Unit – 2 Gene Expression

I) Salient features of Genetic code
II) Biological expression of gene: Protein synthesis -Transcription and Translation processes
III) Regulation of gene expression: Lac operon and Ara operon

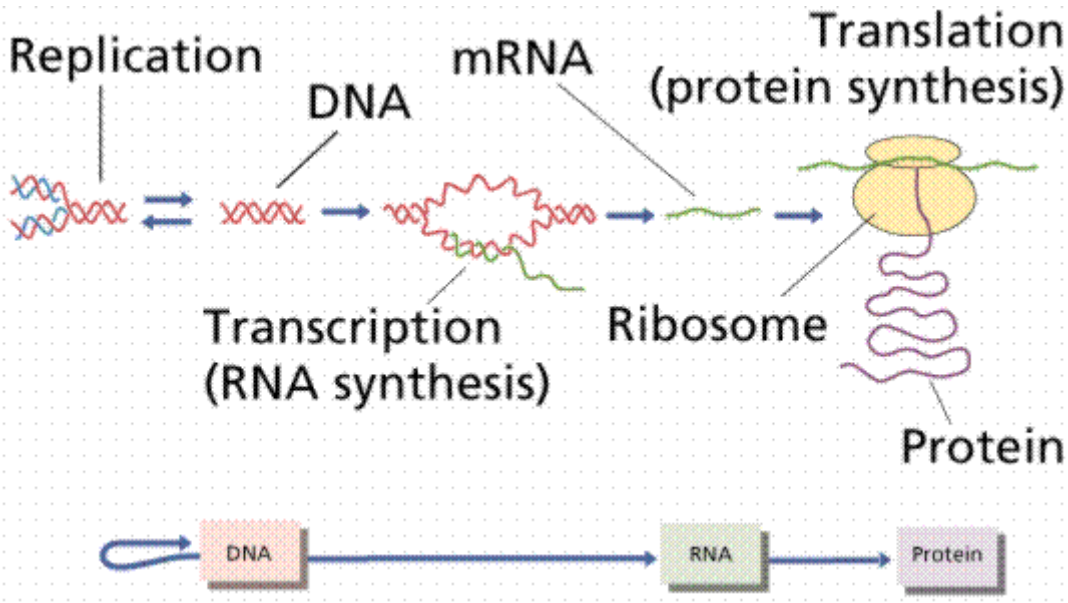
==========

I) Salient features of Genetic code

THE GENETIC CODE

The relationship between genes and proteins is desirable for an understanding of the genetic code.

1. All metabolic reactions are catalyzed by specific enzymes. All enzymes are proteins.
2. The action of an enzyme depends upon the sequence of amino acids constituting it.
3. The one-gene one-enzyme hypothesis proposed by Beadle and Tatum in 1940s states that the synthesis of an enzyme (actually of a polypeptide chain) is controlled by a particular gene.
4. The gene, which is almost always a segment of a DNA strand from which mRNA is formed (transcription), which in turn forms polypeptide chain (translation).
5. Messenger RNA thus acts as an intermediate in conveying information from the sequence of nucleotides in DNA to the sequence of amino acids in the polypeptide chain (Sequence Hypothesis).
6. Each amino acid is specified by a sequence of three bases (the codon) on mRNA.
7. Each tRNA molecule has a sequence of three bases (the anticodon) which reads a codon of mRNA. Transfer RNA molecules thus serve as adaptors in proteins synthesis by reading mRNA codons in a sequence (Click's Adaptor Hypothesis).
8. The relationship between the sequence of bases in DNA|RNA and the sequence of amino acids in a polypeptide chain is called the genetic code. The code indicates which codons specify which amino acids.



Amino acids Involved in protein synthesis.

About 150 amino acids are found in nature of which only 20 are specified by the genetic code. Only these 20 amino acids take part in protein synthesis. Among the other amino acids found in proteins are cystine and hydroxyprolin.

The genes of a cell contain coded information for the maintenance and reproduction of the cell. They direct the arrangement of the 20 types of amino acids into the polypeptide chains of the protein molecules. A polypeptide chain typically contains about 100-300 amino acids and is formed by specific arrangement of the 20 types of amino acids.

Salient features / Properties of genetic code

1. **The genetic code is a triplet code**
2. **The code is non- overlapping**
3. **The code is comma less**
4. **The code has polarity**
5. **Codons and anticodons**
6. **Initiation codons**
7. **Termination codons /Nonsense codons**
8. **The code is degenerate**
9. **The wobble hypothesis**
10. **The code is universal**

1. The genetic code is a triplet code.

DNA contains four kinds of nucleotides (A, T, G and C), and proteins are synthesized from 20 different types of amino acids. A basic problem regarding the genetic code was: how many bases of DNA specify one amino acid?

In a singlet code each base or letter would specify one amino acid. Only 4 of the 20 types of amino acids would be coded unambiguously (not confusing) by a singlet code (Table). In a two-letter or doublet code two bases would specify one amino acid. Here 16 (4 x 4) of the 20 amino acids can be specified, but there would be ambiguous (confusion) determination of a number of amino acids. A triplet or three-letter code was first suggested by the physicist Gamow in 1954. According to the triplet code three letters or bases specify one amino acid. Thus 64 (4x4x4) distinct triplets of purine and /or pyrimidine bases determine the 20 amino acids. These triplets have been called codons. Since there are 64 codons and only 20 amino acids it is obvious that there are many more codons than there are amino acids, i.e. the code is degenerate (many numbers). Experimental evidence shows that the code is a triplet one and that 61 of the 64 codons code for individual amino acids during protein synthesis.

Table: - The maximum possible number of codons in the singlet, doublet and triplet codes.

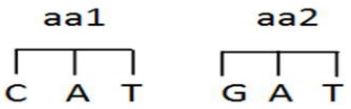
Type of code	**Number of bases in codon**	**Number of codons**	**Ambiguous/ Degenerate**
Singlet code	1	4 X 1= 4	Ambiguous (confusing)
Doublet code	2	4 X 4 = 16	Ambiguous
Triplet code	3	4 X 4 X 4 = 64	Degenerate (many numbers)
Quadruplet code	4	4x4x4x4 = 256	Degenerate

2. The code is non- overlapping.

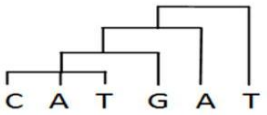
Since the DNA molecule is a long chain of nucleotides, it could be read either in an overlapping or non-overlapping manner. The genetic code could thus be overlapping or non-overlapping. The reading of the code by these two different ways would yield different results. In the non-overlapping code six nucleotides would code for two amino acids, while in the overlapping code up to four amino acids could be coded. In the non-overlapping code each letter is read only once while in the overlapping code it would be read three times, each time as a part of a different word. Mutational changes in one letter would affect

only one word in the non-overlapping code, while it would affect three words in the overlapping code.

Non- overlapping code. (C, A, T **&** G are bases, aa1 and aa2 are amino acids)



Overlapping code



Studies on gene mutations show that the code is of the non-overlapping type. In the tobacco mosaic virus (TMV) mutation of one base of the nucleic acid into another results in the alteration of only a single amino acid. Similarly, studies on normal and sickle cell haemoglobins show that a single mutational change results in the substitution of only one amino acid.

3. The code is comma less.

Is the genetic code read in an uninterrupted manner from one end of the nucleic acid chain to the other? Or are there bases (commas) between successive codons? A code with commas could be represented as follows (the X represents a base acting as a comma).

UUU	X	CUC	X	GUA	X	UCC	X	ACC	.....	Bases
Phe		Leu		Val		Ser		Thr	.....	Amino acids

A mutation resulting in an addition or deletion of a base would affect only one amino acid of the polypeptide chain. The total genetic message would be only slightly changed.

UUU	X -UC	X	GUA	X	UCC	X	ACC	.....	Bases
Phe	Changed		Val		Ser		Thr	.....	Amino acids
	a a								

A comma less code would not have the comma bases and can be represented thus:

UUU	CUC	GUA	UCC	ACC	.....	Bases
Phe	Leu	Val	Ser	Thr	.....	Amino acids

In such a code any mutation involving a deletion of a base (—C) would result in a drastic change in the genetic message.

UUU	UCG	UAU	CCA	CC	.....	Bases
Phe	Ser	Tyr	Pro		.....	Amino acids

The entire series of amino acids following the deletion would change. All the available evidence indicates that the code is comma less, i.e. there are no demarcating signals between codons. The work of **Khorana** and his associates cited below gives clear evidence of a comma less code. Long synthetic polynucleotides with specific repeating sequences were used for translation of protein chains. Thus the repeating sequence CUCUCU...--contains the codons CUC (for leucine) and UCU (for serine). When this sequence is used for translation of proteins, no other amino acid is incorporated into the protein though others are also present. This result can only be explained by a comma less triplet code where there would have to be alternate translation of CUC and UCU codons.

4. The code has polarity

If a gene is to specify the same protein repeatedly it is essential that the code must be read between fixed start and end points. These points are the initiation and the termination codons, respectively. It is also essential that the code must be read in a fixed direction. In other words the code must have polarity. It is obvious that if the code is read in opposite directions it would specify two different proteins, since the codons would have reversed base sequences. Thus if the message given below is read from left to right the first codon, UUG, would specify leucine.

Codons	UUG	AUC	GUC	UCG	CCA	ACA	AGG
⟶	Leu	Ile	Val	Ser	Pro	Thr	Arg
	Val	Leu	Leu	Ala	Thr	Thr	Gly ⟵

If read from right to left the codon would become GUU and would specify valine. It is thus seen that the sequence of amino acids constituting the protein would undergo a drastic change if the code is read in the opposite direction. The available evidence indicates that the message in mRNA is read in the 5' to 3' direction. The polypeptide chain is synthesized in the N ⟶ C direction, i.e. from the amino (NH_2) terminal to the carboxyl (COOH) terminal.

5. Codons and anticodons

During translation the codons of mRNA pair with complementary anticodons of tRNA. Since mRNA is read in a polar manner in the 5' → 3' direction, the codons are also written in the 5'⟶3' direction. Thus the codon AUG is written as 5'AUG3'. The corresponding anticodon on tRNA should therefore be written as 5'CAU3'. In such a configuration the first bases of both codon and anticodon would be the ones at the 5' end and third bases at the 3' end.

Base number		1	2	3	
Codon (mRNA)	5'	A	U	G	3'
Anticodon (tRNA)	3'	U	A	C	5'
Base number		3	2	1	

Often, however, the anticodon is written in the 3' ⟶ 5' direction so as to bring about an easier correlation between the bases of the codon and anticodon. Thus the anticodon for AUG is written as 3' UAC 5' or, more simply, UAC. Here the first letter in the codon is at the 5' end and the first letter of the anticodon at the 3' end.

6. Initiation codons

The starting amino acid in the synthesis of most protein chains is methionine (eukaryotes) or N-formylmethionine (prokaryotes). Methionyl or N-formyl methionyl-tRNA specifically binds to initiation sites containing the AUG codon. This codon is therefore called the initiation codon. Less often, GUG also serves as the initiation codon in bacterial protein synthesis. Normally GUG is the codon for valine. In the phage MS2, GUG is the initiation codon for the A protein. GUG has been found to initiate protein synthesis when the normal AUG codon is lost by deletion. However, initiation by GUG is less efficient since it has a lower affinity for fMet-tRNA.

Both AUG and GUG codons show ambiguity (doubt) in one sense, since each of them codes for two different amino acids. When these two codons are at initiation positions of mRNA they code for N-formyl methionine. In internal positions AUG codes for methionine and GUG for valine.

7. Termination codons / Nonsense codons

Three of the 64 codons do not specify any tRNA, and hence called nonsense codons.

These codons are

UAG (amber),

UAA (ochre)

UGA (opal or umber).

Since they bring about termination of poly-peptide chain synthesis they are also called termination codons, UAG was the first termination codon to be discovered. It was named 'amber' after a graduate student named Bernstein (the German for 'amber') who helped in the discovery of a class of mutations. Apparently to give uniformity the other two termination codons were also named after colours.

Termination codons do not code for any amino acids and hence cause termination and release of polypeptide chains. Apparently no tRNA species has anticodons complementary to the termination codons. There are mRNAs with single termination codons and also mRNAs with two successive termination codons (e.g. MS2coat protein mRNA). Termination codons are not read by any tRNA molecules but by proteins called release factors. In prokaryotes there are three release factors RF-1, RF-2 and RF-3. RF-1 recognizes UAG and UAA, while RF-2 recognizes UAA and UGA. RF-3 stimulates RF-1 and RF-2. In eukaryotes a single release factor (RF) recognizes all three termination codons.

8. The code is degenerate (many numbers)

As mentioned previously, there are 64 possible codons in a triplet code, of which 61 have been shown to code amino acids. Since only 20 amino acids take part in protein synthesis, it is obvious that there are many more codons than amino acid types. Except for tryptophan and methionine, which have a single codon each, all other amino acids involved in protein synthesis have more than one codon. Phenylalanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartic acid, glutamic acid and cysteine have two codons each. Isoleucine has three codons. Valine, praline, threonine, alanim and glycine have four codons each. Leucine, arginine and serine have six codons each. This variability in the number of codons for different amino acids may at least partially account for the unequal distribution of the different amino acids in protein. In general, the

frequency of appearance of amino acids in proteins roughly corresponds to the number of available codons.

Table: - Number of codons coding for different amino acids. Ammo acids in categories 2-5 are coded by more than one codon. Such codons are called degenerate.

	Amino acids	**Number of codons**
1.	Tryptophan, methiotiine	1
2	Phenylalanine, tyrosine, histidine, glutamine, asparagines, lysine, aspartic acid, glutamic acid, cysteine	2
3.	Isoleucine	3
4	Valine, proline, threonine, alanine, glycine	4
5.	Leucine, arginine, serine.	6

9. The wobble hypothesis

The triplet code is a degenerate one with many more codons than the number of amino acid types coded. An explanation for this degeneracy is provided by the 'wobble hypothesis' proposed by Crick (1966). Since there are 61 codons specifying amino acids, the cell should contain 61 different tRNA molecules, each with a different anticodon. Actually, however, the number of tRNA molecule types discovered is much less than 61. This implies that the anticodons of some tRNAs read more than one codon on mRNA.

According to the wobble hypothesis only the first two positions of a triplet codon on mRNA have a precise pairing with the bases of the tRNA anticodon. The pairing of the third position bases of the codon may be ambiguous, and varies according to the nucleotide present in this position. Thus a single tRNA type is able to recognize two or more codons differing only in the third base. The anticodon UCG of serine tRNA recognizes two codons, AGC and AGU. The bonding between UCG and AGC follows the usual Watson-Crick pairing pattern. In UCG-AGU pairing, however, hydrogen bonding takes place between G and U. This is a departure from the usual Watson-Crick pairing mechanism where G pairs with C and A with U. Such interaction between the third bases is referred to as 'wobble pairing'.

mRNA codons (serine)	5'	AGC	3'		5'	AGU	3'
tRNA anticodon	3'	UCG	5'		3'	UCG	5'

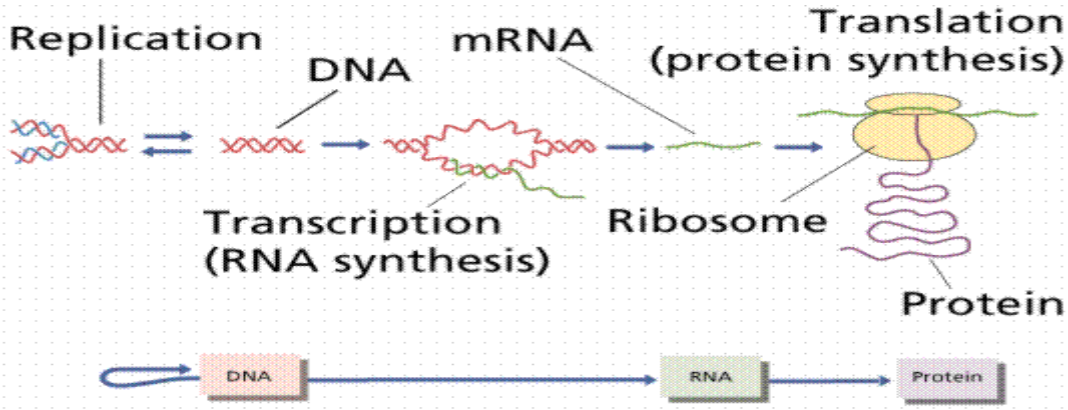
10.The code is universal.

The genetic code is valid for all organisms ranging from bacteria to man. It is essentially the same for all organisms and is therefore said to be universal. The universality of the code was demonstrated by Marshall, Caskey and Nirenberg (1967), who found that *E. coli* (bacterium), *Xenopus laevis* (amphibian) and *Guinea pig* (mammal) amino acyl tRNAs use almost the same code. This showed that the code is essentially universal.

Other evidence for the universality of the code comes from a study of gene mutations. Such mutations result in amino acid substitutions. Amino acid substitutions resulting from gene mutations are known for coat protein in tobacco mosaic virus (TMV), a chain of tryptophan synthetase in *E. coli* and haemoglobin in man. A change in a single base can account for nearly all amino acid substitutions. This proves the universality of the code.

The code has remained constant since the time it was fixed when complex bacteria evolved (about three billion years ago). Any mutation altering the code reading would change the reading of mRNA. This in turn may change the amino acid sequence of the proteins synthesized by the organism. As many of these changes could be lethal, there would be a strong selection pressure against such a mutation, hence the constancy of the code over a long period of time. Changes in proteins take place only with respect to the positions of particular amino acids. Moreover, only a few such changes take place at a time. Most mutations result in only a single amino acid substitution.

II) Biological expression of gene: Protein synthesis –Transcription and Translation processes



- **TRANCRIPTION (RNA Synthesis)**

Transcription is the process of creating a complementary RNA copy of a sequence of DNA. During transcription, a DNA sequence is read by RNA polymerase, which produces a complementary, antiparallel RNA strand. Transcription results in an RNA complement that includes uracil (U) in all instances where thymine (T) would have occurred in a DNA complement and instead of deoxy-ribose sugar, ribose sugar.

Transcription is the first step leading to gene expression. The part / stretch of DNA transcribed into an RNA molecule is called a transcription unit and encodes at least one gene. If the gene transcribed encodes a protein, the result of transcription is messenger RNA (mRNA), which will then be used to create that protein via the process of translation. Alternatively, the transcribed gene may encode for either ribosomal RNA (rRNA) or transfer RNA (tRNA), other components of the protein-assembly process.

DNA is read from 3' → 5' during transcription. Meanwhile, the complementary RNA is created from the 5' → 3' direction. DNA is arranged as two antiparallel strands in a double helix, only one of the two DNA strands, called the template strand, is used for transcription. This is because RNA is only single-stranded. The use of only the 3' → 5' strand eliminates the need for the Okazaki fragments seen in DNA replication.

Transcription can be explained easily in 4 or 5 steps, each moving like a wave along the DNA.

1. Helicase unwinds / unzips the DNA by breaking the hydrogen bonds between complementary nucleotides.
2. RNA nucleotides (A, G, C, and U) are paired with complementary DNA bases.
3. Ribose sugar-phosphate backbone forms with assistance from RNA polymerase.
4. Hydrogen bonds of the untwisted RNA+DNA helix break, freeing the newly synthesized RNA strand.
5. If the cell has a nucleus, the RNA is further processed (addition of a 3' poly-A tail and a 5' cap) and exits through nucleus to the cytoplasm through the nuclear pore complex.

❖ Structure of RNA polymerase (RNAP or RNApol) in prokaryotes

RNA polymerase is an enzyme that produces RNA. In cells, RNAP is needed for constructing RNA chains from DNA genes as templates, a process called transcription. In chemical terms, RNAP is a nucleotidyl transferase that polymerizes ribonucleotides at the 3' end of an RNA transcript.

RNAP is a relatively large molecule. The core enzyme has 5 subunits (**α2ββ'ω)** (~400 kDa) and complete holoenzyme has 6 subunits (**α2β β'ω σ)** (~480 kDa):

- **2 α (Alpha):** - The two α subunits assemble the enzyme and bind regulatory factors. Each subunit has two domains: α CTD (C-Terminal Domain) binds the UP element of the extended promoter, and α NTD (N-Terminal Domain) binds the rest of the polymerase. This subunit is not used on promoters without an UP element.
- 1 **β (Beta):** - This has the polymerase activity (catalyzes the synthesis of RNA), which includes chain initiation and elongation.
- **1 β' (Beta prime)**: - Binds to DNA (nonspecifically).
- **1 ω (Omega)**: Restores denatured RNA polymerase to its functional form in vitro. It has been observed to offer a protective/chaperone function to the β' subunit and to promote assembly.
- **1σ (Sigma)**: In order to bind promoter-specific regions, the holoenzyme requires another subunit, sigma (σ).The sigma factor greatly reduces the affinity of RNAP for nonspecific DNA while increasing specificity for certain promoter regions, that way, transcription is initiated at the right region.

There are many proteins that can bind to RNAP and modify its behavior. For instance, GreA and GreB from *E. coli* and in most other prokaryotes can enhance the ability of RNAP to cleave the RNA template near the growing end of the chain. This cleavage can rescue a stalled polymerase molecule, and is likely involved in proofreading the occasional mistakes made by RNAP. A separate cofactor, Mfd, is involved in transcription-coupled repair, the process in which RNAP recognizes damaged bases in the DNA template and recruits enzymes to restore the DNA. Other cofactors are known to play regulatory roles; i.e. they help RNAP choose whether or not to express certain genes.

❖ RNA polymerases in eukaryotes

Eukaryotes have several types of RNAP, characterized by the type of RNA they synthesize:

- **RNA polymerase I**: - Synthesizes a pre-rRNA 45S (35S in yeast), which matures into 28S, 18S and 5.8S rRNAs which will form the major RNA sections of the ribosome.
- **RNA polymerase II**: - Synthesizes precursors of mRNAs and most snRNA and microRNAs. This is the most studied type, and due to the high level of control required over transcription a range of transcription factors are required for its binding to promoters.
- **RNA polymerase III**: - Synthesizes tRNAs, rRNA 5S and other small RNAs found in the nucleus and cytosol.
- **RNA polymerase IV**: - Synthesizes siRNA in plants.
- **RNA polymerase V**: - Synthesizes RNAs involved in siRNA-directed heterochromatin formation in plants.

There are other RNA polymerase types in mitochondria and chloroplasts. And there is RNA-dependent RNA polymerases involved in RNA interference

❖ Transcription Process (Mechanism of transcription)

Transcription is divided into 5 stages in bacteria

1. Pre- initiation
2. Initiation
3. Promoter clearance
4. Elongation and
5. Termination

1. Pre-initiation

In eukaryotes for the initiation of transcription, RNA polymerase requires the presence of a **core promoter sequence** in the DNA. Promoters are regions of DNA that promote transcription and, in eukaryotes, are found at -30, -75, and -90 base pairs upstream from the **Transcription Start Site** (abbreviated to TSS). Core promoters are sequences within the promoter that are essential for transcription initiation. RNA polymerase is able to bind to core promoters in the presence of various specific transcription factors.

The most common type of core promoter in eukaryotes is a short DNA sequence known as a **TATA box** (Thymine Adenine box), found 25-30 base pairs upstream from the TSS. The TATA box, as a core promoter, is the binding site for a transcription factor known as TATA-binding protein (TBP), which is itself a subunit of another transcription factor, called Transcription Factor II D (TFIID). After TFIID binds to the TATA box via the TBP, five more **transcription factors** and RNA polymerase combine around the TATA box in a series of stages to form a pre-initiation complex. One transcription factor, DNA helicase, has helicase activity and so is involved in the separating of opposing strands of double-stranded DNA to provide access to a single-stranded DNA template. However, only a low or basal rate of transcription is driven by the pre-initiation complex alone. Other proteins known as activators and repressors, along with any associated co-activators or co-repressors are responsible for modulating transcription rate.

Thus, pre-initiation complex contains:

i) Core Promoter Sequence (TATA Box)

ii) Transcription Factors

iii) DNA Helicase

iv) RNA Polymerase

v) Activators and Repressors

2. Initiation

In bacteria, transcription begins with the binding of RNA polymerase to the promoter in DNA. RNA polymerase is a core enzyme consisting of five subunits: 2 α subunits, 1 β subunit, 1 β' subunit, and 1 ω subunit. At the start of initiation, the core enzyme is associated with a sigma factor to form

holoenzyme that aids in finding the appropriate -35 and -10 base pairs downstream of promoter sequences. Transcription initiation is more complex in eukaryotes. Eukaryotic RNA polymerase does not directly recognize the core promoter sequences. Instead, a collection of proteins called transcription factors mediate the binding of RNA polymerase and the initiation of transcription. Only after certain transcription factors are attached to the promoter region then the RNA polymerase binds to it. The completed assembly of transcription factors and RNA polymerase bind to the promoter, form a transcription initiation complex.

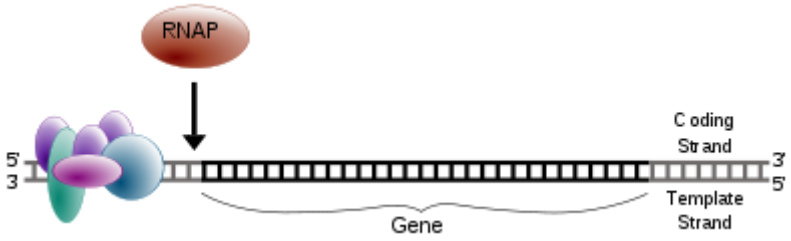


Diagram of transcription initiation. RNAP = RNA polymerase

3. Promoter clearance

After the first bond is synthesized, the RNA polymerase must clear the promoter. During this time there is a tendency to release the RNA transcript and produce truncated transcripts. This is called abortive initiation and is common for both eukaryotes and prokaryotes. Abortive initiation continues to occur until the σ factor rearranges, resulting in the transcription elongation complex (which gives a 35 bp moving footprint). The σ factor is released before 80 nucleotides of mRNA are synthesized. Once the transcript reaches approximately 23 nucleotides.

4. Elongation

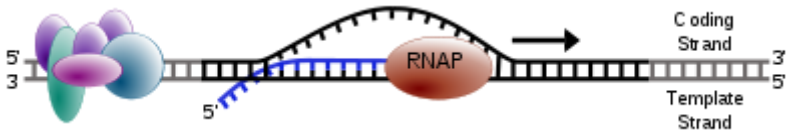


Diagram of transcription elongation

One strand of the DNA, the template strand (or noncoding strand), is used as a template for RNA synthesis. As transcription proceeds, RNA polymerase traverses the template strand and uses base pairing complementarily with the DNA template to create an RNA copy. Although RNA polymerase traverses the template strand from 3' → 5', the coding (non-template) strand and newly-

formed RNA can also be used as reference points, so transcription can be described as occurring 5' → 3'. This produces an RNA molecule from 5' → 3', an exact copy of the coding strand (except that thymines are replaced with uracils, and the nucleotides are composed of a ribose (5-carbon) sugar where DNA has deoxyribose (one less oxygen atom) in its sugar-phosphate backbone

Unlike DNA replication, mRNA transcription can involve multiple RNA polymerases on a single DNA template and multiple rounds of transcription (amplification of particular mRNA), so many mRNA molecules can be rapidly produced from a single copy of a gene.

Elongation also involves a proofreading mechanism that can replace incorrectly incorporated bases. In eukaryotes, this may correspond with short pauses during transcription that allow appropriate RNA editing factors to bind. These pauses may be intrinsic to the RNA polymerase or due to chromatin structure.

5. Termination

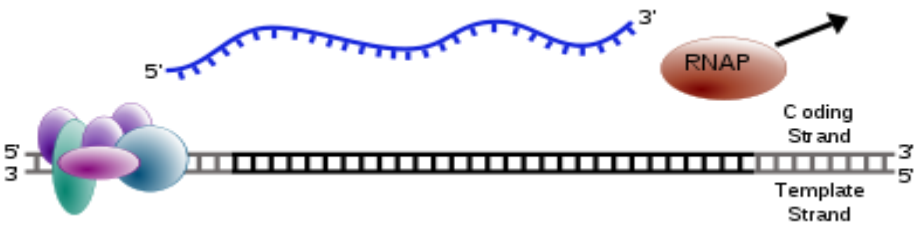


Diagram of transcription termination

Bacteria use two different strategies for transcription termination.

a) **Rho-independent transcription termination:** - RNA transcription stops when the newly synthesized RNA molecule forms a G-C-rich hairpin loop followed by a run of Us. When the hairpin forms, the mechanical stress breaks the weak rU-dA (U of RNA and A of DNA) bonds, now filling the DNA-RNA hybrid. This pulls the poly-U transcript out of the active site of the RNA polymerase, in effect, terminating transcription.
b) **Rho-dependent type of termination**: - a protein factor called "Rho" destabilizes the interaction between the template and the mRNA, thus releasing the newly synthesized mRNA from the elongation complex.

Transcription termination in eukaryotes is less understood but involves cleavage of the new transcript followed by template-independent addition of As at its new 3' end, in a process called polyadenylation

Post transcriptional modification

Prokaryotes have the ability to start translation as soon as the mRNA is free from the RNA polymerase. Eukaryotes on the other hand do modify their mRNAs (pre-mRNAs) before they leave the nucleus.

Post-transcriptional modification is a process in which primary transcript RNA is converted into mature RNA in eukaryotic cells. A notable example is the conversion of precursor messenger RNA (Pre-mRNA) into mature messenger RNA (mRNA), which includes splicing and occurs prior to protein synthesis. This process is important for the correct translation of the genomes of eukaryotes. The primary RNA transcript contains both **exons**, which are coding sections of the primary RNA transcript and **introns**, which are the non coding sections of the primary RNA transcript. In splicing introns are cut and exons are brought closer in mature m-RNA.

mRNA processing

The pre-mRNA molecule undergoes three main modifications which occur in the cell nucleus before the RNA is translated. These modifications are

1. **5' capping**
2. **3' polyadenylation and**
3. **RNA splicing**

1. 5’ Capping

Capping of the pre-mRNA involves the addition of **7-methylguanosine (m^7G)** to the 5' end. To achieve this, the terminal 5' phosphate is removed first, which is done with the help of a **phosphatase enzyme**. The enzyme **(guanine-N^7-methyltransferase** ("cap MTase") transfers a methyl group from S-adenosyl methionine to the guanine ring. Then enzyme **guanosyl transferase** then catalyses the reaction, which produces the diphosphate 5' end. The diphosphate 5' prime end then attacks the phosphorus atom of a GTP molecule in order to add the **guanine** residue in a 5'5' triphosphate link.

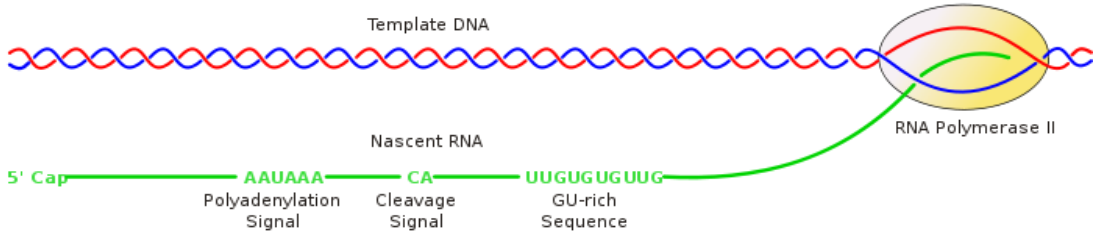
This type of cap, with just the (m^7G) in position is called a **cap 0 structure**. The ribose of the adjacent nucleotide may also be methylated to give a **cap 1**. Methylation of nucleotides downstream of the RNA molecule produce **cap 2**, **cap 3** structures and so on. In these cases the methyl groups are added to the 2' OH groups of the ribose sugar. The cap protects the 5' end of the primary RNA transcript from attack by ribonucleases.

2. Cleavage and 3’ Polyadenylation

The pre-mRNA processing at the 3' end of the RNA molecule involves cleavage of its 3' end and then the addition of about 200 adenine residues to form a poly (A) tail. The cleavage and adenylation reactions occur if a polyadenylation signal sequence first (5'- AAUAAA-3') second **(5'-CA-3') and third GU-rich sequence** is located near the 3' end of the pre-mRNA molecule. The second signal is the site of cleavage.

After the synthesis of the sequence elements, two multi subunit proteins called cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CStF) are transferred from RNA Polymerase II to the RNA molecule. These two factors bind to the sequence elements. A protein complex forms that contains additional cleavage factors and the enzyme Polyadenylate Polymerase (PAP). This complex cleaves the RNA between the polyadenylation sequence and the GU-rich sequence at the cleavage site marked by the (5'-CA-3') sequences.

Poly (A) polymerase then adds about 200 adenine units to the new 3' end of the RNA molecule using ATP as a precursor. As the poly (A) tails is synthesized, it binds multiple copies of poly (A) binding protein, which protects the 3'end from ribonuclease digestion.



3. RNA Splicing

RNA splicing is the process by which introns (regions of RNA that do not code for protein) are removed from the pre-mRNA and the remaining exons are connected to re-form a single continuous molecule. Most RNA splicing occurs after the complete synthesis and end-capping of the pre-mRNA. Transcripts with many exons can be spliced co-transcriptionally. The splicing reaction is catalyzed by a large protein complex called the **spliceosome** assembled from proteins and small nuclear RNA molecules that recognize splice sites in the pre-mRNA sequence. Many pre-mRNAs, including those encoding antibodies, can be spliced in multiple ways to produce different mature mRNAs that encode different protein sequences. This process is known as alternative splicing, and allows production of a large variety of proteins from a limited amount of DNA.

==========

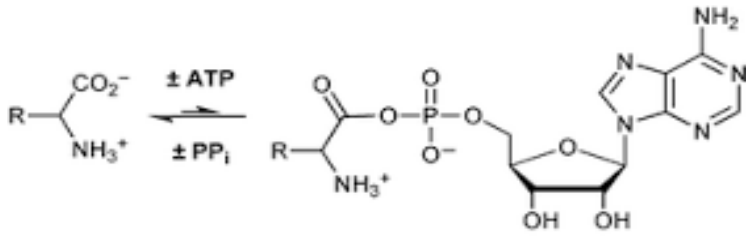
- **Translation (Protein Synthesis)**

During the translation process protein are made by the ribosomes on mRNA strands. The main steps in translation are-

A) Activation of amino acids
B) Transfer of the amino acid to t RNA
C) Initiation of synthesis
D) Elongation of polypeptide chain
E) Chain termination

A) Activation of amino acid

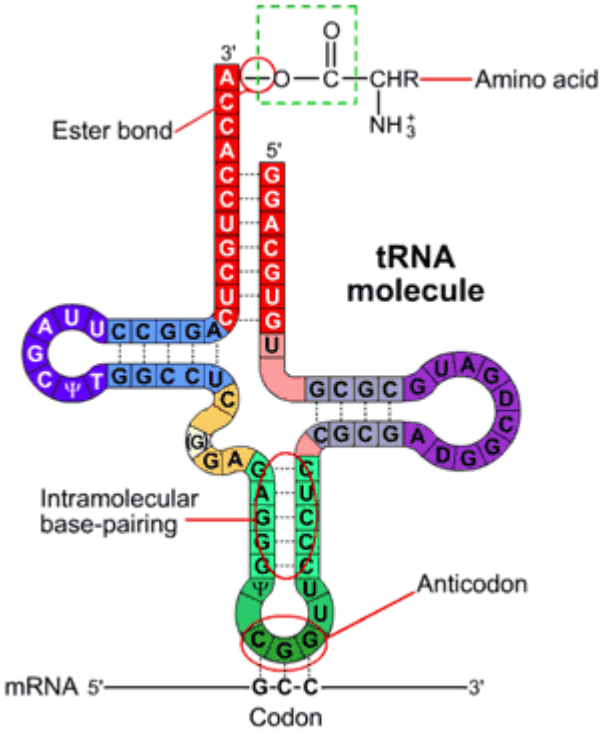
The first step in translation is the activation of amino acids. Only the L-amino acids take part in protein synthesis. The activation of amino acids takes place through their carboxyl groups. Each amino acid is catalyzed by its own specific activating enzyme called **aminoacyl t-RNA synthatase** to form an **amino acyl adenylate (aaa)** in presence of ATP. A high-energy ester bond is formed between the α-phosphate of ATP (first phosphate) & carboxyl group of the amino acid. The β & γ phosphate (second and third phosphate) of ATP break away as inorganic pyrophosphate (ppi). The amino acyl AMP remains bound to the activating enzyme.



AA-AMP

B) Transfer of amino acid to t-RNA

The transfer of the activated amino acid to t-RNA is specific. The t-RNA is named after the amino acid for which it is specific. The activated amino acid is transferred to its specific t-RNA. A high-energy ester bond is formed between the carboxyl group of amino acid & 3'-OH group of t-RNA. The amino acyl AMP attached to amino-acyl t-RNA synthatase reacts with a specific t-RNA to form an amino acyl t-RNA complex.



C) Initiation of synthesis

Polymerization of activated amino acid into proteins is mediated by 70 S ribosomes in prokaryotes and by 80S ribosomes in eucaryotes. The 70S ribosomes dissociates reversibly into a 50S & 30S subunits and 80S ribosome dissociates into 60S and 40S subunits when the concentration of **Mg^{+2}** ions in the suspending buffer is lowered from 10^{-2} to 10^{-4}. The large subunit (50S) contains one molecule each of 23S & 5S r-RNA associated with 35 different proteins. The small subunit (30S) contains a single molecule of 16S r-RNA & 21 different proteins.

This step consists of

1. **Initiation factors**
2. **Formylation of methionine**
3. **Formation of the 30s initiation complex**
4. **Binding of F met t-RNA with m-RNA 30s complex**
5. **Attachment of 50s subunit to form complete initiation complex**

1) Initiation factors: -

In prokaryotes three factors IF-1, IF-2, IF-3, are required for initiation of protein synthesis. These factors are found in the 30s subunit of the ribosome. The IF-1 & IF-2 are required for the binding of initiation t-RNA (f Met-t RNA) to the 30s subunit. IF-2 is involved in the binding of GTP & contains the SH groups necessary for this binding. IF-3 (a protein rich in lysine

& Arginine bases) is required for the binding of the 30s ribosomal subunit to the initiation sequence of m-RNA.

In Eukaryotes IF-1 is absent. eIF-2, eIF-2', eIF-$2a_1$, eIF-$2g_2$, eIF-$2g_3$, eIF-3 are present. eIF-2 and eIF-3 forms a complex with met t-RNA & GTP. This complex reacts with 40s subunits. eIF-2 promotes the AUG dependant binding of met t-RNA to 40s subunits in absence of eIF-2a. The accessory factors eIF $2a_2$ & eIF-$2a_3$ are required for binding of met t-RNA along with GTP. eIF-3 is required for binding of 40s unit of ribosome with m-RNA.

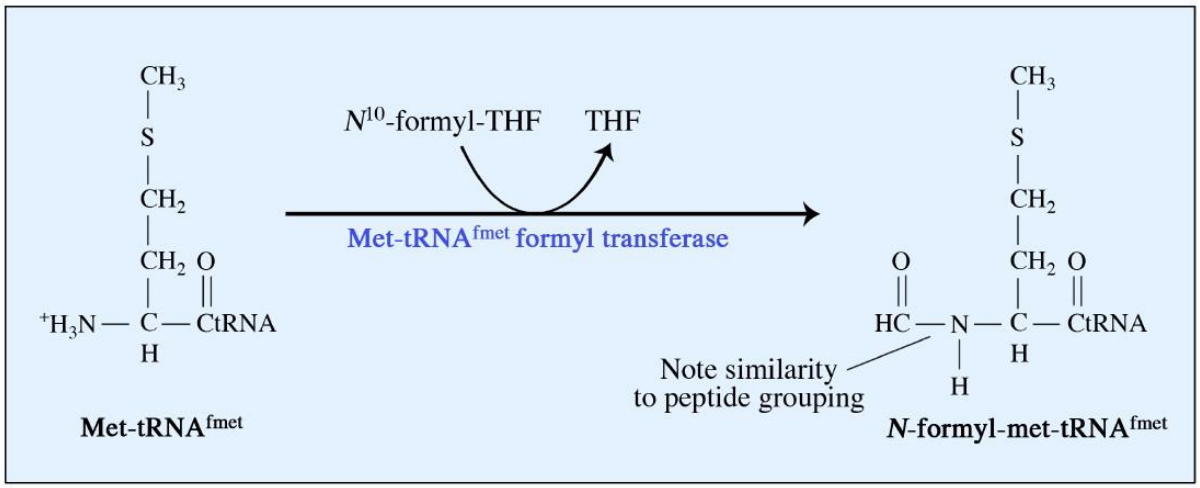
2) Formylation of methionine: -

In Eukaryotes the starting N terminal amino acid is methionine. In prokaryotes methionine carries formyl group (-CHO) and called as N-formyl methionine.

In Eukaryotes the initiation t-RNA forms complex with methionine. This t-RNA is called as met t-RNA.

In prokaryotes a formyl group is added to amino group of methionine to form N-formyl methionyl t-RNA or F met t-RNA.

This reaction is catalyzed by transformylase enzyme.



3. Formation of the 30s initiation complex: -

The first step in protein synthesis is the formation of the 30s initiation complex.

This complex consists of

- m-RNA
- 30s ribosomal subunit.

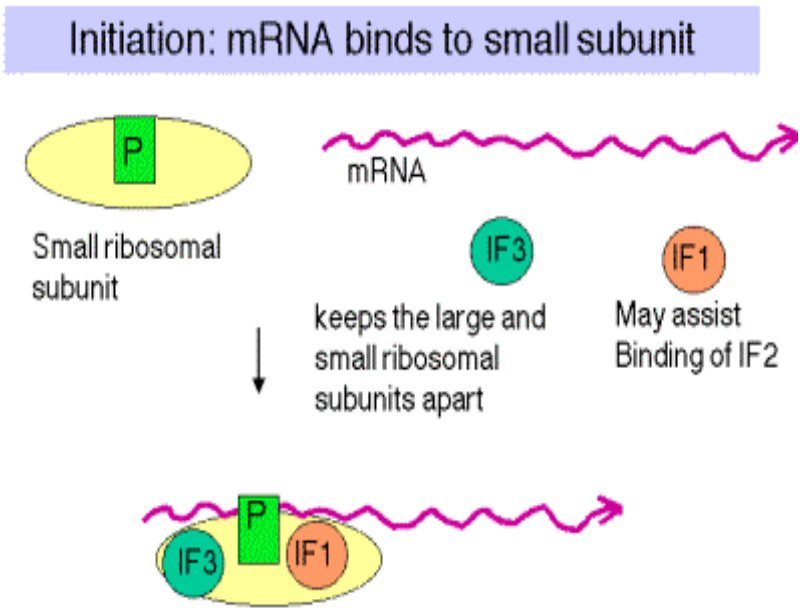
- Initiator t-RNA with attached amino acid N-formyl methionine (N-formylmetheonine t-RNA).
- GTP
- Three initiation factors (IF-1, IF-2, IF-3) in prokaryotes and (EE-2, EF-3) eucaryotes.

In formation of m-RNA–30s subunits complex, m-RNA becomes attached to 30s subunit to form m-RNA 30s complex. This process requires the initiation factor IF-3. Only one molecule of IF-3 binds to each 30s subunit.

m-RNA + 30s subunit + IF-3 ⟶ m RNA-30s complex

Segments of 16s r-RNA in 30s subunit provides a binding site for IF-3. Shine and dalgarno have shown that a nucleotide sequence near the 3' end of 16s r-RNA base pairs directly with a complementary sequence in m-RNA. m-RNA molecules contain a sequence of 3 to 7 purine nucleotides proceeding the initiation codon AUG. This polypurine sequence on m-RNA pairs with a complementary polypyrimidine sequence near the 3’ end of 16s r-RNA. It has been suggested that hydrogen bonding between the complementary regions of 16s r-RNA & m-RNA comprises part of the recognition sequence for chain initiation.

IF-3 has several functions i) it is essential for binding of m-RNA to the 30s subunit. ii) It affects 30s subunit conformation. iii) It prevents re-association of dissociated 30s & 50s subunits. This is important otherwise there would be interference in the formation of 30s initiation complex.



4) Binding of F met -tRNA with m-RNA 30s complex: -

In prokaryotes the binding of the F-met t-RNA complex with m-RNA 30s subunit complex requires the IF-2 & IF-1 and GTP. IF-2 is required for recognition and binding of F met t-RNA.

The F-met t-RNA binds with its UAC anticodon to AUG codon of m-RNA or rarely to GUG or UUG. These codons are called as initiations codons.

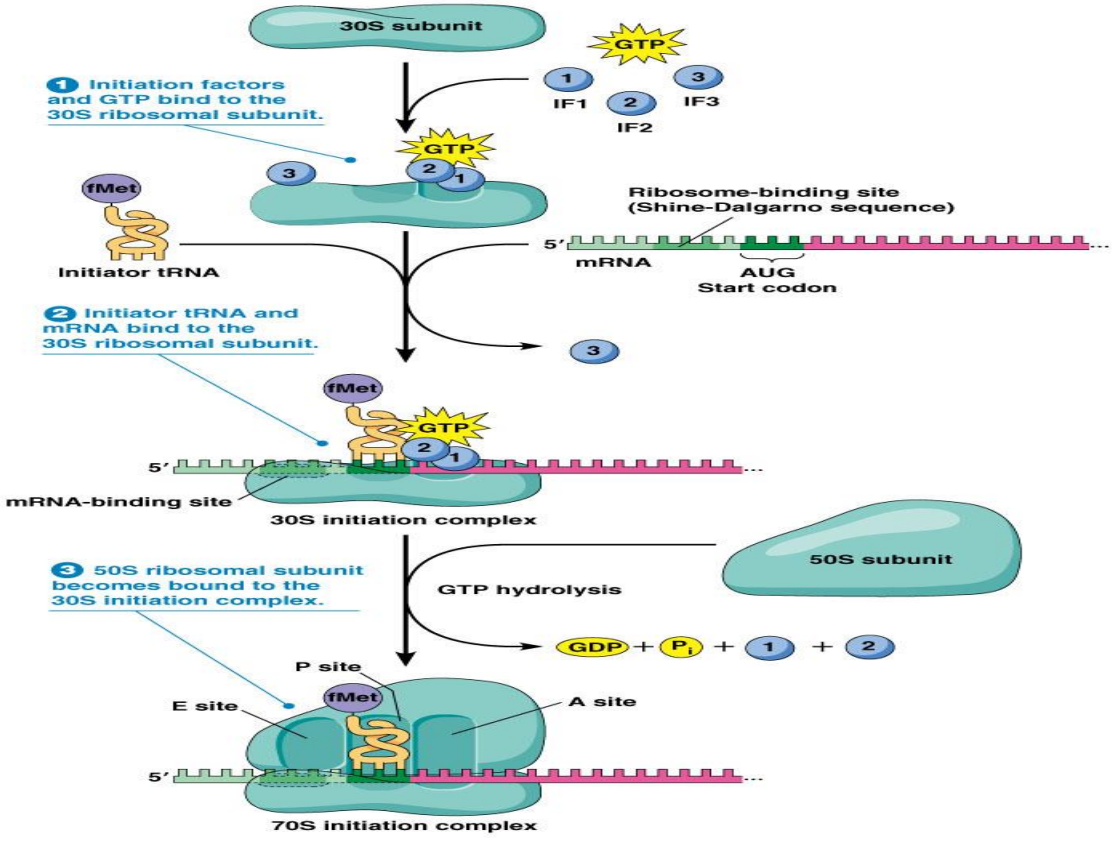
In Eukaryotes the t-RNA binds to the 40s ribosomal subunit first and then the m-RNA attaches the AUG initiation codon.

The m-RNA strand has many AUG codons, only one of these functions as initiation codon. What factors determine that which AUG will function as the initiation codon? One possibility is that conformation of t-RNA and the second is a specific sequence of nucleotides near AUG is required for it to function as an initiator.

The 70s ribosome has two binding sites for t-RNA, the peptidyl or polymerisation site (P site) and the aminoacyl or acceptor (A site). Each ribosome thus functionally accommodates two codons at a time. The P & A sites are located in the 50s subunit. The F-met t-RNA binds P site, all other t-RNAs first bind to A site and then shift to P site.

5) Attachment of 50s subunit to from complete initiation complex: -

During the formation of 70s initiation complex GTP is converted GDP+Pi and the initiation factors IF-1 & IF-2 are released.



D) Elongation of polypeptide chain

This step consists of

1. **Elongation factors**
2. **Binding of aminoacyl t-RNA to the A site**
3. **Peptide bond formation**
4. **Translocation**

1. Elongation factors: -

The two factors isolated initially were called EF-T & EF-G. EF-T consists of two proteins which were called EF-Tu (temperature unstable) and EF-Ts (temperature stable). T-refers to transferase activity, EF-Tu & EF-Ts are required for binding the aminoacyl t-RNA to the ribosome. EF-G is involved in translocation of m-RNA.

In Eukaryotes EF-1 & EF-2 are present which correspond to EF-T and EF-G of prokaryotes.

2. Binding of aminoacyl t-RNA to the A site: -

In prokaryotes EF-Tu & EF-Ts are required for binding of aminoacyl t-RNA to ribosome. EF-Tu forms a ternary complex with aminoacyl t-RNA & GTP. The formation of complex is catalysed by EF-Ts.

aa t-RNA + EF-Tu + GTP $\xrightarrow{\text{EF-Ts}}$ (aa-t-RNA. EF-Tu. GTP)
Ternary complex

Transfer of aa t-RNA from the ternary complex to the A site of the ribosome now takes place.

(aa-t-RNA. EF-Tu. GTP) + Ribosome $\longrightarrow$ (aa t-RNA ribosome)
+ (EF-Tu GDP)+ Pi

Recycling of GDP now takes place from the (EF-Tu GDP) complex.

(EF-Tu-GDP) + GTP $\longrightarrow$ (EF-tu-GTP) + GDP

(EF-Tu-GTP) now reacts with another aa t-RNA. The net result is the aa t-RNA enters the A site.

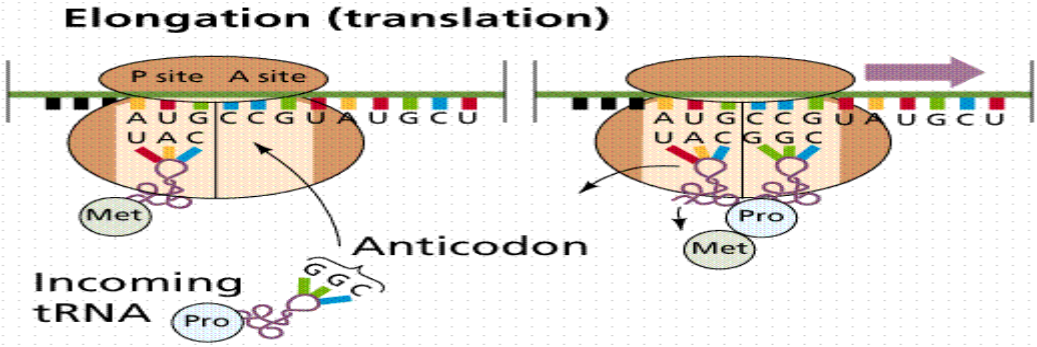


Fig. A

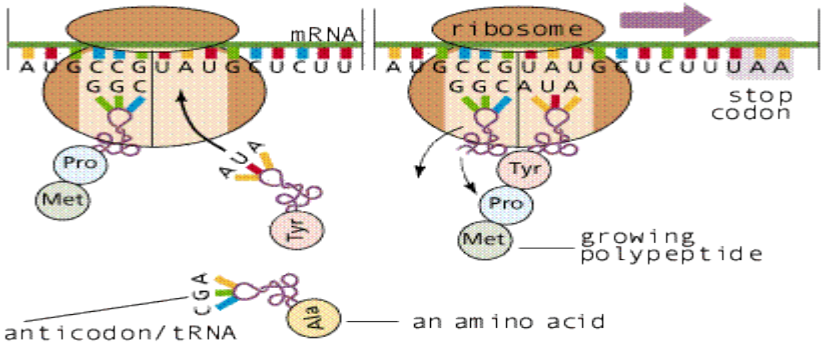


Fig. B

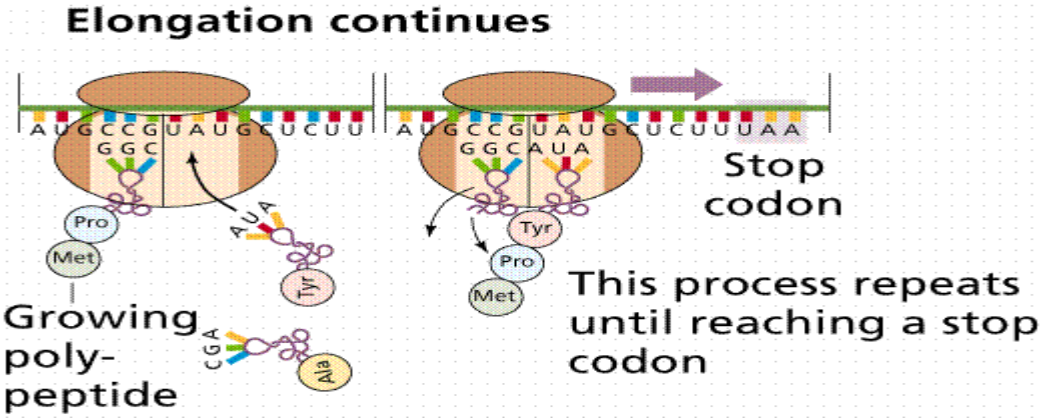


Fig. C

A) F met- tRNA at P sites & Pro. t-RNA at A site
B) Pro.t-RNA translocated from A to P sites. The enzyme peptydal transferase catalyses peptide bond formation between F met and Pro. t-RNA of F met is given out of ribosome. Tyr- t-RNA Occupied vacant A site.
c) Tyr- t-RNA translocated to P site. Tyr- t-RNA is given out of ribosome. Ala- t-RNA Occupies vacant A site.

3. **Peptide bond formation: -**

The starting amino acid unites by peptide bond with second amino acid (aa_2). The t-RNA molecules of f-Met t-RNA or Met t-RNA are discharged

from the P site. The second amino acid (aa_2) remains attached to its t-RNA through carboxyl group. Attached to it is methionine /N-formylmetheonine by means of the peptide bond.

Peptide bond formation does not require any external source of energy. It is catalysed by peptidyl transferase complex located in the large submit of Ribosome.

The t-RNA molecule to which the growing polypeptide chain is attached is called peptidyl t-RNA. It is bound to P site. The incoming t-RNA molecule bearing an amino acid molecule is called as aminoacyl t-RNA. It binds to A site. During the formation of peptide bond there is transfer of the polypeptide chain from peptidyl t-RNA to aminoacyl t-RNA. The carboxyl esters link of peptidyl t-RNA is broken and new peptide bond is formed with the amino acid of amino acid t-RNA.

The α-amino group of one amino acid is bonded to the α-carboxyl group of the other. During this process there is elimination of water.

4. **Translocation: -**

The movement of the ribosome relative to mRNA is called translocation. During the translocation the ribosomes moves along the m-RNA strand by a codon at a time. The movement of ribosome relative to m-RNA is in the 5' to 3' direction.

During the first translocation movement the aa_2-tRNA complex with attached to methionine /N-formylmetheonine, Shift from A site to P site. The A site now occupied by the aa_3-tRNA complex. This processes repeated & elongation of the polypeptide chain is thus brought about by step-by-step addition of amino acids.

In prokaryotes translocation is brought by elongation factor. G (EF-G) or translocase. EF-G rapidly hydrolyses GTP to GDP and Pi. Energy released during hydrolysis of GTP is utilised for translocation and unloading of deacylated t-RNA from P site.

TRANSLOCATION REACTION: -

Three reactions are involved in translocation-

a) Movement of acylated t-RNA from A site to the P site. It requires EFG & GTP.
b) Movement of m-RNA relative to the ribosome it also requires EF-G & GTP.
c) Release of deacylated t-RNA from the P site of the Ribosome.

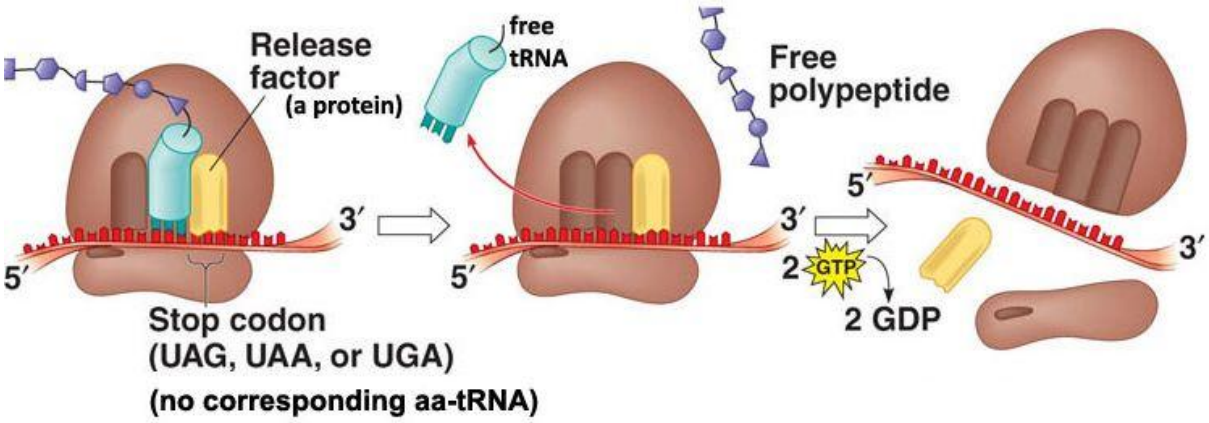
E). Chain termination: -

Elongation of the polypeptide chain continuous until a termination codon on m-RNA is reached. The termination codon may be UAA (ochre) UAG

(amber) or UGA (opal or umber). These codons have been called nonsense codons because ordinarily no t-RNA anticodon pairs with them. The termination codon provides signal to the ribosomes for attachment of release factors.

In prokaryotes there are three release factors EF_1 EF_2 EF_3. The release factors interact with peptidyl transferase causing hydrolysis of the polypeptide chain at the P site. The residual t-RNA is discharged from P site. In eukaryotes single release factor (EF) recognises all the terminal codons.

The first step in chain termination is the binding of release factor to the A site containing a termination codon. This requires GTP & activates peptidyl transferase system. Hydrolysis of peptidyl t-RNA at P site Results in the release of the polypeptide chain from ribosome. The released RF is recycled.



<u>Processing of polypeptide chain</u>: -

Starting amino acid in polypeptide chain is either N formylmetheonine or methionine. Many microorganisms synthesise enzyme which catalyse cleavage of formyl residue by deformylase or of methionine by an amino peptidase from the polypeptide chain.

Rate of protein synthesis: -

In *E coli* transcription of enzyme controlling tryptophan synthesis takes place at the rate of about 28 nucleotides per second. Protein is translated at the rate of about 7 amino acids per second.

==========

III) REGULATION OF GENE EXPRESION

For a bacterium to function, it is not necessary that all of its genes are transcribed at all times. To conserve energy and resources bacteria regulate the activity of their genes so that only those gene products necessary for the cell's functions are produced. For example, it would be wasteful for a bacterium to produce enzyme requires to synthesize an amino acid that was already available to it from its environment. Regulation of gene expression allows bacteria to respond to changes in their environment, typically to the presence or absence of nutrients.

Bacteria regulate expression for their genes in order to control the amount of gene product present. The steady state concentration of a gene product is determined by the balance between the rate of synthesis and the rate of degradation of the expressed protein. In practice, changes in the rate of synthesis are what alter the amount of gene product. The rate of synthesis could potentially be altered by a numbers of factors:

- Changes in the rate of gene transcription.
- Changes in the mRNA turnover time.
- Changes in the rate of translation.

In practice, all three mechanisms probably influence gene expression but the best understood examples are those involving the regulation of gene transcription.

ORGANIZATION OF THE BACTERIAL GENES

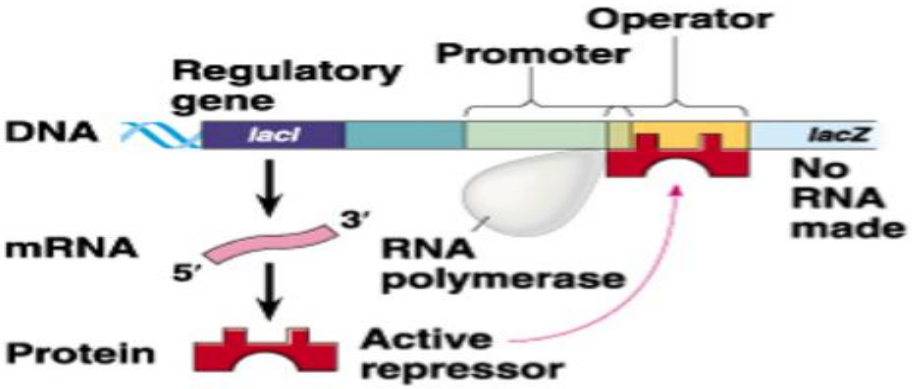
An important feature, which determines how gene transcription is regulated in bacteria, is the organization of genes as **operons**. These are transcriptional units in which several genes, usually encoding protein with related functions, are regulated together. Other genes also occur which encode regulatory proteins that control gene expression in operons. Many different operons have been identified in *E. coli.* Most contain genes that encode proteins involved in the biosynthesis of amino acids or the metabolism of nutrients. Operons are classified as inducible or repressible.

Inducible operons contain genes that encode enzymes involved in **metabolic pathways**. Expression of the genes is controlled by a substrate of the pathway. An example of an inducible operon is the ***lac operon*** and **ara operon** which encodes enzymes required for the metabolism of lactose and arabinose.

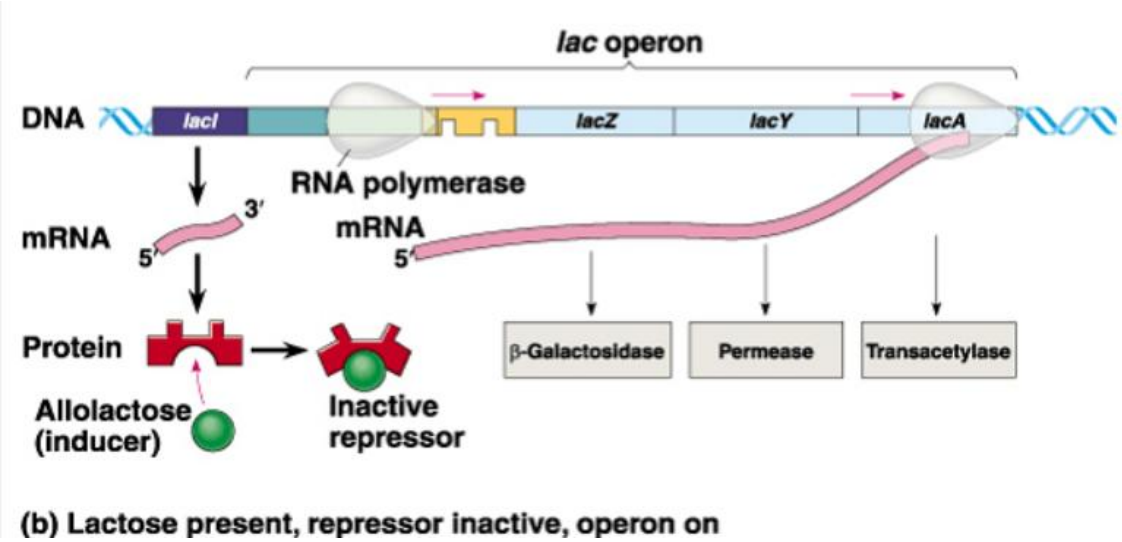
Repressible operons contain genes that encode enzymes involved in **biosynthetic pathways** and gene expression is controlled by the end product of the pathway which may repress expression of the operon or control it by an alternative mechanism called **attenuation**. An example of repressible operon is the ***trp operon***, which encodes enzymes involved in the biosynthesis of tryptophan.

<u>Lac operon</u>: -

This operon contains three gene-encoding enzymes required by the *E. coli* bacterium for the utilization of the disaccharide sugar lactose. These are ***lactose permease***, which transports lactose into the cell, ***β-galactosidase***, which hydrolyses lactose into its component sugar (glucose and galactose) and ***β-galactoside transacetylase***, which also involved in the hydrolysis of lactose. These enzymes are normally present in *E. coli* at very low levels but in the presence of lactose their levels rise rapidly. The three genes in the *lac* operon are known as *lac Z Y* and *A* which encode β-galactosidase, lactose permease and β-galactoside transacetylase respectively. The genes are sequential and are transcribed as a single mRNA from a single promoter. Another regulatory gene, *lac I,* which is expressed separately, lies upstream of the operon and encodes protein called the ***lac repressor*** which regulates the expression of the *lac Z, Y and A* genes. In the absence of lactose the lac repressor binds to a DNA sequence called the **operator** positioned between the lac promoter and the beginning of the *lac Z* genes. When bound to the operator, the lac repressor blocks the path of the RNA polymerase bound to the lac promoter upstream of it and prevents transcription of the *lac* genes. When the cell encounters lactose a few molecules of the *lac* enzymes present in the cell allow lactose to be taken up and metabolised. **Allolactose,** an isomer of lactose produced as an intermediate during the metabolism of lactose, acts as an ***inducer***. It binds to the lactose repressor and changes its conformation such that it can no longer bind to the operator. The path of the RNA polymerase is no longer blocked and the operon is transcribed. Large numbers of enzyme molecules are produced which take up lactose and metabolise it. The presence of lactose thus induces the expression of the enzymes needed to metabolise it. When the lactose is used up the *lac* repressor returns to its original conformation and again, binds the *lac* operator preventing transcription and switching off the operon (Fig.).



(a) Lactose absent, repressor active, operon off



(b) Lactose present, repressor inactive, operon on

CATABOLITE REPRESSION: -

This term describes an additional regulatory mechanism, which allows lac operon to sense the presence of glucose, an alternative and preferred energy, source to lactose. If glucose and lactose are both present, cells will use up the glucose first and will not expend energy splitting lactose into its components sugars. The presence of glucose in the cell switches off lac operon by a mechanism called **catabolite repression, which** involves a regulatory protein called the catabolite activator protein (CAP). CAP binds to a DNA sequence upstream of the lac promoter and enhances binding of the RNA polymerase leading to enhanced transcription of the operon. However CAP only binds in the presence of a derivative of ATP called cyclic adenosine monophosphate (cAMP) whose levels are influenced by glucose. The enzyme **adenylate cyclase** catalyses the formation of cAMP and is inhibited by glucose. When glucose is available to the cell adenylate cyclase is inhibited and cAMP levels are low. Under these conditions CAP does not bind upstream of the promoter and the lac operon is transcribed at a very low level.

Conversely, when glucose is low, adenylate cyclase is not inhibited; cAMP is higher and CAP binds increasing the level of transcription from the operon. If glucose and lactose are present together the lac operon will only transcribed at low level. However when the glucose is used up catabolite repression will end and transcription from the lac operon increases allowing the available lactose to be used up.

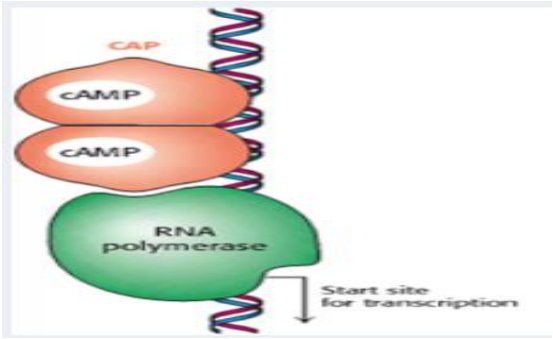
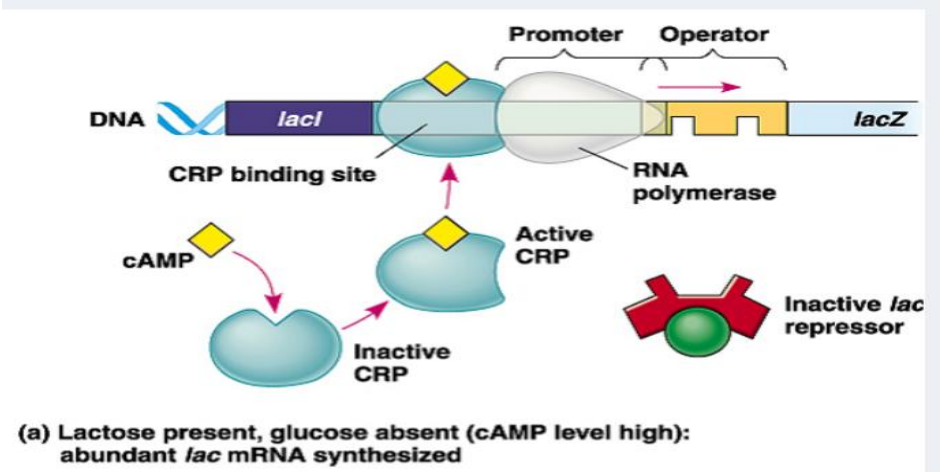


Fig. The CAP binding site on DNA is adjacent to the position at which RNA polymerase binds



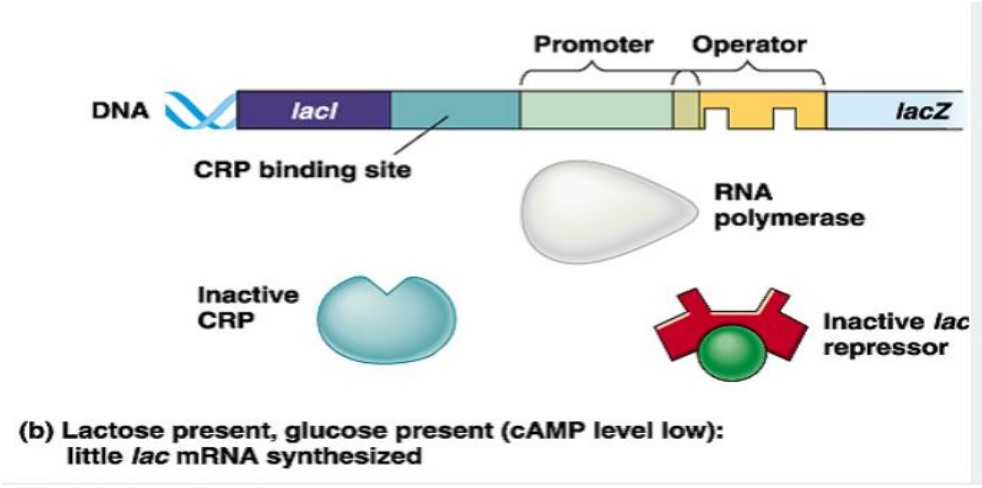


Fig. Catabolite repression and the role of CAP-cAMP complex

- **Arabinose operon (ara operon or araBAD operon)**

The L-arabinose operon, also called the *ara* or araBAD operon, is an **operon** that encodes **enzymes** needed for the **catabolism** of **arabinose** in ***Escherichia coli***. Arabinose is a five carbon sugar that can be used by *E. coli* as an alternative carbon source. The enzymes necessary for the metabolism of arabinose are coded for by the arabinose operon. The arabinose operon has a complex regulatory system. It was studied and explained by a scientist, Ellis Englesberg.

Arabinose operon can be regulated both positively and negatively in a similar manner as the lactose operon. Therefore the arabinose operon is also an inducible operon. In *E. coli* cells growing in the absence of arabinose, the three different enzymes involved in its metabolism are present in the cell in very small amounts and there is no expression of the operon. This is an adaptive mechanism that ensures that these enzymes needed to catabolize arabinose are only produced in sufficient amounts when arabinose is present in the environment. The arabinose operon also exhibits catabolite repression. A cAMP-CAP complex must be formed in order for the positive expression of the arabinose operon to occur. High levels of glucose in the environment will repress the arabinose operon due to low levels of the cAMP molecule. This is similar to the conditions necessary for lactose to be utilized as a carbon source. The arabinose operon will only express its genes if arabinose is the best carbon source present in the environment.

Structure and Mechanism

The arabinose operon consists of

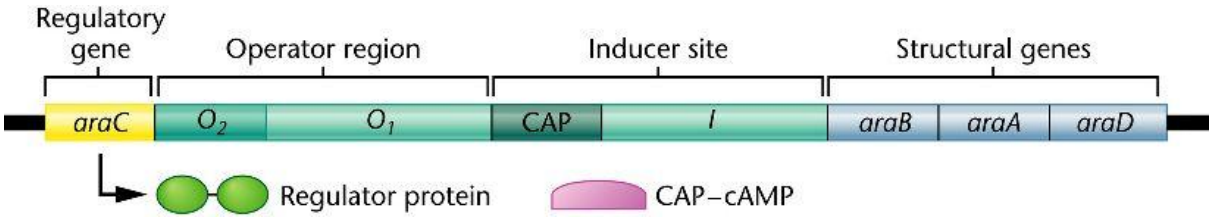
- three structural genes, B, A and D that code for the catabolic enzymes:

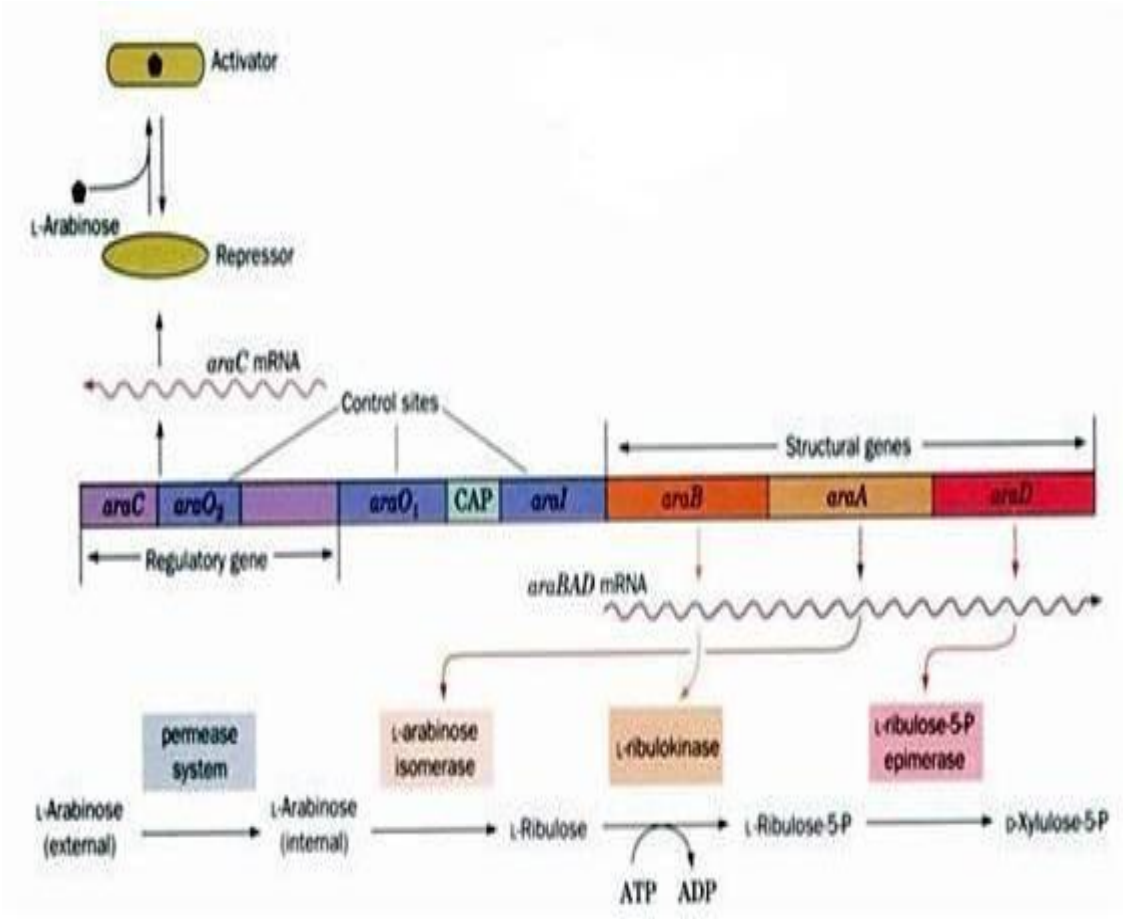
araB gene – produces Ribulokinase - phosphorylates ribulose

araA gene- produces Arabinose isomerase - coverts arabinose to ribulose

araD gene– produces Ribulose 5 phosphate epimerase - converts ribulose-5-phosphate to xylulose-5-phosphate which can then be metabolized via the pentose phosphate pathway

- **C gene**- a conventional gene that produces a protein product that combines with arabinose and acts positively to "turn on" the arabinose operon.
- Two promoter sites, Pc and Pbad
- Two operator sites, *O1* and *O2*
- Inducer site *ara I*





❖ ***araO₁*** is an operator site. **AraC** binds to this site and represses its own transcription from the $\mathbf{P_C}$ promoter. In the presence of arabinose, however, **AraC** bound at this site helps to activate expression of the $\mathbf{P_{BAD}}$ promoter.

- ***araO$_2$*** is also an operator site. **AraC** bound at this site can simultaneously bind to the ***araI*** site to repress transcription from the **P_{BAD}** promoter.

- ***araI*** is also the inducer site. **AraC** bound at this site can simultaneously bind to the ***araO$_2$*** site to repress transcription from the **P_{BAD}** promoter. In the presence of arabinose, however, **AraC** bound at this site helps to activate expression of the **P_{BAD}** promoter.

- **CAP (Catabolite Activator Protein)** binds to the **CAP** binding site. It does not directly assist RNA polymerase to bind to the promoter in this case. Instead, in the presence of arabinose, it promotes the rearrangement of **AraC** when arabinose is present from a state in which it represses transcription of the **P_{BAD}** promoter to one in which it activates transcription of the **P_{BAD}** promoter.

- **Regulation of the arabinose operon**

It is, clearly, much more complex than the lactose operon.

✓ **When arabinose is absent**

There is no need to express the structural genes. **AraC** does this by binding simultaneously to ***araI*** and ***araO$_2$***. As a result the intervening DNA is **looped**. These two events block access to the **P_{BAD}** promoter which is, in any case, a very weak promoter (unlike the ***lac*** promoter):

AraC also prevents its own expression. Thus, it is an autoregulator of its own expression. This makes sense; there is no need to over-express **AraC**. If the concentration falls too low then transcription of ***araC*** resumes until the amount of **AraC** is sufficient to prevent more transcription again.

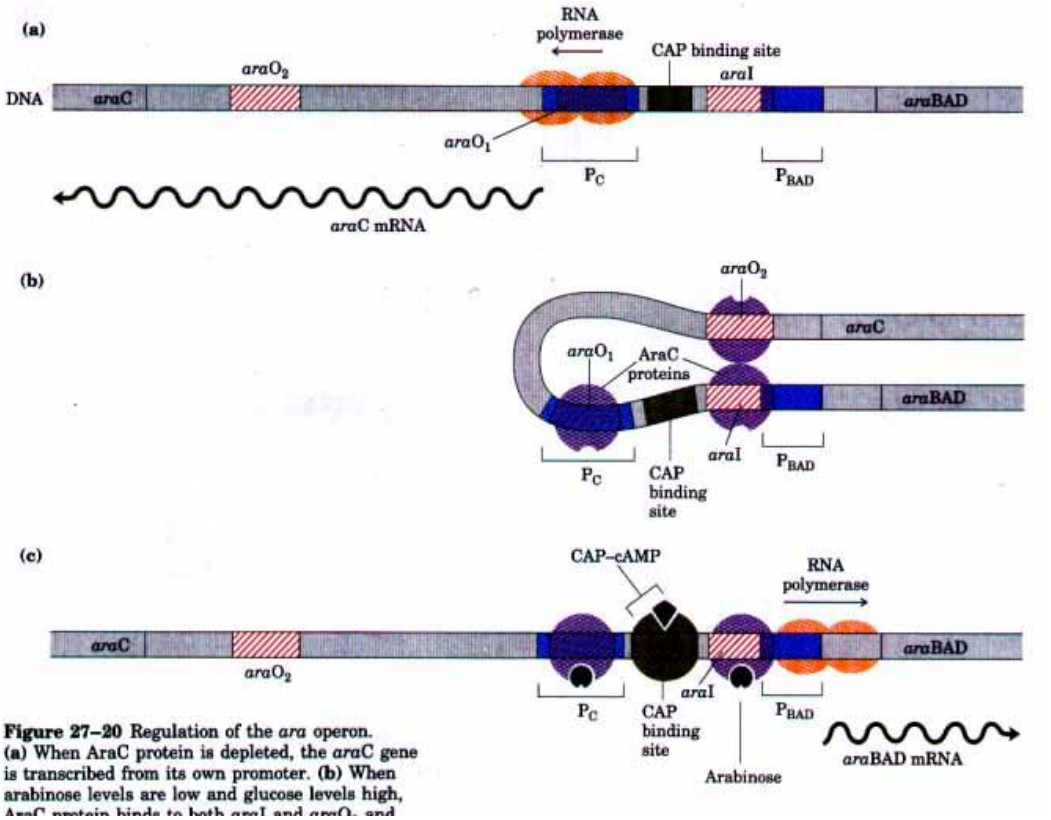
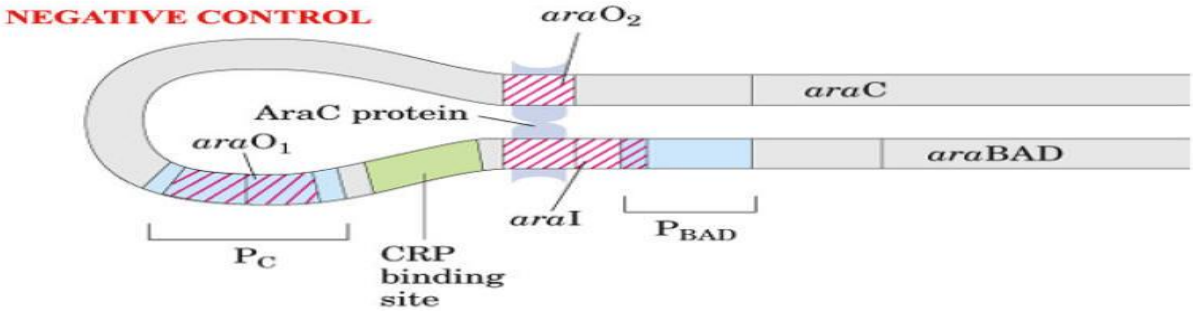


Figure 27–20 Regulation of the *ara* operon. **(a)** When AraC protein is depleted, the *ara*C gene is transcribed from its own promoter. **(b)** When arabinose levels are low and glucose levels high, AraC protein binds to both *araI* and *araO*$_2$ and



✓ **When arabinose is present**

It binds to **AraC** and allosterically induces it to bind to ***araI*** instead ***araO$_2$***. If **glucose** is also absent, then the presence of **CAP** bound to its site between ***araO$_1$*** and ***araI*** helps to break the DNA loop and also helps **AraC** to bind to ***araI***:

The ***ara*** operon demonstrates both negative and positive control. It shows a different function for **CAP**. It also shows how a protein can act as a switch with its activity being radically altered upon the binding of a small molecule.

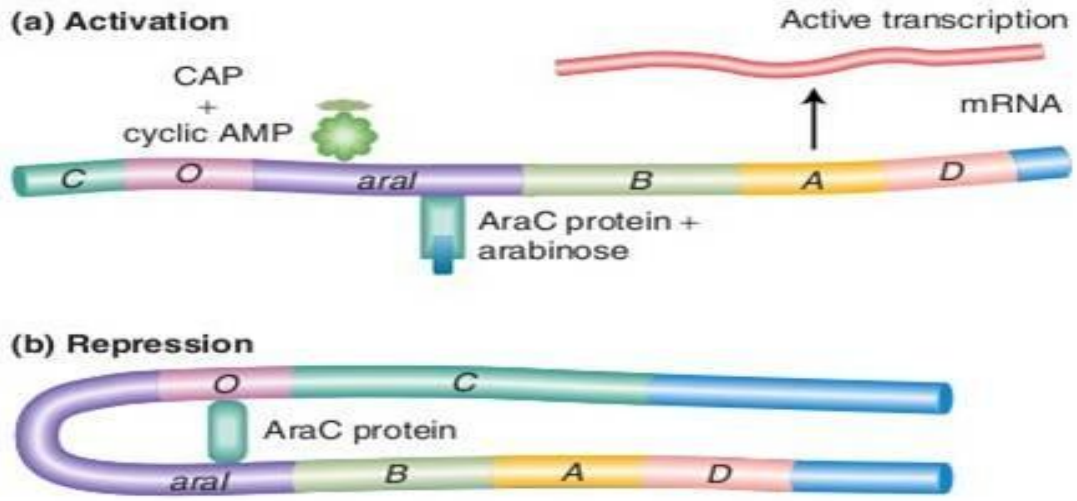


Figure **Dual control of the *ara* operon.** (a) In the presence of arabinose, the AraC protein binds to the *araI* region. The CAP–cAMP complex binds to a site adjacent to *araI*. This binding stimulates the transcription of the *araB*, *araA*, and *araD* genes. (b) In the absence of arabinose, the AraC protein binds to both the *araI* and the *araO* regions, forming a DNA loop. This binding prevents transcription of the *ara* operon.