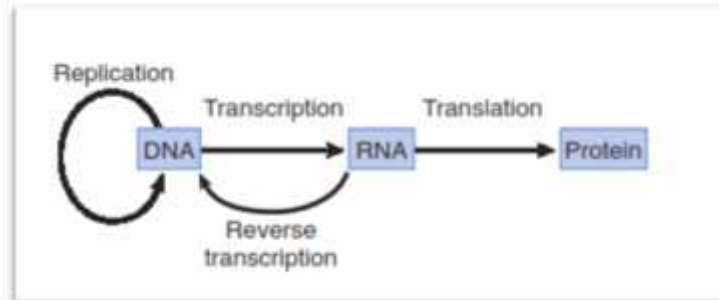


Nucleic Acid Structure and Organization

🌈 Central Dogma of Molecular Biology 📄

An organism must be able to store and preserve its genetic information, pass that information along to future generations, and express that information as it carries out all the processes of life. The major steps involved in handling genetic information are illustrated by the central dogma of molecular biology



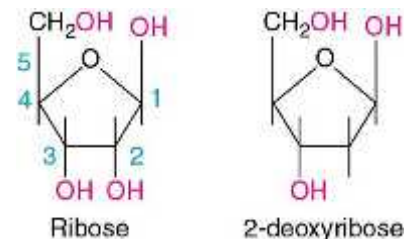
🌈 Nucleotide Structure 📄

Nucleic acids (DNA and RNA) are assembled from nucleotides, which consist of three components: a nitrogenous base, a five-carbon sugar (pentose), and phosphate.

- **Five-Carbon Sugars** 📄

Nucleic acids are classified according to the pentose they contain.

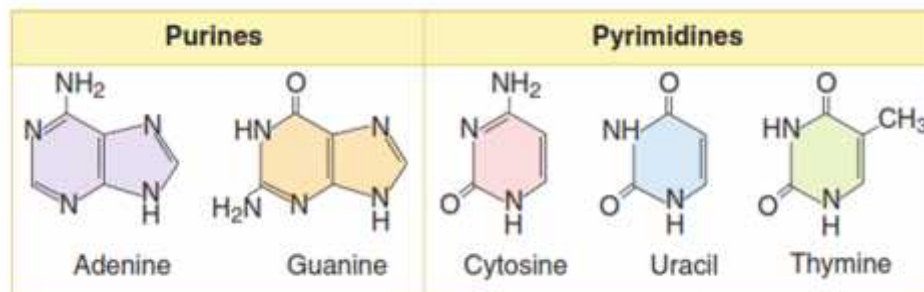
If the pentose is ribose, the nucleic acid is RNA (ribonucleic acid); if the pentose is deoxyribose, the nucleic acid is DNA (deoxyribonucleic acid).



- **Nitrogenous Bases**

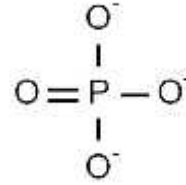
There are two types of nitrogen-containing bases commonly found in nucleotides:

1. **Purines** contain two rings in their structure. The two purines commonly found in nucleic acids are **Adenine (A)** and **Guanine (G)**; both are found in DNA and RNA.
2. **Pyrimidines** have only one ring. **Cytosine (C)** is present in both DNA and RNA. **Thymine (T)** is usually found only in DNA, whereas **Uracil (U)** is found only in RNA.



- **Phosphate group**

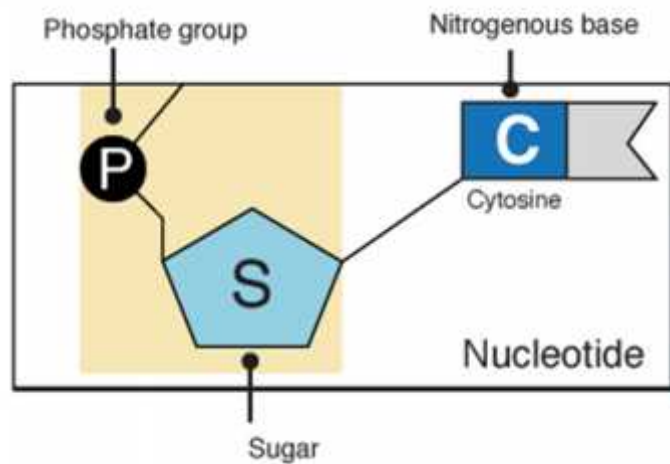
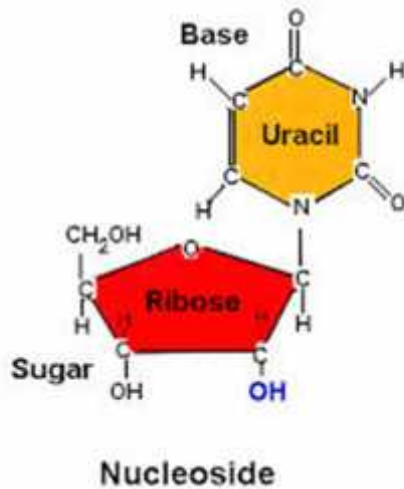
Phosphorus atom surrounded by four oxygen atoms



- **Nucleosides and Nucleotides**

Nucleosides are formed by covalently linking a base to the number 1 (1) carbon of a sugar. The numbers identifying the carbons of the sugar are labeled with “primes” in nucleosides and nucleotides to distinguish them from the carbons of the purine or pyrimidine base.

Nucleotides are formed when one or more phosphate groups is attached to the 5 Carbon of a nucleoside



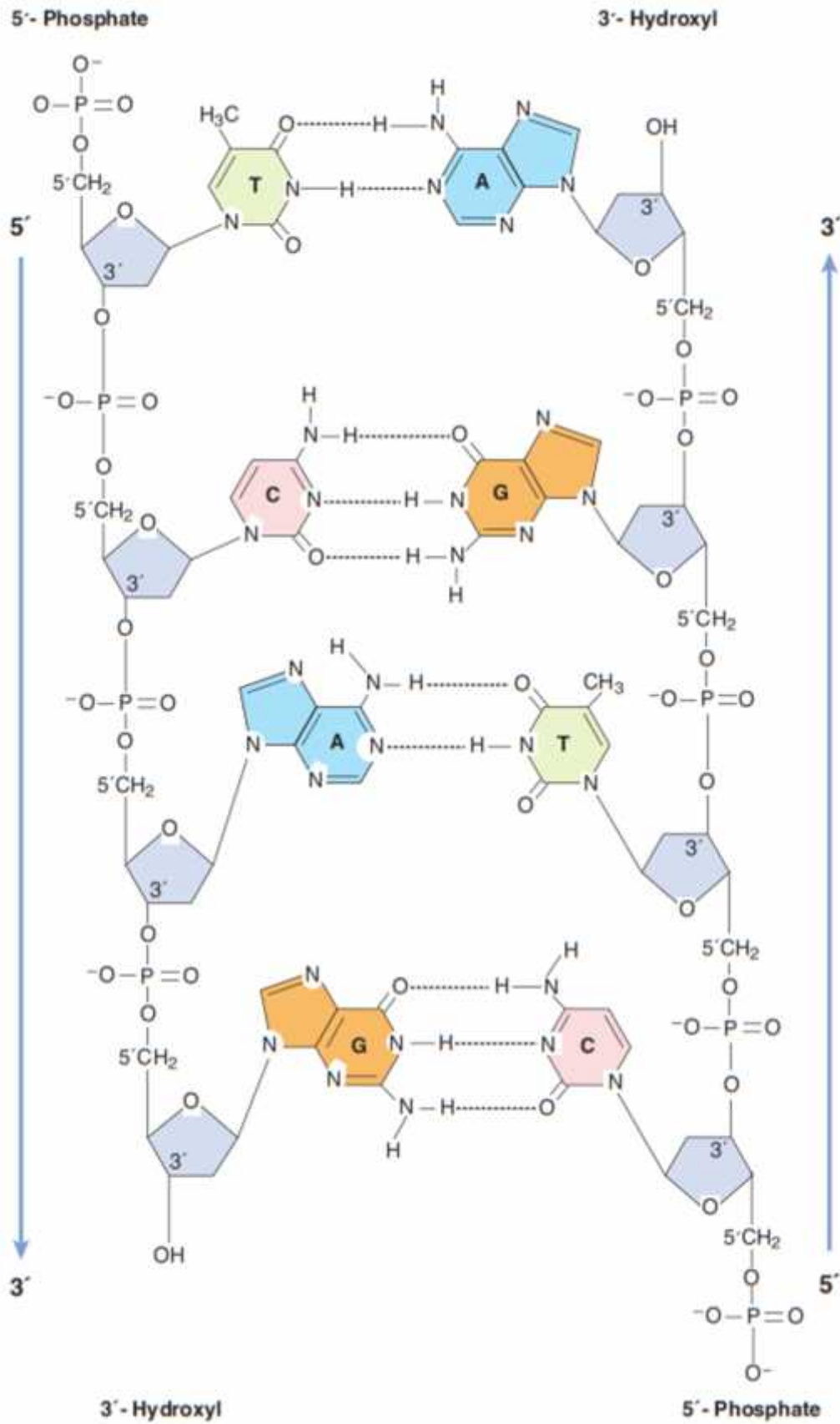
- **Nucleic Acids**

Nucleic acids are polymers of nucleotides joined by 3, 5-phosphodiester bonds; that is, a phosphate group links the 3 carbon of a sugar to the 5 carbon of the next sugar in the chain. Each strand has a distinct 5 end and 3 end, and thus has polarity. A phosphate group is often found at the 5 end, and a hydroxyl group is often found at the 3 end.

The base sequence of a nucleic acid strand is written by convention, in the 5 → 3 direction (left to right). According to this convention, the sequence of the strand on the left in the below Figure must be written

5-TCAG-3 or TCAG

In eukaryotes, DNA is generally double-stranded (dsDNA) and RNA is generally single-stranded (ssRNA).



DNA Structure

The figure here shows an example of a double-stranded DNA (dsDNA) molecule. Some of the features of double-stranded DNA include:

- The two strands are antiparallel (opposite in direction).
- The two strands are complementary. **A** always pairs with **T** (two hydrogen bonds), and **G** always pairs with **C** (three hydrogen bonds). Thus, the base sequence on one strand defines the base sequence on the other strand.
- Because of the specific base pairing, the amount of **A** equals the amount of **T** and the amount of **G** equals the amount of **C**. Thus, total purines equals total pyrimidines. These properties are known as **Chargaff's rules**.

Most DNA occurs in nature as a right-handed double-helical molecule known as Watson-Crick DNA or B-DNA. The hydrophilic sugar-phosphate backbone of each strand is on the outside of the double helix. The hydrogen bonded base pairs are stacked in the center of the molecule.

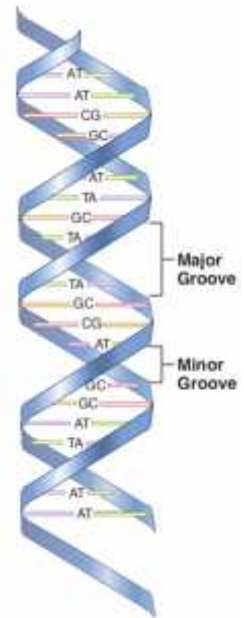
There are about 10 base pairs per complete turn of the helix.

Note: Using Chargaff's Rules, in dsDNA; % A = % T (% U) and % G = % C

Example; a sample of DNA has 10% G; what is the % T?

10% G + 10% C = 20% therefore, % A + % T must total 80%; 40% A and 40% T

Answer: 40% T



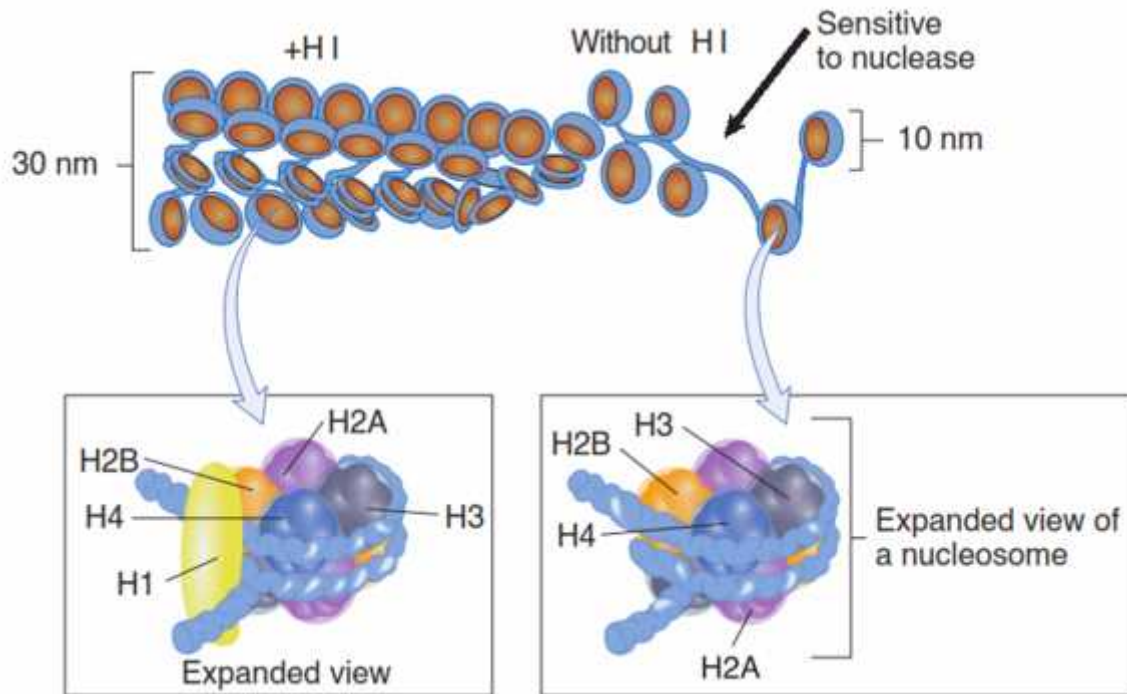
Organization of DNA

Large DNA molecules (about 2 meters length) must be packaged in such a way that they can fit inside the nucleus (about 6 μm) and still be functional.

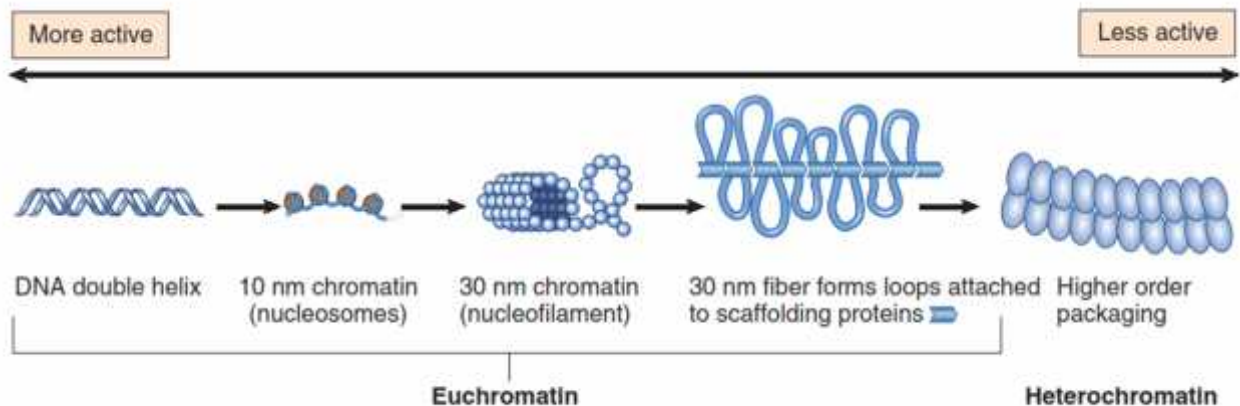
Nucleosomes and Chromatin

Nuclear DNA in eukaryotes is found in chromatin associated with histones and nonhistone proteins. The basic packaging unit of chromatin is the **nucleosome**:

- Histones are rich in lysine and arginine, which confer a positive charge on the proteins.
- Two copies each of histones H2A, H2B, H3, and H4 aggregate to form the **histone octamer**.
- DNA is wound around the outside of this octamer to form a **nucleosome** (a series of nucleosomes is sometimes called “beads on a string”, but is more properly referred to as a **10nm chromatin fiber**).
- Histone H1 is associated with the linker DNA found between nucleosomes to help package them into a solenoid-like structure, which is a **thick 30-nm fiber**.
- Further condensation occurs to eventually form the chromosome. Each eukaryotic chromosome in G₀ or G₁ contains one linear molecule of double-stranded DNA.



Cells in interphase contain two types of chromatin: euchromatin (more opened and available for gene expression) and heterochromatin (much more highly condensed and associated with areas of the chromosomes that are not expressed.)



Euchromatin generally corresponds to the nucleosomes (10-nm fibers) loosely associated with each other (looped 30-nm fibers).

Heterochromatin is more highly condensed, producing interphase heterochromatin as well as chromatin characteristic of mitotic chromosomes.

During mitosis, all the DNA is highly condensed to allow separation of the sister chromatids. This is the only time in the cell cycle when the chromosome structure is visible.

DNA Replication and Repair

🔗 Overview of DNA Replication

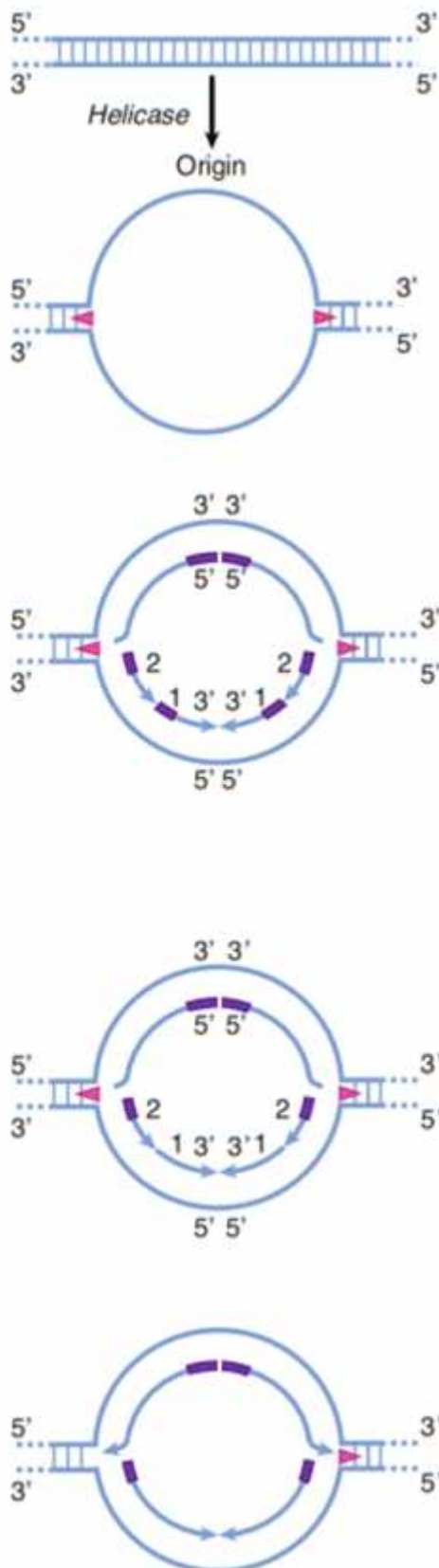
Genetic information is transmitted from parent to progeny by replication of parental DNA, a process in which two daughter DNA molecules are produced that are each identical to the parental DNA molecule. During DNA replication, the two complementary strands of parental DNA are pulled apart. Each of these parental strands is then used as a template for the synthesis of a new complementary strand (semiconservative replication).

During cell division, each daughter cell receives one of the two identical DNA molecules.

🔗 Steps of DNA Replication

The sequence of events is as follows:

1. The base sequence at the origin of replication is recognized.
2. Helicase breaks the hydrogen bonds holding the base pairs together. This allows the two parental strands of DNA to begin unwinding and forms two replication forks.
3. Single-stranded DNA binding protein (SSB) binds to the single-stranded portion of each DNA strand, preventing them from reassociating and protecting them from degradation by nucleases.
4. Primase synthesizes a short (10 nucleotides) RNA primer in the 5' → 3' direction, beginning at the origin on each parental strand. The parental strand is used as a template for this process.
5. DNA polymerase III begins synthesizing DNA in the 5' → 3' direction, beginning at the 3' end of each RNA primer. The newly synthesized strand is complementary and antiparallel to the parental strand used as a template. This strand can be made continuously in one long piece and is known as the “leading strand.”
 - The “lagging strand” is synthesized discontinuously as a series of small fragments (about 1,000 nucleotides long) known as Okazaki fragments. Each Okazaki fragment is initiated by the synthesis of an RNA primer by primase, and then completed by the synthesis of DNA using DNA polymerase III. Each fragment is made in the 5' → 3' direction.
 - There is a leading and a lagging strand for each of the two replication forks on the chromosome.
6. RNA primers are removed by RNAase H and an uncharacterized DNA polymerase fills in the gap with DNA.
7. DNA polymerases have the ability to “proofread” their work by means of a 3' → 5' exonuclease activity. If DNA polymerase makes a mistake during DNA synthesis, the resulting unpaired base at the 3' end of the growing strand is removed before synthesis continues.
8. DNA ligase seals the “nicks” between Okazaki fragments, converting them to a continuous strand of DNA.
9. DNA gyrase (DNA topoisomerase II) provides a “swivel” in front of each replication fork. As helicase unwinds the DNA at the replication forks, the DNA ahead of it becomes overwound and positive supercoils form. DNA gyrase inserts negative supercoils by nicking both strands of DNA, passing the DNA strands through the nick, and then resealing both strands.



Leading Strand Synthesis (Continuous)

1. *Primase* synthesizes the primer (—) 5' to 3'.
2. *DNA polymerases* α and δ extend the primer, moving **into** the replication fork (Leading strand synthesis).
3. *Helicase* (◀) continues to unwind the DNA.

Lagging Strand Synthesis (Discontinuous)

1. *Primase* synthesizes the primer (—) 5' to 3'.
2. *DNA polymerases* α and δ extend the primer, moving **away from** the replication fork (Lagging strand synthesis)
3. Synthesis stops when *DNA polymerase* encounters the primer of the leading strand on the other side of the diagram (not shown), or the primer of the previous (Okasaki) fragment.
4. As *helicase* opens more of the replication fork, a third Okasaki fragment will be added.

RNase H (5' exonuclease activity) digests the RNA primer from fragment 1. In the eukaryotic cell, *DNA polymerase* extends the next fragment (2), to fill in the gap.

In prokaryotic cells *DNA polymerase 1* has both the 5' exonuclease activity to remove primers, and the *DNA polymerase* activity to extend the next fragment (2) to fill in the gap.

In both types of cells *DNA ligase* connects fragments 1 and 2 by making a phosphodiester bond.

This whole process repeats to remove all RNA primers from both the leading and lagging strands.

Steps and Proteins Involved in DNA Replication

	Step in Replication	Protein involved
1	Unwinding of DNA double helix	Helicase
2	Stabilization of unwound template strands	Single-stranded DNA-binding protein (SSB)
3	Synthesis of RNA primers	Primase
4	Synthesis of DNA Leading strand Lagging strand (Okazaki fragments)	DNA polymerase DNA polymerase
5	Removal of RNA primers	RNase H (5' 3' exonuclease)
6	Replacement of RNA with DNA	DNA polymerase
7	Joining of Okazaki fragments	DNA ligase
8	Removal of positive supercoils ahead of advancing replication forks	DNA topoisomerase II (DNA gyrase)
9	Synthesis of telomeres	Telomerase

Telomerase

Telomeres are repetitive sequences at the ends of linear DNA molecules in eukaryotic chromosomes. With each round of replication in most normal cells, the telomeres are shortened because DNA polymerase cannot complete synthesis of the 5' end of each strand. This contributes to the aging of cells, because eventually the telomeres become so short that the chromosomes cannot function properly and the cells die.

Telomerase is an enzyme in eukaryotes used to maintain the telomeres. It contains a short RNA template complementary to the DNA telomere sequence, as well as telomerase reverse transcriptase activity (hTERT). Telomerase is thus able to replace telomere sequences that would otherwise be lost during replication. Normally telomerase activity is present only in embryonic cells, germ (reproductive) cells, and stem cells, but not in somatic cells.

Cancer cells often have relatively high levels of telomerase, preventing the telomeres from becoming shortened and contributing to the immortality of malignant cells.

Reverse Transcriptase

Reverse transcriptase is an RNA-dependent DNA polymerase that requires an RNA template to direct the synthesis of new DNA. Retroviruses, most notably HIV, use this enzyme to replicate their RNA genomes. DNA synthesis by reverse transcriptase in retroviruses can be inhibited by AZT.

DNA Repair

The structure of DNA can be damaged in a number of ways through exposure to chemicals or radiation. Incorrect bases can also be incorporated during replication. Multiple repair systems have evolved, allowing cells to maintain the sequence stability of their genomes. If cells are allowed to replicate their DNA using a damaged template, there is a high risk of introducing stable mutations into the new DNA. Thus any defect in DNA repair carries an increased risk of cancer. Most DNA repair occurs in the G₁ phase of the eukaryotic cell cycle. Mismatch repair occurs in the G₂ phase to correct replication errors.

Damage	Cause	Repair Enzymes
Thymine dimers (G ₁)	UV radiation	DNA polymerase DNA ligase
Mismatched base (G ₂)	DNA replication errors	DNA polymerase DNA ligase
Cytosine deamination (G ₁)	Spontaneous/ heat	DNA polymerase DNA ligase

Diseases Associated With DNA Repair

- Inherited mutations that result in defective DNA repair mechanisms are associated with a predisposition to the development of cancer.
- Xeroderma pigmentosum is an autosomal recessive disorder, characterized by extreme sensitivity to sunlight, skin freckling and ulcerations, and skin cancer. The most common deficiency occurs in the excinuclease enzyme.
- Hereditary nonpolyposis colorectal cancer results from a deficiency in the ability to repair mismatched base pairs in DNA that are accidentally introduced during replication.

Tumor Suppressor Genes and DNA Repair

DNA repair may not occur properly when certain tumor suppressor genes have been inactivated through mutation or deletion:

- The *p53* gene encodes a protein that prevents a cell with damaged DNA from entering the S phase. Inactivation or deletion associated with Li Fraumeni syndrome and many solid tumors.
- *ATM* gene encodes a kinase essential for p53 activity. *ATM* is inactivated in ataxia telangiectasia, characterized by hypersensitivity to x-rays and predisposition to lymphomas.
- *BRCA-1* (breast, prostate, and ovarian cancer) and *BRCA-2* (breast cancer).
- *Rb* The retinoblastoma gene was the first tumor suppressor gene cloned, and is a negative regulator of the cell cycle through its ability to bind the transcription factor E2F and repress transcription of genes required for S phase.

Transcription and RNA Processing

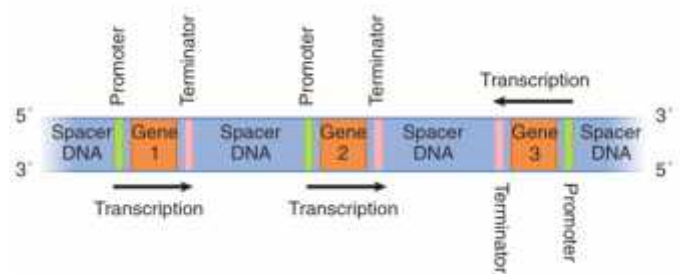
Overview of Transcription

The first stage in the expression of genetic information is transcription of the information in the base sequence of a double-stranded DNA molecule to form the base sequence of a single-stranded molecule of RNA.

For any particular gene, only one strand of the DNA molecule, called the **template strand**, is copied by RNA polymerase as it synthesizes RNA in the 5' to 3' direction. Because RNA polymerase moves in the 3' to 5' direction along the template strand of DNA, the RNA product is antiparallel and complementary to the template.

RNA polymerase recognizes start signals (**promoters**) and stop signals (**terminators**) for each of the thousands of transcription units in the genome of an organism.

The following figure illustrates the arrangement and direction of transcription for several genes on a DNA molecule.



Note: **Gene** is a sequence of DNA that encodes a specific protein.

Types of RNA

RNA molecules play a variety of roles in the cell. The major types of RNA are:

1. **Ribosomal RNA (rRNA)**, which is the most abundant type of RNA in the cell. It is used as a structural component of the ribosome. Ribosomal RNA associates with ribosomal proteins to form the complete, functional ribosome.
2. **Transfer RNA (tRNA)**, which is the second most abundant type of RNA. Its function is to carry amino acids to the ribosome, where they will be linked together during protein synthesis.
3. **Messenger RNA (mRNA)**, which carries the information specifying the amino acid sequence of a protein to the ribosome. Messenger RNA is the only type of RNA that is translated. The mRNA population in a cell is very heterogeneous in size and base sequence, as the cell has essentially a different mRNA molecule for each of the thousands of different proteins made by that cell.
4. **Heterogeneous nuclear RNA (hnRNA or pre-mRNA)**, which is found only in the nucleus of eukaryotic cells. It represents precursors of mRNA, formed during its posttranscriptional processing.
5. **Small nuclear RNA (snRNA)**, which is also only found in the nucleus of eukaryotes. One of its major functions is to participate in splicing (removal of introns) mRNA.
6. **Ribozymes**, which are RNA molecules with enzymatic activity. They are found in both prokaryotes and eukaryotes.

RNA Polymerases

There is a single prokaryotic RNA polymerase that synthesizes all types of RNA in the cell.

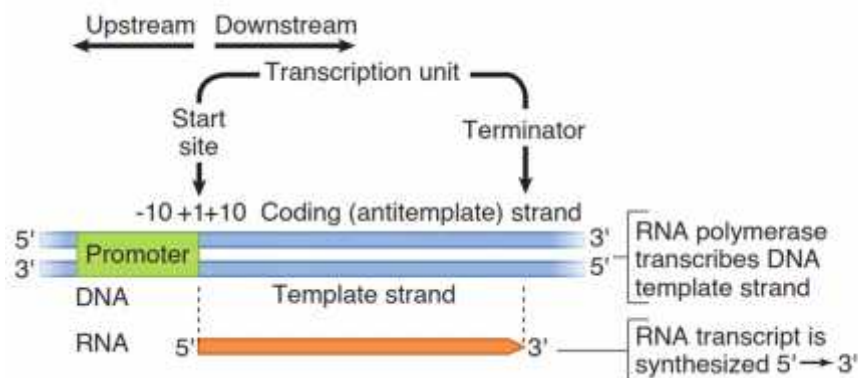
There are three eukaryotic RNA polymerases, distinguished by the particular types of RNA they produce.

1. **RNA polymerase I** is located in the nucleolus and synthesizes most of rRNAs.
2. **RNA polymerase II** is located in the nucleoplasm and synthesizes hnRNA/mRNA and some snRNA.
3. **RNA polymerase III** is located in the nucleoplasm and synthesizes tRNA, some snRNA, and rRNA.

Transcription: Important Concepts and Terminology

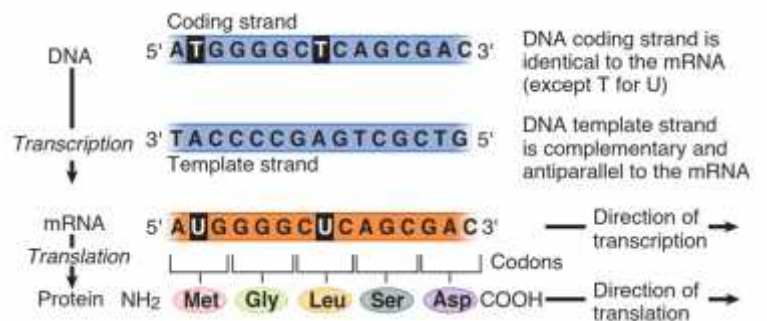
RNA is synthesized by a DNA-dependent RNA polymerase (uses DNA as a template for the synthesis of RNA).

- RNA polymerase locates genes in DNA by searching for promoter regions. The promoter is the binding site for RNA polymerase. Binding establishes where transcription begins, which strand of DNA is used as the template, and in which direction transcription proceeds. No primer is required.
- RNA polymerase moves along the template strand in the 3 to 5 direction as it synthesizes the RNA product in the 5 to 3 direction using NTPs (ATP, GTP, CTP, UTP) as substrates. RNA polymerase does not proofread its work. The RNA product is complementary and antiparallel to the template strand.
- The coding (antitemplate) strand is not used during transcription. It is identical in sequence to the RNA molecule, except that RNA contains uracil instead of the thymine found in DNA.
- By convention, the base sequence of a gene is given from the coding strand (5' → 3').
- In the vicinity of a gene, a numbering system is used to identify the location of important bases. The first base transcribed as RNA is defined as the +1 base of that gene region.
 - To the left (5', or upstream) of this starting point for transcription, bases are -1, -2, -3, etc.
 - To the right (3', or downstream) of this point, bases are +2, +3, etc.
- Transcription ends when RNA polymerase reaches a termination signal.



Flow of Genetic Information from DNA to Protein

For the case of a gene coding for a protein, the relationship among the sequences found in double-stranded DNA, single-stranded mRNA, and protein is illustrated in figure below. Messenger RNA is synthesized in the 5 to 3 direction. It is complementary and antiparallel to the template strand of DNA. The ribosome translates the mRNA in the 5 to 3 direction, as it synthesizes the protein from the amino to the carboxyl terminus.



Sample Questions

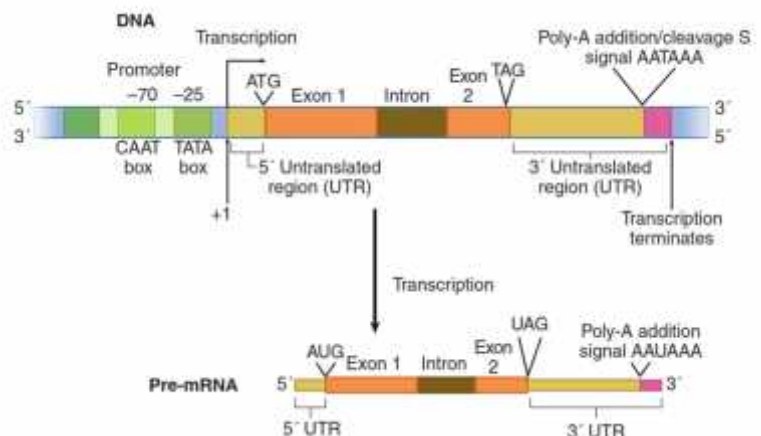
During RNA synthesis, the DNA template sequence TAGC would be transcribed to produce which sequences?

The answer is **GCUA**; RNA is antiparallel and complementary to the template strand. Also remember that, by convention, all base sequences are written in the 5 to 3 direction regardless of the direction in which the sequence may actually be used in the cell.

Production of Eukaryotic Messenger RNA

In eukaryotes, most genes are composed of coding segments (exons) interrupted by noncoding segments (introns). Both exons and introns are transcribed in the nucleus. Introns are removed during processing of the RNA molecule in the nucleus. The mature mRNA is translated in the cytoplasm. The structure and transcription of a typical eukaryotic gene coding for a protein is illustrated in figure below. Transcription of this gene occurs as follows:

1. With the help of proteins called transcription factors, RNA polymerase II recognizes and binds to the promoter region. The basal promoter region of eukaryotic genes usually has two consensus sequences called the TATA box (also called Hogness box) and the CAAT box.
2. RNA polymerase II separates the strands of the DNA over a short region to initiate transcription and read the DNA sequence. The template strand is read in the 3 to 5 direction as the RNA product (the primary transcript) is synthesized in the 5 to 3 direction. Both exons and introns are transcribed.
3. RNA polymerase II ends transcription when it reaches a termination signal. These signals are not well understood in eukaryotes.



Processing of Eukaryotic Pre-Messenger RNA

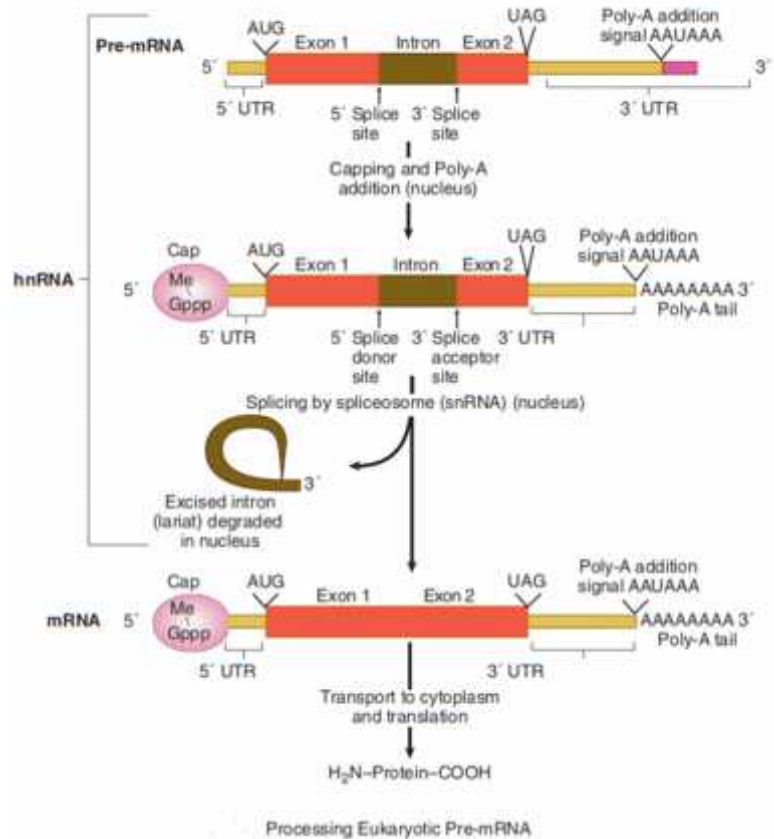
The primary transcript must undergo extensive posttranscriptional processing inside the nucleus to form the mature mRNA molecule. These processing steps include the following:

1. A 7-methylguanosine cap is added to the 5' end while the RNA molecule is still being synthesized. The cap structure serves as a ribosome-binding site and also helps to protect the mRNA chain from degradation.
2. A poly-A tail is attached to the 3' end. In this process, an endonuclease cuts the molecule on the 3' side of the sequence AAUAAA (poly-A addition signal), then poly-A polymerase adds the poly-A tail (about 200 As) to the new 3' end. The poly-A tail protects the message against rapid degradation and aids in its transport to the cytoplasm. A few mRNAs (for example, histone mRNAs) have no poly-A tails.
3. Introns are removed from hnRNA by splicing, accomplished by spliceosomes, which are complexes of snRNA and protein. The hnRNA molecule is cut at splice sites at the 5' (donor) and 3' (acceptor) ends of the intron. The intron is excised in the form of a lariat structure and degraded. Neighboring exons are joined together to assemble the coding region of the mature mRNA.
4. All of the intermediates in this processing pathway are collectively known as hnRNA.
5. The mature mRNA molecule is transported to the cytoplasm, where it is translated to form a protein.

Note: Mutations in splice sites can lead to abnormal proteins. For example, mutations that interfere with proper splicing of β -globin mRNA are responsible for some cases of β -thalassemia.

Transfer RNA (tRNA) Carries Activated Amino Acids for Translation

There are many different specific tRNAs. Each tRNA carries only one type of activated amino acid for making proteins during translation. The genes encoding these tRNAs in eukaryotic cells are transcribed by RNA polymerase III. The tRNAs enter the cytoplasm where they combine with their appropriate amino acids.



Summary of Important Points About Transcription and RNA Processing

	Prokaryotic	Eukaryotic
Gene regions	May be polycistronic Genes are continuous coding regions Very little spacer (noncoding) DNA between genes	Always monocistronic Genes have exons and introns Large spacer (noncoding) DNA between genes
RNA polymerase	Core enzyme: $\alpha_2\beta\beta'$	RNA polymerase I: rRNA RNA polymerase II: mRNA; snRNA RNA polymerase III: tRNA, 5S RNA
Initiation of transcription	Promoter (-10) TATAAT and (-35) sequence Sigma initiation subunit required to recognize promoter	Promoter (-25) TATA and (-70) CAAT Transcription factors (TFIID) bind promoter
mRNA synthesis	Template read 3' to 5'; mRNA synthesized 5' to 3'; gene sequence specified from coding strand 5' to 3'; transcription begins at +1 base	
Termination of transcription	Stem and loop + UUUUU Stem and loop + rho factor	Not well characterized
Relationship of RNA transcript to DNA	RNA is antiparallel and complementary to DNA template strand; RNA is identical (except U substitutes for T) to DNA coding strand	
Posttranscriptional processing of hnRNA (pre-mRNA)	None	In nucleus: 5' cap (7-MeG) 3' tail (poly-A sequence) Removal of introns from pre-RNA <ul style="list-style-type: none"> • Alternative splicing yields variants of protein product
Ribosomes	70S (30S and 50S) rRNA and protein	80S (40S and 60S) rRNA and protein
tRNA	Cloverleaf secondary structure <ul style="list-style-type: none"> • Acceptor arm (CCA) carries amino acid • Anticodon arm; anticodon complementary and antiparallel to codon in mRNA 	

The Genetic Code, Mutation, and Translation

Overview of Translation

The second stage in gene expression is translating the nucleotide sequence of a messenger RNA molecule into the amino acid sequence of a protein.

The genetic code is defined as the relationship between the sequence of nucleotides in DNA (or its RNA transcripts) and the sequence of amino acids in a protein.

Each amino acid is specified by one or more nucleotide triplets (codons) in the DNA.

The Genetic Code

Most genetic code tables designate the codons for amino acids as mRNA sequences.

Important features of the genetic code include:

- Each codon consists of three bases (triplet). There are 64 codons. They are all written in the 5' to 3' direction.
- 61 codons code for amino acids. The other three (**UAA**, **UGA**, **UAG**) are stop codons (or nonsense codons) that terminate translation.
- There is one start codon (initiation codon), **AUG**, coding for methionine. Protein synthesis begins with methionine (Met) in eukaryotes.
- The code is unambiguous. Each codon specifies no more than one amino acid.
- The code is degenerate. More than one codon can specify a single amino acid. All amino acids, except Met and tryptophan (Trp), have more than one codon.
- For those amino acids having more than one codon, the first two bases in the codon are usually the same. The base in the third position often varies.
- The code is universal (the same in all organisms). Some minor exceptions to this occur in mitochondria.
- The code is commaless (contiguous). There are no spacers or “commas” between codons on an mRNA.
- Neighboring codons on a message are nonoverlapping.

First Position (5' End)	Second Position				Third Position (3' End)
	U	C	A	G	
U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA } Stop UAG }	UGU } Cys UGC } UGA } Stop UGG } Trp	U C A G
C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
A	AUU } AUC } Ile AUA } AUG } Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

🌈 Mutations

A mutation is any permanent, heritable change in the DNA base sequence of an organism.

This altered DNA sequence can be reflected by changes in the base sequence of mRNA, and, sometimes, by changes in the amino acid sequence of a protein.

Mutations can cause genetic diseases. They can also cause changes in enzyme activity, nutritional requirements, antibiotic susceptibility, morphology, antigenicity, and many other properties of cells.

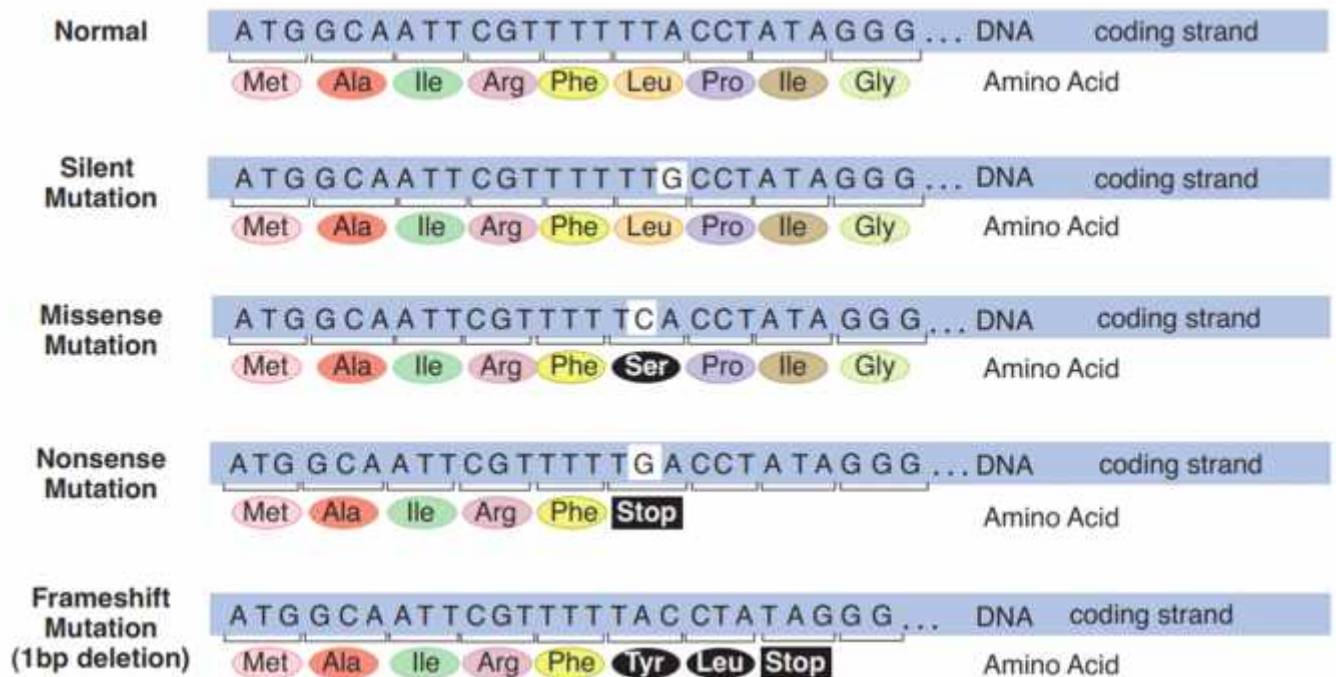
A very common type of mutation is a single base alteration or point mutation.

- A **transition** is a point mutation that replaces a purine-pyrimidine base pair with a different purine-pyrimidine base pair. For example, an A-T base pair becomes a G-C base pair.
- A **transversion** is a point mutation that replaces a purine-pyrimidine base pair with a pyrimidine-purine base pair. For example, an A-T base pair becomes a T-A or a C-G base pair.

Mutations are often classified according to the effect they have on the structure of the gene's protein product. This change in protein structure can be predicted using the genetic code table in conjunction with the base sequence of DNA or mRNA.

Type of Mutation	Effect on Protein
Silent: new codon specifies same amino acids	None
Missense: new codon specifies different A.A	Possible decrease in function; variable effects
Nonsense: new codon is stop codon	Shorter than normal; usually nonfunctional
Frameshift/in-frame: addition or deletion of base(s)	Usually nonfunctional; often shorter than normal
Large segment deletion (unequal crossover in meiosis)	Loss of function; shorter than normal or entirely missing
5' splice site (donor) or 3' splice site (acceptor)	Variable effects ranging from addition or deletion of a few amino acids to deletion of an entire exon
Trinucleotide repeat expansion	Expansions in coding regions cause protein product to be longer than normal and unstable. Disease often shows anticipation in pedigree

Some common types of mutation in DNA



❖ Large Segment Deletions

Large segments of DNA can be deleted from a chromosome during an unequal crossover in meiosis. Crossover or recombination between homologous chromosomes is a normal part of meiosis I that generates genetic diversity in reproductive cells (egg and sperm), a largely beneficial result.

-thalassemia is a well-known example of a genetic disease in which unequal crossover has deleted one or more α -globin genes from chromosome 16.

Cri-du-chat (mental retardation, microcephaly, wide-set eyes, and a characteristic kitten like cry) results from a terminal deletion of the short arm of chromosome 5.

❖ Mutations in Splice Sites

Mutations in splice sites affect the accuracy of intron removal from hnRNA during posttranscriptional processing. If a splice site is lost through mutation, spliceosomes may:

- Delete nucleotides from the adjacent exon.
- Leave nucleotides of the intron in the processed mRNA.
- Use the next normal upstream or downstream splice site, deleting an exon from the processed mRNA.

Mutations in splice sites have now been documented in many different diseases, including α -thalassemia, Gaucher disease, and Tay-Sachs.

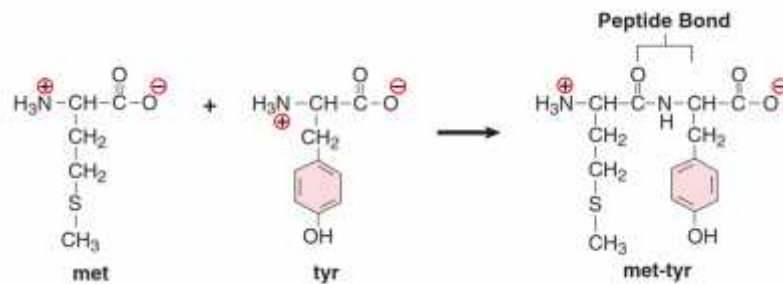
❖ Trinucleotide Repeat Expansion

The mutant alleles in certain diseases, such as Huntington disease, fragile X syndrome, and myotonic dystrophy, differ from their normal counterparts only in the number of tandem copies of a trinucleotide.

In these diseases, the number of repeats often increases with successive generations and correlates with increasing severity and decreasing age of onset, a phenomenon called anticipation. For example, in the normal Huntington allele, there are five tandem repeats of CAG in the coding region. Affected family members may have 30 to 60 of these CAG repeats.

🌈 Translation (Protein Synthesis)

Protein synthesis occurs by peptide bond formation between successive amino acids whose order is specified by a gene and thus by an mRNA.

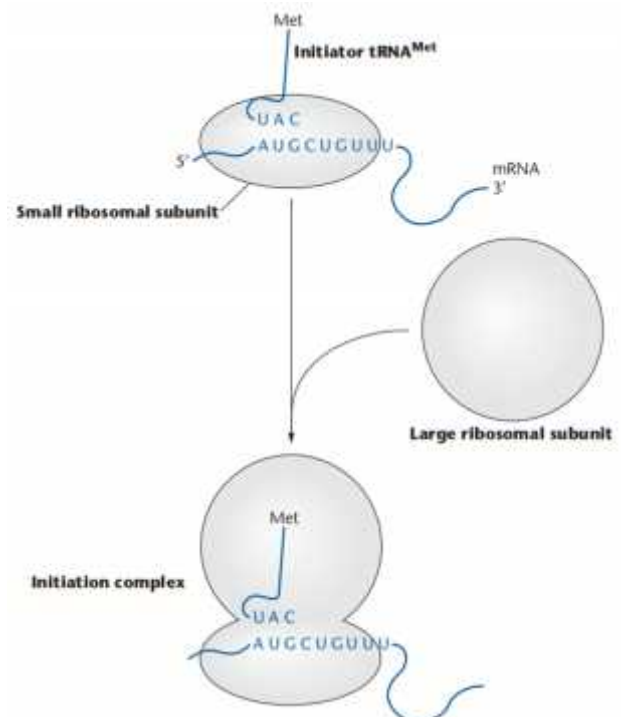


The process of protein synthesis occurs in 3 stages: initiation, elongation, and termination. Special protein factors for initiation (IF), elongation (EF), and termination (release factors), as well as GTP, are required for each of these stages.

✓ Initiation

In eukaryotic organisms, translation is **initiated** by the binding of a specific charged initiator tRNA, *Met-tRNA*, and other factors to the small ribosomal subunit. *No other charged tRNA can bind to a free small ribosomal subunit.* **Next**, the 5' end of an mRNA combines with the initiator tRNA–small ribosomal subunit complex, and the complex migrates along the mRNA until an AUG sequence (**initiator codon**) is encountered.

Then, the UAC anticodon sequence of the initiator *Met-tRNA* base pairs with the AUG sequence of the mRNA, the migration stops, and the larger ribosomal subunit joins the complex.



✓ Elongation

The elongation process entails linking together (by peptide bond formation) the correct amino acid sequence encoded in the mRNA.

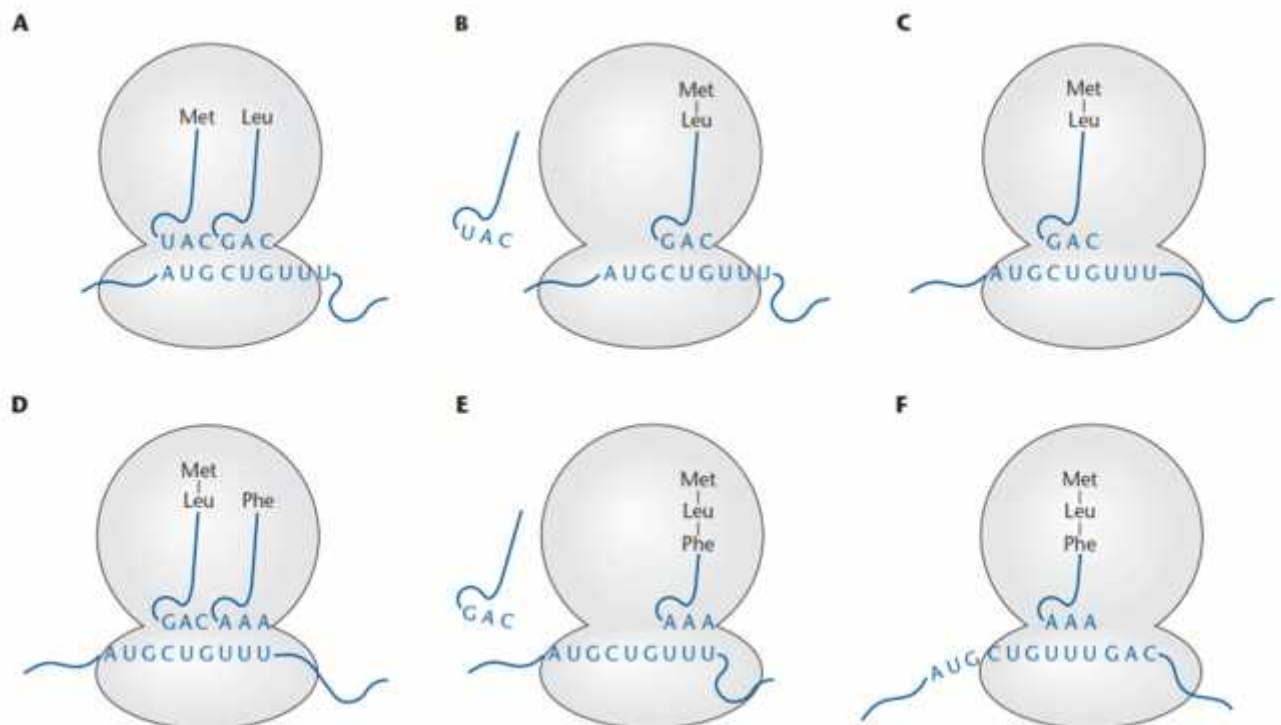
The second set of three nucleotides (triplet, codon) in the mRNA that immediately follows the AUG codon dictates the anticodon sequence and, therefore, which charged tRNA will bind to the ribosome complex. Uncharged tRNAs do not bind efficiently to ribosomes.

If the second nucleotide triplet in the mRNA is CUG (for example), then the charged tRNA with the anticodon sequence GAC will bind. This charged tRNA carries the amino acid leucine. Once in place, a peptide bond is formed between the carboxyl group of the methionine and the amino group of the leucine.

The leucine remains bound to its tRNA. Peptide bond formation is catalyzed by enzymatic activity associated with the large ribosomal subunit.

The formation of the peptide bond “uncharges” the initiator tRNA, because the bond between the carboxyl group of methionine and its tRNA is cleaved to make the carboxyl group available for peptide bond formation.

The uncharged tRNA is ejected from the ribosomal complex. The methionine-leucine-tRNA-mRNA combination shifts (translocates) along the ribosome, and, as a consequence, the next codon of the mRNA is available for binding by a charged tRNA with the appropriate anticodon sequence.

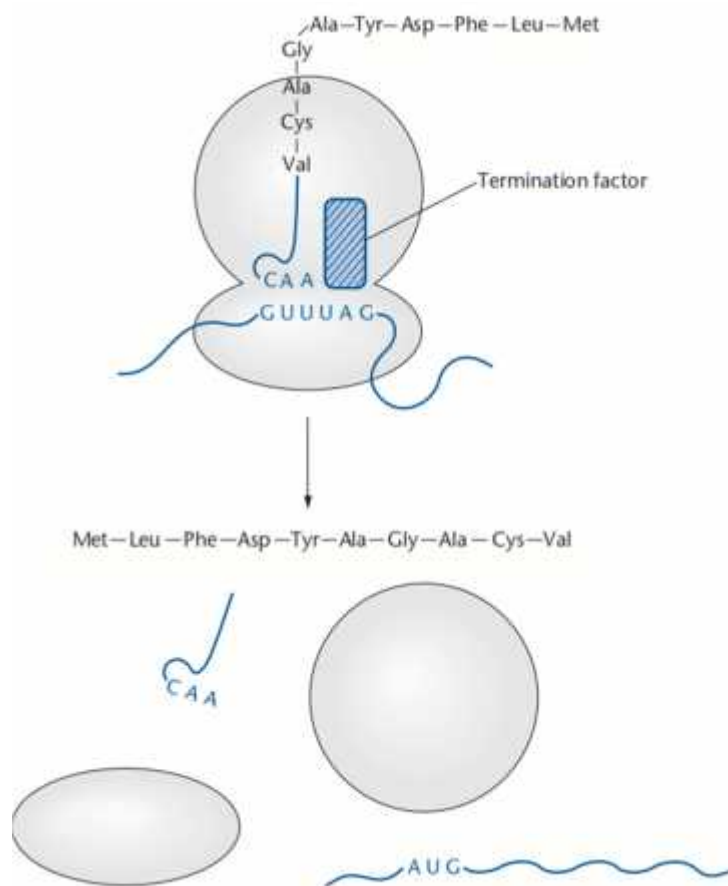


✓ Termination

The elongation process continues until a UAA, UAG, or UGA codon is encountered.

There are no naturally occurring tRNAs with anticodons that are complementary to UAA, UAG, or UGA (stop codons, termination codons). However, a protein (termination factor, release factor) recognizes a stop codon and binds to the ribosome.

After binding of a termination factor, the bond between the last tRNA, which has the complete chain of amino acids linked to it, and its amino acid is broken. This cleavage results in the release of the uncharged tRNA, the complete protein, and the mRNA. In addition, the ribosomal subunits separate from each other. The components of the translation machinery can be used again.



After translation, a protein may be modified in various ways. In some proteins, the methionine at the N-terminus is cleaved off, leaving the second encoded amino acid as the N-terminal moiety. In other cases, the protein is selectively cleaved at specific sites to make smaller protein chains with discrete functions. In other instances, phosphorus, lipids, carbohydrates, or other chemical groups are added enzymatically to specific amino acids to produce modified proteins that can carry out certain cellular activities.

Genetic Strategies in Therapeutics

Overview of Recombinant DNA Technology

Recombinant DNA technology allows a DNA fragment from any source to be joined in vitro with a nucleic acid vector that can replicate autonomously in microorganisms.

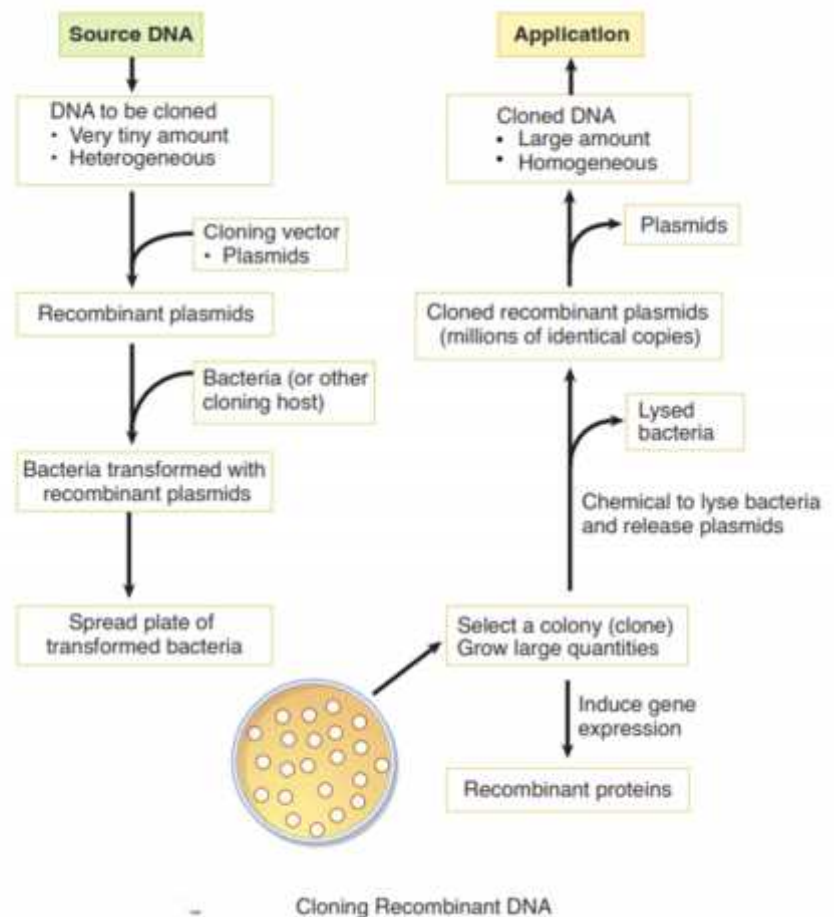
This provides a means of analyzing and altering genes and proteins. It also provides the reagents necessary for genetic testing for carrier detection and prenatal diagnosis of genetic diseases and for gene therapy. Additionally, this technology can provide a source of a specific protein, such as recombinant human insulin, in almost unlimited quantities.

The DNA to be cloned is usually present in a small quantity and is part of a heterogeneous mixture containing other DNA sequences. The goal is to produce a large quantity of homogeneous DNA for one of the above applications.

The general strategy for cloning DNA and isolating the cloned material is shown in the figure

✓ The steps include:

- Ligate the DNA into a piece of nucleic acid (the vector) that can be autonomously replicated in a living organism. The vector containing the new DNA is referred to as a recombinant vector.
- Transfer the recombinant vectors into host cells.
- Grow the host cells in isolated colonies so that each colony contains only one recombinant vector.
- Each cultured colony is a clone; all members are genetically identical.
- Select a colony for study.
- Grow a large quantity of that colony.
- Lyse the host cells and re-isolate the replicated recombinant vectors.
- Remove (by restriction enzyme cutting) the cloned DNA from the vector



The Human Genome Project

The Human Genome Project, initiated in 1991, involved the identification of the entire 3 billion–base-pair human DNA sequence. This project has now been completed.

Although humans appear to be quite different from each other, the sequence of our DNA is, in reality, highly conserved. On average, 2 unrelated individuals share over 99.9% of their DNA sequences. For the Human Genome Project, DNA was obtained from a relatively small number of individuals.

Human Genome Project data can be used to identify:

1. Protein-coding genes
2. Regulatory sequences in noncoding DNA
3. Polymorphic genetic markers dispersed throughout chromosomes

Medical Applications of Recombinant DNA

Recombinant DNA can be used as follows:

1. To produce recombinant proteins, used variously in:
 - a) Replacement therapy (e.g., insulin in diabetes)
 - b) Disease prevention (e.g., vaccines)
 - c) Diagnostic tests (e.g., monoclonal antibodies)
2. To conduct gene therapy in the treatment of genetic diseases.

Examples of Protein Products of Recombinant DNA Technology

Product	Use
Insulin	Diabetes
Growth factor	Growth defects
Epidermal growth factor	Burns, ulcers
Hepatitis B vaccine	Prevention of viral
Erythropoietin	Anemia
Factor VIII	Hemophilia

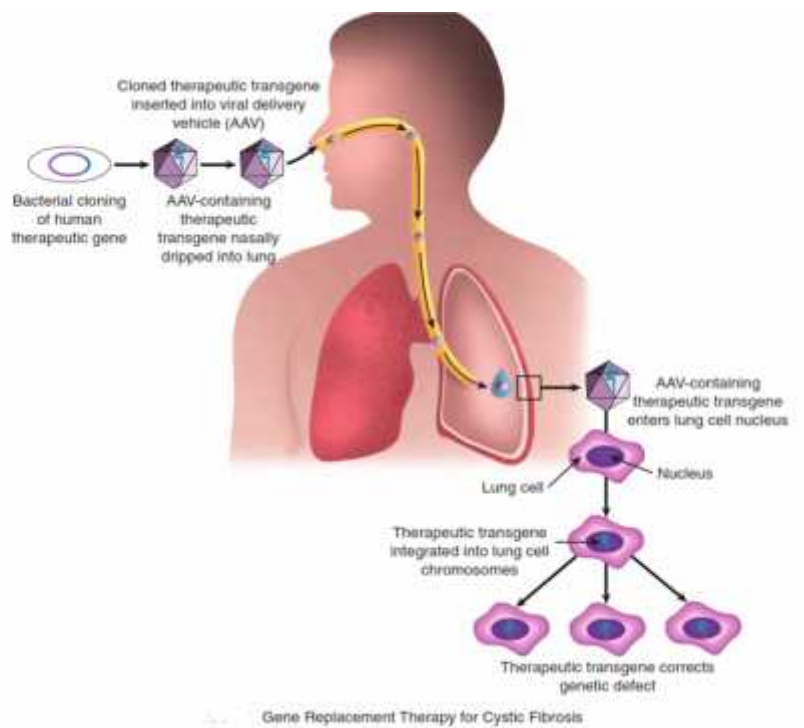
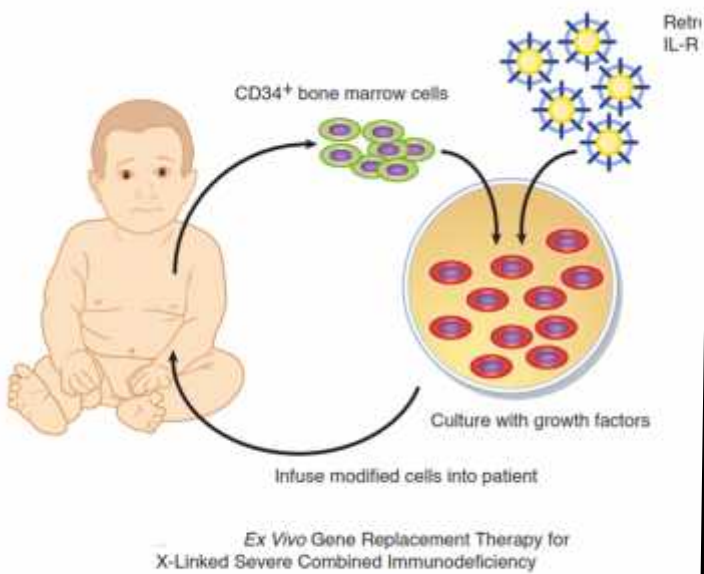
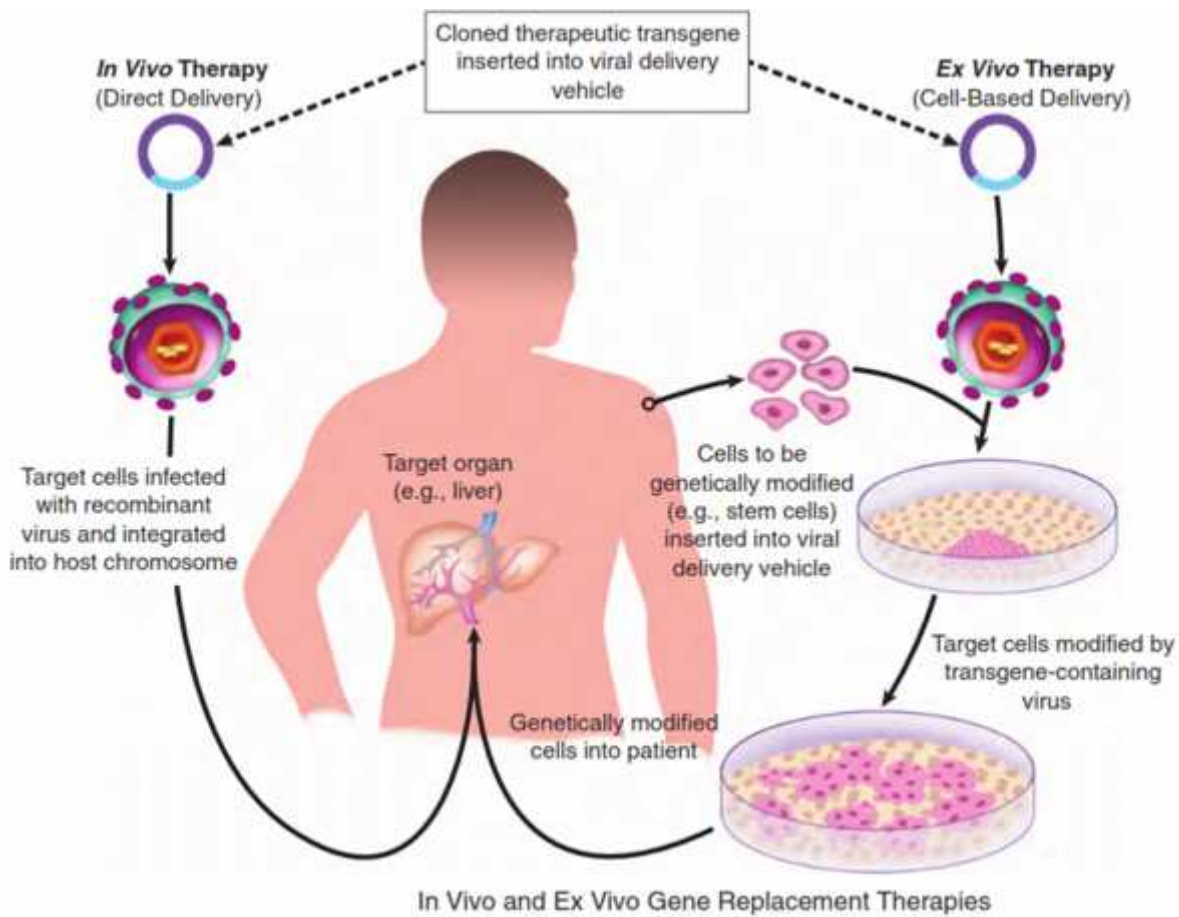
Gene Therapy

Gene therapy now offers potential cures for individuals with inherited diseases. The initial goal is to introduce a normal copy of the gene that is defective into the tissues that give rise to the pathology of the genetic disease.

- Gene transfer requires a delivery vector (retrovirus, adenovirus, and liposome).
- Only tissues giving rise to the disease pathology are targeted for gene therapy.
- The normal gene is not inherited by offspring.

Note

- **Ex Vivo:** Cells modified outside the body, then transplanted back in.
- **In Vivo:** Gene changed in cells still in body.



Techniques of genetic analysis

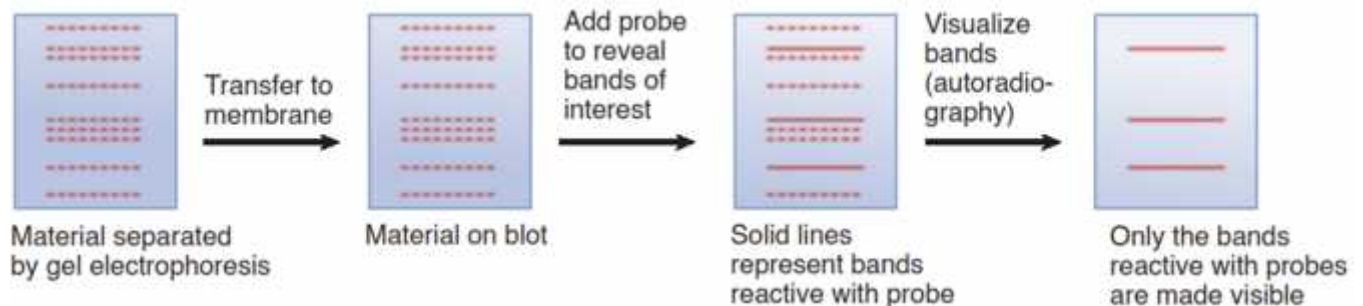
Techniques of genetic analysis are assuming an increasingly larger role in medical diagnosis. These techniques, which once were a specialized part of medical genetics, are now becoming essential tools for every physician to understand.

Blotting techniques allow testing for genetic diseases, gene expression profiling, and routine testing for antigens and antibodies. The polymerase chain reaction (PCR) is now an essential tool in many aspects of genetic testing, forensic medicine, and paternity testing.

Blotting Techniques

Blotting techniques have been developed to detect and visualize specific DNA, RNA, and protein among complex mixtures of contaminating molecules. These techniques have allowed the identification and characterization of the genes involved in numerous inherited diseases.

The general method for performing a blotting technique is illustrated in this figure.



➤ Probes

DNA probes are radioactively labeled single-stranded DNA molecules that are able to specifically hybridize (anneal) to particular denatured DNA sequences.

Examples include:

- Probes that bind to part of a specific gene region. These are often produced by cloning cDNA transcribed from the gene and labeling it with ^{32}P , a radioactive isotope of phosphorus.
- Probes that bind to markers known to be in close proximity (closely linked) to a gene
- Probes that bind specifically to a single allele of a gene—allele-specific oligonucleotide (ASO) probes.

- **Types of Blot Techniques**

Blot type	Material analyzed	Electrophoresis required	Probe used	Purpose
Southern	DNA	Yes	DNA	To determine which restriction fragments of DNA are associated with a particular gene
Northern	RNA	Yes	DNA	To measure sizes and amounts of specific mRNA molecules to answer questions about gene expression
Western	Protein	Yes	Enzyme linked antibody	To measure amount of antigen (proteins) or antibody
Dot (slot)	RNA,DNA, or protein	No	Same as for blots above	To detect specific DNA, RNA, protein, or antibody

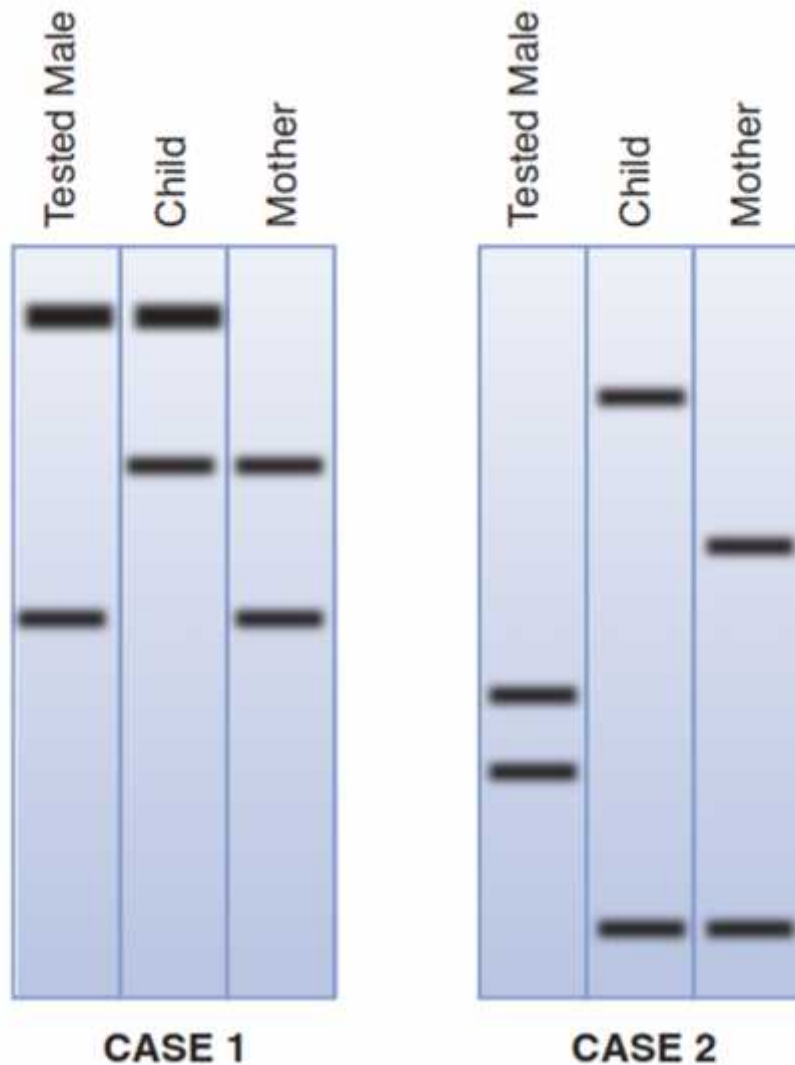
Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) is a technique in which a selected region of a chromosome can be amplified more than a million-fold within a few hours. The technique allows extremely small samples of DNA to be used for further testing.

The PCR has many different applications.

1. Comparing DNA samples in forensic cases
2. Paternity testing
3. Direct mutation testing
4. Diagnosing bacterial and viral infections
5. HIV testing in situations where antibody tests are uninformative (importantly, infants whose mothers are HIV positive)

- **Paternity testing using PCR amplification**



Case 1: The tested male in case 1 may be the father, as he shares a band with the child. We cannot be certain, however, because many other men in the population could have this same band. Matches are required at several different loci to indicate with high probability that a tested male is the father.

Case 2: The tested male in case 2 cannot be the father, as neither of his bands is shared with the child.

✓ In practice, 9 to 10 different polymorphisms are necessary to indicate a match.