Introduction to Bioinformatics

DNA Structure

Ali Sharifi-Zarchi   Somayyeh Koohi

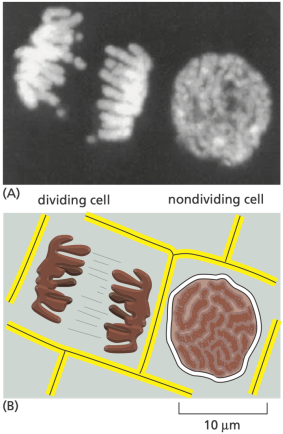
CE Department

Sharif University of Technology

October 13, 2025

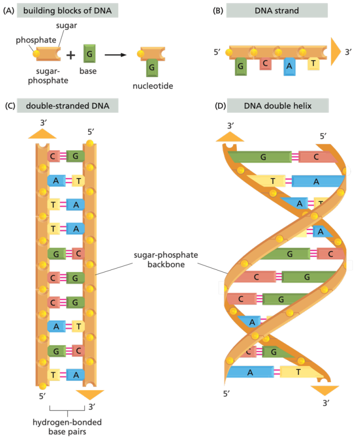
Presentation Overview

Nucleus

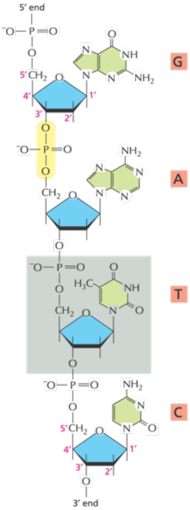


The building blocks of DNA

DNA consists of two long polynucleotide chains held together by hydrogen bonds between bases. Each nucleotide consists of a nitrogen-containing base, a deoxyribose sugar, and a phosphate group.

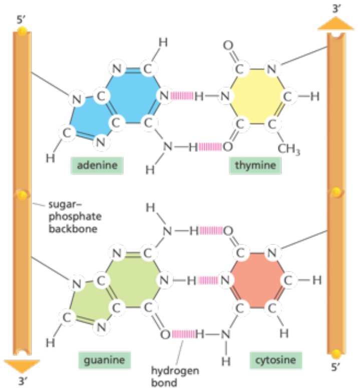


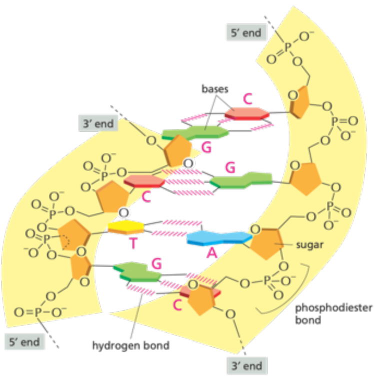
Base



- The specificity of each nucleotide is determined by its organic base.
- The sugar–phosphate groups form the exterior backbone of the strand, while the bases are positioned inside the helix.
- Among the four bases, adenine and guanine are purines with two rings, whereas thymine and cytosine are pyrimidines with one ring.
- Each strand has a chemical direction ($5' \rightarrow 3'$), and the two strands run antiparallel.

Sugar and Phosphate

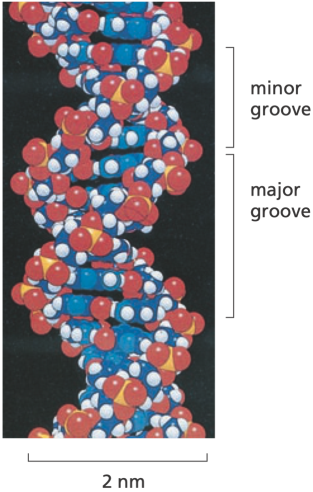




Sugar
Nucleic acids contain two types of pentoses (5-carbon sugars).
DNA contains 2′-deoxy-D-ribose, while RNA contains D-ribose.

Phosphate
Phosphoric acid (H_3PO_4) gives nucleic acids an overall negative charge.

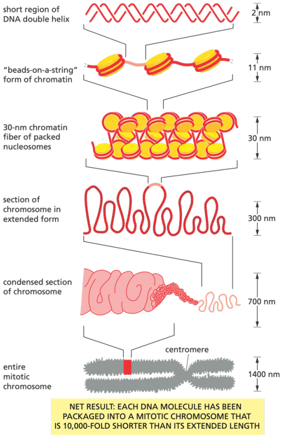
DNA structure



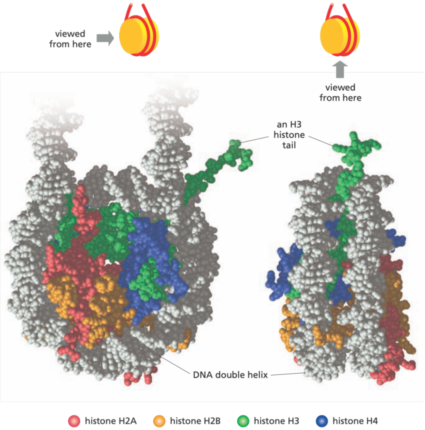
- Two DNA chains coil around the same axis, forming a right-handed helix.
- Alternating deoxyribose and phosphate groups form the hydrophilic backbone, which is positioned outward toward the surrounding water.
- Nucleotides are covalently linked via phosphodiester bonds: the 3′-hydroxyl of one nucleotide connects to the 5′-phosphate of the next
- The offset pairing of the two strands creates a major groove and a minor groove on the surface of the duplex.

Super-helix Structure of Chromosomes

Because of DNA's length, the DNA molecules must be carefully packed and preserved.



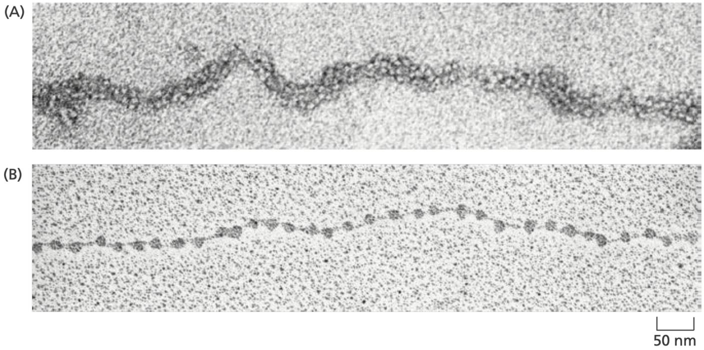
Histone Proteins



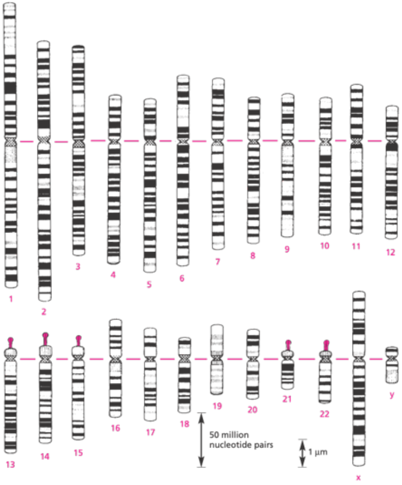
- Histones are specialized, positively charged proteins and rich in lysine and arginine.
- Responsible for the first level of DNA compaction, forming nucleosomes.

Nucleosomes

A nucleosome core particle consists of 147 nucleotide pairs of DNA tightly wrapped around a core of eight histone molecules (two each of H2A, H2B, H3, and H4), with the positive charges on the histones ensuring tight binding to the negative sugar–phosphate backbone of the DNA.

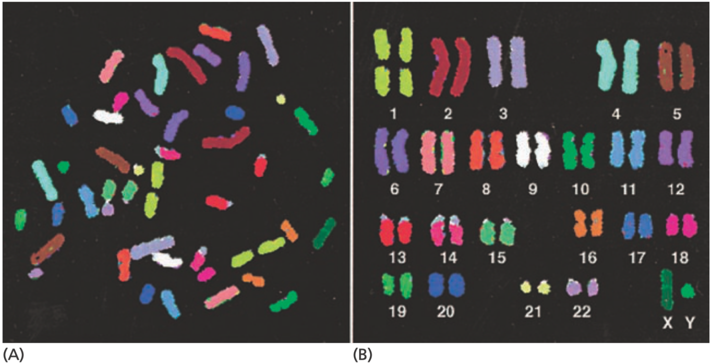


Chromosomes



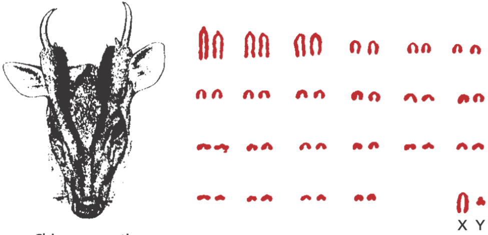
- DNA is wrapped and folded by specialized proteins (forming chromatin) to package the cell's immense DNA length compactly inside the nucleus.
- Human cells have 46 chromosomes (23 pairs), including homologous pairs and sex chromosomes.
- Chromosomes contain specialized DNA sequences—replication origins, centromeres, and telomeres to ensure proper replication and segregation during cell division.
- Chromosomes can be distinguished by size, banding patterns, or fluorescent painting techniques.

Chromosomes



Close Animals can have different Chromosome Numbers

Even closely related species can have vastly different chromosome numbers and sizes, showing that gene number, chromosome number, and genome size are not directly correlated.



Chinese muntjac

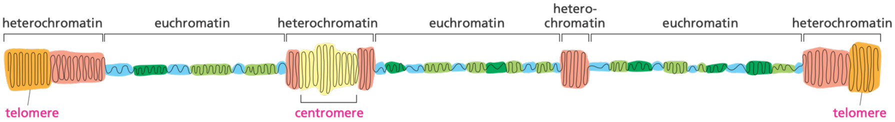
Indian muntjac

Genes

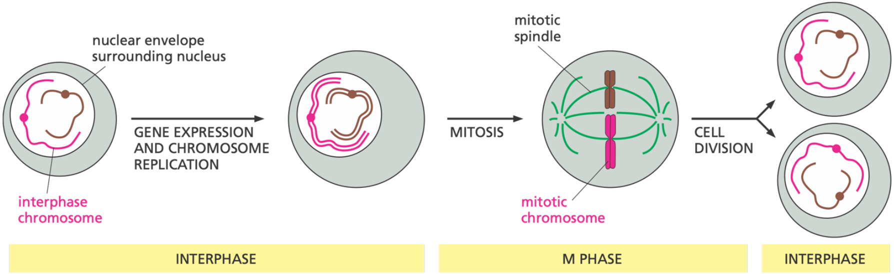
- Genes are segments of DNA carrying hereditary information, encoding proteins or functional RNAs.
- The genome is the complete set of genetic material. It varies widely in size and organization from compact, gene-rich bacterial genomes (~500 genes) to the human genome (~24,000 genes) with abundant, often conserved and functional, non-coding DNA.

Heterochromatin and Euchromatin

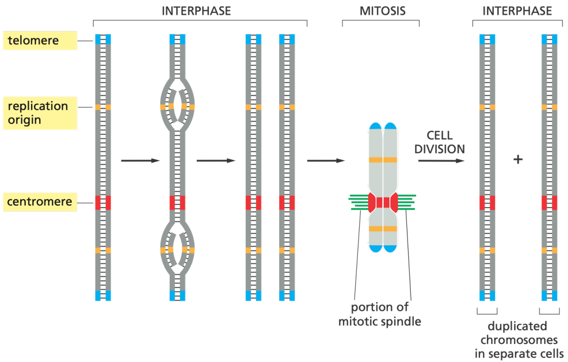
- Chromatin is composed of DNA and histone proteins. During interphase it is unevenly packed and appears in two main forms.
- Heterochromatin is highly condensed and gene-silent (about 10% of interphase chromatin), mostly at centromeres and telomeres.
- Euchromatin is less condensed and open, containing most active genes. In general, active genes reside in euchromatin and silent genes in heterochromatin; a cell typically expresses only about half of its genes.
- Chromatin structure is regulated by histone modifications and associated proteins.



Cell Division



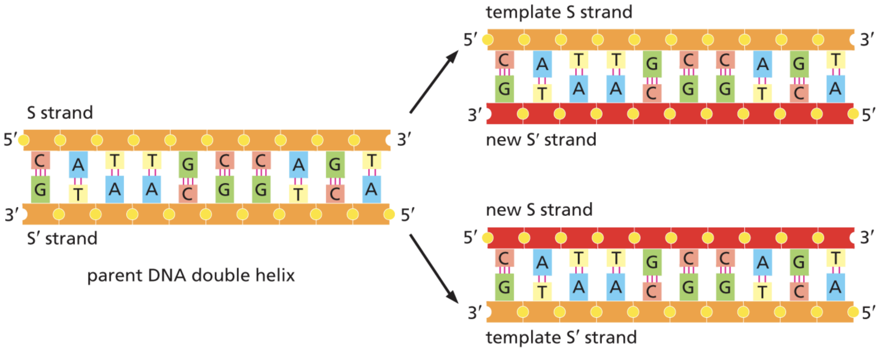
Cell Division Elements



The Fundamental Need for DNA Replication

- A cell must accurately copy its vast quantity of genetic information to survive and proliferate.
- The fundamental process, **DNA Replication**, must occur before a cell can divide to produce two genetically identical daughter cells.
- Cells also constantly inspect and repair their genetic material, as DNA is subjected to unavoidable damage.
- **Mutations** occur randomly and are often detrimental, making continuous repair systems essential for life.

The Fundamental Need for DNA Replication



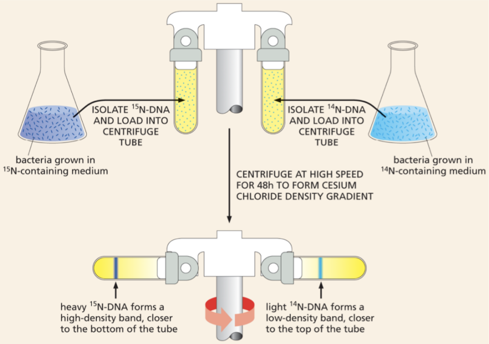
Mechanism: The Template Principle

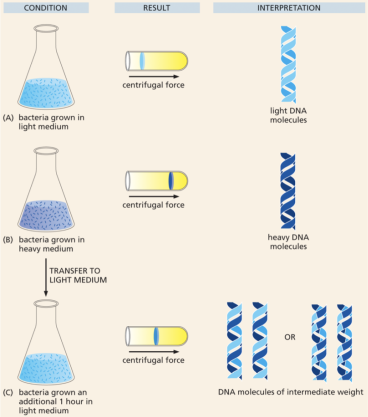
- Each strand of the DNA double helix is complementary to its partner strand.
- This structure allows each parent strand to serve as a **template** (or mold) for the synthesis of a new partner strand.
- The genetic information is accurately copied by separating the two strands, and each separated strand directs the production of a new complementary strand.
- This results in a copy of the original DNA molecule that is nearly identical in sequence.

The Semiconservative Model

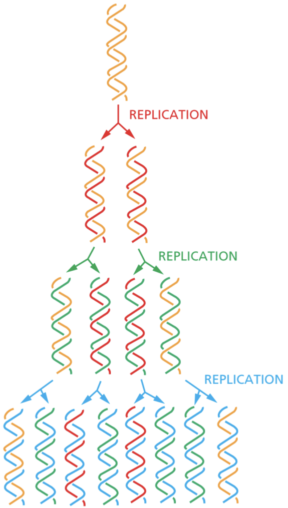
- DNA replication is defined as **semiconservative**.
- This means each daughter DNA double helix is composed of one conserved (old) parent strand and one newly synthesized strand.
- This model, proposed by Watson and Crick, was definitively proven by the **Meselson-Stahl experiment**.
- Their experiment used heavy ($\mathbf{^{15}N}$) and light ($\mathbf{^{14}N}$) nitrogen isotopes to show that replicated DNA molecules are hybrids.

The Semiconservative Model





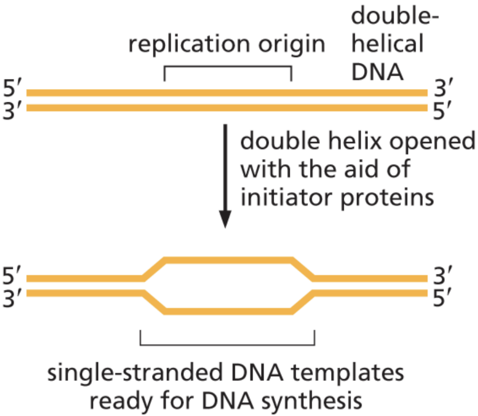
The Semiconservative Model



Initiation and Replication Forks

- DNA synthesis begins at specific sites called **replication origins**.
- **Initiator proteins** bind here and locally pry the two DNA strands apart, breaking hydrogen bonds.
- The opening of the double helix creates a Y-shaped junction known as a **replication fork**.
- Two replication forks form at each origin and move away in opposite directions, a process called **bidirectional replication**.

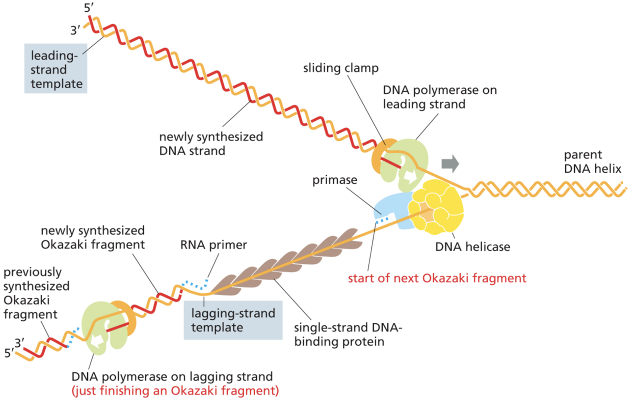
Initiation and Replication Forks



DNA Polymerase: The Core Engine

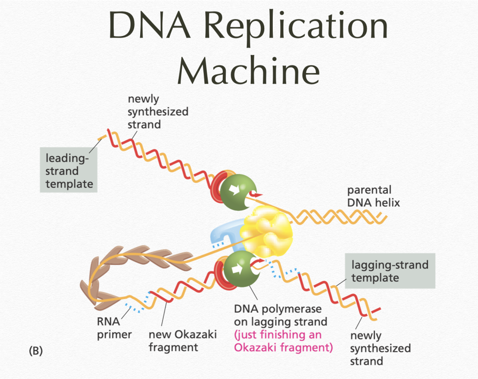
- The enzyme at the heart of the replication machine is **DNA polymerase**.
- This enzyme catalyzes the addition of new nucleotides to the **3’ end** of a growing DNA strand.
- Complementary base-pairing with the template strand determines which of the four bases (A, G, T, or C) is added.
- The energy for this polymerization is provided by the incoming deoxyribonucleoside triphosphate itself.

DNA Polymerase: The Core Engine



(A)

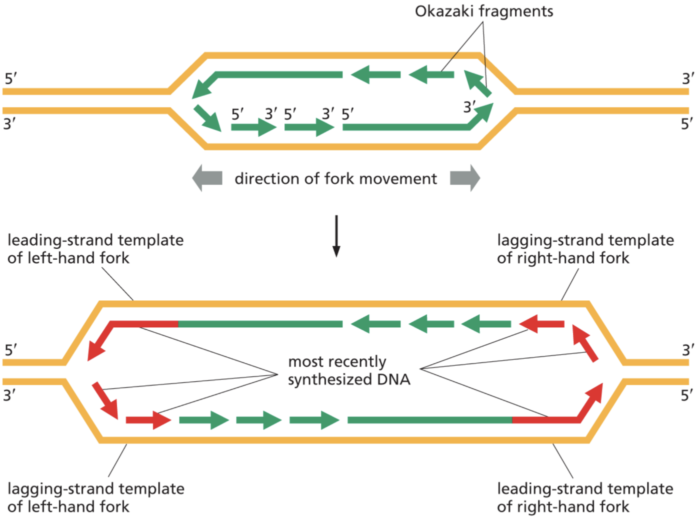
DNA Polymerase: The Core Engine



Asymmetrical Synthesis: Leading and Lagging Strands

- All known DNA polymerases can only synthesize DNA in the **5’-to-3’ direction**.
- However, the two parent strands at the replication fork run in opposite (**antiparallel**) directions.
- This results in an **asymmetrical replication fork**.
- One new strand, the **leading strand**, is synthesized continuously in the direction of the fork.

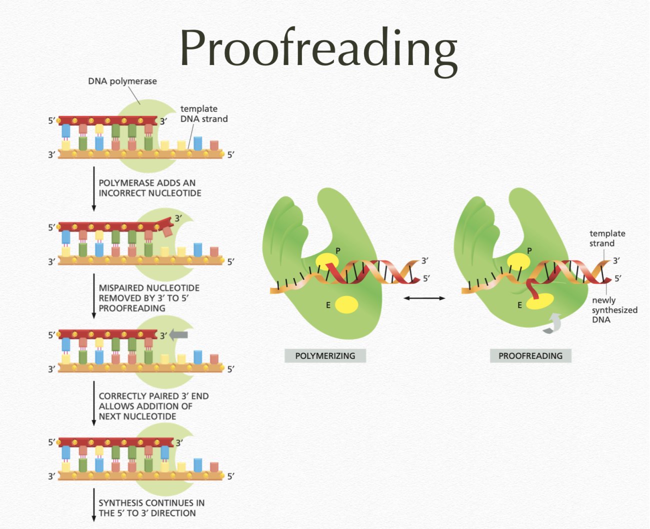
Asymmetrical Synthesis: Leading and Lagging Strands



Proofreading for High Fidelity

- DNA polymerase achieves extraordinary accuracy, making only about one error per $\mathbf{10^7}$ nucleotide pairs copied.
- This is achieved through a mechanism called **proofreading**.
- Before adding the next base, the polymerase checks the previously added nucleotide.
- If an incorrect base is detected, the enzyme pauses to **clip it off and then tries again**.

Proofreading for High Fidelity



The Necessity of DNA Repair

- To survive and reproduce, individual organisms must maintain **genetic stability**.
- DNA is constantly damaged by chemical reactions within the cell and by environmental factors.
- The **DNA Repair** process continually scans the genome and immediately corrects most damage.
- Failure in repair mechanisms leads to accumulating DNA damage, resulting in diseases like **cancer**.

The Basic Three-Step Repair Pathway

- Most DNA damage is repaired using the **undamaged strand as a template**.
- Step 1 (**Excision**): Nucleases recognize the damage and cut out the damaged segment.
- Step 2 (**Resynthesis**): A repair DNA polymerase fills the resulting gap using the complementary strand as a guide.
- Step 3 (**Ligation**): DNA ligase seals the final break in the sugar-phosphate backbone.

Mismatch Repair (MMR)

- **Mismatch Repair (MMR)** is a backup system that corrects copying mistakes that escape DNA polymerase proofreading.
- It improves replication accuracy to an incredibly low rate of **one mistake per 10^9 nucleotides** copied.
- MMR proteins must distinguish between the correct parent strand and the newly synthesized strand containing the error.
- Defects in MMR genes strongly predispose individuals to certain types of **cancer, such as colon cancer**.

Telomeres and Chromosome Ends

- **Telomeres** are repetitive nucleotide sequences that cap the linear ends of eukaryotic chromosomes.
- They counteract the tendency of chromosomes to shorten with every round of replication.
- The enzyme **telomerase** elongates these ends by synthesizing the repetitive sequences.
- Telomeres and telomerase are essential for preventing the loss of genetic information at chromosome ends.

References

Alberts, B. *et al.*
Essential Cell Biology.