

République Algérienne démocratique et populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université de Mohamed Kheider –BISKRA-
Faculté des Sciences de la Nature et de la Vie et des sciences de la terre et de l'univers
Département des Sciences de la Nature et de la Vie



Course and tutorial workbook

Molecular biology

3rd year LMD. Speciality: biochemistry

Prepared by: Dr. Asma MEDDOUR

First edition
Universitary year 2025/2026

Table of content

Table of content	1
Table of figures	5
Table of tables	6
Biology molecular courses definition	8
Introduction.....	9
1. Support of genetic information: DNA	10
1.1. Structure and Dynamics of DNA	10
1.1.1. Primary Structure: The Building Blocks.....	10
1.1.2. Secondary Structure of DNA: The Watson-Crick Double Helix.....	13
1.1.3. Alternative forms of double helix	15
1.1.4. DNA Dynamics and Stability:	17
1.1.5. Physicochemical and Biological Properties of Nucleic Acids.....	17
1.1.6. Topological Manipulations (DNA Topo-isomers).....	21
1.1.7. DNA-Protein Interactions	23
1.2. Structure and organization of prokaryotic and eukaryotic genomes:.....	23
1.2.1. Viral genomes	23
1.2.2. Structure and organization of prokaryotic genomes	24
1.2.3. Structure and Organization of Eukaryotic Genomes	25
1.2.4. Key Differences: Prokaryotic vs. Eukaryotic Genomes	29
1.3. Organization of organellar DNA	30
1.3.1. Mitochondrial DNA (mtDNA)	30
1.3.2. Chloroplast DNA (cpDNA)	30
1.3.3. Comparison: mtDNA vs cpDNA	31
1.3.4. Evolutionary Significance.....	31
1.4. Tutorial (Annex).....	31
2. Mutations, Mutagenesis and Detection:.....	32
2.1. Gene Mutations	32
2.1.1. Definition:	32
2.1.2. Classification.....	32
2.2. Mutagenesis.....	37
2.2.1. Spontaneous Mutagenesis.....	37
2.2.2. Induced Mutagenesis	42
2.3. Genotypic diagnosis	44
2.3.1. Principles of Mutation Detection	44

2.3.2.	Methods for Detecting Known Mutations	44
2.3.3.	Methods for Scanning Unknown Mutations	48
2.3.4.	Detection of Specific Mutation Types	52
2.4.	Tutorial (Annex).....	55
3.	Transmission and Conservation of Genetic Information	56
3.1.	DNA Replication and its regulation	56
3.1.1.	The semi-conservative model	56
3.1.2.	Characteristics of replication	57
3.1.3.	Molecular tools of replication:.....	58
3.1.4.	Molecular Mechanism of DNA Replication	58
3.1.5.	Regulation of replication:	68
3.2.	DNA Repair and Detection of Mutagenic Potential.....	71
3.2.1.	Mutations and DNA lesions:.....	72
3.2.2.	DNA Repair Mechanisms	73
3.2.3.	Detection of mutagenic potential	81
3.2.4.	Deficiency of the DNA repair systems and genetic diseases:.....	83
3.3.	Restriction-Modification Systems.....	84
3.3.1.	Types of Restriction-Modification Systems	84
3.3.2.	Biological Significance	85
3.4.	Tutorial (Annex).....	85
4.1.	Transcription and RNA Maturation	86
4.1.1.	Transcription	86
4.1.2.	RNA Maturation	93
4.2.	Translation and Protein Maturation.....	98
4.2.1.	Definitions.....	98
4.2.2.	The different partners of translation	99
4.2.3.	Mechanism of translation in prokaryotes:.....	100
4.2.4.	Mechanism of translation in eukaryotes:	104
4.2.5.	Protein Maturation	105
4.3.	Regulation of gene expression	107
4.3.1.	Overview	107
4.3.2.	Definitions.....	108
4.3.3.	Gene regulation in prokaryotes	108
4.3.4.	Gene Regulation in Eukaryotes	118
4.4.	Mechanisms of gene regulation by extracellular signals	121
4.4.1.	Transcription Factors: targets of signaling pathways	121
4.4.2.	Signals-transduced from Membrane to Nucleus.....	122

4.5. Tutorial (Annex).....	123
5. Methodology and molecular biology	124
5.1. Extraction and purification of nucleic acids.....	124
5.1.1. Extraction of nucleic acids.....	124
5.1.2. Purification of nucleic acids.....	125
5.1.3. Concentration.....	125
5.1.4. Quantification and qualification	125
5.1.5. Conservation:	125
5.2. Polymerase chain reaction (PCR) strategy.....	125
5.2.1. Definition	125
5.2.2. PCR variants:	126
5.2.3. Applications of PCR	126
5.3. Electrophoresis of Nucleic Acids.....	126
5.3.1. Migration speed	126
5.3.2. Gel supports	126
5.3.3. Visualization	127
5.3.4. Size determination	127
5.3.5. Special variants	127
5.4. Sequencing	127
5.4.1. Definition	127
5.4.2. Principle:	127
5.4.3. Automated sequencing.....	127
5.4.4. Sequencing technology	128
5.5. Molecular Cloning.....	128
5.5.1. Definition	128
5.5.2. Steps of cloning.....	128
5.5.3. Key technical points.....	128
5.5.4. Introduction of recombinant DNA into host cells.....	129
5.5.5. Ideal cloning host.....	129
5.5.6. Selection of transformed colonies.....	129
5.5.7. Gene expression	129
5.6. Molecular Hybridization technics	129
5.6.1. Definition	129
5.6.2. Stringency	129
5.6.3. Probes.....	130
5.6.4. Types of hybridization	130
5.7. Tutorial (Annex).....	131

Annex: Tutorials	132
Tutorial 1.....	132
Tutorial 2.....	134
Tutorial 3.....	136
Tutorial 4.....	138
Tutorial 5.....	140
References :.....	141

Table of figures

Figure 1. The Central Dogma of Molecular Biology.....	9
Figure 2. The bases of DNA and RNA	11
Figure 3. The sugars of DNA and RNA.....	11
Figure 4. structure of three nucleotides (with one, two and three phosphate)	11
Figure 5. Primary structure of DNA.	12
Figure 6. Hydrogen bonding between A-T and C-G pairs.....	13
Figure 7. Secondary structure of DNA.	13
Figure 8. Various representations of the DNA double helix.....	14
Figure 9. Alternative double-helical structures of DNA: B-DNA, A-DNA, and Z-DNA.	15
Figure 10. Alternative double-helical structures of DNA.....	16
Figure 11. DNA denaturation curve.....	19
Figure 12. The denaturation, renaturation, and hybridization of double-stranded DNA molecules.	20
Figure 13. DNA topoisomers.....	22
Figure 14. Supercoiling occurs in nature.	22
Figure 15. Structure of nucleosome	26
Figure 16. Chromatin structure: multiple levels of organization	27
Figure 17. Chromosome organization.....	28
Figure 18. Somatic and germline mutations	33
Figure 19. Structural aberrations of human chromosomes.	35
Figure 20. Meiotic non-disjunction.....	36
Figure 21. Pairing of adenine and guanine tautomers in DNA.....	37
Figure 22. Depurination mutation: release of purine base.	38
Figure 23. Examples of common DNA base modifications resulting from deamination:.....	39
Figure 24. Schematic representation of transposon	40
Figure 25. Effect of Retrotransposon on VvmybA1 Gene and Anthocyanin Production.....	40
Figure 26. Mutations caused by transposable elements.....	41
Figure 27. Pyrimidine dimer.	42
Figure 28. PCR components and process.....	45
Figure 29. Amplification Refractory Mutation System	46
Figure 30. Real time PCR	46
Figure 31. RFLP.....	47
Figure 32. Mico-array CGH.....	48
Figure 33. Sanger sequencing.	49
Figure 34. DGGE	51
Figure 35. HRM	52
Figure 36. Melting curve of HRM	52
Figure 37. Southern blot.....	53
Figure 38. Preparation of karyotype	54
Figure 39. Human karyotype	54
Figure 40. Preparation of FISH test	55
Figure 41. Meselson and Stahl Experiment (1958).....	57
Figure 42. Theta replication form	61
Figure 43. Rolling circle replication	61
Figure 44. Eukaryotic chromosome replication	62
Figure 45. DNA replication	64
Figure 46. Problem of telomer in eukaryotic DNA replication	67
Figure 47. Telomer DNA replication	68

Figure 48. Binding of Dna A protein to the origin <i>oriC</i>	70
Figure 49. Methylation of bacterial origin regulates replication during initiation.....	70
Figure 50. Mismatch Repair (MMR) system in <i>E. coli</i>	74
Figure 51. Direct reversal repair by MGMT	75
Figure 52. Base excision repair.....	76
Figure 53. Nucleotide excision repair in <i>E. coli</i>	77
Figure 54. Nucleotide excision repair in eukaryotes.....	78
Figure 55. Repair by Homologous Recombination	79
Figure 56. Repair by non-homologous end joining	80
Figure 57. Repair by SOS system	81
Figure 58. Ames test	82
Figure 59. Structure of prokaryotic gene	87
Figure 60. Structure of eukaryotic gene.....	87
Figure 61. Prokaryotic promotor.....	89
Figure 62. Prokaryotic transcription	90
Figure 63. Eukaryotic initiation transcription	92
Figure 64. mRNA Maturation in Eukaryotes.....	93
Figure 65. Capping of mRNA.....	93
Figure 66. Polyadenylation	94
Figure 67. Splicing.....	95
Figure 68. Structure of tRNA.....	96
Figure 69. tRNA activation.....	97
Figure 70. Prokaryote and eukaryote ribosomes.....	98
Figure 71. The genetic code.....	99
Figure 72. Structure of ribosome	100
Figure 73. The 30S initiation complex.....	101
Figure 74. The 70S initiation complex.....	101
Figure 75. Formation of peptide bond	102
Figure 76. Translation elongation in prokaryote.....	103
Figure 77. The 80S initiation complex.....	105
Figure 78. Translation elongation in eukaryote	105
Figure 79. Main points of gene regulation in prokaryotes.....	108
Figure 80. Positive and negative control.....	110
Figure 81. Lac operon	111
Figure 82. Lac operon in absence of lactose	111
Figure 83. Lac operon in the presence of lactose in the medium and absence of glucose....	112
Figure 84. Control of operon expression in the presence of glucose and lactose:.....	112
Figure 85. Control of operon expression in regulator gene mutation: Lac I ⁺ , Lac I ⁻ , Lac I ^s	113
Figure 86. Control of operon expression in operon mutation: O, O _c	114
Figure 87. The tryptophan operon of <i>E. coli</i>	115
Figure 88. Regulation of the tryptophan operon of <i>E. coli</i> by co-repressor	116
Figure 89. Regulation of the tryptophan operon of <i>E. coli</i> by attenuation	117
Figure 90. Regulation of the tryptophan operon of <i>E. coli</i> by attenuation in low levels of Trp	117
Figure 91. Regulation of the tryptophan operon of <i>E. coli</i> by attenuation in high levels of Trp	118
Figure 92. A generalized polymerase II promoter.	120
Figure 93. Intracellular pathways underlying regulation of gene expression.	122

Table of tables

Table 1. Nomenclature of Nucleosides and Nucleotides	12
Table 2. Alternative forms of the DNA double helix.....	17
Table 3. Key Differences: Prokaryotic vs. Eukaryotic Genomes	29
Table 4. Comparison between mtDNA vs cpDNA	31
Table 5. Some types of lesions and their repair mechanisms	73
Table 6. Examples of deficiency of the DNA repair systems and genetic diseases.....	83
Table 7. Gene Structure in Prokaryotes vs Eukaryotes	88
Table 8. the sub-units of RNA Polymerase of <i>E. coli</i>	88

Biology molecular courses definition

Fundamental Teaching Unit 1 (UEF1.6): Molecular Biology and Genetic Engineering (UEF 1.6.1): Molecular Biology (BM)

Credits: 5. Coefficient: 3

Teaching Objectives

The course aims to understand, in molecular terms, how genetic information of a living organism is expressed and regulated at the levels of transcription, translation, post-translational modifications, and import into subcellular compartments. It also presents how genetic engineering tools are used to isolate and characterize genes, modify them, and transfer them between species.

Recommended Prerequisites

Basic concepts in genetics and biochemistry acquired in L2 (second year)

Evaluation Method:

Final exam 60% + continuous assessment 40%

Course Content:

1. The Support of Genetic Information, DNA:

- Structure and dynamics of DNA
- Structure and organization of prokaryotic and eukaryotic genomes

2. Mutations, Mutagenesis and Detection:

- Gene mutations
- Mutagenesis
- Genotypic diagnosis

3. Transmission and Conservation of Genetic Information:

- DNA replication and its regulation
- DNA repair and detection of mutagenic potential
- Restriction-modification systems

4. Expression of Genetic Information and Its Control:

- Transcription and RNA maturation
- Translation and protein maturation
- Regulation of gene expression
- Gene regulation pathways by extracellular signals

5. Methodology and Molecular Biology

Introduction

Molecular biology is a science dedicated to the study of molecules carrying hereditary messages, particularly nucleic acids (DNA: Deoxyribonucleic acid, RNA: Ribonucleic acid), including their: structure, synthesis, alterations (mutations), expression and the control of their expression.

Molecular biology can be defined also, as a set of techniques used to conduct these studies.

In this course, we will be concerned with the structure, properties, and biological synthesis of the molecules that have been particularly important in molecular biology: DNA, RNA, and protein.

The structures of DNA and RNA molecules make them well suited for their major biological roles of storing and transmitting, and expression information into proteins. This information is fundamental to the growth and survival of cells, tissues and organisms because it specifies the structure of the molecules that make up a cell (**Robert Schleif, 1993**).

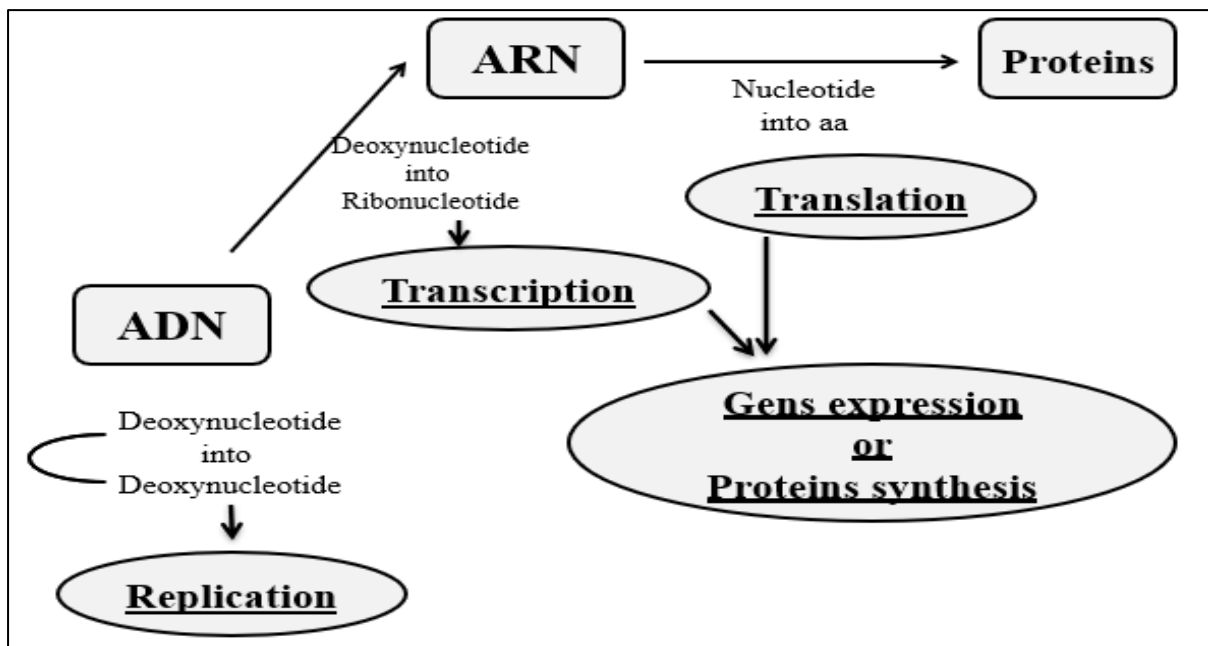


Figure 1. The Central Dogma of Molecular Biology

1. Support of genetic information: DNA

In the 1940s, several types of research showed that the molecule carrying biological information in chromosomes is DNA.

Finally, the detailed molecular structure of DNA was established by James Watson and Francis Crick in the 1950s. From this structure, they deduced that DNA contains information written in the form of a genetic code.

DNA is a linear sequence of four molecular building blocks called nucleotides. The specific sequence of nucleotides constitutes the language of the code. DNA, as element of the chromosome, is passed on intact form from one generation to the next, so that all cells in each generation contain an identical set of DNA with the same informative nucleotide sequence. As a result, one of the greatest secrets of life was solved: the architectural blueprint of life is DNA. This discovery was a key step in the genetics revolution.

To understand how DNA functions, we need to know its structure and organisation within cells.

1.1. Structure and Dynamics of DNA

By the mid-1940s, biochemists knew the fundamental chemical structures of DNA and RNA.

1.1.1. Primary Structure: The Building Blocks

DNA (deoxyribonucleic acid) is a polymer composed of **nucleotides**, each containing:

- Nitrogenous bases: four bases found in DNA are the purines adenine and guanine, and the pyrimidines cytosine and thymine (or uracil for RNA) (figure 2),
- Phosphoric acid: H_3PO_4 ,
- Sugar (Pentose C5): deoxyribose for DNA (hence the name deoxyribonucleic acid) and ribose for RNA.

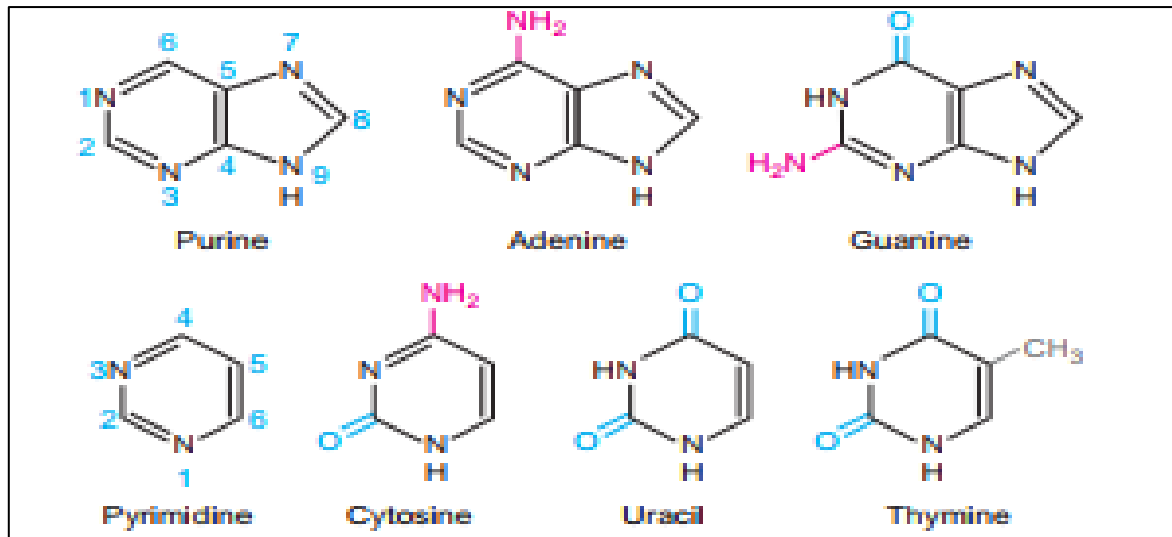


Figure 2. The bases of DNA and RNA

Figure 3 depicts the structures of the sugars found in nucleic acids. They differ in only one place. Where ribose contains a hydroxyl (OH) group in the 2-position, deoxyribose lacks the oxygen and simply has a hydrogen (H).

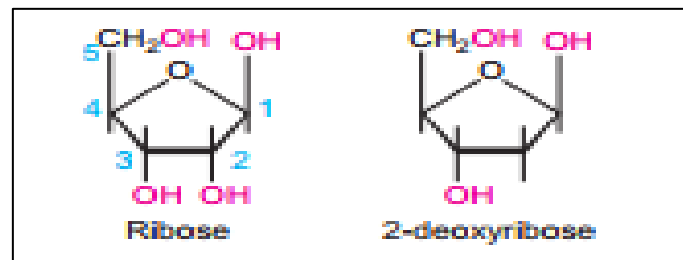


Figure 3. The sugars of DNA and RNA

The bases (N1 of pyrimidine or N9 of purine) and sugars (anomeric carbon C1') in RNA and DNA are joined together by β -N-glycosidic bonds into units called nucleosides. The subunits of DNA and RNA are nucleotides, which are nucleosides with a phosphate group attached (with C5') through a phosphor-ester bond. The names of the nucleosides and nucleotides derive from the corresponding bases (table1).

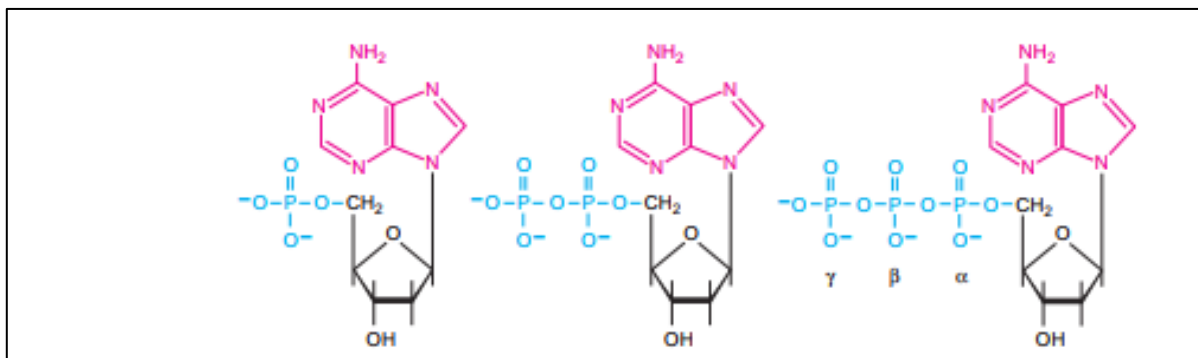


Figure 4. structure of three nucleotides (with one, two and three phosphate)

Table 1. Nomenclature of Nucleosides and Nucleotides

Base	Nucleoside		Nucleotide	
	RNA	DNA	RNA	DNA
Adenine (A)	Adenosine	Deoxyadenosine	Adenosine 5'-triphosphate (ATP)	Deoxyadenosine 5'-triphosphate (dATP)
Guanine (G)	Guanosine	Deoxyguanosine	Guanosine 5'-triphosphate (GTP)	Deoxyguanosine 5'-triphosphate (dGTP)
Cytosine (C)	Cytidine	Deoxycytidine	Cytidine 5'-triphosphate (CTP)	Deoxycytidine 5'-triphosphate (dCTP)
Thymine (T)	—	Deoxythymidine	—	Deoxythymidine 5'-triphosphate (dTTP)
Uracil (U)	Uridine	—	Uridine 5'-triphosphate (UTP)	—
Generic (N)	Nucleoside	Deoxynucleoside	Nucleoside 5'-triphosphate (NTP)	Deoxynucleoside 5'-triphosphate (dNTP)

* Nucleotides may contain one phosphate unit (monophosphate), two such units (diphosphate), or three (triphosphate). The triphosphate form shown in the table serves as the precursor building block for nucleic acid synthesis.

Nucleotides link through **phosphodiester bonds** between the 3'-OH of one sugar and the 5'-phosphate of the next, creating a sugar-phosphate backbone with a directional polarity (5' to 3'). It's the primary chemical structure of DNA and RNA.

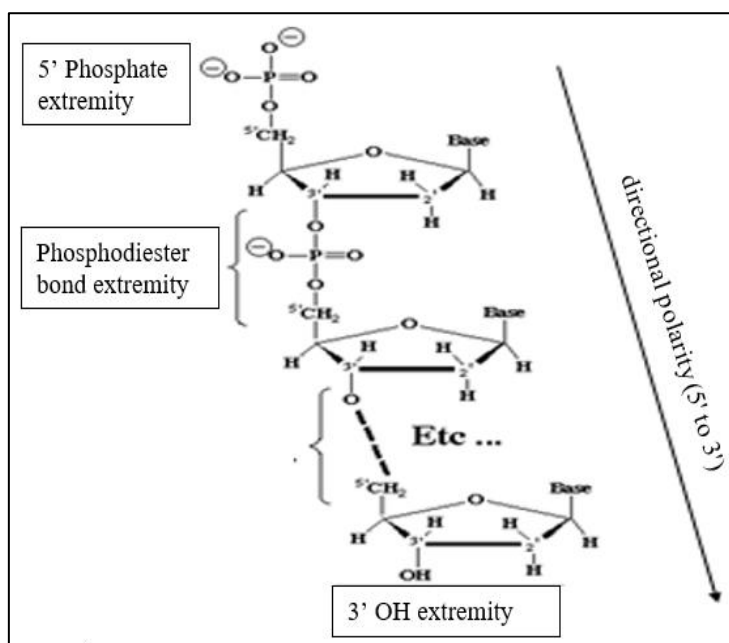


Figure 5. Primary structure of DNA.

1.1.2. Secondary Structure of DNA: The Watson-Crick Double Helix

Clearly the phosphate-sugar-phosphate-sugar along the backbones of DNA are regular, because, as Watson and Crick, the pairs of molecules: adenine-thymine (A-T) and guanine-cytosine (G-C) do possess regular shapes. The deoxyribose residues on both A-T and G-C pairs are separated by the same distance and are stabilized by strong hydrogen bonds. The A-T pair generally can form two hydrogen bonds and the G-C base pair can form three hydrogen bonds between the respective bases. Finally, the base pairs A-T and G-C can stack via hydrophobic interactions.

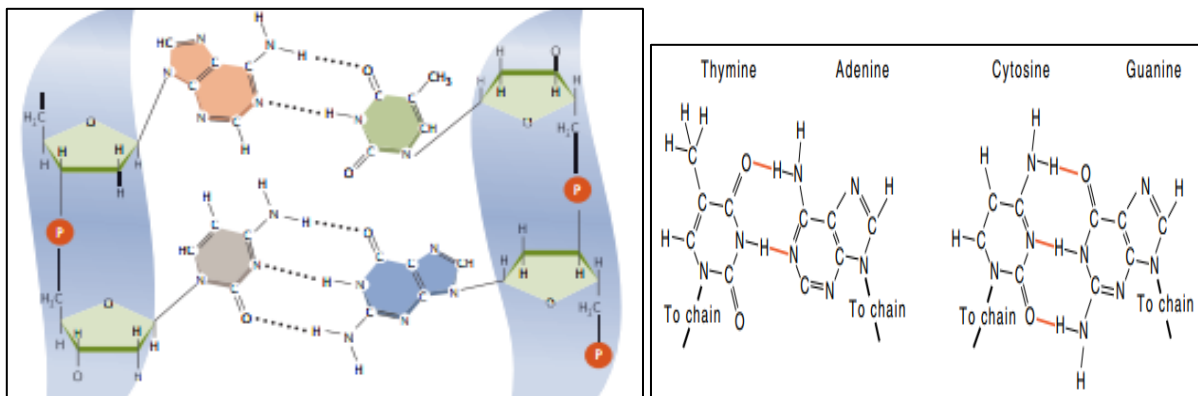


Figure 6. Hydrogen bonding between A-T and C-G pairs.

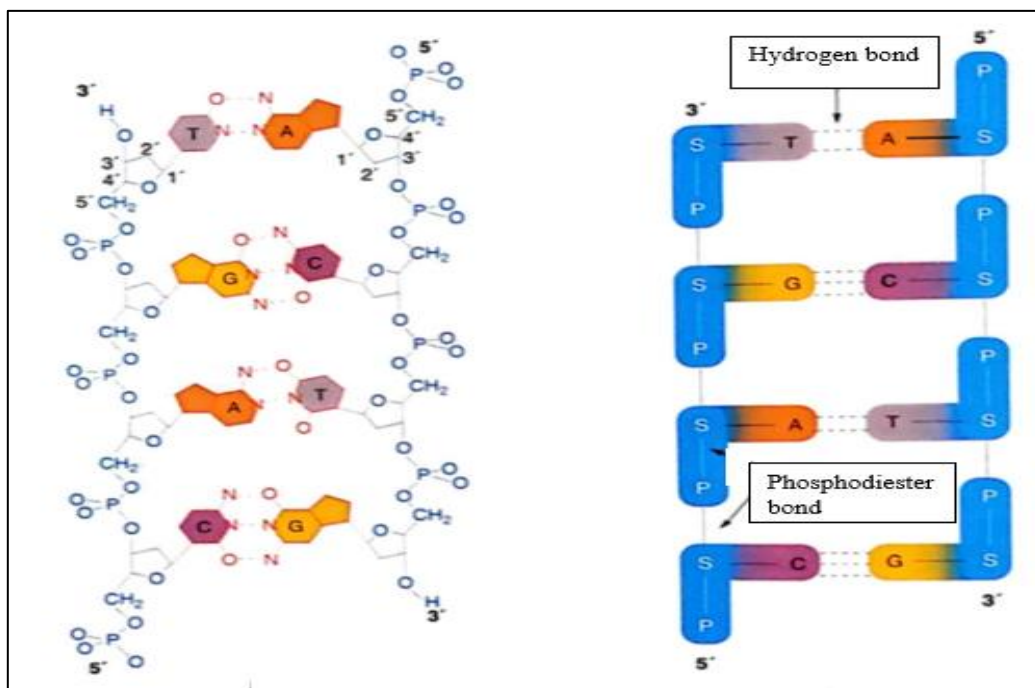


Figure 7. Secondary structure of DNA.

The famous double helix model, discovered by Watson and Crick in 1953, based on X-ray diffraction photos taken by Rosalind Franklin and Wilkins, features are:

- **Two antiparallel strands** running in opposite directions (one 5'→3', the other 3'→5')

- **Complementary base pairing** via hydrogen bonds:
 - Adenine pairs with thymine (A=T, 2 hydrogen bonds)
 - Guanine pairs with cytosine (G=C, 3 hydrogen bonds)
- **Helicoidal structure**
- **Major and minor grooves** where proteins can access base information
- **Right-handed helix** (B-form DNA, most common in cells)
 - Diameter: ~2 nm
 - One complete turn: ~10 base pairs
 - Rise per base pair: 0.34 nm

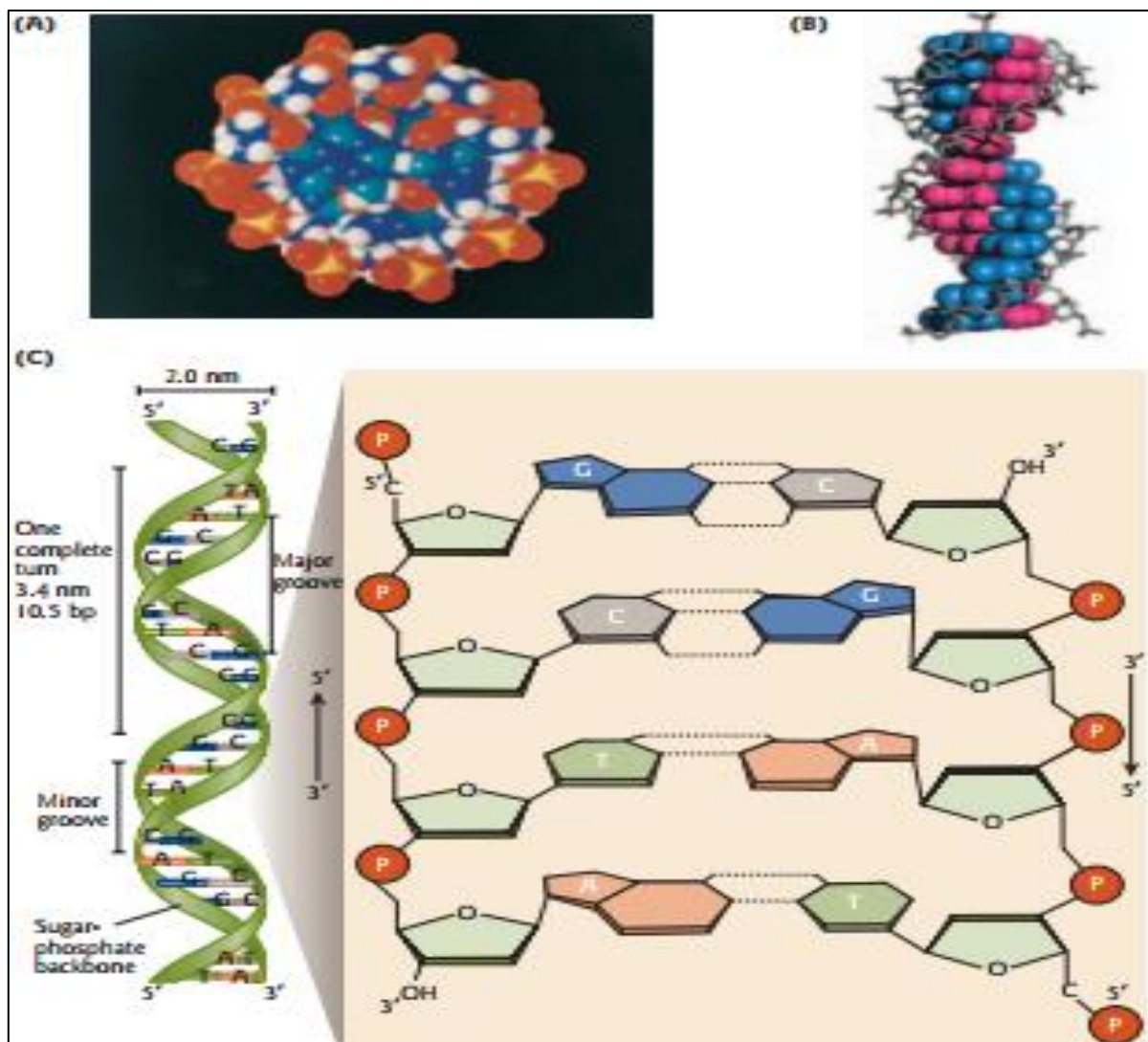


Figure 8. Various representations of the DNA double helix.

(A) The helix axis is most apparent from a view directly down the axis. The sugar–phosphate backbone is on the outside of the helix where the polar phosphate groups (red and yellow) can interact with the polar environment. The bases (blue) are inside, stacking perpendicular to the helix axis. (B) This picture shows how DNA bases are stacked together along the double helix. The atoms of the sugar–phosphate backbone are shown in gray, (C) The ribbon-like strands represent the sugar–phosphate backbones, and the horizontal rungs are the nitrogenous base pairs, of which there are 10.5 per complete turn. The major and minor grooves are apparent. The inset highlights the antiparallel nature of the two strands of the helix.

1.1.3. Alternative forms of double helix

Double-stranded DNA can have many alternative forms, the most famous are three: A, B, and Z.

- The B form is generally the most common form in the cell.
- The A-DNA form is the most compact.
- The Z form is identified in chromosomal regions that have a left-handed double helix.

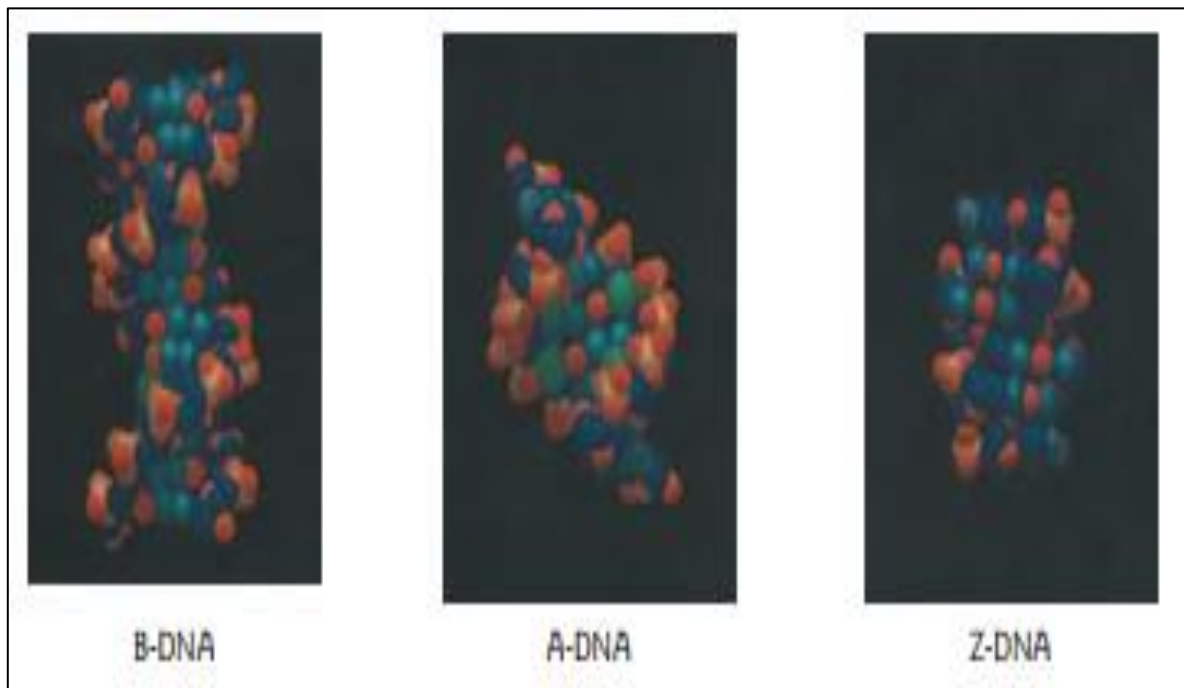


Figure 9. Alternative double-helical structures of DNA: B-DNA, A-DNA, and Z-DNA.

1.1.3.1. B-DNA:

B-DNA is a right-handed helix; it turns in a clockwise manner when viewed down its axis. The bases are stacked almost exactly perpendicular to the main axis with 10.5 bases per turn. The major groove is wide and of moderate depth, while the minor groove is of moderate depth but is much narrower. B-DNA occurs under conditions of high humidity (95%) and relatively low salt. Since the inside of a cell is mostly water with relatively low salt concentration, it follows that the predominant form *in vivo* is B-DNA.

1.1.3.2. A-DNA

If the water content is decreased and the salt concentration increased during crystal formation, the A form of DNA (A-DNA) will occur. In this right-handed helix the bases are tilted with respect to the axis and there are more (11) bases per turn than in B-DNA. The major groove of A-DNA is deep and narrow, while the minor groove is shallow and broad. The A-DNA is present in DNA/RNA hybrids formed during transcription. However, RNA adopts an A-form helix when it forms double-stranded regions.

1.1.3.3. Z-DNA

In 1979, Alexander Rich and his colleagues found that oligonucleotides composed of repeating GC sequences on one strand, with the complementary CG sequences on the other, formed a left-handed helix (turns counter-clockwise away). Because the backbone formed a zig-zag structure, they called the structure Z-DNA. It has 12 base pairs per turn. The minor groove is very deep and narrow. In contrast, the major groove is shallow to the point of being virtually non-existent. ZDNA is formed under conditions of high salt or in the presence of alcohol. This form of double helix can be stabilized in physiologically normal conditions, if methyl groups are added to the cytosines.

Z-DNA is thought to be present transiently in short sections in cells and to play a role in regulating gene expression. The crystal structure of a B–Z junction minimizes structural distortion of the helix, by extruding a base pair at each junction point. Base extrusion occurs in conjunction with a sharp turn in the sugar–phosphate backbone and a slight bend (11°). Thus, short sections of Z-DNA within a cell are more energetically favorable and stable than previously imagined.

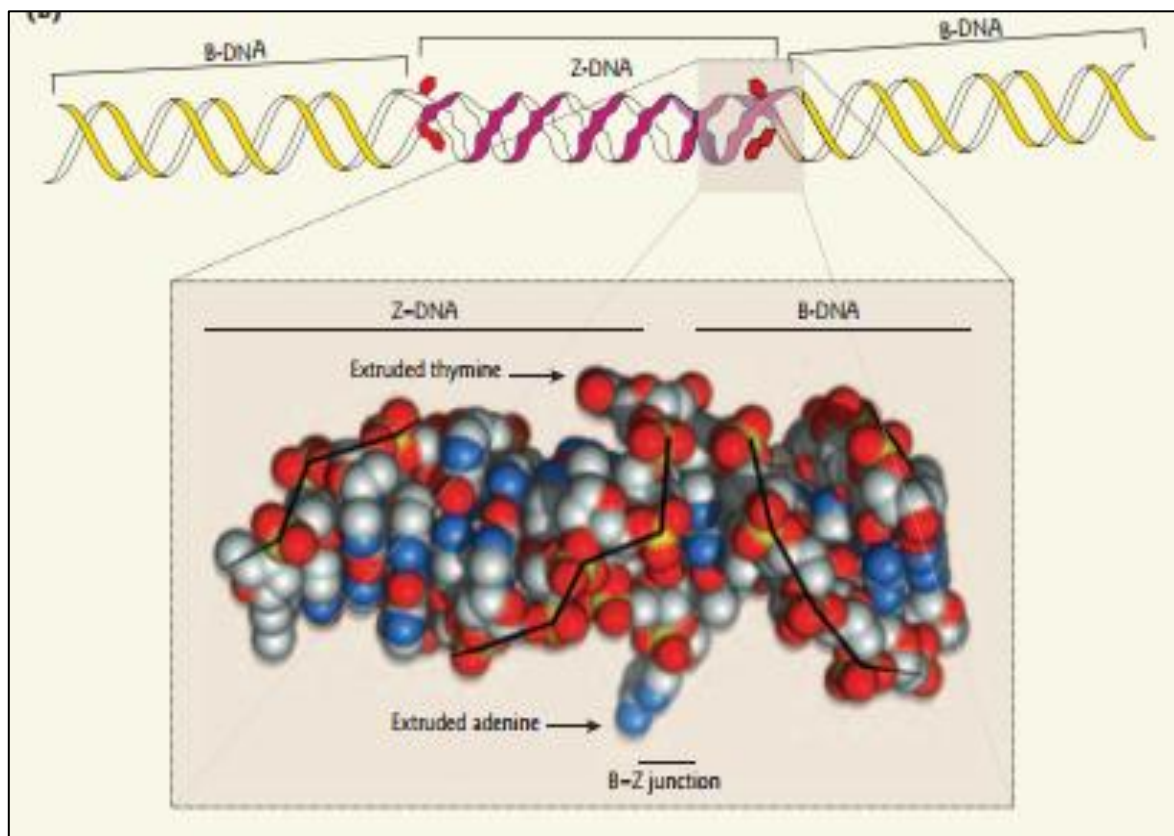


Figure 10. Alternative double-helical structures of DNA

The structure of the B–Z junction. A region of Z-DNA is connected to B-DNA through a junction in which one base pair is flipped out, or extruded, from the DNA helix.

Table 2. Alternative forms of the DNA double helix.

	Z-DNA	A-DNA	B-DNA
Orientation	Left-handed	Right-handed	Right-handed
Major groove	Very shallow, virtually non-existent, sometimes called a “single groove”	Deep and narrow	Moderate depth, wide
Minor groove	Very deep and narrow	Shallow and broad (superficial)	Moderate depth, narrow
Bases/turn	12	11	10.5 ~ 10
Conditions	High MgCl ₂ (> 3 m), NaCl, or ethanol In the presence of methylated cytosine: high humidity and low salt	Low humidity (75%), high salt	High humidity (95%), low salt

* Other forms of DNA have been crystallized, including B', C, C', C'', D, E, and T. All of these are right-handed structures, and occur under unique conditions. Ex: C-DNA forms in the presence of lithium salts and low humidity.

1.1.4. DNA Dynamics and Stability:

DNA stability depends on:

- **Hydrogen bonding** between complementary bases
- **Base stacking interactions** (hydrophobic forces between adjacent bases)
- **Ionic interactions** (phosphate backbone repulsion minimized by cations like Mg²⁺)

1.1.5. Physicochemical and Biological Properties of Nucleic Acids

1.1.5.1. Size: three characteristics can be used to express size:

- **Length:** in μm or nm
- **Molecular weight** in Daltons (Da): 1 Da = mass of a hydrogen atom
- **Number of nucleotides or bases (b)** for single-stranded molecules or base pairs for double-stranded molecules (bp). One bp corresponds on average to 660 Da.

1.1.5.2. Charge: DNA is negatively charged due to the presence of phosphate groups.

1.1.5.3. DNA Density:

After ultracentrifugation of a cell lysate in a density gradient formed by a cesium chloride solution, bands will appear, from bottom to top, of macromolecules: RNA, DNA, proteins. Therefore, the density of RNA > DNA > proteins.

1.1.5.4. DNA Solubility:

- DNA is soluble in water and saline solutions at low concentration

- At higher saline concentration, DNA precipitates (neutralization of negative charges)
- This precipitation is facilitated in the presence of ethanol (or isopropanol) because DNA is insoluble in it

Note: The solubility properties of RNA are equivalent to those of DNA. However, there is a small difference: RNA does not precipitate at 0.1 mol.L^{-1} in NaCl but remains soluble. This allows easy separation from DNA. RNA is precipitable by heavy cations (example: Li^+).

1.1.5.5. Optical Density:

DNA shows maximum light absorption in ultraviolet at a wavelength of 260 nm. One absorbance unit at 260 nm (OD read on spectrophotometer) corresponds to:

- 50 $\mu\text{g/mL}$ of a double-stranded DNA solution
- 37 $\mu\text{g/mL}$ of a single-stranded DNA solution
- 40 $\mu\text{g/mL}$ of an RNA solution

These values apply to perfectly pure nucleic acids in homogeneous solution.

A260 nm / A280 nm Ratio and Purity Control of DNA Solution:

Proteins absorb at 260 nm but their maximum absorption is at 280 nm. The A260 nm / A280 nm ratio is a way to assess possible contamination of the DNA solution:

- A ratio between 1.8 and 2 corresponds to a pure DNA solution
- A ratio less than 1.7 indicates proteins contamination
- A ratio greater than 2 indicates RNA contamination

Notes:

Possible phenol contamination can be detected by measuring absorption at 270 nm

RNA and single-stranded DNA (ssDNA) absorb more at 260 nm than double-stranded DNA (dsDNA) (hyperchromic effect)

1.1.5.6. Denaturation or Melting:

Separation of the two strands of double-stranded DNA under the effect of heat or weakly concentrated saline solutions

Denaturation is accompanied by a sharp increase in UV absorbance: the **hyperchromic effect**

Melting temperature or T_m : The temperature corresponding to denaturation of 50% of DNA molecules in a solution.

$$T_m = 2(A+T) + 4(C+G)$$

Factors affecting T_m :

- T_m depends on the length of the DNA molecule; small molecules are less stable

- T_m also depends on richness in C-G base pairs, which are more stable than A-T base pairs

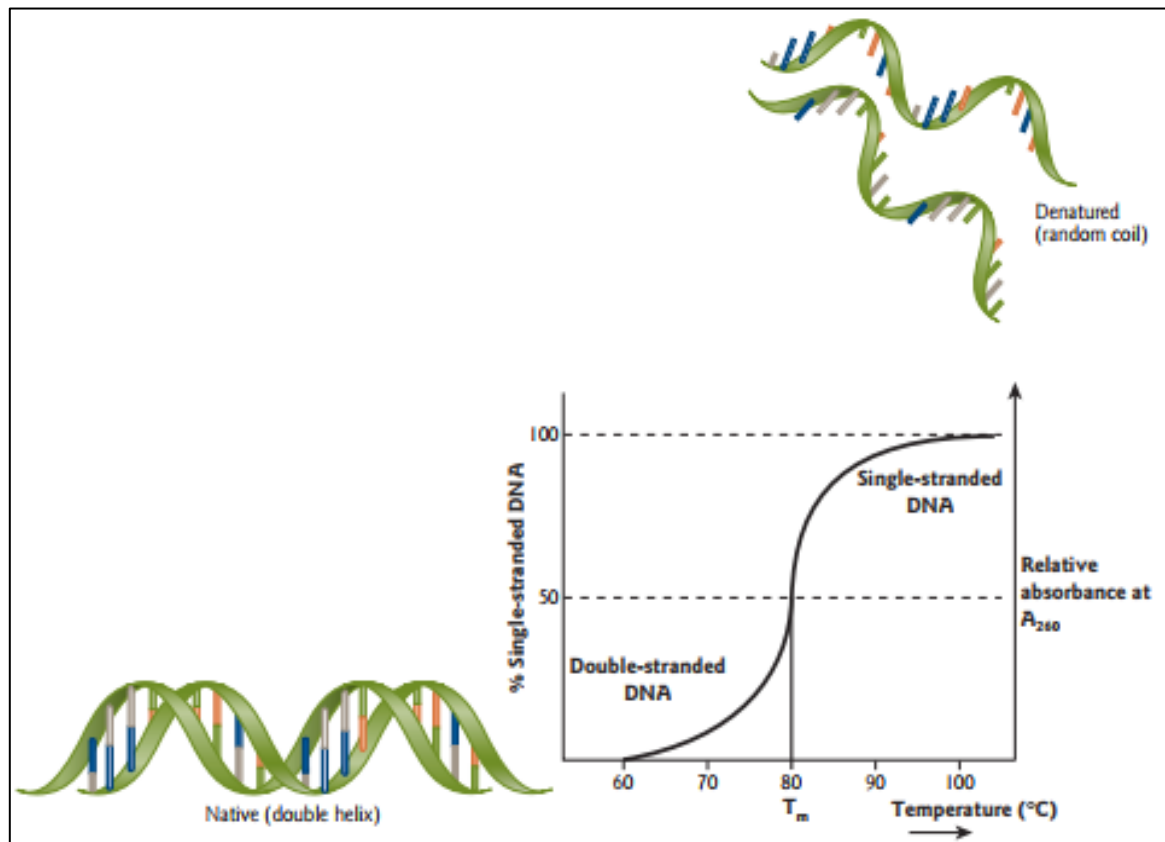


Figure 11. DNA denaturation curve.

Double-stranded DNA is heated and its melting is measured by the increase in absorbance at 260 nm. The point at which 50% of the DNA is single-stranded is the melting temperature, or T_m . In this example, the T_m is about 80°C.

1.1.5.7. Renaturation and Hybridization of DNA:

Melting is a reversible phenomenon: by slowly lowering the temperature, reassociation of the two strands occurs. When cooling is too rapid, reassociation is irreversible

Two single-stranded DNAs from two different origins when heated can also reassociate if there is complementarity between them; this is **hybridization**

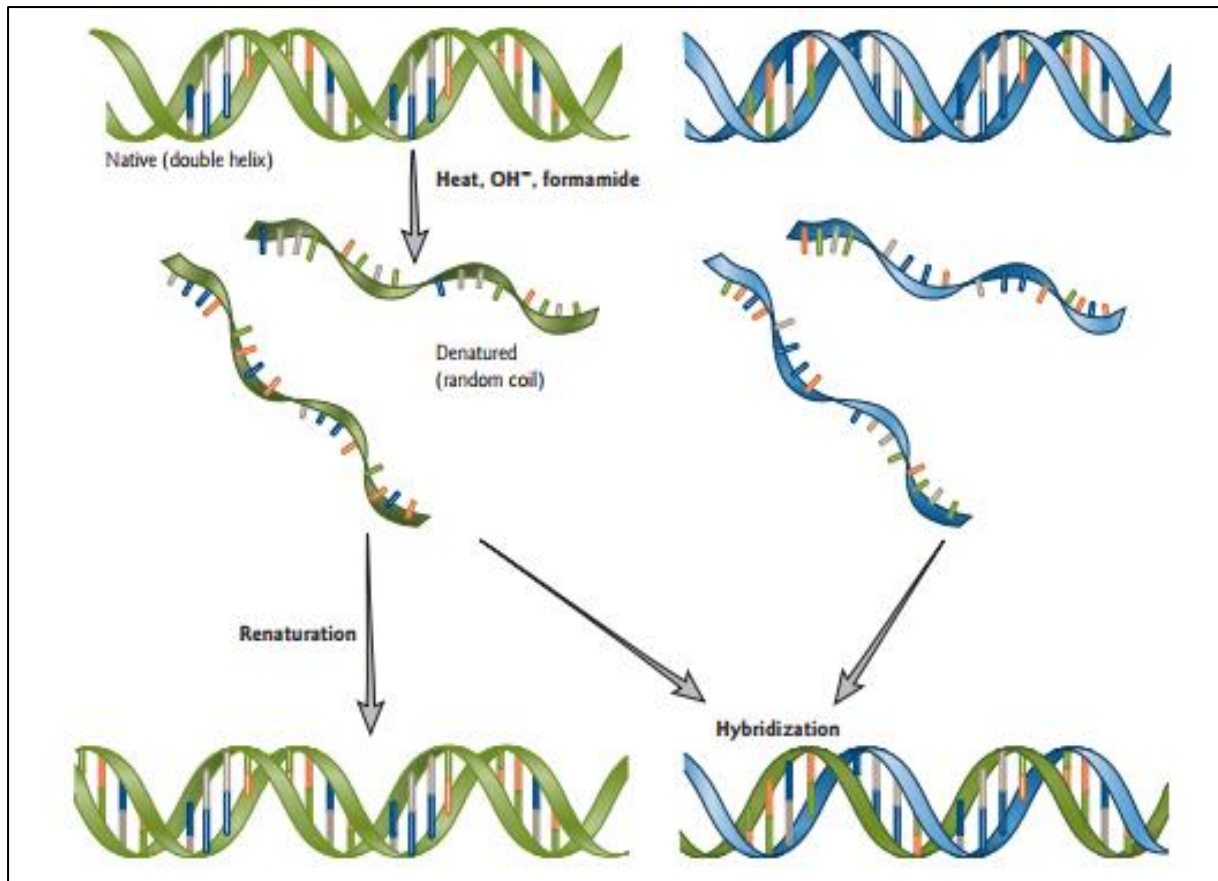


Figure 12. The denaturation, renaturation, and hybridization of double-stranded DNA molecules.

1.1.5.8. Hydrolysis of Nucleic Acids

Degradation of a polynucleotide can be chemical or enzymatic:

A. Chemical Hydrolysis:

Acid treatment affects DNA and RNA in the same way

Degradation of the phosphodiester backbone is obtained under conditions: concentrated acid and heating, which do not resist the bonds; this degradation leads to liberation of a mixture of phosphates, sugars, and bases (complete hydrolysis)

DNA and RNA react differently to alkaline hydrolysis:

DNA resists to basic pH because it has no OH on C2'

RNA is completely hydrolysed to its ribonucleotides in a few minutes at 37°C and pH 11. It's the presence of the free 2' hydroxyl that allows this hydrolysis, which gives a 2':3' P cyclic intermediate

B. Enzymatic Hydrolysis:

By specific phosphor-diesterases called nucleases. Nucleases have levels of specificity and are classified by:

- Their mode of chain attack: extremity (exonucleases) or interior (endonucleases)
- Their specificity for substrate: DNA (DNAase), RNA (RNAase), or both, and for structure; single or double strand
- Their recognition site specificity: bases or their sequence
- The type of phosphodiester bond cleavage (3' or 5' side...)

1.1.6. Topological Manipulations (DNA Topo-isomers)

Almost all DNA within both prokaryotic and eukaryotic cells exists in the negative supercoiled state. Negative supercoiling makes it easier to separate the DNA strands during replication and transcription. DNA can exist in:

- **Circular form**, without free ends, and in this form, DNA undergoes physical constraints that are lessened by numerous supercoiling to stabilize the DNA molecule as a whole
- **Linear form**, DNA ends are free and physical constraints introduced by its various ligands are easily adjusted. However, DNA winding around histone proteins and its organization contribute to circularizing DNA

Then, the circularity of DNA can therefore be considered a universal fact and an essential attribute of active DNA.

Topo-isomers:

Two DNAs having exactly the same number of nucleotides and the same base sequence, but differing by the linking number (the number of turns one strand makes around the other strand) are called **topo-isomers**.

DNA can exist in states:

- **Relaxed state**: tension on the double helix is minimal
- **Supercoiled state**: the axis of the double helix can wind on itself forming a super-helix. Two forms of supercoiling are theoretically possible:
 - **Positive supercoiling**: the linking number has been increased
 - **Negative supercoiling**: the linking number has been decreased

Negatively supercoiled DNA forms have the double advantage of being more compact, but more accessible to transcription and replication enzymes

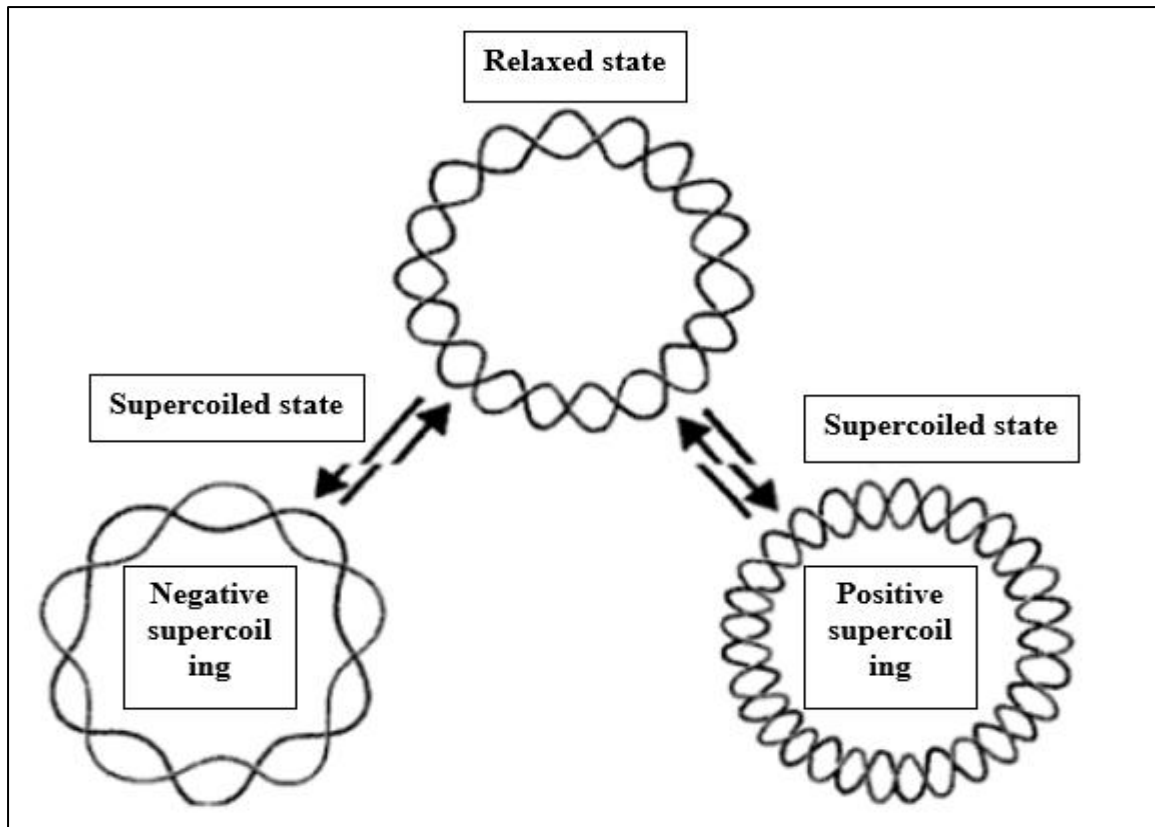


Figure 13. DNA topoisomers

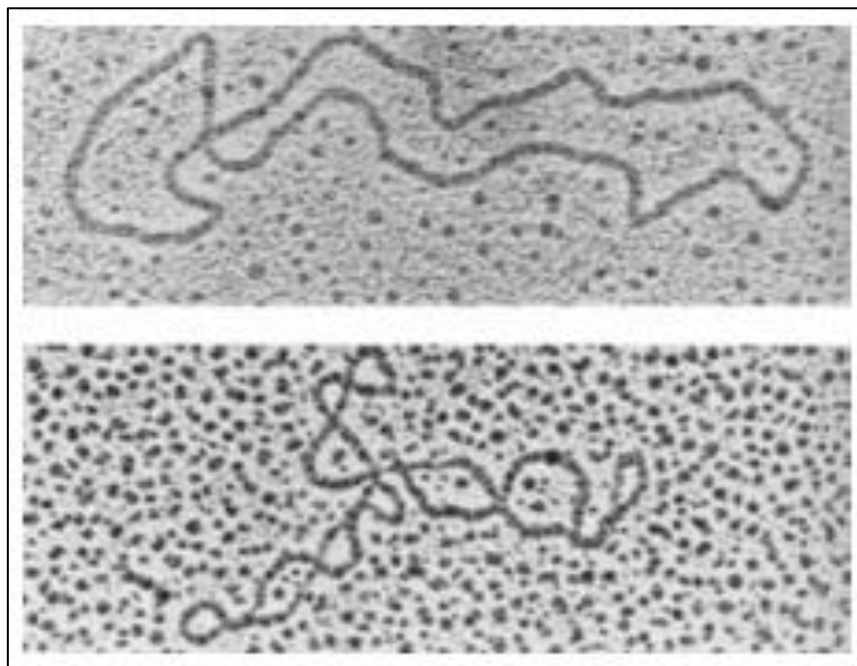


Figure 14. Supercoiling occurs in nature.

(A) The DNA of the bacteriophage PM2 in two topological forms: relaxed circle (upper panel) and supercoiled (lower panel). The latter is the native form. (Reproduced with permission from Wang, J.C. 1982. DNA topoisomerases. *Scientific American* 247:97.)

In very long DNA of eukaryotic cells, despite the linearity of the molecule, topological constraints remain within chromatin. It is therefore important that these constraints be reduced so that fundamental processes are made possible. A special class of enzymes, called **topoisomerases**, can introduce specific breaks, reduce constraints, relax supercoiling, and make physiological processes possible.

DNA topoisomerases fall into two major categories, type I and type II...

1.1.7. DNA-Protein Interactions

A DNA-binding protein is a protein containing a DNA-binding domain and therefore has an affinity for DNA

The sequence-specific DNA-binding proteins generally interact with the major groove of B-DNA because it exposes base pairs more, allowing better identification. However, there are ligands that interact with the minor groove

Among DNA-binding proteins are:

- Transcription factors
- Various polymerases
- Nucleases
- Histones...

1.2. Structure and organization of prokaryotic and eukaryotic genomes:

The term genome (a contraction of the words gene and chromosome) refers to all genetic information, encoded by DNA in most organisms and sometimes by RNA for certain viruses, that is hereditarily transmissible. More precisely, the term designates the DNA sequence corresponding to **a haploid set of chromosomes** (n chromosomes). A human somatic cell, diploid ($2n$ chromosomes), therefore contains two genomes (one paternal genome and one maternal genome) while a germ (sexual) cell, haploid (ovum or spermatozoon), contains only one. The term genome applies both to nuclear DNA (nuclear genome) and to organelle DNA (mitochondrial genome, chloroplast genome).

1.2.1. Viral genomes

Although viruses are not considered living organisms but rather molecular parasites, they reproduce by infecting living cells into which they inject their genome consisting of double-stranded or sometimes single-stranded DNA, protected within a particle or capsid of protein nature. Some viruses, such as retroviruses to which HIV belongs for example, have a genome made of RNA, which exists either as a single strand or double strand. Regardless of the nucleic acid, for some viruses it is in linear form and for others, in circular form. In all

cases, the viral genome exists at least in one phase of its cellular cycle in the form of double-stranded DNA.

1.2.2. Structure and organization of prokaryotic genomes

The genome of most prokaryotic organisms corresponds to a single chromosome composed of often circular DNA and very few associated proteins. However, a bacterium undergoing division may simultaneously contain several copies of this single chromosome.

1.2.2.1. General Characteristics

Prokaryotic genomes (bacteria and archaea) are typically:

- **Small and compact:** *E. coli* has average 4.6 million base pairs
- **Circular chromosomes:** usually one main circular DNA molecule
- **Located in the nucleoid:** not membrane-bound, but organized region
- **Gene-dense:** 85-95% of genome codes for proteins or functional RNAs
- **Minimal non-coding DNA:** Few non coding DNA, little repetitive DNA

1.2.2.2. Genome Organization

Prokaryotic genomes are typically organized as a single, circular, double-stranded DNA molecule (chromosome) located in a non-membrane-bound cytoplasmic region called the **nucleoid**. They are characterized by high coding density, lack of introns, supercoiled packaging, and often feature supplementary, independent circular DNA molecules known as **plasmids**. Genes are frequently arranged in **operons** for coordinated regulation.

Bacterial chromosome: the DNA is condensed through supercoiling (negatively supercoiled *in vivo*) and associated with non-histone, nucleoid-associated proteins (NAPs). This creates a dense, compact nucleoid structure rather than a membrane-bound nucleus, to fit into the small cell. Topoisomerases (DNA gyrase, topoisomerase I) manage supercoiling during replication and transcription.

Plasmids: Small, circular, extrachromosomal DNA molecules that replicate independently. They often carry genes for antibiotic resistance, virulence factors, or metabolic capabilities. Plasmids are crucial for horizontal gene transfer.

Gene organization (Operons): Prokaryotic genes (lack introns) are often clustered into operons: functional units containing multiple genes under the control of a single promoter (transcribed together into a single messenger RNA). This allows for the simultaneous expression of proteins with related functions. Exemple: *lac* operon, *trp* operon coordinated regulation of metabolic pathways

Polycistronic mRNA: One transcript encodes multiple proteins, increasing efficiency.

1.2.2.3. DNA Packaging in Prokaryotes

Despite lacking histones (with some archaeal exceptions), prokaryotes organize DNA using:

- **Nucleoid-associated proteins (NAPs):** HU, IHF, Fis proteins that bend and compact DNA
- **Supercoiling:** Creates organized loops and domains
- **Condensation proteins:** SMC (structural maintenance of chromosomes) proteins

1.2.3. Structure and Organization of Eukaryotic Genomes

1.2.3.1. General Characteristics

Eukaryotic genomes are:

- **Larger and more complex:** human genome is average of 3.2 milliards base pairs
- **Linear chromosomes:** multiple chromosomes contained in membrane-bound nucleus
- **Gene-sparse:** Only average of 1.5% of human genome codes for proteins
- **Rich in non-coding DNA:** introns, regulatory regions, repetitive elements, transposons, ...
- **Highly regulated:** complex chromatin structure controls gene accessibility

1.2.3.2. Chromatin structure

In eukaryotic cells, unlike prokaryotes, DNA is very frequently associated with basic proteins, the histones. The result of the regular assembly between DNA and histones is designated by the term **nucleosome**. The formation of nucleosomes is the first step in a process that ensures the packaging of DNA into compact structures that enormously reduce the volume occupied by the molecule. The whole is designated by the term chromatin.

Each class of histones consists of small proteins rich in positively charged amino acids (lysine and arginine). Their amino acid sequences have remained remarkably identical throughout evolution. Histones H2A, H2B, H3, and H4 (11 to 15 kd) are in equal quantity and each in duplicate; they form a histone octamer constituting a discoidal protein "core" of 8 proteins around which nucleosomal DNA with a size of 147 base pairs is wound. Histone H1 (20 kd) is not part of the "core." It is attached on one hand to the linker DNA located between nucleosomes and, on the other hand, to nucleosomal DNA, which has the effect of solidifying the whole.

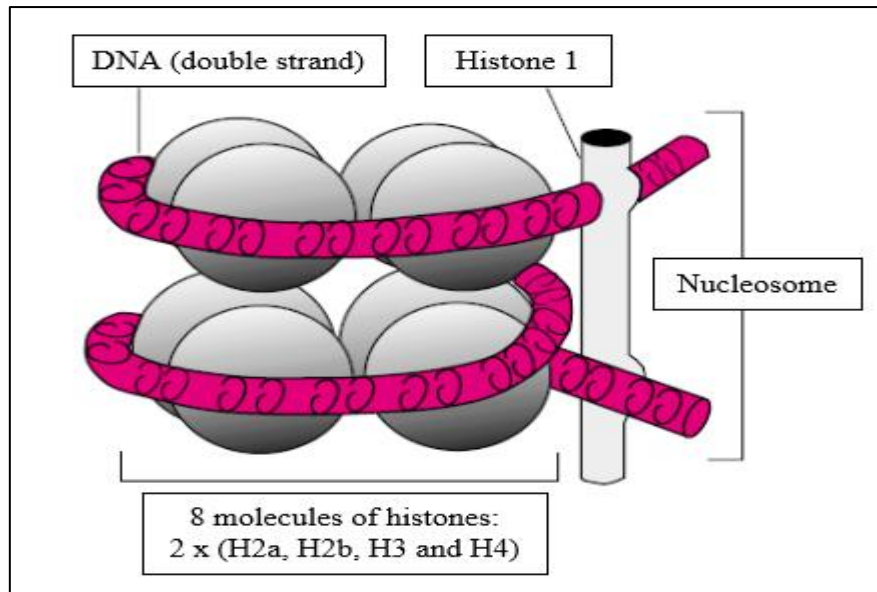


Figure 15. Structure of nucleosome

Numerous contacts exist between the histones of the octamer and nucleosomal DNA. They are independent of the DNA base sequence, which explains why almost all DNA in eukaryotic cells is organized into nucleosomes. The core histones have free N-terminal ends, designated by the term "tails." These tails emerge from the nucleosome at very precise positions. They facilitate the formation of higher-order structures resulting from the supercoiling of nucleosomes giving the final structure of chromatin. Certain amino acids in histone tails are sometimes modified by acetyl and methyl groups (lysine, arginine) or phosphate (serine). These modifications facilitate or alter chromatin accessibility. They would constitute an "epigenetic code" that could be read by proteins involved in regulating gene expression and in the various other functions of DNA. Nucleosomes are dynamic structures, constantly modified by numerous protein remodelling complexes capable of moving them along the DNA molecule, transferring them from one DNA molecule to another, or modifying them to increase DNA accessibility.

Multiple levels of organization of chromatin are described:

Level 1: The nucleosome: Basic structural unit of chromatin

The fundamental repeating unit:

- **Core:** Octamer of histone proteins (2 copies each of H2A, H2B, H3, H4)
- **DNA wrapping:** ~147 base pairs wrap around the histone core
- **Linker DNA:** ~20-80 base pairs connect nucleosomes
- **Histone H1:** Binds linker DNA, stabilizing higher-order structure
- Gives "beads-on-a-string" appearance (**11 nm fiber**)

Level 2: The 30-nm Fiber (Secondary structure)

Nucleosomes compact further into a solenoid structure (~6 nucleosomes per turn).

Level 3: Looped Domains: loop-shaped folding of the 300 nm fiber

The 30-nm fiber forms loops (40,000-200,000 base pairs) attached to a protein scaffold (nuclear matrix).

Level 4: Condensed Chromatin: interphase spirals

- **Euchromatin:** Loosely packed, transcriptionally active, light-staining
- **Heterochromatin:** Densely packed, transcriptionally silent, dark-staining
 - **Constitutive:** Always condensed (centromeres, telomeres)
 - **Facultative:** Can switch between condensed/open states

Level 5: Metaphase Chromosome

Maximum condensation during cell division (~10,000-fold compaction overall).

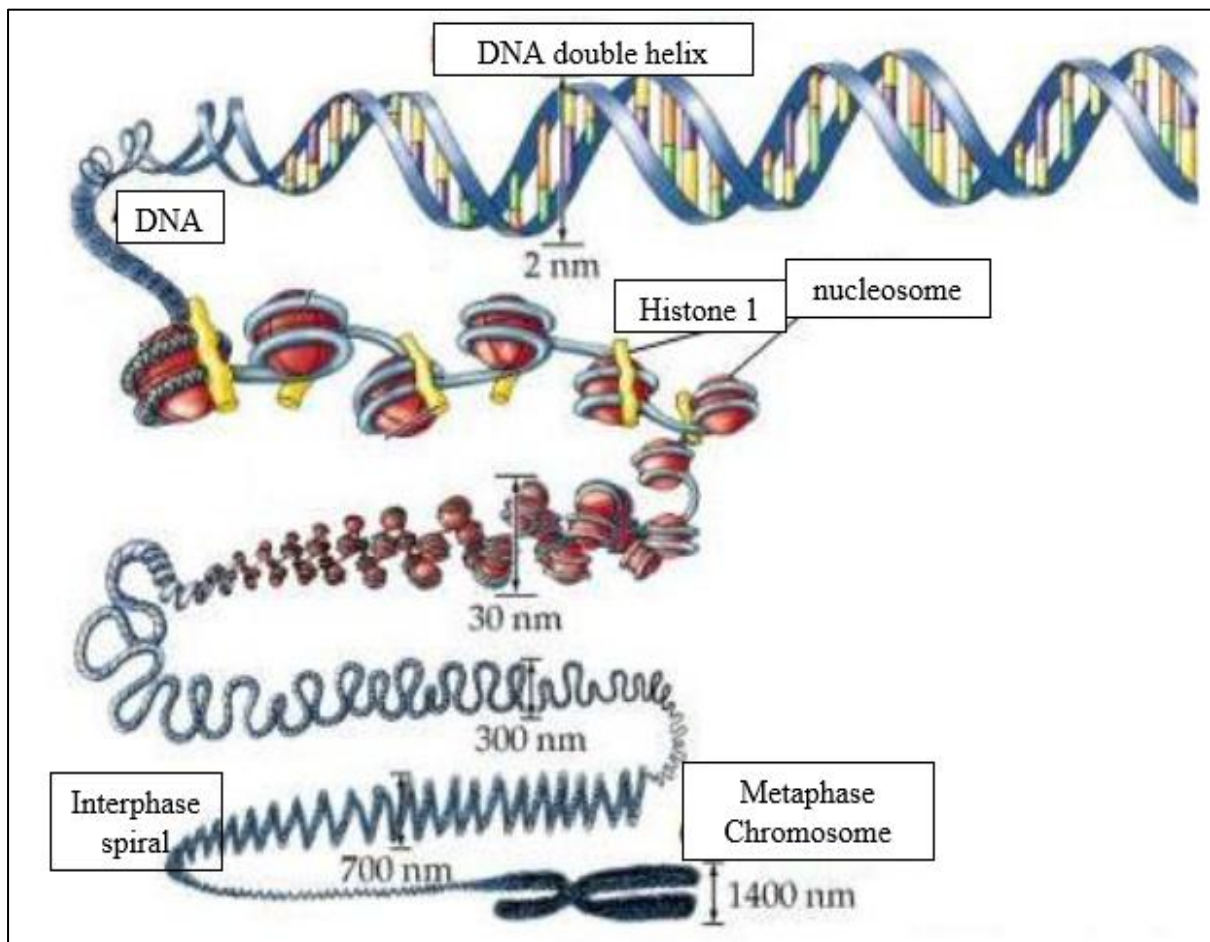


Figure 16. Chromatin structure: multiple levels of organization

1.2.3.3. Chromosome Features

Chromosomes are rod-shaped structures in the cell nucleus consisting of tightly coiled DNA and histone proteins, essential for genetic transmission during division. Key features

include a centromere dividing them into p (short) and q (long) arms, protective telomeres, and unique banding patterns used for identification. Humans possess 46 chromosomes, organized into 23 pairs: 22 pairs of autosomal chromosomes (numbered 1-22) and one pair of sex chromosomes (XX or XY). They are visualized during metaphase, when condensation is highest, allowing for karyotyping

A. Centromere: a constricted, heterochromatin region joining sister chromatids and acting as the attachment site for spindle fibers via the kinetochore (in chromosome segregation). Contains repetitive satellite DNA. Its position can be:

- **Metacentric:** Centromere is in the middle; arms are nearly equal in length.
- **Submetacentric:** Centromere is off-center, creating unequal p and q arms.
- **Acrocentric:** Centromere is close to one end, resulting in a very short p arm.
- **Telocentric:** Centromere is at the very end (not found in humans).

B. Arms (p and q): The centromere divides the chromosome into a short arm (p, for petit) and a long arm (q).

C. Telomeres: Protective caps at chromosome ends with repetitive sequences (TTAGGG in humans). Prevent degradation and end-to-end fusion. Maintained by telomerase enzyme.

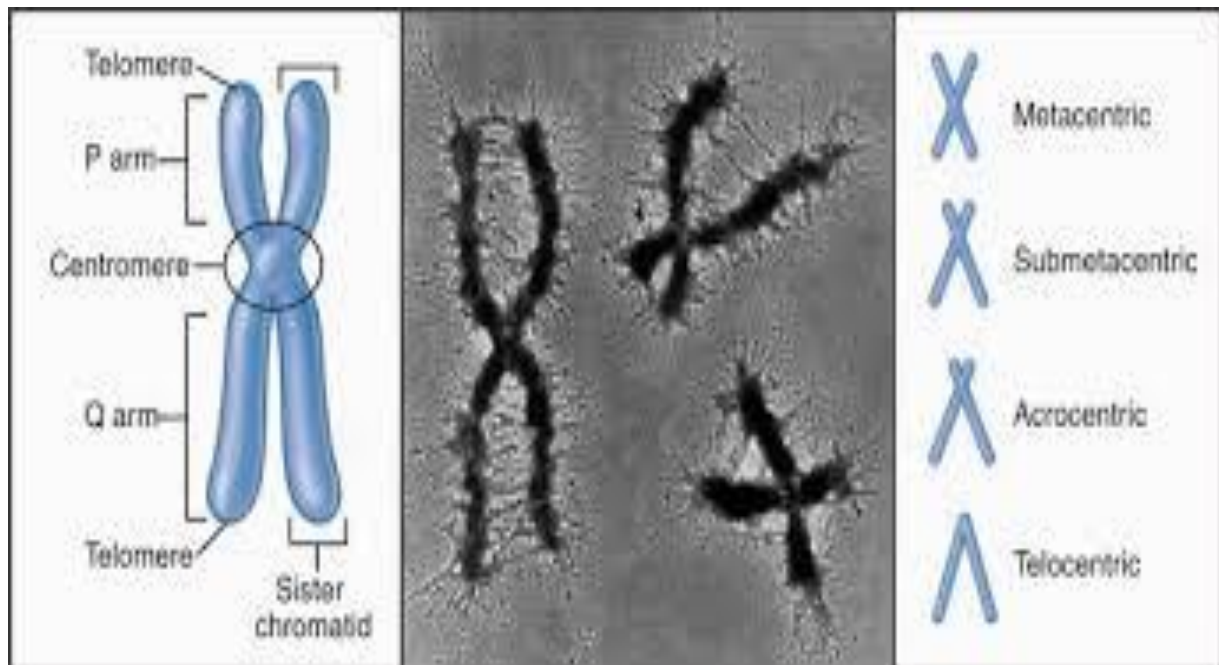


Figure 17. Chromosome organization

1.2.3.4. Genome Content

A. Protein-coding genes: ~20,000-25,000 in humans, often interrupted by introns

- **Introns and exons:** Genes split into coding sequences (exons) and non-coding intervening sequences (introns) that are spliced out
- **Regulatory elements:** Promoters, enhancers, silencers, insulators—can be far from genes they regulate

B. Repetitive DNA:

- **Tandem repeats:** Satellite DNA (centromeres), minisatellites, microsatellites
- **Interspersed repeats:** Transposable elements (SINEs: short interspersed nucleotide, LINEs: long interspersed elements, DNA transposons) make up ~45% of human genome

C. Non-coding RNA genes: rRNA, tRNA, snRNA, miRNA, lncRNA genes

1.2.3.5. Epigenetic Modifications of genome

Chemical modifications that don't change DNA sequence but affect gene expression:

- A. DNA methylation:** Addition of methyl groups to cytosine (especially in CpG dinucleotides), generally associated with gene silencing
- B. Histone modifications:** Acetylation (usually activating), methylation (can activate or repress), phosphorylation, ubiquitination affect chromatin structure and gene accessibility
- C. Histone code:** Specific combinations of modifications create binding sites for regulatory proteins

1.2.4. Key Differences: Prokaryotic vs. Eukaryotic Genomes

Table 3. Key Differences: Prokaryotic vs. Eukaryotic Genomes

Feature	Prokaryotes	Eukaryotes
Chromosome structure	Circular	Linear
Number	Usually, 1 chromosome	Multiple chromosomes
Location	Nucleoid (no membrane)	Nucleus (membrane-bound)
Size	Smaller (0.5-10 Mb typical)	Larger (Mb to Gb)
Gene density	High (85-95% coding)	Low (~1.5% coding in humans)
Gene structure	No introns (usually)	Introns common
mRNA	Polycistronic	Monocistronic
Organization	Operons	Individual gene regulation
Packaging	NAPs, supercoiling	Histones, chromatin
Extrachromosomal DNA	Plasmids common	Mitochondrial/chloroplast DNA

1.3. Organization of organellar DNA

Organellar DNA refers to the genetic material found in **mitochondria** and **chloroplasts** - semi-autonomous organelles that retain their own genomes separate from nuclear DNA.

1.3.1. Mitochondrial DNA (mtDNA)

1.3.1.1. General Structure

- **Form:** circular, double-stranded DNA molecule
- **Size:** varies widely across species: humans: ~16,569 bp, mammals: 16-17 kb, plants: 200-2,500 kb (much larger)
- **Packaging:** no histones; organized by TFAM (Transcription Factor A, Mitochondrial) proteins into nucleoid structures

1.3.1.2. Unique Features

- **Modified Genetic Code:** mtDNA uses a slightly different genetic code than nuclear DNA (Example: UGA codes for tryptophan instead of stop)
- **Compact Organization:** Genes lack introns, minimal intergenic sequences, some genes overlap
- **High Mutation Rate:** 10-17× higher than nuclear DNA due to lack of protective histones, limited DNA repair, and exposure to reactive oxygen species
- **Maternal Inheritance:** Exclusively inherited from the mother in most organisms
- **Multiple Copies:** 100s to 1000s of copies per cell, with multiple copies per mitochondrion.

1.3.2. Chloroplast DNA (cpDNA)

1.3.2.1. General Structure

- **Form:** circular, double-stranded DNA
- **Size:** typically, 120-200 kb
- **Packaging:** no histones; organized by proteins like SiR (Sulfite Reductase) into nucleoid structures
- **Unique Architecture:** quadripartite structure consisting of:
 - Large Single Copy (LSC) region: houses most of the essential photosynthetic apparatus gene
 - Small Single Copy (SSC) region: contains genes for photosynthetic electron transport and other essential functions
 - Two Inverted Repeat (IR) regions (IRA and IRB): (Two identical regions oriented in opposite directions) contain identical copies of: rRNA genes, several tRNA genes and some protein-coding genes



- **Copy Number:** Varies with developmental stage and environmental conditions
 - **Mature leaves:** 20-40 copies per chloroplast
 - **Young leaves:** Up to 200+ copies per chloroplast

1.3.3. Comparison: mtDNA vs cpDNA

Table 4. Comparison between mtDNA vs cpDNA

Feature	Mitochondrial DNA	Chloroplast DNA
Size	Different: 16-17 kb (mammals); 200-2,500 kb (plants)	120-200 kb
Shape	Circular	Circular with quadripartite structure
Gene Number	Different: 37 (humans)	100-120
Inverted Repeats	No	Yes (IRA, IRB)
Genetic Code	Modified (variant)	Universal (mostly)
Mutation Rate	High (10-17× nuclear DNA)	Lower than mtDNA
Inheritance	Maternal (typically)	Maternal (typically)
Primary Function	Oxidative phosphorylation	Photosynthesis

1.3.4. Evolutionary Significance

1.3.4.1. Gene Transfer to Nucleus

During evolution, most organellar genes were transferred to the nuclear genome. Only essential genes that must be rapidly regulated in response to the organellar environment remain in organellar DNA.

1.3.4.2. Endosymbiotic Origin

- **Mitochondria:** Derived from α -proteobacteria (~2 billion years ago)
- **Chloroplasts:** Originated from cyanobacteria (~1.5 billion years ago)

Evidence: circular DNA, 70S ribosomes, double membranes, independent division

1.4. Tutorial (Annex)

2. Mutations, Mutagenesis and Detection:

The term “mutation” refers to any change (accident) occurring in the DNA sequence. Mutations are permanent changes in genetic material, unlike DNA lesions that may be repairable.

We also speak of “variants”. However, non-pathogenic DNA variations, called “polymorphisms” are also mutations by definition.

Mutations can occur in:

- A chromosome: chromosomal mutations
- A gene: genic mutations, particularly in protein-coding regions (coding sequences) or outside the gene, for example in the region that regulates gene expression.

Mutations are the source of genetic variability. They are also responsible for genetic lesions that contribute to cell death, genetic diseases, and cancer.

2.1. Gene Mutations

2.1.1. Definition:

Mutation is a permanent modification in the genetic material sequence. It is a very rare event. A mutation can randomly affect any gene, in any cell, at any time. It is a random event. They are caused by:

- An error occurring during DNA replication
- Or various lesions caused by chemical or physical agents

Mutation introduces modification in genetic information, in the genotype of a cell: it is transmissible to its descendants (of the cell).

Then, A **mutation** is any heritable change in the DNA sequence. Mutations are the ultimate source of genetic variation and can be classified in multiple ways.

2.1.2. Classification

2.1.2.1. Classification by cell Type

For an organism, the consequences of a mutation will differ depending on the cell type affected:

- A. Somatic mutations:** occur in body cells, affect only the individual, not passed to offspring. Can lead to cancer or mosaicism (muted cell and its issues).
- B. Germline mutations:** the individual themselves will not be affected but some of their descendants may be. They occur in gametes (sex cells), heritable and passed to offspring, affect all cells of the next generation. In humans, such a mutation can thus be at the origin

of a hereditary genetic disease, with the defective copy of the gene being transmitted from generation to generation.

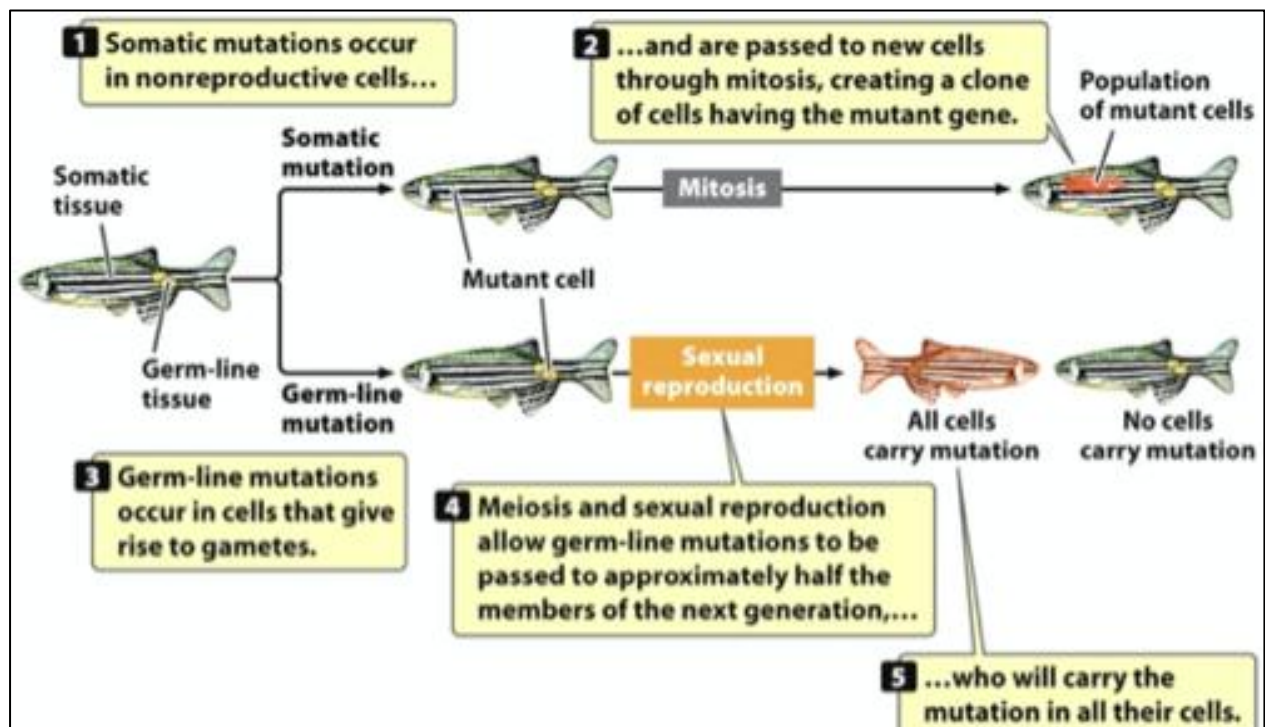


Figure 18. Somatic and germline mutations

2.1.2.2. Classification by Origin

A. Spontaneous mutations: occur naturally without external influence

- DNA replication errors (despite proofreading activity of DNA polymerase, the neo-synthesized DNA can contain errors ~ 1 in 10^{10} bases)
- Spontaneous base modifications (depurination, deamination)
- Transposable element movement

B. Induced mutations: caused by external mutagens (see mutagenesis part)

- Physical agents (radiation, UV light, ...)
- Chemical agents (various mutagens)
- Biological agents (viruses, transposons)

2.1.2.3. Classification by molecular nature

A. Point Mutations (Single Nucleotide Changes): are substitution mutations, corresponds to incorrect incorporation of a nucleotide on the daughter strand. One base is then replaced by another. Two types of this mutation are distinguished:

- **Transitions:** Purine to purine ($A \leftrightarrow G$) or pyrimidine to pyrimidine ($C \leftrightarrow T$): more common ($\sim 2/3$ of point mutations), often result from deamination or tautomeric shifts

- **Transversions:** Purine to pyrimidine or vice versa (A/G ↔ C/T): less common but more structurally disruptive

Consequences: With mismatches and after two replication cycles, the DNA has 25% mutant phenotype.

B. Insertions and Deletions (Indels): they can be:

- **In-frame:** multiples of 3 nucleotides, add/remove amino acids without shifting reading frame
- **Frameshift mutations:** not multiples of 3, shift the reading frame downstream: usually devastating; alter all subsequent amino acids; often create premature stop codons

C. Effects on Protein Coding:

- **Silent (synonymous) mutations:** change codon but not amino acid due to genetic code degeneracy. Example: GAA → GAG (both code for glutamate)
- **Missense (non-synonymous) mutations:** change one amino acid for another
 - **Conservative:** Similar amino acid properties (, leucine → isoleucine)
 - **Non-conservative:** different properties (Example: glutamate → valine in sickle cell anaemia). Effects range from negligible to severe depending on location and change
- **Nonsense mutations:** change amino acid codon to stop codon (UAA, UAG, UGA): produces truncated, usually non-functional protein. Example: UAC (tyrosine) → UAG (stop)
- **Read-frame mutations (Frameshift mutation):** serious mutations resulting from insertion or deletion of one or more nucleotides, causing a shift in the reading frame

D. Chromosomal mutations (Large-Scale)

- **Structural chromosome mutations:** due to chromosome breakage/reunion. They can be:
 - Balanced: without loss or gain of genetic material
 - Unbalanced: with loss or gain of genetic material
- **Deletions:** loss of chromosome segment, can range from few bases to entire genes or chromosomal regions.
- **Duplications:** copying of chromosome segment, can lead to gene dosage effects. It's a source of evolutionary innovation
- **Inversions:** reversal of chromosome segment orientation, can disrupt genes at breakpoints or alter gene regulation
 - **Pericentric:** includes centromere

- **Paracentric:** doesn't include centromere
- **Translocations:** transfer of segment between non-homologous chromosomes
 - **Reciprocal:** two-way exchange
 - **Robertsonian:** fusion of two chromosomes at centromeres

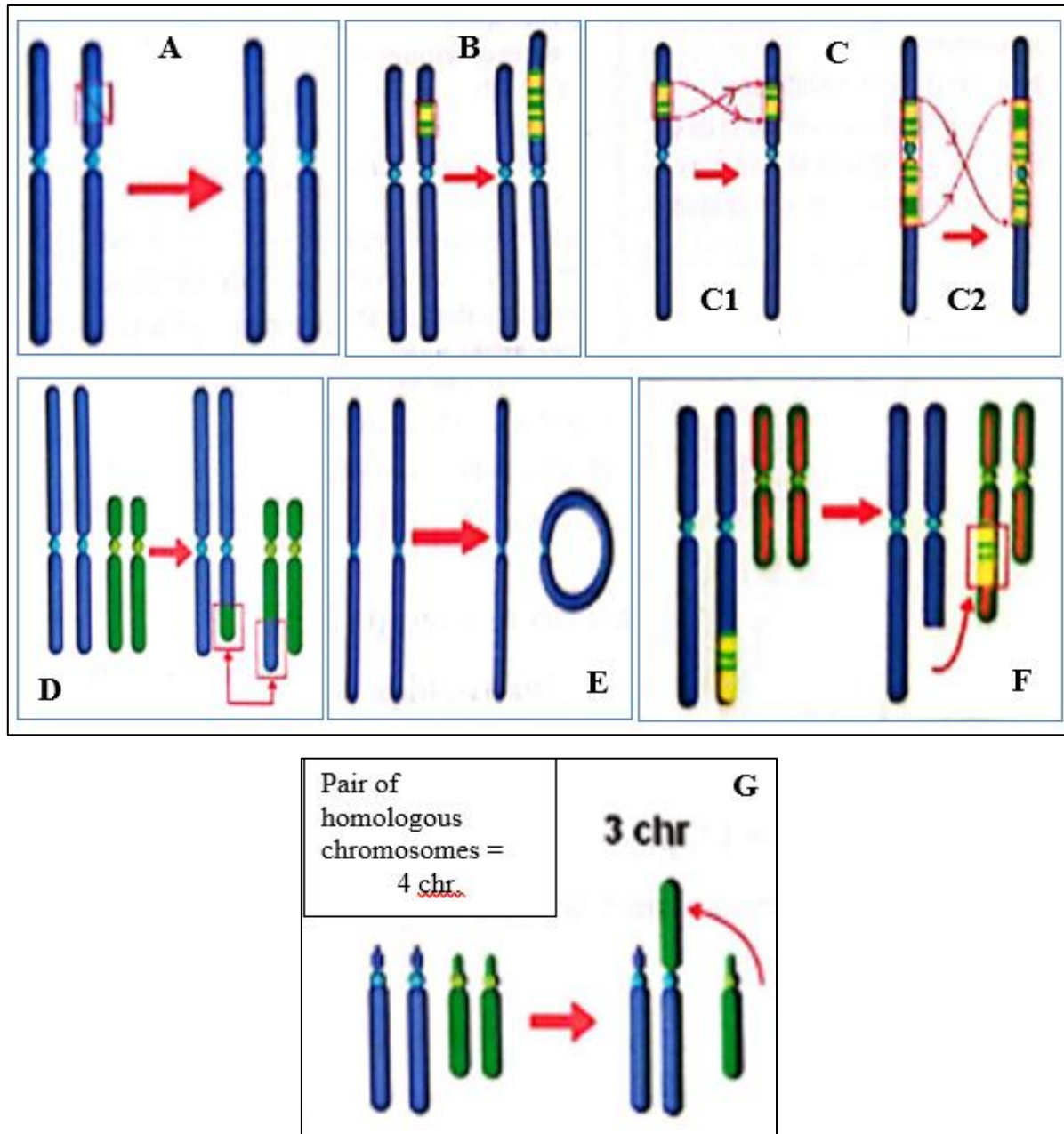


Figure 19. Structural aberrations of human chromosomes.

- (A) deletion, (B) duplication, (C) inversion: (C1) paracentric, (C2) pericentric, (D) reciprocal translocation, (E) ring-chromosome, (F) insertion, (G) Robertsonian translocation.
- **Numerical mutations:** one or more extra or missing chromosomes. They can be:
- **Aneuploidy:** abnormal chromosome number

- **Monosomy:** loss of one chromosome ($2n-1$), usually lethal except Turner syndrome (45, X)
- **Trisomy:** Gain of one chromosome ($2n+1$). Example: Down syndrome (trisomy 21)
- **Polyploidy:** Complete extra chromosome sets ($3n$, $4n$, etc.): usually lethal in animals, common in plants. They result from:
 - Meiotic non-disjunction (meiosis 1 or 2) during gametogenesis
 - Or a fertilization abnormality
 - Or mitotic non-disjunction (after fertilization, responsible for cellular mosaicism in the same individual)

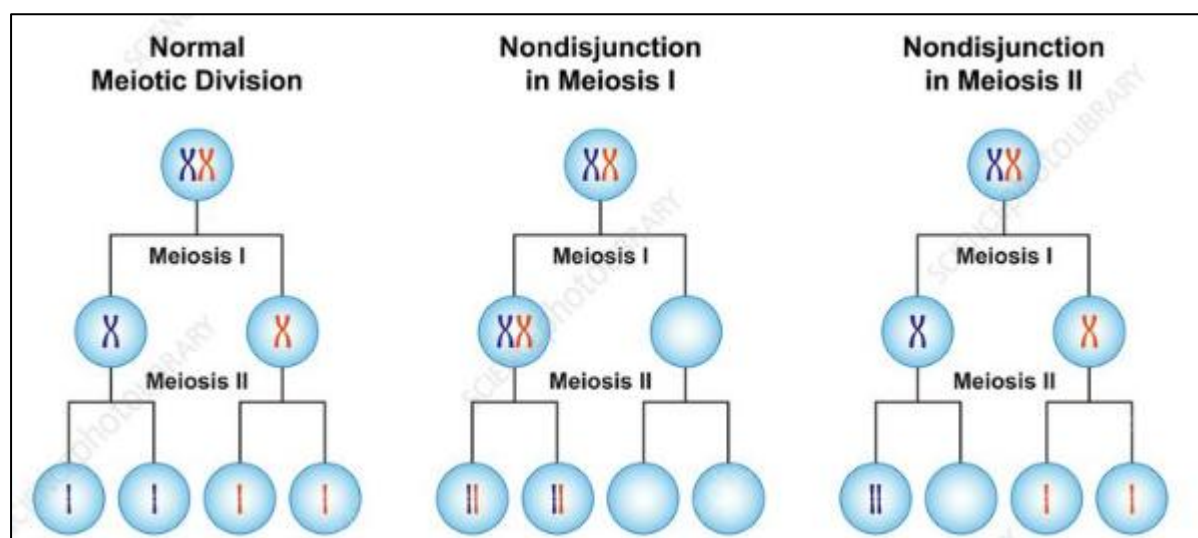


Figure 20. Meiotic non-disjunction.

2.1.2.4. Classification by Effect on Function

- A. Loss-of-function:** reduce or eliminate gene product activity. Often recessive (one functional copy usually sufficient)
- **Null (amorphic):** complete loss of function
 - **Hypo-morphic:** partial loss of function
- B. Gain-of-function:** increase normal activity or confer new function in inappropriate tissue or inappropriate time. Usually, dominant
- **Hypermorphic:** increased normal function
 - **Neo-morphic:** novel function
- C. Dominant-negative (Forward mutation):** mutant product interferes with wild-type product: common in multimeric proteins
- D. Conditional mutations:** manifest only under certain conditions

E. Reverse mutation: restores the wild-type function

F. Lethal mutation: causing premature death of affected organism

2.2. Mutagenesis

2.2.1. Spontaneous Mutagenesis

2.2.1.1. DNA Replication Errors

A. Tautomeric shifts:

Each DNA base can exist in different tautomeric forms: the most common is the **keto form** (present at physiological pH). A rare **enol form** (C=O for G and T) or **imino form** (N=H for A and C), which form unusual pairings outside of AT and GC, not detected by the proofreading function, leading to a mismatch after two replications (The number of hydrogen bonds between the paired bases is different):

- The imino form of cytosine pairs with adenine
- The imino form of adenine pairs with cytosine
- The enol form of thymine pairs with guanine
- The enol form of guanine pairs with thymine

If occurs during replication, causes mispair (e.g., A-C instead of A-T)

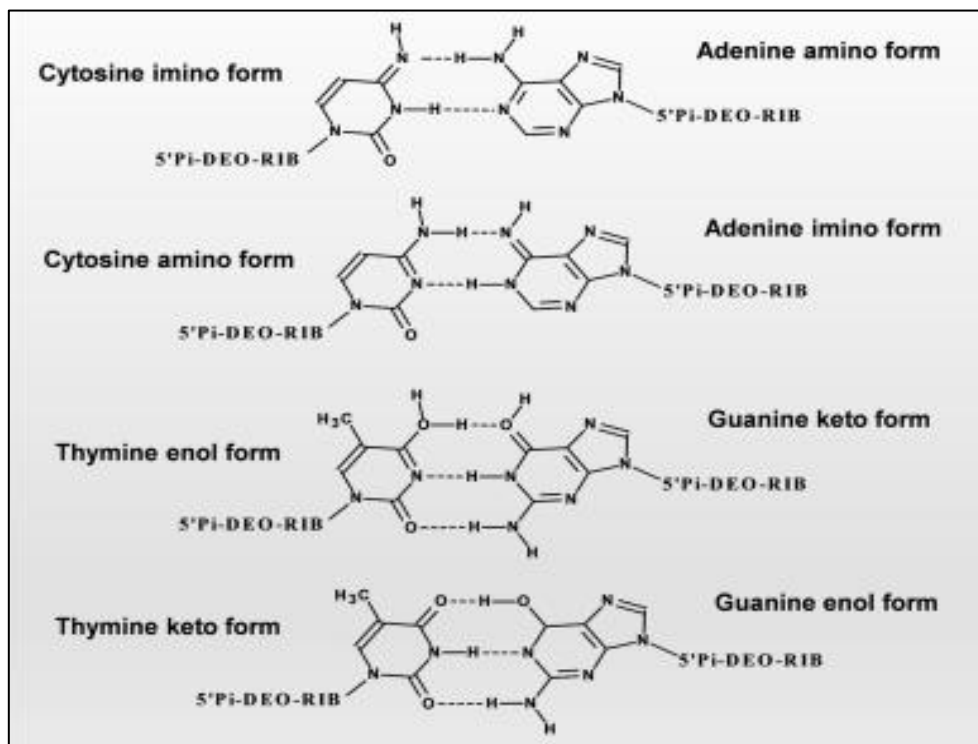


Figure 21. Pairing of adenine and guanine tautomers in DNA

B. DNA polymerase errors: Despite 3'→5' exonuclease proofreading

- Error rate after proofreading: $\sim 10^{-10}$ per base per replication
- Mismatch repair provides additional correction

C. Slippage: In repetitive sequences (microsatellites, STRs)

- Template strand loops out → deletion
- New strand loops out → insertion
- Basis for microsatellite instability in some cancers

2.2.1.2. Spontaneous Chemical Changes

A. Spontaneous depurination: interruption of an N-glycosidic bond between the purine base (A or G) and the deoxyribose, resulting in the loss of a guanine or adenine residue from DNA and creation of an apurinic (AP) site.

Depurination is a very common phenomenon: on the order of 5,000 to 10,000 depurinations per day per cell in mammals.

Note: Apyrimidinic (AP) sites are much less frequent and occur much more slowly.

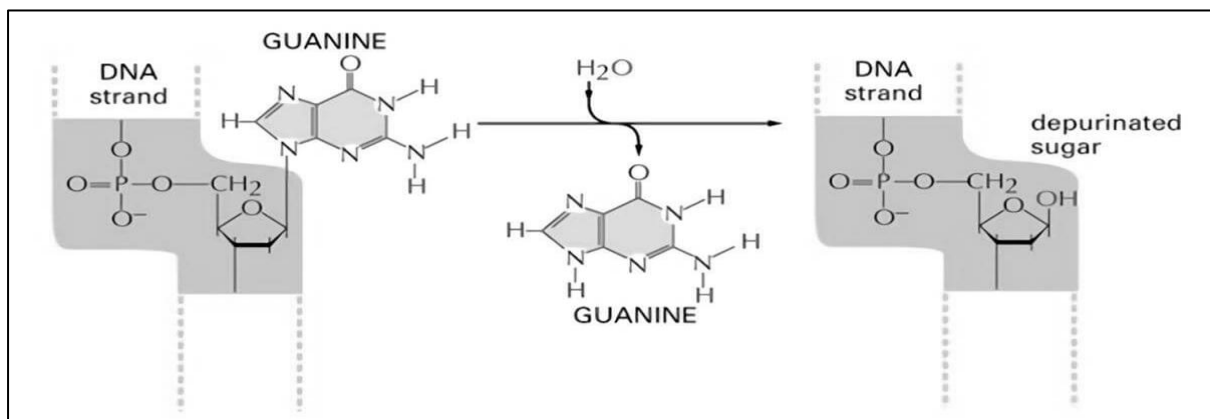


Figure 22. Depurination mutation: release of purine base.

B. Spontaneous deamination of natural bases: also called oxidative deamination, it leads to the formation of modified base (deaminated base), which establishes specific pairings with another base, resulting in mismatch.

The most frequent deamination is that of cytosine into uracil, with about 100 bases per day per cell undergoing this deamination. Uracil can then pair with adenine, so on the daughter strand, a guanine will be incorporated. After two cell cycles, there will be a 25% mutation.

- The spontaneous deamination of adenine produces hypoxanthine (replacing an AT with a GC after two replications if there is no repair).
- The deamination of guanine produces xanthine, which remains paired with cytosine.
- Thymine is the only base that does not undergo deamination.
- A mutation appears stably after 2 replications.

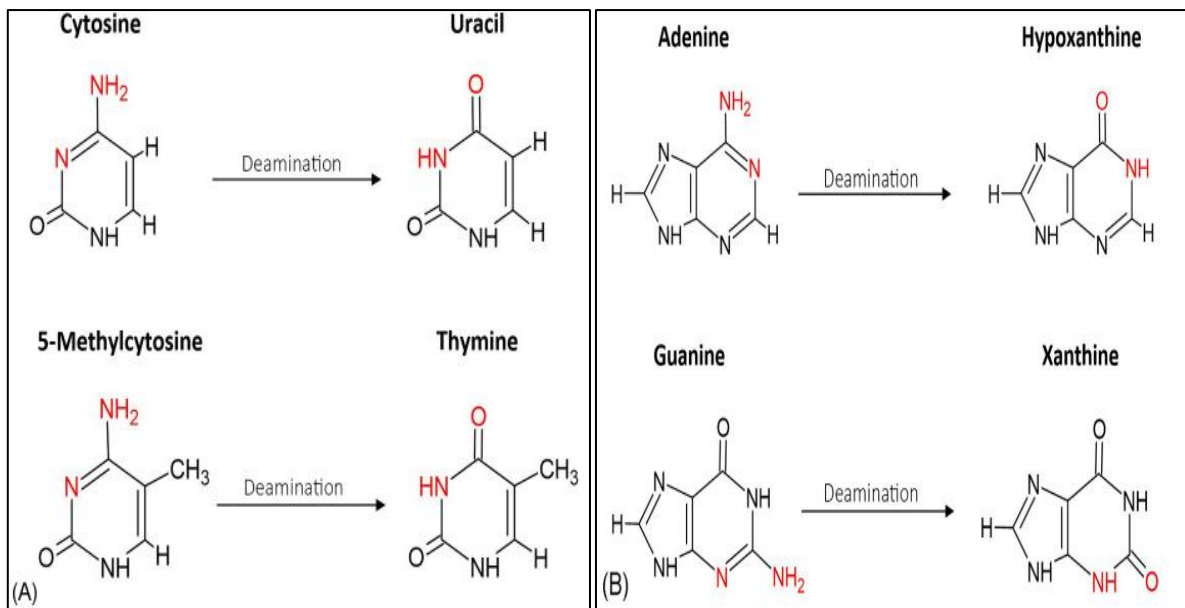


Figure 23. Examples of common DNA base modifications resulting from deamination:
(A) Deamination of pyrimidine bases and (B) deamination of purine bases.

C. Oxidative Damage:

Bases altered by oxidation: the production of highly reactive chemical molecules (**Reactive oxygen species (ROS)**): leads to the oxidation of DNA bases. This oxidation is often the cause of genetic diseases. It is the superoxide radicals ($O_2\cdot$), hydroxyl radicals ($OH\cdot$), and hydrogen peroxide ($H_2O_2\cdot$) that are normally generated by the aerobic system (there are many in the mitochondria, where the respiratory chain is found). Example: the oxidation of guanine into oxo-guanine allows pairing with adenine (GC becomes TA after two replications).

2.2.1.3. Transposable Elements

Transposable element (TE) movement, or "jumping," causes mutations by inserting, deleting, or rearranging DNA, acting as a major driver of genomic structural variation and evolution. These elements, which make up about half of the human genome, propagate by creating new copies, disrupting gene function, or altering gene regulation.

A. DNA transposons: "Cut and paste" mechanism (rare in mammals, active in bacteria)

B. Retrotransposons: "Copy and paste" via RNA intermediate

- **LINEs** (Long Interspersed Nuclear Elements): Autonomous, encode reverse transcriptase
- **SINEs** (Short Interspersed Nuclear Elements): Non-autonomous

TE Can insert into genes causing disruption or alter gene regulation as example below:

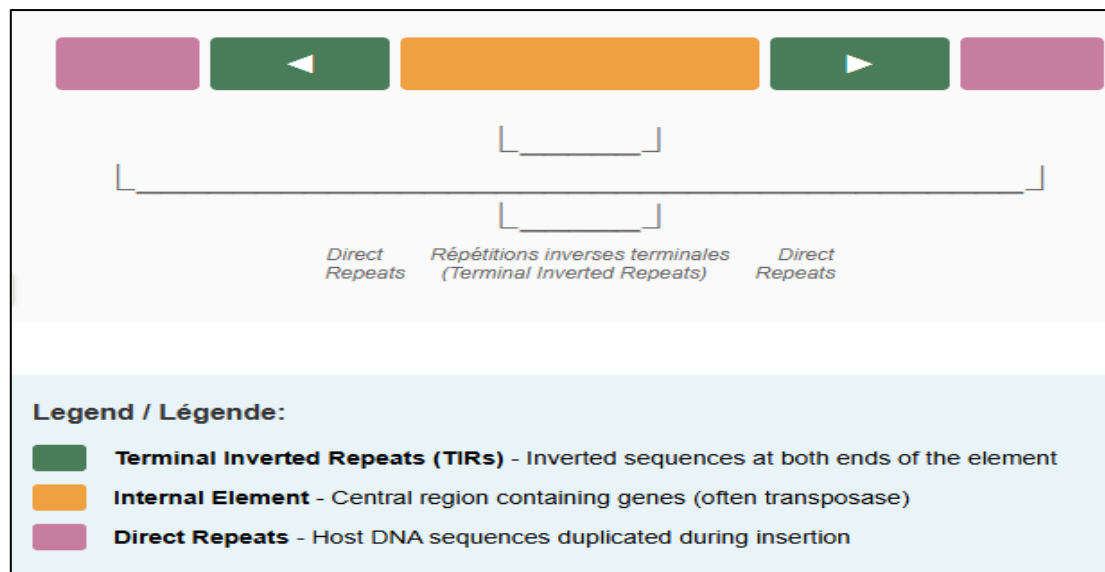


Figure 24. Schematic representation of transposon

Example: Effect of Retrotransposon on VvmybA1 Gene and Anthocyanin Production:

- In purple/black grapes, the VvmybA1 gene controls anthocyanin pigment synthesis.
- In white grapes, a retrotransposon integrated near the VvmybA1 gene and blocks anthocyanin synthesis
- In red grapes, a second mutation eliminated most of the retrotransposon but left a segment. Anthocyanin function is partially restored.

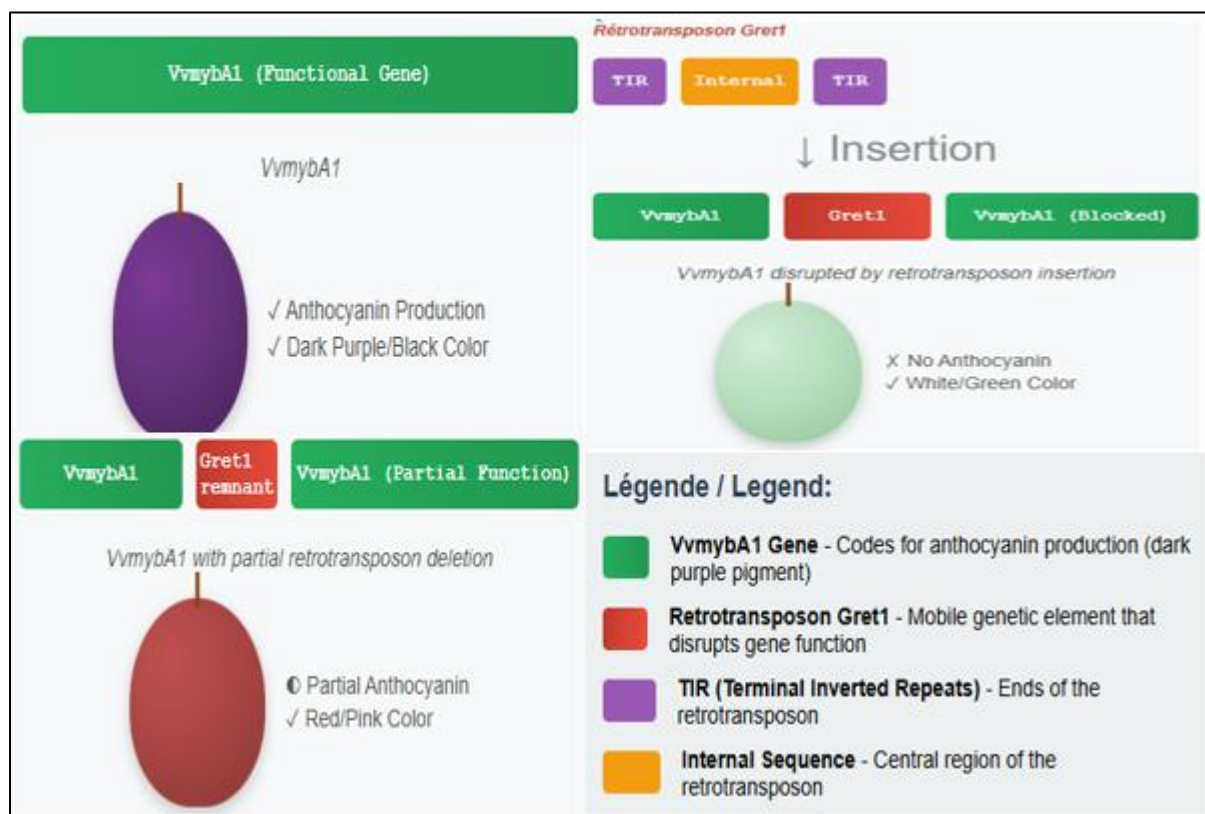


Figure 25. Effect of Retrotransposon on VvmybA1 Gene and Anthocyanin Production

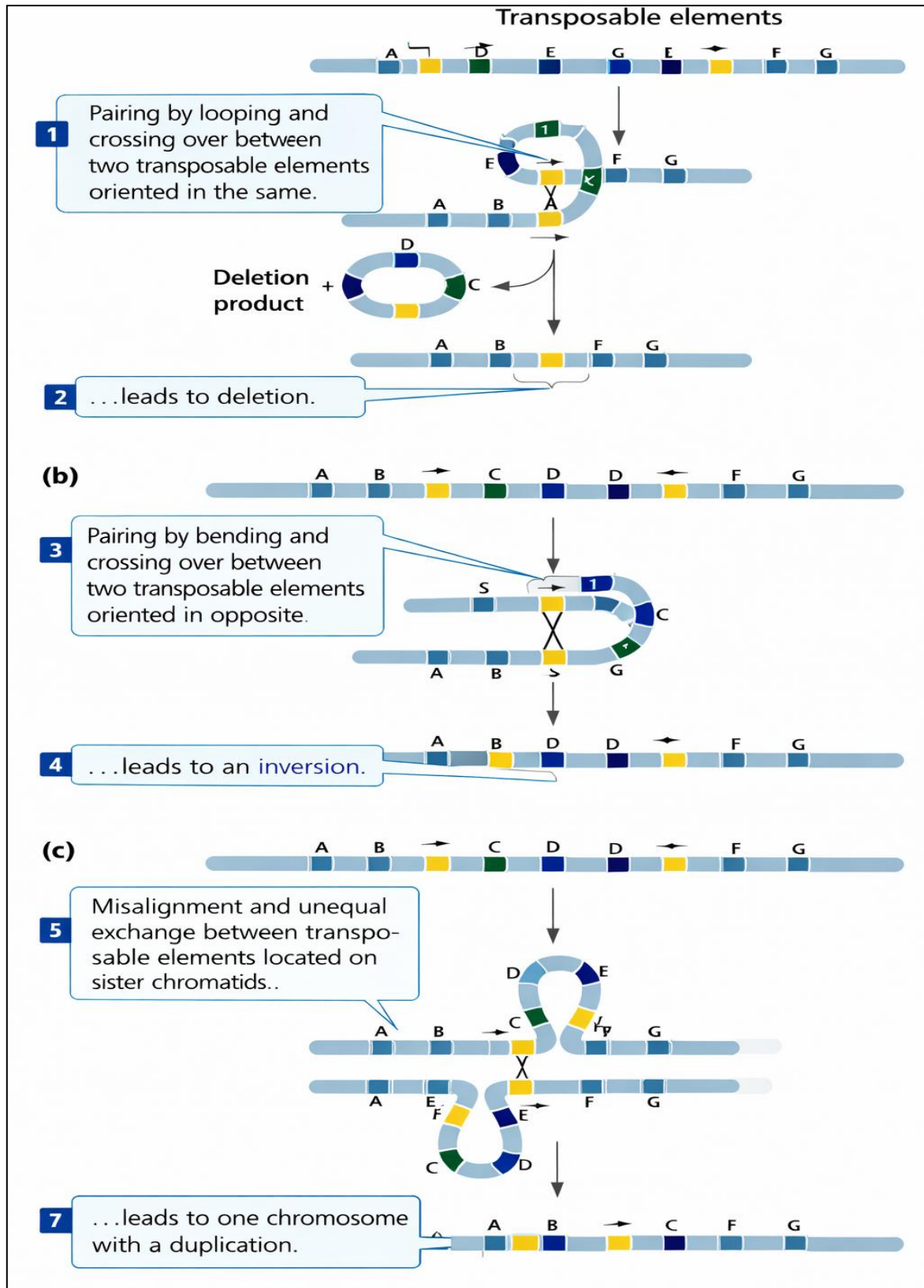


Figure 26. Mutations caused by transposable elements.

- (a) Pairing and crossing over between two TE oriented in the same direction leads to a deletion product. (b) Pairing and crossing over between two TE oriented in opposite directions leads to an inversion. (c) Misalignment and unequal exchange between TE on sister chromatids lead to one chromosome with a deletion and one chromosome with a duplication.

2.2.2. Induced Mutagenesis

2.2.2.1. Physical Mutagens

Mutations and lesions induced by physical mutagenic agents: are very significant lesions, which, in the absence of repair, lead to a blockage of replication, and therefore of cell division.

A. Ionizing Radiation

Damage from ionizing radiation can lead to strand breaks, affecting either a single strand (single-strand break) or both strands (double-strand break), to deletions, translocations, chromosome aberrations. Effects are dose-dependent and cumulative.

The mechanisms of action of damage caused by these radiations are multiple:

- **Direct action:** strand break followed by the creation of AP sites
 - **Indirect action:** involves the creation of oxidative agents that will cause oxidative damage.
- Ionizing radiation capable of inducing this type of damage includes cosmic rays, radioactivity, gamma rays and X-rays (used for therapeutic purposes: cancer radiotherapy uses this principle: the radiation is localized and causes irreparable breaks, which leads to the death of irradiated cells: they undergo apoptosis).

B. Non-Ionizing Radiation (UV light)

Lesions caused by UV light: Lead to the creation of covalent bonds between two adjacent pyrimidines on the same DNA strand: formation of a pyrimidine dimer. This causes DNA distortion, which blocks replication and transcription if there is no repair.

Solar UV rays are capable of causing this type of damage to skin cells, and if there is no repair, as in some predisposed individuals, skin cancer will develop: melanomas.

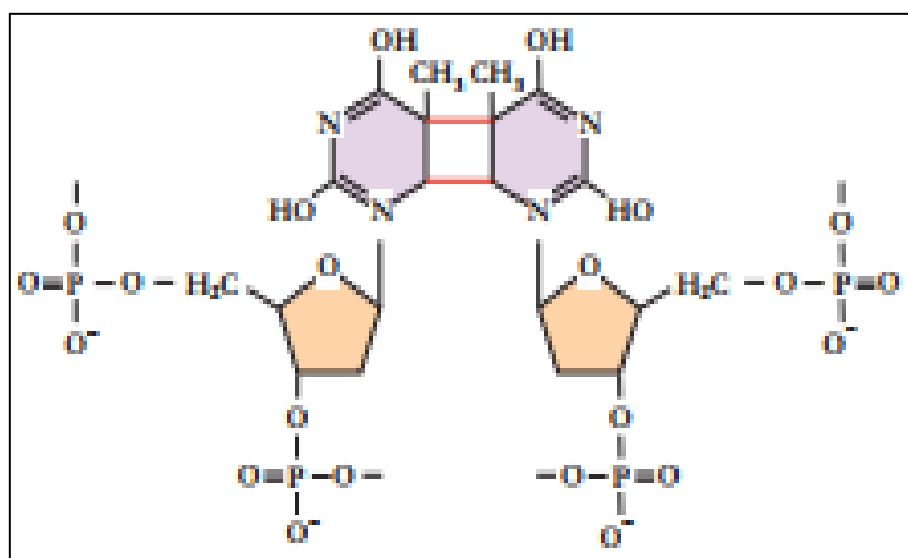


Figure 27. Pyrimidine dimer.

2.2.2.2. Chemical Mutagens

A. Base Analogs:

They are chemical compounds that resemble normal nitrogenous bases and can sometimes be incorporated in their place during replication on the daughter strand. They have different pairing properties, which leads to mutations after replication.

Examples:

- 5-Bromo-uracil (5BU): thymine analog, normally pairs with adenine and can tautomerize more easily → pairs with guanine. It causes A-T → G-C or G-C → A-T transitions
- 2-Aminopurine (2AP): adenine analog, pairs with thymine normally, but can pair with cytosine. It induces transitions

B. Alkylating Agents:

They alter the bases by adding ethyl or methyl groups to the oxygen atom of bases (guanine or thymine). The modified bases have different pairing properties.

Examples:

- **Ethyl methane-sulfonate (EMS)**: Common lab mutagen: alkylates guanine (O-6 position) → pairs with thymine instead of cytosine: G-C → A-T transitions
- **Nitrosoguanidine (NTG)**: Similar mechanism, very potent
- **Mustard gas**: Bifunctional alkylating agent: cross-links DNA strands and causes breaks and rearrangements

C. Deaminating Agents:

They deaminate bases

Examples:

- **Nitrous acid (HNO₂)**: Oxidatively deaminates bases: Cytosine → uracil (pairs with A) → C-G → T-A transition and Adenine → hypoxanthine (pairs with C) → A-T → G-C transition
- **Nitrosamines**: Found in tobacco, cured meats. Metabolically activated to produce alkylating agents

D. Intercalating Agents:

They slip between base pairs in DNA double helix. They insert between bases → distorts helix → causes insertions/deletions (frameshift mutations) during replication.

Examples:

- **Acridine dyes** (proflavin, acridine orange)
- **Ethidium bromide**: Common in molecular biology labs
- **Aflatoxin B1**: Fungal toxin, potent carcinogen

E. Oxidizing Agents:

They cause oxidative base damage (8-oxoG) and strand breaks

Examples:

- **Hydrogen peroxide (H₂O₂)**
- **Bleomycin:** Anticancer drug that generates free radicals

C. Bulky Adduct Formers (Distortion of DNA molecule)

They distort helix, block replication, cause mutations if bypassed

Examples:

- **Benzo[a]pyrene:** In tobacco smoke, forms large DNA adducts
- **Aflatoxin B1:** Metabolically activated, binds guanine

2.2.2.3. Biological Mutagens:

A. Viruses:

They can integrate into host genome. Their insertion can disrupt genes or activate oncogenes

Examples:

- Retroviruses (HIV, HTLV)
- HPV (human papillomavirus) → cervical cancer

B. Transposons:

They already covered under spontaneous

2.3. Genotypic diagnosis

2.3.1. Principles of Mutation Detection

Genotypic diagnosis aims to identify specific genetic variants associated with disease, predisposition, or traits. Methods vary based on:

- Known vs. unknown mutations
- Single gene vs. genome-wide analysis
- Detection sensitivity required
- Sample availability and quality

2.3.2. Methods for Detecting Known Mutations

2.3.2.1. PCR-Based Methods

➤ Polymerase chain reaction (PCR)

PCR is a technique that allows the *in vitro* enzymatic amplification of DNA sequences by repeated elongation reactions in the presence of specific nucleotide primers and a DNA polymerase. This method was developed in 1985 by Kary Mullis, who was awarded the Nobel Prize in Chemistry in 1993.

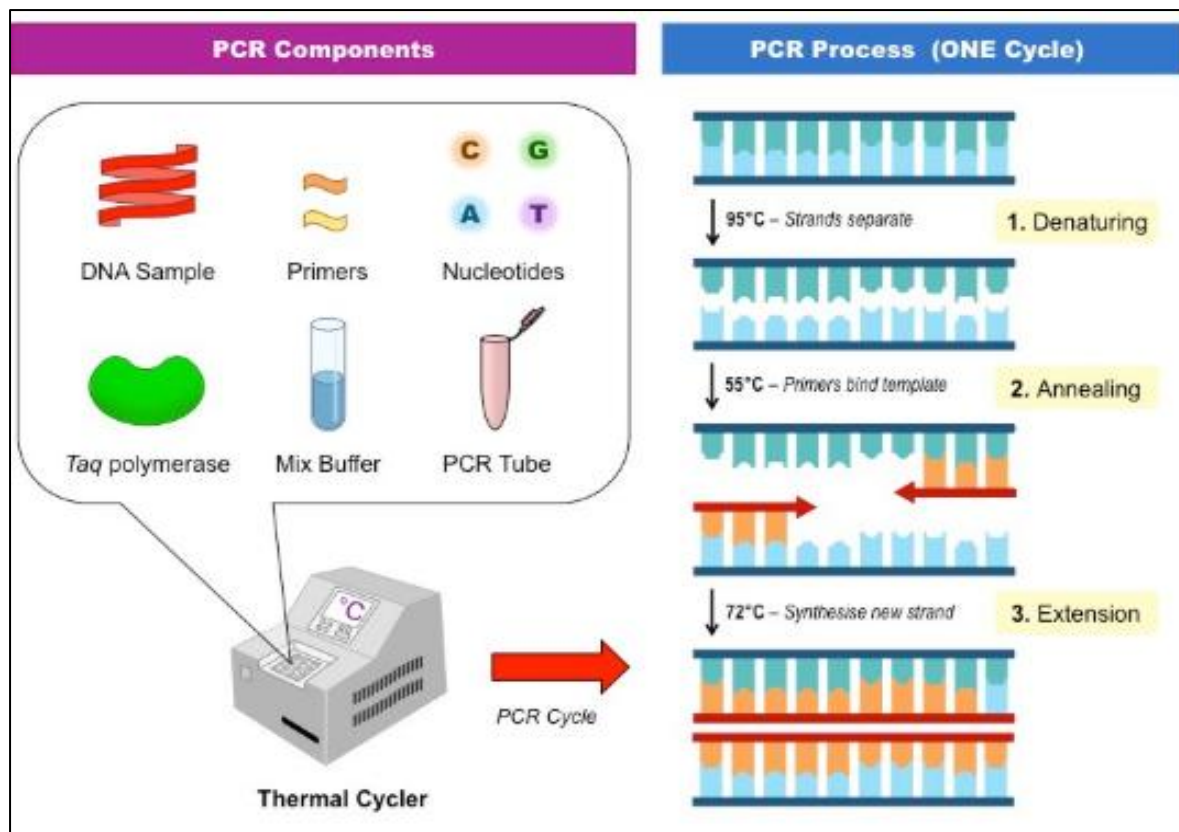


Figure 28. PCR components and process.

A. Allele-Specific PCR (AS-PCR)

- Uses primers specific to wild-type or mutant allele
- Only amplifies if exact match present
- Simple, fast, inexpensive for known point mutations

B. Amplification Refractory Mutation System (ARMS)

- Variant of AS-PCR
- Primer 3' end matches variant nucleotide
- Two reactions: one for normal, one for mutant allele
- Widely used for SNP genotyping

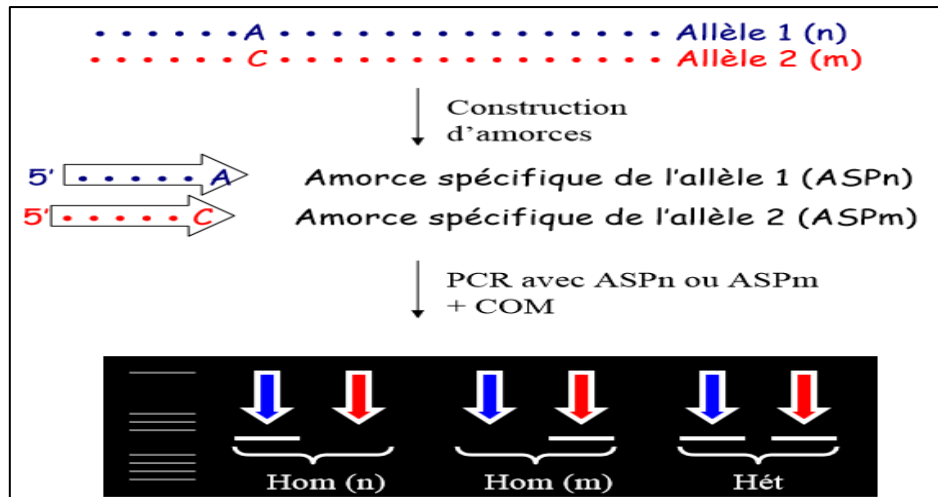


Figure 29. Amplification Refractory Mutation System

C. Real-Time PCR with Probes

The principle of Real-Time PCR is based on the ability to monitor the amount of DNA present in the reaction at any moment with quantitative and specific fluorescent probes (their hybridization and cleavage generates fluorescent signal), rather than at the end of the PCR (end-point PCR). Example: TaqMan probes and

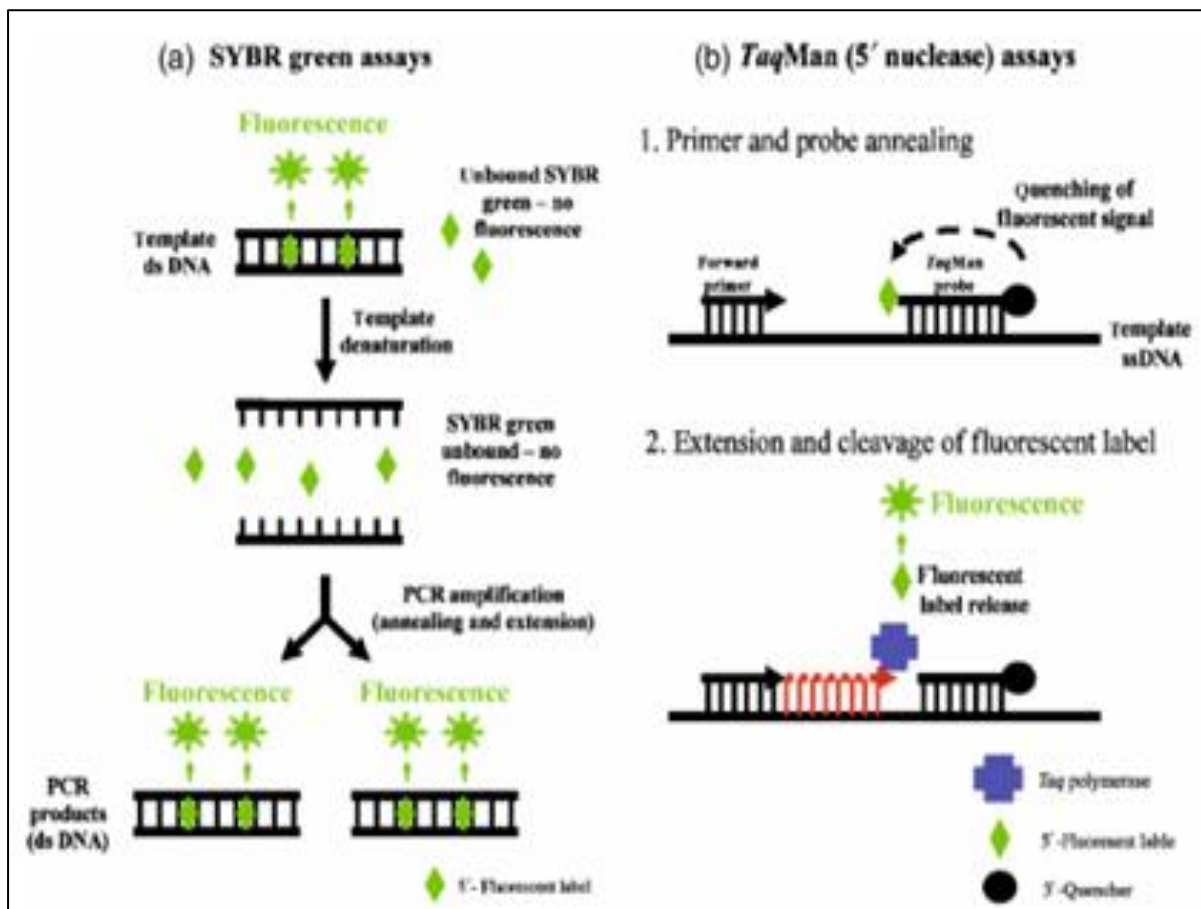


Figure 30. Real time PCR

D. Digital PCR (dPCR)

- Partitions sample into thousands of individual reactions
- Absolute quantification of mutant vs. wild-type alleles
- Extremely sensitive for detecting rare mutations

2.3.2.2. Restriction Fragment Length Polymorphism (RFLP)

Principle: mutation creates or destroys restriction enzyme recognition site

Process:

1. PCR amplify region containing mutation
2. Digest with appropriate restriction enzyme
3. Analyze fragments by gel electrophoresis

Advantages: Simple, inexpensive

Disadvantages: Requires convenient restriction site, labour-intensive

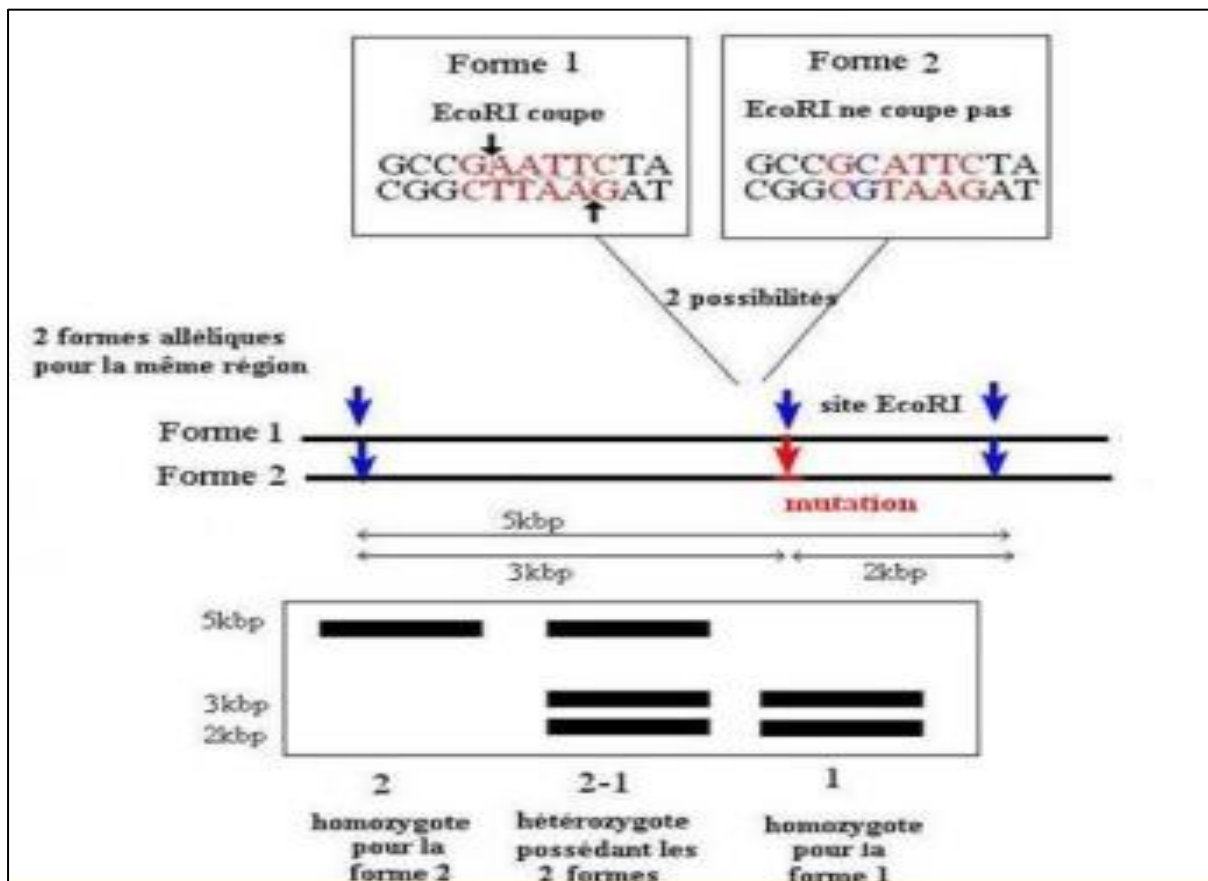


Figure 31. RFLP

2.3.2.3. Hybridization-Based Methods

A. Allele-Specific Oligonucleotide (ASO) Hybridization

- Short probes complementary to normal or mutant sequence
- Dot blot or reverse dot blot format

- Stringent conditions distinguish single-base differences
- Can screen multiple mutations simultaneously

B. Microarrays (SNP Chips)

- Thousands to millions of probes on solid support
- Hybridize labelled genomic DNA
- Detects known SNPs across genome
- Examples: Illumina arrays, Affymetrix chips

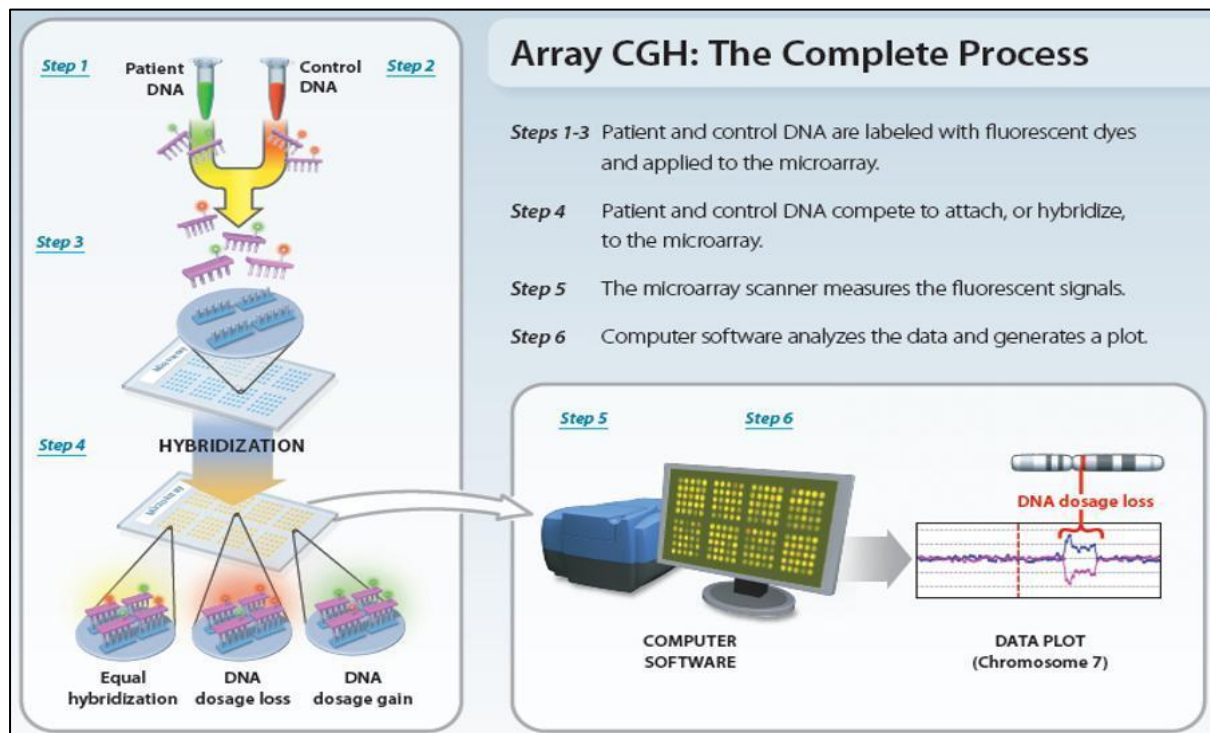


Figure 32. Mico-array CGH

2.3.3. Methods for Scanning Unknown Mutations

2.3.3.1. Sanger Sequencing

The gold standard for decades, still widely used:

Process :

1. PCR amplify target region
2. Chain-termination sequencing with fluorescent di-deoxynucleotides
3. Capillary electrophoresis separates fragments
4. Chromatogram shows exact sequence

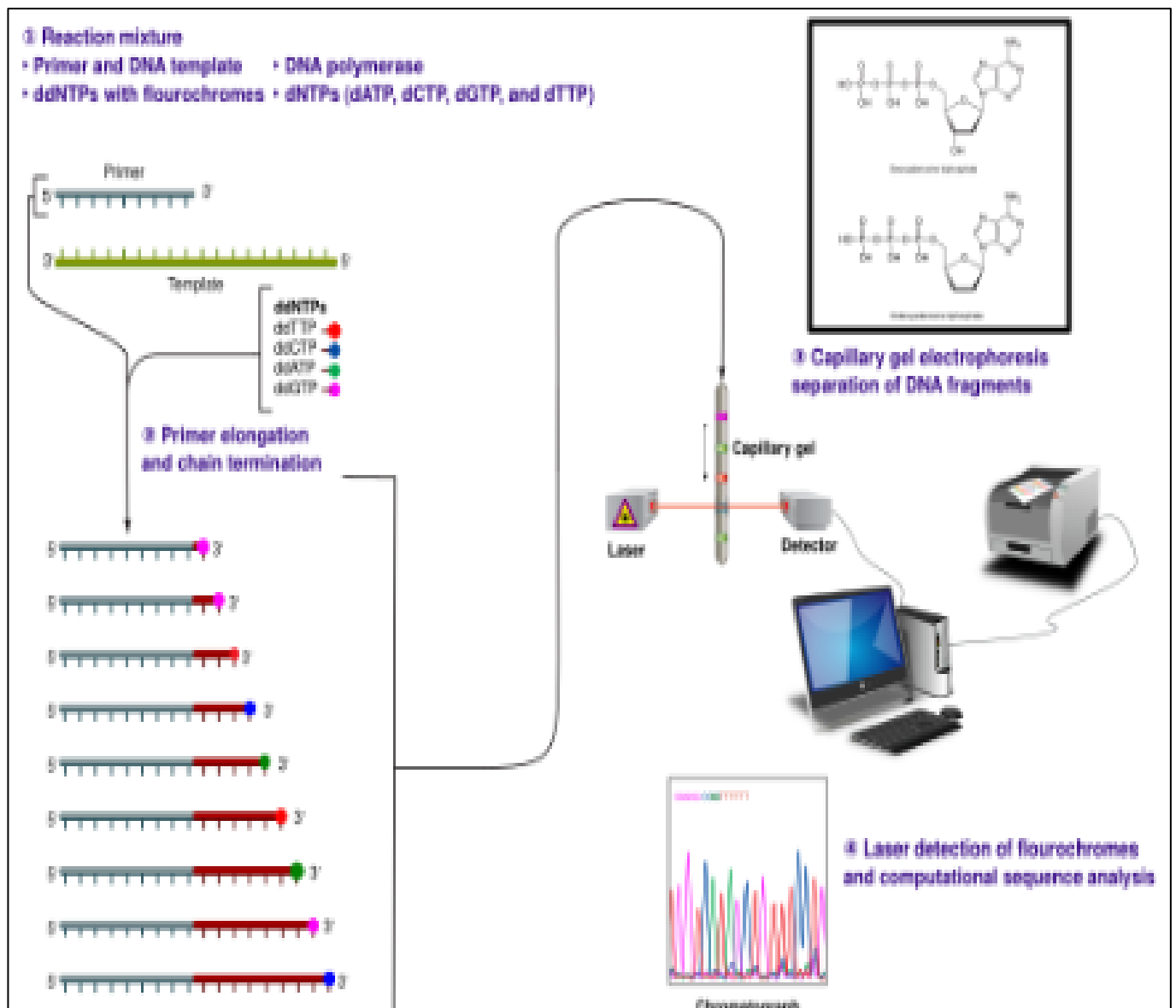


Figure 33. Sanger sequencing.

Advantages:

- High accuracy (~99.9%)
- Reads ~800-1000 bases
- Detects all mutation types in sequenced region

Disadvantages:

- Labour-intensive for multiple genes
- Expensive for large-scale screening

Applications: Confirmatory testing, small gene panels, validating NGS results

2.3.3.2. Next-Generation Sequencing (NGS)

Revolutionary technology enabling massive parallel sequencing:

Platforms: Illumina (most common), Ion Torrent, PacBio, Oxford Nanopore

Advantages:

- Massively parallel (millions of reads simultaneously)
- Neutral variant detection
- Decreasing costs (human genome ~\$600-1000 now)
- Detects all variant types

Disadvantages:

- Complex data analysis (bioinformatics expertise required)
- Variants of uncertain significance (VUS) interpretation
- Ethical issues (incidental findings)
- May miss some structural variants

Approaches:

A. ATargeted Gene Panels

- Sequence specific set of genes (10s to 100s)
- Example: Cancer panels (BRCA1/2, TP53, etc.)
- Cost-effective for specific conditions

B. Whole Exome Sequencing (WES)

- Sequences all protein-coding regions (~1-2% of genome)
- ~20,000 genes, ~30 million bases
- Identifies most disease-causing variants (85% in exons)
- More affordable than whole genome

C. Whole Genome Sequencing (WGS)

- Sequences entire genome (~3 billion bases)
- Detects variants in non-coding regulatory regions
- Identifies structural variants better than WES
- Most comprehensive but most expensive

D. RNA Sequencing (RNA-seq)

- Sequences transcriptome
- Detects expression changes, splice variants, fusion genes
- Useful in cancer diagnosis

2.3.3.3. Heteroduplex Analysis Methods

Detect mismatches between wild-type and mutant DNA strands without knowing exact mutation:

A. Single-Strand Conformation Polymorphism (SSCP)

- Denature PCR product → single strands

- Each sequence adopts unique 3D structure
- Different mobility on non-denaturing gel
- Detects ~80-95% of mutations
- Doesn't identify exact mutation (needs sequencing)

B. Denaturing Gradient Gel Electrophoresis (DGGE)

- PCR products migrate through increasing denaturant gradient
- Heteroduplexes (mismatches) denature at different positions than homoduplexes
- Sensitive but technically demanding

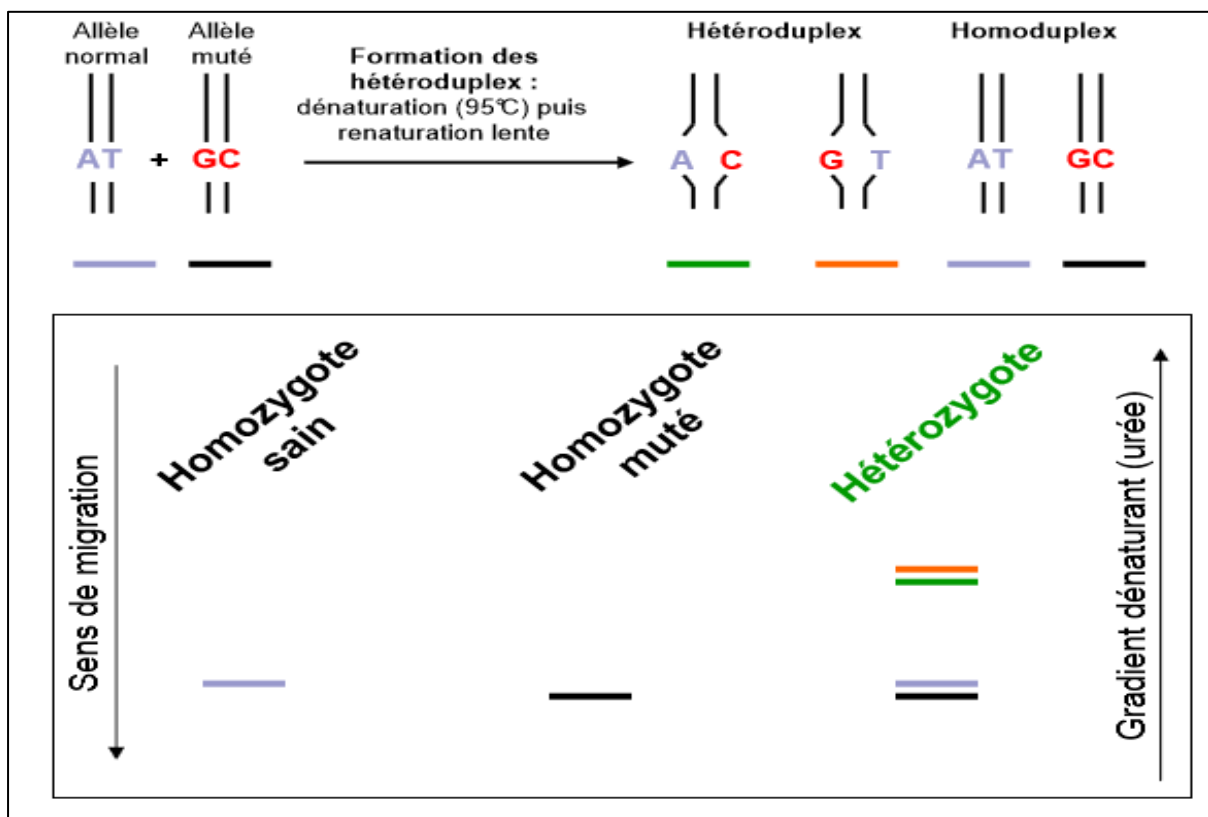


Figure 34. DGGE

C. High-Resolution Melting (HRM)

- Monitor PCR product melting in real-time
- Different sequences have different melting curves
- Distinguishes homozygotes, heterozygotes, wild-type
- Fast, closed-tube (no post-PCR handling)

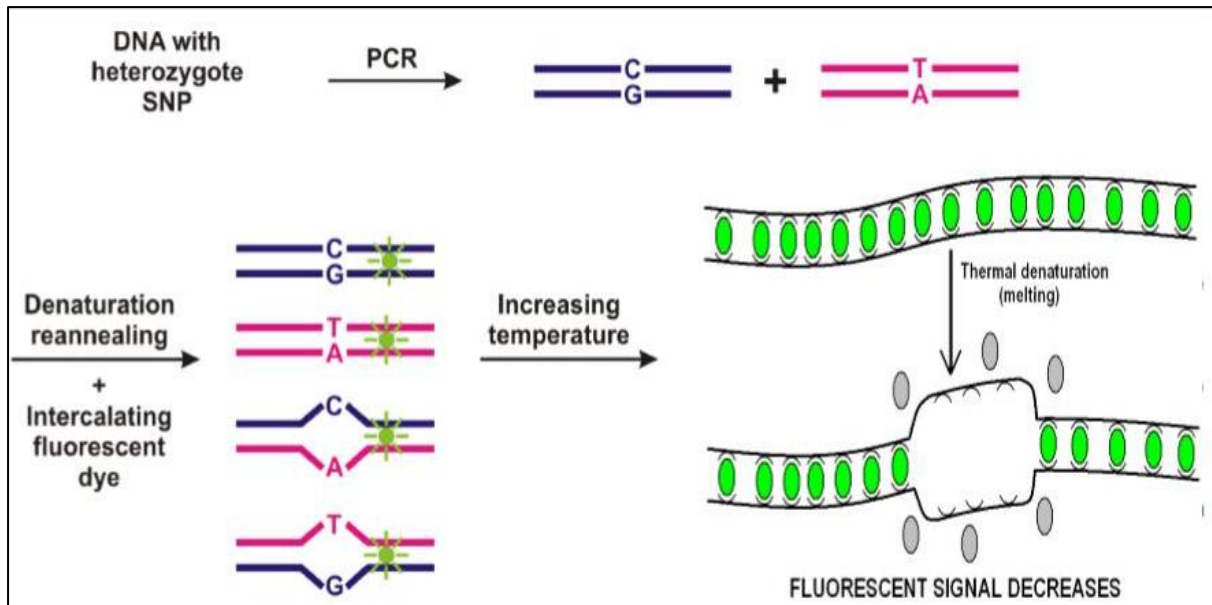


Figure 35. HRM

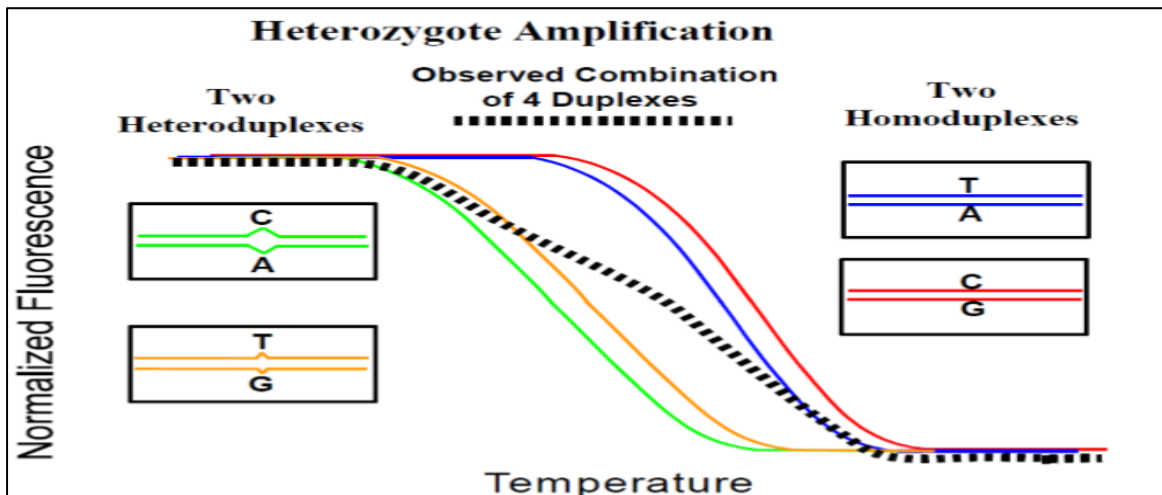


Figure 36. Melting curve of HRM

2.3.3.4. Enzymatic Mismatch Cleavage

T7 Endonuclease I, CEL I/II nucleases

- Enzymes that cleave mismatched DNA
- Hybridize wild-type and test DNA → heteroduplexes if mutation present
- Add enzyme → cleaves at mismatch
- Analyze cleavage products

2.3.4. Detection of Specific Mutation Types

2.3.4.1. Repeat Expansion Detection

Many genetic diseases caused by trinucleotide or other repeat expansions:

Diseases:

- Huntington disease (CAG repeats in HTT)

- Fragile X syndrome (CGG repeats in FMR1)
- Myotonic dystrophy (CTG repeats)
- Friedreich ataxia (GAA repeats)

Methods:

PCR: For small expansions

- Can amplify and size by gel electrophoresis
- Limited by PCR difficulty with long repeats

Southern Blot: For large expansions

- Digests genomic DNA with restriction enzymes
- Probe hybridizes to region
- Determines fragment size (correlates with repeat number)
- Labour-intensive but definitive

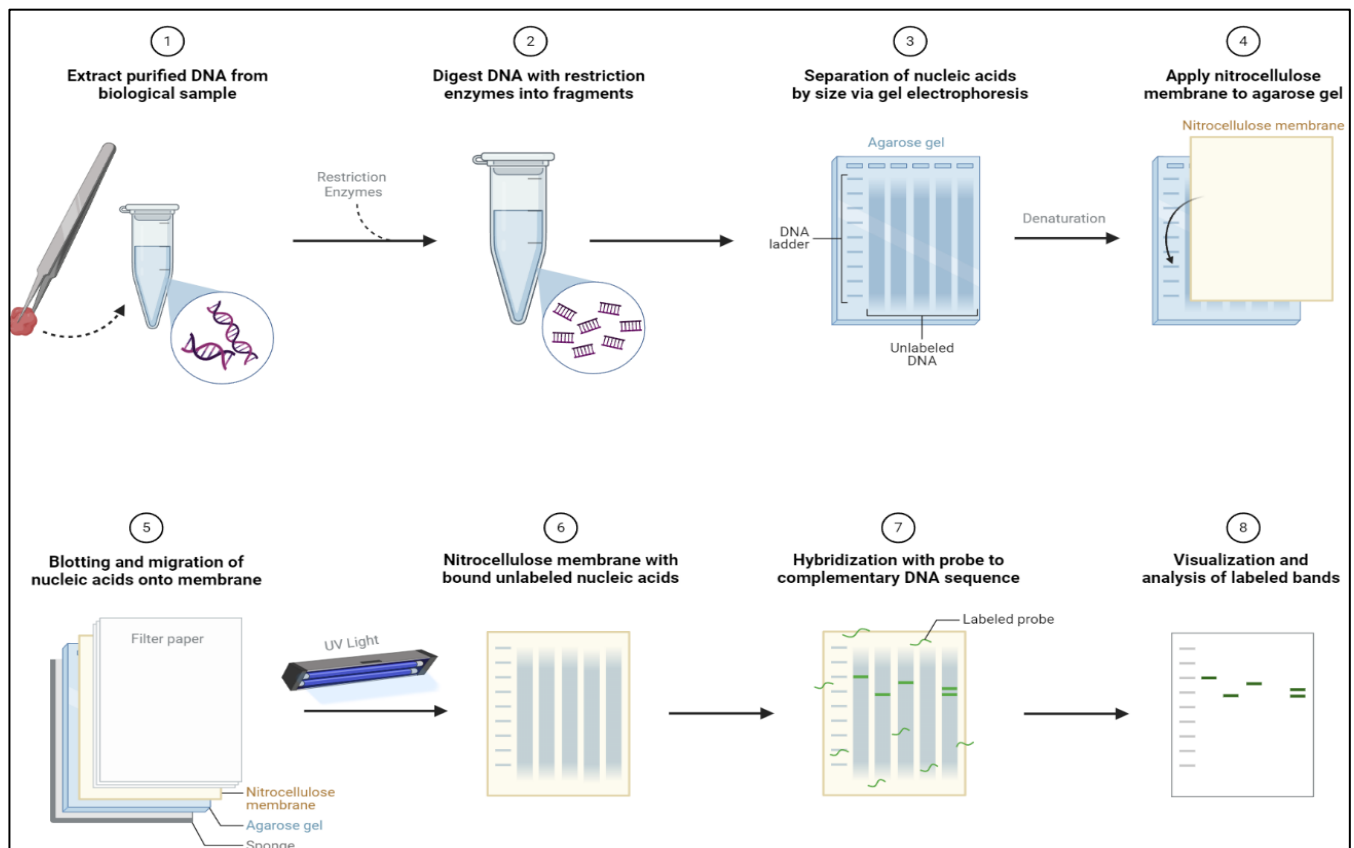


Figure 37. Southern blot.

2.3.4.2. Large Deletion/Duplication Detection

A. Chromosomal Microarray Analysis (CMA)

- Array CGH (comparative genomic hybridization)
- SNP (single nucleotide polymorphism) arrays
- Detects sub-microscopic deletions/duplications genome-wide

- First-tier test for developmental delay/autism

B. Quantitative PCR (qPCR)

- Real-time PCR comparing target gene to reference
- Determines copy number

2.3.4.3. Chromosomal Abnormality Detection

A. Karyotyping

- Stain and photograph chromosomes during metaphase
- Identifies large structural abnormalities, aneuploidy
- Resolution ~5-10 Mb
- Classic method, still used

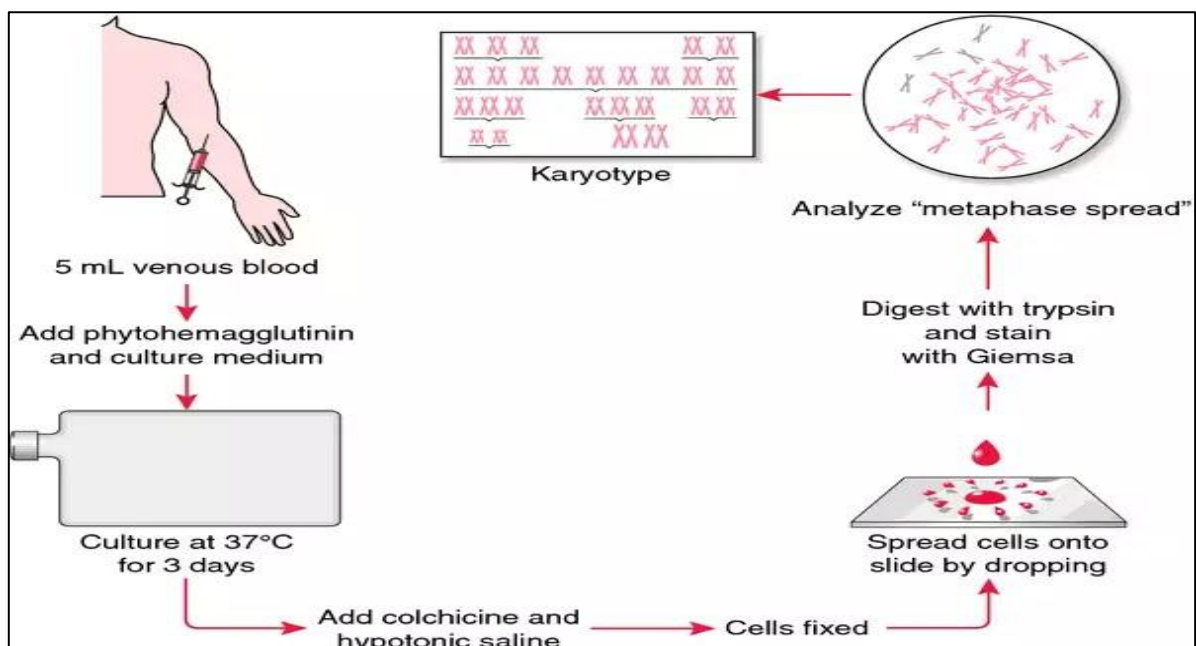


Figure 38. Preparation of karyotype

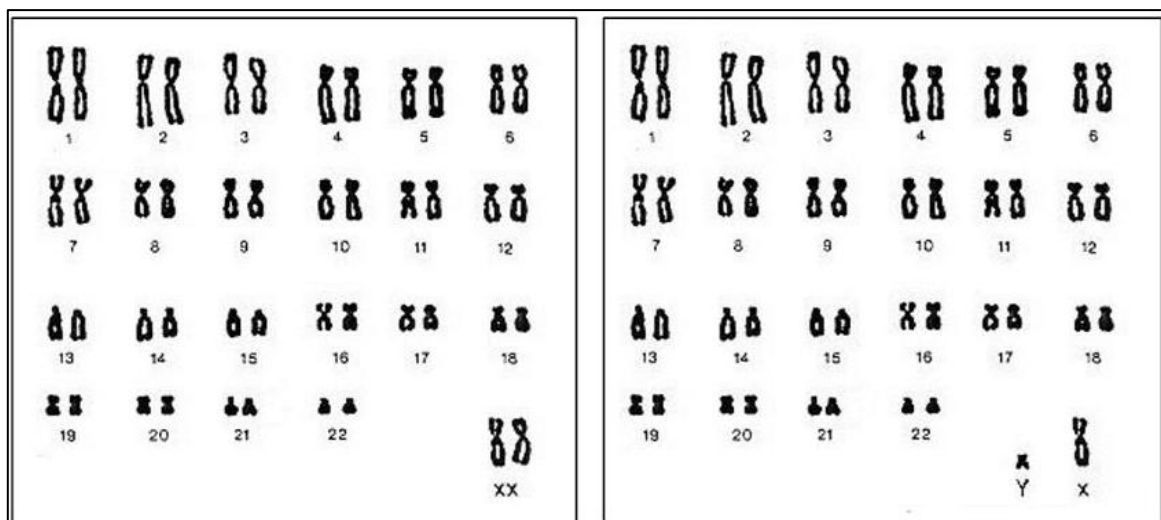


Figure 39. Human karyotype

B. Fluorescence In Situ Hybridization (FISH)

- Fluorescent probes hybridize to specific chromosomal regions
- Visualize under microscope
- Detects deletions, duplications, translocations
- Rapid prenatal aneuploidy testing

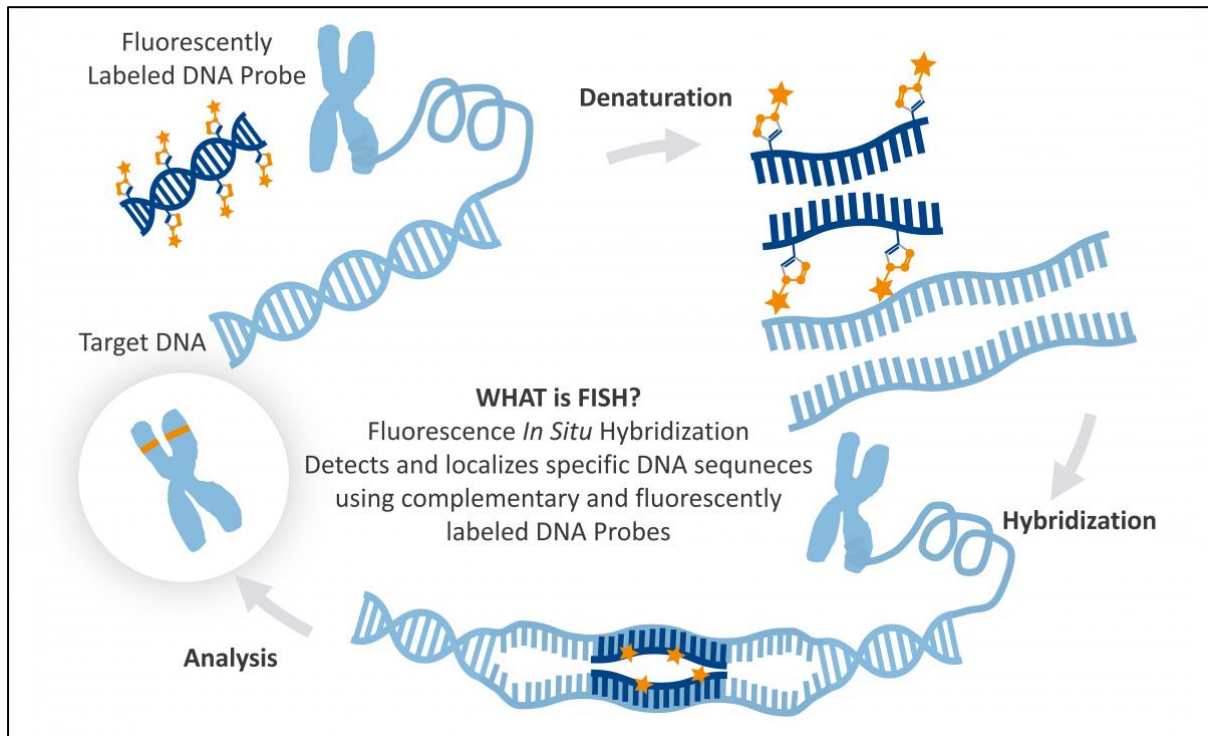


Figure 40. Preparation of FISH test

C. Spectral Karyotyping (SKY)

- Each chromosome painted different colour
- Identifies complex rearrangements

D. Chromosomal Microarray: See above

2.4. Tutorial (Annex)

3. Transmission and Conservation of Genetic Information

DNA replication is a critically important mechanism that allows the duplication of the genetic heritage to obtain 2 copies before division, each transmitted to one of the two daughter cells during cell division. DNA replication is:

- coordinated: synchronized with cell division: occurs before cell division, during the S phase of interphase.
- complete: entire genome replicated once and only once per cell cycle
- It is a complex and ordered phenomenon under the control of several enzymes and proteins, including DNA polymerases.
- must be as faithful as possible, so that the genetic heritage transmitted to daughter cells is as perfectly identical to that of the mother cell as possible. It is a semi-conservative cellular mechanism.
- accurate: error rate < 1 in 10^9 bases after proofreading and repair. It is an essential event where errors can occur, which is why there are control mechanisms complemented by DNA repair mechanisms.

3.1. DNA Replication and its regulation

3.1.1. The semi-conservative model

Meselson-Stahl experiment (1958):

E. coli bacteria were cultured for 14 generations in a medium containing ammonium chloride radioactive $^{15}\text{NH}_4\text{Cl}$ (heavy nitrogen ^{15}N) as the sole nitrogen source. Bacteria were then transferred to a new medium containing only light ^{14}N and collected after one population doubling (20 min).

Heavy and light DNA can be separated by CsCl density gradient centrifugation — two bands form: one for ^{15}N -DNA and one for ^{14}N -DNA.

After one generation: the isolated DNA band appeared exactly between the two bands, indicating hybrid double helices (one ^{14}N strand + one parental ^{15}N strand).

After a second division in ^{14}N medium (40 min): two bands appeared — one with the density of light DNA and one with the density of hybrid DNA, confirming semi-conservative replication.

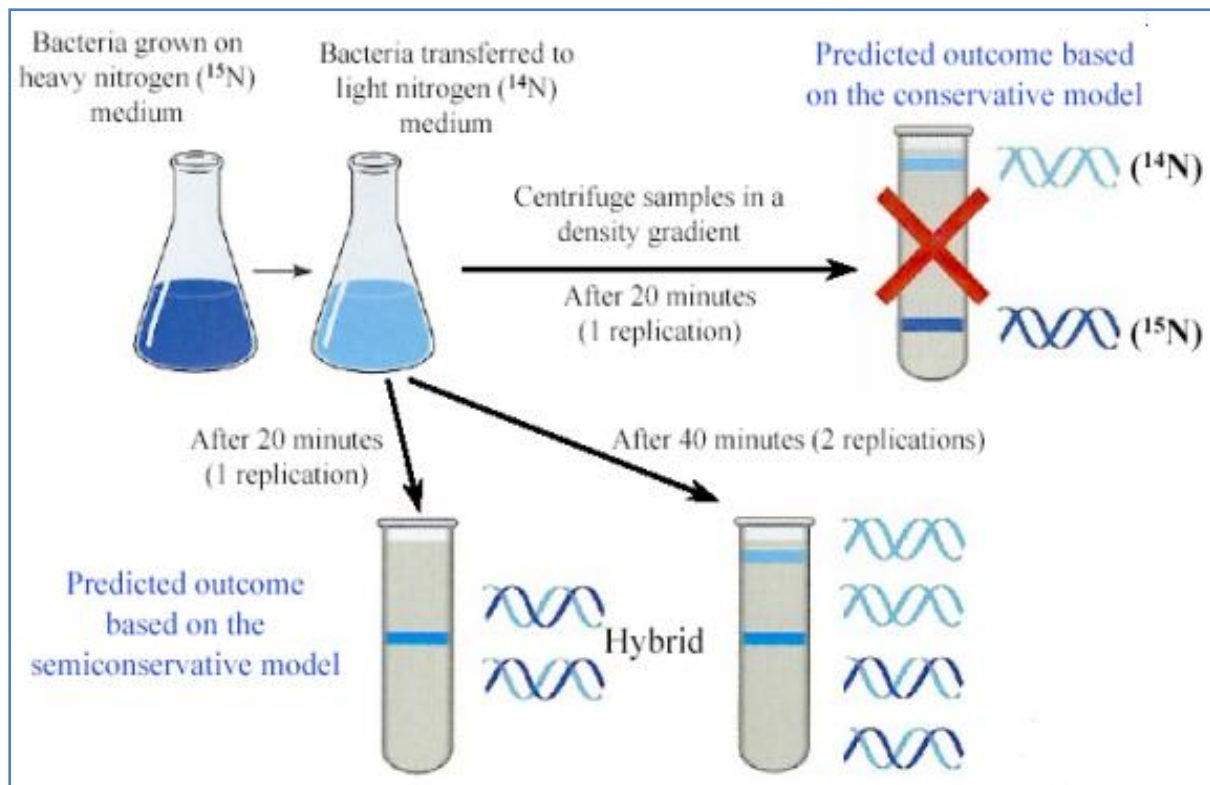


Figure 41. Meselson and Stahl Experiment (1958).

It demonstrates the semi-conservative nature of replication. Separation of labeled DNA by density using CsCl centrifugation, with tracking of heavy DNA proportion over time. Result: After one generation, DNA appeared at an intermediate density — confirming that each daughter molecule contains one parental strand and one newly synthesized strand.

Then: this experiment proved DNA replication is semi-conservative: each daughter molecule contains one original (parental) strand, serve as template, and one newly synthesized strand. Complementary base pairing ensures accurate copying. Alternative models (conservative, dispersive) were disproven.

3.1.2. Characteristics of replication

- **Semi-conservative:** each newly synthesized DNA molecule has one parental strand and one newly formed strand.
- **Oriented:** each parental strand serves as template for the synthesis of a complementary antiparallel strand (daughter strand).
- **Bidirectional:** both DNA strands are unwound in opposite directions and serve as templates.
- **Unidirectional polymerization:** always proceeds in the $5' \rightarrow 3'$ direction. A phosphodiester bond forms between the 3'-OH end of the growing strand and the 5'-phosphate of the added nucleotide.

- **Semi-discontinuous replication:** at a replication fork, both daughter strands are synthesized simultaneously. The leading strand is read in the direction of the fork; the lagging strand is read in the opposite direction and synthesized discontinuously in Okazaki fragments.

3.1.3. Molecular tools of replication:

- **Parental DNA:** each of the two strands serves as a template to synthesize the antiparallel strand.
- **dNTPs (dATP, dCTP, dTTP, dGTP):** enter the reaction as triphosphates; incorporated as monophosphates — pyrophosphate release provides the energy for replication.
- **Mg²⁺ ions:** stabilize dNTPs by protecting them from enzymatic hydrolysis; also important for DNA polymerase activity.
- **DNA polymerases:** responsible for nucleotide polymerization. They move along single-stranded DNA and add nucleotides to the 3' end. Fidelity is very high (~1 error per 1,000,000 nucleotides), corrected by 3'→5' exonuclease activity.
- **Primases:** RNA polymerases capable of synthesizing RNA primers.
- **DnaA proteins:** recognize the replication origin (oriC) and open the double helix.
- **Helicases (DnaB):** enzymes that use ATP hydrolysis energy to break hydrogen bonds between the two DNA strands.
- **SSB proteins (Single Strand Binding) in prokaryotic cell and RPA (Replication Protein A) in eukaryotic cell:** stabilize the open conformation of the double helix.
- **DnaC:** Allows helicase and primase to bind to SSB-coated DNA (primosome assembly).
- **Topoisomerases:** relieve supercoiling upstream of the replication fork and release tension in the presence of ATP.
- **DNA ligases:** enzymes capable of reconstituting the phosphodiester bond between the 3'-OH carbon and the 5'-phosphate of two adjacent nucleotides on a DNA strand. They join DNA strands or Okazaki fragments synthesized by DNA polymerases. It's ATP dependent.

3.1.4. Molecular Mechanism of DNA Replication

3.1.4.1. Key Enzymes and Proteins

A. DNA Polymerases: catalyze phosphodiester bond formation

- Add nucleotides to 3'-OH group (5'→3' synthesis direction): require primer with 3'-OH group (cannot start de novo)

- Use dNTPs (dATP, dGTP, dCTP, dTTP) as substrates
- Energy from breaking high-energy phosphate bonds
- Fidelity mechanisms (base selection, proofreading)
 - **Prokaryotic DNA Polymerases (*E. coli*)**
 - ✓ **DNA Polymerase III:** Primary replicative enzyme
- Synthesizes both leading and lagging strands
- 5'→3' polymerase activity
- 3'→5' exonuclease activity (proofreading)
- Part of replisome complex
- Highly processive (~500,000 nucleotides without dissociating)
 - ✓ **DNA Polymerase I:** "Clean-up" enzyme
- 5'→3' polymerase activity
- 3'→5' exonuclease (proofreading)
- **5'→3' exonuclease** (unique feature): Removes RNA primers
- Nick translation: Simultaneously removes primers and fills gaps
- Less processive (~20-100 nucleotides)
 - ✓ **DNA Polymerase II:** DNA repair, restart stalled replication forks
- **Eukaryotic DNA Polymerases:** more complex, with specialized roles:
 - ✓ **DNA Polymerase α :** Primase-polymerase complex
- Associated with primase enzyme
- Synthesizes RNA-DNA primer (~30 nucleotides: ~10 RNA + ~20 DNA)
- No proofreading activity
- Initiates each Okazaki fragment
 - ✓ **DNA Polymerase δ :** Lagging strand synthesis
- High processivity
- 3'→5' exonuclease (proofreading)
- Extends from Polymerase α primers
- Also involved in DNA repair
 - ✓ **DNA Polymerase ϵ :** Leading strand synthesis
- High processivity
- 3'→5' exonuclease (proofreading)
- May also synthesize lagging strand

Other polymerases:

- ✓ **Polymerase γ :** Mitochondrial DNA replication

- ✓ **Polymerase β :** Base excision repair
- ✓ **Polymerase η , ι , κ , etc.:** Translesion synthesis (bypass damaged bases)

B. Topoisomerases: Relieve tension from unwinding: unwinding creates positive supercoiling ahead of fork, that would halt replication if not relieved

➤ **Type I Topoisomerases:**

- Make transient single-strand break
- Allow rotation to relieve tension
- Reseal break
- No ATP required

➤ **Type II Topoisomerases:**

- Make transient double-strand break
- Pass another DNA segment through
- Reseal break
- Require ATP

✓ **Prokaryotes:** DNA gyrase (introduces negative supercoils), Topoisomerase IV.

✓ **Eukaryotes:** Topoisomerase II

3.1.4.2. Mods of replication:

For the **circular DNA** as that's of bacterial chromosomal and plasmids, two modes are known:

A. Theta (θ) replication: most common plasmid replication mode

- DNA begins to unwind at the origin of replication ori (a single origin in plasmid), exposing single strands that serve as templates, generating a replication bubble.
- The bubble expands progressively as both ends unwind simultaneously.
- The two replication forks eventually meet, aided by topoisomerase IV.
- Result: two identical circular DNA molecules.

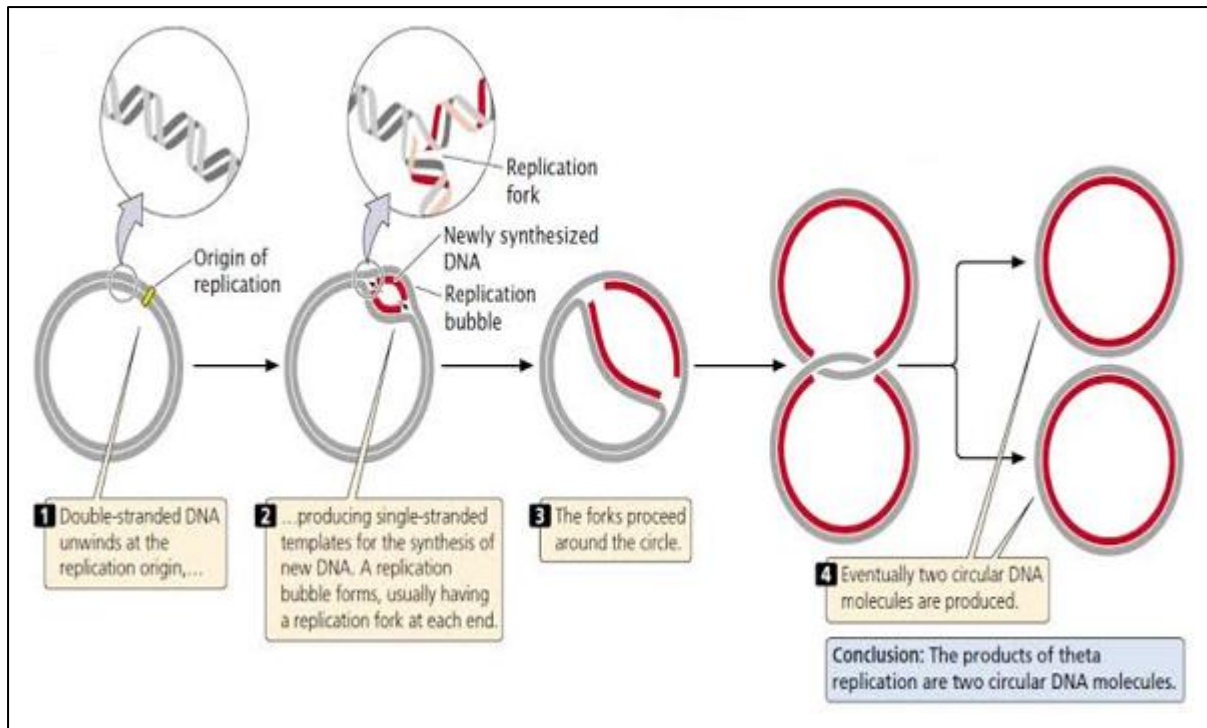


Figure 42. Theta replication form

B. Rolling circle replication: is an asymmetric replication process

- Nick introduction in one strand at ori:
- 3'-OH terminus serves as primer
- Displacement synthesis of the intact strand
- Rolling of the circular template
- Complementary strand synthesis (often later)
- Circularization of linear product

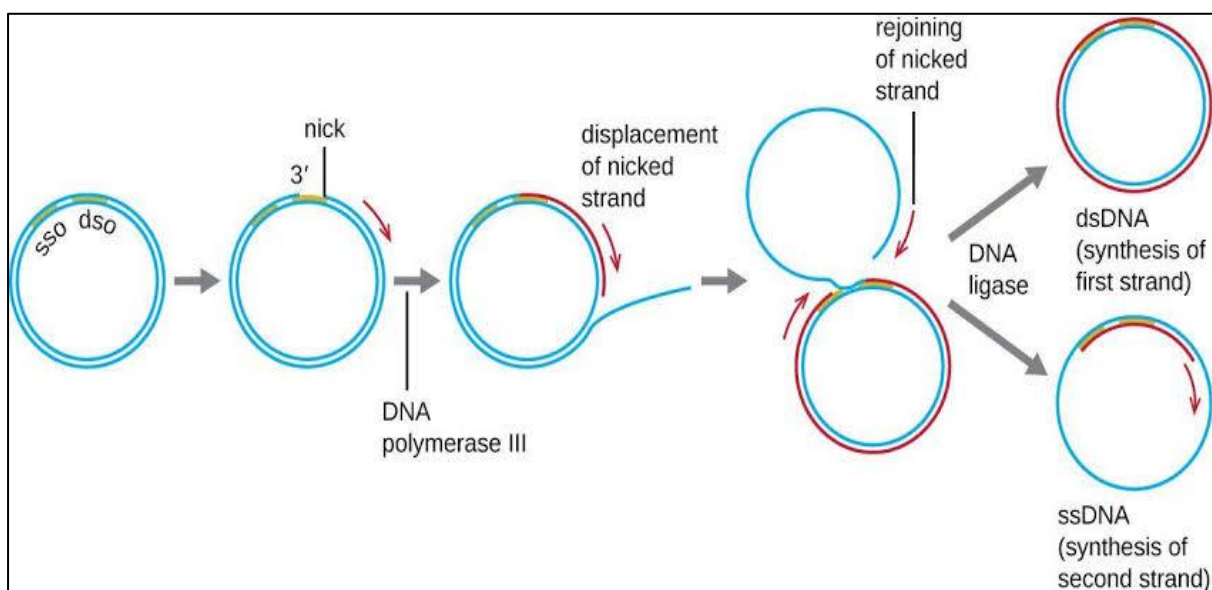


Figure 43. Rolling circle replication

C. Linear eukaryotic chromosome replication

For the eukaryotic chromosome replication it possess in multiple origins:

- Replication proceeds simultaneously from many origins.
- At each origin, DNA is unwound and a replication bubble forms with two diverging forks.
- Forks from adjacent replicons meet and replicated DNAs fuse.
- Fusion of all replicons produces two identical DNA molecules.

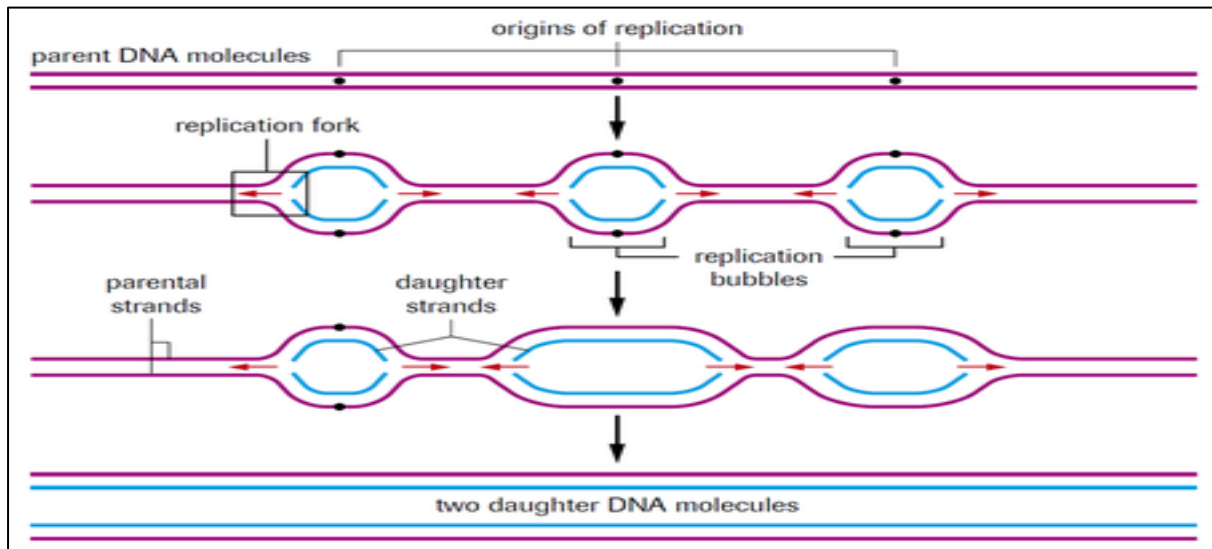


Figure 44. Eukaryotic chromosome replication

3.1.4.3. Mechanism of DNA replication:

Three stages are distinguished during replication, in both prokaryotes and eukaryotes:

- Initiation
- Elongation of polynucleotide chains
- Termination

A. DNA Replication in prokaryotes

It has been studied in detail in *E. coli*. Replication occurs in three stages, coordinated on both strands:

➤ Initiation

The *E. coli* chromosome contains a single origin of replication (*oriC*). The minimal sequence required for *oriC* function is 245 bp and contains several essential sites.

- An initiation protein, DnaA, binds to *oriC* and causes the unwinding of a short DNA segment, allowing helicase and other proteins to attach, including single-strand binding proteins (SSB).

- DNA helicase (DnaB) removes hydrogen bonds between bases of the two DNA strands. It cannot initiate unwinding alone; initiation proteins first separate the strands at the origin, creating a short single-stranded region where helicase attaches. On the lagging strand template, helicase moves in the 5'→3' direction.
- SSB proteins stabilize the exposed single strands and prevent re-annealing or secondary structure formation.
- A topoisomerase, DNA gyrase, reduces torsional strain ahead of the replication fork. It introduces a double-strand break, passes another segment of the helix through, then reseals the break, eliminating positive supercoils.
- Primase (DnaG) synthesizes RNA primers (10–12 nucleotides) that provide a 3'-OH group for DNA polymerase attachment. DNA polymerases cannot initiate synthesis without a primer. On the leading strand, one primer suffices; on the lagging strand, each Okazaki fragment requires a new primer.
- Primase and helicase form the **primosome**, which moves along the lagging strand template.

➤ **Elongation of nucleotide chains**

Once DNA is unwound and primers are synthesized, DNA polymerases elongate the chains by catalyzing nucleotide polymerization.

The **antiparallel** nature of DNA strands and **5'→3' synthesis** direction create asymmetry:

- ✓ **Leading Strand:** synthesized **continuously** in 5'→3' direction, same direction as replication fork movement. It requires only one primer per origin
- ✓ **Lagging Strand:** synthesized **discontinuously** in 5'→3' direction, opposite to fork movement. It creates **Okazaki fragments** (**prokaryotes**: 1,000-2,000 nucleotides, **eukaryotes**: 100-200 nucleotides), each fragment requires its own RNA primer. It synthesized in direction away from fork, then new primer closer to fork

E. coli has five DNA polymerases; only DNA polymerase I and III are involved in replication, while the others specialize in DNA repair.

- **DNA polymerase III** is the main replication enzyme, with two activities: polymerase activity (5'→3'), elongating the DNA chain and exonuclease activity (3'→5'), removing incorrectly incorporated nucleotides (proofreading), reducing the error rate to 1 in 10⁹.

- **DNA polymerase I** also has 5'→3' polymerase and 3'→5' exonuclease activities, enabling synthesis and error correction. Additionally, it has a 5'→3' exonuclease activity that removes RNA primers and replaces them with DNA nucleotides.
- DNA polymerase III continues synthesis as long as a template strand is available. DNA polymerase I follows, degrading RNA primers and replacing them with DNA. After the last RNA nucleotide is replaced, a nick remains in the sugar-phosphate backbone. DNA ligase seals this nick by catalyzing a phosphodiester bond.

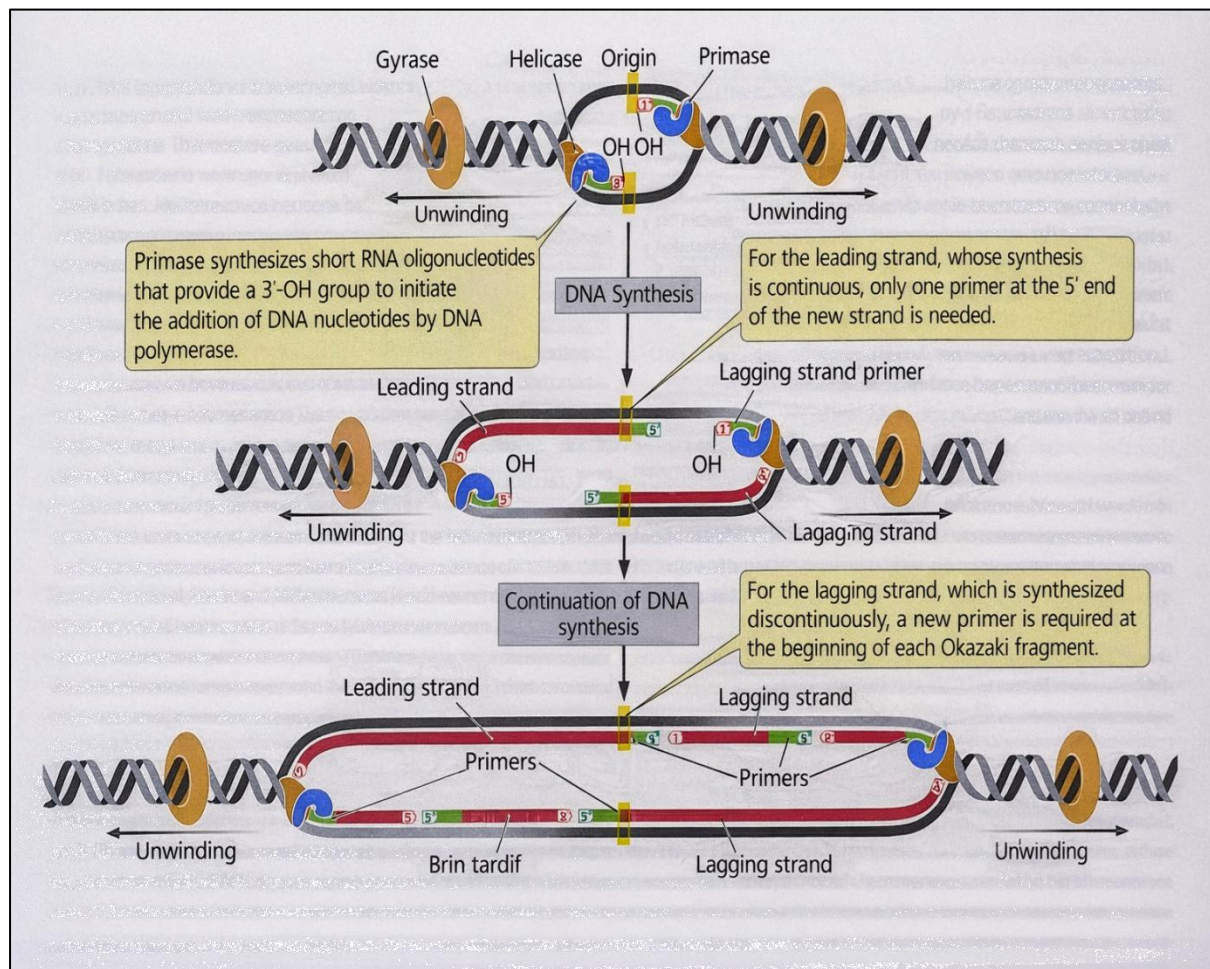


Figure 45. DNA replication

➤ Termination

Replication ends when two replication forks meet, or when termination sequences halt the process. In *E. coli*, a protein called **Tus** binds to termination sequences and stops replication by blocking helicase.

B. DNA replication in eukaryotes

Although DNA replication in eukaryotes resembles that of prokaryotes in several aspects, it nevertheless faces additional challenges related to the organization of the eukaryotic cell:

- The large size of genomes requires replication to be initiated at multiple origins.
- Chromosomes are linear rather than circular.
- The DNA used as a template is associated with histones within nucleosomes, and nucleosome assembly must occur immediately after replication.

➤ **Origins of replication**

The sequences of replication origins vary greatly among eukaryotes, although they are rich in A–T base pairs. In yeast, origins contain about 100 to 120 base pairs of DNA. A multi-protein complex, the Origin Recognition Complex (ORC), binds to these origins and unwinds the DNA in these regions.

➤ **Replication licensing**

Eukaryotes use thousands of replication origins, allowing the entire genome to be replicated within a reasonable time. The whole genome must be replicated once per cell cycle. This regulation is achieved by dividing initiation into two steps:

1. In the first step, origins receive a “replication license”, meaning they gain permission to initiate replication. This occurs early in the cell cycle, in G1 phase, with the binding of a licensing factor to the origins: **Pre-replication complex (pre-RC) Formation:**

- ORC binds origin
- Essential licensing factor Cdc6 and Cdt1 recruit MCM2-7 helicase complex
- Origin is "licensed" (ready but not active)
- S-phase kinases (CDK, DDK) activate licensed origins
- MCM2-7 becomes active helicase

2. In the second step, the replication machinery initiates replication at each licensed origin.

As soon as a replication fork moves away from an origin, the licensing factor is removed, leaving the origin in an unlicensed state where replication cannot be re-initiated until the license is renewed.

Replication occurs only once per cell cycle because the licensing factor is active only after the cell completes mitosis and before replication begins.

➤ **DNA polymerases**

Eukaryotic cells possess many different DNA polymerases:

- DNA polymerase α , which also has primase activity, initiates nuclear DNA synthesis by producing an RNA primer followed by a short DNA sequence.
- After DNA polymerase α synthesizes about 30–40 nucleotides, DNA polymerase δ completes replication of the lagging strand, and DNA polymerase ϵ synthesizes the leading strand after primer removal. Both enzymes have proofreading activity.

- DNA polymerase γ replicates mitochondrial DNA and chloroplast DNA.
- Other DNA polymerases (ζ , η , θ , ι , κ , λ , μ , σ , ϕ , Rev1) are involved in DNA repair.

➤ **DNA unwinding**

Several helicases, single-strand DNA-binding proteins, and topoisomerases have been isolated from eukaryotic cells, performing functions equivalent to their prokaryotic counterparts.

The small size of Okazaki fragments reflects the amount of DNA unwound from each nucleosome as the replication fork progresses. In addition to DNA duplication, histone content is also doubled during the S phase.

Termination

For linear DNA, replication forks stop:

- Either when they meet another replication fork
- Or when they reach the end of a chromosome

Replication at DNA ends

The ends of linear chromosomes (telomeres) cannot be fully replicated on the lagging strand because there is no replicated DNA to provide a 3'-OH group. When the primer is removed, it cannot be replaced with DNA, leaving a gap at the telomere. This would lead to chromosome shortening at each replication. Over generations, this shortening results in telomere loss, chromosome instability, and ultimately cell death.

Telomeric DNA is characterized by many repetitions of a short, G-rich sequence. The single-stranded overhang at the telomere can be extended by telomerase.

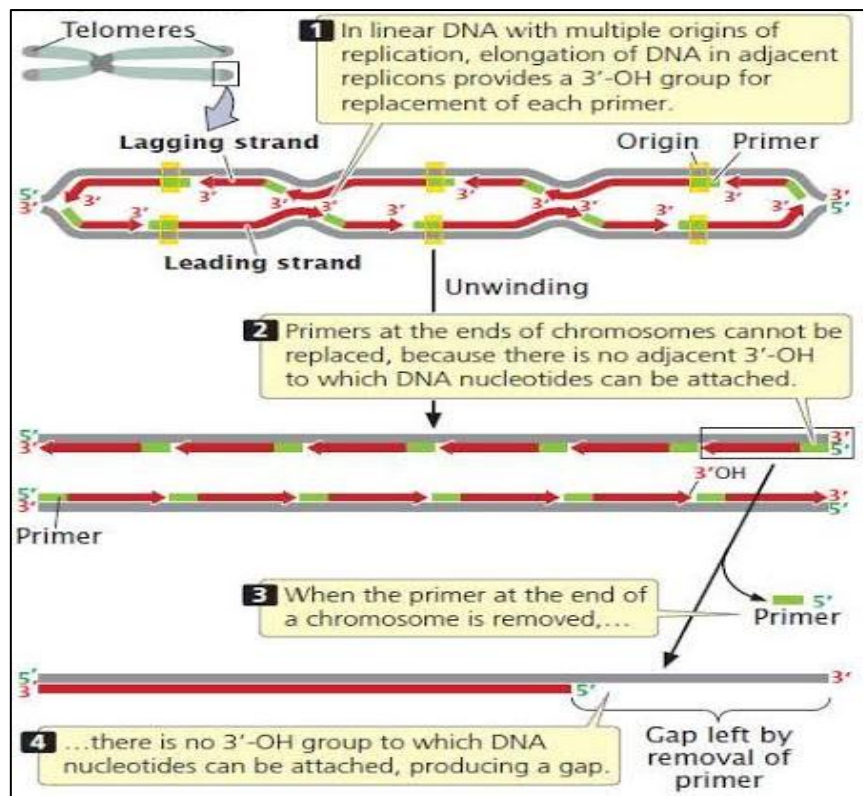


Figure 46. Problem of telomere in eukaryotic DNA replication

➤ Telomerase

Telomerase is a reverse transcriptase composed of two subunits:

- The protein subunit (TERT, Telomerase Reverse Transcriptase): an enzyme with reverse transcriptase activity that synthesizes telomeric sequences using the RNA subunit as a template.
- The RNA subunit: contains the sequence that serves as a template for telomeric repeat synthesis; about 14–16 nucleotides are added to the 3' end of the G-rich strand each time as the enzyme moves along the DNA strand and repeats the process.

The elongated 3' end of the template strand can then serve for the synthesis of a new primer. The 3'-OH end of this primer becomes the starting point for DNA polymerase δ to synthesize the complementary strand.

After removal of the RNA primer, the daughter strand will always be shorter at its 5' end than the parental strand at its 3' end. However, this is not problematic because telomeres consist of many repeated sequences. The loss of a few nucleotides at each replication is not harmful since it occurs in a region that does not contain genetic information (the telomere).

Notes:

Telomere length = "molecular clock"

Short telomeres trigger senescence or apoptosis

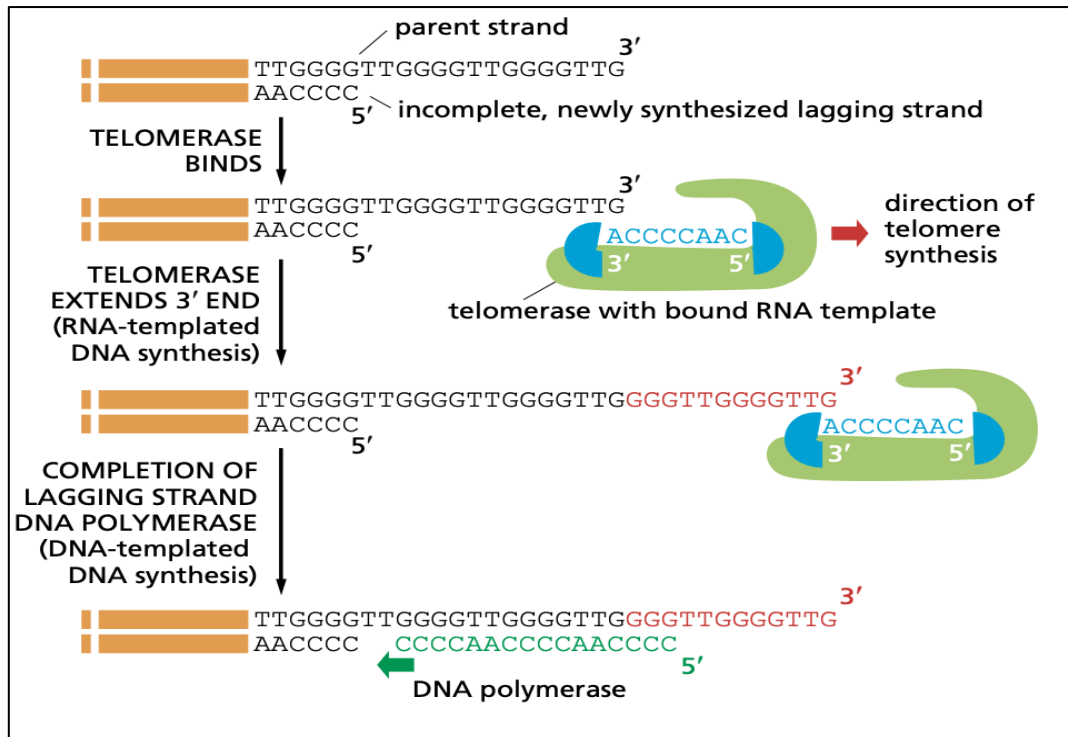


Figure 47. Telomere DNA replication

3.1.5. Regulation of replication:

3.1.5.1. Cell Cycle Control

DNA replication restricted to **S phase** (Synthesis phase):

G1 Phase: Gap 1, growth, origin licensing **S Phase:** DNA synthesis **G2 Phase:** Gap 2, preparation for mitosis **M Phase:** Mitosis (cell division)

DNA replication **must be coordinated with cell division (cell cycle control)** to correctly distribute chromosomes to the two daughter cells. Regulation generally occurs at the initiation phase.

➤ Key Regulators:

A. Cyclin-Dependent Kinases (CDKs):

- Ser/Thr kinases activated by cyclins
- **G1/S-CDK** (CDK2-Cyclin E): Triggers S phase entry
- **S-CDK** (CDK2-Cyclin A): Promotes replication, prevents re-replication

B. Checkpoints: Ensure order and fidelity

- **G1/S checkpoint:** DNA damage check before replication
- **Intra-S checkpoint:** Monitors replication stress (example: transcription-replication conflicts)
- **G2/M checkpoint:** Ensures complete replication before mitosis

C. Origin licensing (G1 phase), **firing** (S phase) and **prevention of re-replication**:

D. Temporal Replication Program: not all origins fire simultaneously:

- **Early-replicating regions:** in the euchromatin (gene-rich, transcriptionally active), origins fire early in S phase
- **Late-replicating regions:** in the heterochromatin (gene-poor, transcriptionally silent), origins fire late in S phase

3.1.5.2. Regulation of replication in prokaryotes:

Replication of the *E. coli* chromosome is controlled at the initiation step. Cell division is blocked when replication is disrupted.

- A '**replicator**' initiates replication by binding to a target once it reaches a sufficient cellular concentration. In *E. coli*, this role is played by **the DnaA protein**, whose synthesis is controlled. The target is oriC, which is subject to a second level of regulation.
 - After initiation, the origin is hemi-methylated (the new strand has not yet been methylated). In this hemi-methylated state, oriC is sequestered in a complex with membrane proteins, preventing immediate re-use of oriC sequences.
- ✓ **DnaA:** (Only used at initiation)

It is sometimes called as **licensing factor in bacteria**. It binds to **the 4 copies of 9 bp repeats**. Further cooperative binding brings in **20 to 40 DnaA monomers**. It Melts the DNA at the 3- 13 bp repeats

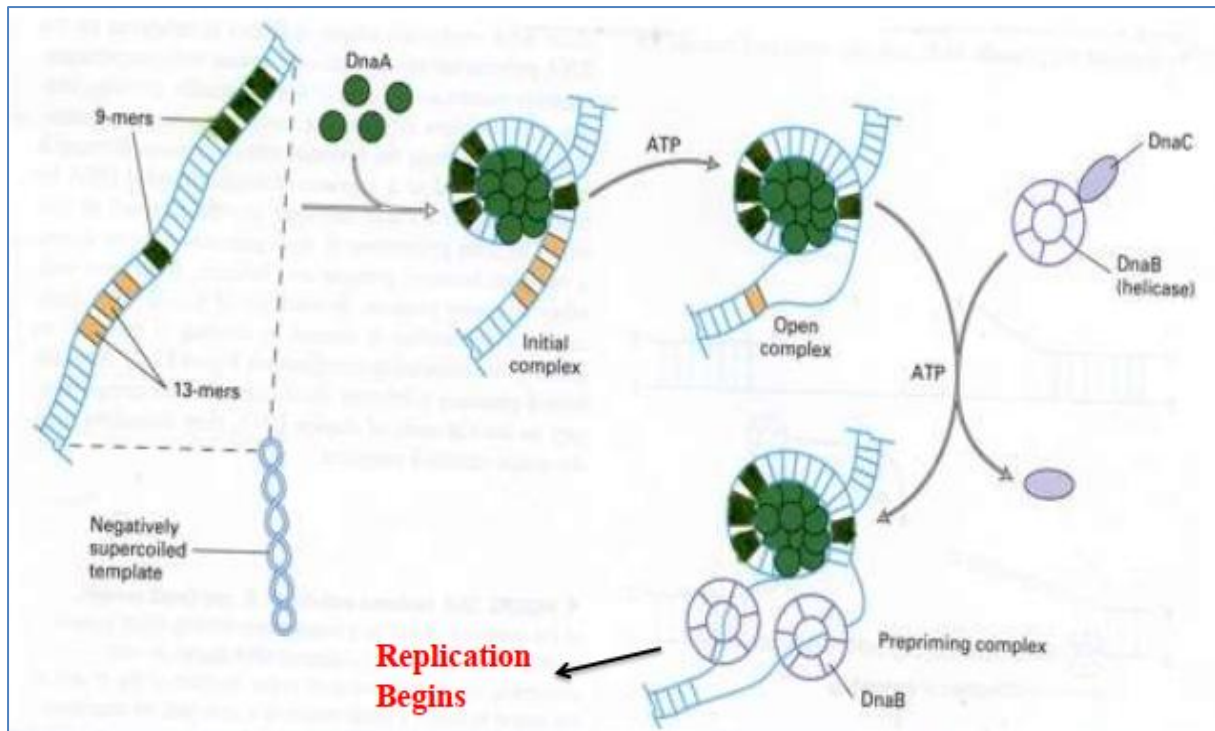


Figure 48. Binding of Dna A protein to the origin oriC

✓ **Methylation of bacterial origin regulates replication during initiation:**

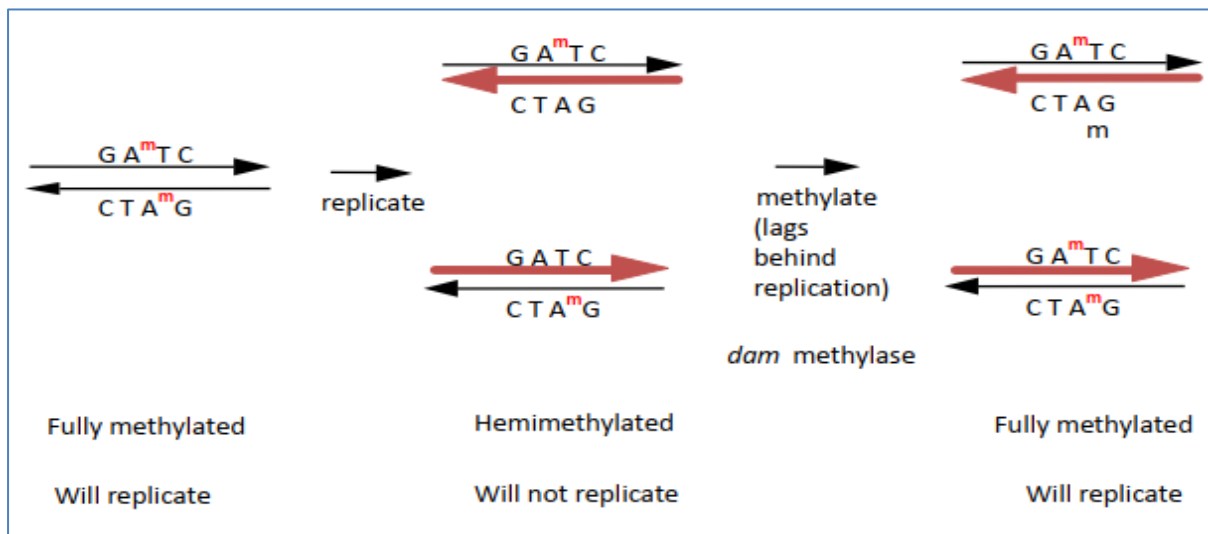


Figure 49. Methylation of bacterial origin regulates replication during initiation

3.1.5.3. Regulation of replication in eukaryotes (during cell cycle):

Study of ARS sequences in *S. cerevisiae* revealed their function as chromosomal replication origins and allowed identification of interacting proteins.

➤ **ORC (Origin Recognition Complex):**

- Present throughout the cell cycle on ARS sequences in *S. cerevisiae*; changes structure after mitosis.

- Only the post-mitotic form is capable of initiating replication from an ARS, after interaction with a signal protein specifically activated at the end of G1 phase (licensing factor).
- After initiation, the ARS/ORC complex changes structure and appears inactivated, preventing new replication from the same origin before cell division.

Since proteins similar to the ORC complex are present in animal cells, this mechanism is likely at least partially conserved in eukaryotes.

➤ **Disappearance of the nuclear membrane**

In animal cells, the disappearance of the nuclear membrane during mitosis also plays a role in preventing origin re-use.

➤ **Nucleosome assembly and chromatin reconstitution during DNA replication is an interesting problem.**

Nucleosomes on replicating DNA are destabilized by the fork and reconstituted on the two post-replicative double helices.

- Newly synthesized histones H3 and H4 are acetylated and deposited as tetramers on replicated DNA: CAF-1 (Chromatin Assembly Factor 1) is responsible for positioning the H3/H4 tetramer on replicated DNA; its activity depends on replication, possibly through indirect interaction with the fork. CAF-1 may be a component of the replisome in animal cells.
- Two H2A/H2B histone dimers then complete the nucleosome in a replication-independent reaction.
- Study of the cellular response to agents that disrupt replication revealed control pathways that inhibit, slow, or accelerate mitosis during S phase. Example: in wild-type yeast, hydroxyurea causes replication arrest.

3.2. DNA Repair and Detection of Mutagenic Potential

Newly synthesized DNA may contain errors (~ 1 per 10^6), despite the proofreading functions of the replication DNA polymerase. It may also sustain damage from many physical and chemical agents: UV or ionizing radiation, heat, mutagenic compounds, oxygen radicals, etc.

- In germ cells: mutations will be transmitted to offspring.
- In somatic cells: mutations will not be transmitted to offspring.

Numerous repair mechanisms must intervene to restore DNA structure, each designed to repair specific lesions based on the complementarity of the two DNA strands.

Failure to repair DNA lesions normally leads to cell death. Tolerating lesions characterizes the genomic instability of cancer cells and contributes to oncogenesis — these lesions can provide an adaptive advantage that cancer cells exploit to proliferate and spread throughout the organism.

3.2.1. Mutations and DNA lesions:

A. Mutations occurring during DNA replication:

Are alterations not corrected by the DNA polymerase proofreading system can lead to:

- Deletions
- Substitutions
- Insertions

B. Mutations occurring outside of DNA replication:

- **Spontaneous lesions**
- **DNA alterations due to mutagenic agents:**
 - ✓ **Physical agents**
 - ✓ **Chemical agents**

3.2.1.1. Mutations and lesions during replication: (see chapter 2)

A. Substitution mutations: (transition or transversion)

Incorrect nucleotide incorporation in the daughter strand. After two replication cycles with mismatches, DNA contains 25% mutant phenotype.

B. Deletion / Insertion mutations:

A nucleotide is missed (or incorporated –extra nucleotide-) by DNA polymerase. After two successive cell cycles, the sequence is completely different from the parental sequences (25% mutant DNA). Causes a frameshift.

3.2.1.2. Mutations and lesions outside replication: (see chapter 2)

A. Spontaneous lesions:

Are high-frequency events that lead to major DNA damage if not repaired. A mutation appears stably after 2 replication cycles.

Examples: depurination, oxidative base modifications, deamination of natural bases, ...

B. Physical agents:

Are very significant: in the absence of repair, they cause a block in replication and therefore in cell division.

Examples: UV light, ionizing radiation, ...

C. Chemical agents:

Examples: base analogue incorporation, alkylating agents, intercalating agents ...

3.2.2. DNA Repair Mechanisms

Each type of lesion can be handled by a specialized repair mechanism. The first step of any repair is recognition of the lesion and determination of whether it is lethal. This recognition is generally followed by cell cycle arrest, allowing either implementation of the appropriate repair mechanism or initiation of a cell death program if lesions cannot be correctly repaired.

It can also be a form of lesion tolerance that allows DNA polymerases to continue replication without repairing the error, this phenomenon is called translesion synthesis (TLS).

Table 5. Some types of lesions and their repair mechanisms

Lesions	Repair Mechanism
O ⁶ -methylguanine	Direct reversal repair
Base mismatch	Mismatch repair (MMR)
Abasic sites / 8-oxoguanine / N ³ -methyladenine / N ⁷ -methylguanine	Base excision repair (BER)
UV-induced lesions (thymine dimers) / Single-strand breaks / N ³ -methyladenine / N ⁷ -methylguanine / Cisplatin adducts / Dichloroethyl adducts (nitrogen mustards) / Carcinogen adducts (benzopyrene)	Nucleotide excision repair (NER)
Lesions caused by ionizing radiation / Double-strand breaks	Homologous recombination (HRR) / Non-homologous end joining (NHEJ)

3.2.2.1. Correction of replication mismatches

a. Proofreading function (3'→5' exonuclease activity) of DNA polymerase:

Corrects mismatches immediately upon recognition. It reduces error rate from 1 in 10⁵ to 10⁷.

b. Mismatch Repair (MMR) system:

Corrects mismatches missed by proofreading. Using methylation-guided recognition, achieves an error rate of 1 per 10⁹ to 10¹⁰ bases.

When a mismatch is detected, the system also identifies which strand to correct — based on the time window before methylation of the daughter strand. The unmethylated (daughter) strand nucleotide at the mismatch is the one corrected.

This system exists in all prokaryotic and eukaryotic cells. It must act within the time window before daughter strand methylation. Once the window passes, DNA methylase

enzymes mirror-methylate the daughter strand. The complex can no longer distinguish the daughter strand.

➤ **In prokaryotes:**

The repair system detects methylation of adenines in GATC sequences and involves MUT enzymes.

Example: in *E. coli* – MUT multi-protein system:

- Recognizes a mismatch and a nearby GATC sequence (up to 1000 bp away) to identify the methylated strand.
- Positions itself at the mismatch and detects the methylated strand.
- Using endonuclease activity: excises the single-stranded DNA fragment containing the mismatch (on the non-methylated strand), creating a gap.
- DNA polymerase III positions itself at the 3'-OH end and synthesizes the complementary antiparallel DNA.
- The final phosphodiester bond is made by DNA ligase.

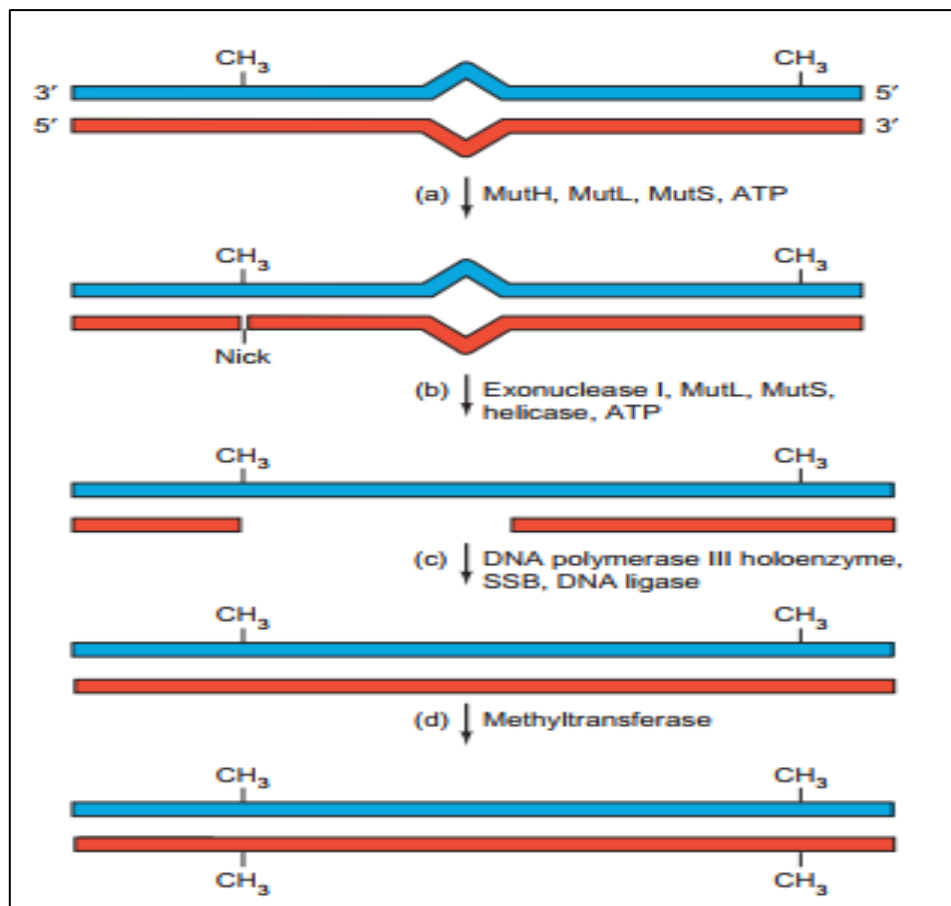


Figure 50. Mismatch Repair (MMR) system in *E. coli*

➤ **In eukaryotes:**

The system detects methylation of cytosines in CG sequences and involves hMSH, hMLH, and hPMS enzymes.

3.2.2.2. Repair of Other DNA Lesions :

Other DNA alterations can cause cell cycle arrest, allowing the cell time to repair damage before transmitting the most intact genome possible to daughter cells.

a. Direct reversal repair:

➤ **O6-methylguanine methyltransferase enzyme (MGMT):**

MGMT is a suicide enzyme, it can only be used for a single repair event. This mechanism therefore depends on the availability of MGMT.

MGMT repairs alkylated (modified) bases formed at the O6 position of guanine, restoring the correct base, and transfers the alkyl group to the thiol group of a cysteine residue.

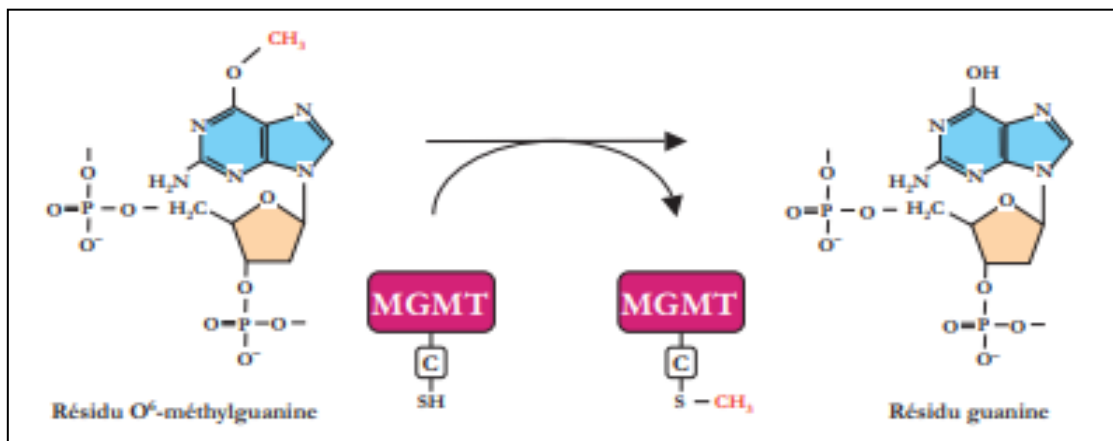


Figure 51. Direct reversal repair by MGMT

➤ **Photolyase (Photoreactivation) mechanism:**

Photolyases are enzymes repair UV-induced lesions (thymine dimers / photo-dimers). They are activated by visible light. Photolyase binds to the thymine dimer and cleaves it, restoring two adjacent thymine.

Note: Photolyase is not present in humans. It exists only in certain organisms: bacteria, some lower eukaryotes (example: Drosophila, plants).

b. Base excision repair (BER):

Also called the short-patch repair mechanism. This system repairs spontaneous deaminations, depurinations, or depyrimidinations. It restores AP sites (in the case of spontaneous deaminations, an AP site is first created).

This mechanism primarily aims to replace nitrogenous bases altered by endogenous oxidative processes:

✓ **Step 1 – DNA glycosylases:**

Depending on the damaged base, glycosylases hydrolyze the N-glycosidic bond, removing the base and creating an AP site.

✓ **Step 2 – AP endonuclease:**

Cuts the deoxyribose-phosphate backbone at the AP site by cleaving the phosphodiester bond and removing the AP site.

✓ **Step 3 – DNA polymerase:**

Positions itself at the 3'-OH end and adds the complementary nucleotide by polymerization.

✓ **Step 4 – DNA ligase:**

Forms the final phosphodiester bond.

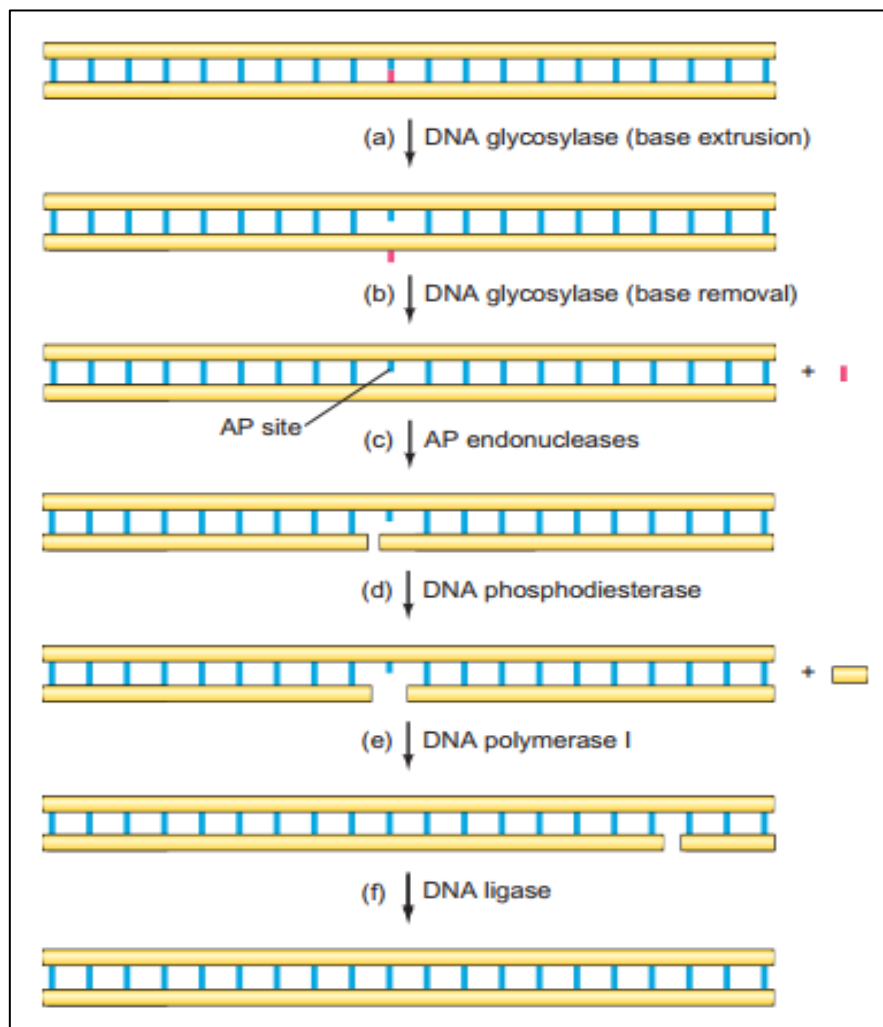


Figure 52. Base excision repair

c. Nucleotide excision repair (NER)

Also called general excision repair; based on the same principle as BER but addresses bulkier lesions (e.g., thymine dimers).

➤ **In *E. coli*:**

- Repair is carried out by a multi-enzyme complex: the UvrABC endonuclease, encoded by *UvrA*, *UvrB*, and *UvrC* genes.
- The complex positions itself at the lesion, recognizes the modification, and induces a distortion of the double helix.
- DNA helicase separates and dissociates the two strands, allowing the endonuclease to cleave a fragment of ~13 nucleotides containing the lesion.
- DNA polymerase I binds at the free 3'-OH end to synthesize new DNA; DNA ligase seals the final bond.

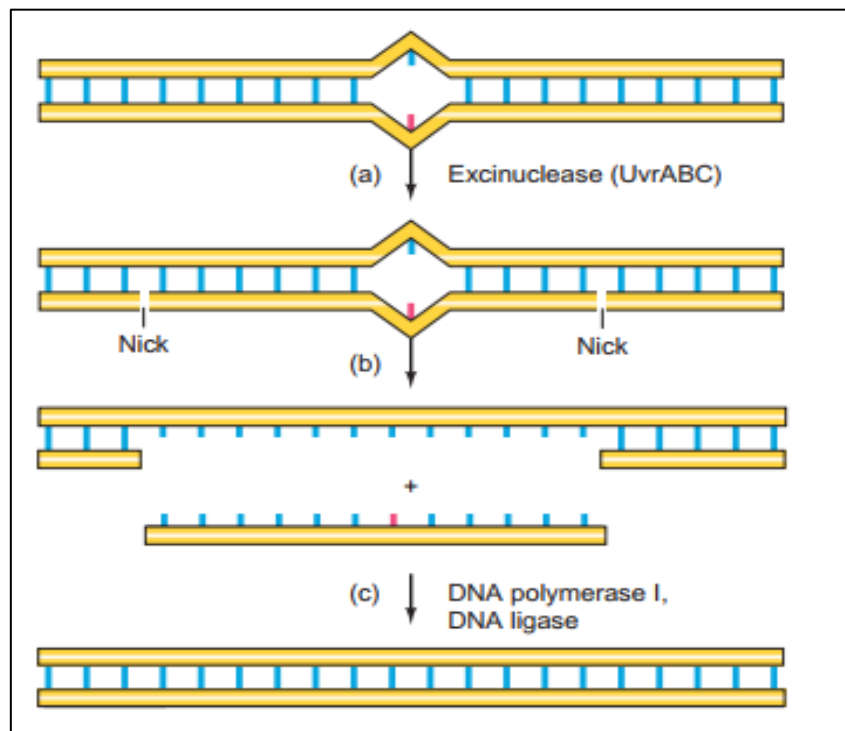


Figure 53. Nucleotide excision repair in *E. coli*

➤ **In Humans:**

- NER is carried out by ERCC1 and XP proteins.
- XPC recognizes the lesion. Some XP proteins have helicase activity (XPB, XPD).
- Excision is performed by XPF, ERCC1, and XPG.
- DNA synthesis is carried out by DNA polymerases δ and ϵ ; DNA ligase seals the final bond.

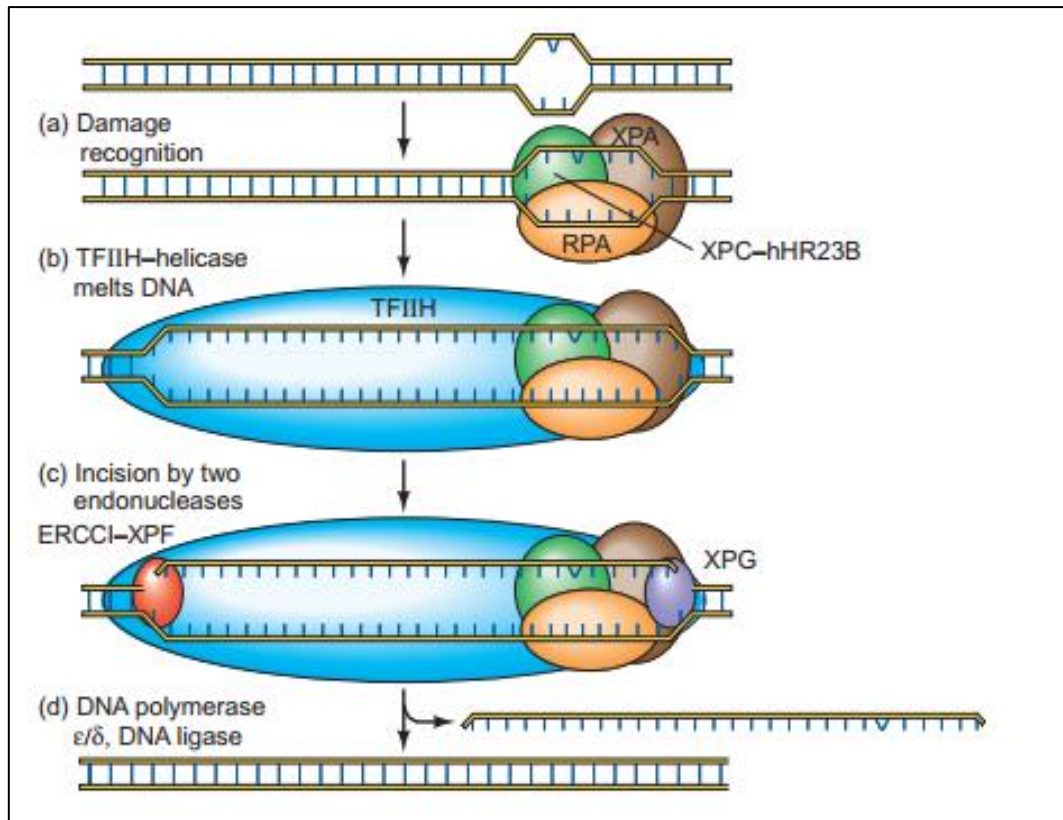


Figure 54. Nucleotide excision repair in eukaryotes

D. Repair by Homologous Recombination:

A constitutive repair system that intervenes when previous systems have failed to repair lesions. Acts against major DNA lesions, especially those induced by ionizing radiation or oxidative agents when the NER system is overwhelmed or defective.

➤ **Mechanism in prokaryotes:**

- If no repair occurs before replication, a gap (post-replicative gap) is left in the daughter strand opposite the dimer.
- This gap is filled with the sequence from the opposite parental strand (identical to the daughter strand), retrieved by the REC A protein.
- This provides the correct sequence and creates a second gap on the opposite parental strand.
- The dimer is then excised by excision enzymes.
- Both gaps are corrected by DNA polymerase, which synthesizes the complementary antiparallel sequence of each parental strand.

➤ **Mechanism in humans:**

Equivalent mechanism but involving RAD51 protein (eukaryotic homolog of REC A), associated with BRCA1 and BRCA2 proteins.

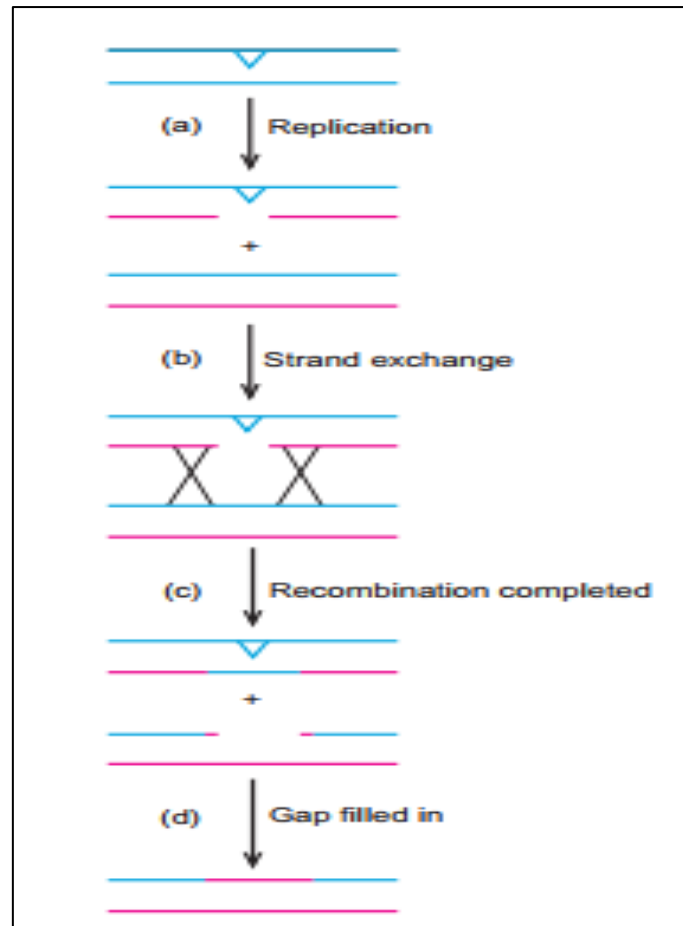


Figure 55. Repair by Homologous Recombination

E. Repair by non-homologous end joining

A constitutive repair system, **directly** ligates broken ends without requirement for homologous template. Often introduces small insertions/deletions or mutations

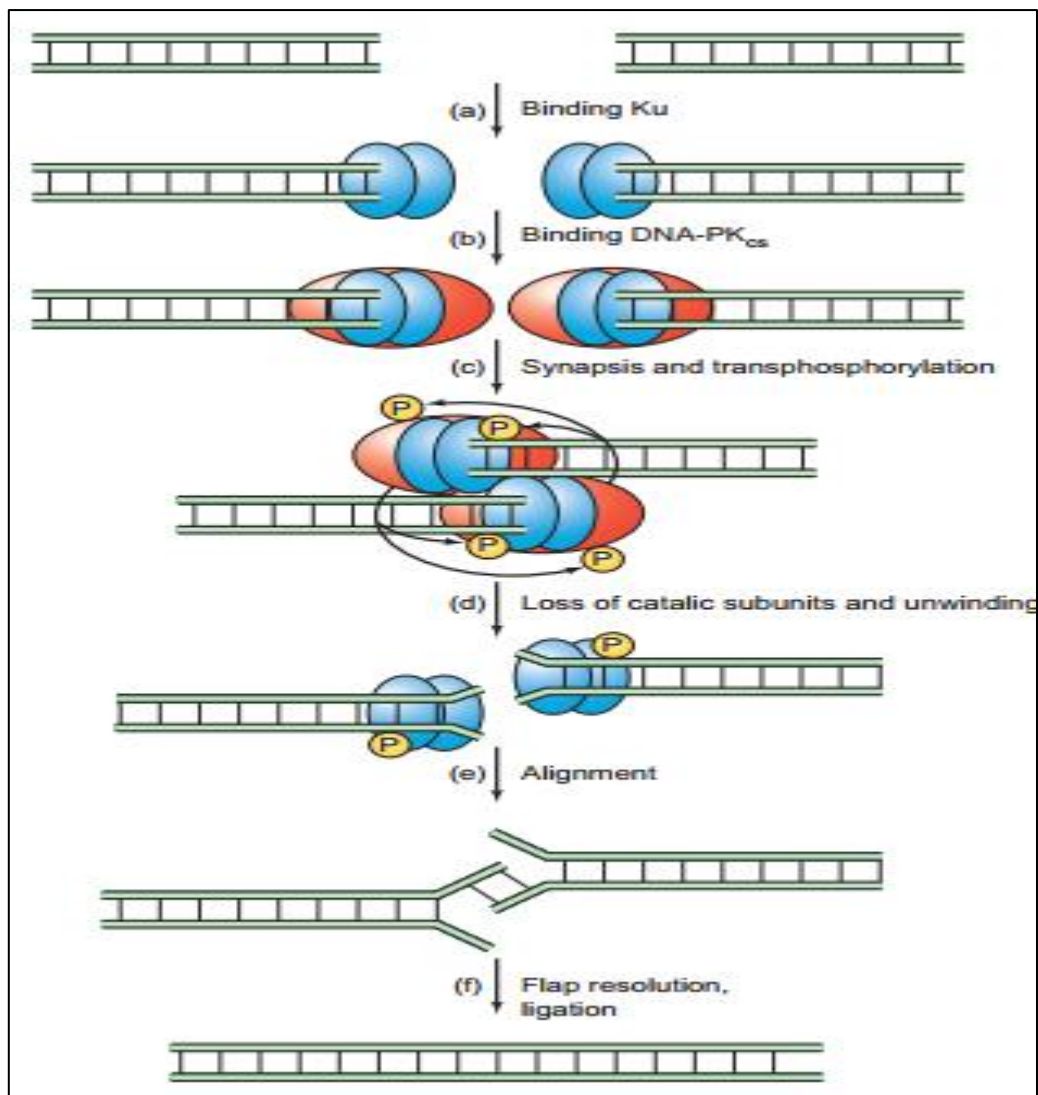


Figure 56. Repair by non-homologous end joining

F. The SOS System

If DNA damage is too extensive, the SOS system is activated — the last resort for attempting DNA repair. It is a homologous recombination repair system, but it is inducible.

➤ Mechanism:

- When recombination systems are overwhelmed, replication is stopped.
- The SOS system activates the second function of the REC A protein: its proteolytic (protein-degrading) activity.
- This activity leads to degradation of its own repressor, LEX A. REC A proteins and about twenty other proteins encoded by SOS genes are then synthesized.
- This stimulates the homologous recombination system.

Note: Unlike other repair systems, the SOS system is inducible, whereas excision repair systems are constitutive.

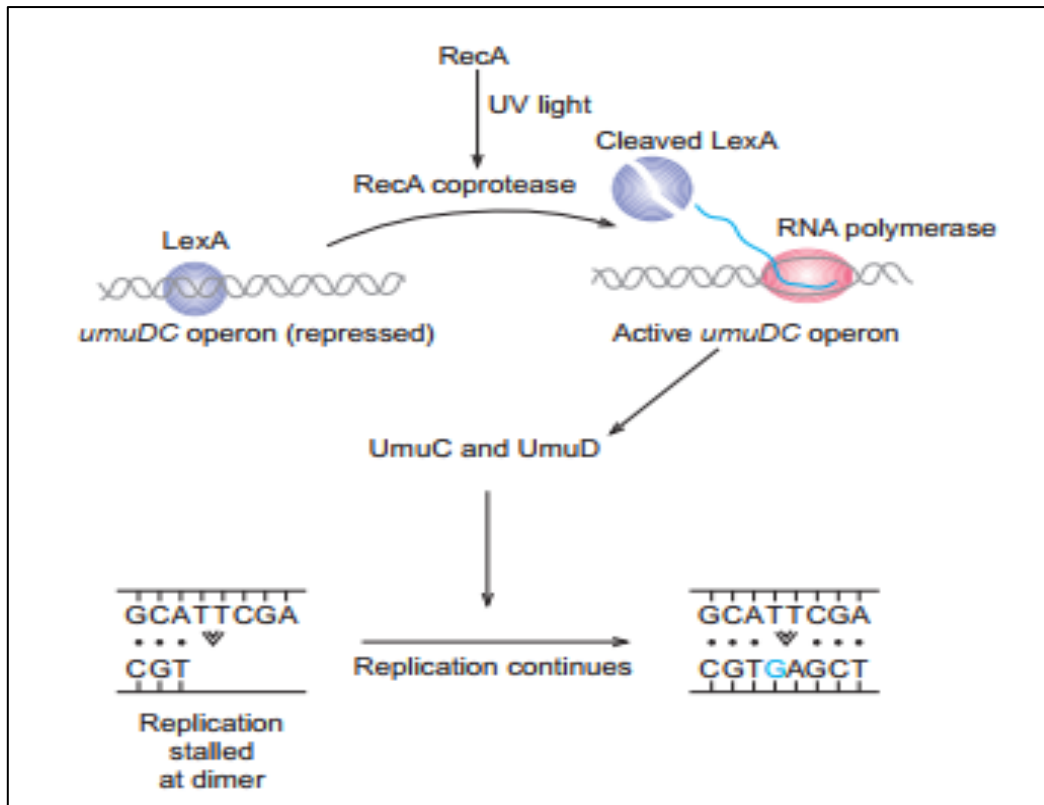


Figure 57. Repair by SOS system

3.2.3. Detection of mutagenic potential

Assessing whether chemicals or environmental factors cause mutations

3.2.3.1. The Ames Test: developed by Bruce Ames in the 1970s.

This test uses mutant bacterial strains *Salmonella typhimurium* *His*⁻ mutants: unable to synthesize histidine (they require histidine in the medium), to detect whether a substance can cause mutations and therefore potentially be carcinogenic.

➤ Protocol

- Preparation: the *His*⁻ bacteria are cultured with the test substance
- Plating: on minimal medium without histidine (or with very little)
- Incubation: the only colonies that can grow are those that have undergone a reverse mutation (*His*⁻ → *His*⁺), allowing them to synthesize histidine
- Counting: revertant colonies are counted

➤ Metabolic Activation System (S9)

Some substances are only mutagenic after hepatic metabolism. Therefore, we add: S9 fraction: rat liver extract containing metabolic enzymes, it allows detection of pro-mutagens (substances becoming mutagenic after metabolism)

➤ Interpretation of Results

- **Negative control** (without substance): Low number of spontaneous revertants
- **Test substance:**
 - If number of revertants $\gg$ negative control $\rightarrow$ substance is **mutagenic**
 - If number of revertants $\approx$ negative control $\rightarrow$ substance is **non-mutagenic**

➤ **Advantages of the Ames Test**

- Fast (48-72 hours), inexpensive and sensitive
- Allows testing of numerous substances
- Good indicator of potential carcinogenicity

➤ **Limitations**

- Test on bacteria, not eukaryotic cells. Requires confirmation with mammalian tests
- Not all mutagens are carcinogens and not all carcinogens are mutagens

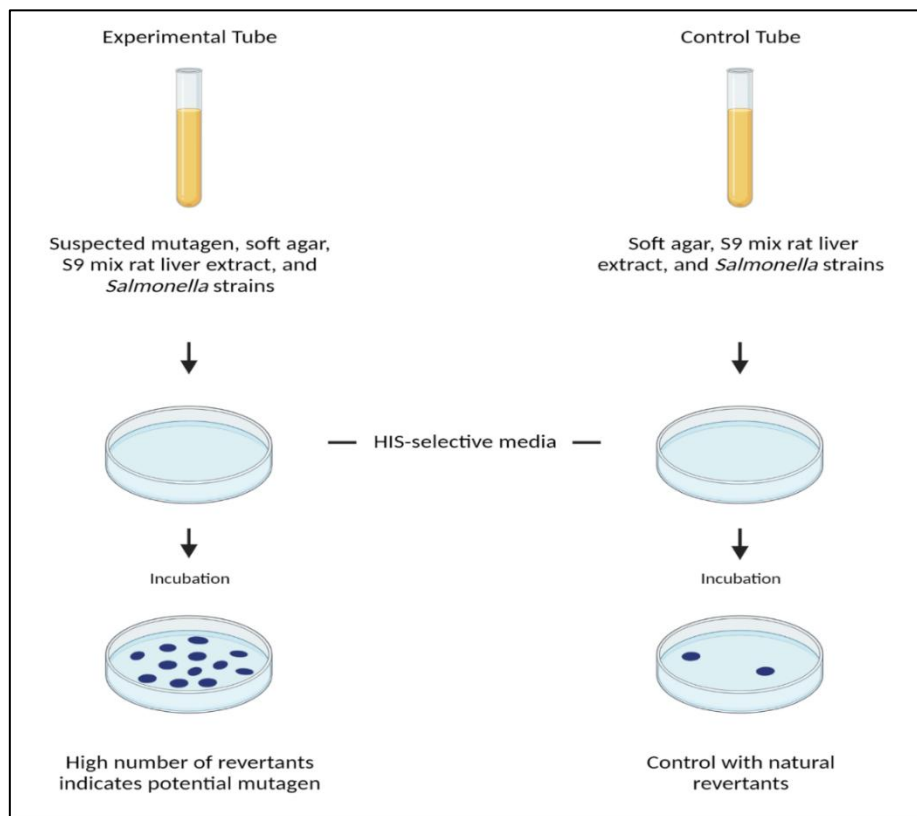


Figure 58. Ames test

3.2.3.2. Mammalian cell mutagenesis Assays

a. HPRT assay (Hypoxanthine-Guanine Phosphoribosyl-transferase):

- Uses mammalian cells in culture (example: Chinese hamster ovary cells)
- **HPRT gene**: Non-essential, X-linked
- Cells treated with test compound
- Select for *HPRT*⁻ mutants with 6-thioguanine (toxic if HPRT functional)

- Count resistant colonies (mutants)

B. Tk assay (Thymidine kinase):

- Mouse lymphoma cells
- Selects *TK*⁻ mutants with trifluorothymidine

C. Chromosome aberration test:

- Mammalian cells treated with compound
- Examine metaphase chromosomes
- Score structural aberrations (breaks, gaps, exchanges)

3.2.3.3. *In Vivo* mutagenicity tests

Rodent Bone Marrow Micronucleus Test:

- Animals exposed to test compound
- Bone marrow cells examined for micronuclei in erythrocytes
- Integrates absorption, metabolism, distribution, excretion (more physiological)

3.2.4. Deficiency of the DNA repair systems and genetic diseases:

Several human genetic diseases are linked to deficiencies in DNA repair systems. These deficiencies often lead to an increased incidence of specific types of cancers, due to elevated mutation rates.

Table 6. Examples of deficiency of the DNA repair systems and genetic diseases

Disease	Symptoms	Genetic deficiency
Xeroderma pigmentosum	Pigmented spots on the skin, hypersensitivity to light, predisposition to skin cancer	Nucleotide excision repair system
Cockayne syndrome	Dwarfism, light sensitivity, premature aging, deafness, intellectual disability	Nucleotide excision repair system
Trichothiodystrophy	Very brittle hair, dry and scaly skin (ichthyosis), short stature, intellectual deficit, characteristic facial features	Nucleotide excision repair system
Hereditary non-polyposis colorectal cancer	Predisposition to colon cancer	Mismatch repair system
Fanconi anemia	Increased skin pigmentation (melanodermia), skeletal, cardiac, and renal abnormalities, predisposition to leukemia	Deficiency in interstrand cross-link removal and repair

3.3. Restriction-Modification Systems

Bacterial defence mechanisms against foreign DNA (bacteriophages). Self vs. non-self-recognition: Host DNA methylated (protected), foreign DNA not methylated (targeted). It is an "Immune system" of bacteria:

- **Restriction enzymes:** Cut foreign DNA
- **Modification enzymes:** Methylate host DNA to protect from restriction

3.3.1. Types of Restriction-Modification Systems

3.3.1.1. Type I: Complex multifunctional enzymes:

Single complex with restriction, modification, and DNA recognition activities; It has three subunits: R (restriction), M (modification), S (specificity)

➤ **Recognition:**

Bipartite, asymmetric sequences (~ 15 bp with spacer). Example: EcoKI: AAC(N₆)GTGC

➤ **Mechanism:**

- Binds recognition site, if unmethylated, translocates DNA (ATP-dependent)
- Cleaves at random site ~1000 bp away from recognition sequence
- Both strands cut, variable location

➤ **Modification:** methylates adenine in recognition sequence

3.3.1.2. Type II: Most commonly used in molecular biology

Separate enzymes: restriction endonuclease: cleaves DNA and methyltransferase: methylates DNA

➤ **Recognition:**

Short, palindromic sequences (4-8 bp). Example: EcoRI: 5'-GAATTC-3' (6 bp)

➤ **Cleavage:** at recognition sequence, creates:

- **Blunt ends:** Straight cut
- **Sticky ends (cohesive ends):** Overhangs

➤ **Mechanism:**

- Homodimer (two identical subunits), binds palindromic sequence and cleaves both strands
- Requires Mg²⁺ cofactor

➤ **Modification:**

- Methyltransferase adds methyl to adenine (m⁶A) or cytosine (m⁵C or m⁴C) within recognition sequence. Methylation blocks restriction enzyme

➤ **Applications in Molecular Biology:**

- **Cloning:** Cut vector and insert, ligate together
- **Restriction mapping:** Determine restriction site locations in DNA
- **RFLP analysis:** Detect polymorphisms
- **Molecular diagnostics:** Detect mutations affecting restriction sites

3.3.1.3. Type III: Intermediate complexity:

Two subunits: Restriction/Modification (Res/Mod). ATP-dependent

➤ **Recognition:** Asymmetric, 5-6 bp sequences. Example: EcoP1I: AGACC

➤ **Cleavage:**

- Requires two oppositely oriented recognition sites
- Cleaves ~25-27 bp 3' to recognition sites

➤ **Modification:** methylates adenine

3.3.1.4. Type IV: Recognize modified DNA (opposite of Types I-III)

- Target methylated, hydroxymethylated, or glucosyl-hydroxymethylated DNA
- Cut modified (foreign) DNA
- Example: **McrBC** system in *E. coli* cuts DNA with methylated cytosines
- **Function:** Defence against phages that modify their DNA to evade Types I-III

3.3.2. Biological Significance

- Evolutionary Arms Race
- Horizontal Gene Transfer Barrier
- Genome Evolution

3.4. Tutorial (Annex)

4. Expression of Genetic Information and Its Control

Gene expression = the entire process that decodes the information carried by a gene and translates it into a protein. It occurs in two stages: transcription and translation.

In prokaryotes, transcription and translation occur simultaneously in the same space; translation begins before transcription is finished.

In eukaryotes, the steps are separated in time and space: RNA is synthesized in the nucleus, then exported to the cytoplasm for translation.

4.1. Transcription and RNA Maturation

4.1.1. Transcription

4.1.1.1. Definition:

Transcription is copying genetic material (DNA) into RNA or synthesis of RNA in the 5' → 3' direction by RNA polymerase using the 3' → 5' DNA template strand in three steps: initiation, elongation and termination.

- **Template strand** (antisense strand): DNA strand used as template (3'→5')
- **Coding strand** (sense strand, non-template): Same sequence as RNA (except T→U) (5'→3')

In prokaryotes: one RNA polymerase synthesizes all RNAs.

In eukaryotes: three RNA polymerases exist:

- RNA polymerase I → rRNA
- RNA polymerase II → mRNA + some snRNA
- RNA polymerase III → tRNA, 5S rRNA, most snRNA

4.1.1.2. Key differences from replication:

- Only one DNA strand transcribed (template strand)
- RNA polymerase doesn't require primer
- Uses ribonucleoside triphosphates (ATP, GTP, CTP, UTP); uracil instead of thymine
- Lower fidelity (error rate $\sim 10^{-4}$ to 10^{-5})
- Only specific genes transcribed (not entire genome)

4.1.1.3. Structure of a gene

A gene is a functional unit of genome. The structure of a gene refers to the organized regions of DNA that are necessary for producing a functional RNA or protein.

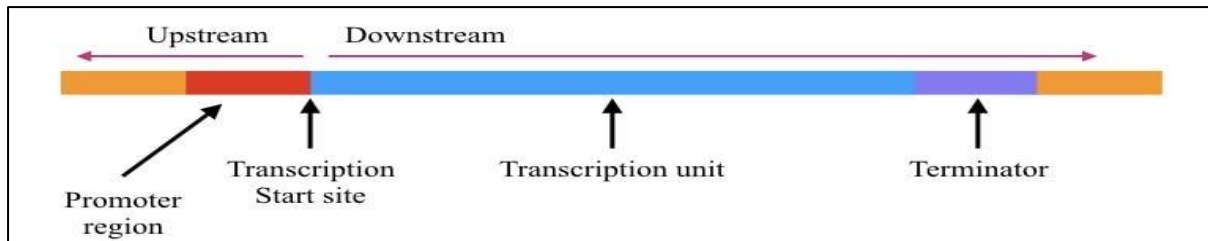


Figure 59. Structure of prokaryotic gene

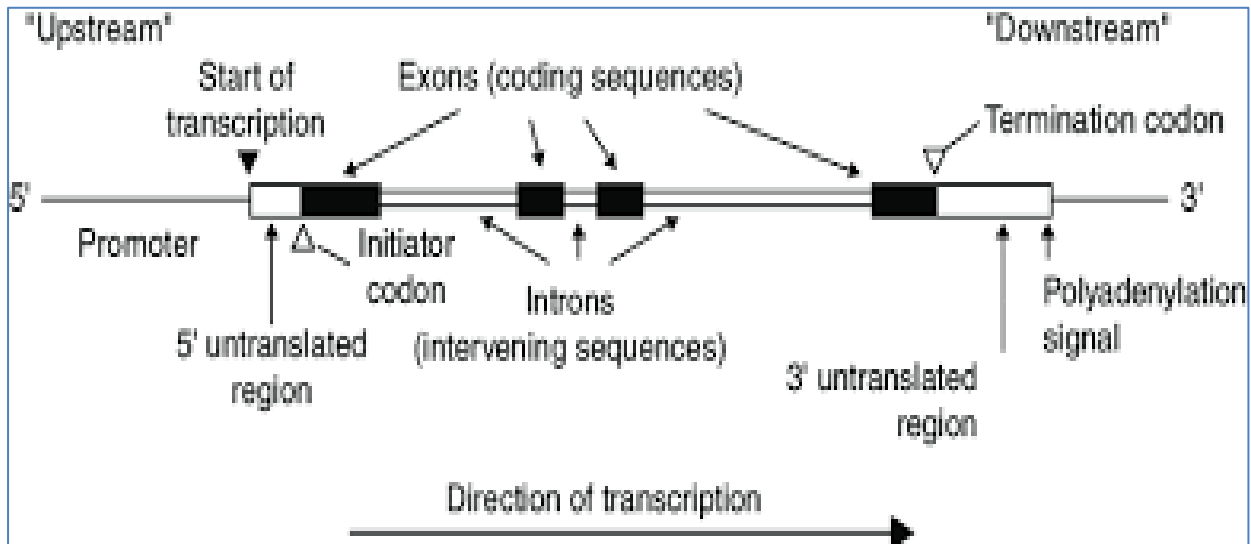


Figure 60. Structure of eukaryotic gene

A. Basic Components of a Gene

- **Promoter:** located **upstream (before)** the gene, is the binding site for RNA polymerase. It determines where transcription starts
- **Transcription Start Site (TSS):** the exact point where transcription begins. Marked as **+1 position**
- **5' Untranslated Region (5' UTR):** region between TSS and coding sequence. Not translated into protein
- **Coding Region (Gene Body):** in eukaryotes, it contains **exons** (coding sequences) and **introns** (non-coding sequences: removed during splicing). In prokaryotes: usually it is **continuous** (no introns)
- **3' Untranslated Region (3' UTR):** located after the coding sequence. Important for: mRNA stability.
- **Terminator:** signals the **end of transcription**. RNA polymerase stops here

B. Extra Regulatory Elements

- **Enhancers** → increase gene expression (can be far away)
- **Silencers** → decrease gene expression

➤ **Operator (prokaryotes)** → binding site for repressors

Table 7. Gene Structure in Prokaryotes vs Eukaryotes

Feature	Prokaryotes	Eukaryotes
Introns	Absent	Present
Organization	Operons	Single genes
mRNA	Polycistronic	Monocistronic
Regulation	Simple	Complex

Then, a gene is not just a coding sequence, it includes regulatory parts and coding protein parts. All of these work together to control **when, where, and how much protein is produced**.

4.1.1.4. Prokaryotic transcription

RNA polymerase have not exonuclease activity.

Transcription most of the time produces polycistronic mRNAs.

A. RNA Polymerase of *E. coli*

Is a multi-subunit complex (~450 kD).

Core enzyme ($\alpha_2\beta\beta'$) contains catalytic sites.

Holoenzyme ($\alpha_2\beta\beta'\sigma$) is required for initiation; the core enzyme continues elongation.

Table 8. the sub-units of RNA Polymerase of *E. coli*

Sub-unit	Number	mass (kd)	Role
α	2	37	Assembly the subunits, promoter binding
β	1	151	phosphodiester bond formation
β'	1	155	binds DNA template
σ	1	50	recognizes promoter, initiates synthesis

B. Promoter structure

It is a **Cis** sequences located upstream of the gene on the DNA molecule. It is known by:

- **-10 box (Pribnow box):** consensus sequence : 5'-TATAAT-3'. Situated ~10 bp upstream of transcription start site (+1), it is AT-rich (easier to unwind)
- **-35 box:** consensus sequence : 5'-TTGACA-3'. Situated ~35 bp upstream of +1.

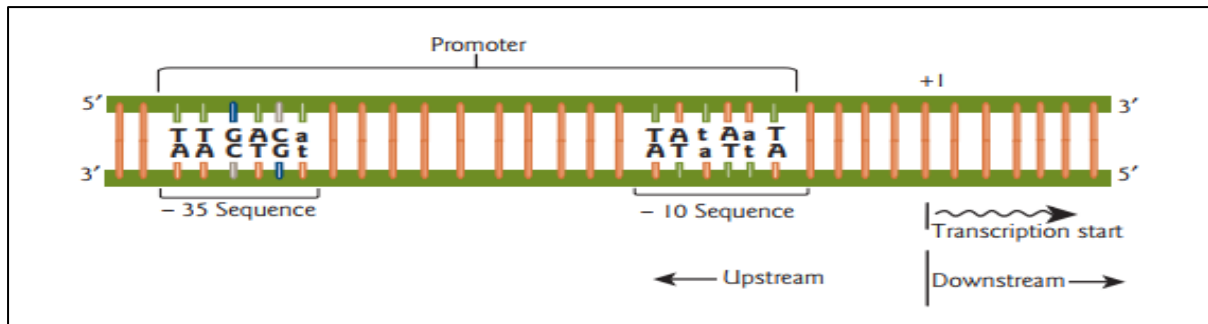


Figure 61. Prokaryotic promotor

C. Three stages of transcription

➤ Initiation

requires promoter sequences: These sequences allow the recognition of the enzymatic complex.

The holoenzyme binds to the DNA (double strand) and moves along the molecule until it encounters a promoter thanks to the σ subunit.

It binds loosely to the -35 and -10 regions: this is the **closed promoter complex**. Then, it separates the two strands (~13 bp) near the -10 region, creating transcription bubble: indeed, only one of the two strands serves as a template. This is the **open promoter complex**. RNA synthesis can begin, without a primer, by the incorporation of nucleotides.

➤ Elongation:

Elongation occurs in the 5'→3' direction. A free nucleotide with a 5'P is brought in, and the enzyme catalyses the formation of the phosphodiester bond with the 3'OH end of the previous nucleotide (thanks to the β subunit).

Transcription reading and synthesis take place in a transcription bubble. The RNA forms a hybrid with the DNA over a constant length (about 12 base pairs). RNA gradually detaches as elongation progresses. The bubble moves along with the enzyme as elongation proceeds. The double helix opens ahead and rewinds behind.

After about 10 nucleotides are added, the σ subunit detaches, leaving only the core enzyme to continue elongation.

➤ Terminaison:

There are termination signals that lead to the stop of polymerization, the dissociation of the RNA-DNA hybrid as well as the release of the enzyme and the closure of the bubble:

Termination sequences consists of about 40 bp, rich in G and C, followed by an A-rich sequence (template strand): The RNA forms a stable hairpin, double-stranded, because it contains a lot of G and C (3 bonds), which slows down RNA polymerase. Then there is the

region with A paired with U, the RNA-DNA interaction is weaker, which facilitates the detachment of the RNA and the enzyme.

The ρ (rho) factor. It is a hexamer: 6 identical subunits. It is capable of hydrolysing ATP. It can bind to single-stranded RNA (at specific sequences) and then it moves along the RNA thanks to ATP hydrolysis. Thus, it eventually encounters the RNA polymerase enzyme and causes the separation between the enzyme and the RNA.

The sigma subunit reassociates with the core enzyme, and the holoenzyme goes in search of a new promoter.

Translation then begins even though transcription is not yet complete.

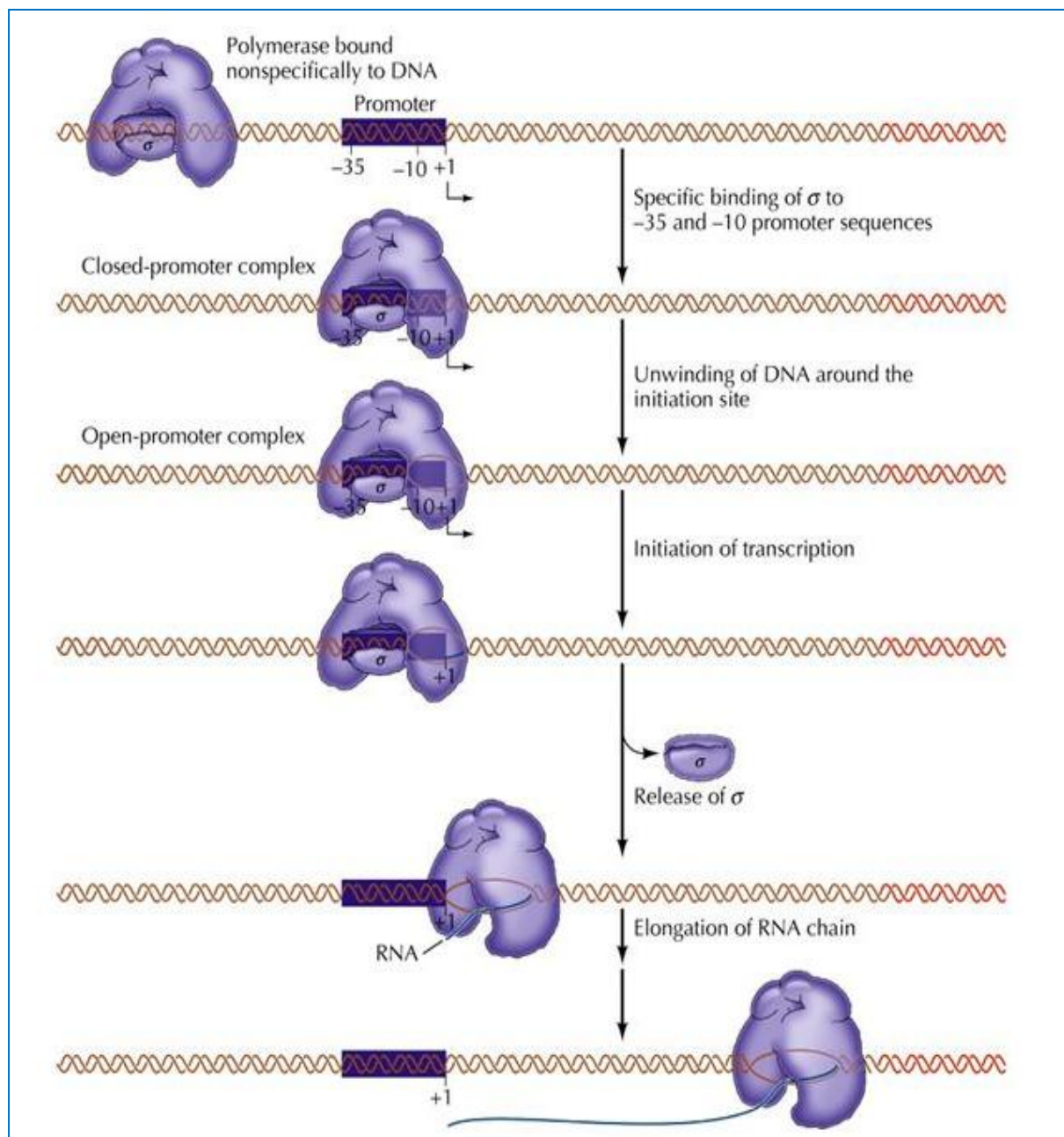


Figure 62. Prokaryotic transcription

4.1.1.5. Eukaryotic transcription

Much more complex than prokaryotic. It is formed in the same way as in prokaryotes with some differences: There are 3 types of RNA polymerase: RNA polymerase II, which ensures the synthesis of mRNA.

The initiation is quite different; the initiation and regulation of transcription are closely linked. The promoter sequences are not recognized by the RNA polymerase itself, but by TF proteins (Transcription Factors), which are the general transcription factors (essential for any transcription). For RNA polymerase II: there are TFII A, B, D, E, F, and H. TFII D binds first by recognizing the sequences, then the others arrive in cooperation with TBP (TATA Binding Protein): this is the transcription initiation complex.

RNA polymerase II is also equipped with protein termination factors, and thus recognizes one or more termination signals (specific signals) carried by the template strand that indicate the end of transcription (TTATTT for example).

The necessity to bypass DNA packaging. mRNA (the primary transcript) undergoes modifications in the nucleus before being exported to the cytoplasm.

A. RNA Polymerase II Promoters:

It is a **Cis** sequences located upstream of the gene on the DNA molecule. It is known by:

- **TATA box:** consensus sequence : 5'-TATAAA-3'. Situated ~25 to 30 bp upstream of TSS, Found in ~10-20% of mammalian genes. It binds TBP (TATA-Binding Protein)
- **Proximal promoter elements:** situated at ~50-200 bp upstream of TSS. They are binding sites of different factors, tissue-specific elements, examples: CAAT box, GC box.

B. Enhancers and silencers:

They are regulatory elements far from promoter (can be kb away, upstream or downstream), contain binding sites for transcription factors. They can function in orientation-independent manner

C. General Transcription Factors (GTFs)

Required for basal transcription by Pol II:

- **TFIID:** contains **TBP** (TATA-Binding Protein), **TAFs** (TBP-Associated Factors); it recognizes core promoter
- **TFIIA:** stabilizes TFIID-DNA interaction
- **TFIIB:** recruited after TFIID, it determines TSS position and recruits RNA polymerase II.
- **TFIIF:** binds RNA polymerase II, escorts to promoter

- **TFIIE**: recruited with RNA polymerase II, recruits TFIIH
- **TFIIH**: it has helicase activity (unwinds DNA at TSS), **kinase activity** (CTD kinase):
Phosphorylates RNA polymerase II CTD

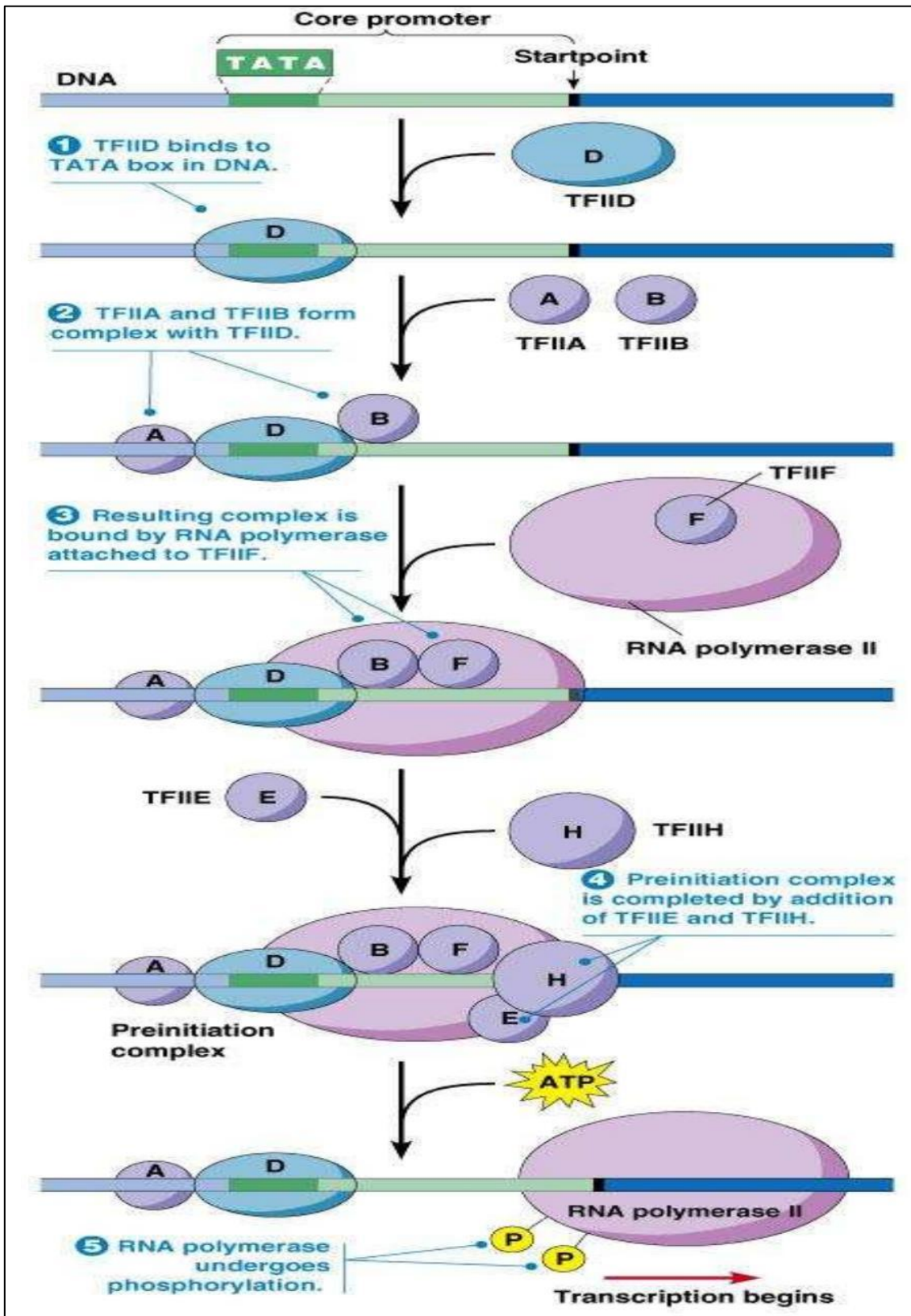


Figure 63. Eukaryotic initiation transcription

4.1.2. RNA Maturation

4.1.2.1. mRNA Maturation in Eukaryotes

Prokaryotic mRNAs used directly (no processing, coupled transcription-translation).

Eukaryotic pre-mRNAs extensively processed:

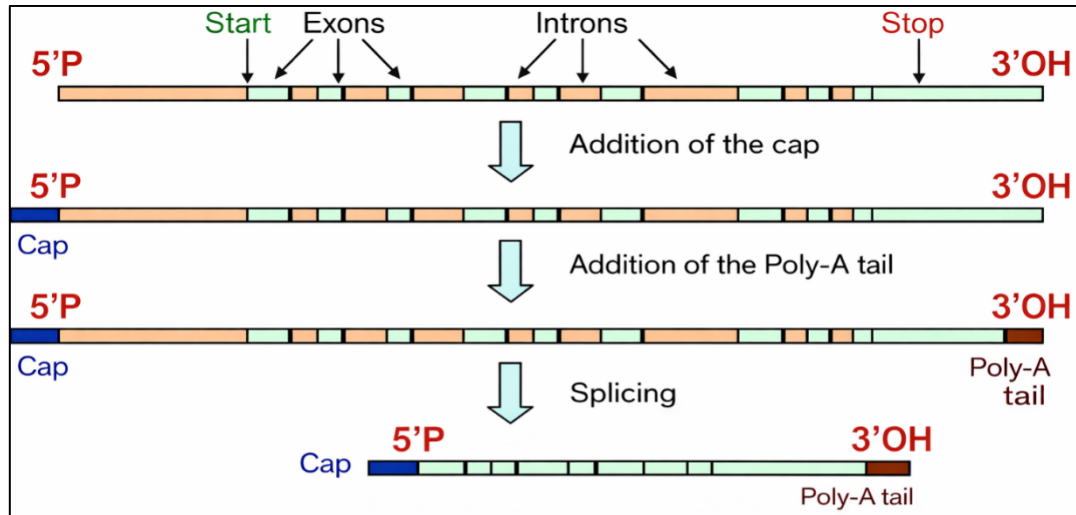


Figure 64. mRNA Maturation in Eukaryotes

A. 5' Capping

Occurs very early (when transcript ~20-30 nt); The 5' end is immediately covered with a cap: **7-methyl-guanosine triphosphate (m7G-P-P-P)**. It is **Inverted guanosine**: 5'-5' triphosphate linkage (unusual), methylated at N⁷ position of guanine. The first 1-2 nucleotides of RNA also methylated (2'-O-methylation)

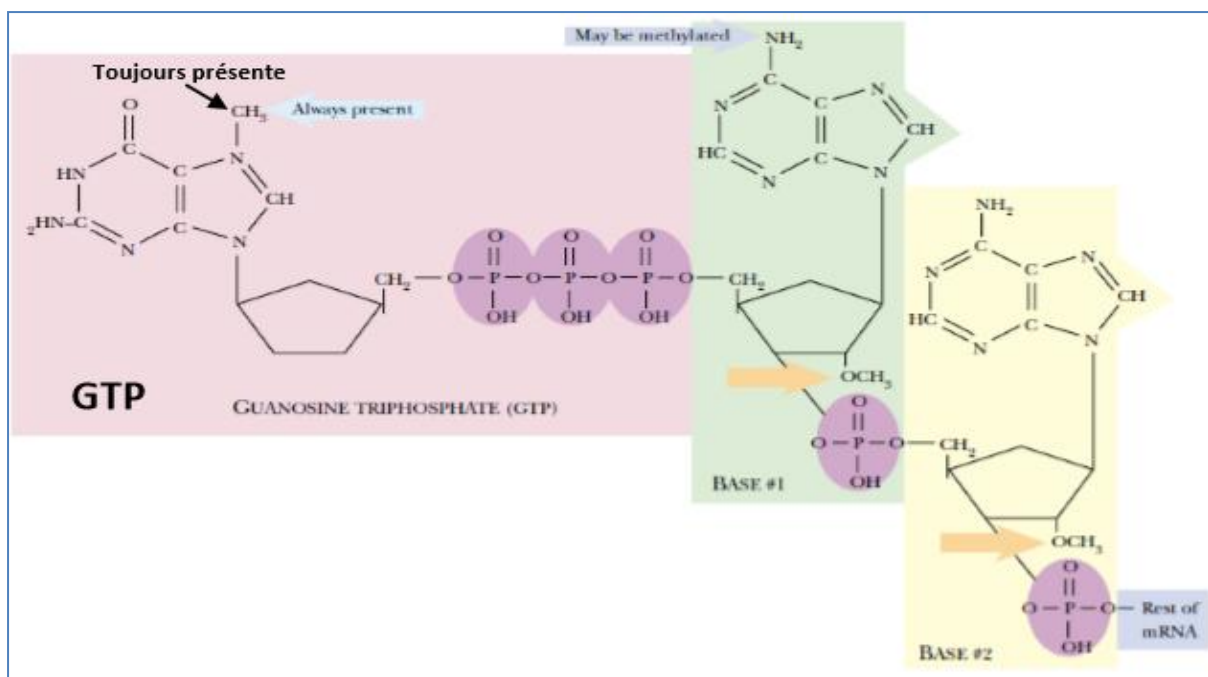


Figure 65. Capping of mRNA

➤ **Addition of the cap at the 5' end of nascent RNAs:**

The first two reactions are catalysed by a capping enzyme that associates with RNA polymerase shortly after the initiation of transcription. Two different methyltransferases catalyse reactions 3 and 4. S-adenosylmethionine (S-Ado-Met) is the source of the methyl group CH₃.

➤ **Functions:**

- Protects from 5' exonucleases
- Translation: Recognized by ribosome (cap-dependent translation)
- mRNA stability
- mRNA export from nucleus

B. Polyadenylation:

At the 3' OH end, a poly A polymerase adds between 150 and 200 A residues.

Once transcription is complete, the mRNA will be cleaved 15-30 nucleotides downstream of the signal sequence, then AMPs will be added by a poly(A) polymerase to form a poly(A) tail.

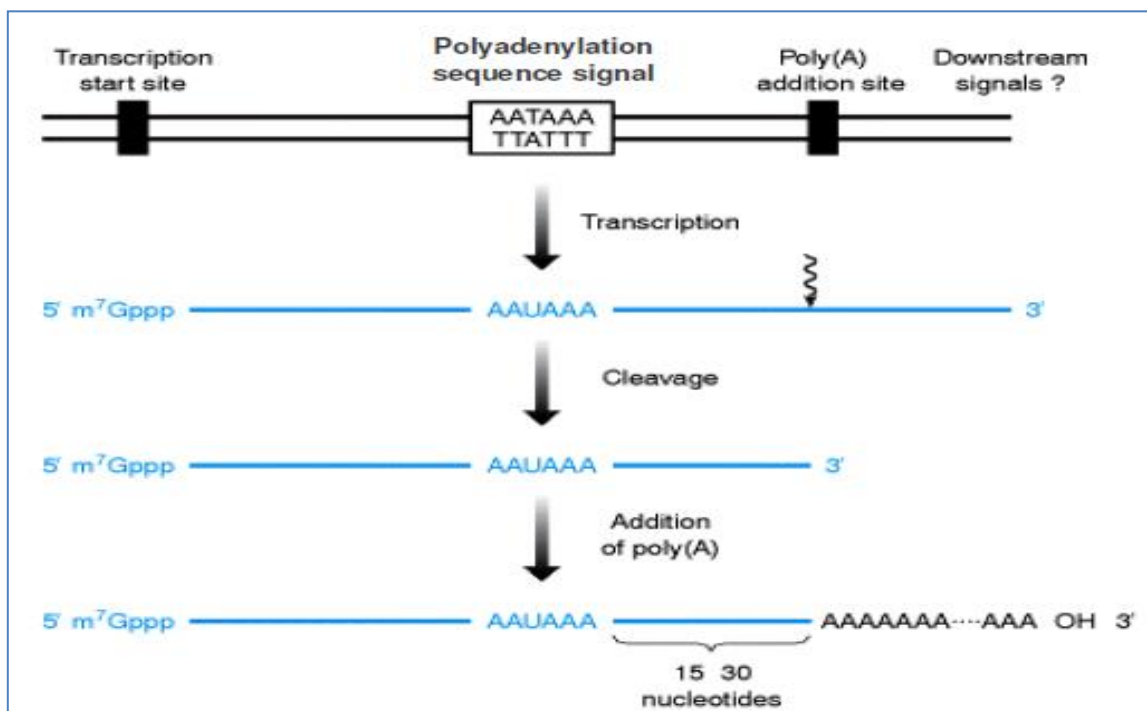


Figure 66. Polyadenylation

➤ **Functions of poly(A) tail:**

- mRNA stability
- Protects from 3' exonucleases
- Regulation: Tail length regulated (developmental, cell-type specific)

B. Splicing

consists of removing introns and retaining only the exons. It takes place in the nucleus after the completion of transcription.

Two types are distinguished:

- **Constitutive splicing:** all introns are removed
- **Alternative splicing:** exons are also excised differentially, producing mRNAs of different sizes coding for proteins with similar functions (isoforms)

➤ Principle of splicing:

The enzymatic complex responsible is the **spliceosome = proteins + snRNA**.

The snRNAs recognize particular sequences at the junction of introns and exons; in fact, they are complementary to these sequences.

Next, the spliceosome performs a cleavage at the intron–exon junction (5' side). The free end of the intron forms a bond with a particular A nucleotide (branch site): a lariat intermediate is obtained. The other end of the intron is cleaved. The two free ends of the exons are joined together.

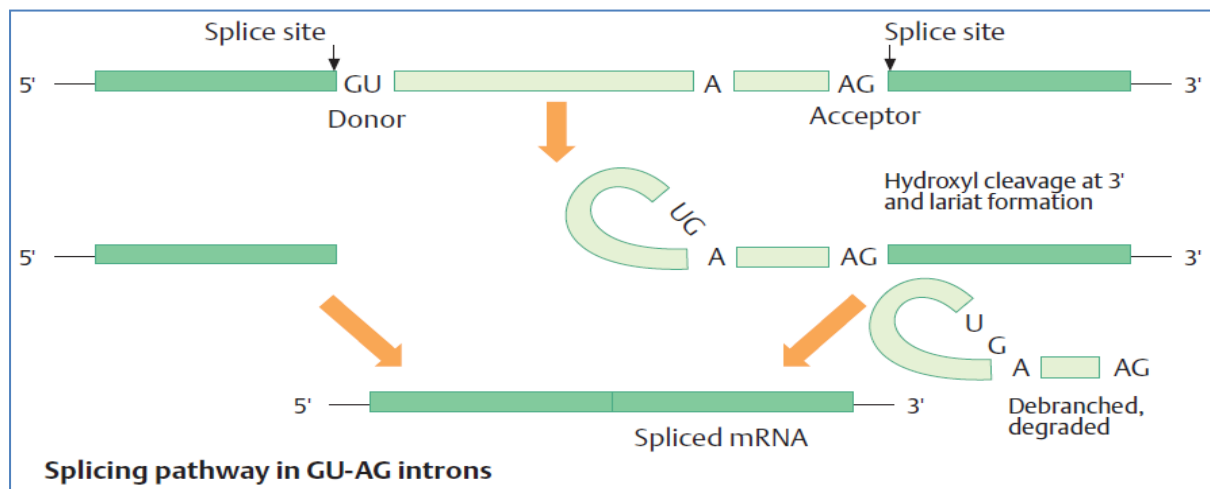


Figure 67. Splicing

4.1.2.2. Transfer RNA maturation tRNA:

These modifications concern prokaryotic and eukaryotic tRNA.

tRNA are RNA containing between 70 and 90 nucleotides. There are roughly as many tRNA as amino acids.

tRNA genes form sequences aligned in tandem. Most of the time, a long RNA precursor containing multiple tRNAs is synthesized (polycistronic RNA). Then this precursor is cleaved by a ribonuclease. These RNAs undergo chemical modifications:

- Some bases are methylated or, on the contrary, demethylated.
- There are also highly modified (atypical) bases.
- tRNA precursors undergo cleavages: fragments are removed, for example at the anticodon loop.
- The 3' end is always identical CCA–OH, these 3 nucleotides are added after transcription. This is the site where the amino acid will be attached.
- There is also the acquisition of a 3D structure: a good portion of the bases are paired, forming 3 hairpins (cloverleaf structure). This can be referred to as a secondary structure. The tertiary structure: the arms fold, and the tRNA acquires an L-shape. Hydrophobic interactions stabilize this structure. It is noted that the 3' CCA end is opposite the anticodon loop.

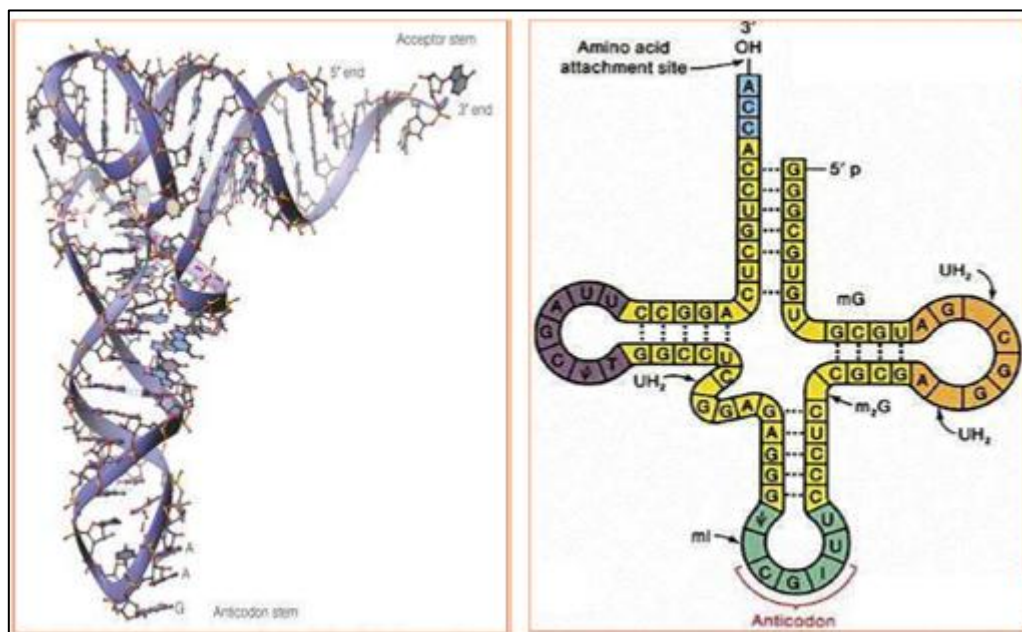


Figure 68. Structure of tRNA.

➤ **Activation of transfer RNAs:**

Activation of transfer RNAs means the association between the tRNA and its amino acid. This process involves the recognition of the amino acid corresponding to the tRNA anticodon.

The activation is carried out by **aminoacyl-tRNA synthetase**. This enzyme catalyzes the formation of the covalent bond between the tRNA and its amino acid. There are about twenty enzymes, each specific to a tRNA–aa pair:

- **Step 1:** the amino acid reacts with an ATP at its COOH group. Energy storage.
- **Step 2:** bond between the COOH of the amino acid and the 3'OH end of the tRNA. The breaking of the ATP–aa bond provides the energy for the aa-tRNA bond.

This is a crucial step: it is the true anticodon–aa association. Furthermore, it is a very faithful association thanks to the properties of the enzyme, which has:

- a recognition site for the anticodon
- a recognition site for the amino acid
- a catalytic site where the bond is formed
- a verification site – correction: the bond is hydrolyzed if the amino acid is too bulky ... it checks the nature of the amino acid mainly through weak interactions.

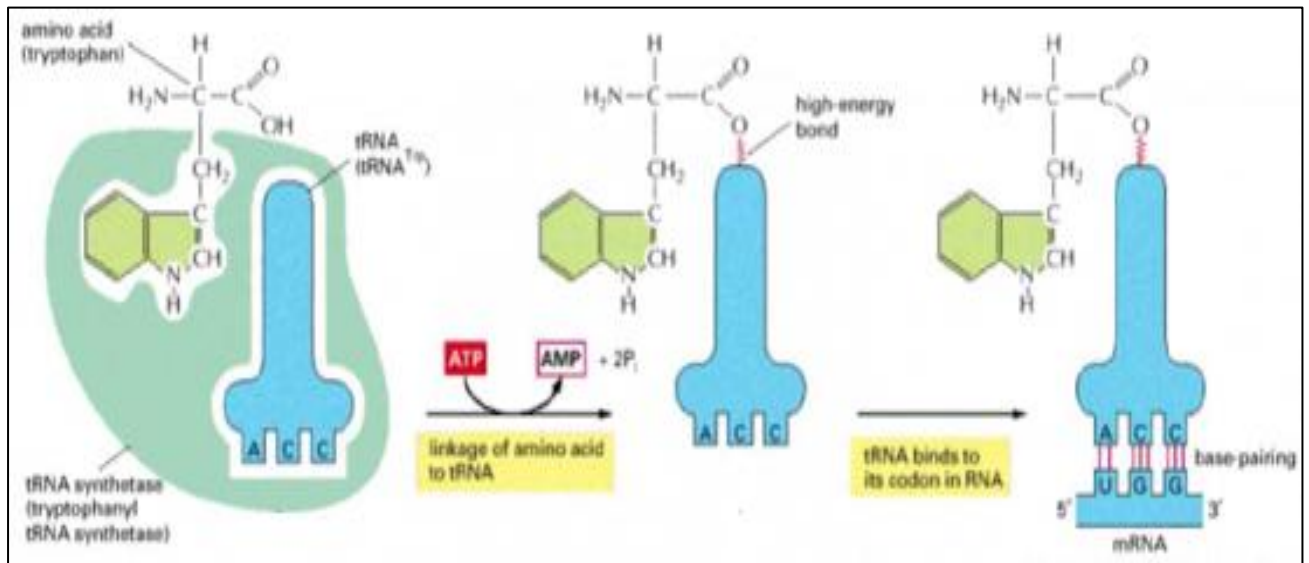


Figure 69. tRNA activation.

4.1.2.3. Ribosomal RNA maturation rRNA:

80 to 90% of synthesized RNA are rRNA.

In eukaryotes, rRNA genes form tandemly repeated families. Thanks to these many copies, the cell produces large amounts of rRNA. There are 4 types of rRNA: 18S, 5.8S, and 28S rRNAs synthesized via a single precursor which is then cleaved into 3 pieces by the RNA polymerase I. The 5S rRNA is synthesized separately by RNA pol III.

There are chemical modifications such as methylations on the ribose at specific positions. Then there is the acquisition of a spatial structure: formation of double-stranded loops through pairing. Then these loops themselves fold according to a complex organization. However, RNA corresponds to 2/3 of the ribosomes, very large structures: molecular weight in Megadalton, the rest being proteins (synthesized in the cytoplasm and imported into the nucleus, where assembly will occur):

- 5.8S, 28S, and 5S rRNA associate with about 50 proteins, resulting in the large subunit (60S).
- 18S rRNA associate with about 33 proteins and form the small subunit (40S)

Note: The structure of a prokaryotic ribosome: $30S + 50S = 70S$.

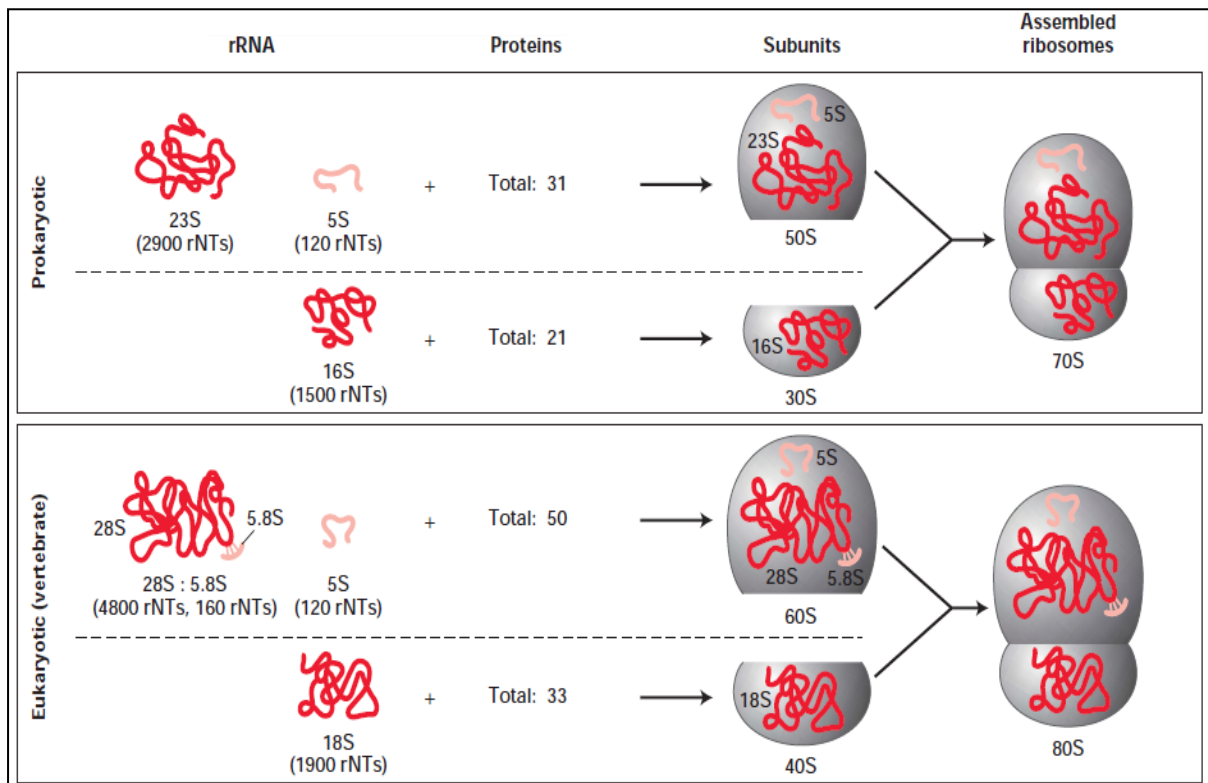


Figure 70. Prokaryote and eukaryote ribosomes

4.2. Translation and Protein Maturation

4.2.1. Definitions

Genetic translation is the stage of protein synthesis by ribosomes, based on the genetic information contained in mRNA. It is polarized: ribosomes move in the 5' to 3' direction on the mRNA, and synthesize the corresponding polypeptide from the NH₂ terminal end to the COOH terminal end.

The mRNA sequence is decoded by three nucleotides (codon) which correspond to a particular amino acid (aa) or to stop signals.

Among the main characteristics of the genetic code, we can note:

- It is redundant: an amino acid can be specified by several triplets.
- Some codons are stop codons.
- It is unambiguous: a given triplet specifies only one aa or is a stop codon.
- It is non-overlapping: a nucleotide belongs to one codon only.
- It is universal: it is the same for all species, except for some differences, for example between the nuclear system and the mitochondria.

It follows from all of these properties that an mRNA has 3 potential reading frames. But only one of these reading frames allows the correct protein to be made.

Note: The expression of genetic information is a rapid phenomenon; in bacteria, the association of amino acids can reach 350 aa per minute. It is also amplified because a gene can be transcribed multiple times at the same time to produce multiple mRNAs, and likewise, an mRNA can be read by multiple ribosomes simultaneously (in humans, an mRNA can carry 5 to 20 successive ribosomes).

Nucleotide base					
First	Second				Third
	Uracil (U)	Cytosine (C)	Adenine (A)	Guanine (G)	
Uracil (U)	F Phenylalanine (Phe)	S Serine (Ser)	Y Tyrosine (Tyr)	C Cysteine (Cys)	U
	F Phenylalanine (Phe)	S Serine (Ser)	Y Tyrosine (Tyr)	C Cysteine (Cys)	C
	L Leucine (Leu)	S Serine (Ser)	Stop Codon	Stop Codon	A
	L Leucine (Leu)	S Serine (Ser)	Stop Codon	W Tryptophan (Trp)	G
Cytosine (C)	L Leucine (Leu)	P Proline (Pro)	H Histidine (His)	R Arginine (Arg)	U
	L Leucine (Leu)	P Proline (Pro)	H Histidine (His)	R Arginine (Arg)	C
	L Leucine (Leu)	P Proline (Pro)	Q Glutamine (Gln)	R Arginine (Arg)	A
	L Leucine (Leu)	P Proline (Pro)	Q Glutamine (Gln)	R Arginine (Arg)	G
Adenine (A)	I Isoleucine (Ile)	T Threonine (Thr)	N Asparagine (Asn)	S Serine (Ser)	U
	I Isoleucine (Ile)	T Threonine (Thr)	N Asparagine (Asn)	S Serine (Ser)	C
	I Isoleucine (Ile)	T Threonine (Thr)	K Lysine (Lys)	R Arginine (Arg)	A
	Start (Methionine)	T Threonine (Thr)	K Lysine (Lys)	R Arginine (Arg)	G
Guanine (G)	V Valine (Val)	A Alanine (Ala)	D Aspartic acid (Asp)	G Glycine (Gly)	U
	V Valine (Val)	A Alanine (Ala)	D Aspartic acid (Asp)	G Glycine (Gly)	C
	V Valine (Val)	A Alanine (Ala)	E Glutamic acid (Glu)	G Glycine (Gly)	A
	V Valine (Val)	A Alanine (Ala)	E Glutamic acid (Glu)	G Glycine (Gly)	G

Figure 71. The genetic code

4.2.2. The different partners of translation

- mRNA.
- tRNAs.
- Ribosomes.

A ribosome can bind an mRNA and 3 tRNA at the A, P, and E sites (A for aminoacyl-tRNA, P for peptidyl-tRNA, and E for exit)

More than one ribosome can bind to an mRNA → **a polyribosome (polysome)** → synthesis of several protein molecules

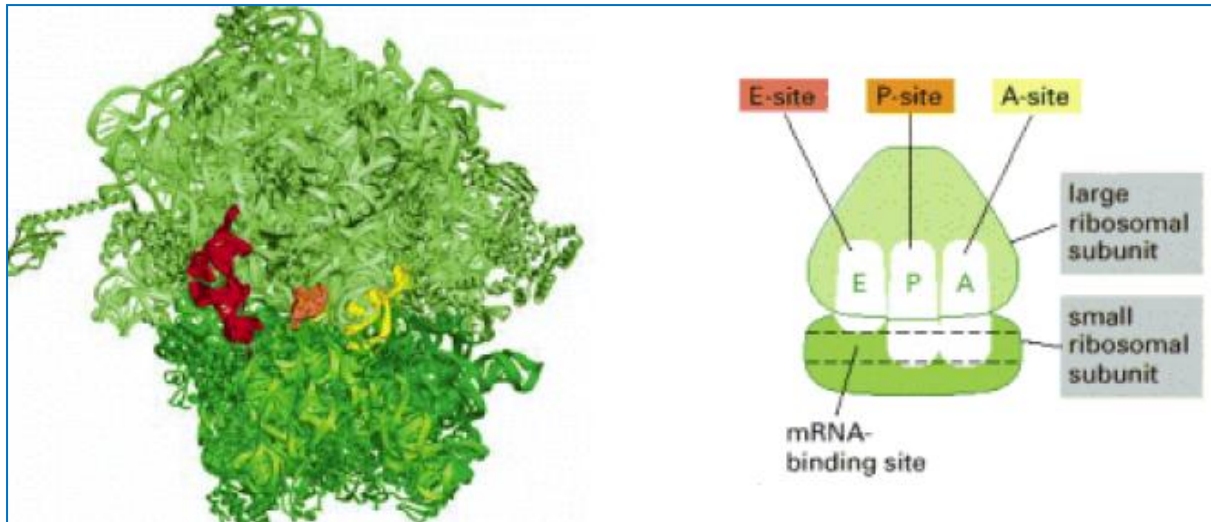


Figure 72. Structure of ribosome

4.2.3. Mechanism of translation in prokaryotes:

Translation occurs in three steps: initiation, elongation, and termination

4.2.3.1. Initiation:

The signal for the start of translation is very important: it determines the reading frame. The initiator codon is generally AUG (corresponding to methionine). A special tRNA, called initiator, is necessary: it has the anticodon for methionine UAC and carries a modified methionine: **a formylated methionine** (which will be removed at the end of translation).

The first step of initiation is the binding of the mRNA to the 30S subunit. The factors IF1 and IF3 stimulate this binding. Then the formyl-Met tRNA binds to the corresponding codon (AUG). This step is activated by IF2. Synthesis must begin at the 5' end and with the initiator codon.

The true initiator codon is associated with a particular sequence located a few nucleotides upstream (a purine-rich sequence, called the Shine-Dalgarno sequence). This sequence is actually complementary to a portion of the rRNA (16S) of the small subunit: it is thus recognized. **The 30S initiation complex** is then formed.

Finally, the large subunit (50S) assembles. This requires the hydrolysis of a GTP (bound to IF2). Then IF2, IF3, and IF1 are released. This is **the 70S initiation complex**. The initiator tRNA is then at the P site.

Note: polycistronic mRNA: each gene has own Shine-Dalgarno and start codon

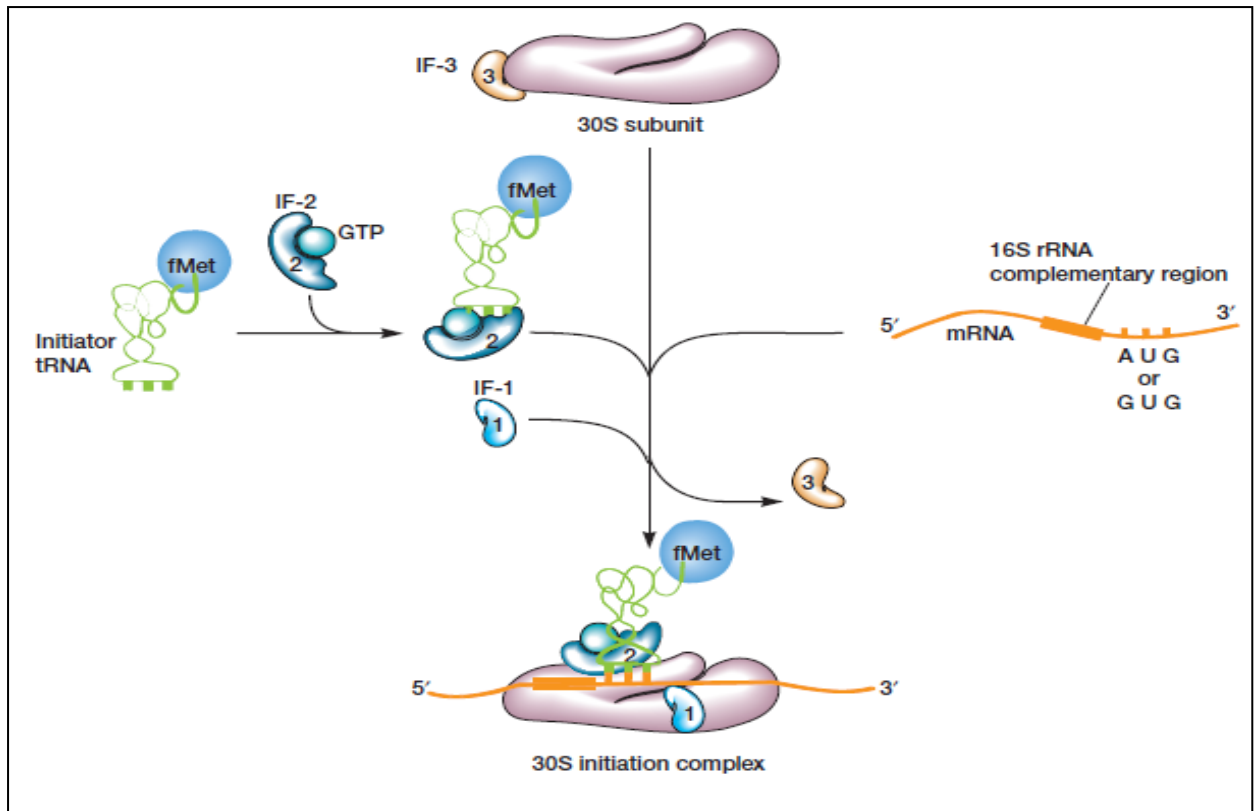


Figure 73. The 30S initiation complex

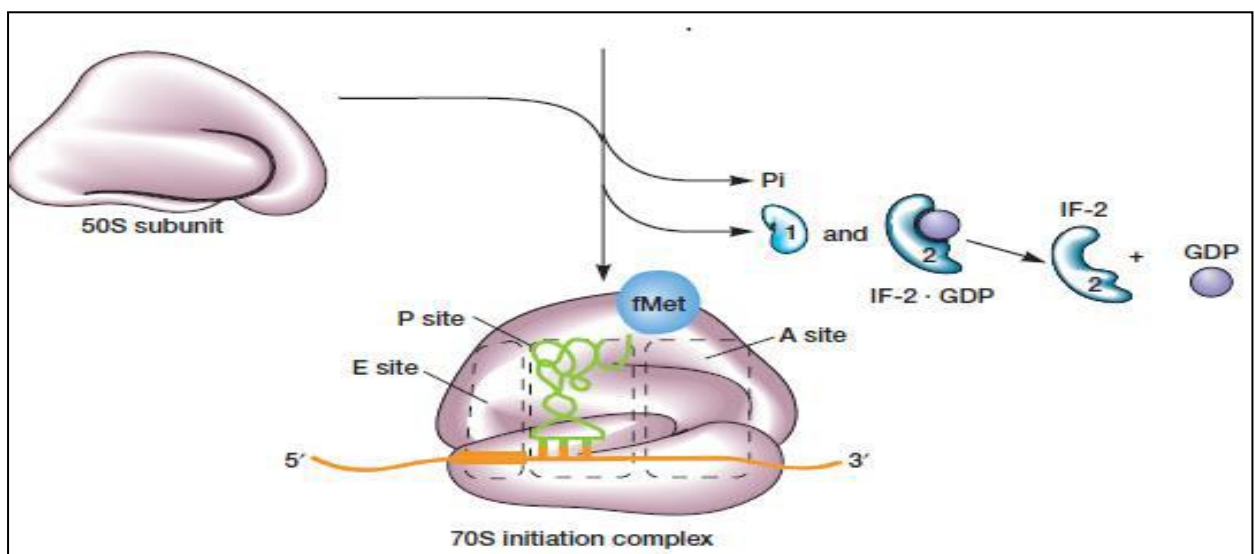


Figure 74. The 70S initiation complex

4.2.3.2. Elongation:

After initiation, the amino acids are added according to a 3-step cycle.

Our starting point is a peptide that already has 3 amino acids. The last one is still attached to its tRNA and is present at the P site. The peptide bond between amino acid 2 and amino acid 3 is formed. The tRNA carrying the next amino acid arrives at the A site. There is a codon-anticodon interaction. In fact, it is a trial-and-error system: if there is complementarity,

the tRNA-amino acid complex remains bound. The binding of the tRNA is facilitated by an elongation factor protein (EF-Tu). EF-Tu has a binding site for GTP and also binds to the tRNA-amino acid. When the tRNA-amino acid positions itself at the A site, the GTP is hydrolyzed and EF-Tu is released. This process adds amino acids, but: if the tRNA is incorrect, the binding is weak and does not remain. Then another one is tried. If the tRNA is correct, the codon-anticodon bond is strong enough, and the tRNA remains. This is an error-checking system: the fidelity is high. It is observed that the process is energy-intensive.

The second step: the breaking of the covalent bond between the tRNA and the amino acid at the P site: exergonic; the energy released allows the formation of the peptide bond between the COOH of amino acid 3 (P site) and the NH₂ of amino acid 4 (A site) by the peptidyl transferase activity of the large subunit. At the same time, a conformational change allows the shift of the large subunit. This is the translocation favored by the EF-G factor, which uses GTP hydrolysis. The tRNA at the A site moves to the P site, and the one at the P site moves to the E site.

The third step: the small subunit shifts by 3 nucleotides. The ribosome is ready to receive a new tRNA at the A site...

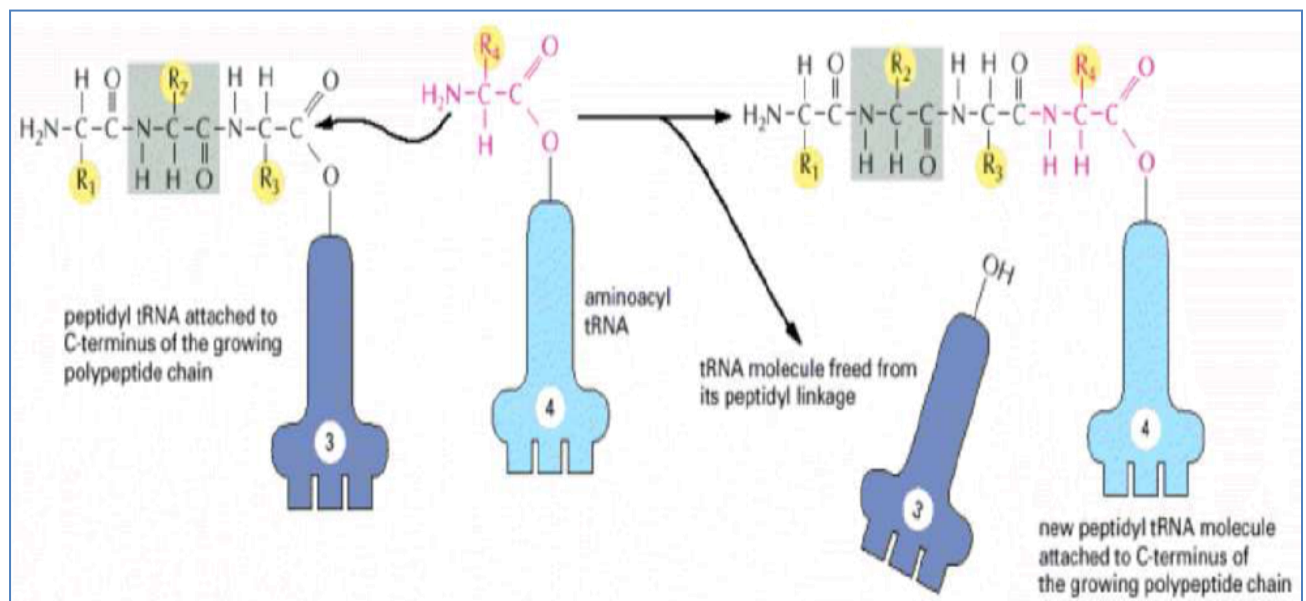


Figure 75. Formation of peptide bond

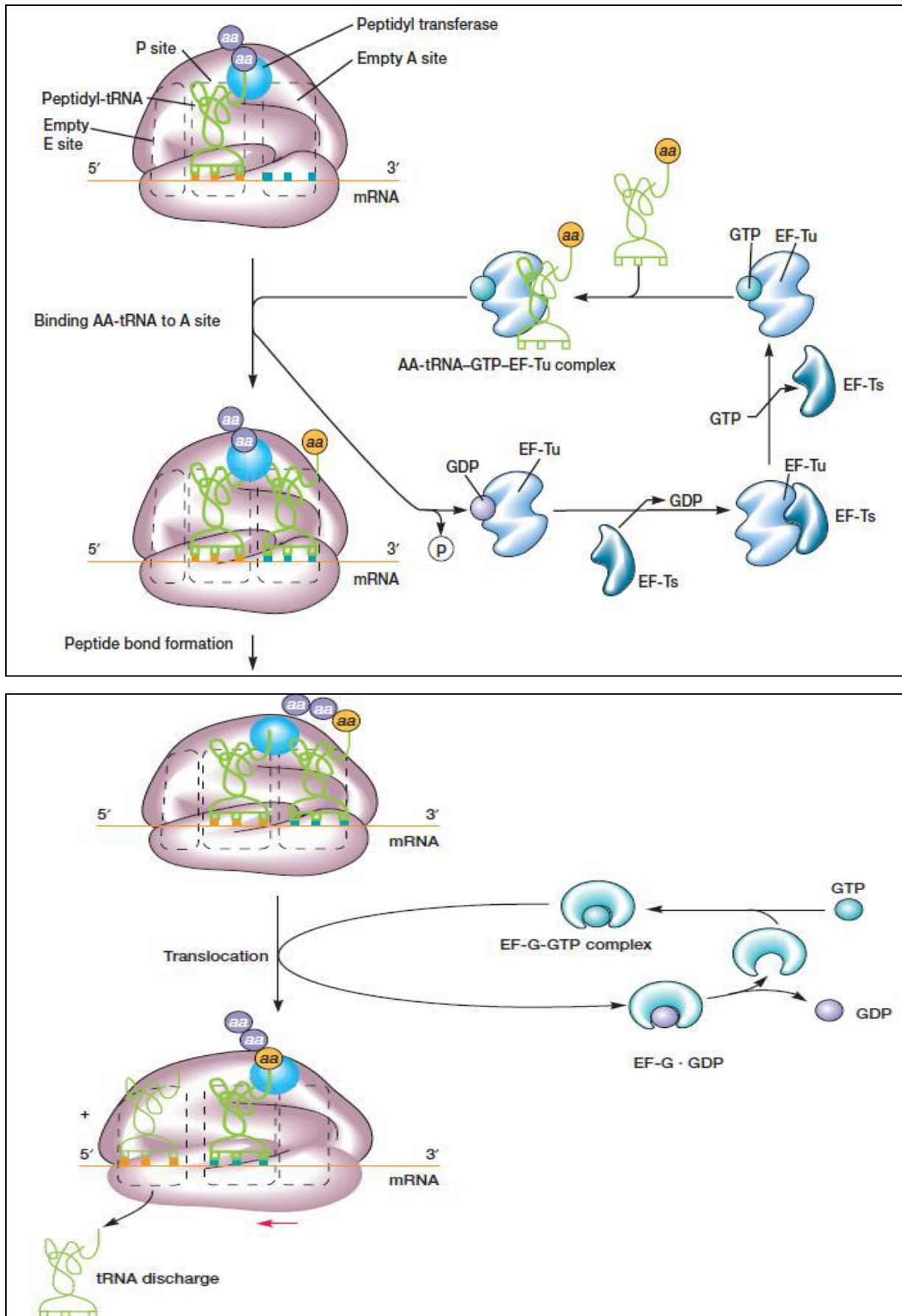


Figure 76. Translation elongation in prokaryote

4.2.3.3. Termination:

The end of the mRNA coding sequence corresponds to one of the three stop codons (UAA, UAG, UGA): they are not recognized by any tRNA. However, they are recognized by release factors RF.

The release factor RF has a 3D structure that mimics a tRNA. Thus, it can bind to the A site facing the stop codon. A GTP is hydrolyzed, which causes the addition of a water molecule to the protein. The last amino acid is released from its tRNA. The polypeptide chain separates from the ribosome, which dissociates from the mRNA, and the two subunits separate.

At the end of translation, the primary structure of the protein has been established.

4.2.4. Mechanism of translation in eukaryotes:

Translation occurs in the same way as in prokaryotes with initiation is similar to that in prokaryotes in several aspects: both use an initiation codon and an initiator tRNA and both use initiation factors to form a complex with the small ribosomal subunit that binds the mRNA before the addition of the large subunit. However, eukaryotes use a completely different method to recognize the mRNA and the initiation codon.

In eukaryotes, the small subunit is already associated with an initiator tRNA. Unlike in prokaryotes, the binding of the initiator tRNA to the small subunit always precedes the association with the mRNA.

The recognition of mRNA involves several auxiliary factors. The ribosome recruited at the 5' cap of the mRNA scans the mRNA in the 5' → 3' direction until it reaches the **Kozak sequence 5'CCACC3'**; this sequence allows the identification of the start codon (AUG).

Finally, the large subunit of the ribosome is recruited after base pairing of the initiator tRNA with the start codon. At the end of a translation cycle, the eukaryotic ribosome dissociates into its constituent subunits.

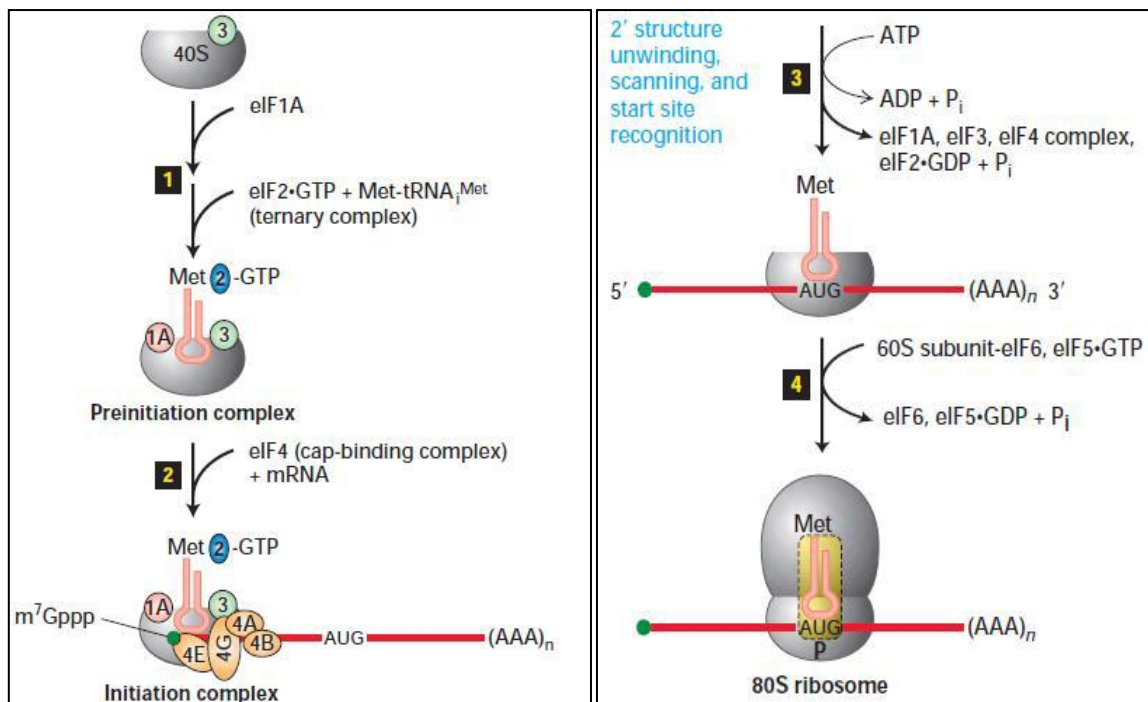


Figure 77. The 80S initiation complex

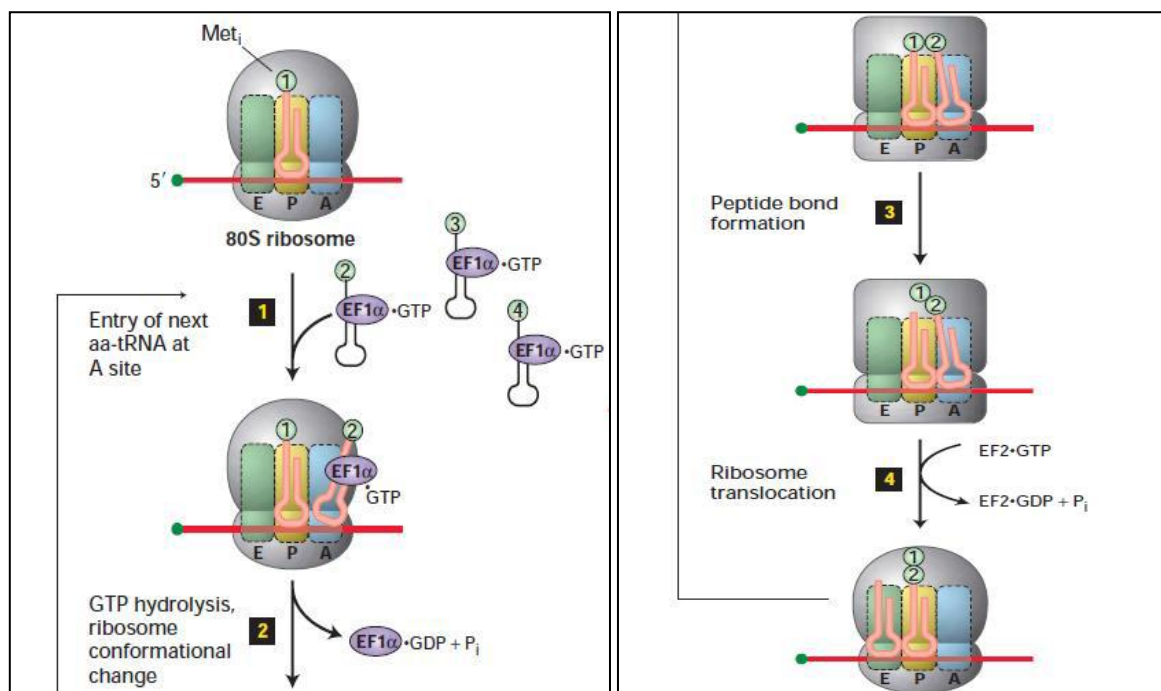


Figure 78. Translation elongation in eukaryote

4.2.5. Protein Maturation

4.2.5.1. Protein folding

The translation of information from a gene occurs in two steps:

- Through the genetic code starting from codons to obtain a sequence in aa.

- Through the folding of the aa sequence which will give a three-dimensional structure to the protein, providing the formation of active sites and stability.

Folding takes place when the polypeptide sequence emerges from the ribosome.

Newly synthesized polypeptide chains by the ribosomes can undergo post-translational modifications by chemically modifying the side chains of certain aa. This may include, for example, covalently attaching functional groups such as phosphate (phosphorylation), acetate (acetylation), or methyl (methylation), as well as sugars, oligosaccharides, or polysaccharides (glycosylation), or lipids (prenylation).

Other forms of post-translational modifications include cleaving certain peptide bonds, for example, from a protein precursor to form a functional protein, as is the case with insulin, which is cleaved twice after the formation of disulfide bridges.

4.2.5.2. Protein targeting

Often the site of synthesis and the site of action of proteins are different. To reach their site of action (or targeting):

➤ **Cytosolic proteins:**

Are not localized in a particular organelle, remaining in the cytosol. They can have catalytic functions (glycolysis enzymes) and structural functions (cytoskeleton elements). They can be either free (soluble in the cytoplasm) but also attached to the inner surface of the plasma membrane.

➤ **Nuclear proteins:**

Are transported from the cytosol to the nucleus through the pores of the nuclear envelope. They are numerous: histones, DNA-binding proteins in the broad sense, ...

➤ **Proteins of semi-autonomous organelles:**

Which have their own genetic information, but this information is not sufficient to synthesize all their proteins (plastids, mitochondria). Some of their proteins are encoded by their own DNA and synthesized on-site, while most of their proteins are encoded by nuclear genes, translated in the cytosol, and then sent to the organelle.

➤ **Proteins of the reticuloendothelial system:**

Include the endoplasmic reticulum (ER), the Golgi apparatus, and the vesicles that derive from the Golgi apparatus (lysosomes). The proteins of these compartments are synthesized at the level of the rough ER and then directed to their final destination via the Golgi apparatus.

➤ **Secreted proteins:**

Are the same as the proteins of the ER system but pass entirely through the system to be released into the extracellular environment.

4.2.5.3. Antibiotics and Protein Synthesis

Antibiotics are molecules that specifically target bacteria, by either:

- blocking their growth (bacteriostatic antibiotics) or
- destroying them (bactericidal antibiotics).

They act by inhibiting an essential step in bacterial development: the synthesis of their cell wall, DNA, RNA, or proteins.

The selectivity of these agents is the result of differences between the 70S ribosomes of prokaryotes and the 80S ribosomes of eukaryotes. Since mitochondrial ribosomes are similar to prokaryotic ribosomes, these molecules can have some toxicity. They are mainly bacteriostatic.

Examples:

- **Aminoglycosides** disrupt protein synthesis at the ribosome.
- **Macrolides, lincosamides, and streptogramins** inhibit protein synthesis at the ribosome. They all bind to the 50S subunit.
- **Tetracyclines** bind to the 30S subunit and possibly to the 50S subunit. They block the elongation phase of protein synthesis.
- **Chloramphenicol** binds to the 50S ribosomal subunit, blocking translocation. A second binding site is located on the 30S subunit.
- **Fusidic acid** inhibits the elongation phase of protein synthesis: it blocks translocation.

4.3. Regulation of gene expression

4.3.1. Overview

- Not all genes of an organism are expressed at the same time.
- Gene expression regulation exists in both prokaryotes and eukaryotes.
- This regulation is essential for survival.
- Every step, from genes to proteins, can be regulated. Cells must precisely regulate gene expression during growth, differentiation, and response to stimuli.
- In prokaryotes: regulation allows adaptation to the environment.
- In eukaryotes: regulation allows specific gene expression in different cell types despite identical genetic material.

4.3.2. Definitions

- **Genetic regulation:** mechanisms allowing activation of necessary genes and repression of unnecessary ones
- **Structural gene:** codes for a protein (structural or enzymatic)
- **Regulatory gene:** codes for proteins regulating other genes
- **Transcription factor:** protein regulating transcription
- **Activator:** stimulates transcription
- **Repressor:** inhibits transcription
- **Operator:** DNA site where repressor binds
- **Trans-acting elements:** diffusible products (activators/repressors)
- **Cis-acting elements:** DNA sequences acting locally (promoter/operator)

4.3.3. Gene regulation in prokaryotes

The main goal of regulation in prokaryotes is to adapt to environmental and nutritional changes. Because the transcription and translation are coupled, regulation mainly occurs at transcription initiation

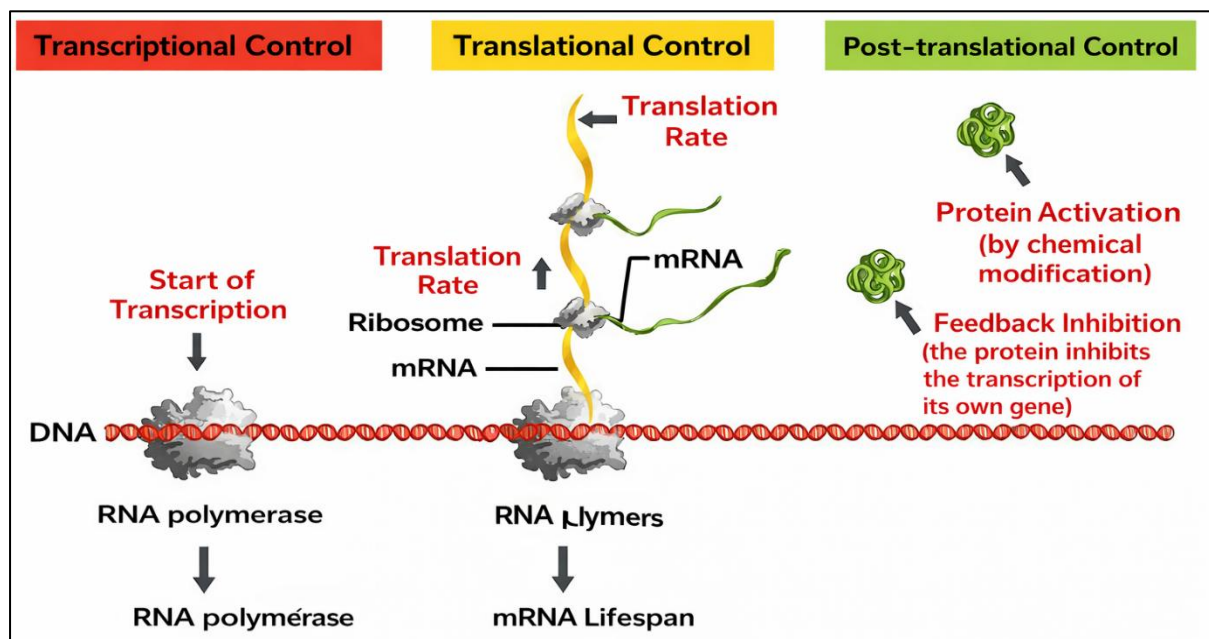


Figure 79. Main points of gene regulation in prokaryotes

4.3.3.1. Operons

A. Definition:

A group of genes co-transcribed together from a single promoter and co-translated into multiple proteins. This unit includes structural genes, one or more regulatory genes encoding regulatory proteins, and control elements present in the DNA sequence.

B. Types:

- **Inducible operons** code for enzymes of the catabolic pathway (degradation pathway).
Example: lactose operon.

- **Repressible operons** code for enzymes of the anabolic pathway (biosynthesis).
Example: tryptophan operon.

C. Modes of regulation:

Two modes of regulation of the expression of a target gene by a regulatory molecule:

- Positively: the interaction activates the transcription of the gene.
- Negatively: the interaction inhibits the transcription of the gene.
- **Negative Control**

- Inducible operon:

When the repressor protein binds to the operator, RNA polymerase can no longer initiate transcription, which prevents gene expression.

When the inducer binds to the repressor, the latter becomes inactivated and RNA polymerase can initiate transcription.

- Repressible operon:

When the co-repressor binds to the repressor protein, the latter is activated and binds to the operator; RNA polymerase can no longer initiate transcription, which prevents gene expression.

➤ **Positive Control**

- Inducible operon:

Activator helps RNA polymerase bind promoter

- Repressible operon:

Co-repressor can inhibit activator → no transcription

Summary:

Basic principle of transcription regulation: The binding of a specific protein to a given site on the DNA influences all the events that make up the assembly of the initiation complex and/or the initiation of RNA synthesis by RNA polymerase.

- Negative control → repressor blocks transcription
- Positive control → activator stimulates transcription
- Inducible operon → inducer activates transcription
- Repressible operon → co-repressor inhibits transcription

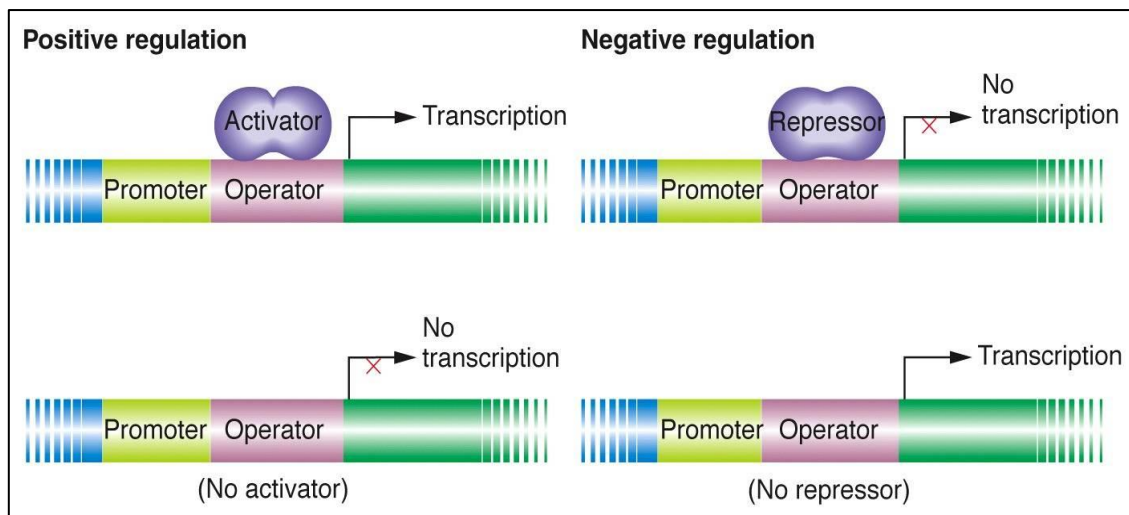


Figure 80. Positive and negative control

D. Example of a catabolic operon: the lactose operon of *E. coli*

The *lac* operon consists of three structural genes, *lacZ*, *lacY*, and *lacA*.

The *lacZ* gene encodes β -galactosidase: cleaves lactose into galactose and glucose, both of which are used by the cell as energy sources.

LacY codes for lactose permease, a membrane-bound protein that is part of the transport system to lactose into the cell.

LacA codes for a transacetylase which rids the cell of toxic thiogalactosides that get taken up by the permease.

Upstream of the *lac* operon is the regulatory gene that codes for Lac repressor. The Lac repressor is constitutively transcribed under control of its own promoter. In the absence of lactose, the Lac repressor binds as a tetramer to the operator DNA sequence. Because the *lac* operator sequence overlaps with the promoter region, the Lac repressor blocks RNA polymerase from binding to the promoter. As a consequence, transcription of the *lac* operon structural genes is repressed.

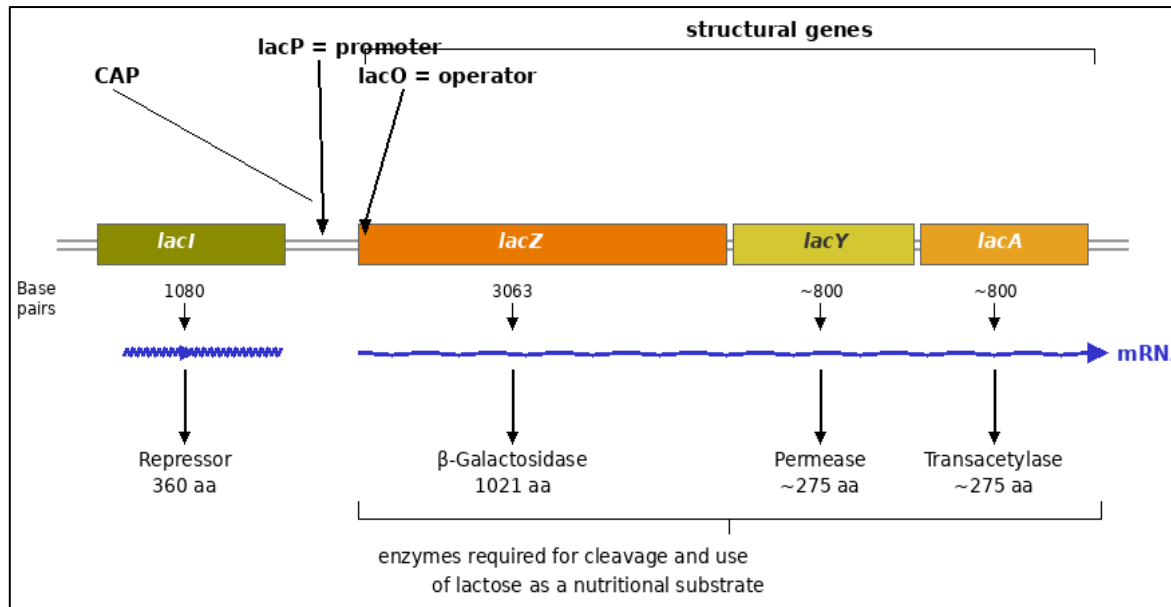


Figure 81. Lac operon

➤ **Conditions:**

➤ **Absence of lactose:**

transcription of the *lac I* gene and synthesis of the repressor

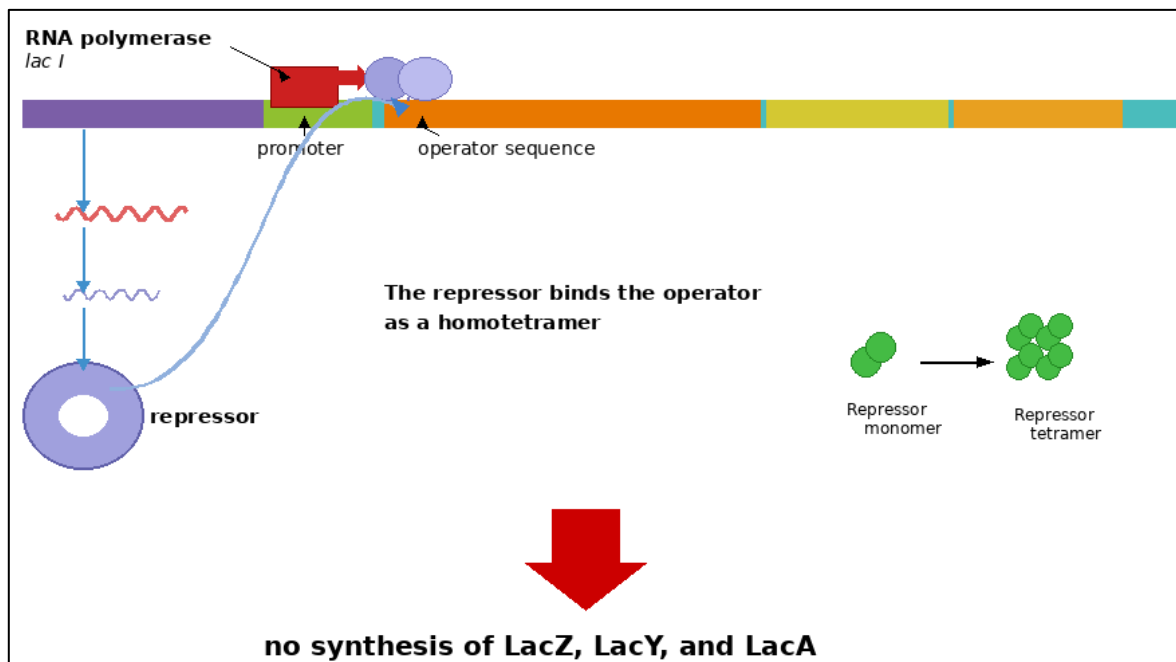


Figure 82. Lac operon in absence of lactose

➤ **Presence of lactose in the medium and absence of glucose:**

the binding of lactose to the repressor leads to the dissociation of the repressor-operator complex, allowing RNA polymerase to transcribe the structural genes

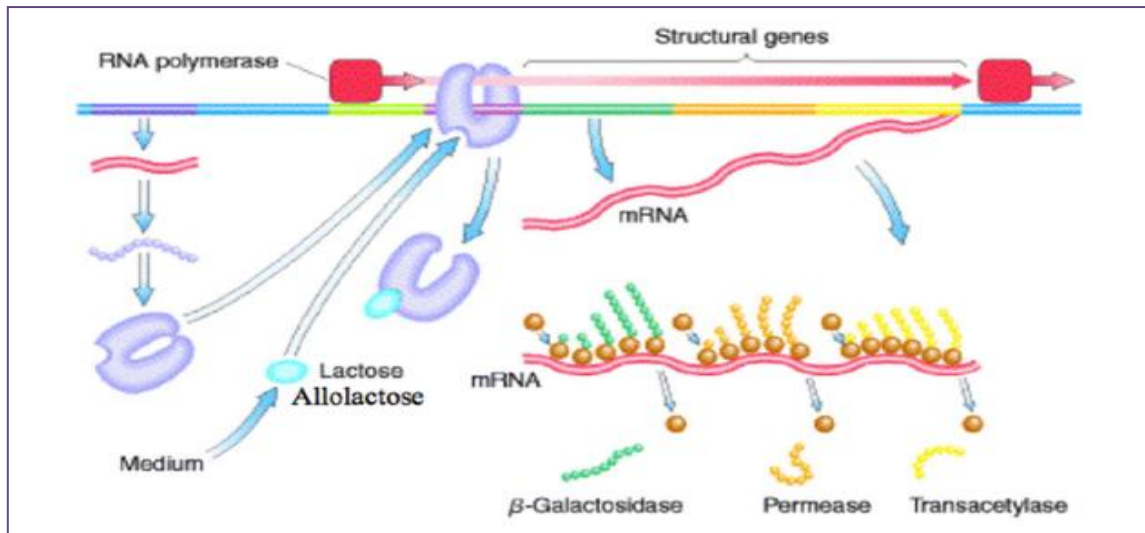


Figure 83. Lac operon in the presence of lactose in the medium and absence of glucose

➤ **Control of operon expression in the presence of glucose and lactose:**

The presence of an activator site (CAP site) of the lactose operon is activated by the binding of the CAP protein when it is associated with cAMP (CAP-cAMP). The presence of glucose is accompanied by a low concentration of cAMP: in the presence of glucose (low cAMP): CAP does not bind to its site

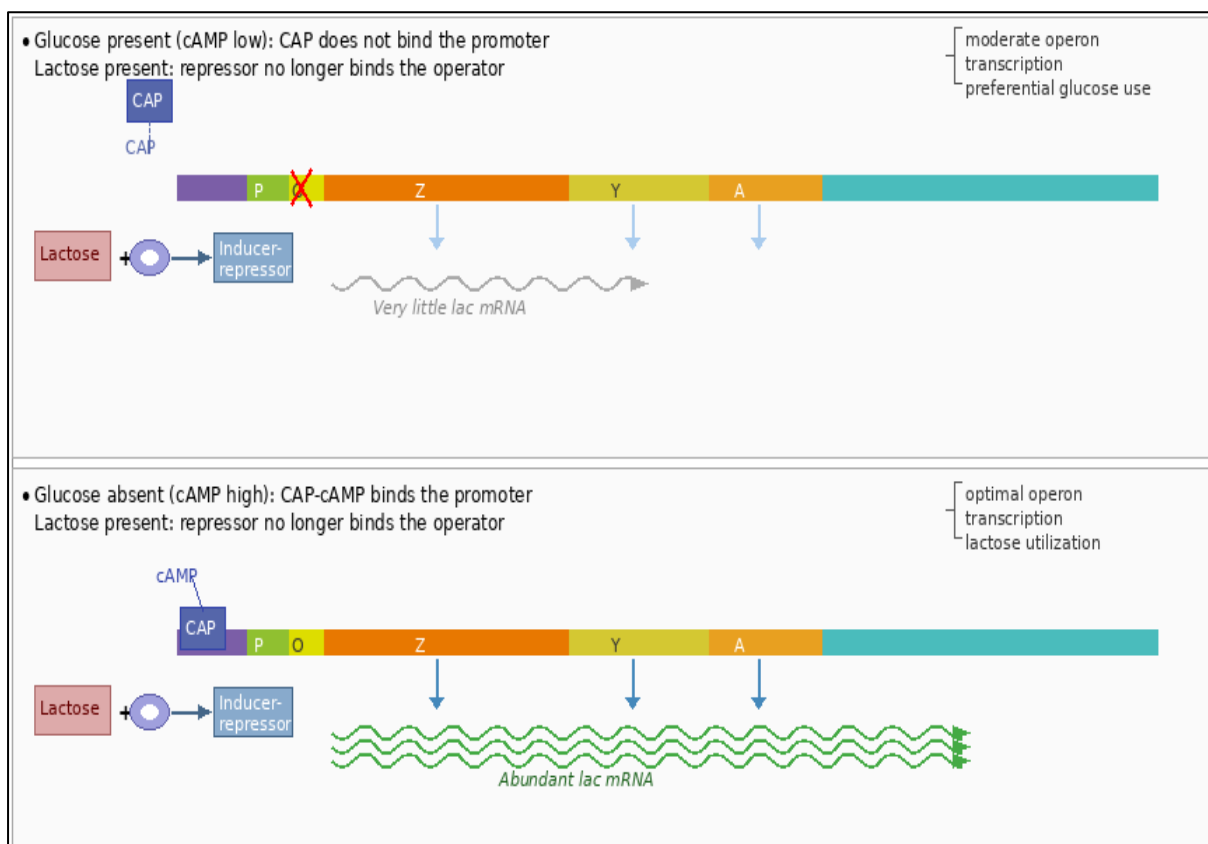


Figure 84. Control of operon expression in the presence of glucose and lactose:

E. Analysis of mutations of the lactose operon:

In 1961, Jacob and Monod presented the operon model which described the genetic regulation of lactose metabolism (Nobel Prize in Physiology or Medicine in 1965). They elucidated the structure and functioning of this operon by analyzing mutations in merodiploid strains (partially diploid) for the Lac genes constructed by bacterial conjugation: structural genes gene mutation, regulator gene mutation, operon mutation

a. Regulator gene mutation: *Lac I*⁺, *Lac I*⁻, *Lac I*^s

Three states of *lacI*:

- **wild-type (*Lac I*⁺)**
- **Null mutant (*Lac I*⁻):** The repressor is either not made or cannot bind to the operator. This leads to **constitutive** expression (genes are always ON).
- **Super-repressor (*Lac I*^s):** The repressor binds to the operator but cannot bind the inducer (lactose/allolactose). The repressor stays stuck on the DNA. This is **uninducible** and **dominant** over the wild-type.

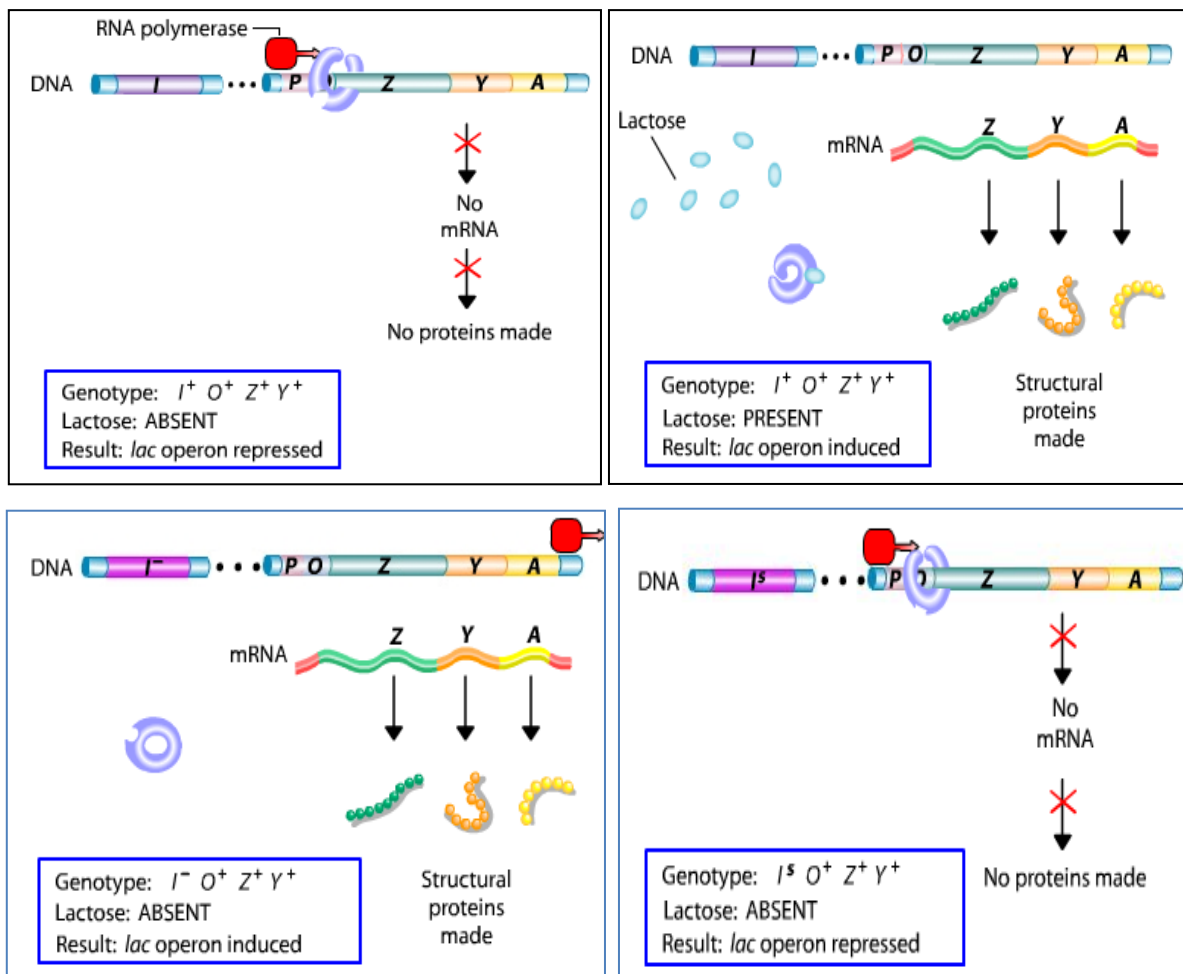


Figure 85. Control of operon expression in regulator gene mutation: *Lac I*⁺, *Lac I*⁻, *Lac I*^s

b. Operator mutation: O or O⁺, O^c

- **wild-type (O or O⁺)**
- **Operator constitutive (O^c):** The DNA sequence is altered so the repressor can no longer "see" or bind to it. The operon is expressed **constitutively**

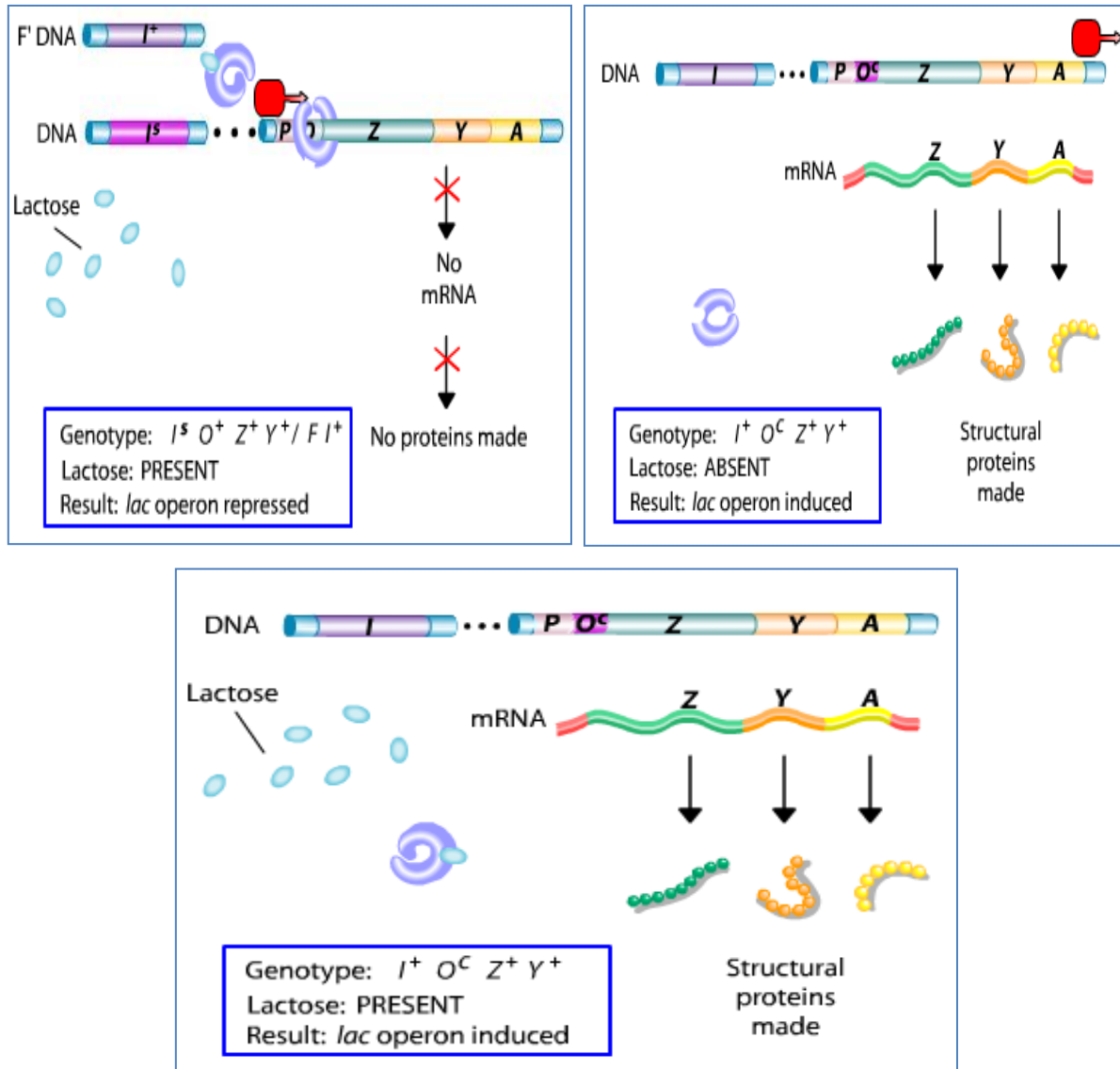


Figure 86. Control of operon expression in operon mutation: O, O^c

c. Promoter Mutations (P, P⁻)

- **wild-type (O or O⁺)**
- **Mutant promoter (P⁻):** Prevents RNA polymerase from binding. No transcription occurs even if lactose is present. The operon is **permanently OFF**.

d. Mutations in structural genes:

Studies are conducted on LacZ and LacY.

- LacZ + and LacY + are functional, whereas LacZ - and LacY - are not.

- The diploid LacZ + LacY - / LacZ - and LacY + is functional.
- Mutations in these genes are independent and affect only the concerned product.

F. Example of an anabolic operon: the tryptophan operon of *E. coli*

The *E. coli* **trp operon** contains the genes for the enzymes that the bacterium needs to make the amino acid tryptophan.

Like the *lac* operon, it is subject to negative control by a repressor. When the tryptophan concentration is high, the products of the *trp* operon are not needed any longer, and we would expect the *trp* operon to be repressed. That is what happens. The *trp* operon also exhibits an extra level of control, called attenuation, not seen in the *lac* operon.

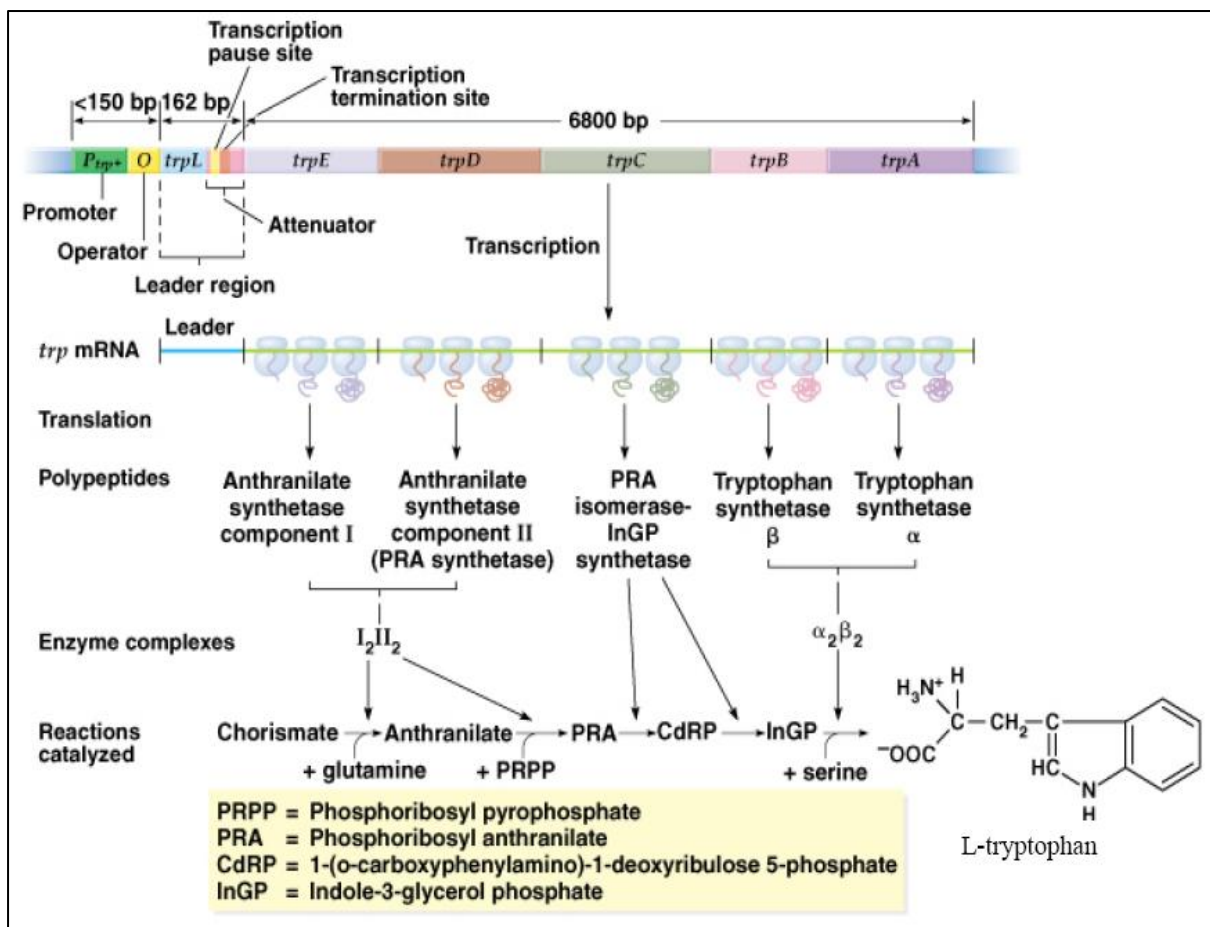


Figure 87. The tryptophan operon of *E. coli*

➤ Conditions:

Tryptophan acts as a **co-repressor**

- High tryptophan → operon OFF
- Low tryptophan → operon ON
- Regulation by **attenuation**: is a fine regulation mechanism where ribosome position determines transcription continuation:

- Low tryptophan → full transcription
- High tryptophan → early termination

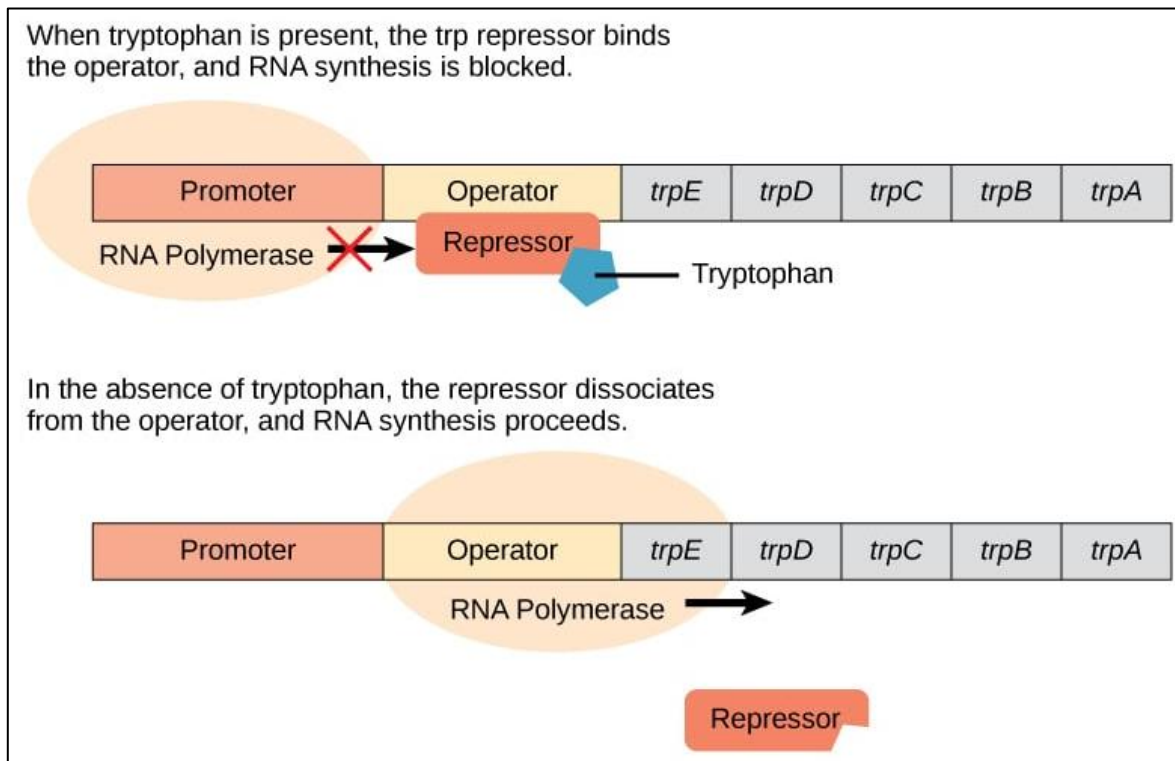


Figure 88. Regulation of the tryptophan operon of *E. coli* by co-repressor

➤ **Regulation by attenuation:**

Transcription is also controlled by attenuation, a process that results in the translation of a small polypeptide: When cells are deprived of tryptophan, the operon genes are expressed at the highest rate; When tryptophan deprivation is less severe, the operon genes are expressed at a level lower than the maximum.

A small coding sequence upstream of the trp operon contains two consecutive codons for tryptophan. When the amount of tryptophan in the cell is limited, the ribosome stops at these trp codons.

The ribosome's ability to read these codons regulates a stem-loop choice (terminator or anti-terminator). The position of the ribosome plays an important role in the attenuation phenomenon.

Attenuation: a mechanism that controls the ability of RNA polymerase to read an attenuator, which is a terminator placed at the beginning of transcription.

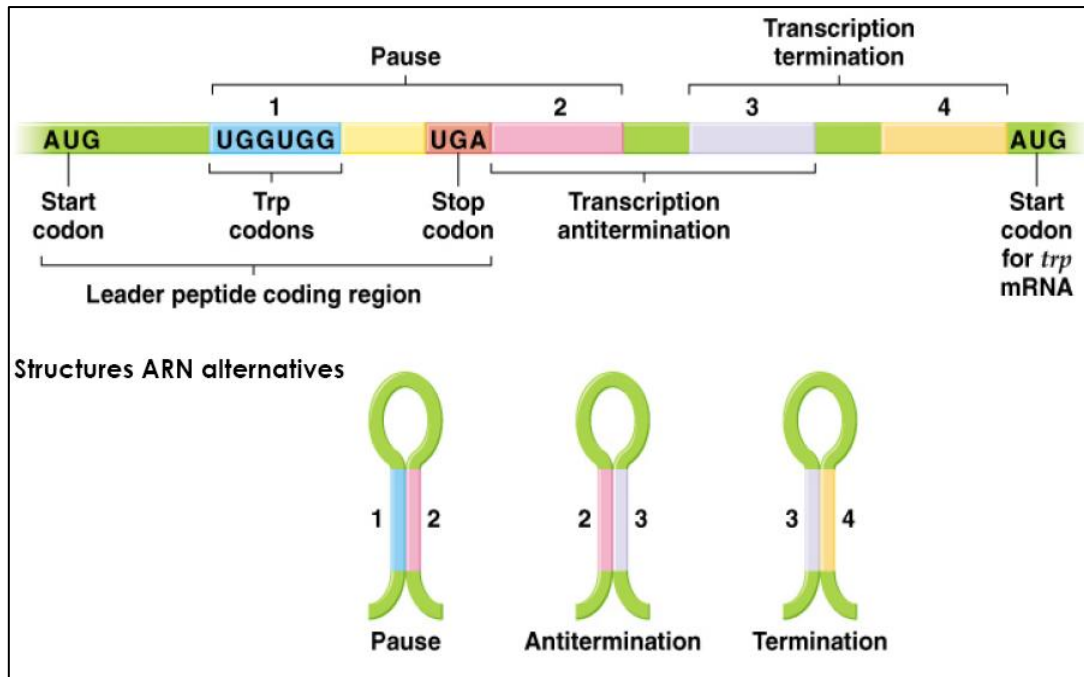


Figure 89. Regulation of the tryptophan operon of *E. coli* by attenuation

Tryptophan is absent (or in insufficient quantity):

Trp-tRNA are unavailable, the ribosome stops at the Trp codons which covers region 1. Region 2 pairs with region 3, so region 3 cannot pair with region 4. RNA polymerase continues to transcribe the entire coding sequence, allowing the complete synthesis of the mRNA.

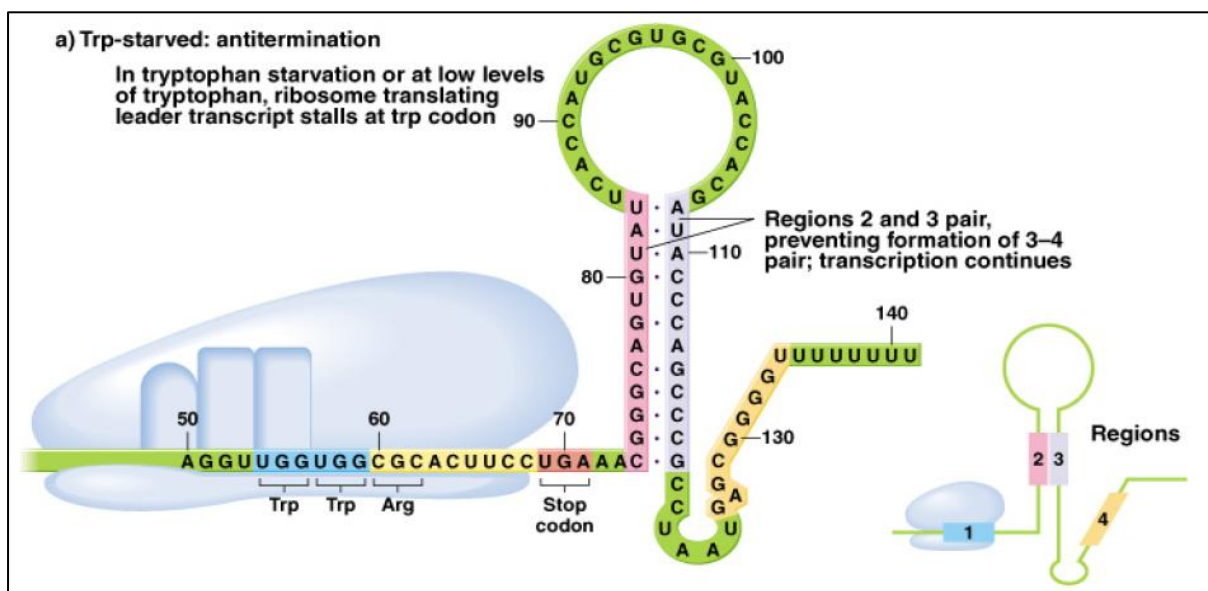


Figure 90. Regulation of the tryptophan operon of *E. coli* by attenuation in low levels of Trp

Tryptophan is abundant:

The ribosome does not stop at the trp codons; it continues to translate the leader sequence, stopping at region 2. Region 2 cannot pair with region 3; which then pairs with region

4. This 3-4 pairing constitutes the 'attenuator' sequence and functions as a termination signal. Transcription ends before RNA polymerase reaches the genes for tryptophan synthesis.

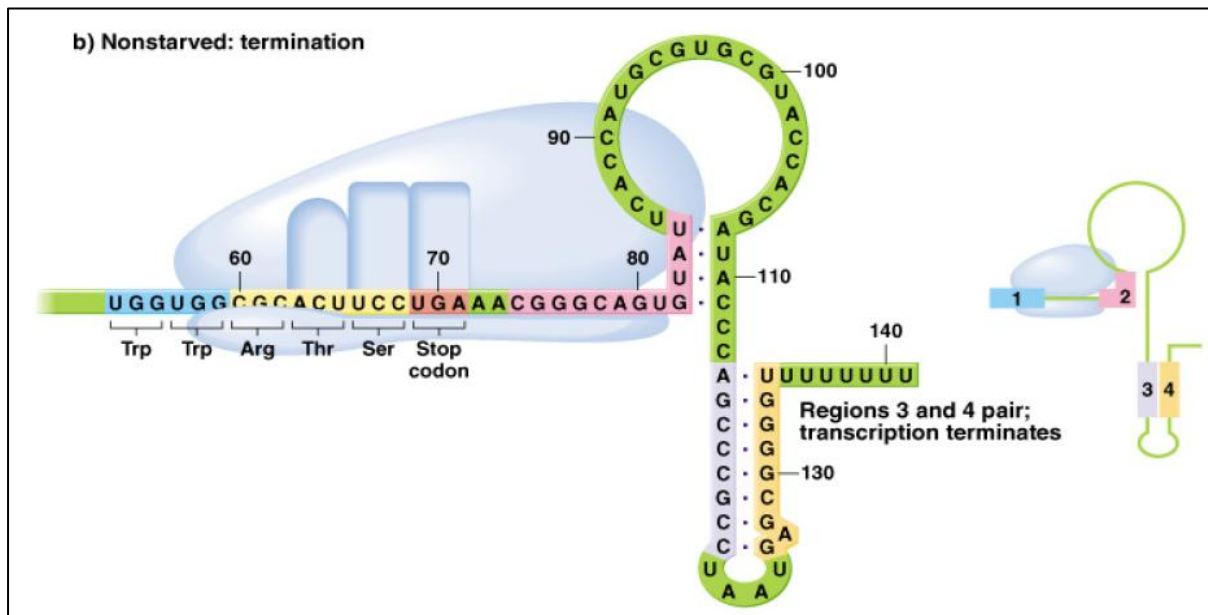


Figure 91. Regulation of the tryptophan operon of *E. coli* by attenuation in high levels of Trp

4.3.3.2. Regulons

A functional set of genes (possibly forming operons) that can be located at different positions along the double-stranded DNA and whose expression is regulated by the same transcriptional repressor.

4.3.3.3. Riboswitches

The 5' region (untranslated: 5' UTR) of an mRNA that can bind a small target molecule, the binding of which affects the activity of the gene(s) located downstream. The mRNA containing a riboswitch is directly involved in regulating its own activity depending on the presence or absence of the target molecule.

4.3.4. Gene Regulation in Eukaryotes

Precise regulation of gene expression is essential for cell differentiation and tissue formation, for the production and maintenance of cells and tissues.

Cells differentiate into specific cell types through combinations of expressed and repressed genes.

In eukaryotes, gene expression is controlled essentially at the following checkpoints:

- Activation of a gene's structure
- Transcription initiation
- Transcript maturation
- mRNA translation

➤ Post-translational modification

4.3.4.1. Gene activation: Genes are active in the cells that express them:

- DNA can exist in two structural forms: euchromatin or heterochromatin. In the latter form, genes are not accessible to polymerases.
- **Gene positioning along the chromosome:** during development, by altering which regions are condensed, the cell controls which genes are expressed, thereby enabling normal organism growth.
- **Chromatin remodeling:** by manipulating how DNA is wound around nucleosomes, gene expression is regulated.
- **Histone modification:** histones can be methylated, acetylated, and/or phosphorylated. This alters DNA expression at the level of these histones:

- **Acetylation** adding an acetyl group to lysine residues on H3 and H4 tails, neutralizing the positive charge. This weakens the electrostatic attraction between histones and DNA and making genes accessible to transcription. It is the most direct activating signal.
- **Methylation** is context-dependent: trimethylation of lysine 4 on histone H3 is a hallmark of active promoters, while trimethylation of lysine 9 and lysine 27 on histone H3 compact chromatin into heterochromatin.
- **Phosphorylation** adding a negatively charged phosphate group to serine or threonine residues, helps chromosome condensation.
- **Base methylation:** DNA can be methylated at the base level, particularly at GC repeats. Base methylation is recognized by enzymes and triggers DNA condensation, thereby leading to gene inactivation.

Note: Methylation is maintained during cell division

4.3.4.2. Initiation of transcription:

Eukaryotic genes contain regulatory sequences, often located upstream of the gene's promoter. The following terms are used:

- **Cis-regulator:** the DNA sequence that regulates the gene
- **Trans-regulator:** the transcription factor that binds specifically to the cis-regulator to activate or inhibit it is called a trans-regulator.

Cis-regulators are sequences of 6 to 15 nucleotides and come in two types:

- **Fixed-location elements:** promoters
- **Variable-location elements:** enhancers and silencers

Trans-regulators recognize these sequences and activate or inhibit their expression. These proteins possess a DNA-binding domain, an action domain (repression or activation), and often a domain for interacting with other ligands.

Transcription can also be regulated by extracellular signals (steroid hormones, thyroid hormones, etc., which act at the level of nuclear receptors).

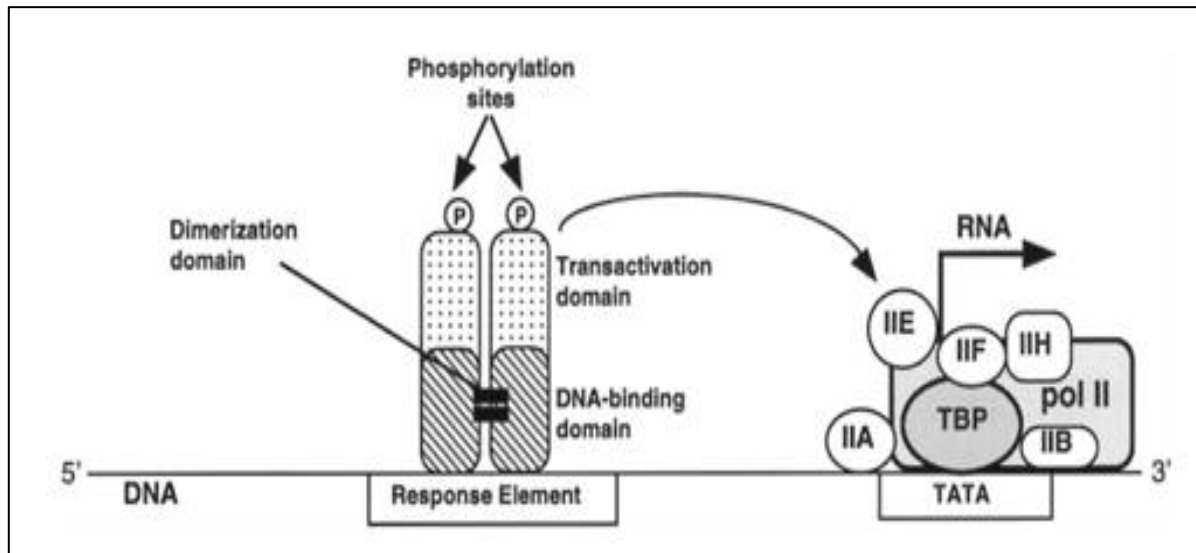


Figure 92. A generalized polymerase II promoter.

The figure shows two regulatory elements: the promoter (TATA) element and a hypothetical activator or response element. The promoter element is shown binding general factors, TFIIA, B, E, F, and H. Each of these transcription “factors” is actually composed of multiple individual proteins complexed together. This basal transcription apparatus recruits RNA polymerase II. It also forms the substrate for interactions with various activator proteins that bind to activator elements such as the one shown. Typical activator proteins contain DNA-binding domains, dimerization domains, and transcription activation domains. The latter interact with the basal transcription apparatus and may be modified by phosphorylation.

4.3.4.3. Transcript Maturation

➤ Alternative Splicing:

Through alternative splicing, a single gene can encode different (though similar) proteins. Depending on conditions on sex, on tissue, ...

➤ Modification of the cleavage-polyadenylation site:

the lifespan of the mRNA depends on the poly-A tail, because the longer it is, the longer the mRNA’s lifespan

4.3.4.4. Translation of mRNA

Translation of an mRNA begins at the first AUG codon encountered. However, a ribosome may “skip” an initiation codon.

Some proteins regulate their own translation by binding to the translation initiation sequence of the mRNA. The ribosome cannot then bind to the mRNA and therefore cannot translate it.

Another possibility involves the use of antisense RNA, which pairs with the mRNA to form a double-stranded RNA that prevents translation.

Finally, a third mechanism utilizes the ability of mRNA to form secondary structures.

Note: These translational regulatory mechanisms can also be found in prokaryotes.

4.3.4.5. Post-translational modification

This encompasses the various mechanisms involved in the maturation and activation of the proteins produced (active or inactive form of the protein depending on the cell's needs).

4.4. Mechanisms of gene regulation by extracellular signals

extracellular signals act directly or indirectly on transcription factors,

Activation of transcription factors by extracellular signals always involves a nuclear translocation step.

4.4.1. Transcription Factors: targets of signaling pathways

Genes contain response elements, short DNA sequences in their regulatory part, that bind transcription factors activated or inhibited by physiologic signals, most commonly through phosphorylation.

Two general mechanisms of transcriptional regulation

4.4.1.1. Rapid activation of pre-existing factors.

Transcription factors already present at significant basal levels are quickly activated (or repressed) by signaling cascades, allowing immediate regulation of target genes without waiting for new protein synthesis.

4.4.1.2. Signal-induced expression of new factors.

Transcription factors expressed at very low basal levels are themselves *induced* by a physiologic signal. Once produced, they go on to regulate a secondary wave of additional target genes.

The two general mechanisms of transcriptional regulation by extracellular signals are illustrated in this figure:

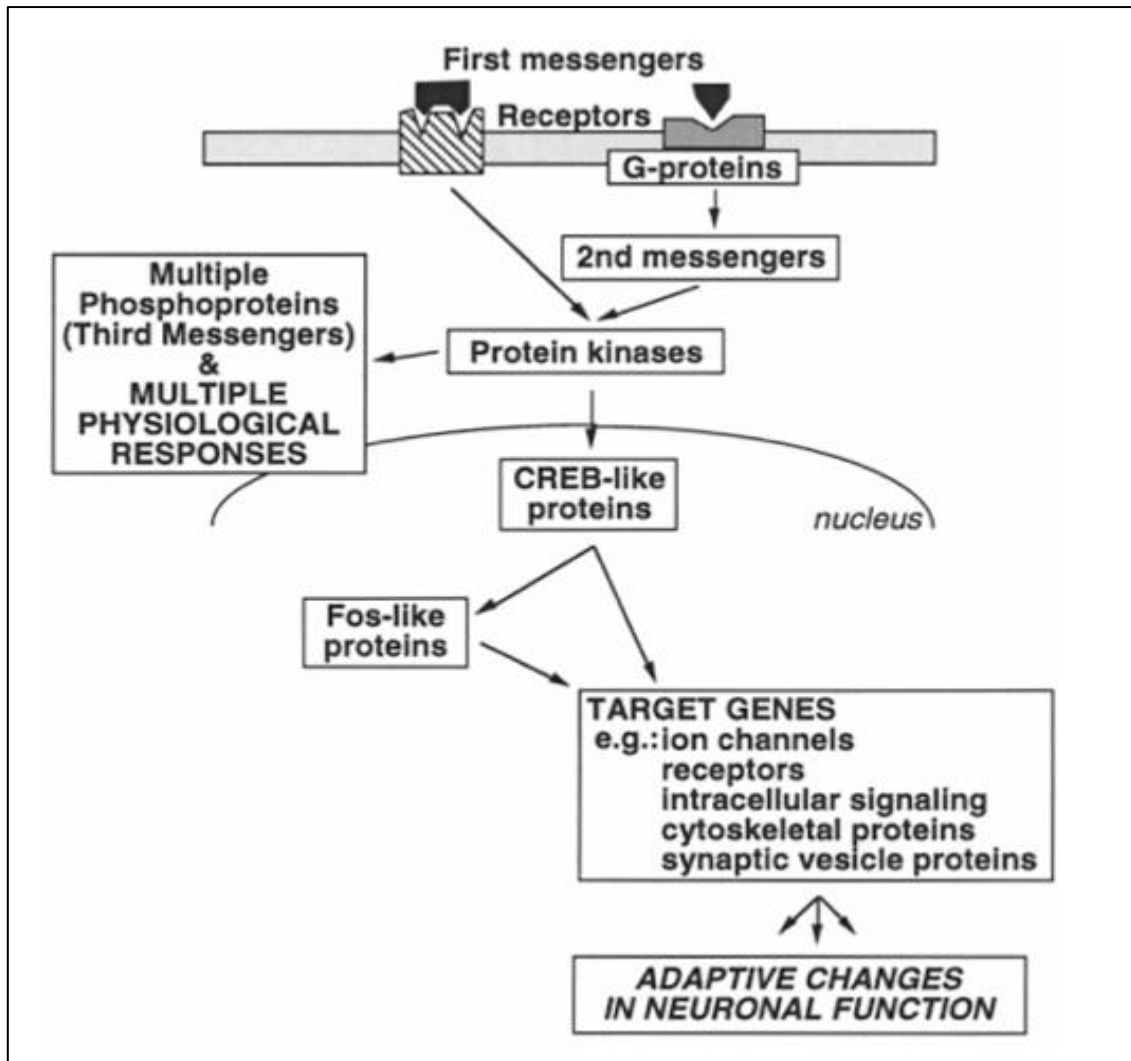


Figure 93. Intracellular pathways underlying regulation of gene expression.

Activation of neurotransmitter, hormone, or neurotrophic factor receptors leads to the activation of specific second messenger and protein phosphorylation pathways, which produce multiple effects on neuronal function through the phosphorylation of numerous proteins. Among the effects of these intracellular pathways on neuronal function is the regulation of gene expression. This can be accomplished by two basic types of mechanisms. In one case, transcription factors, already in the nucleus, are phosphorylated by protein kinases; this alters their transcriptional activity and leads to alterations in the expression of specific target genes. CREB is an example of a transcription factor that functions in this manner. Among the target genes for CREB (cAMP Response Element-Binding protein) family transcription factors are those for other transcription factors, for example, Fos and Jun family proteins. Increased expression of Fos and Jun then leads to alterations in the expression of additional target genes.

4.4.2. Signals-transduced from Membrane to Nucleus

Several distinct mechanisms are existing:

4.4.2.1. Transcription factor translocation itself

Involves migration of signaling molecules themselves into the nucleus. Examples:

- Steroid hormone receptors, ligand binding releases the factor from cytoplasmic chaperones, allowing direct nuclear entry.

- NF- κ B, normally retained in the cytoplasm by I κ B, which masks its nuclear localization signal. Signal-regulated phosphorylation of I κ B by PKC causes dissociation of NF- κ B, freeing it to enter the nucleus and bind DNA; I κ B is then proteolyzed in the cytoplasm.

4.4.2.2. Phosphorylation enabling multimerization and translocation

Activated transcription factors migrate to the nucleus following their activation in the cytoplasm. Example: STATs (Signal Transducer and Activator of Transcription): phosphorylation by cytoplasmic protein tyrosine kinases allows STAT monomers to multimerize, which both creates an effective DNA-binding site and enables nuclear translocation.

4.4.2.3. Nuclear translocation of activated kinases or second messengers

Activation of a single intracellular signaling pathway is sufficient for transcription factor activation. Example: CREB (cAMP Response Element-Binding protein) is already bound to its cAMP-response elements (CREs) in the nucleus under basal conditions. Activation does not require CREB itself to move. Instead, activated protein kinase A (PKA) translocates into the nucleus and phosphorylates CREB directly. Alternatively, Ca²⁺/calmodulin (a second messenger complex) enters the nucleus and activates kinases there, which then phosphorylate CREB. In both cases, phosphorylation converts CREB into a transcriptional activator by enabling it to recruit the coactivator CBP into the transcription complex.

4.5. Tutorial (Annex)

5. Methodology and molecular biology

Molecular biology methodology refers to the collection of techniques used to study and manipulate DNA, RNA, and proteins, as fundamental molecules of gene expression to understand cellular functions.

The essential techniques include:

- nucleic acid extraction,
- amplification technics (PCR and its derived technics),
- gel electrophoresis technics.
- sequencing technics
- cloning tools and methodology
- hybridisation methods

These methods are critical for gene expression studies, disease diagnostics, and biotechnological applications.

Note: these technics are the subject of the module of genetic engineering

5.1. Extraction and purification of nucleic acids

Extraction and purification are the first and essential steps in molecular biology studies.

5.1.1. Extraction of nucleic acids

The general protocol follows the same basic scheme: cell lysis, elimination of proteins and lipids, and selective separation of DNA or RNA.

5.1.1.1. Cell lysis

Can be:

- Mechanical, example: freeze-thaw cycles, osmotic shock, ...
- Chemical: enzymatic fragmentation of cell walls and membrane disruption (detergents like SDS or Triton X-100) example: lysozyme for bacteria, chitinase for fungi, cellulase for plants; membrane disruption by detergents like SDS or Triton X-100).

Chemical lysis is preferred for prokaryotes and mechanical lysis is reserved for eukaryotic cells.

5.1.1.2. Protein elimination

Is performed either by:

- Enzymatic hydrolysis using proteinase K (a non-specific endoprotease active up to 65°C, often used with SDS)
- Precipitation using chaotropic agents such as thiocyanate de guanidine.

5.1.1.3. RNA elimination during DNA extraction

Is done either by:

- Alkaline hydrolysis with NaOH (RNA is degraded at high pH whereas DNA is protected by the absence of a 2'-OH group on deoxyribose)
- Enzymatic digestion with RNase.

5.1.2. Purification of nucleic acids

Purification methods include phenol-chloroform extraction (phenol deproteinizes powerfully; chloroform removes residual phenol traces), isopycnic centrifugation on cesium chloride gradient (separates DNA, RNA, and proteins by density), and chromatography on silica or anion-exchange columns.

Note: now, commercial kits are widely used in extraction and purification of nucleic acids

5.1.3. Concentration

Is achieved by alcohol precipitation:

- Ethanol: requires prior addition of NaCl to neutralize phosphate groups
- Isopropanol: no salt needed, but it eliminates small DNA fragments

5.1.4. Quantification and qualification

Quantification is done by UV absorptiometry at 260 nm:

- 1 OD = 50 µg/mL dsDNA
- 1 OD = 37 µg/mL ssDNA
- 1 OD = 40 µg/mL RNA)

The A260/A280 ratio assesses purity:

- 1.8–2.0 = pure DNA
- <1.7 = protein contamination
- >2.0 = RNA contamination

5.1.5. Conservation:

DNA is stored at 4°C in Tris-EDTA buffer (pH 8) or at -20°C in aliquots.

RNA, which is extremely labile, is stored at -80°C with ethanol.

5.2. Polymerase chain reaction (PCR) strategy

5.2.1. Definition

PCR is an *in vitro* enzymatic amplification technique developed in 1985 by Kary Mullis (Nobel Prize 1993) for amplifying specific DNA sequences, using *Thermus aquaticus* (Taq) polymerase, primers, and heat cycling to produce billions of copies of a target sequence. After *n* cycles, 2^{*n*} copies are produced. The reaction is run in a thermocycler for 25–35 cycles.

Each PCR cycle consists of three steps:

- Denaturation (~94–95°C, 1–2 min): separate double-strand DNA
- Hybridization or annealing (40–65°C, 1 min): primers bind to target

- Elongation (72°C, 1–2 min): Taq polymerase extends
- **Primer** must be: length 18–30 nucleotides; no inter- or intra-primer complementarity; GC content 40–60%; melting temperatures (T_m) of both primers should be equal and between 52–60°C.
- **T_m calculation (Wallace rule):** $T_m = 4(G+C) + 2(A+T)$ for short sequences; for longer sequences: $T_m = 81.5 + 16.6 \cdot \log[M^+] + 0.41 \cdot (\%GC) - 675/L$
- **Hybridization temperature:** $T_h = 0.3 \cdot T_m(\text{primer}) + 0.7 \cdot T_m(\text{amplicon}) - 14.9$

5.2.2. PCR variants:

The most used techniques are:

5.2.2.1. RT-PCR

Converts mRNA to cDNA via reverse transcriptase before standard PCR, used for analyzing gene expression especially from rare mRNAs.

5.2.2.2. Multiplex PCR

Amplifies several different DNA segments simultaneously using multiple primer pairs.

5.2.2.3. Real-time quantitative PCR (qPCR)

Measures fluorescence as amplification proceeds, either using TaqMan hydrolysis probes (specific, uses quencher-reporter mechanism) or SYBR Green I (non-specific intercalating agent that fluoresces when bound to dsDNA).

5.2.3. Applications of PCR

Molecular cloning of PCR products, diagnosis of infectious agents (hepatitis B/C/D, HIV), mutation analysis, direct sequencing of amplified fragments, forensic analysis from minimal biological material ...

5.3. Electrophoresis of Nucleic Acids

Gel electrophoresis separates DNA, RNA, or proteins based on size, electric charge, and physical properties under an electric field, using agarose or polyacrylamide (PAGE) gels.

SDS-PAGE is specifically used to separate proteins by mass.

Nucleic acids are uniformly poly-anionic at pH 8 and migrate toward the positive pole (anode).

5.3.1. Migration speed

depends on: molecular size (smaller = faster), gel concentration, applied voltage (limited to ~5V/cm to prevent melting), and DNA conformation (supercoiled CCC migrates faster than relaxed OC or linear form).

5.3.2. Gel supports

Two type of gel are used:

- Agarose gel (most used): a natural linear polysaccharide from algae; it has large pores for separating large molecules (0.5–30 kb depending on concentration 0.5–1.5%).
- Polyacrylamide gel: used for small fragments with single-base resolution (4–11% for 10–800 bp).

5.3.3. Visualization

Is done with ethidium bromide (BrEt), an intercalating agent that emits orange fluorescence under UV light, or by incorporation of radioactive isotopes detected by autoradiography.

5.3.4. Size determination

Uses commercial molecular weight markers (standard ladders), commonly fragments of viral genomes or synthetic oligonucleotides of known size.

5.3.5. Special variants

- RNA gels require denaturing agents (formamide) to disrupt secondary structures that would impair size estimation.
- Pulsed-field gel electrophoresis (PFGE) separates very large DNA fragments (>20 kb up to several megabases) by alternating the electric field orientation, causing size-dependent reorientation delays ...

5.4. Sequencing

5.4.1. Definition

DNA sequencing determines the number, nature, and order of nucleotides in a gene.

5.4.2. Principle:

The enzymatic method by Sanger-Coulson (1977, first generation) uses dideoxynucleotides (ddNTPs): modified nucleotides that incorporate into a growing chain but block further elongation by lacking the 3'-OH group.

A primer is hybridized to single-stranded template DNA. DNA polymerase synthesizes the complementary strand in the presence of normal dNTPs plus one ddNTP per reaction (4 separate tubes: ddGTP, ddATP, ddTTP, ddCTP). Each ddNTP incorporation terminates the chain at a specific position, generating a family of fragments of different sizes all terminating with the same base. The four tube contents are separated by denaturing polyacrylamide gel electrophoresis and read by autoradiography (radioactive labeling) or laser fluorescence detection (fluorescent labeling).

5.4.3. Automated sequencing

Uses four ddNTPs each labeled with a **different fluorochrome** (green for ddA, blue for ddC, yellow for ddG, red for ddT), allowing a single reaction and single capillary

electrophoresis run. Laser detection identifies each fluorescent ddNTP as it exits the capillary, generating a chromatogram. Modern automated sequencers can read 300–800 bases per run and analyze up to 96 samples in a few hours

5.4.4. Sequencing technology

Has evolved through three generations/ from Sanger's elegant chain-termination chemistry producing one accurate sequence at a time, through massively parallel NGS (next generation sequencing) platforms that democratized genomics, to third-generation single-molecule long-read technologies that are resolving the final complexities of genome structure and enabling real-time, portable diagnostics. Each generation builds on the last, and in modern genomics laboratories, all three are used in complementary ways depending on the question being asked.

5.5. Molecular Cloning

5.5.1. Definition

Molecular cloning inserts a foreign DNA fragment into a vector (plasmid, phage, ...) to replicate it in a host cell. It is an *in vivo* amplification

5.5.2. Steps of cloning

- The cloning vector is cut at a unique restriction site (in the polylinker) by a restriction endonuclease.
- The donor DNA is digested with the same or a compatible enzyme, generating complementary cohesive ends.
- The linearized vector and donor fragments are mixed, complementary ends hybridize via hydrogen bonds.
- DNA ligase creates covalent phosphodiester bonds between 5'-P and 3'-OH ends to form the recombinant plasmid.

5.5.3. Key technical points

- To prevent vector self-recircularization (ligation without insert), the vector is treated with alkaline phosphatase to remove 5'-phosphate groups.
- When insert and vector are cut by two different enzymes, directional cloning is achieved since only one orientation of insertion is possible.
- When blunt ends are used, ligation is less efficient but allows cloning of any fragment including PCR products. If incompatible ends exist, DNA polymerase can fill in 5'-protruding ends (5'→3' polymerase activity) or polish 3'-protruding ends (3'→5' exonuclease activity) to create blunt ends.

5.5.4. Introduction of recombinant DNA into host cells

Is achieved by different methods, the most used are:

- Transformation: bacteria made competent with CaCl_2 , then heat-shocked or electroporated
- Conjugation: cell-to-cell DNA transfer
- Transduction: phage-mediated transfer
- Electroporation: electric pulse creates temporary membrane pores
- Microinjection
- Biolistic: particle gun, tungsten or gold, beads coated with DNA penetrate cell walls without lysis.

5.5.5. Ideal cloning host

The cell host must:

- grow rapidly in inexpensive media
- be non-pathogenic
- incorporate foreign DNA
- remain stable in culture
- possess appropriate enzymes for vector replication.

Escherichia coli is the most commonly used host. Other hosts include *Bacillus subtilis* and *Saccharomyces cerevisiae*.

5.5.6. Selection of transformed colonies

Uses antibiotic resistance genes carried on the vector: only bacteria that have integrated a plasmid survive on antibiotic-containing medium.

Colonies of interest are further verified by in situ hybridization, DNA sequencing, or size control by gel electrophoresis.

5.5.7. Gene expression

Requires the cloned gene to be under the control of a promoter and have terminators; it must be in the correct reading frame for efficient translation.

5.6. Molecular Hybridization technics

5.6.1. Definition

Molecular hybridization is the property of a single-stranded nucleic acid to spontaneously, specifically, and reversibly associate with another complementary single-stranded molecule through hydrogen bonds.

5.6.2. Stringency

Refers to the degree of prevention of non-specific hybridizations.

High stringency (low NaCl + high temperature >65°C) allows only perfectly homologous sequences to hybridize. Increasing stringency destabilizes the double helix and increases hybridization specificity.

5.6.3. Probes

Are labeled single-stranded nucleic acid molecules of known sequence used to detect complementary target sequences. Probes can be DNA, RNA, or antibodies (for protein detection). They carry either "hot" signals (radioactive: ^{32}P or ^3H) or "cold" signals (fluorescent, colorimetric via digoxigenin/alkaline phosphatase system).

5.6.4. Types of hybridization

5.6.4.1. *In situ* hybridization on tissue sections

Detects the spatial localization of specific mRNA within a tissue: tissue is chemically fixed, embedded in paraffin, sectioned, and incubated with a riboprobes.

5.6.4.2. *In situ* hybridization on chromosomes (FISH)

Determines on which chromosome and in which region a gene of interest is located.

5.6.4.3. *In situ* hybridization on bacterial colonies

Screens recombinant colonies by lysing bacteria directly on a filter and probing with a labeled DNA probe.

5.6.4.4. Southern blotting (DNA/DNA hybridization)

Developed by Edwin Southern in 1975. DNA fragments are separated by agarose gel electrophoresis, denatured, then transferred by capillarity onto a nitrocellulose or nylon membrane, hybridized with a labeled probe, washed, and detected by autoradiography.

Applications include detection of specific genes, diagnosis of genetic diseases, and restriction fragment length polymorphism (RFLP) analysis.

5.6.4.5. Northern blotting (RNA/DNA or RNA/RNA hybridization)

RNA is isolated, separated by denaturing agarose gel electrophoresis, transferred to a nylon membrane, and hybridized with a labeled single-stranded DNA or RNA probe.

Reveals the size and abundance of specific mRNA transcripts in a tissue under given conditions or following stress.

5.6.4.6. Dot/Slot blotting

Targets are applied directly to a membrane at known concentrations without prior electrophoresis. Faster than Southern/Northern but cannot detect non-specific hybridization since fragments are not size-separated.

5.6.4.7. Western blotting (protein/antibody hybridization)

Also called immunoblotting. Proteins are separated by SDS-PAGE, transferred to a nitrocellulose membrane, incubated with a primary antibody specific to the target protein, then with an enzyme-conjugated or fluorescently labeled secondary antibody for detection.

Reveals protein size, expression level, and post-translational modification state.

5.6.4.8. Enzyme-Linked Immunosorbent Assay (ELISA)

IT is a quantitative immunological assay for detecting and measuring peptides, proteins, or antigens in biological samples at picogram/mL concentrations. Performed on multi-well microtiter plates. Success depends on the affinity and specificity of antibody/antigen interactions, which are influenced by pH, temperature, and ionic strength.

- Direct ELISA — antigen is bound directly to the microplate well, then detected by an enzyme-conjugated primary antibody. Simple and fast but less sensitive.
- Indirect ELISA — antigen is bound to the plate; a non-conjugated primary antibody binds the antigen; a conjugated secondary antibody binds the primary antibody and generates the signal. More sensitive and flexible — one secondary antibody can detect multiple primary antibodies.
- Sandwich ELISA — a capture antibody is first bound to the plate and captures the antigen; a primary antibody then binds the captured antigen; a conjugated secondary antibody detects the primary antibody. Highly sensitive and specific; ideal for complex biological samples.

In all formats, an enzyme substrate is added at the end — the enzymatic reaction produces a colorimetric, fluorescent, or chemiluminescent signal proportional to the quantity of target molecule present.

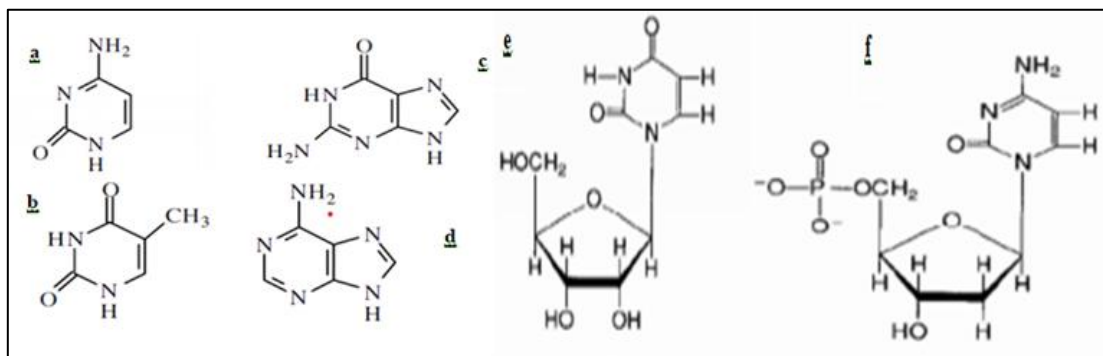
5.7. Tutorial (Annex)

Annex: Tutorials

Tutorial 1

Exercise 1:

Give the names of the structures proposed in the diagram below:



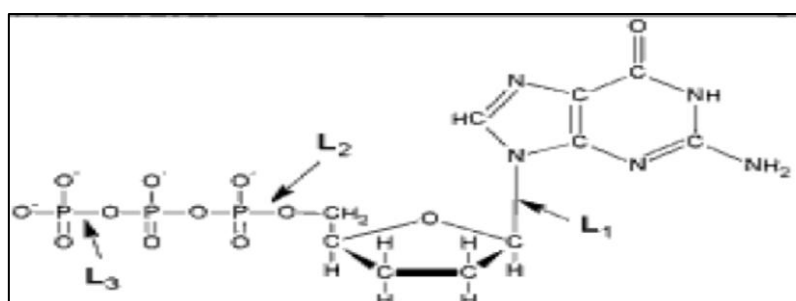
Exercise 2 :

The two strands of the DNA double helix can be separated by heating. If you raise the temperature of a solution containing the following three DNA molecules, in what order do you suppose they would melt? Explain your answer.'

- A. 5' GCGGGCCAGCCCCGAGTGGGTAGCCCCAGG 3'
3' CGCCCGGTCTGGGCTCACCCATCGGGTCC 5'
- B. 5' ATTATAAAATATTTAGATACTATATTTACA 3'
3' TAATATTTTATAAATCTATGATATAAATGT 5'
- C. 5' AGAGCTAGATCGAT 3'
3' TCTCGATCTAGCTA 5'

Exercise 3 :

Consider the following nucleotide:



- Is the nitrogenous base constituting this molecule of purine or pyrimidine nature? Name this base and the bonds L1, L2, and L3?
- Illustrate using a diagram the hydrogen bonds with the complementary base?

- Is it possible to synthesize an artificial polynucleotide from this molecule using a polymerase? Why?

Exercise 4 :

- What is a nucleotide ?
- What do the values and symbols 5' and 3' correspond to on a nucleic acid strand ?
- What is the charge of a nucleotide at physiological pH ?
- What are the three essential characteristics of double-stranded DNA ?
- According to Chargaff's study on base composition, which of the following propositions could characterize any non-viral DNA sample ? (1) $A+T = G+C$; (2) $A/T = 1$; (3) $G=C$; (4) $A+G = T+C$.
- If the (C) content of a double-stranded DNA preparation is 15%, what is the (A) content ?
- For what reason do the two strands of DNA molecule remain associated? What part of the molecule is involved ? What happens when the temperature is increased ? What is the phenomenon called?

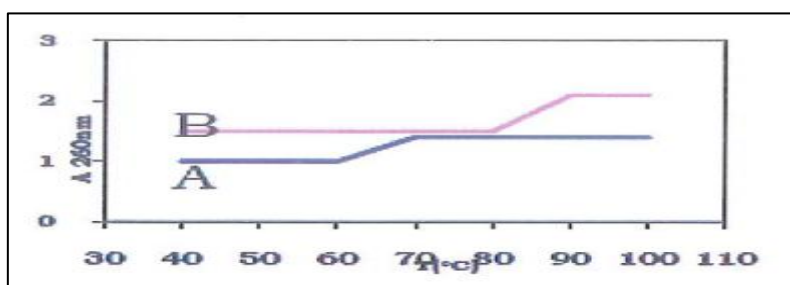
Exercise 5 :

Two double-stranded DNAs A and B of different origins are composed of 647 bp. DNA A is resistant to the action of exonucleases while DNA B is hydrolysed by these enzymes.

- What is the molecular weight expressed in Daltons of these DNAs?
- Give the length of these DNAs in μm ?
- Why is DNA A insensitive to the action of exonucleases?

Solutions of DNA A and B are heated and their absorbance at 260 nm is measured as a function of temperature increase. The results are presented in the figure below:

- For what reason is absorbance measured at 260 nm?
- What do the differences recorded for DNAs A and B at 40°C mean?
- What does the increase in absorbance reflect as the temperature rises?
- How do the two DNAs differ?



Tutorial 2

Exercise 1:

A laboratory is studying mutation rates in *E. coli* under different conditions:

Condition A (Control): Bacteria grown in standard medium. Mutation rate: 1×10^{-8} mutations per base pair per generation

Condition B: Bacteria exposed to UV radiation. Mutation rate: 5×10^{-6} mutations per base pair per generation

Condition C: Bacteria exposed to chemical mutagen EMS (ethyl methanesulfonate). Mutation rate: 3×10^{-6} mutations per base pair per generation

1. Which condition represents spontaneous mutations? Why?
2. Calculate how many times higher the mutation rate is in conditions B and C compared to A.
3. What are the main sources of spontaneous mutations?
4. Give 2 examples of induced mutations and their mechanisms.

Exercise 2:

You perform an Ames test using strain TA100 (detects base-pair substitutions) to test three chemical substances: X, Y, and Z. Results (number of revertant colonies) are indicated in the table below:

Condition	Without S9	With S9
Negative control (no substance)	25	28
Positive control (known mutagen)	850	920
Substance X	30	680
Substance Y	720	750
Substance Z	22	26

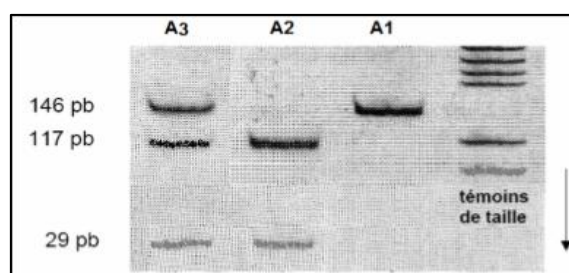
- What is the purpose of the negative control?
- What is the purpose of the positive control?
- What does the S9 fraction represent and why is it important?
- Analyze each substance (X, Y, Z):
 - Is it mutagenic?
 - Does it require metabolic activation?
 - Classify it as: direct mutagen, promutagen, or non-mutagen
- Which substance poses the greatest mutagenic risk in the human body? Explain.

Exercise 3:

Hemochromatosis is a disease characterized by an excess of iron in the body. In its genetic form, the disease is autosomal recessive, caused by two types of mutations in the HFE gene, the most common being the Cys282Tyr mutation (conversion of GTGC to GTAC).

To determine whether two children, A2 and A3, from a family known to have this form of the disease are affected or not, you ask a molecular biology laboratory to perform the diagnosis.

1. What will the laboratory do exactly? Knowing that this laboratory has, in addition to several devices necessary for the different techniques:
 - Two primers that can bind on either side of this gene.
 - A restriction enzyme RsaI that cuts at: GT/AC.
2. The result is shown in the following gel, knowing that A1 represents a healthy control. Comment on the result.



Exercise 4:

A patient with **chronic myeloid leukemia (CML)** undergoes FISH analysis. Two probes are used:

- **Red probe** → binds ABL gene (chromosome 9)
- **Green probe** → binds BCR gene (chromosome 22)

1. What signal pattern would you expect in:

Cell Type	Signal Pattern	Interpretation
Normal cell		No translocation
CML cell		BCR-ABL translocation present

2. Why is detecting the BCR-ABL fusion clinically important?

Tutorial 3

Exercise 1:

A DNA analysis is carried out on several cells (A, T, C, G expressed in %, approximate values):

Origine de l'ADN		A	C	T	G
Sea urchin cell		33,2	16,7	33,3	16,8
Wheat cell		27,3	22,7	27,2	22,8
Cells from the same human individual	Skin cell	30,5	19,9	29,8	19,8
	Intestinal cell	30,5	19,9	29,8	19,8

1. What do the letters A, T, C, G represent?
2. For all cells, compare the percentages of the "letters" — what do you notice?
3. Using your knowledge, explain the results.
4. For the different cells, compare the percentage of each "letter" — what do you notice?
5. Using your knowledge, explain these results.
6. Compare the percentage of the different letters in human cells — what do you notice?
7. Using your knowledge, explain these results.
8. Would the same results be obtained for cells from another human?

Exercise 2:

Given data:

The average size of a human chromosome is approximately **150×10^6 bp** (bp: base pairs, base $\approx$ nucleotide);

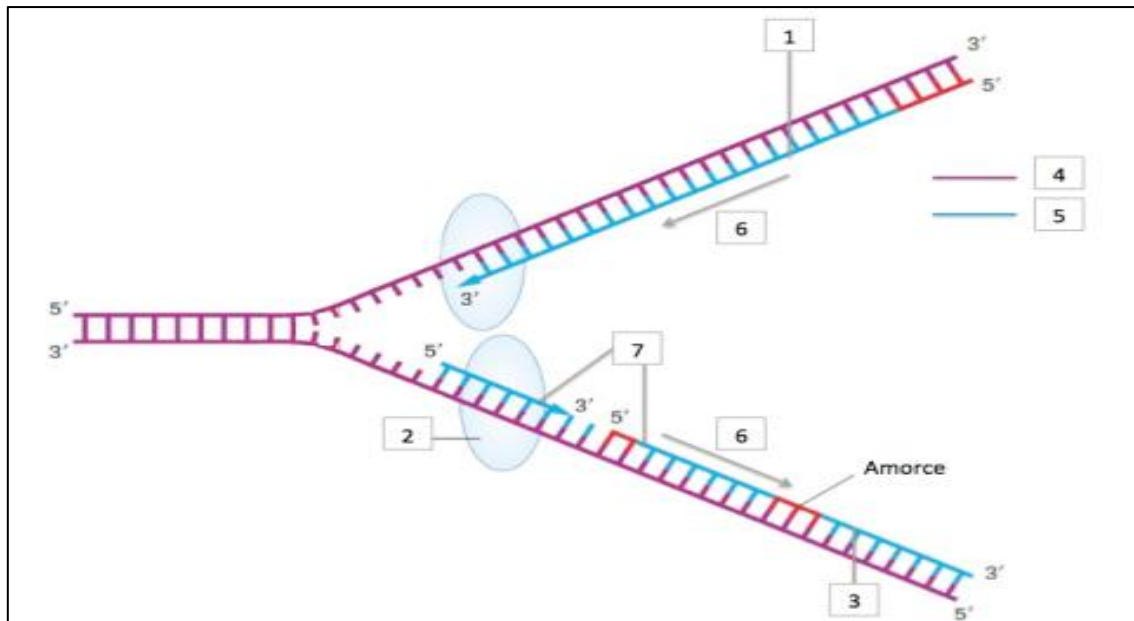
Replication speed: approximately **100 bp/s** at each replication fork

Actual replication duration: approximately **8 hours**

1. Using the information from Document 1, calculate the theoretical duration of replication of a human chromosome if it had only a single replication origin.
2. Based on your result, what can you conclude in relation to the information in Document 1?

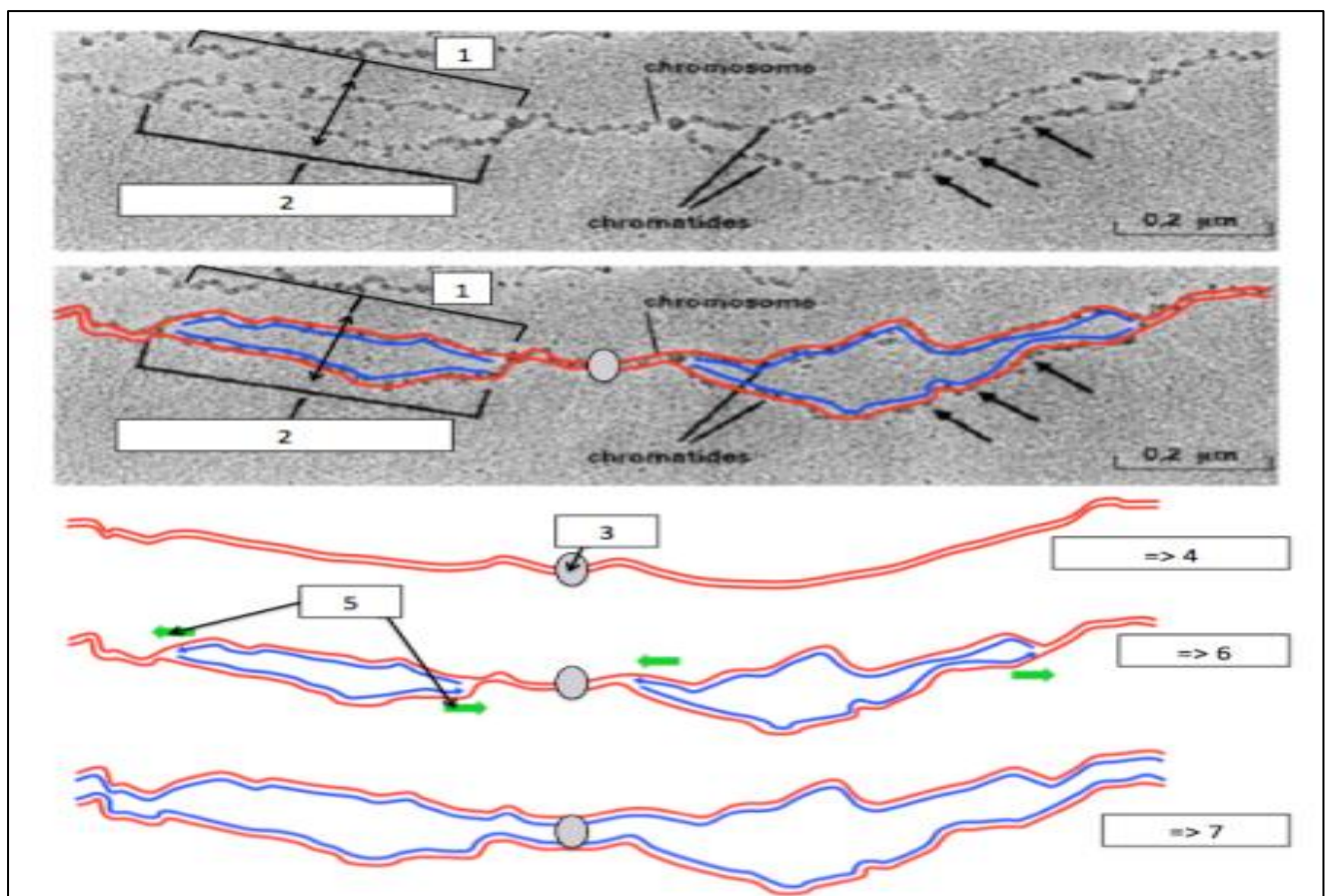
Exercise 3:

Label the following diagram and explain what it illustrates.



Exercise 4:

Label the following figures.



Tutorial 4

Exercise 1:

Given the following DNA sequence:

5'- ATTTACGGGCCTTAATGGCATAACCGCCTAATGGTTAACCGCTAGCGCG -
3'

1. Give the corresponding double-stranded DNA sequence.
2. Under what condition would this double-stranded DNA be transcribed *in vivo*?
3. Give the sequence of the possible transcript and polypeptide.

Exercise 2:

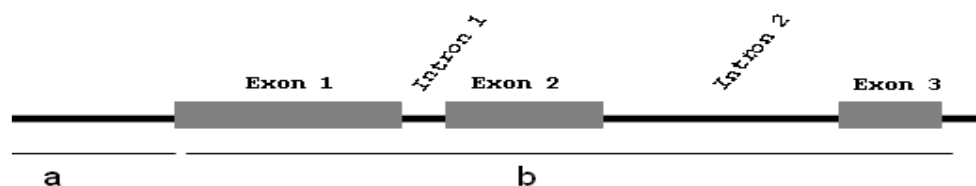
Given the following sequence:

...T
TGACAGTGTG...GAGGGCTGGGGTCTATAATCC...CGAACT...TCATGACGAGGCCGCG
ATT

1. Define the underlined motifs in the sequence.
2. Give the sequence of the complementary fragment.
3. Give the direction of the possible transcription. Define the sense strand and the antisense strand.

Exercise 3:

A schematic representation of a gene encoding protein X is provided:



(Exon 1 = 426 bp, Exon 2 = 300 bp, Exon 3 = 110 bp, Intron 1 = 58 bp, Intron 2 = 670 bp)

- Define parts **a** and **b** and indicate the transcription initiation site and the stop codon.
- If the promoter size is 198 bp, what is the size of the gene (excluding regulatory sequences)?
- Without counting the cap and the poly(A) tail, what will be the size of:
 - The primary transcript
 - The mRNA after constitutive splicing
 - The mRNA after alternative splicing retaining only exons 1 and 3
- Schematically represent the RNAs described in question 3.

- What is the amino acid sequence corresponding to the following end sequences:
5'.....ATGGCACCGGACTGG.....TGGCCGACCAACTGA.....3'

Exercise 4:

A human protein is composed of **576 amino acids**.

- What is the **minimum** number of:
 - Nucleotides contained in the primary transcript encoding this protein?
 - Nucleotide pairs contained in the structural gene encoding this protein?
- Why could the number of nucleotides be higher than the values you just calculated in both cases?

Tutorial 5

References :

AMEUR AMEUR Abdelkader (2017). Génétique Générale « Utile en Biologie : Génétique Biochimie Immunologie Ecologie Microbiologie Zoologie Botanique ».Editions Al-Djazair. 108 pages.

Anita Göndör Assistant Professor, Department of Microbiology, Tumor, and Cell Biology (MTC), Karolinska Institutet, Solna, Sweden. Chromatin Regulation and Dynamics.

Breno Lisboa (2025). Concepts of genetics. TAB. 276 pages.

Caroline S. Hill and Richard Treisman. Transcriptional Regulation by Extracellular Signals: Mechanisms and Specificity. Cell, Vol. 80, 199-211, (January 27, 1995).

Cindy J. Smith & A. Mark Osborn. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. FEMS Microbiol Ecol 67 (2009) 6–20

Denis Tagu, Christian Moussard. 2003. Principes des techniques de biologie moléculaire. Editions Quae, 176 pages

Dorothy K. Grange. CLINICAL MATERNAL–FETAL MEDICINE ONLINE: chapter 43 Basic genetics and patterns of inheritance 21 pages

Émile Biémont.2008. Spectroscopie moléculaire: Structures moléculaires et analyse spectrale De Boeck Supérieur 428 pages

Eric J. Nestler Steven E. Hyman. Chapter 17: Regulation of Gene Expression. Neuropsychopharmacology: The Fifth Generation of Progress. 2017- 229.

Ferreira, T. & Rodriguez, S. (2024). Mitochondrial DNA: Inherent Complexities Relevant to Genetic Analyses. Genes 15(5):617.

Gérard Coutouly, Emile Klein, Eric Barbieri, MostafaKriat. 2006. Travaux dirigés de biochimie, biologie moléculaire et bioinformatique. Wolters Kluwer France. 346 pages

Jared Nordman and Terry L. Orr-Weaver. Regulation of DNA replication during development. Development 139, 455-464 (2012) doi:10.1242/dev.061838

Karl-Uwe Reusswig and Boris Pfander. Control of Eukaryotic DNA Replication Initiation—Mechanisms to Ensure Smooth Transitions; Genes (2019), 10, 99; 1-19.

Lizabeth Allison.(2007). Fundamental molecular biology. Library of Congress Cataloging-in-Publication Data. 748 pages.

Luchetta, Philippe. 2013. Biologie moléculaire en 30 fiches. Dunod, Paris 2^{ème} édition. 155pages.

Mauricio Díaz-Muñoz, Rolando Hernández-Muñoz, Armando Butanda-Ochoa. Structure–activity features of purines and their receptors: implications in cell physiopathology. Molecular Biomedicine (2022) 3:5. 17 pages

Natassia Correa (2025). An introduction to molecular biology. TAB. 380 pages.

Nejat Mahdiah, Bahareh Rabbani. An Overview of Mutation Detection Methods in Genetic Disorders. Iran J Pediatr Aug 2013; Vol 23 (No 4), Pp: 375-388

Peter C. Turner. 2000. L'essentiel de biologie moléculaire. Dunod, Paris. 345 pages

Robert Schleif. (1993) Genetics and molecular biology. 2nd ed. Library of Congress Cataloging . 715 pages.

Sakamoto, W. (2018). Chloroplast DNA Dynamics: Copy Number, Quality Control and Degradation. Plant Cell Physiol 59(6):1120-1127.

Willow Adams (2025). Fundamentals of cell and molecular biology. Intelliz press. 320 pages.