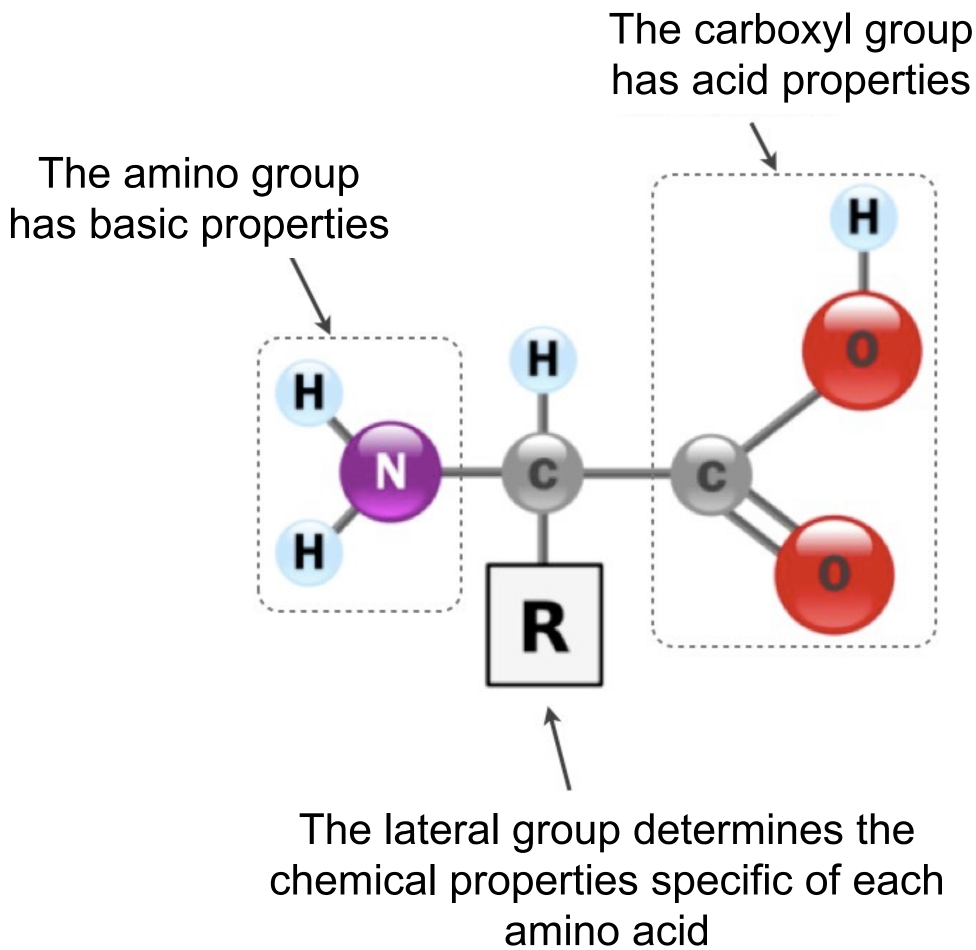
Astrobiology
Lecture 4

Metabolic & genetic macromolecules

SISSA, Academic Year 2023
Giovanni Vladilo (INAF-OATs)

Biological amino acids

Constituents of proteins



Biological macromolecules

Proteins

Proteins are polymers of amino acids

Short chains of amino acids are called peptydes

Long, unbranched peptyde chains are called polypeptides

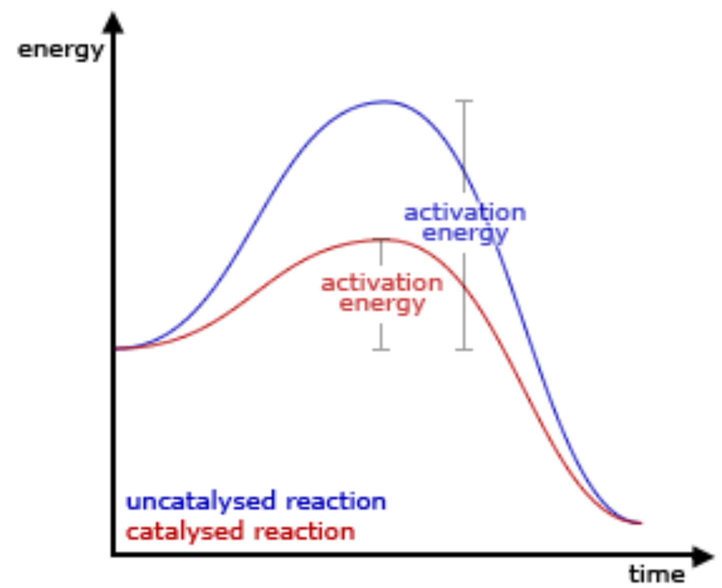
Proteins are formed by one or more chains of polypeptides

Molecular masses of proteins vary between ~10^3 e ~10^6 atomic mass units

They contribute to about half the mass of the cell

Proteins play fundamental functions in living organisms

Mostly structural and enzymatic (i.e., catalytic) functions



Terrestrial biological amino acids

Proteins use only 20 types of amino acids

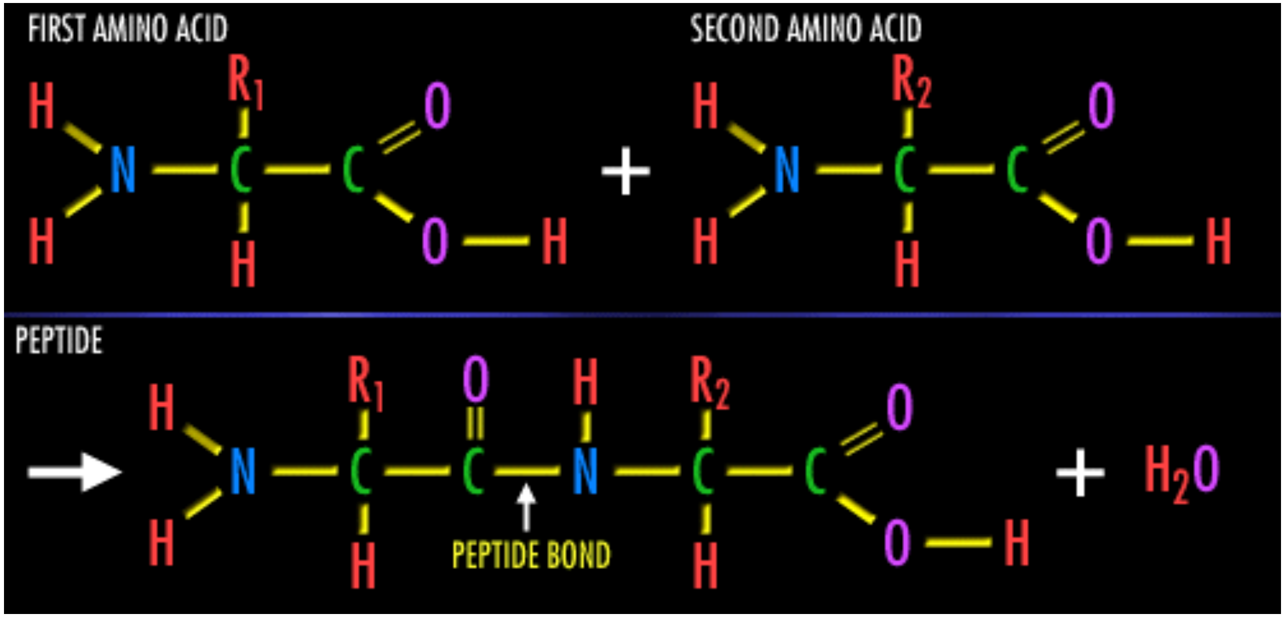
Organic chemistry allows for the existence of thousands of amino acids

Apparently, terrestrial life has "chosen" a short list of amino acids, sufficiently representative of the different types of chemical properties that are required to build up the variety of proteins necessary to living organisms

Table 7.2 *The Twenty Amino Acids Found in Living Organisms*

*Amino Acid**	*Chemical Formula*	*Number of Atoms*
L-Alanine	$C_3H_7O_2N$	13
L-Arginine	$C_6H_{15}O_2N_4$	27
L-Asparagine	$C_4H_8O_3N_2$	17
L-Aspartic Acid	$C_4H_6O_4N$	15
L-Cysteine	$C_3H_7O_2NS$	14
L-Glutamic Acid	$C_5H_8O_4N$	18
L-Glutamine	$C_5H_{10}O_3N_2$	20
Glycine	$C_2H_5O_2N$	10
L-Histidine	$C_6H_9O_2N_3$	20
L-Isoleucine	$C_6H_{13}O_2N$	22
L-Leucine	$C_6H_{13}O_2N$	22
L-Lysine	$C_6H_{15}O_2N_2$	25
L-Methionine	$C_5H_{11}O_2NS$	20
L-Phenylalanine	$C_9H_{11}O_2N$	23
L-Proline	$C_5H_9O_2N$	17
L-Serine	$C_3H_7O_3N$	14
L-Threonine	$C_4H_9O_3N$	17
L-Tryptophan	$C_{11}H_{12}O_2N_2$	27
L-Tyrosine	$C_9H_{11}O_3N$	24
L-Valine	$C_5H_{11}O_2N$	19

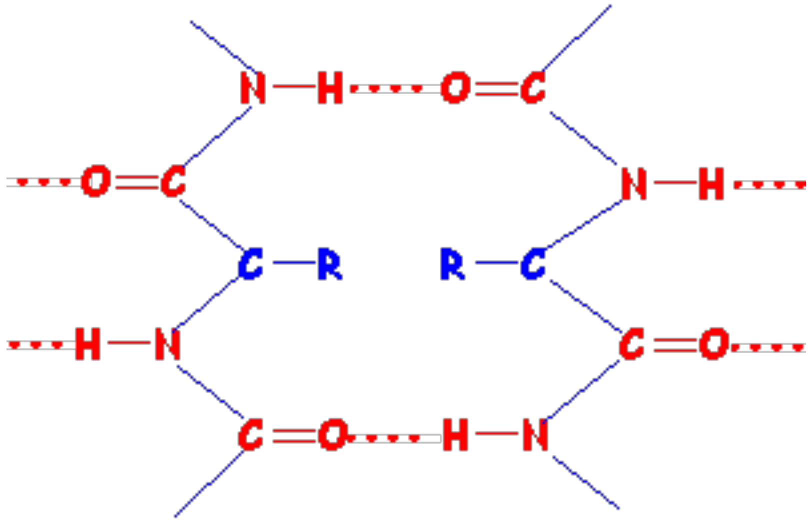
From amino acids to polypeptides

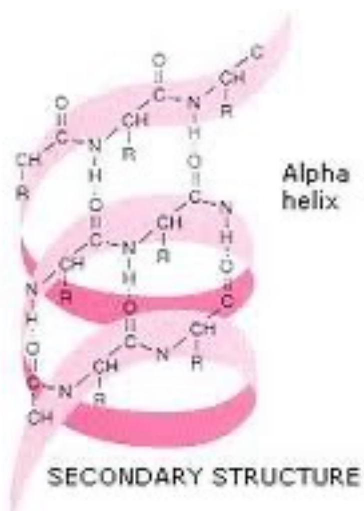


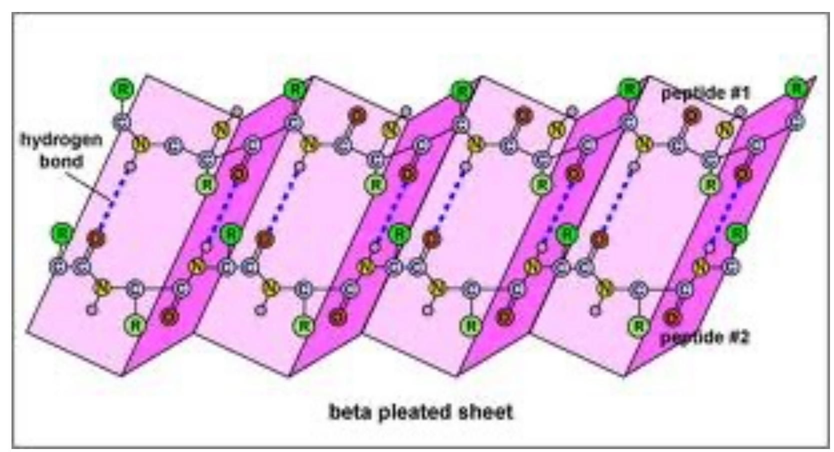
Amino acids are bound to each other with peptide bonds

- The carboxyl end ties to the amino end of the next molecule
  - A sequence OC-NH is formed (peptide bond)
- A water molecule is released each time a peptide bond is created
- A water molecule is required to break a peptide bond

From polypeptides to 3D proteins
Importance of hydrogen bonds as intramolecular forces



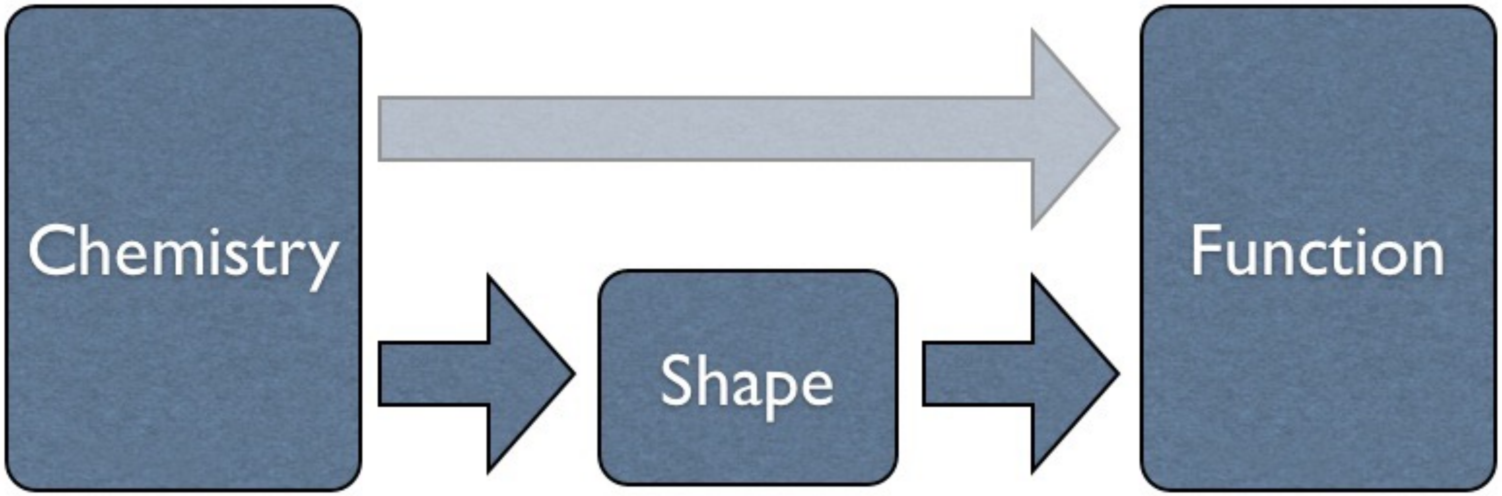


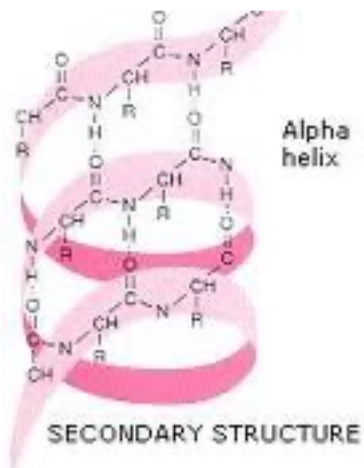


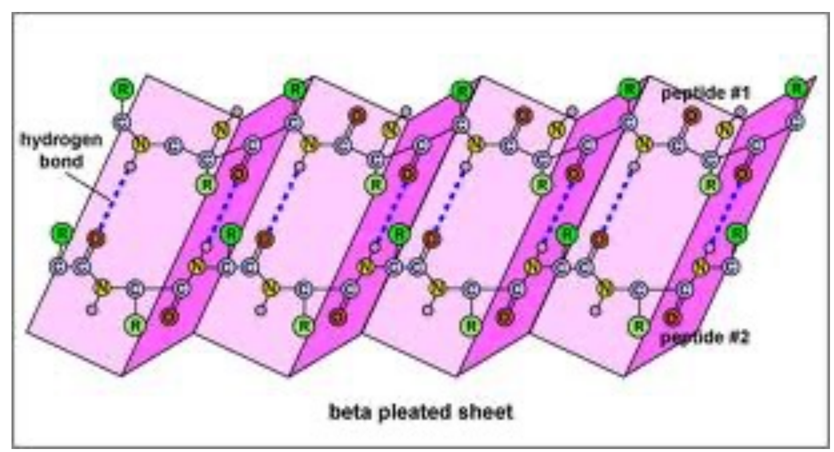
Proteins

The importance of 3D structures

The shape of proteins determines, in large part, their function







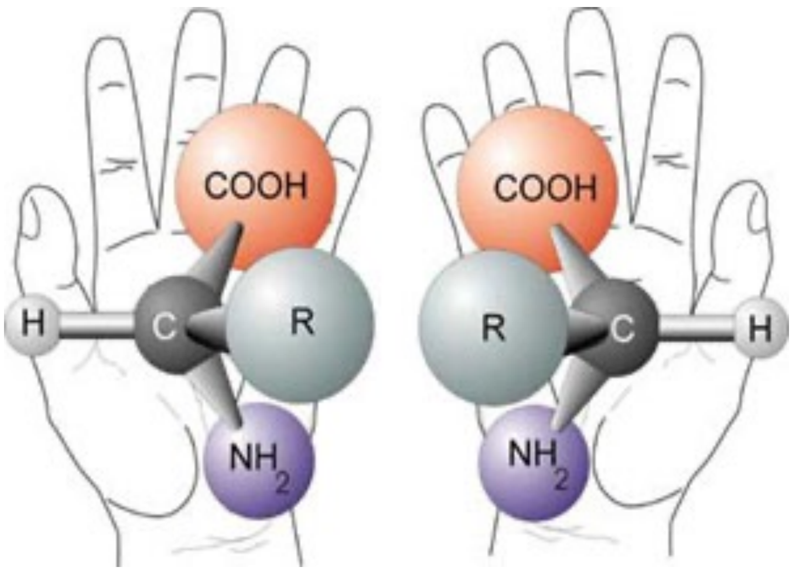
Chirality

Molecules with same chemical formula but different structure are called <u>isomers</u>

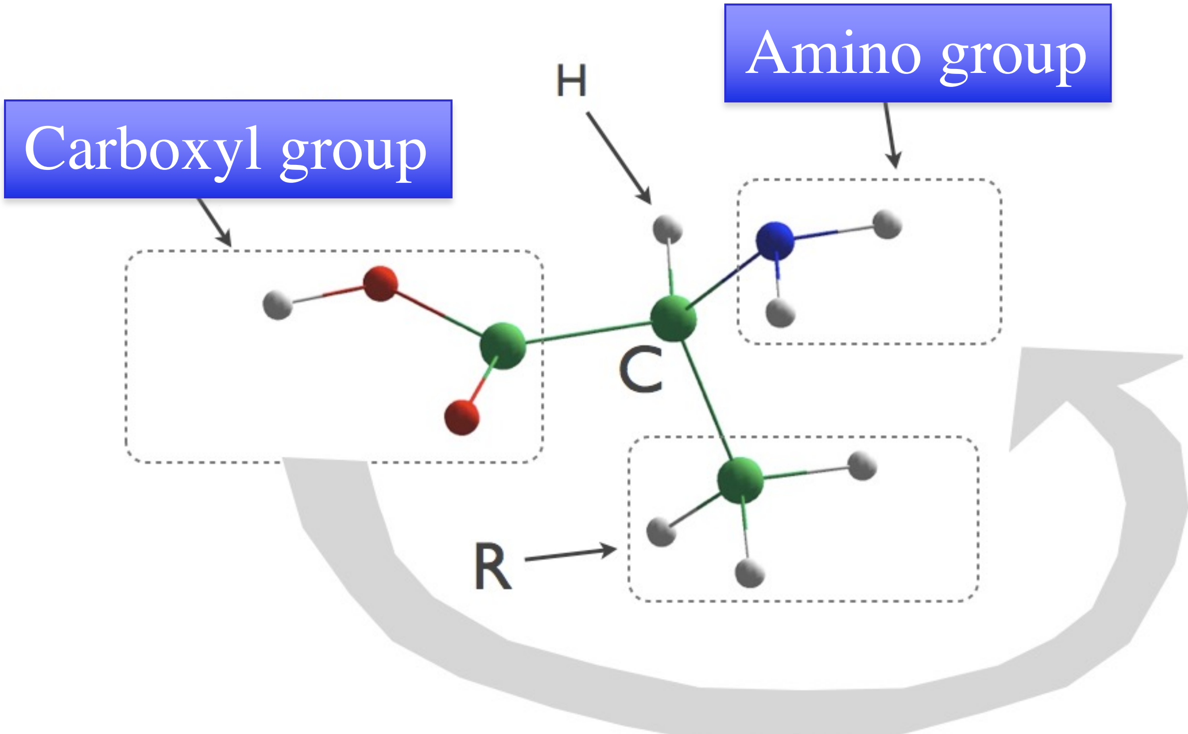
<u>Chiral</u> molecules are isomers with a center of symmetry (“stereocenter”)
They cannot be superimposed to their mirror image
The two mirror images of a chiral molecule are called <u>enantiomers</u>

Amino acids are chiral

The carbon atom at the center of the amino acid is the <u>stereocenter</u>



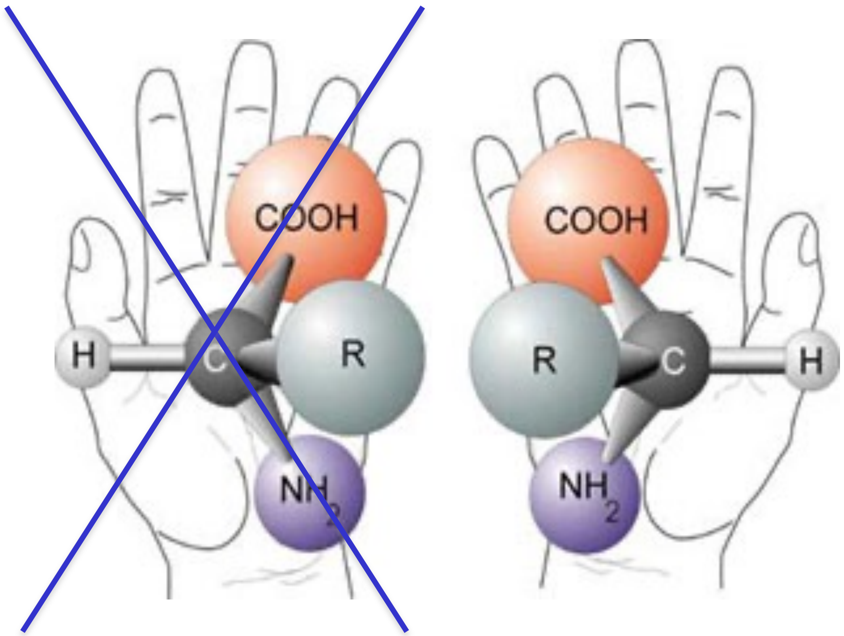
The two enantiomers of amino acids are called L and D according to the "CORN" convention



Example of L-type amino acid: L-alanine

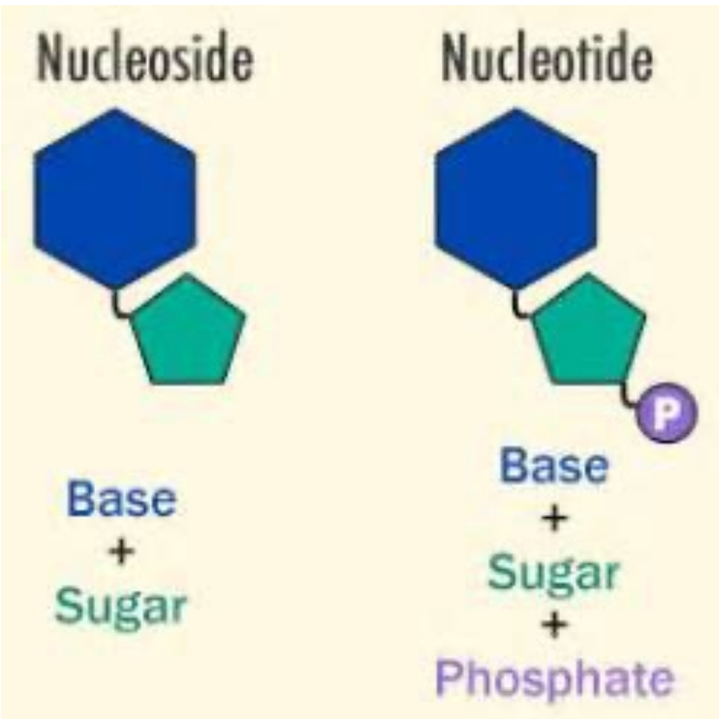
Homochirality of biomolecules

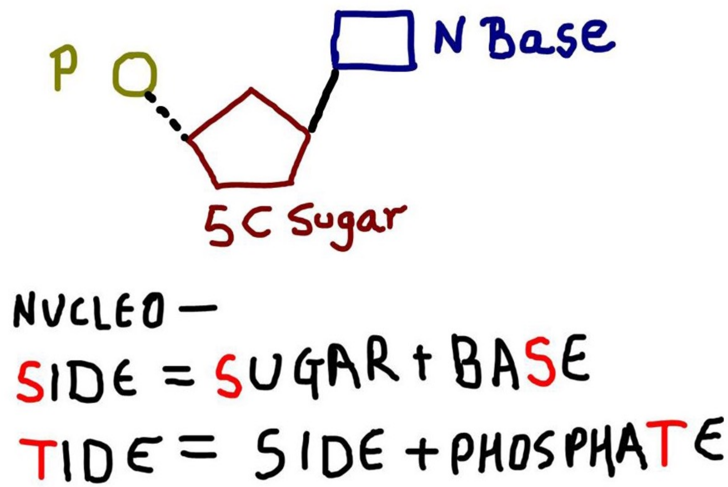
Protein amino acids are homochiral:
they only show the L-type enantiomer



Nucleosides and nucleotides

Constituents of the DNA and RNA

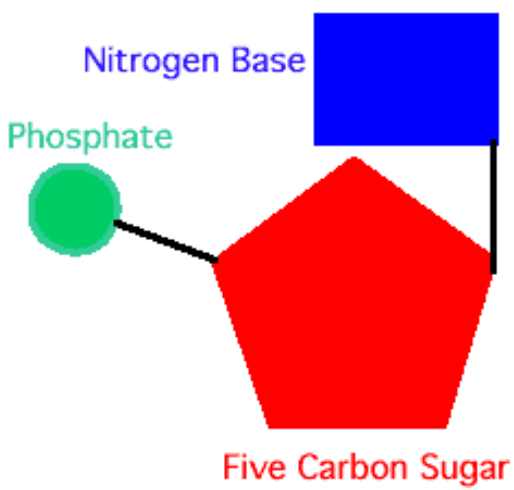




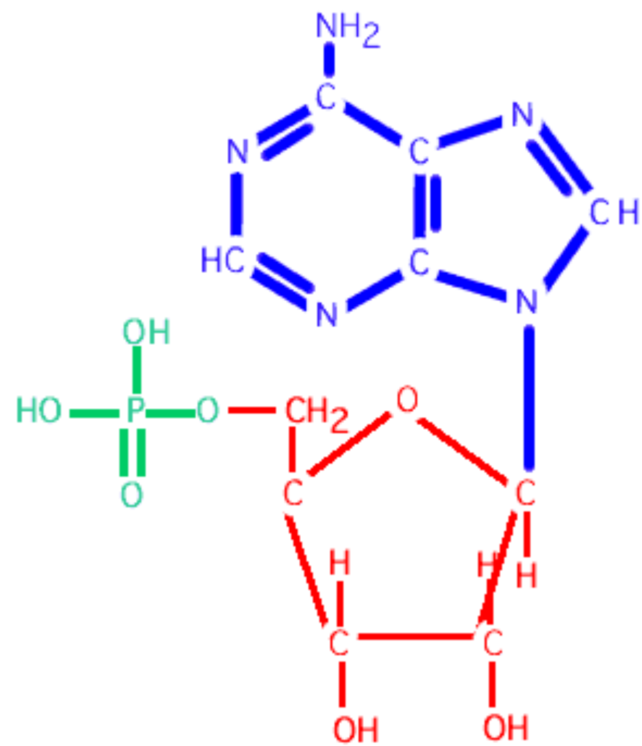
Nucleotides

Constituents of the DNA and RNA

Basic Nucleotide Structure



Example

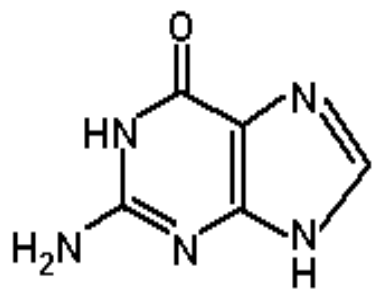


Adenosine 5' phosphoric acid

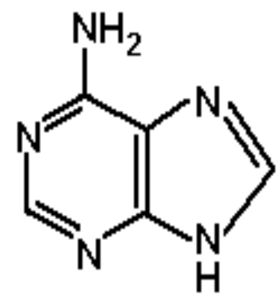
Nitrogen bases

aromatic rings with N substitutions

Purines

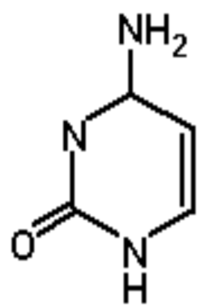


Guanin

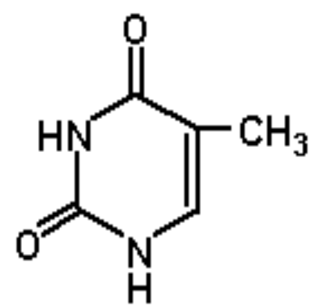


Adenin

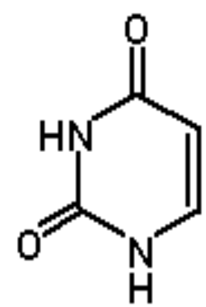
Pyrimidines



Cytosin



Thymin



Uracil

Nucleic acids

Nucleic acids are polymers of nucleotides

Depending on the type of organism,
they may contain ~ $10^6 - 10^8$ nucleotides

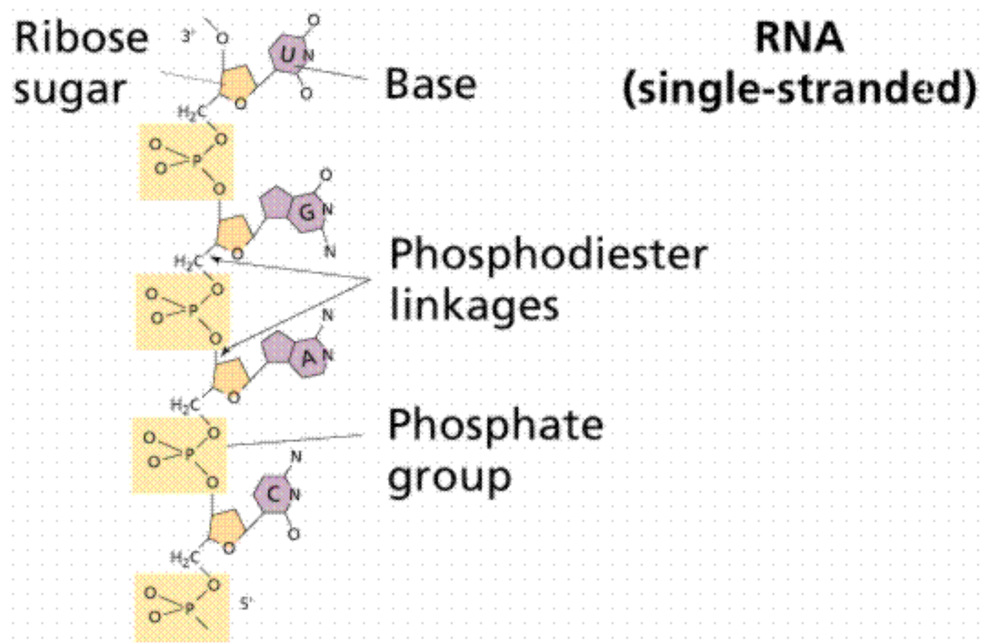
Nucleic acids store and use the genetic information

There are two types: DNA and RNA
DNA preserves the information
RNA, which comes in different forms, uses the information
for driving metabolic/replication processes

Nucleic acids: RNA

RNA has a single strand of nucleotides

The backbone of the strand is made up of a sequence of phosphate groups and <u>ribose</u> sugars



Has 4 types of nucleobases

Purines

- Adenine, Guanine

Pyrimidines

- Cytosine, Uracyl

RNA drives the synthesis of proteins

The order of the nitrogen bases on the backbone determines the sequence in which amino acids are assembled

Nucleic acids: DNA

DNA has two strands that form a double helix structure

The backbone of each strand is made up of a sequence of phosphate groups and deoxyribose sugars

DNA has 4 types of nucleobases

- 2 purins
  - Adenine e Guanine
- 2 pyrimidins
  - Cytosine e Thymine
    - Thymine replaces Uracyl, which is instead used in the RNA

The complementarity of purines and pyrimidines is essential for pairing the two strands

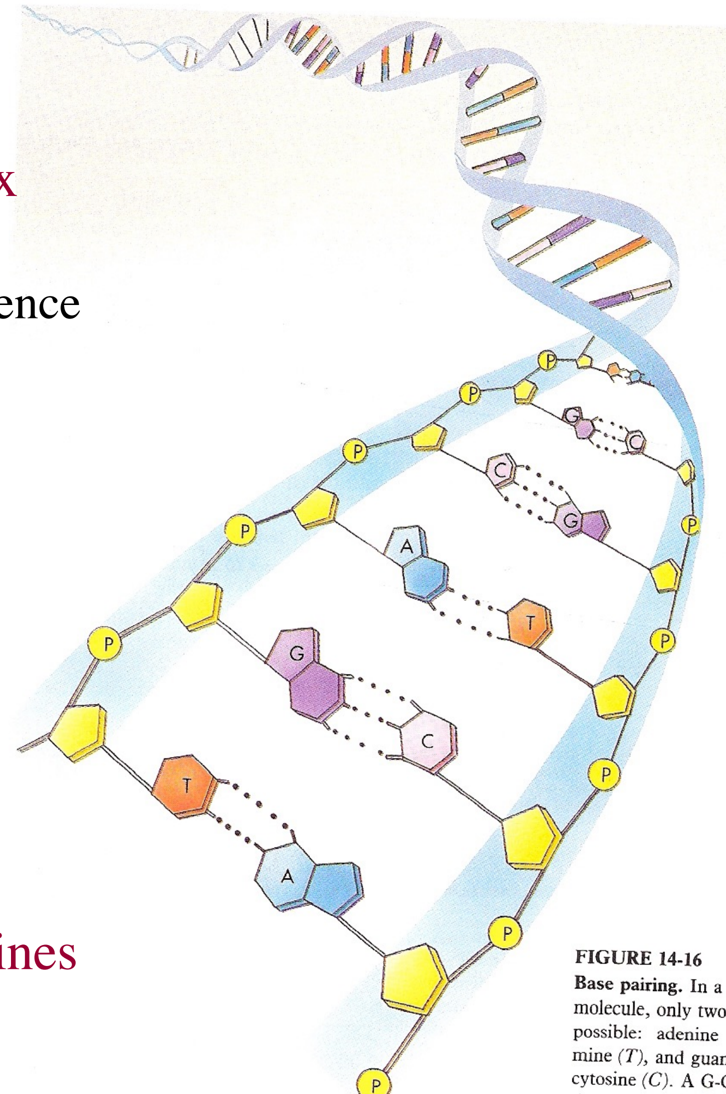


FIGURE 14-16
Base pairing. In a DNA molecule, only two base 1 possible: adenine (*A*) v mine (*T*), and guanine (*G* cytosine (*C*). A G-C base three hydrogen bonds; an base pair, only two.

Hydrogen bonds in the DNA

The two strands of the DNA are tied up by hydrogen bonds

G-C pairs have 3 bonds

T-A pairs have 2 bonds

Hydrogen bonds as intramolecular forces

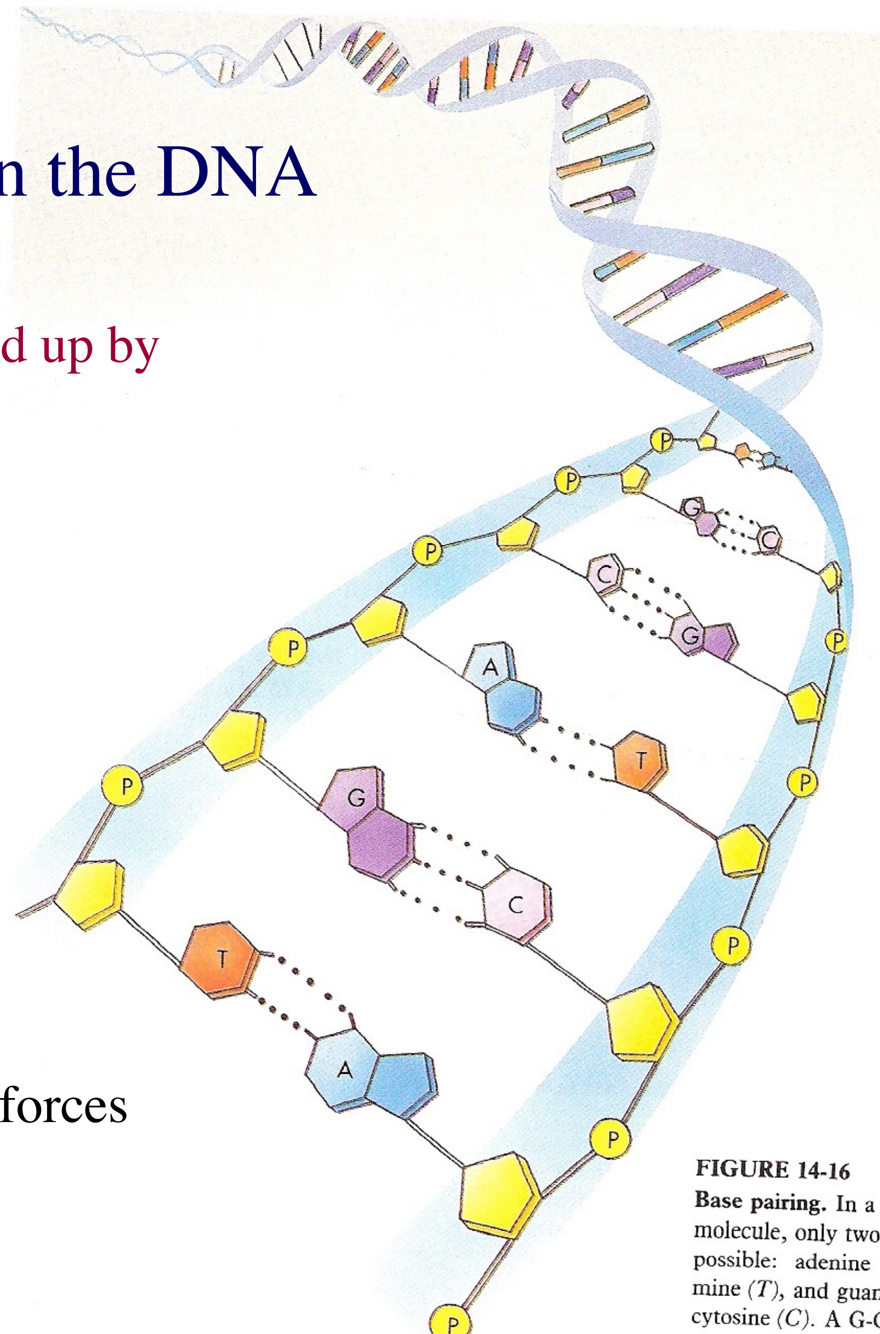


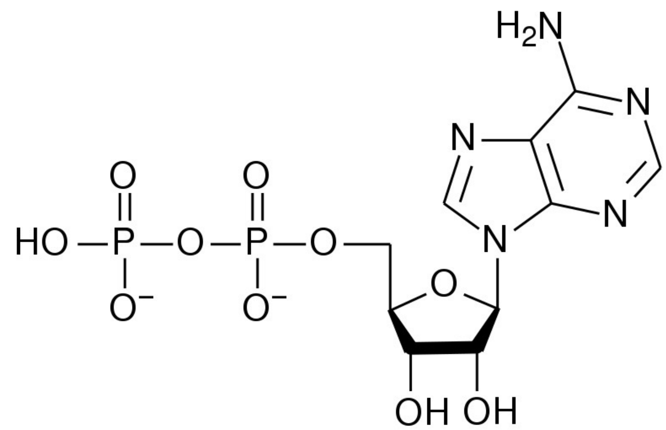
FIGURE 14-16
Base pairing. In a DNA molecule, only two base [cut off] possible: adenine (*A*) [cut off] mine (*T*), and guanine (*G*[cut off] cytosine (*C*). A G-C base [cut off] three hydrogen bonds; an[cut off] base pair, only two.

ADP & ATP

Energy exchange molecules

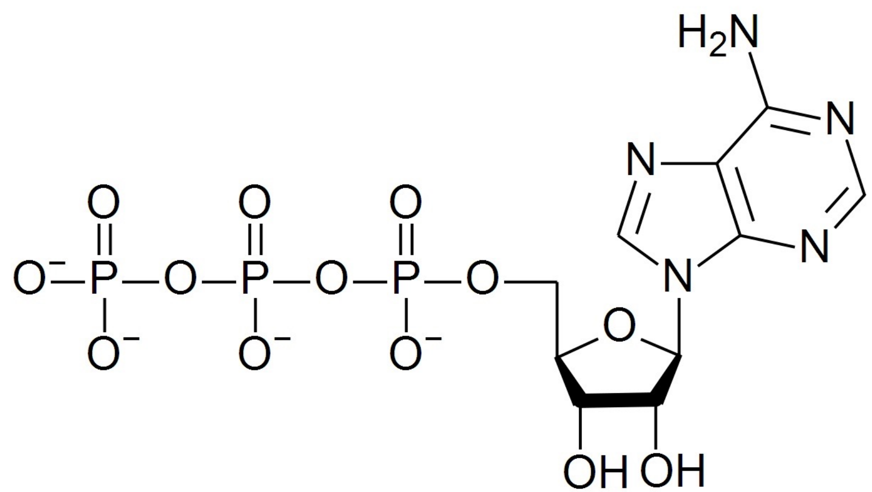
ADP

Adenosine diphosphate



ATP

Adenosine triphosphate



Genetic information

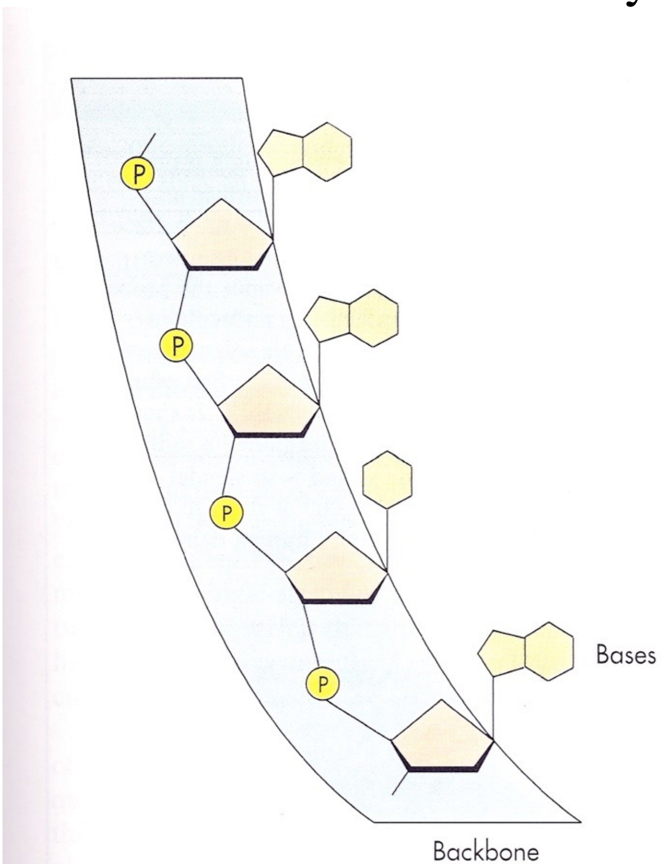
The <u>order</u> of the nucleobases attached to the backbone of the nucleic acids determines the genetic information, which is therefore stored in <u>digital form</u>

The <u>order of the nucleobases</u> is not constrained by chemical laws

We believe this is the result of natural selection at early stages of life evolution

Digital information is more stable than analogic information

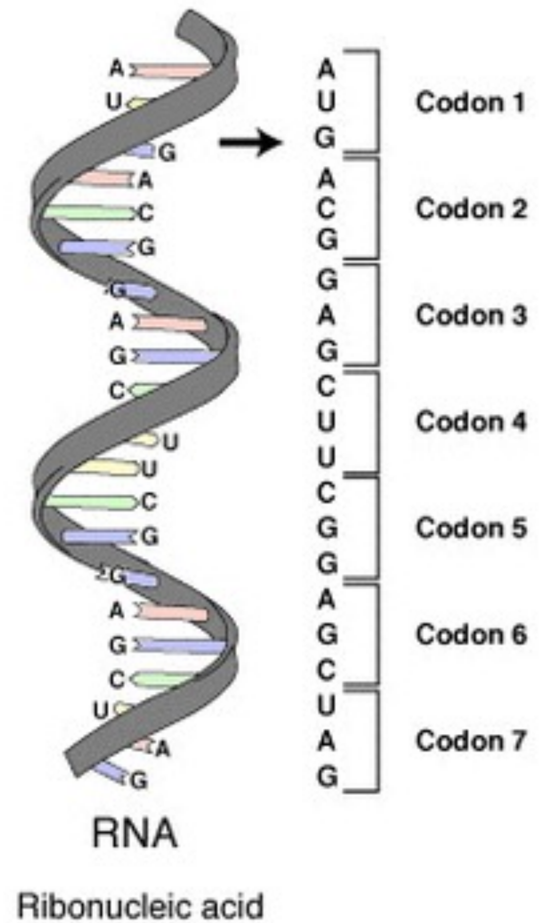
Terrestrial life has started to use digital information more than 3.5 billon years ago



The genetic information and genetic code

The digital information is coded in triplets of nucleobases called codons

Each codon uses 3 of the 4 nucleobases and can express 4^3=64 possible combinations, equivalent to 6 bits of information ($64=2^6$)



Each codon uniquely identifies a single amino acid

Some aminoacids are coded by more than one codon (example of unplanned evolution)

Some codons are used as a “stop” signal of the sequence

Correspondence between RNA codons and amino acids

A=Adenine
G=Guanine
C=Citosine
U=Uracyl

Second letter

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

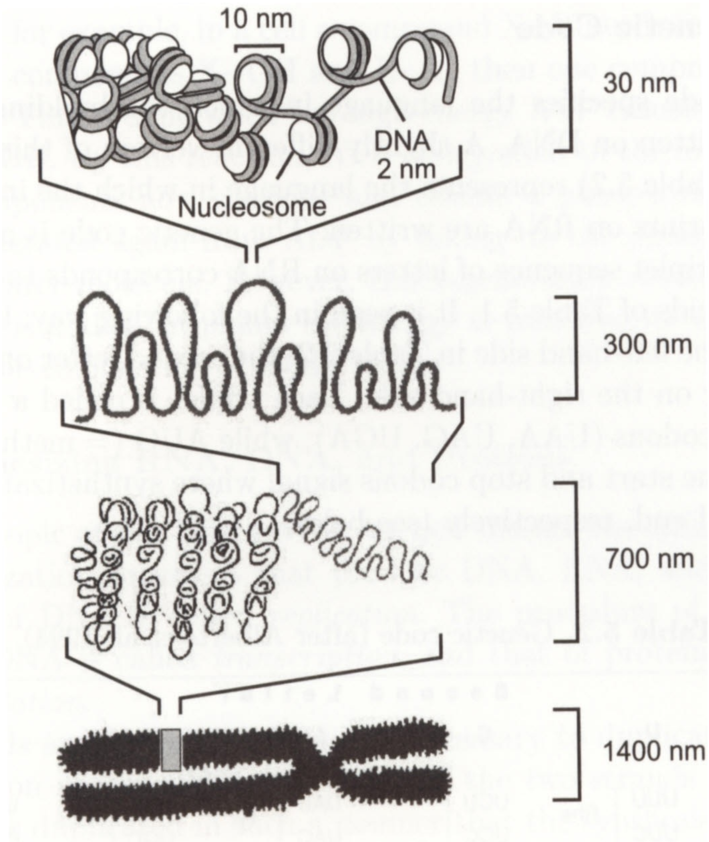
Genes

From the structural point of view, a gene is a sequence of nucleobases along a strand of a nucleic acid

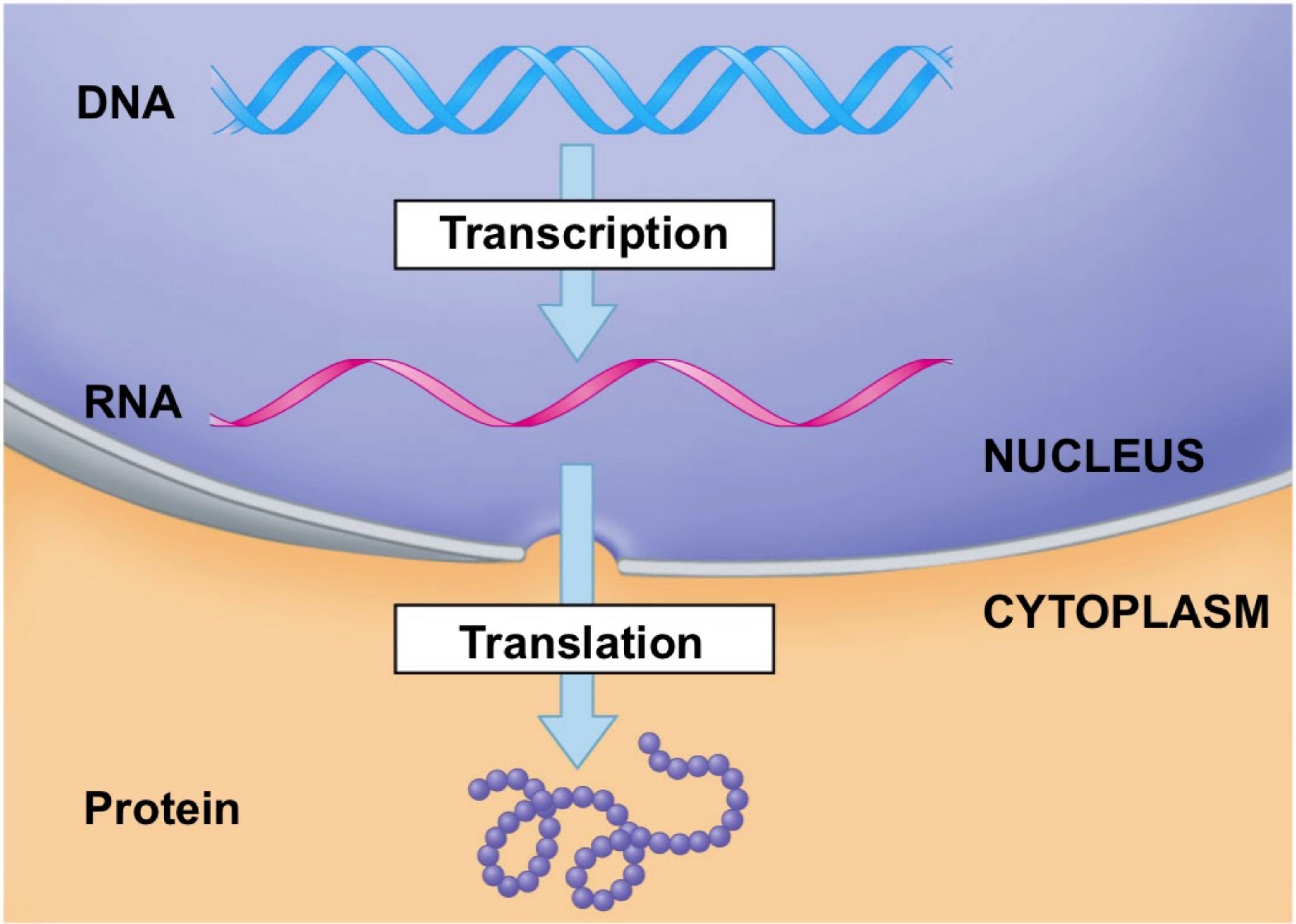
From the point of view of the information content, a gene is a sequence of instructions with a specific function

> Typically, a sequence that specifies how to build up a specific amino acid

In complex organisms, the number of genes is extremely high and this is why DNA needs to be stored in very compact structures, such as chromosomes



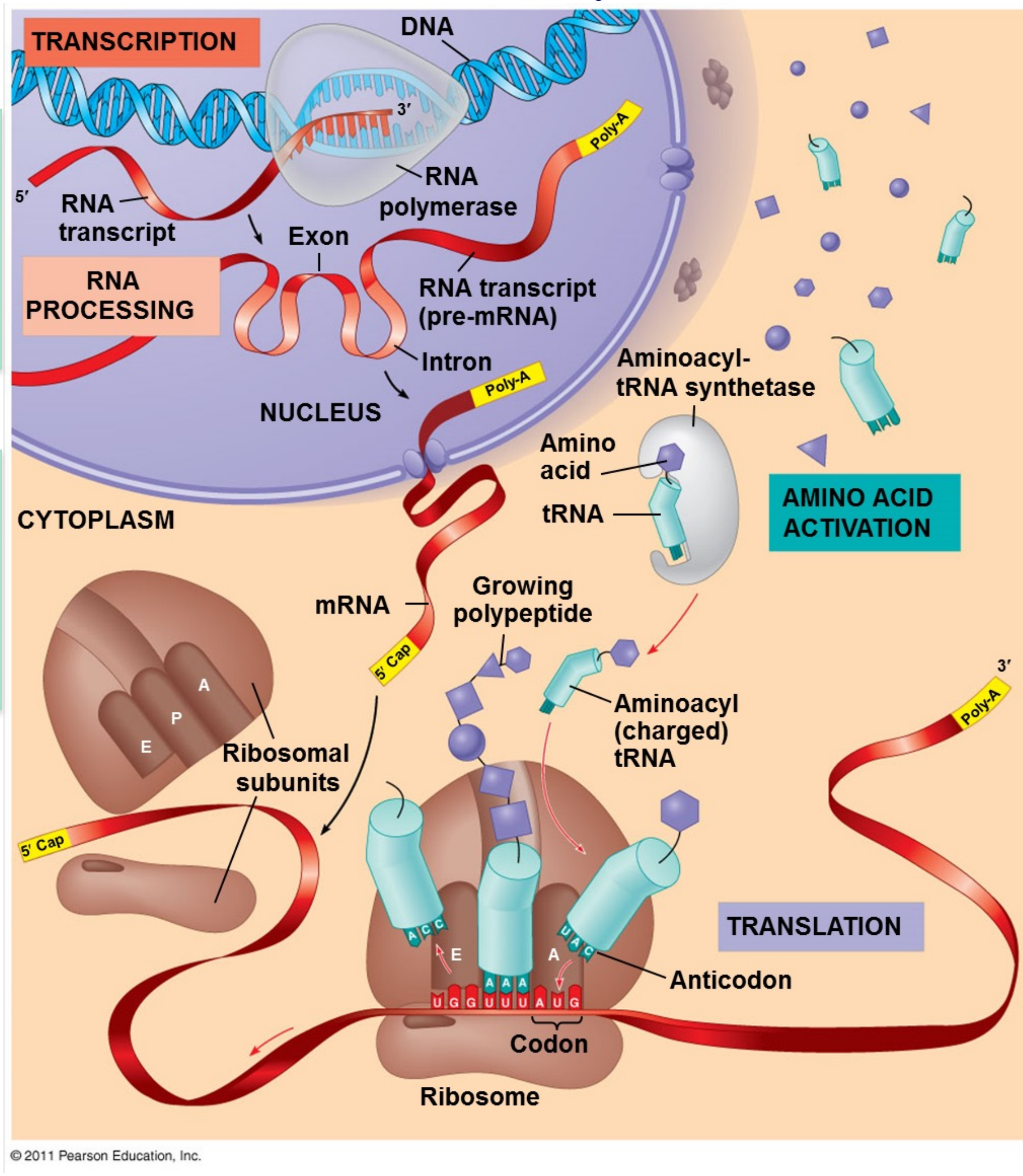
The "central dogma"



The molecular machinery of terrestrial life

Exons:
coding DNA
Introns:
non-coding DNA

mRNA:
messenger RNA
tRNA:
transfer RNA



The functioning of the molecular machinery requires the presence of nuclei acids and proteins (enzymes)

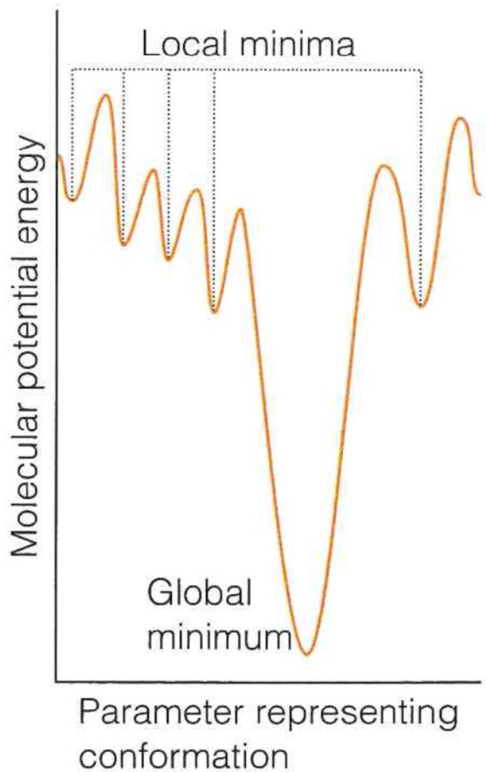
The information stored in the DNA is not sufficient, by itself, to generate an organism

Large part of the information required for the assemblage of macromolecules, organs, etc.
are implicitely contained in the laws of physics and chemistry

Example: folding of proteins

Protein folding

Spontaneous formation of geometrical configurations with lowest potential energy



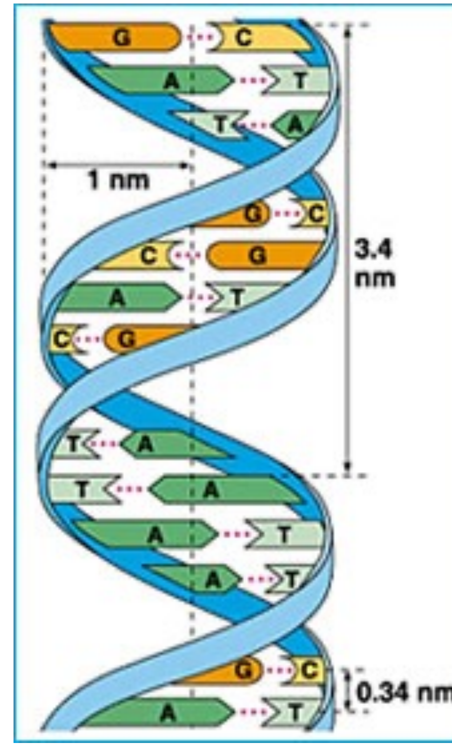
Molecular mechanics simulations

A plot of potential energy againts the
molecular geometry often shows
several local minima and a global minimum

Atkins & De Paula,
Physical Chemistry for
Life Sciences, 2011, p. 451

Appendix:
the DNA as a digital storage device

- The density of information (MB/unit volume) stored in the DNA can be calculated as follows
- The radius of the helix is $r_h \sim 1$ nm, while the perpendicular distance between adjacent nucleobasis $d_n \sim 0.34$ nm
- The volume occupied by a codon (3 nucleobasis) is therefore $V_{codon} \cong 3\, d_n\, \pi\, r_h^2 \cong 3.2$ nm^3
- Each codon has 6 bits of information (64 combinations=2^6), corresponding to 0.75 bytes
- The density of information is therefore
  0.75 B/(3.2 nm^3) = 0.23 B/nm^3 = 2.3x10^5 TB/mm^3
- This is the maximum density of information that can be obtained by compactified DNA strands



- Exercise: check that this density of information is largest than that of present-day storage devices by several orders of magnitude