DNA
The Language of Life

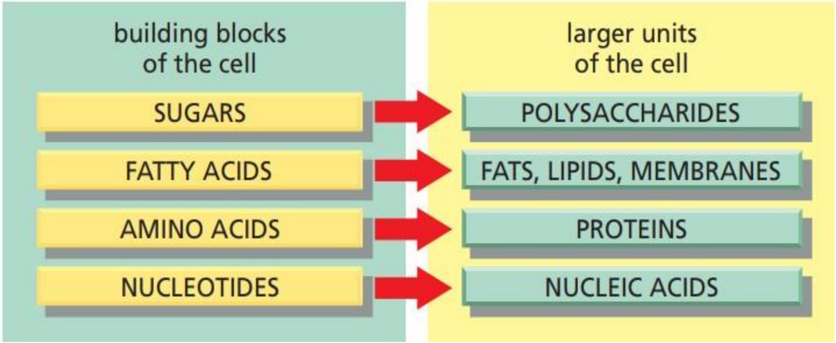
Introduction to Bioinformatics Course

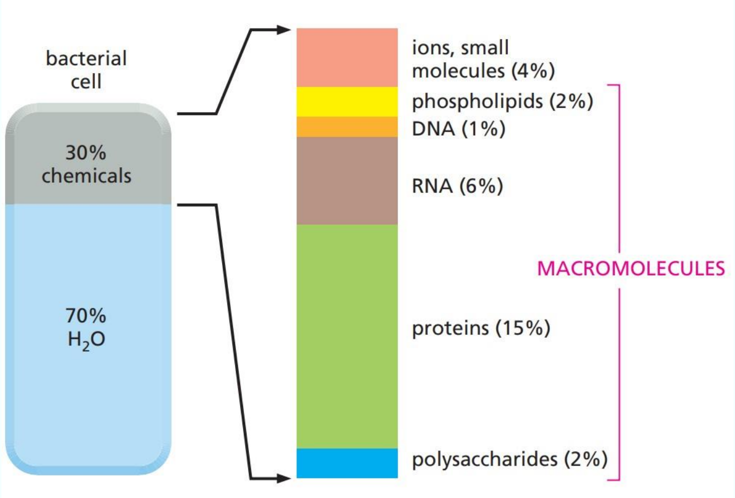
Niloufar Razani
Department of Computer Engineering
Sharif University of Technology

Fall 2023

Major Families of Organic Molecules in Cells

- If we disregard water, nearly all of the molecules in a cell are based on **carbon**.
- The small and large carbon compounds made by cells are called **organic molecules**. All other molecules, including water, are said to be inorganic by contrast.
- Sugars are energy sources for Cells and subunits of Polysaccharides.
- Fatty acids are Components of Cell membranes.
- Amino acids are the subunits of Proteins.
- **Nucleotides are the subunits of DNA and RNA.**

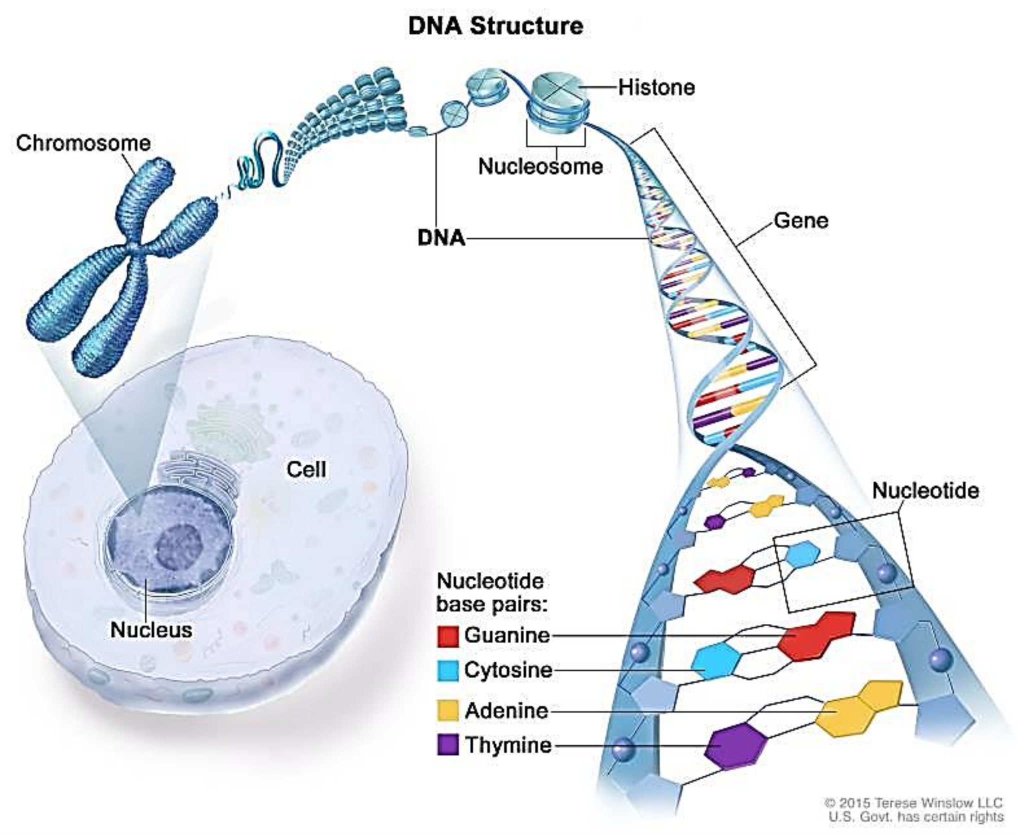




The Structure and Function of DNA

- Life depends on the ability of cells to store, retrieve, and translate the genetic instructions required to make and maintain a living organism.

- What kind of molecule could be capable of such accurate and almost unlimited replication, and also be able to direct the development of an organism and the daily life of a cell?

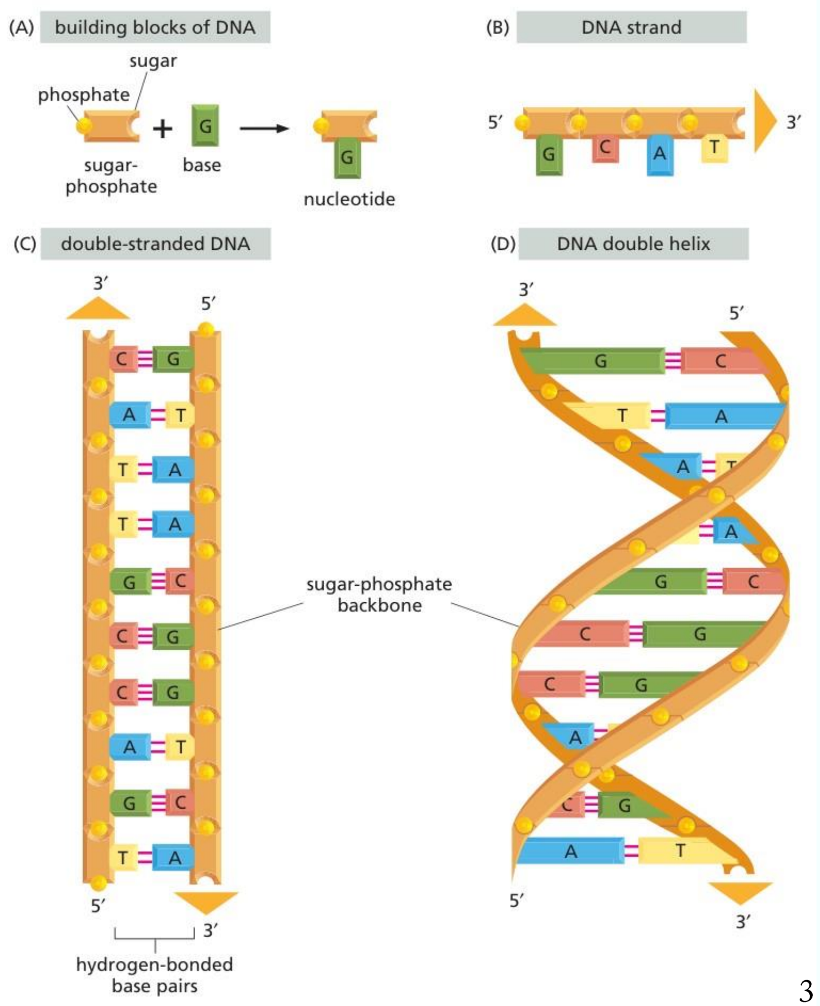


The Structure and Function of DNA

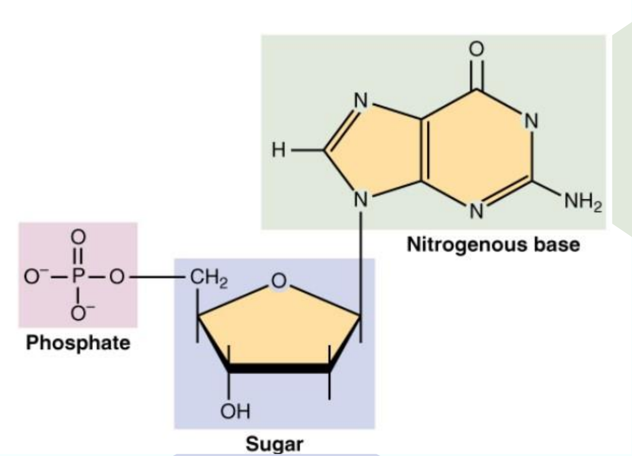
- A molecule of deoxyribonucleic acid (DNA) consists of two long poly-nucleotide chains.

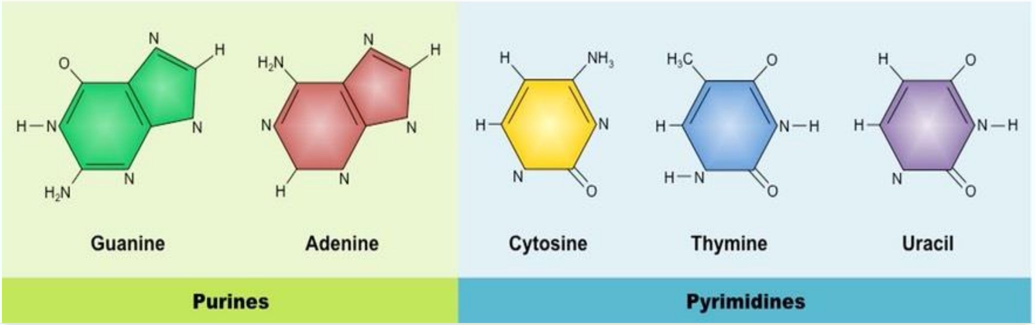
- The nucleotides are covalently linked together into polynucleotide chains, with a sugar–phosphate backbone from which the bases (A, C, G, and T) extend.

- A DNA molecule is composed of two polynucleotide chains (DNA strands) held together by hydrogen bonds between the paired bases.

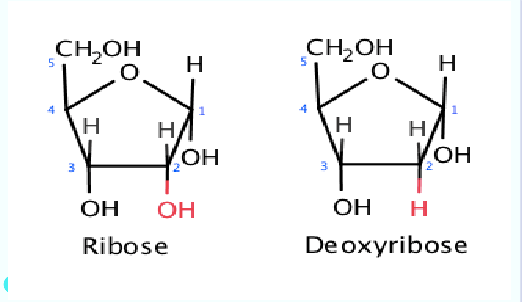


Nucleotides



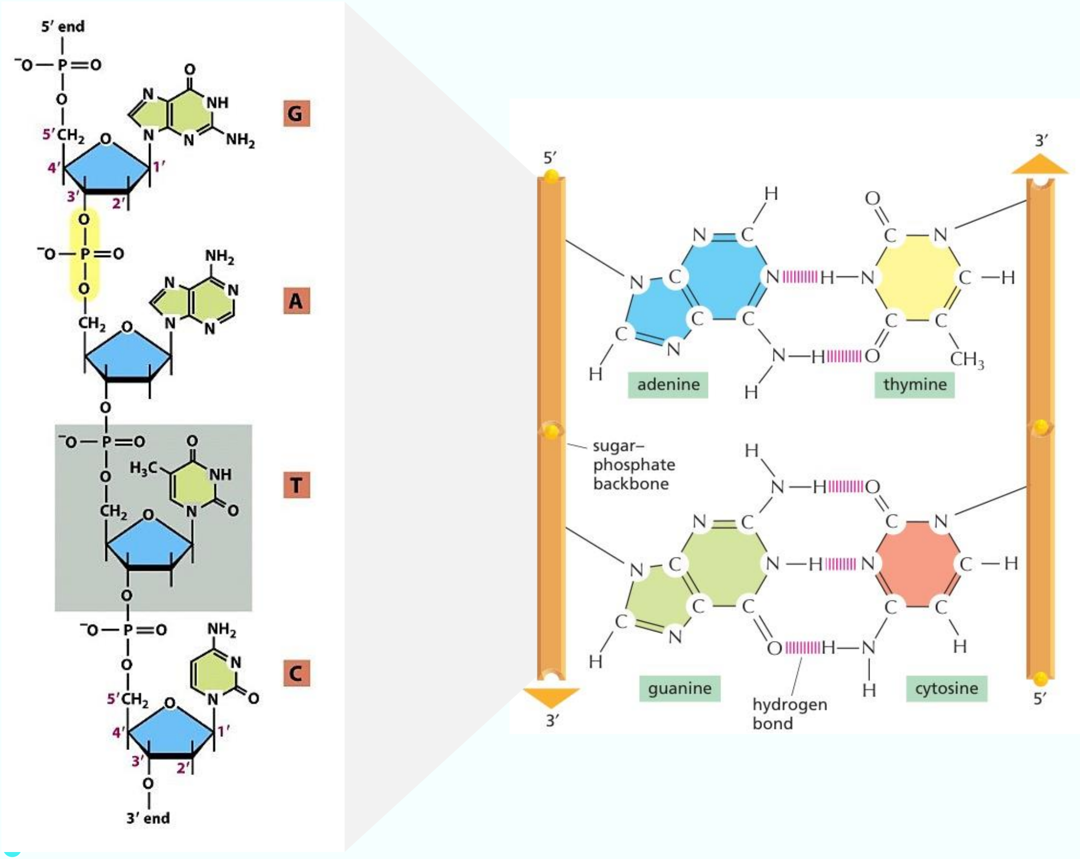


There are four different nitrogenous bases in DNA: adenine (A), guanine (G), thymine (T), and cytosine (C). In RNA, thymine is replaced with uracil (U)



A **pentose sugar** is a sugar molecule that contains five atoms of carbon. The pentose sugar in DNA is deoxyribose sugar, and the pentose sugar in RNA is ribose sugar.

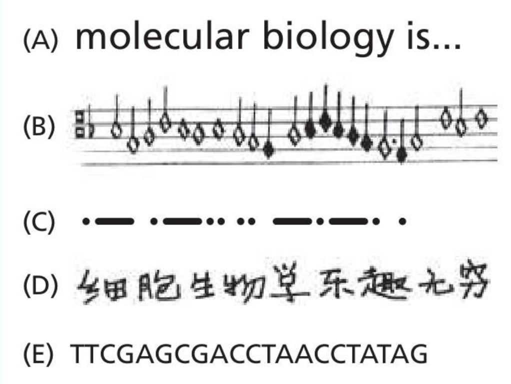
Nucleotide Bonding in DNA

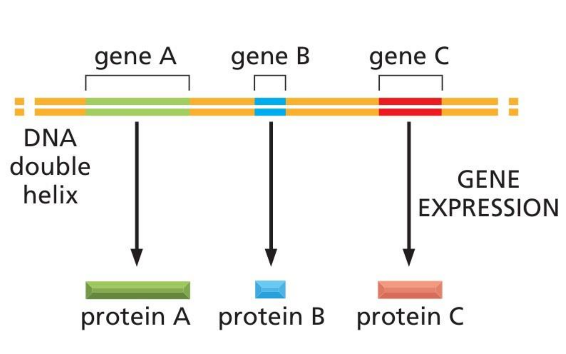


- A always pairs with T, and G always pairs with C.
- This complementary base-pairing enables the base pairs to be packed in the energetically most favorable arrangement.
- Each base pair is of similar width, thus holding the sugar–phosphate back-bones an equal distance apart along the DNA molecule.
- The members of each base pair can fit together within the double helix because the two strands of the helix run antiparallel to each other—that is, they are oriented with opposite polarities.

The Structure of DNA Provides a mechanism for heredity

- DNA encodes information in the order, or sequence, of the nucleotides along each strand.
- Each base—A, C, T, or G—can be considered as a letter in a four-letter alphabet that is used to spell out biological messages in the chemical structure of the DNA. For example genes contain the instructions for producing proteins. The DNA messages, therefore, must somehow encode proteins.
- Organisms differ from one another because their respective DNA molecules have different nucleotide sequences and, consequently, carry different biological messages.
- The complete set of information in an organism's DNA is called its genome.
- The complete sequence of the human genome would fill more than 1000 books each with about 900 pages.





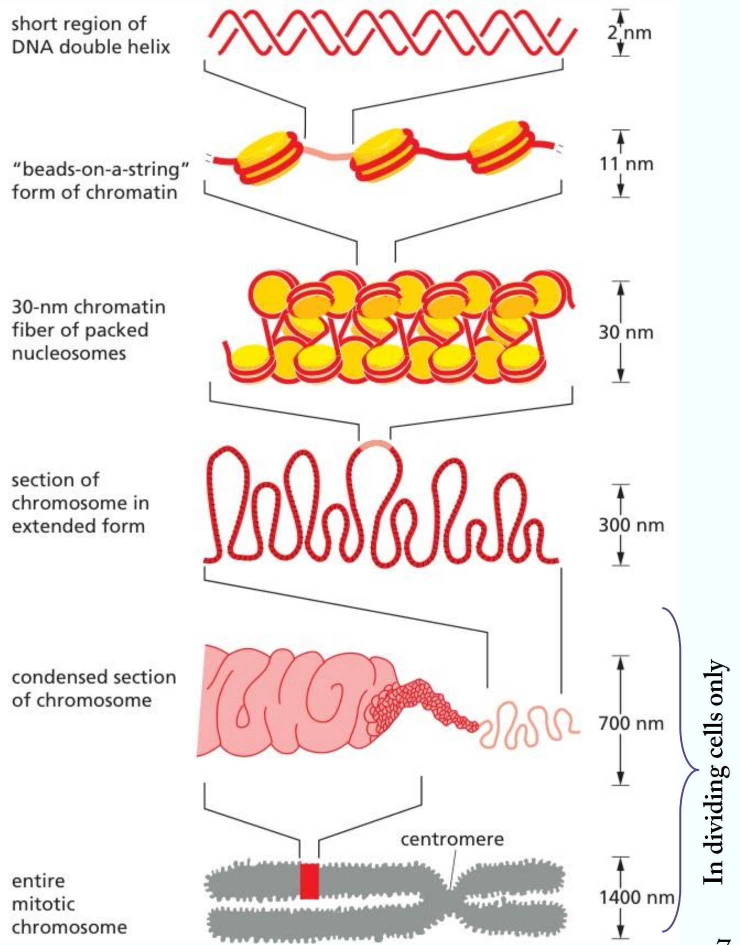
How can all this information be packed neatly into every cell nucleus?

The Structure of Eukaryotic Chromosomes

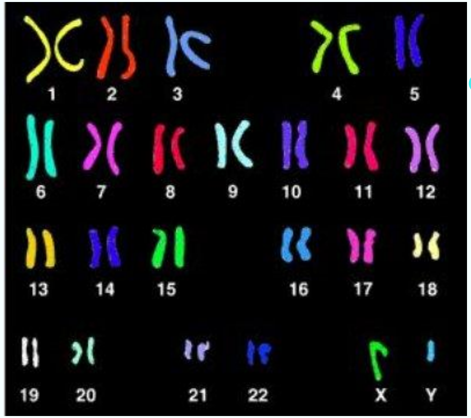
- Each human cell contains about 2 meter of DNA; yet the cell nucleus is only 5–8 micrometer in diameter.

- In eukaryotic cells, very long double-stranded DNA molecules are packaged into structures called chromosomes.

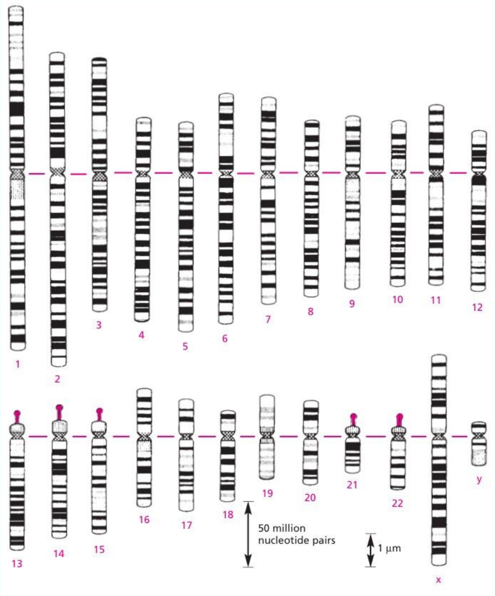
- The complex task of packaging DNA is accomplished by specialized proteins that bind to and fold the DNA, generating a series of coils and loops that provide increasingly higher levels of organization and prevent the DNA from becoming an unmanageable tangle.



Eukaryotic DNA is Packaged into multiple chromosomes

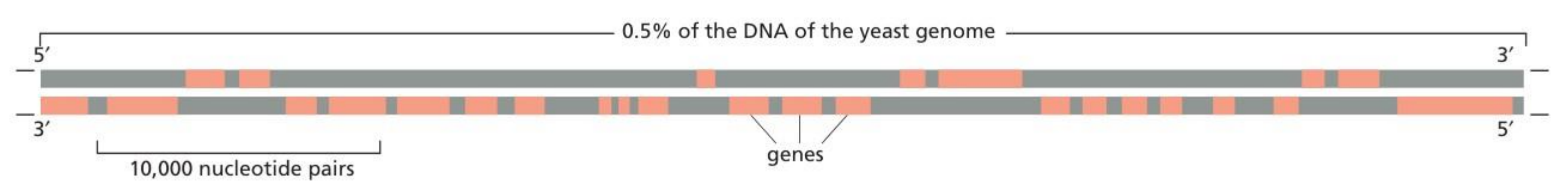


- The DNA in the nucleus is distributed among a set of different chromosomes. The human genome is parceled out into 23 chromosomes.
- Each chromosome consists of a single, enormously long, linear DNA molecule.
- The complex of DNA and protein is called chromatin.
- human cells each contain two copies of each chromosome, one inherited from the mother and one from the father.
- The maternal and paternal chromosomes of a pair are called homologous chromosomes (homologs). The only nonhomologous chromosome pairs are the sex chromosomes in males.
- A display of the full set of 46 human chromosomes is called the human karyotype.



Chromosomes contain long Strings of genes

- The most important function of chromosomes is to carry the genes—the functional units of heredity.
- A gene is usually defined as a segment of DNA that contains the instructions for making a particular protein.
- Some genes direct the production of an RNA molecule, instead of a protein, as their final product.
- In general, the more complex an organism, the larger is its genome. **But this relationship does not always hold true.**
- For example, the total number of genes ranges from less than 500 for a simple bacterium to about 25,000 for humans.
- The human genome is 30 times smaller than that of some plants and at least 60 times smaller than some species of amoeba.

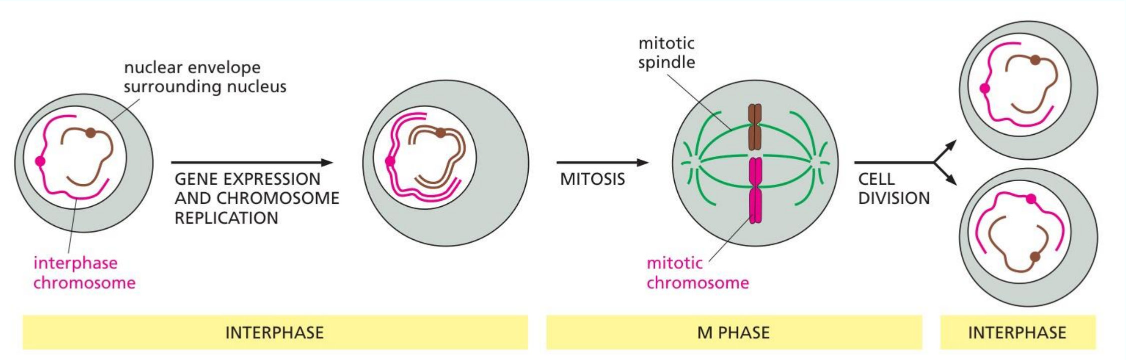


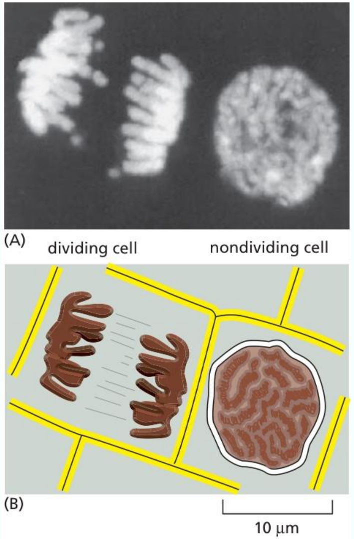
Chromosomes exist in different States Throughout the life of a cell

- A Chromosome must be able to be replicated, and the replicated copies must be separated and partitioned into daughter cells at each cell division.

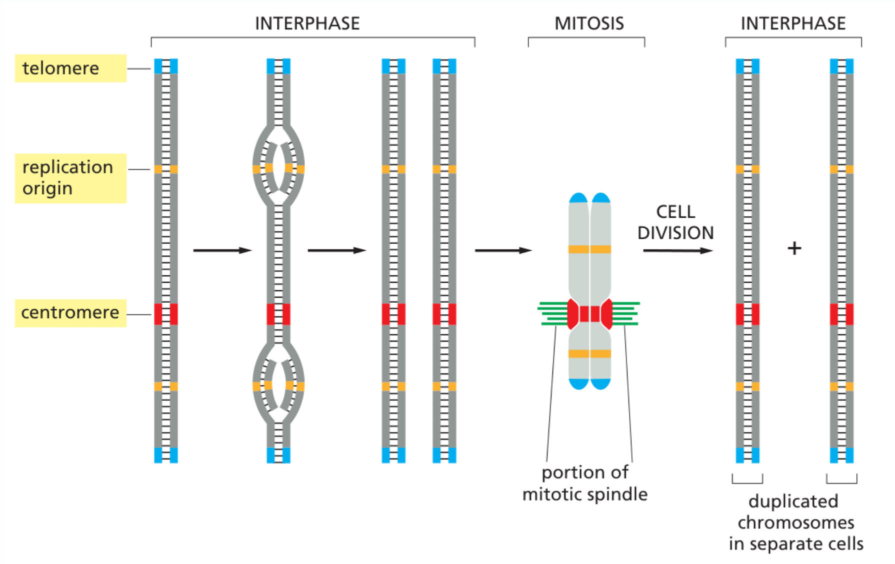
- Two broad stages of the cell cycle are: **Interphase**, when chromosomes are duplicated; and **Mitosis**, when they are distributed to the two daughter nuclei.

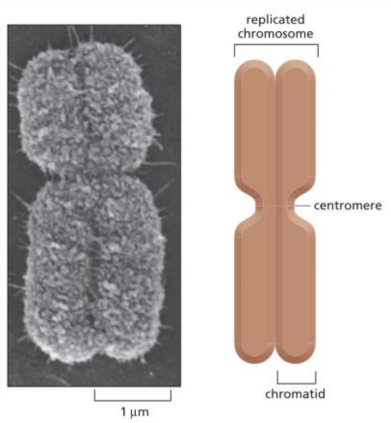
- During interphase, the chromosomes are extended as long, thin, tangled threads of DNA in the nucleus.





- Specialized DNA sequences found in all eukaryotes ensure that the interphase chromosomes replicate efficiently.
- One type of nucleotide sequence acts as a replication origin, where duplication of the DNA begins.
- Another DNA sequence forms the **telomeres** found at each of the two ends of a chromosome.

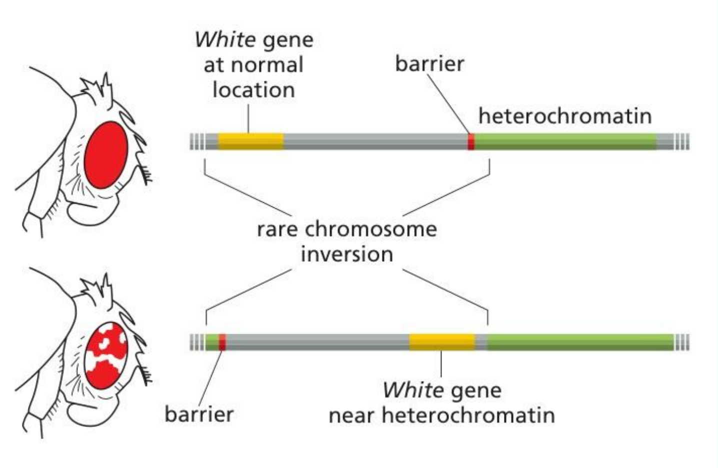




- When the cell cycle reaches M phase, the DNA coils up, adopting a more and more compact structure, ultimately forming highly compacted, or condensed, mitotic chromosomes.
- Once the chromosomes have condensed, it is the presence of the third specialized DNA sequence, the **centromere**, that allows one copy of each duplicated chromosome to be apportioned to each daughter cell

Interphase chromosomes contain Both condensed and more extended Forms of chromatin

- The chromatin in chromosomes is not uniformly packed.
- regions of the chromosome that contain genes that are being expressed are generally more extended, while those that contain quiescent genes are more compact.
- The most highly condensed form of interphase chromatin is called **Heterochromatin**. it is concentrated around the centromere region and in the telomeres.
- Most DNA that is permanently folded into heterochromatin in the cell does not contain genes.
- The rest of the interphase chromatin is called **Euchromatin.**



Genes that accidentally become packaged into heterochromatin usually fail to be expressed.

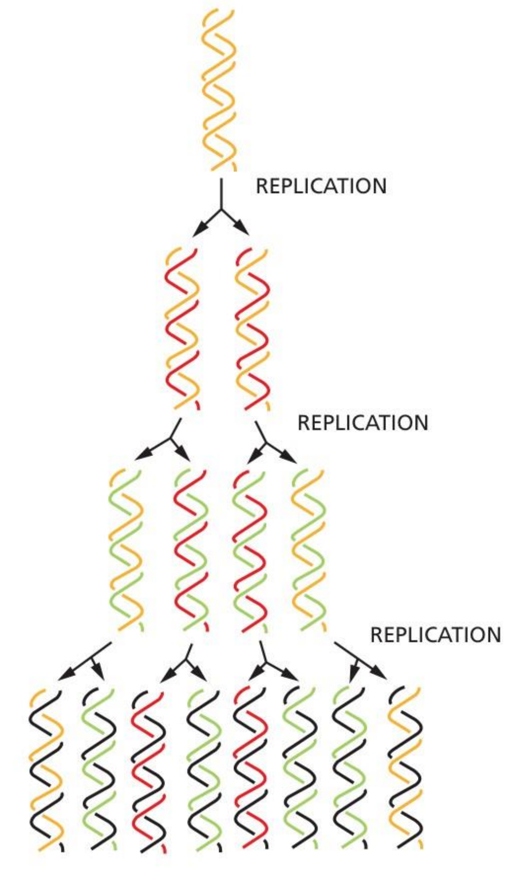
DNA Replication

- The ability of a cell to maintain order in a chaotic environment depends on the accurate duplication of the vast quantity of genetic information carried in its DNA.

- Maintaining order in a cell also requires the continual surveillance and repair of its genetic information.

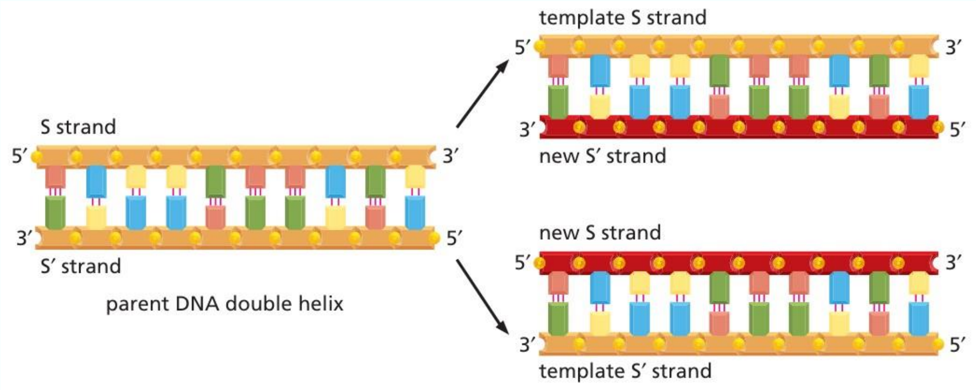
- Protein are the machines that replicate and repair the cell's DNA.

- Despite these systems for protecting the genetic instructions from copying errors and accidental damage, permanent changes, or mutations, sometimes do occur.

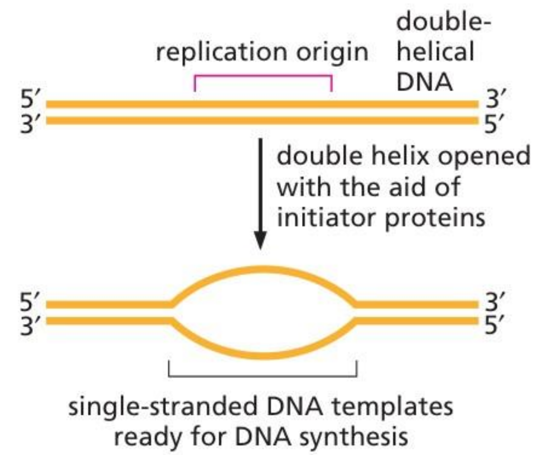


Base-Pairing Enables DNA Replication

- Each strand can act as a template, or mold, for the synthesis of a new complementary strand.
- If we designate the two DNA strands as S and S', strand S can serve as a template for making a new strand S', while strand S' can serve as a template for making a new strand S.
- The copying must be carried out with speed and accuracy: in about 8 hours, a dividing animal cell will copy the whole DNA in it.
- Each of the daughter DNA double helices ends up with one of the original (old) strands plus one strand that is completely new; this style of replication is said to be **Semiconservative**



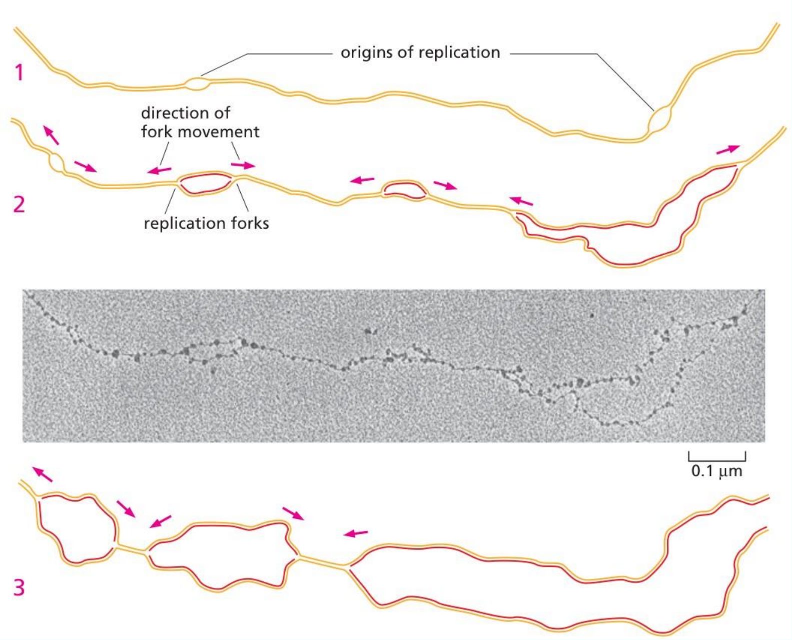
DNA Synthesis Begins at Replication Origins



- The DNA double helix is normally very stable.
- The process of DNA replication is begun by initiator proteins that bind to the DNA and pry the two strands apart, breaking the hydrogen bonds between the bases.
- Although the hydrogen bonds collectively make the DNA helix very stable, individually each hydrogen bond is weak.
- The positions at which the DNA is first opened are called replication origins, and they are usually marked by a particular sequence of nucleotides.
- They are composed of DNA sequences that attract the initiator proteins, as well as stretches of DNA that are especially easy to open.
- The human genome has approximately 10,000 such origins.

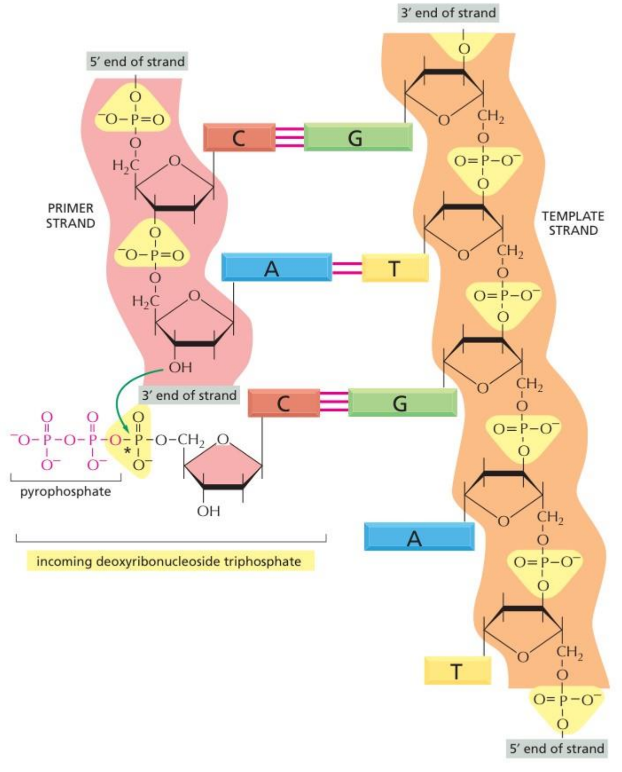
New DNA Synthesis Occurs at Replication Forks

- DNA molecules in the process of being replicated contain Y-shaped junctions called **replication forks.**
- Two replication forks are formed at each replication origin, and they move away from the origin in opposite directions, unzipping the DNA as they go.
- DNA replication in bacterial and eukaryotic chromosomes is therefore termed **Bidirectional**.
- The forks move very rapidly—at about 1000 nucleotide pairs per second in bacteria and 100 nucleotide pairs per second in humans.



New DNA Synthesis Occurs at Replication Forks

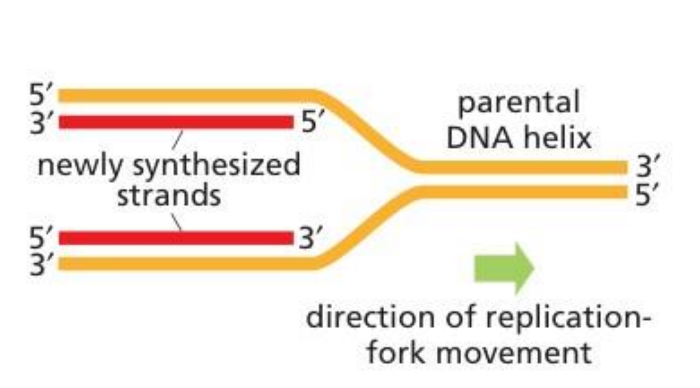
- At the heart of the replication machine is an enzyme called **DNA polymerase**, which synthesizes new DNA using one of the old strands as a template.
- This enzyme catalyzes the addition of nucleotides to the 3’ end of a growing DNA.
- Nucleotides enter the reaction initially as nucleoside triphosphates, which provide the energy for polymerization.
- DNA polymerase does not dissociate from the DNA each time it adds a new nucleotide to the growing chain.



The Replication Fork is asymmetrical

- The 5’-to-3’ direction of the DNA polymerization mechanism poses a problem at the replication fork.

- At the replication fork, one new DNA strand is being made on a template that runs in one direction (3’ to 5’), whereas the other new strand is being made on a template that runs in the opposite direction (5’ to 3’).



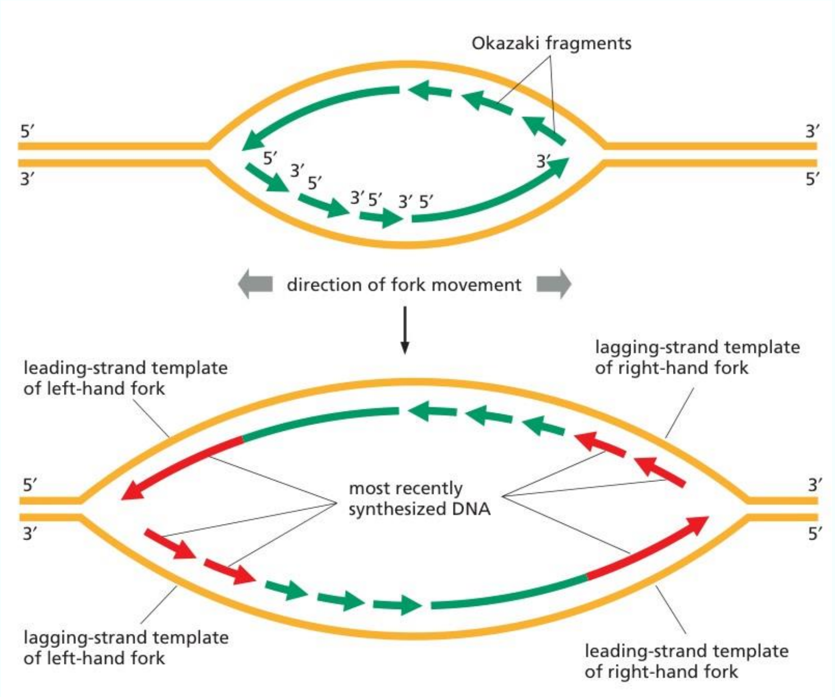
- The replication fork is therefore asymmetrical.

- DNA polymerase, however, can catalyze the growth of the DNA chain in only one direction: it can add new subunits only to the 3’ end of the chain.

- This can easily account for the synthesis of one of the two strands of DNA at the replication fork, but not the other.

The Replication Fork is asymmetrical

- The problem is solved by the use of a '**backstitching**' maneuver.
- The DNA strand whose 5' end must grow is made discontinuously, in successive separate small pieces, with the DNA polymerase moving backward with respect to the direction of the replication fork.
- The resulting small DNA pieces—called **Okazaki fragments.**
- The DNA strand that is synthesized discontinuously in this way is called the **lagging strand**; the other strand, which is synthesized continuously, is called the **leading strand**.



DNA Polymerase is self-correcting

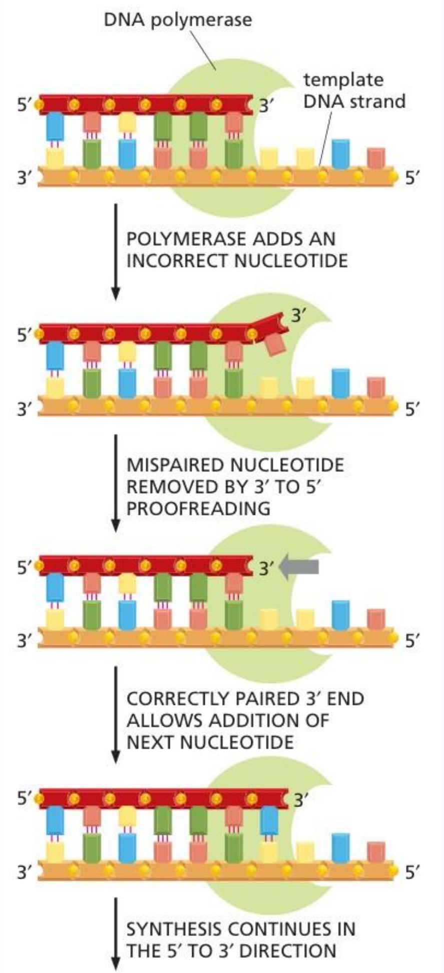
- DNA polymerase is so accurate that it makes only about one error in every 10^7 nucleotide pairs it copies.

- Although A-T and C-G are by far the most stable base pairs, other, less stable base pairs—for example, G-T and C-A—can also be formed.

- This catastrophe is avoided because DNA polymerase has two special qualities:

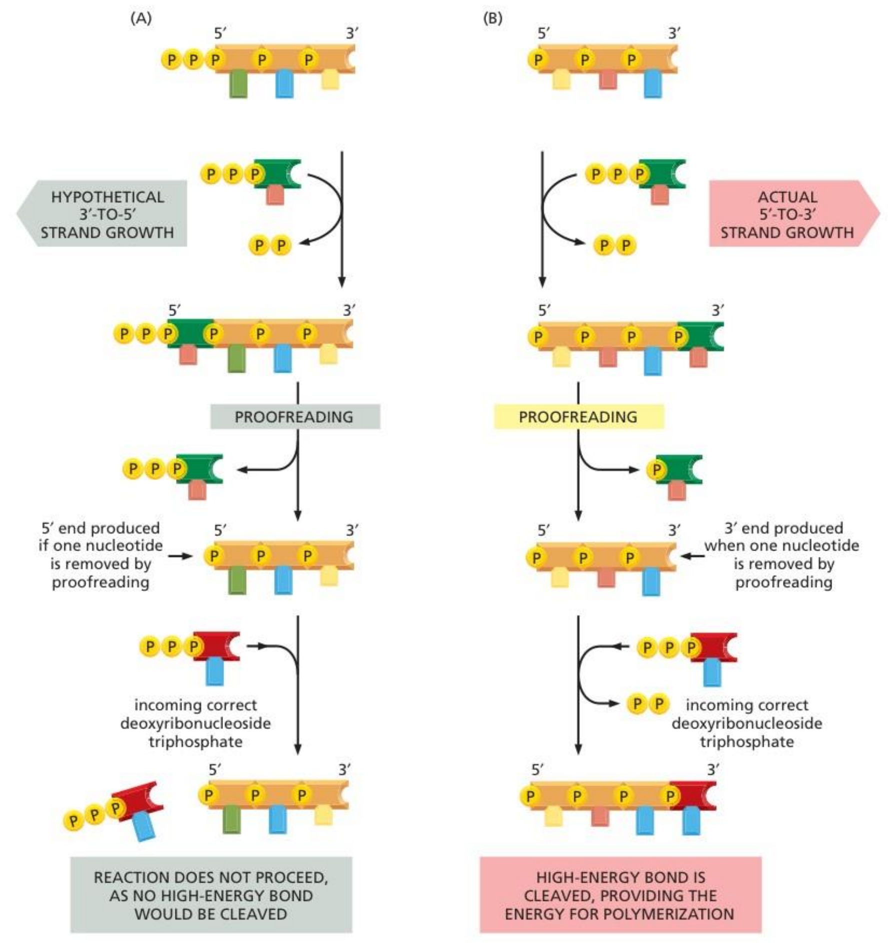
  First, the enzyme carefully monitors the base-pairing between each incoming nucleotide and the template strand.

  Second, when DNA polymerase makes a rare mistake and adds the wrong nucleotide, it can correct the error through an activity called **proofreading**.



DNA Polymerase is Self-correcting

- Proofreading takes place at the same time as DNA synthesis.
- Thus, DNA polymerase possesses a highly accurate 5'-to-3' polymerization activity, as well as a 3'-to-5' proofreading activity.
- This proofreading mechanism explains why DNA polymerases synthesize DNA only in the 5'-to-3' direction, despite the need that this imposes for a cumbersome backstitching mechanism at the replication fork.
- A hypothetical DNA polymerase that synthesized in the 3'-to-5' direction (and would thereby circumvent the need for backstitching) would be unable to proofread.



Short Lengths of RNA act as Primers for DNA Synthesis

- the polymerase can join a nucleotide only to a base-paired nucleotide in a DNA double helix, it cannot start a completely new DNA strand.

- A different enzyme is needed—one that can begin a new polynucleotide chain simply by joining two nucleotides together without the need for a base-paired end.

- This enzyme makes a short length (about 10 nucleotides) of a RNA (ribonucleic acid) using the DNA strand as a template and provides a base-paired 3' end as a starting point for DNA polymerase.

- It thus serves as a **primer** for DNA synthesis, and the enzyme that synthesizes the RNA primer is known as **Primase**.

- However, because U can form a base pair with A, the RNA primer is synthesized on the DNA strand by complementary base-pairing in exactly the same way as is DNA.

- To produce a continuous new DNA strand from the many separate pieces of nucleic acid made on the lagging strand, three additional enzymes are needed. These act quickly to remove the RNA primer, replace it with DNA, and join the DNA fragments together.

Proteins at a Replication Fork Cooperate to Form a Replication Machine

- Other proteins together with DNA polymerase and primase, form the protein machine that powers the replication fork forward and synthesizes new DNA behind it.

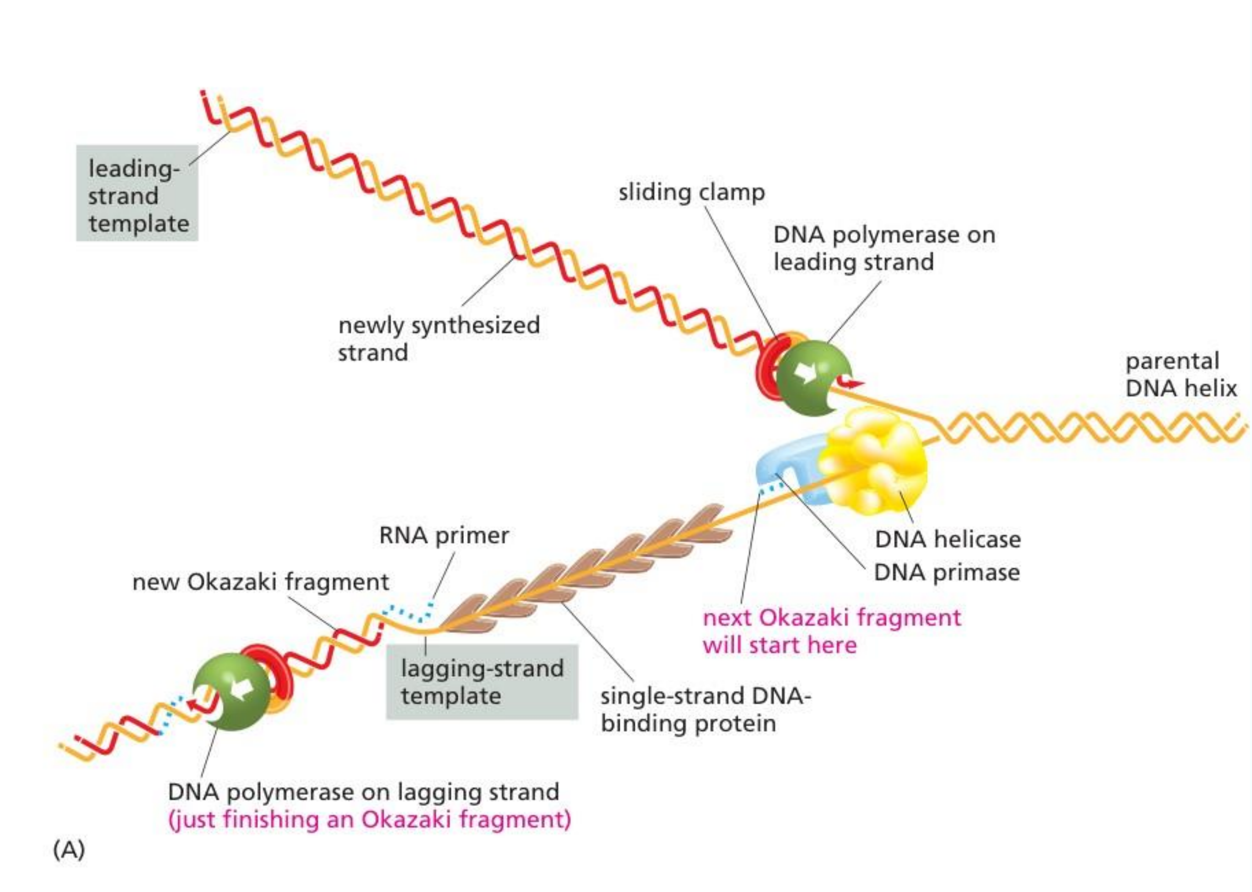
- At the very front of the replication machine is the **Helicase**, a protein that uses the energy of ATP hydrolysis to pry apart the double helix as it speeds along the DNA.

- The single-strand binding protein clings to the single-stranded DNA exposed by the helicase, transiently preventing it from re-forming base pairs and keeping it in an elongated form so that it can readily serve as a template for DNA polymerase.

- An additional replication protein, called a **Sliding Clamp**, keeps the polymerase firmly attached to the template while it is synthesizing new strands of DNA.

- Most of the proteins involved in DNA replication are held together in a large multienzyme complex that moves as a unit along the DNA, enabling DNA to be synthesized on both strands in a coordinated manner.

Proteins at a Replication Fork Cooperate to Form a Replication Machine



RESOURCES

- Alberts, Bruce, et al. *Essential cell biology*, Chapter2, Chemical Components of Cells.
- Alberts, Bruce, et al. *Essential cell biology*, Chapter5, DNA and Chromosomes.
- Alberts, Bruce, et al. *Essential cell biology*, Chapter6, DNA Replication, Repair, and Recombination