Ministry of Higher Education and Scientific Research
University of Baghdad
College of Science
Department of Biology

Theoretical

Molecular Biology

2023-2022

المرحلة الرابعة ـ الدراستين الصباحية والمسائية

الفصل الدراسي الاول

تدريسيي المادة

ا. د. هيفاء هادي حساني
ا. د. غصون علي عبد الحسن
ا.م.د. احمد سالم كاظم
ا.م.د. محمد عبد الرحمن
م.د. عذراء صالح فليح

Lecture 1

Molecular biology is the branch of biology that deals with the nature of biological phenomena at the molecular level through the study of the life molecules (DNA, RNA and proteins).

DNA as a Genetic Material rather than Proteins: Classic Experiments

I. Frederick Griffith: Bacterial Transformation

In 1928, British bacteriologist Frederick Griffith conducted a series of experiments using *Streptococcus pneumoniae* bacteria and mice. Griffith wasn't trying to identify the genetic material, but rather, trying to develop a vaccine against pneumonia. In his experiments, Griffith used two related strains of bacteria, known as R and S.

- **R bacteria** form rough "R" appearance colonies when grown in a petri dish. The R bacteria were non-virulent.
- **S bacteria** form rounded smooth "S" appearance colonies due to a polysaccharide coat produced by the bacteria. This coat protected the S bacteria from the mouse immune system, making them virulent. Mice injected with live S bacteria developed pneumonia and died (Figure 1).

As part of his experiments, Griffith tried injecting mice with heat-killed S bacteria. Unsurprisingly, the heat-killed S bacteria did not cause disease in mice.

The experiments took an unexpected turn, however, when harmless R bacteria were combined with harmless heat-killed S bacteria and injected into a mouse. Not only did the mouse develop pneumonia and die, but when Griffith took a blood sample from the dead mouse and cultured it, he found that it contained living S bacteria!

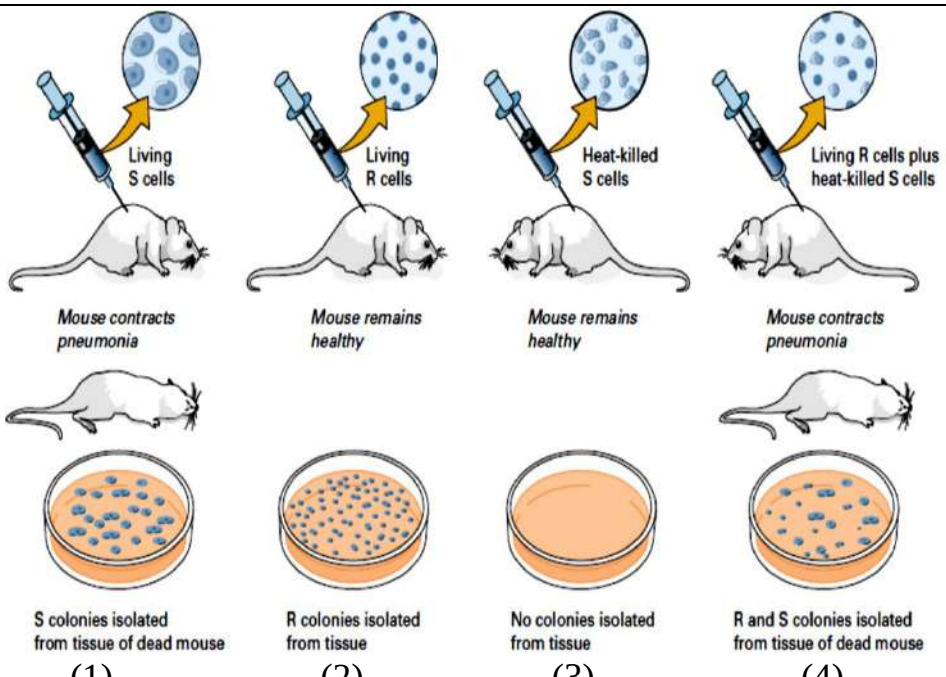


Figure 1. Diagram illustrating Frederick Griffith's experiment with S and R bacteria.

1. When live rough strain is injected into a mouse, the mouse lives.
2. When live smooth strain is injected into a mouse, the mouse gets pneumonia and dies.
3. When heat-killed smooth cells are injected into a mouse, the mouse lives.
4. When live rough strain & heat-killed smooth strain are injected into a mouse as a mixture, the mouse gets pneumonia and dies.

Griffith concluded that the R-strain bacteria must have taken up what he called a "transforming principle" from the heat-killed S bacteria, which allowed them to "transform" into smooth-coated bacteria and become virulent.

2. Avery, McCarty, and MacLeod: Identifying the transforming principle

In 1944, three Canadian and American researchers; Avery, McCarty, and MacLeod, set out to identify Griffith's "transforming principle".

To do so, they began with large cultures of heat-killed S cells and extracted the transforming principle by enzymatically destroying the other cellular components. By this method, they were able to obtain small amounts of highly purified transforming principle, which they could then analyse through other tests to determine its identity (Figure 2).

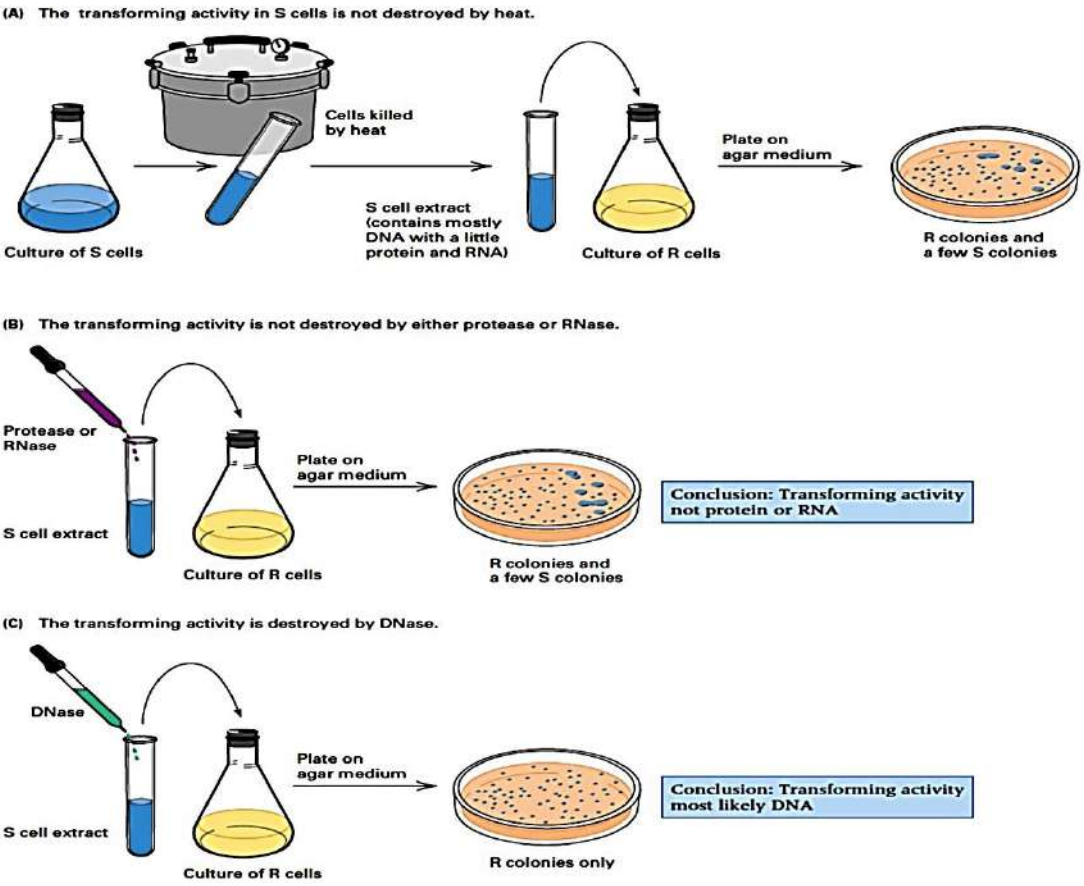


Figure 2. Diagram illustrating Avery, McCarty, and MacLeod

The evidence suggested to Avery and his colleagues that the transforming principle might be DNA:

1. The purified substance gave a negative result in chemical tests known to detect proteins, but a strongly positive result in a chemical test known to detect DNA.
2. The elemental composition of the purified transforming principle closely resembled DNA in its ratio of nitrogen and phosphorous.
3. Protein- and RNA-degrading enzymes had little effect on the transforming principle, but enzymes able to degrade DNA eliminated the transforming activity.

These results all pointed to DNA as the likely transforming principle. However, Avery was cautious in interpreting his results. He realized that it was still possible that some contaminating substance present in small amounts, not DNA, was the actual transforming principle.

Because of this possibility, debate over DNA's role continued until 1952, when Alfred Hershey and Martha Chase used a different approach to conclusively identify DNA as the genetic material.

3. Hershey-Chase experiments

In their experiments, Hershey and Chase studied **bacteriophages**, viruses that attack bacteria and composed of outer structures made of proteins and an inner core consisting of DNA.

Hershey and Chase knew that the bacteriophages attached to the surface of the host bacterial cell and injected some substance (either DNA or protein) into the host. To establish whether the phage injected DNA or protein into host bacteria, Hershey and Chase prepared two different batches of phage. In each batch, the bacteriophages were produced in the presence of a specific radioactive element:

- One sample was produced in the presence of, S^{35}, a radioactive isotope of sulphur. Sulphur is found in many proteins and is absent from DNA, so only phage proteins were radioactively labelled by this treatment.
- The other sample was produced in the presence of P^{32}, a radioactive isotope of phosphorous. Phosphorous is found in DNA and not in proteins, so only phage DNA (and not phage proteins) were radioactively labelled by this treatment.

Each batch of bacteriophages was used to infect a different culture of bacteria. After infection, each culture was blended to remove any remaining phage from the outside of the bacterial cells. Finally, the cultures were centrifuged to separate the bacteria from the phage debris (Figure 3).

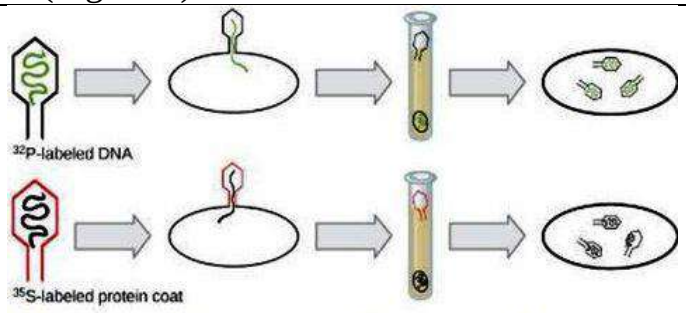


Figure 3. Diagram illustrating Hershey-Chase experiments

1. One batch of phage was labelled with S^3, which is incorporated into the protein coat. Another batch was labelled with P^{32}, which is incorporated into the DNA.
2. Bacteria were infected with the phage.

3. The cultures were blended and centrifuged to separate the phage from the bacteria.
4. Radioactivity was measured in the pellet and liquid (supernatant) for each experiment. P^{32} was found in the pellet (inside the bacteria), while S^{35} was found in the supernatant (outside of the bacteria).
   When Hershey and Chase measured radioactivity in the pellet and supernatant from both of their experiments, they found that a large amount of P^{32} appeared in the pellet, whereas almost all of the S^{35} appeared in the supernatant. Based on this and similar experiments, Hershey and Chase concluded that DNA, not protein, was injected into host cells and made up the genetic material of the phage.

The Molecular Composition of DNA

A DNA molecule is composed of smaller structures (DNA monomers) that are linked together in non-random sequences. The DNA monomers, which are called nucleotides, compose a DNA polymer forming the DNA chains or strands.
The nucleotides that compose DNA are called **deoxyribonucleotides**. The three components of a deoxyribonucleotide are a five-carbon sugar called **deoxyribose**, a phosphate group, and a **nitrogenous base**, a nitrogen-containing ring structure that is responsible for **complementary base pairing** between nucleic acid strands (Figure 4). The carbon atoms of the five-carbon deoxyribose are numbered 1′, 2′, 3′, 4′, and 5′ (1′ is read as “one prime” and so on). A **nucleoside** comprises the five-carbon sugar and nitrogenous base.

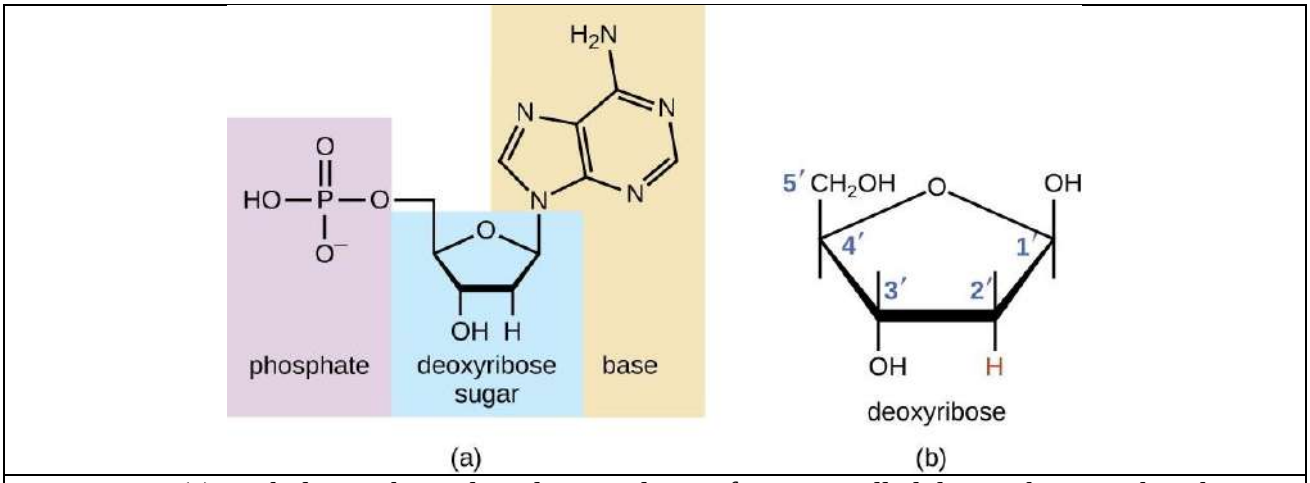


Figure 4. (a) Each deoxyribonucleotide is made up of a sugar called deoxyribose, a phosphate group, and a nitrogenous base in this case, adenine. (b) The five carbons within deoxyribose are designated as 1′, 2′, 3′, 4′, and 5′.

The deoxyribonucleotide is named according to the **nitrogenous bases** (Figure 5). The nitrogenous bases **adenine** (A) and **guanine** (G) are the **purines**; they have a double-ring structure with a six-carbon ring fused to a five-carbon ring. The **pyrimidines**, **cytosine** (C) and **thymine** (T), are smaller nitrogenous bases that have only a six-carbon ring structure.

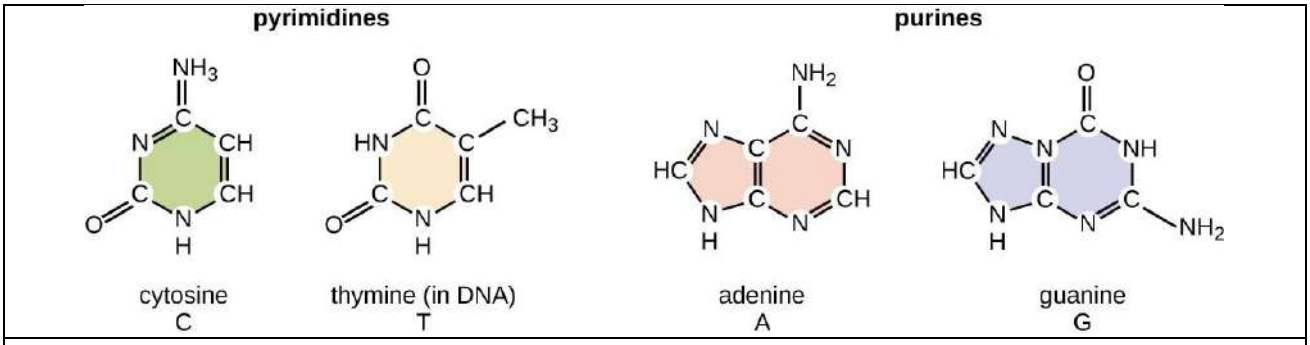


Figure 5. Nitrogenous bases within DNA are categorized into the two-ringed purines adenine and guanine and the single-ringed pyrimidines cytosine and thymine. Thymine is unique to DNA.

Individual nucleoside triphosphates combine with each other by covalent bonds known as 5′-3′ **phosphodiester bonds**, or linkages whereby the phosphate group attached to the 5′ carbon of the sugar of one nucleotide bonds to the hydroxyl group of the 3′ carbon of the sugar of the next nucleotide. Phosphodiester bonding between nucleotides forms the **sugar-phosphate backbone**, the alternating sugar-phosphate structure composing the framework of a nucleic acid strand (Figure 6).

During the polymerization process, deoxynucleotide triphosphates (dNTP) are used. To construct the sugar-phosphate backbone, the two terminal phosphates are released from the dNTP as a pyrophosphate. The resulting strand of nucleic acid has a free phosphate group at the 5′ carbon end and a free hydroxyl group at the 3′ carbon end. The two unused phosphate groups from the nucleotide triphosphate are released as pyrophosphate during phosphodiester bond formation. Pyrophosphate is subsequently hydrolysed, releasing the energy used to drive nucleotide polymerization.

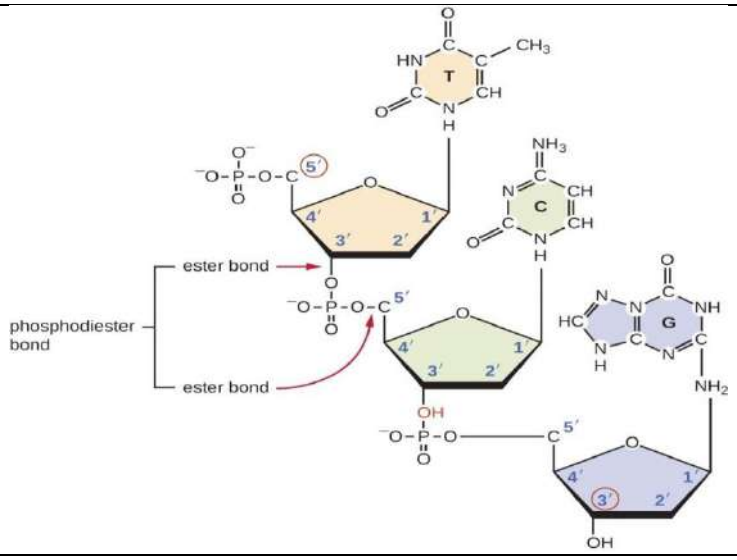


Figure 6. Phosphodiester bonds form between the phosphate group attached to the 5′ carbon of one nucleotide and the hydroxyl group of the 3′ carbon in the next nucleotide, bringing about polymerization of nucleotides into nucleic acid strands. Note the 5′ and 3′ ends of this nucleic acid strand.

Discovering the Double Helix

By the early 1950s, considerable evidence had accumulated indicating that DNA was the genetic material of cells, and now the race was on to discover its three-dimensional structure. Around this time, Austrian biochemist Erwin **Chargaff** (1905–2002) examined the content of **DNA** in different species and discovered that adenine, thymine, guanine, and cytosine were not found in equal quantities, and that it varied from species to species, but not between individuals of the same species. He found that the amount of adenine was very close to equalling the amount of thymine, and the amount of cytosine was very close to equalling the amount of guanine, or A = T and G = C. These relationships are also known as **Chargaff's rules**.

British researchers Rosalind **Franklin** (1920–1958) and her graduate student R.G. **Gosling** were also using X-ray diffraction to understand the structure of DNA (Figure 7). They were clearly show the overall double-helix structure of DNA.

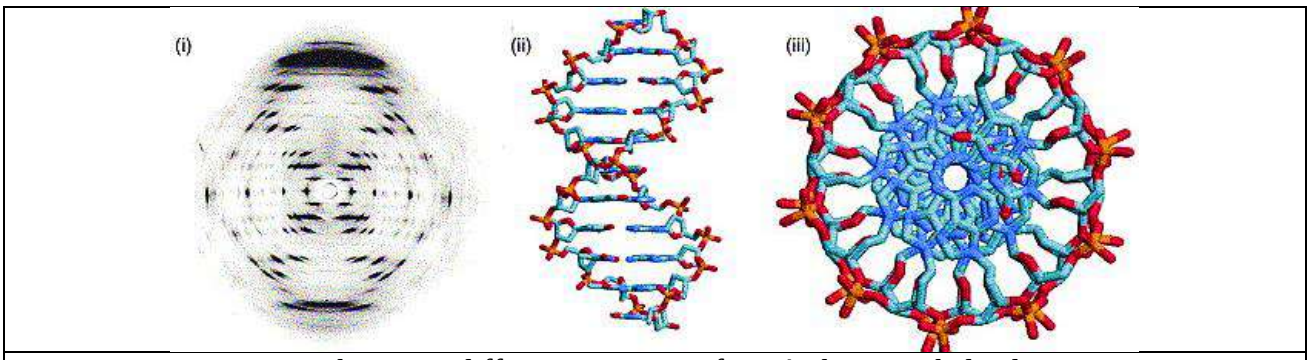


Figure 7. The X-ray diffraction pattern of DNA shows its helical nature.

James Watson, an American scientist, and Francis Crick, a British scientist, were working together in the 1950s to discover DNA's structure. They used **Chargaff's rules** and **Franklin** and **Wilkins' X-ray diffraction** images of DNA fibres to piece together the purine-pyrimidine pairing of the double helical DNA molecule. In April 1953, **Watson and Crick** published their model of the DNA **double helix** in *Nature*.

Lecture 2

DNA Molecular Structure

Watson and Crick proposed that **DNA** is made up of two strands that are twisted around each other to form a right-handed helix. The two DNA strands are **antiparallel**, such that the 3′ end of one strand faces the 5′ end of the other (Figure 8). The 3′ end of each strand has a free hydroxyl group, while the 5′ end of each strand has a free phosphate group. The sugar and phosphate of the polymerized nucleotides form the backbone of the structure, whereas the nitrogenous bases are stacked inside. These nitrogenous bases on the interior of the molecule interact with each other, base pairing.

Analysis of the diffraction patterns of DNA has determined that there are approximately 10 bases per turn in DNA. The asymmetrical spacing of the sugar-phosphate backbones generates major grooves (where the backbone is far apart) and minor grooves (where the backbone is close together) (Figure 8). These grooves are locations where proteins can bind to DNA. The binding of these proteins can alter the structure of DNA, regulate **replication**, or regulate **transcription** of DNA into RNA.

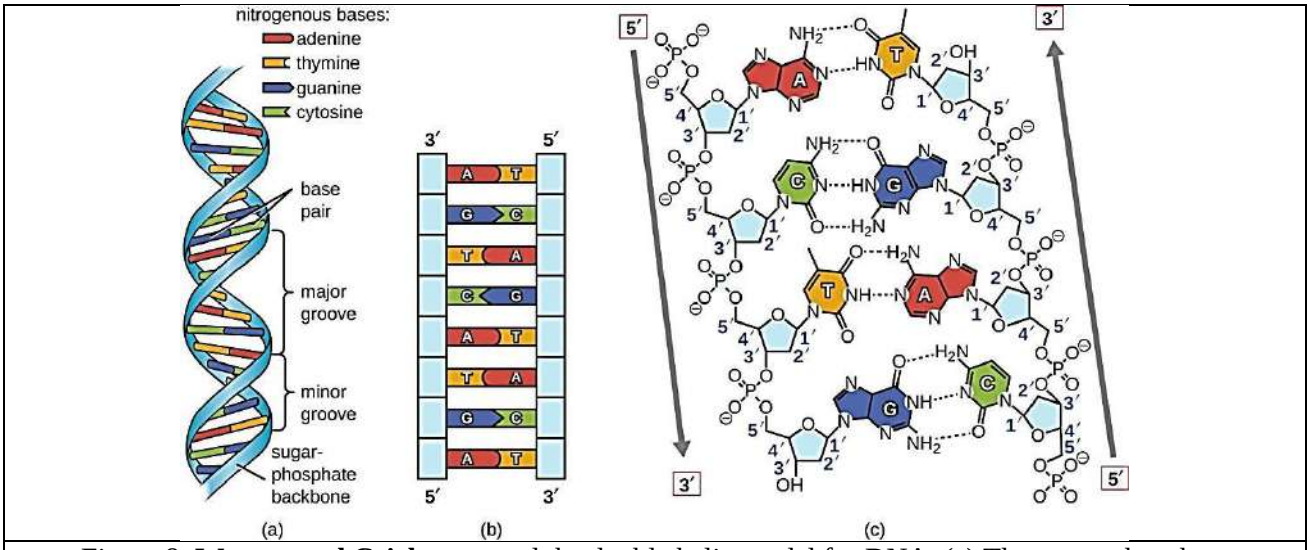


Figure 8. **Watson and Crick** proposed the double helix model for DNA. (a) The sugar-phosphate backbones are on the outside of the double helix and purines and pyrimidines form the "rungs" of the DNA helix ladder. (b) The two DNA strands are antiparallel to each other. (c) The direction of each strand is identified by numbering the carbons (1 through 5) in each sugar molecule. The 5′ end is the one where carbon #5 is not bound to another nucleotide; the 3′ end is the one where carbon #3 is not bound to another nucleotide.

Base pairing takes place between a purine and pyrimidine. In DNA, adenine (A) and thymine (T) are **complementary base pairs**, and cytosine (C) and guanine (G) are also complementary base pairs, explaining **Chargaff's rules** (Figure 9). The base pairs are stabilized by hydrogen bonds; adenine and thymine form two hydrogen bonds between them, whereas cytosine and guanine form three hydrogen bonds between them.

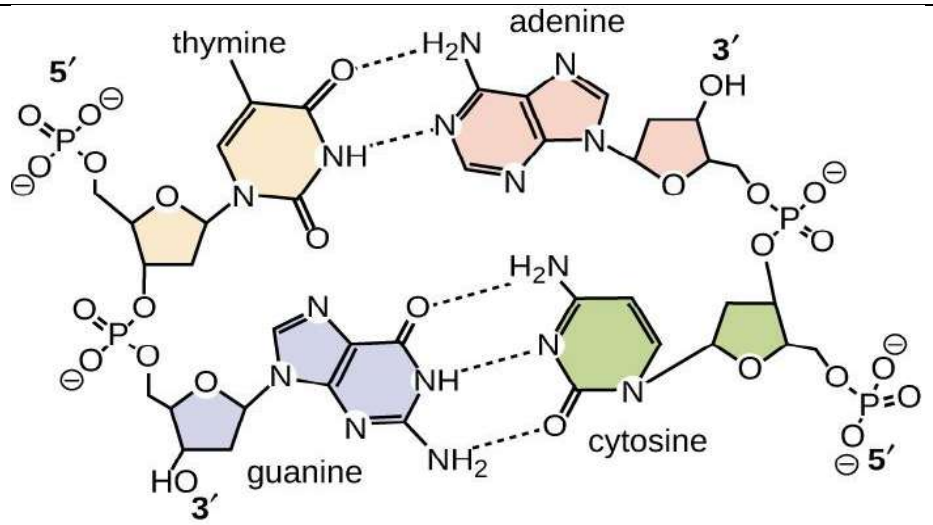


Figure 9. Hydrogen bonds form between complementary nitrogenous bases on the interior of DNA

DNA replication

Initially, three alternative models were proposed for DNA replication (Figure 1).

1- Conservative replication; the entire double-stranded DNA molecule serves as a template for a whole new molecule of DNA, and the original DNA molecule is fully conserved during replication.

2- Dispersive replication; both nucleotide strands break down (disperse) into fragments, which serve as templates for the synthesis of new DNA fragments, and then somehow reassemble into two complete DNA molecules. In this model, none of the original molecule is conserved.

3- Semi-conservative replication: is intermediate between these two models; the two nucleotide strands unwind, and each serve as a template for a new DNA molecule.

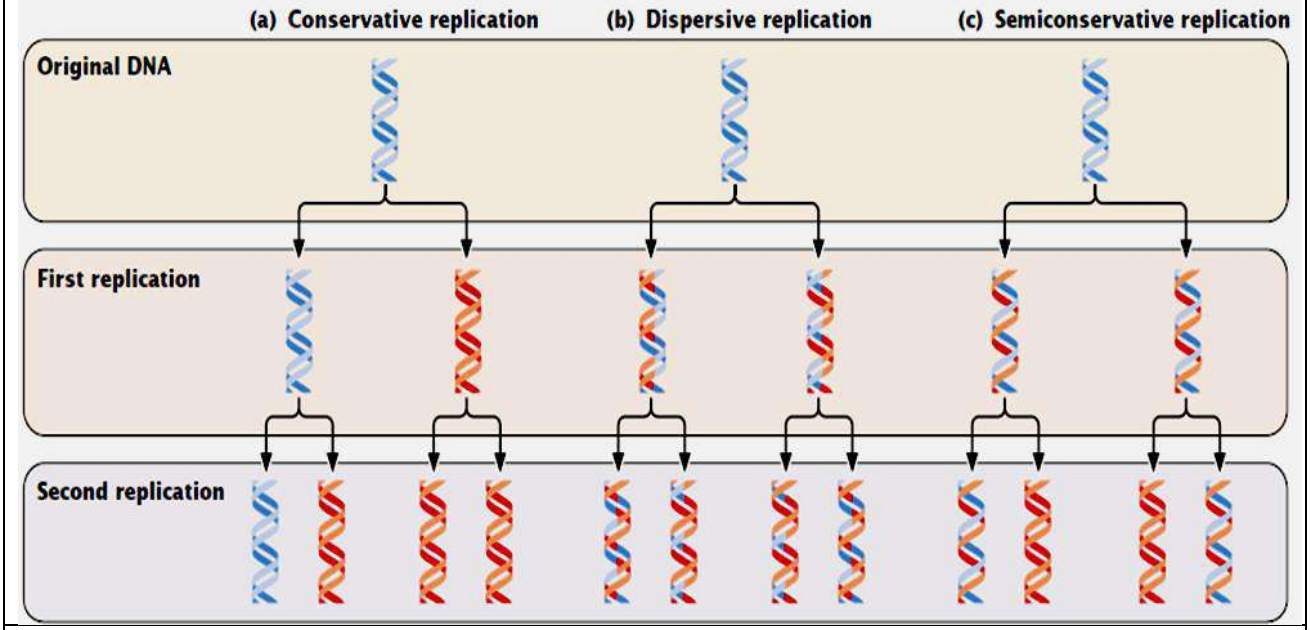


Figure 1. Diagram showing three proposed models for DNA replication

Meselson and Stahl experiment

Meselson and Stahl were interested in understanding how DNA replicates. They grew *E. coli* for several generations in a medium containing a "heavy" isotope of nitrogen (^{15}N) that incorporates into nitrogenous bases, and eventually into the DNA. The *E. coli* culture was then shifted into medium containing ^{14}N and allowed to grow for one generation. The cells were harvested, and the DNA was isolated. The DNA was centrifuged at high speeds in an ultracentrifuge. Some cells were allowed to grow for one more life cycle in ^{14}N and spun again. During the density gradient centrifugation, the DNA is loaded into Cesium Chloride gradient and spun at high speeds of 50,000 rpm. Under these circumstances, the DNA will form a band according to its density in the gradient. DNA grown in ^{15}N will band at a higher density position than that grown in ^{14}N (Figure 2).

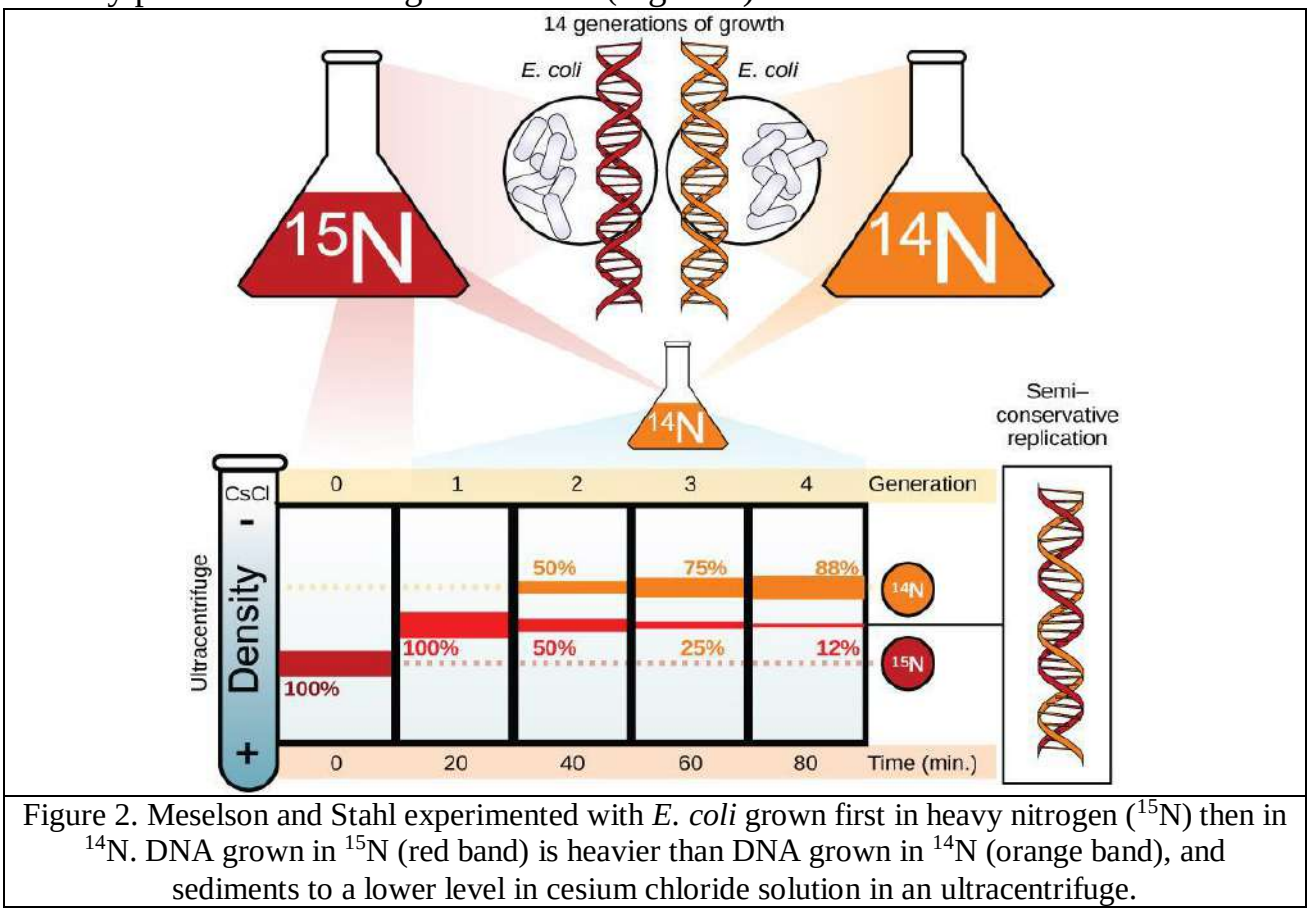


Figure 2. Meselson and Stahl experimented with *E. coli* grown first in heavy nitrogen (^{15}N) then in ^{14}N. DNA grown in ^{15}N (red band) is heavier than DNA grown in ^{14}N (orange band), and sediments to a lower level in cesium chloride solution in an ultracentrifuge.

Meselson and Stahl noted that after one generation of growth in ^{14}N after they had been shifted from ^{15}N, the single band observed was intermediate in position in between DNA of cells grown exclusively in ^{15}N and ^{14}N. This suggested either a semi-conservative or dispersive mode of replication. The DNA harvested from cells grown for two generations in ^{14}N formed two bands: one DNA band was at the intermediate position between ^{15}N and ^{14}N, and the other corresponded to the band of ^{14}N DNA. These results could only be explained if DNA replicates in a semi-conservative manner. Therefore, the other two models (dispersive & conservative) were ruled out.

DNA Replication in Prokaryotes

The replication occurs in three main stages: initiation, elongation, and termination. DNA replication employs a large number of proteins and enzymes, each of which plays a critical role during the process. One of the key players is **DNA polymerase**, which adds nucleotides one by one to the growing DNA chain that are complementary to the template strand. The addition of nucleotides requires energy; this energy is obtained from the nucleotides that have three phosphates attached to them; (dATP, dTTP, dCTP & dGTP). When the bond between the phosphates is broken, the energy released is used to form the phosphodiester bond between the incoming nucleotide and the growing chain, and then pyrophosphate group is released. In prokaryotes, three main types of polymerases are known: DNA pol I, DNA pol II, and DNA pol III. It is now known that DNA pol III is the enzyme required for DNA synthesis, DNA pol I has exonuclease activity, removes RNA primer and replaces with newly synthesized DNA, and DNA pol II is required for DNA repair.

There are specific nucleotide sequences called **origins of replication** where replication begins. In *E. coli*, which has a single origin of replication (called oriC) on its one chromosome (as do most prokaryotes); it is approximately 245 base pairs long and is rich in AT sequences. The origin of replication is recognized by certain proteins that bind to this site. An enzyme called **helicase** unwinds the DNA by breaking the hydrogen bonds between the nitrogenous base pairs. ATP hydrolysis is required for this process. As the DNA opens up, Y-shaped structures called replication forks are formed. Two replication forks are formed at the origin of replication and these get extended bi- directionally as replication proceeds. **Single-strand binding (SSB) proteins** coat the single strands of DNA near the replication fork to prevent the single-stranded DNA from rewinding back into a double helix. DNA polymerase is able to add nucleotides **only in the 5' to 3' direction** (a new DNA strand can be only extended in this direction). It also requires a **free 3'-OH group** to which it can add nucleotides by forming a phosphodiester bond between the 3'-OH end and the 5' phosphate of the next nucleotide. This essentially means that it cannot add nucleotides if a free 3'-OH group is not available. Then how does it add the first nucleotide? The problem is solved with the help of a primer that provides the free 3'-OH end. Another enzyme, **RNA primase**, synthesizes an RNA primer that is about five to ten nucleotides long and complementary to the DNA. Because this sequence primes the DNA synthesis, it is appropriately called the primer. DNA polymerase can now extend this RNA primer, adding nucleotides one by one that are complementary to the template strand.

E. coli needs 42 minutes to replicate the whole genomic DNA. The speed of the replication fork progression is about 1,000 bp per second. DNA polymerase can only extend in the 5' to 3' direction, which poses a slight problem at the replication fork. As we know, the DNA double helix is anti-parallel; that is, one strand is in the 5' to 3' direction and the other is oriented in the 3' to 5' direction. One strand, which is complementary to the 3' to 5' parental DNA strand, is synthesized continuously towards the replication fork because the polymerase can add nucleotides in this direction. This continuously synthesized strand is known as the **leading strand**. The

other strand, complementary to the 5' to 3' parental DNA is extended away from the replication fork, in small fragments known as **Okazaki fragments**, each requiring a primer to start the synthesis. The strand with the Okazaki fragments is known as the **lagging strand**.

The leading strand can be extended by one primer alone, whereas the lagging strand needs a new primer for each of the short Okazaki fragments. The overall direction of the lagging strand will be 3' to 5', and that of the leading strand 5' to 3'. A protein called the sliding clamp holds the DNA polymerase in place as it continues to add nucleotides. The sliding clamp is a ring-shaped protein that binds to the DNA and holds the polymerase in place. **Topoisomerase** prevents the over-winding of the DNA double helix ahead of the replication fork as the DNA is opening up; it does so by causing temporary nicks in the DNA helix and then resealing it. As synthesis proceeds, the RNA primers are replaced by DNA. The primers are removed by the exonuclease activity of DNA pol I, and the gaps are filled in by deoxyribonucleotides. The nicks that remain between the newly synthesized DNA (that replaced the RNA primer) and the previously synthesized DNA are sealed by the enzyme **DNA ligase** that catalyses the formation of phosphodiester linkage between the 3'-OH end of one nucleotide and the 5' phosphate end of the other fragment (Figure 3). Once the chromosome has been completely replicated, the two DNA copies move into two different cells during cell division.

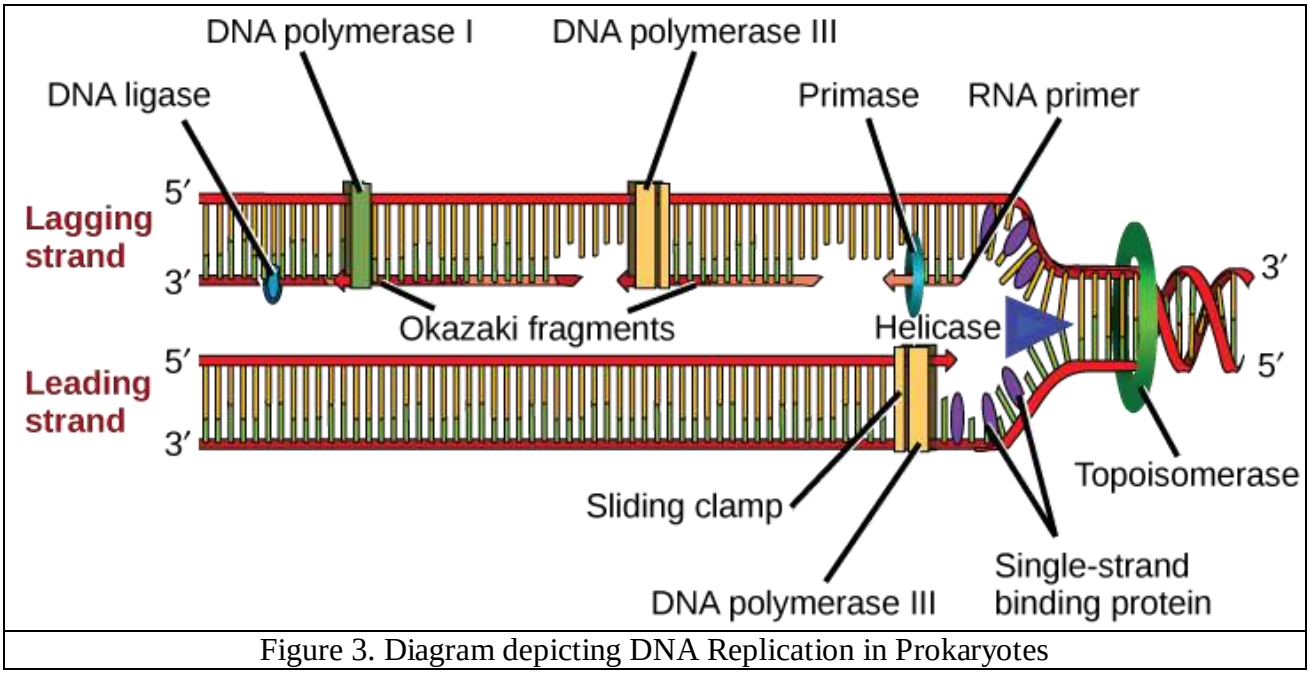


Figure 3. Diagram depicting DNA Replication in Prokaryotes

Lecture 3

DNA Replication in Eukaryotes

Because eukaryotic chromosomes are more complex than that of prokaryotes, DNA replication is a very complicated process, but the principles are quite the same.

Chromosomes Structure: Chromatin

Chromatin is a fibrous substance, which at the time of cell division becomes visible in the form of chromosomes. It is composed of DNA, RNA, histones, and other proteins. Each human cell has 1.8 meters of DNA. The role of chromatin is to package DNA into a smaller volume to fit in the nucleus and to control gene expression and DNA replication. Processes like DNA repair, replication, recombination, and transcription take place in the chromatin.

Histones and Nucleosomes

The most abundant proteins in chromatin are the **histones,** which are five major types: **H1**, **H2A**, **H2B**, **H3**, and **H4**. All histones have a high percentage of positively charged amino acids, arginine and lysine, that give them a net positive charge. The positive charges attract the negative charges on the phosphates of DNA. If lysine is acetylated, its charge will be neutralized. This leads to weakening of histones binding with DNA and plays a crucial role in the regulation of DNA replication and transcription. The interaction between DNA - histones forms the basic structural element of chromatin "**Nucleosome**", which consists of 146 bp DNA sequence that coils about two times around **histone octamer** (eight histone proteins; two copies of each of H2A, H2B, H3, and H4) in a left-handed direction. Nucleosome cores are separated by linker DNA of variable length and are associated with the linker histone H1. The next level of chromatin organization is the 30-nm chromatin fibre, which is composed of packed nucleosome arrays "**Solenoid**".

Chromatin Structure Changes

Chromatin structure is changed dynamically. It is less condensed, wide (**Euchromatin**) in regions where transcription takes place, while in transcriptionally inactive regions it is more condensed (**Heterochromatin**). In condensed chromatin each nucleosome binds with the next one by linker histone H1. This leads to the creation of **solenoid** (Figure 4). In transcriptionally active chromatin, histone H1 is released from DNA and nucleosomes can change position, which means they move along the DNA strand. Solenoid forms furthermore complex structures ensuring multi-level organization of chromatin in the nucleus. Important levels of organization of chromatin are so called domains or loops, which are formed by interactions between nucleosome fibres and protein structures in nucleus (nuclear skeleton).

The highest level of compaction of chromatin is chromosomes condensing during cell division. The compacted DNA molecule is 40,000 times shorter than an unpacked.

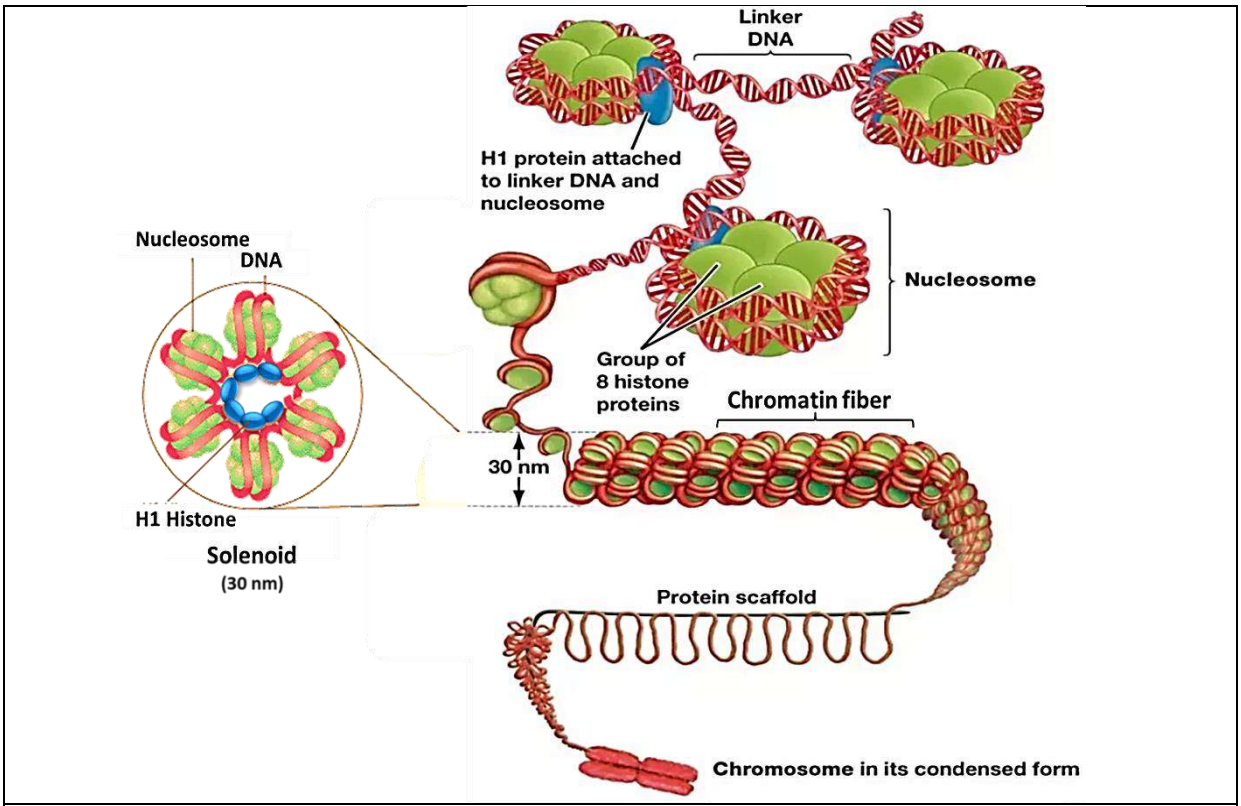


Figure 4. Stages of chromatin organization; Histones, Nucleosomes, Solenoid, Chromatin fibre and Chromosome

Eukaryotic DNA Replication

DNA is replicated during the Synthesis (S) phase of the cell cycle. DNA replication is triggered by a **Transcription factor** (in yeast called MCB binding factor, in mammals is E2F) that regulates the expression of enzymes necessary for replication: DNA polymerases, DNA primases and cyclins.

The replication occurs in three main stages: initiation, elongation, and termination. **Initiation**: Before replication can start, the DNA has to be made available as a template. The chromatin may undergo some chemical modifications such as histone acetylation, so the chromatin will be temporarily disassembled to allow the DNA replication machinery access to the DNA template.

At the origin of replication, a pre-replication complex is made with other initiator proteins. A sliding clamp protein known as PCNA (Proliferating Cell Nuclear Antigen) holds the DNA pol in place so that it does not slide off the DNA. Other proteins involved in replication are **helicases** and single-strand DNA binding (**SSB**) proteins. **Helicase** unwinds double helix into single strands allowing each strand to be copied by breaking H-bonds. SSB proteins bind to single-stranded DNA, preventing from the duplex restoration (Figure 5).

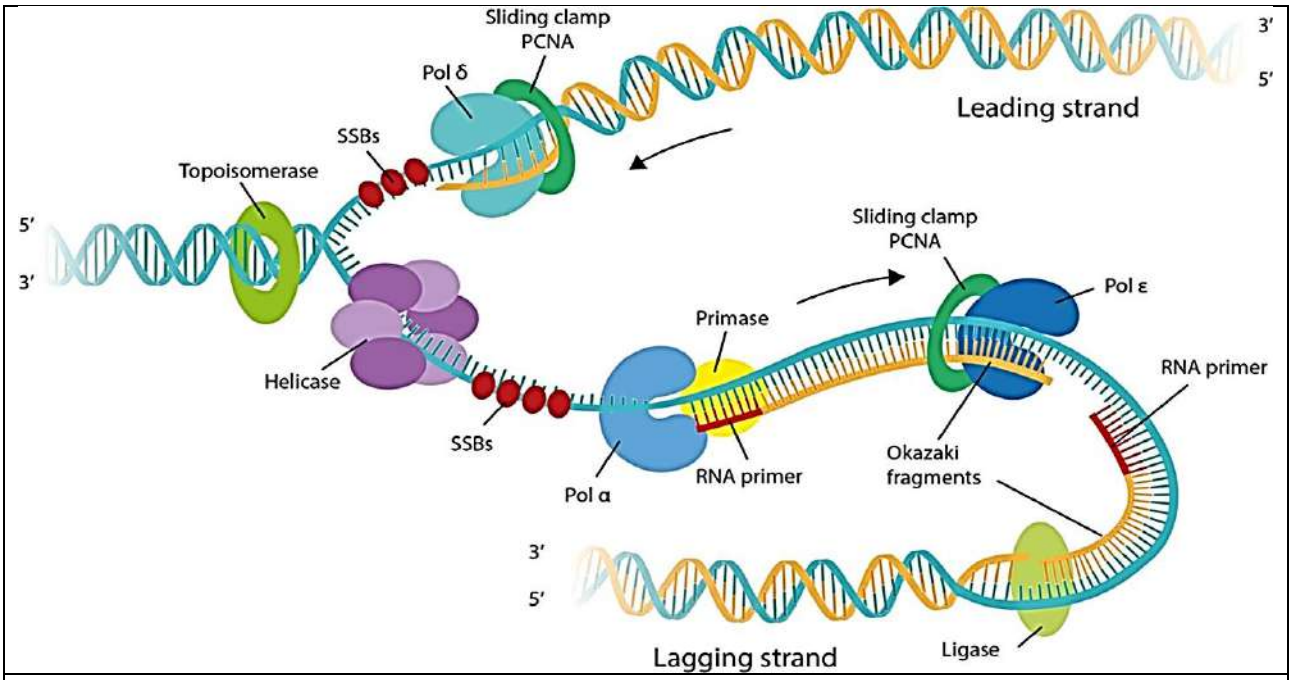


Figure 5. DNA Replication in Eukaryotes

Elongation: As in prokaryotes, DNA is synthesized only in the 5` to 3` direction. Because two chains of DNA are anti-parallel, only one strand (leading strand) can be synthesized by DNA polymerase continuously. The second strand (lagging strand) is synthesized in fragments.

In mammals there are five main DNA polymerases (α, β, γ, δ and ε) and several smaller. **γ polymerase** replicates mitochondrial DNA, the others are present in the nucleus, where they perform different functions:
– **DNA pol α**: initiates DNA synthesis on both the leading and lagging strands.
– **DNA pol β and ε**: DNA repair.
– **DNA pol δ**: synthesis of leading and lagging strands.
To start the synthesis of DNA, a short RNA molecule (primer) is required. The primer is about 5-12 nucleotides in length and is complementary to one strand of the template. It is synthesized by a specific enzyme "**Primase**", which is the RNA polymerase. The use of ribonucleotides for the synthesis of primer means that this fragment is temporary; it is removed in the final phase of replication. RNA polymerases, unlike DNA polymerases, can synthesize new strand without a primer, because they do not check whether the nucleotide inserted into the chain is correct (they are much less accurate) (Figure 6).

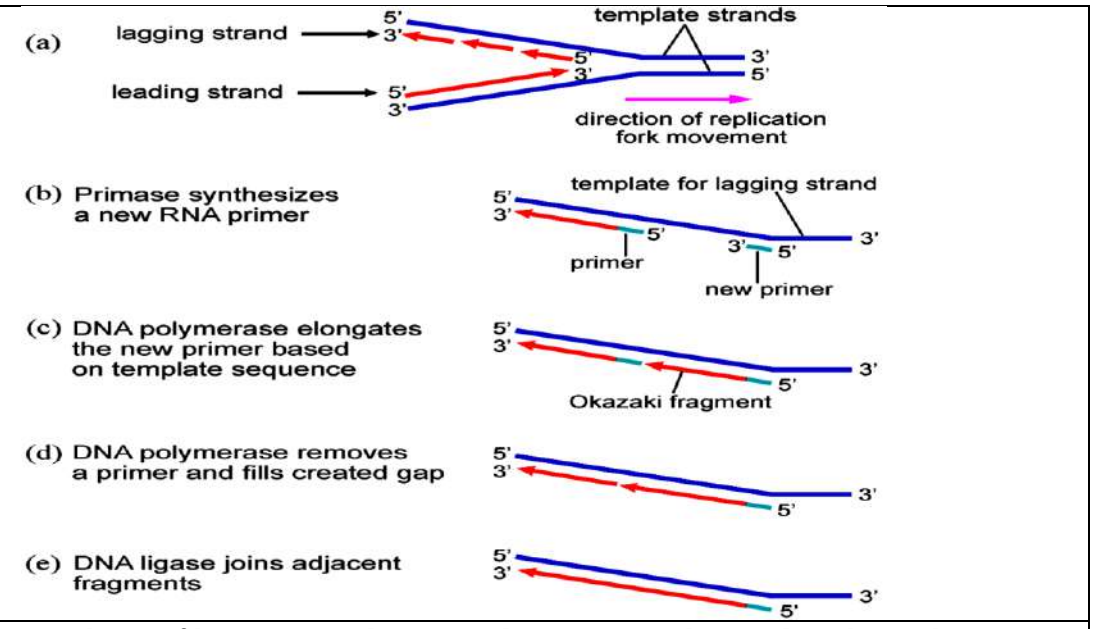


Figure 6. Synthesis of lagging strand: DNA polymerase attaches nucleotides to the new primer in the direction 5' to 3' until it encounters the 5' end of the previous primer. DNA ligase joins adjacent Okazaki fragments. The entire lagging strand is synthesized by repeating steps.

Termination: Unlike bacterial chromosomes, eukaryotic chromosomal DNA has many replication origins on each chromosome, which are also rich in a sequences of A:T pairs. Unwinding of DNA at these origins also leads to the formation of a replication fork for each. Therefore, multiple origins of replication are formed on the eukaryotic chromosome; humans can have up to 100,000 origins of replication operating simultaneously on chromosomal DNA. The rate of replication is approximately 100 nucleotides per second, much slower than prokaryotic replication due to the association of DNA with histones. Replication of the entire human genome takes approximately 8 hours, but Drosophila genome takes 3-4 minutes only.

Topoisomerases

During DNA replication, DNA ahead of the replication bubble becomes positively supercoiled, while DNA behind the replication fork becomes entangled forming pre-catenanes. Enzymes called topoisomerases play an essential role in resolving these topological problems (Figure 1).

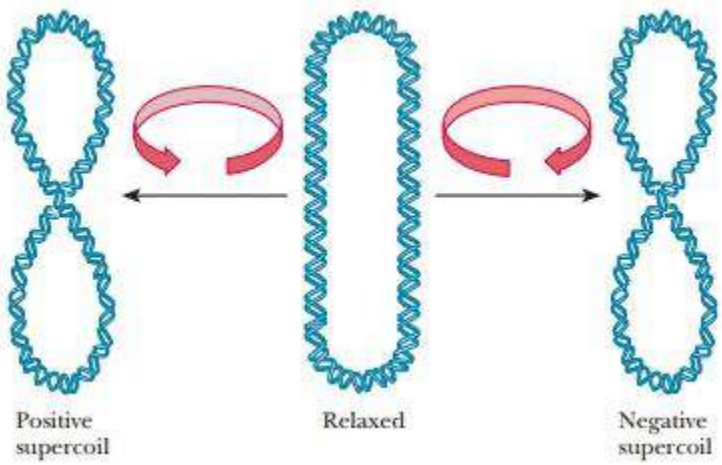


Figure 1. positive supercoil DNA (left-handed, counter clockwise); negative supercoil DNA (right-handed, clockwise).

Topoisomerases regulate DNA topology by making temporary single- strand DNA breaks (Type I DNA topoisomerases) or double-strand DNA breaks (Type II DNA topoisomerases). The DNA double helix needs to be unwound for processes such as DNA replication or transcription to take place, giving rise to the accumulation of positive supercoils ahead of transcription bubbles and replication forks, and negative supercoils pre-catenanes behind. Topoisomerases relax the supercoiled DNA and decatenate it to allow DNA replication and transcription (Figure 2).

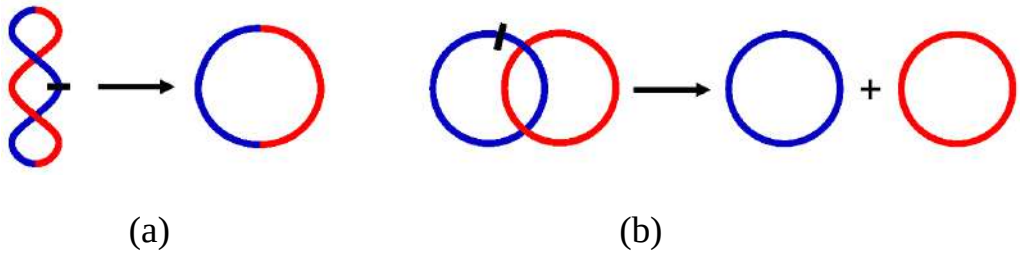


Figure 1. Topo II functions: (a) removing of supercoils; (b) separation of linked molecules of catenanes. These actions require cutting of both strands of DNA (marked with black line). After relocation of the DNA strands they are again connected.

Type II DNA topoisomerases are divided into sub-types, IIA and IIB, based on structural and evolutionary considerations. Type IIA is found in bacteria and

eukaryotes, whereas IIB was discovered in archaea and recently in plants and plasmodial parasites. Most bacteria have two type IIA topoisomerases, DNA gyrase and topoisomerase IV.

Topoisomerase inhibitors are effective therapeutics used as anti-cancer chemotherapy (e.g., Etoposide) and as anti-bacterial agents (e.g., Fluoroquinolones). Without topoisomerases, DNA cannot replicate correctly.

Lecture 4

Telomerase and Cellular Senescence

Synthesis of lagging strand requires synthesis of multiple primers that are finally removed. In bacteria, which have a circular DNA molecule, there is no problem with the replication of the whole molecule. In case of linear DNA molecule, lagging strand is always shorter than the template (at least by the length of the primer). Such situation takes place at the end of chromosomes.

At the ends of eukaryotic chromosomes, **telomere regions** are found. **Telomeres** minimize the problem of shortening of the DNA strands during every cell division (the second function of telomeres is to protect chromosomes from fusing with each other). Telomeres contain hundreds of **tandem repeated sequences**, thus their shortening after each cell division is not detrimental to the cell. Cells can be divided a certain number of times before the DNA loss prevents further division (this is known as the **Hayflick limit**). However, after reaching a certain telomere length, the cell may stop to divide and die. This is a normal process in **somatic cells**. Steady shortening of telomeres with each replication in somatic cells may have a role in **senescence** and in the **prevention of cancer**.

Within the **germ cell line**, which passes DNA to the next generation, the repetitive sequences of the telomere region are extended by **telomerase**. Telomerase is **ribonucleoprotein enzyme** that catalyses the elongation of 3` end of DNA. This enzyme maintains the length of the telomere. The telomere length specifies the number of divisions a cell can undergo before it finally dies. For example, telomerase is activated in **embryonic cell lines** and the telomere length is maintained at a constant level; therefore, these cells have an unlimited fission potential. **Stem cells** are characterized by a lower telomerase activity, which enables only partial compensation for the shortening of telomeres. Somatic cells are usually characterized by the absence of telomerase activity. Telomere shortening leads to the attainment of the **Hayflick limit**, the transition of cells to a state of **senescence**. Telomerase is a **reverse transcriptase**. It consists of two major components: **telomerase RNA** (TER) and **telomerase reverse transcriptase** (TERT). TER is a **non-coding RNA**, and it contains the region which serves as a template for telomere synthesis (Figure 3).

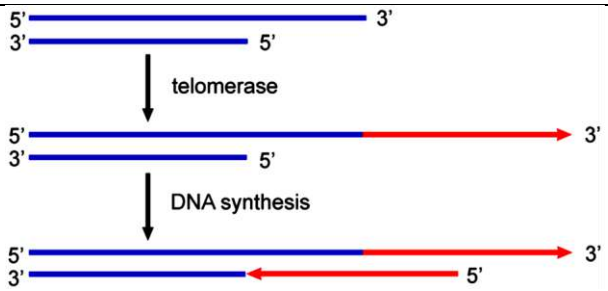


Figure 3. Scheme of telomere elongation by telomerase. Telomerase first extends the longer strand of DNA. Then, using the same mechanism as synthesizing the lagging strand, the shorter strand is extended.

That is quite unique; any other of the known polymerases does not contain polynucleotide template (Figure 4). Telomerase can become mistakenly active in somatic cells, sometimes leading to cancer formation.

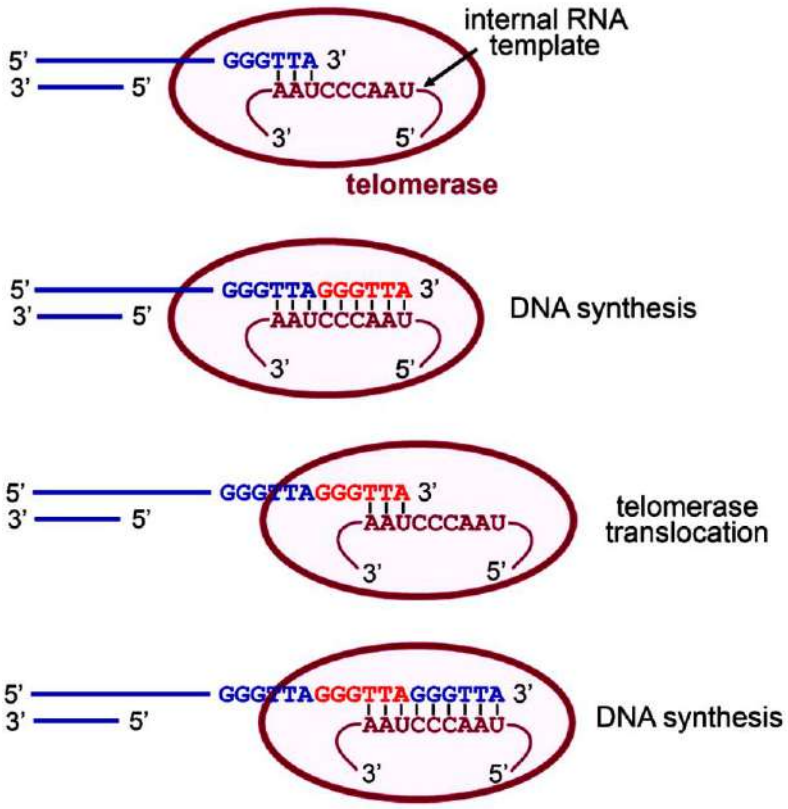


Figure 4. illustrating the mechanism of telomere extension by telomerase.

Mutation

Mutation is a permanent change in the DNA sequence. Effects of the mutation could be observed only when it occurs in coding regions of the genes. Changes in the non-coding regions of DNA generally do not affect the gene function.

Mutations have a wide range of effects. Some mutations are not expressed; these are known as **silent mutations**.

Point mutations are those mutations that affect a single base pair. The most common nucleotide mutations are **substitutions**, in which one base is replaced by another. These can be of two types, either **transitions** or **transversions**. Transition substitution refers to a purine or pyrimidine being replaced by a base of the same kind; for example, a purine such as adenine may be replaced by the purine guanine. Transversion substitution refers to a purine being replaced by a pyrimidine, or vice versa; for example, cytosine, a pyrimidine, is replaced by adenine, a purine. Mutations can also be the result of the addition of a base, known as an **insertion**, or the removal of a base, also known as **deletion**. Sometimes a piece of DNA from one chromosome may get translocated to another chromosome or to another region of the same chromosome; this is also known as **translocation**. A **frameshift mutation** is a genetic mutation caused by insertions or deletions of a number of nucleotides in a DNA sequence. Due to the triplet nature of gene expression by codons, the insertion or deletion can change the reading frame, resulting in a completely different translation from the original .These mutation types are shown in Figure 5.

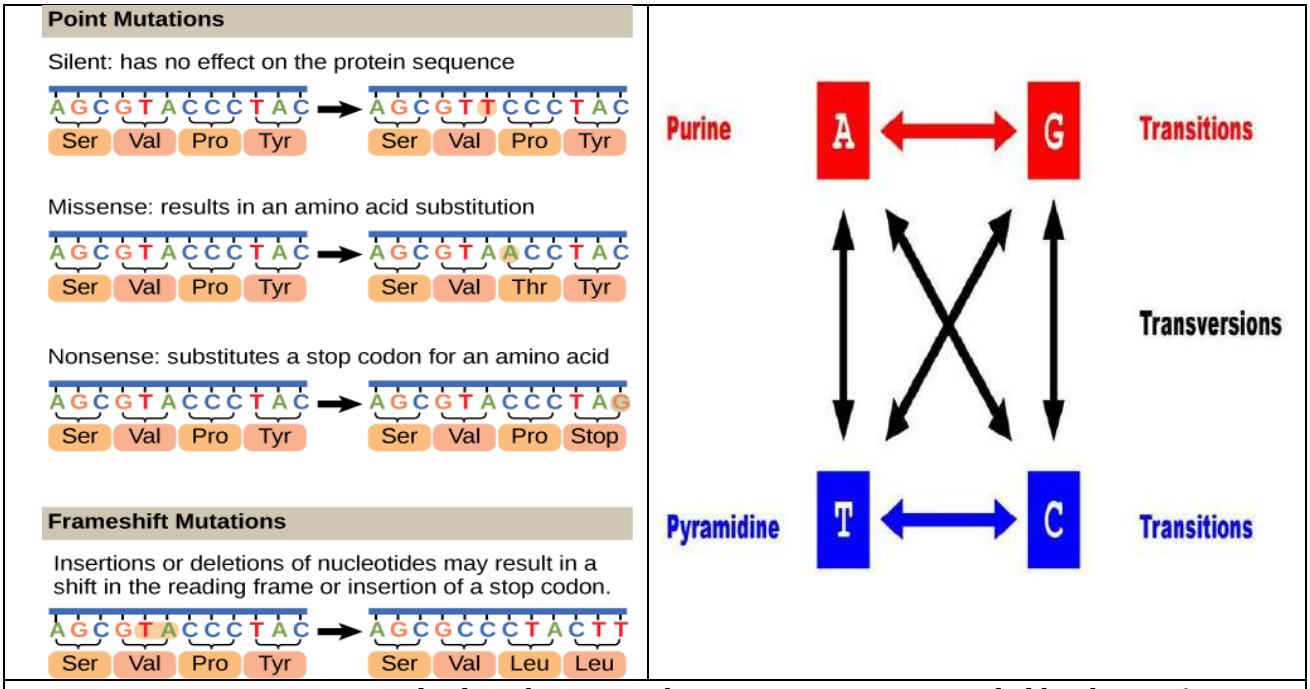


Figure 5. Mutations can lead to changes in the protein sequence encoded by the DNA.

A frameshift mutation that results in the insertion of three nucleotides is often less deleterious than a mutation that results in the insertion of one nucleotide.

Missense mutation refers to a change in one amino acid in a protein, arising from a point mutation in a single nucleotide. **Nonsense mutation** in which a codon is changed to a premature stop codon that results in truncation of the resulting protein.

Mutations in repair genes have been known to cause cancer. Many **mutated repair genes** have been implicated in certain forms of **pancreatic cancer**, **colon cancer**,

and colorectal cancer. Mutations can affect either somatic cells or germ cells. If many mutations accumulate in a somatic cell, they may lead to problems such as the **uncontrolled cell division** observed in cancer. If a mutation takes place in germ cells, the mutation will be passed on to the next generation, as in the case of **hemophilia** and **xeroderma pigmentosa**.
The existence of mutations in the coding regions of DNA can lead to changes in the structure of an encoded protein or to a decrease or complete loss in its expression. Because a change in the DNA sequence affects all copies of the encoded protein, mutations can be particularly damaging to a cell or organism. In contrast, any alterations in the sequences of RNA or protein molecules that occur during their synthesis are less serious because many copies of each RNA and protein are synthesized. A fundamental genetic difference between organisms is whether their cells carry a single chromosome of each (referred to as **haploid**) or have two copies of each chromosome (referred to as **diploid**). Many simple **unicellular organisms** (e.g. Bacteria) are haploid, whereas **complex multicellular organisms** (e.g. flies, mice, humans) are diploid.

Different forms of a gene, whether it is normal or mutant, are referred to as **alleles**. Since diploid organisms carry two copies of each gene, they may carry identical alleles, which are **homozygous** for a gene, or carry different alleles, that are **heterozygous** for a gene.

A **recessive** mutation is one in which both alleles must be mutant in order for the mutant **phenotype** to be observed; that is, the individual must be homozygous for the mutant allele to show the mutant phenotype. In contrast, the phenotypic consequences of a **dominant** mutation are observed in a heterozygous individual carrying one mutant and one normal allele. Recessive mutations result in a loss of function, whereas dominant mutations often, but not always, result in a gain of function.

Mutations can be advantageous and lead to an evolutionary advantage, and can also be deleterious, causing disease (e.g. cancers), structural abnormalities, developmental delays, or other effects.

Bacterial DNA Mutations & Antibiotic resistance

Generally, antibiotics act as synthesis inhibitors of cell wall, protein, RNA and DNA. When an antibiotic loses the capacity to kill or control bacterial growth, antibiotic resistance occurs. This can occur through genetic mutations.

DNA Synthesis Inhibitors

In gram-negative bacteria, such as *Helicobacter pylori*, mutation resistance occurs relatively quickly to fluoroquinolones. Ciprofloxacin, a fluoroquinolones example, inhibits DNA synthesis by targeting DNA topoisomerase II and IV lead to accumulate of substitution mutations in the coding regions for particular subunits of DNA topoisomerase II.

Protein Synthesis Inhibitors

Linezolid prevents protein synthesis and is active against resistant Gram-positives. Linezolid inhibits the formation of the 70S ribosomal initiation complex through binding to the *23SrRNA* gene of the 50S subunit causing mutation in this region.

Cancer Mutation

Cancers arise from a single cell which has undergone mutation. Mutations in genes give the cell increased growth advantages compared to others and allow them to escape normal controls on proliferation.

The initial mutation will cause cells to divide to produce a genetically **homogeneous clone**. In turn, additional mutations occur, which further enhance the cells' growth potential. These mutations give rise to **subclones** within the tumour each with differing properties so that most tumours are **heterogeneous**.

Cancer cells contain many alterations which accumulate as tumours develop. Considerable information has been gathered on the regulation of cell growth and proliferation leading to the identification of the **proto-oncogenes** and the tumour suppressor genes. The proto-oncogenes encode proteins which are important in the control of **cell proliferation**, **differentiation**, **cell cycle control** and **apoptosis** (**programmed cell death**).

Mutations in these genes act dominantly and lead to increase in function. In contrast the **tumour suppressor genes** inhibit cell proliferation by arresting progression through the cell cycle and block differentiation.

Mutations leading to increased genomic instability suggest defects in mismatch and excision repair pathways. Genes involved in DNA repair, when mutated, also predispose the patient (make them vulnerable) to developing cancer (e.g. The failure of DNA repairs in xeroderma pigmentosum). In addition, many other genes encoding proteins, such as **proteinases** or other enzymes capable of disrupting tissues, and **vascular permeability factors**, have been shown to be involved in **carcinogenesis**.

Common Gene Mutation in Cancerous Cells

TP53 (Tumour suppressor gene)

The *TP53* gene (also called ***p53***) is located on the short arm of chromosome 17. Most cancer cells show a defect either in the *TP53* gene or in the pathway leading to *TP53* activation. In tumours without mutations, p53 is inactivated by **viral proteins** interacting with it or more commonly as a result of alterations of components of the p53 pathway.

Loss of p53 function in tumours can occur by a number of different routes including mutation of the gene itself as well as mutation of genes which regulate p53. Mutation of p53 is found in around 50% of cancers. A total of 74% of these mutations are missense, a much higher proportion that seen in other tumour suppressor genes.

EGFR (proto-oncogene)

The *EGFR* gene provides instructions for expressing a receptor protein (**tyrosine kinase glycoprotein**) called the **Epidermal Growth Factor Receptor**, which spans the cell membrane so that one end of the protein remains inside the cell and the other end projects from the outer surface of the cell. This positioning allows the receptor to attach (bind) to other proteins, called **ligands**, outside the cell and to receive signals that help the cell respond to its environment.

EGFR mutations are most common in people with **lung adenocarcinoma** (a form of non-small cell lung cancer,) are more common with lung cancer in non-smokers, and are more common in women than in men. *EGFR* is overexpressed in approximately 80% of esophageal squamous cell cancers. A point mutation in the transmembrane domain of the *EGFR* results in valine to glutamic acid substitution causing **receptor dimerization** and **kinase activation** in the absence of ligands.

Lecture 5

DNA Repair

DNA replication is a highly accurate process, but errors can occasionally occur, such as a DNA polymerase inserting a wrong base. Uncorrected errors may sometimes lead to serious consequences, such as cancer. Repair mechanisms correct the errors. In rare cases, errors are not corrected, leading to mutations; in other cases, repair enzymes are themselves mutated or defective. DNA repair processes exist in both prokaryotic and eukaryotic organisms. Most of the errors during DNA replication are promptly corrected by DNA polymerase by proofreading the base that has just been added (Figure1). **Proofreading** activity that assists most of the replicative polymerases is responsible for removal of incorrectly incorporated nucleotides from the primer terminus before further primer extension. The polymerase checks whether the newly added base has paired correctly with the base in the template strand. If it is the right base, the next nucleotide is added. If an incorrect base has been added, the enzyme makes a cut at the phosphodiester bond and releases the wrong nucleotide. This is performed by the exonuclease action of **DNA pol III**. Once the incorrect nucleotide has been removed, a new one will be added again.

Proofreading

1. DNA polymerase adds a new base to the 3' end of the growing, new strand. (The template has a G, and the polymerase incorrectly adds a T rather than a C to the new strand.)
2. DNA polymerase detects that the bases are mispaired.
3. DNA polymerase uses 3' to 5' exonuclease activity to remove the incorrect T from the 3' end of the new strand.

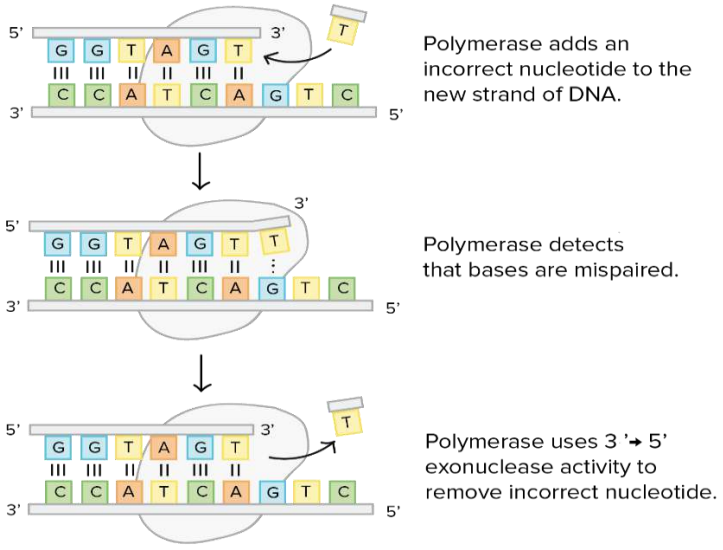


Figure1: Proofreading by DNA polymerase corrects errors during replication.

Mismatch Repair

DNA damage accumulates in cells over time as a result of exposure to **exogenous** chemicals and physical agents (i.e., cigarette smoke, asbestos, ultraviolet light …etc.), as well as **endogenous** reactive metabolites including **reactive oxygen species** (ROS). Another source of DNA damage is errors that occur during normal DNA replication. Nucleotide misincorporation generates DNA base-base mismatches during DNA synthesis. DNA damage, if unrepaired, has the potential to generate mutations in somatic or germline cells, which can alter cellular phenotype and cause dysfunction and diseases such as cancer. To prevent such deleterious effects, cells possess multiple mechanisms to repair DNA damage and thus prevent mutations. One such system is the critical pathway known as DNA mismatch repair

DNA **mismatch repair** is a highly conserved biological pathway that plays a key role in maintaining genomic stability. Many errors are corrected by proofreading, but few slip through. Mismatch repair happens right after new DNA has been made, and its job is to remove and replace mis-paired bases (ones that were not fixed by proofreading during replication but are instead corrected after replication is completed). The enzymes recognize the incorrectly added nucleotide and excise it; this is then replaced by the correct base. If this remains uncorrected, it may lead to more permanent damage (Figure 2).

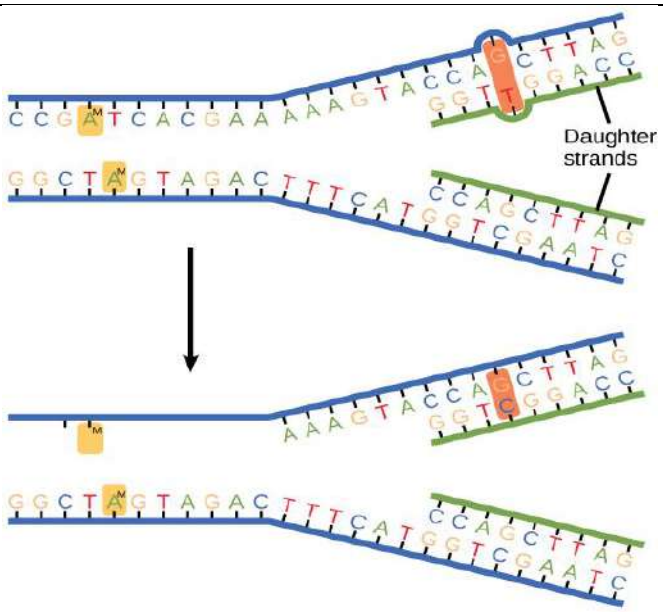


Figure 2: Detection of the incorrectly added nucleotide;
In mismatch repair, the incorrectly added base is detected after replication. The mismatch repair proteins detect this base and remove it from the newly synthesized strand by nuclease action. The gap is now filled with the correctly paired base.

How do mismatch repair enzymes recognize which of the two bases is the incorrect one?

In *E. coli*, after replication, the nitrogenous base adenine acquires a methyl group (CH3); the parental DNA strand will have methyl groups, whereas the newly synthesized strand lacks them. Based on the methylation status, exonucleases can remove the wrongly incorporated bases from the newly synthesized, non-methylated strand. The resulting single-stranded gap undergoes repair DNA resynthesis and ligation by DNA polymerase III, and DNA ligase respectively.

In human, the mechanism is similar to that of *E. coli*, but the DNA polymerase δ binds and nicks the mispairs in newly replicated DNA. Therefore, the processes that allow the original strand to be identified in mismatch repair involve recognition of nicks (single-stranded breaks) that are found only in the newly synthesized DNA.

Mismatch repair mechanism.

1. A mismatch is detected in newly synthesized DNA. There is a G in the new strand paired with a T in the template (old) strand.
2. The new DNA strand is cut, and a patch of DNA containing the mispaired nucleotide and its neighbours is removed.
3. The missing patch is replaced with correct nucleotides by a DNA polymerase.
4. A DNA ligase seals the remaining gap in the DNA backbone (Figure 3).

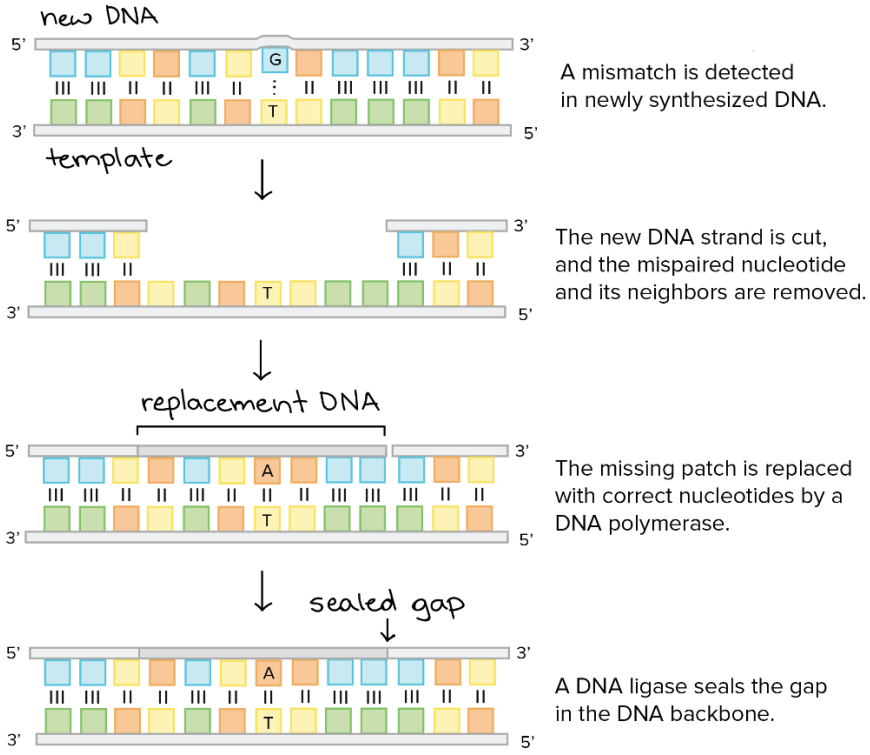


Figure 3: Mismatch repair mechanism

Other DNA damage repair mechanisms

- **Double-stranded break repair:** Two major pathways, **non-homologous end joining** and **homologous recombination**, are used to repair double-stranded breaks in DNA (that is, when an entire chromosome splits into two pieces).

Direct Reversal of DNA damage

One of the most important mechanisms of direct DNA reversal repair is reversal of alkylation damage by DNA alkyltransferases. In the alkylation damage, a methyl (CH_3) group attaches to an oxygen atom in the guanine by guanine methylation (Figure 4). The methyl-bearing guanine, if not fixed, will pair with thymine (T) rather than cytosine (C) during DNA replication. Fortunately, humans and many other organisms have DNA alkyltransferases that can remove the methyl group, reversing the reaction and returning the base to normal.

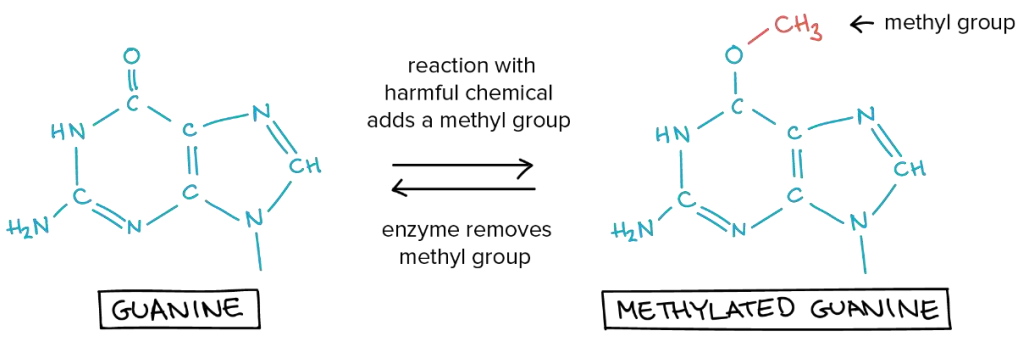


Figure 4: Methylation of guanine

Excision repair: Damage to one or a few bases of DNA is often fixed by removal (excision) and replacement of the damaged region. In **base excision repair**, just the damaged base is removed. In **nucleotide excision repair**, as in the mismatch repair mentioned above, a patch of nucleotides is removed.

I. Base excision repair of a deaminated cytosine

1. Deamination converts a cytosine base into a uracil. This results in a double helix in which a G in one strand is paired with a U in the other. The U was formerly a C but was converted to U via deamination.
2. The uracil (deaminated cytosines) is detected and removed by glycosylase, leaving a base-less nucleotide.
3. The base-less nucleotide is removed, leaving a 1-nucleotide hole in the DNA backbone.
4. The hole is filled with the right base by a DNA polymerase, and the gap is sealed by a ligase (Figure 5).

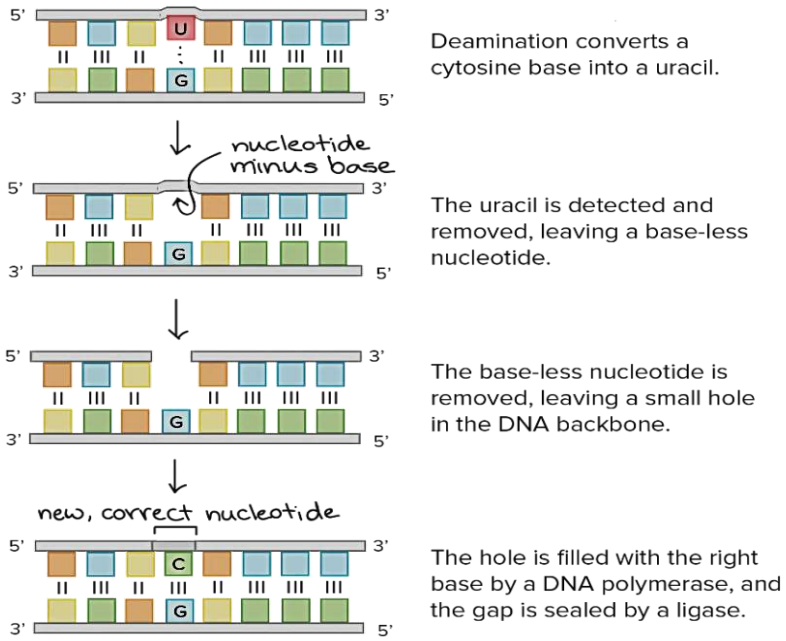


Figure 5: Base excision repair of a deaminated cytosine

II. Nucleotide excision repair of a thymine dimer.

1. UV radiation produces a thymine dimer. In a thymine dimer, two **Ts** that are next to each other in the same strand link up via a chemical reaction between the bases. This creates a distortion in the shape of the double helix.
2. Once the dimer has been detected, the surrounding DNA is opened by helicase to form a bubble.
3. Enzymes cut the damaged region (thymine dimer plus neighbouring regions of same strand) out of the bubble.
4. DNA polymerase replaces the excised (cut-out) DNA, and a ligase seals the backbone (Figure 6).

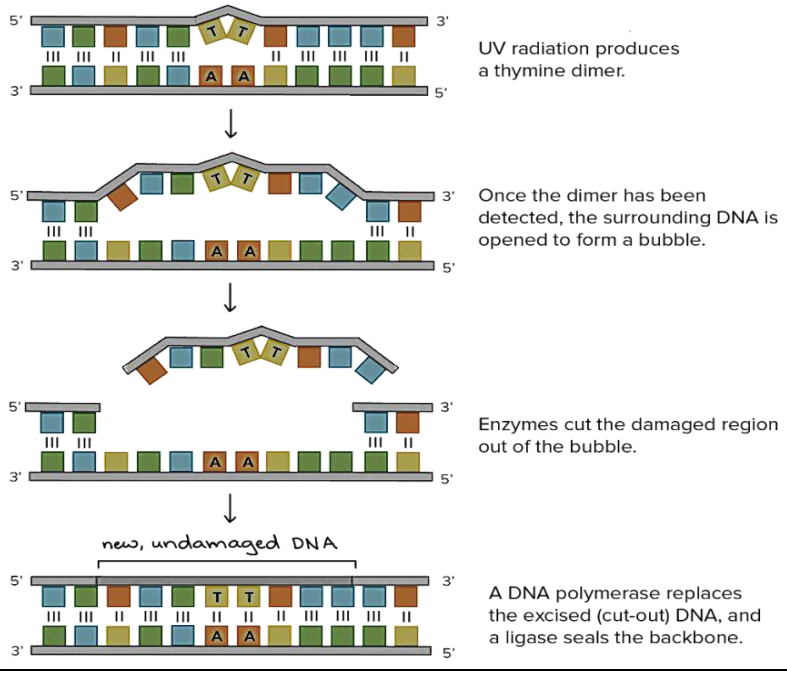


Figure 6: Nucleotide excision repair of a thymine dimer

Double-stranded break repair

Some types of environmental factors, such as high-energy radiation, can cause double-stranded breaks in DNA (splitting a chromosome in two), which are dangerous because large segments of chromosomes, and the hundreds of genes they contain, may be lost if the break is not repaired. Two pathways involved in the repair of double-stranded DNA breaks are the non-homologous end joining and homologous recombination pathways.

In **non-homologous end joining**, the two broken ends of the chromosome are simply glued back together (Figure 7). This repair mechanism is "messy" and typically involves the loss, or sometimes addition, of a few nucleotides at the cut site. So, non-homologous end joining tends to produce a mutation, but this is better than the alternative (loss of an entire chromosome arm).

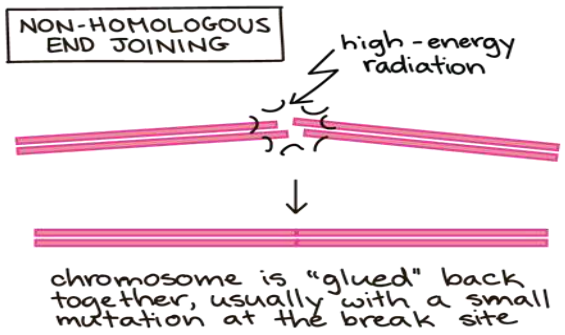


Figure 7: Non-Homologous End Joining

In **homologous recombination**, information from the homologous chromosome that matches the damaged one (or from a sister chromatid, if the DNA has been copied) is used to repair the break (Figure 8). In this process, the two homologous chromosomes come together, and the undamaged region of the homologue or chromatid is used as a template to replace the damaged region of the broken chromosome using sequences copied from this homologue. Homologous recombination is "cleaner" than non-homologous end joining and does not usually cause mutations.

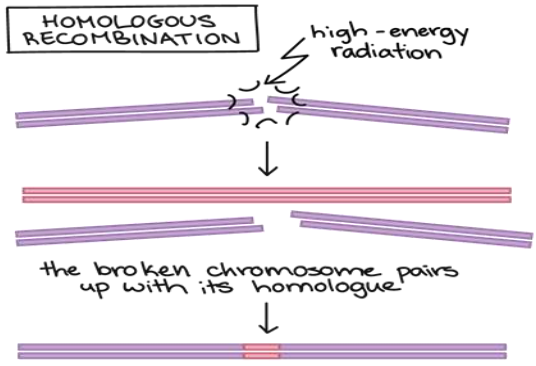


Figure 8: Homologous Recombination

DNA proofreading and repair in human disease

In many cases, mutations in genes that encode proofreading and repair proteins are associated with heredity cancers (cancers that run in families). For example:

- **Hereditary nonpolyposis colorectal cancer** (also called **Lynch syndrome**) is caused by mutations in genes encoding certain mismatch repair proteins. Since mismatched bases are not repaired in the cells of people with this syndrome, mutations accumulate much more rapidly than in the cells of an unaffected person. This can lead to the development of tumours in the colon.
- People with **xeroderma pigmentosum** are extremely sensitive to UV light. This condition is caused by mutations affecting the nucleotide excision repair pathway. When this pathway doesn't work, thymine dimers and other forms of UV damage can't be repaired. People with xeroderma pigmentosum develop severe sunburns from just a few minutes in the sun, and about half will get skin cancer by the age of 10 unless they avoid the sun (Figure 9).

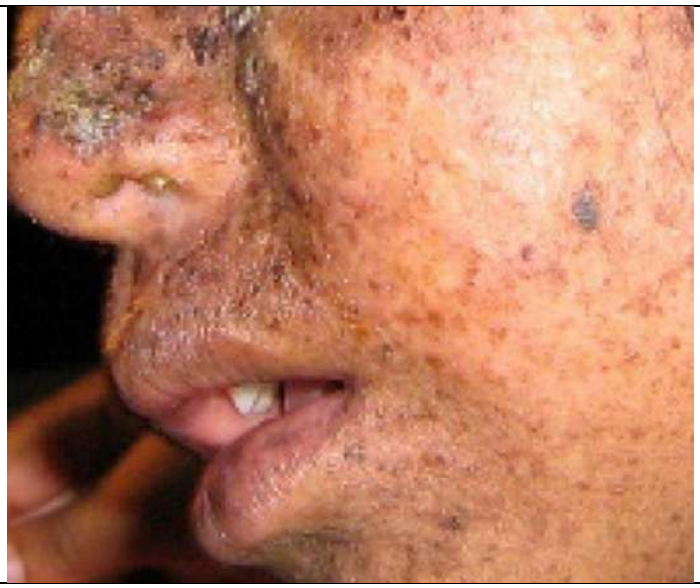


Figure 9: Xeroderma pigmentosa is a condition in which thymine dimerization from exposure to UV is not repaired. Exposure to sunlight results in skin lesions

Lecture 6

Transcription

In the early 1950s, Francis Crick suggested that there is a unidirectional flow of genetic information from DNA through RNA to protein. This is known as the **central dogma** of molecular biology (Figure 1), DNA is transcribed to messenger RNA, that contains the same sequence information as the template DNA, and subsequently this RNA message is translated into a protein sequence.

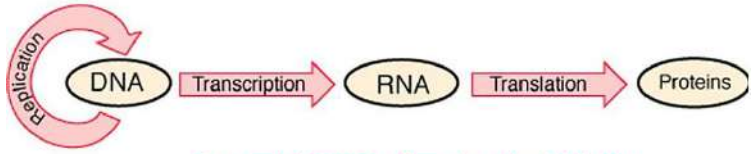


Figure 1: Central dogma of molecular biology

Gene expression, a process by which information from a gene is used in the synthesis of a functional gene product, such as protein or RNA (some genes code for functional RNA molecules, such as tRNA and rRNA). Genes are expressed by being first transcribed into RNA and may then subsequently be translated into protein. In prokaryotes, the gene expression occurs in the cytoplasm while in eukaryote, the transcription occurs in nucleus and translation in the cytoplasm.

Transcription

Transcription is a process of making an RNA strand from a DNA template, and the RNA molecule that is made is called transcript. It is the first step in gene expression. The ribonucleic acids (RNA) are an important class of molecules. They serve as the intermediary in the flow of information from DNA to protein in the flow of genetic information. Some viruses even use RNA, instead of DNA, to carry their genetic information. The features of RNA are:

1. It is a polymer, made up of a limited number of nucleotides.
2. RNA nucleotides contain ribose (An important structural feature of RNA that distinguishes it from DNA is the presence of a hydroxyl group at the 2' position of the ribose sugar).
3. RNA has the base uracil rather than thymine that is present in DNA. In addition to the four nitrogen bases, RNA also has some unusual bases.
4. RNA molecules are usually single stranded and do not form double helices. The single RNA strand is folded upon itself by hydrogen bonds. This helps in the stability of the molecule.

5. The enzyme used in transcription of RNA is RNA polymerase. A significant difference between DNA polymerase and RNA polymerase is that RNA polymerase can initiate chain growth without a primer.
6. The nucleotide in an RNA molecule is complimentary to the sequence of bases in the DNA template. Thus, the bases C, T, G, and A in a DNA strand cause G, A, C, and U, respectively.
7. Nucleotides are added only to the 3'-OH end of the growing chain. Therefore, direction of grown RNA chain is 5'-to-3'.
8. Only one strand of DNA serves as a template for RNA synthesis.

Types of RNA

There are three types of cellular RNA have been distinguished, messenger RNA (mRNA), ribosomal RNA (rRNA) and transfer RNA (tRNA).

Ribosomal RNA

rRNA molecules are the major functional components of the ribosomes (approximately 60% rRNA and 40% protein) and is essential for protein synthesis in all living organisms. It comprises about 80% of the total RNA of the cell. The ribosome consists of proteins and RNA (nucleoprotein). The ribosomal RNAs form two subunits, the large subunit and small subunit. The large subunit rRNA acts as a ribozyme, catalyzing peptide bond formation. A small and a large subunit consist of rRNA of various size and a small quantity of proteins. The 70S ribosome of prokaryotes consists of a 30S subunit and a 50S subunit. The 30S subunit contains 16S rRNA, while the 50S subunit contains 23S and 5S rRNA. The 80S eukaryote ribosome consists of a 40S and a 60S subunit.

Messenger RNA

mRNA accounts for just 5% of the total RNA in the cell. It carries the genetic code copied from the DNA during transcription in the form of triplets of nucleotides called codons. Each codon specifies a particular amino acid, except the stop codons. In prokaryote, mRNA is mature at transcription and requires no processing while in eukaryotic, mRNA derived from transcription is called primary transcript mRNA (known as pre-mRNA) requires extensive processing to became mature mRNA. This mature mRNA is then translated into a polymer of amino acids.

The pre-mRNA is submitted to processing that involves (Figure 2):

- **Polyadenylation:** Polyadenylation occurs during and/or immediately after transcription of DNA into RNA. mRNA molecules are polyadenylated at the 3' end. The poly(A) tail protecting mRNA from degradation by exonucleases.

- **Capping:** addition of 7-methylguanosine on the 5' end of the mRNA to protect mRNA from cleaving by exonucleases. It also serves as a recognition site of the mRNA prior to translation for the small ribosomal subunit.
- **Splicing:** removal of introns and merging of the adjacent exons.

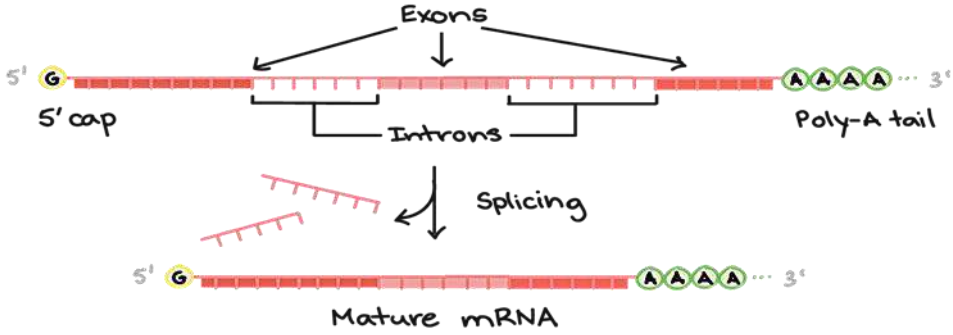


Figure 2: processing of pre-mRNA to mature RNA

Transfer RNA

A transfer RNA is an adaptor molecule that serves as the physical link between the mRNA and the amino acid sequence of proteins. tRNA does this by carrying an amino acid to the ribosome. tRNA constitutes 15% of the total RNA and is directly involved in the translation of the mRNA. The structure of tRNA can be decomposed into its primary structure, its secondary structure (usually visualized as the cloverleaf structure), and its tertiary structure (L-shaped 3D structure) shown in figure 3.

The tRNA structure consists of the following:

1. A 5'-terminal phosphate group.
2. **The acceptor stem**: 3'-terminal nucleotide which contains the CCA 3'-terminal group used to attach the amino acid.
3. **The D arm**: stem ending in a loop that often contains dihydrouridine.
4. **The anticodon arm**: is contains the anticodon.
5. **The T arm**: containing the sequence TΨC where Ψ is pseudouridine, a modified uridine.

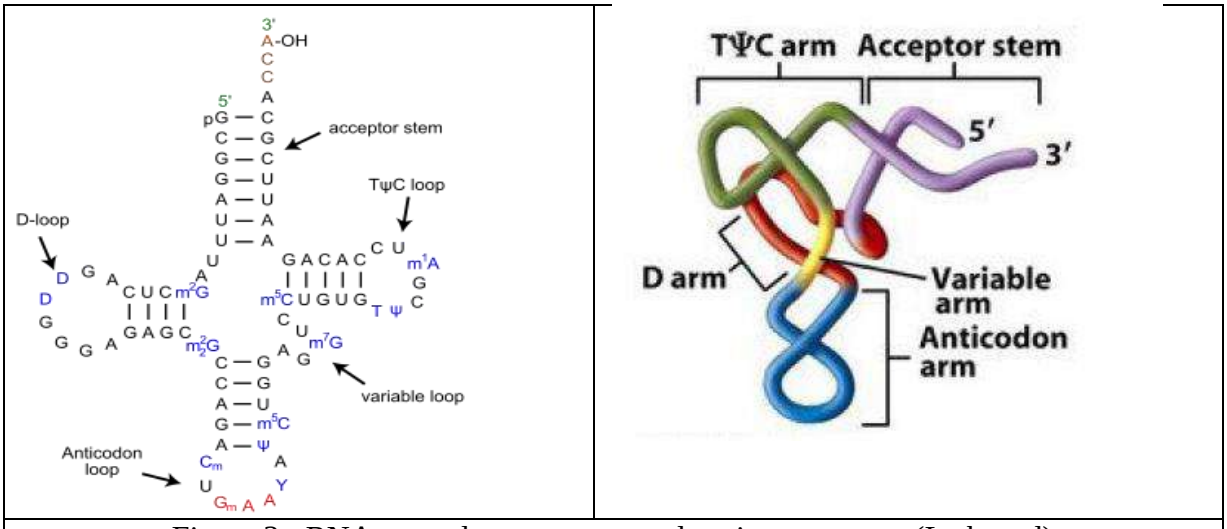


Figure 3: tRNA secondary structure and tertiary structure (L-shaped)

RNA transcription requires the following components:

- The enzyme RNA polymerase (DNA dependent RNA polymerase).
- DNA template
- All four types of ribonucleoside triphosphates (ATP, GTP and UTP)
- Divalent metal ions Mg++ or Mn++ as a co-factors
- No primer is needed for RNA synthesis

RNA polymerase

RNA polymerases (DNA dependent RNA polymerases) are multisubunit enzymes responsible for the polymerisation of ribonucleotides into a sequence complementary to the template DNA. Bacteria and Archaea each have a single RNA polymerase (Figure 4) transcribes all types of RNA, while the eukaryotes contain three enzymes: RNA polymerase I (synthesizes rRNA), RNA polymerase II (synthesizes mRNA), and RNA polymerase III (synthesizes tRNA).

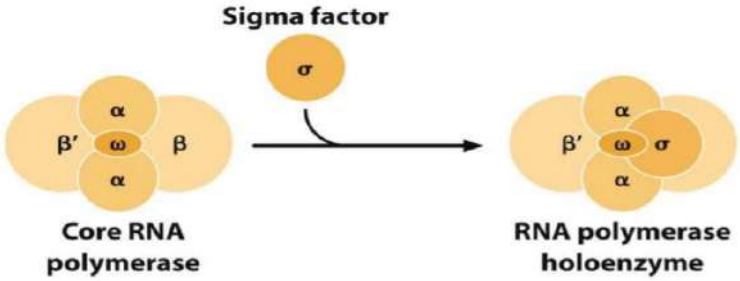


Figure 4: RNA polymerase

In prokaryotes, the enzyme consists of five subunit types. Two alpha (α) subunits, beta (β) subunit, beta prime (β′), omega (ω) and sigma (σ). The σ subunit can dissociate from the rest of the complex, leaving the **core enzyme.** The complete enzyme with σ is termed the **RNA polymerase holoenzyme** and is necessary for correct initiation of transcription, whereas the core enzyme can continue transcription after initiation.

A single RNA polymerase performs multiple functions in transcription process included:

1. Search and binds to promoter site
2. Unwinds a short stretch of double helical DNA.
3. Selects correct ribonucleotide and catalyze the formation of phosphodiester bond.
4. Detects termination signals.

Lecture 7

Promoter recognition

A **promoter** is a regulatory region of DNA located upstream (5' region) of of a gene, providing a control point for regulated gene transcription. The sigma factor plays an important role in recognizing promoter sequences, and after successful initiation it is released from the holoenzyme. The promoter contains specific DNA sequences called **consensus sequences** that are recognized by proteins. In prokaryotes, RNA polymerase only requires the associated protein sigma factor to bind the promoter. On the other hand, the process in eukaryotes is much more complex. Eukaryotes require a minimum of seven transcription factors in order for the binding of RNA polymerase II to the promoter. In prokaryotes, the promoter consists of two short sequences at -10 and -35 positions upstream from the transcription start site (+1) (Figure 5).

- The sequence at **-10** is called the Pribnow box, or the -10 element, and usually consists of the six nucleotides **5´-TATAAT-3´** This box composed of A and T bases which are easier to separate and melt.
- The other sequence at **-35** (the -35 element) usually consists of the six nucleotides **5´-TTGACA-3´**. Its presence allows a very high transcription rate.

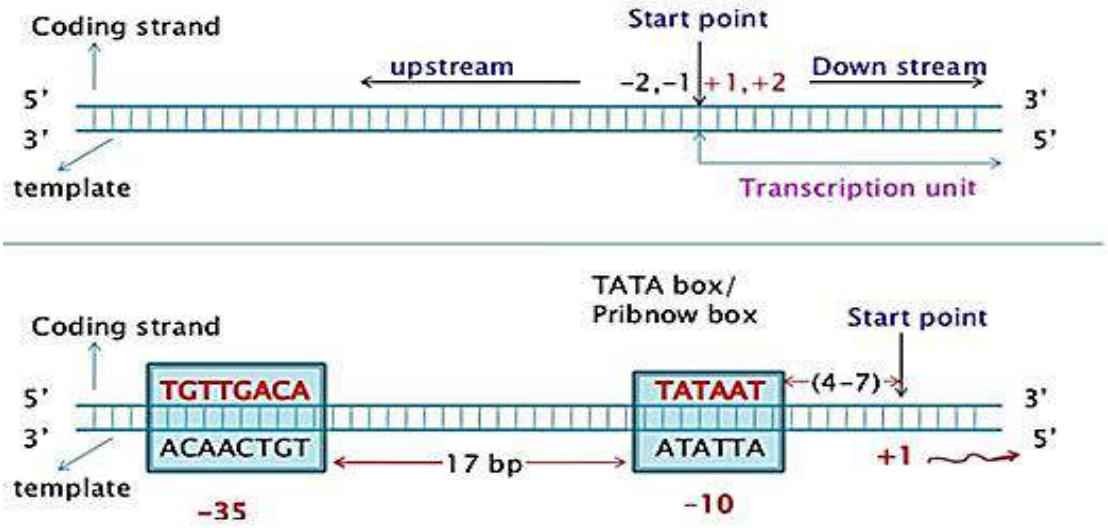


Figure 5: promoter region

The stretch of DNA that is transcribed into RNA is also called a **transcription unit** (Figure 6) that has the following parts:

1. Promoter region.
2. Start point (initiation site)
3. Coding segment
4. Terminator sequence.

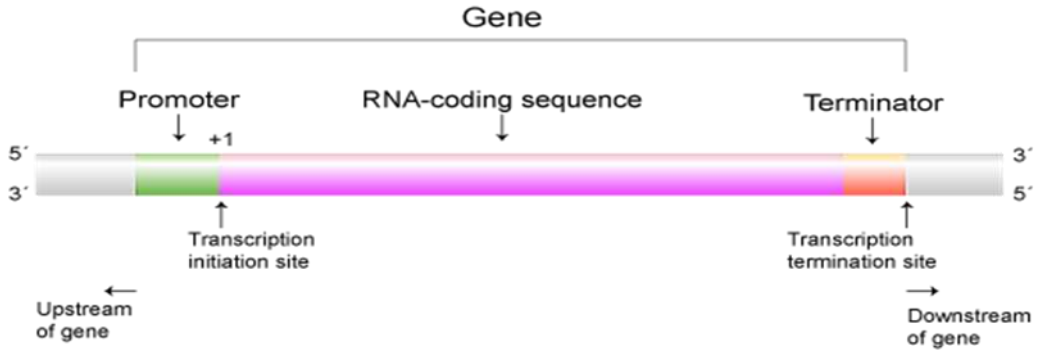


Figure 6: transcription unit

The nucleotide on the DNA template strand that corresponds to the site from which the first 5′ RNA nucleotide is transcribed is called the +1 nucleotide, or the initiation site. Nucleotides prior initiation site are given "-" numbers and are designated **upstream** whereas nucleotides following the initiation site are denoted with "+" numbering and are called **downstream** nucleotides.

The DNA strand whose sequence matches that of the RNA is known as the **coding strand** and the complementary strand on which the RNA was synthesized is the **template strand** (Figure 7**)**. The only difference is that in RNA all of the T nucleotides are replaced with U nucleotides.

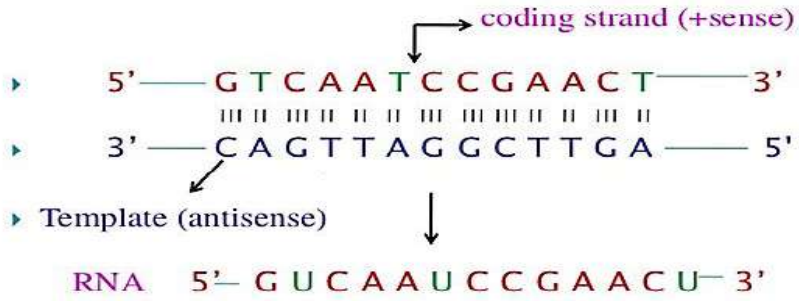


Figure 7: coding and template strand

Transcription process

Initiation

The first step in transcription is initiation, RNA polymerase recognize and bind specifically to promoter regions −35 and −10 regions. The resulting structure is termed a **closed promoter complex**. Then, the enzyme binds more tightly, unwinding bases near the −10 region to form an **open promoter complex**. The region of unwinding is called a **transcription bubble**. This initiation step, the formation of an open complex, requires the sigma factor, several phosphodiester bonds are made by RNA polymerization activity in the 5´ to 3´ direction. DNA-RNA hybrid is formed.

Elongation

Transcription elongation begins with the release of the polymerase σ subunit. RNA polymerase moves progressively along the transcribed DNA strand, adding nucleotides to the growing RNA chain. Only one DNA strand, the template strand, is transcribed. Almost 40 nucleotides/ sec are added for the RNA growing chain (Figure 8).

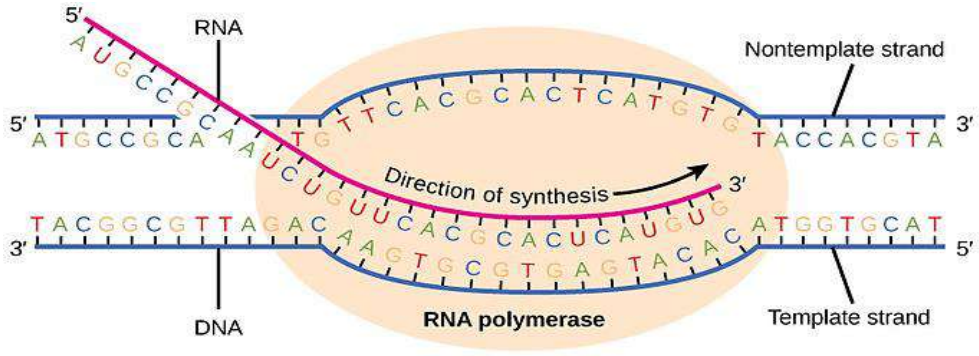


Figure 8: Elongation of transcription process

Termination

RNA polymerase continues transcribing DNA until it reaches a termination signal. There are two types of transcriptional termination in prokaryotes, the **Rho-independent terminator** (intrinsic terminator) is a region of DNA with two inverted repeats separated by about six bases, followed by a stretch of As (Figure 9). When RNA polymerase makes these sequences, the two inverted repeats form a hairpin structure.

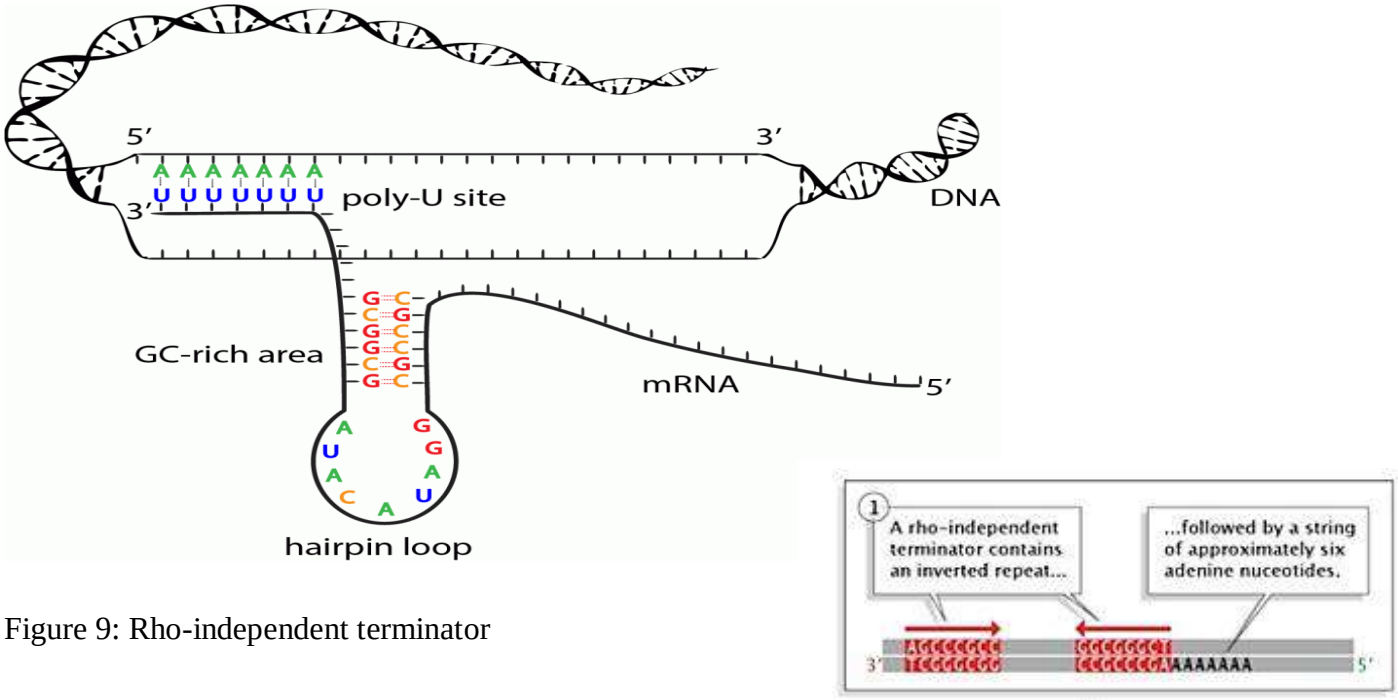


Figure 9: Rho-independent terminator

The secondary structure causes RNA polymerase to pause. As the stretch of As is transcribed into Us, the DNA/RNA hybrid molecule becomes unstable (A/U base pairs have only two hydrogen bonds) because of the strong binding between C and G in hairpin structure. The RNA transcript free from the DNA template and RNA polymerase released.

Rho-dependent terminators that have two inverted repeats but lack the string of As. Rho (ρ) protein is a special helicase that unwinds DNA/RNA hybrid double helices. Rho binds upstream of the termination site in a region containing many cytosines and use its ATPase activity to provide the energy to translocate along the RNA until it reaches the RNA–DNA helical region, where it unwinds the hybrid duplex structure. The RNA polymerase is then released (Figure 10).

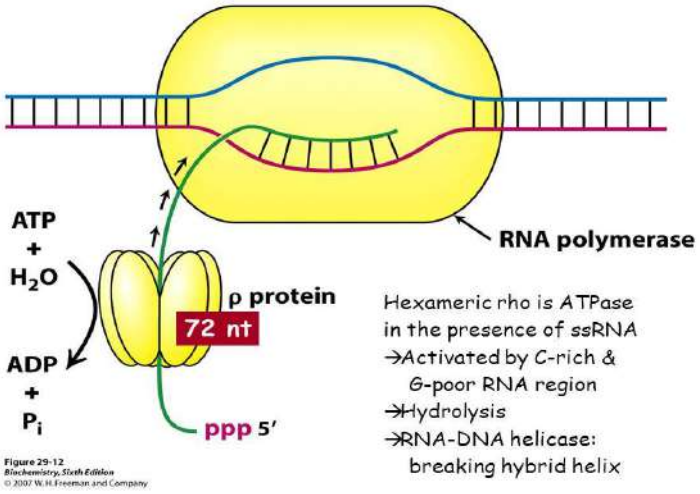


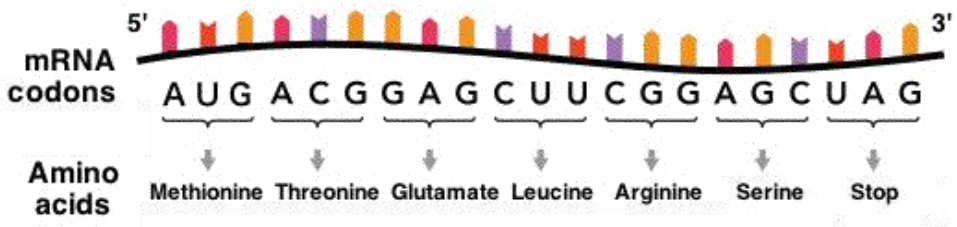
Figure 10: Rho-dependent terminator

Translation

Translation is the process by which ribosomes read the genetic message in mRNA and produce a protein product according to the message's instructions. Ribosomes therefore serve as protein factories. Transfer RNAs (tRNAs) play an equally important role as adapters that can bind an amino acid at one end and interact with the mRNA at the other.

The genetic code

In an mRNA, the instructions for building a polypeptide are RNA nucleotides (As, Us, Cs, and Gs) read in groups of three. These groups of three are called **codons**.



The letters A, G, T and C correspond to the nucleotides found in DNA. They are organized into codons. The collection of codons is called genetic code. For 20 amino acids there should be 20 codons. Each codon should have 3 nucleotides.

- 1 Nucleotide gives 4 combinations.
- 2 Nucleotides give 16 combinations.
- 3 Nucleotides give 64 combinations. Three out of these are non-sense codons and the 61 codons for 20 amino acids, each codon is specific for only one amino acid.

One codon, AUG, specifies the amino acid methionine and also acts as a **start codon** to signal the start of protein synthesis. **Stop codons** are UAA, UAG, and UGA. (Figure 1).

The genetic code is described as degenerate, or redundant, because a single amino acid may be coded for by more than one codon (synonymous), such as leucine have six codons and proline have 4 codons. The code evolved in such a way as to minimize the deleterious effects of mutations

Genetic Code in all living organisms are the same (Universal) with an exception to universality is found in mitochondrial codons such as AUA code for methionine.

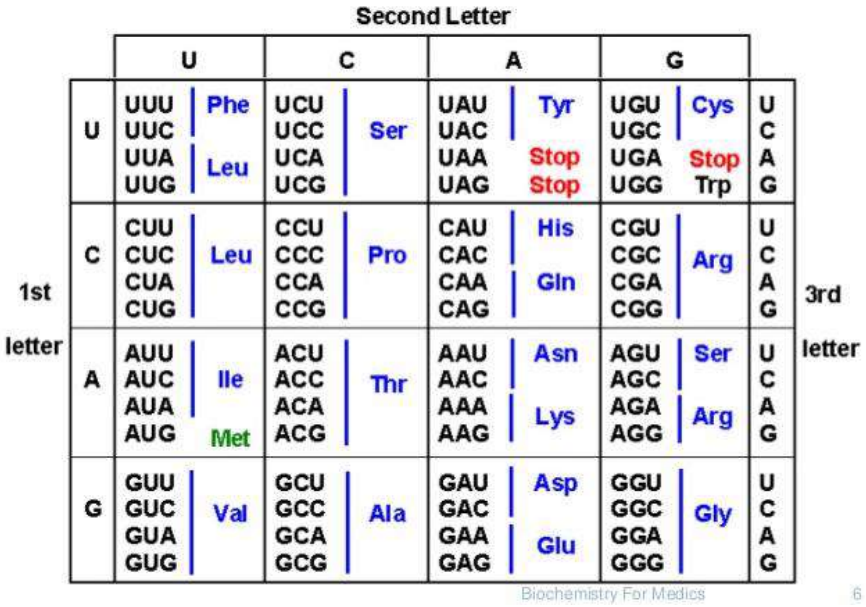


Figure 1: Genetic code

Lecture 8

Wobble Hypothesis

The Wobble Hypothesis, by Francis Crick in 1966, states that the 3rd base in an mRNA codon can undergo non-Watson-Crick base pairing with the 1st base of a tRNA anticodon. Movement (wobble) of the base in the anticodon position is necessary for small conformational adjustments that affect the overall pairing geometry of anticodons of tRNA. The mRNA codon's first 2 bases form hydrogen bonds with their corresponding bases on the tRNA anticodon in the usual Watson-Crick manner, in that they only form base pairs with complimentary bases (Figure2).

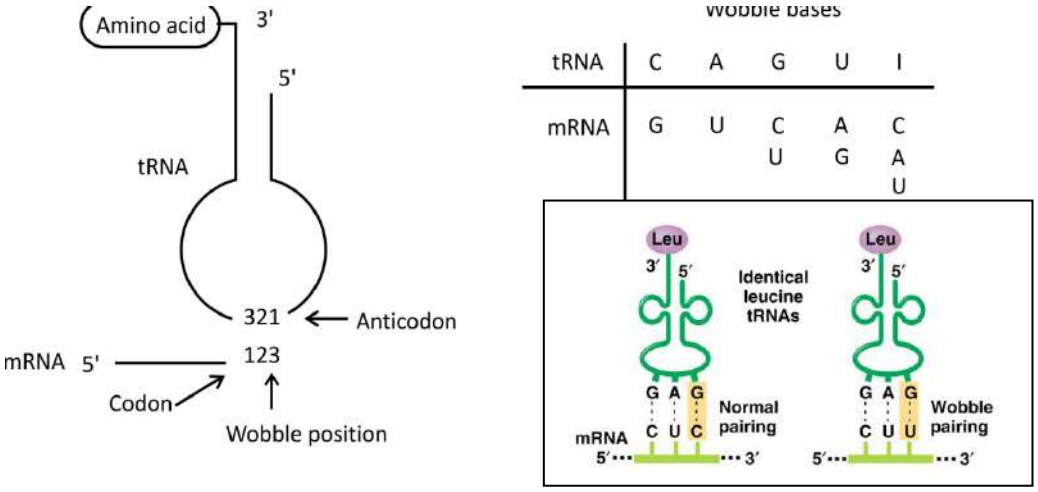


Figure 2: Wobble Hypothesis

In the genetic code, there are 61 possible sense codons. For translation, each of these codons requires a tRNA molecule with a complementary anticodon. If each tRNA molecule is paired with its complementary mRNA codon using usual Watson-Crick base pairing, then 61 types of tRNA molecule would be required. In the standard genetic code, three of these 64 mRNA codons (UAA, UAG and UGA) are stop codons. There are 40-60 tRNA in the cell; most organisms have fewer than 45 types of tRNA. Therefore, tRNA types must pair with more than one codon. For example, the G base in the 1st position of tRNA anticodon may pair with either C or U base in the 3rd position of the mRNA codon.

tRNA Charging

This process is called **tRNA charging;** the tRNA is said to be charged with an amino acid. All tRNAs have the same three bases (CCA) at their 3´ ends, and the terminal adenosine is the target for charging. An amino acid is attached by an ester bond

between its carboxyl group and the 2´- or 3´-hydroxyl group of the terminal adenosine of the tRNA (Figure 3).

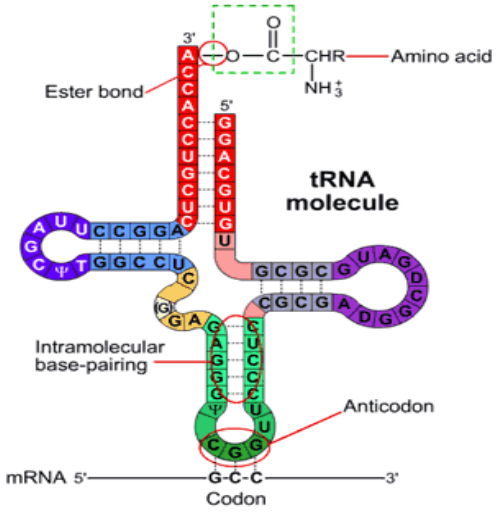


Figure 3:tRNA charging

Charging takes place in two steps, both catalyzed by the enzyme **aminoacyl-tRNA synthetase (**20 synthetases exist, one for each amino acid**).**

In the first reaction, the amino acid is activated, using energy from ATP; the product of the reaction is aminoacyl-AMP.

amino acid + ATP → aminoacyl-AMP + pyrophosphate (PPi)

In the second reaction of charging, the energy in the aminoacyl-AMP is used to transfer the amino acid to a tRNA, forming aminoacyl-tRNA (acylated or charged tRNA).

aminoacyl-AMP + tRNA → aminoacyl-tRNA + AMP

The process of translation involves the following steps:

Initiation

Initiation of translation in prokaryotes involves the assembly of the components of the translation system, which are:

1. Two ribosomal subunits (50S and 30S subunits).
2. Mature mRNA to be translated.
3. tRNA charged with N-formylmethionine (the first amino acid).
4. Guanosine triphosphate (GTP) as a source of energy.
5. Three prokaryotic initiation factors IF1, IF2, and IF3, which help the assembly of the initiation complex.

<u>The ribosome has three active sites (Figure 4):</u>

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 uncharged tRNA after it gives its amino acid to the growing peptide chain.

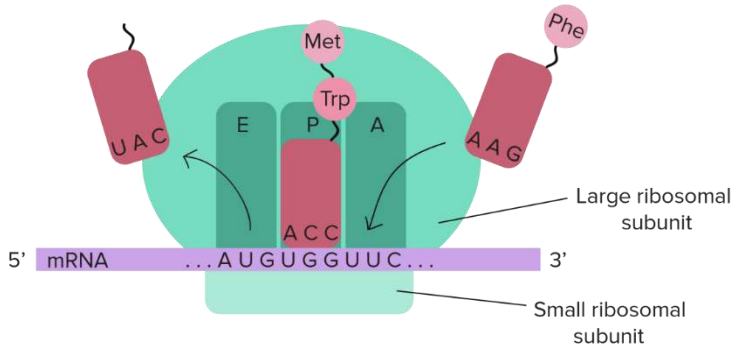


Figure 4: Ribosome active sites

The selection of an initiation site (usually an AUG codon) depends on the interaction between the 30S subunit and the mRNA template. The 30S subunit binds to the mRNA in a site called **ribosome binding site** that consist of two kinds of functional sequences:

1. Short sequence of purine-rich region (5´-AGGAGGU-3´) is called **Shine-Dalgarno sequence or leader sequence** upstream of the initiation codon. The Shine-Dalgarno sequence is complementary to a pyrimidine rich region on the

16S rRNA component of the 30S subunit (Figure 5).

Figure 5: Shine-Dalgarno sequence

2. **Initiation codon**: the initiation codon in mRNA of prokaryote is AUG that recognized by special tRNA UAC anticodon. This tRNA carries the N-formyl methionine (tRNA ^{fmet}) which is a modified methionine functions only at the initiation site and removed during maturation of protein after translation.

Small ribosomal subunit (30S) binds to mRNA. An initiator tRNA with the anticodon UAC base pairs with the start codon AUG, then large ribosomal subunit (50S) completes the initiation complex, the energy molecule GTP provides energy and initiation factors needed for assembly of initiation complex (Figure 6).

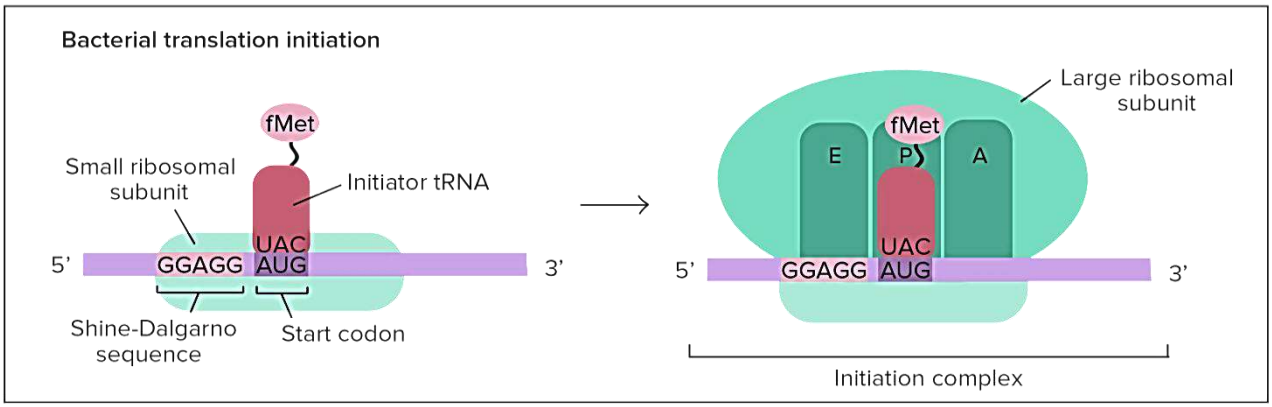


Figure 6: initiation complex in translation process

In eukaryotes, translation includes a different set of 12 initiation factors (eIFs). the initiator tRNA carrying methionine (tRNA $^{met)}$ attaches the small ribosomal subunit (40S). Together, they bind to the 5' end of the mRNA by recognizing the 5' G cap of the mRNA searching the AUG initiation codon. The ribosome begins translation at an AUG that is located within the Kozak consensus sequence 5´-ACC**AUG**G-3´ (Figure 7).

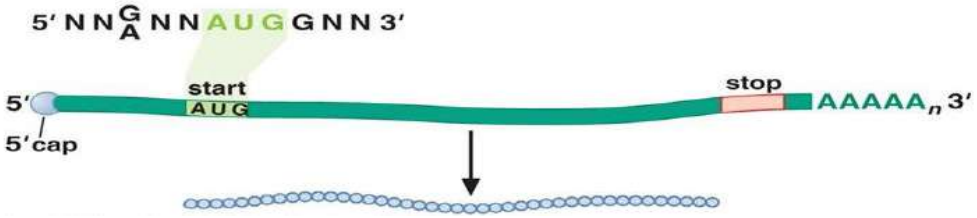


Figure 7: Kozak consensus sequence

1. **Elongation**

Elongation of the polypeptide chain involves addition of amino acids to the carboxyl end of the growing chain. Elongation starts when the fMet-tRNA enters the P site, causing a conformational change which opens the A site for the new aminoacyl-tRNA to bind (new amino acid). This binding is facilitated by elongation factor and energy (GTP).

Now the P site contains the first amino acid (fmet) in the beginning of the peptide chain and the A site has the next amino acid to be added to the peptide chain. **A peptide bond** is formed between the **carboxyl group** of fmet at the P site and the **amino group** of the newly arrived amino acid at the A site. This reaction is catalyzed by the **peptidyl transferase** activity of the 23S rRNA molecule in the large ribosomal subunit. The fmet is separated from it is tRNA (Figure 8).

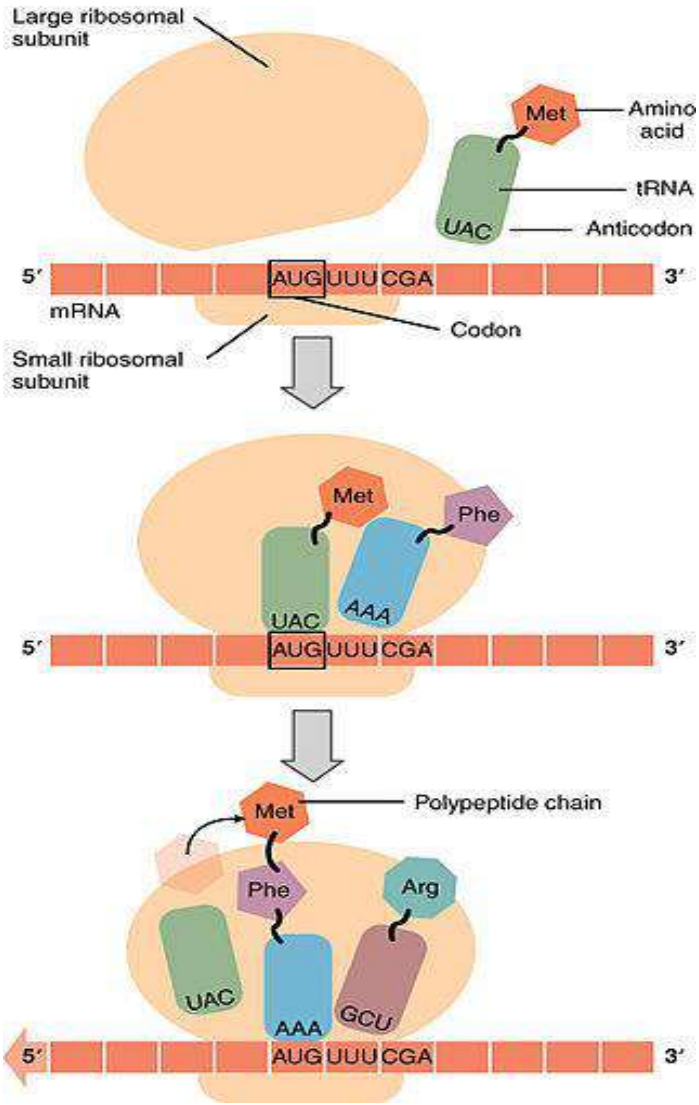


Figure 8: Elongation in in translation process

The next step in elongation is the movement of the ribosome three nucleotide (one codon) down the mRNA in the 5´ to 3´ direction. Therefore, tRNA carrying the elongated polypeptide translocates from the A site to the P site. This step is called **translocation** which requires **translocase** enzyme. The discharged tRNA (tRNA with no amino acids) moves from the P site to the E site and leaves the ribosome. The ribosome continues to translate the remaining codons on the mRNA as more aminoacyl-tRNA bind to the A site, until the ribosome reaches a stop codon on mRNA.

Proteins in prokaryotes are synthesized at a rate of only 18 amino acid residues per second. In bacteria, translation initiation occurs as soon as the 5' end of an mRNA is synthesized, and translation and transcription are coupled.

A number of ribosomes translating the same mRNA transcript is called a polysome or polyribosome (Figure 9).

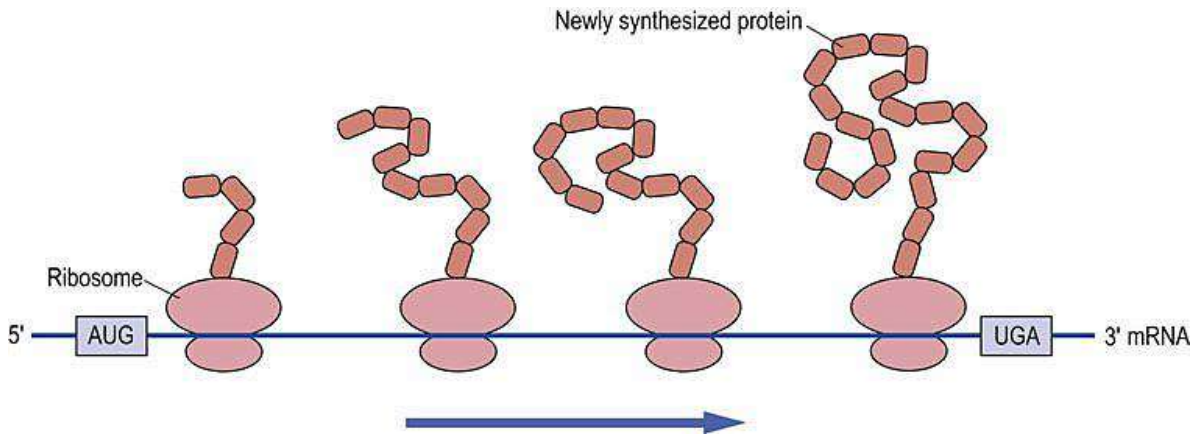


Figure 9: polyribosome

2. Termination

Termination occurs when one of the three termination codons (UAA, UGA and UAG) moves into the A site. These codons are not recognized by any tRNAs. Instead, they are recognized by proteins called release factors (RF1, RF2 and RF-3). The release factor promotes hydrolysis of the bond between the tRNA and the polypeptide, freeing it from the ribosome and components of the ribosomal complex dissociate. Translation is complete. Proteins must be folded into the proper three-dimensional conformations to work properly.

Lecture 9

REGULATION OF GENE EXPRESSION

Each cell of a living organism contains thousands of genes. All genes do not function at a time, but they are function according to requirements of the cell. Thus, there exists an **on-off mechanism** which regulates gene expression in all living cells which is called regulation of gene expression.

The *E. coli* genome contains over 3000 genes. Some of these are active all the time because their products are in constant desired. But some of them are turned off most of the time because their products are rarely needed.

There are various stages at which the expression of a gene can be regulated but most common is the initiation of transcription. Other levels of gene regulation are transcriptional elongation, mRNA processing and post translation stage.

Gene Regulation in Prokaryotes:

In bacteria the expression of genes is controlled by extracellular signals often present in the medium in which bacteria are grown. These signals are carried to the genes by two types of regulatory proteins which are DNA binding proteins.

1. Positive regulators called activators.
2. Negative regulators called repressors.

Gene expression can be positively or negatively controlled. In positive control binding of activator protein triggers the transcription whereas in negative control, binding of repressor protein inhibits the transcription (Figure 1).

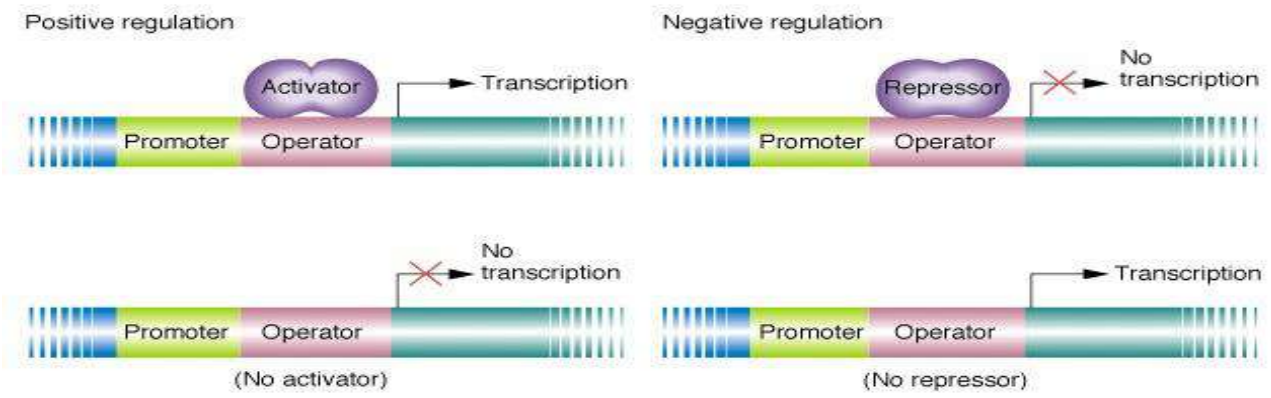


Figure 1: Positive and Negative regulation

The Operon

The operon refers to a cluster of genes which act together and code for various enzymes of a particular biochemical pathway. These genes are under the control of a single promoter. The genes are transcribed together into an mRNA strand and translated together in the cytoplasm. The expression of prokaryotic operons leads to

the generation of polycistronic mRNAs, while eukaryotic operons lead to monocistronic mRNAs (Figure 2).

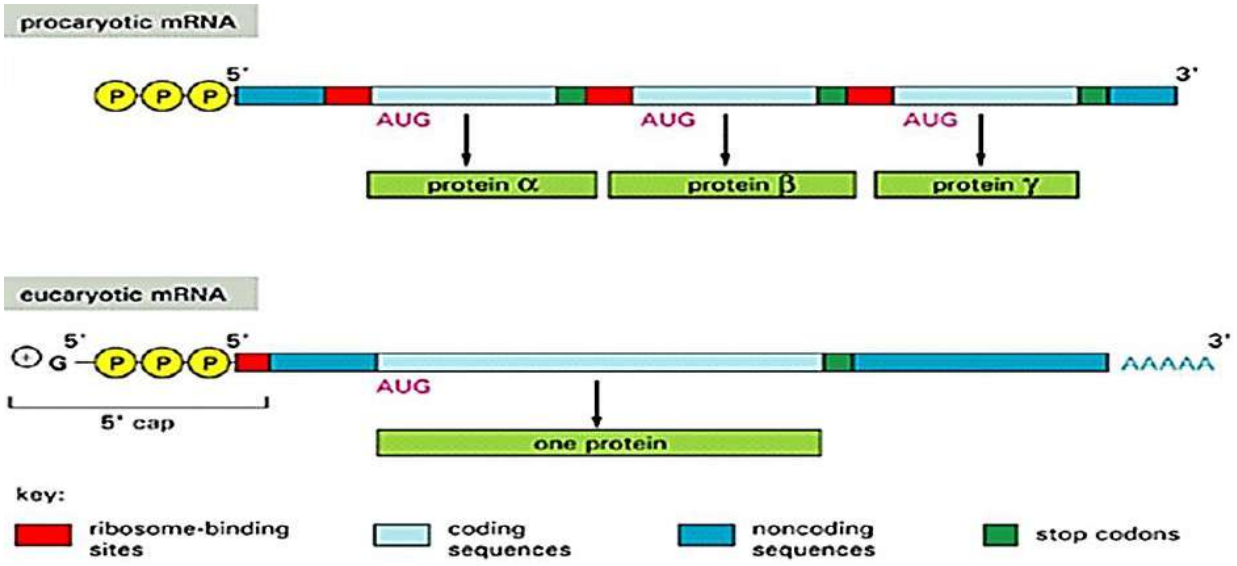


Figure 2: polycistronic and monocistronic mRNAs.

- **Polycistronic mRNA**: One single mRNA strand carries information from more than one gene (in prokaryotes).
- **Monocistronic mRNA**: one single mRNA strand carries information from only one gene (in eukaryotes).

An operon is made up of the following DNA components:

- **Promoter:** a nucleotide sequence that enables a gene to be transcribed. The promoter is recognized by RNA polymerase, which then initiates transcription.
- **Operator**: a segment of DNA that a repressor binds to it. In the case of a repressor, the repressor protein physically prevents the RNA polymerase from transcribing the genes.
- **Regulatory gene:** is a sequence of DNA encoding a regulatory protein (such as a repressor). Repressors are small regulatory proteins that turn off transcription. They bind to the operator region of an operon and prevent RNA polymerase from binding to promoter.
- **Structural genes:** the genes that are regulated by the operon. In all organisms, structural genes can be classified into two groups:

1. **Constitutive genes**, also called housekeeping genes: encoding RNA and proteins having essential functions such as rRNA, ribosomal proteins, proteins of cellular respiratory system, etc... These genes are expressed continually and with a stable amount.

2. **Inducible genes**: encoding proteins necessary for the survival of the organism in changing environment. They are expressed only when their proteins are needed by the cell.

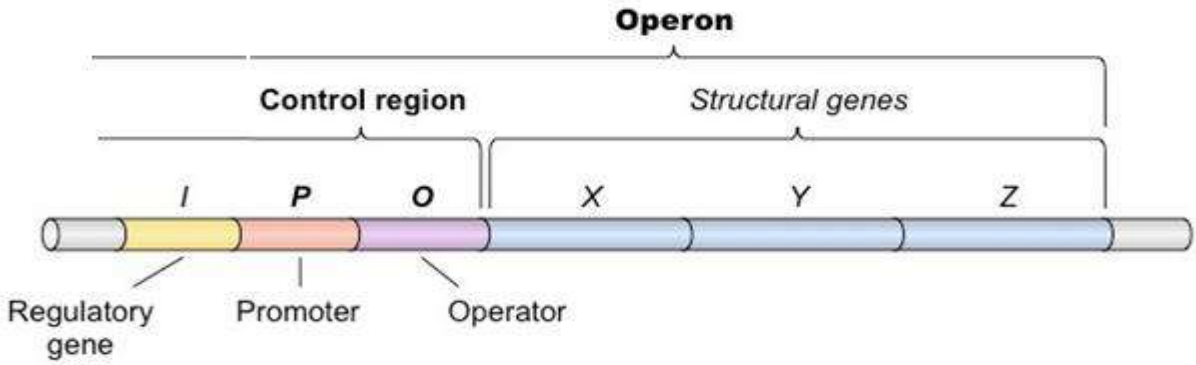


Figure 3: Operon.

Operon regulation can be either negative or positive by induction or repression. The on-off mechanism of operon depending on the metabolic pathways:

- In **catabolic pathway** (degradation of macromolecules into structural units), when substrate degradation is occurred, the operon is "switched on". These operons are characterized as inducible operons.
- In **anabolic pathway** (synthesis of macromolecules from small ones), when a product is needed by the cell, the corresponding operon is "switched off". These operons are considered as repressible operons.

Induction	Repression
1. It turns the operon on.	1. It turns the operon off.
2. It starts transcription and translation.	2. It stops transcription and translation.
3. It is caused by a new metabolite, which needs enzymes to get metabolised.	3. It is caused by an excess of existing metabolite.
4. It operates in a catabolic pathway.	4. It operates in an anabolic pathway.
5. Repressor is prevented by the inducer from joining the operator gene.	5. Aporepressor is enabled by a co-repressor to join the operator gene.

lac operon (inducible operons)

The first discovery of a gene regulation system is widely considered to be the identification in 1961 of the lac operon, discovered by Jacques Monod, in which some enzymes involved in lactose metabolism are expressed by *E. coli* only in the presence of lactose and absence of glucose.

The lac operon is a negative inducible operon, composed of (Figure 4):

1. **Regulatory sequences**: (1) the operator (O) which binds the repressor protein.
(2) the promoter (P) containing two binding sites, one for the RNA polymerase, the other for CAP-cAMP complex
2. **Structural genes** involved in lactose metabolism: gene Z, Y and A
3. **Regulator gene (gene I):** It is located away from the structural genes and constantly transcribes mRNA to produce repressor protein. It is the key element of operon because the function of the operon is dependent on it.

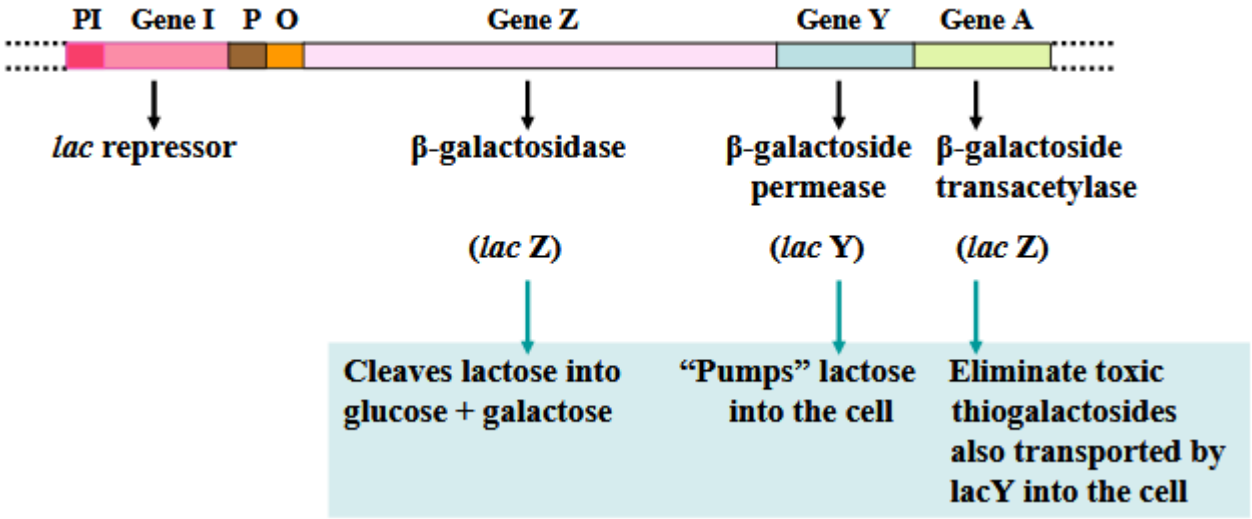


Figure 4: lac operon

Lac operon has two controls:

Negative control involves the binding of a repressor to the operator to prevent transcription. In negative inducible operons, a regulatory repressor protein is normally bound to the operator, which prevents the transcription of the genes on the operon. If an inducer molecule is present, it binds to the repressor and changes its conformation so that it is unable to bind to the operator. This allows for expression of the operon.

Positive control, an activator protein stimulates transcription by binding to DNA. In positive inducible operons, activator proteins are normally unable to bind to the

DNA. When an inducer is bound with the activator protein, it undergoes a change in conformation so that it can bind to the DNA and activate transcription.

Negative control in *E. coli*

When *E. coli* is grown in a medium in absence of lactose, the regulator gene produces a repressor protein (lac repressor) that binds the operator gene and prevents the binding of RNA polymerase to promoter. It stops the transcription of mRNA from structural genes and thus protein synthesis is switched off. Hence, no enzymes are produced.

If lactose is added to the medium, lactose (allolactose, inducer molecule) forms a complex with lac repressor, causes a conformation change of the repressor releasing from the operator. RNA polymerase can then be bind to the promoter and initiates the transcription of the three structural genes (Figure 5).

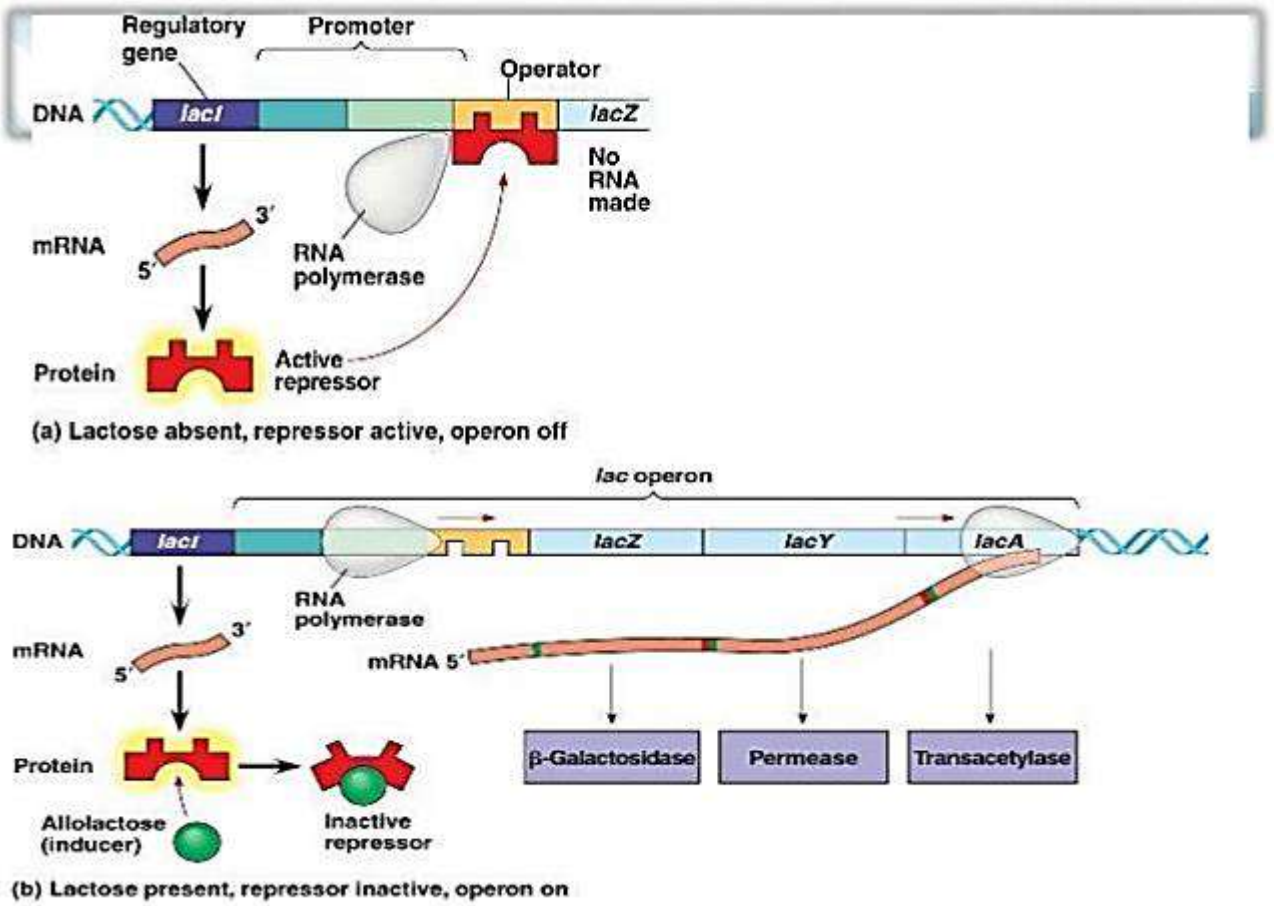


Figure 5: negative control of lac operon.

Positive control

In the presence of **abundant glucose**, cAMP levels will be low, so CAP **(catabolite activator protein)** will not bind to the *lac* operon to enhance transcription. In the

absence of lactose, the *lac* repressor will remain bound to the *lac* operator, preventing transcription (Figure 6A).

(Figure 6A).

In the **absence of glucose and the presence of lactose**, cAMP levels will be high, so CAP protein and cAMP will bind to the *lac* operon to enhance transcription. In the presence of lactose, the *lac* repressor will not bind to the *lac* operator, allowing transcription. The presence of CAP-cAMP will cause abundant transcription (Figure 6B).

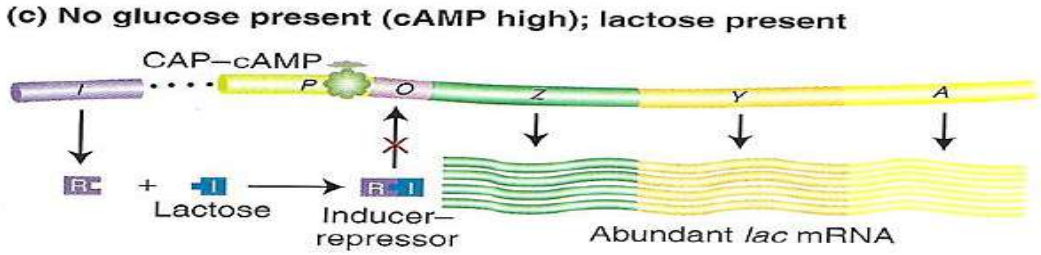


Figure 6 B: positive control of lac operon.

Tryptophan Operon (repressible operons)

The tryptophan (trp) operon is a negative control repressible operon involved in the synthesis of tryptophan, composed of (Figure 7):

1. Regulatory sequences: the operator and promoter regions.
2. Structural genes include trpE, D, C, B, A involved in the synthesis of tryptophan.
3. Regulatory gene (trpR) for the repressor.

A special gene, trpL, encodes the Leader peptide which underlies a regulation mechanism called "**attenuation**".

1) When *E. coli* is grown in a medium having tryptophan:

The regulatory gene produces repressor protein known as **apo-repressor** binds with tryptophan to form a repressor-co-repressor complex. This functional repressor protein binds with operator gene. The operator gene is switched off. RNA polymerase cannot be transferred to structural genes. So, there is no formation of any enzymes for expression of character.

2) When *E. coli* is grown in a medium without tryptophan:

The apo-repressor protein cannot bind to the operator lead to switch on of transcription. The RNA polymerase moves forward, and structural genes produce five enzymes which help in the formation of tryptophan amino acid.

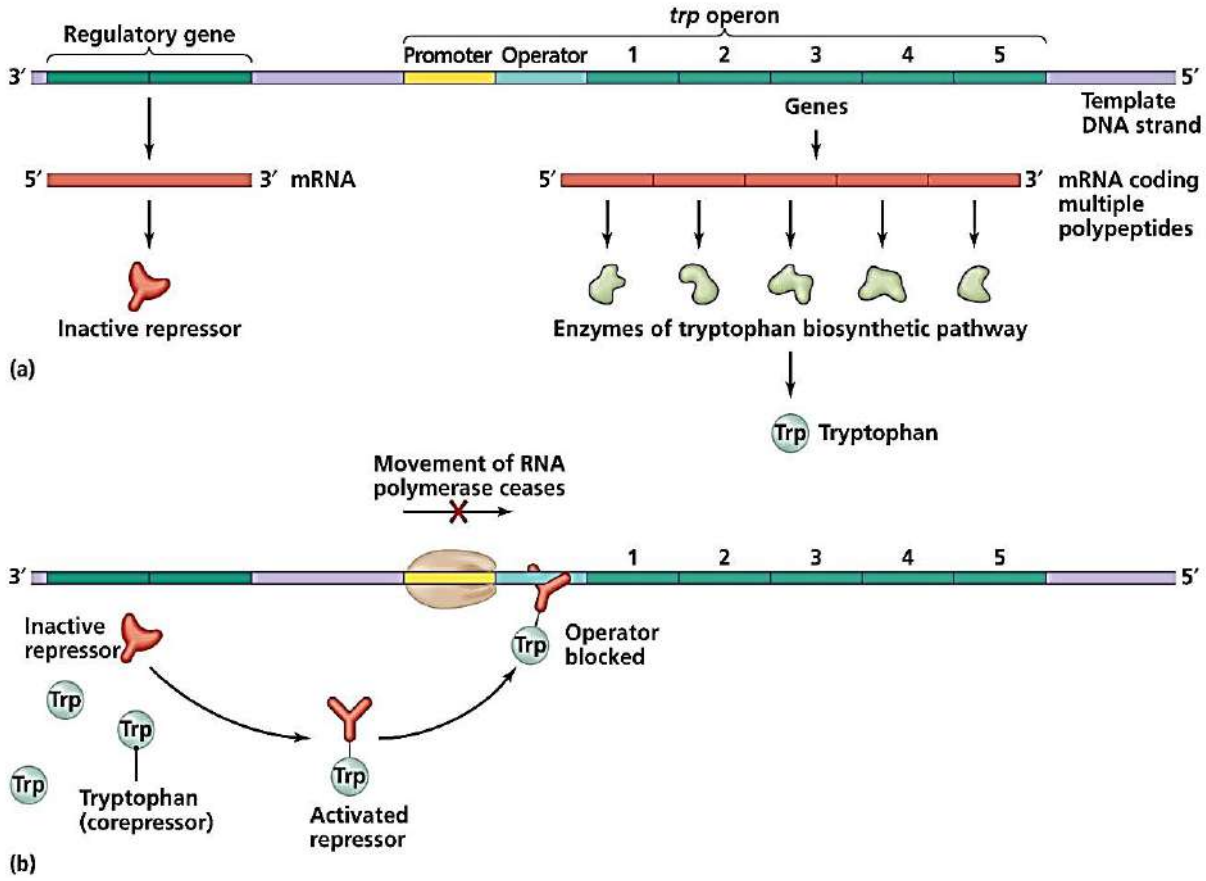


Figure 7: control of Trp operon.

Lecture 10

Gene Transfer in Bacteria

Mutation and gene transfer from one individual to another, play an important role in the diversity and evolution of new variants among organisms in nature.

Gene transfer in bacteria can transfer mainly in two ways:

1. **Vertical gene transfer:** transfer of gene from mother to daughter cell or parents to offspring (from one generation to another). It occurs during the binary fission in bacteria. DNA inherited from parental organism.
2. **Horizontal gene transfer:** also known as lateral gene transfer of gene between cells of the same generation in two different species. DNA acquired from unrelated individuals. There are three types of horizontal gene transfer included: (Figure 1)
   - **Conjugation** donor DNA is transferred from one bacterial cell to another by direct contact.
   - **Transformation** the uptake of free DNA molecules released from one bacterium (the donor cell) by another bacterium (the recipient cell).
   - **Transduction** the bacterial genes are carried from a donor cell to a recipient cell by a bacteriophage.

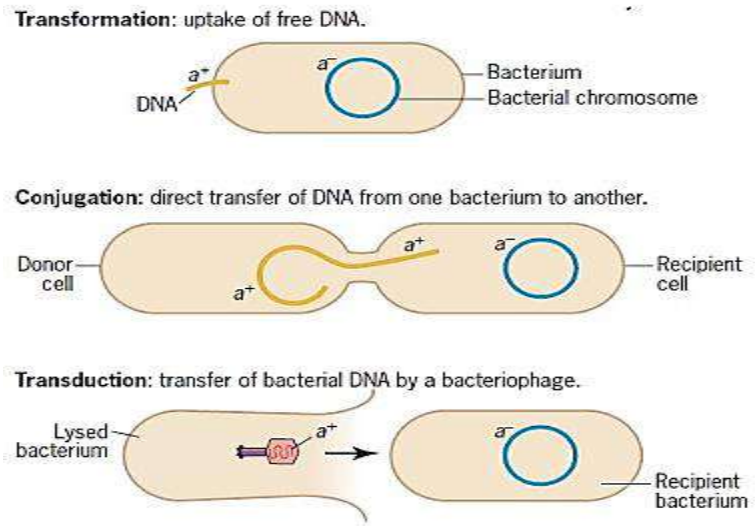


Figure 1: The three types of gene transfer in bacteria.

Two important things are noticed in three mechanisms:

1. Transfer of gene goes in only one direction, from donor cell to recipient cell.
2. Most recipients receive only 3% or less of donor DNA.

The different methods of gene transfer in bacteria allow geneticists to make detailed maps of bacterial genes.

Table 7.6 Natural Mechanisms of Horizontal Genetic Transfer in Bacteria

Mechanism	Requirements	State of Donor	State of Recipient
Transformation	Free DNA in the environment and a competent recipient	Dead	Living
Transduction	Bacteriophage	Killed by bacteriophage	Living
Conjugation	Cell-to-cell contact and F plasmid (either in cytosol or incorporated into chromosome of donor)	Living	Living

Figure1: horizontal gene transfer mechanisms

Conjugation

Conjugation is a process in which DNA is transferred from a bacterial donor cell to a recipient cell by cell-to-cell contact. It has been observed in many bacterial species and is best understood in *E. coli*, in which it was discovered by Joshua Lederberg in 1951. When bacteria conjugate, DNA is transferred to a recipient cell from a donor cell under the control of a set of genes present in a non-chromosomal, circular DNA molecule called a **plasmid that contain F factor** or the **fertility factor.**

Plasmids

Plasmids are circular DNA molecules that can replicate independently of the chromosome and range in size from a few kilo-bases (Kb) to a few hundred kilobases. The F factor is approximately 100 kb in length and contains many genes for its maintenance in the cell and its transmission between cells. Plasmids have been observed in many bacterial species and are usually not essential for growth of the cells but essential at specific conditions. They contain genes for antibiotic resistance, heavy metal resistance, nitrogen fixation and pollutant degradation.

The number of copies of a particular plasmid in a cell varies from one plasmid to the many copies.

High copy number plasmids are found in as many as 50 copies per host cell is called **relaxed plasmids**, whereas low copy number plasmids are present to the extent of 1 or 2 copies per cell is called **stringent plasmids**.

Cells that contain F are donors and are designated F+ ("F plus"); those lacking F are recipients and are designated F- ("F minus"). The F plasmid is a low copy number plasmid (Figure 2).

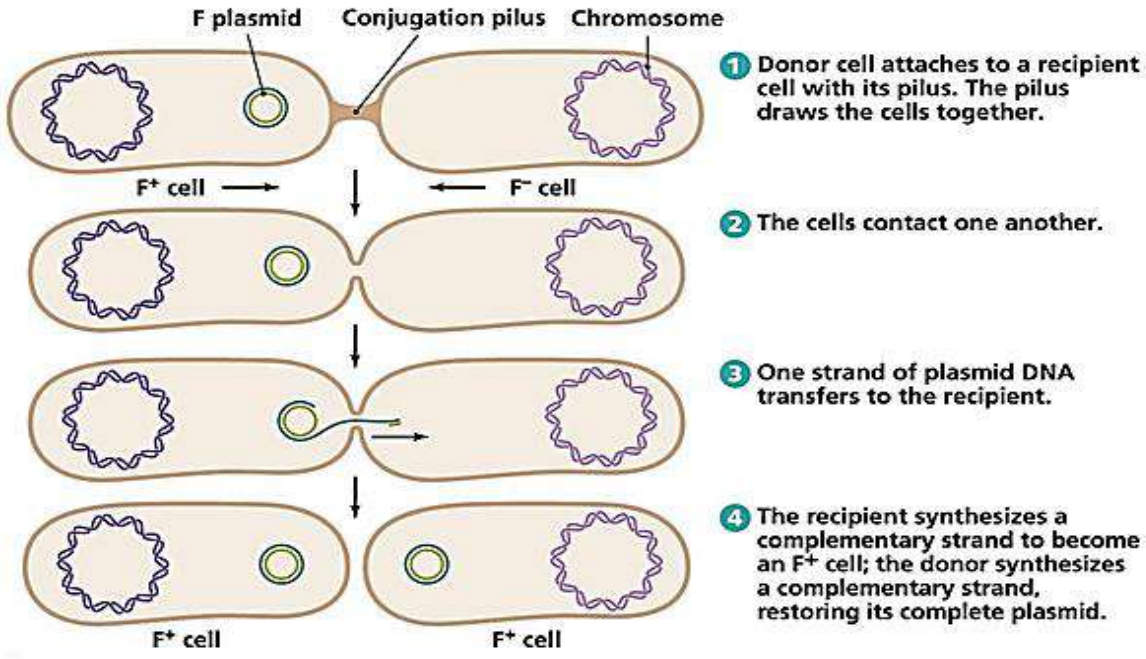


Figure 2: conjugation between F+ and F-

Conjugation mechanism

1. Donor cell produces pilus.
2. Pilus extended forming cytoplasmic bridge attaches to recipient cell and brings the two cells together.
3. The plasmid is nicked by the **relaxase** enzyme (DNase) in one strand and this DNA strand is then transferred to the recipient cell. Relaxase may work alone or in a complex of over a dozen proteins known collectively as a **relaxosome**.
4. Both cells synthesize a complementary strand to produce a double stranded circular plasmid and also reproduce pili; both cells are now viable donor for the F-factor.

Hfr Cells

The F plasmid is an **episome** which means a plasmid that can integrate itself into the bacterial chromosome by homologous recombination produce **Hfr cell (high frequency recombination)** because both the F factor plasmid and the chromosome have similar sequences, allowing the F factor to insert itself into the genome of the cell.

1. Hfr conjugation occurs when an integrated F plasmid from the Hfr donor forms a pilus and attaches to a recipient F- cell.
2. A nick in one strand of the Hfr cell's chromosome is created.

3. DNA begins to be transferred from the Hfr cell to the recipient cell while the second strand of its chromosome is being replicated.
4. The pilus detaches from the recipient cell. The Hfr cell often separate before the complete transfer of the entire chromosome due to its large size.
5. The F- cell remains F- because the entire F factor sequence was not received and is degraded by enzymes. Some of the donor's chromosomal DNA may also be transferred with the plasmid DNA. These DNA regions replace homologous regions in the recipient chromosome. The result is that some F- cells become recombinants containing one or more genes from the Hfr donor cell. The amount of chromosomal DNA that is transferred depends on how long the two conjugating bacteria remain in contact. In common laboratory strains of *E. coli* the transfer of the entire bacterial chromosome takes about 100 minutes in 37ºC.
6. In very rare cases (1 in 10000 Hfr cell transfer), the F factor will be completely transferred and the F- cell will become an Hfr cell

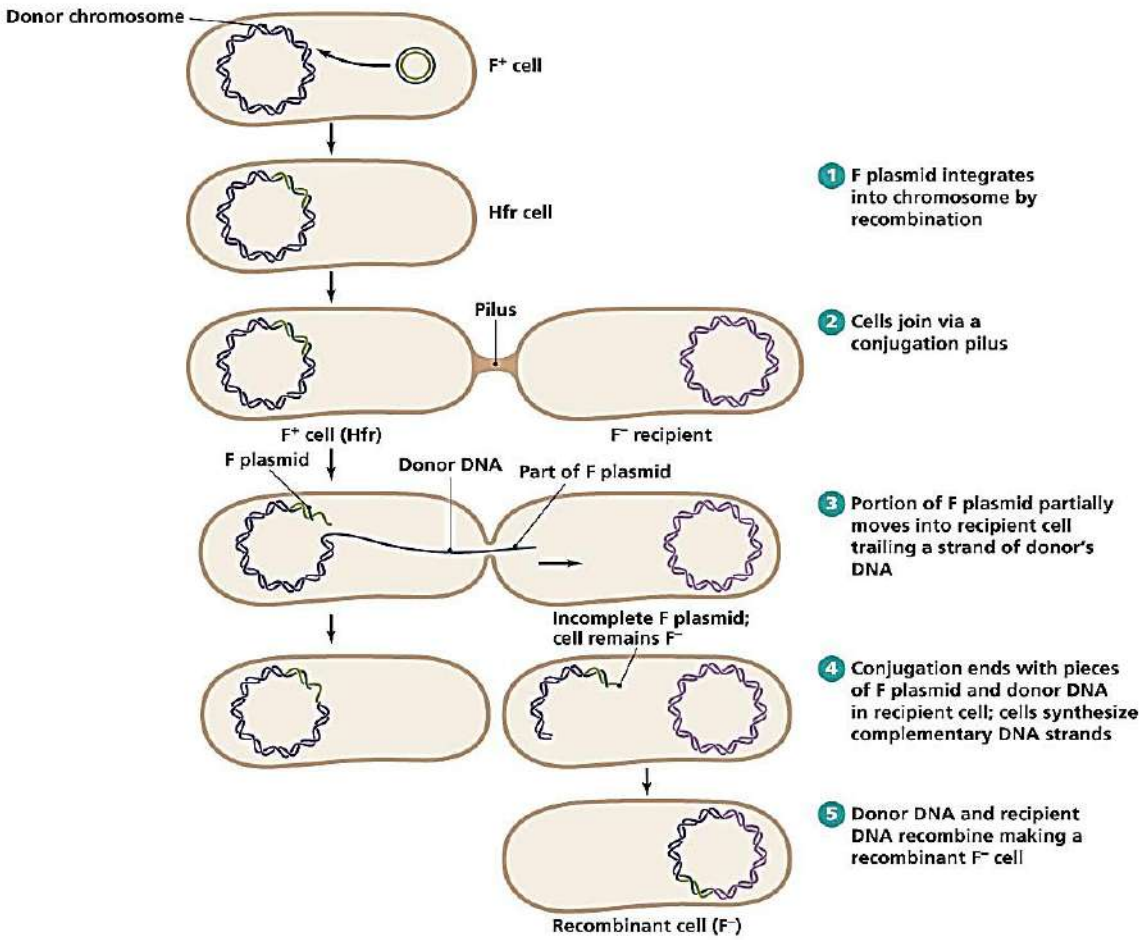


Figure 3: conjugation between Hfr and F-

Several differences between F transfer and Hfr transfer:

• It takes 100 minutes under the usual conditions for an entire bacterial chromosome to be transferred, in contrast with about 2 minutes for the transfer of F result of the relative sizes of F and the chromosome (100 kb versus 4700 kb).

• During transfer of Hfr DNA into a recipient cell, the mating pair usually breaks apart before the entire chromosome is transferred. Under usual conditions, several hundred genes are transferred before the cells separate.

• In a mating between Hfr and F- cells, the F- recipient remains F- because cell separation usually takes place before the final segment of F is transferred.

• In Hfr transfer, some regions in the transferred DNA fragment become incorporated into the recipient chromosome.

F' Plasmids

Occasionally, F is excised from Hfr DNA. This excision creates a plasmid containing a fragment of chromosomal DNA, which is called an **F'plasmid** ("F prime"). (Figure 4)

Conjugation between **F'** and F- resulted in the two cells became **F'** and the recipient cell become partially diploid that is called merozygote "merodiploid" part diploid and part haploid. It is diploid for the bacterial gene carried by the F' (one copy on the F' and the other on the chromosome), and haploid for all other genes.

Conjugation with an F' is much faster and more efficient than with an Hfr, because only a very small piece of DNA is transferred.

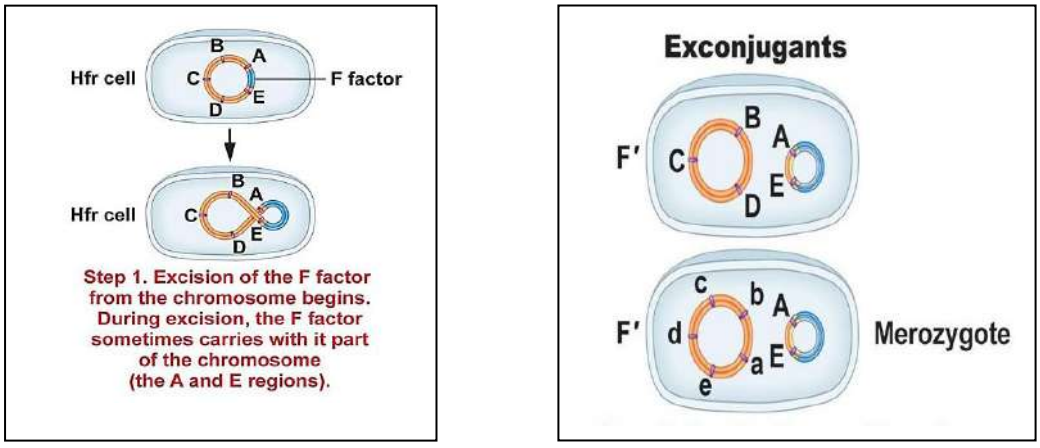


Figure 4: conjugation between F prime and F-

Bacterial Transformation

Bacterial transformation is a process in which recipient cells acquire genes from free DNA molecules in the surrounding medium. For transformation to take place, the recipient bacteria must be in a state of **competence** (competence is the ability of a

cell to alter its genetics by taking up extracellular (naked) DNA from its environment naturally or it may be induced in a laboratory) which is called **competent cell**.
Transformation in bacteria was first demonstrated in 1928 by the Frederick Griffith. It was discovered that a non-virulent strain of *Streptococcus pneumoniae* could be made virulent after being exposed to heat-killed virulent strains.
A few bacteria, such as *Neisseria gonorrhoeae, Neisseria meningitidis, Hemophilus influenzae, Legionella pneomophila, Streptococcus pneumoniae,* and *Helicobacter pylori* tend to be naturally competence.

However, many bacterial species can be made competent to take up DNA, provided that the cells are subjected to an appropriate chemical treatment (for example, treatment with CaCl2 or MgCl2) is called artificial competence.

Mechanism of Natural Competence

A cell that is naturally competent takes DNA into its cytoplasm by a protein-mediated process. First, the long molecule of double-stranded DNA is recognized by a receptor on the surface of the competent cell. A cell-surface endonuclease digests the DNA into small fragments. An exonuclease then degrades one strand of the DNA. The remaining single-stranded fragment is taken into the cytoplasm of the bacterium and displaces the corresponding region of the host chromosome by recombination (Figure 5).
In the case of artificially induced competence, the mechanism is quite different. Double-stranded DNA enters the cell through a cell wall.

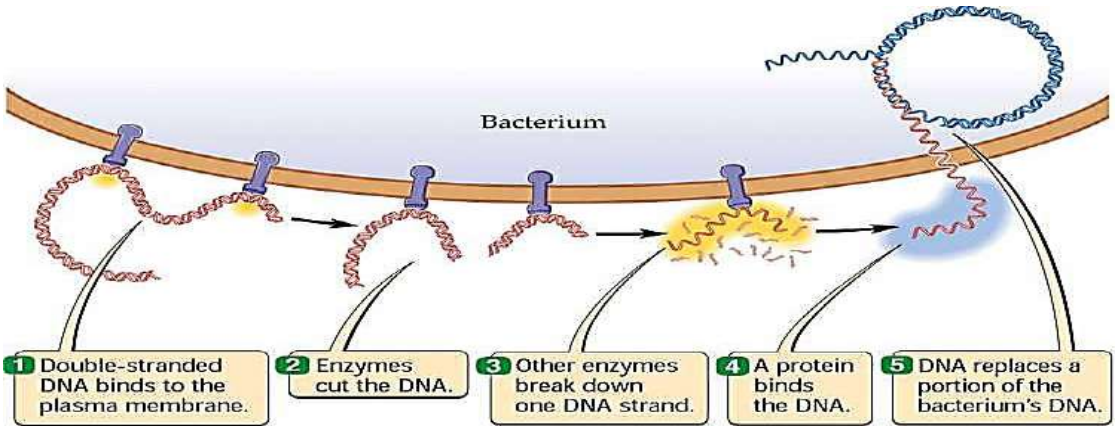


Figure 5: Natural transformation

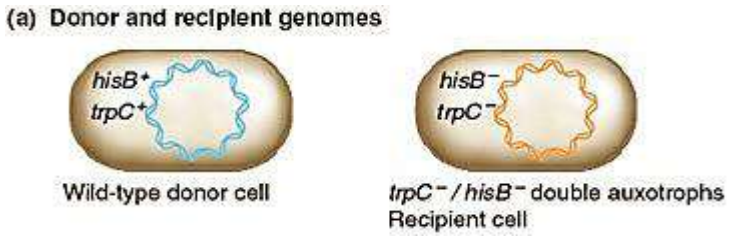
Natural transformation experiment

In one study of natural transformation, investigators isolated *B. subtilis* bacteria with two mutations *trpC* and *hisB* that made them Trp- , His- auxotrophs (s the inability of an organism to synthesize a particular organic compound required for its growth).

These double auxotrophs served as the recipient, wild type cells (Trp+, His+) were the donors. The recipient cells were grown in a suitable medium until became competent by starving the cell when growing in medium containing the limited amount of tryptophan and histidine. At the end of log phase, the bacteria became competent and will uptake the DNA added to the media. The DNA enters the cell and recombines with the homologues region of the bacterial chromosome.

After replication of chromosomes and divided of the cell to two daughter cells became transformants carrying the wild type genes or one of them, the results observed as following:

1. Media contain histidine only showed only the growth of Trp+ transformants colony.
2. Media contain tryptophan only showed only the growth of His+ transformants colony.
3. Media does not contain histidine and tryptophan showed the growth of Trp+ and His+ transformants colony. In this situation the two genes lie very close together on the chromosome and they are genetically linked. This is called co-transformation (simultaneous transformation of two or more genes).



Transduction

Transduction involves the transfer of a DNA fragment from one bacterium to another by a bacteriophage.

Transduction happens through either the lytic cycle or the lysogenic cycle. If the lysogenic cycle is adopted, the phage chromosome is integrated into the bacterial chromosome produced **prophage**, where it can stay dormant for thousands of generations. If the lysogen is induced (by UV light for example), the phage genome is excised from the bacterial chromosome and initiates the lytic cycle, which end in lysis of the cell and the release of phage particles. The lytic cycle leads to the production of new phage particles which are released by lysis of the host. There are two forms of transduction:

Generalized transduction

Generalized transduction is the process by which any bacterial DNA may be transferred to another bacterium via a bacteriophage. If bacteriophages undertake the

lytic cycle of infection upon entering a bacterium, the virus will take control of the cell's machinery for use in replicating its own viral DNA. If by chance bacterial chromosomal DNA is inserted into the viral capsid which is usually used to encapsulate the viral DNA.

The new virus capsule now loaded with part bacterial DNA continues to infect another bacterial cell. This bacterial material may become recombined into another bacterium upon infection.

Generalized transduction occurs in a variety of bacteria, including *Staphylococcus*, *Escherichia*, *Salmonella*, and *Pseudomonas*.

Specialized transduction

Specialized transduction is the process by which a restricted set of bacterial genes is transferred to another bacterium. This may occur occasionally during the lysogenic life cycle. The genes that get transferred (donor genes) depend on where the phage genome is located on the chromosome. Specialized transduction occurs when the prophage excises from the chromosome so that bacterial genes lying adjacent to the prophage are included in the excised DNA. The excised DNA is then packaged into a new virus particle, which then transfer the DNA to a new bacterium.

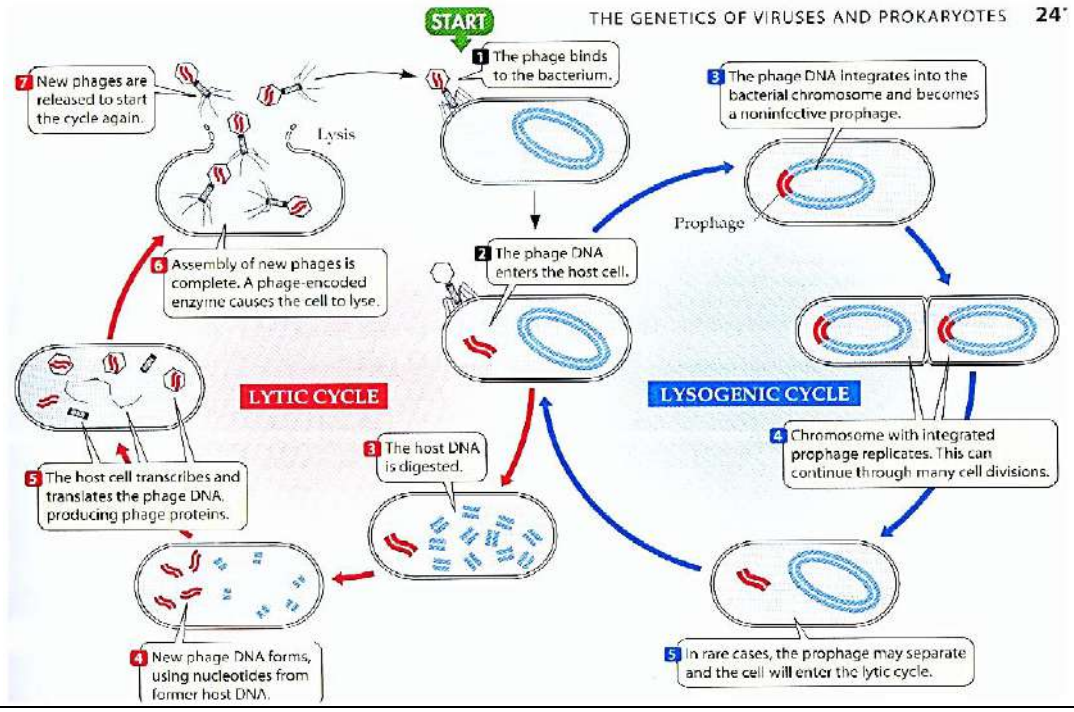


Figure 6: Generalized and Specialized Transduction