Gene Expression and Regulation

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 is very important to cell economy. 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.

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 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 doublestranded, 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. 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. 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. 2). The **template strand** is complementary to the RNA. The nontemplate or **coding strand** contains the same sequence as the RNA except for the substitution of uracil for thymine.

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.

Figure 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.

Figure 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.

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. 3). 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 reform 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 s 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, sigma 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 **terminator** where RNA polymerase and the newly synthesized RNA dissociate from the DNA to end transcription.

Figure 3

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

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. 4). 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 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**.



Figure 4

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.

The lac Operon

The *lac* operon (Figure 5) consists of one regulatory gene (the *i* gene) and three structural genes (*z*, *y*, and *a*). The *i* gene codes for the repressor of the *lac* operon. The *z* gene codes for β**-galactosides** (β-gal), which is primarily responsible for the hydrolysis of the disaccharide, lactose into its monomeric units, galactose and glucose. The gene ***‘y’*** codes for **permease,** which is responsible for the high permeability of bacterial membranes for the β-galactosides. The ‘*a*’ gene encodes a **transacetylase.**

During normal growth on a glucose-based medium, the *lac* repressor is bound to the operator region of the *lac* operon, preventing transcription. However, in the presence of an inducer of the *lac* operon, the repressor protein binds the inducer and is rendered incapable of interacting with the operator region of the operon. RNA polymerase is thus able to bind at the promoter region, and transcription of the operon ensues. The *lac* operon is repressed, even in the presence of lactose, if glucose is also present. This repression is maintained until the glucose supply is exhausted. The repression of the *lac* operon under these conditions is called **catabolite repression** and is a result of the low levels of cAMP that result from an adequate glucose supply. The repression of the *lac* operon is relieved in the presence of glucose if excess cAMP is added. As the level of glucose in the medium falls, the level of cAMP increases. Simultaneously there is an increase in inducer binding to the *lac* repressor. The net result is an increase in transcription from the operon. The ability of cAMP to activate expression from the *lac* operon results from an interaction of cAMP with a protein called **CRP** (for **cAMP receptor protein**). The protein is also called **CAP** (for **catabolite activator protein**). The cAMP-CRP complex binds to a region of the *lac* operon just upstream of the region bound by RNA polymerase and that somewhat overlaps the repressor-binding site of the operator region. The binding of the cAMP-CRP complex to the *lac* operon stimulates RNA polymerase activity 20 to 50-fold.

Figure 5

Regulation of the lac operon in e. coli. The repressor of the operon is synthesized from the i gene. The repressor protein binds to the operator region of the operon and prevents RNA polymerase from transcribing the operon. In the presence of an inducer (such as the natural inducer, allolactose) the repressor is inactivated by interaction with the inducer. This allows RNA polymerase access to the operon and transcription proceeds. The resultant mRNA encodes the -galactosidase, permease, and transacetylase activities necessary for utilization of -galactosides (such as lactose) as an energy source. The lac operon is additionally regulated through binding of the cAMP-receptor protein, CRP (also termed the catabolite activator protein, CAP) to sequences near the promoter domain of the operon. The result is a 50-fold enhancement of polymerase activity.

The trp Operon

The *trp* operon encodes the genes for the synthesis of tryptophan. This cluster of genes, like the *lac* operon, is regulated by a repressor that binds to the operator sequences. The activity of the *trp* repressor for binding the operator region is enhanced when it binds tryptophan; in this capacity, tryptophan is known as a **co-repressor.** Since the activity of the *trp* repressor is enhanced in the presence of tryptophan, the rate of expression of the *trp* operon is graded in response to the level of tryptophan in the cell.