

Introduction to Bioinformatics

DNA Structure

Ali Sharifi-Zarchi Somayyeh Koohi

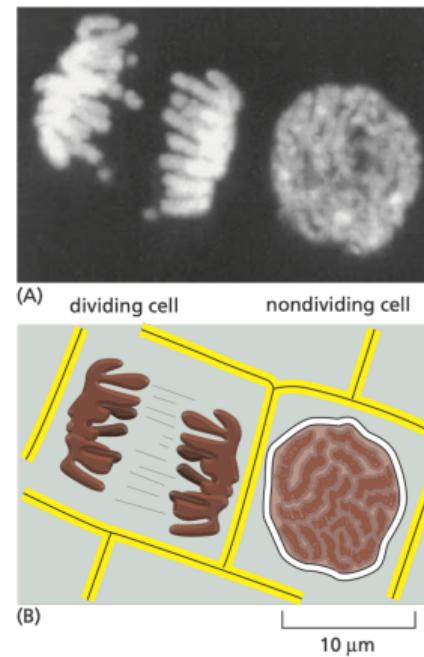
CE Department
Sharif University of Technology

October 18, 2025

Presentation Overview

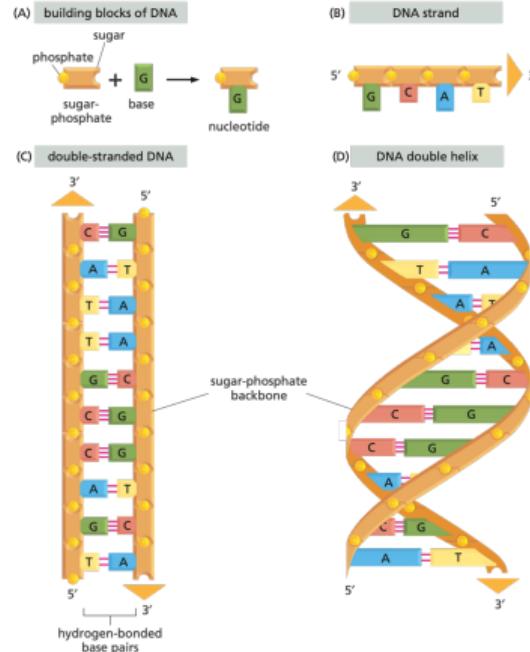
- ① DNA and Chromosomes
- ② DNA Replication and Repair
- ③ Mitosis and Meiosis
- ④ Recommended Videos
- ⑤ Referencing

Nucleus

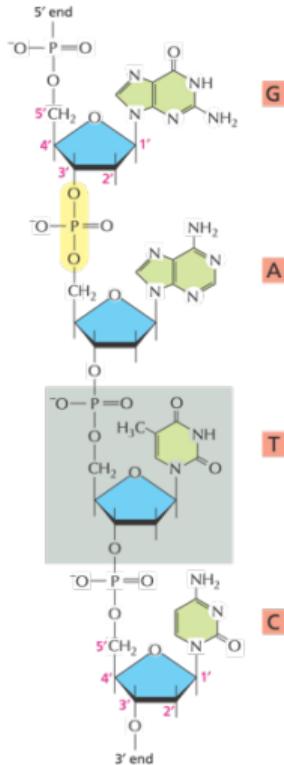


The building blocks of DNA

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.

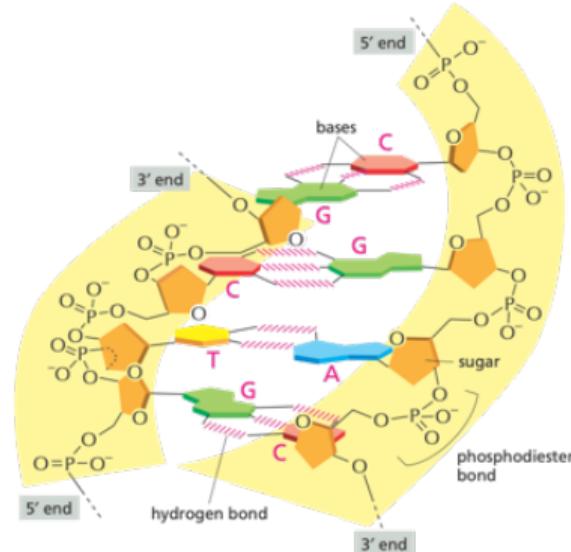
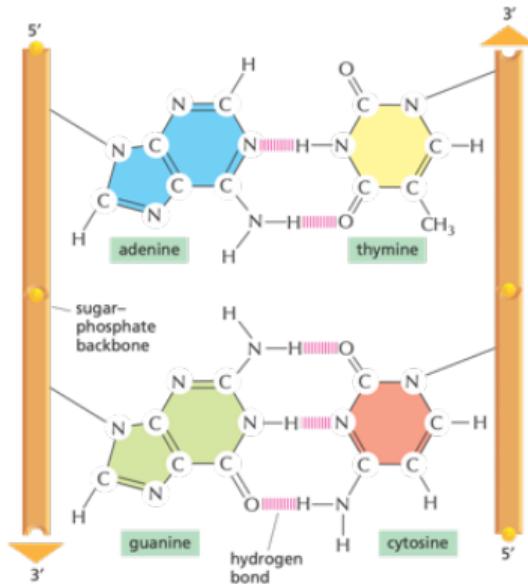


Base



- 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' → 3'), and the two strands run antiparallel.

Sugar and Phosphate



Sugar

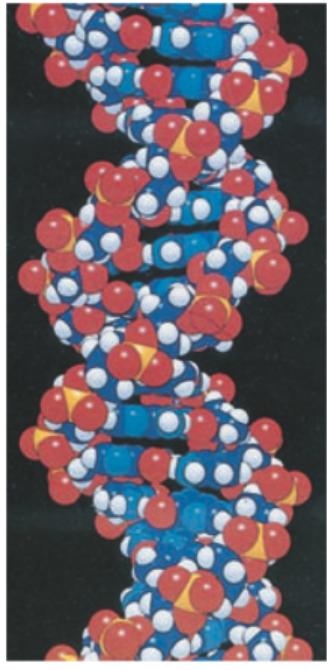
Nucleic acids contain two types of pentoses (5-carbon sugars).

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

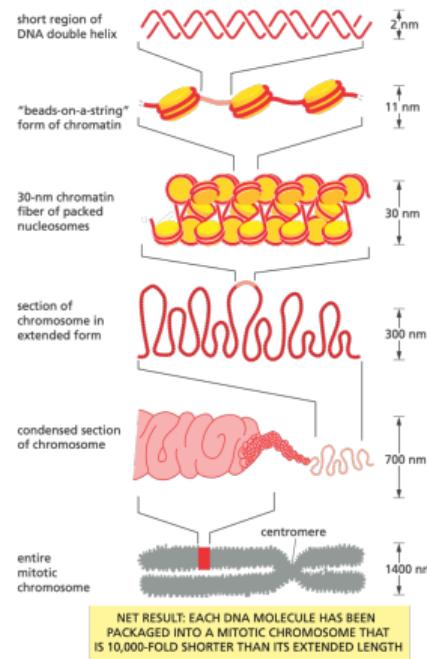


2 nm

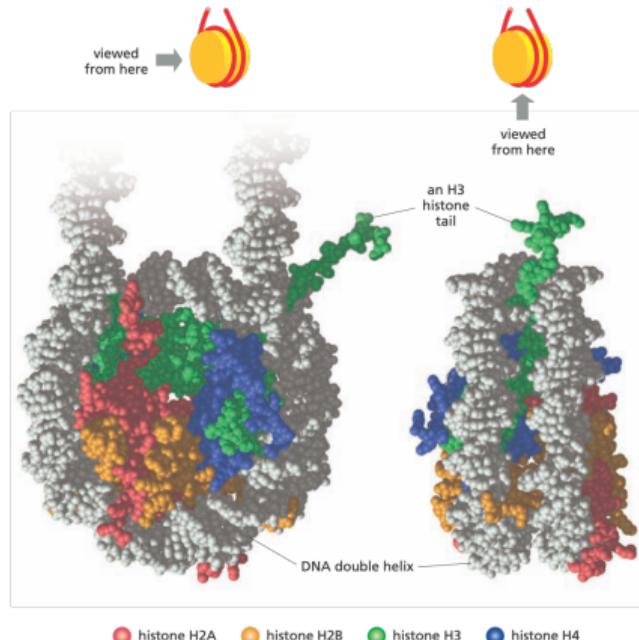
- Two DNA chains coil around the same axis, forming a right-handed helix.
- Alternating deoxyribose and phosphate groups form the hydrophilic backbone, which is positioned outward toward the surrounding water.
- Nucleotides are covalently linked via phosphodiester bonds: the 3'-hydroxyl of one nucleotide 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.



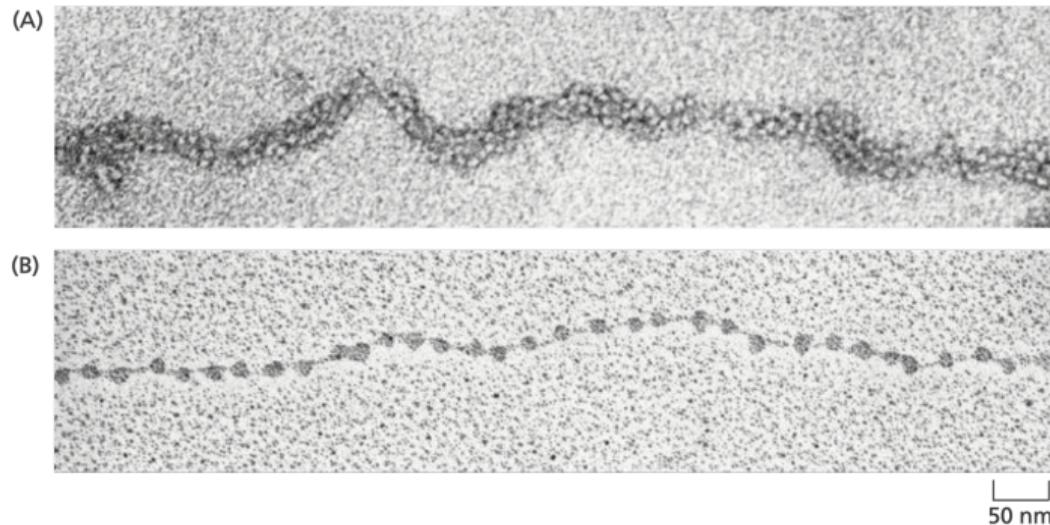
Histone Proteins



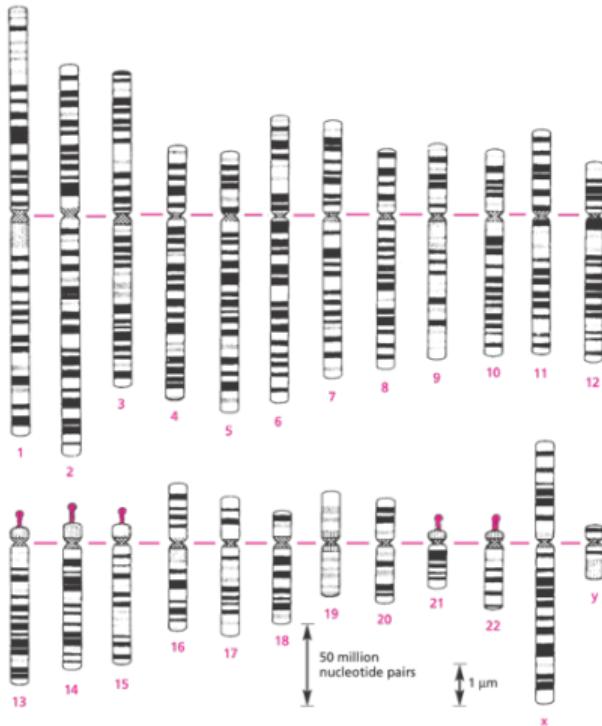
- Histones are specialized, positively charged proteins and rich in lysine and arginine.
- Responsible for the first level of DNA compaction, forming nucleosomes.

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.

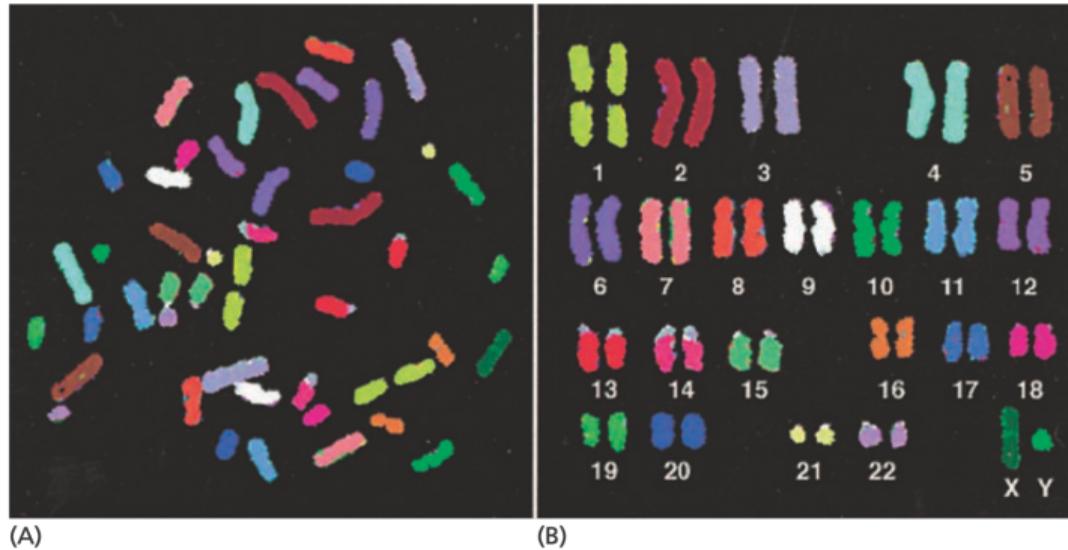


Chromosomes



- 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



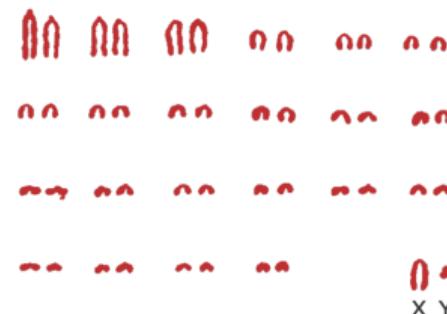
Chromosomes

Close Animals can have different Chromosome Numbers

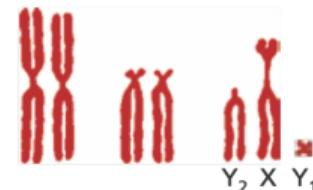
Even closely related species can have vastly different chromosome numbers and sizes, showing that gene number, chromosome number, and genome size are not directly correlated.



Chinese muntjac

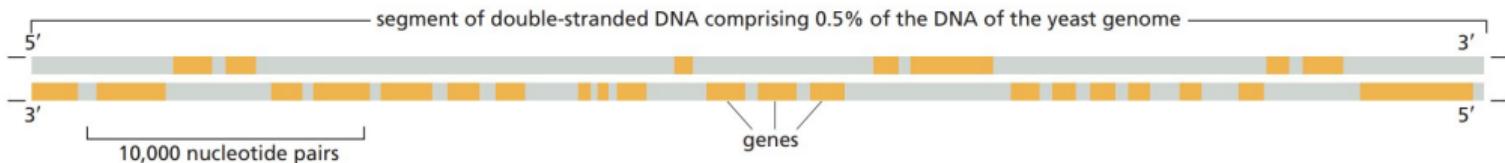


Indian muntjac



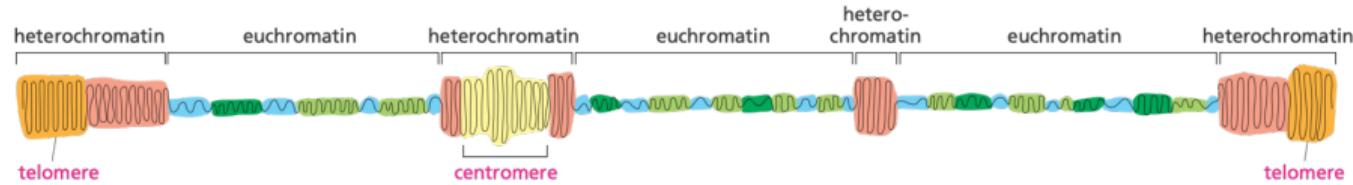
Genes

- 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

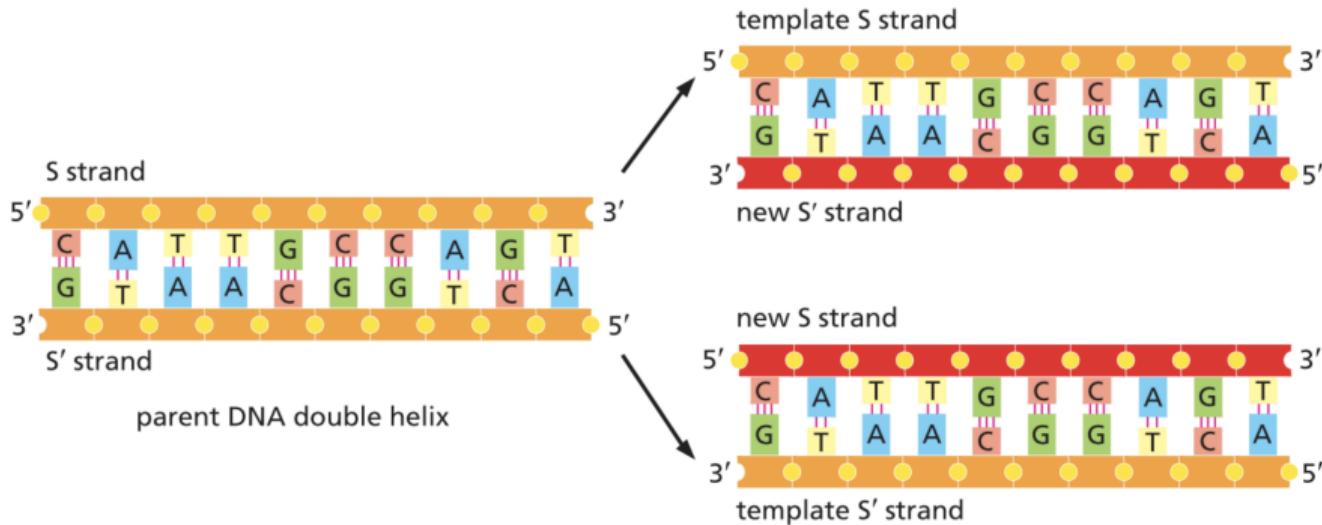
- Chromatin is composed of DNA and histone proteins. During interphase it is unevenly packed and appears in two main forms.
- Heterochromatin is highly condensed and gene-silent (about 10% of interphase chromatin), mostly at centromeres and telomeres.
- Euchromatin is less condensed and open, containing 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.



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



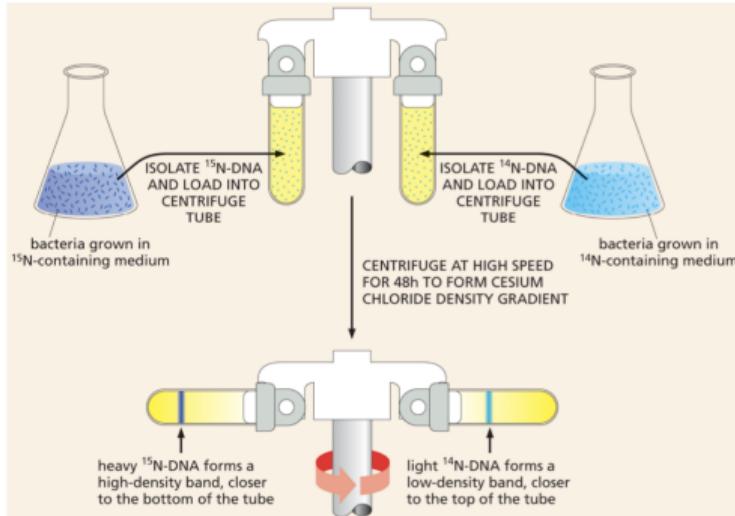
Mechanism: The Template Principle

- 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

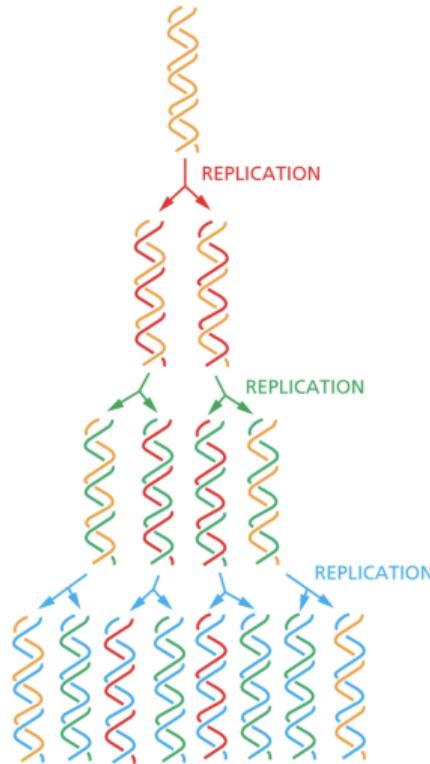
- DNA replication is defined as **semiconservative**.
- This means each daughter DNA double helix is composed of one conserved (old) parent strand and one newly synthesized strand.
- This model, proposed by Watson and Crick, was definitively proven by the **Meselson-Stahl experiment**.
- Their experiment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

The Semiconservative Model



CONDITION	RESULT	INTERPRETATION
(A) bacteria grown in light medium	centrifugal force	light DNA molecules
(B) bacteria grown in heavy medium	centrifugal force	heavy DNA molecules
(C) bacteria grown an additional 1 hour in light medium	TRANSFER TO LIGHT MEDIUM centrifugal force	OR DNA molecules of intermediate weight

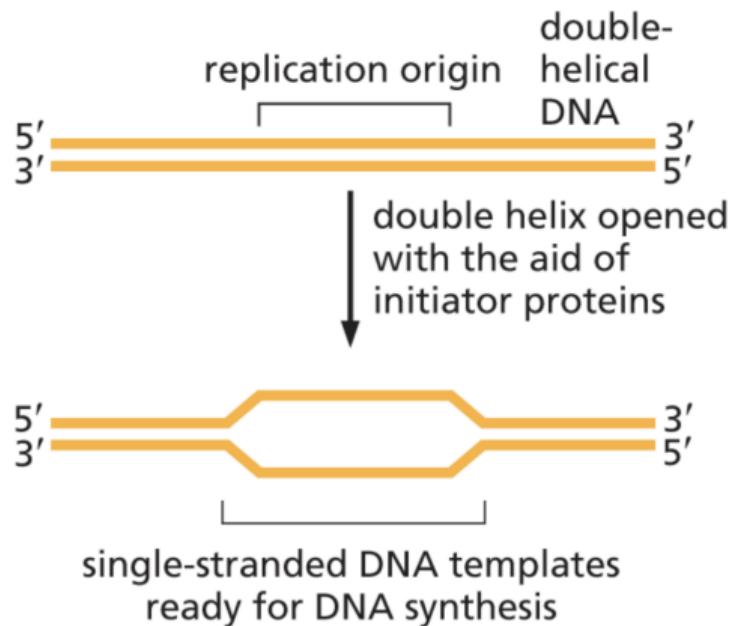
The Semiconservative Model



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**.

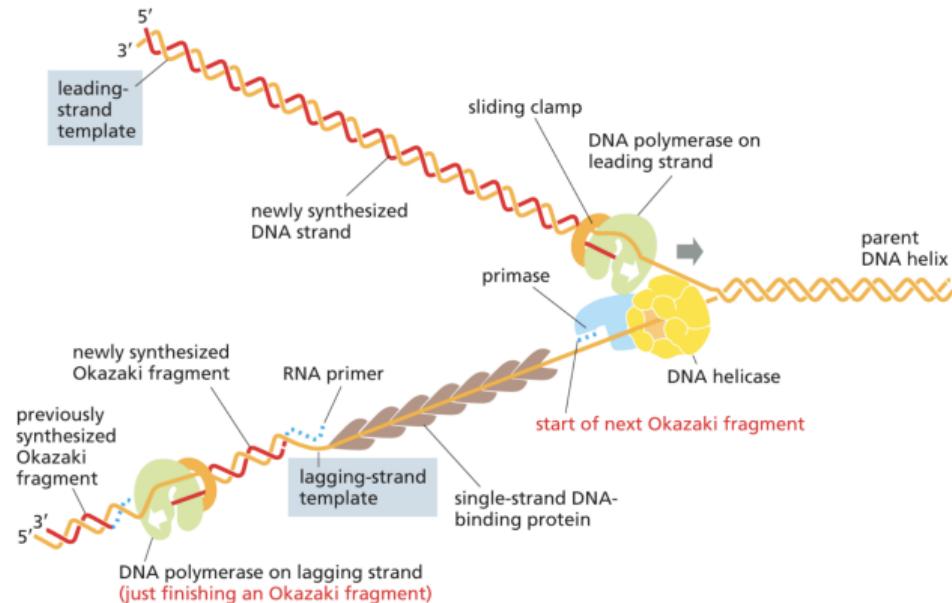
Initiation and Replication Forks



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.

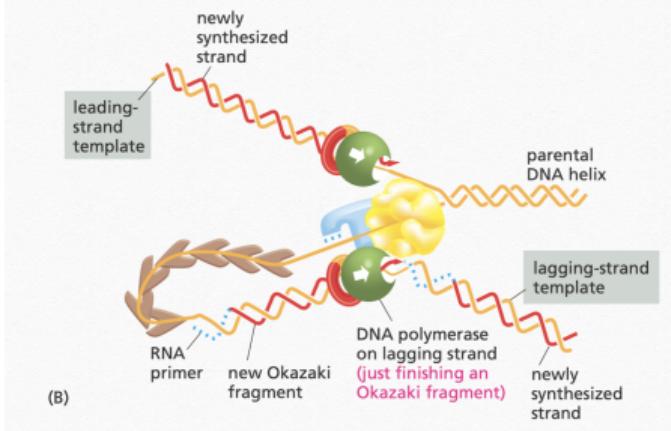
DNA Polymerase: The Core Engine



(A)

DNA Polymerase: The Core Engine

DNA Replication Machine

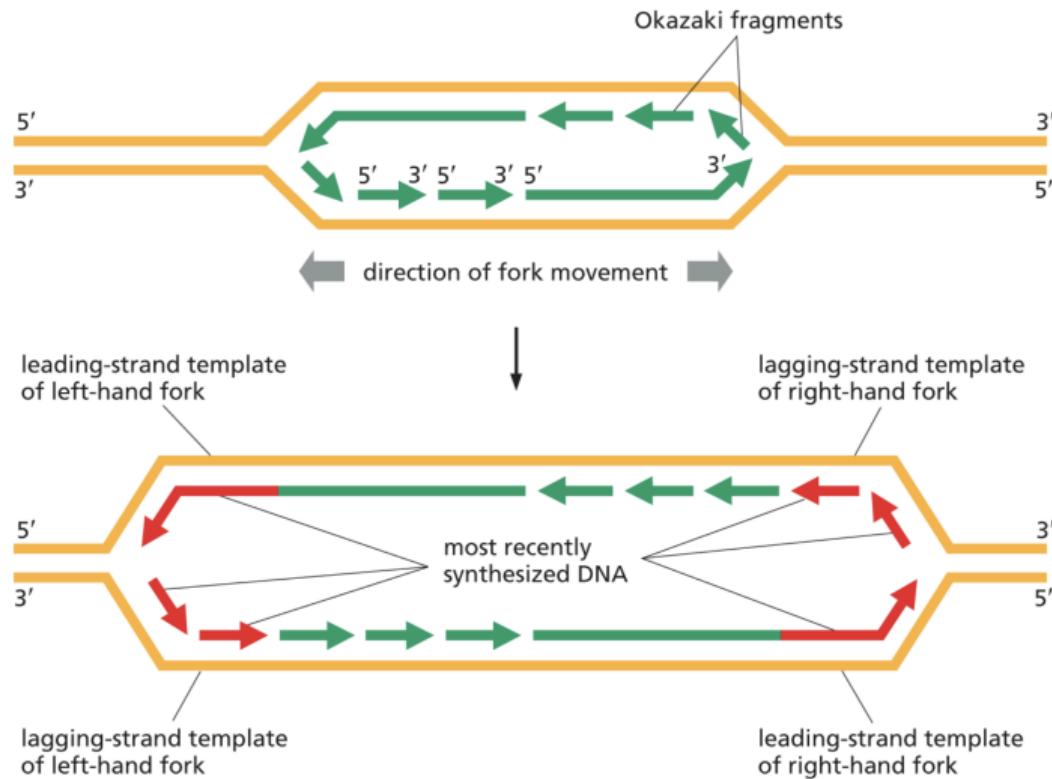


(B)

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.

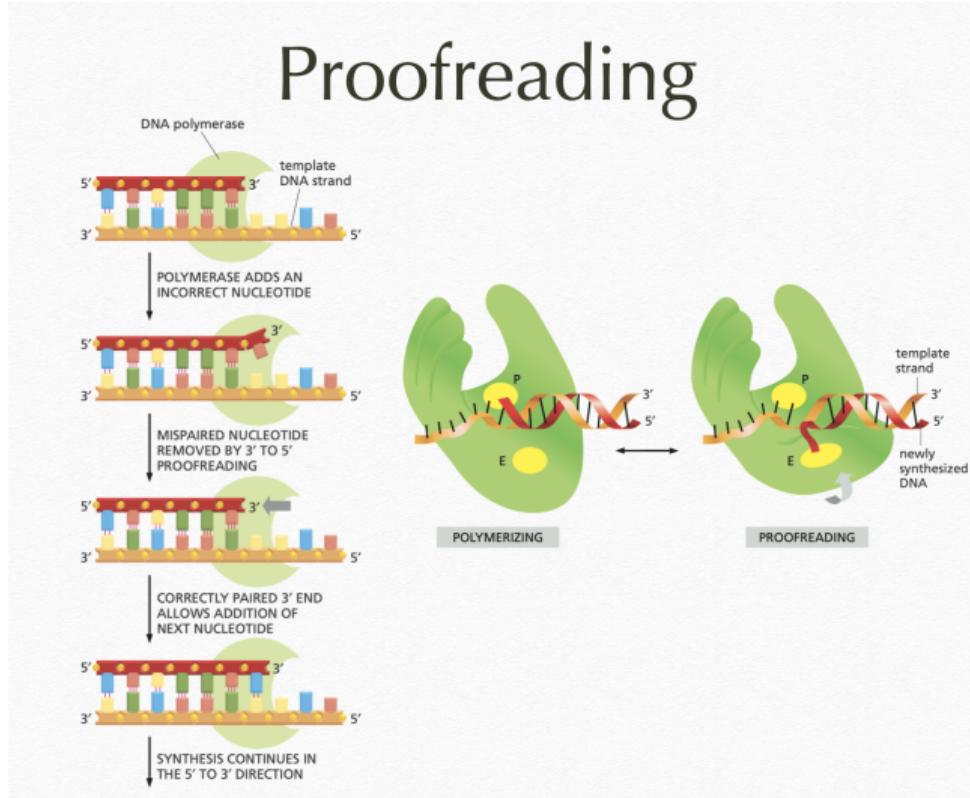
Asymmetrical Synthesis: Leading and Lagging Strands



Proofreading for High Fidelity

- DNA polymerase achieves extraordinary accuracy, making only about one error per 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.

Algorithmic Detection of Replication Origin

How can we find the replication origin in a genome — not in the lab, but using algorithms?

- Biological experiments are expensive and time consuming.
- But the DNA sequence itself contains compositional asymmetries.
- These asymmetries can be analyzed algorithmically to find where replication starts and ends.

Is the Problem Well-Defined for Computer Scientists?

- In biology, the question is “*Where does replication start?*”
- In computer science, we must **formalize** this question.
- Defining the problem precisely — and mapping it into computational terms — is a key step in solving any bioinformatics problem.

Algorithmic Detection of Replication Origin: The Biological Idea

- During DNA replication:
 - **Leading strand** tends to accumulate more **G (guanine)** nucleotides.
 - **Lagging strand** tends to accumulate more **C (cytosine)** nucleotides.
- This produces a **strand asymmetry** over evolutionary time.
- The switch in this bias marks the **origin (oriC)** and **terminus (terC)**.

Algorithmic Detection of Replication Origin: GC Skew Concept

- Quantify base composition bias as:

$$GC_skew = \frac{G - C}{G + C}$$

- Plotting the **cumulative GC skew** over the genome produces a smooth curve.
- Minimum point** \Rightarrow **Origin of replication (oriC)**
- Maximum point** \Rightarrow **Replication terminus (terC)**

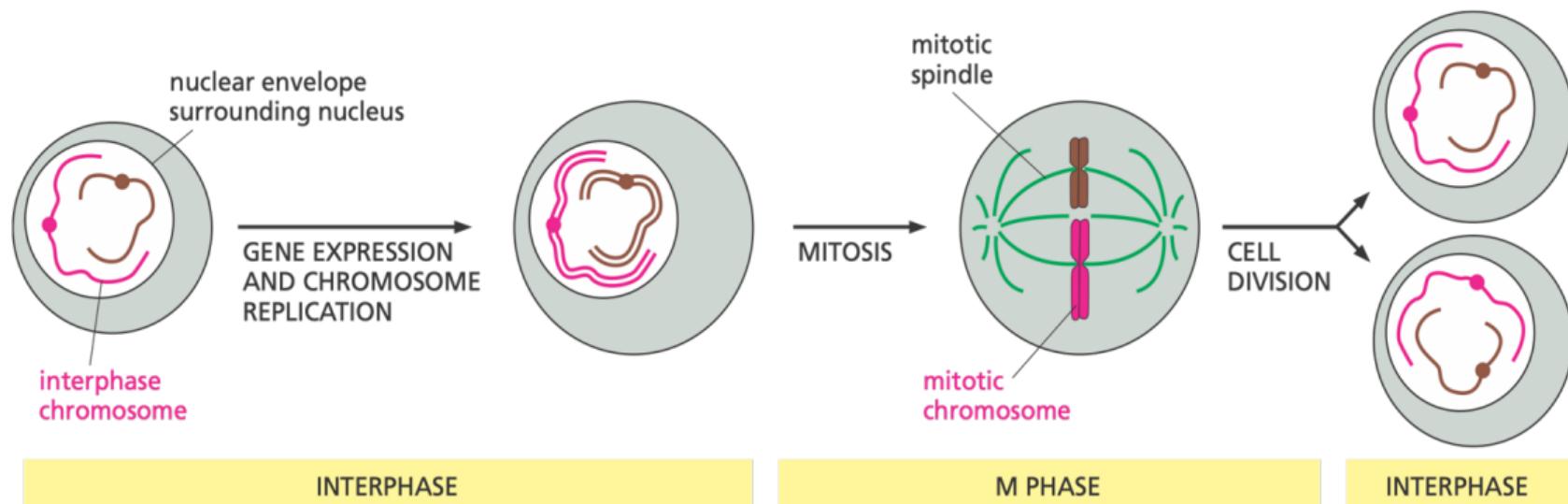
Algorithmic Detection of Replication Origin: Algorithm Steps

- ① Fetch the genome sequence.
- ② Score each base:
 - +1 for G
 - -1 for C
 - 0 for A or T
- ③ Compute the cumulative sum \Rightarrow **cumulative GC skew**.
- ④ Find the **minimum** and **maximum** points \Rightarrow **oriC and terC**.
- ⑤ Plot the GC skew to visualize replication landmarks.

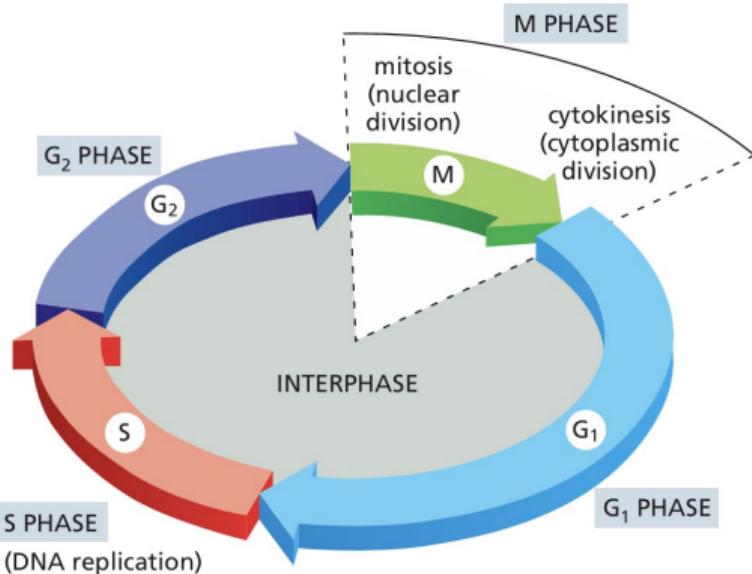
Cell Division

The cell cycle is a series of steps where a cell grows, duplicates its DNA and organelles, and divides. There are two types of cell division:

- **Mitosis:** produces two identical somatic cells.
- **Meiosis:** produces gametes with half the chromosome number.



Cell Cycle Phases

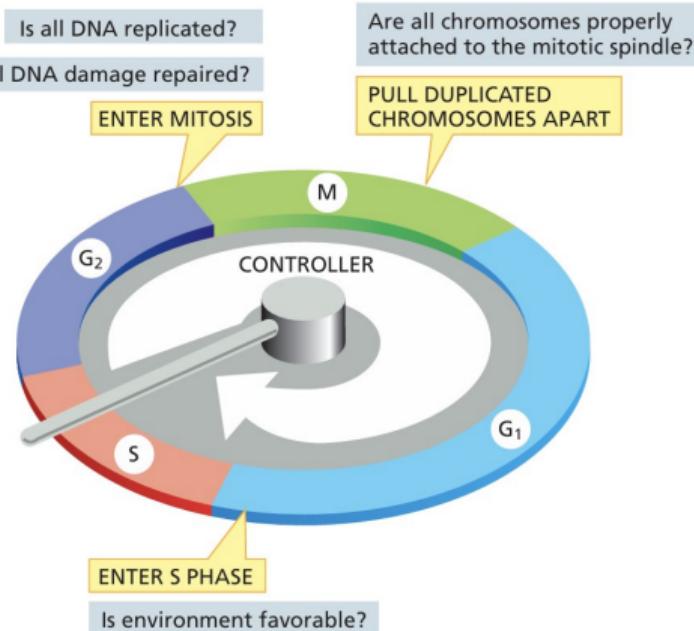


The cell cycle consists of four main phases:

- G₁ phase: cell growth and preparation for DNA replication.
- S phase: DNA synthesis (replication).
- G₂ phase: growth and preparation for mitosis.
- M phase: includes mitosis (nuclear division) and cytokinesis (cell division).

Some cells (like early embryos) shorten or skip G₁ and G₂, dividing rapidly without growth.

Cell Cycle Control System



The cell cycle control system ensures that each stage is completed before the next begins, using checkpoints at:

- G₁ → S: checks environmental conditions and signals.
- G₂ → M: ensures DNA is fully replicated and undamaged.
- M phase: confirms chromosomes are properly attached to the spindle.

Malfunction of the control system can cause uncontrolled cell division and cancer.

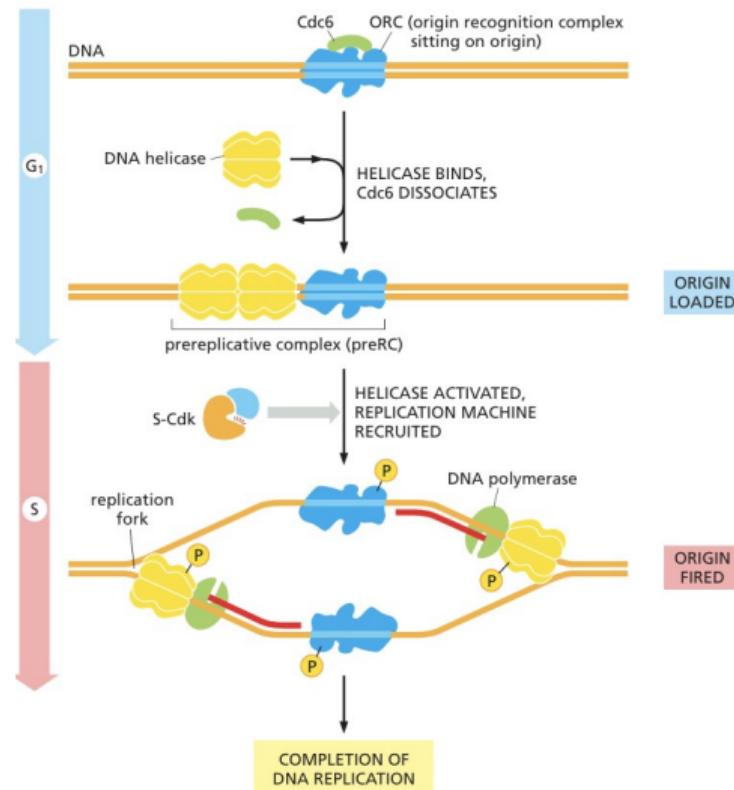
Cell Cycle: G1 PHASE

- G1 is a phase of cell growth and decision-making.
- Cells may proceed to S phase, pause, enter G0, or exit the cycle.
- The cell-cycle control system confirms that conditions are favorable for proliferation before committing to DNA replication.
- Once past the crucial G1 to S transition, the cell is committed to completing the rest of the cell cycle.
- Time in G1/G0 varies among cell types, affecting division rates.

Cell Cycle: S PHASE

- The phase of Interphase dedicated to DNA replication.
- Ensures accurate, one-time duplication of the entire genome.
- S-Cdk complex activates DNA synthesis at the end of G1.
- Prevents re-replication by inactivating proteins like Cdc6 and ORC.
- Cohesin proteins link newly formed sister chromatids for mitosis.

Cell Cycle: G1 and S PHASE



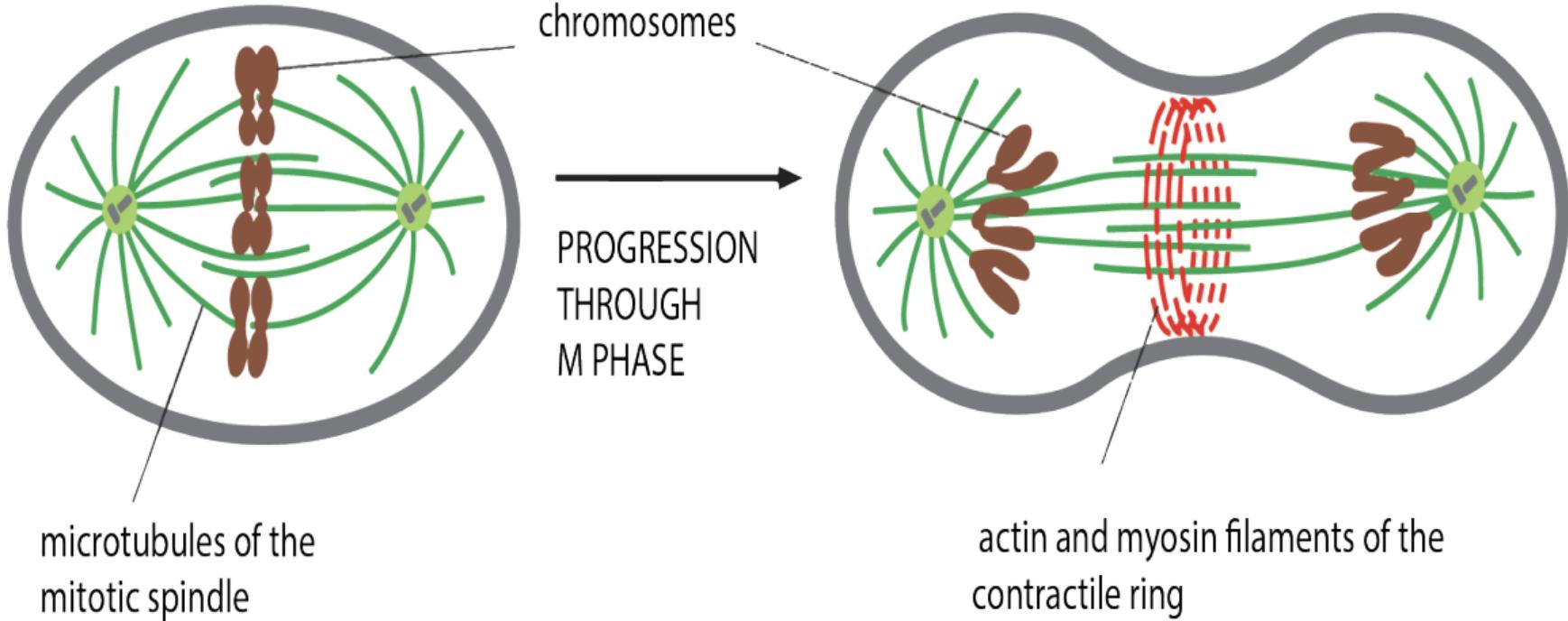
Cell Cycle: G2 PHASE

- The second Interphase gap, between S phase and M phase.
- The cell continues growth, gene transcription, and protein synthesis.
- Serves as a checkpoint to ensure DNA is fully replicated and undamaged.
- M-Cdk complex builds up but stays inactive until the G2-M transition.
- If issues are detected, the cycle is paused for DNA repair before mitosis.

Cell Cycle: M PHASE

- M phase includes mitosis (nuclear division) and cytokinesis (cytoplasmic division).
- Ensures accurate chromosome segregation and equal distribution of cell components.
- Mitosis stages: Prophase, Prometaphase, Metaphase, Anaphase, Telophase, followed by Cytokinesis.
- Outcome: two genetically identical daughter cells.

Cell Cycle: M PHASE



Ali Sharifi-Zarchi, Somayyeh Koohi (SUT)

What is Mitosis?

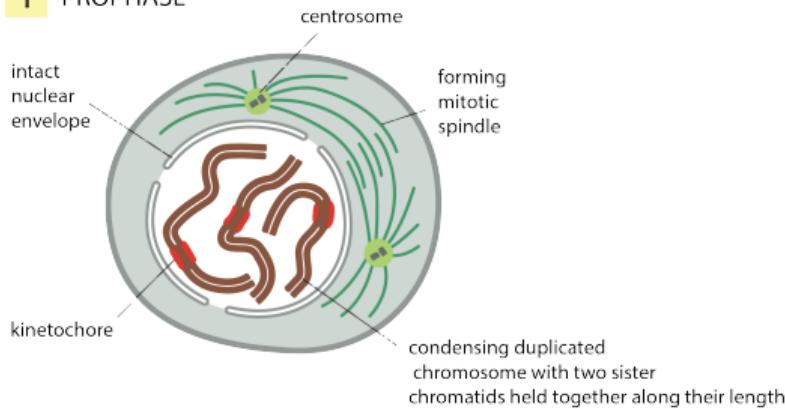
A type of cell division that results in two daughter cells each having the same number and kind of chromosomes as the present nucleus.

Characteristics of Mitosis:

- A diploid cell gives rise to a diploid cell.
- Chromosomes number remains the same.
- The DNA remains identically the same.
- One cell ($2n$) gives rise to two cells ($2n$).

M PHASE: PROPHASE

1 PROPHASE

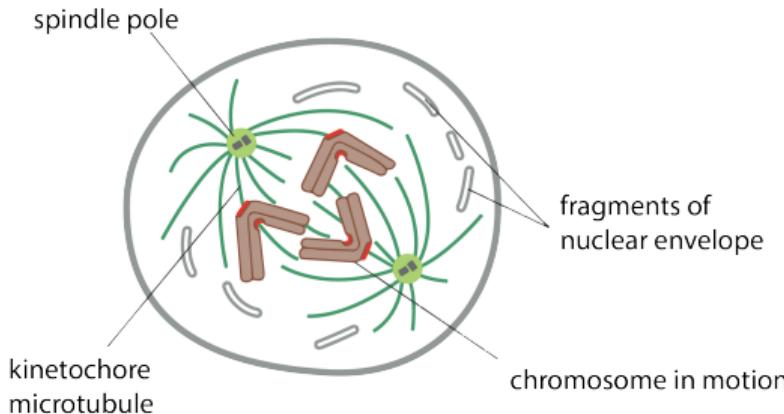


At prophase, the duplicated chromosomes, each consisting of two closely associated sister chromatids, condense.

Outside the nucleus, the mitotic spindle assembles between the two centrosomes, which have begun to move apart.

M PHASE: PROMETAPHASE

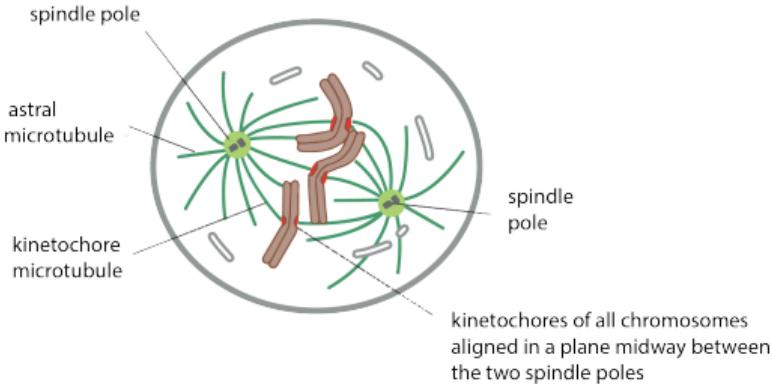
2 PROMETAPHASE



Prometaphase starts abruptly with the breakdown of the nuclear envelope. Chromosomes can now attach to spindle microtubules via their kinetochores and undergo active movement.

M PHASE: METAPHASE

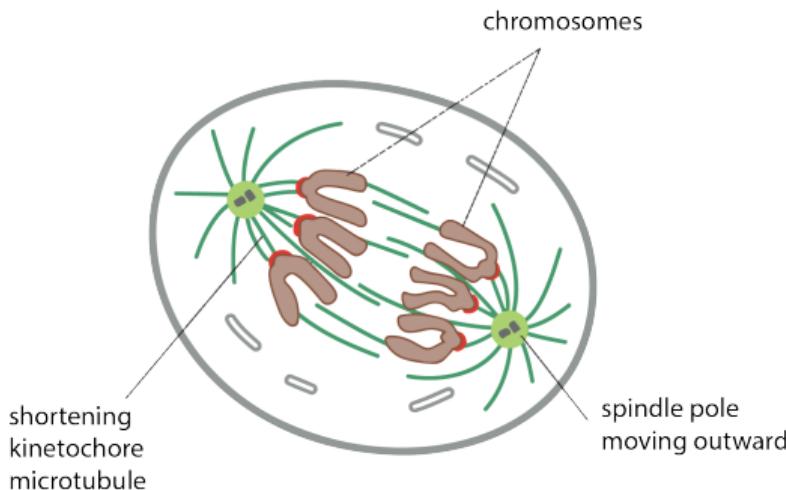
3 METAPHASE



At metaphase, the chromosomes are aligned at the equator of the spindle, midway between the spindle poles. The kinetochore microtubules on each sister chromatid attach to opposite poles of the spindle.

M PHASE: ANAPHASE

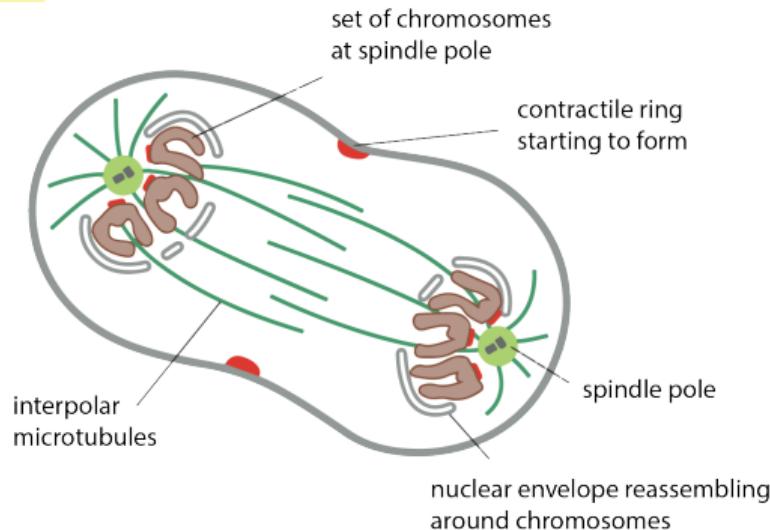
4 ANAPHASE



At anaphase, the sister chromatids synchronously separate and are pulled slowly toward the spindle pole to which they are attached. The kinetochore microtubules get shorter, and the spindle poles also move apart, both contributing to chromosome segregation.

M PHASE: TELOPHASE

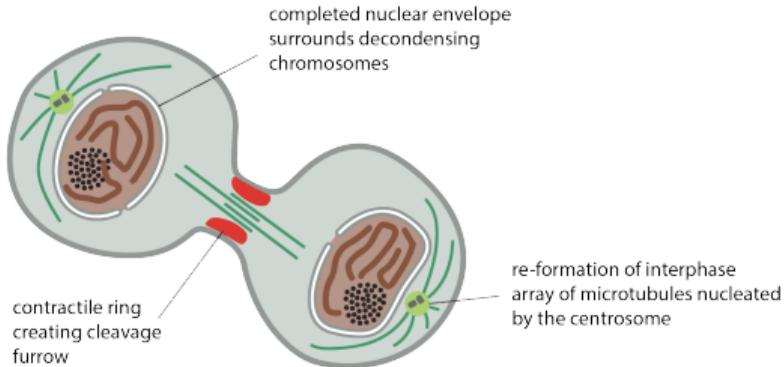
5 TELOPHASE



During telophase, the two sets of chromosomes arrive at the poles of the spindle. A new nuclear envelope reassembles around each set, completing the formation of two nuclei and marking the end of mitosis. The division of the cytoplasm begins with the assembly of the contractile ring.

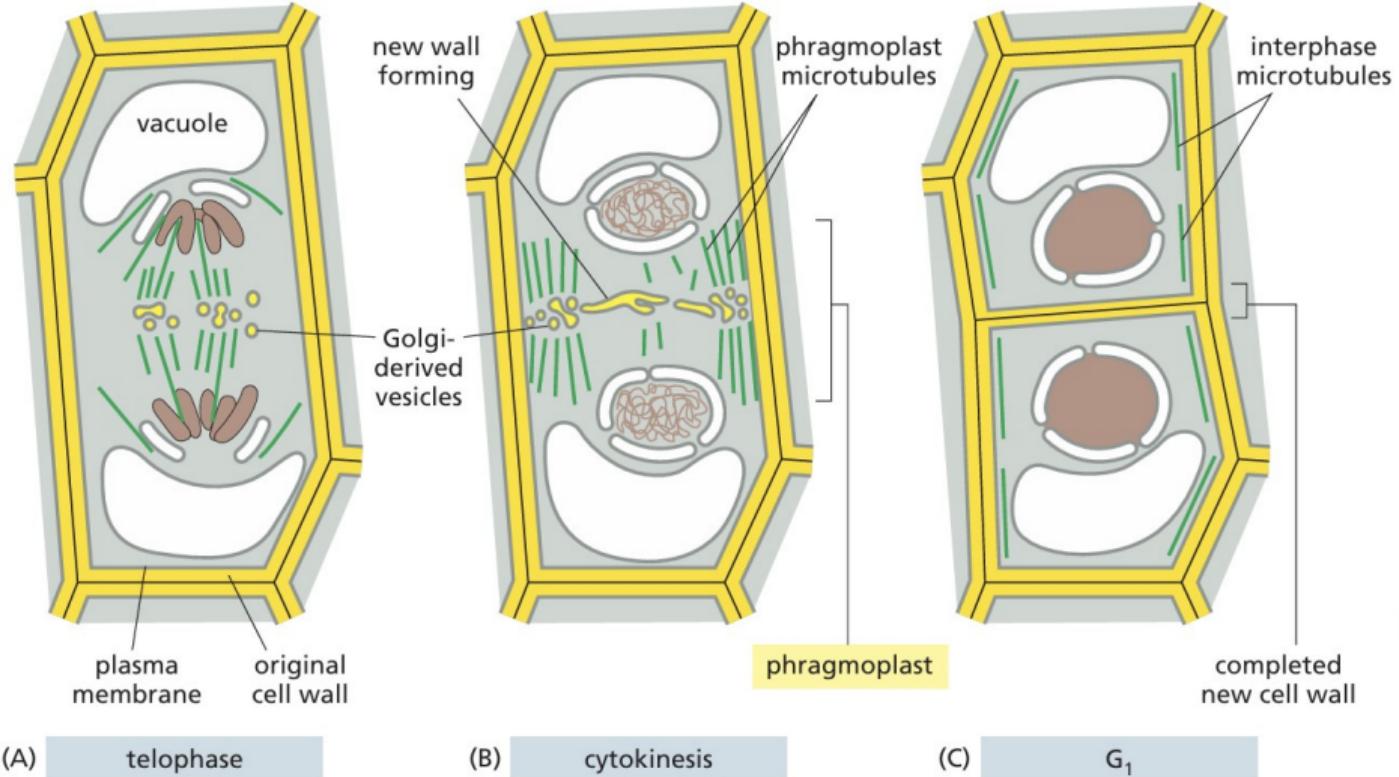
M PHASE: CYTOKINESIS

CYTOKINESIS



During cytokinesis of an animal cell, the cytoplasm is divided in two by a contractile ring of actin and myosin filaments, which pinches the cell into two daughters, each with one nucleus.

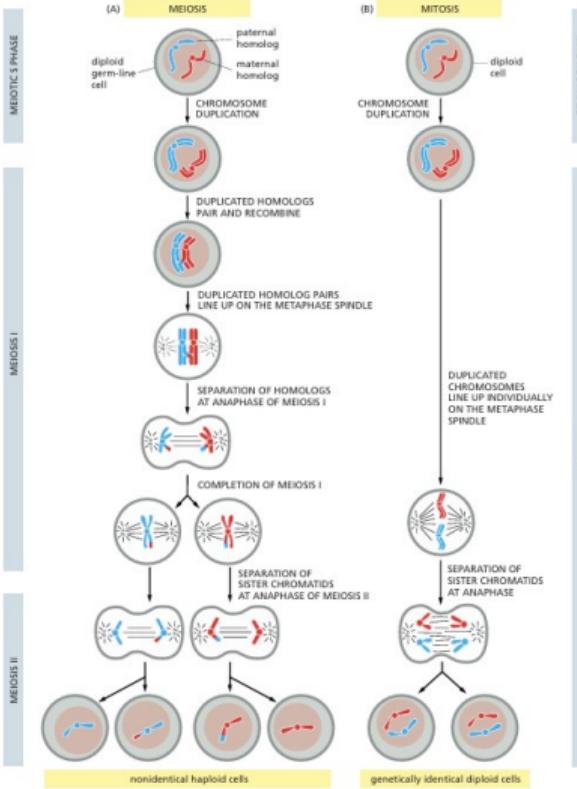
M PHASE: Cytokinesis



What is Meiosis?

- Meiosis is a type of cell division producing four haploid gametes from a diploid germ-line cell.
- It ensures that each gamete receives half the number of chromosomes, maintaining genetic stability across generations.
- The process involves one DNA replication followed by two divisions:
 - **Meiosis I:** homologous chromosomes pair, recombine, and segregate.
 - **Meiosis II:** sister chromatids separate, forming four unique haploid cells.

Meiosis

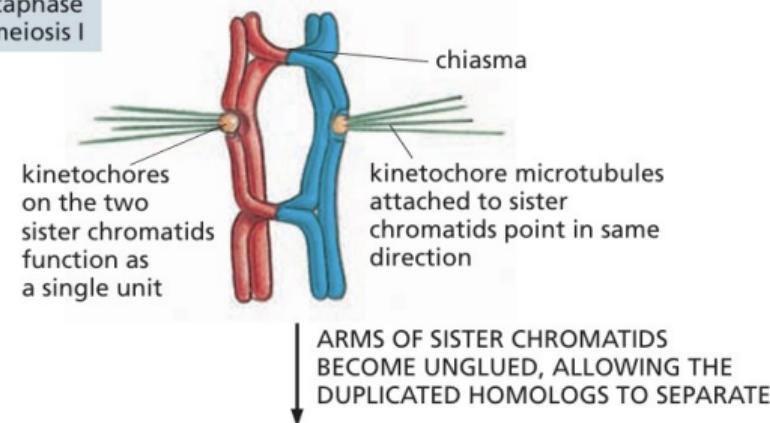


Meiosis I

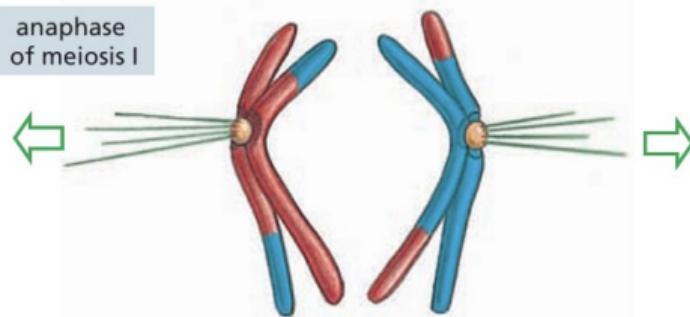
- Pairing and Bivalent Formation: Duplicated homologous chromosomes (maternal + paternal) pair along their length, forming a bivalent with all four sister chromatids together.
- Crossing-Over: Non-sister chromatids exchange DNA segments, forming chiasmata to hold homologs and ensure proper segregation.
- Metaphase I: The duplicated homolog pairs (bivalents) line up on the meiotic spindle.
- Anaphase I: Homologs separate to opposite poles; sister chromatids stay joined at the centromere.

Meiosis I

(A) metaphase of meiosis I



(B) anaphase of meiosis I

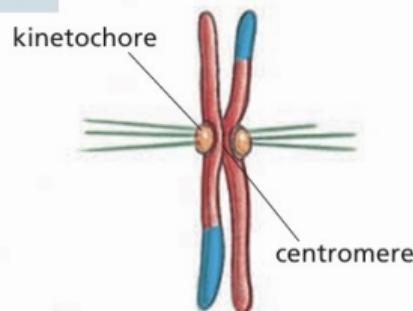


Meiosis II

- The second nuclear division following Meiosis I, occurring without DNA replication, separates the remaining sister chromatids and mechanically resembles mitosis.
- Metaphase II: Chromosomes align at the spindle equator; sister kinetochores attach to opposite spindle poles.
- Anaphase II: Centromeric cohesins are degraded, allowing sister chromatids to separate.
- Outcome: Four genetically dissimilar haploid nuclei are produced.

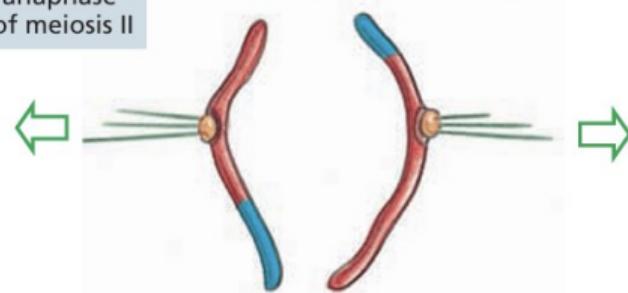
Meiosis II

(A) metaphase
of meiosis II



(B) anaphase
of meiosis II

COHESINS AT CENTROMERE
ARE DEGRADED; SISTER
CHROMATIDS SEPARATE



Recommended Videos

- **DNA and Chromosomes:**

- Chromosome Coiling
- DNA Structure
- DNA Structure and Replication

- **DNA Replication and Repair:**

- DNA Polymerase
- DNA Replication Machine
- DNA Replication Machine 2
- DNA Replication Machine 3
- DNA Replication Machine 4
- Finding Replication Origin

- **Mitosis and Meiosis:**

- Mitosis
- Meiosis
- Axolotl Embryo Development

References



Alberts, B. et al.
Essential Cell Biology.