The Central Dogma

All organisms, from the simplest bacteria to ourselves, use the same basic mechanism of reading and expressing genes, so fundamental to life as we know it that it is often referred to as the "Central Dogma": Information passes from the genes (DNA) to an RNA copy of the gene, and the RNA copy directs the sequential assembly of a chain of amino acids. Said briefly, DNA → RNA → protein

The information encoded in genes is expressed in two phases: transcription, in which an RNA polymerase enzyme assembles an mRNA molecule whose nucleotide sequence is complementary to the DNA nucleotide sequence of the gene; and translation, in which a ribosome assembles a polypeptide, whose amino acid sequence is specified by the nucleotide sequence in the mRNA.

Transcription: An Overview

The first step of the Central Dogma is the transfer of information from DNA to RNA, which occurs when an mRNA copy of the gene is produced. Like all classes of RNA, mRNA is formed on a DNA template. Because the DNA sequence in the gene is transcribed into an RNA sequence, this stage is called **transcription.** Transcription is initiated when the enzyme **RNA polymerase** binds to a particular binding site called a **promoter** located at the beginning of a gene. Starting there, the RNA polymerase moves along the DNA strand into the gene. As it encounters each DNA nucleotide, it adds the corresponding complementary RNA nucleotide to a growing mRNA strand. Thus, guanine (G), cytosine (C), thymine (T), and adenine (A) in the DNA would signal the addition of C, G, A, and uracil (U), respectively, to the mRNA.

When the RNA polymerase arrives at a transcriptional "stop" signal at the opposite end of the gene, it disengages from the DNA and releases the newly assembled RNA chain. This chain is a complementary transcript of the gene from which it was copied.

RNA Polymerase and Transcription

The principal enzyme responsible for RNA synthesis is **RNA polymerase,** which catalyzes the polymerization of ribonucleoside 5'-triphosphates (NTPs) as directed by a DNA template. The synthesis of RNA is similar to that of DNA, and like DNA polymerase, RNA polymerase catalyzes the growth of RNA chains always in the 5' to 3' direction. Unlike DNA polymerase, however, RNA polymerase does not require a preformed primer to initiate the synthesis of RNA. RNA polymerase, like DNA polymerase, is a complex enzyme made up of multiple polypeptide chains. The intact bacterial enzyme consists of five different types of subunits, called α, β, β', σ and ω. The (σ subunit is relatively weakly bound and can be separated from the other subunits, yielding a core polymerase consisting of two α, one β one β', and one ω.

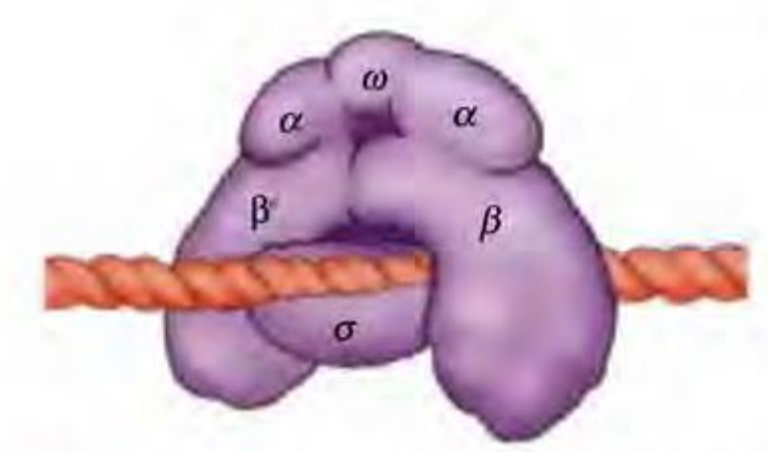


Fig. E. coli RNA polymerase

The core polymerase is fully capable of catalyzing the polymerization of NTPs into RNA, indicating that σ is not required for the basic catalytic activity of the enzyme. However, the core polymerase does not bind specifically to the DNA sequences that signal the normal initiation of transcription; therefore the *(*σ subunit is required to identify the correct sites for transcription initiation.

The DNA sequence to which RNA polymerase binds to initiate transcription of a gene is called the **promoter.** The DNA sequences involved in promoter function were first identified by comparisons of the nucleotide sequences of a series of different genes isolated from E. *coli.* These comparisons revealed that the region upstream of the transcription initiation site contains two sets of sequences that are similar in a variety of genes. These common sequences encompass six nucleotides each and are located approximately 10 and 35 base pairs upstream of the transcription start site. They are called the -10 and -35 elements, denoting their position relative to the transcription initiation site, which is defined as the+ 1 position.

Several types of experimental evidence support the functional importance of the -10 and -35 promoter elements. First, genes with promoters that differ from the consensus sequences are transcribed less efficiently than genes whose promoters match the consensus sequences more closely. Second, mutations introduced in either the -35 or - 10 consensus sequences have strong effects on promoter function. Third, the sites at which RNA polymerase binds to promoters have been directly identified by **footprint**ing experiments, which are widely used to determine the sites at which proteins bind to DNA. Footprinting analysis has shown that RNA polymerase generally binds to promoters over approximately a 60-base-pair region, extending from -40 to +20 (i.e., from 40 nucleotides upstream to 20 nucleotides downstream of the transcription start site). The σ subunit binds specifically to sequences in both the -35 and -10 promoter regions, substantiating the importance of these sequences in promoter function.

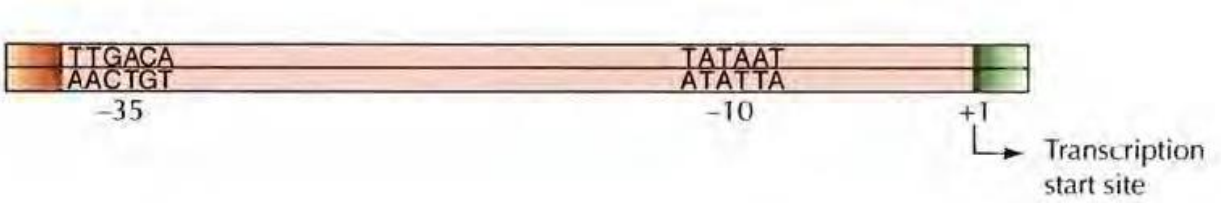


FIG. Sequences of *E.coli* promoters

In the absence of σ, RNA polymerase binds nonspecifically to DNA with low affinity. The role of σ is to direct the polymerase to promoters by binding specifically to both the -35 and -10 sequences, leading to the initiation of transcription at the beginning of a gene The initial binding between the polymerase and a promoter is referred to as a dosed-promoter complex because the DNA is not unwound. The polymerase then unwinds 12-14 bases of DNA, from about -12 to +2, to form an open-promoter complex in which single-stranded DNA is available as a template for transcription.

Transcription is initiated by the joining of two free NTPs. After addition of about the first 10 nucleotides, σ is released from the polymerase, which then leaves the promoter and moves along the template DNA to continue elongation of the growing RNA chain. During elongation, the polymerase remains associated with its template while it continues synthesis of mRNAs. As it travels, the polymerase unwinds the template DNA ahead of it and rewinds the DNA behind it, maintaining an unwound region of about 15 base pairs in the region of transcription. Within this unwound portion of DNA, 8-9 bases of the growing RNA chain are bound to the complementary template DNA strand. RNA synthesis continues until the polymerase encounters a termination signal, at which point transcription stops, the RNA is released from the polymerase, and the enzyme dissociates from its DNA template. There are two alternative mechanisms for termination of transcription in *E. coli.* The simplest and most common type of termination signal consists of a symmetrical inverted repeat of a GC-rich sequence followed by approximately seven A residues. Transcription of the GC-rich inverted repeat results in the formation of a segment of RNA that can form a stable stem loop structure by complementary base pairing. The formation of such a self complementary structure in the RNA disrupts its association with the DNA template and terminates transcription. Because hydrogen bonding between A and U is weaker than that between G and C, the presence of A residues downstream of the inverted repeat sequences is thought to facilitate the dissociation of the RNA from its template. Alternatively, the transcription of some genes is terminated by a specific termination protein (called Rho), which binds extended segments (greater than 60 nucleotides) of single stranded RNA.

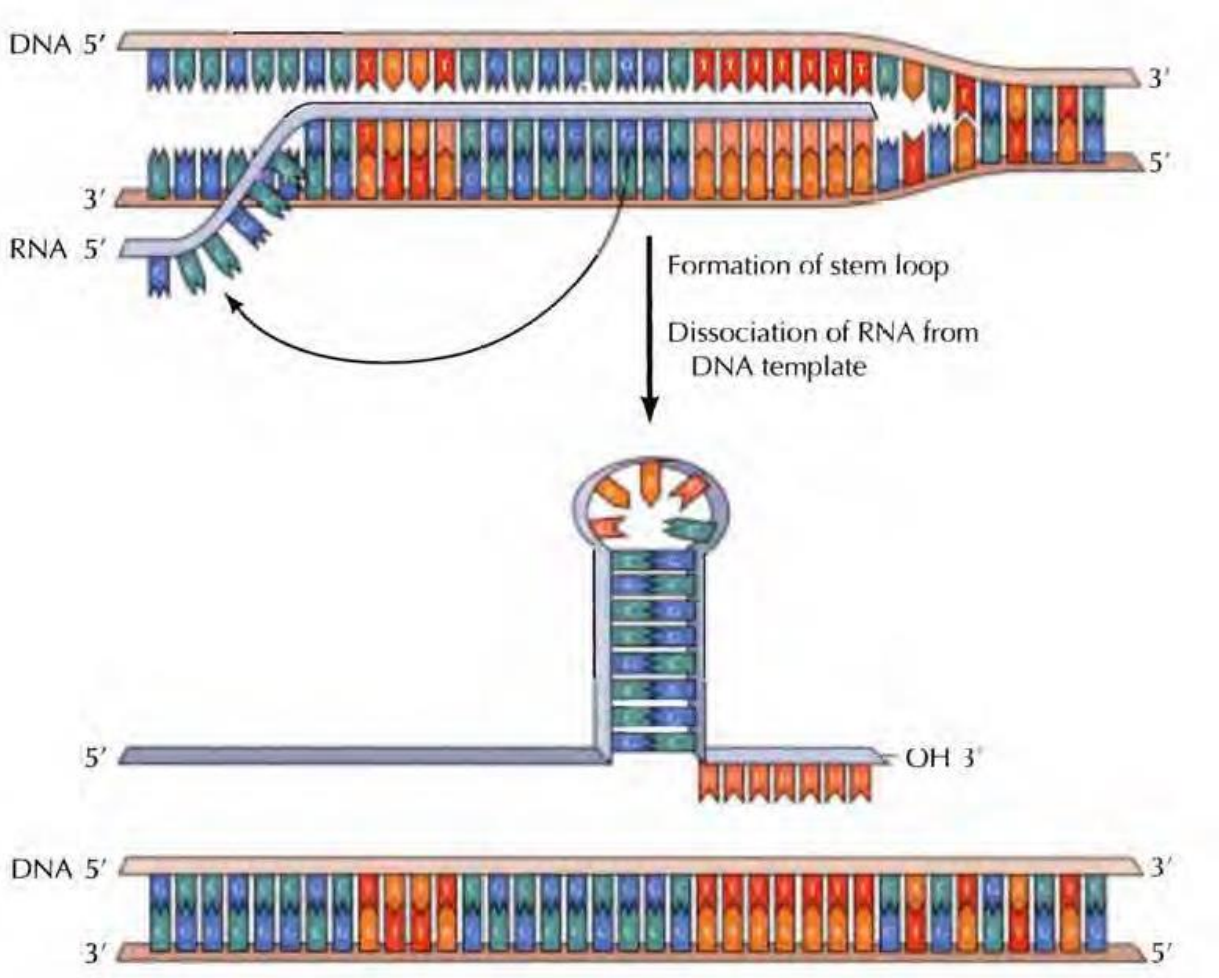


Fig. Transcription termination

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

Kinds of RNA

The class of RNA found in ribosomes is called **ribosomal RNA (rRNA).** During polypeptide synthesis, rRNA provides the site where polypeptides are assembled. In addition to rRNA, there are two other major classes of RNA in cells. **Transfer RNA (tRNA)** molecules both transport the amino acids to the ribosome for use in building the polypeptides and position each amino acid at the correct place on the elongating polypeptide chain. Human cells contain about 45 different kinds of tRNA molecules. **Messenger RNA (mRNA)** molecules are long strands of RNA that are transcribed from DNA and that travel to the ribosomes to direct precisely *which* amino acids are assembled into polypeptides. These RNA molecules, together with ribosomal proteins and certain enzymes, constitute a system that reads the genetic messages encoded by nucleotide sequences in the DNA and produces the polypeptides that those sequences specify.

The structure of tRNA

In 1965, after a 7-year effort, Robert Holley reported the first known base sequence of a biologically significant nucleic acid, that of yeast **alanine tRNA**(**tRNAAla**.

Almost all known tRNAs, as Holley first recognized, may be schematically arranged in the so-called cloverleaf secondary structure. Comparisons of tRNAs from various species have revealed many common structural features:

1. Each is a single chain containing between *73 and 93 ribonucleotides*.
2. They contain *many unusual bases,* typically between 7 and 15 per molecule. Some of these bases are methylated or dimethylated derivatives of A, U, C, and G formed by enzymatic modification of a precursor tRNA. Some methylations prevent the formation of certain base pairs, thereby rendering such bases accessible for interactions with other bases. In addition, methylation imparts a hydrophobic character to some regions of tRNAs, which may be important for their interaction with synthetases and ribosomal proteins.
3. The molecule is L-shaped.
4. About half the nucleotides in tRNAs are base-paired to form double helices. The four helical regions are arranged to form two apparently continuous segments of double helix.
5. The 5^- end of a tRNA is phosphorylated. The 5^- terminal residue is usually pG.
6. An activated amino acid is attached to a hydroxyl group of the adenosine residue in the amino acid-attachment site, located at the end of the 3^- CCA component of the acceptor stem. This single-stranded region can change conformation in the course of amino acid activation and protein synthesis.
7. The anticodon loop, which is present in a loop near the center of the sequence, is at the other end of the L, making accessible the three bases that make up the anticodon.

Transfer RNAs Have Characteristic Structural Features

Transfer RNAs are relatively small and consist of a single strand of RNA folded into a precise three-dimensional structure. The tRNAs in bacteria and in the cytosol of eukaryotes have between 73 and 93 nucleotide residues, corresponding to molecular

weights of 24,000 to 31,000. Mitochondria and chloroplasts contain distinctive, somewhat smaller tRNAs. Cells have at least one kind of tRNA for each amino acid; at least 32 tRNAs are required to recognize all the amino acid codons (some recognize more than one codon), but some cells use more than 32. Yeast alanine tRNA ($tRNA^{Ala}$), the first nucleic acid to be completely sequenced, contains 76 nucleotide residues, 10 of which have modified bases. Comparisons of tRNAs from various species have revealed many common structural features. Eight or more of the nucleotide residues have modified bases and sugars, many of which are methylated derivatives of the principal bases. Most tRNAs have a guanylate (pG) residue at the 5′ end, and all have the trinucleotide sequence CCA(3′) at the 3′ end. When drawn in two dimensions, the hydrogenbonding pattern of all tRNAs forms a cloverleaf structure with four arms; the longer tRNAs have a short fifth arm, or extra arm. In three dimensions, a tRNA has the form of a twisted L. Two of the arms of a tRNA are critical for its adaptor function. The **amino acid arm** can carry a specific amino acid esterified by its carboxyl group to the 2′- or 3′ hydroxyl group of the A residue at the 3′ end of the tRNA. The **anticodon arm** contains the anticodon. The other major arms are the **D arm**, which contains the unusual nucleotide dihydrouridine (D), and the **TψC arm**, which contains ribothymidine (T), not usually present in RNAs, and pseudouridine, which has an unusual carbon–carbon bond between the base and ribose. The D and TψC arms contribute important interactions for the overall folding of tRNA molecules, and the TψC arm interacts with the large subunit rRNA.

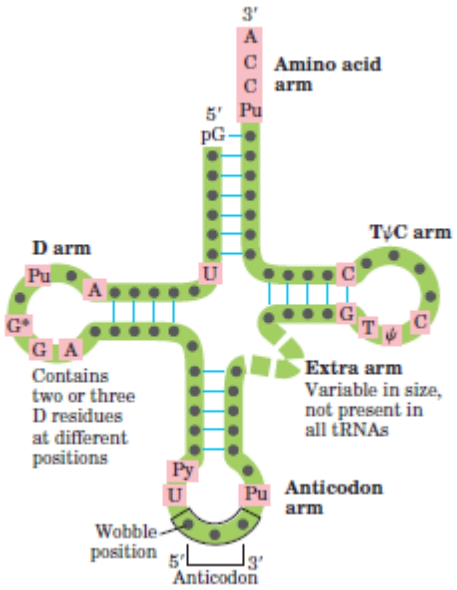


FIGURE General cloverleaf secondary structure of tRNAs. Transfer RNAs vary in length from 73 to 93 nucleotides. Extra nucleotides occur in the extra arm or in the D arm. At the end of the anticodon arm is the anticodon loop, which always contains seven unpaired nucleotides. The D arm contains two or three D (5,6-dihydrouridine) residues, depending on the tRNA. In some tRNAs, the D arm has only three hydrogen-bonded base pairs.

Wobble hypothesis

Examination of codon-anticodon pairings led Crick to conclude that the third base of most codons pairs rather loosely with the corresponding base of its anticodon; to use his picturesque word, the third base of such codons (and the first base of their corresponding anticodons) "wobbles." Crick proposed a set of four relationships called the **wobble hypothesis**:

1. The first two bases of an mRNA codon always form strong Watson-Crick base pairs with the corresponding bases of the tRNA anticodon and confer most of the coding specificity.
2. The first base of the anticodon (reading in the $5^{-}\rightarrow3^{-}$ direction; this pairs with the third base of the codon) determines the number of codons recognized by the tRNA. When the first base of the anticodon is C or A, base pairing is specific and only one codon is recognized by that tRNA. When the first base is U or G, binding is less specific and two different codons may be read. When inosine (I) is the first (wobble) nucleotide of an anticodon, three different codons can be recognized—the maximum number for any tRNA.
3. When an amino acid is specified by several different codons, the codons that differ in either of the first two bases require different tRNAs.
4. A minimum of 32 tRNAs are required to translate all 61 codons (31 to encode the amino acids and 1 for initiation).

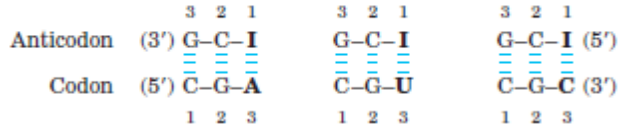


Fig. pairing relationship of codon anticodon

Wobble Allows Some tRNAs to Recognize More than One Codon

When several different codons specify one amino acid, the difference between them usually lies at the third base position (at the 3^{-} end). For example, alanine is coded by the triplets GCU, GCC, GCA, and GCG. The codons for most amino acids can be symbolized by $XY^{A}{}_{G}$ or $XY^{U}{}_{C}$. The first two letters of each codon are the primary determinants of specificity, a feature that has some interesting consequences. Transfer RNAs base-pair with mRNA codons at a three-base sequence on the tRNA called the **anticodon**. The first base of the codon in mRNA (read in the $5^{-}\rightarrow3^{-}$ direction) pairs with the third base of the anticodon. If the anticodon triplet of a tRNA recognized only one codon triplet through Watson-Crick base pairing at all three positions, cells would have a different tRNA for each amino acid codon. This is not the case, however, because the anticodons in some tRNAs include the nucleotide inosinate (designated I), which contains the uncommon base hypoxanthine. Inosinate can form hydrogen bonds with three different nucleotides (U, C, and A, although these pairings are much weaker than the hydrogen bonds of Watson-Crick base pairs (G=C and A=U). In yeast, one $tRNA^{Arg}$ has the anticodon (5^{-}) ICG, which recognizes three arginine codons: (5^{-})CGA, (5^{-})CGU, and

(5-)CGC. The first two bases are identical (CG) and form strong Watson-Crick base pairs with the corresponding bases of the anticodon, but the third base (A, U, or C) forms rather weak hydrogen bonds with the I residue at the first position of the anticodon.

The Genetic Code

■ The particular amino acid sequence of a protein is constructed through the translation of information encoded in mRNA. This process is carried out by ribosomes.
■ Amino acids are specified by mRNA codons consisting of nucleotide triplets. Translation requires adaptor molecules, the tRNAs, that recognize codons and insert amino acids into their appropriate sequential positions in the polypeptide.
■ The base sequences of the codons were deduced from experiments using synthetic mRNAs of known composition and sequence.
■ The codon AUG signals initiation of translation. The triplets UAA, UAG, and UGA are signals for termination.
■ The genetic code is degenerate: it has multiple codons for almost every amino acid.
■ The standard genetic code is universal in all species, with some minor deviations in mitochondria and a few single-celled organisms.
■ The third position in each codon is much less specific than the first and second and is said to wobble.
■ Translational frameshifting and RNA editing affect how the genetic code is read during translation.

Genetic information, stored in the chromosomes and transmitted to daughter cells through DNA replication, is expressed through transcription to RNA and, in the case of messenger RNA (mRNA), subsequent translation into proteins. The pathway of protein synthesis is called translation because the "language" of the nucleotide sequence on the mRNA is translated into the "language" of an amino acid sequence. The process of translation requires a genetic code, through which the information contained in the nucleic acid sequence is expressed to produce a specific sequence of amino acids.

THE GENETIC CODE

The correspondence between the sequence of bases in a codon and the amino acid residue it specifies is known as the **genetic code**. Its near universality among all forms of life is compelling evidence that life on Earth arose from a common ancestor. The genetic code is a dictionary that identifies the correspondence between a sequence of nucleotide bases and a sequence of amino acids. Each individual "word" in the code is composed of three nucleotide bases. These genetic words are called codons. There are four possible bases (U, C, A, and G) that can occupy each of the three positions in a codon, and hence there are 4^3-64 possible codons. Of these codons, 61 specify amino acids (of which there are only 20) and the remaining three, UAA, UAG, and UGA, are **Stop codons** that instruct the ribosome to cease polypeptide synthesis and release the resulting transcript.

A. Codons

Codons are presented in the mRNA language of adenine (A), guanine (G), cytosine (C), and uracil (U). Their nucleotide sequences are always written from the 5'-end to the 3'-end. The four nucleotide bases are used to produce the three-base codons. There are, therefore, 64 different combinations of bases, taken three at a time (a triplet code).

1. How to translate a codon: This table (or "dictionary") can be used to translate any codon and, thus, to determine which amino acids are coded for by an mRNA sequence. For example, the codon 5'-AUG-3' codes for methionine. [Note: AUG is the initiation (start) codon for translation.] Sixty-one of the 64 codons code for the 20 common amino acids.

2. Termination ("stop" or "nonsense") codons: Three of the codons, UAG, UGA, and UAA, do not code for amino acids, but rather are termination codons. When one of these codons appears in an mRNA sequence, synthesis of the polypeptide coded for by that mRNA stops.

Characteristics of the genetic code

Usage of the genetic code is remarkably consistent throughout all living organisms. It is assumed that once the standard genetic code evolved in primitive organisms, any mutation that altered its meaning would have caused the alteration of most, if not all, protein sequences, resulting in lethality. Characteristics of the genetic code include the following:

1. Specificity: The genetic code is specific (unambiguous) that is, a particular codon always codes for the same amino acid.

2. Universality: The genetic code is virtually universal, that is, its specificity has been conserved from very early stages of evolution, with only slight differences in the manner in which the code is translated. (An exception occurs in mitochondria)

3. Degeneracy: The genetic code is degenerate (sometimes called redundant). Although each codon corresponds to a single amino acid, a given amino acid may have more than one triplet coding for it. For example, arginine is specified by six different codons. Only Met and Trp have just one coding triplet.

4. Non-overlapping and comma less: The genetic code is non-overlapping and comma less, that is, the code is read from a fixed starting point as a continuous sequence of bases, taken three at a time. For example, AGCUGGAUACAU is read as AGC/UGG/AUA/CAU without any "punctuation" between the codons.

Consequences of altering the nucleotide sequence:

Changing a single nucleotide base on the mRNA chain (a "point mutation") can lead to any one of three results.

1. Silent mutation: The codon containing the changed base may code for the same amino acid. For example, if the serine codon UCA is given a different third base—U—to become UCU, it still codes for serine. This is termed a "silent" mutation.

2. Missense mutation: The codon containing the changed base may code for a different amino acid. For example, if the serine codon UCA is given a different first base—C—to become CCA, it will code for a different amino acid, in this case, proline. The substitution of an incorrect amino acid is called a "missense" mutation.

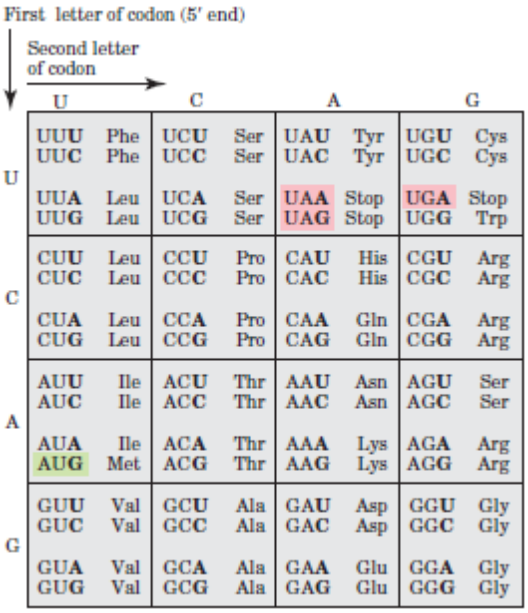
3. Nonsense mutation: The codon containing the changed base may become a termination codon. For example, if the serine codon UCA is given a different second base—A—to become UAA, the new codon causes termination of translation at that point, and the production of a shortened (truncated) protein. The creation of a termination codon at an inappropriate place is called a "nonsense" mutation.

4. Other mutations: These can alter the amount or structure of the protein produced by translation.

a. Trinucleotide repeat expansion: Occasionally, a sequence of three bases that is repeated in tandem will become amplified in number, so that too many copies of the triplet occur. If this happens within the coding region of a gene, the protein will contain many extra copies of one amino acid. For example, amplification of the CAG codon leads to the insertion of many extra glutamine residues in the huntingtin protein, causing the neurodegenerative disorder.

b. Splice site mutations: Mutations at splice sites can alter the way in which introns are removed from pre mRNA molecules, producing aberrant proteins.

c. Frame-shift mutations: If one or two nucleotides are either deleted from or added to the coding region of a message sequence, a frame-shift mutation occurs and the reading frame is altered. This can result in a product with a radically different amino acid sequence, or a truncated product due to the creation of a termination codon. If three nucletides are added, a new amino acid is added to the peptide or, if three nucleotides are deleted, an amino acid is lost. Loss of three nucleotides maintains the reading frame, but can result in serious pathology. For example, cystic fibrosis (CF), a hereditary disease that primarily affects the pulmonary and digestive systems, is most commonly caused by deletion of three nucleotides from the coding region of a gene, resulting in the loss of phenylalanine.



First letter of codon (5′ end)	Second letter of codon: U	C	A	G
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys
	UUA Leu	UCA Ser	UAA Stop	UGA Stop
	UUG Leu	UCG Ser	UAG Stop	UGG Trp
C	CUU Leu	CCU Pro	CAU His	CGU Arg
	CUC Leu	CCC Pro	CAC His	CGC Arg
	CUA Leu	CCA Pro	CAA Gln	CGA Arg
	CUG Leu	CCG Pro	CAG Gln	CGG Arg
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser
	AUC Ile	ACC Thr	AAC Asn	AGC Ser
	AUA Ile	ACA Thr	AAA Lys	AGA Arg
	AUG Met	ACG Thr	AAG Lys	AGG Arg
G	GUU Val	GCU Ala	GAU Asp	GGU Gly
	GUC Val	GCC Ala	GAC Asp	GGC Gly
	GUA Val	GCA Ala	GAA Glu	GGA Gly
	GUG Val	GCG Ala	GAG Glu	GGG Gly

Amino acid	Number of codons	Amino acid	Number o codons
Met	1	Tyr	2
Trp	1	Ile	3
Asn	2	Ala	4
Asp	2	Gly	4
Cys	2	Pro	4
Gln	2	Thr	4
Glu	2	Val	4
His	2	Arg	6
Lys	2	Leu	6
Phe	2	Ser	6

Fig. Dictionary of amino acid code words in mRNA Table: degeneracy of genetic code

Cells Use RNA to Make Protein

Proteins are synthesized in the cytoplasm, on large RNA protein aggregates called **ribosomes.** These polypeptide-making factories are very complex, composed of several RNA molecules and over 50 different proteins. Protein synthesis involves three different sites on the ribosome surface, called the P, A, and E sites.

Translation

In prokaryotes, translation begins when the initial portion of an mRNA molecule binds to an rRNA molecule in a ribosome. The mRNA lies on the ribosome in such a way that only one of its codons is exposed at the polypeptide making site at any time. A tRNA molecule possessing the complementary three-nucleotide sequence, or anticodon, binds to the exposed codon on the mRNA. Because this tRNA molecule carries a particular amino acid, that amino acid and no other is added to the polypeptide in that position. As the mRNA molecule moves through the ribosome, successive codons on the mRNA are exposed, and a series of tRNA molecules bind one after another to the exposed codons. Each of these tRNA molecules carries an attached amino acid, which it adds to the end of the growing polypeptide chain. There are about 45 different kinds of tRNA molecules. Why are there 45 and not 64 tRNAs (one for each codon)? Because the third base-pair of a tRNA anticodon allows some “wobble,” some tRNAs recognize more than one codon. How do particular amino acids become associated with particular tRNA molecules? The

key translation step, which pairs the three-nucleotide sequences with appropriate amino acids, is carried out by a remarkable set of enzymes called activating enzymes. **Activating Enzymes:** Particular tRNA molecules become attached to specific amino acids through the action of activating enzymes called **aminoacyl-tRNA synthetases,** one of which exists for each of the 20 common amino acids. Therefore, these enzymes must correspond to specific anticodon sequences on a tRNA molecule as well as particular amino acids. Some activating enzymes correspond to only one anticodon and thus only one tRNA molecule. Others recognize two, three, four, or six different tRNA molecules, each with a different anticodon but coding for the same amino acid. If one considers the nucleotide sequence of mRNA a coded message, then the 20 activating enzymes are responsible for decoding that message.

"Start" and "Stop" Signals: There is no tRNA with an anticodon complementary to three of the 64 codons: UAA, UAG, and UGA. These codons, called **nonsense codons,** serve as "stop" signals in the mRNA message, marking the end of a polypeptide. The "start" signal that marks the beginning of a polypeptide within an mRNA message is the codon AUG, which also encodes the amino acid methionine. The ribosome will usually use the first AUG that it encounters in the mRNA to signal the start of translation.

The Process of Translation

Translation is generally divided into five stages:

Stage 1: Activation of Amino Acids For the synthesis of a polypeptide with a defined sequence, two fundamental chemical requirements must be met: (1) the carboxyl group of each amino acid must be activated to facilitate formation of a peptide bond, and (2) a link must be established between each new amino acid and the information in the mRNA that encodes it. Both these requirements are met by attaching the amino acid to a tRNA in the first stage of protein synthesis. Attaching the right amino acid to the right tRNA is critical. This reaction takes place in the cytosol, not on the ribosome. Each of the 20 amino acids is covalently attached to a specific tRNA at the expense of ATP energy, using $Mg2^{+}$ dependent activating enzymes known as aminoacyl-tRNA synthetases. When attached to their amino acid (aminoacylated) the tRNAs are said to be "charged." An individual aminoacyl-tRNA synthetase must be specific not only for a single amino acid but for certain tRNAs as well. The interaction between aminoacyl-tRNA synthetases and tRNAs has been referred to as the "second genetic code," reflecting its critical role in maintaining the accuracy of protein synthesis.

Stage 2: Initiation

In prokaryotes, polypeptide synthesis begins with the formation of an **initiation complex.** First, a tRNA molecule carrying a chemically modified methionine called *N*-formylmethionine (tRNAfMet) binds to the small ribosomal subunit. Proteins called

initiation factors (IF-1, IF-2,IF-3) position the tRNAfMet on the ribosomal surface at the *P site* (for peptidyl), where peptide bonds will form. Nearby, two other sites will form: the *A site* (for aminoacyl), where successive amino acid-bearing tRNAs will bind, and the *E site* (for exit), where empty tRNAs will exit the ribosome (figure 15.12). This initiation complex, guided by another initiation factor, then binds to the anticodon AUG on the mRNA. Proper positioning of the mRNA is critical because it determines the reading frame—that is, which groups of three nucleotides will be read as codons. Moreover, the complex must bind to the beginning of the mRNA molecule, so that all of the transcribed gene will be translated. The correct binding of the fMet-tRNAfMet to the P site in the complete 70S initiation complex is assured by at least three points of recognition and attachment: the codon-anticodon interaction involving the initiation AUG fixed in the P site; interaction between the Shine-Dalgarno sequence in the mRNA and the 16S rRNA; and binding interactions between the ribosomal P site and the fMet-tRNAfMet. The initiation complex is now ready for elongation.

Initiation in eukaryotes is similar, although it differs in two important ways. First, in eukaryotes, the initiating amino acid is methionine rather than *N*-formylmethionine. Second, the initiation complex is far more complicated than in bacteria, containing nine or more protein factors, many consisting of several subunits.

Stage 3: Elongation

After the initiation complex has formed, the large ribosome subunit binds, exposing the mRNA codon adjacent to the initiating AUG codon, and so positioning it for interaction with another amino acid-bearing tRNA molecule. When a tRNA molecule with the appropriate anticodon appears, proteins called elongation factors assist in binding it to the exposed mRNA codon at the A site. When the second tRNA binds to the ribosome, it places its amino acid directly adjacent to the initial methionine, which is still attached to its tRNA molecule, which in turn is still bound to the ribosome. The two amino acids undergo a chemical reaction, catalyzed by *peptidyl transferase,* which releases the initial methionine from its tRNA and attaches it instead by a peptide bond to the second amino acid.

Translocation In a process called **translocation**, the ribosome now moves (translocates) three more nucleotides along the mRNA molecule in the 5′ → 3′ direction, guided by other elongation factors. This movement relocates the initial tRNA to the E site and ejects it from the ribosome, repositions the growing polypeptide chain (at this point containing two amino acids) to the P site, and exposes the next codon on the mRNA at the A site. When a tRNA molecule recognizing that codon appears, it binds to the codon at the A

site, placing its amino acid adjacent to the growing chain. The chain then transfers to the new amino acid, and the entire process is repeated.

Stage 4: Termination

Elongation continues in this fashion until a chain-terminating nonsense codon is exposed (for example, UAA, UAG, UGA). Nonsense codons do not bind to tRNA, but they are recognized by **release factors,** proteins that release the newly made polypeptide from the ribosome.

Stage 5: Folding and Posttranslational Processing

In order to achieve its biologically active form, the new polypeptide must fold into its proper three-dimensional conformation. Before or after folding, the new polypeptide may undergo enzymatic processing, including removal of one or more amino acids (usually from the amino terminus); addition of acetyl, phosphoryl, methyl, carboxyl, or other groups to certain amino acid residues; proteolytic cleavage; and/or attachment of oligosaccharides or prosthetic groups.

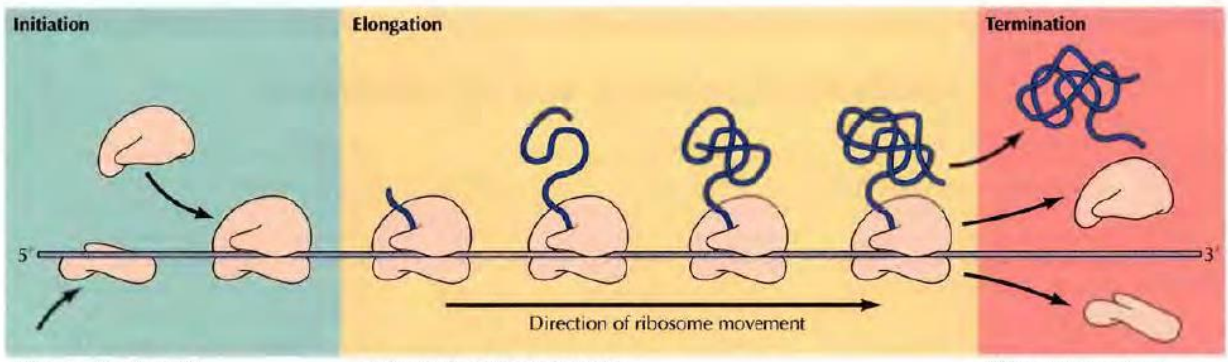


FIGURE 8.9 Overview of translation

Table: Translation factors

Role	Prokaryotes	Eukaryotes
Initiation	IF1, IF2, IF3	eIF1, eIF1A, eIF2, eIF2B, eIF3, eIF4A, eIF4B, eIF4E, eIF4G, eIF4H, eIF5, eIF5B
Elongation	EF-Tu, EF-Ts, EF-G	eEF1α, eEF1$\beta\gamma$, eEF2
Termination	RF1, RF2, RF3	eRF1, eRF3

<u>Gene regulation or Regulation of protein synthesis</u>

Out of some 4000 genes in bacteria only a few are expressed at a given time. The same is true for other organisms, although the number of genes may vary significantly. The requirement for specific enzymes changes with the phase of life cycle in which an

organism is passing through. Thus there ought to be some sort of genetic machinery to regulate the synthesis of certain proteins just for the following reason:

1. Specific proteins are synthesized in those cells only where they are needed, although their genes are present in all the cells of an organism.
2. A particular protein is synthesized only at the time when it is needed by the organism.
3. Only the required amount of proteins is synthesized.

At a particular time certain set of genes are 'switched on' and others 'switched off'. When switched on a gene is active for synthesizing proteins and when switched off it is inactive.

Operon

Jacob and Monod 1961 proposed a scheme for induction and repression of enzyme synthesis popularly known as operon model. An operon is a group or set of genes situated next to each other in the DNA that can be switched on or off in a unified manner. In other words it is the unit of transcription.

Inducible system: In *E. coli* the utilization of lactose is brought about by three enzymes namely β-galactosidase, lac permease, and transacetylase. The enzyme β-galactosidase hydrolyses lactose into glucose and galactose. When lactose is added in the medium the production of the enzymes is enhanced by 1,000 times. This phenomenon is called induction. Such enzymes whose production is induced by adding the substrate are called inducible enzymes and the genetic system which controls their production as inducible system. The substrates whose addition induces the synthesis of enzymes are called inducers. In the absence of the inducer the genes for the synthesis of the enzyme don't function.

Repressible system: The synthesis of amino acid tryptophan needs five enzymes in *E. coli*. If tryptophan is added in the medium the synthesis of enzymes is reduced. The phenomenon is called repression. Such enzymes whose production is checked by the addition of the end product (tryptophan in present case) are repressible enzymes and the governing genetic system as repressible system. Such end products which bring about repression are called co-repressors.

The cell possesses certain molecules called repressors which check the activity of genes. A repressor may be active or inactive. The active repressor becomes inactive by adding an inducer. The inactive repressor is called aporepressor which becomes active if co-repressor is added.

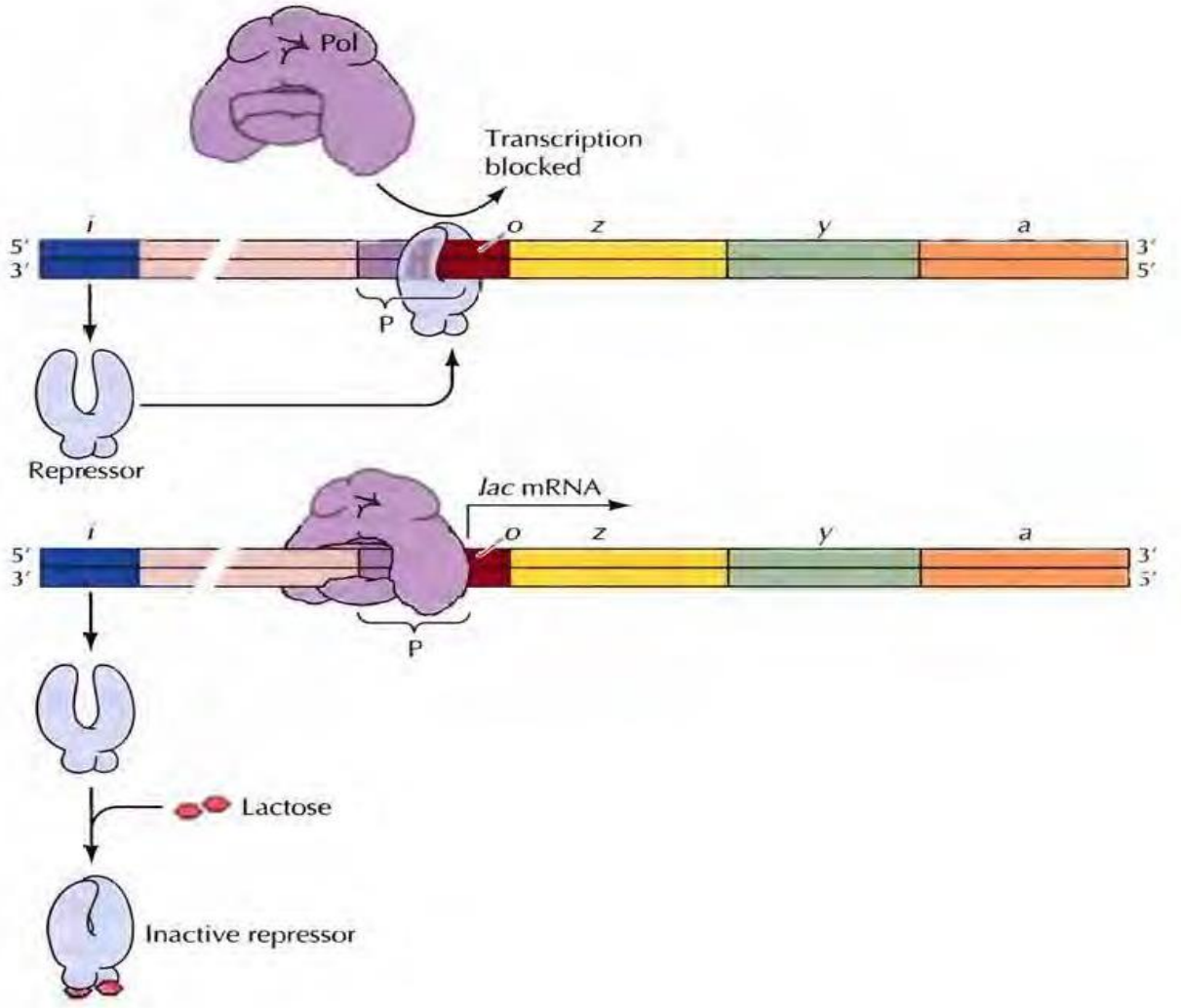
Repressors and Negative Control of Transcription

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

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

The genes encoding, β-galactosidase, permease, and transacetylase are expressed as a single unit, called an operon. Transcription of the operon is controlled by *o* (the operator), which is adjacent to the transcription initiation site. The *i* gene encodes a protein that regulates transcription by binding to the operator. Normal *i* gene product is a repressor, which blocks transcription when bound too. The addition of lactose leads to induction of the operon because lactose binds to the repressor, thereby preventing it from binding to the operator DNA.

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

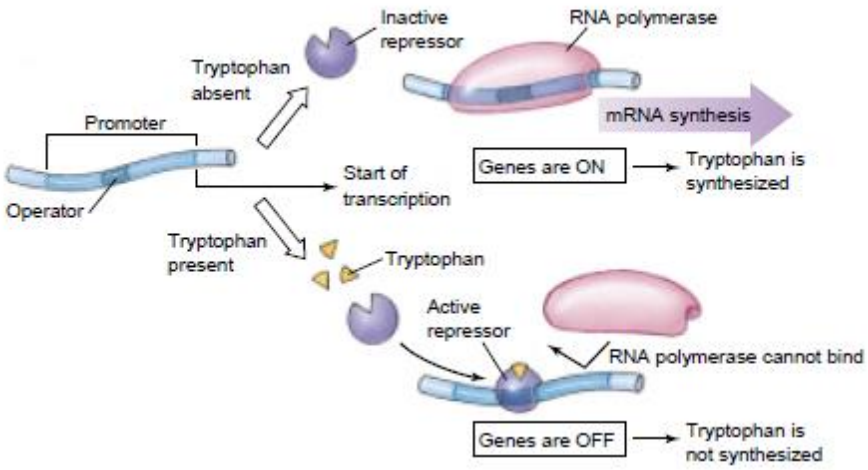


Negative control of the *lac* operon The *i* gene encodes a repressor which, in the absence of lactose (top), binds to the operator (o) and interferes with the binding of RNA polymerase to the promoter, blocking transcription of the three structural genes (z, *β*-galactosidase; *y*, permease; and *a*,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

Repressors Are *OFF* Switches

A typical bacterium possesses genes encoding several thousand proteins, but only some are transcribed at any one time; the others are held in reserve until needed. When the cell encounters a potential food source, for example, it begins to manufacture the enzymes necessary to metabolize that food. Perhaps the best-understood example of this type of transcriptional control is the regulation of tryptophan-producing genes (*trp* genes), which was investigated in the pioneering work of Charles Yanofsky and his students at Stanford University.

Tryptophan Operon: The bacterium *Escherichia coli* uses proteins encoded by a cluster of five genes to manufacture the amino acid tryptophan. All five genes are transcribed together as a unit called an **operon,** producing a single, long piece of mRNA. RNA polymerase binds to a promoter located at the beginning of the first gene, and then proceeds down the DNA, transcribing the genes one after another. Regulatory proteins shut off transcription by binding to an operator site immediately in front of the promoter and often overlapping it. When tryptophan is present in the medium surrounding the bacterium, the cell shuts off transcription of the *trp* genes by means of a tryptophan **repressor,** a helixturn- helix regulatory protein that binds to the operator site located within the *trp* promoter (figure 16.6). Binding of the repressor to the operator prevents RNA polymerase from binding to the promoter. The key to the functioning of this control mechanism is that the tryptophan repressor cannot bind to DNA unless it has first bound to two molecules of tryptophan. The binding of tryptophan to the repressor alters the orientation of a pair of helix turnhelix motifs in the repressor, causing their recognition helices to fit into adjacent major grooves of the DNA (figure 16.7).
Thus, the bacterial cell's synthesis of tryptophan depends upon the absence of tryptophan in the environment. When the environment lacks tryptophan, there is nothing to activate the repressor, so the repressor cannot prevent RNA polymerase from binding to the *trp* promoter. The *trp* genes are transcribed, and the cell proceeds to manufacture tryptophan from other molecules. On the other hand, when tryptophan is present in the environment, it binds to the repressor, which is then able to bind to the *trp* promoter. This blocks transcription of the *trp* genes, and the cell's synthesis of tryptophan halts.



How the *trp* operon is controlled. The tryptophan repressor cannot bind the operator (which is located *within* the promoter) unless tryptophan first binds to the repressor. Therefore, in the absence of tryptophan, the promoter is free to function and RNA polymerase transcribes the operon. In the presence of tryptophan, the tryptophan repressor complex binds tightly to the operator, preventing RNA polymerase from initiating transcription.