The models for DNA replication

There are three possible model for DND replication are suggested:

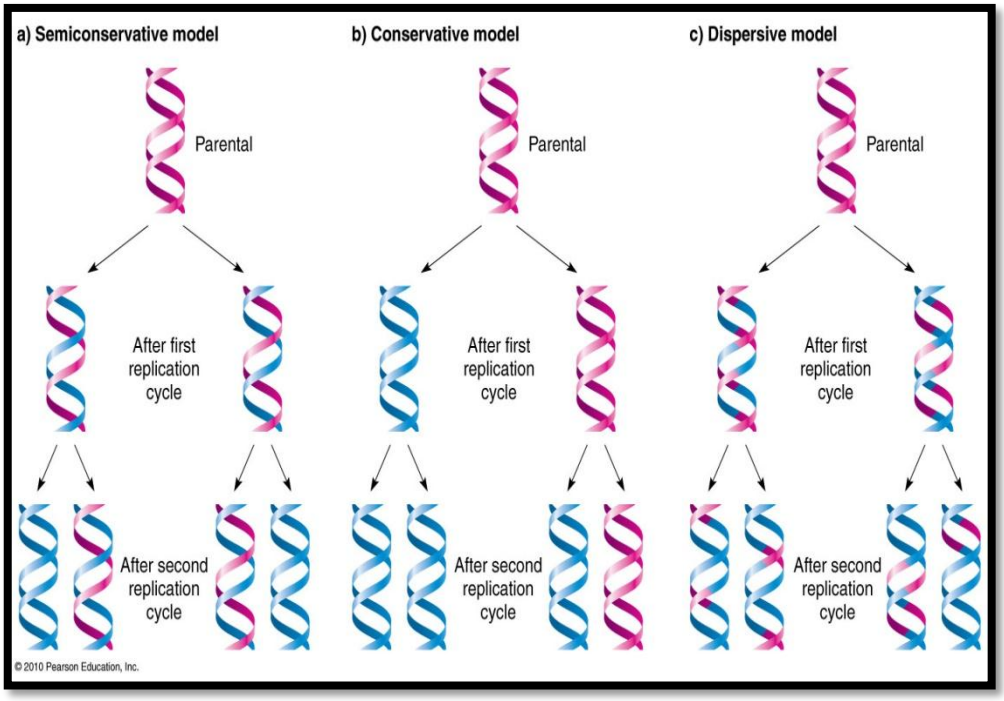
1-Semiconservative replication means that when the double stranded DNA helix was replicated, each of the two double stranded DNA helices consisted of one strand coming from the original helix and one newly synthesized. In the semiconservative hypothesis, proposed by Watson and Crick, the two strands of a DNA molecule separate during replication. Each strand then acts as a template for synthesis of a new strand.

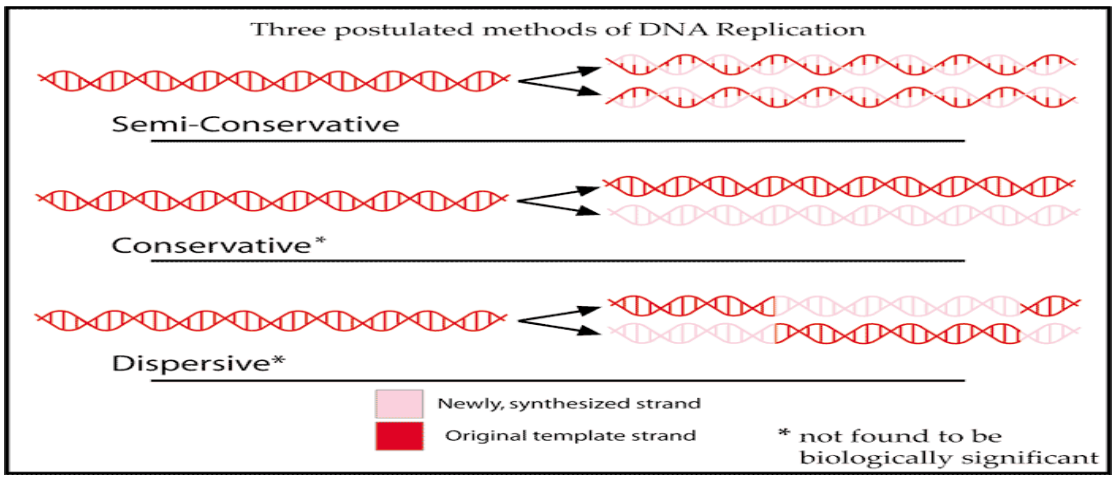
2-The conservative proposed that the entire DNA molecule acted as a template for synthesis of an entirely new one. According to this model, histone proteins bound to the DNA. The old strand is kept it self while new strands will formed from daughter strands.

3- The dispersive hypothesis is suggested by a model proposed by Max Delbrück, which attempts to solve the problem of unwinding the two strands of the double helix by a mechanism that breaks the DNA backbone every 10 nucleotides or so, uncoil the molecule, and attaches the old strand to the end of the newly synthesized one. This would synthesize the DNA in short pieces alternating from one strand to the other. In this model, the each strand of the daughter DNA molecules would be a combination of old and new DNA.

Each of these three models makes a different prediction about the distribution of the "old" DNA in molecules formed after replication.

In the conservative hypothesis, after replication, one molecule is the entirely conserved "old" molecule, and the other is all newly synthesized DNA. The semiconservative hypothesis predicts that each molecule after replication will contain one old and one new strand. The dispersive model predicts that each strand of each new molecule will contain a mixture of old and new DNA.





DNA replication

WHY REPLICATE DNA ?

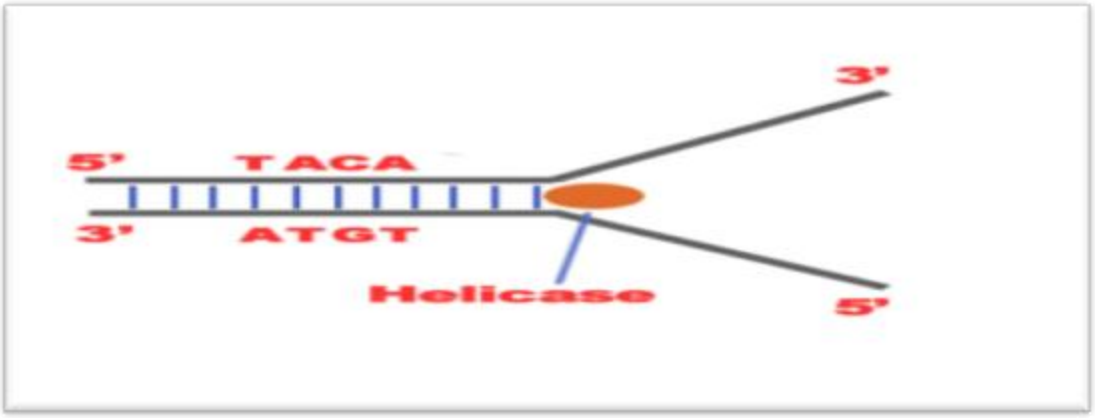
DNA is the genetic material that defines every cell. Before a cell duplicates and is divided into new daughter cells through either mitosis or meiosis, biomolecules and organelles must be copied to be distributed among the cells. DNA, found within the nucleus, must be replicated in order to ensure that each new cell receives the correct number of chromosomes. The process of DNA duplication is called DNA replication. Replication follows several steps that involve multiple proteins called replication enzymes and RNA. In eukaryotic cells, such as animal cells and plant cells, DNA replication occurs in the S phase of interphase during the cell cycle . The process of DNA replication is vital for cell growth, repair, and reproduction in organisms.

- **Steps of DNA Replication**

Step 1: Replication Fork Formation

Before DNA can be replicated, the double stranded molecule must be **"unwind"** into two single strands. DNA has four bases called adenine (A), thymine (T), cytosine (C) and guanine (G) that form pairs between the two strands. Adenine only pairs with thymine and cytosine only binds with guanine. In order to unwind DNA, these interactions between base pairs must be broken. This is performed by an enzyme known as DNA helicase. DNA helicase disrupts the hydrogen bonding between base pairs to separate the strands into a Y shape known as the replication fork. This area will be the template for replication to begin.

B

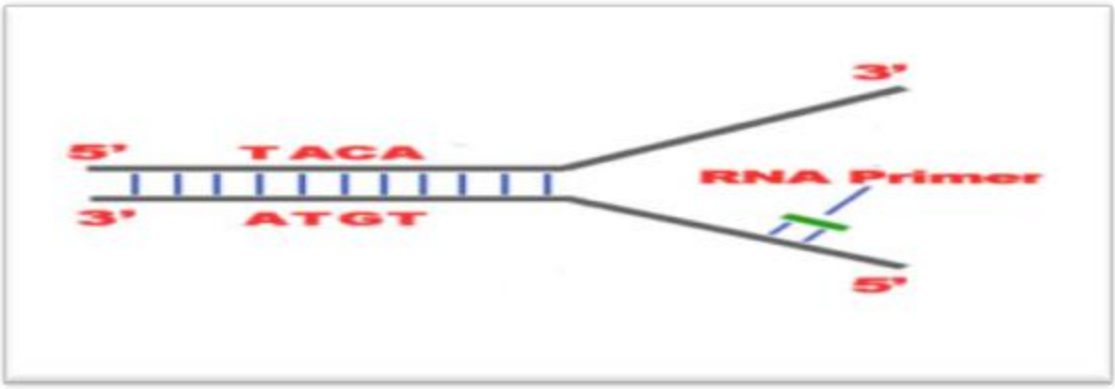


DNA is directional in both strands, signified by a 5' and 3' end. This notation signifies which side group is attached the DNA backbone. The 5' end has a phosphate (P) group attached, while the 3' end has a hydroxyl (OH) group attached. This directionality is important for replication as it only progresses in the 5' to 3' direction. However, the replication fork is bi-directional; one strand is oriented in the 3' to 5' direction (leading strand) while the other is oriented 5' to 3' (lagging strand). The two sides are therefore replicated with two different processes to accommodate the directional difference.

REPLICATION BEGINS

Step 2: Primer Binding

The leading strand is the simplest to replicate. Once the DNA strands have been separated, a short piece of RNA called a primer binds to the 3' end of the strand. The primer always binds as the starting point for replication. Primers are generated by the enzyme RNA primase.



Step 3: Elongation

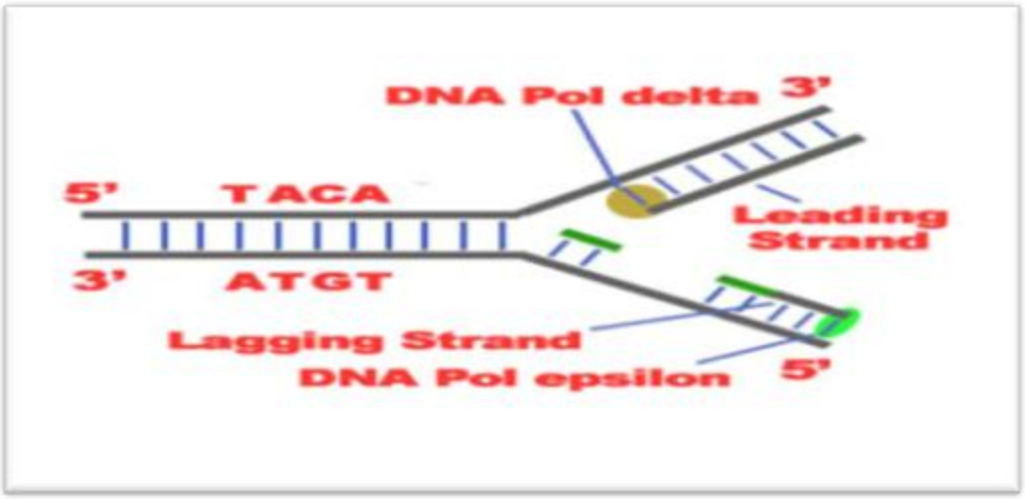
Enzymes known as DNA polymerases are responsible creating the new strand by a process called elongation. There are five different known types of DNA polymerases in bacteria and human cells.

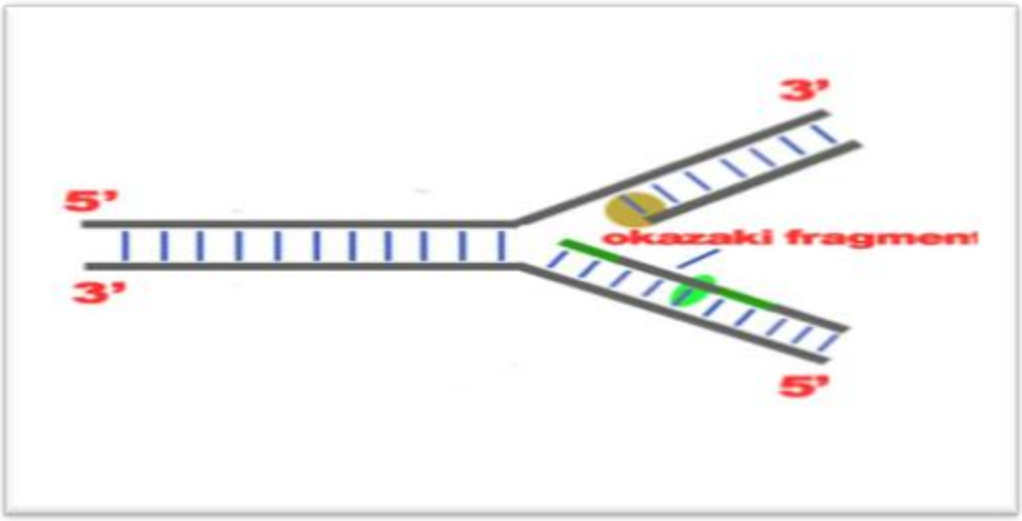
In eukaryotic cells , DNA polymerases alpha (α), delta(δ), and epsilon (ε) are the primary polymerases involved in DNA replication.

a) 5'-3' Template : The 3'-5' proceeding daughter strand -that uses a 5'-3' template- is called leading strand because DNA Polymerase ä can "read" the template and continuously adds nucleotides (complementary to the nucleotides of the template, for example Adenine opposite to Thymine etc).

Because replication proceeds in the 5' to 3' direction on the leading strand, the newly formed strand is continuous.

b) 3'-5'Template : The 3'-5' template cannot be "read" by DNA Polymerase ä . The replication of this template is complicated and the new strand is called lagging strand. The lagging strand begins replication by binding with multiple primers. Each primer is only several bases apart. DNA polymerase then adds pieces of DNA, called **Okazaki fragments** , to the strand between primers . This process of replication is discontinuous as the newly created fragments are disjointed.



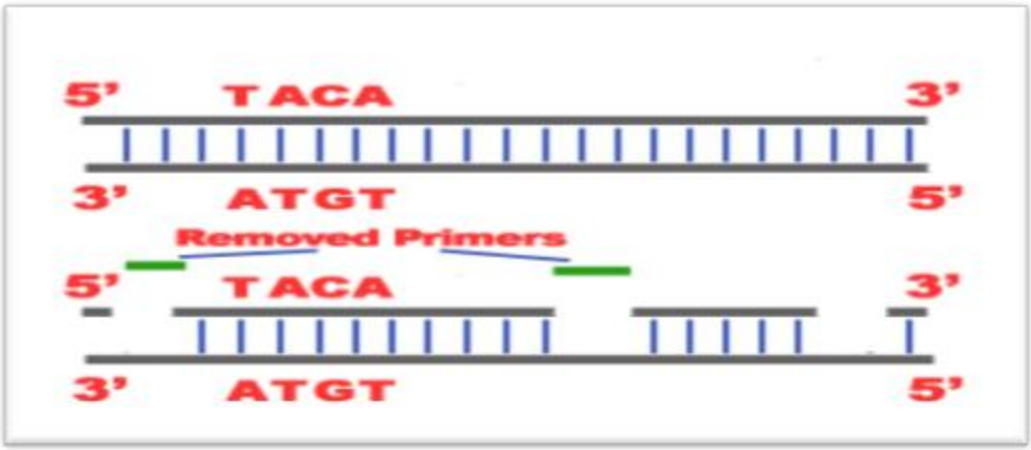


Step 4: Termination

Once both the continuous and discontinuous strands are formed, an enzyme called **exonuclease** removes all RNA primers from the original strands . These primers are then replaced with appropriate bases. Another **exonuclease** "**proofreads**" the newly formed DNA to check, remove and replace any errors. Another enzyme called **DNA ligase** joins Okazaki fragments together forming a single unified strand .

The ends of the parent strands consist of repeated DNA sequences called **telomeres**. Telomeres act as protective caps at the end of chromosomes to prevent nearby chromosomes from fusing. A special type of DNA polymerase enzyme called **telomerase** catalyzes the synthesis of telomere

sequences at the ends of the DNA . Once completed, the parent strand and its complementary DNA strand coils into the familiar double helix shape. In the end, replication produces two DNA molecules, each with one strand from the parent molecule and one new strand .



Replication Enzyme

This is an illustration of a DNA polymerase molecule. DNA polymerase is an enzyme that synthesizes DNA.

DNA replication would not occur without enzymes that catalyze various steps in the process. Enzymes that participate in the eukaryotic DNA replication process include:

1- DNA helicase - unwinds and separates double stranded DNA as it moves along the DNA. It forms the replication fork by breaking hydrogen bonds between nucleotide pairs in DNA .

2- RNA primase - a type of RNA polymerase that generates RNA primers. Primers are short RNA molecules that act as templates for the starting point of DNA replication.

3- DNA polymerases - synthesize new DNA molecules by adding nucleotides to leading and lagging DNA strands.

4-Topoisomerase or DNA Gyrase - unwinds and rewinds DNA strands to prevent the DNA from becoming tangled or supercoiled.

5- Exonucleases - group of enzymes that remove nucleotide bases from the end of a DNA chain.

6- DNA ligase - joins DNA fragments together by forming phosphodiester bonds between nucleotides.

(In bacteria such as E. coli, polymerase III is the main replication enzyme, while polymerase I, II, IV and V are responsible for error checking and repair. DNA polymerase III binds to the strand at the site of the primer and begins adding new base pairs complementary to the strand during replication) .

Gene expression and regulation

Gene expression

Gene expression is the process by which the genetic code - the nucleotide sequence of a gene is used to direct protein synthesis and produce the structures of the cell. Genes that code for amino acid sequences are known as 'structural genes'.

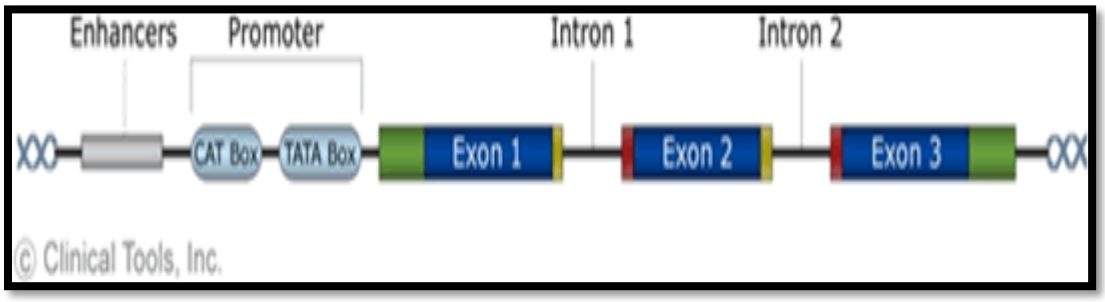
The process of gene expression involves two main stages:

1- Transcription: the production of messenger RNA (mRNA) by the **enzyme RNA polymerase**, and the processing of the resulting mRNA molecule

2- Translation: the use of mRNA to direct protein synthesis, and the subsequent post- translational processing of the protein molecule.

Some genes are responsible for the production of other forms of RNA that play a role in translation, including transfer RNA (tRNA) and ribosomal RNA (rRNA).

A structural gene involves a number of different components:



- **Exons**: Exons code for amino acids and collectively determine the amino acid sequence of the protein product. It is these portions of the gene that are represented in final mature mRNA molecule.
- **Introns**: Introns are portions of the gene that do not code for amino acids, and are removed (spliced) from the mRNA molecule before translation.

Control regions :

- **Start site**: A start site for transcription.
- **A promoter**: A region a few hundred nucleotides 'upstream' of the gene (toward the 5' end). It is not transcribed into mRNA, but plays a role in controlling the transcription of the gene.

Transcription factors bind to specific nucleotide sequences in the promoter region and assist in the binding of RNA polymerases.

- **Enhancers**: Some transcription factors (called activators) bind to regions called 'enhancers' that increase the rate of transcription. These sites may be thousands of nucleotides from the coding sequences or within an intron. Some enhancers are conditional and only work in the presence of other factors as well as transcription factors.
- **Silencers**. Some transcription factors (called repressors) bind to regions called 'silencers' that depress the rate of transcription.
- Note: The term 'gene expression' is sometimes used to refer to the transcription phase alone.

Transcription

Transcription is the process of RNA synthesis, controlled by the interaction of promoters and enhancers. Several different types of RNA are produced, including **messenger RNA (mRNA)**, which specifies the sequence of amino acids in the protein product, plus **transfer RNA (tRNA)** and **ribosomal RNA (rRNA)**, which play a role in the translation process.

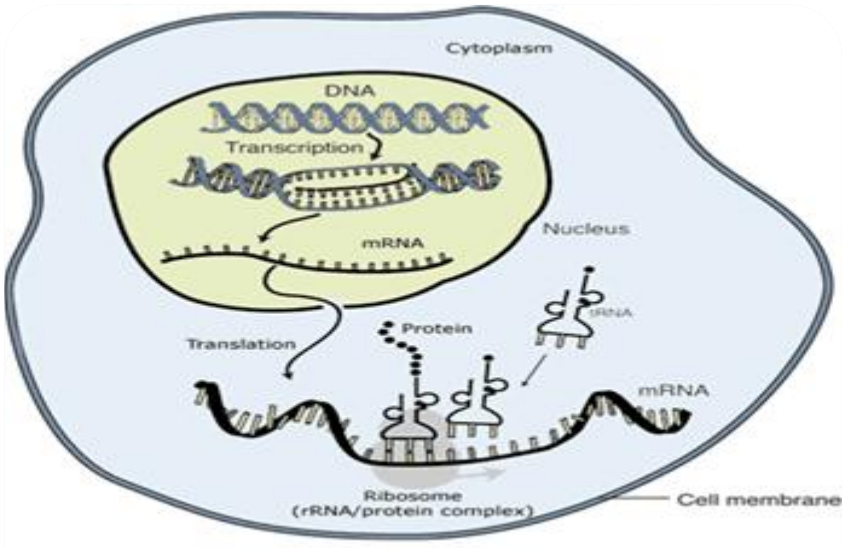


Image adapted from: National Human Genome Research Institute.

RNA polymerase:

The main enzyme involved in transcription is RNA polymerase, which uses a single-stranded DNA template to synthesize a complementary strand of RNA. Specifically, RNA polymerase builds an RNA strand in the 5' to 3' direction, adding each new nucleotide to the 3' end of the strand.
It synthesizes the RNA strand in the 5' to 3' direction, while reading the template DNA strand in the 3' to 5' direction.

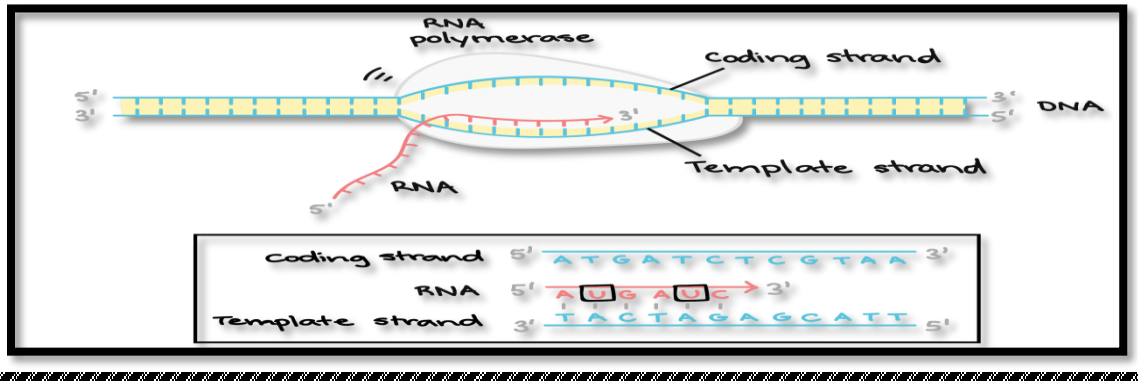
- **Types of RNA polymerase :**
  - ❖ Whereas a single RNA polymerase species synthesizes all RNAs in prokaryotes,
  - ❖ There are three different RNA polymerases in eukaryotic systems:

1- RNA polymerase I synthesizes rRNA.
2- RNA polymerase II synthesizes mRNA. In eukaryotes, the mRNA molecules always code for one protein, whereas in prokaryotes, many mRNAs code for several proteins.
3- RNA polymerase III synthesizes tRNAs as well as small nuclear and cellular RNA molecules.

Transcription involves four steps:

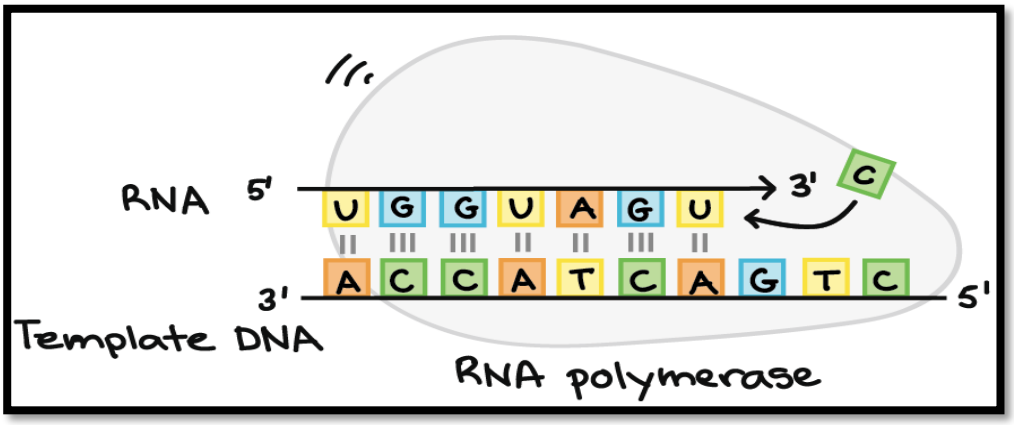
1. **Initiation**:

The DNA molecule unwinds and separates to form a small **open complex**. RNA polymerase binds to the promoter of the **template strand** (also known as the 'sense strand'). The synthesis of RNA proceeds in a 5' to 3' direction, so the template strand must be 3' to 5'.
Example:
coding strand: 5'-ATGATCTCGTAA-3'
Template strand: 3'-TACTAGAGCATT-5'
RNA transcript: 5'-AUGAUCUCGUAA-3'



2. **Elongation**:

RNA polymerase moves along the template strand, synthesising an mRNA molecule. In prokaryotes RNA polymerase is a **holoenzyme** consisting of a number of subunits, including a **sigma factor** (transcription factor) that recognises the promoter. In eukaryotes there are three RNA polymerases: I, II and III. The process includes a proofreading mechanism .



3 . **Termination**:

In prokaryotes there are two ways in which transcription is terminated.

-In **dependent termination**, a protein is responsible for disrupting the complex involving the template strand, RNA polymerase and RNA molecule.

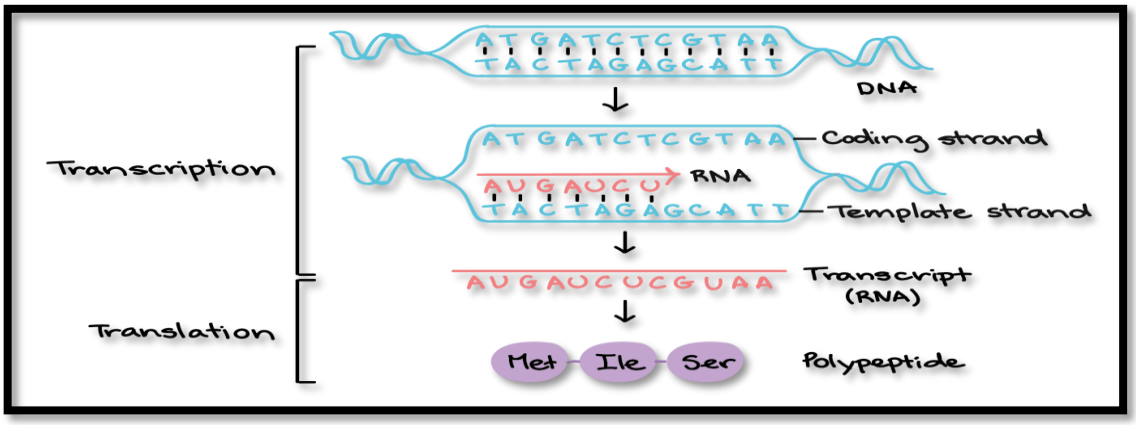
-In **independent termination**, a loop forms at the end of the RNA molecule, causing it to detach itself.

Termination in eukaryotes is more complicated, involving the addition of additional adenine nucleotides at the 3' of the RNA transcript (a process referred to as **polyadenylation**).

4 -Processing:

After transcription the RNA molecule is processed in a number of ways: introns are removed and the exons are spliced together to form a mature mRNA molecule consisting of a single protein-coding sequence.

- RNA synthesis involves the normal base pairing rules, but the base thymine is replaced with the base **uracil**.



- **Differences between RNA transcription and DNA replication:**

There are two fundamental differences between RNA transcription and DNA replication

1-RNA polymerase only makes one RNA strand during transcription, which **means that it copies only one DNA strand in a given gene. While in DNA replication both DNA strands are copied.**

-This makes transcription asymmetrical

-Replication is semi-conservative

2-In transcription ,DNA melting is limited and transient, however during the replication, the two parental DNA strands separated permanently.

Post-transcriptional Modification

Prokaryotic DNA transcription produces messenger RNA, which is necessary for transfer from the cell nucleus to the cytoplasm where translation occurs. In contrast, eukaryotic DNA transcription takes place in a cell's nucleus and produces what is called a primary RNA transcript or pre-messenger RNA. Before eukaryotic products of transcription can be moved into the cytoplasm, where undergo a number of levels of processing known as **post-transcriptional modifications**. In the case of messenger RNA, modification is necessary to convert pre-mRNA into a mature mRNA that is ready for protein translation.

Post-transcriptional modification or Co-transcriptional modification **is the process in eukaryotic cells where primary transcript RNA is converted into mature RNA.**

The process includes three major steps:
1- addition of a 5' cap,
2- addition of a 3' poly-adenylation tail,
3- splicing.

1) 5' Processing :
Capping:
In mRNA, the 5' end is modified by the addition of a 7-methylguanosine cap (a methylated guanine residue) (m7G). The cap is attached the 'wrong way round' forming a 5'-5' pyrophosphate bond, rather than the usual 3'-5' phosphodiester bond. **This has two key roles:**
it enables ribosomal recognition, and thus greater efficiency of protein translation, as well as protecting the mRNA from degradation by 5'-3' exonuclease enzymes.

Note that the 5' cap is the 'front' end of the mRNA and the 3' poly-A tail is the 'back' end. The coding sequence of the transcript lies in between.
After the 5' cap, but before the **initation AUG codon**, is a 5' untranslated region (UTR) or leader sequence. Equally, after the termination codon, but before the 3' untranslated region (UTR) or trailer sequence. These are not protein-coding but may have regulatory roles; for instance, the 3' UTR is known to act as a poly-adenylation signal for the addition of the poly-A tail by various enzymes.

2) 3' Processing :
Cleavage and polyadenylation :
In mRNA, the 3' end is modified by the addition of multiple adenine residues to produce a structure called a poly-adenylated (poly A) tail. The length of a poly-A tail is highly variable . The length of a poly-A tail determines the half-life of mRNA: a longer tail means a longer half-life .
The pre-mRNA processing at the 3' end of the RNA molecule involves cleavage of its 3' end and then the addition of about 250 adenine residues to form a poly(A) tail. The cleavage and adenylation reactions occur if a polyadenylation signal sequence (5'- AAUAAA-3') is located near the 3' end of the pre-mRNA molecule, which is followed by another sequence, which is usually (5'-CA-3') and is the site of cleavage. A GU-rich sequence is also usually present further downstream on the pre-mRNA molecule. After the synthesis of the sequence elements, two multisubunit proteins called cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CStF) are transferred from RNA Polymerase II to the RNA molecule.

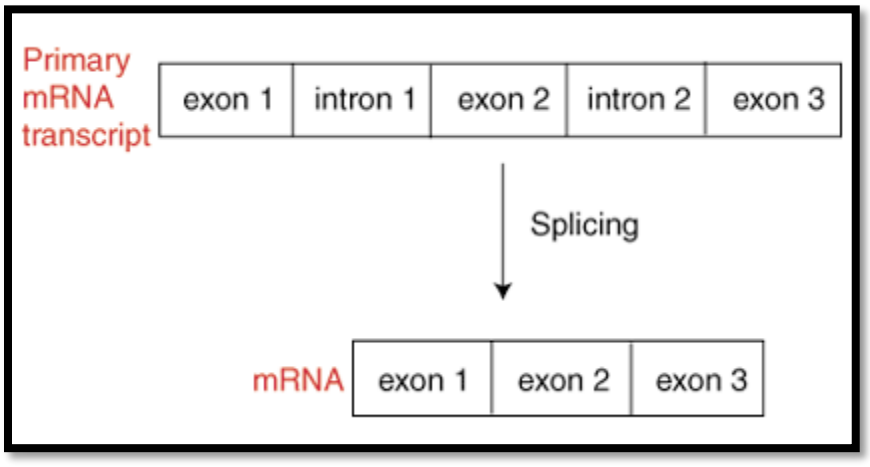
The two factors bind to the sequence elements. A protein complex forms and contains additional cleavage factors and the enzyme Polyadenylate Polymerase (PAP). This complex cleaves the RNA between the polyadenylation sequence and the GU-rich sequence at the cleavage site marked by the (5'-CA-3') sequences. Poly(A) polymerase then adds about 200 adenine units to the new 3' end of the RNA molecule using ATP as a precursor. As the poly(A) tail is synthesised, it binds multiple copies of poly(A) binding protein, which protects the 3'end from ribonuclease digestion.

3- RNA splicing:

RNA splicing is the process by which introns, regions of RNA that do not code for proteins, are removed from the pre-mRNA and the remaining exons connected to re-form a single continuous molecule.

Exons are sections of mRNA which become "expressed" or translated into a protein. They are the coding portions of a mRNA molecule. The splicing reaction is catalyzed by a large protein complex called the spliceosome assembled from proteins and small nuclear RNA molecules that recognize splice sites in the pre-mRNA sequence. Many pre-mRNAs, including those encoding antibodies, can be spliced in multiple ways to produce different mature mRNAs that encode different protein sequences. This process is known as alternative splicing, and allows production of a large variety of proteins from a limited amount of DNA .

Alternative splicing shuffles the composition of exons in a transcript, meaning that numerous different transcripts can be made from the same gene.



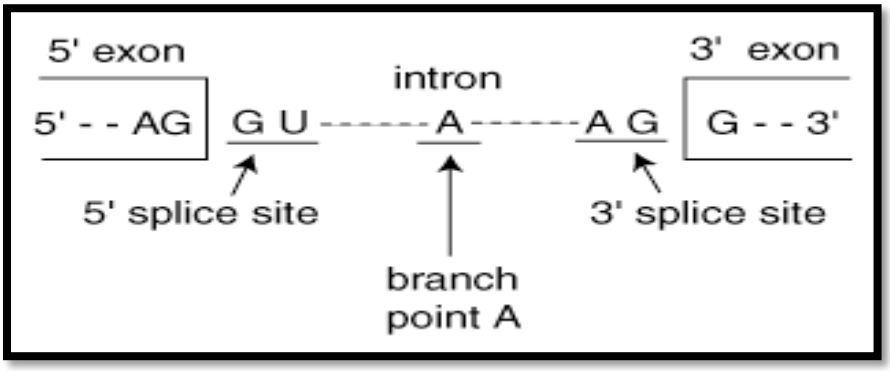
Unlike the sequence of an exon, intron sequences are unimportant. Only small portions of an intron sequence are preserved. These portions, located near the end of each intron, serve to identify a sequence as an intron, identifying the sequence for removal.

There are intron identifying portions:

1- The 5' splice site, consisting of a guanine next to a uracil base at the 5' end of the intron.

2-The 3' splice site, consisting of an adenine next to a guanine at the 3' end of an intron .

3-The branch point A, located about 30 nucleotides from the 3' end, consisting of just one adenine.



In addition to the post-transcriptional modifications already discussed (5' cap, poly A tail addition, and splicing), a fourth type of modification can be made: RNA editing. RNA editing is a modification that changes the mRNA sequence and as a result alters the protein produced by that mRNA. Editing can occur in two ways. First, by changing one nucleotide to another, and second by inserting or deleting a nucleotide or nucleotides.

Translation mRNA into protein & protein synthesis

Translation :

Translation involves "decoding" a messenger RNA (mRNA) and using its information to build a polypeptide ,or chain of amino acids . **Polypeptide is basic unit for the protein.**

In translation the mature mRNA molecule is used as a template to assemble a series of amino acids to produce a polypeptide. The complex in the cytoplasm at which this occurs is called a **ribosome**. Ribosomes are a mixture of ribosomal proteins and ribosomal RNA (rRNA), and consist of a large subunit and a small subunit

The genetic code

These relationships between mRNA codons and amino acids are known as the **genetic code** (See the full codon table) :

In an mRNA, the instructions for building a polypeptide come in groups of three nucleotides called **codons** . some key features of codons :

- There are 61 different codons for amino acids
- Three "stop" codons mark the polypeptide as finished
- One codon, AUG, is a "start" signal to kick off translation (it also specifies the amino acid methionine).

Second letter

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

Genetic code table. Each three-letter sequence of mRNA nucleotides corresponds to a specific amino acid, or to a stop codon. **UGA, UAA,and UAG are stop codons**. **AUG is the codon for methionine, and is also the start codon.**

Codons to amino acids:

In translation, the codons of an mRNA are read in order (from the 5' end to the 3' end) by molecules called **transfer RNAs, or tRNAs**.

Each tRNA has an anticodon , a set of three nucleotides that binds to a matching mRNA codon through base pairing. The other end of the tRNA carries **the amino acid** that's specified by the codon.

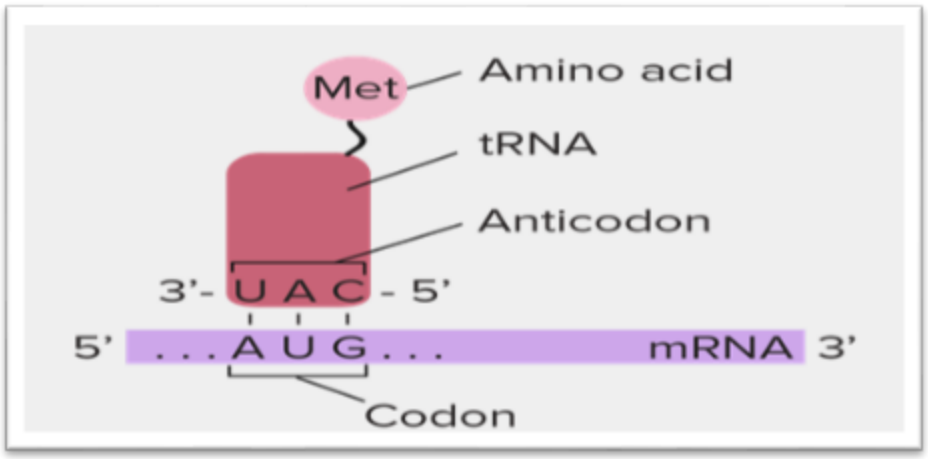
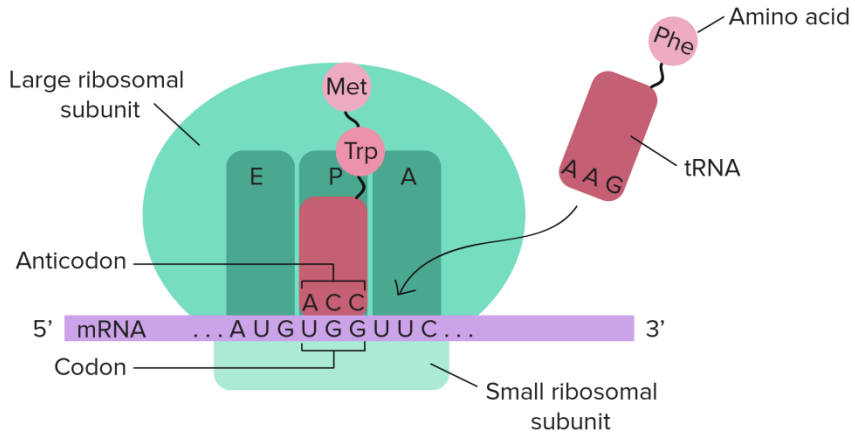


Image showing a tRNA acting as an adapter connecting an mRNA codon to an amino acid.

At one end, the tRNA has an anticodon of 3'-UAC-5', and it binds to a codon in an mRNA that has a sequence of 5'-AUG-3' through complementary base pairing. **The other end** of the tRNA carries the amino acid methionine (Met), which is the the amino acid specified by the mRNA codon AUG.

The ribosome provides where an mRNA can interact with tRNAs bearing amino acids. There are three places on the ribosome where tRNAs bind: the **A, P, and E site**. The **A site** accepts an incoming tRNA bound to an amino acid. The **P site** holds a tRNA that carries a growing polypeptide

(the first amino acid added is methionine (Met)). **The E site** is where a tRNA goes after it is empty, meaning that it has transferred its polypeptide to another tRNA (which now occupies the P site). In the diagram, the empty tRNA has already left the E site and is thus not shown.



Translation involves reading the mRNA nucleotides in groups of three, each of which specifies and amino acid (or provides a stop signal indicating that translation is finished).

The mRNA sequence is:

-'٣AUG AUC UCG UAA-5'

AUG → Methionine , AUC → Isoleucine , UCG → Serine and UAA → "Stop"

Polypeptide sequence:

(N-terminus) Methionine –Isoleucine - Serine (C-terminus).

Translation: has three parts:

initiation, elongation, and termination.

1- Initiation ("beginning"): in this stage, the ribosome gets together with the mRNA and the first tRNA so translation can begin.

In order for translation to start, we need a few key ingredients. These include:

- A ribosome (which comes in two pieces, large and small).
- An mRNA with instructions for the protein we'll build.
- An "initiator" tRNA carrying the first amino acid in the protein, which is almost always methionine (Met).

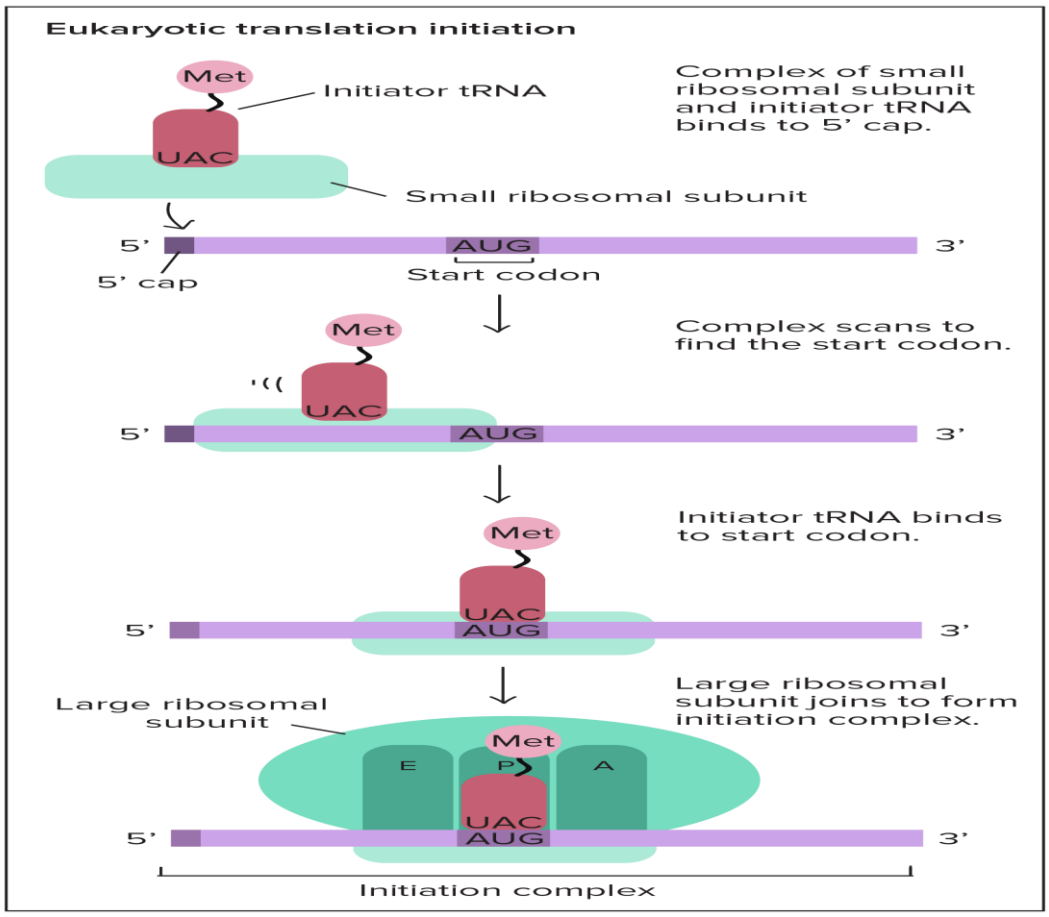
Inside your cells (and the cells of other eukaryotes), translation initiation goes like this:

1- first, the tRNA carrying methionine attaches to the small ribosomal subunit (complex).

2- Together, they bind to the 5' cap end of the mRNA (added during processing in the nucleus).

3- Complex scans from 5' to 3' to find the start codon (AUG). Initiator tRNA binds to start codon

4- Then, they "walk" along the mRNA in the 3' direction, stopping when they reach the start codon (often, but not always, the first AUG).



2- Elongation

Elongation ("middle"): in this stage, amino acids are brought to the ribosome by tRNAs and linked together to form a chain.

Our first, methionine-carrying tRNA starts out in the middle slot of the ribosome, called the P site. Next to it, a fresh codon is exposed in another slot, called the A site. The A site will be the "landing site" for the next tRNA, one whose anticodon is a perfect (complementary) match for the exposed codon.

In the first round of elongation, an incoming amino acid attaches to methionine already present in the ribosome's P site. This action initiates the growth of a polypeptide. The three steps of this first round of elongation are described below.

a- Codon recognition: an incoming tRNA with an anticodon that is complementary to the codon exposed in the A site binds to the mRNA.

b- Peptide bond formation: a peptide bond is formed between the incoming amino acid (carried by a tRNA in the A site) and methionine (a tRNA charged with methionine attached to the P site during initiation). This action passes the polypeptide (the two bonded amino acids) from the tRNA in the P site to the tRNA in the A site. The tRNA in the P site is now "empty" because it does not hold the polypeptide.

c- Translocation: the ribosome moves one codon over on the mRNA toward the 3' end. This shifts the tRNA in the A site to the P site, and the tRNA in the P site to the E site. The empty tRNA in the E site then exits the ribosome.

Once the matching tRNA has landed in the A site, it's time for the action: that is, the formation of the peptide bond that connects one amino acid to another. This step transfers the methionine from the first tRNA onto the amino acid of the second tRNA in the A site..

3- Termination

Termination ("end"): in the last stage, the finished polypeptide is released to go and do its job in the cell.

Translation ends in a process called termination. Termination happens when a stop codon in the mRNA (UAA, UAG, or UGA) enters the A site.

Stop codons are recognized by proteins called release factors, which fit neatly into the P site (though they aren't tRNAs). Release factors mess with the enzyme that normally forms peptide bonds: they make it add a water molecule to the last amino acid of the chain. This reaction separates the chain from the tRNA, and the newly made protein is released.

What next? Luckily, translation "equipment" is very reusable. After the small and large ribosomal subunits separate from the mRNA and from each other, each element can (and usually quickly does) take part in another round of translation.

Post-translational modification(PTM)

Most of the proteins that are translated from mRNA undergo chemical modifications before becoming functional in different body cells. The modifications collectively, are known as post-translational modifications. The protein post translational modifications **play a crucial role in generating the heterogeneity in proteins and also help in utilizing identical proteins for different cellular functions in different cell types.**

PTM refers to the generally enzymatic modification of proteins following protein biosynthesis. Proteins are synthesized by ribosomes translating mRNA into polypeptide chains, which may then undergo PTM to form the mature protein product. PTMs are important components in cell signaling.

Post-translational modifications can occur on the amino acid side chains or at the protein's C- or N- termini . They can extend the chemical repertoire of the 20 standard amino acids by modifying an existing functional group or introducing a new one such as phosphate.

Protein post translational modifications may happen in several ways. Some of them are listed below:

1- Phosphorylation:

principally on **serine, threonine or tyrosine residues ,**and it's the most common post-translational modification **.**

Phosphorylation is a very common mechanism for regulating the activity of enzymes ,and plays critical roles in the regulation of many cellular processes, including **cell cycle, growth and apoptosis.**

2- **glycosylation**:

Protein glycosylation is one of the major post-translational modifications, which can promote protein folding and improve stability as well as serving regulatory functions. Glycosylation encompasses a diverse selection of sugar-moiety additions to proteins that ranges from simple monosaccharide modifications of nuclear transcription factors to highly complex branched polysaccharide changes of cell surface receptors. **Glycosylation in proteins results in addition of a glycosyl group to either asparagines , serine, or threonine.**

3- **lipidation** :

Attachment of lipid molecules, known as **lipidation**, often targets a protein or part of a protein attached to the cell membrane and target proteins to membranes in organelles (endoplasmic reticulum [ER], Golgi apparatus, mitochondria), and vesicles (endosomes, lysosomes)

4- Proteolysis :

Other forms of post-translational modification consist of **cleaving peptide bonds**, as in processing a propeptide to a mature form or removing the initiator methionine residue.

5-**Acetylation**: the addition of an acetyl group, usually at the N-terminus of the protein.

6-Alkylation: The addition of an alkyl group (e.g. methyl, ethyl).

7- **Methylation**: The addition of a methyl group, usually at lysine or arginine residues. (This is a type of alkylation.)

8-Sulfation: The addition of a sulfate group to a tyrosine.

9- Glycylation: Covalent linkage of one to more than 40 glycine residues to the tubulin C-terminal tail of the amino acid sequence.

10-The formation of disulfide bonds from cysteine residues may also be referred to as a post-translational modification. For instance, the peptide hormone insulin is cut twice after disulfide bonds are formed, and a pro-peptide is removed from the middle of the chain; the resulting protein consists of two polypeptide chains connected by disulfide bonds.

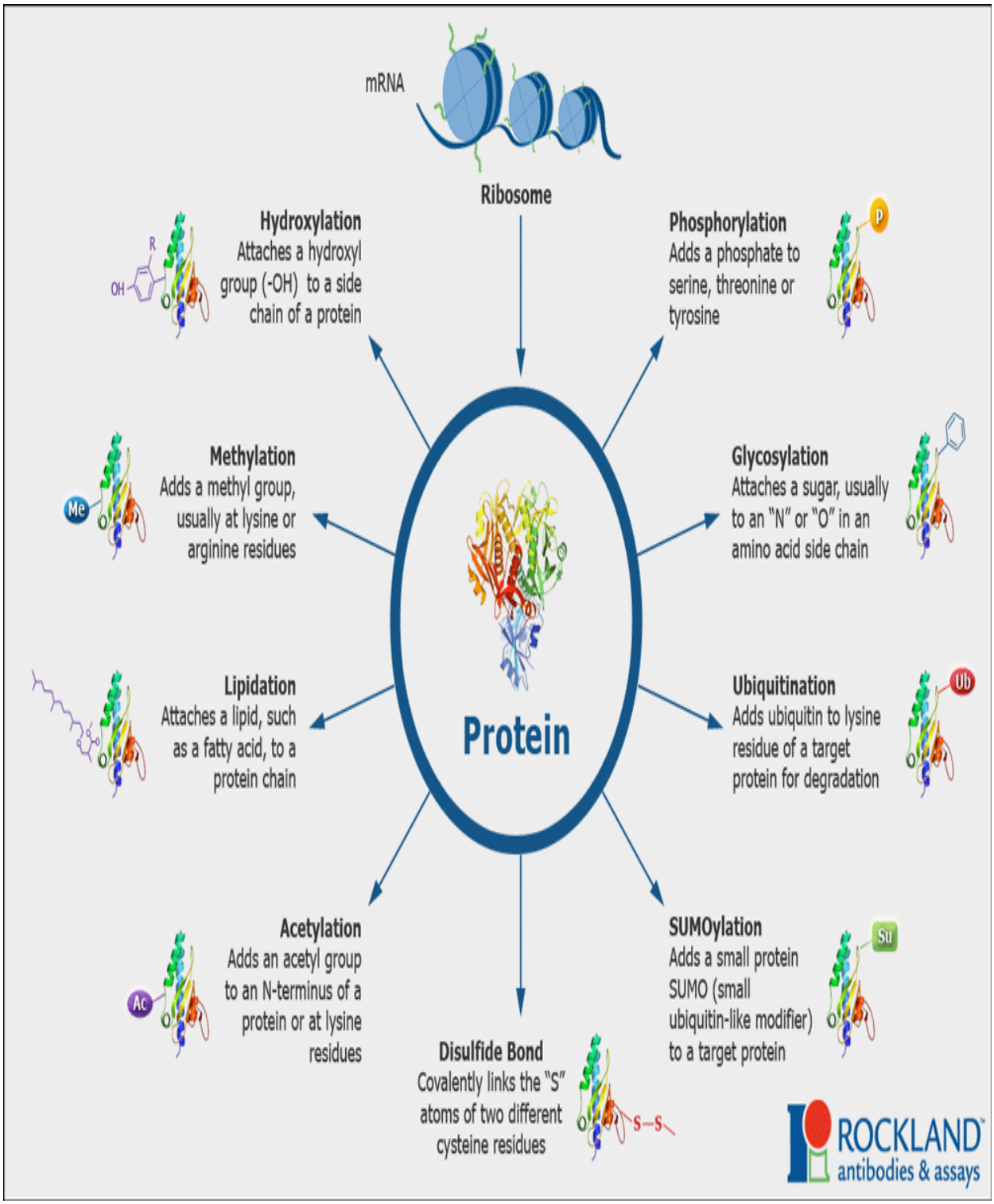
11-Carbonylation:

Some types of post-translational modification are consequences of oxidative stress. **Carbonylation** is one example that targets the modified protein for degradation and can result in the formation of protein aggregates. Specific amino acid modifications can be used as biomarkers indicating oxidative damage.

Sites that often undergo post-translational modification are those that have a functional group that can serve as **a nucleophile in the reaction**: the hydroxyl groups of serine, threonine, and tyrosine; the amine forms of lysine, arginine ,and histidine ;

the carboxylates of aspartate and glutamate; and the N- and C-termini. In addition, although the amide of asparagine is a weak nucleophile, it can serve as an attachment point for glycans.

Post-translational modification of proteins can be experimentally detected by a variety of techniques, including **mass spectrometry, Eastern blotting, and Western blotting.**



DNA Damage and Repair Mechanisms

Damage to cellular DNA is involved in mutagenesis and the development of cancer.

DNA repair is a collection of processes by which a cell identifies and corrects damage to the DNA molecules that encode its genome.

The DNA in a human cell undergoes several thousand to a million damaging events per day, generated by both **external (exogenous) and internal metabolic (endogenous) processes** . Changes to the cellular genome can generate errors in the transcription of DNA and ensuing translation into proteins necessary for signaling and cellular function. Genomic mutations can also be carried over into daughter generations of cells if the mutation is not repaired prior to mitosis.

As a consequence, the DNA repair process is constantly active as it responds to damage in the DNA structure. When normal repair processes fail, and when cellular apoptosis does not occur, irreparable DNA damage may occur, including double-strand breaks ,This can eventually lead to malignant tumors, or cancer.

The rate of DNA repair is dependent on many factors, including

1-The cell type,

2-The age of the cell,

3- The extracellular environment.

A cell that has accumulated a large amount of DNA damage, can enter one of three possible states:

1- The cell may become senescent, stopping mitosis and preventing the cell from evolving further.

2-The cell may become apoptotic. Sufficient DNA damage may trigger an apoptotic signaling cascade, forcing the cell into programmed cell death.

3- The cell may become malignant, i.e., develop immortal characteristics and begin uncontrolled division.

DNA damage:

DNA damage, due to environmental factors and normal metabolic processes inside the cell, occurs at a rate of 10,000 to 1,000,000 molecular lesions per cell per day. While this constitutes only 0.000165% of the human genome's approximately 6 billion bases (3 billion base pairs) .

The vast majority of DNA damage affects the primary structure of the double helix; that is, the bases themselves are chemically modified. These modifications can in turn disrupt the molecules' regular helical structure by introducing non-native chemical bonds that do not fit in the standard double helix. DNA is supercoiled and wound around "packaging" proteins called histones (in eukaryotes), and both superstructures are vulnerable to the effects of DNA damage.

Sources of damage :

DNA damage can be subdivided into two main types:

A) Endogenous damage such as

1- attack by reactive oxygen species produced from normal metabolic byproducts (spontaneous mutation), especially the process of oxidative deamination

2- also includes replication errors

B) Exogenous damage caused by external agents such as:

1- ultraviolet [UV 200–400 nm] radiation from the sun

2- other radiation frequencies, including x-rays and gamma rays.

3- hydrolysis or thermal disruption.

4- certain plant toxins.

5-human-made mutagenic chemicals, especially aromatic compounds that act as DNA intercalating agents.

6- Viruses .

The replication of damaged DNA before cell division can lead to the incorporation of wrong bases opposite damaged ones. Daughter cells that inherit these wrong bases carry mutations from which the original DNA sequence is unrecoverable.

Types of damage :

There are several types of damage to DNA due to endogenous cellular processes:

1- oxidation of bases , generation of DNA strand interruptions from reactive oxygen specie.

2- alkylation of bases (usually methylation), such as formation of 7-methylguanosine, 1-methyladenine, 6-O-Methylguanine .

3-hydrolysis of bases, such as deamination, depurination, and depyrimidination . Hydrolysis reactions can partially or completely cleave the nucleotide base from the DNA strand.

The chemical bond connecting a purine base (adenine or guanine) to the deoxyribosyl phosphate chain can spontaneously break in the process known as **depurination**. An estimated 10,000 depurination events occur per day in a mammalian cell.

Depyrimidination (loss of pyrimidine base from thymine or cytosine) also occurs, but at a rate 20 to 100-fold lower than depurination.

Deamination occurs within the cell with the loss of amine groups from adenine, guanine, and cytosine rings, resulting in uracil . DNA repair enzymes are able to recognize and correct these unnatural bases.

4-mismatch of bases, due to errors in DNA replication, in which the wrong DNA base is stitched into place in a newly forming DNA strand, or a DNA base is skipped over.

5-Monoadduct damage cause by change in single nitrogenous base of DNA .

Damage caused by exogenous agents comes in many forms. Some examples are:

1- **UV-B light** causes crosslinking between adjacent cytosine and thymine bases creating pyrimidine dimers. This is called direct DNA damage.

2- **UV-A light** creates mostly free radicals. The damage caused by free radicals is called indirect DNA damage.

3- **Ionizing radiation** such as that created by radioactive decay or in cosmic rays causes breaks in DNA strands. Intermediate-level ionizing radiation may induce irreparable DNA damage (leading to replicational and transcriptional errors needed for neoplasia or may trigger viral interactions) leading to pre-mature aging and cancer.

4- **Thermal disruption** at elevated temperature increases the rate of depurination (loss of purine bases from the DNA backbone) and single-strand breaks.

5- Industrial chemicals such as vinyl chloride and hydrogen peroxide, and environmental chemicals such as polycyclic aromatic hydrocarbons found in smoke.

-UV damage, alkylation/methylation, X-ray damage and oxidative damage are examples of **induced damage**.

-**Spontaneous damage** can include the loss of a base, deamination .

Mutation and DNA Repair

A **mutation:** is any heritable change in the genetic material, resulting in an alteration of the DNA sequence.

* **Mutations are usually considered as a change that alters gene function and thus the phenotype of the organism.**
* Some results in genetic diseases, but others have no physical effects.
* Some mutations consist of an alteration of the **number or structure of chromosomes** in a cell (abnormalities involve loss or gain of chromosomes or breakage and rejoining of chromatids).
* Other mutations can take place in **coding DNA** or in regulatory sequences **(single-gene mutation).**

❑ Types of mutation

Mutations, which can alter the coding properties of a DNA segment are of several types:

1- Base- pair substitutions

Convert or replaced one type of base pair into another.

* **G-C ⇨ A-T** and **A-T ⇨ G-C** changes are referred to as **transition mutations (replacement of a purine to pyrimidine base pair by a purine to pyrimidine base pair).**
* **T-A ⇨ G-C, G-C ⇨ C-G, A-T ⇨ T-A** are called **transversions (replacement of a pyrimidine - purine base pair by a purine - pyrimidine base pair).**

CATTCACCTGTACCA
GTAAGTGGACATGGT
Normal Sequence

Transition (T-A to C-G)	Transversion (T-A to G-C)
CATCCACCTGTACCA GTAGGTGGACATGGT	CATGCACCTGTACCA GTACGTGGACATGGT
➢ base pair substitutions Transition: pyrimidine- purine bas pair to pyrimidine- purine base pair	➢ base pair substitutions Transversion: pyrimidine- purine base pair to purine- pyrimidine base pair

- This type of mutation (**Base- pair substitutions**) can **change the codon** to that of another amino acid, thus altering the protein.
- In addition, such changes can also **create a stop codon** in which one base pair is replaced by another. This can result in a change in amino acid sequence.
- Although transitions are more common than transversions, both kinds of mutations occur as a **consequence of replication errors**, both can result from chemical damage to DNA, and both have been implicated as causative factors in inherited genetic disease and cancer.

There are two types of base- pair substitutions:

A. Missense mutations: which produce change in a single amino acid

B. Nonsense mutations: which produce one of the three stop codons (**UAA, UAG, or UGA**) in the mRNA.

* Because the stop codons terminate the translation of mRNA, nonsense mutation results in a premature termination of the polypeptide chain.
* Conversely, if a stop codon is altered so that it encodes an amino acid, an abnormally elongated polypeptide is produced.

2- Small insertions / deletions

comprise a second common class of mutation.

* Genetic changes involve insertion or loss of a small number of base pairs **(one to several hundred).**
* These mutations can results in **extra** or **missing** amino acids in a protein.

- **An example** of such mutation is the 3 bp deletion that causes **cystic fibrosis.**

CATTCACCTGTACCA
GTAAGTGGACATGGT
Normal Sequence

Insertion
- CATGTCACCTGTACCA
- GTACAGTGGACATGGT

Deletion
- CATCACCTGTACCA
- GTAGTGGACATGGT

➢ deletions and insertions can involve one or more base pairs

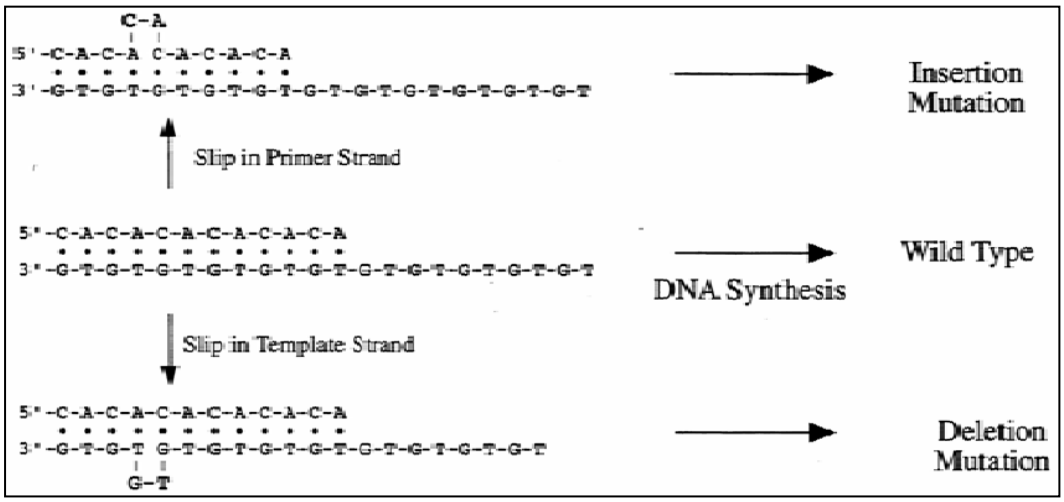
2.1: Frameshift mutation:

Another type of deletion and insertion tend to be especially harmful when the number of **missing or extra base pairs is not a multiple of three. Because codons consist of groups of 3bp,** such insertions or deletions can alter all the downstream codons, and produced truncated polypeptide.

2.2: Promoter mutation:

The affinity of RNA polymerase to bind a promoter site is decrease, resulting in reduced production of mRNA, and final result is decreased production of a protein. Also mutations of **transcription factor genes or enhancer sequences** have the similar effects.

* Repetitive runs of a mono, di-, or trinucleotide sequence are extremely prone to insertion /deletion mutation, an effect that has been attributed to **slippage of template and primer strands during DNA replication**



2.3: Splice site mutation:

Mutation interfere with the splicing of introns as a mature mRNA is formed from the primary mRNA transcript, and alter the splicing signal that is necessary for proper excision of an intron.

* **Mutations can be also caused by expanded tandem repeats DNA sequences (satellite)** which are prone to insertion /deletion as a result of unequal crossover or unequal sister chromatid exchanges

So the principle types of mutations are:

Missens, Nonsense, Frameshift, Promoter, Splice site

❑ Causes of Mutation

* **Mutations** can be induced in our DNA by exposure to a variety of mutagens occurring in our external environment or to mutagens generated in the intracellular environment.
* A large number of agents known to cause **induced mutations.** These mutations due to known environmental causes are called **spontaneous mutations,** which are naturally arising during the process of DNA replication and repair (*endogenous)*.
* **Mutagens:** Agents that cause induced mutations

Types of Mutagens

1. Radiation:

* **Ionizing radiation,** produced by X- rays and nuclear fall-out, can eject electrons from atoms, forming electrically charged ions.
  When these ions are situated within or near the DNA molecule, they can promote chemical reactions that **change DNA bases, and can also break the bonds of the double-stranded DNA.**
* **Nonionizing radiation:** produced by UV radiation which occurs naturally in sunlight.
  It does not form charged ions but **can move electrons from inner to outer orbits within an atom, and the atom becomes chemically unstable**.
* UV radiation causes **the formation of covalent bonds between adjacent pyrimidine base,** and are unable to pair properly with purines during DNA replication, this results in a base pair substitution
* **Sunlight is particularly damaging to DNA**. This figure shows the formation of a thymine dimer, catalyzed by UV light. The thymine dimer bridges two adjacent thymine residues on the same DNA strand. Dimer is (a molecule having two subunits).
* Because UV radiation is absorbed by the epidermis, it does not reach the germ line but can cause skin cancer.

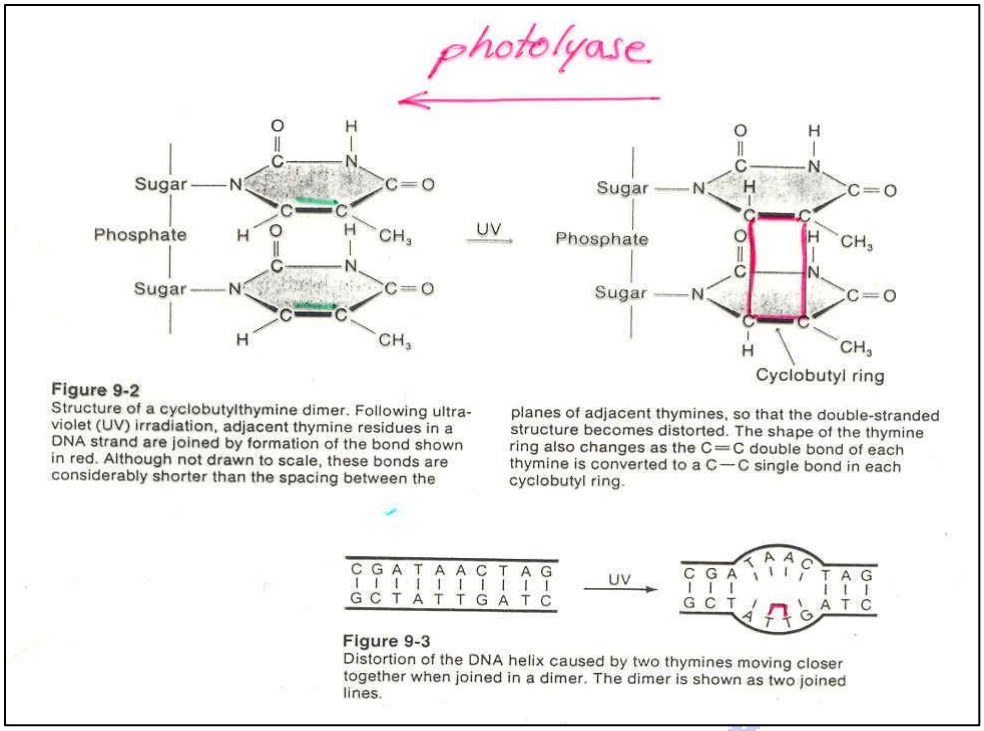
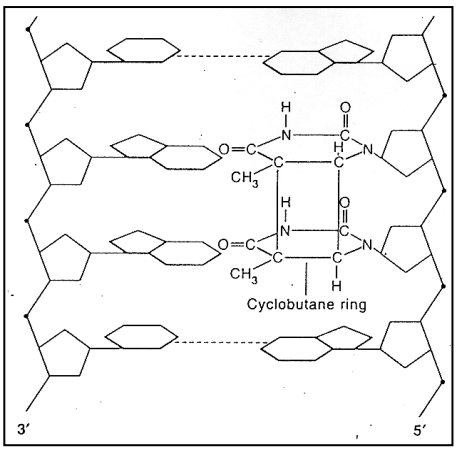


Figure 9-2
Structure of a cyclobutylthymine dimer. Following ultraviolet (UV) irradiation, adjacent thymine residues in a DNA strand are joined by formation of the bond shown in red. Although not drawn to scale, these bonds are considerably shorter than the spacing between the planes of adjacent thymines, so that the double-stranded structure becomes distorted. The shape of the thymine ring also changes as the C=C double bond of each thymine is converted to a C—C single bond in each cyclobutyl ring.

Figure 9-3
Distortion of the DNA helix caused by two thymines moving closer together when joined in a dimer. The dimer is shown as two joined lines.



2. Chemicals: Because of their chemical similarity to DNA bases (base analogs).

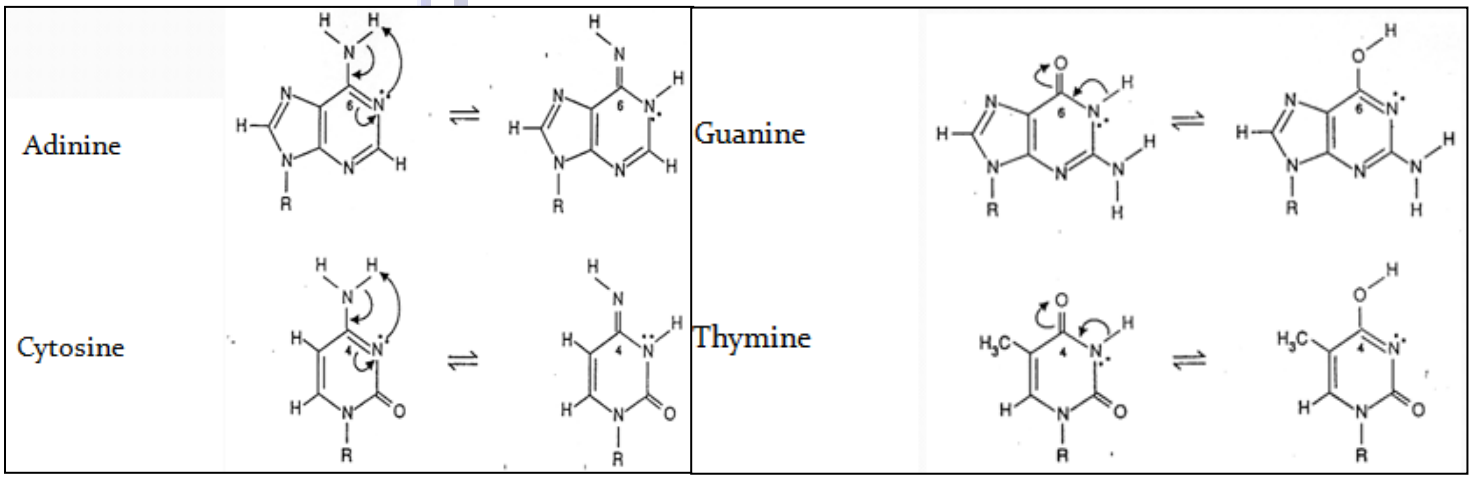
* **Bromouracil:** can be substituted for a true DNA base during replication.

* **Acridin dyes:** can physically insert themselves between existing bases causing frameshift mutation.

* **Nitrous acid:** removes an **amino group from cytosine, converting it to uracil,** and pairs with adenine instead of guanine, as the original cytosine would have done. The end result is a base pair substitution.

3. Spontaneous mutations (those that result from no external cause)

* Spontaneous mutations can occur by rearrangement of bonds and by the repositioning of hydrogens in the purine and pyrimidine bases



Mechanisms of DNA Repair

1. Mutations that occur during DNA replication are repaired when possible by proofreading by the DNA polymerases

2. Mutations that are not repaired by proofreading are repaired by mismatch (post-replication) repair followed by excision repair

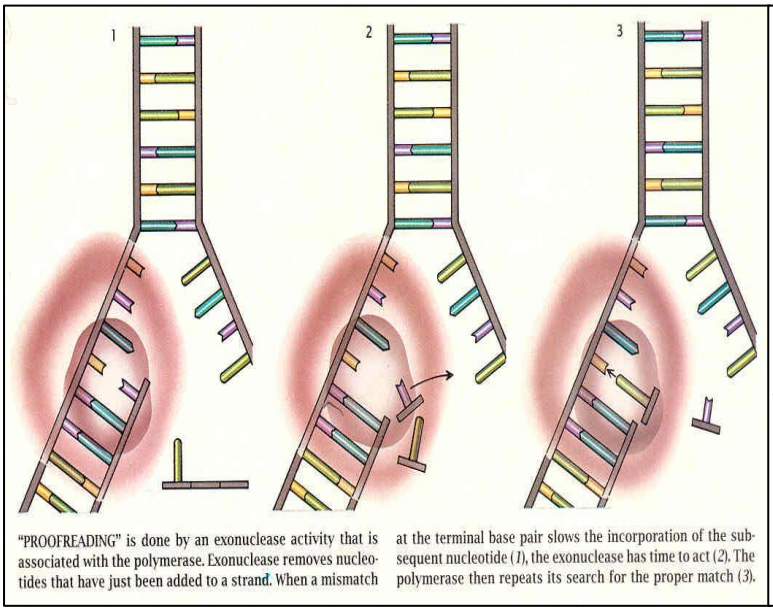
3. Mutations that occur spontaneously any time are repaired by excision repair (base excision or nucleotide excision)

* For DNA to be repaired properly following the **mismatch** incorporation of a nucleotide into the newly synthesized DNA strand, the replication machinery must have a means by which to distinguish between the "old" (template) strand and the "new" (daughter).
* Several dozen enzymes are involved in the repair of damaged DNA. They collectively recognize the altered base, excise it by cutting the DNA strand, replace it with the correct base and reseal the DNA.

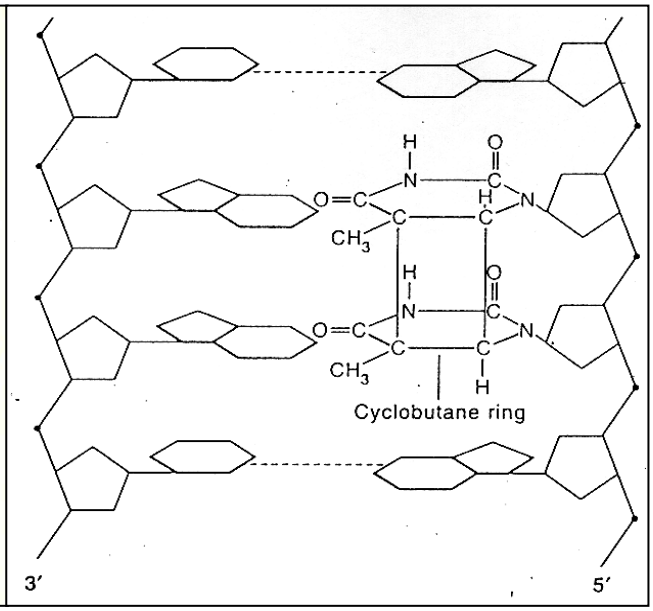
* **There are two types of excision repair:**
- base excision repair and
- nucleotide excision repair .

* Defects in DNA repair system can lead to many types of disease. For example, inherited mutation in genes responsible for **DNA mismatch repair** result in the persistence of cells with replication errors (mismatches) and lead to some kinds of cancers.
* Fail to repair double- stranded DNA breaks can lead specifically to ovarian and /or breast cancer.

* **Nucleotide excision repair** occurs when the DNA lesion is larger, for example when there is **a thymine dimer**.
* In this case, a special repair exconuclease removes about 30 nucleotides, including the lesion. The DNA is then resynthesized and ligated together as with base excision repair.
* **Defect in excision repair** lead to a number of diseases like **xeroderma pigmentosum**



"PROOFREADING" is done by an exonuclease activity that is associated with the polymerase. Exonuclease removes nucleotides that have just been added to a strand. When a mismatch at the terminal base pair slows the incorporation of the subsequent nucleotide (*1*), the exonuclease has time to act (*2*). The polymerase then repeats its search for the proper match (*3*).



xeroderma pigmentosum

* **Xeroderma pigmentosum:** A genetic disease characterized by such extraordinary sensitivity to sunlight that it results in the development of skin cancer at a very early age. Children with xeroderma pigmentosum (XP) can only play outdoors safely after nightfall. They have been called midnight children.
* **Xeroderma pigmentosum (XP)** occurs due to failed to remove pyrimidine dimers in the DNA of the skin cells after exposure to UV radiation, and the resulting replication errors lead to **base pair substitution** in skin cells.
* **XP** usually begins within the first 10 years of life. The skin is dry and scaly (xeroderma), and extensive freckling and abnormal skin pigmentation (pigmentosum) are followed by numerous skin tumors. Patients with XP can develop sever malignancies that some times result in death before 20 years of age.

Mutation Rates

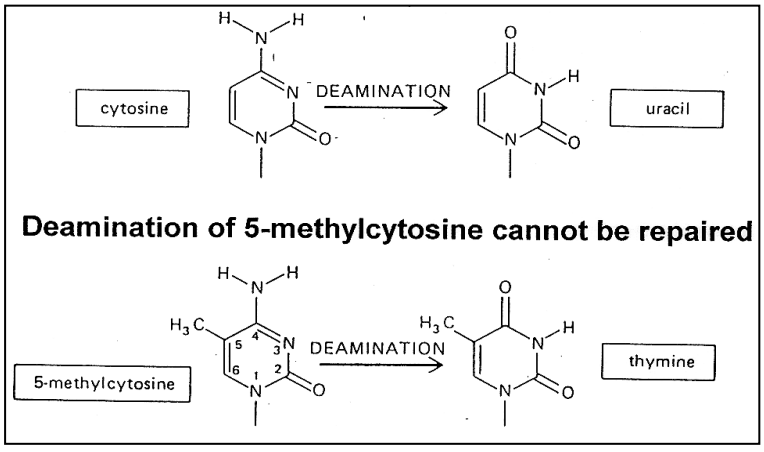
At the level of the gene, the mutation rate ranging **from 10^{-4} to 10^{-7} per locus per cell division.**

There are at least two reasons for the large range of variation:

1. Gene size: larger genes present large **"targets"** for mutation than do smaller genes.

2. Certain nucleotide sequences are susceptible to mutation**.** These are called **mutation hot spots**. The best known example is the two- base (dinucleotide) sequence **CG**. In mammals about 80% of **CG dinucleotides** are **methylated**: a methyl group is attached to the cytosine base. The methylated cytosine, 5-methylcytosine easily loses an amino group, **converting it to thymine**, resulting mutation from cytosine to thymine.
The problem that arises from these methylations is that deamination of **a 5 methyl cytosine** results in the production of **thymine**, which is not foreign to DNA. Thus, while a base pair mismatch is seen in the DNA by the repair machinery, it does not know which of the two strands to repair.

3. Mutation rates also affected with the age of the parent



Molecular consequences of Mutation:

Mutations can produce either:

***** a gain of function or ***** loss of function of the protein product

*** Gain- of- function** result in over expression or inappropriate expression of the product (in the wrong tissue or in the wrong stage of development).
Gain- of- function mutations produce **dominant disorders.**

*** Dominant negative mutations,** result in a protein product that is not only nonfunctional but also inhibits the function of the protein produced by the normal allele

*** Loss- of- function mutations are seen in recessive** diseases. In such diseases, the mutation results in the loss of 50% of the protein product (e.g. metabolic enzymes), but the 50% that remains is sufficient for normal function.

*** Diseases involving haplo insufficiency:** in which **50% of the gene product is insufficient** for normal function and **dominant disorder** can result.

E.g. **familial hypercholesterolemia**, in this disease, a single copy of mutation reduces the number of low-density lipoprotein (LDL) receptors by 50%, so cholesterol levels are double than normal resulting in increase the risk of heart disease.

Molecular biology

Definition and introduction:

Definition:

What is molecular biology?

Molecular biology is the study of living things at the level of the molecules which control them and make them up.

Molecular biology is a specialized branch of biochemistry, the study of the chemistry of molecules which are specifically connected to living processes. Of particular importance **to molecular biology are the nucleic acids (DNA and RNA) and the proteins which are constructed using the genetic instructions encoded in those molecules. Other bio-molecules, such as carbohydrates and lipids may also be studied for the interactions they have with nucleic acids and proteins** . Molecular biology is often separated from the field of cell biology, which concentrates on cellular structures (organelles and the like), molecular pathways within cells and cell life cycles.

The concepts of molecular biology included:

- The attempt to understand biological phenomena in molecular terms
- The study of gene structure and its function at the molecular level
- Molecular biology is a melding of aspects of genetics and biochemistry.
- Field of science concerned with the chemical structures and processes of biological at the molecular level.

Molecular biology having developed out of the related fields of biochemistry, genetics, and biophysics, the discipline is particularly concerned with the study of proteins, nucleic acids, and enzymes

• The term molecular biology was coined in 1938 by Warren Weaver .

Introduction:

Molecular Biology is an old science. It started with the evolution of life. Soon after the cell was recognized as the basic unit of life and its structure was discovered, it became clear that the cellular constituents are nothing but chemical compounds.

some of these compounds are simple chemicals **such as water, NaCl and many other small molecules, certain other molecules are very large having complex chemical structure . Nucleic acids, proteins, complex carbohydrates and lipids are some such molecules.**

These **complex molecules** often have very large molecular weight and are relatively difficult to be synthesized by chemical means. These are there- fore referred as biomacromolecules . These are synthesized in the cell by a number of complex reactions catalyzed by enzymes, the biological catalysts.

The biochemical pathways are highly regulated chemical reactions taking place in the cell . Further, at any given time thousands of different reactions take place simultaneously in a cell. There is a highly regulated coordination between different cellular reactions .

A number of regulatory molecules (usually protein in nature) regulate various reactions. These biochemical reactions can be influenced by both endogeneous factors such as hormones, metabolic status of the cell etc and by exogeneous factors such as the environmental changes.

Modern molecular biology started with the discovery of DNA by Friedrich Meischer in 1869. A series of experiments finally proved **that DNA is the genetic material that transfers the genetic information from one generation to another**. These included the classical experiments by Frederic Griffin (1928), Oswald Avery, Colin McLeod and Maclyn McCarty (1943) and the famous double labeling experiment of Alfred Hershey and Martha Chase (1952) **that proved the role of DNA as carrier of genetic information without any doubt .**

Eukaryotic Cell Structure

A typical eukaryotic cell is 10 μ in diameter, making its volume about 1,000 times that of a bacterial cell. Like bacteria, eukaryotic cells contain cell membranes، cytoplasmic proteins, DNA, and ribosomes. Eukaryotic cells, however, possess many structural features that even more clearly distinguish them from prokaryotic cells. Within the eukaryotic cytoplasm are a number of structural proteins that form networks. Microtubules, actin, intermediate filaments, and thin filaments form, four main categories of fibers found within eukaryotic cells .

Fibers within the cell provide a rigid structural skeleton, participate in vesicle and chromosome movement, and participate in changing the cell shape so that it can move. The DNA of eukaryotic cells does not freely mix with the cytoplasm, but is confined within a nuclear membrane. Normally only small proteins of molecular weight less than 20 to 40,000 can freely enter the nucleus through the nuclear membrane. Larger proteins and nuclear RNAs enter the nucleus through special nuclear pores. These are large structures that actively transport proteins or RNAs into or out of the nucleus.

The DNA itself is tightly complexed with a class of proteins called histones, whose main function appears to be to help DNA retain a condensed state.

Eukaryotic cells also contain specialized organelles such as mitochondria, which perform oxidative phosphorylation to generate the cell's needed chemical energy. In many respects mitochondria resemble bacteria and, in fact, appear to have evolved from bacteria.

Mitochondria contains DNA, usually in the form of a circular chromosome like that of E. coli and ribosomes that often more closely resemble those found in bacteria than the ribosomes located in the cytoplasm of the eukaryotic cell .

Chloroplasts carry out photosynthesis in plant cells, and are another type of specialized organelle found within some eukaryotic cells. **Like mitochondria,**

chloroplasts also contain DNA and ribosomes different from the analogous structures located elsewhere in the cell.

Most eukaryotic cells also contain internal membranes. The nucleus is surrounded by two membranes. The endoplasmic reticulum is another membrane found in eukaryotic cells. The Golgi apparatus is another structure containing membranes.

Nucleic acids

- Molecular biology helps in understanding genetic mutations that can cause certain types of diseases.
- Molecular biology is primarily concerned with the **inter-relationship between the formation macromolecules DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) and how these molecules are used to synthesize polypeptides , the basic components of all proteins.**

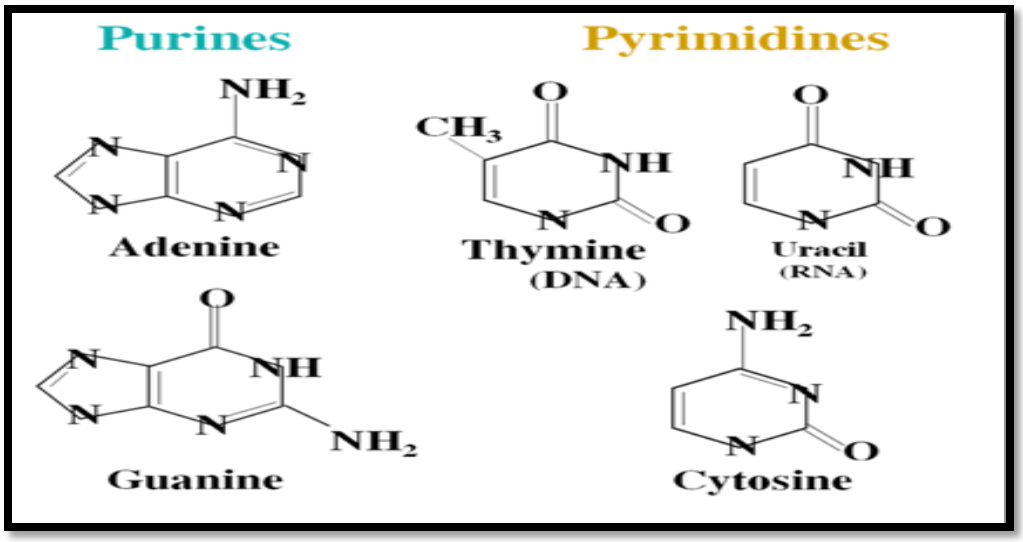
Nucleic acids

Nucleic acids play an important role in the storage and expression of genetic information. They are divided into two major classes: **deoxyribonucleic acid (DNA)** functions solely in information storage, while **ribonucleic acids (RNAs)** are involved in most steps of gene expression and protein biosynthesis. All nucleic acids are made up from **nucleotide** components, which consist of a nitrogenous base, a sugar and a phosphate residue . **A sugar with an attached base is called a nucleoside. A nucleoside with a phosphate group attached at carbon atom 5′ or 3′ constitute a nucleotide , the basic repeat unit of a DNA strand.**

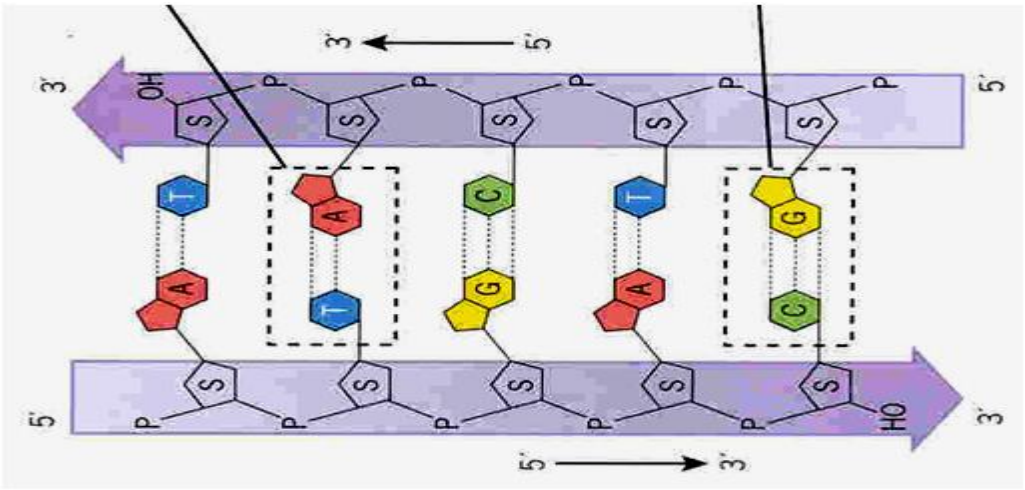
The nitrogenous bases are classified into two types:-

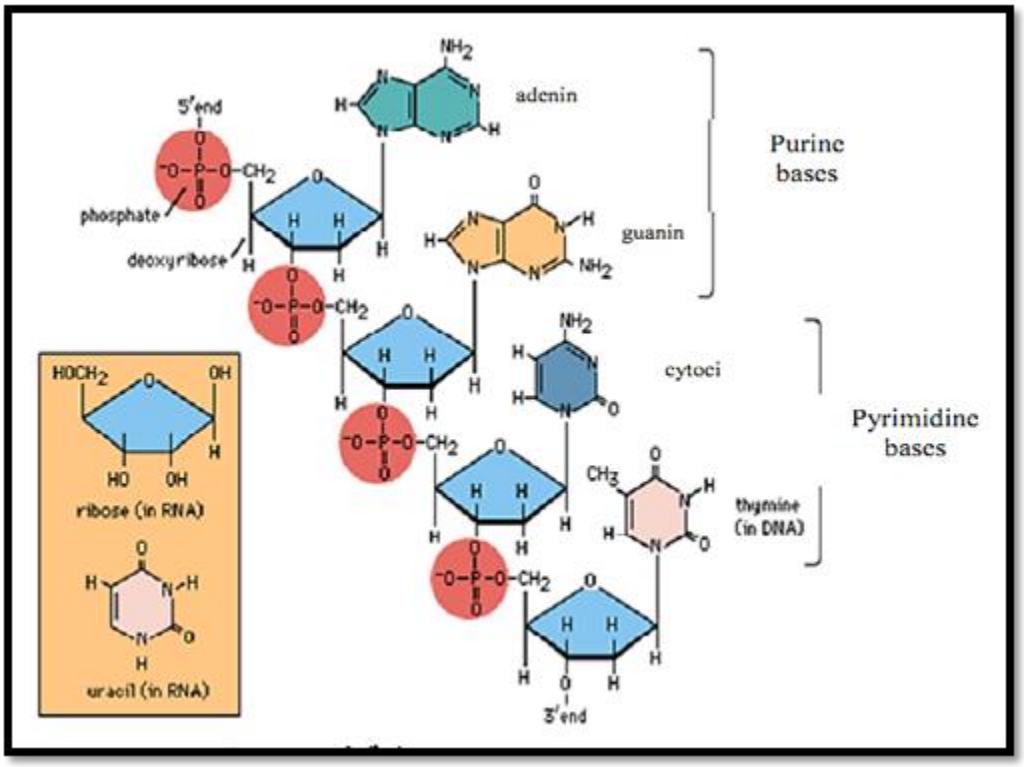
1) **Purines** : Fused 5 & 6 membered rings (heterocyclic) eg: Adenine (A) and Guanine (G).

2) **Pyrimidines** : Six membered rings eg: Thymine (T), Cytosine (C) and Uracil (U).



In a nucleotide the base is joined to 1' carbon of pentose by an N-β-glycosyl bond and a phosphate is esterified to 5' carbon . Phosphate of 5' carbon reacts with –OH group attached to 3' ribose sugar carbon. During this bond formation a water molecule is removed . This process is continued to make a polynucleotide with one end 3' sugar carbon attached -OH free (called 3' end) and at other end 5' sugar carbon attached phosphate free (called 5' end).

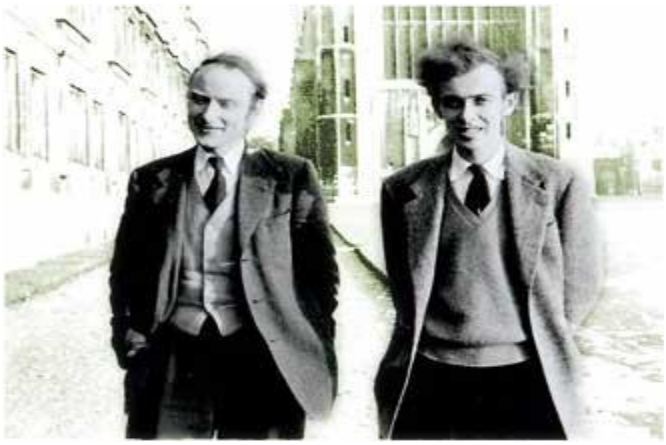




Structure of DNA :

James Watson and Francis Crick discovered the Deoxyribo-nucleic Acid (DNA) structure in the cell in 1953 using X-ray diffraction patterns which showed that the DNA molecule is long, thin, and has a spiral-like shape

Crick Watson

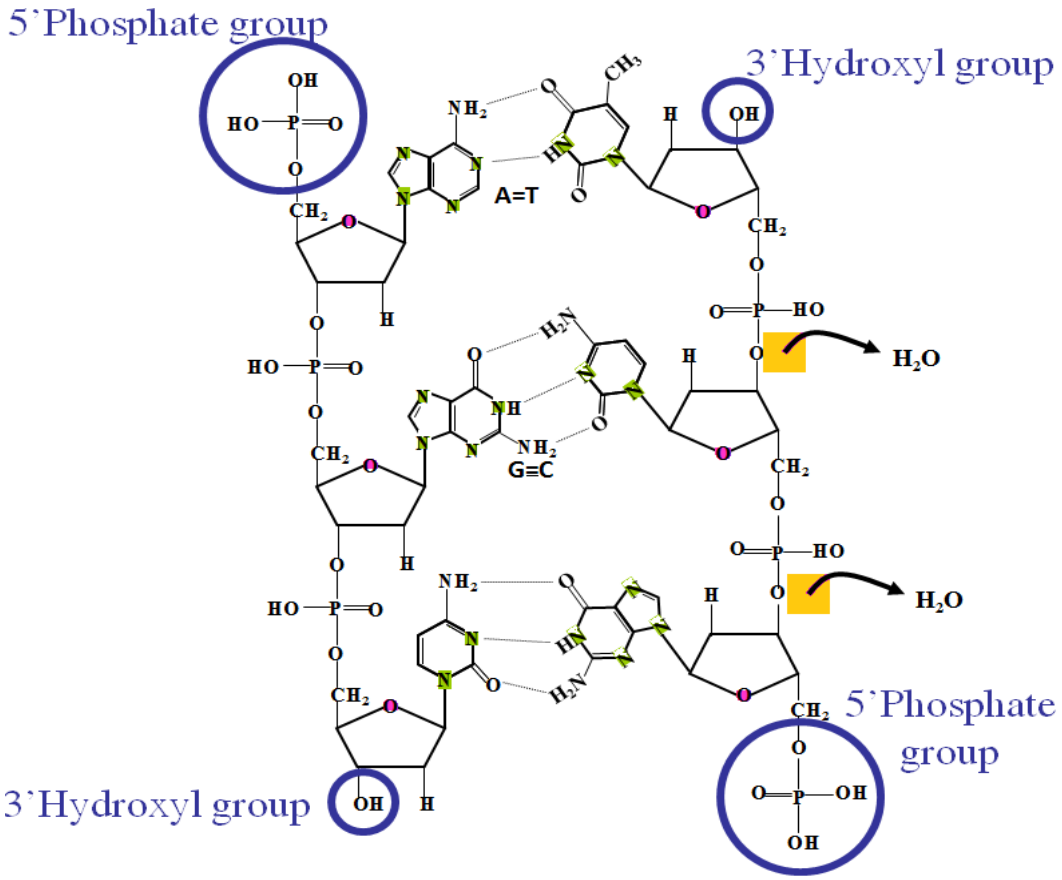


1916-2004 1928

Molecular biology, Physics. Genetics

The DNA is contained in the nuclei of eukaryotic cells, and is composed of small molecules called ***nucleotides*** . Each nucleotide consists of a five-carbon sugar, a phosphate group, and a base. The carbon atoms in a sugar molecule are labeled 1_ to 5_, DNA molecules start at 5_ end and finish at 3_ end. There are four nucleotides in the DNA which are distinguished by the bases they have: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T).

Human DNA consists approximately of three billion bases. Nucleotide A pairs only with T, and C pairs only with G, we can say A and T are complementary and so are G and C . **It has a diameter of 20 Å. Each complete turn of the coil contains 10 bases . Further, two bases are 3.4 Å apart from each other. The length of a complete turn of the helix is 34 Å.**



The strands have two important characteristics:

complementary and antiparallel.

Complementary means

_ A will only base pair with T by 2 hydrogen bonds

_ C will only base pair with G by 3 hydrogen bonds

Antiparallel means they are oriented in different directions.

Types of DNA

In the 1950 the X –RAY diffraction data suggested that DNA is helical strand. The DNA helix can take on several conformations.

1) The most common form is B-DNA, in which the helix is right handed and has just over 10 bp per helical turn.

2) A -DNA is a right handed helix which is shorter and wider than B-DNA .

3) Z-DNA is a left handed helix in which alternating purines and pyrimidines gives rise to a zigzag appearance to the helix.

RNA (Ribonucleic acid):

In some viruses, RNA is the hereditary material , but in all cells genetic information is stored in DNA molecules. Certain regions of the cellular DNA molecules serve as templates for synthesizing RNA molecules. The great majority of the RNA molecules are used to specify the synthesis of polypeptides

Structure of RNA

Ribonucleic acid or RNA is the second major nucleic acid in cells, serves as genetic messenger, passing the information stored in DNA to other parts of cell for protein synthesis. Like DNA it is also a long unbranched

polymer consisting of nucleotides joined by phosphordiester bonds in 5' to 3' direction. RNA differs from DNA in two respects…

1) It has a similar structure to DNA but consists of only one strand and does not form a helix structure like DNA (mostly).

2) Uracil (U) is found in RNA in place of Thymine (T).

3) The sugar units in RNA are ribose rather than deoxyribose (in DNA) .

In addition to 3' - 5', a 2' - 5' linkage is also possible for RNA. This 2' – 5' linkage is important in the removal of introns and joining of exons for the formation of mature RNA during RNA splicing. Due to the presence of an OH group at C-2 , RNA is hydrolysed more rapidly under alkaline conditions.

Three major forms of RNA exist in any living cell…

1) Messenger RNA (mRNA) : is a subtype of RNA. mRNA, molecule in cells that carries codes from the DNA in the nucleus to the sites of protein synthesisin the cytoplasm (the ribosomes). mRNA is created during transcription. During the transcription process, a single strand of DNA is decoded by RNA polymerase, and mRNA is synthesized. Physically, mRNA is a strand of nucleotides known as ribonucleic acid. This class of RNA is the genetic coding templates used by the translational machinery to determine the order of amino acids incorporated into an elongating polypeptide in the process of translation.

2) Transfer RNA (tRNA) : small molecule in cells that carries amino acids to organelles called ribosomes, where they are linked into proteins.This class of small RNA form covalent attachments to individual

amino acids and recognize the encoded sequences of the mRNAs to allow correct insertion of amino acids into the elongating polypeptide chain.

3) Ribosomal RNA (rRNA): molecule in cells that forms part of the protein-synthesizing organelle known as a ribosome . It is synthesized in nucleolus and that is exported to the cytoplasm to help translate the information in messenger RNA (mRNA) into protein. This class of RNA is assembled, together with numerous ribosomal proteins, to form the ribosomes .

* These are the most abundant form of RNA and counts about 80% of total cellular RNA.

Summary of differences between DNA and RNA

	DNA & RNA	
Name	Deoxyribo Nucleic Acid. (DNA)	Ribo Nucleic Acid. (**RNA**)
Definition	A nucleic acid that contains the genetic in used in the development and functioning of all modern living organisms.	The information found in DNA determines which traits are to be created, activated, or deactivated, while the various forms of RNA do the work.
Function	Long-term storage of genetic information; transmission of genetic information to make other cells and new organisms..	RNA is used to transmit genetic information in some organisms and may have been the molecule used to store genetic information in primitive organisms . Used to transfer the genetic code from the nucleus to the ribosomes to make proteins.
Structure	B-form , double helix . DNA is a double-stranded molecule consisting of a long chain of nucleotides. DNA is consist of its phosphate group, five-carbon sugar (the stable 2-deoxyribose), and 4 nitrogen-containing nucleobases : adenine, thymine, cytosine, and guanine.	A-form helix. RNA usually is a single-strand helix consisting of shorter chains of nucleotides . Like DNA, RNA is composed of its phosphate group, five-carbon sugar (the less stable ribose), and 4 nitrogen-containing nucleobases: adenine, uracil (not thymine), guanine, and cytosine.
Composition of Bases and Sugars	deoxyribose sugar , phosphate backbone adenine, guanine, cytosine, thymine base.	ribose sugar ,phosphate backbone ,adenine, guanine cytosine, uracil bases.
Base Pairing	Adenine links to thymine (A-T) and cytosine links to guanine (C-G).	Adenine links to uracil (A-U) and cytosine links to guanine (C-G).
Location	DNA is found in the nucleus of a cell and in mitochondria .	Depending on the type of RNA, this molecule is found in a cell's nucleus, its cytoplasm , and its ribosome.
Stability	Deoxyribose sugar in DNA is	Ribose sugar is more reactive

	less reactive because of C-H bonds. Stable in alkaline conditions.	because of C-OH (hydroxyl) bonds. Not stable in alkaline conditions.
Propagation	DNA is self-replicating.	RNA is synthesized from DNA when needed.
Unique Features	The helix geometry of DNA is of B-Form. DNA is protected in the nucleus, as it is tightly packed. DNA can be damaged by exposure to ultra-violet rays.	The helix geometry of RNA is of A-Form. RNA strands are continually made, broken down and reused. RNA is more resistant to damage by Ultra-violet rays.

DNA is the genetic material

The nucleic acids and proteins became the key molecules that were implicated in having the genetic information, its transfer from one generation to another and carrying out various biochemical processes. Scientists were looking for the material that was responsible for the storage of genetic information and transfer of parental characters to offsprings. A group of scientists believed that DNA is the genetic material while many of the geneticists doubted it and thought that proteins are the key molecules for transfer of genetic information**.**

In 1928 Frederick Griffith carried out the classical experiment where he infected mice with two separate strains of *Pneumococcus*. One strain that contained capsular glycoproteins had smooth surface (*the `S' strain*) and was virulent while the other strain had rough cell surface (*the `R' strain*) and was non-virulent. He showed that infection with live bacteria of `S' strain resulted in disease and lead to death of mice while the mice were not killed if infected with `R' strain. Further, the mice survived when infected with heat-killed bacteria of `S' strain. However, the co-infection with bacteria of heat killed `S' strain and the live `R' strain together resulted in

death of the animals. This suggested that some factor(s) present in the killed `S' bacteria was capable of transforming the `R' bacterium resulting in its (the bacteria of non-virulent `R' strain) conversion (transformated) to a virulent strain. Furthermore, when the DNA from heat killed `S' strain was injected into the mice along with live `R' strain bacteria, the combination was found to be virulent and lethal. It clearly demonstrated that DNA from killed `S' strain bacteria was able to transform the `R' strain bacteria and this was responsible for the pathogenicity of the transformed bacteria. However, doubts were still and a number of scientists still thought that the contaminating proteins might have been responsible
for transformation of the `R' bacteria resulting in the death of mice. The dilemma continued till 1943 and many workers still believed that proteins were the genetic material.

In 1944 Avery, McLead and McCarty extracted the transforming principle from the heat killed bacteria of `S' strain, chemically characterized it and showed it to be DNA. Their analysis included elemental analysis as well as physical characterizations such as optical properties, ultracentrifugal behavior, electrophoretic migration and diffusion properties. Further, they also showed that removal of even the last traces of lipids and proteins from the transforming principle has no effect on its effectiveness as transforming factor. The treatment of this factor with either proteases or RNases did not result in loss of its activity. But treatment with DNase resulted in total loss of its transforming property.

In 1951 Roger Harriot determined the structure of bacteriophages and showed that the phages have needle like structure with a head and a sharp pointed tail. The nucleic acid was shown to be localized in the head region. Experiments

demonstrated that during infection the tail pierces (penetrate) the host bacteria and facilitates the transfer of phage nucleic acid into the host cell. The outer envelope (or the proteins) remains outside the infected host. The phage nucleic acid was implicated to be responsible for the host cell transformation. This observation gave further support to the theory that DNA was the genetic material.

Though all these studies made it clear that DNA is the genetic material and is responsible for the transfer of virulent characters in original experiment of Griffith .

Finally **in 1952 Alfred Hershey and Martha Chase performed their famous double labeling experiment that clearly proved that DNA is the genetic material** . They radio labeled the DNA of phage T_2 with ^{32}P and its protein with ^{35}S.

Bacteria were infected with the double-labeled phage and the fate of radioactivity was carefully followed. It was found that while ^{32}P (i.e. phage DNA) was taken up by the host cells and was detected inside the infected host cells, ^{35}S (i.e. the phage proteins) did not enter the host cells and remained outside in the culture medium.

This confirmed that DNA was the genetic material that was responsible for the transformation of the host bacteria beyond any doubt. It is now well established that DNA is the genetic material for all the living cells except in certain viruses