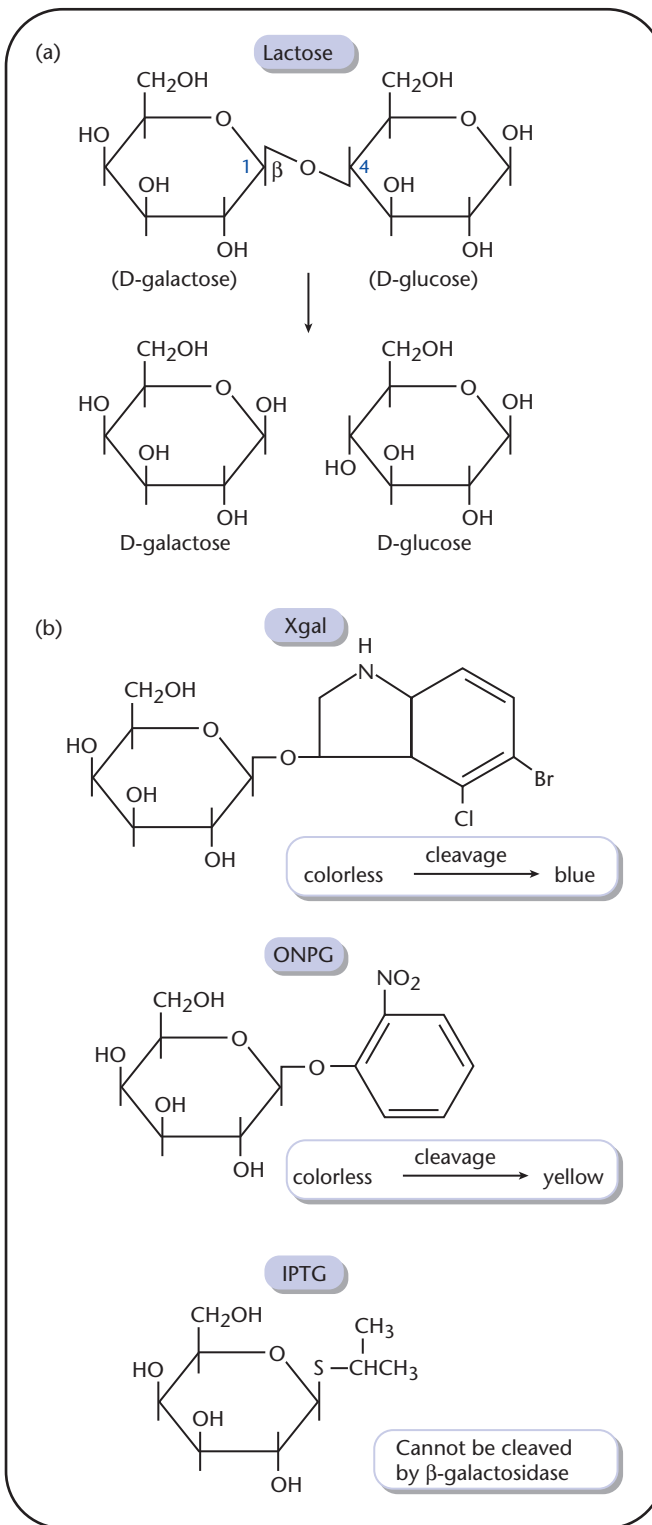
Fig. 12.10 (a) Cleavage of lactose by β-galactosidase. (b) Lactose analogs that are used to study the *lac* operon. X-gal is converted from a colorless compound to a dark blue compound when it is cleaved by β-galactosidase. ONPG is converted from a colorless compound to a yellow compound. IPTG cannot be cleaved by β-galactosidase but it can bind to LacI and induce the *lac* operon.

in medium that does not contain lactose (Table 12.1). If lactose is added to the medium, β-gal, Lac permease, and transacetylase appear at the same time in *E. coli* cells. After lactose is transported into the cell, some of it is converted to allolactose (1–6-O-β-galactopyranosyl-D-glucose) by the few molecules of β-gal found in uninduced cells. Allolactose is a rearranged lactose molecule and an inducer of the *lac* operon.
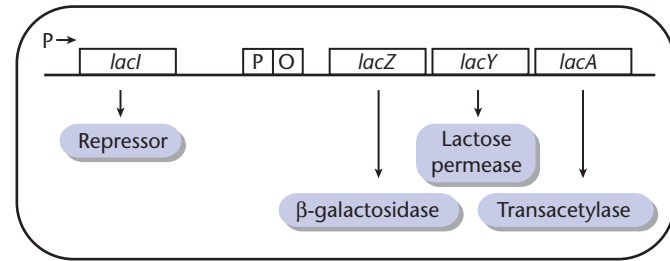
Fig. 12.11 The genes that encode the proteins used to transport and degrade lactose form an operon. The operon is regulated by LacI. The *lacI* gene is located just upstream of *lacZ*, *lacY*, and *lacA*. LacI is expressed from its own promoter that is on all of the time.

lacZ, *lacY*, and *lacA* are coordinately expressed from one promoter called *lacP* that directs expression of a polycistronic mRNA (Fig. 12.11). *lacP* is located upstream of *lacZ*. *lacI* encodes the lactose repressor protein and is located upstream of *lacP*. *lacI* has its own promoter and is constitutively expressed. Lac Repressor or LacI binds to the DNA at the *lac* operator site called *lacO*. *lacO* is located between *lacI* and *lacZ*. In the absence of inducer, Lac repressor is bound to *lacO*, preventing expression of the operon (Fig. 12.12a). In the presence of inducer, inducer binds to Lac repressor and prevents repressor from binding to *lacO*, leading to transcription of the operon (Fig. 12.12b).

How *lac* operon induction takes place and the players that control this induction were elucidated in the extensive genetic analyses undertaken by Francois Jacob, Jacques Monod, and their coworkers. Table 12.1 describes a few of the *lac* mutants isolated by Jacob and Monod and the effect the inducer has on the activity of β-gal and transacetylase. For example, Class I mutations eliminate the inducibility of the *lac* operon resulting in β-gal and transacetylase activity being detectable in the absence of lactose (uninduced). This class of mutation results in a constitutive Lac$^+$ phenotype and maps to either the *lacI* gene or to *lacO*. *lacI* $^-$mutations are recessive or have no phenotype in the presence of a wild-type *lac* operon. Enough LacI is made from the wild-type *lacI* gene to repress both copies of the *lac* operon. Class II mutations result in a super-repressed Lac$^-$ phenotype. Super-repressed mutants have no β-gal or transacetylase activity in the presence of inducer. These mutations map to the *lacI*

Table 12.1 Some classes of *E. coli lac* mutant isolated by Jacob and Monod and the effect inducer has on the relative concentrations of ß-galactosidase and transacetylase in the mutant backgrounds. The mutations were examined when they resided in the chromosome in one copy and in merodiploids containing a second copy of the *lac* operon on an F′.

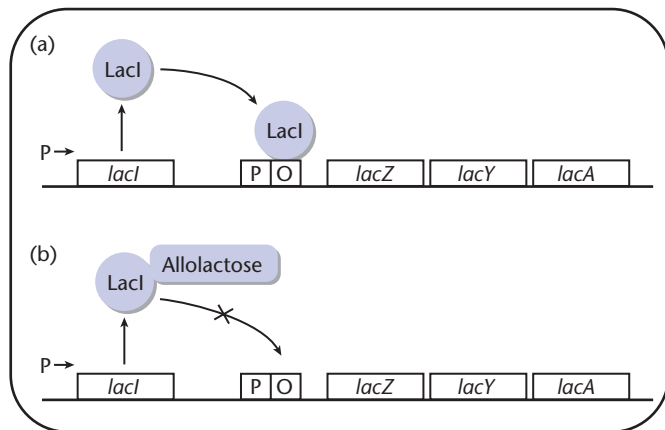| Class	Genotype	ß-galactosidase (*lacZ*)		Transacetylase (*lacA*)	
		Uninduced	Induced	Uninduced	Induced
WT	lacZ A I	0.1	100	1	100
I	lacZ A I⁻	100	100	90	90
	lacZ A I⁻/F′ lacZ A I	1	240	1	270
II	lacZ A Iˢ	0.1	1	1	1
	lacZ A Iˢ/F′ lacZ A I	0.1	2	1	3
III	lacZ A Oᶜ	25	95	15	100
	lacZ⁻ A O /F′ lacZ A⁻ Oᶜ	180	440	1	220
II + III	lacZ A O Iˢ	190	219	150	200
	F′ lacZ A Oᶜ I				

(a)

(b)

Fig. 12.12 A model for the regulation of the *lac* operon. The Lac repressor or LacI binds to the *lac* operator DNA (*lacO*) and prevents transcription of *lacZ*, *lacY*, and *lacA*. If lactose is present inside the cell, it is converted to allolactose. Allolactose binds to LacI and prevents its interaction with *lacO*. This allows expression of the *lacZ*, *lacY*, and *lacA* genes.

gene and result in a repressor that can no longer be inactivated by inducer. Super-repressor mutants are dominant to wild type because the mutant LacI can bind to both copies of the *lac* operon and turn them off. Class III mutations are in *lacO* and prevent Lac repressor from binding and turning off the operon. *lacO^c* mutants are dominant to wild type because wild-type LacI from either copy of the operon cannot repress the *lac* operon containing the *lacO^c* mutant. When super-repressor mutants are combined with *lacO^c* mutants, *lacO^c* mutants are dominant to super-repressor mutants because the super-repressor still cannot bind the mutant *lacO*.

Activation of the *lac* operon by cyclic AMP and the CAP protein

If lactose and glucose are added to a culture of wild-type *E. coli* cells, the *lac* operon is not induced. No *lac* mRNA or gene products are made. This effect of glucose is the result of a second regulatory mechanism, **catabolite repression**. Catabolite repression affects not only *lac* gene expression but also other operons that catabolize specific sugars such as galactose, arabinose, and maltose. Contrary to its name, catabolite repression describes an activating mechanism, involving a complex between cAMP (cyclic adenosine monophosphate) and the catabolite activator protein (CAP; also called cyclic AMP receptor protein or CRP) (Fig. 12.13).

If given the choice of sugars to metabolize, *E. coli* will use glucose first and then other sugars. When levels of glucose are high, there is no need to express high levels of the enzymes needed for the metabolism of lactose or other sugars, which yield less energy than glucose. High levels of intracellular glucose result in low levels of cAMP. Low levels of cAMP mean very few cAMP–CAP complexes. When intracellular levels of glucose drop and other sugars must be metabolized, levels of cAMP increase. Increased cAMP levels mean there are more cAMP–CAP complexes. cAMP–CAP binds to a specific site on the DNA that is located adjacent to the promoter for the *lac* genes (Fig. 12.14) and adjacent to promoters controlling the expression of other sugar metabolizing operons affected by cAMP–CAP. The –10 and –35 motifs in the promoter sequence for *lac* are not a perfect match to the consensus sequence for σ^{70} promoters. cAMP–CAP bound to the CAP binding site increases the binding of σ^{70} RNA polymerase to the *lac* promoter. By having two mechanisms controlling the expression of the *lac* operon, cells ensure that the *lac* gene products are only made when needed.

Regulation of the tryptophan biosynthesis operon by attenuation

The biosynthesis of the aromatic amino acid tryptophan requires the activity of three enzymes, anthranilate synthetase (*trpE*) complexed with phosphoribosyl-anthranilate transferase (*trpD*), phosphoribosyl-anthranilate isomerase (*trpC*), and tryptophan synthetase (*trpA* and *trpB*) to convert chorismate to tryptophan. The five *trp* structural genes encoding the three tryptophan biosynthetic enzymes are physically linked and coordinately regulated as an operon. The *trp* genes are arranged in the
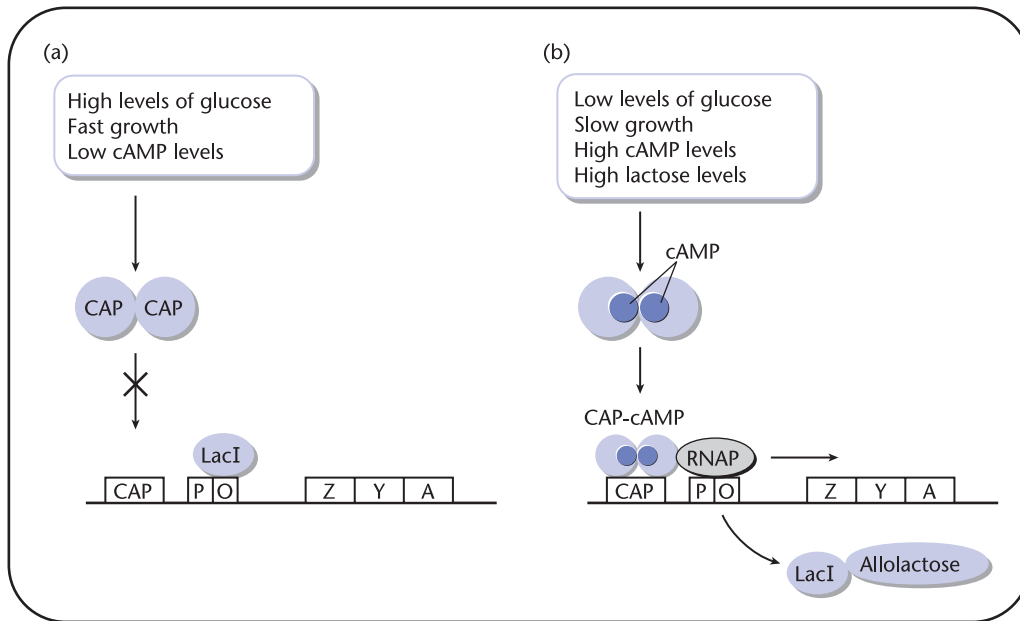
Fig. 12.13 Catabolite repression of the lactose operon. (a) High levels of glucose mean low levels of cAMP. There is not enough cAMP to complex with CAP and no induction of the *lac* operon takes place. LacI remains bound to *lacO*. (b) Low levels of glucose lead to high levels of cAMP. cAMP complexes with CAP and binds in front of *lacP*. If there is lactose present, it is converted to allolactose, allolactose complexes with LacI and prevents LacI from binding to lacO. This leads to induction of the *lac* operon.
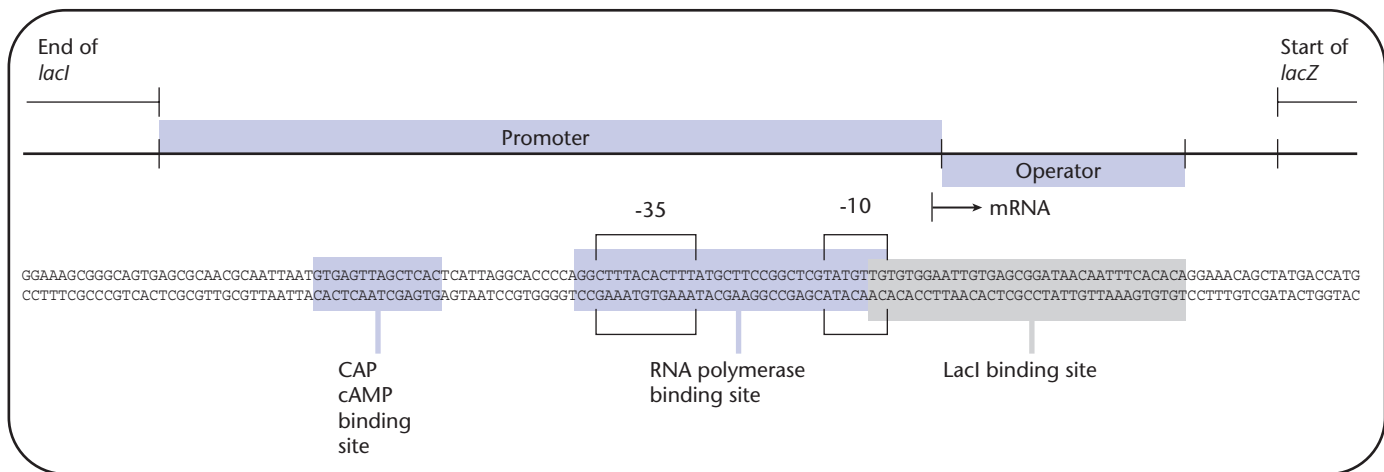
Fig. 12.14 The binding sites for CAP–cAMP, RNA polymerase, and LacI in the *lac* operon of *E. coli*.

order that their encoded enzymes function in the biosynthetic pathway (Fig. 12.15). The goal of a biosynthetic or **anabolic** pathway is to synthesize the end product only when it is needed. In 1959, Cohen and Jacob made the observation that addition of tryptophan to an actively dividing *E. coli* culture repressed the synthesis of trypto-phan synthetase. The *trp* operon is repressed by TrpR, a transcriptional repressor,

Fig. 12.15 The structure of the *trp* operon and the enzymatic activities encoded by the *trp* genes. The regulator of the *trp* operon, TrpR, is encoded by a gene that is not linked to the rest of the operon. The structural genes, *trp EDCBA*, are positioned in the operon in the same order as the enzymes they encode are used in tryptophan synthesis.
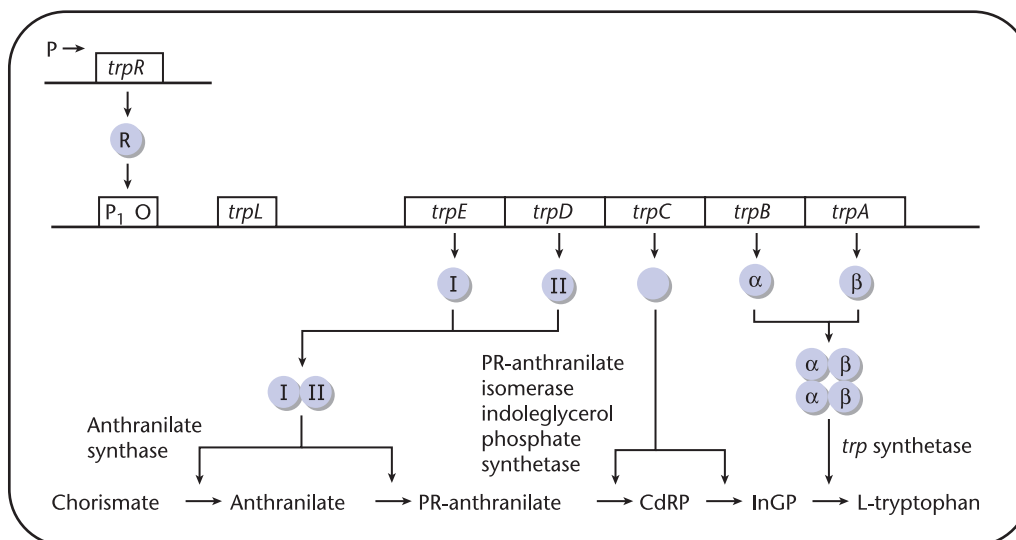
when levels of tryptophan are high in the cell (Fig. 12.15). The gene encoding the TrpR repressor is not located near the *trp* operon.

Even though expression of both the *lac* and *trp* operons are regulated by repression, the mechanism by which the LacI and TrpR repressors work is very different. The LacI repressor when bound to its effector molecule (i.e. lactose) cannot repress transcription. In contrast, the TrpR repressor must be bound by an effector molecule, in this case tryptophan, for it to repress transcription. The repression mechanism for the *trp* operon can be viewed as a coarse control mechanism; an all or none approach to gene expression. However, another regulatory mechanism, **attenuation**, is available to fine-tune gene expression of some biosynthetic operons, especially during times of extremely short supply of end products.

Attenuation was first proposed by Charles Yanofsky and coworkers who observed that under normal growth conditions approximately nine out of 10 *trp* mRNA transcripts terminate before they reach the structural genes of the *trp* operon. This means that even when repression is lifted and transcription is initiated from the *trp* promoter, RNA polymerase abruptly halts, failing to elongate the *trp* mRNA across the structural genes. During conditions of severe tryptophan starvation, the premature termination of transcription is abolished, allowing expression of the *trp* operon at 10-fold higher levels than under normal growth conditions.

What is happening to cause early termination of the *trp* mRNA? The answer lies in the regulatory sequence in front of the *trp* structural genes. Downstream from the *trp* promoter and the TrpR repressor binding site but upstream of the *trp* structural genes, are 162 base pairs. The 162 base pair sequence is called the **leader sequence** or *trpL* (Fig. 12.16a). *trpL* contains a ribosome binding site adjacent to an AUG translational start codon. This means there are two RBS sites and two AUG translational start codons located after the *trp* promoter but in front of the first *trp* structural gene (*trpE*).

The *trp* leader sequence, after it is transcribed into mRNA, can be translated into a short polypeptide consisting of 14 amino acids. Of these 14 amino acids, two are *trp* residues that are located next to each other. The presence of two tandem *trp* residues is very rare. Most *E. coli* proteins contain one *trp* per every 100 amino acids. At the end of *trpL* is a Rho-independent terminator. This signal resides approximately 40 base pairs
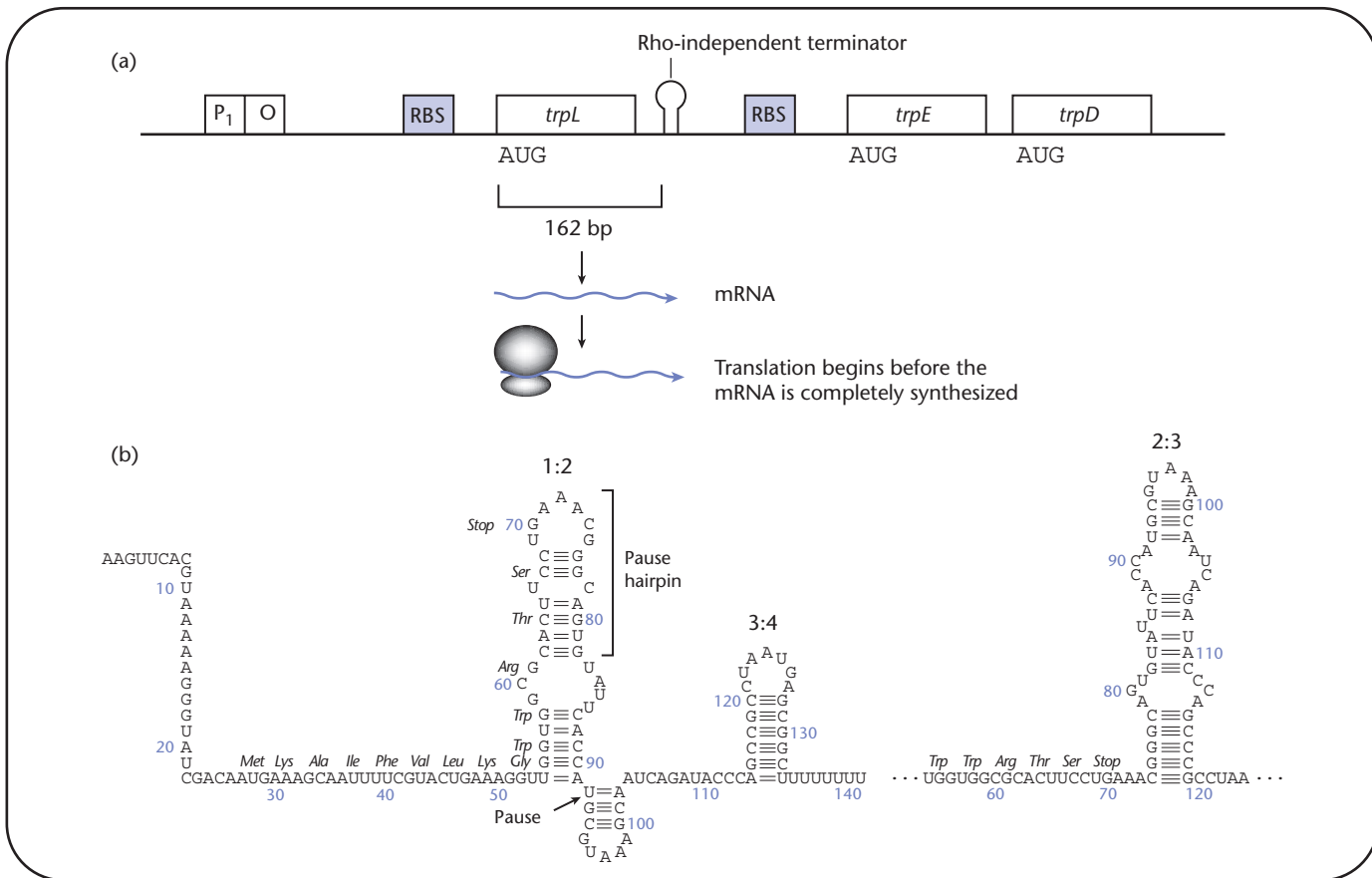
Fig. 12.16 The features of the leader region of the *trp* operon. (a) *trpL* is transcribed into mRNA. (b) The *trpL* mRNA can form three different stem loop structures depending on the prevalence of charged trp–tRNAs inside the cell (see text and Fig. 12.17 for details).

upstream of the signals that control the translation of the *trp* structural genes. The DNA sequence in this region suggests that the encoded mRNA has a very high probability of forming stem loop secondary structures. There are three combinations of stem loops that can form. They have been named the pause loop (1:2), the terminator loop (3:4), and the anti-terminator loop (2:3) (Fig. 12.16b).

Attenuation of the *trp* operon uses the secondary structures in the mRNA to determine the amount of tryptophan in the cell (Fig. 12.17). RNA polymerase initiates transcription at the *trp* promoter after the TrpR repressor has been removed. Because translation can be initiated before the mRNA is completed, a ribosome loads at the *trpL* RBS and initiates translation of TrpL (Fig. 12.17 step 1). As RNA polymerase elongates the *trpL* mRNA, stem loop structures will form in the mRNA. The first secondary structure that can form is the 1:2 pause loop. This large hairpin structure is called the pause loop because it forms behind the transcribing RNA polymerase and causes RNA polymerase's forward movement along the DNA template to pause (Fig. 12.17 step 2). Pausing transcription allows the translating ribosome to stay close to the transcribing RNA polymerase. This ensures that these two complexes stay in close communication during crucial times of abruptly changing levels of tryptophan. As the translating ribosome advances on the mRNA, the ribosome breaks apart the pause loop freeing RNA polymerase to continue transcribing *trpL* (Fig. 12.17 step 3). At this time, the ribosome
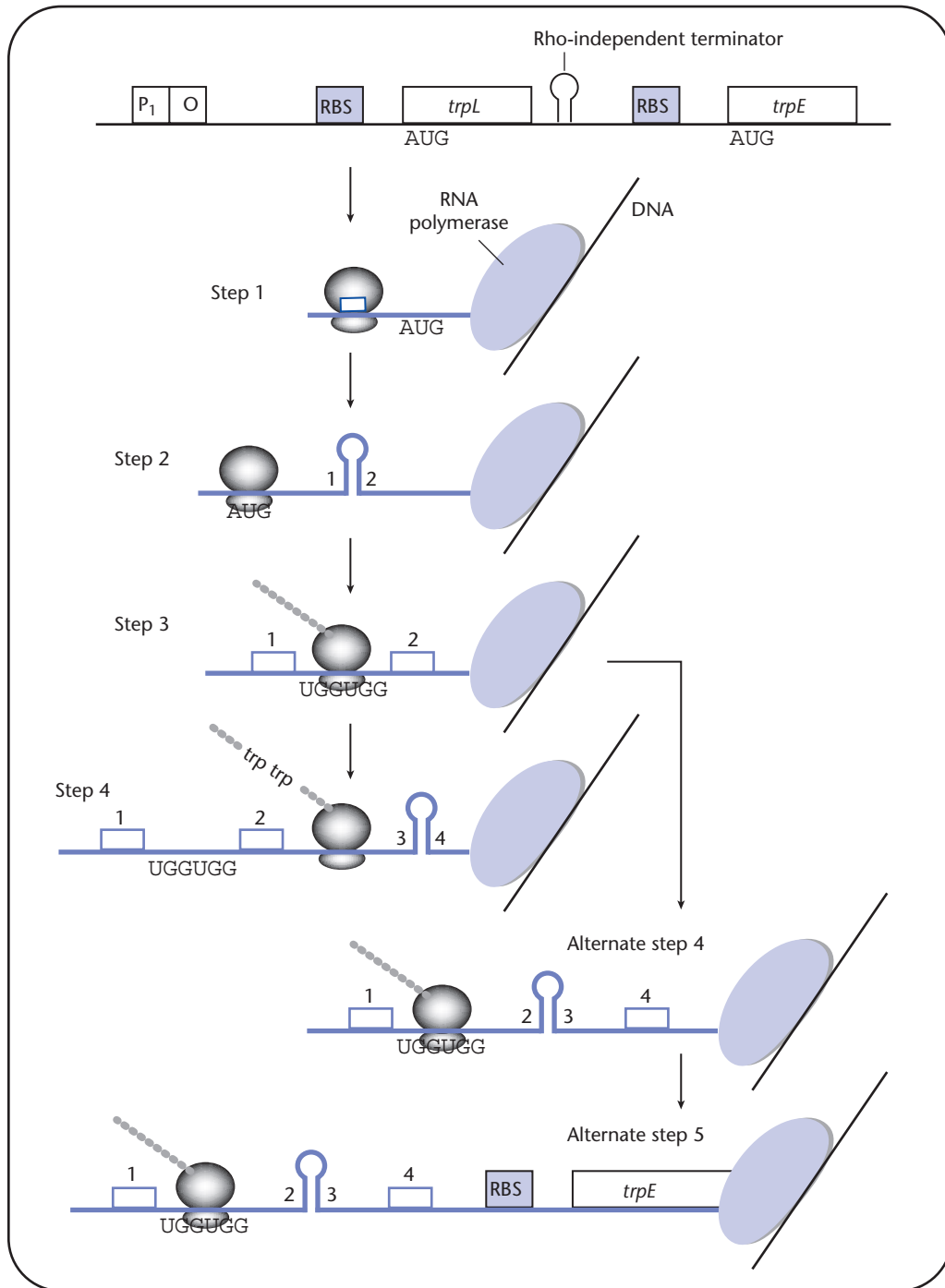
Fig. 12.17 The predictions of the attenuation model. (Step 1) Coupled transcription/translation takes place as for any bacterial gene. (Step 2) RNA polymerase pauses and 1 : 2 stem loop forms. (Step 3) The ribosome disrupts the 1 : 2 stem loop and encounters the two *trp* codons. (Step 4) If enough tryptophan is present, charged trp–tRNAs will be present and the ribosomes will translate *trpL*. This causes RNA polymerase to stop at the terminator composed of a 3 : 4 stem loop. (Alternate Step 4) If not enough *trp* is found in the cell, the ribosome stalls at the two *trp* codons, while RNA polymerase continues. The 2 : 3 stem loop forms. (Alternate Step 5) The 3 : 4 terminator cannot form and RNA polymerase continues transcribing into the *trp* structural genes. This exposes the RBS upstream of *trpE*, allowing translation of *trpE*.

encounters the two *trp* codons. If there are two charged tryptophanyl–tRNAs to pass tryptophan to the ribosome to add to the growing TrpL polypeptide, then the ribosome does not stall, and it is able to stay close to the transcribing RNA polymerase (Fig. 12.17 step 4). If they stay close, very few bases of the mRNA transcript will separate the ribosome from the RNA polymerase and only the 3:4 terminator loop structure is predicted to form. The 3:4 loop is a Rho-independent transcription terminator.

If two charged tryptophanyl–tRNAs are not available to pass tryptophan to the translating ribosome, then the ribosome stalls in its attempt to synthesize the TrpL polypeptide (Fig. 12.17 alternate step 4). The transcribing RNA polymerase continues to move away from the stalled ribosome, creating a longer stretch of bases that separate these two complexes. Because of the longer distance between RNA polymerase and the ribosome, the larger 2:3 anti-terminator stem loop structure forms. The 2:3 loop is not a transcription termination signal and RNA polymerase continues its forward motion, transcribing the *trp* structural genes (Fig. 12.17 alternate step 5). Thus, the leader peptide plays a decisive role in determining if transcription of the *trp* operon will terminate early or will continue into the structural genes. Other *E. coli* biosynthetic operons, including those that synthesize *phe*, *his*, *leu*, *thr*, and *ilv*, are controlled by attenuation. Of this group of operons, only *trp* has a repressor in addition to attenuation.

Regulation of the heat-shock regulon by an alternate sigma factor, mRNA stability, and proteolysis

The coordinately regulated genes of an operon are physically linked and controlled by the same promoter. In contrast, the coordinately regulated genes of a regulon are physically located in different parts of the chromosome and are controlled by their own promoters yet are regulated by the same mechanisms. Regulons are also called multigene systems or global regulatory systems. Table 12.2 describes a few of the regulons identified in bacteria. Three bacterial regulons, heat shock, SOS and eps, will be used to illustrate the regulatory mechanisms and components associated with regulons.

Cells respond to an abrupt increase in temperature by inducing synthesis of a specific group of proteins to cope with this stress. This heat-shock response has been documented in many different cell types and extensively studied at the molecular level in prokaryotic cells such as *E. coli*. *E. coli* is a mesophile, which means its normal growth temperature is between 20°C and 37°C. Shifts in temperature within this range result in very little stress for the bacterium and there is no noticeable adjustment in its gene expression. *E. coli* responds to an abrupt increase in temperature from 30°C to 42°C by producing a set of ~30 different proteins (Table 12.3). These proteins are collectively called the heat-shock proteins (Hsps). Within 5 minutes after a 30°C to 42°C temperature shift, the Hsp proteins increase by 10-fold. After this transient increase, the amount of the Hsp proteins decreases slightly to a steady-state level. The steady-state level is maintained as long as the cells remain at the elevated temperature. Once the temperature is decreased from 42°C to 30°C, the levels of Hsp proteins decrease abruptly, approximately a 10-fold decrease within 5 minutes. Besides a change in temperature, other agents such as ethanol, organic solvents, or certain DNA damaging agents can induce heat-shock gene expression. This suggests that the members of the heat shock regulon deal with cellular damage that can occur in several different ways.

Many of the *hsp* genes encode either proteases or chaperones. Proteases are proteins that degrade proteins especially those that may be abnormal. Abnormal proteins

Table 12.2 Examples of regulons identified in *E. coli*.

Multigene system	Stimulus	Regulator	Regulated genes
Heat shock	Shift to high temperature	σ^{32} (*htpR = rpoH*)	~30 heat-shock proteins (Hsps) including proteases and chaperones
pH shock	Acid tolerance	CadC	Several gene products needed for protection from acid damage
SOS response	Extensive DNA damage	LexA repressor	~20 gene products needed for DNA repair
Oxidation response	H_2O_2	OxyR repressor	~12 gene products needed for protection from H_2O_2 and other oxidants
Porin response	High osmolarity	Sensor EnvZ regulator OmpR	Porin genes
Catabolite repression	Carbon utilization	Crp-cAMP	Many different sugar catabolizing operons

Table 12.3 Heat-shock proteins in *E. coli* and their corresponding functions.

Regulon	Hsp protein	Function
σ^{32}	DnaK	Chaperone
	DnaJ	Chaperone
	GroEL	Chaperone
	GroES	Chaperone
	GrpE	Nucleotide exchange factor
	Lon	Protease
	ClpX	Protease
	ClpP	Protease
	HflB	Protease
	RpoD	Sigma factor σ^{70}
	HtpR	Heat-shock sigma factor σ^{32}
	HtpG	Chaperone
	GapA	Dehydrogenase
	ClpB	Chaperone
	HtrM	Epimerase
	IbpB	Chaperone
	IbpA	Chaperone
	ClpY	Protease
σ^{E}	DegP	Protease
	σ^{E}	Sigma factor
	σ^{32}	Sigma factor
unknown	LysU	Lysyl-tRNA synthetase

σ^{32} regulated Hsps whose function is unknown: HtpY, HslA, HslC, HslK, HtpX, FtsJ, HslO, HslP, HslV, HtrC, HslW, HslX, HslY, HslZ, HtpK, HtpT.
σ^{32} regulated Hsps whose function and regulation are unknown: PspA, HslE, HslF, HslG, HslI, HslJ, HslM, HslQ, HslR.

include incompletely synthesized proteins and proteins that are not folded correctly. Chaperones are proteins that bind to abnormal proteins, unfold them, and try to let them refold into an active configuration. Frequently, if a chaperone cannot fix an abnormal protein, it will target the abnormal protein for degradation by a protease. From looking at the genes that are part of the heat-shock regulon, it is interesting that they mainly deal with misfolded proteins. This suggests that misfolding of proteins is the predominant damage to cells at high temperatures.

The genes encoding Hsps are scattered around the chromosome but are coordinately regulated. Thus, the *hsp* genes constitute a regulon. The regulator of the heat-shock response is an alternative sigma factor, named sigma 32 or σ^{32}. σ^{32} is a 32 kilodalton protein and the gene encoding σ^{32} is *rpoH* (for RNA polymerase subunit heat shock). σ^{32} directed RNA polymerase recognizes the distinctive sequences of heat-shock promoters. The heat shock promoters have different –10 (CCCCAT) and –35 (CTTGAAA) consensus sequences than promoters recognized by σ^{70}. All of the *hsp* genes have a heat-shock promoter. In fact, once the *E. coli* genome was sequenced, heat-shock promoter sequences were used to identify novel potential Hsps.

σ^{32} is made at all temperatures yet the *hsp* gene products are not. σ^{32} is an unstable protein and at low temperatures it has a half-life of about one minute. At high temperatures, 10 times more σ^{32} protein is made and the σ^{32} that is made is five times more stable than the σ^{32} made at low temperatures. The regulation of the *rpoH* gene is at the translational level. There is a significant amount of *rpoH* mRNA in cells at low temperatures yet translation of this mRNA is inhibited. At high temperatures, translational inhibition is relieved and σ^{32} protein is made. Two regions of the *rpoH* mRNA are required for translational inhibition, a region near the +1 site and a region internal to the *rpoH* gene between nucleotides 150 and 250 in the mRNA. How these two regions of the mRNA inhibit translation at low temperatures is not well understood. The two regions are capable of forming a stem loop structure and this may prevent the ribosome from loading at the RBS and inhibit translation in addition to possibly increasing the stability of this mRNA.

σ^{32} is degraded by a specific protease called HflB. Degradation of σ^{32} at 30°C also requires a chaperone composed of three proteins, DnaK, DnaJ, and GrpE. Degradation of σ^{32} at 30°C is decreased ~ 10-fold by mutations in any one of the genes that encode HflB, DnaK, DnaJ, or GrpE. All four of these proteins are part of the heat-shock regulon and so their expression is induced by heat shock. If more chaperone proteins are present at higher temperatures to help degrade σ^{32}, then how does σ^{32} ever interact with core RNA polymerase to activate the heat-shock genes? While the precise answer to this question is unknown, recent experiments suggest that the interactions between DnaK and σ^{32} are temperature dependent and only occur at low temperatures. When the temperature is raised, DnaK and σ^{32} no longer interact yet interaction between σ^{32} and core RNA polymerase remains unchanged. It is possible that this temperature-dependent interaction of DnaK and σ^{32} plays a role in σ^{32} stability at high temperatures. If the temperature is lowered from 42°C to 30°C, translational inhibition of the mRNA returns and σ^{32} again becomes sensitive to degradation. The heat-shock response can be turned off as quickly as it was turned on.

A second heat-induced regulon is controlled by a second alternative sigma factor, σ^{E}. At temperatures around 50°C, σ^{E} controlled promoters are very active. Deletions of the gene encoding σ^{E} are temperature sensitive at 42°C, whereas deletions of the gene encoding σ^{32} are temperature sensitive at 20°C. σ^{E} responds to misfolded outer membrane proteins and σ^{32} responds to misfolded cytoplasmic proteins. The σ^{32} gene also has a σ^{E} promoter so that all of the Hsps are induced by a cascade effect when σ^{E} regu-

Weigle reactivation

If a lysate of lambda is UV irradiated, the DNA in many of the phage is damaged and these phage are not capable of growth. If this irradiated lysate is used to infect a population of *E. coli* that have also been irradiated, many of the damaged phage now survive. This basic observation is known as the Weigle effect or Weigle reactivation. At first it was not clear why two doses of irradiation (one to the phage and one to the bacterium) should fix the damage caused by one dose of irradiation. We now know that irradiating the bacteria induces the SOS response and induces the synthesis of DNA repair enzymes. When the damaged λ DNA is introduced into the SOS induced bacteria, the λ DNA is repaired and is used to produce more phage.

lated genes are expressed. It is thought that the σ^E regulon provides proteins that are needed at more extreme conditions. Given what we know, it is possible to induce the σ^{32} regulon by itself or the σ^{32} and σ^E regulons together. Having several overlapping regulons that respond to different degrees of one condition gives the cell more flexibility in how it responds. Cells can make large amounts of very specific gene products without wasting resources by making gene products that they don't need.

Regulation of the SOS regulon by proteolytic cleavage of the repressor

The SOS regulon permits a rapid response to severe or lethal levels of DNA damage. The goal is to quickly activate DNA repair proteins and deal with the DNA damage in a swift manner. Speed is of the essence to save the cell by restoring its chromosome to an intact double-stranded state. The DNA repair enzymes that are induced result in a higher level of mistakes than normally seen but this is balanced by their speed in re-pairing the majority of the chromosome. During the SOS response, cell division is halted so that damaged chromosomes are not segregated into daughter cells. Thus, in addition to DNA repair enzymes, a cell division inhibitor protein is also expressed at high levels during the SOS response. Table 12.4 describes the genes and proteins of the SOS regulon. The SOS genes are expressed at low levels during normal cell growth. They are induced to high-level expression upon extensive DNA damage.

Weigle and coworkers first reported from their observations of reactivated UV-irradiated λ that some DNA repair systems are inducible (see FYI12.2). Many of these repair mechanisms are activated as part of the SOS regulon. Like the heat-shock regu-lon, the SOS regulon has a mechanism for signaling when the regulon should be on or off. The SOS genes contain a common sequence that in some cases overlaps the pro-moter region and in other cases is adjacent to the promoter region. This common se-quence is bound by a repressor protein called LexA. SOS promoters bound by LexA cannot be used to initiate transcription. As a result of significant DNA damage, the LexA repressor is inactivated and removed so that the SOS genes are expressed.

How is the LexA repressor inactivated and removed from the promoters of SOS genes? Upon exposure to DNA damaging agents, large amounts of single-stranded DNA accumulate (Fig. 12.18a). Single-stranded DNA is bound by the RecA protein (Fig. 12.18b). RecA participates in homologous recombination (see Chapter 5) and in

Table 12.4 Known SOS genes, proteins, and corresponding functions in *E. coli*.

| Gene	Protein	Function
lexA	LexA	Represses transcription of SOS genes
recA	RecA	LexA activator and major homologous recombination protein
recN	RecN	Binds ATP; RecBCD substitute
ruvAB	Ruv	Branch migration and resolution
sulA (*sfiA*)	SulA	Cell division inhibitor
umuC	UmuC	Mutagenesis
umuD	UmuD	Mutagenesis
uvrA	UvrA	Excision repair
uvrB	UvrB	Excision repair
uvrD	UvrD	Excision repair
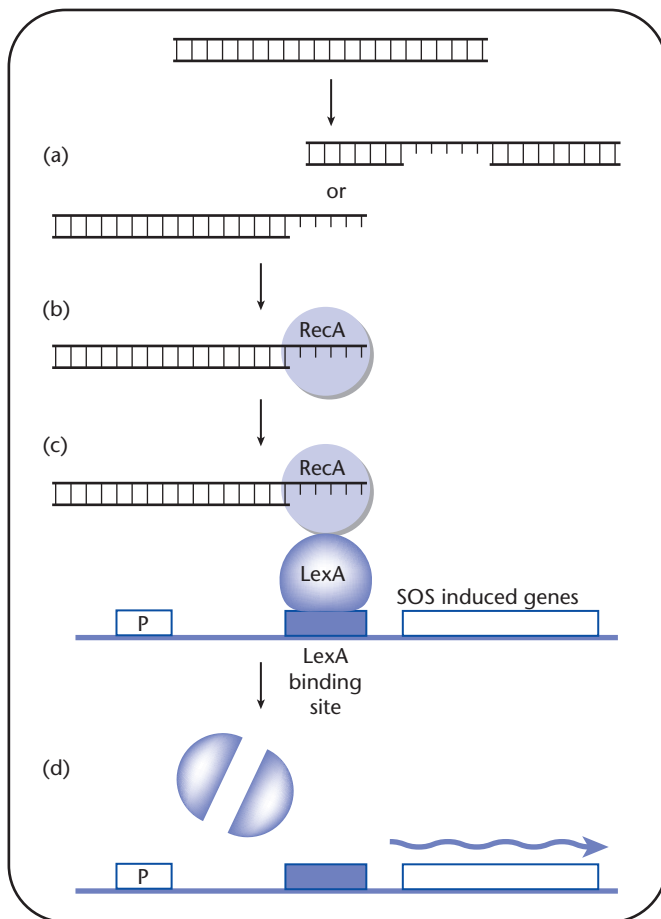
Fig. 12.18 The SOS response. (a) DNA damage leads to single-stranded DNA. (b) RecA binds to the ssDNA. (c) RecA–ssDNA binds to LexA. (d) Lex A undergoes autocleavage, relieving repression of the SOS genes.

post replication DNA repair (see Chapter 4). In the case of extensive DNA damage, RecA complexed to single-stranded DNA binds to LexA and induces LexA to cleave itself (Fig. 12.18c). Since LexA, with help from the RecA–ssDNA complex, cleaves itself, LexA is said to autocleave. Autocleavage of LexA takes place between two specific amino acids that separate the repressor into two domains, the DNA binding domain and the dimerization domain. Two LexA molecules bind to the sequence found in SOS regulon promoters. The disruption of dimerization results in the removal of LexA from the SOS promoters (Fig. 12.18d). Once LexA is removed, SOS genes are expressed at high levels. Eventually the induction signal, RecA complexed to single-stranded DNA, drops because of DNA repair, and LexA is no longer able to undergo autocleavage. This returns the regulon to its prestimulus or uninduced state.

Two-component regulatory systems: signal transduction and the *cps* regulon

Bacteria continually monitor their environment so as to set in motion a proper response to detected changes. There are molecular mechanisms that pick up signals from the environment and transmit this information to the appropriate regulators. This is the concept of **signal transduction**. The molecular mechanism associated with this process in bacteria is the two-component system. The **two-component**

Table 12.5 Examples of two component systems in *E. coli*.

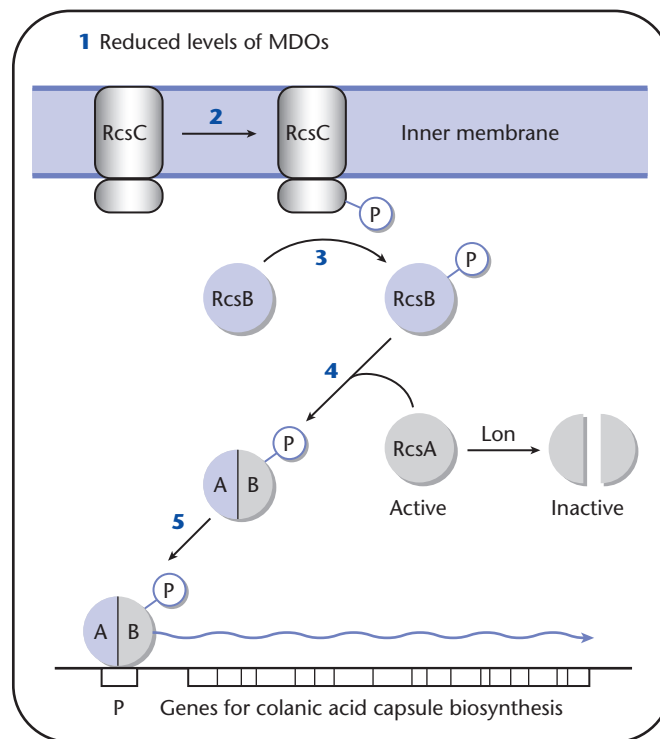| System	Two-component response-regulator/sensor-kinase
Nitrogen limitation	NtrC regulator/NtrB sensor
Phophorus limitation	PhoR regulator/PhoB sensor
Oxygen limitation	ArcA regulator/ArcB sensor
Osmotic upshift	KdpD regulator/KdpE sensor
Capsule expression	RcsB regulator/RcsC sensor
Porin regulation	OmpR regulator/EnvZ sensor

signal transduction system consists of a **sensor-kinase protein** and a **response-regulator protein** (see Fig. 11.4). This system has been identified as a regulatory mechanism for many operons and regulons, including induction of competency in *B. subtilis* (see Chapter 11). The basic components of two-component signal transduction systems are described in Chapter 11. Two-component systems are found in many different bacteria. In *E. coli* there are at least 50 different examples of this type of regulation. Table 12.5 describes a few of these systems. We will use the *E. coli cps* system, which controls synthesis of an exopolysaccharide capsule, to illustrate two-component signal transduction.

 E. coli produces a number of different types of exopolysaccharides. Exopolysaccharides are polymers consisting of sugars linked together, often in a lattice-type structure. They are located on the outside of the bacterial cell and while some remain attached to the bacterial cell, others do not. They often protect bacteria from harsh environments, toxic chemicals, or bacteriophage. Some exopolysaccharides serve as matrices to link a number of different bacteria into a community or biofilm. Other bacterial exopolysaccharides cause health problems in humans.

 An extensively studied *E. coli* exopolysaccharide is the colanic acid capsule. The genes encoding the enzymes that make this polymer are named *cps* (colanic acid polysaccharide). The majority of the *cps* genes are arranged in a single operon. The regulation of the *cps* genes involves several different regulatory mechanisms (Fig. 12.19). One such mechanism is a two-component system involving the sensor-kinase protein, RcsC (regulator of capsule synthesis), and the response-regulator protein, RcsB. This two-component system can be activated by

Fig. 12.19 Regulation of the *cps* genes. (Step 1) Reduced levels of MDOs (membrane-derived oligosaccharides) result in autophosphorylation of RcsC (Step 2). (Step 3) The phosphate group is transferred to RcsB. (Step 4) RcsB binds to RcsA. (Step 5) This complex acts on the *cps* promoter and activates transcription of the *cps* genes. RcsA is also an unstable protein that is degraded by Lon protease. When RcsA is in a complex with RcsB, it is protected from degradation. Expression of *cps* genes is controlled by several different mechanisms to ensure that the capsule in only made when needed.

osmotic shock or by cell desiccation. The signal that RcsC responds to during periods of high osmolarity is the level of membrane-derived oligosaccharides (MDOs). Levels of MDOs are low during periods of high osmolarity, and the reduced levels of MDOs are thought to act as a signal to trigger RcsC autophosphorylation. Once RcsC is phosphorylated it transfers the phosphate to the response-regulator, RcsB. Phosphorylated RcsB interacts with a second positive regulatory protein, RcsA, to activate the expression of the *cps* genes and lead to biosynthesis of the colanic acid capsule.

Interestingly, the heat-shock protease, Lon, also plays a role in regulating the expression of the *cps* genes. RcsA is an unstable protein and is degraded by Lon. This instability keeps the levels of RcsA in the cell very low and the capsule genes turned off. Removing Lon activity increases levels of RcsA and together with phosphorylated RcsB, this complex turns on *cps* genes expression. It is estimated that there are approximately 40 *cps* genes or about 1% of the genes in *E. coli*. Thus, making the colanic acid capsule requires a significant amount of energy and cellular resources. The cell uses several mechanisms to ensure that the capsule is only made when needed.

Summary

Many mechanisms regulate gene activity. The goal is to express genes only when needed. The catabolic machinery needed to metabolize lactose only needs be available when lactose is present. Likewise, tryptophan only needs to be made when the cell runs out of this amino acid. Transcriptional regulation is frequently observed because transcription is the first step in expressing a gene. Transcriptional regulation for the *lac* and *trp* operons involves a repressor, LacI or TrpR, respectively. LacI represses transcription when not bound by the inducer lactose. In contrast, TrpR represses transcription only when bound by tryptophan. Both of these operons have at least one additional regulatory feature. For the *lac* operon, its transcription is positively regulated by cAMP–CAP. For the *trp* operon, its transcription is fine-tuned by the process of attenuation. Attenuation is the mechanism by which the transcribing RNA polymerase pays attention to the translating ribosome, and bases its decision as to whether to transcribe the *trp* structural genes on if the translating ribosome stalls in its movement along the leader mRNA.

Regulons involve large networks of genes, often arranged in multiple operons, yet regulated by a common global regulatory mechanism. The heat-shock, SOS, and *cps* capsule regulons are controlled by several different mechanisms. The heat-shock alternate sigma factor, σ^{32}, directs RNA polymerase to transcribe only heat-shock promoters. σ^{32} levels are controlled by translational repression and protein stability. The SOS regulon is induced upon the removal of the LexA repressor. RecA complexed to single-stranded DNA facilitates LexA autocleavage, effectively removing LexA from SOS promoters. Signal transduction, through the transfer of phosphate groups, from RcsC to RcsB regulates the expression of genes needed to make colanic acid capsular polysaccharide. The second regulator of the capsule genes, RcsA, is regulated by protein stability.

Many examples of operons and regulons are known in *E. coli* and other prokaryotes. Conjugation under the control of the *tra* operon (Chapter 10) and λ's lytic/lysogenic decision-making process controlled by multiple operons (Chapter 7) represent examples of groups of genes coordinately regulated. Areas of intense research today center on the regulation of virulence gene activity in bacterial pathogenicity. The products encoded by virulence genes allow pathogenic bacteria to cause disease. Some virulence genes are organized and controlled as global regulons. Examples include regulation of the toxin genes of *Vibrio cholerae*, the causative agent of cholera and the regulation of the toxin genes of *Corynebacterium diphtheriae*, the causative agent of diphtheria. A sensor regulator pair involved in signal transduction regulates expression of the virulence genes of *Bordetella pertussis*, the causative agent of whooping cough.

Study questions

1 Give two examples each of positive and negative regulation.

2 Is it guaranteed that once mRNA is expressed that the encoded gene product will also be expressed? Why or why not?

3 What steps in the production of a protein can be regulated? What steps in the production of RNA can be regulated?

4 How does the presence of glucose in the medium in which *E. coli* is grown affect the synthesis of *E. coli*'s enzymes involved in the utilization of lactose? What happens at the molecular level to explain this effect?

5 Name a compound that only acts as an inducer of the *lac* operon and explain why it cannot be a substrate. Name a compound that is a substrate for β-galactosidase and explain why it cannot act as an inducer.

6 Describe transcriptional attenuation. Why hasn't this regulatory mechanism been described in eukaryotic cells?

7 Suppose you isolated a new bacterial species in the genus, *Finalococcus* and you wish to learn whether it has a multigene response (e.g. regulon) to the stress of taking a final exam. How would you experimentally demonstrate that this bacterium had a multigene response to the stress of taking a final?

8 What would happen to the regulation of the SOS regulon if the LexA binding regions were deleted? How would this impact the cell?

9 Contrast and compare repression in the *lac* and *trp* operons.

10 Why are super-repressor mutants dominant to wild-type repressor?

Further reading

Gottesman, S. 1995. Regulation of capsule synthesis: modification of the two-component paradigm by an accessory unstable regulator, In *Two-Component Signal Transduction*, eds. J. Hoch and T.J. Silhavy, pp. 253–62. Washington, DC: ASM Press.

Jacob, F. and Monod, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, **3**: 318–56.

Schwartz, D. and Beckwith, J.R. 1970. Mutants missing a factor necessary for the expression of catabolite sensitive operons in *E. coli*. In *The Lactose Operon*, eds. J.R. Beckwith and D. Zipser, pp. 417–22. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Walker, G.C. 1996. SOS regulon. In *Escherichia coli* and *Salmonella typhimurium: Cellular and Molecular Biology*, 2nd edn., eds. F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Hagasanik, W.S. Rexnikoff, M. Riley, M. Schaechter, and H.E. Umbarger. Washington, DC: ASM Press.

Weigle, J. 1953. Induction of mutations in a bacterial virus. *Proceedings of the National Academy of Science, USA*, **39**: 628–36.

Yanofsky, C. 1981. Attenuation in the control of expression of bacterial operons. *Nature*, **289**: 751–8.