

Introduction to Molecular Biology

- Molecular Biology
- (BIO-302)
- In the broader terms, definition of molecular biology includes all aspects of the study of life from a molecular perspective.

Introduction to Molecular Biology

- More precisely, the term “Molecular Biology” refers to the biology of the molecules related to genes, gene products and heredity.
- In other words, the term molecular biology is often substituted for a more appropriate term, **Molecular Genetics.**

Introduction to Molecular Biology

- So molecular biology grew out of the disciplines of genetics and biochemistry.
- In the present age, world is in the midst of two scientific revolutions. One is information technology and the other is Molecular Biology.

Introduction to Molecular Biology

- Both deal with the handling of large amounts of information.
- Molecular Biology has revolutionized the biological sciences as well especially in the fields of Health Sciences and Agricultural Sciences.

History of Molecular Biology

- Molecular Biology takes its roots mainly from the the disciplines of Genetics and Biochemistry.
- Some important discoveries which led to the emergence of Molecular Biology as a discipline include:-

History of Molecular Biology

Biochemistry

- Synthesis of Urea by Friedrich Wohler (1828).
- The first enzyme diastase was discovered by Anselme Payen in 1833.

Genetics

- Mendel's Experiments on garden pea by Gregor Mendel in 1865.
- Chromosome theory of inheritance by Thomas Hunt Morgan in 1910.

History of Molecular Biology

Biochemistry

- The Discovery of DNA by Friedrich Miescher in 1869.
- Eduard Buchner discovered cell-free fermentation in 1897.
- James B. Sumner crystallized enzyme Urease in 1926.

Genetics

- Barbara McClintock and Harriet Creighton provided physical evidence of recombination in 1931.
- One-gene/one-enzyme hypothesis proposed by George Beadle and Edward Tatum in 1941.

History of Molecular Biology

Biochemistry

- James Watson and Francis Crick discovered the double helical structure of DNA in 1953.

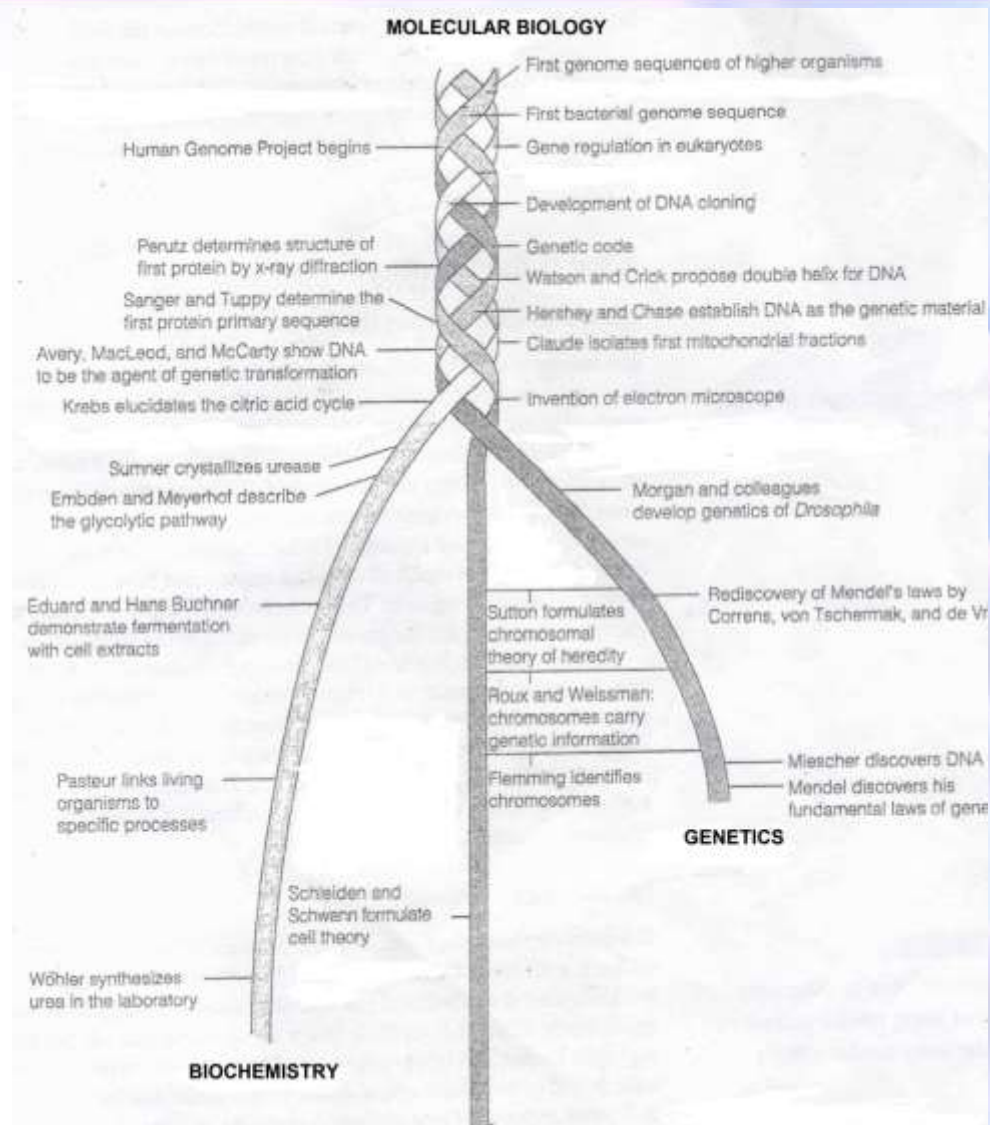
Genetics

- Oswald Avery and his colleagues demonstrated in 1944 that DNA is the hereditary material.
- It was further confirmed by Alfred Hershey and Martha Chase in 1952.

History of Molecular Biology

- At this point, the discoveries in the two fields led to the foundation of a new discipline called Molecular Biology.

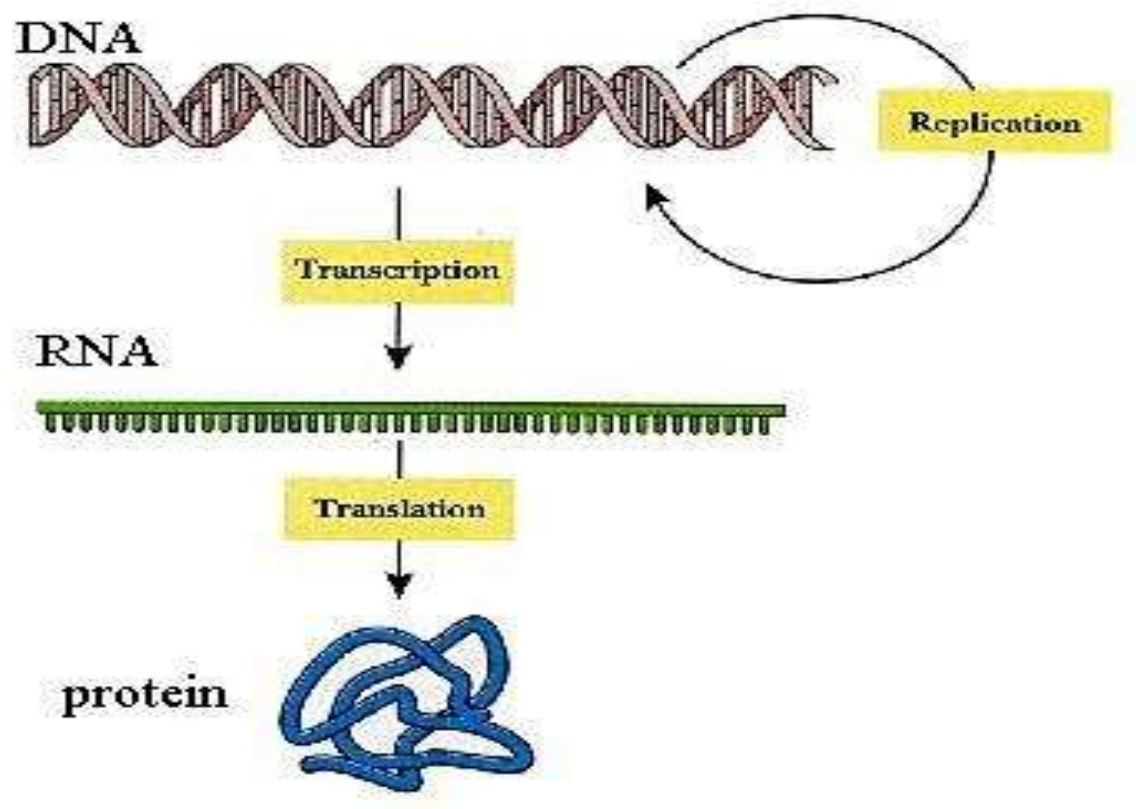
History of Molecular Biology



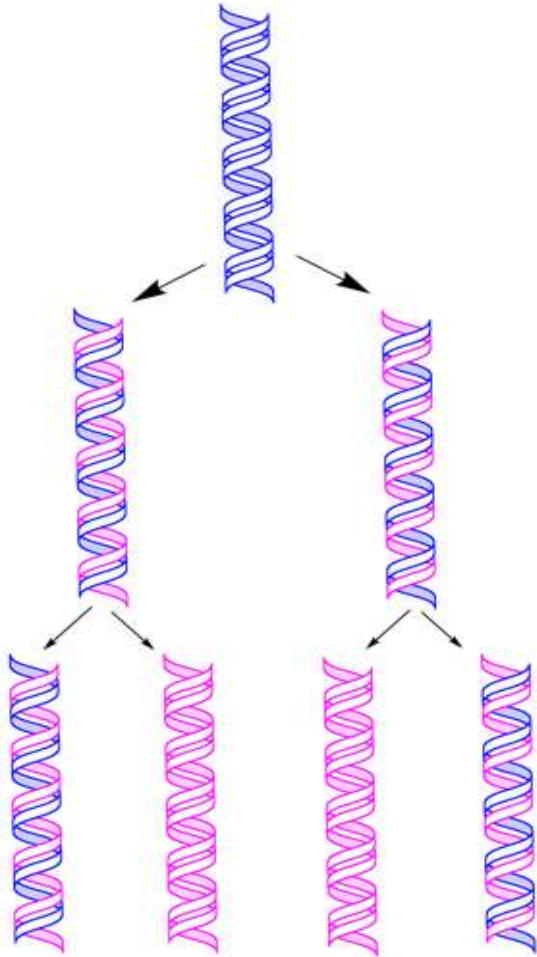
Achievements of Molecular Biology

- In 1957, Francis Crick laid out the “Central dogma of molecular biology” which foretold the relationship between DNA, RNA, and proteins.

Achievements of Molecular Biology



Achievements of Molecular Biology



- In 1958, Mathew Meselson & Franklin Stahl proved that DNA replication was **semi-conservative**.

Achievements of Molecular Biology

- Marshall Nirenberg and Gobind Khorana working independently cracked the code in the early 1960s.
- They found that 3 bases constitute a code word, called a **codon**, that stands for one amino acid.

Achievements of Molecular Biology

Gene Cloning

- Since 1970s, scientists have learned to isolate genes, place them in new organisms and reproduce them by a set of techniques collectively known as **gene cloning**.

Achievements of Molecular Biology

Genetically Modified Organisms (GMOs)

- The technique of Gene cloning led to the creation of a large number of genetically modified organisms with desirable characters.

Achievements of Molecular Biology

Human Genome Project

- The Human Genome Project (HGP) was launched in 1990 and completed in 2003.



Nucleic Acids



Nucleic Acids

- Nucleic acids are important group of biomolecules which are responsible for storage & transmission of hereditary information.
- Like proteins and polysaccharides, nucleic acids are also polymeric compounds.

Nucleic Acids

The repeating units in the nucleic acids are.
Nucleotides.

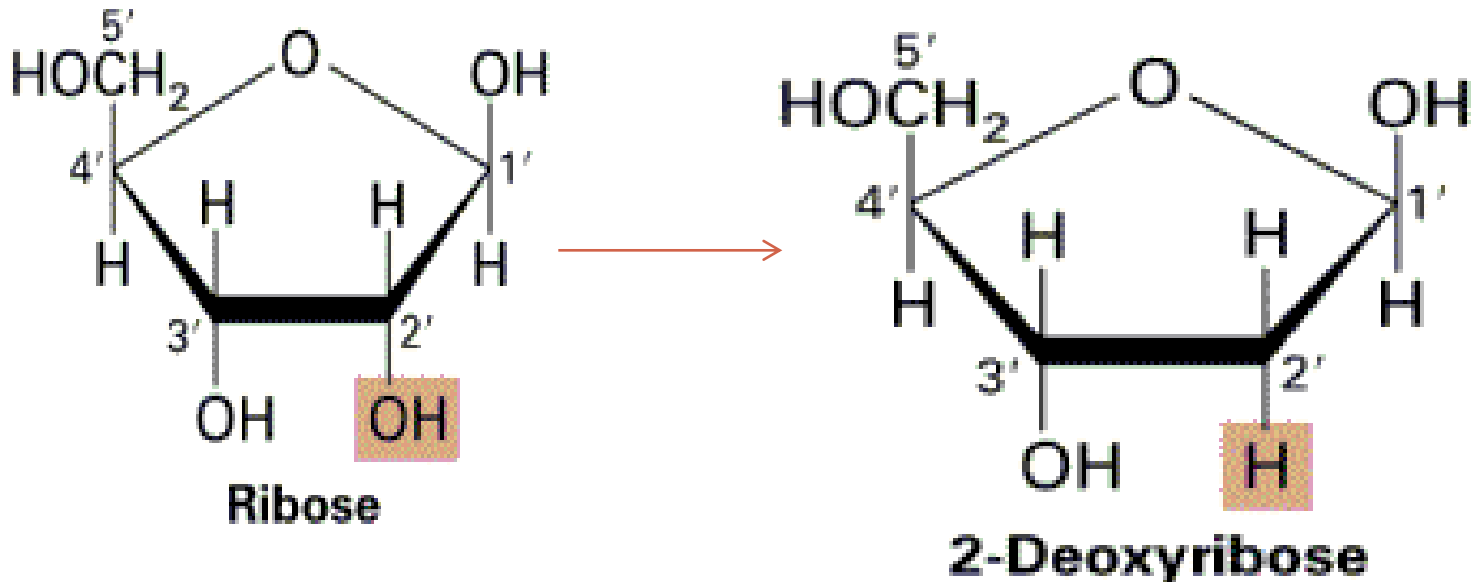


Chemical composition of DNA

- DNA is a polymer of Deoxyribonucleotides.
- Deoxyribonucleotide is composed of three components:
 - Deoxyribose
 - Nitrogenous Base
 - Phosphoric acid

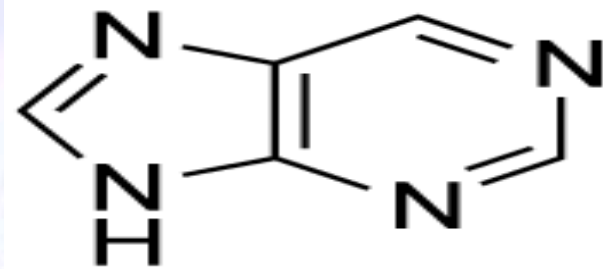
Chemical composition of DNA

Deoxyribose (a pentose sugar derivative)

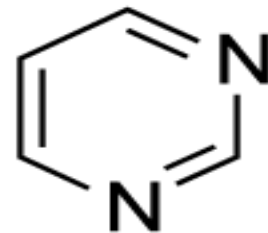


Chemical composition of DNA

Nitrogenous Bases



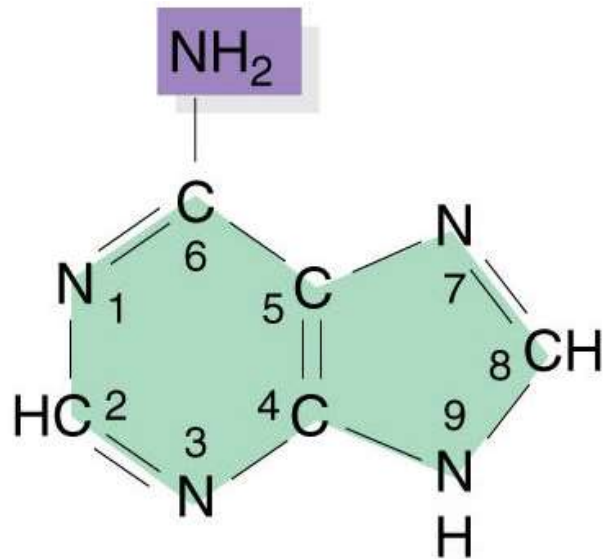
purine



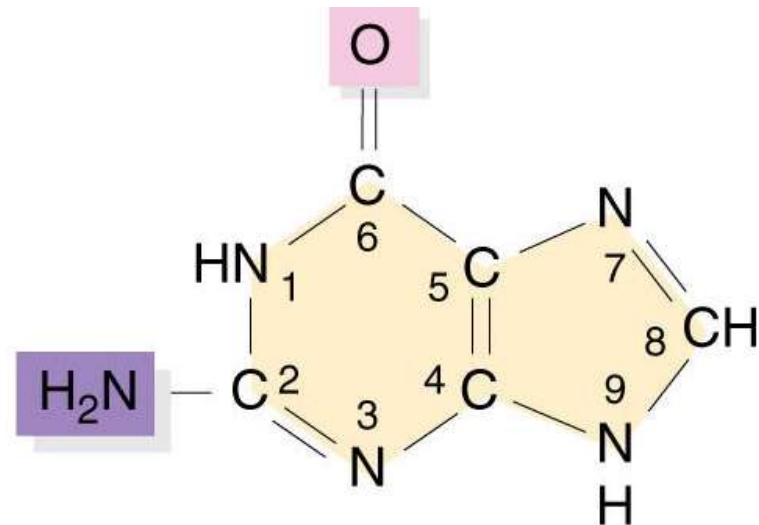
pyrimidine

Chemical composition of DNA

Purines



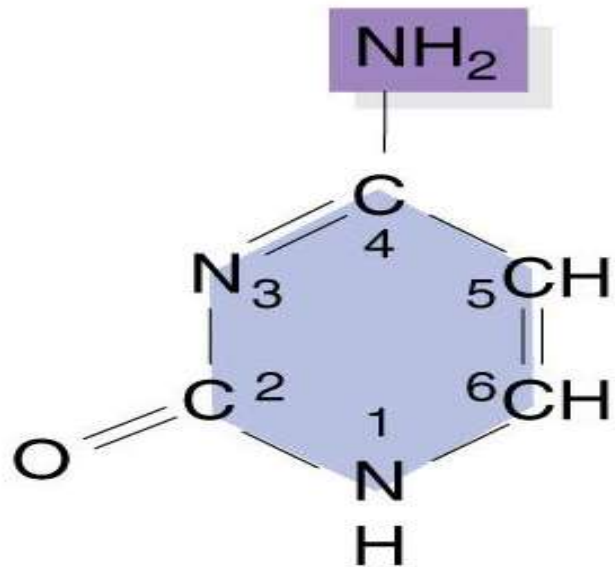
Adenine (A)



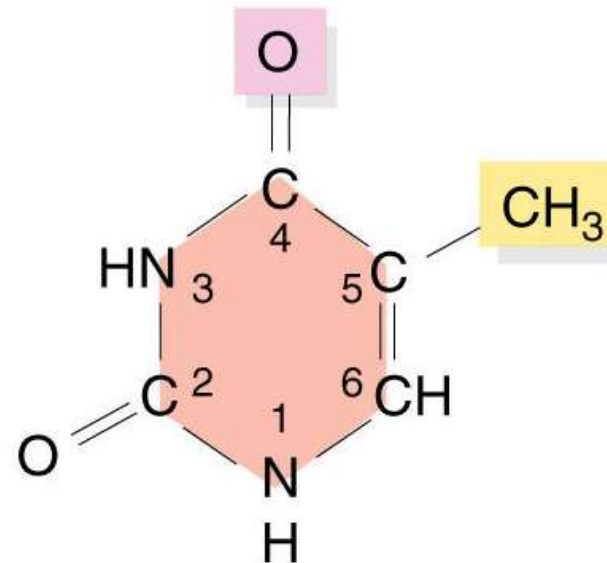
Guanine (G)

Chemical composition of DNA

Pyrimidines



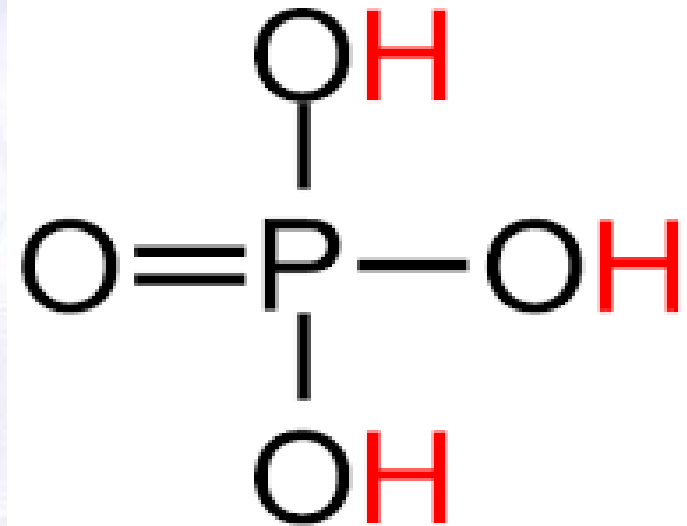
Cytosine (C)



Thymine (T)

Chemical composition of DNA

Phosphoric acid

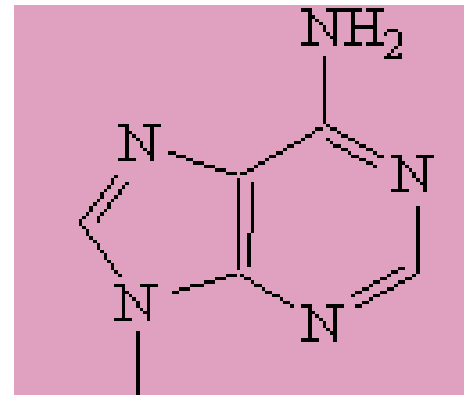
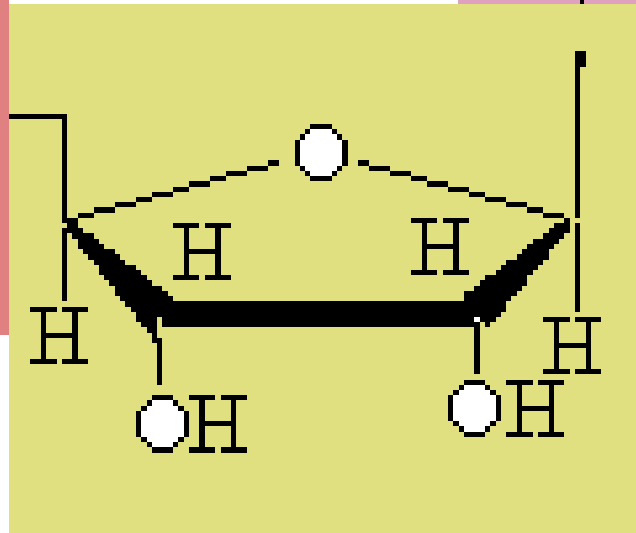
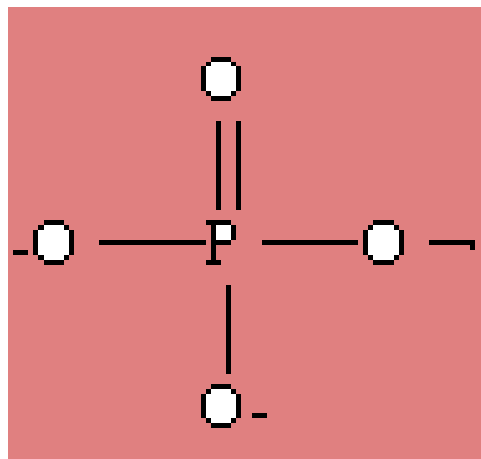


end

Nucleoside & Nucleotide

- A molecule containing all these three components is called a **nucleotide**.
- While a molecule without the phosphate group is called a **nucleoside**.

Nucleoside & Nucleotide



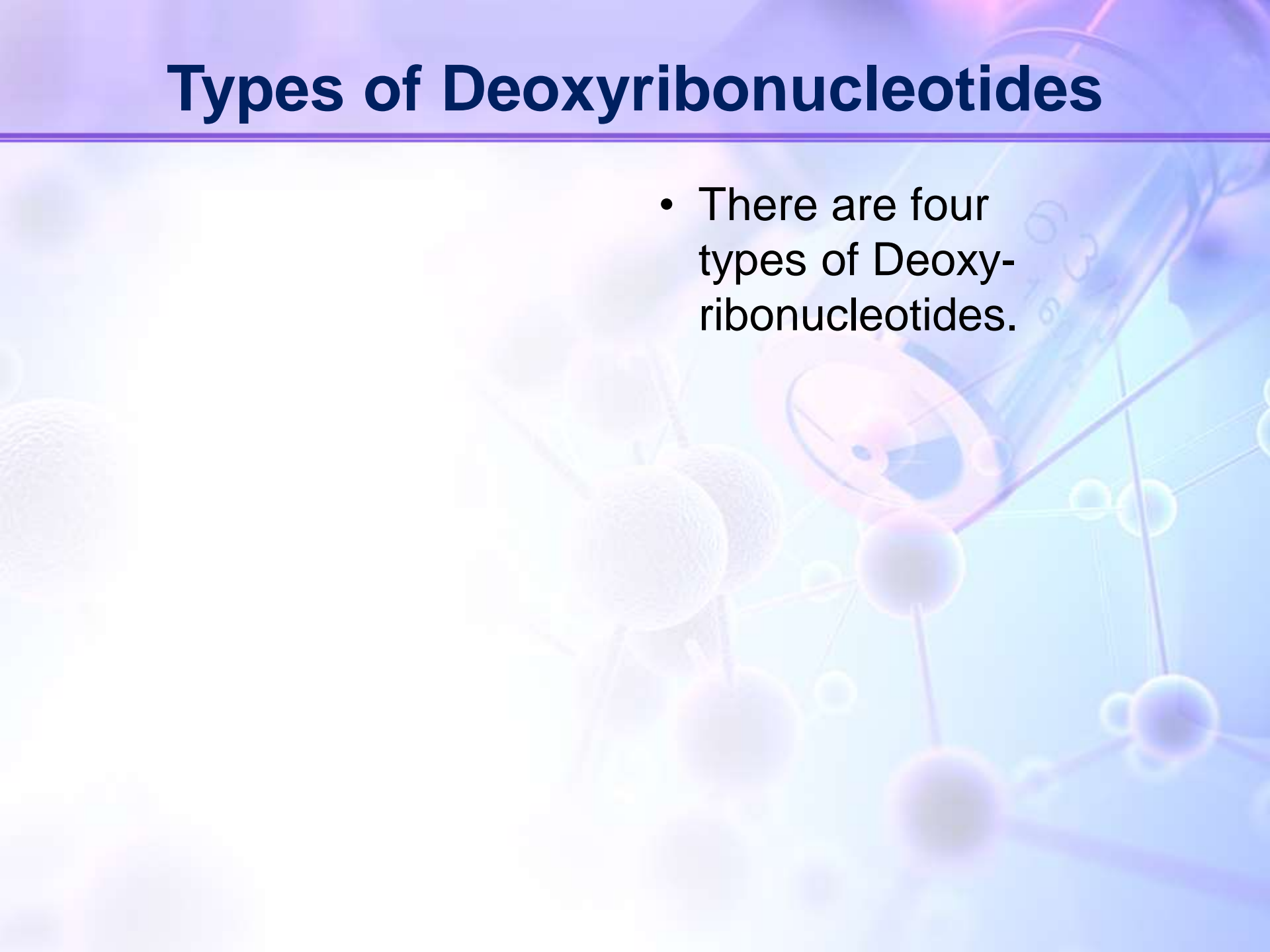
Nucleoside & Nucleotide

Nucleotide =
Nucleoside +
Phosphoric acid
&
Nucleoside =
Nucleotide –
Phosphoric acid

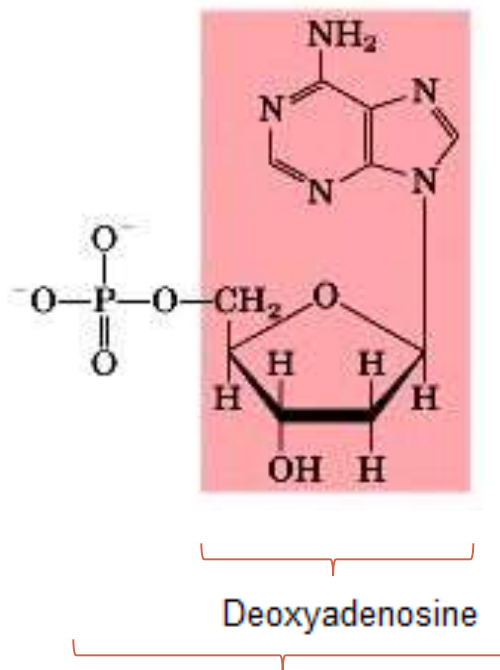
end

Types of Deoxyribonucleotides

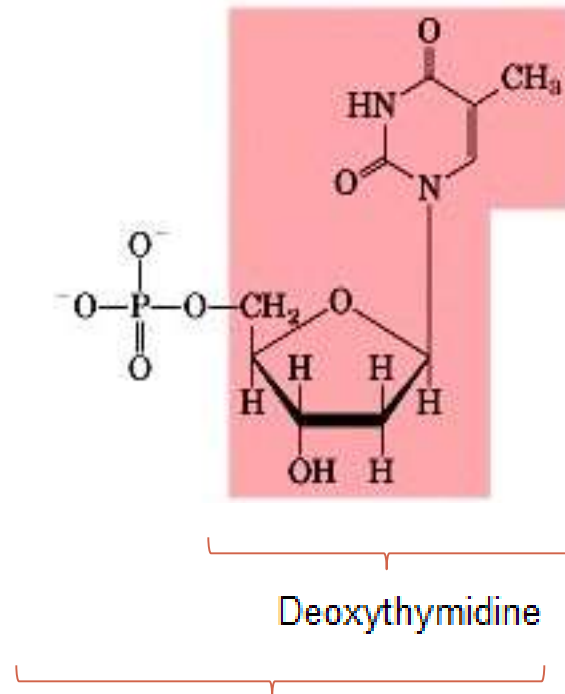
- There are four types of Deoxyribonucleotides.



Types of Deoxyribonucleotides

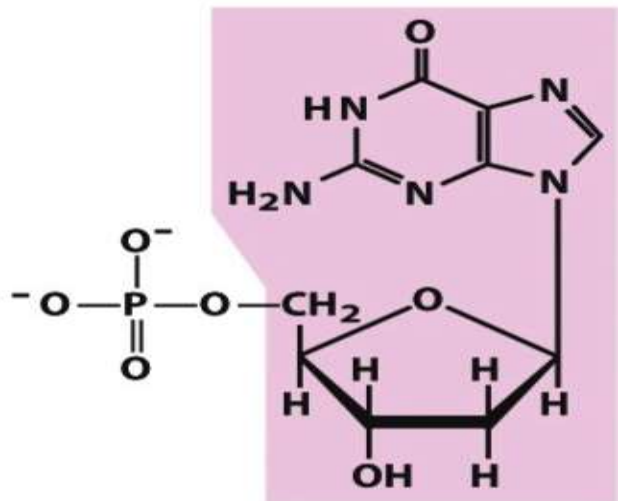


Deoxyadenylate



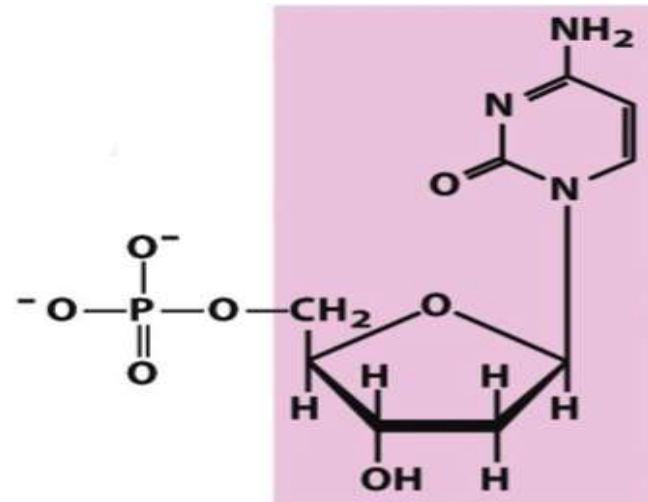
Deoxythymidylate

Types of Deoxyribonucleotides



Deoxyguanosine

Deoxyguanylate



Deoxycytidine

Deoxycytidylate

Types of Deoxyribonucleotides

- These four deoxyribonucleotides make the structural units of DNA.

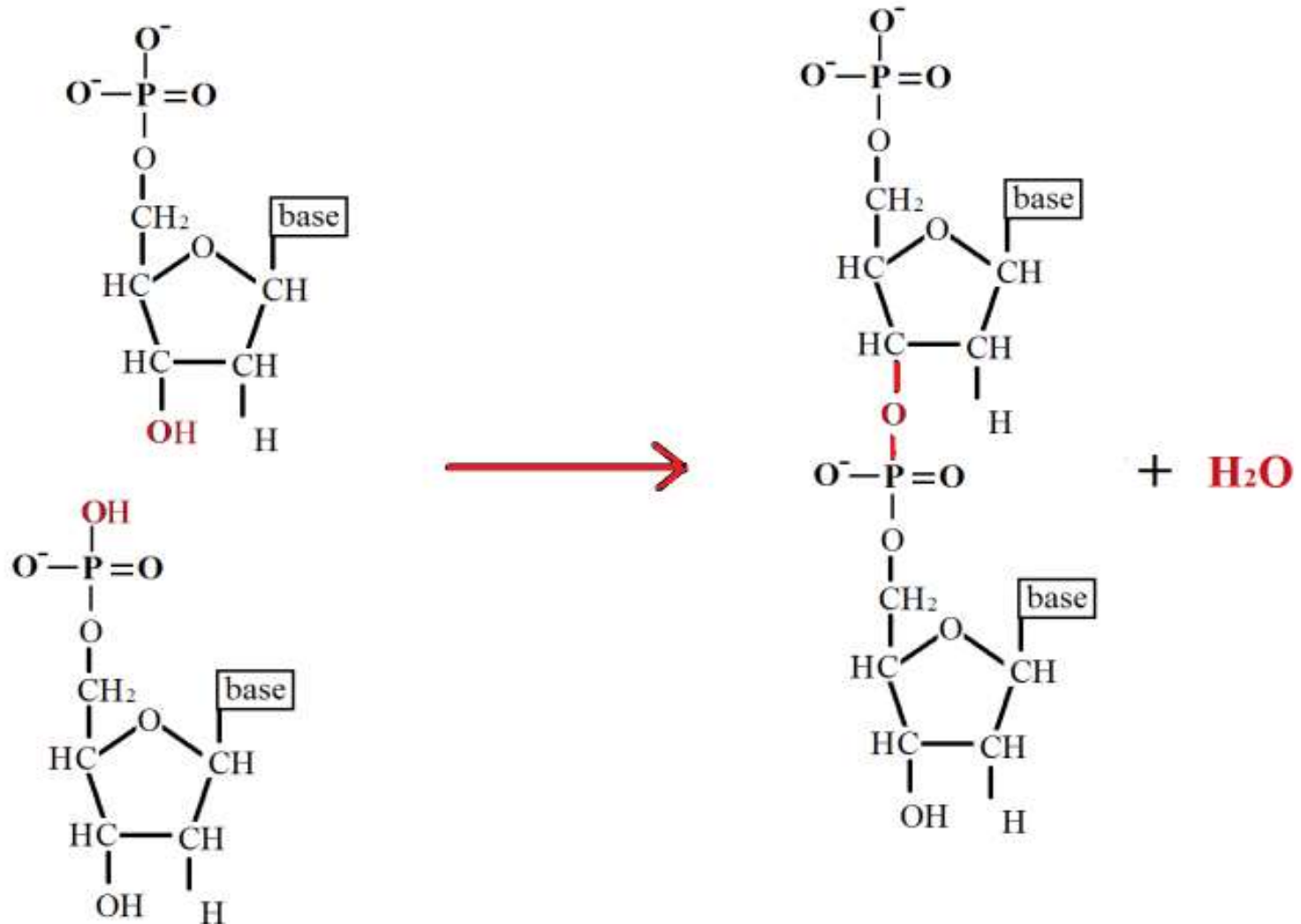
end

How do Deoxyribonucleotides Join?

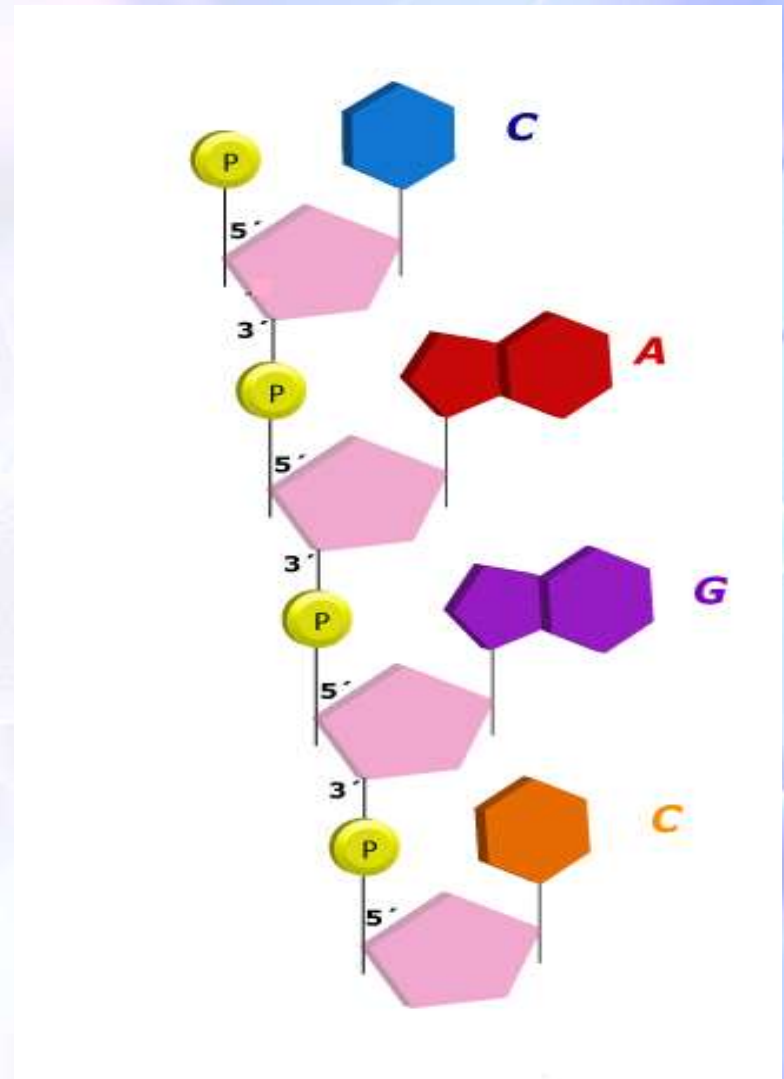
- The successive nucleotides of DNA are joined together through phosphodiester linkages.



How do Deoxyribonucleotides Join?



How do Deoxyribonucleotides Join?



Structure of DNA

Work of Chargaff (Late 1940s)



The discovery of the structure of DNA is one of the greatest events in the history of science.

Structure of DNA

Work of Chargaff (Late 1940s)

- Erwin Chargaff and his colleagues provided a most important clue to the structure of DNA.
- The work of Chargaff led him to following conclusions, also called “Chargaff Rules”:-

Structure of DNA

Work of Chargaff (Late 1940s)

1. Base composition of DNA varies from one species to another.

Structure of DNA

Work of Chargaff (Late 1940s)

2. The DNA isolated from different tissues of the same species have the same base composition.

Structure of DNA

Work of Chargaff (Late 1940s)

3. The base composition of DNA in a given species does not change with an organism's age, nutritional state, or changing environment.

Structure of DNA

Work of Chargaff (Late 1940s)

4. In DNA, the number of adenosine residues is equal to the number of thymidine ($A=T$) and the number of guanosine residues is equal to the number of cytidine ($G=C$).

Structure of DNA

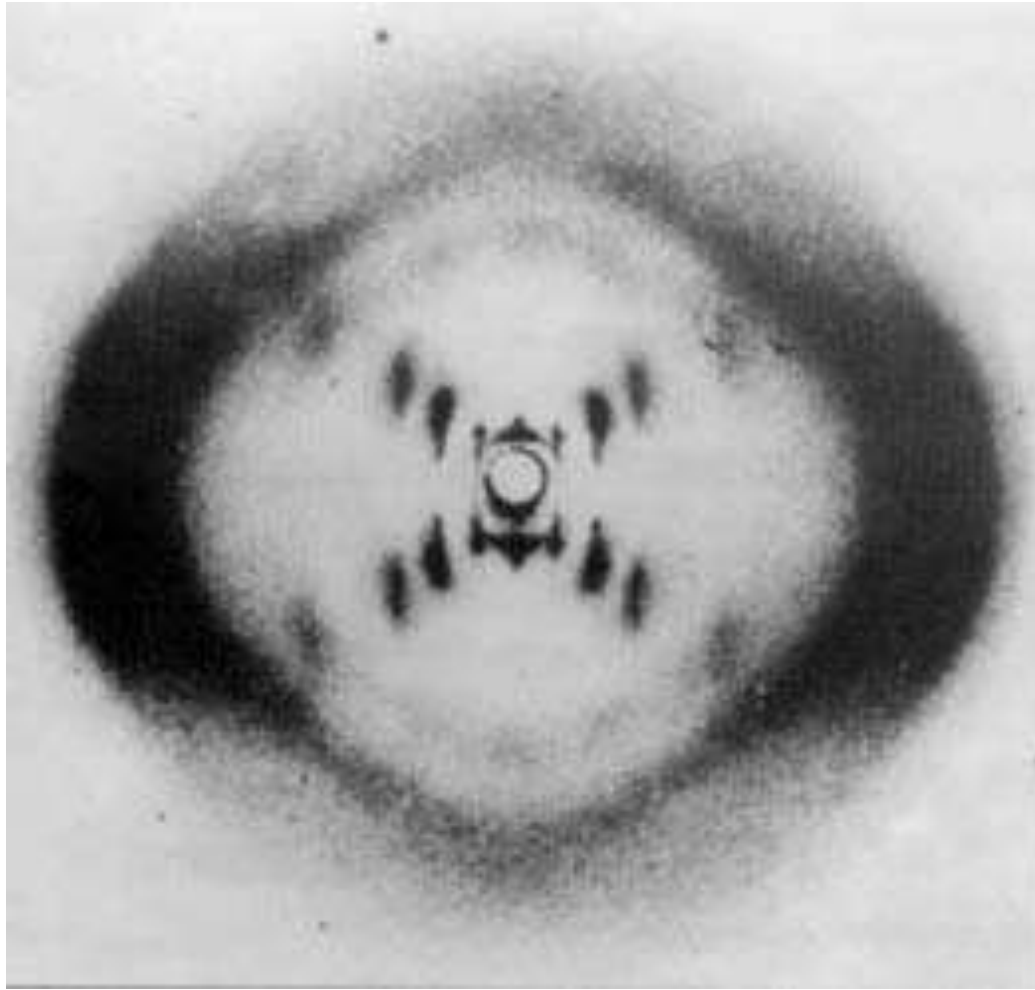
Work of Chargaff (Late 1940s)

- It means that the sum of the purine residues equals the sum of the pyrimidine residues (AG=TC).

Work of Franklin & Wilkins (1950s)

- Rosalind Franklin and Maurice Wilkins performed the x-ray diffraction analysis of DNA fibers.
- They showed that DNA produces a characteristic x-ray diffraction pattern.

Work of Franklin & Wilkins (1950s)



Work of Franklin & Wilkins (1950s)

- From this pattern, they made two important findings:-
 1. DNA molecules are helical.
 2. The helices have two periodicities along their long axis, a primary one of 3.4 \AA and a secondary one of 34 \AA .

Structure of DNA

Work of Watson & Crick

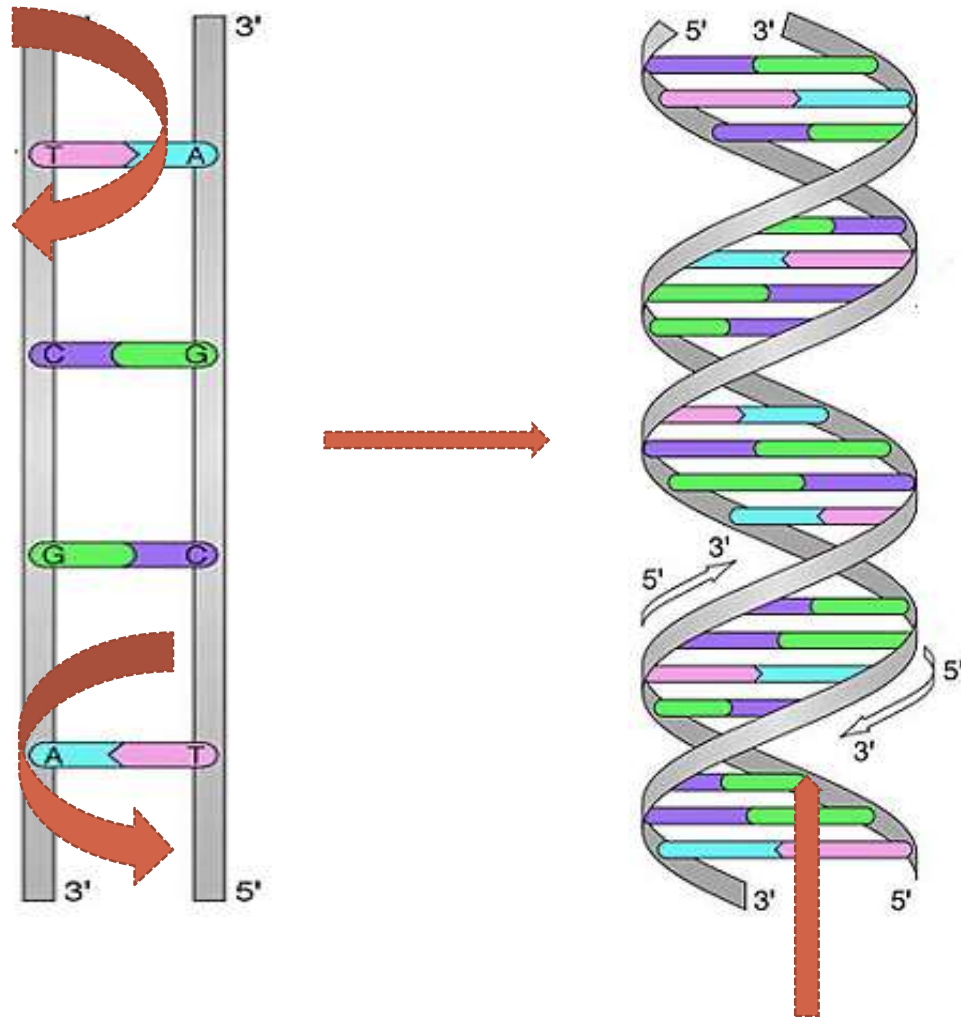
- James Watson and Francis Crick postulated a three dimensional model of DNA structure in 1953. The major features of this model are as follow:-

Structure of DNA

Work of Watson & Crick

1. DNA consists of two helical polynucleotide strands which are wound around the same axis to form a right handed double helix.

Structure of DNA



Structure of DNA

Work of Watson & Crick

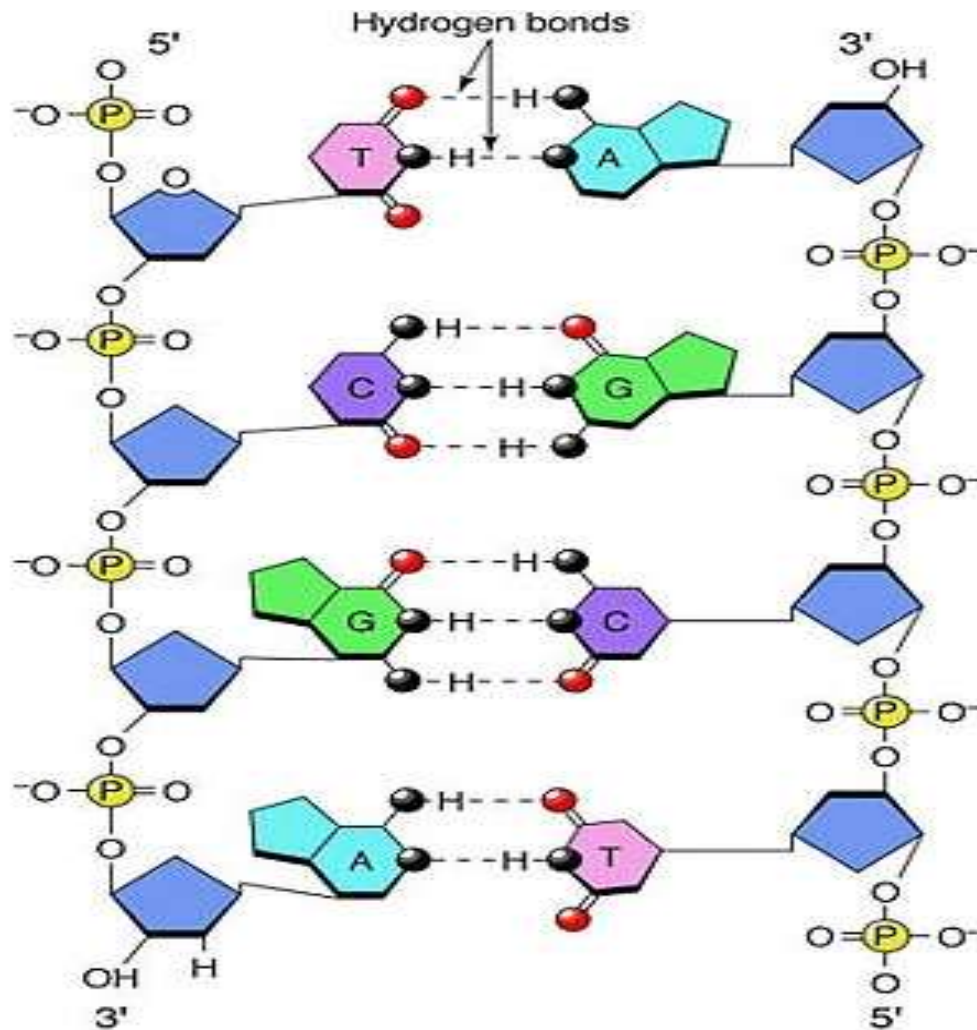
2. The hydrophilic backbones of alternating deoxyribose and phosphate groups are on the outside of the double helix, facing the surrounding water.

Structure of DNA

Work of Watson & Crick

3. The nitrogenous bases of both strands are stacked inside the double helix lying perpendicular to the long axis of the helix.

Structure of DNA

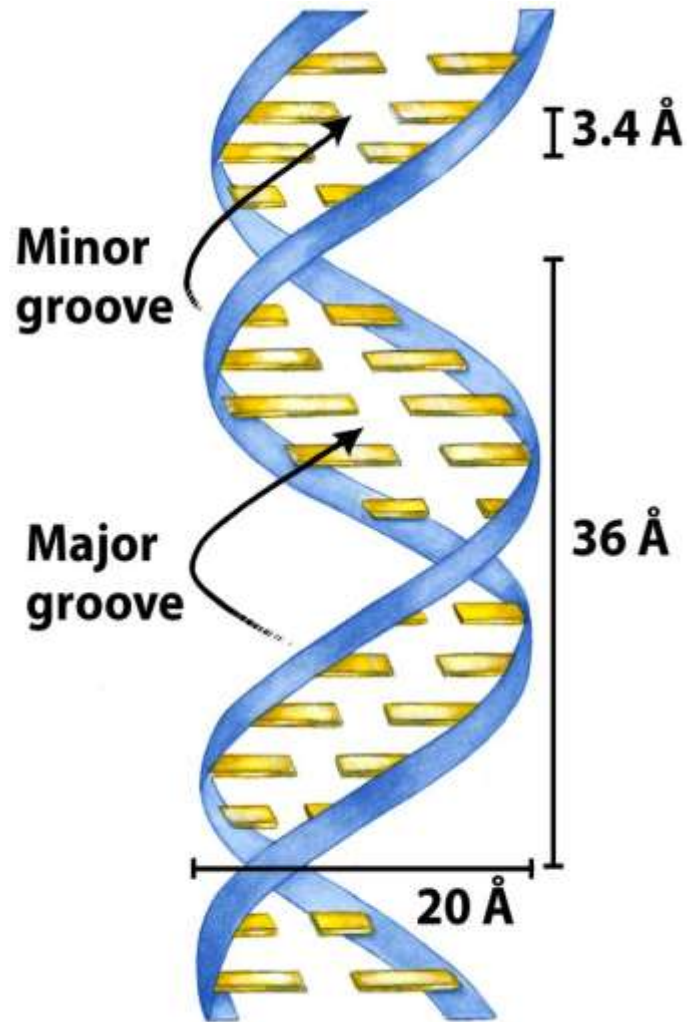


Structure of DNA

Work of Watson & Crick

4. The pairing and coiling of the two strands create a **major groove** and **minor groove** on the surface of the helix.

Structure of DNA

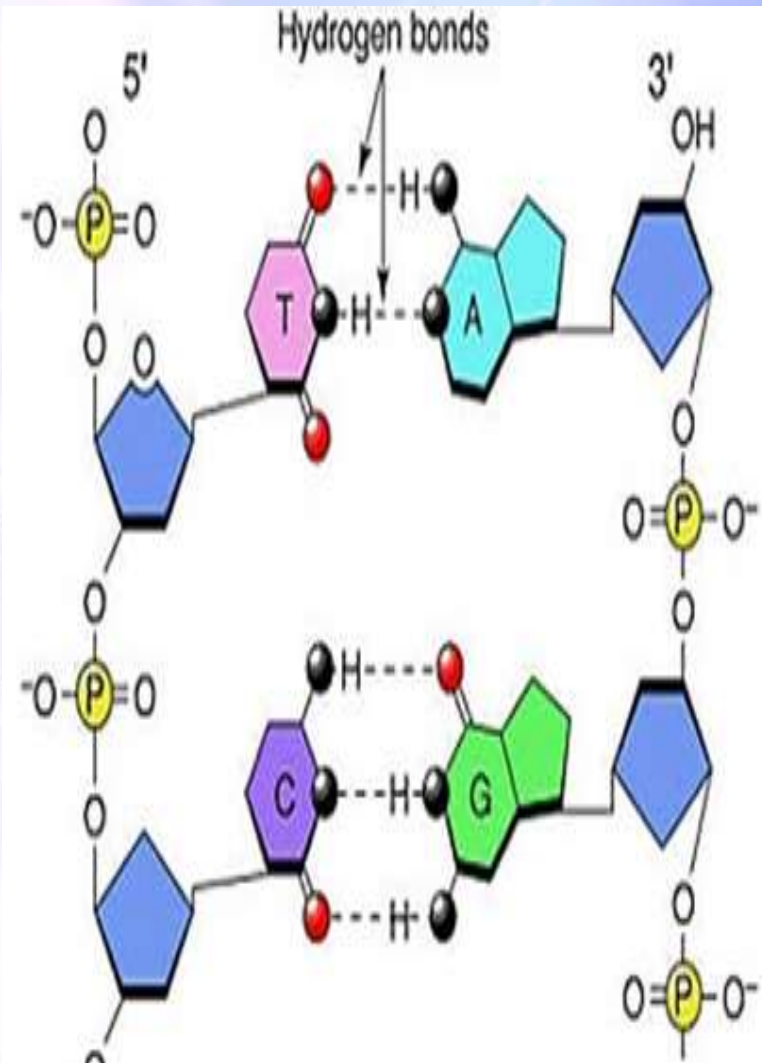


Structure of DNA

Work of Watson & Crick

5. Each nucleotide base of one strand is paired in the same plane with a base of the other strand.

Structure of DNA



Structure of DNA

Work of Watson & Crick

6. They also found that G pairs with C and A pairs with T due to the reason that they fit best within the structure.

Structure of DNA

Work of Watson & Crick

7. The two strands in DNA are present in **antiparallel orientation** i.e; their 5,3-phosphodiester bonds run in the opposite directions.

Structure of DNA

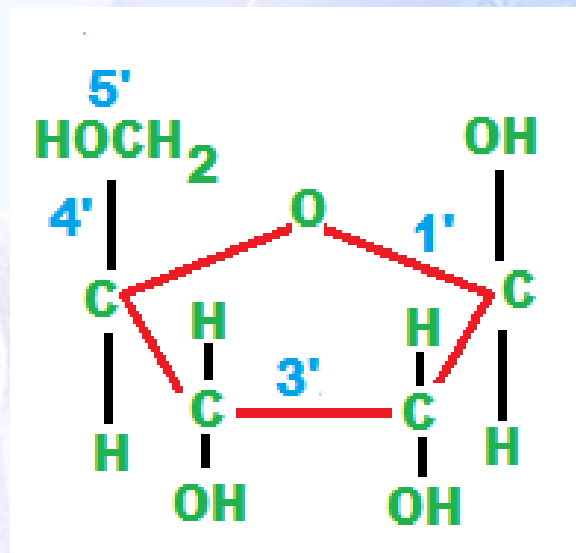
Work of Watson & Crick

8. The two antiparallel strands of double-helical DNA are **complementary** to each other.

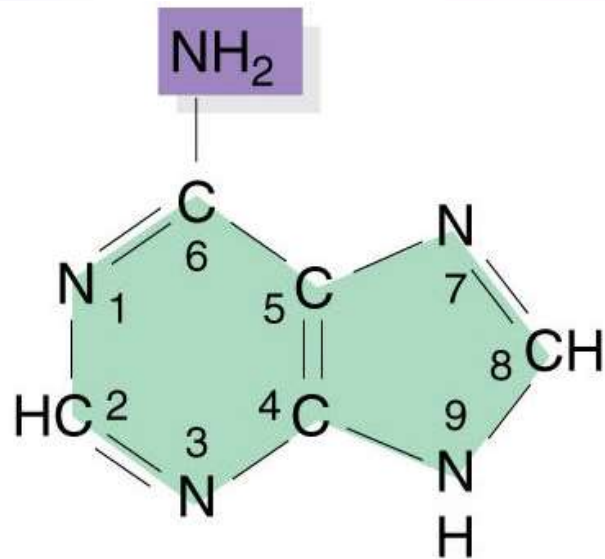
Chemical composition of RNA

- RNA (Ribonucleic acid) is a polymer of ribonucleotides.
- Each ribonucleotide is composed of three components:
 - A ribose sugar
 - A Nitrogenous Base
 - A Phosphoric acid

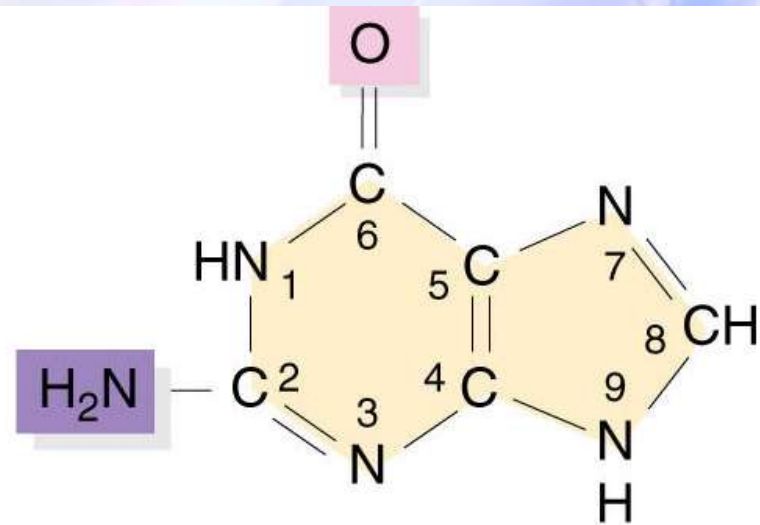
Ribose (a pentose sugar)



Nitrogenous Bases

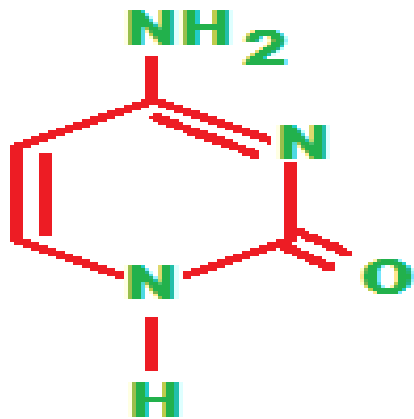


Adenine (A)

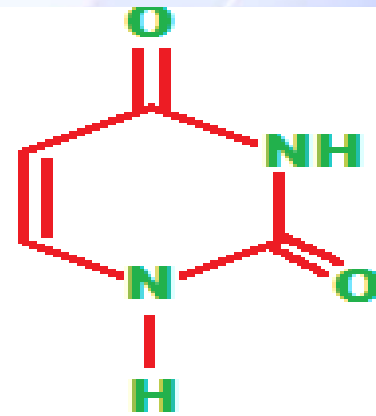


Guanine (G)

Nitrogenous Bases

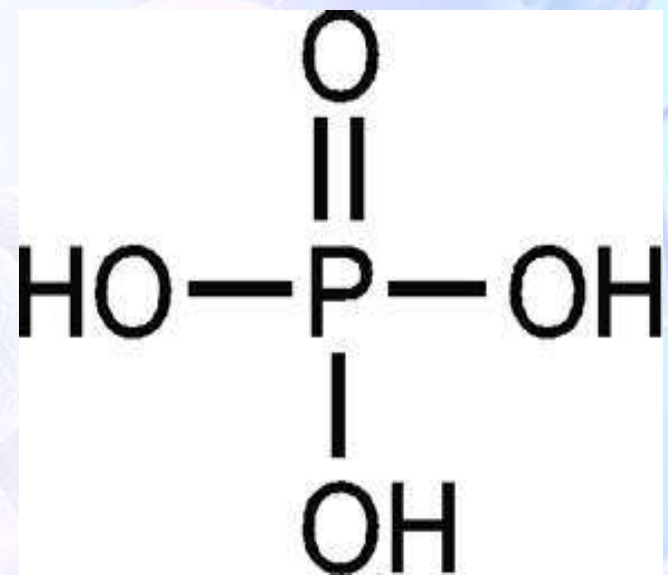


Cytosine (C)

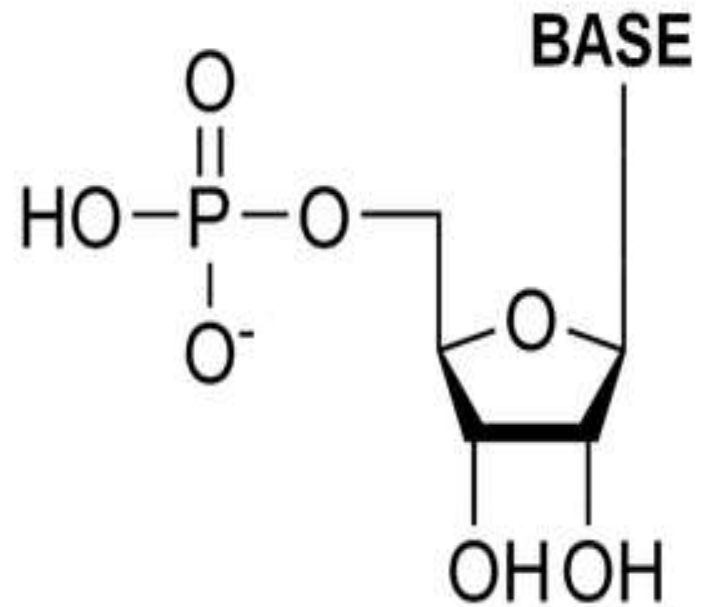


Uracil (U)

Phosphoric acid



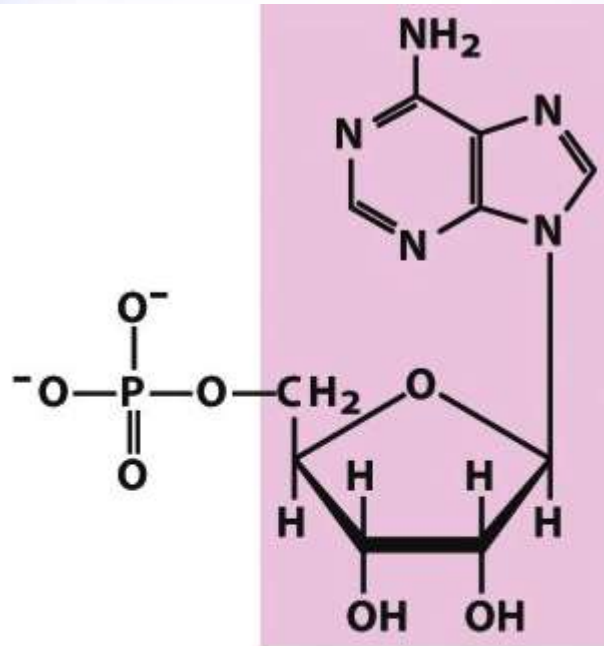
A Ribonucleotide



Types of Ribonucleotides

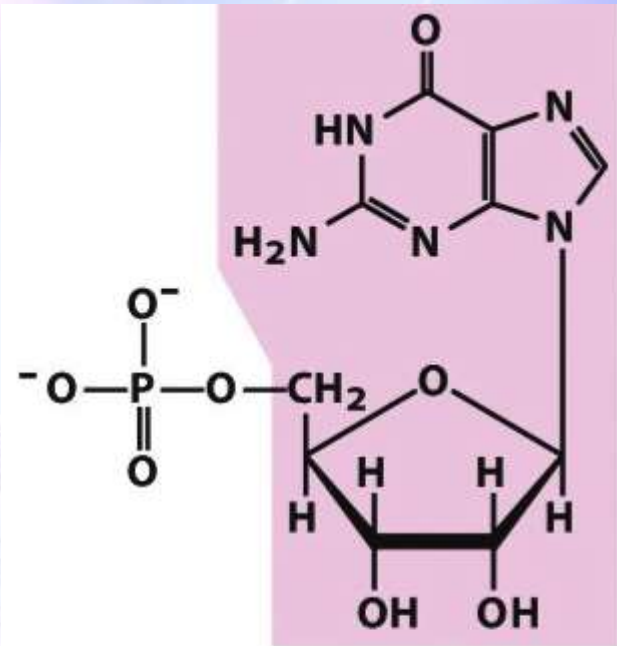
- There are mainly four types of ribonucleotides depending upon the types of nitrogenous bases present in RNA.

Types of Ribonucleotides



**Adenylate (adenosine
5'-monophosphate)**

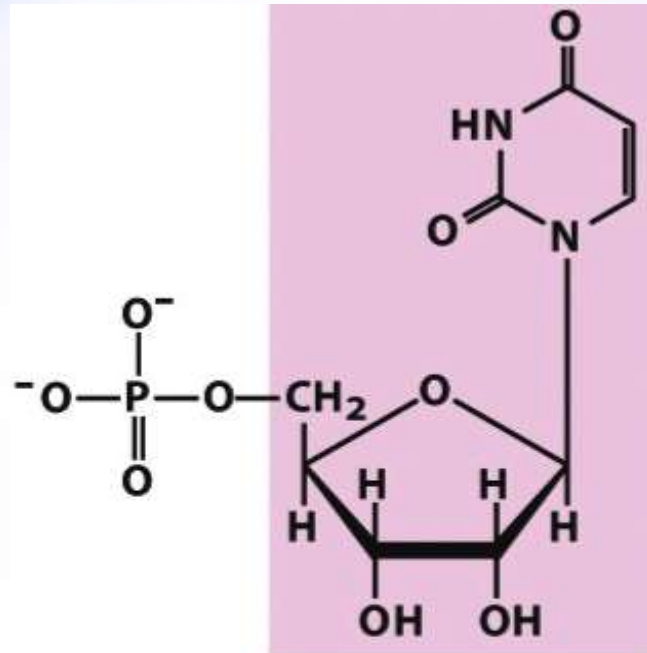
Adenosine



**Guanylate (guanosine
5'-monophosphate)**

Guanosine

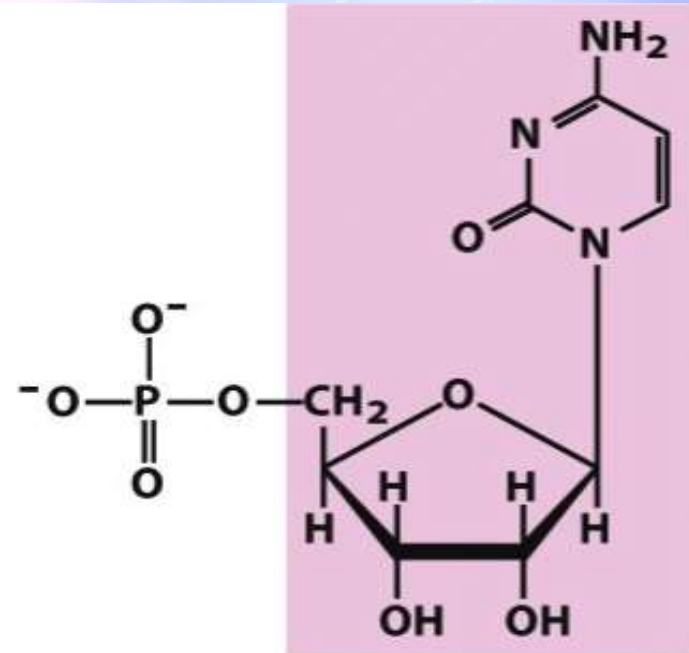
Types of Ribonucleotides



**Uridylate (uridine
5'-monophosphate)**

U, UMP

Uridine



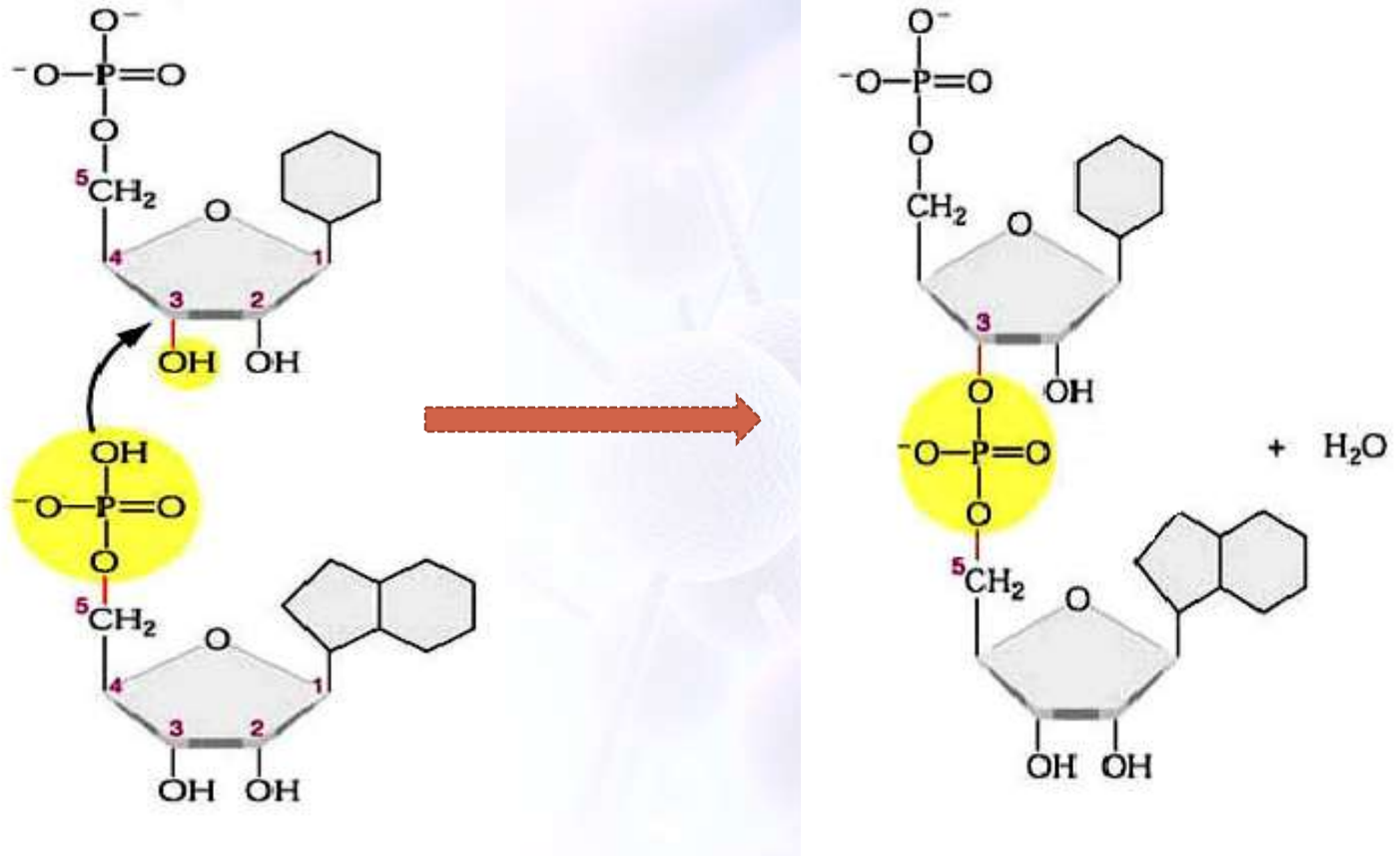
**Cytidylate (cytidine
5'-monophosphate)**

C, CMP

Cytidine

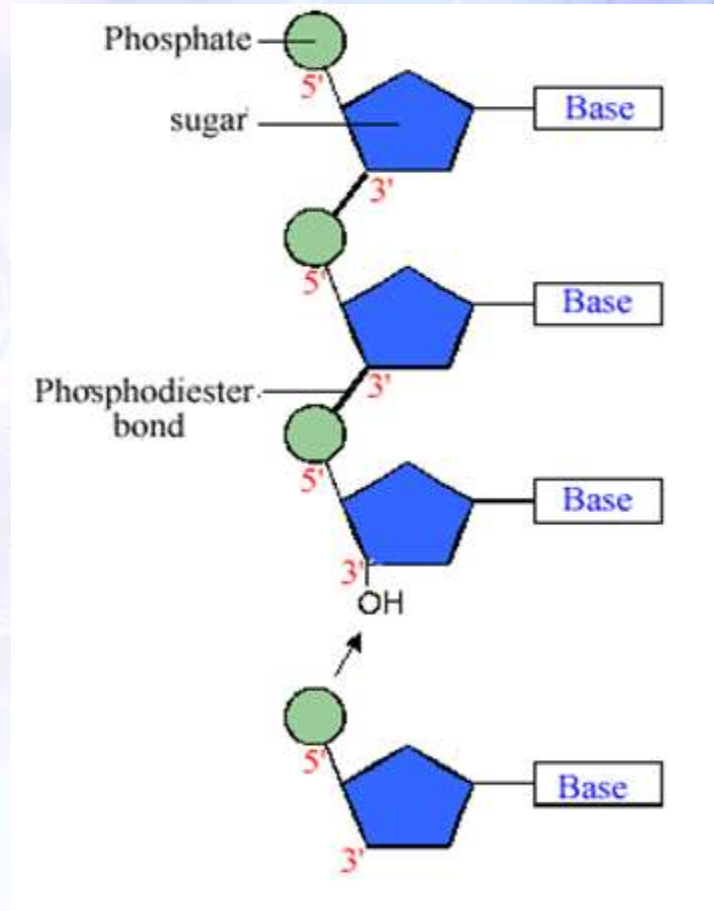
Types of Ribonucleotides

How do Ribonucleotides Join?



Types of Ribonucleotides

A Poly-Ribonucleotide



Types of RNAs

- There are mainly three types of Ribonucleic acids (RNAs) present in the cells of living organisms.
 - Messenger RNA (mRNA)
 - Transfer RNA (tRNA)
 - Ribosomal RNA (rRNA)

Types of RNAs

Messenger RNA (mRNA)

- It is the type of RNA that carries genetic information from DNA to the protein biosynthetic machinery of the ribosome.
- It provides the templates that specify amino acid sequences in polypeptide chains.
- The process of forming mRNA on a DNA template is known as **transcription**.

Types of RNAs

Messenger RNA (mRNA)

- It may be **monocistronic** or **polycistronic**.
- The length of mRNA molecules is variable and it depends on the length of gene.

Types of RNAs

Transfer RNA (tRNA)

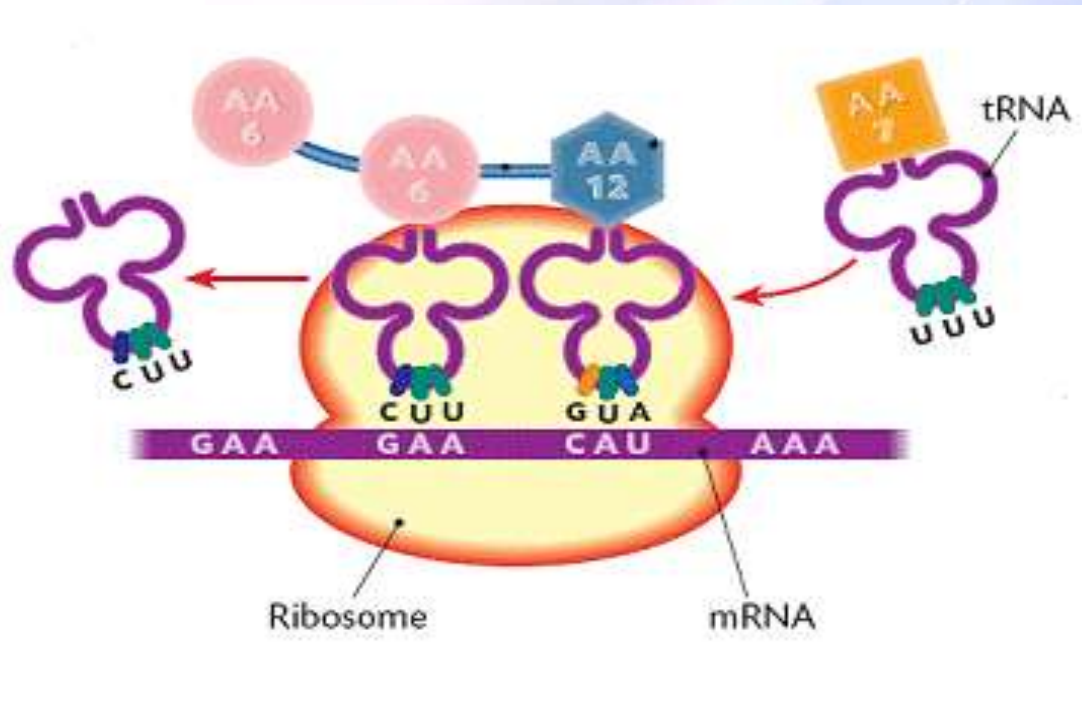
- Transfer RNAs serve as adapter molecules in the process of protein synthesis.
- They are covalently linked to an amino acid at one end.

Types of RNAs

Transfer RNA (tRNA)

- They pair with the mRNA in such a way that amino acids are joined to a growing polypeptide in the correct sequence.

Types of RNAs



Types of RNAs

Ribosomal RNA (rRNA)

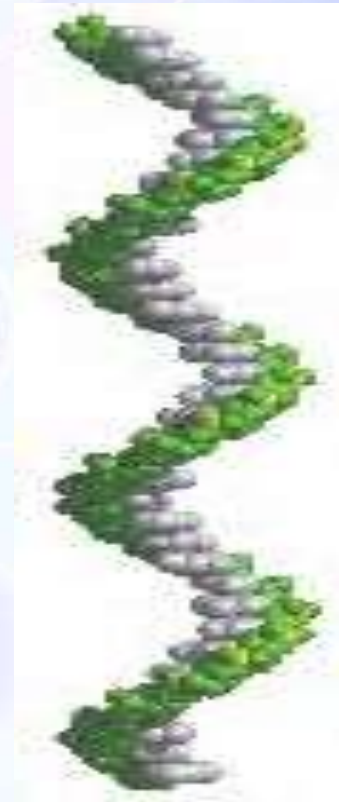
- Ribosomal RNAs are components of ribosomes.
- rRNA is a predominant material in the ribosomes constituting about 60% of its weight.
- It has a number of functions to perform in the ribosomes.

Structures of RNAs

- mRNA is always single stranded when it is formed from DNA.
- But this single strand assumes a double helical conformation soon after its formation.
- This confirmation is achieved mainly due to base stacking interactions.

Structures of RNAs

Messenger RNA (mRNA)



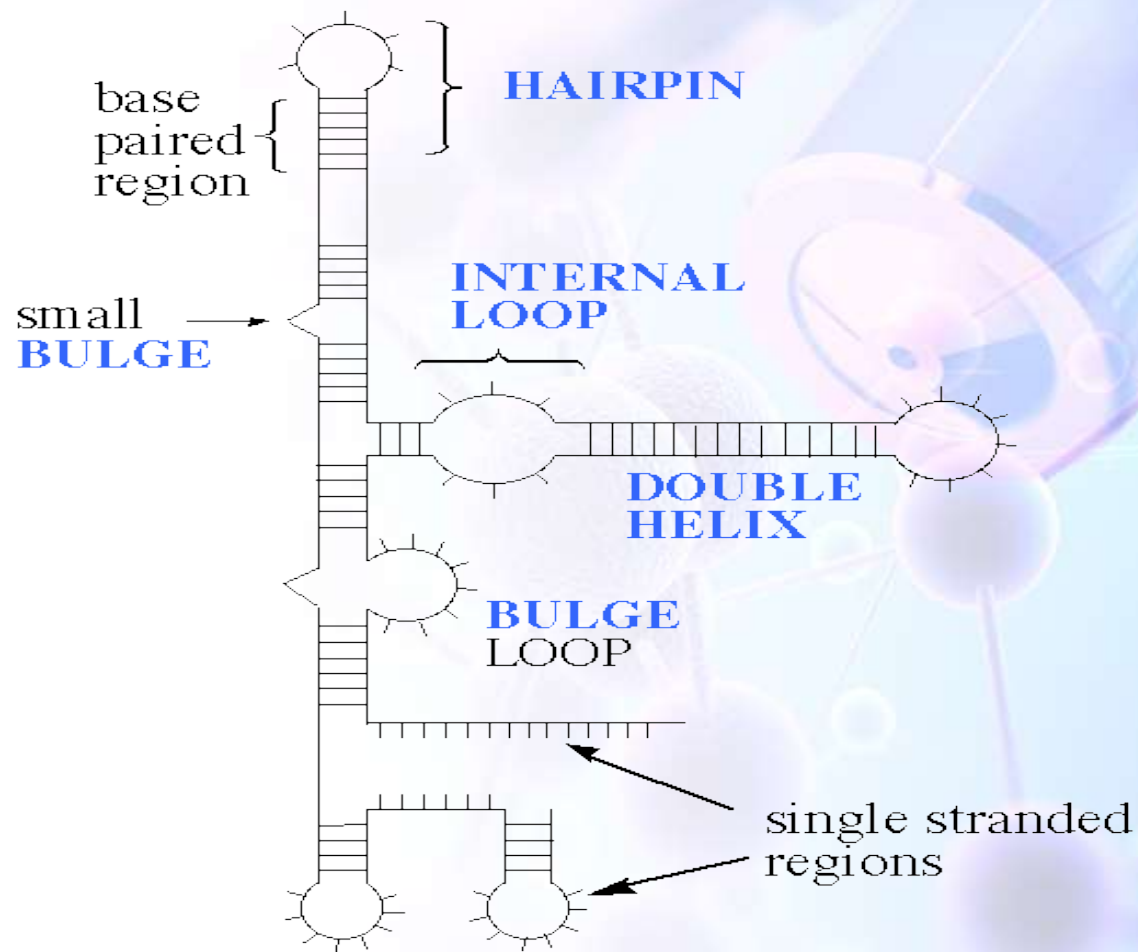
Structures of RNAs

- Self-complementary sequences may occur in the RNA molecules which produce more complex structures.
- So RNA can base-pair with complementary regions of either RNA or DNA.
- RNA has no any regular secondary structure that serves as a reference point. The three-dimensional structures of many RNAs are complex and unique.

Structures of RNAs

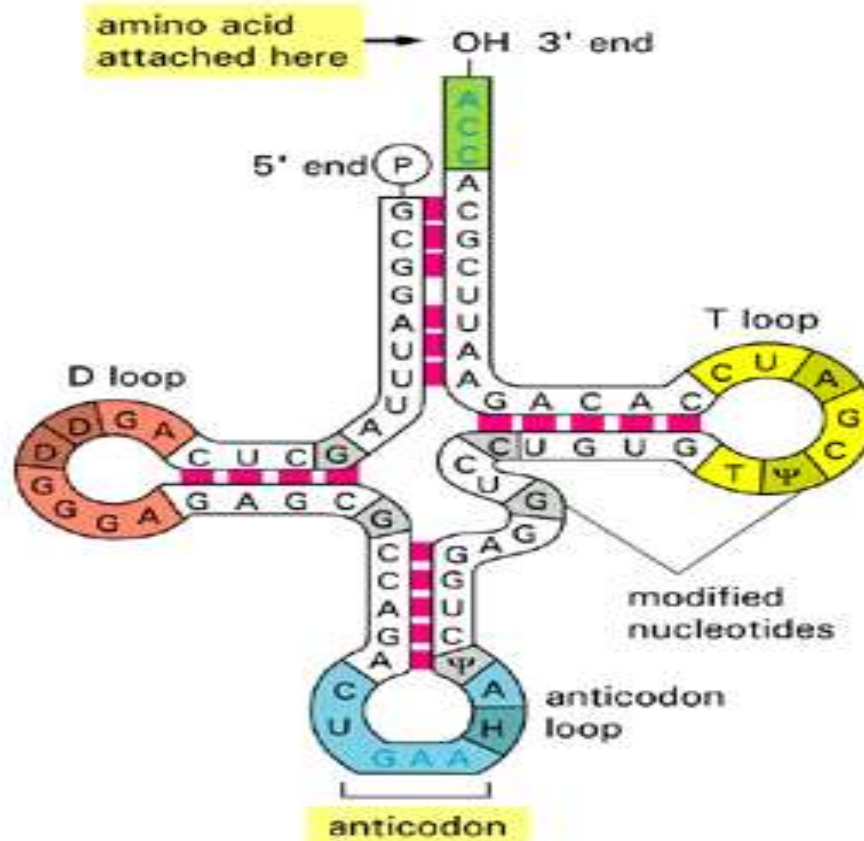
- Breaks in the helix caused by mismatched or unmatched bases in one or both strands are common and result in bulges or internal loops.
- Hairpin loops form between nearby self-complementary sequences.

Structures of RNAs



Structures of RNAs

Transfer RNA (tRNA)



Nature of Genetic Material

- After establishment of the fact that genes are the physical units located on the chromosomes. A major problem for the biologists was to find out the molecules responsible for carrying the hereditary information.

Nature of Genetic Material

Characteristics of Genetic Material

- Genetic material must contain complex information.
- Genetic material must replicate faithfully.
- Genetic material must encode phenotype.

Nature of Genetic Material

- Three sets of experiments provided a pivotal evidence that DNA rather than protein, is the hereditary material.
 - Griffith's Experiments (1928)
 - Avery's Experiments (1944)
 - Hershey-Chase experiments (1952)

Griffith's Experiments

- The first clue that DNA was the carrier of hereditary information came from the work of **Fred Griffith** in 1928.
- He observed the phenomenon of **Transformation** in the bacterium *Streptococcus pneumonia*.

Griffith's Experiments

- Griffith had isolated different strains of *S. pneumonia*.
- In the virulent strain, each bacterium is surrounded by a polysaccharide coat.

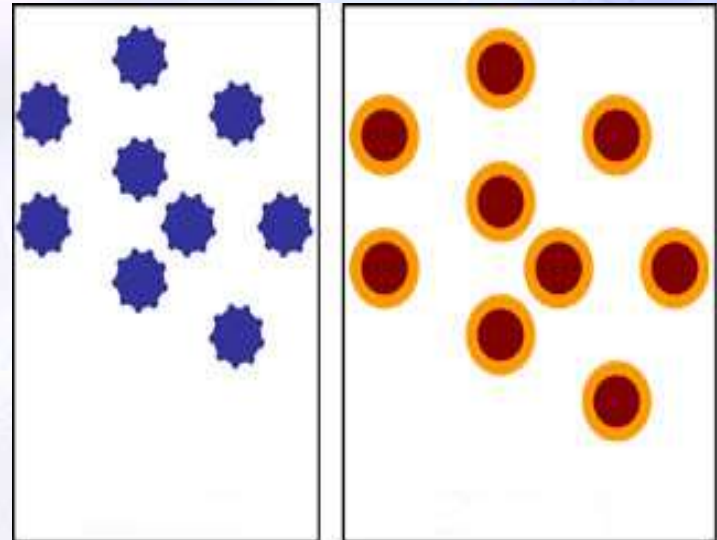
Griffith's Experiments

- This polysaccharide coat makes the bacterial colony appear smooth when grown on an agar plate.
- These forms of *S. pneumonia* are referred to as S forms.

Griffith's Experiments

- Griffith found that these virulent forms occasionally get mutated to non-virulent forms, which lack a polysaccharide coat.
- They produced a rough appearing colony on an agar plate; these forms are referred to as **R forms**.

Griffith's Experiments



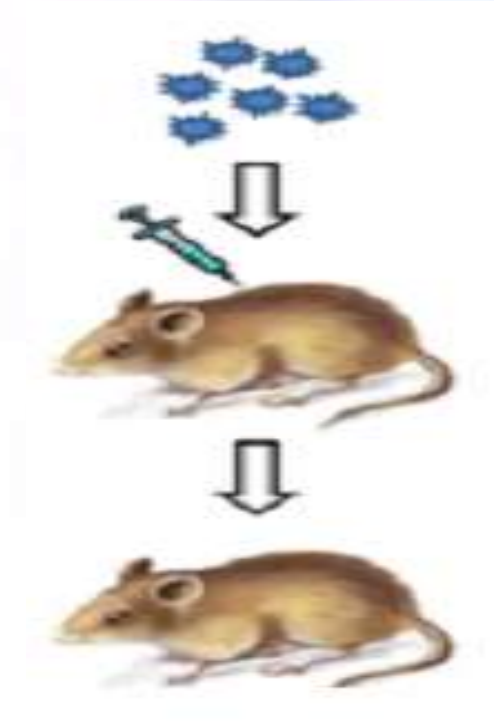
R Forms

S Forms

Transformation Experiments

- Griffith carried out series of experiments using S-Form & R-Form strains of *S. pneumonia* and mouse as an experimental animal.

Transformation Experiments



Transformation Experiments



Transformation Experiments



Transformation Experiments



Transformation Experiments

Possible Interpretations

- It could have been the case that S- Type bacteria were not completely killed and a few live bacteria remained in the culture.

Transformation Experiments

- A second interpretation was that the live R-Type bacteria had mutated to the virulent S form.
- Griffith finally concluded that R-Type bacteria had been *transformed*.

Transformation Experiments

- Griffith theorized that some substance of the dead bacteria might be responsible for that transformation . He called that substance as the **transforming principle**.

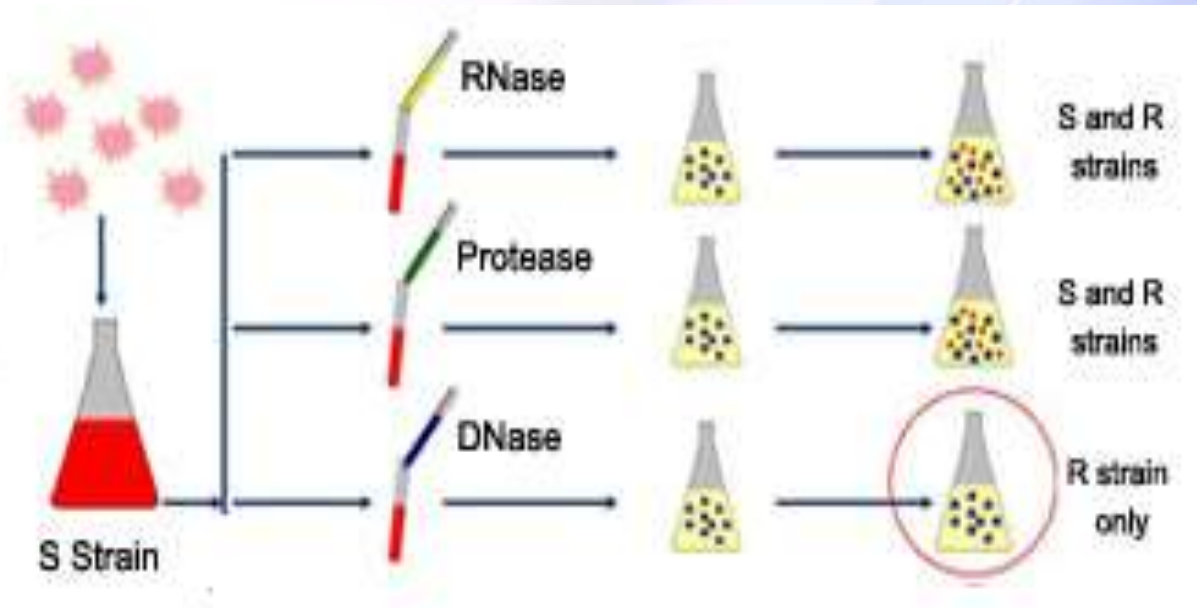
Avery, MacLeod, and McCarty's experiment

- Avery, MacLeod and McCarty succeeded in isolating and purifying the transforming substance in 1944.
- They showed that it had a chemical composition closely matching that of DNA and quite different from that of proteins.

Avery, MacLeod, and McCarty's experiment

- They showed that proteolytic enzymes had no effect on the transforming substance.
- Ribonuclease also had no effect on it.
- However, enzymes capable of destroying DNA, destroyed that substance.

Avery, MacLeod, and McCarty's experiment



Avery, MacLeod, and McCarty's experiment

- Avery, MacLeod, and McCarty further showed that the purified transforming substance precipitated at about the same rate as purified DNA.
- It absorbed ultraviolet light at the same wavelengths as does DNA.

Avery, MacLeod, and McCarty's experiment

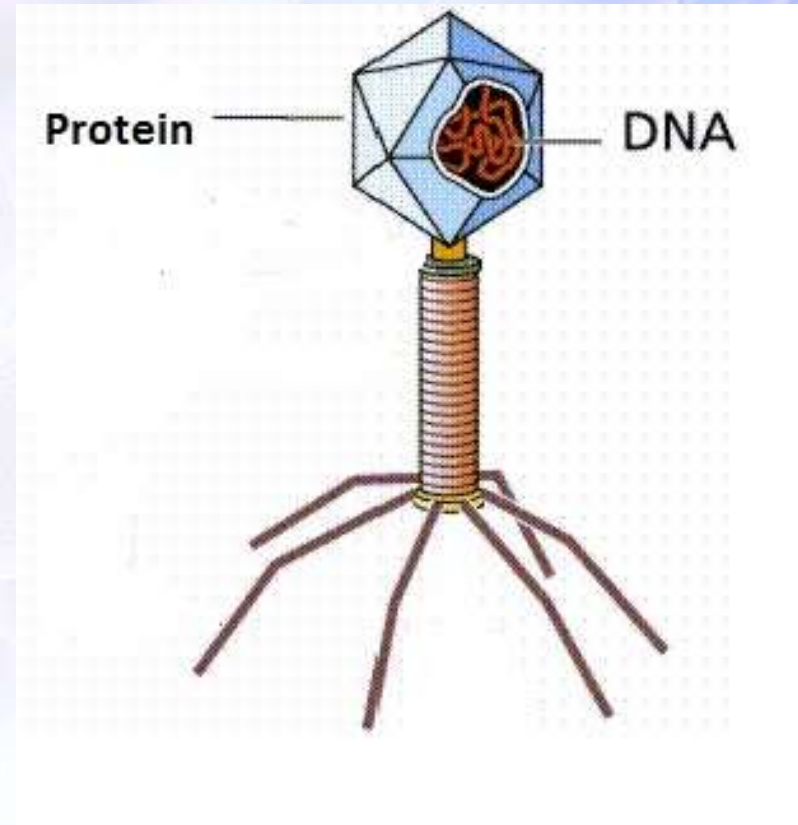
- These findings provided compelling evidence that the transforming principle—and therefore the genetic information—resides in DNA.

Hershey & Chase experiment

- Alfred Hershey and Martha Chase in 1952 provided a second evidence that DNA is the genetic material.
- They were working on T2 which is a *bacteriophage* that infects the bacterium *Escherichia coli*.

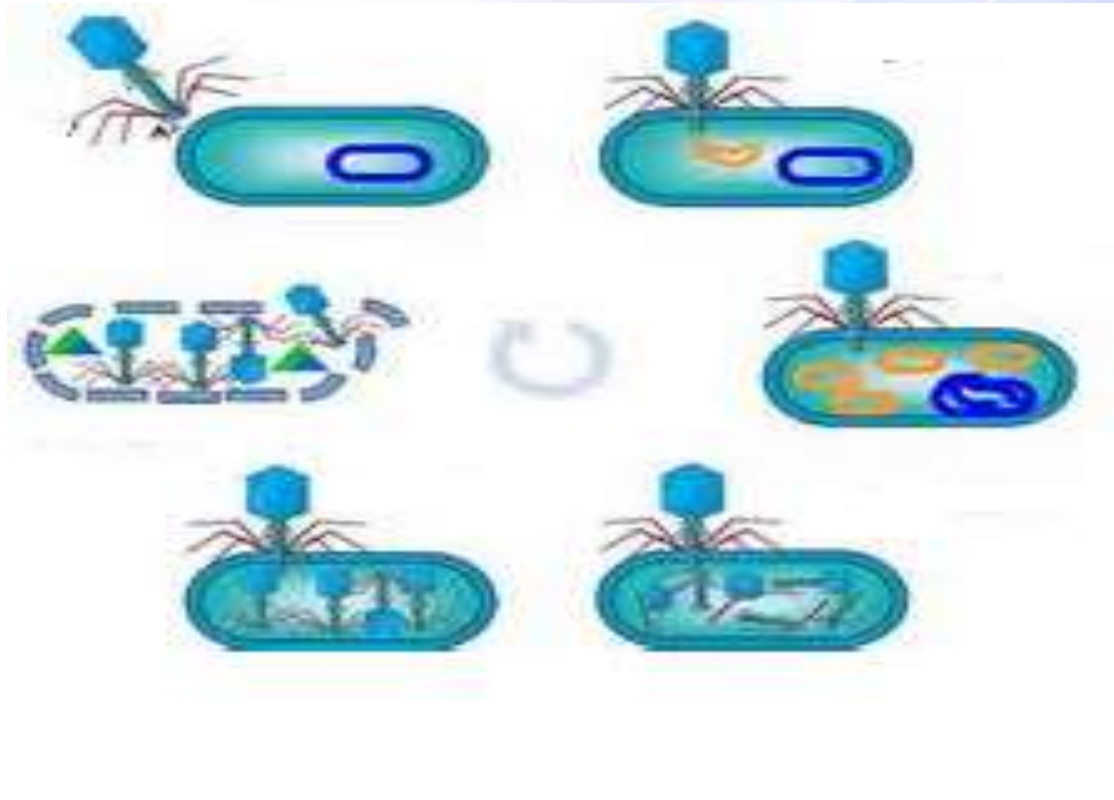
Hershey & Chase experiment

Bacteriophage



Hershey & Chase experiment

Life Cycle of a Bacteriophage



Hershey & Chase experiment

- They used radioactive isotopes of phosphorus and sulfur.
- DNA contains P but not S; so Hershey and Chase used ^{32}P to follow phage DNA during reproduction. Protein contains sulfur but not phosphorus; so they used ^{35}S to follow the protein.

Hershey & Chase experiment

- Hershey and Chase first grew *E. coli* in a medium containing ^{32}P and infected the bacteria with T2 so that all the new phages would have DNA labeled with ^{32}P .

Hershey & Chase experiment

- They grew a second batch of *E. coli* in a medium containing ^{35}S and infected these bacteria with T2 so that all these new phages would have protein labelled with ^{35}S .

Hershey & Chase experiment

- Hershey and Chase then infected separate batches of unlabeled *E. coli* with the ^{35}S - and ^{32}P -labeled phages.
- Then they placed the *E. coli* cells in a blender and sheared off the empty protein coats from the cell walls.

Hershey & Chase experiment

- They separated out the protein coats and cultured the infected bacterial cells.
- Eventually, the cells burst and new phage particles emerged.

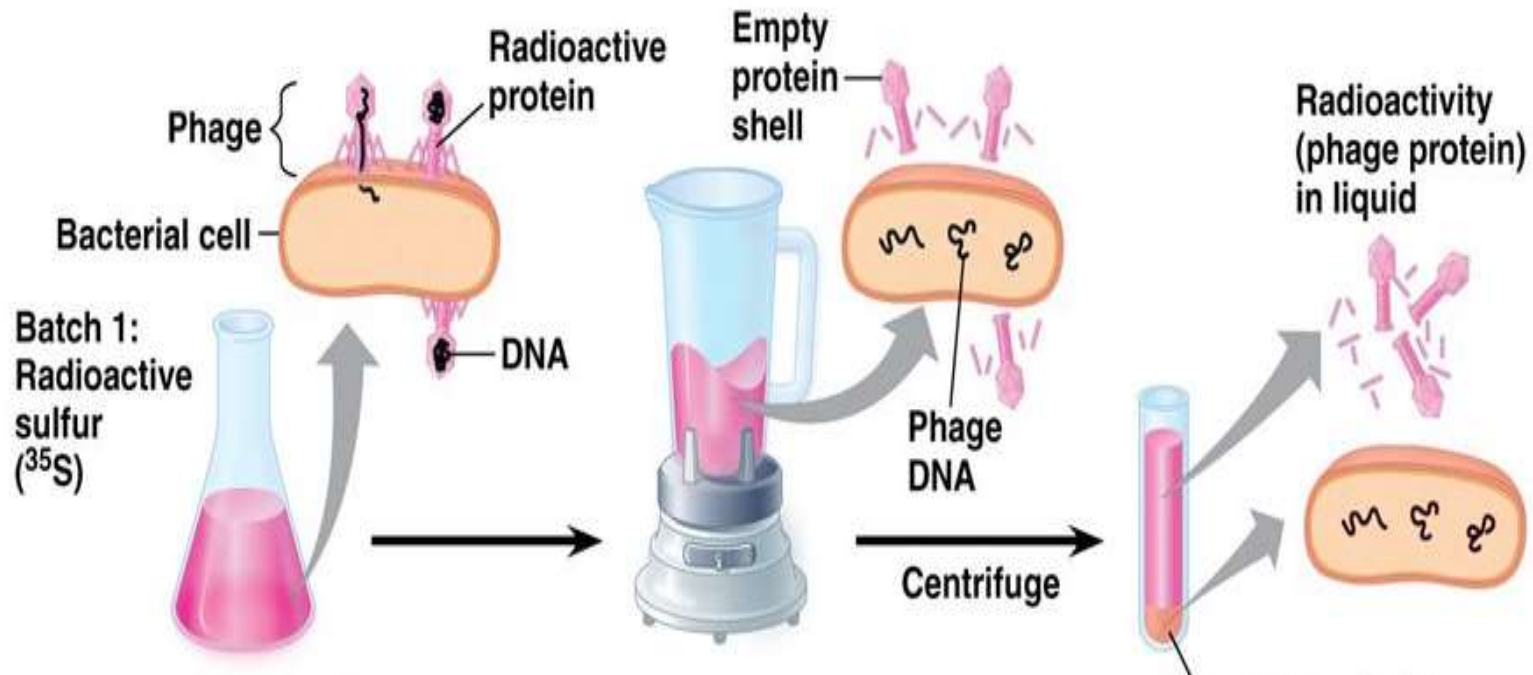
Hershey & Chase experiment

- When phages labeled with ^{35}S infected the bacteria, most of the radioactivity separated with the protein coats.
- When new phages emerged from the cell, they contained almost no radioactivity.
- So?

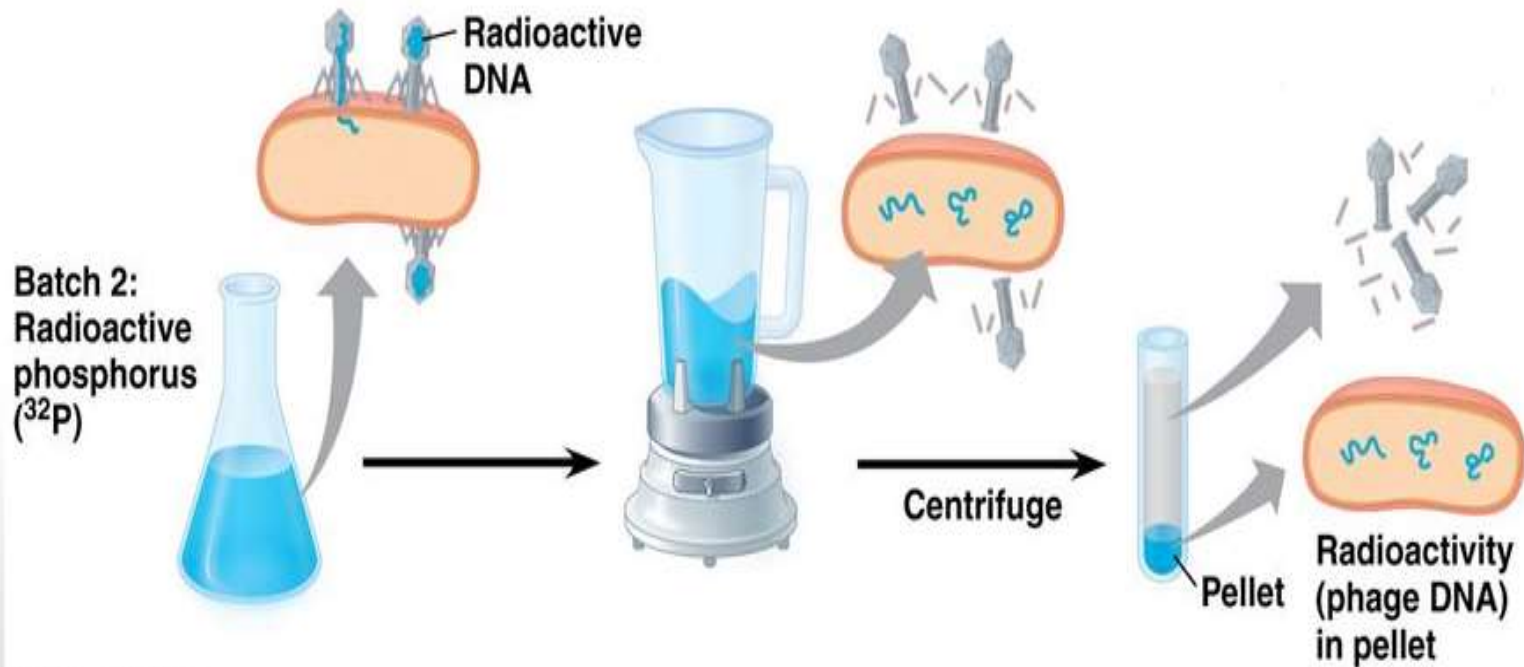
Hershey & Chase experiment

- When phages labelled with ^{32}P infected the bacteria, and removed the protein coat, radioactivity was present in the cells.
- When new phages emerged from the cell, they were also radioactive.
- So?

Hershey & Chase experiment



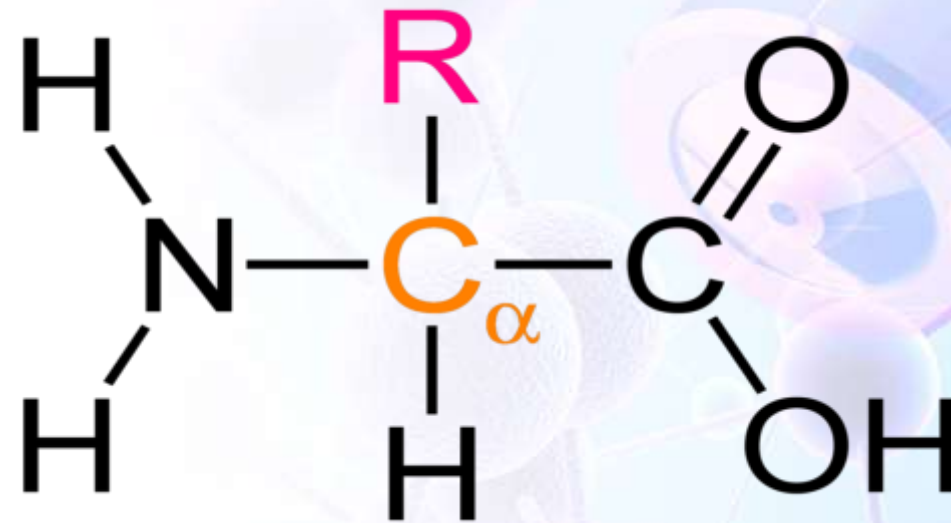
Hershey & Chase experiment



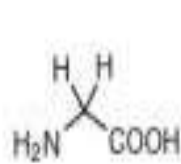
Chemical composition of proteins

- Proteins are polymers of amino acids.
- They range in size from small to very large.
- All the proteins are made up of Twenty different types of amino acids. So these amino acids are called standard amino acids.

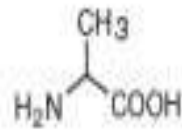
Chemical composition of proteins



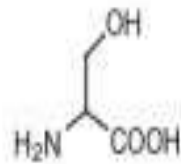
Chemical composition of proteins



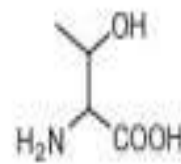
Glycine (Gly, G)



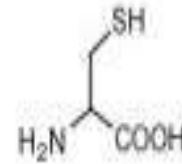
Alanine (Ala, A)



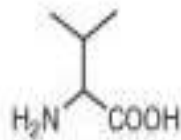
Serine (Ser, S)



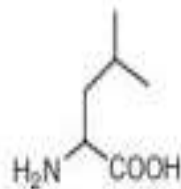
Threonine (Thr, T)



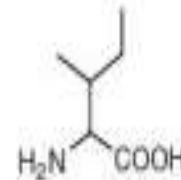
Cysteine (Cys, C)



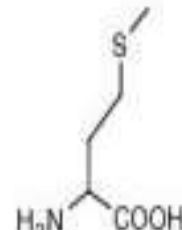
Valine (Val, V)



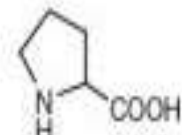
Leucine (Leu, L)



Isoleucine (Ile, I)

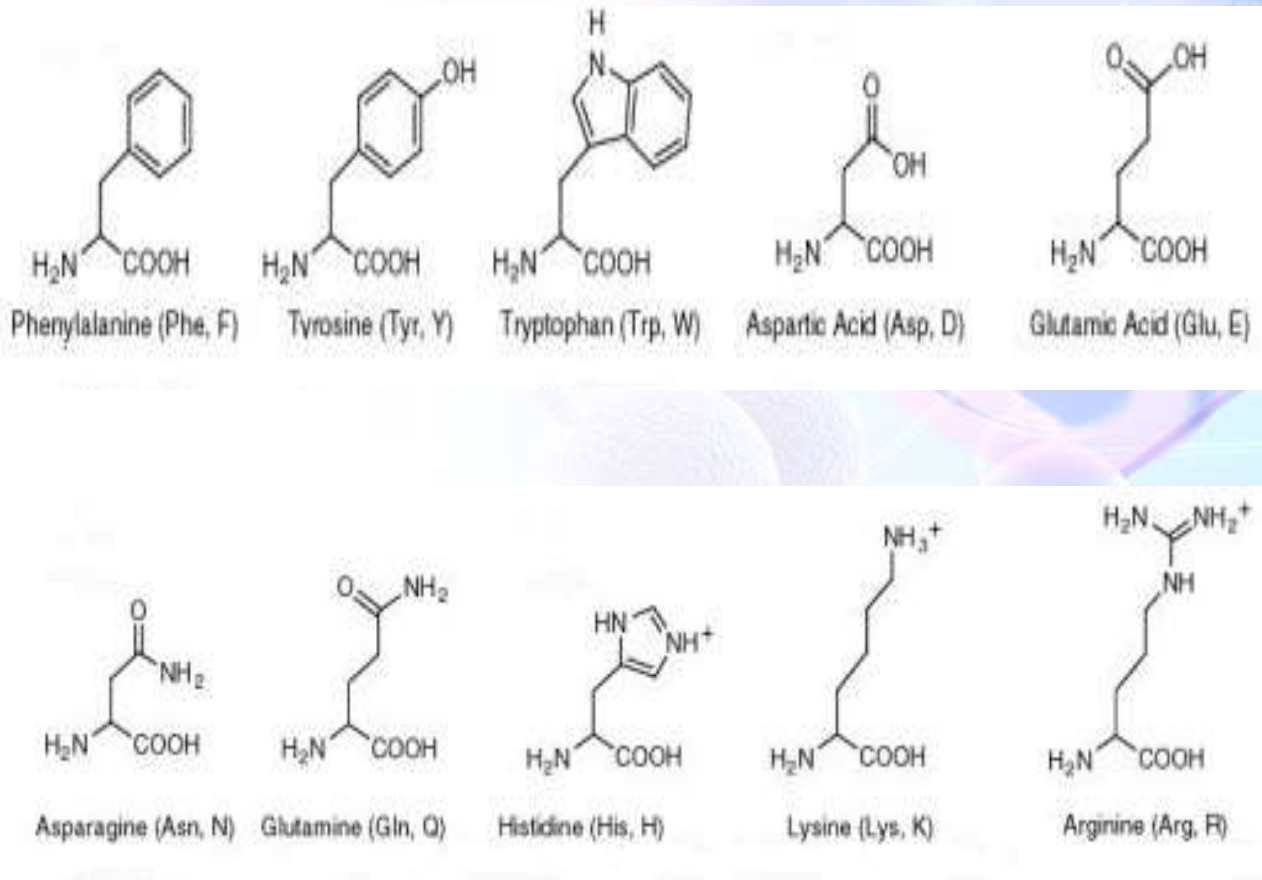


Methionine (Met, M)



Proline (Pro, P)

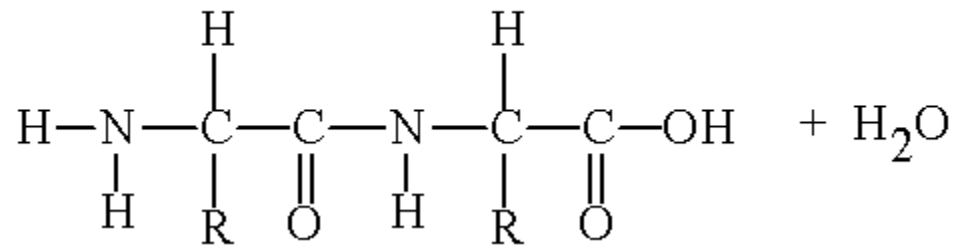
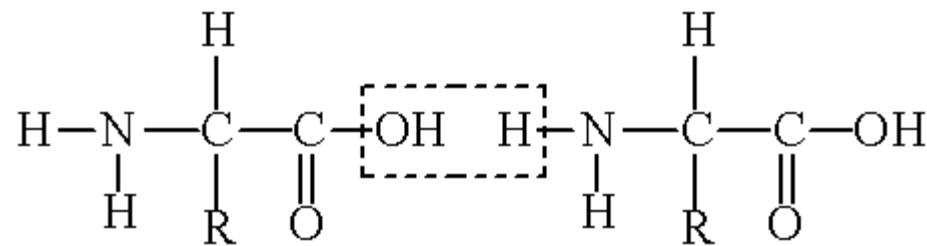
Chemical composition of proteins



Chemical composition of proteins

- In a protein molecule, each amino acid residue is joined to its neighbour by a specific type of covalent bond which is called **Peptide Bond**.

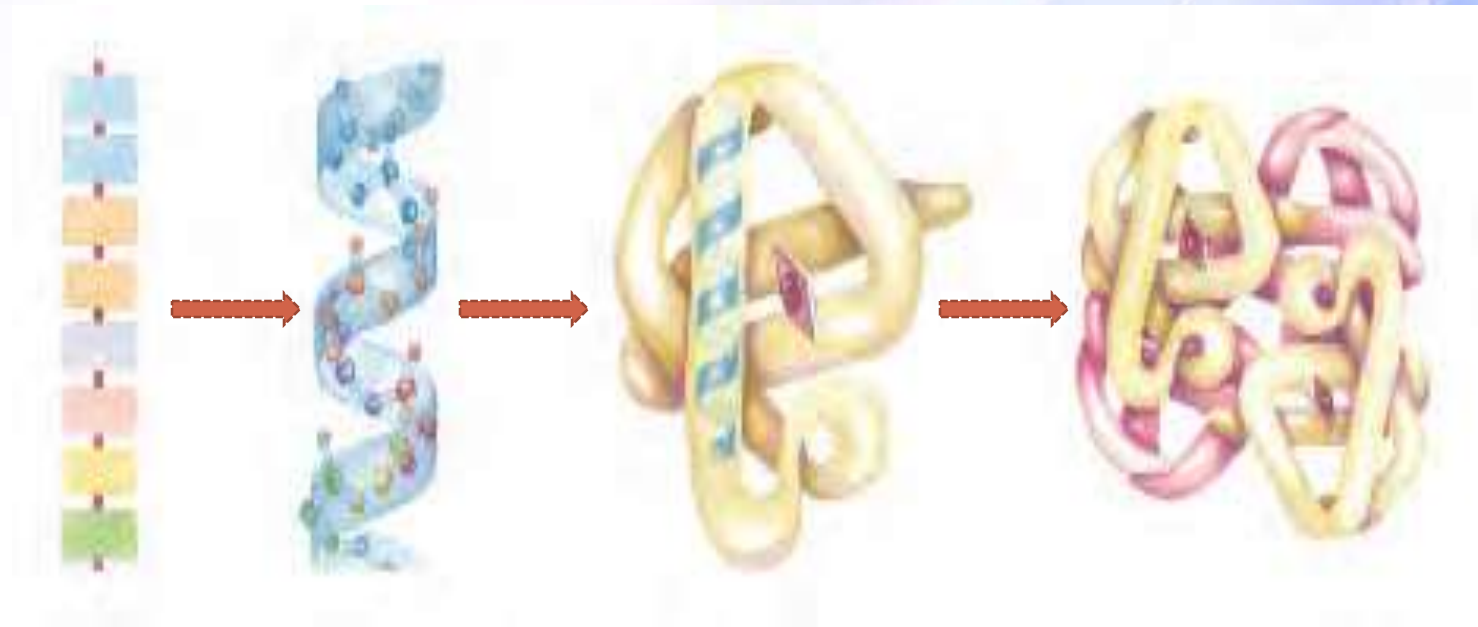
Chemical composition of proteins



Chemical composition of proteins

- Amino acids can successively join to form dipeptides, tripeptides, tetrapeptides, oligopeptides and polypeptides.

Chemical composition of proteins



Primary structure of proteins

- Primary structure or **covalent structure** of protein refers to the amino acid sequence of its polypeptide chain.
- Each type of protein has a unique amino acid sequence.

Primary structure of proteins

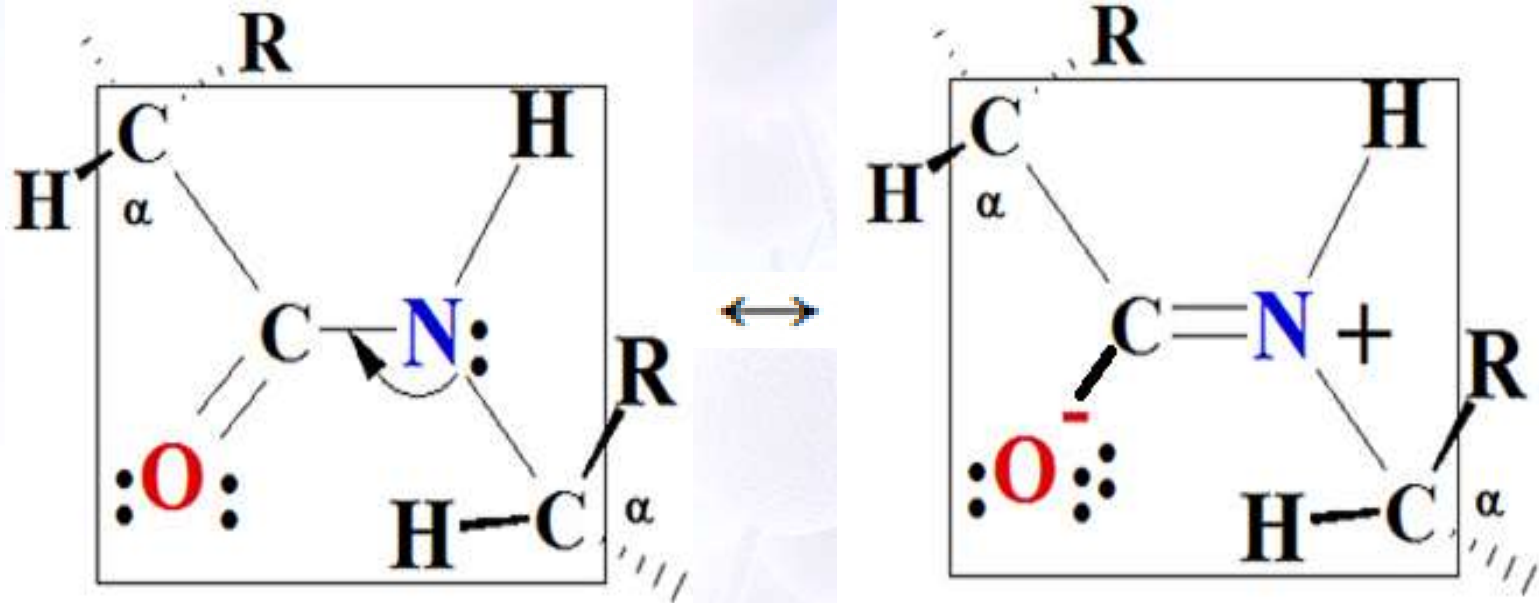
Peptide Bond Is Rigid and Planar

- Linus Pauling and Robert Corey carefully analyzed the peptide bond.
- Their findings laid the foundation for our present understanding of protein structure.

Primary structure of proteins

- They demonstrated that the peptide C - N bond is somewhat shorter than the C - N bond in a simple amine.

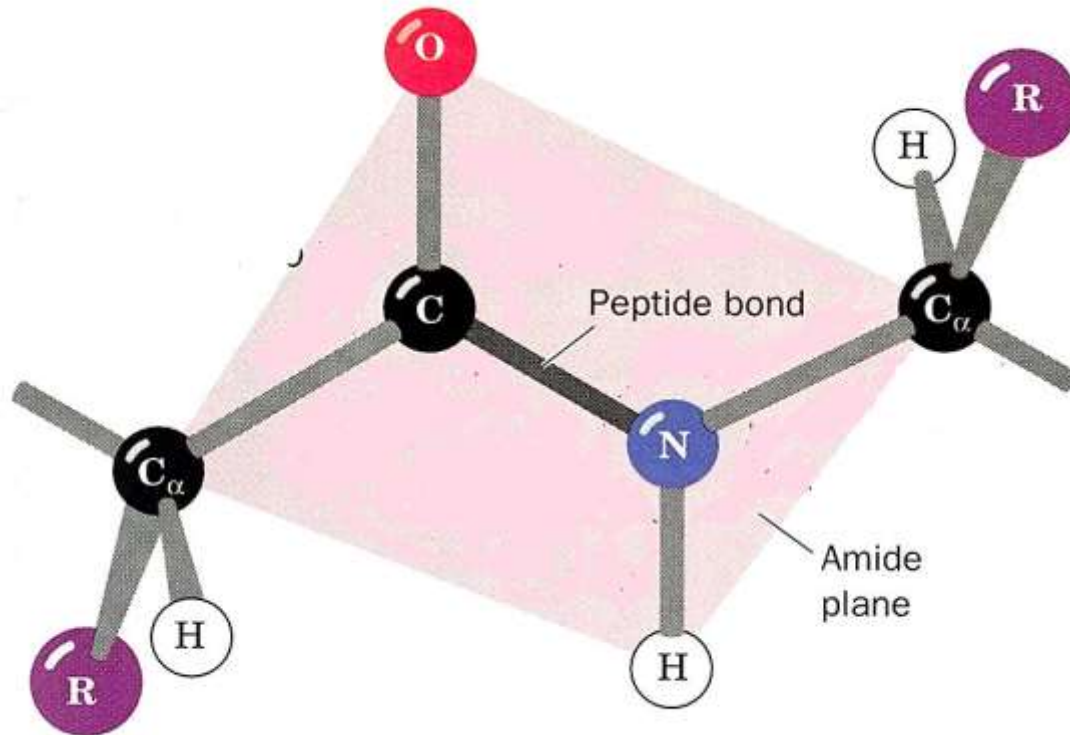
Primary structure of proteins



Primary structure of proteins

- The six atoms of the peptide group are co-planar i.e., lie in a single plane, with the oxygen atom of the carbonyl group and the hydrogen atom of the amide nitrogen trans to each other.

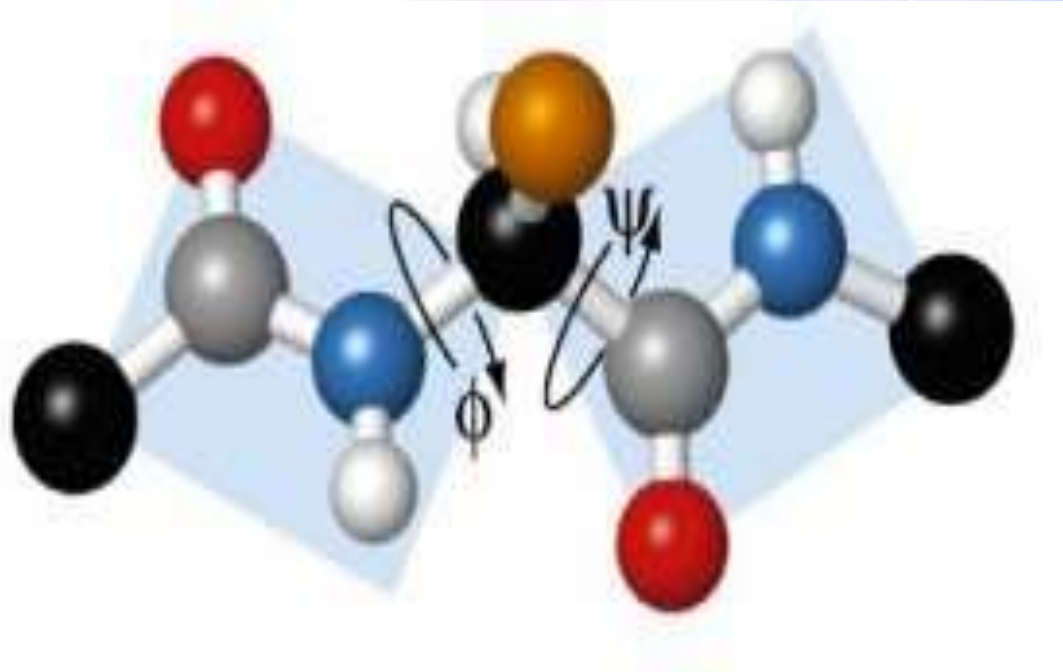
Primary structure of proteins



Primary structure of proteins

- Pauling and Corey concluded that the peptide C - N bonds are unable to rotate freely because of their partial double-bond character.
- Rotation is permitted about the N - α C and the α C - C bonds.

Primary structure of proteins



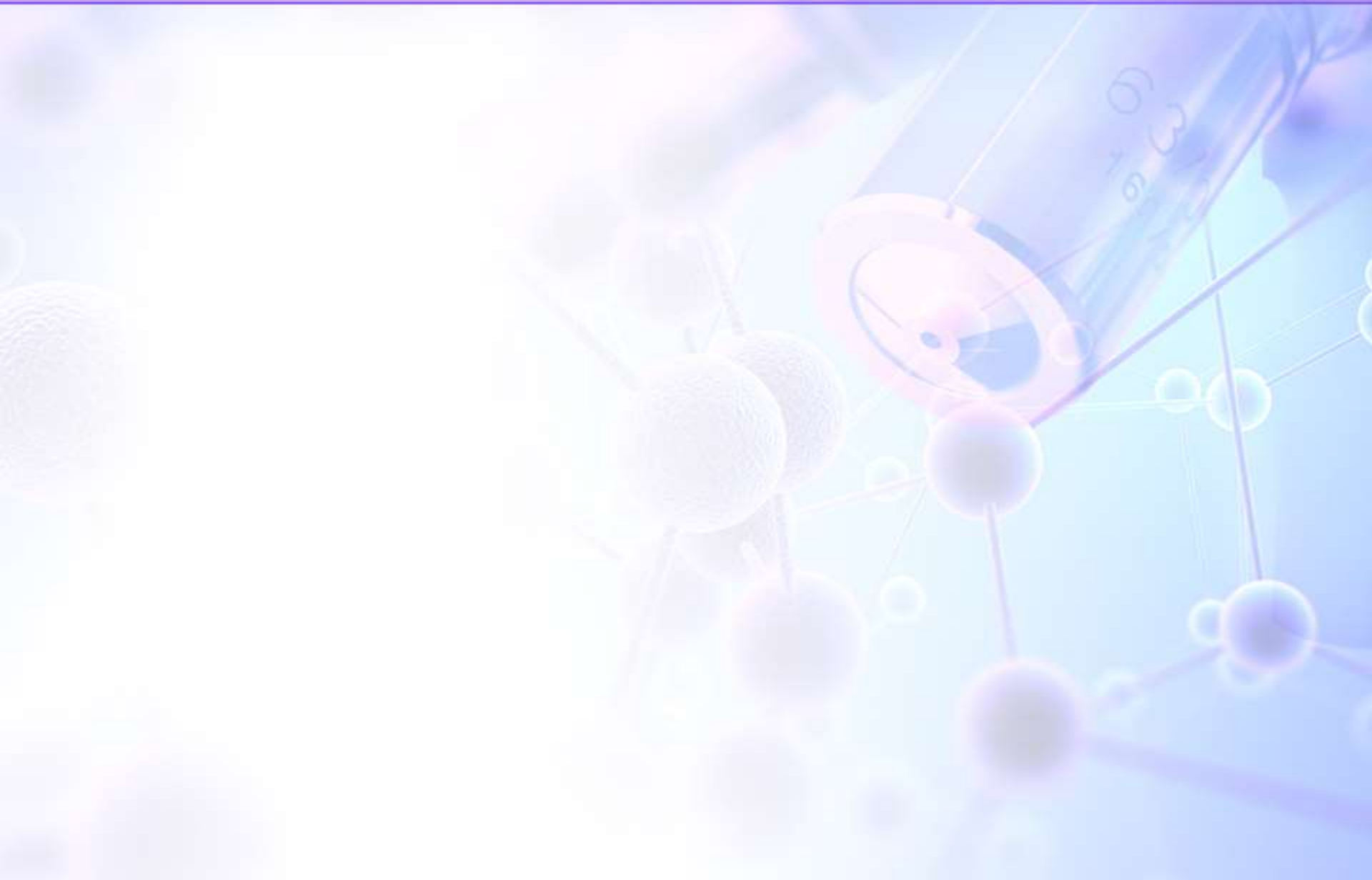
Primary structure of proteins

- The bond angles resulting from rotations at C are labelled ϕ (phi) for the N - α C bond and ψ (psi) for the α C - C bond.
- In principle, ϕ and ψ can have any value between +180 & -180.

Secondary structure of proteins

- Secondary structure of proteins refers to the local conformation of some part of a polypeptide.

Secondary structure of proteins



Secondary structure of proteins

- A few types of secondary structures are particularly stable and occur widely in proteins.

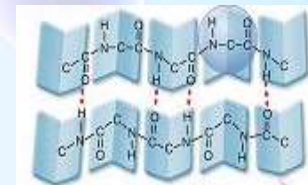
Secondary structure of proteins

- The most prominent are:-

- α -helix



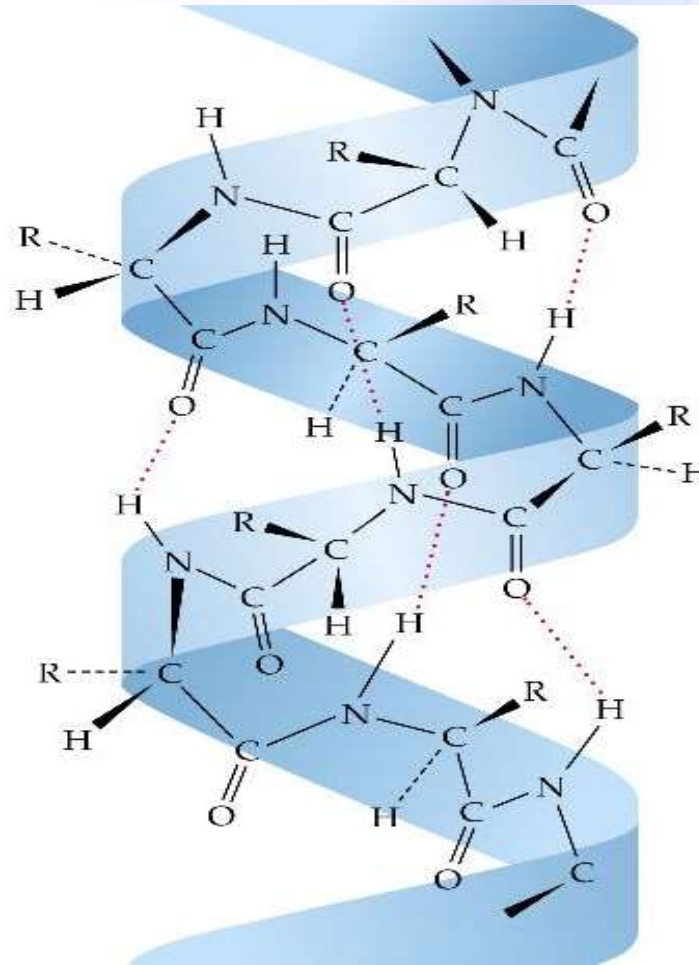
- β - conformations.



α - Helix

- The simplest arrangement which a polypeptide chain could assume with its rigid peptide bonds is a helical structure, which Pauling and Corey called the **α -helix**.

α - Helix



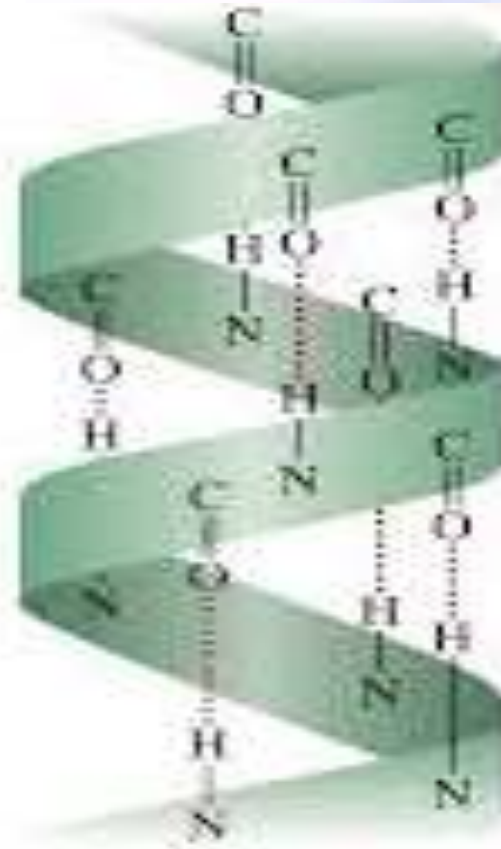
α - Helix

- The helical twist of the α -helix found in all proteins is right-handed.
- The repeating unit is a single turn of the helix, which extends about 5.4 Å (includes 3.6 amino acid residues) along the long axis.

α - Helix

- The amino acid residues in an helix have conformations with $\psi = -45$ to -50 and $\phi = -60$.
- An helix makes optimal use of internal hydrogen bonds.

α - Helix



α - Helix

- About one-fourth of all amino acid residues in polypeptides are found in α -helices while in some proteins it is the predominant structure.

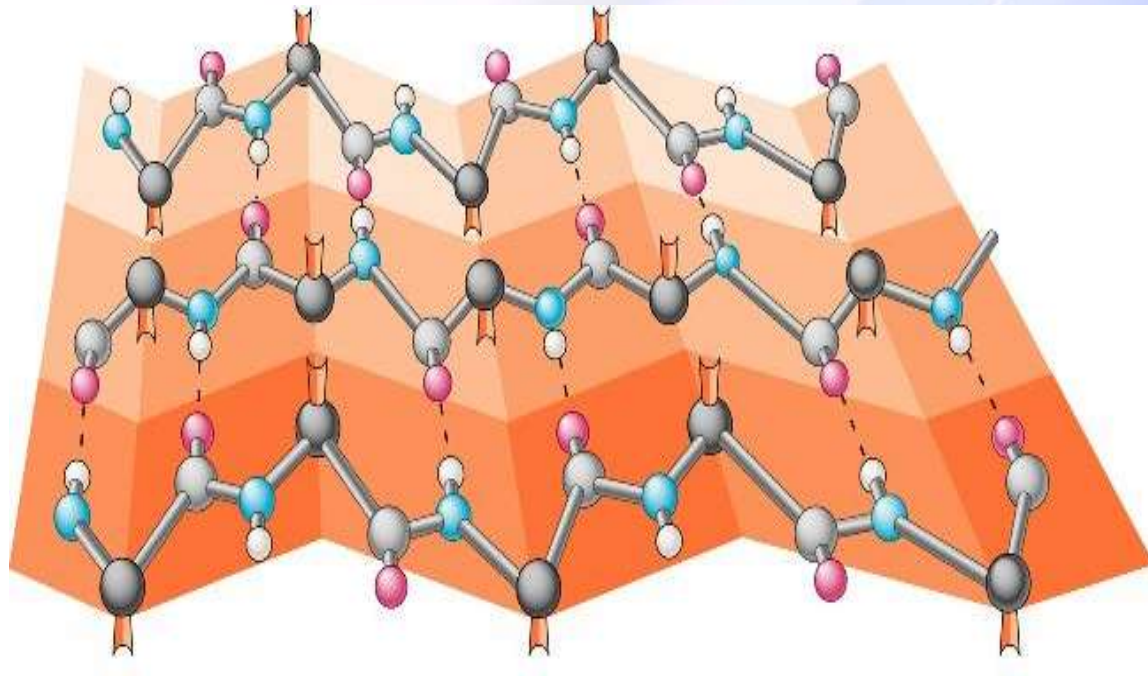
β - Pleated Sheets

- Pauling and Corey predicted a second type of secondary structure which they called **β -sheets**.
- This is a more extended conformation of polypeptide chains.

β - Pleated Sheets

- The backbone of the polypeptide chain is extended into a zigzag structure.
- The zigzag polypeptide chains are arranged side by side to form a structure resembling a series of pleats.

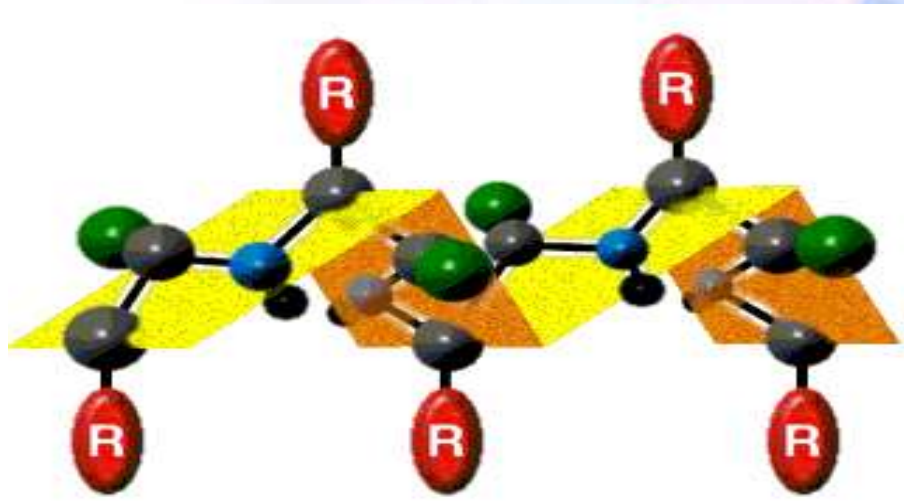
β - Pleated Sheets



β - Pleated Sheets

- The R groups of adjacent amino acids protrude from the zigzag structure in opposite directions.

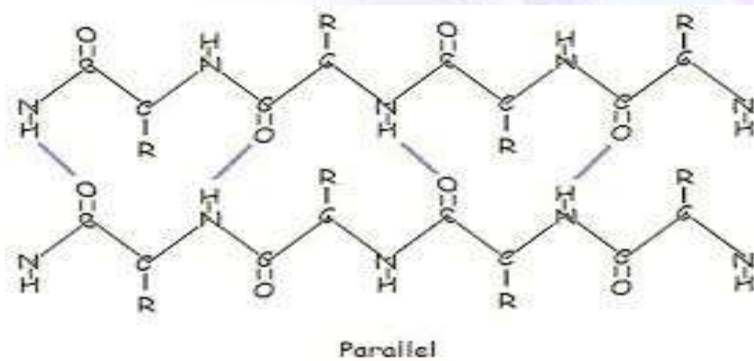
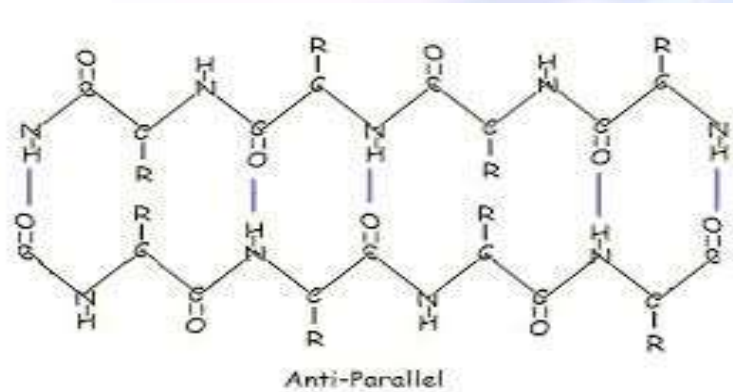
β - Pleated Sheets



β - Pleated Sheets

- Hydrogen bonds are formed between adjacent segments of polypeptide chain.
- The adjacent polypeptide chains in a sheet can be either parallel or antiparallel.

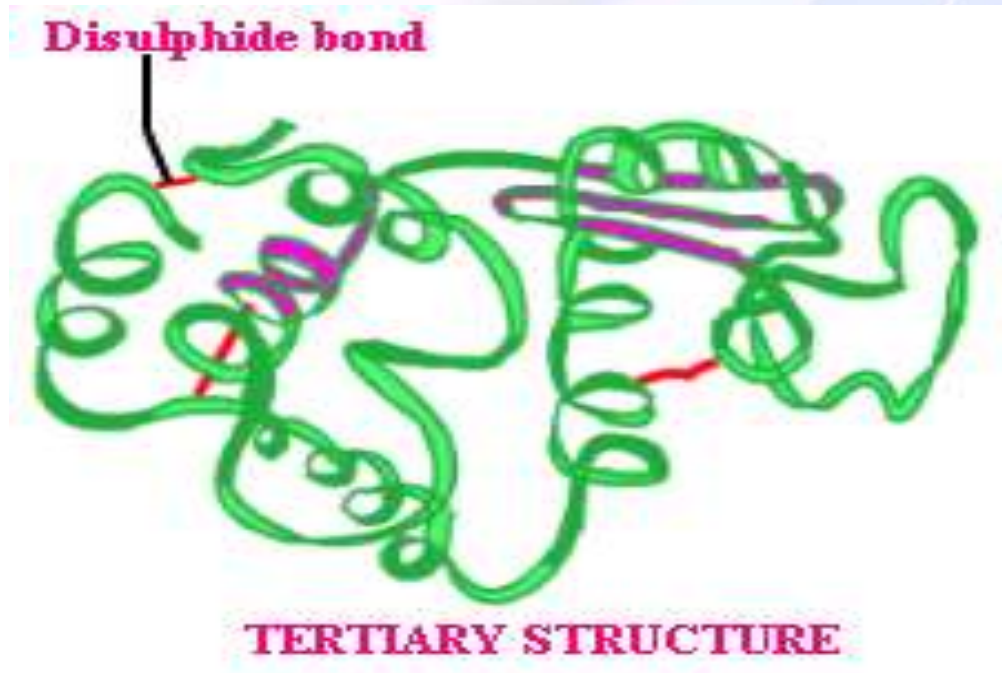
β - Pleated Sheets



Tertiary Structure of Proteins

- The overall three-dimensional arrangement of all atoms in a protein is referred to as the protein's **tertiary structure**.

Tertiary Structure of Proteins



Tertiary Structure of Proteins

- It includes longer-range aspects of amino acid sequence.
- Amino acids that are far apart in the polypeptide chain may interact within the completely folded structure of a protein.

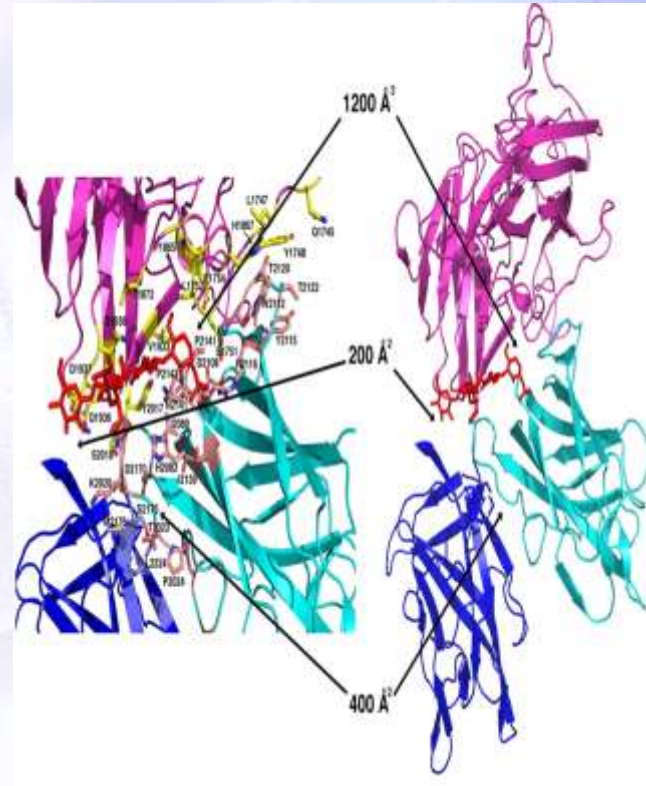
Tertiary Structure of Proteins

- Interacting segments of polypeptide chains are held in their characteristic tertiary positions by different kinds of weak interactions (and sometimes by covalent bonds) between the segments.

Tertiary Structure of Proteins

- Large polypeptide chains usually fold into two or more globular clusters known as **domains**, which often give these proteins a bi- or multilobal appearance.

A 3D ribbon diagram of the 19S proteasome structure. The structure is composed of three main subunits: a blue subunit at the top, an orange subunit in the middle, and a green subunit at the bottom. The orange subunit is the largest and most complex, featuring numerous alpha-helices and beta-sheets. The blue and green subunits are smaller and more compact, with fewer helices and more beta-strands. The entire structure is shown against a white background.



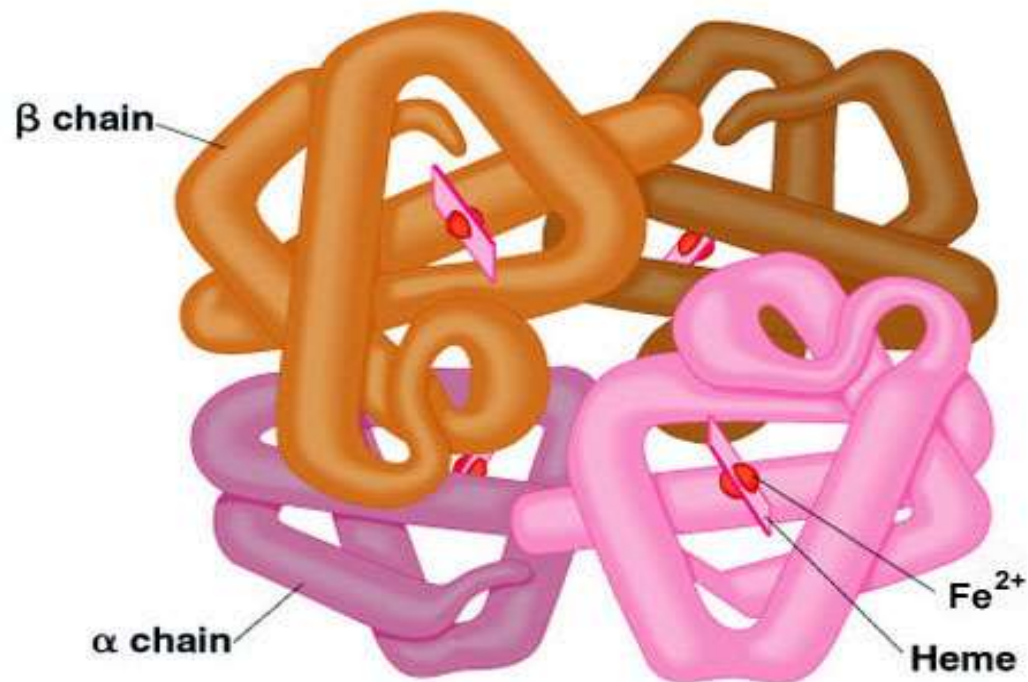
Quaternary Structure of Proteins

- Some proteins contain two or more separate polypeptide chains or subunits, which may be identical or different.
- The spatial arrangement of these subunits is known as a protein's quaternary structure.

Quaternary Structure of Proteins

- A multi-subunit protein is also referred to as a **multimer**.
- A multimer with just a few subunits is called as **oligomer** and a single subunit or a group of subunits, is called a **protomer**.

Quaternary Structure of Proteins



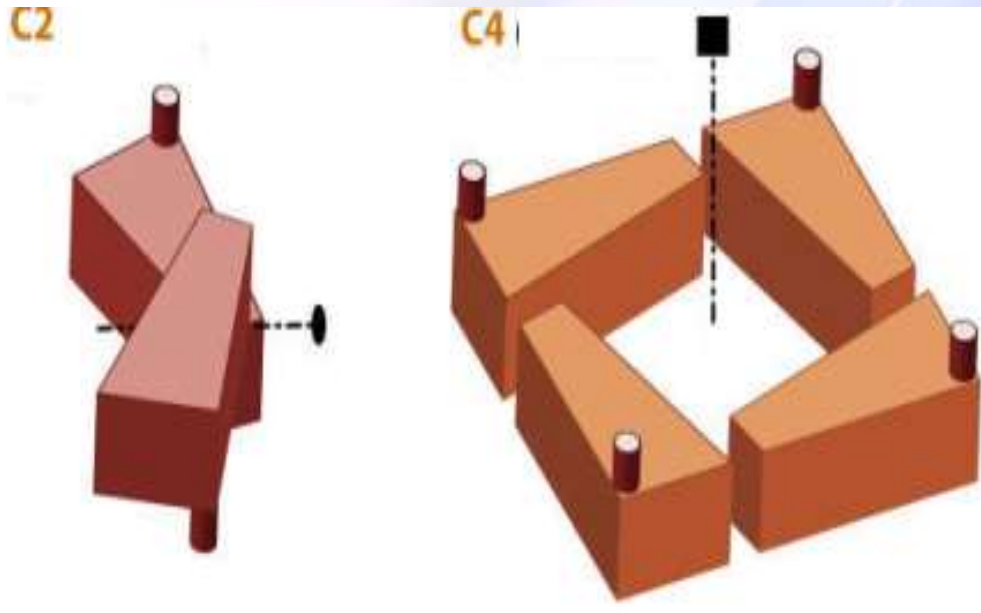
Quaternary Structure of Proteins

- Identical subunits of multimeric proteins are generally arranged in a symmetric patterns.
- Oligomers can have either rotational symmetry or helical symmetry.

Quaternary Structure of Proteins

- There are several forms of rotational symmetry. The simplest is **cyclic symmetry**, involving rotation about a single axis.

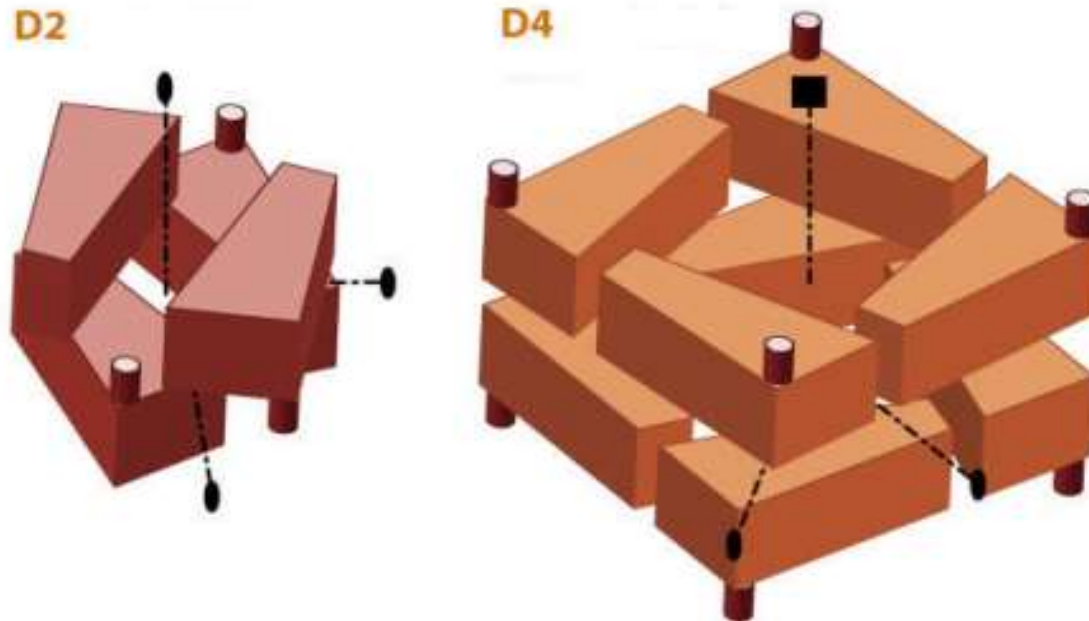
Quaternary Structure of Proteins



Quaternary Structure of Proteins

- A somewhat more complicated rotational symmetry is **dihedral symmetry**, in which a twofold rotational axis is present.

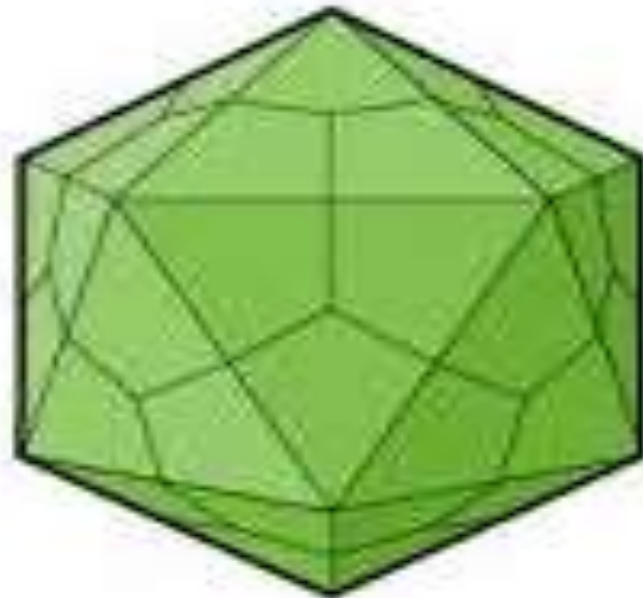
Quaternary Structure of Proteins



Quaternary Structure of Proteins

- More complex rotational symmetries include icosahedral symmetry.
- An icosahedron is a regular 12-cornered polyhedron having 20 triangular faces.

Quaternary Structure of Proteins



Quaternary Structure of Proteins

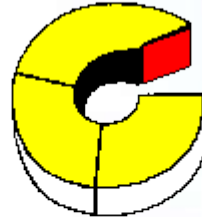
- The other major type of symmetry found in oligomers is **helical symmetry**.

Quaternary Structure of Proteins

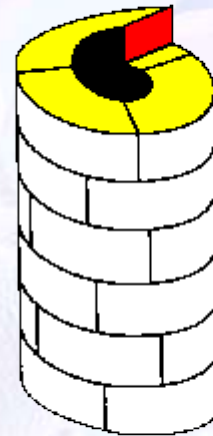
Helical Symmetry



subunit



helix segment



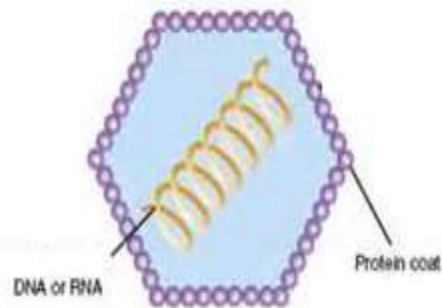
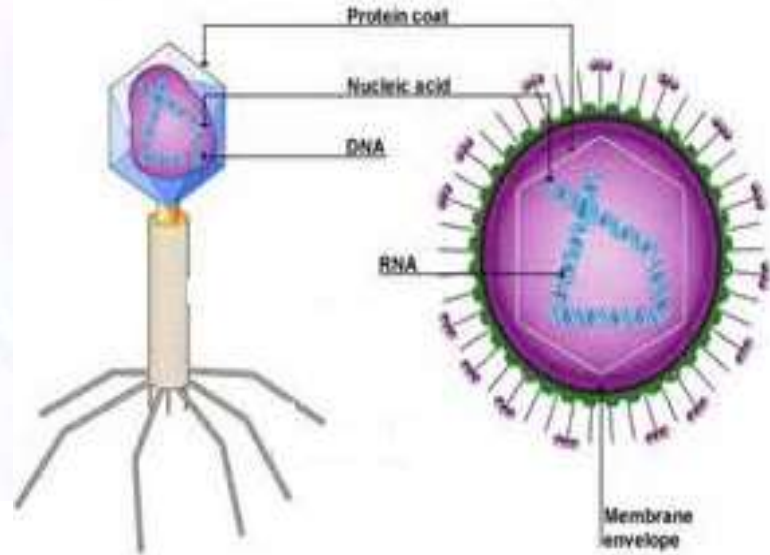
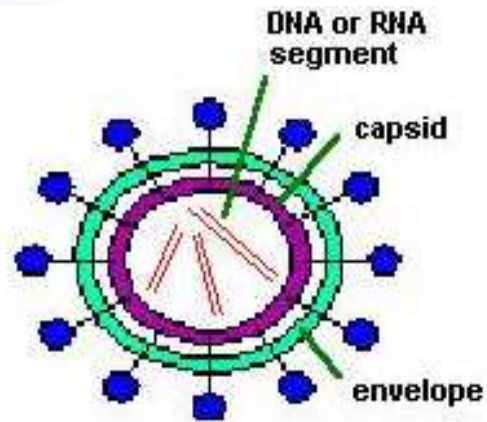
helix

(adapted from Voet and Voet, 1990)

Genetic Materials in Viruses

- Viruses are exceptionally simple and extremely small microorganisms.
- They have a very simple structural organization consisting of a molecule of nucleic acid and a protein coat.

Genetic Materials in Viruses



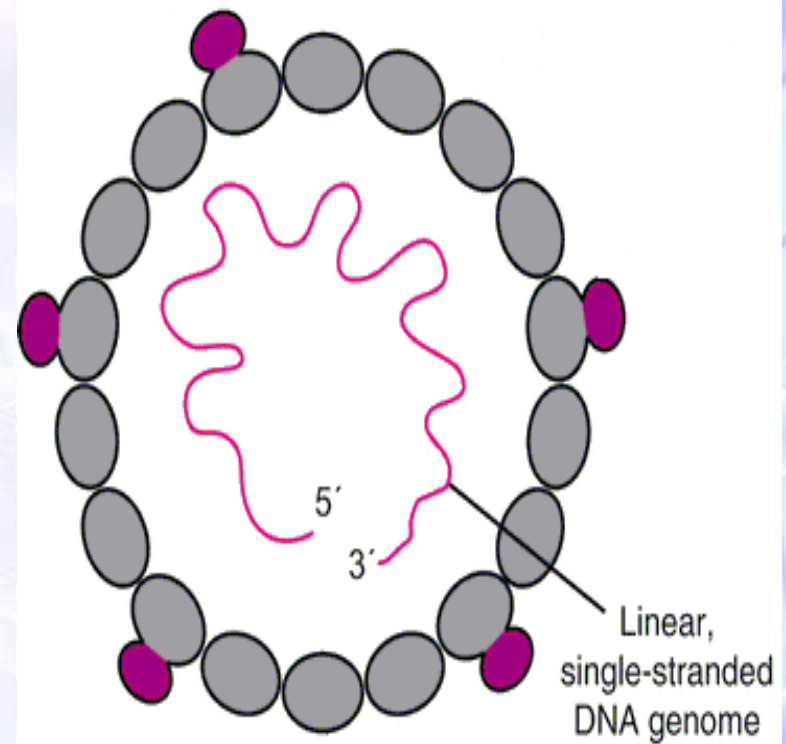
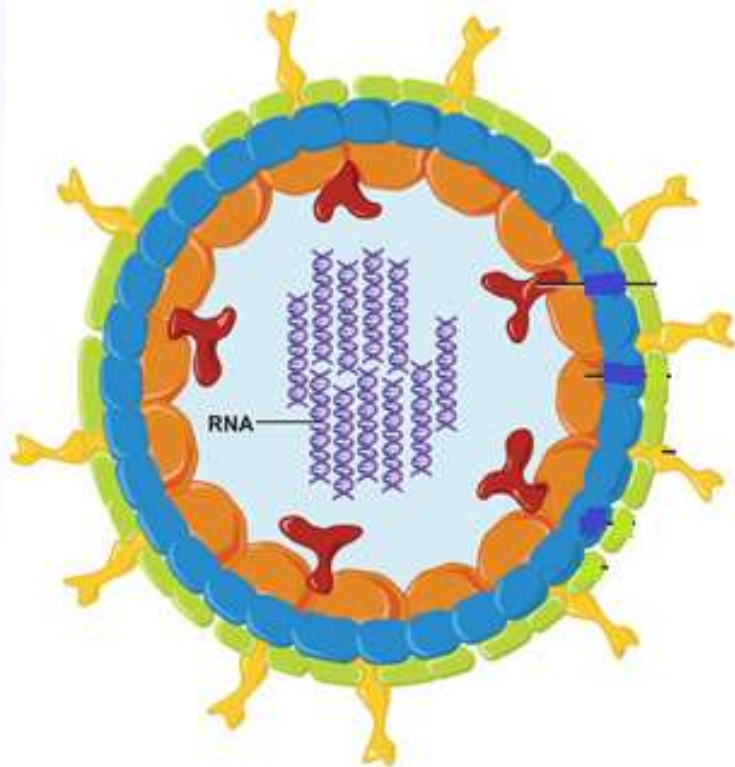
Genetic Materials in Viruses

- Viruses can have either DNA or RNA as a genetic material but never both.
- The nucleic acid of a virus can be single-stranded or double-stranded.

Genetic Materials in Viruses

- There are viruses with the familiar double-stranded DNA, with single-stranded DNA, with double-stranded RNA and with single-stranded RNA.

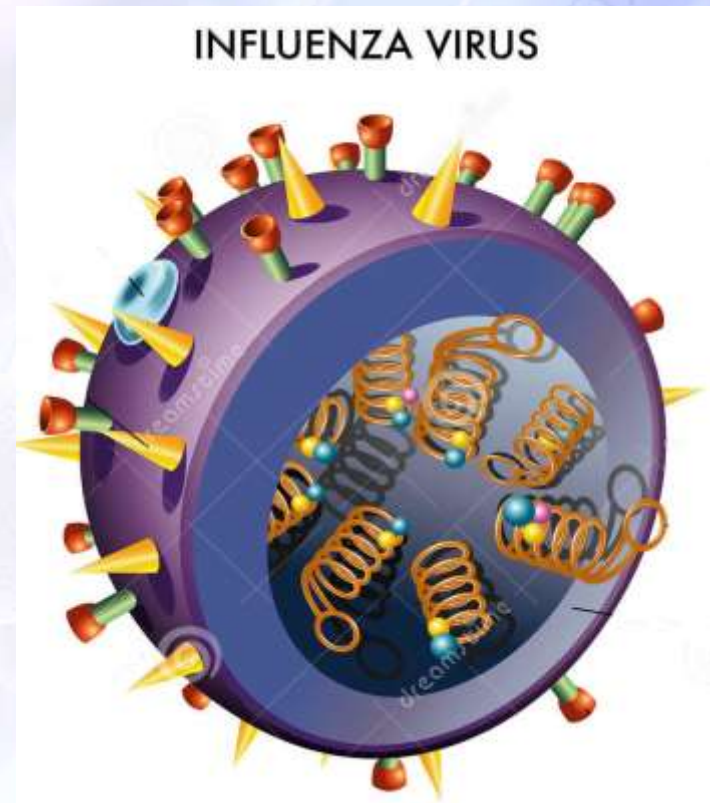
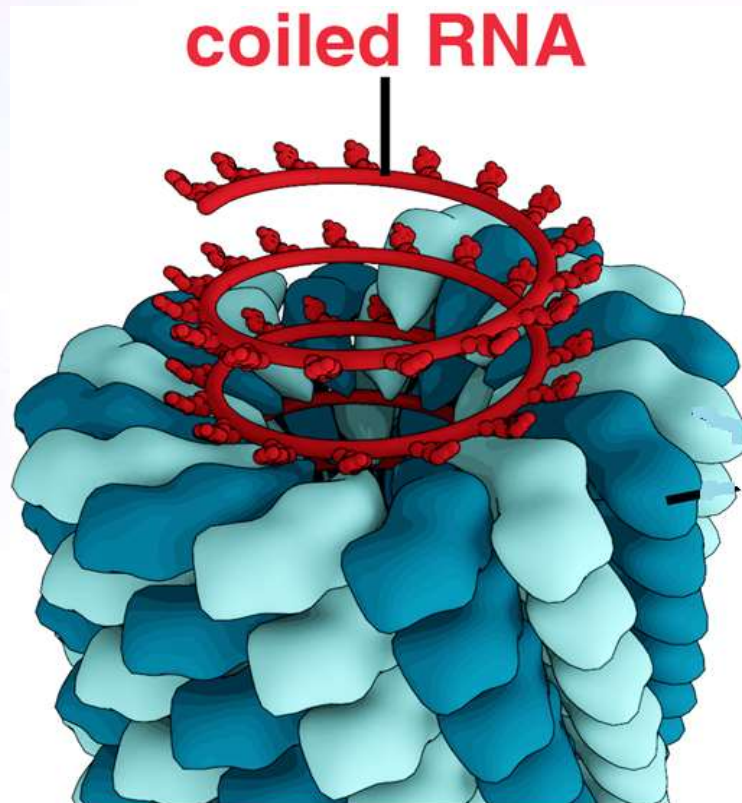
Genetic Materials in Viruses



Genetic Materials in Viruses

- The viral nucleic acid can be linear or circular.
- Some viruses may contain coiled RNA.
- In some viruses , the nucleic acid is in several separate segments.

Genetic Materials in Viruses



Genetic Materials in Viruses

- The percentage of nucleic acid in relation to protein is about 1% for the influenza virus and about 50% for some bacteriophages.

Genetic Materials in Viruses

- The total amount of nucleic acid varies from a few thousand nucleotides to as many as 250,000 nucleotides.
- *E. coli*'s chromosome consists of approx. 4 million nucleotide pairs.

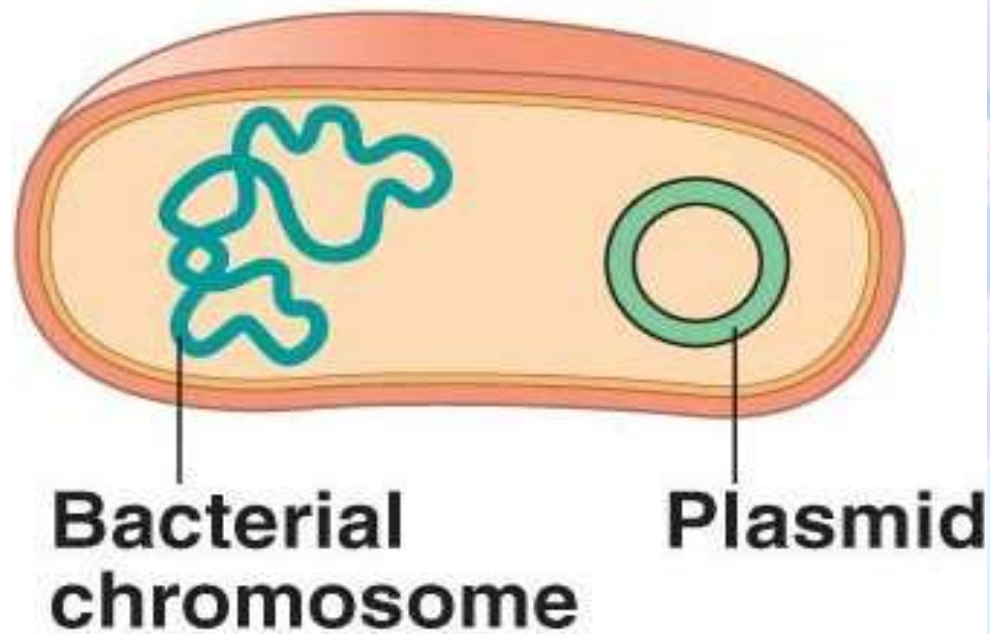
Organization of Genetic Material in Bacteria

- Bacteria typically have a single circular chromosome consisting of a single circular molecule of DNA with associated proteins.

Organization of Genetic Material in Bacteria

- The bacterial chromosome is a very long (up to 1mm).
- It is looped and folded and attached at one or several points to the plasma membrane.

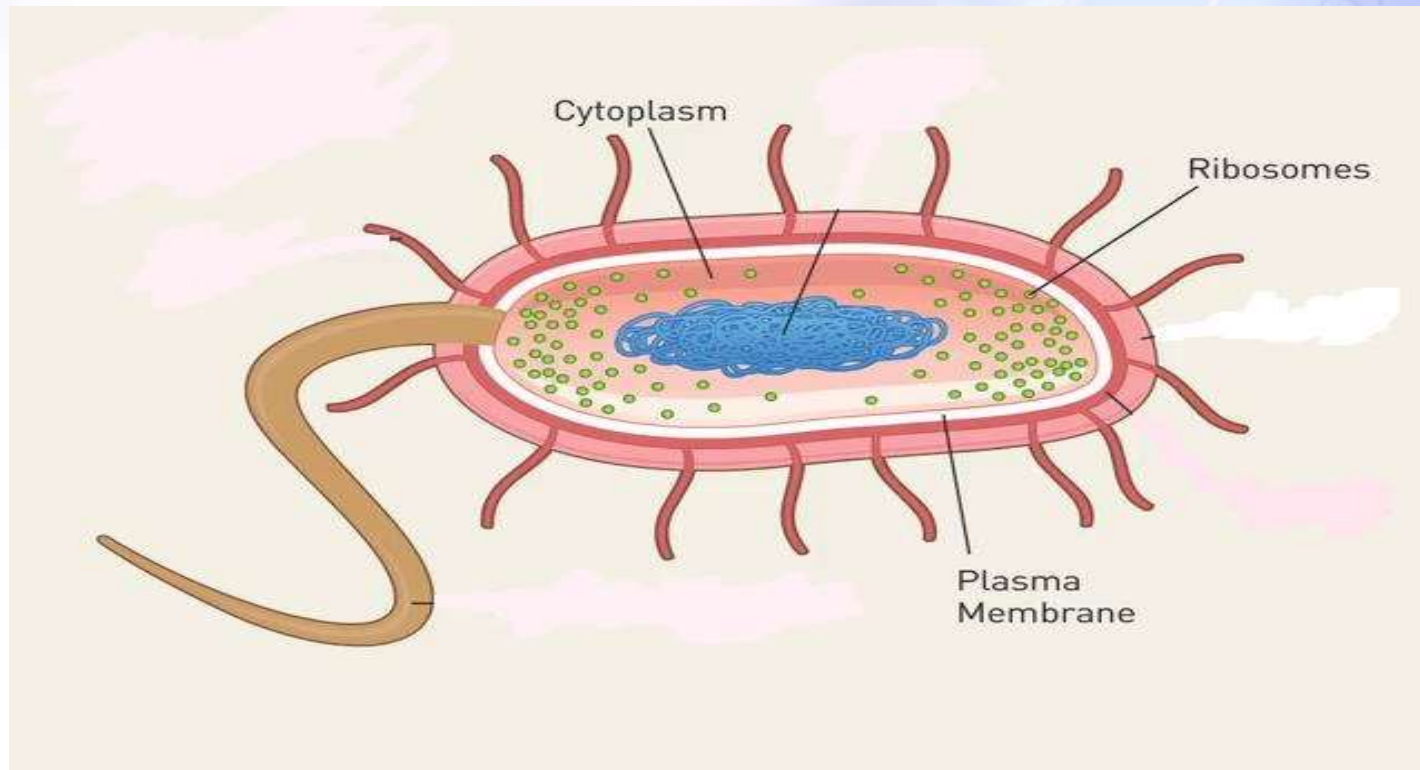
Organization of Genetic Material in Bacteria



Organization of Genetic Material in Bacteria

- Specific proteins interact with the bacterial DNA to form a highly condensed nucleoprotein complex called the **nucleoid**.

Organization of Genetic Material in Bacteria



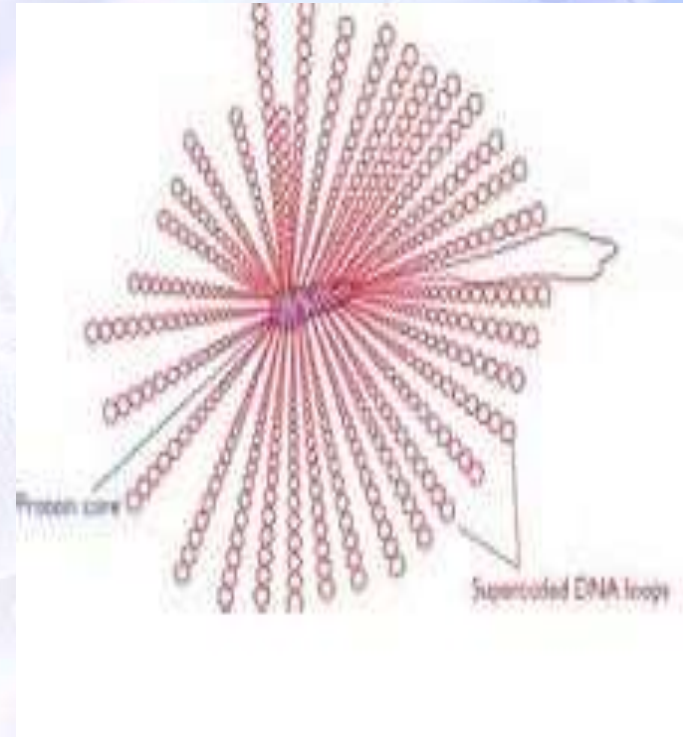
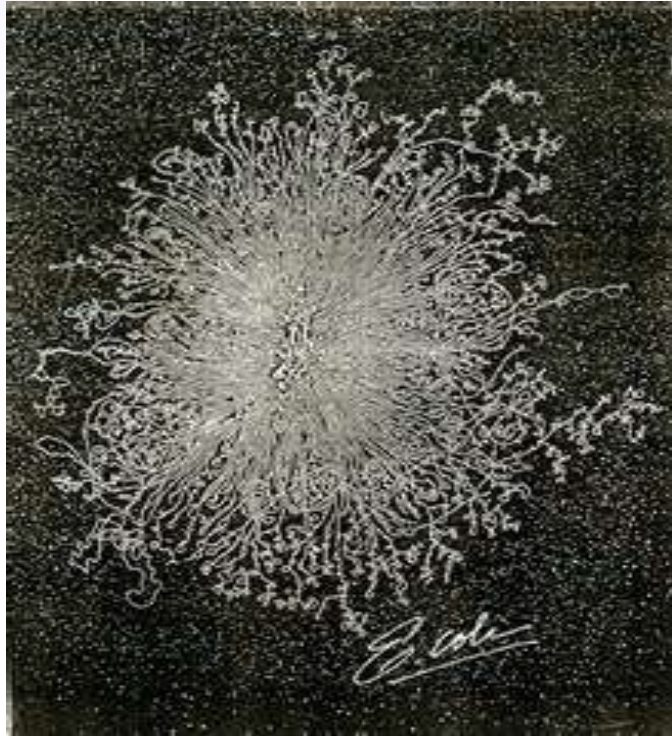
Organization of Genetic Material in Bacteria

- Bacterial chromatin can be released from the cell by gentle lysis of the cell.
- Electron micrograph of the chromatin reveals that it consists of multiple loops which emerge from a central region of the chromatin.

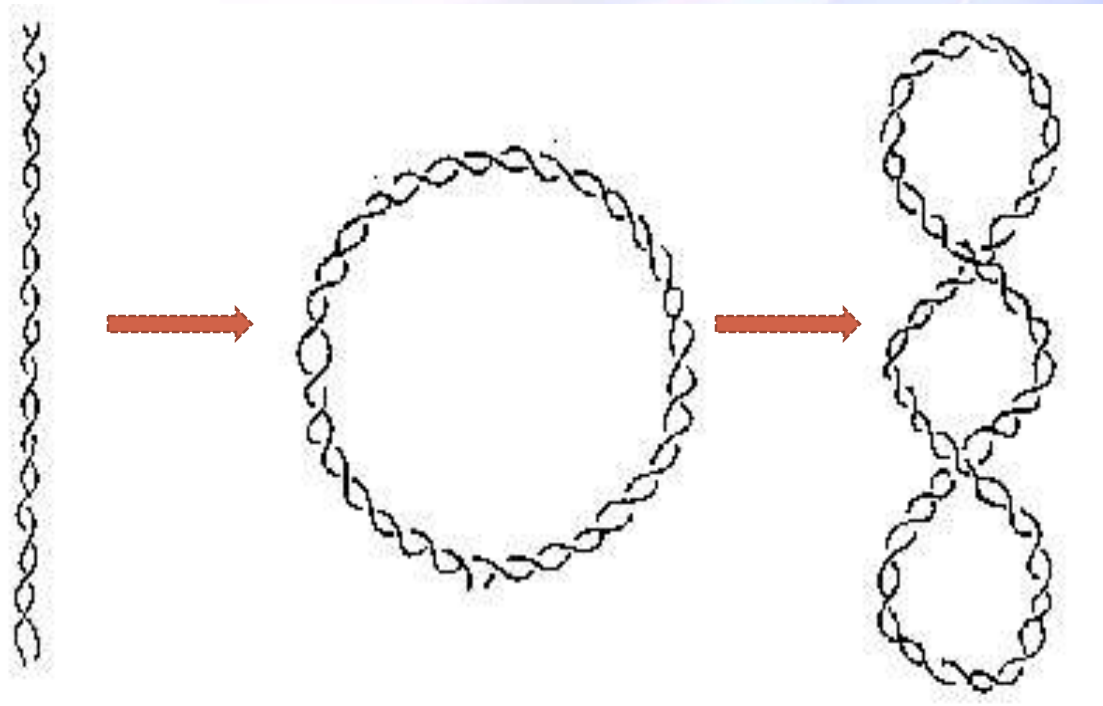
Organization of Genetic Material in Bacteria

- Some of the loops are super-coiled while some are relaxed.
- Relaxed loops are formed as a result of a nick introduced into super-coiled loops by a cellular DNase.

Organization of Genetic Material in Bacteria



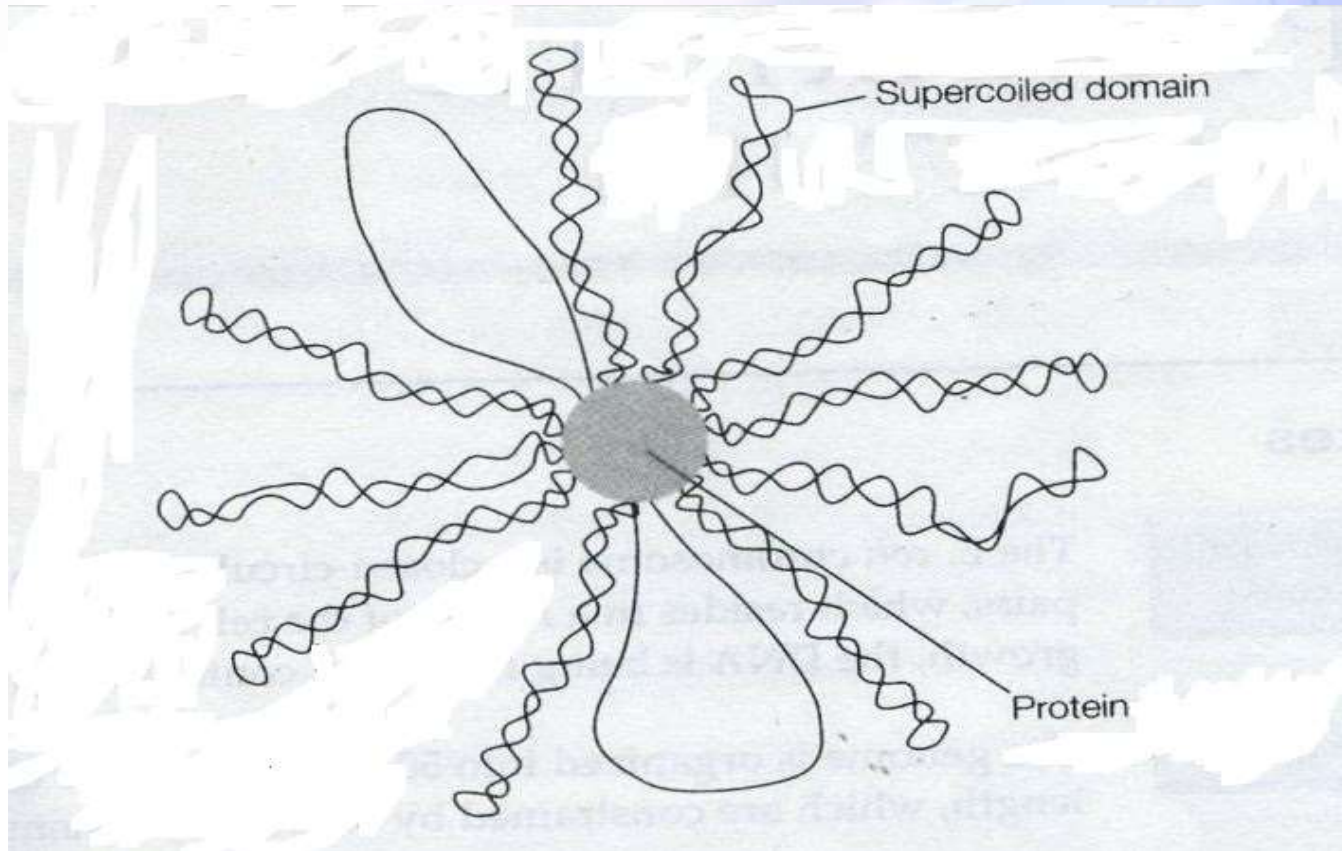
Organization of Genetic Material in Bacteria



Organization of Genetic Material in Bacteria

- If a super-coiled DNA molecule receives a nick, the strain of underwinding is immediately removed, and all the super-coiling is lost.
- Studies confirm that continued nuclease treatment increases number of relaxed loops.

Organization of Genetic Material in Bacteria



Organization of Genetic Material in Bacteria

- The bacterial DNA is arranged in super-coiled loops that are fastened to a central protein matrix, so that each loop is topologically independent from all the others.

Organization of Genetic Material in Bacteria

- So a nick that causes one super-coiled loop to relax would have no effect on other super-coiled loops.
- The super-coiled loops are dynamic structures which change during cell growth & division.

Organization of Genetic Material in Bacteria

- An *E. coli* chromosome is estimated to have about 400 super-coiled loops.
- Each loop has an average length of about 10-20 kbp.

Organization of Genetic Material in Bacteria

- The DNA compaction in a bacterial cell is contributed by supercoiling of loops, macromolecular crowding and DNA-binding proteins.

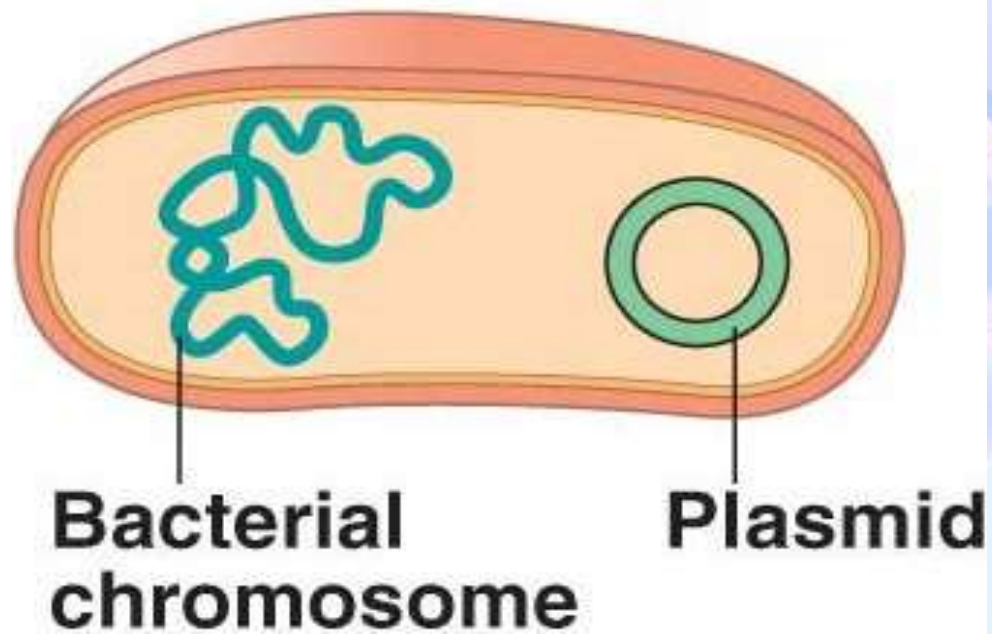
Organization of Genetic Material in Bacteria

- Bacteria typically have a single circular chromosome consisting of a single circular molecule of DNA with associated proteins.

Organization of Genetic Material in Bacteria

- The bacterial chromosome is a very long (up to 1mm).
- It is looped and folded and attached at one or several points to the plasma membrane.

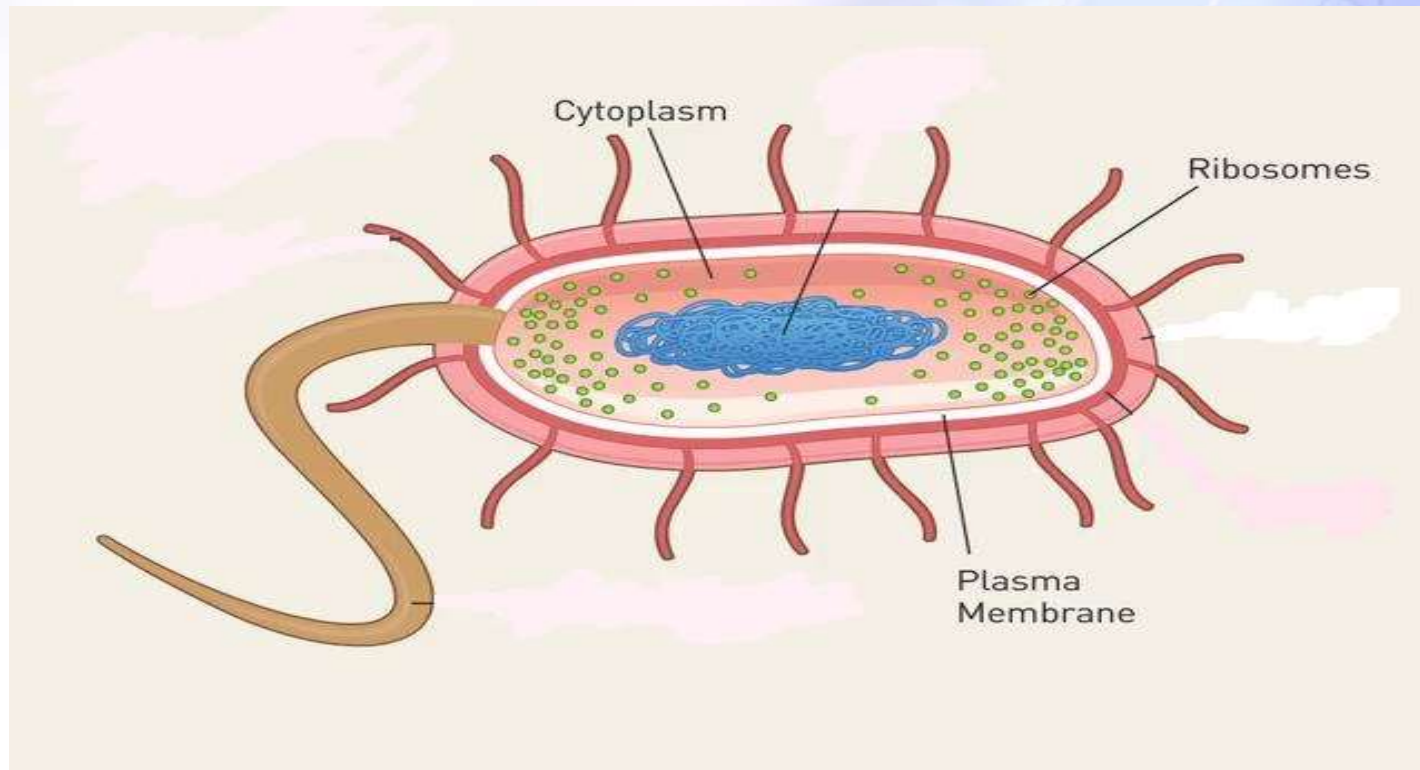
Organization of Genetic Material in Bacteria



Organization of Genetic Material in Bacteria

- Specific proteins interact with the bacterial DNA to form a highly condensed nucleoprotein complex called the **nucleoid**.

Organization of Genetic Material in Bacteria



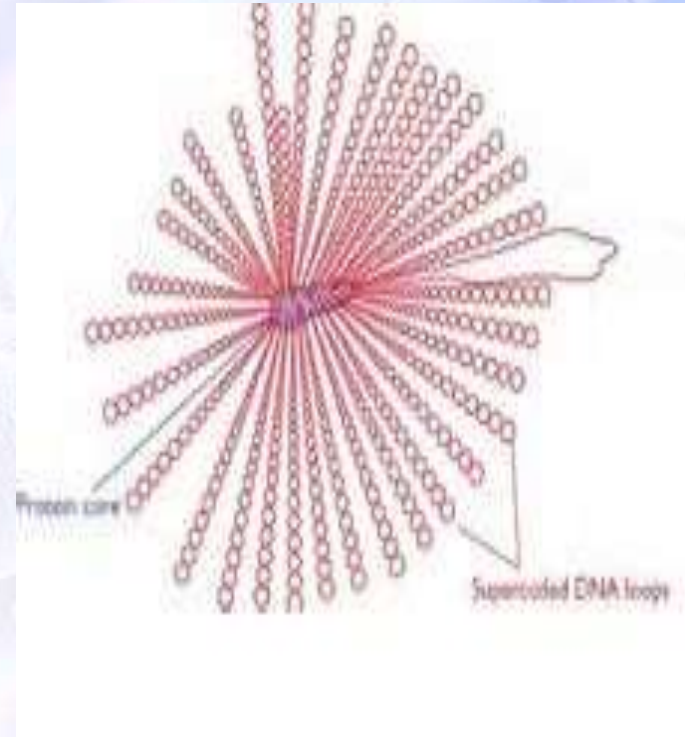
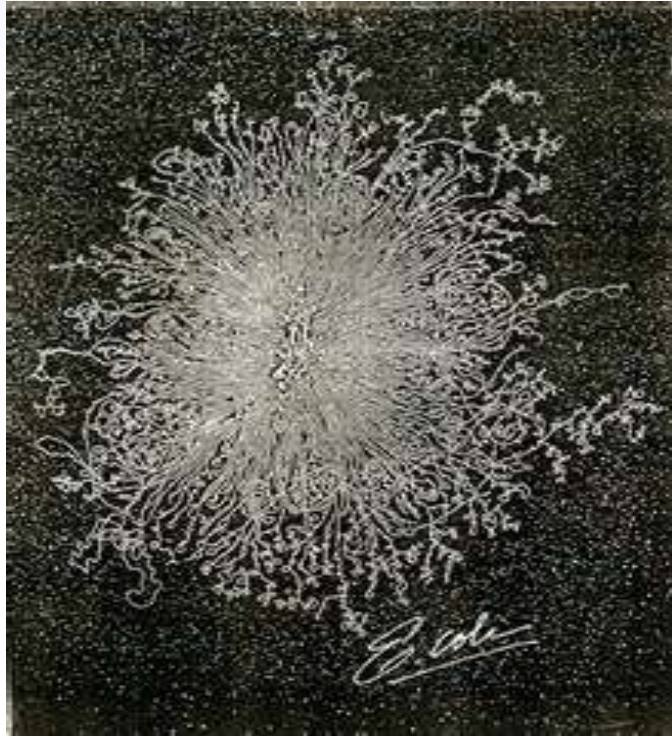
Organization of Genetic Material in Bacteria

- Bacterial chromatin can be released from the cell by gentle lysis of the cell.
- Electron micrograph of the chromatin reveals that it consists of multiple loops which emerge from a central region of the chromatin.

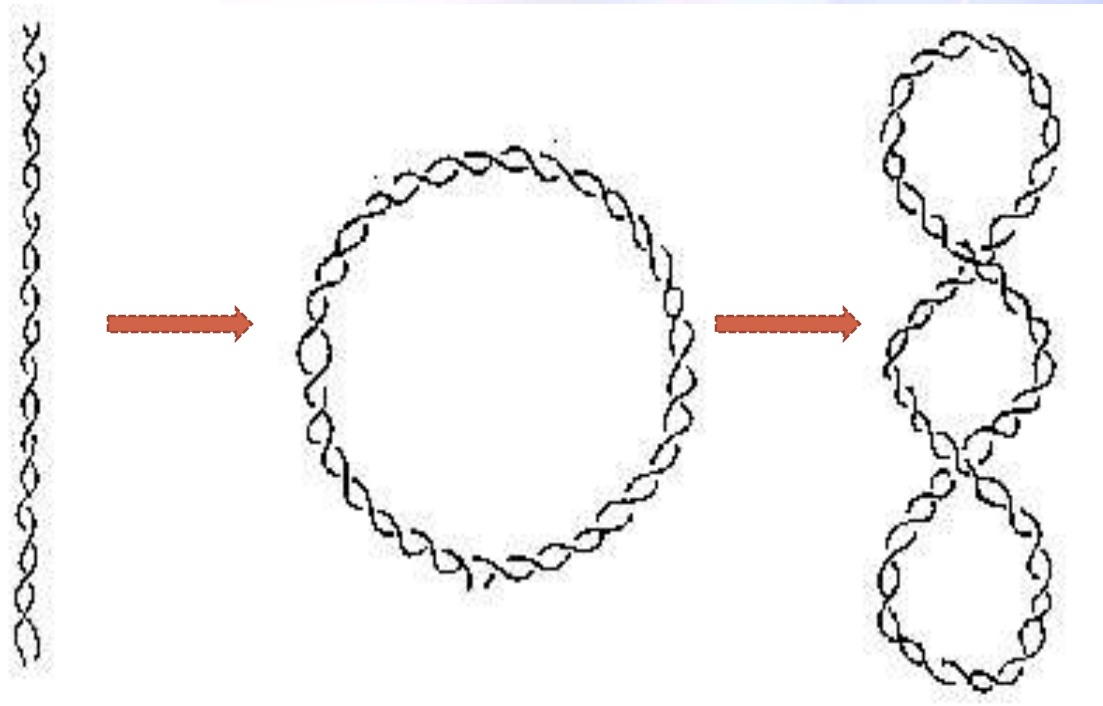
Organization of Genetic Material in Bacteria

- Some of the loops are super-coiled while some are relaxed.
- Relaxed loops are formed as a result of a nick introduced into super-coiled loops by a cellular DNase.

Organization of Genetic Material in Bacteria



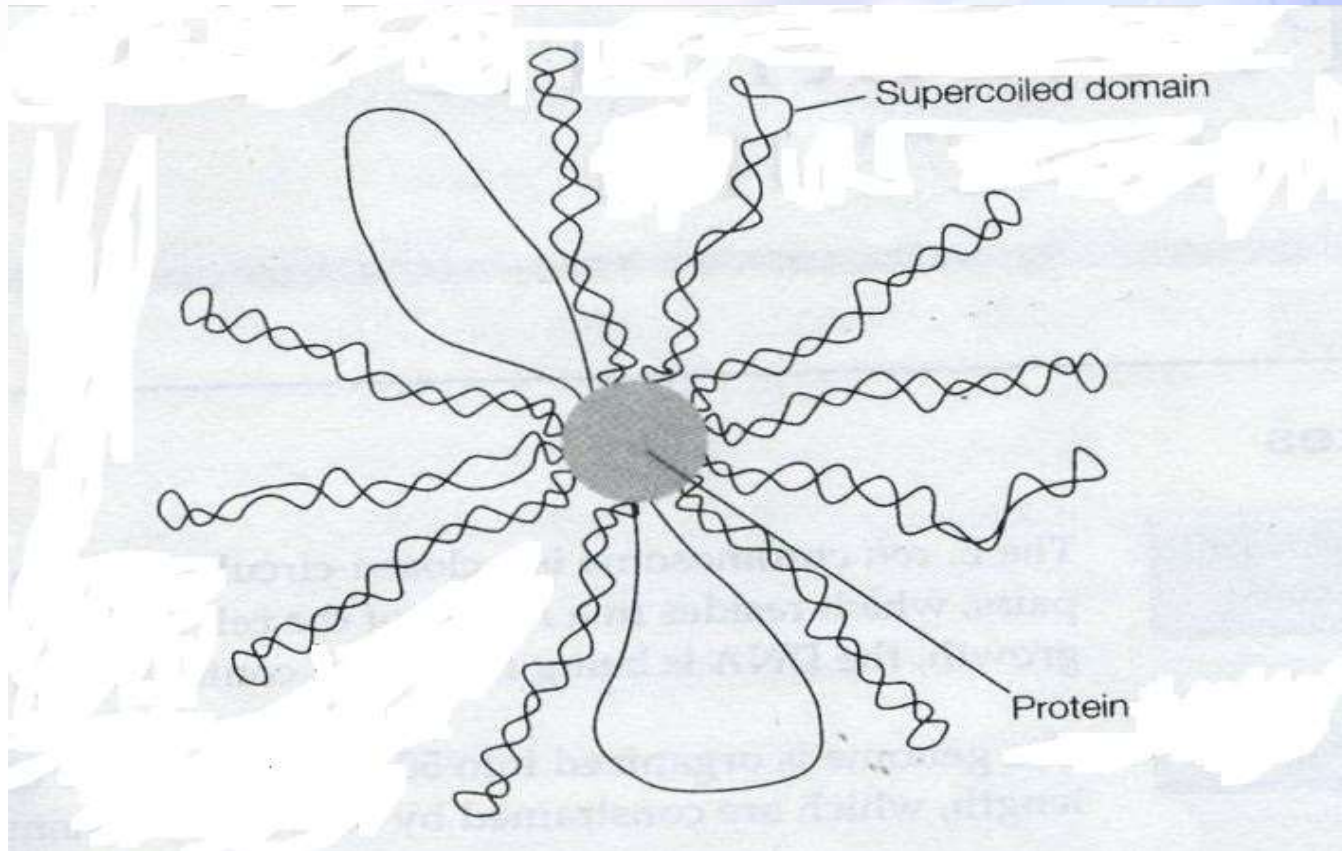
Organization of Genetic Material in Bacteria



Organization of Genetic Material in Bacteria

- If a super-coiled DNA molecule receives a nick, the strain of underwinding is immediately removed, and all the super-coiling is lost.
- Studies confirm that continued nuclease treatment increases number of relaxed loops.

Organization of Genetic Material in Bacteria



Organization of Genetic Material in Bacteria

- The bacterial DNA is arranged in super-coiled loops that are fastened to a central protein matrix, so that each loop is topologically independent from all the others.

Organization of Genetic Material in Bacteria

- So a nick that causes one super-coiled loop to relax would have no effect on other super-coiled loops.
- The super-coiled loops are dynamic structures which change during cell growth & division.

Organization of Genetic Material in Bacteria

- An *E. coli* chromosome is estimated to have about 400 super-coiled loops.
- Each loop has an average length of about 10-20 kbp.

Organization of Genetic Material in Bacteria

- The DNA compaction in a bacterial cell is contributed by supercoiling of loops, macromolecular crowding and DNA-binding proteins.

Organization of Genetic Material in Eukaryotes

- The genetic material (DNA) of eukaryotic organisms is organized in the form of chromosomes.
- The chromosomes of eukaryotic cells are larger and more complex than those of prokaryotes.

Organization of Genetic Material in Eukaryotes



Organization of Genetic Material in Eukaryotes

- Each un-replicated chromosome consists of a single molecule of DNA.
- If stretched out, some human chromosomes would be several centimetres long.

Organization of Genetic Material in Eukaryotes

- To package such a tremendous length of DNA into this small volume, each DNA molecule is coiled again and again and tightly packed around histone proteins.

Organization of Genetic Material in Eukaryotes

- As eukaryotic chromosomes are not circular, so instead of super-coiling, the mechanism of packaging involves winding the DNA around special proteins, the **histones**.

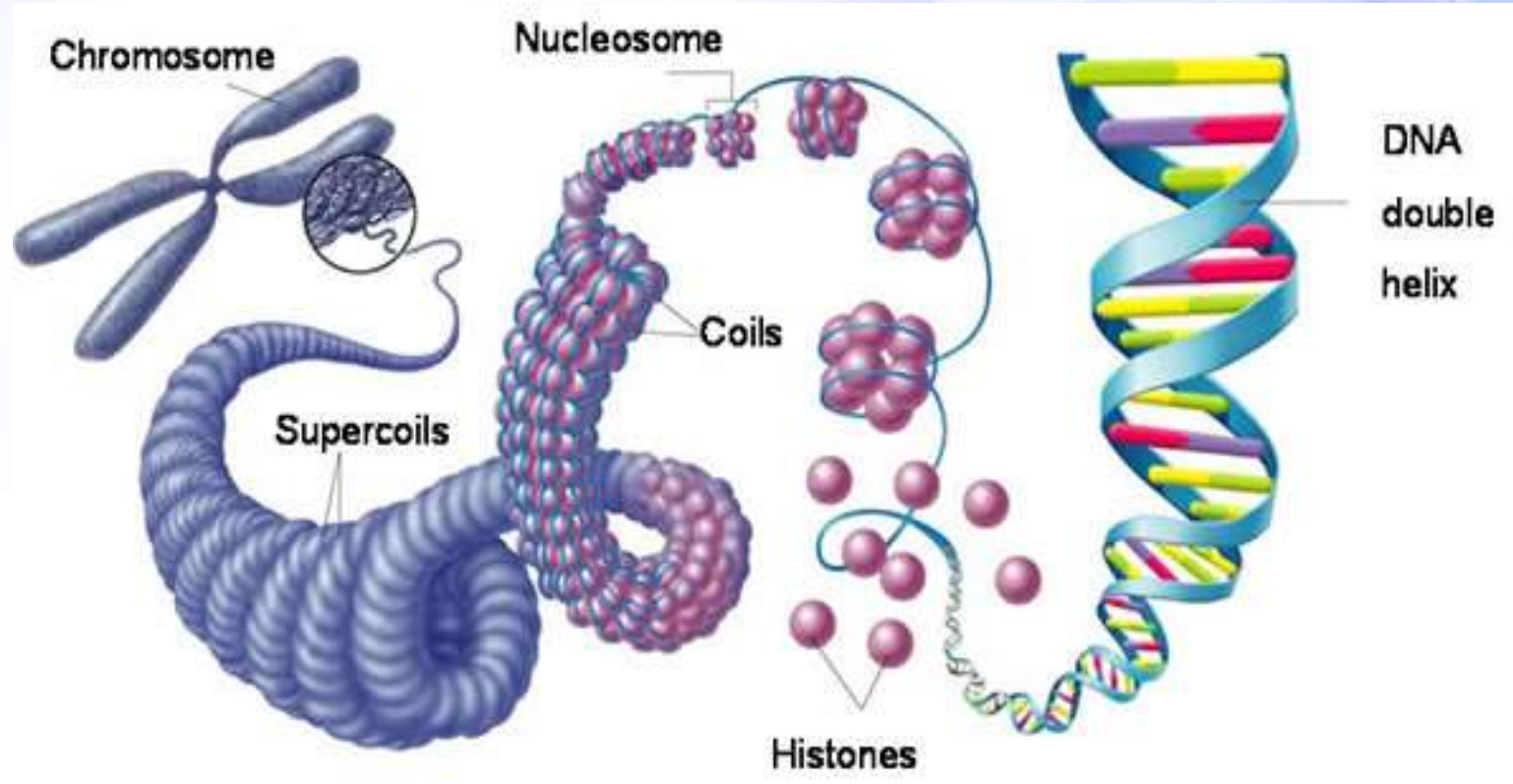
Organization of Genetic Material in Eukaryotes

- DNA with bound histones in the eukaryotes is called as **chromatin**.
- Chromatin consists of roughly spherical subunits, the **nucleosomes**, each containing approx. 200 bp of DNA and nine histones.

Organization of Genetic Material in Eukaryotes

- A condensed mitotic chromosome is about 50,000 times shorter than fully extended DNA.
- Highly condensed chromatin is known as **heterochromatin**.
- The more extended form is known as **euchromatin**.

Organization of Genetic Material in Eukaryotes



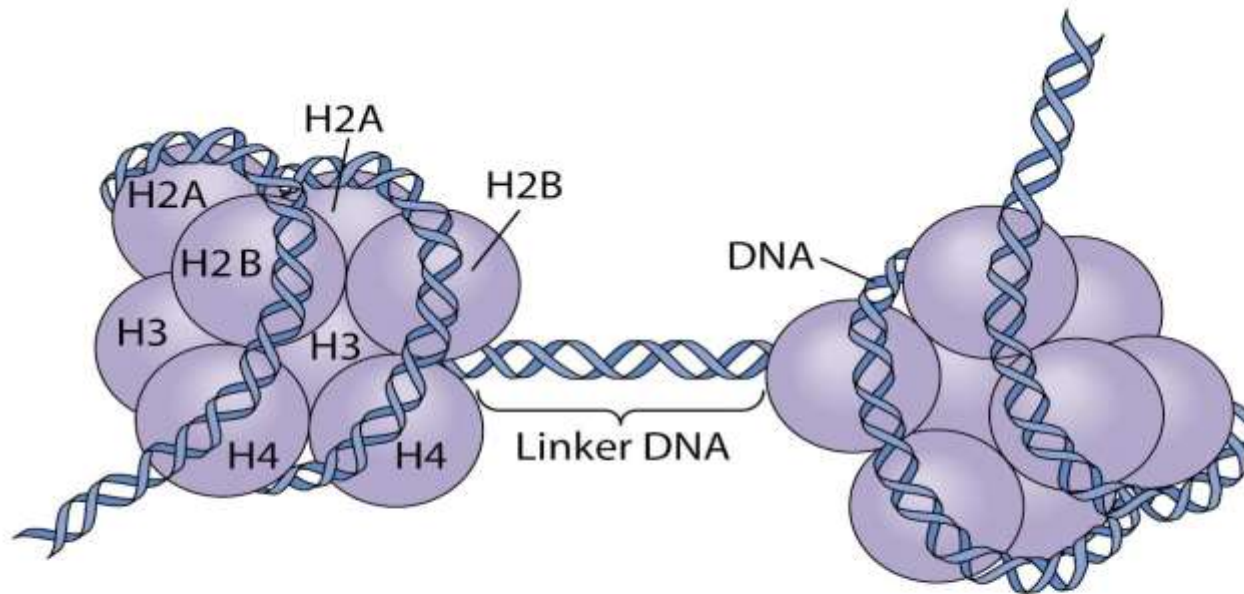
Histone Proteins

- Most abundant proteins in the chromatin are histones.
- There are nine types of histones including two each of H2A, H2B, H3 and H4 and one of H1.

Histone Proteins

- These histones fall in five major classes i.e., H1, H2A, H2B, H3 and H4.
- A typical human cell contains about 60 million copies of each kind of histone.

Histone Proteins



Histone Proteins

- All histones have a high percentage of arginine and lysine but the lysine-to-arginine ratio differs in each type of histone.

Histone Proteins

- The positively charged side chains of lysine and arginine enable histones to bind to the negatively charged phosphate groups of the DNA.

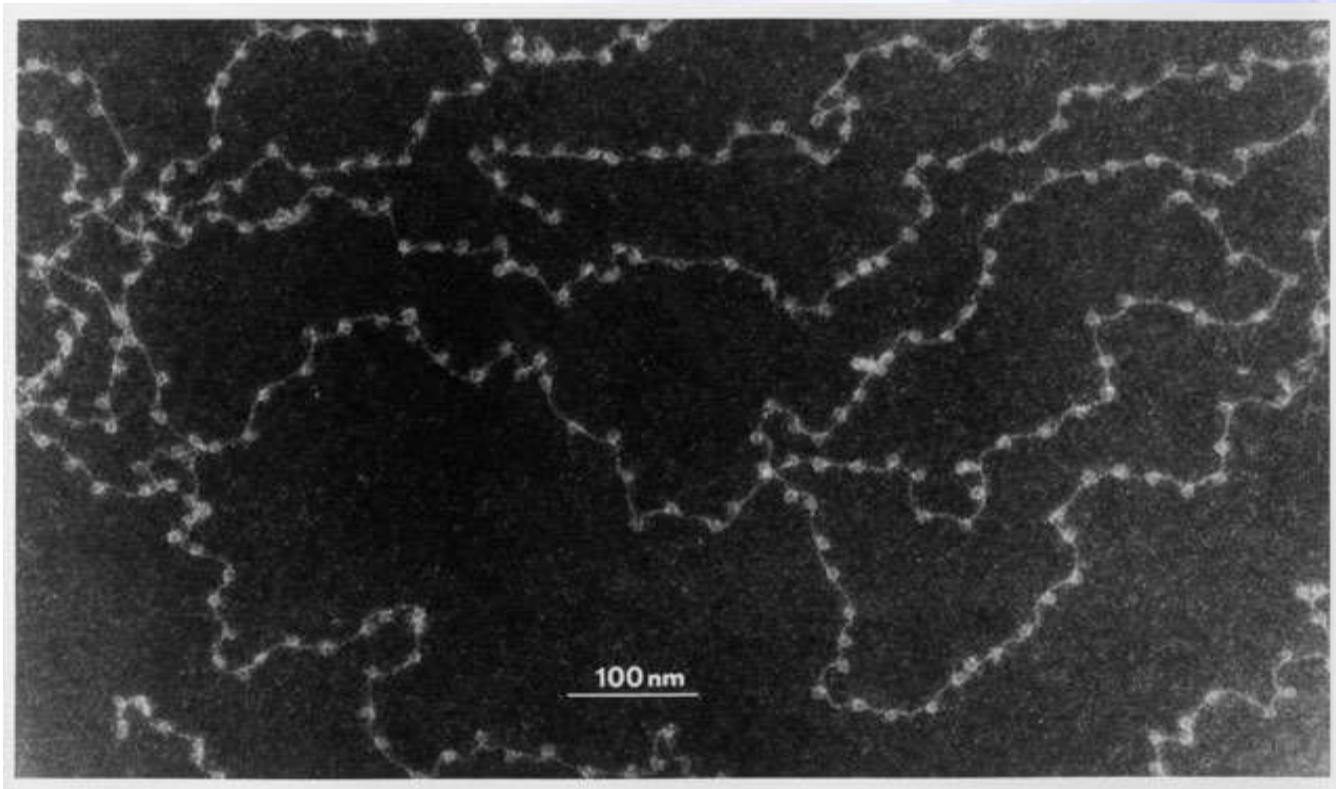
Histone Proteins

- The electrostatic attraction is an important stabilizing force in the chromatin.

The Nucleosome

- The uncondensed chromatin resembles beads on a string when viewed under the electron microscope.
- Each bead is a nucleoprotein complex called **nucleosome**.

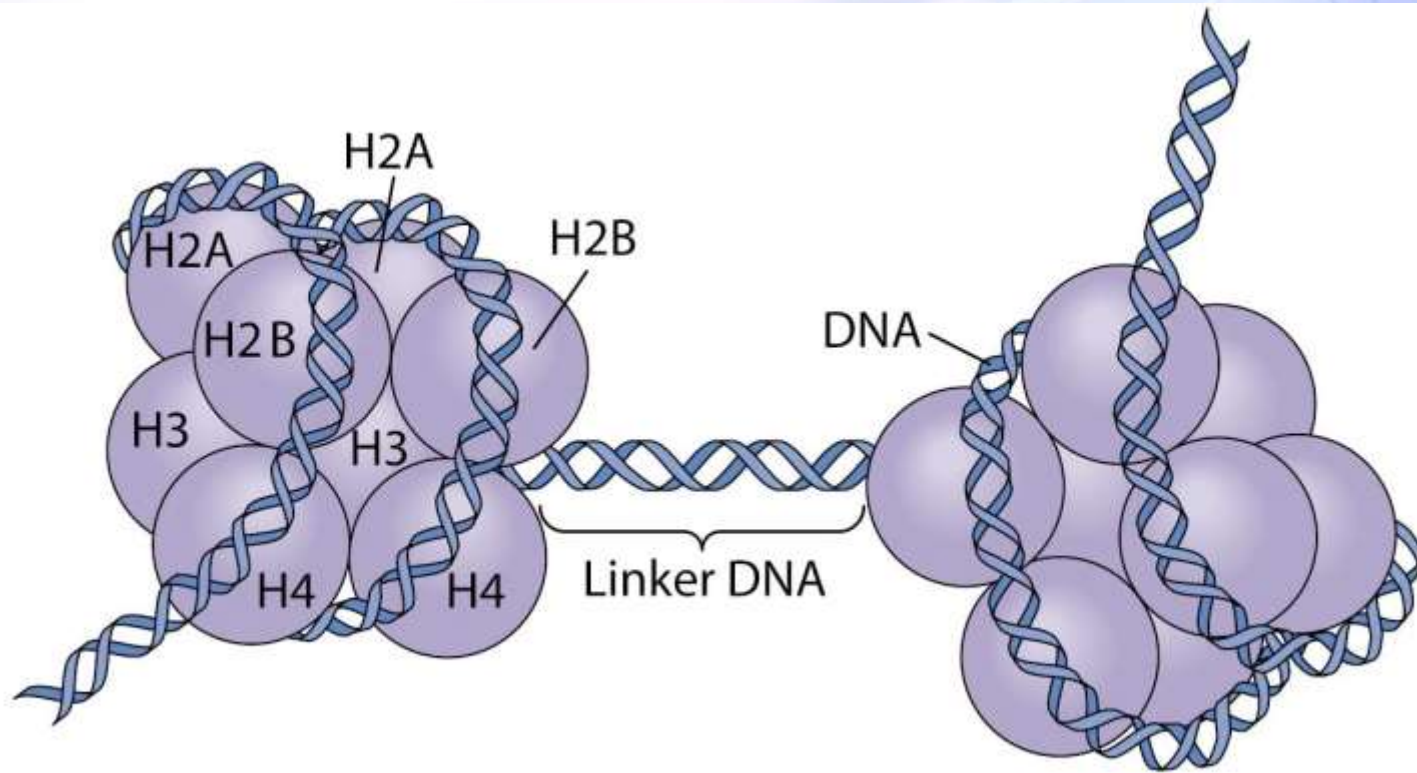
The Nucleosome



The Nucleosome

- Each nucleosome is formed by winding DNA fibre around a protein assembly consisting of eight histone molecules.
- The DNA connecting two nucleosomes is called **linker DNA**.

The Nucleosome



The Nucleosome

- The size of the linker DNA between the nucleosomes varies among different organisms and even different organs of the same organism.

The Nucleosome

- The length of DNA wrapped around nucleosomes also varies from one organism to the other ranging from about 170-240 bp.

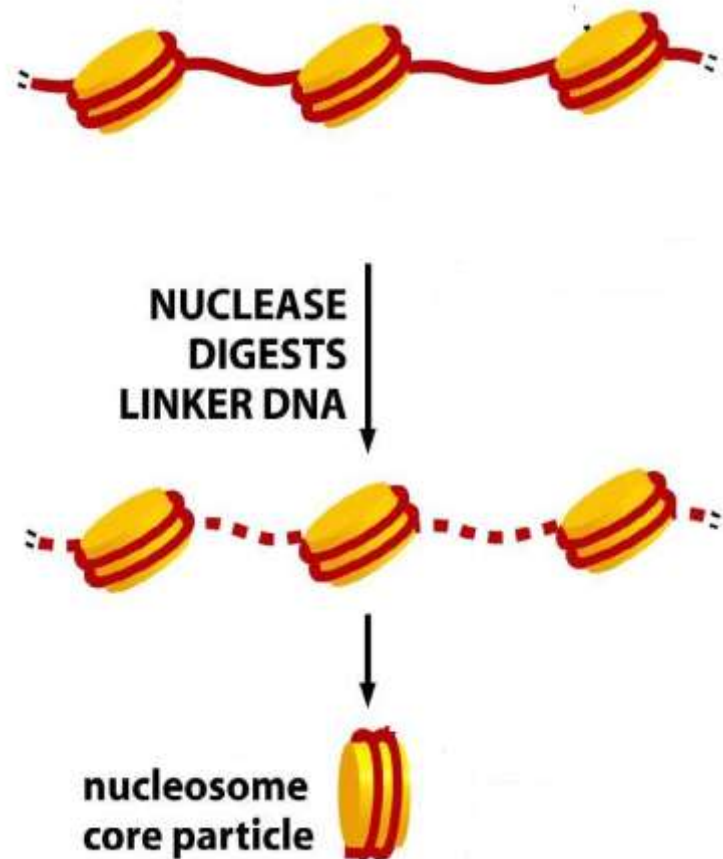
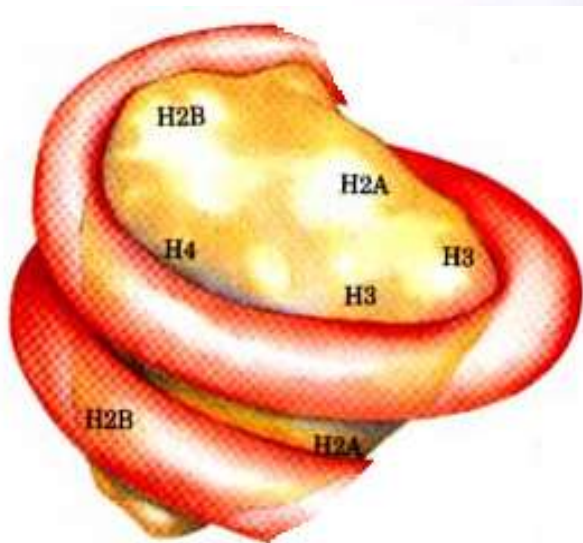
The Nucleosome

- Prolonged nuclease digestion of chromatin cleaves additional nucleotides.
- The structure that remains is the nucleosome core particle.

The Nucleosome

- The nucleosome core particle consists of an octameric protein complex (two copies of each H2A, H2B, H3 & H4) with a 146 bp DNA fragment wound around it.

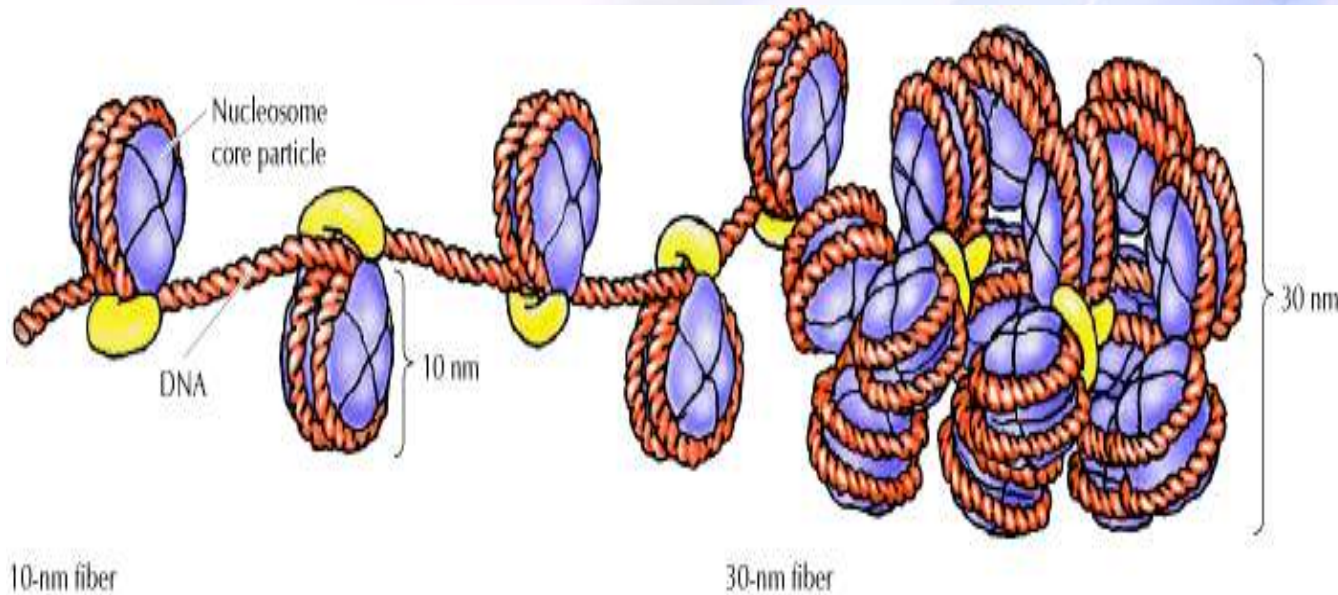
The Nucleosome



The 30-nm Fiber

- It is still unclear that how a chain of nucleosomes folds into higher order structures.
- The next level of organization is a 30-nm fiber.

The 30-nm Fiber



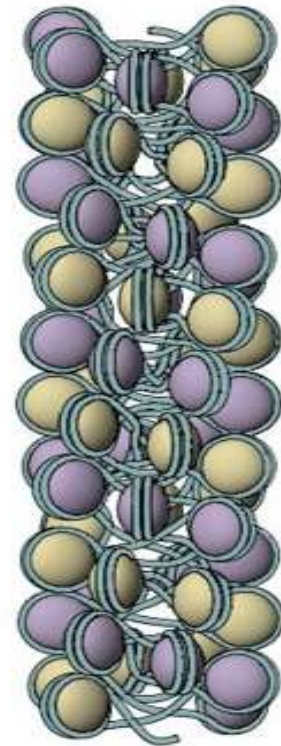
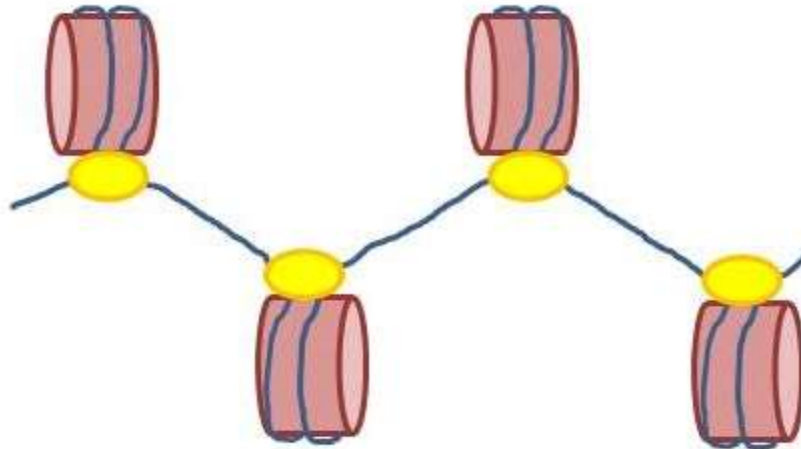
The 30-nm Fiber

- Various models have been proposed to explain how nucleosomes fold to form the 30-nm fiber.
- However, two models gained the most support:-
 - Zigzag Model
 - Solenoid Model

The 30-nm Fiber

- Zigzag model predicts that the linker DNA forms a straight path between successive nucleosomes.
- The nucleosomes lie on opposite sides of the fiber.

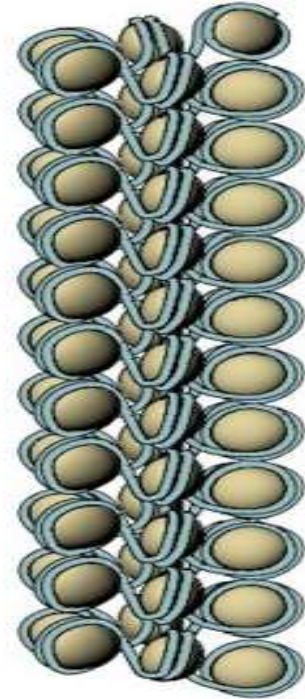
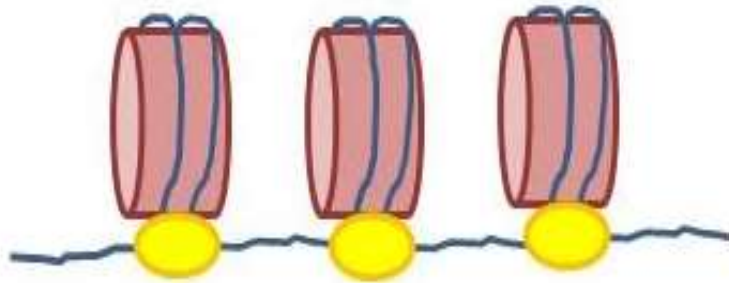
The 30-nm Fiber



The 30-nm Fiber

- The solenoid model predicts that the nucleosome chain forms a helical structure with about 6 nucleosomes per turn.
- Linker DNA is bent to connect neighbouring nucleosomes.

The 30-nm Fiber



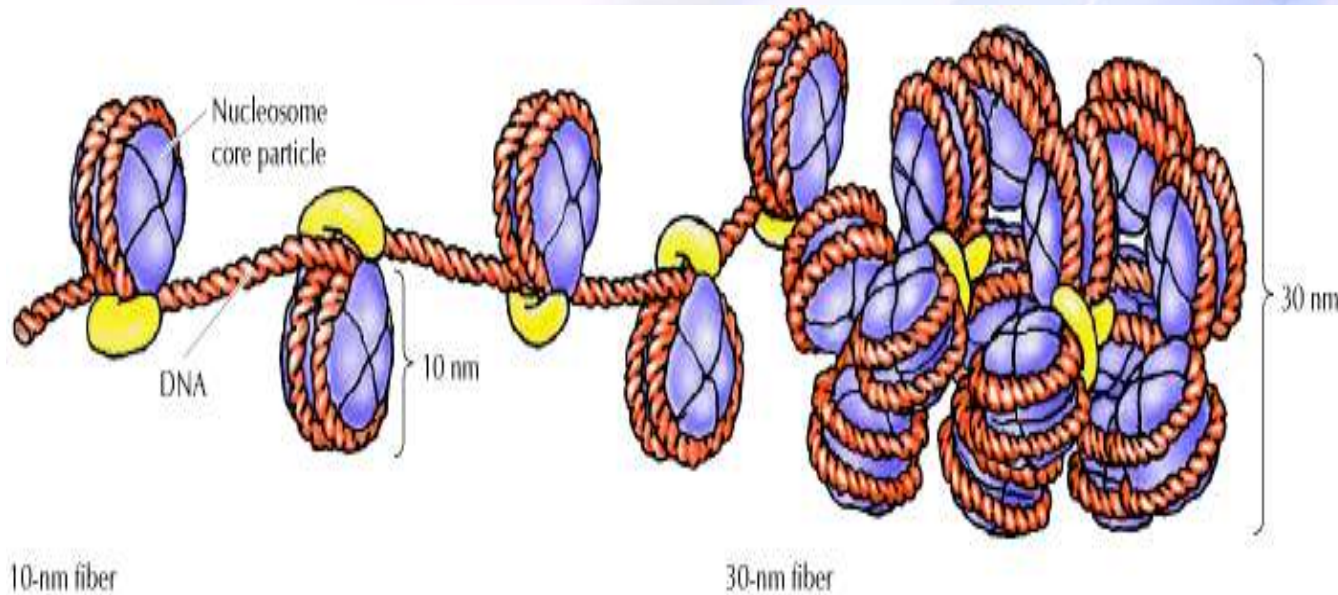
The 30-nm Fiber

- Some of the reconstitution experiments appear to support Zigzag models while some others support Solenoid model.

The 30-nm Fiber

- It is still unclear that how a chain of nucleosomes folds into higher order structures.
- The next level of organization is a 30-nm fiber.

The 30-nm Fiber



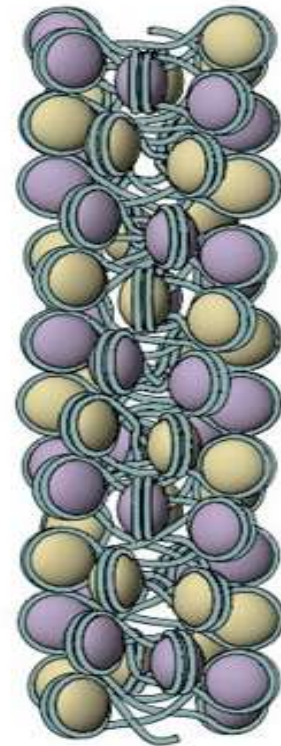
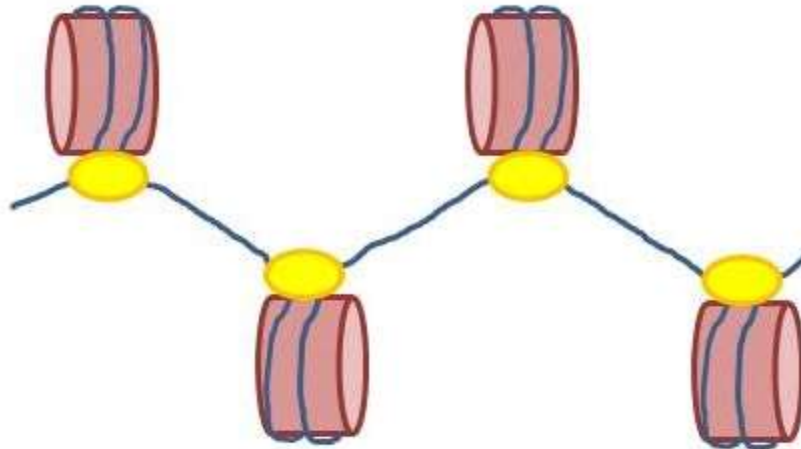
The 30-nm Fiber

- Various models have been proposed to explain how nucleosomes fold to form the 30-nm fiber.
- However, two models gained the most support:-
 - Zigzag Model
 - Solenoid Model

The 30-nm Fiber

- Zigzag model predicts that the linker DNA forms a straight path between successive nucleosomes.
- The nucleosomes lie on opposite sides of the fiber.

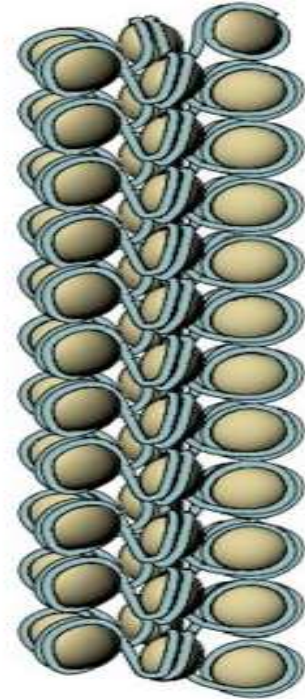
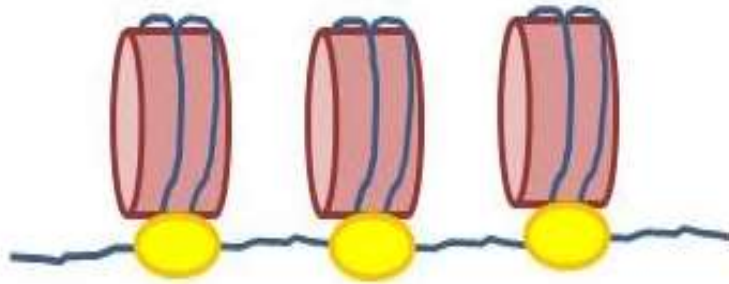
The 30-nm Fiber



The 30-nm Fiber

- The solenoid model predicts that the nucleosome chain forms a helical structure with about 6 nucleosomes per turn.
- Linker DNA is bent to connect neighbouring nucleosomes.

The 30-nm Fiber



The 30-nm Fiber

- Some of the reconstitution experiments appear to support Zigzag models while some others support Solenoid model.

Replication of DNA

- The double-helical model for DNA includes the concept that the two strands are complementary.
- Thus, each strand can in principle serve as the template for making its own partner.

Replication of DNA

- A number of models were proposed to explain the mode of replication of DNA.
- But the **semiconservative model** for DNA replication is the correct one.

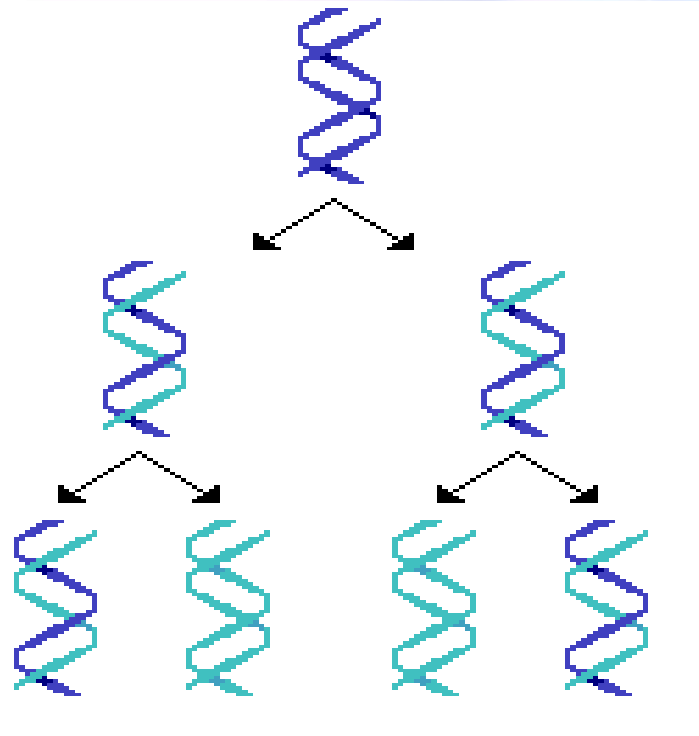
Replication of DNA

- The Watson–Crick model for DNA replication proposed that the two parental strands separate and that each then serves as a template for a new progeny strand.

Replication of DNA

- This is called semiconservative replication because each daughter duplex has one parental strand and one new strand which means that one of the parental strands is “conserved” in each daughter duplex.

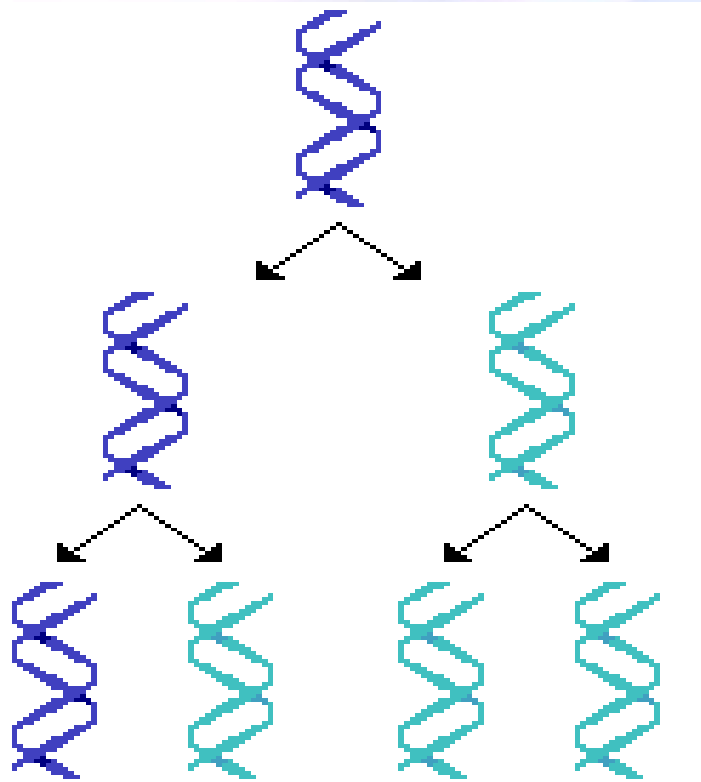
Replication of DNA



Replication of DNA

- Another potential mechanism is **conservative replication**, in which the two parental strands stay together and somehow produce another daughter helix with two completely new strands.

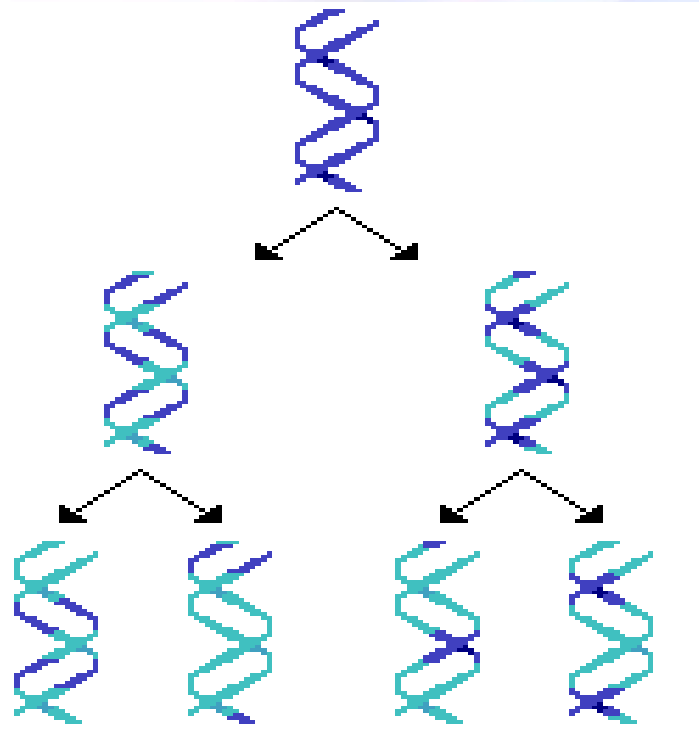
Replication of DNA



Replication of DNA

- Yet another possibility is **dispersive replication**, in which the DNA becomes fragmented so that new and old DNAs coexist in the same strand after replication.

Replication of DNA



Experiment of Meselson & Stahl

- In 1958, Matthew Meselson and Franklin Stahl performed a classic experiment to distinguish among these three possibilities.

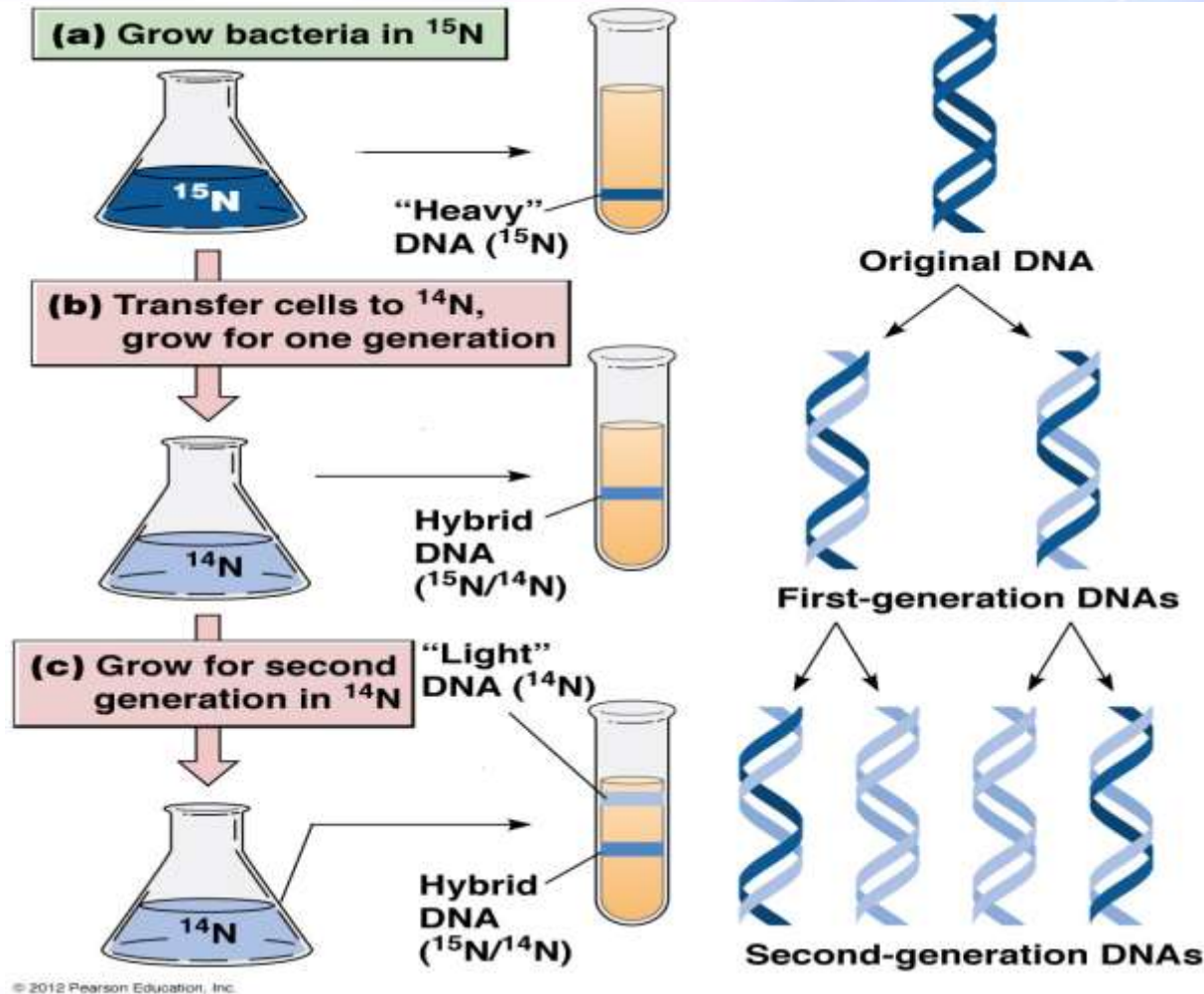
Experiment of Meselson & Stahl

- They labeled *E. coli* DNA with heavy nitrogen (^{15}N) by growing cells in a medium enriched in this nitrogen isotope.
- This made the DNA denser than normal.

Experiment of Meselson & Stahl

- Then they switched the cells to an ordinary medium containing primarily ^{14}N , for various lengths of time.
- Finally, they subjected the DNA to density gradient centrifugation to determine the density of the DNA.

Experiment of Meselson & Stahl



Chemistry of DNA Synthesis

- Two key substrates are required for the synthesis of DNA to proceed.
 - Deoxynucleoside triphosphates
 - Primer:template junction

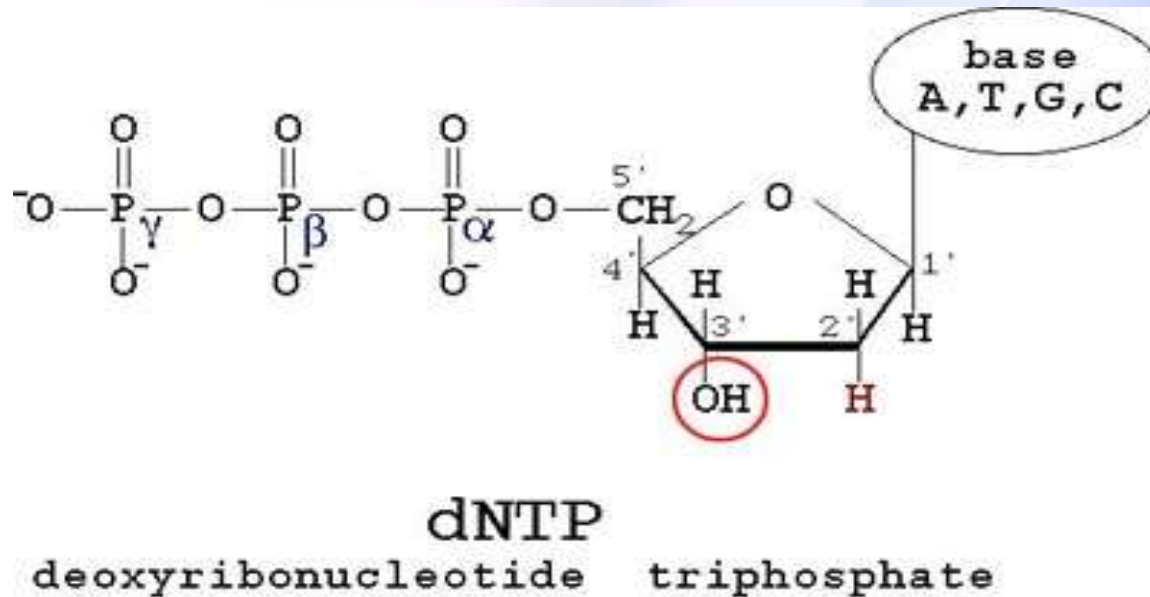
Chemistry of DNA Synthesis

- Four deoxynucleoside triphosphates namely dGTP, dCTP, dATP & dTTP are required.
- Nucleoside triphosphates have three phosphoryl groups attached to the 5' hydroxyl of deoxyribose.

Chemistry of DNA Synthesis

- The innermost phosphoryl group is called the α -phosphate whereas the middle and outermost groups are called β - and γ -phosphates.

Chemistry of DNA Synthesis



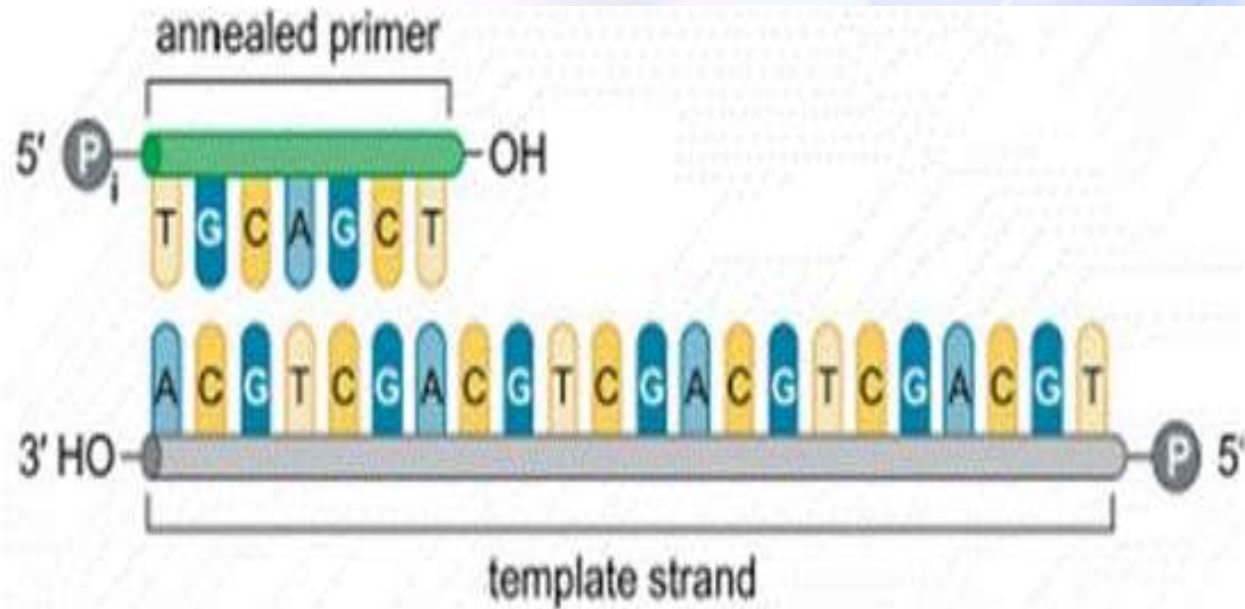
Chemistry of DNA Synthesis

- The second important substrate for DNA synthesis is a particular arrangement of single stranded DNA (ssDNA) and double stranded DNA (dsDNA).

Chemistry of DNA Synthesis

- This particular arrangement is called a primer:template junction.
- It has two components:-
 - The Template
 - The Primer

Chemistry of DNA Synthesis



END

Chemistry of DNA Synthesis

- The new chain of DNA grows by extending the 3' end of the primer.
- The phosphodiester bond is formed in an S_N2 reaction.

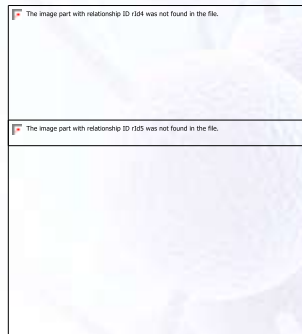
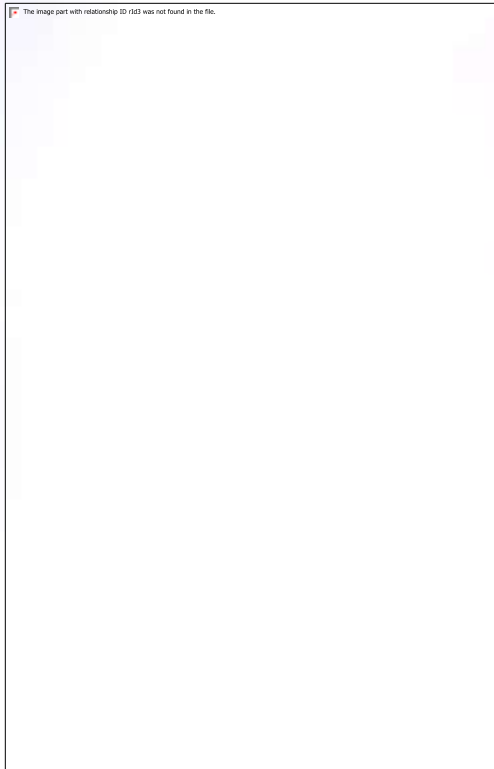
Chemistry of DNA Synthesis

- In this reaction, the hydroxyl group of the 3' end of the primer attacks the α -phosphoryl group of the incoming nucleoside triphosphate.

Chemistry of DNA Synthesis

- The leaving group of the reaction is pyrophosphate which arises from the release of β - and γ -phosphates of the nucleoside.

Chemistry of DNA Synthesis



Chemistry of DNA Synthesis

- The template strand directs which of the four nucleoside triphosphates is added.
- The incoming nucleoside triphosphate base pairs with the template strand.

Chemistry of DNA Synthesis

- What is the driving force for the addition of nucleotide to a growing polynucleotide chain?

Chemistry of DNA Synthesis

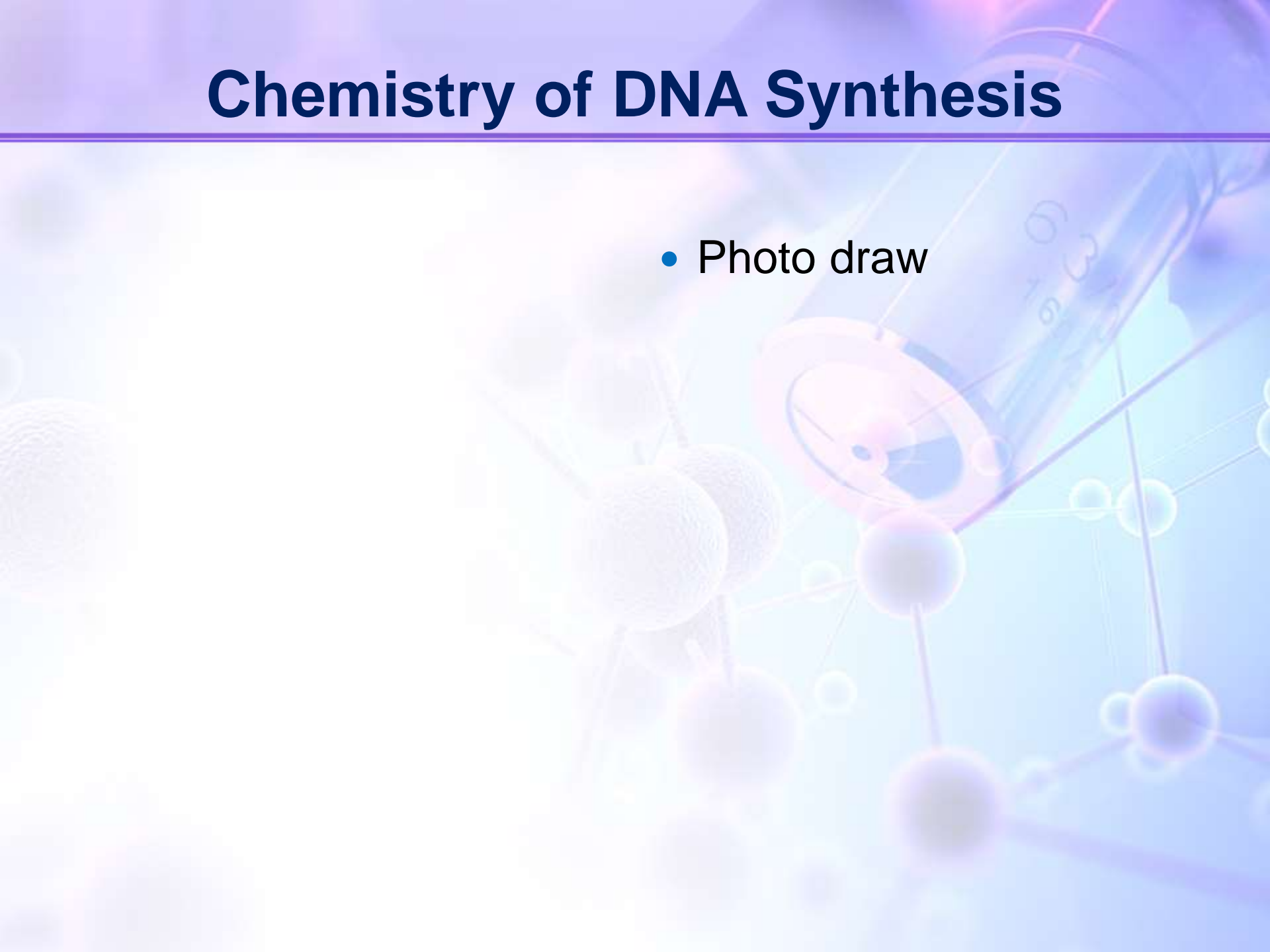
- The free energy for this reaction is provided by the rapid hydrolysis of the pyrophosphate into two phosphate groups by an enzyme known as pyrophosphatase.

Chemistry of DNA Synthesis

- The net result of nucleotide addition and pyrophosphate hydrolysis is the simultaneous breaking of two high energy phosphate bonds.

Chemistry of DNA Synthesis

- Photo draw



Chemistry of DNA Synthesis

- Therefore, DNA synthesis is a coupled process.
- This reaction is highly favourable with high value of K_{eq} which means that its an irreversible reaction.

Chemistry of DNA Synthesis

- Two key substrates are required for the synthesis of DNA to proceed.
 - Deoxynucleoside triphosphates
 - Primer:template junction

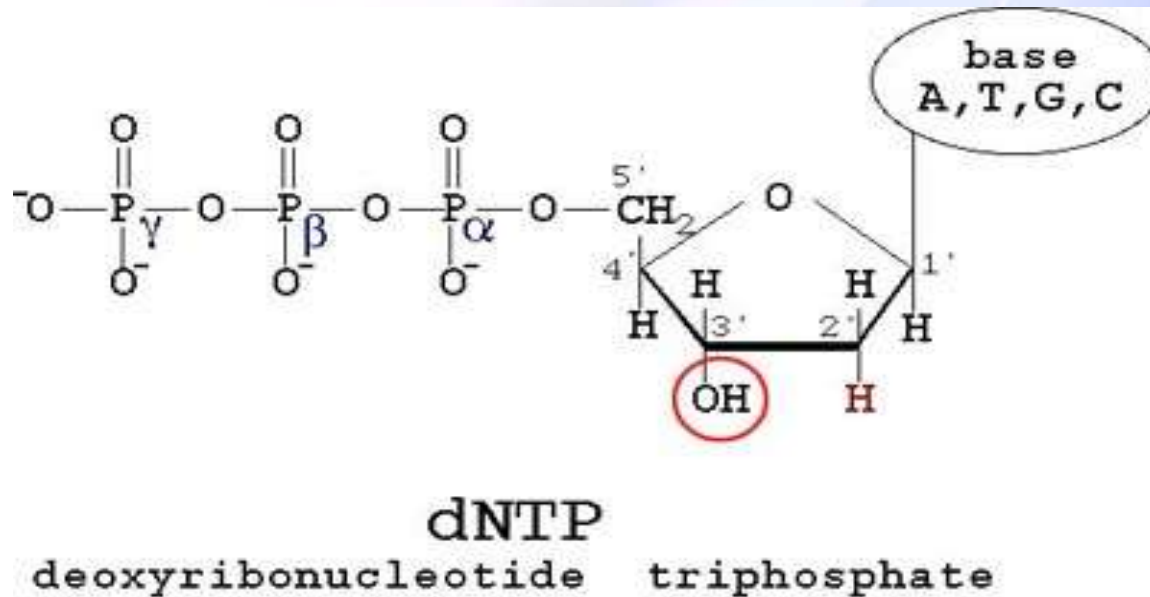
Chemistry of DNA Synthesis

- Four deoxynucleoside triphosphates namely dGTP, dCTP, dATP & dTTP are required.
- Nucleoside triphosphates have three phosphoryl groups attached to the 5' hydroxyl of deoxyribose.

Chemistry of DNA Synthesis

- The innermost phosphoryl group is called the α -phosphate whereas the middle and outermost groups are called β - and γ -phosphates.

Chemistry of DNA Synthesis



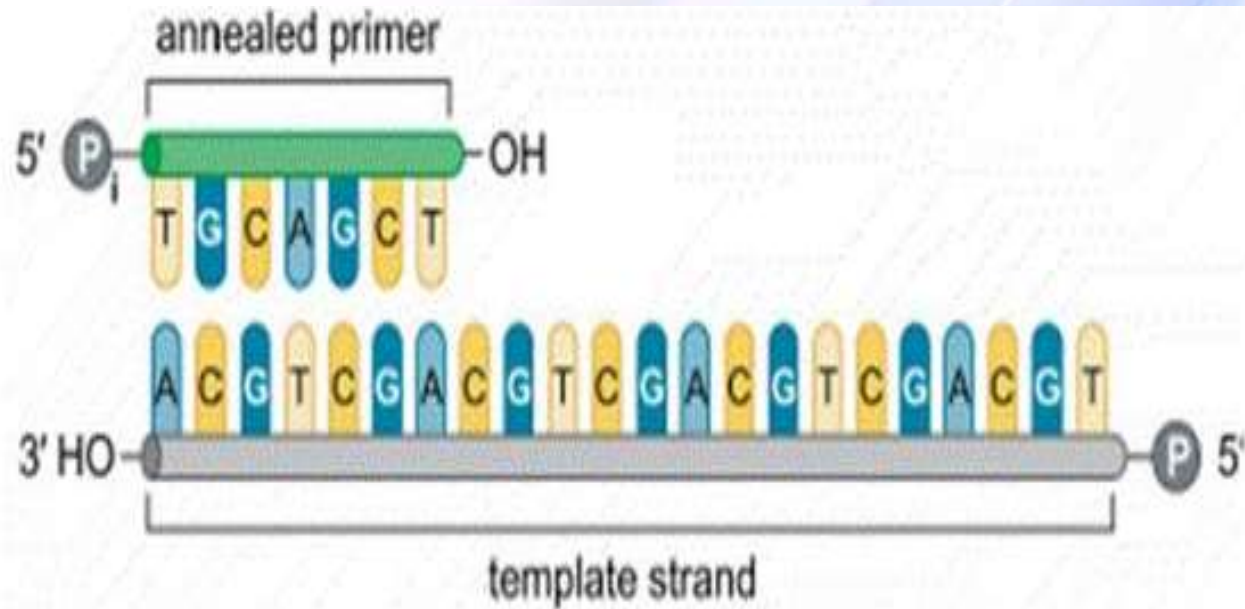
Chemistry of DNA Synthesis

- The second important substrate for DNA synthesis is a particular arrangement of single stranded DNA (ssDNA) and double stranded DNA (dsDNA).

Chemistry of DNA Synthesis

- This particular arrangement is called a primer:template junction.
- It has two components:-
 - The Template
 - The Primer

Chemistry of DNA Synthesis



END

Chemistry of DNA Synthesis

- The new chain of DNA grows by extending the 3' end of the primer.
- The phosphodiester bond is formed in an S_N2 reaction.

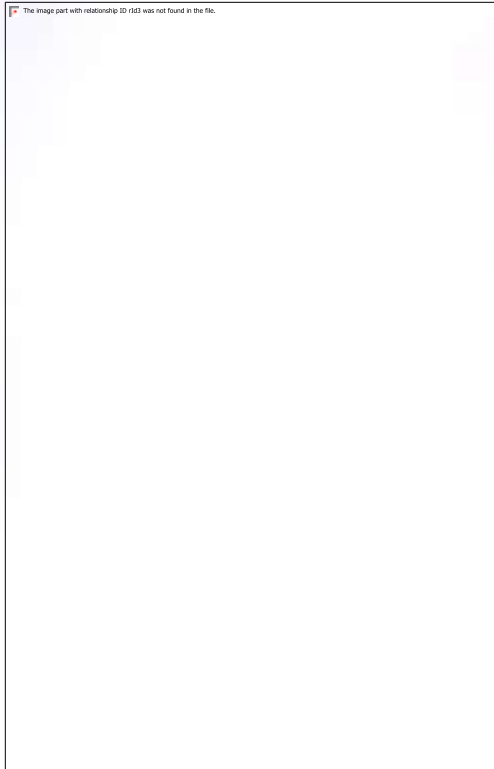
Chemistry of DNA Synthesis

- In this reaction, the hydroxyl group of the 3' end of the primer attacks the α -phosphoryl group of the incoming nucleoside triphosphate.

Chemistry of DNA Synthesis

- The leaving group of the reaction is pyrophosphate which arises from the release of β - and γ -phosphates of the nucleoside.

Chemistry of DNA Synthesis



Chemistry of DNA Synthesis

- The template strand directs which of the four nucleoside triphosphates is added.
- The incoming nucleoside triphosphate base pairs with the template strand.

Chemistry of DNA Synthesis

- What is the driving force for the addition of nucleotide to a growing polynucleotide chain?

Chemistry of DNA Synthesis

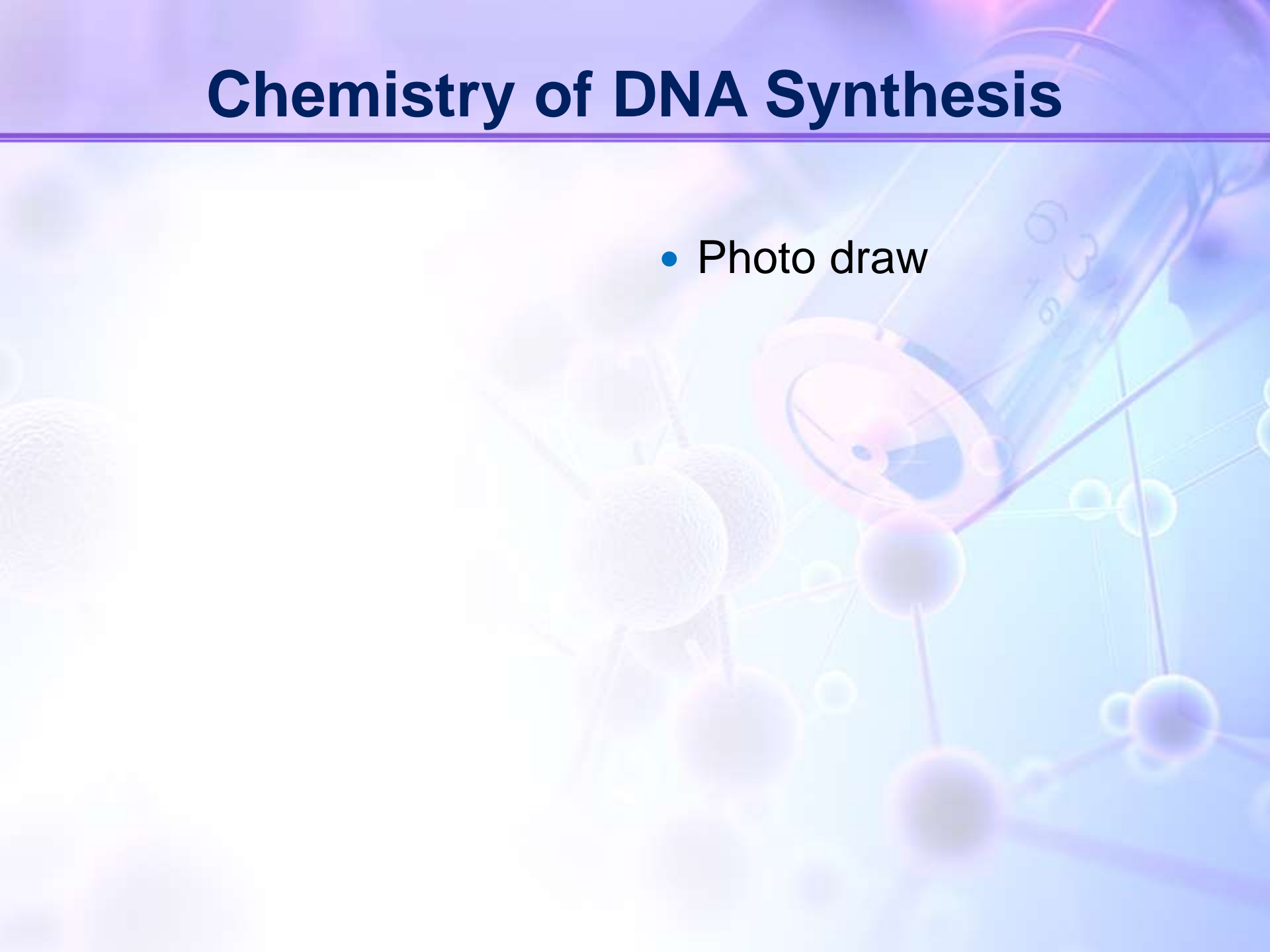
- The free energy for this reaction is provided by the rapid hydrolysis of the pyrophosphate into two phosphate groups by an enzyme known as pyrophosphatase.

Chemistry of DNA Synthesis

- The net result of nucleotide addition and pyrophosphate hydrolysis is the simultaneous breaking of two high energy phosphate bonds.

Chemistry of DNA Synthesis

- Photo draw



Chemistry of DNA Synthesis

- Therefore, DNA synthesis is a coupled process.
- This reaction is highly favourable with high value of K_{eq} which means that its an irreversible reaction.

Mechanism of DNA Polymerase

- The synthesis of DNA is catalyzed by an enzyme **DNA polymerase**.
- It uses a single active site to catalyze the addition of any of four deoxynucleoside triphosphates.

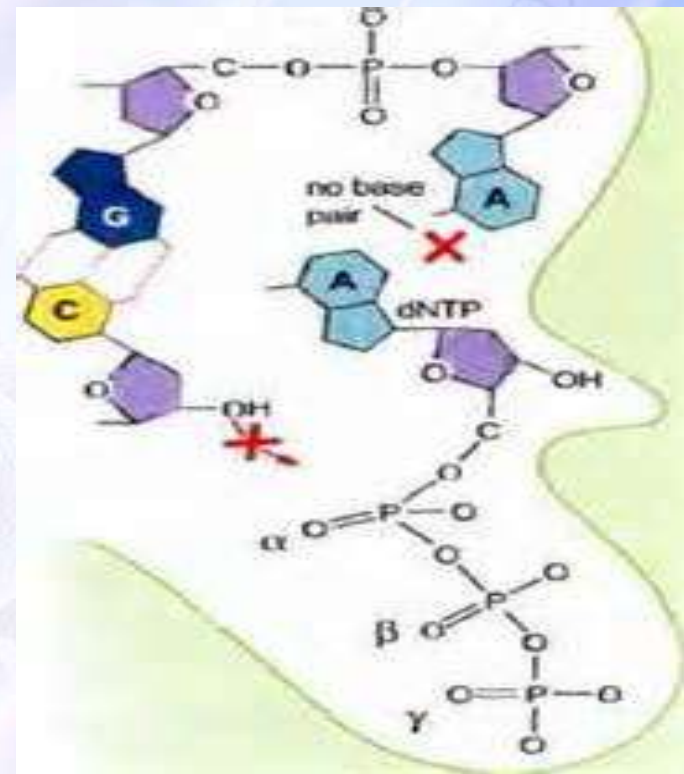
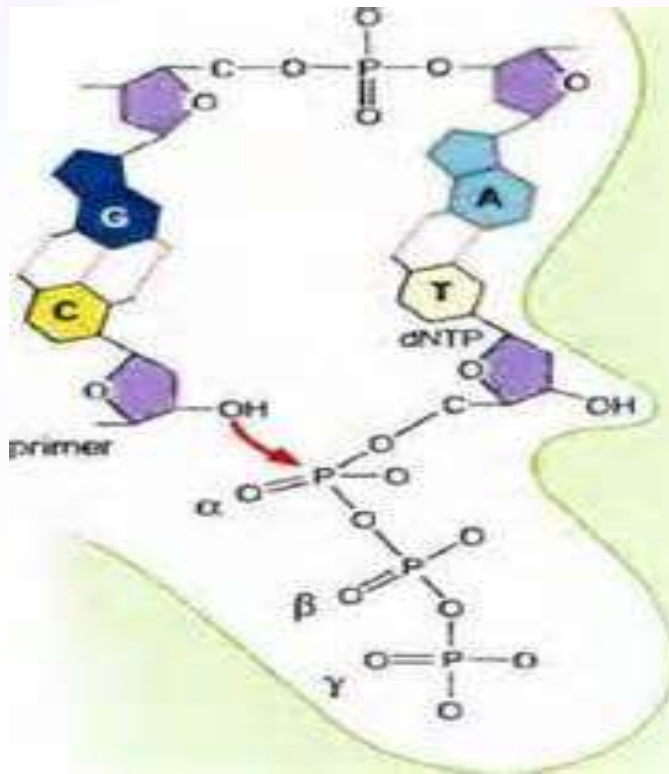
Mechanism of DNA Polymerase

- 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.

Mechanism of DNA Polymerase

- Only when a correct nucleotide comes, the 3'-OH of the primer and the α -phosphate of the nucleotide align in optimum position for catalysis to take place.

Mechanism of DNA Polymerase



Mechanism of DNA Polymerase

- Incorrect base pairing leads to dramatically lower rate of nucleotide addition as a result of catalytically unfavourable alignment of these substrates.

Mechanism of DNA Polymerase

- DNA polymerase shows an impressive ability to distinguish between ribonucleoside (rNTPs) and deoxyribonucleoside triphosphates (dNTPs).

Mechanism of DNA Polymerase

- Although rNTPs are present at approx. ten-fold higher concentration in the cell, yet their incorporation rate is 1000-folds lower than dNTPs.

Mechanism of DNA Polymerase

- This discrimination is mediated by the steric exclusion of rNTPs from the active site of DNA polymerase.

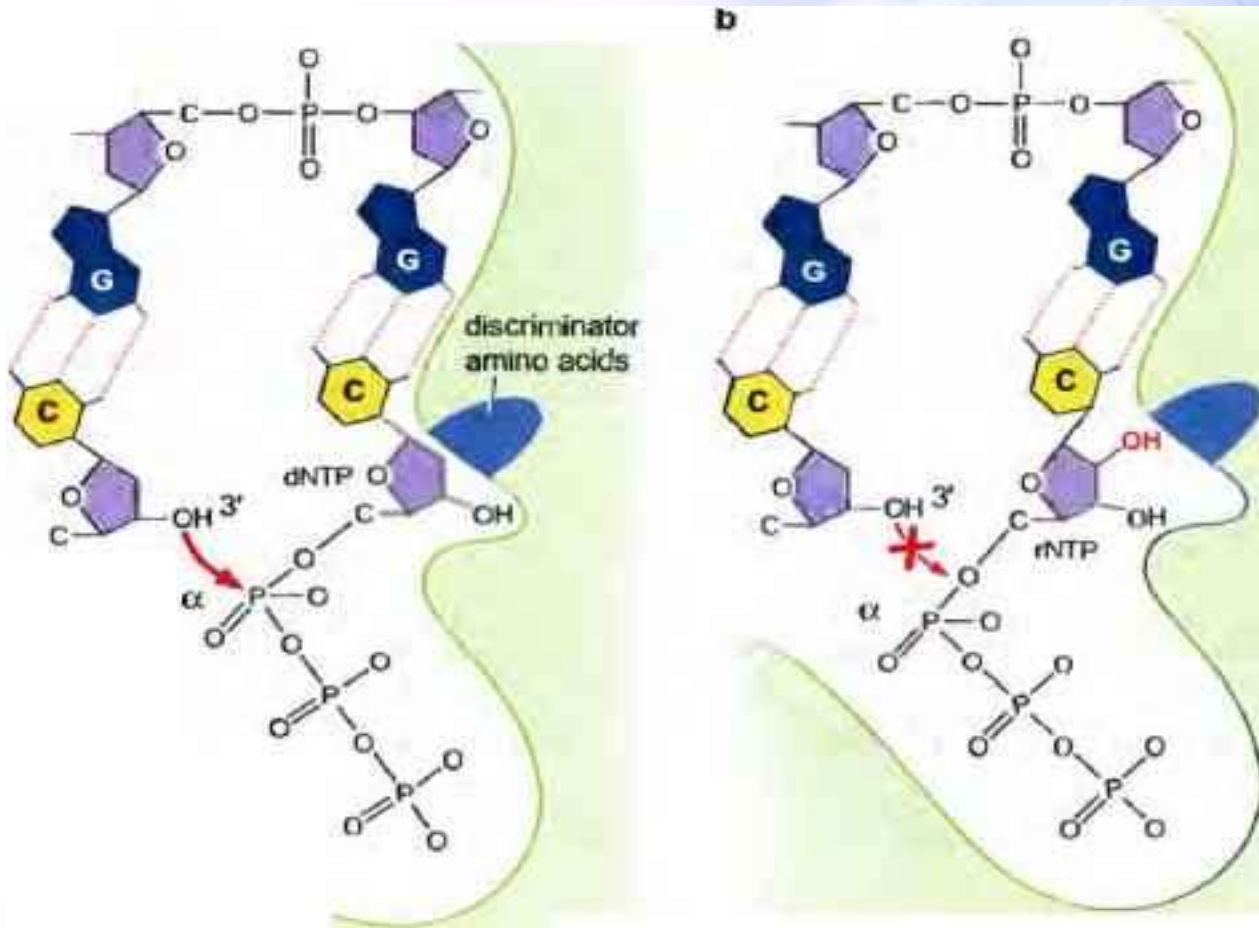
Mechanism of DNA Polymerase

- In DNA polymerase, the nucleotide-binding pocket cannot accommodate a 2'-OH on the incoming nucleotide.

Mechanism of DNA Polymerase

- This space is occupied by two amino acids that make van der Waals contacts with the deoxyribose ring.
- These amino acids are called discriminator amino acids.

Mechanism of DNA Polymerase



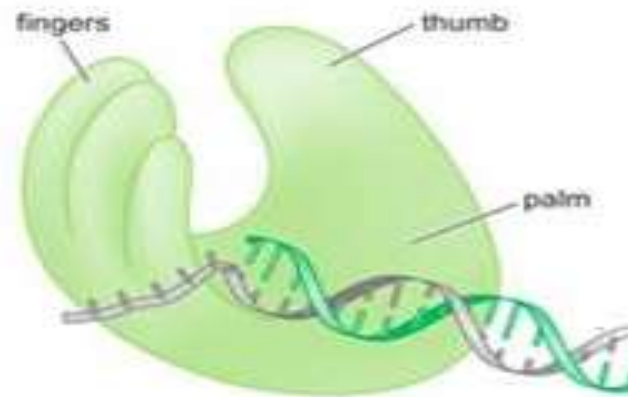
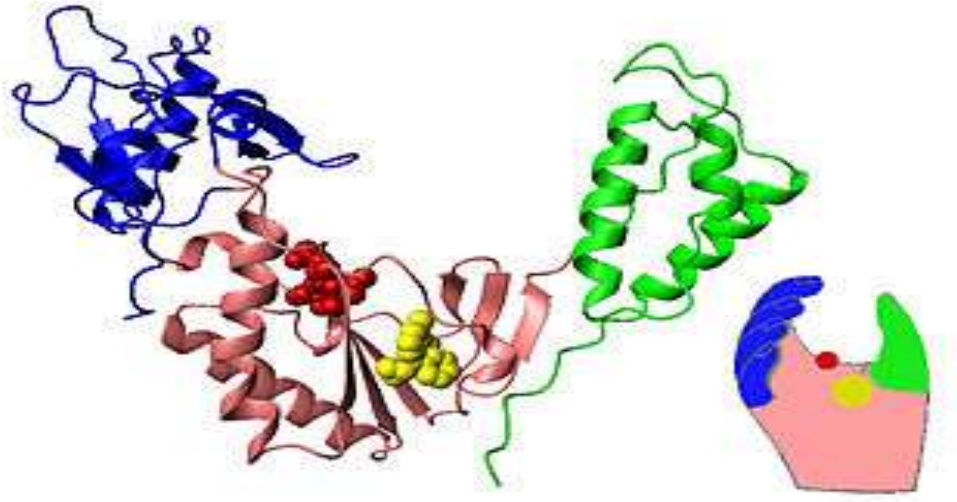
DNA Polymerases Resemble a Hand

- The structural studies on DNA polymerases reveal that the DNA substrate sits in a large cleft that resembles a partially closed right hand.

DNA Polymerases Resemble a Hand

- Based on the hand analogy, the three domains of the DNA polymerase are called the thumb, fingers and palm.

DNA Polymerases Resemble a Hand



DNA Polymerases Resemble a Hand

- The palm domain is composed of a β -sheet and contains the primary elements of the catalytic site.
- This region of DNA polymerase binds two divalent metal ions (Mg^{2+} or Zn^{2+}).

DNA Polymerases Resemble a Hand

- One metal ion reduces the affinity of the 3'-OH for its hydrogen.
- This generates a 3'-O⁻ that is primed for the nucleophilic attack of the α -phosphate of the incoming dNTP.

DNA Polymerases Resemble a Hand

- The second metal ion coordinates the negative charges of the β - and γ -phosphates of the dNTP and stabilizes the pyrophosphate produced by joining the primer and the incoming nucleotide.

DNA Polymerases Resemble a Hand

- In addition to its role in catalysis, the palm domain also monitors the base pairing of the most recently added nucleotides.

DNA Polymerases Resemble a Hand

- The fingers of the polymerase are also important for catalysis.
- Several residues located within the fingers bind to the incoming dNTP.

DNA Polymerases Resemble a Hand

- More importantly, once a correct base pair is formed between the incoming dNTP and the template, the finger domain moves to enclose the dNTP.
- This closed form of the polymerase “hand” stimulates catalysis.

DNA Polymerases Resemble a Hand

- In contrast to the fingers and the palm, the thumb domain is not intimately involved in catalysis.
- Instead, it interacts with the DNA that has been most recently synthesized.

DNA Polymerases Resemble a Hand

- This serves two purposes:-
- First, it maintains the correct position of the primer and the active site.

DNA Polymerases Resemble a Hand

- Second, the thumb helps to maintain a strong association between the DNA polymerase and its substrate.
- This association contributes to the ability of the DNA polymerase to add many dNTPs.

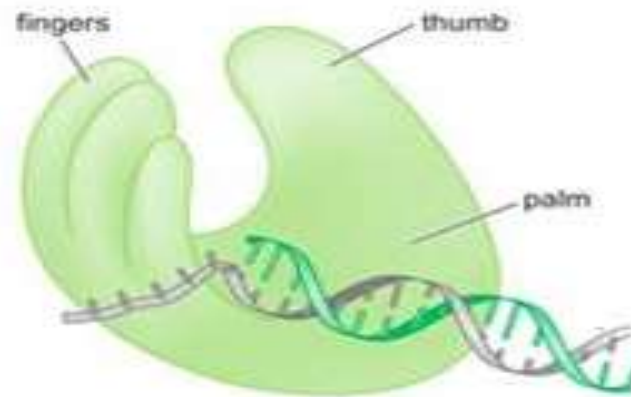
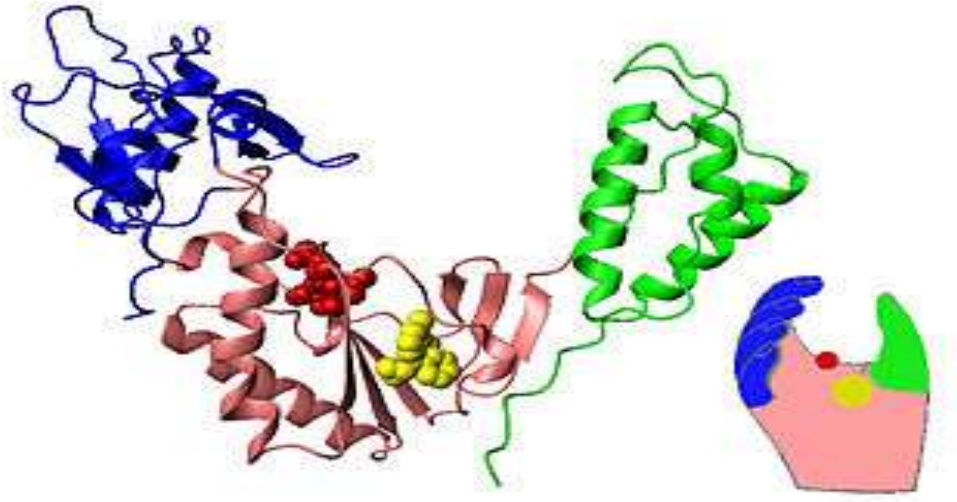
DNA Polymerases Resemble a Hand

- The structural studies on DNA polymerases reveal that the DNA substrate sits in a large cleft that resembles a partially closed right hand.

DNA Polymerases Resemble a Hand

- Based on the hand analogy, the three domains of the DNA polymerase are called the thumb, fingers and palm.

DNA Polymerases Resemble a Hand



DNA Polymerases Resemble a Hand

- The palm domain is composed of a β -sheet and contains the primary elements of the catalytic site.
- This region of DNA polymerase binds two divalent metal ions (Mg^{2+} or Zn^{2+}).

DNA Polymerases Resemble a Hand

- One metal ion reduces the affinity of the 3'-OH for its hydrogen.
- This generates a 3'-O⁻ that is primed for the nucleophilic attack of the α -phosphate of the incoming dNTP.

DNA Polymerases Resemble a Hand

- The second metal ion coordinates the negative charges of the β - and γ -phosphates of the dNTP and stabilizes the pyrophosphate produced by joining the primer and the incoming nucleotide.

DNA Polymerases Resemble a Hand

- In addition to its role in catalysis, the palm domain also monitors the base pairing of the most recently added nucleotides.

DNA Polymerases Resemble a Hand

- The fingers of the polymerase are also important for catalysis.
- Several residues located within the fingers bind to the incoming dNTP.

DNA Polymerases Resemble a Hand

- More importantly, once a correct base pair is formed between the incoming dNTP and the template, the finger domain moves to enclose the dNTP.
- This closed form of the polymerase “hand” stimulates catalysis.

DNA Polymerases Resemble a Hand

- In contrast to the fingers and the palm, the thumb domain is not intimately involved in catalysis.
- Instead, it interacts with the DNA that has been most recently synthesized.

DNA Polymerases Resemble a Hand

- This serves two purposes:-
- First, it maintains the correct position of the primer and the active site.

DNA Polymerases Resemble a Hand

- Second, the thumb helps to maintain a strong association between the DNA polymerase and its substrate.
- This association contributes to the ability of the DNA polymerase to add many dNTPs.

THE REPLICATION FORK

- In the cell, both strands of the DNA duplex are replicated at the same time.
- So it requires separation of the two strands of the double helix to create two template DNAs.

THE REPLICATION FORK

- The junction between the newly separated template strands and the unreplicated duplex DNA is known as the **Replication Fork**.

THE REPLICATION FORK



THE REPLICATION FORK

- The replication fork moves continuously towards the duplex region of unreplicated DNA.
- As the fork moves, it creates two ssDNA templates that each directs the synthesis of a complementary DNA strand.

THE REPLICATION FORK

- The antiparallel nature of DNA creates a complication for the simultaneous replication of the two exposed templates at the replication fork.

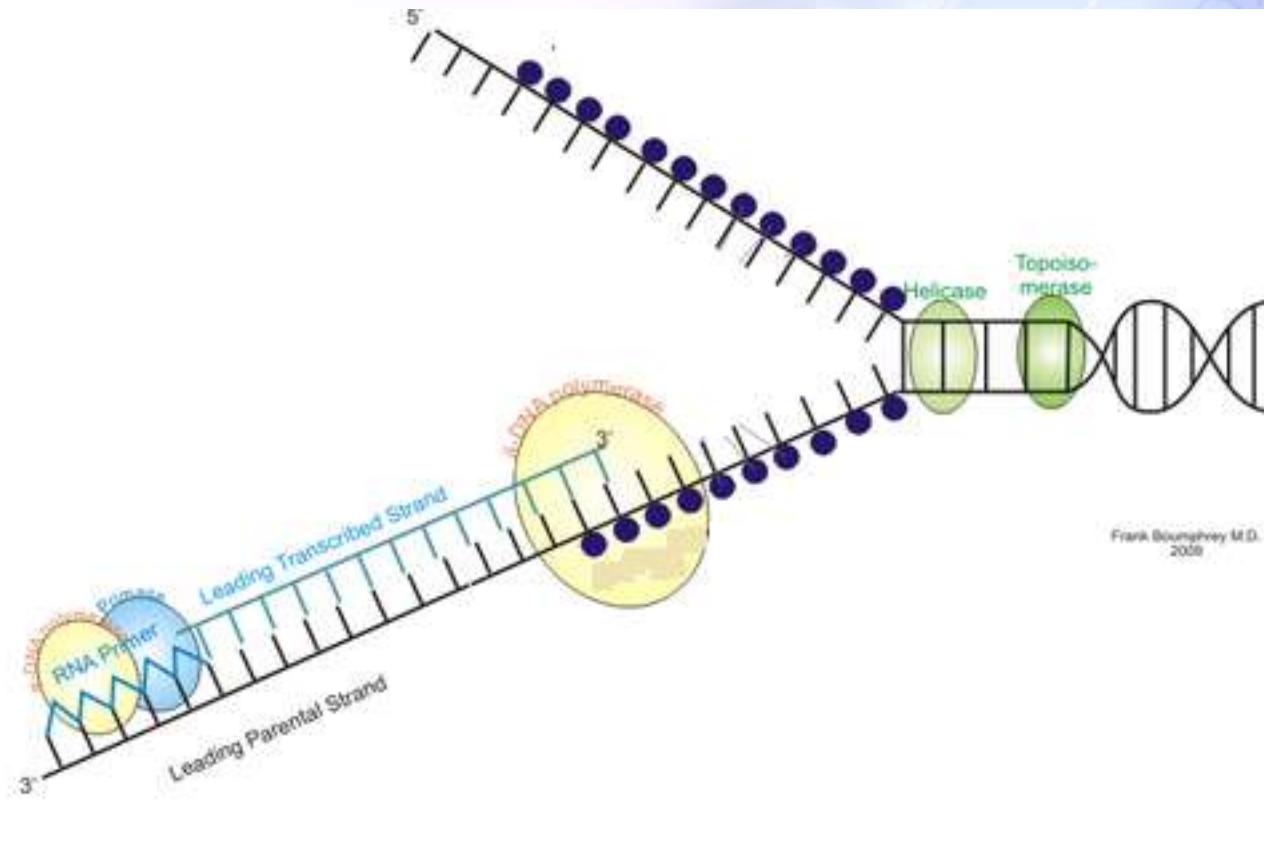
THE REPLICATION FORK

- Because DNA is synthesized only by elongating a 3' end, only one of the two exposed templates can be replicated continuously as the replication fork moves.

THE REPLICATION FORK

- The newly synthesized DNA strand directed by this template is known as the **leading strand**.

THE REPLICATION FORK



THE REPLICATION FORK

- Synthesis of the newDNA strand directed by the other ssDNA template is more complicated.
- This template directs the DNA polymerase to move in the opposite direction of the replication fork.

THE REPLICATION FORK

- The new DNA strand directed by this template is known as the **lagging strand**.
- This strand of DNA must be synthesized in a discontinuous fashion.

THE REPLICATION FORK

- Synthesis of the lagging strand must wait for movement of the replication fork to expose a substantial length of template before it can be replicated.

THE REPLICATION FORK

- Each time a substantial length of the template is exposed, DNA synthesis is initiated and continues until it reaches the 5' end of the previous newly synthesized fragment of lagging strand DNA.

THE REPLICATION FORK

- The resulting short fragments of new DNA formed on the lagging strand are called Okazaki fragments.
- They vary in length from 1000 to 2000 nucleotides in bacteria and from 100 to 400 nucleotides in eukaryotes.

THE REPLICATION FORK

- Photo lagging strand

THE REPLICATION FORK

- Shortly after being synthesized, Okazaki fragments are covalently joined together to generate a continuous, intact strand of new DNA.
- Okazaki fragments are therefore transient intermediates in DNA replication.

THE RNA PRIMER

- All DNA polymerases require a primer with a free 3'-OH.
- They cannot initiate the synthesis of new DNA strand de novo.
- Then how?

THE RNA PRIMER

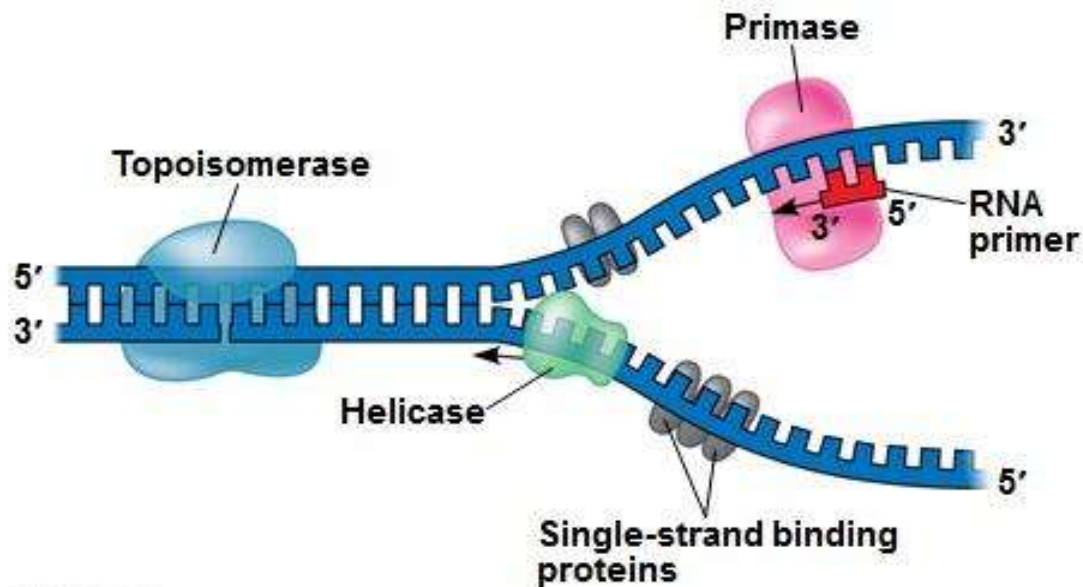
- To accomplish this, the cell takes advantage of the ability of RNA polymerases to do what DNA polymerases cannot: start new RNA chains de novo.

THE RNA PRIMER

- **Primase** is a specialized RNA polymerase dedicated to making short RNA primers (5–10 nucleotides long) on a ssDNA template.
- These primers are then extended by DNA polymerase.

THE RNA PRIMER

Figure 16.13



THE RNA PRIMER

- Both the leading and lagging strands require primase to initiate DNA synthesis.
- Each leading strand requires only a single RNA primer.

THE RNA PRIMER

- The discontinuous synthesis of the lagging strand means that new primers are needed for each Okazaki fragment.
- Synthesis of the lagging strand can require hundreds of Okazaki fragments and their associated RNA primers.

THE RNA PRIMER

- Primase activity is dramatically increased when it associates with another protein that acts at the replication fork called **DNA Helicase**.

THE RNA PRIMER

- This protein unwinds the DNA at the replication fork, creating an ssDNA template that can be acted on by primase.

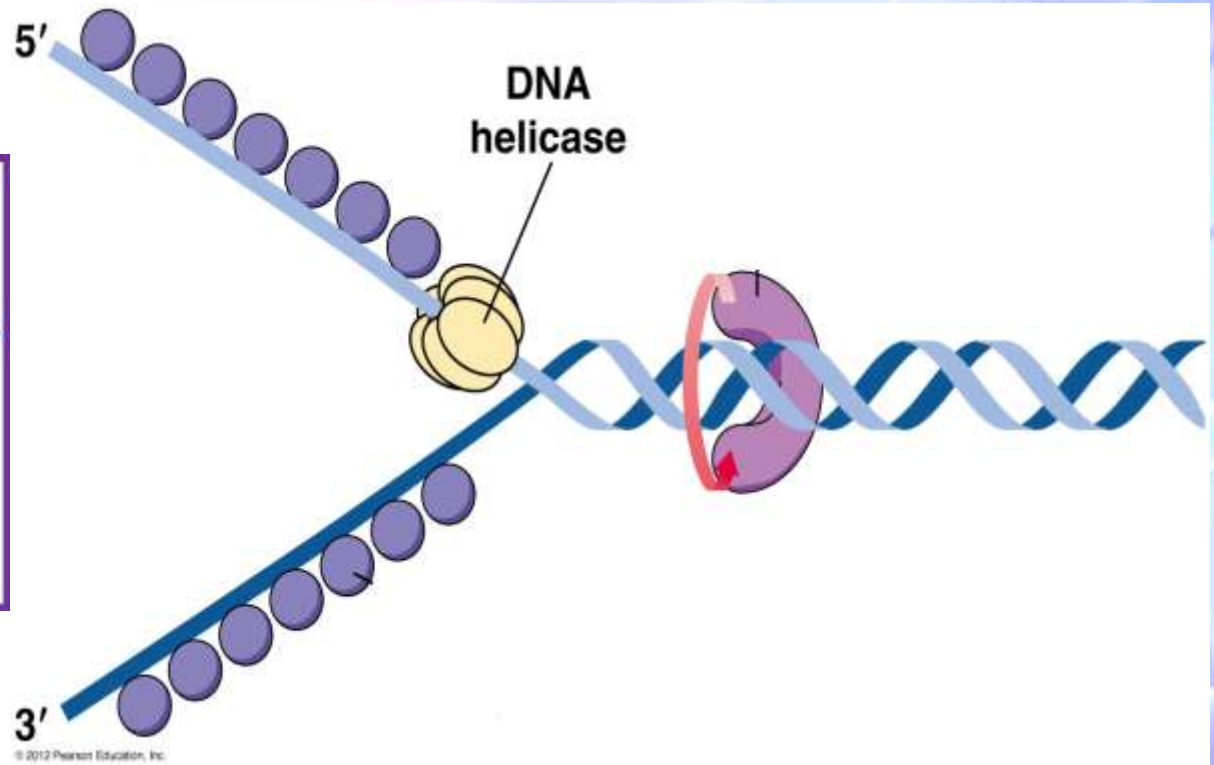
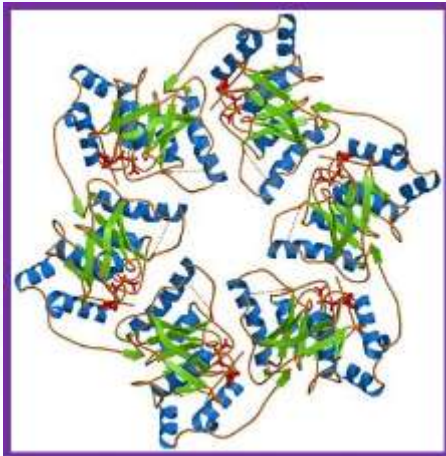
THE DNA HELICASE

- DNA polymerases are unable to separate the two strands of duplex DNA.
- Therefore a third class of enzymes, called **DNA Helicases** catalyze the separation of the two strands of duplex DNA at the replication fork.

THE DNA HELICASE

- DNA helicases are hexameric proteins that assume the shape of a ring.
- This ring encircles one of the two single strands at the replication fork adjacent to the single-stranded:double-stranded junction.

THE DNA HELICASE



THE DNA HELICASE

- DNA helicases found at replication forks exhibit high processivity because they encircle the DNA.
- They associate with the DNA and unwind multiple base pairs of DNA.

THE DNA HELICASE

- Release of the helicase from the DNA therefore requires the opening of the hexameric protein ring, which is a rare event.
- However, the helicase can dissociate when it reaches the end of the DNA strand.

THE DNA HELICASE

- Of course, this arrangement of enzyme and DNA poses problems for the binding of the DNA helicase to the DNA strand in the first place.

THE DNA HELICASE

- Thus, there are specialized mechanisms that open the DNA helicase (hexameric) ring and place it around the DNA before re-forming the ring.

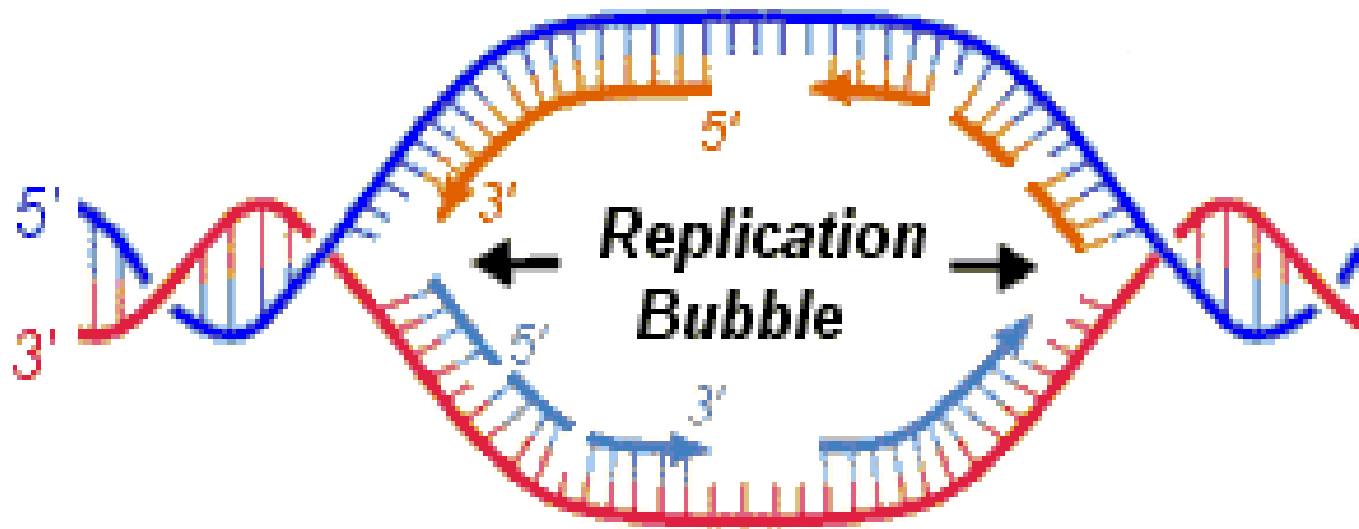
THE DNA HELICASE

- Each DNA helicase moves along ssDNA in a defined direction.
- This property is referred to as the **polarity** of the DNA helicase.

THE DNA HELICASE

- DNA helicases can have a polarity of either $5' \rightarrow 3'$ or $3' \rightarrow 5'$. This direction is always defined according to the strand of DNA bound rather than the strand that is displaced.

THE DNA HELICASE



(c) 2000 Chemis

TOPOISOMERASES

- The action of DNA helicase results in the introduction of supercoils in the DNA duplex just ahead of the replication fork.

TOPOISOMERASES

- As the strands of DNA are separated at the replication fork, the dsDNA in front of the fork becomes increasingly positively supercoiled.

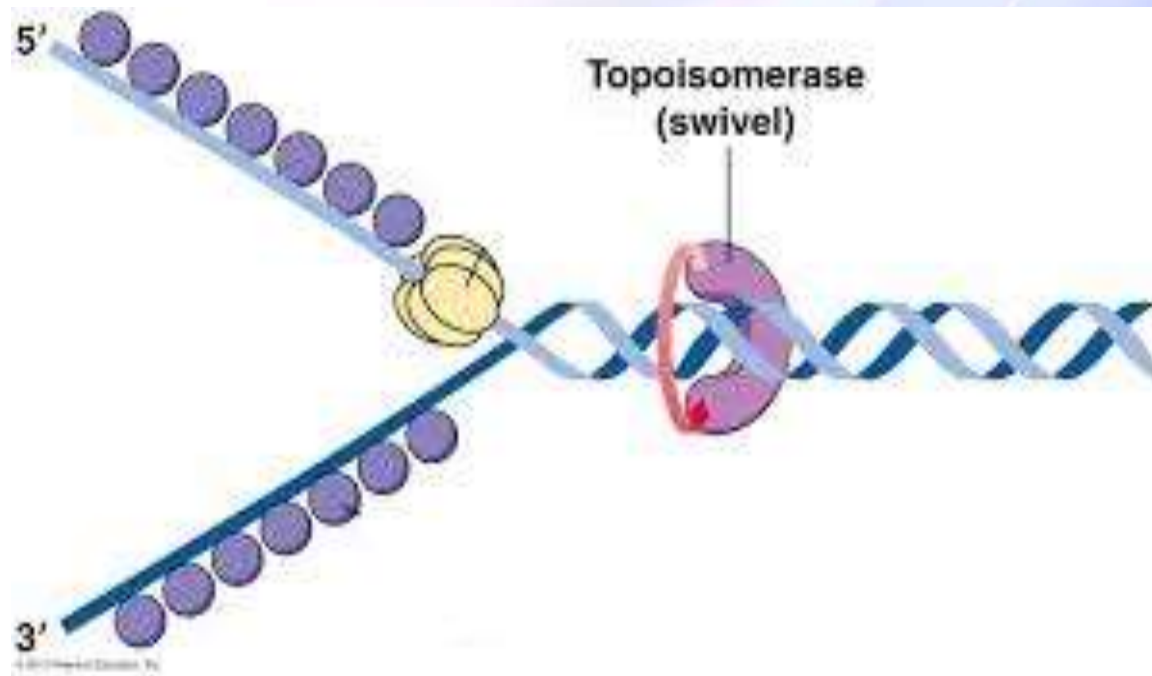
TOPOISOMERASES

- If there were no mechanism to relieve the accumulation of these supercoils, the replication machinery would grind to a halt due to the mounting strain placed on the DNA in front of the replication fork.

TOPOISOMERASES

- These supercoils are removed by the action of **Topoisomerases** that act on the unreplicated dsDNA in front of the replication fork.

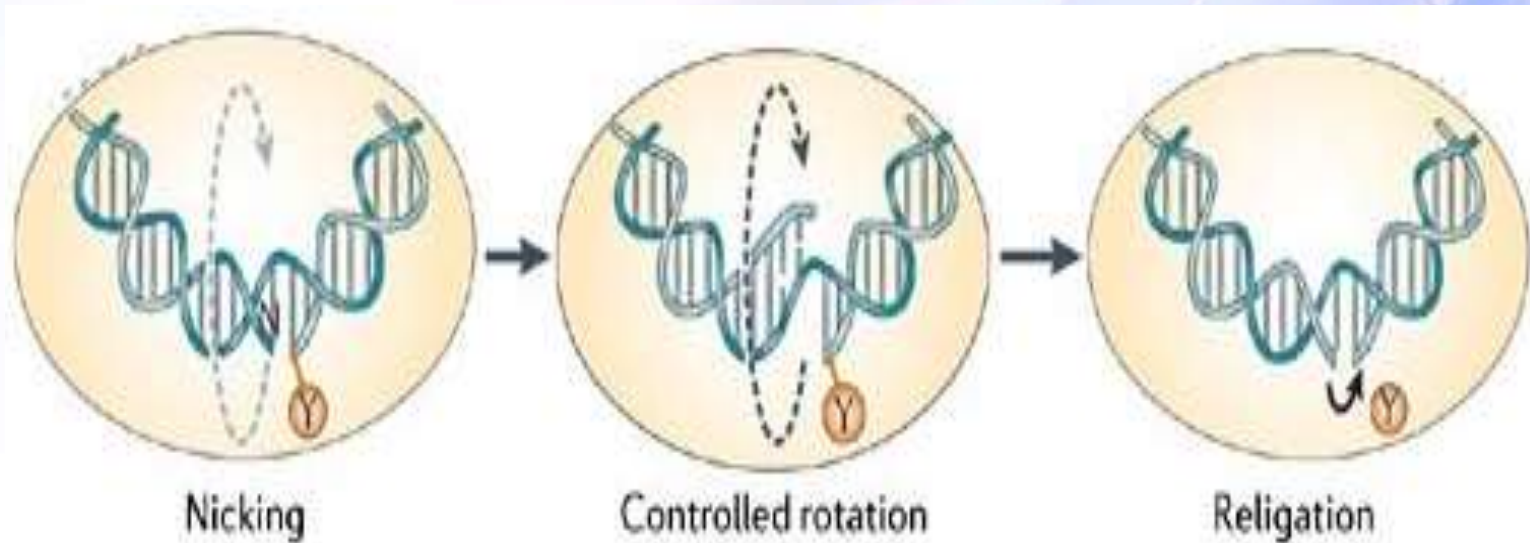
TOPOISOMERASES



TOPOISOMERASES

- These enzymes do this by breaking either one or both strands of the DNA without letting go of the DNA and passing the same number of DNA strands through the break.

TOPOISOMERASES



TOPOISOMERASES

- This action relieves the accumulation of supercoils.
- In this way, topoisomerases act as a “swivelase” that prevents the accumulation of supercoils ahead of the replication fork.

INITIATION OF REPLICATION

- The initial formation of a replication fork requires the separation of the two strands of the DNA duplex to provide the ssDNA.

INITIATION OF REPLICATION

- ssDNA is required for DNA helicase binding and to act as a template for the synthesis of both the RNA primer and new DNA.

INITIATION OF REPLICATION

- Although DNA strand separation (DNA unwinding) is most easily accomplished at chromosome ends, but DNA synthesis generally initiates at internal regions.

INITIATION OF REPLICATION

- As the circular chromosomes lack the chromosome ends so it makes internal DNA unwinding essential for the replication initiation.

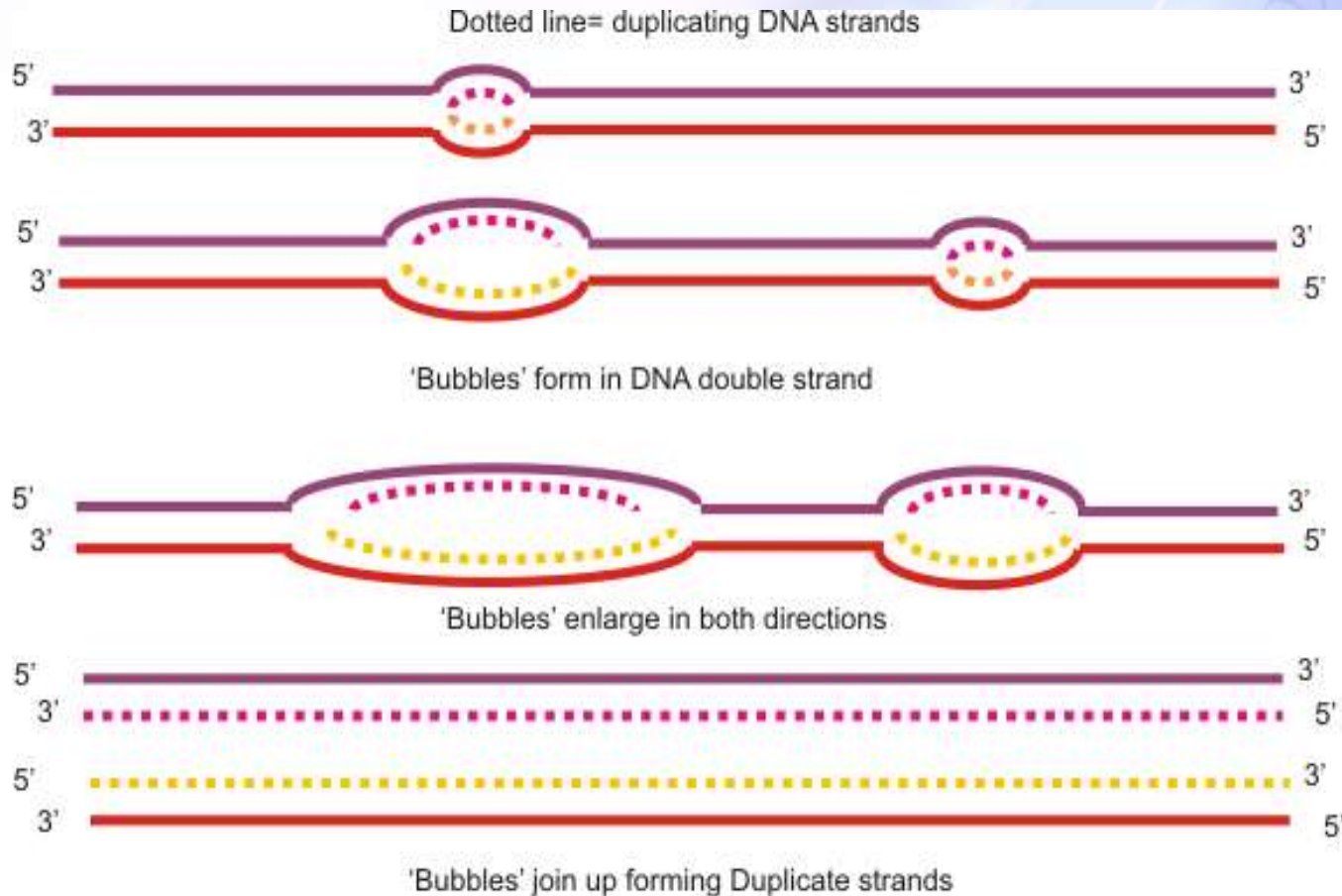
INITIATION OF REPLICATION

- The specific sites at which DNA unwinding and initiation of replication occur are called **Origins of Replication**.

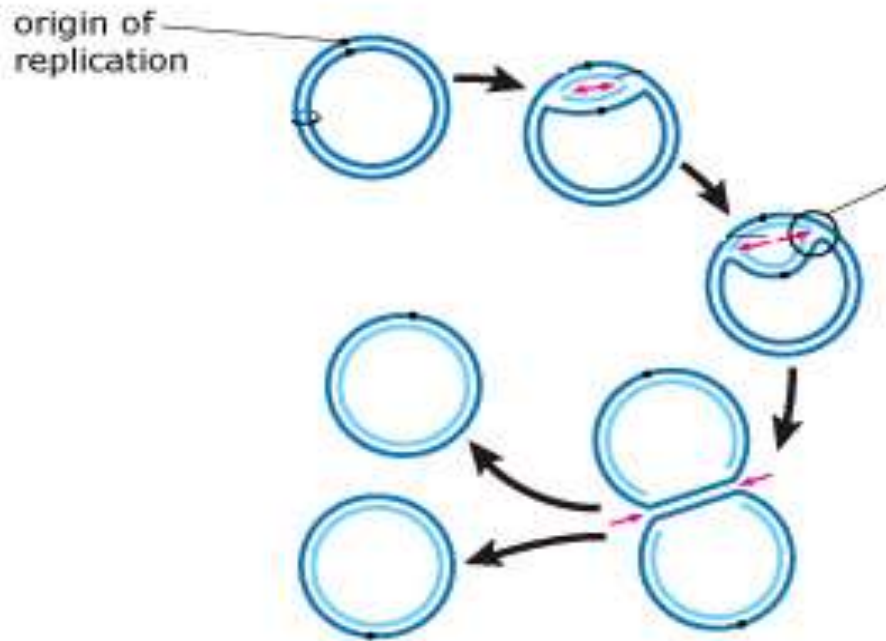
INITIATION OF REPLICATION

- Depending on the organism, there may be as few as one or as many as thousands of origins per chromosome.

INITIATION OF REPLICATION



INITIATION OF REPLICATION



From *Biology* by Campbell and Reece
© 2008 Pearson Education, Inc.

THE REPLICON MODEL

- How the initiation of replication takes place?
- It was explained by Francois Jacob, Sydney Brenner and Jacques Cuzin in 1963.

THE REPLICON MODEL

- This is called as **The Replicon Model** of replication Initiation.
- They defined all of the DNA replicated from a particular origin of replication as a **replicon**.

THE REPLICON MODEL

- As the single chromosome found in *E. coli* cells has only one origin of replication, the entire chromosome is a single replicon.

THE REPLICON MODEL

- In contrast, the presence of multiple origins of replication divides each eukaryotic chromosome into multiple replicons; one for each origin of replication.

THE REPLICON MODEL

- The replicon model proposed two components that controlled the initiation of replication; the replicator and the initiator.

THE REPLICON MODEL

- The replicator is defined as the cis-acting DNA sequences that are sufficient to direct the initiation of DNA replication.

THE REPLICON MODEL

- This is in contrast to the origin of replication, which is the physical site on the DNA where the DNA is unwound and DNA synthesis initiates.

THE REPLICON MODEL

- Although the origin of replication is always part of the replicator, sometimes the origin of replication is only a fraction of the DNA sequences required to direct the initiation of replication (the replicator).

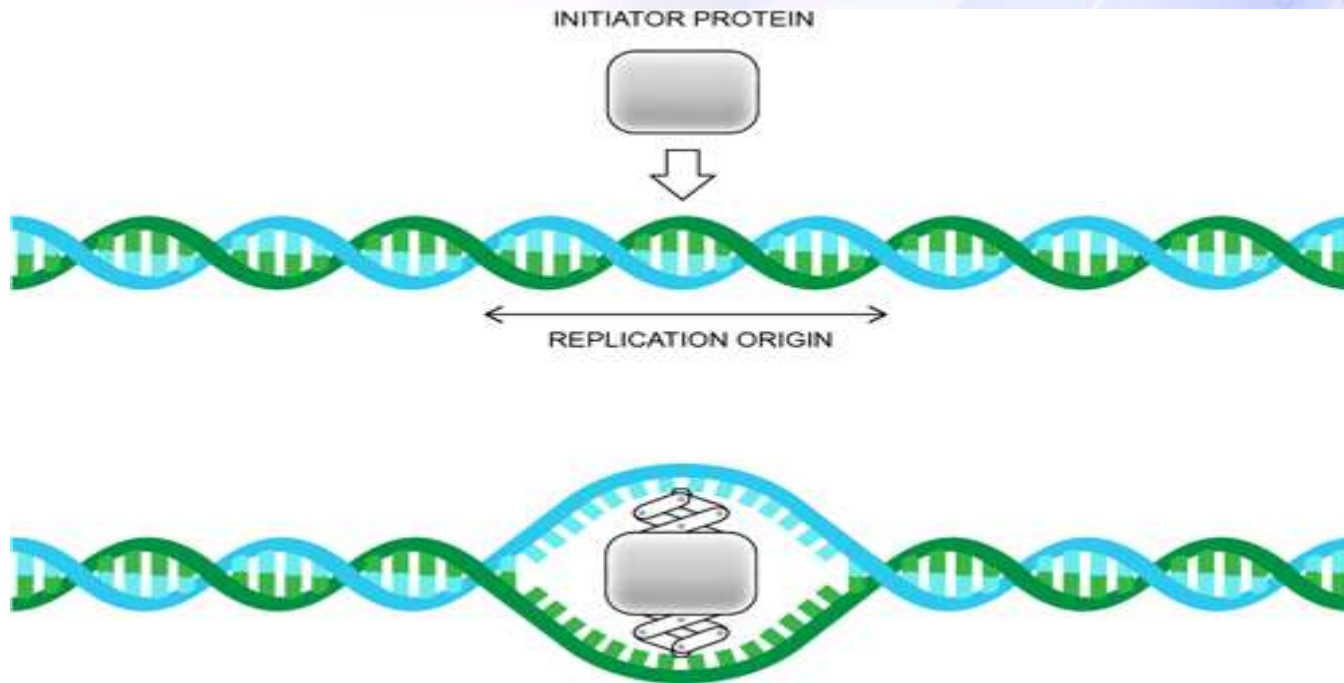
THE REPLICON MODEL

- The second component of the replicon model is the **initiator** protein.
- This protein specifically recognizes a DNA element in the replicator and activates the initiation of replication.

THE REPLICON MODEL

- All initiator proteins select the sites that will become origins of replication.
- The initiator protein is the only sequence-specific DNA binding protein involved in the initiation of replication.

THE REPLICON MODEL



THE REPLICON MODEL

- All the remaining proteins other than initiator protein, required for replication initiation do not bind to a DNA sequence specifically.

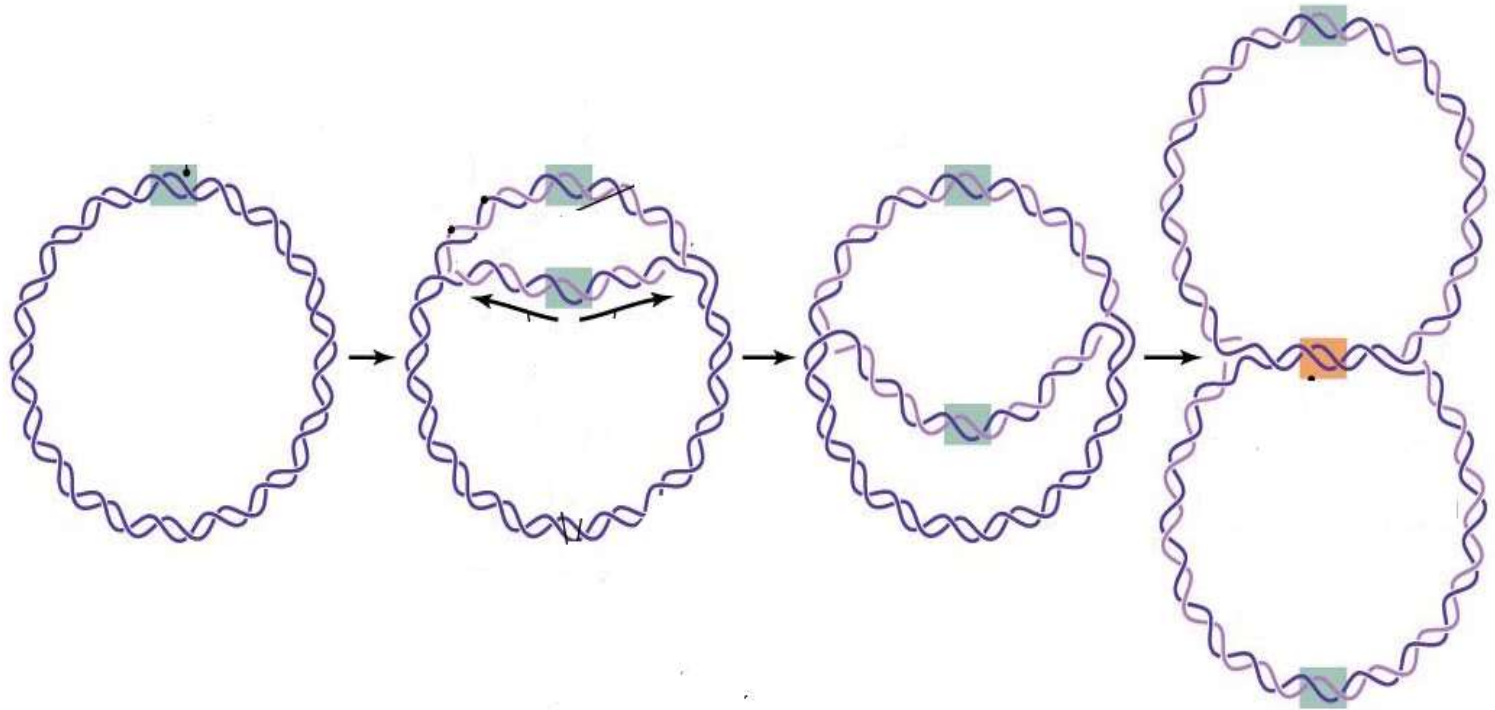
FINISHING REPLICATION

- Completion of DNA replication requires a set of specific events.
- These events are different for circular chromosomes and linear chromosomes.

FINISHING REPLICATION

- In case of circular chromosome, the conventional replication fork machinery replicates the entire molecule, but the resulting daughter molecules are topologically linked to each other.

FINISHING REPLICATION



FINISHING REPLICATION

- While in case of linear chromosome, the replication fork machinery cannot complete replication of the very ends of linear chromosomes.

FINISHING REPLICATION

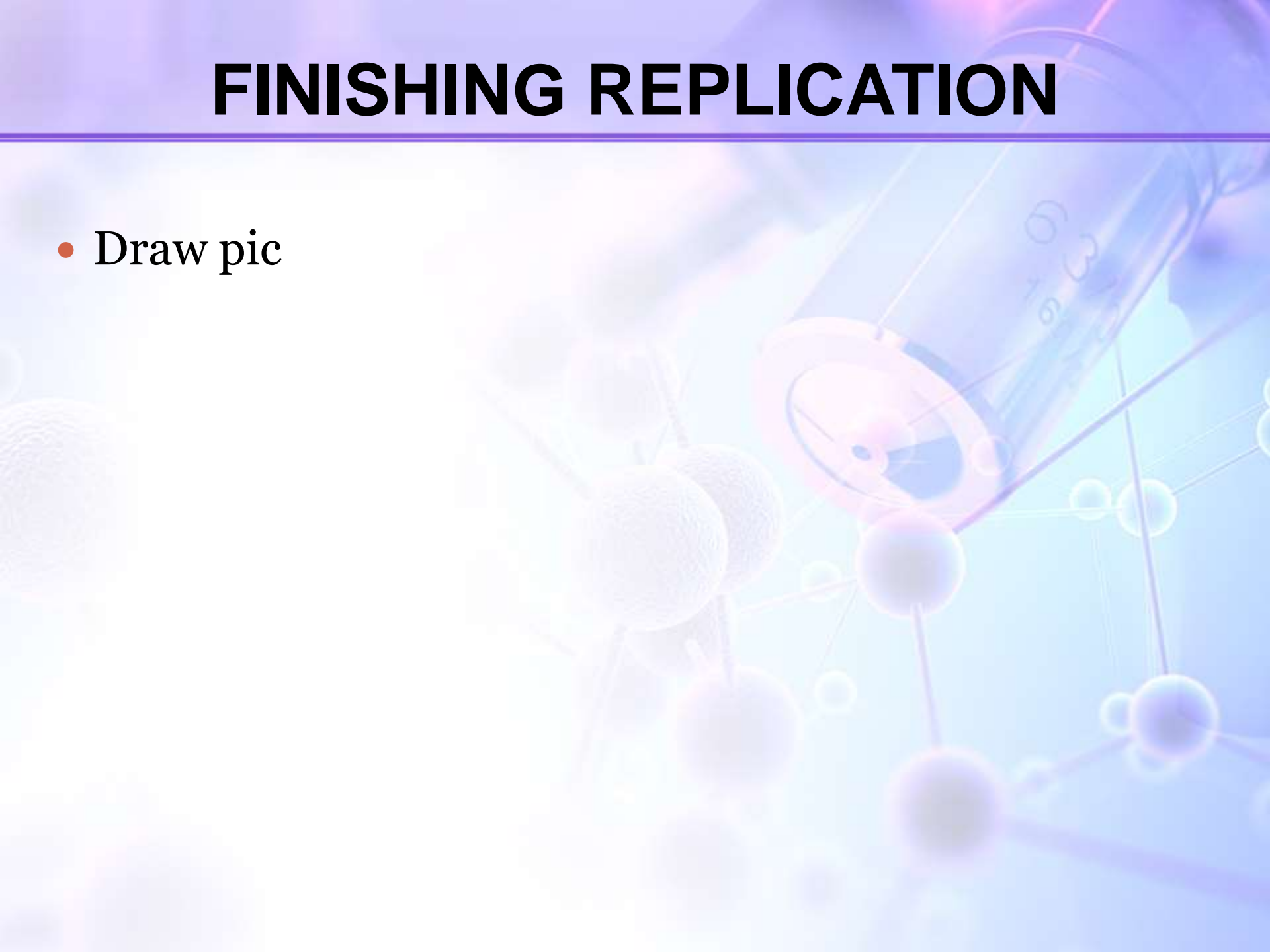
- Therefore, organisms containing linear chromosomes have developed novel strategies to replicate their chromosome ends.

FINISHING REPLICATION

- After replication of a circular chromosome is complete, the resulting daughter DNA molecules remain linked together as **catenanes**.

FINISHING REPLICATION

- Draw pic



FINISHING REPLICATION

- To segregate these chromosomes into separate daughter cells, the two circular DNA molecules must be disengaged from each other or “decatenated.”

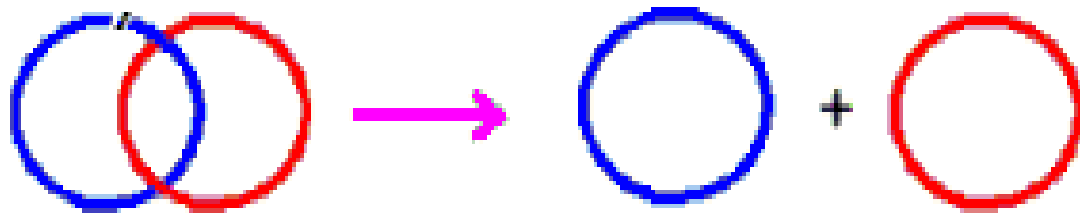
FINISHING REPLICATION

- This separation is accomplished by the action of **Type II Topoisomerases**.

TYPE II TOPOISOMERASES

- Type II Topoisomerases are the enzymes which have the ability to break a dsDNA molecule and pass a second dsDNA molecule through this break.

TYPE II TOPOISOMERASES



DRAW

TYPE II TOPOISOMERASES

- So this reaction can easily decatenate the two circular daughter chromosomes and allow their segregation into separate cells.

TYPE II TOPOISOMERASES

- The activity of type II topoisomerases is also critical to the segregation of large linear molecules.

TYPE II TOPOISOMERASES

- Although there is no inherent topological linkage after the replication of a linear molecule, the large sized chromosomes necessitates the intricate folding of the DNA into loops which are attached to a protein scaffold.

TYPE II TOPOISOMERASES

- These attachments lead to many of the same problems that circular chromosomes have after replication.
- So type II topoisomerases allow these linked DNAs to be separated.

TYPE II TOPOISOMERASES

- So as in the case of circular chromosomes, type II topoisomerases also allow these linked DNAs to be separated.

TELOMERASE

- The requirement for an RNA primer to initiate all new DNA synthesis creates a dilemma for the replication of the ends of linear chromosomes.
- This is called as the **end replication problem**.

TELOMERASE

- This difficulty is not observed during the duplication of the leading-strand template.
- Because it requires only one RNA primer which completes the DNA synthesis up to extreme terminus of the strand.

TELOMERASE

- In contrast, the requirement for multiple primers to complete lagging-strand synthesis means that a complete copy of its template cannot be made.

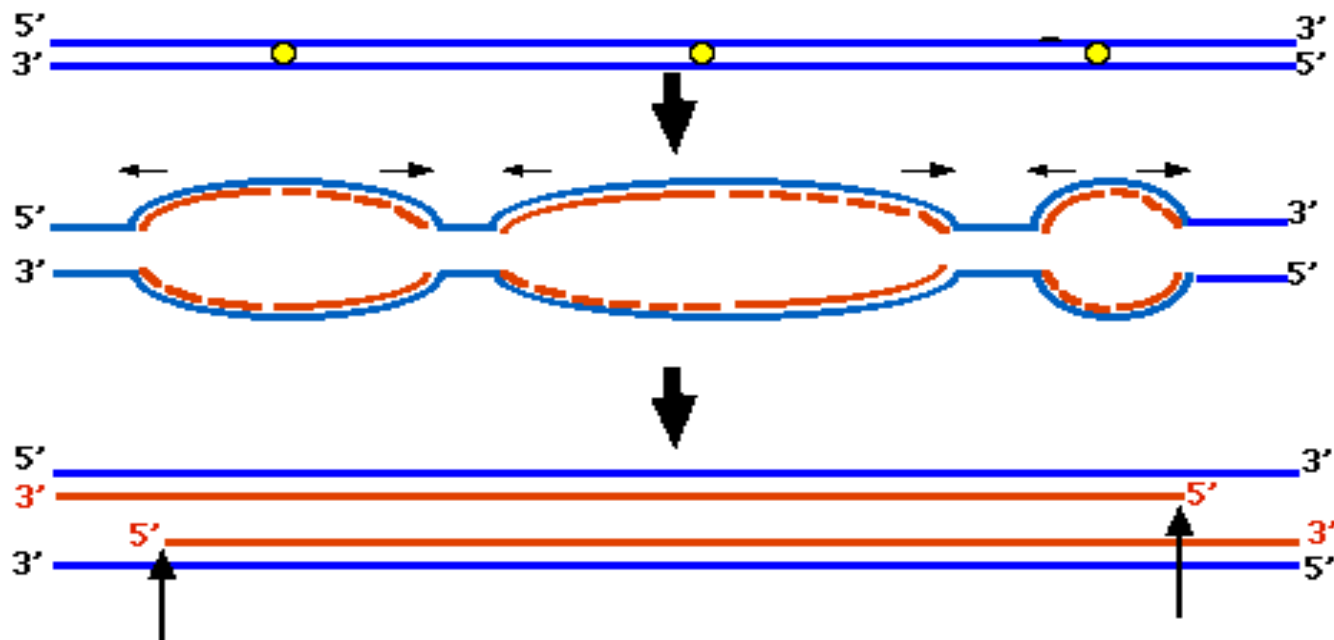
TELOMERASE

- Even if the end of the last RNA primer for Okazaki fragment synthesis anneals to the final base of the lagging-strand template.

TELOMERASE

- Once this RNA molecule is removed, there will remain a short region (the size of the RNA primer) of un-replicated ssDNA at the end of the chromosome.

TELOMERASE



TELOMERASE

- Telomerase is a remarkable enzyme that includes multiple protein subunits and an RNA component.
- Like all other DNA polymerases, telomerase acts to extend the 3' end of its DNA substrate.

TELOMERASE

- But unlike most DNA polymerases, telomerase does not need an exogenous DNA template to direct the addition of new dNTPs.

TELOMERASE

- Instead, the RNA component of telomerase serves as the template for adding the telomeric sequence to the 3' terminus at the end of the chromosome

TELOMERASE

- Telomerase specifically elongates the 3'-OH of telomeric ssDNA sequences using its own RNA as a template.
- As a result, the newly synthesized DNA is single-stranded.

TELOMERASE

- So when telomerase acts on the 3' end of the telomere, it extends only one of the two strands of the chromosome.
- How is then 5' end extended?

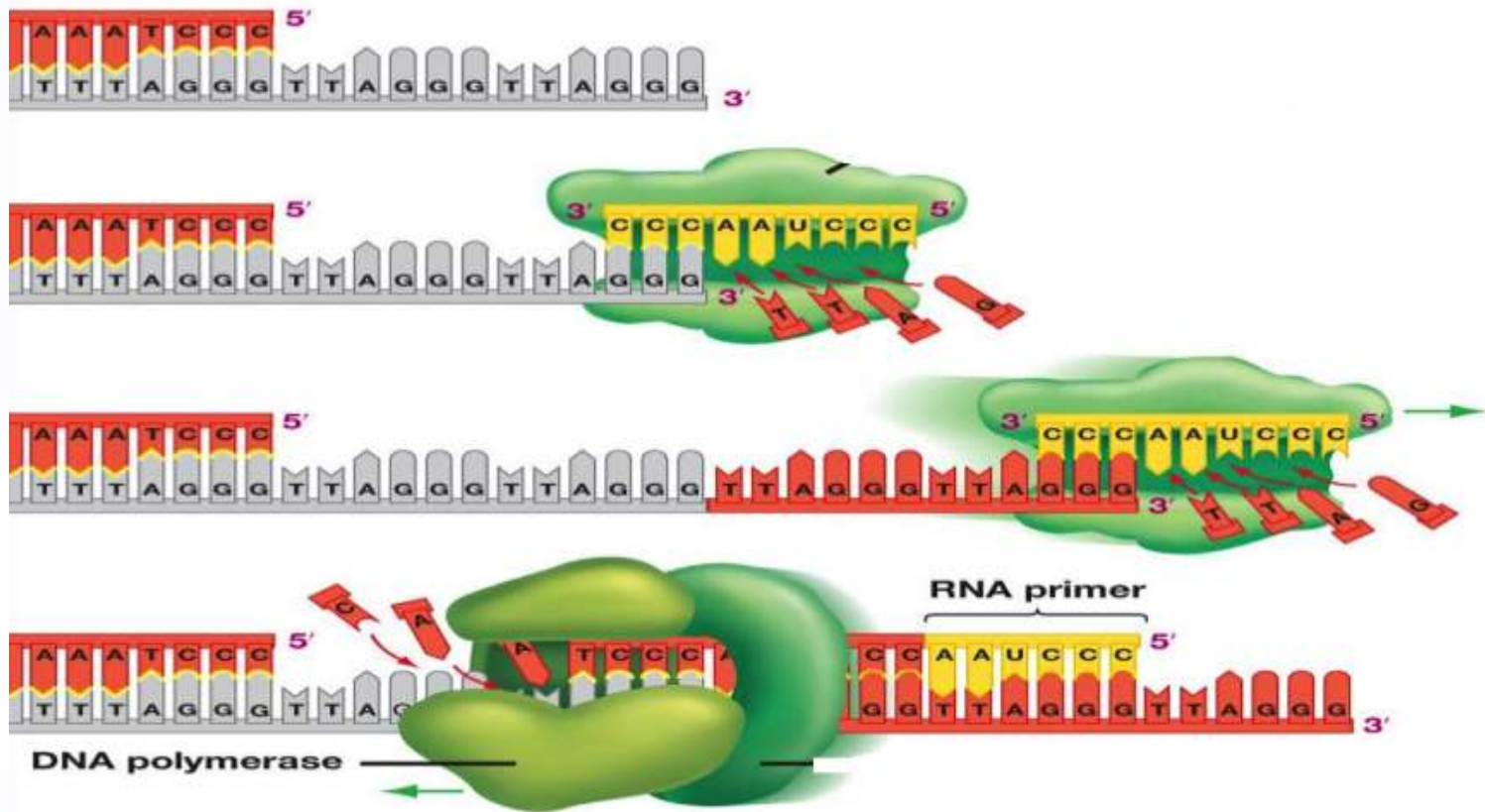
TELOMERASE

- This is accomplished by the lagging-strand DNA replication machinery.
- By providing an extended 3' end, telomerase provides additional template for the lagging-strand replication machinery.

TELOMERASE

- By synthesizing and extending RNA primers using the telomerase extended 3' end as a template, the cell can effectively increase the length of the 5' end of the chromosome as well.

TELOMERASE



DNA MUTATIONS

- DNA mutation can be defined as a permanent transmissible change in the genetic material (DNA/RNA).

DNA MUTATIONS

- In other words, it is a permanent change in the nucleotide sequence of the genome of an organism.

DNA MUTATIONS

- DNA can be easily damaged even under normal physiological conditions.
- Many different kinds of chemical and physical agents can damage DNA.

DNA MUTATIONS

- Some of these agents are endogenous which are produced inside the cells as a result of normal metabolic pathways.

DNA MUTATIONS

- While some others are exogenous agents which come from the surrounding environment.

DNA MUTATIONS

- On one hand, DNA stability is required to ensure that the genetic information may pass accurately from one generation to the next
- It is also required for the correct functioning of thousands of genes.

DNA MUTATIONS

- On the other hand the genetic variation is needed to drive evolution.
- If this variation would be lacking, the new species, including humans, would have not arisen.

DNA MUTATIONS

- So the life and biodiversity depend on a happy balance between DNA damage (mutation) and its repair.

END

NATURE OF MUTATIONS

- DNA mutations may be very simple (single base change) or very complex and may include several thousands of nucleotides.

NATURE OF MUTATIONS

- The simplest mutations are switches of one base for another. There are two kinds of such mutations which include:-
 - Transitions
 - Transversions

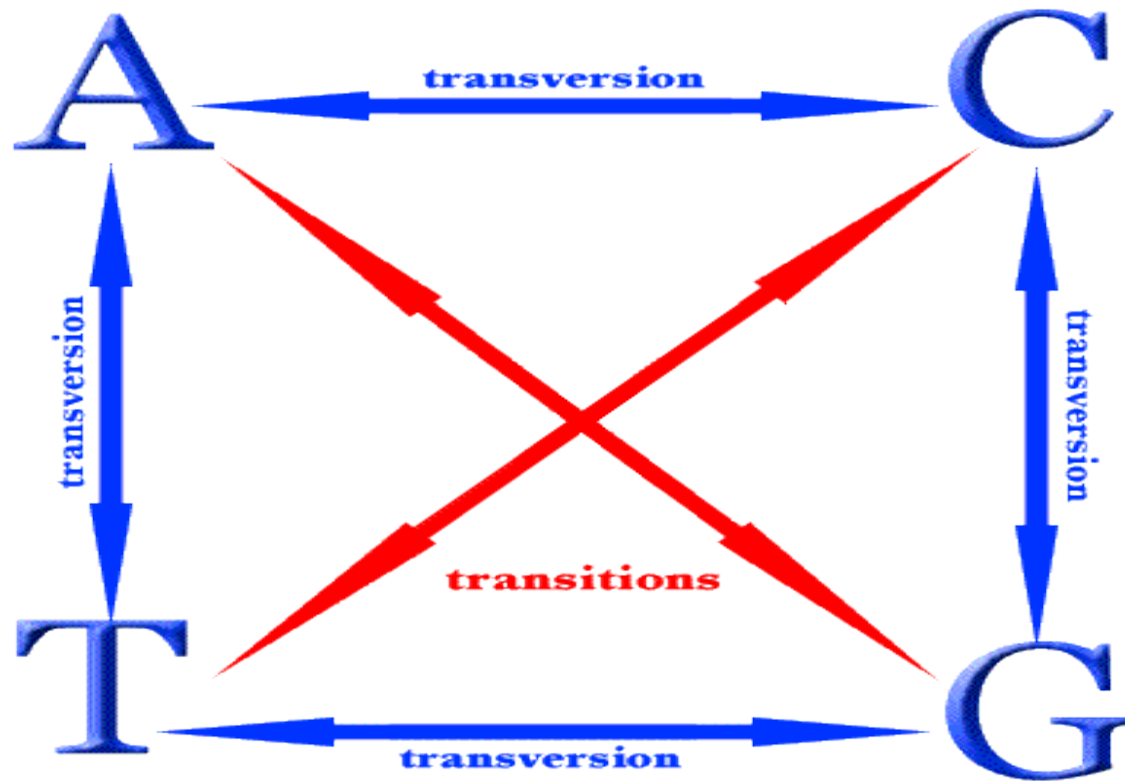
NATURE OF MUTATIONS

- Transitions are pyrimidine-to-pyrimidine and purine-to-purine substitutions, such as thymine (T) to cytosine (C) and adenine (A) to guanine (G).

NATURE OF MUTATIONS

- Transversions are pyrimidine-to-purine and purine-to-pyrimidine substitutions, such as T to G or A and A to C or T.

NATURE OF MUTATIONS

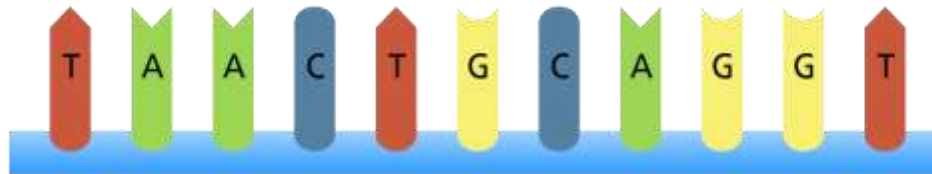


NATURE OF MUTATIONS

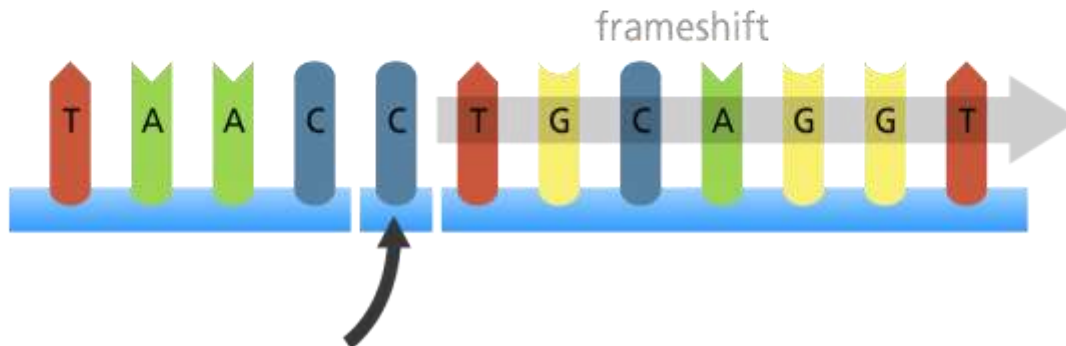
- Other simple mutations are **insertions** or **deletions** of a nucleotide or a small number of nucleotides.

NATURE OF MUTATIONS

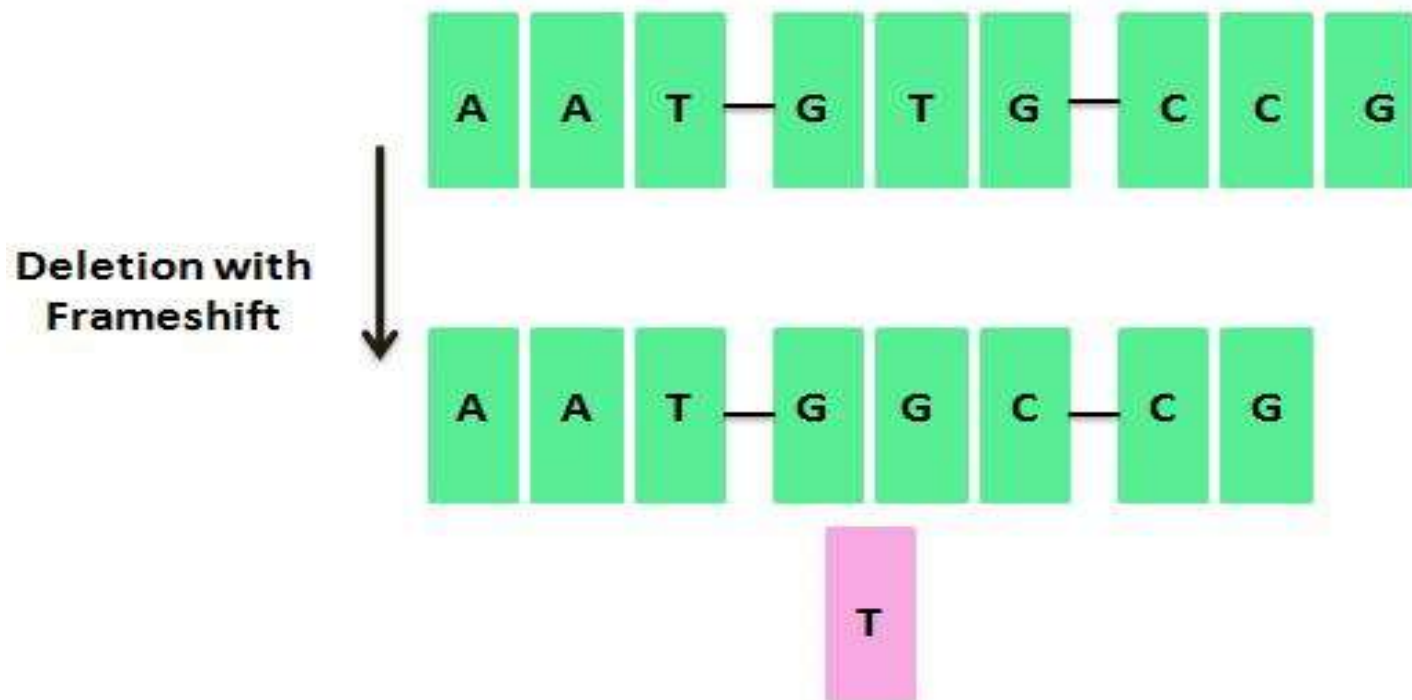
Original sequence



Insertion



NATURE OF MUTATIONS



NATURE OF MUTATIONS

- All such mutations that alter a single nucleotide are called **point mutations**.

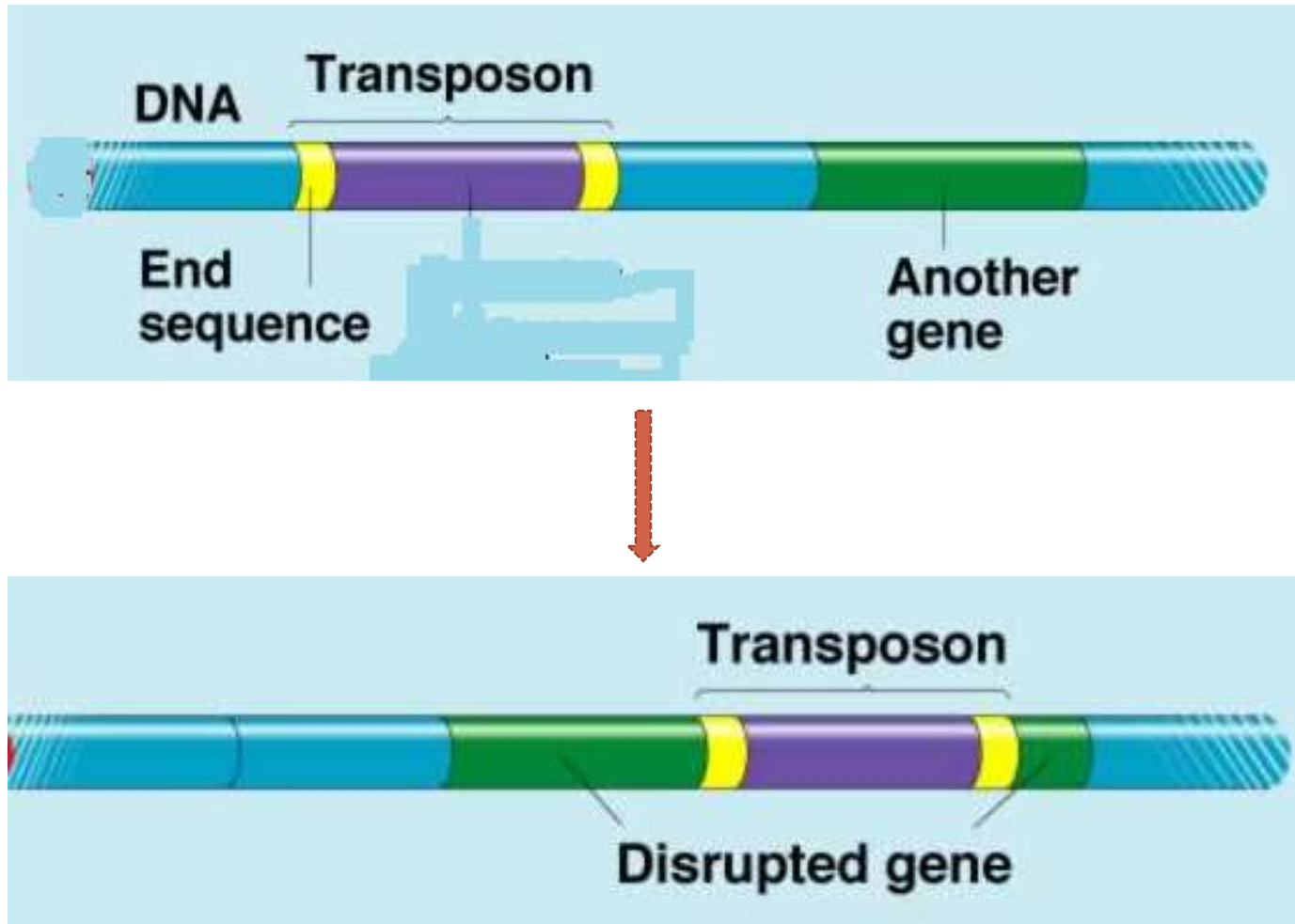
NATURE OF MUTATIONS

- Other kinds of mutations cause more drastic changes in DNA, such as extensive insertions and deletions and gross rearrangements of chromosome structure.

NATURE OF MUTATIONS

- Such changes might be caused, for example, by the insertion of a transposon, which typically places many thousands of nucleotides of foreign DNA in the coding or regulatory sequences of a gene.

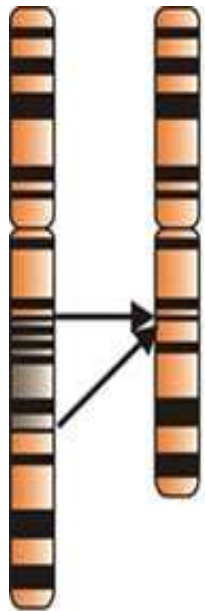
NATURE OF MUTATIONS



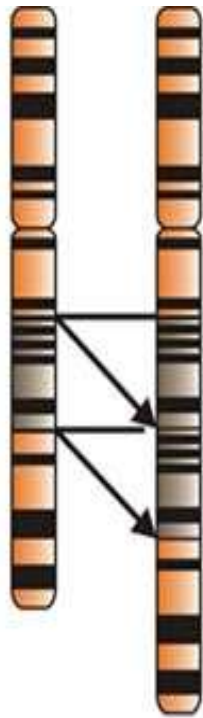
NATURE OF MUTATIONS

- Another type of mutations which are more drastic occur at chromosomal levels.
- These are changes in appearance of the individual chromosomes through mutation-induced rearrangements.

NATURE OF MUTATIONS



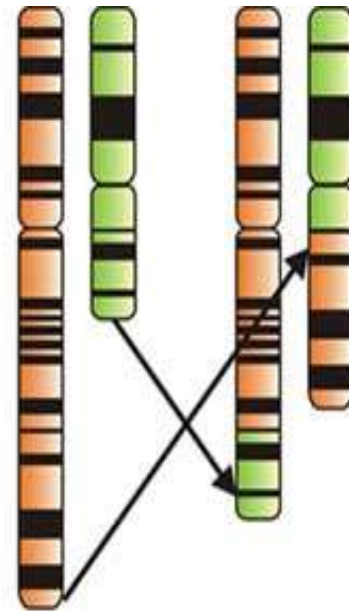
Deletion



Duplication



Inversion



Translocation

REPLICATION ERRORS

- The replication machinery achieves a remarkably high degree of accuracy using a proofreading mechanism, which removes wrongly incorporated nucleotides.

REPLICATION ERRORS

- However, this proofreading is not foolproof.
- Some mis-incorporated nucleotides escape detection and become a mismatch between the newly synthesized strand and the template strand.

REPLICATION ERRORS

- If the misincorporated nucleotide is not subsequently detected and replaced, the sequence change will become permanent in the genome.

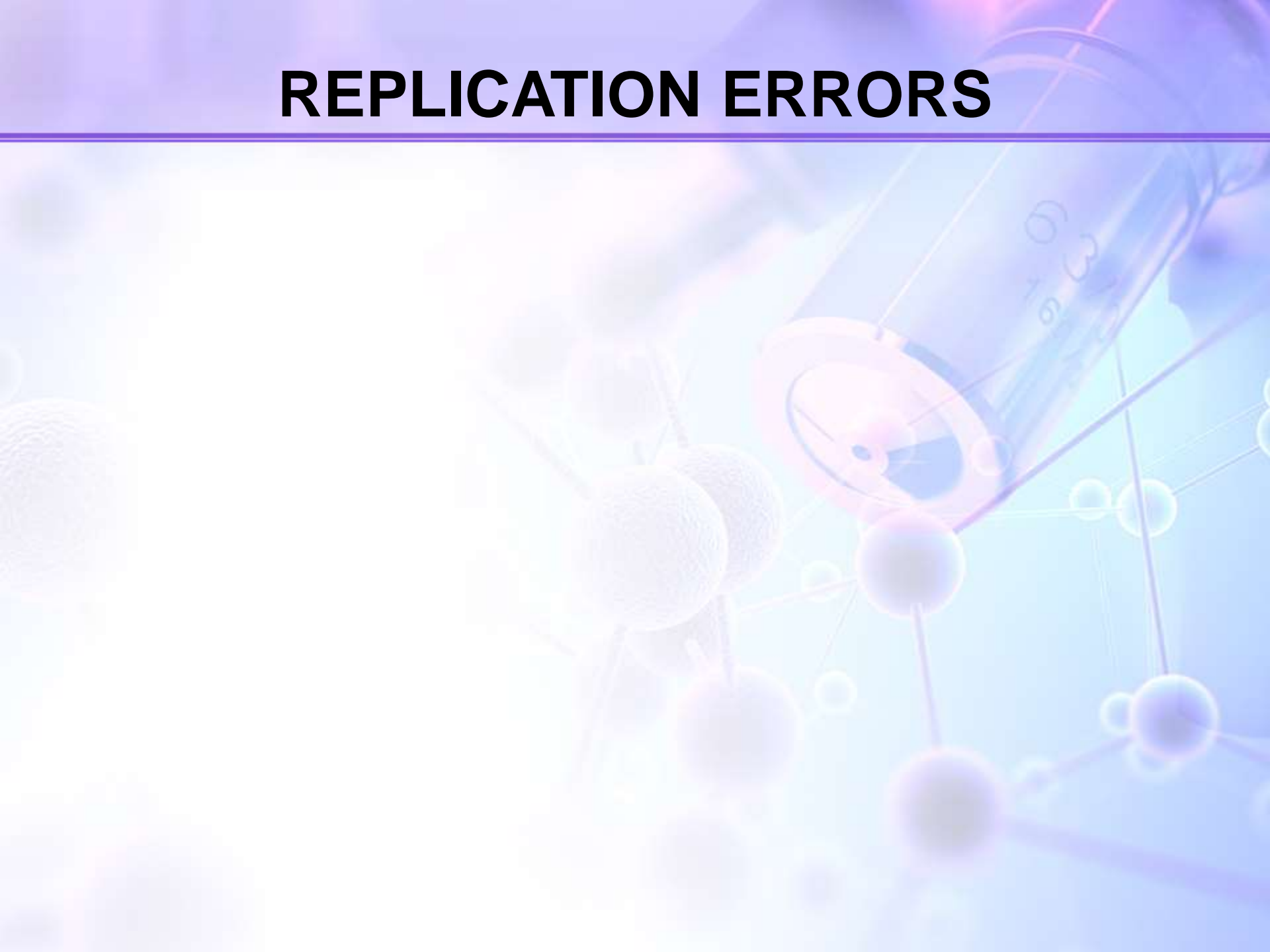
REPLICATION ERRORS

- During a second round of replication, the mis-incorporated nucleotide will direct the incorporation of its complementary nucleotide into the newly synthesized strand.

REPLICATION ERRORS

- At this point, the mismatch will no longer exist; instead, it will have resulted in a permanent change (a mutation) in the DNA sequence.

REPLICATION ERRORS



RADIATION DAMAGE

- Cells are exposed to three types of high energy electromagnetic radiations which can damage their DNA.
- These radiations include:-

RADIATION DAMAGE

- Ultraviolet light
(wavelength 100-400nm)
- X-rays (wavelength 0.01-100nm)
- Gamma rays
(wavelength $<0.01\text{nm}$)
- Later two are ionizing radiations.

RADIATION DAMAGE

- Ultraviolet light is divided into three bands:
 - UV –A (321-400nm)
 - UV –B (296-320nm)
 - UV –C (100-295nm)

RADIATION DAMAGE

- The majority of UV light reaching on earth is UV – A.
- This is least energetic band and so does little damage to DNA.

RADIATION DAMAGE

- UV – B accounts for about 10% of the UV radiation reaching the earth's surface.
- It is responsible for most of the DNA damage in the skin.

RADIATION DAMAGE

- UV – C includes the wavelength of maximum DNA absorbance (260nm).
- So it would cause a great deal of DNA damage to exposed organisms if it were able to penetrate the earth's surface.

RADIATION DAMAGE

- Fortunately, very little UV - C reaches the earth's surface because the ozone layer prevents it from penetration.
- However, during lab studies, the germicidal lamps that produce UV - C light are used.

END

Cyclobutane pyrimidine Dimer

- Two major photoproducts account for nearly all of the UV - induced DNA damage.
- They involve the dimer formation between adjacent pyrimidine bases on the same DNA strand.

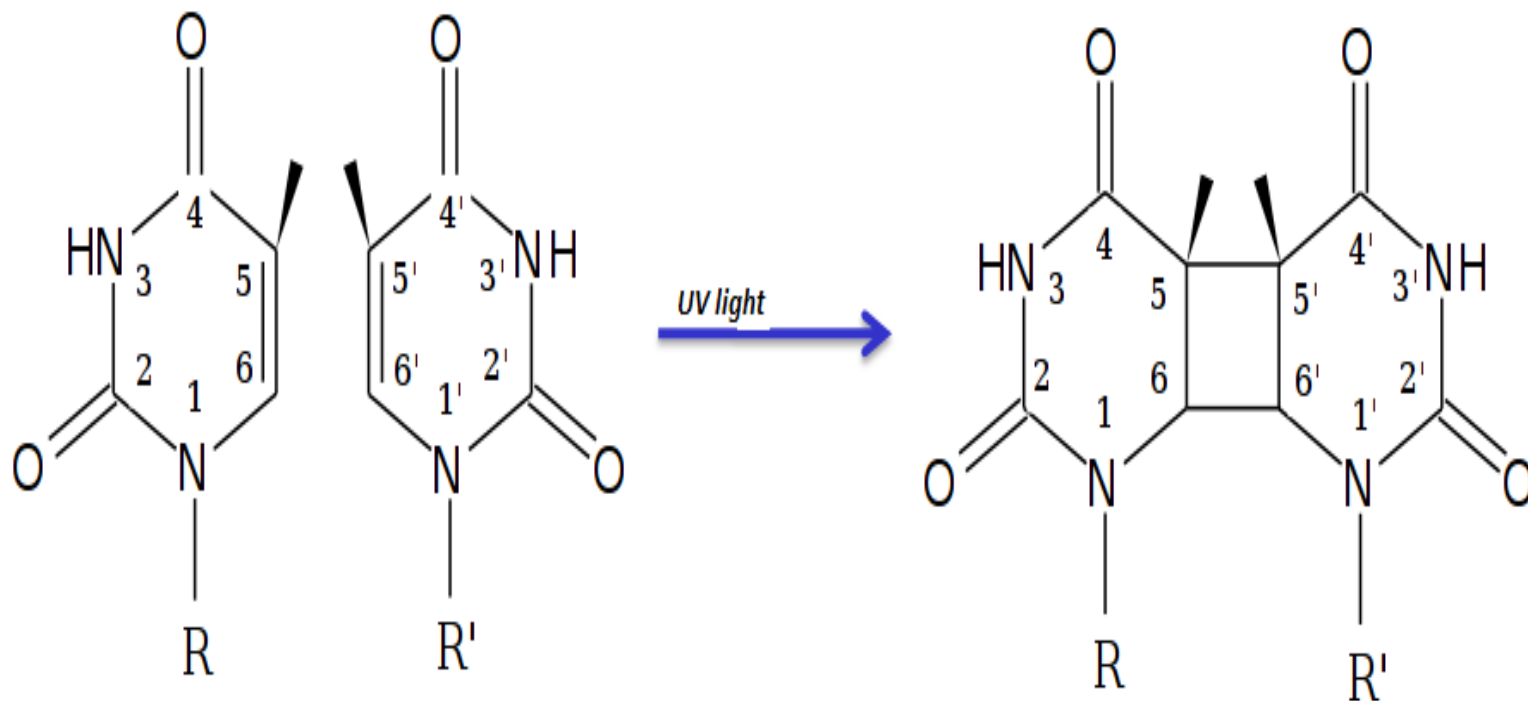
Cyclobutane pyrimidine Dimer

- The first pathway includes the formation of **cyclobutane pyrimidine dimer (CPD)** which accounts for about 75% of all the UV - induced damage.

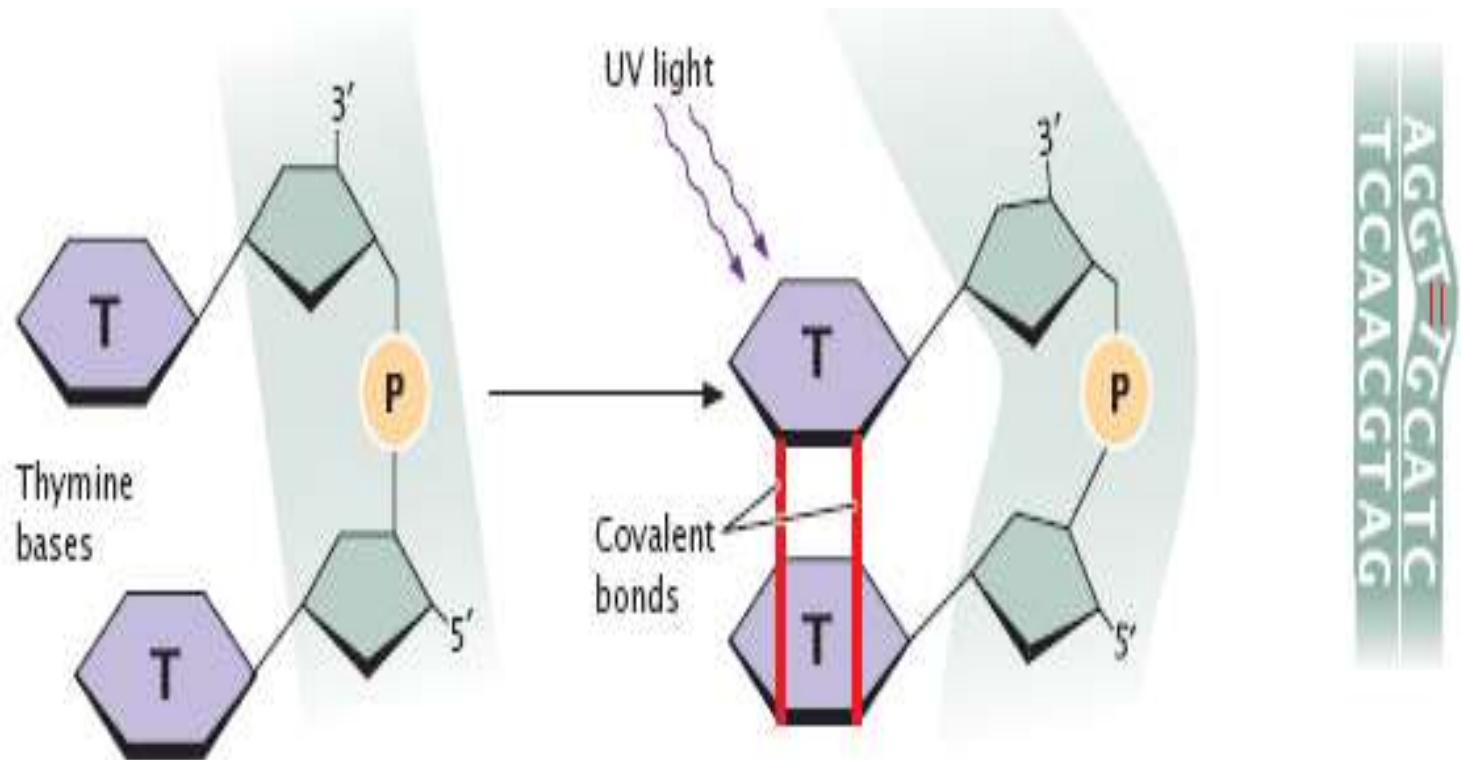
Cyclobutane pyrimidine Dimer

- The cyclobutane ring is generated by forming one bond between C-5 atoms and another between C-6 atoms on adjacent pyrimidine rings.

Cyclobutane pyrimidine Dimer



Cyclobutane pyrimidine Dimer



Cyclobutane pyrimidine Dimer

- The most common cyclobutane pyrimidine dimer is the thymine-thymine ($T < > T$) dimer.
- Cytosine-thymine ($C < > T$) and cytosine-cytosine ($C < > C$) dimers are also formed but at slower rates.

Cyclobutane pyrimidine Dimer

- Structural studies show that:-
- 1) B-DNA can accommodate a single T< >T dimer forcing the helical axis to bend by about 30° towards the major groove.

Cyclobutane pyrimidine Dimer

- 2) The dimer's 3'-thymine can form a normal base pairing with its adenine partner on the complementary strand.

Cyclobutane pyrimidine Dimer

- 3) The interaction between the 5'-thymine and its complementary adenine partner will be weaker than normal because a single hydrogen bond can be formed here.

Cyclobutane pyrimidine Dimer

- Thymine-thymine dimers are often used to study DNA repair systems because they are stable, easy to form and easy to detect.

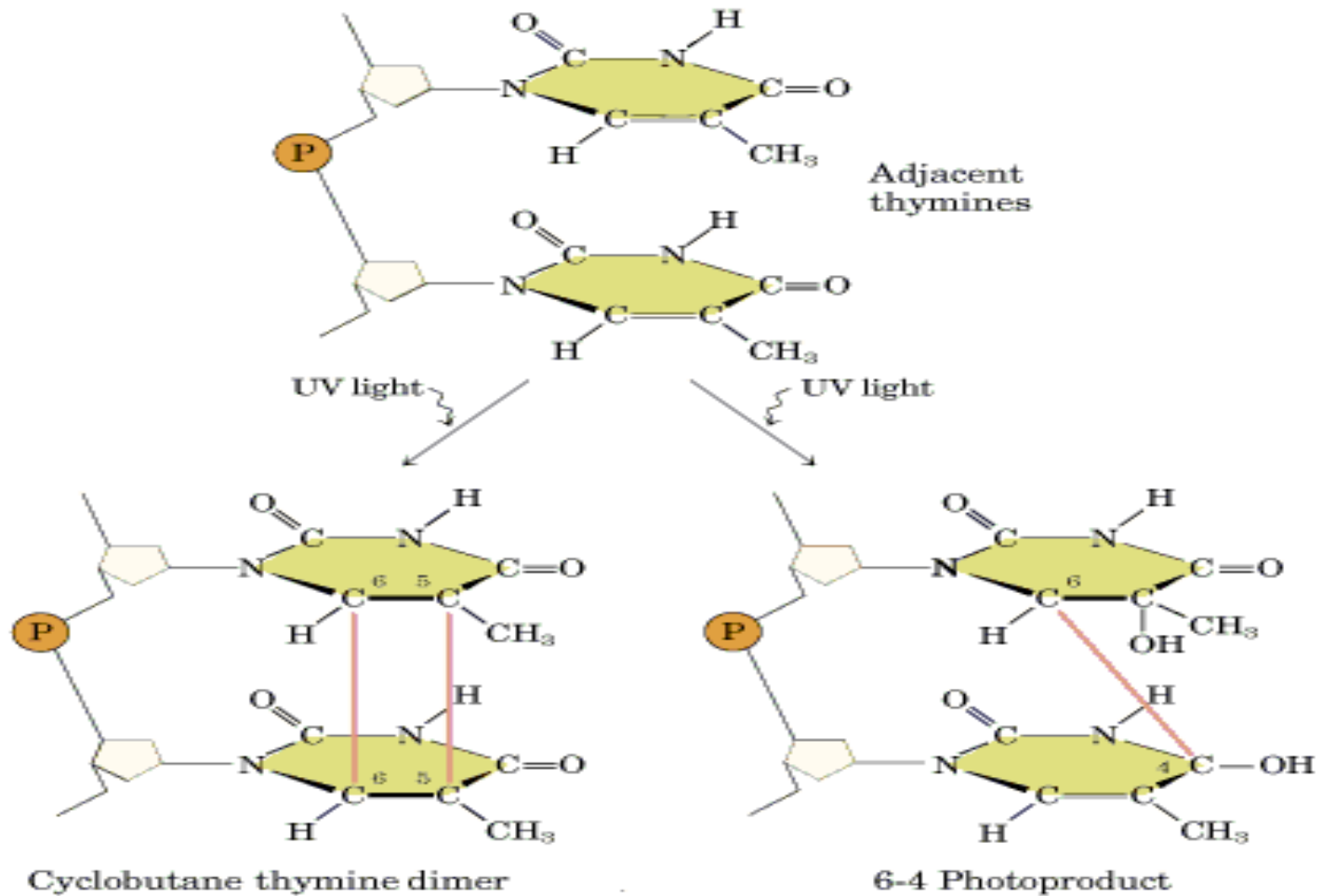
(6-4) Photoproducts

- The second pathway, which accounts for most of the remaining UV-induced DNA damage, produces (6-4) photoproducts.

(6-4) Photoproducts

- A bond is formed between the C-6 atom of the 3'-pyrimidine (either thymine or cytosine) and the C-4 atom of the 5'-pyrimidine (usually cytosine).

(6-4) Photoproducts



(6-4) Photoproducts

- The (6-4) photoproduct causes a major distortion in B-DNA because the two pyrimidine rings are perpendicular to each other.

(6-4) Photoproducts

- If not removed, a pyrimidine dimer or (6-4) photoproduct can interfere with the normal operation of the replication and transcription machinery.

(6-4) Photoproducts

- This interference results in mutations and cell death.
- Even if the lesion is removed, the result can be a mutation.

End

X-rays & gamma rays damage

- Gamma radiations and X-rays (ionizing radiation) are particularly hazardous because they cause double-strand breaks in the DNA, which are difficult to repair.

X-rays & gamma rays damage

- Ionizing radiations directly or indirectly generate many different kinds of DNA lesions.
- Direct damage takes place when DNA or water bound to it absorbs the radiation.

X-rays & gamma rays damage

- Indirect damage takes place when water molecule or any other molecule surrounding the DNA absorb the radiation and form reactive species that then damage the DNA.

X-rays & gamma rays damage

- The DNA lesions caused due to these radiations may be isolated or clustered.
- Many lesions within a few helical turns are called clustered lesions.

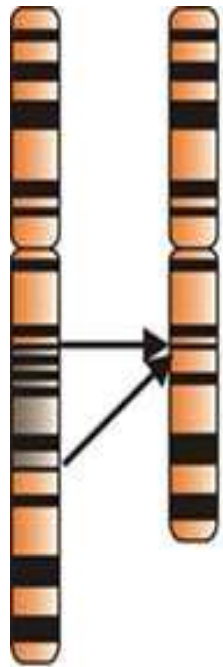
X-rays & gamma rays damage

- One type of clustered lesion, the double-strand break, is generally thought to be the primary reason that ionization radiation is so lethal to cells.

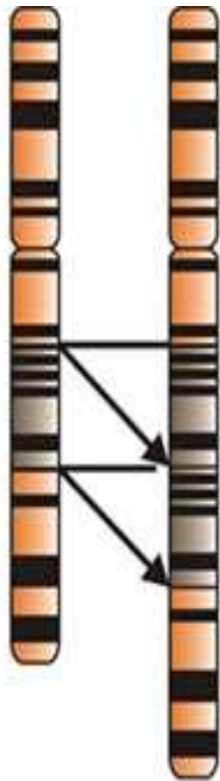
X-rays & gamma rays damage

- Double-strand breaks are also responsible for various chromosomal aberrations such as deletions, duplications, inversions and translocations.

X-rays & gamma rays damage



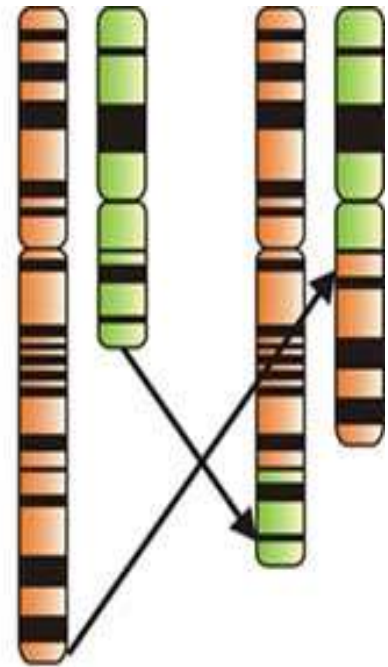
Deletion



Duplication



Inversion



Translocation

X-rays & gamma rays damage

- About 65% of the DNA damage caused by these radiations is due to indirect effects, primarily due to transfer of photons to water.

X-rays & gamma rays damage

- The photon transfer activates the water and causing it to undergo two types of primary reactions.
- In the first of these reactions, the water molecule is ionized.



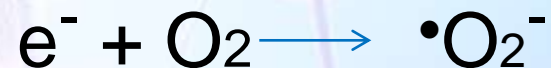
X-rays & gamma rays damage

- The $\text{H}_2\text{O}^{\bullet+}$ that is formed readily dissociates and release a proton and a hydroxyl radical ($\bullet\text{OH}$).



X-rays & gamma rays damage

- The electron generated by the first reaction can combine with molecular oxygen to form a super-oxide radical ($\bullet\text{O}_2^-$).



X-rays & gamma rays damage

- In the second type of primary reaction, excited water molecule (H_2O^*) splits into a hydrogen atom and a hydroxyl radical.



X-rays & gamma rays damage

- So the three highly reactive chemical species i.e., $\bullet\text{H}$, $\bullet\text{OH}$ and $\bullet\text{O}_2^-$ are produced by the two primary pathways.

X-rays & gamma rays damage

- Each of these attacks and damages whatever biomolecule they encounter.
- A wide variety of changes take place when that molecule happens to be DNA.

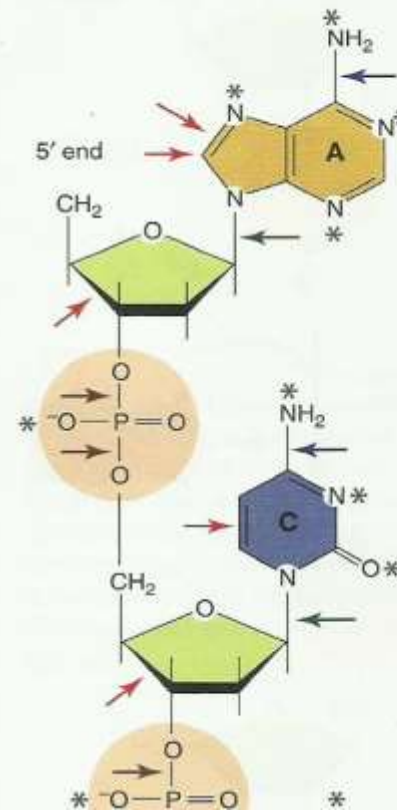
End

DNA INSTABILITY IN WATER

- DNA has three kinds of bonds with the potential for hydrolytic cleavage, namely:
 - Phosphodiester bond
 - N-glycosyl bond
 - Bonds linking exocyclic amine groups to bases

DNA INSTABILITY IN WATER

Hydrolytic Bond Cleavage



DNA INSTABILITY IN WATER

Phosphodiester Bond Cleavage

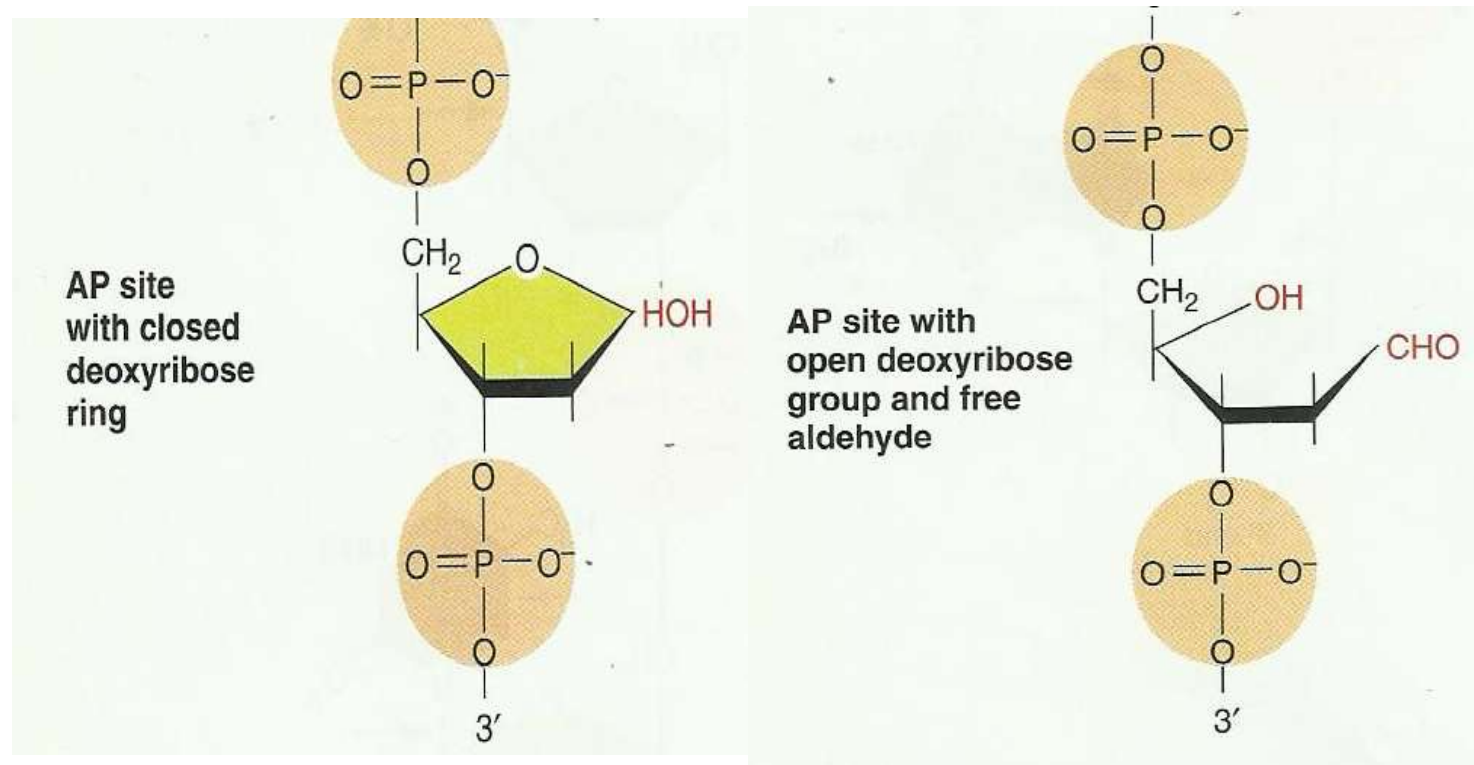
- Spontaneous phosphodiester bond cleavage, which introduces a nick in the DNA strand, is a very rare occurrence and doesn't make a significant contribution to DNA damage.

DNA INSTABILITY IN WATER

N-glycosyl Bond Cleavage

- N-glycosyl bond cleavage leads to the formation of an abasic site, which is also known as an **AP site** (AP for apurinic & apyrimidinic).

DNA INSTABILITY IN WATER



DNA INSTABILITY IN WATER

N-glycosyl Bond Cleavage

- According to current estimates, about 10,000 purine and 500 pyrimidine bases are lost from DNA in a mammalian cell nucleus each day.

DNA INSTABILITY IN WATER

N-glycosyl Bond Cleavage

- Experiments are also showing that purine N-glycosyl bonds are more easily hydrolyzed than pyrimidine N-glycosyl bonds.

DNA INSTABILITY IN WATER

N-glycosyl Bond Cleavage

- AP site formation sensitizes the neighbouring 3'-phosphodiester bond to cleavage which can be attributed to the formation of a free aldehyde group.

DNA INSTABILITY IN WATER

N-glycosyl Bond Cleavage

- A DNA strand with one or more AP sites makes a poor template because it lacks the information required to direct accurate replication and transcription.

End

Water-mediated Deamination

- Water-mediated deamination converts cytosine, guanine and adenine to uracil, xanthine and hypoxanthine, respectively.

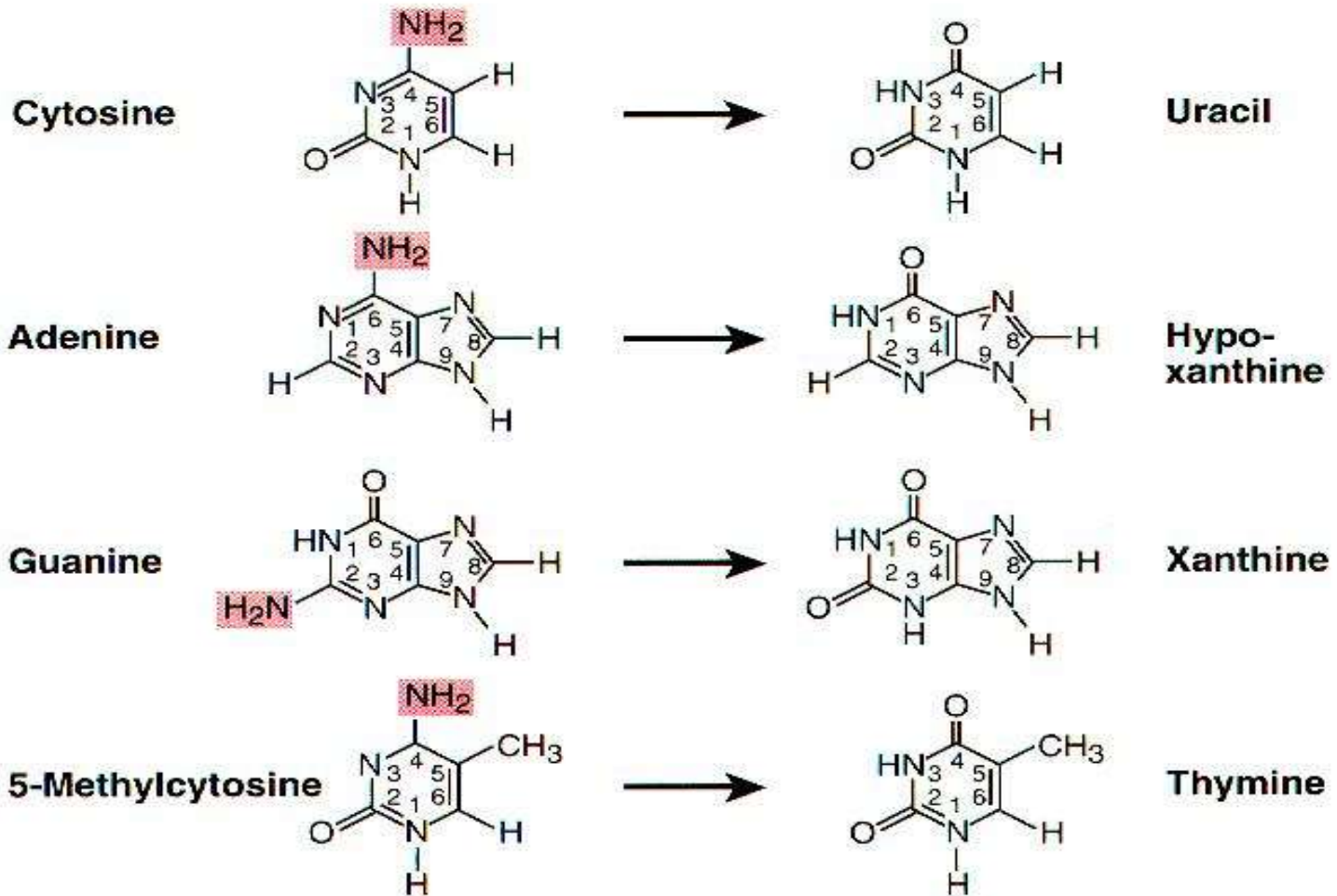
Water-mediated Deamination

- Hydrolytic deamination of cytosine is estimated to take place about 100 to 500 times a day in a mammalian cell.

Water-mediated Deamination

- Whereas, combined guanine and adenine deaminations are estimated to occur at about 1 or 2% of that of cytosine deamination.

Water-mediated Deamination



Water-mediated Deamination

- The conversion of guanine to xanthine may result in mutations or arrested DNA synthesis because xanthine does not make stable base pairs with either cytosine or thymine.

Water-mediated Deamination

- While the conversion of adenine to hypoxanthine will cause a T – A base pair to be replaced by C – G base pair.

Water-mediated Deamination

- Likewise, an uncorrected deamination that converts cytosine to uracil will cause a C – G base pair to be replaced with a T – A base pair.

Water-mediated Deamination

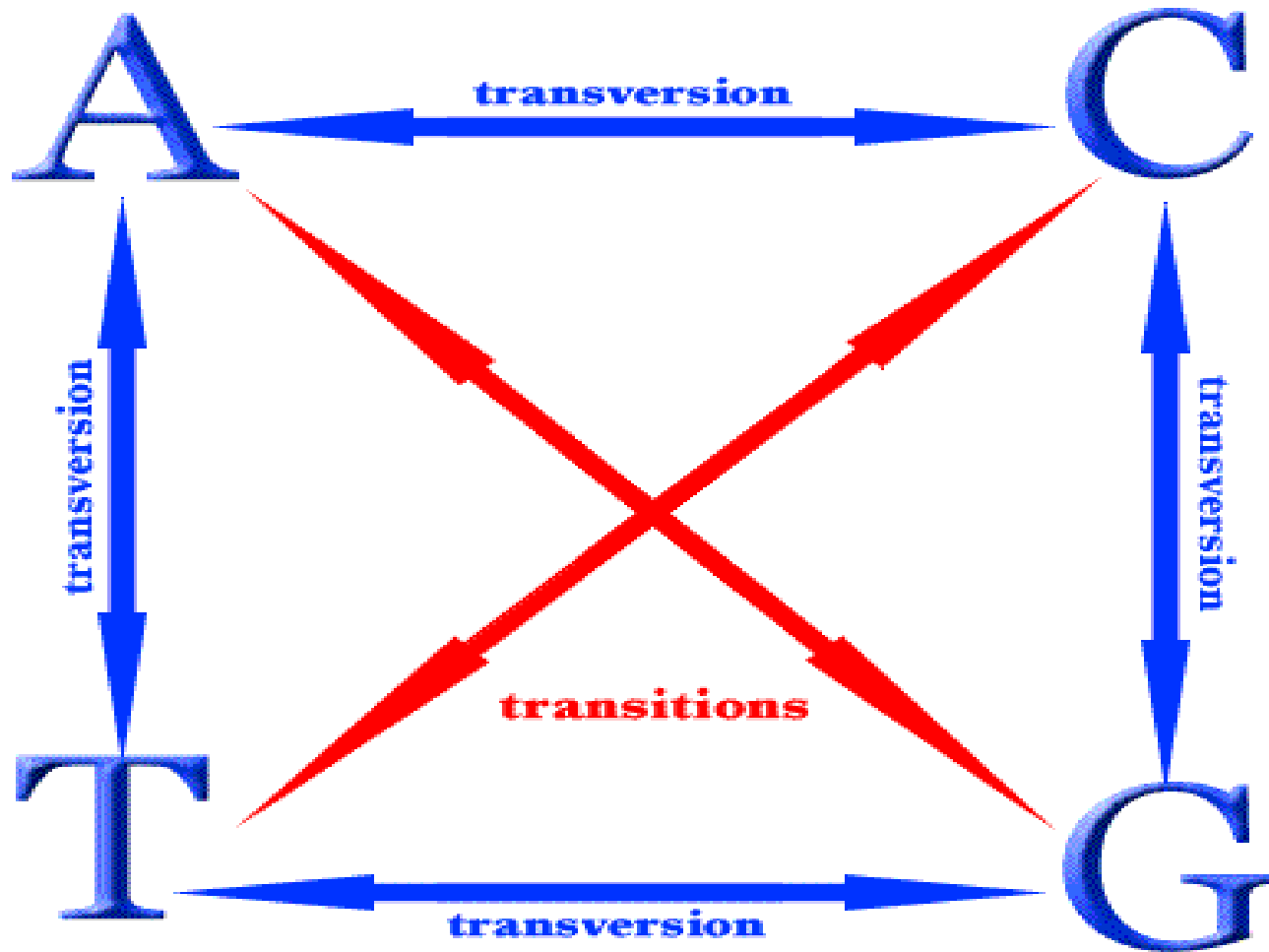
- Mutations of this type in which a pyrimidine on one strand is replaced by a different pyrimidine and a purine on the other strand is replaced by a different purine are called **transitions**.

END

Water-mediated Deamination

- Another type of replacement mutation which is termed as **transversion** mutation involves replacing a pyrimidine on one strand with purine and a purine on the other strand with a pyrimidine.

Water-mediated Deamination



Water-mediated Deamination

- A few cytosine bases in eukaryotic DNA are converted into modified base 5-methylcytosine.
- This modified base is concentrated in so called **CpG islands**.

Water-mediated Deamination

- CpG islands are small segments of DNA often present in regulatory elements called promoters that are located just before the transcription unit they regulate.

Water-mediated Deamination

- The frequency of spontaneous deamination of 5-methylcytosine bases in CpG islands is even greater than those of cytosine.

Water-mediated Deamination

- The product of this deamination is thymine and not uracil.
- So it results in the conversion of a C – G base pair to a T – A base pair.

Water-mediated Deamination

- Nitrous acid (HNO_2) is formed from nitrites which are used as preservatives in processed meat.
- It reacts with amine groups attached to the ring structure in C, A and G and greatly increase their rate of deamination.

Water-mediated Deamination

- Bisulfite (HSO_3^-) used as an additive in fruit juices and dry fruits also greatly increases the rate of cytosine deamination but doesn't affect purine or 5-methylcytosine deamination.

End

Oxidative Damage to DNA

- The reactive oxygen species damage DNA.
- So the reactive oxygen species produced during cellular respiration (ETC) may have the potential to damage DNA.

Oxidative Damage to DNA

- But it is unlikely that they do so because:-
- 1) the respiratory chain doesn't normally release these reactive oxygen species.

Oxidative Damage to DNA

- 2) cells contain superoxide dismutase to convert superoxide radicals into molecular oxygen and hydrogen peroxide. Then catalase to convert hydrogen peroxide to oxygen and water.

Oxidative Damage to DNA

- 3) superoxides and hydroxyl radicals are so reactive that, if released from respiratory chain, they would react with the nearby biomolecules before they had chance to reach nuclear DNA.

Oxidative Damage to DNA

- So the reactive oxygen species produced in this way do not damage the nuclear DNA under normal physiological conditions.

Oxidative Damage to DNA

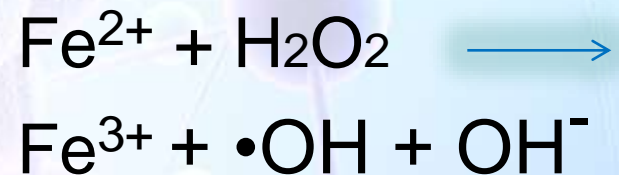
- The primary culprit appears to be the hydroxyl radical, which is produced by ionizing radiations.
- Hydroxyl radicals can also be produced chemically from hydrogen peroxide.

Oxidative Damage to DNA

- Hydrogen peroxide is not as reactive as superoxide and hydroxyl radicals, so it has a much longer half-life in the cell, provided it escapes catalase and peroxidase.

Oxidative Damage to DNA

- If it does escape, hydrogen peroxide can be converted into hydroxide radical by the following reaction:-



Oxidative Damage to DNA

- This reaction is called the Fenton reaction and in this reaction copper, manganese and cobalt can replace iron.

End

Oxidative Damage to DNA

- Hydroxyl radicals generated by any means are known to cause more than 80 different kinds of base damage.
- Two of the oxidized base products are 8-oxoguanine (oxoG) and thymine glycol.

Oxidative Damage to DNA

- 8-Oxoguanine can base pair with adenine or cytosine and if uncorrected, this 8-oxoG – A base pair will be replicated to form a T – A base pair, thus causing a transversion mutation.

Oxidative Damage to DNA

- On the other hand, thymine glycol inhibits DNA replication and is therefore cytotoxic.

Oxidative Damage to DNA

- Hydroxyl radicals produced by the Fenton reaction are tend to be more widely dispersed than those produced by ionizing radiations and therefore, much less likely to produce double-stranded breaks.

Oxidative Damage to DNA

- Cells can repair single-stranded breaks much more easily than they can repair double-stranded breaks.

Oxidative Damage to DNA

- Reactive oxygen species can also convert various biomolecules into reactive species that can then damage DNA.

Oxidative Damage to DNA

- For example, polyunsaturated fatty acid oxidation produces two aldehyde products, **malondialdehyde** and **4- hydroxynonenal**, which contribute to base damage.

END

Oxidative Damage to DNA

- The reactive oxygen species damage DNA.
- So the reactive oxygen species produced during cellular respiration (ETC) may have the potential to damage DNA.

Oxidative Damage to DNA

- But it is unlikely that they do so because:-
- 1) the respiratory chain doesn't normally release these reactive oxygen species.

Oxidative Damage to DNA

- 2) cells contain superoxide dismutase to convert superoxide radicals into molecular oxygen and hydrogen peroxide. Then catalase to convert hydrogen peroxide to oxygen and water.

Oxidative Damage to DNA

- 3) superoxides and hydroxyl radicals are so reactive that, if released from respiratory chain, they would react with the nearby biomolecules before they had chance to reach nuclear DNA.

Oxidative Damage to DNA

- So the reactive oxygen species produced in this way do not damage the nuclear DNA under normal physiological conditions.

Oxidative Damage to DNA

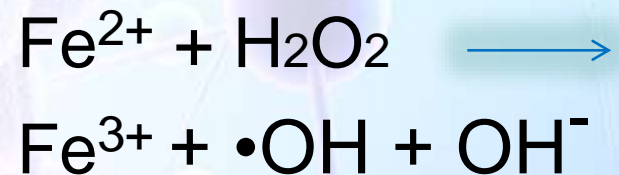
- The primary culprit appears to be the hydroxyl radical, which is produced by ionizing radiations.
- Hydroxyl radicals can also be produced chemically from hydrogen peroxide.

Oxidative Damage to DNA

- Hydrogen peroxide is not as reactive as superoxide and hydroxyl radicals, so it has a much longer half-life in the cell, provided it escapes catalase and peroxidase.

Oxidative Damage to DNA

- If it does escape, hydrogen peroxide can be converted into hydroxide radical by the following reaction:-



Oxidative Damage to DNA

- This reaction is called the Fenton reaction and in this reaction copper, manganese and cobalt can replace iron.

Oxidative Damage to DNA

- Hydroxyl radicals generated by any means are known to cause more than 80 different kinds of base damage.
- Two of the oxidized base products are 8-oxoguanine (oxoG) and thymine glycol.S

Oxidative Damage to DNA

- 8-Oxoguanine can base pair with adenine or cytosine and if uncorrected, this 8-oxoG – A base pair will be replicated to form a T – A base pair, thus causing a transversion mutation.

Oxidative Damage to DNA

- On the other hand, thymine glycol inhibits DNA replication and is therefore cytotoxic.

Oxidative Damage to DNA

- Hydroxyl radicals produced by the Fenton reaction are tend to be more widely dispersed than those produced by ionizing radiations and therefore, much less likely to produce double-stranded breaks.

Oxidative Damage to DNA

- Cells can repair single-stranded breaks much more easily than they can repair double-stranded breaks.

Oxidative Damage to DNA

- Reactive oxygen species can also convert various biomolecules into reactive species that can then damage DNA.

Oxidative Damage to DNA

- For example, polyunsaturated fatty acid oxidation produces two aldehyde products, **malondialdehyde** and **4- hydroxynonenal**, which contribute to base damage.

Oxidative Damage to DNA

- The reactive oxygen species damage DNA.
- So the reactive oxygen species produced during cellular respiration (ETC) may have the potential to damage DNA.

Oxidative Damage to DNA

- But it is unlikely that they do so because:-
- 1) the respiratory chain doesn't normally release these reactive oxygen species.

Oxidative Damage to DNA

- 2) cells contain superoxide dismutase to convert superoxide radicals into molecular oxygen and hydrogen peroxide. Then catalase to convert hydrogen peroxide to oxygen and water.

Oxidative Damage to DNA

- 3) superoxides and hydroxyl radicals are so reactive that, if released from respiratory chain, they would react with the nearby biomolecules before they had chance to reach nuclear DNA.

Oxidative Damage to DNA

- So the reactive oxygen species produced in this way do not damage the nuclear DNA under normal physiological conditions.

Oxidative Damage to DNA

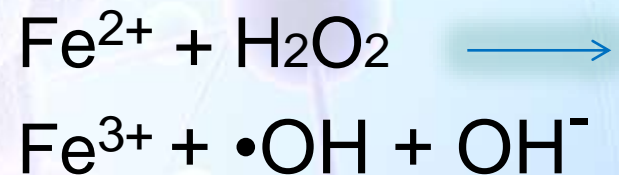
- The primary culprit appears to be the hydroxyl radical, which is produced by ionizing radiations.
- Hydroxyl radicals can also be produced chemically from hydrogen peroxide.

Oxidative Damage to DNA

- Hydrogen peroxide is not as reactive as superoxide and hydroxyl radicals, so it has a much longer half-life in the cell, provided it escapes catalase and peroxidase.

Oxidative Damage to DNA

- If it does escape, hydrogen peroxide can be converted into hydroxide radical by the following reaction:-



Oxidative Damage to DNA

- This reaction is called the Fenton reaction and in this reaction copper, manganese and cobalt can replace iron.

Oxidative Damage to DNA

- Hydroxyl radicals generated by any means are known to cause more than 80 different kinds of base damage.
- Two of the oxidized base products are 8-oxoguanine (oxoG) and thymine glycol.S

Oxidative Damage to DNA

- 8-Oxoguanine can base pair with adenine or cytosine and if uncorrected, this 8-oxoG – A base pair will be replicated to form a T – A base pair, thus causing a transversion mutation.

Oxidative Damage to DNA

- On the other hand, thymine glycol inhibits DNA replication and is therefore cytotoxic.

Oxidative Damage to DNA

- Hydroxyl radicals produced by the Fenton reaction are tend to be more widely dispersed than those produced by ionizing radiations and therefore, much less likely to produce double-stranded breaks.

Oxidative Damage to DNA

- Cells can repair single-stranded breaks much more easily than they can repair double-stranded breaks.

Oxidative Damage to DNA

- Reactive oxygen species can also convert various biomolecules into reactive species that can then damage DNA.

Oxidative Damage to DNA

- For example, polyunsaturated fatty acid oxidation produces two aldehyde products, **malondialdehyde** and **4- hydroxynonenal**, which contribute to base damage.

Alkylation Damage to DNA

- DNA has electron-rich atoms that are readily attacked by electron-seeking chemicals called **electrophiles** or electrophilic agents.

Alkylation Damage to DNA

- Alkylating agents are a highly reactive group of electrophiles which transfer methyl, ethyl, or alkyl groups to the electron-rich atoms in the DNA and damage it.

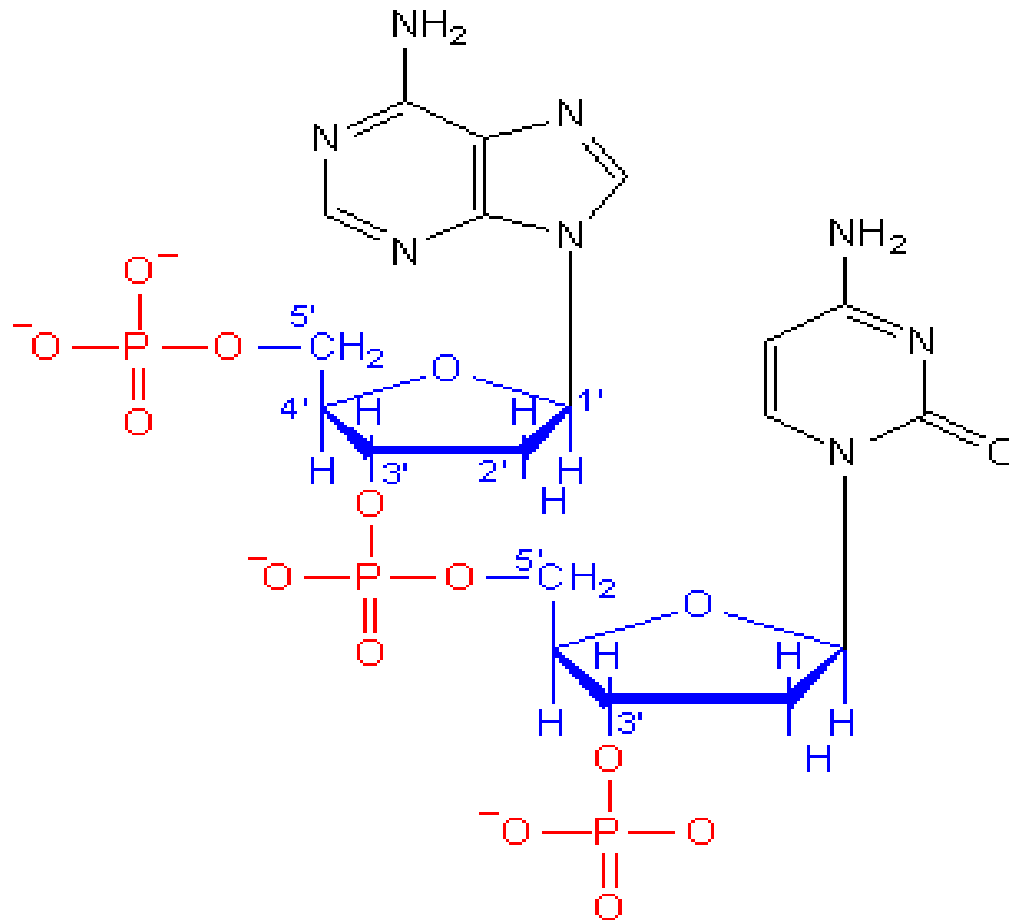
Alkylation Damage to DNA

- Alkylation in DNA takes place at:-
- A) nitrogen and oxygen atoms external to the base ring systems;
- B) non-bridging oxygen atoms in phosphate groups; and

Alkylation Damage to DNA

- C) nitrogen atoms in the base ring systems except those linked to deoxyribose.

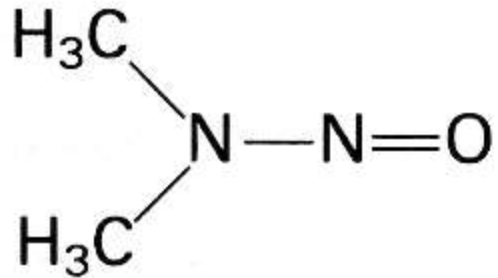
Alkylation Damage to DNA



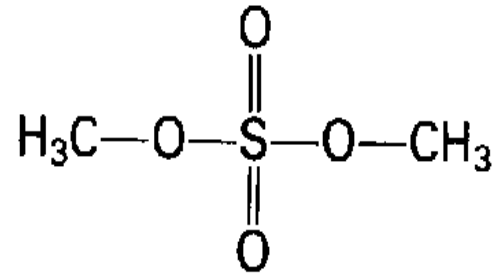
Alkylation Damage to DNA

- Many different kinds of naturally occurring and synthetic chemical agents are known to transfer alkylating agents to DNA.

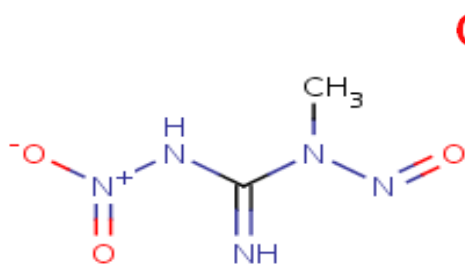
Alkylation Damage to DNA



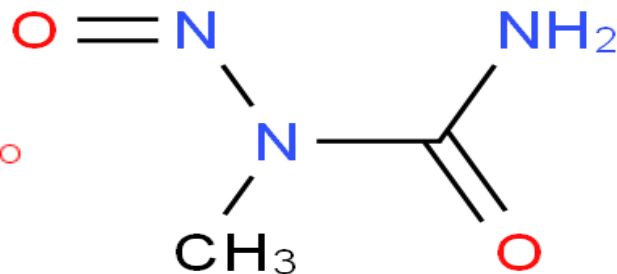
Dimethylnitrosamine



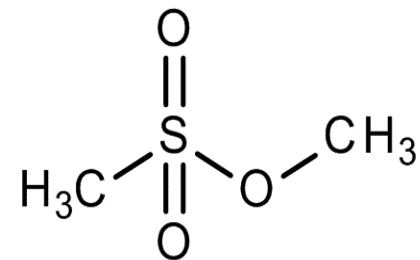
Dimethyl sulfate (DMS)



NTG



MNU



MMS

Alkylation Damage to DNA

- The product formed by attaching a chemical group to DNA is called an **adduct**.
- If the chemical group attaches to a single site on the DNA then the product is termed as **monoadduct**.

END

Alkylation Damage to DNA

- The exposure of DNA to dimethylnitrosamine leads to the production of a mono-adduct in which a single methyl group attaches to DNA.

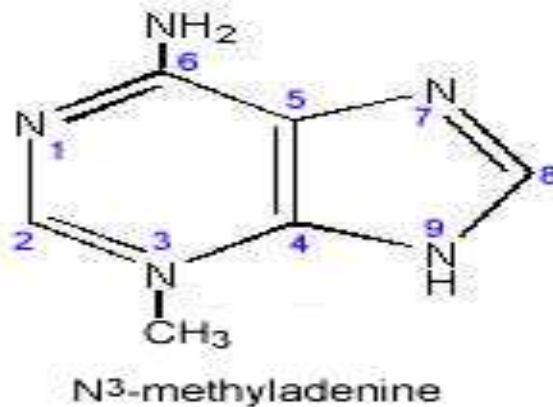
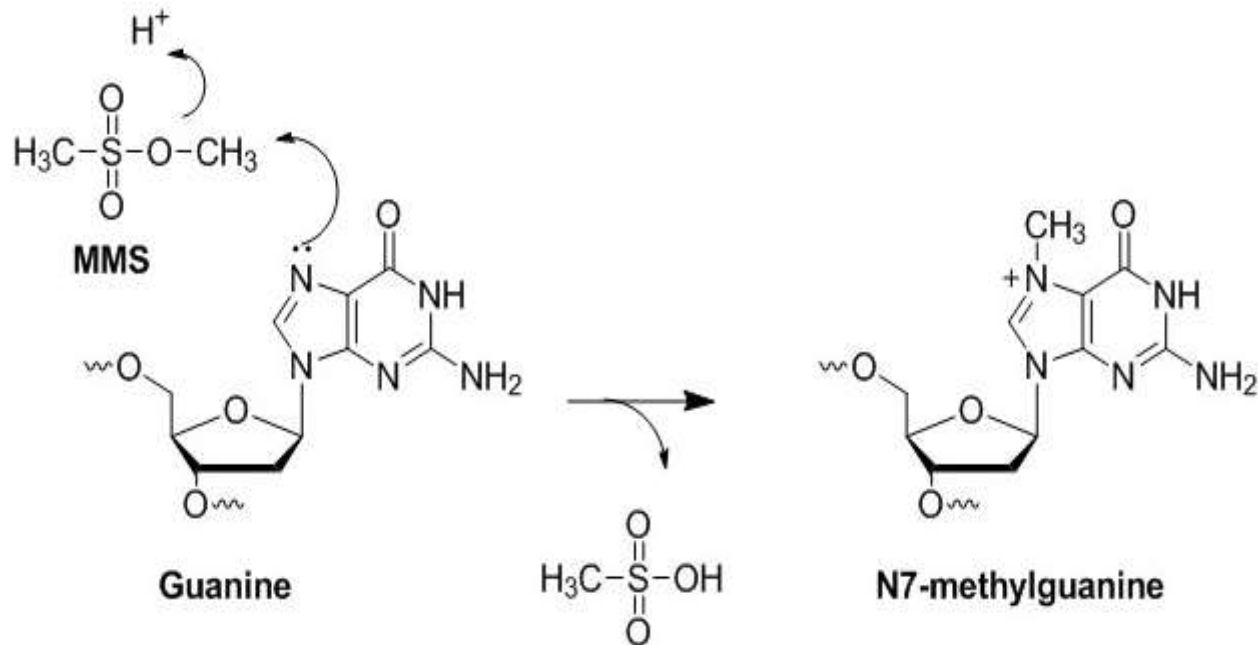
Alkylation Damage to DNA

- When DNA is exposed to methyl methane sulfonate (MMS) or N-methyl-N-nitrosourea (MNU), methylation takes place most frequently at:-

Alkylation Damage to DNA

- i) N-7 position in guanine and
- ii) next most frequently at N-3 in adenine.

Alkylation Damage to DNA



Alkylation Damage to DNA

- N-methylguanine forms a base pair with cytosine, but it is readily removed from the DNA with the resultant formation of an abasic site.

Alkylation Damage to DNA

- Methylation at N-3 in adenine is of great practical significance because N-3 methyl adenine formation blocks DNA replication but does not appear to lead to mutations.

Alkylation Damage to DNA

- Therefore, a methylating agent that can transfer methyl group exclusively to N-3 position in adenine would have the potential to kill cancer cells without causing mutations.

Alkylation Damage to DNA

- Methylation at O-6 in guanine and O-4 in thymine are much less frequent events than either of the above described methylations.

Alkylation Damage to DNA

- O⁶-methylguanine and O⁴-methylthymine formation are quite important because the methylated bases mispair during DNA replication, resulting in transition mutations.

Alkylation Damage to DNA

- The phosphate groups in DNA backbone can also be methylated.
- The resulting neutral phosphodiester is easily cleaved by water to produce single strand breaks.

END

Alkylation Damage to DNA

- DNA has electron-rich atoms that are readily attacked by electron-seeking chemicals called **electrophiles** or electrophilic agents.

Alkylation Damage to DNA

- Alkylating agents are a highly reactive group of electrophiles which transfer methyl, ethyl, or alkyl groups to the electron-rich atoms in the DNA and damage it.

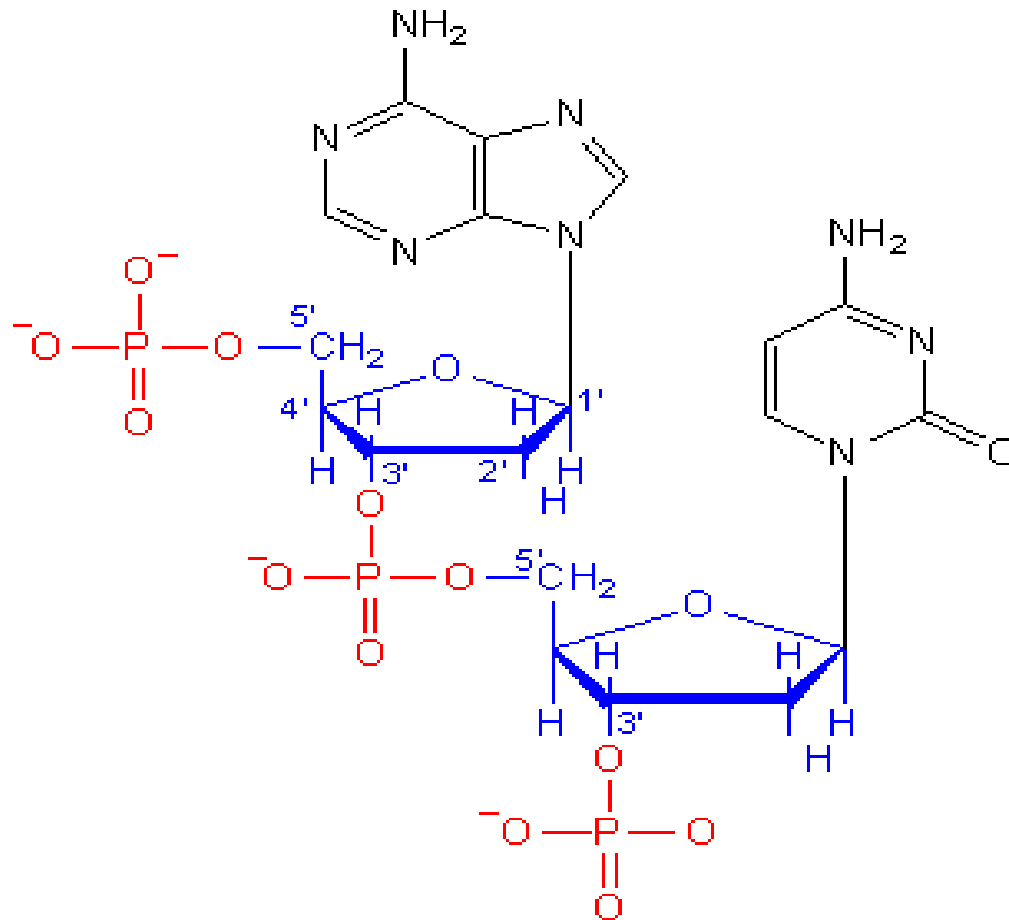
Alkylation Damage to DNA

- Alkylation in DNA takes place at:-
- A) nitrogen and oxygen atoms external to the base ring systems;
- B) non-bridging oxygen atoms in phosphate groups; and

Alkylation Damage to DNA

- C) nitrogen atoms in the base ring systems except those linked to deoxyribose.

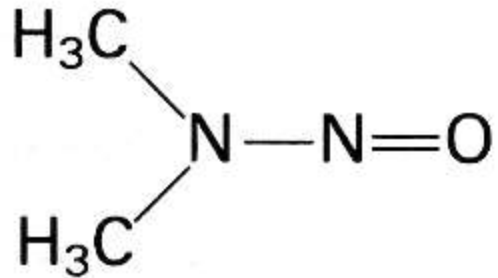
Alkylation Damage to DNA



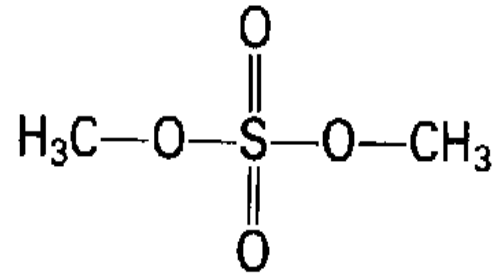
Alkylation Damage to DNA

- Many different kinds of naturally occurring and synthetic chemical agents are known to transfer alkylating agents to DNA.

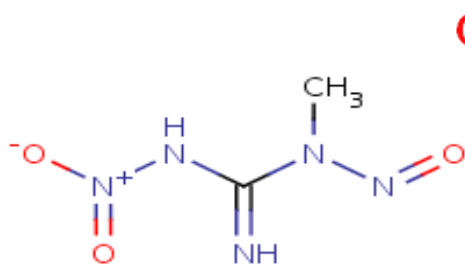
Alkylation Damage to DNA



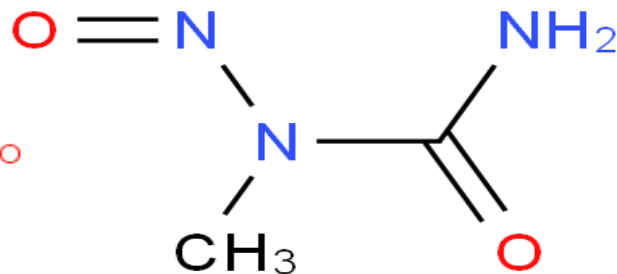
Dimethylnitrosamine



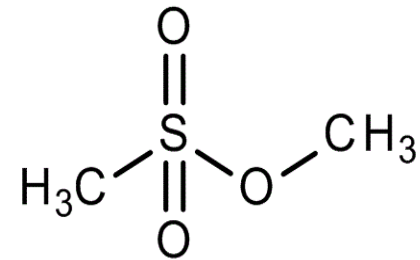
Dimethyl sulfate (DMS)



NTG



MNU



MMS

Alkylation Damage to DNA

- The product formed by attaching a chemical group to DNA is called an **adduct**.
- If the chemical group attaches to a single site on the DNA then the product is termed as **monoadduct**.

END

Alkylation Damage to DNA

- The exposure of DNA to dimethylnitrosamine leads to the production of a mono-adduct in which a single methyl group attaches to DNA.

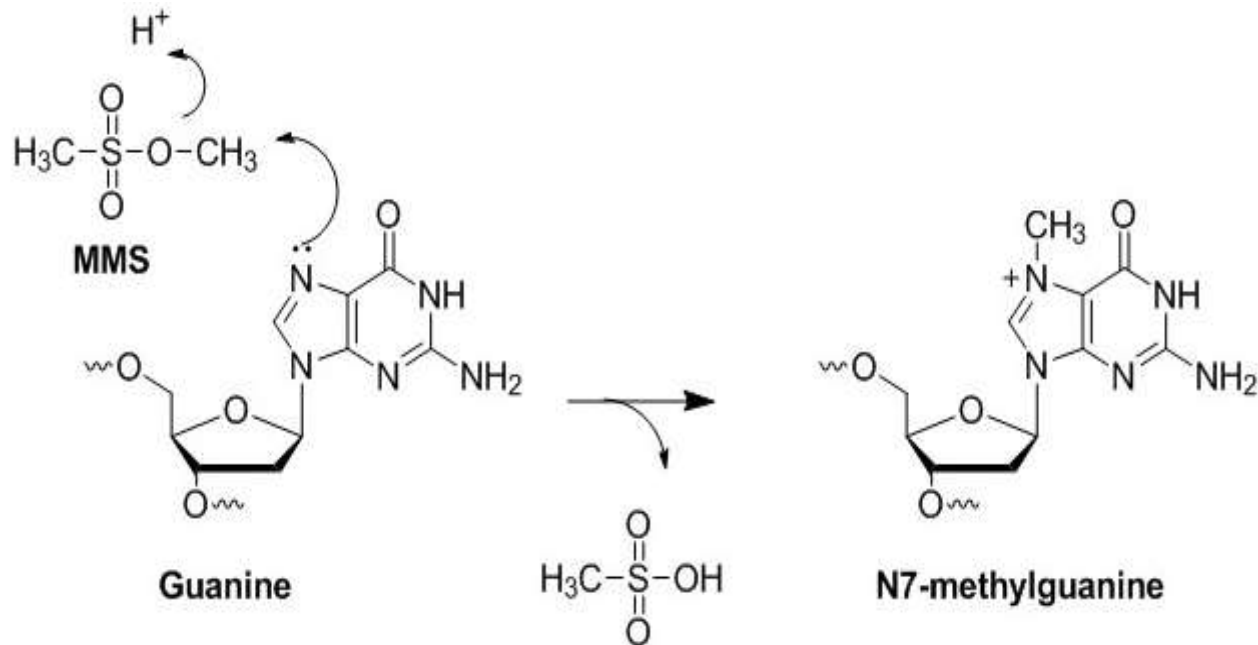
Alkylation Damage to DNA

- When DNA is exposed to methyl methane sulfonate (MMS) or N-methyl-N-nitrosourea (MNU), methylation takes place most frequently at:-

Alkylation Damage to DNA

- i) N-7 position in guanine and
- ii) next most frequently at N-3 in adenine.

Alkylation Damage to DNA



Alkylation Damage to DNA

- N-methylguanine forms a base pair with cytosine, but it is readily removed from the DNA with the resultant formation of an abasic site.

Alkylation Damage to DNA

- Methylation at N-3 in adenine is of great practical significance because N-3 methyl adenine formation blocks DNA replication but does not appear to lead to mutations.

Alkylation Damage to DNA

- Therefore, a methylating agent that can transfer methyl group exclusively to N-3 position in adenine would have the potential to kill cancer cells without causing mutations.

Alkylation Damage to DNA

- Methylation at O-6 in guanine and O-4 in thymine are much less frequent events than either of the above described methylations.

Alkylation Damage to DNA

- O⁶-methylguanine and O⁴-methylthymine formation are quite important because the methylated bases mispair during DNA replication, resulting in transition mutations.

Alkylation Damage to DNA

- The phosphate groups in DNA backbone can also be methylated.
- The resulting neutral phosphodiester is easily cleaved by water to produce single strand breaks.

END

DNA Damage by PAHs

- Many environmental agents become active alkylating agents only after they are metabolized in the cells.

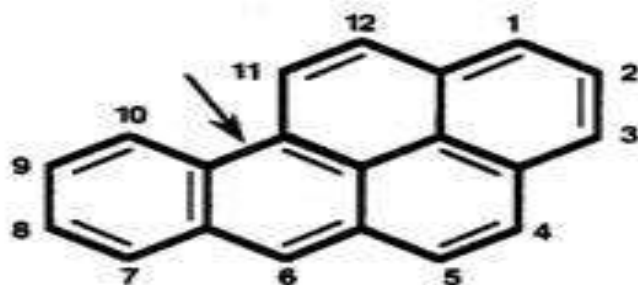
DNA Damage by PAHs

- One such agent is a mixture of polycyclic aromatic hydrocarbons (PAHs) formed by the incomplete combustion of the burning wood or coal used as fuel.

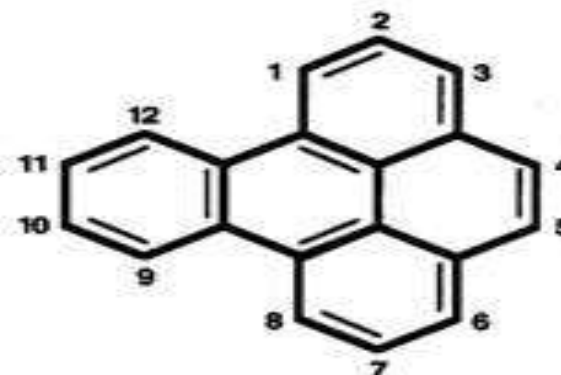
DNA Damage by PAHs

- Similar type of polycyclic aromatic hydrocarbons are also present in tobacco smoke and charbroiled meats.
- There are more than 100 different types of PAH compounds.

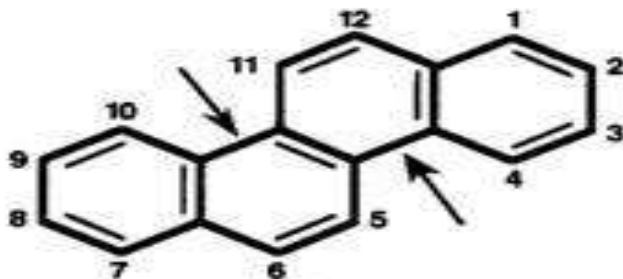
DNA Damage by PAHs



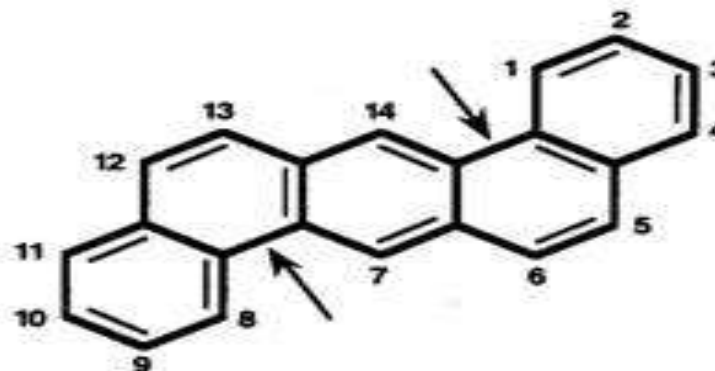
Benzo[a]pyrene



Benzo[e]pyrene



Chrysene



Dibenz[a,h]anthracene

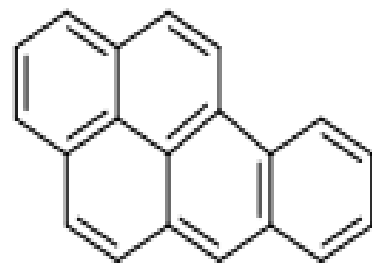
DNA Damage by PAHs

- The common structural feature in all PAHs is the presence of two or more fused aromatic rings.
- These are not able to damage DNA unless they are metabolically activated in the cell.

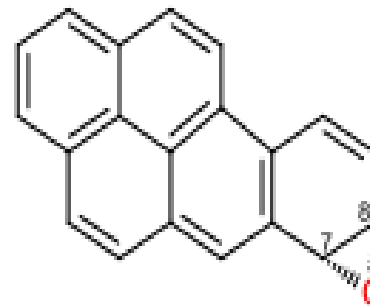
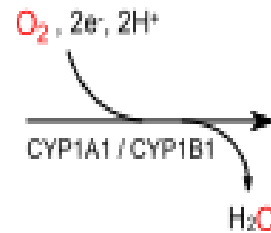
DNA Damage by PAHs

- One of the polycyclic aromatic hydrocarbon Benzo[a]pyrene is converted into an active epoxide alkylating agent through a metabolic pathway.

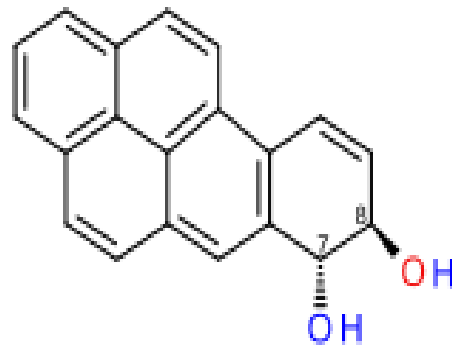
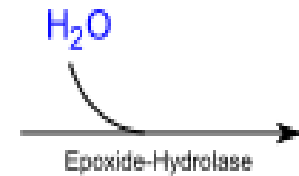
DNA Damage by PAHs



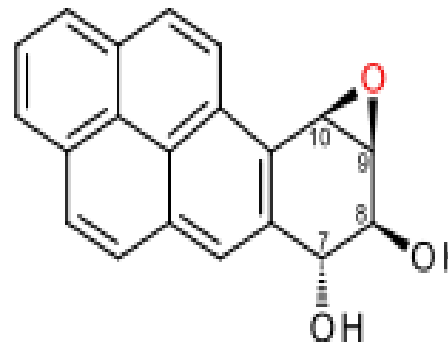
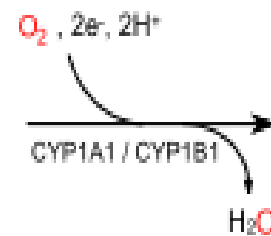
benzo[a]pyrene



(+)benzo[a]pyrene-7,8-epoxide



(-)benzo[a]pyrene-7,8-dihydrodiol



(+)benzo[a]pyrene-7,8-dihydrodiol-9,10-epoxide

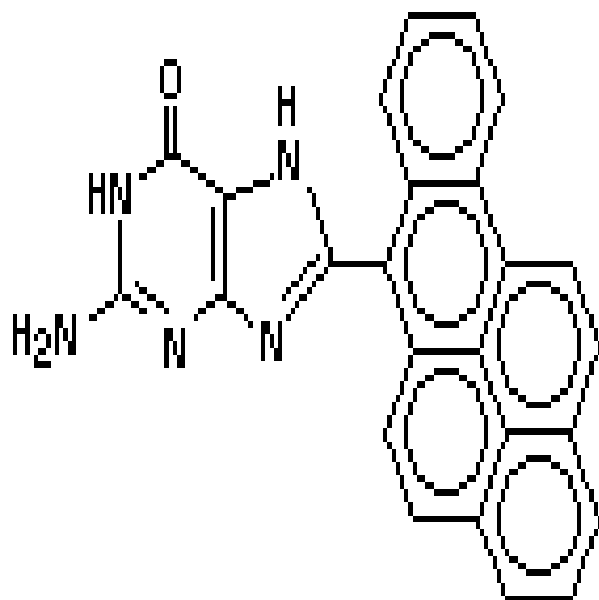
DNA Damage by PAHs

- Cytochrome P450 enzymes being not highly specific can act on PAHs such as benzo[a]pyrene and add oxygen atoms to form reactive three-membered epoxide rings.

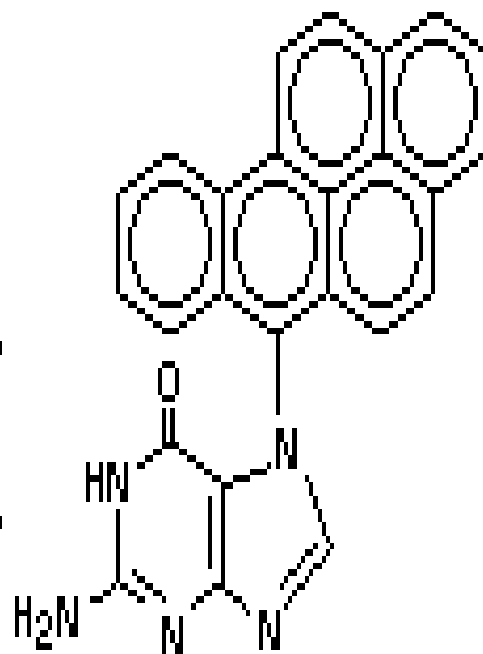
DNA Damage by PAHs

- These epoxides then alkylate DNA, causing replication errors that result in mutations, which ultimately convert a normal cell into a cancer cell.

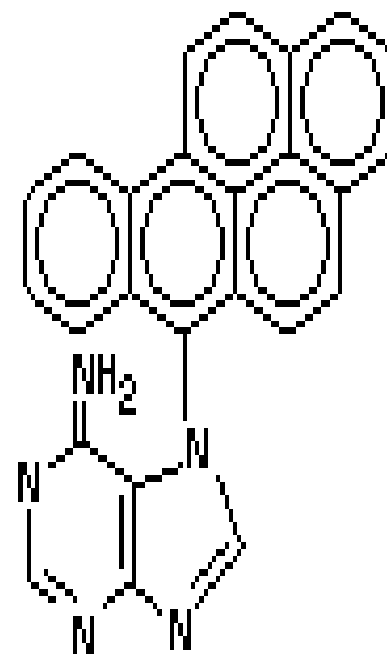
DNA Damage by PAHs



Bap-8-C8Gua



BuP-6-N7Gua



BuP-6-97Adc

DNA Damage by Aflatoxins

- Another class of chemical carcinogens that must be activated before damaging DNA are called **aflatoxins**.

DNA Damage by Aflatoxins

- They are produced from *Aspergillus flavus* and *A. parasiticus*, fungi that grow on peanut and other grains such as rice and corn.

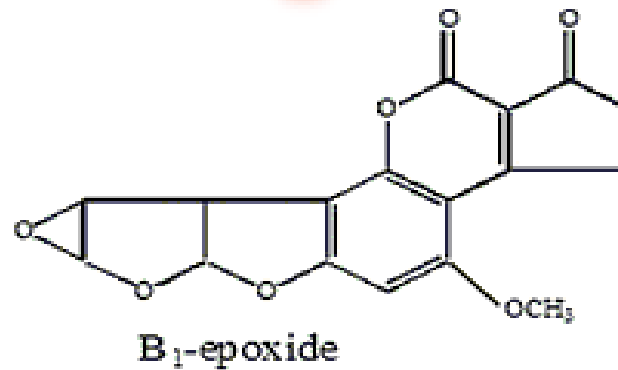
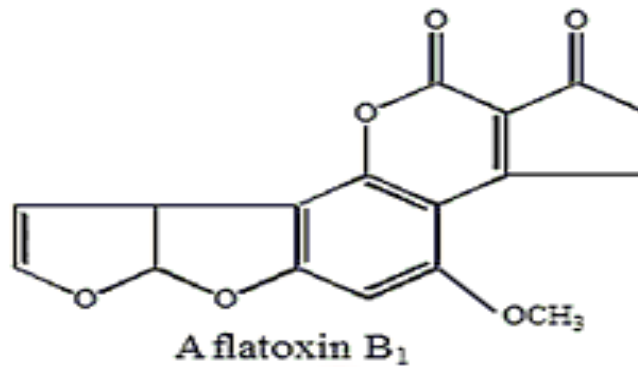
DNA Damage by Aflatoxins

- Animals feeding on contaminated peanuts or grains containing aflatoxins exhibit markedly increased rates of liver diseases including liver cancer.

DNA Damage by Aflatoxins

- Aflatoxin B1, the most potent toxin produced by *A. flavus* presents a particularly serious health threat in the United States.
- Cytochrome P450 converts it into an epoxide derivative that damages DNA.

DNA Damage by Aflatoxins



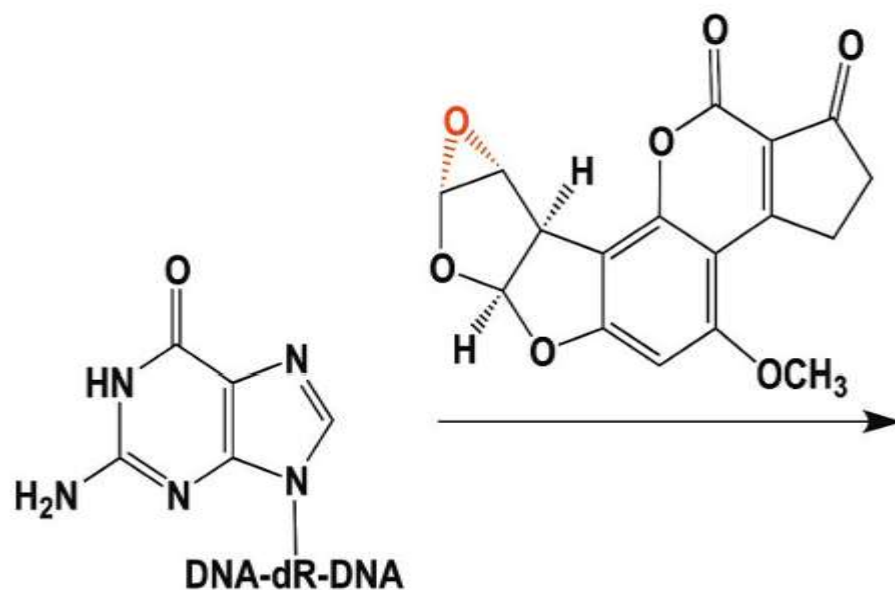
DNA Damage by Aflatoxins

- Under ideal conditions, a small tripeptide glutathione will attack the epoxide ring, making the aflatoxin derivative soluble so that it can be excreted in the urine.

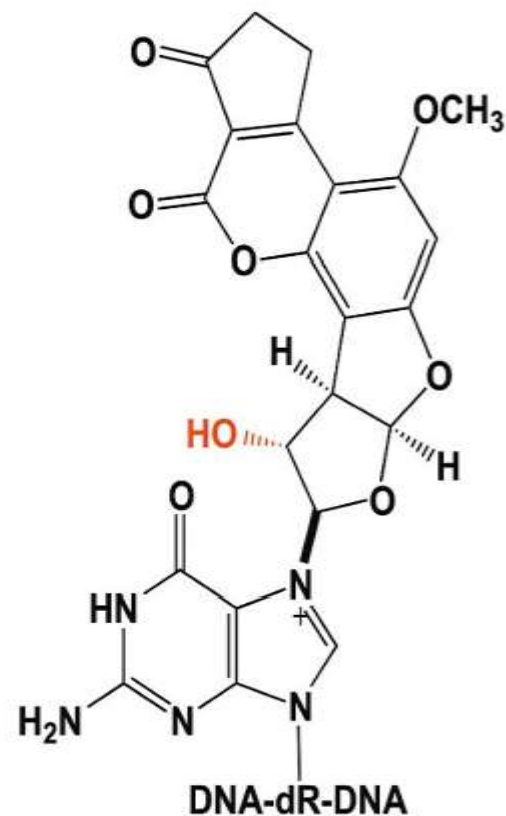
DNA Damage by Aflatoxins

- If the reactive epoxide derivatives escape the attack of glutathione, they are free to attack guanine rings in DNA.

DNA Damage by Aflatoxins



Guanine Base in DNA (dG)



AFB₁-N7-dG-Adduct

DNA Damage by Aflatoxins

- The flat aflatoxin ring system inserts between DNA bases causing helical distortion that in turn leads to replication errors.

END

Chemical Cross-Linking Agents

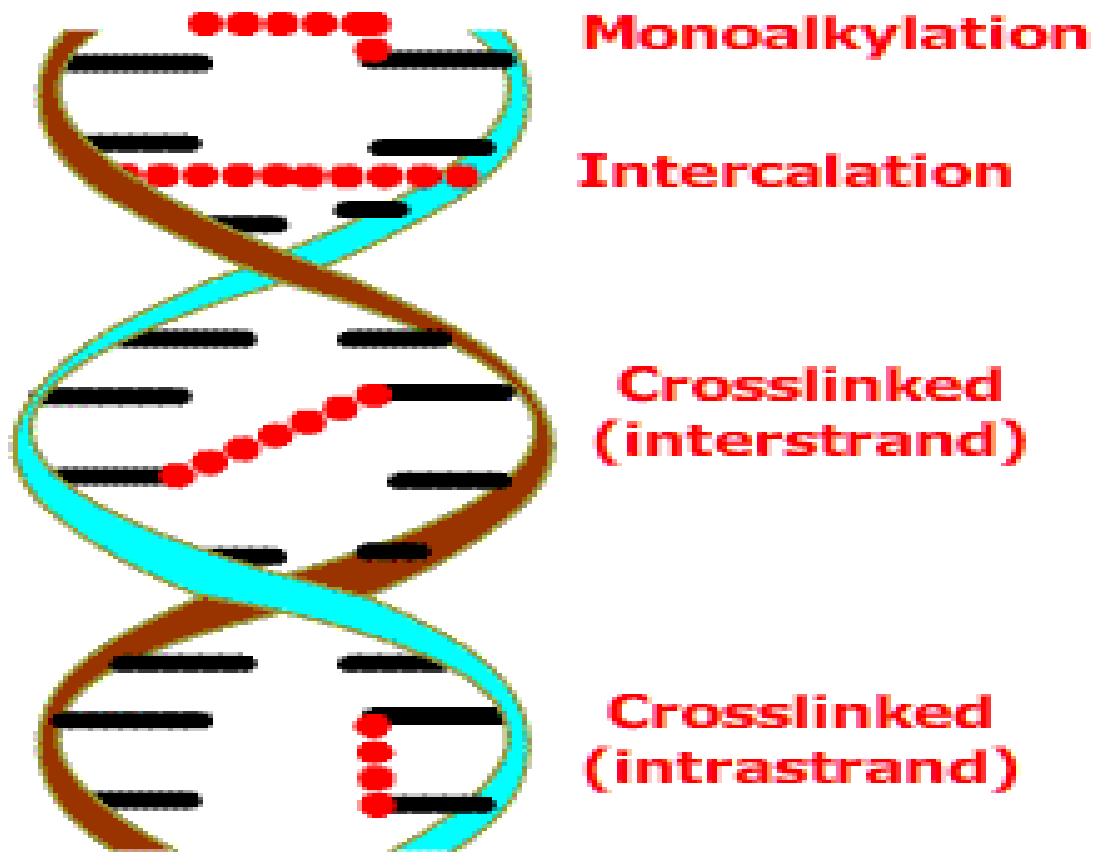
- Many alkylating agents, having two reactive sites can form intra-strand or inter-strand cross-links in addition to forming mono-adducts.

Chemical Cross-Linking Agents

- Inter-strand cross-links are of special interest because they prevent strand separation and, if not corrected, are lethal to the cells.

Chemical Cross-Linking Agents

Alkylated DNA



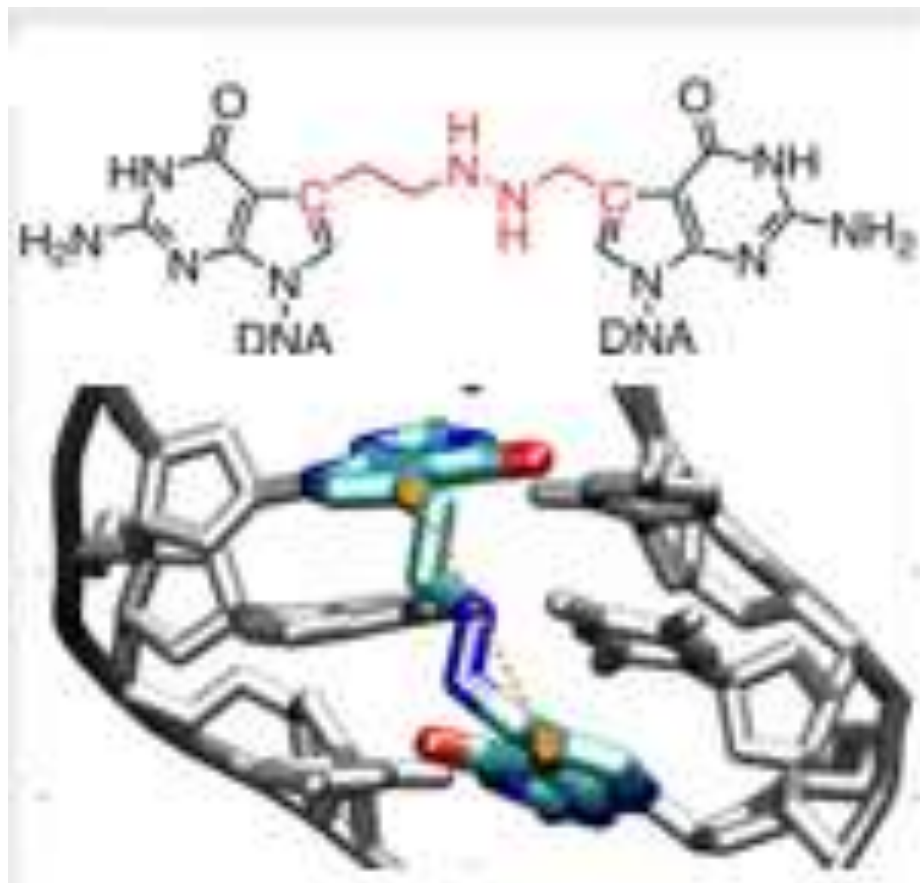
Chemical Cross-Linking Agents

- One of the simplest cross-linking agents, nitrogen mustard gas (bis[2-chloroethyl] methylamine) damages DNA by forming inter-strand cross-links.

Chemical Cross-Linking Agents

- It does so by attacking N-7 on two guanines, which are on opposite strands of DNA double helix.

Chemical Cross-Linking Agents



Chemical Cross-Linking Agents

- Although a very toxic substance, nitrogen mustard gas has found clinical application as a chemotherapeutic agent for treating certain forms of cancers.

END

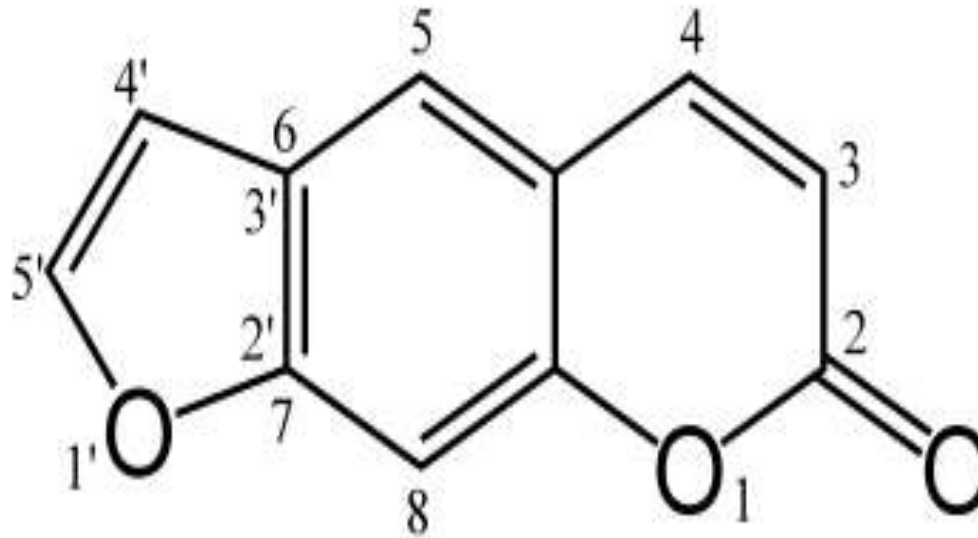
DNA damage by Psoralen

- Psoralen can form mono-adducts or cross-links.
- It is a naturally occurring substance that can alkylate DNA if photo-activated.

DNA damage by Psoralen

- The planar psoralen molecule, which consist of a furan ring fused to a heterobicyclic ring system called coumarin, intercalates into the DNA molecule.

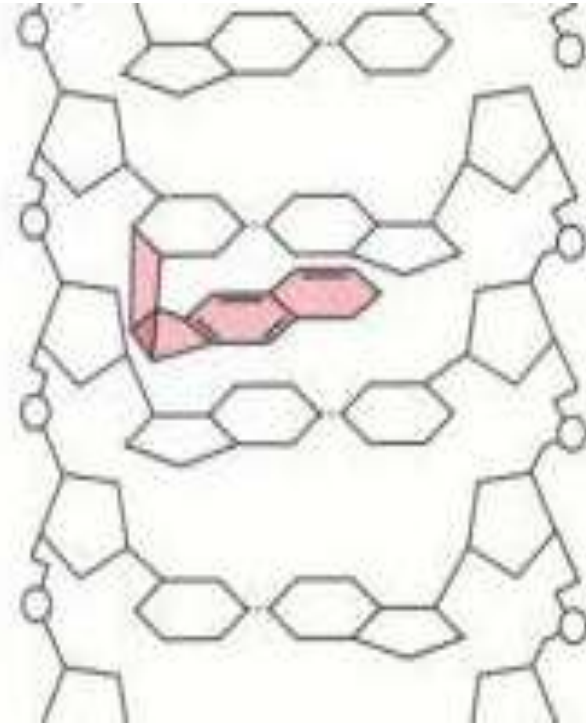
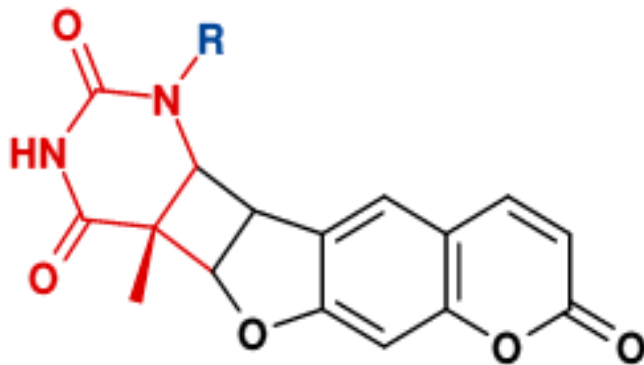
DNA damage by Psoralen



DNA damage by Psoralen

- Upon exposure to light with a wavelength of 400 to 450nm, the furan ring in psoralen becomes activated and adds across the 5,6 double bond in a pyrimidine base to form the 4',5', monoadduct.

DNA damage by Psoralen



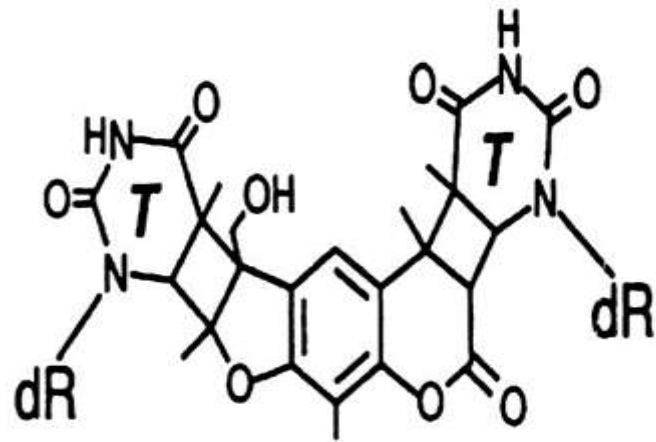
DNA damage by Psoralen

- The planar tricyclic psoralen derivative in the monoadduct is in position to combine with a second pyrimidine base on the opposite DNA strand.

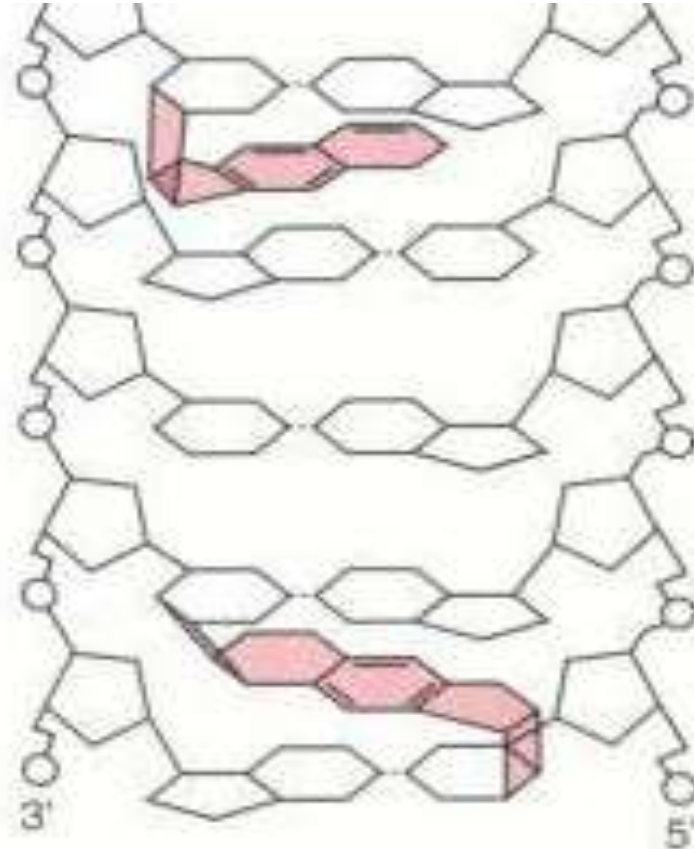
DNA damage by Psoralen

- But to do so, it must be activated by light with the wavelength of 320 to 400nm.
- The resulting photoproduct contain a cross-link between pyrimidine bases on opposite strands of the DNA duplex.

DNA damage by Psoralen



Thymine-Psoralen-Thymine crosslink



DNA damage by Psoralen

- If not properly repaired, psoralen damage causes mutations which are lethal to cells.

END

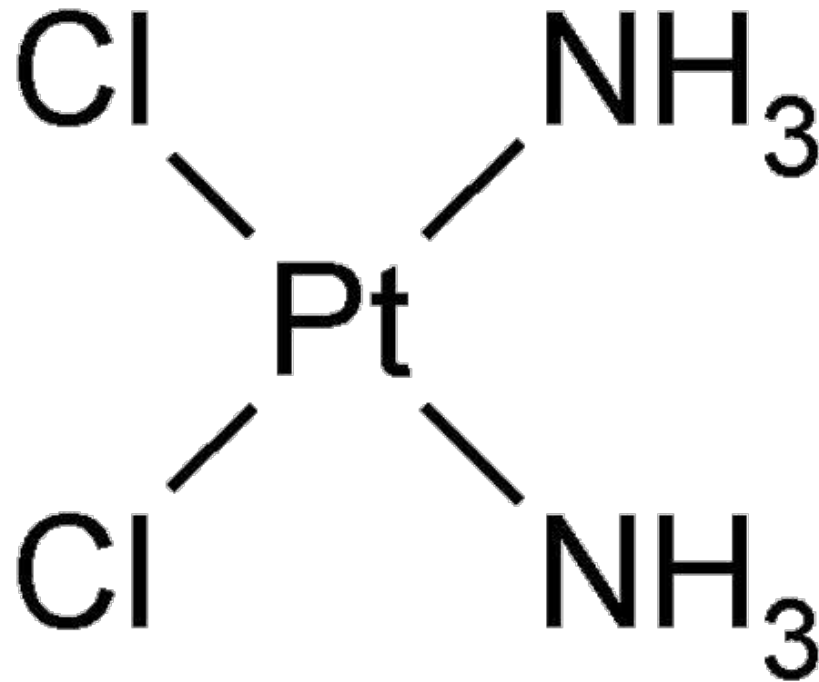
DNA damage by Cisplatin

- While examining the effects of electric currents on *E. coli*, a new compound viz; *cis*-diamminedichloroplatinum, better known as Cisplatin was found to block cell division in *E. coli*.

DNA damage by Cisplatin

- Efforts were made to see if cisplatin would also inhibit cell division in other kinds of cells.
- It was revealed that cisplatin blocks the division of tumor cells.

DNA damage by Cisplatin



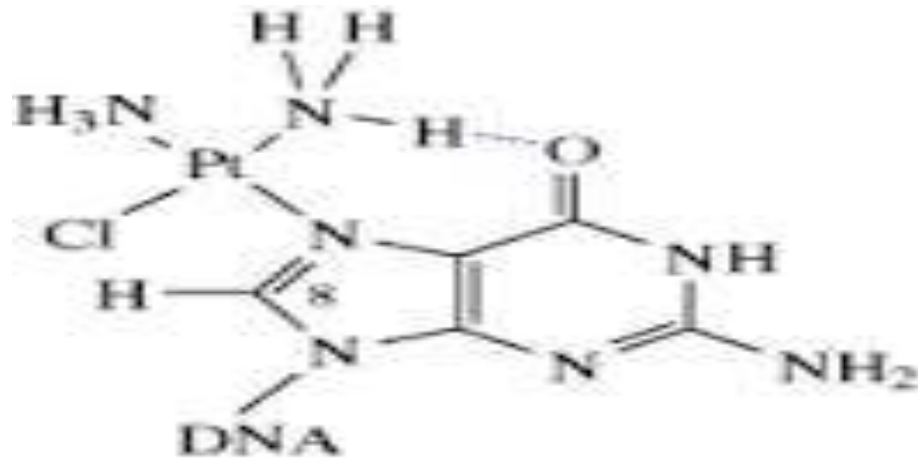
DNA damage by Cisplatin

- After cisplatin enters the cell by passive diffusion or active transport, it undergoes hydrolysis to produce a highly reactive and charged complex ($\text{Pt}(\text{NH}_3)_2\text{ClH}_2\text{O}^+$).

DNA damage by Cisplatin

- This complex coordinates to the N-7 atom of either a guanine or adenine base in DNA.

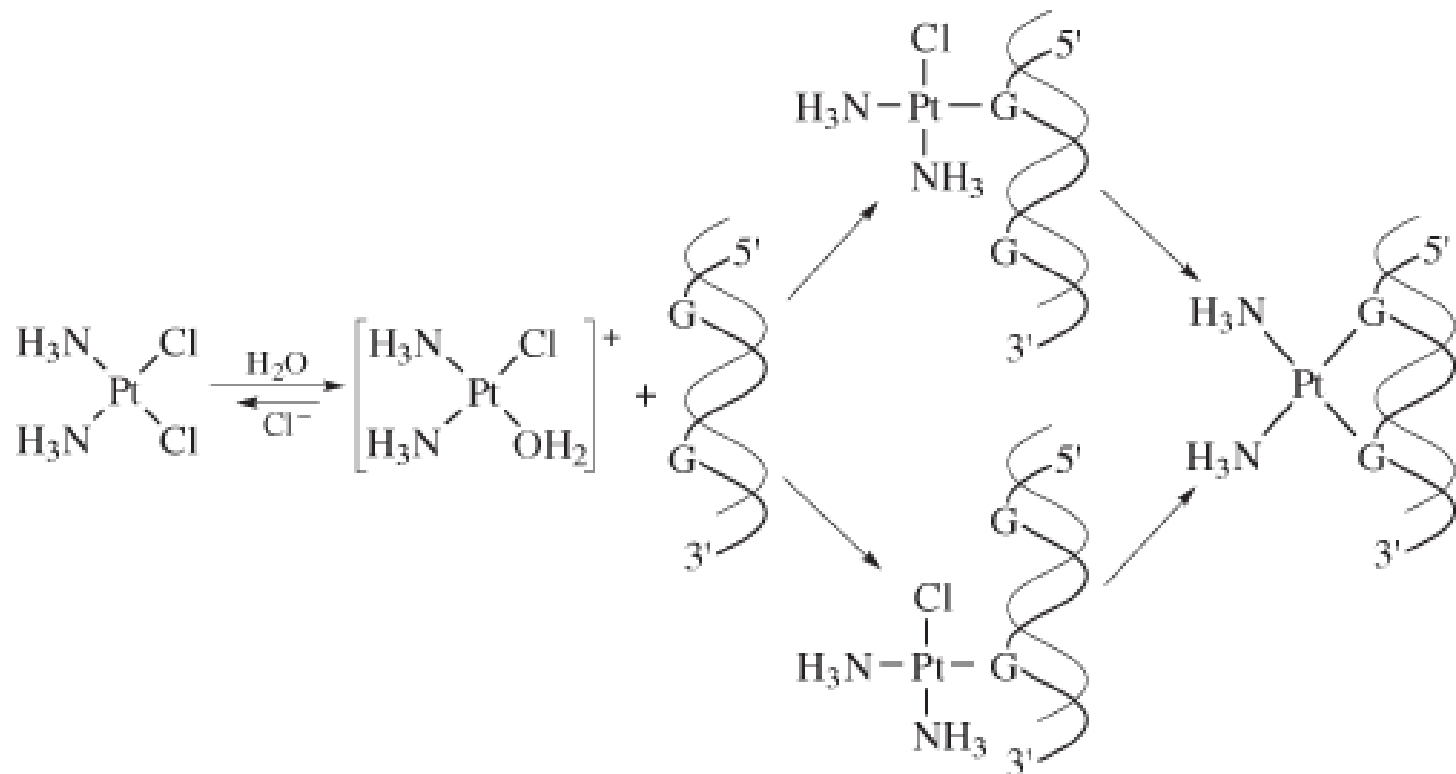
DNA damage by Cisplatin



DNA damage by Cisplatin

- Then the remaining chloride ligand is displaced by hydrolysis , allowing the platinum to coordinate to a second purine base on the same or opposite strand of the double stranded DNA.

DNA damage by Cisplatin



DNA damage by Cisplatin

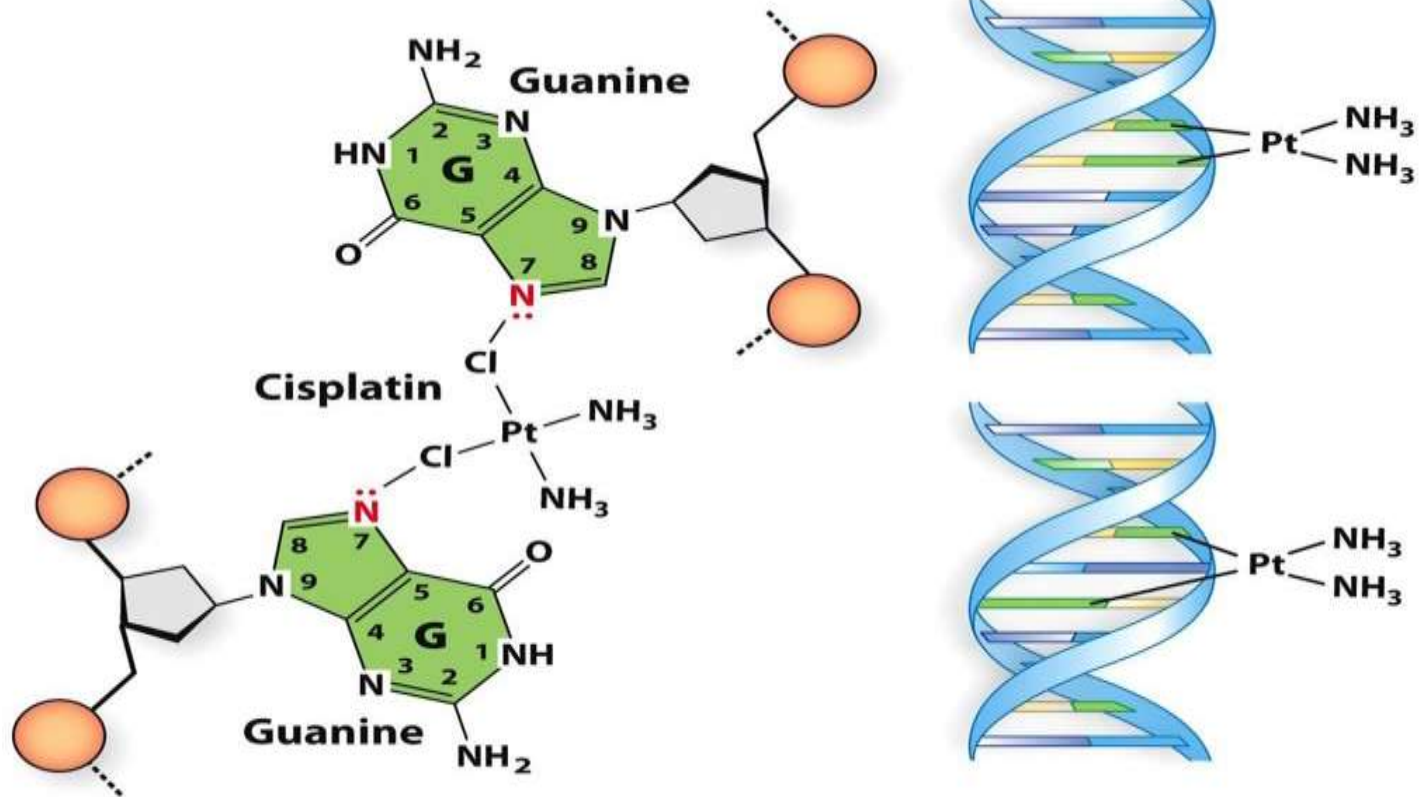


Figure 12-14a
Molecular Biology: Principles and Practice
© 2012 W. H. Freeman and Company

DNA damage by Cisplatin

- Cisplatin's cytotoxic effects appear to be due to inter-strand cross-linking which blocks the replication and transcription machinery.

DNA damage by Cisplatin

- Cisplatin is a very effective chemotherapeutic agent for treating cancers of the bladder, ovaries and testicles, but has a number of side effects.

END

Base Analogs and Intercalating Agents

- Mutations are also caused by compounds that substitute for normal bases called **Base analogs** or the compounds that slip between the bases called **Intercalating agents** and cause errors in replication.

Base Analogs and Intercalating Agents

- Base analogs are structurally similar to proper bases but differ in ways that make them treacherous to the cell.

Base Analogs and Intercalating Agents

- Thus, base analogs are similar enough to the proper bases to get taken up by cells, converted into nucleoside triphosphates, and incorporated into DNA during replication.

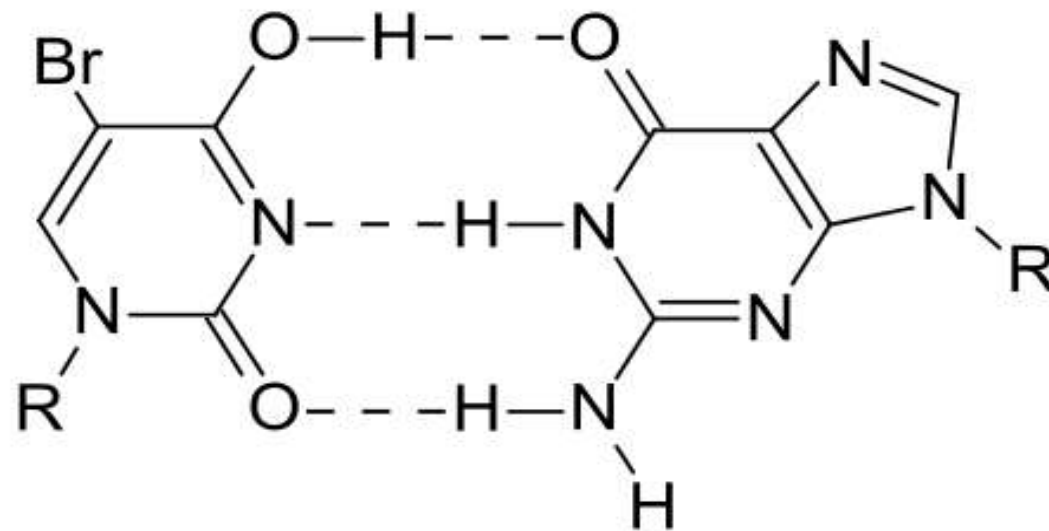
Base Analogs and Intercalating Agents

- But, because of the structural differences from the proper bases, the analogs base-pair inaccurately, leading to frequent mistakes during the replication process.

Base Analogs and Intercalating Agents

- One of the most mutagenic base analogs is 5-bromouracil, an analog of thymine. The presence of the bromo substituent allows the base to mispair with guanine.

Base Analogs and Intercalating Agents



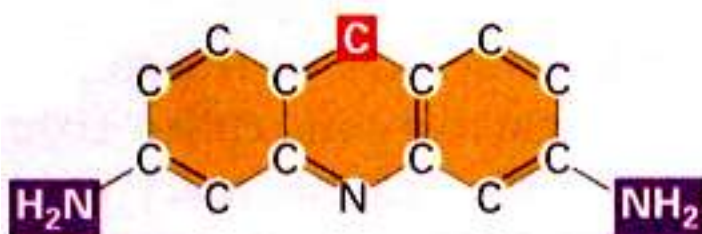
5-BrU

Guanine

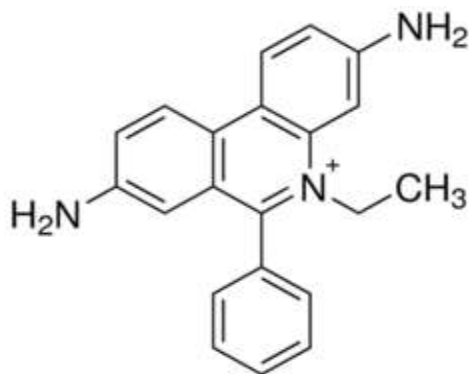
Base Analogs and Intercalating Agents

- **Intercalating agents** are flat molecules containing several polycyclic rings that bind to the purine or pyrimidine bases of DNA, just as the bases bind or stack with each other in the double helix.

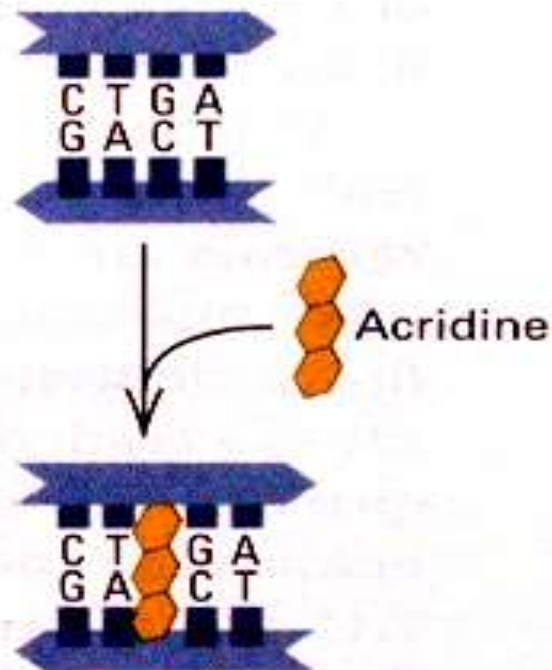
Base Analogs and Intercalating Agents



Proflavin



Ethidium



Base Analogs and Intercalating Agents

- Intercalating agents, such as proflavin, acridine, and ethidium, cause the deletion or addition of a base pair or even a few base pairs.

Base Analogs and Intercalating Agents

- These intercalating agents also cause short insertions or deletions.
- But how do these agents cause short insertions and deletions?

Base Analogs and Intercalating Agents

- One possibility in the case of insertions is that, by slipping between the bases in the template strand, these mutagens cause the DNA polymerase to insert an extra nucleotide opposite the intercalated molecule.

Base Analogs and Intercalating Agents

- Conversely, in the case of deletions, the distortion to the template caused by the presence of an intercalated molecule might cause the polymerase to skip a single or few nucleotides.

END

Direct Reversal of DNA Damage

- The first clue to the existence of an enzyme that catalyzes the direct reversal of DNA damage was reported by Albert Kelner in 1949.

Direct Reversal of DNA Damage

- In his experiments, Kelner first irradiated bacteria with UV light at doses that killed most of the bacteria.
- Then he tested the survivors to isolate the desired mutants.

Direct Reversal of DNA Damage

- Even though Kelner was very careful in his experimentation, he noticed a great deal of variation in the number of survivors from one experiment to the other.

Direct Reversal of DNA Damage

- Finally, he found that cells placed in dark after UV treatment had a very low survival rate whereas those placed in light had a high survival rate.

Direct Reversal of DNA Damage

- Exposure to light thus reversed the UV light's bactericidal effects.
- Similar phenomenon was also observed by Renato Dulbecco while studying UV-irradiated phage T2.

Direct Reversal of DNA Damage

- Dulbecco prepared multiple plates each containing the same number of UV-irradiated phages and sensitive bacteria, and placed the plates in a stack under a florescent bulb in the lab.

Direct Reversal of DNA Damage

- Each of the stacked plates should have about the same number of plaques.
- However, the plaque number decreased dramatically, going from top to bottom of the stack.

Direct Reversal of DNA Damage

- Dulbecco explained this by proposing that the plates on the top of the stack were exposed to more light from the bulb as compared to the plates on the bottom of the stack.

Direct Reversal of DNA Damage

- He tested this hypothesis by exposing some of the plates to more fluorescent light while keeping others in dark.

Direct Reversal of DNA Damage

- As expected, the number of plaques on plates exposed to light was much higher than those left in the dark.

Direct Reversal of DNA Damage

- The bacteria were somehow using the visible light to repair the UV damaged DNA in the phage.

Direct Reversal of DNA Damage

- The chemical basis for this light-dependant phenomenon, which Dulbecco called **photoreactivation**, remained to be elucidated.

Photoreactivation

- Claud S. Rupert and co-workers devised an in vitro photoreactivation system in 1957, taking a major step toward determining the chemical mechanism of photoreactivation.

Photoreactivation

- They used a straightforward approach in which they isolated DNA from the gram-negative bacteria *Haemophilus influenzae*.

Photoreactivation

- They irradiated this DNA with UV light to inactivate its transforming ability.
- Then they demonstrated that a cell-free *E. coli* extract restored the transforming activity in the presence of light.

Photoreactivation

- Although, this study had the potential to open the way for the purification and characterization of photoreactivation enzyme, investigators still needed to establish the chemical nature of this repair.

CPD Photolyase

- The problem of the chemical nature of photoreactivation was resolved over next few years when investigators came to know that UV irradiation induces the formation of cyclobutan pyrimidine dimer in DNA.

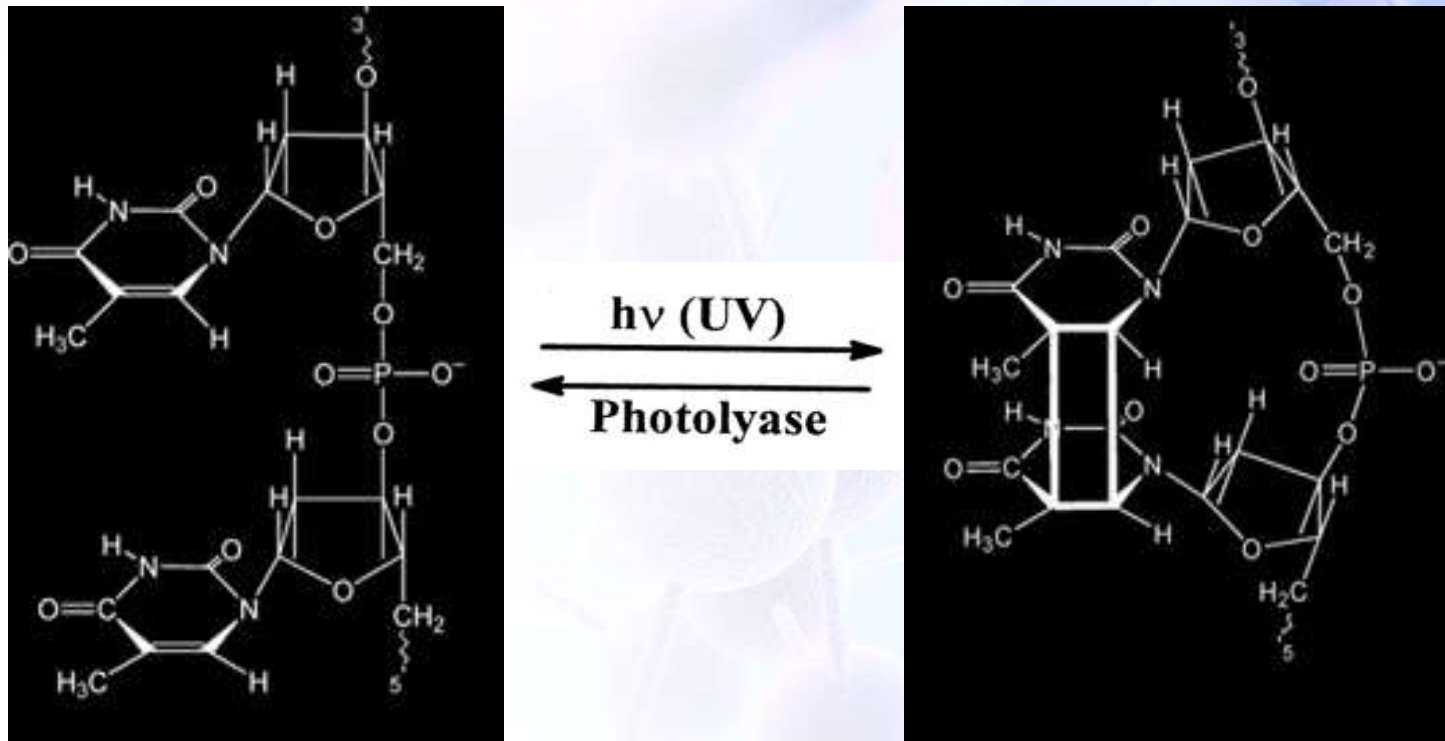
CPD Photolyase

- Further studies showed that the photoreactivation enzyme reverses the UV induced damage by using the energy provided by blue light (350-450nm) to drive cyclobutane ring disruption.

CPD Photolyase

- When it was recognized that the photoreactivation enzyme catalyzes the disruption of carbon – carbon bonds, it was given a more descriptive name of **CPD photolyase**.

CPD Photolyase



CPD Photolyase

- The bacteria that lack CPD photolyase can't repair cyclobutane pyrimidine lesions through photoreaction thus can't reverse the UV induced damage in DNA.

CPD Photolyase

- CPD photolyases are present in a wide variety of organisms including bacteria, archaea, plants, and animals but not in humans and other placental mammals.

CPD Photolyase

- These are monomeric proteins ranging in size from about 450-550 amino acid residues.
- All CPD photolyases have two domains, designated as N – and C – terminal domains.

CPD Photolyase

- A light absorbing pigment, or **chromophore pigment** binds to each domain through non-covalent bonding.

CPD Photolyase

- This chromophore factor acts as a photoantenna to capture light with wavelengths that would not otherwise be available.

Mechanism of CPD Photolyase

- CPD photolyase can bind to DNA in the dark by recognizing the altered DNA structure caused by a CPD formation rather than a specific nucleotide sequence.

Mechanism of CPD Photolyase

- This binding is about 10^5 tighter when a DNA segment contains a CPD than when it does not.
- Half of the binding energy appears to come from interactions between enzyme and the DNA back bone.

Mechanism of CPD Photolyase

- The other half of energy comes from interactions between the FADH^- at the active site and the CPD.
- However, the light harvesting antenna pigment does not influence binding.

Mechanism of CPD Photolyase

- Once the enzyme – DNA complex is formed, the CPD is flipped out of the DNA double helix and into the enzyme's active site.

Mechanism of CPD Photolyase

- After the CPD flips into the enzyme's active site, the energy of an absorbed photon is transferred from the light harvesting antenna pigment to the FADH⁻.

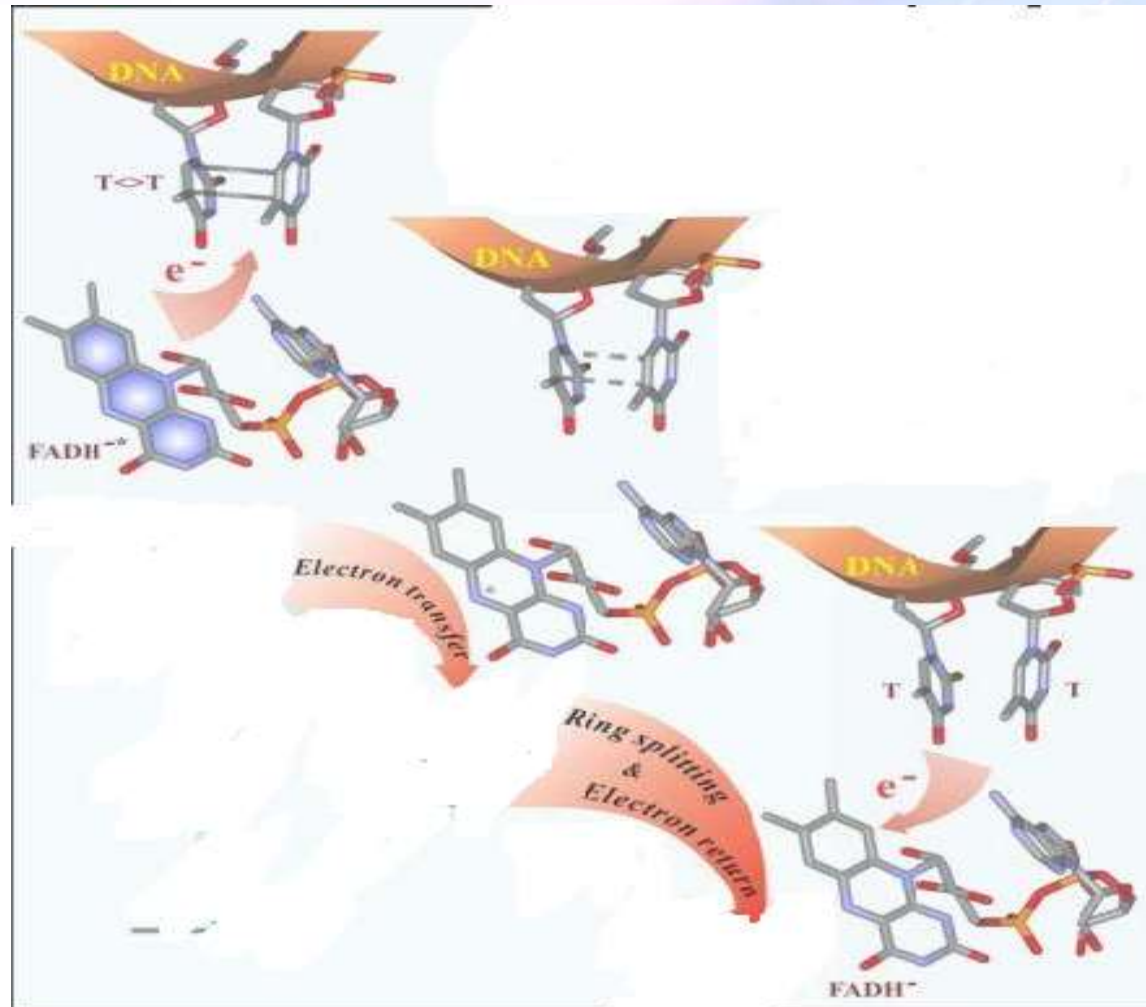
Mechanism of CPD Photolyase

- FADH^- then transfers an electron to the CPD to induce cyclobutane ring cleavage.

Mechanism of CPD Photolyase

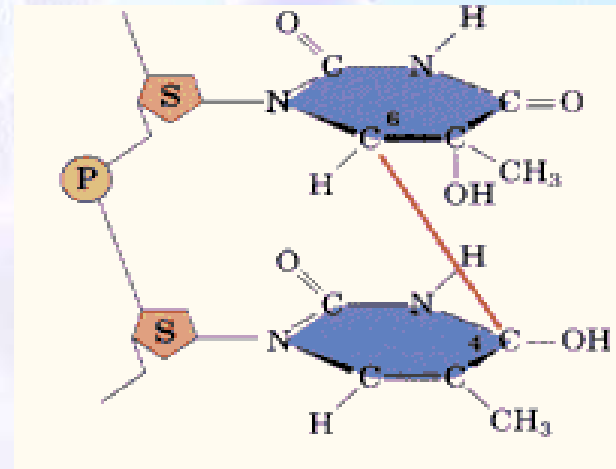
- The catalytic cycle is completed when the electron is transferred back from the repaired thymine to the FADH cofactor.

Mechanism of CPD Photolyase



(6-4) Photolyase

- UV irradiation also induces the formation of another type of pyrimidine dimer, the (6-4) photoproduct.



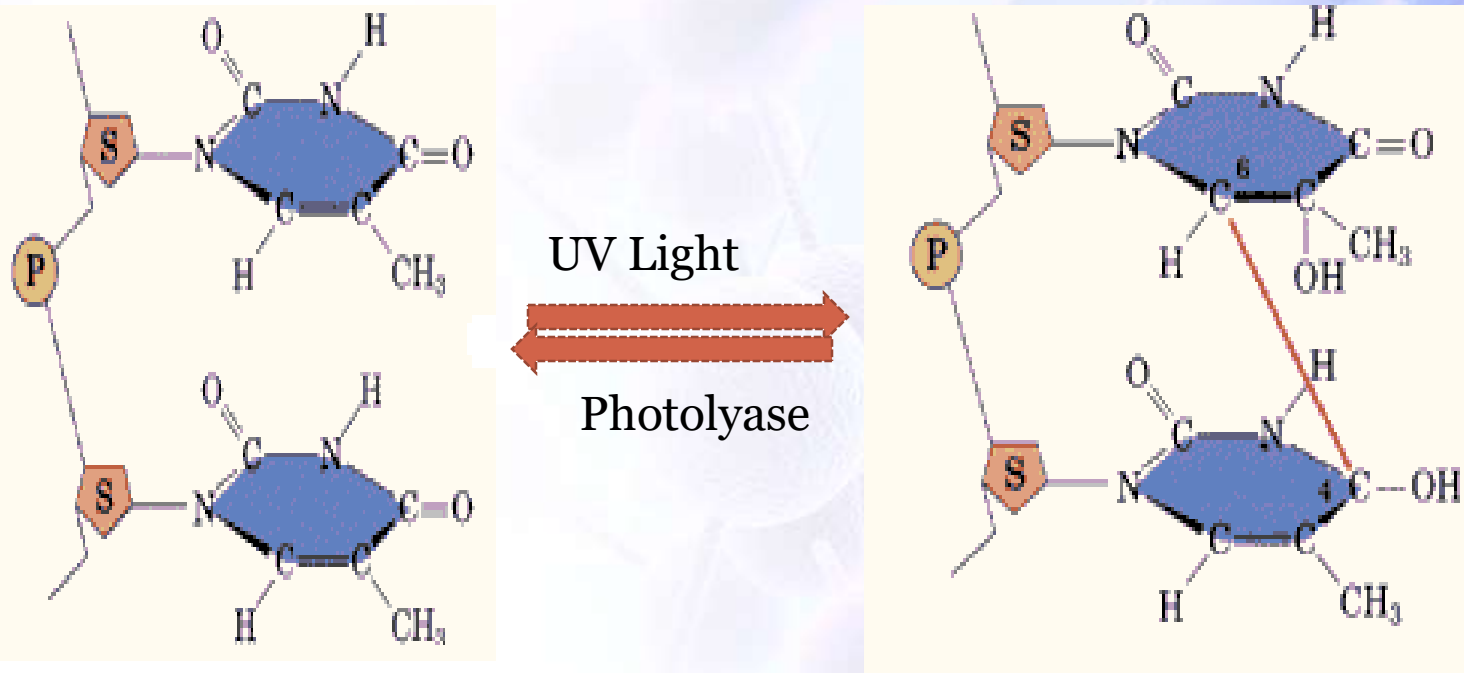
(6-4) Photolyase

- Takeshi Todo, Taisei Nomura and coworkers reported in 1993 that *Drosophila melanogaster* has a photolyase that reverses (6-4) photoproduct lesions in the DNA.

(6-4) Photolyase

- This photolyase was designated as (6 – 4) photolyase.
- It is widely distributed in plants and animals, but has not been detected in bacteria and mammals.

(6-4) Photolyase



(6-4) Photolyase

- Considerably less information is known about the mechanism of action of (6-4) photolyase as compared to CPD photolyase; however, the two enzymes seem to work in a similar way.

(6-4) Photolyase

- The (6-4) photolyase binds to damaged DNA, causing the (6-4) photoproduct to flip out of the DNA and into the active site of the enzyme.

(6-4) Photolyase

- In the active site of the enzyme, it undergoes a rearrangement to produce a product that receives an electron from an excited FADH^- molecule.

(6-4) Photolyase

- The final outcome is that the organisms containing (6-4) photolyase can use light energy to convert (6-4) photoproducts back to normal pyrimidine rings.

(6-4) Photolyase

- Organisms can also repair dimer lesions introduced by UV light by excising damaged nucleotides and replacing them with normal nucleotides.

(6-4) Photolyase

- This type of excision repair is the major pathway for repairing UV-induced damage to DNA in organisms such as humans that lack both the types of photolyases.

Damage Reversal by Dealkylation

- Another means of direct reversal of DNA damage is by dealkylation.
- Direct dealkylation reactions have probably been most extensively studied in *E. coli*.

Damage Reversal by Dealkylation

- Three different proteins can catalyze the direct removal of alkyl groups attached to oxygen atoms in DNA.
- These include:-

Damage Reversal by Dealkylation

- 1) O⁶-alkylguanine
DNA alkyltransferase I
- 2) O⁶-alkylguanine
DNA alkyltransferase II
- 3) Alkylguanine DNA
alkyltransferase

Damage Reversal by Dealkylation

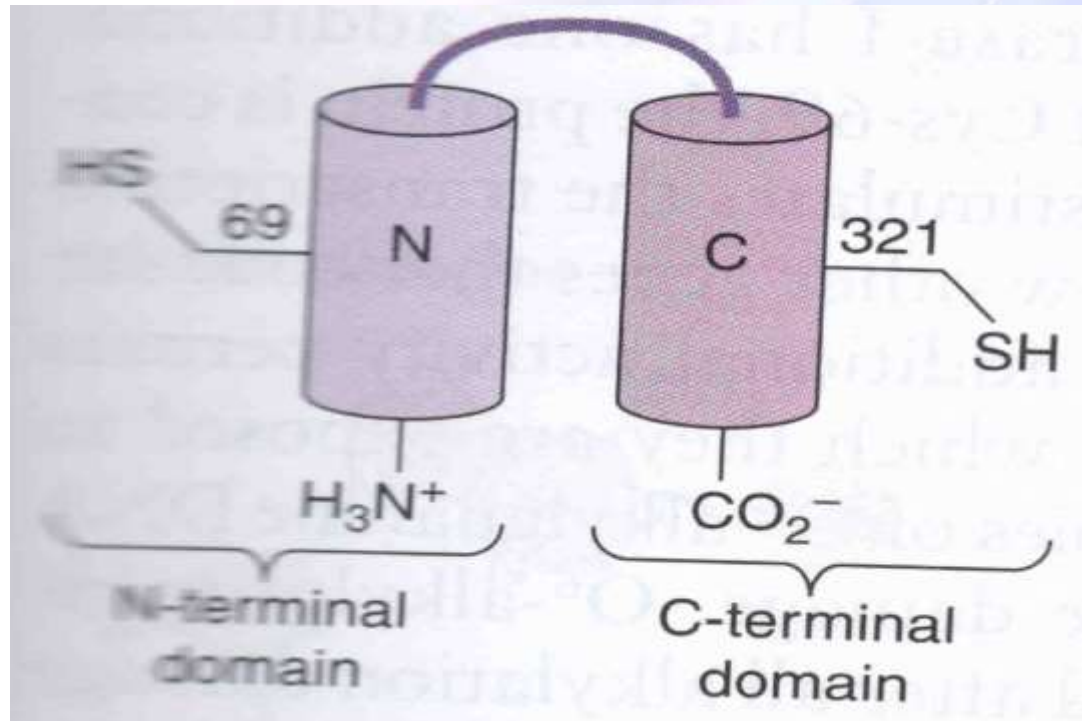
O⁶-alkylguanine DNA alkyltransferase I

- O⁶-alkylguanine DNA alkyltransferase I can remove methyl & other alkyl groups attached to O-6 in guanine.
- It can also remove alkyl groups attached to O-4 of thymine and to phosphotriesters.

Damage Reversal by Dealkylation

- This enzyme is a monomer that has a flexible linker connecting its N- and C- terminal domains.
- Each domain has an active site that performs a specific function.

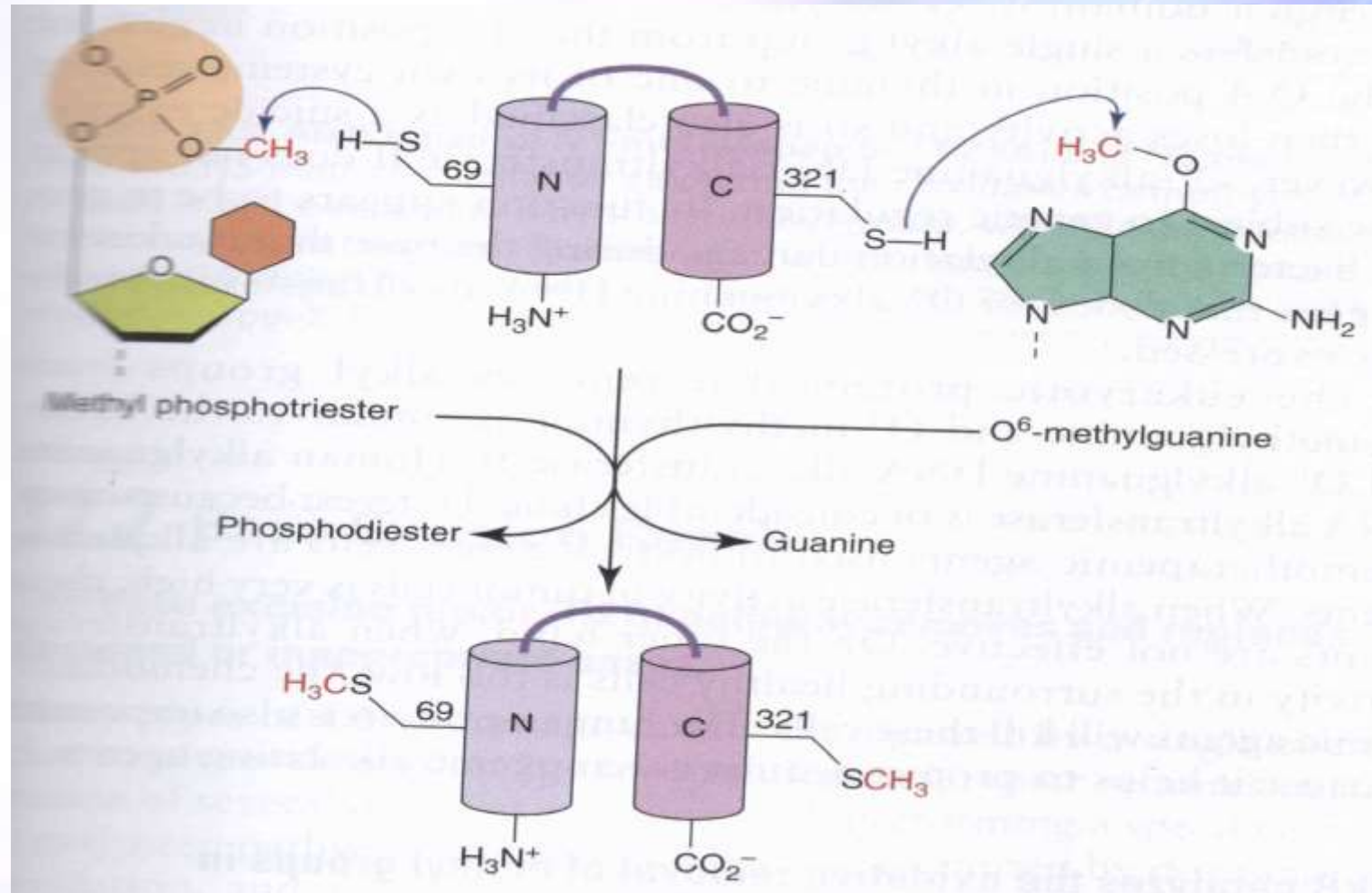
Damage Reversal by Dealkylation



Damage Reversal by Dealkylation

- The N- terminal domain transfers an alkyl group from an S_p phosphotriester to one of its own cysteine residues, Cys-69.

Damage Reversal by Dealkylation



Damage Reversal by Dealkylation

- The C- terminal domain transfers an alkyl group from either O⁶-alkylguanine or O⁴-alkylthymine to one of its own cysteine residues, Cys-321.

Damage Reversal by Dealkylation

- Once alkylated, the protein can't be regenerated and therefore behaves more like an alkyl transferring agent than an enzyme.

Damage Reversal by Dealkylation

- Such proteins as O⁶-alkylguanine DNA alkyltransferase I, which lose their activity after acting one time, are called **suicide enzymes**.

Damage Reversal by Dealkylation

- O⁶-alkylguanine DNA alkyltransferase I, has one additional remarkable function.
- After methylation at Cys-69, this protein is converted to a transcriptional activator.

Damage Reversal by Dealkylation

- This activator stimulates the transcription of the gene that codes for it as well as some other genes that code for the proteins that repair DNA damage.

Damage Reversal by Dealkylation

- This additional activity permits the bacteria to adapt to environments in which they are exposed to alkylating agents by synthesizing more copies of these enzymes which can repair damaged DNA.

Damage Reversal by Dealkylation

- O⁶-alkylguanine DNA alkyltransferase I that is synthesized after all the alkylation damage has been repaired will remain unmethylated.

Damage Reversal by Dealkylation

- This unmethylated form of the enzyme blocks transcription of the same genes that were activated by the methylated proteins.

Dealkylation Enzymes

O⁶-alkylguanine DNA alkyltransferase II

- O⁶-alkylguanine DNA alkyltransferase II is another E. coli's alkyl transferring enzyme.
- It has properties very similar to that of C-terminal domain of O⁶-alkylguanine DNA alkyltransferase I.

Dealkylation Enzymes

- It also transfers a single alkyl group from the O – 6 position in guanine or the O-4 position in thymine to one of its own cysteine residues.

Dealkylation Enzymes

- After this transfer, the enzyme loses its activity and so is also classified as a suicide enzyme.
- However, this enzyme doesn't appear to be subjected to genetic regulation.

Dealkylation Enzymes

- Its function appears to be to protect bacteria from alkylation damage during the time that it takes for the gene of O⁶-alkylguanine DNA alkyltransferase I to be fully expressed.

Dealkylation Enzymes

Alkylguanine DNA alkyltransferase

- The eukaryotic protein that removes alkyl group from O⁶-methylguanine or O⁴-methylthymine is similar to the bacterial O⁶-alkylguanine DNA alkyltransferase II.

Dealkylation Enzymes

- Human alkylguanine DNA alkyltransferase is of considerable clinical interest because many chemotherapeutic agents used to destroy cancer cells are alkylating agents.

Dealkylation Enzymes

- When alkyltransferase activity is very high in the tumor cells, these agents are not effective.

Dealkylation Enzymes

- On the other hand, when alkyltransferase activity in the surrounding healthy cells is too low, the chemotherapeutic agents will kill these cells.

Dealkylation Enzymes

- The human protein/enzyme is also important because it helps to protect the cells against carcinogenic alkylating agents.

Dealkylation Enzymes

AlkB

- An entirely different type of alkylation damage repair activity was found in *E. Coli* in 2002 which is carried out by a protein AlkB.

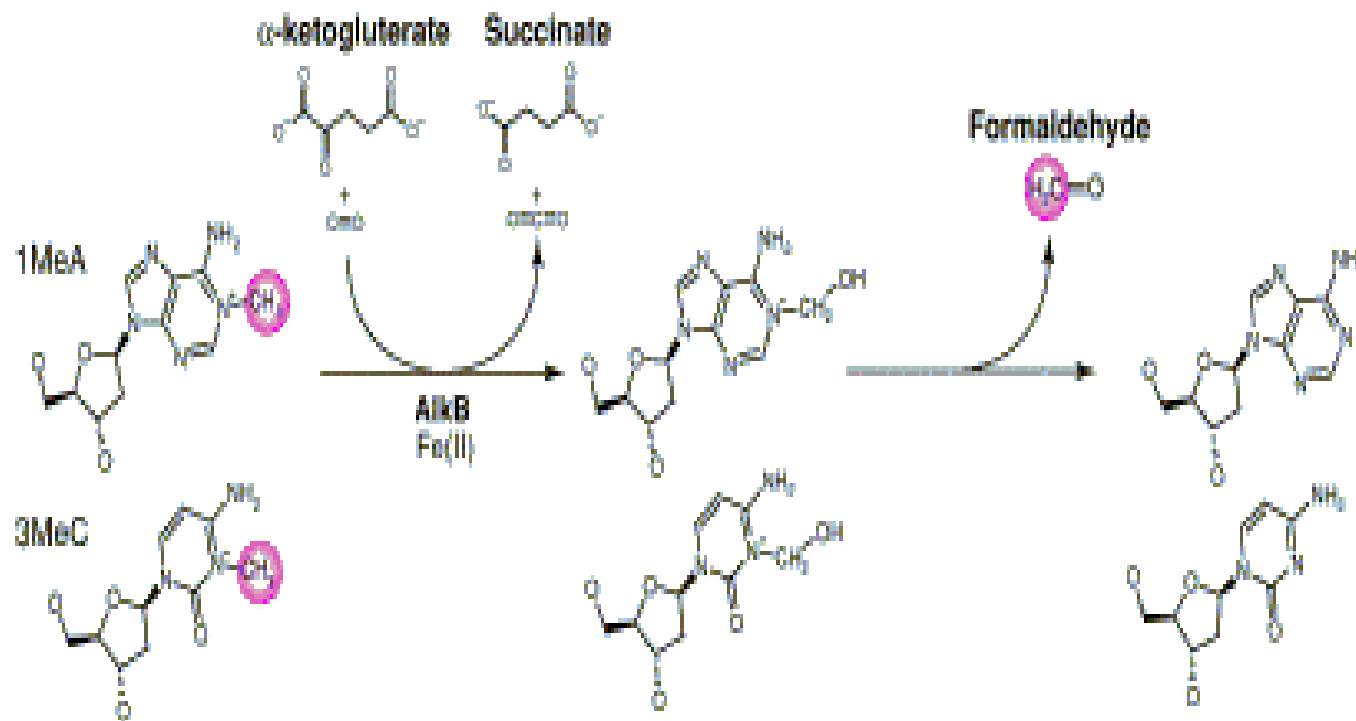
Dealkylation Enzymes

- AlkB catalyzes the direct conversion of 1-methyladenine, 1-methylguanine, 3-methylcytosine and 3-methylthymine to adenine, guanine, cytosine and thymine, respectively.

Dealkylation Enzymes

- The AlkB catalyzed reaction requires, Fe^{2+} , molecular oxygen, and α - ketoglutarate.
- A similar type of enzyme has been found in other organisms including human & mammals.

Dealkylation Enzymes



Base Excision Repair

- Many types of DNA damages can't be repaired by a single enzyme that catalyzes direct damage reversal.

Base Excision Repair

- Instead, repair requires the participation of several different enzymes, each performing a specific task in multistep pathway.

Base Excision Repair

- Damage to DNA bases caused by deamination, oxidation, and alkylation is mainly repaired by one such multistep pathway which is called, **Base Excision Repair (BER)**.

Base Excision Repair

- BER pathway is same in all organisms.
- Enzymes involved in the base excision repair pathway also participate in the repair of single-strand break in DNA.

Base Excision Repair

- Base excision repair derives its name from the first step in the pathway, N-glycosyl bond cleavage.
- This cleavage excises the damaged or inappropriate base from the DNA to form an abasic site.

Base Excision Repair

- Because no single enzyme can distinguish the four bases present in DNA from a wide variety of altered bases, cells must use many different enzymes to perform this function.

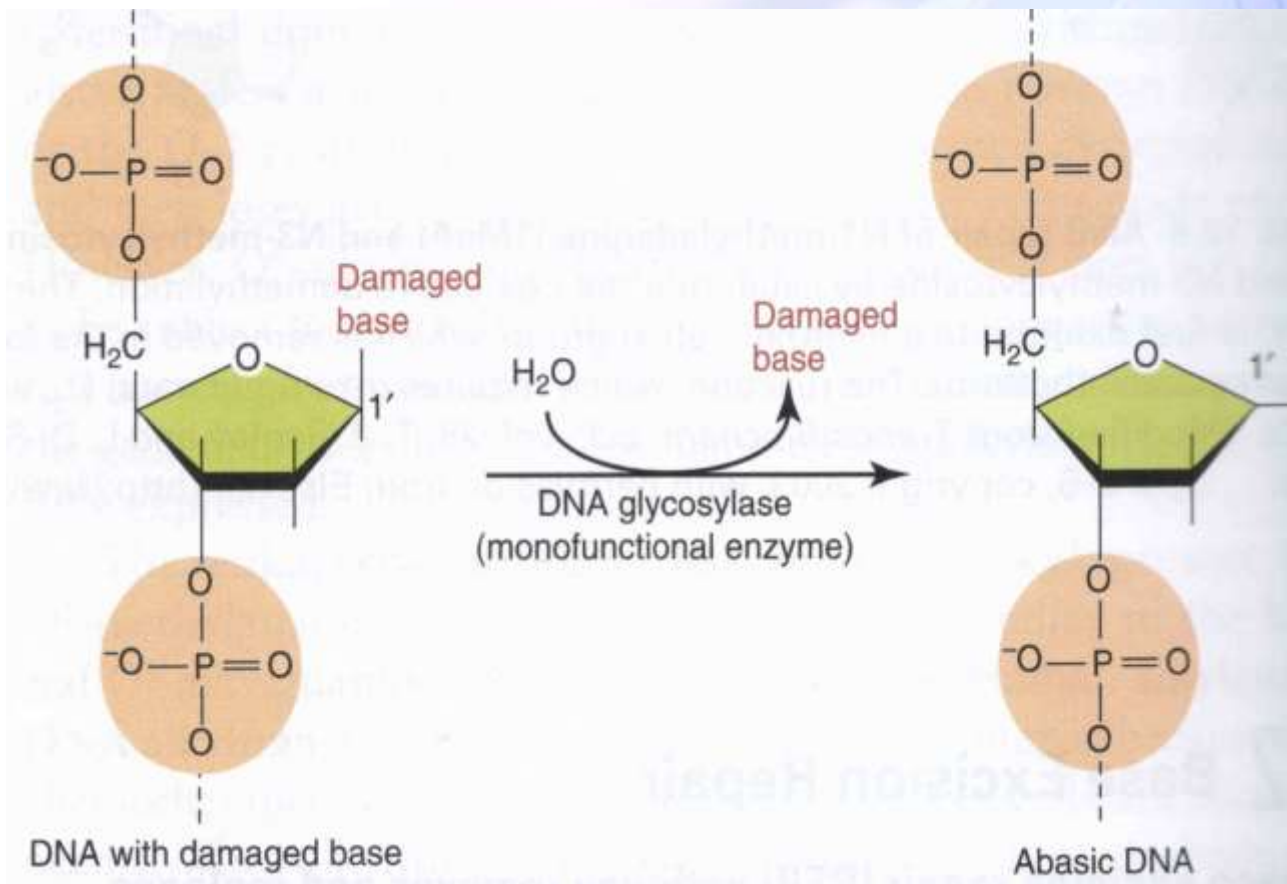
Base Excision Repair

- Some N-glycosylases are monofunctional enzymes with only the ability to excise a damaged base.
- Such enzymes are called **DNA glycosylases**.

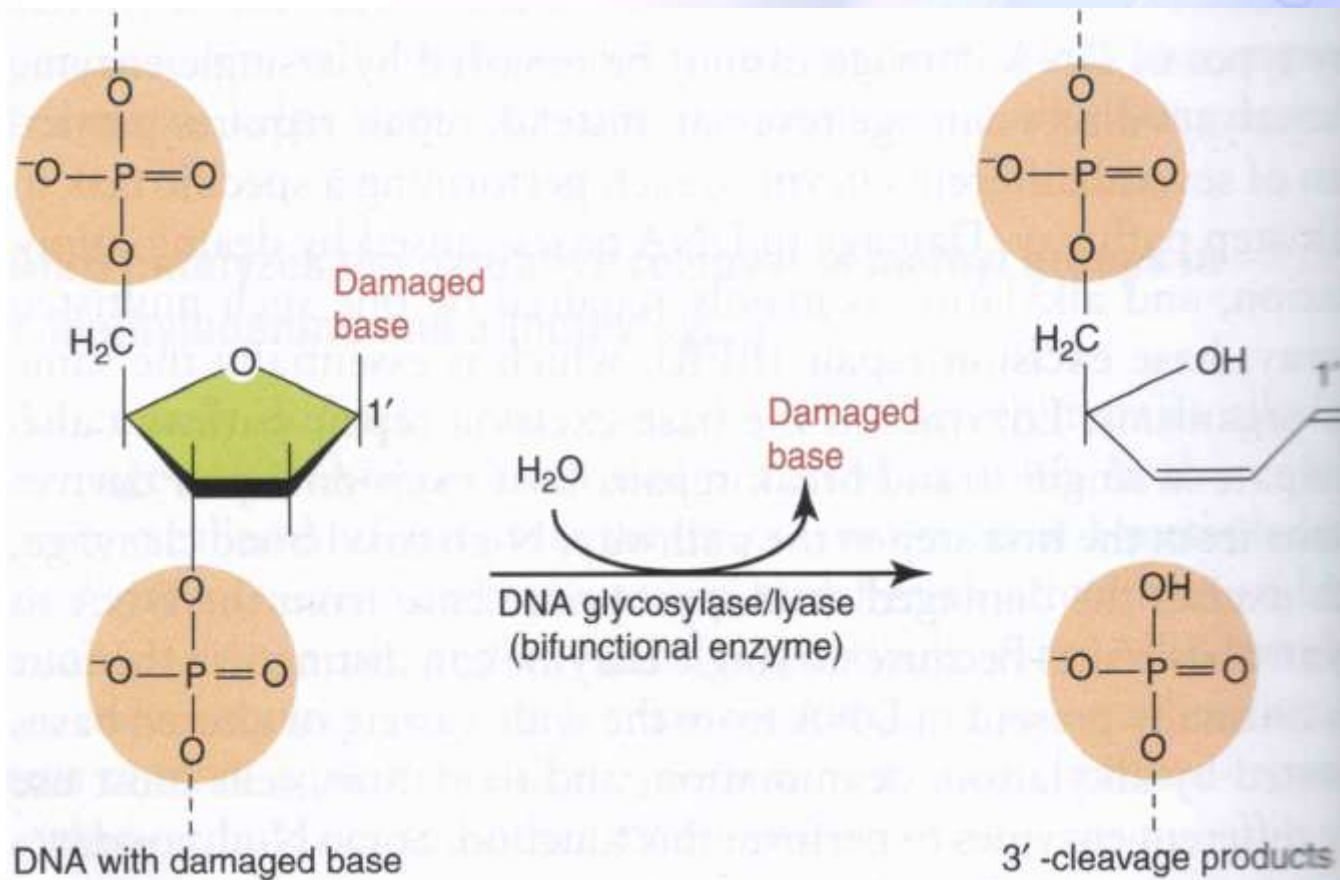
Base Excision Repair

- Other N-glycosylases have an AP lyase activity that cleaves the bond between sugar and the phosphate 3' to the damaged site.
- This enzymes is designated as DNA glycosylase/lyase.

Base Excision Repair



Base Excision Repair



Base Excision Repair

- Both types of enzymes detect the damaged base, flip it out of the DNA helix into an active site pocket, and then cleave the N-glycosyl bond.

Base Excision Repair

- Although some glycosylases excise a specific base, most have a somewhat broader specificity.
- The *E.coli* enzyme **Uracil N-glycosylase (Ung)** is specific for uracil.

Base Excision Repair

- Whereas another *E. coli* N-glycosylase, 3-methyladenine DNA glycosylase (AlkA), acts on 3- or 7-methyl purines, 3- or 7-ethylpurine, ethenoadenine, and O²-methyl pyrimidine.

Base Excision Repair

- A null mutation in a gene that codes for a DNA glycosylase or DNA glycosylase/lyase is not lethal which probably reflects overlapping functional abilities among the glycosylases.

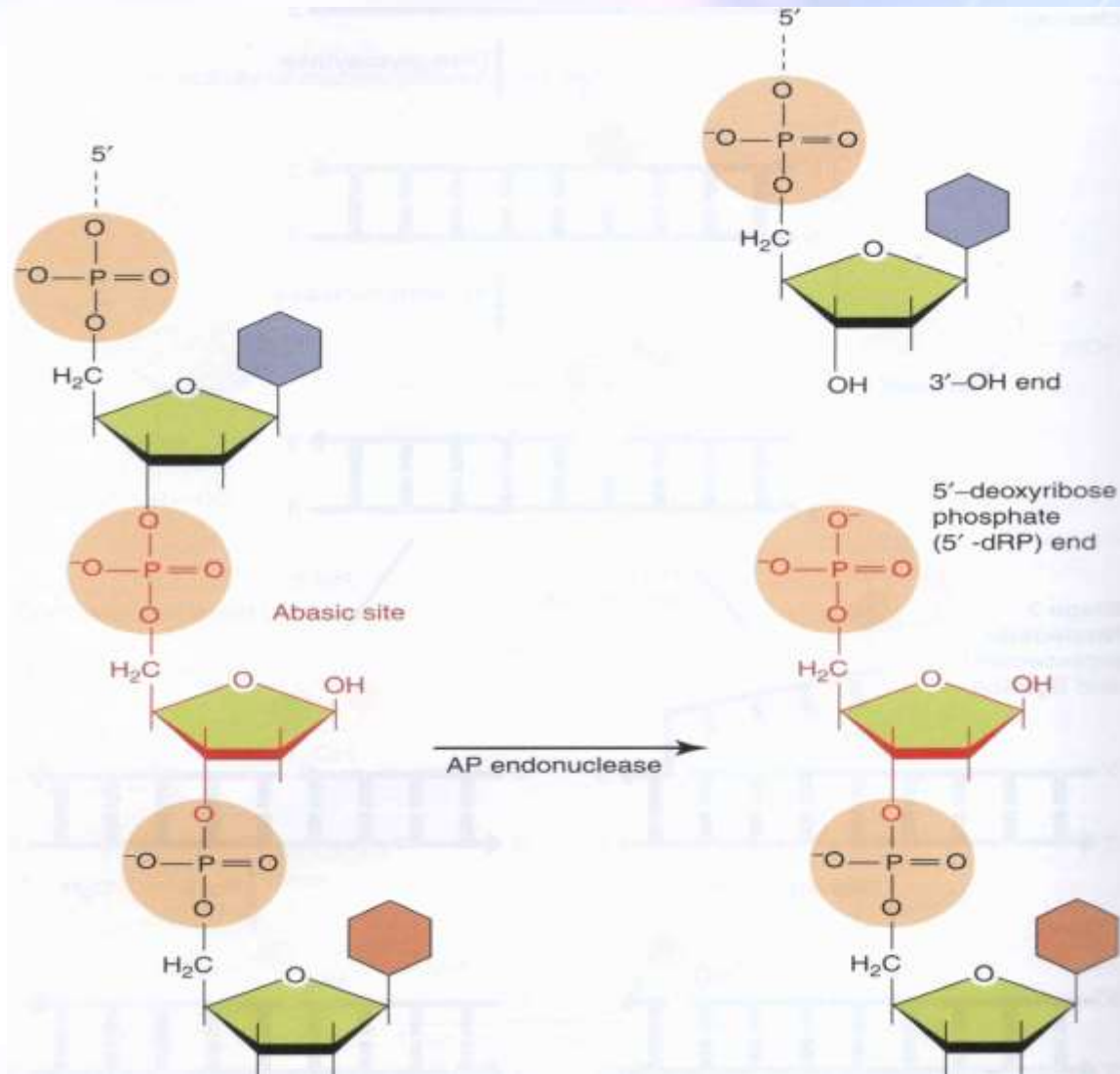
Base Excision Repair Pathway

- The DNA glycosylase catalyzes base excision to produce an AP (apurinic and apyrimidininc) site.

Base Excision Repair Pathway

- The next enzyme in the pathway, **AP endonuclease**, hydrolyzes the phosphodiester bond 5' to the AP site to generate a nick.

Base Excision Repair Pathway



Base Excision Repair Pathway

- *E.coli* has two well-characterized AP endonucleases:-
 - Exonuclease III (Xth)
 - Endonuclease IV (Nfo)
- Exonuclease III (Xth) despite its name accounts for most of the bacterial AP endonuclease activity.

Base Excision Repair Pathway

- Both are multifunctional enzymes that have 3'-phosphate and 3'-repair phosphodiesterase activities.
- The former activity removes phosphate groups from the 3' end of a DNA strand.

Base Excision Repair Pathway

- The latter activity i.e., 3'-repair phosphodiesterase activity removes the 3'-unsaturated aldehydic group produced by DNA glycosylase/lyase action.

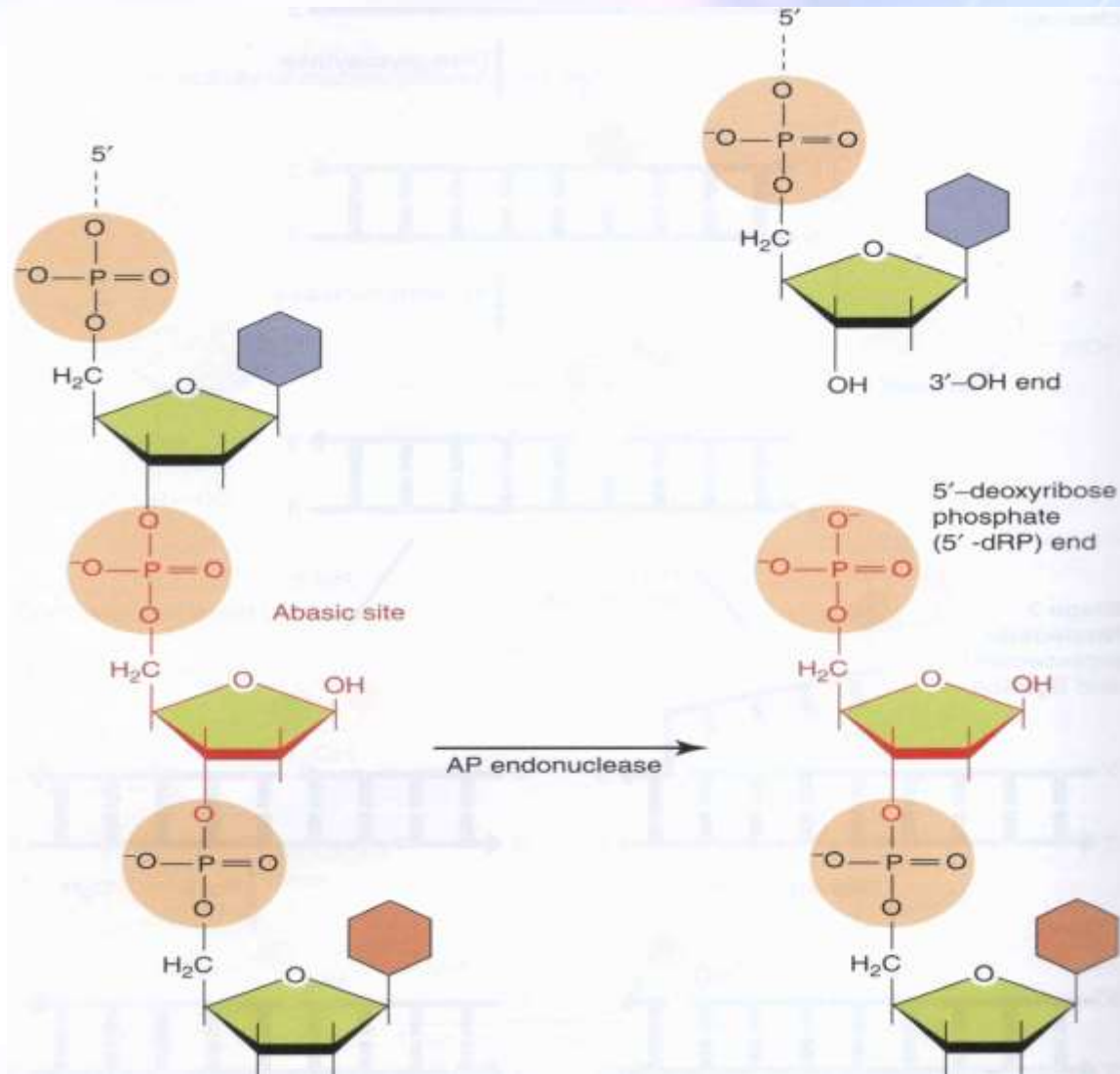
Base Excision Repair Pathway

- These activities are important because DNA polymerase can't attach new nucleotides to a blocked 3'-end.
- The mammalian AP endonuclease, **APE1** is homologous to *E. coli* exonuclease III.

Short Patch Repair

- Base excision by DNA glycosylase and strand cleavage by AP endonuclease introduces a gap with a 5'-deoxyribose phosphate (5'-dRP) on one side and a 3'-OH on the other.

Short Patch Repair



Short Patch Repair

- Additional enzymes are required to fill in the gap and remove the 5'-deoxyribose phosphate.
- Cells can repair the damage by two different pathways viz. short patch repair & long patch repair.

Short Patch Repair

- In the **short patch repair**, only a single nucleotide is replaced.
- This repair pathway involves the following enzyme catalyzed reactions:-

Short Patch Repair

- (1) DNA polymerase adds a deoxyribonucleotide to the 3'-OH end.
- (2) Deoxyribose phosphate lyase (dRPase) removes 5'-deoxyribose phosphate from the 5'-end; and

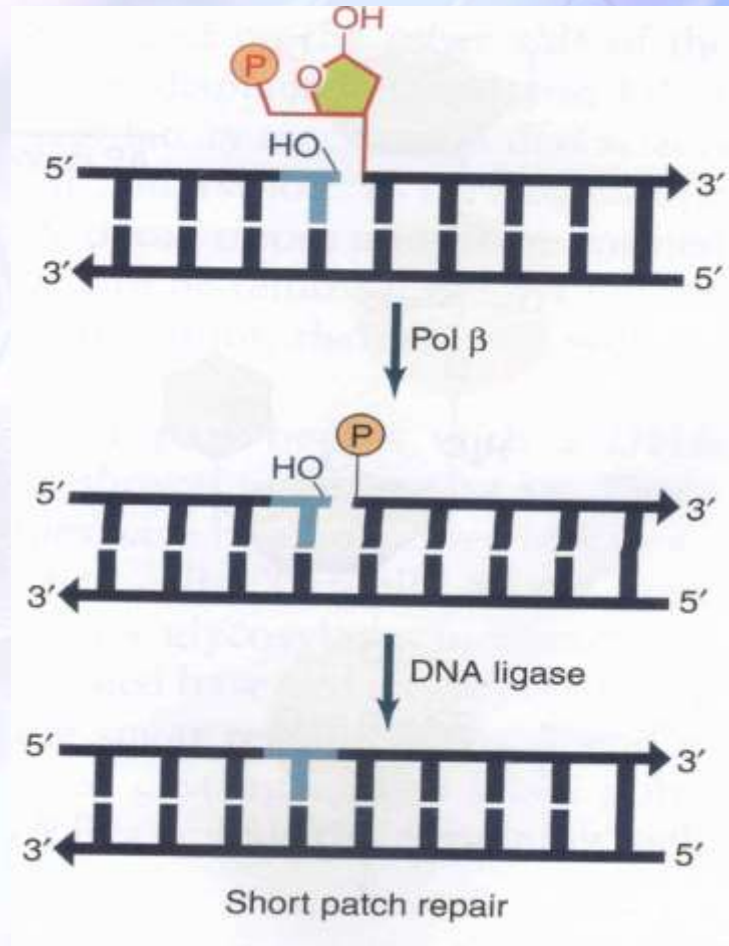
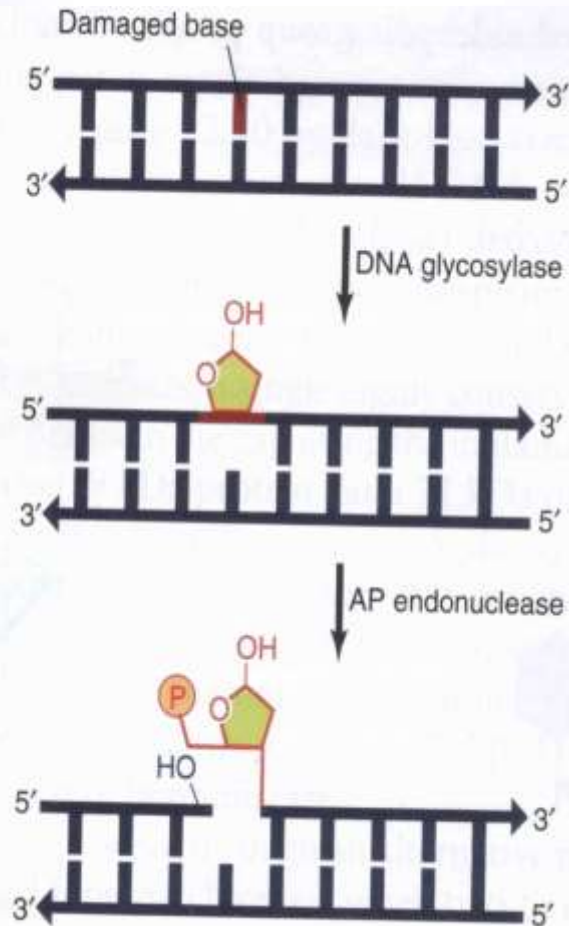
Short Patch Repair

- (3) DNA ligase joins the adjacent ends.
- In *E. coli*, DNA polymerase I fills in the gap.

Short Patch Repair

- While in eukaryotes, a single highly conserved enzyme, DNA polymerase β (Pol β) fills in the gap using the undamaged DNA strand as the template.

Short Patch Repair



Short Patch Repair

- The eukaryotic enzyme differs from its bacterial counterpart in one very important respect, i.e., it lacks 3' → 5' proofreading activity._

Short Patch Repair

- Two mammalian 3' exonucleases, **TREX1** and **TREX2** may correct the errors introduced by Pol β .

Long Patch Repair

- The alternative pathway to repair the damage in mammalian cells is termed as **long patch repair**.
- In this pathway, 2 – 8 nucleotides are replaced.

Long Patch Repair

- In this case, DNA polymerase δ or ϵ catalyzes chain extension with the assistance of the RFC (replication factor C) clamp loader and the PCNA sliding clamp.

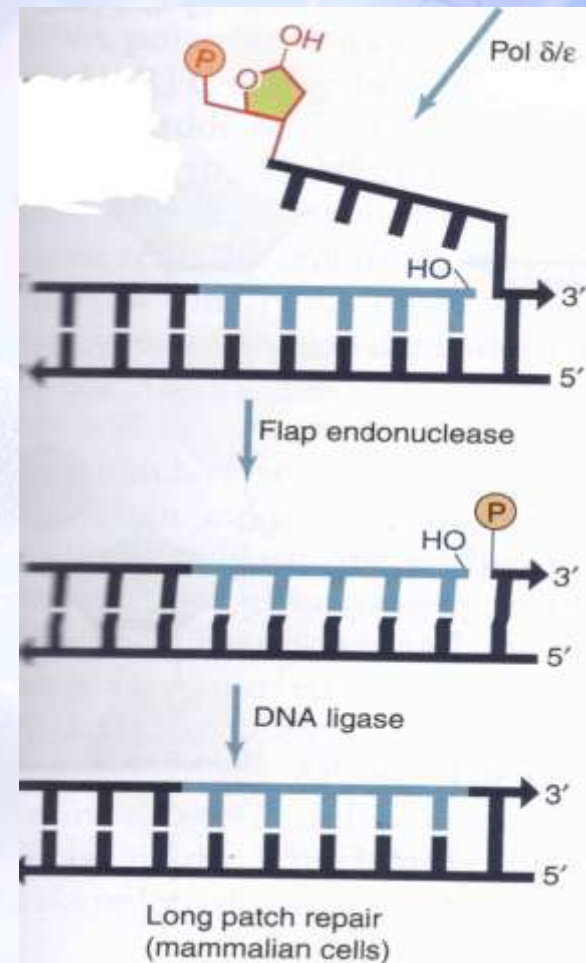
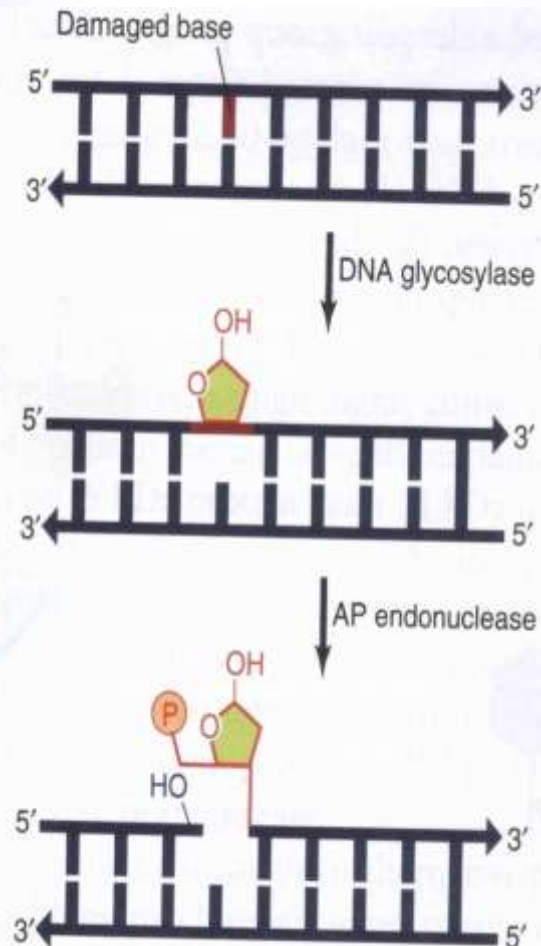
Long Patch Repair

- As the polymerase adds nucleotides to the 3'-OH end on one side of the gap, it displaces the 5'-deoxyribose phosphate end on the other side of the gap.

Long Patch Repair

- The flap endonuclease cleaves the displaced strand and DNA ligase seals the remaining nick.
- The regulatory mechanism that selects long or short patch repair is not well understood.

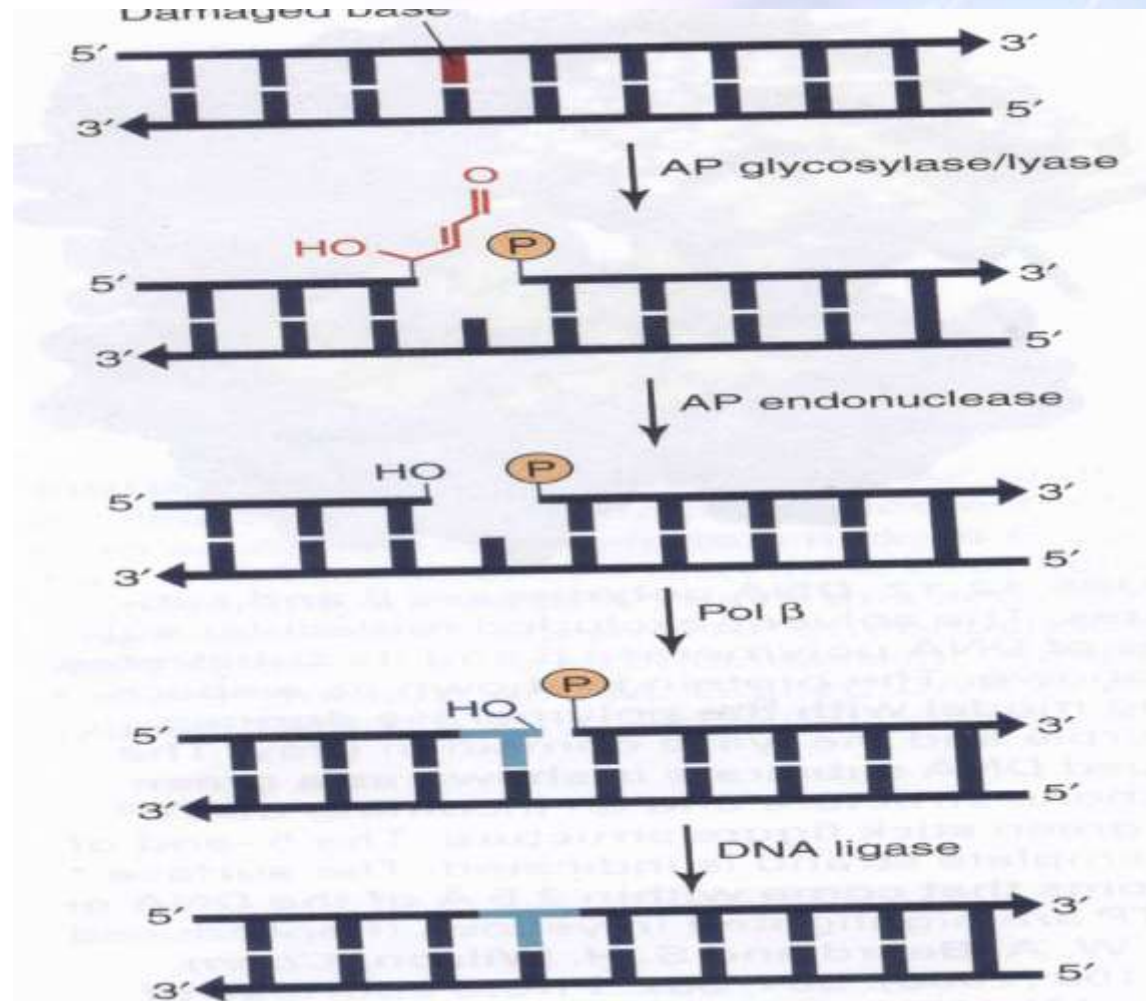
Long Patch Repair



Long Patch Repair

- When eukaryotic base excision repair begins with a DNA glycosylase/lyase, the pathway is shown below:-

Long Patch Repair



Long Patch Repair

- Three distinct DNA glycosylase/lyase enzymes have been identified in *E. coli*, three in *S. cerevisiae*, and six in human cells.

Long Patch Repair

- Virtually all oxidized bases are removed by bifunctional DNA glycosylases in mammals.
- DNA glycosylase/lyase excises the damaged base and cleaves the DNA strand 3' of the AP site.

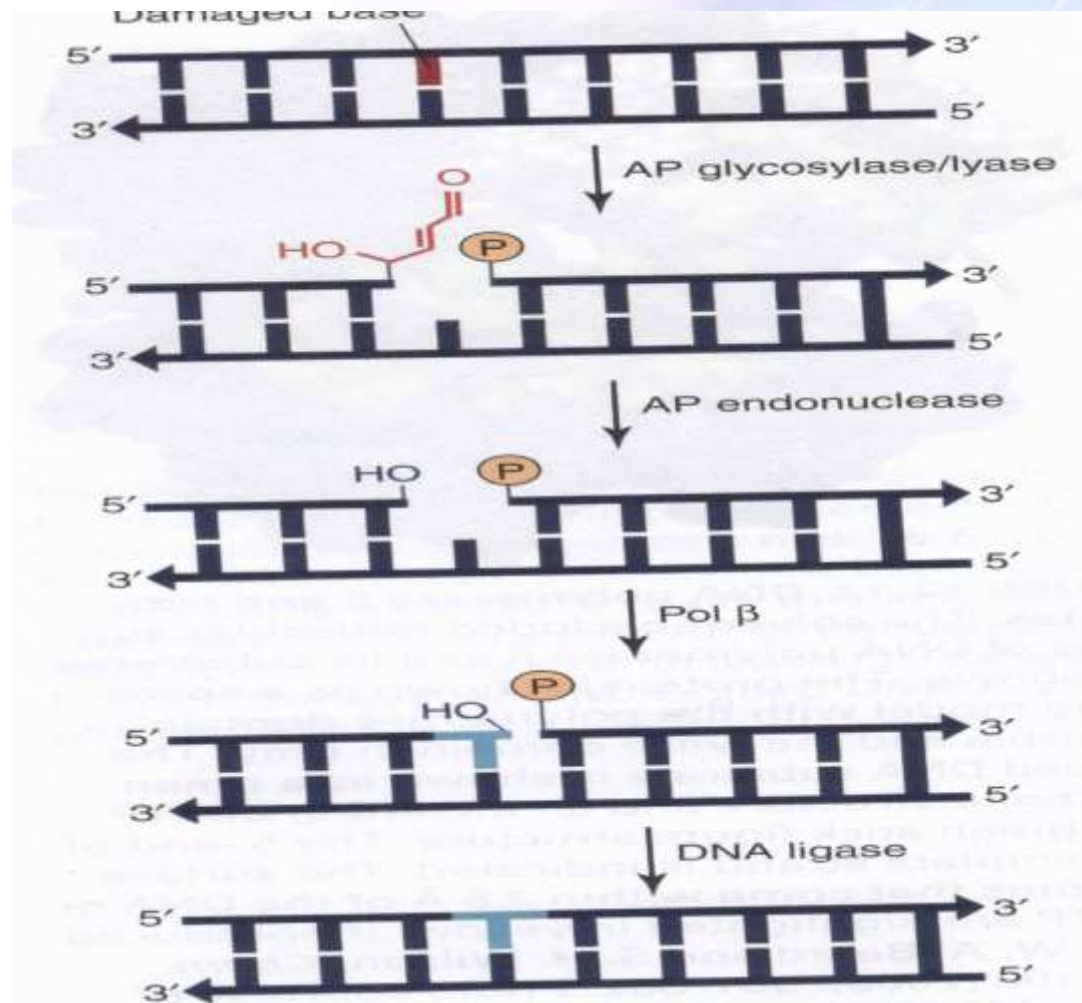
Long Patch Repair

- The resulting sugar residue at the 3'-end is removed by the AP endonuclease catalyzed cleavage.

Long Patch Repair

- Then DNA polymerase β adds a nucleotide and DNA ligase seals the remaining nick to complete short patch repair.

Long Patch Repair



Nucleotide Excision Repair

- Nucleotide excision repair (NER) pathway removes bulky adducts from DNA by excising an oligonucleotide bearing the lesion and replacing it with new DNA.

Nucleotide Excision Repair

- This repair mechanism excises UV-induced cyclobutane pyrimidine dimers, (6-4) photoproducts, damaged bases formed by alkylating agents and certain types of cross-links.

Nucleotide Excision Repair

- The efficiency of repair for different kinds of lesions can vary a lot.
- In general, there is a direct relationship between the amount of helical distortions produced by lesion and the efficiency of this repair.

Nucleotide Excision Repair

- The basic nucleotide excision repair pathway is same in all the organisms.
- It involves the following steps:-
 - 1. damage recognition

Nucleotide Excision Repair

- 2. an incision in the damaged DNA strand on each side of the lesion,
- 3. excision of the oligonucleotide created by the incision,

Nucleotide Excision Repair

- 4. synthesis of new DNA to replace the excised DNA segment using the undamaged DNA strand as a template, and
- 5. ligation of the remaining nick.

Nucleotide Excision Repair

- Although the basic nucleotide excision repair pathway is similar in all the organisms, there are considerable differences in the proteins that carry out the various steps.

Nucleotide Excision Repair of UV-induced Damage

- UV-irradiated *E. coli* can regain their ability to survive after incubation in dark.
- However, they recover more slowly than when incubated in the light.

Nucleotide Excision Repair of UV-induced Damage

- This observation suggests that the bacteria use some process other than photoreactivation to repair the UV-induced DNA damage.

Nucleotide Excision Repair of UV-induced Damage

- Richard Setlow & William Carrier and, working independantly, Richard Boyce & Paul Howard-Flanders, used a similar approach to investigate this alternative process in 1964.

Nucleotide Excision Repair of UV-induced Damage

- Both groups cultured *E. coli* in the presence of [³H]thymine to label the DNA and then irradiated the cells with UV light to induce the formation of cyclobutane thymine dimer.

Nucleotide Excision Repair of UV-induced Damage

- Then they:-
- 1) incubated the UV-irradiated cells in the dark so that the photoreactivation could not take place;
- 2) removed samples after various incubation times and added TCA to them;

Nucleotide Excision Repair of UV-induced Damage

- 3) separated acid-insoluble DNA from acid soluble nucleotides;
- 4) digested the DNA and oligonucleotides to release intact thymine cyclobutane dimers; and

Nucleotide Excision Repair of UV-induced Damage

- 5) detected the released dimers by chromatography.
- The experiments revealed that as the incubation time in the dark increases,

Nucleotide Excision Repair of UV-induced Damage

- Cyclobutane thymine dimers disappear from the acid-insoluble DNA and appear in the acid soluble oligonucleotide fraction.

Nucleotide Excision Repair of UV-induced Damage

- These results were correctly interpreted to mean that bacteria can excise an oligonucleotide containing a lesion and replace the excised oligonucleotides with newly synthesized DNA.

Nucleotide Excision Repair of UV-induced Damage

- Subsequent studies showed that eukaryotes and the archaea also have nucleotide excision repair pathways.

UvrA, UvrB, and UvrC Proteins

- Genetic studies revealed that three *E. coli* genes viz., *uvrA*, *uvrB*, and *uvrC*, code for proteins that are essential for damage recognition, incision, and excision.

UvrA, UvrB, and UvrC Proteins

- All three genes have been cloned and the proteins that they encode (UvrA, UvrB, and UvrC) have been purified and characterized.

UvrA, UvrB, and UvrC Proteins

- Under normal physiological conditions, *E. coli* has about 25 molecules of UvrA, 250 molecules of UvrB, and 10 molecules of UvrC.

UvrA, UvrB, and UvrC Proteins

- After UV damage of DNA, UvrA and UvrB levels increase ten- and four folds, but the UvrC level remains the same.

UvrA, UvrB, and UvrC Proteins

- Although UvrA, UvrB, and UvrC do not combine to form a stable ternary complex, the polypeptides nevertheless are said to be part of a **UvrABC damage-specific endonuclease**.

UvrA, UvrB, and UvrC Proteins

- The UvrABC damage-specific endonuclease is called as UvrABC endonuclease in short but some investigators preferably term them as **UvrABC excinuclease**.

UvrA, UvrB, and UvrC Proteins

- **UvrABC excinuclease** is termed so because the proteins participate in incision and excision reactions.

UvrA, UvrB, and UvrC Proteins

- The three polypeptides work in the order suggested by their names, that is their order of action is UvrA, UvrB, and the UvrC.

The NER Pathway

- The crystal structure of UvrA•DNA complex reveals that UvrA does not make direct contact with the modified thymine but does bind to DNA regions on either side of the lesion.

The NER Pathway

- Based on this information, it appears that UvrA makes an important contribution towards recognition of DNA damaged lesion.
- This UvrA•DNA complex is formed in vitro in the absence of UvrB.

The NER Pathway

- The recognition process appears to be more complicated in vivo where UvrA is a part of a $(UvrA)_2 \bullet UvrB$ complex.

The NER Pathway

- UvrB plays a central role in nucleotide excision repair by interacting with both the UvrA and UvrC, although not at the same time.
- UvrB has at least two catalytic sites that are essential for its function.

The NER Pathway

- UvrB combines with UvrA to form a (UvrA)₂•UvrB complex either in solution or on DNA.
- Initially, this protein complex binds to DNA at some distance from the damaged site.

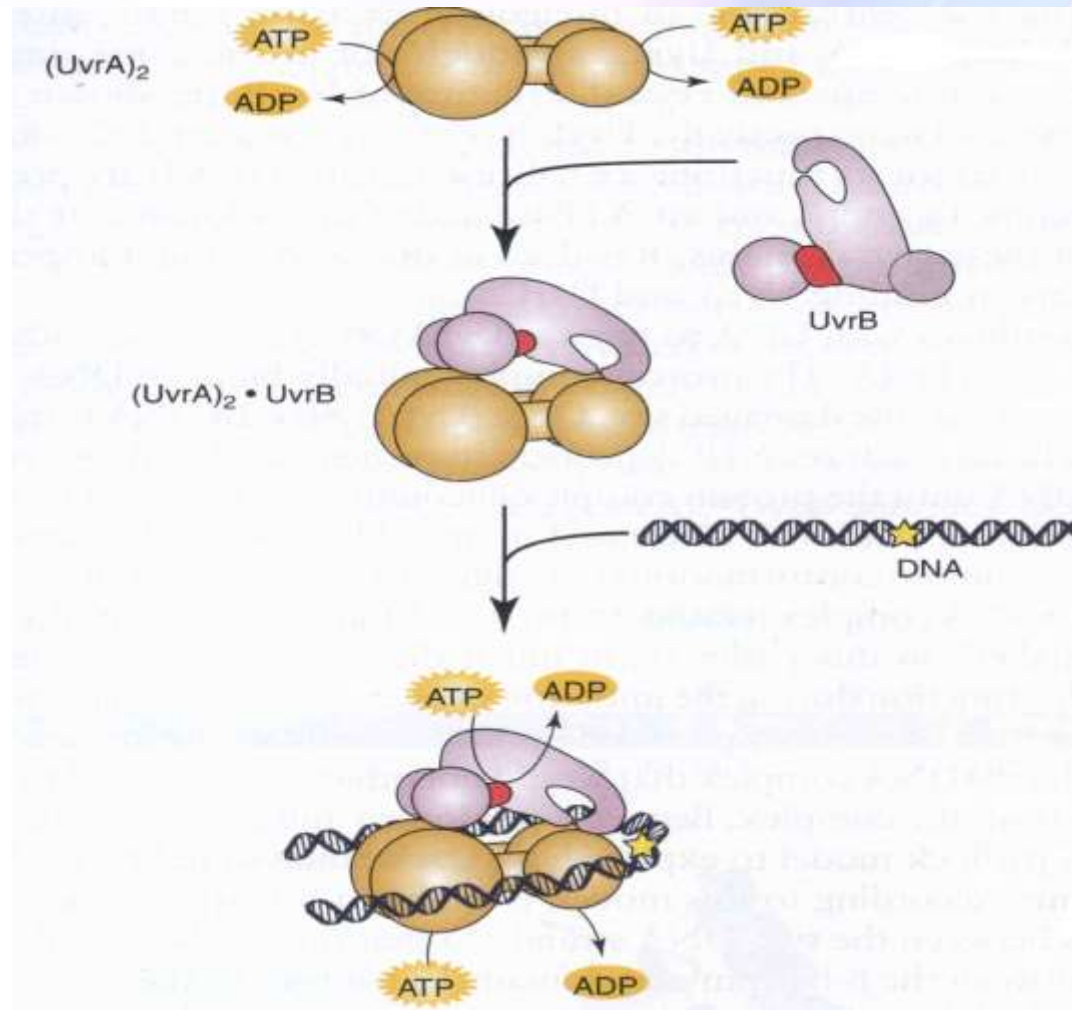
The NER Pathway

- Once (UvrA)₂•UvrB • DNA complex is formed, the UvrB helicase catalyzes ATP-dependant movement of (UvrA)₂•UvrB along the DNA until the protein complex encounters a bulky adduct or helix distortion.

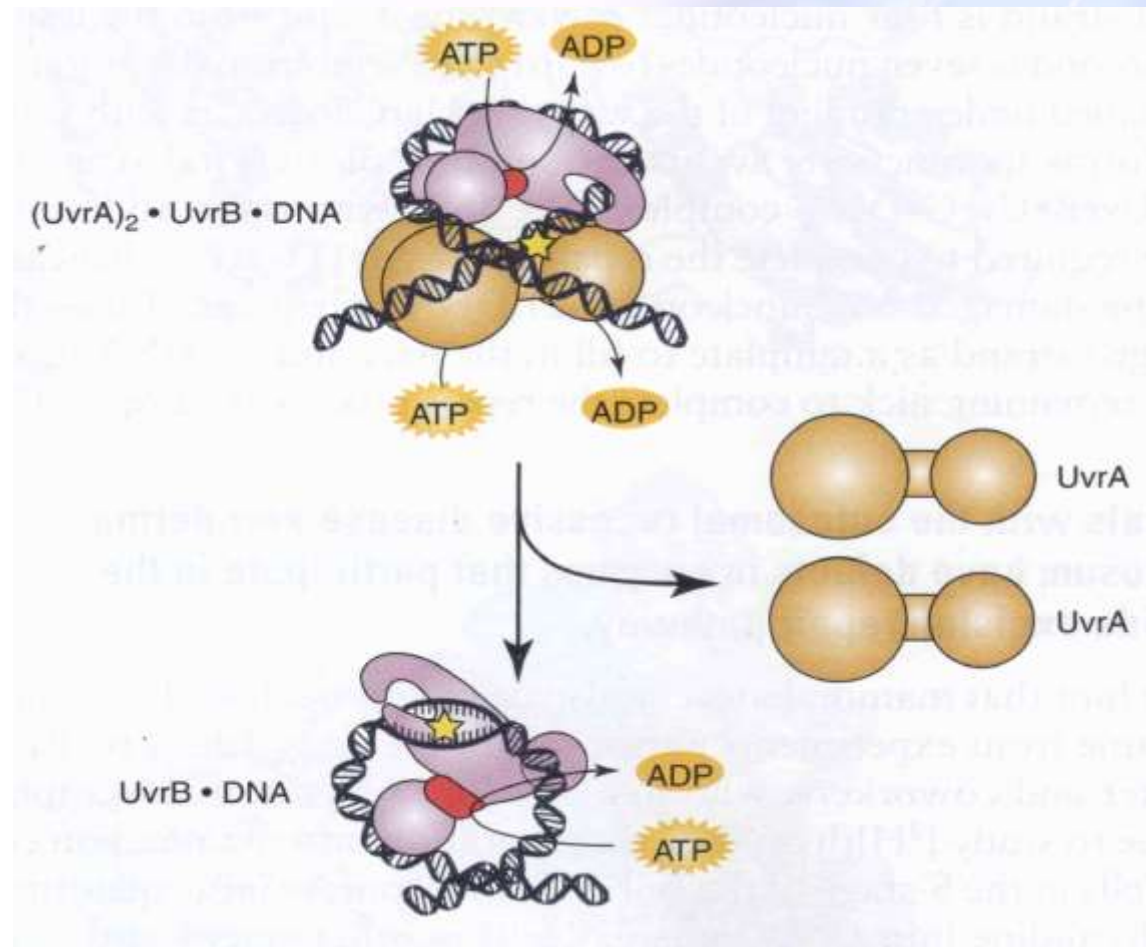
The NER Pathway

- The UvrA is released in an ATP-dependant reaction with a concomitant conformational change in the UvrB that produces a stable UvrB•DNA complex.

The NER Pathway



The NER Pathway



The NER Pathway

- UvrA functions as a “molecular match maker” in this pathway.
- It uses energy of ATP to facilitate the formation of UvrB•DNA complex and then dissociates from the complex.

The NER Pathway

- UvrC, which has a flexible linker that connects its N- and C-terminal domains, binds to the UvrB•DNA complex and makes two incisions, one on each side of the lesion.

The NER Pathway

- The first incision of the damaged strand is four nucleotides towards the 3'-end from the lesion and the second is seven nucleotides toward the 5'-end from the lesion.

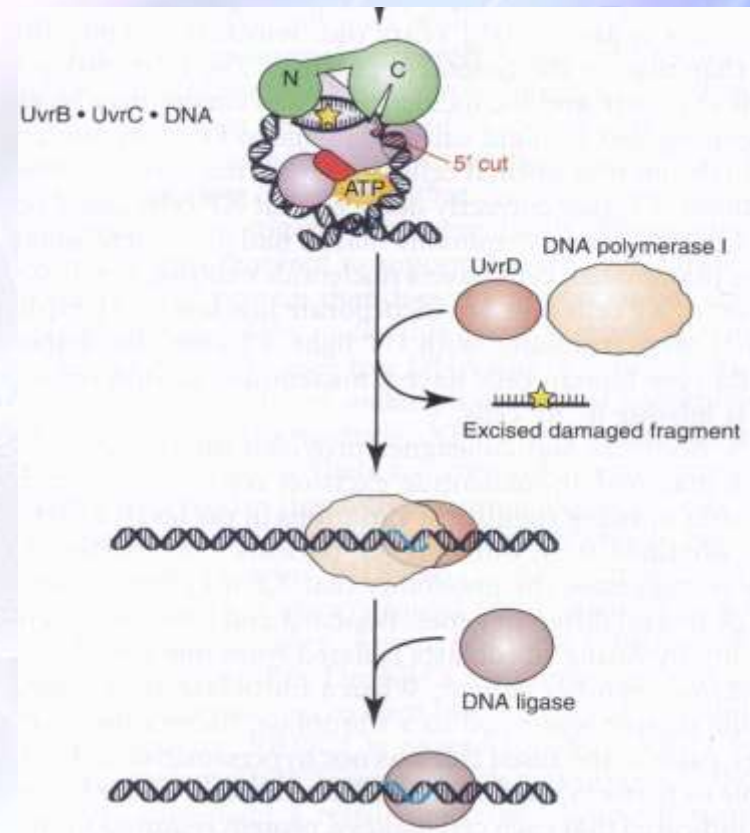
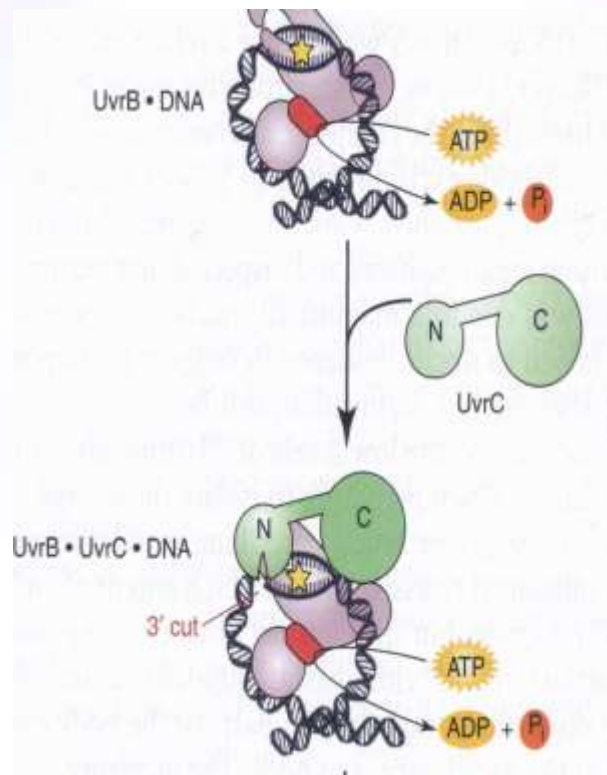
The NER Pathway

- How UvrC interacts with UvrB and performs its function is still to be described.
- Following three more steps are required to complete the repair process:-

The NER Pathway

- 1). UvrD, a helicase, excises the damaged oligonucleotide;
- 2). DNA polymerase I uses the undamaged strand as a template to fill the gap;
- 3). DNA ligase seals the remaining nick to complete the repair process.

The NER Pathway



Mismatch Repair

- The mismatch repair system corrects rare base pair mismatches and short insertions or deletions that appear in DNA following replication.

Mismatch Repair

- DNA polymerases introduce about one mispaired nucleotide per 10^5 nucleotides.
- However, the 3' → 5' proofreading exonuclease increases replication fidelity by 100- fold by removing mispaired nucleotides.

Mismatch Repair

- Although an error frequency of one nucleotide in 10^7 may seem extremely low, it would result in a high mutation rate.

Mismatch Repair

- The second type of error that occurs during replication is the short insertions and deletions, which result from the fact that repeated-sequence motifs sometimes dissociate and then re-anneal incorrectly.

Mismatch Repair

- As a result, the newly synthesized strand will have a different number of repeats than the template strand.
- Introduction of an insertion or deletion into the newly synthesized DNA is likely to produce a mutation.

Mismatch Repair

- Cells with a non-functional mismatch repair system have a high rate of mutation due to their inability to efficiently repair base pair mismatches, short insertions or deletions that arise during replication.

Mismatch Repair System in *E. coli*

- Let us begin the examination of mismatch repair by considering the *E. coli* mismatch repair system because this system has been the most extensively studied.

Mismatch Repair System in *E. coli*

- Although this system provides valuable informations for studying mismatch repair in other organisms, it differs from the mismatch repair systems of gram-positive bacteria & eukaryotes in one important respect.

Mismatch Repair System in *E. coli*

- The *E. coli* mismatch repair system can distinguish a newly synthesized strand from a parental strand because only the latter has methyl groups attached to sites with the sequence GATC.

Mismatch Repair System in *E. coli*

- *E. coli* has a deoxy-adenosine methylase that transfers methyl groups from S-adenosylmethionine molecules to deoxyadenosines in GATC sequences.

Mismatch Repair System in *E. coli*

- The time of methylation by deoxyadenosine methylase, however, lags behind that of nucleotide addition at the replication fork about two minutes, so the newly synthesized strand is transiently unmethylated.

Mismatch Repair System in *E. coli*

- The *E. coli* mismatch repair system exploits this period of transient unmethylation to identify and cut GATC sites in a newly synthesized strand with a mismatch.

MutS, MutL, & MutH Proteins

- Genetic and biochemical studies have demonstrated that three *E. coli* proteins viz., MutS, MutL, and MutH are dedicated to mismatch repair.

MutS, MutL, & MutH Proteins

- Although these proteins are essential for mismatch repair, they are not sufficient.
- Several additional enzymes and protein factors also make important contributions.

MutS, MutL, & MutH Proteins

- Among these enzymes and protein factors are:-
- i. DNA helicase II (UvrD),
- ii. Single-stranded DNA binding protein (SSB)
- iii. 5' \longrightarrow 3' exonucleases (ExoVII or RecJ)

MutS, MutL, & MutH Proteins

- iv. 3'→5' exonucleases (ExoI, ExoVII, or ExoX),
- v. DNA polymerase III holoenzyme,
- vi. DNA ligase, and
- vii. Deoxyadenosine methylase.

MutS, MutL, & MutH Proteins

- The process of repair begins when **MutS** (a homodimer or homotetramer) binds to the mismatch.
- MutS recruits MutL (a homodimer) in an ATP-dependant fashion.

MutS, MutL, & MutH Proteins

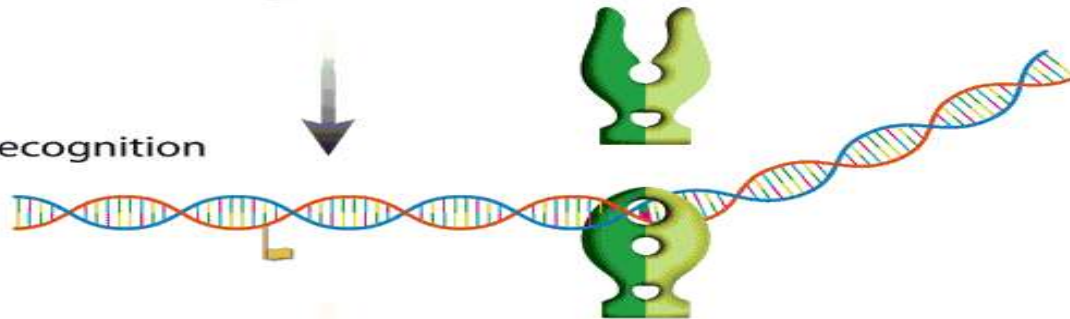
- Then the MutS•MutL complex activates MutH, which makes an incision at the nearest unmethylated GATC site, either 5' or 3' to the mismatch, in the newly synthesized strand.

MutS, MutL, & MutH Proteins

DNA with replication error



Mismatch recognition



Protein complex assembly



Daughter strand nicking



MutS, MutL, & MutH Proteins

- MutH shares sequence homology with the type II restriction endonuclease, Sau3AI.
- Both enzymes recognize and cleave GATC sequences.

MutS, MutL, & MutH Proteins

- MutH does not bind or cleave fully methylated GATC sites, whereas Sau3AI cleaves fully, hemi and unmethylated GATC sites.

MutS, MutL, & MutH Proteins

- Next in the process is that helicase (UvrD) unwinds the DNA and SSB binds to the resulting single strands.

MutS, MutL, & MutH Proteins

- When the incision is 5' to the mismatch, ExoVII or RecJ exonucleases hydrolyze the nicked strand in a 5'→3' direction.

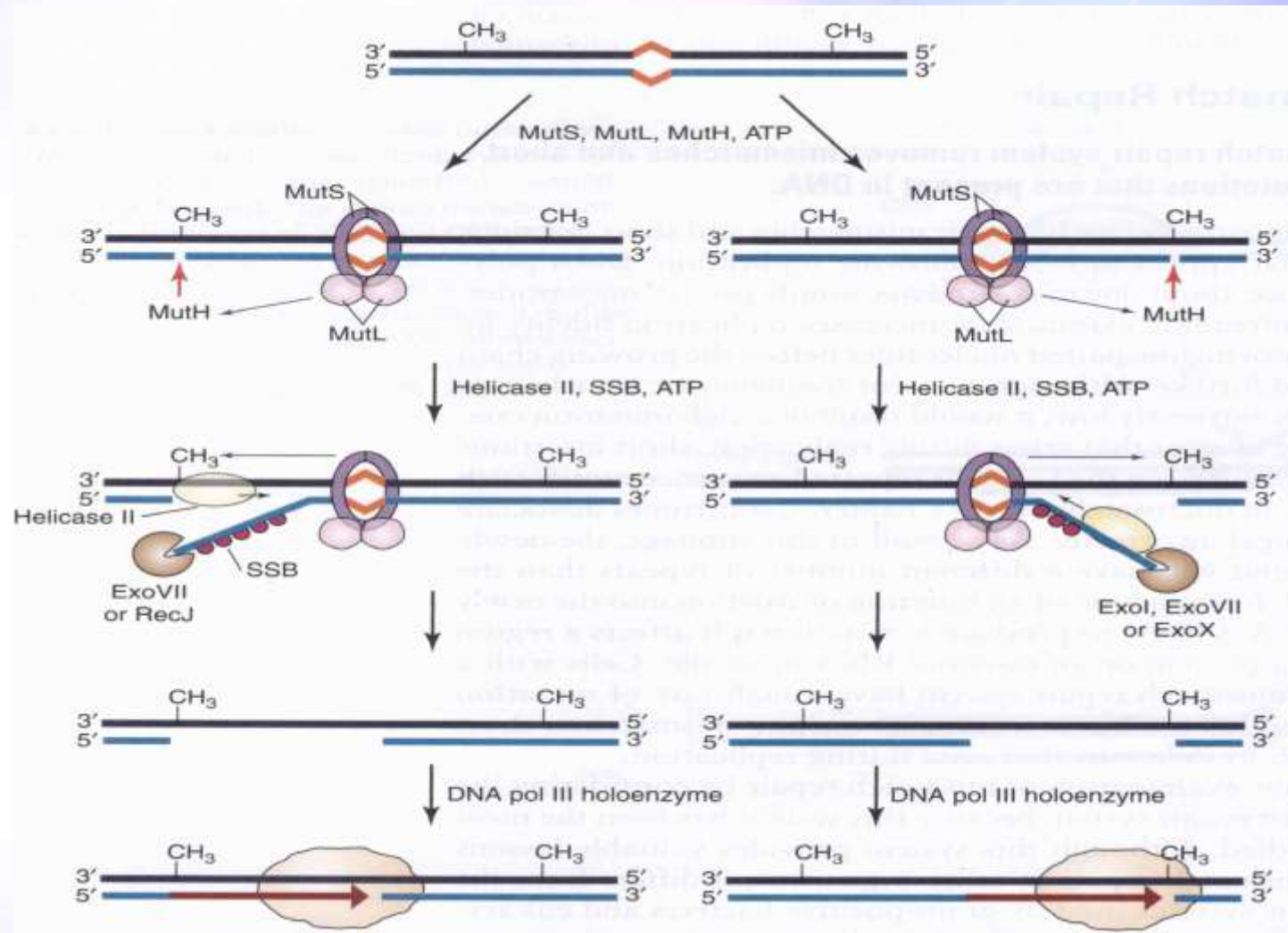
MutS, MutL, & MutH Proteins

- When the incision is 3' to the mismatch, ExoI, ExoVII or ExoX exonucleases hydrolyze the nicked strand in a 3'→5' direction.

MutS, MutL, & MutH Proteins

- DNA polymerase III holoenzyme fills the gap with new DNA.
- DNA ligase seals the remaining nick and deoxyadenosine methylase adds a methyl group to the GATC site.

MutS, MutL, & MutH Proteins



MutS, MutL, & MutH Proteins

- All organisms that have a mismatch repair system have MutS and MutL homologs.
- However, MutH is present only in *E. coli* and some other gram-negative bacteria.

Mismatch Repair in Eukaryotes

- Eukaryotes have proteins that are homologous to MutS and MutL but lack homologs to MutH.

Mismatch Repair in Eukaryotes

- There are three human MutS homologs designated as MSH2, MSH3, and MSH6 which participate in mismatch repair.

Mismatch Repair in Eukaryotes

- MSH2 and MSH6 combine to form a heterodimer called **MutS α** , and MSH2 and MSH3 combine to form a second heterodimer called **MutS β** .

Mismatch Repair in Eukaryotes

- The structures of MutS α and MutS β are thought to be similar to the MutS homodimer in bacteria.

Mismatch Repair in Eukaryotes

- **MutS α** initiates mismatch repair at single mismatches and small insertions/deletion loops, whereas **MutS β** only initiates mismatch repair at insertion/deletion loops of various sizes.

Mismatch Repair in Eukaryotes

- Mammalian homologs of the bacterial MutL protein that participate in mismatch repair are designated as MLH1 and PMS2 and the heterodimer containing these two subunits is called MutL α .

Mismatch Repair in Eukaryotes

- Paul Modrich and coworkers have reconstituted the human mismatch repair system *in vitro*.
- A strand break on either side of the mismatch is sufficient to direct repair.

Human Mismatch Repair System

- Paul Modrich and coworkers have reconstituted the human mismatch repair system *in vitro*.
- A strand break on either side of the mismatch is sufficient to direct repair.

Human Mismatch Repair System

- Purified human MutS α , MutL α , ExoI (a 5' \rightarrow 3' exonuclease), RPA, PCNA, RFC, and DNA polymerase δ are required for bidirectional repair.

Human Mismatch Repair System

- MutS α , ExoI, and RPA are adequate to excise a mismatch when the nick is on the 5' side of the mismatch but MutL α , PCNA, and RFC are also required when the nick is on the 3' side of the mismatch.

Human Mismatch Repair System

- The observation that the mismatch repair system can degrade newly synthesized strands with a nick on the 3'-side of the mismatch was very puzzling because ExoI degrades DNA in a 5' → 3' direction.

Human Mismatch Repair System

- Modrich and co-workers solved the puzzle by demonstrating that MutL α is a latent endonuclease that is activated in a mismatch.

Human Mismatch Repair System

- Once activated, MutL α preferentially makes incisions in the strand that already has a nick, that is, the discontinuous strand during replication.

Human Mismatch Repair System

- The endonuclease activity appears to require an amino acid motif present in the PMS2 but not the MLH1 subunit.

Human Mismatch Repair System

- In the human mismatch repair pathway, MutS α , PCNA, and RFC cooperate to activate the latent MutL α endonuclease.

Human Mismatch Repair System

- This MutL α endonuclease then nicks the discontinuous strand of a DNA duplex on both the 5' and 3' sides of the mismatch.

Human Mismatch Repair System

- When the original nick is on the 3' side of the mismatch, MutL α incisions produce a new 5' terminus on the far side of the mismatch that can serve as an entry site for MutS α -activated ExoI.

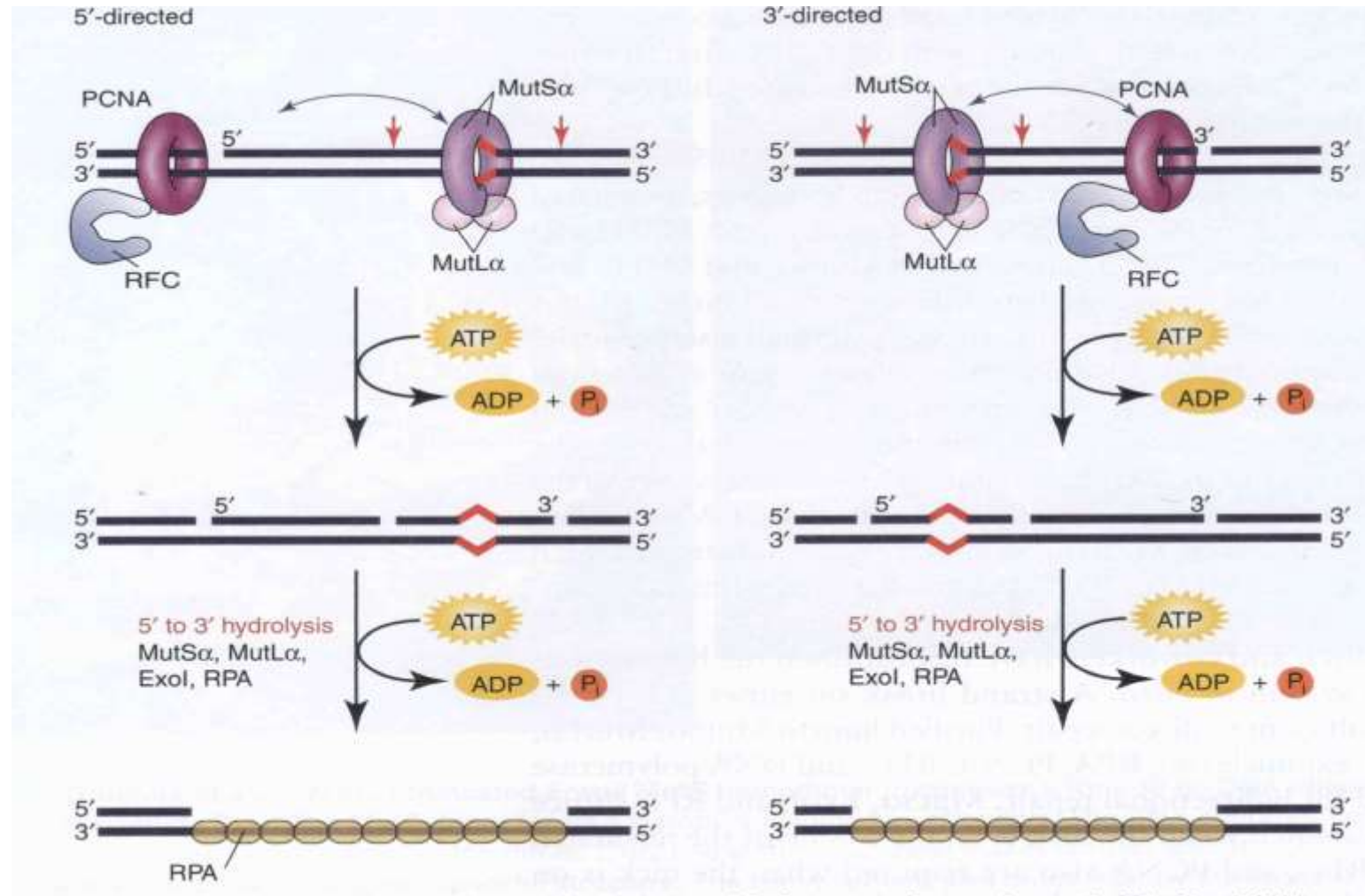
Human Mismatch Repair System

- This activated ExoI then removes the mismatch using its 5' → 3' exonuclease activity.
- RPA stimulates ExoI activity in the presence of MutS α as long as the mismatch is present.

Human Mismatch Repair System

- Once the mismatch has been removed, RPA inhibits the exonuclease, probably by displacing ExoI from the DNA.

Human Mismatch Repair System



Human Mismatch Repair System

- Individuals with a non-functional mismatch repair system due to defective MutS α or MutL α suffer from nonpolyposis colon cancer (HNPCC), an autosomal recessive syndrom.

end

The Central Dogma

- The central dogma outlines the flow of genetic information during growth and division of the cells.
- Genetic information flows from DNA to RNA to protein during cell growth.

The Central Dogma

- In addition, all living cells must replicate their DNA when they divide.
- During cell division each daughter cell receives a copy of the genome of the parent cell.

The Central Dogma

- Replication is the process by which two identical copies of DNA are made from an original molecule of DNA.
- So Replication occurs in the cells prior to cell division.

The Central Dogma

- An important point is that information does not flow from protein to RNA or DNA.
- However, flow of information from RNA “backwards” to DNA is possible in certain special circumstances due to the operation of reverse transcriptase.

The Central Dogma

- By the end of 1953, the working hypothesis was adopted that chromosomal DNA functions as the template for the synthesis of RNA molecules.

The Central Dogma

- These RNA molecules, the subsequently move to the cytoplasm, where they determine the arrangement of amino acids within proteins.

The Central Dogma

- In 1956, Francis Crick referred to this pathway for the flow of genetic information as the **Central Dogma**.

The Central Dogma



The Central Dogma

- An important point in the above equation is that the two arrows are unidirectional which means that RNA sequences are never determined by protein templates nor was DNA then imagined ever to be made on RNA templates.

The Central Dogma

- The idea that proteins never serve as templates for RNA has stood the test of time.
- However, RNA chains sometimes do act as templates for the synthesis of DNA chains of complementary sequence.

The Central Dogma

- Such reversals of the normal flow of information are very rare events compared with the enormous number of RNA molecules made on DNA templates.

The Central Dogma

- Thus, the central dogma as originally proclaimed more than 50 years ago still remains essentially valid.

end

The Adaptor Hypothesis of Crick

- At first it seemed simplest to believe that the RNA templates for protein synthesis were folded up to create cavities on their outer surfaces specific for the 20 different amino acids.

The Adaptor Hypothesis of Crick

- The cavities would be so shaped that only one given amino acid would fit, and in this way RNA would provide the information to order amino acids during protein synthesis.

The Adaptor Hypothesis of Crick

- By 1955, however, Crick became disenchanted with this conventional wisdom, arguing that it would never work.

The Adaptor Hypothesis of Crick

- In the first place, the specific chemical groups on the four bases of RNA (A, U, G, and C) should mostly interact with water-soluble groups.

The Adaptor Hypothesis of Crick

- Yet, the specific side groups of many amino acids (e.g., leucine, valine, and phenylalanine) strongly prefer interactions with water-insoluble (hydrophobic) groups.

The Adaptor Hypothesis of Crick

- In the second place, even if somehow RNA could be folded so as to display some hydrophobic surfaces, it seemed unlikely that

The Adaptor Hypothesis of Crick

- an RNA template would be used to discriminate accurately between chemically very similar amino acids like glycine and alanine or valine and isoleucine, both pairs differing only by the presence of single methyl groups.

The Adaptor Hypothesis of Crick

- Crick thus proposed that prior to incorporation into proteins, amino acids are first attached to specific adaptor molecules, which in turn possess unique surfaces that can bind specifically to bases on the RNA templates.

end

Discovery of Transfer RNA

- The discovery of how proteins are synthesized required the development of cell-free extracts capable of making proteins from amino acid precursors as directed by added RNA molecules.

Discovery of Transfer RNA

- These were first effectively developed in 1953 by Paul C. Zamecnik and his collaborators. Key to their success were the recently available radioactively tagged amino acids.

Discovery of Transfer RNA

- They used these labelled amino acids to mark the trace amounts of newly synthesized proteins.
- Early on, the cellular site of protein synthesis was pinpointed to be the ribosomes.

Discovery of Transfer RNA

- Several years later, Zamecnik, with his collaborator Mahlon B. Hoagland, made an important discovery that prior to their incorporation into proteins, amino acids are first attached to what we now call transfer RNA (tRNA) molecules.

Discovery of Transfer RNA

- Transfer RNA accounts for some 10% of all cellular RNA.
- To nearly everyone except Crick, this discovery was totally unexpected.

Discovery of Transfer RNA

- He had, of course, previously speculated that his proposed “adaptors” might be short RNA chains,

Discovery of Transfer RNA

- because their bases would be able to base-pair and “read” the appropriate groups on the RNA molecules that served as the templates for protein synthesis.

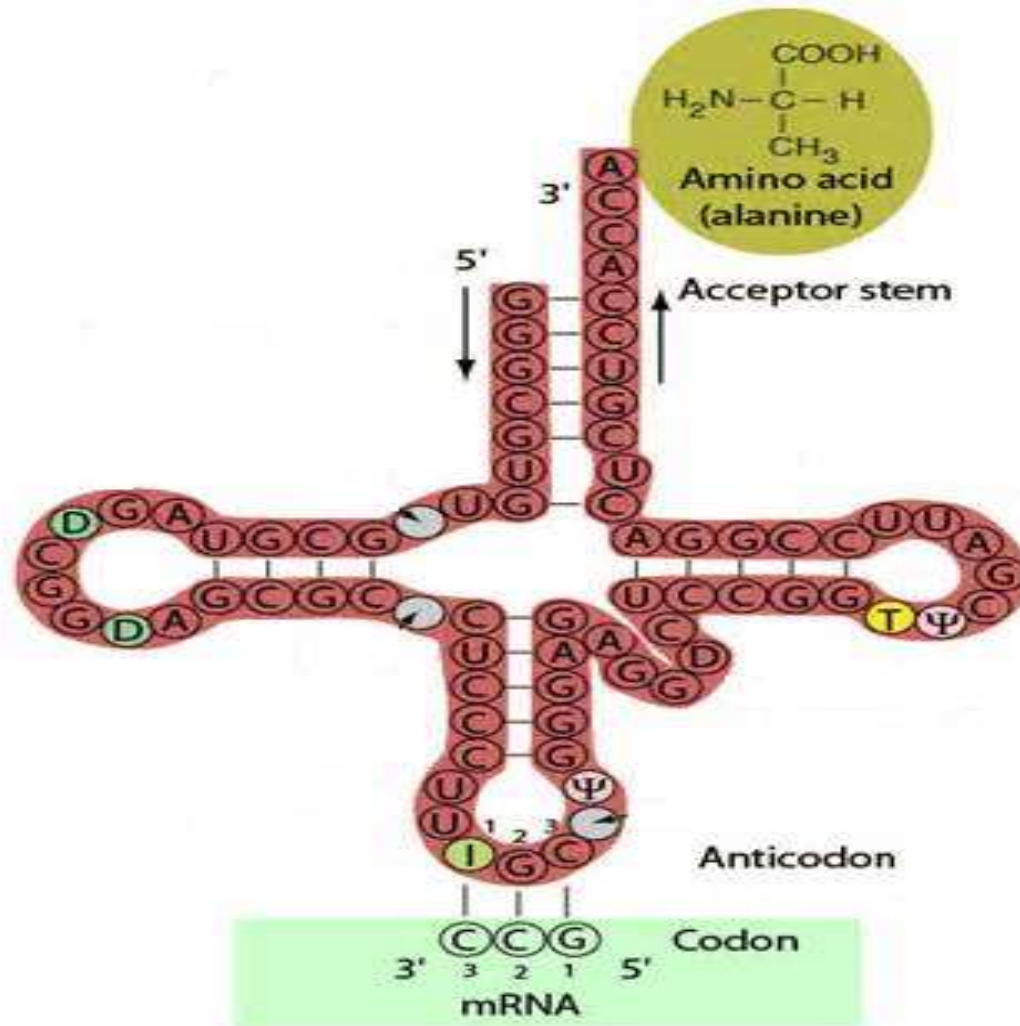
Discovery of Transfer RNA

- The transfer RNA molecules of Zamecnik and Hoagland are in fact the adaptor molecules postulated by Crick.

Discovery of Transfer RNA

- Each transfer RNA contains a sequence of adjacent bases (the anticodon) that bind specifically during protein synthesis to successive groups of bases (codons) along the RNA template.

Discovery of Transfer RNA



Discovery of Messenger RNA

- Cells infected with phage T4 provided the ideal system to find the true template for protein synthesis.
- Following infection by this virus, cells stop synthesizing *E. coli* RNA; the only RNA synthesized is transcribed off the T4 DNA.

Discovery of Messenger RNA

- Most strikingly, not only does T4 RNA have a base composition very similar to T4 DNA, but it does not bind to the ribosomal proteins that normally associate with rRNA to form ribosomes.

Discovery of Messenger RNA

- Instead, after first attaching to previously existing ribosomes, T4 RNA moves across their surface to bring its bases into positions where they can bind to the appropriate tRNA—amino acid precursors for protein synthesis.

Discovery of Messenger RNA

- In so acting, T4 RNA orders the amino acids and is thus the long-sought-for RNA template for protein synthesis.

Discovery of Messenger RNA

- Because it carries the information from DNA to the ribosomal sites of protein synthesis, it is called messenger RNA (mRNA).

Discovery of Messenger RNA

- The observation of T4 RNA binding to *E. coli* ribosomes, first made in 1960, was soon followed with evidence for a separate messenger class of RNA within uninfected *E. Coli* cells.

Discovery of Messenger RNA

- Instead, the rRNA components of ribosomes, together with some 50 different ribosomal proteins that bind to them, serve as the factories for protein synthesis.

Discovery of Messenger RNA

- They function to bring the tRNA–amino acid precursors into positions where they can read off the information provided by the mRNA templates.

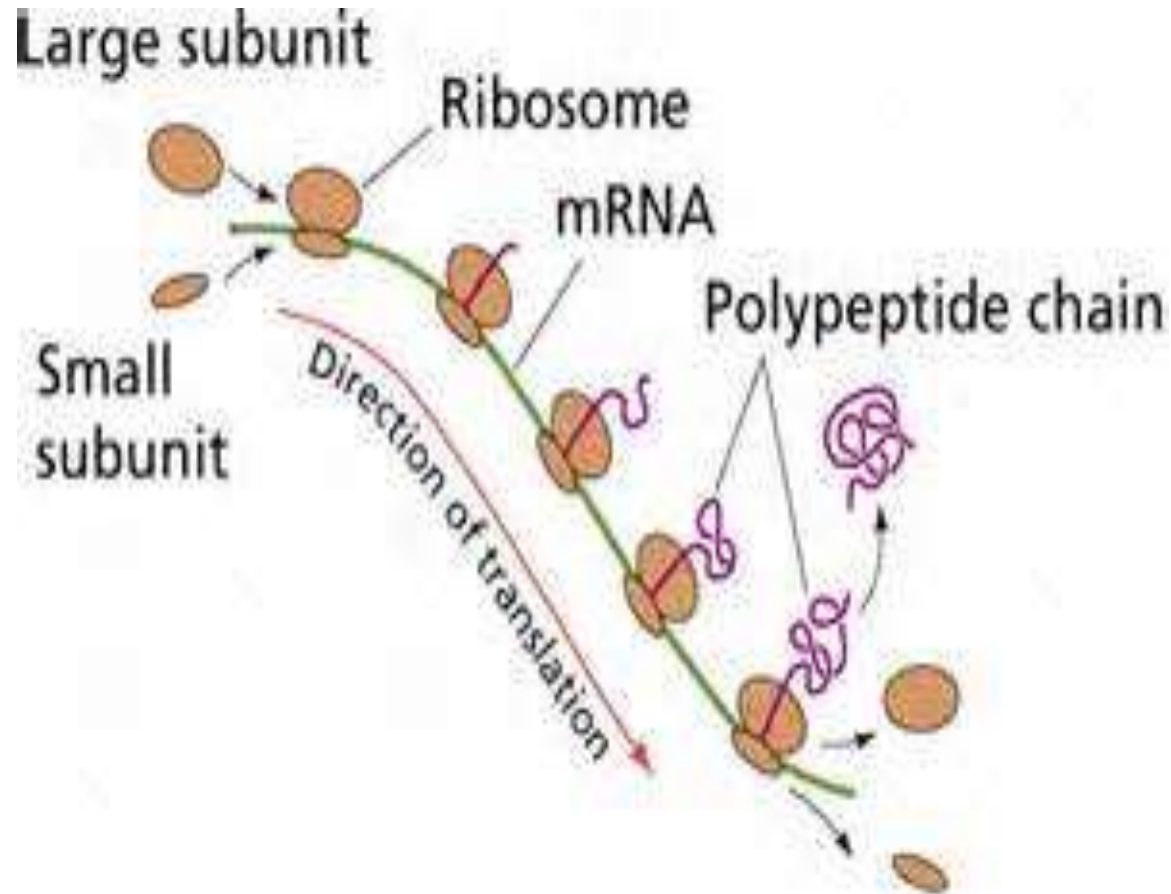
Discovery of Messenger RNA

- Only a few percent of total cellular RNA is mRNA.
- This RNA shows the expected large variations in length and nucleotide composition required to encode the many different proteins found in a given cell.

Discovery of Messenger RNA

- Hence, it is easy to understand why mRNA was first overlooked. Because only a small segment of mRNA is attached at a given moment to a ribosome, a single mRNA molecule can simultaneously be read by several ribosomes.

Discovery of Messenger RNA



Synthesis of RNA upon DNA Templates

- As mRNA was being discovered, the first of the enzymes that synthesize (or transcribe) RNA using DNA templates was being independently isolated in the labs of biochemists Jerard Hurwitz and Samuel B. Weiss.

Synthesis of RNA upon DNA Templates

- These enzymes called **RNA Polymerases** function only in the presence of DNA, which serves as the template upon which single-stranded RNA chains are made, and use the nucleotides ATP, GTP, CTP, and UTP as precursors.

Synthesis of RNA upon DNA Templates

- These enzymes make RNA using appropriate segments of chromosomal DNA as their templates.

Synthesis of RNA upon DNA Templates

- Direct evidence that DNA lines up the correct ribonucleotide precursors came from seeing how the RNA base composition varied with the addition of DNA molecules of different AT/GC ratios.

Synthesis of RNA upon DNA Templates

- In every enzymatic synthesis, the RNA AU/GC ratio was roughly similar to the DNA's AT/GC ratio.
- During transcription, only one of the two strands of DNA is used as a template to make RNA.

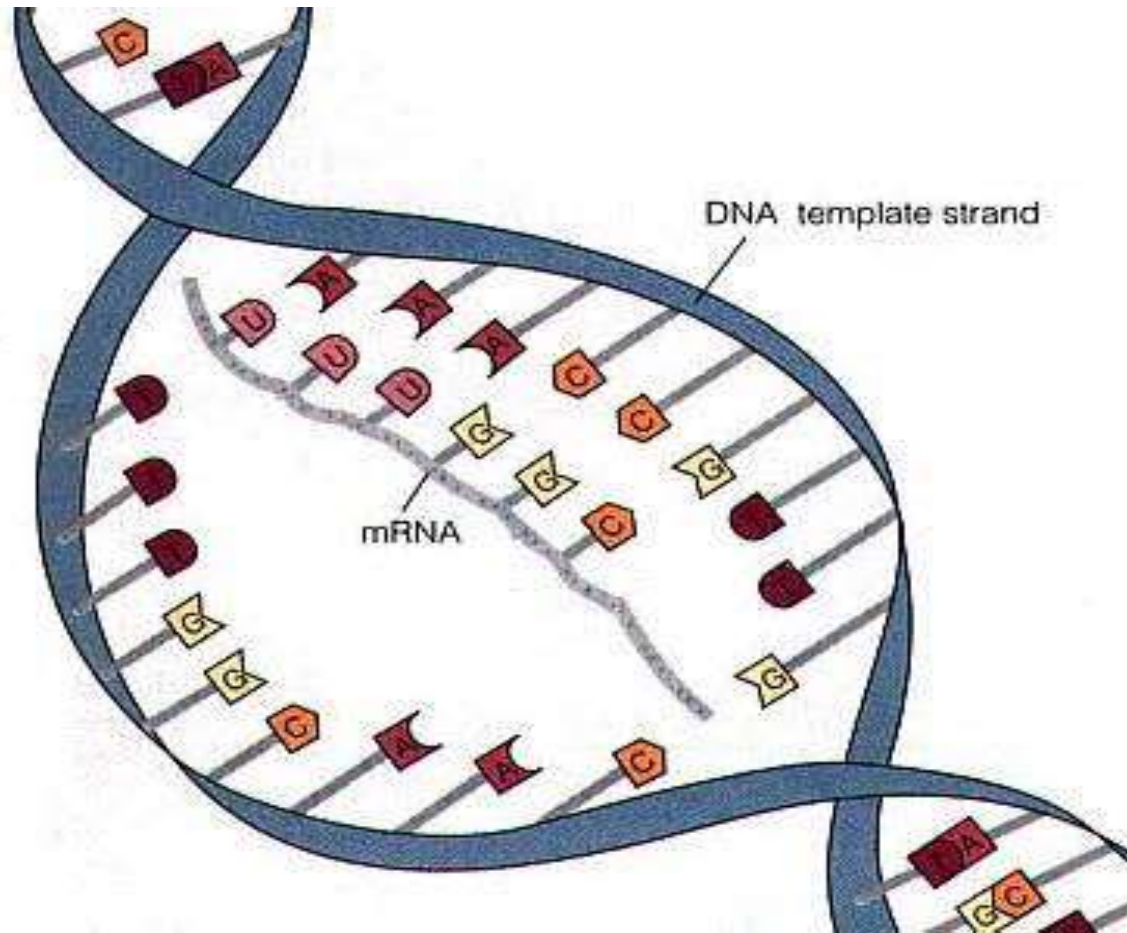
Synthesis of RNA upon DNA Templates

- This makes sense, because the messages carried by the two DNA strands, being complementary but not identical, are expected to code for completely different polypeptides.

Synthesis of RNA upon DNA Templates

- The synthesis of RNA always proceeds in a fixed direction, beginning at the 5' end and concluding with the 3'-end nucleotide.

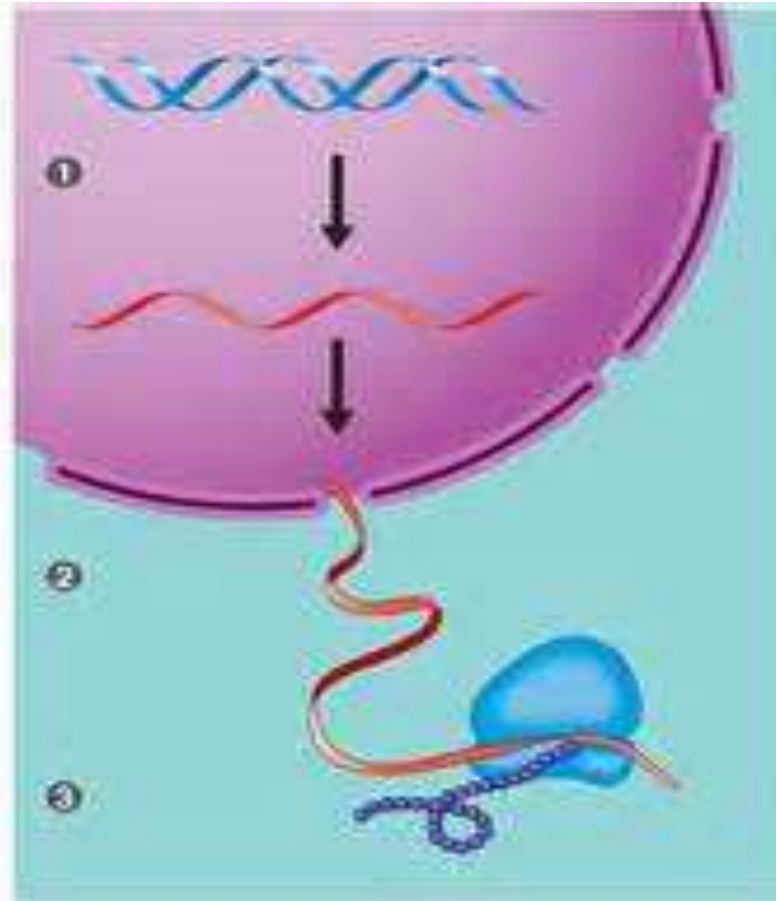
Synthesis of RNA upon DNA Templates



Synthesis of RNA upon DNA Templates

- By this time, there was firm evidence for the postulated movement of RNA from the DNA-containing nucleus to the ribosome-containing cytoplasm of eukaryotic cells.

Synthesis of RNA upon DNA Templates



© 2000 Pearson Education, Inc.

Synthesis of RNA upon DNA Templates

- By briefly exposing cells to radioactively labelled precursors, then adding a large excess of unlabeled ribonucleotides (a “pulse chase” experiment), mRNA synthesized during a short time window was labelled.

Synthesis of RNA upon DNA Templates

- These studies showed that mRNA is synthesized in the nucleus.
- Within an hour, most of this RNA had left the nucleus and was observed in the cytoplasm.

end

Establishing the Genetic Code

- Given the existence of 20 amino acids but only four bases, groups of several nucleotides must somehow specify a given amino acid.
- Groups of two nucleotides, however, would specify only 16 (4×4) amino acids.

Establishing the Genetic Code

- So from 1954, most attention was given to how triplets (groups of three) might work, even though they obviously would provide more permutations ($4 \times 4 \times 4$) than needed if each amino acid was specified by only a single triplet.

Establishing the Genetic Code

- It was assumed that successive groups of nucleotides along a DNA chain code for successive amino acids along a given polypeptide chain.

Establishing the Genetic Code

- An elegant mutational analysis on bacterial proteins, carried out in the early 1960s by Charles Yanofsky and Sydney Brenner, showed that co-linearity does in fact exist.

Establishing the Genetic Code

- Equally important were the genetic analyses by Brenner and Crick, which in 1961 first established that groups of three nucleotides are used to specify individual amino acids.

Establishing the Genetic Code

- But which specific groups of three bases (codons) determine which specific amino acids could only be learned by biochemical analysis.

Establishing the Genetic Code

- The major breakthrough came when Marshall Nirenberg and Heinrich Matthaei, then working together, observed in 1961 that the addition of the synthetic polynucleotide poly U (UUUUU . . .)

Establishing the Genetic Code

- to a cell-free system capable of making proteins leads to the synthesis of polypeptide chains containing only the amino acid phenylalanine.
- The nucleotide groups UUU thus must specify phenylalanine.

Establishing the Genetic Code

- Use of increasingly more complex polynucleotides as synthetic messenger RNAs rapidly led to the identification of more and more codons.

Establishing the Genetic Code

- Particularly important in completing the code was the use of polynucleotides like AGUAGU, put together by organic chemist Har Gobind Khorana.
- In this way, more specific sets of codons were tested.

Establishing the Genetic Code

- Completion of the code in 1966 revealed that 61 out of the 64 possible permuted groups corresponded to amino acids, with most amino acids being encoded by more than one nucleotide triplet.

Establishing the Genetic Code

- Completion of the code in 1966 revealed that 61 out of the 64 possible permuted groups corresponded to amino acids, with most amino acids being encoded by more than one nucleotide triplet.

Establishing the Genetic Code

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

TRANSCRIPTION

- Up to this point, we have learnt that how the genome has been maintained i.e., how the genetic material is organized, protected, and replicated.

TRANSCRIPTION

- We now turn to the question of how that genetic material is **expressed**—that is, how the series of bases in the DNA directs the production of the RNAs and proteins that perform cellular functions and define cellular identity.

TRANSCRIPTION

- The basic processes responsible for gene expression include transcription, RNA processing, and translation.
- Transcription is chemically and enzymatically, very similar to DNA replication.

TRANSCRIPTION

- Both involve enzymes that synthesize a new strand of nucleic acid complementary to a DNA template strand.

TRANSCRIPTION

- However, the main difference is that in case of transcription, the new strand is made from ribonucleotides rather than deoxyribonucleotides.
- There are some other important differences include:-

TRANSCRIPTION

- 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.

TRANSCRIPTION

- Transcription is less accurate than replication (one mistake/10,000 nucleotides, compared with one / 10 million). This difference reflects the lack of extensive proofreading mechanisms for transcription.

TRANSCRIPTION

- Transcription selectively copies only certain parts of the genome and makes any number of copies of that section. In contrast, replication must copy the entire genome and do so once (and only once) every cell division.

end

RNA POLYMERASES

- RNA polymerase performs essentially the same reaction in all cells, from bacteria to humans that is the synthesis of RNA.
- From bacteria to mammals, the cellular RNA polymerases are made up of multiple subunits.

RNA POLYMERASES

- Bacteria have only a single RNA polymerase, whereas eukaryotic cells have three: RNA polymerases I, II, and III (RNA Pol I, II, and III).

RNA POLYMERASES

- Recently, two more DNA-dependent RNA polymerases have been identified in recent years, and have been called as **Pol IV** and **Pol V**.
- These are found only in plants, where they transcribe **small interfering RNAs**.

RNA POLYMERASES

- The bacterial RNA polymerase **core enzyme** alone is capable of synthesizing RNA and comprises two copies of the α subunit and one each of the β , β' and ω subunits. This enzyme is closely related to the eukaryotic polymerases.

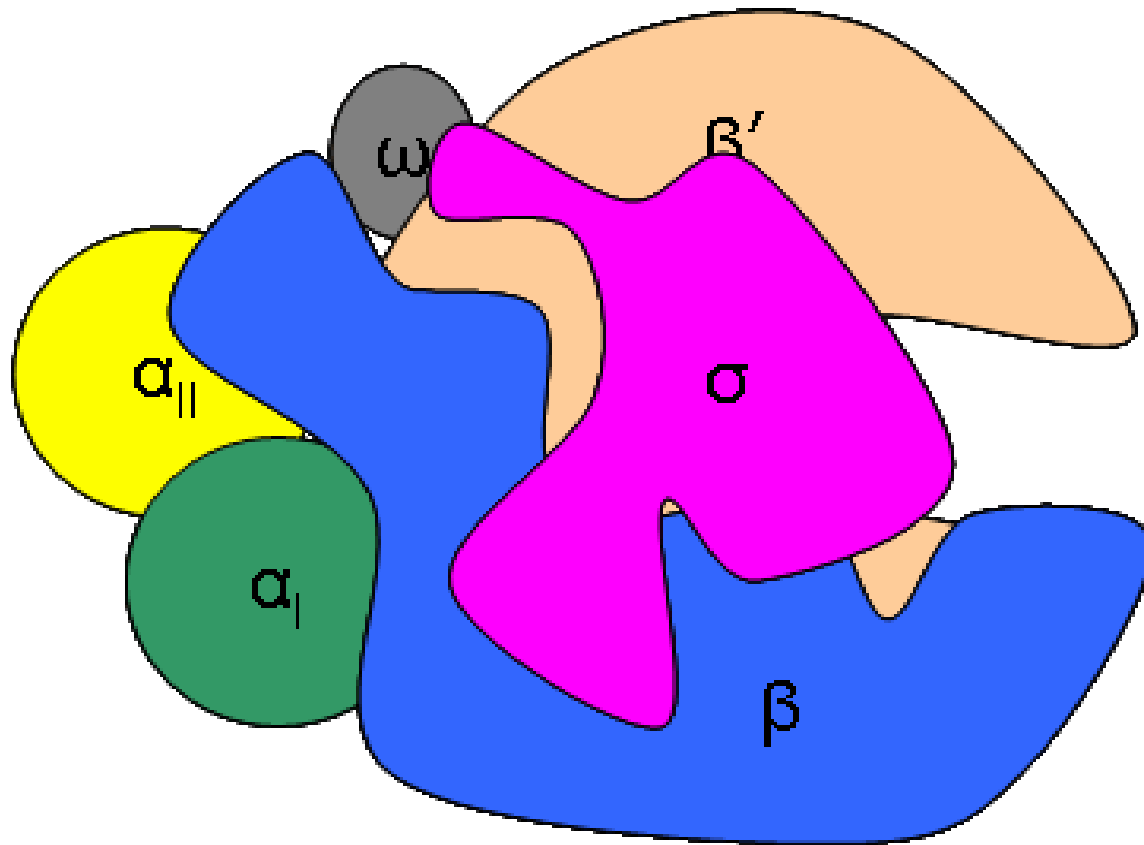
RNA POLYMERASES

- The structure of a bacterial RNA polymerase core enzyme is similar to that of the yeast Pol II enzyme.
- Overall, the shape of each enzyme resembles a crab claw.

RNA POLYMERASES

- This is reminiscent of the “hand” structure of DNA polymerases.
- The two pincers of the crab claw are made up predominantly of the two largest subunits of each enzyme (β and β').

RNA POLYMERASES



RNA POLYMERASES

- The active site, which is made up of regions from both these subunits, is found at the base of the pincers within a region called the “active center cleft”.

END

Transcription Proceeds in Series of Steps

- To transcribe a gene, RNA polymerase proceeds through a series of well defined steps grouped into three phases:
 - **Initiation**
 - **Elongation**
 - **Termination**

Transcription Proceeds in Series of Steps

- **Initiation**
- A **promoter** is the DNA sequence that initially binds to the RNA polymerase (together with any initiation factors required).

Transcription Proceeds in Series of Steps

- Once formed, the promoter–polymerase complex undergoes structural changes required for initiation to proceed.
- Then the DNA around the point where transcription will start unwinds.

Transcription Proceeds in Series of Steps

- The base pairs are disrupted, producing a “transcription bubble” of single-stranded DNA.
- Again, like DNA replication, transcription always occurs in a 5'-to-3' direction: the new ribonucleotide is added to the 3' end of the growing chain.

Transcription Proceeds in Series of Steps

- However, unlike replication, only one of the DNA strands acts as a template on which the RNA strand is built.
- The initiation can itself be broken down into a series of defined steps.

Transcription Proceeds in Series of Steps

- The first step is the initial binding of polymerase to a promoter to form what is called a **closed complex**.
- In this form, the DNA remains double-stranded, and the enzyme is bound to one face of the helix.

Transcription Proceeds in Series of Steps

- In the second step of initiation, the closed complex undergoes a transition to the **open complex** in which the DNA strands separate over a distance of 13 bp around the start site to form the transcription bubble.

Transcription Proceeds in Series of Steps

- In the next stage of initiation, polymerase enters the phase of initial transcription followed by promoter escape.
- The opening up of the DNA frees the template strand.

Transcription Proceeds in Series of Steps

- The first two ribonucleotides are brought into the active site, aligned on the template strand, and joined together.
- In the same way, subsequent ribonucleotides are incorporated into the growing RNA chain.

Transcription Proceeds in Series of Steps

- Incorporation of the first 10 or so ribonucleotides is a rather inefficient process, and at that stage, the enzyme often releases short transcripts (each of less than 10 or so nucleotides) and then begins synthesis again.

Transcription Proceeds in Series of Steps

- In this phase, the polymerase–promoter complex is called the **initial transcribing complex**.
- Once an enzyme makes a transcript longer than 10 nucleotides, it is said to have **escaped** the promoter.

Transcription Proceeds in Series of Steps

- At this point, it has formed a stable ternary complex, containing enzyme, DNA, and RNA.
- This is the transition to the elongation phase.

END

Transcription Proceeds in Series of Steps

- **Elongation**
- Once the RNA polymerase has synthesized a short stretch of RNA (10 bases), it shifts into the elongation phase.

Transcription Proceeds in Series of Steps

- During elongation, the enzyme performs an impressive range of tasks in addition to the catalysis of RNA synthesis.
- 1) It unwinds the DNA in front and re-anneals it behind.

Transcription Proceeds in Series of Steps

- 2) It dissociates the growing RNA chain from the template as it moves along.
- 3) And it performs the proofreading functions.

Transcription Proceeds in Series of Steps

- Recall that during replication, in contrast, several different enzymes are required to catalyze a similar range of functions.

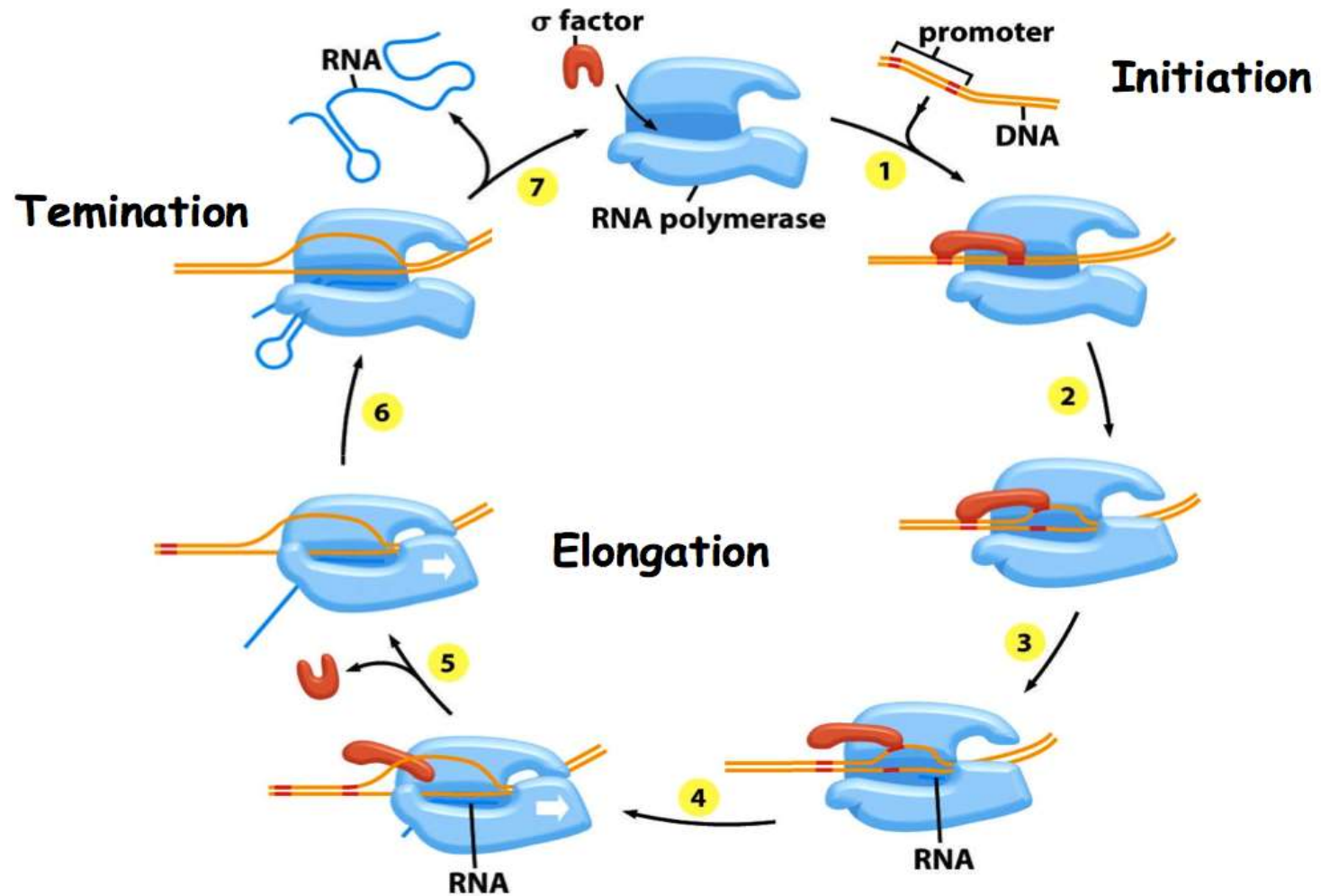
Transcription Proceeds in Series of Steps

- **Termination**
- Once the polymerase has transcribed the length of the gene (or genes), it must stop and release the RNA product (as well as dissociating from the DNA itself). This step is called termination.

Transcription Proceeds in Series of Steps

- In some cells, specific, well characterized sequences trigger termination. In others, it is less clear what instructs the enzyme to cease transcribing and dissociate from the template.

Transcription Proceeds in Series of Steps



Transcription Cycle in Bacteria

- The bacterial core RNA polymerase can, in principle, initiate transcription at any point on a DNA molecule.
- It, however, initiates transcription only at promoters.

Transcription Cycle in Bacteria

- It is the addition of an initiation factor called σ that converts core enzyme ($\alpha_2 \beta \beta' \omega$) into the form that initiates only at promoters. This form of the enzyme is called the RNA polymerase **holoenzyme**.

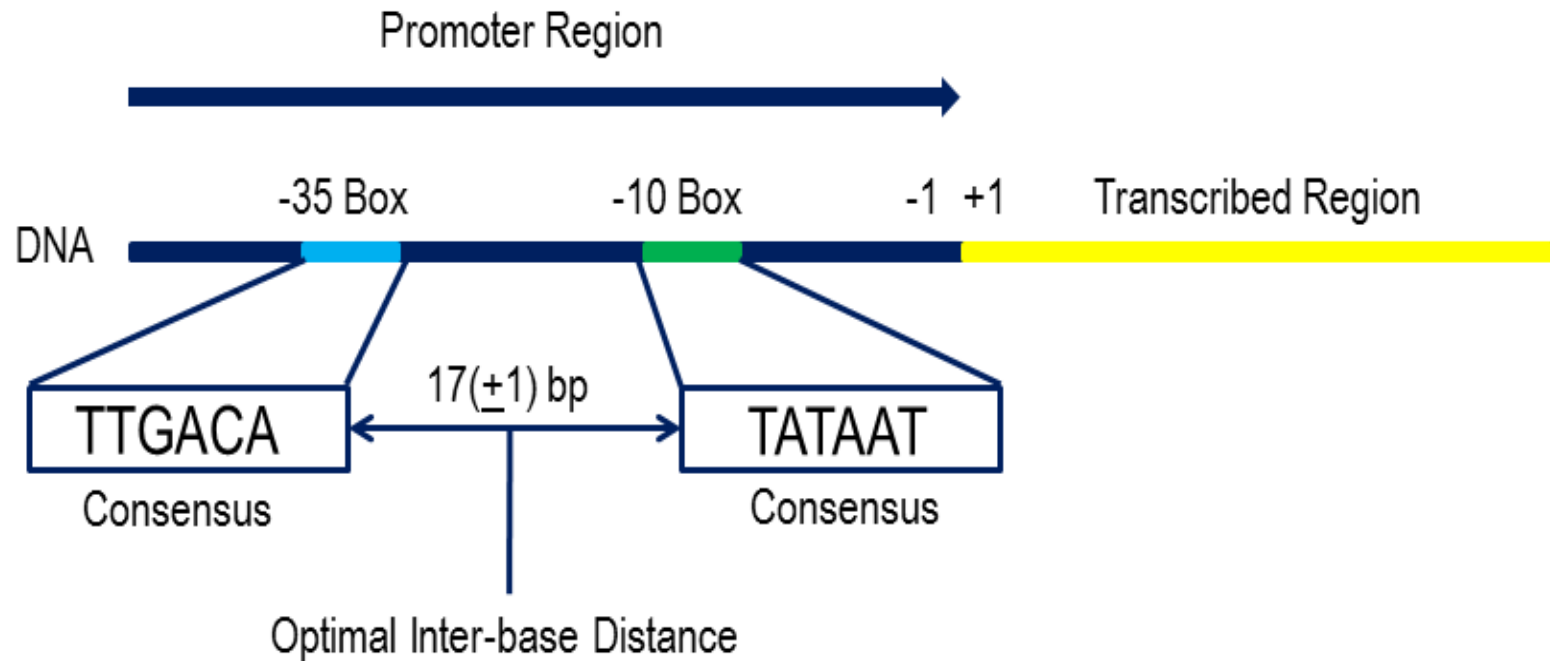
Transcription Cycle in Bacteria

- In the case of *Escherichia coli*, the predominant σ factor is called σ^{70} .
- Promoters recognized by polymerase have two conserved sequences each of 6 nucleotides, separated by a nonspecific stretch of 17–19 nucleotides.

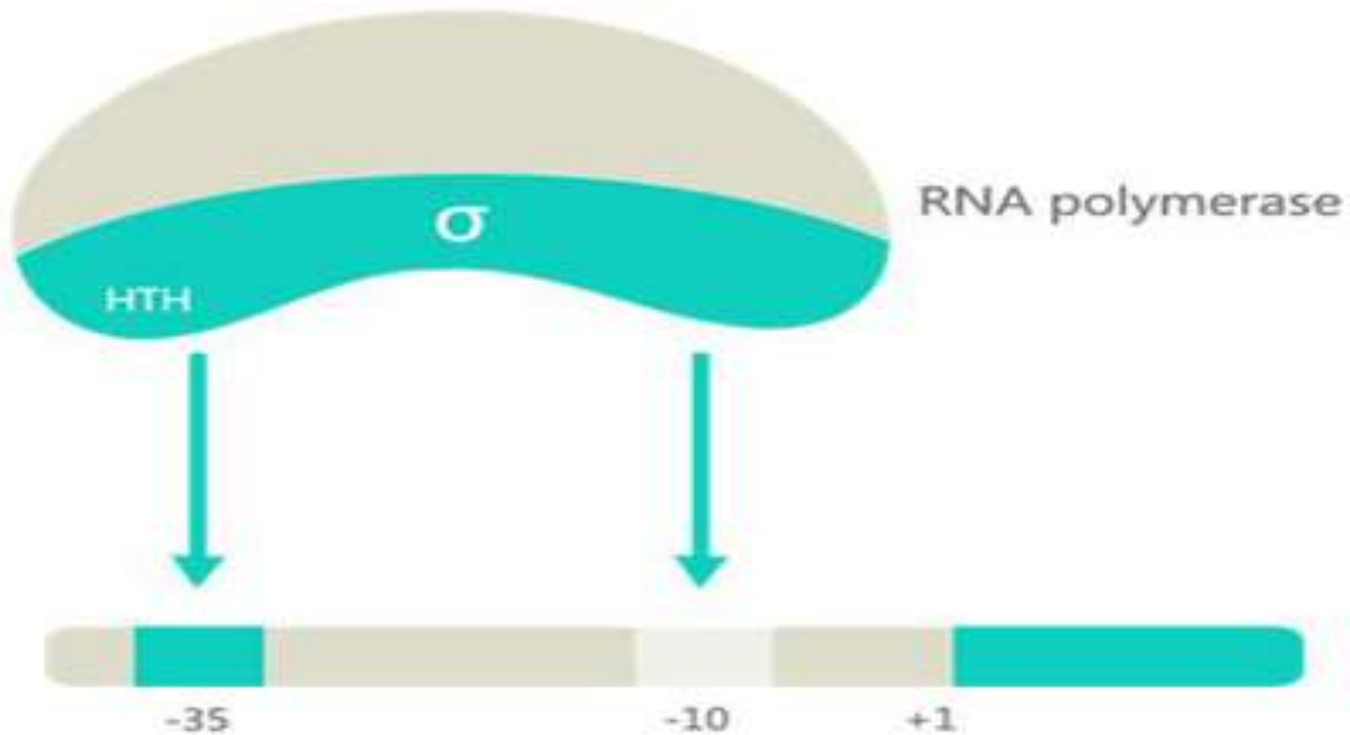
Transcription Cycle in Bacteria

- The two defined sequences are centered, respectively, at ~10 bp and at ~35 bp upstream of the site where RNA synthesis starts.
- The sequences are thus called the **minus 35** and **minus 10** regions or element.

Transcription Cycle in Bacteria



Transcription Cycle in Bacteria



Transcription Cycle in Bacteria

- 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 Cycle in Bacteria

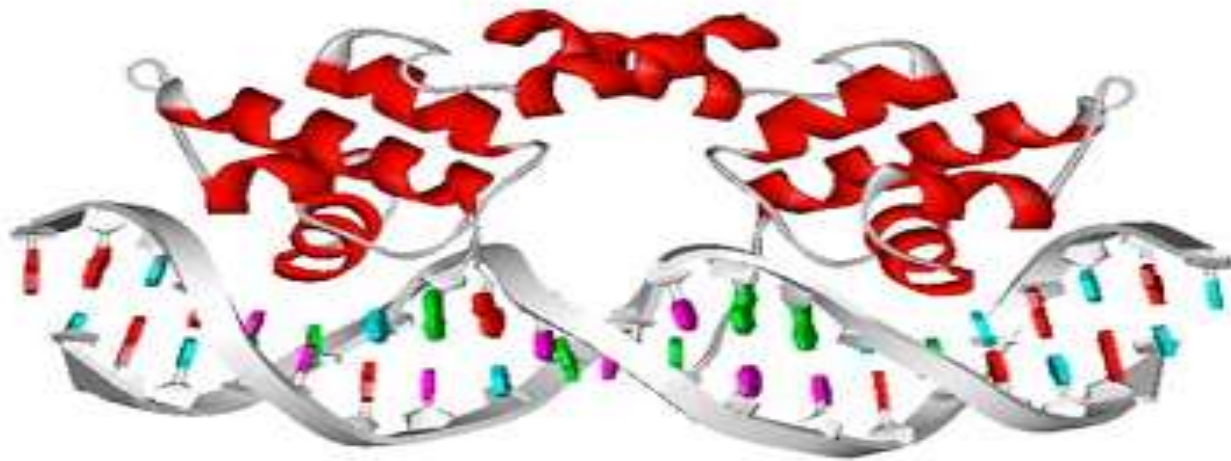
- Two helices within region 4 form a common DNA-binding motif called a **helix-turn-helix**.
- One of these helices inserts into the major groove and interacts with bases in the -35 region.

Transcription Cycle in Bacteria

- The other helix lies across the top of the groove, making contacts with the DNA backbone.

Transcription Cycle in Bacteria

Helix-turn-helix



Transcription Cycle in Bacteria

- The -10 region has a more elaborate role in transcription initiation, because it is within that element that DNA melting is initiated in the transition from the closed to the open complex.

Transcription Cycle in Bacteria

- Thus, the region of σ that interacts with the -10 region is doing more than simply binding DNA.

Transcription Cycle in Bacteria

- The α helix involved in recognition of the -10 region contains several essential aromatic amino acids that can interact with bases on the non-template strand in a manner that stabilizes the melted DNA.

Transcription Cycle in Bacteria

- Two bases in the non-template strand are flipped out and inserted into pockets within the σ protein where they make favorable contacts that stabilize the unwound state of the promoter region.

Transcription Cycle in Bacteria

- The σ subunit is positioned within the holoenzyme structure in such a way as to make feasible the recognition of various promoter elements.

Transcription Cycle in Bacteria

- Thus, the DNA-binding regions point away from the body of the enzyme, rather than being embedded.

END

Transition to the Open Complex

- Transition to the Open Complex involves structural changes in RNA Polymerase and in the Promoter DNA.
- The initial binding of RNA polymerase to the promoter DNA in the closed complex leaves the DNA in double-stranded form.

Transition to the Open Complex

- The next stage in initiation requires the enzyme to become more intimately engaged with the promoter, in the open complex.

Transition to the Open Complex

- The transition from the closed to the open complex involves structural changes in the enzyme and the opening of the DNA double helix to reveal the template and non-template strands.

Transition to the Open Complex

- This “melting” occurs between positions –11 and +2, with respect to the transcription start site.
- In case of the bacterial enzyme bearing $\sigma 70$, this transition is often called **isomerization**.

Transition to the Open Complex

- This transition does not require energy derived from ATP hydrolysis and is instead the result of a spontaneous conformational change in the DNA–enzyme complex to a more energetically favourable form.

Transition to the Open Complex

- Two bases in the non-template strand of the -10 element flip out from their base-stacking interactions and instead insert into pockets within the σ protein where they make more favourable interactions.

Transition to the Open Complex

- By stabilizing the single-stranded form of the -10 element, these interactions drive melting of the promoter region.

Transition to the Open Complex

- Isomerization is essentially irreversible and, once complete, typically guarantees that transcription will subsequently initiate.
- Formation of the closed complex, in contrast, is readily reversible.

Transition to the Open Complex

- The examination of the structure of the holoenzyme in more detail reveals the active site of the enzyme, which is made up of regions from both the β and β' subunits, is found at the base of the pincers within the active center cleft.

END

Transition to the Open Complex

- There are five channels into the enzyme, as shown in the illustration of the open complex.
- The NTP-uptake channel allows ribonucleotides to enter the active center.

Transition to the Open Complex

- The RNA-exit channel allows the growing RNA chain to leave the enzyme as it is synthesized during elongation.
- The remaining three channels allow DNA entry and exit from the enzyme.

Transition to the Open Complex

- The downstream DNA (i.e., DNA ahead of the enzyme, yet to be transcribed) enters the active center cleft in double-stranded form through the downstream DNA channel (between the pincers).

Transition to the Open Complex

- Within the active center cleft, the DNA strands separate from position +3.

Transition to the Open Complex

- The non-template strand exits the active center cleft through the non-template-strand (NT) channel and travels across the surface of the enzyme.

Transition to the Open Complex

- The template strand, in contrast, follows a path through the active center cleft and exits through the template-strand (T) channel.
- The double helix reforms at -11 in the upstream DNA behind the enzyme.

Transition to the Open Complex

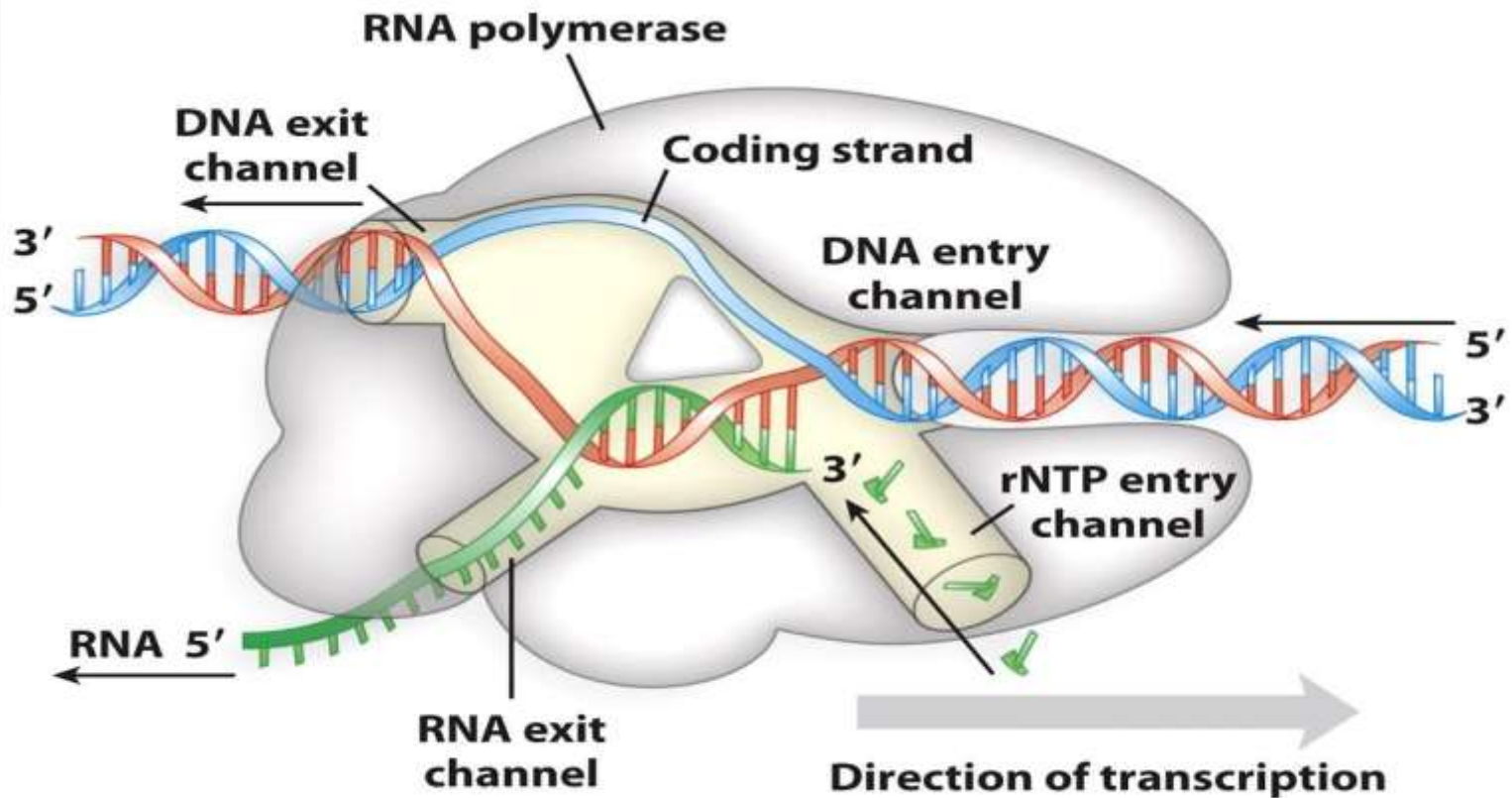


Figure 15-14
Molecular Biology: Principles and Practice
© 2012 W. H. Freeman and Company

Transition to the Open Complex

- Two striking structural changes are seen in the enzyme upon isomerization from the closed to the open complex.

Transition to the Open Complex

- First, the pincers at the front of the enzyme clamp down tightly on the downstream DNA.
- Second, there is a major shift in the position of the amino-terminal region of σ .

Transition to the Open Complex

- When not bound to DNA, σ region 1.1 lies within the active center cleft of the holoenzyme, blocking the path that, in the open complex, is followed by the template DNA strand.

Transition to the Open Complex

- In the open complex, region 1.1 shifts some 50Å° and is now found on the outside of the enzyme, allowing the DNA access to the cleft.

Transition to the Open Complex

- Region 1.1 of σ is highly negatively charged (just like DNA). Thus, in the holoenzyme, region 1.1 acts as a molecular mimic of DNA.

Transition to the Open Complex

- The space in the active center cleft, which may be occupied either by region 1.1 or by DNA, is highly positively charged.

END

Initial Transcription

- RNA polymerase can initiate a new RNA chain on a DNA template and thus does not need a primer.

Initial Transcription

- This requires that the DNA template be brought into the polymerase active site and held stably in a helical conformation.

Initial Transcription

- And that the initiating ribonucleotide be brought into the active site and held stably on the template while the next NTP is presented with correct geometry for the polymerization to take place.

Initial Transcription

- This is particularly difficult because RNA polymerase starts most transcripts with an A, and that ribonucleotide binds the template nucleotide (T) with only two hydrogen bonds.

Initial Transcription

- Thus, the enzyme has to make specific interactions with the DNA template strand, the initiating ribonucleotide, and the second ribonucleotide to hold them all rigidly in the correct orientation to allow chemical attack by the incoming NTP.

Initial Transcription

- This is the reason that most transcripts start with the same nucleotide.

Initial Transcription

- During initial transcription, RNA polymerase produces and releases short RNA transcripts of <10 nucleotides (abortive synthesis) before escaping the promoter, entering the elongation phase, and synthesizing the proper transcript.

Initial Transcription

- It has long been unclear how the enzyme's active site translocates along the DNA template during initial abortive cycles of transcription.
- Three general models were proposed to explain this.

Initial Transcription

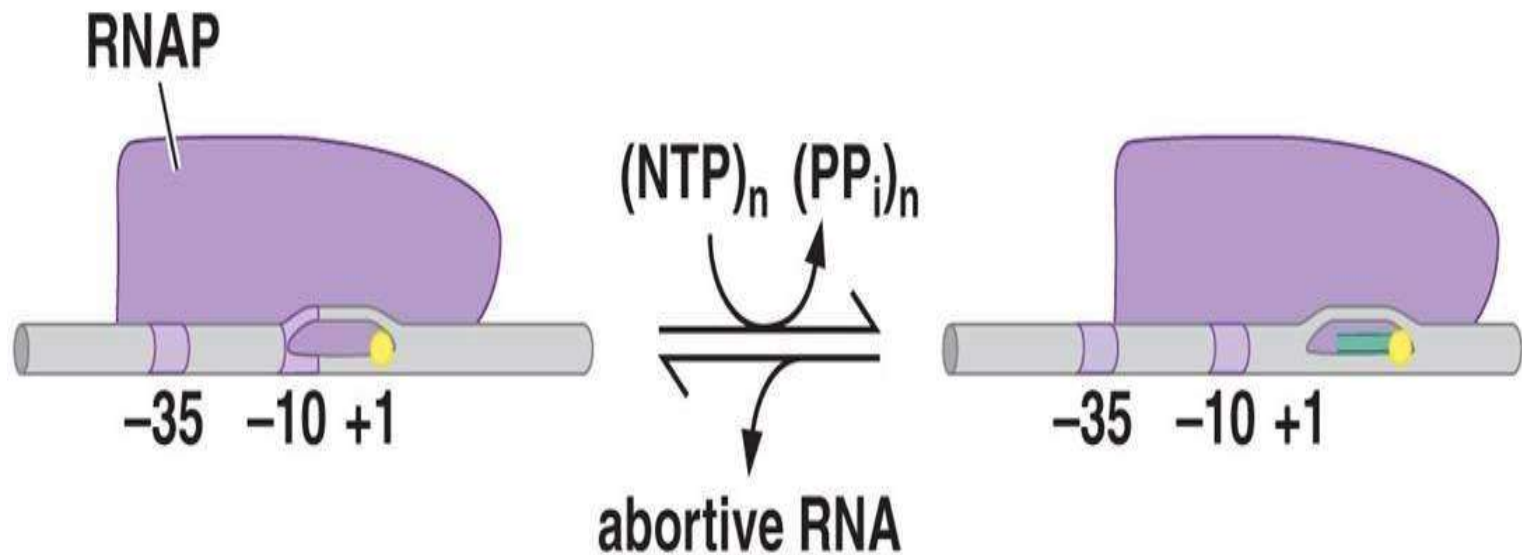
- 1. The “transient excursion” model proposes transient cycles of forward and reverse translocation of RNA polymerase.

Initial Transcription

- Thus, polymerase is thought to leave the promoter and translocate a short way along the DNA template, synthesizing a short transcript before aborting transcription, releasing the transcript, and returning to its original location on promoter.

Initial Transcription

“transient excursions”



Initial Transcription

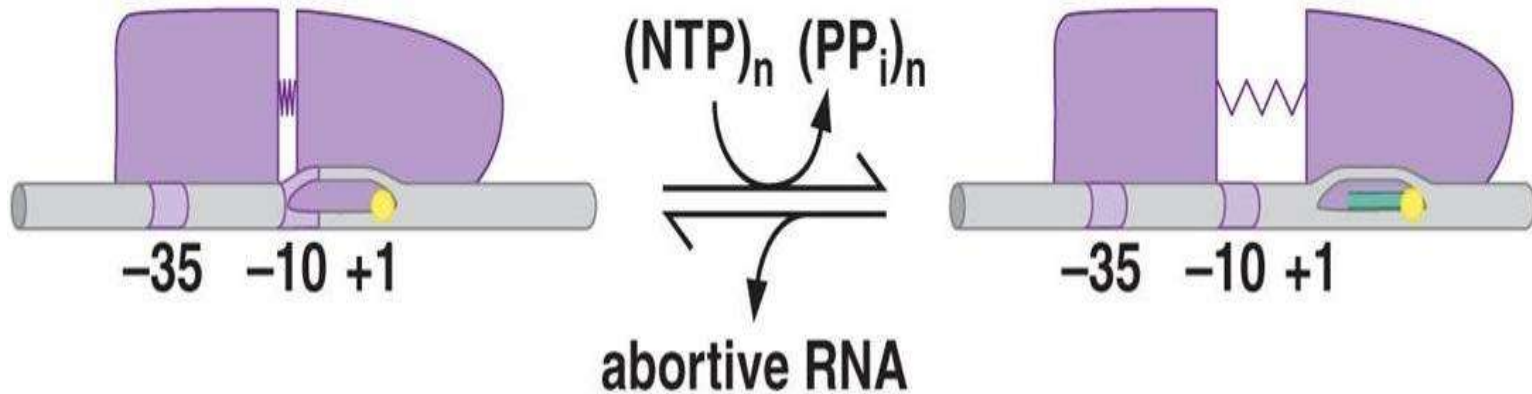
- 2. “Inchworming” invokes a flexible element within the polymerase that allows a module at the front of the enzyme, containing the active site.

Initial Transcription

- The module moves downstream, synthesizing a short transcript before aborting and retracting to the body of the enzyme still at the promoter.

Initial Transcription

“inchworming”



Initial Transcription

- 3. “Scrunching”
proposes that DNA downstream from the stationary, promoter-bound, polymerase is unwound and pulled into the enzyme.

Initial Transcription

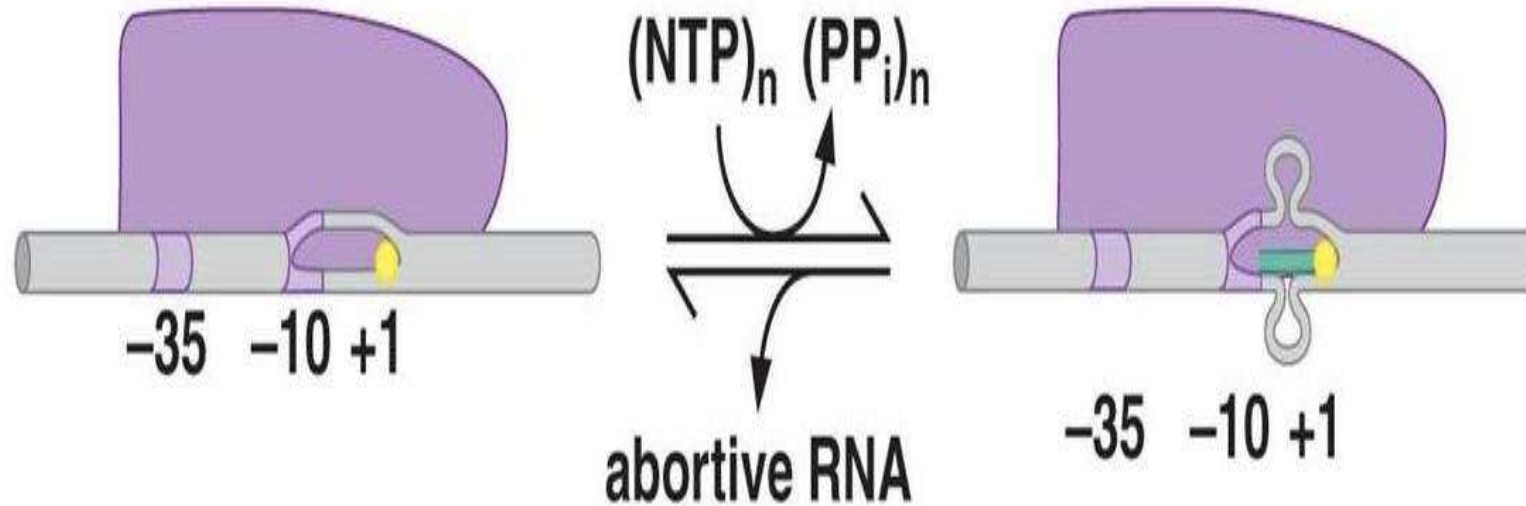
- The DNA thus accumulated within the enzyme is accommodated as single-stranded bulges.

Initial Transcription

- The DNA thus accumulated within the enzyme is accommodated as single-stranded bulges.

Initial Transcription

“scrunching”



© 2004 W. H. Freeman & Co. All rights reserved. This material is available at <http://www.ck12.org>

Initial Transcription

- It is now believed that the third model—scrunching—reflects what actually happens.

Initial Transcription

- The experiments have shown that during initial transcription, the polymerase remains stationary on the promoter, unwinds downstream DNA, and pulls that DNA into itself.

Initial Transcription

- Only the scrunching model is consistent with these results.

END

Initial Transcription

- RNA polymerase can initiate a new RNA chain on a DNA template and thus does not need a primer.

Initial Transcription

- This requires that the DNA template be brought into the polymerase active site and held stably in a helical conformation.

Initial Transcription

- And that the initiating ribonucleotide be brought into the active site and held stably on the template while the next NTP is presented with correct geometry for the polymerization to take place.

Initial Transcription

- This is particularly difficult because RNA polymerase starts most transcripts with an A, and that ribonucleotide binds the template nucleotide (T) with only two hydrogen bonds.

Initial Transcription

- Thus, the enzyme has to make specific interactions with the DNA template strand, the initiating ribonucleotide, and the second ribonucleotide to hold them all rigidly in the correct orientation to allow chemical attack by the incoming NTP.

Initial Transcription

- This is the reason that most transcripts start with the same nucleotide.

Initial Transcription

- During initial transcription, RNA polymerase produces and releases short RNA transcripts of <10 nucleotides (abortive synthesis) before escaping the promoter, entering the elongation phase, and synthesizing the proper transcript.

Initial Transcription

- It has long been unclear how the enzyme's active site translocates along the DNA template during initial abortive cycles of transcription.
- Three general models were proposed to explain this.

Initial Transcription

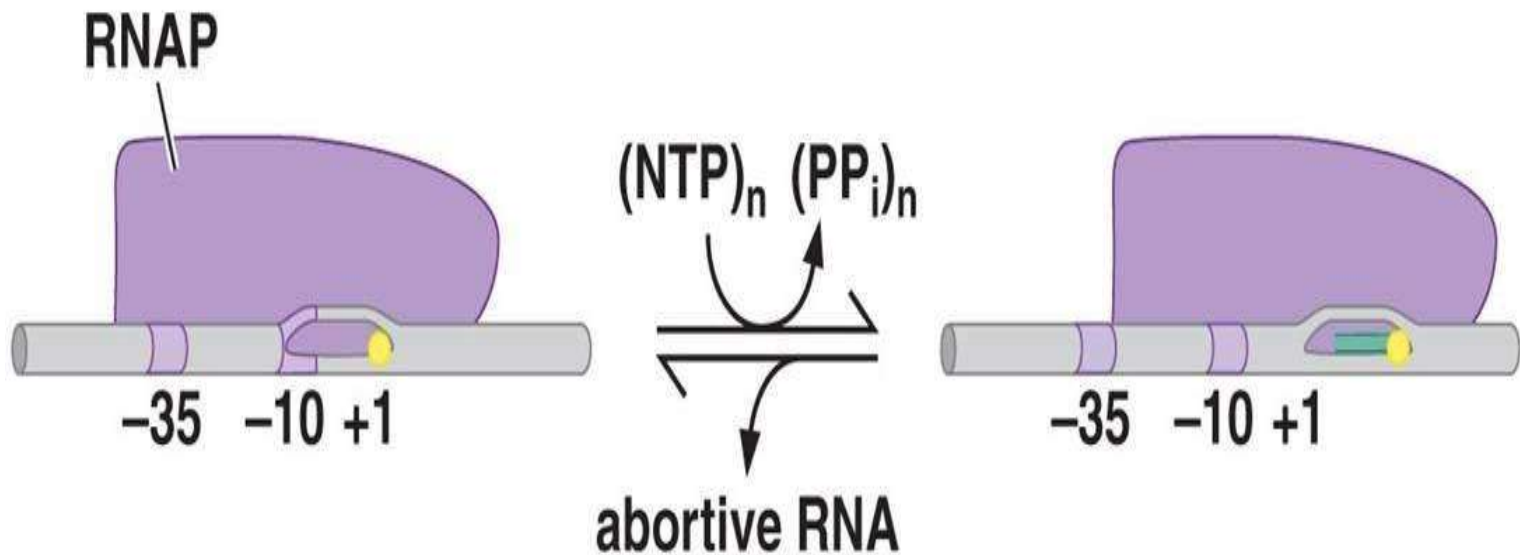
- 1. The “transient excursion” model proposes transient cycles of forward and reverse translocation of RNA polymerase.

Initial Transcription

- Thus, polymerase is thought to leave the promoter and translocate a short way along the DNA template, synthesizing a short transcript before aborting transcription, releasing the transcript, and returning to its original location on promoter.

Initial Transcription

“transient excursions”



Initial Transcription

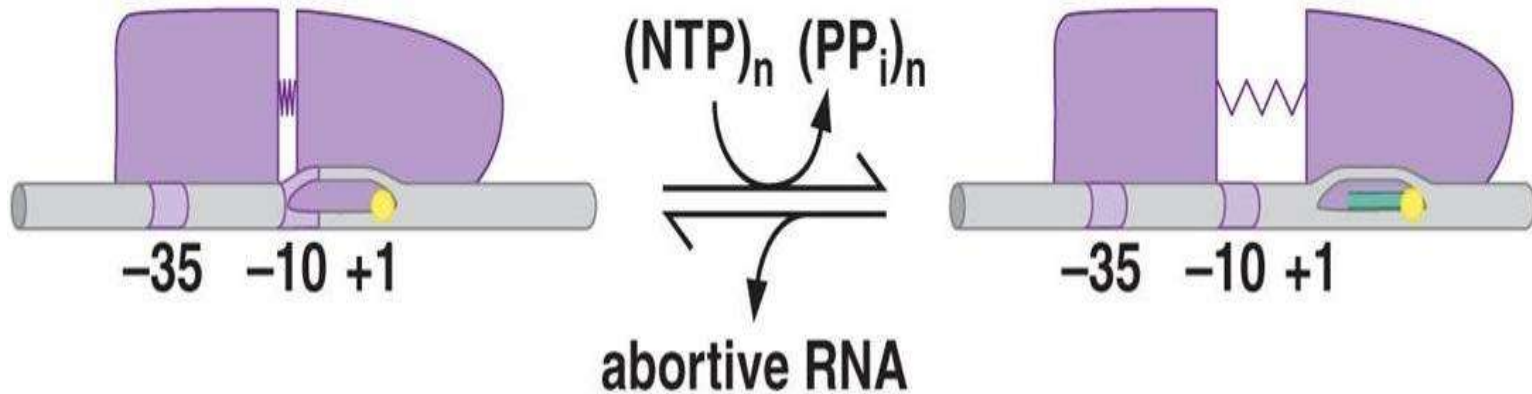
- 2. “Inchworming” invokes a flexible element within the polymerase that allows a module at the front of the enzyme, containing the active site.

Initial Transcription

- The module moves downstream, synthesizing a short transcript before aborting and retracting to the body of the enzyme still at the promoter.

Initial Transcription

“inchworming”



Initial Transcription

- 3. “Scrunching”
proposes that DNA downstream from the stationary, promoter-bound, polymerase is unwound and pulled into the enzyme.

Initial Transcription

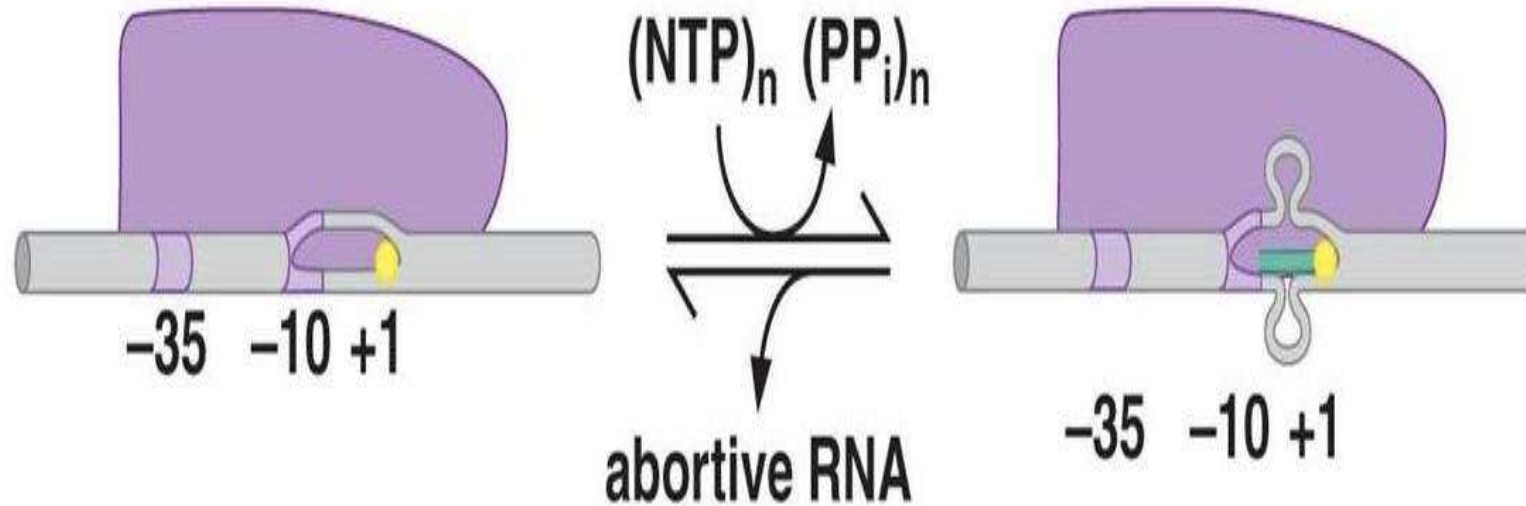
- The DNA thus accumulated within the enzyme is accommodated as single-stranded bulges.

Initial Transcription

- The DNA thus accumulated within the enzyme is accommodated as single-stranded bulges.

Initial Transcription

“scrunching”



© 2004 W. H. Freeman & Co. All rights reserved. This material is available at <http://www.ck12.org>

Initial Transcription

- It is now believed that the third model—scrunching—reflects what actually happens.

Initial Transcription

- The experiments have shown that during initial transcription, the polymerase remains stationary on the promoter, unwinds downstream DNA, and pulls that DNA into itself.

Initial Transcription

- Only the scrunching model is consistent with these results.

END

The Elongating Polymerase

- The Elongating Polymerase Is a processive machine that synthesizes and proofreads RNA.
- DNA passes through the elongating enzyme in a manner very similar to its passage through the open complex.

The Elongating Polymerase

- Thus, double-stranded DNA enters the front of the enzyme between the pincers.
- At the opening of the catalytic cleft, the strands separate to follow different paths through the enzyme.

The Elongating Polymerase

- The strands exit via their respective channels and re-form a double helix behind the elongating polymerase.

The Elongating Polymerase

- Ribonucleotides enter the active site through their defined channel and are added to the growing RNA chain under the guidance of the template DNA strand.

The Elongating Polymerase

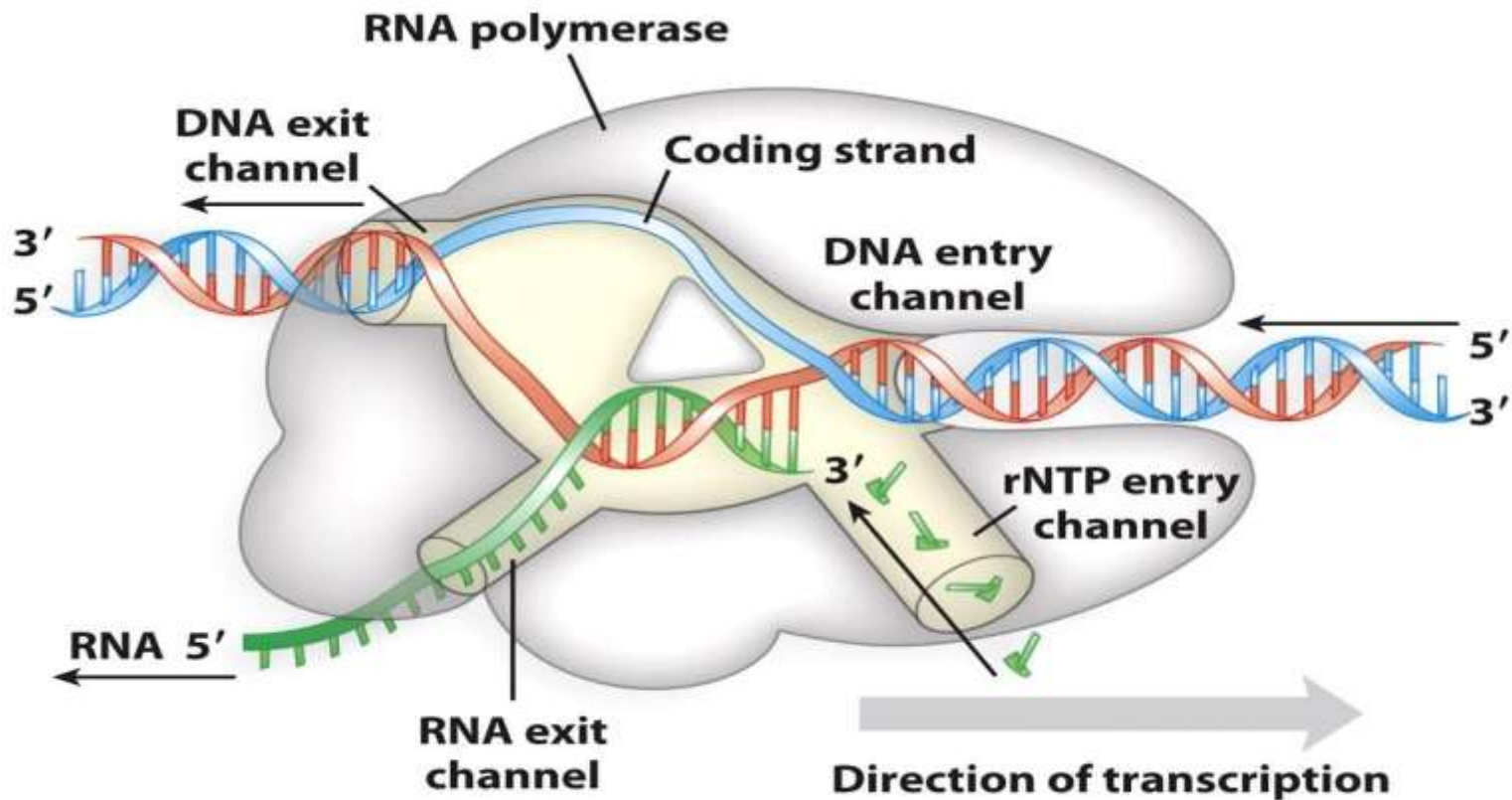


Figure 15-14
Molecular Biology: Principles and Practice
© 2012 W. H. Freeman and Company

The Elongating Polymerase

- Only 8 or 9 nucleotides of the growing RNA chain remain base-paired to the DNA template at any given time.
- The remainder of the RNA chain is peeled off and directed out of the enzyme through the RNA exit channel.

The Elongating Polymerase

- During elongation, the enzyme adds one nucleotide at a time to the growing RNA transcript.
- During elongation, polymerase uses a step mechanism: using single-molecule techniques.

The Elongating Polymerase

- It was shown that the enzyme steps forward as a molecular motor, advancing in a single step a distance equivalent to a base pair for every nucleotide it adds to the growing RNA chain.

The Elongating Polymerase

- In addition, the size of the bubble, i.e., the length of DNA that is not double-helical, remains constant throughout elongation.
- 1 bp is separated ahead of the processing enzyme, 1 bp is formed behind it.

The Elongating Polymerase

- As well as synthesizing the transcript, RNA polymerase performs two proofreading functions on that growing transcript.
- The first of these is called pyrophosphorolytic editing.

The Elongating Polymerase

- In this, the enzyme uses its active site, in a simple back-reaction, to catalyze the removal of an incorrectly inserted ribonucleotide, by reincorporation of PPI.

The Elongating Polymerase

- The enzyme can then incorporate another ribonucleotide in its place in the growing RNA chain.

The Elongating Polymerase

- In the second proofreading mechanism, called **hydrolytic editing**, the polymerase backtracks by one or more nucleotides and cleaves the RNA product, removing the error-containing sequence.

The Elongating Polymerase

- Another group of proteins—the **Nus proteins**—joins polymerase in the elongation phase and promotes the processes of elongation and termination.

END

Termination of Transcription

- The termination of transcription is a normal and important function at the ends of gene/s.
- However, an unusual termination may be triggered by damaged DNA or by other unanticipated hindrances.

Termination of Transcription

- The sequences called **terminators** trigger the elongating polymerase to dissociate from the DNA and release the RNA chain it has made.

Termination of Transcription

- In bacteria, terminators come in two types:-
 - 1. Rho-Dependant
 - 2. Rho-Independent

Termination of Transcription

- The first, requires a protein called Rho to induce termination.
- The second causes termination without the involvement of other factors.

Termination of Transcription

- Rho-dependent terminators have rather ill-defined RNA elements called **rut sites**.
- The work of these sites requires the action of the Rho factor.

Termination of Transcription

- Rho, which is a ring-shaped protein with six identical subunits, binds to single-stranded RNA as it exits the polymerase.

Termination of Transcription

- The protein also has an ATPase activity, and once attached to the transcript, Rho uses the energy derived from ATP hydrolysis to induce termination.

Termination of Transcription

- The precise mechanism of termination remains to be determined, and models include the following:-

Termination of Transcription

- i. Rho pushes polymerase forward relative to the DNA and RNA, resulting in termination in a manner analogous to termination by the protein TRCF (transcription-repair coupling factor).

Termination of Transcription

- **Rho** binds double-stranded DNA upstream of the polymerase and translocates along the DNA until it collides the RNA polymerase.

Termination of Transcription

- The collision pushes polymerase forward, causing dissociation of the ternary complex of RNA polymerase, template DNA, and RNA transcript.

END

Termination of Transcription

- ii. Rho pulls RNA out of the polymerase, resulting in termination; or
- iii. Rho induces a conformational change in polymerase, causing the enzyme to terminate the transcription.

Termination of Transcription

- Most recent experiments suggest that the last of the above may be correct and that the conformational change causes the elongating complex to stall, with dissociation following slowly.

Termination of Transcription

- Recent studies also have suggested that Rho binds to RNA polymerase throughout the transcription cycle.
- Thus, Rho doesn't reach polymerase by translocating along a nascent, rut-containing transcript like TRCF.

Termination of Transcription

- Rather, Rho binds polymerase early in transcription and then at some point also binds the RNA transcript being exuded from that elongating enzyme.

Termination of Transcription

- The role of translocation by Rho is thus perhaps to tighten the resulting RNA loop, and when sufficiently tight, polymerase elongation ceases.
- How is Rho directed to work on particular RNA transcripts?

Termination of Transcription

- First, there is some specificity in the sites it binds (the rut sites).
- These sites consist of stretches of ~40 nucleotides that do not fold into a secondary structure. They are also rich in C residues.

Termination of Transcription

- The second level of specificity is that Rho fails to bind any transcript that is being translated.

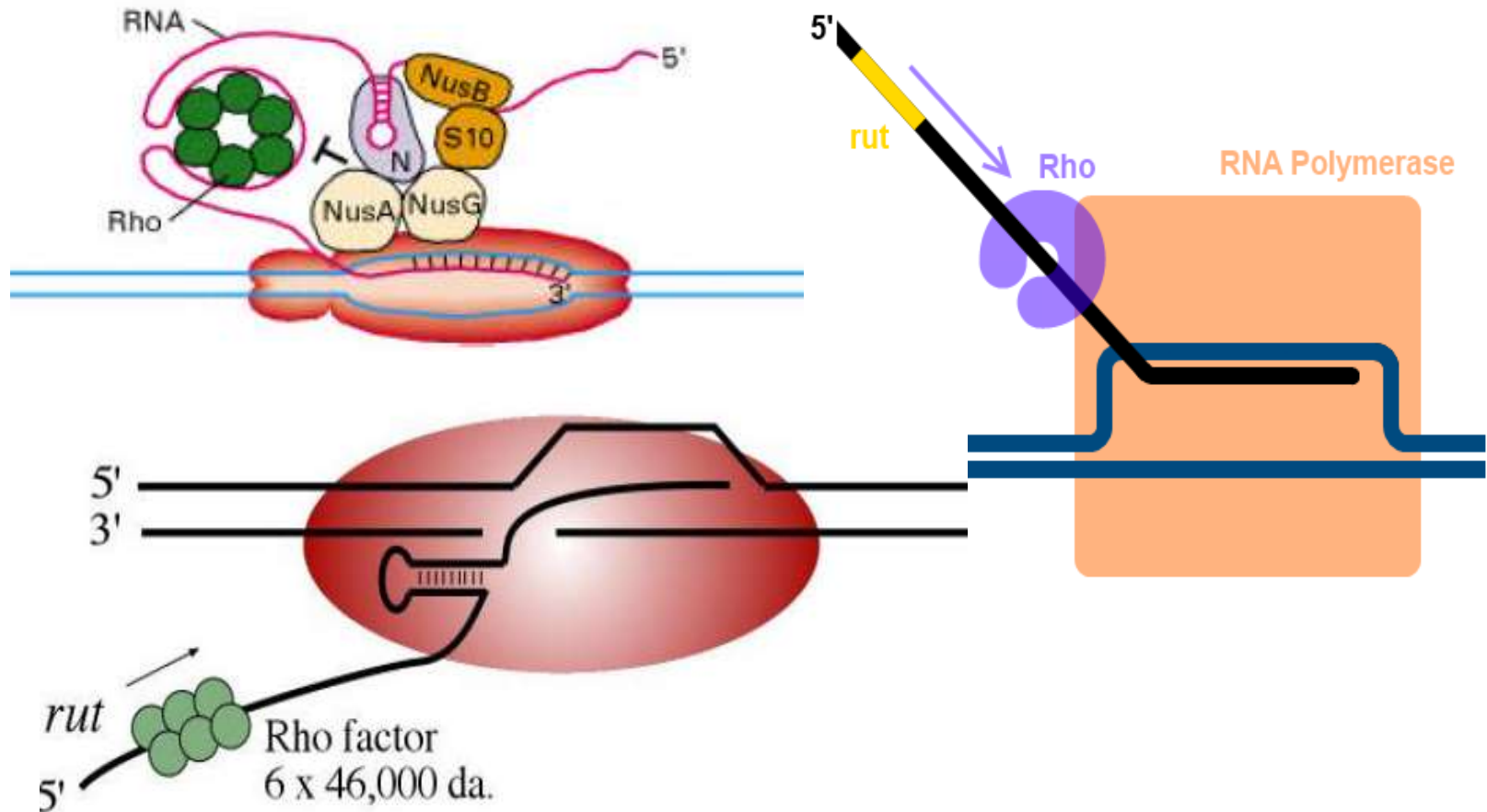
Termination of Transcription

- In bacteria, transcription and translation are tightly coupled—translation initiates on growing RNA transcripts as soon as they start exiting polymerase, while they are still being synthesized.

Termination of Transcription

- Thus, Rho typically terminates only those transcripts which are still being transcribed beyond the end of a gene or operon.

Termination of Transcription



Rho-independent Terminators

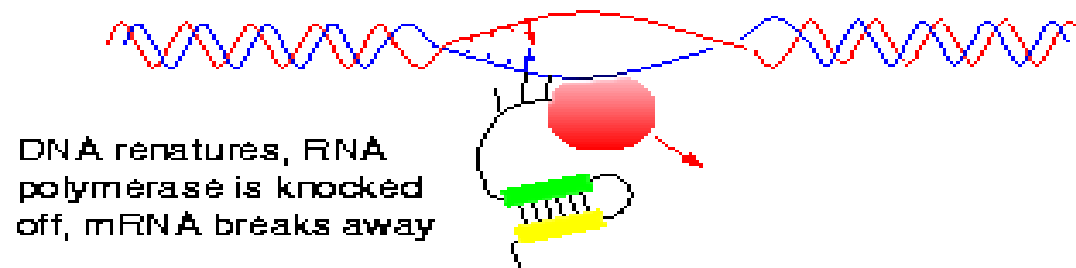
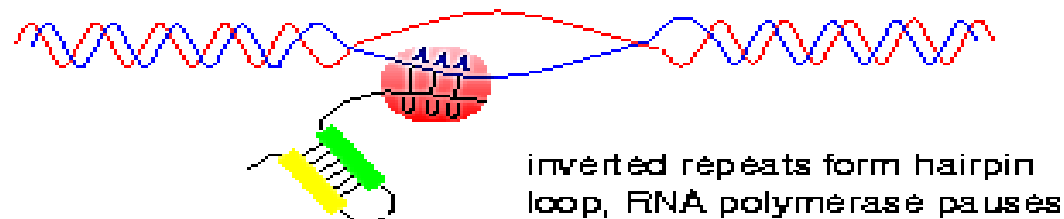
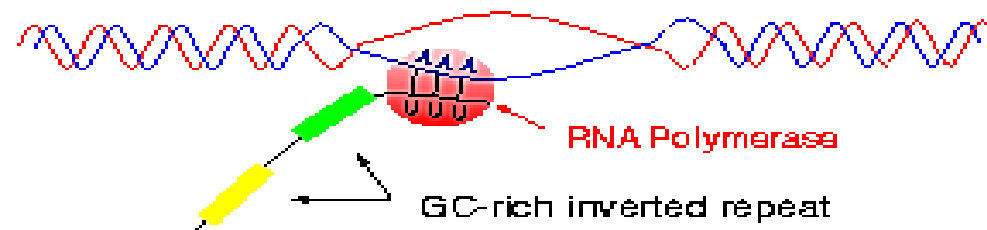
- Rho-independent terminators are also called **intrinsic terminators** because they need no other factors to work.

Rho-independent Terminators

- They consist of two sequence elements:-
- A short inverted repeat of 20 nucleotides followed by a stretch of about eight A:T base pairs.

Rho-independent Terminators

Chain Termination



Rho-independent Terminators

- These elements do not affect the polymerase until they have been transcribed—that is, they function in the RNA rather than in the DNA.

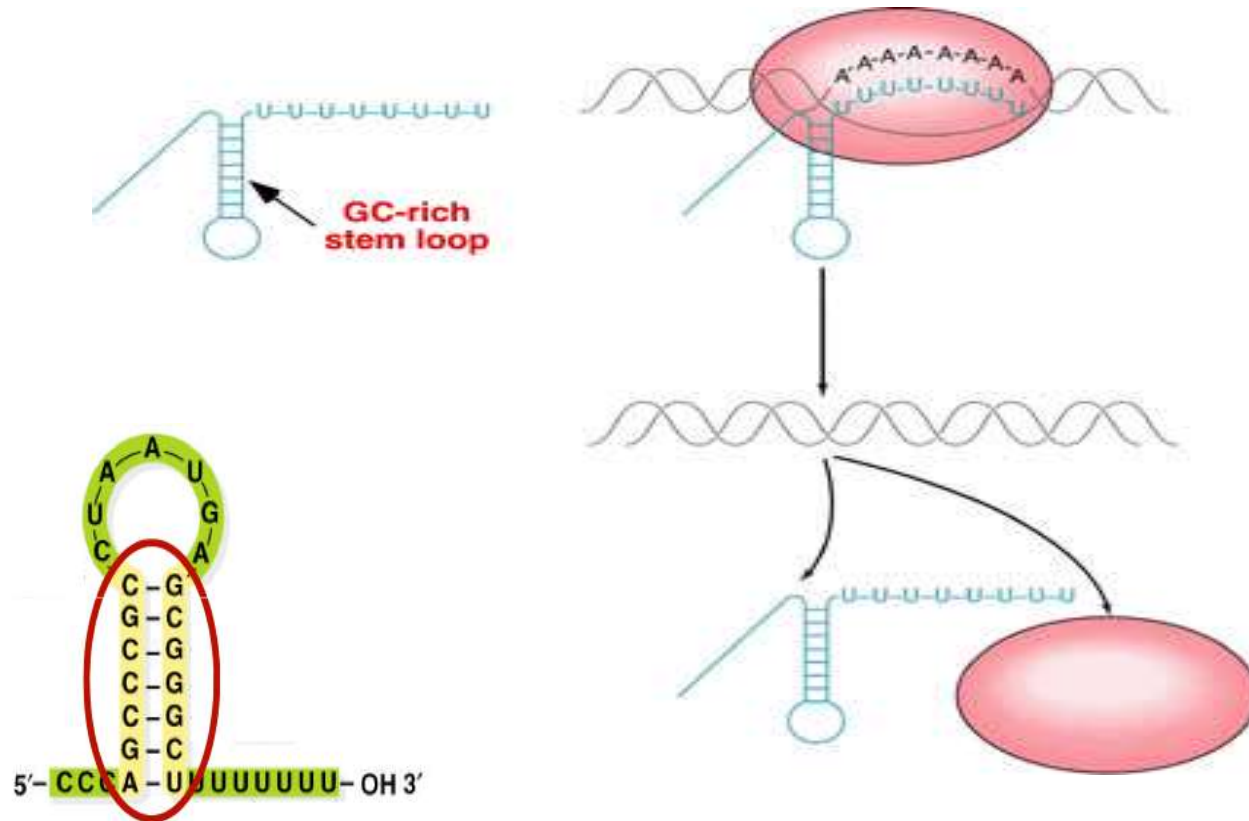
Rho-independent Terminators

- When polymerase transcribes an inverted repeat sequence, the resulting RNA can form a stem-loop structure often called a “hairpin” by base-pairing with itself.

Rho-independent Terminators

- Formation of the hairpin causes termination by disrupting the elongation complex.

Rho-independent Terminators



Rho-independent Terminators

- The hairpin induces termination by either pushing polymerase forward relative to the DNA and RNA, wresting the transcript from polymerase, or inducing a conformational change in polymerase.

Rho-independent Terminators

- The hairpin works as an efficient terminator only when it is followed by a stretch of A:U base pairs.

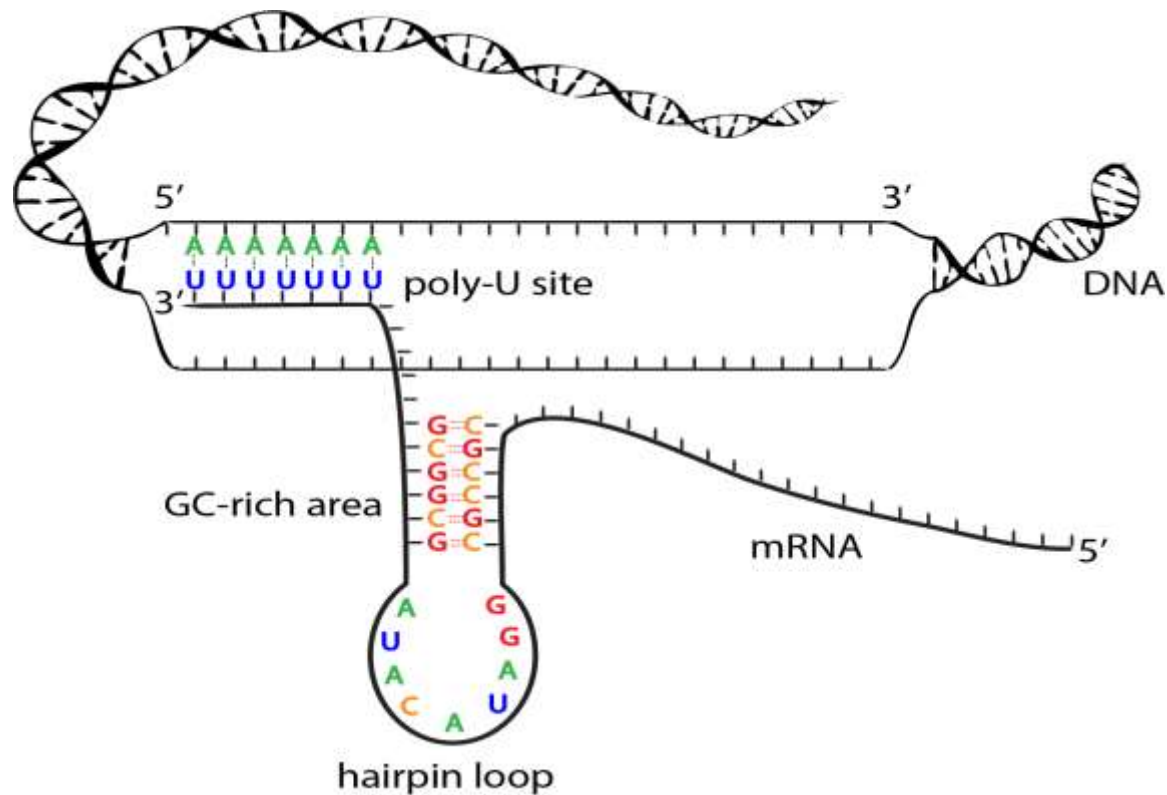
Rho-independent Terminators

- This is because, under those circumstances, at the time the hairpin forms, the growing RNA chain will be held on the template at the active site by only A:U base pairs.

Rho-independent Terminators

- Because A:U base pairs are the weakest of all 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.

Rho-independent Terminators



TRANSCRIPTION IN EUKARYOTES

- Transcription in eukaryotes is undertaken by polymerases closely related to the RNA polymerases found in prokaryotes.

TRANSCRIPTION IN EUKARYOTES

- The process of transcription is identical in both the prokaryotes and eukaryotes. There are, however, differences in the machinery used in each case.

TRANSCRIPTION IN EUKARYOTES

- Bacteria have only one RNA polymerase but all eukaryotes have at least three different ones i.e., Pol I, II, and III; and plants also have a Pol IV and a Pol V.

TRANSCRIPTION IN EUKARYOTES

- In addition, whereas bacteria require only one additional initiation factor (σ), several initiation factors are required for efficient and promoter-specific initiation in eukaryotes.

TRANSCRIPTION IN EUKARYOTES

- These are called the general transcription factors (GTFs).
- In vitro, the general transcription factors are all that are required, together with Pol II, to initiate transcription on a DNA template.

TRANSCRIPTION IN EUKARYOTES

- In vivo, however, the general transcription factors are not alone sufficient to bind promoter sequences and elicit significant expression.

TRANSCRIPTION IN EUKARYOTES

- Rather, additional factors are required, including DNA-binding regulatory proteins, the so-called Mediator complex, and often chromatin-modifying enzymes.

TRANSCRIPTION IN EUKARYOTES

- The eukaryotic **core promotor** refers to the minimal set of sequence elements required for accurate transcription initiation by the Pol II machinery.

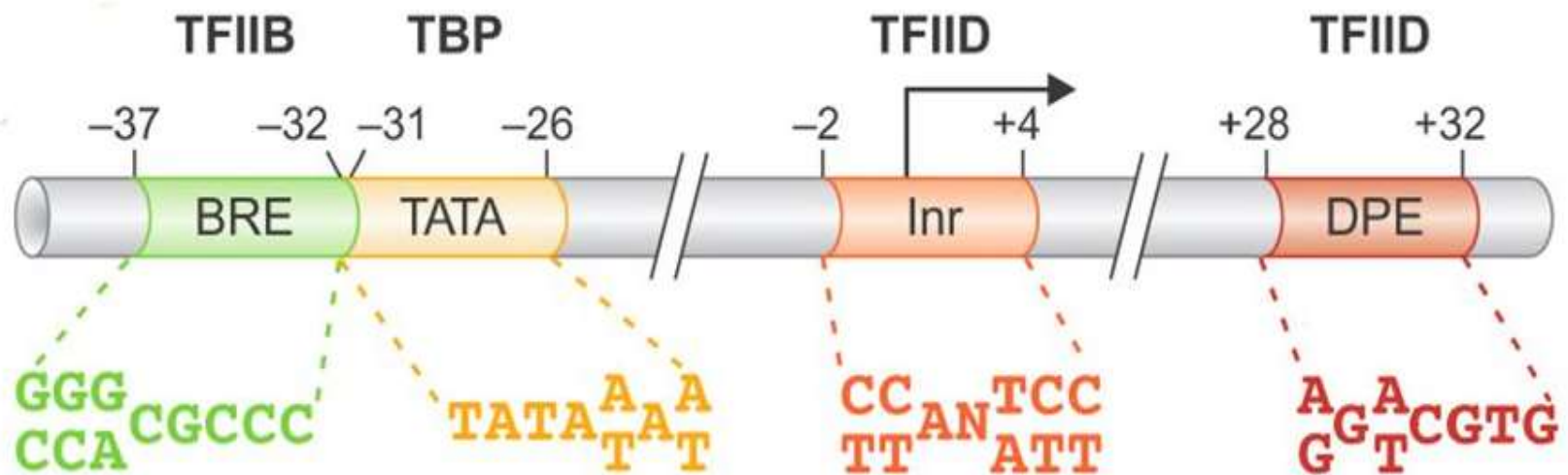
TRANSCRIPTION IN EUKARYOTES

- A core promoter is typically ~40–60 nucleotides long, extending either upstream or downstream from the transcription start site.

TRANSCRIPTION IN EUKARYOTES

- The elements found in Pol II core promoter include the TFIIB recognition element (BRE), the TATA element (or box), the initiator (Inr), and the downstream promoter elements (known as DPE, DCE, and MTE).

TRANSCRIPTION IN EUKARYOTES



Pol II core promoter

TRANSCRIPTION IN EUKARYOTES

- Typically, a promoter includes some subset of these elements. Thus, for example, promoters typically have either a TATA element or a DPE element, not both.

TRANSCRIPTION IN EUKARYOTES

- Often, a TATA-containing promoter also contains a DCE.
- The Inr is the most common element, found in combination with both TATA and DPEs.

TRANSCRIPTION IN EUKARYOTES

- Typically upstream of the core promoter, there are other sequence elements required for accurate and efficient transcription in vivo.
- Together, these elements constitute the **regulatory sequences**.

TRANSCRIPTION IN EUKARYOTES

- These elements include promoter proximal elements, upstream activator sequences (UASs), enhancers, and a series of other elements called silencers, boundary elements, and insulators.

TRANSCRIPTION IN EUKARYOTES

- All of these DNA elements bind regulatory proteins (activators and repressors), which help or hinder transcription from the core promoter.

TRANSCRIPTION IN EUKARYOTES

- Some of these regulatory sequences can be located many tens or even hundreds of kilobases from the core promoters on which they act.

END

Formation of pre-initiation Complex

- The general transcription factors collectively perform the functions performed by σ in bacterial transcription.

Formation of pre-initiation Complex

- Thus, the general transcription factors help polymerase bind to the promoter and melt the DNA.
- This is comparable to the transition from the closed to the open complex in the bacterial case.

Formation of pre-initiation Complex

- They also help polymerase escape from the promoter and embark on the elongation phase.

Formation of pre-initiation Complex

- The complete set of general transcription factors and polymerase, bound together at the promoter and poised for initiation, is called the **pre-initiation complex**.

Formation of pre-initiation Complex

- Many Pol II promoters contain a so-called TATA element (some 30 bp upstream of the transcription start site).
- This is where preinitiation complex formation begins.

Formation of pre-initiation Complex

- The TATA element is recognized by the general transcription factor called **TFIID** (“TFII” denotes a transcription factor for Pol II).

Formation of pre-initiation Complex

- Like many of the general transcription factors, TFIID is, in fact, a multi-subunit complex.
- The component of TFIID that binds to the TATA DNA sequence is called **TBP** (TATA-binding protein).

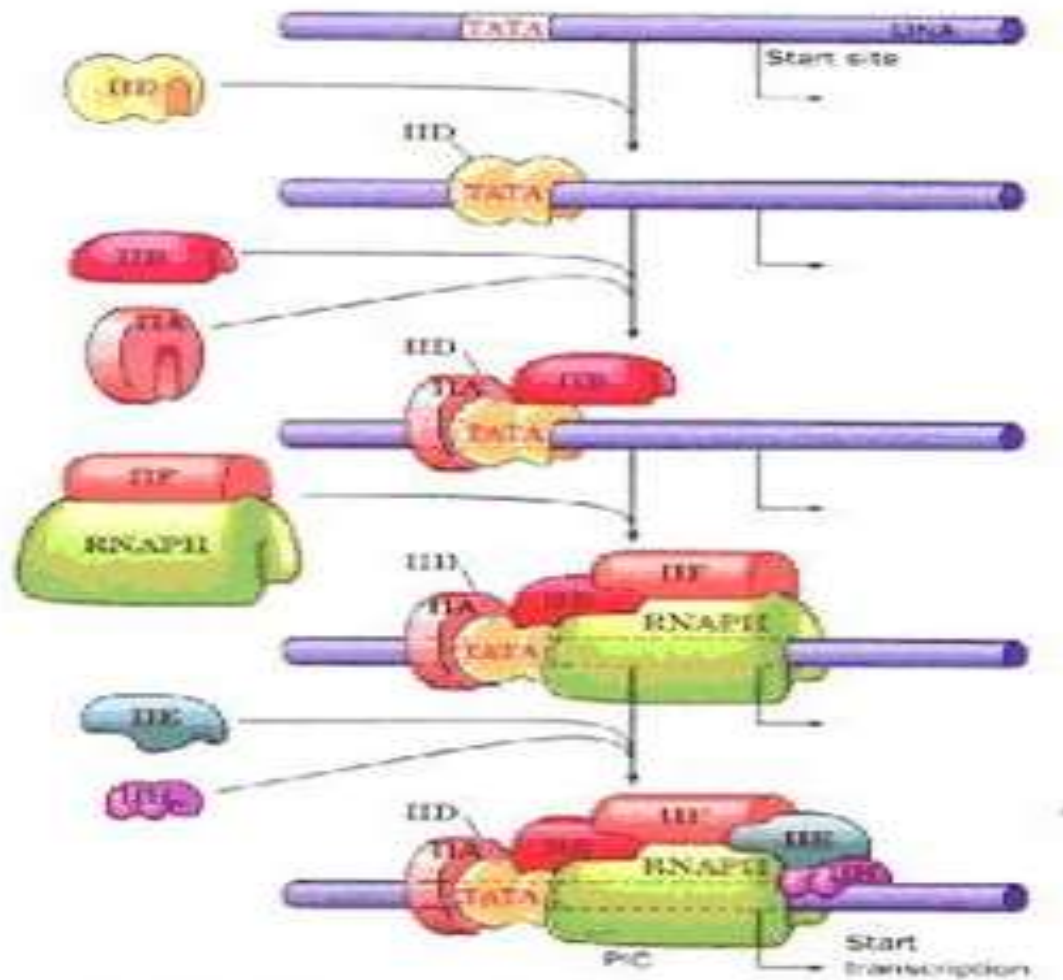
Formation of pre-initiation Complex

- The other subunits in this complex are called **TAFs**, for TBP-associated factors.
- Some TAFs recognize other core promoter elements such as the Inr, DPE, and DCE, although the strongest binding is between TBP and TATA.

Formation of pre-initiation Complex

- Thus, TFIID is a critical factor in promoter recognition and establishment of pre-initiation complex.

Formation of pre-initiation Complex



Formation of pre-initiation Complex

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

Formation of pre-initiation Complex

- In vitro, these proteins assemble at the promoter in the following order: TFIIA, TFIIB, TFIIF together with polymerase, and then TFIIE and TFIIH.

Formation of pre-initiation Complex

- Formation of the preinitiation complex containing these components is followed by promoter melting.
- In contrast to the situation in bacteria, promoter melting in eukaryotes requires hydrolysis of ATP and is mediated by TFIIF.

END

Promoter Escape

- The formation of pre-initiation complex follows a period of abortive initiation before the polymerase escapes the promoter and enters the elongation phase.

Promoter Escape

- During abortive initiation, the polymerase synthesizes a series of short transcripts.
- In eukaryotes, promoter escape involves two steps not seen in bacteria:-

Promoter Escape

- One is ATP hydrolysis (in addition to the earlier ATP hydrolysis needed for DNA melting), and the other is phosphorylation of the polymerase.

Promoter Escape

- The large subunit of Pol II has a carboxy-terminal domain (CTD), which is referred to as the “tail”.
- The CTD contains a series of repeats of the heptapeptide sequence: Tyr-Ser-Pro-Thr-Ser-Pro-Ser.

Promoter Escape

- There are 27 of these repeats in the yeast Pol II CTD, 32 in the worm *Caenorhabditis elegans*, 45 in the fly *Drosophila*, and 52 in humans.
- The number of repeats correlate with the complexity of the genome.

Promoter Escape

- Each repeat contains sites for phosphorylation by specific kinases, including one that is a subunit of TFIIF.

Promoter Escape

- The form of Pol II recruited to the promoter initially contains a largely unphosphorylated tail, but the species found in the elongation complex bears multiple phosphoryl groups on its tail.

Promoter Escape

- Addition of these phosphates helps polymerase shed most of the general transcription factors used for initiation, and which the enzyme leaves behind as it escapes the promoter.

Promoter Escape

- Regulating the phosphorylation state of the CTD of Pol II controls subsequent steps—elongation and even processing of the RNA— as well.

Promoter Escape

- Indeed, in addition to TFIIF, several other kinases have been identified that act on the CTD, as well as a number of phosphatases that remove the phosphates added by those kinases.

END

General Transcription Factors

- We do not know in detail the functions of all of the other general transcription factors.
- Some of these factors are in fact complexes made up of two or more subunits.

General Transcription Factors

- **TAFs**
- TBP is associated with about 10 TAFs. Two of the TAFs bind DNA elements at the promoter, for example, the initiator element (Inr) and the downstream promoter elements.

General Transcription Factors

- Several of the TAFs have structural homology with histone proteins, and it has been proposed that they might bind DNA in a similar manner, although evidence for such a binding has not been obtained.

General Transcription Factors

- For example, TAF42 and TAF62 from *Drosophila* have been shown to form a structure similar to that of the H3.H4 tetramer of histones.

General Transcription Factors

- Another TAF appears to regulate the binding of TBP to DNA. It does this using an inhibitory flap that binds to the DNA-binding surface of TBP.
- This flap must be displaced for TBP to bind TATA.

General Transcription Factors

- **TFIIB**
- This protein, a single polypeptide chain, enters the preinitiation complex after TBP.
- The crystal structure of the ternary complex of TFIIB–TBP–DNA shows specific TFIIB–TBP and TFIIB–DNA contacts.

General Transcription Factors

- These include base-specific interactions with the major groove upstream (to the BRE) and the minor groove downstream of the TATA element.

General Transcription Factors

- The asymmetric binding of TFIIB to the TBP–TATA complex accounts for the asymmetry in the rest of the assembly of the preinitiation complex and the unidirectional transcription that results.

General Transcription Factors

- TFIIB also contacts Pol II in the preinitiation complex.
- Thus, this protein appears to bridge the TATA-bound TBP and polymerase.

General Transcription Factors

- Structural studies suggest that segments of TFIIB insert into the RNA-exit channel and active center cleft of Pol II in a manner analogous to the σ region $3/4$ linker in the bacterial case.

General Transcription Factors

- These regions of TFIIB(called the **linker** and **reader**) aid in open complex formation, perhaps by stabilizing the melted DNA until the RNA:DNA hybrid takes over that role.

General Transcription Factors

- **TFIIF**
- This two-subunit factor associates with Pol II and is recruited to the promoter together with that enzyme (and other factors).

General Transcription Factors

- Binding of Pol II–TFIIF stabilizes the DNA–TBP–TFIIB complex and is required before TFIIE and TFIIH are recruited to the preinitiation complex.

General Transcription Factors

- In yeast, this factor includes a third subunit, but the function of the third subunit is not known.

END

General Transcription Factors

- **TFIIE and TFIIH**
- TFIIE, which consists of two subunits, binds next and has roles in the recruitment and regulation of TFIIH.

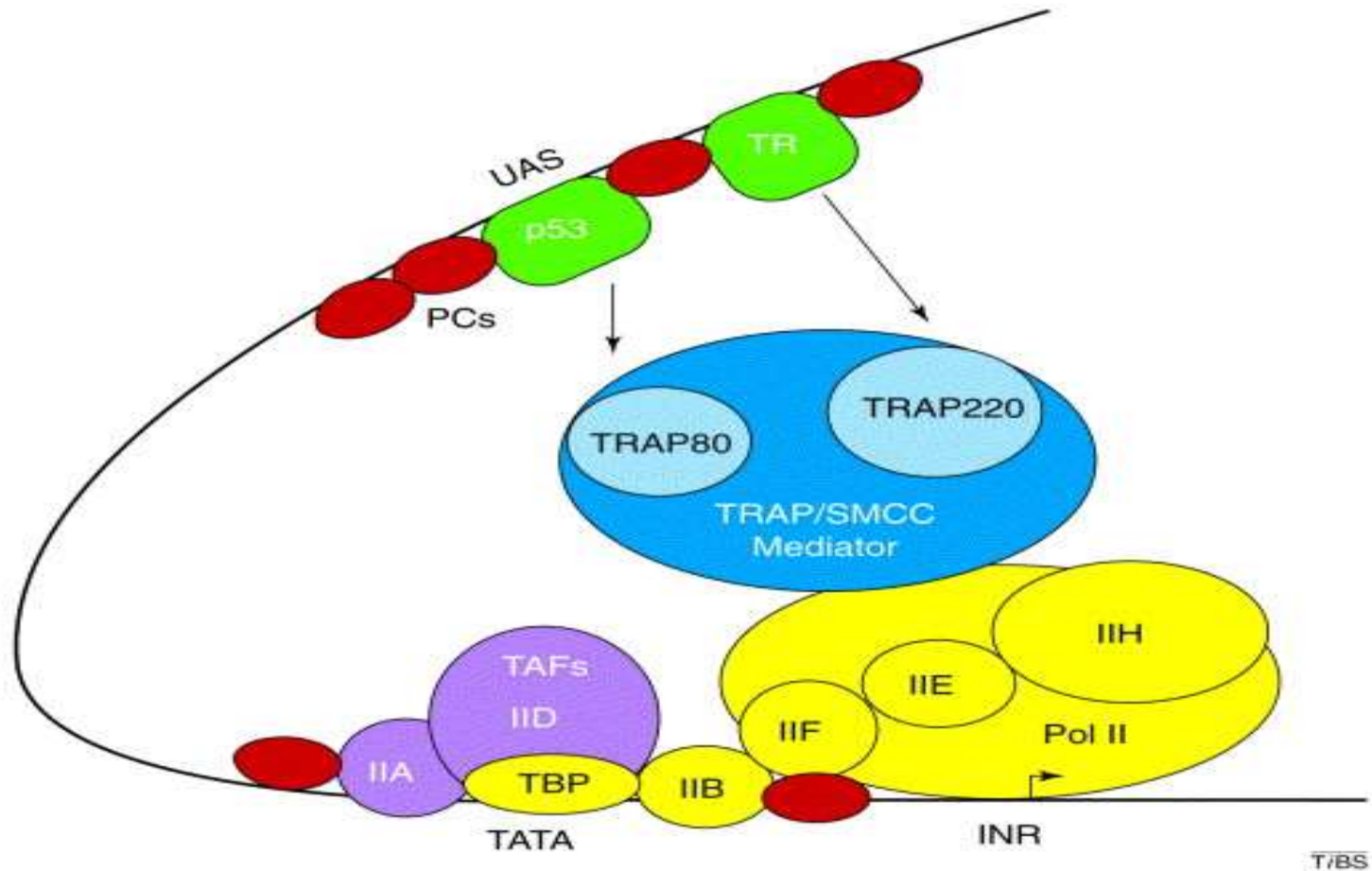
General Transcription Factors

- TFIIH controls the ATP-dependent transition of the preinitiation complex to the open complex.

General Transcription Factors

- It is the largest and most complex of the general transcription factors having 10 subunits and a molecular mass comparable to that of the polymerase itself.

General Transcription Factors



General Transcription Factors

- Within TFIIH are two subunits that function as ATPases and another that is a protein kinase, with roles in promoter melting and escape.

General Transcription Factors

- Together with other factors, the ATPase subunits are also involved in nucleotide excision repair.
- How does TFIIH mediate promoter melting?

General Transcription Factors

- It is now believed that a subunit of TFIIF acts as an ATP-driven translocator of double-stranded DNA.

General Transcription Factors

- This subunit binds to DNA downstream from polymerase and feeds double-stranded DNA, with a right-handed threading, into the cleft of the polymerase.

General Transcription Factors

- This action drives the melting of the DNA because the upstream promoter DNA is held in a fixed position by TFIID and the rest of the GTFs.

END

The Mediator Complex

- High regulated levels of transcription in vivo, require transcriptional regulatory proteins, the **Mediator complex**, and nucleosome modifying enzymes in addition to the above described general transcription factors.

The Mediator Complex

- One reason for these additional requirements is that the DNA template in vivo is packaged into chromatin.
- This condition complicates binding to the promoter of polymerase and its associated factors.

The Mediator Complex

- Transcriptional regulatory proteins called **activators** help recruit polymerase to the promoter, stabilizing its binding there.

The Mediator Complex

- This recruitment is mediated through interactions between DNA-bound activators, chromatin-modifying and -remodeling factors, and parts of the transcription machinery.

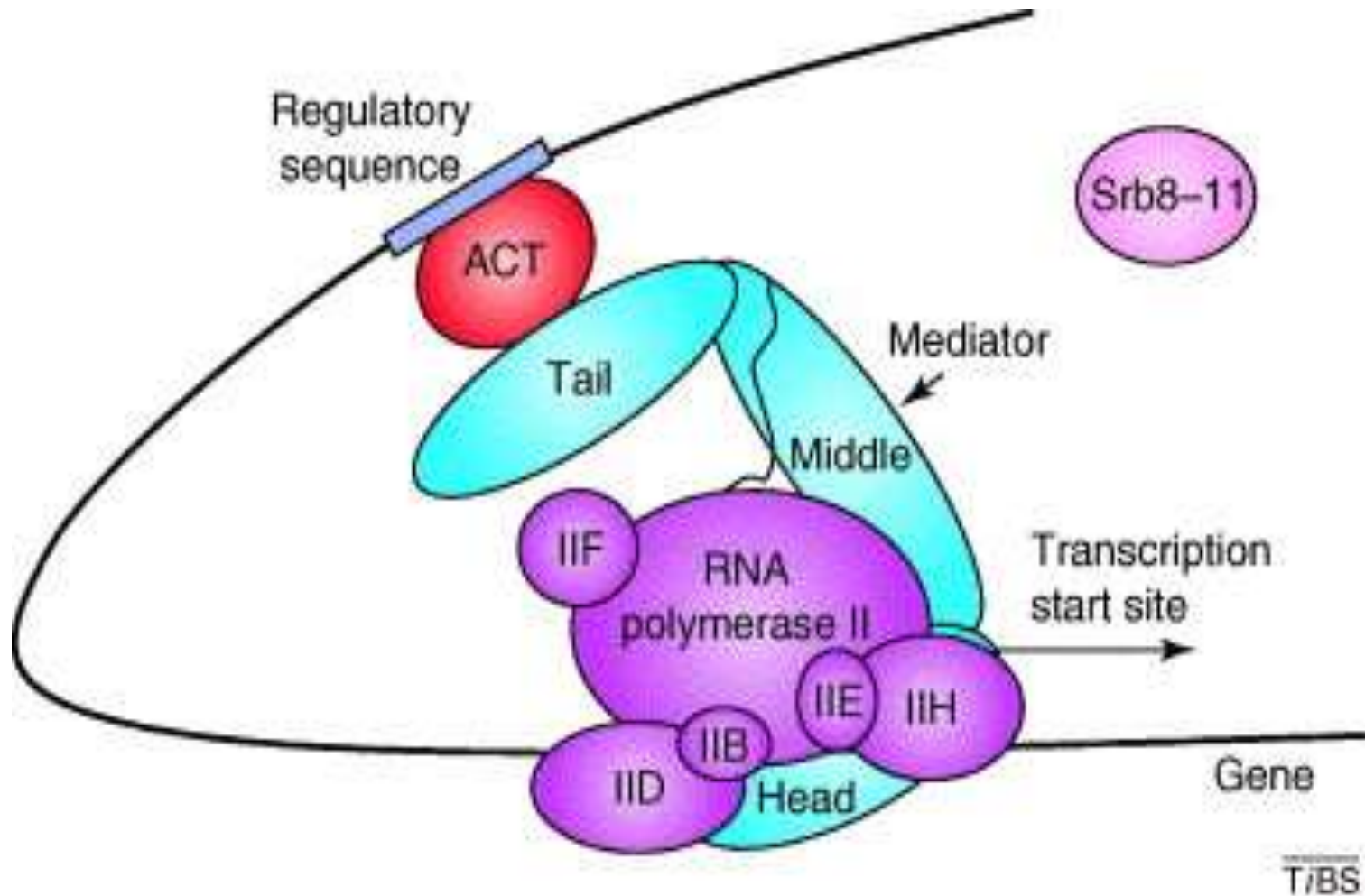
The Mediator Complex

- One such interaction is with the Mediator complex.
- Mediator is associated with the basic transcription machinery, most likely touching the CTD “tail” of the large polymerase subunit through one surface,

The Mediator Complex

- while presenting other surfaces for interaction with DNA-bound activators.
- This explains the need for Mediator to achieve significant transcription in vivo.

The Mediator Complex



The Mediator Complex

- Despite this central role in transcriptional activation, deletion of individual subunits of Mediator often leads to loss of expression of only a small subset of genes, different for each subunit.

The Mediator Complex

- This result likely reflects the fact that different activators are believed to interact with different Mediator subunits to bring polymerase to different genes.
- Mediator also aids initiation by regulating the CTD kinase in TFIIF.

The Mediator Complex

- The need for nucleosome modifiers and remodelers also differs at different promoters or even at the same promoter under different circumstances.

The Mediator Complex

- When and where required, these complexes are also typically recruited by the DNA-bound activators, or sometimes by regulatory RNAs.

The Mediator Complex

- The yeast and human Mediators each include more than 20 subunits, of which seven show significant sequence homology between the two organisms.

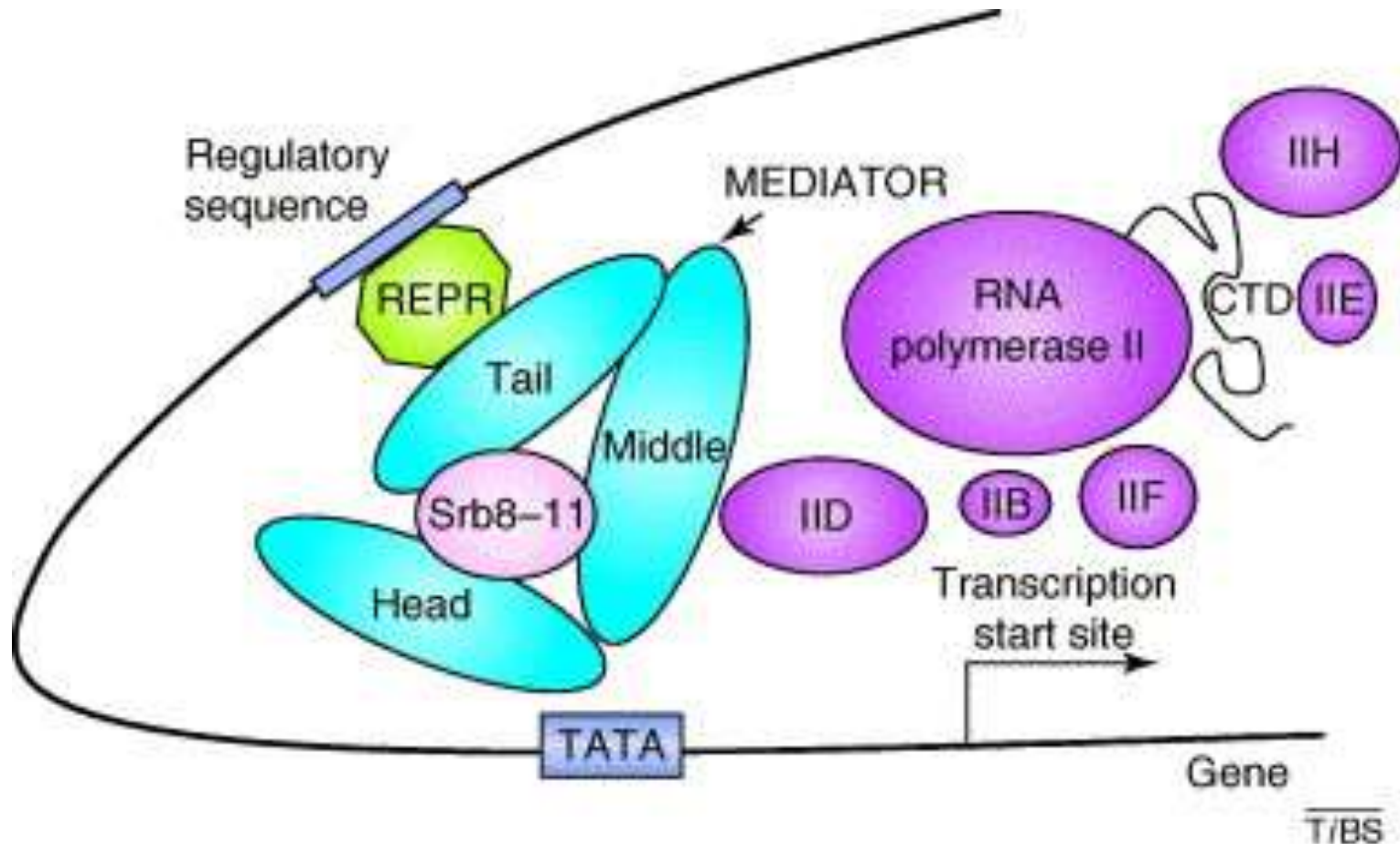
The Mediator Complex

- The Mediator from both yeast and humans is organized in modules, each containing a subset of the subunits.

The Mediator Complex

- These modules—called head, middle (or arm), and tail—can be dissociated from one another under certain conditions in vitro.

The Mediator Complex



The Mediator Complex

- Crystal structure of the head module of yeast Mediator reveals that it contains seven subunits (Med17/Srb4, Med11, Med22/Srb6, Med6, Med8, Med18/Srb5, and Med20/Srb2).

The Mediator Complex

- It forms a three-domain structure that binds the transcription complex in such a way as to juxtapose TFIID and the CTD tail of RNA polymerase, promoting phosphorylation of the latter by the former.

END

RNA Elongation and Proofreading

- Once polymerase has escaped the promoter and initiated transcription, it shifts into the elongation phase.

RNA Elongation and Proofreading

- This transition involves the Pol II enzyme shedding most of its initiation factors such as, the general transcription factors and Mediator.

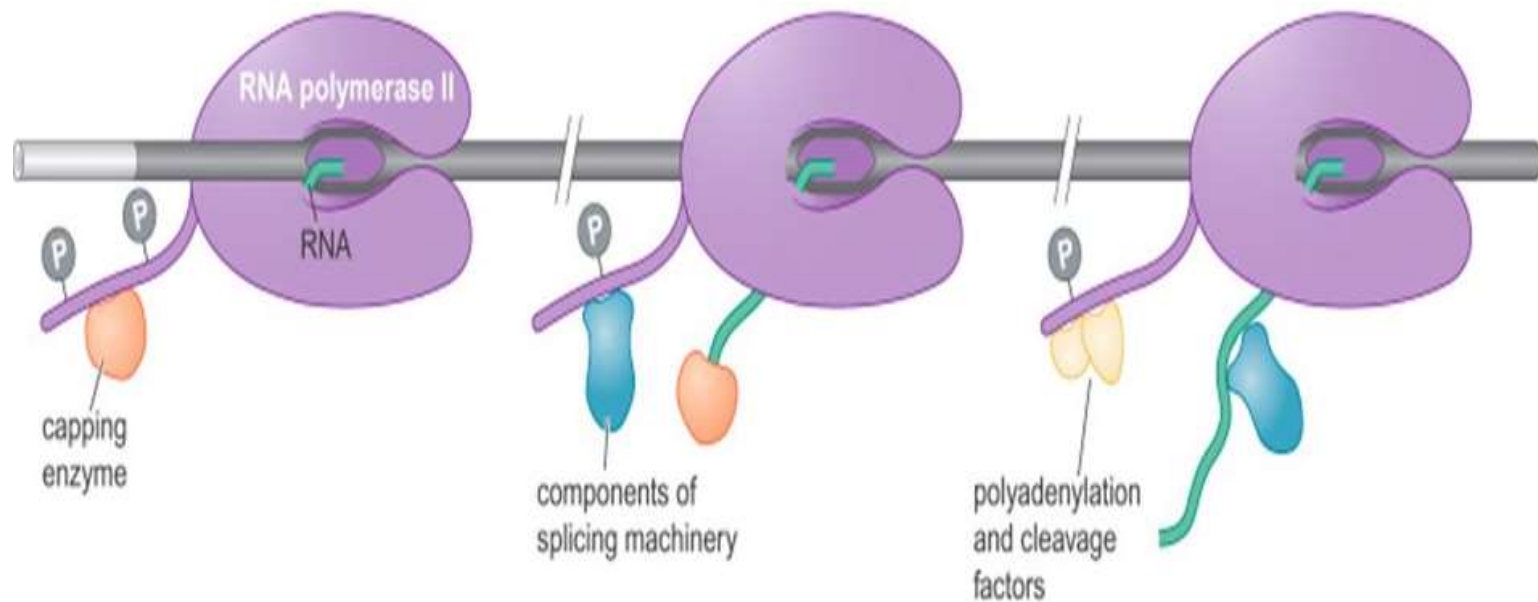
RNA Elongation and Proofreading

- In their place, another set of factors is recruited. Some of these (such as TFIIS and SPT5) are **elongation factors** (i.e., factors that stimulate elongation).
- Others are required for RNA processing.

RNA Elongation and Proofreading

- The enzymes involved in RNA processing are, like several of the initiation factors recruited to the carboxy-terminal (CTD) tail of the large subunit of Pol II

RNA Elongation and Proofreading



RNA Elongation and Proofreading

- In this case, however, the factors favor the phosphorylated form of the CTD. Thus, phosphorylation of the CTD leads to an exchange of initiation factors for those factors required for elongation and RNA processing.

RNA Elongation and Proofreading

- As is evident from the crystal structure of yeast Pol II, the polymerase CTD lies directly adjacent to the channel through which the newly synthesized RNA exits the enzyme.

RNA Elongation and Proofreading

- The CTD tail is also very long.
- It could potentially extend 800 Å from the body of the enzyme—that is, about seven times the length of the rest of the enzyme.

RNA Elongation and Proofreading

- Together, these features allow the tail to bind several components of the elongation and processing machinery and deliver them to the emerging RNA.

RNA Elongation and Proofreading

- Various proteins are thought to stimulate elongation by Pol II.
- One of these, the **kinase P-TEFb**, is recruited to polymerase by transcriptional activators.

RNA Elongation and Proofreading

- Once bound to Pol II, this protein phosphorylates the serine residue at position 2 of the CTD repeats. That phosphorylation event correlates with elongation.

RNA Elongation and Proofreading

- In addition, P-TEFb phosphorylates and thereby activates another protein, called **SPT5**, itself an elongation factor.
- Finally, **TAT-SF1**, yet another elongation factor, is recruited by P-TEFb.

RNA Elongation and Proofreading

- Thus, P-TEFb stimulates elongation in three separate ways.
- SPT5 is comparable to the bacterial elongation factor NusG.

RNA Elongation and Proofreading

- Indeed, this is the only universally conserved transcription factor across all three kingdoms of life—from bacteria, through Archaea, to eukaryotes.

END

RNA Elongation and Proofreading

- NusG/SPT5 factors bind to their respective RNA polymerases at the tip of the clamp, overlapping the region contacted by σ region 4 (in bacteria) and TFIIB (in eukaryotes).

RNA Elongation and Proofreading

- This overlapping and binding raises the interesting possibility that displacing initiation factors may be part of the function of these elongation regulators.

RNA Elongation and Proofreading

- This also suggests that regulating the rate of elongation is an ancient mechanism of regulating gene expression.

RNA Elongation and Proofreading

- There are some promoters in higher eukaryotes where the preinitiation complex is recruited effectively, but polymerase remains paused just after initiating transcription.

RNA Elongation and Proofreading

- Such promoters seem to be associated with genes poised to be expressed either rapidly or in a highly coordinated fashion.

RNA Elongation and Proofreading

- And their expression is regulated through recruitment by specific activators of the PTEFb kinase, which then releases them from their pause.

RNA Elongation and Proofreading

- Yet another class of elongation factor is the so-called **ELL family**.
- These also bind to elongating polymerase and suppress transient pausing by the enzyme.

RNA Elongation and Proofreading

- The first human ELL protein was originally identified as the product of a gene that undergoes translocations in acute myeloid leukemia.

RNA Elongation and Proofreading

- Another factor that does not affect initiation, but stimulates elongation, is **TFIIS**.

RNA Elongation and Proofreading

- This factor, like ELL, stimulates the overall rate of elongation by limiting the length of time that polymerase pauses when it encounters sequences that would otherwise tend to slow the enzyme's progress.

RNA Elongation and Proofreading

- It is a feature of polymerase that it does not transcribe through all sequences at a constant rate.
- Rather, it pauses periodically, sometimes for rather long periods, before resuming transcription.

RNA Elongation and Proofreading

- In the presence of TFIIIS, the length of time that polymerase pauses at any given site is reduced.
- TFIIIS also contributes to proofreading by polymerase.

RNA Elongation and Proofreading

- TFIIIS stimulates an inherent Rnase activity in polymerase (not part of the active site), allowing an alternative approach to removing misincorporated bases through local limited RNA degradation.

RNA Elongation and Proofreading

- This feature is comparable to the hydrolytic editing in the bacterial case stimulated by the Gre factors.

END

Transcription Termination

- The final RNA processing event, polyadenylation of the 3' end of the mRNA, is intimately linked with the termination of transcription, although exactly how is still not quite clear.

Transcription Termination

- Recently, however, an enzyme that degrades the second RNA as it emerges from the polymerase has been identified, and this enzyme may itself trigger termination. This is called the **torpedo model** of termination.

Transcription Termination

- The free end of the second RNA is uncapped and thus can be distinguished from genuine transcripts.

Transcription Termination

- This new RNA is recognized by an RNase called, in yeast, **Rat1** (in humans, **Xrn2**) that is loaded onto the end of the RNA by another protein (**Rtt103**) that binds the CTD of RNA polymerase.

Transcription Termination

- The Rat1 enzyme is very processive and quickly degrades the RNA in a 5'-to-3' direction, until it catches up to the still-transcribing polymerase from which the RNA is being spewed.

Transcription Termination

- Termination may not require very specific interaction between Rat1 and polymerase and might, in fact, be triggered in a manner rather similar to that of Rho-dependent termination in bacteria.

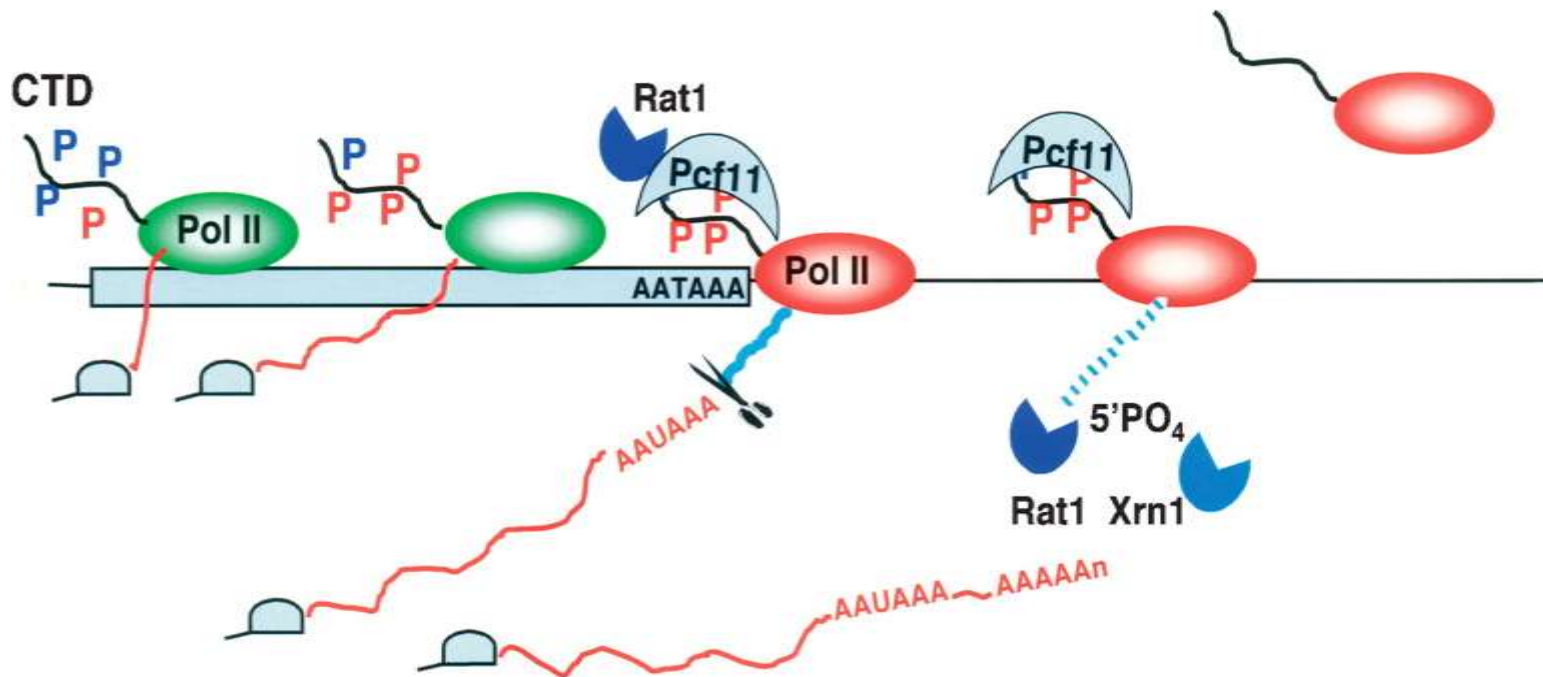
Transcription Termination

- The highly processing RNase polymerase either pushes polymerase forward and/or pulls the remains of the nascent RNA transcript from the enzyme.

Transcription Termination

- It is also possible that other factors are needed in addition to Rat1 to dislodge polymerase as, in vitro, Rat1 is alone insufficient to carry out this function, even after it has degraded the transcript.

Transcription Termination



Transcription Termination

- Although the torpedo model for termination is now the favored one, there is an alternative called the **allosteric model**.

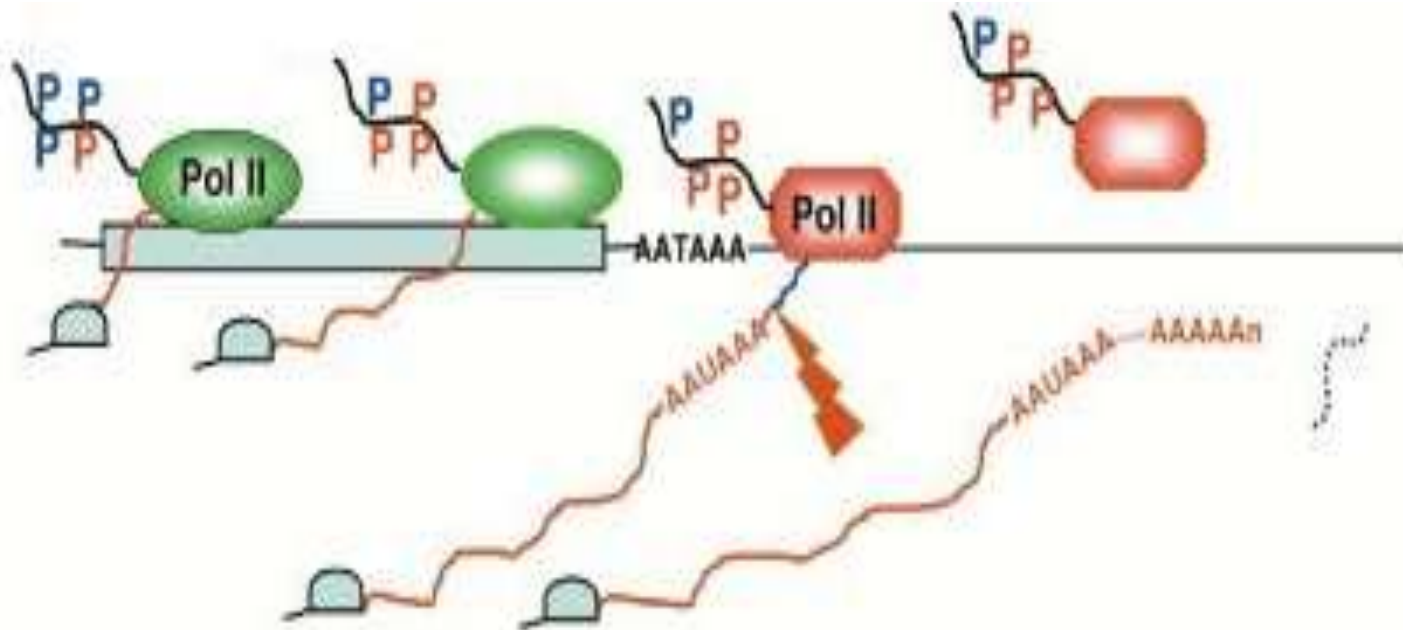
Transcription Termination

- According to this model, termination depends on a conformational change in the elongating polymerase that reduces the processivity of the enzyme leading to spontaneous termination soon afterward.

Transcription Termination

- This conformational change would be linked to polyadenylation and could be triggered by the transfer of the 3'-processing enzymes from the CTD tail of polymerase to the RNA.

Transcription Termination



Polymerases I and III

- All eukaryotes have two other RNA polymerases— Pol I and Pol III—in addition to Pol II.

Polymerases I and III

- These enzymes are related to Pol II and even share several subunits but they initiate transcription from distinct promoters and transcribe distinct genes.

Polymerases I and III

- Those genes encode specialized RNAs rather than proteins. Each of these enzymes also works with its own unique set of general transcription factors.

Polymerases I and III

- TBP, however, is universal—it is involved in initiating transcription by Pol I and Pol III, as well as Pol II.

Polymerases I and III

- Although TBP is the only GTF that is used by Pol I and Pol III as well as by Pol II, it has emerged recently that some of the other GTFs, in fact, have structurally and functionally equivalent components in the other systems.

Polymerases I and III

- Thus, for example, TFIIIF seems to have a counterpart in two subunits within Pol I (A49/34.5), and also in Pol II (C37/53).
- Likewise, TFIIIE-like subunits are found in Pol I and Pol III enzymes.

Polymerases I and III

- In addition, both these other systems include additional factors comparable to TFIIB: the TAF1B factor in the Pol I system, and the Brf1 subunit of TFIIB in the case of Pol III.

Polymerases I and III

- Pol I is required for the expression of only one gene, that encoding the rRNA precursor.

Polymerases I and III

- There are many copies of that gene in each cell, and, indeed, it is expressed at far higher levels than any other gene, perhaps explaining why it has its own dedicated polymerase.

Polymerases I and III

- The promoter for the rRNA gene comprises two parts: the core element and the UCE (upstream control element).

END

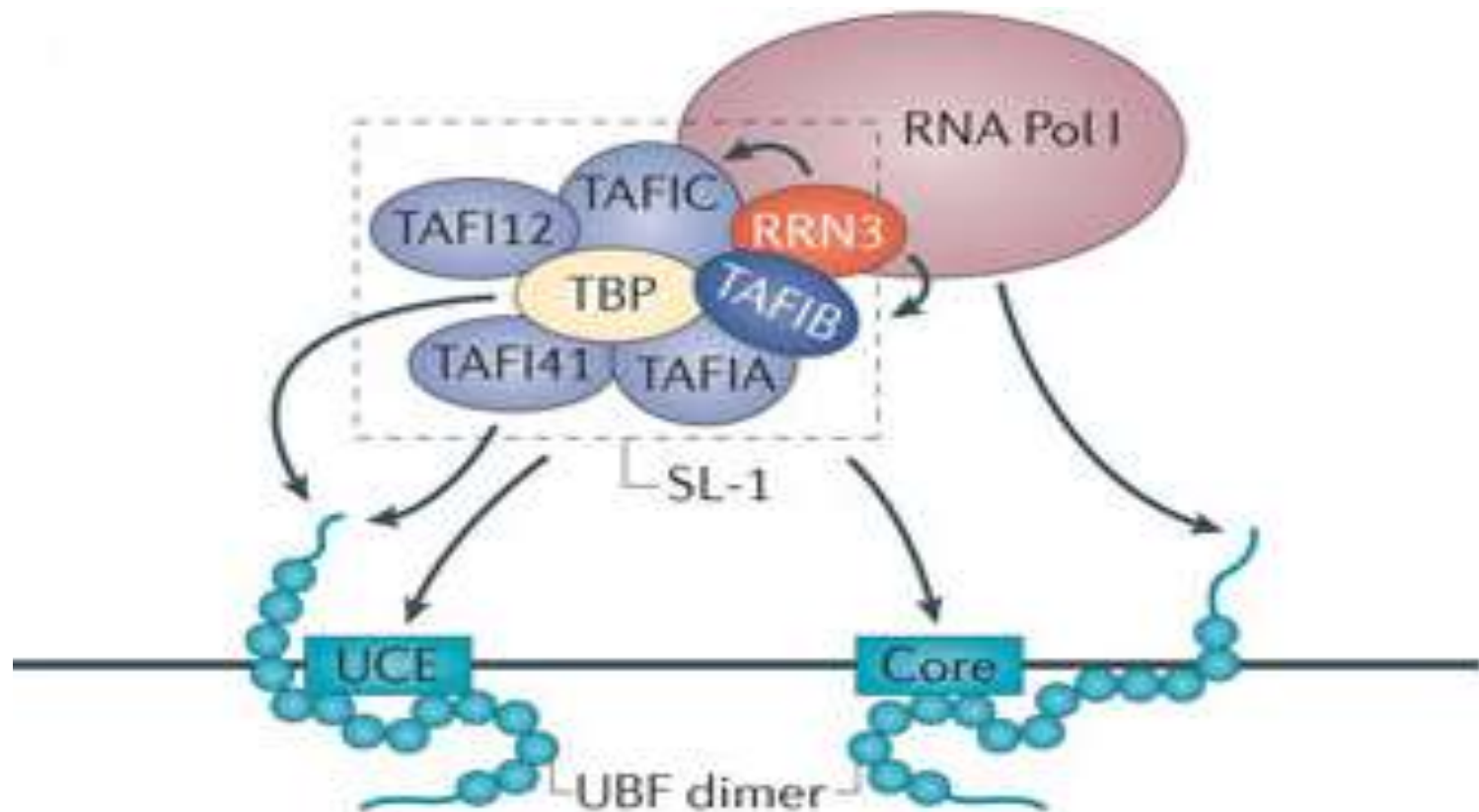
Polymerases I and III

- The former is located around the start site of transcription, and the latter between 100 and 150 bp upstream (in humans).
- In addition to Pol I, initiation requires two other factors, called SL1 and UBF.

Polymerases I and III

- SL1 comprises TBP and three TAFs specific for Pol I transcription. This complex binds to the core element.

Polymerases I and III



Polymerases I and III

- SL1 binds DNA only in the presence of UBF. This factor binds to UCE, bringing in SL1 and stimulating transcription from the core promoter by recruiting Pol I.

Polymerases I and 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).

Polymerases I and III

- 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.

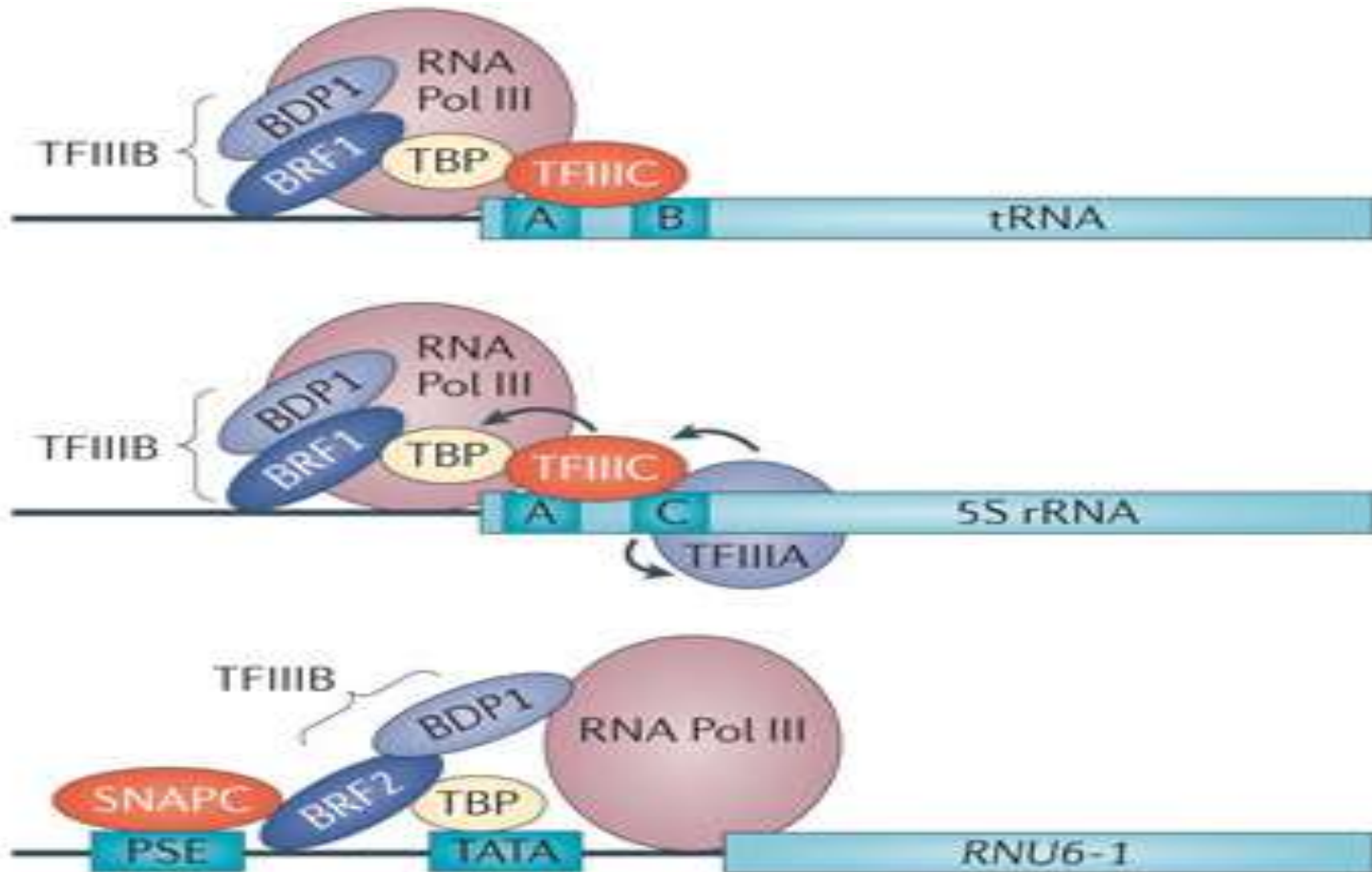
Polymerases I and III

- Others contain Box A and Box C (e.g., the 5S rRNA gene); and still others contain a TATA element like those of Pol II.

Polymerases I and III

- Transcription by Pol III also requires transcription factors in addition to polymerase.
- In this case, the factors are called TFIIIB and TFIIIC for the tRNA genes and those plus TFIIIA for the 5S rRNA gene.

Polymerases I and III



Polymerases I and III

- The TFIIIC complex binds to the promoter region. This complex recruits TFIIIB to the DNA just upstream of the start site, where it, in turn, recruits Pol III to the start site of transcription.

Polymerases I and III

- The enzyme then initiates, presumably displacing TFIIIC from the DNA template as it goes.
- Pol III also uses TBP like other polymerases. In this case, that ubiquitous factor is found within the TFIIB complex.

END

RNA Splicing

- The coding sequence of a gene is a series of three nucleotide codons that specifies the linear sequence of amino acids in its polypeptide product.

RNA Splicing

- It is generally assumed that the coding sequence is contiguous; i.e., the codon for one amino acid is immediately adjacent to the codon for the next amino acid in the polypeptide chain.

RNA Splicing

- This is true in the vast majority of cases in bacteria and their phage. But it is rarely so for eukaryotic genes.
- In those cases, the coding sequence is interrupted by stretches of non-coding sequences.

RNA Splicing

- Many eukaryotic genes are thus mosaics, consisting of blocks of coding sequences separated from each other by blocks of non-coding sequences.

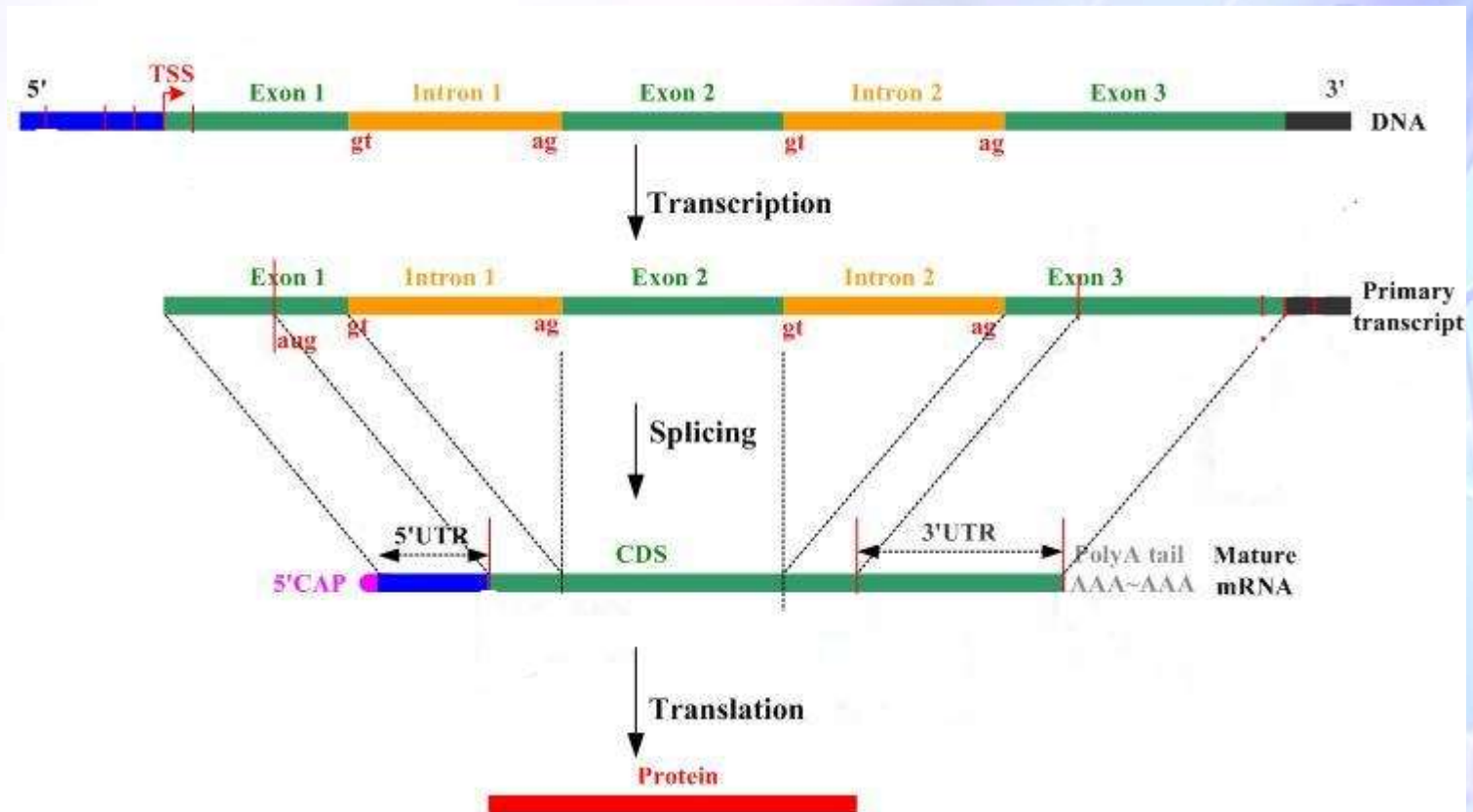
RNA Splicing

- The coding sequences are called **exons** and the intervening sequences are called **introns**.

RNA Splicing

- Once transcribed into an RNA transcript, the introns must be removed and the exons joined together to create the mRNA for that gene.

RNA Splicing



RNA Splicing

- The number of introns found within a gene varies enormously—from one in the case of most yeast genes (and a few human genes), to 50 in the case of the chicken pro α 2 collagen gene,

RNA Splicing

- to as many as 363 in the case of the Titin gene of humans.
- The sizes of the exons and introns vary as well.
- Indeed, introns are very often much longer than the exons they separate.

RNA Splicing

- Thus, for example, exons are typically on the order of 150 nucleotides, whereas introns—although they too can be short—can be as long as 800,000 nucleotides (800 kb).

RNA Splicing

- As another example, the mammalian gene for the enzyme dihydrofolate reductase is more than 31 kb long, and within it are dispersed six exons that correspond to 2 kb of mRNA.

RNA Splicing

- Thus, in this case, the coding portion of the gene is, 10% of its total length.

RNA Splicing

- Like the uninterrupted genes of prokaryotes, the split genes of eukaryotes are transcribed into a single RNA copy of the entire gene - the primary transcript that contains introns as well as exons.

RNA Splicing

- Because the length and number of introns, the primary transcript (or **pre-mRNA**) can be very long indeed.

RNA Splicing

- As already mentioned, the primary transcripts of intron-containing genes must have their introns removed before they can be translated into proteins.

RNA Splicing

- The process of intron removal is called **RNA Splicing**.
- It converts the pre-mRNA into mature mRNA containing only exons.

RNA Splicing

- RNA Splicing 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.

RNA Splicing

- The triplet-nucleotide codons of mRNA are translated in a fixed reading frame that is set by the first codon.
- Lack of precision in splicing will change the reading frames of exons.

RNA 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.

RNA Splicing

- Alternative splicing strategy enables a gene to give rise to more than one polypeptide product.
- These alternative products are called **isoforms**.

RNA Splicing

- It is estimated that 90% or more of the protein-coding genes in the human genome are spliced in alternative ways to generate more than one isoform.

How splicing site is determined?

- Let us consider the molecular mechanisms of the splicing reaction.
- How are the introns and exons distinguished?
- How are introns removed?
- How are exons join with high precision?

How splicing site is determined?

- The borders between introns and exons are marked by specific nucleotide sequences within the pre-mRNAs.
- These sequences delineate where splicing will occur.

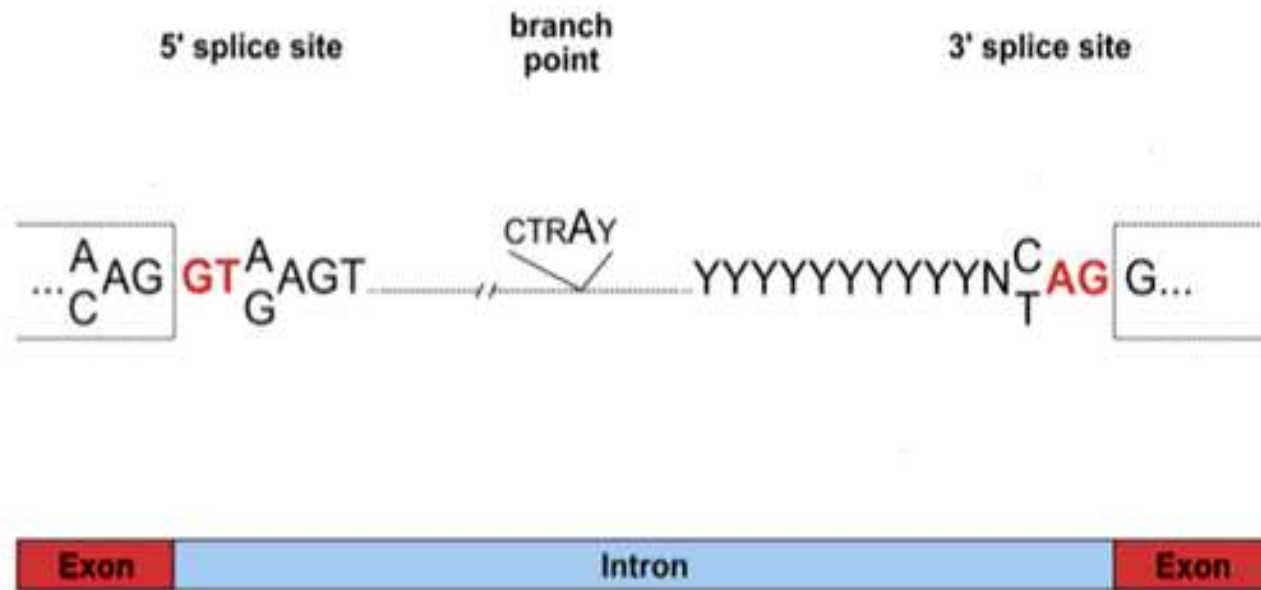
How splicing site is determined?

- Thus the exon – intron boundary - that is, the boundary at the 5' end of the intron — is marked by a sequence called the **5' splice site**.
- The intron – exon boundary at the 3' end of the intron is marked by the **3' splice site**.

How splicing site is determined?

- The 5' and 3' splice sites were sometimes referred to as the **donor** and **acceptor** sites, respectively.

How splicing site is determined?



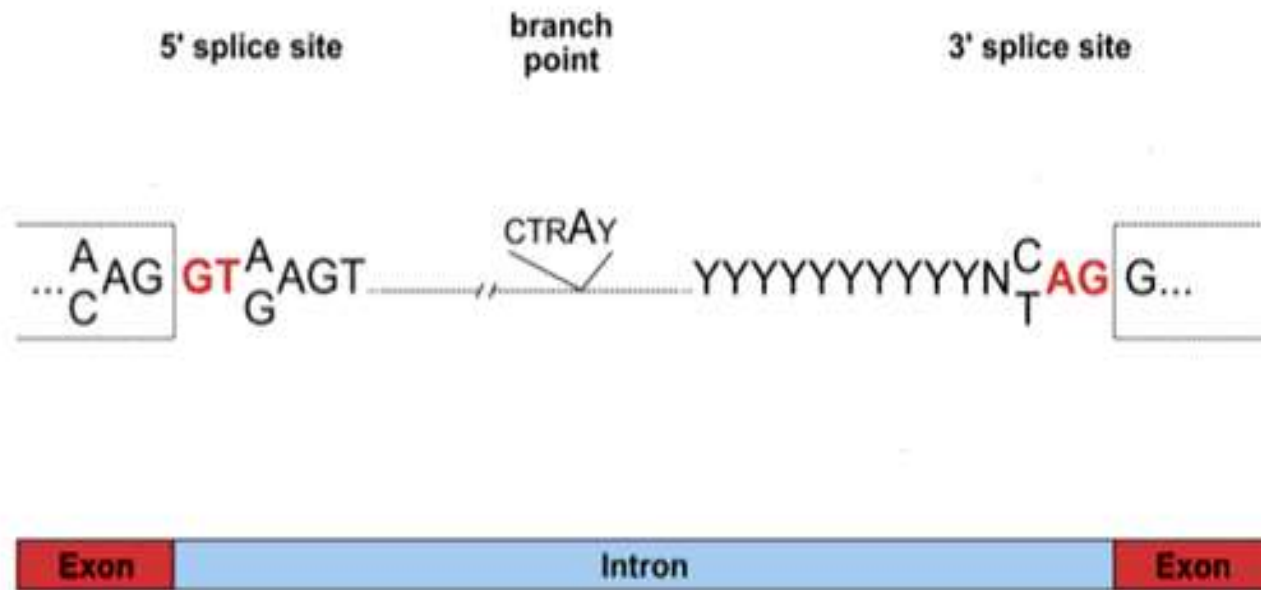
How splicing site is determined?

- The figure shows a third sequence necessary for splicing. This is called the **branchpoint site** (or branchpoint sequence).

How splicing site is determined?

- It is found entirely within the intron, usually close to its 3' end, and is followed by a polypyrimidine tract (Py tract).

How splicing site is determined?



How splicing site is determined?

- The most highly conserved sequences are the GU in the 5' splice site, the AG in the 3' splice site, and the A at the branch site.

How splicing site is determined?

- These highly conserved nucleotide sequences are all found within the intron itself.

How splicing site is determined?

- This is perhaps because the sequence of most exons, in contrast to the introns, is constrained by the need to encode the specific amino acids of the protein product.

Removal of Introns

- An intron is removed through two successive transesterification reactions in which phosphodiester linkages within the pre-mRNA are broken and new ones are formed.

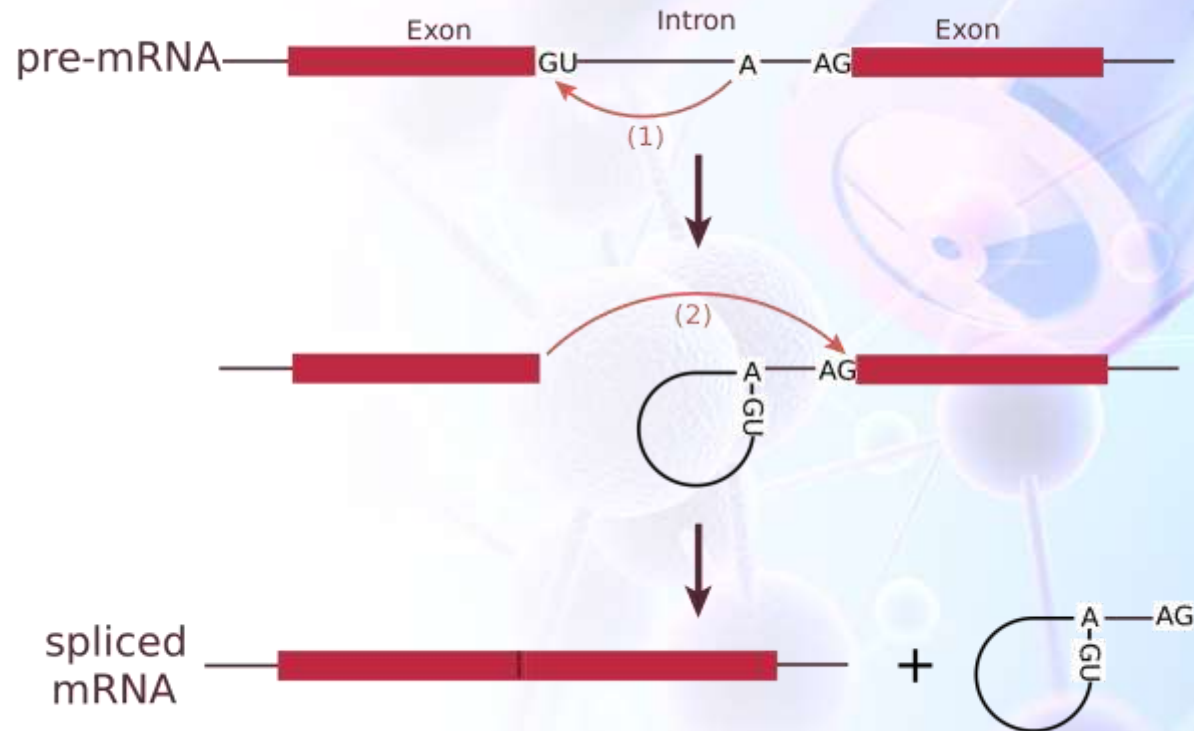
Removal of Introns

- The first reaction is triggered by the 2'-OH of the conserved A at the branch site. This group acts as a nucleophile to attack the phosphoryl group of the conserved G in the 5' splice site.

Removal of Introns

- As a consequence of this first reaction, the phosphodiester bond between the sugar and the phosphate at the 5' junction between the intron and the exon is cleaved.
- The freed 5' end of the intron is joined to the A within the branch site.

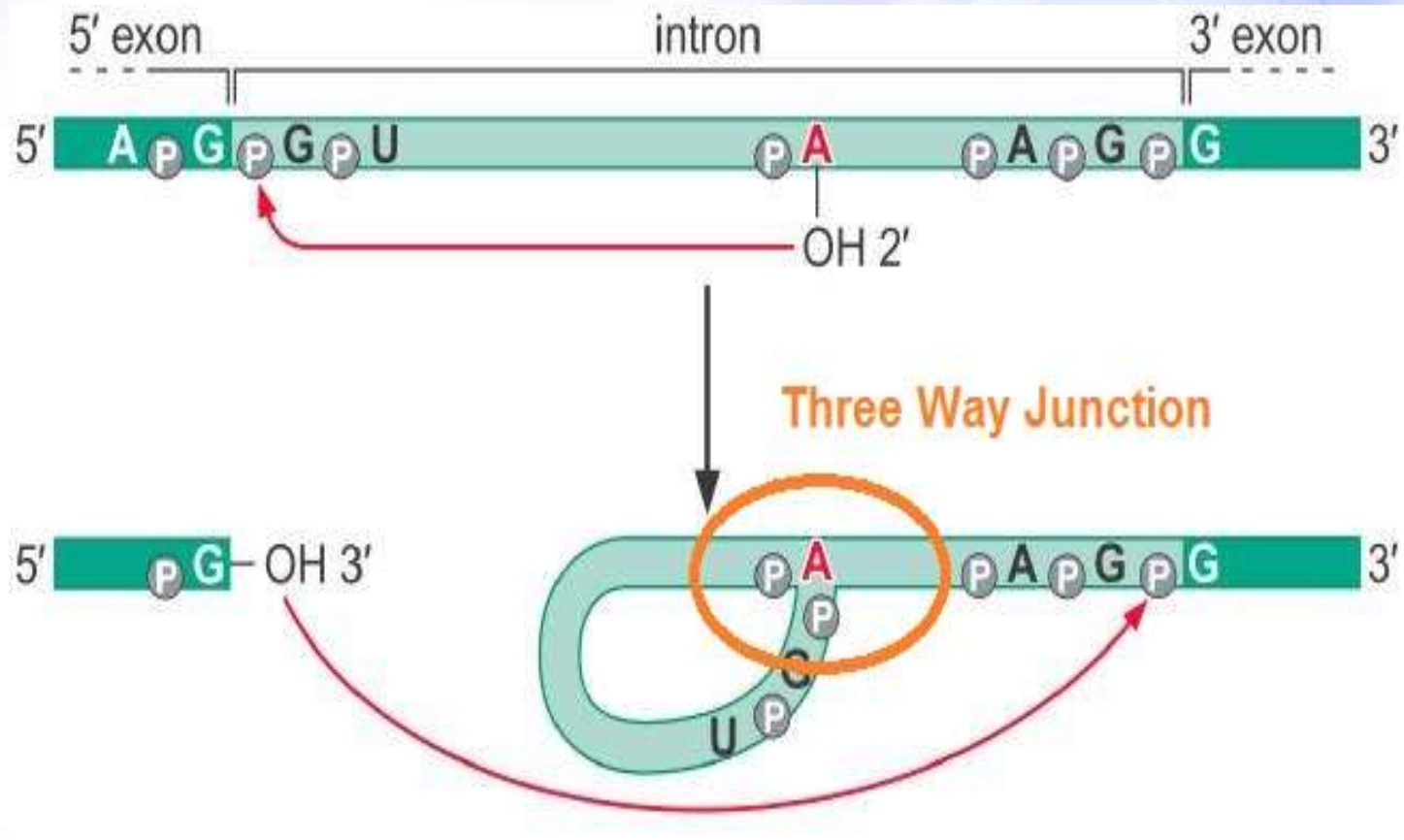
Removal of Introns



Removal of Introns

- Thus, in addition to the 5' and 3' backbone linkages, a third phosphodiester extends from the 2'-OH of that A to create a three-way junction (hence its description as a branchpoint).

Removal of Introns



Removal of Introns

- Note that the 5' exon is a leaving group in the first transesterification reaction.
- In the second reaction, the 5' exon reverses its role and becomes a nucleophile that attacks the phosphoryl group at the 3' splice site.

Removal of Introns

- This second reaction has two consequences.
- First, and most importantly, it joins the 5' and 3' exons; thus, this is the step in which the two coding sequences are actually “spliced” together.

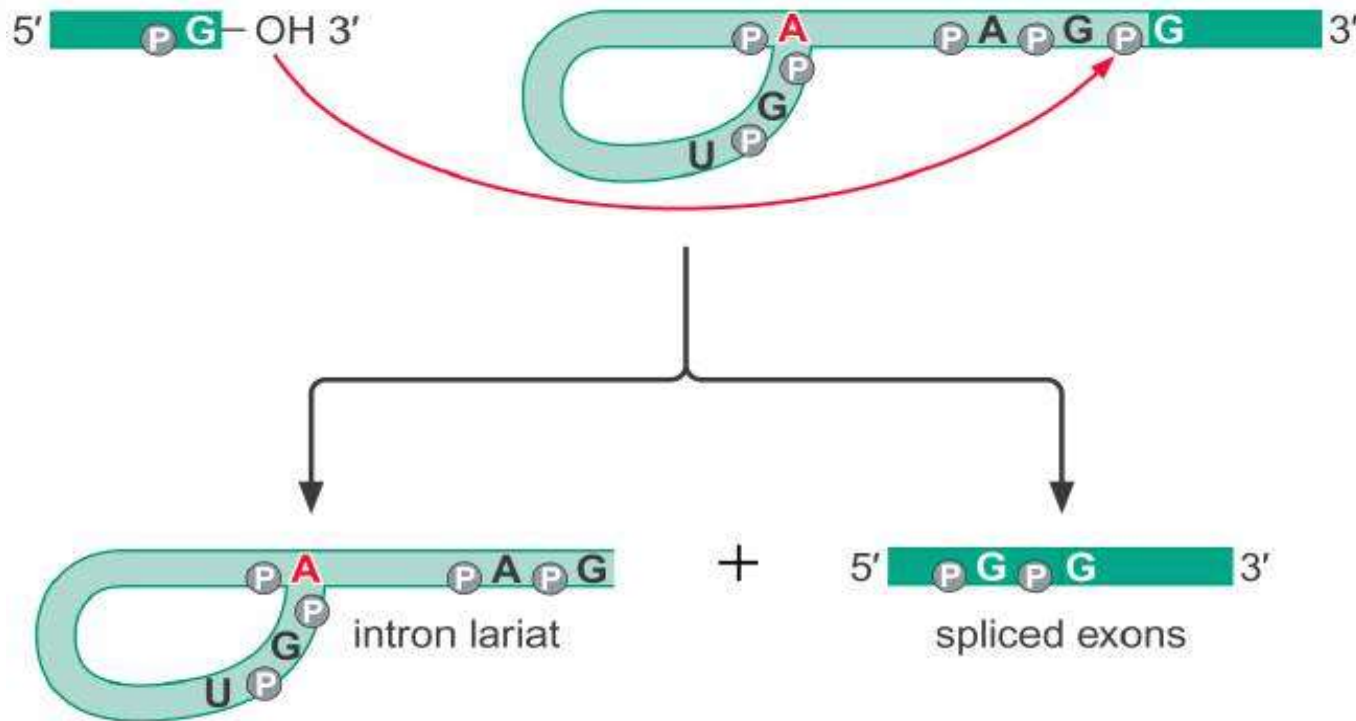
Removal of Introns

- Second, this same reaction liberates the intron, which serves as a leaving group.

Removal of Introns

- Because the 5' end of the intron had been joined to branchpoint A in the first transesterification reaction, the newly liberated intron has the shape of a **Lariat**.

Removal of Introns



Removal of Introns

- In the two reaction steps, there is no net gain in the number of chemical bonds—two phosphodiester bonds are broken, and two new ones are made.

Removal of Introns

- Because it is just a question of shuffling bonds, no energy input is demanded by the chemistry of this process.
- But a large amount of ATP is consumed during the splicing reaction.

Removal of Introns

- This energy is required, to properly assemble and operate the splicing machinery.
- Another point regarding the splicing reaction is direction: what ensures that splicing only goes forward—that is, toward the products.

Removal of Introns

- In principle, the reactions could go in the other direction, and indeed this can be forced to happen under special circumstances.
- But in practice, this does not happen in the cell.

The Spliceosome

- The transesterification reactions are mediated by a huge molecular “machine” called the **Spliceosome**.
- This complex comprises about 150 proteins and five RNAs and is similar in size to a ribosome.

The Spliceosome

- In performing even a single splicing reaction, the spliceosome hydrolyzes several molecules of ATP.

The Spliceosome

- Strikingly, it is believed that many of the functions of the spliceosome are performed by its RNA components rather than the proteins.

The Spliceosome

- Thus, RNAs locate the sequence elements at the intron – exon borders and likely participate in catalysis of the splicing reaction itself.

The Spliceosome

- The five RNAs (U1, U2, U4, U5, and U6) are collectively called **small nuclear RNAs (snRNAs)**. Each of these RNAs is between 100 and 300 nucleotides long in most eukaryotes and is complexed with several proteins.

The Spliceosome

- These RNA – protein complexes are called **small nuclear ribonuclear proteins** (snRNPs - pronounced “snurps”).

The Spliceosome

- The spliceosome is the large complex made up of these snRNPs, but the exact makeup differs at different stages of the splicing reaction.

The Spliceosome

- Different snRNPs come and go at different times, each performing particular functions in the reaction.

The Spliceosome

- There are also many proteins within the spliceosome that are not part of the snRNPs, and others besides that are only loosely bound to the spliceosome.
- The snRNPs have three roles in splicing:-

The Spliceosome

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

The Spliceosome

- To perform these functions, RNA–RNA, RNA–protein, and protein–protein interactions are all important.
- Let us consider some of the RNA– RNA interactions.

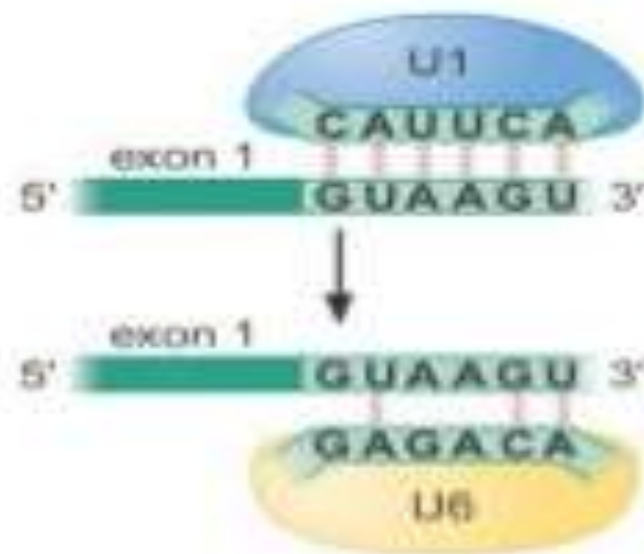
The Spliceosome

- These interactions operate within individual snRNPs, between different snRNPs, and between snRNPs and the pre-mRNA.

The Spliceosome

- The figure below shows the interaction, through complementary base pairing, of the U1 snRNA and the 5' splice site in the pre-mRNA.
- Subsequently in the reaction, that splice site is recognized by the U6 snRNA.

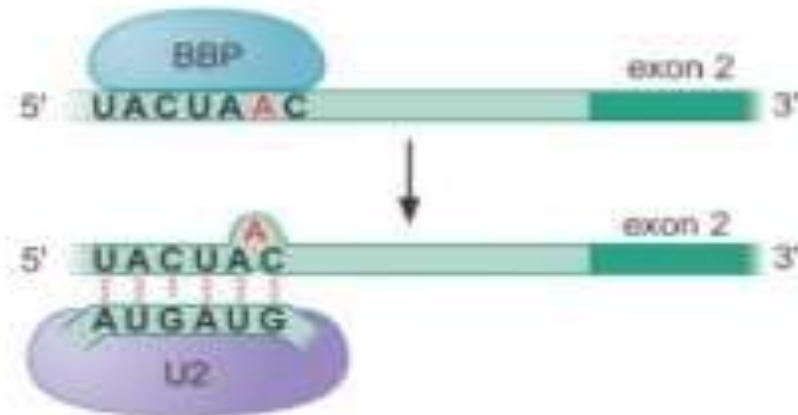
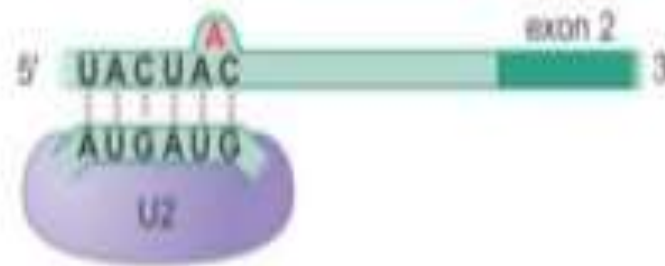
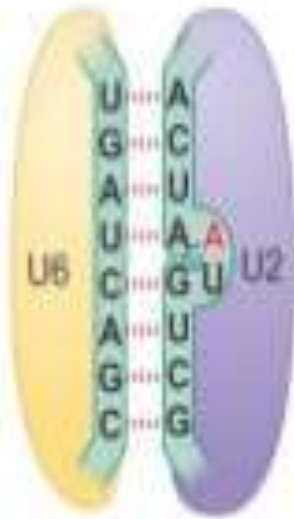
The Spliceosome



The Spliceosome

- In another example, shown below, the branch site is recognized by the U2 snRNA.
- A third example shows an interaction between U2 and U6 snRNAs.
- This brings the 5' splice site and the branch site together.

The Spliceosome



The Spliceosome

- It is these and other similar interactions, and the rearrangements they lead to, that drive the splicing reaction and contribute to its precision.

The Spliceosome

- Some non-snRNPs are also involved in splicing. One example, U2AF (U2 auxiliary factor), recognizes the polypyrimidine (Py) tract/3' splice site.

The Spliceosome

- And then in the initial step of the splicing reaction, it helps another protein, branchpoint-binding protein (BBP), bind to the branch site.
- BBP (also called SF1) is then displaced by the U2 snRNP.

The Spliceosome

- Other proteins involved in the splicing reaction include RNA-annealing factors, which help load snRNPs onto the mRNA, and DEAD-box helicase proteins.

The Spliceosome

- The latter use their ATPase activity to dissociate given RNA - RNA interactions, allowing alternative pairs to form and thereby driving the rearrangements that occur through the splicing reaction.

The Spliceosome

- They are also required to remove spliced mRNA from the spliceosome and trigger spliceosome disassembly.

The Splicing Pathway

- In a splicing pathway, a number of the steps may differ slightly in their order or might even reverse.
- But the pathway reveals the series of events undertaken by the spliceosome to drive the splicing reaction in the cell.

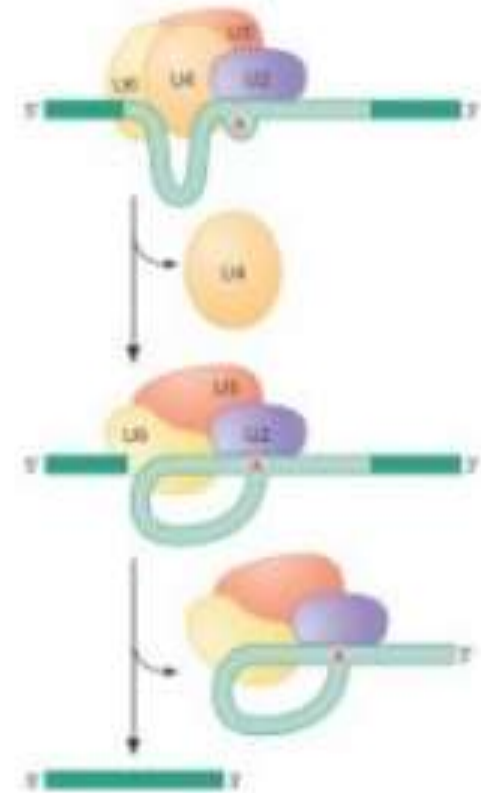
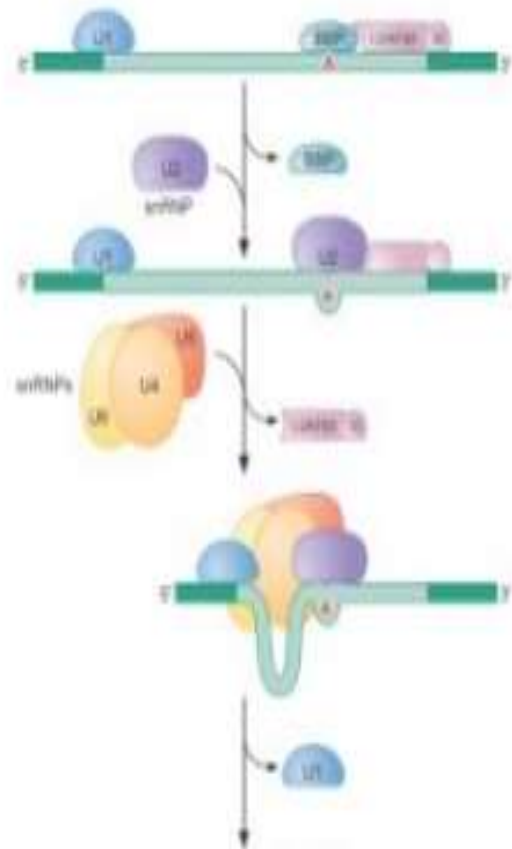
The Splicing Pathway

- Initially, the 5' splice site is recognized by the U1 snRNP.
- U2AF is made up of two subunits, the larger of which (65) binds to the Py tract and the smaller (35) binds to the 3' splice site.

The Splicing Pathway

- The former subunit interacts with BBP (SF1) and helps that protein bind to the branch site. This arrangement of proteins and RNA is called the early (E) complex.

The Splicing Pathway



The Splicing Pathway

- The former subunit interacts with BBP (SF1) and helps that protein bind to the branch site. This arrangement of proteins and RNA is called the **early (E) complex**.

The Splicing Pathway

- U2 snRNP then binds to the branch site, aided by U2AF and displacing BBP (SF1). This arrangement is called the **A complex**.

The Splicing Pathway

- The base pairing between the U2 snRNA and the branch site is such that the branch site A residue is extruded from the resulting stretch of double-helical RNA as a single nucleotide bulge.

The Splicing Pathway

- This A residue is thus unpaired and available to react with the 5' splice site.
- The next step is a rearrangement of the A complex to bring together all three splice sites.

The Splicing Pathway

- This is achieved as follows: the U4 and U6 snRNPs, along with the U5 snRNP, join the complex.
- Together, these three snRNPs are called the **tri-snRNP particle**.

The Splicing Pathway

- Within this particle, the U4 and U6 snRNPs are held together by complementary base pairing between their RNA components, and the U5 snRNP is more loosely associated through protein-protein interactions.

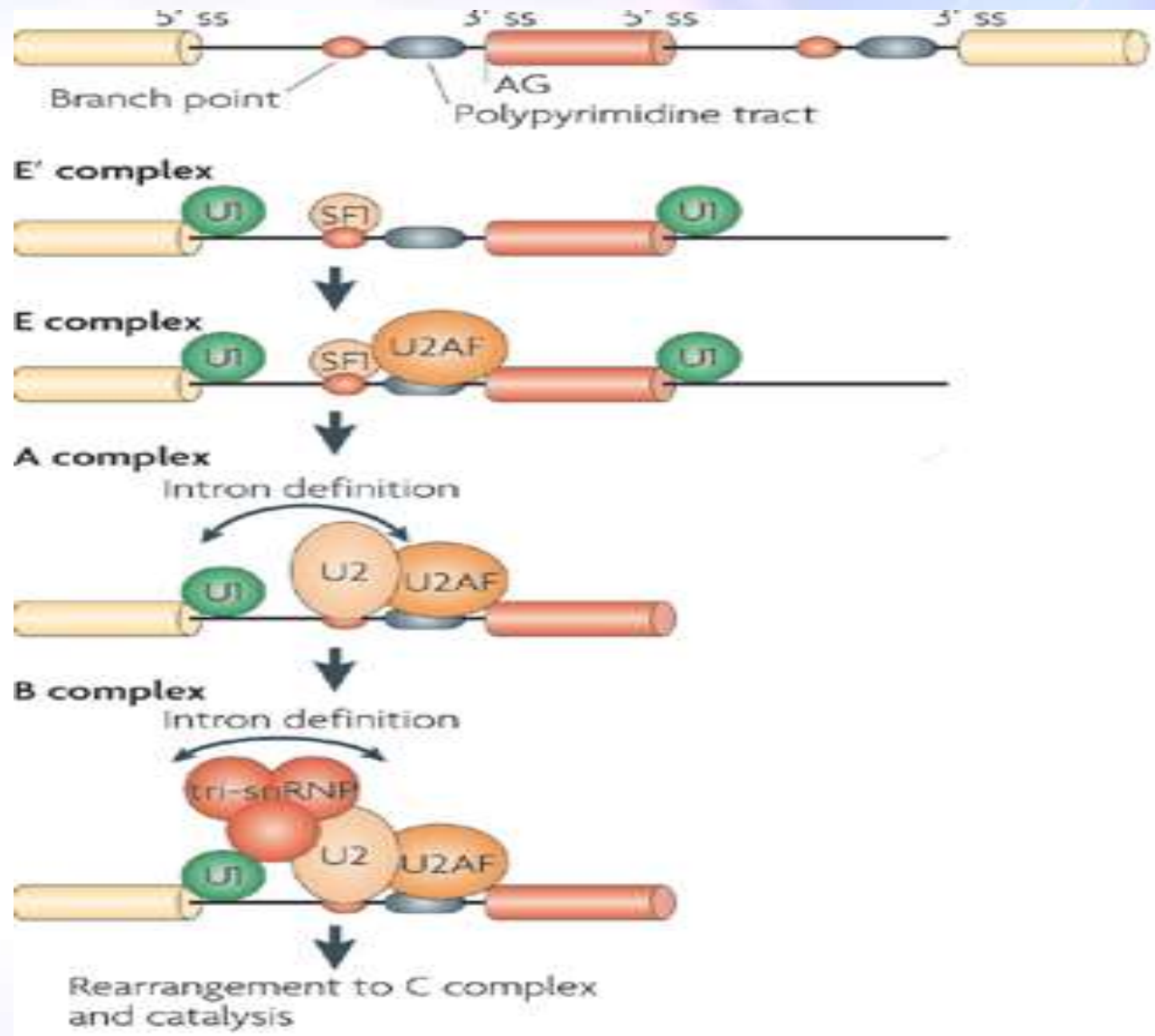
The Splicing Pathway

- With the entry of the tri-snRNP, the A complex is converted into the B complex.
- In the next step, U1 leaves the complex, and U6 replaces it at the 5' splice site.

The Splicing Pathway

- This requires that the base pairing between the U1 snRNA and the pre-mRNA be broken, allowing the U6 RNA to anneal with the same region.
- Those steps complete the assembly pathway.

The Splicing Pathway



The Splicing Pathway

- The next rearrangement triggers catalysis and occurs as follows:-
- U4 is released from the complex, allowing U6 to interact with U2 (through the RNA:RNA base pairing).

The Splicing Pathway

- This arrangement, called the **C complex**, produces the active site. That is, the rearrangement brings together within the spliceosome those components that together form the active site.

The Splicing Pathway

- The same rearrangement also ensures that the substrate RNA is properly positioned to be acted upon.

The Splicing Pathway

- It is striking not only that the active site is primarily formed of RNA, but also that it is only formed at this stage of spliceosome assembly.
- Presumably, this strategy lessens the chance of aberrant splicing.

The Splicing Pathway

- Linking the formation of the active site to the successful completion of earlier steps in spliceosome assembly makes it highly likely that the active site is available only at legitimate splice sites.

The Splicing Pathway

- Formation of the active site juxtaposes the 5' splice site of the pre-mRNA and the branch site, facilitating the first transesterification reaction.

The Splicing Pathway

- The second reaction, between the 5' and 3' splice sites, is aided by the U5 snRNP, which helps to bring the two exons together.

The Splicing Pathway

- The final step involves release of the mRNA product and the snRNPs. The snRNPs are initially still bound to the lariat, but they get recycled after rapid degradation of that piece of RNA.

Self-Splicing Introns

- Self-Splicing Introns Reveal That RNA Can Catalyze RNA Splicing.
- There are total three classes of splicing found in the cells:-
 - Nuclear pre-mRNA
 - Group II introns
 - Group I introns

Self-Splicing Introns

- Thus far, we have dealt only with nuclear pre-mRNA splicing, that mediated by the spliceosome found in all eukaryotes.
- The so-called Group I and Group II are self-splicing introns.

Self-Splicing Introns

- By “self- splicing” we mean that the intron itself folds into a specific conformation within the precursor RNA and catalyzes the chemistry of its own release.

Self-Splicing Introns

- In terms of a practical definition, “self-splicing” refers to introns that can remove themselves from RNAs in the test tube in the absence of any proteins or other RNA molecules.

Self-Splicing Introns

- The self splicing introns are grouped into two classes on the basis of their structure and splicing mechanism.
- Strictly speaking, self-splicing introns are not enzymes because they mediate only one round of RNA processing.

Self-Splicing Introns

- In the case of group II introns, the chemistry of splicing and the RNA intermediates produced are the same as those for nuclear pre-mRNAs.

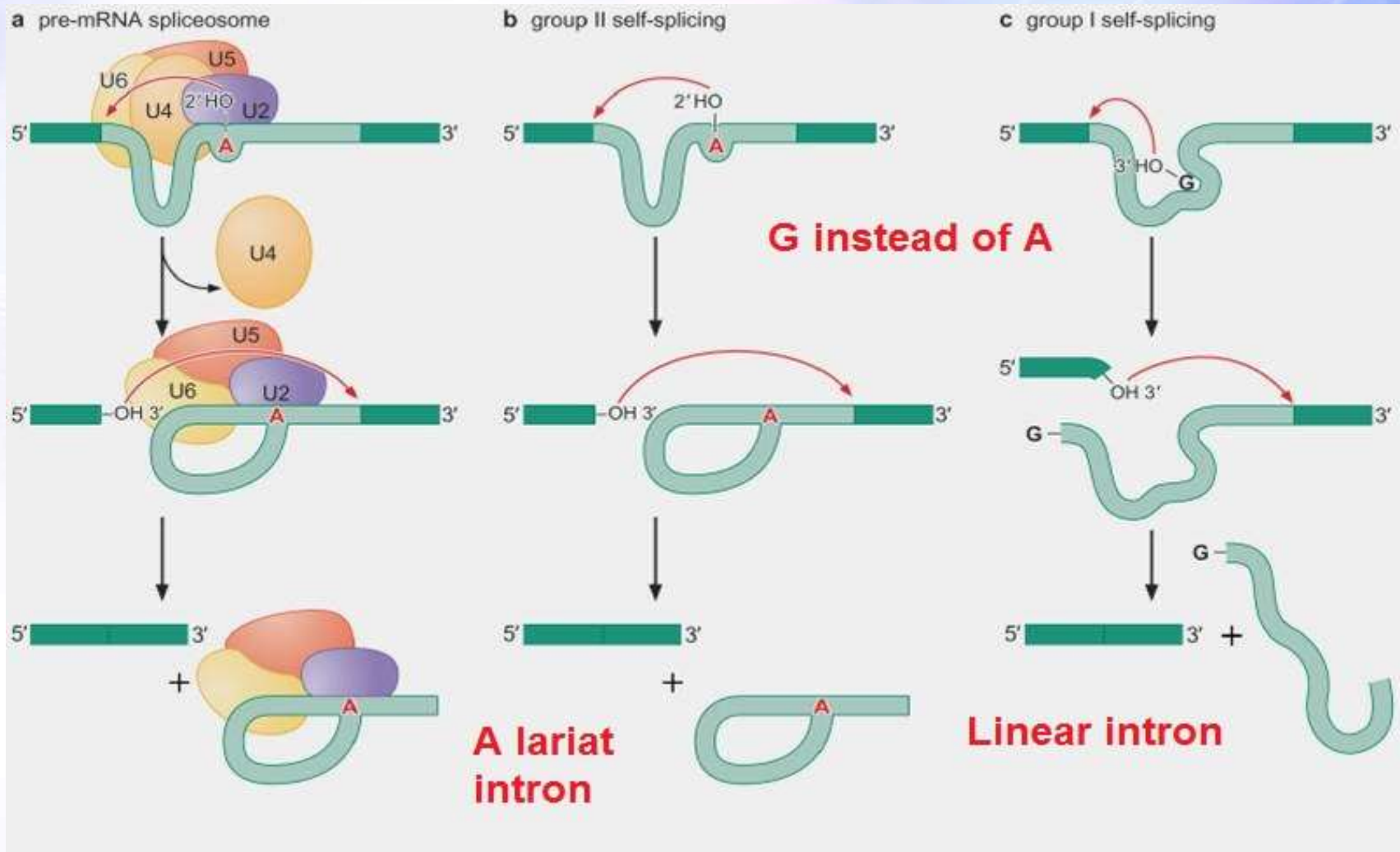
Self-Splicing Introns

- For example, the intron uses an A residue within the branch site to attack the phosphodiester bond at the boundary between its 5' end and the end of the 5' exon - that is, at the 5' splice site.

Self-Splicing Introns

- This reaction produces the branched lariat and is followed by a second reaction in which the newly freed 3'-OH of the exon attacks the 3' splice site, releasing the intron as a lariat and fusing the 3' and 5' exons.

Self-Splicing Introns



Group I Introns

- Group I introns splice by a different pathway.
- Instead of a branchpoint A residue, they use a free G nucleotide or nucleoside.

Group I Introns

- This G species is bound by the RNA, and its 3'-OH group is presented to the 5' splice site.
- The same type of transesterification reaction that leads to the lariat formation fuses the G to the 5' end of the intron.

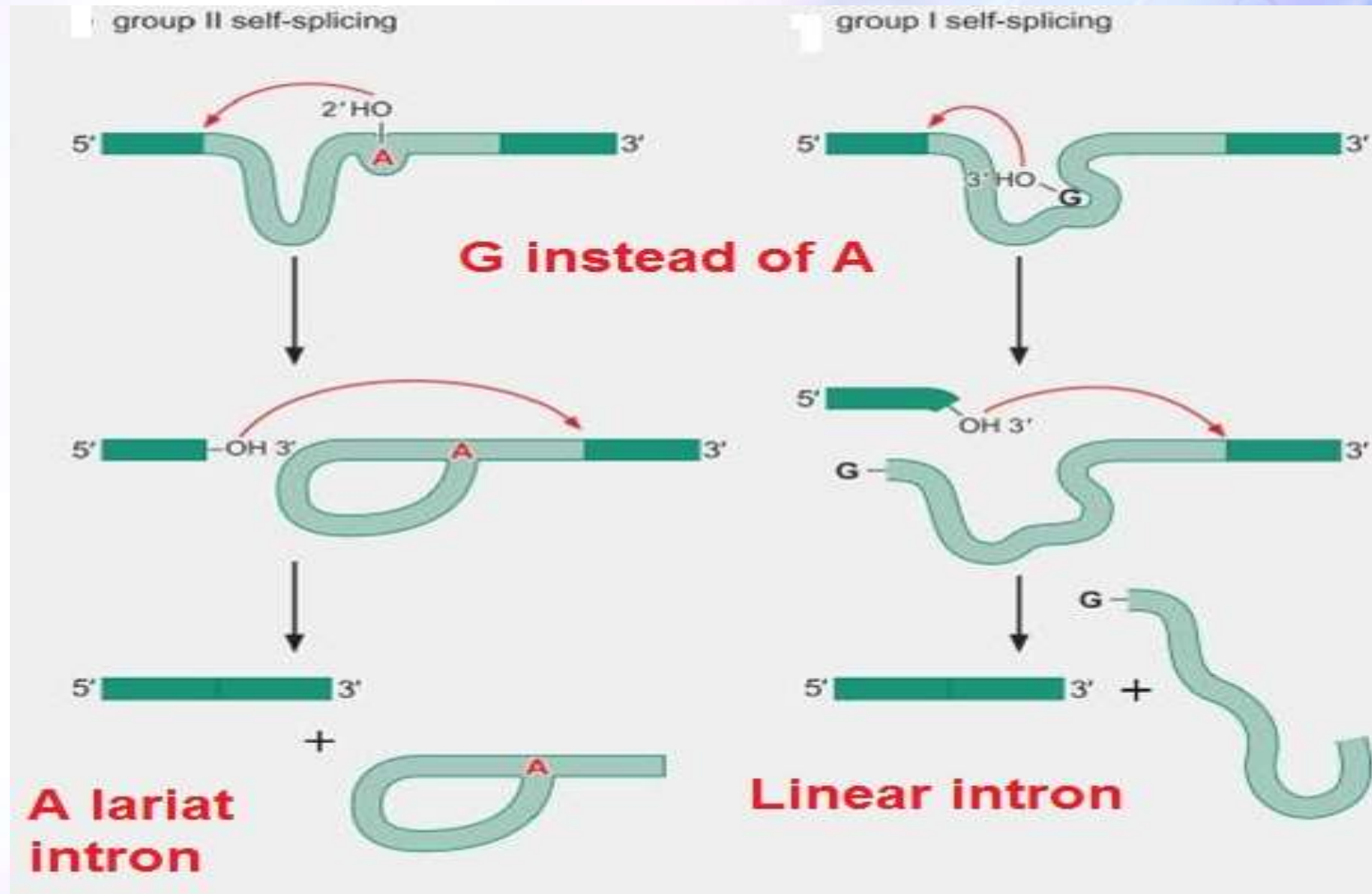
Group I Introns

- The second reaction now proceeds just as it does in the earlier examples: the freed 3' end of the exon attacks the 3' splice site.

Group I Introns

- This fuses the two exons and releases the intron, although, in this case, the intron is linear rather than a lariat structure.

Group I Introns



Group I Introns

- Group I introns, which are smaller than group II introns, share a conserved secondary structure.
- The structure of group I introns includes a binding pocket that will accommodate any guanine nucleotide or nucleoside.

Group I Introns

- In addition, group I introns contain an “internal guide sequence” that base-pairs with the 5' splice site sequence and thereby determines the precise site at which nucleophilic attack by the G nucleotide takes place

Group I Introns

- A typical self-splicing intron is between 400 and 1000 nucleotides long, and, in contrast to introns removed by spliceosomes, much of the sequence of a self-splicing intron is critical for the splicing reaction.

Group I Introns

- This sequence requirement holds because the intron must fold into a precise structure to perform the reaction chemistry.

Group I Introns

- In addition, in vivo, the intron is complexed with several proteins that help stabilize the correct structure — partly by shielding regions of the backbone from each other.

Group I Introns

- Thus, the folding requires certain sections of the RNA backbone to be in close proximity to other sections, and the negative charges provided by the phosphates in those backbone regions would repel each other if not shielded.

Group I Introns

- The similar chemistry seen in self- and spliceosome-mediated splicing is believed to reflect an evolutionary relationship.

Group I Introns

- Perhaps ancestral group II – like self-splicing introns were the starting point for the evolution of modern pre-mRNA splicing.

Group I Introns

- The catalytic functions provided by the RNA were retained, but the requirement for extensive sequence specificity within the intron itself was relieved.

Group I Introns

- In this way, introns had only to retain the minimum of sequence elements required to target splicing to the correct places.

Group I Introns

- The structure of the catalytic region that performs the first transesterification reaction is very similar in the group II intron and the pre-mRNA/snRNP complex.

Group I Introns

- This observation, supports the speculation that early in the evolution of modern organisms, many catalytic functions in the cell were performed by RNAs, and that these functions have been replaced by proteins.

Finding the Splice Sites Correctly

- We have already seen one mechanism that guards against inappropriate splicing: the active site of the spliceosome is only formed on RNA sequences that pass the test of being recognized by multiple elements.

Finding the Splice Sites Correctly

- Thus, for example, the 5' splice site must be recognized initially by the U1 snRNP and then by the U6 snRNP.
- It is unlikely that both would recognize an incorrect sequence.

Finding the Splice Sites Correctly

- Splice-site recognition is prone to two kinds of errors.
- First, splice sites can be skipped, with components bound at, for example, a given 5' splice site pairing with those at a 3' site beyond the correct one.

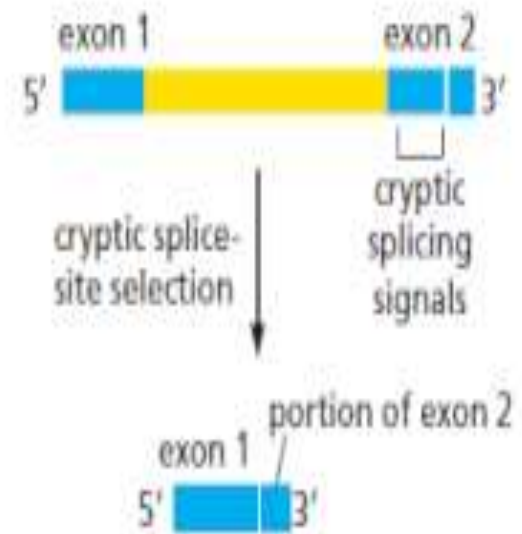
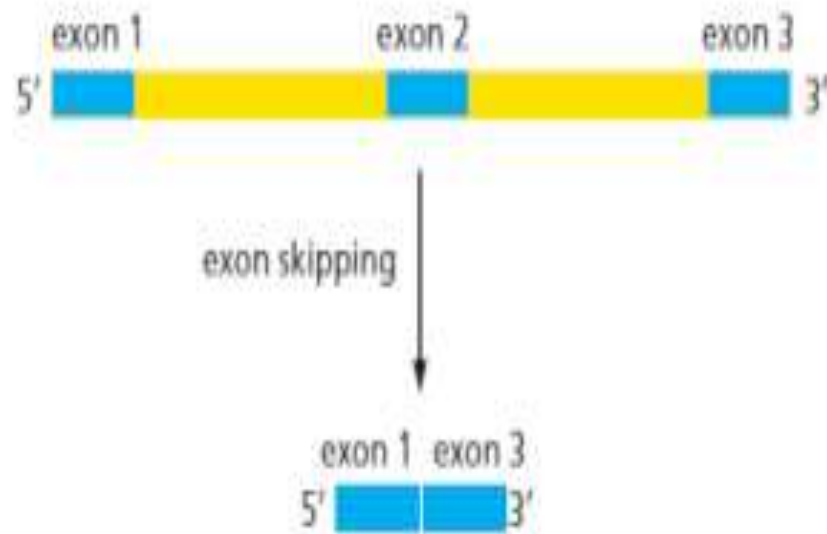
Finding the Splice Sites Correctly

- Second, other sites, close in sequence but not legitimate splice sites, could be mistakenly recognized because the splice site consensus sequences are rather loose.

Finding the Splice Sites Correctly

- Therefore, for example, components at a given 5' splice site might pair with components bound incorrectly at such a “pseudo” 3' splice site.

Finding the Splice Sites Correctly



Finding the Splice Sites Correctly

- Two ways in which the accuracy of splice-site selection can be enhanced are as follows:-

Finding the Splice Sites Correctly

- First, while transcribing a gene to produce the RNA, RNA polymerase II carries with it various proteins with roles in RNA processing.
- These include proteins involved in splicing.

Finding the Splice Sites Correctly

- When a 5' splice site is encountered in the newly synthesized RNA, the factors that recognize that site are transferred from the polymerase carboxy-terminal "tail" onto the RNA.

Finding the Splice Sites Correctly

- Once in place, the 5' splice site components are poised to interact with those other factors that bind to the next 3' splice site to be synthesized.

Finding the Splice Sites Correctly

- Thus, the correct 3' splice site can be recognized before any competing sites further downstream have been transcribed. This co-transcriptional loading process greatly diminishes the likelihood of exon skipping.

Finding the Splice Sites Correctly

- A second mechanism guards against the use of incorrect sites by ensuring that splice sites close to exons are recognized preferentially.

Finding the Splice Sites Correctly

- So-called **SR** (Serine Arginine Rich) proteins bind to the sequences called **exonic splicing enhancers (ESEs)** within the exons.

Finding the Splice Sites Correctly

- SR proteins bound to these sites interact with the components of the splicing machinery, recruiting them to the nearby splice sites.

Finding the Splice Sites Correctly

- In this way, the machinery binds more efficiently to those splice sites than to incorrect sites not close to exons.

Finding the Splice Sites Correctly

- Specifically, the SR proteins recruit the U2AF proteins to the 3' splice site and U1 snRNP to the 5' splice site.
- These factors demarcate the splice sites for the rest of the machinery to assemble correctly.

Finding the Splice Sites Correctly

- SR proteins are essential for splicing.
- They not only ensure the accuracy and efficiency of constitutive splicing but also regulate alternative splicing.

Alternative Spliceosome

- Higher eukaryotes (mammals, plants, etc.) use the major splicing machinery to direct splicing of the majority of their pre-mRNAs.

Alternative Spliceosome

- But in these organisms (unlike in yeast), some pre-mRNAs are spliced by an alternative, low - abundance form of the spliceosome.

Alternative Spliceosome

- This rare form of spliceosome contains some components common to the major spliceosome, but it contains other unique components as well.

Alternative Spliceosome

- Thus, U11 and U12 components of the alternative spliceosome have the same roles in the splicing reaction as U1 and U2 of the major form, but they recognize distinct sequences.

Alternative Spliceosome

- U4 and U6 have equivalent counterparts in both spliceosome forms — although these snRNPs are distinct, they share the same names.

Alternative Spliceosome

- Finally, the identical U5 component is found in both the major and the alternative or so called **minor spliceosome**.

Alternative Spliceosome

- The minor spliceosome recognizes rarely occurring introns having consensus sequences distinct from the sequences of most pre-mRNA introns.

Alternative Spliceosome

- It should be emphasized that although these introns are rare, they are widely distributed; approximately 800 human genes contain at least one minor intron.

Alternative Spliceosome

- Furthermore, mutations in minor snRNAs have recently been found to underlie some rare human genetic diseases.

Alternative Spliceosome

- The minor form of the spliceosome is also known as the AT-AC spliceosome, because the termini of the originally identified rare introns contain AU at the 5' splice site and AC at the 3' site (or AT and AC in DNA).

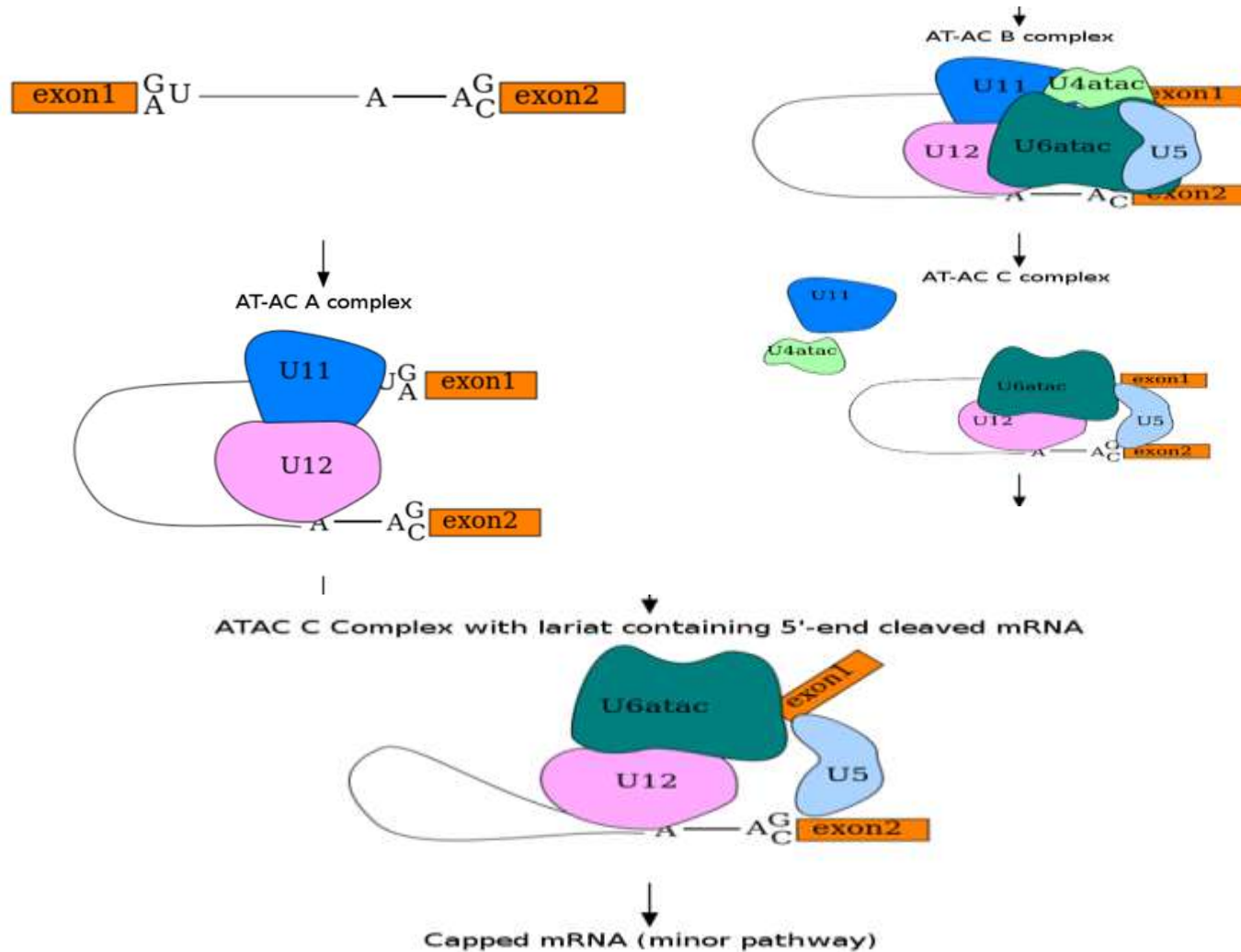
Alternative Spliceosome

- Later it was found that many introns spliced by this pathway have GT-AG termini (like mainstream introns), but otherwise their consensus sequences are distinct from those of the major pathway.

Alternative Spliceosome

- Despite the different splice site and branchsite sequences recognized by the two systems, these major and minor forms of spliceosomes both remove introns using the same chemical pathway.

Alternative Spliceosome



Alternative Spliceosome

- Consistent with this conserved mechanism, the differences in splice-site sequences recognized by these snRNPs are mirrored by complementary differences in the sequences of their snRNAs.

Alternative Spliceosome

- Thus, it is the ability of the snRNAs and splice-site sequences to base-pair that is conserved, not any particular sequence within either.

END

Alternative Splicing

- Many genes in higher eukaryotes encode RNAs that can be spliced in alternative ways to generate two or more different mRNAs and thus different protein products (isoforms).

Alternative Splicing

- It is now believed that at least 40% of *Drosophila* genes and as many as 90% of human genes undergo alternative splicing.

Alternative Splicing

- Many alternatively spliced genes generate only two alternative products, but in some cases,

Alternative Splicing

- the number of potential alternatives that can be generated from a single gene is breathtaking - hundreds (e.g., in the human *Slo* gene) or even many thousands (for the *Drosophila Dscam* gene).

Alternative Splicing

- Alternative splicing is sometimes used as a way of generating diversity, with alternative forms being generated stochastically.

Alternative Splicing

- But in many cases, the process is regulated to ensure that different protein products are made in different cell types or in different conditions.

Alternative Splicing

- For a simple case of alternative splicing, consider the gene for the mammalian muscle protein troponin T.
- The pre mRNA made from this gene that contains five exons.

Alternative Splicing

- This pre-RNA is spliced to form two alternative mature mRNAs, each containing four exons.
- A different exon is eliminated from each of the two mRNAs, thus the two messages have three exons in common.

Alternative Splicing



Alternative Splicing

- Alternative splicing can occur in a number of ways. Thus, in addition to alternative exons, exons can be extended.
- In other cases, exons can be skipped (deliberately), or introns can be retained in the mature message.

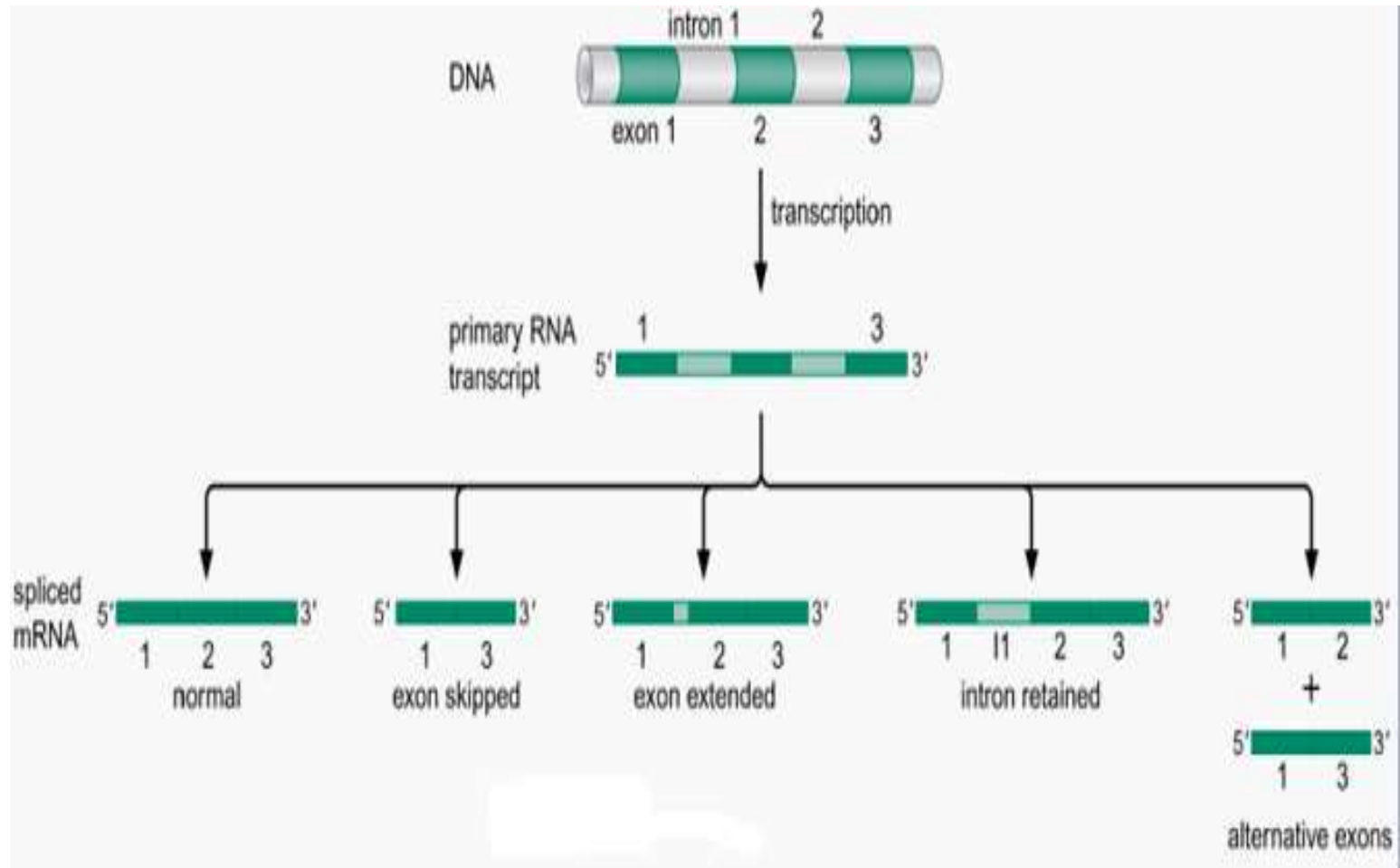
Alternative Splicing

- Some alternative splicing results from transcription of a gene from alternative promoters, allowing one transcript to include a 5' exon not present in the other.

Alternative Splicing

- Similarly, alternative poly-A sites allow 3' terminal exons to be extended or alternative 3' terminal exons to be used in some transcripts of a given gene. There are even cases of alternative *trans* splicing.

Alternative Splicing



Alternative Splicing

- The case of T antigen of the monkey virus SV40 shows an example of an extended exon.
- The T-antigen gene encodes two protein products: the large T antigen (T-ag) and the small t antigen (t-ag).

Alternative Splicing

- The two proteins result from alternative splicing of the pre-mRNAs from the same gene.
- Thus the gene has two exons, and different mature mRNAs result from the use of two different 5' splice sites.

Alternative Splicing

- In the mRNA encoding T-ag, exon 1 is spliced directly to exon 2, deleting the intron that lies between.
- The mRNA for t-ag, on the other hand, is formed using the alternative 5' splice site within the intron.

Alternative Splicing

- Thus, in this case, the mRNA includes some of the intron as well.
- The reason this larger message encodes the smaller protein is because there is an in-frame stop codon within the region of the intron retained in this mRNA.

Alternative Splicing

- Both forms of T antigens are made in a cell infected by SV40 but have different functions.
- Large T induces transformation and cell cycle re-entry, whereas small t blocks the apoptotic response of cells forced down that path.

Alternative Splicing

- The ratio of the two forms produced differs depending on the level of the splicing regulator SF2/ASF.

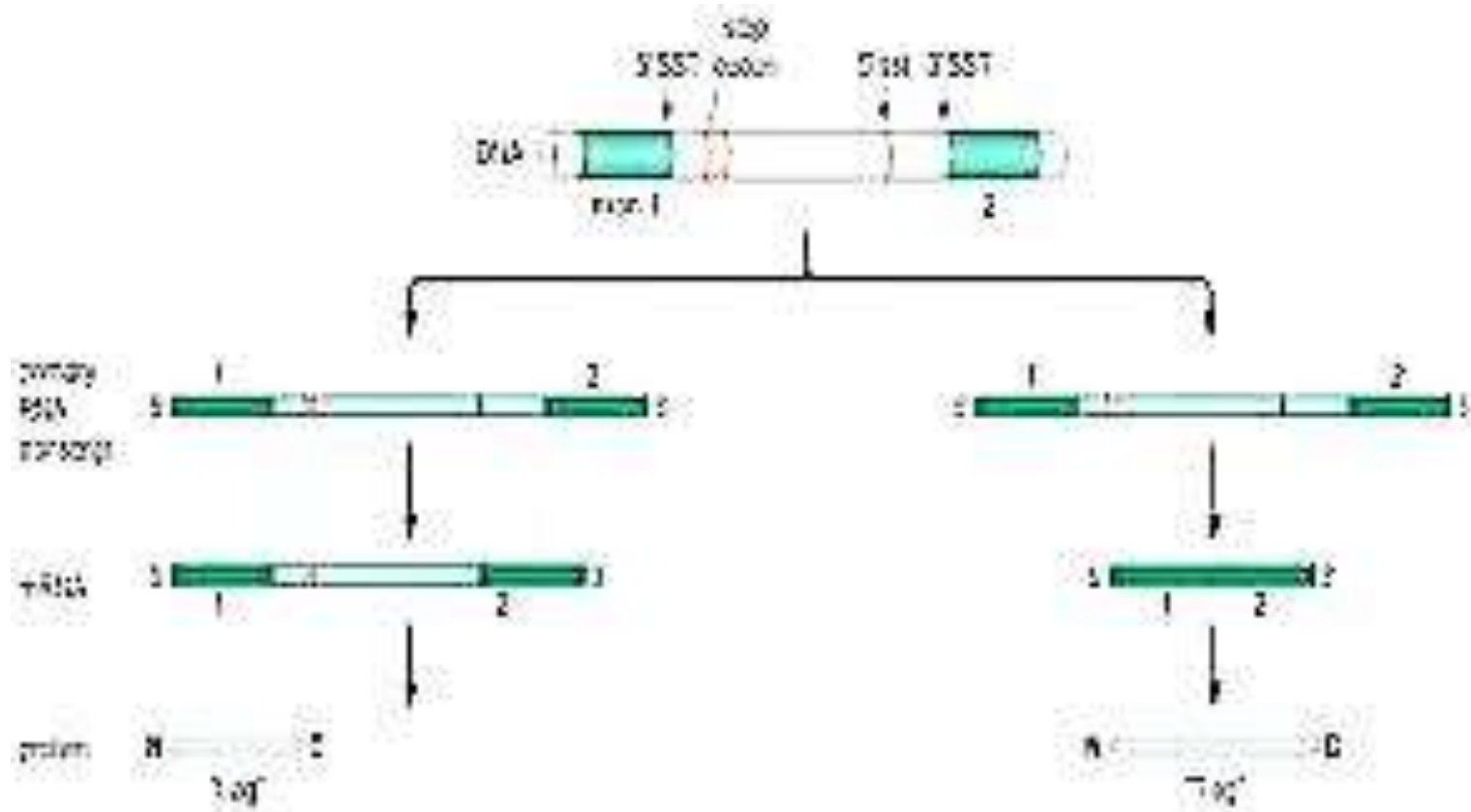
Alternative Splicing

- When present at high levels, this protein directs the machinery to favor use of the 5' splice site that generates more of the t-ag mRNA.

Alternative Splicing

- SF2/ASF is an SR protein and, when abundant, presumably binds sites within exon 2 and helps the spliceosome assemble there.

Alternative Splicing



Alternative Splicing

- In genome-wide studies, the most commonly seen forms of alternative splicing are cases in which complete exons are included or excluded from the mature message. Such exons are often called **cassette exons**.

Alternative Splicing

- In ~10% of cases, cassette exons come in pairs, only one of which is included in the spliced message.
- In these cases, there must be mechanisms that ensure that the exons are spliced in a mutually exclusive fashion.

END

Alternative Splicing

- Many genes in higher eukaryotes encode RNAs that can be spliced in alternative ways to generate two or more different mRNAs and thus different protein products (isoforms).

Alternative Splicing

- It is now believed that at least 40% of *Drosophila* genes and as many as 90% of human genes undergo alternative splicing.

Alternative Splicing

- Many alternatively spliced genes generate only two alternative products, but in some cases,

Alternative Splicing

- the number of potential alternatives that can be generated from a single gene is breathtaking - hundreds (e.g., in the human *Slo* gene) or even many thousands (for the *Drosophila Dscam* gene).

Alternative Splicing

- Alternative splicing is sometimes used as a way of generating diversity, with alternative forms being generated stochastically.

Alternative Splicing

- But in many cases, the process is regulated to ensure that different protein products are made in different cell types or in different conditions.

Alternative Splicing

- For a simple case of alternative splicing, consider the gene for the mammalian muscle protein troponin T.
- The pre mRNA made from this gene that contains five exons.

Alternative Splicing

- This pre-RNA is spliced to form two alternative mature mRNAs, each containing four exons.
- A different exon is eliminated from each of the two mRNAs, thus the two messages have three exons in common.

Alternative Splicing



Alternative Splicing

- Alternative splicing can occur in a number of ways. Thus, in addition to alternative exons, exons can be extended.
- In other cases, exons can be skipped (deliberately), or introns can be retained in the mature message.

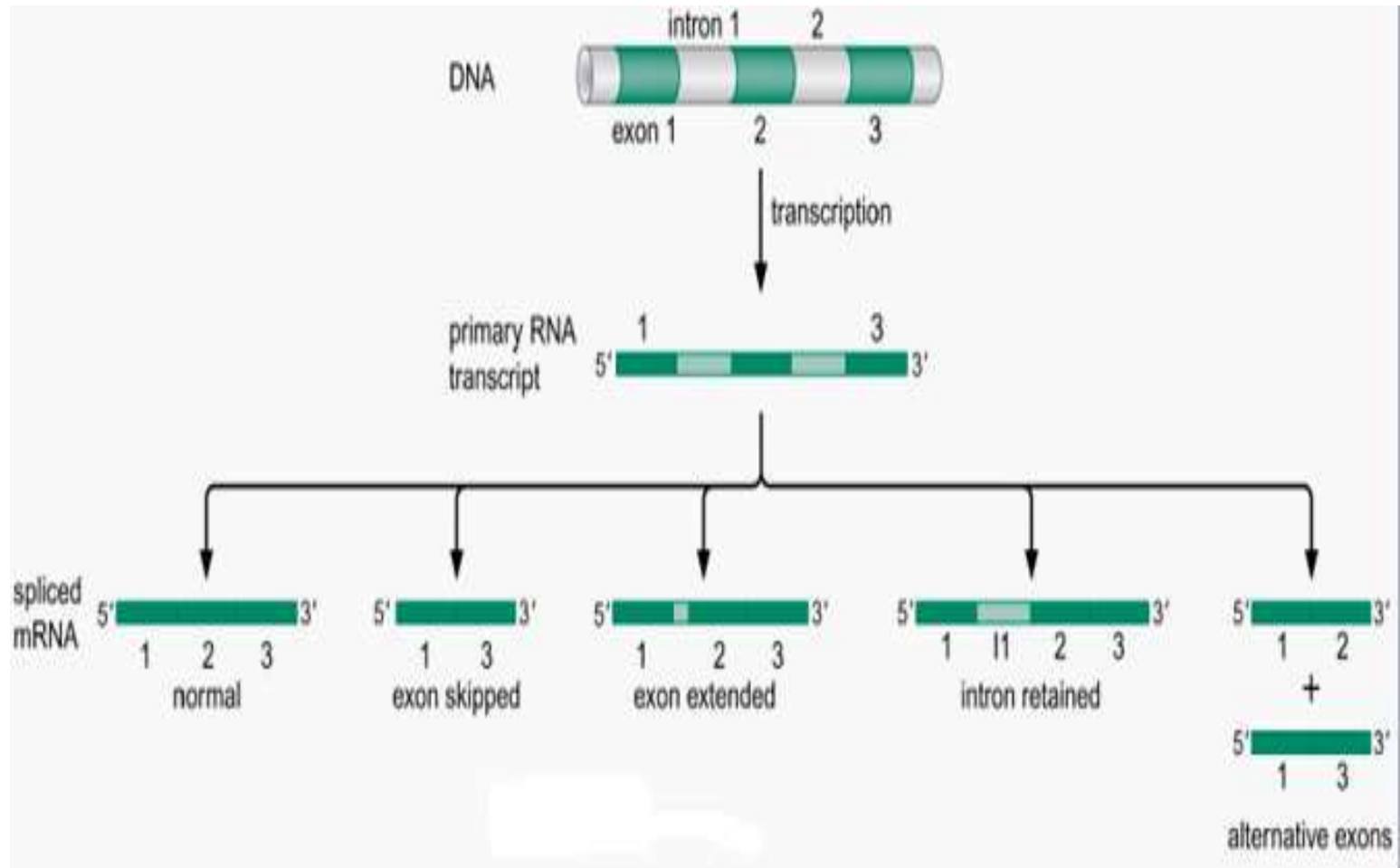
Alternative Splicing

- Some alternative splicing results from transcription of a gene from alternative promoters, allowing one transcript to include a 5' exon not present in the other.

Alternative Splicing

- Similarly, alternative poly-A sites allow 3' terminal exons to be extended or alternative 3' terminal exons to be used in some transcripts of a given gene. There are even cases of alternative *trans* splicing.

Alternative Splicing



Alternative Splicing

- The case of T antigen of the monkey virus SV40 shows an example of an extended exon.
- The T-antigen gene encodes two protein products: the large T antigen (T-ag) and the small t antigen (t-ag).

Alternative Splicing

- The two proteins result from alternative splicing of the pre-mRNAs from the same gene.
- Thus the gene has two exons, and different mature mRNAs result from the use of two different 5' splice sites.

Alternative Splicing

- In the mRNA encoding T-ag, exon 1 is spliced directly to exon 2, deleting the intron that lies between.
- The mRNA for t-ag, on the other hand, is formed using the alternative 5' splice site within the intron.

Alternative Splicing

- Thus, in this case, the mRNA includes some of the intron as well.
- The reason this larger message encodes the smaller protein is because there is an in-frame stop codon within the region of the intron retained in this mRNA.

Alternative Splicing

- Both forms of T antigens are made in a cell infected by SV40 but have different functions.
- Large T induces transformation and cell cycle re-entry, whereas small t blocks the apoptotic response of cells forced down that path.

Alternative Splicing

- The ratio of the two forms produced differs depending on the level of the splicing regulator SF2/ASF.

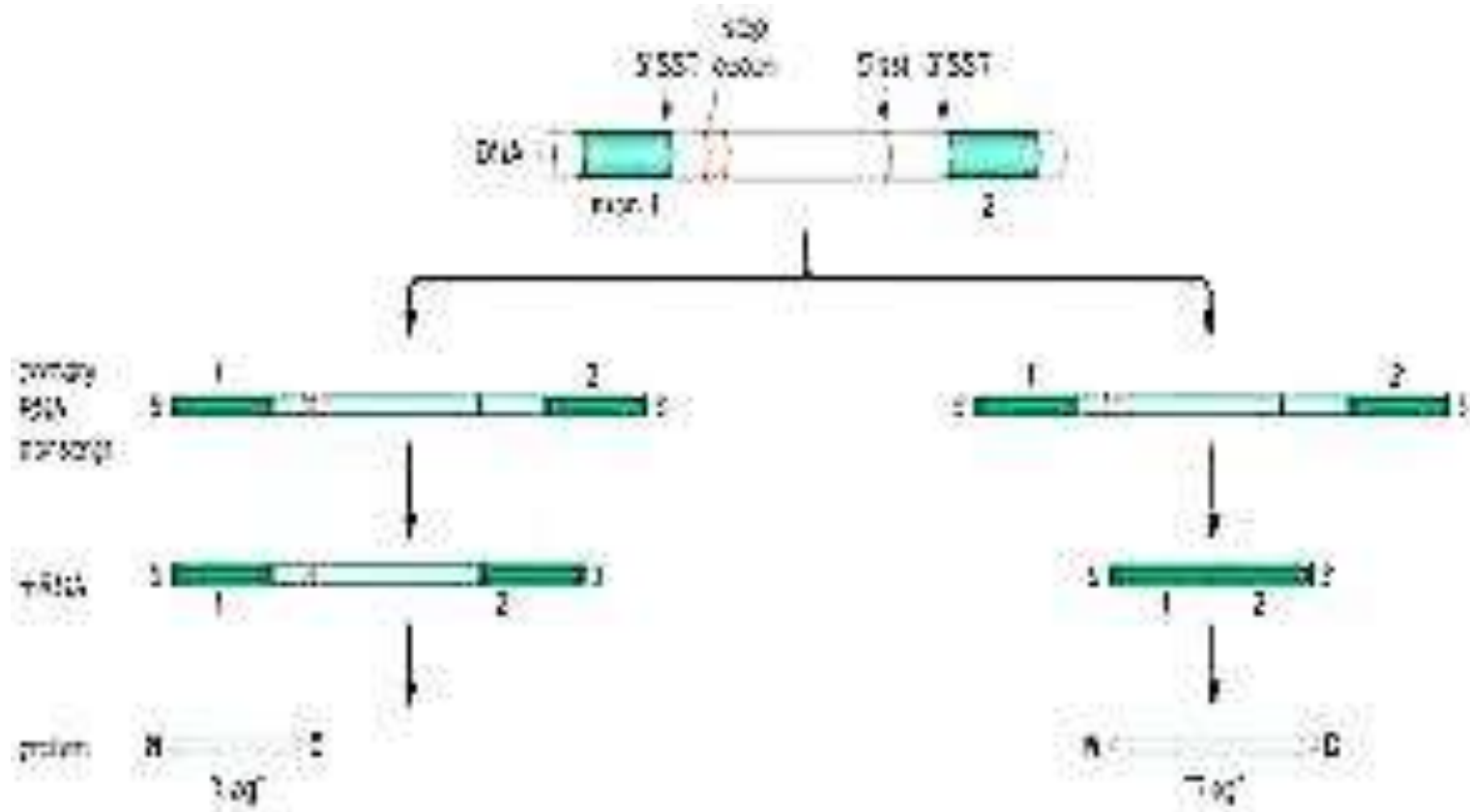
Alternative Splicing

- When present at high levels, this protein directs the machinery to favor use of the 5' splice site that generates more of the t-ag mRNA.

Alternative Splicing

- SF2/ASF is an SR protein and, when abundant, presumably binds sites within exon 2 and helps the spliceosome assemble there.

Alternative Splicing



Alternative Splicing

- In genome-wide studies, the most commonly seen forms of alternative splicing are cases in which complete exons are included or excluded from the mature message. Such exons are often called **cassette exons**.

Alternative Splicing

- In ~10% of cases, cassette exons come in pairs, only one of which is included in the spliced message.
- In these cases, there must be mechanisms that ensure that the exons are spliced in a mutually exclusive fashion.

END

Mutually Exclusive Splicing

- There are several mechanisms to ensure that selection of alternative exons is mutually exclusive — that is, that when one is chosen, the other is not.

Mutually Exclusive Splicing

- ***Steric Hindrance***
- Consider two alternative exons separated by an intron. If the splice sites within the intron are too close together, splicing factors cannot bind to both sites at the same time.

Mutually Exclusive Splicing

- Thus the binding of U1 snRNP to the 5' splice site of the intron between two alternative exons (exons 2 and 3) prevents the binding of U2 snRNP to the branchpoint within that same intron.

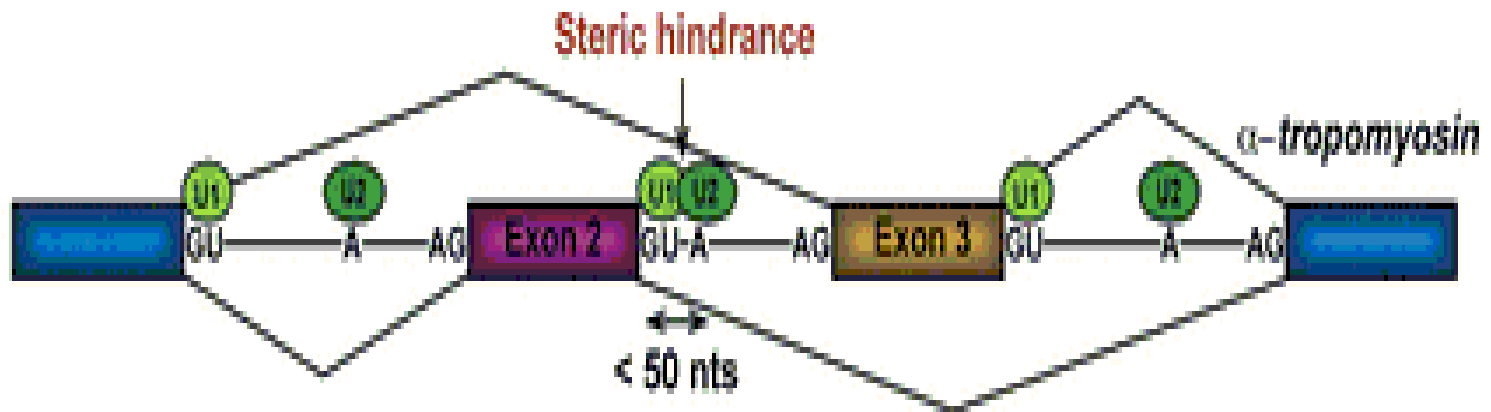
Mutually Exclusive Splicing

- Alternatively, binding of U2 snRNP to the branchpoint excludes use of the 5' splice site. The splicing of exons 3 and 4 of α -troponin is made mutually exclusive by this mechanism.

Mutually Exclusive Splicing

- This arrangement can arise through the relative positions of the splice sites within an intron or because the intron is simply too small to work; in *Drosophila*, any intron under 59 nucleotides falls into that category.

Mutually Exclusive Splicing



Mutually Exclusive Splicing

- *Combinations of Major and Minor Splice Sites*
- The minor spliceosome recognizes splice sites distinct from those recognized by the major spliceosome.

Mutually Exclusive Splicing

- Neither spliceosome can remove an intron that contains a combination of sites (i.e., a 5' splice site of one type and a 3' of the other).

Mutually Exclusive Splicing

- Thus, by judicious arrangement of 5' and 3' splice sites recognized by these alternative spliceosomes, mutual exclusion can be achieved.
- The human *JNK1* gene is an example of this.

Mutually Exclusive Splicing

- *Nonsense-Mediated Decay*
- Rather than forcing the splicing machinery to splice in a mutually exclusive fashion, this mechanism instead ensures that only messages that have one or another exon survive.

Mutually Exclusive Splicing

- In other words, although not ensuring mutually exclusive splicing, the consequences of this mechanism amount to the same thing.

Mutually Exclusive Splicing

- If the stop codon lies downstream (3') of all the exon-exon boundaries, the mRNA is spared from nonsense-mediated decay.

Mutually Exclusive Splicing

- If, on the other hand, a stop codon is located upstream (5') to an exon-exon boundary, the mRNA is degraded.

Mutually Exclusive Splicing

- Translating ribosomes, in conjunction with other surveillance proteins, assess this relationship for each individual mRNA.

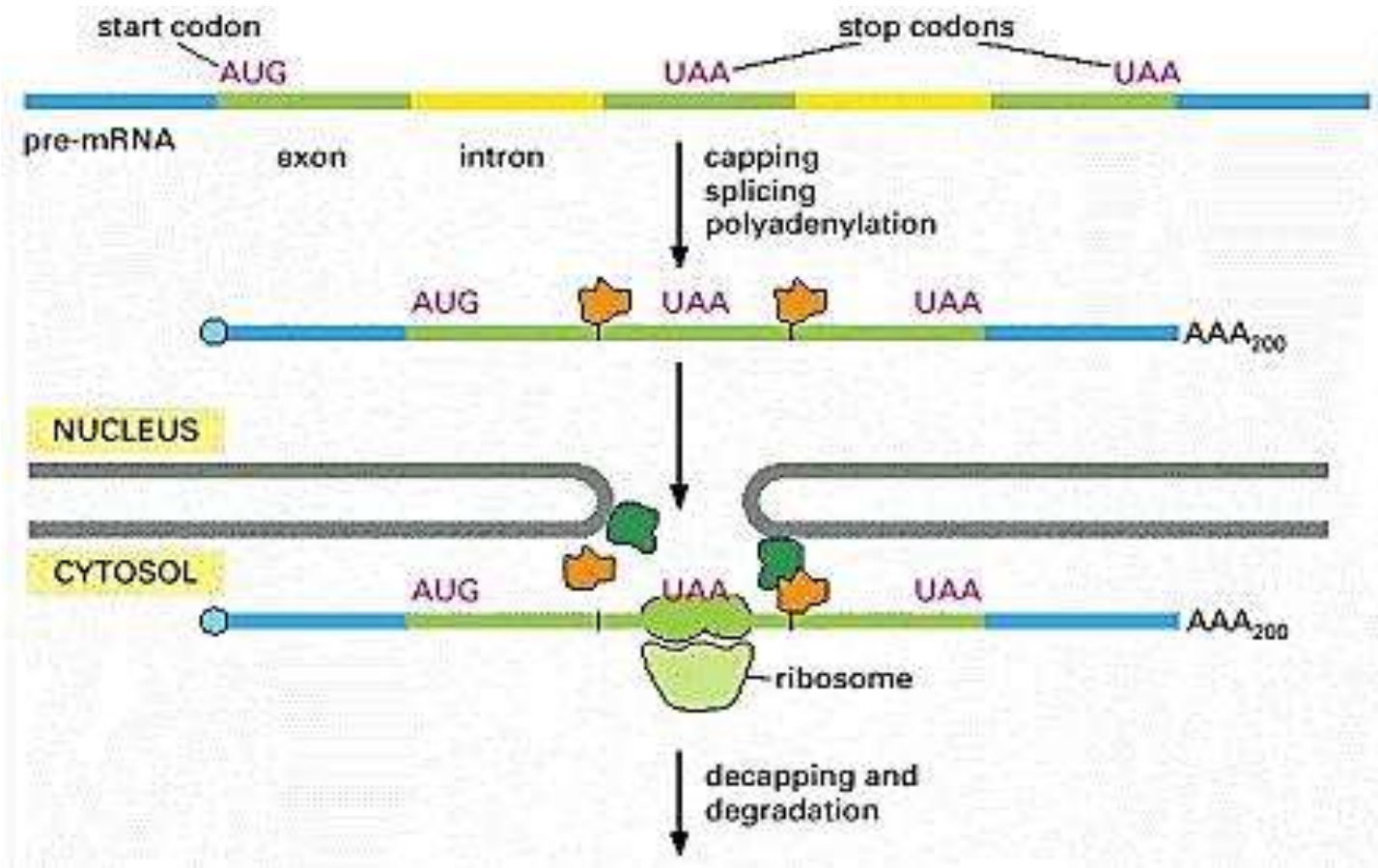
Mutually Exclusive Splicing

- Exactly how this is accomplished is not understood in detail, but it is easy to understand why ribosomes must play a part.

Mutually Exclusive Splicing

- Only in-frame termination codons trigger nonsense-mediated decay, and it is the relationship between the ribosome and the mRNA that defines the reading frame.

Mutually Exclusive Splicing



Regulation by Activators and Repressors

- Proteins that regulate splicing bind to specific sites called **exonic (or intronic) splicing enhancers** (ESE or ISE) or **silencers** (ESS or ISS).

Regulation by Activators and Repressors

- The former enhance, and the latter repress, splicing at nearby splice sites.
- You have already encountered enhancers and the SR proteins that bind to them.

Regulation by Activators and Repressors

- Indeed, these elements and proteins are important in directing the splicing machinery to many exons, even when alternative splicing is not involved.

Regulation by Activators and Repressors

- In addition, in the example of T-antigen splicing described before, it was an SR protein that ensured that alternative splicing occurred.

Regulation by Activators and Repressors

- But this protein family - which is large and diverse – has specific roles in regulated alternative splicing as well, directing the splicing machinery to different splice sites under different conditions.

Regulation by Activators and Repressors

- Thus, the presence or activity of a given SR protein can determine whether a particular splice site is used in a particular cell type or at a particular stage of the development.

Regulation by Activators and Repressors

- The SR proteins bind RNA using one domain—for example, the well characterized RNA-recognition motif (RRM).
- Each SR protein has another domain, rich in arginine and serine, called an **RS domain**.

Regulation by Activators and Repressors

- The RS domain, found at the carboxy-terminal end of the protein, mediates interactions between the SR protein and proteins within the splicing machinery, recruiting that machinery to a nearby splice site.

Regulation by Activators and Repressors

- This activator regulates the alternative splicing of a set of pre-mRNAs in the fly ovary.
- It works by binding to sites near the 3' splice site of specific exons in those pre-mRNAs and recruiting the U2AF splicing factor.

Regulation by Activators and Repressors

- Most silencers are recognized by members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family.
- These bind RNA but lack the RS domains and thus cannot recruit the splicing machinery.

Regulation by Activators and Repressors

- Instead, by blocking specific splice sites, they repress the use of those sites.
- In case of *Dscam*, for example, Hrp36 inhibits inclusion of exon 6 variants in the mRNA.

Regulation by Activators and Repressors

- Another example is hnRNPA1, which binds to an exonic silencer element within an exon of the HIV, *tat* pre- RNA and represses the inclusion of that exon in the final mRNA.

Regulation by Activators and Repressors

- By binding to its site, the repressor blocks binding of the activator SC35 (an SR protein) to a nearby enhancer element.

Regulation by Activators and Repressors

- In this case, blocking is not direct - the two binding sites do not overlap - but hnRNPA1 promotes cooperative binding of additional molecules of hnRNPA1 to adjacent sequences, spreading over the enhancer site.

Regulation by Activators and Repressors

- When present, another SR protein (SF2/ASF) can overcome this repression because it has a higher affinity for the enhancer sequence than does SC35 and therefore displaces the repressors bound there.

Regulation by Activators and Repressors

- Another mammalian splicing repressor is the hnRNPI protein.
- In some cases this protein blocks the binding of the basic splicing machinery by binding directly to the Py tract (hnRNPI is also called the polypyrimidine tract-binding protein).

Regulation by Activators and Repressors

- In other cases it excludes a given exon from the mature mRNA by binding to sequences that flank that exon.

Regulation by Activators and Repressors

- This exclusion occurs either because molecules of hnRNPI at each end of the exon interact and loop out the exon, which is then passed over by the spliceosome.

Regulation by Activators and Repressors

- or because the molecules of hnRNPI at each end bind cooperatively with other molecules of hnRNPI, coating the RNA across the whole exon. This too would render the exon invisible to the splicing machinery.

Regulation by Activators and Repressors

- We have emphasized alternative splicing as a way in which multiple protein products can be produced from a single gene. These different proteins are called **isoforms**.

Regulation by Activators and Repressors

- They can have similar functions, distinct functions, or even antagonistic functions.
- But even some genes that encode only a single functional protein show alternative splicing.

Regulation by Activators and Repressors

- In those cases, alternative splicing is used simply as a way of switching expression of the gene on and off.
- This is achieved in two ways.

Regulation by Activators and Repressors

- Most straight forwardly, an exon contains a stop codon, and, when incorporated into mRNA, this prematurely terminates translation generating a truncated polypeptide.

Regulation by Activators and Repressors

- Typically, such an incomplete protein is non-functional and rapidly degraded.
- Alternative splicing determines whether or not the exon with the stop codon is included in a given mRNA.

Regulation by Activators and Repressors

- The second way alternative splicing can be used as an on/off switch is by regulating the use of an intron, which, when retained in the mRNA, ensures that species is not transported out of the nucleus and so is never translated.

END

Regulation by Activators and Repressors

- Proteins that regulate splicing bind to specific sites called **exonic (or intronic) splicing enhancers** (ESE or ISE) or **silencers** (ESS or ISS).

Regulation by Activators and Repressors

- The former enhance, and the latter repress, splicing at nearby splice sites.
- You have already encountered enhancers and the SR proteins that bind to them.

Regulation by Activators and Repressors

- Indeed, these elements and proteins are important in directing the splicing machinery to many exons, even when alternative splicing is not involved.

Regulation by Activators and Repressors

- In addition, in the example of T-antigen splicing described before, it was an SR protein that ensured that alternative splicing occurred.

Regulation by Activators and Repressors

- But this protein family - which is large and diverse – has specific roles in regulated alternative splicing as well, directing the splicing machinery to different splice sites under different conditions.

Regulation by Activators and Repressors

- Thus, the presence or activity of a given SR protein can determine whether a particular splice site is used in a particular cell type or at a particular stage of the development.

Regulation by Activators and Repressors

- The SR proteins bind RNA using one domain—for example, the well characterized RNA-recognition motif (RRM).
- Each SR protein has another domain, rich in arginine and serine, called an **RS domain**.

Regulation by Activators and Repressors

- The RS domain, found at the carboxy-terminal end of the protein, mediates interactions between the SR protein and proteins within the splicing machinery, recruiting that machinery to a nearby splice site.

Regulation by Activators and Repressors

- This activator regulates the alternative splicing of a set of pre-mRNAs in the fly ovary.
- It works by binding to sites near the 3' splice site of specific exons in those pre-mRNAs and recruiting the U2AF splicing factor.

Regulation by Activators and Repressors

- Most silencers are recognized by members of the heterogeneous nuclear ribonucleoprotein (hnRNP) family.
- These bind RNA but lack the RS domains and thus cannot recruit the splicing machinery.

Regulation by Activators and Repressors

- Instead, by blocking specific splice sites, they repress the use of those sites.
- In case of *Dscam*, for example, Hrp36 inhibits inclusion of exon 6 variants in the mRNA.

Regulation by Activators and Repressors

- Another example is hnRNPA1, which binds to an exonic silencer element within an exon of the HIV, *tat* pre- RNA and represses the inclusion of that exon in the final mRNA.

Regulation by Activators and Repressors

- By binding to its site, the repressor blocks binding of the activator SC35 (an SR protein) to a nearby enhancer element.

Regulation by Activators and Repressors

- In this case, blocking is not direct - the two binding sites do not overlap - but hnRNPA1 promotes cooperative binding of additional molecules of hnRNPA1 to adjacent sequences, spreading over the enhancer site.

Regulation by Activators and Repressors

- When present, another SR protein (SF2/ASF) can overcome this repression because it has a higher affinity for the enhancer sequence than does SC35 and therefore displaces the repressors bound there.

Regulation by Activators and Repressors

- Another mammalian splicing repressor is the hnRNPI protein.
- In some cases this protein blocks the binding of the basic splicing machinery by binding directly to the Py tract (hnRNPI is also called the polypyrimidine tract-binding protein).

Regulation by Activators and Repressors

- In other cases it excludes a given exon from the mature mRNA by binding to sequences that flank that exon.

Regulation by Activators and Repressors

- This exclusion occurs either because molecules of hnRNPI at each end of the exon interact and loop out the exon, which is then passed over by the spliceosome.

Regulation by Activators and Repressors

- or because the molecules of hnRNPI at each end bind cooperatively with other molecules of hnRNPI, coating the RNA across the whole exon. This too would render the exon invisible to the splicing machinery.

Regulation by Activators and Repressors

- We have emphasized alternative splicing as a way in which multiple protein products can be produced from a single gene. These different proteins are called **isoforms**.

Regulation by Activators and Repressors

- They can have similar functions, distinct functions, or even antagonistic functions.
- But even some genes that encode only a single functional protein show alternative splicing.

Regulation by Activators and Repressors

- In those cases, alternative splicing is used simply as a way of switching expression of the gene on and off.
- This is achieved in two ways.

Regulation by Activators and Repressors

- Most straight forwardly, an exon contains a stop codon, and, when incorporated into mRNA, this prematurely terminates translation generating a truncated polypeptide.

Regulation by Activators and Repressors

- Typically, such an incomplete protein is non-functional and rapidly degraded.
- Alternative splicing determines whether or not the exon with the stop codon is included in a given mRNA.

Regulation by Activators and Repressors

- The second way alternative splicing can be used as an on/off switch is by regulating the use of an intron, which, when retained in the mRNA, ensures that species is not transported out of the nucleus and so is never translated.

END

Mechanism of RNA Editing

- One important clue about the mechanism of RNA editing is that partially edited transcripts have been isolated, and these are always edited at their 3'-ends but not at their 5'-ends.

Mechanism of RNA Editing

- This suggests strongly that editing proceeds in a 3'→5' direction.
- Kenneth Stuart and colleagues first reported this phenomenon in 1988.

Mechanism of RNA Editing

- Their experimental tool was RT-PCR, starting with reverse transcriptase to make the first DNA strand from an RNA template, followed by standard PCR.

Mechanism of RNA Editing

- In one experiment, Stuart and coworkers used pairs of PCR primers in which both were edited primers, both unedited primers, or one of each.

Mechanism of RNA Editing

- A completely edited RNA will hybridize only to edited primers and give a PCR signal, whereas it will not hybridize to unedited primers, so any PCR protocol including at least one unedited primer will not give a signal from this RNA.

Mechanism of RNA Editing

- By contrast, a completely unedited RNA will react only with unedited primers.

Mechanism of RNA Editing

- But the real test is to use an unedited 5'-primer and an edited 3'-primer to detect 3'-edited transcripts, or an edited 5'-primer and an unedited 3'-primer to detect 5'-edited transcripts.

Mechanism of RNA Editing

- If editing goes from 3' to 5' in the transcript, then 3'-edited transcripts, but not 5'-edited transcripts, should be detected.
- The signals were present only when both primers were edited, or the 3'-primer was edited.

Mechanism of RNA Editing

- No signal was observed when only the 5'-primer was edited. Thus, 3'-editing occurred in the absence of 5'-editing, but 5'-editing did not occur without 3'-editing.
- This is consistent with editing in the 3'→5' direction.

Mechanism of RNA Editing

- This experiment is valuable, but it has a flaw: None of the lanes involving the unedited 3'-primer shows a signal.
- What determines where the editing system should add or delete UMPs?

Mechanism of RNA Editing

- Larry Simpson and colleagues found the answer in 1990 when they discovered **guide RNAs (gRNAs)** encoded in *Leishmania* maxicircles.

Mechanism of RNA Editing

- They found in maxicircle DNA, seven short sequences that could produce short RNAs (gRNAs) complementary to parts of five different edited mitochondrial mRNAs.

Mechanism of RNA Editing

- In principle, such gRNAs could direct the insertion and deletion of UMPs over a stretch of several dozen nucleotides in the mRNA.

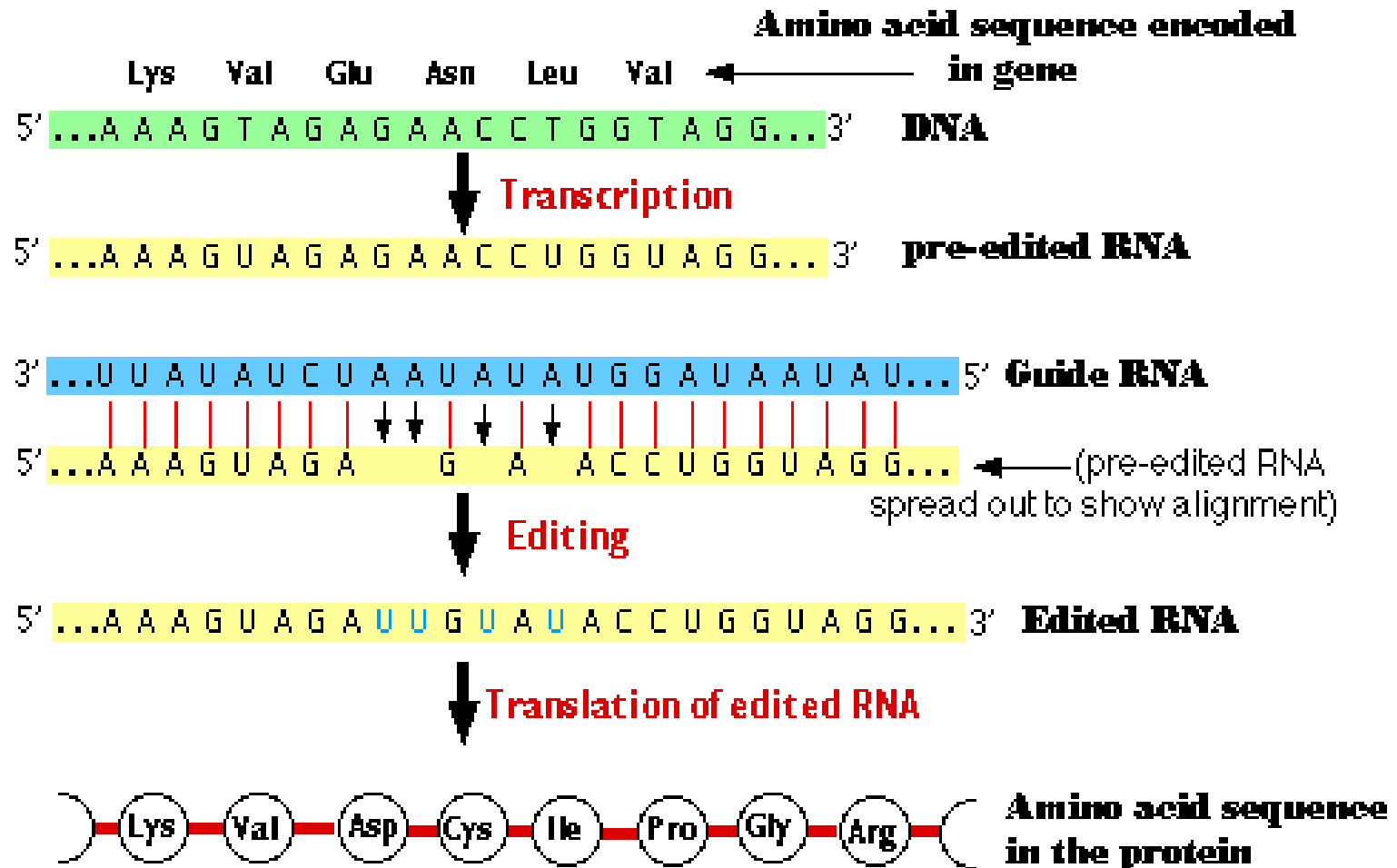
Mechanism of RNA Editing

- Once that editing is done, another gRNA could hybridize near the 5'-end of the newly edited region and direct editing of a new segment.

Mechanism of RNA Editing

- Working in this way from the 3'-end of the mRNA toward the 5'-end, successive gRNAs bind to regions edited by their predecessor gRNAs and direct further editing until they have finished the whole editing job.

Mechanism of RNA Editing



Mechanism of RNA Editing

- Later in 1990, Nancy Sturm and Larry Simpson found that minicircles also encode gRNAs.
- But besides the coding potential, they found direct evidence for the existence of gRNAs.

Mechanism of RNA Editing

- The precise mechanism of editing, the cutting and pasting required to insert and delete UMPs, remained unclear for several years, but the enzyme activities found in kinetoplasts provided some hints.

Mechanism of RNA Editing

- For example, kinetoplasts have a terminal **uridylyl transferase (TUTase)** that could add extra UMPs (uridylates) to the mRNA during editing.

Mechanism of RNA Editing

- Because the mRNA has to be cut to accept these new UMPs, it must also be ligated together again, and kinetoplasts also contain an **RNA ligase**.

Mechanism of RNA Editing

- Uridylates at the ends of gRNAs could be transferred to the pre-mRNA by transesterification. That is, the uridylates could be plucked off of the ends of gRNAs and transferred directly to the pre-mRNA.

Mechanism of RNA Editing

- Then, in 1994, Scott Seiwert and Stuart used a mitochondrial extract and a gRNA to edit a synthetic pre-mRNA.
- They found that deletion of UMPs required three enzymatic activities.

Mechanism of RNA Editing

- 1. An endonuclease that follows directions from the gRNA and cuts the pre-mRNA at the site where a UMP needs to be removed.
- 2. a 3'-exonuclease that is specific for terminal uridines; and
- 3. an RNA ligase.

Mechanism of RNA Editing

- In 1996, using a similar in vitro system, Stuart and colleagues demonstrated that UMP insertion follows a similar three-step pathway.

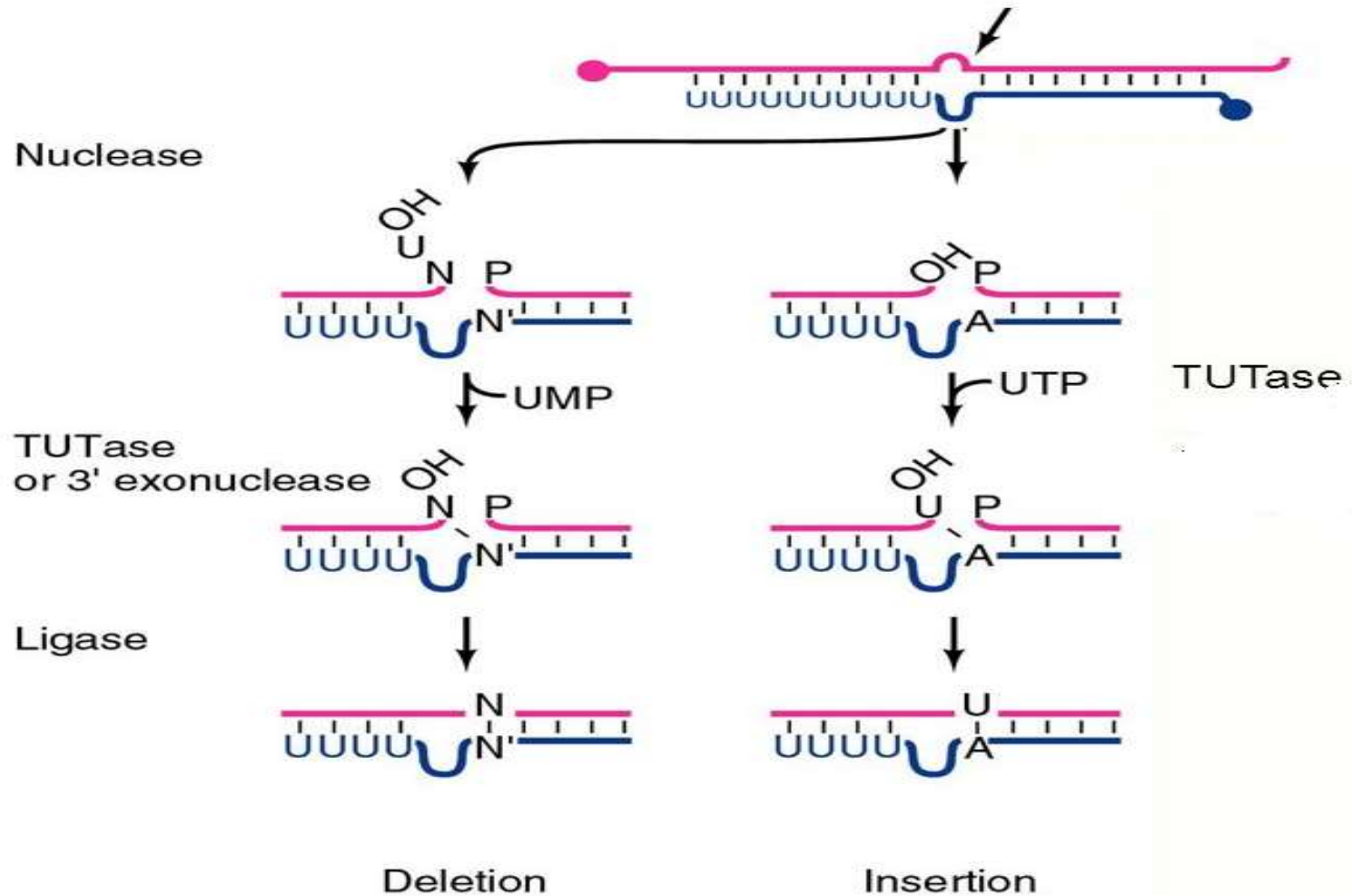
Mechanism of RNA Editing

- 1. a gRNA-directed endonuclease cuts at the site where UMP insertion is required.
- 2. an enzyme (probably TUTase) transfers UMPs from UTP (not from gRNA), as directed by the gRNA.

Mechanism of RNA Editing

- 3. an RNA ligase puts the two pieces of RNA back together.

Mechanism of RNA Editing



Mechanism of RNA Editing

- One important clue about the mechanism of RNA editing is that partially edited transcripts have been isolated, and these are always edited at their 3'-ends but not at their 5'-ends.

Mechanism of RNA Editing

- This suggests strongly that editing proceeds in a 3'→5' direction.
- Kenneth Stuart and colleagues first reported this phenomenon in 1988.

Mechanism of RNA Editing

- Their experimental tool was RT-PCR, starting with reverse transcriptase to make the first DNA strand from an RNA template, followed by standard PCR.

Mechanism of RNA Editing

- In one experiment, Stuart and coworkers used pairs of PCR primers in which both were edited primers, both unedited primers, or one of each.

Mechanism of RNA Editing

- A completely edited RNA will hybridize only to edited primers and give a PCR signal, whereas it will not hybridize to unedited primers, so any PCR protocol including at least one unedited primer will not give a signal from this RNA.

Mechanism of RNA Editing

- By contrast, a completely unedited RNA will react only with unedited primers.

Mechanism of RNA Editing

- But the real test is to use an unedited 5'-primer and an edited 3'-primer to detect 3'-edited transcripts, or an edited 5'-primer and an unedited 3'-primer to detect 5'-edited transcripts.

Mechanism of RNA Editing

- If editing goes from 3' to 5' in the transcript, then 3'-edited transcripts, but not 5'-edited transcripts, should be detected.
- The signals were present only when both primers were edited, or the 3'-primer was edited.

Mechanism of RNA Editing

- No signal was observed when only the 5'-primer was edited. Thus, 3'-editing occurred in the absence of 5'-editing, but 5'-editing did not occur without 3'-editing.
- This is consistent with editing in the 3'→5' direction.

Mechanism of RNA Editing

- This experiment is valuable, but it has a flaw: None of the lanes involving the unedited 3'-primer shows a signal.
- What determines where the editing system should add or delete UMPs?

Mechanism of RNA Editing

- Larry Simpson and colleagues found the answer in 1990 when they discovered **guide RNAs (gRNAs)** encoded in *Leishmania* maxicircles.

Mechanism of RNA Editing

- They found in maxicircle DNA, seven short sequences that could produce short RNAs (gRNAs) complementary to parts of five different edited mitochondrial mRNAs.

Mechanism of RNA Editing

- In principle, such gRNAs could direct the insertion and deletion of UMPs over a stretch of several dozen nucleotides in the mRNA.

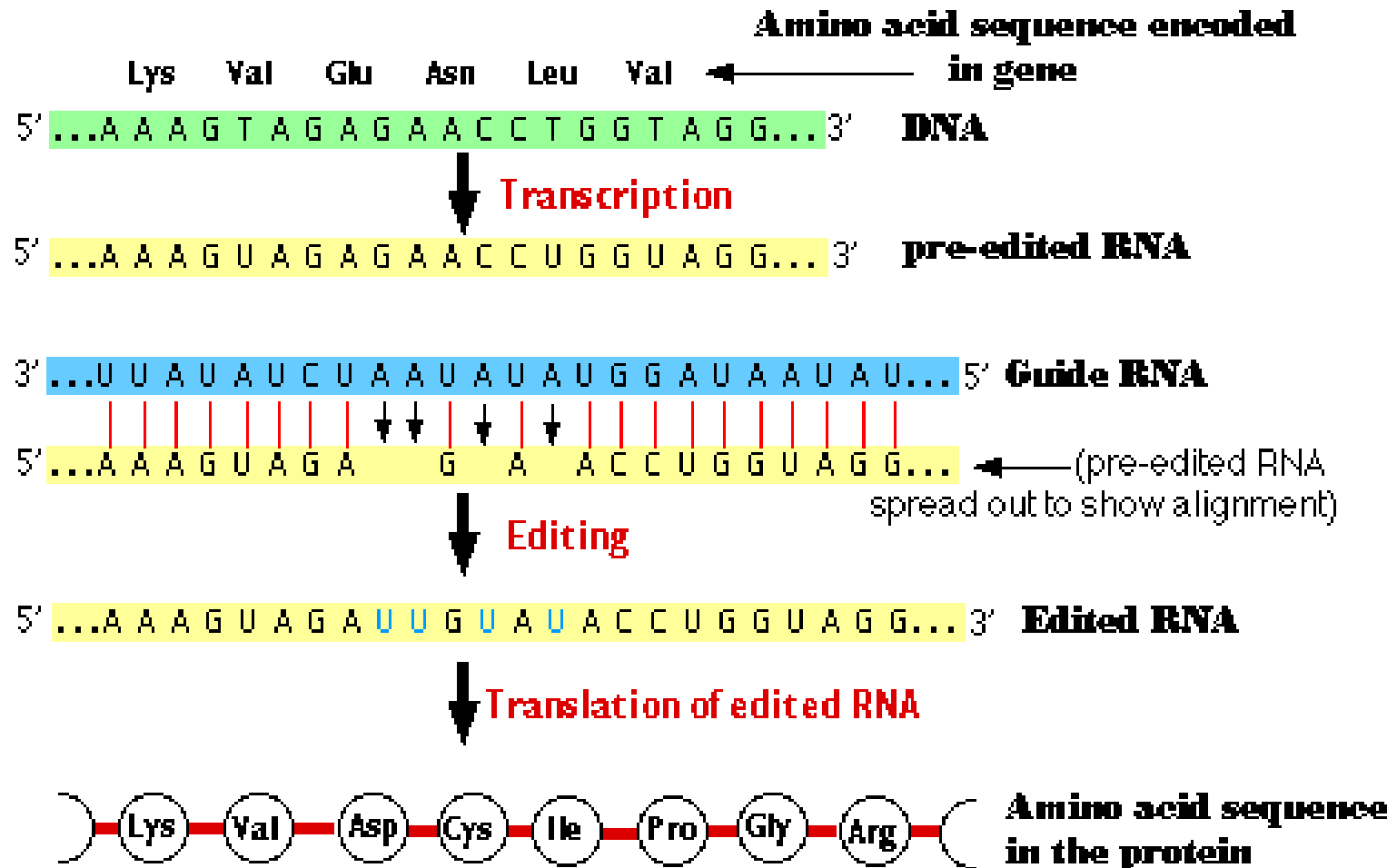
Mechanism of RNA Editing

- Once that editing is done, another gRNA could hybridize near the 5'-end of the newly edited region and direct editing of a new segment.

Mechanism of RNA Editing

- Working in this way from the 3'-end of the mRNA toward the 5'-end, successive gRNAs bind to regions edited by their predecessor gRNAs and direct further editing until they have finished the whole editing job.

Mechanism of RNA Editing



Mechanism of RNA Editing

- Later in 1990, Nancy Sturm and Larry Simpson found that minicircles also encode gRNAs.
- But besides the coding potential, they found direct evidence for the existence of gRNAs.

Mechanism of RNA Editing

- The precise mechanism of editing, the cutting and pasting required to insert and delete UMPs, remained unclear for several years, but the enzyme activities found in kinetoplasts provided some hints.

Mechanism of RNA Editing

- For example, kinetoplasts have a terminal **uridylyl transferase (TUTase)** that could add extra UMPs (uridylates) to the mRNA during editing.

Mechanism of RNA Editing

- Because the mRNA has to be cut to accept these new UMPs, it must also be ligated together again, and kinetoplasts also contain an **RNA ligase**.

Mechanism of RNA Editing

- Uridylates at the ends of gRNAs could be transferred to the pre-mRNA by transesterification. That is, the uridylates could be plucked off of the ends of gRNAs and transferred directly to the pre-mRNA.

Mechanism of RNA Editing

- Then, in 1994, Scott Seiwert and Stuart used a mitochondrial extract and a gRNA to edit a synthetic pre-mRNA.
- They found that deletion of UMPs required three enzymatic activities.

Mechanism of RNA Editing

- 1. An endonuclease that follows directions from the gRNA and cuts the pre-mRNA at the site where a UMP needs to be removed.
- 2. a 3'-exonuclease that is specific for terminal uridines; and
- 3. an RNA ligase.

Mechanism of RNA Editing

- In 1996, using a similar in vitro system, Stuart and colleagues demonstrated that UMP insertion follows a similar three-step pathway.

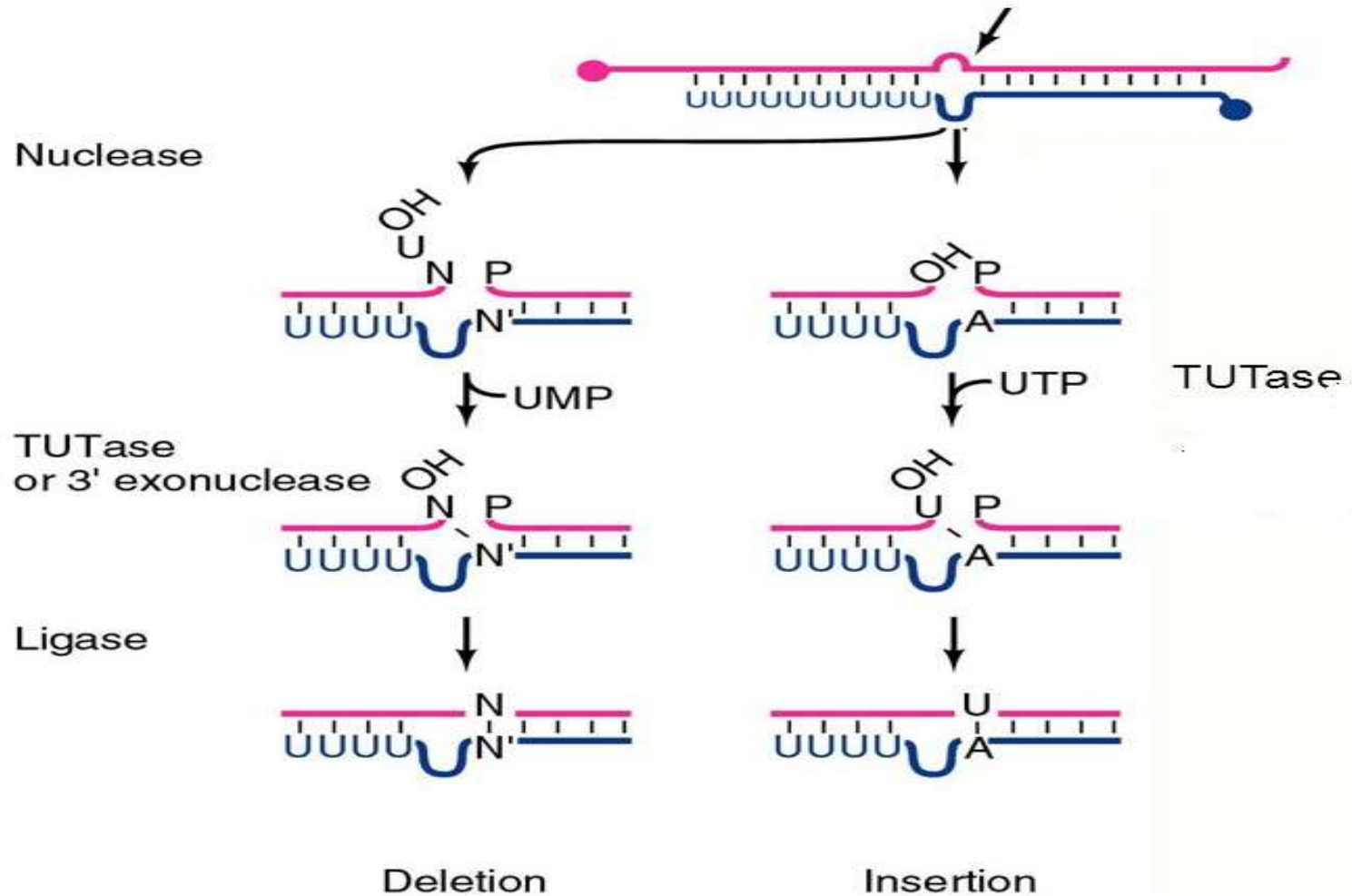
Mechanism of RNA Editing

- 1. a gRNA-directed endonuclease cuts at the site where UMP insertion is required.
- 2. an enzyme (probably TUTase) transfers UMPs from UTP (not from gRNA), as directed by the gRNA.

Mechanism of RNA Editing

- 3. an RNA ligase puts the two pieces of RNA back together.

Mechanism of RNA Editing



RNA Editing by Nucleotide Deamination

- RNA editing also plays a vital role in higher organisms even mammals.

RNA Editing by Nucleotide Deamination

- Another kind of RNA editing takes place through the deamination of adenosine, which converts adenosine to inosine, which has an oxygen in place of adenine's amino group.

RNA Editing by Nucleotide Deamination

- Because inosine forms base pairs with cytidine in the same way as guanosine, the deamination of adenosine changes the meaning of a codon.

RNA Editing by Nucleotide Deamination

- For example, an ACG (threonine) codon becomes an ICG codon, which would be read by the ribosome as GCG (alanine).

RNA Editing by Nucleotide Deamination

- This kind of RNA editing is directed by an enzyme called adenosine deaminase acting on RNA (ADAR).
- Humans and mice contain three ADAR genes: ADAR1, ADAR2, and ADAR3.

RNA Editing by Nucleotide Deamination

- The products of the first two are ubiquitous in the body, but the third gene product is found only in the brain.

RNA Editing by Nucleotide Deamination

- These enzymes are very specific. It would be disastrous if they deaminated every adenosine in an mRNA, so they select only certain adenosines in certain mRNAs.

RNA Editing by Nucleotide Deamination

- For example, ADAR2 deaminates one adenosine in the glutamate-sensitive ion-channel receptor subunit B (GluR-B) mRNA, with greater than 99% efficiency.
- This alteration in the mRNA changes a glutamine codon to an arginine codon.

RNA Editing by Nucleotide Deamination

- We know that an ion channel containing the GluR-B protein with a glutamine instead of an arginine is too permeable to calcium ions.
- Therefore, mice with a defective ADAR2 gene would have serious problems.

RNA Editing by Nucleotide Deamination

- Peter Seeburg and colleagues demonstrated that the only critical target of ADAR2 is the GluR-B transcript.

RNA Editing by Nucleotide Deamination

- The *Drosophila* genome contains only one ADAR gene.
- When this gene is mutated so the flies lack all ADAR activity, they do not carry out any mRNA editing at known editing sites.

RNA Editing by Nucleotide Deamination

- These mutant flies are viable, but they have difficulty in walking, cannot fly, and suffer progressive neural degeneration, particularly in the brain.

RNA Editing by Nucleotide Deamination

- Thus, the phenotype of this mutation is similar to the phenotype of mutations in the gene for ADAR2 in mammals.

RNA Editing by Nucleotide Deamination

- ADAR1 also appears to be essential for mammalian life.
- Kazuko Nishikura and coworkers mutated mouse stem cells to heterozygous mutant ($\text{ADAR}^{+/-}$).

RNA Editing by Nucleotide Deamination

- They found that no embryo containing this mutated gene survived to birth.
- Thus, even heterozygous mutations in ADAR1 appear to be embryonic lethal.

RNA Editing by Nucleotide Deamination

- Interestingly, certain tumors lose ADAR activity. In particular, a very malignant human brain tumor called glioblastoma multiforme (GBM) has very low ADAR2 activity, and a corresponding underediting in the GluR-B mRNA.

RNA Editing by Nucleotide Deamination

- Another kind of editing is carried out by cytidine deaminase acting on RNA (CDAR), which converts cytidine to uridine.

RNA Editing by Nucleotide Deamination

- This C→U editing is defective in about 25% of the benign peripheral nerve sheath tumors found in neurofibromatosis type I patients.
- C→U editing also appears to occur in HIV transcripts in human cells.

RNA Editing by Nucleotide Deamination

- Still another kind of editing that occurs in HIV-infected human cells is G→A editing.
- But this kind of editing cannot be explained by a single-step deamination, and it is unclear how it is accomplished.

END

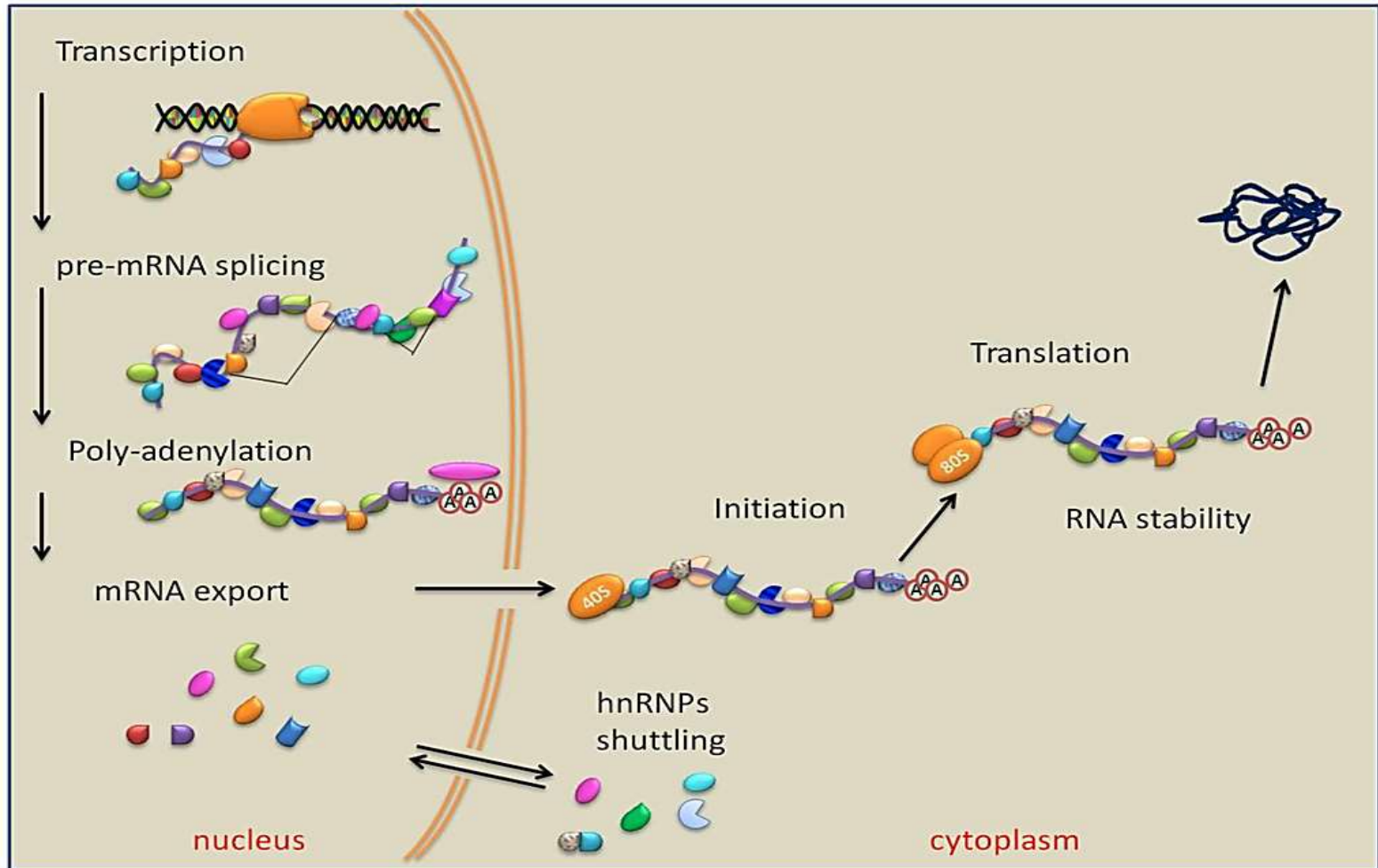
mRNA Transport

- Once fully processed-capped, intron-free, and polyadenylated mRNA is transported out of the nucleus and into the cytoplasm where it is translated to give its protein product.

mRNA Transport

- Movement from the nucleus to the cytoplasm is not a passive process.
- Indeed, it must be carefully regulated: the fully processed mRNAs represent only a small proportion of the RNA found in the nucleus.

mRNA Transport



mRNA Transport

- Many of the other RNAs would be detrimental to the cell if exported.
- These include, for example, damaged or misprocessed RNAs, and liberated introns.

mRNA Transport

- How are RNA selection and transport achieved?
- From the moment an RNA molecule starts to be transcribed, it becomes associated with proteins of various sorts.

mRNA Transport

- Initially proteins involved in capping, then splicing factors, and finally the proteins that mediate polyadenylation.

mRNA Transport

- Some of these proteins are replaced at various steps along the processing path, but others (including some SR proteins) are not; and, moreover, additional proteins join.

mRNA Transport

- As a result, a typical mature mRNA carries a collection of proteins that identifies it as being mRNA destined for transport.

mRNA Transport

- Other RNAs not only lack the particular signature collection required for transport, but have their own alternative set of proteins that actively blocks export.

mRNA Transport

- Thus, for example, excised introns will often carry hnRNPs, and these probably mark such an RNA for nuclear retention and destruction.

mRNA Transport

- Mature mRNAs carry residual SR proteins, and even another group of proteins that bind specifically to exon-exon junctions (which are only found in spliced species of course).

mRNA Transport

- The mRNAs do also contain some hnRNPs, but fewer than are typically bound to introns, and in a different context as well.

mRNA Transport

- This emphasizes the fact that it is the set of proteins, not any individual kind of protein, that marks RNAs for either export or retention in the nucleus.

mRNA Transport

- Max Birnstiel and colleagues in 1991 demonstrated that transcripts of a bacterial neomycin gene transplanted into monkey COS1 cells remained in the nucleus.

mRNA Transport

- They reasoned that the lack of a polyadenylation signal in the bacterial gene would have left the transcripts without a mature 3'-end, and that might be the reason for defective transport to the cytoplasm.

mRNA Transport

- To test this hypothesis, they provided the neomycin gene with the strong polyadenylation signal from a mammalian β -globin gene.

mRNA Transport

- This allowed for polyadenylation of the neomycin transcripts, which were then efficiently transported out of the nucleus into the cytoplasm.

mRNA Transport

- This allowed for polyadenylation of the neomycin transcripts, which were then efficiently transported out of the nucleus into the cytoplasm.

mRNA Transport

- Export takes place through a special structure in the nuclear membrane called the nuclear pore complex.

mRNA Transport

- Small molecules - those under about 50 Kd -can pass through these pores unaided; but larger molecules and complexes, including mRNAs and their associated proteins, require active transport.

mRNA Transport

- Some of the proteins associated with the RNA carry nuclear export signals that are recognized by export receptors that guide the RNA out through the pore.

mRNA Transport

- Once in the cytoplasm, the proteins are discarded, and are then recognized for import back into the nucleus where they associate with another mRNA and repeat the cycle.

END

Translation

- The genetic information contained within the order of nucleotides in messenger RNA (mRNA) is interpreted to generate the linear sequences of amino acids in proteins. This process is known as **translation**.

Translation

- Translation is among the most highly conserved across all organisms and among the most energetically costly for the cell.

Translation

- In rapidly growing bacterial cells, up to 80% of the cell's energy and 50% of the cell's dry weight are dedicated to protein synthesis.

Translation

- Indeed, the synthesis of a single protein requires the coordinated action of well over 100 proteins and RNAs.

Translation

- Unlike the complementarity between the DNA template and the ribonucleotides of the mRNA, the side chains of amino acids have little or no specific affinity for the purine and pyrimidine bases found in RNA.

Translation

- For example, the hydrophobic side chains of the amino acids alanine, valine, leucine, and isoleucine cannot form hydrogen bonds with the amino and keto groups of the nucleotide bases.

Translation

- Similarly, it is hard to imagine that how the bases of RNA can have unique affinities for the aromatic amino acids phenylalanine, tyrosine, and tryptophan.

Translation

- So Francis H. Crick in 1955 proposed that before their incorporation into polypeptides, amino acids must attach to a special adaptor molecule that is capable of directly interacting with and recognizing the coding units of the mRNA.

Translation

- Paul C. Zamecnik and Mahlon B. Hoagland (1957) showed that before their incorporation into proteins, amino acids are attached to a class of RNA molecules (representing 15% of all cellular RNA).

Translation

- These RNAs are called **transfer RNAs (tRNAs)** because their attached amino acid is subsequently transferred to the growing polypeptide chain.

Translation

- The machinery responsible for translating the language of mRNAs into the language of proteins is composed of four primary components:-

Translation

- mRNAs, tRNAs, aminoacyl-tRNA synthetases and the ribosomes.

Translation

- Together, these components accomplish the extraordinary task of translating a code written in a four-base alphabet into a second code written in the language of the 20 amino acids.

Translation

- The mRNA provides the information that must be interpreted by the translation machinery and is the template for translation.

Translation

- The protein-coding region of the mRNA consists of an ordered series of three nucleotide- long units called **codons** that specify the order of amino acids.

Translation

- The tRNAs provide the physical interface between the amino acids being added to the growing polypeptide chain and the codons in the mRNA.

Translation

- Enzymes called aminoacyl-tRNA synthetases couple amino acids to specific tRNAs that recognize the appropriate codon(s).

Translation

- The final major player in translation is the ribosome, a remarkable, multimegadalton machine composed of both RNA and protein.

Translation

- The ribosome coordinates the correct recognition of the mRNA by each tRNA and catalyzes peptide-bond formation between the growing polypeptide chain and the amino acid attached to the selected tRNA.

END

Messenger RNA

- The protein-coding region(s) of each mRNA is composed of a contiguous, non-overlapping string of codons called an **open reading frame** (commonly known as an **ORF**).

Messenger RNA

- Each ORF specifies a single protein and starts and ends at internal sites within the mRNA.
- That is, the ends of an ORF are distinct from the ends of the mRNA.

Messenger RNA

- Translation starts at the 5' end of the ORF and proceeds one codon at a time to the 3' end. The first and last codons of an ORF are known as the **start** and **stop codons**.

Messenger RNA

- In bacteria, the start codon is usually 5'-AUG-3', but 5'-GUG-3' and sometimes even 5'-UUG-3' are also used.
- Eukaryotic cells always use 5'-AUG-3' as the start codon.

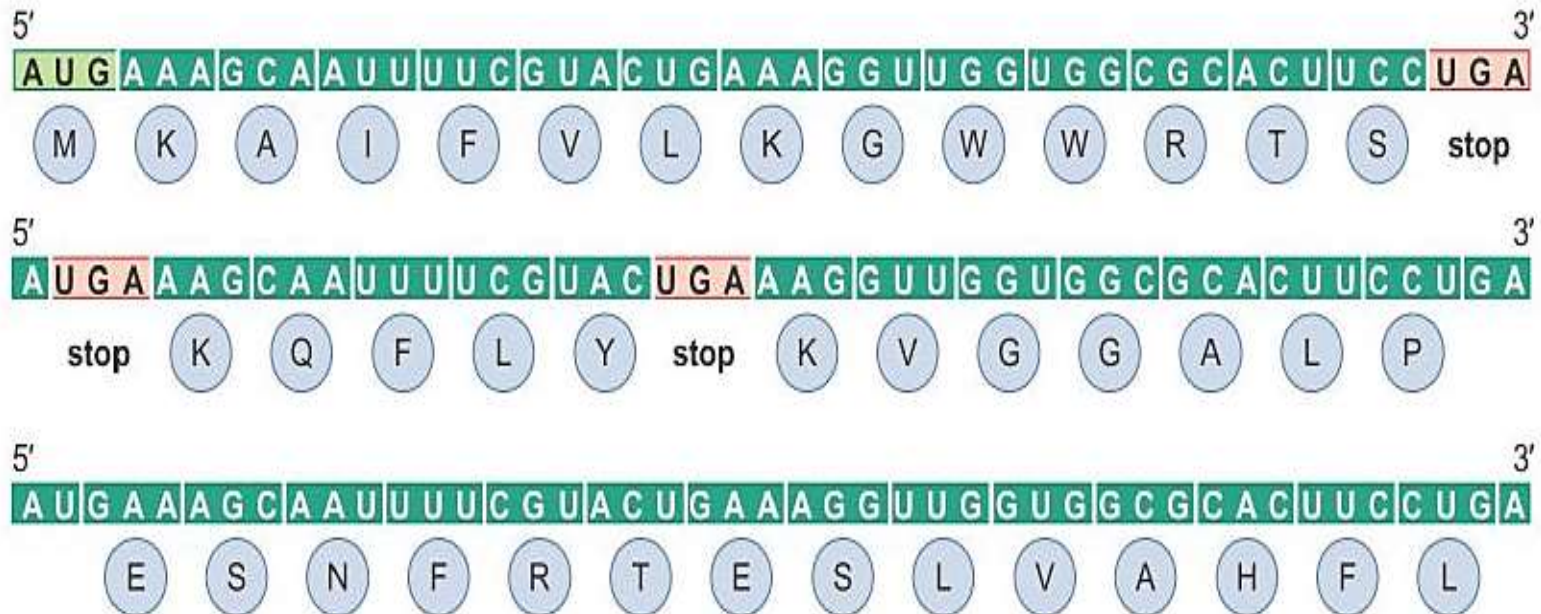
Messenger RNA

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

Messenger RNA

- Because each codon is immediately adjacent to the next codon, and because codons are three nucleotides long, any stretch of mRNA could be translated in three different reading frames.

Messenger RNA



Three possible reading frames of the *E. coli trp* leader sequence

Messenger RNA

- Once translation starts, however, the reading frame is determined.
- Thus, by setting the location of the first codon, the start codon determines the location of all following codons.

Messenger RNA

- Stop codons, of which there are three (5'-UAG-3', 5'-UGA-3', and 5'-UAA-3'), define the end of the ORF and signal termination of polypeptide synthesis.

Messenger RNA

- You can now understand the origin of the term open reading frame. It is a contiguous stretch of codons “read” in a particular frame (as set by the first codon) that is “open” to translation because it lacks a stop codon.

Messenger RNA

- mRNAs contain at least one ORF. The number of ORFs per mRNA is different between eukaryotes and prokaryotes.
- Eukaryotic mRNAs almost always contain a single ORF.

Messenger RNA

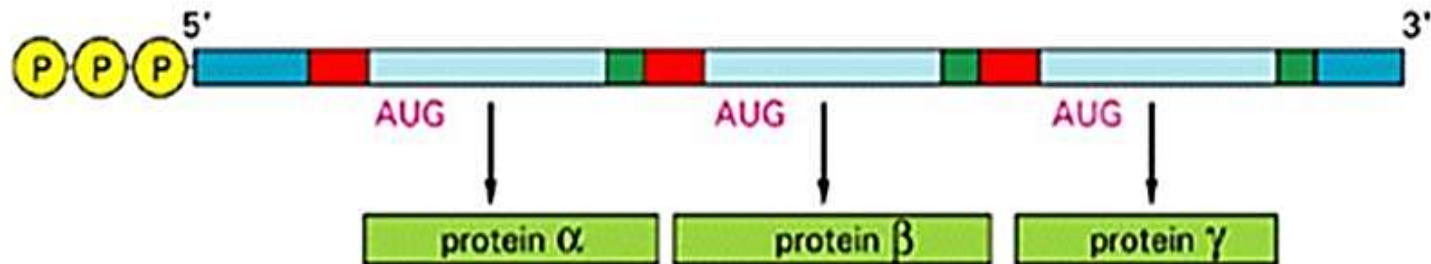
- In contrast, prokaryotic mRNAs frequently contain two or more ORFs.
- mRNAs containing multiple ORFs are known as polycistronic RNAs and those encoding a single ORF are known as monocistronic RNAs.

Messenger RNA

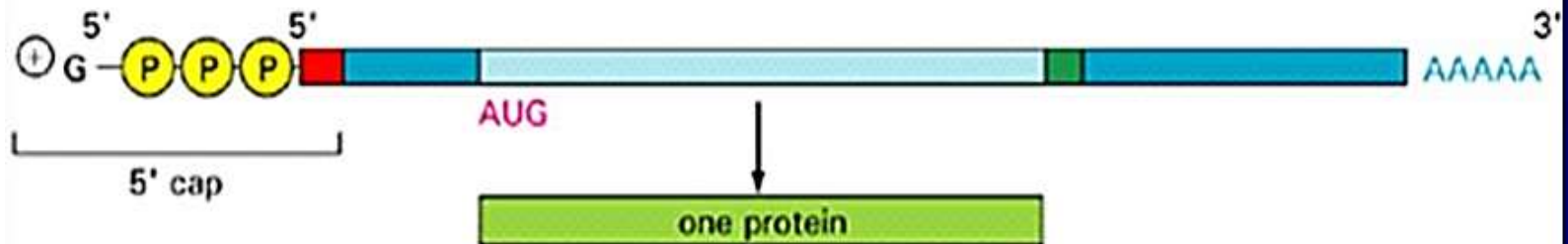
- The polycistronic mRNAs found in bacteria often encode proteins that perform related functions, such as different steps in the biosynthesis of an amino acid or nucleotide.

Messenger RNA

procaryotic mRNA



eucaryotic mRNA



Prokaryotic mRNAs

- For translation to occur, the ribosome must be recruited to the mRNA.
- Prokaryotic mRNAs have a ribosome-binding site that recruits the translational machinery.

Prokaryotic mRNAs

- To facilitate binding by a ribosome, many prokaryotic ORFs contain a short sequence upstream (on the 5' side) of the start codon called the **ribosome-binding site (RBS)**.

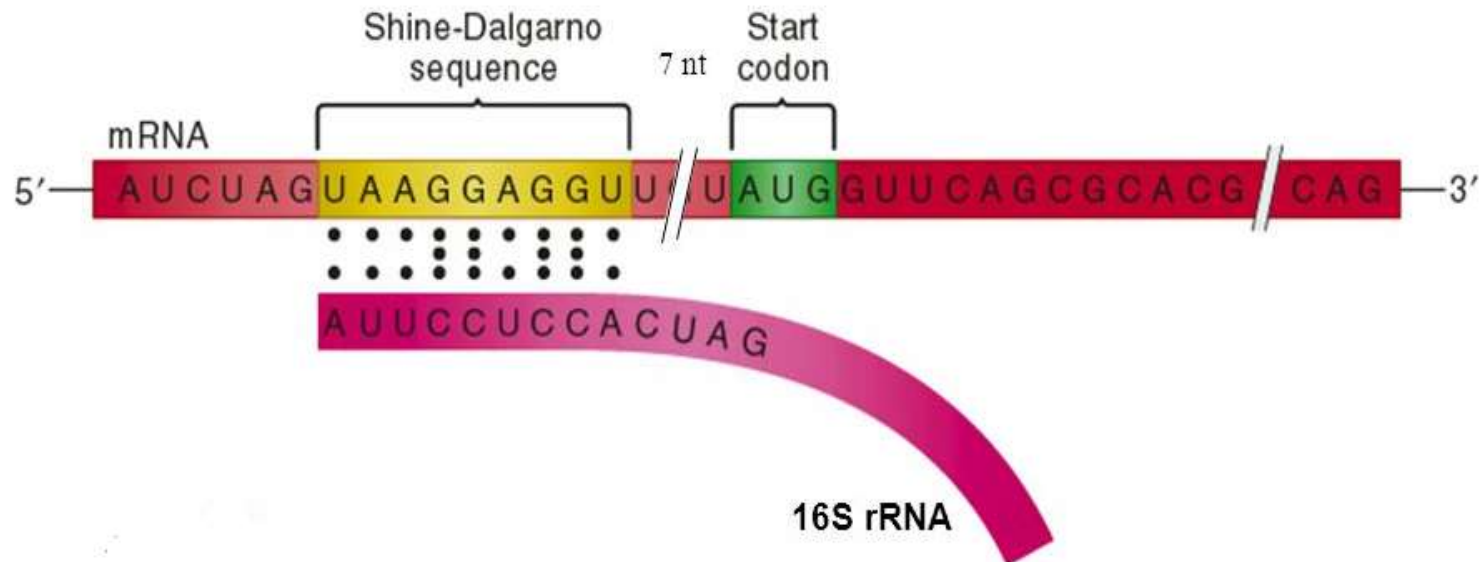
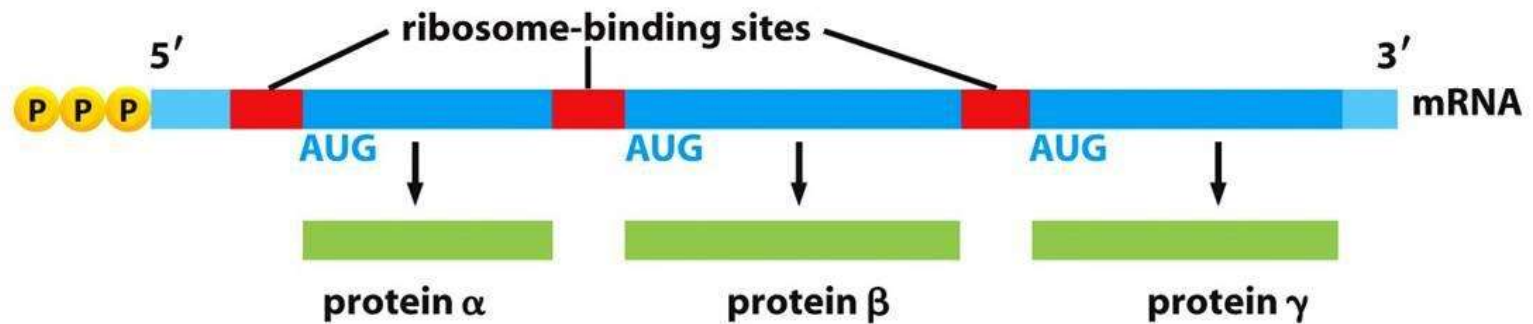
Prokaryotic mRNAs

- This element is also referred to as a **Shine Öalgarno sequence** after the scientists who discovered it by comparing the sequences of multiple mRNAs.

Prokaryotic mRNAs

- The RBS, typically located 3–9 bp on the 5' side of the start codon, is complementary to a sequence located near the 3' end of one of the ribosomal RNA components, the 16S ribosomal RNA (rRNA).

Prokaryotic mRNAs



Prokaryotic mRNAs

- The RBS base-pairs with this RNA, thereby aligning the ribosome with the beginning of the ORF. The core of this region of the 16S rRNA has the sequence 5'-CCUCCU-3'.

Prokaryotic mRNAs

- Not surprisingly, prokaryotic RBS are most often a subset of the sequence 5'-AGGAGG-3'.

Prokaryotic mRNAs

- The extent of complementarity and the spacing between the RBS and the start codon has a strong influence on how actively a particular ORF is translated.

Prokaryotic mRNAs

- High complementarity and proper spacing promote active translation, whereas limited complementarity and/or poor spacing generally support lower levels of translation.

Prokaryotic mRNAs

- Some prokaryotic ORFs lack a strong RBS but are nonetheless actively translated.
- These ORFs are not the first ORF in an mRNA but instead are located just after another ORF in a polycistronic message.

Prokaryotic mRNAs

- In these cases, the start codon of the downstream ORF often overlaps the 3' end of the upstream ORF.

Prokaryotic mRNAs

- Thus, a ribosome that has just completed translating the upstream ORF is positioned to begin translating from the start codon for the downstream ORF.

Prokaryotic mRNAs

- This phenomenon of linked translation between overlapping ORFs is known as **translational coupling**.
- So in this situation translation of the downstream ORF requires translation of the upstream ORF.

Prokaryotic mRNAs

- Indeed, with two translationally coupled genes, a mutation that leads to a premature stop codon in the upstream ORF also prevents translation of the downstream ORF.

END

Eukaryotic mRNAs

- Unlike their prokaryotic counterparts, eukaryotic mRNAs recruit ribosomes using a specific chemical modification called the **5' cap**, which is located at the extreme 5' end of the mRNA.

Eukaryotic mRNAs

- The 5' cap is a methylated guanine nucleotide that is joined to the 5' end of the mRNA via an unusual 5'-to-5' linkage.

Eukaryotic mRNAs

- Created in three steps, the guanine nucleotide of the 5' cap is connected to the 5' end of the mRNA through three phosphate groups.

Eukaryotic mRNAs

- The resulting 5' cap is required to recruit the ribosome to the mRNA. Once bound to the mRNA, the ribosome moves in a 5' → 3' direction until it encounters a 5'-AUG-3' start codon, a process called **scanning**.

Eukaryotic mRNAs

- Two other features of eukaryotic mRNAs stimulate translation. One feature is the presence, in some mRNAs, of a purine three bases upstream of the start codon and a guanine immediately downstream (5'-G/ANNAUGG-3').

Eukaryotic mRNAs

- This sequence was originally identified by Marilyn Kozak and is referred to as the Kozak sequence. Many eukaryotic mRNA lack these bases, but their presence increases the efficiency of translation.

Eukaryotic mRNAs

- In contrast to the situation in prokaryotes, these bases are thought to interact with the initiator tRNA, not with an RNA component of the ribosome.

Eukaryotic mRNAs

- A second feature that contributes to efficient translation is the presence of a poly-A tail at the extreme 3' end of the mRNA.
- This tail is added enzymatically by the enzyme poly-A polymerase.

Eukaryotic mRNAs

- Despite its location at the 3' end of the mRNA, the poly-A tail enhances the level of translation of them RNA by enhancing the recruitment of key translation initiation factors.

Eukaryotic mRNAs

- Importantly, in addition to their roles in translation, these 5'- and 3'-end modifications also protect eukaryotic mRNAs from rapid degradation.

END

Transfer RNA

- The heart of protein synthesis is the “translation” of nucleotide sequence information (in the form of codons) into amino acids.

Transfer RNA

- This is accomplished by tRNA molecules, which act as adaptors between codons and the amino acids they specify.

Transfer RNA

- There are many types of tRNA molecules, but each is attached to a specific amino acid, and each recognizes a particular codon, or codons, in the mRNA (most tRNAs recognize more than one codon).

Transfer RNA

- tRNA molecules are between 75 and 95 ribonucleotides in length.
- Although the exact sequence varies, all tRNAs have certain features in common.

Transfer RNA

- First, all tRNAs end at the 3' terminus with the sequence 5'-CCA-3'.
- Consistent with this absolute conservation, the 3' end of this sequence is the site that is attached to the cognate amino acid.

Transfer RNA

- A second striking aspect of tRNAs is the presence of several unusual bases in their primary structure.
- These unusual features are created post-transcriptionally by enzymatic modification of normal bases in the polynucleotide chain.

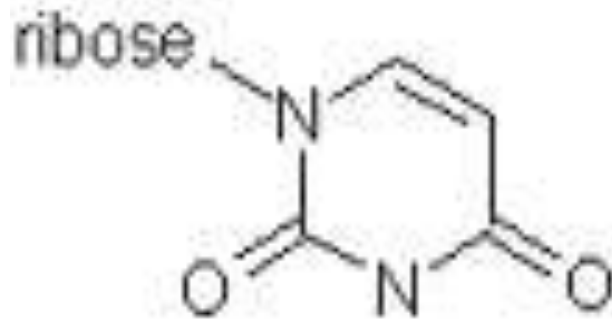
Transfer RNA

- For example, pseudouridine (ψ U) is derived from uridine by an isomerization in which the site of attachment of the uracil base to the ribose is switched from the nitrogen at ring position 1 to the carbon at ring position 5.

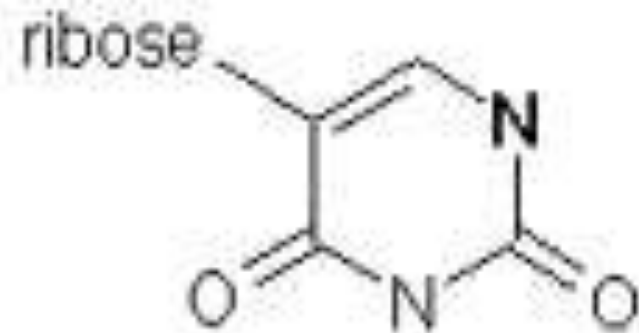
Transfer RNA

- Likewise, dihydrouridine (D) is derived from uridine by enzymatic reduction of the double bond between the carbons at positions 5 and 6.

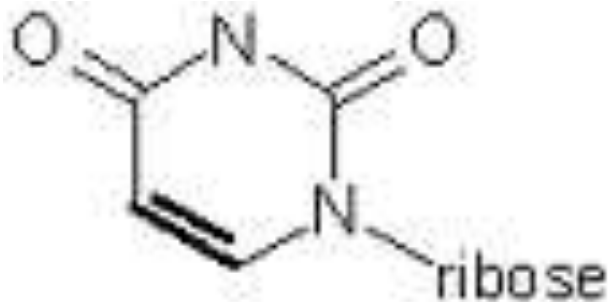
Transfer RNA



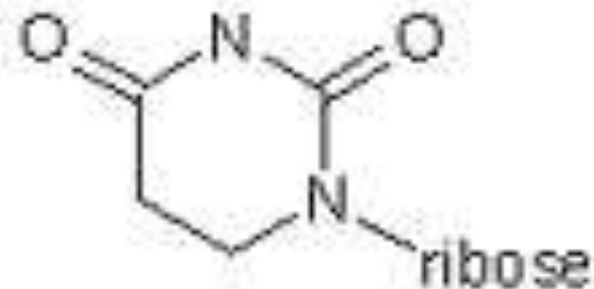
uridine



Pseudouridine



uridine



dihydrouridine

Transfer RNA

- Other unusual bases found in tRNA include hypoxanthine, thymine, and methylguanine.
- These modified bases are not essential for tRNA function, but cells lacking these modified bases show reduced rates of growth.

Transfer RNA

- This observation suggests that the modified bases lead to improved tRNA function.
- For example, hypoxanthine plays an important role in the process of codon recognition by certain tRNAs.

END

Secondary Structure of tRNA

- RNA molecules typically contain regions of self complementarity that enable them to form limited stretches of double helix that are held together by base pairing.

Secondary Structure of tRNA

- tRNA molecules show a characteristic and highly conserved pattern of single-stranded and double stranded regions (secondary structure) that can be illustrated as a cloverleaf.

Secondary Structure of tRNA

- The principal features of the tRNA cloverleaf are an acceptor stem, three stem-loops (referred to as the ψ U loop, the D loop, and the anticodon loop), and a fourth variable loop.

Secondary Structure of tRNA

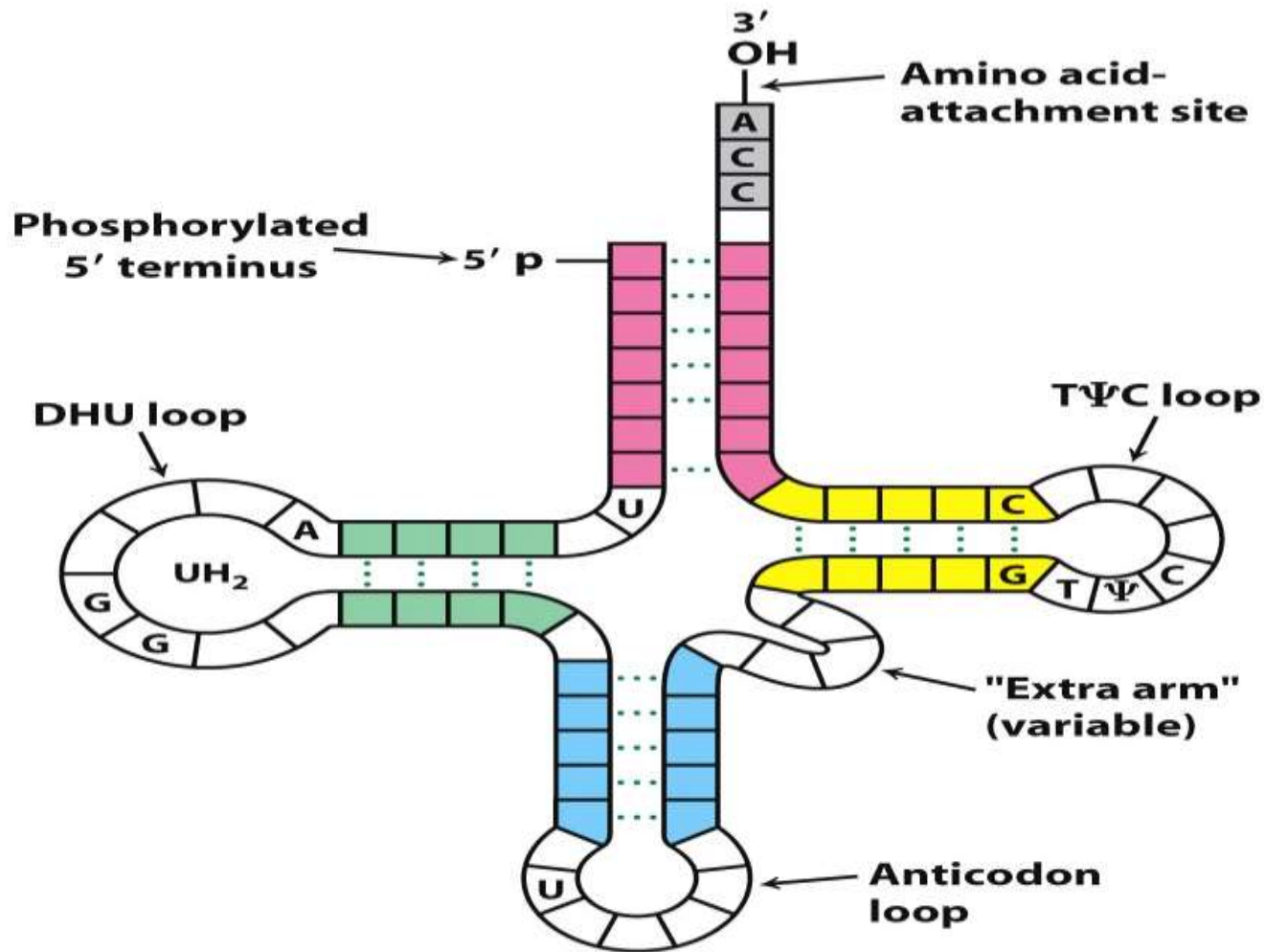


Figure 30.3
Biochemistry, Seventh Edition
© 2012 W. H. Freeman and Company

Secondary Structure of tRNA

- **The Acceptor Stem:**
- It is so-named because it is the site of attachment of the amino acid, is formed by pairing between the 5' and 3' ends of the tRNA molecule.

Secondary Structure of tRNA

- **The ψ U Loop:**
- It is so-named because of the characteristic presence of the unusual base ψ U in the loop. The modified base is often found within the sequence 5'-TCUCG-3'.

Secondary Structure of tRNA

- **The D Loop:**
- It takes its name from the characteristic presence of dihydrouridines in the loop.

Secondary Structure of tRNA

- **The Anticodon Loop:**
- As its name implies, contains the anticodon, a three-nucleotide-long sequence that is responsible for recognizing the codon by base pairing with the mRNA.

Secondary Structure of tRNA

- **The Variable Loop:**
- It sits between the anticodon loop and the ψ U loop and, as its name implies, varies in size from 3 to 21 bases.

END

Attachment of Amino Acids to tRNA

- tRNA molecules to which an amino acid is attached are said to be **charged**, and tRNAs that lack an amino acid are said to be **uncharged**.

Attachment of Amino Acids to tRNA

- Charging requires an acyl linkage between the carboxyl group of the amino acid and the 2'- or 3'-hydroxyl group of the adenosine nucleotide that protrudes from the acceptor stem at the 3' end of the tRNA.

Attachment of Amino Acids to tRNA

- This acyl linkage is a high-energy bond because its hydrolysis results in a large change in free energy.

Attachment of Amino Acids to tRNA

- This is significant for protein synthesis: the energy released when this acyl bond is broken is coupled to the formation of the peptide bonds that link amino acids to each other in polypeptide chains.

Attachment of Amino Acids to tRNA

- All aminoacyl-tRNA synthetases attach an amino acid to a tRNA in two enzymatic steps:
 - Adenylation
 - tRNA charging

Attachment of Amino Acids to tRNA

- Step one is **adenylation** in which the amino acid reacts with ATP to become adenylylated with the concomitant release of pyrophosphate.

Attachment of Amino Acids to tRNA

- Adenylylation refers to transfer of AMP, as opposed to adenylation, which would indicate the transfer of adenine.

Attachment of Amino Acids to tRNA

- The principal driving force for the adenylation reaction is the subsequent hydrolysis of pyrophosphate by pyrophosphatase.

Attachment of Amino Acids to tRNA

- As a result of adenylation, the amino acid is attached to adenylic acid via a high-energy ester bond in which the carbonyl group of the amino acid is joined to the phosphoryl group of AMP.

Attachment of Amino Acids to tRNA

- Step two is tRNA Charging in which the adenylylated amino acid, which remains tightly bound to the synthetase, reacts with tRNA.

Attachment of Amino Acids to tRNA

- This reaction results in the transfer of the amino acid to the 3' end of the tRNA via the 2'- or 3'-hydroxyl and the release of AMP.

Attachment of Amino Acids to tRNA

- There are two classes of tRNA synthetases:
- Class I enzymes attach the amino acid to the 2'-OH of the tRNA and are generally monomeric.

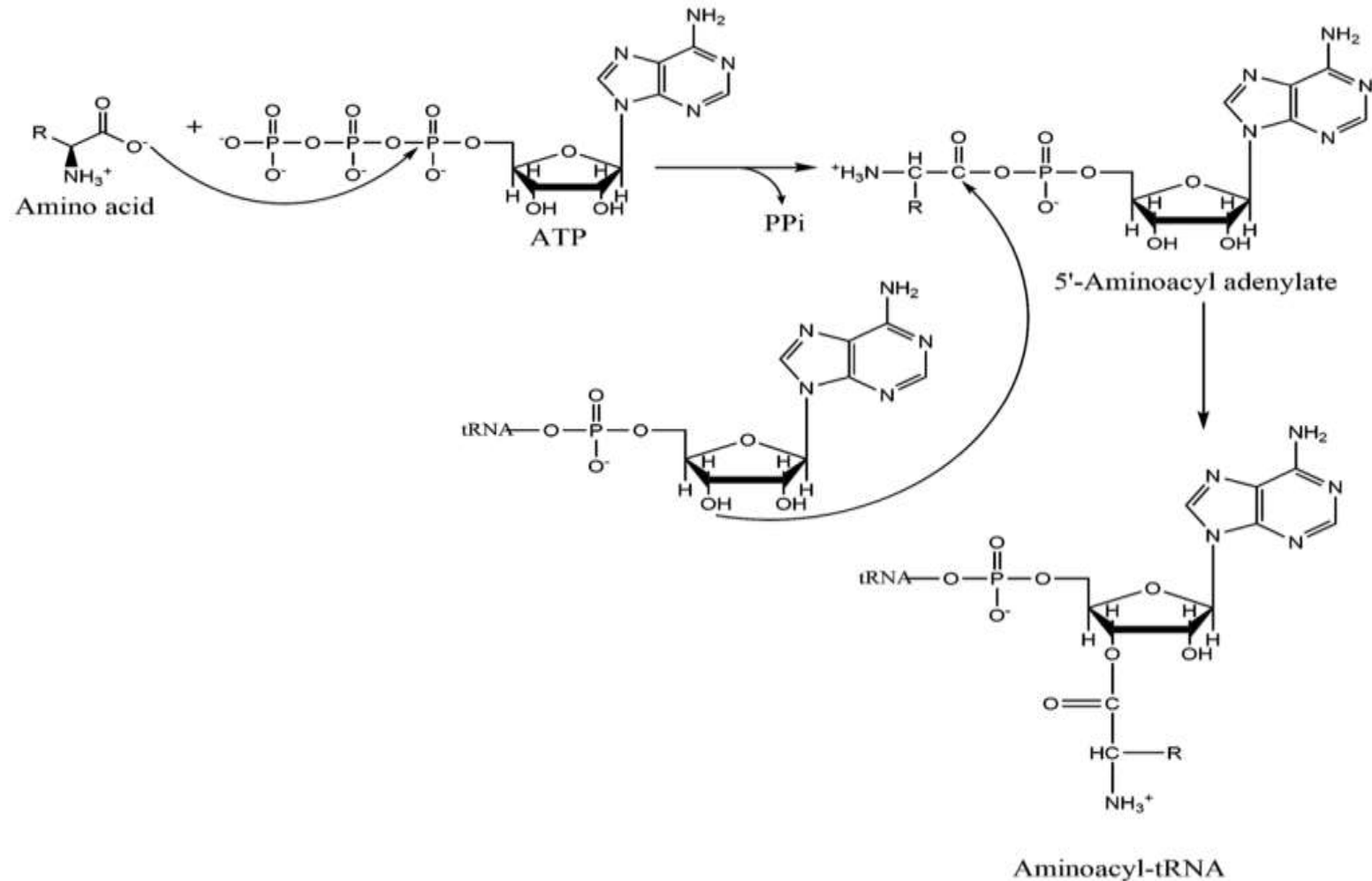
Attachment of Amino Acids to tRNA

- Class II enzymes attach the amino acid to the 3'-OH of the tRNA and are typically dimeric or tetrameric.

Attachment of Amino Acids to tRNA

- Although the initial coupling between the tRNA and the amino acid is different, once released from the synthetase, the amino acid rapidly equilibrates between attachment at the 3'-OH and the 2'-OH.

Attachment of Amino Acids to tRNA



Attachment of Amino Acids to tRNA

- Each of the 20 amino acids is attached to the appropriate tRNA by a single, dedicated tRNA synthetase.

Attachment of Amino Acids to tRNA

- Because most amino acids are specified by more than one codon, it is not uncommon for one synthetase to recognize and charge more than one tRNA (known as **isoaccepting tRNAs**).

Attachment of Amino Acids to tRNA

- Nevertheless, the same tRNA synthetase is responsible for charging all tRNAs for a particular amino acid.
- Thus, one and only one tRNA synthetase attaches each amino acid to all of the appropriate tRNAs.

END

The Ribosomes

- The ribosome is the macromolecular machine that directs the synthesis of proteins.
- The ribosome is larger and more complex than the minimal machinery required for DNA or RNA synthesis.

The Ribosomes

- The machinery for polymerizing amino acids is composed of at least three RNA molecules and more than 50 different proteins, with an overall molecular mass of >2.5 MDa.

The Ribosomes

- Compared with the speed of DNA replication i.e., 200 – 1000 nucleotides per second; translation takes place at a rate of only two to 20 amino acids per second.

The Ribosomes

- In prokaryotes, the transcription machinery and the translation machinery are located in the same compartment. Thus, the ribosome can commence translation of the mRNA as it emerges from the RNA polymerase.

The Ribosomes

- This situation allows the ribosome to proceed in tandem with the RNA polymerase as it elongates the transcript.

The Ribosomes

- Recall that the 5' end of an RNA is synthesized first, and thus the ribosome, which begins translation at the 5' end of the mRNA, can start translating a nascent transcript as soon as it emerges from the RNA polymerase.

END

Formation of Peptide Bonds

- Each new amino acid is added to the carboxyl terminus of the growing polypeptide chain (often referred to as synthesis in the amino- to carboxy-terminal direction).

Formation of Peptide Bonds

- The ribosome catalyzes a single chemical reaction — the formation of a peptide bond.

Formation of Peptide Bonds

- This reaction occurs between the amino acid residue at the carboxy-terminal end of the growing polypeptide and the incoming amino acid to be added to the chain.

Formation of Peptide Bonds

- Both the growing chain and the incoming amino acid are attached to tRNAs; as a result, during peptide-bond formation, the growing polypeptide is continuously attached to a tRNA.

Formation of Peptide Bonds

- The actual substrates for each round of amino acid addition are two charged species of tRNAs — an aminoacyl-tRNA and a **peptidyl-tRNA**.

Formation of Peptide Bonds

- As you know the aminoacyl-tRNA is attached at its 3' end to the carboxyl group of the amino acid. The peptidyl-tRNA is attached in exactly the same manner (at its 3' end) to the carboxyl terminus of the growing polypeptide chain.

Formation of Peptide Bonds

- The bond between the aminoacyl-tRNA and the amino acid is not broken during the formation of the next peptide bond.

Formation of Peptide Bonds

- Instead, the bond between the peptidyl-tRNA and the growing polypeptide chain is broken as the growing chain is attached to the amino group of the amino acid attached to the aminoacyl-tRNA to form a new peptide bond.

Formation of Peptide Bonds

- To catalyze peptide-bond formation, the 3' ends of these two tRNAs are brought into close proximity by the ribosome.

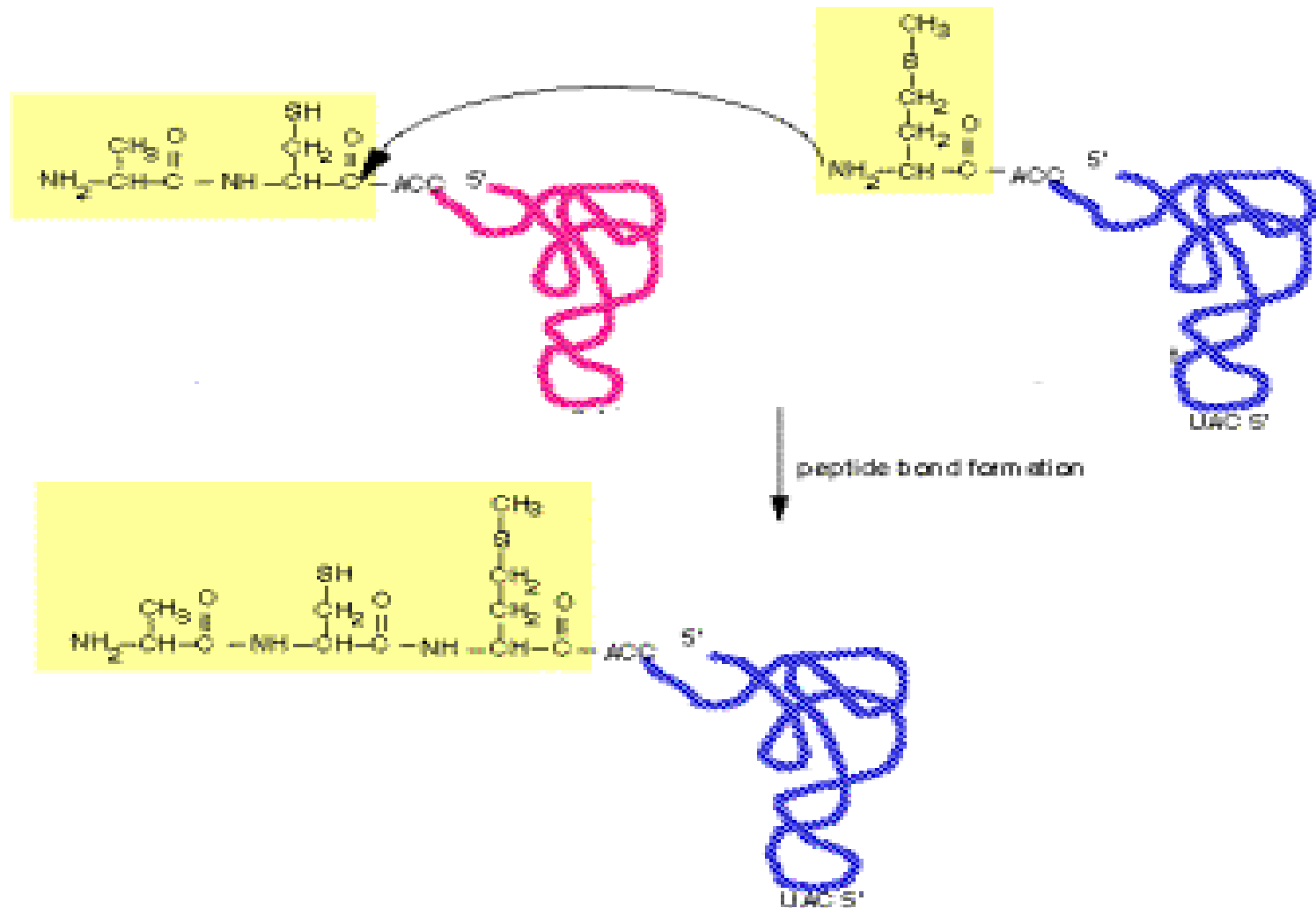
Formation of Peptide Bonds

- The resulting tRNA positioning allows the amino group of the amino acid attached to aminoacyl-tRNA to attack the carbonyl group of the most carboxy-terminal amino acid attached to the peptidyl-tRNA.

Formation of Peptide Bonds

- The result of this nucleophilic attack is the formation of a new peptide bond between the amino acids attached to the tRNAs and the release of the polypeptide chain from the peptidyl tRNA.

Formation of Peptide Bonds



Formation of Peptide Bonds

- There are two consequences of this method of polypeptide synthesis. First, this mechanism of peptide-bond formation requires that the amino terminus of the protein be synthesized before the carboxyl terminus.

Formation of Peptide Bonds

- Second, the growing polypeptide chain is transferred from the peptidyl-tRNA to the aminoacyl-tRNA. For this reason, the reaction to form a new peptide bond is called the **peptidyl transferase reaction**.

Formation of Peptide Bonds

- Interestingly, peptide-bond formation takes place without the simultaneous hydrolysis of a nucleoside triphosphate.

Formation of Peptide Bonds

- This is because peptide-bond formation is driven by breaking the high-energy acyl bond that joins the growing polypeptide chain to the tRNA.

Formation of Peptide Bonds

- Recall that this bond was created during the tRNA synthetase – catalyzed reaction that is responsible for charging tRNA.
- And the charging reaction involves the hydrolysis of a molecule of ATP.

Formation of Peptide Bonds

- Thus, the energy for peptide-bond formation originates from the molecule of ATP that was hydrolyzed during the tRNA charging reaction.

END

Binding Sites on Ribosomes for tRNA

- The ribosome is composed of two subassemblies of RNA and protein known as the large and small subunits.

Binding Sites on Ribosomes for tRNA

- The large subunit contains the **peptidyl transferase center**, which is responsible for the formation of peptide bonds.

Binding Sites on Ribosomes for tRNA

- The small subunit contains the **decoding center** in which charged tRNAs read or "decode" the codon units of the mRNA.

Binding Sites on Ribosomes for tRNA

- Both the decoding center and the peptidyl transferase center are buried within the intact ribosome.

Binding Sites on Ribosomes for tRNA

- Yet, mRNA must be threaded through the decoding center during translation, and the nascent polypeptide chain must escape from the peptidyl transferase center.

Binding Sites on Ribosomes for tRNA

- How do these polymers enter and exit the ribosome?
- The answer is provided by the structure of the ribosome, which reveals that there are "tunnels" in and out of the ribosome.

Binding Sites on Ribosomes for tRNA

- To perform the peptidyl transferase reaction, the ribosome must be able to bind at least two tRNAs simultaneously.
- In fact, the ribosome contains three tRNA-binding sites, called the A-, P-, and E-sites.

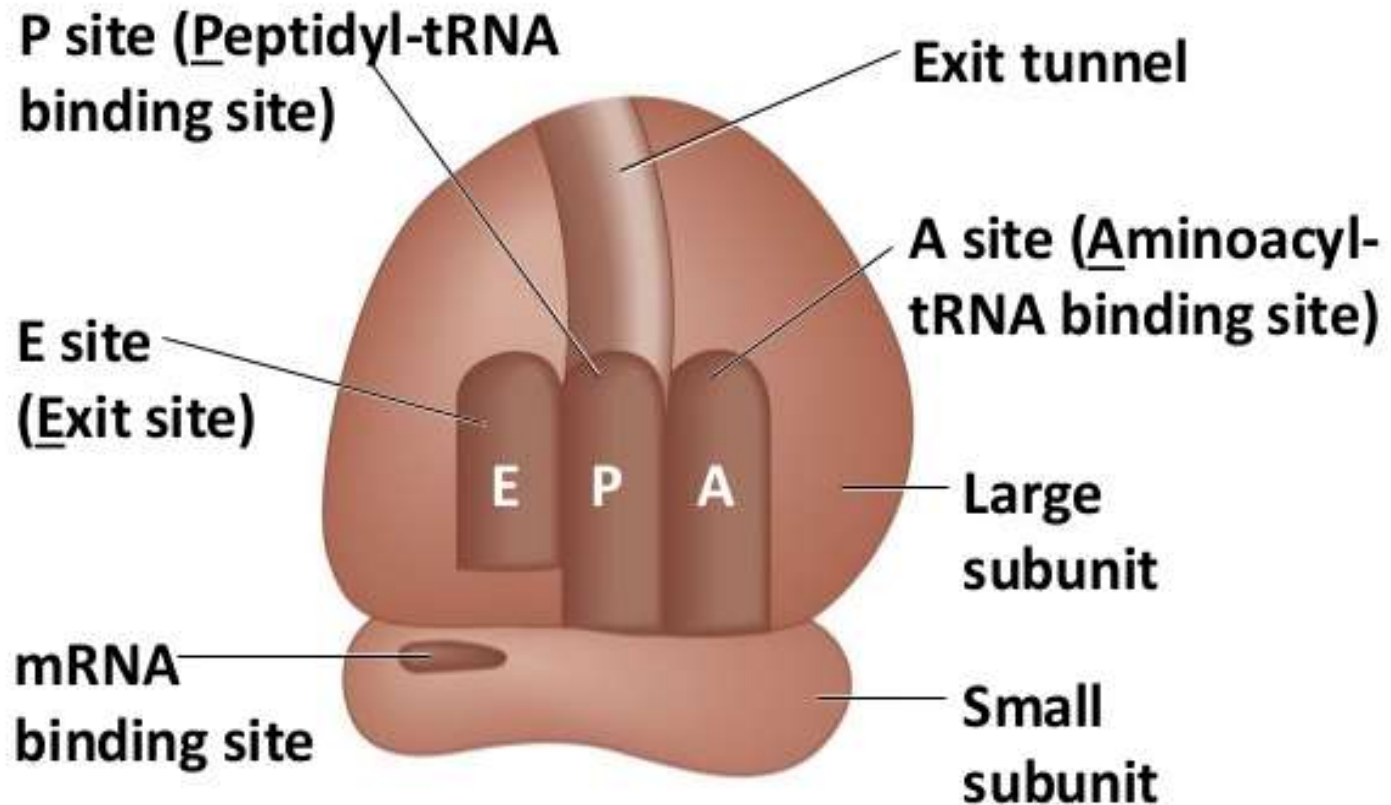
Binding Sites on Ribosomes for tRNA

- The **A-site** is the binding site for the aminoacylated-tRNA, the **P-site** is the binding site for the peptidyl-tRNA, and

Binding Sites on Ribosomes for tRNA

- The **E-site** is the binding site for the tRNA that is released after the growing polypeptide chain has been transferred to the aminoacyl-tRNA (E is for “exiting”).

Binding Sites on Ribosomes for tRNA



Binding Sites on Ribosomes for tRNA

- Each tRNA binding site is formed at the interface between the large and the small subunits of the ribosome.

Binding Sites on Ribosomes for tRNA

- In this way, the bound tRNAs can span the distance between the peptidyl transferase center in the large subunit and the decoding center in the small subunit.

Binding Sites on Ribosomes for tRNA

- The 3' ends of the tRNAs that are coupled to the amino acid or to the growing peptide chain are adjacent to the large subunit.
- The anticodon loops of the bound tRNAs are located adjacent to the small subunit.

END

Initiation of Translation

- For translation to be successfully initiated, three events must occur:-
 - i) the ribosome must be recruited to the mRNA.
 - ii) a charged tRNA must be placed into the P-site of the ribosome.

Initiation of Translation

- iii) the ribosome must be precisely positioned over the start codon.

Initiation of Translation

- The correct positioning of the ribosome over the start codon is critical because this establishes the reading frame for the translation of the mRNA.

Initiation of Translation

- In prokaryotes, the assembly of the ribosome on an mRNA occurs one subunit at a time. The small subunit associates with the mRNA first.

Initiation of Translation

- In prokaryotes, the association of the small subunit with the mRNA is mediated by base-pairing interactions between the RBS and the 16S rRNA.

Initiation of Translation

- For ideally positioned RBSs, the small subunit is positioned on the mRNA such that the start codon will be in the P-site when the large subunit joins the complex.

Initiation of Translation

- The large subunit joins its partner only at the very end of the initiation process, just before the formation of the first peptide bond.
- Thus, many of the key events of translation initiation occur in the absence of the full ribosome.

Initiation of Translation

- Translation initiation is the only time a tRNA binds to the P-site without previously occupying the A-site. This event requires a special tRNA known as the **initiator tRNA**.

Initiation of Translation

- The initiator tRNA base-pairs with the start codon (AUG or GUG). AUG and GUG have a different meaning when they occur within an ORF, where they are read by tRNAs for methionine and valine, respectively.

Initiation of Translation

- Although the initiator tRNA is first charged with a methionine, a formyl group is rapidly added to the methionine amino group by a separate enzyme (**Met-tRNA transformylase**).

Initiation of Translation

- Thus rather than valine or methionine, the initiator tRNA is coupled to N-formyl methionine. The charged initiator tRNA is referred to as **fMet-tRNA^{fMet}**.

Initiation of Translation

- Because N-formyl methionine is the first amino acid to be incorporated into a polypeptide chain, one might think that all prokaryotic proteins have a formyl group at their amino termini.

Initiation of Translation

- This is not the case, however, because an enzyme known as a **deformylase** removes the formyl group from the amino terminus during or after the synthesis of the polypeptide chain.

Initiation of Translation

- In fact, many mature prokaryotic proteins do not even start with a methionine; aminopeptidases often remove the amino-terminal methionine as well as one or two additional amino acids.

END

The Initiation Factors

- The initiation of prokaryotic translation commences with the small subunit and is catalyzed by three translation initiation factors called IF1, IF2, and IF3.
- Each factor facilitates a key step in the initiation process.

The Initiation Factors

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

The Initiation Factors

- **IF2:**
- It is a GTPase that interacts with three key components of the initiation machinery: the small subunit, IF1, and the charged initiator tRNA (fMet-tRNA^{fMet}).

The Initiation Factors

- By interacting with these components, IF2 facilitates the association of fMet-tRNA^{fMet} with the small subunit and prevents other charged tRNAs from associating with the small subunit.

The Initiation Factors

- **IF3:**
- It binds to the small subunit and blocks it from re-associating with a large subunit. Because initiation requires a free small subunit, the binding of IF3 is critical for a new cycle of translation.

The Initiation Factors

- IF3 becomes associated with the small subunit at the end of a previous round of translation when it helps to dissociate the 70S ribosome into its large and small subunits.

The Initiation Factors

- Each of the initiation factors binds at, or near, one of the three tRNA binding sites on the small subunit.

The Initiation Factors

- Consistent with its role in blocking the binding of charged tRNAs to the A-site, IF1 binds directly to the portion of the small subunit that will become the A-site.

The Initiation Factors

- IF2 binds to IF1 and reaches over the A-site into the P-site to contact the fMet - tRNA^{fMet}.
- Finally, IF3 occupies the part of the small subunit that will become the E-site.

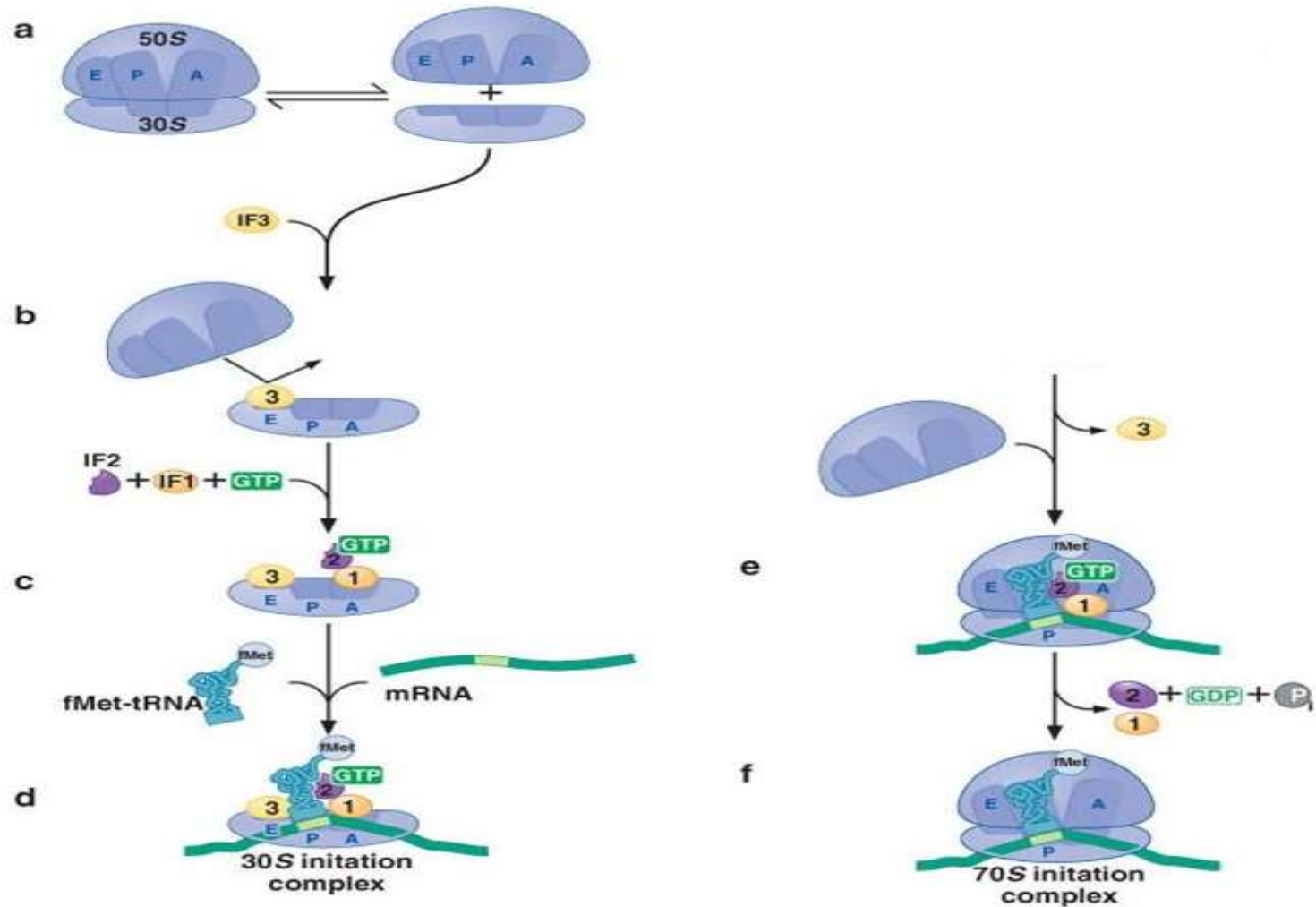
The Initiation Factors

- Thus, of the three potential tRNA-binding sites on the small subunit, only the P-site is capable of binding a tRNA in the presence of the initiation factors.

The Initiation Factors

- With all three initiation factors bound, the small subunit is prepared to bind to the mRNA and the initiator tRNA .
- These two RNAs can bind in either order and independently of each other.

The Initiation Factors



The Initiation Factors

- Binding fMet-tRNA^{fMet} to the small subunit is facilitated by its interactions with IF2 bound to GTP and base pairing between the anticodon and the start codon of the mRNA.

The Initiation Factors

- Similarly, base pairing between the fMet-tRNA^{fMet} and the mRNA serves to position the start codon in the P-site.

The Initiation Factors

- The last step of initiation involves the association of the large subunit to create the **70S initiation complex**.
- When the start codon and fMet-tRNA^{fMet} base-pair, the small subunit undergoes a change in conformation.

The Initiation Factors

- This altered conformation results in the release of IF3.
- In the absence of IF3, the large subunit is free to bind to the small subunit with its cargo of IF1, IF2, mRNA, and fMet-tRNA^{fMet}.

The Initiation Factors

- In particular, IF2 acts as an initial docking site of the large subunit, and this interaction subsequently stimulates the GTPase activity of IF2.GTP.

The Initiation Factors

- IF2 bound to GDP has reduced affinity for the ribosome and the initiator tRNA, leading to the release of IF2.GDP as well as IF1 from the ribosome.

The Initiation Factors

- Thus, the net result of initiation is the formation of an intact (70S) ribosome assembled at the start site of the mRNA with fMet-tRNA^{fMet} in the P-site and an empty A-site.

The Initiation Factors

- The ribosome – mRNA complex is now poised to accept a charged tRNA into the A-site and commence polypeptide synthesis.

END

Translation Elongation

- Once the ribosome is assembled with the charged initiator tRNA in the P site, polypeptide synthesis can begin.
- There are three key events that must occur for the correct addition of each amino acid.

Translation Elongation

- First, the correct aminoacyl-tRNA is loaded into the A site of the ribosome as dictated by the A-site codon.

Translation Elongation

- Second, 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.

Translation Elongation

- This peptidyl transferase reaction results in the transfer of the growing polypeptide from the tRNA in the P site to the amino acid moiety of the charged tRNA in the A site.

Translation Elongation

- Third, the resulting peptidyl-tRNA in the A site and its associated codon must be translocated to the P site so that the ribosome is poised for another cycle of codon recognition and peptide bond formation.

Translation Elongation

- As with the original positioning of the mRNA, this shift must occur precisely to maintain the correct reading frame of the message.
- Two auxiliary proteins known as **elongation factors** control these events.

Translation Elongation

- Both of these factors use the energy of GTP binding and hydrolysis to enhance the rate and accuracy of ribosome function.

Translation Elongation

- Unlike the initiation of translation, the mechanism of elongation is highly conserved between prokaryotic and eukaryotic cells.

Translation Elongation

- Aminoacyl-tRNAs do not bind to the ribosome on their own. Instead, they are "escorted" to the ribosome by the elongation factor EF-Tu.

Translation Elongation

- Once a tRNA is aminocylated, EF-Tu binds to the tRNA's 3' end, masking the coupled amino acid. This interaction prevents the bound aminoacyl-tRNA from participating in peptide bond formation until it is released from EF-Tu.

Translation Elongation

- Like the initiation factor IF2, the elongation factor EF-Tu binds and hydrolyzes GTP and the type of guanine nucleotide bound governs its function.

Translation Elongation

- EF-Tu can only bind to an aminoacyl-tRNA when it is associated with GTP. EF-Tu bound to GDP, or lacking any bound nucleotide, shows little affinity for aminoacyl-tRNAs.

Translation Elongation

- Thus, when EF-Tu hydrolyzes its bound GTP, any associated aminoacyl-tRNA is released.

Translation Elongation

- The trigger that activates the EF-Tu GTPase is the same domain on the large subunit of the ribosome that activates the IF2 GTPase when the large subunit joins the initiation complex.

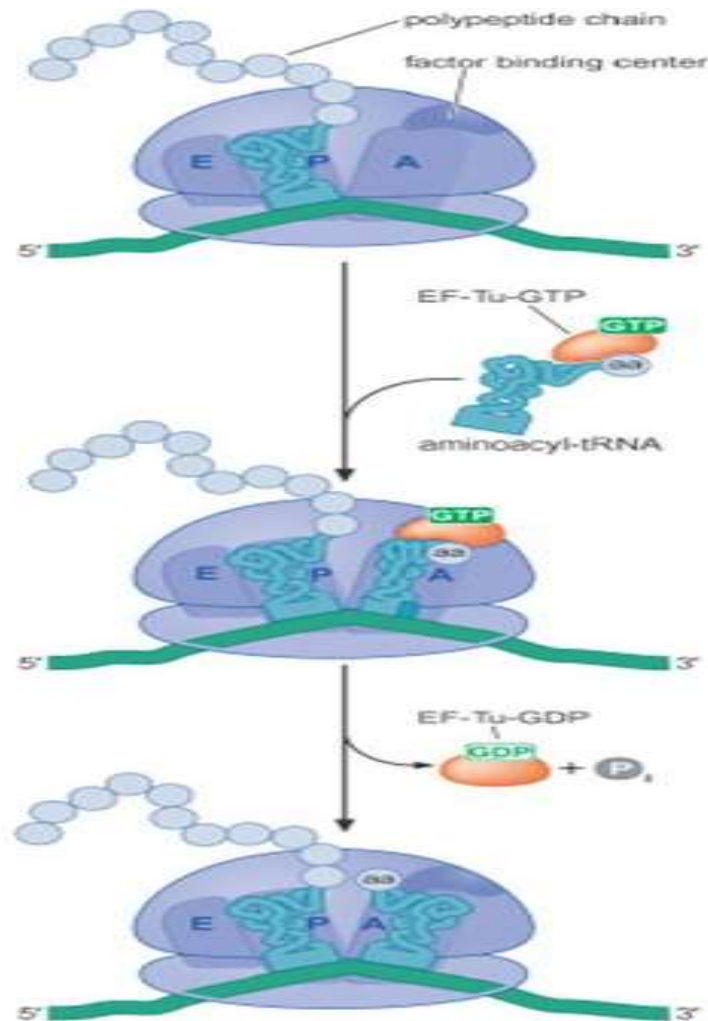
Translation Elongation

- This domain is known as the **factor binding center**.
- EF-Tu only interacts with the factor binding center after the tRNA is loaded into the A site and a correct codon-anticodon match is made.

Translation Elongation

- At this point, EF-Tu hydrolyzes its bound GTP and is released from the ribosome.
- The control of GTP hydrolysis by EF-Tu is critical to the specificity of translation.

Translation Elongation



Pearson Education, Inc., publishing as Benjamin

Translation Elongation

- The error rate of translation is between 10^{-3} to 10^{-4} .
- The ultimate basis for the selection of the correct aminoacyl-tRNA is the base pairing between the charged tRNA and the codon displayed in the A site of the ribosome.

Translation Elongation

- However, in some cases, the base pairing in the anticodon-codon interaction may be mismatched, yet the ribosome rarely allows such mismatched aminoacyl-tRNAs to continue in the translation process.

END

The Ribosome Is a Ribozyme

- Once the correctly charged tRNA has been placed in the A site and has rotated into the peptidyl transferase center, peptide bond formation takes place.

The Ribosome Is a Ribozyme

- This reaction is catalyzed by RNA, specifically the 23 S rRNA component of the large subunit.

The Ribosome Is a Ribozyme

- Early evidence for this came from experiments in which it was shown that a large subunit that had been largely stripped of its proteins was still able to carry out peptide bond formation.

The Ribosome Is a Ribozyme

- Proof that the peptidyl transferase is entirely composed of RNA has come from the high-resolution, three-dimensional structure of the ribosome, which reveals that no amino acid is located closer than 18 Å from the active site.

The Ribosome Is a Ribozyme

- Because catalysis requires distances in the 1 - 3 Å range, it is clear that the peptidyl transferase center is a ribozyme. That is an enzyme composed of RNA.

The Ribosome Is a Ribozyme

- How does the 23 S rRNA catalyze peptide bond formation?
- The exact mechanism remains to be determined, but some answers to this question are beginning to emerge.

The Ribosome Is a Ribozyme

- First, base-pairing between the 23 S rRNA and the CCA ends of the tRNAs in the A and the P sites help to position the alpha-amino group of the aminoacyl-tRNA to attack the carbonyl group of the growing polypeptide attached to the peptidyl-tRNA.

The Ribosome Is a Ribozyme

- These interactions are also likely to stabilize the aminoacyl-tRNA after accommodation.

The Ribosome Is a Ribozyme

- Because close proximity of substrates is rarely sufficient to generate high levels of catalysis, it is hypothesized that other elements of the ribosomal RNA change the chemical environment of the peptidyl transferase active site.

The Ribosome Is a Ribozyme

- For example, it has been proposed that nucleotides in the peptidyl transferase center accept a hydrogen from the alpha amino group of the aminoacyl-tRNA, making the associated nitrogen a stronger nucleophile.

The Ribosome Is a Ribozyme

- This is a common mechanism used by many proteins to stimulate nucleophilic attack of carbonyl groups.

END

Translocation in the Large Subunit

- Once the peptidyl transferase reaction has occurred, the tRNA in the P-site is deacetylated (no longer attached to an amino acid), and the growing polypeptide chain is linked to the tRNA in the A-site.

Translocation in the Large Subunit

- For a new round of peptide chain elongation to occur, the P-site tRNA must move to the E-site and the A-site tRNA must move to the P-site.
- At the same time, the mRNA must move by three nucleotides to expose the next codon.

Translocation in the Large Subunit

- These movements are coordinated within the ribosome and are collectively referred to as **translocation**.
- The initial steps of translocation are coupled to the peptidyl transferase reaction.

Translocation in the Large Subunit

- Once the growing peptide chain has been transferred to the A-site tRNA, the A- and P-site tRNAs have a preference to occupy new positions in the large subunit.

Translocation in the Large Subunit

- The 3' end of the A-site tRNA is bound to the growing polypeptide chain and prefers to bind in the P-site of the large subunit.

Translocation in the Large Subunit

- The now deacetylated P-site tRNA is no longer attached to the growing polypeptide chain and prefers to bind in the E-site of the large subunit.

Translocation in the Large Subunit

- In contrast, at this time, the anticodons of these tRNAs remain in their initial location in the small subunit bound to the mRNA.

Translocation in the Large Subunit

- Thus, translocation is initiated in the large subunit before the small subunit, and the tRNAs are said to be in “hybrid states.”

Translocation in the Large Subunit

- Their 3' ends have shifted into a new location, but their anticodon ends are still in their pre-peptidyl transfer position.

Translocation in the Large Subunit

- Importantly, this change is associated with a counter clockwise rotation of the small subunit relative to the large subunit facilitating interaction of the tRNAs with distinct tRNA-binding sites in the different subunits.

Translocation in the Large Subunit

- The completion of translocation requires the action of a second elongation factor called **EF-G**.
- Initial binding of EF-G to the ribosome occurs when associated with GTP.

Translocation in the Large Subunit

- After the peptidyl transferase reaction, EF-G–GTP binds to and stabilizes the ribosome in the rotated, hybrid state.

Translocation in the Large Subunit

- When EF-G–GTP binds, it contacts the factor-binding center of the large subunit, which stimulates GTP hydrolysis.
- GTP hydrolysis changes the conformation of EF-G with two consequences.

Translocation in the Large Subunit

- First, interactions between EF-G–GDP and the ribosome are thought to “unlock” the ribosome.

Translocation in the Large Subunit

- Structural studies reveal that there are “gates” that separate the A-, P-, and E-sites and EF-G–GDP is said to unlock the ribosome by opening these gates.

Translocation in the Large Subunit

- Second, the changed EF-G–GDP conformation binds to the A-site of the decoding center.
- This interaction competes with the tRNA for binding to the A-site of the decoding center.

Translocation in the Large Subunit

- Because the ribosome is unlocked, the formerly A-site tRNA can move into the P-site, allowing EF-G–GDP to bind the A-site.

Translocation in the Large Subunit

- Completion of translocation is accompanied by a clockwise rotation of the small subunit back to its starting position.
- The resulting ribosome structure has dramatically reduced affinity for EF-G-GDP.

Translocation in the Large Subunit

- Release of EF-G results in the return of the ribosome to a “locked” state in which the tRNAs and mRNA are once again tightly associated with the small subunit decoding center and the gates between the A-, P- and E-sites are closed.

Translocation in the Large Subunit

- Together, these events result in the translocation of the A-site tRNA into the P-site, the P-site tRNA into the E-site, and the movement of the mRNA by exactly 3 bp.
- The ribosome is now ready for a new cycle of amino acid addition to begin.

END

Termination of Translation

- The ribosome's cycle of aminoacyl-tRNA binding, peptide-bond formation, and translocation continues until one of the three stop codons enters the A-site.

Termination of Translation

- It was initially postulated that there would be one or more chain terminating tRNAs that would recognize these codons.
- However, this is not the case.

Termination of Translation

- Instead, stop codons are recognized by proteins called **release factors (RFs)** that activate the hydrolysis of the polypeptide from the peptidyl-tRNA.

Termination of Translation

- There are two classes of release factors.
- Class I release factors recognize the stop codons and trigger hydrolysis of the peptide chain from the tRNA in the P-site.

Termination of Translation

- Prokaryotes have two class I release factors called RF1 and RF2.
- RF1 recognizes the stop codon UAG and RF2 recognizes the stop codon UGA.
- The third stop codon, UAA, is recognized by both RF1 and RF2.

Termination of Translation

- In eukaryotic cells, there is a single class I release factor called **eRF1** that recognizes all three stop codons.
- Class II release factors stimulate the dissociation of the class I factors from the ribosome after release of the polypeptide chain.

Termination of Translation

- Prokaryotes and eukaryotes have only one class II factor called **RF3** and **eRF3**, respectively.
- Like EF-G, IF2, and EF-Tu, class II release factors are regulated by GTP binding and hydrolysis.

Termination of Translation

- How do release factors recognize stop codons?
- Because release factors are composed entirely of protein, protein–RNA interaction must mediate stop codon recognition.

Termination of Translation

- Experiments in which short coding regions were genetically swapped between RF1 and RF2 (having different stop-codon specificity) identified a three-amino-acid sequence that is critical for release factor specificity.

Termination of Translation

- Exchange of these three amino acids between RF1 and RF2 swaps the stop-codon specificity of the two complexes.

Termination of Translation

- For this reason, this three-amino-acid sequence is called a peptide anticodon and must interact with and recognize stop codons.
- A 3D structure of RF1 bound to the ribosome confirms that RF1 binds to the A-site of the ribosome.

Termination of Translation

- In this structure, the peptide anticodon is located very near the anticodon, but it is likely that there are additional protein regions that contribute to codon recognition.

Termination of Translation

- A region of class I release factors that stimulates polypeptide release has also been identified.
- All class I factors share a conserved three-amino-acid sequence (glycine, glycine, glutamine) that is essential for polypeptide release.

Termination of Translation

- Moreover, the structure of RF1 bound to the ribosome confirms that the GGQ motif is located in close proximity to the peptidyl transferase center.

Termination of Translation

- It remains unclear whether the GGG motif is directly involved in the release of polypeptide from the peptidyl-tRNA or it induces a change in the peptidyl transferase center that allows the center itself to catalyze hydrolysis.

Termination of Translation

- Studies of the conserved bases found adjacent to the CCA ends in the peptidyl transferase center indicate that several of these residues are required for peptide hydrolysis.

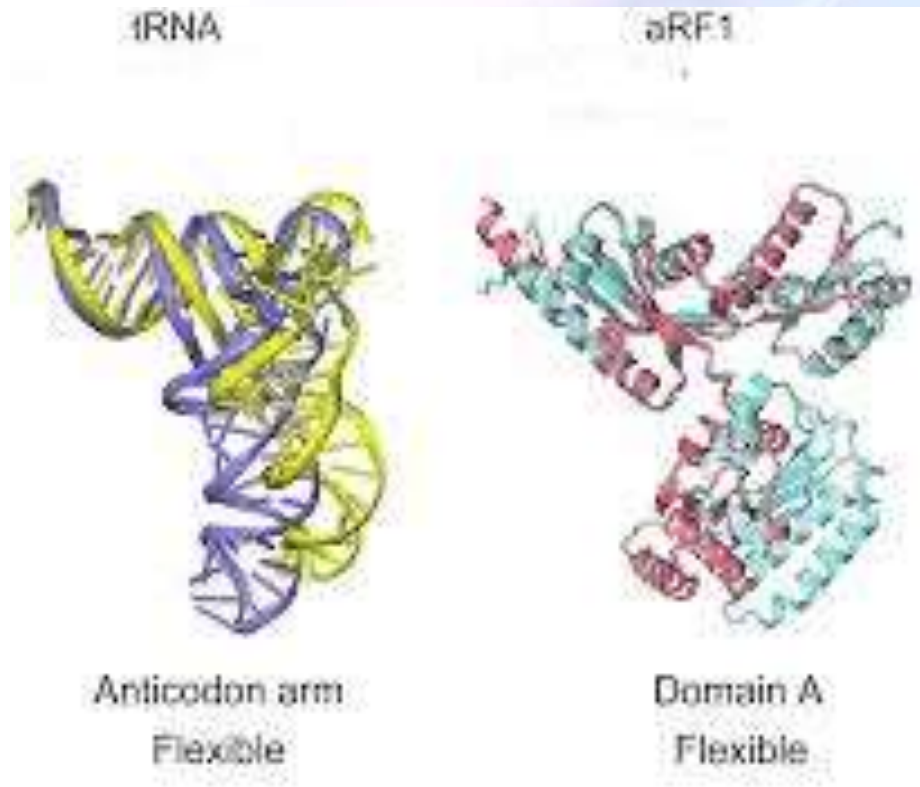
Termination of Translation

- Indeed, these bases appear to play a more important role in peptide release than they do in peptide-bond formation.

Termination of Translation

- Together, these studies have led to the hypothesis that class I release factors functionally mimic a tRNA, having a peptide anticodon that interacts with the stop codon and a GGQ motif that reaches into the peptidyl transferase center.

Termination of Translation



Termination of Translation

- Once the class I release factor has triggered the hydrolysis of the peptidyl tRNA linkage, it must be removed from the ribosome.

Termination of Translation

- This step is stimulated by the class II release factor, RF3.
- RF3 is a GTP-binding protein but, unlike the other GTP-binding proteins involved in translation, this factor has a higher affinity for GDP than GTP.

Termination of Translation

- Thus, free RF3 is predominantly in the GDP-bound form.
- RF3-GDP binds to the ribosome in a manner that depends on the presence of a class I release factor.

Termination of Translation

- After the class I release factor stimulates polypeptide release, a change in the conformation of the ribosome and the class I release factor stimulates RF3 to exchange its bound GDP for a GTP.

Termination of Translation

- The binding of GTP to RF3 leads to the formation of a high-affinity interaction with the ribosome that favors the rotated hybrid state.
- This change in conformation displaces the class I factor from the ribosome.

Termination of Translation

- These changes also allow RF3 to associate with the factor-binding center of the large subunit. As with other GTP-binding proteins involved in translation, this interaction stimulates the hydrolysis of GTP.

Termination of Translation

- In the absence of a bound class I factor, the resulting RF3.GDP has a low affinity for the ribosome and is released.

Termination of Translation

- The ribosome's cycle of aminoacyl-tRNA binding, peptide-bond formation, and translocation continues until one of the three stop codons enters the A-site.

Termination of Translation

- It was initially postulated that there would be one or more chain terminating tRNAs that would recognize these codons.
- However, this is not the case.

Termination of Translation

- Instead, stop codons are recognized by proteins called **release factors (RFs)** that activate the hydrolysis of the polypeptide from the peptidyl-tRNA.

Termination of Translation

- There are two classes of release factors.
- Class I release factors recognize the stop codons and trigger hydrolysis of the peptide chain from the tRNA in the P-site.

Termination of Translation

- Prokaryotes have two class I release factors called RF1 and RF2.
- RF1 recognizes the stop codon UAG and RF2 recognizes the stop codon UGA.
- The third stop codon, UAA, is recognized by both RF1 and RF2.

Termination of Translation

- In eukaryotic cells, there is a single class I release factor called **eRF1** that recognizes all three stop codons.
- Class II release factors stimulate the dissociation of the class I factors from the ribosome after release of the polypeptide chain.

Termination of Translation

- Prokaryotes and eukaryotes have only one class II factor called **RF3** and **eRF3**, respectively.
- Like EF-G, IF2, and EF-Tu, class II release factors are regulated by GTP binding and hydrolysis.

Termination of Translation

- How do release factors recognize stop codons?
- Because release factors are composed entirely of protein, protein–RNA interaction must mediate stop codon recognition.

Termination of Translation

- Experiments in which short coding regions were genetically swapped between RF1 and RF2 (having different stop-codon specificity) identified a three-amino-acid sequence that is critical for release factor specificity.

Termination of Translation

- Exchange of these three amino acids between RF1 and RF2 swaps the stop-codon specificity of the two complexes.

Termination of Translation

- For this reason, this three-amino-acid sequence is called a peptide anticodon and must interact with and recognize stop codons.
- A 3D structure of RF1 bound to the ribosome confirms that RF1 binds to the A-site of the ribosome.

Termination of Translation

- In this structure, the peptide anticodon is located very near the anticodon, but it is likely that there are additional protein regions that contribute to codon recognition.

Termination of Translation

- A region of class I release factors that stimulates polypeptide release has also been identified.
- All class I factors share a conserved three-amino-acid sequence (glycine, glycine, glutamine) that is essential for polypeptide release.

Termination of Translation

- Moreover, the structure of RF1 bound to the ribosome confirms that the GGQ motif is located in close proximity to the peptidyl transferase center.

Termination of Translation

- It remains unclear whether the GGG motif is directly involved in the release of polypeptide from the peptidyl-tRNA or it induces a change in the peptidyl transferase center that allows the center itself to catalyze hydrolysis.

Termination of Translation

- Studies of the conserved bases found adjacent to the CCA ends in the peptidyl transferase center indicate that several of these residues are required for peptide hydrolysis.

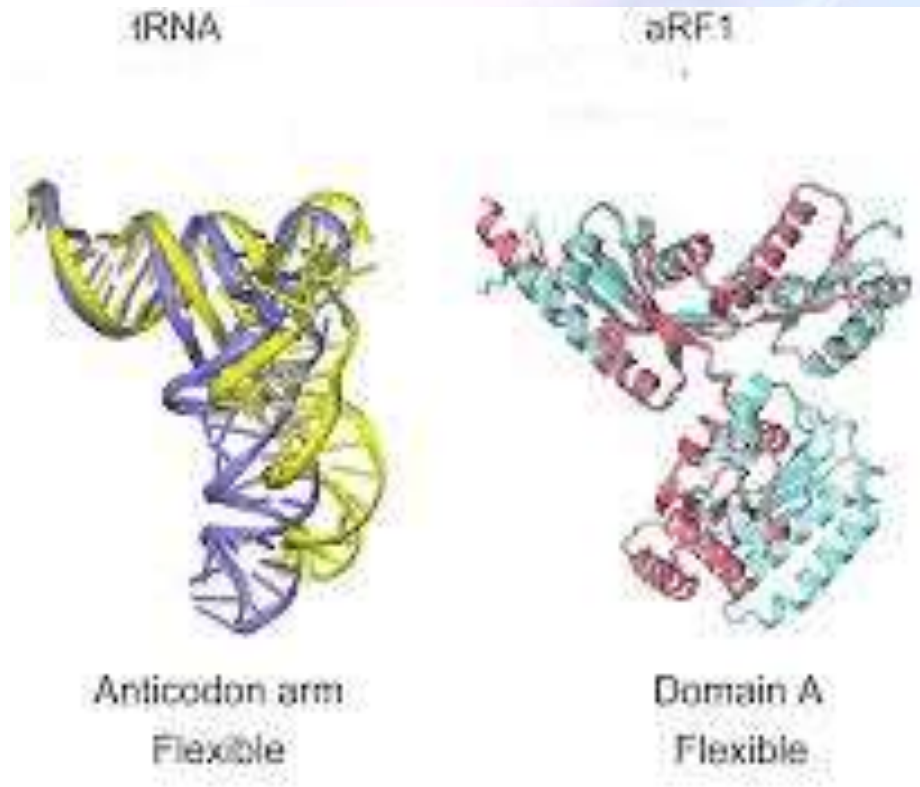
Termination of Translation

- Indeed, these bases appear to play a more important role in peptide release than they do in peptide-bond formation.

Termination of Translation

- Together, these studies have led to the hypothesis that class I release factors functionally mimic a tRNA, having a peptide anticodon that interacts with the stop codon and a GGQ motif that reaches into the peptidyl transferase center.

Termination of Translation



Termination of Translation

- Once the class I release factor has triggered the hydrolysis of the peptidyl tRNA linkage, it must be removed from the ribosome.

Termination of Translation

- This step is stimulated by the class II release factor, RF3.
- RF3 is a GTP-binding protein but, unlike the other GTP-binding proteins involved in translation, this factor has a higher affinity for GDP than GTP.

Termination of Translation

- Thus, free RF3 is predominantly in the GDP-bound form.
- RF3-GDP binds to the ribosome in a manner that depends on the presence of a class I release factor.

Termination of Translation

- After the class I release factor stimulates polypeptide release, a change in the conformation of the ribosome and the class I release factor stimulates RF3 to exchange its bound GDP for a GTP.

Termination of Translation

- The binding of GTP to RF3 leads to the formation of a high-affinity interaction with the ribosome that favors the rotated hybrid state.
- This change in conformation displaces the class I factor from the ribosome.

Termination of Translation

- These changes also allow RF3 to associate with the factor-binding center of the large subunit. As with other GTP-binding proteins involved in translation, this interaction stimulates the hydrolysis of GTP.

Termination of Translation

- In the absence of a bound class I factor, the resulting RF3.GDP has a low affinity for the ribosome and is released.

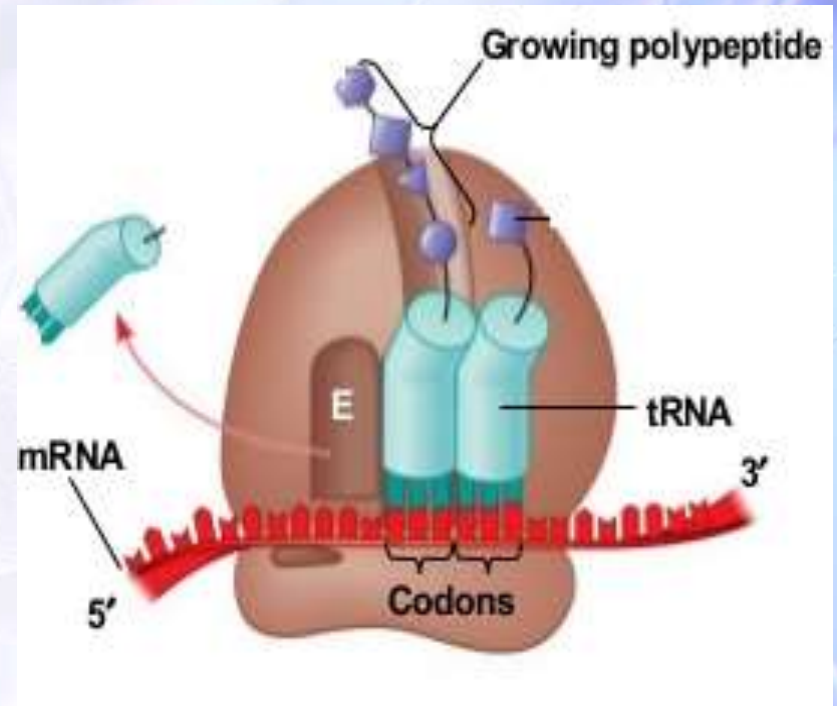
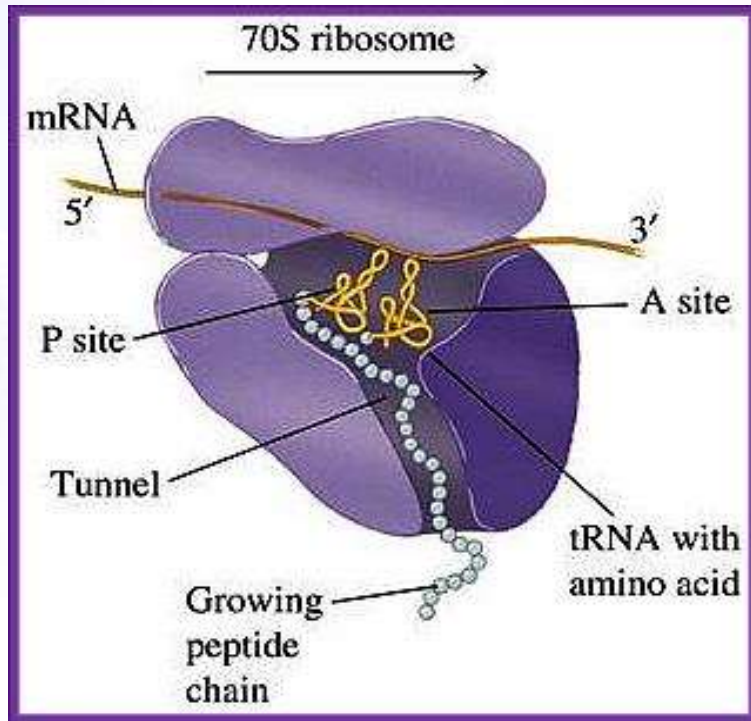
Nascent polypeptide processing & folding

- The nascent protein passes through a peptide exit tunnel that extends from the peptide transferase centre to the ribosome surface.

Nascent polypeptide processing & folding

- The exit tunnel in the bacterial ribosome is about 80-100 Å^o long and about 10 Å^o in its diameter at its narrowest point but widens to about twice that diameter at the rim of the exit pore.

Nascent polypeptide processing & folding



Nascent polypeptide processing & folding

- The exit tunnel can accommodate an α helix with about 60 residues or an extended peptide with about half that number of residues.

Nascent polypeptide processing & folding

- The space within the exit tunnel may permit the nascent polypeptide chain to assume an α -helical conformation but is too narrow to permit more extensive folding.

Nascent polypeptide processing & folding

- As the nascent peptide chain emerges from the tunnel, it interacts with enzymes that catalyze co-translational modifications, chaperones that assist in folding and prevent mis-folding,

Nascent polypeptide processing & folding

- and the signal recognition particle that facilitates transport across the cell membrane.
- Peptide deformylase and methionine aminopeptidase bind at the rim of the bacterial ribosome's exit pore.

Nascent polypeptide processing & folding

- The deformylase cleaves the N-terminal formyl group from the nascent polypeptide as it emerges from the exit tunnel.

Nascent polypeptide processing & folding

- Then the aminopeptidase recognizes about 60% of the different nascent polypeptides and removes their N-terminal methionine.

Nascent polypeptide processing & folding

- Nascent eukaryotic polypeptides do not have an N-terminal formyl group but do begin with methionine.
- Ribosome-bound methionine aminopeptidases remove the N-terminal methionine.

Nascent polypeptide processing & folding

- In bacteria, co-translational protein folding is assisted by a 48 kDa chaperone called the **trigger factor**, which binds at the bacterial ribosome's exit pore.

Nascent polypeptide processing & folding

- The trigger factor protein transiently associates with the L23 protein on the 50S subunit.
- Its residence time on the ribosome depends on whether the ribosome has a nascent protein in the exit tunnel.

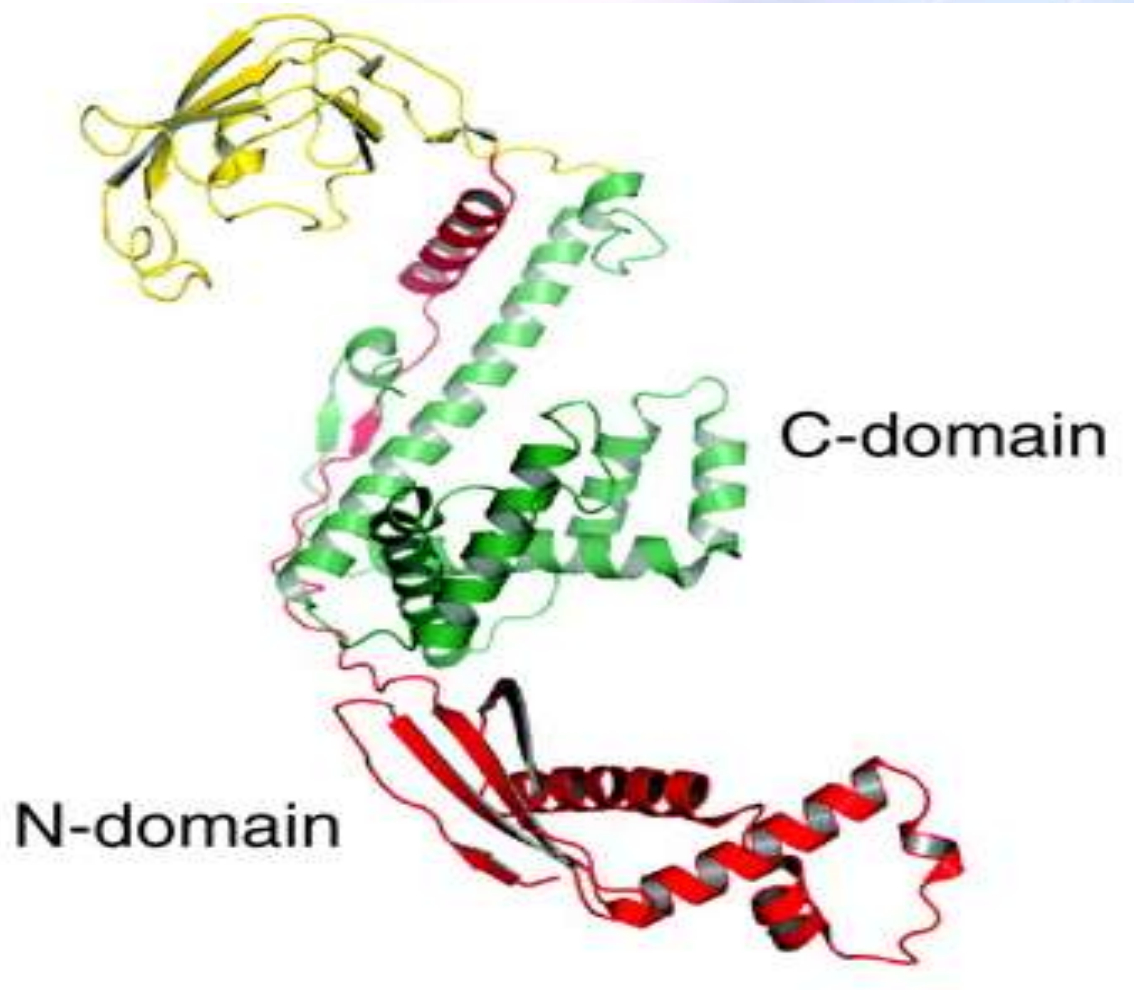
Nascent polypeptide processing & folding

- If a nascent protein is not present, the average residence time is about 11 – 15 seconds but this time increases several – fold when a nascent protein is present.

Nascent polypeptide processing & folding

- The trigger factor contains three domains that arrange to form a characteristic elongated dragon-shaped structure.
- The N-terminal domain forms the tail that binds to the 50S subunit.

Nascent polypeptide processing & folding



Nascent polypeptide processing & folding

- The middle domain which forms the dragon's head has peptidyl prolyl *cis/trans* isomerase activity that is not essential for the trigger factor's chaperone's function.

Nascent polypeptide processing & folding

- The C-terminal domain the central body of the dragon and is responsible for the trigger factor's chaperone activity.

Nascent polypeptide processing & folding

- The trigger factor binds to the hydrophobic patches as they emerge from the ribosome and sometimes remain associated with the segment even after polypeptide chain completion.

Nascent polypeptide processing & folding

- Moreover, a single nascent polypeptide chain or free polypeptide may have two or more trigger factors associated with it.

Nascent polypeptide processing & folding

- A chaperone also associates with the large subunit of the eukaryotic ribosome.
- This chaperone consists of three different subunits that differ in both sequence and structure from the trigger factor.

Nascent polypeptide processing & folding

- In yeast, deletion of any single subunit results in slow growth and cold sensitivity.
- Protein folding is very important for the normal functioning of the cells.

Nascent polypeptide processing & folding

- The accumulation of toxic proteins, which result from misfolding may lead to several neurological disorders including Alzheimer's, Huntington's and Parkinson's diseases.

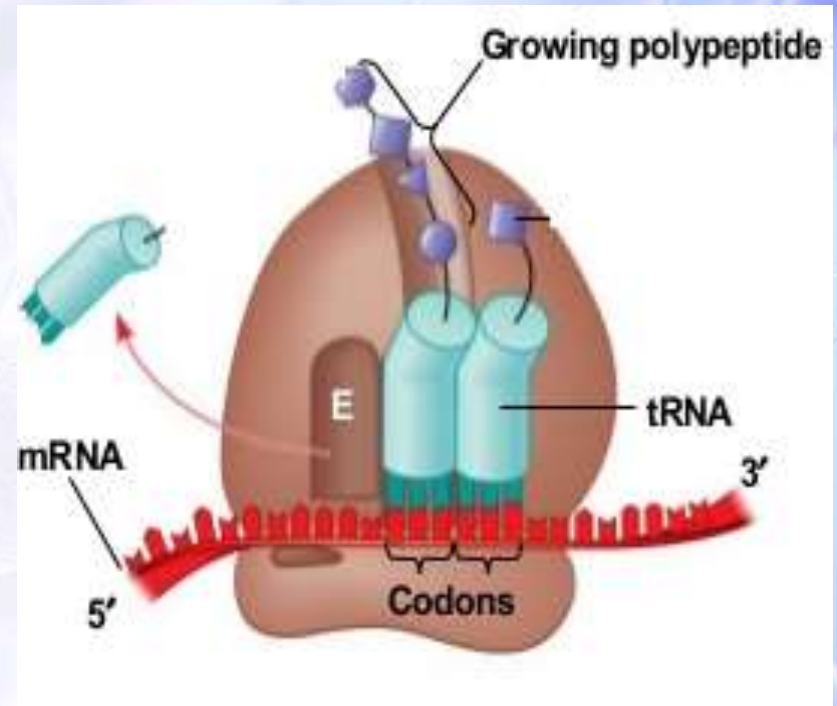
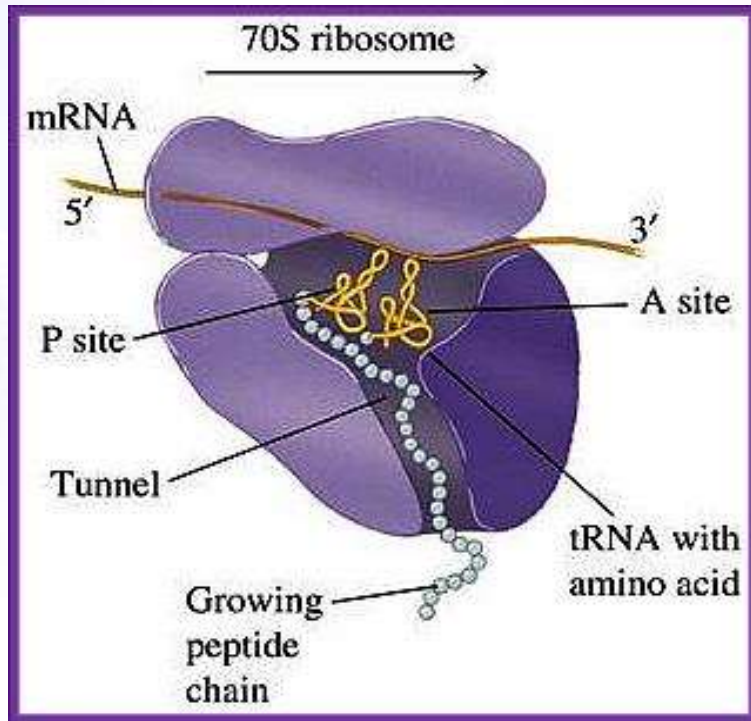
Nascent polypeptide processing & folding

- The nascent protein passes through a peptide exit tunnel that extends from the peptide transferase centre to the ribosome surface.

Nascent polypeptide processing & folding

- The exit tunnel in the bacterial ribosome is about 80-100 Å^o long and about 10 Å^o in its diameter at its narrowest point but widens to about twice that diameter at the rim of the exit pore.

Nascent polypeptide processing & folding



Nascent polypeptide processing & folding

- The exit tunnel can accommodate an α helix with about 60 residues or an extended peptide with about half that number of residues.

Nascent polypeptide processing & folding

- The space within the exit tunnel may permit the nascent polypeptide chain to assume an α -helical conformation but is too narrow to permit more extensive folding.

Nascent polypeptide processing & folding

- As the nascent peptide chain emerges from the tunnel, it interacts with enzymes that catalyze co-translational modifications, chaperones that assist in folding and prevent mis-folding,

Nascent polypeptide processing & folding

- and the signal recognition particle that facilitates transport across the cell membrane.
- Peptide deformylase and methionine aminopeptidase bind at the rim of the bacterial ribosome's exit pore.

Nascent polypeptide processing & folding

- The deformylase cleaves the N-terminal formyl group from the nascent polypeptide as it emerges from the exit tunnel.

Nascent polypeptide processing & folding

- Then the aminopeptidase recognizes about 60% of the different nascent polypeptides and removes their N-terminal methionine.

Nascent polypeptide processing & folding

- Nascent eukaryotic polypeptides do not have an N-terminal formyl group but do begin with methionine.
- Ribosome-bound methionine aminopeptidases remove the N-terminal methionine.

Nascent polypeptide processing & folding

- In bacteria, co-translational protein folding is assisted by a 48 kDa chaperone called the **trigger factor**, which binds at the bacterial ribosome's exit pore.

Nascent polypeptide processing & folding

- The trigger factor protein transiently associates with the L23 protein on the 50S subunit.
- Its residence time on the ribosome depends on whether the ribosome has a nascent protein in the exit tunnel.

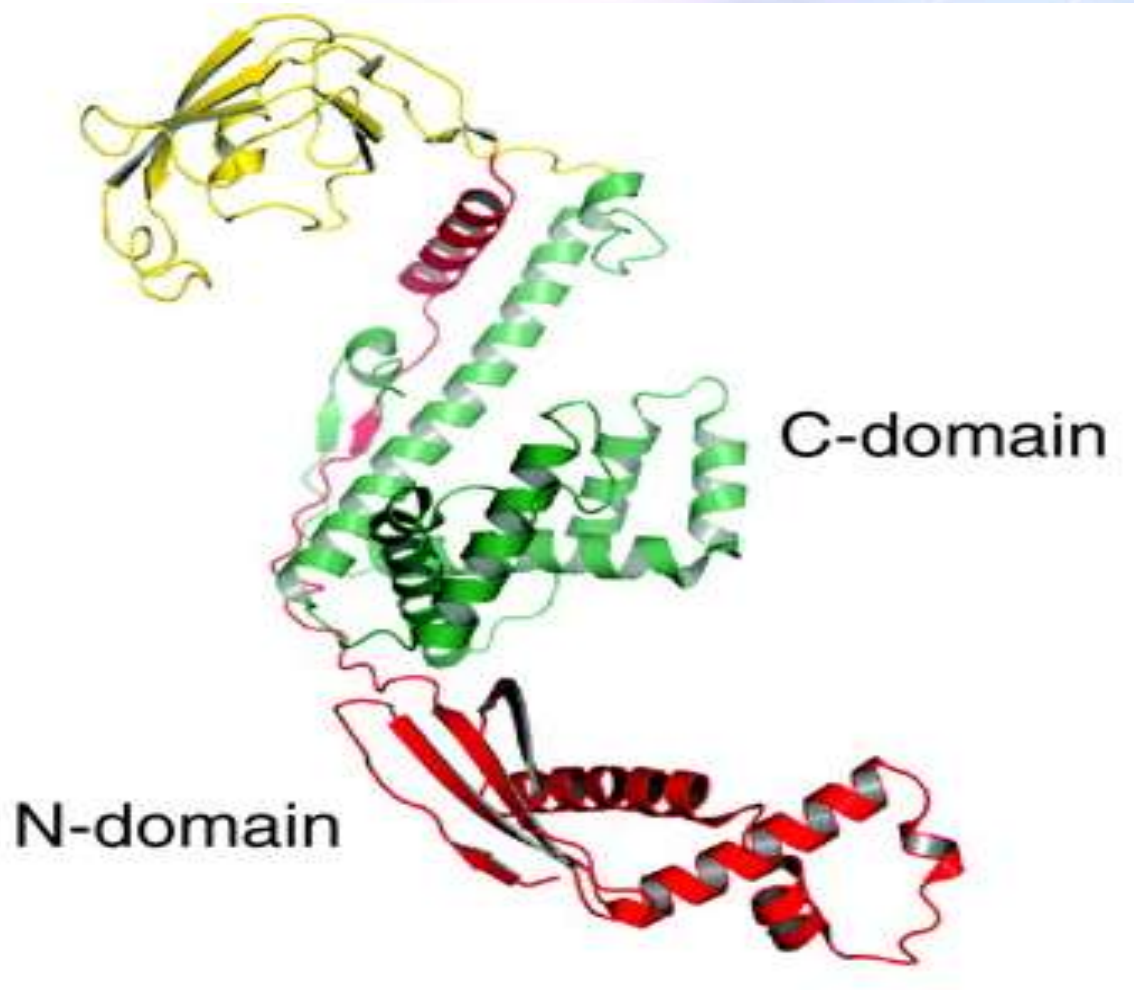
Nascent polypeptide processing & folding

- If a nascent protein is not present, the average residence time is about 11 – 15 seconds but this time increases several – fold when a nascent protein is present.

Nascent polypeptide processing & folding

- The trigger factor contains three domains that arrange to form a characteristic elongated dragon-shaped structure.
- The N-terminal domain forms the tail that binds to the 50S subunit.

Nascent polypeptide processing & folding



Nascent polypeptide processing & folding

- The middle domain which forms the dragon's head has peptidyl prolyl *cis/trans* isomerase activity that is not essential for the trigger factor's chaperone's function.

Nascent polypeptide processing & folding

- The C-terminal domain the central body of the dragon and is responsible for the trigger factor's chaperone activity.

Nascent polypeptide processing & folding

- The trigger factor binds to the hydrophobic patches as they emerge from the ribosome and sometimes remain associated with the segment even after polypeptide chain completion.

Nascent polypeptide processing & folding

- Moreover, a single nascent polypeptide chain or free polypeptide may have two or more trigger factors associated with it.

Nascent polypeptide processing & folding

- A chaperone also associates with the large subunit of the eukaryotic ribosome.
- This chaperone consists of three different subunits that differ in both sequence and structure from the trigger factor.

Nascent polypeptide processing & folding

- In yeast, deletion of any single subunit results in slow growth and cold sensitivity.
- Protein folding is very important for the normal functioning of the cells.

Nascent polypeptide processing & folding

- The accumulation of toxic proteins, which result from misfolding may lead to several neurological disorders including Alzheimer's, Huntington's and Parkinson's diseases.

Signal Sequence

- The biologists observed that some ribosomes appear to exist free in the cytoplasm while others are bound to the endoplasmic reticulum in eukaryotes or cell membrane in prokaryotes.

Signal Sequence

- The free ribosomes synthesize cytoplasmic and mitochondrial proteins whereas, the membrane-bound ribosomes synthesize integral membrane proteins, lysosomal proteins and secretory proteins.

Signal Sequence

- The reason why some ribosomes bind to the endoplasmic reticulum was not clear at first.
- One possibility was that cells have two distinct kinds of ribosomes, free and bound.

Signal Sequence

- Studies by Gunter Blobel and David D. Sabitini in 1971 indicated that free and bound ribosomes appeared to be the same.
- Therefore, there must be some other explanation for the same.

Signal Sequence

- Blobel and Sabitini proposed the signal hypothesis to explain how cells determine whether a protein will be synthesized on a free or membrane-bound ribosome.
- According to this hypothesis:-

Signal Sequence

- 1) Free and membrane-bound ribosomes are identical.
- 2) Protein synthesis always begins on free ribosomes.

Signal Sequence

- 3) Nascent secretory, trans-membrane, or lysosomal proteins have sequences of 20-30 amino acids at their amino terminus that act as signals to bind the ribosomes to the endoplasmic reticulum.

Signal Sequence

- Ribosomes only bind to the endoplasmic reticulum when synthesizing proteins with signal sequences.
- Although, signal sequences vary from one protein to another, certain common features can be recognized.

Signal Sequence

- Signal sequences can be divided into three parts:
- a) a short, positively charged N-terminal region,
- b) a central region containing 7-13 hydrophobic amino acid residues,

Signal Sequence

- c) and a more polar C-terminal region that includes a cleavage site.
- Only polypeptides that have signal sequence can be inserted into the endoplasmic reticulum membrane.

Signal Sequence

- The amino end of a nascent polypeptide with a signal sequence requires the assistance of three components to pass into the lumen of the endoplasmic reticulum or to be integrated into the membrane.

Signal Sequence

- These components are the signal recognition particle (SRP), the SRP receptor (SR), and a protein conducting channel or translocon.
- The components required for SRP cycle are as follows:-

Signal Sequence

- 1) Signal recognition particle (SRP) consists of six proteins (SRP9, 14, 19, 54, 68, and 72 named according to their molecular masses) and a 7S RNA.

Signal Sequence

- The SRP54 subunit is especially noteworthy because the guanine nucleotide binding site and the signal sequence binding site are located within it.

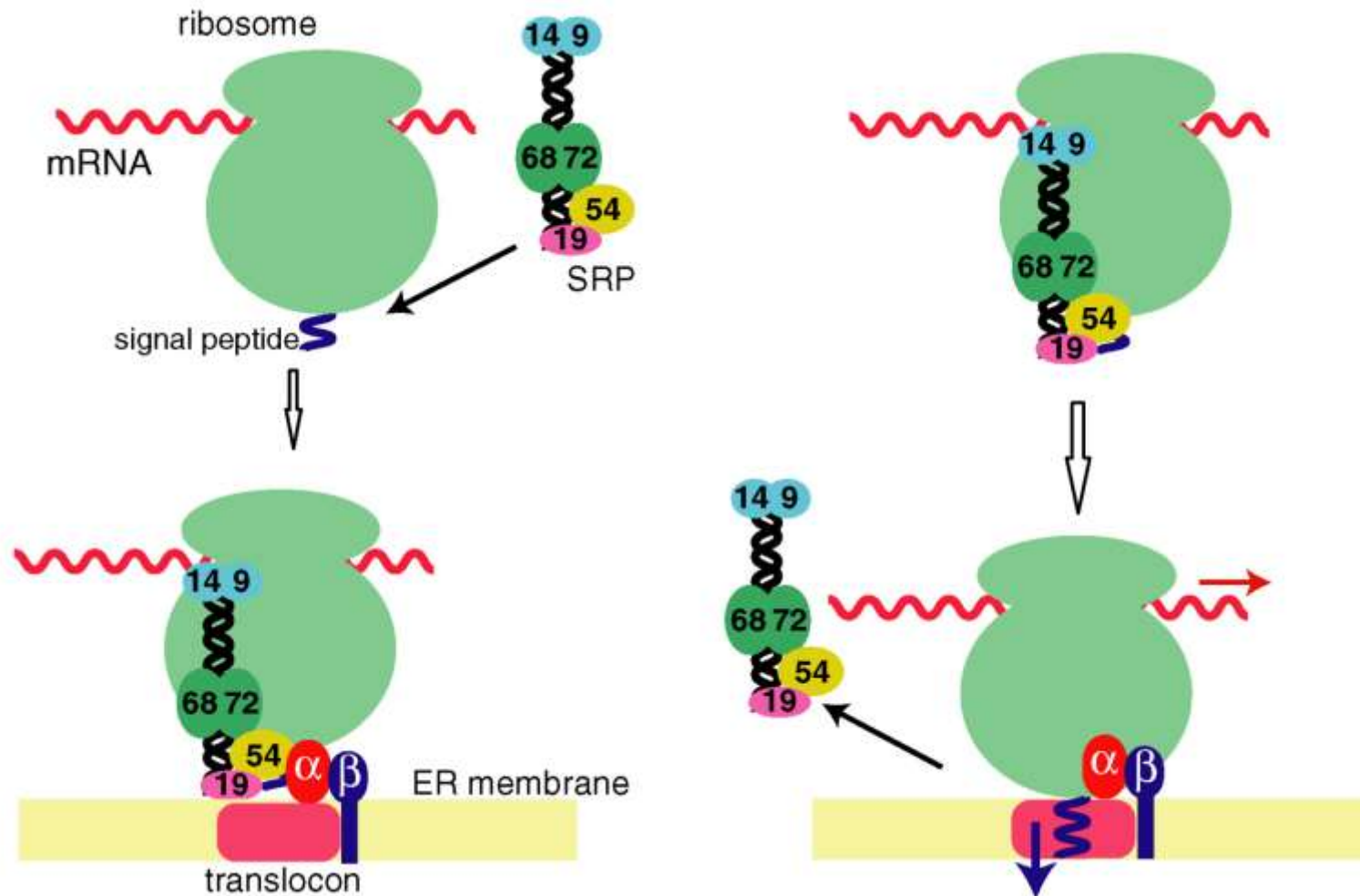
Signal Sequence

- 2) SRP receptors contains two GTP-binding polypeptide subunits, a peripheral membrane protein $SR\alpha$ and an integral membrane protein SR.

Signal Sequence

- 3) Translocon. The protein conducting channel, which is made of three polypeptides (Sec61 α , Sec61 β and Sec61 γ) acts as a passive conduit for polypeptides.

Signal Sequence



Signal Sequence

- Secretory and lysosomal proteins pass completely through the endoplasmic reticulum membrane and are directed to their ultimate destination by biochemical modification such as

Signal Sequence

- glycosylation that take place in the lumen of the endoplasmic reticulum and in the Golgi apparatus.
- Integral membrane proteins have one or more hydrophobic sequences that

Signal Sequence

- allow them to move laterally across the translocon and into the membrane, causing the polypeptide chain to embed itself within the membrane.

Regulation by RNAs in bacteria

- Small RNAs are involved in regulating the replication of plasmids, and others are involved in regulating gene expression.

Regulation by RNAs in bacteria

- Some of these RNAs control transcription, for example the 6S RNA of *Escherichia coli*.
- This RNA binds to the σ^{70} subunit of RNA polymerase and down-regulates transcription from many σ^{70} promoters.

Regulation by RNAs in bacteria

- The 6S RNA accumulates at high levels in stationary phase of bacterial growth.
- In stationary phase, an alternative σ factor, σ^S , is made.

Regulation by RNAs in bacteria

- This σ factor competes with σ^{70} for core polymerase and directs the enzyme to promoters expressing genes for the multiple stress responses needed to survive stationary phase.

Regulation by RNAs in bacteria

- By down-regulating transcription from σ^{70} promoters, 6S RNA helps this shift in expression to σ^S promoters.

Regulation by RNAs in bacteria

- In recent years, attention has focused on small RNA molecules in bacteria that regulate translation and mRNA degradation.

Regulation by RNAs in bacteria

- Interest in these small RNAs (small interfering and microRNAs) has been heightened by their similarity to RNAs that regulate gene expression in eukaryotes.

Regulation by RNAs in bacteria

- One class of bacterial regulatory RNAs (called sRNAs) acts in *trans* to control translation of target genes, rather as microRNAs do in eukaryotes.

Regulation by RNAs in bacteria

- They are, however, larger (80-110 nucleotides) than those eukaryotic regulatory RNAs (21 to 30 nucleotides), and they are not generally formed by processing of larger double-stranded RNA (dsRNA) precursors

Regulation by RNAs in bacteria

- (as those eukaryotic RNA regulators are); instead, they are encoded in their final form by small genes.
- Many of these genes have been identified by bioinformatics, with more than 100 sRNAs being uncovered in *E. coli*.

Regulation by RNAs in bacteria

- Most sRNAs work by base pairing with complementary sequences within target mRNAs and directing destruction of the mRNA, inhibiting its translation or even in some cases stimulating translation.

Regulation by RNAs in bacteria

- Binding of an sRNA to its target mRNA is in most cases aided by the bacterial protein Hfq.

Regulation by RNAs in bacteria

- This RNA chaperone is needed because the complementarity between the sRNAs and their target mRNAs is typically imperfect and short, and thus their interaction is weak.

Regulation by RNAs in bacteria

- Hfq facilitates base pairing.
- Also, by binding the sRNAs even before they are paired with their targets, Hfq increases the stability of these regulators.
- A well-studied sRNA from *E. coli* is the 81-nucleotide RybB RNA.

Regulation by RNAs in bacteria

- This sRNA binds several target mRNAs and triggers their destruction because the double-strand stretch of heteroduplex formed upon pairing is recognized as a substrate by the nuclease RNase E.

Regulation by RNAs in bacteria

- Most of the mRNAs targeted by RybB encode iron storage proteins. Free iron is required by the cell under certain circumstances, but high levels are toxic.

Regulation by RNAs in bacteria

- RybB regulates the levels of free iron by controlling the levels of iron storage proteins.
- RybB is expressed from a promoter recognized by a special σ factor called σ^E (like σ^S , a stress response σ factor).

Regulation by RNAs in bacteria

- Expression of the gene encoding σ^E is itself regulated by RybB, and thus this sRNA is part of an autonegative regulatory loop for σ^E .

Regulation by RNAs in bacteria

- The stationary-phase σ factor σ^S is encoded by the *rpoS* gene of *E. coli*.
- Translation of *rpoS* mRNA is stimulated by two sRNAs: DsrA and RprA.

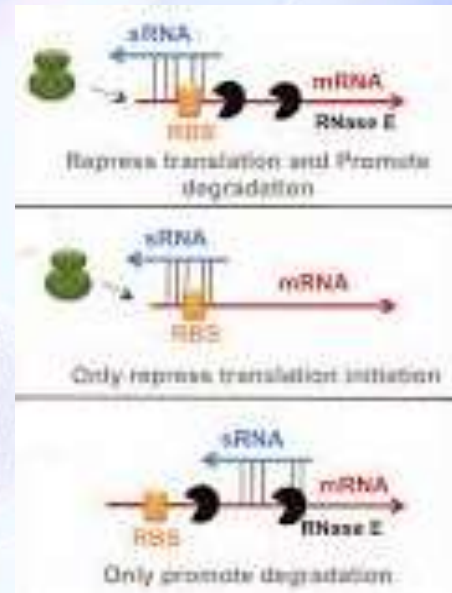
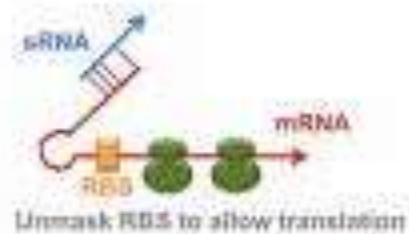
Regulation by RNAs in bacteria

- Activation is achieved by a switch in alternative RNA base pairing: the small RNAs bind to a region of the mRNA that otherwise would pair with the ribosome-binding site, inhibiting translation.

Regulation by RNAs in bacteria

- The rpoS gene is also acted on negatively by another small RNA, OxyS.

Regulation by RNAs in bacteria



Regulation by RNAs in bacteria

- Other regulatory RNAs act simply as “antisense” RNAs: they are encoded by the strand opposite the coding strand of a gene and act through homologous base pairing to inhibit expression of the mRNA produced from that gene.

Regulation by RNAs in bacteria

- These tend to be associated with genes encoding potentially toxic products, and also in regulation of some phage genes (as in λ).

Regulation by RNAs in bacteria

- These RNAs are often said to act in *cis* because they act only on the gene from which they are made (in contrast to the *trans*-acting sRNAs).

Regulation by RNAs in bacteria

- Let us consider other examples of gene regulation mediated through alternative RNA pairing that truly operate in *cis*.
- These are RNA regulatory elements that control expression of the genes within whose mRNAs they reside.

Regulation by RNAs in bacteria

- The most striking examples are the so-called that control metabolic operons and in biosynthetic operons.

Regulation by RNAs in bacteria

- The *trp* genes of *E. coli* are the classic examples of the latter mechanism and are where RNA-mediated regulation was discovered.

Regulation by RNAs in bacteria

- Small RNAs are involved in regulating the replication of plasmids, and others are involved in regulating gene expression.

Regulation by RNAs in bacteria

- Some of these RNAs control transcription, for example the 6S RNA of *Escherichia coli*.
- This RNA binds to the σ^{70} subunit of RNA polymerase and down-regulates transcription from many σ^{70} promoters.

Regulation by RNAs in bacteria

- The 6S RNA accumulates at high levels in stationary phase of bacterial growth.
- In stationary phase, an alternative σ factor, σ^S , is made.

Regulation by RNAs in bacteria

- This σ factor competes with σ^{70} for core polymerase and directs the enzyme to promoters expressing genes for the multiple stress responses needed to survive stationary phase.

Regulation by RNAs in bacteria

- By down-regulating transcription from σ^{70} promoters, 6S RNA helps this shift in expression to σ^S promoters.

Regulation by RNAs in bacteria

- In recent years, attention has focused on small RNA molecules in bacteria that regulate translation and mRNA degradation.

Regulation by RNAs in bacteria

- Interest in these small RNAs (small interfering and microRNAs) has been heightened by their similarity to RNAs that regulate gene expression in eukaryotes.

Regulation by RNAs in bacteria

- One class of bacterial regulatory RNAs (called sRNAs) acts in *trans* to control translation of target genes, rather as microRNAs do in eukaryotes.

Regulation by RNAs in bacteria

- They are, however, larger (80-110 nucleotides) than those eukaryotic regulatory RNAs (21 to 30 nucleotides), and they are not generally formed by processing of larger double-stranded RNA (dsRNA) precursors

Regulation by RNAs in bacteria

- (as those eukaryotic RNA regulators are); instead, they are encoded in their final form by small genes.
- Many of these genes have been identified by bioinformatics, with more than 100 sRNAs being uncovered in *E. coli*.

Regulation by RNAs in bacteria

- Most sRNAs work by base pairing with complementary sequences within target mRNAs and directing destruction of the mRNA, inhibiting its translation or even in some cases stimulating translation.

Regulation by RNAs in bacteria

- Binding of an sRNA to its target mRNA is in most cases aided by the bacterial protein Hfq.

Regulation by RNAs in bacteria

- This RNA chaperone is needed because the complementarity between the sRNAs and their target mRNAs is typically imperfect and short, and thus their interaction is weak.

Regulation by RNAs in bacteria

- Hfq facilitates base pairing.
- Also, by binding the sRNAs even before they are paired with their targets, Hfq increases the stability of these regulators.
- A well-studied sRNA from *E. coli* is the 81-nucleotide RybB RNA.

Regulation by RNAs in bacteria

- This sRNA binds several target mRNAs and triggers their destruction because the double-strand stretch of heteroduplex formed upon pairing is recognized as a substrate by the nuclease RNase E.

Regulation by RNAs in bacteria

- Most of the mRNAs targeted by RybB encode iron storage proteins. Free iron is required by the cell under certain circumstances, but high levels are toxic.

Regulation by RNAs in bacteria

- RybB regulates the levels of free iron by controlling the levels of iron storage proteins.
- RybB is expressed from a promoter recognized by a special σ factor called σ^E (like σ^S , a stress response σ factor).

Regulation by RNAs in bacteria

- Expression of the gene encoding σ^E is itself regulated by RybB, and thus this sRNA is part of an autonegative regulatory loop for σ^E .

Regulation by RNAs in bacteria

- The stationary-phase σ factor σ^S is encoded by the *rpoS* gene of *E. coli*.
- Translation of *rpoS* mRNA is stimulated by two sRNAs: DsrA and RprA.

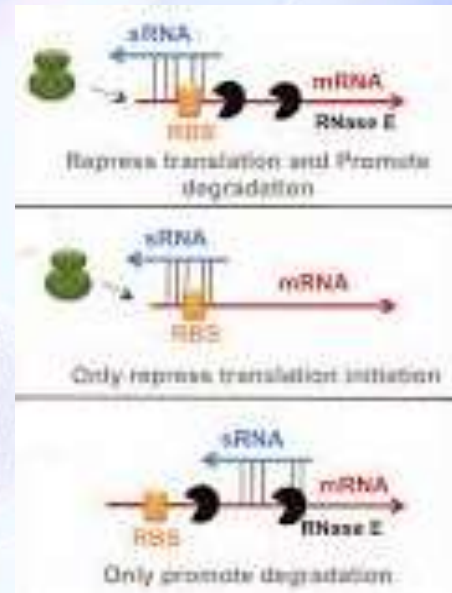
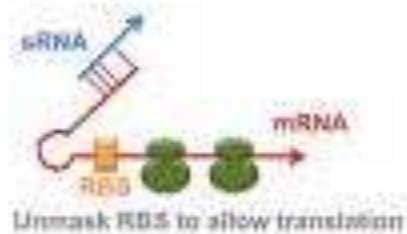
Regulation by RNAs in bacteria

- Activation is achieved by a switch in alternative RNA base pairing: the small RNAs bind to a region of the mRNA that otherwise would pair with the ribosome-binding site, inhibiting translation.

Regulation by RNAs in bacteria

- The rpoS gene is also acted on negatively by another small RNA, OxyS.

Regulation by RNAs in bacteria



Regulation by RNAs in bacteria

- Other regulatory RNAs act simply as “antisense” RNAs: they are encoded by the strand opposite the coding strand of a gene and act through homologous base pairing to inhibit expression of the mRNA produced from that gene.

Regulation by RNAs in bacteria

- These tend to be associated with genes encoding potentially toxic products, and also in regulation of some phage genes (as in λ).

Regulation by RNAs in bacteria

- These RNAs are often said to act in *cis* because they act only on the gene from which they are made (in contrast to the *trans*-acting sRNAs).

Regulation by RNAs in bacteria

- Let us consider other examples of gene regulation mediated through alternative RNA pairing that truly operate in *cis*.
- These are RNA regulatory elements that control expression of the genes within whose mRNAs they reside.

Regulation by RNAs in bacteria

- The most striking examples are the so-called that control metabolic operons and in biosynthetic operons.

Regulation by RNAs in bacteria

- The *trp* genes of *E. coli* are the classic examples of the latter mechanism and are where RNA-mediated regulation was discovered.

Riboswitches

- Riboswitches control gene expression in response to changes in the concentrations of small molecules.
- They do so through changes in RNA secondary structure.

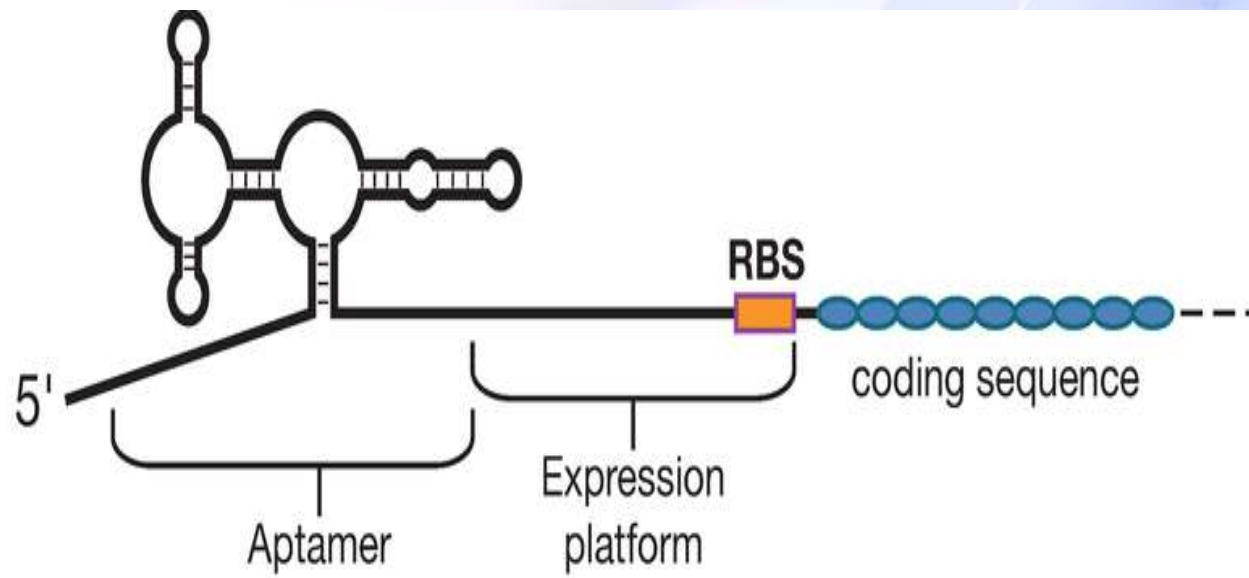
Riboswitches

- These regulatory elements are typically found within the 5'-untranslated regions (5'-UTRs) of the genes they control.
- They can regulate expression at the level of transcription or translation.

Riboswitches

- Each riboswitch is made up of two components:-
- the **aptamer** and,
- the **expression platform**.

Riboswitches



Riboswitches

- The aptamer binds the small-molecule ligand and, in response, undergoes a conformational change, which, in turn, causes a change in the secondary structure of the adjoining expression platform.

Riboswitches

- These conformational changes alter expression of the associated gene by either terminating transcription or inhibiting the initiation of translation.

Riboswitches

- Riboswitches are typically found upstream of genes involved in the synthesis of the metabolite ligand recognized by the riboswitch in question.

Riboswitches

- For example, in *Bacillus subtilis*, many genes involved in the use of the amino acid methionine have a 200-nucleotide-long untranslated leader RNA that acts as a SAM (S-adenosyl methionine) – sensing riboswitch.

Riboswitches

- RNA polymerase initiates transcription at the promoter and transcribes through this leader region before entering the coding sequence of the downstream genes.

Riboswitches

- Once transcribed into RNA, the leader region can adopt alternative structures through alternative patterns of intramolecular base pairing.
- One arrangement includes a stem-loop transcriptional Terminator .

Riboswitches

- SAM – the ligand for this riboswitch – binds to the aptamer and stabilizes the secondary structure that includes this transcriptional terminator.

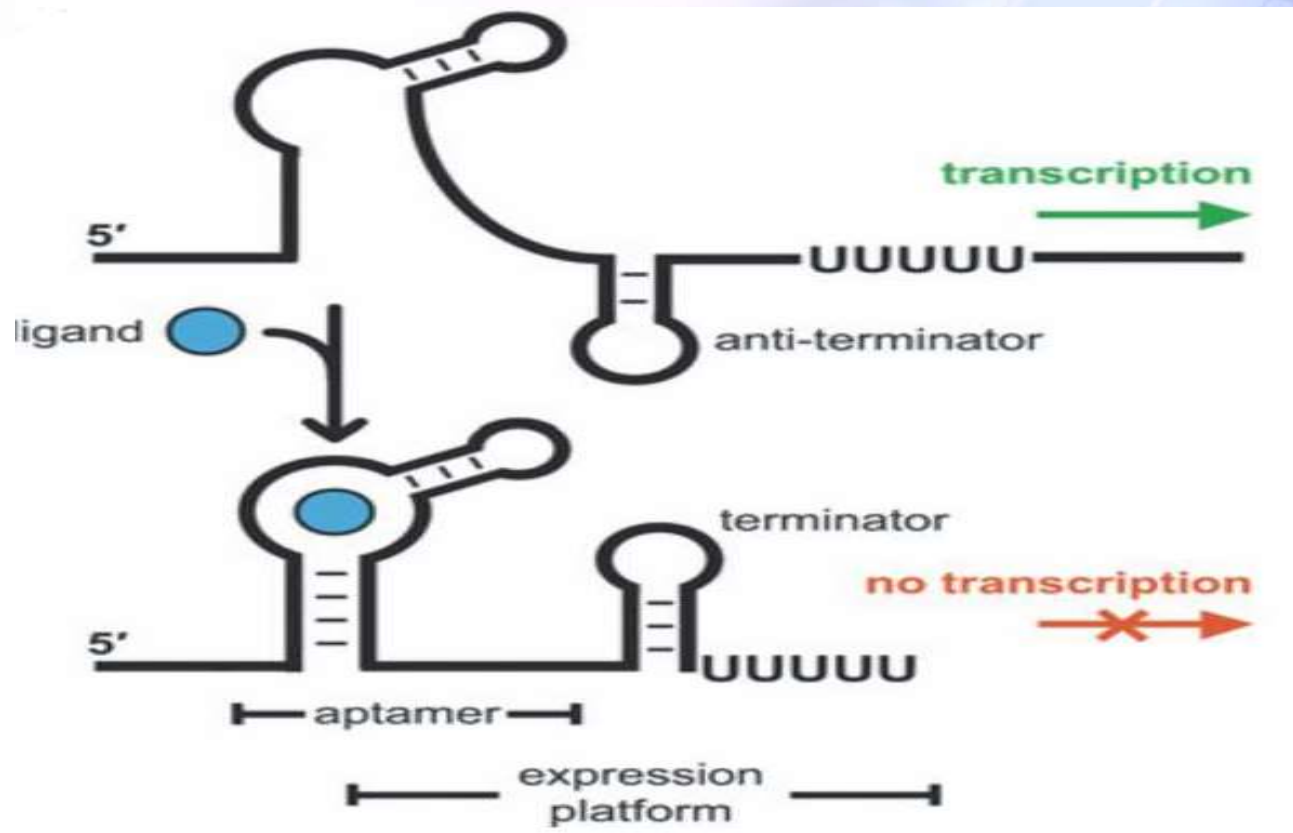
Riboswitches

- Under these circumstances, transcription is terminated before polymerase has a chance to transcribe the downstream protein-coding segment of the gene.

Riboswitches

- This form of transcriptional regulation is also called **attenuation**.

Riboswitches



Riboswitches

- In another case - at another gene - a SAM-sensing riboswitch can work by regulating translation.

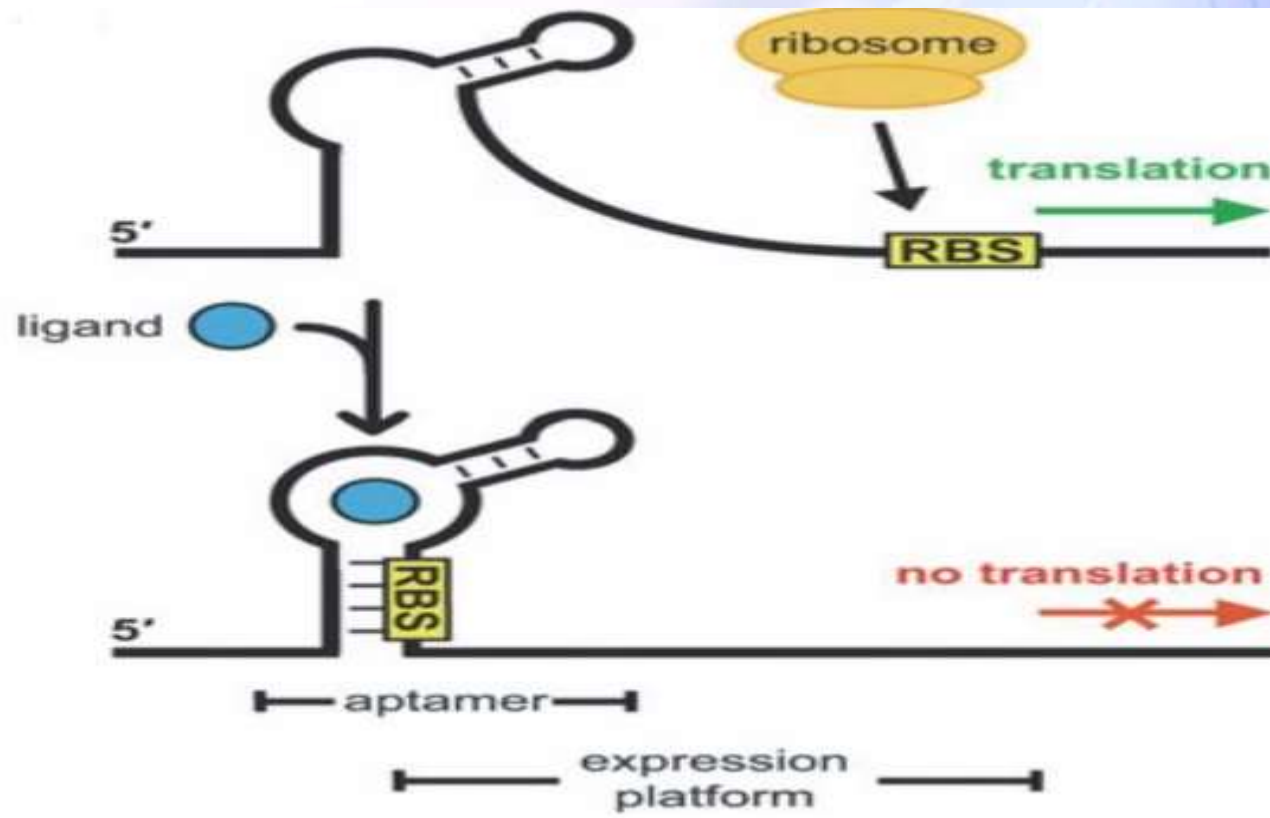
Riboswitches

- In that case, the alternative secondary structure stabilized by SAM binding to the aptamer includes a stem-loop that, although not a transcriptional terminator, does include the ribosome-binding site.

Riboswitches

- This conformational change sequesters the RBS and blocks ribosomes from initiating translation.
- This form of translation inhibition is thus essentially identical to that described for trans-acting sRNAs.

Riboswitches



Riboswitches

- Many riboswitches have been identified, and current whole-genome sequencing results suggest there are probably many hundreds or thousands found across bacterial species.

Riboswitches

- Even well-characterized examples respond to a range of different metabolites, including lysine, vitamin B12, coenzyme thiamine pyrophosphate (TPP), flavin mononucleotide (FMN), and guanine.

Riboswitches

- Although most prevalent in bacteria, riboswitches are found in other organisms as well, including archaea, fungi, and plants.

Riboswitches

- In some cases in these higher organisms, riboswitches are even involved in controlling alternative splicing.

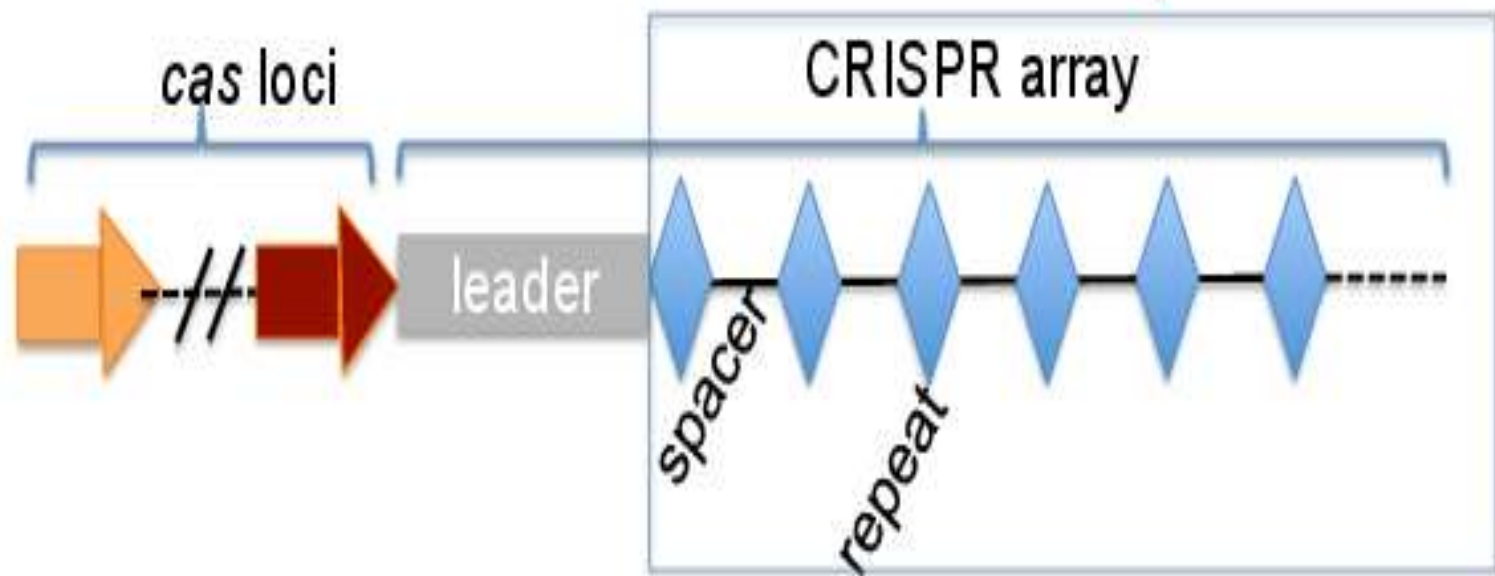
Riboswitches

- For example, in case of the fungus *Neurospora crassa*, three TPP aptamers were identified, two of which inhibited, and the third stimulated, expression of genes through regulation of RNA splicing.

CRISPRs

- The particular stretches of unusual but characteristically organized sequence were noticed in the genomes of several bacteria which were named as of Clustered Regularly Interspaced Short Palindromic Repeats (or CRISPRs).

CRISPRs



CRISPRs

- They consist of **repeated sequences** (each 30 bp long and highly conserved within a given cluster) interleaved with **spacer sequences** of similar length but highly divergent sequence

CRISPRs

- At one end of the array is a so-called leader sequence, often A-T rich and 500 bp in length.
- CRISPRs have been found in half of all bacterial genomes sequenced, and essentially all genomes of Archaea.

CRISPRs

- In many cases, there is only one cluster per genome, but not uncommonly there are more and the number can range up to 20 or more - and in one case almost 400 were detected in a species of *Chloroflexus*.

CRISPRs

- How do they arise, and what do they do?
- The first clue to their origin came from the striking observation that a significant number of the spacer sequences were identical to regions of known phage or plasmids.

CRISPRs

- This quickly led to the proposal that these arrays are involved in some sort of defense mechanism against foreign nucleic acids entering the cell.

CRISPRs

- Experimental support for this model came when resistant bacterial cells that arose in populations challenged with a given phage were found to have incorporated spacer sequences derived from that phage.

CRISPRs

- Likewise, reduced sensitivity to a phage could be conferred or revoked by addition or removal of relevant spacer sequences.

CRISPRs

- Furthermore, bacteria were increasingly insensitive to infection by a given phage the more spacer sequences they acquired from that phage.
- A set of conserved protein-coding genes is tightly associated with the CRISPR sequences.

CRISPRs

- The two most highly conserved members (*cas1* and *cas2*, for “CRISPR associated”) are found in all cases, but other *cas* genes, and more distantly conserved genes, are less so.

CRISPRs

- These genes encode proteins involved in different aspects of CRISPR function.
- Acquisition by a cell of spacer regions from a given phage confers decreased sensitivity to further infection by that phage.

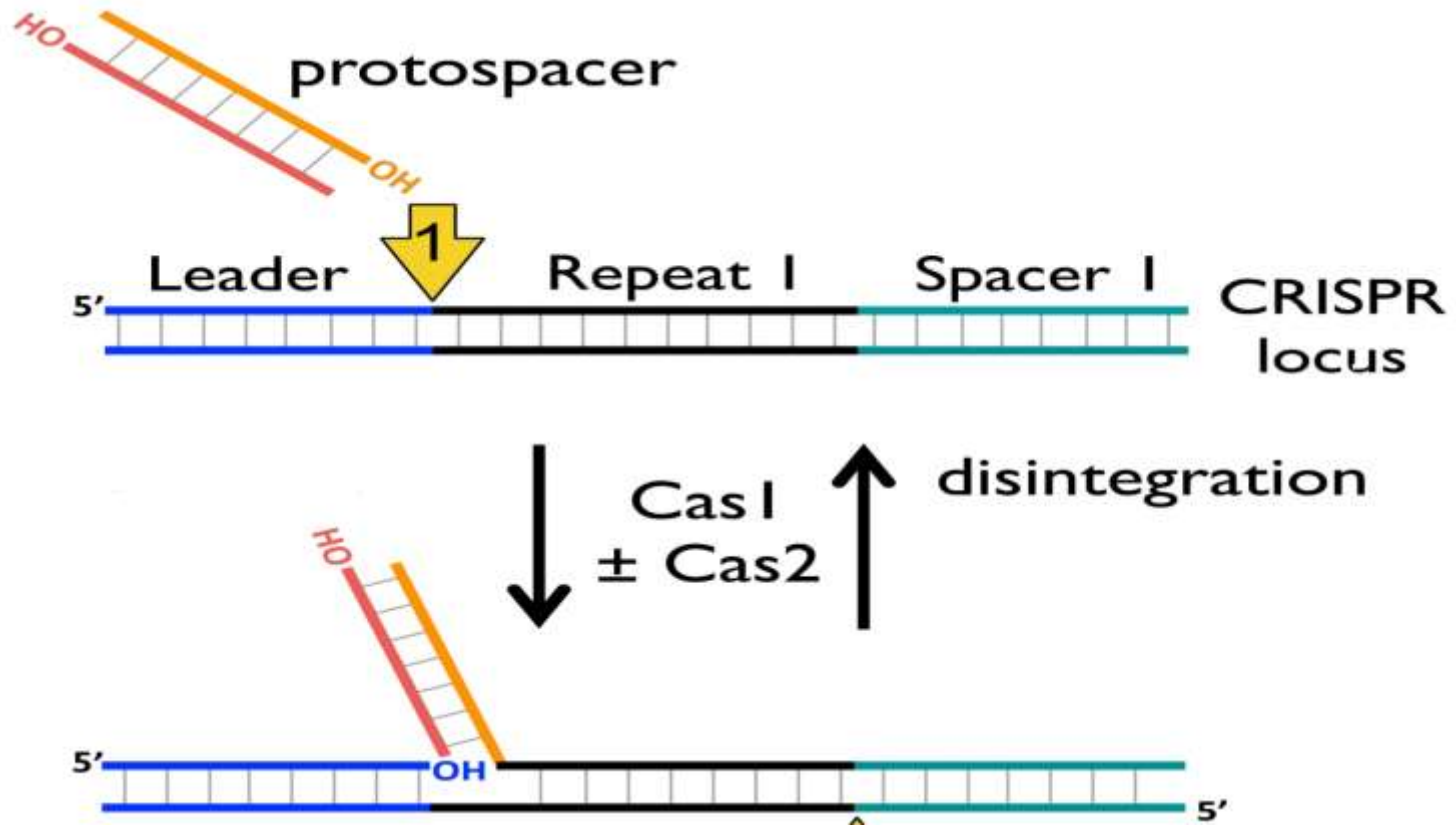
CRISPRs

- The sequence in the virus that will become a new spacer is called the proto-spacer and is found close to a PAM (proto-spacer adjacent motif) sequence.

CRISPRs

- When a new spacer is added to a CRISPR array, it is incorporated at the proximal end, near the leader sequence.
- Some of the *cas* genes encode proteins required for this acquisition process.

CRISPRs



CRISPRs

- Thus, the antiviral defense mechanism is not impaired by their absence, but the cell cannot acquire resistance to new viruses.
- The products of the *cas1*, *cas2*, and *cas4* genes fall into this category.

CRISPRs

- *Cas1* is a putative integrase, whereas *Cas2* is a ribonuclease.
- In contrast, of other *Cas* proteins, *Cas6* is involved in expression and processing of the CRISPR cluster and *Cas3* in the interference of viral infection.

END

Regulatory RNAs in Eukaryotes

- Eukaryotic regulatory RNAs come in many flavors characterized by their size (“long” or “short”), their origin, and the mechanisms by which they are generated and regulate gene expression.

Regulatory RNAs in Eukaryotes

- It is now believed that between 30% and 70% of genes in higher eukaryotes are regulated to some extent by RNAs, with roles ranging from development to cellular homeostasis and protection of the cell from viruses and transposons.

Regulatory RNAs in Eukaryotes

- Furthermore, one form of regulation viz., RNA interference (RNAi) has been adapted for use as a powerful experimental tool to manipulate gene expression in many organisms.

Regulatory RNAs in Eukaryotes

- Several types of very short RNAs repress or silence the expression of genes with homology to those short RNAs.

Regulatory RNAs in Eukaryotes

- Depending on the origin and context, these RNAs act by inhibiting translation of the mRNA, destruction of the mRNA, or even by transcriptional silencing of the promoter that directs expression of that mRNA.

Regulatory RNAs in Eukaryotes

- These short RNAs are often generated by special enzymes from longer double-stranded RNAs (dsRNAs) of various origins.

Regulatory RNAs in Eukaryotes

- The small RNAs have different names depending on their origin.
- Those made artificially or produced in vivo from dsRNA precursors are typically called **small interfering RNAs (SiRNAs)**.

Regulatory RNAs in Eukaryotes

- Another group of regulatory RNAs is the **microRNAs (miRNAs)**.
- These miRNAs are derived from precursor RNAs that are encoded by genes expressed in cells where those miRNAs have specific regulatory functions.

Regulatory RNAs in Eukaryotes

- A third class of short regulatory RNAs is the piwi-interaction RNAs (piRNAs), which are expressed predominantly in the germline and have features distinct from miRNAs.

Regulatory RNAs in Eukaryotes

- Both siRNAs and miRNAs are generated from longer RNA molecules by the enzyme **Dicer**, an RNase III–like enzyme that recognizes and digests longer dsRNAs or the stem-loop structures formed by miRNA precursors.

Regulatory RNAs in Eukaryotes

- The siRNA and miRNA products are typically 21–23 nucleotides long.
- The piRNAs (which are 24–34 nucleotides long) are derived in a manner that does not involve a dsRNA precursor.

Regulatory RNAs in Eukaryotes

- Instead, the piRNAs are generated by processing long single-stranded transcripts covering so-called piRNA clusters found in the genome.
- This processing does not require Dicer.

Regulatory RNAs in Eukaryotes

- These small RNAs inhibit expression of homologous target genes in three ways:-
- a) they trigger destruction of the mRNA encoded by the target gene,

Regulatory RNAs in Eukaryotes

- b) they inhibit translation of the mRNA, or
- c) they induce chromatin modifications within the target gene and thereby silence its transcription.

Regulatory RNAs in Eukaryotes

- Remarkably, whichever route is used in any given case, much of the same machinery is required.

Regulatory RNAs in Eukaryotes

- This machinery includes a complex called the **RNA-induced silencing complex (RISC)**.
- A RISC contains, in addition to the small RNA, various proteins including a member of the **Argonaut** family.

Regulatory RNAs in Eukaryotes

- The small RNA must be denatured to give a **guide RNA** - the strand that gives the RISC specificity and a **passenger RNA**, which usually gets discarded.

Regulatory RNAs in Eukaryotes

- The resulting complex, the mature RISC, is then directed to target RNAs containing sequences complementary to the guide RNA.
- These target RNAs are degraded or their translation is inhibited.

Regulatory RNAs in Eukaryotes

- Typically, the choice depends in part on how closely the guide RNA matches the target mRNA:-
- if the sequences are highly complementary (as is usually the case with siRNAs), the target is degraded;

Regulatory RNAs in Eukaryotes

- if the match is not as good (i.e., if there are several base-pairing mismatches, as is often the case with miRNAs), the response is more often inhibition of translation.

Regulatory RNAs in Eukaryotes

- In those cases in which the target RNA is degraded, Argonaute is the catalytic subunit that performs the initial mRNA cleavage; for this reason, Argonaute is often called **Slicer** and mRNA cleavage is called **slicing**.

Regulatory RNAs in Eukaryotes

- A RISC can also be directed into the nucleus, where it recruits other proteins that modify the chromatin around the promoter of the gene complementary to the guide RNA. This modification leads to silencing of transcription.

Regulatory RNAs in Eukaryotes

- A distinction worth making between miRNAs and siRNAs is that the former act like traditional trans-acting regulators: they are encoded by a gene but act on other genes.

Regulatory RNAs in Eukaryotes

- In contrast, siRNAs are typically generated by transcripts of the regions on which they act and are thus described as working in *cis*.

END

Promoters

- A gene consists of a transcriptional region and a regulatory region.
- The regulatory region can be divided into *cis*-regulatory (or *cis*-acting) elements and *trans*-regulatory (or *trans*-acting) elements.

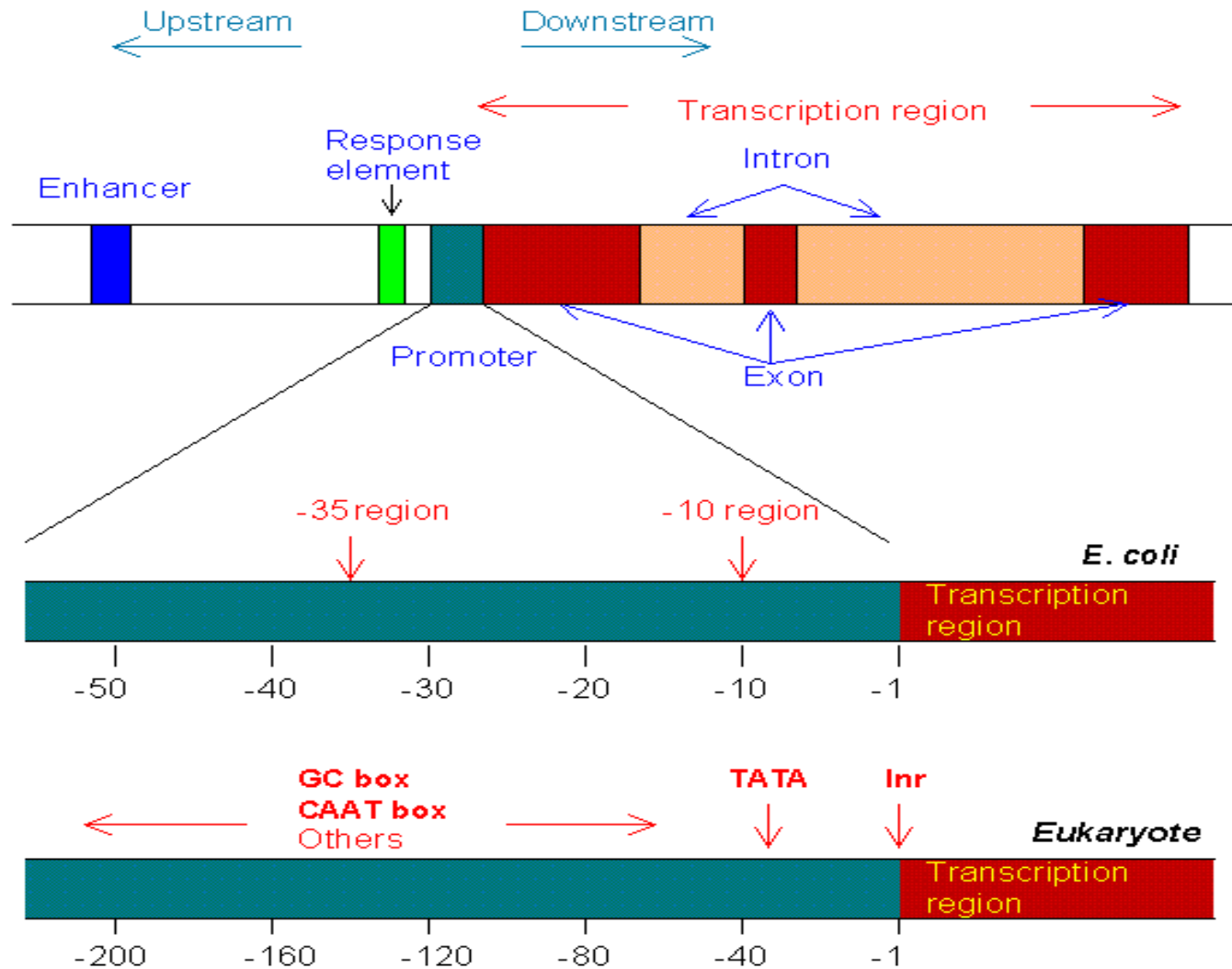
Promoters

- The *cis*-regulatory elements are the binding sites of transcription factors which are the proteins that, upon binding with *cis*-regulatory elements, can affect (either enhance or repress) transcription.

Promoters

- The cis-acting elements may be divided into the following four types:
- Promoter
- Enhancer
- Silencer
- Response element

Promoters



Promoters

- Three classes of promoters can be distinguished on the basis of their relative locations.
- Class II Promoters
- Class I Promoters
- Class III Promoters

Promoters

- **Class II Promoter**
- The promoters recognized by RNA polymerase II are called class II promoters because these are the most complex and best studied.

Promoters

- Class II promoters can be considered as having two parts: the **core promoter** and the **proximal promoter**.

Promoters

- The core promoter attracts general transcription factors and RNA polymerase II at a basal level and sets the transcription start site and direction of transcription.

Promoters

- The proximal promoter helps attract general transcription factors and RNA polymerase and includes promoter elements that can extend from about 37 bp up to 250 bp upstream of the transcription start site.

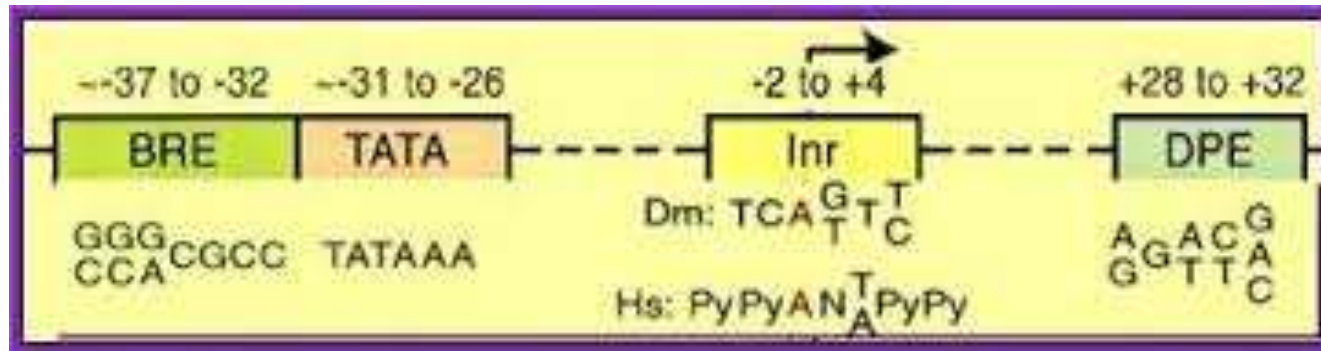
Promoters

- The core promoter is modular and can contain almost any combination of the following elements.
- The TATA box is centered at approximately position 228 and has the consensus sequence TATA(A/T)AA(G/A);

Promoters

- the **TFIIB recognition element** (BRE) lies just upstream of the TATA box and has the consensus sequence (G/C)(G/C)(G/A)CGCC; the **initiator (Inr)** is centered on the transcription start site, and the **downstream core element** (DCE) having three parts.

Promoters



Promoters

- **Class I Promoter**
- We can refer to this promoter in the singular because almost all species have only one kind of gene recognized by polymerase I: the rRNA precursor gene.

Promoters

- The promoter has two critical regions, one of which is the **core element**, also known as the **initiator (rINR)**, is located at the start of transcription, between positions - 45 and +20.
- The other is the **upstream promoter element (UPE)**, located between positions -156 and - 107.

Promoters

- **Class III Promoters**
- RNA polymerase III transcribes a variety of genes that encode small RNAs. These include:-
 - 1) the “classical” class III genes, including the 5S rRNA and tRNA genes; and

Promoters

- 2) some relatively recently discovered class III genes, including the U6 snRNA gene, the 7SL RNA gene, the 7SK RNA gene, and the Epstein – Barr virus EBER2 gene.

Promoters

- By contrast, the “classical” class III genes have promoters located entirely within the genes themselves.

END

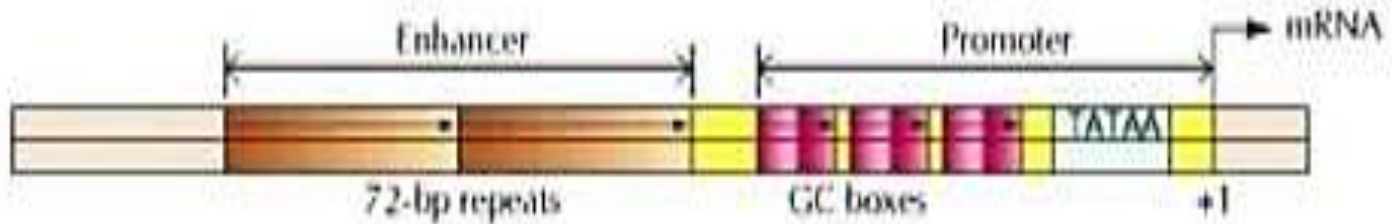
Enhancers & Silencers

- Enhancers are elements that stimulate transcription while Silencers, by contrast, depress transcription.
- Chambon and colleagues discovered the first enhancer in the 5'-flanking region of the SV40 early gene.

Enhancers & Silencers

- This DNA region contains a conspicuous duplication of a 72-bp sequence, called the *72-bp repeat*.
- When Benoist and Chambon made deletion mutations in this region, they observed profoundly depressed transcription in vivo.

Enhancers & Silencers



The SV40 enhancer

Enhancers & Silencers

- This behaviour suggested that the 72-bp repeats constituted another upstream promoter element.

Enhancers & Silencers

- However, Paul Berg and his colleagues discovered that the 72-bp repeats still stimulated transcription even if they were inverted or moved all the way around to the opposite side of the circular SV40 genome, over 2 kb away from the promoter.

Enhancers & Silencers

- The latter behaviour, at least, is very un-promoter-like.
- Thus, such orientation- and position independent DNA elements are called enhancers to distinguish them from promoter elements.

Enhancers & Silencers

- Enhancers act through proteins that bind to them.
- These proteins have several names: transcription factors, enhancer-binding proteins, or activators.

Enhancers & Silencers

- These proteins appear to stimulate transcription by interacting with general transcription factors at the promoter.
- This interaction promotes formation of a pre-initiation complex, which is necessary for transcription.

Enhancers & Silencers

- Thus, enhancers usually allow a gene to be induced (or sometimes repressed) by activators that frequently require help from other molecules (e.g., hormones and co-activator proteins) to exert their effects.

Enhancers & Silencers

- Enhancers are not the only DNA elements that can act at a distance to modulate transcription. Silencers also do this, but - as their name implies - they inhibit rather than stimulate transcription.

Enhancers & Silencers

- Yeast chromosome III contains three loci of very similar sequence: MAT, HML, and HMR.
- Though MAT is expressed, the other two loci are not, and silencers located at least 1 kb away seem to be responsible for this genetic inactivity.

Enhancers & Silencers

- They seem to be responding to an external negative influence: a silencer.
- How do silencers work?

Enhancers & Silencers

- The available data indicate that they cause the chromatin to coil up into a condensed, inaccessible, and therefore inactive form, thereby preventing transcription of neighbouring genes.

Enhancers & Silencers

- Sometimes the same DNA element can have both enhancer and silencer activity, depending on the protein bound to it.

Enhancers & Silencers

- For example, the thyroid hormone response element acts as a silencer when the thyroid hormone receptor binds to it without its ligand, thyroid hormone.

Enhancers & Silencers

- But it acts as an enhancer when the thyroid hormone receptor binds along with thyroid hormone.

END

Monocistronic or Polycistronic mRNA

- In a bacterial cell, transcription and translation can proceed simultaneously because both these processes take place in the same cellular compartment so the protein synthetic machinery can start reading 5' end of mRNA before the 3' end is formed.

Monocistronic or Polycistronic mRNA

- While in eukaryotic cells, both these processes occur in different compartments.
- So primary transcript is converted to mature mRNA before the protein synthesis starts in the cytoplasm.

Monocistronic or Polycistronic mRNA

- The segment of mRNA that codes for a polypeptide chain is called an **open reading frame (ORF)** because the protein synthesis starts at a specific start codon and stops at a specific termination/stop codon.

Monocistronic or Polycistronic mRNA

- A DNA segment corresponding to an ORF plus the translational start and stop signals for protein synthesis is called a cistron and the mRNA encoding a single polypeptide is called *monocistronic mRNA*.

Monocistronic or Polycistronic mRNA

- Although the terms *cistron* and *gene* are sometimes used interchangeably to describe bacterial DNA segments that specify polypeptides,

Monocistronic or Polycistronic mRNA

- the term gene has a broader meaning because it also includes the promoter region and applies to DNA segment that code for RNA molecules such as tRNA and rRNA that are not translated.

Monocistronic or Polycistronic mRNA

- Bacterial mRNA molecules often contain two or more cistrons, thus are *polycistronic* and are actually more common than bacterial monocistronic mRNA molecules.

Monocistronic or Polycistronic mRNA

- Each cistron within a polycistronic mRNA specifies a specific polypeptide chain.
- Furthermore, cistrons contained in polycistronic mRNA often specify proteins for a single metabolic pathway.

Monocistronic or Polycistronic mRNA

- For example, one *Escherichia coli* mRNA has eight cistrons, each coding for a different enzyme required for histidine synthesis.
- Using polycistronic mRNA is a way for the cell to regulate synthesis of related protein coordinately.

Monocistronic or Polycistronic mRNA

- With a polycistronic mRNA molecule, the synthesis of several related proteins – in similar quantities and at same time – can be regulated by a single signal.

Monocistronic or Polycistronic mRNA

- In addition to reading frames and start and stop codons, other regions in mRNA are also significant.
- For example mRNA translation rarely starts exactly at the 5' end and stops exactly at 3' end .

Monocistronic or Polycistronic mRNA

- Instead, initiation of synthesis of the first polypeptide chain of a polycistronic mRNA may begin hundreds of nucleotides from 5'-RNA terminus.

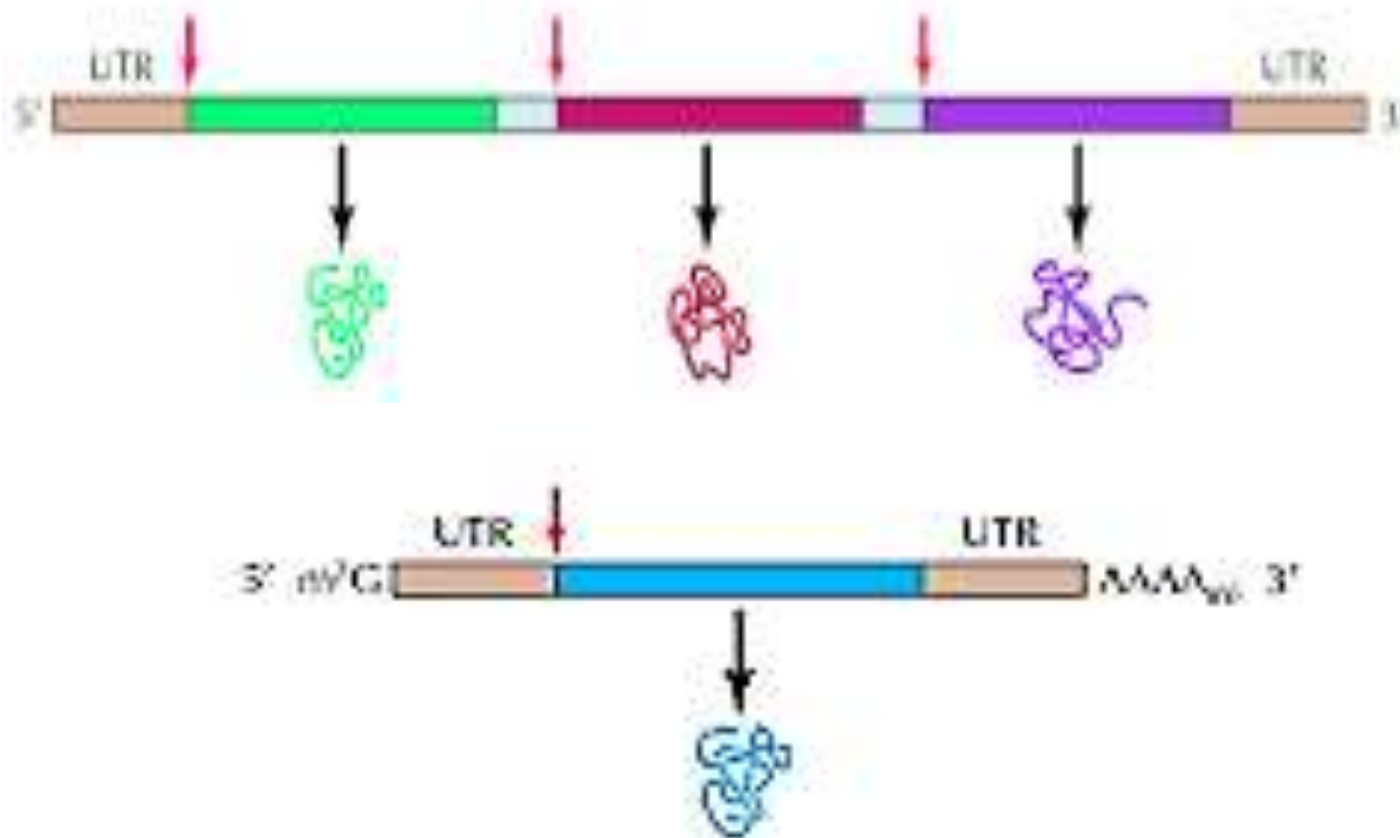
Monocistronic or Polycistronic mRNA

- The untranslated RNA sequence before the coding region is called the 5' – untranslated region (5'- UTR) or 5' – leader while the untranslated sequence after the coding region is called 3'– untranslated region (3'- UTR).

Monocistronic or Polycistronic mRNA

- Polycistronic mRNA molecules also usually contain intercistronic sequences (**spacers**) that are tens of bases long.

Monocistronic or Polycistronic mRNA



mRNA has a short life span

- An important characteristic of bacterial mRNA is that its lifetime is short compared to other types of bacterial RNA molecules.
- The half life of a typical bacterial mRNA is a few minutes.

mRNA has a short life span

- Although mRNA's short lifetime may seem wasteful, it has an important regulatory function.
- A cell can turn off the synthesis of a protein that is no longer needed by turning off synthesis of mRNA that encodes the protein.

mRNA has a short life span

- Soon after, none of that particular mRNA will remain and protein synthesis will cease.
- Of course, bacterial cells save energy by not being forced to synthesize proteins that they no longer need.

mRNA has a short life span

- The particular mRNA and their life span can be analyzed by the pulse-chase experiment.
- Bacteria are briefly cultured in a medium that contains a radioactive precursor for RNA such as [^3H] uridine.

mRNA has a short life span

- Then the bacteria are switched to a medium that contains a high concentration of non radioactive uridine while no [^3H] uridine and samples are removed at specific times for analyses.

mRNA has a short life span

- The RNA are isolated and different types are separated by gel electrophoresis and detected by their radioactivity.

mRNA has a short life span

- A typical radioactive mRNA molecule will decrease with a half life of a few minutes, whereas radioactive rRNA and tRNA will remain through many generations.

mRNA has a short life span

- However, bacteria also contain some longlived mRNA molecules which can not be analyzed by this technique.

END

Controlling the rate of mRNA synthesis

- Bacterial cells can control gene expression by regulating the rate at which specific genes are transcribed.
- They vary the promoter sequence so that RNA polymerase initiates transcription at some genes more efficiently than at others.

Controlling the rate of mRNA synthesis

- However, this mechanism is not ideally suited for all situations.
- For example, the products of many genes are needed under some physiological conditions but not others, so the products of such genes must be synthesized only when circumstances demand it.

Controlling the rate of mRNA synthesis

- Bacterial cells increase their efficiency by:-
- 1) Selecting the catabolic pathway for energy production that yields the greatest amount of energy per unit time, and
- 2) synthesizing molecules only as the need arises.

Controlling the rate of mRNA synthesis

- Bacteria accomplish both of these objectives by turning on the transcription of specific genes when their products are needed and turning off their transcription when their products are not needed.

Controlling the rate of mRNA synthesis

- Actually, there are no known examples of switching a system completely off.
- When transcription is in the “off” state, there always remains a basal level of gene expression.

Controlling the rate of mRNA synthesis

- So the term “*off*” means *very slow*.
- There are also some examples in which activity is switched from fully on to partly on rather than off.

Controlling the rate of mRNA synthesis

- In bacterial systems in which several enzymes act in sequence in a single metabolic pathway, it is often the case that either all of these enzymes are present or all are absent.

Controlling the rate of mRNA synthesis

- This phenomenon, which is called **coordinate regulation**, results from control of the synthesis of a single polycistronic mRNA that encodes all of the gene products.
- There are several mechanisms for such regulation.

Controlling the rate of mRNA synthesis

- The molecular mechanism of mRNA regulation can be divided into two major categories:-
 - a) negative regulation
 - b) positive regulation

Controlling the rate of mRNA synthesis

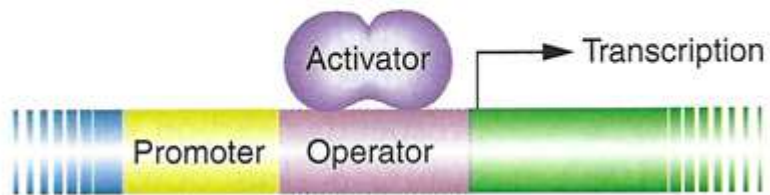
- In negative regulation, a **repressor** turns off the transcription of one or more gene.
- In positive regulation, an **activator** turns on the transcription of one or more gene.

Controlling the rate of mRNA synthesis

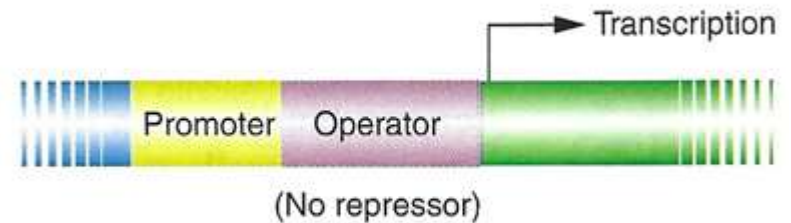
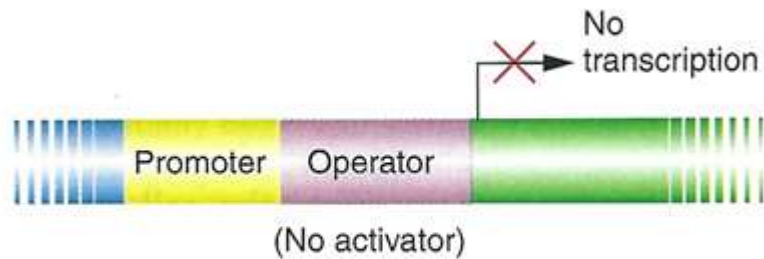
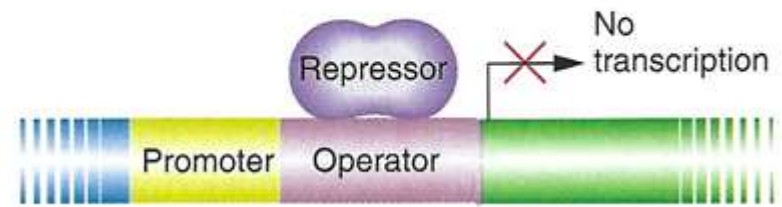
- Negative regulation and positive regulation are not mutually exclusive.
- Many genes respond to both types of regulation.

Controlling the rate of mRNA synthesis

Positive regulation



Negative regulation



Lactose Operon

- In *E. coli*, two proteins are necessary for lactose metabolism.
- These include the enzyme β -galactosidase and a carrier molecule, lactose permease, which transports lactose (and other galactosides) into the cell.

Lactose Operon

- The existence of the two proteins was first shown by a combination of genetic experiments and biochemical analysis.
- First, hundreds of Lac⁻ mutants (unable to use lactose as a carbon source) were isolated.

Lactose Operon

- By genetic manipulation, some of these mutations were moved from the *E. coli* chromosome to an *F'**lac* plasmid (a plasmid carrying the genes for lactose utilization).

Lactose Operon

- Then partial diploids having the genotypes $F' lac^- / lac^+$ or $F' lac^+ / lac^-$ were constructed.
- It was observed that these diploids always have a Lac^+ phenotype (can make β -galactosidase).

Lactose Operon

- This shows that none of the *lac*⁻ mutants make an inhibitor that blocks *lac* gene function.
- Partial diploids were also constructed in which both the chromosome and the F'*lac* plasmid were *lac*⁻.

Lactose Operon

- Using different pairs of lac^- mutants, some pairs were observed to have a Lac^+ phenotype while others were observed to have a Lac^- phenotype.

Lactose Operon

- This complementation test showed that all of the mutants initially isolated fell into two groups which were called *lacZ* and *lacY*.

Lactose Operon

- The existence of two complementation groups was good evidence that there are at least two genes in the *lac* system.
- The *lacZ* gene is the structural gene for β -galactosidase.

Lactose Operon

- This enzyme catalyzes the hydrolysis of o-nitrophenyl- β -galactoside (a lactose analog) and results in the production of o-nitrophenoxide which is a yellow product.

Lactose Operon

- The function of the *lacY* gene product as lactose permease was strongly suggested by experiments that showed that *lacZ*^{+ *lacY*⁻ cells can not transport lactose into the cell while *lacY*⁻ *lacZ*⁺ cells can do so.}

Lactose Operon

- Investigators discovered a third gene, lacA, the product of which is a β -galactoside transacetylase, which transfers an acetyl group from acetyl-CoA to lactose analog.

Lactose Operon

- The reason that *lacA* was not detected at the same time as *lacZ* and *lacY* is that its gene product, transacetylase, is not required for lactose catabolism.

Lactose Operon

- The precise role of *lacA* is still a matter of supposition.
- One hypothesis is that transacetylase detoxifies lactose analogs that would harm cells.

END

lac structural genes are regulated

- When *E. coli* with a *lac*⁺ genotype is cultured in a lactose-free medium, the intracellular concentration of β -galactosidase, permease and tranacetylase are exceedingly low.

***lac* structural genes are regulated**

- When lactose is added to the growth medium, however, the concentration of these protein increases simultaneously to about 10^5 molecules per cell (about 1% of total cellular protein).

***lac* structural genes are regulated**

- Furthermore, lactose addition triggers the synthesis of *lac* mRNA as evidenced in studies in which mRNA, labeled with [³²P] phosphate at various times after lactose addition, is hybridized to DNA that carries *lac* genes.

lac structural genes are regulated

- Enzymes such as β -galactosidase, permease and tranacetylase are said to be **inducible enzymes** because their rate of synthesis increases in response to the addition of a small molecule (lactose) to the medium.

lac structural genes are regulated

- Other enzymes, called **repressible enzymes**, exhibit a decreased rate of synthesis in response to the addition of a small molecule in the medium.

***lac* structural genes are regulated**

- For instance, the addition of tryptophan to the growth medium causes *E. coli* to greatly decrease the rate at which it produces enzymes needed for tryptophan synthesis.

lac structural genes are regulated

- Still other enzymes, called **constitutive enzyme**, are synthesized at fixed rates under all growth conditions.

lac structural genes are regulated

- **Constitutive enzyme** usually perform basic cellular “housekeeping” functions needed for normal cell maintenance.

***lac* structural genes are regulated**

- Lactose is rarely used in experiments to study induction of the lactose enzymes because the β -galactosidase that is synthesized catalyzes lactose cleavage.

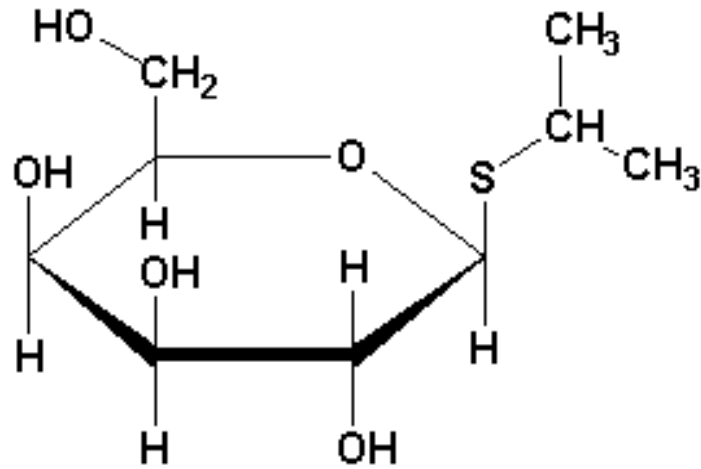
***lac* structural genes are regulated**

- As a result, the lactose concentration continually decreases, which complicates the analysis of kinetic experiments.

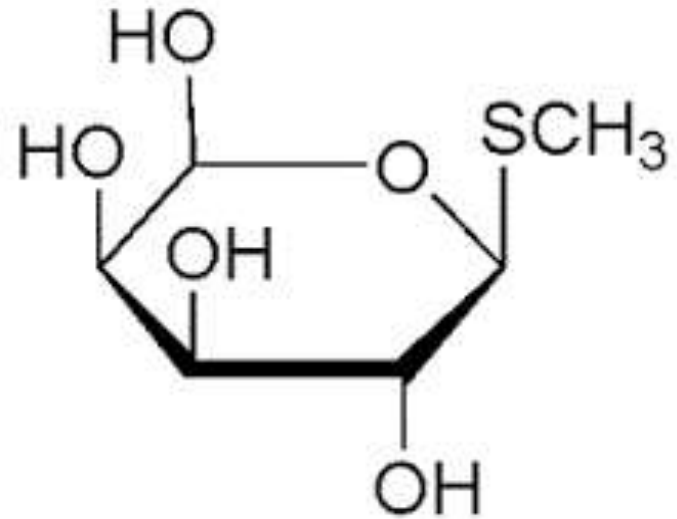
***lac* structural genes are regulated**

- Instead, two sulfur-containing lactose analogs, isopropylthiogalactoside (IPTG) and thiomethylgalactoside (TMG) are used.

lac structural genes are regulated



Isopropyl Thiogalactoside (IPTG)



lac structural genes are regulated

- These are effective inducers without being substrates of β -galactosidase.
- Inducers having this property are called gratuitous inducers.

END

Regulation of *lac* mRNA

- Two French investigators Jacques Monod and Francois Jacob performed a series of genetic and biochemical experiments in the late 1950's that elucidated the mechanism of regulation of *lac* system.

Regulation of *lac* mRNA

- The began by isolating constitutive *E. coli* mutants that make *lac* mRNA (make all the three enzymes) in the presence as well as absence of the inducer.

Regulation of *lac* mRNA

- Then they constructed a variety of partial diploid cells containing constitutive mutants and observed the cell's ability to synthesize β -galactosidase.

Regulation of *lac* mRNA

- These mutations appeared to be of two types, termed *lacI* and *lacO^c*.
- The *lacI*⁻ mutations behaved like typical minus mutations in most genes and are recessive.

Regulation of *lac* mRNA

- Because *lac* mRNA synthesis is off in a *lacI*⁺ cell and is on in a *lacI*⁻ mutant, the *lacI* gene is apparently a regulatory gene.

Regulation of *lac* mRNA

- The *lacI* gene codes for a product that acts as an inhibitor to keep the *lac* structural genes turned off.
- The *lacI*⁻ mutant lacks the inhibitor and thus is constitutive.

Regulation of *lac* mRNA

- A $lacI^+ / lacI^-$ partial diploid has one good copy of the *lacI* gene product, so the system is inhibited.
- Monod & Jacob called the *lacI*-gene product, the **Lac repressor**.

Regulation of *lac* mRNA

- The most likely explanation for the constitutive lactose system is that the mutant strain synthesizes a truncated repressor protein that can't block transcription of the *lac* genes.

Regulation of *lac* mRNA

- This conclusion was confirmed when the Lac repressor was purified and characterized.
- Genetic mapping experiments placed the *lacI* gene adjacent to the *lacZ* gene and established the order of the genes as *lacI lacZ lacY lacA*.

Regulation of *lac* mRNA

- A striking property of the *lacO^c* mutations is that in certain cases, they are dominant.
- The significance of the dominance of the *lacO^c* mutations becomes clear from the properties of the partial diploids.

Regulation of *lac* mRNA

- Both the diploids $lacO^c lacZ^- / lacO^+ lacZ^+$ and $lacO^c lacZ^+ / lacO^+ lacZ^-$ are Lac^+ , because there is a functional *lacZ* gene.
- In case of former, β -galactosidase synthesis is inducible even though a $lacO^c$ mutation is present.

Regulation of *lac* mRNA

- In case of former, β -galactosidase synthesis is inducible even though a lacO^c mutation is present.

Regulation of *lac* mRNA

- An immunological test capable of detecting a mutant β -galactosidase showed that the mutant enzyme is synthesized constitutively in a $lacO^c lacZ^- / lacO^+ lacZ^+$ partial diploid,

Regulation of *lac* mRNA

- whereas the wild type enzyme is synthesized only if an inducer is added.

Regulation of *lac* mRNA

- This experiment takes advantage of the fact that a purified antibody to β -galactosidase will also react with mutant protein as long as the structural differences between wild and mutant proteins are not too great.

Regulation of *lac* mRNA

- A reaction of this type in which an antibody that is raised in response to one protein is used to detect a closely related protein is called a **cross-reaction** and closely related protein is called **cross-reacting material (CRM)**.

Regulation of *lac* mRNA

- Thus the presence of CRM, which can be detected by a variety of standard immunological procedures, is indicative of the presence of mutant protein.

Regulation of *lac* mRNA

- Genetic mapping experiments showed that all the *lacO^c* mutations are located between genes *lacI* and *lacZ*, so the gene order of the five elements of the *lac* system is *lacI lacO lacZ lacY lacA*.

Regulation of *lac* mRNA

- Together these experiments lead to the conclusion that *lacO^c* mutations define a site or a noncoding region of the DNA rather than a gene and that the *lacO* region determines whether synthesis of the

Regulation of *lac* mRNA

- product of the adjacent *lacZ* gene is inducible or constitutive.
- The *lacO* site is called the **operator**.

END

The operon model

- Monod and Jacob proposed the **operon model** in 1961 to explain how the lac system is regulated.

The operon model

- The term operon refers to two or more contiguous genes and the genetic elements that regulate their transcription in a coordinate fashion.

The operon model

- Promoters had not yet been discovered when Monod & Jacob proposed the operon model but were readily incorporated into the operon model after their discovery.
- The five major features of the model are:-

The operon model

- 1. The products of the *lacZ*, *lacY* and *lacA* genes are encoded in a single polycistronic *lac* mRNA molecule.
- 2. The promoter for this mRNA molecule is immediately adjacent to the *lac* region.

The operon model

- Promoter mutations (p-) that are completely incapable of making β -galactosidase, permease, and transacetylase have been isolated. The promoter is located between *lacI* and *lacO*.

The operon model

- 3. The operator is a sequence of bases (in the DNA) to which the repressor protein binds.
- 4. When the repressor protein is bound to the operator, *lac* mRNA transcription can't take place.

The operon model

- 5. Inducers stimulate lac mRNA synthesis by binding to the repressor. This binding alters the repressor's conformation so it can't bind to the operator. In the presence of an inducer, therefore,

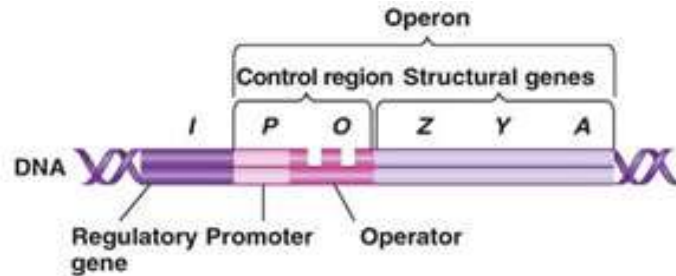
The operon model

- The operator is unoccupied and the promoter is available for initiation of mRNA synthesis. This state is called **derepression**.

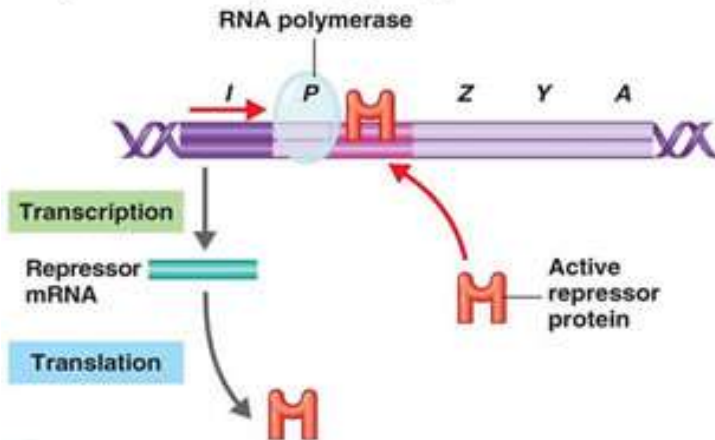
The operon model

- This simple model explains many of the features of the *lac* system and of other negatively regulated genetic systems.
- However, *lac* operon is also subjected to positive regulation.

The operon model

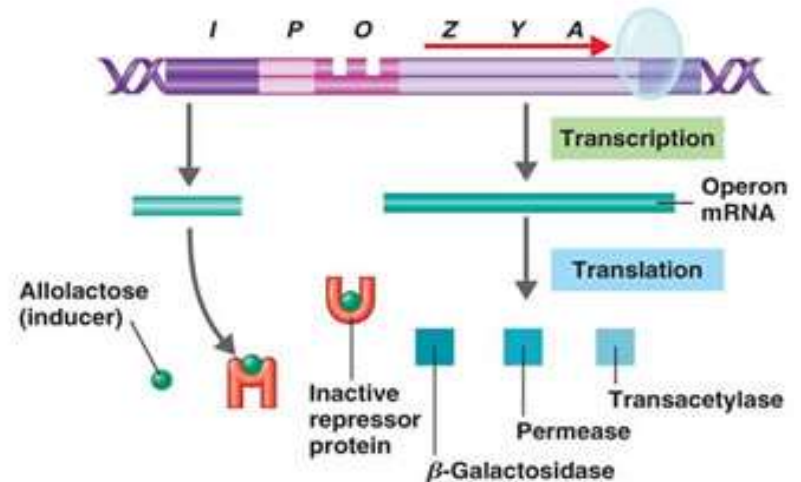


- 1 Structure of the operon.** The operon consists of the promoter (*P*) and operator (*O*) sites and structural genes that code for the protein. The operon is regulated by the product of the regulatory gene (*I*).



- 2 Repressor active, operon off.** The repressor protein binds with the operator, preventing transcription from the operon.

Copyright © 2010 Pearson Education, Inc.



- 3 Repressor inactive, operon on.** When the inducer allolactose binds to the repressor protein, the inactivated repressor can no longer block transcription. The structural genes are transcribed, ultimately resulting in the production of the enzymes needed for lactose catabolism.

Allolactose; the inducer of lactose operon

- Two related problems became evident as the operon model was tested.

Allolactose; the inducer of lactose operon

- First, inducers must enter a cell if they are to bind to repressor molecules, yet lactose transport requires permease, and permease synthesis requires induction.

Allolactose; the inducer of lactose operon

- Thus, it must be explained how the inducer gets into the cell in the first place.

Allolactose; the inducer of lactose operon

- Second, the isolated Lac repressor does not bind lactose but does bind a lactose isomer called **allolactose**.
- Remarkably, β -galactosidase also converts a small proportion of lactose to allolactose.

Allolactose; the inducer of lactose operon

- Therefore, induction of the synthesis of β -galactosidase by lactose requires that β -galactosidase be present.

Allolactose; the inducer of lactose operon

- Both problems are solved in the same way; in the uninduced state, a small amount of *Lac* mRNA is synthesized (roughly one mRNA molecule per cell per generation).

Allolactose; the inducer of lactose operon

- This synthesis, called the **basal synthesis**, occurs because the binding of the repressor to the operator is never infinitely strong.

Allolactose; the inducer of lactose operon

- Thus, even though the repressor binds tightly to the operator, it occasionally comes off and an RNA polymerase molecule can initiate transcription during the instant that the operator is free.

Allolactose; the inducer of lactose operon

- When lactose is added to a *Lac*⁺ culture growing in lactose free medium, the few permease molecules transport a few lactose molecules into the cell and few β -galactosidase molecules convert some of lactoses to allolactose.

Allolactose; the inducer of lactose operon

- An allolactose molecule then binds to a repressor molecule that is sitting on the operator, and the repressor is inactivated and falls off the operator.

Allolactose; the inducer of lactose operon

- Synthesis of *Lac* mRNA begins and these mRNA molecules are translated to produce hundreds of β -galactosidase and permease molecules which allow lactose molecules to enter into the cells.

Allolactose; the inducer of lactose operon

- Most of the lactose molecules are cleaved to yield glucose and galactose, but some are converted to allolactose molecules, which binds to and inactivate all of the intracellular repressor molecules.

Allolactose; the inducer of lactose operon

- Thus lac mRNA is synthesized at high rate and the permease and β -galactosidase concentrations become quite high.
- The glucose produced is used as a carbon source for energy.

Allolactose; the inducer of lactose operon

- The glucose produced by cleavage reaction is used as a carbon source for energy while the galactose produced is converted to glucose-1-phosphate by a set of enzymes which are also inducible.

Allolactose; the inducer of lactose operon

- Ultimately, all of the lactose in the growth medium and within the cells is consumed.
- Then the allolactose concentration drops so that there is not sufficient allolactose to bind to a repressor.

Allolactose; the inducer of lactose operon

- The repressor binds to the operator, re-establishing repression and thereby blocking further synthesis of *lac* mRNA.
- In bacteria, most mRNA molecules have half life of only a few minutes.

Allolactose; the inducer of lactose operon

- Hence, in less than one generation, there is little remaining *lac* mRNA and synthesis of β -galactosidase and permease ceases.
- These proteins are quite stable but are gradually diluted as the cell divide.

Allolactose; the inducer of lactose operon

- Note that if lactose were added again one generation after the original lactose had been depleted, cleavage of lactose would begin immediately as the cells would already have adequate β -galactosidase and permease.

END

The *Lac* repressor

- Walter Gilbert and Benno Muller-Hill successfully isolated the *Lac* Repressor from *E. coli* extracts in 1966.
- The *Lac* repressor is a homotetramer with a molecular mass of 154 kDa.

The *Lac* repressor

- Each subunit of the *Lac* repressor is made up of 360 amino acids and can bind one molecule of IPTG.
- Crude cell extracts bind about 20-40 molecules of IPTG per cell, so there are roughly 5-10 repressor molecules per cell.

The *Lac* repressor

- The specific binding of repressor to the operator sequence and the inhibition of this binding by an inducer have been demonstrated by using purified repressor.

The *Lac* repressor

- An important procedure for studying repressor-operator binding is the nitrocellulose filter assay.
- Protein stick to these filters but DNA does not.

The *Lac* repressor

- If a mixture of repressor and radioactive *lac* DNA is passed through such a filter, radioactivity will be retained on the filter if the protein and the *lac* DNA form a complex.

The *Lac* repressor

- The results indicate that the *Lac* repressor binds to DNA with a normal *lac* operator but fails to bind to DNA with a *lacO^c* mutant operator.

END

The *Lac* repressor

- Furthermore, IPTG prevents the *Lac* repressor from binding to DNA with a normal *lac* operator.
- These studies confirm the major predictions of the operon model.

The *Lac* repressor

- After the *lac* operon was sequenced, it eventually became clear that it has two additional operators.
- The original operator with its centre of symmetry at position +11 is now designated as *lacO*₁.

The *Lac* repressor

- Auxiliary operators *lacO*₂ and *lacO*₃ have their centers of symmetry at positions +412 and -82, respectively.
- Thus, *lacO*₃ is upstream of the *lac* promoter while *lacO*₂ is located in *lacZ*.

The *Lac* repressor

- The discoveries of *lacO*₂ and *lacO*₃ led investigators to ask whether these auxiliary operators participate in the *lac* operon regulation.

The *Lac* repressor

- Muller-Hill and coworkers performed series of genetic experiments in which they altered one or more *lac* operators and then determined the alteration's effect on repression.

The *Lac* repressor

- Point mutations in *lacO*₁ caused a 5 – 50 folds decrease in repression, but some repression was still observed.

The *Lac* repressor

- Destruction of either *lacO*₂ or *lacO*₃ caused a twofold decrease in repression and destruction of both auxiliary operators caused a 70-fold decrease in repression.

The *Lac* repressor

- The ability to repress the *lac* system was completely lost in cells in which *lacO*₁ and either of the auxiliary operators were destroyed.

The *Lac* repressor

- These results indicate that *lacO*₁ play a major role in repression but the two auxiliary operators also make important contributions.

END

Catabolite Repression

- The function of the β -galactosidase in lactose metabolism is to hydrolyze lactose to form glucose and galactose.

Catabolite Repression

- If the growth medium contains both glucose and lactose, then in the interest of efficiency, there is no need for a cell to turn on the *lac* operon.
- Monod in 1940s showed that cells behave according to this logic.

Catabolite Repression

- *E. coli* cells incubated in the presence of glucose and lactose do not start to make β -galactosidase until all the exogenous glucose is consumed.

Catabolite Repression

- These findings were later extended to lactose permease and transacetylase, the two other proteins specified by *lac* operon.
- The reason that lactose enzymes are not made when glucose is present is that no *lac* mRNA is made.

Catabolite Repression

- Transcription-level inhibition of the lactose enzymes and a variety of other inducible enzymes by glucose (or other readily used carbon source) is called **catabolite repression**.

Catabolite Repression

- The mechanism by which glucose inhibits β -galactosidase synthesis remained a complete mystery for about 20 year after Monod first observed the phenomenon.

Catabolite Repression

- Richard S. Makman and Earl W. Sutherland found an important clue to the mystery in 1965 when they observed that the intracellular concentration of 3',5'-cyclic adenylate (cAMP) drops from

Catabolite Repression

- about 10^{-4} to 10^{-7} when glucose is added to a growing culture of *E. coli*.
- Genetic studies confirmed the involvement of cAMP in the catabolite repression.

Catabolite Repression

- Two mutant classes were isolated that could not synthesize *lac* enzymes when cultured in a medium containing lactose but not glucose.

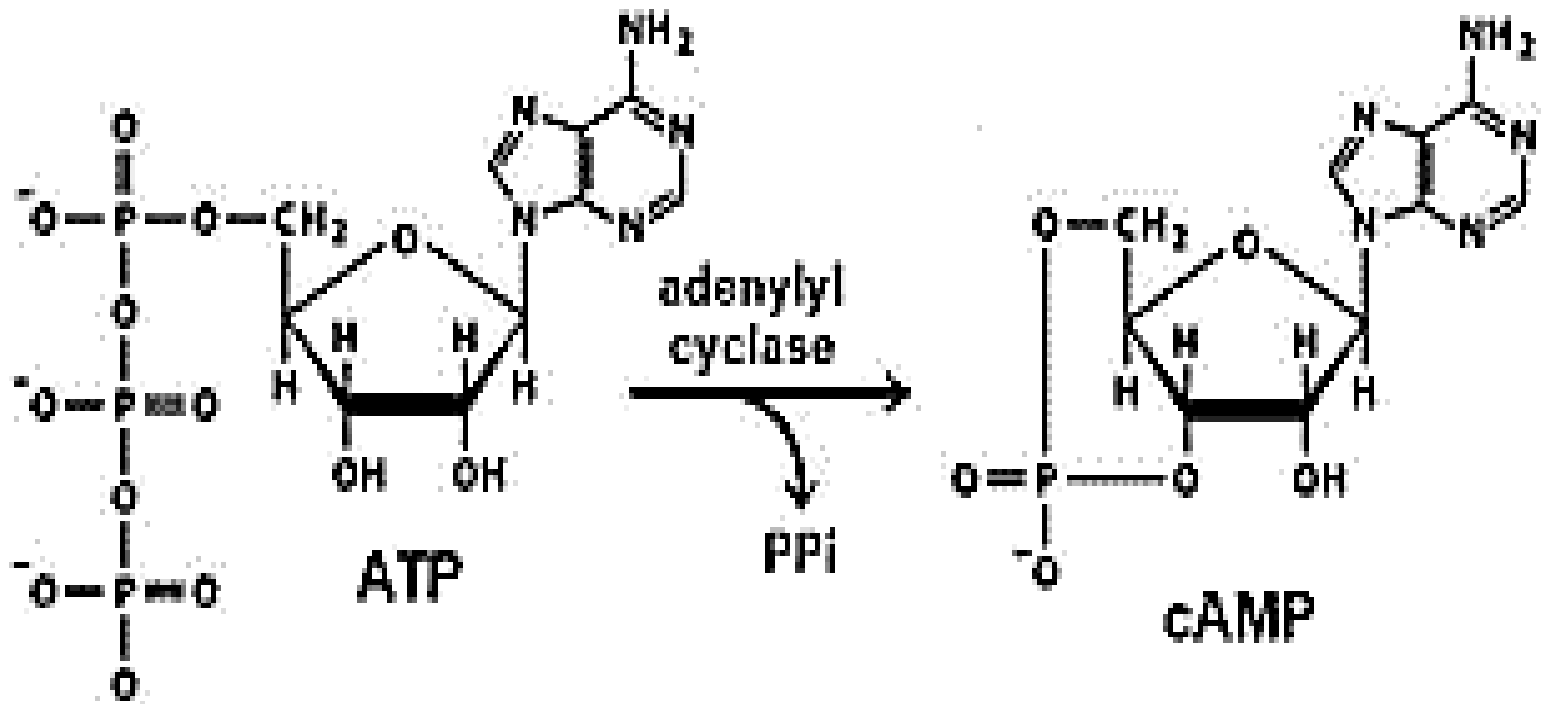
Catabolite Repression

- Class I mutants regained the ability to synthesize *lac* enzymes when cAMP was added to the growth medium but class II mutants did not.

Catabolite Repression

- Subsequent studies showed that class I mutants have defects in **adenylate cyclase**, the enzyme that converts ATP to cAMP.
- Adenylate cyclase exists in an active form that is phosphorylated and an inactive form that is dephosphorylated.

Catabolite Repression



Catabolite Repression

- Class II mutants have defects in a protein that binds cAMP.
- This protein, called the **cAMP receptor protein (CRP)** or the **catabolite activator protein (CAP)**, is encoded by the *crp* gene.

Catabolite Repression

- *In vitro* studies have shown that CRP and cAMP form a **cAMP•CRP complex**, which is needed to activate the *lac* system.

END

Catabolite Repression

- The **cAMP•CRP** is a **positive regulator** or **activator**, in contrast to the repressor, and the *lac* operon is independently regulated both positively and negatively.

Catabolite Repression

- Based on the information available, it may be proposed that glucose somehow inhibits phosphorylation of adenylate cyclase, thereby preventing cAMP formation.

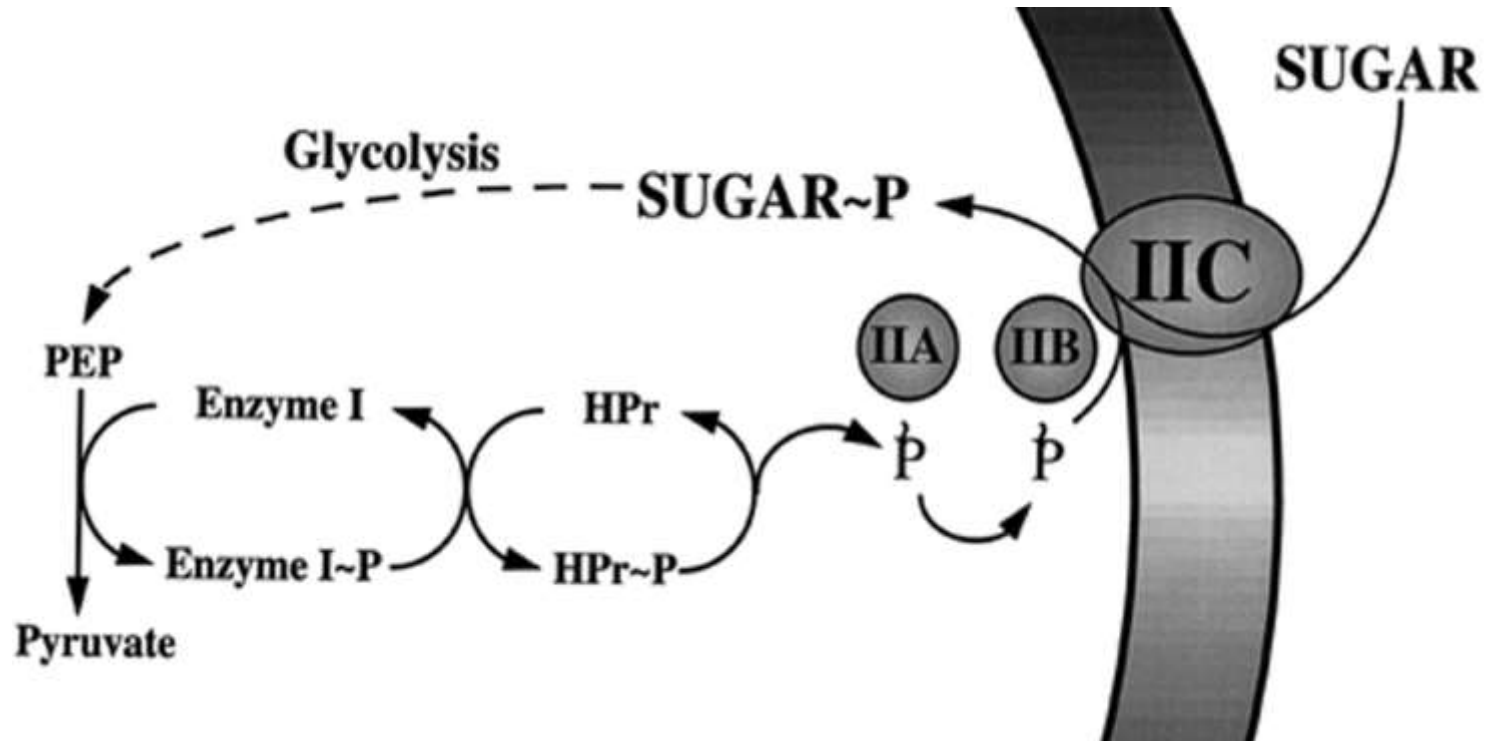
Catabolite Repression

- The next challenge was to find the link between glucose metabolism and adenylate kinase phosphorylation.
- Studies indicated that the link is a glucose-specific, **phosphoenolpyruvate-dependant phosphotransferase system (PTS)**.

Catabolite Repression

- This system uses energy supplied by phosphoenolpyruvate (PEP) to phosphorylate glucose as it transports the sugar across the inner cell membrane.
- The system requires four proteins.

Catabolite Repression



Catabolite Repression

- Two of these, enzyme I (E-I) and the histidine containing protein (Hpr), are also components of other sugar transporter systems and therefore are unlikely to be direct participants in a glucose-specific phenomenon.

Catabolite Repression

- The two other proteins, enzyme II (EIIA) and enzyme IIBC (EBIIC), are specific for the glucose transport system and are more likely participants in a glucose-specific phenomenon.

Catabolite Repression

- EIIA participates in the catabolite repression by two different mechanisms.
- The first mechanism is based on the fact that EIIA can transfer a phosphoryl group from HPr-P to either EIIBC or adenylate cyclase.

Catabolite Repression

- The preferred substrate is EIIBC, which then transfers the phosphoryl group to glucose to form glucose-6-phosphate.

Catabolite Repression

- However, when glucose is unavailable, EIIBC will be fully phosphorylated and EIIA-P has no other alternative but to transfer its phosphoryl group to adenylate cyclase.

Catabolite Repression

- Phosphorylation changes the inactive dephosphorylated form of adenylate cyclase to the active phosphorylated form of adenylate cyclase, which then converts ATP to cAMP.

Catabolite Repression

- Thus, glucose interferes with the conversion of the inactive form of adenylate cyclase to the active form.

Catabolite Repression

- The second mechanism by which glucose influences lac operon transcription is called **inducer exclusion**, which also involves the glucose transport system.

Catabolite Repression

- When glucose is present, EIIA-P transfers its phosphate group through EIIB to the sugar and the dephosphorylated form of EIIA binds to lactose permease and inactivates it.

Catabolite Repression

- Lactose permease inactivation prevents from entering the cell and being converted to allolactose.
- In the absence of allolactose, the repressor remains bound to the operator and the *lac* operon is turned off.

Catabolite Repression

- The cAMP modulation and inducer exclusion mechanisms may not be the only ones that contribute to glucose's ability to inhibit *lac* mRNA formation.
- For example, regulation of CRP synthesis may also be important.

END

The cAMP•CRP complex

- In the absence of cAMP•CRP complex, the *lac* promoter is quite weak because its – 10 box differs significantly from the consensus sequence.

The cAMP•CRP complex

- A mutant *lac* promoter with a – 10 box that has the consensus sequence does not require cAMP•CRP complex for transcription activation.

The cAMP•CRP complex

- It therefore seems reasonable to propose that interactions between cAMP•CRP complex and RNA polymerase holoenzyme increase the holoenzyme's affinity for the *lac* promoter.

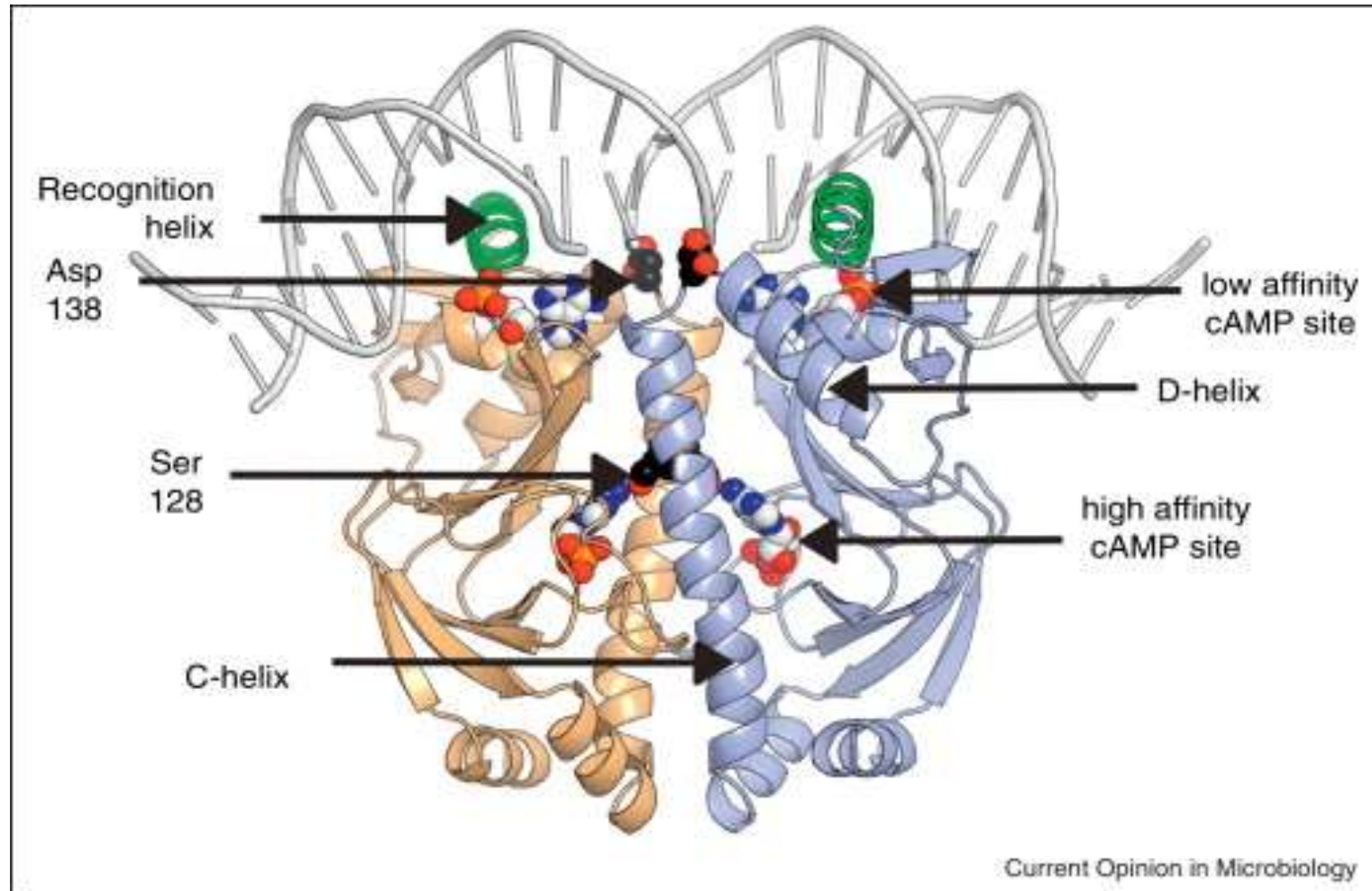
The cAMP•CRP complex

- Thomas Steitz and coworkers determined the crystal structure of the cAMP•CRP complex bound to DNA.
- CRP consists of two identical polypeptide chains of 209 amino acid residues.

The cAMP•CRP complex

- Each chain consists of an N-terminal domain and a C-terminal domain.
- The N-terminal domain contains a pocket for binding cAMP while the C-terminal domain contains a motif that binds to DNA.

The cAMP•CRP complex



The cAMP•CRP complex

- In the absence of cAMP, CRP•DNA interactions are non specific and weak.
- The cAMP•CRP complex however, binds very tightly to a specific DNA sequence designated as the **activator site (AS)**.

The cAMP•CRP complex

- In the *lac* operon, the center of AS is 61.5 bp upstream from the transcription start site.
- Promoters like the *lac* promoter that have AS at position – 61.5 are designated as **class I cAMP-dependant promoter**.

The cAMP•CRP complex

- Binding of cAMP•CRP to the consensus sequence is so tight that an operon containing this sequence would be permanently switched on.
- It is therefore not surprising that actual cAMP•CRP activator sites differ from the consensus sequence.

The cAMP•CRP complex

- Activator sites in different operons compete for the cAMP•CRP with the activator preferentially binding to sequences that most closely resemble the consensus sequence.

The cAMP•CRP complex

- cAMP•CRP bound to DNA migrates in an anomalous fashion when subjected to non denaturing gel electrophoresis.
- The most likely explanation of this is that the protein causes the DNA to bend when it binds to it.

The cAMP•CRP complex

- X-ray crystallography studies confirmed this interpretation, showing that the cAMP•CRP complex sharply bends DNA by an angle of between 80° to 90° .

The cAMP•CRP complex

- In addition to specific contacts with AS, the cAMP•CRP complex also makes specific contacts with RNA polymerase holoenzyme.

END


Galactose Operon

- The *gal* operon, one of the best studied catabolite sensitive operon, contains four structural genes *galK*, *galT*, *galE* and *galM* which specifies the enzymes galactokinase, galactose transferase, galactose epimerase and mutarotase.

Galactose Operon

- Mutarotase converts the β -D-galactose formed when β -galactosidase hydrolyzes lactose into α -D-galactose.

Galactose Operon

- Then galactokinase, galactose transferase and epimerase act in a sequence of steps to yield the overall reaction:-
- Galactose + ATP  Glucose-1-phosphate + ADP

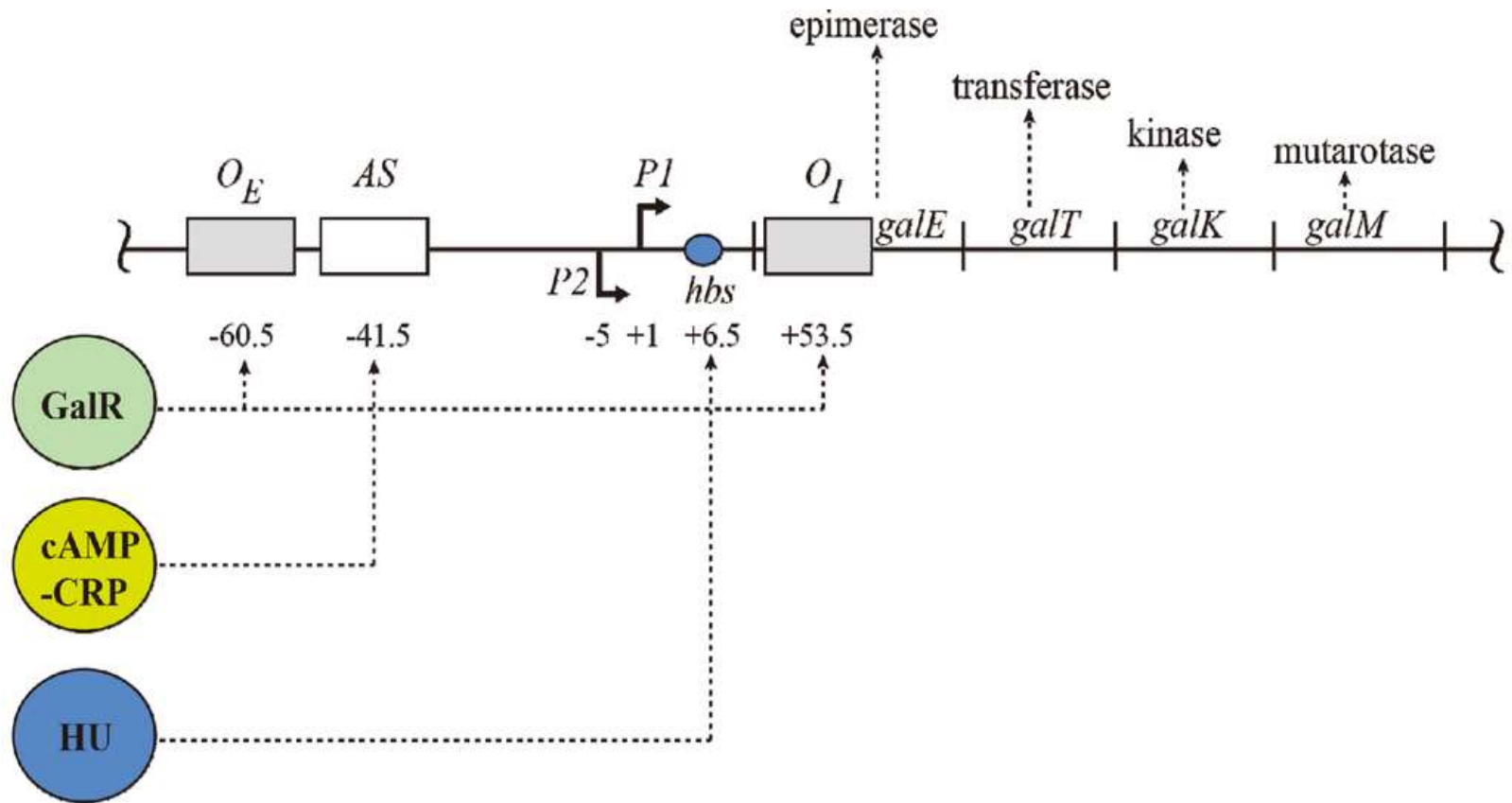
Galactose Operon

- The galactose (*gal*) operon, like the *lac* operon, is regulated by a repressor and cAMP•CRP.
- The structural gene for the repressor, *galR*, is located far from structural genes for the galactose enzymes.

Galactose Operon

- The *gal* operon also has two operators, $galO_E$ (O_E) and $galO_I$ (O_I), a promoter region, and a cAMP•CRP AS.
- The promoter region contain two promoters P_1 and P_2 .

Galactose Operon



Galactose Operon

- One operators, $galO_E$, is upstream from the promoter region and the other $galO_I$, is in $galE$.
- A mutation in either operator causes at least partial constitutivity, indicating that both operators participate in repression.

Galactose Operon

- The Gal repressor, which binds to both operators, is made of two identical polypeptide chains.
- A null mutation in *galR* makes both promoters constitutive.

Galactose Operon

- The amino acid sequence of Gal repressor is similar to that of Lac repressor.
- There is one important difference, however,

Galactose Operon

- the Gal repressor does not have the C-terminal residues that allow the Lac repressor to form a four helical bundle and become a stable tetramer.

Galactose Operon

- The Gal repressor has a C-terminal domains that binds the inducer galactose and an NTD with a helix-turn-helix motif that binds to a $galO_E$ or $galO_I$ half-site.

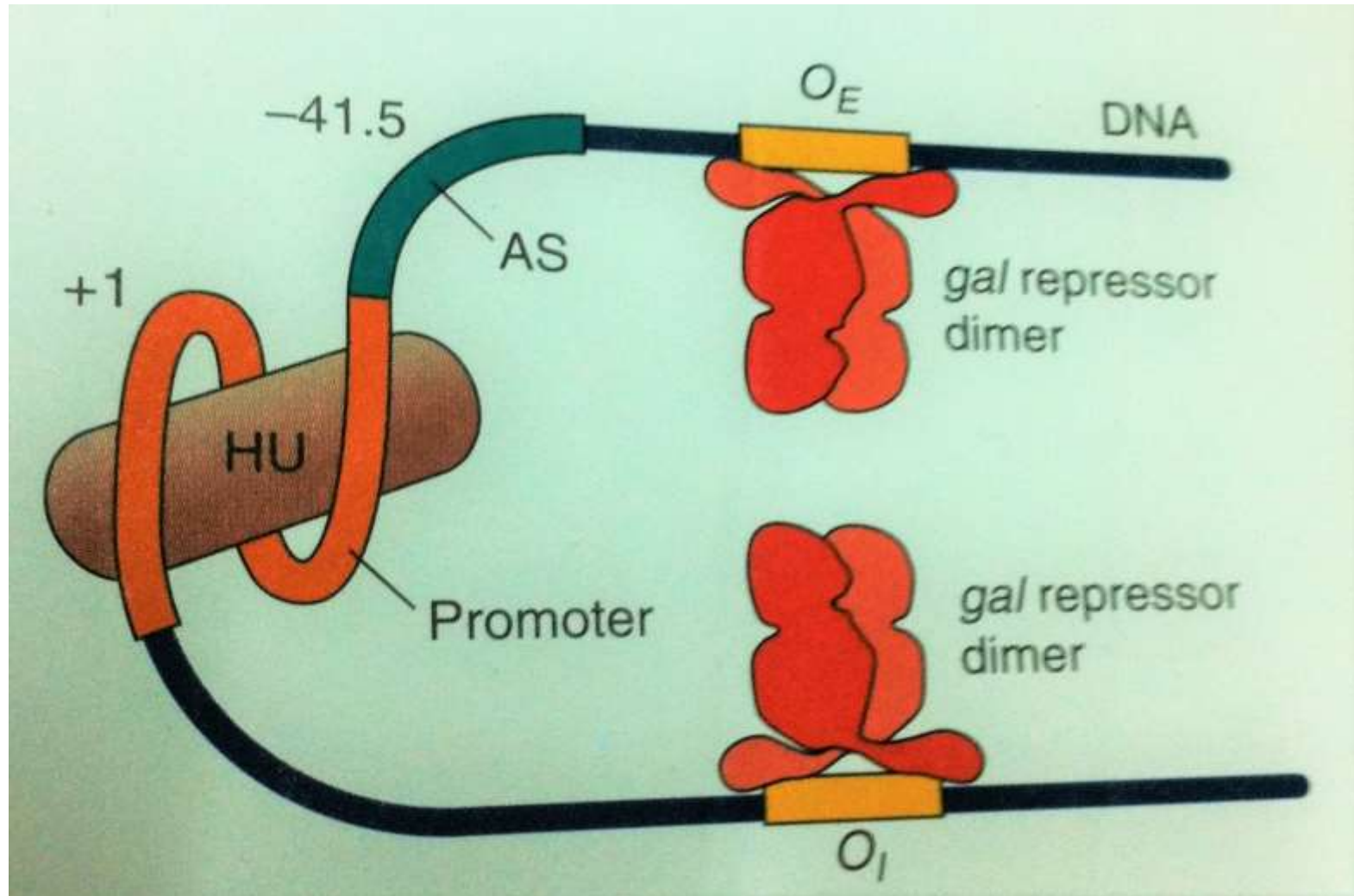
Galactose Operon

- The Gal repressor is a dimer and not a tetramer and the studies which indicated that both $galO_E$ and $galO_I$ are required for repression, suggested some type of cooperative interaction.

Galactose Operon

- Sanker Adhya and coworkers solved the problem by showing that a histone-like protein called HU binds to the *gal* promoter region only when the Gal repressors are bound at both $galO_E$ or $galO_I$ causing the DNA to loop.

Galactose Operon



Galactose Operon

- The two *gal* promoter (P_1 & P_2) are separated by 5 bp and as a result, the transcription initiation site for P_1 is 5 bp upstream from that for P_2 .

END

Galactose Operon

- Polycistronic mRNA molecules formed in response to either promoter contain all the information required to synthesize the four galactose enzymes.

Galactose Operon

- Binding cAMP•CRP to AS repress transcription from P_2 but activates transcription from P_1 .

Galactose Operon

- Moreover, P_2 can support *gal* mRNA transcription in the presence of glucose when intracellular cAMP•CRP levels are low but P_1 cannot do so.

Galactose Operon

- The fact that AS is 20 bp closer to P_1 in the *gal* operon than it is to P_{lac} in the *lac* operon leads one to predict that cAMP•CRP and RNA polymerase interactions should be different in the *gal* and *lac* operons.

Galactose Operon

- This prediction is correct.
- Promoters like P_1 in *gal* operon that have AS at position – 41.5 are called **class II cAMP-dependant promoters**.

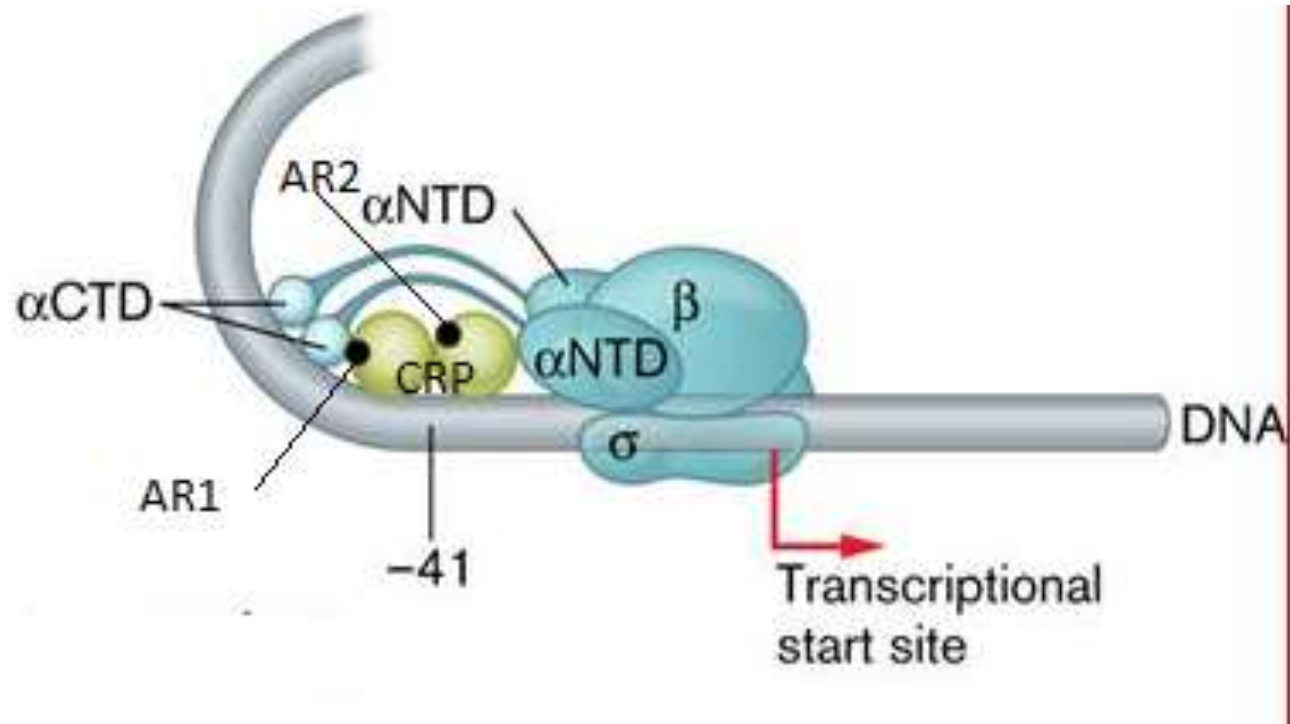
Galactose Operon

- Three activator regions on CRP – **AR1**, **AR2** and **AR3** – interact with RNA polymerase holoenzyme.

Galactose Operon

- AR1 on the upstream unit of the CRP dimer and AR2 and AR3 on the downstream subunit of dimer interact with α CTD, α NTD, and the σ^{70} subunit, respectively.

Galactose Operon



Galactose Operon

- Why does the *gal* operon has two promoters?
- The answer appears to be that galactose has two roles in cellular metabolism.
- It is both a carbon source and a precursor for lipopolysachharide synthesis.

Galactose Operon

- When galactose is not available in the growth medium, cells require the epimerase specified by the *gal* operon to convert glucose to galactose, which is then used to make lipopolysachharides.

Galactose Operon

- Synthesis from the P_2 promoter permits the low level of epimerase formation required to convert glucose to galactose-1-phosphate so that the lipopolysachharides can be formed.

Galactose Operon

- If P_1 were the only promoter, then epimerase could not be made when glucose is present because P_1 require cAMP•CRP activation.

Galactose Operon

- On the other hand, if P_2 were the only promoter, then galactose could not fully induce the operon when galactose was the sole carbon source because cAMP•CRP inhibits P_2 .

Galactose Operon

- Thus, for the sake of both necessity and economy, a cAMP•CRP – independent promoter (P_2) is needed for background constitutive synthesis and a cAMP•CRP – dependent promoter

Galactose Operon

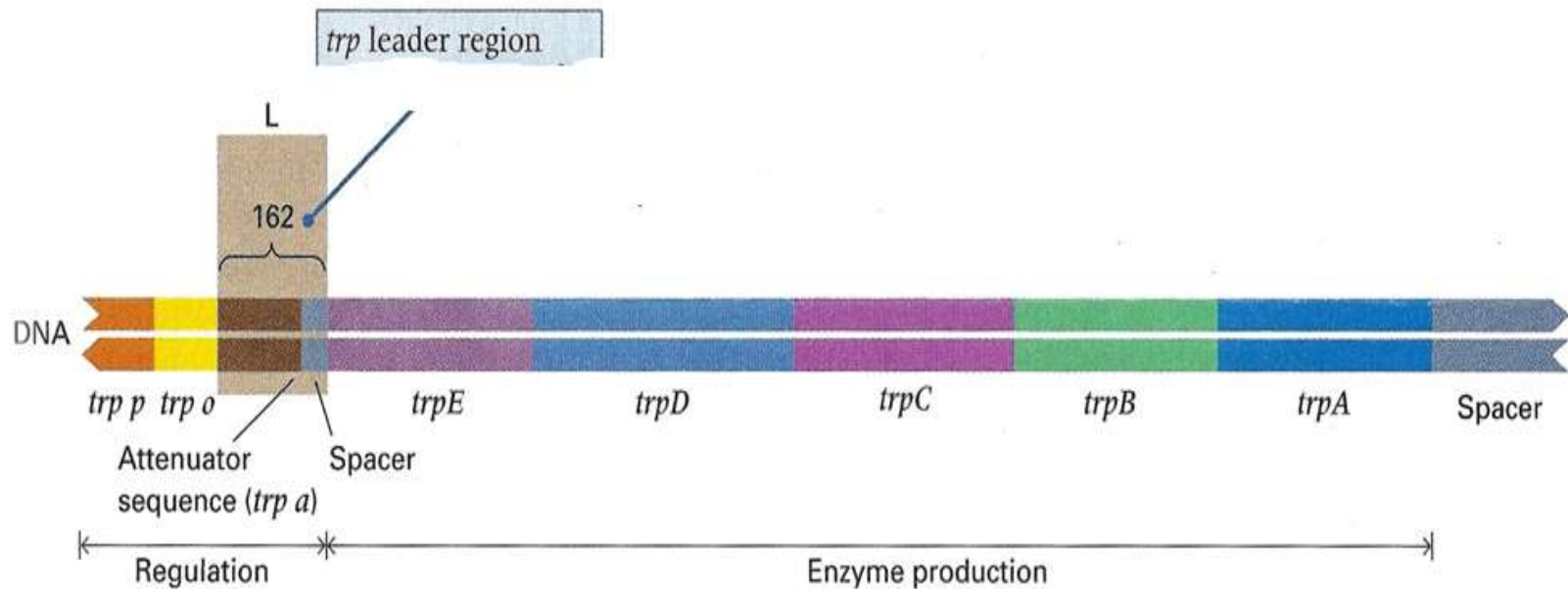
- (P_1) is needed to regulate high-level synthesis.
- Furthermore, the regulation is efficient only if P_2 is inhibited by cAMP•CRP.

END

Tryptophan Operon

- The *trp* operon consists of a promoter, an operator, a leader (*trpL*), an attenuator, and five structural genes designated as *trpE*, *trpD*, *trpC*, *trpB* and *trpA*.

Tryptophan Operon



Tryptophan Operon

- Consistent with the fact that the *trp* operon specifies biosynthetic rather than degradative enzymes, it does not have a cAMP•CRP activation site.

Tryptophan Operon

- The operator is part of a coarse on-off control, and the leader and attenuator allow for finer control.
- The leader sequence, which codes for a short peptide is the first region of the *trp* operon to be transcribed.

Tryptophan Operon

- RNA polymerase then moves forward to transcribe the five structural genes in the order *trpE* → *trpA* to form the polycistronic *trp* mRNA.

Tryptophan Operon

- The five polypeptides specified by the structural genes form the three enzymes that are essential for tryptophane biosynthesis.

Tryptophan Operon

- Early studies revealed that the *trp* system is turned off when tryptophane is added to an *E. coli* culture, suggesting that the operon is repressed rather than induced.

Tryptophan Operon

- Repression, like induction, involves negative regulation of transcription initiation by a regulatory protein.
- However, the regulatory protein does not bind to the operator until after forming a complex with tryptophan.

Tryptophan Operon

- The structural gene for the Trp regulatory protein, *trpR*, is located a considerable distance away from the *trp* operon.

Tryptophan Operon

- This distance does not present a problem because the regulatory protein diffuses throughout the cell.
- The biologically active form of the regulatory protein is a homodimer.

Tryptophan Operon

- Each subunit contains a helix-turn-helix motif that can bind to a *trp* operator half-site.
- The *trp* operator and promoter regions have significant overlap.
- Therefore, binding of the TrpR•tryptophane complex and RNA polymerase are mutually exclusive.

Tryptophan Operon

- The TrpR•tryptophane complex functions as a coarse on-off switch that turns the tryptophane operon off when tryptophane levels are high.

Tryptophan Operon

- A fine control mechanism also exists that allows cells to regulate their tryptophane enzyme concentration according to the tryptophane concentration.

Tryptophan Operon

- An important clue to the existence of this fine control mechanism came from an experiment performed by Charles Yanofsky in 1972, which showed that *E. coli* mutants that lack a functional TrpR protein

Tryptophan Operon

- increase *trp* operon transcription after being starved for tryptophan.
- If TrpR•tryptophan complex were the only regulatory factor, then transcription of the *trp* operon should not have increased.

Tryptophan Operon

- Further studies revealed that *trp* mRNA has a 162 nucleotide sequence before the first codon in *trpE*, designated as the **leader** or *trpL*, which plays an essential role in the fine control mechanism.

Tryptophan Operon

- Constitutive mutants exhibit a sixfold increase in tryptophane enzyme synthesis when bases 123 – 150 within the leader are deleted.
- This 28 base sequence is called the **attenuator**.

Tryptophan Operon

- The attenuator can fold into a stem-and-loop structure with the potential to function as a rho-independent transcription terminator.

Tryptophan Operon

- Deleting the attenuator removes the transcription termination site, allowing RNA polymerase to complete *trp* mRNA synthesis.

END

Tryptophan Operon

- The *trp* leader has four complementary segments that can interact to form two sets of mutually exclusive hairpin structures.
- Segment 1 & 2 can base pair to form hairpin 1•2 while segment 3 & 4 can form hairpin 3•4.

Tryptophan Operon

- Alternatively, segment 2 & 3 can base pair to form hairpin 2•3.
- The 3•4 hairpin, which contains the attenuator sequence, is a transcription terminator.

Tryptophan Operon

- It follows that RNA polymerase would be able to synthesize full length *trp* mRNA if conditions were somehow favorable for 2•3 hairpin formation rather than 3•4 hairpin formation.

Tryptophan Operon

- RNA polymerase molecules that escape repression begin synthesizing *trp* mRNA.
- RNA polymerase continues transcribing *trp* operon until it encounters a pause site located just after segment 2 of the leader sequence.

Tryptophan Operon

- Because the leader sequence is at the 5' end of *trp* mRNA, it is the first sequence available for translation.
- Ribosomes begin translating the leader sequence at its AUG start codon.

Tryptophan Operon

- When the moving ribosome reaches the paused RNA polymerase, the paused RNA polymerase is released.
- Pausing serves the important role of synchronizing the transcription and translation process.

Tryptophan Operon

- The subsequent fate of the RNA polymerase depends on the tryptophan concentration i.e.,
 1. high tryptophan concentration or
 2. low tryptophan concentration.

Tryptophan Operon

- In the first case, ribosome moves past the tryptophan codon in the segment 1 and continues to translate the leader region until it encounters the stop codon (UGA) between segment 1 & 2 and falls off the nascent mRNA molecule.

Tryptophan Operon

- Once free of the ribosome, segment 1 pairs with segment 2 to form the 1•2 hairpin.
- RNA polymerase continues to transcribe the leader region, synthesizing segment 3 and then segment 4.

Tryptophan Operon

- These two segments pair to form the rho independent transcription terminator, which causes RNA polymerase to fall off the DNA template, preventing *trpE* transcription.

Tryptophan Operon

- In the second case when the concentration of tryptophan is very low, the bulky ribosome will pause at the tryptophane codon (UGG) on segment 1, preventing segment 1 from pairing with segment 2 to form the 1•2 hairpin.

Tryptophan Operon

- Segment 2 is therefore free to interact with segment 3 to form 2•3 hairpin (the so called **antiterminator**) as soon as the RNA polymerase completes the synthesis of segment 3.

Tryptophan Operon

- Then RNA polymerase can continue to synthesize the complete *trp* mRNA because the 3•4 hairpin (the transcription terminator) is not formed.

Tryptophan Operon

- Thus, if tryptophan is present in excess, transcription termination occurs at the attenuator and little enzyme is synthesized.

Tryptophan Operon

- If tryptophan is absent, transcription termination does not occur and tryptophan enzymes are formed.

END

Gene Expression Is Controlled by Regulatory Proteins

- Genes are very often controlled by extracellular signals; in the case of bacteria, these signals typically mean molecules present in the growth medium.

Gene Expression Is Controlled by Regulatory Proteins

- These signals are communicated to genes by regulatory proteins, which come in two types: positive regulators, or **activators**, and negative regulators, or **repressors**.

Gene Expression Is Controlled by Regulatory Proteins

- 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.

Gene Expression Is Controlled by Regulatory Proteins

- How do these regulators work?
- First, RNA polymerase binds to the promoter in a closed complex (in which the DNA strands remain together).

Gene Expression Is Controlled by Regulatory Proteins

- The polymerase–promoter complex then undergoes a transition to an open complex in which the DNA at the start site of transcription is unwound and the polymerase is positioned to initiate transcription.

Gene Expression Is Controlled by Regulatory Proteins

- This is followed by promoter escape, the step in which polymerase leaves the promoter and starts transcribing.
- Polymerase then proceeds through the elongation phase before finally terminating.

Gene Expression Is Controlled by Regulatory Proteins

- Which steps are stimulated by activators and inhibited by repressors depends on the promoter and regulators in question.

Gene Expression Is Controlled by Regulatory Proteins

- The most common step at which gene expression is regulated is the initiation of transcription.
- There are two reasons why this might make sense.

Gene Expression Is Controlled by Regulatory Proteins

- First, transcription initiation is the most energetically efficient step to regulate.
- By this we mean that deciding whether or not to express a gene at the first step ensures that no energy or resources are wasted.

Gene Expression Is Controlled by Regulatory Proteins

- Second, regulation at this first step is easier to do.
- There is only a single copy of each gene and so typically only a single promoter on a single DNA molecule must be regulated to control expression of a given gene.

Gene Expression Is Controlled by Regulatory Proteins

- Why then is not all regulation focused on the step of transcription initiation?
- Regulating later steps can have two advantages.

Gene Expression Is Controlled by Regulatory Proteins

- First, it allows for more inputs: if a gene is regulated at more than one step, more signals can modulate its expression, or the same signals can do so even more effectively.

Gene Expression Is Controlled by Regulatory Proteins

- Second, regulation at steps later than transcription initiation can reduce the response time.
- Thus, consider again the example of translational regulation.

Gene Expression Is Controlled by Regulatory Proteins

- If a signal relieves repression of this step, the protein product encoded by the gene will be produced immediately upon receipt of that signal.
- This reduced response time might obviously be advantageous in some situations.

Gene Expression Is Controlled by Regulatory Proteins

- Many promoters are regulated by activators that help RNA polymerase bind DNA and by repressors that block that binding.

Gene Expression Is Controlled by Regulatory Proteins

- At many promoters, in the absence of regulatory proteins, RNA polymerase binds only weakly.
- This is because one or more of the promoter elements is absent or imperfect.

Gene Expression Is Controlled by Regulatory Proteins

- When polymerase does occasionally bind, however, it spontaneously undergoes a transition to the open complex and initiates transcription.

Gene Expression Is Controlled by Regulatory Proteins

- This gives a low level of **constitutive expression** called the **basal level**.
- Binding of RNA polymerase is the rate-limiting step in this case.

Gene Expression Is Controlled by Regulatory Proteins

- To control expression from such a promoter, a repressor need only bind to a site overlapping the region bound by polymerase.

Gene Expression Is Controlled by Regulatory Proteins

- In that way, the repressor blocks polymerase binding to the promoter, thereby preventing transcription, although it is important to note that repression can work in other ways as well.

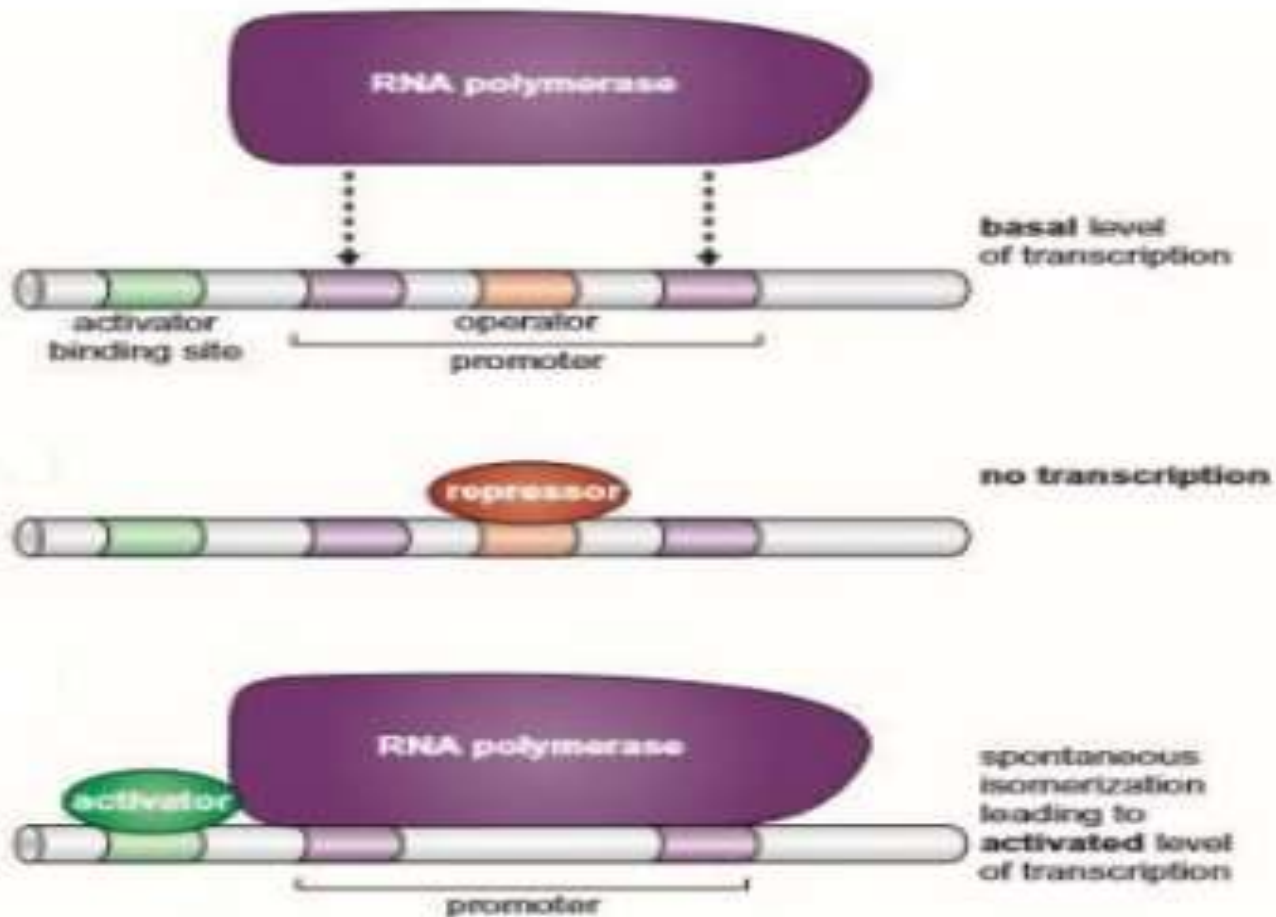
Gene Expression Is Controlled by Regulatory Proteins

- The site on DNA where a repressor binds is called an **operator**.
- To activate transcription from this promoter, an activator can just help the polymerase bind the promoter.
- Typically, this is achieved as follows:

Gene Expression Is Controlled by Regulatory Proteins

- the activator uses one surface to bind to a site on the DNA near the promoter; with another surface, the activator simultaneously interacts with RNA polymerase, bringing the enzyme to the promoter.

Gene Expression Is Controlled by Regulatory Proteins



Gene Expression Is Controlled by Regulatory Proteins

- This mechanism, often called **recruitment**, is an example of **cooperative binding** of proteins to DNA.
- The interactions between the activator and polymerase, and between activator and DNA, serve merely “adhesive” roles.

Gene Expression Is Controlled by Regulatory Proteins

- The enzyme is active and the activator simply brings it to the nearby promoter.
- Once there, it spontaneously isomerizes to the open complex and initiates transcription.

END

Some Activators and Repressors Work by Allostery

- Let us consider another class of promoter in which RNA polymerase binds efficiently unaided and forms a stable closed complex.

Some Activators and Repressors Work by Allostery

- But that closed complex does not spontaneously undergo transition to the open complex.
- At this promoter, an activator must stimulate the transition from a closed to open complex, since that transition is the rate-limiting step.

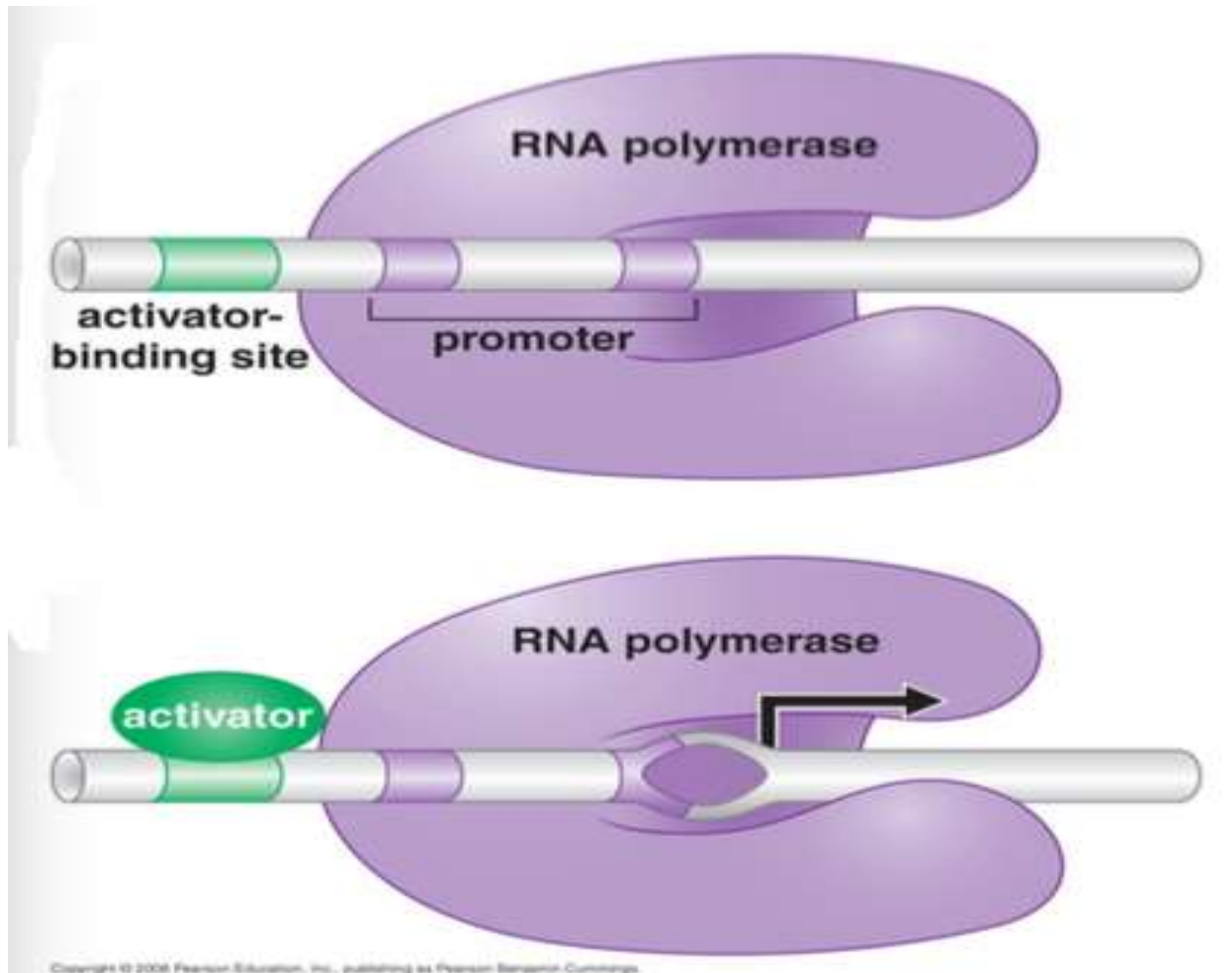
Some Activators and Repressors Work by Allostery

- Activators that stimulate this kind of promoter work by triggering a conformational change in either RNA polymerase or DNA.

Some Activators and Repressors Work by Allostery

- It means that, they interact with the stable closed complex and induce a conformational change that causes transition to the open complex.
- This mechanism is an example of **allostery**.

Some Activators and Repressors Work by Allostery



Some Activators and Repressors Work by Allostery

- For example, the *glnA* promoter and *NtrC* the activator.
- In still another class of promoter, transcription initiation is limited at the step of promoter escape.

Some Activators and Repressors Work by Allostery

- One example of such a promoter directs expression of the *malT* gene.
- In the absence of an activator, it undergoes abortive initiation, and only in the presence of an activator will it efficiently escape into elongation.

Some Activators and Repressors Work by Allostery

- Repressors can work in ways other than just blocking the binding of RNA polymerase.
- For example, some repressors interact with polymerase at the promoter and inhibit transition to the open complex, or promoter escape.

Some Activators and Repressors Work by Allostery

- Generally, DNA-binding proteins that interact with each other bind to adjacent sites.
- But some proteins interact with each other even when bound to sites well separated on the DNA.

Some Activators and Repressors Work by Allostery

- To accommodate this interaction, the DNA between the sites loops out, bringing the sites into proximity with one another.
- Distant DNA sites can be brought closer together to help loop formation.

Some Activators and Repressors Work by Allostery

- In bacteria, for example, there are cases in which a protein binds between an activator-binding site and the promoter and helps the activator interact with polymerase by bending the DNA in a favorable direction.

Some Activators and Repressors Work by Allostery

- There are also cases where such a protein hinders loop formation and activation by bending the DNA in an unfavorable direction.

END

Cooperative Binding and Allostery

- It is already pointed out that gene activation can be mediated by simple cooperative binding.
- The activator interacts simultaneously with DNA and with polymerase and so recruits the enzyme to the promoter.

Cooperative Binding and Allostery

- How activation can be mediated by allosteric events: an activator interacts with polymerase already bound to the promoter and, by inducing a conformational change in the enzyme or the promoter, stimulates transcription initiation.

Cooperative Binding and Allostery

- Both cooperative binding and allostery have additional roles in gene regulation.
- For example, groups of regulators often bind DNA cooperatively:-

Cooperative Binding and Allostery

- two or more activators and/or repressors interact with each other and with DNA and thereby help each other bind near a gene they all regulate.

Cooperative Binding and Allostery

- This kind of interaction can produce sensitive switches that allow a gene to go from completely off to fully on in response to only small changes in conditions.

Cooperative Binding and Allostery

- Cooperative binding of activators can also serve to integrate signals: some genes are activated only when multiple signals (and thus multiple regulators) are simultaneously present.

Cooperative Binding and Allostery

- A particularly striking and well-understood example of cooperativity in gene regulation is provided by bacteriophage λ .

Cooperative Binding and Allostery

- Allostery is not only a mechanism of gene activation, but also often the way regulators are controlled by their specific signals.

Cooperative Binding and Allostery

- Thus a typical bacterial regulator can adopt two conformations: in one, it can bind DNA; in the other, it cannot.

Cooperative Binding and Allostery

- Binding of a signal molecule locks the regulatory protein in one or another conformation, thereby determining whether or not it can act.

Cooperative Binding and Allostery

- The bulk of gene regulation takes place at the initiation of transcription.
- This is true in eukaryotes just as it is in bacteria. But regulation is certainly not restricted to that step in either class of organisms.

Cooperative Binding and Allostery

- In bacteria, gene regulation also occurs at the level of transcriptional elongation and termination and even at translational levels.

Cooperative Binding and Allostery

- There are also cases involving regulation by RNAs e.g., attenuation, riboswitches, and small RNAs.

END

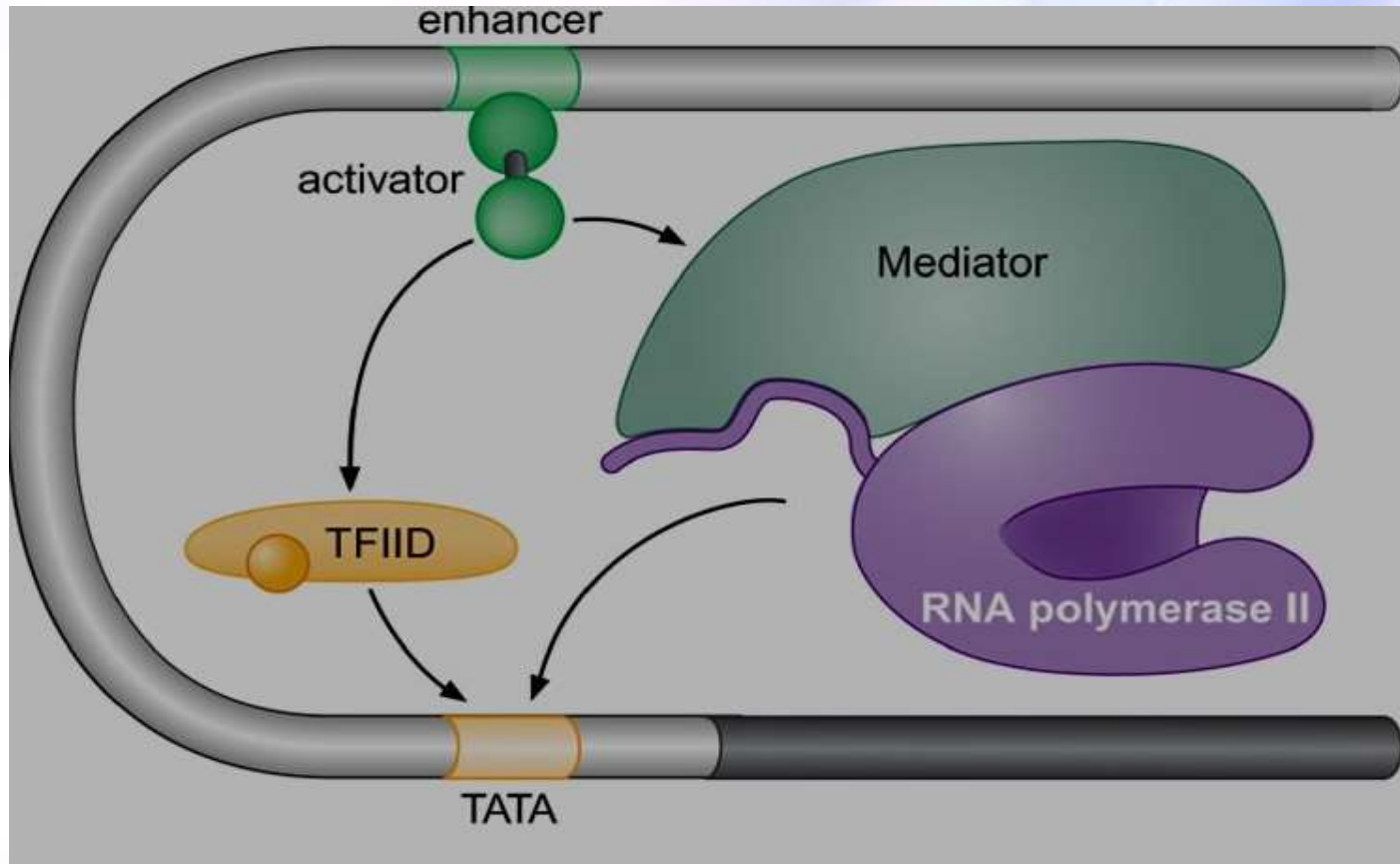
Activators Recruit the Transcriptional Machinery

- The eukaryotic transcriptional machinery contains numerous proteins in addition to RNA polymerase.
- Many of these proteins come in preformed complexes such as mediator and the TFIID Complex.

Activators Recruit the Transcriptional Machinery

- Activators interact with one or more of these complexes and recruit them to the gene.
- Other components that are not directly recruited by the activator bind cooperatively with those that are recruited.

Activators Recruit the Transcriptional Machinery



Activators Recruit the Transcriptional Machinery

- At most genes , the transcriptional machinery appears at the promoter only upon activation of the gene.
- That is, the machinery is not pre-bound, confirming that the role of the activator is to recruit it.

Activators Recruit the Transcriptional Machinery

- In bacteria, we saw that genes activated by recruitment (such as the lac genes) can be activated in so-called activator bypass experiments.

Activators Recruit the Transcriptional Machinery

- In such an experiment, activation is observed when RNA polymerase is recruited to the promoter without using the natural activator –polymerase interaction.
- Similar experiments work in yeast.

Activators Recruit the Transcriptional Machinery

- Thus, the GAL1 gene (normally activated by Gal4) can be activated equally well by a fusion protein containing the DNA-binding domain of the bacterial protein LexA fused directly to a component of the Mediator complex.

Activators Recruit the Transcriptional Machinery

- It is important to note that these experiments do not exclude the possibility that at least some activators not only recruit parts of the transcriptional machinery but also induce allosteric changes in them.

Activators Recruit the Transcriptional Machinery

- Such changes might stimulate the efficiency of transcription initiation.

Activators Recruit the Transcriptional Machinery

- Nevertheless, the recruitment of the machinery to one or another gene is the basis of specificity; that is, which gene is activated depends on which gene has the machinery recruited to it.

Activators Recruit the Transcriptional Machinery

- In addition, the success of the activator bypass experiments suggests that any allosteric events that occur during initiation do not require the activator to do anything beyond recruiting proteins to the gene.

Activators Recruit Nucleosome Modifiers

- In addition to direct recruitment of the transcriptional machinery, recruitment of nucleosome modifiers can help activate a gene packaged within chromatin.

Activators Recruit Nucleosome Modifiers

- Nucleosome modifiers come in two types:
- 1) those that add chemical groups to the tails of histones, such as **histone acetyltransferases** (HATs), which add acetyl groups, and

Activators Recruit Nucleosome Modifiers

- 2) those that displace (or “remodel”) the nucleosomes, such as the ATP-dependent activity of **SWI/SNF**.
- How do these modifications help activate a gene?

Activators Recruit Nucleosome Modifiers

- Two basic models explain how changes in nucleosomes can help the transcriptional machinery bind at the promoter.

Activators Recruit Nucleosome Modifiers

- First, remodeling, and certain modifications, can uncover DNA-binding sites that would otherwise remain inaccessible within the nucleosome.

Activators Recruit Nucleosome Modifiers

- For example, by removing or increasing the mobility of nucleosomes, remodelers are proposed to free up binding sites for regulators and for the transcriptional machinery.

Activators Recruit Nucleosome Modifiers

- Similarly, the addition of acetyl groups to histone tails alters the interactions between those tails and adjacent nucleosomes.
- This modification is often said to “loosen” chromatin structure, freeing up sites.

Activators Recruit Nucleosome Modifiers

- But adding acetyl groups also helps binding of the transcriptional machinery (and other proteins) in another way: it creates specific binding sites on nucleosomes for proteins bearing so-called **bromodomains**.

Activators Recruit Nucleosome Modifiers

- One component of the TFIID complex bears bromodomains and thus binds to acetylated nucleosomes better than to unacetylated nucleosomes.

Activators Recruit Nucleosome Modifiers

- Thus, a gene bearing acetylated nucleosomes at its promoter will likely have a higher affinity for the transcriptional machinery than one with unacetylated nucleosomes.

Activators Recruit Nucleosome Modifiers

- Other proteins contain **chromodomains**.
- These recognize methylated nucleosomes, examples of which we will encounter later.

Activators Recruit Nucleosome Modifiers

- Some components of the transcriptional machinery are more stringently required at some genes than at others, and the same applies to nucleosome modifiers as well.

Activators Recruit Nucleosome Modifiers

- Thus, although all genes absolutely require RNA polymerase itself, a given gene may depend on another particular component of the transcriptional machinery, or a nucleosome modifier, or it may not.

Activators Recruit Nucleosome Modifiers

- In addition, what is needed to activate a given gene can vary depending on circumstances, such as the stage of the cell cycle.
- For example, Gal4 usually activates the *GAL1* gene efficiently in the absence of a histone acetylase.

Activators Recruit Nucleosome Modifiers

- During mitosis, however, when chromatin is more condensed, activation is eliminated unless that acetylase is recruited to the gene.

Activators Recruit Nucleosome Modifiers

- In yeast, recent experiments have provided good evidence for particular activator–target interactions at specific genes.

Activators Recruit Nucleosome Modifiers

- The acidic activator Gcn4 is known to interact with Gal11; it also interacts with the TAF12 subunit of TFIID and other complexes involved in transcription, including the nucleosome remodeler SWI/SNF.

Activators Recruit Nucleosome Modifiers

- And Gal4 appears to contact at least three components: Mediator, TFIID, and a third complex called SAGA (Spt-Ada-Gcn5-acetyltransferase).

Activators Recruit Nucleosome Modifiers

- The last of these complexes harbors acetylation activity and seems to be capable of interacting with the transcriptional machinery as well.

Activators Recruit Nucleosome Modifiers

- The ability of an acidic activator such as Gal4 to work at genes with different requirements can be explained by its ability to interact with multiple targets.

Activators Recruit Additional Factors

- The elaborate transcriptional machinery of a eukaryotic cell contains numerous proteins required for initiation.
- It also contains some that aid in elongation.

Activators Recruit Additional Factors

- At some genes, sequences downstream from the promoter cause pausing or stalling of the polymerase soon after initiation.

Activators Recruit Additional Factors

- At those genes, the presence or absence of certain elongation factors greatly influences the level at which the gene is expressed.

Activators Recruit Additional Factors

- One example is the *HSP70* gene from *Drosophila*.
- This gene, activated by heat shock, is controlled by two activators working together.

Activators Recruit Additional Factors

- The GAGA binding factor is believed to recruit enough of the transcription machinery to the promoter for initiation of transcription.

Activators Recruit Additional Factors

- But, in the absence of a second activator, HSF, most of the initiated polymerases stall some 25–50 bp downstream from the promoter.

Activators Recruit Additional Factors

- In response to heat shock, HSF binds to specific sites at the promoter and recruits a kinase, P-TEFb (positive transcription elongation factor), to the stalled polymerases.

Activators Recruit Additional Factors

- The kinase phosphorylates the carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase (the so-called CTD “tail”), freeing the enzyme from the stall and allowing transcription to proceed through the gene.

Activators Recruit Additional Factors

- Recent studies suggest that P-TEFb is part of a larger complex, the SEC (super elongation complex), which releases paused Pol II from the proximal promoter.

Activators Recruit Additional Factors

- Phosphorylation of the CTD tail on Ser5 of the heptad repeat is an important step in the early stages of transcription at all genes, and the kinase TFIIF can perform that phosphorylation.
- Whether P-TEFb is also needed at most genes is not clear.

Activators Recruit Additional Factors

- P-TEFb has been implicated in the phosphorylation of Ser2 of the CTD heptad repeat, and this modification is associated with the release of activated Pol II from promoter sequences.

Activators Recruit Additional Factors

- A strong acidic activator like Gal4 is able to recruit PTEFb/SEC along with the rest of the machinery.

Activators Recruit Additional Factors

- It may be that only at certain genes is the recruitment of the machinery partitioned between regulators in the way we see at this *HSP70* gene, allowing an extra layer of control.

Activators Recruit Additional Factors

- The human immunodeficiency virus (HIV), which causes AIDS, transcribes its genes from a promoter controlled by P-TEFb (and SEC).

Activators Recruit Additional Factors

- Again, polymerase initiates transcription at that promoter, under the control of the activator SP1, but stalls soon afterward.
- In this case, P-TEFb is brought to the stalled polymerase by an RNA-binding protein called TAT.

Activators Recruit Additional Factors

- TAT recognizes a specific sequence near the start of the HIV RNA and present in the nascent transcript made by the stalled polymerase.
- Another domain of TAT interacts with P-TEFb and recruits it to the stalled polymerase.

Activators Recruit Additional Factors

- This results in the release of polymerase, the transcription of the viral genome, and infection of the host cell, typically a T-lymphocyte.
- It is now believed that paused polymerase is more commonly seen, particularly during development.

Activators Recruit Additional Factors

- Thus, recent studies in human embryonic stem cells and the early *Drosophila* embryo suggest that roughly one-third of all protein-coding genes contain paused Pol II before their activation during development.

Activators Recruit Additional Factors

- Such genes might be particularly dependent on recruitment of SEC for their expression.

Activators Recruit Additional Factors

- It is possible that paused Pol II is also away of excluding inhibitory nucleosomes from the promoter region, rendering the promoter “poised” for rapid activation by upstream regulatory sequences.

Single integration & combinatorial control

- The *lac* genes of *E. coli* are efficiently expressed only when lactose is present and glucose absent.
- The two signals are communicated to the gene through separate regulators - one an activator and the other a repressor.

Single integration & combinatorial control

- In multicellular organisms signal integration is used extensively.
- In some cases, numerous signals are required to switch a gene on.

Single integration & combinatorial control

- But just as in bacteria, each signal is transmitted to the gene by a separate regulator, so at many genes multiple activators must work together to switch the gene on.

Single integration & combinatorial control

- When multiple activators work together, they do so synergistically.
- That is, the effect of, say, two activators working together is usually much greater than the sum of each of them working alone.

Single integration & combinatorial control

- Synergy can result from multiple activators recruiting a single component of the transcriptional machinery; multiple activators each recruiting a different component;

Single integration & combinatorial control

- or multiple activators helping each other bind to their sites upstream of the gene they control.
- Let us see all three strategies:-

Single integration & combinatorial control

- Two activators can recruit a single complex-for example, the Mediator-by touching different parts of it.
- The combined binding energy will have an exponential effect on recruitment.

Single integration & combinatorial control

- Synergy can also result from activators helping each other bind under conditions where the binding of one depends on binding of the other.

Single integration & combinatorial control

- This co-operativity can be seen in bacteria, whereby the two activators touch each other when they bind their sites on DNA.

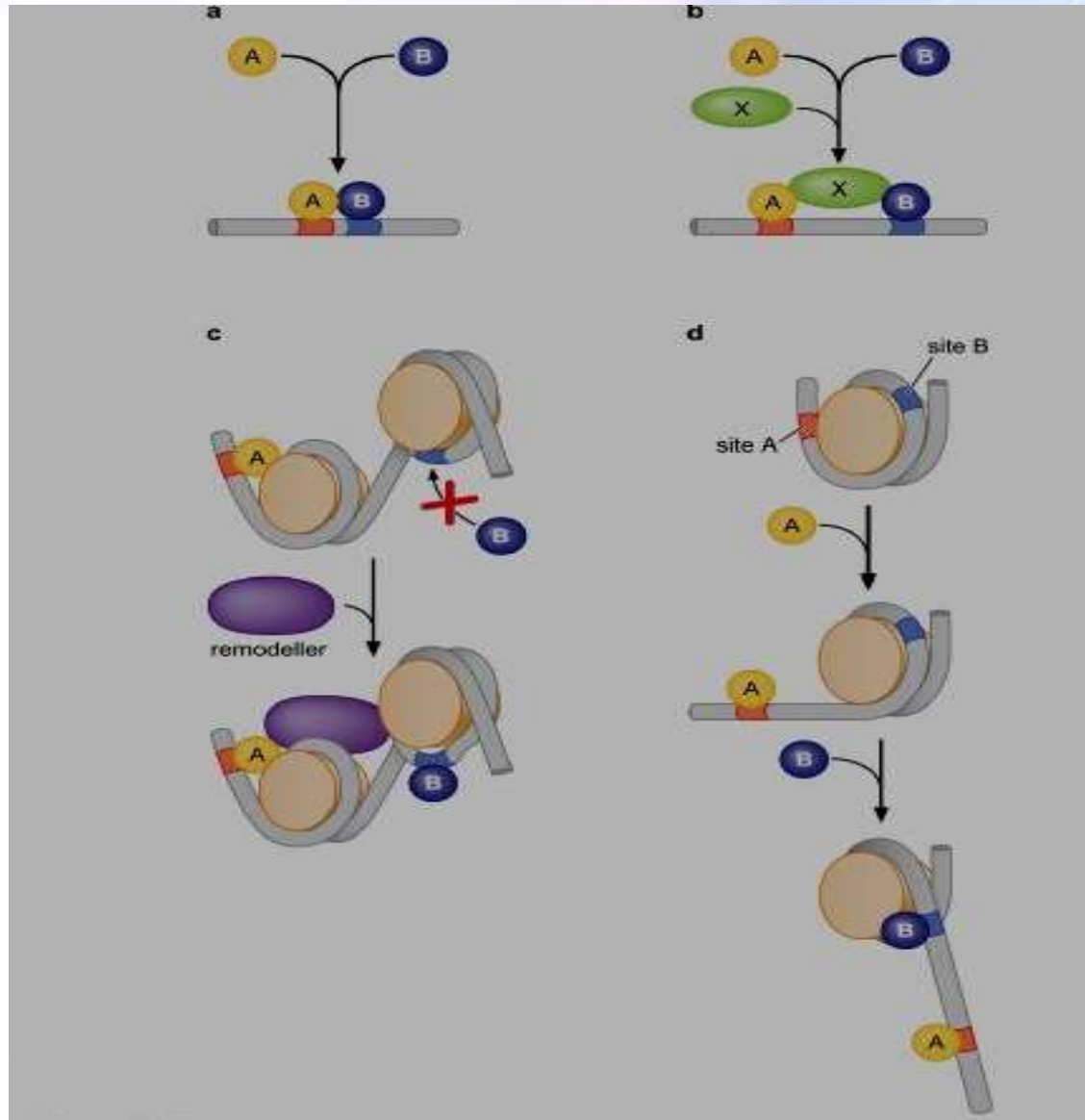
Single integration & combinatorial control

- But it can work in other ways as well: one activator can recruit something that helps the second activator bind.
- The different ways activators help each other bind DNA include "classical" cooperative binding;

Single integration & combinatorial control

- recruitment of a modifier by one activator to help a second bind; and binding of one activator to nucleosomal DNA uncovering the binding site for another.

Single integration & combinatorial control



Single integration & combinatorial control

- Synergy is critical for signal integration by activators.
- Consider a gene whose product is only needed when two signals are received.

Single integration & combinatorial control

- Each signal is communicated to the gene by a separate activator.
- The gene must be efficiently expressed when both activators are present but be relatively impervious to the action of either activator alone.

Single integration & combinatorial control

- Each signal is communicated to the gene by a separate activator.
- The gene must be efficiently expressed when both activators are present but be relatively impervious to the action of either activator alone.

Single integration & combinatorial control

- We have seen simple cases of combinatorial control in bacteria.
- For example, CAP is involved in regulating many genes, in collaboration with other regulators.

Single integration & combinatorial control

- At the *lac* genes it works with the Lac repressor; at the *gal* genes with the Gal repressor.
- There is also extensive combinatorial control in eukaryotes.

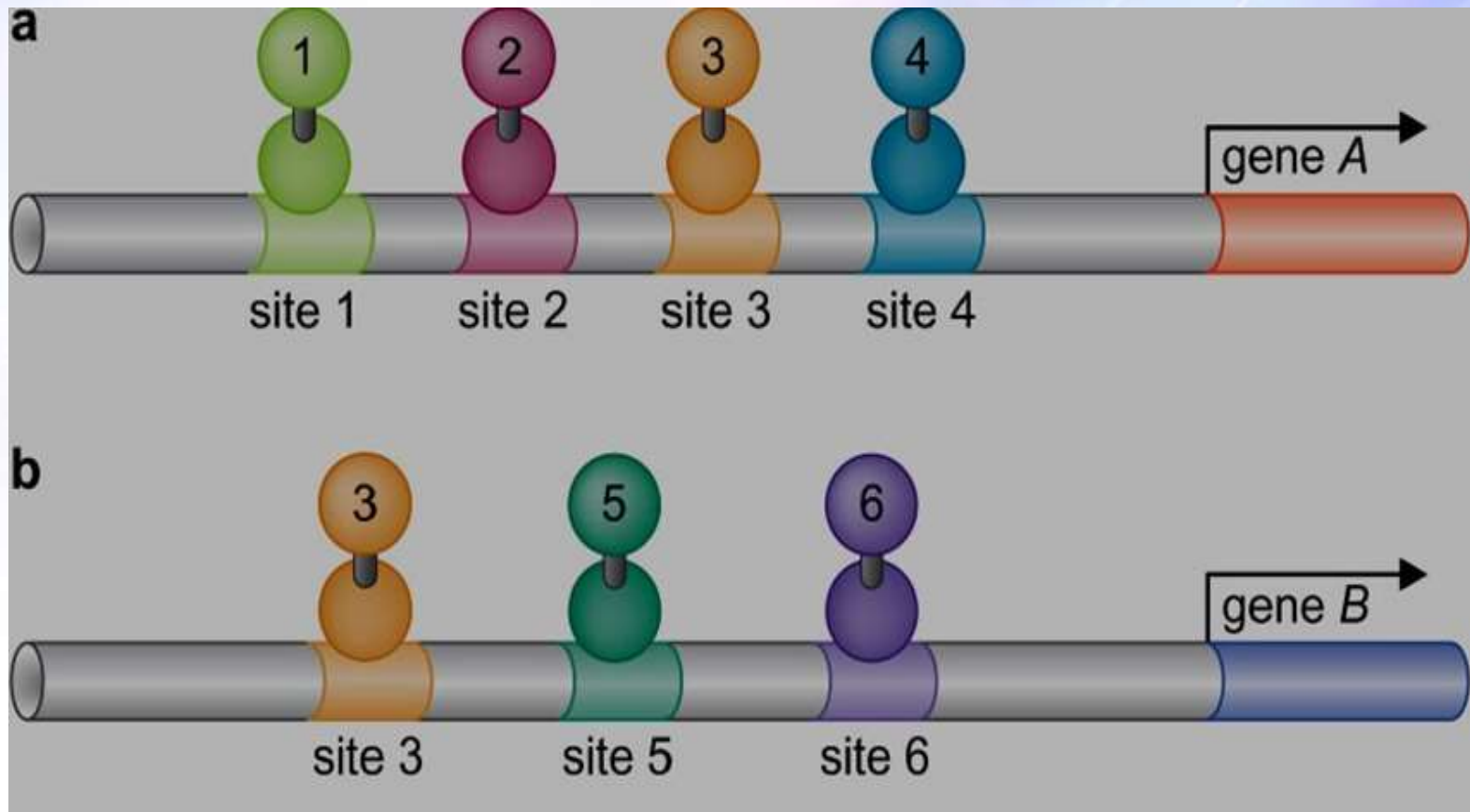
Single integration & combinatorial control

- Let us consider a generic case.
- Gene A is controlled by four signals (1, 2, 3, and 4), each working through a separate activator (activators 1, 2, 3, and 4).
- Gene B is controlled by three signals (3, 5, and 6), working through activators 3, 5, and 6.

Single integration & combinatorial control

- Note that there is one signal in common between these two cases, and the activator through which that signal works is the same at both genes.

Single integration & combinatorial control



Single integration & combinatorial control

- In complex multicellular organisms, such as *Drosophila* and humans, combinatorial control involves many more regulators and genes as shown in this example.

Single integration & combinatorial control

- How is it that the regulators can intermix so promiscuously?
- As you know, multiple activators work synergistically.

Single integration & combinatorial control

- In fact, even multiple copies of a single activator work synergistically, suggesting that a given activator can interact with multiple targets.

Single integration & combinatorial control

- This provides an explanation for why different regulators can work together in so many combinations: because each can use any of an array of targets, the combinations that work together are unrestricted.

Single integration & combinatorial control

- The examples of signal integration such as the *HO* gene in yeast and the human β -interferon gene- involve activators that also regulate other genes in examples of combinatorial control.

Transcriptional Repressors

- In bacteria, we saw that many repressors work by binding to sites that overlap the promoter and thus block binding of RNA polymerase.

Transcriptional Repressors

- But we also saw other ways they can work: they can bind to sites adjacent to promoters and, by interacting with polymerase bound there, inhibit the enzyme from initiating transcription.

Transcriptional Repressors

- They can also interfere with the action of activators.
- In eukaryotes, we see all of these except the first.
- We also see another form of repression, perhaps the most
- common in eukaryotes, that works as follows.

Transcriptional Repressors

- As with activators, repressors can recruit nucleosome modifiers, but in this case, the enzymes have effects opposite to those recruited by activators - they compact the chromatin or remove groups recognized by the transcriptional machinery.

Transcriptional Repressors

- Therefore, for example, histone deacetylases repress transcription by removing acetyl groups from the tails of histones in *S. cerevisiae*; as we have already seen, the presence of acetyl groups helps transcription.

Transcriptional Repressors

- Paradoxically, the histone deacetylase Rpd3 is also recruited to active genes to ensure transcription fidelity.

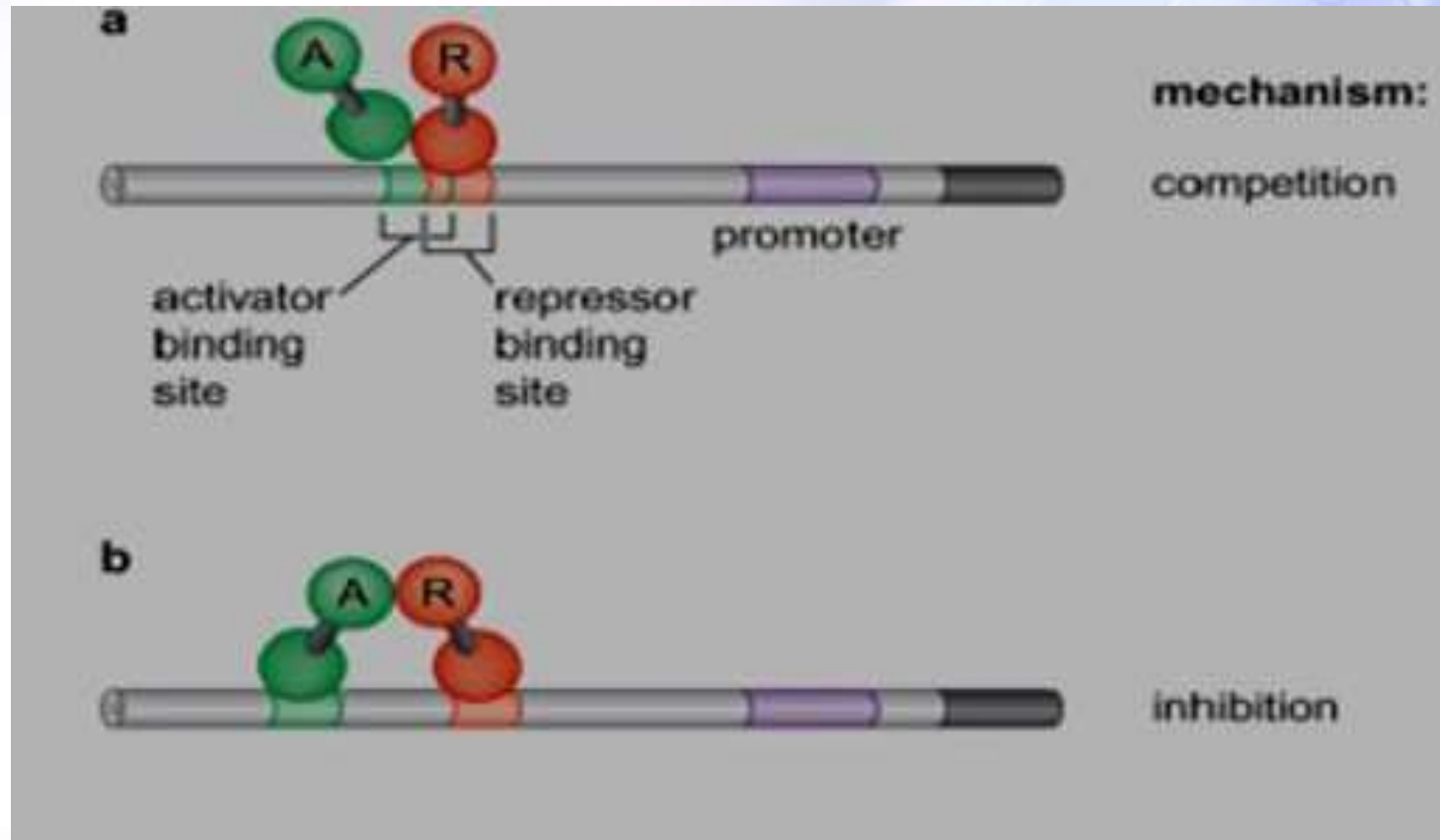
Transcriptional Repressors

- Nucleosomes are deacetylated behind elongating Pol II to prevent the use of “cryptic” promoters within the transcription unit.

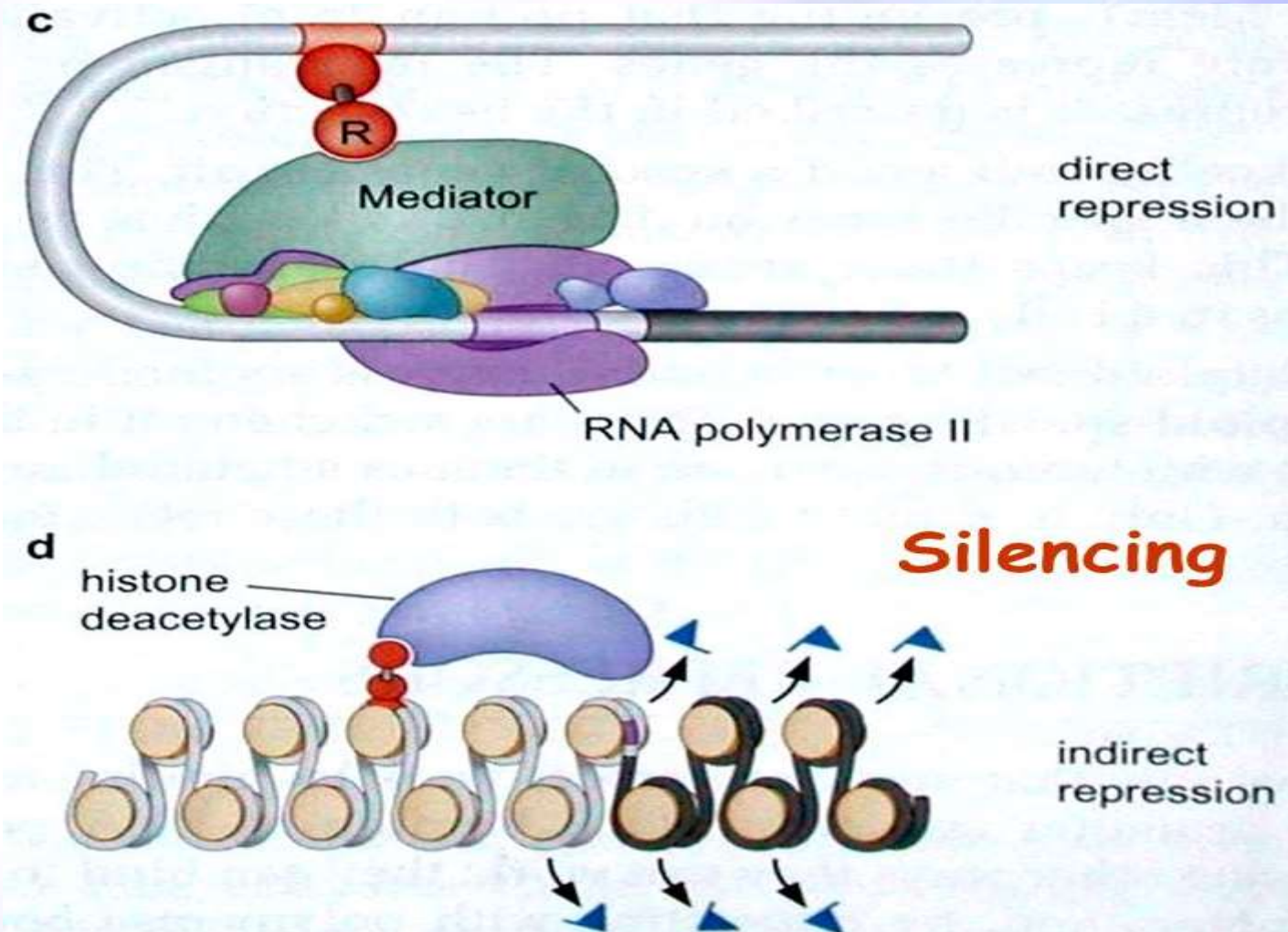
Transcriptional Repressors

- Other enzymes add methyl groups to histone tails, and this frequently represses transcription, although in some cases it is associated with an actively transcribed gene.

Transcriptional Repressors



Transcriptional Repressors



Transcriptional Repressors

- Histone (and DNA) modifications also form the basis of a type of repression called **silencing**.
- Let us consider one specific example, the repressor called Mig1, which, like Gal4, is involved in controlling the GAL genes of the yeast *S. cerevisiae*.

Transcriptional Repressors

- The *GAL* genes but with the addition of a site between the Gal4-binding sites and the promoter: this is where, in the presence of glucose, Mig1 binds and switches off the *GAL* genes.

Transcriptional Repressors

- Thus, just as in *E. coli*, the cell only makes the enzymes needed to metabolize galactose if the preferred energy source, glucose, is not present.
- How does Mig1 repress the *GAL* genes?

Transcriptional Repressors

- Mig1 recruits a “repressing complex” containing the Tup1 protein.
- This complex is recruited by many yeast DNA-binding proteins that repress transcription, including the $\alpha 2$ protein involved in controlling the mating-type specific genes.

Transcriptional Repressors

- Tup1 also has counterparts in mammalian cells.
- Two mechanisms have been proposed to explain the repressing effect of Tup1.

Transcriptional Repressors

- First, Tup1 acts on nucleosomes either through recruiting histone deacetylases and/or by positioning a nucleosome at or near the transcription start site.

Transcriptional Repressors

- Second, Tup1 interacts directly with the transcriptional machinery at the promoter and inhibits initiation.

Signal Transduction Pathways

- As we have seen, whether or not a given gene is expressed very often depends on environmental signals.
- Signals come in many forms:-

Signal Transduction Pathways

- They can be small molecules such as sugars as in bacteria, but they can also be proteins released by one cell and received by another as in multicellular organisms.

Signal Transduction Pathways

- There are various ways that signals are detected by a cell and communicated to a gene.
- In bacteria, we saw that signals control the activities of regulators by inducing allosteric changes in those regulators.

Signal Transduction Pathways

- Often, this effect is direct: a small molecular signal, such as a sugar, enters the cell and binds the transcriptional regulator directly.
- But we saw one example where the effect of the signal is indirect (control of the activator NtrC).

Signal Transduction Pathways

- In that case, the signal (low ammonia levels) induces a kinase that phosphorylates NtrC.
- This type of indirect signaling is an example of a **signal transduction pathway**.

Signal Transduction Pathways

- The term **signal** refers to the initiating ligand itself—the sugar or protein, for example.

Signal Transduction Pathways

- It can also refer to the “information” as it passes from detection of that ligand to the regulators that directly control the genes — that is, as it passes along a signal transduction pathway.

Signal Transduction Pathways

- In the simplest of bacterial cases, there was no distinction, of course, but once a signal transduction pathway is involved, there is.

Signal Transduction Pathways

- In addition, in eukaryotes, that most signals are communicated to genes through signal transduction pathways, sometimes very elaborate ones.

Signal Transduction Pathways

- In a signal transduction pathway, the initiating ligand is typically detected by a specific **cell surface receptor**: the ligand binds to an extracellular domain of the receptor, and this binding is communicated to the intracellular domain.

Signal Transduction Pathways

- From there, the signal is relayed to the relevant transcriptional regulator, often through a cascade of kinases.

Signal Transduction Pathways

- How is the binding of ligand to the extracellular domain communicated to the intracellular domain?

Signal Transduction Pathways

- This can be through an allosteric change in the receptor, whereby binding of ligand alters the shape (and thus activity) of the intracellular domain.

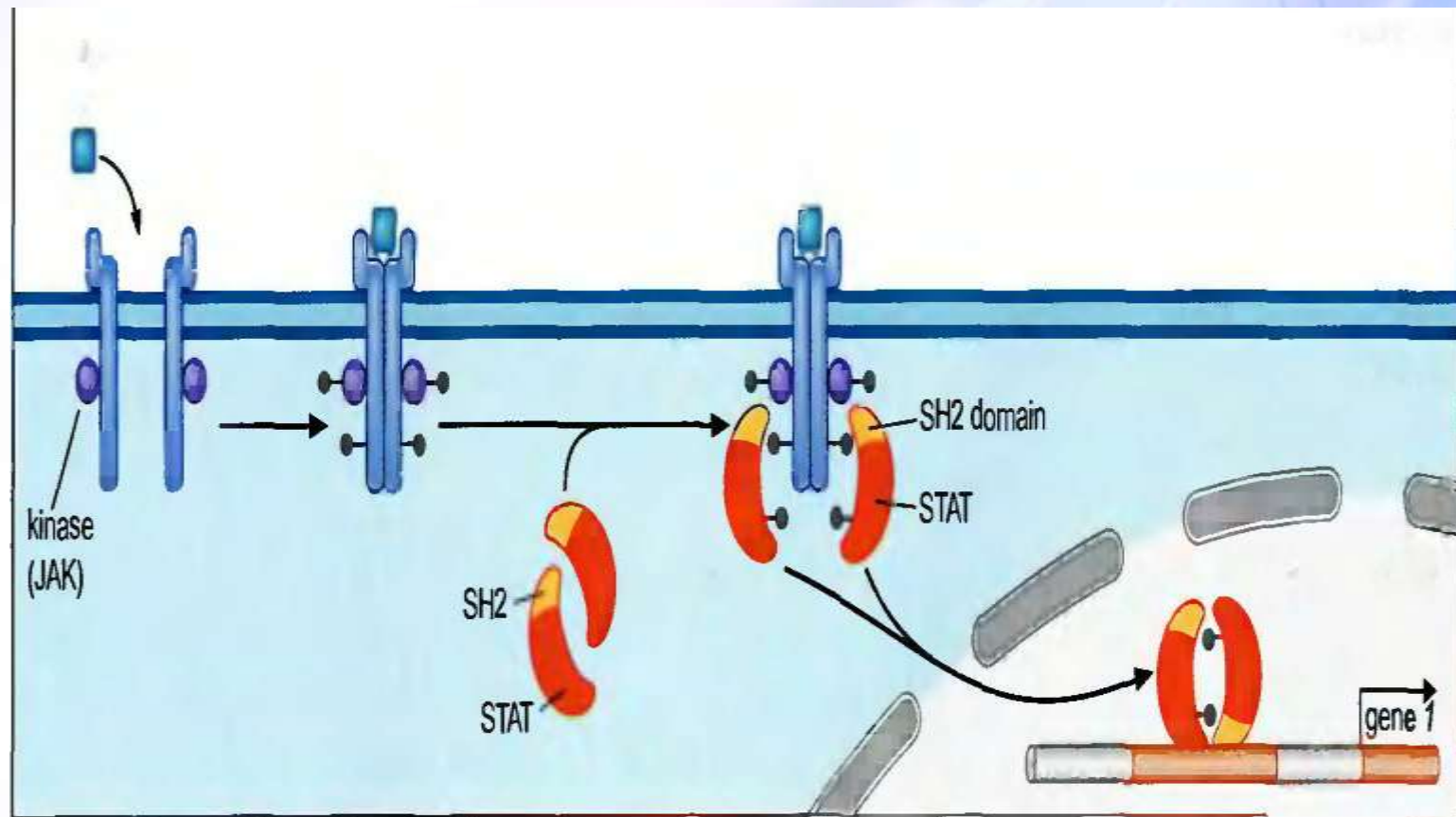
Signal Transduction Pathways

- Alternatively, the ligand can act simply to bring together two or more receptor chains, allowing interactions between the intracellular domains of those receptors to activate each other.

Signal Transduction Pathways

- The **STAT** (signal transducer and activator of transcription) pathway is a relatively simple example of signal transduction pathway.
- In this example, a kinase is bound to the intracellular domain of a receptor.

Signal Transduction Pathways



Signal Transduction Pathways

- When the receptor is activated by its ligand (a cytokine), it brings together two receptor chains and triggers the kinase in each chain to phosphorylate a particular sequence in the intracellular domain of the opposing receptor.

Signal Transduction Pathways

- This phosphorylated site is then recognized by a particular STAT protein that, once bound, gets phosphorylated itself.
- Once phosphorylated, the STAT dimerizes, moves to the nucleus, and binds DNA.

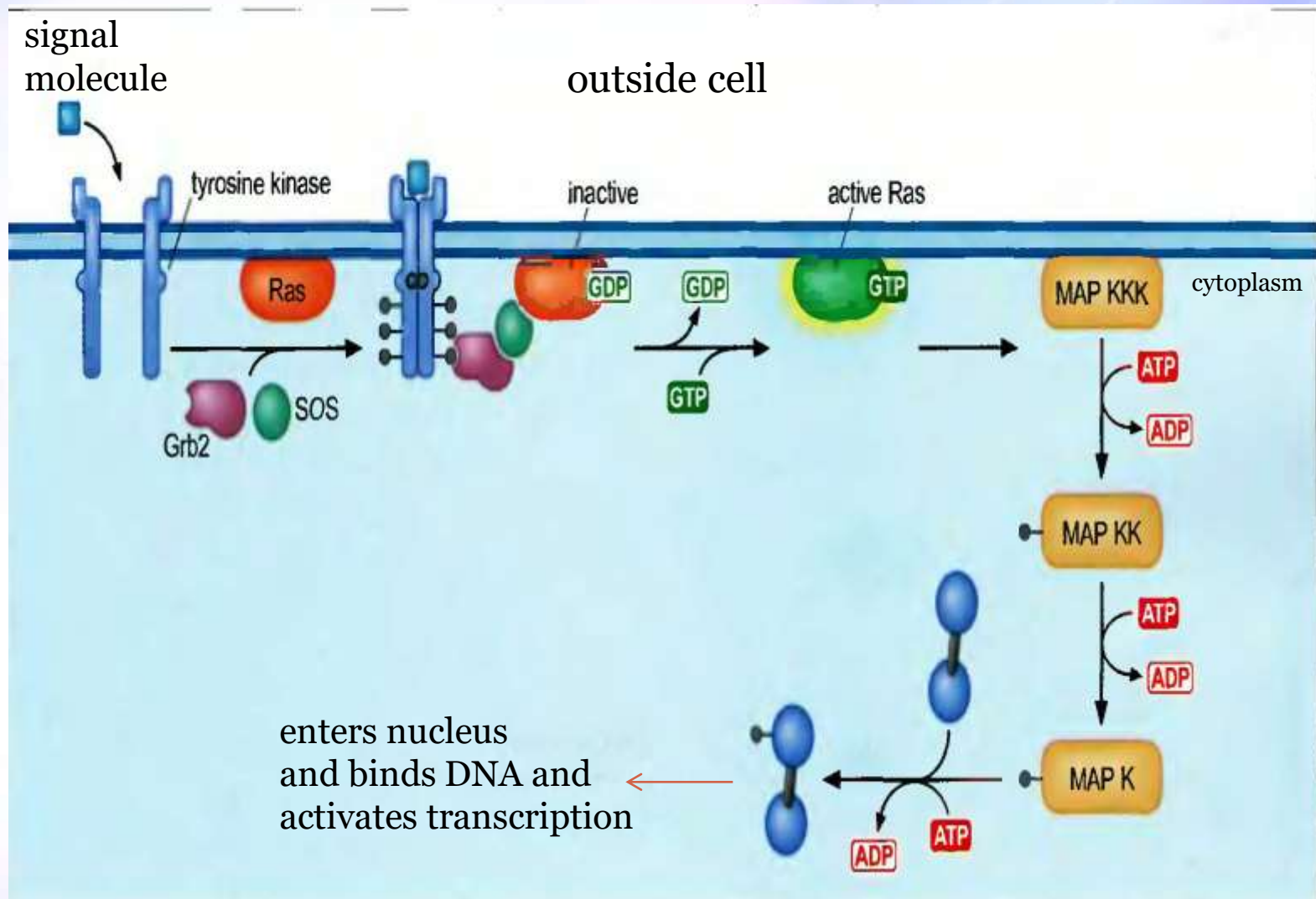
Signal Transduction Pathways

- Another example is the mitogen-activated protein kinase (MAPK) pathway that controls activators such as Jun, one of the activators that works at the interferon- β enhancer.

Signal Transduction Pathways

- In this case, the activated receptor induces a cascade of signaling events, ending in activation of an MAPK that phosphorylates Jun (and other transcriptional regulators).

Signal Transduction Pathways



Signal Transduction Pathways

- The most common way in which information is passed through signal transduction pathways is via phosphorylation, but proteolysis, dephosphorylation, and other modifications are also used.

Plasmids as Vectors

- **Plasmids** are autonomous self-replicating molecules of DNA (or very rarely RNA).
- They are not chromosomes, although they do reside inside living cells and carry genetic information.

Plasmids as Vectors

- They are not regarded as part of the cell's genome for two reasons.
- First, a particular plasmid may be found in cells of different species and may move from one host species to another.

Plasmids as Vectors

- Second, a plasmid may sometimes be present and sometimes absent from the cells of a particular host species.
- They are not needed for cell growth and division under normal conditions.

Plasmids as Vectors

- The DNA segment to be cloned is first linked to a **vector DNA**, which is a vehicle for carrying foreign DNA into a suitable host cell, such as the bacterium *E. coli*.
- Such carriers are called **vectors**.

Plasmids as Vectors

- Since the mid-1970s, many vectors have been developed which mainly fall into two major classes:-
 - Plasmids and
 - Phages.

Plasmids as Vectors

- Bacteriophages lambda and M 13 are commonly used vectors and have the advantage of being able to accept large pieces (up to 40 and 10 kb, respectively) of foreign DNA.

Plasmids as Vectors

- However, the most commonly used vectors for transferring foreign DNA into cells are bacterial plasmids, which accept smaller pieces (5-10 kb) of foreign DNA.

Plasmids as Vectors

- A typical plasmid vector is a relatively small circular DNA molecule that contains its own origin of replication (*ori*) so it can replicate inside a host cell independent of the host chromosome.

Plasmids as Vectors

- Although plasmids occur naturally in bacteria, where they usually confer useful properties on the cell such as antibiotic resistance,

Plasmids as Vectors

- yet, most plasmid vectors in use today have been specially modified or constructed for the purpose of transferring foreign DNA into cells.

Plasmids as Vectors

- Some, such as *pUC* 18 and the Bluescript vectors, have been engineered to make many hundreds of copies of themselves, and any foreign DNA they contain, once they are introduced into a host cell.

Plasmids as Vectors

- This process of producing a large quantity of identical copies of any chosen DNA is referred to as **cloning**, and the cells derived from a host cell containing a recombinant plasmid is referred to as a **cell clone**.

Plasmids as Vectors

- For cloning to work, the vector must have a cutting site for a restriction enzyme that makes it possible to insert the desired fragment of DNA and, second, the vector must be selectable so that cells that have taken up a vector can be identified.

Plasmids as Vectors

- To overcome the difficulty of inserting fragments into vectors, most plasmids have been constructed with polylinkers (multiple cloning sites)-pieces of DNA that contain one or more unique restriction enzyme recognition sites.

Plasmids as Vectors

- Moreover, most plasmids contain antibiotic-resistance genes or other **selectable Markers** that enable cells containing plasmids to be isolated from cells that have not taken up a plasmid.

Plasmids as Vectors

- In the early years of the cloning era, Boyer and his colleagues developed a set of very popular vectors known as the pBR plasmid series.

Plasmids as Vectors

- Nowadays, one can choose from many plasmid cloning vectors besides the pBR plasmids.
- One useful, though somewhat dated, class of plasmids is the *pUC series*.

Plasmids as Vectors

- Nowadays, one can choose from many plasmid cloning vectors besides the pBR plasmids.
- One useful, though somewhat dated, class of plasmids is the *pUC series*.

Plasmids as Vectors

- One of the most notable plasmids, termed pBR322 after its developers Bolivar and Rodriguez (pBR), was widely adopted and illustrates its desirable features of a cloning vector as indicated below;-

Plasmids as Vectors

- The plasmid is much smaller than a natural plasmid which makes it more resistant to damage by shearing and increases the efficiency of uptake by bacteria, a process termed transformation.

Plasmids as Vectors

- A bacterial origin of DNA replication ensures that the plasmid will be replicated by the host cell.
- Most plasmids, including pBR322, have a relaxed origin of replication and which means replication is not tightly linked to cell division.

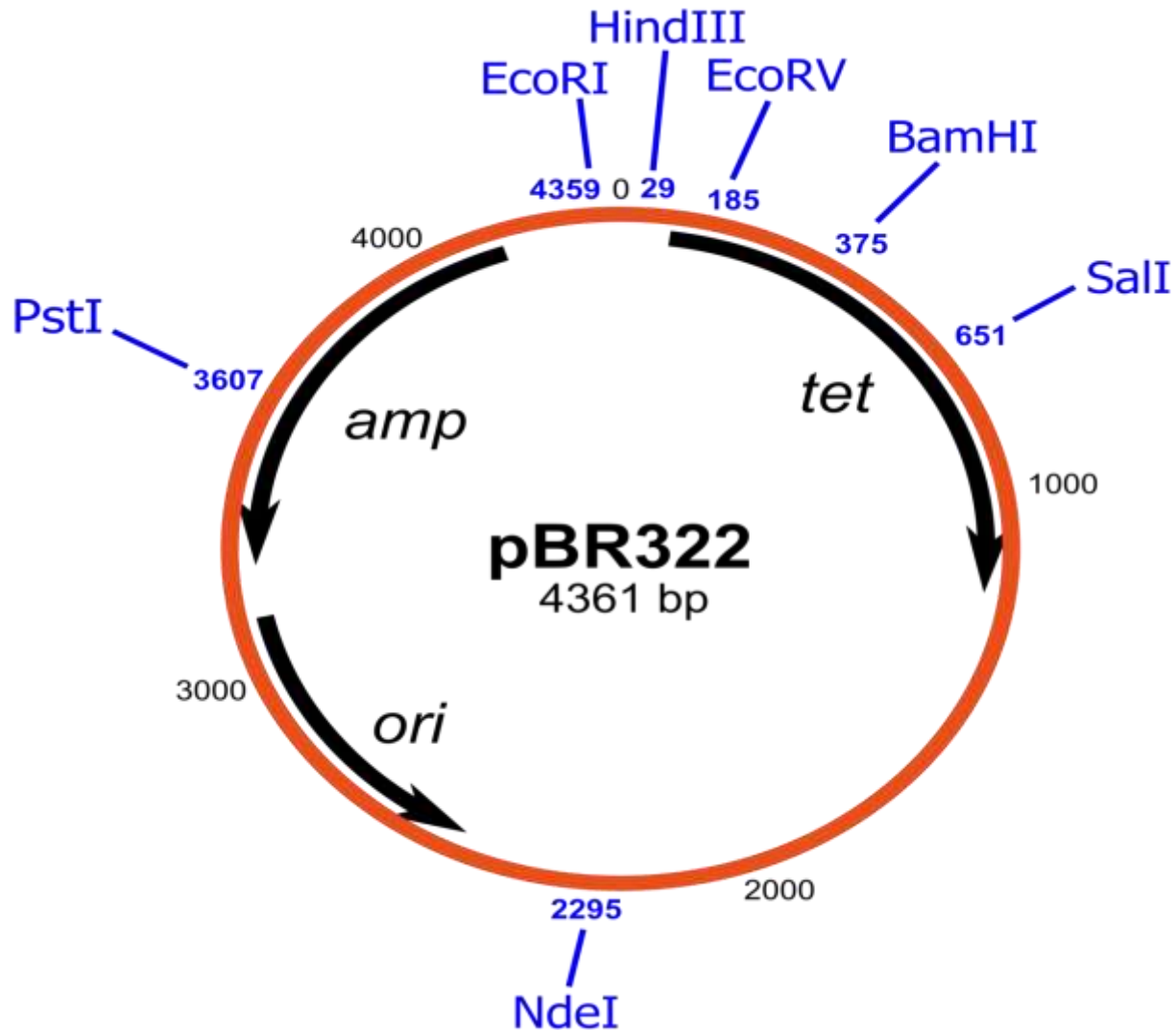
Plasmids as Vectors

- The genes coding for resistance to antibiotics have been introduced which allow the selection of cells which contain plasmid.

Plasmids as Vectors

- There are single recognition sites for a number of restriction enzymes at various points around the plasmid which can be used to open or linearize the circular plasmid.

Plasmids as Vectors



Plasmid Selection Systems

- Insertional inactivation is a useful selection method for identifying recombinant vectors with inserts. For example, a fragment of chromosomal DNA digested with *Bam*H1 would be isolated and purified.

Plasmid Selection Systems

- The plasmid pBR322 would also be digested at a single site, using *Bam*H1, after which both samples would be deproteinized to inactivate the restriction enzyme.

Plasmid Selection Systems

- BamH 1 cleaves to give sticky ends, and so it is possible to obtain ligation between the plasmid and digested DNA fragments in the presence of T4 DNA ligase.

Plasmid Selection Systems

- The products of this ligation will include plasmid containing a single fragment of the DNA as an insert, but there will also be unwanted products,

Plasmid Selection Systems

- such as plasmid which has recircularized without an insert, dimers of plasmid, fragments joined to each other, and plasmid with an insert composed of more than one fragment.

Plasmid Selection Systems

- Most of these unwanted molecules can be eliminated during subsequent steps. The products of such reactions are usually identified by agarose gel electrophoresis.

Plasmid Selection Systems

- The ligated DNA must now be used to transform *E. coli*. Bacteria do not normally take up DNA from their surroundings, but can be induced to do so by prior treatment with Ca^{2+} at 4°C.

Plasmid Selection Systems

- These cells are then termed **competent**, since DNA added to the suspension of cells will be taken up during a brief increase in temperature, termed a **heat shock**.
- Plasmids can also be introduced into the bacterial cells by electroporation.

Plasmid Selection Systems

- In this process the cells are subjected to pulses of a high voltage gradient, causing many of them to take up DNA from the surrounding solution.
- This technique has proved to be very useful.

Plasmid Selection Systems

- After a brief incubation to allow expression of the antibiotic resistance genes the cells are plated onto medium containing the antibiotic, e.g. Ampicillin.

Plasmid Selection Systems

- Colonies which grow on these plates must be derived from cells which contain plasmid, since this carries the gene for resistance to ampicillin.
- Next the colonies are replica plated, using a sterile velvet pad, onto plates containing tetracycline in their medium.

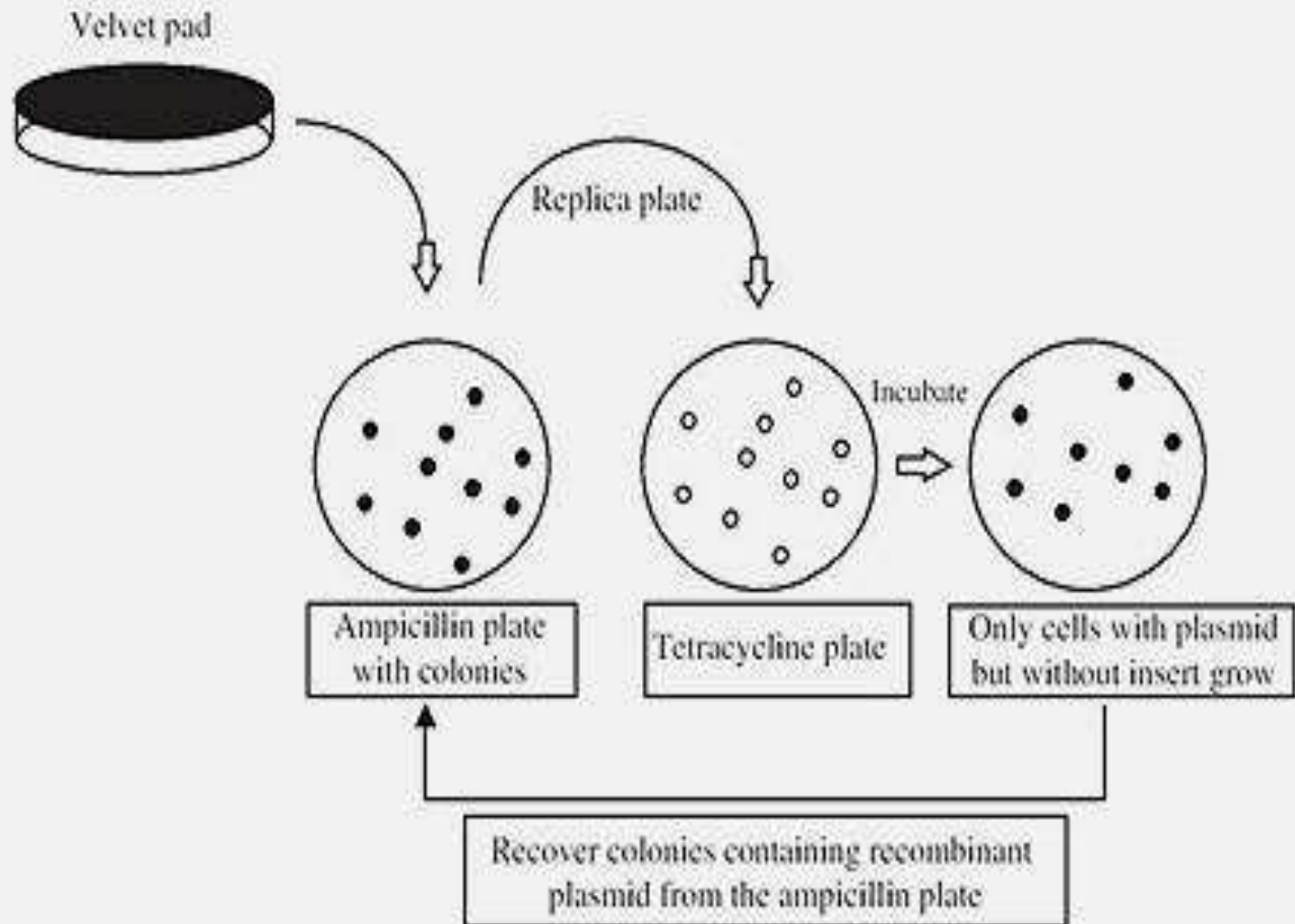
Plasmid Selection Systems

- Since the *Bam*H1 site lies within the tetracycline resistance gene, this gene will be inactivated by the presence of insert, but will be intact in those plasmids which have merely recircularized.

Plasmid Selection Systems

- Thus colonies which grow on ampicillin but not on tetracycline must contain plasmids with inserts.
- This illustrates the importance of a second gene for antibiotic resistance in a vector.

Plasmid Selection Systems



Plasmid Selection Systems

- If the digested plasmid is treated with the enzyme alkaline phosphatase prior to ligation, recircularization will be prevented, since this enzyme removes the 5' phosphate groups which are essential for ligation.

END

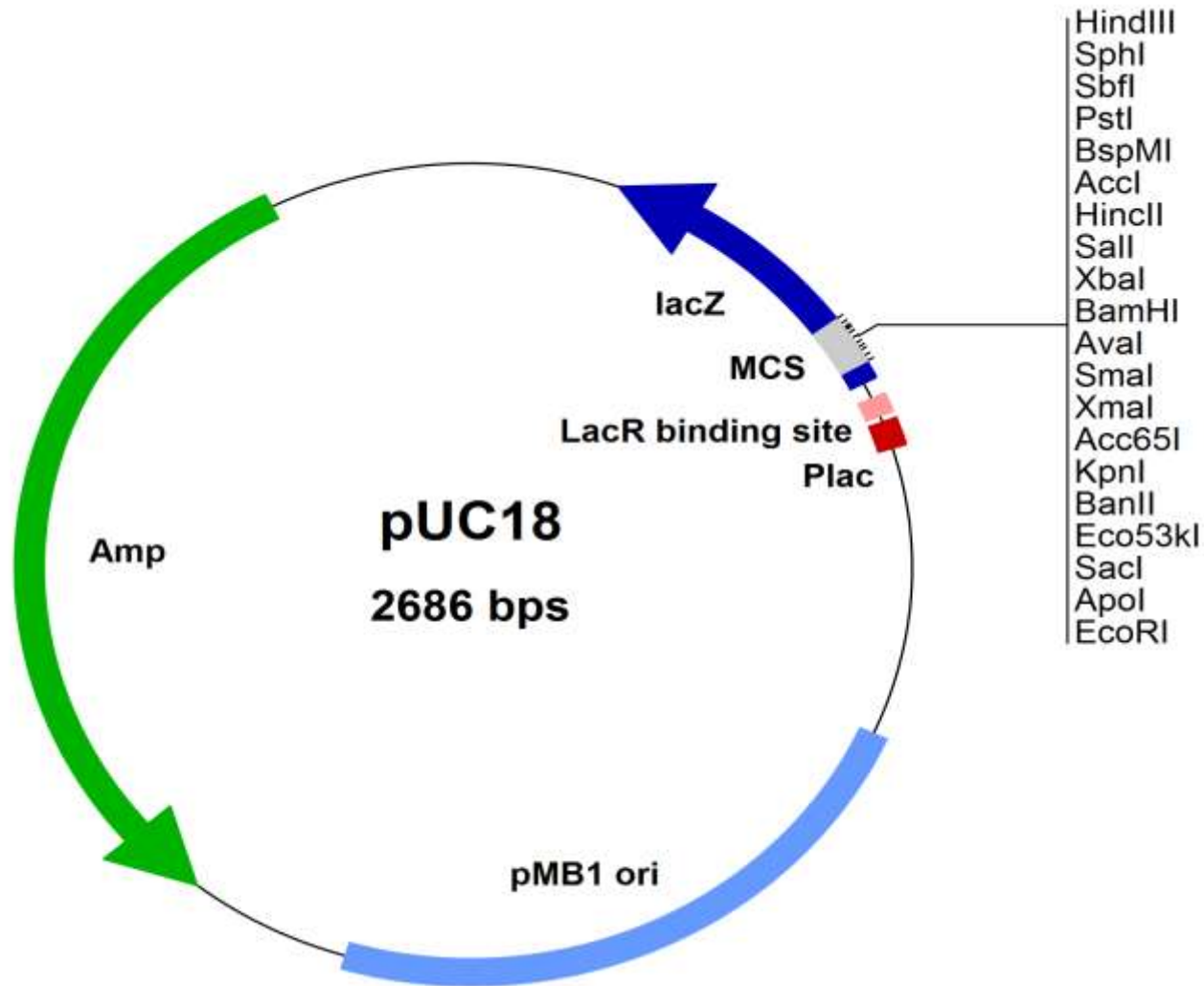
pUC Plasmid Cloning Vectors

- The valuable features of pBR322 have been enhanced by the construction of a series of plasmids termed *pUC* (produced at the University of California)

***pUC* Plasmid Cloning Vectors**

- These plasmids are based on pBR322, from which about 40% of the DNA has been deleted.

pUC Plasmid Cloning Vectors



pUC Plasmid Cloning Vectors

- There is an antibiotic resistance gene for ampicillin and origin of replication for *E. coli*.
- *coli*.
- In addition the most popular restriction sites are concentrated into a region termed the **multiple cloning site** (MCS).

pUC Plasmid Cloning Vectors

- Moreover, the MCS is part of a gene (*LacZ*) in its own right and codes for a portion of a polypeptide called β -galactosidase.

pUC Plasmid Cloning Vectors

- When the *pUC* plasmid has been used to transform the host cell *E. coli* the gene may be switched on by adding the inducer IPTG (isopropyl- β -D-thiogalactopyranoside)

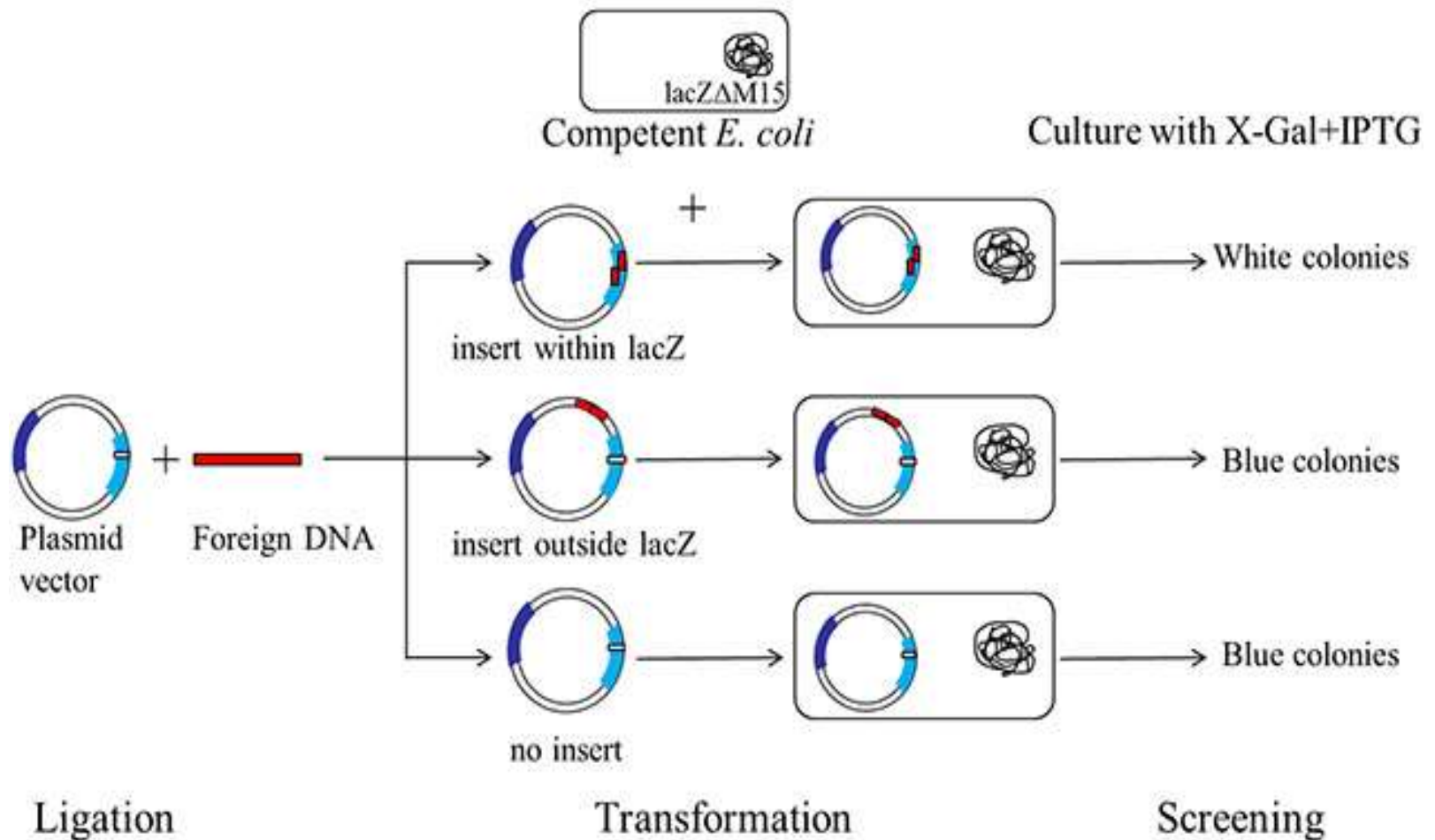
***pUC* Plasmid Cloning Vectors**

- Its presence causes the enzyme β -galactosidase to be produced.
- The functional enzyme is able to hydrolyse a colourless substance called X-gal into a blue insoluble material.

***pUC* Plasmid Cloning Vectors**

- However if the gene is disrupted by the insertion of a foreign fragment of DNA, a nonfunctional enzyme results which is unable to carry out hydrolysis of Xgal.

pUC Plasmid Cloning Vectors



pUC Plasmid Cloning Vectors

- Thus, a recombinant *pUC* plasmid may be easily detected since it is white or colourless in the presence of X-gal, whereas an intact non-recombinant *pUC* plasmid will be blue as its gene is fully functional and not disrupted.

pUC Plasmid Cloning Vectors

- This elegant system, termed **blue/white selection**, allows the initial identification of recombinants to be undertaken very quickly and has been included in a number of subsequent vector systems.

pUC Plasmid Cloning Vectors

- This selection method and insertional inactivation of antibiotic resistance genes do not, however, provide any information on the character of the DNA insert, only the status of the vector.

END

Virus-based Cloning Vectors

- A useful feature of any cloning vector is the amount of DNA it may accept or may be inserted before it becomes unviable.

Virus-based Cloning Vectors

- Inserts greater than 5 kb in size increase plasmid size to the point at which efficient transformation of bacterial cells decreases markedly.

Virus-based Cloning Vectors

- Thus, bacteriophages (bacterial viruses) have been adapted as vectors in order to propagate larger fragments of DNA in bacterial cells.

Virus-based Cloning Vectors

- Cloning vectors derived from one bacteriophage are commonly used since they offer approximately 1 6-fold advantage in cloning efficiency in comparison with the most efficient plasmid cloning vectors.

Virus-based Cloning Vectors

- Bacteriophage λ is a linear double-stranded phage approximately 49 kb in length. It infects *E. coli* with great efficiency by injecting its DNA through the cell membrane.

Virus-based Cloning Vectors

- In the wild-type phage λ the DNA follows one of two possible modes of replication.
- Firstly, it may enter the lysogenic life cycle. Alternatively, it may follow a lytic life cycle.

Virus-based Cloning Vectors

- At the extreme ends of phage DNA, are 12 bp sequences termed **cos (cohesive) sites**. Although they are asymmetrical, they are similar to restriction sites and allow the phage DNA to be circularized.

Virus-based Cloning Vectors

- Phage may be replicated very efficiently in this way, resulting in concatemers of many phage genomes which are cleaved at the cos sites and inserted into newly formed phage protein heads.

Virus-based Cloning Vectors

- Phage λ has been used extensively in the production of gene libraries, mainly because of its efficient entry into the *E. coli* cell and the fact that larger fragments of DNA may be stably integrated.

Virus-based Cloning Vectors

- For the cloning of long DNA fragments, up to approximately 25 kb, much of the nonessential λ DNA that codes for the lysogenic life cycle is removed and replaced by the foreign DNA insert.

Virus-based Cloning Vectors

- The recombinant phage is then assembled into pre-formed viral protein particles, a process known as **in vitro packaging**.
- These newly formed phages are used to infect bacterial cells which have been plated out on agar.

Virus-based Cloning Vectors

- Once inside the host cells, the recombinant viral DNA is replicated.
- All the genes needed for normal lytic growth are still present in the phage DNA.

Virus-based Cloning Vectors

- So multiplication of the virus takes place by cycles of cell lysis and infection of surrounding cells, giving rise to plaques of lysed cells on a background, or lawn, of bacterial cells.

Virus-based Cloning Vectors

- The viral DNA including the cloned foreign DNA can be recovered from the viruses from these plaques and analysed further by restriction mapping and agarose gel electrophoresis.

END

λ Phage Vectors

- Fred Blattner and his colleagues constructed the first phage vectors by modifying the well-known λ phage.

λ Phage Vectors

- They took out the region in the middle of the phage DNA, but retained the genes needed for phage replication. The missing phage genes could then be replaced with foreign DNA.

λ Phage Vectors

- Blattner named these vectors **Charon phages** after Charon, the boatman on the river Styx in classical mythology.
- In general, two types of λ phage vectors have been developed, **λ insertion vectors** and **λ replacement vectors**.

λ Phage Vectors

- One clear advantage of the λ phages over plasmid vectors is that they can accommodate much more foreign DNA.
- For example, Charon 4 can accept up to about 20 kb of DNA.

λ Phage Vectors

- λ insertion vectors accept less DNA than the replacement type since the foreign DNA is simply inserted into a region of the phage genome with appropriate restriction sites; common examples are λ gt10 and λ charon16A.

λ Phage Vectors

- In a replacement vector, a central region of DNA not essential for lytic growth is removed (a stuffer fragment) by a double digestion with, for example, *EcoRI* and *BamHI*.

λ Phage Vectors

- This leaves two DNA fragments, called right and left arms.
- The central stuffer fragment is replaced by inserting foreign DNA between the arms to form a functional recombinant λ phage.

λ Phage Vectors

- The most notable examples of λ replacement vectors are λ EMBL and λ ZAP.
- λ ZAP is a commercially produced cloning vector which includes unique cloning sites clustered into a multiple cloning site (MCS).

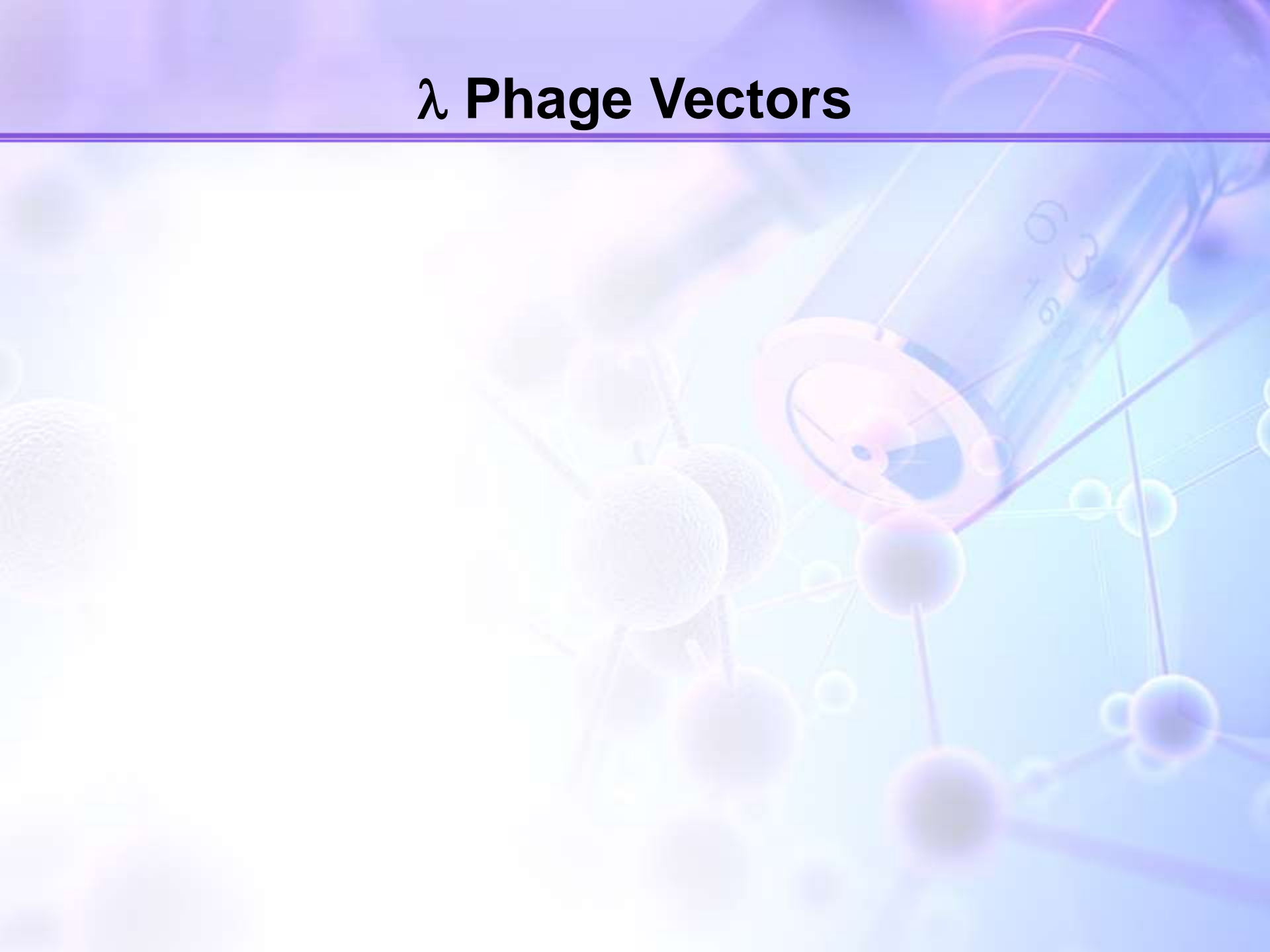
λ Phage Vectors

- Furthermore, the MCS is located within a *lacZ* region providing a blue/white screening system based on insertional inactivation.
- It is also possible to express foreign cloned DNA from this vector.

λ Phage Vectors

- This is a very useful feature of some λ vectors since it is then possible to screen for protein product rather than the DNA inserted into the vector.
- This screening is undertaken with antibody probes directed against the protein of interest.

λ Phage Vectors



λ Phage Vectors

- Another feature that make this a useful cloning vector is the ability to produce RNA transcripts, termed cRNA or **riboprobes**.

λ Phage Vectors

- This is possible because two promoters for RNA polymerase enzymes exist in the vector, a T7 and a T3 promoter, which flank the MCS.

λ Phage Vectors

- One of the most useful features of λ ZAP is that it has been designed to allow automatic in vivo excision of a small 2.9 kb colony-producing vector, the phagemid pBluescript SK.

λ Phage Vectors

- This technique is sometimes called single-stranded DNA rescue and occurs as the result of a ‘superinfection’ process where helper phages are added to the cells which are grown for an additional period of approximately 4 hours.

λ Phage Vectors

- A common use for λ replacement vectors is in constructing genomic libraries.
- Suppose we wanted to clone the entire human genome.

λ Phage Vectors

- A common use for λ replacement vectors is in constructing genomic libraries.
- Suppose we wanted to clone the entire human genome.
- This would obviously require a great many clones.

λ Phage Vectors

- So the larger the insert in each clone, the fewer total clones would be needed.
- In fact, such genomic libraries have been constructed for the human genome and for genomes of a variety of other organisms.

END

M13 Phage Vectors

- Another phage used as a cloning vector is the filamentous (long, thin, filament-like) phage M13.
- M13 is a filamentous coliphage with a genome composed of single-stranded circular DNA.

M13 Phage Vectors

M13 Bacteriophage



M13 Phage Vectors

- Joachim Messing and his co-workers endowed the phage DNA with the same β -galactosidase gene fragment and multiple cloning sites found in the *pUC* family of vectors.

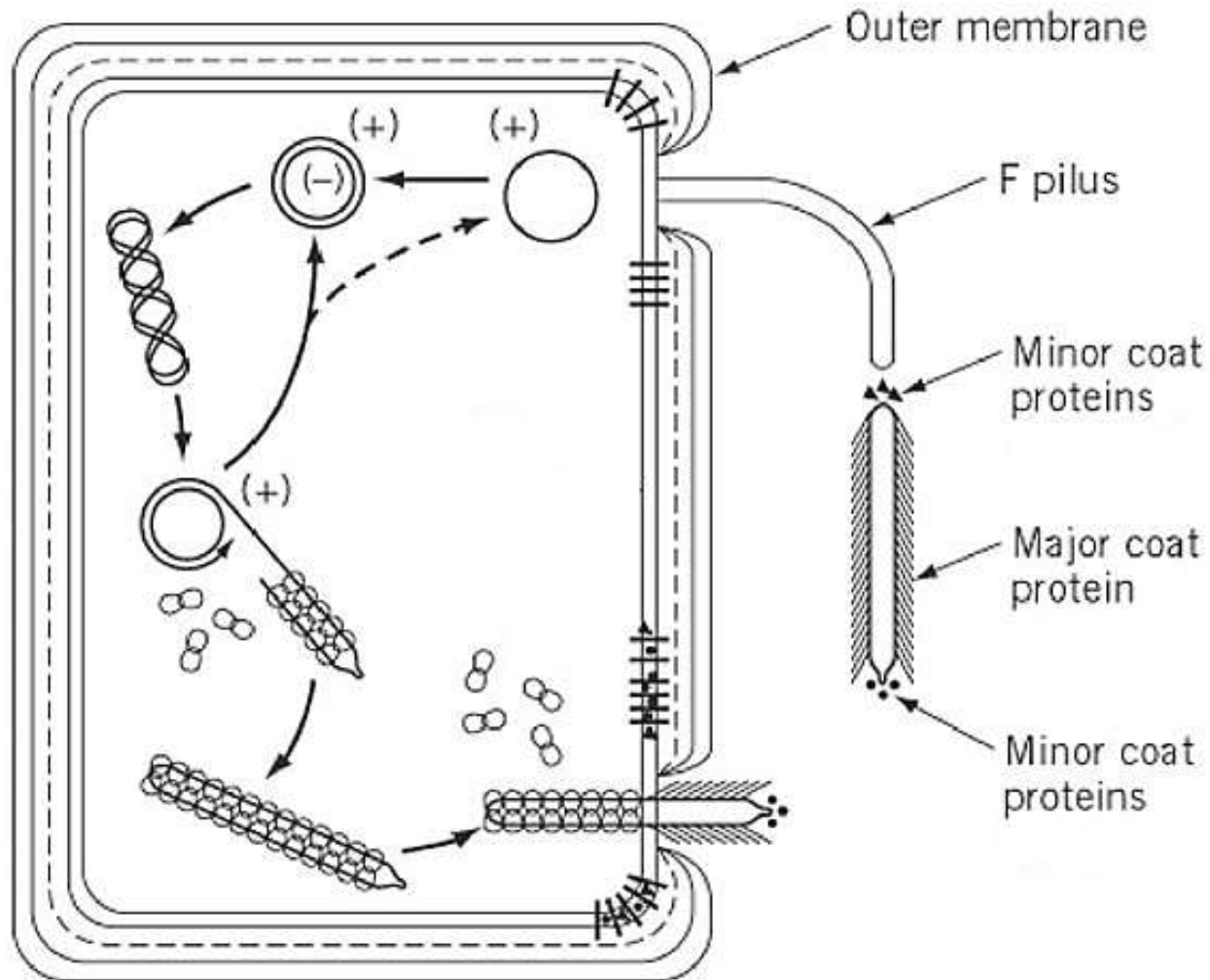
M13 Phage Vectors

- In fact, the M13 vectors were engineered first; then the useful cloning sites were simply transferred to the *pUC* plasmids.

M13 Phage Vectors

- The DNA in the phage particle itself is single-stranded, but after infecting an *E. coli* cell, the DNA is converted to a double-stranded replicative form (RF).
- This double-stranded replicative form of the phage DNA is used for cloning.

M13 Phage Vectors



M13 Phage Vectors

- After it is cut with one or two restriction enzymes at its multiple cloning site, foreign DNA with compatible ends can be inserted.

M13 Phage Vectors

- This recombinant DNA is then used to transform host cells, giving rise to progeny phages that bear single-stranded recombinant DNA.

M13 Phage Vectors

- The phage particles, containing phage DNA, are secreted from the transformed cells and can be collected from the growth medium.

M13 Phage Vectors

- The main factor is that the genome of this phage is a single-stranded DNA, so DNA fragments cloned into this vector can be recovered in single-stranded form.

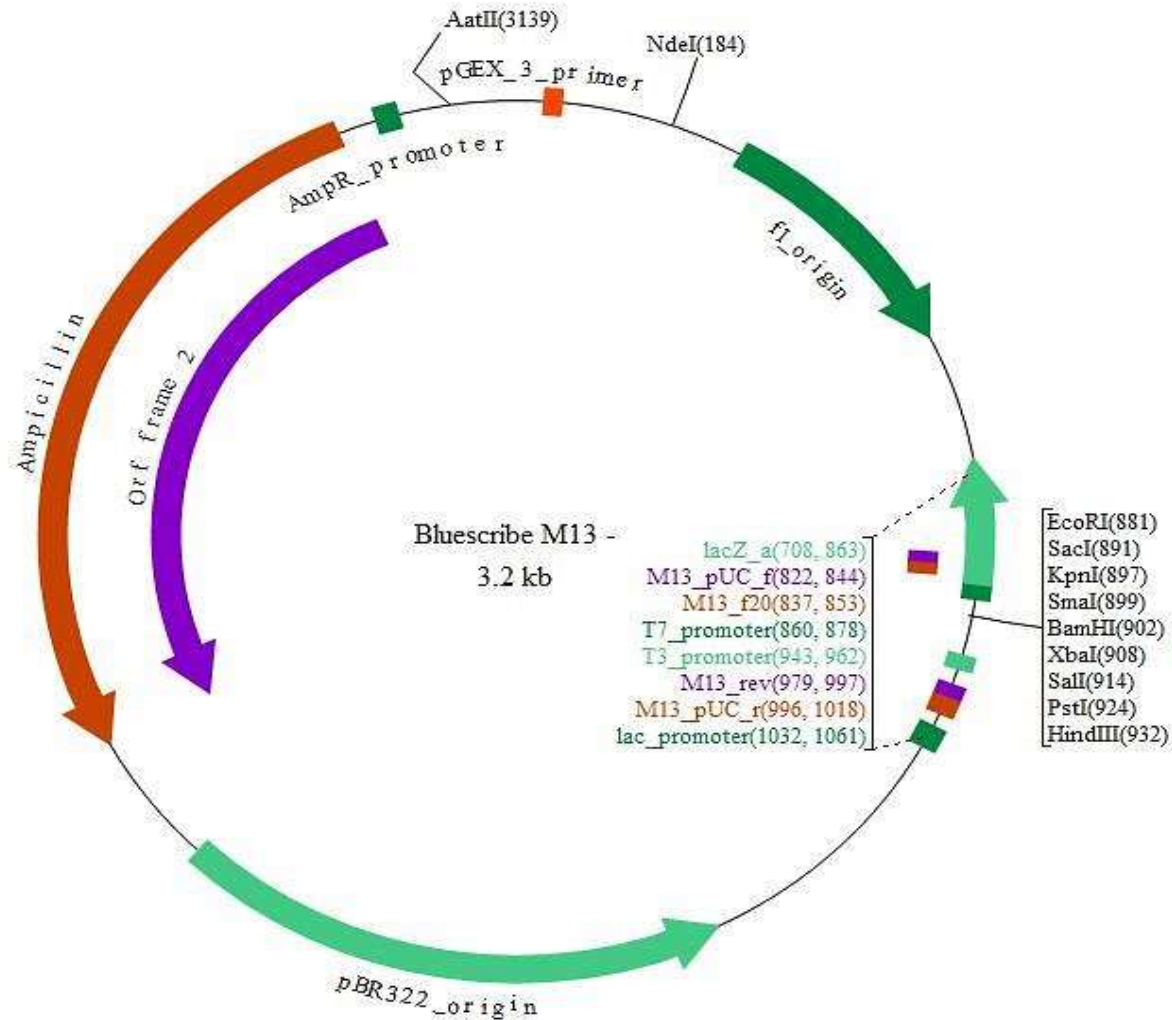
M13 Phage Vectors

- Single-stranded DNA can be an aid to site-directed mutagenesis, by which we can introduce specific, premeditated alterations into a gene.

M13 Phage Vectors

- The nature of these vectors makes them ideal for techniques such as chain termination sequencing and *in vitro* mutagenesis, since both methods require single-stranded DNA.

M13 Phage Vectors



Cosmids & Phagemids

- Another vector designed especially for cloning large DNA fragments is called a cosmid.
- Cosmids behave both as plasmids and as phages.

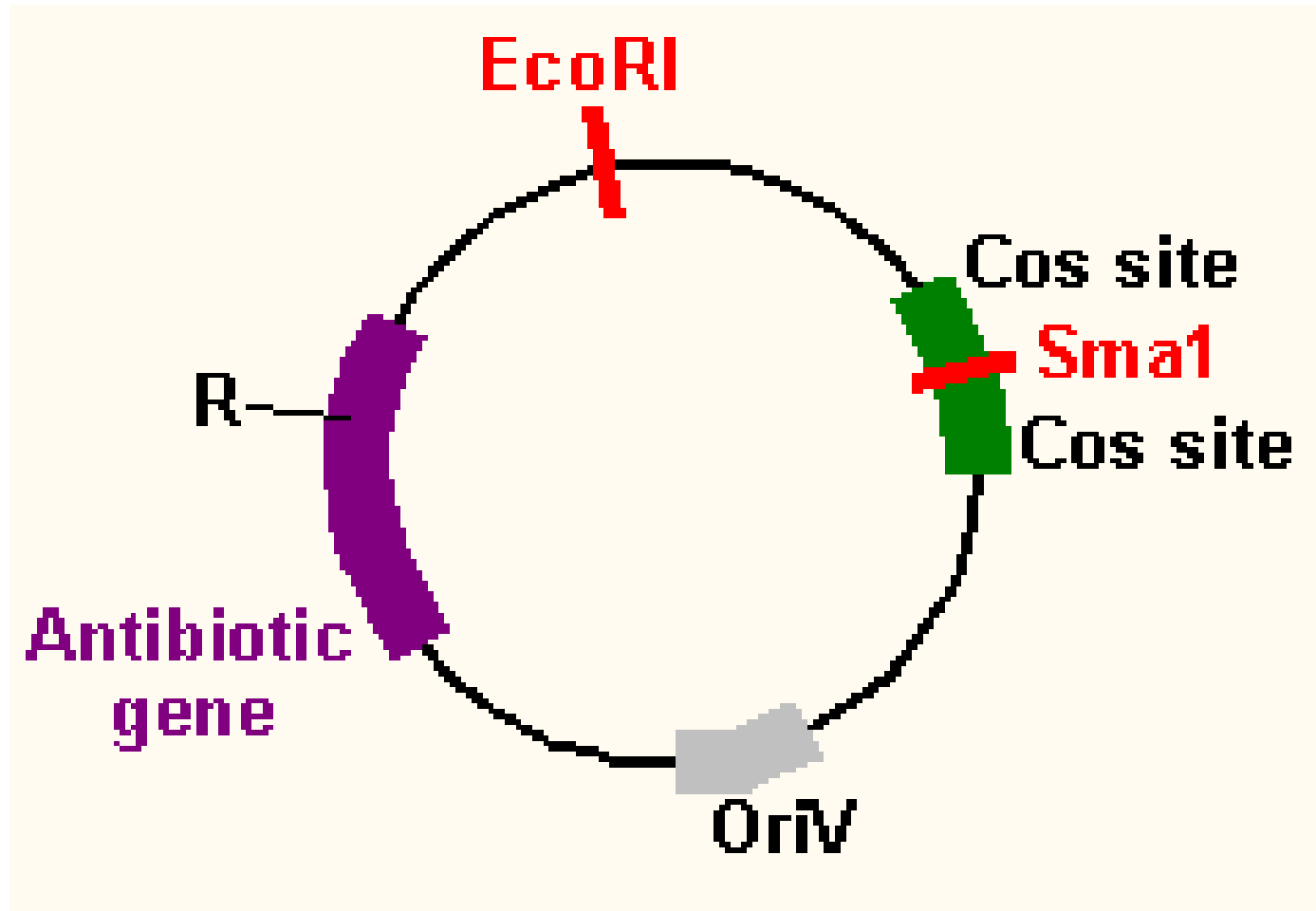
Cosmids & Phagemids

- They contain the cos sites, or cohesive ends, of λ phage DNA, which allow the DNA to be packaged into λ phage heads (hence the “cos” part of the name “cosmid”).

Cosmids & Phagemids

- They also contain a plasmid origin of replication, so they can replicate as plasmids in bacteria (hence the “mid” part of the name).

Cosmids & Phagemids



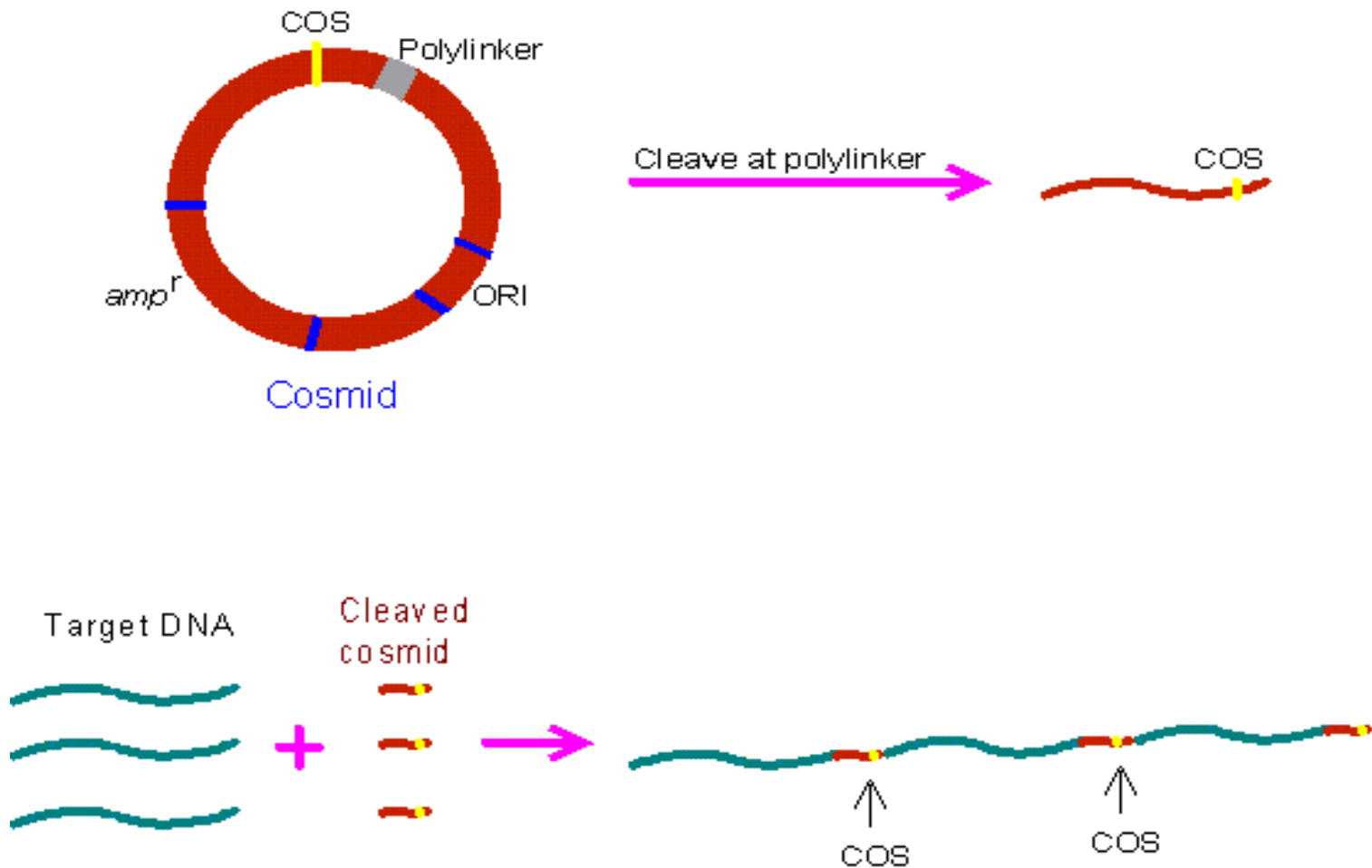
Cosmids & Phagemids

- Because almost the entire λ genome, except for the cos sites, has been removed from the cosmids, they have room for large inserts (40–50 kb).

Cosmids & Phagemids

- Once these inserts are in place, the recombinant cosmids are packaged into phage particles in vitro.

Cosmids & Phagemids



Cosmids & Phagemids

- These particles cannot replicate as phages because they have almost no phage DNA, but they are infectious, so they carry their recombinant DNA into bacterial cells.

Cosmids & Phagemids

- Once inside the cell, the DNA recircularizes through its cos sites, and from then on behaves exactly like a plasmid and replicate because it has a plasmid origin of replication.

Cosmids & Phagemids

- Cosmids are especially useful for the analysis of highly complex genomes and are an important part of various genome mapping projects.

Cosmids & Phagemids

- Another class of vectors that produce single-stranded DNA has also been developed.
- These are like the cosmids in that they have characteristics of both phages and plasmids; thus, they are called **phagemids**.

Cosmids & Phagemids

- One popular variety goes by the trade name pBluescript (pBS).
- Like the pUC vectors, pBluescript has a multiple cloning site inserted into the *lacZ* gene, so clones with inserts can be distinguished by white versus blue staining with X-gal.

Cosmids & Phagemids

- This vector also has the origin of replication of the single-stranded phage f1, which is related to M13.

Cosmids & Phagemids

- This means that a cell harboring a recombinant phagemid, if infected by an f1 helper phage that supplies the single-stranded phage DNA replication machinery, will produce and package single-stranded phagemid DNA.

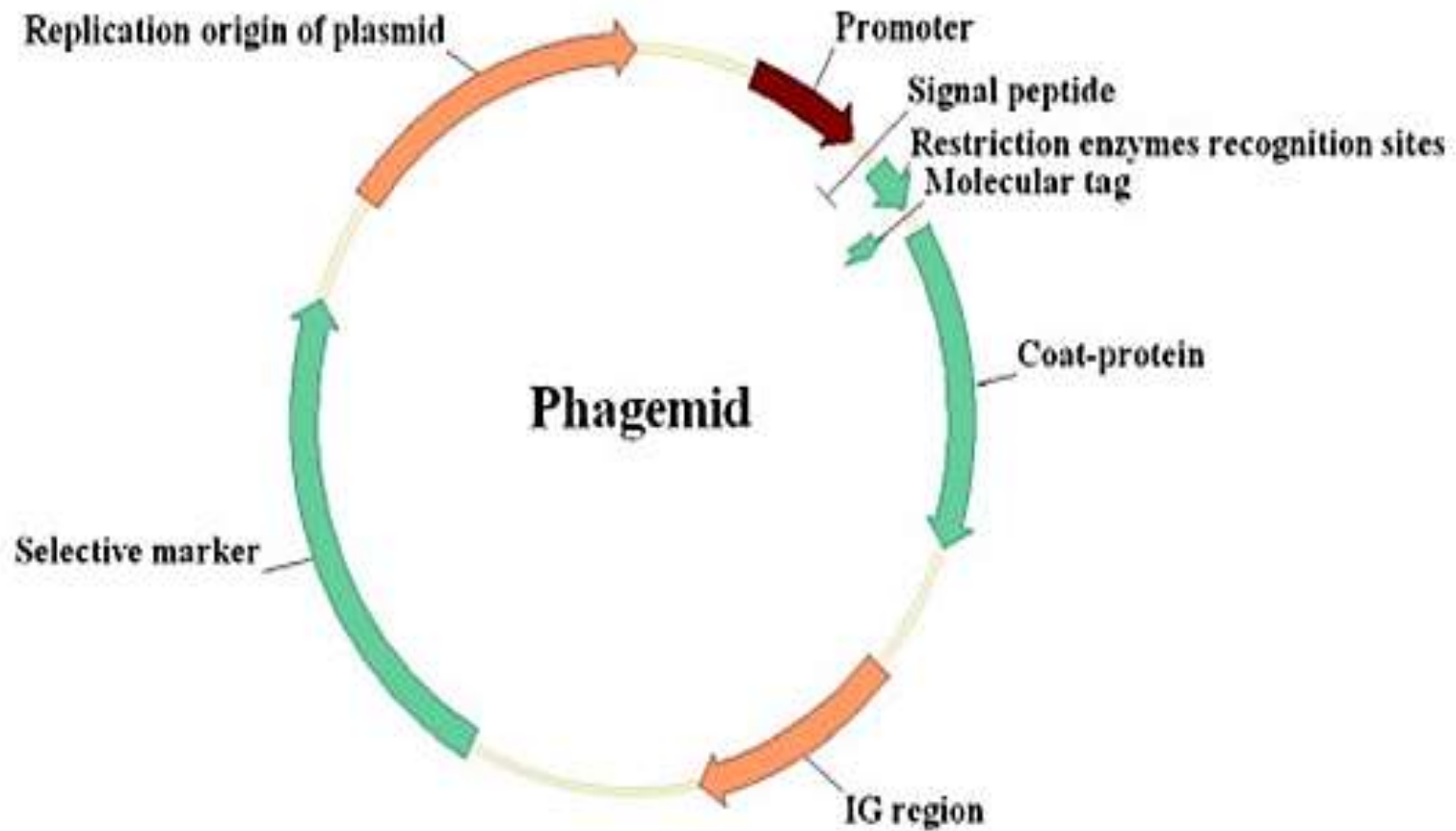
Cosmids & Phagemids

- A final useful feature of this class of vectors is that the multiple cloning site is flanked by two different phage RNA polymerase promoters.
- For example, pBS has a T3 promoter on one side and a T7 promoter on the other.

Cosmids & Phagemids

- This allows one to isolate the double-stranded recombinant phagemid DNA and transcribe it in vitro with either of the phage polymerases to produce pure RNA transcripts corresponding to either strand of the insert.

Cosmids & Phagemids



Large Insert Capacity Cloning Vectors & YAC vectors

- Vectors which accept larger fragments of DNA than phage A or cosmids have the distinct advantage that fewer clones need to be screened when searching for the foreign DNA of interest.

Large Insert Capacity Cloning Vectors & YAC vectors

- In addition they have had an enormous impact in the mapping of the genomes of organisms such as the mouse and are used extensively in the human genome mapping project.

Large Insert Capacity Cloning Vectors & YAC vectors

- Further developments have given rise to the production of large insert capacity vectors based on bacterial and mammalian artificial chromosomes (BACs and MACs) and on the virus P1, P1 artificial chromosomes (PACs).

Large Insert Capacity Cloning Vectors & YAC vectors

- However, perhaps most significant is the development of vectors based on **yeast artificial chromosomes** or **YACs** which could accommodate hundreds of thousands of kilobases each.

Large Insert Capacity Cloning Vectors & YAC vectors

- YACs containing a megabase or more are known as “**megaYACs.**”
- YACs are linear molecules composed of a centromere, telomere and a replication origin termed an ARS element.

Large Insert Capacity Cloning Vectors & YAC vectors

- A YAC contains a left and right yeast chromosomal telomere, which are both necessary to protect the chromosome's ends, and a yeast centromere,

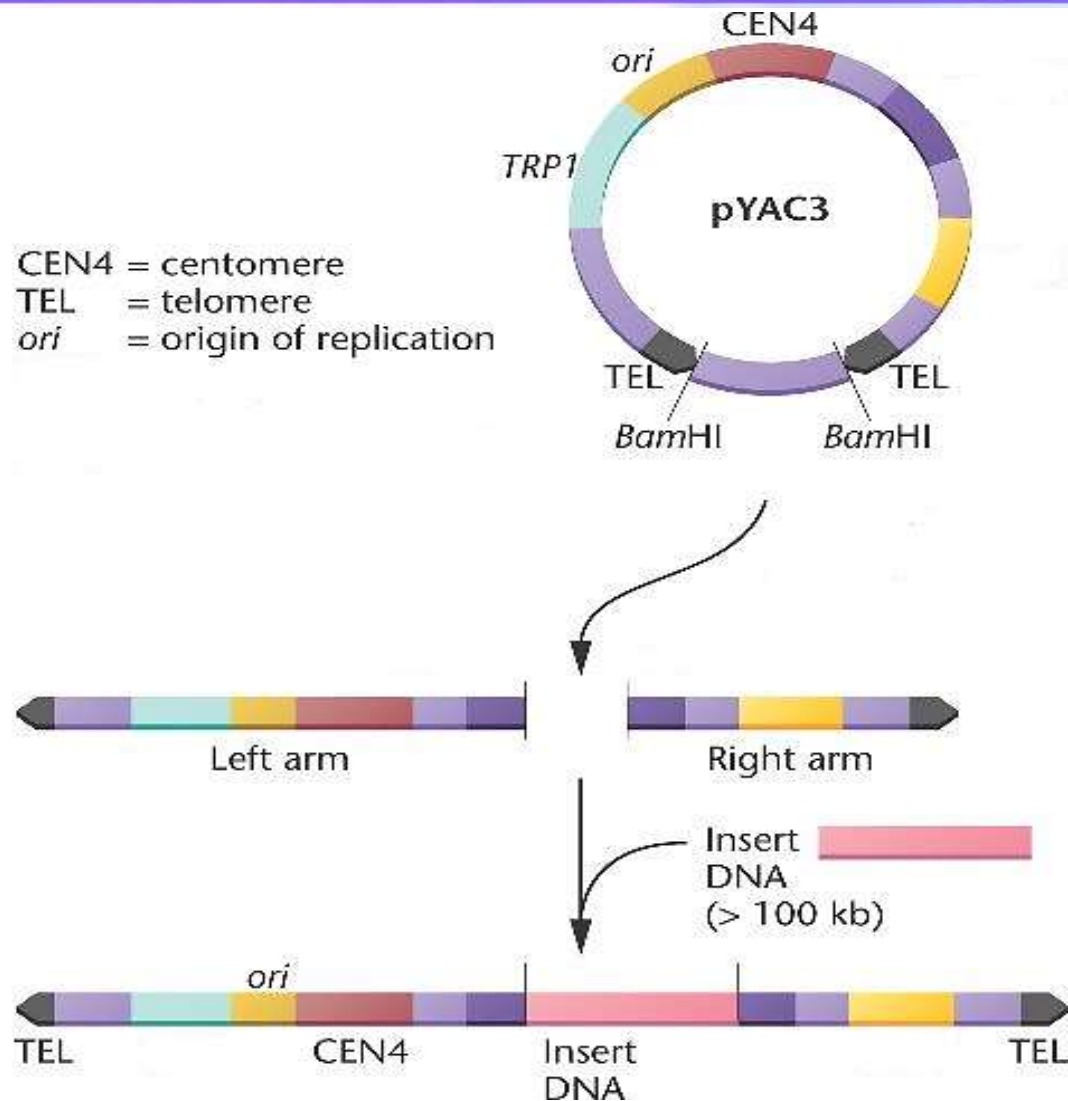
Large Insert Capacity Cloning Vectors & YAC vectors

- which is necessary for segregation of sister chromatids to opposite poles of the dividing yeast cell.

Large Insert Capacity Cloning Vectors & YAC vectors

- The centromere is placed adjacent to the left telomere, and a huge piece of human (or any other) DNA can be placed in between the centromere and the right telomere.

Large Insert Capacity Cloning Vectors & YAC vectors



Large Insert Capacity Cloning Vectors & YAC vectors

- The large DNA inserts are prepared by slightly digesting long pieces of human DNA with a restriction enzyme.

Large Insert Capacity Cloning Vectors & YAC vectors

- The YACs, with their huge DNA inserts, can then be introduced into yeast cells, where they will replicate just as if they were normal yeast chromosomes.

Large Insert Capacity Cloning Vectors & YAC vectors

- Using YACs, geneticists made great strides in the mapping phase of the Human Genome Project.
- They produced a genetic map of the whole genome that provided an average resolution of 0.7 centimorgan.

Large Insert Capacity Cloning Vectors & YAC vectors

- A centimorgan (cM) is the distance that yields a 1% recombination frequency between two markers and corresponds to an average of about 1 Mb in humans.

Large Insert Capacity Cloning Vectors & YAC vectors

- These researchers also produced relatively high-resolution physical maps of two of the smallest chromosomes, 21 and Y.

Large Insert Capacity Cloning Vectors & YAC vectors

- These maps were especially useful in that they represented long stretches of overlapping DNA segments cloned in YACs.

Large Insert Capacity Cloning Vectors & YAC vectors

- Thus, in the days before the human genome was sequenced, if you were interested in a disease gene that mapped to one of these chromosomes, you had a much simplified task.

Large Insert Capacity Cloning Vectors & YAC vectors

- You needed only to discover two markers flanking the gene of interest, look on the map to find which YAC or YACs contained these markers, obtain the YACs, and begin your final search for the gene.

END