GENETICS

By Jamie

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.

THE BEGINNING: DNA & RNA

Introduction

Nucleic acids, and DNA in particular, are key macromolecules for the continuity of life. DNA bears the hereditary information that's passed on from parents to children, providing instructions for how (and when) to make the many proteins needed to build and maintain functioning cells, tissues, and organisms.

How DNA carries this information, and how it is put into action by cells and organisms, is complex, fascinating, and fairly mind-blowing, and we'll explore it in more detail in the section on molecular biology. Here, we'll just take a quick look at nucleic acids from the macromolecule perspective.

Roles of DNA and RNA in cells

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).

DNA in cells

In eukaryotes, such as plants and animals, DNA is found in the **nucleus**, a specialized, membrane-bound vault in the cell, as well as in certain other

types of organelles (such as mitochondria and the chloroplasts of plants). 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**.

In eukaryotes, DNA is typically broken up into a number of very long, linear pieces called **chromosomes**, while in prokaryotes such as bacteria, chromosomes are much smaller and often circular (ring-shaped). A chromosome may contain tens of thousands of **genes**, each providing instructions on how to make a particular product needed by the cell.

From DNA to RNA to proteins

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. This progression from DNA to RNA to protein is called the "central dogma" of molecular biology.

Importantly, not all genes encode protein products. For instance, some genes specify **ribosomal RNAs** (**rRNAs**), which serve as structural components of ribosomes, or **transfer RNAs** (**tRNAs**), cloverleaf-shaped RNA molecules that bring amino acids to the ribosome for protein synthesis. Still other RNA molecules, such as tiny **microRNAs** (**miRNAs**), act as regulators of other genes, and new types of non-protein-coding RNAs are being discovered all the time.

Nucleotides

DNA and RNA are polymers (in the case of DNA, often very long 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.

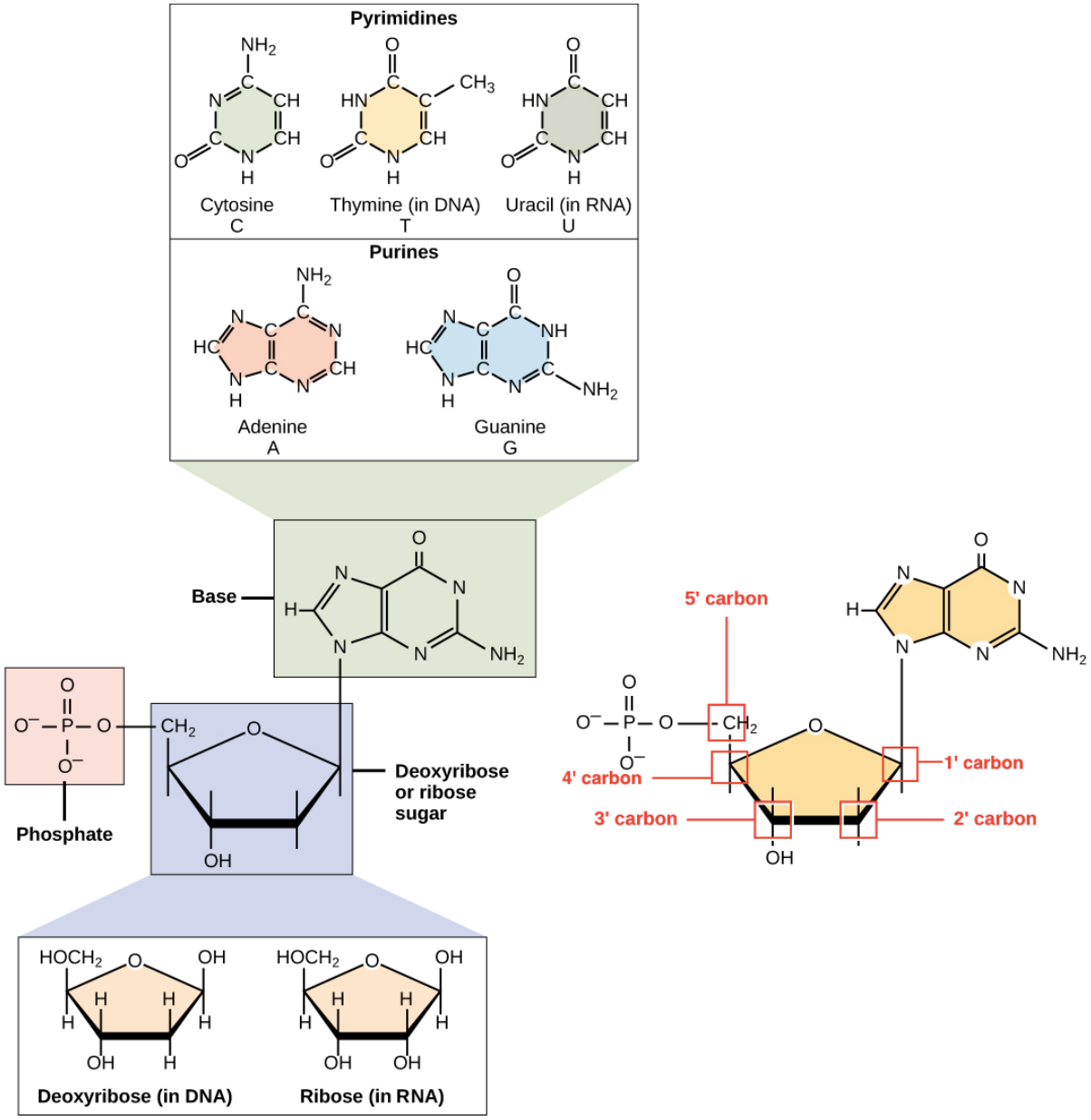


Image of the components of DNA and RNA, including the sugar (deoxyribose or ribose), phosphate group, and nitrogenous base. Bases include the pyrimidine bases (cytosine, thymine in DNA, and uracil in RNA, one ring) and the purine bases (adenine and guanine, two rings). The phosphate group is attached 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).

Nitrogenous bases

The nitrogenous bases of nucleotides are organic (carbon-based) molecules made up of nitrogen-containing ring structures.
[Why is it called a base?]

Each nucleotide in DNA contains one of four possible nitrogenous bases: adenine (A), guanine (G) cytosine (C), and thymine (T). Adenine and guanine are **purines**, meaning that their structures contain two fused carbon-nitrogen rings. Cytosine and thymine, in contrast, are **pyrimidines** and have a single carbon-nitrogen ring. RNA nucleotides may also bear adenine, guanine and cytosine bases, but instead of thymine they have another pyrimidine base called uracil (U). As shown in the figure above, each base has a unique structure, with its own set of functional groups attached to the ring structure.

In molecular biology shorthand, the nitrogenous bases are often just referred to by their one-letter symbols, A, T, G, C, and U. DNA contains A, T, G, and C, while RNA contains A, U, G, and C (that is, U is swapped in for T).

Sugars

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. The carbon atoms of a nucleotide's sugar molecule are numbered as 1′, 2′, 3′, 4′, and 5′ (1′ is read as "one prime"), as shown in the figure above. In a nucleotide, the sugar occupies

a central position, with the base attached to its 1′ carbon and the phosphate group (or groups) attached to its 5′ carbon.

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 accepted^11start superscript, 1, end superscript

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.

Polynucleotide chains

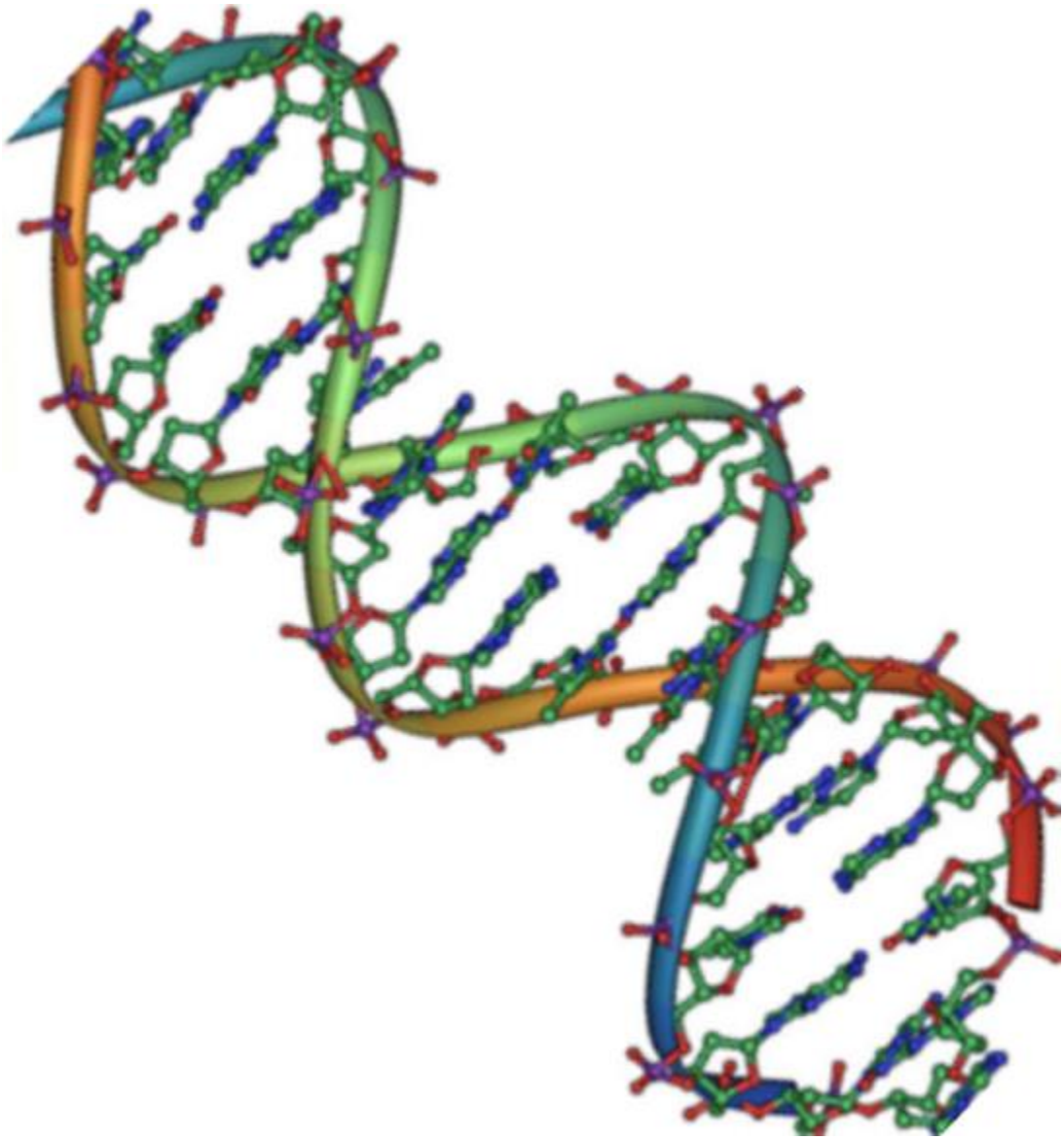
A consequence of the structure of nucleotides is that a polynucleotide chain has **directionality** – that is, it has two ends that are different from each other. At the **5’ end**, or beginning, of the chain, the 5’ phosphate group of the first nucleotide in the chain sticks out. At the other end, called the **3’ end**, the 3’ hydroxyl of the last nucleotide added to the chain is exposed. DNA sequences are usually written in the 5' to 3' direction, meaning that the nucleotide at the 5' end comes first and the nucleotide at the 3' end comes last.

As new nucleotides are added to a strand of DNA or RNA, the strand grows at its 3’ end, with the 5′ phosphate of an incoming nucleotide attaching to the hydroxyl group at the 3’ end of the chain. This makes a

chain with each sugar joined to its neighbors by a set of bonds called a **phosphodiester linkage**.

Properties of DNA

Deoxyribonucleic acid, or DNA, chains are typically found in a **double helix**, a structure in which two matching (complementary) chains are stuck together, as shown in the diagram at left. The sugars and phosphates lie on the outside of the helix, forming the backbone of the DNA; this portion of the molecule is sometimes called the **sugar-phosphate** backbone. The nitrogenous bases extend into the interior, like the steps of a staircase, in pairs; the bases of a pair are bound to each other by hydrogen bonds.



Structural model of a DNA double helix.

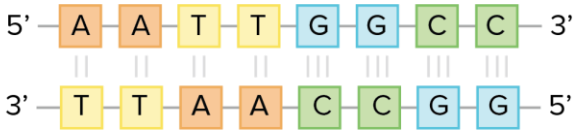
Image credit: Jerome Walker/Dennis Myts.

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.

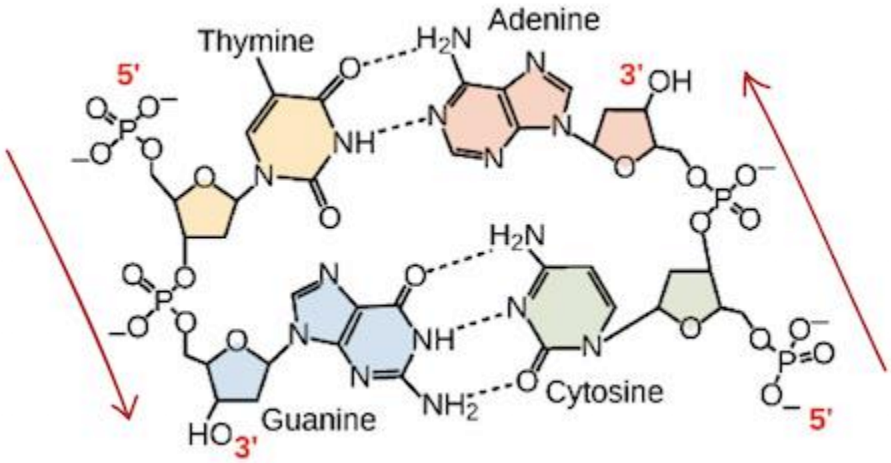
For instance, if you know that the sequence of one strand is 5'-AATTGGCC-3', the complementary strand must have the sequence 3'-TTAACCGG-5'. This allows each base to match up with its partner:



5'-AATTGGCC-3' 3'-TTAACCGG-5'

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. Here, we'll take a look at four major types of RNA: messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), and regulatory RNAs.

Messenger RNA (mRNA)

Messenger RNA (**mRNA**) is an intermediate between a protein-coding gene and its protein product. If a cell needs to make a particular protein, the gene encoding the protein will be turned “on,” meaning an RNA-polymerizing enzyme will come and make an RNA copy, or transcript, of the gene’s DNA sequence. The transcript carries the same information as the DNA sequence of its gene. However, in the RNA molecule, the base T is replaced with U. For instance, if a DNA coding strand has the sequence 5’-AATTGCGC-3’, the sequence of the corresponding RNA will be 5’-AAUUGCGC-3’.

Once an mRNA has been produced, it will associate with a ribosome, a molecular machine that specializes in assembling proteins out of amino

acids. The ribosome uses the information in the mRNA to make a protein of a specific sequence, “reading out” the mRNA’s nucleotides in groups of three (called **codons**) and adding a particular amino acid for each codon.

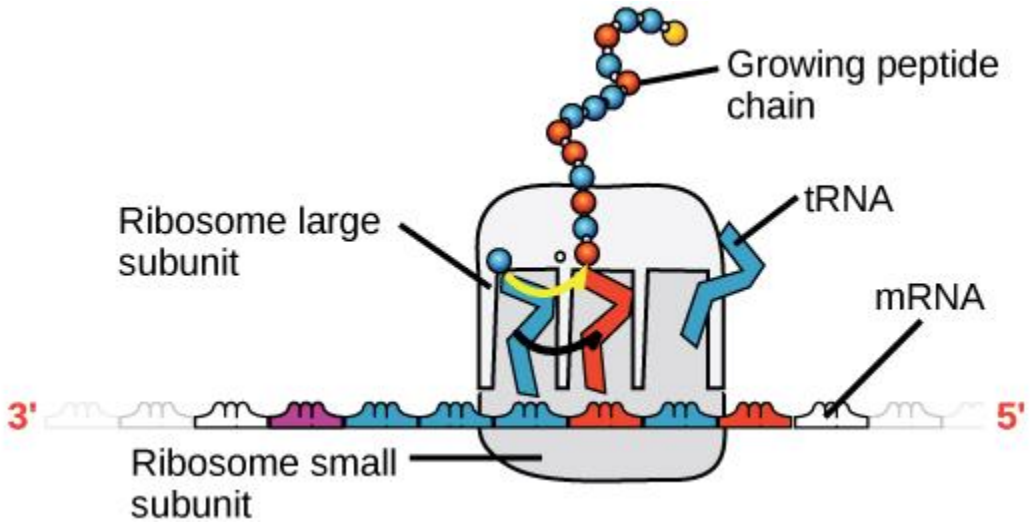


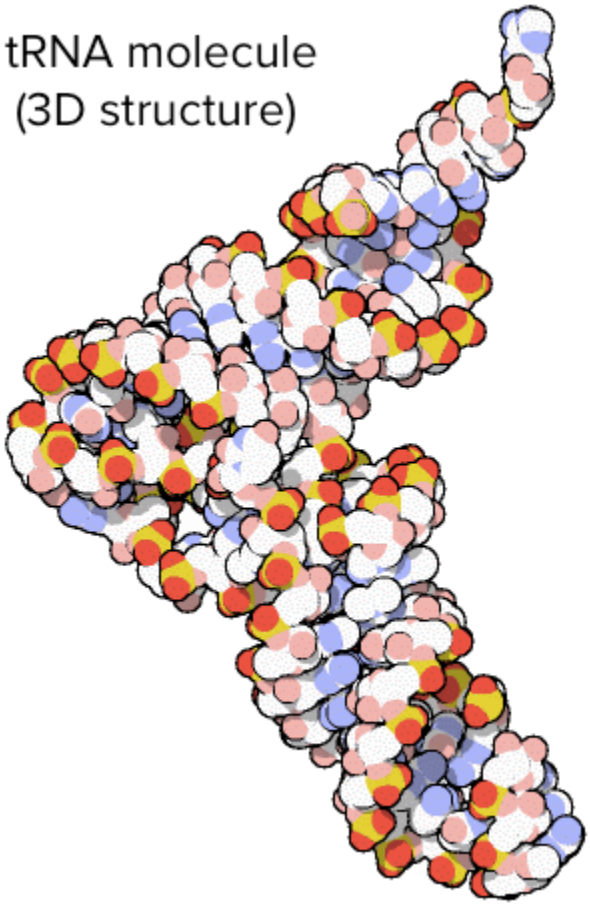
Image of a ribosome (made of proteins and rRNA) bound to an mRNA, with tRNAs bringing amino acids to be added to the growing chain. The tRNA that binds, and thus the amino acid that's added, at a given moment is determined by the sequence of the mRNA that is being "read" at that time.

Image credit: OpenStax Biology.

Ribosomal RNA (rRNA) and transfer RNA (tRNA)

Ribosomal RNA (**rRNA**) is a major component of ribosomes, where it helps mRNA bind in the right spot so its sequence information can be read out. Some rRNAs also act as enzymes, meaning that they help accelerate (catalyze) chemical reactions – in this case, the formation of bonds that link amino acids to form a protein. RNAs that act as enzymes are known as **ribozymes**.

Transfer RNAs (**tRNAs**) are also involved in protein synthesis, but their job is to act as carriers – to bring amino acids to the ribosome, ensuring that the amino acid added to the chain is the one specified by the mRNA. Transfer RNAs consist of a single strand of RNA, but this strand has complementary segments that stick together to make double-stranded regions. This base-pairing creates a complex 3D structure important to the function of the molecule.



Structure of a tRNA. The overall molecule has a shape somewhat like an L.

Image modified from Protein Data Bank (work of the U.S. government).

Regulatory RNA (miRNAs and siRNAs)

Some types of non-coding RNAs (RNAs that do not encode proteins) help regulate the expression of other genes. Such RNAs may be called regulatory RNAs. For example, **microRNAs** (**miRNAs**) and **small interfering RNAs siRNAs** are small regulatory RNA molecules about 22 nucleotides long. They bind to specific mRNA molecules (with partly or fully complementary sequences) and reduce their stability or interfere with their translation, providing a way for the cell to decrease or fine-tune levels of these mRNAs.

These are just some examples out of many types of noncoding and regulatory RNAs. Scientists are still discovering new varieties of noncoding RNA.

[More about regulatory RNAs]

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.

MOVING ON: PROKARYOTES

Key points:

- **Prokaryotes** are single-celled organisms belonging to the domains Bacteria and Archaea.
- Prokaryotic cells are much smaller than eukaryotic cells, have no nucleus, and lack organelles.
- All prokaryotic cells are encased by a **cell wall**. Many also have a **capsule** or **slime layer** made of polysaccharide.
- Prokaryotes often have appendages (protrusions) on their surface. **Flagella** and some **pili** are used for

locomotion, **fimbriae** help the cell stick to a surface, and **sex pili** are used for DNA exchange.
- Most prokaryotic cells have a single circular chromosome. They may also have smaller pieces of circular DNA called **plasmids**.

Introduction

Bacteria often get a bad rap: they're described as unsafe "bugs" that cause disease. Although some types of bacteria do cause disease (as you know if you've ever been prescribed antibiotics), many other are harmless, or even beneficial.

Bacteria are classified as **prokaryotes**, along with another group of single-celled organisms, the archaea. Prokaryotes are tiny, but in a very real sense, they dominate the Earth. They live nearly everywhere – on every surface, on land and in water, and even inside of our bodies.

To emphasize that last point: you probably have about the same number of prokaryotic cells in your body as human cells^1₁start superscript, 1, end superscript! That may sound gross, but many of our prokaryotic "sidekicks" play important roles in keeping us healthy.

In this article, we'll look at what prokaryotes are and what exactly makes them different from eukaryotes (such as you, a houseplant, or a fungus). Then, we'll take a closer look at the structures these efficient, omnipresent little organisms use to survive.

What are prokaryotes?

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 3.53.53, point, 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

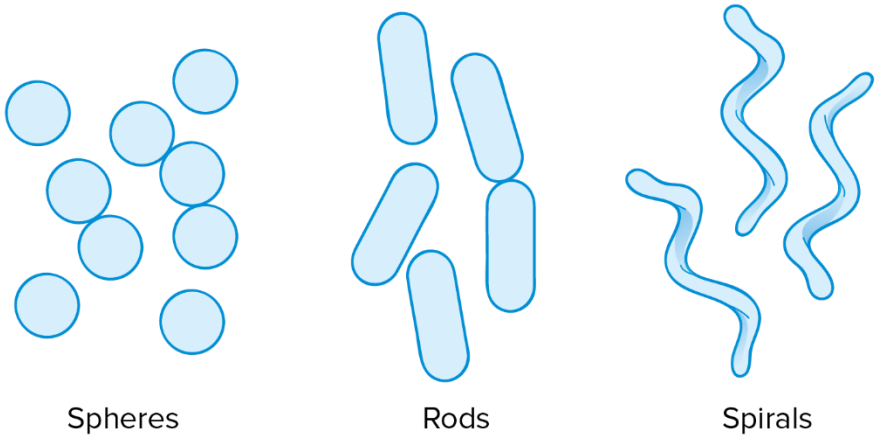
Prokaryotes and eukaryotes are similar in some fundamental ways, reflecting their shared evolutionary ancestry. For instance, both you and the bacteria in your gut decode genes into proteins through transcription and translation(Opens in a new window). Similarly, you and your

prokaryotic inhabitants both pass genetic information on to your offspring in the form of DNA.

In other ways, prokaryotes and eukaryotes are quite different. That may be obvious when we're comparing humans to bacteria. But for me at least, 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**, a membrane-bound chamber where DNA is stored, 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 (as shown below). In the following sections, we'll walk through the structure of a prokaryotic cell, starting on the outside and moving towards the inside of the cell.



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).

The capsule

Many prokaryotes have a sticky outermost layer called the **capsule**, which is usually made of polysaccharides (sugar polymers).

The capsule helps prokaryotes cling to each other and to various surfaces in their environment, and also helps prevent the cell from drying out. In the case of disease-causing prokaryotes that have colonized the body of a host organism, the capsule or slime layer may also protect against the host's immune system.

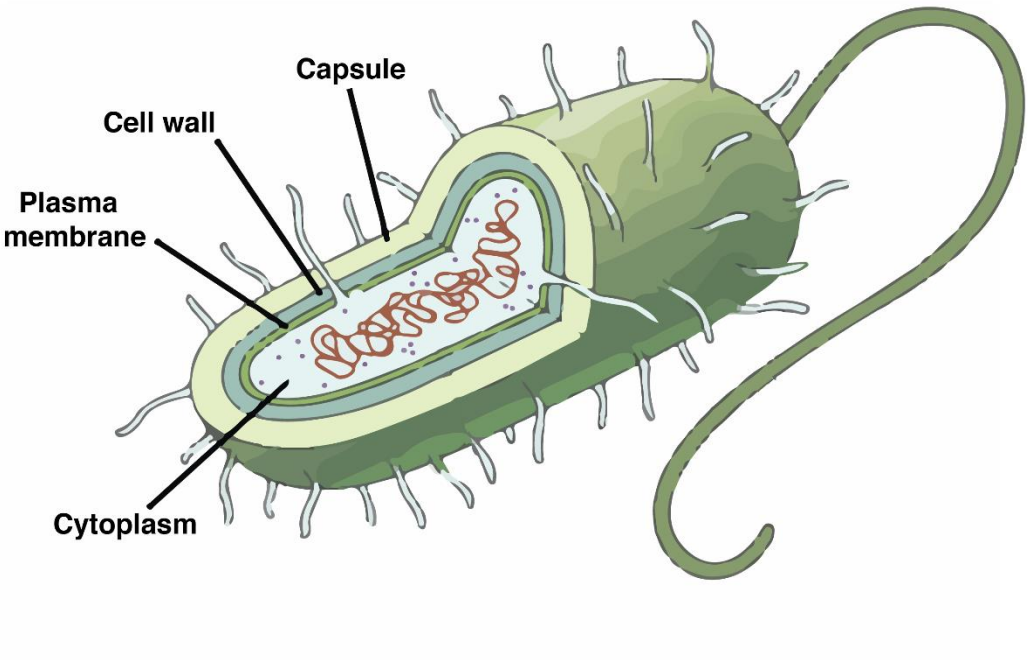
Remember Griffith's experiment, which demonstrated the existence of a "transforming principle" (DNA) that could turn rough, harmless bacteria into smooth, pathogenic bacteria? The smooth bacteria were smooth (and capable of causing disease) because they had a capsule!

The cell wall

All prokaryotic cells have a stiff **cell wall**, located underneath the capsule (if there is one). This structure maintains the cell's shape, protects the cell interior, and prevents the cell from bursting when it takes up water.

The cell wall of most bacteria contains **peptidoglycan**, a polymer of linked sugars and polypeptides. Peptidoglycan is unusual in that it contains not only L-amino acids, the type normally used to make proteins, but also D-amino acids ("mirror images" of the L-amino acids).

Archaeal cell walls don't contain peptidoglycan, but some include a similar molecule called pseudopeptidoglycan, while others are composed of proteins or other types of polymers



The external structures of the prokaryotic cell include a plasma membrane, cell wall, and capsule (or slime layer).

Image modified from "Structure of Prokaryotes: Figure 2," by OpenStax College, Biology (CC BY 3.0).

Some of the antibiotics used to treat bacterial infections in humans and other animals act by targeting the bacterial cell wall. For instance, some antibiotics contain D-amino acids similar to those used in peptidoglycan synthesis, "faking out" the enzymes that build the bacterial cell wall (but not affecting human cells, which don't have a cell wall or utilize D-amino acids to make polypeptides)

[Cell walls: Gram-positive and gram-negative bacteria]

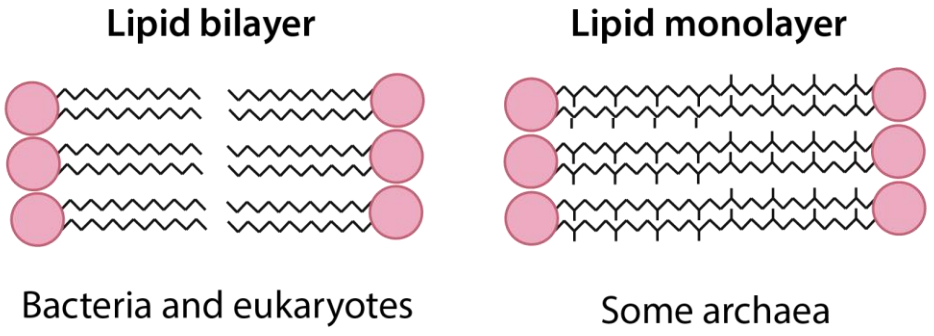
The plasma membrane

Underneath the cell wall lies the **plasma membrane**. The basic building block of the plasma membrane is the **phospholipid**, a lipid composed of a glycerol molecule attached a hydrophilic (water-attracting) phosphate head and to two hydrophobic (water-repelling) fatty acid tails. The phospholipids of a eukaryotic or bacterial membrane are organized into two layers, forming a structure called a **phospholipid bilayer**.

[See a diagram]

The plasma membranes of archaea have some unique properties, different from those of both bacteria and eukaryotes. For instance, in some species, the opposing phospholipid tails are joined into a single tail, forming a monolayer instead of a bilayer (as shown below). This modification may stabilize the membrane at high temperatures, allowing the archaea to live happily in boiling hot springs.

[More about archaeal plasma membranes]



The plasma membrane of bacterial and eukaryotic (and some archaeal) cells is composed of a phospholipid bilayer. The tails of opposite-facing phospholipids remain separated, forming two separate layers.

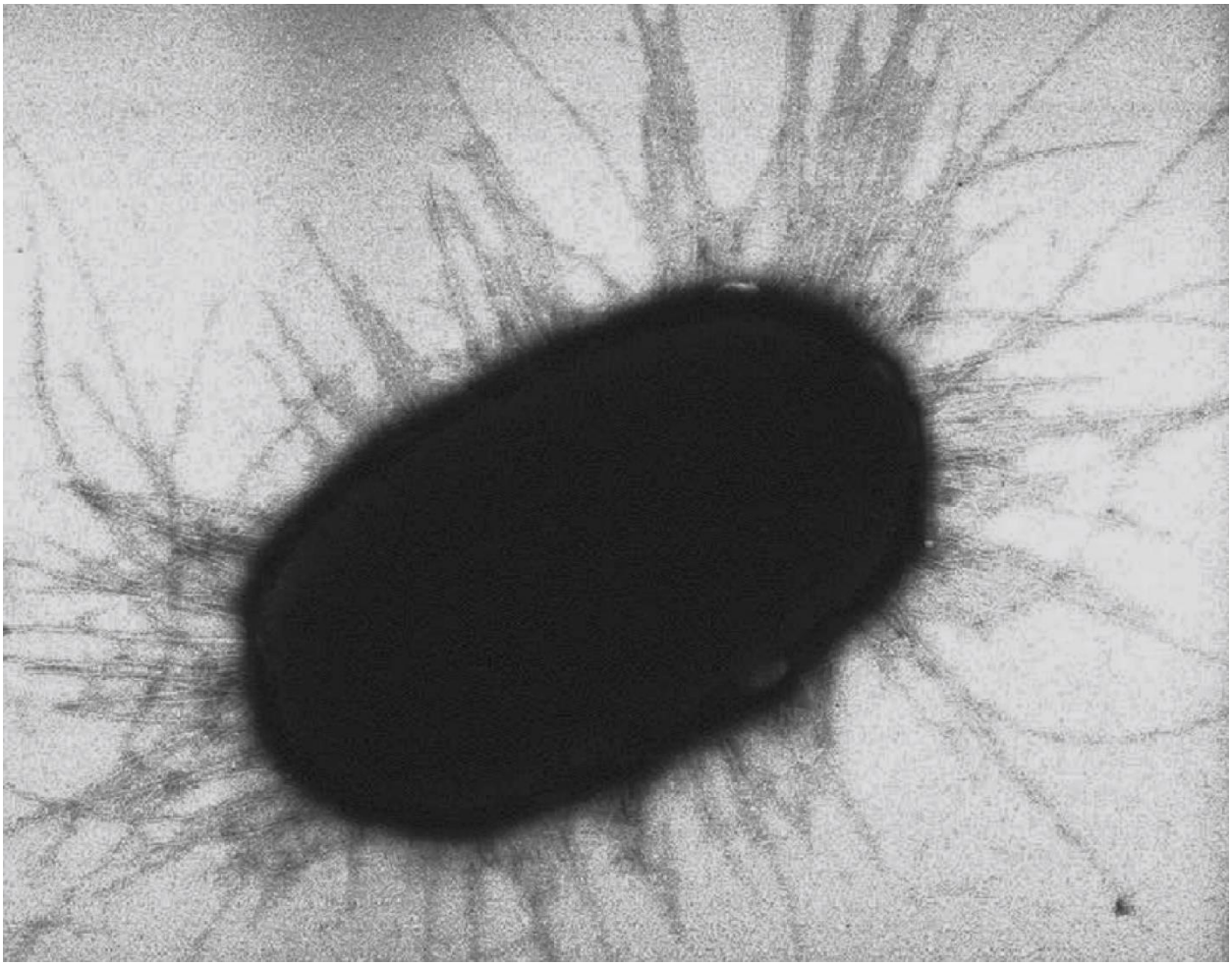
The plasma membrane of some archaeal cells is composed of a phospholipid monolayer. The tails of opposite-facing phospholipids become united, forming a single layer.

Image modified from "Archaea membrane," by Fransciscosp2 (public domain).

Appendages

Prokaryotic cells often have appendages (protrusions from the cell surface) that allow the cell to stick to surfaces, move around, or transfer DNA to other cells.

Thin filaments called **fimbriae** (singular: **fimbria**), like those shown in the picture below, are used for adhesion—that is, they help cells stick to objects and surfaces in their environment.



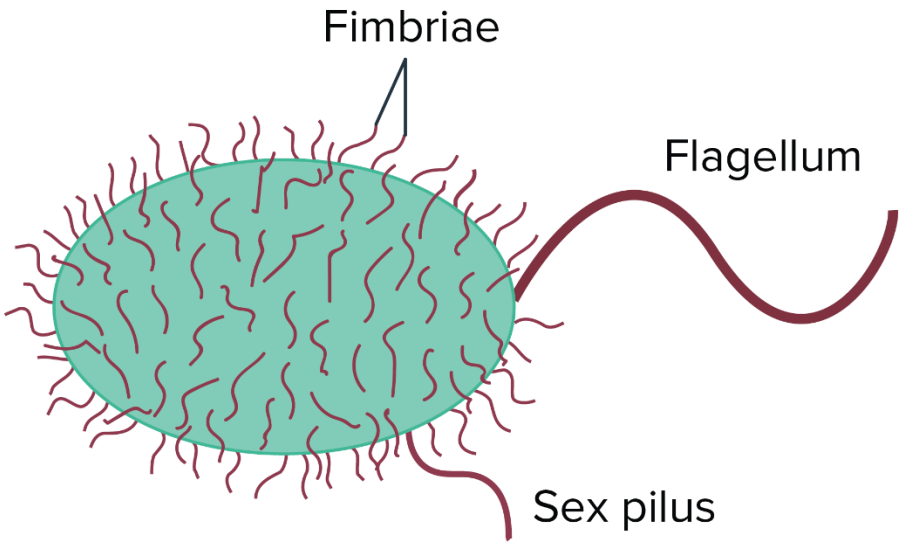
A fimbria (plural: fimbriae) is a type of appendage of prokaryotic cells. These hair-like protrusions allow prokaryotes to stick to surfaces in their environment and to each other.

Image modified from "E. coli fimbriae.png," by Manu Forero (CC BY 2.5).

Longer appendages, called **pili** (singular: **pilus**), come in several types that have different roles. For instance, a **sex pilus** holds two bacterial cells together and allows DNA to be transferred between them in a process called conjugation. Another class of bacterial pili, called **type IV pili**, help the bacterium move around its environment

The most common appendages used for getting around, however, are **flagella** (singular: **flagellum**). These tail-like structures whip around like propellers to move cells through watery environments.

[Can't eukaryotic cells have flagella too?]



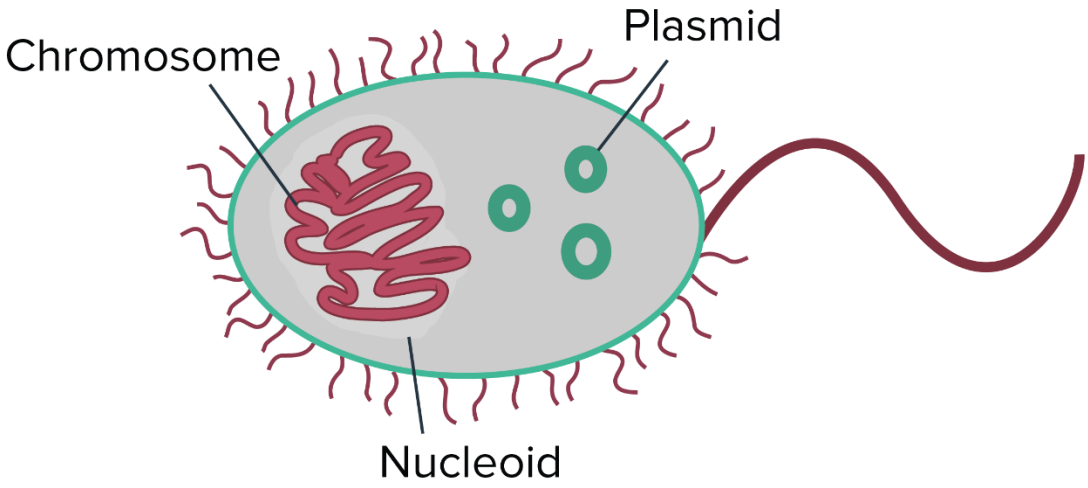
Bacteria may have various types of surface structures. These include fimbriae, short protrusions found all over the surface of the bacterium; a flagellum, found at the back of the bacterium and used for propulsion; and a sex pilus, used to grab on to other bacteria for exchange of genetic material.

Chromosome and plasmids

Most prokaryotes have a single circular chromosome, and thus a single copy of their genetic material. Eukaryotes like humans, in contrast, tend to have multiple rod-shaped chromosomes and two copies of their genetic material (on homologous chromosomes).

Also, prokaryotic genomes are generally much smaller than eukaryotic genomes. For instance, the *E. coli* genome is less than half the size of the genome of yeast (a simple, single-celled eukaryote), and almost 700700700 times smaller than the human genome!

By definition, prokaryotes lack a membrane-bound nucleus to hold their chromosomes. Instead, the chromosome of a prokaryote is found in a part of the cytoplasm called a **nucleoid**.



Prokaryotes generally have a single circular chromosome that occupies a region of the cytoplasm called a nucleoid. They also may contain small rings of double-stranded extra-chromosomal DNA called plasmids.

In addition to the chromosome, many prokaryotes have **plasmids**, which are small rings of double-stranded extra-chromosomal ("outside the chromosome") DNA. Plasmids carry a small number of non-essential genes and are copied independently of the chromosome inside the cell. They can be transferred to other prokaryotes in a population, sometimes spreading genes that are beneficial to survival.

For instance, some plasmids carry genes that make bacteria resistant to antibiotics. (These genes are called **R genes**.) When the plasmids carrying R genes are exchanged in a population, they can quickly make the population resistant to antibiotic drugs. While beneficial to the bacteria, this process can make it difficult for doctors to treat harmful bacterial infections.

[More on the antibiotic resistance problem]

Internal compartments

Prokaryotes aren't "supposed" to have internal compartments like the organelles of eukaryotes, and for the most part, they don't. However, prokaryotic cells sometimes need to increase membrane surface area for reactions or concentrate a substrate around its enzyme, just like eukaryotic cells. Because of this, some prokaryotes have membrane folds or compartments functionally similar to those of eukaryotes.

For example, photosynthetic bacteria often have extensive membrane folds to increase surface area for the light-dependent reactions, similar to the thylakoid membranes of a plant cell. These bacteria may also have carboxysomes, protein-enclosed cellular compartments where carbon dioxide is concentrated for fixation in the Calvin cycle

REPLICATION OF DNA

Key points:

- DNA replication is **semiconservative**. Each strand in the double helix acts as a template for synthesis of a new, complementary strand.
- New DNA is made by enzymes called **DNA polymerases**, which require a template and a **primer** (starter) and synthesize DNA in the 5' to 3' direction.
- During DNA replication, one new strand (the **leading strand**) is made as a continuous piece. The other (the **lagging strand**) is made in small pieces.

- DNA replication requires other enzymes in addition to DNA polymerase, including **DNA primase**, **DNA helicase**, **DNA ligase**, and **topoisomerase**.

Introduction

DNA replication, or the copying of a cell's DNA, is no simple task! There are about 333 billion base pairs of DNA in your genome, all of which must be accurately copied when any one of your trillions of cells divides.

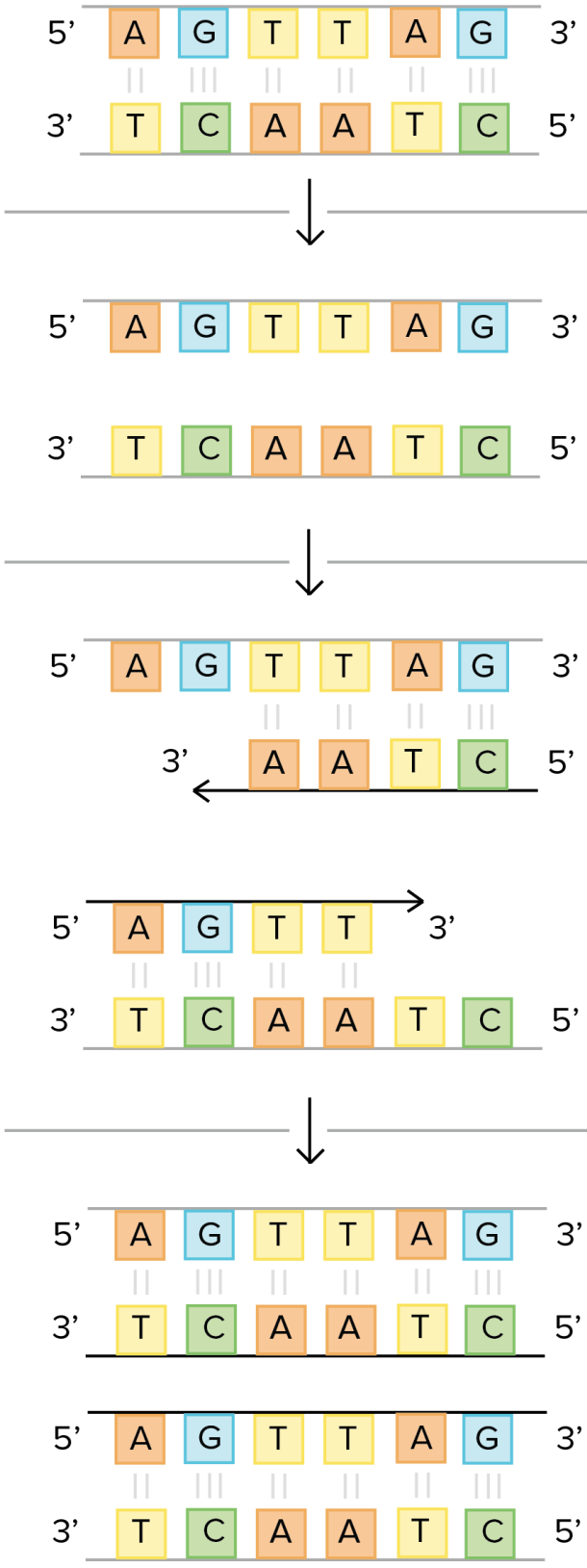
The basic mechanisms of DNA replication are similar across organisms. In this article, we'll focus on DNA replication as it takes place in the bacterium *E. coli*, but the mechanisms of replication are similar in humans and other eukaryotes.

Let's take a look at the proteins and enzymes that carry out replication, seeing how they work together to ensure accurate and complete 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.


5’ A G T T A G 3’
3’ T C A A T C 5’
DNA double helix
5’ A G T T A G 3’
3’ T C A A T C 5’
Hydrogen bonds break and helix opens
5’ A G T T A G 3’
3’ A A T C 5’
5’ A G T T 3’
3’ T C A A T C 5’
Each strand of DNA acts as a template for synthesis of a new, complementary strand
5’ A G T T A G 3’
3’ T C A A T C 5’
5’ A G T T A G 3’
3’ T C A A T C 5’
Replication produces two identical DNA double helices, each with one new and one old strand

Schematic of Watson and Crick's basic model of DNA replication.

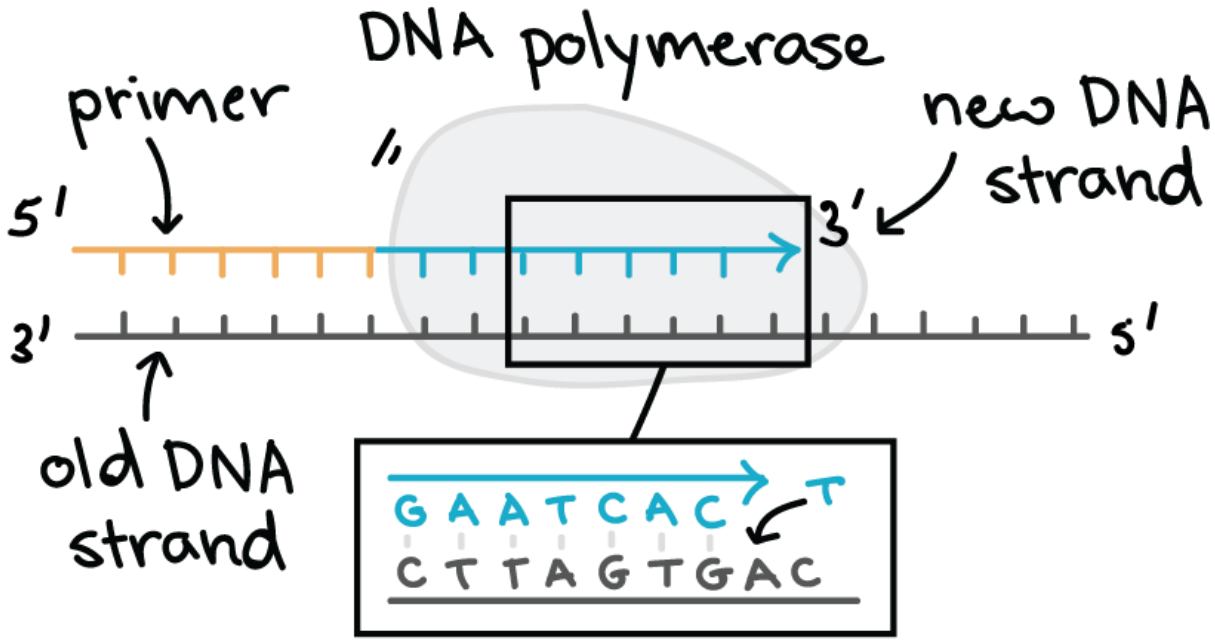
1. DNA double helix.
2. Hydrogen bonds break and helix opens.
3. Each strand of DNA acts as a template for synthesis of a new, complementary strand.
4. Replication produces two identical DNA double helices, each with one new and one old strand.

In a sense, that's all there is to DNA replication! But what's actually most interesting about this process is how it's carried out in a cell.

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]

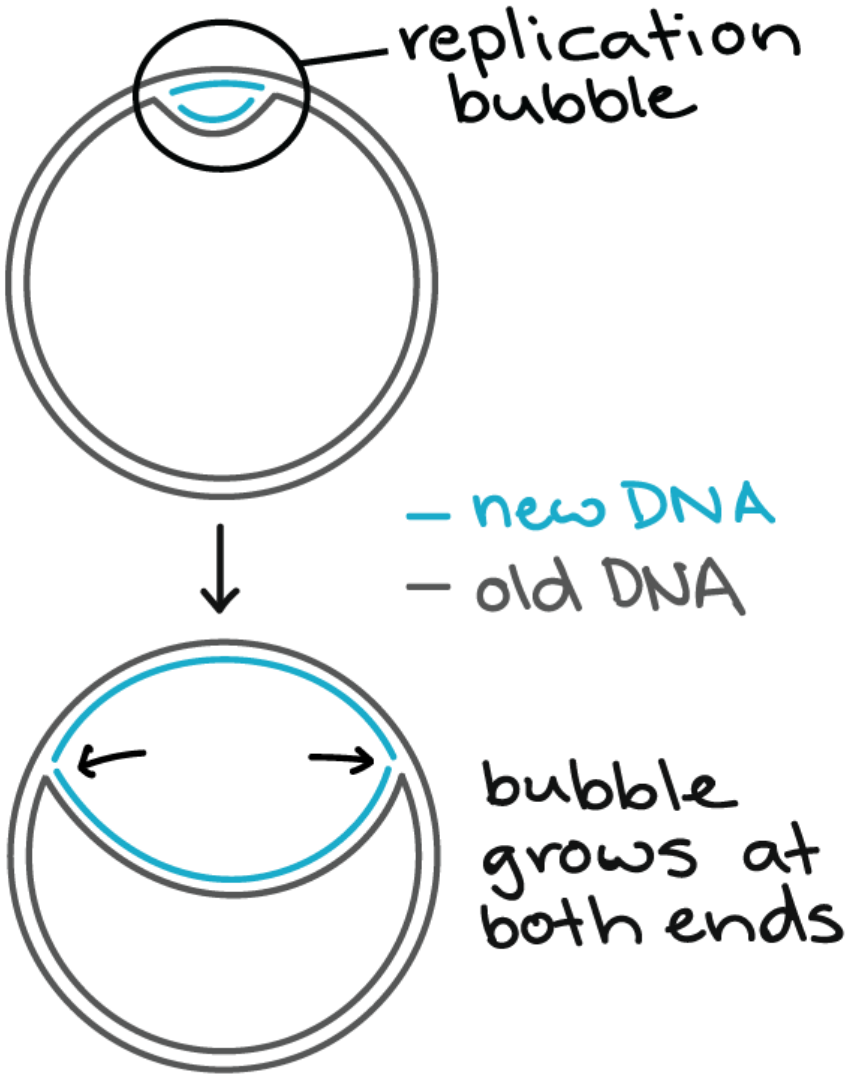
In prokaryotes such as *E. coli*, there are two main DNA polymerases involved in DNA replication: DNA pol III (the major DNA-maker), and DNA pol I, which plays a crucial supporting role we'll examine later.

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, bind to this site, 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.

Diagram based on similar illustration in Reece et al. ^2₂squared.

How does replication actually get going at the forks? **Helicase** is the first replication enzyme to load on at the origin of replication^3$_3$cubed. 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, i.e., gets it started.

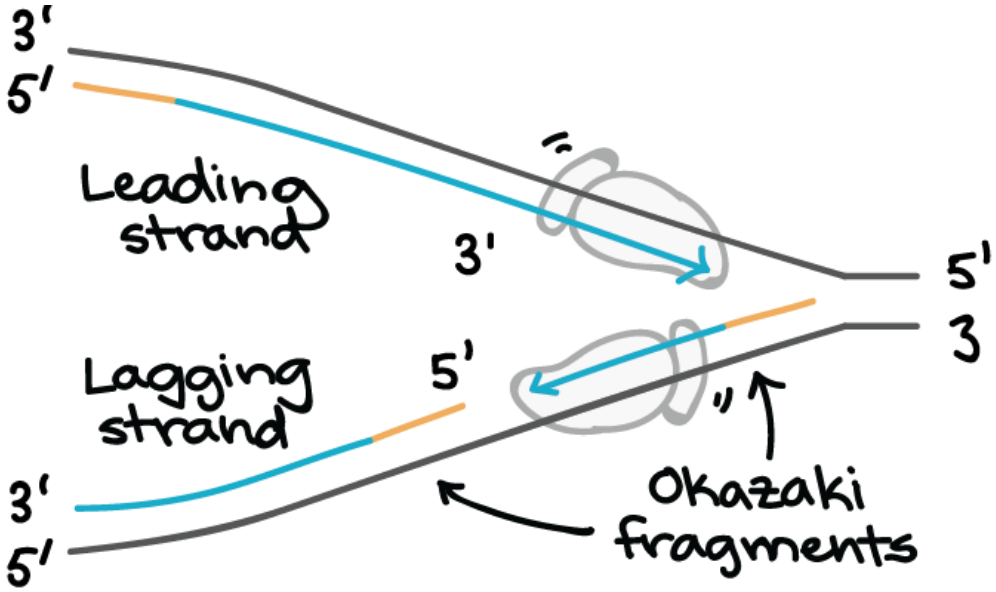
Once the RNA primer is in place, DNA polymerase "extends" it, adding nucleotides one by one to make a new DNA strand that's complementary to the template strand.

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 anti-parallel; 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^4start superscript, 4, end superscript.

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.

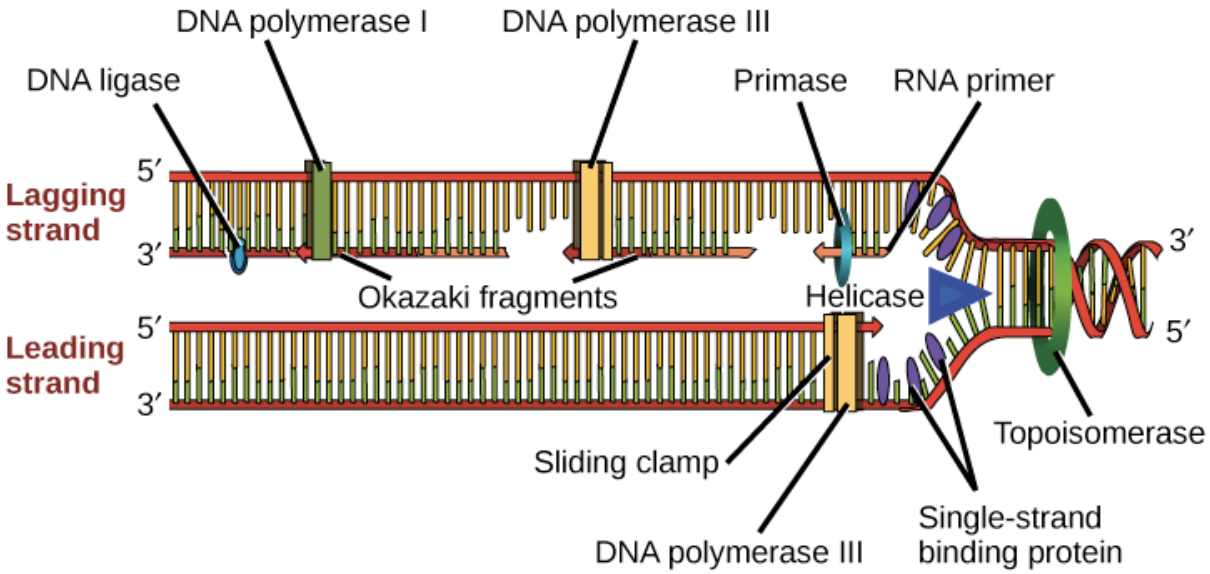


Illustration shows the replication fork. Helicase unwinds the helix, and single-strand binding proteins prevent the helix from re-forming. Topoisomerase prevents the DNA from getting too tightly coiled ahead of the replication fork. DNA primase forms an RNA primer, and DNA polymerase extends the DNA 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 100,100,100,000000000 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.