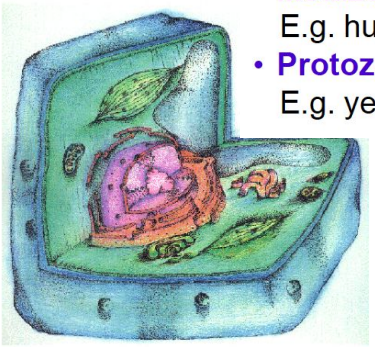
Introduction to molecular biology

Based on slides from
Sandrine Dudoit and Robert Gentleman
(Bioconductor course)

Cells

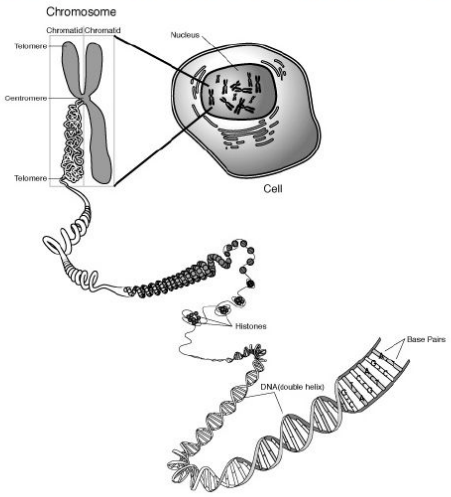
- **Cells**: the fundamental working units of every living organism.
- **Metazoa**: multicellular organisms. E.g. humans: trillions of cells.
- **Protozoa**: unicellular organisms. E.g. yeast, bacteria.



Cells

- Each cell contains a complete copy of an organism's **genome**, or blueprint for all cellular structures and activities.
- Cells are of many different types (e.g. blood, skin, nerve cells), but all can be traced back to a single cell, the fertilized egg.

Chromosomes and DNA

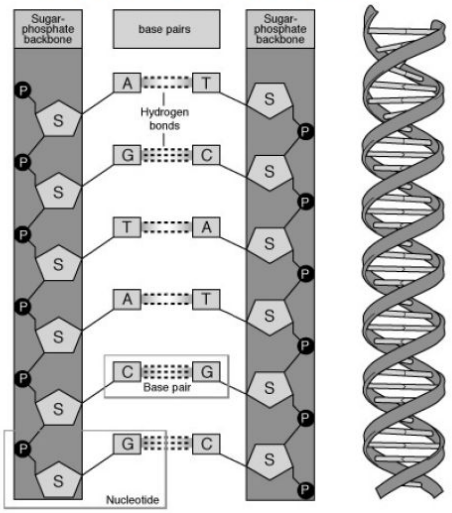


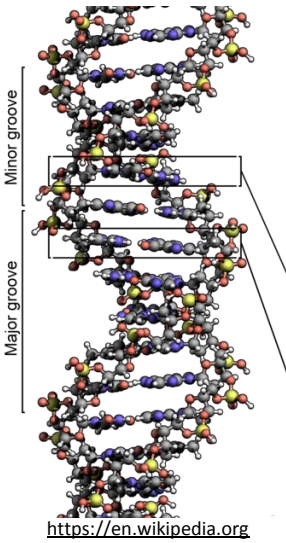
DNA structure

"We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest."

J.D. Watson & F. H. C. Crick. (1953). Molecular structure of Nucleic Acids. *Nature*. **171**: 737-738.

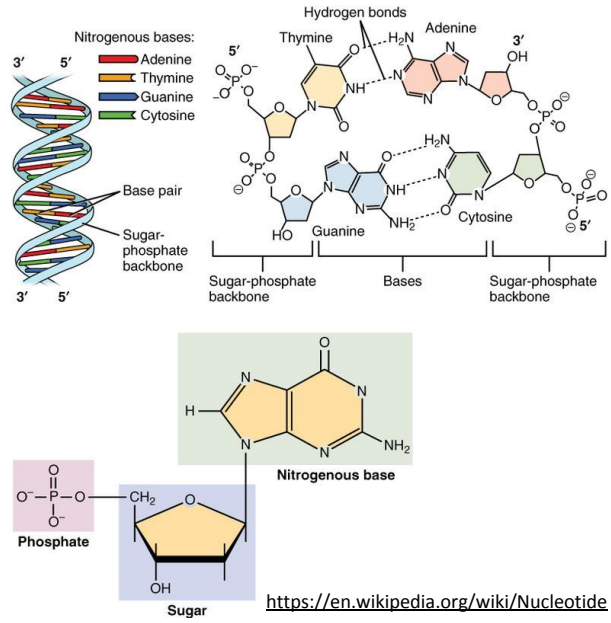
DNA structure





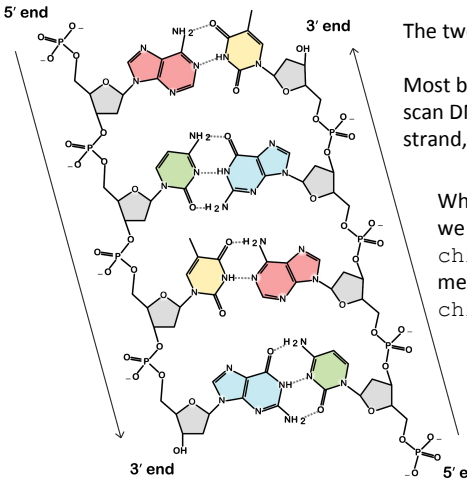
https://en.wikipedia.org

Nucleotides



https://en.wikipedia.org/wiki/Nucleotide

DNA = Two antiparallel strands



The two strands run in opposite directions

Most biological processes that scan DNA proceed along one given strand, from 5' to 3'

When representing DNA as string, we only show one strand, from 5' to 3'.

```
chr1 = ACGCAGCATCAGCATCGCA
```

means

```
chr1 = ACGCAGCATCAGCATCGCA
       ||||||||||||||||||||
       TGCGTCGTAGTCGTAGCGT
```

which strand we choose is arbitrary

DNA structure

- A **deoxyribonucleic acid** or **DNA** molecule is a double-stranded polymer composed of four basic molecular units called nucleotides.
- Each **nucleotide** comprises
  - a phosphate group;
  - a deoxyribose sugar;
  - one of four nitrogen bases:
    - purines: **adenine (A)** and **guanine (G)**,
    - pyrimidines: **cytosine (C)** and **thymine (T)**.

DNA structure

- Polynucleotide chains are **directional** molecules, with slightly different structures marking the two ends of the chains, the so-called **3' end** and **5' end**.
- The 3' and 5' notation refers to the numbering of carbon atoms in the sugar ring.
- The 3' end carries a sugar group and the 5' end carries a phosphate group.
- The two complementary strands of DNA are **antiparallel** (i.e, 5' end to 3' end directions for each strand are opposite)

Genomes

The *genome* of a cell is the entirety of its DNA content.

A genome is made of one or more *chromosomes*: contiguous piece of double-stranded DNA

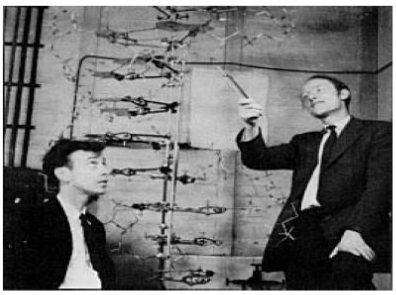
In bacteria (prokaryotes):

- One circular circular chromosome (2-10 Mb)
- Some small chromosomes called plasmids

In human (eukaryote):

- 23 pairs of chromosomes = 22 autosomes pairs + 1 pair of sex chromosome (XX or XY)
- Each chromosome is 50 – 250 Mb
- Total genome size: 3,000,000,000 bp
- Total length of DNA in one nucleus: 2 meters!

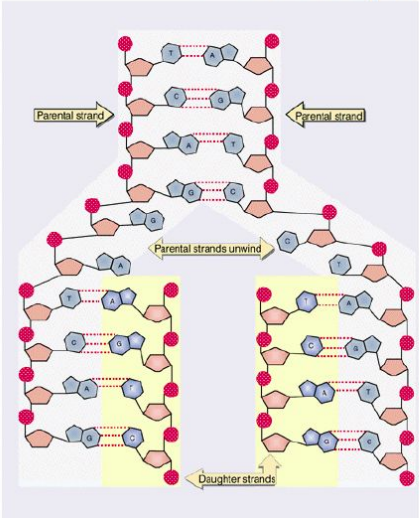
DNA replication



"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

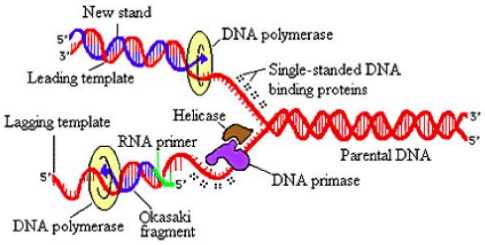
J.D. Watson & F. H. C. Crick. (1953). Molecular structure of Nucleic Acids. *Nature*. **171**: 737-738.

DNA replication



Base pairing provides the mechanism for DNA replication.

DNA replication



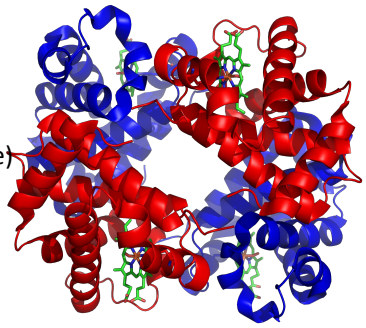
Collaboration of Proteins at the Replication Fork

Useful video: https://www.youtube.com/watch?v=TNKWgcFPHqw

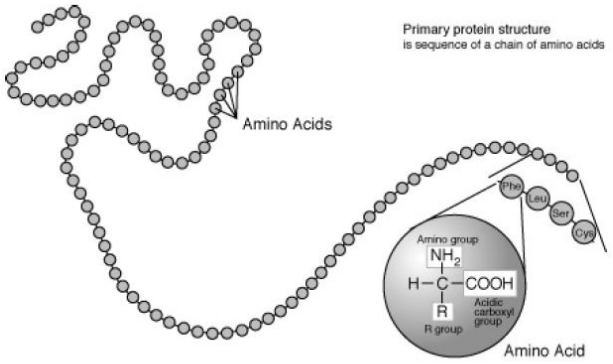
Proteins

Proteins are molecules that perform a huge diversity of functions in the cell:

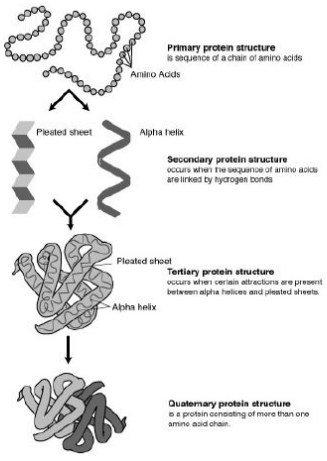
- Structure (actin, tubulin)
- DNA replication (DNA polymerase) + repairs
- DNA transcription (DNA transcriptase)
- Transport of small molecules (hemoglobin)
- Signaling (kinases)
- Regulation (transcription factors)
- Catalyze reactions (enzymes)
- Etc. etc.

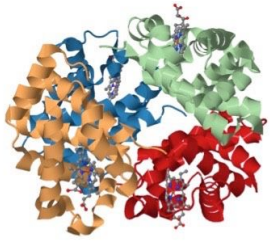


Proteins



Proteins

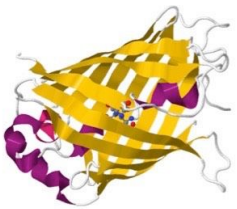




Haemoglobin *(quaternary)*
4 subunits with heme groups

Aquaporin *(quaternary)*
4 subunits forming transmembrane pores

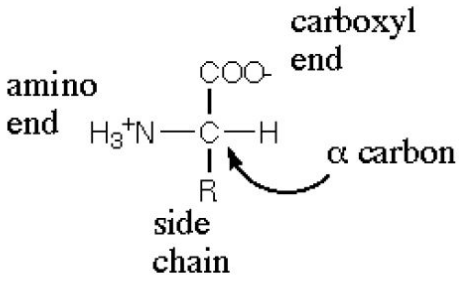
Keratin *(secondary)*
Insoluble helical strands



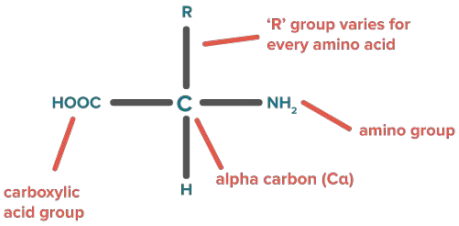
Green Fluorescent Protein *(tertiary)*
Beta barrel protecting inner chromophore

Source: https://ib.bioninja.com.au/higher-level/topic-7-nucleic-acids/73-translation/visualising-proteins.html

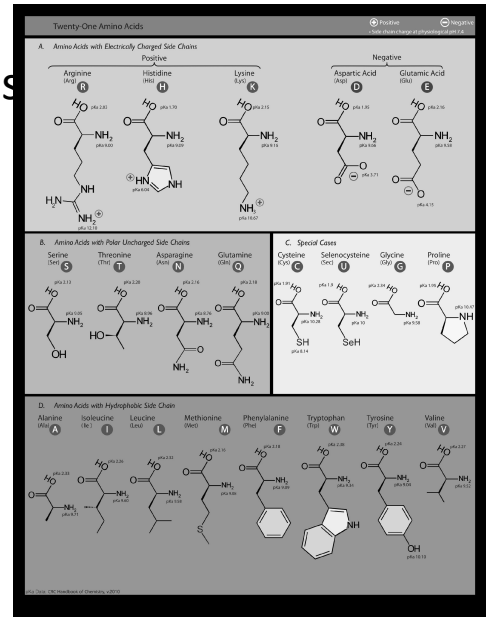
Amino acids

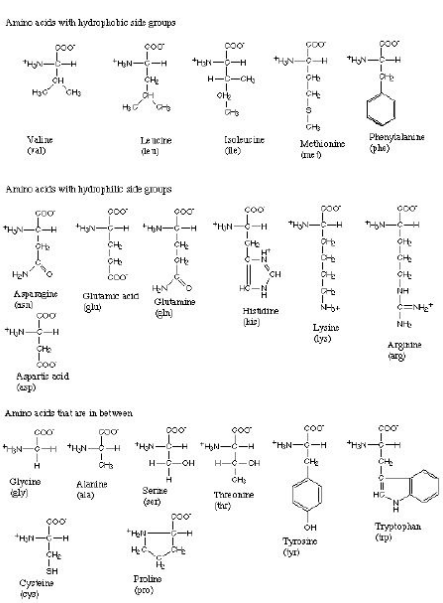


Amino acids



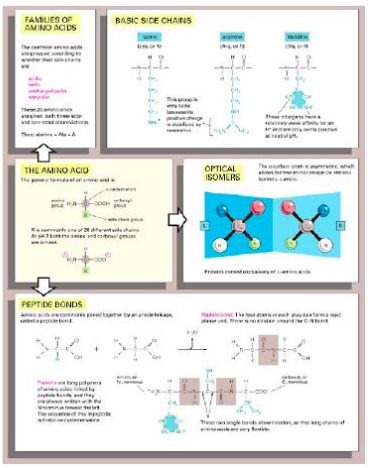
20 amino acids

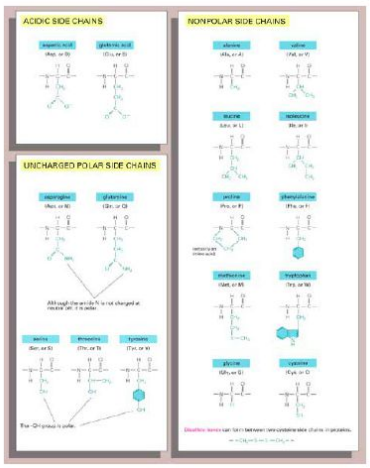




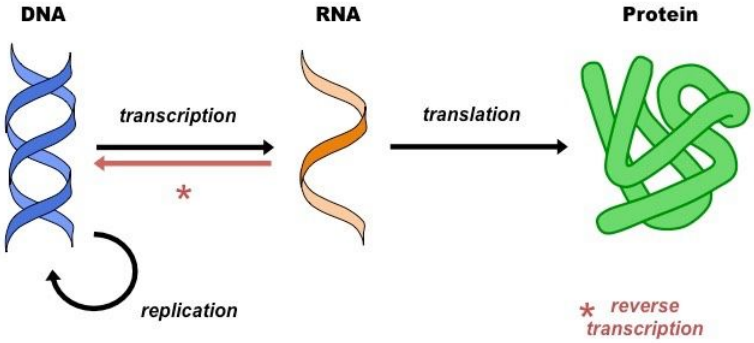
Amino acids

Amino acids





From DNA to proteins
The Central Dogma of Molecular Biology



Central dogma

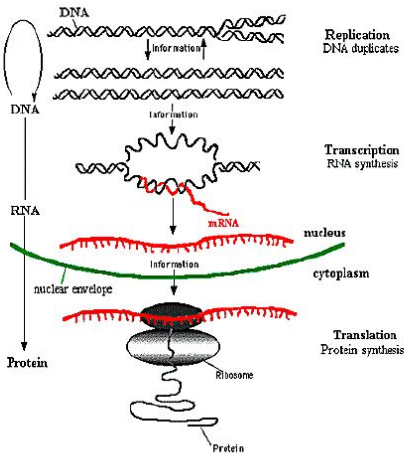
The **expression** of the genetic information stored in the DNA molecule occurs in two stages:

- (i) **transcription**, during which DNA is transcribed into mRNA;
- (ii) **translation**, during which mRNA is translated to produce a protein.

DNA ➔ mRNA ➔ protein

Other important aspects of regulation: methylation, alternative splicing, etc.

Central dogma



The Central Dogma of Molecular Biology

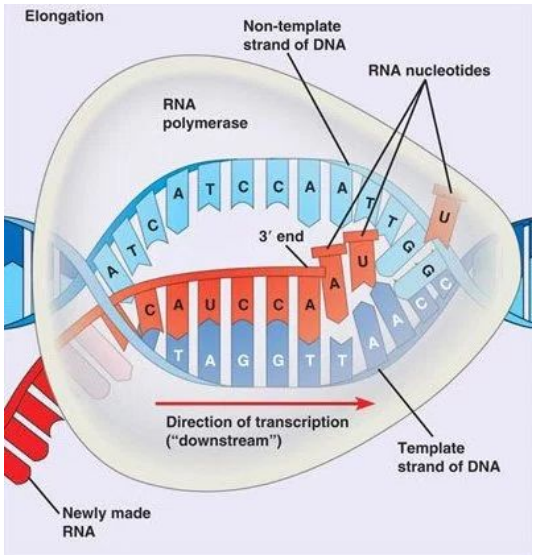
RNA

- A **ribonucleic acid** or **RNA** molecule is a nucleic acid similar to DNA, but
  - single-stranded;
  - ribose sugar rather than deoxyribose sugar;
  - **uracil (U)** replaces thymine (T) as one of the bases.
- RNA plays an important role in protein synthesis and other chemical activities of the cell.
- Several classes of RNA molecules, including **messenger RNA** (**mRNA**), transfer RNA (tRNA), ribosomal RNA (rRNA), and other small RNAs.

Transcription

- Analogous to DNA replication: several steps and many enzymes.
- **RNA polymerase** synthesizes an RNA strand complementary to one of the two DNA strands.
- The RNA polymerase recruits **rNTPs** (ribonucleotide triphosphate) in the same way that DNA polymerase recruits dNTPs (deoxunucleotide triphospate).
- However, synthesis is **single stranded** and only proceeds in the 5' to 3' direction of mRNA (no Okazaki fragments).

Transcription



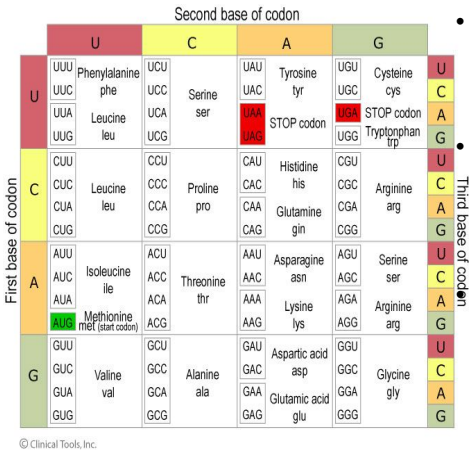
Translation

- Goal: Produce an protein sequence from a messenger RNA template
- But how to encode a protein sequence made of amino acids (20 different varieties) from a DNA sequence made of nucleotides (4 different varieties)?

Useful videos:

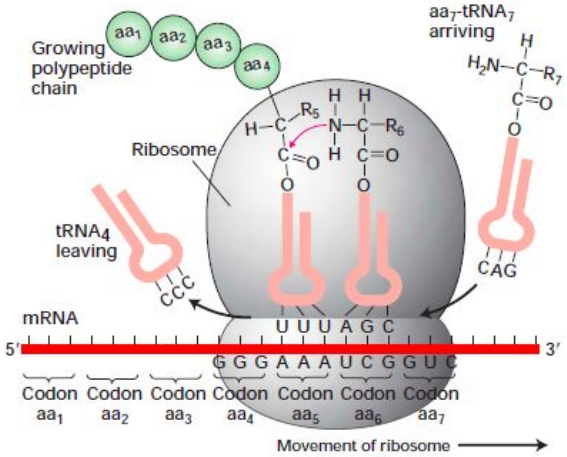
https://www.youtube.com/watch?v=2BwWavExcFI [1:34 to 1:50]
https://www.youtube.com/watch?v=5MfSYnItYvg

The genetic code



- 61 of the 4^3=64 triplets of nucleotide (called a codon) are translated to amino acids
- Much redundancy: most aa are encoded by multiple codons
  - Most redundancy is a 3rd codon position
- Special cases:
  - ATG encode Met but also means START
  - TAA|TAG|TGA = STOP

Translation

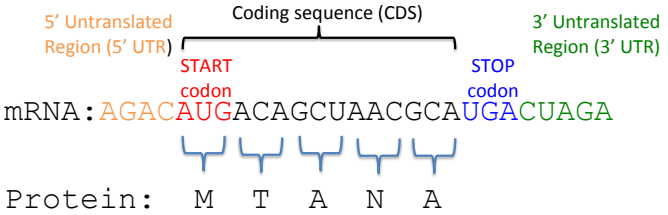


Translation

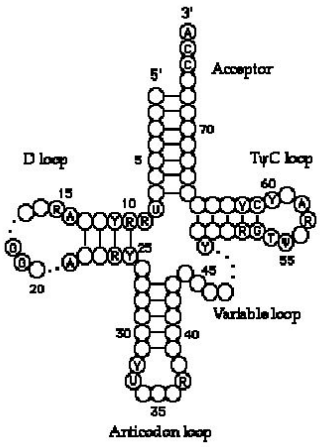
- **Ribosome**:
  - cellular factory responsible for protein synthesis;
  - a large subunit and a small subunit;
  - structural RNA and about 80 different proteins.
- **transfer RNA (tRNA)**:
  - adaptor molecule, between mRNA and protein;
  - specific **anticodon** and **acceptor site**;
  - specific **charger protein**, can only bind to that particular tRNA and attach the correct amino acid to the acceptor site.

Translation

1) Ribosome searches for the first START codon (but there are many exceptions)
2) From there, non-overlapping triplets (codons) are translated to an amino acid
3) Until the ribosome encounters an in-frame STOP codon



tRNA



- The tRNA has an **anticodon** on its mRNA-binding end that is complementary to the codon on the mRNA.
- Each tRNA only binds the appropriate amino acid for its anticodon.