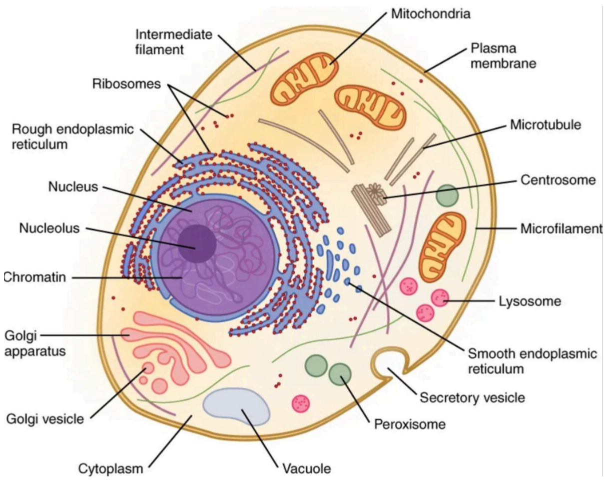
Coder's survival guide to Biology

written by Asia Pedroia

What's a cell?

The cell is the basic structural and functional unit of all forms of life or organisms. Imagine that our body is composed of around 30.000.000.000.000 cells !
There are many different types of cells. They can share common structural components but also have specialized features depending on their role, environment, and level of specialization within the organism.
For example muscle cells are specialized for contraction and neurons are specialized for electrical signaling.

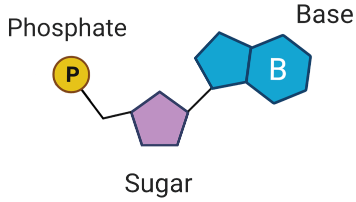


Despite their differences, most human cells share a common internal organization.

Here on the right you can see a typical eukaryotic cell and its structures. In particular, you should notice the nucleus. It is extremely important because it contains the DNA, which stores the genetic information of the organism.

DNA

What does it really mean that the DNA stores the genetic information of an organism?
This means that DNA contains the instructions that tell a cell how to build and maintain itself.



These instructions specify which proteins the cell needs to produce. Proteins are the molecules that form cellular structures and carry out most biological functions.

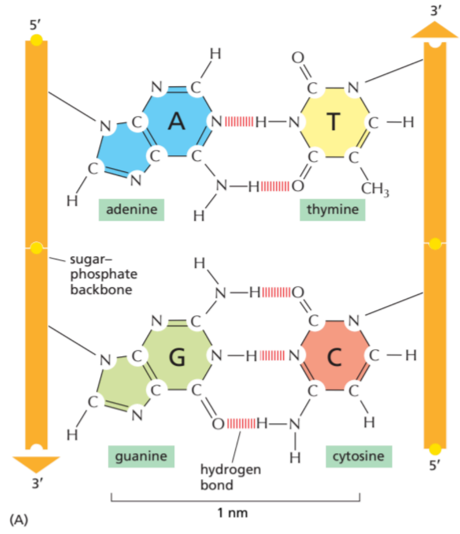
DNA stands for Deoxyribonucleic Acid. It has a shape that looks like a twisted ladder and scientists call it a double helix. It is made of small units called nucleotides and each of them has three parts: a sugar (called deoxyribose), a phosphate group, and a nitrogen base.

There are four types of nitrogen bases in DNA:

- Adenine (A)
- Thymine (T)
- Cytosine (C)
- Guanine (G)

These four bases always pair in a specific way:
A pairs with T
C pairs with G

The information in DNA is used to build proteins which are large molecules made of smaller units called amino acids.



Three consecutive nucleotides form a unit called a codon which corresponds to one specific amino acid.

When you look at the following genetic code table, you might notice that there is no T (Thymine), but instead there is a U. Why?

The central Dogma

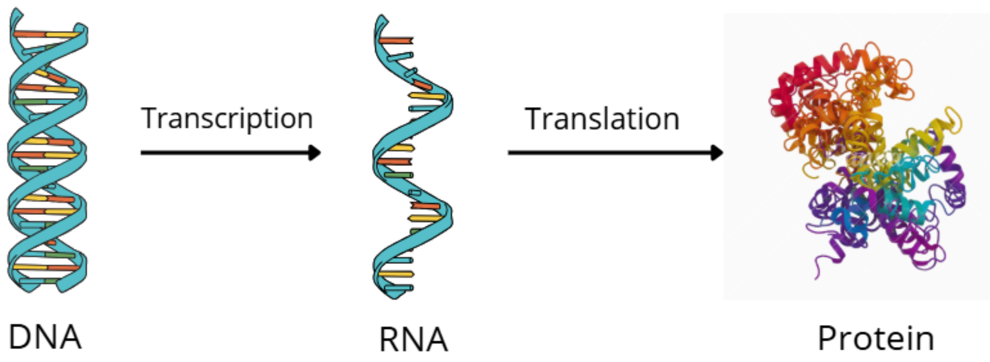
The letter U corresponds to Uracyl, another nitrogen base. It appears when the DNA is being manipulated to create proteins.

Let's see more in detail of this process:

Codon Chart

Second base in codon

First base in codon	U	C	A	G	Last base in codon
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC	UCC	UAC	UGC	C
	UUA Leu	UCA	UAA STOP	UGA STOP	A
	UUG	UCG	UAG	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC	CCC	CAC	CGC	C
	CUA	CCA	CAA Gln	CGA	A
	CUG	CCG	CAG	CGG	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC	ACC	AAC	AGC	C
	AUA	ACA	AAA Lys	AGA Arg	A
	AUG Met (start)	ACG	AAG	AGG	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC	GCC	GAC	GGC	C
	GUA	GCA	GAA Glu	GGA	A
	GUG	GCG	GAG	GGG	G



One of the two strands of the DNA is transcribed in one single filament called RNA. Each nucleotide is transcribed into its pair:

- C is transcribed into G
- G is transcribed into C

- T is transcribed into A

except for one of them:

- A is transcribed into U (instead of T)

After transcription, the RNA molecule is translated into a protein. During translation, the sequence of nucleotides in the mRNA is read in groups of three bases, the codons. Each of them specifies a particular amino acid which is then linked to another amino acid to form a protein. Both transcription and translation happen inside the cell and are catalyzed by the cell's machinery.

Proteins

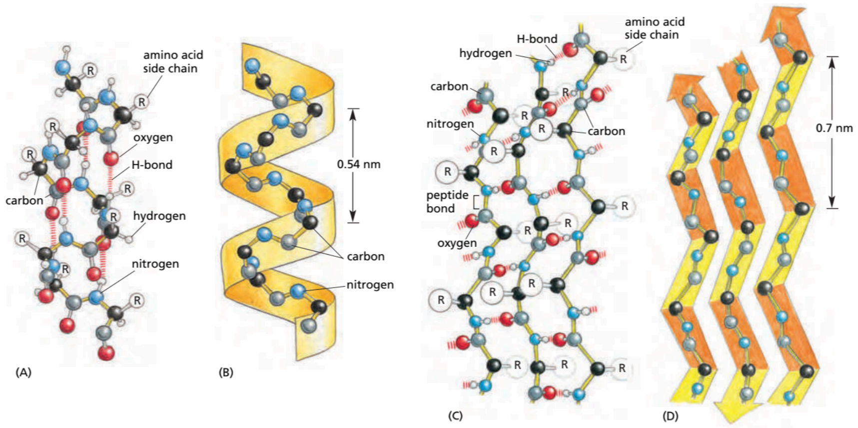
The primary structure is the simplest level. It is just the sequence of amino acids determined directly by the DNA sequence. If even one amino acid changes, the protein's function can sometimes change.

The secondary structure refers to local folding patterns of the amino acid chain. Because of interactions between nearby atoms, parts of the chain fold into regular shapes such as:

- Alpha helix (a spiral shape)
- Beta sheet (a folded, sheet-like structure)

These structures are stabilized mainly by hydrogen bonds.

The tertiary structure is the overall three-dimensional shape of a single protein chain. It determines the protein's function and, very importantly, if the shape of the protein changes because of a mutation, also its function could be affected!



At this level, the entire amino acid chain folds into a compact structure.
This folding is determined by physics, it happens because of electrostatic interactions between atoms that may be far apart in the sequence.

This principle connects directly to what's known as the thermodynamic hypothesis (or Anfinsen's principle). It states that a protein's native structure, its final, functional shape, is completely determined by the sequence of amino acids and the basic laws of physics, particularly the drive to reach the lowest possible free energy. In other words, proteins fold spontaneously into the most energetically stable conformation allowed by their chemical interactions without any external “instruction.” This elegant idea laid the foundation for computational protein design, where researchers use algorithms, physical models and later AI to predict or invent amino acid sequences that will fold into specific, stable 3D structures with desired functions. By simulating these physical principles in silico, coders can now help design real, functional molecules.

Good luck and Enjoy the Ride!