

# NUCLEIC ACIDS

## 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 mammals like you and me. 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.

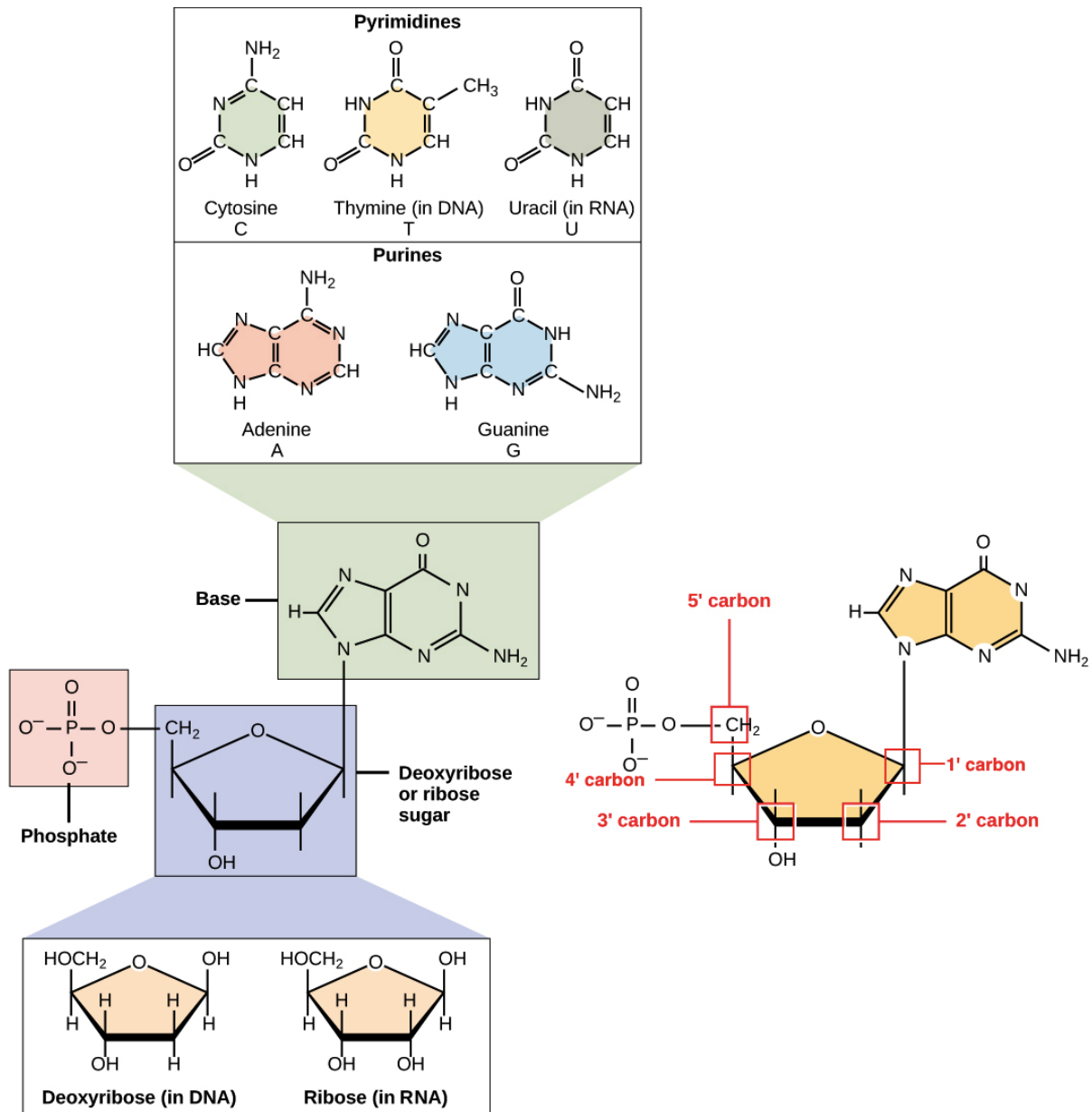


Image of the components of DNA and RNA, including the sugar (deoxyribose or ribose), phosphate group, and nitrogenous base. Bases include the pyrimidine bases (cytosine, thymine in DNA, and uracil in RNA, one ring) and the purine bases (adenine and guanine, two rings). The phosphate group is attached to the 5' carbon. The 2' carbon bears a hydroxyl group in ribose, but no hydroxyl (just hydrogen) in deoxyribose.

# Nitrogenous bases

The nitrogenous bases of nucleotides are organic (carbon-based) molecules made up of nitrogen-containing ring structures.

[\[Why is it called a base?\]](#)

^+start superscript, plus, end superscript

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<sup>1</sup>

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

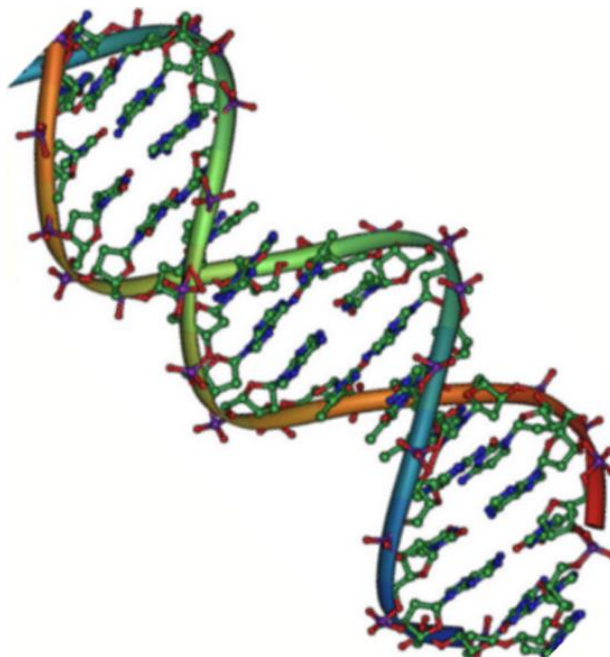
A consequence of the structure of nucleotides is that a polynucleotide chain has **directionality** – that is, it has two ends that are different from each other. At the **5’ end**, or beginning, of the chain, the 5’ phosphate group of the first nucleotide in the chain sticks out. At the other end, called 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.

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:

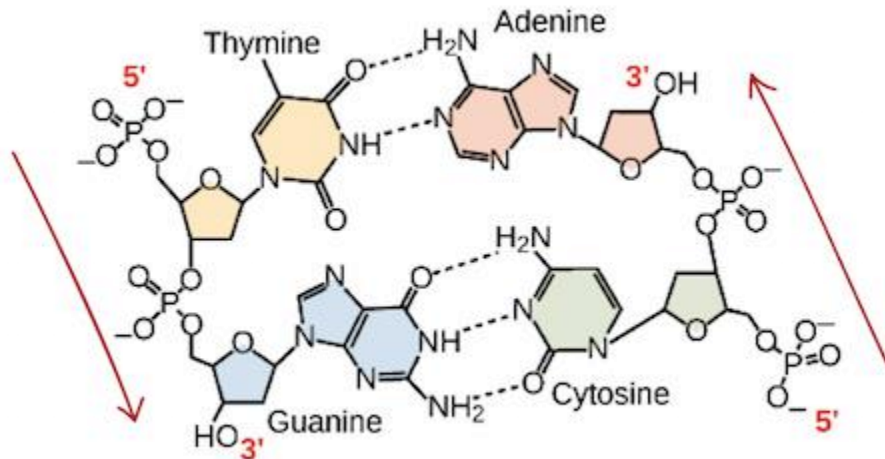


5'-AATTGGCC-3' 3'-TTAACCGG-5'

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.

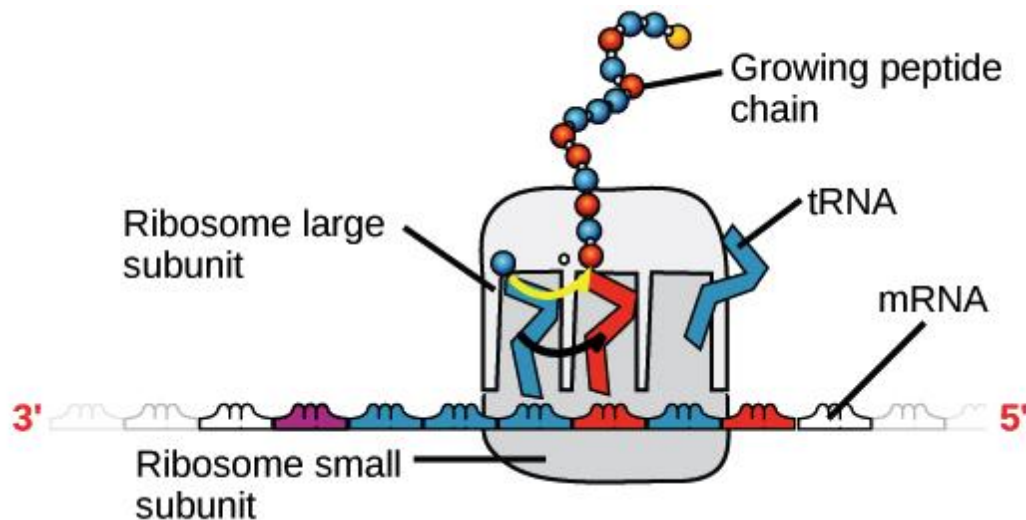


Image of a ribosome (made of proteins and rRNA) bound to an mRNA, with tRNAs bringing amino acids to be added to the growing chain. The tRNA that binds, and thus the amino acid that's added, at a given moment

is determined by the sequence of the mRNA that is being "read" at that time.

## 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.

## 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.

## 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