SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

**SCHOOL OF BIO AND CHEMICAL ENGINEERING**

**DEPARTMENT OF BIOTECHNOLOGY**

# SBC2101 – MOLECULAR BIOLOGY I

# Sathyabama Institute of Science and Technology
# B.Sc BIOCHEMISTRY
# SBC2101: MOLECULAR BIOLOGY I

| | | |
|---|---|---|
| **Teaching Scheme:**<br>**TH: 04 Hours/Week** | **Credit**<br>**03** | **Examination Scheme:**<br>**In-Sem (Paper): 50 Marks**<br>**End-Sem (Paper): 50 Marks** |

Molecular Biology I is an introductory course deals with the fundamental principles of DNA structure, organization and its function in both prokaryotes and eukaryotes.

**Learning outcomes;**

- Discuss the most significant discoveries and theories through the historical progress of biological scientific discoveries, and their impacts on the development of molecular biology.
- Describe the general principles of gene organization in both prokaryotic and eukaryotic organisms.
- Elucidate the mechanisms of DNA replication, RNA synthesis and processing, and protein synthesis.
- Compare and contrast the fundamental structure, properties and processes in which nucleic acids play a part between prokaryotes and eukaryotes
- Illustrate how gene expression is regulated at the transcriptional and post-transcriptional level
- Demonstrate the ability to effectively communicate and work collaboratively with peers in the small group to successfully address problems sets in molecular biology.

**Unit I Basic concepts in Molecular Biology**

History of significant discoveries in Molecular Biology. Central dogma – DNA as genetic material. Evidence to prove DNA as genetic material- Griffth's transforming principle; Avery,Mc Leod and McCarty's experiment; Hershey and Chase experiment. Chargaff's law. Watson and Crick model of DNA (B form). Other different forms of DNA- A,Z . Physical Properties of DNA- Denaturation , Renaturation , hyper and hypochromic effect, Cot curve significance.

**Unit II DNA Replication**

Meselson-Stahl's experiment to prove semi-conservative model of replication. DNA replication in prokaryotes and eukaryotes. Types of DNA polymerases , enzymes and proteins associated with DNA replication. Mechanism of replication- Intiation, Elongation and Termination. Other modes of replication: rolling circle replication, D-loop mechanism.

**Unit III Genetic code and Transcription**

Genetic code properties; initiation and termination codon; Wobble hypothesis. Transcription in prokaryotes and eukaryotes. RNA polymerases; mechanism of transcription- initiation,elongation and termination.Types of RNA – rRNA,mRNA,tRNA, snRNA. Post transcriptional modifications; 5'capping and poly A tailing.RNA splicing- Splicing pathways and Spliceosome. Concept of Alternate Splicing.

**Unit IV Protein Synthesis**

Translation in prokaryotes and eukaryotes. Mechanism of translation; initiation, elongation and termination. Co and post translational modifications of proteins. Import in to nucleus.

**Unit V Gene regulation**

Regulation of gene expression- regulatory proteins, promoters, enhancers. Operon concept- negative and positive regulation- Lac, Trp , Ara operon- structure and function.

**Textbooks:**

1. Friefelder, David And George M. Malacinski "Essentials Of Molecular Biology" 6th Edition, Panima Publishing, 1993.

2. Karp, G. (2010). Cell and Molecular Biology: Concepts and Experiments. VI Edition. John Wiley & Sons. Inc.
3. De Robertis, E.D.P. and De Robertis, E.M.F. (2006). Cell and Molecular Biology.VIII Edition. Lippincott Williams and Wilkins, Philadelphia.
4. Watson, J. D., Baker T.A., Bell, S. P., Gann, A., Levine, M., and Losick, R., (2008) Molecular Biology of the Gene (VI Edition.). Cold Spring Harbour Lab. Press, Pearson Pub.
5. Darnell, Lodish and Baltimore. Molecular Cell Biology, Scientific American Publishing Inc, 2000

**Reference books:**

- Tropp, Burton E. "Molecular Biology: Genes To Proteins". 3rd Edition. Jones And Bartlett, 2008.
- Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter. Molecular biology of the Cell. 4th ed. Garland publishing Inc, 2002
- Becker, W.M., Kleinsmith, L.J., Hardin. J. and Bertoni, G. P. (2009). The World of the Cell. VII Edition. Pearson Benjamin Cummings Publishing, San Francisco.
- Benjamin Lewin. Gene VII. Oxford University Press, Nelson Cox.

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks : 100** **Exam Duration : 3 Hrs.**

**PART A :** 10 questions of 2 marks each - No choice **20 Marks**

**PART B :** 2 questions from each unit of internal choice; each carrying 16 marks **80 Marks**

# UNIT – I – BASIC CONCEPTS IN MOLECULAR BIOLOGY

## DNA: The genetic material

The physical nature of the gene fascinated scientists for many years. A series of experiments beginning in the 1920s finally revealed that DNA was the genetic material.

## Discovery of transformation

A puzzling observation was made by Frederick Griffith in the course of experiments on the bacterium *Streptococcus pneumoniae* in 1928. This bacterium, which causes pneumonia in humans, is normally lethal in mice. However, different strains of this bacterial species have evolved that differ in virulence (in the ability to cause disease or death). In his experiments, Griffith used two strains that are distinguishable by the appearance of their colonies when grown in laboratory cultures. In onestrain, a normazl virulent type, the cells are enclosed in a polysaccharide capsule, giving colonies a smooth appearance; hence, this strain is labeled *S*. In Griffith's other strain, a mutant nonvirulent type that grows in mice but is not lethal, the polysaccharide coat is absent, giving colonies a rough appearance; this strain is called *R*.

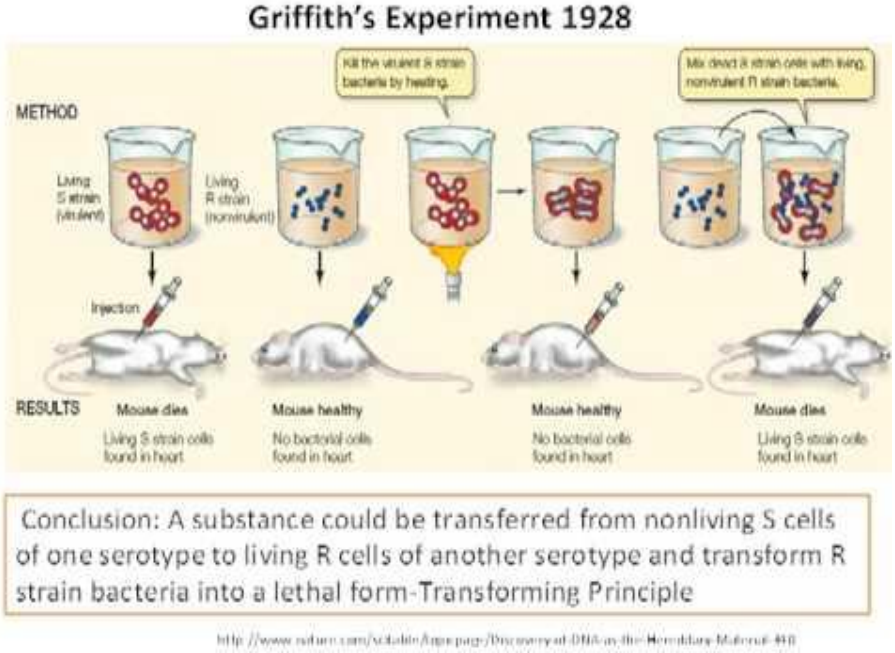



**Figure 1. Griffith Experiment**

Griffith killed some virulent cells by boiling them and injected the heat-killed cells into mice. The mice survived, showing that the carcasses of the cells do not cause death. However, mice injected with a mixture of heat-killed virulent cells and live nonvirulent cells did die. Furthermore, live

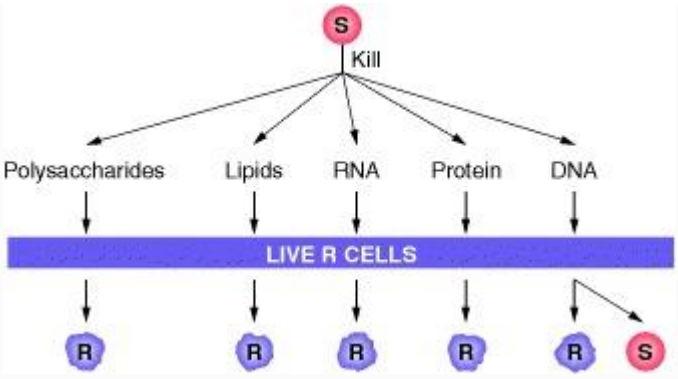
cells could be recovered from the dead mice; these cells gave smooth colonies and were virulent on subsequent injection. Somehow, the cell debris of the boiled S cells had converted the live R cells into live S cells. The pr0cess is called transformation. Griffith's experiment is summarized in Figure .

The first demonstration of bacterial transformation. (a) Mouse dies after injection with the virulent S strain. (b) Mouse survives after injection with the R strain. (c) Mouse survives after injection with heat-killed S strain. (d) Mouse dies after injection with a mixture of heat-killed S strain and live R strain. The heat-killed S strain somehow transforms the R strain into virulence. Parts a, b, and c act as control experiments for this demonstration. .

This same basic technique was then used to determine the nature of the *transforming principle*—the agent in the cell debris that is specifically responsible for transformation. In 1944, Oswald Avery, C. M. MacLeod, and M. McCarty separated the classes of molecules found in the debris of the dead S cells and tested them for transforming ability, one at a time. These tests showed that the polysaccharides themselves do not transform the rough cells. Therefore, the polysaccharide coat, although undoubtedly concerned with the pathogenic reaction, is only the phenotypic expression of virulence. In screening the different groups, Avery and his colleagues found that only one class of molecules, DNA, induced the transformation of R cells (Figure given below). They deduced that DNA is the agent that determines the polysaccharide character and hence the pathogenic character (see pages 219–220 for a description of the mechanism of transformation). Furthermore, it seemed that providing R cells with S DNA was tantamount to providing these cells with S genes.

**Figure 2**

Demonstration that DNA is the transforming agent. DNA is the only agent that produces smooth (S) colonies when added to live rough (R) cells.




**MESSAGE**

**Figure 2.** The demonstration that DNA is the transforming principle was the first demonstration that genes are composed of DNA.

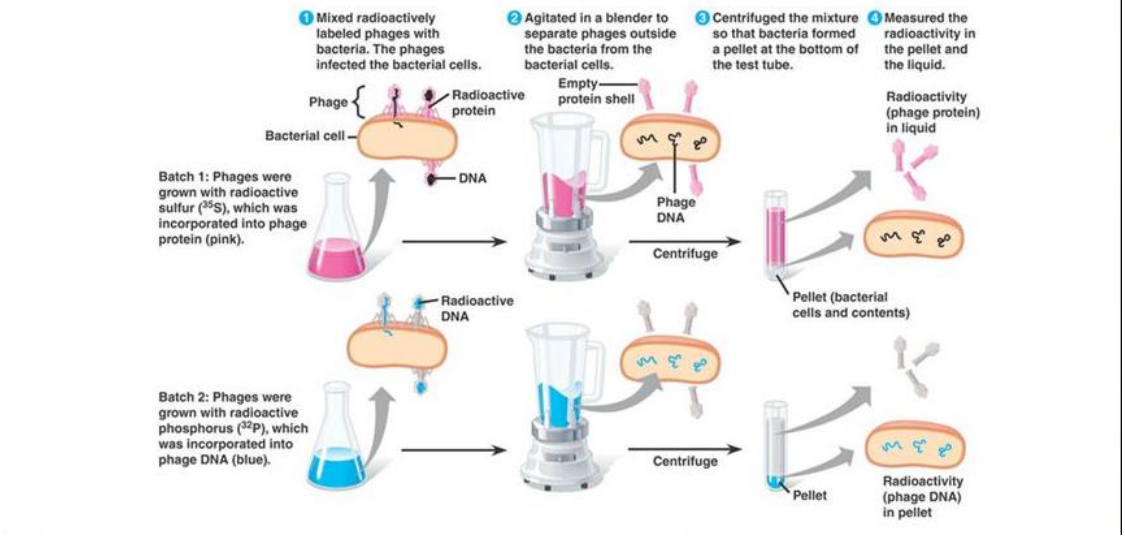
**Hershey-Chase experiment**

The experiments conducted by Avery and his colleagues were definitive, but many scientists were very reluctant to acceptDNA (rather than proteins) as the genetic material. The clincher was provided in 1952 by Alfred Hershey and Martha Chase with the use of the phage (virus) T2. They reasoned that phage infection must entail the introduction (injection) into the bacterium of the specific information that dictates viral reproduction. The phage is relatively simple in molecular constitution. Most of its structure is protein, with DNA contained inside the protein sheath of its "head."

Phosphorus is not found in proteins but is an integral part of DNA; conversely, sulfur is present in proteins but never in DNA. Hershey and Chase incorporated the radioisotope of phosphorus ($^{32}P$) into phage DNA and that of sulfur ($^{35}S$) into the proteins of a separate phage culture. They then used each phage culture independently to infect *E. coli* with many virus particles per cell. After sufficient time for injection to take place, they sheared the empty phage carcasses (called *ghosts*) off the bacterial cells by agitation in a kitchen blender. They used centrifugation to separate the bacterial cells from the phage ghosts and then measured the radioactivity in the two fractions. When the $^{32}P$-labeled phages were used, most of the radioactivity ended up inside the bacterial cells, indicating that the phage DNA entered the cells. $^{32}P$ can also be recovered from phage progeny. When the $^{35}S$-labeled phages were used, most of the radioactive material ended up in the phage ghosts, indicating that the phage protein never entered the bacterial cell (Figure 8-3). The conclusion is inescapable: DNA is the hereditary material; the phage proteins are mere structural packaging that is discarded after delivering the viral DNA to the bacterial cell.

# The Hershey-Chase experiment

- *Analysis of the Hershey-Chase experiment provided evidence that DNA is the genetic material*



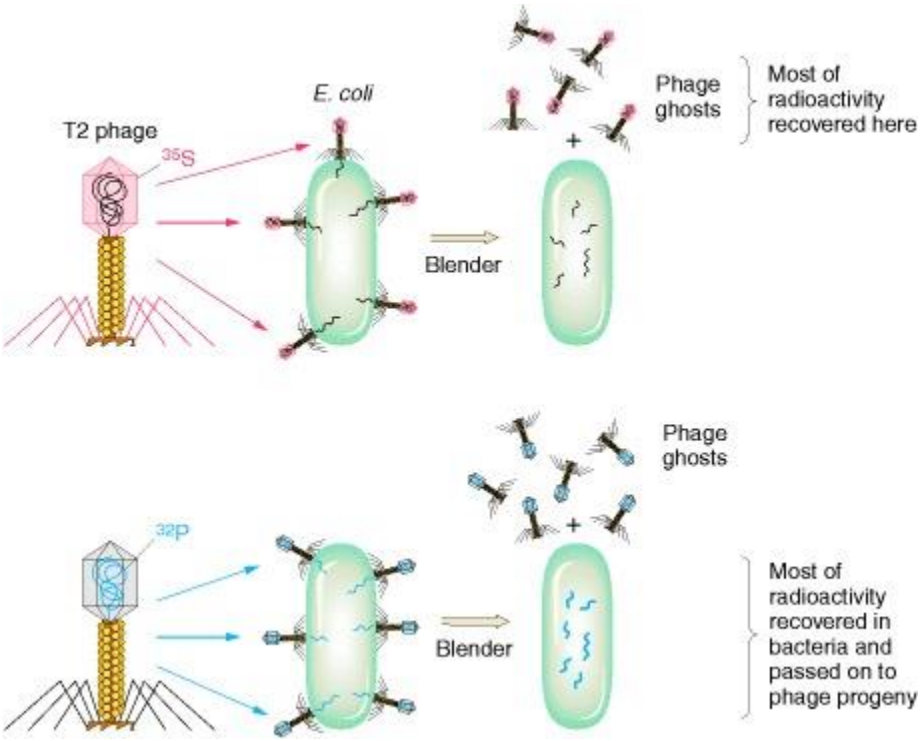




**Figure 3. Hershey Chase Experiment**

The Hershey-Chase experiment, which demonstrated that the genetic material of phage is DNA, not protein. The experiment uses two sets of T2 bacteriophages. In one set, the protein coat is labeled with radioactive sulfur ($^{35}S$), not found in DNA. ther set, the DNA is labeled with radioactive phosphorus ($^{32}P$), not found in protein. Only the $^{32}P$ is injected into the *E. coli*, indicating that DNA is the agent necessary for the production of new phages.

Why such reluctance to accept this conclusion? DNA was thought to be a rather simple chemical. How could all the information about an organism's features be stored in such a simple molecule? How could such information be passed on from one generation to the next? Clearly, the genetic material must have both the ability to encode specific information and the capacity to duplicate that information precisely. What kind of structure could allow such complex functions in so simple a molecule?

**RNA as genetic material**

RNA is the genetic material in viruses was demonstrated in 1956 with the experiments conducted on tobacco plant by A.Gierer and G.Schramm.

All viruses are not limited to bacterial hosts. Viruses that infect and parasite plant cells, some animal cells contain RNA only. In these viruses RNA act as genetic material. One plant virus, Tobacco mosaic virus(TMV), that ontains RNA, not DNA was an important tool for genetic Experiments. TMV infects tobacco, causing the infected regions on leaves to become discoloured and bristled. Different strains of TMV produce clearly different inherited lesions on the infected leaves.The common virus produces a green mosaic disease, but a variant Holmes rib grass(TMV-HR), produces ring spot lesions. Moreover , the amino acid compositions of the proteins of these two strains differ.

H.Fraenkel-Conrat and B.Singer first developed the techniques for separating TMV particles into RNA and protein. They found that virus could be broken into component parts and they could again be reassembled or reconsitituted to form functional virus.From the two strains of TMV they were able to reconstitute viruses with the RNA from TMV common enclosed in TMV-HR protein and TMV-HR RNA with TMV common protein. When these reassembled viruses were used to infect tobacco leaves, the progency viruses produced were always found to be phenotypically and genotypically identical to the parent strain from which

the RNA had been obtained. The reassembled viruses with the TMV-common RNA and TMV-HR protein produced a green mosaic disease characteristic of TMV-common. Recovered virus had protein characteristic of TMV common.This proved that specificity of virus proteins was determined by RNA alone and that proteins carried no genetic information. Hence RNA carries genetic information not proteins.

The genetic RNA is usually found to be single stranded but in some it is double stranded as in reovirus, wound tumor virus.

**Typical tobacco mosaic virus (TMV) particle**

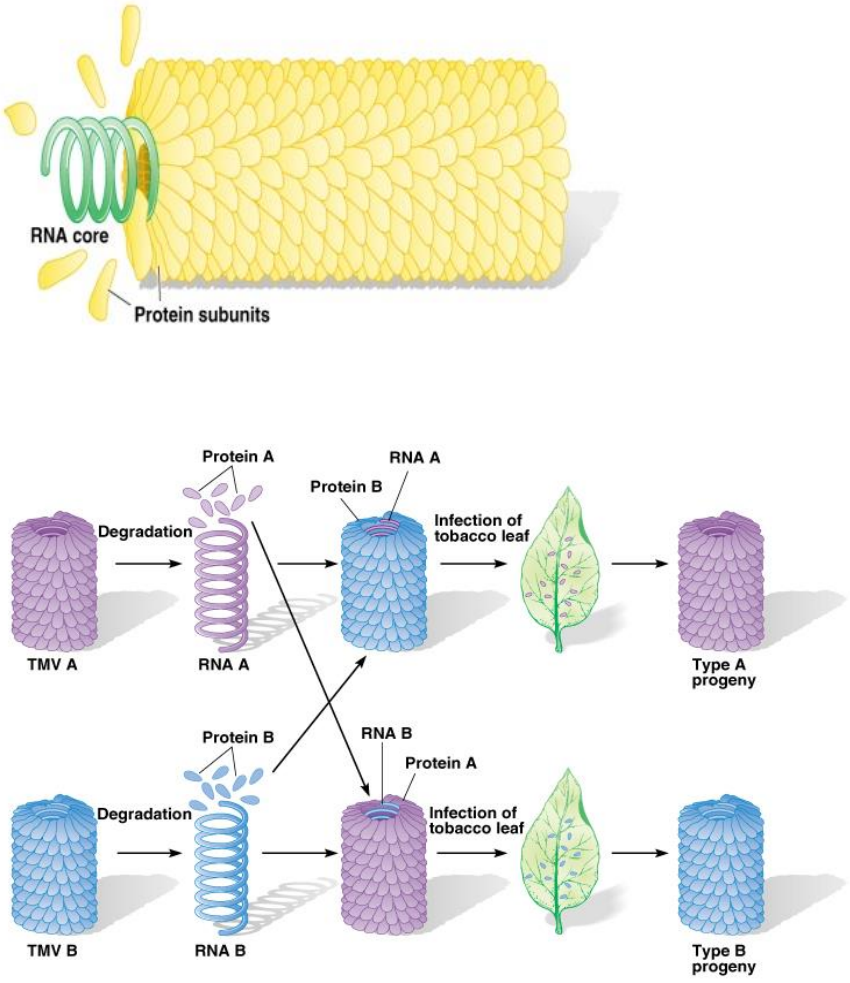



**Figure 4. Demonstration that RNA is the genetic material in tobacco mosaic virus (TMV)**

SUGARS INNUCLEIC ACIDS

Ribose

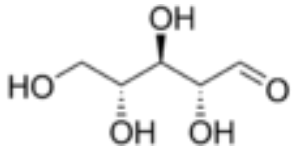



**Figure 5a Ribose**

**Ribose** is a simple sugar and carbohydrate with molecular formula $C_5H_{10}O_5$ the linear-form composition $H{-}(C{=}O){-}(CHOH)_4{-}H$.

In its linear form, ribose can be recognised as the pentose sugar with all of its hydroxyl functional groups on the same side in its Fischer projection.

Fischer chose the name "ribose" as it is a partial rearrangement of the name of another sugar, arabinose, of which ribose is an epimer at the 2' carbon;

both names also relate to gum arabic, from which arabinose was first isolated and from which they prepared l-ribose.

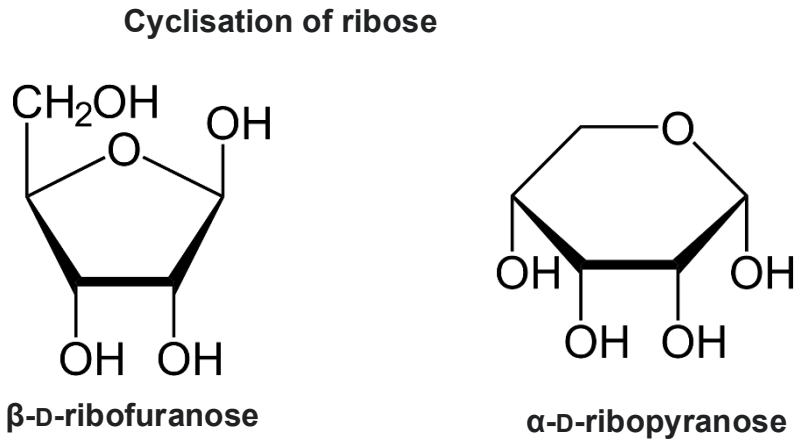



**Figure 5b. Ribose**

In each case, there are two possible geometric outcomes, named as α- and β- and known as anomers, depending on the stereochemistry at the hemiacetal carbon atom (the "anomeric carbon")

At room temperature, about 76% of d-ribose is present in pyranose forms (α:β = 1:2) and 24% in the furanose forms (α:β = 1:3), with only about 0.1% of the linear form present.

- The ribonucleosides adenosine, cytidine, guanosine, and uridine are all derivatives of β-d-ribofuranose.
- Metabolically-important species that include phosphorylated ribose include ADP, ATP, coenzyme A, and NADH.
- cAMP and cGMP serve as secondary messengers in some signaling pathways and are also ribose derivatives.
- The ribose moiety appears in some pharmaceutical agents, including the antibiotics neomycin and paromomycin.

**Structure**

- Ribose is an aldopentose (a monosaccharide containing five carbon atoms) that, in its open chain form, has an aldehyde functional group at one end.
- In the conventional numbering scheme for monosaccharides, the carbon atoms are numbered from C1' (in the aldehyde group) to C5'.
- The deoxyribose derivative found in DNA differs from ribose by having a hydrogen atom in place of the hydroxyl group at C2'. This hydroxyl group performs a function in RNA splicing.
- Like many monosaccharides, ribose exists in an equilibrium among 5 forms—the linear form H−(C=O)−$(CHOH)_4$–H and either of the two ring forms: *α- or β-ribofuranose* ("C3'-endo"), with a five-membered tetrahydrofuran ring, and *α- or β-ribopyranose* ("C2'-endo"), with a six-membered tetrahydropyran ring.

- The “d-” in the name d-ribose refers to the stereochemistry of the chiral carbon atom farthest away from the aldehyde group (C4'). In d-ribose, as in all d-sugars, this carbon atom has the same configuration as in d-glyceraldehyde.

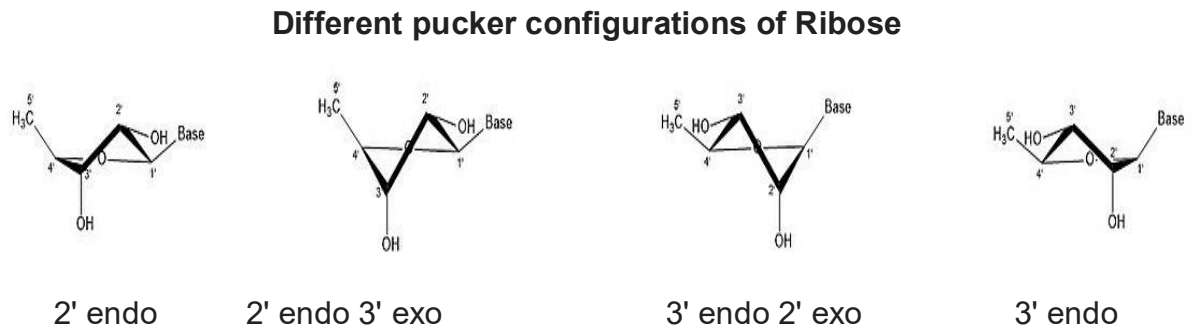



**Figure 6. Different Pucker configurations of Ribose**

A ribose molecule is typically represented as a planar molecule on paper. Despite this, it is typically non-planar in nature.

Even between hydrogen atoms, the many constituents on a ribose molecule cause steric hindrance and strain between them.

To relieve this crowding and ring strain, the ring puckers, i.e. becomes non-planar.

This puckering is achieved by displacing an atom from the plane, relieving the strain and yielding a more stable configuration.

When only a single atom is displaced, it is referred to as an "envelope" pucker.

When two atoms are displaced, it is referred to as a "twist" pucker, in reference to the zigzag orientation

In an "endo" pucker, the major displacement of atoms is on the β-face, the same side as the C4'-C5' bond and the base.

In an "exo" pucker, the major displacement of atoms is on the α-face, on the opposite side of the ring.

The major forms of ribose are the 3'-endo pucker (commonly adopted by RNA and A-form DNA) and 2'-endo pucker (commonly adopted by B-form DNA).

## FUNCTIONS IN BIOCHEMISTRY

- Ribose plays many important roles in metabolism, which means that it is involved in a lot of biochemistry.
- Ribose is used as a building block for a lot of the signals and products throughout the metabolic pathway.
- One of the most important products of the metabolic pathway is adenosine triphosphate (ATP), which provides energy that drives processes in cells.
- ATP is derived from ribose; it contains one ribose, three phosphate groups, and an adenine base.
- ATP is created during cellular respiration from adenosine diphosphate (ATP with one less phosphate group).

### Signalling Pathway

Ribose also plays a major role in signaling pathways because it is a building block in secondary signaling molecules such as cyclic adenosine monophosphate (cAMP) which is derived from ATP.

One specific case in which cAMP is used is in cAMP-dependent signaling pathways. In cAMP signaling pathways, either a stimulative or inhibitory hormone receptor is activated by a signal molecule.

These receptors are linked to a stimulative or inhibitory regulative G-protein. When a stimulative G-protein is activated, adenylyl cyclase catalyzes ATP into cAMP by using $Mg^{2+}$ or $Mn^{2+}$. cAMP, a secondary messenger, then goes on to activate protein kinase A, which is an enzyme that regulates cell metabolism.

Protein kinase A regulates metabolic enzymes by phosphorylation which causes a change in the cell depending on the original signal molecule. The opposite occurs when an inhibitory G-protein is activated; the G-protein inhibits adenylyl cyclase and ATP is not converted to cAMP.

### Metabolism

- Ribose is referred to as the "molecular currency" because of its involvement in intracellular energy transfers. For example, nicotinamide adenine dinucleotide (NAD), flavin adenine

dinucleotide (FAD), and nicotinamide adenine dinucleotide phosphate (NADP) all contain the d-ribofuranose moiety.

- They can each be derived from d-ribose after it is converted to d-ribose 5-phosphate by the enzyme ribokinase.
- NAD, FAD, and NADP act as electron acceptors in biochemical redox reactions in major metabolic pathways including glycolysis, the citric acid cycle, fermentation, and the electron transport chain.

**DEOXYRIBOSE**

Deoxyribose, or more precisely 2-deoxyribose, is a monosaccharide with idealized formula $H{-}(C{=}O){-}(CH_2){-}(CHOH)_3{-}H$.

Its name indicates that it is a deoxy sugar, meaning that it is derived from the sugar ribose by loss of an oxygen atom.

Since the pentose sugars arabinose and ribose only differ by the stereochemistry at C2′, 2-deoxyribose and 2-deoxyarabinose are equivalent, although the latter term is rarely used because ribose, not arabinose, is the precursor to deoxyribose.

**STRUCTURE**

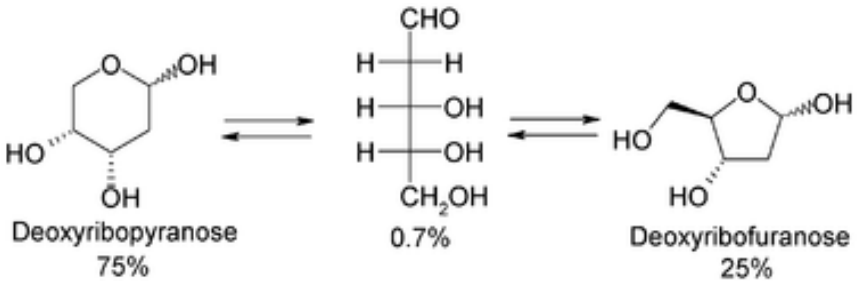



**Figure 7. Deoxyribose**

Several isomers exist with the formula $H{-}(C{=}O){-}(CH_2){-}(CHOH)_3{-}H$, but in deoxyribose all the hydroxyl groups are on the same side in the Fischer projection.

The term "2-deoxyribose" may refer to either of two enantiomers:

- the biologically important d-2-deoxyribose and

- to the rarely encountered mirror image l-2-deoxyribose.

d-2-deoxyribose is a precursor to the nucleic acid DNA. 2-deoxyribose is an aldopentose, that is, a monosaccharide with five carbon atoms and having an aldehyde functional group.

In aqueous solution, deoxyribose primarily exists as a mixture of three structures: the linear form $H{-}(C{=}O){-}(CH_2){-}(CHOH)_3{-}H$ and two ring forms, deoxyribofuranose ("C3′-endo"), with a five-membered ring, and deoxyribopyranose ("C2′-endo"), with a six-membered ring.

**BIOLOGICAL IMPORTANCE**

As a component of DNA, 2-deoxyribose derivatives have an important role in biology.

The DNA (deoxyribonucleic acid) molecule, which is the main repository of genetic information in life, consists of a long chain of deoxyribose-containing units called nucleotides, linked via phosphate groups.

In the standard nucleic acid nomenclature, a DNA nucleotide consists of a deoxyribose molecule with an organic base (usually adenine, thymine, guanine or cytosine) attached to the 1′ ribose carbon. The 5′ hydroxyl of each deoxyribose unit is replaced by a phosphate (forming a nucleotide) that is attached to the 3′ carbon of the deoxyribose in the preceding unit.

The absence of the 2′ hydroxyl group in deoxyribose is apparently responsible for the increased mechanical flexibility of DNA compared to RNA, which allows it to assume the double-helix conformation, and also (in the eukaryotes) to be compactly coiled within the small cell nucleus.

The double-stranded DNA molecules are also typically much longer than RNA molecules. The backbone of RNA and DNA are structurally similar, but RNA is single stranded, and made from ribose as opposed to deoxyribose.

Other biologically important derivatives of deoxyribose include mono-, di-, and triphosphates, as well as 3′-5′ cyclic monophosphates.

**NITROGENOUS BASES**

ADENINE

**Adenine** is a nucleobase (a purine derivative).

It is one of the four nucleobases in the nucleic acid of DNA that are represented by the letters G–C–A–T.


**Figure 8. Adenine**

The three others are guanine, cytosine and thymine.

Its derivatives have a variety of roles in biochemistry including cellular respiration, in the form of both the energy-rich adenosine triphosphate (ATP) and the cofactors nicotinamide adenine dinucleotide (NAD) and flavin adenine dinucleotide (FAD).

It also has functions in protein synthesis and as a chemical component of DNA and RNA.

The shape of adenine is complementary to either thymine in DNA or uracil in RNA.

When connected into DNA, a covalent bond is formed between deoxyribose sugar and the bottom left nitrogen (thereby removing the existing hydrogen atom).

Adenosine is adenine reacted with ribose, as used in RNA and ATP;

deoxyadenosine is adenine attached to deoxyribose, as used to form DNA.

**Function**

Adenine is one of the two purine nucleobases (the other being guanine) used in forming nucleotides of the nucleic acids.

In DNA, adenine binds to thymine via two hydrogen bonds to assist in stabilizing the nucleic acid structures.

In RNA, which is used for protein synthesis, adenine binds to uracil.

Adenine forms adenosine, a nucleoside, when attached to ribose, and deoxyadenosine when attached to deoxyribose.

It forms adenosine triphosphate (ATP), a nucleoside triphosphate, when three phosphate groups are added to adenosine.

Adenosine triphosphate is used in cellular metabolism as one of the basic methods of transferring chemical energy between chemical reactions.

GUANINE

**Guanine** is one of the four main nucleobases found in the nucleic acids DNA and RNA, the others being adenine, cytosine, and thymine (uracil in RNA).

In DNA, guanine is paired with cytosine. The guanine nucleoside is called guanosine.

With the formula $C_5H_5N_5O$, guanine is a derivative of purine, consisting of a fused pyrimidine-imidazole ring system with conjugated double bonds.

This unsaturated arrangement means the bicyclic molecule is planar.

**Properties**

Guanine, along with adenine and cytosine, is present in both DNA and RNA, whereas thymine is usually seen only in DNA, and uracil only in RNA. Guanine has two tautomeric forms, the major keto form and rare enol form.

It binds to cytosine through three hydrogen bonds. In cytosine, the amino group acts as the hydrogen bond donor and the C-2 carbonyl and the N-3 amine as the hydrogen-bond acceptors.

Guanine has the C-6 carbonyl group that acts as the hydrogen bond acceptor, while a group at N-1 and the amino group at C-2 act as the hydrogen bond donors.

Guanine can be hydrolyzed with strong acid to glycine, ammonia, carbon dioxide, and carbon monoxide.

First, guanine gets deaminated to become xanthine.

Guanine oxidizes more readily than adenine, the other purine-derivative base in DNA.

Its high melting point of 350 °C reflects the intermolecular hydrogen bonding between the oxo and amino groups in the molecules in the crystal.

Because of this intermolecular bonding, guanine is relatively insoluble in water, but it is soluble in dilute acids and bases.

**Other occurrences and biological uses**

Guanine crystals are rhombic platelets composed of multiple transparent layers, but they have a high index of refraction that partially reflects and transmits light from layer to layer, thus producing a pearly luster. It can be applied by spray, painting, or dipping.

It may irritate the eyes.

Guanine has a very wide variety of biological uses that include a range of functions ranging in both complexity and versatility. These include camouflage, display, and vision among other purposes.

Spiders, scorpions, and some amphibians convert ammonia, as a product of protein metabolism in the cells, to guanine, as it can be excreted with minimal water loss.

Guanine is also found in specialized skin cells of fish called iridocytes (e.g., the sturgeon), as well as being present in the reflective deposits of the eyes of deep-sea fish and some reptiles, such as crocodiles.

On 8 August 2011, a report, based on NASA studies with meteorites found on Earth, was published suggesting building blocks of DNA and RNA (guanine, adenine and related organic molecules) may have been formed extra-terrestrially in outer space.

CYTOSINE

**Cytosine** C is one of the four main bases found in DNA and RNA, along with adenine, guanine, and thymine (uracil in RNA).

It is a pyrimidine derivative, with a heterocyclic aromatic ring and two substituents attached (an amine group at position 4 and a keto group at position 2).

The nucleoside of cytosine is cytidine.

In Watson-Crick base pairing, it forms three hydrogen bonds with guanine.

**Chemical Reactions**

Cytosine can be found as part of DNA, as part of RNA, or as a part of a nucleotide.

As cytidine triphosphate (CTP), it can act as a co-factor to enzymes, and can transfer a phosphate to convert adenosine diphosphate (ADP) to adenosine triphosphate (ATP).

In DNA and RNA, cytosine is paired with guanine.

However, it is inherently unstable, and can change into uracil (spontaneous deamination). This can lead to a point mutation if not repaired by the DNA repair enzymes such as uracil glycosylase, which cleaves a uracil in DNA.

Cytosine can also be methylated into 5-methylcytosine by an enzyme called DNA methyltransferase or be methylated and hydroxylated to make 5-hydroxymethylcytosine.

The difference in rates of deamination of cytosine and 5-methylcytosine (to uracil and thymine) forms the basis of bisulfite sequencing.

**Biological function**

When found third in a codon of RNA, cytosine is synonymous with uracil, as they are interchangeable as the third base.

When found as the second base in a codon, the third is always interchangeable. For example, UCU, UCC, UCA and UCG are all serine, regardless of the third base.

Active enzymatic deamination of cytosine or 5-methylcytosine by the APOBEC family of cytosine deaminases could have both beneficial and detrimental implications on various cellular processes as well as on organismal evolution.

The implications of deamination on 5-hydroxymethylcytosine, on the other hand, remains less understood.

THYMINE

**Thymine** T is one of the four nucleobases in the nucleic acid of DNA that are represented by the letters G–C–A–T.

The others are adenine, guanine, and cytosine.

Thymine is also known as **5 -methyluracil,** a pyrimidine nucleobase.

In RNA, thymine is replaced by the nucleobase uracil.

Thymine was first isolated in 1893 by Albrecht Kossel and Albert Neumann from calves' thymus glands, hence its name.

**Derivation**

As its alternate name (5-methyluracil) suggests, thymine may be derived by methylation of uracil at the 5th carbon.

In RNA, thymine is replaced with uracil in most cases.

In DNA, thymine (T) binds to adenine (A) via two hydrogen bonds, thereby stabilizing the nucleic acid structures.

Thymine combined with deoxyribose creates the nucleoside deoxythymidine, which is synonymous with the term thymidine.

Thymidine can be phosphorylated with up to three phosphoric acid groups, producing dTMP (**d**eoxy**t**hymidine **m**ono**p**hosphate), dTDP, or dTTP (for the **d**i- and **t**ri-phosphates, respectively).

One of the common mutations of DNA involves two adjacent thymines or cytosine, which, in presence of ultraviolet light, may form thymine dimers, causing "kinks" in the DNA molecule that inhibit normal function.

Thymine could also be a target for actions of 5-fluorouracil (5-FU) in cancer treatment. 5-FU can be a metabolic analog of thymine (in DNA synthesis) or uracil (in RNA synthesis). Substitution of this analog inhibits DNA synthesis in actively dividing cells.

Thymine bases are frequently oxidized to hydantoins over time after the death of an organism.

URACIL

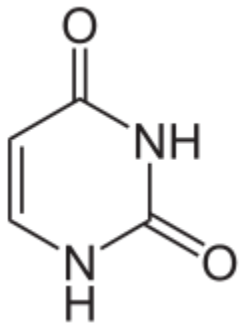



**Figure 9: Uracil, U** is one of the four nucleobases in the nucleic acid RNA that are represented by the letters A, G, C and U.

The others are adenine (A), cytosine (C), and guanine (G).

In RNA, uracil binds to adenine via two hydrogen bonds.

In DNA, the uracil nucleobase is replaced by thymine.

Uracil is a demethylated form of thymine.

Uracil is a common and naturally occurring pyrimidine derivative.

**Properties**

In RNA, uracil base-pairs with adenine and replaces thymine during DNA transcription. Methylation of uracil produces thymine.

In DNA, the evolutionary substitution of thymine for uracil may have increased DNA stability and improved the efficiency of DNA replication (discussed below). Uracil pairs with adenine through hydrogen bonding.

When base pairing with adenine, uracil acts as both a hydrogen bond acceptor and a hydrogen bond donor.

In RNA, uracil binds with a ribose sugar to form the ribonucleoside uridine. When a phosphate attaches to uridine, uridine 5'-monophosphate is produced.

**Reactions**

Uracil readily undergoes regular reactions including oxidation, nitration, and alkylation. While in the presence of phenol (PhOH) and sodium hypochlorite (NaOCl), uracil can be visualized in ultraviolet light.

Uracil also has the capability to react with elemental halogens because of the presence of more than one strongly electron donating group.

**Uses**

Uracil's use in the body is to help carry out the synthesis of many enzymes necessary for cell function through bonding with riboses and phosphates.

Uracil serves as allosteric regulator and coenzyme for reactions in animals and in plants.

UMP controls the activity of carbamoyl phosphate synthetase and aspartate transcarbamoylase in plants, while UDP and UTP requlate CPSase II activity in animals.

UDP-glucose regulates the conversion of glucose to galactose in the liver and other tissues in the process of carbohydrate metabolism.

Uracil is also involved in the biosynthesis of polysaccharides and the transportation of sugars containing aldehydes.

Uracil is important for the detoxification of many carcinogens, for instance those found in tobacco smoke.

## PHOSPHODIESTER BOND

A **phosphodiester bond** occurs when exactly two of the hydroxyl groups in phosphoric acid react with hydroxyl groups on other molecules to form two ester bonds.

Phosphodiester bonds are central to all life on Earth as they make up the backbone of the strands of nucleic acid. In DNA and RNA, the phosphodiester bond is the linkage between the 3' carbon atom of one sugar molecule and the 5' carbon atom of another, deoxyribose in DNA and ribose in RNA.

Strong covalent bonds form between the phosphate group and two 5-carbon ring carbohydrates (pentoses) over two ester bonds.

The phosphate groups in the phosphodiester bond are negatively charged.

Because the phosphate groups have a $pK_a$ near 0, they are negatively charged at pH 7.

This repulsion forces the phosphates to take opposite sides of the DNA strands and is neutralized by proteins (histones), metal ions such as magnesium, and polyamines.

In order for the phosphodiester bond to be formed and the nucleotides to be joined, the tri-phosphate or di-phosphate forms of the nucleotide building blocks are broken apart to give off energy required to drive the enzyme-catalyzed reaction.

When a single phosphate or two phosphates known as pyrophosphates break away and catalyze the reaction, the phosphodiester bond is formed.

Hydrolysis of phosphodiester bonds can be catalyzed by the action of phosphodiesterases which play an important role in repairing DNA sequences.

The phosphodiester linkage between two ribonucleotides can be broken by alkaline hydrolysis, whereas the linkage between two deoxyribonucleotides is more stable under these conditions.

The relative ease of RNA hydrolysis is an effect of the presence of the 2' hydroxyl group.

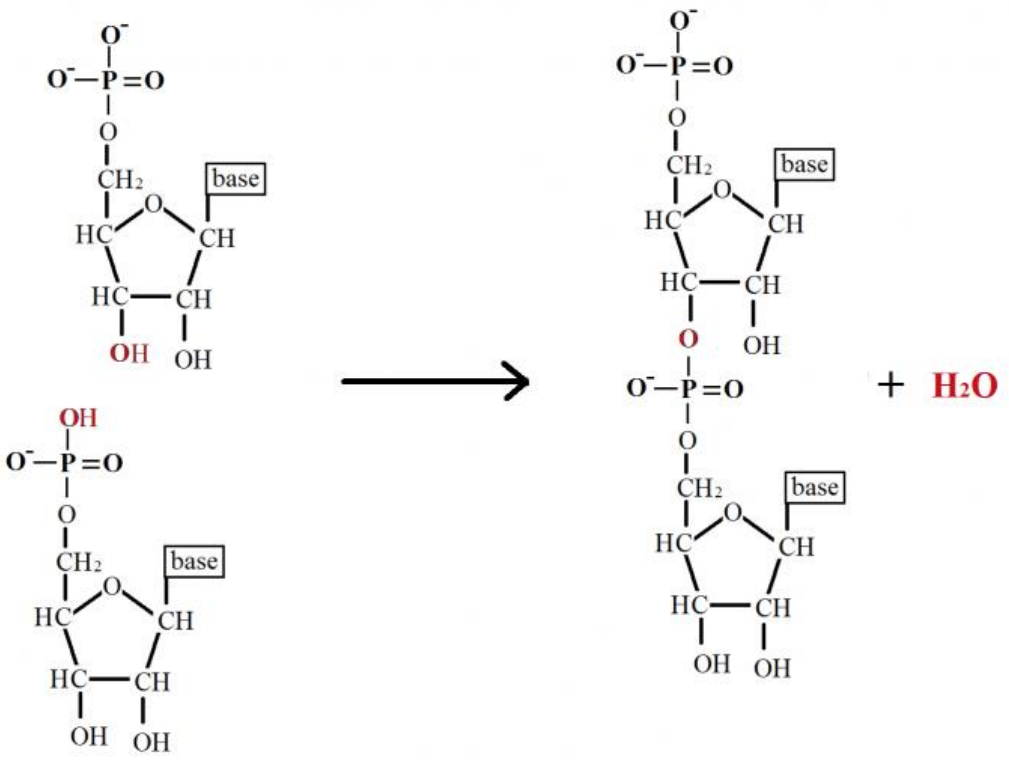



**Figure 10. Phosphodiester Bond Formation**

In phosphodiester formation, two hydroxyl (OH) groups on the phosphate molecule bind to the 3' and 5' carbons on two independent pentose sugars.

These are two condensation reactions, so two molecules of water are produced.

The phosphate is then bonded to the sugars by two ester bonds, hence the nomenclature of phosphodiester bond. This reaction is catalysed by ligases, such as DNA ligase during DNA replication.

## NUCLEOTIDES

**Nucleic acids** are linear, unbranched polymers of **nucleotides**.

**1.**

A five-carbon sugar (hence a **pentose**). Two kinds are found:

**deoxyribose**, which has a hydrogen atom attached to its #2 carbon atom (designated 2'), and

**ribose**, which has a hydroxyl group there.

Deoxyribose-containing nucleotides, the **deoxyribonucleotides**, are the monomers of deoxyribonucleic acids (**DNA)**.

Ribose-containing nucleotides, the **ribonucleotides**, are the monomers of ribonucleic acids (**RNA)**.

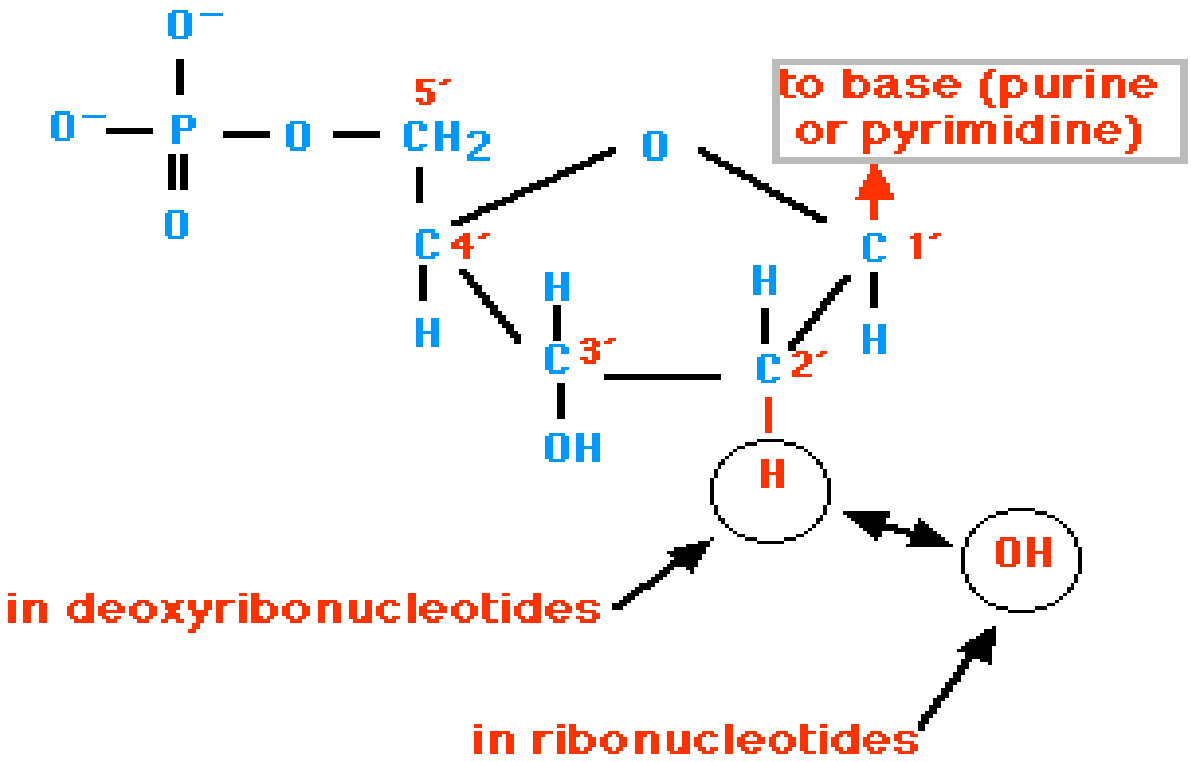



**Figure 11. Nucleotide**

**2.**

A nitrogen-containing ring structure called a **nucleobase** (or simply a base). The nucleobase is attached to the 1' carbon atom of the pentose. In **DNA**, four different nucleobases are found:

two **purines**, called **adenine** (**A**) and **guanine** (**G**)

two **pyrimidines**, called **thymine** (**T**) and **cytosine** (**C**)

**RNA** contains:

The same purines, **adenine** (**A**) and **guanine** (**G**).

RNA also uses the pyrimidine **cytosine** (**C**), but instead of thymine, it uses the pyrimidine **uracil** (**U**).

The combination of a nucleobase and a pentose is called a **nucleoside**.

**3.**

One (as shown in the first figure), two, or three **phosphate** groups. These are attached to the 5' carbon atom of the pentose. The product in each case is called a **nucleotide**.

Both DNA and RNA are assembled from **nucleoside triphosphates**.

For **DNA**, these are **dATP**, **dGTP**, **dCTP**, and **dTTP**.

For **RNA**, these are **ATP**, **GTP**, **CTP**, and **UTP**.

In both cases, as each nucleotide is attached, the second and third phosphates are removed.

| The nucleosides and their mono-, di-, and triphosphates | | | | | |
|---|---|---|---|---|---|
| | **Nucleobase** | **Nucleoside** | **Nucleotides** | | |
| **DNA** | Adenine (A) | Deoxyadenosine | dAMP | dADP | dATP |
| | Guanine (G) | Deoxyguanosine | dGMP | dGDP | dGTP |
| | Cytosine (C) | Deoxycytidine | dCMP | dCDP | dCTP |
| | Thymine (T) | Deoxythymidine | dTMP | dTDP | dTTP |
| **RNA** | Adenine (A) | Adenosine | AMP | ADP | ATP |
| | Guanine (G) | Guanosine | GMP | GDP | GTP |
| | Cytosine (C) | Cytidine | CMP | CDP | CTP |
| | Uracil (U) | Uridine | UMP | UDP | UTP |

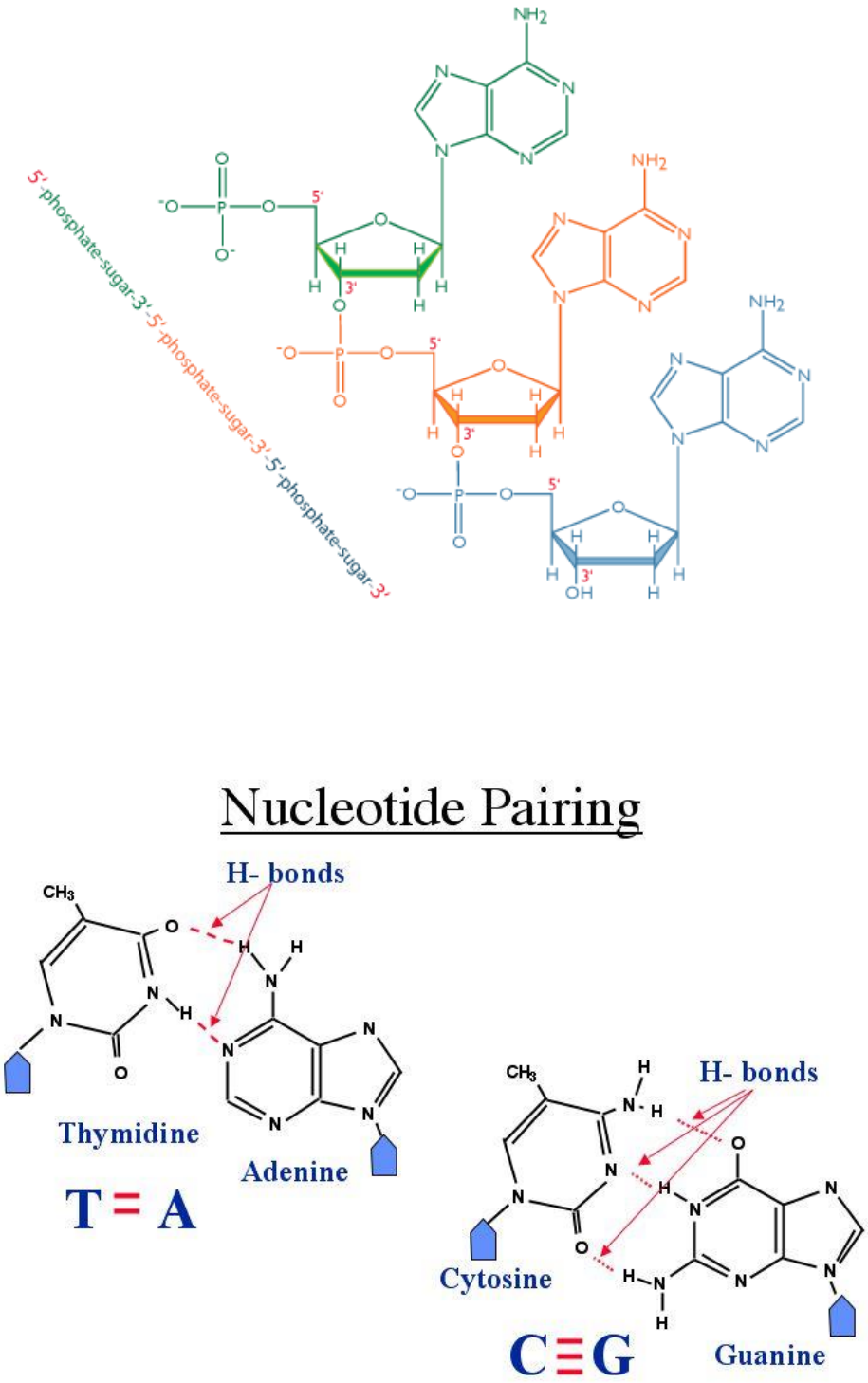



**Figure 12. Base pairing**

The nucleic acids, both DNA and RNA, consist of polymers of nucleotides. The nucleotides are linked covalently between the 3' carbon atom of the pentose and the phosphate group attached to the 5' carbon of the adjacent pentose.

Most intact DNA molecules are made up of **two** strands of polymer, forming a "double helix".

RNA molecules, while single-stranded, usually contain regions where two portions of the strand twist around each other to form helical regions

The two strands of DNA and the helical regions of RNA are held together by base pairing.

**DOUBLE HELIX**

The double helix of DNA has these features:

- It contains two polynucleotide strands wound around each other.
- The backbone of each consists of alternating deoxyribose and phosphate groups.
- The phosphate group bonded to the 5' carbon atom of one deoxyribose is covalently bonded to the 3' carbon of the next.
- The two strands are "antiparallel"; that is, one strand runs 5′ to 3′ while the other runs 3′ to 5′.
- The DNA strands are assembled in the 5′ to 3′ direction [More] and, by convention, we "read" them the same way.
- The purine or pyrimidine attached to each deoxyribose projects in toward the axis of the helix.
- Each base forms hydrogen bonds with the one directly opposite it, forming **base pairs** (also called nucleotide pairs).
- 3.4 Å separate the planes in which adjacent base pairs are located.
- The double helix makes a complete turn in just over 10 nucleotide pairs, so each turn takes a little more (35.7 Å to be exact) than the 34 Å shown in the diagram.
- There is an average of 25 hydrogen bonds within each complete turn of the double helix providing a stability of binding about as strong as what a covalent bond would provide.
- The diameter of the helix is 20 Å.

- The helix can be virtually any length; when fully stretched, some DNA molecules are as much as 5 cm (2 inches!) long.
- The path taken by the two backbones forms a major (wider) groove (from "34 A" to the top of the arrow) and a minor (narrower) groove (the one below).

This structure of DNA was worked out by **Francis Crick** and **James D. Watson** in 1953.

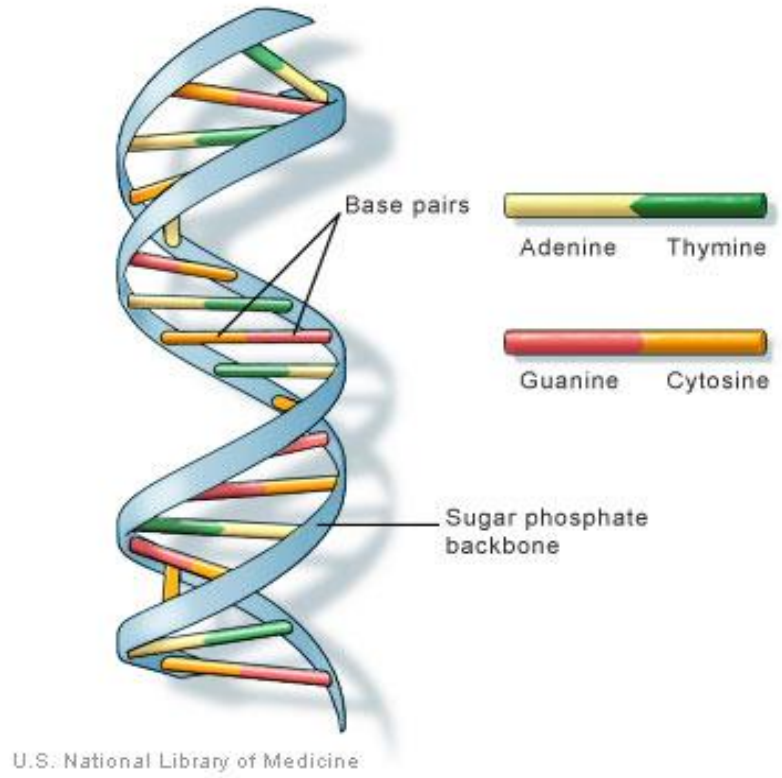



**Figure 13. DNA structure**

## DNA REPLICATION

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material' – **Watson & Crick** ***Nature (1953)***

**Process of duplication of the entire genome prior to cell division**

**Biological significance**

- extreme accuracy of DNA replication is necessary in order to preserve the integrity of the genome in successive generations
- In eukaryotes , replication only occurs during the S phase of the cell cycle.
- Replication rate in eukaryotes is slower resulting in a higher fidelity/accuracy of replication in eukaryotes

## BASIC RULES OF REPLICATION

A. Semi-conservative

B. Starts at the 'origin'

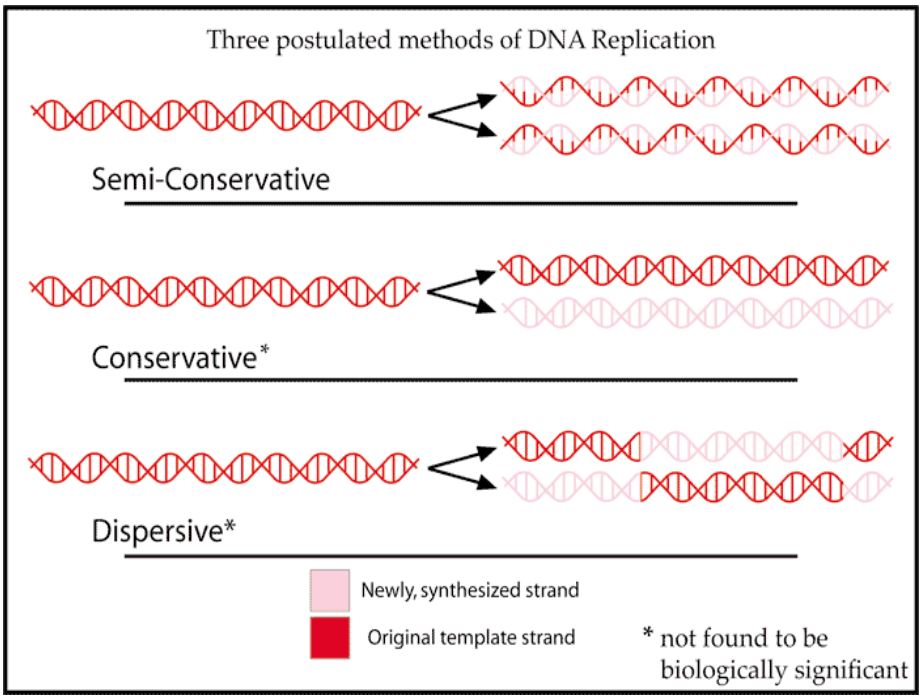
C. Synthesis always in the 5-3' direction

D. Can be uni or bidirectional

E. Semi-discontinuous

F. RNA primers required

A. Semiconservative



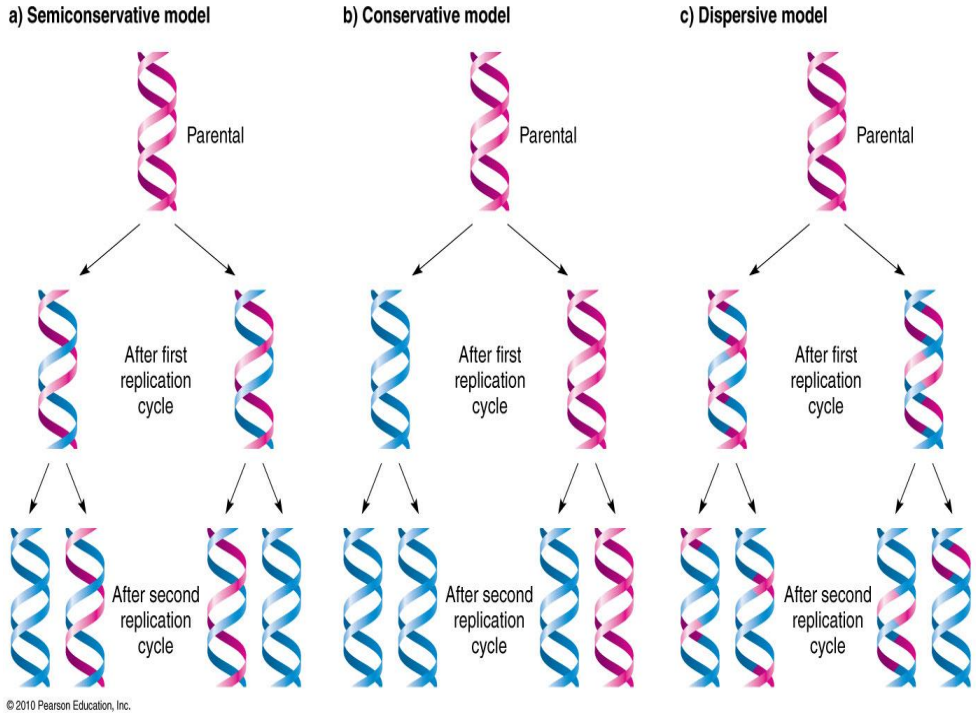




**Figure 14. 3 MODES OF REPLICATION**

**Meselson and Stahl's experiment. M. Meselson** and **F.W. Stahl** (1958) verified the semiconservative nature of DNA replication in a series of elegant experiments using isotopically labelled DNA and a form of isopycnic density gradient centrifugation. They cultured *Escherichai coli* cells in a medium in which the nitrogen was 15N (a 'heavy' isotope of nitrogen, but not a radioisotope) instead of commonly occurring and lighter 14N. In time, the purines and pyrimidines of DNA in new cells contained 15N (where 14N normally occurs) and, thus, the DNA molecules were denser. DNA in which the nitrogen atoms are heavy (15N) can be distinguished from DNA containing light nitrogen (14N), because during isopycnic centrifugation, the two different DNAs band at different density positions in the centrifuge tube.

Depending on its content of 15N and 14N, the DNA bands at a specific position in the density gradient. Because the DNA synthezised by *E.coli.* cells grown in 15N would be denser than 14Ncontaining DNA, it would band further down the tube. *E. coli* cells grown for sometime in the presence of 15N-medium were washed free of the medium and transferred to 14N-containing medium and allowed to continue to grow for specific lengths of time (*i.e.*, for various numbers of generation time). DNA isolated from cells grown for one generation of time in the 14N medium had a density intermediate to that of the DNA from cells grown only in 15N-containing medium (identified as *generation O;* Fig. 4.3) and that of DNA from cells grown only in 14N-containing medium (*the controls*). Such a result immediately ruled out the possibility that DNA replication was conservative, because the conservative replication would have yielded two DNA bands in the density gradient for **generation 1** (*i.e.*, F1 cells).

The single band of intermediate density (identified as ''**hybrid DNA'')** consisted of DNA molecules in which one strand contained 15N and the other contained 14N. When the incubation in the 14N-medium was carried out for two generations of time (*i.e.* **generation 2**), two DNA bands were formed — one at the same density position as the DNA from cells grown exclusively in 14N medium (*i.e.*, light controls) and the other of intermediate density. Subsequent generations produced greater numbers of DNA molecules

MESELSON – STAHL EXPERIMENT

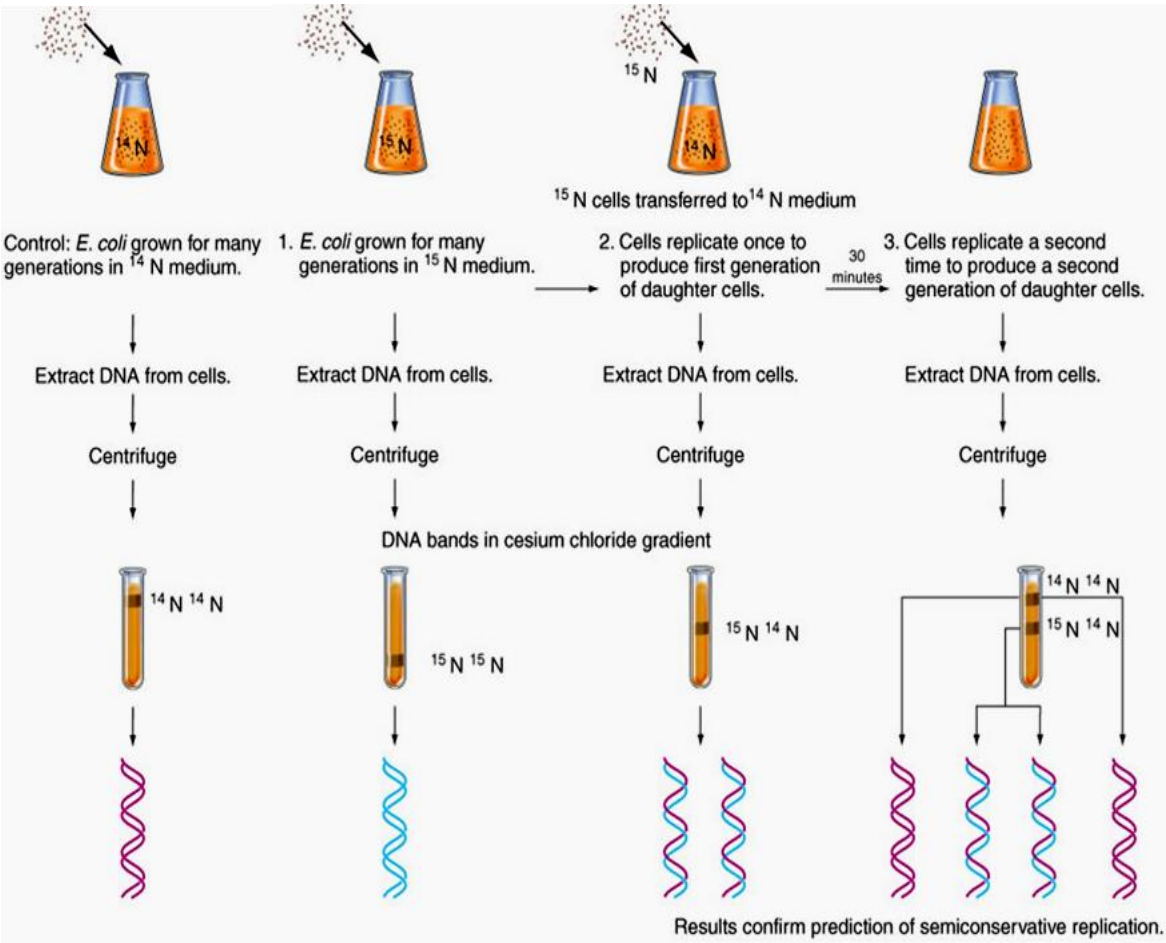



**Figure 15a. Semi-conservative replication:**

One strand of duplex passed on unchanged to each of the daughter cells. This 'conserved' strand acts as a template for the synthesis of a new, complementary strand by the enzyme DNA polymerase

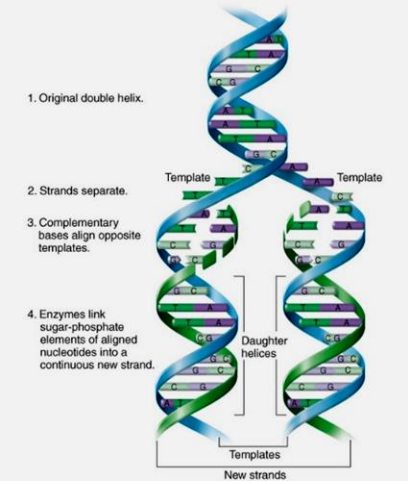



**Figure 15b.**

**Semi-conservative replication:**

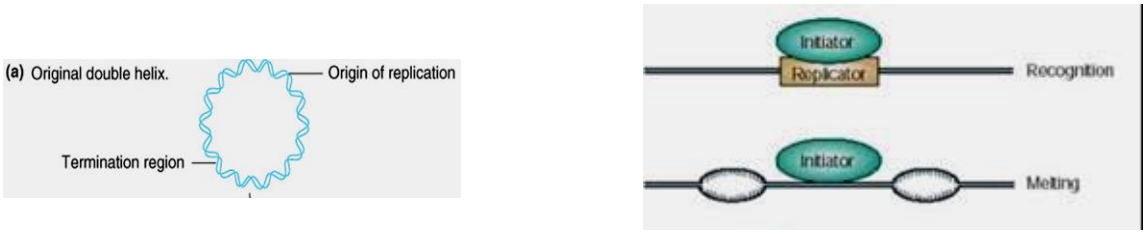
B. STARTS AT ORIGIN OF REPLICATION

Initiator proteins identify specific base sequences on DNA called sites of origin

Prokaryotes – single origin site E.g *E.coli* - *oriC*

Eukaryotes – multiple sites of origin (replicator)

E.g. yeast - ARS (autonomously replicating sequences)




**Prokaryotes**

**Eukaryotes**

**Figure 16. Origin of replication**

C. SYNTHESIS ALWAYS IN 5'-3'DIRECTION

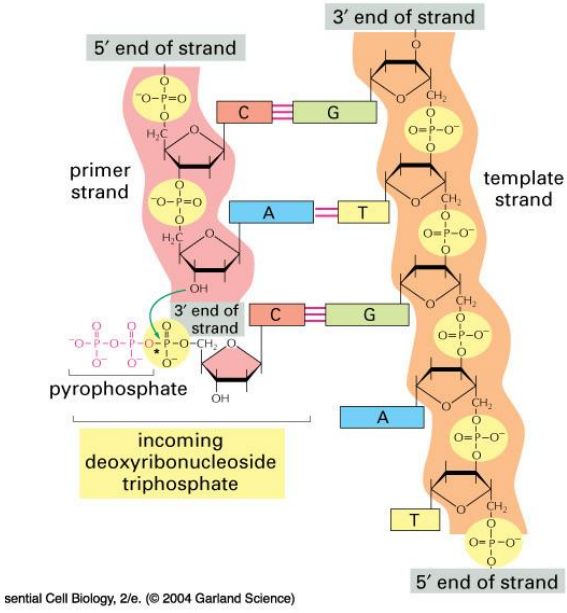



**Figure 17. Direction of Replication**

D. UNI OR BIDIRECTIONAL

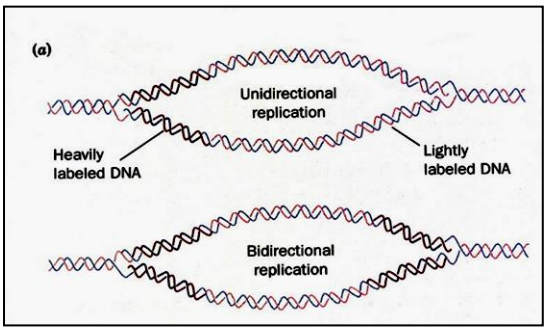



**Figure 18. Uni or Bi Directional Replication**

E. SEMI DISCONTINUOUS REPLICATION

**Anti parallel** strands replicated simultaneously

Leading strand synthesis **continuously** in 5'– 3'

Lagging strand synthesis in **fragments** in 5'-3'

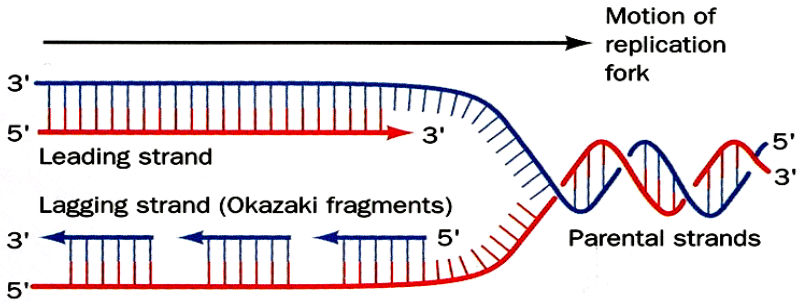



**Figure 19. Semidiscontinuous replication**

F. RNA PRIMERS REQUIRED

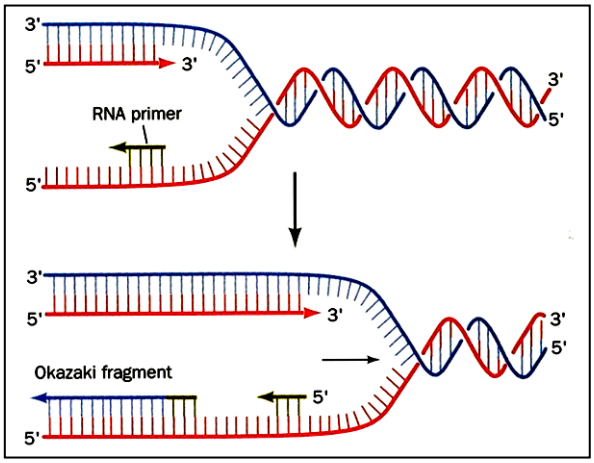



**Figure 20. RNA primers**

THE MECHANISM OF DNA REPLICATION

Arthur Kornberg, a Nobel prize winner and other biochemists deduced steps of replication

- Initiation
  - Proteins bind to DNA and open up double helix
  - Prepare DNA for complementary base pairing
- Elongation
  - Proteins connect the correct sequences of nucleotides into a continuous new strand of DNA
- Termination
  - Proteins release the replication complex

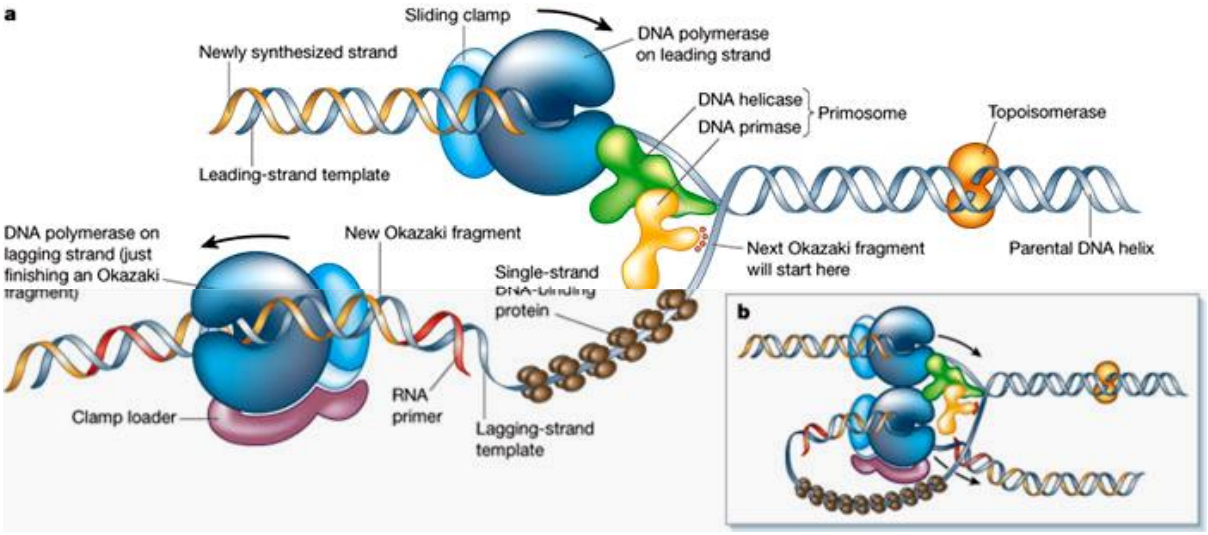



**Figure 21. DNA replication**

Bacterial Replication Requires a Large Number of Enzymes and Proteins

Replication takes place in four stages: initiation, unwinding, elongation, and termination. The following discussion of the process of replication will focus on bacterial systems, where replication has been most thoroughly studied and is best understood. Although many aspects of replication in eukaryotic cells are similar to those in prokaryotic cells, there are some important differences. We will compare bacterial and eukaryotic replication later in the chapter.

Initiation

The circular chromosome of *E. coli* has a single replication origin (*oriC*). The minimal sequence required for *oriC* to function consists of 245 bp that contain several critical sites. An **initiator protein** (known as DnaA in *E. coli*) binds to *oriC* and causes a short section of DNA to unwind. This unwinding allows helicase and other single- strand-binding proteins to attach to the polynucleotide strand.

Unwinding

Because DNA synthesis requires a single-stranded template and because double-stranded DNA must be unwound before DNA synthesis can take place, the cell relies on several proteins and enzymes to accomplish the unwinding.

**DNA helicase** A **DNA helicase** breaks the hydrogen bonds that exist between the bases of the two nucleotide strands of a DNA molecule. Helicase cannot *initiate* the unwinding of double-stranded DNA; the initiator protein first separates DNA strands at the origin, providing a short stretch of single-stranded DNA to which a helicase binds. Helicase binds to the lagging-strand template at each replication fork and moves in the 5'→3′ direction along this strand, thus also moving the replication fork.

**Single-strand-binding proteins** After DNA has been unwound by helicase, **single-strand-binding proteins** (SSBs) attach tightly to the exposed single-stranded DNA. These proteins protect the single-stranded nucleotide chains and prevent the formation of secondary structures such as hairpins (see Figure 10.17) that interfere with replication.

Unlike many DNA-binding proteins, SSBs are indifferent to base sequence: they will bind to any single-stranded DNA. Single- strand-binding proteins form tetramers (groups of four); each tetramer covers from 35 to 65 nucleotides.

**DNA gyrase** Another protein essential for the unwinding process is the enzyme **DNA gyrase**, a topoisomerase. Topoisomerases control the supercoiling of DNA. In replication, DNA gyrase reduces the torsional strain (torque) that builds up ahead of the replication fork as a result of unwinding. It reduces torque by making a double-stranded break in one segment of the DNA helix, passing another segment of the helix through the break, and then resealing the broken ends of the DNA. This action removes a twist in the DNA and reduces the supercoiling.

A group of antibiotics called 4-quinolones kill bacteria by binding to DNA gyrase and inhibiting its action. The inhibition of DNA gyrase results in the cessation of DNA synthesis and bacterial

growth. Many bacteria have acquired resistance to quinolones through mutations in the gene for DNA gyrase.

Elongation

During the elongation phase of replication, single-stranded DNA is used as a template for the synthesis of DNA. This process requires a series of enzymes.

**Synthesis of primers** All DNA polymerases require a nucleotide with a 3′-OH group to which a new nucleotide can be added. Because of this requirement, DNA polymerases cannot initiate DNA synthesis on a bare template; rather, they require a primer—an existing 3′-OH group—to get started. How, then, does DNA synthesis begin?

An enzyme called **primase** synthesizes short stretches of nucleotides, or **primers**, to get DNA replication started. Primase synthesizes a short stretch of RNA nucleotides (about 10–12 nucleotides long), which provides a 3′-OH group to which DNA polymerase can attach DNA nucleotides. (Because primase is an RNA polymerase, it does not require a 3′-OH group to which nucleotides can be added.) All DNA molecules initially have short RNA primers embedded within them; these primers are later removed and replaced by DNA nucleotides. On the leading strand, where DNA synthesis is continuous, a primer is required only at the 5′ end of the newly synthesized strand. On the lagging strand, where replication is discontinuous, a new primer must be generated at the beginning of each Okazaki fragment. Primase forms a complex with helicase at

the replication fork and moves along the template of the lagging strand. The single primer on the leading strand is probably synthesized by the primase–helicase complex on the template of the lagging strand of the *other* replication fork, at the opposite end of the replication bubble.

Two of them, DNA polymerase I and DNA polymerase III, carry out DNA synthesis in replication; the other three have specialized functions in DNA repair. **DNA polymerase III** is a large multiprotein complex that

acts as the main workhorse of replication. DNA polymerase III synthesizes nucleotide strands by adding new nucleotides to the 3′ end of a growing DNA molecule. This enzyme has two enzymatic activities (see Table 12.3). Its 5′→3′ polymerase activity allows it to add new nucleotides in the 5′→3′ direction. Its 3′→5′ exonuclease activity allows it to remove nucleotides in the 3′→5′ direction, enabling it to correct

errors. If a nucleotide having an incorrect base is inserted into the growing DNA molecule, DNA polymerase III uses its 3′→5′ exonuclease activity to back up and remove the incorrect nucleotide.

It then resumes its 5′→3′ polymerase activity. These two functions together allow DNA polymerase III to efficiently and accurately synthesize new DNA molecules.

DNA polymerase III has high processivity, which means that it is capable of adding many nucleotides to the growing DNA strand without releasing the template: it normally holds on to the template and continues synthesizing DNA until the template has been completely replicated. The high processivity of DNA polymerase III is ensured by one of the polypeptides that constitutes the enzyme. This polypeptide, termed the β subunit, serves as a clamp for the polymerase enzyme: it encircles the DNA and keeps the DNA polymerase attached to

the template strand during replication. DNA polymerase III adds DNA nucleotides to the primer, synthesizing the DNA of both the leading and the lagging strands.

The first *E. coli* polymerase to be discovered, **DNA polymerase I**, also has 5′→3′ polymerase and 3′→5′ exonuclease activities (see Table 12.3), permitting the enzyme to synthesize DNA and to correct errors. Unlike DNA polymerase III, however, DNA polymerase I also possesses 5′→3′ exonuclease activity, which is used to remove the primers laid down by primase and to replace them with DNA nucleotides by synthesizing in a 5′→3′ direction. DNA polymerase I has lower processivity than DNA polymerase III. The removal and replacement of primers appear to constitute the main function of DNA polymerase I. After DNA polymerase

III has initiated synthesis at the primer and moved downstream,

DNA polymerase I removes the RNA nucleotides of the primer, replacing them with DNA nucleotides. DNA

polymerases II, IV, and V function in DNA repair.

Despite their differences, all of *E. coli*'s DNA polymerases

**1.** synthesize any sequence specified by the template strand;

**2.** synthesize in the 5′→3′ direction by adding nucleotides to a 3′-OH group;

**3.** use dNTPs to synthesize new DNA;

**4.** require a primer to initiate synthesis;

**5.** catalyze the formation of a phosphodiester bond by joining the 5′-phosphate group of the incoming nucleotide to the 3′-OH group of the preceding nucleotide on the growing strand, cleaving off two phosphates in the process;

**6.** produce newly synthesized strands that are complementary and antiparallel to the template strands; and

**7.** are associated with a number of other proteins.

**DNA ligase** After DNA polymerase III attaches a DNA nucleotide to the 3′-OH group on the last nucleotide of the RNA primer, each new DNA nucleotide then provides the 3′-OH group needed for the next DNA nucleotide to be added. This process continues as long as template is available. DNA polymerase I follows DNA polymerase III and, using its 5′→3′ exonuclease activity, removes the RNA primer.

It then uses its 5′→3′ polymerase activity to replace the RNA nucleotides with DNA nucleotides. DNA polymerase I attaches the first nucleotide to the OH group at the 3′ end of the preceding Okazaki fragment and then continues, in the 5′→3′ direction along the nucleotide strand, removing and replacing, one

at a time, the RNA nucleotides of the primer.

After polymerase I has replaced the last nucleotide of the RNA primer with a DNA nucleotide, a nick remains in

the sugar–phosphate backbone of the new DNA strand. The 3′-OH group of the last nucleotide to have been added by DNA polymerase I is not attached to the 5′-phosphate group of the first nucleotide added by DNA polymerase III. This nick is sealed by the enzyme **DNA ligase**, which catalyzes the formation of a phosphodiester bond without adding another nucleotide to the strand. Some of the major enzymes and proteins required for prokaryotic DNA replication are summarized in **Table**.

**Elongation at the replication fork**

Now that the major enzymatic components of elongation—DNA polymerases, helicase, primase, and ligase—have been introduced, let's consider how these components interact at the replication fork. Because the synthesis of both strands takes place simultaneously, two units of DNA polymerase III must be present at the replication fork, one for each strand. In one model of the replication process, the two units of DNA polymerase III are connected; the lagging-strand template loops around so that it is in position for 5′→3′ replication. In this way, the DNA polymerase III complex is able to carry out 5′→3′ replication simultaneously on both templates, even though they run in opposite directions. After about 1000 bp of new DNA has been synthesized, DNA polymerase III releases the lagging strand template, and a new loop forms (see Figure 12.15).

Primase synthesizes a new primer on the lagging strand and DNA polymerase III then synthesizes a new Okazaki fragment.

In summary, each active replication fork requires five basic components

**1.** helicase to unwind the DNA,

**2.** single-strand-binding proteins to protect the single nucleotide strands and prevent secondary structures,

**3.** the topoisomerase gyrase to remove strain ahead of the replication fork,

**4.** primase to synthesize primers with a 3′-OH group at the beginning of each DNA fragment, and

**5.** DNA polymerase to synthesize the leading and lagging nucleotide strands.

Termination

In some DNA molecules, replication is terminated whenever two replication forks meet. In others, specific termination sequences block further replication. A termination protein, called Tus in *E. coli*, binds to these sequences. Tus blocks the movement of helicase, thus stalling the replication fork and preventing further DNA replication.

The Fidelity of DNA Replication

Overall, the error rate in replication is less than one mistake per billion nucleotides. How is this incredible accuracy achieved?

DNA polymerases are very particular in pairing nucleotides with their complements on the template strand. Errors in nucleotide selection by DNA polymerase arise only about once per 100,000 nucleotides. Most of the errors that do arise in nucleotide selection are corrected in a second process called **proofreading**. When a DNA polymerase inserts an incorrect nucleotide into the growing strand, the 3′-OH group of the mispaired nucleotide is not correctly positioned in the active site of the DNA polymerase for accepting the next

nucleotide. The incorrect positioning stalls the polymerization reaction, and the 3′→5′ exonuclease activity of DNA polymerase removes the incorrectly paired nucleotide. DNA polymerase then inserts the correct nucleotide. Together, proofreading and nucleotide selection result in an error rate of only one in 10 million nucleotides.

A third process, called **mismatch repair** (discussed further in Chapter 18), corrects errors after replication is

complete. Any incorrectly paired nucleotides remaining after replication produce a deformity in the secondary structure of the DNA; the deformity is recognized by enzymes that excise an incorrectly paired nucleotide and use the original nucleotide strand as a template to replace the incorrect nucleotide. Mismatch repair requires the ability to distinguish between the old and the new strands of DNA, because the enzymes need some way of determining which of the two

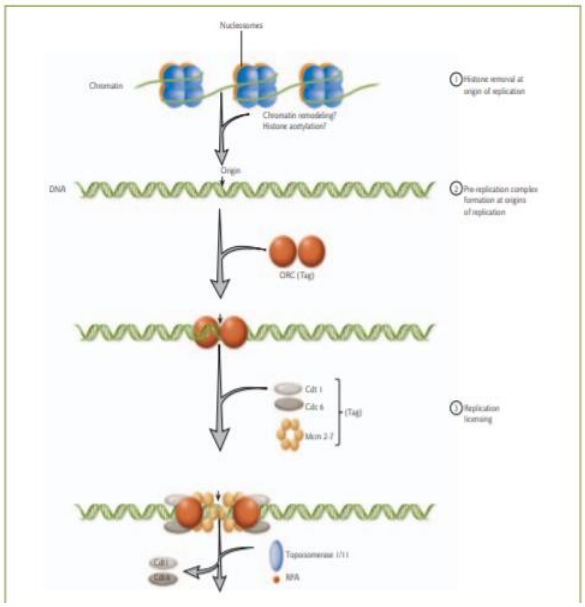
incorrectly paired bases to remove. In *E. coli*, methyl groups (−CH3) are added to particular nucleotide sequences, but only *after* replication. Thus, immediately after DNA synthesis,
only the old DNA strand is methylated. It can therefore be distinguished from the newly synthesized strand, and mismatch repair takes place preferentially on the unmethylated nucleotide strand. No single process could produce this level of accuracy; a series of processes are required, each process catching errors missed by the preceding ones.

The Basic Rules of Replication

Bacterial replication requires a number of enzymes , proteins, and DNA sequences that function together to synthesize a new DNA molecule. These components are important, but we must not become so immersed in the details of the process that we lose sight of the general principles of replication.

1. Replication is always semiconservative.
2. Replication begins at sequences called origins.
3. DNA synthesis is initiated by short segments of RNA called primers.
4. The elongation of DNA strands is always in the 5′→3′ direction.
5. New DNA is synthesized from dNTPs; in the polymerization of
DNA, two phosphate groups are cleaved from a dNTP and the resulting nucleotide is added to the 3′-OH group of the growing nucleotide strand.
6. Replication is continuous on the leading strand and discontinuous on the lagging strand.
7. New nucleotide strands are complementary and antiparallel to their template strands.
8. Replication takes place at very high rates and is astonishingly accurate, thanks to precise nucleotide selection, proofreading,
and repair mechanisms.

**EUKARYOTIC REPLICATION**



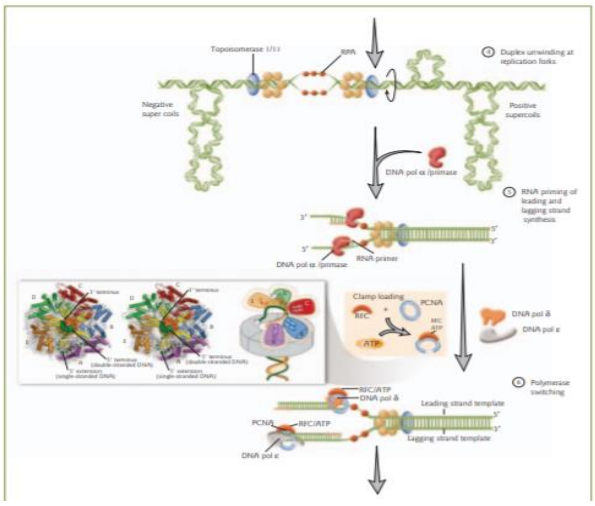
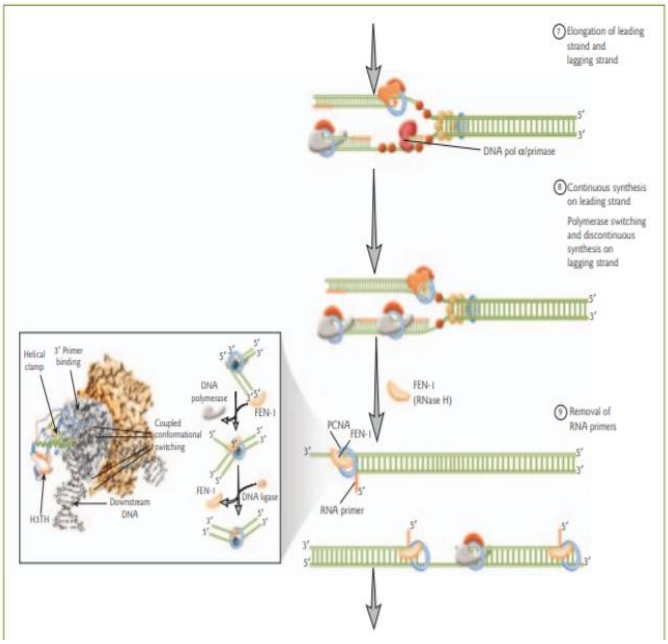




**Figure 22a. Eukaryotic replication**

**Figure 22 b. Eukaryotic replication**



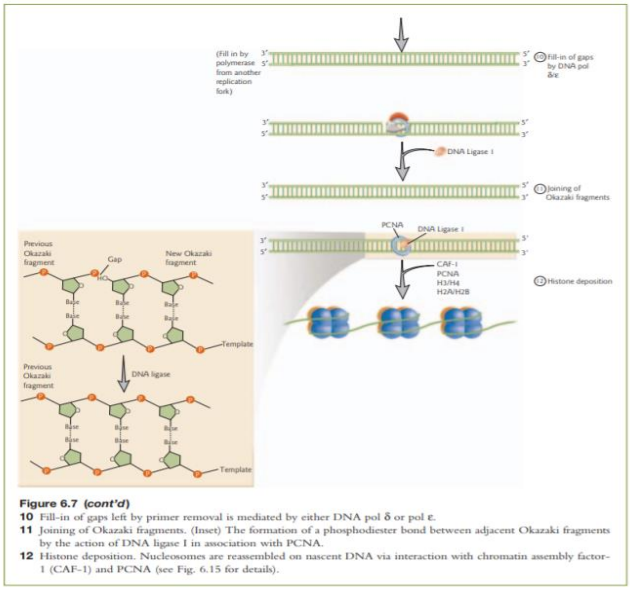




**Figure 6.7 (*cont'd*)**
**10** Fill-in of gaps left by primer removal is mediated by either DNA pol δ or pol ε.
**11** Joining of Okazaki fragments. (Inset) The formation of a phosphodiester bond between adjacent Okazaki fragments by the action of DNA ligase I in association with PCNA.
**12** Histone deposition. Nucleosomes are reassembled on nascent DNA via interaction with chromatin assembly factor-1 (CAF-1) and PCNA (see Fig. 6.15 for details).

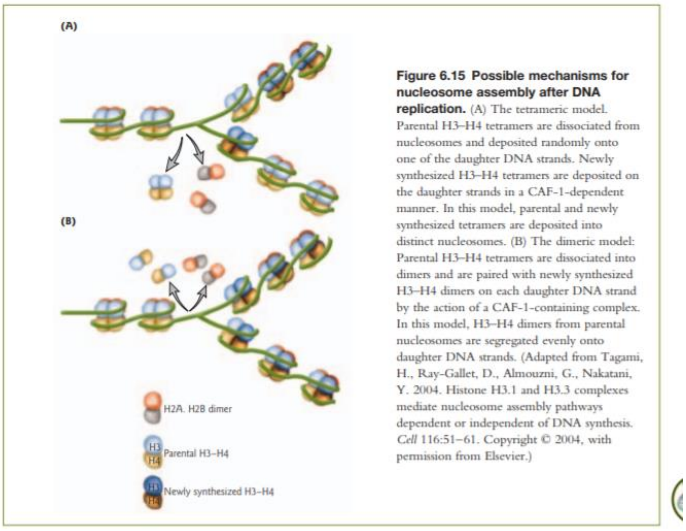



**Figure 6.15 Possible mechanisms for nucleosome assembly after DNA replication.** (A) The tetrameric model. Parental H3–H4 tetramers are dissociated from nucleosomes and deposited randomly onto one of the daughter DNA strands. Newly synthesized H3–H4 tetramers are deposited on the daughter strands in a CAF-1-dependent manner. In this model, parental and newly synthesized tetramers are deposited into distinct nucleosomes. (B) The dimeric model: Parental H3–H4 tetramers are dissociated into dimers and are paired with newly synthesized H3–H4 dimers on each daughter DNA strand by the action of a CAF-1-containing complex. In this model, H3–H4 dimers from parental nucleosomes are segregated evenly onto daughter DNA strands. (Adapted from Tagami, H., Ray-Gallet, D., Almouzni, G., Nakatani, Y. 2004. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* 116:51–61. Copyright © 2004, with permission from Elsevier.)

**Figure 22c. Eukaryotic replication DNA Replication in Eukaryotes**

Eukaryotic DNA replication requires two different DNA polymerase enzymes, namely **DNA polymerase** α and **DNA polymerase** δ. DNA polymerase δ synthesizes the DNA on the leading strand (continuous DNA synthesis), whereas DNA polymerase α synthesizes the DNA on the lagging strand (discontinuous DNA synthesis). Besides these two enzymes, six more factors are involved in eukaryotic DNA replication: (1) T antigen; (2) replication factor A or RF-A (also called RP-A or eukaryotic SSB); (3) topoisomerase I; (4) topoisomerase II; (5) proliferating - cell nuclear antigen (PCNA, also called cyclin), and (6) replication factor Cor RF-C.

The process of eukaryotic DNA replication involves the following steps:

1. Before the onset of DNA synthesis, there is a presynthetic stage of 8-1 0 minutes duration for the formation of unwound DNA complex. This step needs only three purified proteins, namely T antigen (T-ag or tumour antigen), RF-A and topiosomerases I and II.
2. The T-antigen, using its DNA-binding domain, forms a multi-subunit complex with site I and site II in the presence of A TP and caused local unwinding.
3. More extensive duplex unwinding occurs due to association ofRF-A and a topoisomerase with the help of DNA helicase component ofT-ago Topoisomerases help in unwinding of DNA by altering topology of DNA at the replication fork.
4. RF-A or SSB proteins bind to unwound single stranded DNA.
5. The primer RNA synthesis is performed by primase which is tightly associated with DNA Polymerase α.
6. DNA polymerase α helps in synthesis of an okazaki fragment in 5' to 3' direction.

7. Replication factor C (or RF-C) and PCNA (cyclin) help in switching of DNA polymerases so that pol α is replaced by pol δ which then continuously synthesized DNA on the leading strand.
8. Another okazaki fragment is then synthesized from the replication fork on the lagging strand by pol α - primase complex and this step is repeated again and again, till the .entire DNA molecule is covered.
9. The RNA primers are removed and the gaps are filled as in prokaryotic DNA replication. Recently, role of DNA polymerase E in DNA replication has been stressed upon, so that three DNA polymerases (α, β and ε) are now known to be involved in eukaryotic DNA replication.
**A. Sugino** and coworkers have proposed that DNA polymerase a might function at both the leading and lagging strands (since polymerase a has a primase activity), whereas polymerase E and polymerase δ are involved in elongation of the leading and lagging-strands respectively.

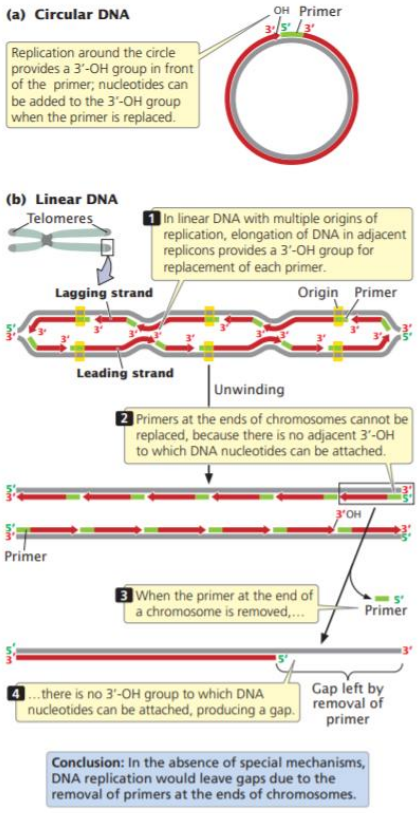
**Replication at the Ends of Chromosomes**

A fundamental difference between eukaryotic and bacterial replication arises because eukaryotic chromosomes are linear and thus have ends.

As already stated, the 3′-OH group needed for replication by DNA polymerases is provided at the initiation of replication by RNA primers that are synthesized by primase.

This solution is temporary because, eventually, the primers must be removed and replaced by DNA nucleotides. In a circular DNA molecule, elongation around the circle eventually provides a 3′-OH group immediately in front of the primer.

After the primer has been removed, the replacement DNA nucleotides can be added to this 3′-OH group.




12.18 DNA synthesis at the ends of circular and linear chromosomes must differ.

**Figure 23. Telomere replication**

**The end-replication problem**

In linear chromosomes with multiple origins, the elongation of DNA in adjacent replicons also provides a 3′-OH group preceding each primer (Figure 12.18b).

At the very end of a linear chromosome, however, there is no adjacent stretch of replicated DNA to provide this crucial 3′-OH group. When the primer at the end of the chromosome has been removed, it cannot be replaced by DNA nucleotides, which produces a gap at the end of the chromosome, suggesting that the chromosome should become progressively shorter with each round of replication.

Chromosome shortening would mean that, when an organism reproduced, it would pass on shorter chromosomes than it had inherited.

Chromosomes would become shorter with each new generation and would eventually destabilize. This situation has been termed the end-replication problem.

Chromosome shortening does in fact take place in many somatic cells but, in single-celled organisms, germ cells, and early embryonic cells, chromosmes do not shorten and self-destruct. So how are the ends of linear chromosomes replicated?

**Telomeres**

To prevent the loss of genes as chromosome ends wear down, the tips of eukaryotic chromosomes have specialized DNA “caps” called **telomeres**. Telomeres consist of hundreds or thousands of repeats of the same short DNA sequence, which varies between organisms but is 5'-TTAGGG-3' in humans and other mammals.

Telomeres protect the end of the chromosome from DNA damage or from fusion with neighbouring chromosomes.

The ends of chromosomes— the telomeres—possess several unique features, one of which is the presence of many copies of a short repeated sequence.

In the protozoan Tetrahymena, this telomeric repeat is TTGGGG, with this G-rich strand typically protruding beyond the C-rich strand: toward ← 5′–TTGGGGTTGGGG–3′ → end of centromere 3′–AACCCC–5′ chromosome

Telomeres need to be protected from a cell's DNA repair systems because they have single-stranded overhangs, which "look like" damaged DNA. The overhang at the lagging strand end of the chromosome is due to incomplete end replication.

A telomere is a region of repetitive sequences at each end of the chromosomes of most eukaryotes.

In some species (including humans), the single-stranded overhangs bind to complementary repeats in the nearby double-stranded DNA, causing the telomere ends to form protective loops ^33 cubed.

Proteins associated with the telomere ends also help protect them and prevent them from triggering DNA repair pathways.

The single-stranded protruding end of the telomere, known as the G overhang can be extended by telomerase, an enzyme with both a protein and an RNA component (also known as a ribonucleoprotein).

## MODELS OF DNA REPLICATION

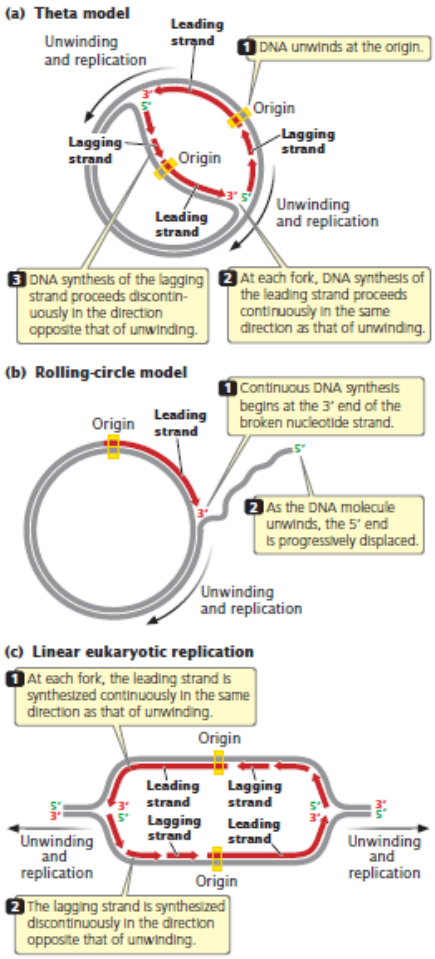



**Figure 24. Models of DNA replication**

### The Direction of Replication in Different Models of Replication

In the theta model, the DNA unwinds at one particular location, the origin, and a replication bubble is formed. If the bubble has two forks, one at each end, synthesis takes place simultaneously at both forks (bidirectional replication). At each fork, synthesis on one of the template strands proceeds in the same direction as that of unwinding; this newly replicated strand is the leading strand with continuous replication. On the other template strand, synthesis proceeds in the direction opposite that of unwinding; this newly synthesized strand is the lagging strand with discontinuous replication. Focus on just one of the template strands within the bubble. Notice that synthesis on this template strand is continuous at one fork but discontinuous at the other. This difference arises because DNA synthesis is always in the same direction (5′→ 3′), but the two forks are moving in opposite directions. Replication in the rolling-circle model (**Figure 12.10b**) is somewhat different,

because there is no replication bubble. Replication begins at the 3′ end of the broken nucleotide strand. Continuous replication takes place on the circular template as new
nucleotides are added to this 3′ end.

The replication of linear molecules of DNA, such as those found in eukaryotic cells, produces a series of replication bubbles. DNA synthesis in these bubbles is the same as that in the single replication bubble of the theta model; it begins at
the center of each replication bubble and proceeds at two forks, one at each end of the bubble. At both forks, synthesis of the leading strand proceeds in the same direction as that of unwinding, whereas synthesis of the lagging strand proceeds in the direction opposite that of unwinding.

# UNIT – III – GENETIC CODE AND TRANSCRIPTION

**RIBONUCLEIC ACID**

- RNA is a polymer of ribonucleotides linked together by 3'-5' phosphodiester linkage.
- The major role of RNA is to participate in protein synthesis, which requires three classes of RNA.

STRUCTURE OF RNA

- Back bone is sugar and phosphate group
- Nitrogenous bases linked to sugar moiety project from the backbone
- Nitrogenous bases are linked to pentose sugar through N-glycosidic linkage to form a nucleoside
- Phosphate group is linked with 3'OH of nucleoside through phosphoester linkage
- 2 nucleotides are linked through 3'-5'-phosphodiester linkage to form a dinucleotide
- More and more such groups will be linked to form a poly nucleotide chain
- Such a polymer has a free phosphate moiety at 5' end of ribose sugar and it is called as 5'-end of polynucleotide chain
- At other end, ribose has free 3'-OH group which is called as the 3'-end of polynucleotide chain
- In RNA, every nucleotide has an additional-OH present at 2'-position of ribose

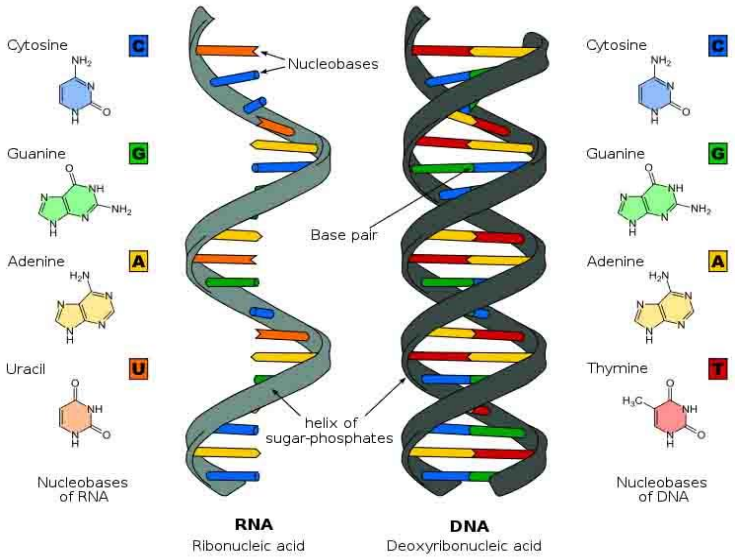



**Figure 25. RNA**

- ❖ Synthesis of RNA is usually catalyzed by an enzyme—RNA polymerase
- ❖ By using DNA as a template
- ❖ The process is known as transcription
- ❖ There are also a number of RNA-dependent RNA polymerases that use RNA as their template for synthesis of a new strand of RNA
- ❖ A number of RNA viruses (such as poliovirus) use this type of enzyme to replicate their genetic material

BASIC PRINCIPLES OF TRANSCRIPTION AND TRANSLATION

- RNA is the bridge between genes and the proteins for which they code
- Transcription is the synthesis of RNA using information in DNA
- Transcription produces messenger RNA (mRNA)
- Translation is the synthesis of a polypeptide, using information in the mRNA
- Ribosomes are the sites of translation
- In prokaryotes, translation of mRNA can begin before transcription has finished
- In a eukaryotic cell, the nuclear envelope separates transcription from translation
- Eukaryotic RNA transcripts are modified through RNA processing to yield the finished mRNA
- A primary transcript is the initial RNA transcript from any gene prior to processing
- The *central dogma* is the concept that cells are governed by a cellular chain of command: DNA → RNA → protein


**Figure 26. Central Dogma**

TYPES OF RNA

In all prokaryotic and eukaryotic organisms, three main classes of RNA molecules exist-

1) Messenger RNA(m RNA)
2) Transfer RNA (t RNA)
3) Ribosomal RNA (r RNA)

The other are –

- small nuclear RNA (SnRNA),
- micro RNA(mi RNA) and
- small interfering RNA(Si RNA) and
- heterogeneous nuclear RNA (hnRNA)

Heterogeneous nuclear RNA (hnRNA)

- In mammalian nuclei , hnRNA is the immediate product of gene transcription
- The nuclear product is heterogeneous in size (Variable) and is very large.
- Molecular weight may be more than $10^{7,}$ while the molecular weight of m RNA is less than $2x\ 10^{6}$
- 75 % of hnRNA is degraded in the nucleus, only 25% is processed to mature m RNA

Messenger RNA

- Messenger RNA (mRNA) carries information about a protein sequence to the ribosomes, the protein synthesis factories in the cell
- It is coded so that every three nucleotides (a codon) correspond to one amino acid
- . In eukaryotic cells, once precursor mRNA (pre-mRNA) has been transcribed from DNA, it is processed to mature mRNA
- This removes its introns—non-coding sections of the pre-mRNA
- The mRNA is then exported from the nucleus to the cytoplasm, where it is bound to ribosomes and translated into its corresponding protein form with the help of tRNA

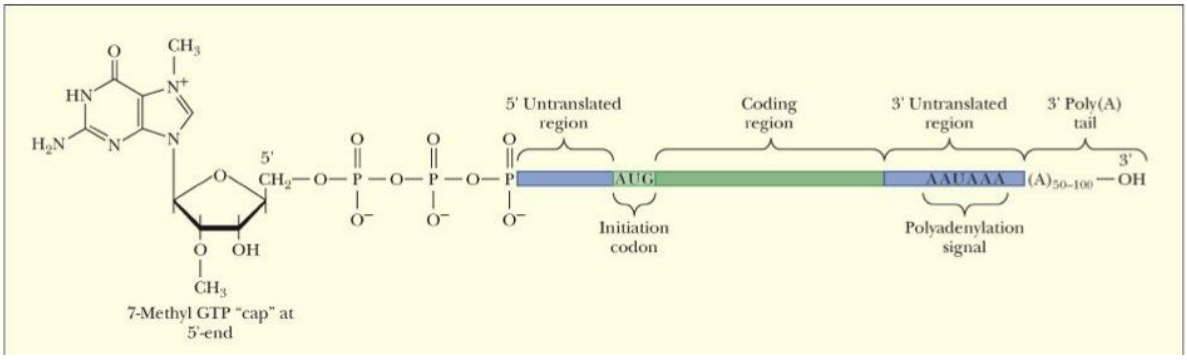



**Figure 27. Structure of mRNA**

Ribosomal RNA

The mammalian ribosome contains two major nucleoprotein subunits—a larger one with a molecular weight of 2.8 x $10^6$ (60S) and a smaller subunit with a molecular weight of 1.4 x $10^6$ (40S).

- Is a structural and functional component of ribosomes w/ are "platforms" on which protein synthesis occur.
- Consist of about 35% protein and 65% ribosomal RNA.
- Complexed w/ proteins, the rRNA forms the cellular structures called the ribosomes.
- Ribosomal RNA (rRNA) is the catalytic component of the ribosomes
- Eukaryotic ribosomes contain four different rRNA molecules: 18S, 5.8S, 28S and 5S rRNA
- Three of the rRNA molecules are synthesized in the nucleolus, and one is synthesized elsewhere.
- In the cytoplasm, ribosomal RNA and protein combine to form a nucleoprotein called a ribosome
- The ribosome binds mRNA and carries out protein synthesis
- Several ribosomes may be attached to a single mRNA at any time.
- Nearly all the RNA found in a typical eukaryotic cell is rRNA.

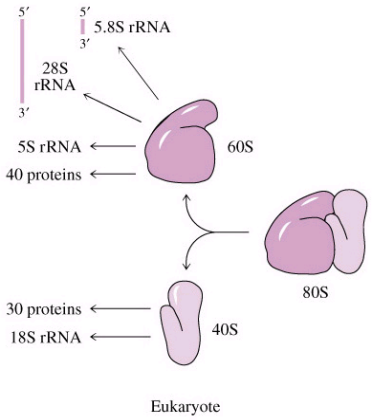



**Figure 28. rNA**

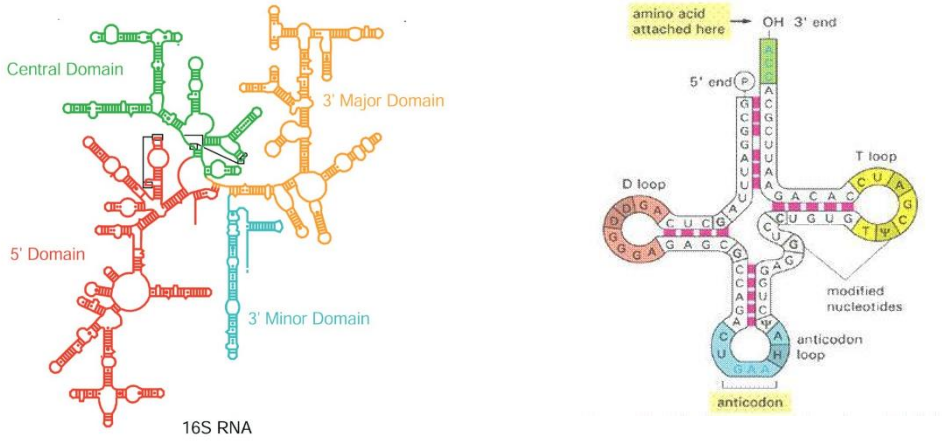



**Figure 29.** tRNA

1) Primary structure- The nucleotide sequence of all the t RNA molecules allows extensive intrastand complimentarity that generates a secondary structure.

2) Secondary structure- Each single t- RNA shows extensive internal base pairing and acquires a clover leaf like structure. The structure is stabilized by hydrogen bonding between the bases and is a consistent feature.

- ❑ The L shaped tertiary structure is formed by further folding of the clover leaf due to hydrogen bonds between T and D arms.
- ❑ The base paired double helical stems get arranged in to two double helical columns, continuous and perpendicular to one another.

Small nuclear RNA (SiRNA)

- ❑ Most of these molecules are complexed with proteins to form small nuclear ribonucleoproteins particles{snurps} and are distributed in the nucleus, in the cytoplasm, or in both.
- ❑ They range in size from 20 to 300 nucleotides and are present in 100,000–1,000,000 copies per cell.

Function:

- ✓ Help w/ the processing of the initial mRNA into mature form that is ready for export out of the nucleus-{ splicing}.
- ❑ Of the several snRNAs, U1, U2, U4, U5, and U6 are involved in intron removal and the processing of hnRNA into mRNA
- ❑ The U7 snRNA is involved in production of the correct 3' ends of histone mRNA—which lacks a poly(A) tail.

MicroRNA, mi RNA, Small interferingRNA, siRNA

- ❑ These two classes of RNAs represent a subset of small RNAs; both play important roles in gene regulation.
- ❑ miRNAs and siRNAs cause inhibition of gene expression by decreasing specific protein production albeit apparently via distinct mechanisms

Five different types of RNA, each encoded by different genes:

1. mRNA — Messenger RNA, encodes the amino acid sequence of a polypeptide.
2. tRNA — Transfer RNA, transports amino acids to ribosomes during translation.
3. rRNA — Ribosomal RNA, forms complexes called ribosomes with protein, the structure on which mRNA is translated.

| | |
|---|---|
| 4. snRNA | Small nuclear RNA, forms complexes with proteins used in eukaryotic RNA processing (e.g., exon splicing and intron removal). |
| 5. miRNA/siRNA | Micro RNA/small interfering RNA, short ~22 nt RNA sequences that bind to 3' UTR target mRNAs and result in gene silencing. |

Transcription

When a protein is needed by a cell, the genetic code for that protein must be read from the DNA and processed.

A two step process:

1. Transcription = synthesis of a single-stranded RNA molecule using the DNA template (1 strand of DNA is transcribed).
2. Translation = conversion of a messenger RNA sequence into the amino acid sequence of a polypeptide (i.e., protein synthesis)

✓ Both processes occur throughout the cell cycle. Transcription occurs in the nucleus, whereas translation occurs in the cytoplasm.

Materials required

✓ The enzyme RNA polymerase or DNA directed RNA polymerase

✓ DNA template – the transcription unit

✓ All the four types of ribonucleoside triphosphates (ATP, CTP, GTP and UTP)

✓ Divalent metal ions $Mg^{2+}$ or $Mn^{2+}$ as a cofactor

✓ No primer is needed for RNA synthesis

**Transcription: How is an RNA strand synthesized?**

1. **Regulated by gene regulatory elements within each gene.**
2. **DNA unwinds next to a gene.**

3. **RNA is transcribed 5' to 3' from the template (3' to 5').**

4. **Similar to DNA synthesis, except:**

    - ✓ **NTPs instead of dNTPs (no deoxy-)**
    - ✓ **No primer**
    - ✓ **No proofreading**
    - ✓ **Adds Uracil (U) instead of thymine (T)**
    - ✓ **RNA polymerase**

RNA POLYMERASE

α - Two α polypeptide chains, coded by gene rpoA, assembly of core enzyme & and help in the probably in the recognition of promoter

β – one copy of β subunit, , coded by gene rpoB, binds with the incoming nucleotides & helps in the formation of the first phosphodiester bond

β' – one copy of β' subunit, , coded by gene rpoC, binds with the template strand or antisense DNA strand

σ - Single polypeptide chain, loosely attached to the core enzyme.

σ Subunit

- Recognises the start signal on DNA molecules and directs the core enzyme of RNA polymerase to bind to the promoter region upstream of initiation codon
- Recognises two special sequences of bases in the promoter region of the coding strand (i.e. Antitemplate strand) of DNA - -10 sequence and -35 sequence
- Recognises of promoter sequence, facilitates opening or melting of DNA helix
- Separates from core enzyme once about 10nt are joined to initiate RNA synthesis

FUNCTIONS OF RNA POLYMERASE

- Unwinds about 15 bases of DNA around the initiation site to form an open promoter – DNA complex and provides single strand of DNA to act as template for transcription

- Catalyses the formation of phosphodiester bonds between successive nt of a polynucleotide chain during synthesis of RNA
- Lacks proof reading 3'-5' exonuclease activity
- Therefore, one error for every $10^4$ to $10^5$ ribonucleotides incorporated is introduced during RNA transcription
- But the mistake during transcription is not serious because of its high turnover and Wobble pairing during translation

**Three Steps to Transcription:**

1. **Initiation**
2. **Elongation**
3. **Termination**

✓ **Occur in both prokaryotes and eukaryotes.**

✓ **Elongation is conserved in prokaryotes and eukaryotes.**

✓ **Initiation and termination proceed differently.**

**INITIATION**

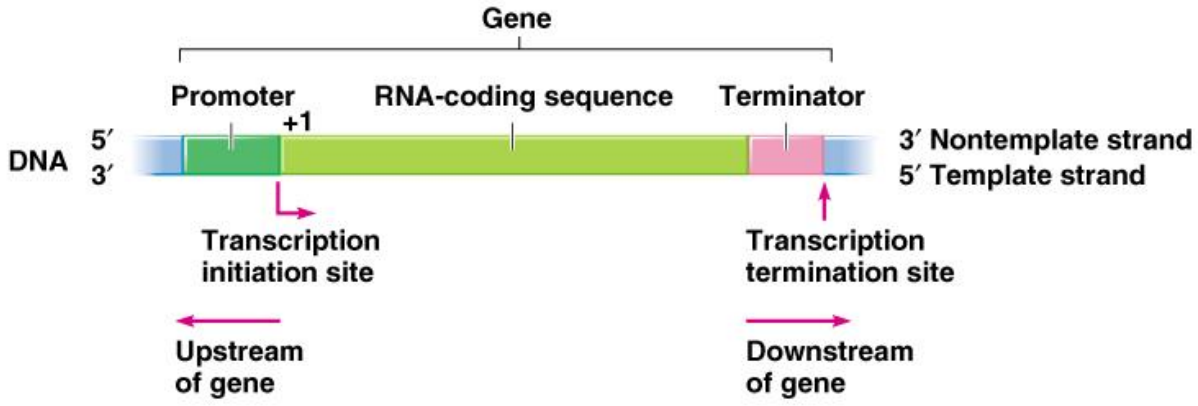



**Figure 30. Transcription Unit**

Each gene has three regions:

1. 5' Promoter, attracts RNA polymerase

-10 bp 5'-TATAAT-3'

-35 bp 5'-TTGACA-3'

2. Transcribed sequence (transcript) or RNA coding sequence
3. 3' Terminator, signals the stop point

Step 1-Initiation, *E. coli* model:

1. RNA polymerase combines with sigma factor (a polypeptide) to create RNA polymerase holoenzyme
    - ✓ Recognizes promoters and initiates transcription.
    - ✓ Sigma factor required for efficient binding and transcription.
    - ✓ Different sigma factors recognize different promoter sequences.
2. RNA polymerase holoenzyme binds promoters and untwists DNA
    - ✓ Binds loosely to the -35 promoter (DNA is d.s.)
    - ✓ Binds tightly to the -10 promoter and untwists
3. Different types and levels of sigma factors influence the level and dynamics of gene expression (how much and efficiency).

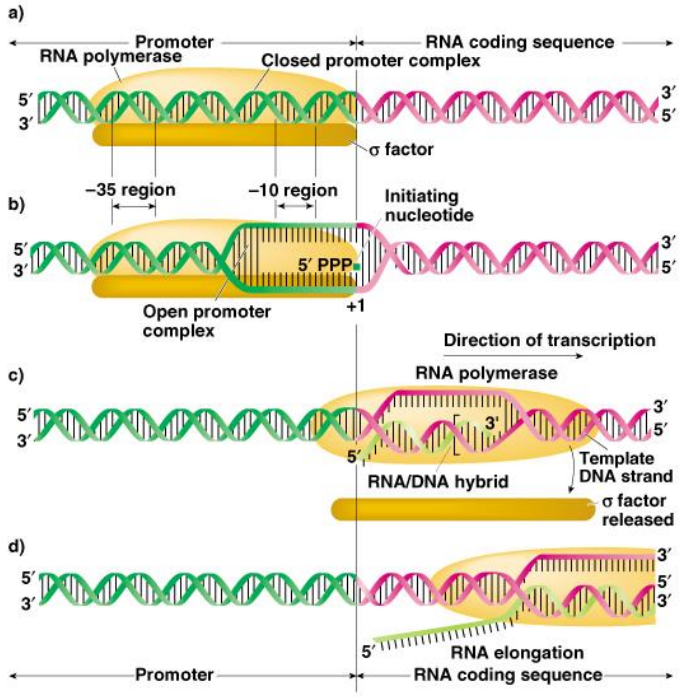



**Figure 31. Transcription**

Step 2-Elongation, *E. coli* model:

1. After 8-9 bp of RNA synthesis occurs, sigma factor is released and recycled for other reactions.
2. RNA polymerase completes the transcription at 30-50 bp/second (and order of magnitude slower than DNA polymerase).
3. DNA untwists rapidly, and re-anneals behind the enzyme.
4. Part of the new RNA strand is hybrid DNA-RNA, but most RNA is displaced as the helix reforms.

Step 3-Termination, *E. coli* model:

Two types of terminator sequences occur in prokaryotes:

1. Type I (ρ-independent)

Palindromic, inverse repeat forms a hairpin loop and is believed to physically destabilize the DNA-RNA hybrid.

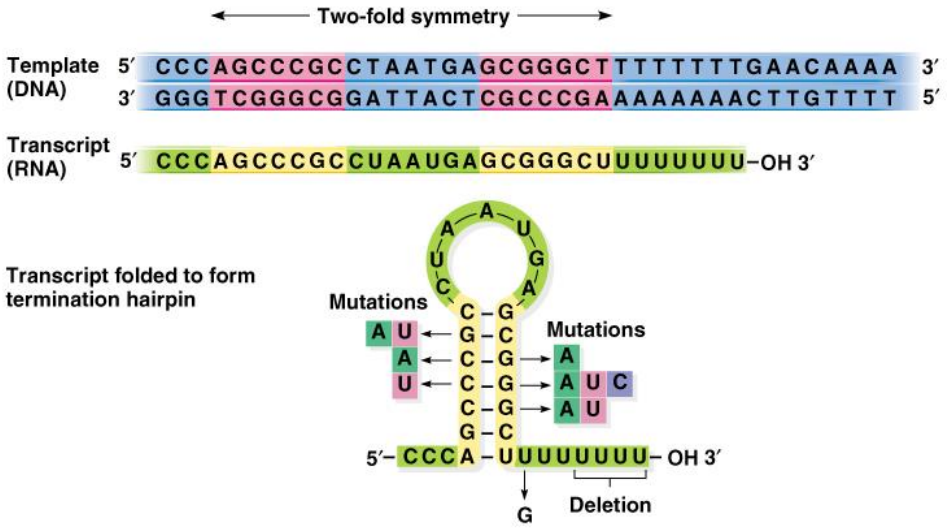



**Figure 32. Transcription Termination**

2. Type II (ρ-dependent)

Involves ρ factor proteins that break the hydrogen bonds between the template DNA and RNA.

**The Basic Rules of Transcription**

Before we examine the process of eukaryotic transcription, let's pause to summarize some of the general principles of bacterial transcription.

1. Transcription is a selective process; only certain parts of the DNA are transcribed at any one time.

2. RNA is transcribed from single-stranded DNA. Within a gene, only one of the two DNA strands—the template strand—is normally copied into RNA.

3. Ribonucleoside triphosphates are used as the substrates in RNA synthesis. Two phosphate groups are cleaved from a ribonucleoside triphosphate, and the resulting nucleotide is joined to the 3′-OH group of the growing RNA strand.

4. RNA molecules are antiparallel and complementary to the DNA template strand. Transcription is always in the 5′→3′ direction, meaning that the RNA molecule grows at the 3′ end.

5. Transcription depends on RNA polymerase—a complex, multimeric enzyme. RNA polymerase consists of a core enzyme, which is capable of synthesizing RNA, and other subunits that may join transiently to perform additional functions.

6. A sigma factor enables the core enzyme of RNA polymerase to bind to a promoter and initiate transcription.

7. Promoters contain short sequences crucial in the binding of RNA polymerase to DNA; these consensus sequences are interspersed with nucleotides that play no known role in transcription.

8. RNA polymerase binds to DNA at a promoter, begins transcribing at the start site of the gene, and ends transcription after a terminator has been transcribed.

## EUKARYOTIC TRANSCRIPTION

**Table 5-3.** Properties of three different eukaryotic nuclear RNA polymerases.

| Enzymes | Location | Product and abundance | Sensitivity to α - amanitin |
|---|---|---|---|
| 1. RNA polymerase I | Nucleolus | rRNA (50–70%) (except 5S rRNA) | Not sensitive |
| 2. RNA polymerase II | Nucleoplasm | hnRNA (mRNA) (20–40%) | Sensitive |
| 3. RNA polymerase III | Nucleoplasm | tRNA(~ 10%) (and 5S rRNA) | Inhibited in animals at high levels; not in yeast and insects. |

Most eukaryotic cells

possess three distinct types of RNA polymerase, each of which is responsible for transcribing a different class of RNA: **RNA polymerase I** transcribes rRNA; **RNA polymerase II** transcribes pre-mRNAs, snoRNAs, some miRNAs, and some snRNAs; and **RNA polymerase III** transcribes other small RNA molecules—specifically tRNAs, small rRNA, some miRNAs, and some snRNAs (**Table 13.3**). RNA polymerases I, II, and III are found in all eukaryotes. Two additional RNA polymerases, **RNA polymerase IV** and **RNA polymerase V**, have been found in plants. RNA polymerases IV and V transcribe RNAs that play a role in DNA methylation and chromatin structure.

All eukaryotic polymerases are large, multimeric enzymes, typically consisting of more than a dozen subunits. Some sub- units are common to all RNA polymerases, whereas others are limited to one of the polymerases. As in bacterial cells, a number of accessory proteins bind to the core enzyme and affect its function.

**Table 11.2 Eukaryotic promoter elements.**

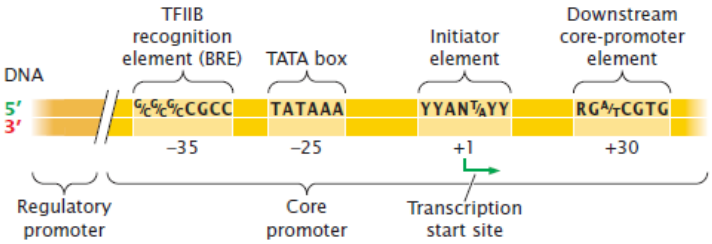
| Promoter | Position | Transcription factor | Consensus sequence |
|---|---|---|---|
| **Upstream core promoter elements** | | | |
| **TFIIB recognition element (BRE)** | −37 to −32 | TFIIB | (G/C)(G/C)(G/A)CGCC |
| **TATA box** | −31 to −26 | TBP | TATA(A/T)AA(G/A) |
| **Initiator (Inr)** | −2 to +4 | TAF1 ($TAF_{II}250$)<br>TAF2 ($TAF_{II}150$) | $PyPyA_{+1}N(T/A)PyPy$ |
| **Downstream core promoter elements** | | | |
| **Motif ten element (MTE)** | +18 to +27 | TFIID | C(G/A)A(A/G)C(G/C)<br>(C/A/G)AACG(G/C) |
| **Downstream promoter element (DPE)** | +28 to +32 | TAF9 ($TAF_{II}40$)<br>TAF6 ($TAF_{II}60$) | (A/G)G(A/T)(C/T)(G/A/C) |
| **Proximal promoter elements** | | | |
| **CAAT box** | −200 to −70 | CBF, NF1, C/EBP | CCAAT |
| **GC box** | −200 to −70 | Sp1 | GGGCGG |

Most, but not all, CAAT and GC boxes are located between −200 and −70.
CBF, CAAT-binding protein; C/EBP, CAAT/enhancer-binding protein; N, any (A, T, C, or G); Py, pyrimidine (C or T).

| | | | |
|---|---|---|---|
| **General transcription machinery** | RNA pol II (pol II, RNAPII) | RNA polymerase II | Catalysis of RNA synthesis |
| | General transcription factors: | | |
| | TFIIB | Transcription factor for RNA polymerase II B | Stabilization of TBP–DNA interactions, recruitment of RNA pol II–TFIIF, start site selection by RNA pol II |
| | TFIID: | | |
| | TBP | TATA-binding protein | Core promoter recognition, TFIIB recruitment |
| | TAF | TBP-associated factor | Core promoter recognition/selectivity |
| | TFIIE | Transcription factor for RNA polymerase II E | TFIIH recruitment |
| | TFIIF | Transcription factor for RNA polymerase II F | Recruitment of RNA pol II to promoter DNA–TBP–TFIIB complex |
| | TFIIH | Transcription factor for RNA polymerase II H | Promoter melting, helicase, RNA pol II CTD kinase |
| | Mediator | Mediator | Transduces regulatory information from activator and repressor proteins to RNA pol II |

**Table 11.1 Proteins that regulate transcription.**

| Category | Acronym | Derivation of name | Function |
|---|---|---|---|
| **Transcription factors (activators or repressors)** | Some examples mentioned in text: | | |
| | CBF | CAAT binding factor | Binds CAAT box |
| | C/EBP | CAAT/enhancer-binding protein | Binds CAAT box |
| | CREB | cAMP response element-binding protein | Binds the cAMP response element |
| | CTCF | CCCTC-binding factor | Binds insulator element (CCCTC) and mediates enhancer blocking activity |
| | FOG-1 | Friend of GATA-1 | Required for developmental expression of β-globin genes |
| | GATA-1 | GATA-binding protein | Required for developmental expression of β-globin genes |
| | NF-E2 | Nuclear factor erythroid-derived 2 | Required for developmental expression of β-globin genes |
| | NF-κB | Nuclear factor of kappa light polypeptide enhancer in B cells | Central mediator of human stress and immune responses |
| | USF1, USF2 | Upstream stimulatory factor 1 and 2 | Bind insulator element, recruit chromatin-modifying enzymes |
| | SATB1 | Special AT-rich binding protein 1 | Matrix attachment region (MAR) binding protein required for T-cell-specific gene regulation |
| | Sp1 | SV40 early and late promoter-binding protein 1 | Binds GC box |
| **General transcription machinery** | RNA pol II (pol II, RNAPII) | RNA polymerase II | Catalysis of RNA synthesis |
| | General transcription factors: | | |
| | TFIIB | Transcription factor for RNA polymerase II B | Stabilization of TBP–DNA interactions, recruitment of RNA pol II–TFIIF, start site selection by RNA pol II |
| | TFIID: | | |
| | TBP | TATA-binding protein | Core promoter recognition, TFIIB recruitment |
| | TAF | TBP-associated factor | Core promoter recognition/selectivity |
| | TFIIE | Transcription factor for RNA polymerase II E | TFIIH recruitment |
| | TFIIF | Transcription factor for RNA polymerase II F | Recruitment of RNA pol II to promoter DNA–TBP–TFIIB complex |
| | TFIIH | Transcription factor for RNA polymerase II H | Promoter melting, helicase, RNA pol II CTD kinase |
| | Mediator | Mediator | Transduces regulatory information from activator and repressor proteins to RNA pol II |
| **Coactivators and corepressors** | Chromatin modification complexes: | | |
| | HAT | Histone acetyltransferase | Acetylates histones |
| | HDAC | Histone deacetylase | Deacetylates histones |
| | CBP | CREB-binding protein | HAT activity |
| | HMT | Histone methyltransferase | Methylates histones |
| | LSD1 | Lysine-specific demethylase 1 | Demethylates histones |
| | Chromatin remodeling complexes: | | |
| | SWI/SNF | Mating-type switching defective/sucrose nonfermenters | ATP-dependent chromatin remodeling (sliding and disassembly) |
| | ISWI | Imitation Swi2 | ATP-dependent chromatin remodeling (sliding) |
| | SWR1 | Swi2/Snf2-related 1 | ATP-dependent chromatin remodeling (histone replacement) |
| **Elongation factors** | FACT | Facilitates chromatin transcription | Transcription-dependent nucleosome alterations |
| | Elongator | Elongator | Exact function in elongation unknown |
| | TFIIS | Transcription factor for RNA polymerase II S | Facilitates RNA pol II passage through regions that cause transcriptional arrest |




**13.15 The promoters of genes transcribed by RNA polymerase II consist of a core promoter and a regulatory promoter that contain consensus sequences.** Not all the consensus sequences shown are found in all promoters.

**Figure 33. Transcription promoters**

A significant difference between bacterial and eukaryotic transcription is the existence of three different eukaryotic RNA polymerases, which recognize different types of promoters. In bacterial cells, the holoenzyme (RNA polymerase  plus the sigma factor) recognizes and binds directly to sequences in the promoter. In eukaryotic cells, promoter recognition is carried out by accessory proteins that bind to the promoter and then recruit a specific RNA polymerase (I, II, or III) to the promoter.

One class of accessory proteins comprises **general transcription factors**, which, along with RNA polymerase, form the **basal transcription apparatus**—a group of proteins that assemble near the start site and are sufficient to initiate minimal levels of transcription. Another class of accessory proteins consists of **transcriptional activator proteins**, which bind to specific DNA sequences and bring about higher levels of transcription by stimulating the assembly of the basal transcription apparatus at the start site.

A promoter for a gene transcribed by RNA polymerase II typically consists of two primary parts: the core

promoter and the regulatory promoter.

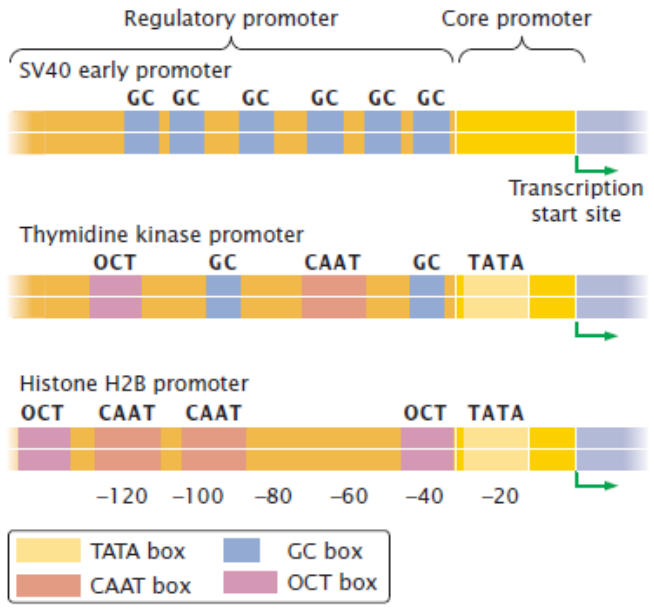
**Core promoter** The **core promoter** is located immediately upstream of the gene and is the site to which the basal transcription apparatus binds. The core promoter typically includes one or more consensus sequences. One of the most common of these sequences is the **TATA box**, which has the consensus sequence TATAAA and is located from −25 to −30 bp upstream of the start site. Additional consensus sequences that may be found in the core promoters of genes transcribed by RNA polymerase II are shown in Figure 13.15. These consensus sequences are recognized by transcription factors that bind to them and serve as a platform for the assembly of the basal transcription apparatus.

**Regulatory promoter** The **regulatory promoter** is located immediately upstream of the core promoter. A variety of different consensus sequences can be found in the regulatory promoters,

and they can be mixed and matched in different combinations. Transcriptional activator proteins bind to these sequences and either directly or indirectly make contact with the basal transcription apparatus and affect the rate at which transcription is initiated.

Transcriptional activator proteins also regulate transcription by binding to more-distant sequences called **enhancers**. The DNA between an enhancer and the promoter loops out, and so transcriptional activator proteins bound to the enhancer can interact with the basal transcription machinery at the core promoter.

**Polymerase I and III promoters** RNA polymerase I and RNA polymerase III each recognize promoters that are distinct from those recognized by RNA polymerase II. For example, promoters for small rRNA and tRNA genes, transcribed by RNA polymerase III, contain **internal promoters** that are downstream of the start site and are transcribed into the RNA.




**13.16 Consensus sequences in the promoters of three eukaryotic genes.** These promoters illustrate the principle that consensus sequences can be mixed and matched in different combinations to yield a functional eukaryotic promoter.

**Figure 34. Consensus sequences**

## Initiation of Eukaryotic Transcription

For the eukaryotic transcription the regulatory DNA sequences (such as promoters, enhancers and silencers) for genes transcribed by each of the three RNA polymerases differ. Various transcription factors are also involved in the formation of a transcription complex which are needed for initiation of transcription. Generally, each of RNA polymerase is believed to have its own set

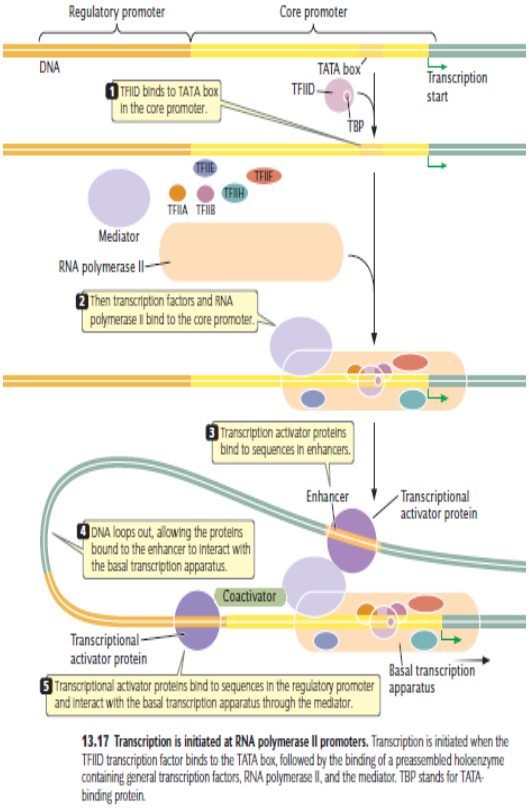
of transcription factors, however, TF II D or a part of it (*e.g.*, TBP=TATA binding protein) is required for all the three RNA polymerases.

The **transcription factors** (**TFs**) can be defined as proteins, which are needed for initiation of transcription, but are not part of the RNA polymerase. They help in DNA binding of a RNA polymerase to constitute the so-called **pre-initiation complex** or **transcription complex**. After the formation of this complex initiation of transcription occurs. All known transcription factors may recognize either **DNA sequences, another factor or RNA polymerase.**

**Formation of transcriptosome with RNA pol II.**

A promoter sequence which is responsible for constitutive expression of common genes (also **called house keeping genes**) in all cells, is called **generic promoter.**

The generic promoter cannot bring about regulated expression (i.e., tissue or stimulus specific expression of genes, called **luxary genes**).




**13.17 Transcription is initiated at RNA polymerase II promoters.** Transcription is initiated when the TFIID transcription factor binds to the TATA box, followed by the binding of a preassembled holoenzyme containing general transcription factors, RNA polymerase II, and the mediator. TBP stands for TATA-binding protein.

**Figure 35. Transcription initiation**

Initiation of transcription on the generic promoter by RNA polymerase II requires the action of diverse transcription factors (TFs) in the following order :

(i) **TF II D** binds at TATA box;

(ii) the step (i) permits the association of **TF IIA** and **TF IIB**;

(iii) TF II B forms the so-called DB complex and RNA polymerase II associates to promoter site;

(iv) RNA pol II is accompanied to the promoter by **TF II F** to form a transcription complex ;

(v) orderly addition of **TF II E, TF II H** and **TF II J** helps the initiation process.

**Elongation of RNA Chain in Eukaryotes**

There are certain accessory proteins of transcription, called **elongation factors**, which enhance the overall activity of RNA polymerase II and lead to increase in the elongation rate. At least two such proteins are known:

(1) The **TF II F** accelerates RNA chain growth relatively uniformly in concord with RNA polymerase II.

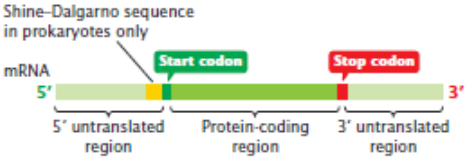
(2) The **TF II S** (also called **S II**) helps in elongation of RNA chain by unburdening the obstruction in the path of such elongation. TF II S is known to act by first causing hydrolytic cleavage at 3´ end of RNA chain, thereby, helping in the forward movement of RNA polymerase through any block to elongation.

**Termination of Eukaryotic Transcription**

(1) In eukaryotes, the actual termination of RNA polymerase II activity during transcription may take place through termination sites similar to those found in prokaryotes.

(2) However, the nature of individual sites is not known. Such termination sites are believed to be present away (sometimes up to one kilobase away from the site of the 3´end of mRNA).

## POST TRANSCRIPTIONAL MODIFICATIONS




**14.5** Three primary regions of mature mRNA are the 5′ untranslated region, the protein-coding region, and the 3′ untranslated region.

**Figure 36. Mature mRNA**

Pre-mRNA Processing

In bacterial cells, transcription and translation take place simultaneously; while the 3′ end of an mRNA is undergoing transcription, ribosomes attach to the Shine–Dalgarno sequence near the 5′ end and begin translation. Because transcription and translation are coupled, bacterial mRNA has little opportunity to be modified before protein synthesis. In contrast, transcription and translation are separated in both time and space in eukaryotic cells. Transcription takes place in the nucleus, whereas translation takes place in the cytoplasm; this separation provides an opportunity for eukaryotic RNA to be modified before it is translated. Indeed, eukaryotic mRNA is extensively altered after transcription. Changes are made to the 5′ end, the 3′ end, and the protein coding section of the RNA molecule.

**Table 14.2** Posttranscriptional modifications to eukaryotic pre-mRNA

| Modification | Function |
|---|---|
| Addition of 5′ cap | Facilitates binding of ribosome to 5′ end of mRNA, increases mRNA stability, enhances RNA splicing |
| 3′ cleavage and addition of poly(A) tail | Increases stability of mRNA, facilitates binding of ribosome to mRNA |
| RNA splicing | Removes noncoding introns from pre-mRNA, facilitates export of mRNA to cytoplasm, allows for multiple proteins to be produced through alternative splicing |
| RNA editing | Alters nucleotide sequence of mRNA |

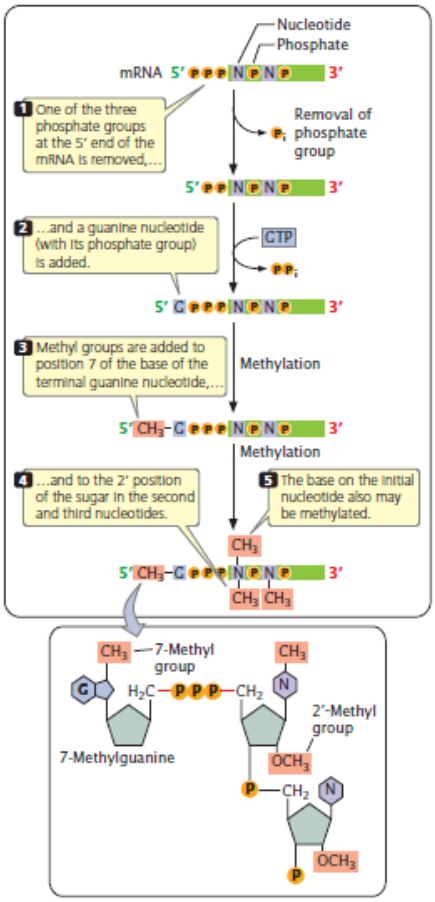
The Addition of the 5′ Cap

One type of modification of eukaryotic pre-mRNA is the addition at its 5′ end of a structure called a **5′ cap**. The cap consists of an extra nucleotide at the 5′ end of the mRNA and methyl groups

(CH3) on the base in the newly added nucleotide and on the 2′-OH group of the sugar of one or more nucleotides at the 5′ end (**Figure 14.6**). The addition of the cap takes place rapidly after the initiation of transcription and, as will be discussed in more depth in Chapter 15, the 5′ cap functions in the initiation of translation. Cap-binding proteins recognize the cap and attach to it; a ribosome then binds to these proteins and moves downstream along the mRNA until the start codon is reached and translation begins. The presence of a 5′ cap also increases the stability of mRNA and influences the removal of introns.

Three phosphate groups are present at the 5′ end of all RNA molecules because phosphate groups are not cleaved from the first ribonucleoside triphosphate in the transcription reaction. The 5′ end of pre-mRNA can be represented as 5′–pppNpNpN . . . , in which the letter "N" represents a ribonucleotide and "p" represents a phosphate. Shortly after the initiation of transcription, one of these phosphate groups is removed and a guanine nucleotide is added. This guanine nucleotide is attached to the premRNA by a unique 5′–5′ bond, which is quite different from the usual 5′–3′ phosphodiester bond that joins all the other nucleotides in RNA. One or more methyl groups are then added to the 5′ end; the first of these methyl groups is added

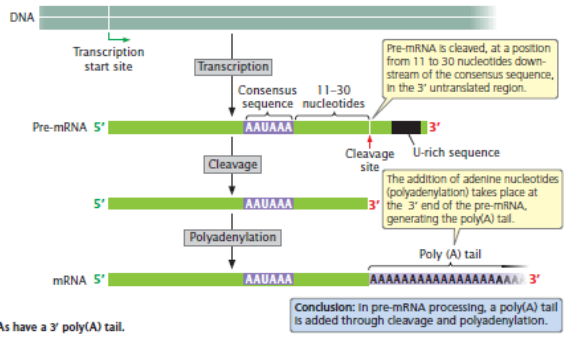
to position 7 of the base of the terminal guanine nucleotide, making the base 7-methylguanine. Next, a methyl group may be added to the 2′ position of the sugar in the second and third nucleotides, as shown in Figure 14.6. Rarely, additional methyl groups may be attached to the bases of the second and third nucleotides of the pre-mRNA. Several different enzymes take part in the addition of the 5′ cap. The initial step is carried out by an enzyme that associates with RNA polymerase II. Because neither RNA polymerase I nor RNA polymerase III have this associated enzyme, RNA molecules transcribed by these polymerases (rRNAs, tRNAs, and some snRNAs) are not capped.




**14.6 Most eukaryotic mRNAs have a 5′ cap.** The cap consists of a nucleotide with 7-methylguanine attached to the pre-mRNA by a unique 5′–5′ bond (shown in detail in the bottom box).

**Figure 37. 5'Cap of mRNA**

The Addition of the Poly(A) Tail




**14.7 Most eukaryotic mRNAs have a 3′ poly(A) tail.**

**Figure 38. Poly (A) tail of mRNA**

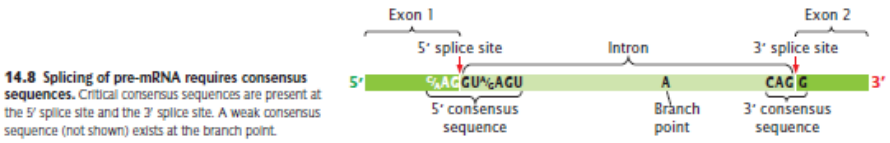
A second type of modification to eukaryotic mRNA is the addition of 50 to 250 or more adenine nucleotides at the 3′ end, forming a **poly(A) tail**. These nucleotides are not encoded in the DNA but are added after transcription in a process termed polyadenylation. Many eukaryotic genes

transcribed by RNA polymerase II are transcribed well beyond the end of the coding sequence; most of the extra material at the 3′ end is then

cleaved and the poly(A) tail is added. For some pre-mRNA molecules, more than 1000 nucleotides may be removed from the 3′ end before polyadenylation. Processing of the 3′ end of pre-mRNA requires sequences both upstream and downstream of the cleavage site. The consensus sequence AAUAAA is usually from 11 to 30 nucleotides upstream of the cleavage site and determines the point at which cleavage will take place. A sequence rich in uracil nucleotides (or in guanine and uracil nucleotides) is typically downstream of the cleavage site. A large number of proteins take part in finding the cleavage site and removing the 3′ end. After cleavage has been completed, adenine nucleotides are added to the new 3′ end, creating the poly(A) tail.

The poly(A) tail confers stability on many mRNAs, increasing the time during which the mRNA remains intact

and available for translation before it is degraded by cellular enzymes. The stability conferred by the poly(A) tail depends on the proteins that attach to the tail and on its length. The poly(A) tail also facilitates attachment of the ribosome to the mRNA. Poly(U) tails are added to the 3′ ends of some mRNAs, microRNAs, and small nuclear RNAs. Although the function of poly(U) tails is still under investigation, evidence suggests that poly(U) tails on some mRNAs may facilitate their degradation.




**14.8 Splicing of pre-mRNA requires consensus sequences.** Critical consensus sequences are present at the 5′ splice site and the 3′ splice site. A weak consensus sequence (not shown) exists at the branch point.

**Figure 39. Splicing**

**RNA Splicing**

The other major type of modification of eukaryotic premRNA is the removal of introns by **RNA splicing**. This

modification takes place in the nucleus, before the RNA moves to the cytoplasm.

**Consensus sequences and the spliceosome**

Splicing requires the presence of three sequences in the intron. One end of the intron is referred to as the **5′ splice site**, and the other end is the **3′ splice site** (**Figure 14.8**); these splice sites possess short consensus sequences. Most introns in premRNAs begin with GU and end with AG, indicating that these sequences play a crucial role in splicing. Indeed, changing a single nucleotide at either of these sites prevents splicing.

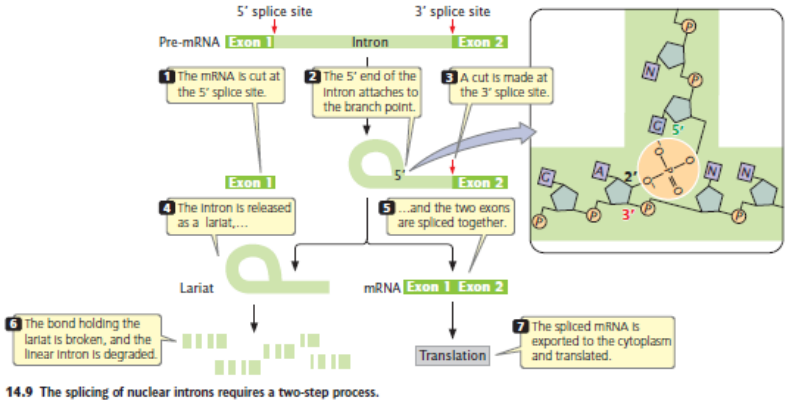
The third sequence important for splicing is at the **branch point**, which is an adenine nucleotide that lies

from 18 to 40 nucleotides upstream of the 3′ splice site The sequence surrounding the branch point does not have a strong consensus. The deletion or mutation of the adenine nucleotide at the branch point prevents splicing. Splicing takes place within a large structure called the **spliceosome**, which is one of the largest and most complex of all molecular complexes. The spliceosome consists of five RNA molecules and almost 300 proteins.

The RNA components are small nuclear RNAs ranging in length from 107 to 210 nucleotides; these snRNAs associate with proteins to form small ribonucleoprotein particles (snRNPs).

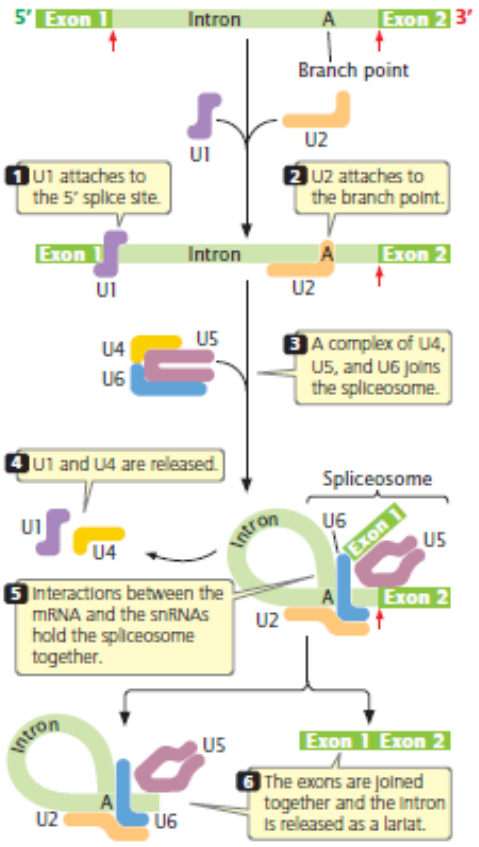
Each snRNP contains a single snRNA molecule and multiple proteins.

The spliceosome is composed of five snRNPs (U1, U2, U4, U5, and U6), and some proteins not associated with an snRNA.




14.9 The splicing of nuclear introns requires a two-step process.

**Figure 40a. mRNA splicing**




**14.10 RNA splicing takes place within the spliceosome.** The spliceosome assembles sequentially.

**Figure 40b. mRNA splicing**

**The process of splicing** Before splicing takes place, an intron lies between an upstream exon (exon 1) and a

downstream exon (exon 2), PremRNA is spliced in two distinct steps. In the first step of splicing, the pre-mRNA is cut at the 5′ splice site. This cut frees exon 1 from the intron, and the 5′ end of the intron attaches to the branch point; that is, the intron folds back on itself, forming a structure called a **lariat**. In this reaction, the guanine nucleotide in the consensus sequence at the 5′splice site bonds with the adenine nucleotide at the branch point through a transesterification reaction. The result is that the 5′ phosphate group of the guanine nucleotide is now attached to the 2′-OH group of the adenine nucleotide at the branch point.

In the second step of RNA splicing, a cut is made at the 3′ splice site and, simultaneously, the 3′ end of exon 1 becomes covalently attached (spliced) to the 5′ end of exon 2. The intron is released as a lariat. The intron becomes linear when the bond breaks at the branch point and is then rapidly degraded by nuclear enzymes. The mature mRNA consisting of the exons spliced together is exported to the cytoplasm, where it is translated. These splicing reactions take place within the

spliceosome, which assembles on the pre-mRNA in a step-bystep fashion and carries out the splicing reactions. A key feature of the process is a series of interactions between the mRNA and the snRNAs and between different snRNAs. These interactions depend on complementary base pairing between the different RNA molecules and bring the essential components of the pre-mRNA transcript and the spliceosome close together, which make splicing possible.

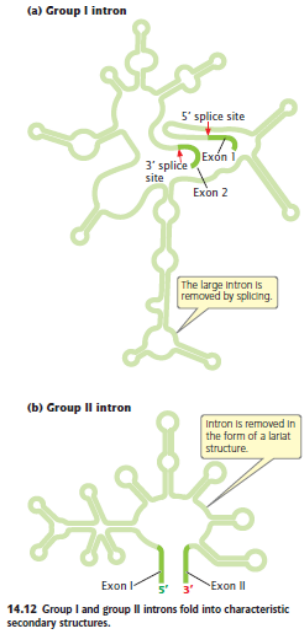
**Self-splicing introns**

Some introns are self-splicing, meaning that they possess the ability to remove themselves from an RNA molecule. These self-splicing introns fall into two major categories.

Group I introns are found in a variety of genes, including some rRNA genes in protists, some mitochondrial genes in fungi, and even some bacteriophage genes. Although the lengths of group I introns vary, all of them fold into a common secondary structure with nine looped stems, which are necessary for splicing.

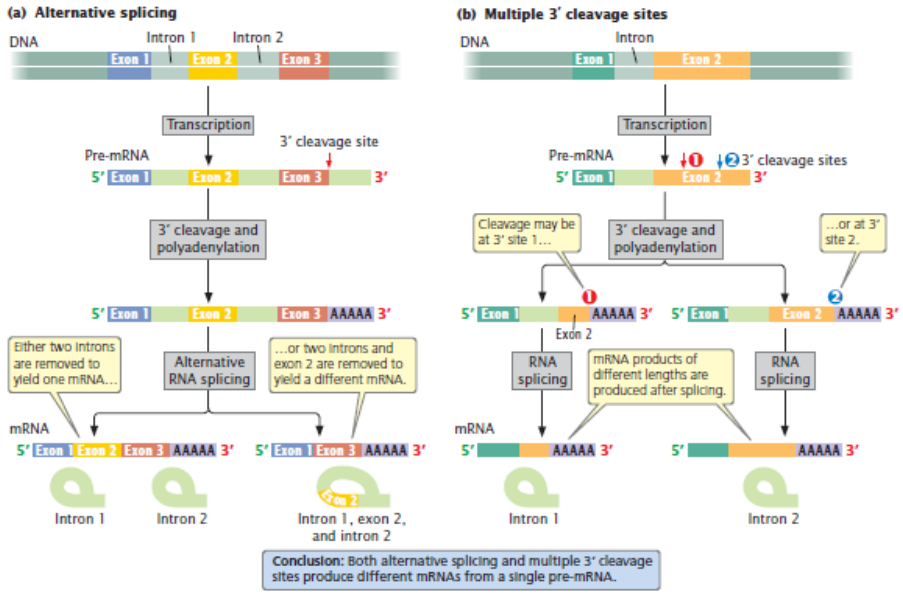
Group II introns, present in some mitochondrial genes, also have the ability to self-splice. All group II introns also fold into secondary structures. The splicing of group II introns is accomplished by a mechanism that has some similarities to the spliceosomal-mediated splicing of nuclear genes, and splicing generates a lariat structure.

Because of these similarities, group II introns and nuclear pre-mRNA introns have been suggested to be evolutionarily related; perhaps the nuclear introns evolved from self-splicing group II introns and later adopted the proteins and snRNAs of the spliceosome to carry out the splicing reaction.




14.12 Group I and group II introns fold into characteristic secondary structures.

**Figure 41. rRNA splicing**




**14.13 Eukaryotic cells have alternative pathways for processing pre-mRNA.** (a) With alternative splicing, pre-mRNA can be spliced in different ways to produce different mRNAs. (b) With multiple 3′ cleavage sites, there are two or more potential sites for cleavage and polyadenylation; use of the different sites produces mRNAs of different lengths.

**Figure 42. Alternative splicing**

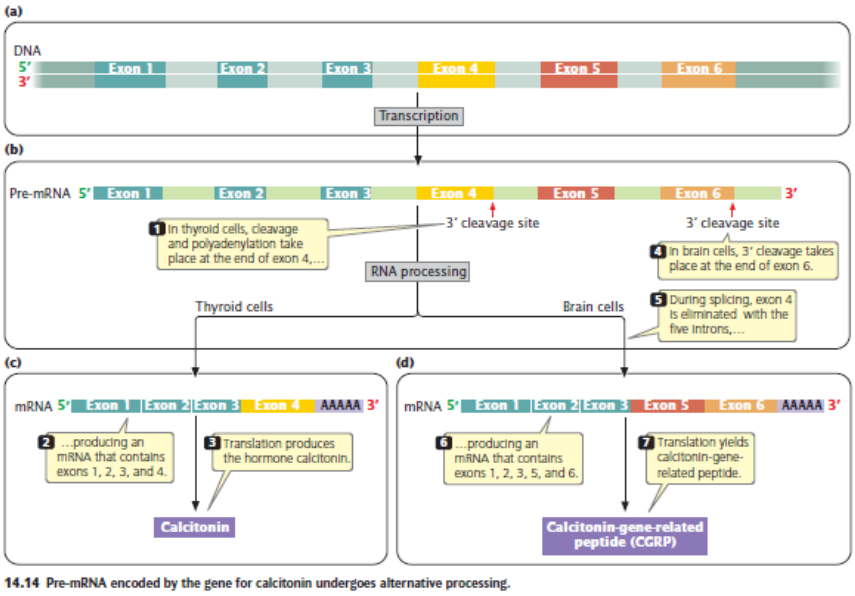
Alternative Processing Pathways

A finding that complicates the view of a gene as a sequence of nucleotides that specifies the amino acid sequence of a protein (see section on The Concept of the Gene Revisited) is the existence of **alternative processing pathways**. In these pathways, a single pre-mRNA is processed in different ways to produce alternative types of mRNA, resulting in the production of different proteins from the same DNA sequence.

One type of alternative processing is **alternative splicing,** in which the same pre-mRNA can be spliced in more than one way to yield multiple mRNAs that are translated into different amino acid sequences and thus different proteins. Another type of alternative processing requires the use of **multiple 3′ cleavage sites**, where two or more potential sites for cleavage and polyadenylation are present in the pre-mRNA. In the example, cleavage at the first site produces a relatively short mRNA compared with the mRNA produced through cleavage at the second site. The use of an alternative cleavage site may or may not produce a different protein, depending on whether the position of the site is before or after the termination codon.

Both alternative splicing and multiple 3′ cleavage sites can exist in the same pre-mRNA transcript. An example is seen in the mammalian gene that encodes calcitonin; this gene contains six exons and five introns. The entire gene is transcribed into pre-mRNA.

There are two possible 3′ cleavage sites. In cells of the thyroid gland, 3′ cleavage and polyadenylation take place after the fourth exon to produce a mature mRNA consisting of exons 1, 2, 3, and 4. This mRNA is translated into the hormone calcitonin.




14.14 Pre-mRNA encoded by the gene for calcitonin undergoes alternative processing.

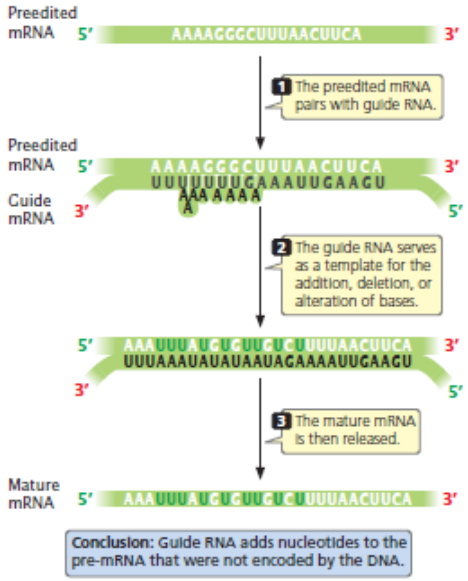
**Figure 43. Alternate splicing**

RNA Editing

In humans, for example, a gene is transcribed into mRNA that encodes a lipid-transporting polypeptide called apolipoprotein-B100, which has 4563 amino acids and is synthesized in liver cells.

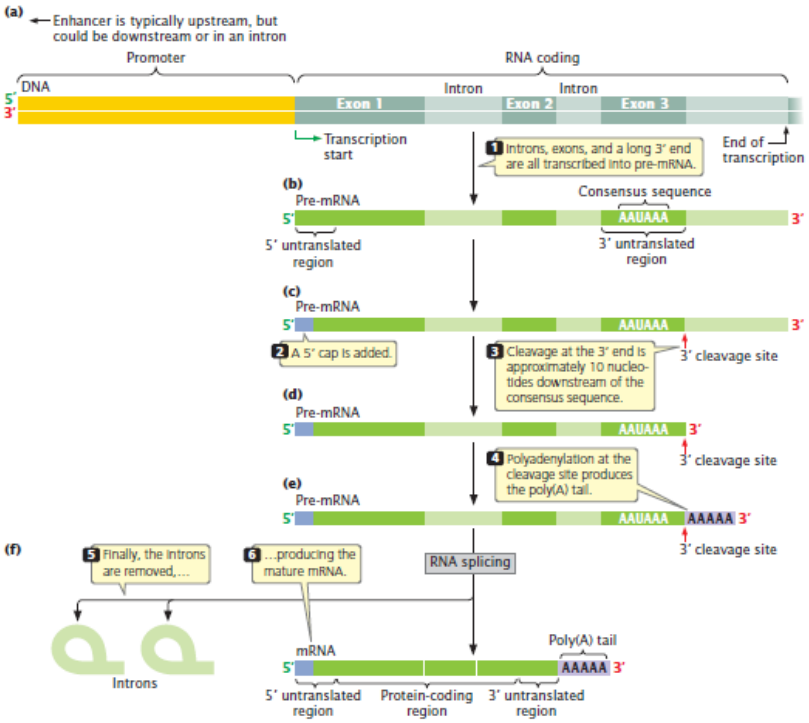
A truncated form of the protein called apolipoprotein-B48—with only 2153 amino acids—is synthesized in intestinal cells through editing of the apolipoprotein-B100 RNA.

In this editing, an enzyme deaminates a  cytosine base, converting it into uracil. This conversion changes a codon that specifies the amino acid glutamine into a stop codon that prematurely terminates translation, resulting in the shortened protein.



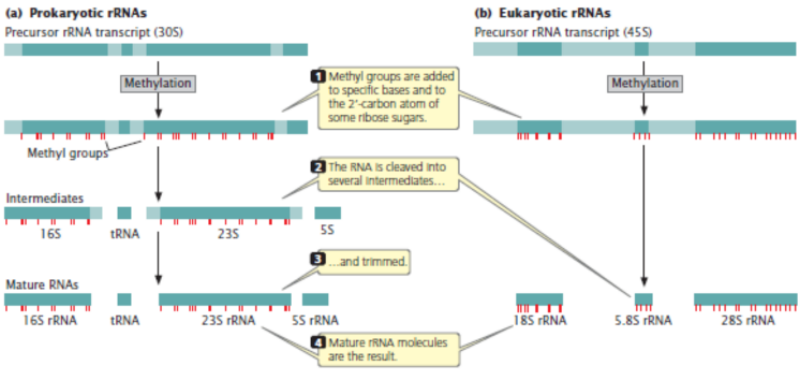

**14.16 RNA editing is carried out by guide RNAs.** The guide mRNA has sequences that are partly complementary to those of the preedited mRNA and pairs with it. After pairing, the mRNA undergoes cleavage and new nucleotides are added, with sequences in the gRNA serving as a template. The ends of the mRNA are then joined together.

## Figure 44. RNA editing




**14.17 Mature eukaryotic mRNA is produced when pre-mRNA is transcribed and undergoes several types of processing.**

## Figure 45. mRNA processing




**14.22 Ribosomal RNA is processed after transcription.** Prokaryotic rRNA (a) and eukaryotic rRNA (b) are produced from precursor RNA transcripts that are methylated, cleaved, and processed to produce mature rRNAs. Eukaryotic 5S rRNA is transcribed separately from a different gene.

**Figure 46. rRNA processing**

**RNA interference** (RNAi) is a powerful and precise mechanism used by eukaryotic cells to limit the invasion of foreign genes (from viruses and transposons) and to censor the expression of their own genes.

RNA interference is triggered by double-stranded RNA molecules, which may arise in several ways:

by the transcription of inverted repeats into an RNA molecule that then base pairs with itself to form double-stranded RNA;

by the simultaneous transcription of two different RNA molecules that are complementary to one another and that pair, forming double-stranded RNA;

or by infection by viruses that make double-stranded RNA. T

these double-stranded RNA molecules are chopped up by an enzyme appropriately called Dicer, resulting in tiny RNA molecules that are unwound to produce siRNAs and miRNAs (see Figure 14.23).

Some geneticists speculate that RNA interference evolved as a defense mechanism against RNA viruses and transposable elements that move through RNA intermediates (see Chapter 11); indeed, some have called RNAi the immune system of the genome.

However, RNA interference is also responsible for regulating a number of key genetic and developmental processes, including changes in chromatin structure, translation, cell fate and proliferation, and cell death.

Geneticists also use the RNAi machinery as an effective tool for blocking the expression of specific genes

**Types of Small RNAs**

Two abundant classes of RNA molecules that function in RNA interference in eukaryotes are small interfering RNAs and microRNAs.

They have a number of features in common and their functions overlap considerably. Both are about 22 nucleotides long.

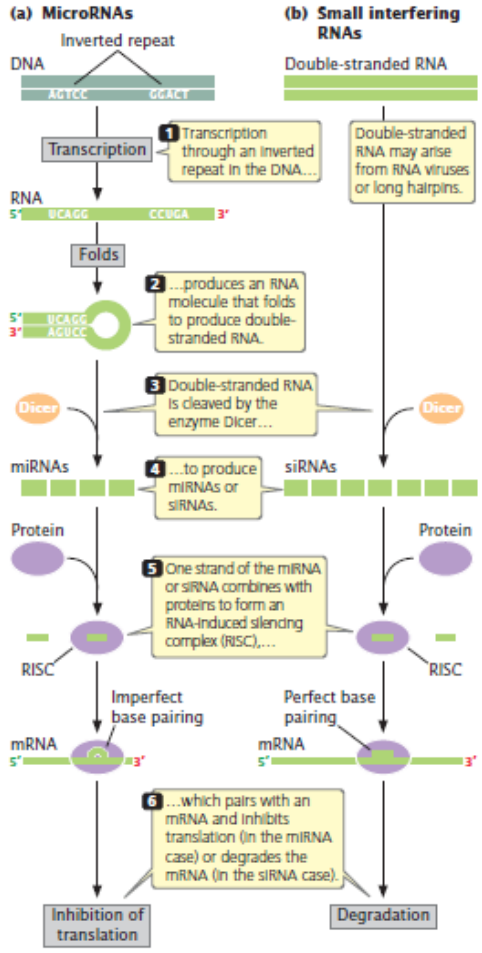
Small interfering RNAs arise from the cleavage of mRNAs,RNA transposons, and RNA viruses. Some miRNAs are cleaved from RNA molecules transcribed from sequences that encode miRNA only, but others are encoded in the introns and exons of mRNAs.

Each miRNA is cleaved from a single-stranded RNA precursor that forms small hairpins, whereas multiple siRNAs are produced from the cleavage of an RNA duplex consisting of two different RNA molecules.

Usually, siRNAs have exact complementarity with their target mRNA or DNA sequences and suppress gene expression by degrading mRNA or inhibiting transcription, whereas miRNAs often have limited complementarity with their target mRNAs and often suppress gene expression by inhibiting translation.

Finally, miRNAs usually silence genes that are distinct from those from which the miRNAs were transcribed, whereas siRNAs typically silence the genes from which the siRNAs were transcribed.

Note, however, that these differences between siRNAs and miRNAs are not hard and fast, and scientists are increasingly finding small RNAs that exhibit characteristics of both. For example, some miRNAs have exact complementarity with mRNA sequences and cleave these sequences, characteristics that are usually associated with siRNAs.




**14.23 Small interfering RNAs and microRNAs are produced from double-stranded RNAs.**

**Figure 47. siRNA and miRNA**

**Table 14.5** Differences between siRNAs, miRNAs, and piRNAs

| Feature | siRNA | miRNA | piRNA |
|---|---|---|---|
| Origin | mRNA, transposon, or virus | RNA transcribed from distinct gene | Transposons |
| Cleavage of | RNA duplex or single-stranded RNA that forms long hairpins | Single-stranded RNA that forms short hairpins of double-stranded RNA | Single-stranded RNA from transposons |
| Size | 21–25 nucleotides | 21–25 nucleotides | 24–31 nucleotides |
| Action | Degradation of mRNA, inhibition of transcription, chromatin modification | Degradation of mRNA, inhibition of translation, chromatin modification | Degradation of RNA, chromatin modification |
| Target | Genes from which they were transcribed | Genes other than those from which they were transcribed | Transposons |

## Processing and Function of MicroRNAs

MicroRNAs have been found in all eukaryotic organisms examined to date, as well as viruses; they control the expression of genes taking part in many biologicalprocesses, including growth, development, and metabolism.

Humans have more than 450 distinct miRNAs; scientists estimate that more than one-third of all human genes are regulated by miRNAs.

Most miRNA genes are found in regions of noncoding DNA or within the introns of other genes.

The genes that encode miRNAs are transcribed into longer precursors, called primary miRNA (pri-miRNA), that range from several hundred to several thousand nucleotides in length.

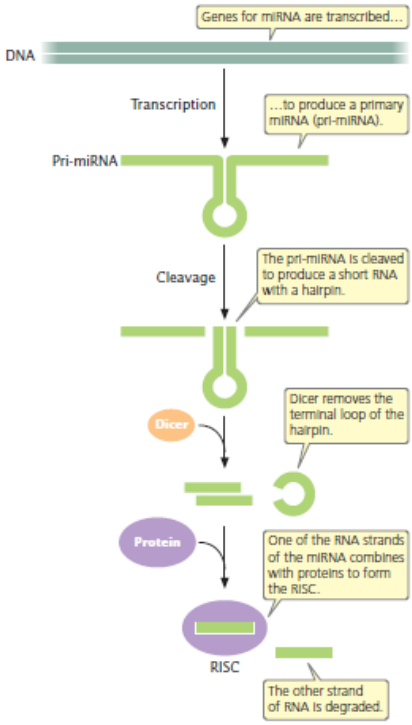
The pri-miRNA is then cleaved into one or more smaller RNA molecules with a hairpin.

Dicer (endoribonuclease Dicer or helicase with RNase motif) binds to this hairpin structure and removes the terminal loop. One of the miRNA strands is incorporated into the RISC (RNA induced Silencing Complex); the other strand is released and degraded.

The RISC attaches to a complementary sequence on the mRNA, usually in the 3′ untranslated region of the mRNA.

The region of close complementarity, called the seed region, is quite short, usually only about seven nucleotides long.

Because the seed sequence is so short, each miRNA can potentially pair with sequences on hundreds of different mRNAs. Furthermore, a single mRNA molecule may possess multiple miRNA-binding sites. The inhibition of translation may require binding



14.24 MicroRNAs are cleaved from larger precursors (pri-miRNAs).

**Figure 48. miRNA processing**

**INHIBITORS OF TRANSCRIPTION**

**Rifampicin**- binds with Beta subunit of prokaryotic RNA polymerase,

• It is an inhibitor of prokaryotic transcription initiation.

• It binds only to bacterial RNA polymerase but not to eukaryotic RNA polymerases.

• Therefore, Rifampicin is a powerful drug for treatment of bacterial infections.

• Used for the treatment of tuberculosis and leprosy

**Actinomycin D**

• Actinomycin D- Intercalates with DNA strands

• Actinomycins inhibit both DNA synthesis and RNA synthesis by blocking chain elongation.

• They interact with G·C base pairs as they require the 2-amino group of guanine for binding.

• Actinomycins are used as anticancer Drugs

**Mitomycin**

• Mitomycin- Intercalates with DNA strands

• blocks transcription,

• used as anticancer drug

**Alpha amanitin**

• Alpha amanitin is a molecule made from the “death cap” mushroom and is a known potent inhibitor RNA polymerase.

• One single mushroom could very easily lead to a fast death in 10 days.

• The mechanism of action is that alpha amanitin inhibits RNA polymerase –II at both the initiation and elongation states of transcription.

## GENETIC CODE

The first person to suggest the existence of a relation between genotype and proteins was English physician Archibald Garrod. In 1908, Garrod correctly proposed that genes encode enzymes, but, unfortunately, his theory made little impression on his contemporaries. Not until the 1940s, when George Beadle and Edward Tatum examined the genetic basis of biochemical pathways in *Neurospora*, did the relation between genes and proteins become widely accepted.

### The One Gene, One Enzyme Hypothesis

Beadle and Tatum used the bread mold *Neurospora* to study the biochemical results of mutations. *Neurospora* is easy to cultivate in the laboratory. The main vegetative part of the fungus is haploid, which allows the effects of recessive mutations to be easily observed.

Wild-type *Neurospora* grows on minimal medium, which contains only inorganic salts, nitrogen, a carbon source such as sucrose, and the vitamin biotin. The fungus can synthesize all the biological molecules that it needs from these basic compounds. However, mutations may arise that disrupt fungal growth by destroying the fungus's ability to synthesize one or more essential biological molecules. These nutritionally deficient mutants, termed auxotrophs, will not grow on minimal medium, but they can grow on medium that contains the substance that they are no longer able to synthesize.

Beadle and Tatum first irradiated spores of *Neurospora* to induce mutations. After irradiation, they placed individual spores into different culture tubes containing complete medium (medium having all the biological substances needed for growth). Next, they transferred spores from each culture to tubes containing minimal medium. Fungi containing auxotrophic mutations grew on complete medium but would not grow on minimal medium, which allowed Beadle and Tatum to identify cultures that contained mutations.

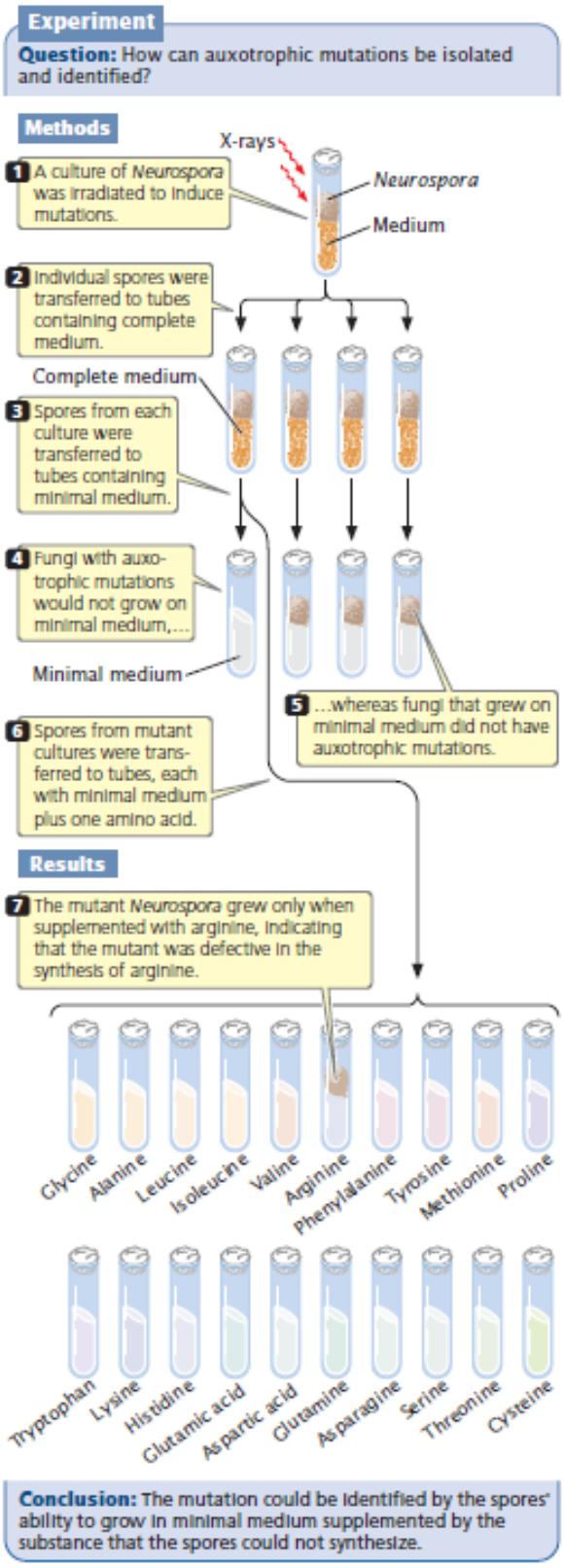
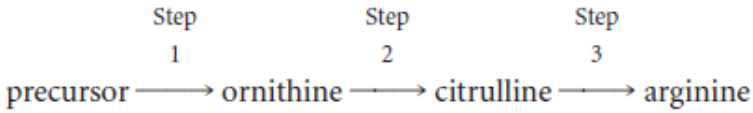



**Figure 49. Beadle and Tatum experiment**

After they had determined that a particular culture had an auxotrophic mutation, Beadle and Tatum set out to determine the specific *effect* of the mutation. They transferred spores of each mutant

strain from complete medium to a series of tubes, each of which possessed minimal medium plus one of a variety of essential biological molecules, such as an amino acid. If the spores in a tube grew, Beadle and Tatum were able to identify the added substance as the biological molecule whose synthesis had been affected by the mutation. For example, an auxotrophic mutant that would grow only on minimal medium to which arginine had been added must have possessed a mutation that disrupts the synthesis of arginine.

Adrian Srb and Norman H. Horowitz patiently applied this procedure to genetically dissect the multistep biochemical pathway of arginine synthesis. They first isolated a series of auxotrophic mutants whose growth required arginine. They then tested these mutants for their ability to grow on minimal medium supplemented with three compounds: ornithine, citrulline, and arginine. From the results, they were able to place the mutants into three groups on the basis of which of the substances allowed growth. Group I mutants grew on minimal medium supplemented with ornithine, citrulline, or arginine. Group II mutants grew on minimal medium supplemented with either arginine or citrulline but did not grow on medium supplemented only with ornithine. Finally, group III mutants grew only on medium supplemented with arginine. Biochemical pathway leading to the amino acid arginine has at least three steps:




They concluded that the mutations in group I affect step 1 of this pathway, mutations in group II affect step 2, and

mutations in group III affect step 3. But how did they know that the order of the compounds in the biochemical pathway was correct?

Notice that, if step 1 is blocked by a mutation, then the addition of either ornithine or citrulline allows growth, because these compounds can still be converted into arginine (see Figure 15.3). Similarly, if step 2 is blocked, the addition of citrulline allows growth, but the addition of ornithine has no effect. If step 3 is blocked, the spores will grow only if arginine is added to the medium. The underlying principle is that an auxotrophic mutant cannot synthesize any compound that comes after the step blocked by a mutation.

Notice that, if step 1 is blocked by a mutation, then the addition of either ornithine or citrulline allows growth, because these compounds can still be converted into arginine. Similarly, if step 2 is blocked, the addition of citrulline allows growth, but the addition of ornithine has no effect. If step 3 is blocked, the spores will grow only if arginine is added to the medium. The underlying principle is that an auxotrophic mutant cannot synthesize any compound that comes after the step blocked by a mutation.

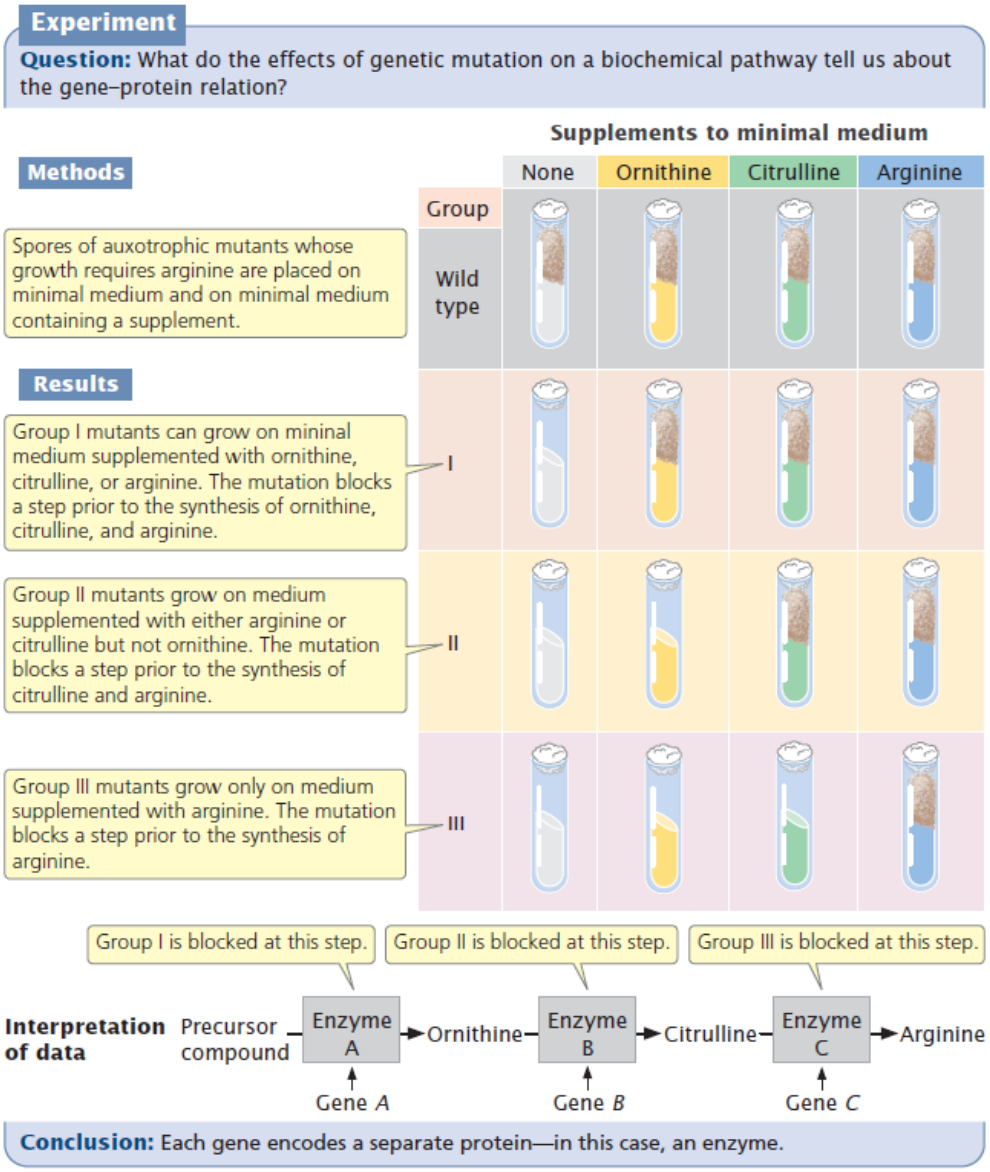


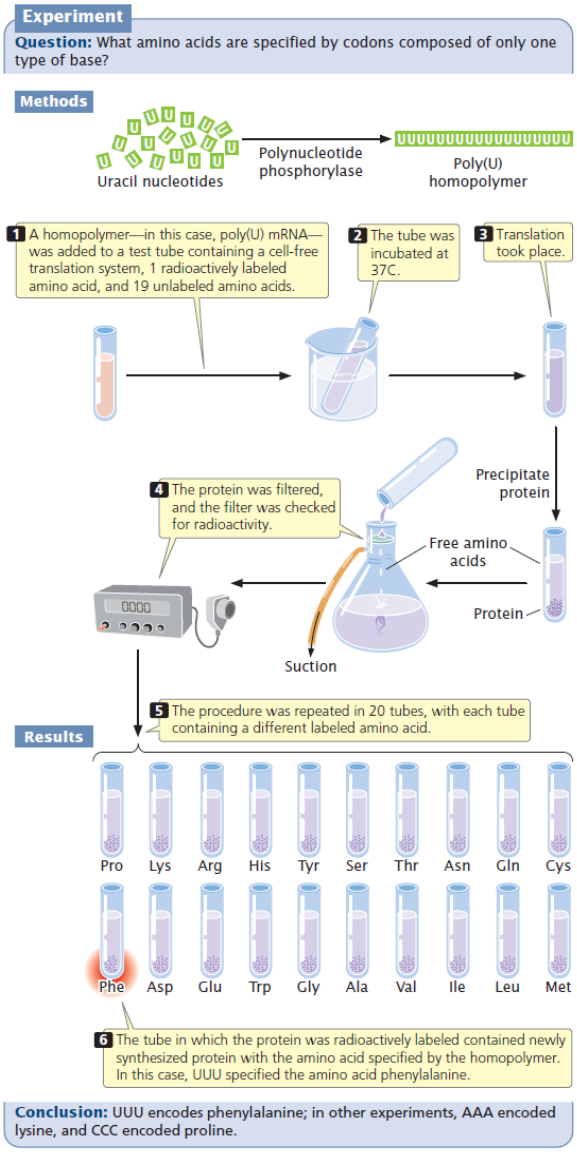

**Figure 50. Adrian Srb and Norman H. Horowitz experiment**

In 1953, James Watson and Francis Crick solved the structure of DNA and identified the base sequence as the carrier of genetic information. However, the way in which the base sequence of

DNA specifies the amino acid sequences of proteins (the genetic code) was not immediately obvious and remained elusive for another 10 years.

One of the first questions about the genetic code to be addressed was: How many nucleotides are necessary to specify a single amino acid? This basic unit of the genetic code—the set of bases that encode a single amino acid—is a *codon* (see p. 380 in Chapter 14). Many early investigators recognized that codons must contain a minimum of three nucleotides. Each nucleotide position in mRNA can be occupied by one of four bases: A, G, C, or U. If a codon consisted of a single nucleotide, only four different codons (A, G, C, and U) would be possible, which is not enough to encode the 20 different amino acids commonly found in proteins. If codons were made up of two nucleotides each (i.e., GU, AC, etc.), there would be $4 \times 4 = 16$ possible codons—still not enough to encode all 20 amino acids. With three nucleotides per codon, there are $4 \times 4 \times 4 = 64$ possible codons, which is more than enough to specify 20 different amino acids.

Therefore, a *triplet code* requiring three nucleotides per codon is the most efficient way to encode all 20 amino acids. Using mutations in bacteriophage, Francis Crick and his colleagues confirmed in 1961 that the genetic code is indeed a triplet code.




15.8 Nirenberg and Matthaei developed a method for identifying the amino acid specified by a homopolymer.

**Figure 51. Nirenberg and Matthaei experiment**

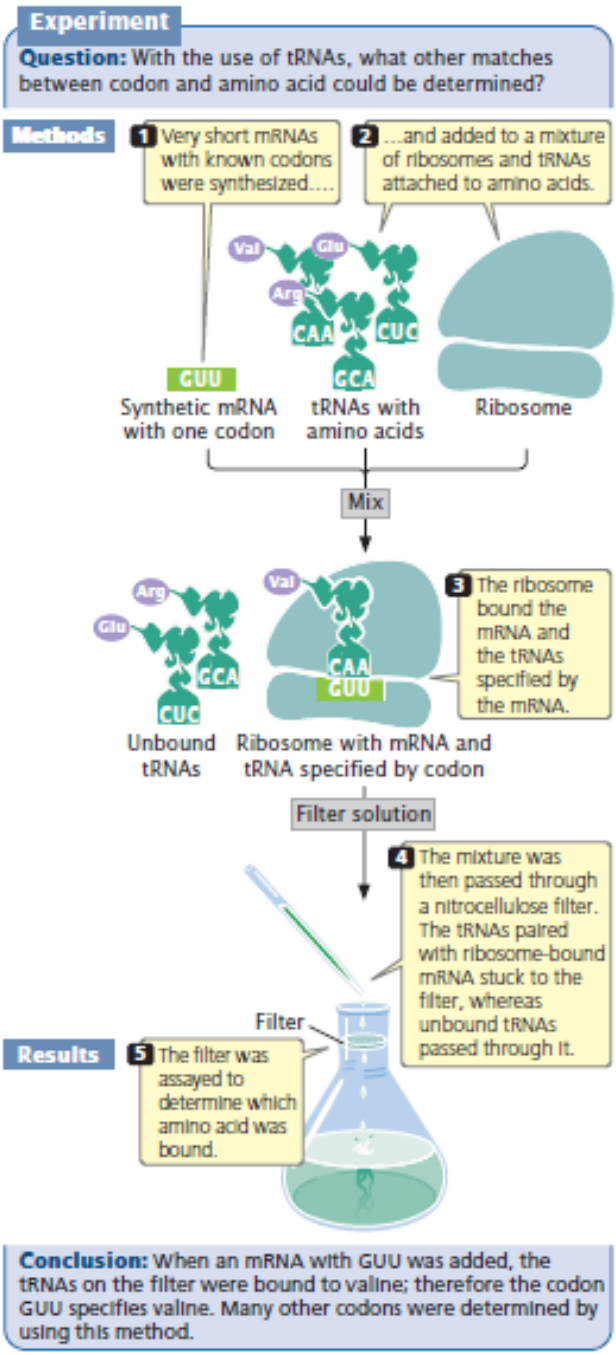
**The use of homopolymers** The first clues to the genetic code came in 1961, from the work of Marshall Nirenberg and Johann Heinrich Matthaei. These investigators created synthetic RNAs by using an enzyme called polynucleotide phosphorylase. Unlike RNA polymerase, polynucleotide phosphorylase does not require a template; it randomly links together any RNA nucleotides that happen to be available. The first synthetic mRNAs used by Nirenberg and Matthaei were homopolymers, RNA molecules consisting of a single type of nucleotide. For example, by adding polynucleotide phosphorylase to a solution of uracil nucleotides, they generated RNA molecules that consisted entirely of uracil nucleotides and thus contained only UUU codons. These poly(U)

RNAs were then added to 20 tubes, each containing the components necessary for translation and all 20 amino acids. A different amino acid was radioactively labeled in each of the 20 tubes. Radioactive protein appeared in only one of the tubes—the one containing labeled phenylalanine. This result showed that the codon UUU specifies the amino acid phenylalanine. The results of similar experiments using poly(C) and poly(A) RNA demonstrated that CCC

**The use of homopolymers** The first clues to the genetic code came in 1961, from the work of Marshall Nirenberg and Johann Heinrich Matthaei. These investigators created synthetic RNAs by using an enzyme called polynucleotide phosphorylase. Unlike RNA polymerase, polynucleotide phosphorylase does not require a template; it randomly links together any RNA nucleotides that happen to be available. The first synthetic mRNAs used by Nirenberg and Matthaei were homopolymers, RNA molecules consisting of a single type of nucleotide. For example, by adding polynucleotide phosphorylase to a solution of uracil nucleotides, they generated RNA molecules that consisted entirely of uracil nucleotides and thus contained only UUU codons. These poly(U) RNAs were then added to 20 tubes, each containing the components necessary for translation and all 20 amino acids. A different amino acid was radioactively labeled in each of the 20 tubes. Radioactive protein appeared in only one of the tubes—the one containing labeled phenylalanine (see Figure 15.8). This result showed that the codon UUU specifies the amino acid phenylalanine. The results of similar experiments using poly(C) and poly(A) RNA demonstrated that CCC

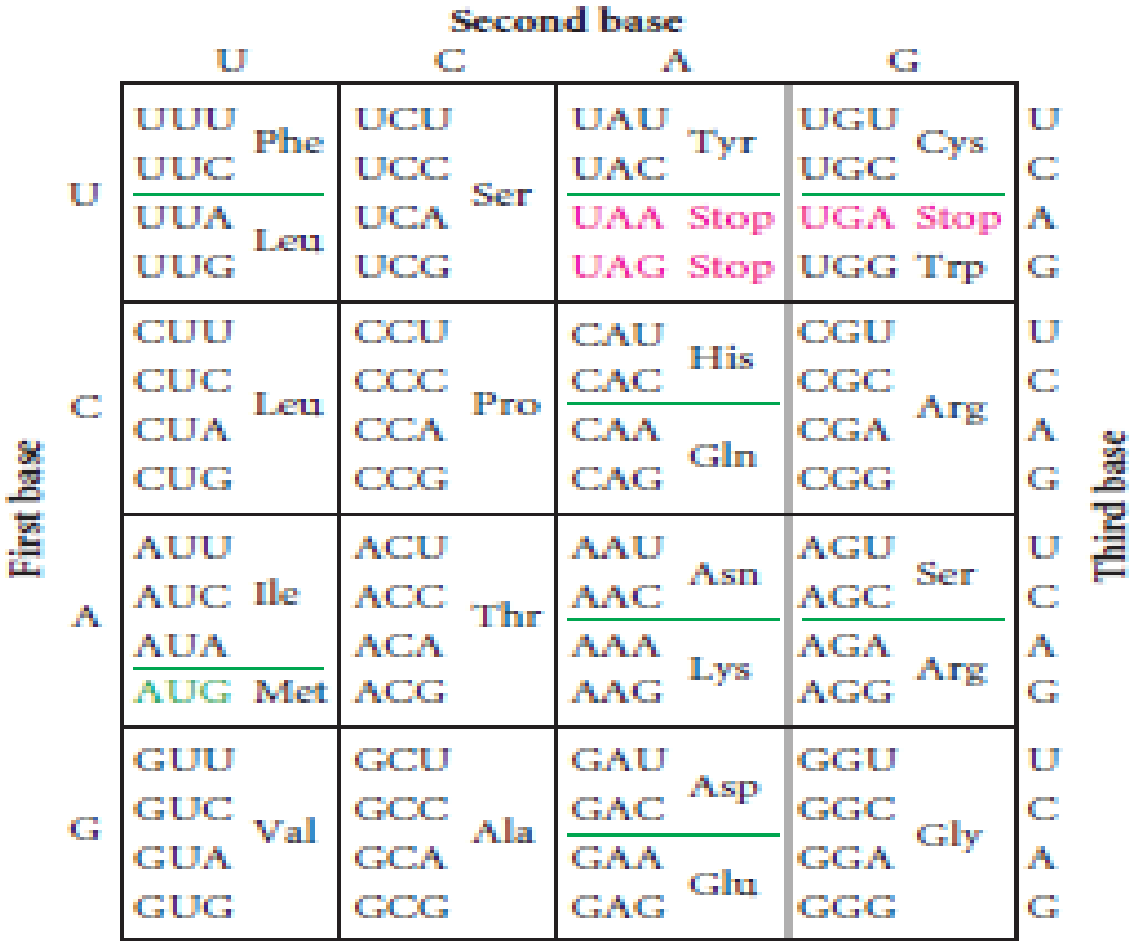
**The use of random copolymers** To gain information about additional codons, Nirenberg and his colleagues created synthetic RNAs containing two or three different bases.

Because polynucleotide phosphorylase incorporates nucleotides randomly, these RNAs contained random mixtures of the bases and are thus called random copolymers. For example, when adenine and cytosine nucleotides are mixed with polynucleotide phosphorylase, the RNA molecules produced have eight different codons: AAA, AAC, ACC, ACA, CAA, CCA, CAC, and CCC. These poly(AC) RNAs produced proteins containing six different amino acids: asparagine, glutamine, histidine, lysine, proline, and threonine.




**15.9 Nirenberg and Leder used ribosome-bound tRNAs to provide additional information about the genetic code.**

**Figure 52. Nirenberg and Leder experiment**



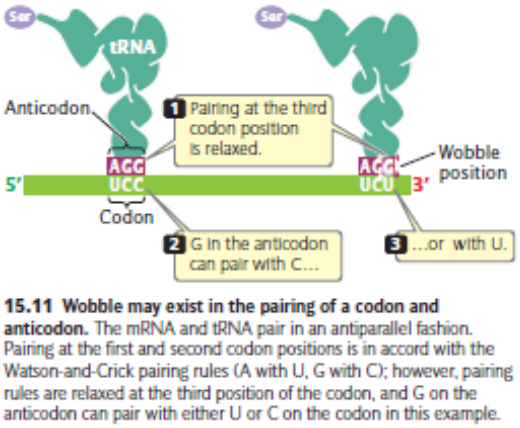

**15.10 The genetic code consists of 64 codons.** The amino acids specified by each codon are given in their three-letter abbreviation. The codons are written 5′⟶3′, as they appear in the mRNA. AUG is an initiation codon; UAA, UAG, and UGA are termination (stop) codons.

**The Degeneracy of the Code**

One amino acid is encoded by three consecutive nucleotides in mRNA, and each nucleotide can have one of four possible bases (A, G, C, and U) at each nucleotide position, thus permitting 43 = 64 possible codons. Three of these codons are stop codons, specifying the end of translation. Thus, 61 codons, called **sense codons**, encode amino acids. Because there are 61 sense codons and only 20 different amino acids commonly found in proteins, the code contains more information than is needed to specify the amino acids and is said to be a **degenerate code**.

The degeneracy of the genetic code means that amino acids may be specified by more than one codon. Only tryptophan and methionine are encoded by a single codon. Other amino acids are specified by two codons, and some, such as leucine, are specified by six different codons. Codons

that specify the same amino acid are said to be **synonymous**, just as synonymous words are different words that have the same meaning.




**15.11 Wobble may exist in the pairing of a codon and anticodon.** The mRNA and tRNA pair in an antiparallel fashion. Pairing at the first and second codon positions is in accord with the Watson-and-Crick pairing rules (A with U, G with C); however, pairing rules are relaxed at the third position of the codon, and G on the anticodon can pair with either U or C on the codon in this example.

**Figure 53. Wobble hypothesis**

The cells of most organisms possess from about 30 to 50 different tRNAs, and yet there are only 20 different amino acids in proteins. Thus, some amino acids are carried by more than one tRNA. Different tRNAs that accept the same amino acid but have different anticodons are called **isoaccepting tRNAs**. Some synonymous codons specify different isoacceptors.

Many synonymous codons differ only in the third position. For example, alanine is encoded by the codons GCU, GCC, GCA, and GCG, all of which begin with GC. When the codon on the mRNA and the anticodon of the tRNA join, the first (5′) base of the codon pairs with the third (3′) base of the anticodon, strictly according to Watson-and-Crick rules: A with U; C with G. Next, the middle bases of codon and anticodon pair, also strictly following the Watson-and-Crick rules. After these pairs have hydrogen bonded, the third bases pair weakly: there may be flexibility, or **wobble**, in their pairing.

the code is generally **nonoverlapping**. An overlapping code is one in which a single nucleotide may be included in more than one codon, as follows:

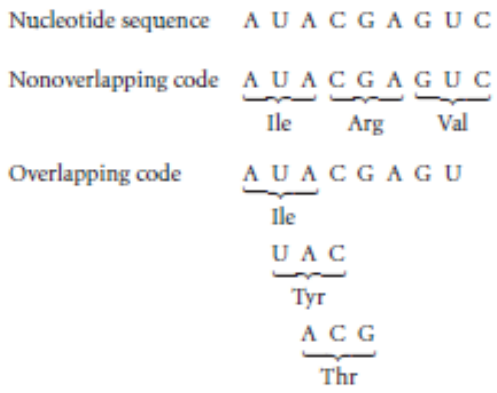



**Figure 54. Codes are non overlapping**

**Termination Codons**

Three codons—UAA, UAG, and UGA—do not encode amino acids. These codons signal the end of the protein in both bacterial and eukaryotic cells and are called **stop codons, termination codons,** or **nonsense codons**. No tRNA molecules have anticodons that pair with termination codons.

**The Universality of the Code**

For many years the genetic code was assumed to be **universal**, meaning that each codon specifies the same amino acid in all organisms. We now know that the genetic code is almost, but not completely, universal; a few exceptions have been found. Most of these exceptions are termination codons, but there are a few cases in which one sense codon substitutes for another. Most exceptions are found in mitochondrial genes; a few nonuniversal codons have also been detected in the nuclear genes of protozoans and in bacterial DNA

**Table 15.3** Some exceptions to the universal genetic code

| Genome | Codon | Universal Code | Altered Code |
|---|---|---|---|
| Bacterial DNA | | | |
| *Mycoplasma capricolum* | UGA | Stop | Trp |
| Mitochondrial DNA | | | |
| Human | UGA | Stop | Trp |
| Human | AUA | Ile | Met |
| Human | AGA, AGG | Arg | Stop |
| Yeast | UGA | Stop | Trp |
| Trypanosomes | UGA | Stop | Trp |
| Plants | CGG | Arg | Trp |
| Nuclear DNA | | | |
| *Tetrahymena* | UAA | Stop | Gln |
| *Paramecium* | UAG | Stop | Gln |

**Characteristics of Genetic code**

1. The genetic code consists of a sequence of nucleotides in DNA or RNA. There are four letters in the code, corresponding to the four bases—A, G, C, and U (T in DNA).

2. The genetic code is a triplet code. Each amino acid is encoded by a sequence of three consecutive nucleotides, called a codon.

3. The genetic code is degenerate; that is, of 64 codons, 61 codons encode only 20 amino acids in proteins (3 codons are termination codons). Some codons are synonymous, specifying the same amino acid.

4. Isoaccepting tRNAs are tRNAs with different anticodons that accept the same amino acid; wobble allows the anticodon on one type of tRNA to pair with more than one type of codon on mRNA.

5. The code is generally nonoverlapping; each nucleotide in an mRNA sequence belongs to a single reading frame.

6. The reading frame is set by an initiation codon, which is usually AUG.

7. When a reading frame has been set, codons are read as successive groups of three nucleotides.

8. Any one of three termination codons (UAA, UAG, and UGA) can signal the end of a protein; no amino acids are encoded by the termination codons.

9. The code is almost universal.

# UNIT – IV – PROTEIN SYNTHESIS

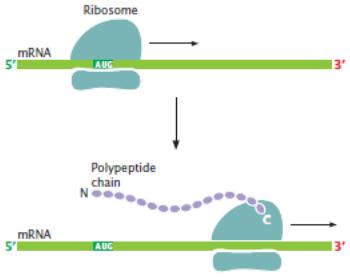
## TRANSLATION

### Amino Acids Are Assembled into a Protein Through the Mechanism of Translation

Now that we are familiar with the genetic code, we can begin to study the mechanism by which amino acids are assembled into proteins. Because more is known about translation in bacteria, we will focus primarily on bacterial translation. In most respects, eukaryotic translation is similar, although some significant differences will be noted as we proceed through the stages of translation. Translation takes place on ribosomes; indeed, ribosomes can be thought of as moving protein-synthesizing machines. Through a variety of techniques, a detailed view of the structure of the ribosome has been produced in recent years, which has greatly improved our understanding of the translational process. A ribosome attaches near the 5′ end of an mRNA strand and moves toward the 3′ end, translating the codons as it goes. Synthesis begins at the amino end of the protein, and the protein is elongated by the addition of new amino acids to the carboxyl end. Protein synthesis includes a series of RNA–RNA interactions: interactions between the mRNA and the rRNA that hold the mRNA in the ribosome, interactions between the codon on the mRNA and the anticodon on the tRNA, and interactions

between the tRNA and the rRNAs of the ribosome Protein synthesis can be conveniently divided into four stages: (1) tRNA charging, in which tRNAs bind to amino acids; (2) initiation, in which the components necessary for translation are assembled at the ribosome; (3) elongation, in which amino acids are joined, one at a time, to the growing polypeptide chain; and (4) termination, in which protein synthesis halts at the termination codon and the translation

components are released from the ribosome.




**15.12 The translation of an mRNA molecule takes place on a ribosome.** The letter N represents the amino end of the protein; C represents the carboxyl end.

**Figure 55. Translation**

### The Binding of Amino Acids to Transfer RNAs

The first stage of translation is the binding of tRNA molecules to their appropriate amino acids, called tRNA charging. Each tRNA is specific for a particular amino acid. All tRNAs have the sequence CCA at the 3′ end, and the carboxyl group (COO–) of the amino acid is attached to the

adenine nucleotide at the 3′ end of the tRNA. If each tRNA is specific for a particular amino acid but all amino acids are attached to the same nucleotide (A) at the 3′ end of a tRNA, how does a tRNA link up with its appropriate amino acid?

tRNA is a set of enzymes called **aminoacyl-tRNA synthe tases**. A cell has 20 different aminoacyl-tRNA synthetases, one for each of the 20 amino acids. Each synthetase recognizes a particular amino acid, as well as all the tRNAs that accept that amino acid. Recognition of the appropriate amino acid by a synthetase is based on the different sizes, charges, and R groups of the amino acids. The recognition of tRNAs by a synthetase depends on the differing nucleotide sequences of the tRNAs. Researchers have identified which nucleotides are important in recognition by altering different nucleotides in a particular tRNA and determining whether the altered tRNA is still recognized by its synthetase.

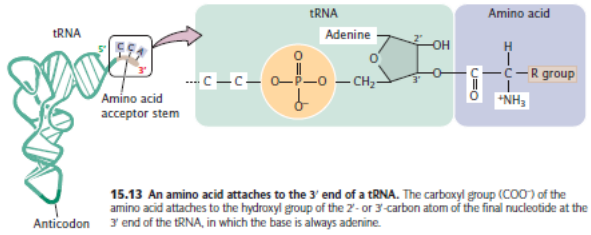



**Figure 56. tRNA charging**

The attachment of a tRNA to its appropriate amino acid, termed **tRNA charging**, requires energy, which is supplied by adenosine triphosphate (ATP):

amino acid + tRNA + ATP → aminoacyl-tRNA + AMP + PPi

This reaction takes place in two steps. To identify the resulting aminoacylated tRNA, we write the three-letter abbreviation for the amino acid in front of the tRNA; for example, the amino acid alanine (Ala) attaches to its tRNA (tRNAAla), giving rise to its aminoacyl-tRNA (Ala-tRNAAla). Errors in tRNA charging are rare; they occur in only about 1 in 10,000 to 1 in 100,000 reactions. This fidelity is due in part to the presence of editing (proofreading) activity

in many of the synthetases. Editing activity detects and removes incorrectly paired amino acids from the tRNAs. Some antifungal chemical agents work by trapping tRNAs in the editing site of the enzyme, preventing their release and thus inhibiting the process of translation in the fungi.

**The Initiation of Translation**

The second stage in the process of protein synthesis is initiation. At this stage, all the components necessary for protein synthesis assemble: (1) mRNA; (2) the small and large subunits of the ribosome; (3) a set of three proteins called initiation factors; (4) initiator tRNA with *N*-

formylmethionine attached (fMet-tRNAfMet); and (5) guanosine triphosphate (GTP). Initiation comprises three major steps. First, mRNA binds to the small subunit of the ribosome. Second, initiator tRNA binds to the mRNA through base pairing between the codon and the anticodon. Third, the large ribosome joins the initiation complex. Let's look at each of these steps more closely.

**Initiation in bacteria** The functional ribosome of bacteria exists as two subunits, the small 30S subunit and the large 50S subunit. An mRNA molecule can bind to the small ribosome subunit only when the subunits are separate. **Initiation factor 3** (IF-3) binds to the small subunit of the ribosome and prevents the large subunit from binding during initiation . Another factor, **initiation factor 1** (IF-1), enhances the disassociation of the large and small ribosomal subunits.

Key sequences on the mRNA required for ribosome binding have been identified in experiments designed to allow the ribosome to bind to mRNA but not proceed with protein synthesis; the ribosome is thereby stalled at the initiation site. After the ribosome is allowed to attach to the mRNA, ribonuclease is added, which degrades all the mRNA except the region covered by the ribosome. The intact mRNA can be separated from the ribosome and studied. The sequence covered by the ribosome during initiation is from 30 to 40 nucleotides long and includes the AUG initiation codon. Within the ribosome-binding site is the Shine–Dalgarno consensus sequence, which is complementary to a sequence of nucleotides at the 3′ end of 16S rRNA (part of the small subunit of the ribosome). During initiation, the nucleotide in the Shine–Dalgarno sequence pair with their complementary nucleotides in the 16S rRNA, allowing the small subunit of the ribosome to attach to the mRNA and positioning the ribosome directly over the initiation codon.

Next, the initiator tRNA, fMet-tRNAfMet, attaches to the initiation codon. This step requires **initiation factor 2** (IF-2), which forms a complex with GTP. At this point, the initiation complex consists of (1) the small subunit of the ribosome; (2) the mRNA; (3) the initiator tRNA with its amino acid (fMet-tRNAfMet); (4) one molecule of GTP; and (5) several initiation factors. These components are collectively known as the **30S initiation complex**. In the final step of initiation, IF-3 dissociates from the small subunit, allowing the large subunit of the ribosome to join the initiation complex. The molecule of GTP (provided by IF-2) is hydrolyzed to guanosine diphosphate (GDP), and the initiation factors dissociate. When the large subunit has joined the initiation complex, the complex is called the **70S initiation complex**.

**Initiation in eukaryotes**

Similar events take place in the initiation of translation in eukaryotic cells, but there are some important differences. In bacterial cells, sequences in 16S rRNA of the small subunit of the ribosome bind to the Shine–Dalgarno sequence in mRNA. No analogous consensus sequence exists in eukaryotic mRNA. Instead, the cap at the 5′ end of eukaryotic mRNA plays a critical role in the initiation of translation. In a series of steps, the small subunit of the eukaryotic ribosome, initiation factors, and the initiator tRNA with its amino acid (Met-tRNAi Met) form an initiation complex that recognizes the cap and binds there.

The initiation complex then moves along (scans) the mRNA until it locates the first AUG codon. The identification of the start codon is facilitated by the presence of a consensus sequence (called the Kozak sequence) that surrounds the start codon:

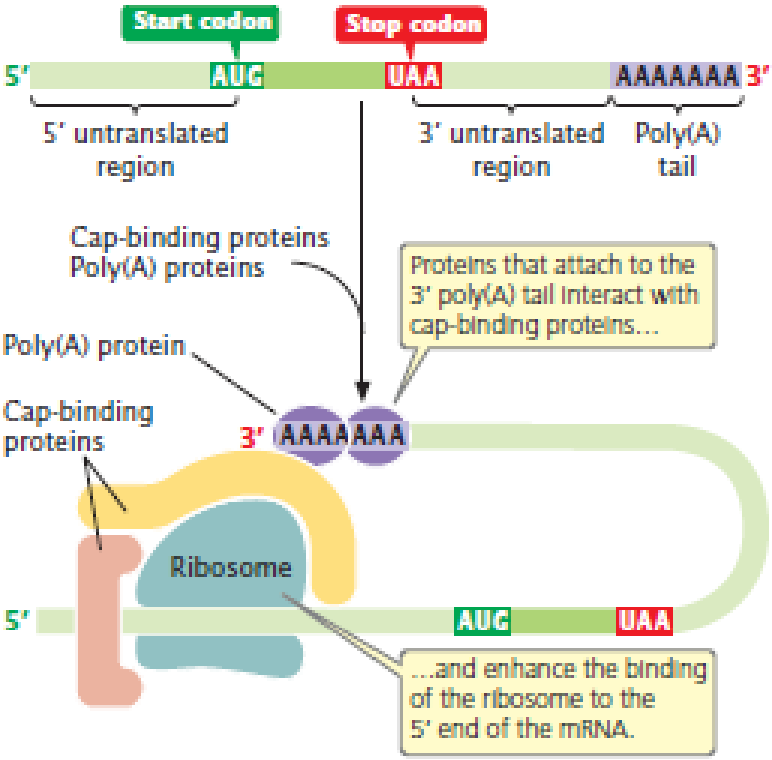
5 –ACCAUGG–3

**Kozak sequence**

**Start codon**

Another important difference is that eukaryotic initiation requires at least seven initiation factors. Some factors keep the ribosomal subunits separated, just as IF-3 does in bacterial cells. Others recognize the 5′ cap on mRNA and allow the small subunit of the ribosome to bind there. Still others possess RNA helicase activity, which is used to unwind secondary structures that may exist in the 5′ untranslated region of mRNA, allowing the small subunit to move down the mRNA until the initiation codon is reached.

Other initiation factors help bring Met-tRNAi Met to the initiation complex. The poly(A) tail at the 3′ end of eukaryotic mRNA also plays a role in the initiation of translation. During initiation, proteins that attach to the poly(A) tail interact with proteins that bind to the 5′ cap, enhancing the binding of the small subunit of the ribosome to the 5′ end of the mRNA. This interaction indicates that the 3′ end of mRNA bends over and associates with the 5′ cap during the initiation of translation, forming a circular structure known as the closed loop. A few eukaryotic mRNAs contain internal ribosome entry sites, where ribosomes can bind directly without first attaching to the 5′ cap.




**15.18 The poly(A) tail of eukaryotic mRNA plays a role in the initiation of translation.**

**Figure 57. Initiation of translation**

**Elongation**

The next stage in protein synthesis is elongation, in which amino acids are joined to create a polypeptide chain. Elongation requires (1) the 70S complex just described; (2) tRNAs charged with their amino acids; (3) several elongation factors; and (4) GTP.

A ribosome has three sites that can be occupied by tRNAs; the **aminoacyl**, or **A**, **site**, the **peptidyl**, or **P**, **site**, and the **exit**, or **E**, **site**. The initiator tRNA immediately occupies the P site (the only site to which the fMet-tRNAfMet is capable of binding), but all other tRNAs first enter the A site. After initiation, the ribosome is attached to the mRNA, and fMet-tRNAfMet is positioned over the AUG start codon in the P site; the adjacent A site is unoccupied (see Figure 15.19a).

Elongation takes place in three steps. In the first step, a charged tRNA binds to the A site. This binding takes place when **elongation factor Tu** (EF-Tu) joins with GTP and then with a charged tRNA to form a three-part complex. This complex enters the A site of the ribosome, where the anticodon on the tRNA pairs with the codon on the mRNA. After the charged tRNA is in the A site, GTP is cleaved to GDP, and the EF-Tu–GDP complex is released. **Elongation factor Ts** (EF-

Ts) regenerates EF-Tu–GDP to EF-Tu–GTP. In eukaryotic cells, a similar set of reactions delivers the charged tRNA to the A site.

The second step of elongation is the formation of a peptide bond between the amino acids that are attached to tRNAs in the P and A sites. The formation of this peptide bond releases the amino acid in the P site from its tRNA. Evidence indicates that the catalytic activity is a property of the ribosomal RNA in the large subunit of the ribosome; this rRNA acts as a ribozyme.

The third step in elongation is **translocation**, the movement of the ribosome down the mRNA in the 5′→3′ direction. This step positions the ribosome over the hydrolysis of GTP to GDP. Because the tRNAs in the P and A sites are still attached to the mRNA through codon– anticodon pairing, they do not move with the ribosome as it translocates. Consequently, the ribosome shifts so that the tRNA that previously occupied the P site now occupies the E site, from which it moves into the cytoplasm where it can be recharged with another amino acid. Translocation also causes the tRNA that occupied the A site (which is attached to the growing polypeptide chain) to be in the P site, leaving the A site open. Thus, the progress of each tRNA through the ribosome in the course of elongation can be summarized as follows:

cytoplasm → A site → P site → E site → cytoplasm.

As discussed earlier, the initiator tRNA is an exception: it attaches directly to the P site and never occupies the A site. After translocation, the A site of the ribosome is empty and ready to receive the tRNA specified by the next codon. The elongation cycle (see Figure 15.19b through e) repeats itself: a charged tRNA and its amino acid occupy the A site, a peptide bond is formed between the amino acids in the A and P sites, and the ribosome translocates to the next codon. Throughout the cycle, the polypeptide chain remains attached to the tRNA in the P site. Messenger RNAs, although single stranded, often contain secondary structures formed by pairing of complementary bases on different parts of the mRNA. As the ribosome moves along the mRNA, these secondary structures are unwound by helicase activity located in the small subunit of the ribosome.

Recently, researchers have developed methods for following a single ribosome as it translates individual codons of an mRNA molecule. These studies revealed that translation does not take place in a smooth continuous fashion. Each translocation step typically requires less than a tenth of a second, but sometimes there are distinct pauses, often lasting a few seconds, between each translocation event when the ribosome moves from one codon to another. Thus, translation takes place in a series of quick translocations interrupted by brief pauses. In addition to the short pauses between translocation events, translation may be interrupted by longer pauses—lasting from 1 to

2 minutes—that may play a role in regulating the process of translation. Elongation in eukaryotic cells takes place in a similar manner. Eukaryotes possess at least three elongation factors, one of which also acts in initiation and termination. Another

of the elongation factors used in eukaryotes, called elongation factor 2 (EF-2) is the target of a toxin produced by bacteria that causes diphtheria, a disease that, until recently, was a leading killer of children. The diphtheria toxin inhibits EF-2, preventing translocation of the ribosome along the mRNA, and protein synthesis ceases.

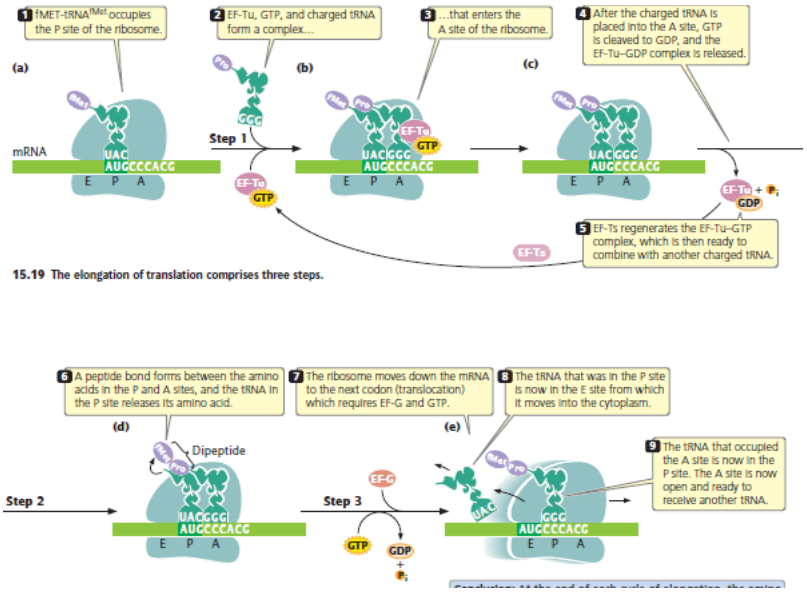


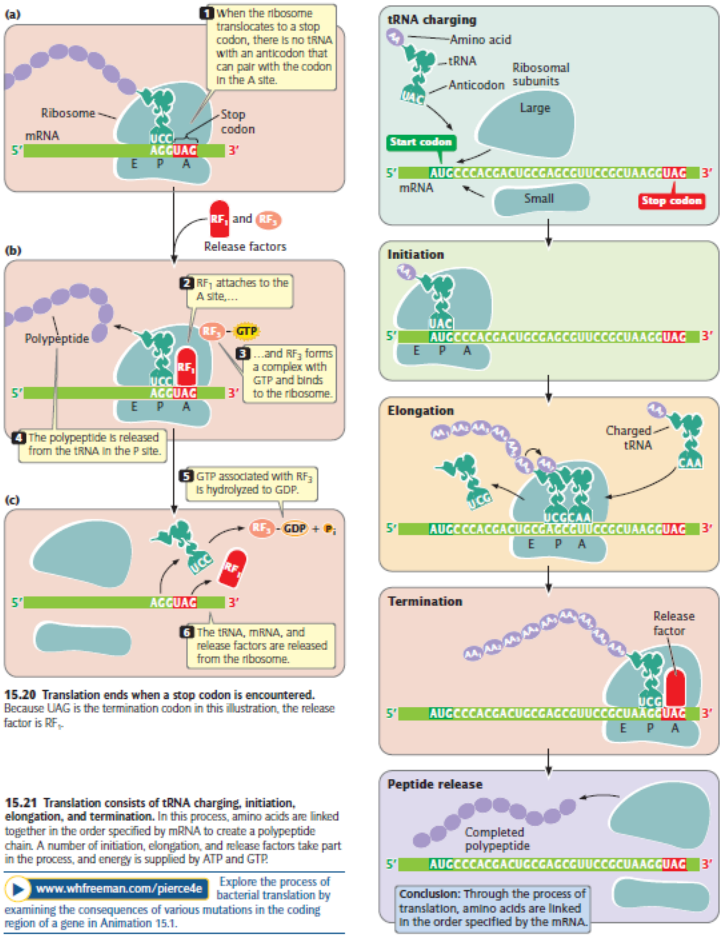

**Figure 58. Elongation**

**Termination**

Protein synthesis terminates when the ribosome translocates to a termination codon. Because there are no tRNAs with anticodons complementary to the termination codons, no tRNA enters the A site of the ribosome when a termination codon is encountered. Instead, proteins called **release factors** bind to the ribosome. *E. coli* has three release factors—RF1, RF2, and RF3. Release factor 1 binds to the termination codons UAA and UAG, and RF2 binds to UGA and UAA. The binding of release factor RF1 or RF2 to the A site of the ribosome promotes the cleavage of the tRNA in the P site from the polypeptide chain and the release of the polypeptide. Release factor 3 binds to the ribosome and forms a complex with GTP. This binding brings about a conformational change in the ribosome, releasing RF1 or RF2 from the A site and causing the tRNA in the P site to move to the E site; in the process, GTP is hydrolyzed to GDP. Additional factors help bring about the

release of the tRNA from the P site, the release of the mRNA from the ribosome, and the dissociation of the ribosome.

Translation in eukaryotic cells terminates in a similar way, except that there are two release factors: eRF1, which recognizes all three termination codons, and eRF2, which binds GTP and stimulates the release of the polypeptide from the ribosome.




**15.20 Translation ends when a stop codon is encountered.** Because UAG is the termination codon in this illustration, the release factor is $RF_1$.

**15.21 Translation consists of tRNA charging, initiation, elongation, and termination.** In this process, amino acids are linked together in the order specified by mRNA to create a polypeptide chain. A number of initiation, elongation, and release factors take part in the process, and energy is supplied by ATP and GTP.

www.whfreeman.com/pierce4e Explore the process of bacterial translation by examining the consequences of various mutations in the coding region of a gene in Animation 15.1.

**Figure 59. Termination**

## POST TRANSLATIONAL MODIFICATIONS

After synthesis is completed, proteins can be modified by various methods such as phosphorylation, glycosylation, ADP ribosylation, hydroxylation, and addition of other groups.

**Proteolysis**

As the newly synthesized protein is released in the lumen of the ER, signal peptidases cleave peptide sequence. Apart from signal peptide, some polypeptide sequence of the protein is also cleaved resulting in the final sequence.

Example:

Insulin is synthesized in the cells in its inactive form which cannot perform its function. Post translational modifications ensure proper function which involves the removal of the part of protein to convert it into a three dimensional and fully active form.

**Phosphorylation**

Phosphoryalation is the addition of one or more phosphate groups to the protein. Post Translational Phosphorylation is one of the most common protein modifications that occur in animal cells. Majority of phosphorylation occurs as a mechanism to regulate the biological activity of a protein. In animal cells Serine, tyrosine and thereonine are the amino acids that subjected to the phosphorylation.

**Glycosylation**

Glycosylation is the addition of carbohydrate molecules to the polypeptide chain and modifying it into glycoproteins. Many of the proteins that are destined to become a part of plasma membrane or to be secreted from the cell, have carbohydrate chains attached to the amide nitrogen of asparagine(N linked) or the hydroxyl groups of serine, threonine(O linked). N glycosylation occurs in ER and O glycosylation occurs in the golgi complex.

**Sulfation**

Sulfate modidication takes place by the addition of sulphate molecules and these modifications of proteins occurs at tyrosine residues. Tyrosine sulfation accomplished via the activity of tyrosylproteinsulfotransferases (TPST) which are membrane associated enzymes of trans-Golgi network. There are two known TPSTs. TPST-1 TPST-2 The universal phosphate donor is 3'-phosphoadenosyl- 5'-phosphosulphate (PSPA).

**Methylation**

The transfer of one-carbon methyl groups to nitrogen or oxygen to amino acid side chains increases the hydrophobicity of the protein and can neutralize a negative amino acid charge when bound to carboxylic acids. Methylation is mediated by methyltransferases and S-adenosyl methionine (SAM) is the primary methyl group donor.

**Hydroxylation**

The biological process of addition of a hydroxy group to a protein amino acid is called Hydroxylation. Protein hydroxylation is one type of PTM that involves the conversion of –CH group into –COH group and these hydroxylated amino acids are involved in the regulation of some important factors called transcription factors. Among the 20 amino acids, the two amino acids regulated by this method are proline and lysine.

**Others**

**a) SUMOylation**

SUMO (small ubiquitin related modifier) proteins are 100 amino acid residue proteins which bind to the target protein in the same way as ubiquitin. They also confer the transcription regulatory activity of the protein and help in the transport of the target protein from cytosol to the nucleus.

b) Disulfide bond formation

Stabilizes protein structure and involved in redox processes.

c) Lipidylation, Acetylation, Prenylation etc.

**SIGNIFICANCE**

Proteins are synthesized by ribosomes translating mRNA into polypeptide chains, which may then undergo modifications to form the mature protein product.

Post-translational modifications of proteins, which are not gene- template based, can regulate the protein functions, by causing changes in protein activity, their cellular locations and dynamic interactions with other proteins.

PTMs have significant biological functions which include:

Aids in proper protein folding – few lectin molecules called calnexin binds to glycosylated proteins and assist in its folding.

Confers stability to the protein- glycosylation can modify the stability of the protein by increasing protein half life.

It protects the protein against cleavage by proteolytic enzyme by blocking the cleavage sites.

Protein sorting or translocation- If phosphorylated mannose residues are present in the protein it always goes to lysosome.

It regulates protein activity and function- phosphorylation of protein is a reversible PTM which activates the protein.

Acetylation regulates many diverse functions, including DNA recognition, protein-protein interaction and protein stability.

Redox-dependent PTM of proteins is emerging as a key signaling system conserved through evolution, influences many aspects of cellular homeostasis.

PTMs are important components in cell signaling, as for example when prohormones are converted to hormones.

It significantly increases the diversity and complexity in the proteome.

**INHIBITING PROTEIN SYNTHESIS**

- Protein synthesis inhibitors usually act at the ribosome level, taking advantage of the major differences between prokaryotic and eukaryotic ribosome structures.
- Protein synthesis inhibitors work at different stages of prokaryotic mRNA translation into proteins like initiation, elongation (including aminoacyl tRNA entry, proofreading, peptidyl transfer, and ribosomal translocation), and termination.
- By targeting different stages of the mRNA translation, antimicrobial drugs can be changed if resistance develops.

A protein synthesis inhibitor is a substance that stops or slows the growth or proliferation of cells by disrupting the processes that lead directly to the generation of new proteins. It usually refers to substances, such as antimicrobial drugs, that act at the ribosome level. The substances take advantage of the major differences between prokaryotic and eukaryotic ribosome structures which differ in their size, sequence, structure, and the ratio of protein to RNA. The differences in structure allow some antibiotics to kill bacteria by inhibiting their ribosomes, while leaving human ribosomes unaffected.

Translation in prokaryotes involves the assembly of the components of the translation system which are: the two ribosomal subunits (the large 50S & small 30S subunits), the mRNA to be translated, the first aminoacyl tRNA, GTP (as a source of energy), and three initiation factors that help the assembly of the initiation complex. The ribosome has three sites: the A site, the P site, and the E site (not shown in ). The A site is the point of entry for the aminoacyl tRNA. The P site is where the peptidyl tRNA is formed in the ribosome. The E site which is the exit site of the now uncharged tRNA after it gives its amino acid to the growing peptide chain.

In general, protein synthesis inhibitors work at different stages of prokaryotic mRNA translation into proteins like initiation, elongation (including aminoacyl tRNA entry, proofreading, peptidyl transfer, and ribosomal translocation), and termination. The following is a list of common antibacterial drugs and the stages which they target.

- Linezolid acts at the initiation stage, probably by preventing the formation of the initiation complex, although the mechanism is not fully understood.
- Tetracyclines and Tigecycline (a glycylcycline related to tetracyclines) block the A site on the ribosome, preventing the binding of aminoacyl tRNAs.
- Aminoglycosides, among other potential mechanisms of action, interfere with the proofreading process, causing an increased rate of error in synthesis with premature termination.
- Chloramphenicol blocks the peptidyl transfer step of elongation on the 50S ribosomal subunit in both bacteria and mitochondria.
- Macrolides, clindamycin, and aminoglycosides have evidence of inhibition of ribosomal translocation.
- Streptogramins also cause premature release of the peptide chain.

By targeting different stages of the mRNA translation, antimicrobial drugs can be changed if resistance develops to one or many of the drugs.

# UNIT –V – GENE REGULATION

**The Operon**

Within its tiny cell, the bacterium *E. coli* contains all the genetic information it needs to metabolize, grow, and reproduce. It can synthesize every organic molecule it needs from glucose and a number of inorganic ions.

Many of the genes in *E. coli* are expressed constitutively; that is, they are always turned "on". Others, however, are active only when their products are needed by the cell, so their expression must be regulated.

Two examples:

- If the amino acid tryptophan (**Trp**) is added to the culture, the bacteria soon stop producing the five enzymes previously needed to synthesize Trp from intermediates produced during the respiration of glucose. In this case, the presence of the products of enzyme action **represses** enzyme synthesis.
- Conversely, adding a new substrate to the culture medium may **induce** the formation of new enzymes capable of metabolizing that substrate. If we take a culture of *E. coli* that is feeding on glucose and transfer some of the cells to a medium contain lactose instead, a revealing sequence of events takes place.
    - At first the cells are quiescent: they do not metabolize the lactose, their other metabolic activities decline, and cell division ceases.
    - Soon, however, the culture begins growing rapidly again with the lactose being rapidly consumed. What has happened? During the quiescent interval, the cells began to produce **three enzymes**.

The three enzymes are

- a **permease** that transports lactose across the plasma membrane from the culture medium into the interior of the cell
- **beta-galactosidase** which converts lactose into the intermediate allolactose and then hydrolyzes this into glucose and galactose. Once in the presence of lactose, the quantity of beta-galactosidase in the cells rises from a tiny amount to almost 2% of the weight of the cell.

- a **transacetylase** whose function is still uncertain.

**The *lac* operon**

The capacity to respond to the presence of lactose was always there. The genes for the three induced enzymes are part of the genome of the cell. But until lactose was added to the culture medium, these genes were not expressed (β-galactosidase was expressed weakly — just enough to convert lactose into allolactose).

The most direct way to control the expression of a gene is to **regulate its rate of transcription**; that is, the rate at which RNA polymerase transcribes the gene into molecules of messenger RNA (mRNA).

Gene transcription begins at a particular nucleotide shown in the figure as "**+1**". RNA polymerase actually binds to a site "upstream" (i.e., on the 5' side) of this site and opens the double helix so that transcription of one strand can begin.

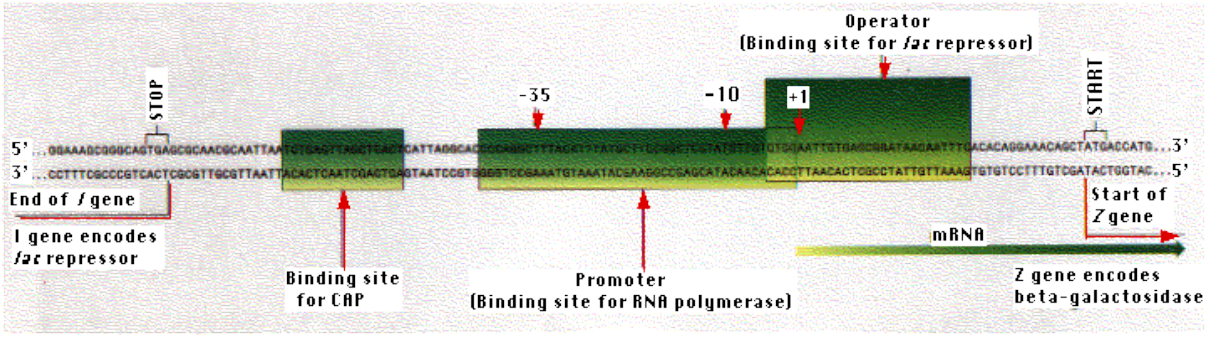



**Figure 60. Transcription**

The binding site for RNA polymerase is called the **promoter**. In bacteria, two features of the promoter appear to be important:

- a sequence of TATAAT (or something similar) centered 10 nucleotides upstream of the **+1** site and
- another sequence (TTGACA or something quite close to it) centered 35 nucleotides upstream.

The exact DNA sequence between the two regions does not seem to be important.

Each of the three enzymes synthesized in response to lactose is encoded by a separate gene. The three genes are arranged in tandem on the bacterial chromosome.

**The *lac* operon.** In the absence of lactose, the repressor protein encoded by the I gene binds to the *lac* operator and prevents transcription. Binding of allolactose to the repressor causes it to leave the operator. This enables RNA polymerase to transcribe the three genes of the operon. The single mRNA molecule that results is then translated into the three proteins.

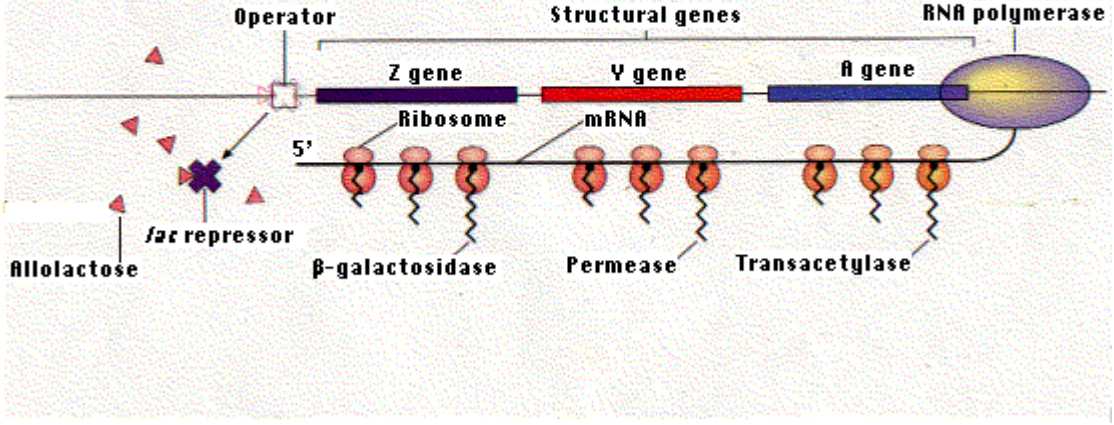



**Figure 61. Lac Operon**

The ***lac* repressor** binds to a specific sequence of two dozen nucleotides called the **operator**. Most of the operator is downstream of the promoter. When the repressor is bound to the operator, RNA polymerase is unable to proceed downstream with its task of gene transcription.

The operon is the combination of the

- operator and
- the three protein-encoding genes associated with it.

The gene encoding the *lac* repressor is called the *I* gene. It happens to be located just upstream of the *lac* promoter. However, its precise location is probably not important because it achieves its

effect by means of its protein product, which is free to diffuse throughout the cell. And, in fact, the genes for some repressors are not located close to the operators they control.

Although repressors are free to diffuse through the cell, how does — for example — the lac repressor find the single stretch of 24 base pairs of the operator out of the 4.6 million base pairs of DNA in the *E. coli* genome? It turns out the repressor is free to bind **anywhere** on the DNA using both

- hydrogen bonds and
- ionic (electrostatic) interactions between its positively-charged amino acids (Lys, Arg) and the negative charges on the deoxyribose-phosphate backbone of the DNA.

Once astride the DNA, the repressor can move along it until it encounters the operator sequence. Now an allosteric change in the tertiary structure of the protein allows the same amino acids to establish bonds — mostly hydrogen bonds and hydrophobic interactions — with particular bases in the operator sequence.

The *lac* repressor is made up of four identical polypeptides (thus a "homotetramer"). Part of the molecule has a site (or sites) that enable it to recognize and bind to the 24 base pairs of the lac operator. Another part of the repressor contains sites that bind to allolactose. When allolactose unites with the repressor, it causes a change in the shape of the molecule, so that it can no longer remain attached to the DNA sequence of the operator. Thus, when lactose is added to the culture medium,

- it causes the repressor to be released from the operator
- RNA polymerase can now begin transcribing the 3 genes of the operon into a **single** molecule of **messenger RNA**.

Hardly does transcription begin, before ribosomes attach to the growing mRNA molecule and move down it to **translate** the message into the three proteins. You can see why punctuation codons — UAA, UAG, or UGA — are needed to terminate translation between the portions of the mRNA coding for each of the three enzymes.

This mechanism is characteristic of bacteria, but differs in several respects from that found in eukaryotes:

- Genes in eukaryotes are not linked in operons (except for nematodes like *C. elegans* and tunicates like *Ciona intestinalis*).
- Primary transcripts in eukaryotes contain the transcript of only a single gene (with the above exceptions).
- Transcription and translation are not physically linked in eukaryotes as they are in bacteria; transcription occurs in the nucleus while translation occurs in the cytosol (with a few exceptions).

**Corepressors**

As mentioned above, the synthesis of tryptophan from precursors available in the cell requires 5 enzymes. The genes encoding these are clustered together in a single operon with its own promoter and operator. In this case, however, the **presence** of tryptophan in the cell **shuts down** the operon. When Trp is present, it binds to a site on the Trp repressor and **enables** the Trp repressor to bind to the operator. When Trp is not present, the repressor leaves its operator, and transcription of the 5 enzyme-encoding genes begins.

The usefulness to the cell of this control mechanism is clear. The presence in the cell of an essential metabolite, in this case tryptophan, turns off its own manufacture and thus stops unneeded protein synthesis.

As its name suggests, repressors are **negative control** mechanisms, shutting down operons

- in the absence of a substrate (lactose in our example) or
- the presence of an essential metabolite (tryptophan is our example).

However, some gene transcription in *E. coli* is under positive control.

**Positive Control of Transcription: CAP**

Absence of the lac repressor is essential but not sufficient for effective transcription of the lac operon. The activity of RNA polymerase also depends on the presence of another DNA-binding protein called **catabolite activator protein** or **CAP**. Like the lac repressor, CAP has two types of binding sites:

- One binds the nucleotide cyclic AMP; the other
- binds a sequence of 16 base pairs upstream of the promoter

However, CAP can bind to DNA only when cAMP is bound to CAP. so when cAMP levels in the cell are low, CAP fails to bind DNA and thus RNA polymerase cannot begin its work, even in the absence of the repressor.

So the *lac* operon is under both **negative** (the **repressor**) and **positive** (**CAP**) control.

Why?

It turns out that it is not simply a matter of belt and suspenders. This dual system enables the cell to **make choices**. What, for example, should the cell do when fed both glucose and lactose? Presented with such a choice, *E. coli* (for reasons about which we can only speculate) chooses glucose. It makes its choice by using the interplay between these two control devices.

- Although the presence of lactose removes the repressor,
- the presence of glucose lowers the level of cAMP in the cell and thus removes CAP.

Without CAP, binding of RNA polymerase is inhibited even though there is no repressor to interfere with it if it could bind. The molecular basis for its choices is shown in the graphic.

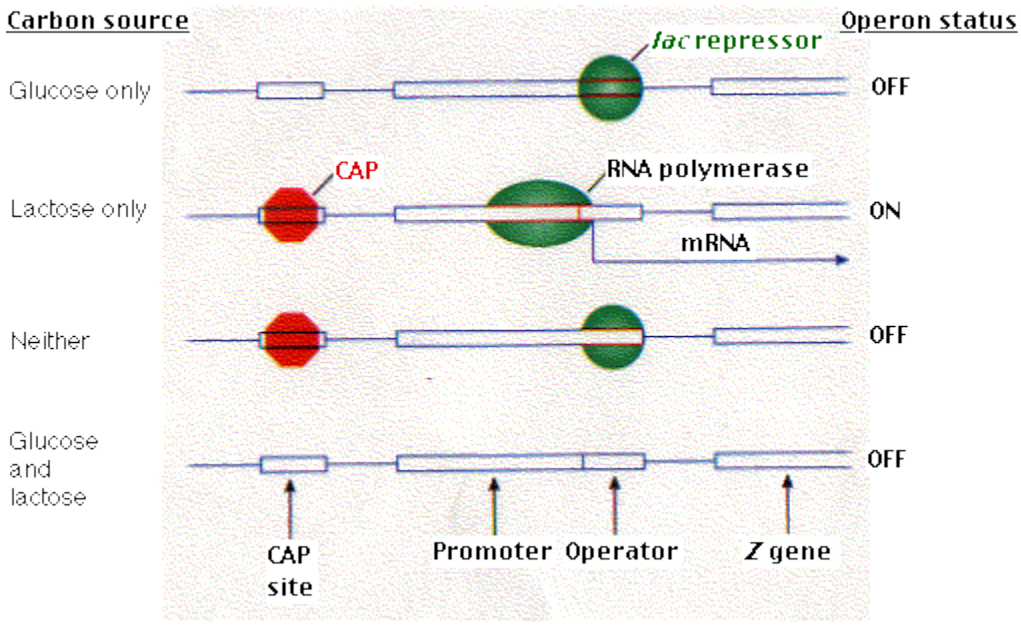



**Figure 62. Carbon source and operon status**

CAP consists of two identical polypeptides (hence it is a homodimer). Toward the C-terminal, each has two regions of alpha helix with a sharp bend between them. The longer of these is called the **recognition helix** because it is responsible for recognizing and binding to a particular sequence of bases in DNA.

The graphic shows a model of CAP. The two monomers are identical. Each monomer recognizes a sequence of nucleotides in DNA by means of the region of alpha helix labeled **F**. Note that the two recognition helices are spaced 34Å apart, which is the distance that it takes the DNA molecule (on the left) to make precisely one complete turn.

The recognition helices of each polypeptide of CAP are, of course, identical. But their orientation in the dimer is such that the sequence of bases they recognize must run in the opposite direction for each recognition helix to bind properly. This arrangement of two identical sequences of base pairs running in opposite directions is called an **inverted repeat**.

The strategy illustrated by CAP and its binding site has turned out to be used widely. As more and more DNA-regulating proteins have been discovered, many turn out to share the traits we find in CAP:

- They usually contain two subunits. Therefore, they are **dimers**.
- They recognize and bind to DNA sequences with **inverted repeats**.
- In bacteria, recognition and binding to a particular sequence of DNA is accomplished by a segment of alpha helix. Hence these proteins are often described as **helix-turn-helix** proteins. The Trp repressor shown above is a member of this group.

**Riboswitches**

**Protein** repressors and corepressors are not the only way in which bacteria control gene transcription. It turns out that the regulation of the level of certain metabolites can also be controlled by riboswitches. A riboswitch is section of the 5'-untranslated region (5'-UTR) in a

molecule of messenger RNA (mRNA) which has a specific binding site for the metabolite (or a close relative).

Some of the metabolites that bind to riboswitches:

- the purines adenine and guanine
- the amino acids glycine and lysine
- flavin mononucleotide (the prosthetic group of NADH dehydrogenase)
- S-adenosyl methionine (that donates methyl groups to many molecules, including
    - DNA
    - the cap at the 5' end of messenger RNA [Link]
- tRNAs. When these are bound to their amino acid (aminoacyl-tRNA), they bind to the riboswitch in the mRNA that encodes the enzyme (an aminoacyl-tRNA synthetase) responsible for loading the amino acid onto the tRNA. This causes transcription of the mRNA to terminate prematurely. tRNAs with no amino acid attached also bind to the riboswitch but in such a way that transcription of the mRNA continues. Its translation (in bacteria, translation begins while transcription is still going on [Link]) produces the aminoacyl-tRNA synthetase used to load the amino acid onto the tRNA. Thus these riboswitches regulate the level of aminoacyl-tRNAs producing more when needed, less when not (a kind of feedback inhibition.)

In each case, the riboswitch regulates **transcription** of genes involved in the metabolism of that molecule. The metabolite binds to the growing mRNA and induces an allosteric change that

- for some genes causes further synthesis of the mRNA to terminate before forming a functional product and
- for other genes, enhances completion of synthesis of the mRNA.
- In both cases, one result is to control the level of that metabolite.

Some riboswitches control mRNA **translation** rather than its transcription.

It has been suggested that these regulatory mechanisms, which **do not involve any protein**, are a relic from an "RNA world".

Tryptophan (Trp) Operon

Many protein-coding genes in bacteria are clustered together in operons which serve as transcriptional units that are coordinately regulated.

It was Jacob and Monod in 1961 who proposed the operon model for the regulation of transcription.

The operon model proposes three elements:

A set of structural genes (i.e. genes encoding the proteins to be regulated);

An operator site, which is a DNA sequence that regulates transcription of the structural genes.

A regulator gene which encodes a protein that recognizes the operator sequence.

Operons are thus clusters of structural genes under the control of a single operator site and regulator gene which ensures that expression of the structural genes is coordinately controlled.

Trp Operon

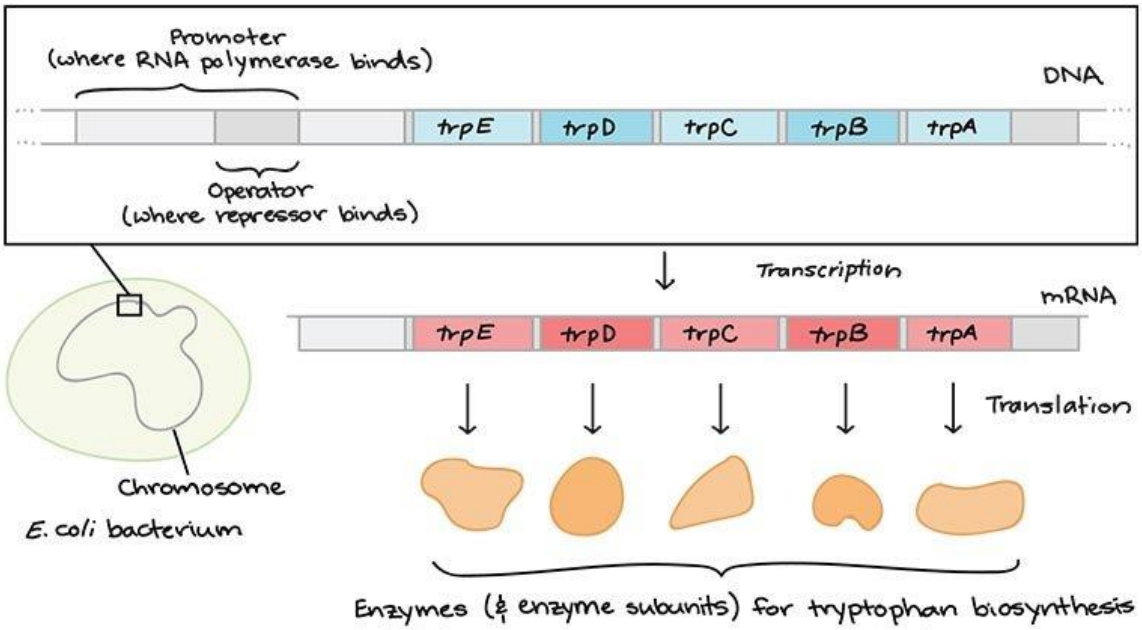



**Figure 63a. trp Operon**

The tryptophan operon is the regulation of transcription of the gene responsible for biosynthesis of tryptophan.

The tryptophan (trp) operon contains five structural genes encoding enzymes for tryptophan biosynthesis with an upstream trp promoter (Ptrp) and trp operator sequence (Otrp).

Structural genes are TrpE, TrpD, TrpC, TrpB and TrpA

trpE: It enodes the enzyme Anthranilate synthase I

trpD: It encodes the enzyme Anthranilate synthase II

trpC: It encodes the enzyme N-5'-Phosphoribosyl anthranilate isomerase and Indole-3-glycerolphosphate synthase

trpB: It encodes the enzyme tryptophan synthase-B sub unit

trpA: It encode the enzyme tryptophan synthase-A sub unit

The trp operator region partly overlaps the trp promoter.

The operon is regulated such that transcription occurs when tryptophan in the cell is in short supply.

Ever Wonder What Extraterrestrials Might Be Like If We Met Them?

In the Absence of Tryptophan

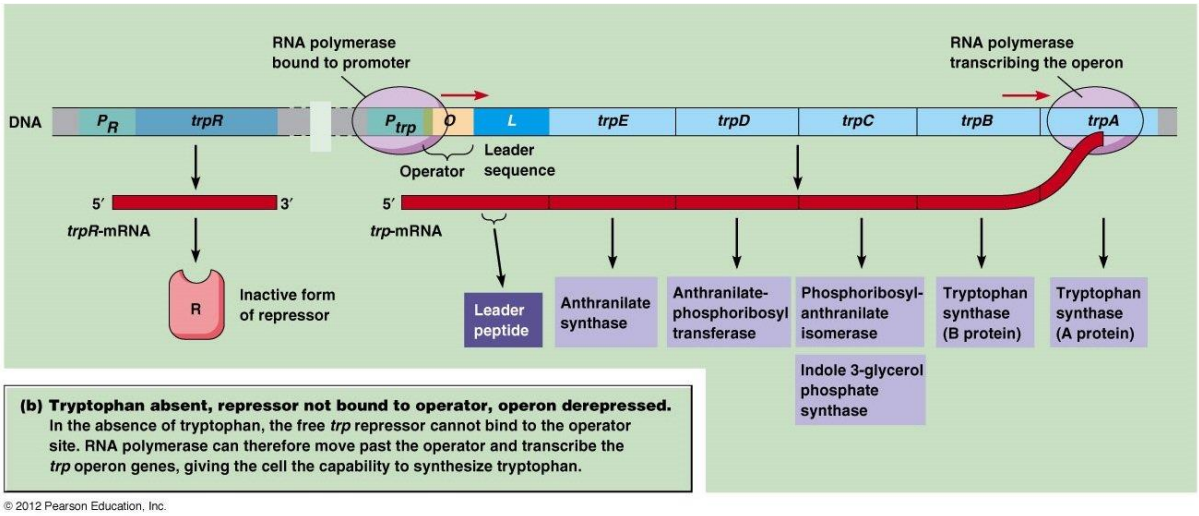



**Figure 63b. trp Operon**

In the absence of tryptophan, a trp repressor protein encoded by a separate operon, trpR, is synthesized and forms a dimer.

However, this is inactive and so is unable to bind to the trp operator and the structural genes of the trp operon are transcribed.

In the Presence of Tryptophan

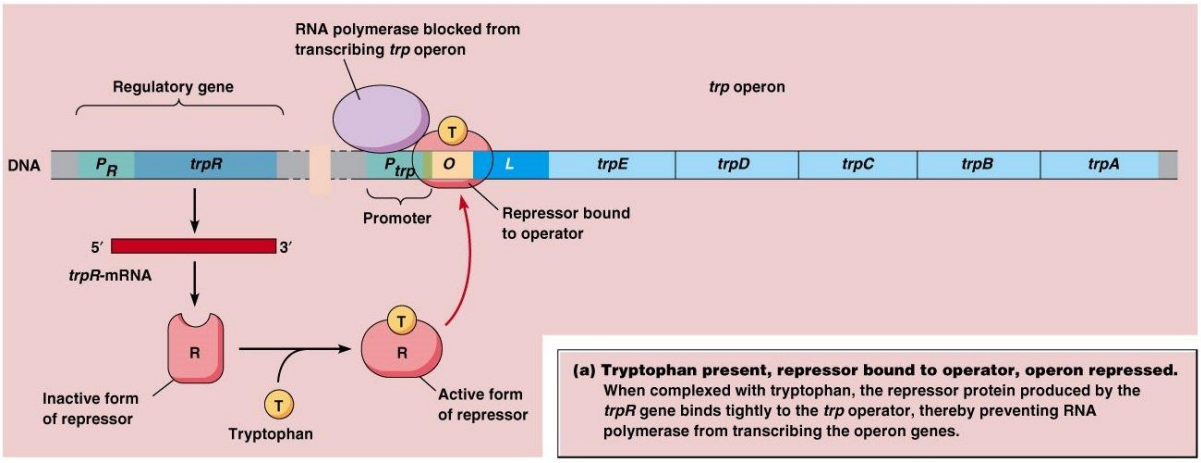



**Figure 63c. trp Operon**

When tryptophan is present, the enzymes for tryptophan biosynthesis are not needed and so expression of these genes is turned off.

This is achieved by tryptophan binding to the repressor to activate it so that it now binds to the operator and stops transcription of the structural genes.

Binding of repressor protein to operator overlaps the promoter, so RNA polymerase cannot bind to the prometer. Hence transcription is halted.

In this role, tryptophan is said to be a co-repressor. This is negative control, because the bound repressor prevents transcription.

Trp Operon Attenuation

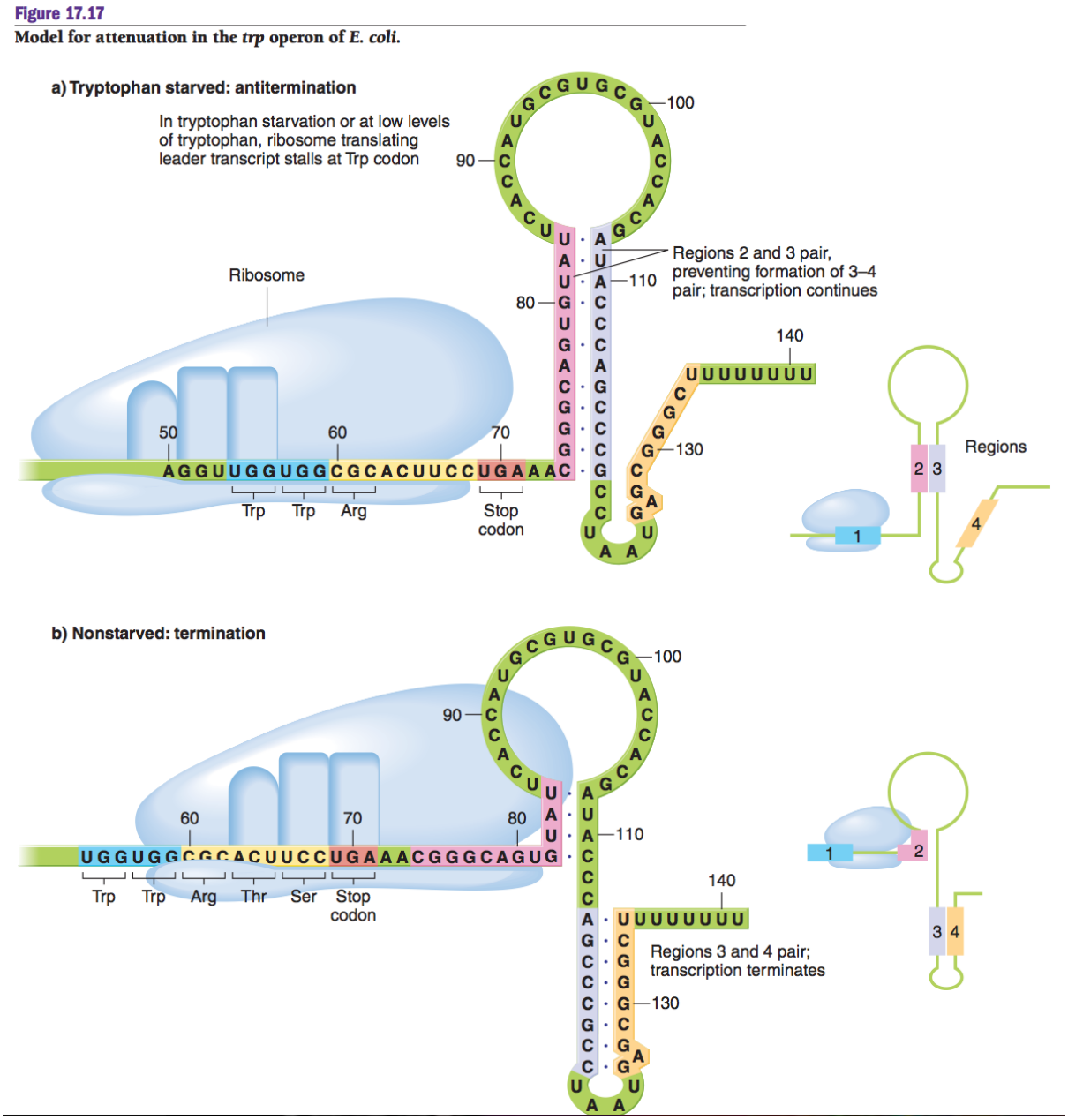



**Figure 64a. Attenuation**

A second mechanism, called attenuation, is also used to control expression of the trp operon.

The 5′ end of the polycistronic mRNA transcribed from the trp operon has a leader sequence upstream of the coding region of the trpE structural gene.

This leader sequence encodes a 14 amino acid leader peptide containing two tryptophan residues.

The function of the leader sequence is to fine tune expression of the trp operon based on the availability of tryptophan inside the cell.

The leader sequence contains four regions (numbered 1–4) that can form a variety of base paired stem-loop ('hairpin') secondary structures.

The regions are: Region 1, region 2, region 3 and Region 4. Region 3 is complementary to both region 2 and region 4.

If region 3 and region 4 base pair with each other, they form a loop like structure called attenuator and it function as transcriptional termination. If pairing occur between region 3 and region 2, then no such attenuator form so that transcription continues.

Attenuation depends on the fact that, in bacteria, ribosomes attach to mRNA as it is being synthesized and so translation starts even before transcription of the whole mRNA is complete.

When Trypophan is abundant

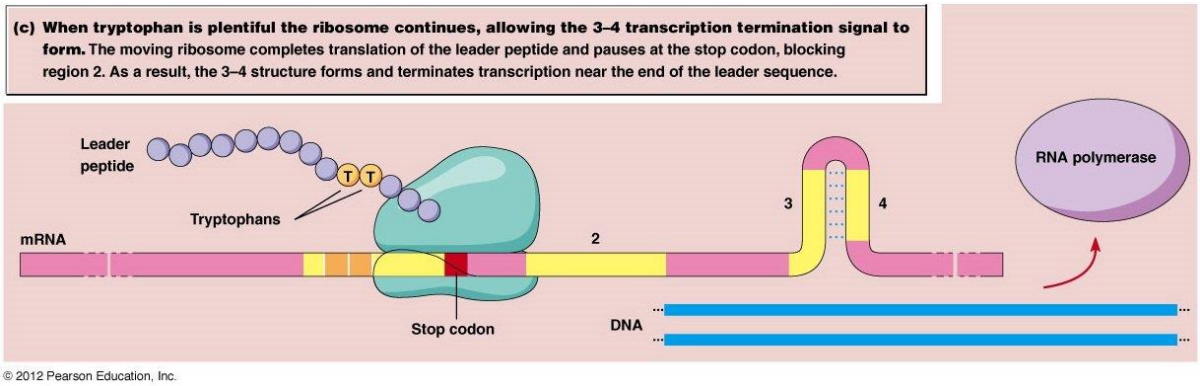



**Figure 64b. Attenuation**

When tryptophan is abundant, ribosomes bind to the trp polycistronic mRNA that is being transcribed and begin to translate the leader sequence.

Now, the two trp codons for the leader peptide lie within sequence 1, and the translational Stop codon lies between sequence 1 and 2.

During translation, the ribosomes follow very closely behind the RNA polymerase and synthesize the leader peptide, with translation stopping eventually between sequences 1 and 2.

At this point, the position of the ribosome prevents sequence 2 from interacting with sequence 3.

Instead sequence 3 base pairs with sequence 4 to form a 3:4 stem loop which acts as a transcription terminator.

Therefore, when tryptophan is present, further transcription of the trp operon is prevented.

GWhen Trypophan is scarce

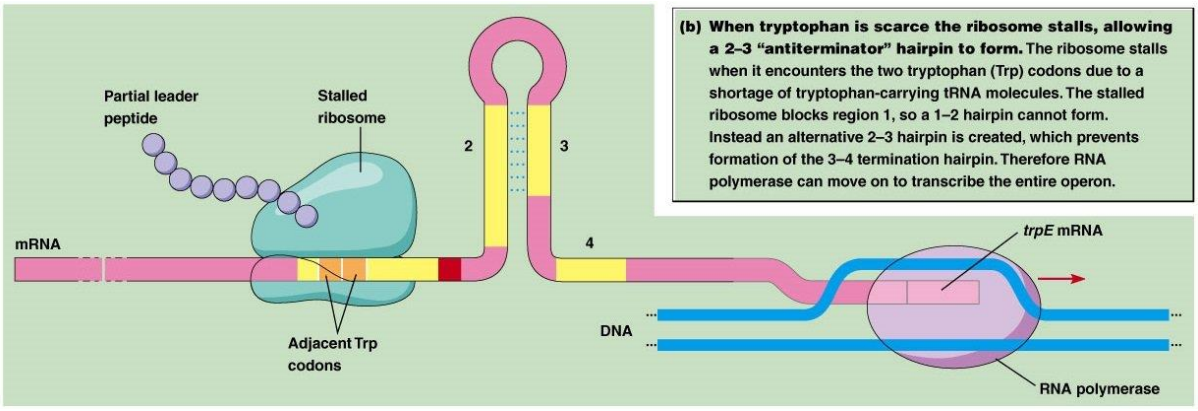



**Figure 64c. Attenuation**

If, however, tryptophan is in short supply, the ribosome will pause at the two trp codons contained within sequence 1.

This leaves sequence 2 free to base pair with sequence 3 to form a 2:3 structure (also called the anti-terminator), so the 3:4 structure cannot form and transcription continues to the end of the trp operon.

Hence the availability of tryptophan controls whether transcription of this operon will stop early (attenuation) or continue to synthesize a complete polycistronic mRNA.

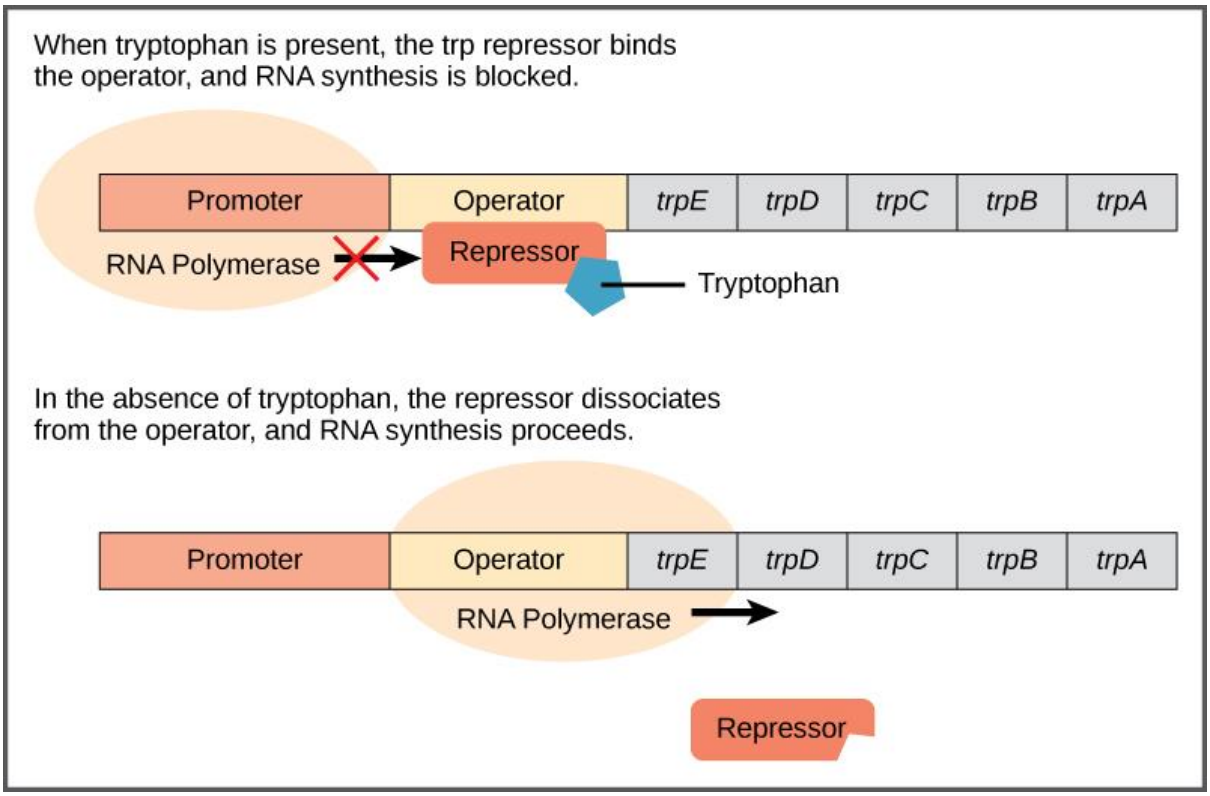



**Figure 65. Regulation of Trp Operon**

Overall, for the trp operon, repression via the trp repressor determines whether transcription will occur or not and attenuation then fine tunes transcription.

Dual Positive And Negative Control: the Arabinose Operon

Prokaryotic control of transcription seems to mix and match different aspects of positive and negative regulation in different ways. Thus the general themes from our detailed examination of the lactose operon should be on how positive and negative regulation can work through the activity of sequence-specific, allosterically regulated DNA-binding proteins. Even in the metabolism of other sugars, we see transcriptional regulatory mechanisms that are certainly different in detail. The arabinose operon is such an example (Figure 14-16).

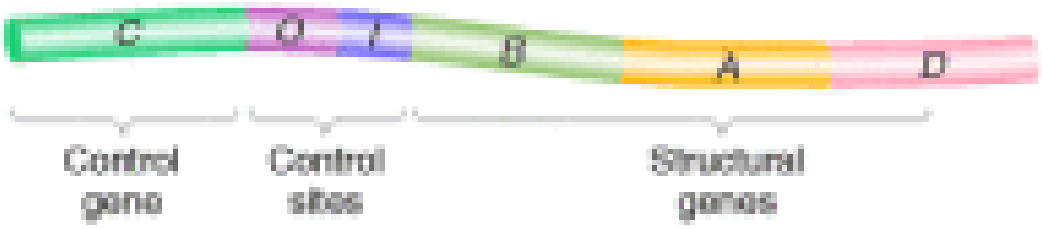



**Figure 66. ara Operon**

Map of the ara region. The B, A, and D genes together with the I and O sites constitute the ara operon.

The structural genes (araB, araA, and araD) that encode the metabolic enzymes that break down arabinose are transcribed as a multigenic mRNA. Transcription is activated at araI, the initiator region, which contains both an operator site and a promoter. The araC gene encodes an activator protein that, when bound to arabinose, activates transcription of the ara operon, perhaps by helping RNA polymerase bind to the promoter, located within in the araI region. An additional activation event is mediated by the same CAP–cAMP catabolite repression system that regulates lac operon expression.

In the presence of arabinose, both the CAP–cAMP complex and the AraC–arabinose complex must bind to the initiator region in order for RNA polymerase to bind to the promoter and transcribe the ara operon (Figure 14-17a). In the absence of arabinose, the AraC protein assumes a different conformation and represses the ara operon by binding both to araI and to a second operator region, araO, thereby forming a loop (Figure 14-17b) that prevents transcription. Thus, the AraC protein has two conformations, one that acts as an activator and the other that acts

as a repressor. The two conformations, dependent on whether the allosteric effector has bound to the protein, also differ in their abilities to bind a specific target site in the araO region of the operon.

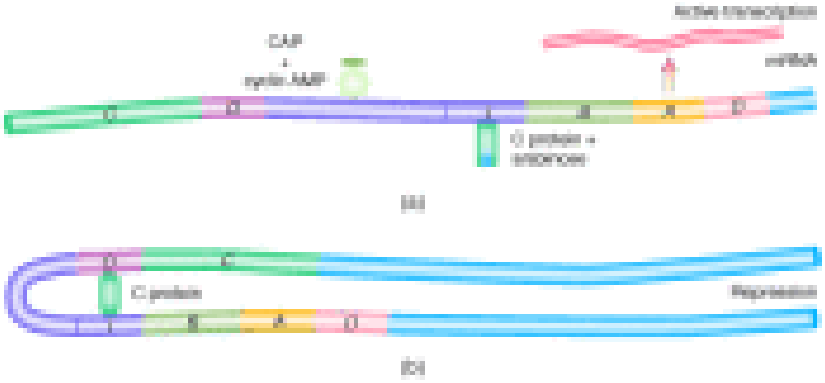


**Figure 67.** Dual control of the ara operon. (a) In the presence of arabinose, the AracC protein binds to the araI region and, when bound to cAMP, the CAP protein binds to a site adjacent to araI. This stimulates the transcription of the araB, araA, and araD genes. (more...)

MESSAGE

Operon transcription can be regulated by both activation and repression. Operons regulating the metabolism of similar compounds, such as sugars, can be regulated in quite different ways.

## REFERENCES

### Unit I


Cell Biology, Genetics, Molecular Biology, Evolution and Ecology, P.S. Verma, V.K. Agarwal, S. Chand & Company Ltd, 2005

Principles of Molecular Biology, Veer Bala Rastogi, Medtech, 2016

Molecular Biology of the Cell. 4th edition. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.

http://www.ncbi.nlm.nih.gov/books/NBK22104/

https://www.google.co.in/search?q=dna+as+genetic+material&biw=1280&bih=890&tbm=isch&imgil=TUQEOSzOlQrxuM%253A%253B5_P9-NeMFLTYMM%253Bhttps%25253A%25252F%25252Fwww.youtube.com%25252Fwatch%25253Fv%2525253DATtWiTqn1Ok&source=iu&pf=m&fir=TUQEOSzOlQrxuM%253A%252C5_P9-NeMFLTYMM%252C_&dpr=1&usg=__t7Dy9JbSGZodbUp0eyZeNP5aO7w%3D&ved=0ahUKEwj3-vLAyqPKAhUVCY4KHTtrBisQyjcIUA&ei=k5SUVrflMpWSuAS71pnYAg#imgrc=TUQEOSzOlQrxuM%3A&usg=__t7Dy9JbSGZodbUp0eyZeNP5aO7w%3D

**Unit II**

Cell Biology, Genetics, Molecular Biology, Evolution and Ecology, P.S. Verma, V.K. Agarwal, S. Chand & Company Ltd, 2005

Principles of Molecular Biology, Veer Bala Rastogi, Medtech, 2016

Molecular Biology of the Cell. 4th edition. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.

http://www.ncbi.nlm.nih.gov/books/NBK22104/

https://www.google.co.in/search?q=dna+as+genetic+material&biw=1280&bih=890&tbm=isch&imgil=TUQEOSzOlQrxuM%253A%253B5_P9-NeMFLTYMM%253Bhttps%25253A%25252F%25252Fwww.youtube.com%25252Fwatch%25253Fv%2525253DATtWiTqn1Ok&source=iu&pf=m&fir=TUQEOSzOlQrxuM%253A%252C5_P9-NeMFLTYMM%252C_&dpr=1&usg=__t7Dy9JbSGZodbUp0eyZeNP5aO7w%3D&ved=0ahUKEwj3-vLAyqPKAhUVCY4KHTtrBisQyjcIUA&ei=k5SUVrflMpWSuAS71pnYAg#imgrc=TUQEOSzOlQrxuM%3A&usg=__t7Dy9JbSGZodbUp0eyZeNP5aO7w%3D

http://biologyntechnology.blogspot.in/2009/10/rna-as-genetic-material.html

www.bio.miami.edu/mccracken/genetics/lectures-2013/chapter_02.ppt

http://www.bx.psu.edu/~ross/workmg/RepairDNACh7.pdf

**Unit III**

Cell Biology, Genetics, Molecular Biology, Evolution and Ecology, P.S. Verma, V.K. Agarwal, S. Chand & Company Ltd, 2005

Principles of Molecular Biology, Veer Bala Rastogi, Medtech, 2016

Molecular Biology of the Cell. 4th edition. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.

www.namrata.com

http://highered.mcgrawhill. com/olcweb/cgi/pluginpop.cgi?it=swf:: 535::535::/sites/dl/free/0072437316/120077/bio30.swf::How%20Spliceosomes%20Process%20RNA

**Unit IV**

Cell Biology, Genetics, Molecular Biology, Evolution and Ecology, P.S. Verma, V.K. Agarwal, S. Chand & Company Ltd, 2005

Principles of Molecular Biology, Veer Bala Rastogi, Medtech, 2016

https://microbenotes.com/post-translational-modification/

https://bio.libretexts.org/Bookshelves/Microbiology/Book%3A_Microbiology_(Boundless)/13%3A_Antimicrobial_Drugs/13.2%3A_Functions_of_Antimicrobial_Drugs/13.2D%3A_Inhibiting_Protein_Synthesis

**Unit V**

Cell Biology, Genetics, Molecular Biology, Evolution and Ecology, P.S. Verma, V.K. Agarwal, S. Chand & Company Ltd, 2005

Principles of Molecular Biology, Veer Bala Rastogi, Medtech, 2016

Molecular Biology of the Cell. 4th edition. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.

https://microbenotes.com/tryptophan-trp-operon/

https://www.ncbi.nlm.nih.gov/books/NBK21277/