UNIT CODE/TITLE:
SBC 222: Nucleic acid and protein synthesis

PREREQUISITE
None

PURPOSE
To introduce learners to the types, structure and basic functions of nucleic acids including the basic techniques applied in their study.

EXPECTED LEARNING OUTCOME
By the end of this course, the learners should be able to;

By the end of this course, learners should be able to;

i) Describe the structure of DNA molecule.
ii) Explain the role of RNA in protein synthesis.
iii) Discuss the techniques applied in the study of DNA and RNA.
iv) Explain Gel electrophoresis technique.

COURSE CONTENT*(Existing)*
DNA as the hereditary material. Primary structure of DNA. Structure of and types of RNA. Transcription. Translation. RNA and protein synthesis. Ribosome structure. Protein synthesis. Nucleic acid isolation and analysis: Gel electrophoresis and nucleotide sequencing, Nucleic acid hybridisation, nucleic acid probes, Application Southern and Northern blotting. DNA replication and repair. Post-translational modifications.. Practical include isolation and characterization of DNA, blotting techniques, polymerase chain reaction.

TEACHING ORGANIZATIONN
Lectures- 35 hours per semester
Group discussion
Assignment

MODE OF DELIVERY
Face to face full time

ASSESSMENT
Continuous assessment test -30%
End of semester examination -70%

REFERENCES
UNIT CODE/ TITLE
SBC 222: Nucleic acid and protein synthesis
PREREQUISITE
None
PURPOSES
To introduce learners to the types, structure and basic functions of nucleic acids including the basic techniques applied in their study
EXPECTED LEARNING OUTCOMES
By the end of this course, learners should be able to;
i) Describe the structure of DNA molecule.
ii) Explain the role of RNA in protein synthesis.
iii) Discuss the techniques applied in the study of DNA and RNA.
iv) Explain Gel electrophoresis technique.
COURSE CONTENT
Central Dogma Theory, DNA as the hereditary material. Primary structure of DNA. Structure of and types of RNA. Transcription. Translation. RNA and protein synthesis. Ribosome structure. Protein synthesis. Nucleic acid isolation and analysis: Gel electrophoresis and nucleotide sequencing, Nucleic acid hybridisation, nucleic acid probes, Application Southern and Northern blotting. DNA replication and repair. Post-translational modifications.. Practical include isolation and characterization of DNA, blotting techniques, polymerase chain reaction.

TEACHING ORGANIZATION
Lecture - 35 hours per semester

Group discussion
Assignment
MODE OF DELIVERY
Face to face full time
ASSESSMENT
Continuous assessment test - 30%
End of semester examination - 70%
REFERENCES
i) H.F., Lodish and J. E., Darnell (1995). Molecular cell biology. Scientific American Books, New York.
ii) ii) L. G., Davis, M. D., Dibner and J. F., Battey (1986). Basic Methods in Molecular Biology. ISBN 978-0-444-01082-7.
iii) J. A. Moore (1972). Heredity and Development, 2nd Edition. Oxfoed University Press, London.
Recommended Textbooks:
i) W. B., Coleman and G. J. Tsongalis Molecular Diagnostics. ISBN-13: 978-1588293565 ISBN-10: 1588293564.
ii) R. Sinden (2012). DNA Structure and Function. 1st Edition. Academic Press. Pp 398. ISBN: 9780126457506
E-Materials
i) https://bio.libretexts.org/Bookshelves/Microbiology/Book%3A_Microbiology_(OpenStax)/10%3A_Biochemistry_of_the_Genome/10.2%3A_Structure_and_Function_of_DNA
ii) https://www.ncbi.nlm.nih.gov/books/NBK21603/
Journals
i) Nucleic acid research Journal. Online ISSN 1362-4962 Print ISSN 0305-1048.
ii) Europen Journal of Biochemistry. ISSN: 0014-2956.

THE CENTRAL DOGMA THEORY

The Central Dogma of Molecular Biology is a fundamental concept that outlines the flow of genetic information within a biological system. Proposed by Francis Crick in 1958, it describes the sequential processes by which genetic information is stored, replicated, and expressed in living organisms. The central dogma has three main components:

1. **Replication:** The first step in the central dogma is DNA replication. During this process, the genetic information encoded in DNA is copied to form an identical DNA molecule. DNA replication occurs before cell division, ensuring that each daughter cell receives a complete and accurate set of genetic instructions. DNA replication is highly accurate and is performed by enzymes called DNA polymerases.

2. **Transcription:** Transcription is the second step in the central dogma. During transcription, a specific segment of DNA, known as a gene, is used as a template to synthesize a complementary RNA molecule. This RNA molecule, called messenger RNA (mRNA), carries the genetic code from the DNA in the cell's nucleus to the ribosomes in the cytoplasm, where protein synthesis will occur. Transcription is catalyzed by an enzyme called RNA polymerase.

3. **Translation:** Translation is the final step in the central dogma. It takes place on ribosomes in the cytoplasm, where the genetic information encoded in the mRNA is used to synthesize proteins. In this process, transfer RNA (tRNA) molecules recognize and bind to specific codons (three-nucleotide sequences) on the mRNA. Each tRNA carries a corresponding amino acid. As the ribosome moves along the mRNA, it facilitates the bonding of amino acids in the correct sequence, forming a polypeptide chain. This chain eventually folds into a functional protein based on the genetic code provided by the mRNA.

Key points about the Central Dogma of Molecular Biology:

- Information flows unidirectionally, from DNA to RNA to protein.
- DNA serves as the permanent repository of genetic information.
- RNA acts as an intermediary molecule that carries genetic instructions from DNA to the protein-synthesizing machinery.
- Proteins are the ultimate functional products of genetic information and play essential roles in all cellular processes.

While the central dogma provides a fundamental framework for understanding the flow of genetic information, it's important to note that there are exceptions and additional layers of complexity in molecular biology. For example, some viruses can reverse transcribe RNA into DNA (reverse transcription), and the field of epigenetics explores modifications to DNA and chromatin structure that can influence gene expression without altering the underlying DNA

sequence. Nonetheless, the Central Dogma remains a foundational concept for understanding the basic processes of genetics and molecular biology.

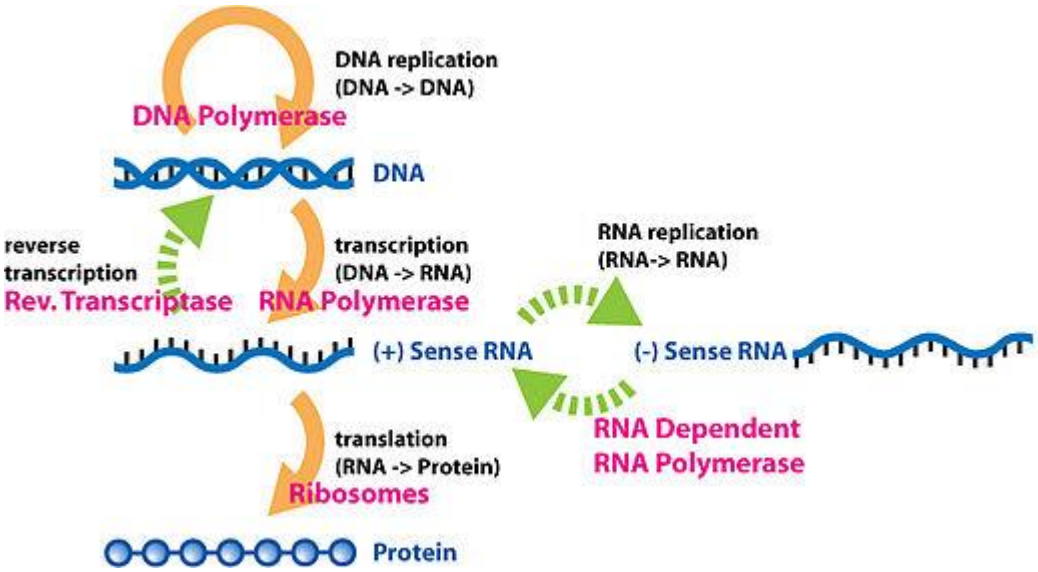


Fig. 1. The Central Dogma Theory highlighted in yellow. Unusual flows of information highlighted in green

DNA AS A HEREDITARY MATERIAL

DNA (deoxyribonucleic acid) is the hereditary material in most living organisms, serving as the molecule that carries genetic information from one generation to the next. The discovery of DNA as the hereditary material is one of the most significant milestones in the history of biology and genetics. Here's an overview of how DNA functions as the hereditary material:

1. Storage of Genetic Information:

- DNA is a long, chain-like molecule composed of four different nucleotide bases: adenine (A), cytosine (C), guanine (G), and thymine (T).
- The sequence of these nucleotide bases in DNA encodes the genetic instructions for building and maintaining an organism.
- Genes, which are specific segments of DNA, contain the information necessary for the synthesis of proteins and the regulation of various cellular processes.

2. Replication:

- One of DNA's crucial functions is to replicate itself accurately during cell division. This ensures that genetic information is faithfully passed from parent cells to daughter cells.
- DNA replication involves the separation of the DNA double helix into two complementary strands and the synthesis of new DNA strands based on the existing ones. Each daughter cell receives a copy of the original DNA.

3. Transmission of Traits:

- DNA carries the genetic code that determines an organism's traits, including its physical characteristics, biochemical properties, and susceptibility to various diseases.
- The inheritance of DNA from parents to offspring is responsible for the passing down of traits from one generation to the next.

4. Genetic Diversity:

- DNA contributes to genetic diversity within a population. Mutations, which are changes in the DNA sequence, can occur naturally or as a result of external factors (e.g., radiation, chemicals).
- Genetic diversity arises from these mutations and provides the raw material for evolution by natural selection.

5. Protein Synthesis:

- DNA carries the instructions for the synthesis of proteins, which are the workhorses of the cell. Proteins perform a wide range of functions, including enzymatic reactions, structural support, and regulation of cellular processes.
- The process of protein synthesis involves the transcription of DNA into messenger RNA (mRNA) and the subsequent translation of mRNA into a specific protein sequence.

6. Epigenetics:

- In addition to the DNA sequence itself, the structure and chemical modifications of DNA can influence gene expression without altering the underlying genetic code.
- Epigenetic modifications, such as DNA methylation and histone modifications, can be heritable and play a role in gene regulation and cellular differentiation.

7. Conservation of Information:

- DNA's stability and ability to replicate accurately ensure the preservation of genetic information over generations.
- The conservation of genetic information allows for the continuation of species and the gradual evolution of life forms over time.

The discovery of DNA as the hereditary material, often attributed to the work of scientists such as Gregor Mendel, Rosalind Franklin, James Watson, Francis Crick, and others, has revolutionized our understanding of genetics, inheritance, and the diversity of life on Earth. It has also had profound implications for fields such as medicine, biotechnology, and forensics, as well as for our understanding of human evolution and the genetic basis of diseases.

Experiments affirming DNA is the main genetic material

By the early 1900's, biochemists had isolated hundreds of different chemicals from living cells. Which of these was the genetic material? Proteins seemed like promising candidates, since they were abundant, diverse, and complex molecules. However, a few key experiments demonstrated that DNA, rather than protein, is the genetic material.

A) Frederick Griffith: Bacterial transformation

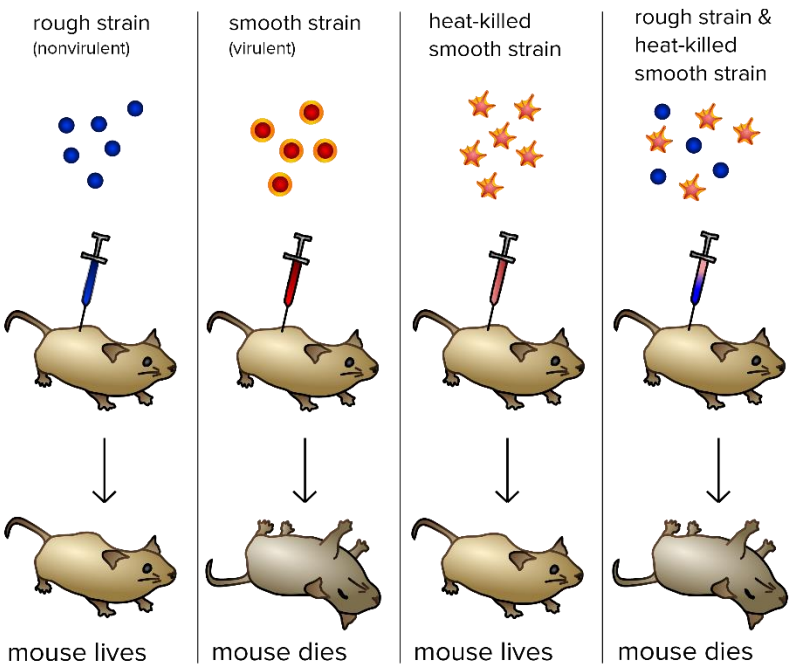


Fig. 2 Griffith experiment

In 1928, British bacteriologist Frederick Griffith conducted a series of experiments using *Streptococcus pneumoniae* bacteria and mice. Griffith wasn't trying to identify the genetic material, but rather, trying to develop a vaccine against pneumonia. In his experiments, Griffith used two related strains of bacteria, known as R and S.

- **R strain.** When grown in a petri dish, the R bacteria formed colonies, or clumps of related bacteria, that had well-defined edges and a rough appearance (hence the abbreviation "R"). The R bacteria were nonvirulent, meaning that they did not cause sickness when injected into a mouse.

- **S strain.** S bacteria formed colonies that were rounded and smooth (hence the abbreviation "S"). The smooth appearance was due to a polysaccharide, or sugar-based, coat produced by the bacteria. This coat protected the S bacteria from the mouse immune system, making them virulent (capable of causing disease). Mice injected with live S bacteria developed pneumonia and died.

As part of his experiments, Griffith tried injecting mice with heat-killed S bacteria (that is, S bacteria that had been heated to high temperatures, causing the cells to die). Unsurprisingly, the heat-killed S bacteria did not cause disease in mice.

The experiments took an unexpected turn, however, when harmless R bacteria were combined with harmless heat-killed S bacteria and injected into a mouse. Not only did the mouse develop pnenumonia and die, but when Griffith took a blood sample from the dead mouse, he found that it contained living S bacteria! Griffith concluded that the R-strain bacteria must have taken up what he called a "transforming principle" from the heat-killed S bacteria, which allowed them to "transform" into smooth-coated bacteria and become virulent.

B) Avery, McCarty, and MacLeod: Identifying the transforming principle

Background:

- The experiment built upon the work of Frederick Griffith, who had observed the transformation of non-virulent (R) bacteria into virulent (S) bacteria when exposed to heat-killed S bacteria. Griffith called this phenomenon "transformation" but did not identify the transforming substance.

Objective:

- To determine the nature of the "transforming principle" responsible for the transformation of non-virulent bacteria into virulent bacteria.

Experimental Procedure:

1. **Isolation of Components:**
    - Avery, MacLeod, and McCarty began by isolating various cellular components from the virulent (S) strain of Streptococcus pneumoniae bacteria. They extracted proteins, lipids, RNA, and DNA from the heat-killed S bacteria.
2. **Enzymatic Treatment:**
    - They treated each isolated component with specific enzymes to selectively break down one type of molecule while leaving others intact:
        - Enzyme DNase was used to degrade DNA.
        - Other enzymes were used to break down proteins, lipids, or RNA.
3. **Mixing with Non-Virulent Bacteria:**
    - After enzymatic treatment, they mixed each treated component separately with non-virulent (R) bacteria of the same strain.
    - Each mixture contained one of the treated components and R bacteria.
4. **Injection into Mice:**
    - Each mixture of treated components and R bacteria was injected into separate mice.

Results:

- When the mixture containing DNase-treated extracts of the heat-killed S bacteria was injected into mice, the R bacteria remained non-virulent. No transformation occurred.
- In contrast, when mixtures containing extracts treated with enzymes targeting proteins, lipids, or RNA were injected into mice, the R bacteria underwent transformation and became virulent, causing disease in the mice.
- The key observation was that only the destruction of DNA prevented transformation.

Conclusion:

- Avery, MacLeod, and McCarty's experiment conclusively demonstrated that DNA was the "transforming principle" responsible for the genetic transformation of non-virulent bacteria into virulent bacteria.
- Their results provided strong evidence that DNA carried the genetic information necessary for the transmission of genetic traits and laid the foundation for the recognition of DNA as the hereditary material in living organisms.
- This experiment was a milestone in the history of molecular biology and genetics, ultimately leading to the elucidation of the structure of DNA by James Watson and Francis Crick in 1953.

C) **The Hershey-Chase experiments**

The Hershey-Chase experiment, conducted by Martha Chase and Alfred Hershey in 1952, provided crucial evidence confirming that DNA, not protein, is the genetic material that carries hereditary information in bacteriophages (viruses that infect bacteria). This experiment played a significant role in elucidating the nature of genetic material and further supported the role of DNA in inheritance.

Experiment Overview:

1. **Bacteriophages:** Hershey and Chase used T2 bacteriophages for their experiment. These viruses infect Escherichia coli (E. coli) bacteria and consist of a protein coat and DNA core. The protein coat contains sulfur, while the DNA core contains phosphorus.
2. **Radioactive Labeling:** To distinguish between the protein coat and the DNA core, Hershey and Chase used radioactive isotopes. They labeled the protein coat with radioactive sulfur-35 (S-35) and the DNA with radioactive phosphorus-32 (P-32). Importantly, these labels allowed them to trace the fate of each component.

Experimental Procedure:

1. **Infection:** Hershey and Chase allowed the labeled T2 bacteriophages to infect E. coli bacteria. The viruses attached to the bacterial cell surface and injected their genetic material (either labeled DNA or labeled protein) into the host cell.

2. **Blender Experiment:** After allowing time for the infection process to occur, they took a crucial step. To separate the protein coat from the bacterial cell, they used a blender to shear off the protein coats from the surface of the infected bacterial cells. This step was critical because it ensured that only the labeled component (either DNA or protein) inside the bacterial cells would be analyzed further.

3. **Centrifugation:** Hershey and Chase then subjected the blended mixture to centrifugation. Centrifugation is a process that spins the mixture at high speeds, causing the denser components to separate from the less dense ones. In this case, the heavy, radioactive P-32-labeled DNA settled at the bottom of the tube, while the lighter protein coats and bacteria formed a supernatant.

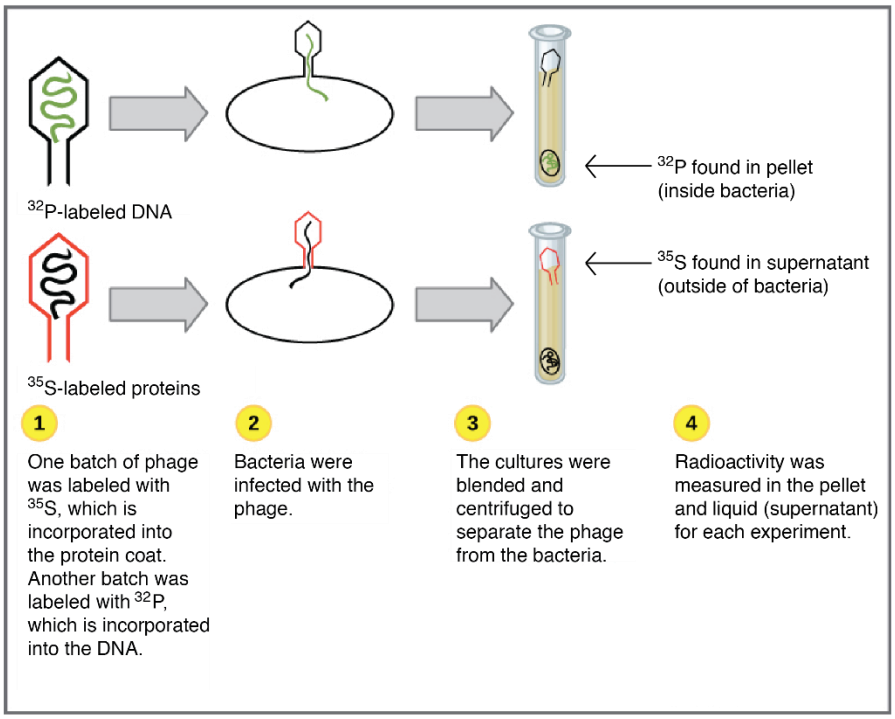


Fig. 3 The Hershey-Chase experiments

Results and Conclusions:

The key findings of the Hershey-Chase experiment were as follows:

- When they analyzed the bacterial cells for radioactivity, they found that P-32 (labeled DNA) entered the bacterial cells, while S-35 (labeled protein) remained outside the cells.

- This result demonstrated that the genetic material responsible for viral replication and heredity was DNA, not protein.

- The experiment supported the concept that DNA, with its ability to replicate and store genetic information, was the molecule responsible for transmitting genetic traits from one generation to the next.

In summary, the Hershey-Chase experiment provided strong evidence supporting the idea that DNA, not protein, is the genetic material responsible for transmitting hereditary information in living organisms. This groundbreaking experiment significantly contributed to our understanding of the molecular basis of genetics and laid the foundation for subsequent discoveries in molecular biology and genetics.

STRUCTURE OF DNA

Who Discovered DNA?

DNA was first recognized and identified by the Swiss biologist **Johannes Friedrich Miescher** in 1869 during his research on white blood cells.

The double helix structure of a DNA molecule was later discovered through the experimental data by James Watson and Francis Crick. Finally, it was proved that DNA is responsible for storing genetic information in living organisms.

DNA Types

There are three different DNA types:

- **A-DNA:** It is a right-handed double helix similar to the B-DNA form. Dehydrated DNA takes an A form that protects the DNA during extreme conditions such as desiccation. Protein binding also removes the solvent from DNA, and the DNA takes an A form.
- **B-DNA:** This is the most common DNA conformation and is a right-handed helix. The majority of DNA has a B type conformation under normal physiological conditions.
- **Z-DNA:** Z-DNA is a left-handed DNA where the double helix winds to the left in a zig-zag pattern. It was discovered by Andres Wang and Alexander Rich. It is found ahead of the start site of a gene and hence, is believed to play some role in gene regulation.

The components of DNA

From the work of biochemist Phoebus Levene and others, scientists in Watson and Crick's time knew that DNA was composed of subunits called **nucleotides**. A nucleotide is made up of a sugar (deoxyribose), a phosphate group, and one of four nitrogenous bases: adenine (A), thymine (T), guanine (G) or cytosine (C).

C and T bases, which have just one ring, are called **pyrimidines**, while A and G bases, which have two rings, are called **purines**.

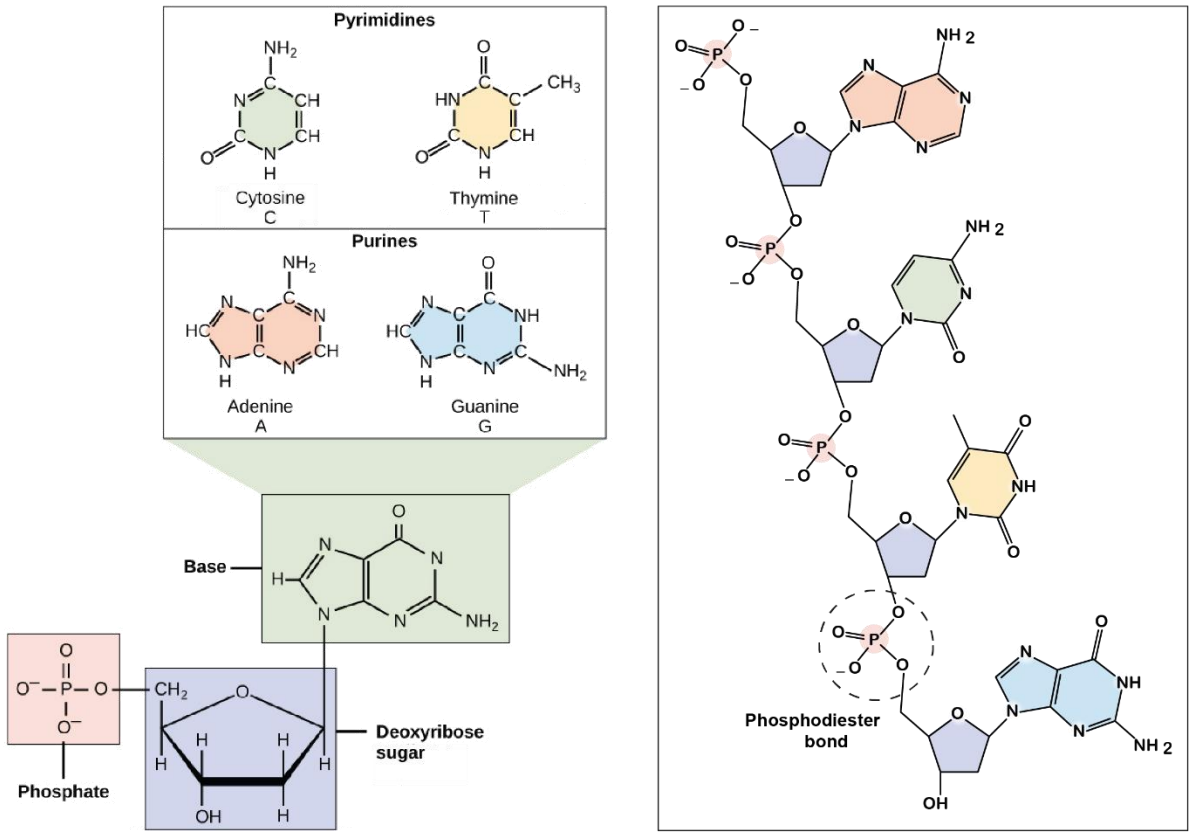


Fig. 4 The components of DNA

DNA nucleotides assemble in chains linked by covalent bonds, which form between the deoxyribose sugar of one nucleotide and the phosphate group of the next. This arrangement makes an alternating chain of deoxyribose sugar and phosphate groups in the DNA polymer, a structure known as the **sugar-phosphate backbone.**

Chargaff's rules

One other key piece of information related to the structure of DNA came from Austrian biochemist Erwin Chargaff. Chargaff analyzed the DNA of different species, determining its composition of A, T, C, and G bases. He made several key observations:

- A, T, C, and G were not found in equal quantities (as some models at the time would have predicted)
- The amounts of the bases varied among species, but not between individuals of the same species
- The amount of A always equalled the amount of T, and the amount of C always equalled the amount of G (A = T and G = C)

These findings, called **Chargaff's rules**, turned out to be crucial to Watson and Crick's model of the DNA double helix.

Watson, Crick, and Rosalind Franklin

Until the 1950s, the structure of DNA remained a mystery. In the early 1950s, American biologist James Watson and British physicist Francis Crick came up with their famous model of the DNA double helix. They were the first to cross the finish line in this scientific "race," with others such as Linus Pauling (who discovered protein secondary structure) also trying to find the correct model. Rather than carrying out new experiments in the lab, Watson and Crick mostly collected and analyzed existing pieces of data, putting them together in new and insightful ways. Some of their most crucial clues to DNA's structure came from Rosalind Franklin, a chemist working in the lab of physicist Maurice Wilkins. Franklin was an expert in a powerful technique for determining the structure of molecules, known as **X-ray crystallography**. When the crystallized form of a molecule such as DNA is exposed to X-rays, some of the rays are deflected by the atoms in the crystal, forming a **diffraction pattern** that gives clues about the molecule's structure.

Franklin’s crystallography gave Watson and Crick important clues to the structure of DNA. Some of these came from the famous “image 51,” a remarkably clear and striking X-ray diffraction image of DNA produced by Franklin and her graduate student.

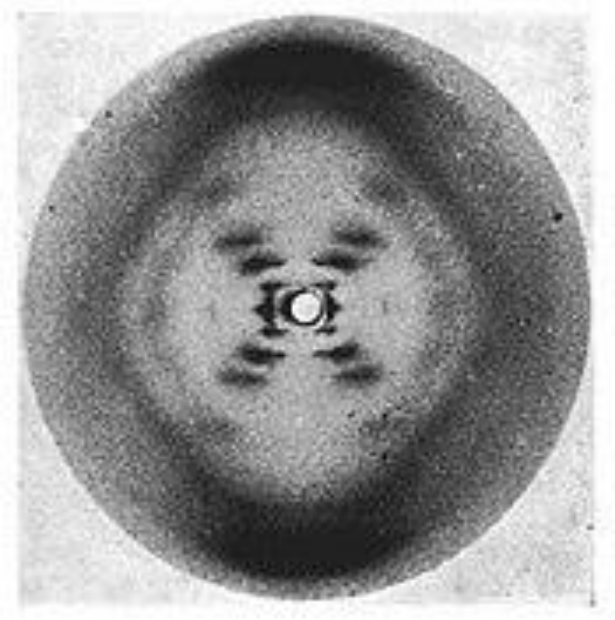


Fig. 5 Photo 51, showing X-ray diffraction pattern of DNA

To Watson, the X-shaped diffraction pattern of Franklin's image immediately suggested a helical, two-stranded structure for DNA. Watson and Crick brought together data from a number of researchers (including Franklin, Wilkins, Chargaff, and others) to assemble their celebrated model of the 3D structure of DNA. In 1962, James Watson, Francis Crick, and Maurice Wilkins were awarded the Nobel Prize in Medicine. Unfortunately, by then Franklin had died, and Nobel prizes are not awarded posthumously.

Watson and Crick's model of DNA

James Watson and Francis Crick are renowned for proposing the first accurate model of the DNA double helix in 1953, a landmark discovery that transformed our understanding of genetics and molecular biology. Their model of DNA incorporated several key features:

1. Double Helix Structure:

- Watson and Crick's model depicted DNA as a double helix, consisting of two long chains (strands) of nucleotides wound around each other in a spiral formation.
- The two DNA strands run in opposite directions, creating an anti-parallel arrangement.

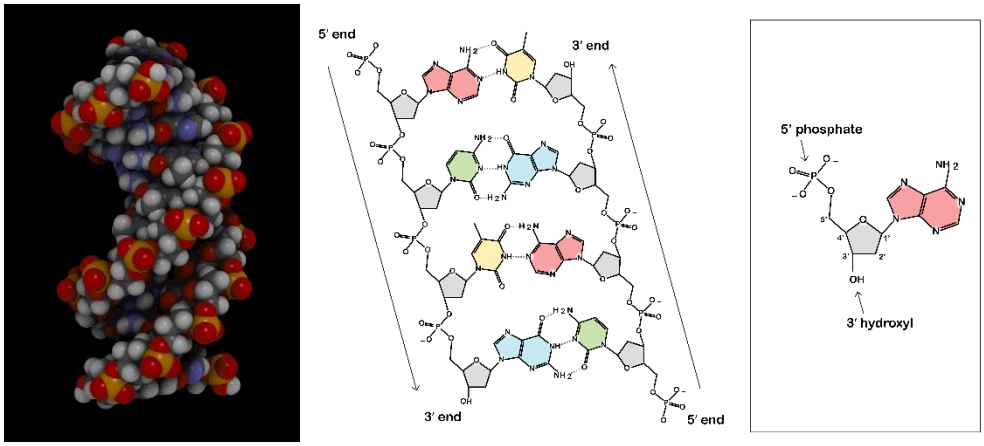


Fig. 6 DNA structure

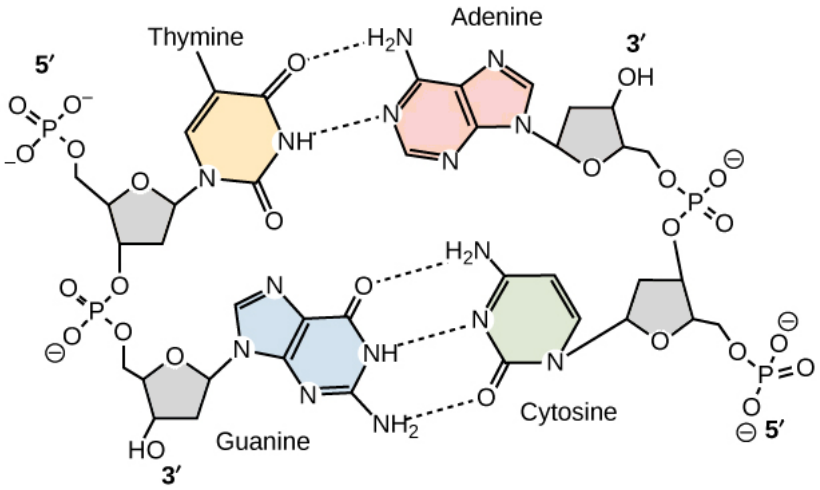
2. Complementary Base Pairing:

- The model explained that the nitrogenous bases in DNA are responsible for the specific pairing between the two DNA strands.
- Adenine (A) on one strand pairs with thymine (T) on the other strand via two hydrogen bonds.
- Cytosine (C) on one strand pairs with guanine (G) on the other strand via three hydrogen bonds.
- This complementary base pairing ensured that the two DNA strands are held together in a stable and precise manner.

3. Antiparallel Orientation:

- Watson and Crick's model emphasized that the two DNA strands are oriented in opposite directions, with one strand running from 5' to 3' and the other from 3' to 5'.
- This antiparallel arrangement ensures that the phosphate groups and deoxyribose sugars are aligned in a repeating pattern along the backbone.

4. Base Pairing Rules:

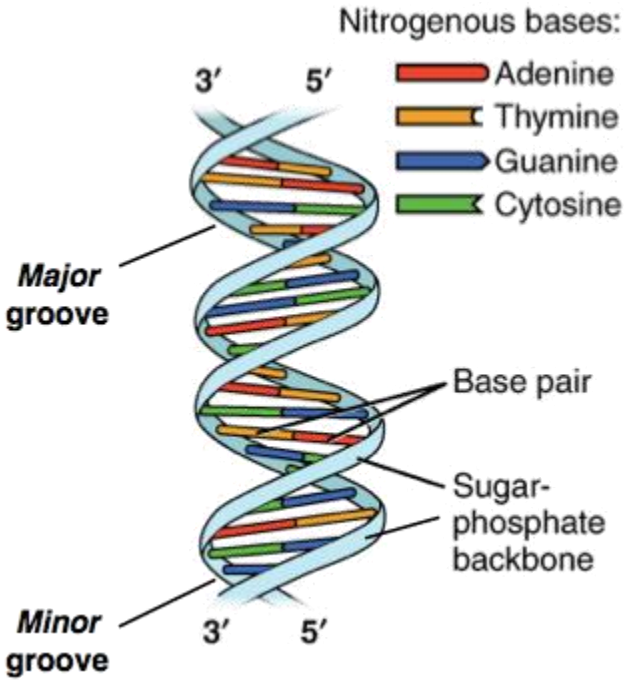
- The model adhered to Chargaff's rules, which state that the amount of adenine (A) is equal to the amount of thymine (T), and the amount of cytosine (C) is equal to the amount of guanine (G) in a given DNA molecule.
- These base pairing rules allow for the precise replication of DNA during cell division.
- 

5. Molecular Dimensions:

- Watson and Crick's model considered the dimensions and spacing between the components of DNA.
- They determined that one complete turn of the DNA double helix is approximately 34 angstroms (Å) in length, and each base pair contributes about 3.4 Å to this length.

6. Illustration of the Helix:

- Watson and Crick created a famous illustration of the right hand DNA double helix, showing the two strands spiraling around a central axis.
- All helices have a handedness, which is a property that describes how their grooves are oriented in space.
- The twisting of the DNA double helix and the geometry of the bases creates a wider gap (called the **major groove**) and a narrower gap (called the **minor groove**) that run along the length of the molecule, as shown in the figure above. These grooves are important binding sites for proteins that maintain DNA and regulate gene activity.
- This iconic image has become a symbol of modern biology.

- 

7. Replication and Information Storage:

- The model implied that the DNA double helix's structure facilitated the accurate replication of DNA, as each strand could serve as a template for the synthesis of a new complementary strand.
- It also underscored that the specific sequence of nucleotides along the DNA strands encoded genetic information.

Watson and Crick's model of the DNA double helix revolutionized the field of molecular biology and provided the foundation for understanding the structure, replication, and function of DNA. Their groundbreaking work earned them the Nobel Prize in Physiology or Medicine in 1962 and has had far-reaching implications in genetics, genomics, biotechnology, and medicine.

Structure of and types of RNA

Structurally speaking, ribonucleic acid (RNA), is quite similar to DNA. However, whereas DNA molecules are typically long and double stranded, RNA molecules are much shorter and are typically single stranded. RNA molecules perform a variety of roles in the cell but are mainly involved in the process of protein synthesis (translation) and its regulation.

RNA Structure

RNA is typically single stranded and is made of ribonucleotides that are linked by phosphodiester bonds. A ribonucleotide in the RNA chain contains ribose (the pentose sugar), one of the four nitrogenous bases (A, U, G, and C), and a phosphate group. The subtle structural difference between the sugars gives DNA added stability, making DNA more suitable for storage of genetic information, whereas the relative instability of RNA makes it more suitable for its more short-term functions.

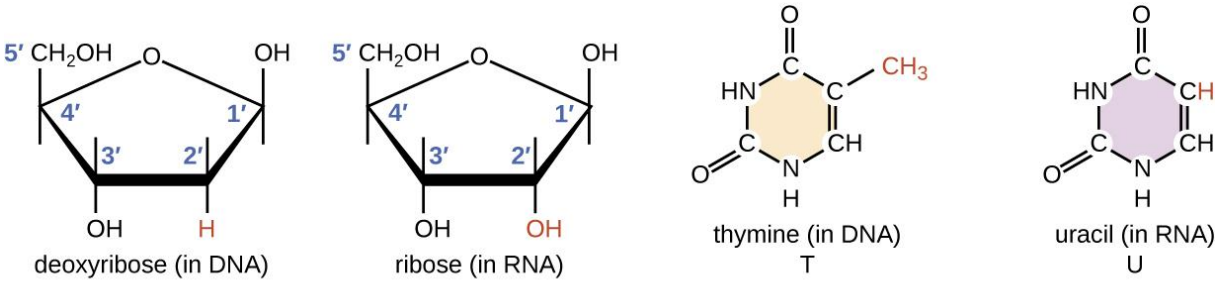


Figure (a) Ribonucleotides contain the pentose sugar ribose instead of the deoxyribose found in deoxyribonucleotides. (b) RNA contains the pyrimidine uracil in place of thymine found in DNA.

The RNA-specific pyrimidine uracil forms a complementary base pair with adenine and is used instead of the thymine used in DNA. Even though RNA is single stranded, most types of RNA molecules show extensive intramolecular base pairing between complementary sequences within the RNA strand, creating a predictable three-dimensional structure essential for their function.

RNA has three primary structural forms:

1. **Messenger RNA (mRNA):** mRNA carries genetic information from the DNA in the cell's nucleus to the ribosomes in the cytoplasm, where protein synthesis occurs. It is synthesized during transcription and serves as a template for protein synthesis.

   Structure: mRNA is single-stranded and consists of a ribose sugar-phosphate backbone with four nitrogenous bases (adenine, cytosine, guanine, and uracil).

   Function: mRNA carries genetic information from the DNA in the cell's nucleus to the ribosomes in the cytoplasm. It serves as a template for protein synthesis and is transcribed from DNA during the process of transcription.

2. **Transfer RNA (tRNA):** tRNA is responsible for bringing amino acids to the ribosome during protein synthesis. Each tRNA molecule has an anticodon that is complementary to a specific mRNA codon, ensuring the correct amino acid is added to the growing protein chain.

   Structure: tRNA is also a single-stranded molecule with a cloverleaf-like secondary structure. It has an anticodon region that pairs with the codon on mRNA and a specific amino acid attachment site.

   Function: tRNA is responsible for bringing amino acids to the ribosome during protein synthesis. It reads the genetic code on mRNA and ensures the correct amino acid is added to the growing protein chain.

3. **Ribosomal RNA (rRNA):** rRNA is a major component of ribosomes, the cellular machinery responsible for protein synthesis. It helps catalyze the formation of peptide bonds between amino acids, thus facilitating the assembly of the protein.

   Structure: rRNA is found in ribosomes and is responsible for the structural and functional components of these cellular organelles.

   Function: rRNA helps catalyze the formation of peptide bonds between amino acids, which is essential for the assembly of the protein during translation.

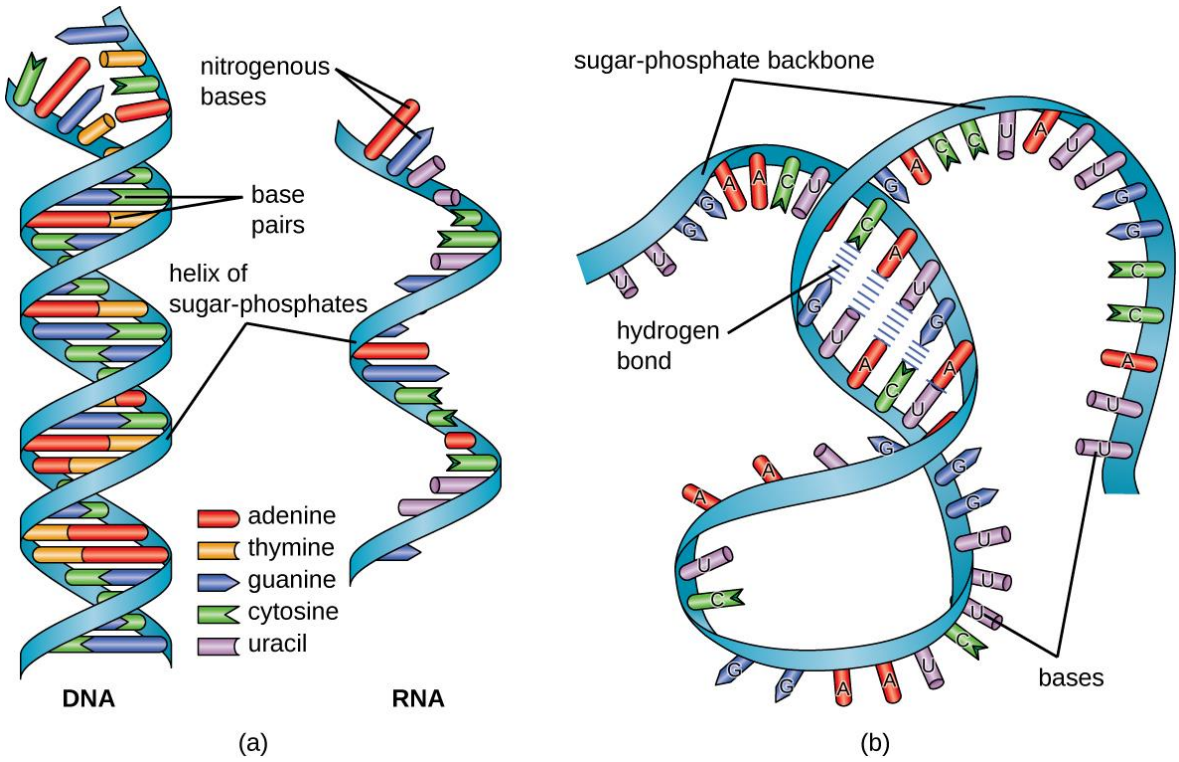


Figure (a) DNA is typically double stranded, whereas RNA is typically single stranded. (b) Although it is single stranded, RNA can fold upon itself, with the folds stabilized by short areas of complementary base pairing within the molecule, forming a three-dimensional structure.

Other forms of RNA include:

1. **Small Nuclear RNA (snRNA):** snRNA is involved in the splicing of pre-mRNA during the processing of mRNA. It plays a crucial role in removing introns and joining exons, resulting in the mature mRNA.

   Structure: snRNA molecules are typically short and single-stranded.

   Function: snRNA plays a crucial role in the splicing of pre-mRNA during RNA processing. It assists in the removal of introns and the joining of exons, leading to the production of mature mRNA.

2. **Small Nucleolar RNA (snoRNA):** snoRNA guides chemical modifications, such as methylation and pseudouridylation, of ribosomal RNA (rRNA) and small nuclear RNA (snRNA). These modifications are essential for the proper function of these RNAs.

   Structure: snoRNAs are usually single-stranded and fold into specific secondary structures.

   Function: snoRNA guides the chemical modification (methylation and pseudouridylation) of ribosomal RNA (rRNA) and small nuclear RNA (snRNA), ensuring proper RNA function.

3. **MicroRNA (miRNA):** miRNAs are small RNA molecules that regulate gene expression by binding to mRNA and either degrading it or inhibiting its translation. They play a significant role in post-transcriptional gene regulation.

   Structure: miRNAs are small, single-stranded RNA molecules, typically around 22 nucleotides in length.

   Function: miRNAs regulate gene expression by binding to complementary sequences on mRNA molecules. This binding can lead to mRNA degradation or translational inhibition, influencing protein production.

4. **Long Non-Coding RNA (lncRNA):** lncRNAs are longer RNA molecules that do not code for proteins. They have diverse roles in gene regulation, chromatin structure, and other cellular processes.

   Structure: lncRNAs are typically longer RNA molecules with diverse structural characteristics.

   Function: lncRNAs have a wide range of roles in gene regulation, chromatin structure, and other cellular processes. They do not code for proteins but are involved in regulating gene expression.

5. **Small Interfering RNA (siRNA):** siRNAs are short double-stranded RNA molecules that play a role in gene silencing by guiding the degradation of specific mRNA molecules.

   Structure: siRNAs are typically small, double-stranded RNA molecules.

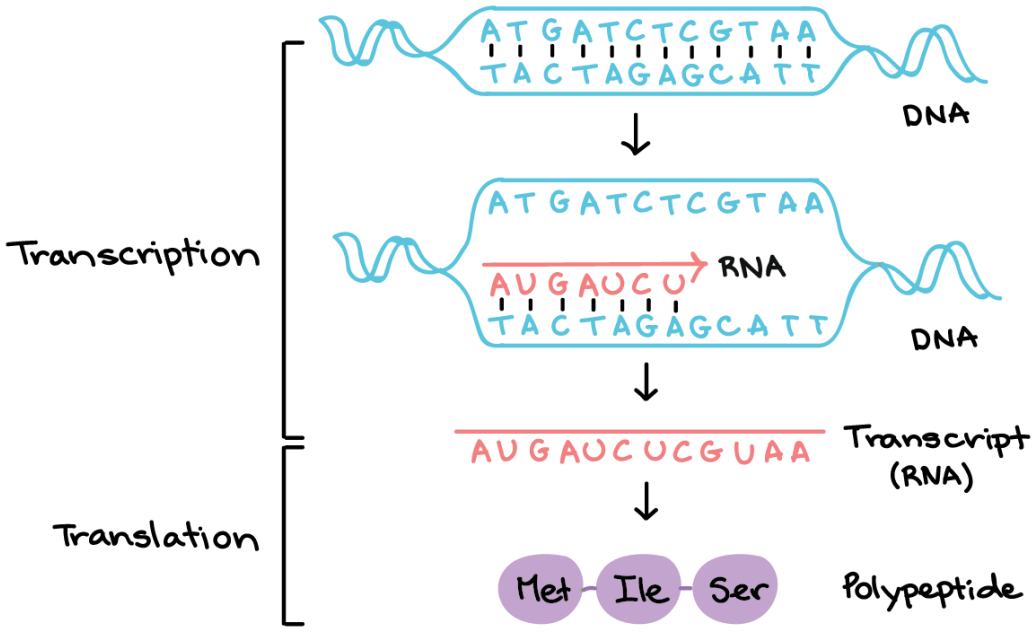
   Function: siRNAs play a role in gene silencing by guiding the degradation of specific mRNA molecules. They are involved in RNA interference (RNAi), a mechanism for controlling gene expression.

The various types of RNA work together to control gene expression, regulate protein synthesis, and ensure the proper functioning of cells and organisms. Each type of RNA plays a unique and vital role in these processes.

FLOW OF GENETIC INFORMATION

A) Transcription

"Transcription is the first step of gene expression that involves the formation of RNA molecule from DNA."



Key points:

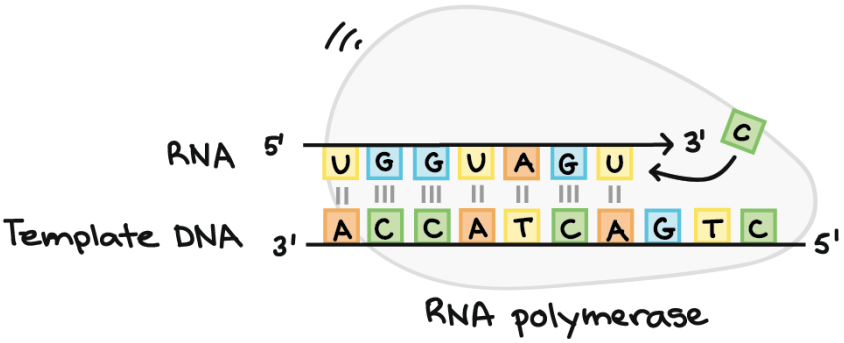
- **Transcription** is the first step in gene expression. It involves copying a gene's DNA sequence to make an RNA molecule.
- Transcription is performed by enzymes called **RNA polymerases**, which link nucleotides to form an RNA strand (using a DNA strand as a template).

- Transcription has three stages: initiation, elongation, and termination.
- In eukaryotes, RNA molecules must be processed after transcription: they are **spliced** and have a **5' cap** and **poly-A tail** put on their ends.
- Transcription is controlled separately for each gene in your genome.

Transcription is the first step in gene expression, in which information from a gene is used to construct a functional product such as a protein. The goal of transcription is to make a RNA copy of a gene's DNA sequence. For a protein-coding gene, the RNA copy, or **transcript**, carries the information needed to build a polypeptide (protein or protein subunit). Eukaryotic transcripts need to go through some processing steps before translation into proteins.

RNA polymerase

The RNA polymerase is the main enzyme involved in transcription. It uses single-strand DNA to synthesize a complementary RNA strand. The DNA-dependent RNA polymerase binds to the promoter and catalyses the polymerization in the 5’ to 3’ direction on the template strand. Once it reaches the terminator sequence, the process terminates and the newly synthesised RNA strand is released. Transcription Unit is a stretch of a DNA transcribed into an RNA molecule. Its function is to encode at least one gene. Suppose if gene encodes protein than mRNA is produced by transcription. A protein encoded by the DNA transcription unit may comprise a coding sequence. Compared to DNA replication, transcription has a lower copying fidelity.



Stages of Transcription

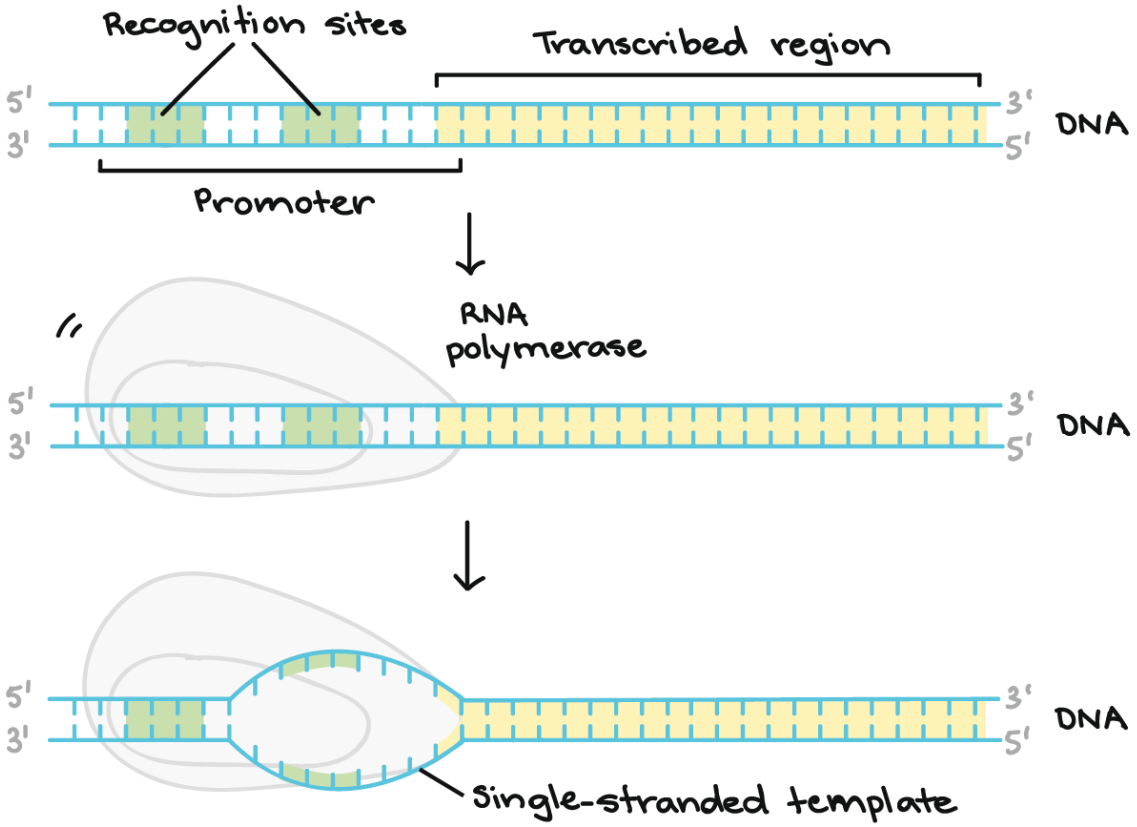
Transcription proceeds in enzymatically catalysed steps i.e.

1. Initiation
2. Elongation
3. Termination

1. Initiation. RNA polymerase binds to a sequence of DNA called the **promoter**, found near the beginning of a gene. Each gene (or group of co-transcribed genes, in bacteria) has its own

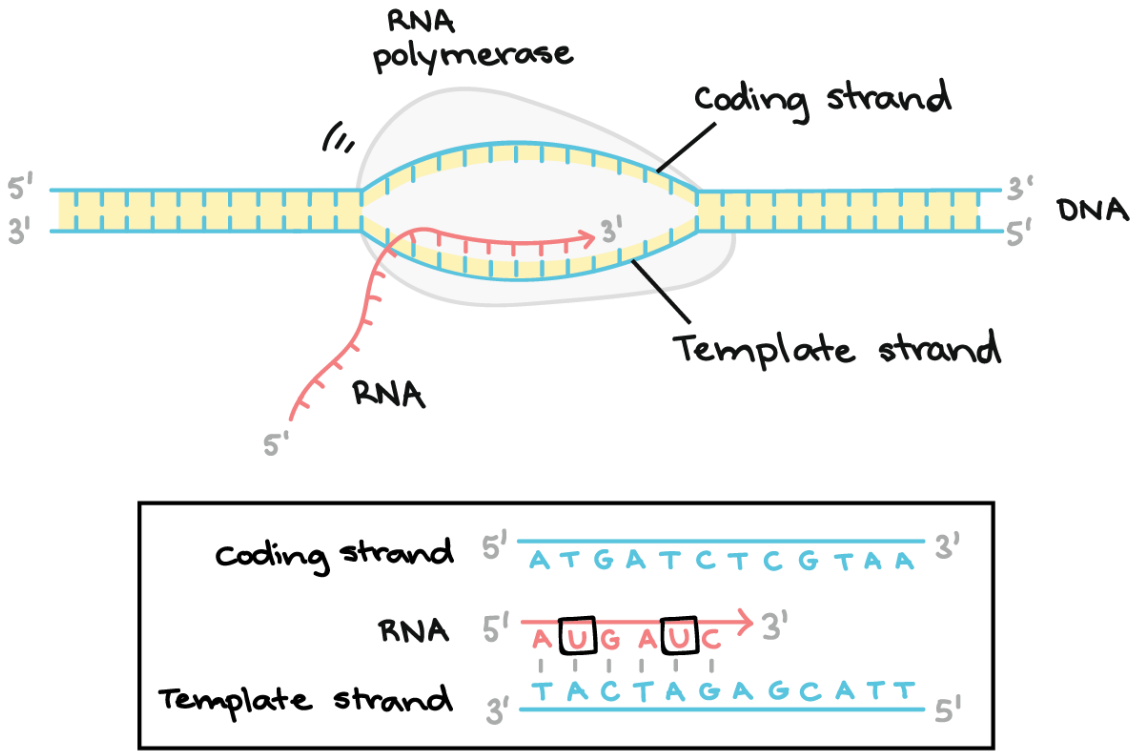
promoter. Once bound, RNA polymerase separates the DNA strands, providing the single-stranded template needed for transcription.

- In eukaryotes, transcription is initiated by several transcription factors and RNA polymerase. Specific transcription factors recognize promoter elements, including the TATA box, CAAT box, and GC box.
- RNA polymerase II is responsible for transcribing protein-coding genes (mRNA) and is recruited to the promoter by these transcription factors.

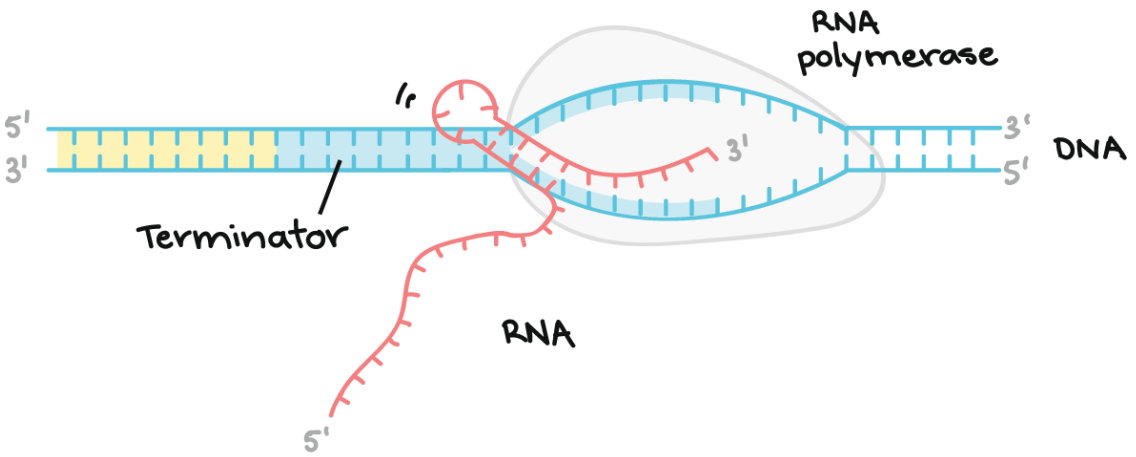


2. DNA Unwinding and Elongation:

- After the formation of the pre-initiation complex, RNA polymerase II unwinds the DNA and starts transcription. Unlike prokaryotes, eukaryotic transcription occurs in the nucleus. One strand of DNA, the **template strand**, acts as a template for RNA polymerase. As it "reads" this template one base at a time, the polymerase builds an RNA molecule out of complementary nucleotides, making a chain that grows from 5' to 3'. The RNA transcript carries the same information as the non-template (**coding**) strand of DNA, but it contains the base uracil (U) instead of thymine (T).



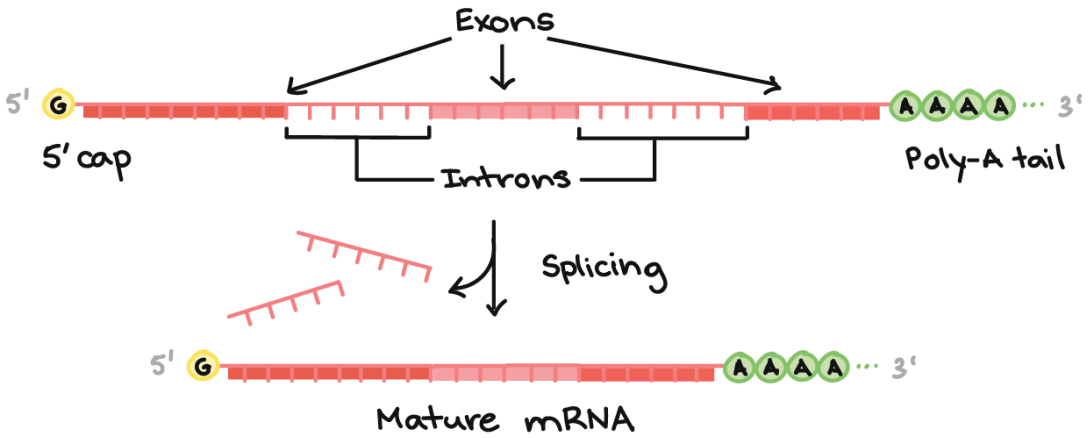
3. Termination. Sequences called **terminators** signal that the RNA transcript is complete. Once they are transcribed, they cause the transcript to be released from the RNA polymerase. An example of a termination mechanism involving formation of a hairpin in the RNA is shown below.



Eukaryotic RNA modifications

In bacteria, RNA transcripts can act as **messenger RNAs** (**mRNAs**) right away. In eukaryotes, the transcript of a protein-coding gene is called a **pre-mRNA** and must go through extra processing before it can direct translation.

- **5' Capping:** Eukaryotic pre-mRNAs must have their ends modified, by addition of a **5' cap** (at the beginning). A 7-methylguanosine cap is added to the 5' end of the pre-mRNA, protecting it from degradation and facilitating export from the nucleus.
- **Polyadenylation: 3' poly-A tail** (at the end). A poly(A) tail is added to the 3' end of the mRNA, contributing to stability and aiding in translation.
- **Splicing:** Many eukaryotic pre-mRNAs undergo **splicing**. Introns (non-coding regions) are removed, and exons (coding regions) are joined together to create a continuous coding sequence. This process is carried out by the spliceosome, composed of small nuclear ribonucleoproteins (snRNPs).



B) Translation: The synthesis of proteins

Translation involves translating the sequence of a messenger RNA (mRNA) molecule to a sequence of amino acids during protein synthesis. It is the process in which ribosomes synthesize proteins after the process of transcription of DNA to RNA. In an mRNA, the instructions for building a polypeptide come in groups of three nucleotides called codons as prescribed by the **genetic code (Table below)**. The genetic code is a triplet code read continuously from a fixed starting point in each mRNA. Specifically, it is defined by the following:

1. A group of three bases codes for one amino acid.

2. The code is not overlapping.

3. The base sequence is read from a fixed starting point without punctuation. That is, the

mRNA sequences contain no “commas” signifying appropriate groupings of triplets.

If the reading frame is displaced by one base, it remains shifted throughout the subsequent message; no “commas” are present to restore the “correct” frame.

4. The code is degenerate, meaning that, in most cases, each amino acid can be coded

by any of several triplets. Recall that a triplet code yields 64 codons for 20 amino

acids. Most codons (61 of 64) code for some amino acid.

The genetic code

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

Third-Base Degeneracy Is Color-Coded

Third-Base Relationship	Third Bases with Same Meaning	Number of Codons
Third base irrelevant	U, C, A, G	32 (8 families)
Purines	A or G	12 (6 pairs)
Pyrimidines	U or C	14 (7 pairs)
Three out of four	U, C, A	3 (AUX = Ile)
Unique definitions	G only	2 (AUG = Met) (UGG = Trp)
Unique definition	A only	1 (UGA = Stop)

*AUG signals translation initiation as well as coding for Met residues.

Several noteworthy features characterize the genetic code:

1. *All the codons have meaning.* Of the 64 codons, 61 specify particular amino acids. The remaining 3—UAA, UAG, and UGA—specify no amino acid and, thus, they are nonsense codons. Nonsense codons serve as termination codons; they are “stop” signals indicating that the end of the protein has been reached.

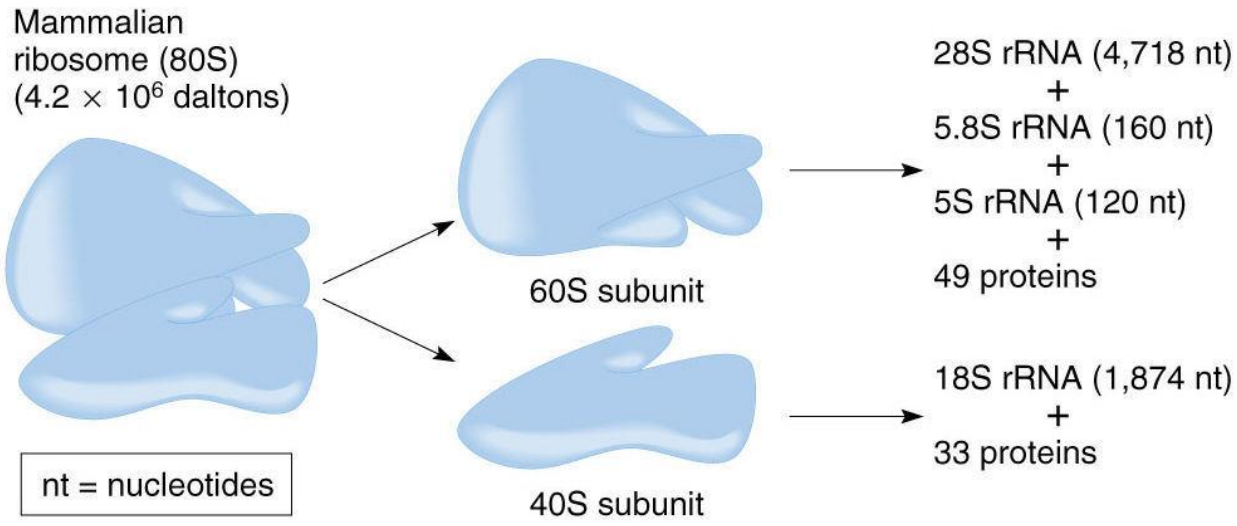
2. *The genetic code is unambiguous.* Each of the 61 “sense” codons encodes only one amino acid.

3. *The genetic code is degenerate.* With the exception of **Met (*AUG, is a “start” signal to kick off translation)* and Trp**, every amino acid is coded by more than one codon. Several—Arg, Leu, and Ser—are represented by six different codons. Codons coding for the same amino acid are called synonymous codons.

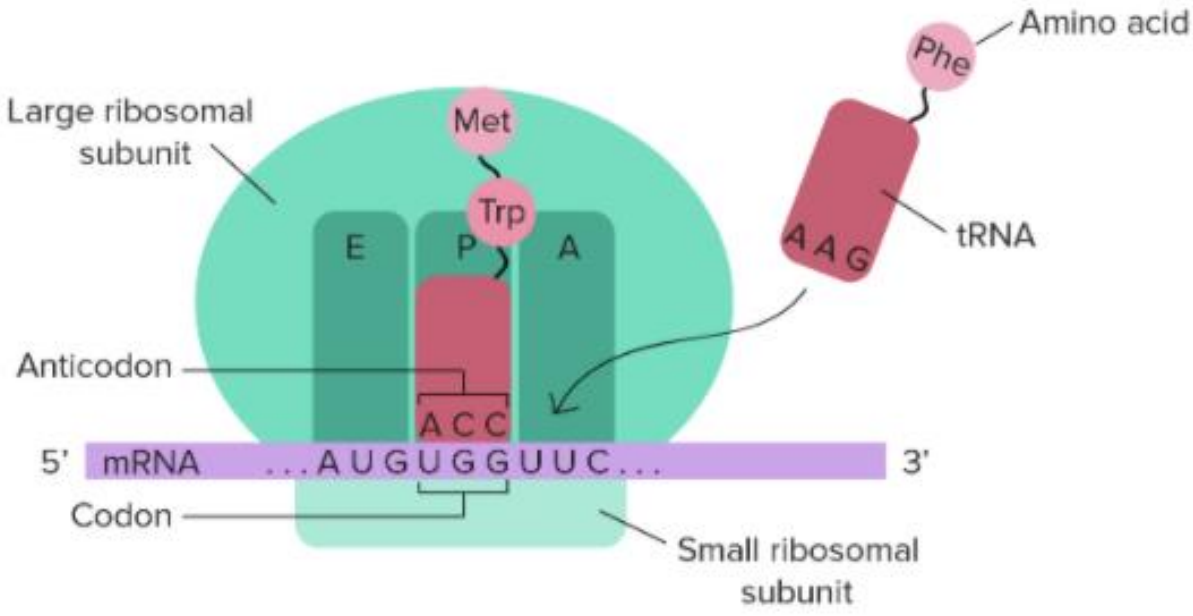
4. Codons representing the same amino acid or chemically similar amino acids tend to be similar in sequence. Often the third base in a codon is irrelevant, so, for example, all four codons in the GGX family specify Gly, and the UCX family specifies Ser. This feature is known as **third-base degeneracy**. Note also that codons with a pyrimidine as second base likely encode amino acids with hydrophobic side chains, and codons with a purine in the second-base position typically specify polar or charged amino acids. The two negatively charged amino acids, Asp and Glu, are

encoded by GAX codons; GA–pyrimidine gives Asp and GA–purine specifies Glu.

The Ribosomes



Ribosomes are essential cellular structures involved in protein synthesis, which is a crucial process in all living organisms. They are composed of two subunits, the small subunit (40S in eukaryotes) and the large subunit (60S in eukaryotes). These subunits work together to read the genetic code in mRNA (messenger RNA) and assemble amino acids into a polypeptide chain, forming a protein. In translation, the codons of an mRNA are read in order (from the 5' end to the 3' end) by molecules called transfer RNAs, or tRNAs. Each tRNA has an anticodon, a set of three nucleotides that binds to a matching mRNA codon through base pairing. The other end of the tRNA carries the amino acid that's specified by the codon.



Ribosomes exist normally as separate subunits that are composed of protein and rRNA. Eukaryotic ribosomes are larger (80S) and more complex than prokaryotic ribosomes (70S). The subunits come together to form a ribosome when they bind to an mRNA, near its 5' end. On binding to an mRNA, the ribosome reads the nucleotide sequence from the 5' to 3' direction, synthesizing the corresponding protein from amino acids in an N-terminal (amino-terminal) to C-terminal (carboxyl terminal) direction. Ribosomes are located in the cytosol, either freely floating or associated with the endoplasmic reticulum. They serve to synthesize proteins. Each prokaryotic ribosome, shown schematically, has three binding sites for tRNAs. The aminoacyl-tRNA binding site (or A site) is where, during elongation, the incoming aminoacyl-tRNA binds. The peptidyl-tRNA binding site (or P site) is where the tRNA linked to the growing polypeptide chain is bound. The exit site (or E site) is a binding site for tRNA following its role in translation and prior to its release from the ribosome. All three sites (A, P and E) are formed by the rRNA molecules in the ribosome.

Here's an overview of ribosome structure:

1. Small Subunit:
    - The small subunit contains ribosomal RNA (rRNA) molecules and a variety of proteins.
    - In eukaryotes, the small subunit contains the 18S rRNA.
    - It is responsible for binding to the mRNA and the initiator tRNA during the translation process.
2. Large Subunit:
    - The large subunit also contains rRNA and numerous proteins.
    - In eukaryotes, the large subunit contains the 5.8S, 25S (or 28S), and 5S rRNAs.
    - It plays a role in catalyzing the formation of peptide bonds between amino acids, resulting in polypeptide chain synthesis.
3. mRNA Binding Site:
    - The mRNA binding site is found between the small and large subunits and provides a location for the mRNA to be read.
4. A (aminoacyl) Site:
    - The A site in the ribosome is where the incoming aminoacyl-tRNA (tRNA carrying an amino acid) binds during translation.
5. P (peptidyl) Site:
    - The P site is where the growing polypeptide chain is held by the ribosome.
6. E (exit) Site:

- The E site is where the tRNA that no longer carries an amino acid is released from the ribosome.

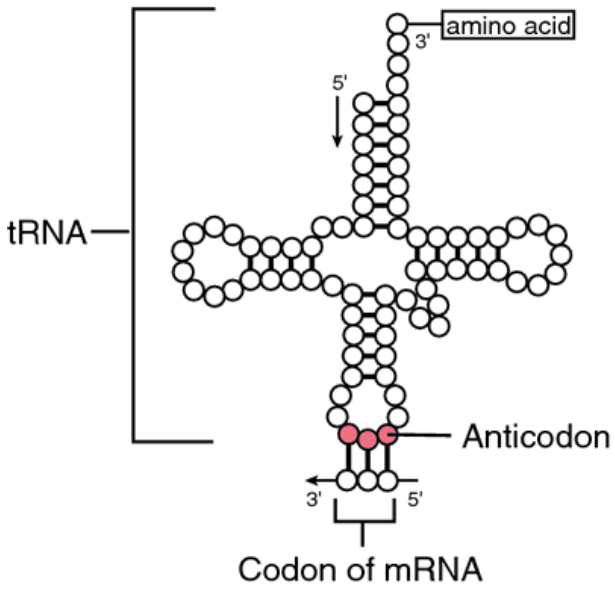
The ribosome's structure allows it to bring together the mRNA and tRNA, ensuring that the genetic information on the mRNA is read and translated into a specific sequence of amino acids, ultimately leading to protein synthesis.

The ribosome's function is highly conserved across all forms of life, although there are differences in the ribosome structure between prokaryotes (bacteria) and eukaryotes (plants, animals, fungi), as well as in the structure of ribosomes in mitochondria and chloroplasts, which have their own unique ribosomes. The exact arrangement and composition of ribosomal proteins and rRNA molecules can vary between species, reflecting evolutionary differences, but the core function of protein synthesis remains the same. Sometimes, group of ribosomes can be bound to an mRNA molecule like "beads" on a "thread" forming a polyribosome (or polysome or ergosome).

Translation

The translation process in ribosomes is a fundamental cellular process that converts the information carried by messenger RNA (mRNA) into a specific sequence of amino acids, ultimately leading to protein synthesis.

Components of Translation

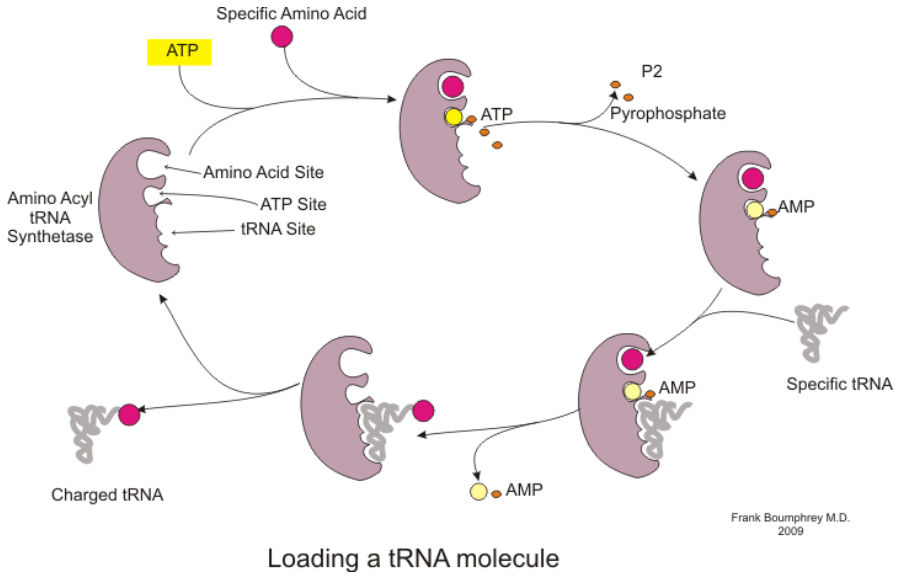


tRNA featuring the anticodon, complementary to specific mRNA

The key components required for translation are mRNA, ribosomes, and transfer RNA (tRNA). During translation, mRNA nucleotide bases are read as codons of three bases. Each 'codon' codes for a particular amino acid. Every tRNA molecule possesses an anticodon that is

complementary to the mRNA codon, and at the opposite end lies the attached amino acid. tRNA molecules are therefore responsible for bringing amino acids to the ribosome in the correct order ready for polypeptide assembly.

Aminoacyl-tRNA synthetases are enzymes that link amino acids to their corresponding tRNA molecules. The resulting complex is charged and is referred to as an aminoacyl-tRNA.



Aminoacyl-tRNA synthetase actively and specifically charging a tRNA

THE PROCESS OF TRANSLATION

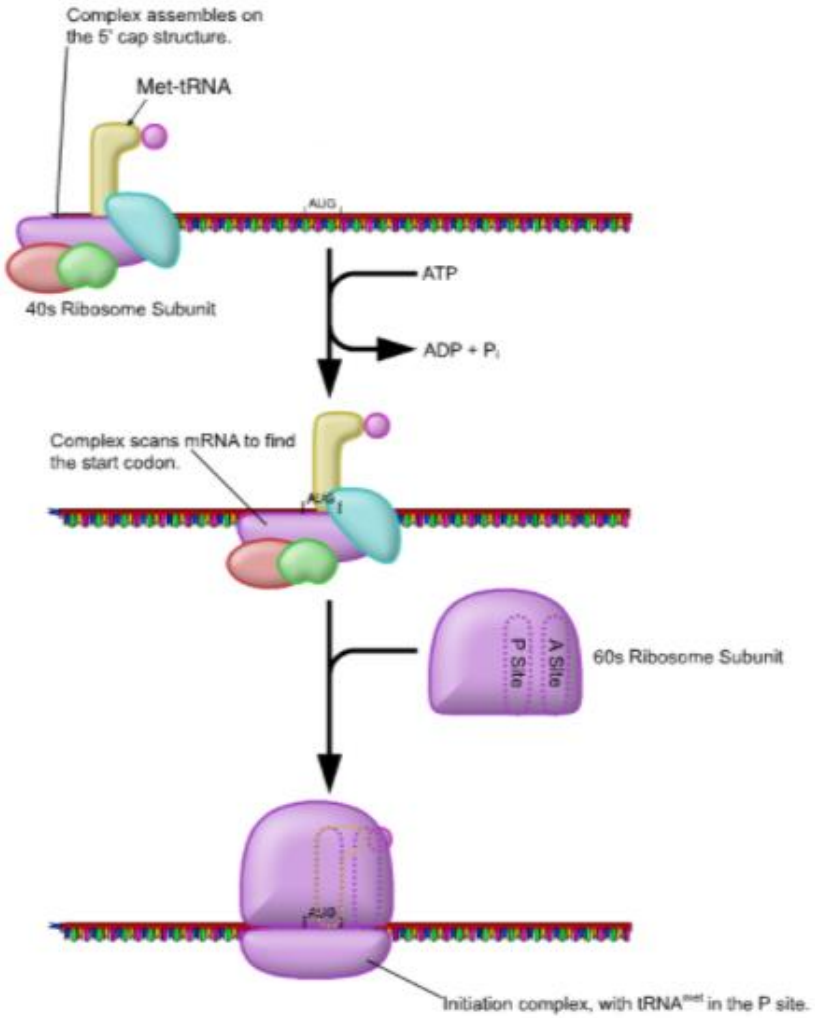
The overall mechanism of protein synthesis in eukaryotes is basically the same as in prokaryotes. However, there are some significant differences:

- Whereas a prokaryotic ribosome has a sedimentation coefficient of 70S and subunits of 30S and 50S, a eukaryotic ribosome has a sedimentation coefficient of 80S with subunits of 40S and 60S.
- The composition of eukaryotic ribosomal subunits is also more complex than prokaryotic subunits but the function of each subunit is essentially the same as in prokaryotes.
- In eukaryotes, each mRNA is monocistronic that is, discounting any subsequent post translational cleavage reactions that may occur; the mRNA encodes a single protein. In prokaryotes, many mRNAs are polycistronic that is they encode several proteins.
- Each coding sequence in a prokaryotic mRNA has its own initiation and termination codons.
- Initiation of protein synthesis in eukaryotes requires at least nine distinct eukaryotic initiation factors (eIFs) compared with the three initiation factors (IFs) in prokaryotes.

- In eukaryotes, the initiating amino acid is methionine, not N-formylmethionine as in prokaryotes.
- As in prokaryotes, a special initiator tRNA is required for initiation and is distinct from the tRNA that recognizes and binds to codons for methionine at internal positions in the mRNA. When charged with methionine ready to begin initiation, this is known as Met-tRNAimet.
- The main difference between initiation of translation in prokaryotes and eukaryotes is that in bacteria, a Shine–Dalgarno sequence lies 5’ to the AUG initiation codon and is the binding site for the 30S ribosomal subunit. In contrast, most eukaryotic mRNAs do not contain Shine–Dalgarno sequences. Instead, a 40S ribosomal subunit attaches at the 5’ end of the mRNA and moves downstream (i.e. in a 5’ to 3’ direction) until it finds the AUG initiation codon. This process is called scanning.
- Prokaryotic translation requires no helicase, presumably because protein synthesis in bacteria can start even as the mRNA is still being synthesized whereas in eukaryotes, transcription in the nucleus and translation in the cytoplasm are separate events which allows time for mRNA secondary structure to form.

Protein synthesis (or translation) takes place in three stages: Initiation, Elongation, Termination.

1. Initiation of Protein Synthesis



Initiation of translation showing charged Met-tRNA and the ribosome subunits at the start codon

- The process begins with the small ribosomal subunit (40S in eukaryotes) binding to the mRNA. In eukaryotes, this initiation complex also includes the initiator tRNA, which carries the amino acid methionine (or, in some cases, formylmethionine in prokaryotes).
- The small ribosomal subunit scans the mRNA until it finds the start codon, typically AUG, which codes for methionine. This signals the correct starting point for translation.
- The large ribosomal subunit (60S in eukaryotes) then joins the complex to form the complete ribosome.

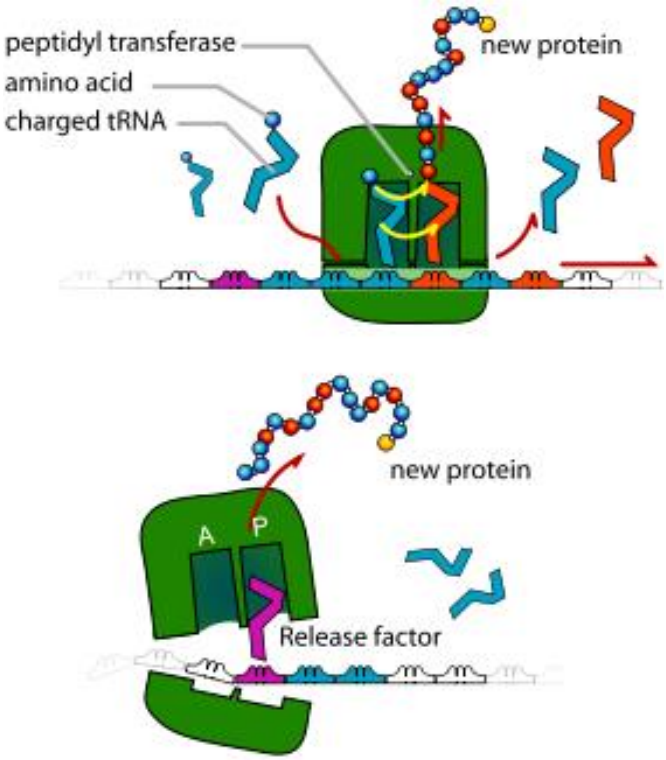
2. Elongation

- During elongation, the ribosome moves along the mRNA in a 5' to 3' direction, reading the codons and adding amino acids to the growing polypeptide chain.
- At each codon, a corresponding transfer RNA (tRNA) molecule binds to the mRNA in the ribosome's A site. The tRNA carries the complementary anticodon sequence and the appropriate amino acid.

- The ribosome catalyzes the formation of a peptide bond between the amino acid carried by the tRNA in the A site and the growing polypeptide chain held by the tRNA in the P site. This process transfers the amino acid from the A site to the P site.
- The ribosome then translocates, moving one codon along the mRNA. This process shifts the uncharged tRNA from the P site to the E (exit) site and the tRNA with the growing polypeptide chain from the A site to the P site.
- The A site is now empty and ready to accept the next tRNA with its corresponding amino acid.

3. Termination

- Translation continues until a stop codon (UAA, UAG, or UGA) is reached in the mRNA. Stop codons do not code for amino acids but signal the end of protein synthesis.
- When a stop codon is encountered in the A site, no tRNA with a complementary anticodon binds. Instead, release factors (proteins) bind to the ribosome, causing the release of the completed polypeptide chain from the P site.
- The small and large ribosomal subunits dissociate, and the mRNA is released, allowing the ribosome to be used in future translation processes.



Termination of translation upon encountering a stop codon at the P site

The resulting polypeptide chain is then subject to post-translational modifications and may fold into its native, functional structure to become a fully functional protein within the cell. It's important to note that the translation process is highly regulated and complex, involving multiple protein factors, initiation and termination factors, and quality control mechanisms to ensure the accuracy and fidelity of protein synthesis. Additionally, the specifics of translation may vary between prokaryotes and eukaryotes, and there are variations in the translation process for specialized organelles like mitochondria and chloroplasts.

Post-translational modifications (PTMs)

Post-translational modifications (PTMs) are covalent chemical modifications that occur on a protein after it has been synthesized during translation. These modifications play crucial roles in regulating the structure, function, localization, and stability of proteins in cells. PTMs can affect a protein's activity, interactions with other molecules, and its overall function. There are numerous types of post-translational modifications, with some of the most common and significant ones including:

1. Phosphorylation:
    - Phosphorylation involves the addition of a phosphate group to specific amino acid residues, most commonly serine, threonine, and tyrosine.
    - Protein kinases catalyze phosphorylation reactions, while protein phosphatases reverse the process.
    - Phosphorylation is a key regulatory mechanism for signal transduction, cellular processes, and protein function.
2. Glycosylation:
    - Glycosylation is the attachment of carbohydrate (glycan) chains to specific amino acid residues, primarily asparagine (N-linked glycosylation) or serine/threonine (O-linked glycosylation).
    - Glycosylation can influence protein folding, stability, and interactions, and it is essential for many cell surface receptors and adhesion molecules.
3. Acetylation:
    - Acetylation involves the addition of an acetyl group to the amino group of lysine residues.
    - Histone acetylation, for example, is involved in regulating chromatin structure and gene expression.

4. Methylation:
    - Methylation can occur on amino acid residues, such as lysine or arginine, and it can involve the addition of one or more methyl groups.
    - Methylation of histone proteins plays a role in chromatin structure and gene regulation.
5. Ubiquitination:
    - Ubiquitination is the attachment of a small protein called ubiquitin to lysine residues on a target protein.
    - It marks proteins for degradation by the proteasome or regulates their cellular localization and activity.
6. SUMOylation:
    - SUMOylation involves the attachment of small ubiquitin-like modifier (SUMO) proteins to specific lysine residues on target proteins.
    - SUMOylation plays a role in a variety of cellular processes, including nuclear-cytoplasmic transport, DNA repair, and protein-protein interactions.
7. Palmitoylation:
    - Palmitoylation is the addition of lipid (palmitate) groups to cysteine residues.
    - It anchors proteins to the cell membrane and can modulate their localization and activity.
8. Hydroxylation:
    - Hydroxylation adds hydroxyl groups to amino acid residues, often proline or lysine.
    - It is important in the synthesis and stabilization of collagen, as well as in the regulation of hypoxia-inducible factors (HIFs).
9. Sulfation:
    - Sulfation involves the addition of sulfate groups to tyrosine or other amino acids.
    - Sulfation is common in proteoglycans and can affect protein-protein interactions.
10. Nitrosylation:
    - Nitrosylation is the addition of a nitric oxide group to cysteine residues.

- It regulates various cellular processes, including vasodilation and neurotransmission.

These are just a few examples of the many post-translational modifications that can occur on proteins. PTMs are highly regulated and finely tuned processes that are essential for the proper functioning of cellular pathways and the diverse roles that proteins play in biology. They allow cells to quickly and dynamically respond to changes in their environment and adapt to various signals and stresses.

Nucleic acid isolation and analysis

The isolation and analysis of nucleic acids, which include DNA and RNA, are fundamental techniques in molecular biology and genetics. These procedures are critical for various applications, including genetic research, diagnostics, forensics, and biotechnology. Here is an overview of nucleic acid isolation and analysis:

Nucleic Acid Isolation:

1. **Sample Collection**
    - The process begins with the collection of biological samples. These samples can vary widely and may include tissues, cells, blood, saliva, or even environmental samples containing microbial DNA.
2. **Cell Lysis**
    - Cells in the sample need to be broken open to release their nucleic acids. This can be achieved through various methods:
        - Mechanical: Homogenization, sonication, or grinding.
        - Chemical: Detergents, chaotropic agents, and proteases can disrupt cell membranes and release nucleic acids.
        - Enzymatic: Enzymes like lysozyme and proteinase K can digest cell components, allowing nucleic acids to be released.
        - DNA extraction protocols generally use EDTA to chelate divalent cations, thereby inhibiting nuclease activity
3. **Nucleic Acid Extraction**
    - After cell lysis, nucleic acids are separated from other cellular components. Several extraction methods are commonly used:
    - Organic Extraction: Phenol-chloroform extraction is a classic method for separating nucleic acids from proteins and lipids. This involves mixing the lysate with phenol and chloroform, followed by centrifugation to separate the layers.
    - Solid-Phase Extraction (SPE): Columns or magnetic beads with specific binding properties can be used for nucleic acid purification. These can include silica-based columns or commercially available kits.
    - Alcohol Precipitation: DNA or RNA can be precipitated from solution using alcohol (usually isopropanol or ethanol). This method relies on the differential solubility of nucleic acids in alcohol and water.

4. **Purification**
    - To ensure the purity of isolated nucleic acids, additional purification steps are often required. Common purification methods include:
        - Size-exclusion chromatography: This separates nucleic acids by size, allowing the removal of contaminants.
        - Silica membrane-based purification: Utilizes the binding properties of silica to isolate nucleic acids efficiently.
        - Magnetic bead purification: Magnetic beads with nucleic acid-binding surfaces are used for isolation.

Gel electrophoresis

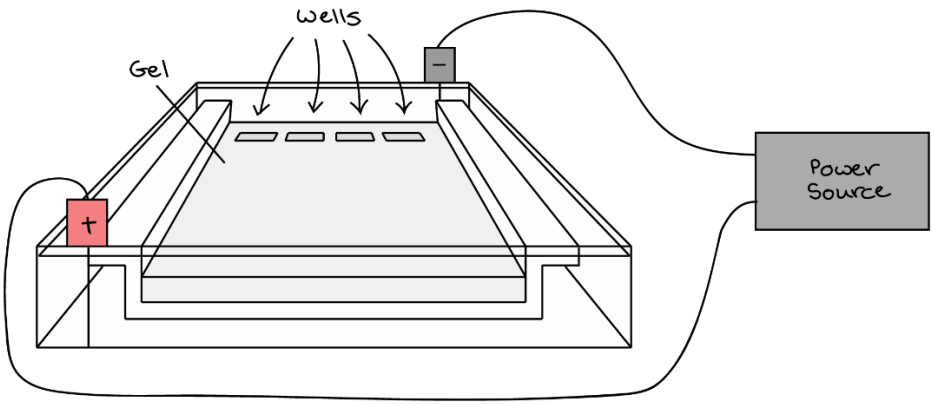
Gel electrophoresis is a technique used to separate DNA fragments (or other macromolecules, such as RNA and proteins) based on their size and charge. Electrophoresis involves running a current through a gel containing the molecules of interest. Based on their size and charge, the molecules will travel through the gel in different directions or at different speeds, allowing them to be separated from one another.

- Gel electrophoresis is a versatile and essential technique in molecular biology.
- It is used for a wide range of applications, including DNA fingerprinting, DNA sequencing, RNA analysis, and protein characterization.
- This method allows researchers to separate and visualize biomolecules based on size and charge, providing valuable insights into their characteristics.

All DNA molecules have the same amount of charge per mass. Because of this, gel electrophoresis of DNA fragments separates them based on size only. Using electrophoresis, we can see how many different DNA fragments are present in a sample and how large they are relative to one another. We can also determine the absolute size of a piece of DNA by examining it next to a standard "yardstick" made up of DNA fragments of known sizes.

As the name suggests, gel electrophoresis involves a gel: a slab of Jello-like material. Gels for DNA separation are often made out of a polysaccharide called **agarose**, which comes as dry, powdered flakes. When the agarose is heated in a buffer (water with some salts in it) and allowed to cool, it will form a solid, slightly squishy gel. At the molecular level, the gel is a matrix of agarose molecules that are held together by hydrogen bonds and form tiny pores.

At one end, the gel has pocket-like indentations called **wells**, which are where the DNA samples will be placed:

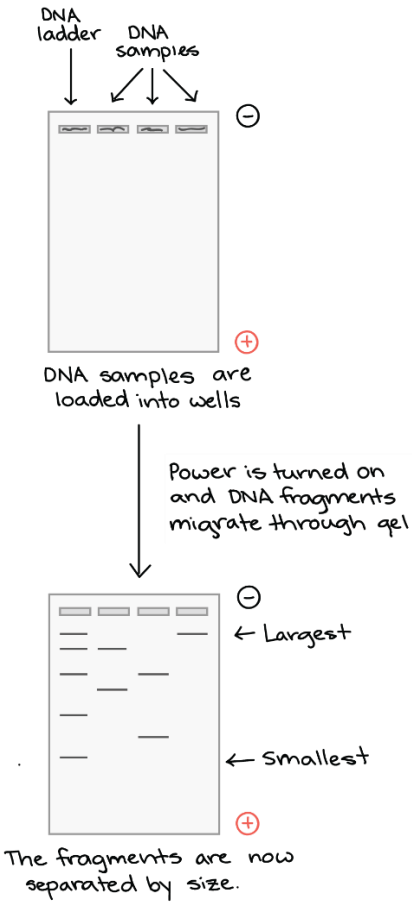


Before the DNA samples are added, the gel must be placed in a **gel box**. One end of the box is hooked to a positive electrode, while the other end is hooked to a negative electrode. The main body of the box, where the gel is placed, is filled with a salt-containing buffer solution that can conduct current. The end of the gel with the wells is positioned towards the negative electrode. The end without wells (towards which the DNA fragments will migrate) is positioned towards the positive electrode.

Once the gel is in the box, each of the DNA samples we want to examine (for instance, each PCR reaction or each restriction-digested plasmid) is carefully transferred into one of the wells. One well is reserved for a **DNA ladder**, a standard reference that contains DNA fragments of known lengths. Commercial DNA ladders come in different size ranges, so we would want to pick one with good "coverage" of the size range of our expected fragments.

Next, the power to the gel box is turned on, and current begins to flow through the gel. The DNA molecules have a negative charge because of the phosphate groups in their sugar-phosphate backbone, so they start moving through the matrix of the gel towards the positive pole.

As the gel runs, shorter pieces of DNA will travel through the pores of the gel matrix faster than longer ones. After the gel has run for awhile, the shortest pieces of DNA will be close to the positive end of the gel, while the longest pieces of DNA will remain near the wells.



Materials Needed:

- Gel electrophoresis apparatus, including a gel casting tray, comb, and chamber with electrodes.
- Agarose or polyacrylamide gel.
- Electrophoresis buffer (e.g., TAE or TBE).
- DNA, RNA, or protein samples mixed with a loading dye.

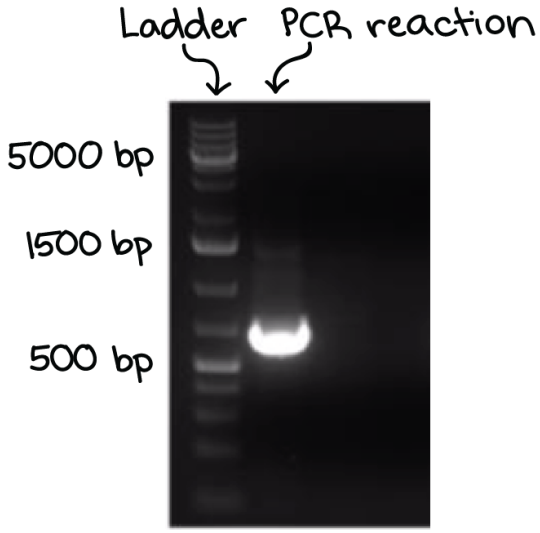
Steps of Gel Electrophoresis:

1. **Preparing the Gel:**
    - Prepare the gel by dissolving agarose or polyacrylamide in an electrophoresis buffer, then pouring it into a gel casting tray. A comb creates wells in the gel.
2. **Sample Loading:**
    - Mix samples with a loading dye to assist in loading. Fill the wells with these sample-loading mixtures.
3. **Running the Electrophoresis:**

- Place the gel in the electrophoresis chamber and connect the electrodes. Apply an electric current to move the molecules through the gel.

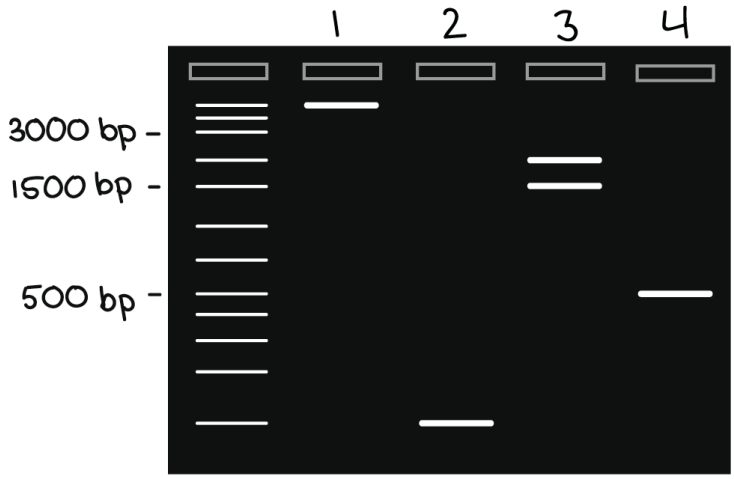
4. **Separation of Molecules:**
    - Smaller molecules move faster, and larger ones move more slowly, leading to separation based on size and charge.
5. **Visualization:**
    - Stain the gel with a dye (e.g., ethidium bromide for DNA) and visualize under UV light. This results in bands or patterns corresponding to different molecules.
    - **Visualization of DNA fragments**

Once the fragments have been separated, we can examine the gel and see what sizes of bands are found on it. When a gel is stained with a DNA-binding dye and placed under UV light, the DNA fragments will glow, allowing us to see the DNA present at different locations along the length of the gel.



A well-defined “line” of DNA on a gel is called a **band**. Each band contains a large number of DNA fragments of the same size that have all traveled as a group to the same position. A single DNA fragment (or even a small group of DNA fragments) would not be visible by itself on a gel. By comparing the bands in a sample to the DNA ladder, we can determine their approximate sizes. For instance, the bright band on the gel above is roughly \[700\] base pairs (bp) in size.

Check your understanding



Which lane matches each description below?

	1	2	3	4
This lane contains the longest DNA fragment.	○	○	○	○
This lane contains the shortest DNA fragment.	○	○	○	○
This lane contains a 1500 base pair (bp) DNA fragment.	○	○	○	○

Applications:

- **DNA Electrophoresis:** Used for DNA size determination, fragment analysis, and purity assessment.
- **RNA Electrophoresis:** Assesses RNA integrity, size, and extraction success.
- **Protein Electrophoresis:** Determines protein size, purity, and concentration.
- **Blotting Techniques:** Combine electrophoresis with blotting to transfer molecules to membranes for further analysis (Southern, Northern, Western blotting).

NUCLEOTIDE SEQUENCING

DNA SEQUENCING

Nucleotide sequencing is a fundamental technique in molecular biology and genetics used to determine the precise order of nucleotides (adenine, thymine, cytosine, and guanine) in a DNA or RNA molecule.

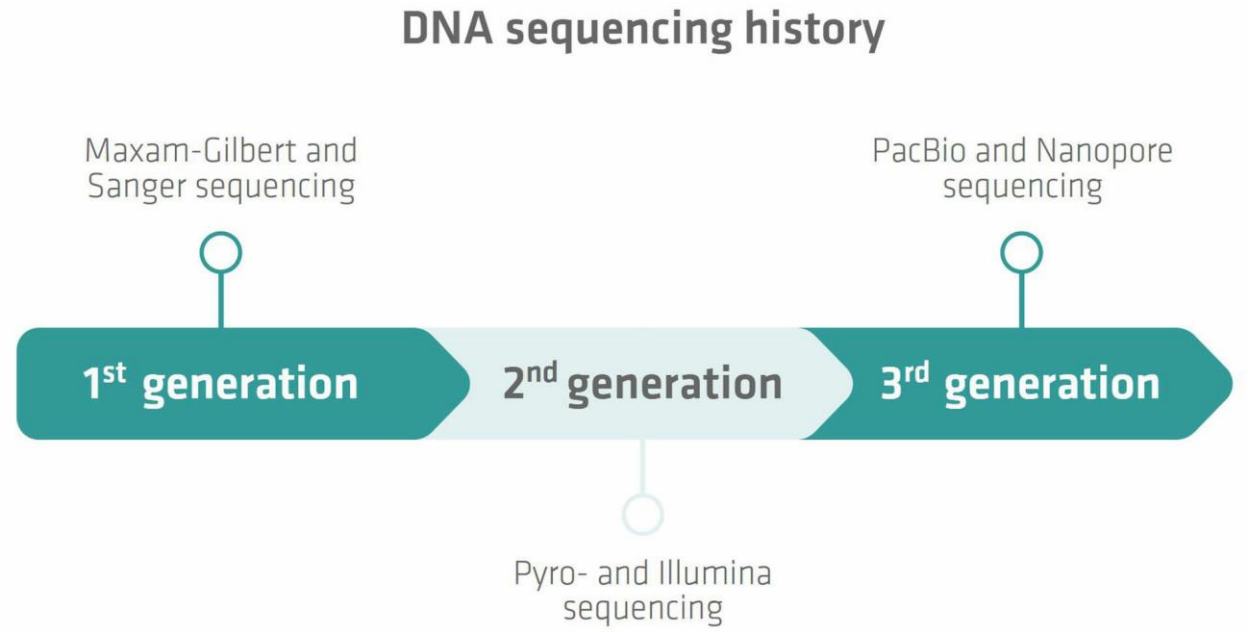
Significance of DNA Sequencing

- Information obtained by DNA sequencing makes it possible to understand or alter the function of genes.
- DNA sequence analysis demonstrates regulatory regions that control gene expression and genetic “hot spots” particularly susceptible to mutation.
- Comparison of DNA sequences shows evolutionary relationships that provide a framework for definite classification of microorganisms including viruses.
- Comparison of DNA sequences facilitates identification of conserved regions, which are useful for development of specific hybridization probes to detect microorganisms including viruses in clinical samples.
- DNA sequencing has become sufficiently fast and inexpensive to allow laboratory determination of microbial sequences for identification of microbes. Sequencing of the 16S ribosomal subunit can be used to identify specific bacteria. Sequencing of viruses can be used to identify the virus and distinguish different strains.
- DNA sequencing shows gene structure that helps research workers to find out the structure of gene products.

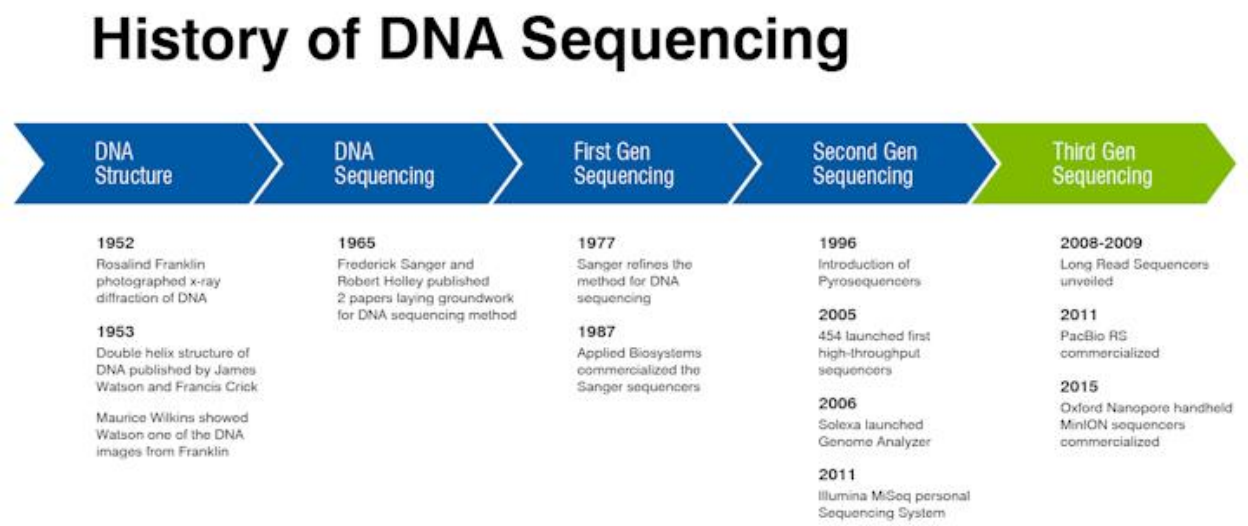
Applications of DNA Sequencing

- **Genomics**: Understanding genetic diseases, population genetics, and evolutionary biology.
- **Clinical Diagnostics**: Identifying mutations associated with diseases, pharmacogenomics.
- **Microbial Ecology**: Studying biodiversity and microbial communities through metagenomics.
- **Forensic Science**: DNA fingerprinting for criminal investigations.

DNA sequencing history



Timeline of DNA Sequencing History



1950s: Foundations

- **1953**: James Watson and Francis Crick publish their model of the DNA double helix, proposing a mechanism for genetic information storage and transmission.

1970s: Early Sequencing Techniques

- **1970**: **Maxam-Gilbert Method** is developed by Allan Maxam and Walter Gilbert. This chemical cleavage method allowed for the sequencing of short DNA fragments.
- **1975**: **Sanger Sequencing** is developed by Frederick Sanger. This method uses chain-terminating dideoxynucleotides (ddNTPs) to terminate DNA synthesis, allowing for the determination of DNA sequences.

1980s: Advancements in Sanger Sequencing

- **1980**: Sanger sequencing is refined and adapted for larger DNA fragments. It becomes the dominant method for DNA sequencing.
- **1986**: Sanger sequencing is used to sequence the first complete DNA molecule, the genome of the bacteriophage φX174.

1990s: The Human Genome Project

- **1990**: The Human Genome Project (HGP) is launched, aiming to sequence the entire human genome.
- **1995**: The first complete bacterial genome, that of Haemophilus influenzae, is sequenced, marking a significant achievement in genomics.
- **1996**: The first eukaryotic genome (that of Saccharomyces cerevisiae, or baker's yeast) is sequenced.

2000s: Next-Generation Sequencing (NGS)

- **2001**: The first draft of the human genome is published by the Human Genome Project, and Celera Genomics, marking a landmark achievement in genetics.
- **2005**: Next-generation sequencing technologies begin to emerge, allowing for massively parallel sequencing. This includes platforms like 454 Pyrosequencing.
- **2007**: Illumina and SOLiD (supported by Applied Biosystems) develop new NGS technologies that further decrease sequencing costs and increase throughput.

2010s: Widespread Adoption of NGS

- **2010**: The first whole-genome sequencing of an individual human (James Watson) is completed using NGS technology.
- **2012**: The emergence of Oxford Nanopore technology, allowing for real-time sequencing of long DNA fragments.
- **2013**: The first NGS-based whole-genome sequencing of a cancer patient’s tumor, enabling personalized medicine approaches.

2020s: Advances and Innovations

- **2020**: The COVID-19 pandemic highlights the importance of rapid sequencing technologies for tracking virus mutations and outbreaks.
- **2021**: Advances in single-cell sequencing technologies enhance our understanding of cellular heterogeneity and gene expression.

- **2023**: Continued development in long-read sequencing technologies (e.g., PacBio and Oxford Nanopore) improve the accuracy of complex genomic regions and structural variants.

TYPES OF DNA SEQUENCING TECHNOLOGIES

Three main methods are widely known to be used to sequence DNA:

1. **The Chemical Method** (also called the Maxam–Gilbert method after its inventors).
2. **The Chain Termination Method** (also known as the Sanger dideoxy method after its inventor).
3. **Next-generation sequencing (NGS)**

- Maxam–Gilbert technique depends on the relative chemical liability of different nucleotide bonds, whereas the Sanger method interrupts elongation of DNA sequences by incorporating dideoxynucleotides into the sequences.
- The chain termination method is the method more usually used because of its speed and simplicity.

Chemical Cleavage Method (Maxam–Gilbert Method)

The Maxam–Gilbert method is a chemical cleavage technique for DNA sequencing developed by Allan Maxam and Walter Gilbert in the 1970s. This method was one of the first to be used for sequencing DNA and is based on the selective cleavage of DNA strands at specific nucleotide bases.

Overview of the Maxam–Gilbert Method

Principle

The Maxam–Gilbert method utilizes chemical reagents to cleave DNA at specific bases, allowing for the identification of the sequence of nucleotides in a given DNA strand.

Key Steps

1. **Preparation of the DNA Sample**
    - The DNA of interest is labeled at one end, usually with a radioactive phosphate. This labeling allows for the detection of the fragments after cleavage.
2. **Chemical Reactions for Cleavage**
    - The labeled DNA is subjected to specific chemicals that cleave the DNA at particular bases:

        - **Guanine (G)**: The DNA is treated with a reagent such as formic acid, resulting in cleavage at guanine residues.
        - **Adenine (A)**: Treatment with a specific reagent cleaves at adenine residues.
        - **Cytosine (C)**: Another reagent cleaves at cytosine residues.
        - **Thymine (T)**: A different reagent cleaves at thymine residues.
    - The specific reagents used in the original method are:
        - **G**: Formic acid and piperidine for guanine.
        - **A**: Hydrazine for adenine.
        - **C**: Hydrazine (with a different reaction condition) for cytosine.
        - **T**: An alkaline solution for thymine.
3. **Fragment Separation**
    - After the cleavage reactions, the resulting DNA fragments are separated by size using polyacrylamide gel electrophoresis. This technique allows for the resolution of fragments based on their length.
4. **Visualization**
    - The gel is then subjected to autoradiography, where the radioactive fragments produce a visual pattern on film. The position of the bands indicates the size of the fragments, which correspond to the nucleotides that were cleaved.
5. **Reading the Sequence**
    - By running four separate reactions (one for each base), the sequence can be deduced from the pattern of bands observed on the gel. The order of the bands indicates the position of the bases in the original DNA strand.

Advantages and Disadvantages

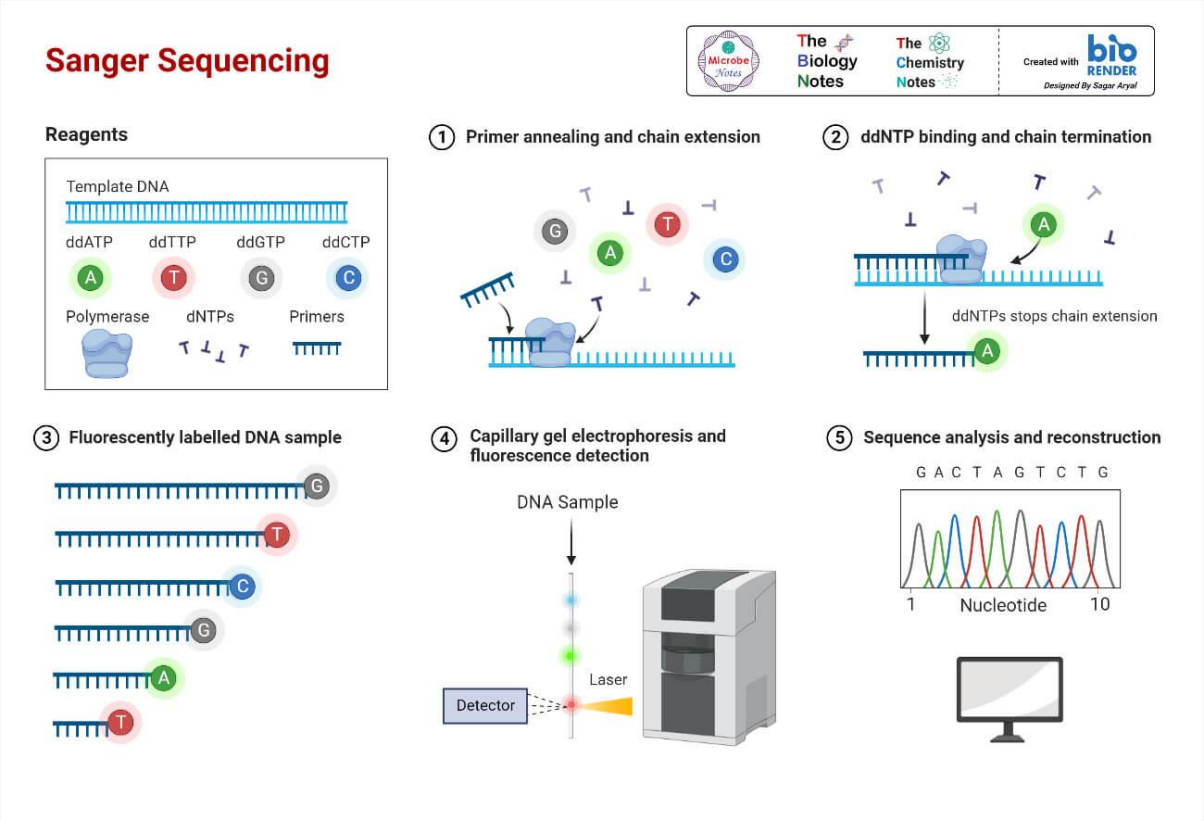
Advantages

- **High Resolution**: Provides high-resolution results due to the ability to separate fragments accurately.
- **Direct Sequence Information**: Cleavage occurs at specific bases, allowing for straightforward determination of the sequence.

Disadvantages

- **Labor Intensive**: Requires multiple steps and is more labor-intensive than modern sequencing techniques.
- **Radioactive Materials**: Uses radioactive labeling, which poses safety and handling concerns.
- **Limited to Short Sequences**: Not suitable for sequencing large regions of DNA or entire genomes.

Chain Termination Method (Sanger Dideoxy Method)



The Chain Termination Method, commonly known as Sanger sequencing, is a widely used technique for determining the nucleotide sequence of DNA. Developed by Frederick Sanger in the 1970s, this method revolutionized molecular biology and laid the foundation for modern sequencing technologies. Here's an overview of the method:

Overview of the Chain Termination Method (Sanger Sequencing)

Principle

The chain termination method relies on the incorporation of dideoxynucleotides (ddNTPs) during DNA synthesis, which terminate the elongation of the DNA strand. By using labeled ddNTPs, the lengths of the terminated fragments correspond to specific nucleotide sequences.

Key Steps

1. **Preparation of the DNA Template**
    - The DNA segment to be sequenced is isolated and often amplified through techniques like PCR to obtain sufficient material.
2. **Reaction Components** The sequencing reaction mixture contains:
    - **Template DNA**: The strand to be sequenced.
    - **DNA Polymerase**: Enzyme that synthesizes the new DNA strand.
    - **Primers**: Short, complementary sequences that initiate DNA synthesis.
    - **Deoxynucleotide triphosphates (dNTPs)**: The regular nucleotides (A, T, C, G) used for DNA synthesis.
    - **Dideoxynucleotide triphosphates (ddNTPs)**: Modified nucleotides that lack a hydroxyl group at the 3' carbon. These are incorporated into the growing DNA strand and cause termination of synthesis.
3. **DNA Synthesis and Chain Termination**
    - The reaction is set up in four separate tubes, each containing one type of ddNTP (ddATP, ddTTP, ddCTP, or ddGTP).
    - When a ddNTP is incorporated into the growing DNA strand, it prevents the addition of any further nucleotides, thus terminating the chain.
    - The resulting DNA fragments vary in length depending on where the ddNTP was incorporated.
4. **Fragment Separation**
    - After the reactions are completed, the resulting fragments are separated by size using polyacrylamide gel electrophoresis or capillary electrophoresis.
    - Shorter fragments travel faster through the gel, allowing for the resolution of different lengths.
5. **Detection and Visualization**
    - The fragments are labeled, often using radioactive or fluorescent dyes.
    - The gel is then analyzed (or a capillary sequencer is used) to read the sequence. The order of the bands corresponds to the sequence of the original DNA template.
6. **Reading the Sequence**
    - The sequence can be read from the gel or electrophoresis output, with the colors or positions indicating the order of nucleotides.

Advantages and Disadvantages

Advantages

- **Accuracy**: Sanger sequencing is known for its high accuracy and reliability.
- **Simplicity**: The method is relatively straightforward and easy to perform in a laboratory setting.
- **Established**: Well-established and widely used for sequencing smaller DNA fragments and validating results from next-generation sequencing.

Disadvantages

- **Throughput**: Low throughput compared to next-generation sequencing methods, making it less suitable for large-scale sequencing projects.
- **Cost**: More expensive per base than newer technologies.
- **Length Limitations**: Typically limited to sequences of around 500-1000 bases in length.

Next-generation sequencing (NGS)

While a number of improvements have been made to Sanger sequencing over the years, new high-throughput techniques have also arisen, termed next-generation sequencing (NGS) technologies. NGS is conducted in a massively parallel fashion.

NGS is also known as high-throughput sequencing, is a set of revolutionary DNA sequencing technologies that have transformed genomics and molecular biology since their introduction in the mid-2000s. NGS methods enable rapid and cost-effective sequencing of DNA or RNA, generating massive amounts of data. Here is an in-depth overview of NGS:

Principle:

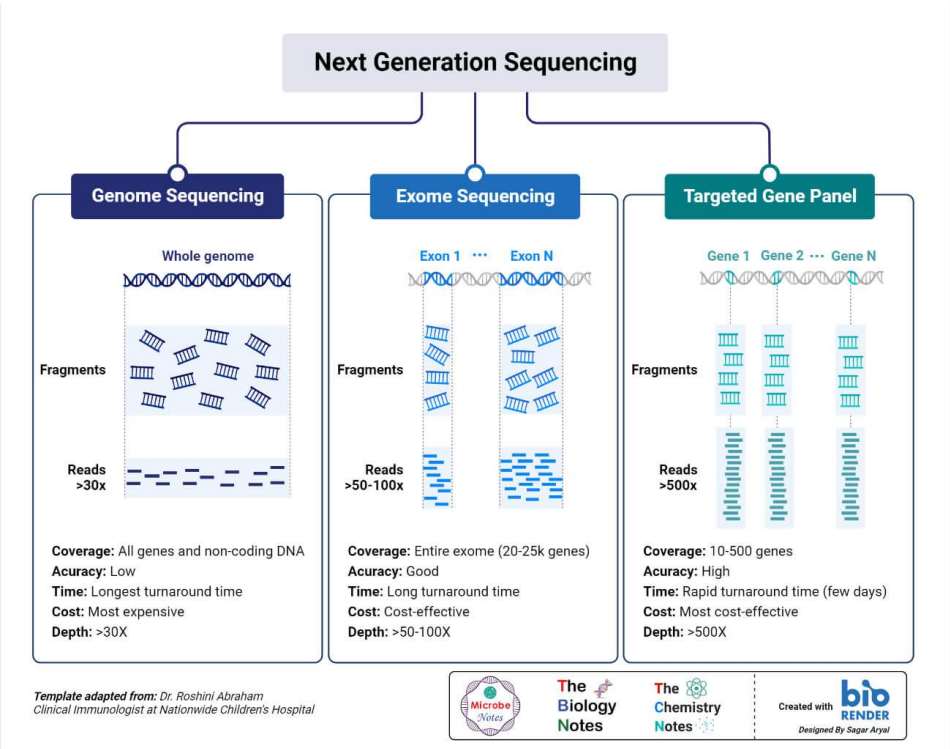
- NGS is based on the parallel sequencing of millions to billions of DNA fragments simultaneously, resulting in a vast amount of sequence data in a single run.

Key Features:

1. **High Throughput**:
    - NGS platforms can generate a large number of sequencing reads in a single run, making it possible to sequence entire genomes, transcriptomes, or targeted regions of interest rapidly.
2. **Parallel Sequencing**:

- NGS methods use solid-phase or solution-phase sequencing, where many individual DNA fragments are sequenced in parallel.
- This parallelism significantly accelerates the sequencing process.

3. **Short Reads**:
    - Most NGS technologies produce short DNA sequences (reads), typically ranging from 50 to 300 base pairs.
    - These short reads are suitable for many applications, including resequencing and transcriptomics.
4. **Applications**:
    - NGS has a wide range of applications, including whole-genome sequencing, exome sequencing, RNA sequencing (RNA-Seq), metagenomics, ChIP-Seq (chromatin immunoprecipitation sequencing), and more.
    - It is used in diverse fields, including genomics, cancer research, personalized medicine, and agriculture.
5. **Library Preparation**:
    - Before sequencing, the DNA or RNA sample is converted into a library of fragments with adapters that allow for sequencing and indexing.
    - Various library preparation methods are available to suit different research objectives.



Next-generation sequencing (NGS) refers to a suite of modern sequencing technologies that allow for the rapid and cost-effective sequencing of entire genomes, exomes, or targeted regions of DNA. NGS has revolutionized genomics by enabling high-throughput sequencing, significantly increasing the speed and scale at which DNA can be analyzed. Here's a comprehensive overview:

Overview of Next-Generation Sequencing (NGS)

Key Features

- **High Throughput**: NGS can sequence millions of fragments simultaneously, generating vast amounts of data in a single run.
- **Cost-Effective**: The cost per base has dramatically decreased, making large-scale sequencing projects feasible.
- **Versatile Applications**: NGS can be applied in genomics, transcriptomics, epigenomics, and metagenomics.

Main NGS Platforms

1. **Illumina Sequencing**
    - **Principle**: Utilizes reversible dye terminators and bridge amplification on a flow cell.
    - **Key Steps**:
        - **Library Preparation**: DNA is fragmented and adapters are ligated.
        - **Cluster Generation**: Amplification of DNA fragments in clusters on the flow cell.
        - **Sequencing by Synthesis**: Fluorescently labeled nucleotides are added, and the incorporated bases are detected at each cycle.
    - **Output**: Short reads (typically 50-300 bp).
2. **Ion Torrent Sequencing**
    - **Principle**: Measures changes in pH as nucleotides are incorporated.
    - **Key Steps**:
        - **Library Preparation**: Similar to Illumina.
        - **Sequencing**: When a nucleotide is added, a proton is released, leading to a change in pH that is detected by a semiconductor sensor.
    - **Output**: Short to medium reads (up to 400 bp).

3. **Pacific Biosciences (PacBio) Sequencing**
    - **Principle**: Single-molecule real-time (SMRT) sequencing that allows for long reads.
    - **Key Steps**:
        - **Library Preparation**: DNA is circularized.
        - **Sequencing**: DNA polymerase incorporates labeled nucleotides, and the emitted signal is recorded in real time.
    - **Output**: Long reads (up to 10,000 bp or more).
4. **Oxford Nanopore Sequencing**
    - **Principle**: Measures changes in ionic current as DNA strands pass through a nanopore.
    - **Key Steps**:
        - **Library Preparation**: DNA is prepared with specific adapters.
        - **Sequencing**: The DNA molecule is pulled through a nanopore, and the sequence is determined based on the disruption of the current.
    - **Output**: Very long reads (up to millions of bases).

Key Steps in NGS Workflow

1. **Sample Preparation**
    - DNA or RNA is extracted and prepared for sequencing by fragmenting and ligating adapters.
2. **Library Construction**
    - Adapters are added to both ends of the DNA fragments, enabling sequencing and allowing for indexing.
3. **Amplification**
    - Library fragments are amplified using PCR to generate sufficient quantities for sequencing.
4. **Sequencing**
    - The prepared library is sequenced using one of the NGS platforms.
5. **Data Analysis**

- Raw sequence data is processed to remove low-quality reads, align sequences to a reference genome, and identify variants.

Applications of NGS

- **Whole Genome Sequencing**: Comprehensive analysis of entire genomes.
- **Exome Sequencing**: Targeted sequencing of coding regions of the genome.
- **RNA-Seq**: Analysis of transcriptomes to study gene expression.
- **Targeted Sequencing**: Sequencing specific regions associated with diseases.
- **Metagenomics**: Analysis of genetic material from environmental samples to study microbial diversity.

Advantages and Disadvantages

Advantages

- **Speed**: Rapid sequencing capabilities allow for large datasets to be generated in days.
- **Scalability**: Suitable for small targeted studies to large population-scale projects.
- **Depth of Coverage**: High coverage enables the detection of rare variants.

Disadvantages

- **Data Management**: Generates large volumes of data that require substantial storage and computational resources.
- **Error Rates**: Certain NGS technologies have higher error rates in homopolymeric regions.
- **Complexity of Analysis**: Requires bioinformatics expertise for data interpretation and variant calling.

Nucleic acid hybridization and nucleic acid probes

Nucleic acid hybridization is a foundational molecular biology technique that plays a crucial role in various aspects of genetics, genomics, diagnostics, and molecular biology research. Here are more detailed insights into nucleic acid hybridization:

Principle:

- Nucleic acid hybridization is based on the principle of complementary base pairing. DNA or RNA strands with complementary sequences will bind together to form stable double-stranded molecules due to the formation of hydrogen bonds between complementary bases.

Key Concepts:

1. **Complementary Base Pairing**:
    - In DNA, the four nitrogenous bases are adenine (A), thymine (T), cytosine (C), and guanine (G). In RNA, thymine is replaced by uracil (U).
    - Complementary base pairing involves A pairing with T (or U in RNA) and C pairing with G. These pairings allow for hydrogen bonding, which stabilizes the double-stranded structure.

Common Techniques and Applications:

1. **Southern Blotting**:
    - Southern blotting is primarily used to detect specific DNA sequences in a complex mixture.
    - The DNA sample is digested with restriction enzymes, separated by gel electrophoresis, and transferred to a membrane (usually nitrocellulose or nylon).
    - A labeled DNA probe, a single-stranded DNA sequence complementary to the target, is used to hybridize with the immobilized DNA on the membrane. The probe binds to the complementary target sequence on the membrane, allowing for detection.
    - This technique is useful for applications like genetic fingerprinting and gene mapping.

Applications of Southern Blotting:

- **Gene Mapping and Localization**:
    - Southern blotting is used to map and locate specific genes on chromosomes. By hybridizing a DNA probe with genomic DNA fragments, researchers can identify the presence and location of target genes.
- **Restriction Fragment Length Polymorphism (RFLP) Analysis**:
    - Southern blotting can be used to detect RFLPs, which are variations in the lengths of DNA fragments generated by restriction enzymes.
    - RFLP analysis is used in genetic fingerprinting, genetic linkage analysis, and forensic genetics.
- **Identification of Copy Number Variations (CNVs)**:

- Southern blotting can help identify copy number variations in the genome. By comparing the intensity of bands in different samples, researchers can determine the number of copies of a specific DNA sequence.

- **Verification of DNA Cloning and Recombinant DNA**:
  - Southern blotting can confirm the successful integration of a recombinant DNA construct into the genome or the presence of specific DNA fragments in cloned vectors.

- **Detection of Epigenetic Modifications**:
  - Southern blotting can be used to study DNA methylation patterns, a common epigenetic modification associated with gene regulation and silencing.

- **Screening for Specific Mutations**:
  - It can be applied to identify specific mutations or polymorphisms in genes responsible for genetic disorders, which is valuable for diagnostic purposes.

2. **Northern Blotting**:

- Northern blotting is similar to Southern blotting but is used to detect specific RNA sequences, typically for gene expression analysis.
- It involves the separation and transfer of RNA molecules to a membrane, followed by hybridization with a labeled RNA or DNA probe.

Applications of Northern Blotting:

- **Gene Expression Analysis**:
  - Northern blotting is primarily used to assess gene expression levels. It allows researchers to determine which genes are actively transcribed and at what level in a given tissue or under specific conditions.

- **Identification of RNA Splice Variants**:
  - Researchers can use Northern blotting to detect and distinguish different splice variants of a gene by targeting specific RNA sequences.

- **Validation of Gene Expression Data**:
  - Northern blotting is used to validate the results obtained from other gene expression analysis methods like RT-PCR or microarrays.

- **Comparative Gene Expression Studies**:
    - Researchers can compare gene expression profiles in different tissues, cell types, or under various experimental conditions to gain insights into gene regulation.
- **Detection of Non-Coding RNAs**:
    - Northern blotting can also be used to study non-coding RNAs, such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs).

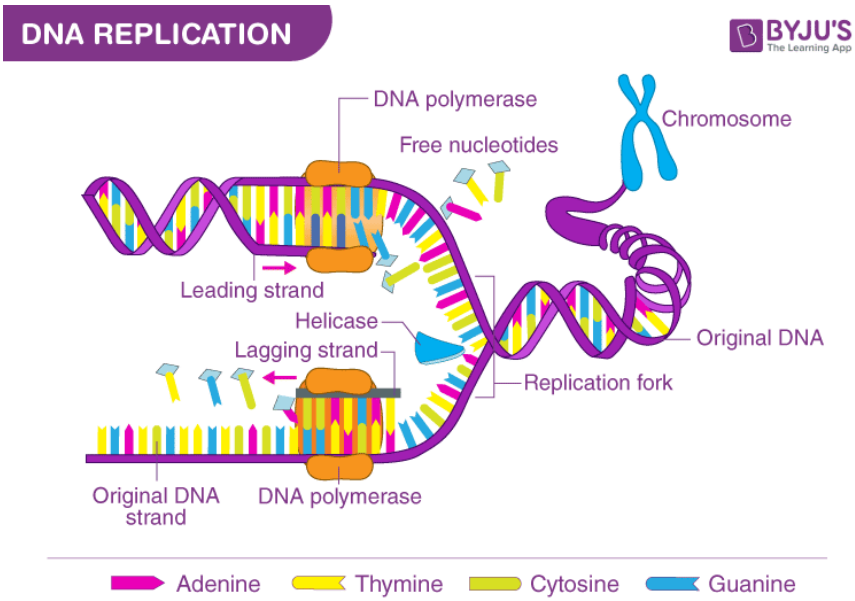
3. **In Situ Hybridization**:
    - In situ hybridization allows the localization of specific DNA or RNA sequences within intact cells or tissue sections.
    - Fluorescent or radioactive probes are used to hybridize with target sequences in situ, and the location of the hybridization is visualized under a microscope.
    - This technique is crucial in developmental biology, histopathology, and identifying gene expression patterns in tissues.
4. **DNA Microarrays**:
    - DNA microarrays, also known as gene chips, are used for high-throughput gene expression analysis.
    - Microarray chips contain thousands of known DNA sequences (probes) that represent specific genes or genomic regions.
    - Labeled cDNA or RNA from a sample is hybridized to these probes. The intensity of the signal from the hybridization reveals the relative abundance of specific transcripts in the sample.
    - Microarrays are valuable for studying gene expression, identifying biomarkers, and understanding the effects of genetic variations.
5. **Polymerase Chain Reaction (PCR)**:
    - In PCR, hybridization plays a critical role during the annealing step.
    - Two short DNA primers are designed to be complementary to the sequences flanking the target DNA region.
    - During the annealing step, these primers hybridize to the complementary target sequences to initiate DNA amplification.

Challenges and Considerations:

- Hybridization efficiency is influenced by factors such as temperature, salt concentration, and probe design.
- Specificity of hybridization depends on the degree of complementarity between the probe and target sequence. Stringency conditions can be adjusted to control mismatch tolerance during hybridization.

DNA REPLICATION AND REPAIR

DNA replication and repair are fundamental processes in molecular biology that are essential for maintaining the integrity and stability of genetic information. DNA replication is the process by which an exact copy of the DNA molecule is synthesized. It occurs before cell division to ensure that each daughter cell receives a complete and identical set of genetic information.



1. **Key Steps**:
    - **Initiation**: Replication begins at specific sites on the DNA molecule called origins of replication. Enzymes, such as helicase, unwind and open the double helix. The origin is characterized by an origin recognition complex (ORC) in eukaryotes.
    - **Elongation**: DNA polymerases synthesize a new DNA strand by adding complementary nucleotides to the template strand. DNA replication is semi-conservative, meaning that each new DNA molecule consists of one old (template) strand and one new strand. DNA polymerase III is the main polymerase in prokaryotes, while eukaryotes use polymerase δ and ε. DNA polymerases add deoxyribonucleotides to the growing strand in a 5' to 3' direction. Leading and lagging strands are synthesized simultaneously: the

leading strand is synthesized continuously, while the lagging strand is synthesized in short fragments (Okazaki fragments).

- **Termination**: Replication proceeds bidirectionally from the origin, creating two replication forks. It continues until it reaches termination sequences, where replication is stopped. Termination proteins in prokaryotes, like Tus, prevent further replication. Eukaryotic termination is less well-defined but often involves meeting of the replication forks.

2. **Enzymes and Proteins Involved**:
    - DNA helicase: Unwinds the double helix.
    - DNA polymerase: Adds nucleotides to the growing DNA strand.
    - Primase: Synthesizes RNA primers for DNA polymerase to start replication.
    - DNA ligase: Seals nicks and joins Okazaki fragments on the lagging strand.
    - Topoisomerases: Relieve supercoiling of DNA ahead of the replication fork.
    - Single-strand binding proteins: Stabilize single-stranded DNA.
3. **Errors and Proofreading**:
    - DNA polymerases have proofreading capabilities and can correct mismatched base pairs during replication.
    - DNA repair mechanisms also help fix errors that escape proofreading.

DNA Repair:

1. **Overview**:
    - DNA is susceptible to various forms of damage, including chemical modifications, UV radiation, and spontaneous mutations. DNA repair mechanisms ensure the correction of these damages to maintain genomic integrity.
2. **Types of DNA Repair Mechanisms**:
    - **Base Excision Repair (BER)**: Corrects individual damaged or mismatched bases. Enzymes recognize the damaged base, remove it, and replace it with the correct base.
    - **Nucleotide Excision Repair (NER)**: Repairs bulky lesions, such as those caused by UV radiation or chemical agents. It involves the removal and replacement of a stretch of nucleotides containing the damage.

- **Mismatch Repair (MMR)**: Corrects mismatched base pairs that were not repaired during replication. It distinguishes between the newly synthesized and the template DNA strand to repair the mismatch.
- **Double-Strand Break Repair (DSBR)**: Addresses breaks in both strands of the DNA molecule, which can be highly deleterious. It includes homologous recombination and non-homologous end-joining mechanisms.
- **Direct Repair**: Some DNA lesions can be directly reversed by specific repair enzymes. For example, the enzyme photolyase can reverse UV-induced damage to DNA.

3. **Repair Enzymes and Pathways**:
    - Repair pathways involve a variety of proteins and enzymes, such as DNA glycosylases, endonucleases, and DNA ligases.
    - The choice of repair pathway depends on the type of DNA damage and the cell cycle phase.
4. **Importance**:
    - DNA repair is crucial in preventing mutations and genomic instability, which can lead to cancer and other genetic diseases.
    - Repair mechanisms also play a role in maintaining the functionality of the genome, ensuring proper gene expression, and protecting against environmental damage.

In summary, DNA replication ensures the faithful duplication of genetic information during cell division, while DNA repair mechanisms safeguard the genome from damage and mutations, allowing for the preservation of genetic integrity. These processes are vital for the proper functioning and long-term survival of cells and organisms.