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

Advanced Molecular Biology

2020-2021

دراسات عليا

ماجستير-فرع الاحياء المجهرية

الفصل الدراسي الاول

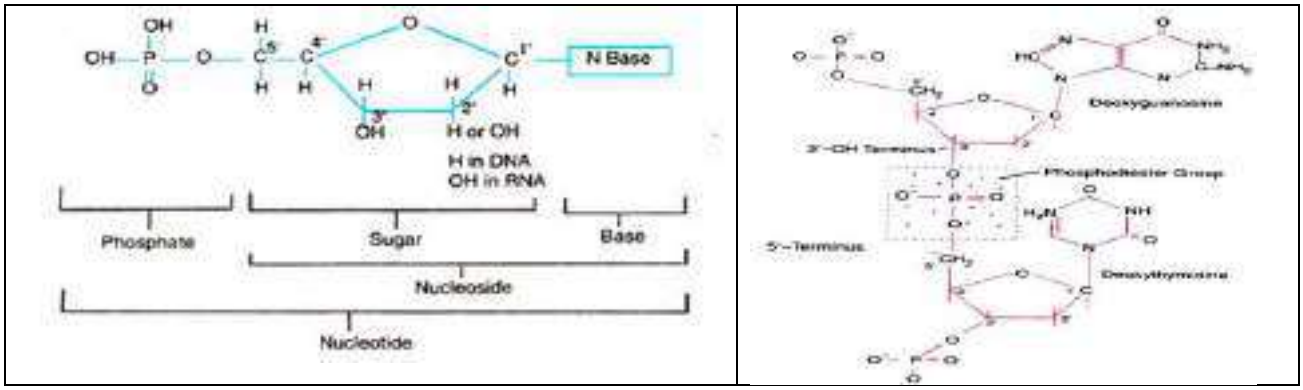
تدريسي المادة:

أ.د. هيفاء هادي حساني ا.م.د. غصون علي عبد الحسن

Lec1. Structure of DNA molecule, RNA molecules and genome

The nucleic acids found in viruses and all living organisms including microorganisms, plants and animals carry the genetic information. A nucleic acid is the polynucleotide or polymer of nucleotide, each nucleotide consists of three main components:

1. Five Carbon Ring
2. Nitrogenous Base: purines (A,G) and pyrimidines (T,C,U).
3. Phosphate Group

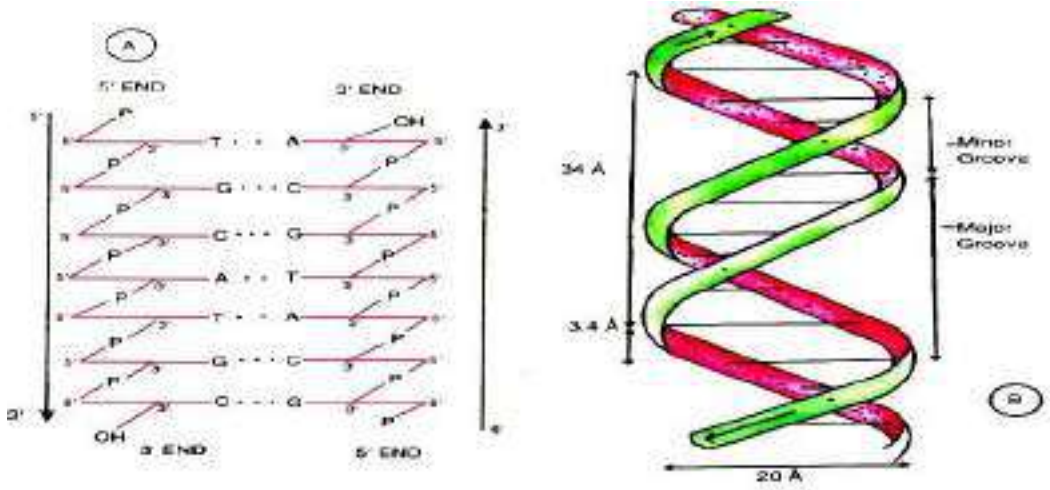


Watson and Crick's Model of Double Helix of DNA

In 1953,Watson and Crick's put the Model of Double Helix of DNA. They concluded that:

1. Polynucleotide chain of DNA has a regular helix. The DNA molecule consists of two strands which are connected together by hydrogen bonds and helically twisted.
2. The helix has a diameter of about 20 Å.
3. The double helix coils in right hand direction makes a full turn at every 34 Å along its length and contains a stack of ten nucleotides per turn because the inter-nucleotide distance is 3.4 Å.

The dimension of purine ring is more than the dimension of pyrimidine ring; therefore, the two chain helix contain a constant diameter provided a complementary relationship existed between the two nucleotide stacks, one harbouring purine and the other pyrimidine bases. The thermodynamic stability was provided by the formation of hydrogen bonds between amino ($-NH_2$) or hydroxyl (-OH) hydrogen's and ketone oxygen or amino-nitrogen of the two bases. The two strands of double helix ran in antiparallel direction i.e. they have opposite polarity. The left hand strand has $5' \rightarrow 3'$ polarity, whereas the right hand has $3' \rightarrow 5'$ polarity as compared to the first one. The polarity is due to the direction of phosphodiester linkage. The turning of double helix results in the appearance of a deep and wide groove called major groove (is the site of bonding of specific protein). Sugar-phosphate makes the backbone of double helix of DNA molecule.



DNA Forms

The most common form of DNA which has right handed helix and proposed by Watson and Crick is called B-form of DNA or B-DNA. In addition, the DNA may be able to exist in other forms of double helical structure. In certain condition, different forms of DNAs are found to be appeared like A-DNA,Z-DNA,C- DNA,D-DNA,E-DNA. These deviations in forms are based on their structural diversity such as spacing between nucleotides and number of nucleotides per turn, rotation per base pair, vertical rise per base pair and helical diameter.

	Forms of DNA					
Characteristic	**A-DNA**	**B-DNA**	**C-DNA**	**D-DNA**	**E-DNA**	**Z-DNA**
Coiling	Right handed	Right handed	Right handed	—	—	Left handed
Pitch	—	34 Å	—	—	—	45 Å
Base pair per turn	11	10	9.33	8	7.5	12
Helical diameter	25.5Å	23.7Å	23.7 Å	—	—	18.4Å
Axial rise per base pair	2.3 Å	3.4Å	3.3 Å	—	—	3.8 Å
Sugar phosphate backbone	—	Regular	—	—	—	Zig-zag
Tilt of base pair	20.2^0	6.3^0	7.8^0	16.7^0	—	7^0

Circular and linear DNA

Almost in all the prokaryotes and a few viruses, the DNA is organized in the form of closed circle. The two ends of the double helix get covalently sealed to form a closed circle. Thus, a closed circle contains two unbroken complementary strands.

Bacterial chromosomal DNA is usually a circular molecule that is a few million nucleotides in length.

- *Escherichia coli* 4.6 million base pairs
- *Haemophilus influenzae* 1.8 million base pairs

Bacterial chromosome is haploid located in the cytoplasm found in region called the nucleoid mostly single and some bacteria have multiple circular chromosomes such as *Brucella melitensis*, *Vibrio* Species and *Agrobacterium tumefaciens* (one circular and one linear). Linear chromosomes found in Gram-positive *Borrelia* and *Streptomyces*.

Bacterial DNA binding proteins are a family of small proteins that bind DNA and are known as histone-like proteins which have many similar traits with the eukaryotic histone proteins. Examples include the HU protein in *Escherichia coli*, a dimer of alpha and beta chains and in other bacteria can be a dimer of identical chains.
HU is a small bacterial histone-like protein that resembles the eukaryotic Histone H2B. HU acts similarly to a histone by inducing negative supercoiling into circular DNA with the assistance of topoisomerase. The protein has been implicated in DNA replication, recombination, and repair. It is also binds to RNA and DNA-RNA hybrids with the same affinity as supercoiled DNA.

Eukaryotic chromosomes are typically linear. Eukaryotic cells contain two copies of each chromosome, therefore, are diploid. Eukaryotic chromosomes contain DNA and protein, tightly packed together to form chromatin. The major proteins involved in chromatin are:

1. **Structural histone proteins (Packaging proteins**): main structural proteins found in eukaryotic cells with high proportion of positively charged amino acids, bound to DNA along most of its length. The positive charge helps histones to bind to DNA and play a crucial role in packaging of long DNA molecules. constitute about 60% of total protein.
2. **Functional Proteins (Non- Histones):** associated with gene regulation and other functions of chromatin.

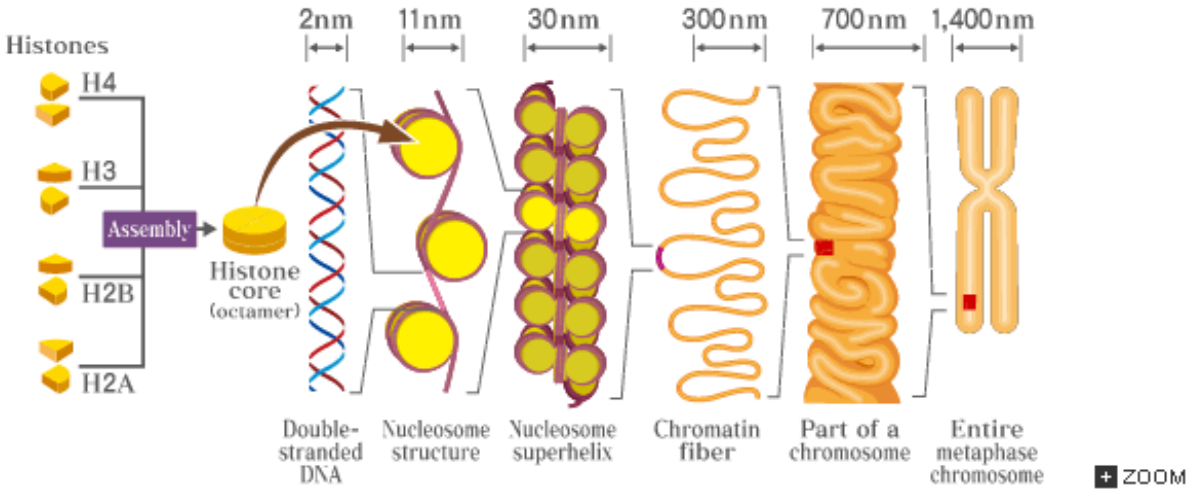
DNA and histone molecules form nucleosomes. The nucleosome consists of a DNA double helix bound to an octamer of core histones (2 dimers of H2A and H2B, and an H3/H4 tetramer) link to DNA by linker histone (H1).

Chemical composition of chromatin are DNA (20-40 %), RNA (5-10 % associated with chromatin as rRNA, mRNA and tRNA) and proteins (55-60%).The functions of chromatin

is to package DNA into smaller volume to fit in the cell, strengthen the DNA to allow mitosis and meiosis and control gene expression and DNA replication. There are two types of chromatin, euchromatin and heterochromatin.

Euchromatin is lightly packed form of chromatin that is rich in gene concentration, takes up light stain and represent most of the chromatin. Consists of structural genes which replicate and transcribe during G1 and S phase of the interphase. Considered genetically active chromatin, since it has a role in their phenotypic expression of the genes.

Heterochromatin is tightly packed form of chromatin that takes up deep stain Heterochromatin consists of highly repetitive DNA sequences. It is late replicating during the s-phase of the cell and considered genetically inctive chromatin.



Packaging of the DNA into chromosomes serves several important functions:

1. The length of a chromosome greatly exceeds the length of the cell, so a chromosome needs to be packaged into a very small space to fit within the cell. For example, the combined length of all of the 3 billion base pairs of DNA of the human genome would measure approximately 2 meters if completely stretched out. Most compaction in eukaryotic cells is the result of the regular association of DNA with histones to form structures called nucleosomes. The formation of nucleosomes is the first step in a process that allows the eukaryotic DNA to be folded into much more compact structures that reduce the linear length by as much as 10,000-fold.
2. Packaging the DNA into chromosomes serves to protect the DNA from damage. Completely naked DNA molecules are relatively unstable in cells. In contrast, chromosomal DNA is extremely stable.

3. Only DNA packaged into a chromosome can be transmitted efficiently to both daughter cells when a cell divides.

Centromere is the specialized DNA sequence of a chromosome that links a pair of sister chromatids. During mitosis, spindle fibers attach to the centromere via the kinetochore (highly complex multiprotein structure that is responsible for the actual events of chromosome segregation).

Telomere is a region of repetitive nucleotide sequences at each end of most eukaryotic chromosome, which protects the ends from deterioration or from fusion with neighboring chromosomes. For vertebrates, the sequence of nucleotides in telomeres is TTAGGG. A protein complex known as shelterin serves to protect the ends of telomeres from being recognized as double-strand breaks. A small fraction of bacterial chromosomes (such as those in *Streptomyces*, *Agrobacterium*, and *Borrelia*) are linear and possess telomeres, which are very different from those of the eukaryotic chromosomes in structure and functions.

The majority of eukaryotic cells are diploid; that is, they contain two copies of each chromosome. The two copies of a given chromosome are called homologs, one being derived from each parent. But not all cells in a eukaryotic organism are diploid; a subset of eukaryotic cells are either haploid or polyploid. Haploid cells contain a single copy of each chromosome and are involved in sexual reproduction (e.g., sperm and eggs are haploid cells).
Polyploid cells have more than two copies of each chromosome. Indeed, some organisms maintain the majority of their adult cells in a polyploid state. This type of global genome amplification allows a cell to generate larger amounts of RNA and, in turn, protein. For example, megakaryocytes are specialized polyploid cells (about 28 copies of each chromosome) that produce thousands of platelets, which lack chromosomes but are an essential component of human blood (there are about 200,000 platelets per milliliter of blood). By becoming polyploid, megakaryocytes can maintain the very high levels of metabolism necessary to produce large numbers of platelets.

A typical chromosome consisted of:

- Structural gene sequences (encoding proteins).
- Noncoding sequences include introns, sequences for non-coding RNAs, regulatory regions, and repetitive DNA.

Prokaryotes appear to use their genomes very efficiently, with only an average of 12% of the genome being taken up by noncoding sequences. In contrast, noncoding DNA can represent about 98% of the genome in eukaryotes, as seen in humans, but the percentage

of noncoding DNA varies between species. It found roles for some of these regions, many of which contribute to the regulation of transcription or translation through the production of small noncoding RNA molecules, DNA packaging, and chromosomal stability.

Extrachromosomal DNA

Although most DNA is contained within a cell's chromosomes, many cells have additional molecules of DNA outside the chromosomes, called extrachromosomal DNA, that are also part of its genome. The genomes of eukaryotic cells would also include the chromosomes from any organelles such as mitochondria or chloroplasts. In some cases, genomes of certain DNA viruses can also be maintained independently in host cells during latent viral infection. In these cases, these viruses are another form of extrachromosomal DNA.

Besides chromosomes, some prokaryotes also have smaller loops of DNA called plasmids that may contain one or a few genes not essential for normal growth. Bacteria can exchange these plasmids with other bacteria in a process known as horizontal gene transfer.

Genome size and density

Genome size is correlated with an organism's complexity. Genome size varies substantially between different organisms because more genes are required to direct the formation of more complex organisms such as prokaryotic cells typically have genomes of 10 Mb, the genomes of single cell eukaryotes are typically 50 Mb and multicellular organisms have even larger genomes that can reach sizes 100,000 Mb.

Genome Density: Gene density is the average number of genes per megabase (Mb) of genomic DNA. For example, if an organism has 5000 genes and a genome size of 50 Mb, then the gene density for that organism is 100 genes/Mb.

There is inverse correlation between organism complexity and gene density that the less complex organism has the higher the gene density. For example, the highest gene densities are found for viruses. Bacterial gene density is consistently near 1000 genes/Mb. Gene density in eukaryotic organisms is consistently lower and more variable than in their prokaryotic. The simple unicellular eukaryote *Saccharomyces cerevisiae* has a gene density 500 genes/Mb. In contrast, the human genome is estimated to have a 50-fold lower gene density.

Repetitive DNA sequences and RNA structure

Assistant prof. Dr. Ghusoon Ali Abdulhasan

Advanced Molecular Biology/ Lec 2

Postgraduate Studies/ Microbiology

2020-2021

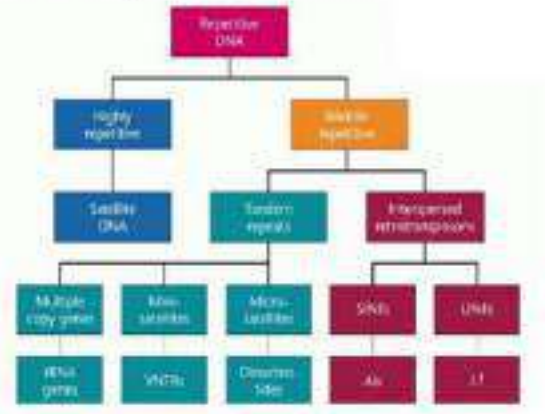
Repetitive DNA sequences

Repeated sequences (also known as repetitive elements, or repeats) are patterns of nucleic acid (DNA or RNA) that occur in multiple copies throughout the genome.

Both moderately (middle) repetitive and highly repetitive DNA sequences are sequences that appear many times within a genome.

These sequences can be arranged within the genome in one of two ways:

- distributed at irregular intervals known as dispersed repeated DNA or interspersed repeated DNA.
- clustered together so that the sequence repeats many times in a row known as tandemly repeated DNA.



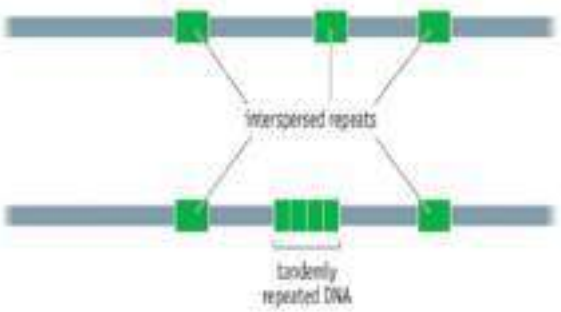
Two types of dispersed repeated sequences are known:

- **Long interspersed elements (LINEs), in which the sequences in the families are about 1,000–7,000 bp long. LINEs are mammalian retrotransposons and consist of repetitive sequences that are adenine-rich at their 3'ends.**
- **Short interspersed elements (SINEs), in which the sequences in the families are 100–400 bp long. SINEs are found in a diverse array of eukaryotic species, including mammals, amphibians, and sea urchins. A well-studied SINE family is the Alu family of certain primates. This family is named for the cleavage site for the restriction enzyme AluI typically found in the repeated sequence.**

In humans, the Alu family is the most abundant SINE family in the genome, consisting of 200–300 bp sequences repeated as many as a million times. One Alu repeat is located every 5,000 bp in the genome.

All eukaryotic organisms have LINEs and SINEs, with a wide variation in their relative proportions. Humans and frogs, for example, have mostly SINEs, whereas Drosophila and birds have mostly LINEs.

LINEs and SINEs found in the moderately repetitive DNA. The distribution of these elements has been implicated in some genetic diseases and cancers.



- **Transposition is a specific form of genetic recombination that moves certain genetic elements from one DNA site to another. These mobile genetic elements are called transposable elements or transposons.**

- **Dispersed throughout eukaryotic genomes are multiple copies of different transposable elements and hence these elements are moderately repetitive DNA (dispersed) sequences.**

- **Eukaryotic transposable elements are divided into two classes based on the properties of their transposition intermediate.**

- **In class 1 elements the transposition intermediate is formed from mRNA. Since these elements resemble retroviruses they are usually referred to as retrotransposons.**

- **In class 2 elements the transposition intermediate is the transposon DNA.**

Both classes of transposon can be subdivided into:

- ✓ **autonomous elements encode the gene products required for transposition.**
- ✓ **non-autonomous elements have no significant coding capacity but retain the DNA sequences necessary for transposition.**

Interspersed Repetitive DNA

These are:

- **Retrotransposons** (class I transposable elements) (**copy and paste**), copy themselves to RNA and then back to DNA (using reverse transcriptase) to integrate into the genome.
- **Transposons** (Class II TEs) (**cut and paste**) uses transposases to make makes a staggered sticky cut.

Tandem Repeats

Some moderately and highly repetitive sequences are clustered together in a tandem array, also known as tandem repeats. In a tandem array, a very short nucleotide sequence is repeated many times in a row.

Depending on the average size of the arrays of repeat units, highly repetitive noncoding DNA belonging to this class can be grouped into three subclasses:

❖**Satellite DNA: repeat unit 100-500 pb. Human satellite DNA is comprised of very large arrays of tandemly repeated DNA. Repeated DNA of this type is not transcribed. Accounts for the bulk of the heterochromatic regions of the genome, being notably found in the centromeres.**

❖**Minisatellite DNA: 6-50 bp. comprises a collection of moderately sized arrays of tandemly repeated DNA sequences which are dispersed over considerable portions of the nuclear genome (euchromatin). they are not normally transcribed.**

✓**In humans, 90% of minisatellites are found at the sub-telomeric region of chromosomes. The telomere sequence itself is a tandem repeat: TTAGGG TTAGGG TTAGGG which are added by a specialized enzyme, telomerase .**

✓**Variation in size (array length) of these regions between individuals in humans. The number of elements in a minisatellite region can vary, and thus these are also known as variable number tandem repeats (VNTRs). These can be used in mapping studies and also formed the basis of genetic fingerprinting (DNA profiling).**

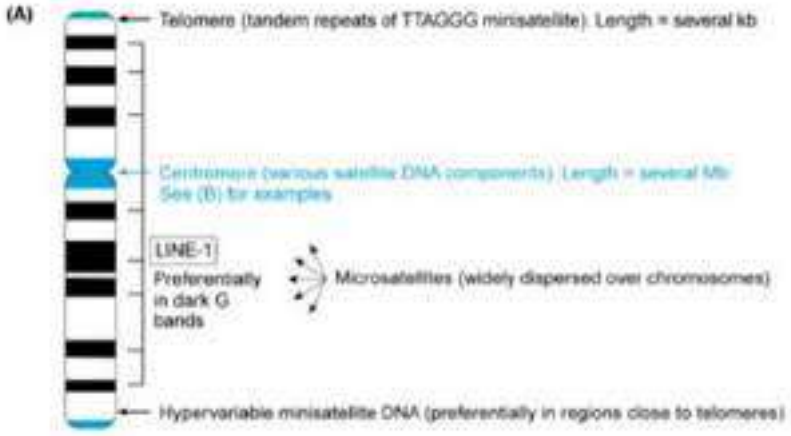
❖**Microsatellite DNA:** usually 4 bp or less, known as Short Tandem Repeat (STR), Simple Sequence length polymorphism (SSLP) and Simple Sequence Repeat (SSR). Comprise 3% of human genome (euchromatin).

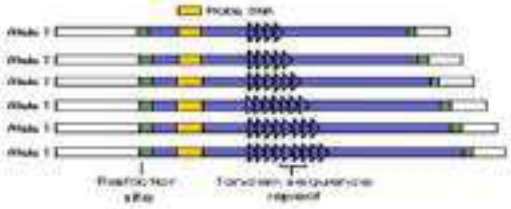
✓repeats are interspersed throughout the genome can be repeated 10 to 100 times. Most is the dinucleotide repeats are common in humans (CA)n sequence where n varies from 5 -50 or more (e.g., CACACACACACACACA). Found on average every 10kb in the human genome.

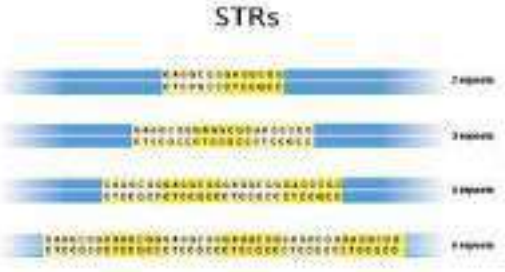
✓Trinucleotide repeats are much rarer . The significance of these repeats in normal genes is not known but they can be the locus for a number of inherited disorders when they undergo unstable expansion. For example, in fragile X syndrome patients can exhibit hundreds or even thousands of the CGG triplet at a particular site, whereas unaffected individuals only have about 30 repeats.

✓Although microsatellite DNA has generally been identified in intergenic DNA or within the introns of genes, a few examples have been recorded within the coding sequences of genes.

✓Similar trinucleotide repeats have been discovered in bacteria and yeast following complete genome sequencing. Some pathogenic bacteria use length variation in simple repeats to change the antigens on their surfaces so that they can evade host immune attack.







- **Repeated DNA sequences are a portion of microbial genomes that are found more than once in the genome of the species. Interspersed repetitive DNA sequences have been described in different microbial genomes, characterized as relatively short (usually <500 bp), non-coding, and dispersed elements in bacterial genomes.**
- **The repetitive sequences are found in different species of Gram-negative and Gram-positive bacteria. Interspersed repetitive sequence elements called Repetitive Extragenic Palindromic (REP) and Enterobacterial Repetitive Intergenic Consensus (ERIC) sequences are present in different species of Enterobacteriaceae family, including *E. coli* and *Salmonella typhimurium*.**
- **Their functions are not completely clear, probably they play important role in regulation of gene expression. Nevertheless, REP and ERIC elements are widely used in identification and genetic analysis of bacteria. For example, using rep-PCR technique it is possible to discriminate between closely related serovars of the same species, which enables to analyze phylogenetic and epidemiological relations among them.**

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) is a family of DNA sequences in bacteria and archaea containing short, repetitive base sequences. The sequences contain snippets of DNA from viruses that have attacked the prokaryote. These snippets are used by the prokaryote to detect and destroy DNA from similar viruses during subsequent attacks. These sequences play a key role in a prokaryotic defense system.

Repeats display diverse roles in terms of bacterial cell physiology and cell–host interactions. They are found positioned upstream or downstream of genes and regulate gene expression at the transcriptional or post-transcriptional levels.

The structure of RNA

RNA differs from DNA in three respects:

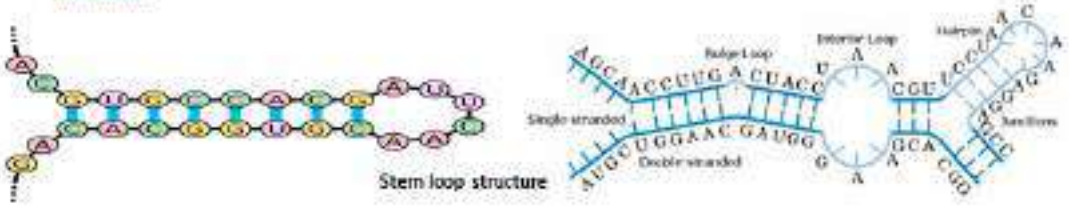
1. The backbone of RNA contains ribose rather than 2'-deoxyribose.
2. RNA contains uracil in place of thymine.
3. RNA is usually found as a single polynucleotide chain.

RNA chains fold back on themselves to form local regions of double helix form base-paired segments between short stretches of complementary sequences similar to A- form DNA.

If the two complementary sequence are near each other, the RNA may adopt a stem-loop structure.

double-helical RNA may also exhibit:

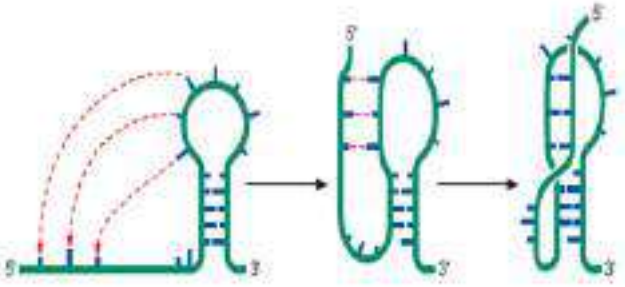
- internal loops (unpaired nucleotides on either side of the stem),
- bulges (an unpaired nucleotide on one side of the bulge), or
- junctions.

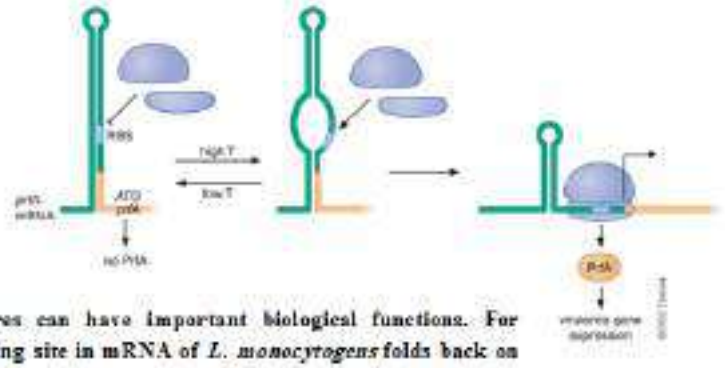


Stem loop structure

The presence of 2′-hydroxyls in the RNA backbone prevents RNA from adopting a B-form helix. Rather, double-helical RNA resembles the A-form structure of DNA.

Base pairing can also take place between sequences that are not contiguous to form complex structures named pseudoknots



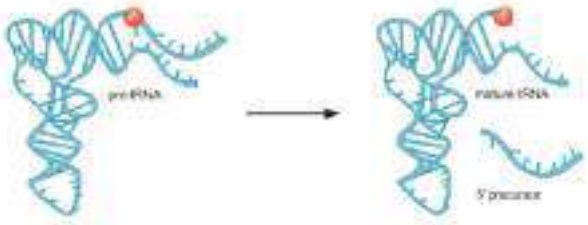


RNA secondary structures can have important biological functions. For Example, Ribosome binding site in mRNA of *L. monocytogens* folds back on itself to forma temperature-sensitive secondary RNA structure that masks the ribosome-binding site, such that it is inaccessible to the ribosome at 30 °C. At 37 °C, however, the structure melts, allowing the translation machinery to gain access to the ribosome-binding site.

RNAs can similarly adopt complex tertiary structures, can also be biological catalysts. Such RNA enzymes are known as ribozymes , and they exhibit many of the features of a classical enzyme, such as:

- an active site.
- a binding site for a substrate.
- a binding site for a co-factor , such as a metal ion.

One of the first ribozymes to be discovered was RNase P, an endoribonuclease that is involved in generating tRNA molecules from larger, precursor RNAs. Specifically, RNase P cleaves off a leader segment from the 5′-end of the precursor RNA in helping to generate the mature and functional tRNA .



Different forms of RNA

All RNA molecules except the RNA genomes of certain viruses are derived from information permanently stored in DNA. During transcription, an enzyme system converts the genetic information in a segment of double-stranded DNA into an RNA strand with a base sequence complementary to one of the DNA strands.

Three major kinds of RNA are produced:

- **Messenger RNAs (mRNAs):** Encode the amino acid sequence of one or more polypeptides specified by a gene or set of genes.
- **Transfer RNAs (tRNAs):** Read the information encoded in the mRNA and transfer the appropriate amino acid to a growing polypeptide chain during protein synthesis.
- **Ribosomal RNAs (rRNAs):** Are constituents of ribosomes, the intricate cellular machines that synthesize proteins.

Many additional specialized RNAs have regulatory or catalytic functions or are precursors to the three main classes of RNA. The sum of all the RNA molecules produced in a cell under a given set of conditions is called the cellular transcriptome.

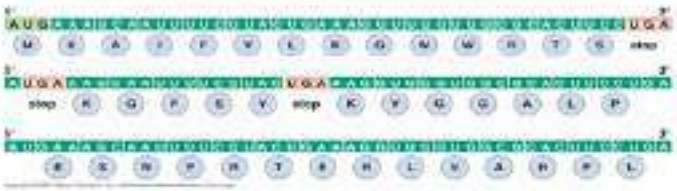
Messenger RNA

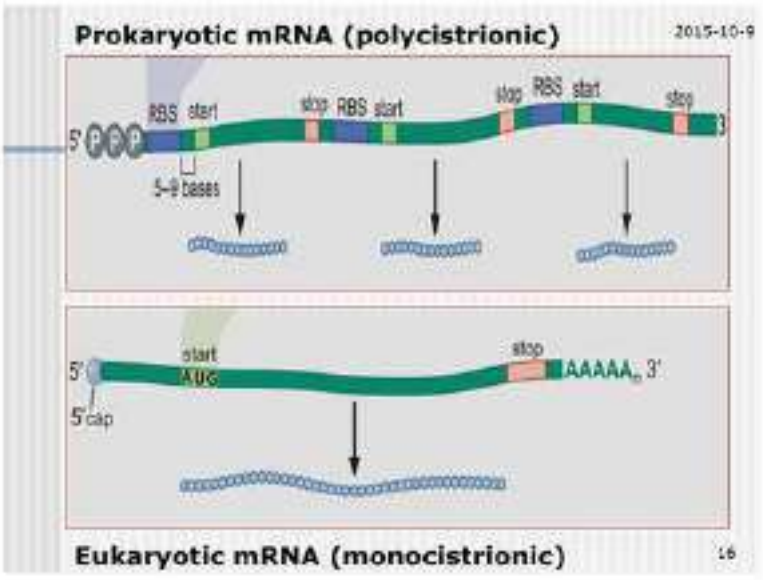
A newly synthesized RNA molecule is called a primary transcript. Perhaps the most extensive processing of primary transcripts occurs in eukaryotic mRNAs.

The protein coding region of each mRNA is composed of a contiguous of codons called an opening reading frame (ORF) that specify a single protein.

ORF is a sequence of codons which starts with start codon, ends with an end codon and has no end codons in-between (open to translation) that can be transcribed into mRNA and later translated into protein.

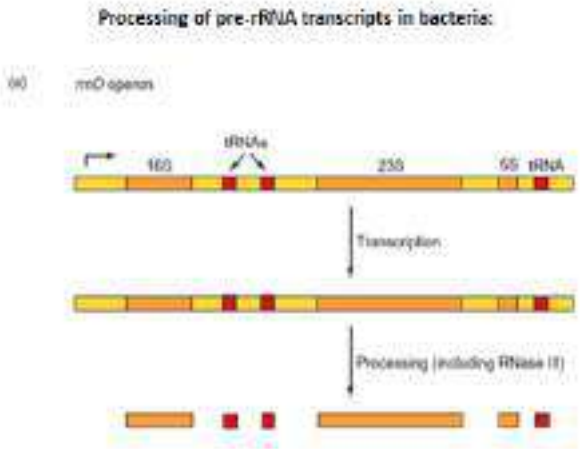
- **Polycistronic mRNAs:** mRNAs that contain multiple ORFs. Most of the mRNA found in bacteria and archaea is polycistronic
- **Monocistronic mRNAs:** mRNAs that contain a single ORF.





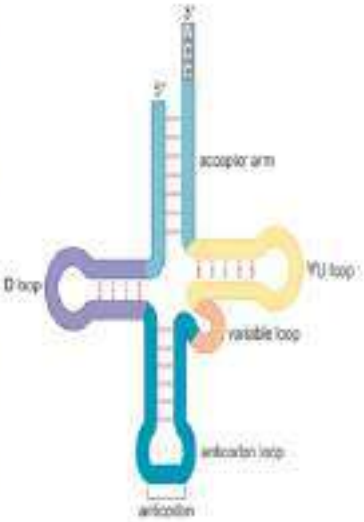
Ribosomal RNAs and tRNAs Processing

- **Ribosomal RNAs and tRNAs undergo posttranscriptional processing. Ribosomal RNAs of bacterial, archaeal, and eukaryotic cells are made from longer precursors called preribosomal RNAs, or pre- rRNAs.**
- **Ribosomal RNAs In bacteria, 16S, 23S, and 5S rRNAs arise from a single 30S RNA precursor of about 6,500 nucleotides. RNA at both ends of the 30S precursor and segments between the rRNAs are removed during processing. The 16S and 23S rRNAs contain modified nucleosides.**
- **The bacterium *E. coli* has seven *rrn* operons that contain rRNA genes. rrnD is one example of these operons. Transcription of the operon yields a 30S precursor, which must be cut up to release the three rRNAs and three tRNAs. The segment between the 16S and 23S rRNA genes generally encodes one or two tRNAs, with different tRNAs produced from different pre- rRNA transcripts.**



- **Translation of nucleotide sequence information into amino acids is accomplished by tRNA. Most cells have 40 to 50 distinct tRNAs. Each tRNA molecule is attached to specific amino acids (20) and each recognizes a particular codon, or codons (61), in the mRNA. All tRNAs end with the sequence CCA at the 3' end, where the aminoacyl tRNA synthetase adds the amino acid.**
- **Transfer RNAs are derived from longer RNA precursors by enzymatic removal of nucleotides from the 5' and 3' ends. The endonuclease RNase P, found in all organisms, removes RNA at the 5' end of tRNAs. The 3' end of tRNAs is processed by one or more nucleases, including the exonuclease RNase D.**

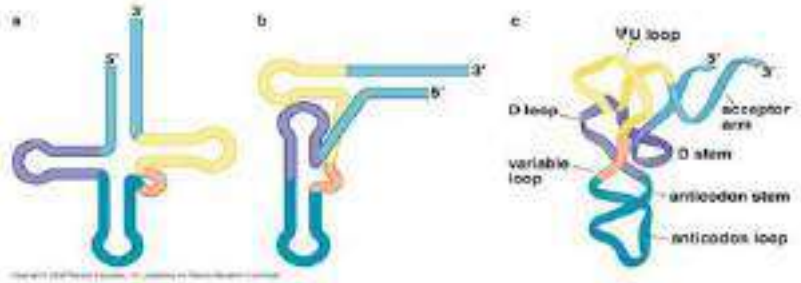
- The clover leaf structure is a common secondary structural representation of tRNA molecules which shows the base paring of various regions to form four stems (arms).
- The acceptor stem is formed by pairing between the 5′ and 3′ ends of the tRNA molecules. It is the site of amino acid attachment.
- The ΨU loop. The unusual base ΨU is often found within the sequence 5′-TΨUCG-3′.
- The D loop: presence of dihydrouridines
- The anticodon loop. the anticodon sequence is responsible for recognizing the codon by base pairing with the mRNA.
- The variable loop. It varies in size from 3 to 21 bases.



tRNAs have an L-shaped three-dimensional structure

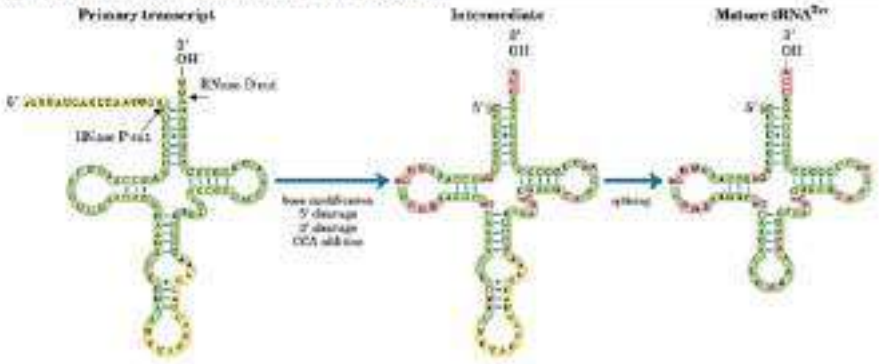
X-ray crystallography reveals an L-shaped tertiary structure. The terminus of the acceptor stem is about 70 Å away from the anticodon loop at the other end.

Three kinds of interactions stabilize this structure: 1. The formation of the two extended regions of base pairing results in additional base stacking interactions. 2. unconventional hydrogen bonds between bases in different helical regions. 3. Interactions between the bases and the sugar-phosphate backbone.



Processing of tRNA in bacteria and eukaryotes

Transfer RNA precursors may undergo further post-transcriptional processing. The 3′-terminal trinucleotide CCA(3′) to which an amino acid is attached during protein synthesis is absent from some bacterial and all eukaryotic tRNA precursors and is added during processing. This addition is carried out by tRNA nucleotidyltransferase. The final type of tRNA processing is the modification of some bases by methylation, deamination, or reduction.



Lec 3. DNA Replication

DNA replication is the biological process of producing two identical replicas of DNA from one original DNA molecule. This process occurs in all living organisms and is the basis for biological inheritance.

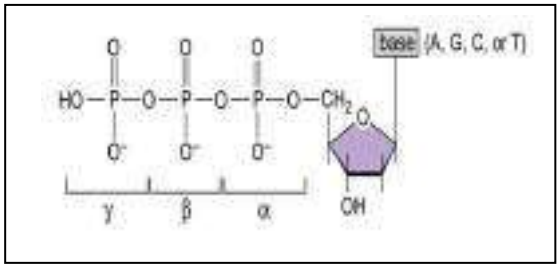
The Watson-Crick model for DNA replication assumed that new strands of DNA are made, they follow the usual base-pairing rules of A with T and G with C. The model also proposed that the two parental strands separate and that each then serves as a template for a new progeny strand. This is called semiconservative replication because each daughter duplex has one parental strand and one new strand. In 1958, Meselson and Stahl provided the experimental support for Watson-Crick model for semiconservative nature of DNA replication which is called Meselson-Stahl experiment.

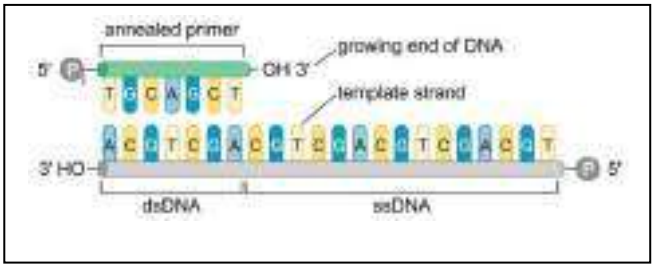
DNA Synthesis Requires Deoxynucleoside Triphosphates and a Primer

For the synthesis of DNA to proceed, two key substrates must be present:

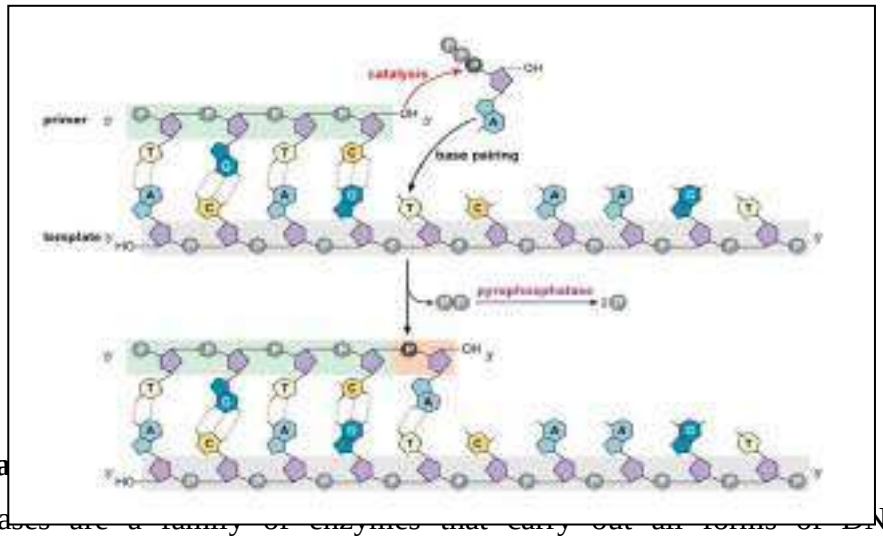
First, new synthesis requires the four deoxynucleoside triphosphates—dGTP, dCTP, dATP, and dTTP. Nucleoside triphosphates have three phosphoryl groups that are attached via the 5′-hydroxyl group. The phosphoryl group proximal to the deoxyribose is called the α-phosphate, whereas the middle and distal groups are called the β-phosphate and the γ-phosphate, respectively.

The **second** is a particular arrangement of single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) called a **primer:template junction**. As suggested by its name, the primer: template junction has two key components. The template provides the ssDNA have a free 3′-OH that directs the addition of each complementary deoxynucleotide.





New DNA synthesis extends the 3´ end of the primer by one nucleotide and releases one molecule of pyrophosphate. Pyrophosphatase rapidly hydrolyzes released pyrophosphate into two phosphate molecules.



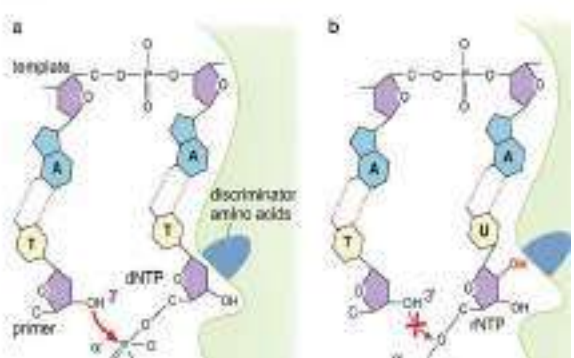
DNA Polymera

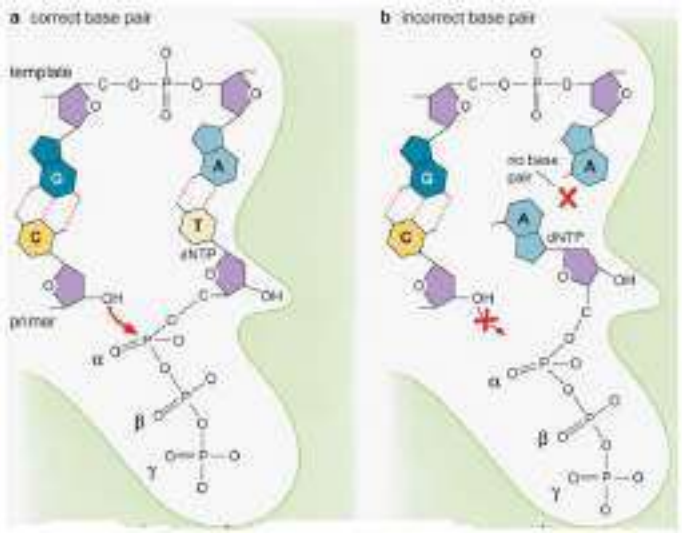
DNA polymerases are a family of enzymes that carry out all forms of DNA replication. DNA polymerases in general cannot initiate synthesis of new strands, but can only extend an existing DNA or RNA strand paired with a template strand. Unlike most enzymes, which have one active site that catalyzes one reaction, DNA polymerase uses a single active site to catalyze the addition of any of the four deoxynucleoside triphosphates (dNTP).

The DNA polymerase monitors the ability of the incoming nucleotide to form an A:T or G:C base pair, rather than detecting the exact nucleotide that enters the active site. Only when a **correct base pair is formed at the 3´-OH of the primer** and the ***α*-phosphate of the incoming nucleoside triphosphate in the optimum position** lead to catalysis reaction to occur. Incorrect base pairing leads to dramatically lower rates of nucleotide addition. The rate of incorporation of an incorrect nucleotide by DNA polymerase arise only about **once per 100,000 nucleotides**.

DNA polymerases show an impressive ability to distinguish between ribonucleoside (rNTPs) and dNTPs. Although rNTPs are present at approximately 10-fold higher concentration in the cell, they are incorporated at a rate that is more than 1000-fold lower than dNTPs.

The discrimination is mediated by the steric exclusion of rNTPs from the DNA polymerase active site. When binding of a correct dNTP to the DNA polymerase, the 3´-OH of the primer and the α-phosphate of the dNTP are in close proximity. In contrast, rNTPs have 2´-OH results in a steric clash with discriminator amino acids. This result in the α-phosphate is incorrectly aligned with the 3´-OH of the primer, dramatically reducing the rate of catalysis.





The structure of DNA polymerase described as resembling a right hand with thumb, fingers, and palm domains. The **palm domain** appears to function in catalyzing the transfer of phosphoryl groups. The **finger domain** functions to bind the dNTP with the template base.

The **thumb domain** plays a potential role in the processivity, translocation, and positioning of the DNA. The **degree of processivity** refers to the average number of nucleotides added each time the enzyme binds a template. Each DNA polymerase has a characteristic processivity that can range from only a few nucleotides to more than 50,000 bases. Some polymerases add over 50,000 nucleotides to a growing DNA strand before dissociating from the template strand, giving a replication rate of up to 1,000 nucleotides per second.

Prokaryotic DNA polymerases

Pol I

DNA polymerase I enzyme belongs to family A polymerases, which is encoded by the *polA* gene. Its ubiquitous among prokaryotes and most abundant polymerase, accounting for >95% of polymerase activity in *E. coli*. This repair polymerase is involved in excision repair with both **3'–5'** and **5'–3'** exonuclease activity and processing of Okazaki fragments generated during lagging strand synthesis. It is removed and replaced primer and has low level of processivity.

Pol II

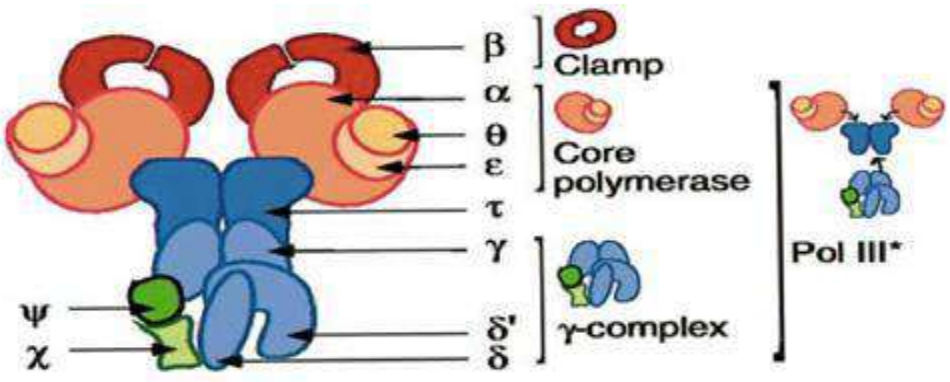
DNA polymerase II, a family B polymerase, is a polB gene product. Pol II has 3'–5' exonuclease activity and participates in DNA repair. It is a minor component of the cell during normal growth but is inducible by the SOS response and is also thought to be a support to Pol III. This enzyme participates in Base-Excision repair and Nucleotide-Excision repair.

Pol III

DNA polymerase III holoenzyme is the major enzyme involved in DNA replication belongs to family C polymerases. DNA polymerase III synthesizes base pairs at a rate of around 1000 nucleotides per second. It called holoenzyme consists of three components:

1. **pol III core**, consists of three subunits: **α**, the polymerase activity center, **ε** have 3'→5' exonuclease proofreader, and **θ**, stimulates the ε subunit's proofreading.
2. **Beta sliding clamp processivity factor**. Two β units act as sliding DNA clamps, they keep the polymerase bound to the DNA that encircule it and allow DNA pol lll to slide along the DNA allowing for high processivity.
3. **Clamp-loading complex** that assembly in a seven-subunit (τ2γδδ′χψ).

The holoenzyme contains two cores, one for each strand, the lagging and leading. The beta sliding clamp processivity factor is also present in duplicate, one for each core.



Pol IV

It is belong to Y family polymerase expressed by the dinB gene that is switched on via SOS induction. During SOS induction, Pol IV production is increased tenfold and one of the functions during this time is to interfere with Pol III holoenzyme processivity. This creates a checkpoint, stops replication, and allows time to repair DNA lesions via the appropriate repair pathway. It does not have exonuclease activity.

Pol V

DNA polymerase V (Pol V) is a Y family DNA polymerase that is involved in SOS response and translesion synthesis DNA repair mechanisms. It does not have exonuclease activity.

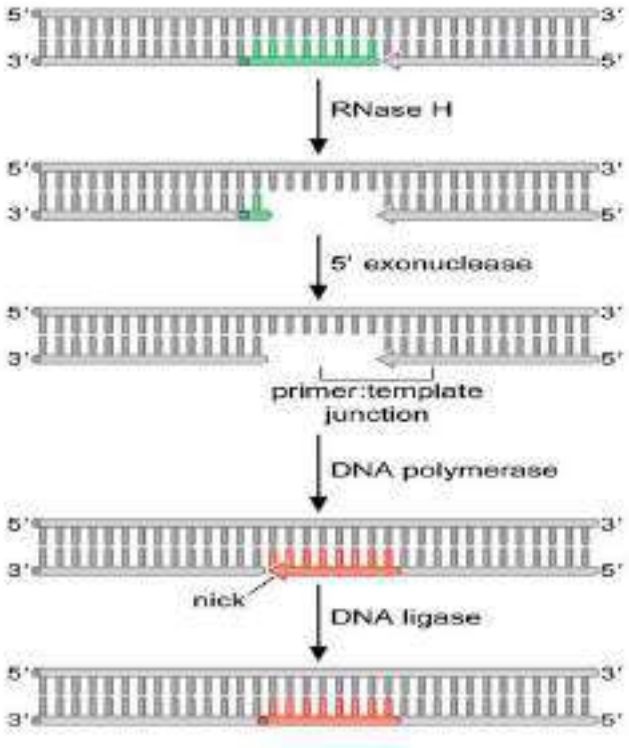
Primase (DnaG)

All DNA polymerases require a primer with a free 3´-OH.They cannot initiate a new DNA strand de novo. **Primase** is a specialized RNA polymerase dedicated to making short RNA primers from 5-10 nucleotides long) on an ssDNA template. dnaG gene coded for this enzyme. These primers are subsequently extended by DNA polymerase. Leading strand requires only a single RNA primer. In contrast, the discontinuous synthesis of the lagging strand means that new primers are needed for each

Okazaki fragment. Primase activity is dramatically increased when it associates with other enzyme acts at the replication fork called DNA helicase.

To complete DNA replication, the RNA primers must be removed and replaced with DNA nucleotides; an enzyme called **RNaseH** recognizes and removes most of each RNA primer. This enzyme specifically degrades RNA that is base-paired with DNA (the H in its name stands for "hybrid" in RNA:DNA hybrid). RNase H removes all of the RNA primer except the ribonucleotide directly linked to the DNA end. This is because RNase H can only cleave bonds between two ribonucleotides. The final ribonucleotide is removed by a 5´ exonuclease of DNA Pol I that leaves a gap in the dsDNA.

DNA Pol I fills this gap until every nucleotide is base-paired, leaving a DNA molecule that is complete except for a break in the phosphodiester backbone between the 3´-OH and 5´-phosphate of the repaired strand. This nick in the DNA can be repaired by an enzyme called **DNA ligase**. DNA ligases use high energy co-factors (such as ATP) to create a phosphodiester bond.



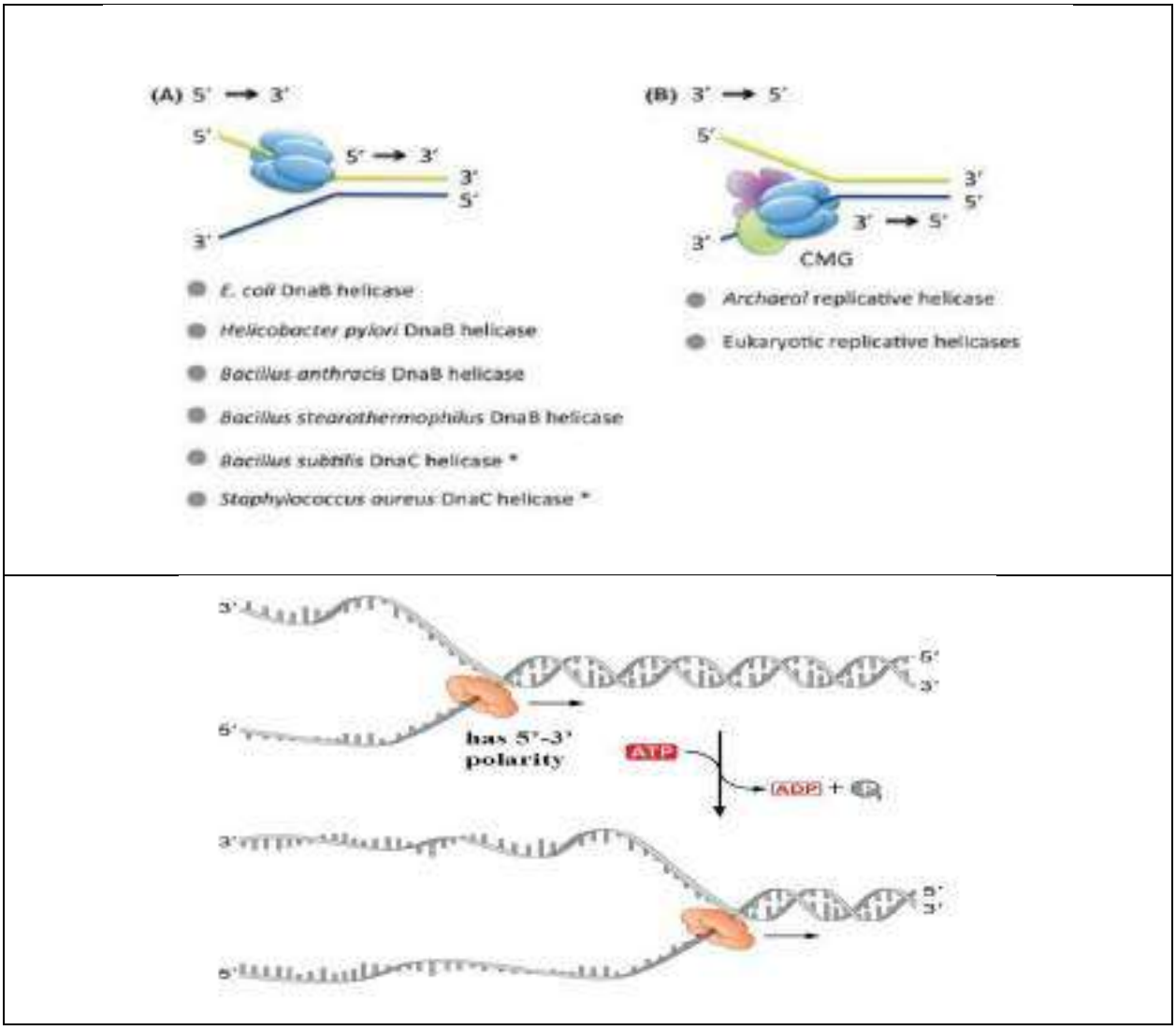
DNA helicase (DnaB)

At the replication fork, **DNA helicases** catalyze the separation of the two strands of duplex DNA through break the hydrogen bonds between the duplex helix. These enzymes bind to and move directionally along ssDNA using the energy of nucleoside triphosphate (usually ATP). Typically, DNA helicases that act at replication forks are hexameric proteins that assume the shape of a ring. These ring-shaped protein complexes encircle one of the two single strands at the replication fork.

DNA helicases exhibit specific polarity, which is defined as the direction of DNA helicase movement on initially bound ssDNA template (3´to 5´or 5´to 3´). For helicases involved in DNA replication, the

polarity of the reaction is strongly indicative of helicase placement on the leading (3´to 5) or lagging (5´to 3´) strands. Like DNA polymerases, DNA helicases act processively. The ring shaped hexameric DNA helicases exhibit high processivity because they encircle the DNA.

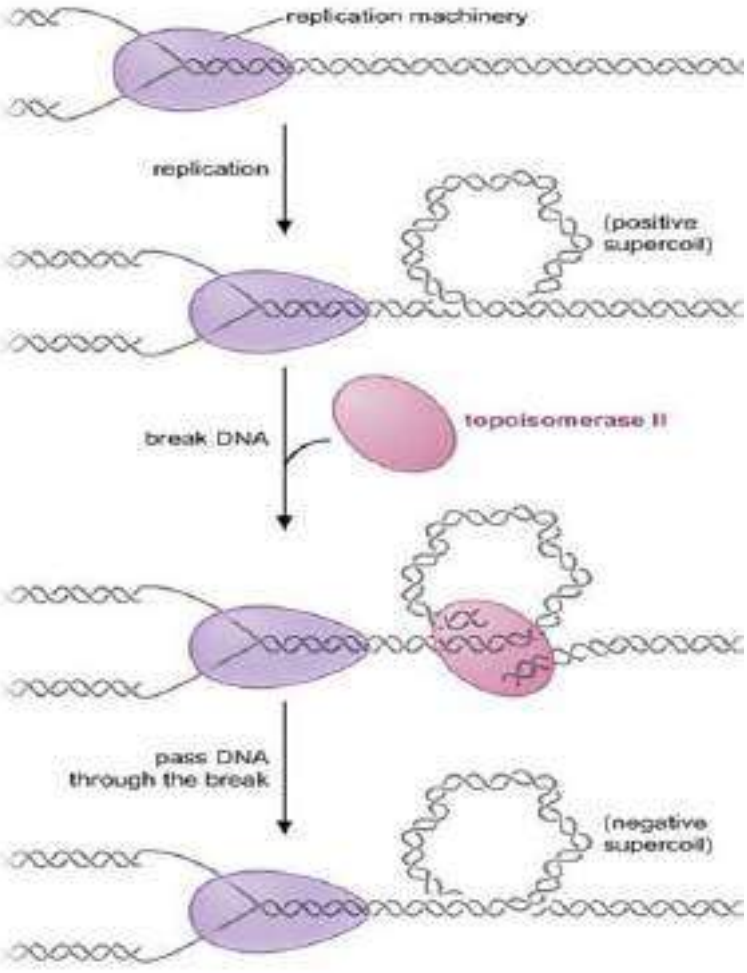
More than one helicase is present in each system because of a variety of different needs for the duplex DNA to unwind in different DNA metabolisms. For example, at least 14 different DNA helicases have been isolated from a simple single cell organism such as *E. coli*.



DNA topoisomerase

As the strands of DNA are separated at the replication fork, the dsDNA in front of the fork becomes increasingly positively supercoiled. These supercoils are removed by **topoisomerases** that act on the unreplicated dsDNA in front of the replication fork by cut the phosphate backbone of either one or both the DNA strands and at the end of these processes, the DNA backbone is resealed again.

Topoisomerase I, cuts one strand of a DNA double helix, relaxation occurs, and then the cut strand is re-ligated. Cutting one strand allows the part of the molecule on one side of the cut to rotate around the



Single-Stranded DNA-Binding Proteins Stabilize ssDNA before Replication

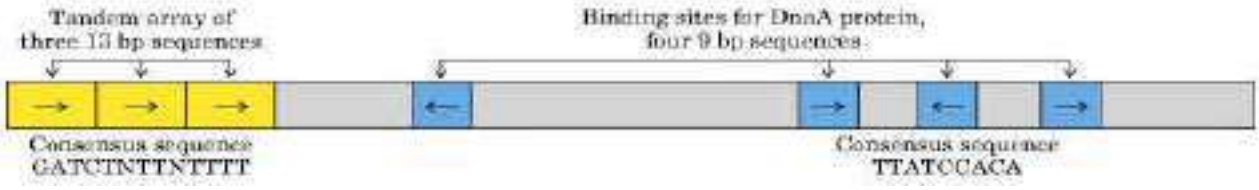
The prokaryotic single-stranded DNA binding proteins bind much more strongly to single stranded than to double-stranded DNA. They aid helicase action by binding tightly to newly formed single-stranded DNA and keeping it from annealing with its partner. By coating the single-stranded DNA, SSBs also protect it from degradation.

Binding of one SSB promotes the binding of another SSB to the immediately adjacent ssDNA which is called **cooperative binding** and occurs because the SSB–SSB interaction strongly stabilizes SSB binding to ssDNA and it inhibits the formation of intramolecular base pairs.

Prokaryotic DNA replication

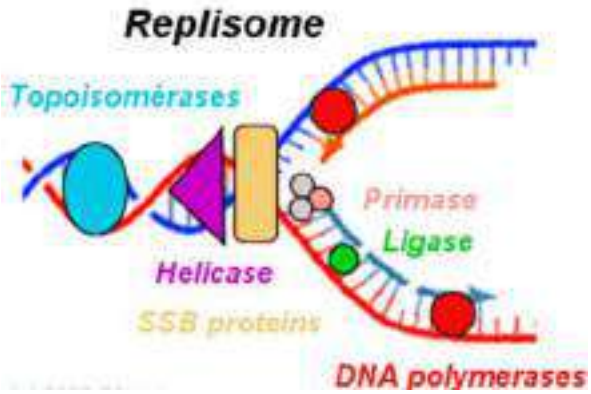
Initiation

Most circular bacterial chromosomes are replicated **bidirectionally**, starting at one point of origin and replicating in two directions away from the origin. In *E. coli,* which has a single origin of replication **(*oriC*)** on its one chromosome, it is approximately 245 base pairs long and is rich in AT sequences. The *oriC* site consists of three of 13 bp sequence are highly conserved in many bacteria and forms the consensus sequences (GATCTNTTNTTTT). Close to *oriC* site, there are four of 9 bp sequences each TTATCCACA.



The initiation stage of chromosome replication is mediated by the bacterial orisome. In all bacteria, orisomes assemble when the initiator protein, DnaA bind to a DNA *oriC* (9 bp repeats) to make nucleoprotein complex binding of DnaA leads to strand separation at the 13 bp repeats. This binding causes the DNA to loop in preparation for melting open by the helicase. A replication fork is formed when helicase separates the DNA strands at the origin of replication. The DNA tends to become more highly coiled ahead of the replication fork. Topoisomerase breaks and reforms DNA's phosphate backbone ahead of the replication fork, thereby relieving the pressure that results from this supercoiling. Single-strand binding proteins bind to the single-stranded DNA to prevent the helix from re-forming. Primase synthesizes an RNA primer.

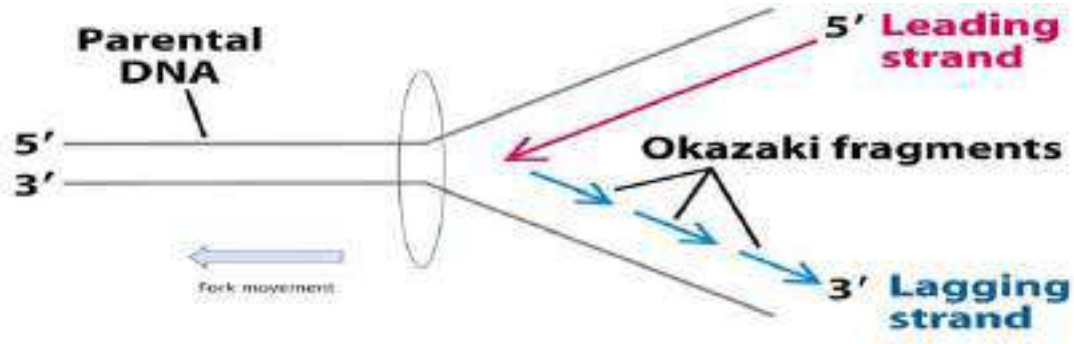
REPLISOME: It is the multiprotein structure that assembles at the bacterial replicating fork to undertake synthesis of DNA. It contains DNA polymerase and other enzymes.



Elongation

Once priming is complete, DNA polymerase III holoenzyme is loaded into the DNA and replication begins. The catalytic mechanism of DNA polymerase III involves the use of two metal ions in the active site, and a region in the active site that can discriminate between deoxyribonucleotides and ribonucleotides. Furthermore, DNA polymerase III must be able to distinguish between correctly paired bases and incorrectly paired bases.

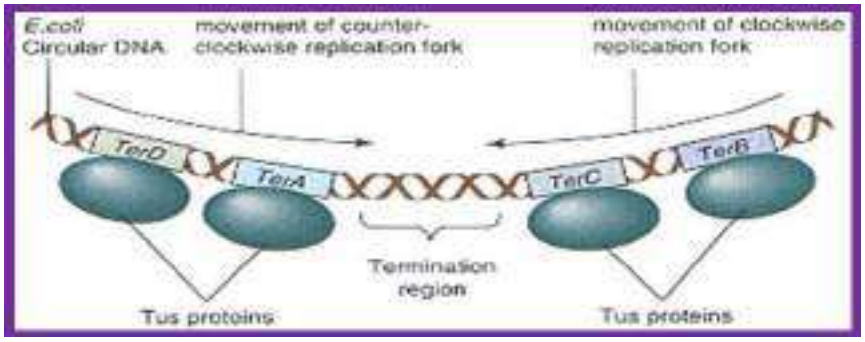
The replication fork moves at the rate of 1000 nucleotides per second. DNA polymerase can only extend in the 5' to 3' direction. One strand, which is complementary to the 3' to 5' parental DNA strand, is synthesized continuously **towards the replication fork** called **leading strand** because the polymerase can add nucleotides in this direction. The other strand is the **lagging str**and that is complementary to the 5' to 3' parental DNA. It is **extended away from the replication fork** in small fragments known as Okazaki fragments (1000-2000 nucleotides), each requiring a primer to start the synthesis. The RNA primers of Okazaki fragments are subsequently degraded by RNase H and DNA Polymerase I (exonuclease), and the gaps (or nicks) are filled with deoxyribonucleotides and sealed by the enzyme ligase.



Termination

The termination region (ter) of E.coli is a ~350 kbp sequence 180˚ from Ori C. *ter* elements are asymmetric patterns of DNA that act as protein binding sites. It contains six sequences, TerA to TerF, which are binding sites for terminal utilization protein (Tus protein). The binding of specific proteins to *ter* elements provides a trap for the going on replication fork, catching the replication fork as it passes

by blocking the helicase.. There are several *ter* elements responsible for stopping each replication fork, with each of these elements being specific for the fork passing in one direction only. The *ter*-protein complex responsible for catching the clockwise replication fork will allow the anticlockwise fork to proceed unchecked, until it is stopped by its own anticlockwise facing *ter* element fork trap. The clockwise replication is stopped at ter B,C,F and counterclockwise replication at ter A,D or E. The process is complete when synthesis from the opposite direction reaches the stopped strand. The Topoisomerase IV separates the replicative strands.

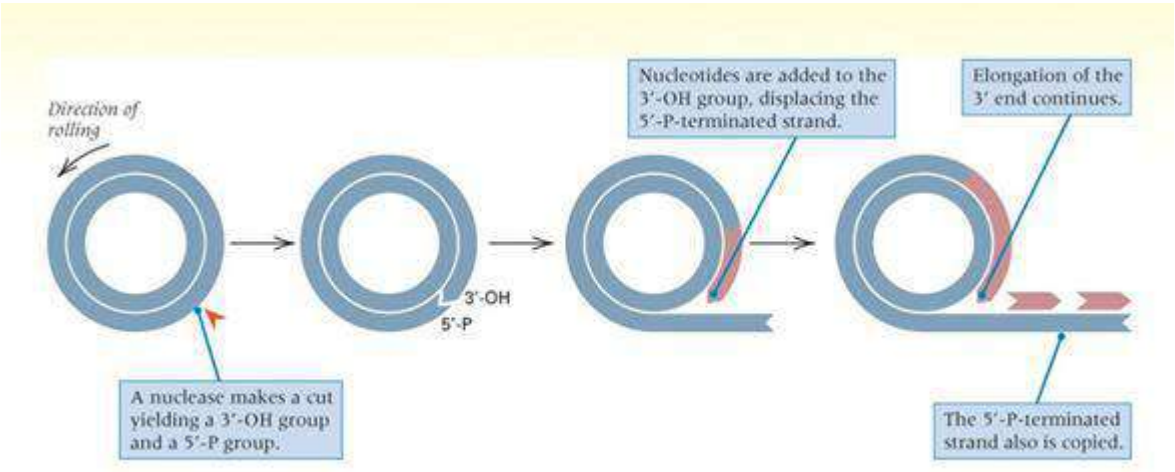


Rolling circle model replication:

In the rolling circle model of replication, a nick is made in one of the strands of the circular DNA, resulting in replication of circle and a tail. It describes a process of **unidirectional** nucleic acid replication that can rapidly synthesize multiple copies of circular molecules of DNA or RNA, such as plasmids (F plasmid or *E.coli* Hfr chromosome during conjugation). The F^+ or Hfr cell retains the circular daughter while passing the linear tail into the F^- cell), the genomes of bacteriophages, and the circular RNA genome of viroids.

A typical DNA rolling circle replication has five steps:

1. Circular dsDNA will be nicked.
2. The 3' end is elongated using "unnicked" DNA as leading strand (template); 5' end is displaced.
3. Displaced DNA is a lagging strand and is made double stranded via a series of Okazaki fragments.
4. Replication of both "unnicked" and displaced ssDNA.
5. Displaced DNA circularizes.

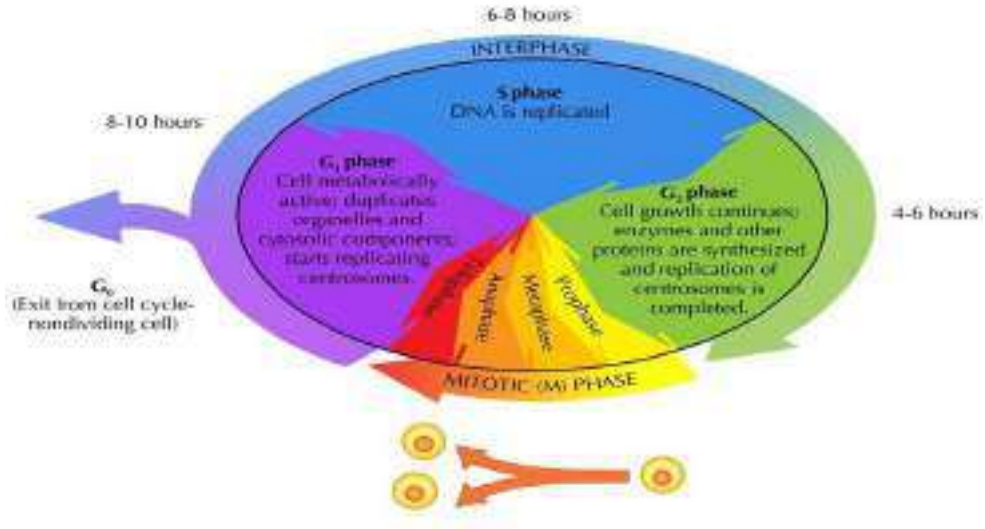


Lec4

Eukaryotic DNA replication

Is a conserved mechanism that restricts DNA replication to once per cell cycle. Eukaryotic DNA replication of chromosomal DNA is central for the duplication of a cell and is necessary for the maintenance of the eukaryotic genome. Replication processes permit the copying of a single DNA double helix into two DNA helices, which are divided into the daughter cells at mitosis. The major enzymatic functions carried out at the replication fork are well conserved from prokaryotes to eukaryotes, but the replication machinery in eukaryotic DNA replication is a much larger complex, coordinating many proteins at the site of replication, forming the replisome.

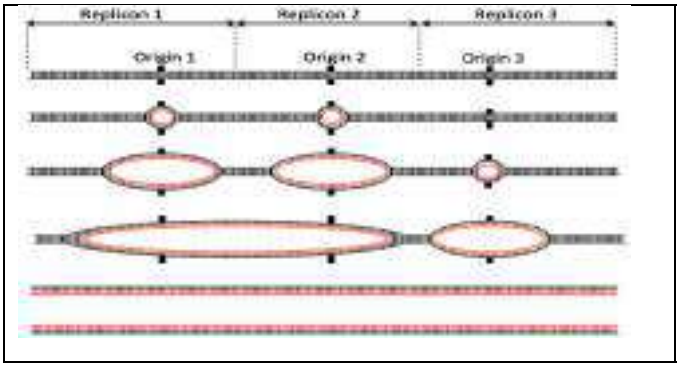
Much of the cell cycle is built around ensuring that DNA replication occurs without errors. In G_1 phase of the cell cycle, many of the DNA replication regulatory processes are initiated. In eukaryotes, the vast majority of DNA synthesis occurs during S phase of the cell cycle, and the entire genome must be unwound and duplicated to form two daughter copies. During G_2, any damaged DNA or replication errors are corrected. Finally, one copy of the genomes is segregated to each daughter cell at mitosis or M phase.

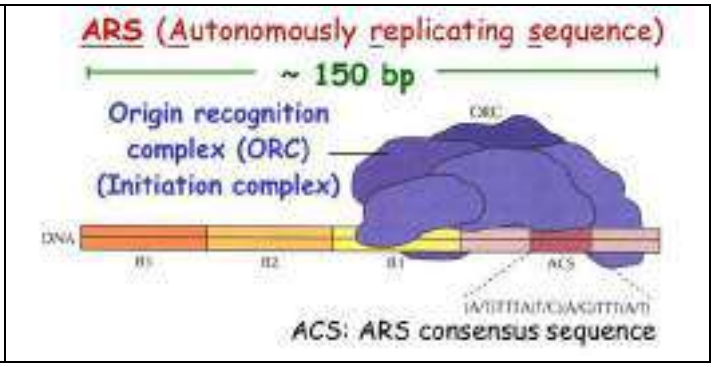


Multiple Origins of Replication in Eukaryotes

The DNA duplex in a eukaryotic chromosome is linear and replicates **bidirectionally.** Furthermore, replication is initiated at many sites in the DNA. Origins of replication are typically assigned names containing ***ori*.** There are significant differences between the replication of origin in eukaryotes and prokaryotes:

- **Bacteria** have a **single** circular molecule of DNA, and typically only a **single** replication origin per circular chromosome.
- **Archaea** have a **single** circular molecule of DNA and **several** origins of replication along this circular chromosome. The origins are generally AT-rich area that vary based on the archaeal species. The singular archaeal ORC protein recognizes the AT-rich area and binds DNA in an ATP-dependent fashion.
- **Eukaryotes** often have **multiple** origins of replication on each linear chromosome that initiate at different times. The large genome sizes of eukaryotic cells, which range from 12 Mbp in *S. cerevisiae* to 3 Gbp in humans, necessitates that DNA replication starts at several hundred (in budding yeast) to tens of thousands (in humans) origins to complete DNA replication of all chromosomes during each cell cycle. In eukaryotes, the budding yeast *Saccharomyces cerevisiae* has the best-characterized replication origins name Autonomously replicating sequences or ARS that recruits replication proteins. In other eukaryotes, including humans, the DNA sequences at the replication origins vary. Despite this sequence variation, all the origins form a base for assembly of a group of proteins known collectively as the pre-replication complex (pre-RC).



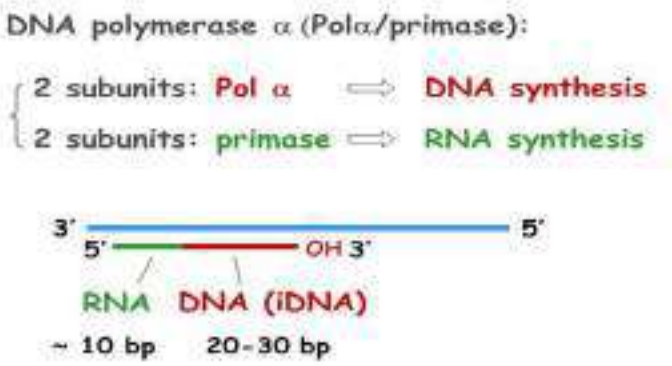


Multiple Eukaryotic DNA Polymerases

Although DNA polymerases all share the same basic catalytic mechanism, eukaryotic cells contain at least 15 distinct polymerases At least three polymerases (α, δ and ε) are needed for chromosome replication, and even more polymerase diversity seems to be needed for DNA repair. Some polymerases consist of a single polypeptide chain, whereas others, such as those involved in chromosome replication, are composed of several different subunits. Eukaryotic DNA polymerases fall into five families, designated A; B (which includes all the eukaryotic polymerases involved in chromosome replication); RT (reverse transcriptase), X and Y, which include enzymes involved in DNA repair or specialized types of replication.

DNA polymerase α (DNA polymerase alpha)

DNA polymerase α holoenzyme has both **primase** and **polymerase** activity. It is involved in the primer synthesis during DNA replication. Primase activity synthesizes ~12-nt RNA primers. The polymerase activity extends the primer to some 20-30 nucleotide lengths, after that **polymerase switching** takes place. In polymerase switching polymerase α is displaced from the template and the synthesis is taken over by polymerase δ and polymerase ε. The polymerase switching is required as polymerase α has low processivity.



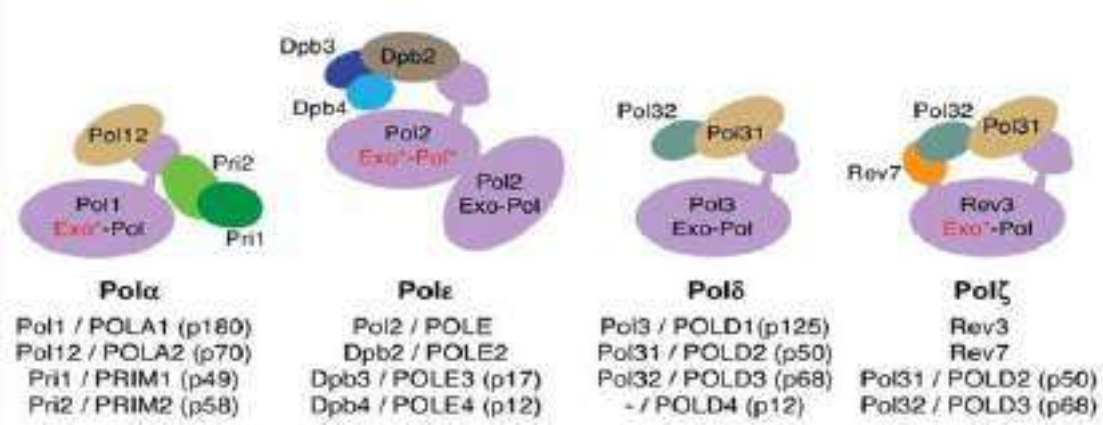
DNA polymerase δ (DNA polymerase delta)

During polymerase switching, **Replication factor C** (RF-C) displaces polymerase α and load Sliding clamps **PCNA** (Proliferating cell nuclear antigen. They are Multi-subunit proteins that prevent the dissociation of the polymerase from template by encircle the double helix of DNA. Act as a sliding platform to recruit other replication proteins such as DNA helicase, nuclease, ligase, and histone chaperones) on the template DNA near the primer strand, following which polymerase δ binds to the PCNA and extends the DNA strand. Pol δ in complex with PCNA is required for **lagging strand synthesis.**

Name of Major polymerase	Family	Activities	Function	Location
α polymerase	B Family	Polymerase Primase 3' to 5' exonuclease	Primer synthesis Repair	Nucleus
Y polymerase	A Family	Polymerase 3' to 5' exonuclease	Mitochondrial DNA replication	Mitochondria
δ polymerase	B Family	Polymerase 3' to 5' exonuclease	Lagging strand synthesis Elongation DNA repair	Nucleus
ε polymerase	B Family	Polymerase 3' to 5' exonuclease 5' to 3' exonuclease	Leading strand synthesis Gap filling on lagging strand	Nucleus

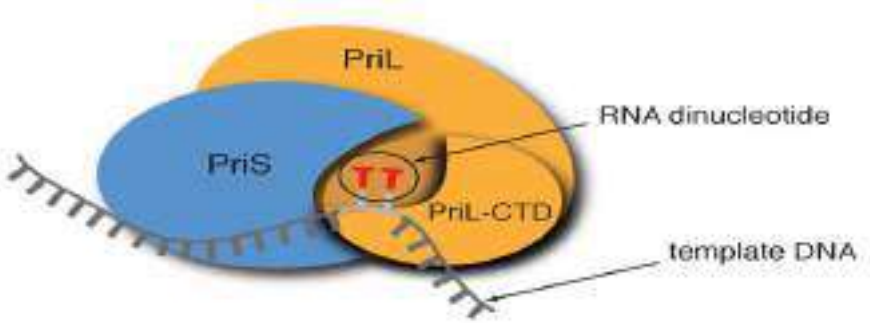
DNA polymerase ε (DNA polymerase epsilon)

PCNA interact with DNA polymerase ε to facilitate **leading strand synthesis,** has a 3' to 5' exonuclease activity. The processivity of DNA polymerase is increased by their association with the sliding clamps. Mitochondria have their own DNA polymerase, **polymerase γ,** which replicates the mitochondrial DNA. **polymerase β** is not processive at all. It usually adds only one nucleotide to a growing DNA chain and then falls off, requiring a new polymerase to bind and add the next nucleotide have a role as a repair enzyme.



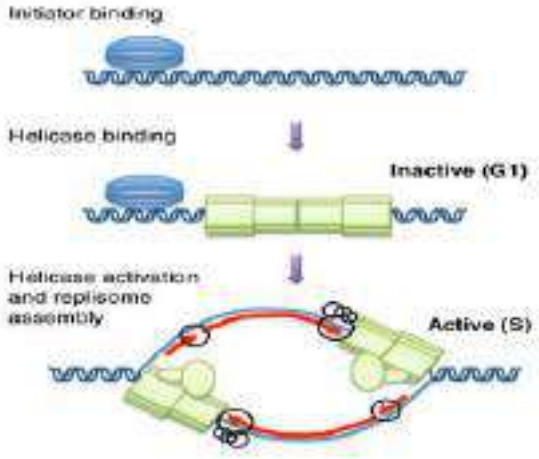
Primase

In bacteria, primase is a single-chain multi-domain polymerase known as DnaG, which interacts directly with the DnaB helicase at the fork to prime DNA synthesis. In archaea and eukaryote, primase is a constitutive heterodimer of a **catalytic subunit**, PriS (Pri1), and a **regulatory subunit**, PriL (Pri2). In eukaryotic replication, primase activity is tightly coupled to that of DNA polymerase α to form the Pol α/primase complex. PriL is composed of two structural domains. Several functions have been proposed for PriL such as stabilization of the PriS, involvement in synthesis initiation, improvement of primase processivity, determination of product size and transfer of the products to DNA polymerase alpha.



Helicase

The catalytic core of the helicase is composed of **six minichromosome maintenance (Mcm2-7) proteins**, forming a heterohexameric ring. The Mcm2-7 forms a ring that encircles the leading stand and tracks along it, and it may seem similar to the prokaryotic DnaB homohexamer in this way, which surrounds the lagging strand.

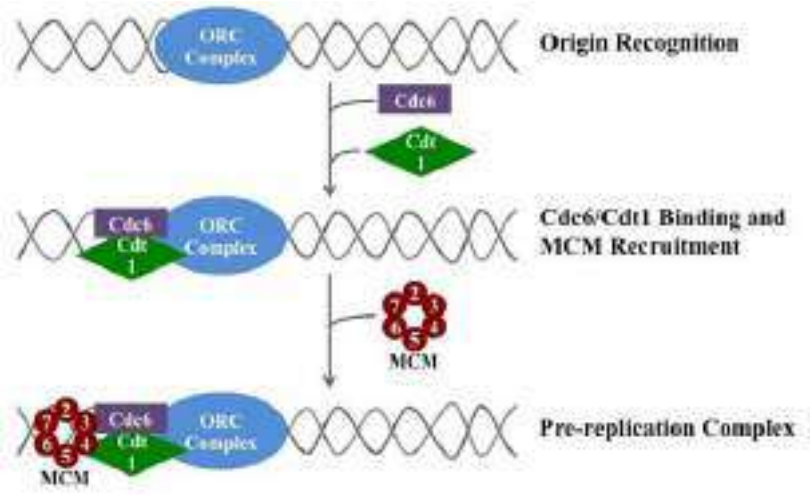
	In eukaryotes, the initiator proteins (origin recognition complex Orc1-6) loads the replicative DNA helicase (Mcm2-7) in an inactive, double-hexameric form around double-stranded DNA (dsDNA) during the G1-phase. Activation of the helicase is occurs during S-phase by recruitment of Cdc45 and 4-subunit GINS to form the active Cdc45/Mcm2-7/GINS (CMG) holo-helicase at DNA replication forks. Since CMG translocates in the 3' to 5' direction along the DNA, it encircles the leading strand template.

Replication protein A (**RPA**) is the major protein that binds to single-stranded DNA (ssDNA) in eukaryotic cells. is a protein of heterotrimer composed of three tightly associated subunits (RPA1, RPA2 and RPA3). In vitro, RPA shows a much higher affinity for ssDNA than RNA or double-stranded DNA. During DNA replication, RPA prevents single-stranded DNA (ssDNA) from winding back on itself or from forming secondary structures. This keeps DNA unwound for the polymerase to replicate it.

Initiation

Initiation of eukaryotic DNA replication is the first stage of DNA synthesis. DNA replication is initiated from specific sequences called origins of replication, and eukaryotic cells have multiple replication origins. To initiate DNA replication, multiple replicative proteins assemble on and dissociate from these replicative origins. The individual factors described below work together to direct the formation of the pre-replication complex (pre-RC). **In** bacteria, the main component of the pre-RC is DnaA. The pre-RC is complete when DnaA occupies all of its binding sites within the bacterial origin of replication (oriC). **The** eukaryotic **pre-RC** is the most complex and highly regulated pre-RC. In most eukaryotes it is composed of six origin recognition complex proteins (ORC1-6), Cdc6, Cdt1, and Mcm2-7.

After the pre-RC is formed it must be activated and the replisome assembled in order for DNA replication to occur. At the transition of the G_1 stage to the S phase of the cell cycle, S phase–specific cyclin-dependent protein kinase (CDK) and DDK transform the pre-RC into an active replication fork. During this transformation, the pre-RC is disassembled with the loss of Cdc6, creating the initiation complex. Cell division cycle 45 (Cdc45) protein is also essential for initiating DNA replication. Cdc45 then recruits key components of the replisome; the replicative DNA polymerase α and its primase. DNA replication can then begin.

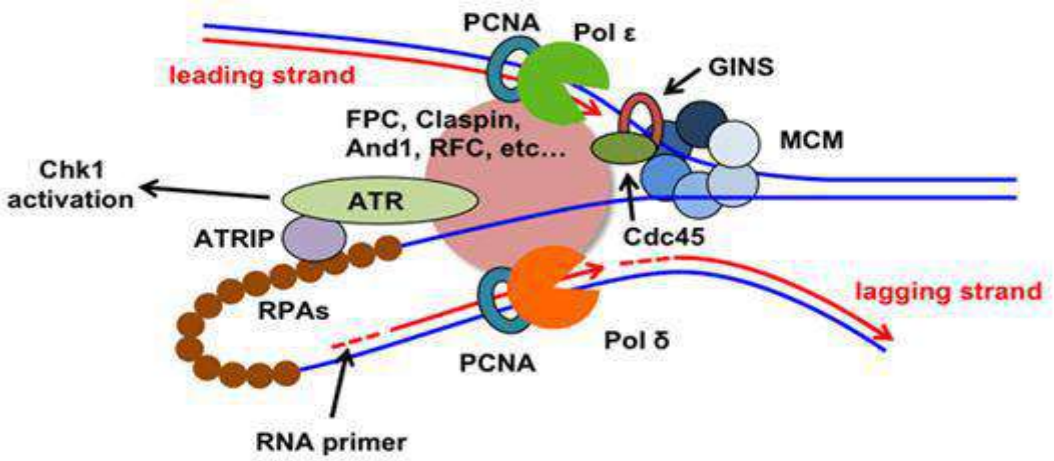


Elongation

Replication on the leading and lagging strands is performed by DNA polymerase ε and δ. Many replisome factors including Claspin, And1, replication factor C clamp loader are responsible for regulating polymerase functions and coordinating DNA synthesis with the unwinding of the template strand by Cdc45-Mcm-GINS complex. Topoisomerases are responsible for removing the positive supercoils ahead of the replication fork.

The DNA polymerase move along single-stranded DNA and allow for the extension of the nascent DNA strand by reading the template strand and allowing for incorporation of the proper purine and pyrimidine bases. The free dNTPs nucleotides are added to an exposed 3'-hydroxyl group on the last incorporated nucleotide. In this reaction, a pyrophosphate is released from the free dNTP, generating energy for the polymerization reaction and exposing the 5' monophosphate, which is then covalently bonded to the 3'

oxygen. Two replicative polymerases synthesize DNA in opposite orientations. Polymerase ε synthesizes DNA on the "leading" DNA strand continuously as it is pointing in the same direction as DNA unwinding. In contrast, polymerase δ synthesizes DNA on the "lagging" strand, which is the opposite DNA template strand in discontinuous manner.



The discontinuous stretches of DNA replication products on the lagging strand are known as Okazaki fragments and are about 100 to 200 bases in length at eukaryotic replication forks. Each Okazaki fragment is preceded by an RNA primer, which is displaced by the procession of the next Okazaki fragment during synthesis. RNase H recognizes the DNA:RNA hybrids that are created by the use of RNA primers and is responsible for removing these from the replicated strand. DNA polymerase α, recognizes these sites and elongates the breaks left by primer removal. DNA synthesis is complete once all RNA primers are removed and nicks are repaired.

Termination

During DNA replication, the double stranded DNA is unwound and DNA polymerase synthesizes new strands. However, as DNA polymerase moves in a unidirectional manner (from 5' to 3'), only the leading strand can be replicated continuously. In the case of the lagging strand, DNA replication is discontinuous. The **final RNA primer** synthesized on the lagging-strand template cannot be replaced by DNA because there is **no 3′-OH end** available for the repair polymerase. Therefore, at 5' end of each daughter strand there is **a gap**. This gap must be filled before next round of replication. Without a mechanism to deal with this problem, DNA would be lost from the ends of all chromosomes each time a cell divides.

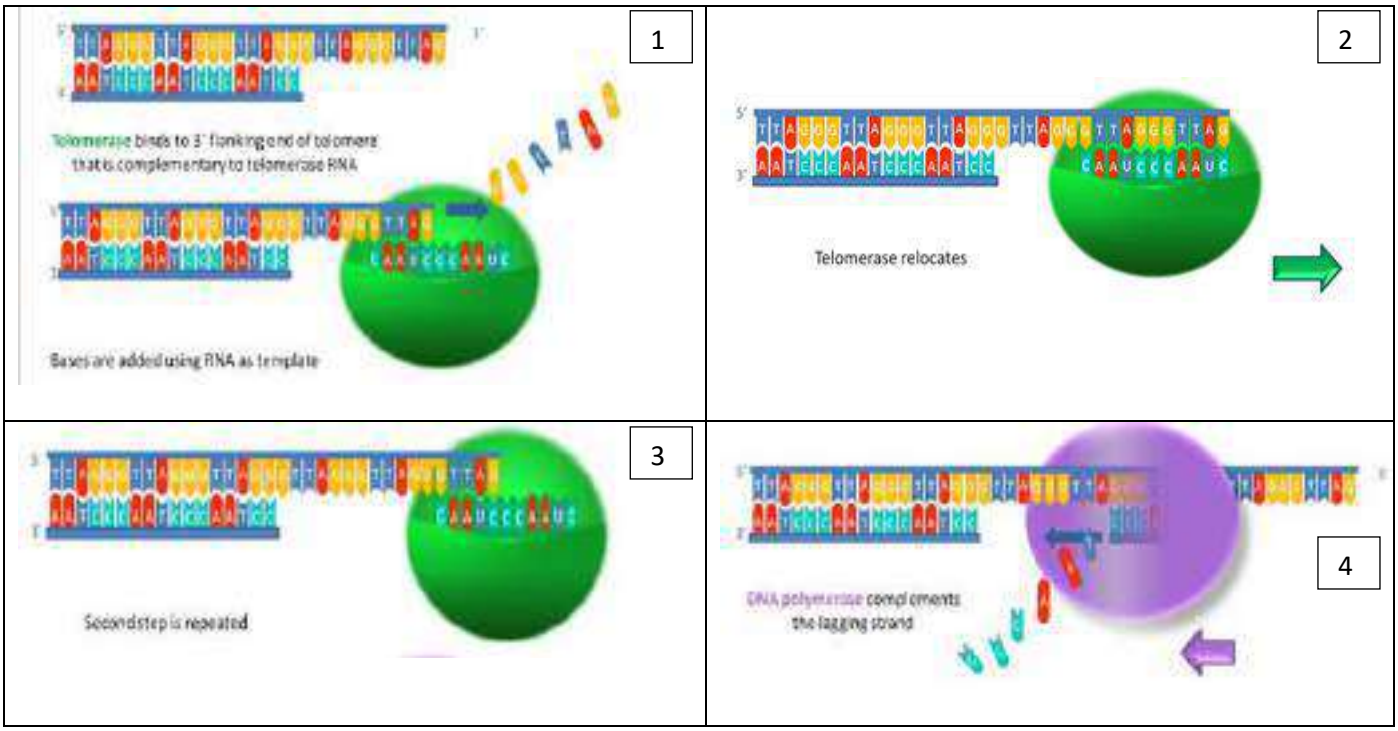
Bacteria solve this "end-replication" problem by having circular DNA molecules as chromosomes. Eukaryotes solve it in a different way: they have specialized nucleotide sequences at the ends of their chromosomes that are incorporated into structures called **telomeres**.

Telomere has G:C rich repeats. These repeats of telomere sequence is different among different organisms. Telomere in human cell consists of repeats of TTAGGG. These telomere sequence do not codes anything but it is essential to **fill the gap in daughter strand** and **maintain the integrity of DNA by protect the end of the chromosome from DNA damage or from fusion with neighbouring chromosomes**.

There is an enzyme found in eukaryotic cell called **telomerase**. Telomerase is a DNA polymerase (RNA dependent DNA polymerase) which adds many copies of telomere sequence at 3'-OH end of template strand. Like other DNA polymerase, terlomerase also adds dNTPs at 3'-OH end. Unlike other DNA polymerase, telomerase adds DNA at 3'-OH end of parent strand not at the daughter strand and also it synthesizes the same sequences over and over in absence of template strand.

Telomerase is a nucleoprotein complex consists of two molecules telomerase reverse transcriptase **(TERT)** and telomerase RNA **(TR)**. TERT is a reverse transcriptase, which is creates single-stranded DNA using single-stranded RNA as a template. TR can add a six-nucleotide repeating sequence, TTAGGG (vertebrates) to the 3' strand of chromosomes. These TTAGGG repeats are called telomeres. The template region of TERT is 3'-AAUCCCCAA-5'. Telomerase is active in normal stem cells, in gametes and most cancer cells, but is normally absent from, or at very low levels in, most somatic cells.

Again telomerase translocates and adds GGGTTA sequence. This process is continued for many time. The parent strand become more longer than daughter strand. Now RNA polymerase (Primase) synthesize RNA primer by copying the parent strand in 5'-3' direction using telomere sequence as template. The DNA polymerase can now extend the primer in 5'-3' direction by adding dNTPs to 3' end. The primer is now removed and it won't be replaced because it is an extra sequence added by copying telomere sequence. Finally the integrity of daughter strand is maintained.



Nucleosomes Assembly

During S phase of the cell division cycle, not only is the entire genomic DNA replicated, but the underlying chromatin structure has to be duplicated as well. It is widely assumed that the chromatin structure is transiently disrupted during passage of a replication fork, and that new nucleosomes are subsequently assembled on the emerging DNA daughter strands.

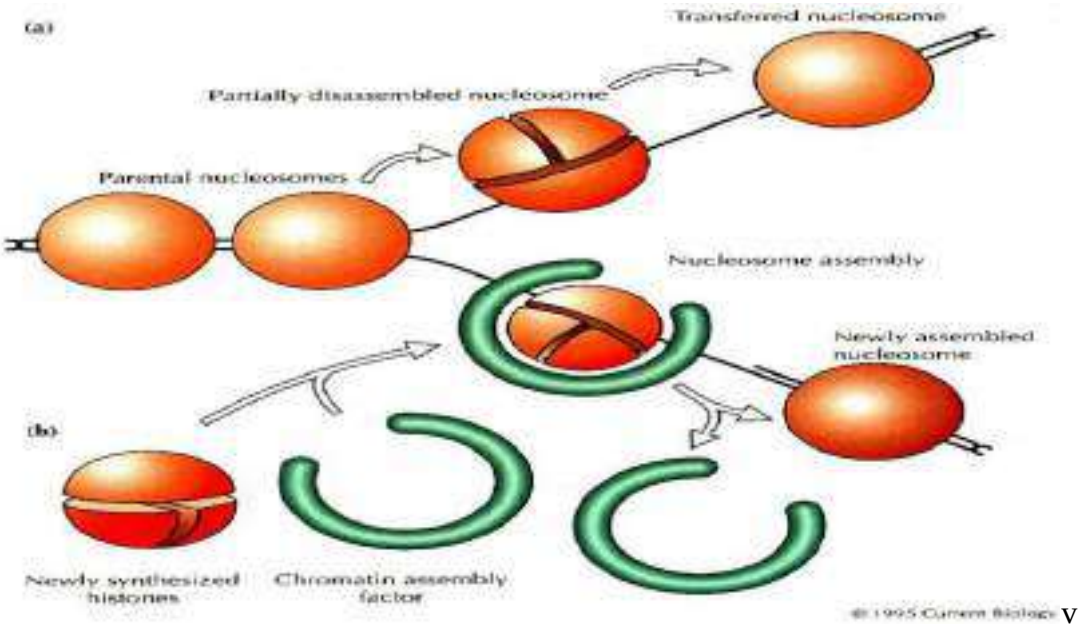
The formation of chromatin during replication consists of two distinct reactions.

In one reaction, histones from the parental nucleosomes are transferred directly onto the replicated DNA and reassembled into nucleosomes, apparently without preference for either of the two DNA daughter strands. This kind of reaction seems to be passively induced by the replication fork.

In second reaction, new nucleosomes are assembled on DNA daughter strands from soluble histone proteins which are newly synthesized during S phase. This latter reaction is mediated by chromatin assembly factor 1 (CAF-1).

Histones H3 and H4 from disassembled old nucleosomes are kept in the vicinity and randomly distributed on the newly synthesized DNA. They are assembled by the chromatin assembly factor-1 (CAF-1) complex. Newly synthesized H3 and H4 are assembled by the replication coupling assembly factor (RCAF).

The old H2A and H2B histone proteins are released and degraded; therefore, newly assembled H2A and H2B proteins are incorporated into new nucleosomes. H2A and H2B are assembled into dimers which are then loaded onto nucleosomes by the nucleosome assembly protein-1 (NAP-1).



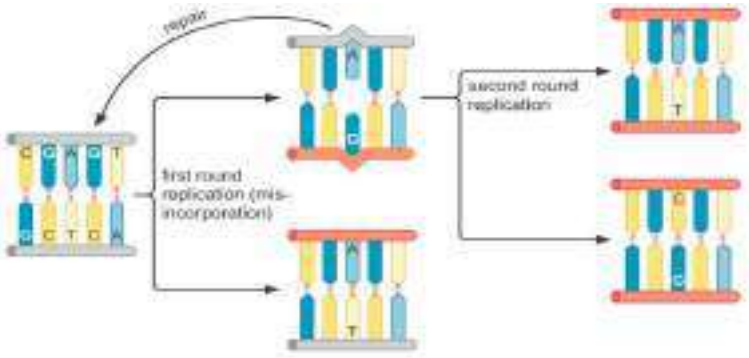
The Mutability and Repair of DNA

Organisms can survive only if their DNA is replicated faithfully and is protected from chemical and physical damage that would change its coding properties and lead to mutations. The genome is under constant attack from endogenous metabolic processes and exogenous environmental factors that can alter its chemical structure. DNA lesions consist of single strand breaks, double strand breaks, inter- and intra-strand crosslinks and base modifications, as well as oxidation and alkylation of bases. DNA damage can lead to multiple lesions including mutations, deletions, insertions, translocations, and loss of chromosomes and essential genetic information. Three important sources of mutation:

1. Inaccuracy in DNA replication.
2. Chemical and physical damage to the genetic material.
3. Transposons are segments of DNA that can move around to different positions in the genome of a single cell. In the process, they may cause mutations and increase (or decrease) the amount of DNA in the genome.

Proofreading

The 3’-5’ exonuclease of DNA Pol has proofreading activity which removes wrongly incorporated nucletides. The proofreading exonuclease is not foolproof. Some misincorporated nucleotides escape detection and become a mismatch between newly synthesized strand and the template strand. In the second round the mutation becomes permanently incorporated in the DNA sequence.



Mismatch repair

Mismatch repair system can detect mismatches that escape from proofreading and repair them. There are two challenges:

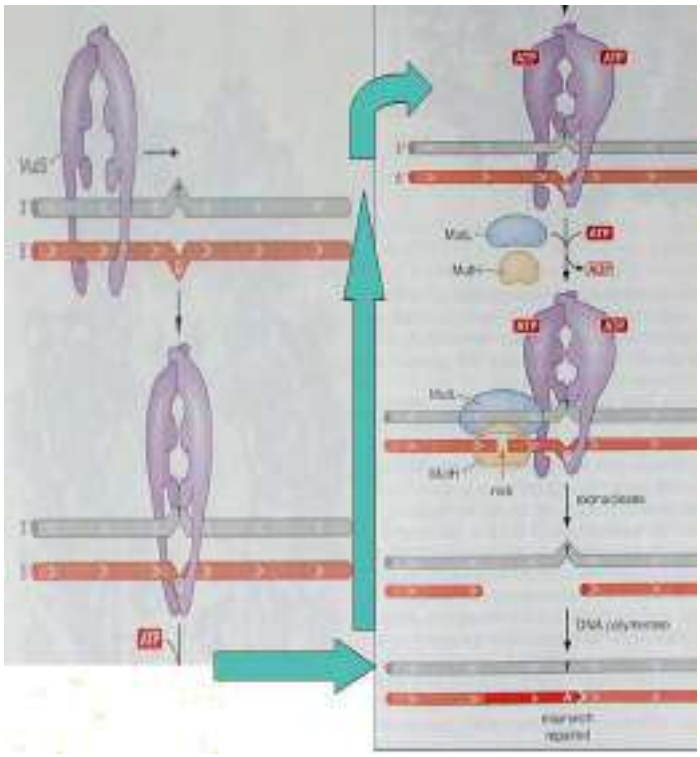
1. It must scan the genome for mismatches, it will rapidly find and repair.
2. The system must correct the mismatches accurately, it must replace the misincorporated nucleotide in the newly synthesized stand not the parental strand.

In *E.coli*, mismatches are detected by a dimer of the mismatch repair protein MutS. MutS scans the DNA, recognizing mismatches from the distortion they cause in the DNA backbone. MutS surround the mismatch-containing DNA, and a conformational change in MutS itself. MutS has an ATPase activity that is required for mismatch repair.

The complex of MutS and the mismatch-containing DNA recruits MutL, a second protein component of the repair system. MutL, in turn, activates MutH, an enzyme that causes a nick on one strand near the site of the mismatch. Nicking is followed by the action of helicase and exonucleases. This action produces a single-strand gap, which is then filled in by DNA polymerase III and sealed with DNA ligase.

The *E. coli* tags the parental strand by transient hemimethylation. The newly synthesized strand is not methylated by Dam methylase in a few minutes after the synthesis.

In eukaryotes, The main task of mismatch repair is to remove base mismatches and small insertion/deletion loops (IDL) introduced during replication. In yeast, single base mismatches are recognized by MUTSα (MSH2/ MSH6) and IDL are sensed by MUTSβ. PCNA protein is also engaged in mismatch repair, maybe supporting the damage detection and strand discrimination steps. Another complex named MUTLα binds both MUTSα and MUTSβ to promote their efficient binding to mismatches. Finally, EXO1 removes these regions and gaps filling and closing are completed by DNA polymerase and DNA ligase, respectively. The inactivation of human mismatch repair homologous proteins is cause of hereditary colorectal cancer.



DNA damage

Mutations arise not only from errors in replication, but also from damage to the DNA. Some damage is caused by environmental factors such as radiation and mutagens. DNA also undergoes spontaneous damage from the action of water. The most frequent and important kind of hydrolytic damage is **spontaneous deamination** of base **cytosine** by generating the unnatural base U in DNA. U pairs with A, instead of G that would have been directed by C.

A and G are also subject to **spontaneous deamination** (converts adenine to hypoxanthine, and guanine is converted to xanthine,).

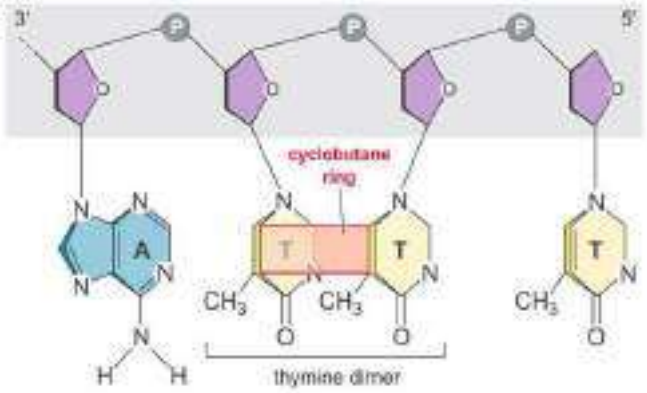
DNA also undergoes **depurination** by spontaneous hydrolysis of the N-glycosyl linkage, and this produces an abasic site (i.e., deoxyribose lacking a base) in the DNA. All of these hydrolytic reactions result in alterations to the DNA that are unnatural. This allows changes to be recognized by the repair systems.

In **alkylation**, methyl or ethyl groups are transferred to reactive sites on the bases and to phosphates in the DNA backbone such as **nitrosamines**.

DNA is also subject to attack from reactive oxygen species (e.g., $O^{-}2$, H2O2, and OH) cause **oxidation** to the nitrogen base. These potent oxidizing agents are generated by ionizing radiation and by chemical agents that generate free radicals.

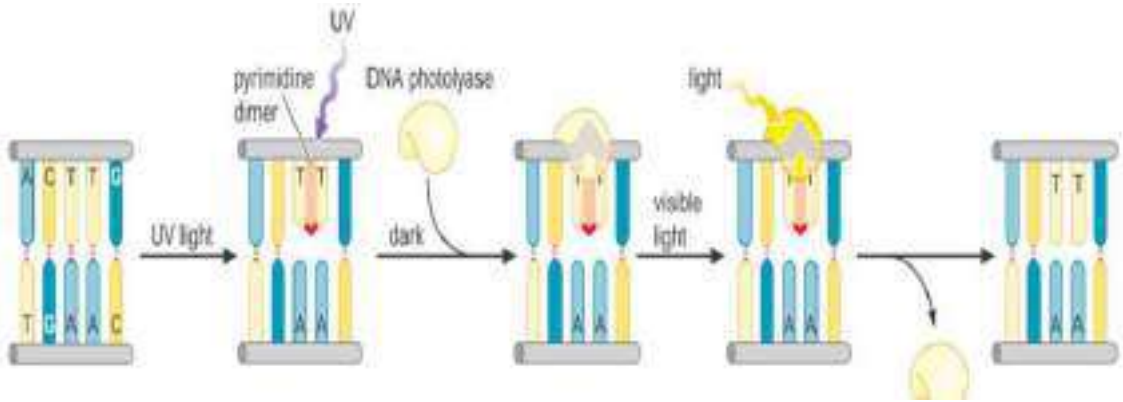
Another type of damage to bases is caused by **ultraviolet light**. Radiation with a wavelength of about 260 nm is strongly absorbed by the bases. One consequence is the photochemical fusion of two pyrimidines

that occupy adjacent positions on the same polynucleotide. Gamma radiation and X-rays are particularly hazardous because they cause double-strand breaks in the DNA which are difficult to repair.



Repair of DNA damage

Repair by simple reversal of damage is **photoreactivation.** Photoreactivation directly reverses the formation of pyrimidine dimers that result from ultraviolet irradiation. In photoreactivation, the enzyme DNA photolyase captures energy from light and uses it to break the covalent bonds linking adjacent pyrimidines.



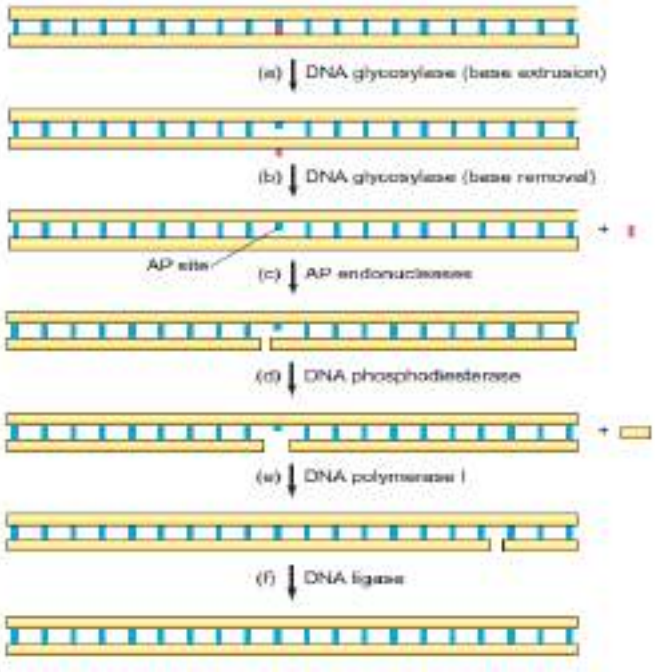
Another repair system is **excision repair systems**, in which the damaged nucleotide is not repaired but removed from the DNA. In excision repair systems, the other, undamaged, strand serves as a template.

Recombinational repair, which is used when both strands are damaged, as when the DNA is broken. In such situations, one strand cannot serve as a template for the repair of the other. Hence, in recombinational repair (known as double-strand break repair), sequence information is retrieved from a second undamaged copy of the chromosome.

Base-Excision repair

Base –Excision Repair mainly repairs non-bulky lesions produced by alkylation, oxidation or deamination of bases. In *E. coli*:

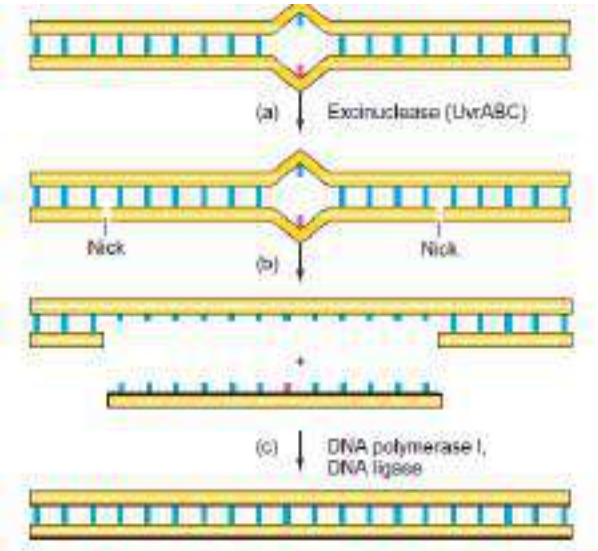
(a) DNA glycosylase releases the damaged base (red).
(b) DNA glycosylase removes the extruded base, leaving an apurinic or apyrimidinic site on the bottom DNA strand.
(c) An AP endonuclease cuts the DNA on side of the AP site.
(d) DNA phosphodiesterase removes the AP-deoxyribose phosphate (yellow block at right) that was left by the DNA glycosylase,
(e) DNA polymerase I fills in the gap and continues repair synthesis for a few nucleotides downstream, degrading DNA and simultaneously replacing it.
(f) DNA ligase seals the nick left by the DNA polymerase



Nucleotide excision repair in *E. coli*

The main function of Nucleotide excision repair pathway is to remove photoproducts induced by ultraviolet light (UV) and other bulky lesions, such as inter- and intra-strand crosslinks. In *E. coli*:

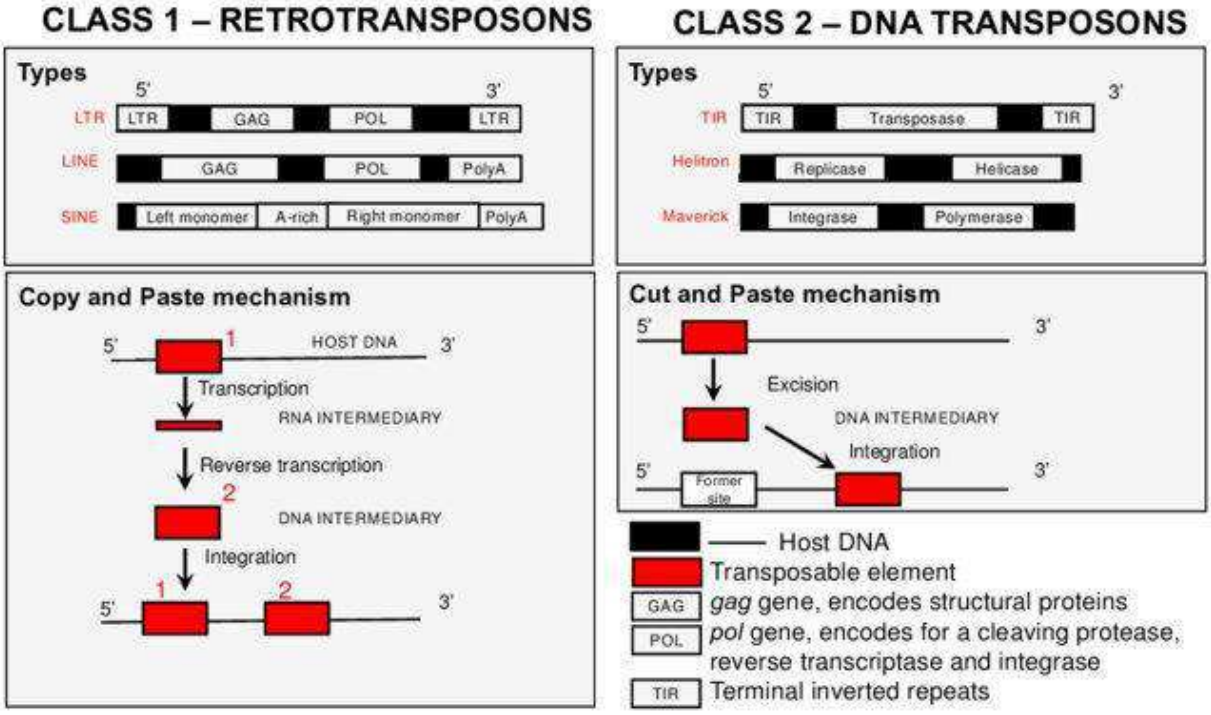
(a) The UvrABC excinuclease cuts on either side of a bulky damaged base (red).
(b) This causes removal of an oligonucleotide 12 nt long.
(c) DNA polymerase I fill in the missing nucleotides, using the top strand as template, and then DNA ligase seals the nick to complete the task, as in base excision repair.

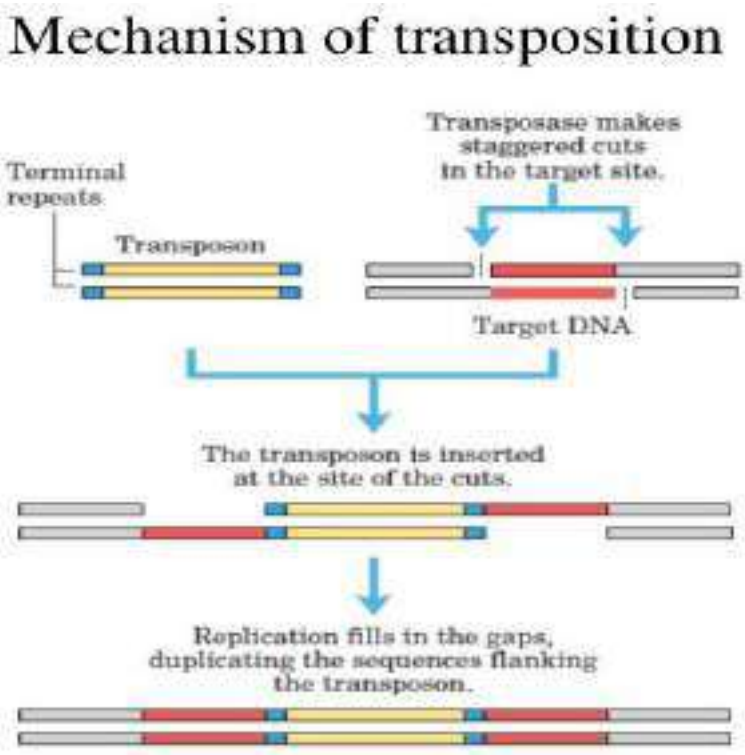


Transposons

These mobile segments of DNA are sometimes called **"jumping genes"**. There are two types:

- **Retrotransposons (Class I)** that first transcribe the DNA into RNA and then use reverse transcriptase to make a DNA copy of the RNA to insert in a new location.
- **Class II Transposons** consisting only of DNA that moves directly from place to place.





Transposons may have an effect on genetic diversity of many organisms. DNA transposons can drive the evolution of genomes by promoting the relocation of sections of DNA sequences. As a result, this can alter gene regulatory regions and phenotypes.

Bacterial transposons are especially good at facilitating horizontal gene transfer between microbes. In bacteria, transposable elements can easily jump between the chromosomal genome and plasmids.

All DNA transposons are inactive in the human genome. Inactivated, or silenced, transposons do not result in a phenotypic outcome and do not move around in the genome. Some are inactive because they have mutations that affect their ability to move between chromosomes, while others are capable of moving but remain inactive due to epigenetic defenses, like DNA methylation . For example, chemical modifications of DNA can constrict certain areas of the genome such that transcription enzymes are unable to reach them.
RNAi, specifically siRNA and miRNA silencing, is a naturally occurring mechanisms that, in addition to regulating eurkaryotic gene expression, prevents transcription of DNA transposons. Another mode of inactivation is overproduction inhibition. When transposase exceeds a threshold concentration, transposon activity is decreased.

transposons are mutagens and their movements are often the causes of genetic disease. They can damage the genome of their host cell in different ways:

- a transposon or a retrotransposon that inserts itself into a functional gene will most likely disable that gene.
- after a DNA transposon leaves a gene, the resulting gap will probably not be repaired correctly.
- additionally, many transposons contain promoters which drive transcription of their own transposase. These promoters can cause aberrant expression of linked genes, causing disease or mutant phenotypes.

Lec 5
Transcription in Prokaryotes and Eukaryotes

Transcription is, chemically and enzymatically, very similar to DNA replication. Both involve enzymes that synthesize a new strand of nucleic acid complementary to a DNA template strand. There are some important differences, of course; most notably, in the case of transcription, the new strand is made from ribonucleotides rather than deoxyribonucleotid. Other mechanistic features of transcription that differ from that of replication include the following:

- RNA polymerase (the enzyme that catalyzes RNA synthesis) does not need a primer; rather, it can initiate transcription de novo.
- The RNA product does not remain base-paired to the template DNA strand: the enzyme displaces the growing chain only a few nucleotides behind where each ribonucleotide is added.
- Transcription, although very accurate, is less accurate than replication (one mistake occurs in 10,000 nucleotides added, compared with one in 10 million for replication).
- Transcription selectively copies only certain parts of the genome and makes anywhere from one to several hundred, or even thousand, copies of any given section. In contrast, replication must copy the entire genome and do so once (and only once) every cell division.

RNA polymerase

RNA polymerase performs essentially the same reaction in all cells, from bacteria to humans. From bacteria to mammals, the cellular RNA polymerases are made up of multiple subunits (although some phage and organelles such as mitochondria and chloroplast do encode single-subunit enzymes that are

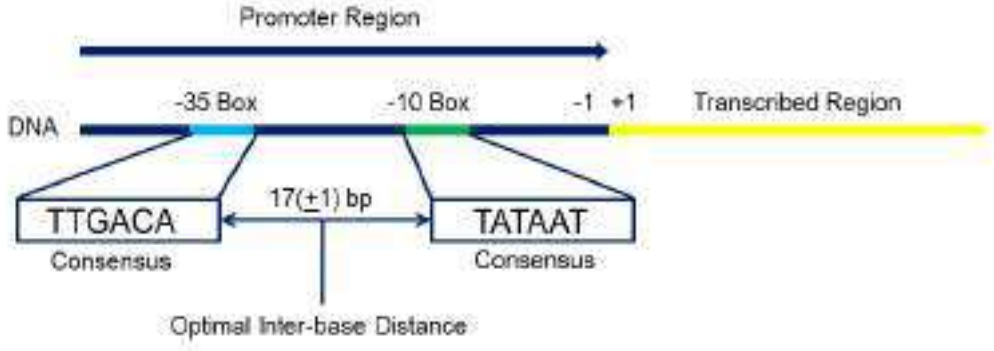
capable of performing the same basic reaction as their more complex multicellular enzymes). Bacteria have only a single RNA polymerase, whereas eukaryotic cells have three RNA polymerases I, II, and III.
The basic enzyme from *E. coli*, called the core enzyme, has one copy of each of three subunits (β, β′and ω) and two copies of α. All of these subunits have homologs in the eukaryotic enzymes. β, β′subunits are involved in phosphodiester bond formation, that these subunits also participate in DNA binding, and that the α-subunit has several activities, including assembly of the core polymerase. ω restores denatured RNA polymerase to its functional form in vitro.

Prokaryotic		Eukaryotic		
Bacterial	**Archaeal**	**RNAP I**	**RNAP II**	**RNAP III**
Core	**Core**	**(Pol I)**	**(Pol II)**	**(Pol III)**
β′	A′/A″	RPA1	RPB1	RPC1
β	B	RPA2	RPB2	RPC2
α′	D	RPC5	RPB3	RPC5
α″	L	RPC9	RPB11	RPC9
ω	K	RPB6	RPB6	RPB6
	[+6 others]	[+9 others]	[+7 others]	[+11 others]

Promoter

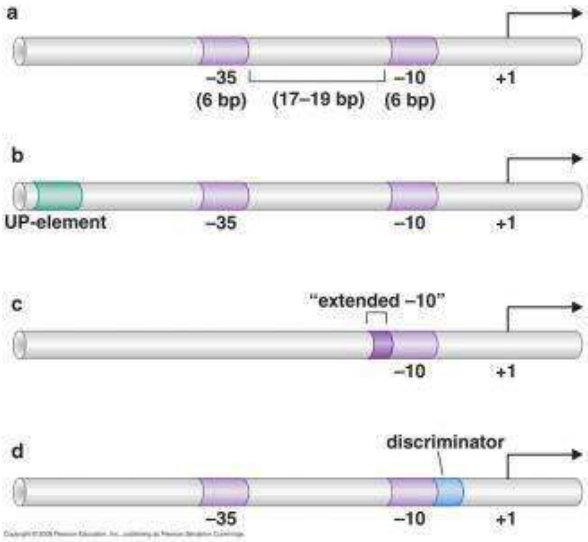
Although RNA polymerases can synthesize RNA unaided, other proteins called initiation factors are required for accurate and efficient initiation. These factors ensure that the enzyme initiates transcription only from appropriate sites on the DNA, called promoters. In bacteria, there is only one initiation factor (σ factor).
Each bacterium has a primary σ-factor that transcribes the genes that required for everyday growth. For example, the primary σ in *E. coli* is called σ70, and the primary s in *B. subtilis* is σ43.These proteins are

named for their molecular masses, 70 and 43 kD, respectively. In addition, bacteria have alternative σ-factors that transcribe specialized genes (heat shock genes and sporulation genes).
In the case of *E. coli*, promoters recognized by polymerase containing σ70 share the following characteristic structure: two conserved sequences, each of 6 nucleotides, separated by a nonspecific stretch of 17–19 nucleotides. The two defined sequences are at -10 bp and at -35 bp upstream of the site where RNA synthesis starts and the DNA nucleotide encoding the beginning of the RNA chain is designated +1.

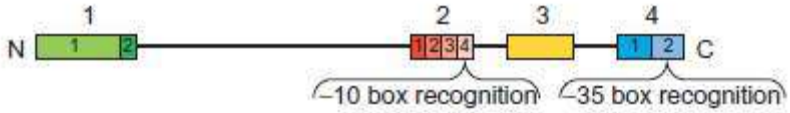


An additional DNA element that binds RNA polymerase is found in some strong promoters, for example, those directing expression of the rRNA genes. This is called an **UP-element** and increases polymerase binding by providing an additional specific interaction between the enzyme and the DNA.

Another class of σ 70 promoters lacks a -35 region and instead has a region called **extended -10 element**. This comprises a standard -10 region with an additional short sequence element at its upstream end. Extra contacts made between polymerase and this additional sequence element compensate for the



The σ 70 factor can be divided into four regions called σ region 1 through σ region 4. The regions that recognize the -10 and -35 elements of the promoter are regions 2 and 4, respectively.



Transcription in prokaryotes

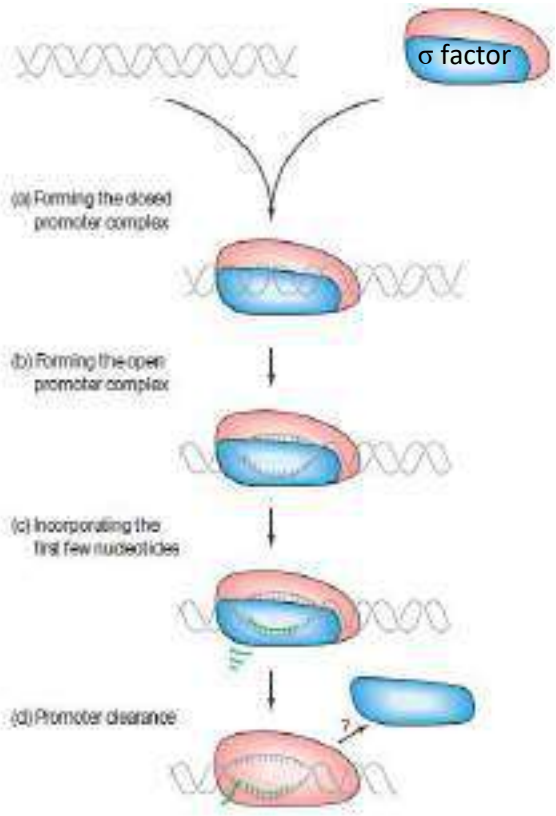
A round of transcription proceeds through three phases called initiation, elongation, and termination.

Initiation

A promoter is the DNA sequence that initially binds the RNA polymerase. Once formed, the promoter polymerase complex undergoes structural changes required for initiation to proceed. Like DNA replication, transcription always occurs in a 5´to 3´ direction. Unlike replication, however, only one of the DNA strands acts as a template on which the RNA strand is built.

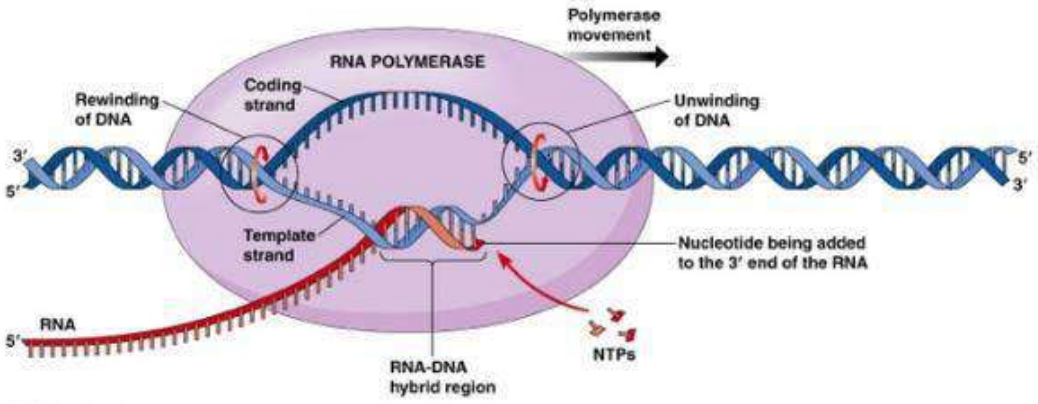
During initiation, RNA polymerase (together with the initiation factor) binds to the promoter in a **closed complex**. In that state, the DNA remains in a double-stranded form. This closed complex then undergoes isomerization to the **open complex**. In that form, the DNA around the transcription start site is unwound, disrupting the base pairs and forming a a transcription bubble of single-stranded DNA. This melting occurs between positions -11 and +2.This transition allows access to the template strand, which determines the order of bases in the new RNA strand.

After the transcript becomes long enough to form a stable hybrid with the template strand, the polymerase changes



Elongation

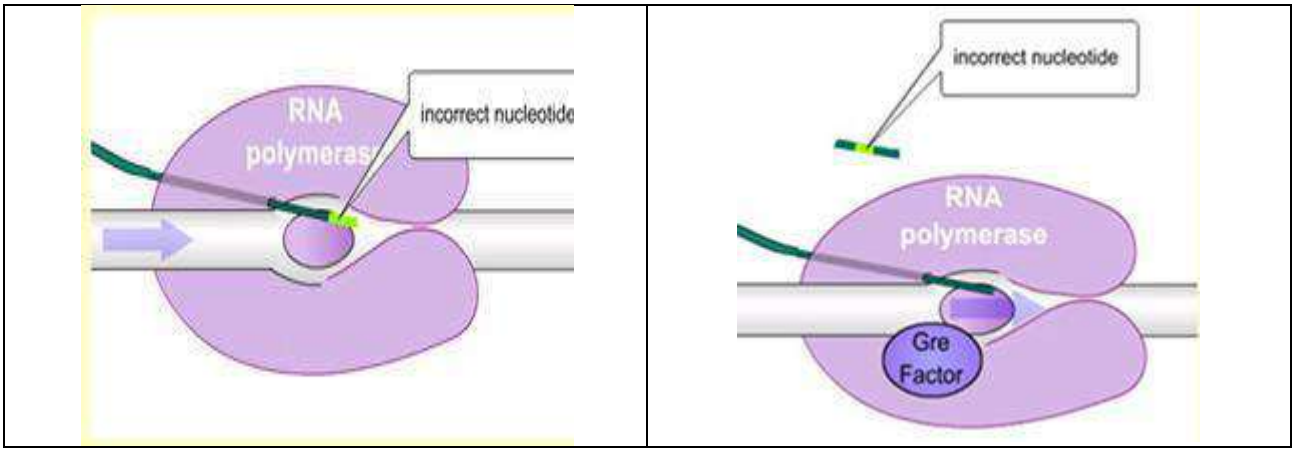
In this phase, the ribonucleotides were added to the 3′end of a growing RNA chain during transcription. The transcription elongation phase begins with the release of the σ subunit from the polymerase. The dissociation of σ allows the core RNA polymerase enzyme to proceed along the DNA template, synthesizing mRNA in the 5′ to 3′ direction at a rate of approximately 40 nucleotides per second. RNA polymerase acts as a stable linker between the DNA template and the nascent RNA strands to ensure that elongation is not interrupted prematurely.



Transcription elongation RNA polymerase has proofreading capabilities from 3´ to 5´ by removing one or several nucleotide and adding correct ones. RNA polymerase performs two proofreading functions: **Pyrophosphorolytic editing**: In this, the enzyme uses its active site to catalyze the removal of an

incorrectly inserted ribonucleotide, by reincorporation of PPi. The enzyme can then incorporate another ribonucleotide in its place in the growing RNA chain.

Hydrolytic editing: the polymerase backtracks by one or more nucleotides and cleaves the RNA product, removing the error containing sequence. Hydrolytic editing is stimulated by Gre factors, which also serves as elongation stimulating factors.

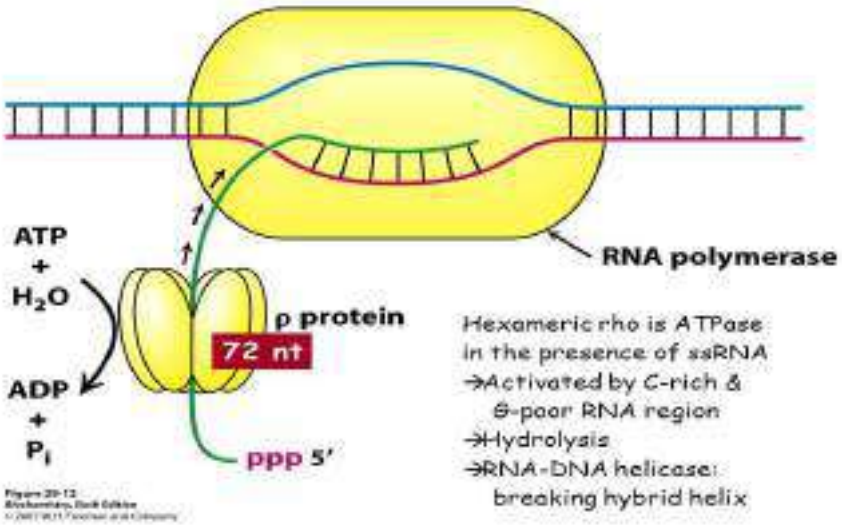


Termination

After RNA polymerase transcribe the gene, it must be stop and release the RNA product as well as dissociating from the DNA itself. This step is called termination. There, sequences called terminators trigger the elongating polymerase to dissociate from the DNA and release the RNA chain. In bacteria, terminators come in two types: Rho-dependent and Rho-independent.

Rho-dependent termination

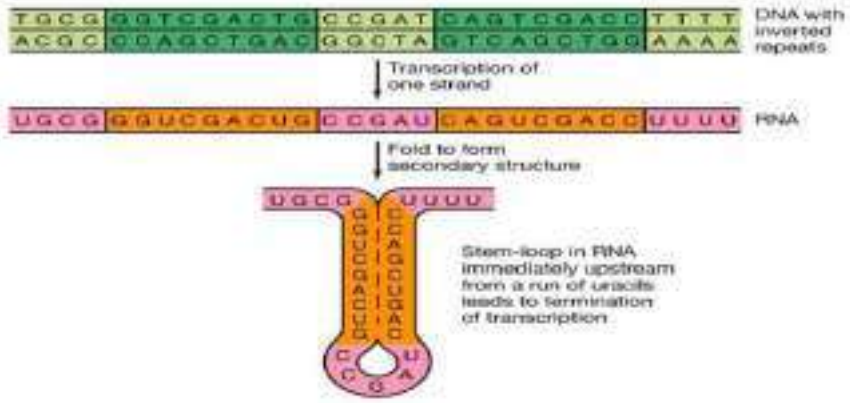
It is requires a protein called Rho to induce termination. Rho, which is a ring-shaped protein with six identical subunits, binds to single-stranded RNA as it exits the polymerase. The protein also has an ATPase activity, and once attached to the transcript, Rho uses the energy derived from ATP hydrolysis to induce termination. Rho induces a conformational change in polymerase, causing the enzyme to terminate. Terminator is a sequence rich in C and poor in G.



Rho-independent terminators, also called intrinsic terminators because they need no other factors to work, consist of two sequence elements: a short inverted repeat (of about 20 nucleotides) followed by a stretch of about eight A:T base pairs. When polymerase transcribes an inverted repeat sequence, the

resulting RNA can form a stem-loop structure (hairpin) by base-pairing with itself. Formation of the hairpin causes termination by disrupting the elongation complex.

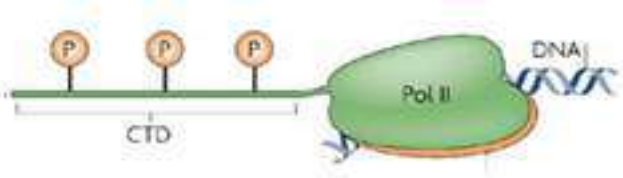
The hairpin works as an efficient terminator only when it is followed by a stretch of A:U base pairs. Because A:U base pairs are the weakest of all base pairs (weaker even than A:T base pairs), they are more easily disrupted by the effects of the stem-loop on the transcribing polymerase, and thus the RNA will more readily dissociate.



Transcription in eukaryotes

Transcription in eukaryotes is undertaken by polymerases closely related to the RNA polymerases found in prokaryotes. All eukaryotes have at least three different ones (Pol I, II, and III; and plants also have a Pol IV and a Pol V). In addition, whereas bacteria require only one additional initiation factor (σ), several initiation factors are required for efficient and promoter-specific initiation in eukaryotes. These are called the **general transcription factors (GTFs)**.

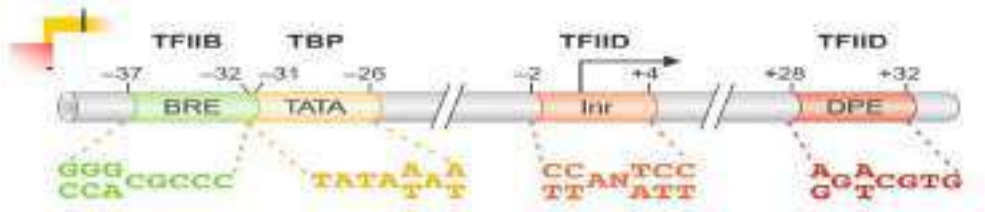
- **RNA polymerase I** resides in the nucleolus and is responsible for synthesizing three of the four types of rRNA found in eukaryotic ribosomes (28S, 18S,and 5.8 S rRNA).
- **RNA polymerase II** is found in the nucleoplasm and synthesizes precursors to mRNA and some small nuclear RNAs (snRNAs) . Pol II differs from the bacterial enzyme in one important way. It is contain tail at the carboxy-terminal end of the large subunit called **carboxy-terminal domain** (**CTD**), and this is absent from the bacterial enzyme. This tail is made up of multiple repeats of a heptapeptide sequence: Tyr-Ser-Pro-Thr-Ser-Pro-Ser.



- **RNA polymerase III** is also a nucleoplasmic enzyme, but it synthesizes a variety of small RNAs, including tRNA precursors and the smallest type of ribosomal RNA, 5S rRNA.

The eukaryotic core promoter refers to the minimal set of sequence elements required for accurate transcription initiation by the Pol II machinery. A core promoter is typically about 40-60 nucleotides long, extending either upstream or downstream from the transcription start site.

The Pol II core promoters contain the **TFIIB recognition element** (BRE), the **TATA element** (or box), the **initiator** (Inr), and the **downstream promoter elements** (known as DPE, DCE, and MTE). Typically, a promoter includes some subset of these elements.



- TFIIB recognition element (BRE)
- The TATA element/box
- Initiator (Inr)
- The downstream promoter element (DPE)

Many Pol II promoters contain a TATA element (some 30 bp upstream of the transcription start site). This is where preinitiation complex formation begins. The TATA element is recognized by the general transcription factor called TFIID that binds to the TATA DNA sequence is called **TBP** (**TATA-binding protein**). TFIID is a critical factor in promoter recognition and preinitiation complex establishment.

Upon binding DNA, TBP extensively distorts the TATA sequence and the resulting TBP–DNA complex provides a platform to recruit other general transcription factors and polymerase itself to the promoter.

Formation of the preinitiation complex containing these components is followed by promoter melting which needed ATP hydrolysis for DNA melting.

In eukaryotes, promoter escape involves two steps not seen in bacteria: one is ATP hydrolysis and the phosphorylation of the polymerase large subunit that has a carboxy-terminal domain (CTD) tail contains a series of repeats of the heptapeptide sequence. Regulating the phosphorylation state of the CTD of Pol II controls subsequent steps of elongation and even processing of the RNA.

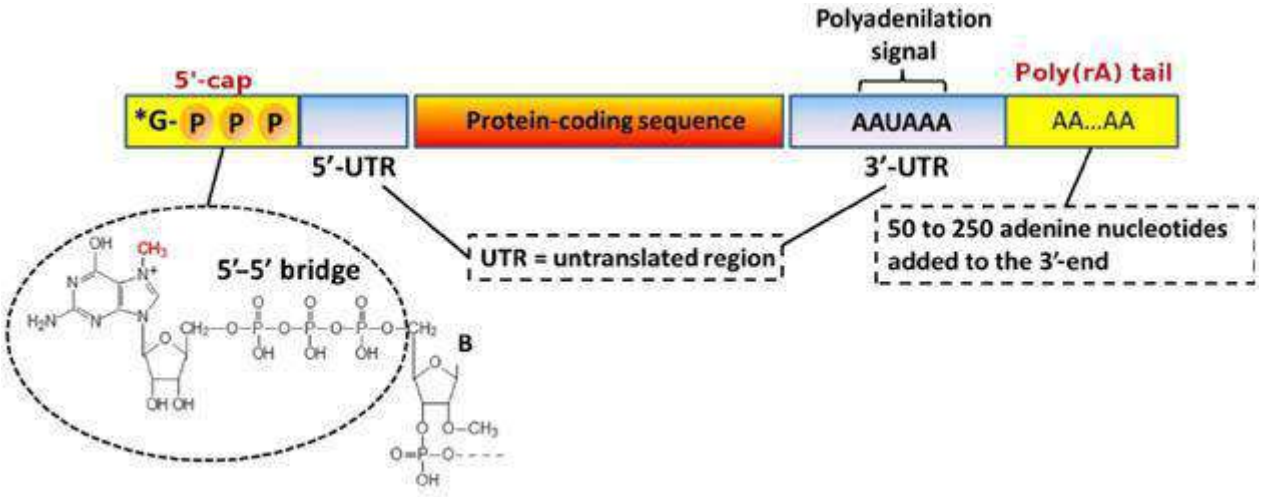
Once polymerase has escaped the promoter and initiated transcription, it shifts into the elongation phase, **TFIIS (elongation factor)** are recruited to stimulate elongation.

The first RNA processing event is **capping**. This involves the addition of a modified guanine base to the 5′ end of the RNA. Specifically, it is a methylated guanine, and it is joined to the RNA transcript by an unusual 5′- 5′ linkage involving three phosphates. This happens when the transcription cycle has progressed only as far as the transition from the initiation to elongation phases.

The final RNA processing event, **polyadenylation** of the 3′ end of the mRNA, is intimately linked with the termination of transcription. The polymerase CTD tail is involved in recruiting some of the enzymes necessary for polyadenylation.

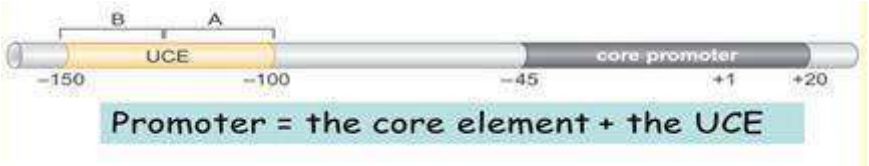
After pol II transcribes the protein-coding region of the mRNA, it encounters two sequence elements: **AAUAAA** and a **GU-rich element**. These act as signals for the assembly of a large 3′ processing complex that cleaves the nascent pre-mRNA, releasing it from the transcription complex, and adds a tail of up to 200 adenosine residues.

Polyadenylation is mediated by an enzyme called **poly-A polymerase**, which adds adenines to the RNA's 3′ end produced by the cleavage. This enzyme uses ATP as a precursor and adds the nucleotides using the same chemistry as RNA polymerase. But it does so without a template. Thus, the long tail of As is found in the RNA but not the DNA then the mature mRNA is then transported from the nucleus.



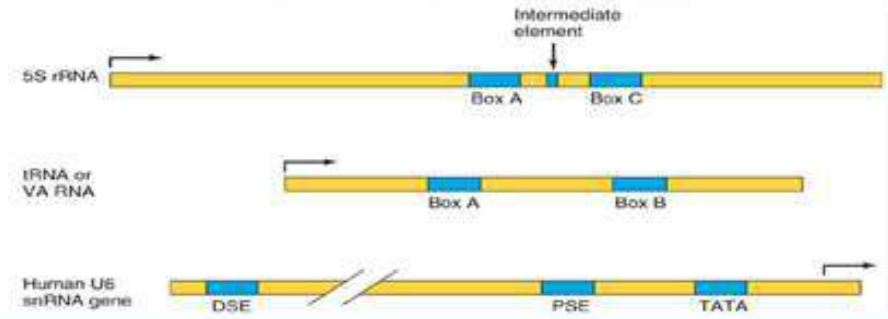
RNA Pol I

Pol I is required for the expression of only one gene, that encoding the rRNA precursor. There are many copies of that gene in each cell, and, indeed, it is expressed at far higher levels than any other gene. The promoter for the rRNA gene comprises two parts: the **core element** and the **UCE (upstream control element)**. Transcription by RNA polymerase I yields a 45S primary transcript (pre-rRNA), which is processed into the mature 28S, 18S, and 5.8S rRNAs found in cytoplasmic ribosomes.



RNA Pol III

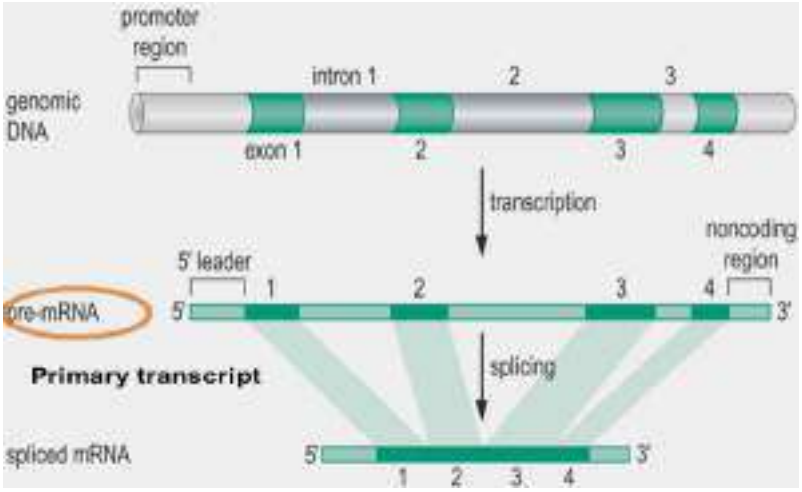
Pol III promoters come in various forms, and the vast majority have the unusual feature of being located downstream from the transcription start site (i.e., within the coding region of the gene). Some Pol III promoters (e.g., those for the tRNA genes) consist of two regions, called Box A and Box B, separated by a short element; others contain Box A and Box C (e.g., the 5S rRNA gene); and still others contain a TATA element.



RNA Splicing

Many eukaryotic genes are thus mosaics, consisting of blocks of coding sequences separated from each other by blocks of non-coding sequences. The coding sequences are called exons and the intervening sequences are called introns. Once transcribed into an RNA transcript (**primary transcript** or **pre-mRNAs**), the introns must be removed and the exons joined together to create the mature mRNA for that gene in processing called **RNA splicing** that must occur with great precision to avoid the loss, or addition, of even a single nucleotide at the sites at which the exons are joined.

The sizes of the exons and introns vary as well. Indeed, introns are very often much longer than the exons they separate. Thus, for example, exons are typically on the order of 150 nucleotides, whereas introns can be as long as 800,000 nucleotides (800 kb).



The chemistry of pre-mRNA splicing

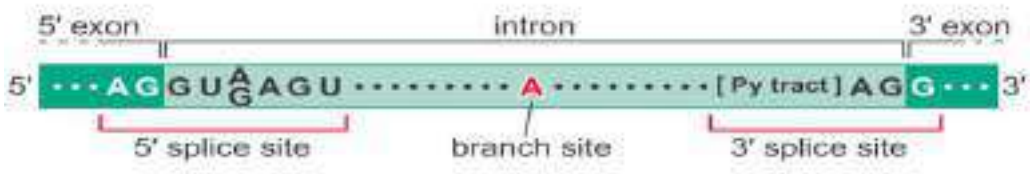
The signals in the pre-mRNA that identify the introns and exons are recognized by a combination of proteins and **small nuclear RNAs** (snRNAs). Splicing occurs in a large complex termed the spliceosome, within which the pre-mRNA assembles together with five **snRNAs (U1, U2, U4, U5, and U6)** and around 150 different proteins.

<u>The snRNPs have three roles in splicing:</u>

- they recognize the 5´ splice site and the branch site.
- they bring those sites together as required.
- they catalyze the RNA cleavage and joining reactions.

The borders between introns and exons are marked by specific nucleotide sequences within the pre-mRNAs. There are three splice sites:

- 5'splice site: the exon-intron boundary at the 5' end of the intron .
- 3' splice site: the exon-intron boundary at the 3' end of the intron .
- Branch point site: an A close to the 3' end of the intron, which is followed by a polypyrimidine tract (Py tract).

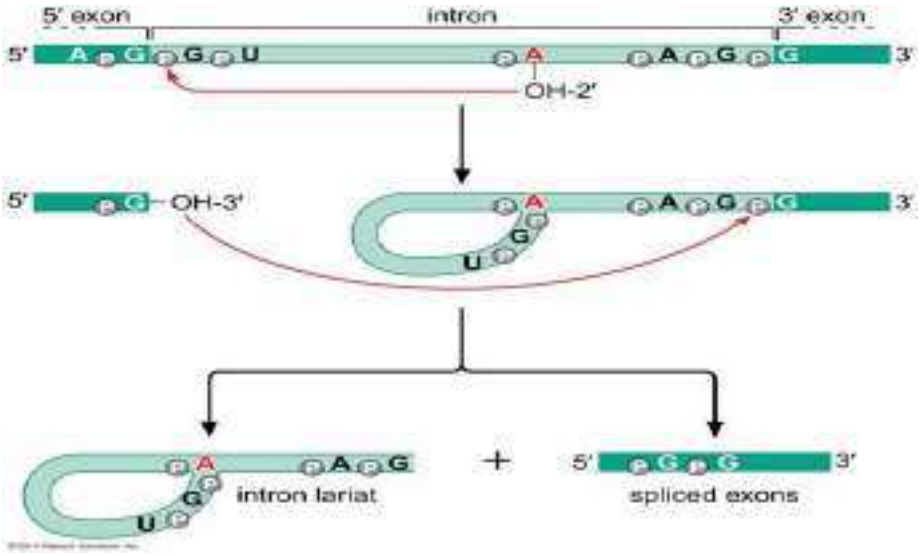


The U1 and U6 snRNAs have sequences that are complementary to the 5′ splice site, while U2 is complementary to the branch point region. the intron is removed in a Form Called a Lariat as the Flanking Exons are joined.

Two successive transesterification Step:

Step 1: The OH of the conserved A at the branch site attacks the phosphoryl group of the conserved G in the 5’ splice site. As a result, the 5’ exon is released and the 5’-end of the intron forms a three-way junction structure.

Step 2: The OH of the 5’ exon attacks the phosphoryl group at the 3’ splice site. As a consequence, the 5’ and 3’ exons are joined and the intron is liberated in the shape of a lariat. The lariat is subsequently linearized (debranched) and degraded.



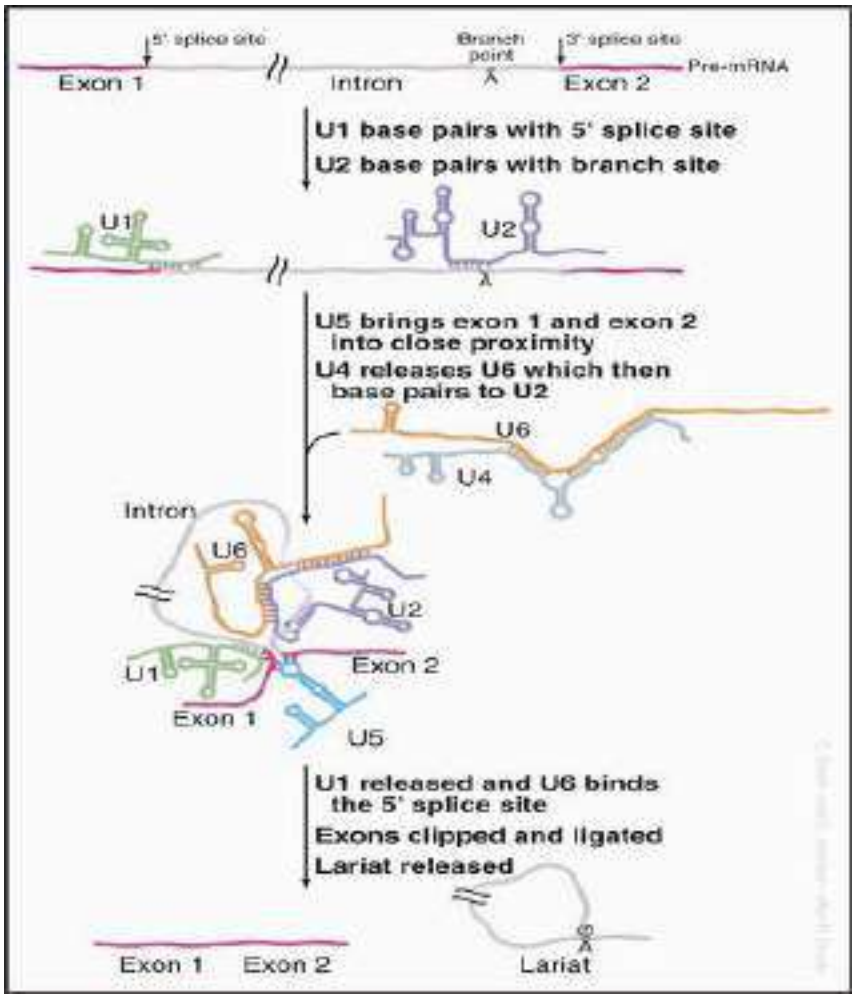
Initially, the 5´ splice site is recognized by the U1 snRNP (using base pairing between its snRNA and the pre-mRNA). U2 snRNP then binds to the branch site, producing **A complex** (pre-spliceosome). The next step is a rearrangement of the A complex to bring together all three splice sites.
This is achieved as follows: the U4 and U6 snRNPs, along with the U5 snRNP, join the complex. The A complex is converted into the **B complex** (mature spliceosome).

In the next step, U1 leaves the complex, and U6 replaces it at the 5´ splice site. The next rearrangement triggers catalysis and occurs as follows: U4 is released from the complex, allowing U6 to interact with U2.This arrangement, called the **C complex**.

Catalysis Step 1: Formation of the C complex produces the active site, with U2 and U6 RNAs being brought together.
Formation of the active site juxtaposes the 5’ splice site of the pre-mRNA and the branch site, allowing the branched A residue to attack the 5’ splice site to accomplish the first transesterfication reaction.

Catalysis Step 2: U5 snRNP helps to bring the two exons together, and aids the second transesterification reaction, in which the 3’-OH of the 5’ exon attacks the 3’ splice site.

Final Step Release of the mRNA product and the snRNPs.

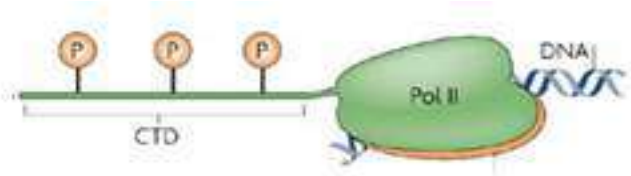

5' splice site
Branch point
3' splice site
Exon 1
Intron
Exon 2
Pre-mRNA
U1 base pairs with 5' splice site
U2 base pairs with branch site
U1
U2
U5 brings exon 1 and exon 2 into close proximity
U4 releases U6 which then base pairs to U2
U6
U4
Intron
U6
U2
U1
Exon 2
Exon 1
U5
U1 released and U6 binds the 5' splice site
Exons clipped and ligated
Lariat released
Exon 1
Exon 2
Lariat

Lec 6

Transcription in eukaryotes

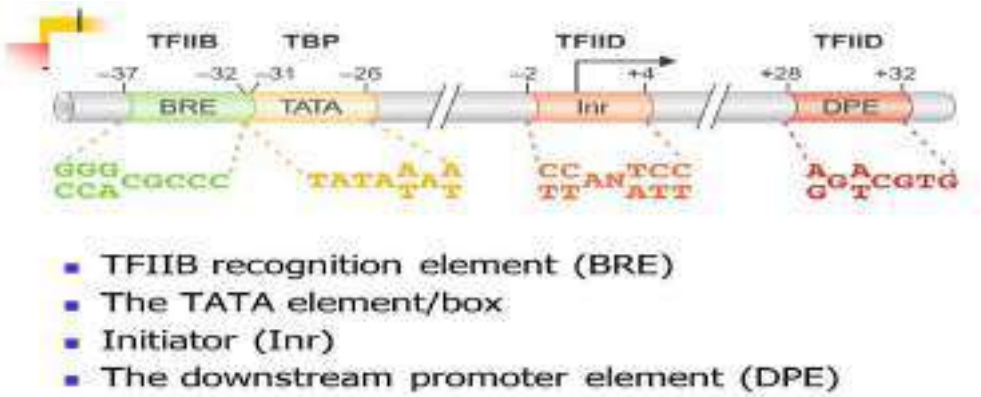
Transcription in eukaryotes is undertaken by polymerases closely related to the RNA polymerases found in prokaryotes. All eukaryotes have at least three different ones (Pol I, II, and III; and plants also have a Pol IV and a Pol V). In addition, whereas bacteria require only one additional initiation factor (σ), several initiation factors are required for efficient and promoter-specific initiation in eukaryotes. These are called the **general transcription factors (GTFs)**.

- **RNA polymerase I** resides in the nucleolus and is responsible for synthesizing three of the four types of rRNA found in eukaryotic ribosomes (28S, 18S,and 5.8 S rRNA).
- **RNA polymerase II** is found in the nucleoplasm and synthesizes precursors to mRNA and some small nuclear RNAs (snRNAs) . Pol II differs from the bacterial enzyme in one important way. It is contain tail at the carboxy-terminal end of the large subunit called **carboxy-terminal domain** (**CTD**), and this is absent from the bacterial enzyme. This tail is made up of multiple repeats of a heptapeptide sequence: Tyr-Ser-Pro-Thr-Ser-Pro-Ser.



- **RNA polymerase III** is also a nucleoplasmic enzyme, but it synthesizes a variety of small RNAs, including tRNA precursors and the smallest type of ribosomal RNA, 5S rRNA.

The eukaryotic core promoter is typically about 40-60 nucleotides long, extending either upstream or downstream from the transcription start site. The Pol II core promoters contain the **TFIIB recognition element** (BRE), the **TATA element** (or box), the **initiator** (Inr), and the **downstream promoter elements** (known as DPE, DCE, and MTE). Typically, a promoter includes some subset of these elements.



- TFIIB recognition element (BRE)
- The TATA element/box
- Initiator (Inr)
- The downstream promoter element (DPE)

Many Pol II promoters contain a TATA element (some 30 bp upstream of the transcription start site). This is where preinitiation complex formation begins. The TATA element is recognized by the general transcription factor called TFIID that binds to the TATA DNA sequence is called **TBP** (**TATA-binding protein**). TFIID is a critical factor in promoter recognition and preinitiation complex establishment.

Upon binding DNA, TBP extensively distorts the TATA sequence and the resulting TBP–DNA complex provides a platform to recruit other general transcription factors and polymerase itself to the promoter.

Formation of the preinitiation complex containing these components is followed by promoter melting which needed ATP hydrolysis for DNA melting.

In eukaryotes, promoter escape involves two steps not seen in bacteria:

1. ATP hydrolysis
2. The phosphorylation of the polymerase large subunit that has a carboxy-terminal domain (CTD) tail contains a series of repeats of the heptapeptide sequence. Regulating the phosphorylation state of the CTD of Pol II controls subsequent steps of elongation and even processing of the RNA.

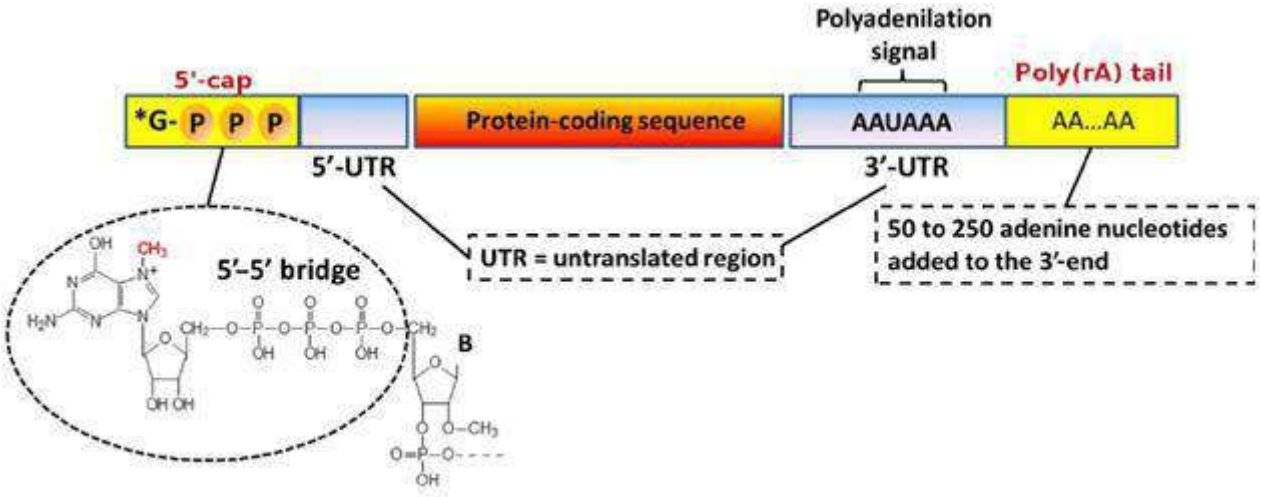
Once polymerase has escaped the promoter and initiated transcription, it shifts into the elongation phase, **TFIIS (elongation factor)** are recruited to stimulate elongation.

The first RNA processing event is **capping**. This involves the addition of a modified guanine base to the 5′ end of the RNA. Specifically, it is a methylated guanine, and it is joined to the RNA transcript by an unusual 5′- 5′ linkage involving three phosphates. This happens when the transcription cycle has progressed only as far as the transition from the initiation to elongation phases.

The final RNA processing event, **polyadenylation** of the 3′ end of the mRNA, is intimately linked with the termination of transcription. The polymerase CTD tail is involved in recruiting some of the enzymes necessary for polyadenylation.

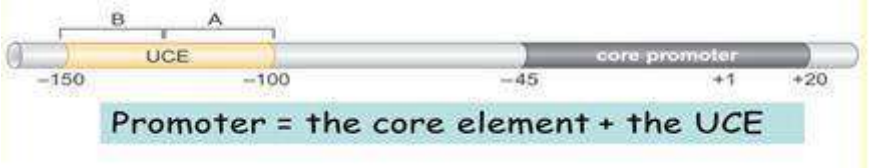
After pol II transcribes the protein-coding region of the mRNA, it encounters two sequence elements: **AAUAAA** and a **GU-rich element**. These act as signals for the assembly of a large 3′ processing complex that cleaves the nascent pre-mRNA, releasing it from the transcription complex, and adds a tail of up to 200 adenosine residues.

Polyadenylation is mediated by an enzyme called **poly-A polymerase**, which adds adenines to the RNA's 3′ end produced by the cleavage. This enzyme uses ATP as a precursor and adds the nucleotides using the same chemistry as RNA polymerase. But it does so without a template. Thus, the long tail of As is found in the RNA but not the DNA then the mature mRNA is then transported from the nucleus.



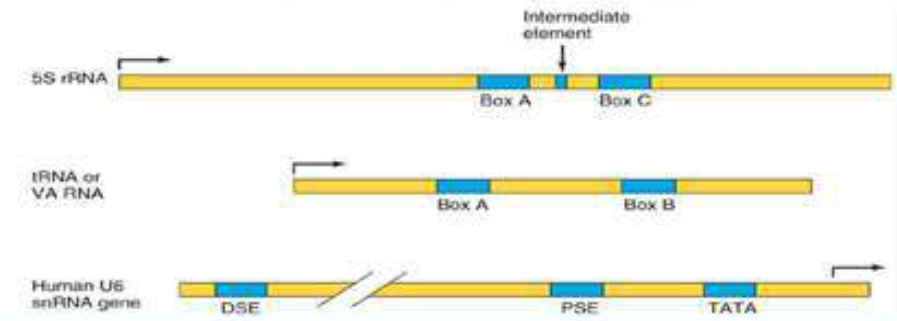
RNA Pol I

Pol I is required for the expression of only one gene, that encoding the rRNA precursor. There are many copies of that gene in each cell, and, indeed, it is expressed at far higher levels than any other gene. The promoter for the rRNA gene comprises two parts: the **core element** and the **UCE (upstream control element)**. Transcription by RNA polymerase I yields a 45S primary transcript (pre-rRNA), which is processed into the mature 28S, 18S, and 5.8S rRNAs found in cytoplasmic ribosomes.



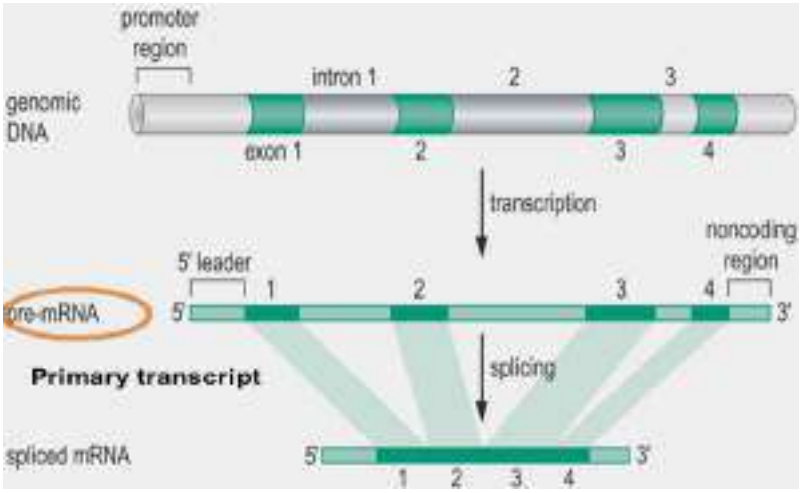
RNA Pol III

Pol III promoters come in various forms, and the vast majority have the unusual feature of being located downstream from the transcription start site (i.e., within the coding region of the gene). Some Pol III promoters (e.g., those for the tRNA genes) consist of two regions, called Box A and Box B, separated by a short element; others contain Box A and Box C (e.g., the 5S rRNA gene); and still others contain a TATA element.



RNA Splicing and RNA types in Eukaryotes

Many eukaryotic genes are consisting of blocks of coding sequences separated from each other by blocks of non-coding sequences. The coding sequences are called **exons** and the intervening sequences are called **introns**. Once transcribed into an RNA transcript (**primary transcript** or **pre-mRNAs**), the introns must be removed and the exons joined together to create the mature mRNA for that gene in processing called **RNA splicing** that must occur with great precision to avoid the loss, or addition, of even a single nucleotide at the sites at which the exons are joined. The sizes of the exons and introns vary as well. Indeed, introns are very often much longer than the exons they separate. Thus, for example, exons are typically on the order of 150 nucleotides, whereas introns can be as long as 800,000 nucleotides (800 kb).



The chemistry of pre-mRNA splicing

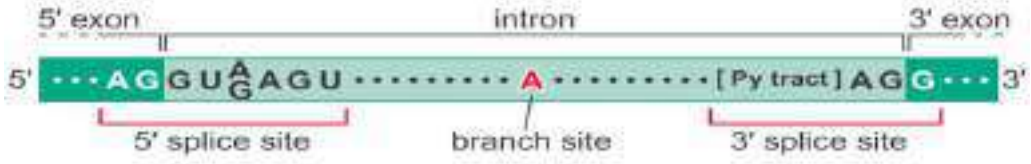
The signals in the pre-mRNA that identify the introns and exons are recognized by a combination of proteins and **small nuclear RNAs (snRNAs).** Splicing occurs in a large complex termed the **spliceosome**, within which the pre-mRNA assembles together with five **snRNAs (U1, U2, U4, U5, and U6)** and around 150 different proteins.

The snRNAs have three roles in splicing:

- they recognize the 5´ splice site and the branch site.
- they bring those sites together as required.
- they catalyze the RNA cleavage and joining reactions.

The borders between introns and exons are marked by specific nucleotide sequences within the pre-mRNAs. There are three splice sites:

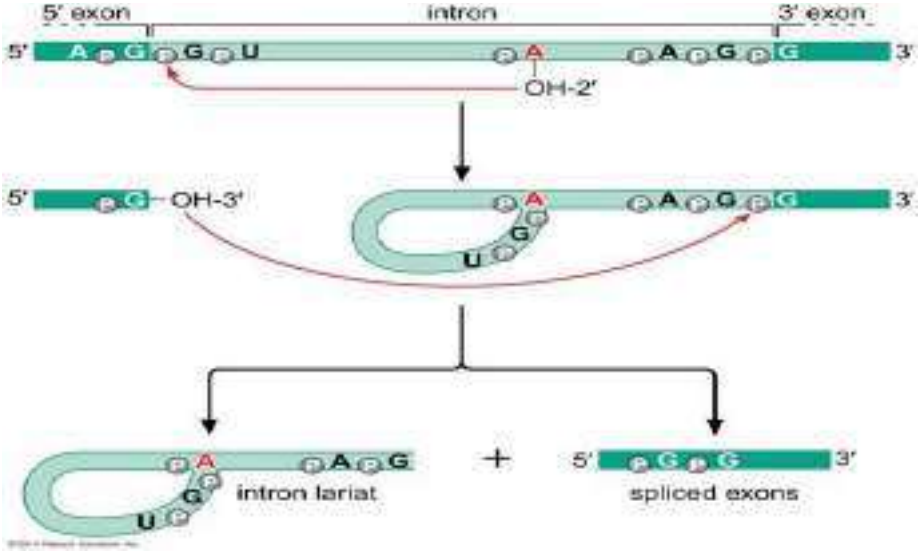
- 5'splice site: the exon-intron boundary at the 5' end of the intron .
- 3' splice site: the exon-intron boundary at the 3' end of the intron .
- Branch point site: an A close to the 3' end of the intron, which is followed by a polypyrimidine tract (Py tract).



Two successive transesterification Step:

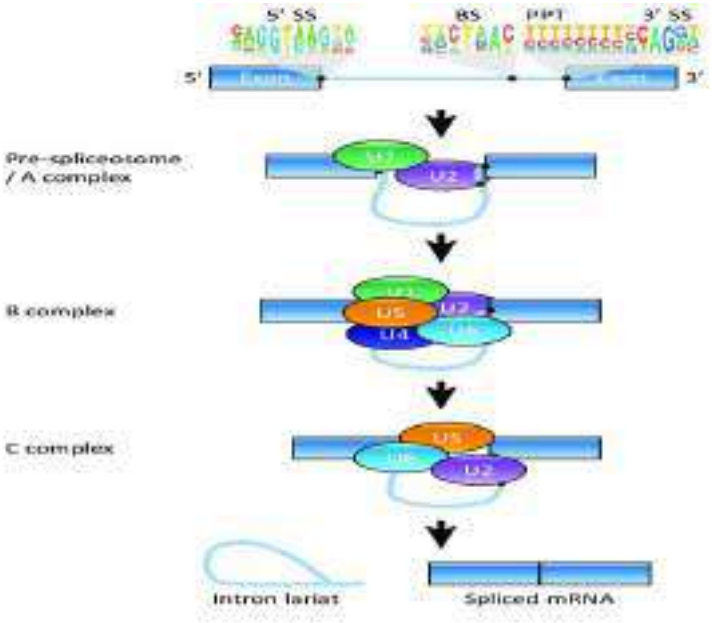
Step 1: The OH of the conserved A at the branch site attacks the phosphoryl group of the conserved G in the 5' splice site. As a result, the 5' exon is released and the 5'-end of the intron forms a three-way junction structure.

Step 2: The OH of the 5' exon binds the phosphoryl group at the 3' splice site. As a consequence, the 5' and 3' exons are joined and the intron is liberated in the shape of a **lariat**. The lariat is subsequently linearized (debranched) and degraded.



Initially, the 5´ splice site is recognized by the U1 snRNP (using base pairing between its snRNA and the pre-mRNA). U2 snRNA then binds to the branch site, producing **A complex (pre-spliceosome)**. The next step is a rearrangement of the A complex to bring together all three splice sites. This is achieved as follows: the U4, U6 and U5 snRNPs join the complex. The A complex is converted into the **B complex (mature spliceosome).**

U1 leaves the complex, and U6 replaces it at the 5´ splice site. The next rearrangement triggers catalysis and occurs as follows: U4 is released from the complex, allowing U6 to interact with U2.This arrangement, called the **C complex** allowing the branched A residue to attack the 5' splice site to accomplish the first transesterfication reaction. U5 snRNA helps to bring the two exons together, and aids the second transesterification reaction, in which the 3'-OH of the 5' exon attacks the 3' splice site. Finally, release of the mRNA product, snRNAs and other proteins.

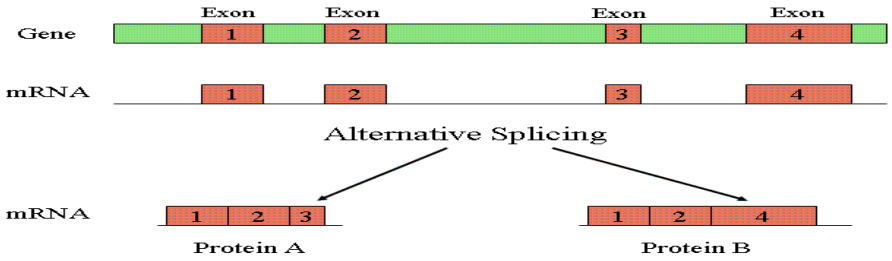


Self-splicing RNA

There are another RNA splicing process called **self-splicing introns** refers to introns that can remove themselves from RNAs in the absence of any proteins or other RNA molecules.

Alternative splicing

Some pre-mRNAs can be spliced in more than one way. Thus, mRNAs containing different selections of exons can be generated from a given pre-mRNA. Called **alternative splicing**, this strategy enables a gene to give rise to more than one polypeptide product. These alternative products are called **isoforms**. It is estimated that 90% or more of the protein-coding genes in the human are spliced in alternative ways to generate more than one isoform produced in response to different signals or in different cell types.

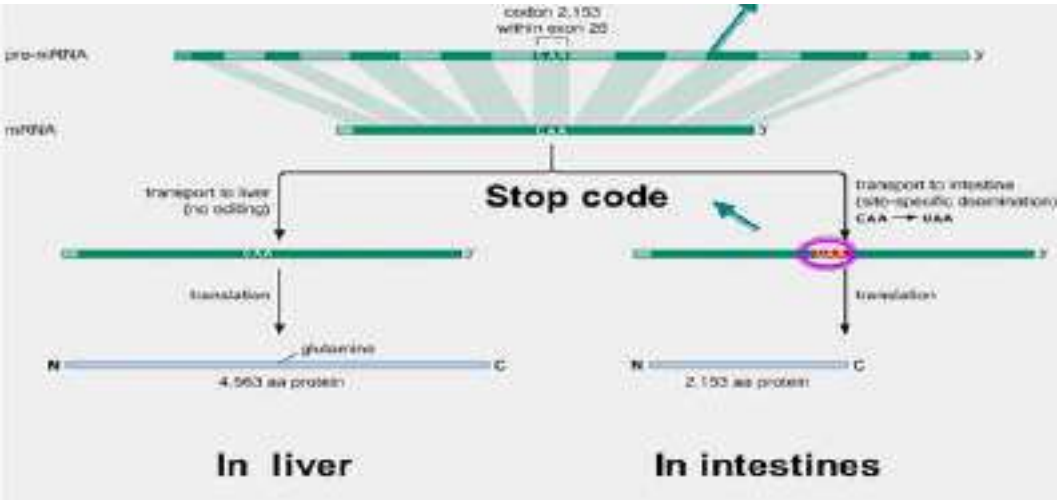


RNA editing

RNA editing is process which information change at the level of mRNA. It is revealed by situations in which the coding sequence in an RNA differs from the sequence of DNA from which it was transcribed. Editing events may include the insertion, deletion, and base substitution of nucleotides within the edited RNA molecule. RNA editing occurs in the cell nucleus and cytosol, as well as within mitochondria and plastids. There are two mechanisms that mediate editing:

1. **Site-specific deamination**

It is targeted cytosine residue within mRNA is converted into uridine by deamination. For a given mRNA species that undergoes editing, that process typically occurs only in certain tissues or cell types and in a regulated manner. The mammalian apolipoprotein B gene has several exons, within one of which is a particular CAA codon that is targeted for editing; it is the C within this codon that gets deaminated by the enzyme cytidine deaminase, converts the C to a U.



In this example, the deamination occurs in a tissue-specific manner: messages are edited in intestinal cells but not in liver cells. The CAA codon, which is translated as glutamine in the unedited message in the liver. The two forms of apolipoprotein B are both involved in lipid metabolism. The longer form, found in the liver, is involved in the transport of endogenously synthesized cholesterol and triglycerides. The smaller version, found in the intestines, is involved in the transport of dietary lipids to various tissues.

2. Guide RNAs (Insertion/deletion U)

RNA editing through the addition and deletion of uracil has been found in the mitochondria of *Trypanosoma brucei*. editing starts with the base-pairing of the unedited primary transcript with a **guide RNA (gRNA**), which contains complementary sequences to the regions around the insertion/deletion points. The newly formed double-stranded region is then enveloped by an **editosome**, a large multi-protein complex that catalyzes the editing.

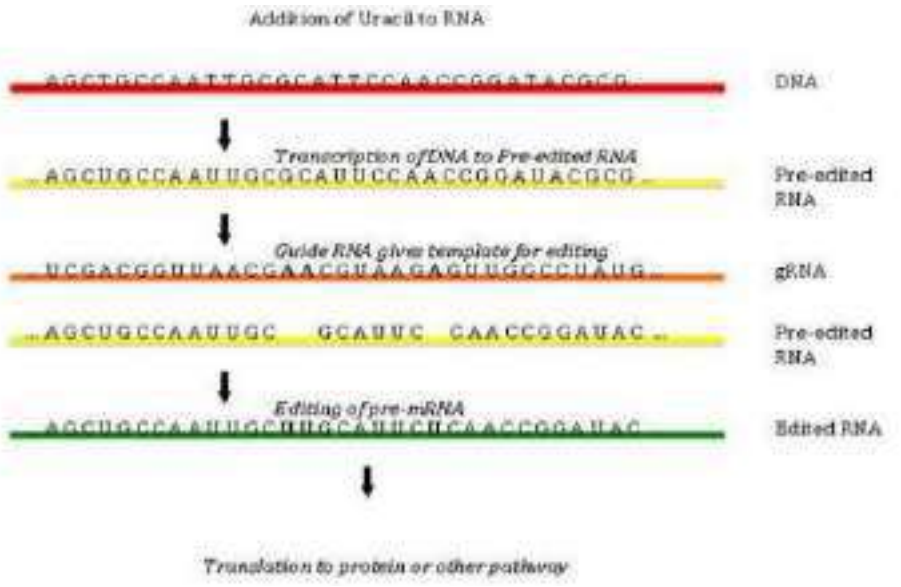
The editosome opens the transcript at the first mismatched nucleotide and starts inserting uridines. The inserted uridines will base-pair with the guide RNA, and insertion will continue as long as A is present in the guide RNA. The inserted nucleotides cause a frameshift and result in a translated protein that differs from its gene.

Significance of RNA Editing

-It is essential in regulating gene expression of organisms.

-RNA editing mutant was reported with strong defects in organelle development.

-It is a mechanism to increase the number of different proteins available without the need to increase the number of genes in the genome.



Other RNA molecules in Eukaryotes

The coding sequence of the mRNA determines the amino acid sequence in the protein that is produced. However, many RNAs do not code for protein (about 97% of the transcriptional output is non-protein-coding in eukaryotes).

These called non-coding RNAs (ncRNA) can be encoded by their own genes (RNA genes). The most prominent examples of non-coding RNAs are transfer RNA (tRNA) and ribosomal RNA (rRNA), both of which are involved in the process of translation. There are also non-coding RNAs involved in gene regulation, RNA processing and other roles. Certain RNAs are able to catalyse chemical reactions such as cutting and ligating other RNA molecules, and the catalysis of peptide bond formation in the ribosome; these are known as ribozymes.

According to the length of RNA chain, RNA includes small RNA and long RNA. Usually, small RNAs are shorter than 200 nt in length, and long RNAs are greater than 200 nt long. Long RNAs, also called large RNAs, mainly include long non-coding RNA (lncRNA) and mRNA.

Small RNAs mainly include 5.8S and 5S rRNA, tRNA, microRNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNAs), Piwi-interacting RNA (piRNA), tRNA-derived small RNA (tsRNA) and small rDNA-derived RNA (srRNA).

Among them, small interfering RNAs (siRNAs) and microRNAs (miRNAs) have attracted considerable attention because their role in gene regulation makes them likely targets for drug discovery and development. Indeed, the therapeutic potential of siRNAs and miRNAs has been demonstrated in the treatment of many different diseases including cancers and infections.

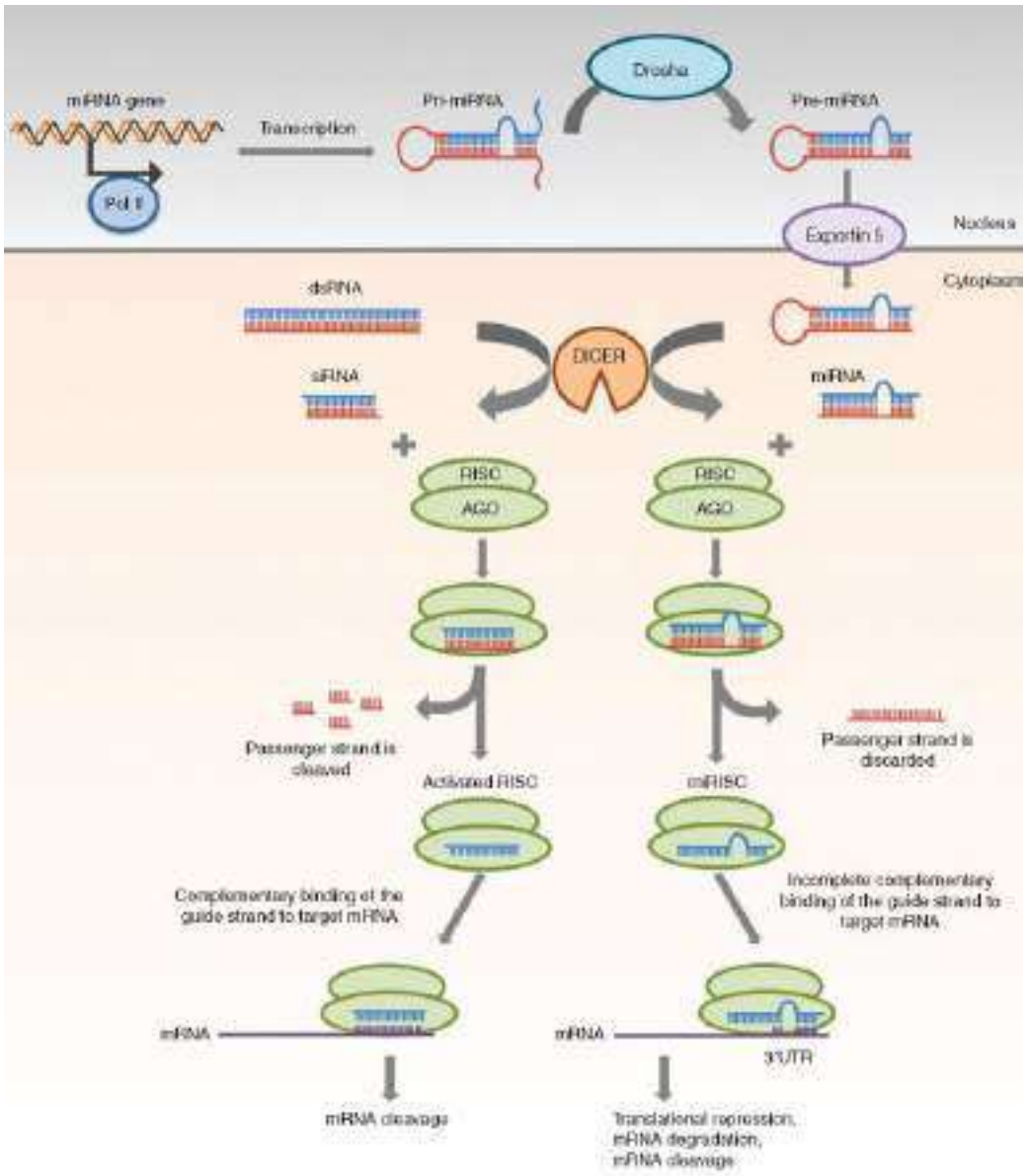
Although the gene silencing effects of siRNAs and miRNAs are distinct, the distinction has been obscured because they are associated with common enzymes (*e.g.*, Dicer and RISC). The major difference between siRNAs and miRNAs is that the siRNAs inhibit the expression of one specific target mRNA while the miRNAs regulate the expression of multiple mRNAs.

Small interfering RNA (siRNA), is a class of double-stranded RNA molecules. It interferes with the expression of specific genes with complementary nucleotide sequences by degrading mRNA after transcription, preventing translation. It plays an important role in gene regulation and innate defense against invading viruses. The mechanism occurs as follows:

In general, the dsRNA (either transcribed from cellular genes or infecting pathogens, or artificially introduced into the cells) is processed by a specialized ribonuclease (RNase) III-like enzyme named **Dicer** in the cytoplasm into a smaller dsRNA molecule. This short dsRNA molecule is known as the siRNA, which has 20–25 nucleotides with 3′ two-nucleotide overhangs. The siRNA interacts with and activates the **RNA-induced silencing complex (RISC)**. The **endonuclease argonaute 2 (AGO2)** component of the RISC cleaves the **passenger strand (sense strand)** of the siRNA while the **guide strand (antisense strand)** remains associated with the RISC. Subsequently, the guide strand guides the active RISC to its target mRNA for cleavage by AGO2. As the guide strand only binds to mRNA that is fully complementary to it, siRNA causes specific gene silencing.

A microRNA (miRNA) is a small non-coding RNA molecule (containing about 22 nucleotides) found in plants, animals and some viruses, that functions in RNA silencing and post-transcriptional regulation of gene expression. miRNAs function via base pairing with complementary sequences within mRNA molecules.
miRNA gene transcription is carried out by RNA polymerase II in the nucleus to give **primary miRNA (pri-miRNA)**, which is 5′ capped, 3′ polyadenylated RNA with double-stranded stem-loop structure. The pri-miRNA is then cleaved by a **Drosha** (an RNase III enzyme located on chromosome 5p) to form

precursor miRNA (pre-miRNA), which is a duplex that contains 70–100 nucleotides with interspersed mismatches and adopts a loop structure. The pre-miRNA is subsequently transported from the nucleus to the cytoplasm, where it is further processed by Dicer into a miRNA duplex of 18–25 nucleotides. The miRNA duplex then associates with the RISC forming a complex called miRISC. The miRNA duplex is unwound, releasing and discarding the passenger strand (sense strand). The mature single-stranded miRNA guides the miRISC to the target mRNAs. The miRNA binds to the target mRNAs through partial complementary base pairing with the consequence that the target gene silencing occurs via translational repression, degradation, and/or cleavage.



Translation in Prokaryotes

Introduction

The genetic information contained within the order of nucleotides in mRNA is interpreted to generate the linear sequences of amino acids in proteins. This process is known a translation. Translation is among the most highly conserved across all organisms and among the most energetically costly for the cell. The synthesis of a single protein requires the coordinated action of well over 100 proteins and RNAs.

Messenger RNA

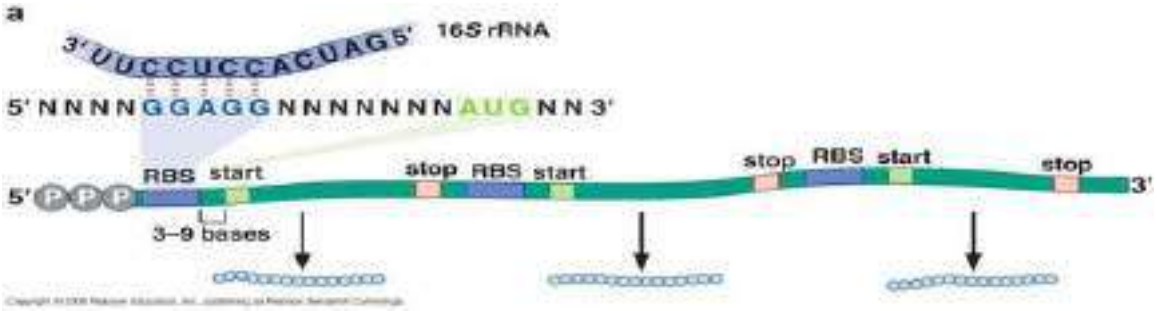
The protein coding region of each mRNA is composed of a contiguous of codons called an opening reading frame that specify a single protein. The start codon (the first codon) of an ORF:

- In bacteria : AUG, GUG, or UUG (5'-3')

- In eukaryotic cells: 5’-AUG-3’

The **start codon** has two important functions, specifies the first amino acid to be incorporated into the growing polypeptide chain and defines the reading frame for all subsequent codons.

RBS (ribosome binding site): a short sequence upstream of the start codon that facilitates binding of a ribosome. **AGGAGG**, also referred to as a Shine-Dalgarno sequence that interacts with 16S rRNA .



In polycistronic mRNA, the start codon of downstream ORF2 overlaps the stop codon of upstream ORF1, for example, with **AUG**A, translation of two ORFs is linked. This is known as **translational coupling** where the translation of one protein influences the translation of another protein.

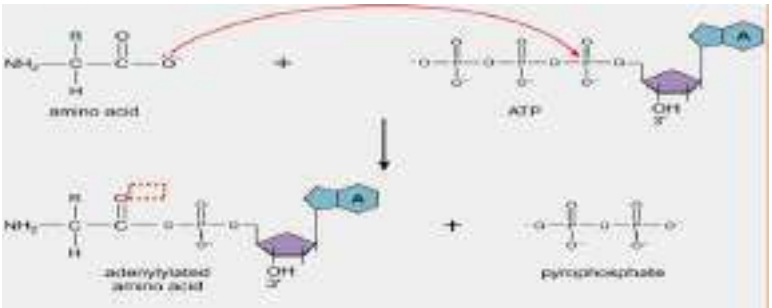
Translational coupling happens when the second gene of two adjacent genes in an operon, doesn't have it's own ribosome binding site. Instead, the ribosome ends translation at a stop codon at the first gene, and then steps back slighly, and starts translating the start codon of the second gene.

Transfer RNA

Translation of nucleotide sequence information into amino acids is accomplished by **tRNA** which acts as adaptors between codons and amino acids. There are many types of tRNA molecules in cell (~40). Each tRNA molecule is attached to specific amino acids (20) and each recognizes a particular codon, or codons (61), in the mRNA. All tRNAs end with the sequence CCA at the 3’ end, where the aminoacyl tRNA synthetase adds the amino acid. **Aminoacyl tRNA synthetase** catalyzing the reaction has three binding sites for ATP, amino acid and tRNA.

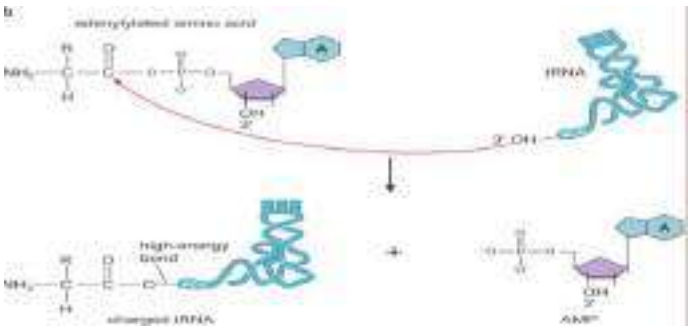
Step 1-Adenylylation of amino acids:

The aminoacyl-tRNA synthetase attaches AMP to the-COOH group of the amino acid utilizing ATP to form adenylated amino acid. Most organisms have 20 different tRNA synthetases.



Step 2- tRNA charging:

Transfer of the adenylated amino acid to the 3’ end of the appropriate tRNA via the 2’ or 3’-OH group, and the AMP is released as a result.



The Ribosome

The ribosome is the macromolecular machine that directs the synthesis of proteins. The ribosome is composed of at least 3 RNAs and more than 50 proteins. **In prokaryotes**, the transcription and the translation are coupled. The typical prokaryotic rate of translation is 20 amino acids /sec (60 nucleotides of mRNA).

The ribosome is composed of a large and a small subunit. **Peptidyl transferase center** is a part of the large subunit that is responsible for the formation of peptide bonds while **Decoding center** is a part of the small subunit where charged tRNA decodes the codon units of the mRNA. In prokaryote, the ribosome (70S) composed of 30S and 50S.The translation components undergo a specific series of events in which the small and large subunits of the ribosome associate with each other and the mRNA, translate the target mRNA, and then dissociate after completing synthesis of the protein. This sequence of association and dissociation is known as the **ribosome cycle**. An mRNA bearing multiple ribosomes is known as a **polyribosome** or a **polysome**.

Polypeptides are synthesized in the N- to C-terminal direction. The ribosome catalyzes the formation of a peptide bond between the amino acids attached to tRNAs. The ribosome has three binding sites for tRNA:

1. The A site is for the aminoacylated-tRNA.
2. The P site is for the peptidyl-tRNA.
3. The E site is for the exiting tRNA.

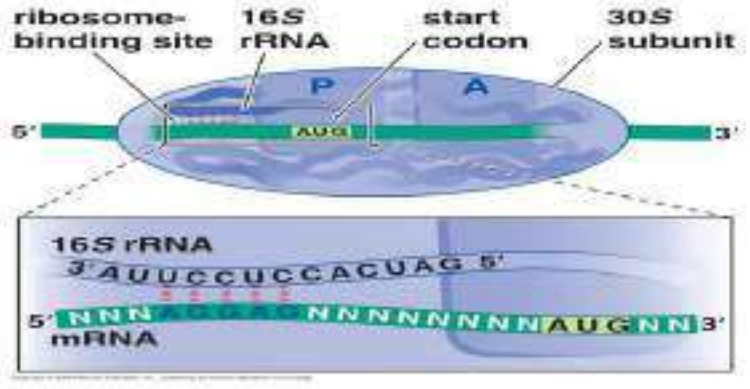
Translation occurs in three stages include:

1. **Initiation**

Prokaryotic mRNAs are initially recruited to the small subunit by base-pairing to rRNA. The small subunit associates with the mRNA first by base-pairing between the RBS and the 16S rRNA. The small subunit is positioned such that the start codon will be in the P site when the large subunit joins the complex. The large subunit joins its partner only at the very end of the initiation process.

a special tRNA that base-pairs with the start codon (usually AUG or GUG). The initiator tRNA gets charged with N-formyl methionine.

Deformylase: removes the formyl group from the amino terminus during or after the synthesis of the polypeptide chain.



Three translation initiation factors direct the assembly of an initiation complex:

IF1 prevents tRNAs from binding to the portion of the small subunit that will become part of the A site.

IF2 is a GTPase.

IF3 binds to the small subunit and blocks it from reassociating with a large subunit.

With all three IFs bound, the small subunit binds to mRNA and the initiator tRNA. Base-pairing between start codon and the initiator leads to conformational change of the small subunit, resulting in the release

of IF3. Binding of the large subunit. Hydrolysis of GTP bound to IF2. Release of IF2/GDP and IF1. Formation of 70S initiation complex.

2. Elongation

Codons are read 5´ to 3´ as the protein is synthesized from the amino end to the carboxyl end with aid of elongation factors such as **EF-Tu** to control these events. The correct addition of amino acids needs three key events to occur:

1. The correct aminoacyl-tRNA is loaded into the A site of the ribosome.
2. A peptide bond is formed between the aminoacyl-tRNA in the A site and the peptide chain that is attached to the peptidyl-tRNA in the P site.
3. The resulting peptidyl-tRNA in the A site and its associated codon must be translocated to the P site.

In the translocation step, the P-site tRNA must move to the E site. The A-site tRNA must move to the P site and the mRNA must move by three nucleotides to expose the next codon. The initial translocation steps are coupled to the peptidyl transferase reaction. It is require an elongation factor EF-G.

3. Termination

Release factors terminate translation in response to stop codons. RFs (release factors) recognize stop codons and activate the hydrolysis of the polypeptide from the peptidyl-tRNA. RF1 for UAG and UAA while RF2 for UGA and UAA. RF3 is a GTP-binding protein that leads to the dissociation of RF1/RF2 after peptide release. The ribosome gets separated from the tRNA and the mRNA and dissociate into its large and small subunits in a process called ribosome recycling.

Lec. 7
Translation in Eukaryotes

Eukaryotic translation is similar to that of prokaryotes but is more complex. Differences in the translation process is due to differences in transcript structure and location of translation. Unlike bacteria, in which translation initiation occurs as soon as the 5' end of an mRNA is synthesized, in eukaryotes such tight coupling between transcription and translation is not possible because transcription and translation are carried out in separate compartments of the cell (the nucleus and cytoplasm). Eukaryotic mRNA precursors must be processed in the nucleus (e.g., capping, polyadenylation, splicing) before they are exported to the cytoplasm for translation. The stages of translation included initiation, elongation and termination.

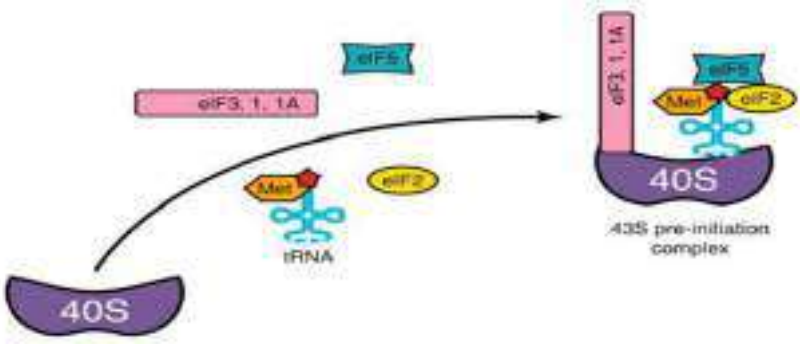
1. **Initiation:** Two types of translation initiation mechanisms: cap dependent translation initiation and cap-independent translation initiation.

Cap dependent translation initiation: In eukaryotes, the small subunit associated with the initiator tRNA is recruited to the 5' cap and scans along the mRNA until it reaches the first AUG. Most mRNAs of the cell are characterized by the 7mG at the 5′ end, called cap. The 3′ end of mRNA contains a polyadenylated tract (poly-A), which is attached to the poly-A- binding protein (PABP). Both cap and poly-A have been observed to play key roles in translation efficiency.

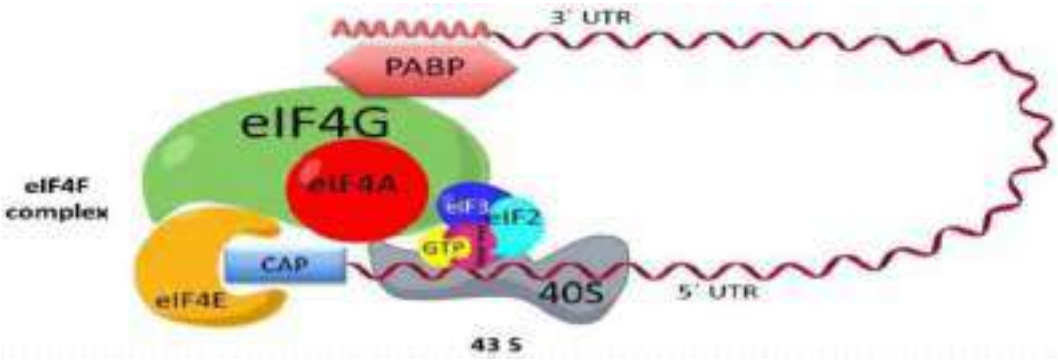
The initiation of translation in eukaryotes is complex, involving at least 12 eukaryotic initiation factors (eIFs), to translate an mRNA, it is important that the mRNAs joining to a protein complex called eIF4F (eIF4A, eIF4E and eIF4G) binds the 5' cap of messenger RNAs (mRNAs) to promote eukaryotic translation initiation. The initiation of translation divided into 4 steps :

a. Ribosomal dissociation. The 80S ribosome dissociates to form 40S and 60S subunits. Two initiating factors namely elF-3 and elF-1A bind to the newly formed 40S subunit and thereby block its reassociation with 60S subunit.

b. Formation of 43S preinitiation complex. A ternary complex containing met-tRNA and elF-2 bound to GTP attaches to 40S ribosomal subunit, eIF2 positions the Met-tRNA in the P site forming the 43S preinitiation complex.



c. Formation of 48S initiation complex. The binding of mRNA (5′ cap) to 43S preinitiation complex results in the formation of 48S initiation complex. The ribosomal initiation complex scans the mRNA for the identification of appropriate initiation codon (5'-AUG). The process is called scanning of mRNA. The mRNA is recruited to the **eIF4F complex** across the interaction of the 3′ end and poly-A- binding protein (PABP) and the 5′ cap and eIF4E.



d. Formation of 80S initiation complex. 48S initiation complex binds to 60S ribosomal subunit to form 80S initiation complex. The binding involves the hydrolysis of GTP (bound to elF- 2). This step is facilitated by the involvement of elF-5. As the 80S complex is formed, the initiation factors bound to 48S initiation complex are released and recycled.

Cap-independent translation initiation: What differentiates cap-independent translation from cap-dependent translation is that cap-independent translation does not require the 5' cap to initiate scanning from the 5' end of the mRNA until the start codon. The ribosome can be trafficked to the start site by direct binding. This method of translation has been found important in conditions that require the translation of specific mRNAs during cellular stress, when overall translation is reduced.

2. **Elongation**

Binding of Aminoacyl t-RNA to A- site. The 80S initiation complex contains met tRNA′ in the P- site and A-site is free. Another Aminoacyl-tRNA is placed in the A-site. This requires proper codon recognition on the mRNA and involvement of elongation factor 1a (EF-1a), this step need supply of energy by GTP. The Aminoacyl-tRNA is placed in the A-site, EF-1a and GDP are recycled to bring another Aminoacyl-tRNA. About six amino acids per second are incorporated during the course of elongation of translation in eukaryotes.

3. **Termination**

Termination of elongation depends on eukaryotic release factors. The process is similar to that of prokaryotic termination, but unlike prokaryotic termination, there is a universal release factor, eRF1, that recognizes all three stop codons. Upon termination, the ribosome is disassembled and the completed polypeptide is released. eRF3 is a ribosome-dependent GTPase that helps eRF1 release the completed polypeptide.

The Genetic Code

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 to impart specificity to each of the amino acid.

- 1 Nucleotide gives 4 combinations.
- 2 Nucleotides give 16 combinations.
- 3 Nucleotides give 64 combinations (Most suited for 20 amino acids). Three out of these are non sense codons and the 61 codons for 20 amino acids.

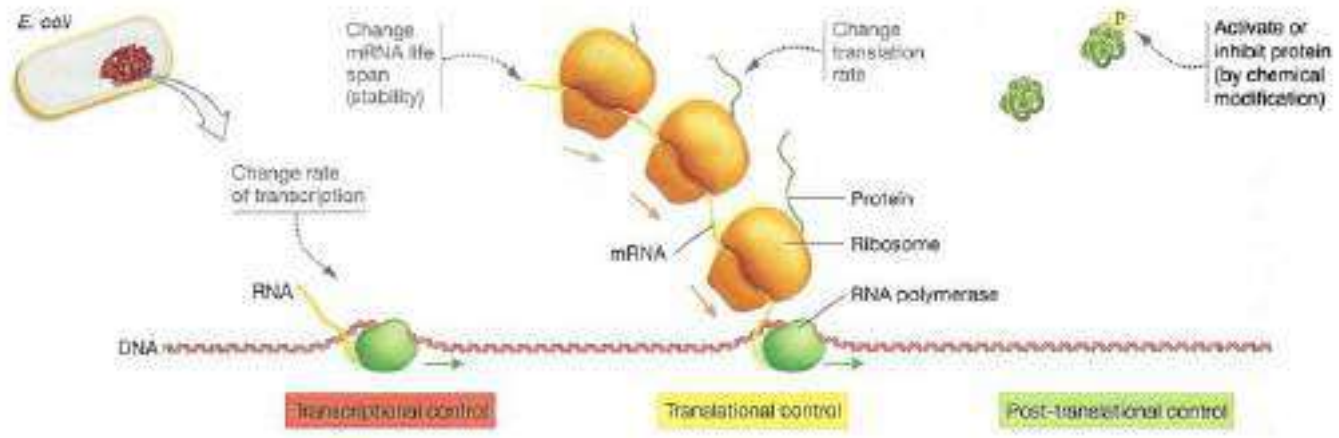
Properties of the Genetic code

-**Tiplate**: 3nucleotide give one amino acid.
-**Specific**: specific codon always codes for the same amino acid. e.g. UUU codes for Phenylalanine, it cannot code for any other amino acid.
-**Non-overlapping**: The genetic code is non-overlapping which means that the same letter does not take part in the formation of more than one codon.

-**Redundant or degeneracy**: one amino acid may be coded for by more than one codon, such as 4 codons to proline.
-**Commaless**: The genetic code is without comma i.e. no punctuations are required between the two codons.
- **Codons are read in a 5' to 3' direction**.
-**The code is nearly universal**: The code is nearly universal exceptions to the universal code are not limited to mitochondria but also found in several prokaryotic genomes and in the nuclear genomes for certain eukaryotes. UGA is not a stop signal but codes for tryptophan; methionine is encoded by both AUG and AUA.

Gene Regulation in Prokaryotes

Regulation of gene expression includes a wide range of mechanisms that are used by cells to increase or decrease the production of specific gene products (protein or RNA). Gene regulation is essential for viruses, prokaryotes and eukaryotes as allowing the cell to express protein when needed .Different potential modes of regulation of gene expression in prokaryote by controlling transcription, controlling translation, or controlling protein function post-translationally.



- **Transcriptional control**: Regulatory proteins affect the ability for RNA polymerase to bind to or transcribe a particular gene.
- **Translational control**: Various proteins may affect rate of translation or enzymes may affect lifetime of an mRNA transcript.
- **Post-translational control**: Translated protein may be modified by phosphorylation, which may change its folding and/or activity.

Gene Expression is Controlled by Regulatory Proteins

Genes are very often controlled by extracellular signals; in the case of bacteria, this typically means molecules present in the growth medium. These signals are communicated to genes by regulatory proteins, which come in two types: **positive regulators, or activators**, and **negative regulators, or repressors**. Typically, these regulators are DNA-binding proteins that recognize specific sites at or near the genes they control. An activator increases transcription of the regulated gene, and repressors decrease or eliminate that transcription. **Most activators and repressors act at the level of transcription initiation**.

Depending on the presence of regulator (activator and repressor) there are two types of regulations. **Positive regulation**, a transcriptional activator protein binds a region at the 5' end of a gene to promote transcription. In the absence of the activator protein, there is no transcription.

Negative regulation, a repressor binds the operator sequence to prevent transcription. Only in the absence of repressor protein can the gene be transcribed.

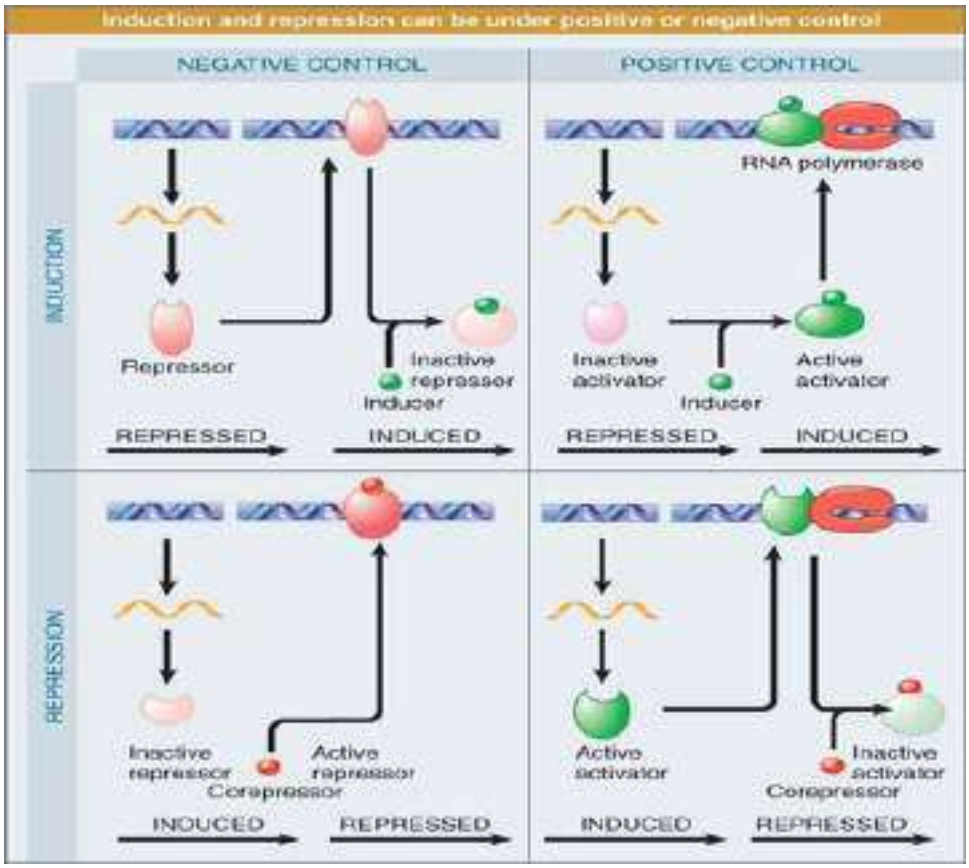
Operon

One strategy bacteria employ to control the expression of their genes: by grouping functionally related genes together for easily regulated. An operon is made up of 4 basic DNA components:

1. **Promoter**: a nucleotide sequence that enables a gene to be transcribed.
2. **Regulator gene**: it is gene product control the operator.
3. **Operator**: a segment of DNA that a repressor binds to it.
4. **Structural genes**: the genes that are regulated by the operon.

Operon regulation can be either negative or positive by induction or repression.

- **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. The lac operon is a negatively controlled inducible operon, where the inducer molecule is allolactose.
- **In negative repressible operons**, transcription of the operon normally takes place. Repressor proteins are produced by a regulator gene, but they are unable to bind to the operator in their normal conformation. However, certain molecules called corepressors are bound by the repressor protein, causing a conformational change to the active site. The activated repressor protein binds to the operator and prevents transcription. The trp operon, involved in the synthesis of tryptophan (which itself acts as the corepressor), is a negatively controlled repressible operon.



Positive control, an activator protein stimulates transcription by binding to DNA (usually at a site other than the operator).

- **In positive inducible operons**, activator proteins are normally unable to bind to the DNA. When an inducer is bound by the activator protein, it undergoes a change in conformation so that it can bind to the DNA and activate transcription.
- **In positive repressible operons**, the activator proteins are normally bound to the DNA segment. However, when an inhibitor is bound by the activator, it is prevented from binding the DNA. This stops activation and transcription of the system.

mexAB-oprM operon

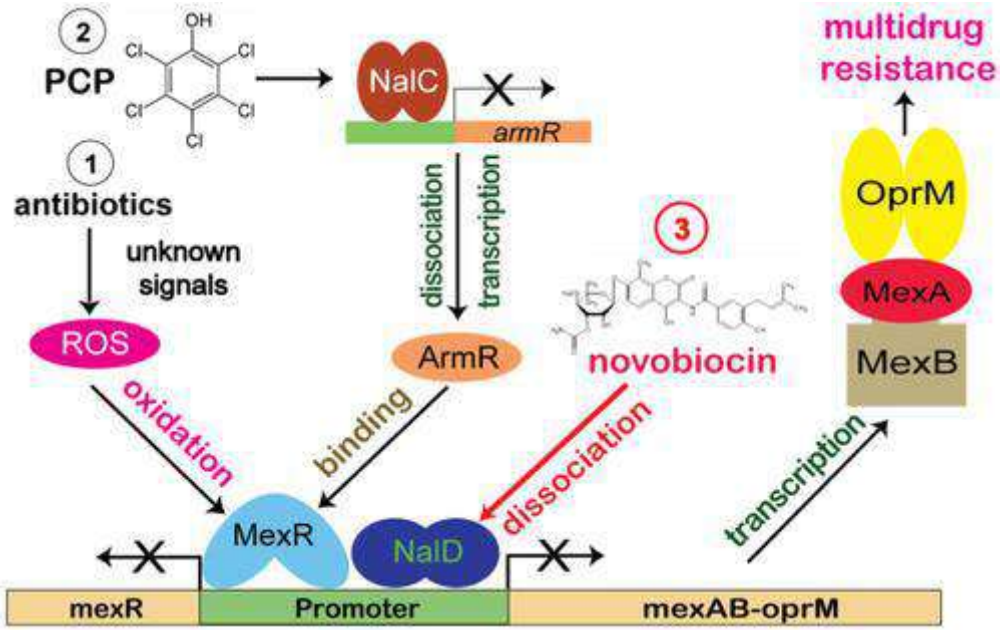
It is multidrug efflux pump operon in *P. aeruginosa*, is able to export drugs from several different classes, exist as a tripartite system consisting of periplasmic membrane fusion protein, outer membrane factor, and cytoplasmic membrane transporter. Expression of the MexAB-OprM is governed mainly by regulatory proteins such as **mexR, nalC and nalD** amongst several others.

mexR is major regulator, it gene is located upstream of the mexAB-oprM operon. MexR binds to this intergenic region as a stable homodimer and represses transcription from the mexAB-oprM operon. Its utilizes an oxidation-sensing mechanism to regulate the virulence and antibiotic resistance of *P. aeruginosa*.

NalD, acts as a secondary repressor of the MexAB-OprM multidrug efflux system by binding to a sequence between mexAB-oprM and the mexR binding site proximal to the mexA promoter. Mutations in mexR and NalD resulting in loss of dimerization and binding capacity of proteins, lead to hyperexpression of MexAB-OprM and increased resistance toward a range of different antibiotics.

ArmR whose expression is regulated by NalC acts as an anti-repressor by inhibiting the MexR-DNA interaction resulting in derepression of mexAB-oprM.

Three known pathways are presented to mediate MexAB-OprM transcriptional expression. MexR is the primary repressor of the MexAB-OprM pump. The oxidation signal could cause MexR conformational change, and leads to its dissociation from the DNA promoter region (**pathway 1**). NalC binding to pentachlorophenol (PCP) results in de-repression of ArmR. The overexpression of ArmR de-represses the MexAB-OprM pump via an allosteric MexR-ArmR interaction which prevents the binding of MexR to its cognate DNA operator (**pathway 2**). NalD is the secondary repressor of the MexAB-OprM pump, novobiocin binds directly to NalD, which leads NalD to dissociate from the DNA promoter (**pathway 3**).



Gene Regulation in Eukaryotes

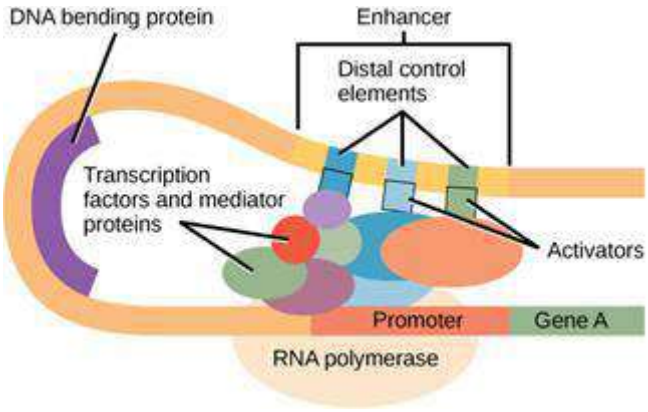
Eukaryotic transcription Gene Regulation

Like prokaryotic cells, the transcription of genes in eukaryotes requires the action of an RNA polymerase to bind to a DNA sequence upstream of a gene in order to initiate transcription. However, unlike prokaryotic cells, the eukaryotic RNA polymerase requires other transcription factors to facilitate transcription initiation. There are two types of transcription factors that regulate eukaryotic transcription: *General (or basal) transcription factors* bind to the core promoter region to assist with the binding of RNA polymerase. *Specific transcription factors* bind to various regions outside of the core promoter

region and interact with the proteins at the core promoter to enhance or repress the activity of the RNA polymerase.

Genes are organized to make the control of gene expression easier. The promoter region is immediately upstream of the coding sequence. This region can be short (only a few nucleotides in length) or quite long (hundreds of nucleotides long). The longer the promoter, the more available space for proteins to bind. This also adds more control to the transcription process. Within the core promoter region, 25 to 35 bases upstream of the transcriptional start site, resides the TATA box. which contains a TATA-binding protein. other binding sites are found in some promoters. Examples of these elements are the CAAT box, with the consensus sequence 5'-CCAAT-3' and the GC box, with the consensus sequence 5'-GGGCGG-3'.

In some eukaryotic genes, there are additional regions that increase or enhance transcription. These regions called **enhancers**. They can be located upstream of a gene, within the coding region of the gene, downstream of a gene, or may be thousands of nucleotides away. Enhancer regions are binding sequences for specific transcription factors. When a protein transcription factor binds to its enhancer sequence, the shape of the protein changes, allowing it to interact with proteins at the promoter site. However, since the enhancer region may be distant from the promoter, the DNA must bend to allow the proteins at the two sites to come into contact. This shape change allows for the interaction of the specific activator proteins bound to the enhancers with the general transcription factors bound to the promoter region and the RNA polymerase.



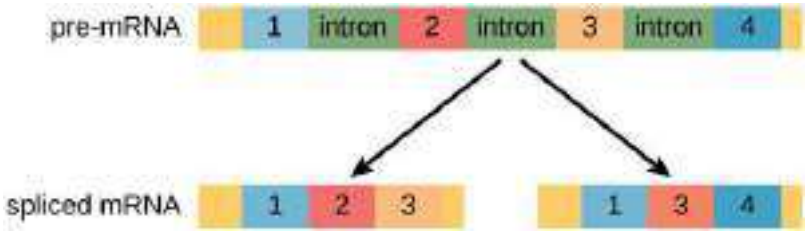
Each enhancer is made up of short DNA sequences called distal control elements. Activators bound to the distal control elements interact with mediator proteins and transcription factors. Two different genes may have the same promoter but different distal control elements, enabling differential gene expression.

Like prokaryotic cells, eukaryotic cells also have mechanisms to prevent transcription. **Transcriptional repressors** can bind to promoter or enhancer regions and block transcription. Like the transcriptional activators, repressors respond to external stimuli to prevent the binding of activating transcription factors.

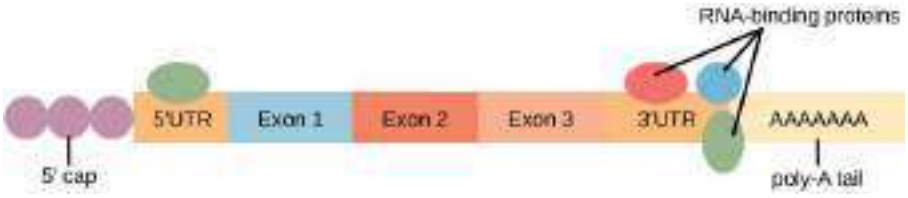
Eukaryotic Post-transcriptional Gene Regulation

RNA is transcribed, but must be processed into a mature form before translation can begin. This processing that takes place after an RNA molecule has been transcribed, but before it is translated into a protein, is called *post-transcriptional modification* which can also be regulated gene expression in the cell.

RNA Splicing After an RNA molecule has been transcribed, the RNA is processed and the introns are removed by splicing which is done by spliceosomes. Alternative splicing is now understood to be a common mechanism of gene regulation in eukaryotes, with the frequency of different splicing alternatives controlled by the cell as a way to control the production of different protein products in different cells or at different stages of development.



Control of RNA Stability Once the RNA is transported to the cytoplasm, the length of time that the RNA resides there can be controlled. Each RNA molecule has a lifespan and decays at a specific rate. This rate of decay can influence how much protein is in the cell. If the decay rate is increased, the RNA will not exist in the cytoplasm as long, shortening the time available for translation of the mRNA to occur. Conversely, if the rate of decay is decreased, the mRNA molecule will reside in the cytoplasm longer and more protein can be translated.



Binding of proteins to the RNA can also influence its stability. Proteins called RNA-binding proteins can bind to the regions of the mRNA just upstream or downstream of the protein-coding region can increase or decrease the stability of an RNA molecule. These regions in the RNA that are not translated into protein are called the untranslated regions, or UTRs. They are not introns, these regions regulate mRNA localization, stability, and protein translation. The region just before the protein-coding region is called the 5' UTR, whereas the region after the coding region is called the 3' UTR.

RNA Stability and microRNAs mature miRNAs recognize a specific sequence and bind to the RNA complementary sequences on an mRNA and either impede translation of the message or lead to the degradation of the mRNA.

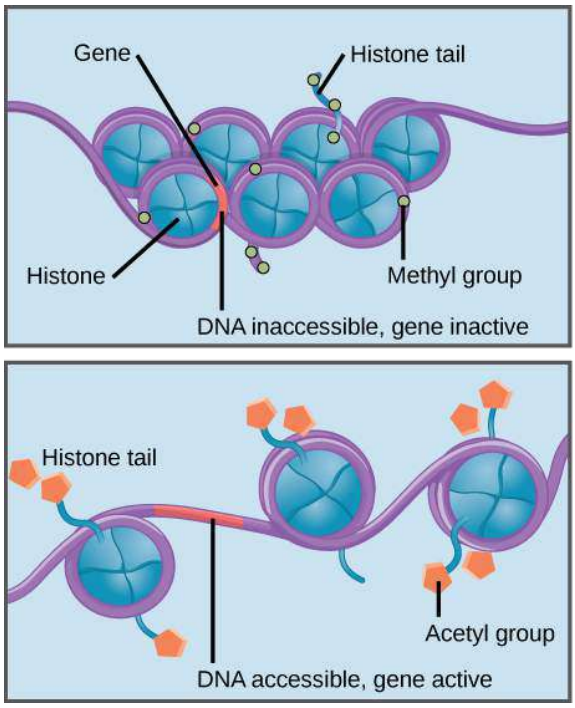
RNA editing RNA editing is any process, other than splicing, which information change at the level of mRNA. It is revealed by situations in which the coding sequence in an RNA differs from the sequence of DNA from which it was transcribed.

Eukaryotic Epigenetic Gene Regulation

Transcriptional access to the DNA can be controlled in two general ways: chromatin remodeling and DNA methylation. Chromatin remodeling changes the way that DNA is associated with chromosomal histones. DNA methylation is associated with developmental changes and gene silencing.

The DNA in the nucleus is precisely wound, folded, and compacted into chromosomes so that it will fit into the nucleus. It is also organized so that specific segments can be accessed as needed by a specific cell type.

If DNA encoding a specific gene is to be transcribed into RNA, the nucleosomes surrounding that region of DNA can slide down the DNA to open that specific chromosomal region and allow for the transcriptional machinery (RNA polymerase) to initiate transcription. Modifications to the histones and DNA affect nucleosome spacing.



Methylation of DNA and histones causes nucleosomes to pack tightly together. Transcription factors cannot bind the DNA, and genes are not expressed.

Histone acetylation results in loose packing of nucleosomes. Transcription factors can bind the DNA and genes are expressed.

Some chemical groups (phosphate, methyl, or acetyl groups) are attached to specific amino acids in histone "tails" at the N-terminus of the protein. These groups do not alter the DNA base sequence, but they do alter how tightly wound the DNA is around the histone proteins. DNA is a negatively charged molecule and unmodified histones are positively charged; therefore, changes in the charge of the histone will change how tightly wound the DNA molecule will be. By adding chemical modifications like acetyl groups, the charge becomes less positive, and the binding of DNA to the histones is relaxed. Altering the location of nucleosomes and the tightness of histone binding opens some regions of chromatin to transcription and closes others.

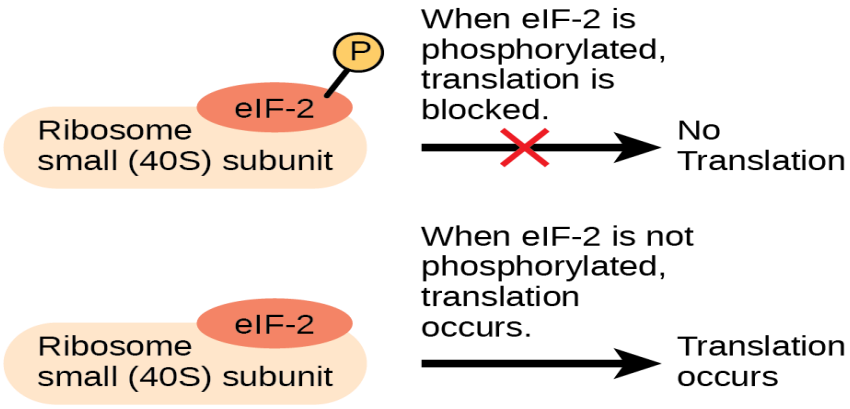
If a gene is to be transcribed, the histone proteins and DNA in the chromosomal region encoding that gene are modified in a way that opens the promoter region to allow RNA polymerase and transcription factors to bind and initiate transcription. If a gene is to remain turned off, or silenced, the histone proteins and DNA have different modifications that signal a closed chromosomal configuration lead to RNA polymerase and transcription factors do not have access to the DNA and transcription cannot occur.

Eukaryotic Translational and Post-translational Gene Regulation

After RNA has been transported to the cytoplasm, it is translated into protein. Control of this process is largely dependent on the RNA molecule. The stability of the RNA will have a large impact on its translation into a protein.

The Initiation Complex and Translation Rate

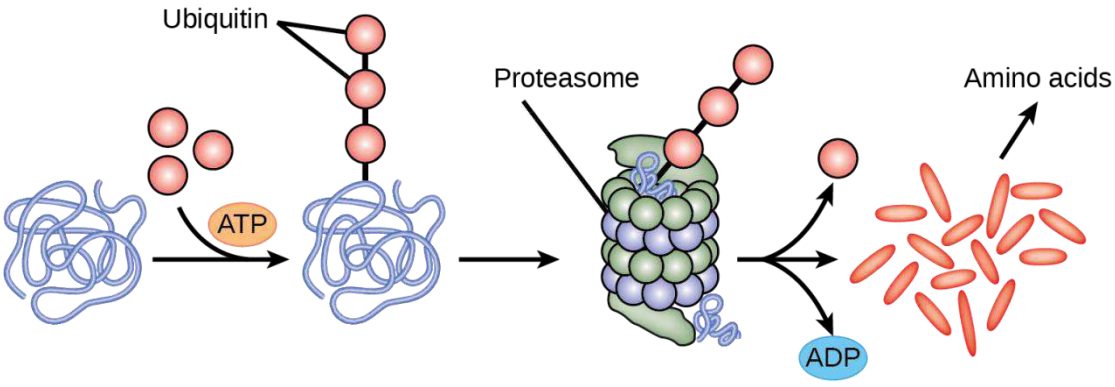
Translation is controlled by proteins that bind and initiate the process. In translation, the complex that assembles to start the process is referred to as the translation initiation complex. The binding of eIF-2 to the RNA is controlled by phosphorylation. If eIF-2 is phosphorylated, it undergoes a conformational change and cannot bind to GTP. Therefore, the initiation complex cannot form properly and translation is impeded. When eIF-2 remains unphosphorylated, the initiation complex can form normally and translation can proceed. Gene expression can be controlled by factors that bind the translation initiation complex.



Chemical Modifications, Protein Activity, and Longevity

Proteins can be chemically modified with the addition of groups including methyl, phosphate, acetyl, and ubiquitin groups. The addition or removal of these groups from proteins regulates their activity or the length of time they exist in the cell. Sometimes these modifications can regulate where a protein is found in the cell—for example, in the nucleus, in the cytoplasm, or attached to the plasma membrane.

The addition of an ubiquitin group to a protein marks that protein for degradation. Ubiquitin acts like a flag indicating that the protein lifespan is complete. These proteins are moved to be degraded. One way to control gene expression, therefore, is to alter the longevity of the protein.



Bacterial Genome

lec. 8

Prof. Dr Hayfa H. Hassani
haiphahassani@gmail.com
hayfa.hassani@sc.uobaghdad.edu.iq
2021

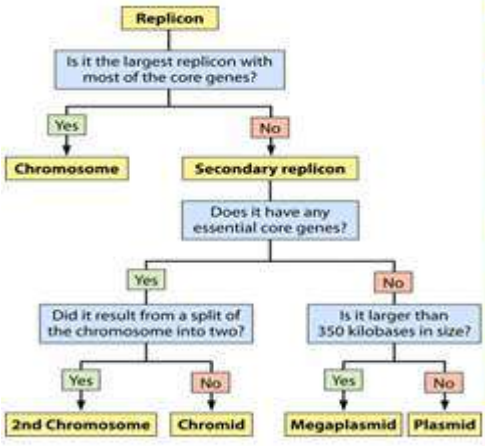
Bacterial Genome

Bacterial Genome Includes:

- **Chromosomes and other replicons** (Plasmids, megaplasmids, chromid, and 2nd chromosomes).

- **Regulators regions such as:**
  - Operons
  - Transposons (retrotransposons & DNA transposons)
  - Repetitive DNA sequences (Microsatellites)
  - ncRNA (Structural & Functional ncRNA)

Classification of bacterial replicons



Replicon refers to any DNA molecule regardless of its specific nature.

Classification of replicon:
Each replicon can be classified based on specific characteristics such as origin and size.
In general, each replicon can be classified into one of the following groups:

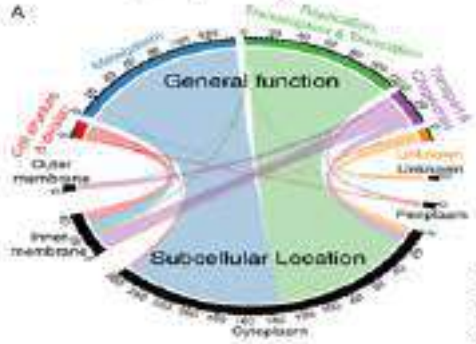
1. **Chromosome** refers to the primary replicon.

2. **secondary replicon** refers to any replicon that is not the primary chromosome of the cell which included:

2nd Chromosome, chromid, megaplasmid, and **plasmid.**

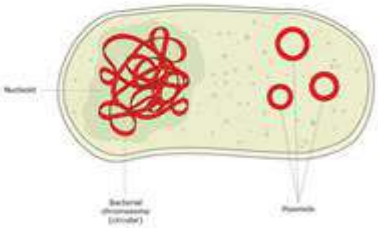
1. Bacterial Chromosome

chromosome is always the largest replicon in the genome and contains the majority of the core/essential genes. The bacterial chromosomes are with average and median sizes of ~3.65 Mb and ~3.46 Mb, respectively.



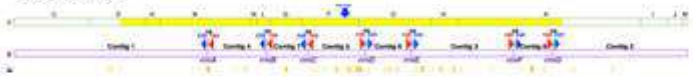
There are core (essential) genes included subcellular location (bottom) and general function (top)

Structure of bacterial genome

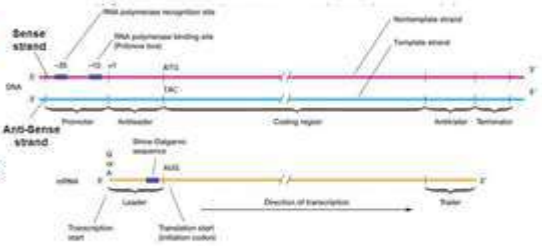


The vast majority of bacterial genomes are circular, consisting usually of large chromosomes and small plasmids.

Some bacteria harbor linear genomes, including some that are industrially important, such as *Streptomyces coelicolor*.



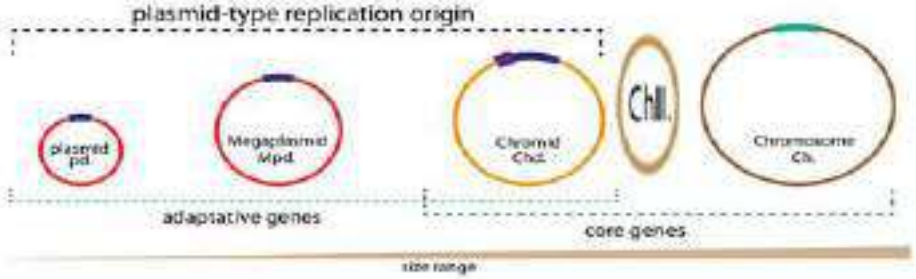
Organization of typical bacterial gene



structural genes of related function are often organized on the genome and transcribed under the control of a single promoter.

Other Circular Replicons in Bacterial Genome

- **Plasmid,**
- **Megaplasmid,**
- **Chromid,**
- **2nd Chromosome**



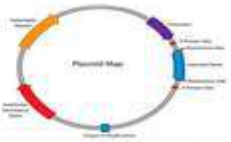
Plasmid and Megaplasmid

Plasmid & Megaplasmid

They are secondary replicons in bacterial genome that carry no core genes, and they are nonessential for cell viability in most environments.

Origin of their genes

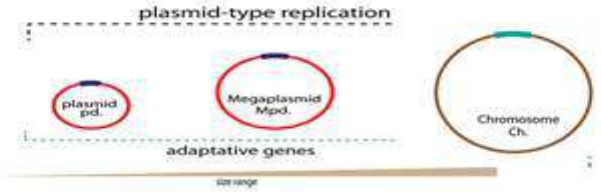
The majority of the genes on these replicons were acquired through horizontal gene transfer (HGT).



Comparison between Plasmid & Megaplasmid:

The distinction between plasmid and megaplasmid is currently based on size.

- Megaplasmid size: the average median size of megaplasmid is ~78.9 kb. As this is equal to roughly 10% of the median bacterial genome size.
- Plasmids size: the average median size of plasmid is ~46.2 kb .



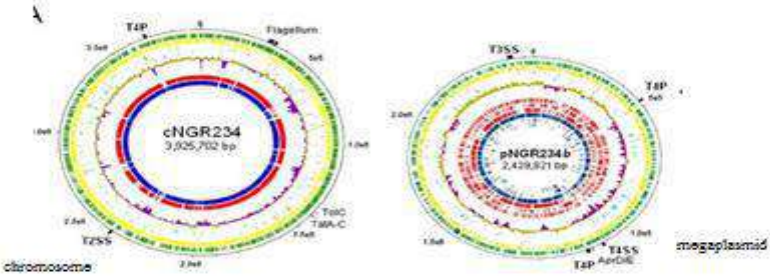
Origin of Megaplasmid

(1) The acquisition of a megaplasmid (orange) expands the niche range (special zone) of the cell.

(2) this replicon accumulates horizontally acquired metabolic genes(purple) and carry a plasmid-type replication origin (*oriC*) that provide a fitness advantage in the novel environment.

(3) Loss of Megaplasmid is favored in the cell's original niche.

Chromosome and Megaplasmid



Map of the genome of *Rhizobium* sp. strain NGR234 chromosome (cNGR234, 3.93 Mbp), megaplasmid (pNGR234 b , 2.43 Mbp).

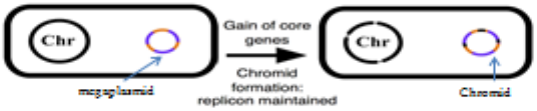
The outer circle shows the coordinates and **the position of the secretion systems and several other selected genes**; the second and the third outer circles indicate the **ORFs on the leading and the lagging strands** (yellow and green).

The next inner circle indicates the **sncRNAs (in light red)**, followed by a circle indicating the positions of **the repeats** (in light blue).

The next inner circle shows the **GC content (dark yellow and purple)**.

Chromid

Chromid results from the transferring of a gene from the chromosome (black) makes the megaplasmid (now a chromid) needed in all environments. .



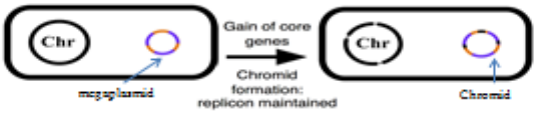
Chromid size has average (~1.52 Mb) and median (~1.26 Mb) sizes.
However, unlike plasmids and megaplasmids, chromids carry at least one gene that is essential for cell viability (i.e., a core gene whose loss would result in cell death) and adaptive genes.

2nd Chromosome

2nd chromosome results from splitting of chromosome, and is carrying core genes, and has non self-transmisible (no –passed) genes.

Chromid

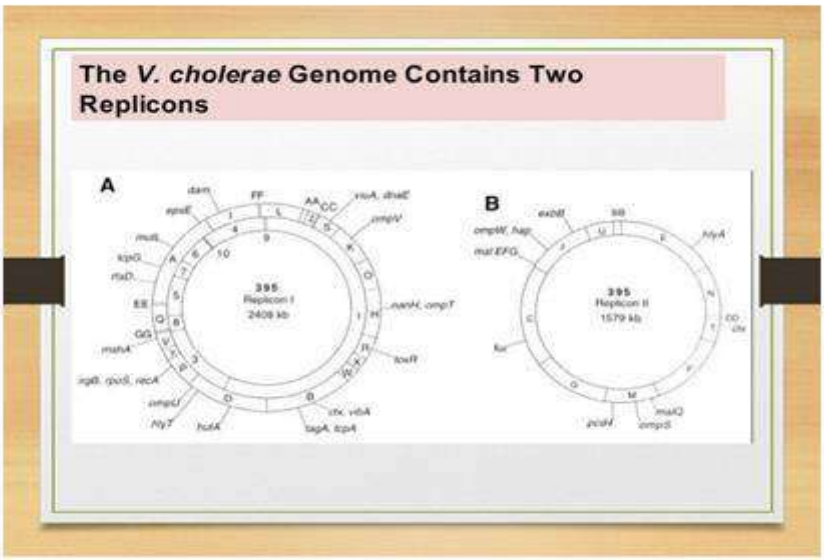
Chromid results from the transferring of a gene from the chromosome (black) makes the megaplasmid (now a chromid) needed in all environments. .



Chromid size has average (~1.52 Mb) and median (~1.26 Mb) sizes.
However, unlike plasmids and megaplasmids, chromids carry at least one gene that is essential for cell viability (i.e., a core gene whose loss would result in cell death) and adaptive genes.

2nd Chromosome

2nd chromosome results from splitting of chromosome, and is carrying core genes, and has non self-transmisible (no –passed) genes.



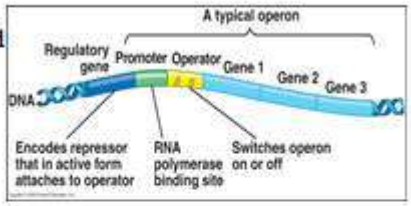
Regulators in Bacterial Genome

(A): Operon
(B): Repetitive DNA Sequences (DNA microsatellite)
(C): Transposons (DNA, retrotransposons)

(D): non-coding RNA (ncRNA)

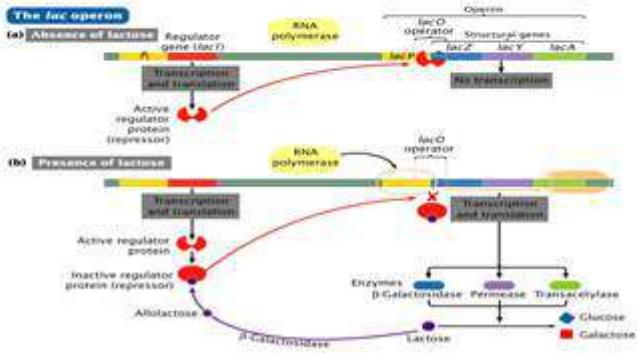
(A) Operon as a model for gene expression in bacteria

Operon is a polycistronic transcription unit.
The operon region includes both the promoter and the operator.



(a) repressor binds to the operator, the structural genes will not be transcribed.

(b) activators may bind to the regulatory region, enhancing transcription.

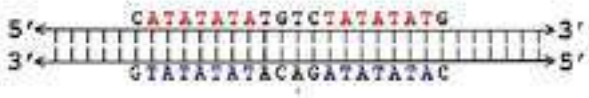


(B): Repetitive DNA Sequences in Bacteria

- Microsatellites
- Transposable genetic elements

Background

All DNA genomes contain amounts of repetitive sequences that are larger than expected for random distribution of bases, but the percentage of repetitive sequences varies greatly across different organisms. For example, genome of *E. coli* contains only 0.7% of repeats while at least 50% of the human genome are repetitive.



repetitive DNA sequences influence the genetic stability including replication, recombination, repair, and transcription.

▪ Microsatellite in Bacteria

Microsatellite sequences consist of up to six bases per repeat.

Type of microsatellite	Example sequence	Shorthand nomenclature
Uninterrupted	CCCCCC CTCTCTCTCTCTCT CTGCTGCTGCTGCTGCTG	$(C)_6$ $(CT)_7$ or $(TC)_7$ $(CTG)_6$
Interrupted	CCACCC CTCTAACTCTCT CTGCTGCTGAAACTGCTG	$(C)_2A(C)_3$ $(CT)_2AA(CT)_3$ $(CTG)_2AAA(CTG)_2$
Compound	CCCAAA CTCTCTCACACA CTGCTGCTGCACGCACGCAG	$(C)_3(A)_3$ $(CT)_3(CA)_3$ $(CTG)_3(CACG)_2$

Uninterrupted microsatellite, means **repeats of the same sequence** that may consist of one base (mononucleotide), two different bases (dinucleotide), or three different bases (trinucleotide).

Interrupted (or "imperfect") microsatellite, when one or more bases interrupt the repeat array (shown by the bases in red).

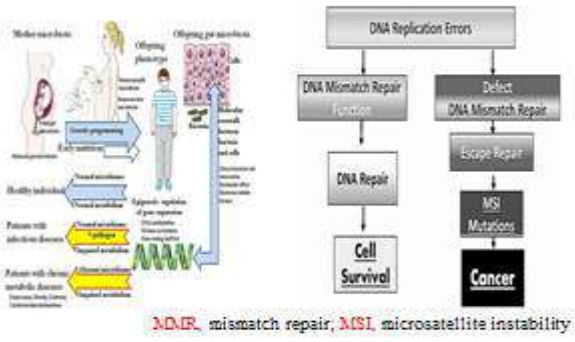
Compound (or "composite") microsatellites, when two types of repeat that neighbor each other are found frequently in genomes.

Role of Microsatellite in Bacteria

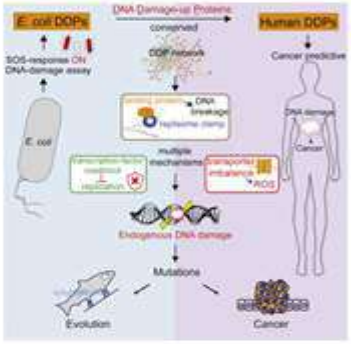
In bacteria, microsatellite variants prompt generation of variations, ensuring the survival of the bacterial population in different environments, and adaptation to pathogenesis and virulence.

Bacterial Repetitive sequences and human Health

Defective mismatch repair (MMR), nucleotide excision repair (NER) and elevated microsatellite instability will increase antigenic variance of the pathogens and this are linked to some human diseases.



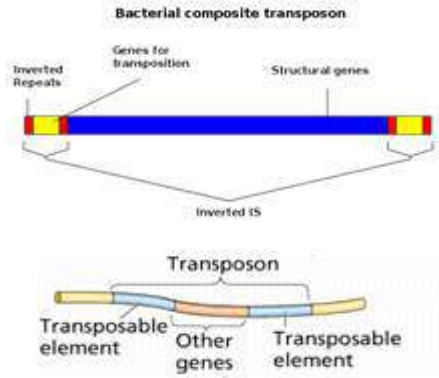
MMR, mismatch repair; MSI, microsatellite instability

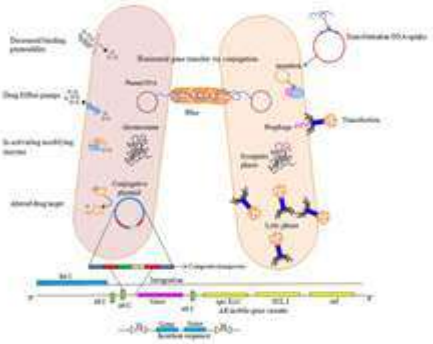


Replication errors can be repaired by a functional DNA MMR mechanism that is capable of recognizing and successfully repairing these errors.
The dysregulated DNA MMR mechanism cannot recognize the DNA replication errors leading to genetic alterations, such as MSI and/or mutations in functional sites of DNA that subsequently may lead to carcinogenesis.

(C) Bacterial Transposon

Transposons are mobile genetic elements (MGE) that can carry additional genetic cargo non-essential for their own transposition. This cargo can include antibiotic resistance genes or heavy metal genes or those increasing metabolic plasticity. MGEs are transposing across or between the chromosome and other replicons in a single cell.





Types of Bacterial Transposon

Classes of Bacterial Transposons		Characteristics of Bacterial Transposons	Role in bacterial cell
1. Retrotransposons		transpose by an RNA intermediate	
a	Retrons (multicopy-single stranded DNA, msDNA))	msDNA, produced from reverse transcript, covalently linked to RNA (msDNA)	Defense agent (anti-phages)
b	Diversity-generating retrotransposons (DGR)	Compose of target gene, variable template, template repeat, and reverse transcriptase gene.	Evolution by creation new variations of bacterial cell tin the target genes through introducing sequence diversity of DNA molecule that result from mutagenic homing.
2. DNA transposons		transpose by an DNA intermediate	
a.	Insertion sequences (ISs)	transposonase gene flank on both ends by Inverted repeat sequences (IR)	Mutation in the genes
b.	Transposons (Tns)	segment of drug resistance genes flank with either Is or IR	Antibacterial resistance
i	Composite transposons	have a central segment containing genes for drug resistance flank on both ends by IS-elements	=
ii	Non-composite	have genes for drug resistance flank on both ends by IR	=

Types of Transposons

MGEs are classified into two major groups based on their structure and transposition mechanism:

1. group I : **Retrotransposons**, which transpose by an RNA intermediate and include:

A. **Retrons:** genetic elements that produce multicopy single-stranded DNA covalently linked to RNA (msDNA) by a reverse transcriptase

B. **Diversity generating retroelements (DGRs):** introducing vast amounts of sequence diversity into target genes.

2. group II: **DNA transposons**, which use a DNA intermediate.

A. **Insetion elements (ISs)**

B. **Transposons** :

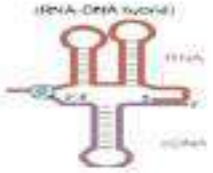
i- Composite transposon

ii- non-composite transposons

Group I: Retrotransposons

They are transpose by an RNA intermediate which include:

A. Retrons (msDNA)



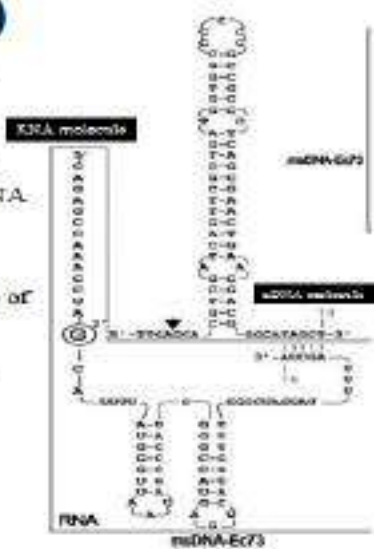
B. Diversity generating retroelements (DGR)

Group (1-A). Retron (msDNA)

Retrons are multicopy single-stranded DNA (msDNA).

Structure of Retrons :

- They compose of RNA molecule (boxed sequence) joined to the reversed transcribed single-stranded DNA (cDNA) molecule.
- The 5′ end of the DNA chain is joined, via a 2′,5′ phosphodiester bond, to a specific guanosine residue of the RNA molecule (circled G).
- Both the RNA chain and the DNA chain fold-up into stable stem-loop structures.



Group (I- B). Diversity-generating retroelements (DGRs)

Diversity-generating retroelements (DGRs) are a unique family of retroelements that generate sequence diversity of DNA to benefit their hosts by introducing variations and accelerating the evolution of target proteins.

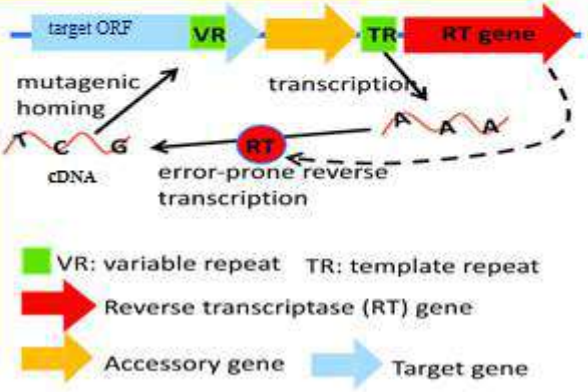
They exist widely in bacteria, archaea, phage and plasmid.

Components and Mechanism of DGR sytsem

MECHANISM

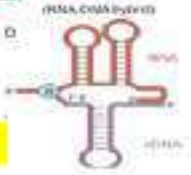
RNA transcript is made from the TR, which is then reversely transcribed to cDNA by the RT in a process termed mutagenic homing.

The mutagenized cDNA replaces the parental VR in the target ORF, thereby altering the host gene.



Function of Bacterial Retron

Retron-containing systems are abundant in genomic "defense islands", suggesting a role for most retrons in phage resistance. The defensive unit is composed of three components: *RT, non-coding RNA (ncRNA), and an effector protein.*

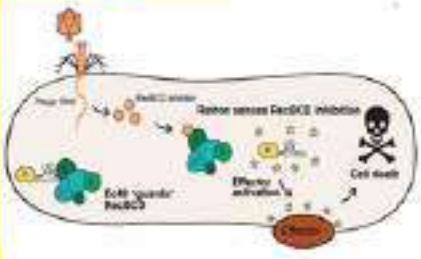


bacterial defence by retron

e.g. retron Ec48 is a retron found in *E. coli* whose reverse transcribed DNA segment is 48 nt long.

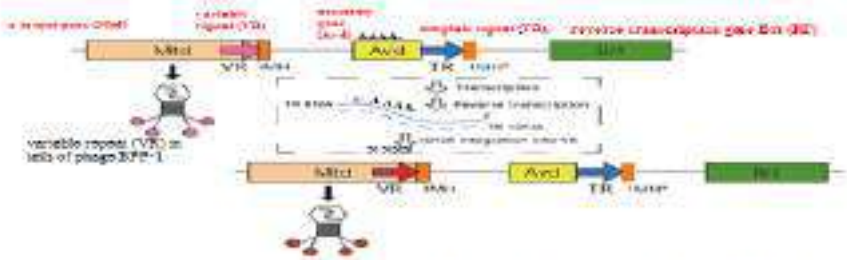
This retron shows evidence that it is a "guardian" of RecBCD (a helicase–nuclease that initiates the repair of double-stranded DNA breaks by homologous recombination. It also degrades linear double-stranded DNA, protecting the bacteria from phages and extrachromosomal DNA).

Inhibition of RecBCD by phage proteins will activate the retron, leading to abortive infection and cell death.



DGR in Bordetella phage BPP-1

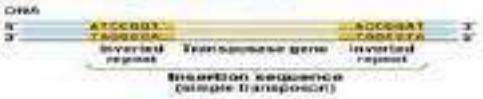
Bordetella phage BPP-1 DGR is mainly composed of a target gene (Mtd) with a variable repeat (VR) in the tail, a template repeat (TR), a reverse transcriptase gene Brt (RT) and an accessory gene (Avd).



Reverse transcription mediated by ***Brt*** gene is the key procedure of DGR mechanism, in which adenine-specific mutagenesis (A-to-N substitution) occurs and TR cDNA is generated.
TR cDNA is then integrated into the homologous VR, region (cDNA integration), which may diverse the target gene.
Avd (Accessory variability determinant) acts as an accessory gene interacts with RT, which is essential for the cDNA synthesis and generate diversity.

Group II: DNA transposon

Class 1: Insertion Sequences (ISs)



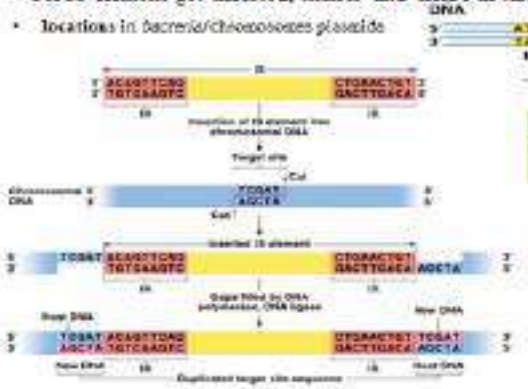
Class 2: Transposons (Tn)

(A): Composite Tns

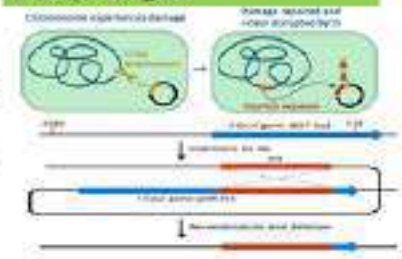
(B): non-composite Tns

Class 1: Insertion Sequences (ISs)

- IS elements contain single open reading frame (ORF) which encodes for the enzyme transposonase, catalyzing its own transposition.
- Once recognizing the target site, it generate staggered cleavage (cut the single strand of DNA) generating sticky and itself get inserted.
- As IS element get inserted, then it and helps in target site duplication.
- locations in bacteria/chromosomes plasmids



Role of ISs in the Bacterial Cells:
mutation that alters or destroys the activity of the gene

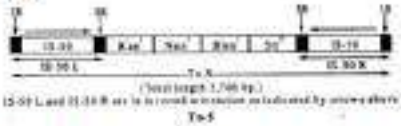


Class 2: Transposons (Tns)

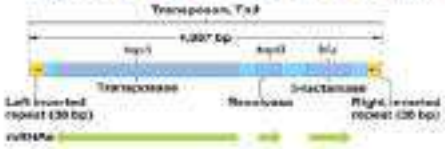
Transposons elements (Tns) are longer than ISs, usually the carrier of additional genes for antibiotic resistance. Transposons can transfer from a plasmid to other plasmids or from a DNA chromosome to plasmid and vice versa that cause the transmission of antibiotic resistance genes in bacteria.

Transposons are distinguished into two main types:

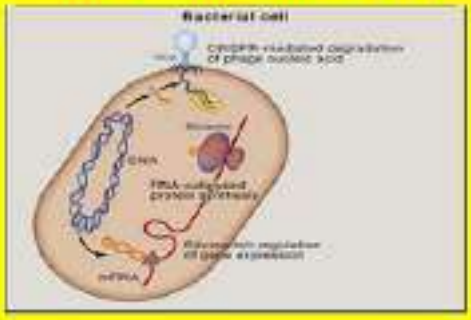
1) **Composite transposons** have a central segment containing drug resistance genes and flank on both ends by IS-elements.



2) **Non-composite transposons** are without IS-elements, but they do have inverted repeats at both ends (IR Left & IR Right). They also have genes for drug resistance.



(D) NON-CODING RNA(NCRNA) IN BACTERIA

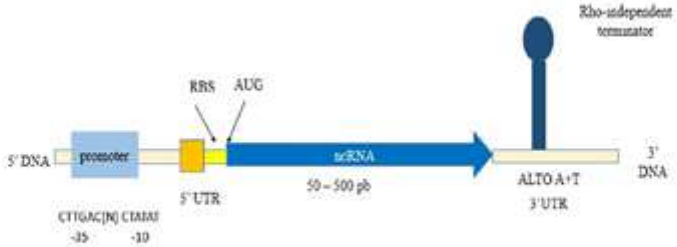


Characteristic of Bacterial ncRNA molecules

Bacterial ncRNAs are small RNA transcripts (typically 50–500 nt) that are not translated into proteins. They are involved in many biological processes, such as metabolic regulation, adaptation to environmental conditions, stress response, regulation of morphology, and cellular behavior.

Structure of bacterial ncRNAs

- CTTGAC (N) (-35) and CTATAT (-10) promoter region,
- ribosome binding site (RBS),
- AUG coding region at the 5'UTR,
- ncRNA vary in size from 50 -500 bp,
- 3'UTR region has a terminator sequence (Rho-independent terminator) which controlled by the specific sequences of RNA.



Intrinsic or Rho-independent termination: is a mechanism for termination of DNA transcription. This termintor is characterized by a RNA hairpin followed by a run of 6–8 U residues.

Role of Bacterial ncRNAs

ncRNA can be divided into two types:

❑ **1: Structural function** (e.g tmRNA)

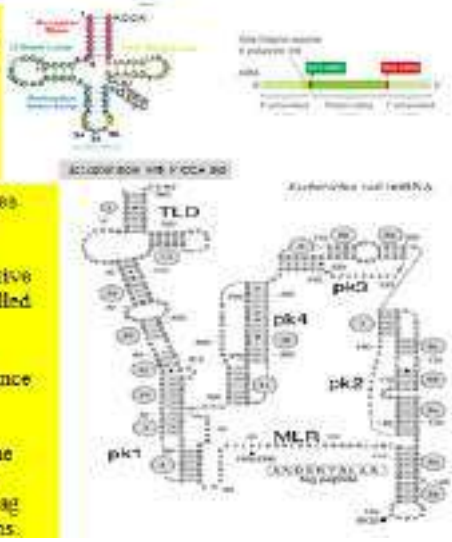
❑ **2: Regulatory function:**

(I) riboswitches

(II) ncRNA(cis and trans)

An example on Type one: tmRNAs

transfer–messenger RNA (tmRNA): tmRNA, previously named 10S RNA, is also known as a long ncRNA (LncRNA) (260 to 430 nt). The gene *ssrA* encodes for tmRNA. tmRNA has dual structural and functional similarities to both tRNA and mRNA.



- **Structure of tmRNA:** tmRNA contains structures similar to the upper half of tRNA, including a complete acceptor stem with **3' CCA** end, a thymidyl stem-loop domain (TL) and a degenerative dihydrouridine stem-loop (DL), both of them called (**TLD**).
- In place of the anticodon there is a coding sequence for a short peptide called a **tag peptide**.
- Prominents are the tRNA-like domain (TLD), the messenger RNA-like region (MLR), and four pseudoknots (pk1 - pk4). The MLR encodes the tag peptide between resume (re-start) and stop codons.

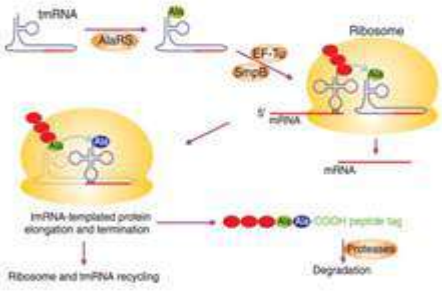
action of tmRNA

Function of tmRNA: It is tagging abnormal proteins in bacteria for degradation.

Trans- translation model by tmRNA:

- . tmRNA is aminoacylated with alanine by alanyl-tRNA synthetase (AlaRS).
- Ala-tmRNA is then taken to stalled ribosomes in a pathway dependent on the proteins SmpB and EF-Tu.
- The " stalled " polypeptide chain is then transferred to the Ala of Ala-tmRNA and protein synthesis resumes, but now using tmRNA as its template.
- tmRNA-templated protein elongation and termination result in the release of a polypeptide with an 11–amino acid COOH-terminal tag.

This tagged polypeptide is subsequently recognized and degraded by COOH-terminal–specific proteases.



Type 2: Regulatory Function of Bacterial ncRNA

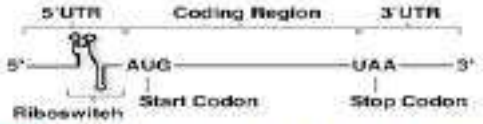
Examples of Regulatory function ncRNAs

(i) riboswitches, which are part of the mRNAs that they regulate Protein-binding sRNAs.

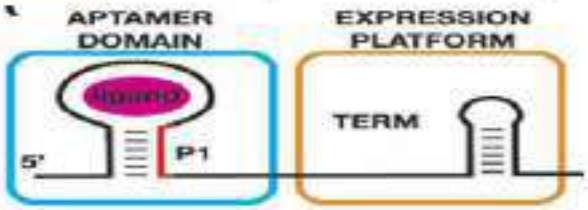
(ii) two classes (cis and trans) that modulate the translation and stability of mRNAs. The first class is encoded in cis on the DNA strand opposite the target RNA, and the second type is encoded elsewhere.

Riboswitch ncRNA as Regulatory molecules

Riboswitches are ncRNA which considered to be cis-encoded elements of RNA.
Location: It locates in the 5'UTR region of a targeted mRNA but less frequent at the 3'end UTR.

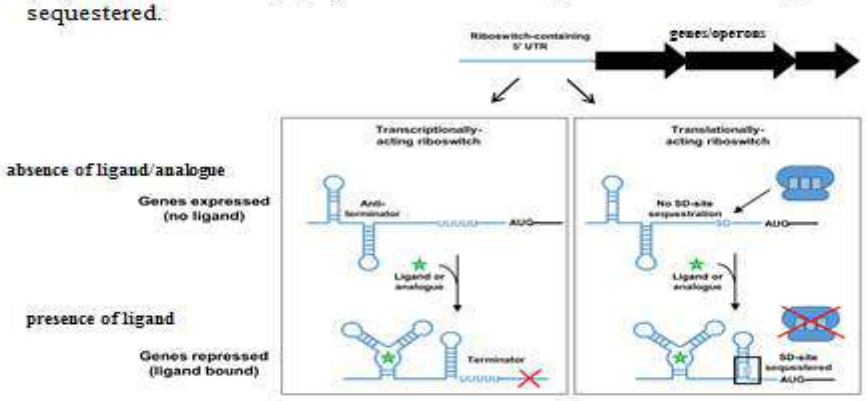


Structure of the riboswitch: It consists of aptamer domain and an expression platform.



Riboswitch action

Riboswitches are found in the 5' UTR of mRNAs (blue sequence), and control the expression of their downstream genes/operons (shown in black). They adopt one of two mutually exclusive configurations: one in the absence of ligand/analogue (shown here as 'gene expressed') and one in the presence of ligand (depicted here as 'gene repressed'). They can influence the transcription (left) or the translation (right) of their associated genes. SD = Shine Dalgarno sequestered.

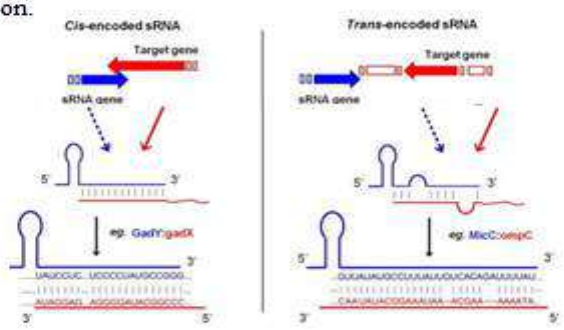


(II) cis-encoded & trans-encoded ncRNAs

- cis-encoded ncRNAs genes are located close to the target mRNA encoding genes .
- trans-encoded ncRNAs genes are located far from their target mRNAs.

Regulation of transcription by Cis & trans-encoded ncRNA:

After transcription, the cis-encoded ncRNAs (blue) make short and perfect base pairing with their targets, while trans-encoded ncRNAs make long and imperfect base pairing region.

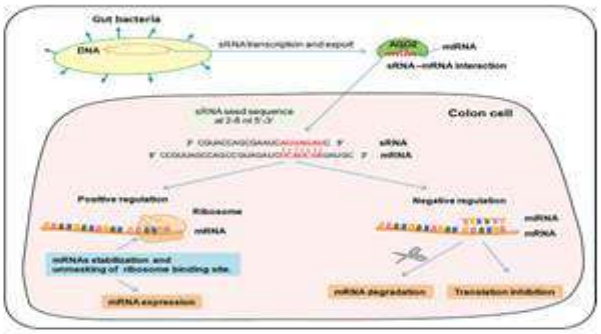


Interaction between bacteria small RNAs (sRNA) and their mRNAs specific target from the colon cell

After transcription of sRNA and export from bacteria, sRNAs enter into the colon cells and based on their seed region complementarities bind to their mRNA targets, producing a positive or negative regulation.

(Left) **Positive regulation**, sRNAs can stabilize mRNAs, unmasking (not binding) the ribosome binding site (RBS), and facilitates gene expression.

(Right) **Negative regulation**, sRNAs leads to destablization and subsequent degradation of mRNA and/or their translational inhibition by binding to ribosome-binding sites.



Epigenetic in Bacteria

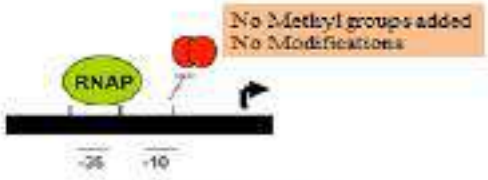
Lec.9

Prof. Dr Hayfa H Hassani
haiphahassani@gmail.com
hayfa.hassani@sc.uobaghdad.edu.iq
2021

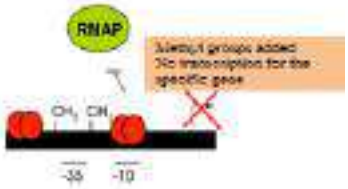
Bacterial Epigenetic Concept

Bacterial Epigenetic is a DNA modification that do not change the DNA sequence can affect gene activity (gene expression). A chemical compound (Methyl group) that added to the gene can regulate their activity; this modification is known as epigenetic change.

Epigenetic changes can help determine whether genes are turned on or off and can influence the production of proteins in certain cells.



gene expression ON

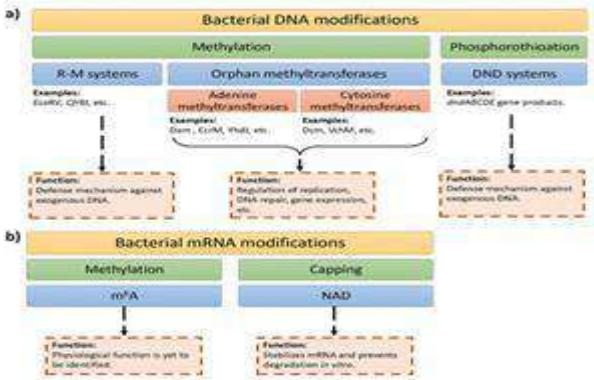


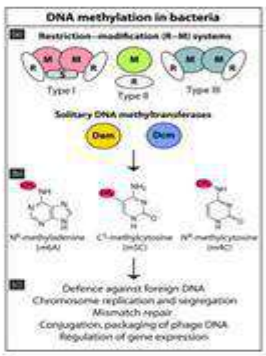
gene expression Off

Epigenetic Systems in Bacterial DNA

Modifications of bacterial DNA is mediated by:

1. **R–M systems**
2. **Solitary (Orphan) DNA methyltransferase**



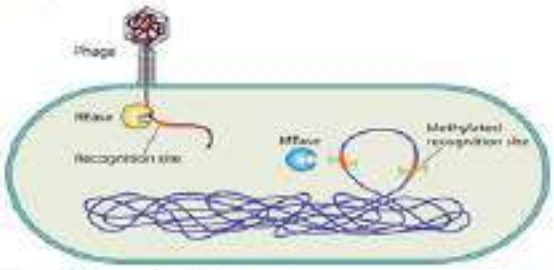


First Mechanism in Bacterial Genome

R-M systems

Restriction-modification (R-M) systems are important components of prokaryotic defense mechanisms against invading genomes.
R-M systems appear as a strategy employed by bacteria to serve as innate immunity.
Composition of R-M Systems: they comprise two contrasting enzymatic activities:

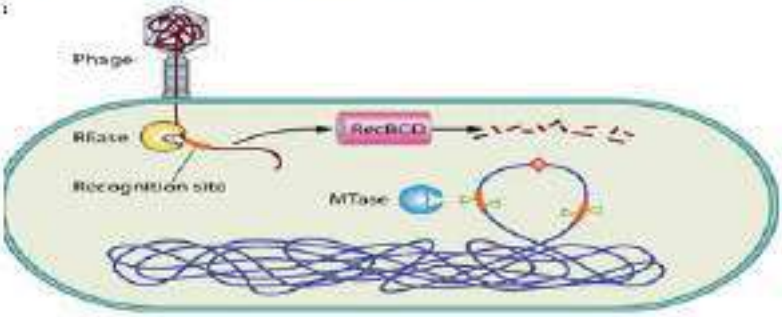
1. Restriction endonuclease (REase)
2. Methyltransferase (MTase)



Mechanism of R-M

R-M systems typically aid resistance to phage infection by two steps:

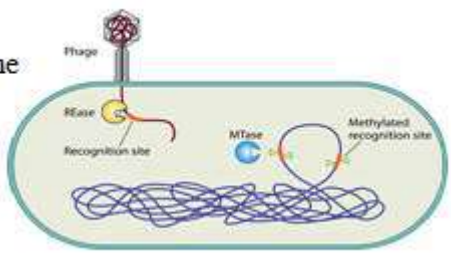
- encoding a methyltransferase (MTase) that methylates a particular sequence of "self"-DNA
- a cognate (associated) restriction endonuclease (REase) that discriminates and destroys non-modified invasive DNA.



Action of REase & MTase

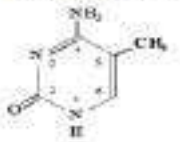
REase: it recognizes and cleaves foreign DNA sequences at specific sites.

MTase: R-M systems discriminate between self and non-self DNA, by transferring methyl groups to the same specific DNA sequence within the host's genome.

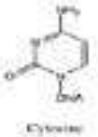


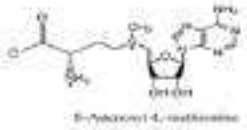
Modified bases by Mtase

MTases : I) transfer the methyl group from S-adenosyl methionine to the **C-5 carbon by Dna methyltransferase (Dcm) .**



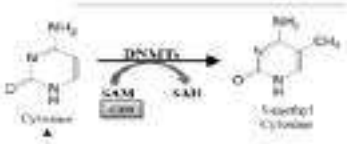
Cytosine and S-Adenosyl-L-methionine (SAM), two extremely important substrates for the DNA Methylation Mechanism





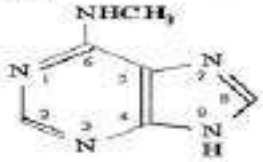
Mechanism of C modification

DNA methylation occurs at the 5′ position of cytosine, within CpG dinucleotides. DNMTs catalyze the transfer of the methyl group to cytosine and generate 5-mC using SAM as methyl donor and producing SAH (Sadenosyl homocysteine).



Modification of other bases such as Adinine (A) or other citation in Cytosine (C)by MTase

II) transfer the methyl group from S-adenosyl methionine to the Adenine by DNA adeninemethyltransfrase (Dam) or to N4 in C by Dcm:



N_6 methyl adenine

N_4 methyl cytosine

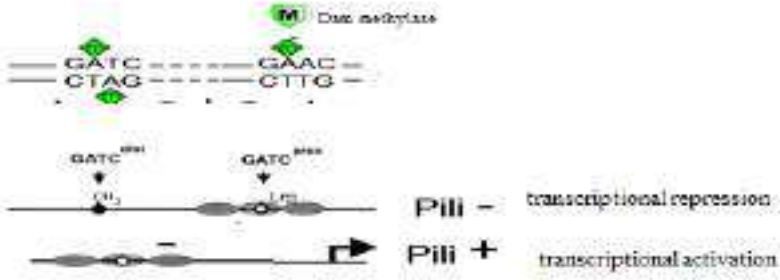
(a) N6 amino group of adenine

(b) N4 amino group of cytosine

Role of DNA methylation in Regulation of Bacterial Genes

Bacteria use DNA adenine methylation (rather than DNA cytosine methylation) as an epigenetic signal. DNA adenine methylation plays roles in the virulence of diverse pathogens of humans and livestock animals, including pathogenic *Escherichia coli, Salmonella, Vibrio, Yersinia, Haemophilus,* and *Brucella.*

e.g. In *E.coli*, adenine methylation at GATC sites by DNA adenine methyltransferase (Dam) provides signals for regulation of gene expression, then transcriptional repression was done.



Types of R-M Systems

There are four types of R-M system within bacterial genomes:

1. **Type I**
2. **Type II**
3. **Type III**
4. **Type IV**

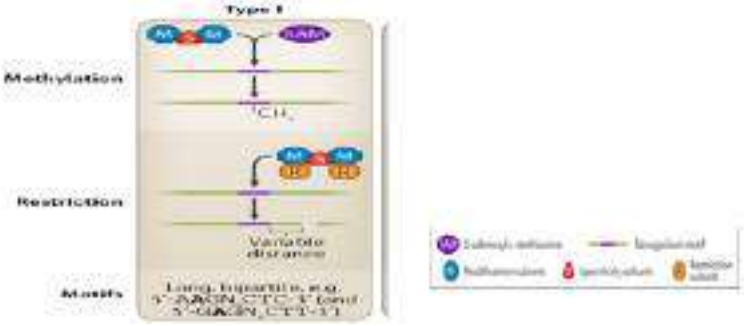
Their classification mainly is based on:

- R-M system subunit composition
- cleavage position
- cofactor requirements
- substrate specificity

Type I R-M system

In type I R-M system,

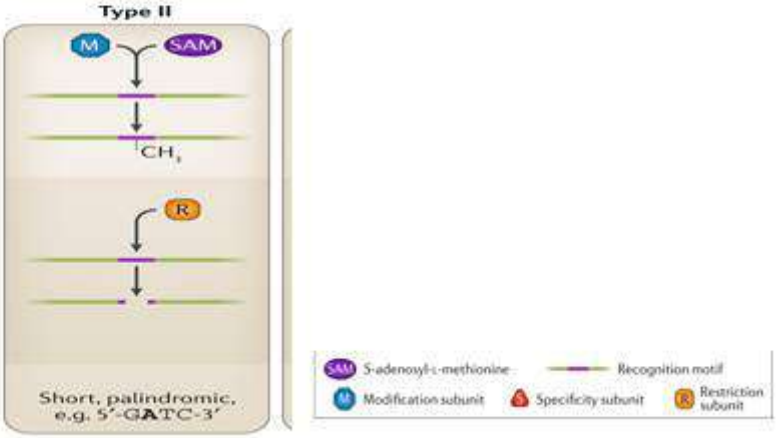
- DNA methylation of both DNA strands occurs at an asymmetrical DNA sequence by a complex of two modification/methyltransferase (Mod, or M) subunits and one specificity (S) subunit, (M_2S_1).
- DNA restriction requires two restriction endonuclease (Res, or R) subunits per complex ($R_2M_2S_2$).
- With complexes ($M_2S_1R_2$) mediating DNA cleavage at a location far from the recognition sites.



Type II R-M system

In type II R-M system,

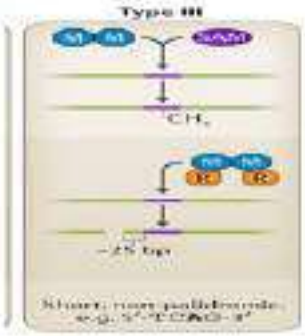
- a single Mod (M) subunit mediates methylation on both strands of the DNA at a palindromic sequence,
- DNA restriction is mediated by one or two Res (R_1 or R_2) subunits in or near the recognition sequence.



Type III R-M System

In type III R–M system,

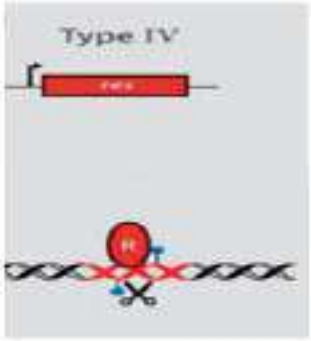
- two Mod (M_2) subunits methylate DNA on one strand at a specific 5–6 bp asymmetrical recognition site,
- $M_2 R_2$ complexes mediate DNA cleavage at a fixed distance (25–27 bp) from the recognition site.



Type IV R-M Systems

In Type IV R-M system

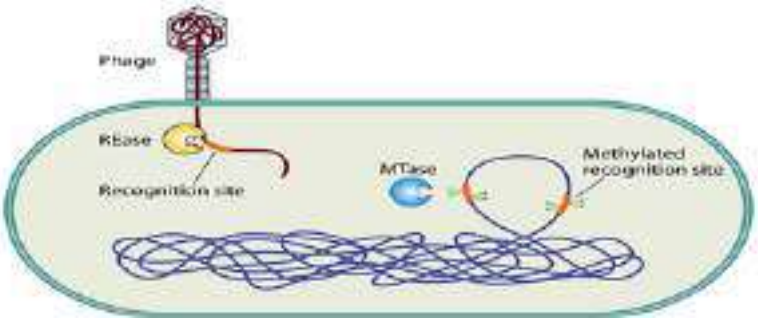
- contains only endonuclease (res) that recognizes methylated DNA.
- The Cleavage occurs within or away from the recognition sequences.



Restriction-modification (R-M) systems as innate immune

Restriction–modification (R-M) systems are often regarded as bacteria's innate immune systems, protecting cells from infection by mobile genetic elements (MGEs) or other foreign DNAs.

R-M systems recognize the methylation status of incoming foreign DNA, e.g., phage genomes. Methylated sequences ( ) are recognized as self, while recognition sequences ( ) on the incoming DNA lacking methylation are recognized as non-self and are cleaved by the restriction endonuclease (REase). The methylation status at the genomic recognition sites is maintained by methyltransferase (MTase) of the R-M system.



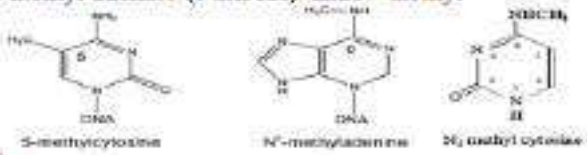
Second Modification System in Bacterial Genome

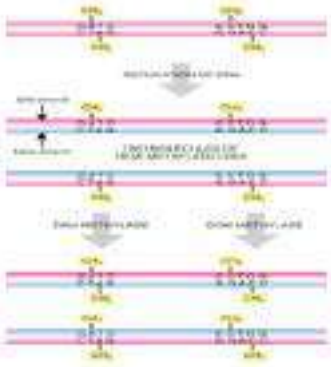
Solitary (Orphan)DNA methyltransferase

Solitary (Orphan) DNA methyltransferase is a modification of DNA which is not part of R-M systems.

Modified Bases by Solitary DNA Methyltransferase

5-methylcytosine (5-meCyt), 6-methyl-adenine (6-meAde) and N^4-methyl-cytosine (m4C) residues.





Methylated sites in the bacterial chromosome

GATC, CCWGG

- About 1.5% of all adenines in the sequence (GATC)
- About 0.75% of all cytosine in sequence (CCWGG)

are methylated

Enzymes responsible for methylaed Bases:

- Dam methylase for methylation adenine
- Dcm methylase for methylation cytosine

DNA methylation in bacteria can mediate by solitary DNA methyltransferases

Dam Dcm

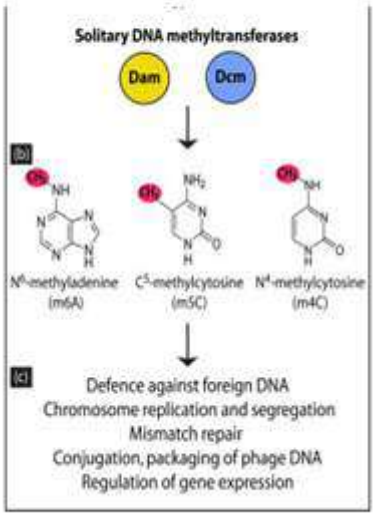
Dam: DNA adenine methyltransferase

Dcm : DNA cytosine methyltransferase

Dam & Dcm are DNA methyltransferases that do not have a restriction enzyme counterpart.

(b) Dam & Dcm generate N^6-methyl-adenine (m6A), C^5-methylcytosine (m5C), and N^4-methyl-cytosine (m4C) residues.

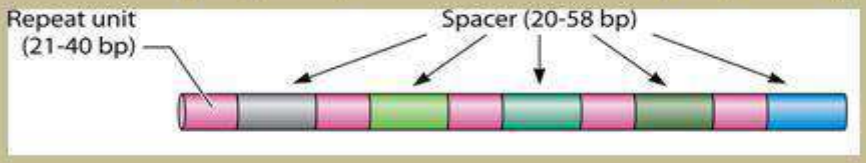
(C) DNA methylation is involved in numerous bacterial cell processes.



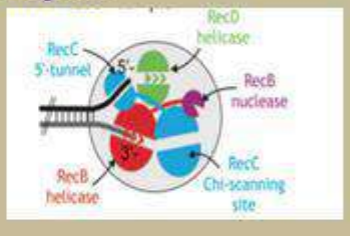
Other Strategies of Bacterial Defense that are not related to Epigenetic

Non-Epigenetic and Non- R-M systems but Considered as Defense System in Bacteria

1. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)



2. RecBCD Enzyme



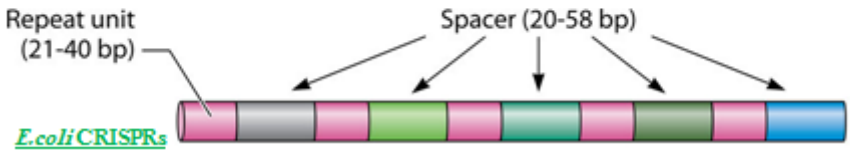
1. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)

CRISPRs found in many eubacteria and archaea.

- CRISPRs are gene loci involved invasive DNA elements (phages or plasmids) in the host cells.
- These loci serve as memory for the bacteria with respect to earlier phage encounters.

Structure of CRISPRs

- **Repeats** consist of between 20 and 49 bp long and highly conserved short DNA repeat sequences which are at least partially palindromic.
- **Spacers** :stretch of variable sequences of between 24 and 75 bp which interspersed repeats. The spacer sequences usually originate from (1) fragments of captured foreign DNA; (2) coding or non-coding DNA; or (3) DNA derivative of RNA from various phages and transposons.

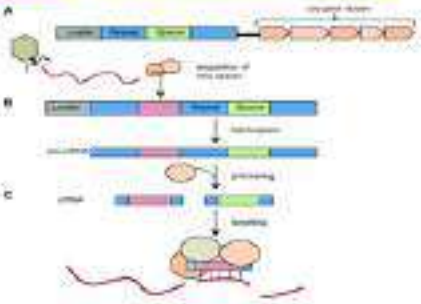


Mechanism of CRISPRs as Adaptive Immunity

CRISPRs use small non-coding RNAs called (crRNA) for defense and function in conjunction with **CRISPR associated (Cas) proteins are involved in adaptive immunity against invading foreign genomes (phages or Plasmids) in the host cells.**

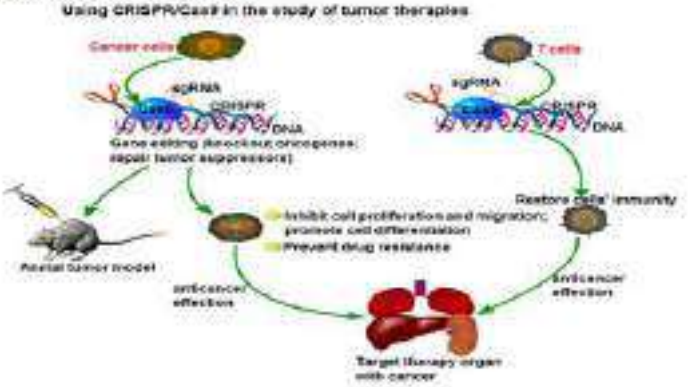
Mechanism of (CRISPR)/CRISPR associated (Cas) adaptive immunity

- **(A) Adaptation:** Upon infection with a foreign element (e.g., phages or plasmids), a part of the foreign genome is typically incorporated into the leader end of the CRISPR array and the repeat is duplicated. The CRISPR arrays are located adjacent to a cluster of *cas* genes.
- **(B) crRNA generation:** The CRISPRs are transcribed into pre-crRNAs that are then processed into mature crRNAs.
- **(C) Interference:** The crRNA, in a complex with Cas proteins, binds and degrades the target nucleic acid of the invading element.



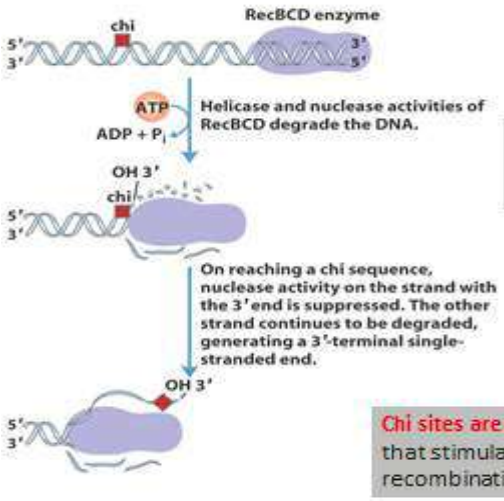
Application of CRISPR system in cancer therapy

CRIPRs system possibly cure the disorders in human genome which caused cancer by switching off the oncogenes or by switching on the tumor suppressor genes.



2. RecBCD Enzyme

RecBCD enzyme is a large (330-kDa) complex with both helicase (DNA-unwinding) and nuclease (DNA hydrolysis) activities.

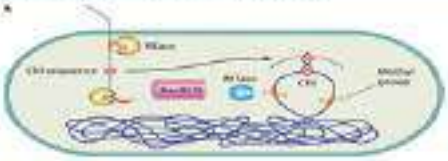


RecBCD Machanism
producing ssDNA during recombination of double-strand DNA breaks.

Chi sites are octameric nucleotide sequences in DNA that stimulate the Rec-BCD pathway of homologous recombination in bacteria.

1) Variation of bacteria by RecBCD

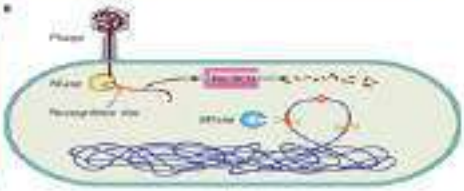
Restriction of incoming DNA from a closely related bacterium (harboring similar Chi sequences) generates DNA fragment which can be utilized as substrates for homologous recombination by the RecBCD pathway.



2) RecBCD as defence System

RecBCD distinguishes the host genome from the phage DNA by Chi sequence, which is absent in phages but present at high frequencies in bacterial genomes. .

Here, the fragments generated by the restriction of phage DNA (lacking the Chi sequence) are recognized as non-self and subjected to further degradation by the RecBCD pathway.



Detection of Methylation in Bacteria

Techniques for Detection of Methylation

Methods can be separated into two categories: those that are used for:

(1)Determination of unknown epigenetic changes

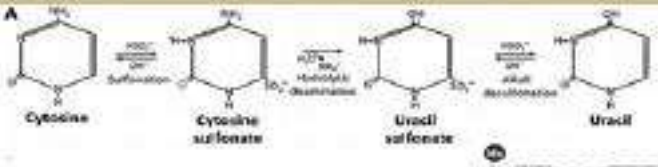
(2)Assessment of DNA methylation within particular regulatory regions/genes of interest

Techniques for Detection of Methylated DNA in Bacteria

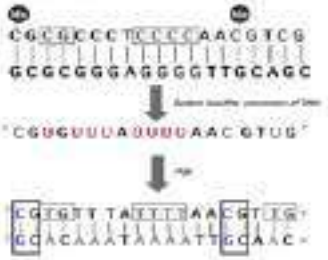
1. **Bisulfite conversion and pyrosequencing**
2. **Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS)**
3. **Enzyme-linked immunosorbent assay (ELISA)**
4. **Restriction fragment length polymorphism (RFLP)Amplification fragment length polymorphism (AFLP),**

❑ Bisulfite conversion and pyrosequencing

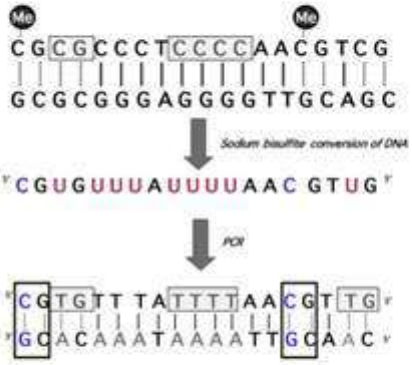
A) The procedure is based on the chemical reaction of single-stranded DNA with sodium bisulfite (HSO_3^-) result in conversion of un-methylated C to U.
The reaction step is as follows: **sulfonation** at the carbon-6 position of cytosine, **hydrolytic deamination** at the carbon-4 position to generate uracil sulfonate, and finally, **desulfonation** under alkaline conditions to generate uracil.



B) The sodium bisulfite treatment converts unmethylated cytosines of the original DNA sequence to uracil, whereas methylated cytosines remain as cytosine. Then, PCR was done.



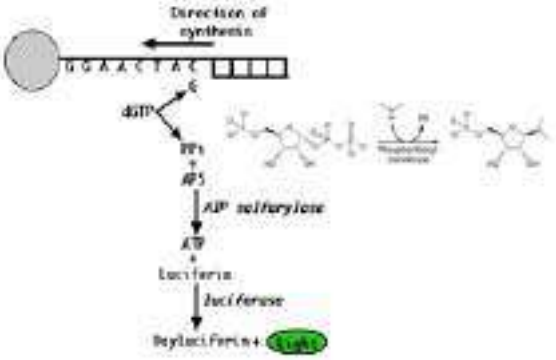
C) In order to detect which cytosines were protected after bisulfite conversion, the DNA is amplified by PCR. PCR will convert the Us to Ts but faithfully replicate the previously methylated cytosines, although they lose the methylation during this process.

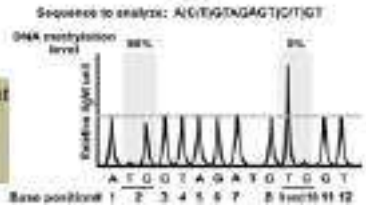


D) PCR Products directly sequenced via **Pyrosequencing**.
The "pyro-" in *pyrosequencing* refers to the release of pyrophosphates (PPi) following the incorporation of each nucleotide.

The release of PPi occurs in an equimolar to the number of nucleotides incorporated and is therefore highly quantifiable.

E) The addition of **ATP sulfurase** converts the PP_i to ATP, which then activates a **luciferase reaction** and produces **light** in proportion to the amount of PP_i generated.

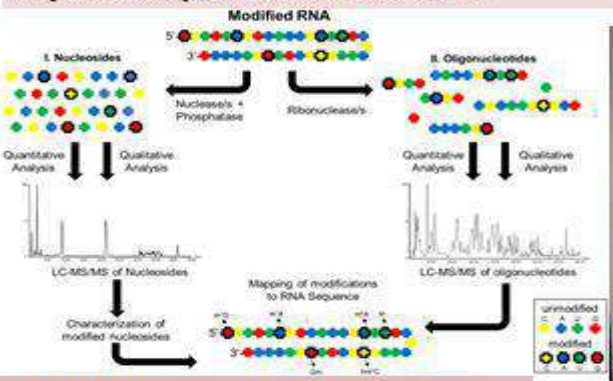




F) The amount of light generated provides a precise read-out of which nucleotide is incorporated and then shown as a pyrogram. The degree of methylation is analyzed as a[C/T] using software [C%=C/(C/T)]

❑ **Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) for detecting levels of methylation ranging from 0.05%–10% in sample.**

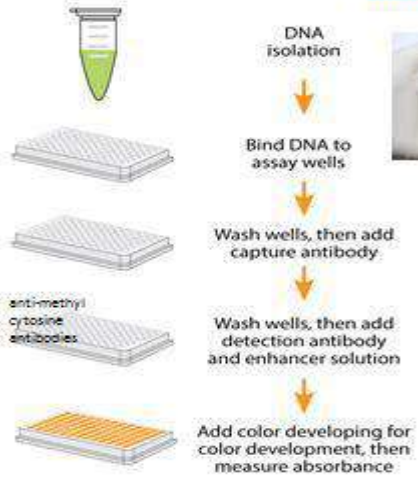
Total hydrolysis of RNA leads to a mixture of both modified and unmodified nucleosides. The modified RNA is subjected to **(I)** nucleosides **(II)** and oligonucleotide analyses. In **(II)** analysis, the RNA is digested with ribonucleases resulting in oligonucleotides of varied length. Their nucleotide sequences are determined by different type of LC-MS/MS analysis to identify the location of modification. .

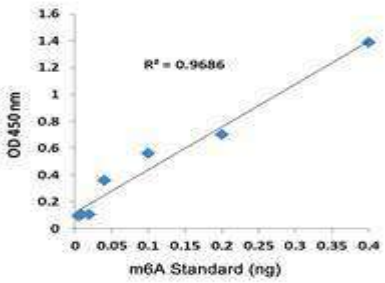


The four colors represent four bases. The bold outline denotes the existence of modification.

LC-MS/MS identifies the resident modifications

❑**Enzyme-linked immunosorbent assay (ELISA) used for the identification of large changes in global DNA methylation.**





❑ Restriction fragment length polymorphism (RFLP), amplification fragment length polymorphism (AFLP) used for detection of fragments that are differentially methylated.

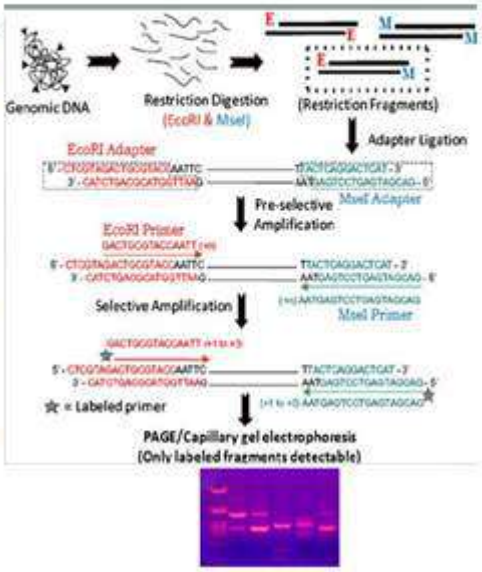
(1) Template fragments are generated by digestion of genomic DNA with a combination of the two restriction enzymes *EcoRI* and *MseI* (red and blue represent ***EcoRI*** and ***MseI*** restriction enzyme sites, respectively).

(2) ligation of the double-stranded *EcoRI*- and *MseI*-specific adapters to the fragment ends.

(3) a pre-selective amplification step using primers that match the adapter sequences and that carry each one selective nucleotide (represented by n*/n*) at their 3 end are used to PCR-amplify subsets of the ***EcoRI/MseI*** template.

(4) Selective PCR-amplification step in which additional selective nucleotides are added to the ***EcoRI*** and ***MseI*** primers;

(5) Electrophoretic size fractionation displays on denaturing polyacrylamide gels of the *EcoRI/MseI* amplification products.



Bacteria Remodel the Host's Epigenome

Crosstalk Between Bacteria and Host Epigenetic Machinery

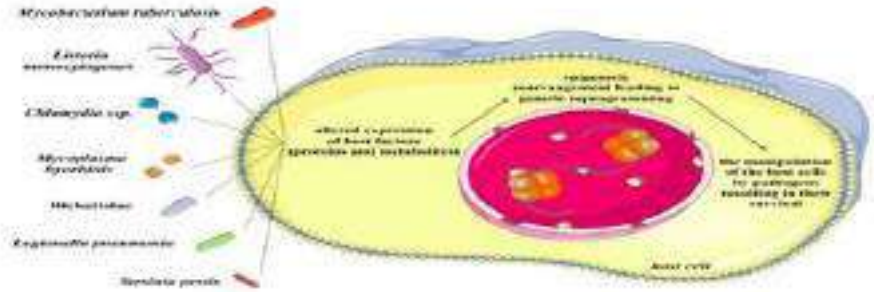
Lec.10

Prof Dr Hayfa H Hassani
haiphahassani@gmil.com

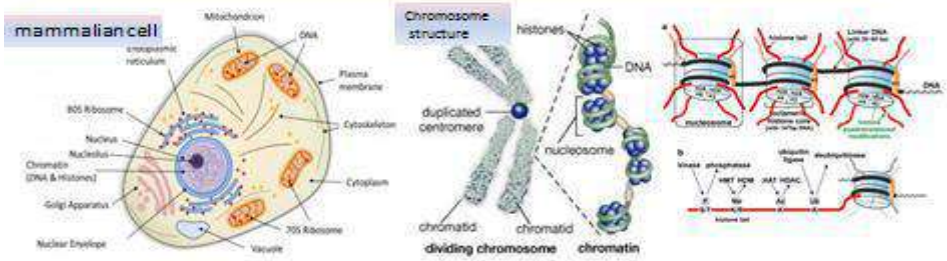
2021

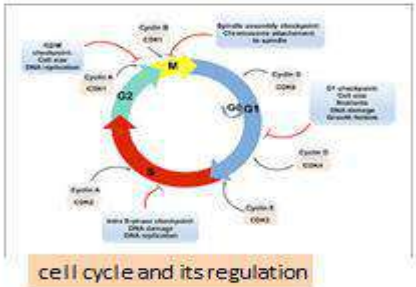
Infections Contribute with Host Epigenetic

Infectious agents (e.g., *Helicobacter pylori*, *Porphyromonas gingivalis*, *Fusobacterium nucleatum*, *Streptococcus bovis*, *Chlamydia pneumoniae*, *Campylobacter rectus*, Epstein–Barr virus (EBV), hepatitis viruses, human papilloma virus (HPV) etc.) have developed a wide variety of epigenetic regulatory mechanisms, through which they are able to effectively use the epigenome of the host for their own benefits, and then they are frequently associated with carcinogenesis.

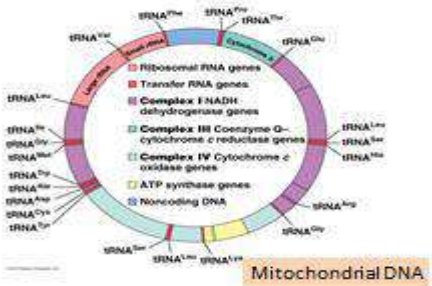


Human Genome and Cell Cycle





cell cycle and its regulation

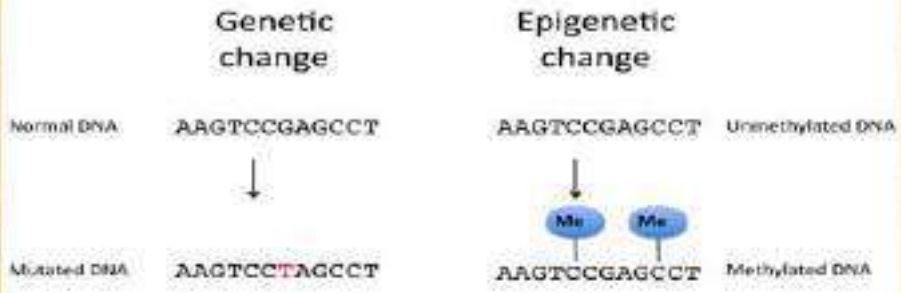


Mitochondrial DNA

Mammalian Epigenetic

Epigenetic Concept

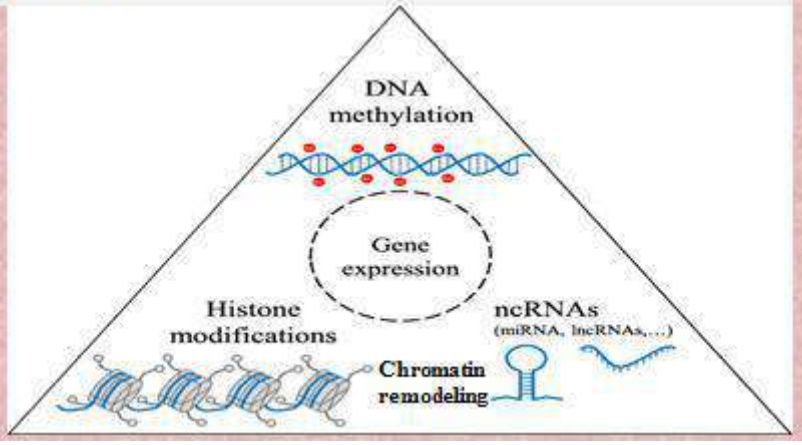
Epigenetic defines as changes in a chromosome without alterations in the DNA sequence" can be triggered by changes in the environment of the cell.



Mammalian Epigenetic

Mechanisms of Epigenetic in mammals:

- DNA methylation
- Histone modifications
- Chromatin remodeling
- ncRNA



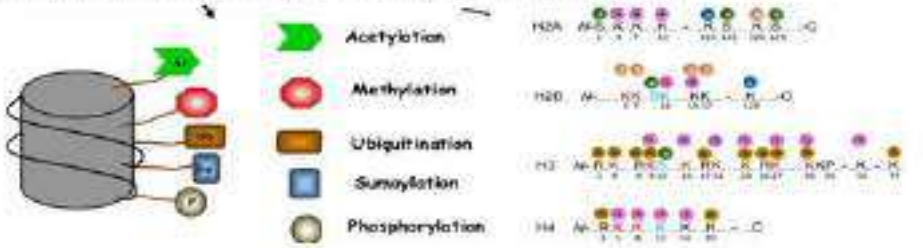
1. DNA Methylation

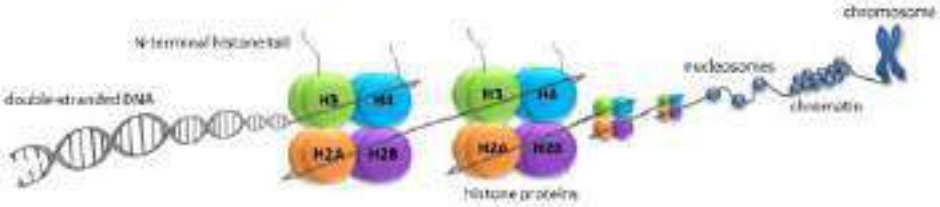
DNA methylation is the a process in which a methyl group is added to the 5th carbon of a cytosine resulting in 5-methylcytosine (5mC). This methyl group is donated from S-adenosyl methionine (SAM), and this reaction is catalyzed by DNMT enzymes.

Methylation process usually occurs in the CpG dinucleotide context.

2. Histone Modifications

The main targets of modifications are the amino acids located in the N-terminal portion of histone tails that shape the nucleosome.
Lysine (K), arginine (R), serine (S), threonine (T), and tyrosine (Y) are modified.





Modification of Histones

Modifications are methylation, acetylation, phosphorylation, ubiquitination (is a process by which by which the carboxyl-terminus of the ubiquitin (polypeptide) molecule binds to the lysine residue in the tails of H2A, H2B) this process controls gene transcription and DNA damage repair), and sumoylation (is a process by which the SUMI (a small ubiquitin-like modifier is a member of the ubiquitin-like protein family) covalentlybinding to the lysine residue of histones ; SUMOylation of histones can regulate gene transcription and affect the occurrence and development of certain diseases.

ubiquitination

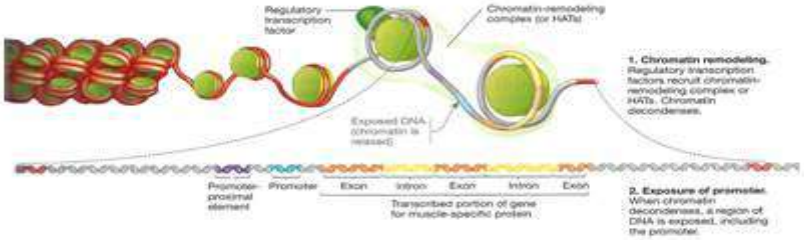
sumyolation

3. Chromatin remodeling

Chromatin remodeling is an important mechanism of regulating eukaryotic gene expression, which makes tightly condensed DNA accessible to various regulatory factors, such as transcription factors and components of DNA replication.

The basic mechanism of chromatin remodeling depends on the three dynamic properties of nucleosomes: reconstruction, enzyme-induced covalent modification, and repositioning.

Chromatin remodeling exposes the promoter



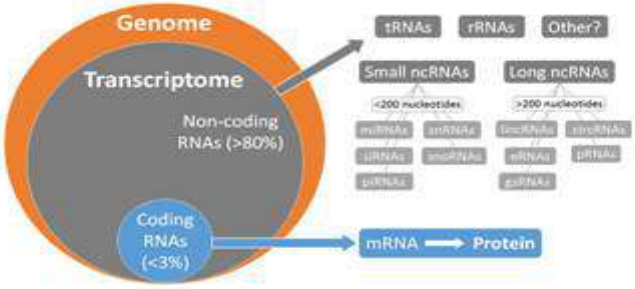
4. Non-coding RNAs (ncRNA)

ncRNAs provide a unique level of post-transcriptional gene regulation that modulates a range of fundamental cellular processes.

Types of ncRNAs

ncRNA transcripts are divided into two major groups based on their length, localization, and/or function:

- long non-coding RNAs (lncRNAs).
- Short ncRNA (shncRNA) or Small ncRNA (snRNA)



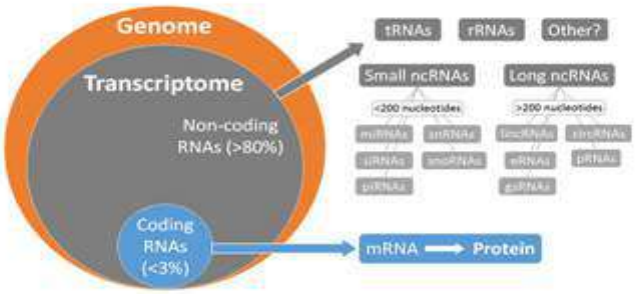
4. Non-coding RNAs (ncRNA)

ncRNAs provide a unique level of post-transcriptional gene regulation that modulates a range of fundamental cellular processes.

Types of ncRNAs

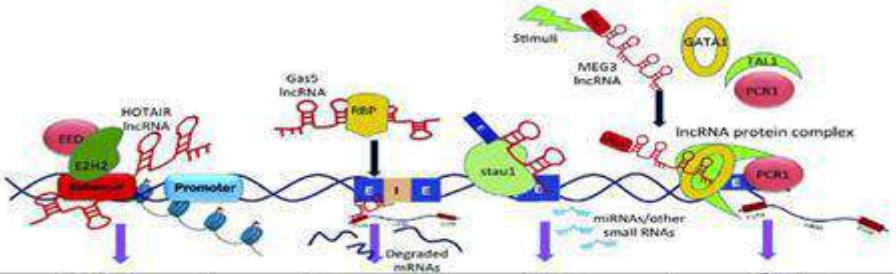
ncRNA transcripts are divided into two major groups based on their length, localization, and/or function:

- long non-coding RNAs (lncRNAs).
- Short ncRNA (shncRNA) or Small ncRNA (snRNA)



Actions of lncRNA

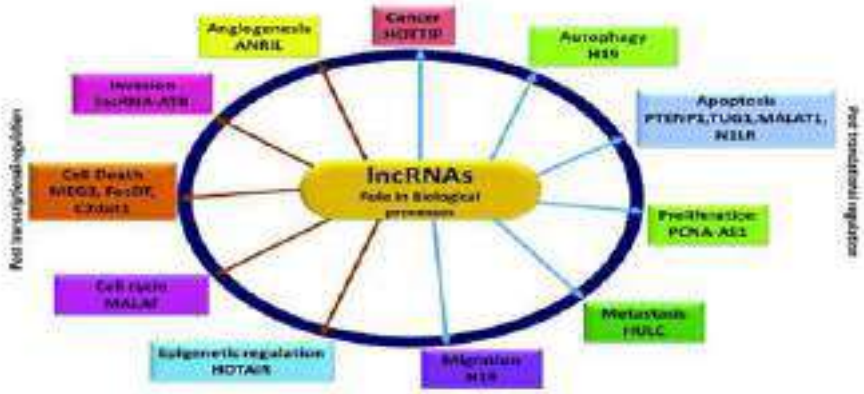
- LncRNAs are RNA transcripts with more than 200 nt long that can not be translated into a protein.
- Many lncRNAs have 5' cap and exons and about 60% of lncRNAs have polyA tails.



Scaffold	Decoy	Guide	Signal
Chromatin modification and epigenetic modification also recruitment of chromatin remodeler. e.g. HOTAIR, ANRIL, and PARTICLE etc.	Transcriptional inhibition / mRNA degradation and RNA-RNA interaction, transcriptional and translation inhibition. e.g. HOTTIP, GAS5, and KCNQ1OT1 etc.	Precursors of miRNAs and other small and medium RNAs or RNA Processing such as splicing, editing, turnover and miRNA sequestration. e.g. MALAT1, GAS5, XIST, FENDRR, and lncHSC1-2 etc.	Transcription/Translation activation for gene activation, modulation of miRNA regulation e.g. NRON, BAND, DLEU2, PACER, and lincRNA COX2

Functions of lncRNA

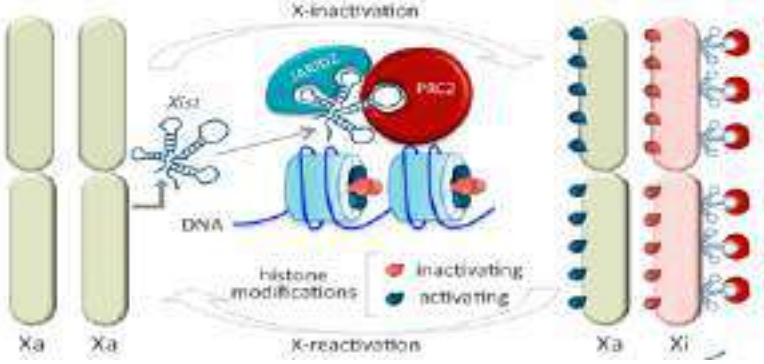
LncRNAs regulate various cellular processes by post-transcriptional and post-translational modifications.



Role of LncRNA in Epigenetic (X-inactivation)

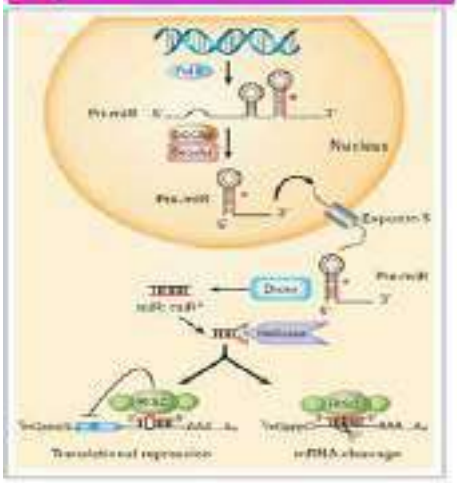
The X-chromosome state in females is linked to cell differentiation. Pluripotent cells have two active X-chromosomes (XaXa) and undergo X-chromosome inactivation when differentiated, resulting in one active and one inactive X-chromosome (XaXi).

The lncRNA *Xist* is the main player in such process. The espression of the Xist gene on one of the X-chromosomes leads to the recruitment of the PRC2 complex to that chromosome, establishing its inactivation.

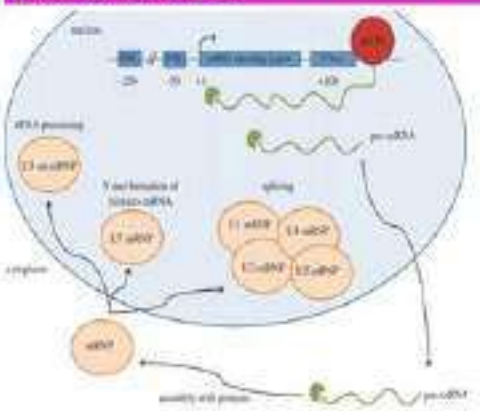


Small ncRNAs (snRNA)
short RNAs less than 300 nts

Biogenesis and Function of snRNA



snRNAs associate with proteins (U1, U2, U4, U5 and U6) called spliceosome to form a small nuclear ribonucleoprotein particles (snRNPs) that contribute with splicing introns from genes before transcription that occur.



Role of snRNA in Splicing of mammalian genes

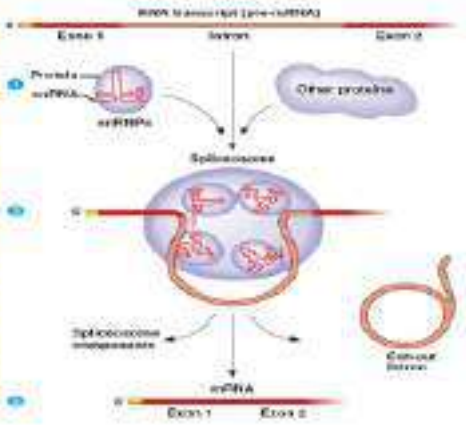
The spliceosome includes the U1, U2, U4, U5 and U6 snRNPs , will recognize the canonical GT/AG splice sites flanking introns to assemble with the pre-mRNA

Pre-mRNA Splicing:

1. Small nuclear RNAs (snRNAs) form a complex called a spliceosome with small nuclear ribonucleoproteins (snRNPs) and other proteins.

2. The snRNAs bind to specific nucleotides in the introns of a pre-mRNA.

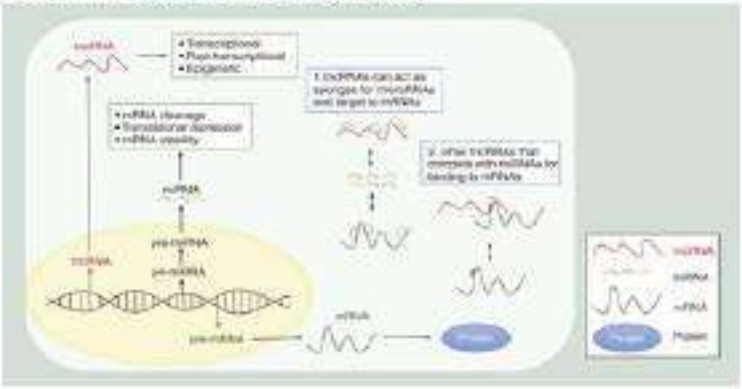
3. The RNA transcript is cut, releasing the introns and splicing the exons together, producing mature mRNA.



Functions of LncRNA and miRNA in mammalian Cells

LncRNAs can be activated or inhibited at different levels: transcriptional, post-transcriptional, and epigenetic.

miRNAs interact with target genes though degrading or inhibiting their mRNAs, repressing gene translation, and stabilizing mRNA.

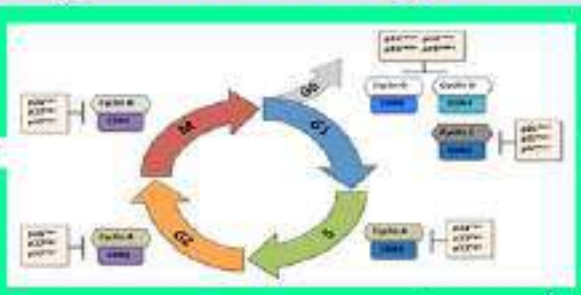


Story Between Bacteria and Our Genome

War and defense

Cell Cycle Progression and Repression

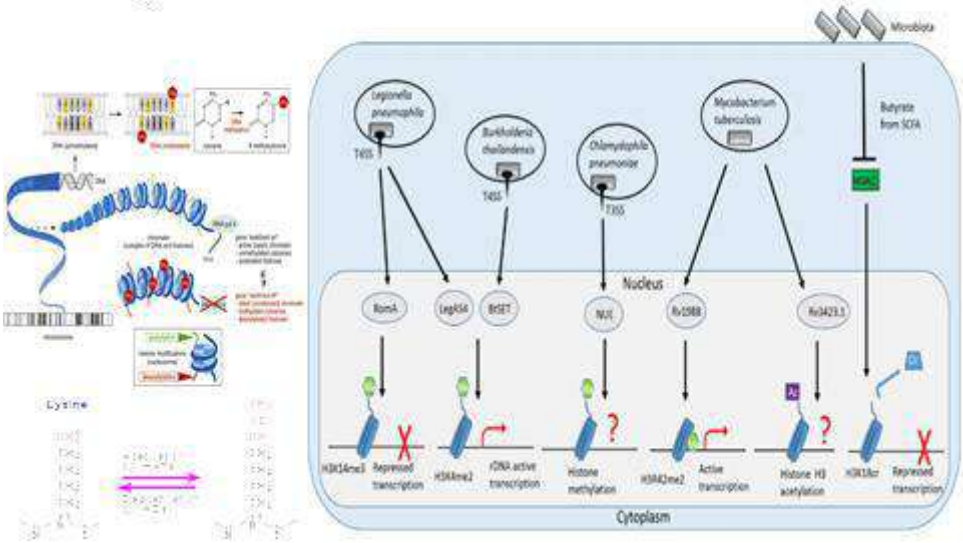
Cell Cycle in Human



prokaryotic cell cycle

Microbial Strategies Remodeling Host Epigenetic

The most common mechanisms by which epigenetics control changes in gene expression involve histone acetylation, histone deacetylation, histone methylation and DNA methylation.



Infection remodels ncRNA expression

MicroRNAs (miRNAs) have emerged as a class of regulatory RNAs in host–pathogen interactions.

Aberrant miRNA expression seems to play a central role in the pathology of several respiratory viruses, promoting development and progression of infection.

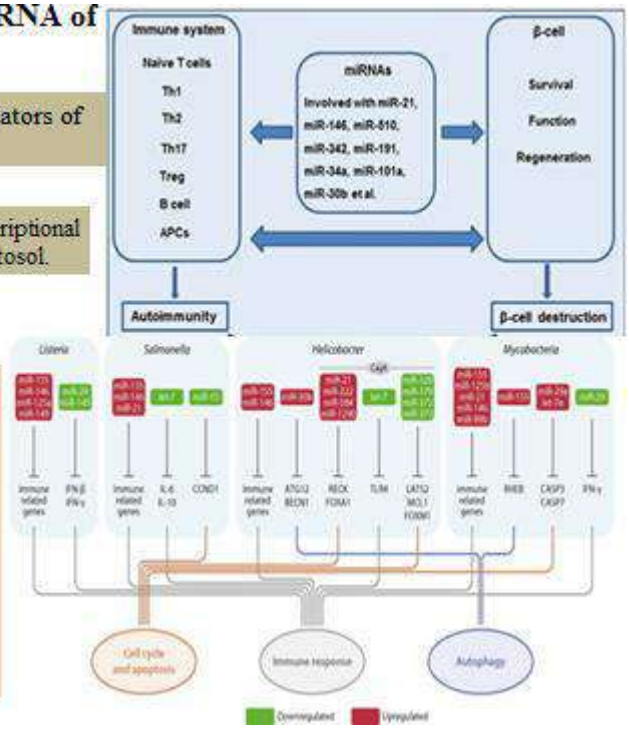
miRNAs may serve as therapeutic and prognostic factors for respiratory viral infectious disease.

Bacteria remodeling ncRNA of the Host

miRNAs are important regulators of immune responses.

miRNAs function in posttranscriptional repression of mRNAs in the cytosol.

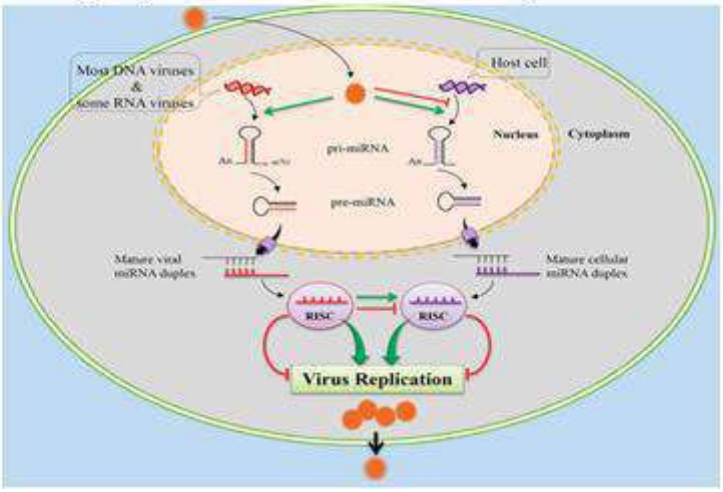
several of miRNAs are induced in response to pathogenic bacteria, such as H. pylori, Salmonella typhimurium, L. monocytogenes, and Mycobacterium bovis, and microbiota that effect, crell cycle, immune response and apoptosis.



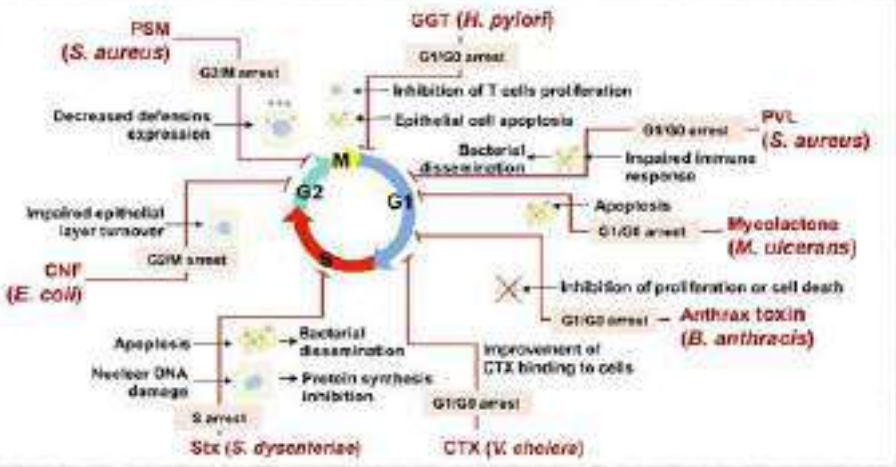
Infections remodeling ncRNA expression

Host cellular miRNAs may have direct effects on viral replication, through positive or negative interactions with viral genomes or other viral factors.

On another hand, miRNAs are utilized by viruses to invade host cells, replicate in host cell, evade host immune response, and establish and maintain virus latency



Smart Weapons That Allow Bacteria to Hijack the Eukaryotic Cell Cycle and Promote Infections



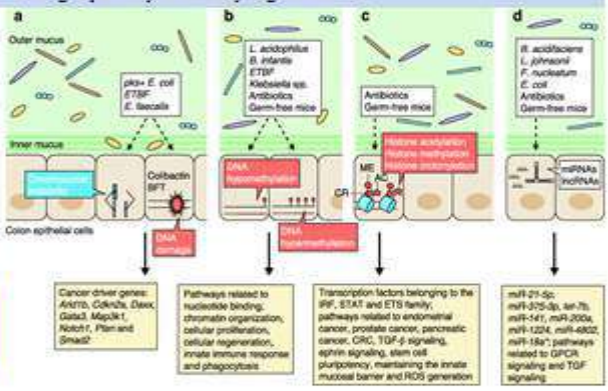
Effect of gut microbiome on colon epithelial cell genome and epigenome.

(a) Enterotoxigenic *Bacteroides fragilis* (BFT) & pks + *E.coli* (colibactin) cause DNA damage in colon epithelial cells (CECs), and *Enterococcus faecalis*, through impact on macrophages, induces chromosomal instability and tumor-inducing DNA mutations in cancer driver genes.

(b) Antibiotics, germ-free mice, and specific microbes (*Bifidobacterium infantis, Lactobacillus acidophilus, Klebsiella* species, and ETBF) show that gut microbes induce both the hypermethylation and the hypomethylation of genes belonging to pathways that are dysregulated in colorectal cancer (CRC).

(c) Antibiotics and germ-free mice show that gut microbes cause changes in the accessibility of transcription factor binding sites, in histone modifications, and in the location of those modified histones. These modifications often affect the promoter and enhancer regions of genes that belong to pathways that are dysregulated in CRC.

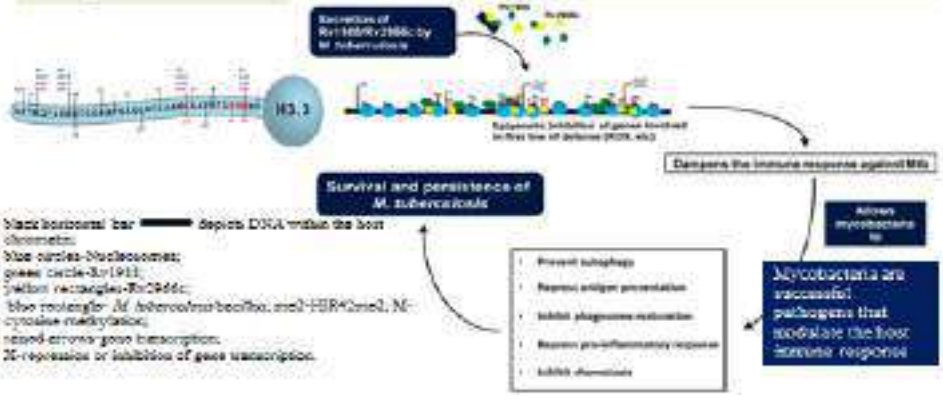
(d) Antibiotics, germ-free mice, and specific microbes (*Bacteroides acidifaciens, Lactobacillus johnsonii, and Fusobacterium nucleatum* show that gut microbes alter the expression of onco-miRNAs and anti-oncomiRNAs in CECs. They also alter the expression of long non-coding RNAs (lncRNAs) that are involved in transforming growth factor (TGF) signaling.



Abbreviations: ETS e26 transformation-specific, IRF interferon regulatory factor, miRNA microRNA, ROS reactive oxygen species, STAT signal transducer and activator of transcription

M. tuberculosis remodeling histone modification of host gene

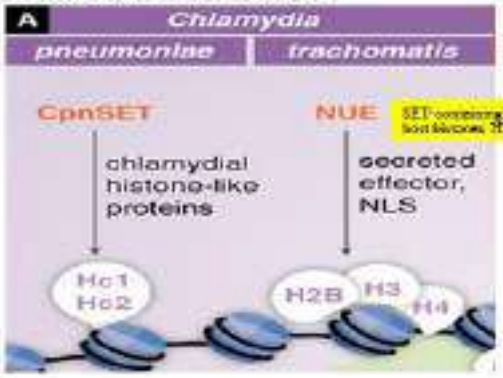
It is a causative agent of tuberculosis (TB) in humans. The mycobacterial protein Rv1988 (methyltransferase) responsible for dimethylation (me2) of arginine (R) present specifically at the 42 position within the core region of histone H3 (H3R42 me2) which involved in first line of defense (e.g. ROS etc.) activity in the host cell. This modification alters the expression of certain host genes, which benefits bacteria and supports the development of infection. Dampening (weak) of the initial host defense could allow mycobacteria to utilize additional multiple factors to ensure its continued survival and persistence in the host cell.



Chlamydia spp. Modulates Host Epigenome

C. pneumoniae **CpnSET methylates chlamydial histone-like proteins Hc1 and Hc2**

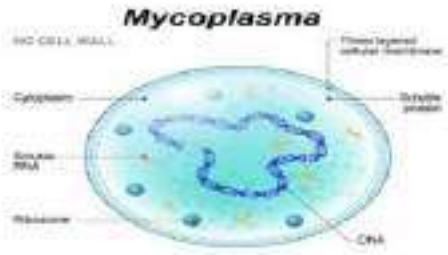
C. trachomatis **containing SET-containing protein called NUE** has been shown to be secreted into the host cell during chlamydial infection, where it enters the nucleus and binds to host chromatin. NUE as a methyltransferase, is able to catalyze methylation of host histones H2B, H3 and H4.



Mycoplasmas

Mycoplasmas are the smallest self-replicating Gram-negative bacteria, which lack the genes coding for the cell wall.

Instead of a cell wall, they possess a three-layered membrane, containing sterol, which is taken up from the environment.



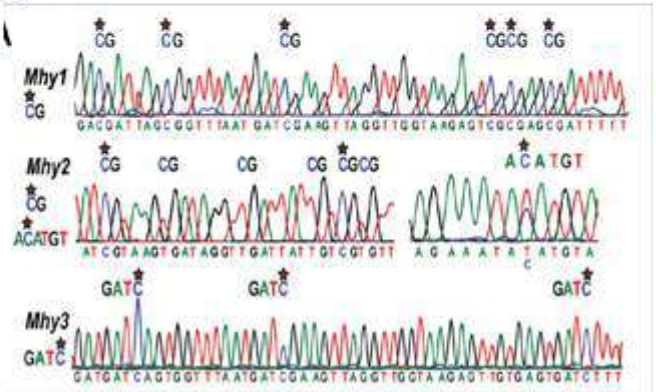
In humans, mycoplasmas are present frequently at mucosal surfaces of respiratory, urogenital tracts, mammary glands and joints.

Mycoplasmas Remodeling Host Epigenetic

Mycoplasmas have been found to produce DNA methyltransferases responsible for the conversion of cytosine to 5mC in the context of CG-dinucleotides.

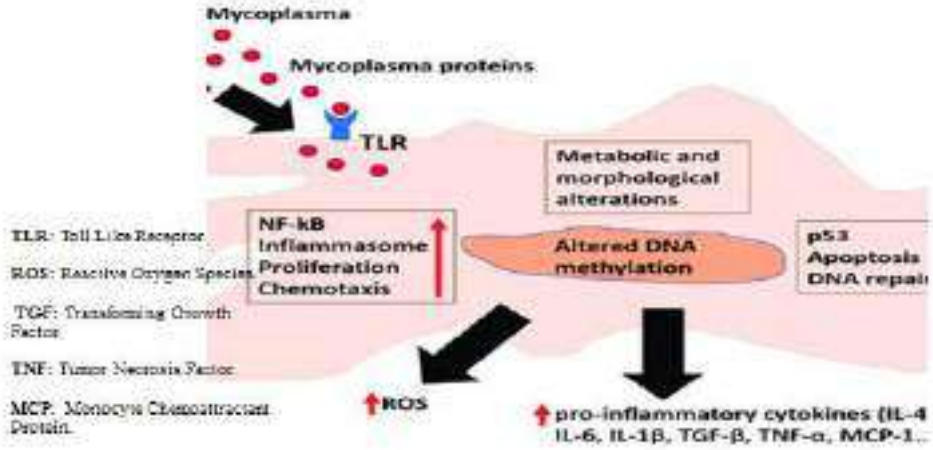
Three DNA methyltransferases have been identified in *Mycoplasma hyorhinis*:

- Mhy1 } promote CG methylation
- Mhy2 }
- Mhy3 acts on GATC sites



Mycoplasmas affect cellular pathways

Mycoplasmas' proteins interact with TLR, where they can alter several pathways responsible for inflammation and DNA repair. In addition, affecting methylation of cellular DNA results in alteration of cellular epigenetic landscape.



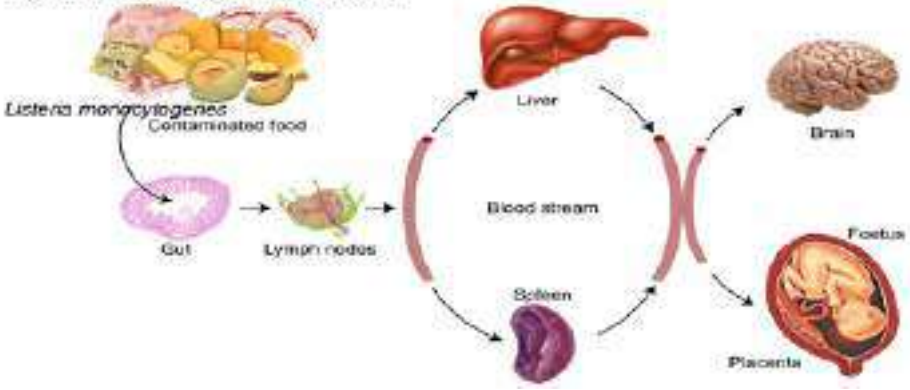
Mycoplasma and cancer Diseases

Due to Mycoplasma frequent persistence as long-term asymptomatic infections they are likely to induce reprogramming of somatic cells and oncogenic cell transformation, resulting in dysregulation of cancer-specific genes.

After translocation of Mycoplasmas to the cell nucleus, the enzymes methylated the host genome at the DNA sequence sites free from pre-existing endogenous methylation, including those in a variety of cancer-associated genes.

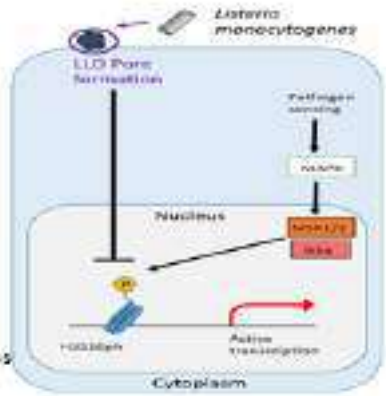
Listeria

Listeria monocytogenes is G+ve facultative pathogen that causes the foodborne disease, listeriosis which is a serious disease for immunocompromised individuals, fetuses, and newborns. This intracellular bacterium is a powerful model to study various aspects of the molecular interactions between pathogen and mammalian cells, especially as it invades different cell types and reaches various organs, such as the liver, spleen, placenta, and brain.



L. monocytogenes Remodeling Epigenetic of the Host

L monocytogenes can induce epigenetic and miRNA modifications in the host to modulate immune defense.
Entry of *L. monocytogenes* into the host cytoplasm depends on escape from the phagosome, a process mediated by the key virulence factor listeriolysin O (LLO) which is a part of pore-forming toxins.



Early in infection by *L monocytogenes*, the LLO induces dephosphorylation of Ser10 on H3 and deacetylation of H4 by secreting LLO. These modifications correlate with transcriptional reprogramming of a subset of host genes, including decreased expression of key immunity factors.

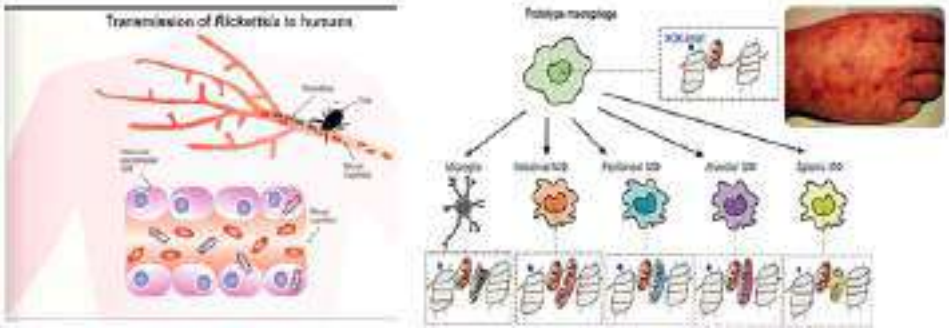
Mitogen-activated protein kinases (MAPKs) govern numerous cellular processes including cell growth, stress response, apoptosis, and differentiation.

Rickettsia

Rickettsia spp. are obligate intracellular Gram-negative bacteria that require a vector for host transmission. Rickettsiae are transmitted to humans via mosquitoes, or by the bite of infected ticks and mites as well as by the feces of infected lice and fleas.

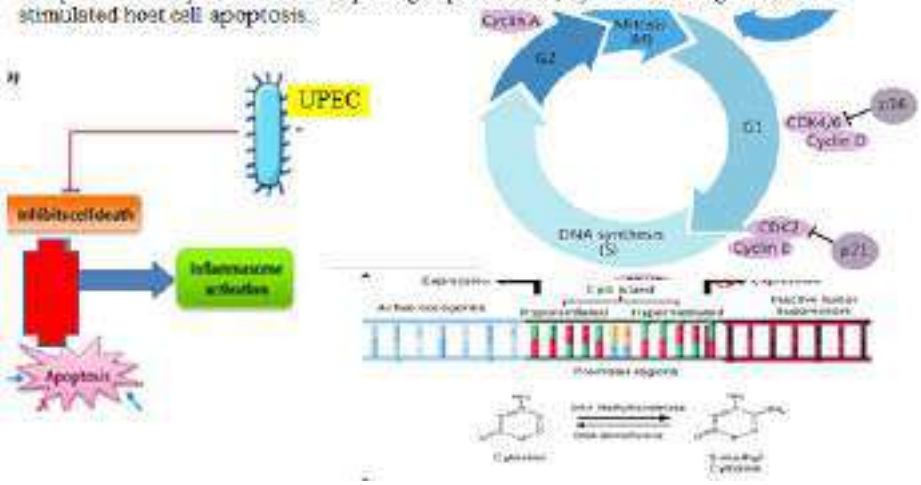
Rickettsia mechanism that allows survival within host
It has several enzymes involved in metabolic pathways suggest that they reprogramming of defense genes through regulation of epigenetic in macrophages.

This model shows the development of different tissue macrophages.
A prototype macrophage has a PU.1 which is a major regulator of the epigenetic landscape of macrophages.
This PU.1 -bound chromatin landscape. Enzymes from Rickettsia will reprogram the epigenetic of macrophages.



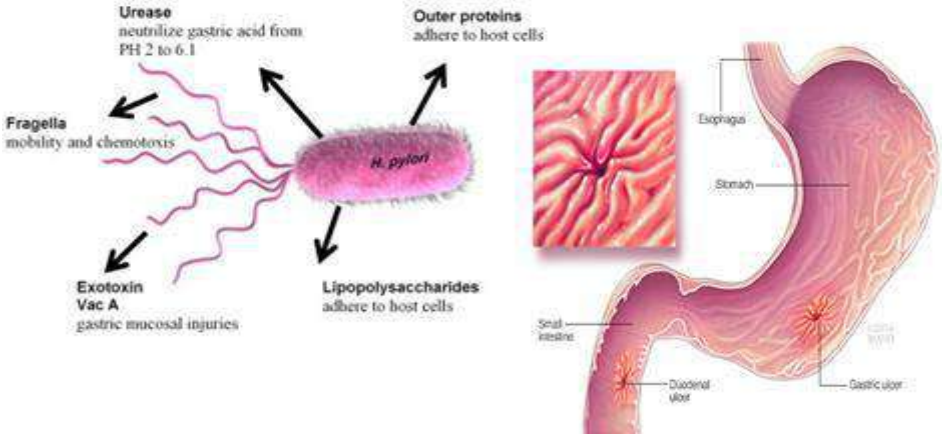
Escherichia coli (UPEC) Remodeling the Host Epigenetic

In human uroepithelial cells, infection with UPEC results in the up-regulation of DNA methyltransferase (DNMT) activity and DNMT1 expression and induces CpG methylation and down-regulation of CDK4/6, a G1-cell-cycle inhibitor regulator. This may increase uroepithelial cell proliferation and pathogen persistence, by counteracting infection-stimulated host cell apoptosis.



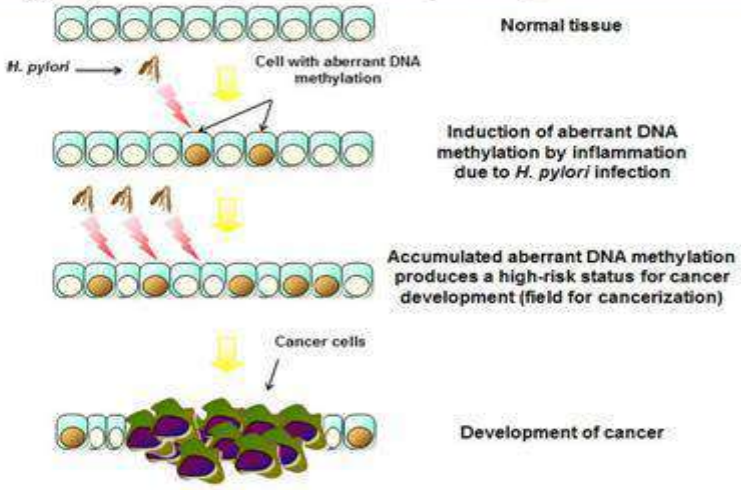
Helicobacter pylori (*H. pylori*)

H. pylori is a spiral-shaped Gram-negative bacteria colonized in the gastrointestinal tract. *H. pylori* infection leads to peptic ulceration, gastritis, and gastric carcinoma.



H. Pylori remodeling Host DNA methylation

H. pylori infection induces aberrant DNA methylation in the human gastric mucosa, strikingly at promoters of genes found methylated in gastric cancer cells.



H. Pylori remodeling Host Epigenetic

H. pylori infection that induces aberrant DNA methylation in the human gastric mucosa, strikingly at promoters of genes found methylated in gastric cancer cells.

H. pylori-associated hypermethylation occurs, for instance, at the E-cadherin gene *CDH1*, tumor-suppressor genes (e.g., *USF1/2* and *WWOX*), DNA repair genes (e.g., *MLH1*), as well as to CpG islands of miRNA genes.

Among signals resulting from chronic inflammation, elevated levels of IL1β and nitric oxide (NO) are proposed to contribute to influence the recruitment of DNMTs at specific loci.

Legionella

Legionella spp., a facultative intracellular Gram-negative bacterium, is an etiologic agent of the atypical pneumonia called the Legionnaires disease.

Legionella Modulates Host Epigenetic

Legionella pathogen modulates the ubiquitin signaling pathway by secreting molecules that mimic certain eukaryotic proteins.

L. pneumophila RomA is secreted into the host cell.

It targets host histone H3 and induces an acetyl/methyl switch at lysine 14 leads to inhibition of apoptosis.