

GENETICS

Information from Khan academy



A ROUGH SUMMARY

DNA

Deoxyribonucleic acid

DNA is made of 4 nucleotides- **Adenine**, **Thymine**, **Guanine** and **Cytosine**. **Adenine** and **Guanine** are single-ring nucleotides, and **Cytosine** and **Thymine** are double-ringed nucleotides. **Adenine** and **Thymine** form hydrogen bonds, and **Guanine** and **Cytosine** form hydrogen bonds. Each nucleotide is made up of a phosphate and five carbon sugar backbone, and it also has a nitrogenous base, which is where genetic information is stored.

RNA

Ribonucleic acid

There are four types of RNA- messenger RNA (mRNA), transfer RNA(tRNA), ribosomal RNA(rRNA) and micro RNA(miRNA). RNA is still made up of four nucleotides- **Adenine**, **Uracil**, **Guanine** and **Cytosine**. **Uracil** is still a double-ringed nucleotide.

Nucleic acids, macromolecules made out of units called nucleotides, come in two naturally occurring varieties: **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. DNA is the genetic material found in living organisms, all the way from single-celled bacteria to multicellular mammal. Some [viruses](#) use RNA, not DNA, as their genetic material, but aren't technically considered to be alive (since they cannot reproduce without help from a host).

In eukaryotes, such as plants and animals, DNA is found in the **nucleus**. In prokaryotes, such as bacteria, the DNA is not enclosed in a membranous envelope, although it's located in a specialized cell region called the **nucleoid**.

Many genes encode protein products, meaning that they specify the sequence of amino acids used to build a particular protein. Before this information can be used for protein synthesis, however, an RNA copy (transcript) of the gene must first be made. This type of RNA is called a **messenger RNA (mRNA)**, as it serves as a messenger between DNA and the ribosomes, molecular machines that read mRNA sequences and use them to build proteins.

DNA and RNA are polymers, and are made up of monomers known as **nucleotides**. When these monomers combine, the resulting chain is called a **polynucleotide** (*poly-* = "many").

Each nucleotide is made up of three parts: a nitrogen-containing ring structure called a nitrogenous base, a five-carbon sugar, and at least one

phosphate group. The sugar molecule has a central position in the nucleotide, with the base attached to one of its carbons and the phosphate group (or groups) attached to another. Let's look at each part of a nucleotide in turn.

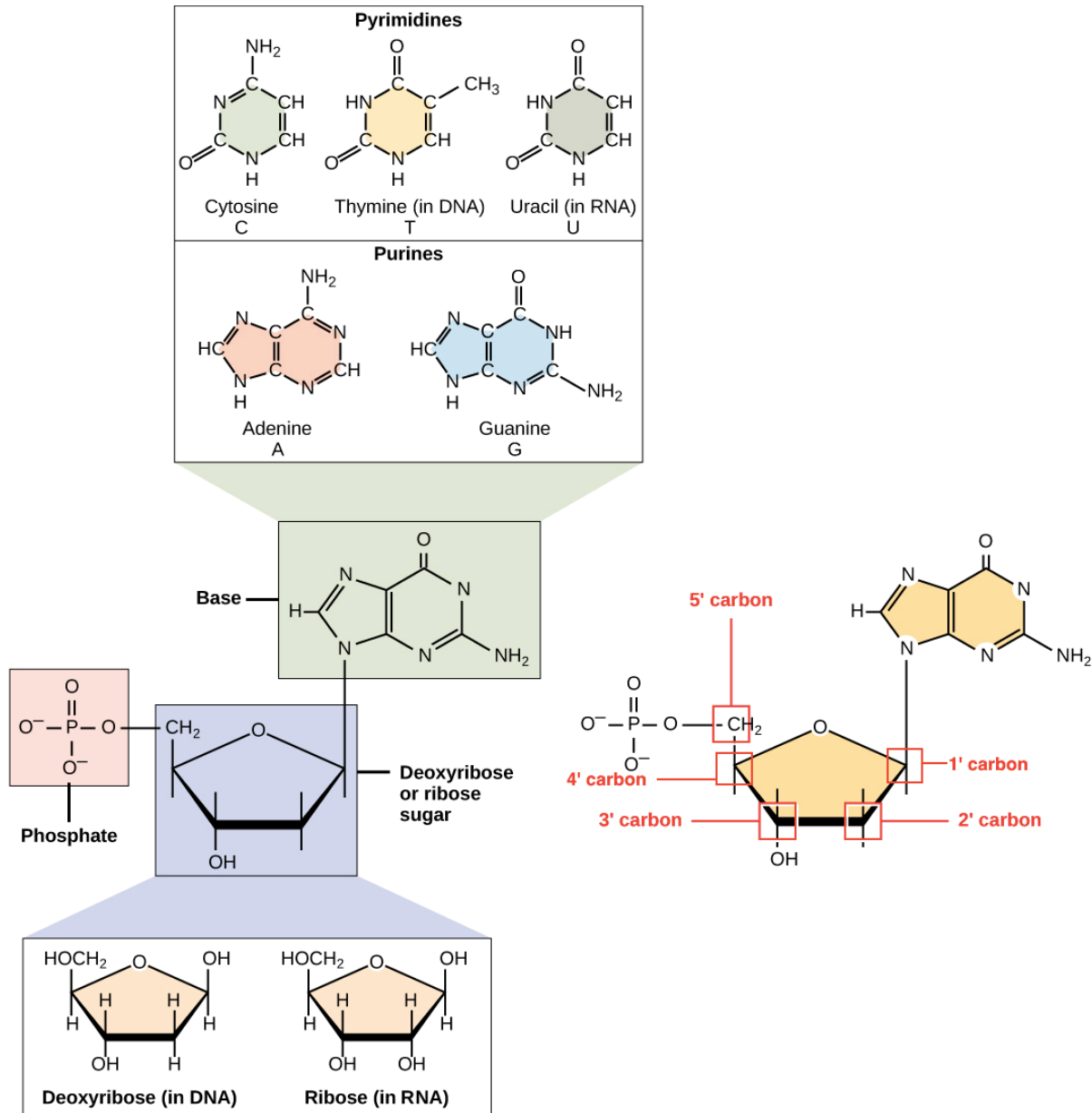


Image of the components of DNA and RNA, including the sugar (deoxyribose or ribose), phosphate group, and nitrogenous base. Bases include the pyrimidine bases (cytosine, thymine in DNA, and uracil in

RNA, one ring) and the purine bases (adenine and guanine, two rings). The phosphate group is attached to the 5' carbon. The 2' carbon bears a hydroxyl group in ribose, but no hydroxyl (just hydrogen) in deoxyribose.

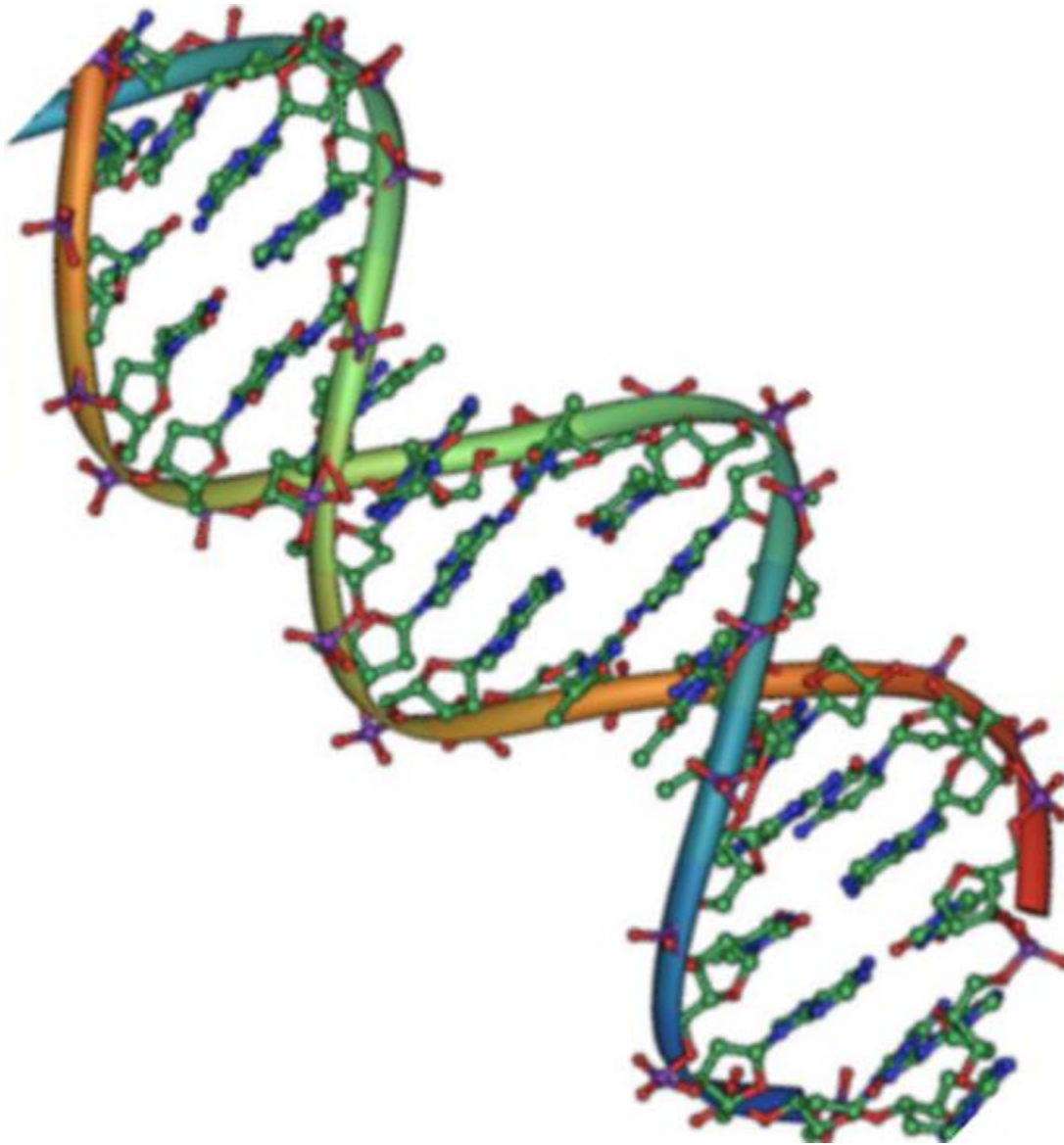
Image modified from "[Nucleic acids: Figure 1](#)," by OpenStax College, Biology ([CC BY 3.0](#)).

In addition to having slightly different sets of bases, DNA and RNA nucleotides also have slightly different sugars. The five-carbon sugar in DNA is called **deoxyribose**, while in RNA, the sugar is **ribose**. These two are very similar in structure, with just one difference: the second carbon of ribose bears a hydroxyl group, while the equivalent carbon of deoxyribose has a hydrogen instead.

Phosphate

Nucleotides may have a single phosphate group, or a chain of up to three phosphate groups, attached to the 5' carbon of the sugar. Some chemistry sources use the term “nucleotide” only for the single-phosphate case, but in molecular biology, the broader definition is generally accepted¹

In a cell, a nucleotide about to be added to the end of a polynucleotide chain will bear a series of three phosphate groups. When the nucleotide joins the growing DNA or RNA chain, it loses two phosphate groups. So, in a chain of DNA or RNA, each nucleotide has just one phosphate group.



Structural model of a DNA double helix.

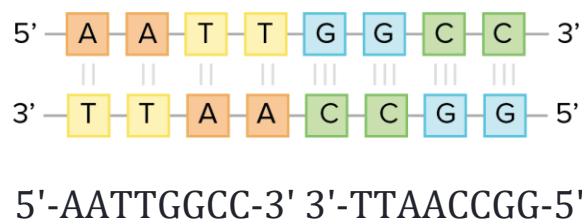
Image credit: Jerome Walker/Dennis Myts.

The two strands of the helix run in opposite directions, meaning that the 5' end of one strand is paired up with the 3' end of its matching strand. (This is referred to as **antiparallel** orientation and is important for the copying of DNA.)

So, can any two bases decide to get together and form a pair in the double helix? The answer is a definite no. Because of the sizes and

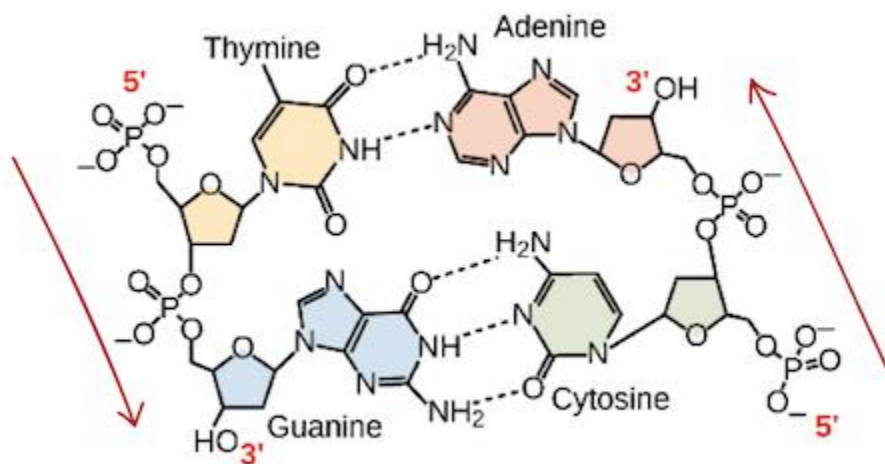
functional groups of the bases, base pairing is highly specific: A can only pair with T, and G can only pair with C, as shown below. This means that the two strands of a DNA double helix have a very predictable relationship to each other.

For instance, if you know that the sequence of one strand is 5'-AATTGGCC-3', the complementary strand must have the sequence 3'-TTAACCGG-5'. This allows each base to match up with its partner:



These two strands are complementary, with each base in one sticking to its partner on the other. The A-T pairs are connected by two hydrogen bonds, while the G-C pairs are connected by three hydrogen bonds.

When two DNA sequences match in this way, such that they can stick to each other in an antiparallel fashion and form a helix, they are said to be **complementary**.



Hydrogen bonding between complementary bases holds DNA strands together in a double helix of antiparallel strands. Thymine forms two hydrogen bonds with adenine, and guanine forms three hydrogen bonds with cytosine.

Image modified from OpenStax Biology.

Properties of RNA

Ribonucleic acid (RNA), unlike DNA, is usually single-stranded. A nucleotide in an RNA chain will contain ribose (the five-carbon sugar), one of the four nitrogenous bases (A, U, G, or C), and a phosphate group.

Summary: Features of DNA and RNA

	DNA	RNA
Function	Repository of genetic information	Involved in protein synthesis and gene regulation; carrier of genetic information in some viruses
Sugar	Deoxyribose	Ribose
Structure	Double helix	Usually single-stranded
Bases	C, T, A, G	C, U, A, G

Table modified from OpenStax Biology.

Prokaryotes are microscopic organisms belonging to the domains Bacteria and Archaea, which are two out of the three major domains of

life. (Eukarya, the third, contains all **eukaryotes**, including animals, plants, and fungi.) Bacteria and archaea are single-celled, while most eukaryotes are multicellular.

Fossils show that prokaryotes were already here on Earth 3.5 billion years ago, and scientists think that prokaryotic ancestors gave rise to all of the life forms present on Earth today

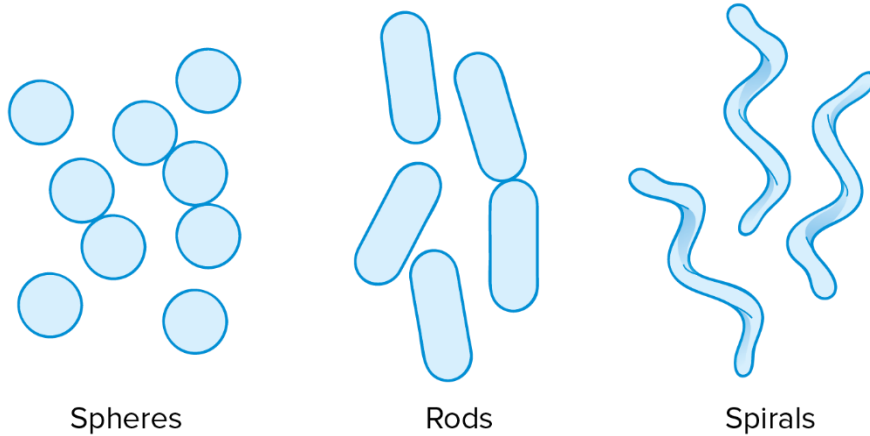
Prokaryotes vs. eukaryotes

Prokaryotes and eukaryotes are quite different. That may be obvious when comparing humans to bacteria. However, it's less obvious when we're comparing a bacterium to a yeast (which is tiny and unicellular, but eukaryotic). What actually separates these categories of organisms?

The most fundamental differences between prokaryotes and eukaryotes relate to how their cells are set up. Specifically:

- Eukaryotic cells have a **nucleus** while prokaryotic cells don't. This is the feature that formally separates the two groups.
- Eukaryotes usually have other membrane-bound organelles in addition to the nucleus, while prokaryotes don't.
- Cells in general are small, but prokaryotic cells are *really* small.

- Many prokaryotic cells have sphere, rod, or spiral shapes



Prokaryotic cells are typically shaped as either spheres (called cocci), rods (called bacilli), or spirals.

_Image modified from "[Bacterial morphology diagram](#)," by Mariana Ruiz Villareal (public domain). _

REPLICATION OF DNA

The basic idea

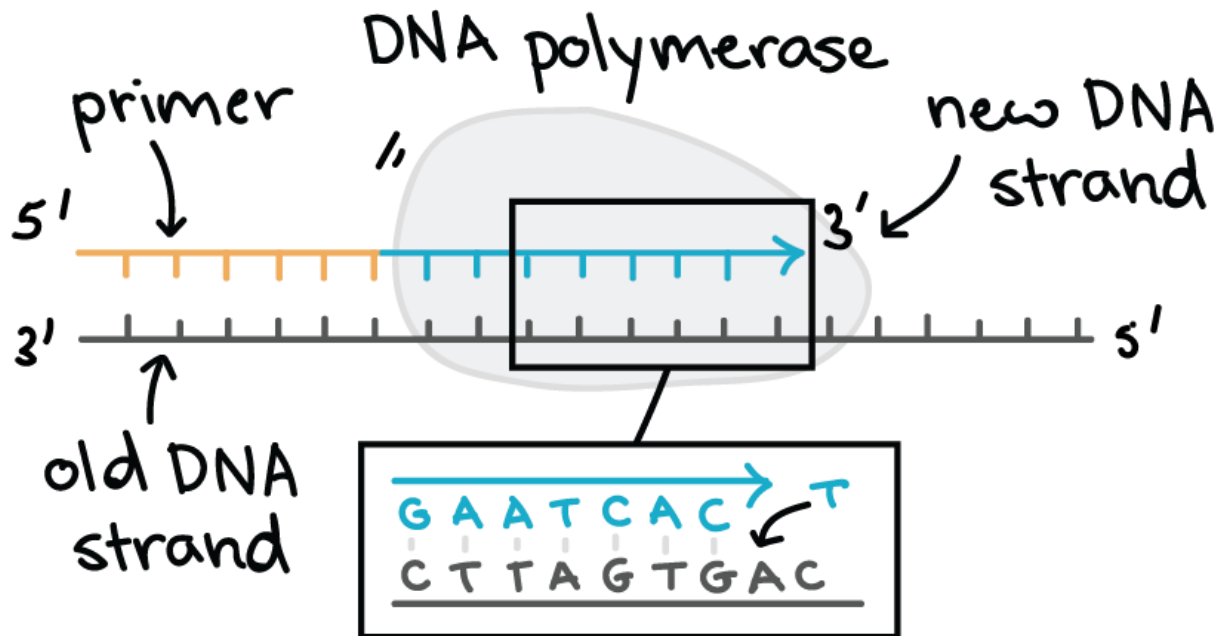
DNA replication is **semiconservative**, meaning that each strand in the DNA double helix acts as a template for the synthesis of a new, complementary strand.

This process takes us from one starting molecule to two "daughter" molecules, with each newly formed double helix containing one new and one old strand.

Cells need to copy their DNA very quickly, and with very few errors (or risk problems such as cancer). To do so, they use a variety of enzymes and proteins, which work together to make sure DNA replication is performed smoothly and accurately.

DNA polymerase

One of the key molecules in DNA replication is the enzyme **DNA polymerase**. DNA polymerases are responsible for synthesizing DNA: they add nucleotides one by one to the growing DNA chain, incorporating only those that are complementary to the template.



Here are some key features of DNA polymerases:

- They always need a template
- They can only add nucleotides to the 3' end of a DNA strand
- They can't start making a DNA chain from scratch, but require a pre-existing chain or short stretch of nucleotides called a **primer**
- They **proofread**, or check their work, removing the vast majority of "wrong" nucleotides that are accidentally added to the chain

The addition of nucleotides requires energy. This energy comes from the nucleotides themselves, which have three phosphates attached to them (much like the energy-carrying molecule ATP). When the bond between

phosphates is broken, the energy released is used to form a bond between the incoming nucleotide and the growing chain.

[\[See the polymerization reaction\]](#)

In prokaryotes such as *E. coli*, there are two main DNA polymerases involved in DNA replication: DNA pol III (the major DNA-maker), and DNA pol I, which plays a crucial supporting role we'll examine later.

Starting DNA replication

How do DNA polymerases and other replication factors know where to begin? Replication always starts at specific locations on the DNA, which are called **origins of replication** and are recognized by their sequence.

E. coli, like most bacteria, has a single origin of replication on its chromosome. The origin is about 245245245 base pairs long and has mostly A/T base pairs (which are held together by fewer hydrogen bonds than G/C base pairs), making the DNA strands easier to separate.

Specialized proteins recognize the origin and open up the DNA. As the DNA opens, two Y-shaped structures called **replication forks** are formed, together making up what's called a **replication bubble**. The replication forks will move in opposite directions as replication proceeds.

Bacterial chromosome. The double-stranded DNA of the circular bacteria chromosome is opened at the origin of replication, forming a replication bubble. Each end of the bubble is a replication fork, a Y-shaped junction where double-stranded DNA is separated into two single strands. New DNA complementary to each single strand is synthesized at each

replication fork. The two forks move in opposite directions around the circumference of the bacterial chromosome, creating a larger and larger replication bubble that grows at both ends.

How does replication actually get going at the forks? **Helicase** is the first replication enzyme to load on at the origin of replication. Helicase's job is to move the replication forks forward by "unwinding" the DNA (breaking the hydrogen bonds between the nitrogenous base pairs).

Proteins called **single-strand binding proteins** coat the separated strands of DNA near the replication fork, keeping them from coming back together into a double helix.

Primers and primase

DNA polymerases can only add nucleotides to the 3' end of an existing DNA strand. (They use the free -OH group found at the 3' end as a "hook," adding a nucleotide to this group in the polymerization reaction.) How, then, does DNA polymerase add the first nucleotide at a new replication fork?

Alone, it can't! The problem is solved with the help of an enzyme called **primase**. Primase makes an RNA **primer**, or short stretch of nucleic acid complementary to the template, that provides a 3' end for DNA polymerase to work on. A typical primer is about five to ten nucleotides long. The primer *primes* DNA synthesis.

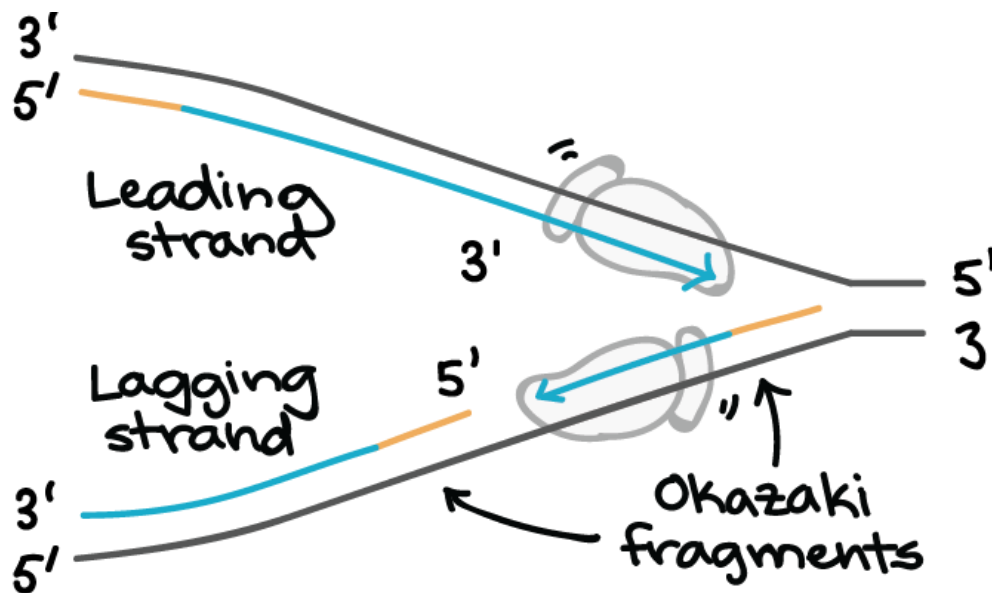
Once the RNA primer is in place, DNA polymerase "extends" it, adding nucleotides one by one to make a new DNA strand that's complementary to the template strand.

Leading and lagging strands

In *E. coli*, the DNA polymerase that handles most of the synthesis is DNA polymerase III. There are two molecules of DNA polymerase III at a replication fork, each of them hard at work on one of the two new DNA strands.

DNA polymerases can only make DNA in the 5' to 3' direction, and this poses a problem during replication. A DNA double helix is always antiparallel; in other words, one strand runs in the 5' to 3' direction, while the other runs in the 3' to 5' direction. This makes it necessary for the two new strands, which are also antiparallel to their templates, to be made in slightly different ways.

One new strand, which runs 5' to 3' towards the replication fork, is the easy one. This strand is made continuously, because the DNA polymerase is moving in the same direction as the replication fork. This continuously synthesized strand is called the **leading strand**.



The other new strand, which runs 5' to 3' away from the fork, is trickier. This strand is made in fragments because, as the fork moves forward, the DNA polymerase (which is moving away from the fork) must come off and reattach on the newly exposed DNA. This tricky strand, which is made in fragments, is called the **lagging strand**.

The small fragments are called **Okazaki fragments**, named for the Japanese scientist who discovered them. The leading strand can be extended from one primer alone, whereas the lagging strand needs a new primer for each of the short Okazaki fragments.

The maintenance and cleanup crew

Some other proteins and enzymes, in addition the main ones above, are needed to keep DNA replication running smoothly. One is a protein called the **sliding clamp**, which holds DNA polymerase III molecules in place as they synthesize DNA. The sliding clamp is a ring-shaped protein and keeps the DNA polymerase of the lagging strand from floating off when it re-starts at a new Okazaki fragment.

Topoisomerase also plays an important maintenance role during DNA replication. This enzyme prevents the DNA double helix ahead of the replication fork from getting too tightly wound as the DNA is opened up. It acts by making temporary nicks in the helix to release the tension, then sealing the nicks to avoid permanent damage.

Finally, there is a little cleanup work to do if we want DNA that doesn't contain any RNA or gaps. The RNA primers are removed and replaced by DNA through the activity of **DNA polymerase I**, the other polymerase

involved in replication. The nicks that remain after the primers are replaced get sealed by the enzyme **DNA ligase**.

Summary of DNA replication in *E. coli*

Let's zoom out and see how the enzymes and proteins involved in replication work together to synthesize new DNA.

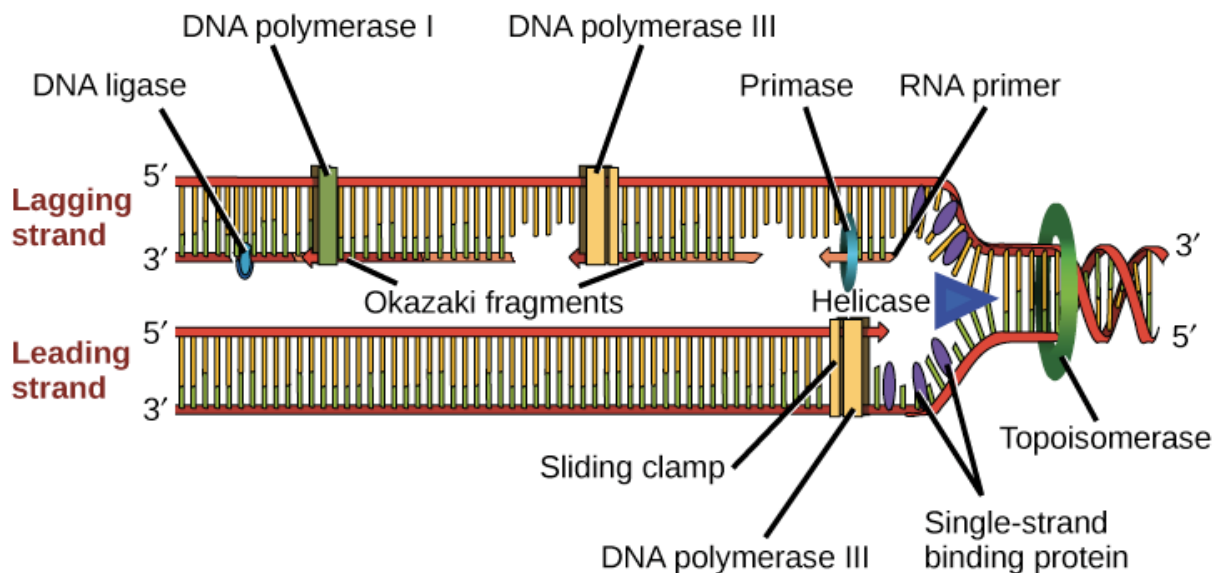


Illustration shows the replication fork. Helicase unwinds the helix, and single-strand binding proteins prevent the helix from re-forming. Topoisomerase prevents the DNA from getting too tightly coiled ahead of the replication fork. DNA primase forms an RNA primer, and DNA polymerase extends the DNA strand from the RNA primer. DNA synthesis occurs only in the 5' to 3' direction. On the leading strand, DNA synthesis occurs continuously. On the lagging strand, DNA synthesis restarts many times as the helix unwinds, resulting in many short fragments called “Okazaki fragments.” DNA ligase joins the Okazaki fragments together into a single DNA molecule.

- **Helicase** opens up the DNA at the replication fork.

- **Single-strand binding proteins** coat the DNA around the replication fork to prevent rewinding of the DNA.
- **Topoisomerase** works at the region ahead of the replication fork to prevent supercoiling.
- **Primase** synthesizes RNA primers complementary to the DNA strand.
- **DNA polymerase III** extends the primers, adding on to the 3' end, to make the bulk of the new DNA.
- RNA primers are removed and replaced with DNA-by-DNA **polymerase I**.
- The gaps between DNA fragments are sealed by **DNA ligase**.

DNA replication in eukaryotes

The basics of DNA replication are similar between bacteria and eukaryotes such as humans, but there are also some differences:

- Eukaryotes usually have multiple linear chromosomes, each with multiple origins of replication. Humans can have up to 100,100,100,000,000,000 origins of replication.
- Most of the *E. coli* enzymes have counterparts in eukaryotic DNA replication, but a single enzyme in *E. coli* may be represented by multiple enzymes in eukaryotes. For instance, there are five human DNA polymerases with important roles in replication.
- Most eukaryotic chromosomes are linear. Because of the way the lagging strand is made, some DNA is lost from the ends of linear chromosomes (the [telomeres](#)) in each round of replication.