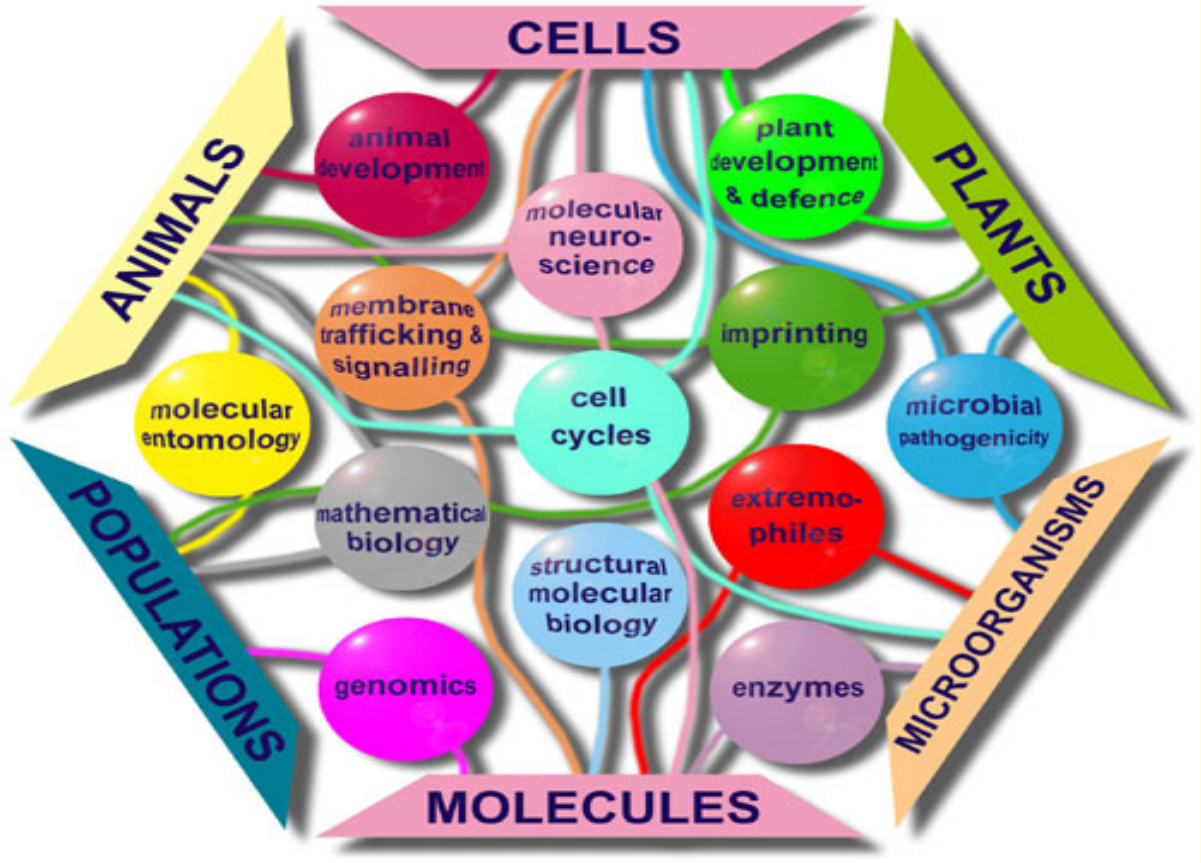
# Basics of Molecular biology

# Basic differences between eukaryotes and prokaryotes

| Attribute | Eukaryotes | Prokaryotes |
|---|---|---|
| **Organisms** | Plants, animals and fungi | bacteria and cyanobacteria |
| **Cell wall** | No (animals); Yes (plants) | yes |
| Chromosome segregation | Mitotic spindle | Cell membrane |
| meiosis | + | _ |
| Ribosome size | 80 s | 70 s |
| **Cell organelle** | | |
| Nuclear membrane | + | Absent |
| Endoplasmic reticulum | + | - |
| Golgi apparatus | + | - |
| Mitochondria | + | - |
| Chloroplast | + | - |

# Molecular biology: definition

- Molecular biology is the study of molecular underpinnings of the process of replication, transcription and translation of the genetic material.



- This field overlaps with other areas of biology and chemistry, particularly genetics and biochemistry. Molecular biology chiefly concerns itself with understanding the interactions between the various systems of a cell, including the interactions between DNA, RNA and protein biosynthesis as well as learning how these interactions are regulated.

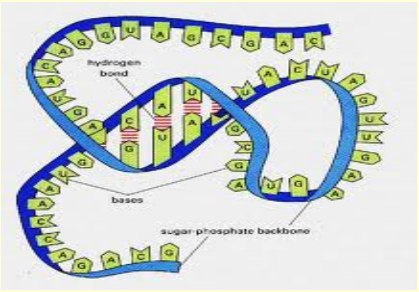
- Much of the work in molecular biology is quantitative, and recently much work has been done at the interface of molecular biology and computer science in bioinformatics and computational biology.

- Since the late 1950s and early 1960s, molecular biologists have learned to characterize, isolate, and manipulate the molecular components of cells and organisms includes DNA, the repository of genetic information; RNA, a close relative of DNA; and proteins, the major structural and enzymatic type of molecule in cells.

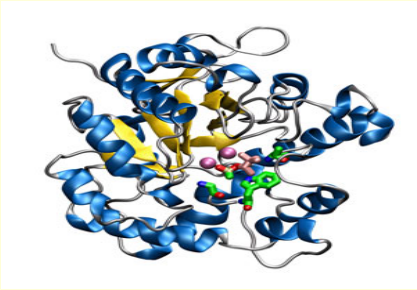
# Components involve in molecular biology

DNA

RNA

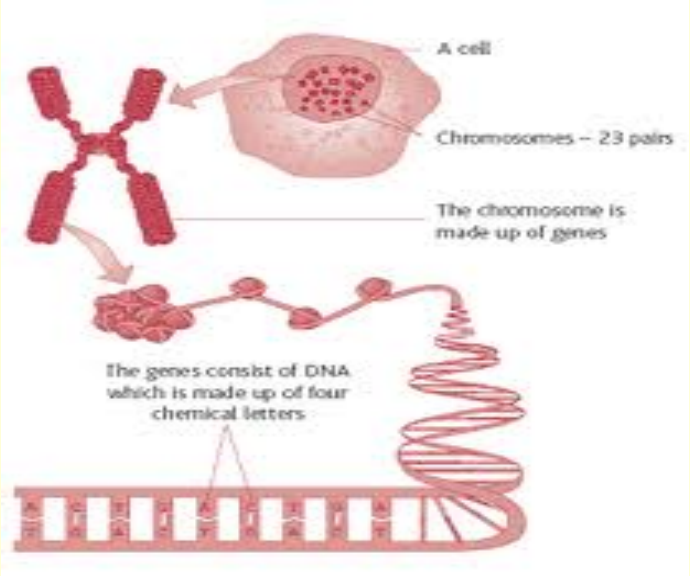


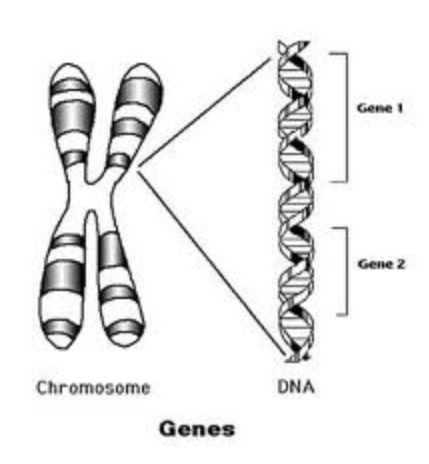

Protein



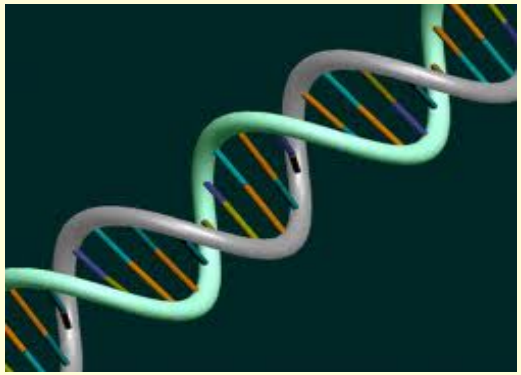
# Gene : Unit of heredity

- The DNA segments that carries genetic information are called genes.
- It is normally a stretch of DNA that codes for a type of protein or for an RNA chain that has a function in the organism.
- Genes hold the information to build and maintain an organism's cells and pass genetic traits to offspring.



# Deoxyribonucleic acid (DNA)



- DNA is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms and some viruses.
- DNA is a set of blueprints needed to construct other components of cells, such as proteins and RNA molecules.

- Two long strands makes the shape of a double helix.
- two strands run in opposite directions to each other and are therefore anti-parallel.
- Chemically, DNA consists of two long polymers of simple units called nucleotides, with backbones made of base, sugars and phosphate groups.

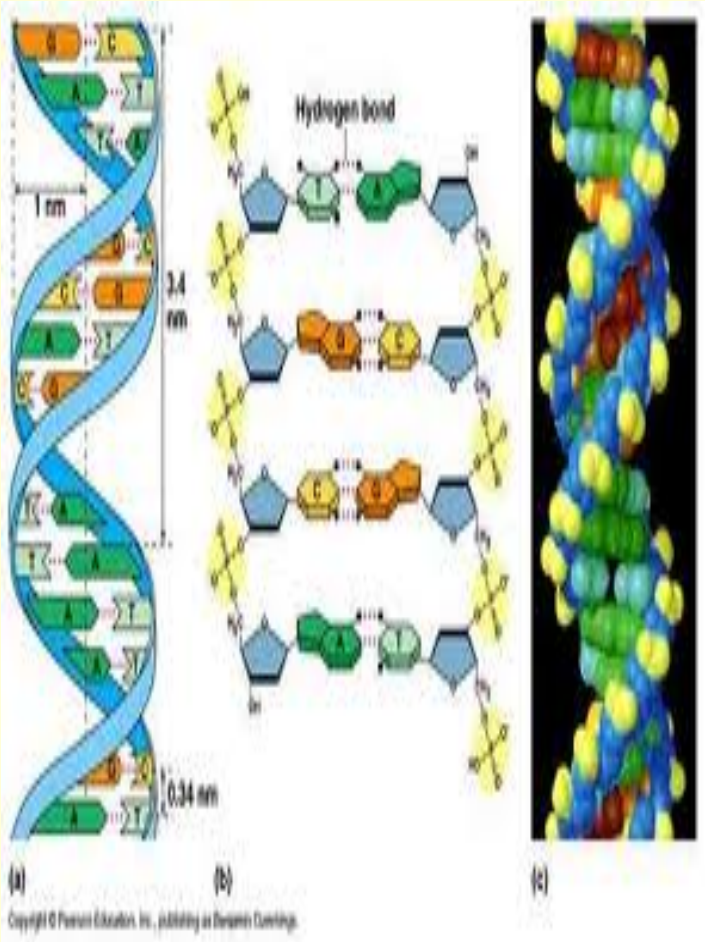
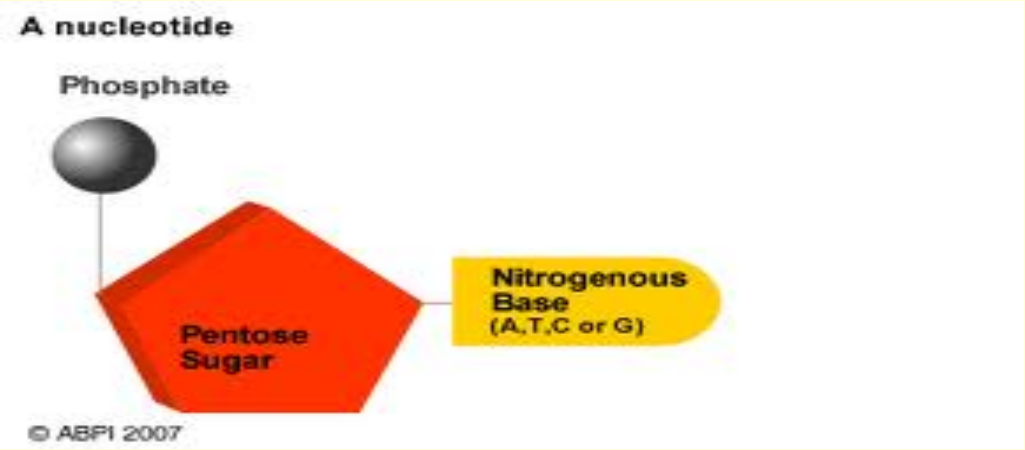



Fig : DNA double helix

# Sugar +Base = nucleoside

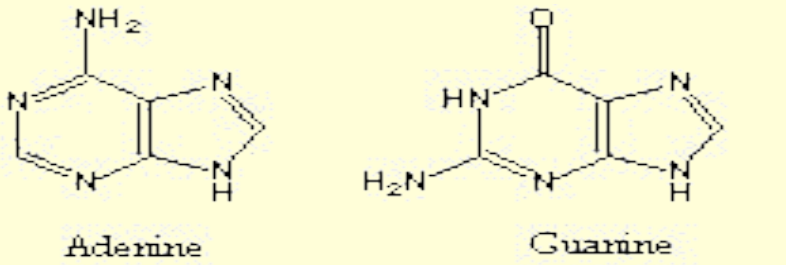

nucleoside

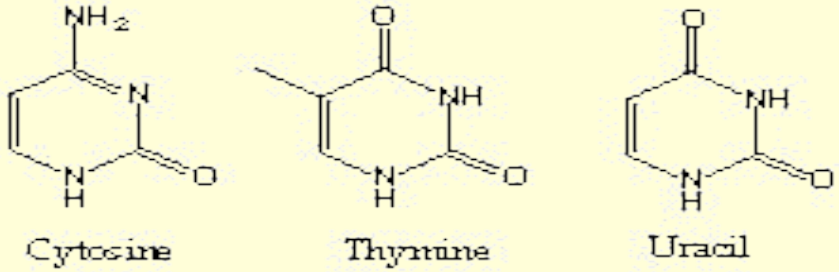
# Phosphate+ sugar + Base  = nucleotide



# Bases

- Types:- **adenine** and **guanine** (fused five- and six-membered heterocyclic compounds) – **Purines**

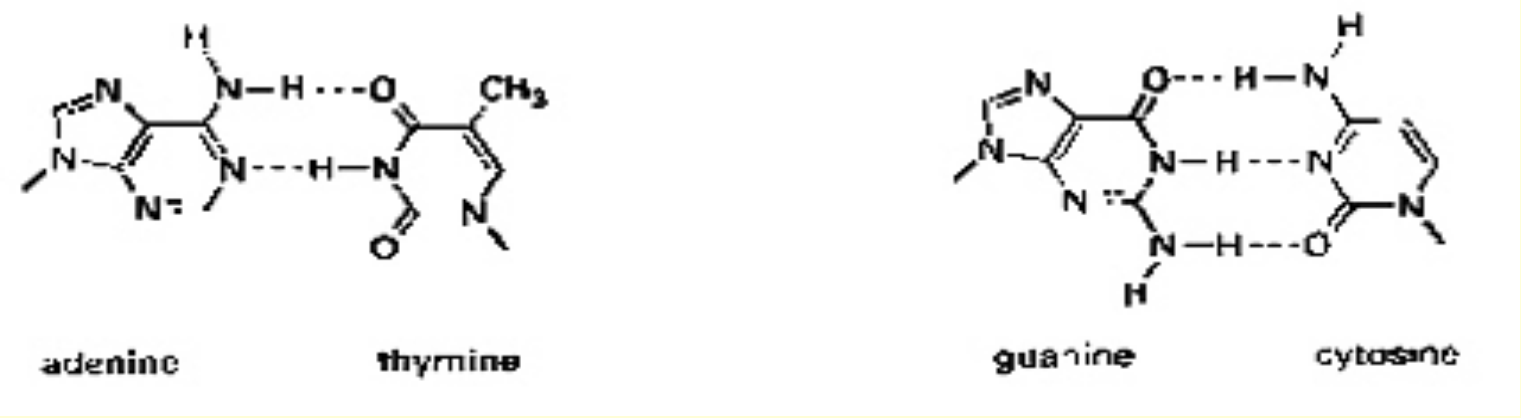


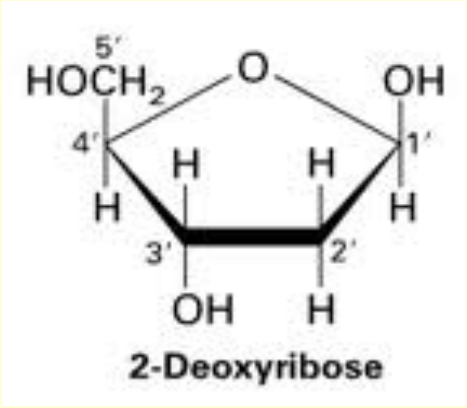

- **cytosine** & **thymine** (six-membered rings)-**Pyrimidines.**
- A fifth pyrimidine base, called uracil (U), usually takes the place of thymine in RNA and differs from thymine by lacking a methyl group on its ring.
- PAIRING : A =T and A=U

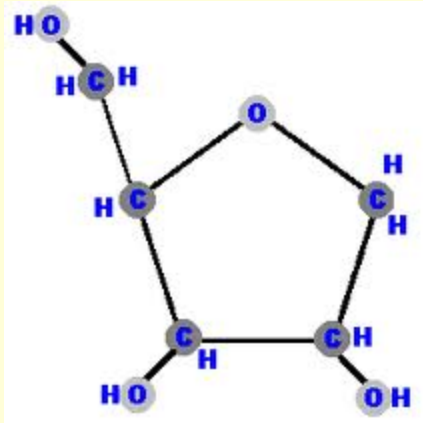
  G≡C

- The DNA double helix is stabilized by hydrogen bonds between the bases attached to the two strands.



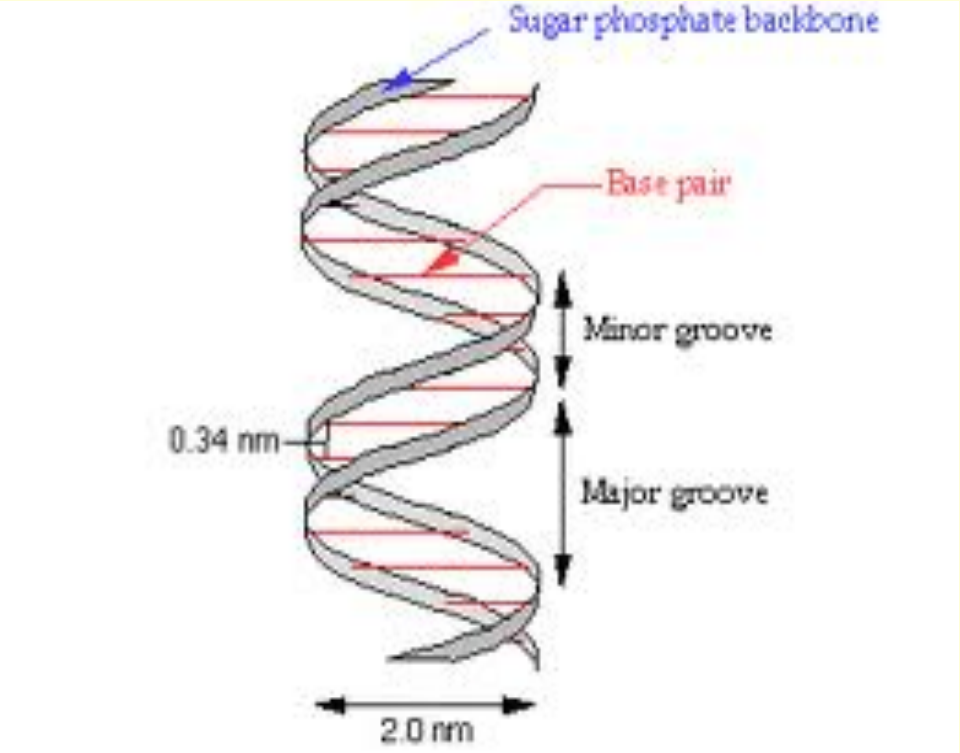

- One major difference between DNA and RNA is the sugar, with the 2-deoxyribose in DNA being replaced by the alternative pentose sugar ribose in RNA.



Ribose

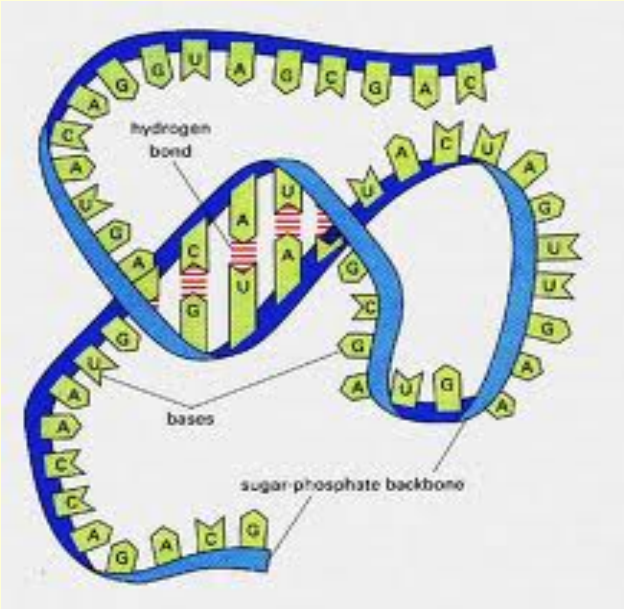
# Size:




- The DNA chain is 22 to 26 Ångströms wide (2.2 to 2.6 nanometres), and one nucleotide unit is 3.3 Å (0.33 nm) long.

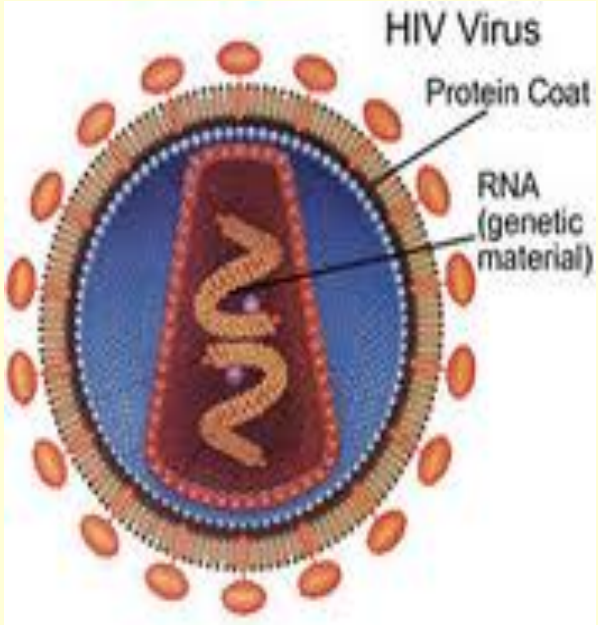
# Ribonucleic acid (RNA)

- **RNA** is a biologically important type of molecule that consists of a long chain of nucleotide units.
- Each nucleotide consists of a nitrogenous base, a ribose sugar, and a phosphate.



# Double-stranded RNA

- Double-stranded RNA (dsRNA) is RNA with two complementary strands, similar to the DNA found in all cells.
- dsRNA forms the genetic material of some viruses (double-stranded RNA viruses).



# Types of RNA

| Type | Abbr | Function | Distribution |
|---|---|---|---|
| Messenger RNA | mRNA | Codes for protein | All organisms |
| Ribosomal RNA | rRNA | Translation | All organisms |
| Transfer RNA | tRNA | Translation | All organisms |

## in post-transcriptional modification

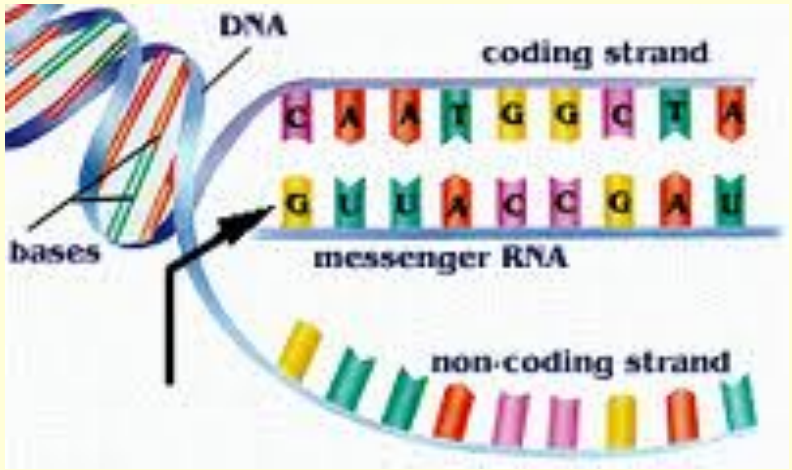
| | | | |
|---|---|---|---|
| Small nuclear RNA | **snRNA** | **Splicing and other functions** | Eukaryotes and archaea |
| Y RNA | | RNA processing, DNA replication | Animals |
| Telomerase RNA | | Telomere synthesis | Most eukaryotes |

## Regulatory RNAs

| | | | |
|---|---|---|---|
| Antisense RNA | aRNA | Transcriptional attenuation / mRNA degradation / mRNA stabilisation / Translation block | All organisms |

# Messenger RNA

- mRNA carries information about a protein sequence to the ribosomes, the protein synthesis factories in the cell.
- It is coded so that every three nucleotides (a codon) correspond to one amino acid.
- In eukaryotic cells, once precursor mRNA (pre-mRNA) has been transcribed from DNA, it is processed to mature mRNA. This removes its introns—non-coding sections of the pre-mRNA.

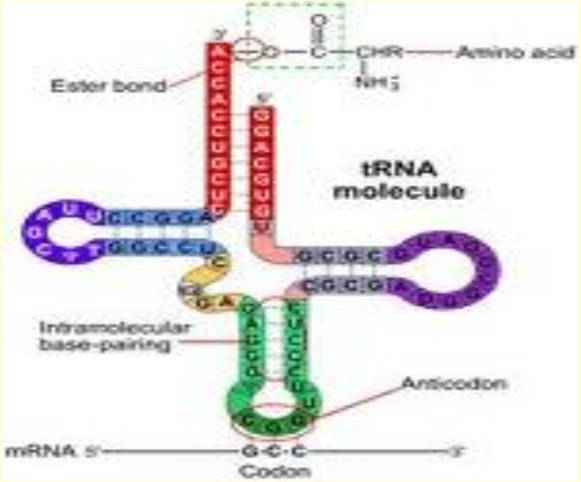


- The mRNA is then exported from the nucleus to the cytoplasm, where it is bound to ribosomes and translated into its corresponding protein form with the help of tRNA.

- In prokaryotic cells, which do not have nucleus and cytoplasm compartments, mRNA can bind to ribosomes while it is being transcribed from DNA.

# Transfer RNA

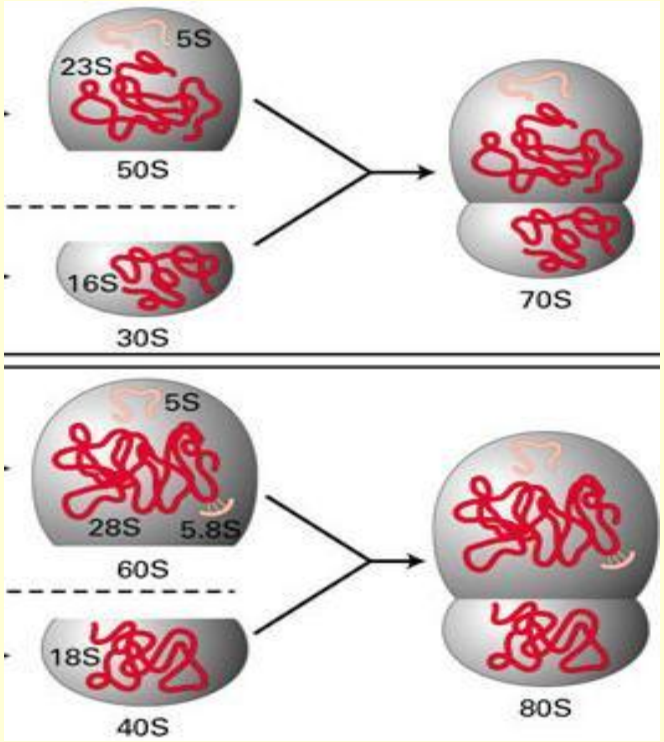
- Transfer RNA (tRNA) is a small RNA chain of about 80 nucleotides that transfers a specific amino acid to a growing polypeptide chain at the ribosomal site of protein synthesis during translation.




- It has sites for amino acid attachment and an anticodon region for codon recognition
- that site binds to a specific sequence on the messenger RNA chain through hydrogen bonding.

# Ribosomal RNA

- Ribosomal RNA (rRNA) is the catalytic component of the ribosomes.
- Eukaryotic ribosomes contain four different rRNA molecules: 18S, 5.8S, 28S and 5S rRNA.
- rRNA molecules are synthesized in the nucleolus.
- In the cytoplasm, ribosomal RNA and protein combine to form a nucleoprotein called a ribosome.
- The ribosome binds mRNA and carries out protein synthesis. Several ribosomes may be attached to a single mRNA at any time.
- rRNA is extremely abundant and makes up 80% of the 10 mg/ml RNA found in a typical eukaryotic cytoplasm.

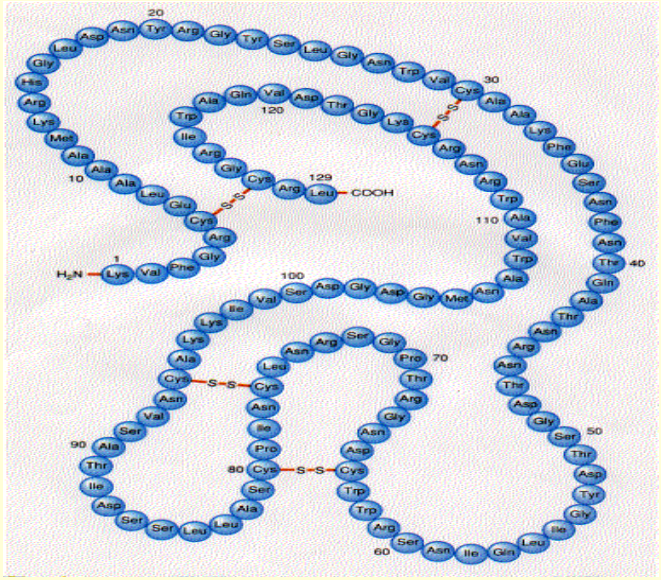


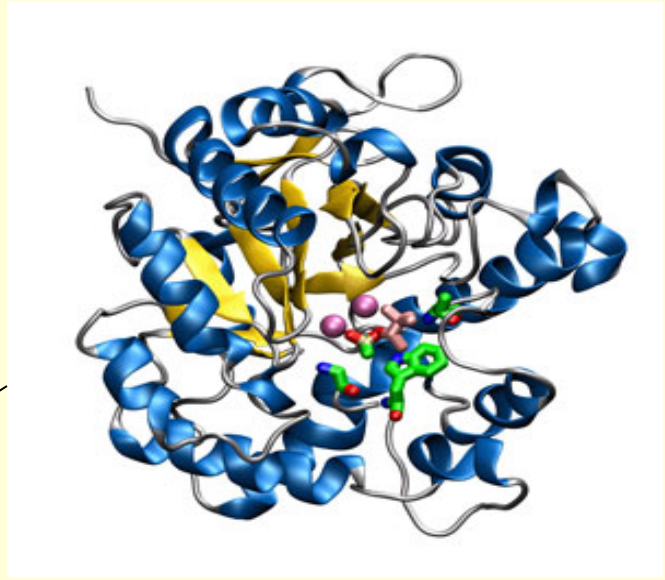
# Difference between RNA & DNA

| RNA | DNA |
| --- | --- |
| RNA nucleotides contain ribose sugar | DNA contains deoxyribose |
| RNA has the base uracil | DNA has the base thymine |
| presence of a hydroxyl group at the 2' position of the ribose sugar. | Lacks of a hydroxyl group at the 2' position of the ribose sugar. |
| RNA is usually single-stranded. | DNA is usually double-stranded. |

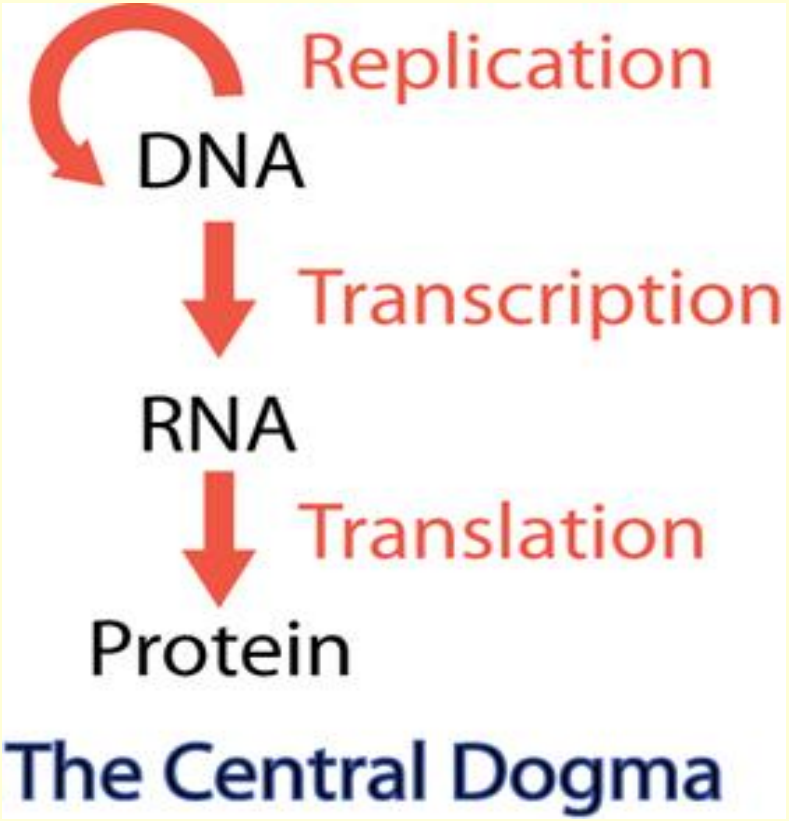
# Protein

- **Proteins** (also known as **polypeptides**) are made of amino acids arranged in a linear chain and folded into a globular form.
- The sequence of amino acids in a protein is defined by the sequence of a gene, which is encoded in the genetic code.
- Genetic code specifies 20 standard amino acids.





**Basic players in molecular biology**: DNA, RNA, and proteins. **What they do is this** :

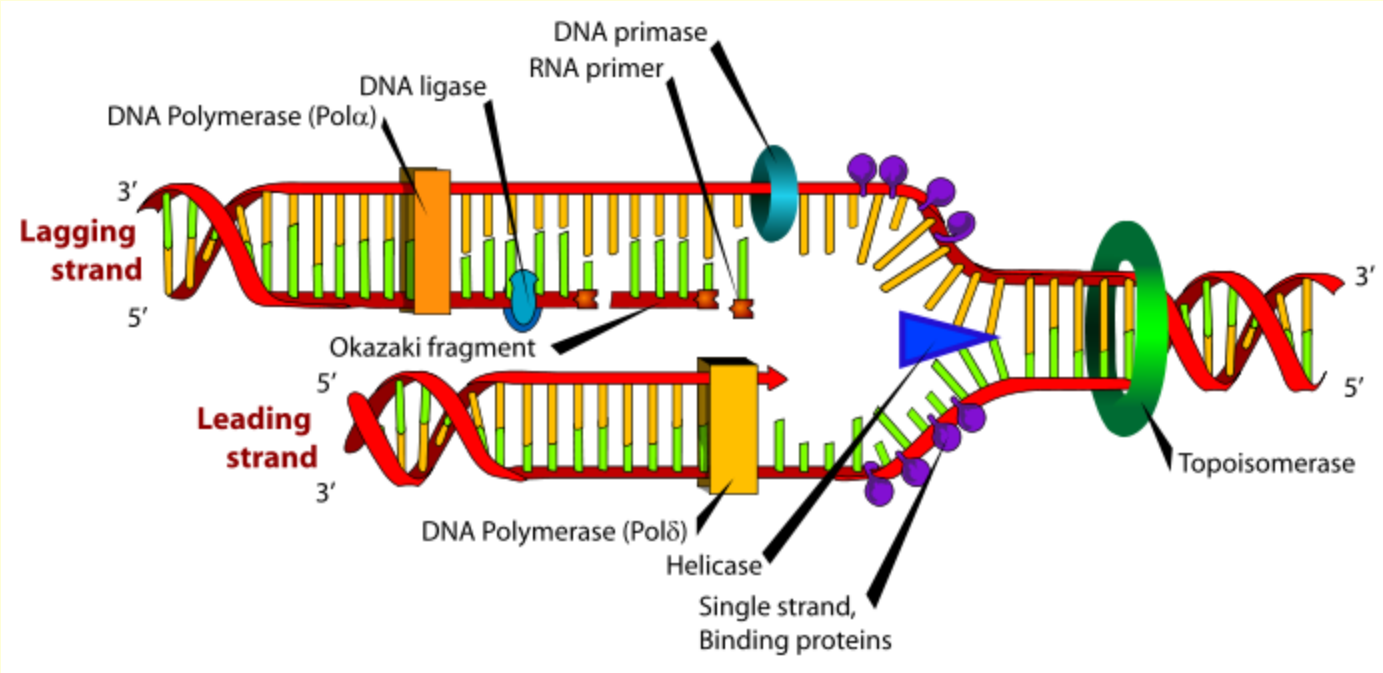


# DNA replication

- DNA replication, the basis for biological inheritance, is a fundamental process occurring in all living organisms to copy their DNA.

- In the process of "replication" each strand of the original double-stranded DNA molecule serves as template for the reproduction of the complementary strand.

- Two identical DNA molecules have been produced from a single double-stranded DNA molecule.




- In a cell, DNA replication begins at specific locations in the genome, called "origins".
- Unwinding of DNA at the origin, and synthesis of new strands, forms a replication fork.
- In addition to DNA polymerase, the enzyme that synthesizes the new DNA by adding nucleotides matched to the template strand, a number of other proteins are associated with the fork and assist in the initiation and continuation of DNA synthesis.
- Cellular proof reading that ensure near perfect fidelity for DNA replication.

# Transcription

- **Transcription**, is the process of creating an equivalent RNA copy of a sequence of DNA.
- Transcription is the first step leading to gene expression.

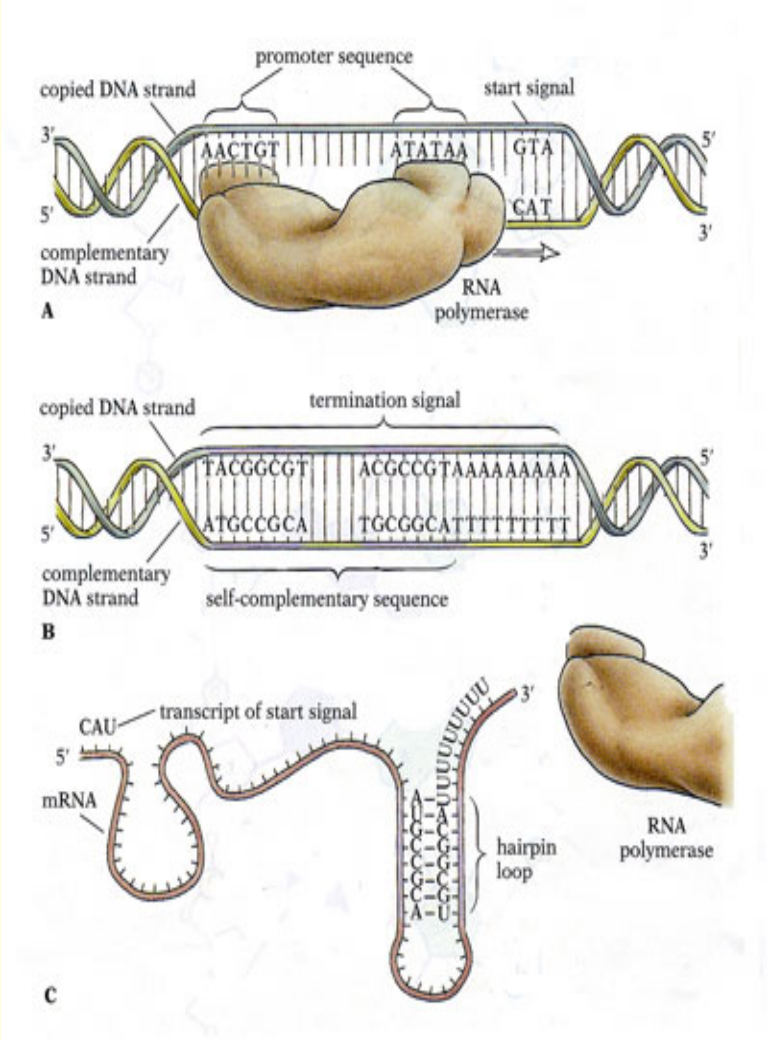
- DNA ⇄ (transcription / reverse transcription) RNA.

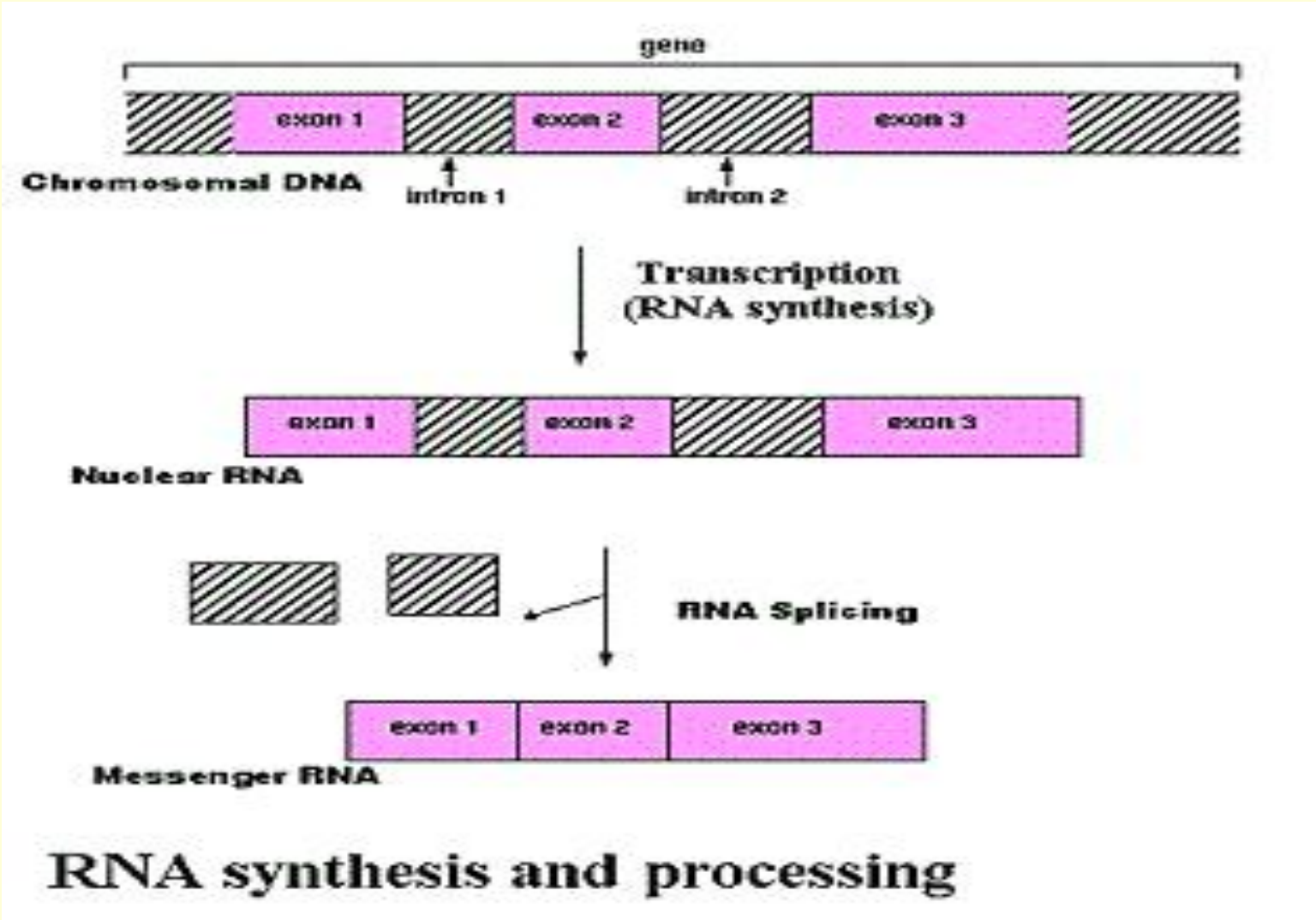
- During transcription, a DNA sequence is read by RNA polymerase, which produces a complementary, antiparallel RNA strand.
- Transcription results in an RNA complement that includes uracil (U) instead of thymine (T).

# Transcription process

- The stretch of DNA transcribed into an RNA molecule is called a transcription *unit* and encodes at least one gene.
- If the gene transcribed encodes for a protein, the result of transcription is messenger RNA (mRNA).
- This mRNA will be used to create that protein via the process of translation.
- Alternatively, the transcribed gene may encode for either rRNA or tRNA, other components of the protein-assembly process, or other ribozymes.
- A DNA transcription unit encoding for protein (the *coding sequence*) and *regulatory sequences* that direct and regulate the synthesis of that protein.

- DNA is read from 3' → 5' during transcription.
- the complementary RNA is created from the 5' → 3' direction.
- only one of the two DNA strands, called the template strand, is used for transcription because RNA is only single-stranded.
- The other DNA strand is called the coding strand.

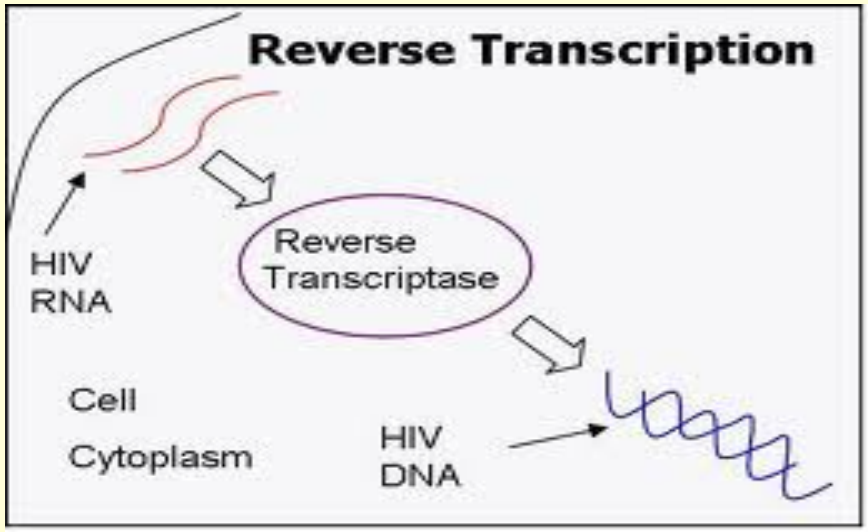


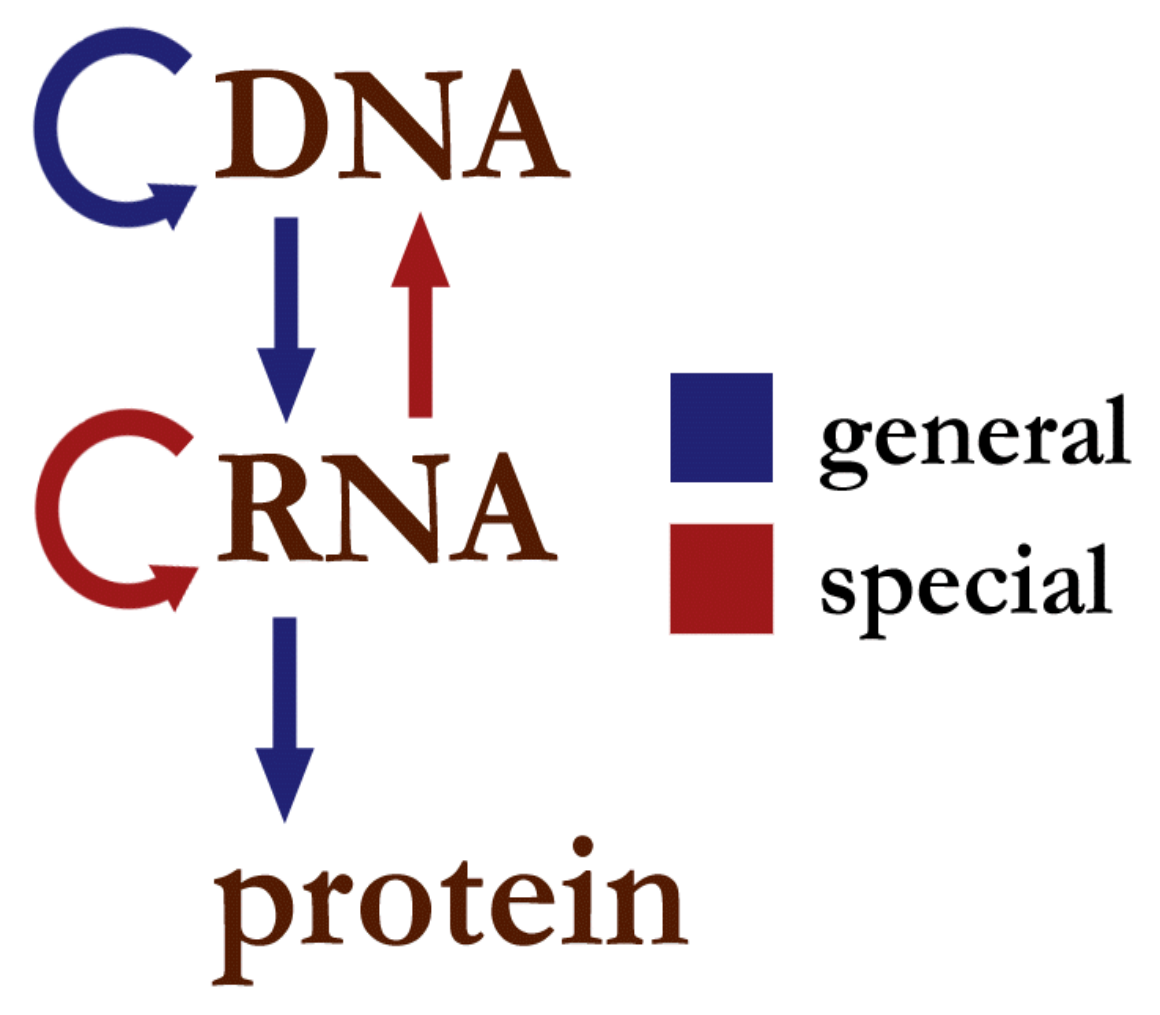



RNA synthesis and processing

# Reverse transcription

- Reverse transcribing viruses replicate their genomes by reverse transcribing DNA copies from their RNA;
- These DNA copies are then transcribed to new RNA.
- Retrotransposans also spread by copying DNA and RNA from one another.




DNA
RNA
protein
general
special

# Translation

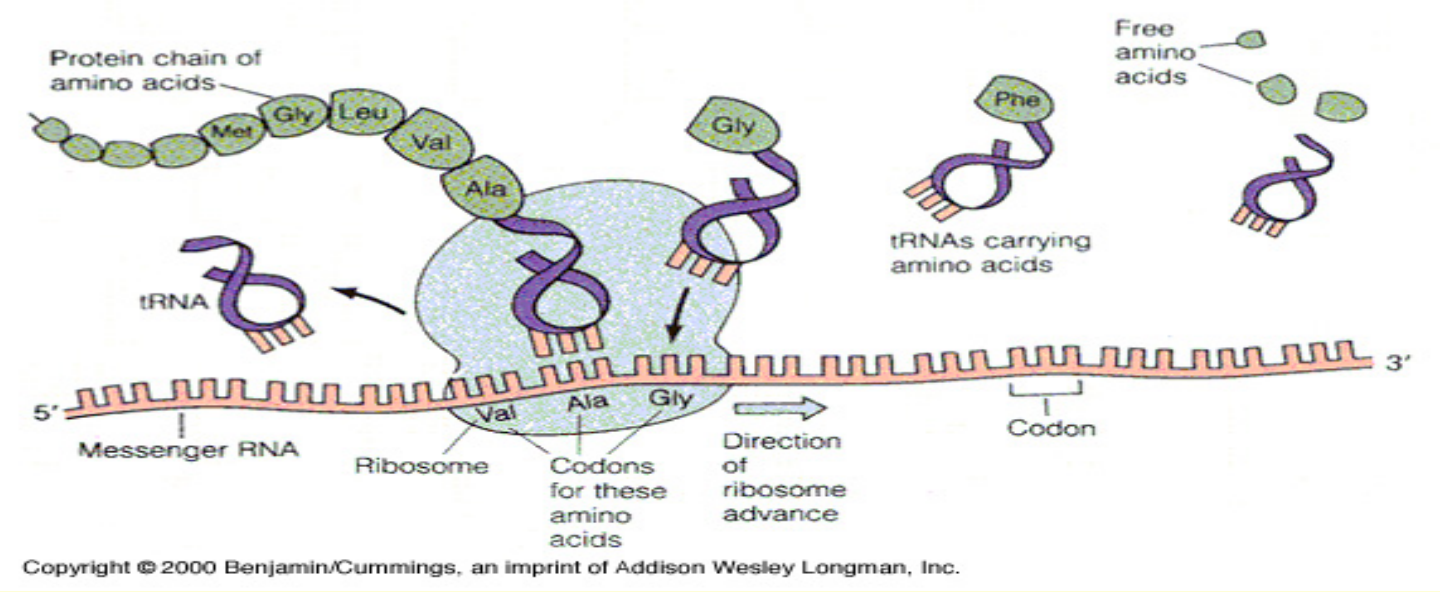
- **Translation** is the first stage of protein biosynthesis .

- In translation, (mRNA) produced by transcription is decoded by the ribosome to produce a specific amino acid chain, or polypeptide, that will later fold into an active protein.

- Translation occurs in the cell's cytoplasm, where the large and small subunits of the ribosome are located, and bind to the mRNA.

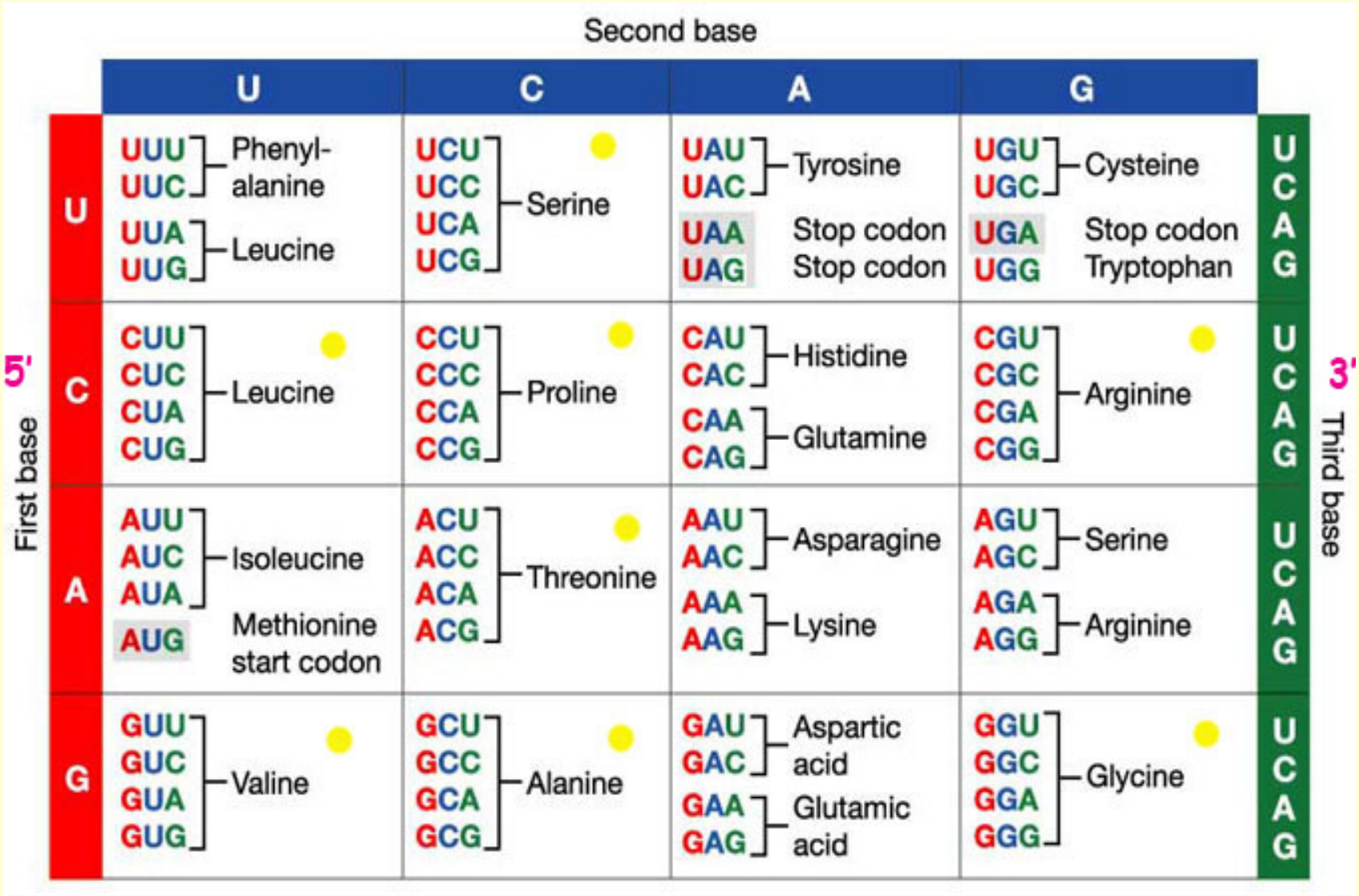
# Translation process

- The ribosome facilitates decoding by inducing the binding of tRNAs with complementary anticodon sequences to mRNA.
- The tRNAs carry specific amino acids that are chained together into a polypeptide as the mRNA passes through and is "read" by the ribosome.

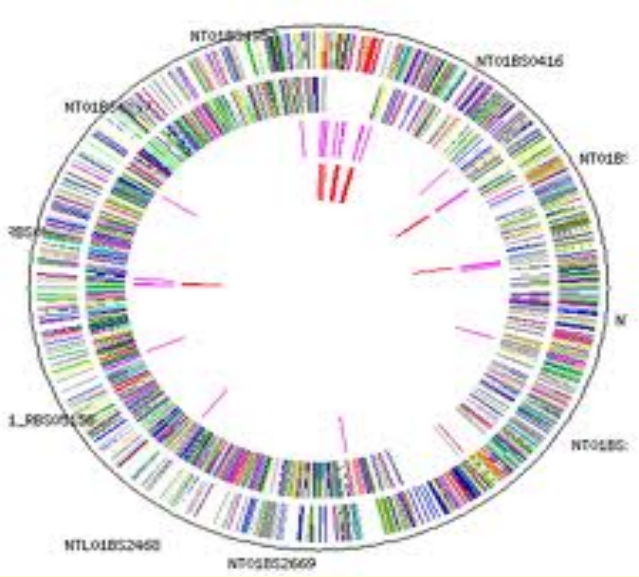



- the entire ribosome/mRNA complex will bind to the outer membrane of the rough endoplasmic reticulum and release the nascent protein polypeptide inside for later vesicle transport and secretion outside of the cell.

# Genetic code



# What is Genome ?



- **Genome** is the entirety of an organism's hereditary information.
- It is encoded either in DNA or, for many types of virus, in RNA.
- The genome includes both the genes and the non-coding sequences of the DNA.

# comparative genome sizes of organisms

| organism | Size (bp) | gene number | average gene density | chromosome number |
|---|---|---|---|---|
| *Homo sapiens* (human) | 3.2 billion | ~25,000 | 1 gene /100,000 bases | 46 |
| *Mus musculus* (mouse) | 2.6 billion | ~25,000 | 1 gene /100,000 bases | 40 |
| *Drosophila melanogaster* (fruit fly) | 137 million | 13,000 | 1 gene / 9,000 bases | 8 |
| *Arabidopsis thaliana* (plant) | 100 million | 25,000 | 1 gene / 4000 bases | 10 |
| *Caenorhabditis elegans* (roundworm) | 97 million | 19,000 | 1 gene / 5000 bases | 12 |
| *Saccharomyces cerevisiae* (yeast) | 12.1 million | 6000 | 1 gene / 2000 bases | 32 |
| *Escherichia coli* (bacteria) | 4.6 million | 3200 | 1 gene / 1400 bases | 1 |
| *H. influenzae* (bacteria) | 1.8 million | 1700 | 1 gene /1000 bases | 1 |

# Why Genome analysis ?

- The prediction of genes in uncharacterised genomic sequences.

- To obtain the complete sequences of as many genomes as possible.

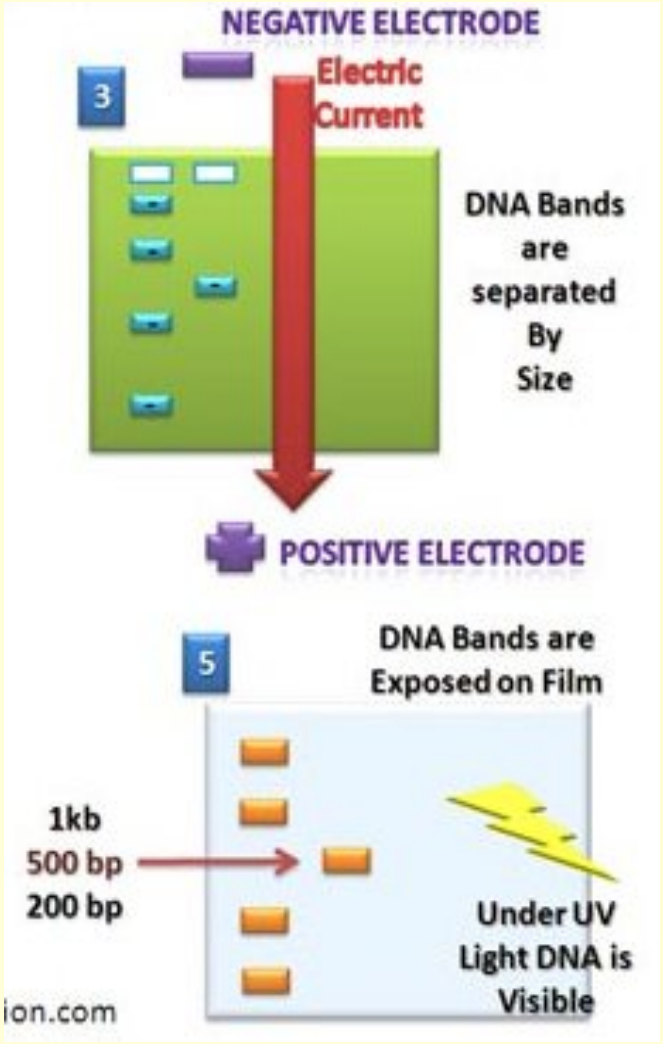
- For Genetic modification.

- Genetic modification to develop new varieties at a faster rate like BT cotton and BT brinjal.

# Tools
# used in
# Molecular Biology

# Gel electrophoresis

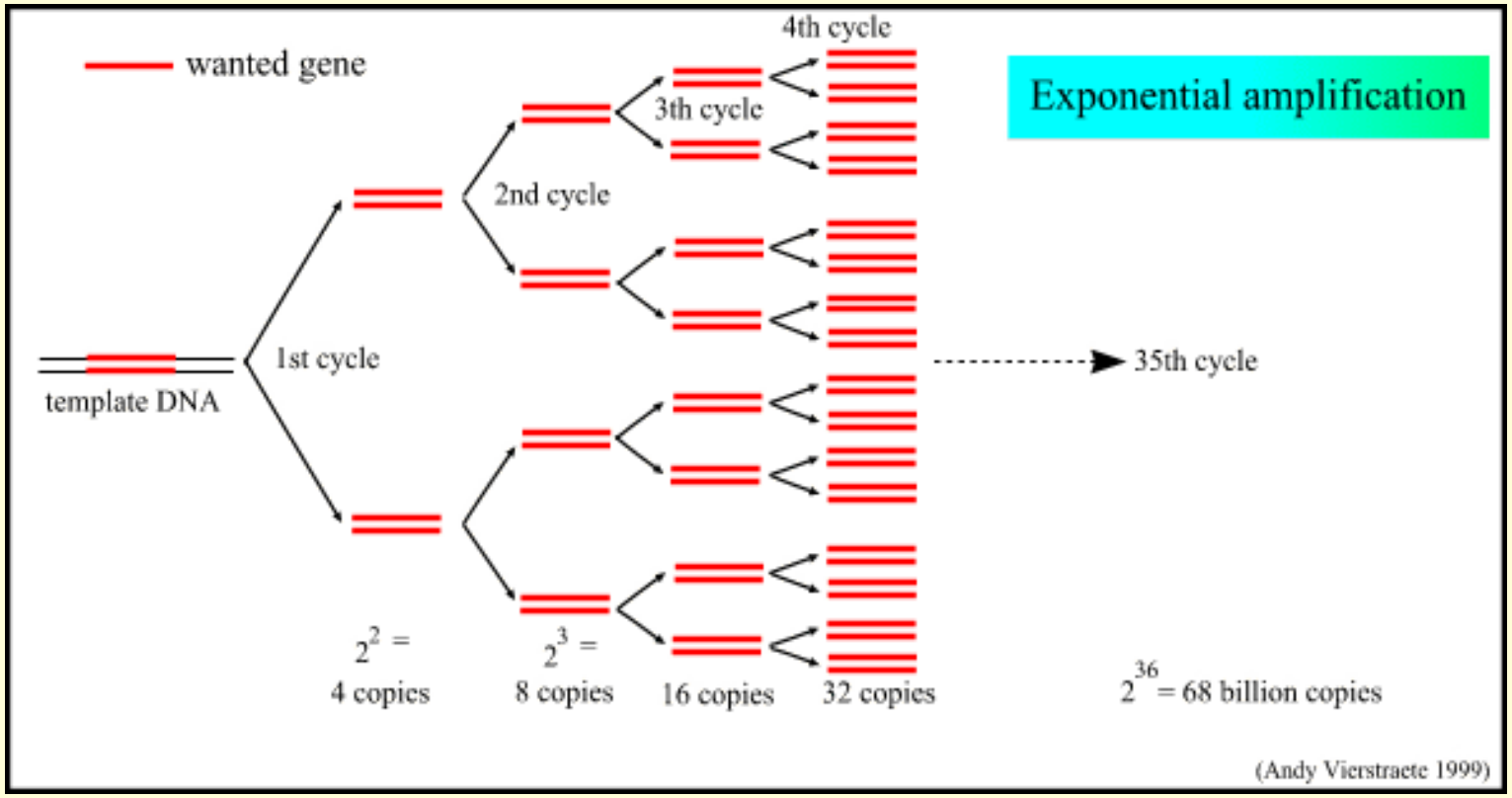
- The basic principle is that DNA, RNA, and proteins can all be separated by means of an electric field.
- In agarose gel electrophoresis, DNA and RNA can be separated on the basis of size by running the DNA through an agarose gel.
- Proteins can be separated on the basis of size by using an SDS-PAGE gel, or on the basis of size and their electric charge by using what is known as a 2D gel electrophoresis.



# Polymerase chain reaction (PCR)

- The polymerase chain reaction is an extremely versatile technique for copying DNA.
- PCR allows a single DNA sequence to be copied (millions of times), or altered in predetermined ways.
- PCR has many variations, like reverse transcription PCR (RT-PCR) for amplification of RNA, and real-time PCR (QPCR) which allow for quantitative measurement of DNA or RNA molecules.

# PCR Analysis




The process follows the principle of DNA replication

# PRIMER

- A **primer** is a strand of nucleic acid that serves as a starting point for DNA synthesis.
- These primers are usually short, chemically synthesized oligonucleotides, with a length of about twenty bases. They are hybredized to a target DNA, which is then copied by the polymerase.
- minimum primer length used in most applications is 18 nucleotides.
- Replication starts at the 3'-end of the primer, and copies the opposite strand.
- In most cases of natural DNA replication, the primer for DNA synthesis and replication is a short strand of RNA .

# Applications of PCR

•A common application of PCR is the study of patterns of gene expression.

•The task of DNA sequencing can also be assisted by PCR.

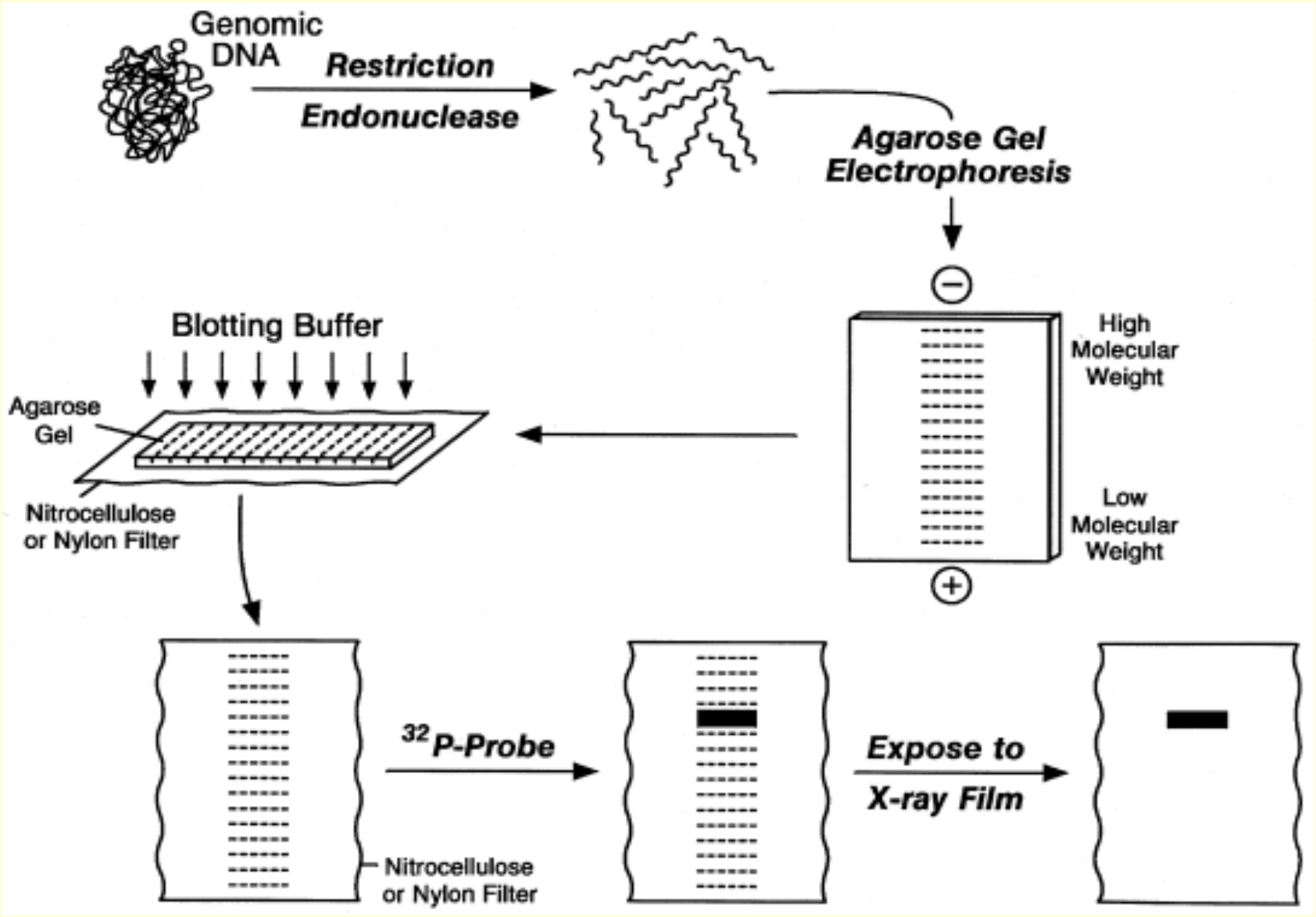
•PCR has numerous applications to the more traditional process of DNA cloning.

•An exciting application of PCR is the phylogenic analysis of DNA from ancient sources

•A common application of PCR is the study of patterns of genetic mapping

•PCR can also used in Parental testing, where an individual is matched with their close relatives.

# Macromolecule blotting & probing



# Southern blotting

- Southern blot is a method for probing for the presence of a specific DNA sequence within a DNA sample.
- DNA samples are separated by gel electrophoresis and then transferred to a membrane by blotting via capillary action.
- The membrane is then exposed to a labeled DNA probe that has a complement base sequence to the sequence on the DNA of interest.
- less commonly used due to the capacity of other techniques, such as PCR.
- Southern blotting are still used for some applications such as measuring transgene copy number in transgenic mice, or in the engineering of gene knockout embryonic stem cell lines.

# Northern blotting

- The northern blot is used to study the expression patterns of a specific type of RNA molecule as relative comparison among a set of different samples of RNA.
- RNA is separated based on size and is then transferred to a membrane then probed with a labeled complement of a sequence of interest.
- The results may be visualized through a variety of ways depending on the label used. Most result in the revelation of bands representing the sizes of the RNA detected in sample.
- The intensity of these bands is related to the amount of the target RNA in the samples analyzed.
- It is used to study when and how much gene expression is occurring by measuring how much of that RNA is present in different samples.
- one of the most basic tools for determining at what time, and under what conditions, certain genes are expressed in living tissues.

# Western blotting

- In western blotting, proteins are first separated by size, in a thin gel sandwiched between two glass plates in a technique known as SDS-PAGE sodium dodecyl sulphate polyacrylamide gel electrophoresis.

- The proteins in the gel are then transferred to a nitrocellulose, nylon or other support membrane.

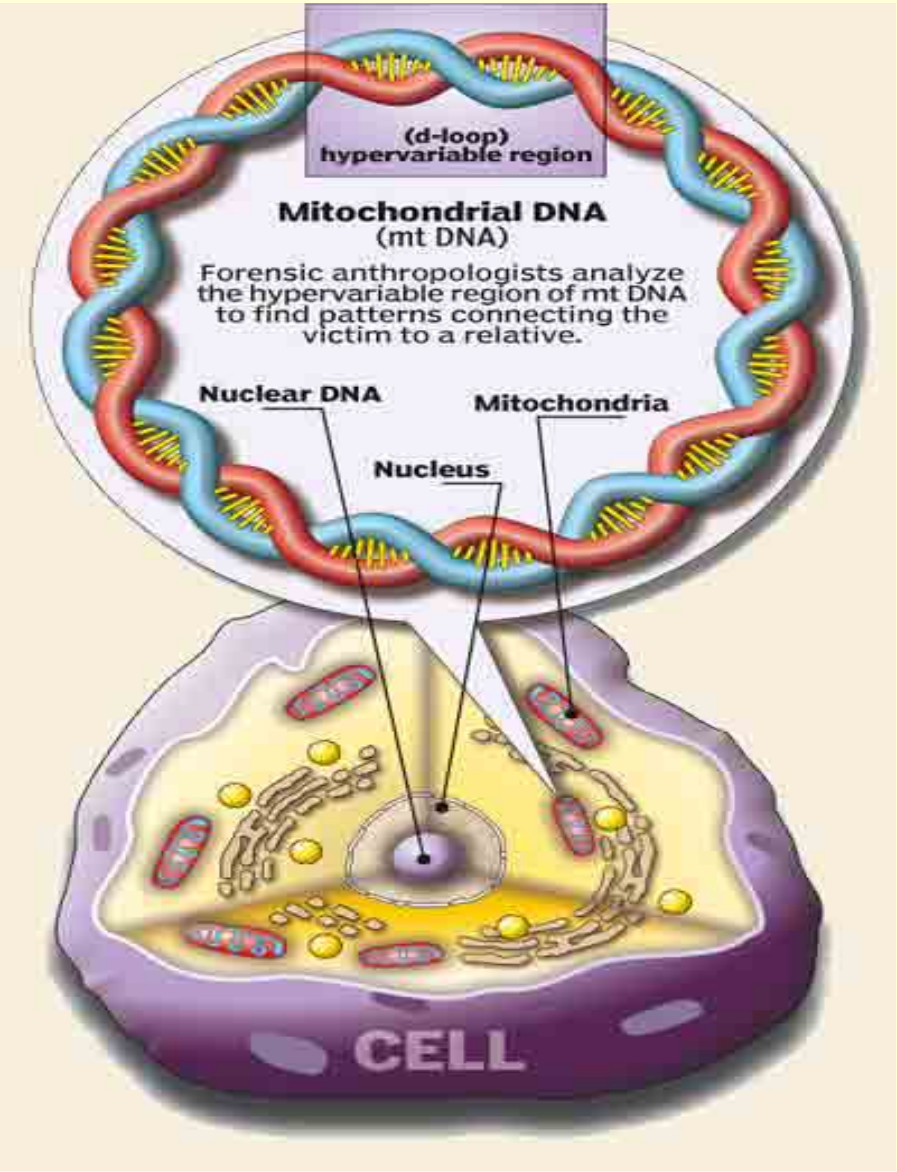
- This membrane probed with solutions of antibodies. Antibodies specifically bind to the protein of interest & visualized by a variety of techniques, including colored products, chemiluminescence, or autoradiography.

- Antibodies are labeled with enzymes. When a chemiluminescent substrate is exposed to the enzyme it allows detection.

- Using western blotting techniques allows not only detection but also quantitative analysis.

# Molecular markers

- Molecular marker are based on naturally occurring polymorphism in DNA sequence(i.e. base pair deletion, substitution ,addition or patterns).
- Genetic markers are sequences of DNA which have been traced to specific locations on the chromosomes and associated with particular traits.
- It can be described as a variation that can be observed.
- A genetic marker may be a short DNA sequence, such as a sequence surrounding a single base-pair change (single nucleotide polymorphism, SNP), a long one, like mini satellites.



# Some commonly used types of genetic markers are

- RFLP (or Restriction fragment length polymorphism)
- AFLP (or Amplified fragment length polymorphism)
- RAPD (or Random amplification of polymorphic DNA)
- VNTR (or Variable number tandem repeat)
- Micro satellite polymorphism, SSR (or Simple sequence repeat)
- SNP (or Single nucleotide polymorphism)
- STR (or Short tandem repeat)
- SFP (or Single feature polymorphism)
- DArT (or Diversity Arrays Technology)
- RAD markers (or Restriction site associated DNA markers)

# There are 5 conditions that characterize a suitable molecular marker

- Must be polymorphic
- Co-dominant inheritance
- Randomly and frequently distributed throughout the genome
- Easy and cheap to detect
- Reproducible

# Molecular markers can be used for several different applications including

- Germplasm characterization,
- Genetic diagnostics,
- Characterization of transformants,
- Study of genome
- Organization and phylogenic analysis.
- Paternity testing and the investigation of crimes.
- Measure the genomic response to selection in livestock

# RFLP (Restriction fragment length polymorphism)

RFLPs involves fragmenting a sample of DNA by a restriction enzyme, which can recognize and cut DNA wherever a specific short sequence occurs. A RFLP occurs when the length of a detected fragment varies between individuals and can be used in genetic analysis.

**Advantages:**

- Variant are co dominant
- Measure variation at the level of DNA sequence, not protein sequence.

**Disadvantage:**

- Requires relatively large amount of DNA

# AFLP ( Amplified fragment length polymorphism)

In this analysis we can amplify restricted fragments and reduces the complexity of material to be analyzed (approx 1000 folds).it can be used for *comparison b/w closely related species only.*

**Advantages:**

- Fast
- Relatively inexpensive
- Highly variable

**Disadvantage:**

- Markers are dominant
- Presence of a band could mean the individual is either homozygous or heterozygous for the Sequence - can't tell which?

# RAPD ( Random amplification of polymorphic DNA)

**R**andom **A**mplification of **P**olymorphic **D**NA. It is a type of PCR reaction, but the segments of DNA that are amplified are random.

**Advantages:**

- Fast
- Relatively inexpensive
- Highly variable

**Disadvantage:**

- Markers are dominant
- Presence of a band could mean the individual is either homozygous or heterozygous for the Sequence - can't tell which?
- Data analysis more complicated

# Micro satellite polymorphism, SSR or Simple sequence repeat

Microsatellites, Simple Sequence Repeats (SSRs), or Short Tandem Repeats (STRs), are repeating sequences of 1-6 base pairs of DNA.

**Advantages:**

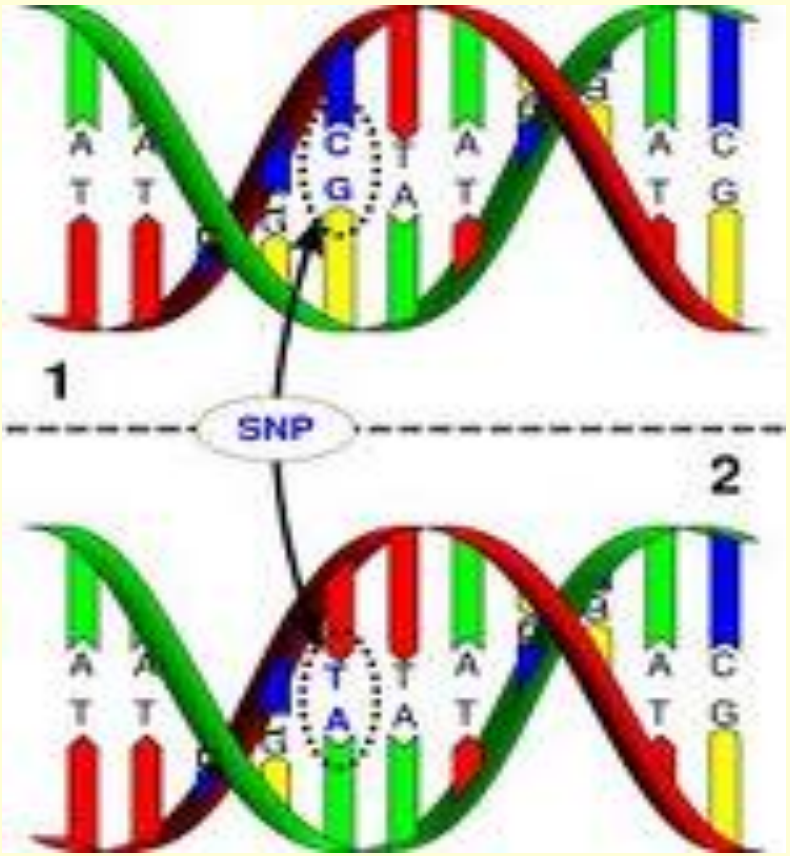
- Highly variable
- Fast evolving
- Co dominant

**Disadvantage:**

- Relatively expensive and time consuming to develop

# SNP

- A **single-nucleotide polymorphism** (**SNP**, pronounced *snip*) is a DNA sequence variation occurring when a single nucleotide — A, T,C, or G — in the genome (or other shared sequence) differs between members of a species or paired chromosomes in an individual.
- Used in biomedical research ,crop and livestock breeding programs.



# STR

- A **short tandem repeat** (STR) in DNA occurs when a pattern of two or more nucleotides are repeated and the repeated sequences are directly adjacent to each other.
- The pattern can range in length from 2 to 16 base pairs (bp) (for example $(CATG)_n$ in a genomic region) and is typically in the non-coding intron region
- Used in forensic cases.
- used for the genetic fingerprinting of individuals

# PRINCIPLES OF DNA ISOLATION & PURIFICATION

DNA can be isolated from any nucleated cell.

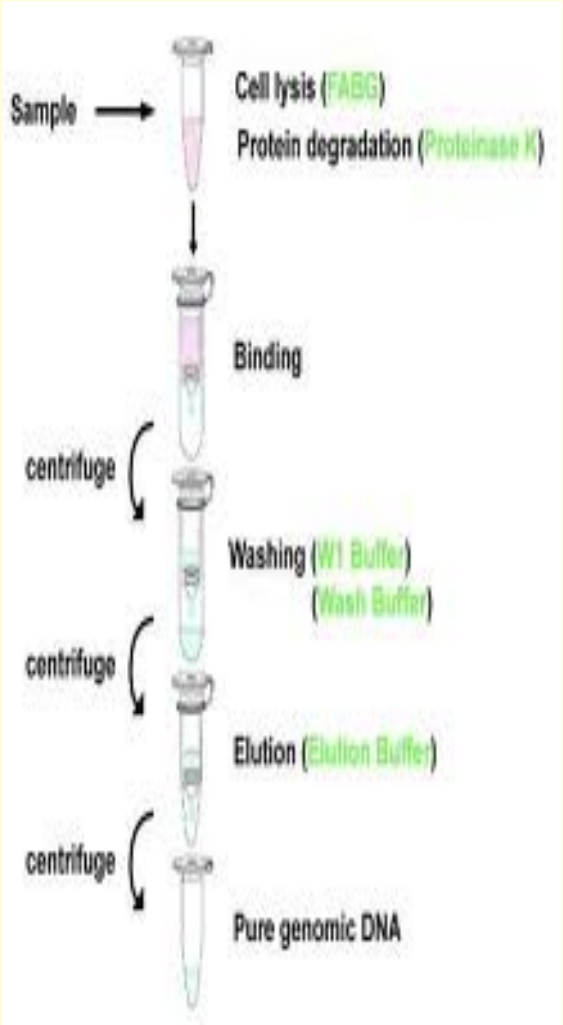
DNA is a giant anion in solution.

# Sources of DNA include

- Blood
- Buccal cells
- Cultured cells (plant and animal)
- Bacteria
- Biopsies
- Forensic samples i.e. body fluids, hair follicles, bone & teeth roots.

**DNA isolation** is a routine procedure to collect DNA for subsequent molecular analysis. There are three basic steps in a DNA extraction:

- **Cell disruption**:- This is commonly achieved by grinding or sonicating the sample. Removing membrane lipids by adding a detergent.
- **Isolation of DNA**:- Removing proteins by adding a protease (optional but almost always done).
- **Precipitating the DNA** :-usually ice-cold ethanol or isopropanol is used. Since DNA is insoluble in these alcohols, it will aggregate together, giving a *pellet* upon centrifugation. This step also removes alcohol soluble salt.



# Basic rules

- **Blood** – first lyse (explode) the red blood cells with a gentle detergent such as Triton-X-100.
- **Wash cells** – haemoglobin (and other pigments) inhibits restriction enzymes and TAQ polymerase.
- Work on **ice** to slow down enzymatic processes.
- Wear **gloves** to protect your samples from you!!
- **Autoclave** all solutions and store in fridge (except SDS and organic solvents!)
- Keep all pellets & supernatants until you have the DNA you want.

# Getting to the DNA

- Cells – **lyse** all cells in presence of :

- **NaCl** so that DNA is stabilised and remains as a double helix,
- **EDTA** which chelates Mg++ and is a co-factor of DNAse which chews up DNA rapidly.
- **anionic detergent SDS** which disrupts the lipid layers, helps to dissolve membranes & binds positive charges of chromosomal proteins (*histones*) to release the DNA into the solution.
- Include a **protease** *(proteinase K*) to digest the proteins
- incubate the solution at an **elevated temperature** (56$^{o}$C to inhibit degradation by DNAses) for 4-24 hrs.

# Getting rid of the protein

- **Organic solvent extraction** using equal volume phenol:chloroform (24:1)

- Protein at the interface after centrifugation (10000 rpm at $10^{o}$ c for 10 min.)

# Precipitating the DNA

- add 2.5 - 3 volumes **ice-cold 95% ethanol** to the DNA & leave at -20$^{o}$C overnight.
- Centrifuge sample at 10000 rpm ,10 min., 4$^{0}$C.
- **Wash** DNA pellet to remove excess salt in 70% EtOH and air-dry.
- **Resuspend** in sterile distilled water(pH7.4)
- Store at 4$^{o}$C or frozen at -20$^{o}$C long term.

# Quantifying the DNA

- The amount of DNA can be quantified using the formula:

$$\text{DNA concentration } (\mu g/ml) = \frac{OD_{260} \times 100 \text{ (dilution factor)} \times 50\ \mu g/ml}{1000}$$

- Nucleic acids have a peak absorbance in the ultraviolet range at about 260 nm

- 1 A260 O.D. unit for dsDNA = 50 µg/ml
- 1 A260 O.D. unit for ssDNA = 33 µg/ml
- 1 A260 O.D. unit for RNA = 40 µg/ml

# DNA purity

- The purity of the DNA is reflected in the OD260:OD 280 ratio and must be between 1.6 and 2.00.

  < 1.6 – protein contaminated

  > 2.0 – chloroform / phenol contaminated

- Repurify sample.

# Summary

- Sample for DNA extraction
- Lysis of cells at elevated temperature + detergent + enzyme in salt buffer
- Removal of cellular proteins
- Precipitation of nucleic acids with ethanol
- Quantitation and purity measurement of DNA

# Future aspects

- For agricultural development and environment protection.
- To ensure food security for ever growing human population.


Thank you