Basics of Molecular biology

- Molecular biology is the study of biology at molecular level.
- This field overlaps with other areas of biology and chemistry, particularly genetics and biochemistry (Fig 1).

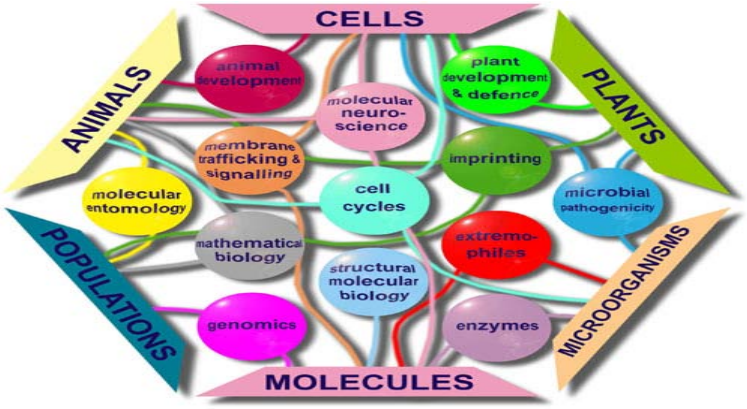

Fig: 1 molecular biology frame with other branches of biology

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

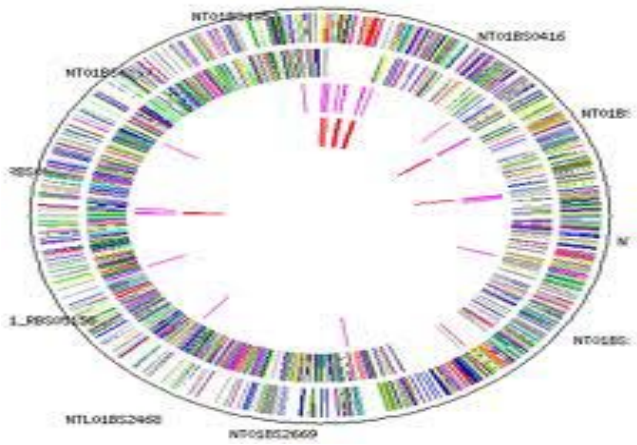

Fig 2: Genome

- **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.

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.

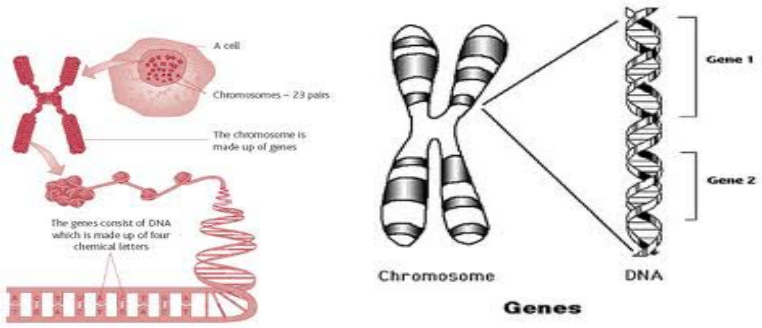

Fig 3: Gene location in cell

- DNA is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms and some viruses.
- The main role of DNA molecules is the long-term storage of information.

- DNA is a set of blueprints needed to construct other components of cells, such as proteins and RNA molecules.
- The DNA segments that carry this genetic information are called genes, but other DNA sequences have structural purposes & involved in regulating the use of this genetic information.
- DNA exist as a pair of molecules that are held tightly together.
- These two long strands make the shape of a double helix.

Chemical structure of DNA

- Chemically, DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds.
- A base linked to a sugar is called a nucleoside and a base linked to a sugar and one or more phosphate groups is called a nucleotide.
- If multiple nucleotides are linked together, as in DNA, this polymer is called a polynucleotide.
- These two strands run in opposite directions to each other and are therefore anti-parallel. Attached to each sugar is one of four types of molecules called bases.

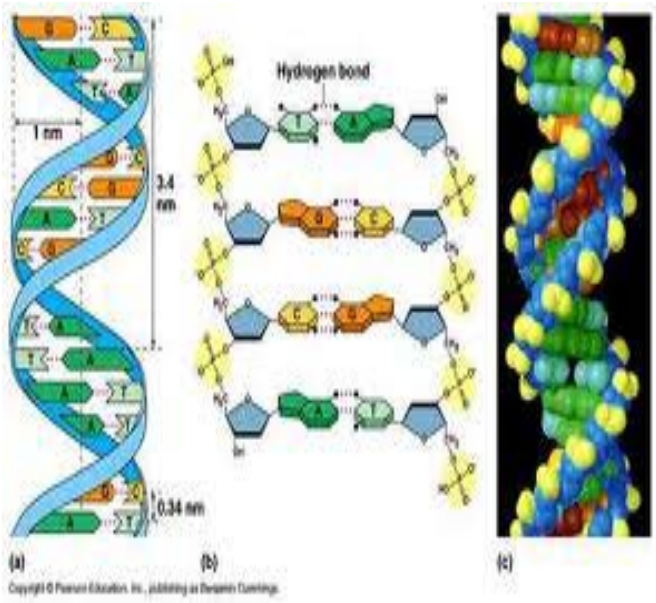

Fig 4 : DNA structure

- Base + sugar + phosphate = nucleotide
- Base + sugar = nucleoside

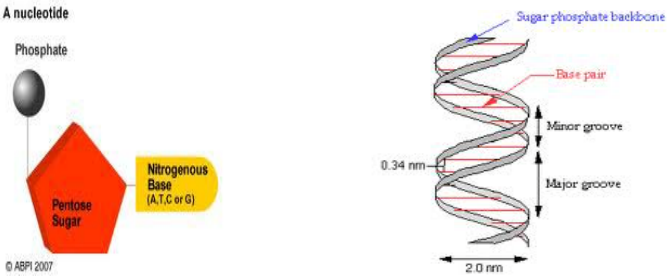

Fig 5: DNA size

- 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.
- The backbone of the DNA strand is made from alternating phosphate and sugar residues.
- The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar rings.

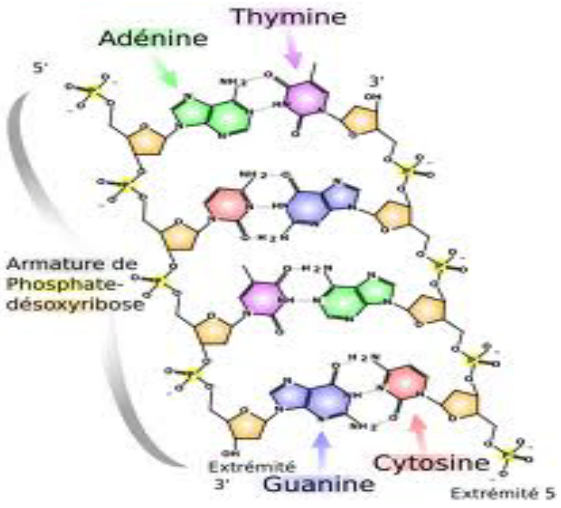

Fig 6: chemical structure of DNA

- These asymmetric bonds mean a strand of DNA has a direction. In a double helix the direction of the nucleotides in one strand is opposite to their direction in the other strand: the strands are *antiparallel*.

- The asymmetric ends of DNA strands are called the 5′ (*five prime*) and 3′ (*three prime*) ends, with the 5' end having a terminal phosphate group and the 3' end a terminal hydroxyl group.
- 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.

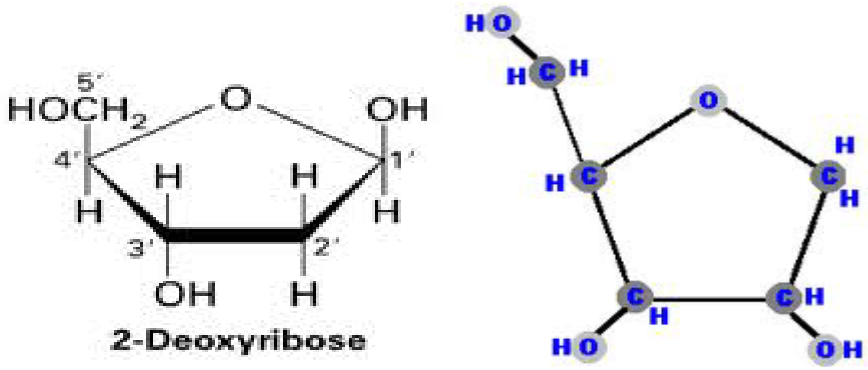

Fig 7: sugars in DNA and RNA

- Bases are classified into two 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.

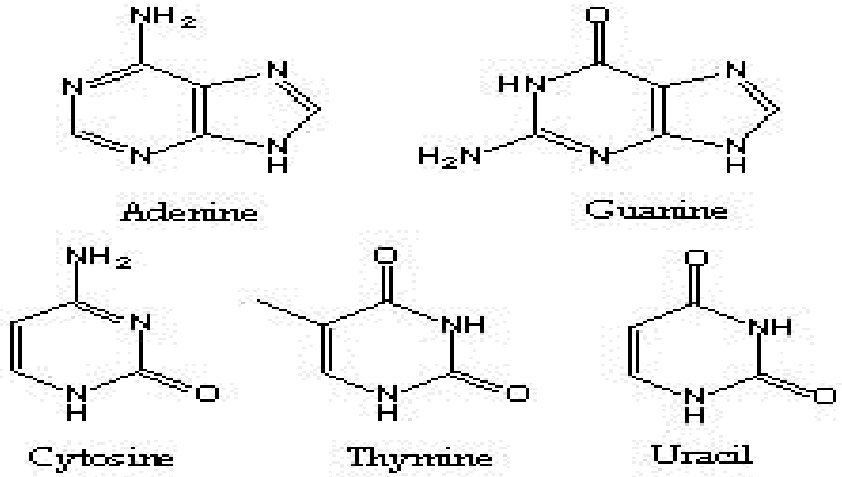

Fig 8: Bases in nucleic acids

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.

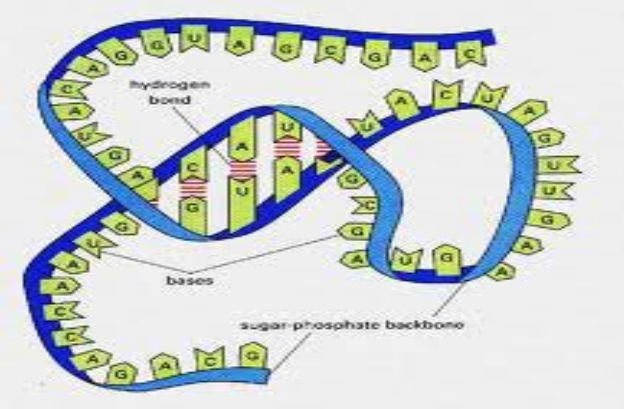

Fig 9: RNA structure

RNA is transcribed from DNA by enzymes called RNA polymerases and is generally further processed by other enzymes.

RNA is central to protein synthesis.

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.

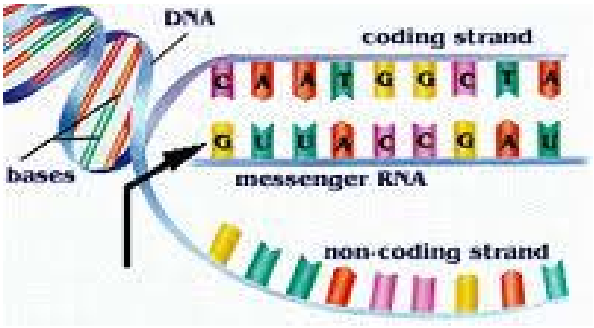

Fig 9: mRNA structure

- 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.
- After a certain amount of time the message degrades into its component nucleotides with the assistance of ribonucleases.

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.
- Three of the rRNA molecules are synthesized in the nucleolus, and one is synthesized elsewhere.
- 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.

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).

- Double-stranded RNA such as viral RNA or siRNA can trigger RNA interference in eukaryotes, as well as interferon response in vertebrates

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.
- Retrotransposons also spread by copying DNA and RNA from one another.

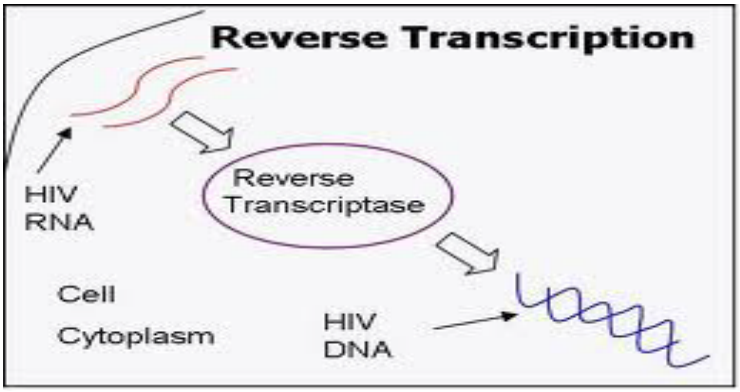

Fig 10: Reverse transcription

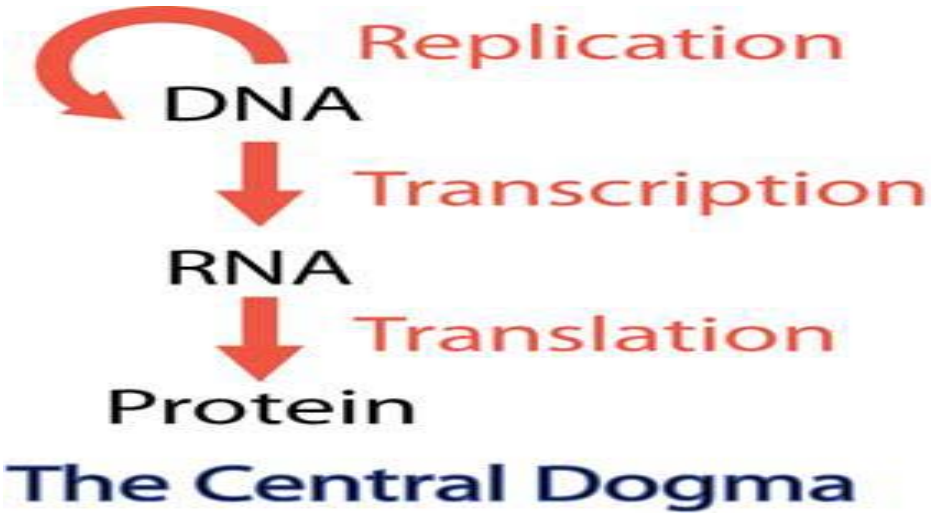

Fig 11: Central dogma of molecular biology

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.
- Cellular proofreading that ensure near perfect fidelity for DNA replication.

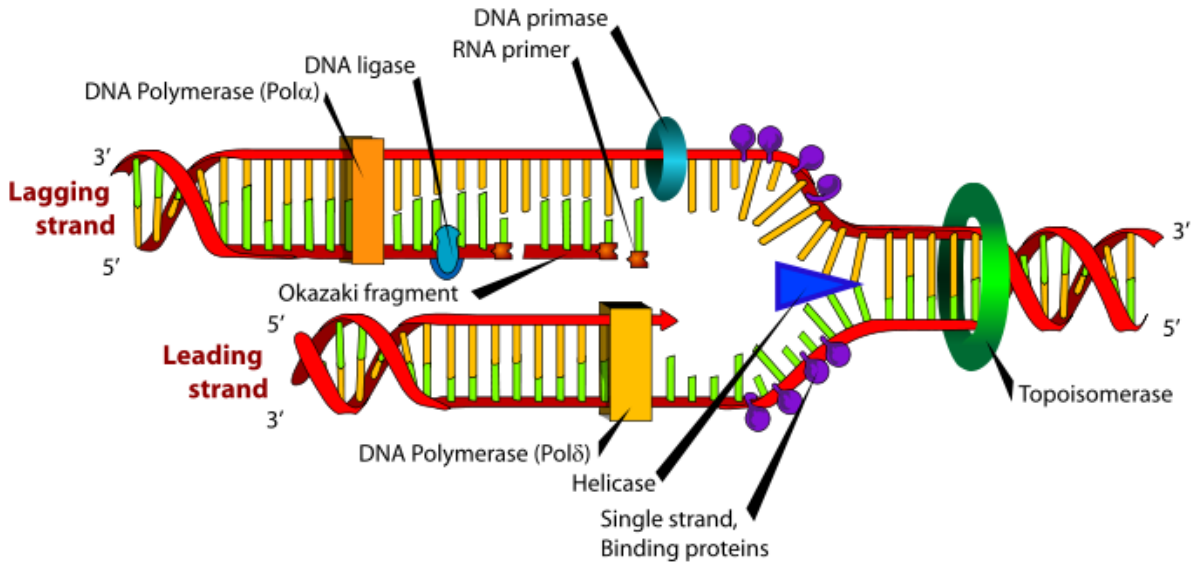

Fig 12: DNA Replication

- 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.

DNA replication- *in vitro*

- DNA replication can also be performed *in vitro* (outside a cell).
- DNA polymerases, isolated from cells, and artificial DNA primers are used to initiate DNA synthesis at known sequences in a template molecule.

- The polymerase chain reaction (PCR), a common laboratory technique, employs such artificial synthesis in a cyclic manner to amplify a specific target DNA fragment from a pool of DNA.

PRIMER

- A **primer** is a strand of nucleic acid that serves as a starting point for DNA synthesis.
- They are required because the enzymes that catalyze replication, DNA polymerases, can only add new nucleotides to an existing strand of DNA.
- The polymerase starts replication 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 (which can be made *de novo*).

Transcription

- **Transcription**, is the process of creating an equivalent RNA copy of a sequence of DNA.
- Both RNA and DNA are nucleic acids, which use base pairs of nucleotides as a complementary language that can be converted back and forth from DNA to RNA in the presence of the correct enzymes.
- During transcription, a DNA sequence is read by RNA polymerase, which produces a complementary, antiparallel RNA strand.
- As opposed to DNA replication, transcription results in an RNA complement that includes uracil (U) in all instances where thymine (T) would have occurred in a DNA complement.
- Transcription is the first step leading to gene expression.
- 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.

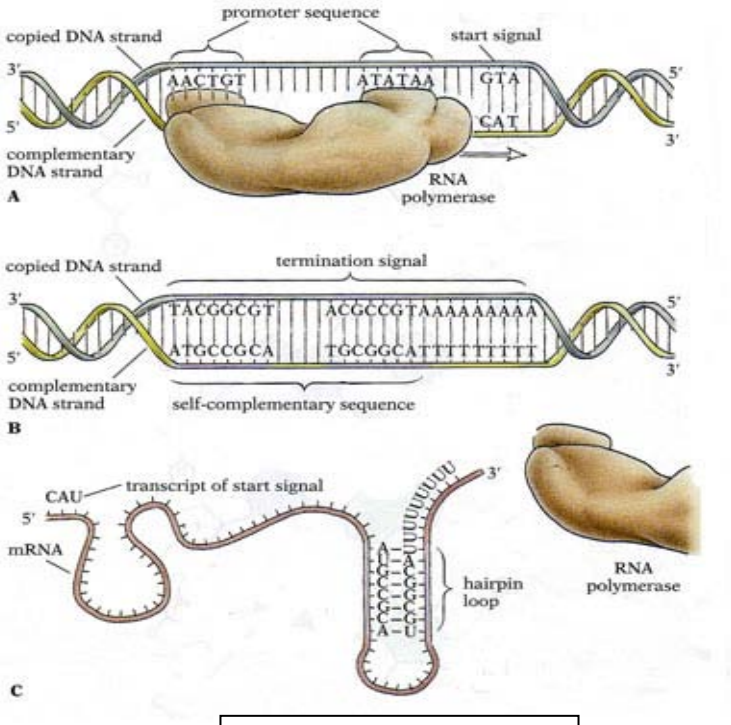

Fig 13: Transcription

- 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, because its sequence is the same as the newly created RNA transcript (except for the substitution of uracil for thymine).

Translation

- **Translation** is the first stage of protein biosynthesis .
- Translation proceeds in four phases: activation, initiation, elongation and termination.
- 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.

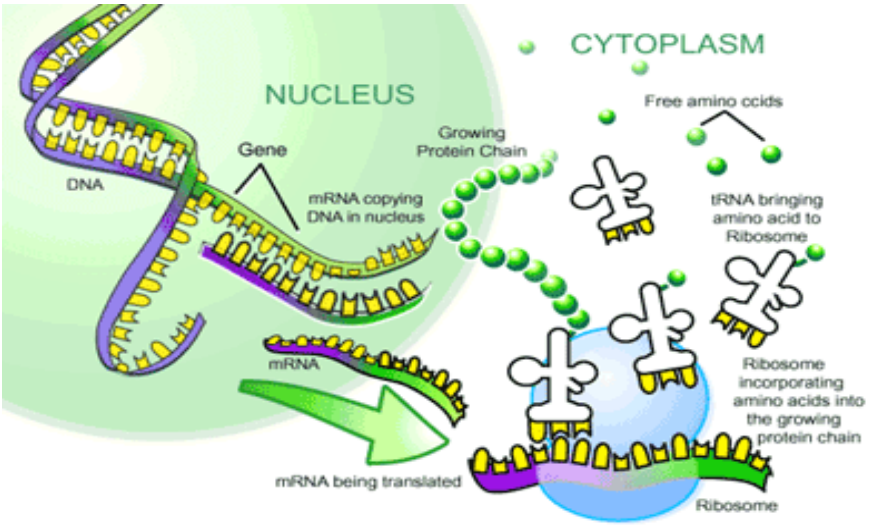

Fig 14: Translation

- 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.

Many types of transcribed RNA, such as transfer RNA, ribosomal RNA, and small nuclear RNA, do not undergo translation into proteins.

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.
- For example, PCR can be used to introduce restriction enzyme sites, or to mutate (change) particular bases of DNA, the latter is a method referred to as "Quick change".
- PCR can also be used to determine whether a particular DNA fragment is found in a cDNA library.

- 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.
- Makes millions of exact copies of DNA from a biological sample
- Allows DNA analysis on biological samples as small as a few skin cells
- Enables even highly degraded samples to be analyzed
- Great care must be taken to prevent contamination with other biological materials during the identifying, collecting, and preserving of a sample

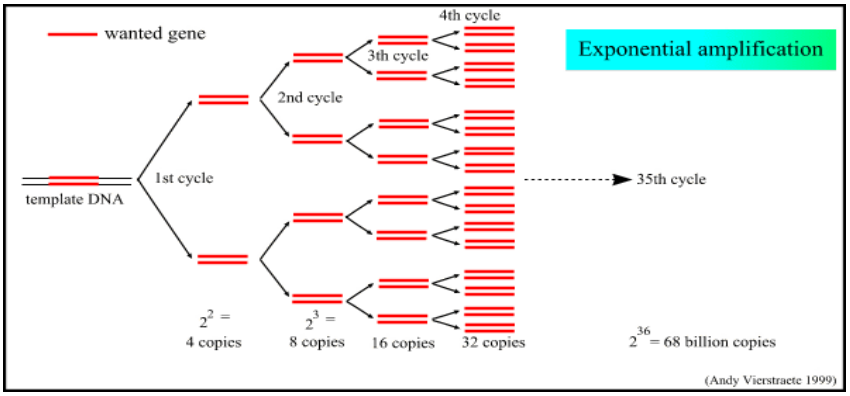

Fig 15: Polymerase chain reaction

- **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.

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.

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 before or after restriction enzyme digestion 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.
- Most original protocols used radioactive labels, however non-radioactive alternatives are now available.
- 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.
- It is essentially a combination of denaturing RNA gel electrophoresis, and a blot.
- 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.

Microarray

- A DNA microarray is a collection of microscopic DNA spots attached to a solid surface, such as glass, plastic or silicon chip forming an array for the purpose of expression profiling, monitoring expression levels for thousands of genes simultaneously .
- The affixed DNA segments are known as probes, thousands of which can be placed in known locations on a single DNA micro array.
- Single nucleotide polymorphisms (SNP) micro arrays -a particular type of DNA micro arrays that are used to identify genetic variation in individuals and across populations.

- Analysis of SNPs utilizing very short amplicons can provide a more sensitive analysis on degraded DNA

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), or a long one, like mini satellites.
- They can be further categorized as dominant or co-dominant. Dominant markers allow for analyzing many loci at one time, e.g. RAPD. A primer amplifying a dominant marker could amplify at many loci in one sample of DNA with one PCR reaction.
- Co-dominant markers analyze one locus at a time. A primer amplifying a co-dominant marker would yield one targeted product.

Some commonly used types of genetic markers are

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

RAD markers (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)

Advantages:

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

Disadvantage:

- Labor intensive
- Requires relatively large amount of DNA

AFLP (Amplified fragment length polymorphism)

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?

Micro satellite polymorphism

Advantages:

- Highly variable
- Fast evolving
- Co dominant

Disadvantage:

- Relatively expensive and time consuming to develop

RAPD (Random amplification of polymorphic DNA)

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

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.

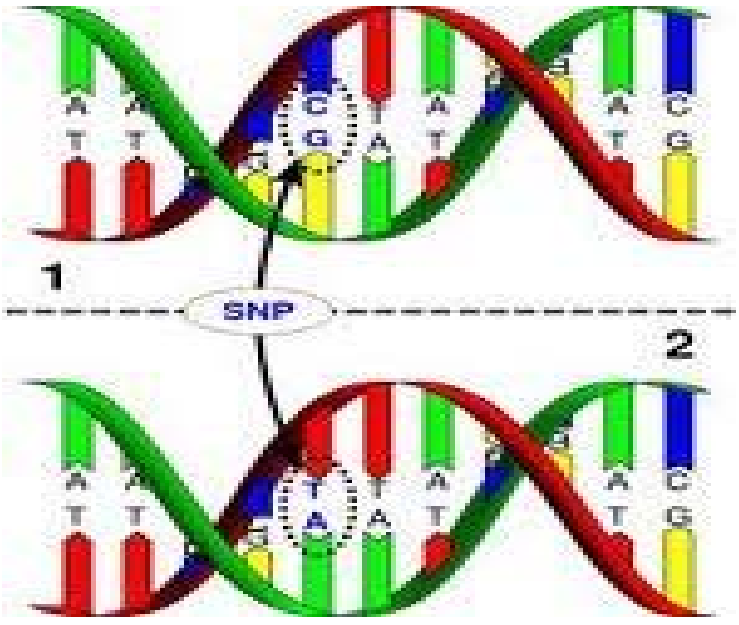

Fig 16: Single nucleotide polymorphism (SNP)

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
- Bacterial plasmids, cosmids
- Biopsies
- Forensic samples i.e. body fluids, hair follicles, bone & teeth roots.

DNA isolation is a routine procedure to collect DNA for subsequent molecular or forensic analysis. There are three basic and one optional steps in a DNA extraction

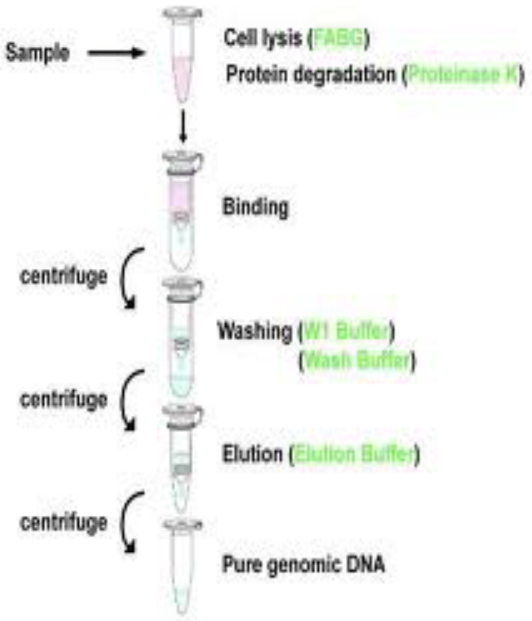

Fig 17: DNA Extraction

- Breaking the cells open, commonly referred to as cell disruption or cell lysis, to expose the DNA within. This is commonly achieved by grinding or sonicating the sample.
- Removing membrane lipids by adding a detergent.
- Removing proteins by adding a protease (optional but almost always done).
- Precipitating the DNA with an alcohol — usually ice-cold ethanol or isopropanol. 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 (10K, 10^{o}, 10')
- **Salt-out** proteins by precipitation with NaCl or Na-acetate – protein pelleted after centrifugation.

Precipitating the DNA

- add 2.5 - 3 volumes **ice-cold 95% ethanol/acetate** to the DNA & leave at -20^{o}C overnight.
- Centrifuge sample at 10K ,10', 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 "absorbance" (O.D.) of a chemical is the:
- product of its (concentration) x (optical path length) x (extinction coefficient, E).
- 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 or 50 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.
- Decreased 260:280 ratio means that contaminating protein is still present.

  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