

# Astrobiology

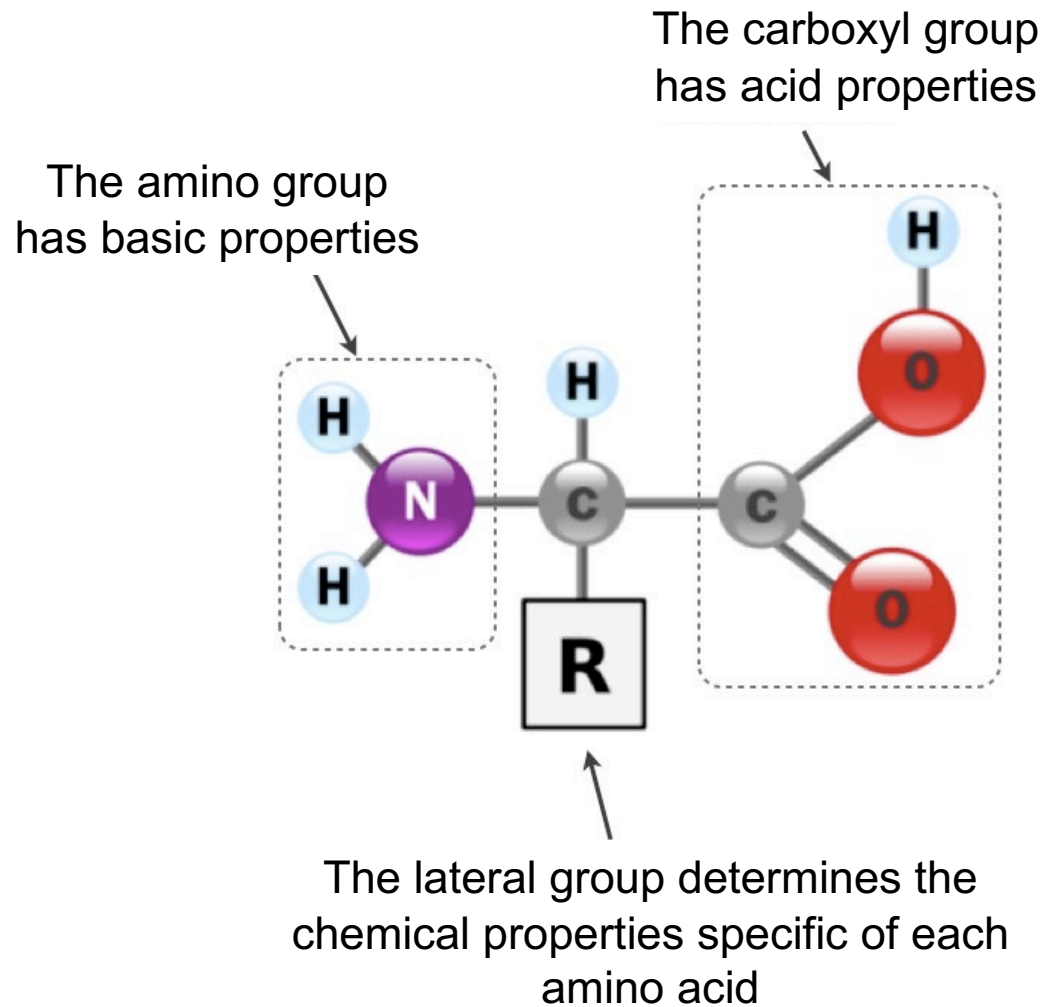
## Lecture 4

# Metabolic & genetic macromolecules

Trieste University, Academic Year 2021-2022  
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 peptides

Long, unbranched peptide chains are called polypeptides

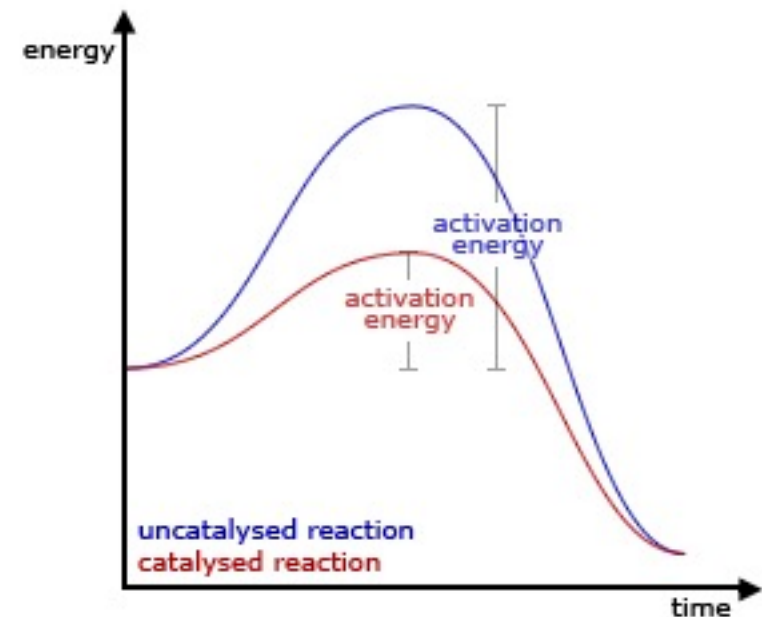
Proteins are formed by one or more chains of polypeptides

Molecular masses of proteins vary between  $\sim 10^3$  e  $\sim 10^6$  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

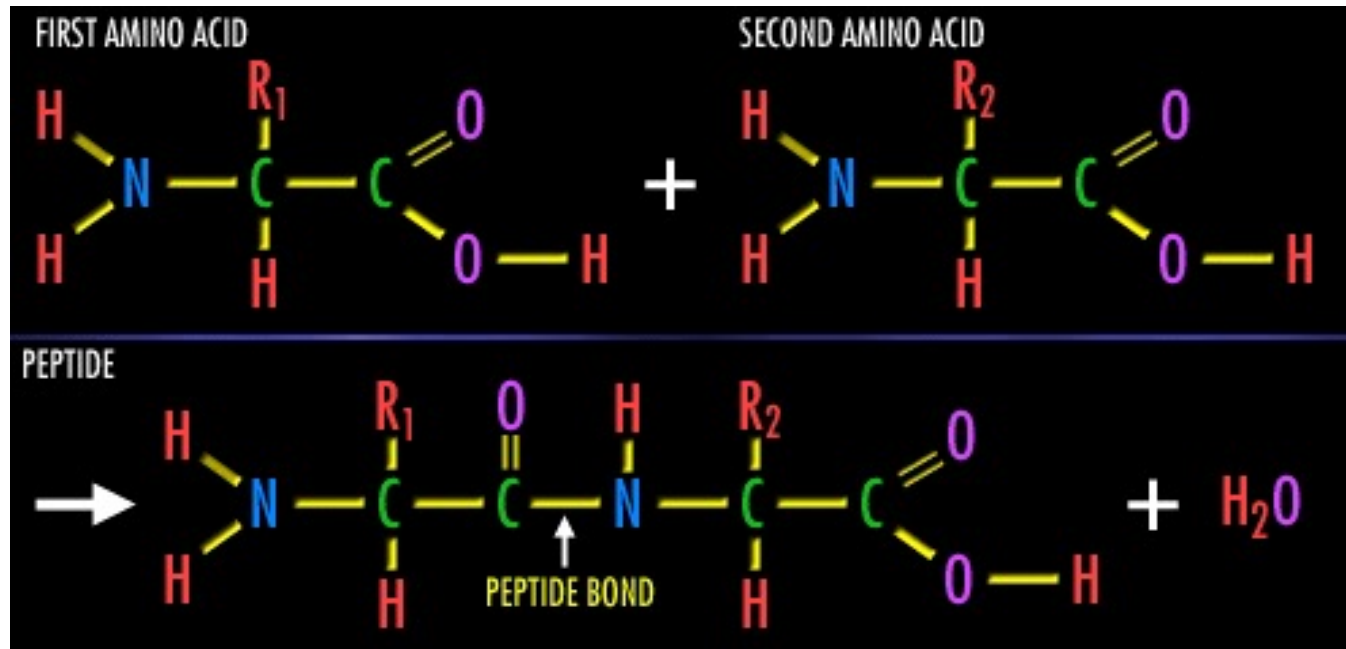
Organic chemistry allows for the existence of thousands of amino acids

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*

<i>Amino Acid*</i>	<i>Chemical Formula</i>	<i>Number of Atoms</i>
L-Alanine	C <sub>3</sub> H <sub>7</sub> O <sub>2</sub> N	13
L-Arginine	C <sub>6</sub> H <sub>15</sub> O <sub>2</sub> N <sub>4</sub>	27
L-Asparagine	C <sub>4</sub> H <sub>8</sub> O <sub>3</sub> N <sub>2</sub>	17
L-Aspartic Acid	C <sub>4</sub> H <sub>6</sub> O <sub>4</sub> N	15
L-Cysteine	C <sub>3</sub> H <sub>7</sub> O <sub>2</sub> NS	14
L-Glutamic Acid	C <sub>5</sub> H <sub>8</sub> O <sub>4</sub> N	18
L-Glutamine	C <sub>5</sub> H <sub>10</sub> O <sub>3</sub> N <sub>2</sub>	20
Glycine	C <sub>2</sub> H <sub>5</sub> O <sub>2</sub> N	10
L-Histidine	C <sub>6</sub> H <sub>9</sub> O <sub>2</sub> N <sub>3</sub>	20
L-Isoleucine	C <sub>6</sub> H <sub>13</sub> O <sub>2</sub> N	22
L-Leucine	C <sub>6</sub> H <sub>13</sub> O <sub>2</sub> N	22
L-Lysine	C <sub>6</sub> H <sub>15</sub> O <sub>2</sub> N <sub>2</sub>	25
L-Methionine	C <sub>5</sub> H <sub>11</sub> O <sub>2</sub> NS	20
L-Phenylalanine	C <sub>9</sub> H <sub>11</sub> O <sub>2</sub> N	23
L-Proline	C <sub>5</sub> H <sub>9</sub> O <sub>2</sub> N	17
L-Serine	C <sub>3</sub> H <sub>7</sub> O <sub>3</sub> N	14
L-Threonine	C <sub>4</sub> H <sub>9</sub> O <sub>3</sub> N	17
L-Tryptophan	C <sub>11</sub> H <sub>12</sub> O <sub>2</sub> N <sub>2</sub>	27
L-Tyrosine	C <sub>9</sub> H <sub>11</sub> O <sub>3</sub> N	24
L-Valine	C <sub>5</sub> H <sub>11</sub> O <sub>2</sub> N	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

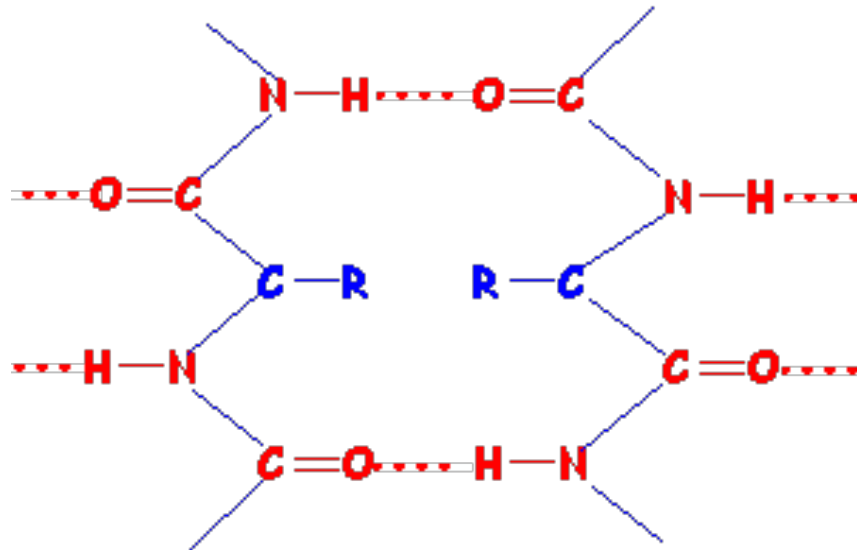
A sequence OC-NH is formed (peptide bond)

A water molecule is released each time a peptide bond is created

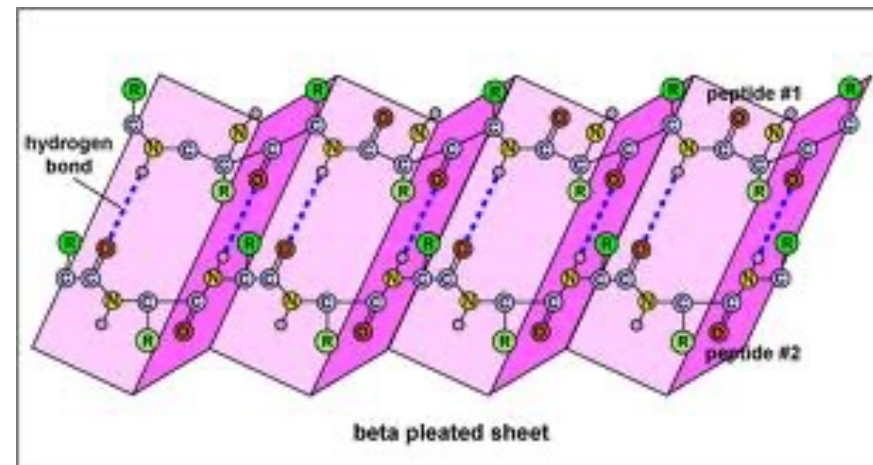
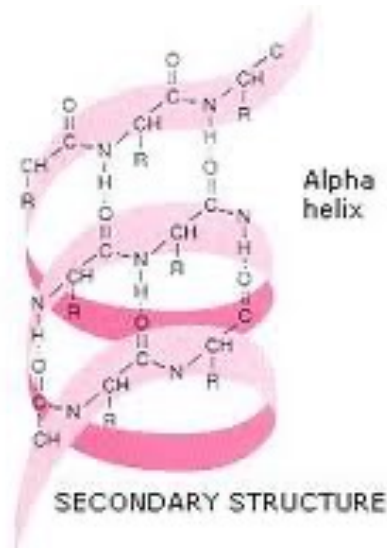
A water molecule is required to break a peptide bond

# From polypeptides to 3D proteins

## Importance of hydrogen bonds as intramolecular forces



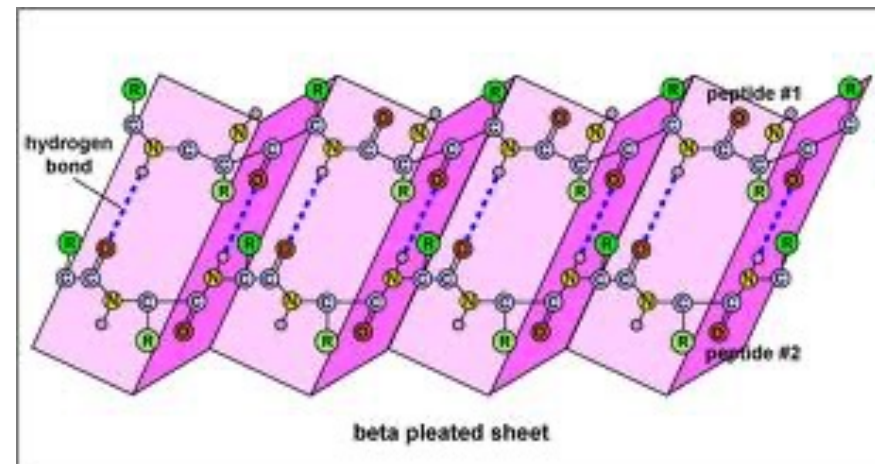
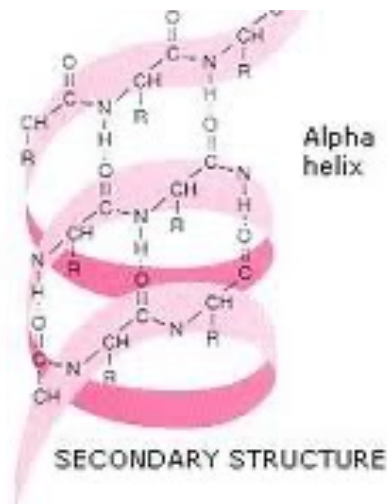
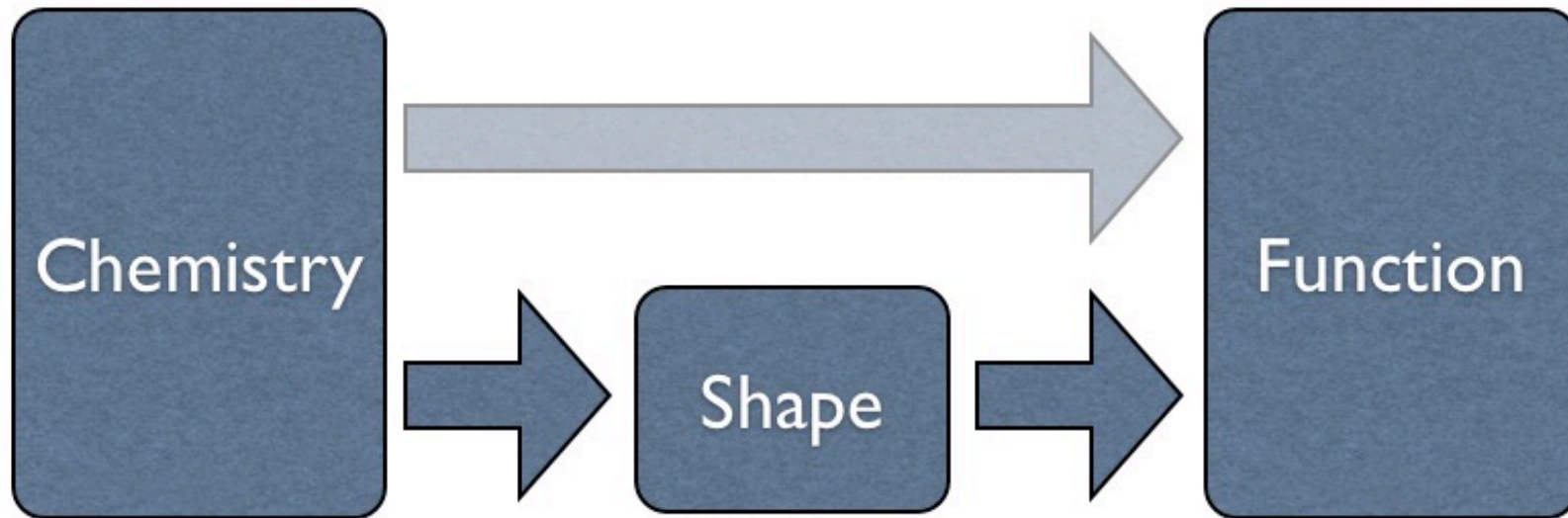
Spontaneous formation of geometrical configurations with lowest potential energy



# 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 isomers

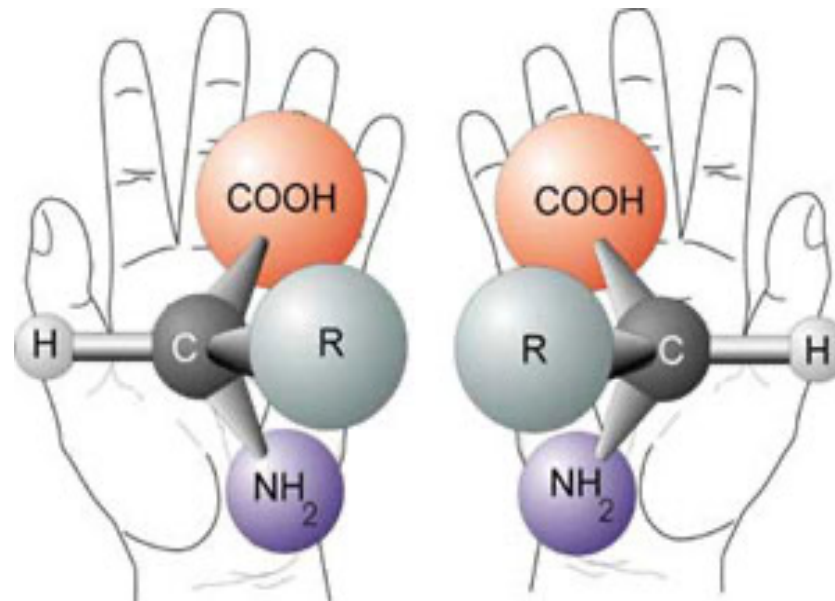
Chiral molecules are isomers with a center of symmetry (“stereocenter”)

They cannot be superimposed to their mirror image

The two mirror images of a chiral molecule are called enantiomers

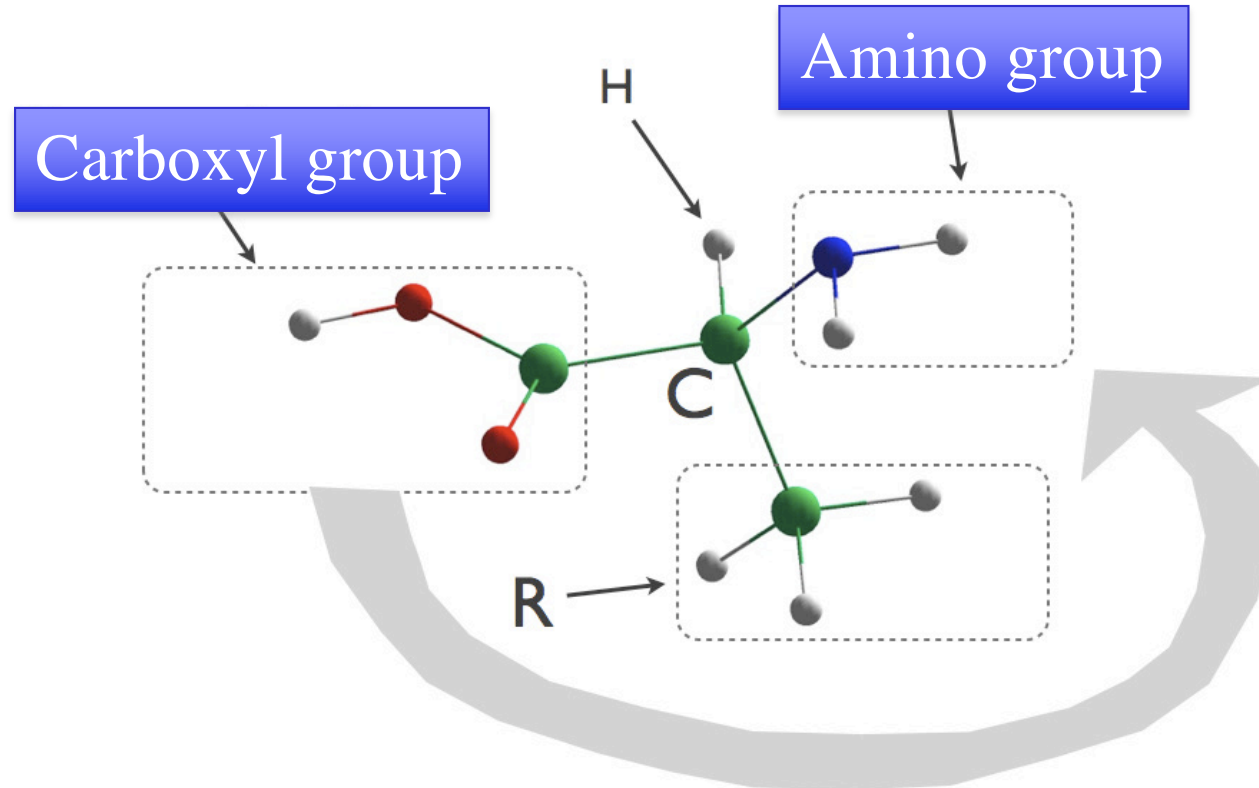
Amino acids are chiral

The carbon atom at the center of the amino acid is the stereocenter





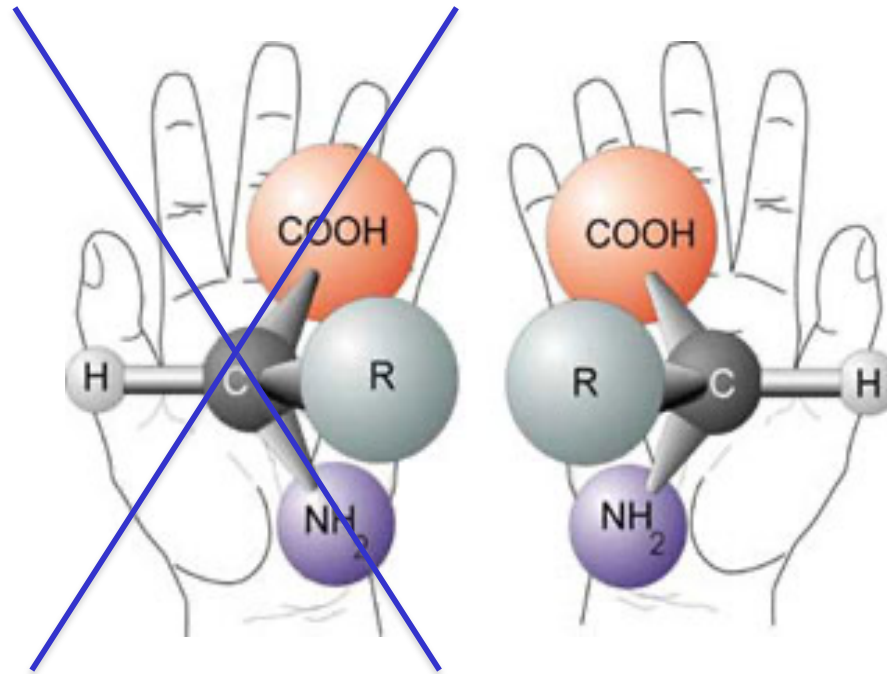
The two enantiomers of amino acids are called L and D according to the “CORN” convention



Example of L-type amino acid: L-alanine

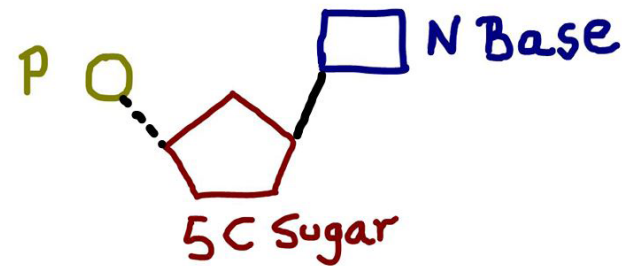
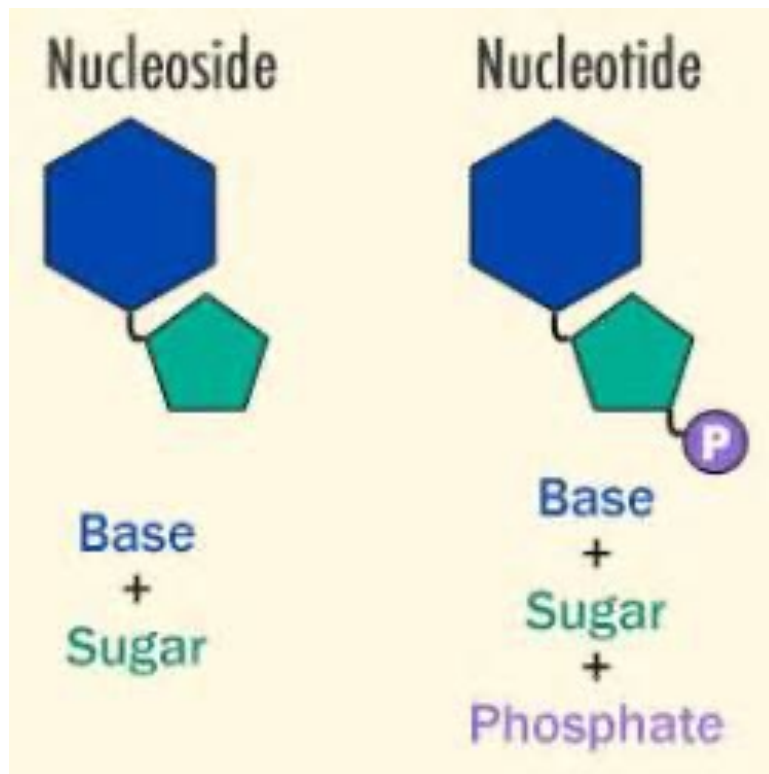
# Homochirality of biomolecules

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



# Nucleosides and nucleotides

Constituents of the DNA and RNA

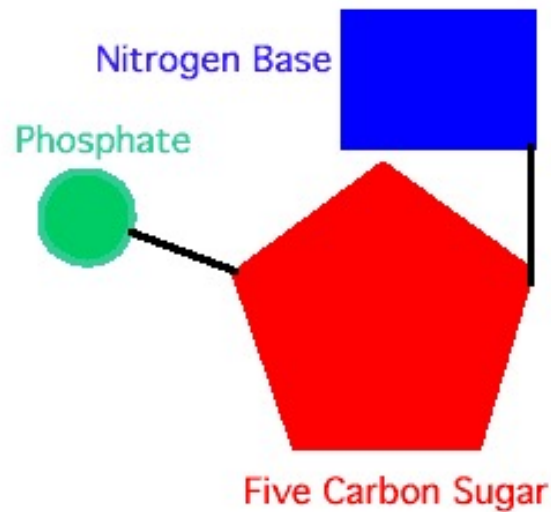


NUCLEO -  
SIDE = SUGAR + BASE  
TIDE = SIDE + PHOSPHATE

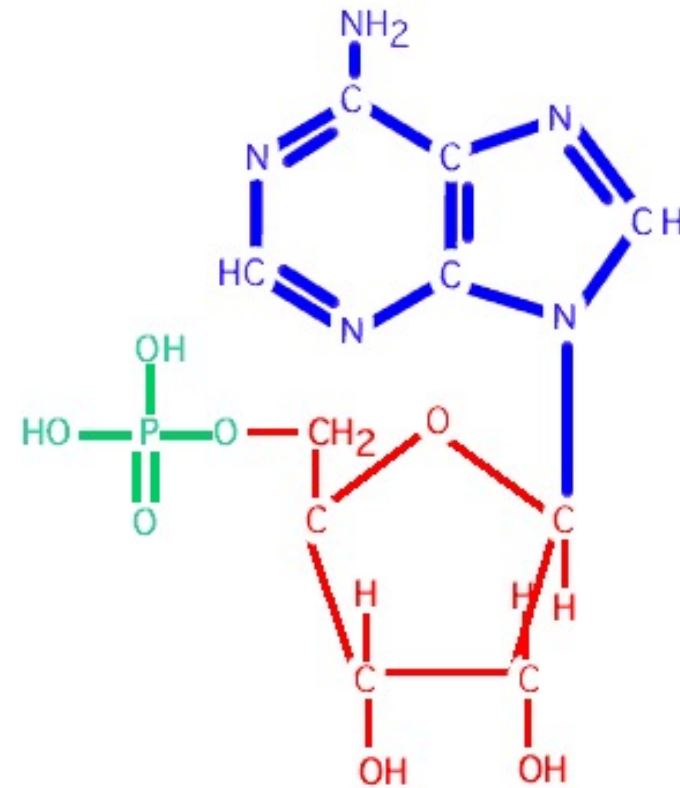
# Nucleotides

## Constituents of the DNA and RNA

Basic Nucleotide Structure



Example

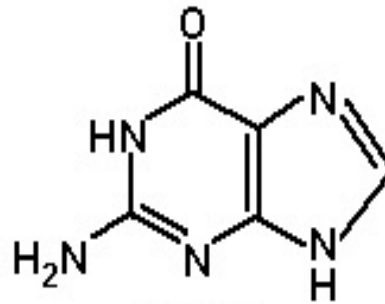


Adenosine 5' phosphoric acid

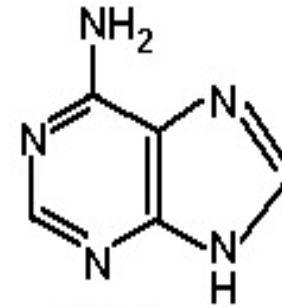
# Nitrogen bases

aromatic rings with N substitutions

Purines

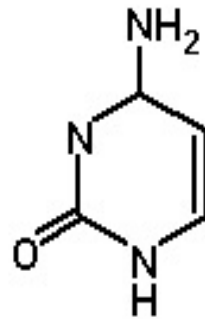


Guanin

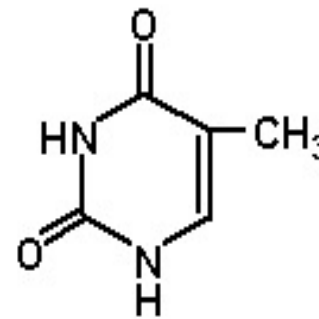


Adenin

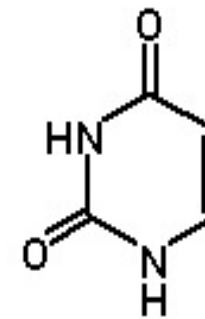
Pyrimidines



Cytosin



Thymin



Uracil

# Biological macromolecules

## Nucleic acids

Nucleic acids are polymers of nucleotides

Depending on the type of organism,  
they may contain  $\sim 10^6 - 10^8$  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 ribose sugars

## Has 4 types of nucleobases

Purines

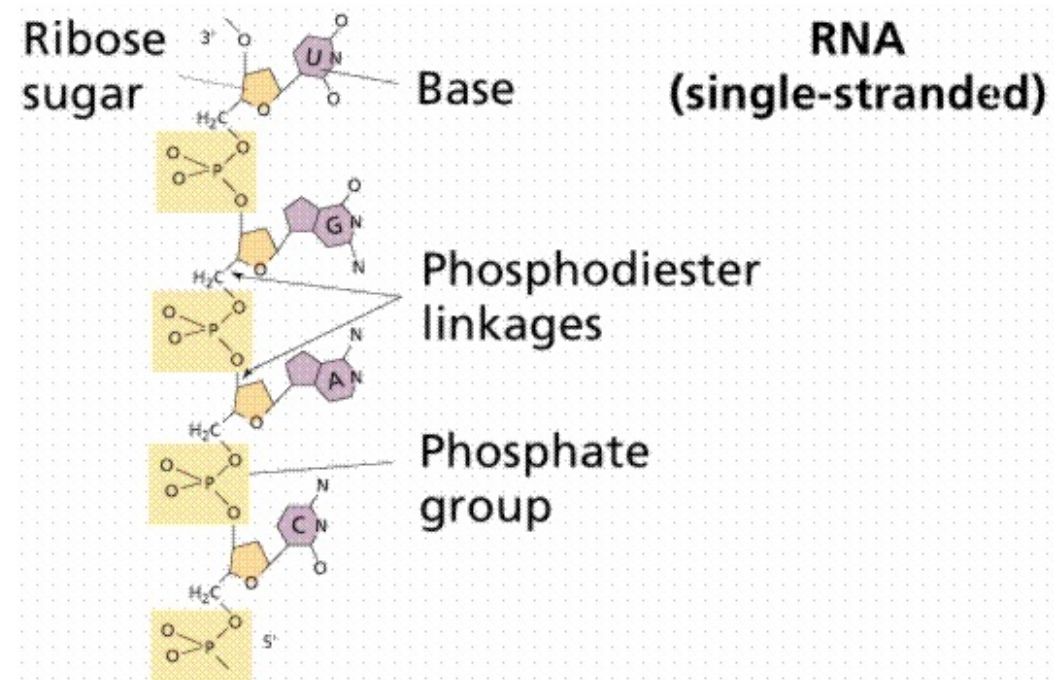
Adenine, Guanine

Pyrimidines

Cytosine, Uracil

## 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

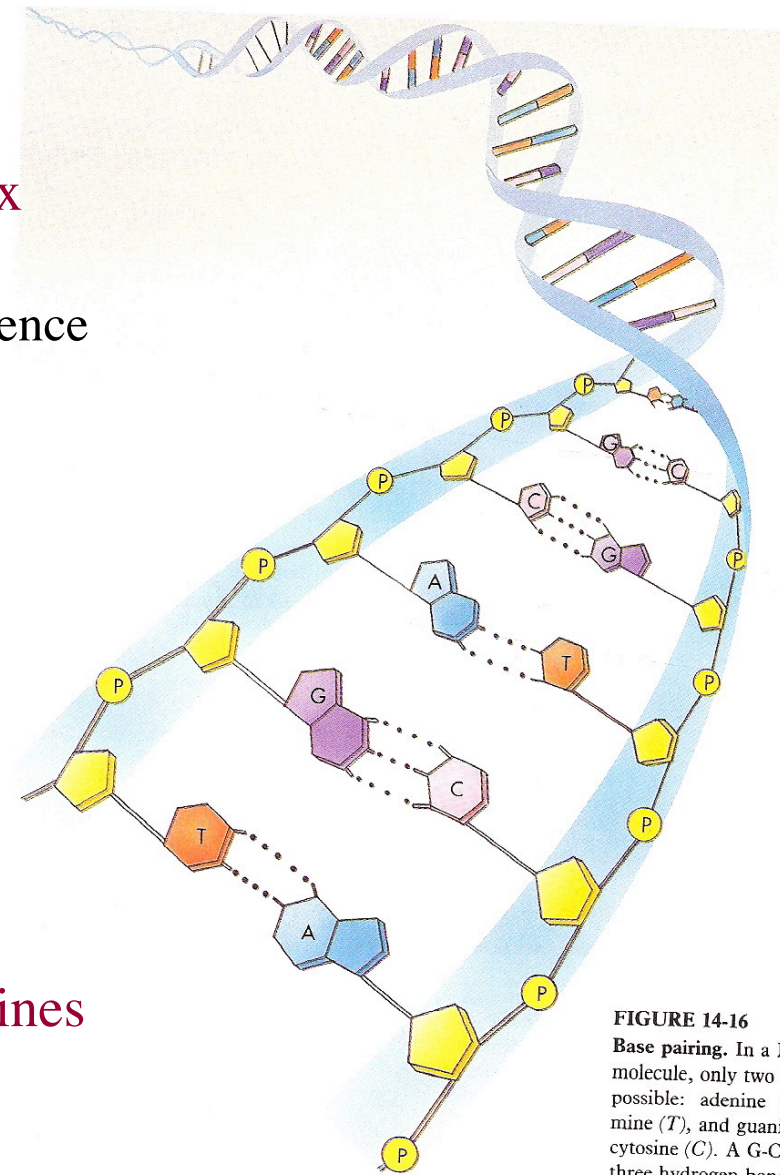
Adenine e Guanine

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 pairings are possible: adenine (A) with thymine (T), and guanine (G) with cytosine (C). A G-C base pair is held together by three hydrogen bonds; an A-T base pair, only two.

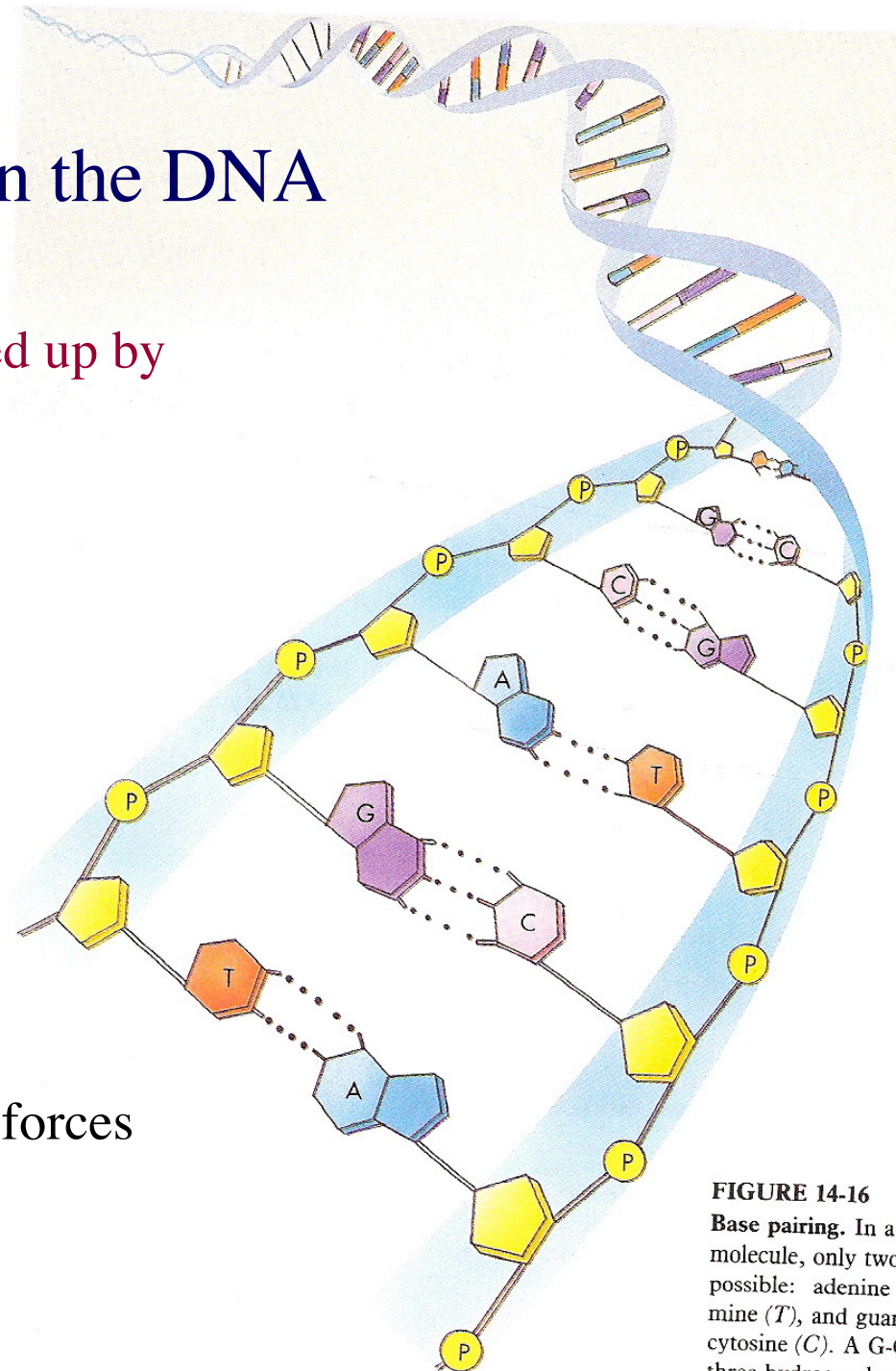


# Hydrogen bonds in the DNA

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

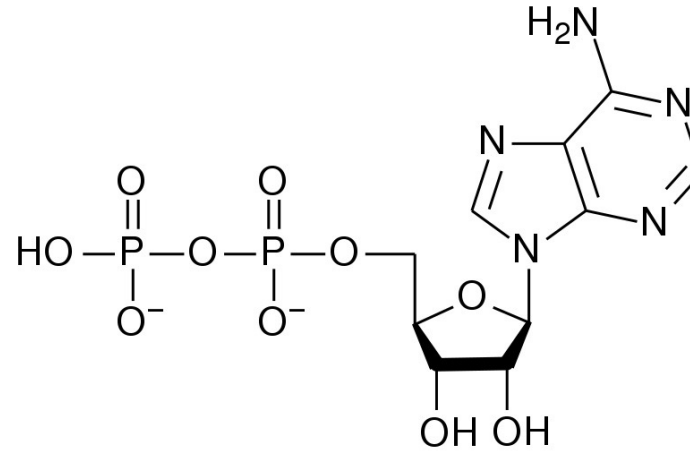
**FIGURE 14-16**

**Base pairing.** In a DNA molecule, only two base pairs are possible: adenine (A) and thymine (T), and guanine (G) and cytosine (C). A G-C base pair has three hydrogen bonds; an A-T base pair, only two.

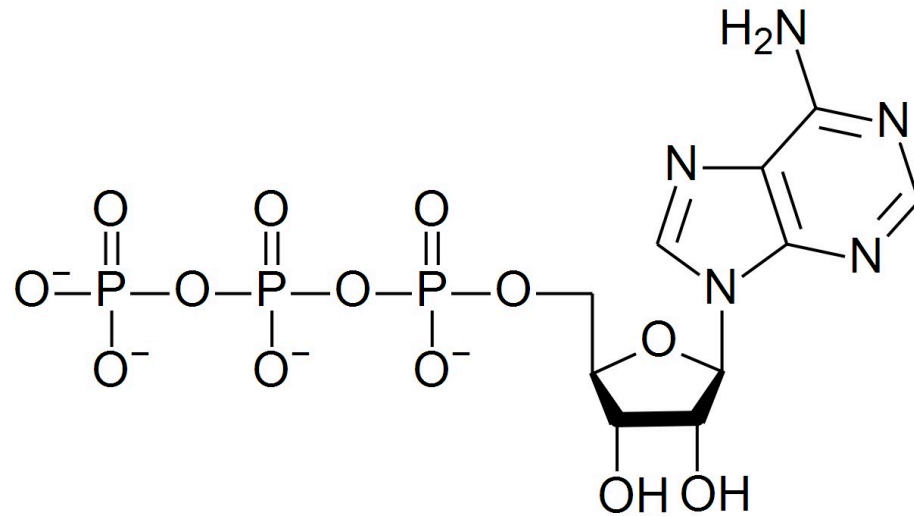
# ADP & ATP

## Energy exchange molecules

**ADP**  
Adenosine diphosphate



**ATP**  
Adenosine triphosphate



# Genetic information

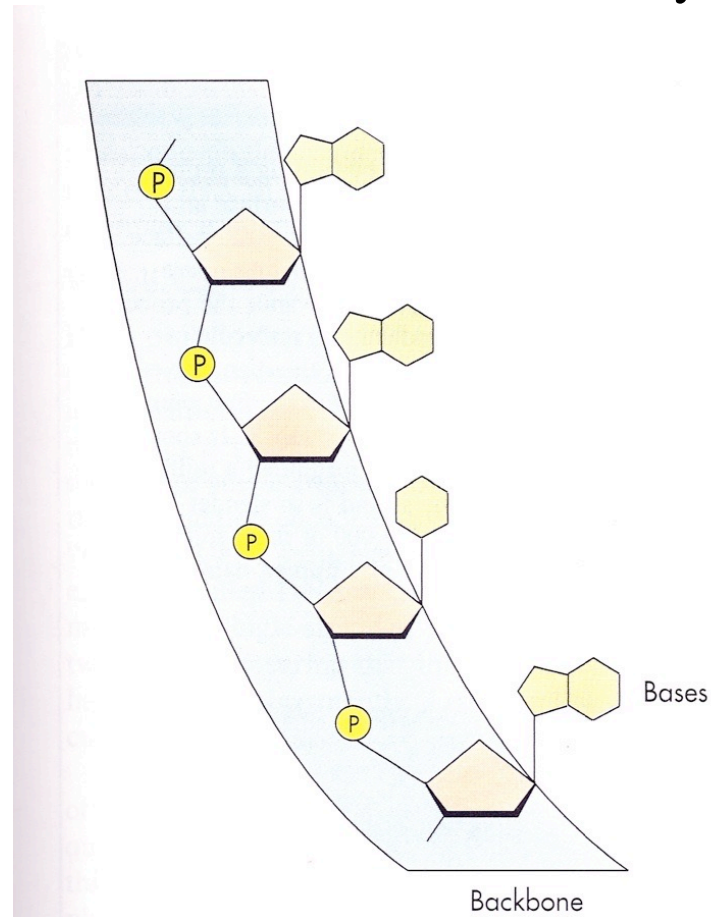
The order of the nucleobases attached to the backbone of the nucleic acids determines the genetic information, which is therefore stored in digital form

The order of the nucleobases 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

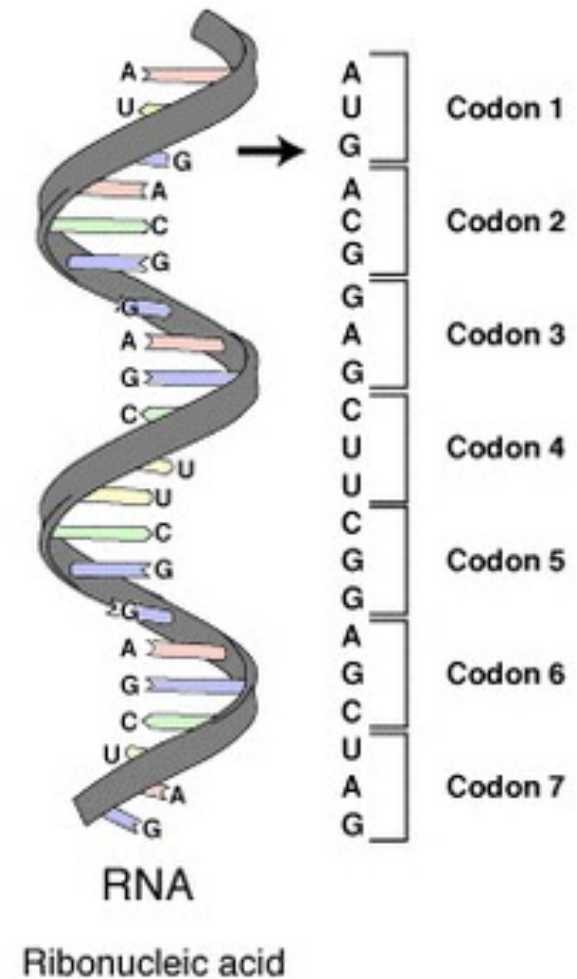
Terrestrial life has started to use digital information more than 3.5 billion years ago



# 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 amino acids 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=Cytosine  
U=Uracyl

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

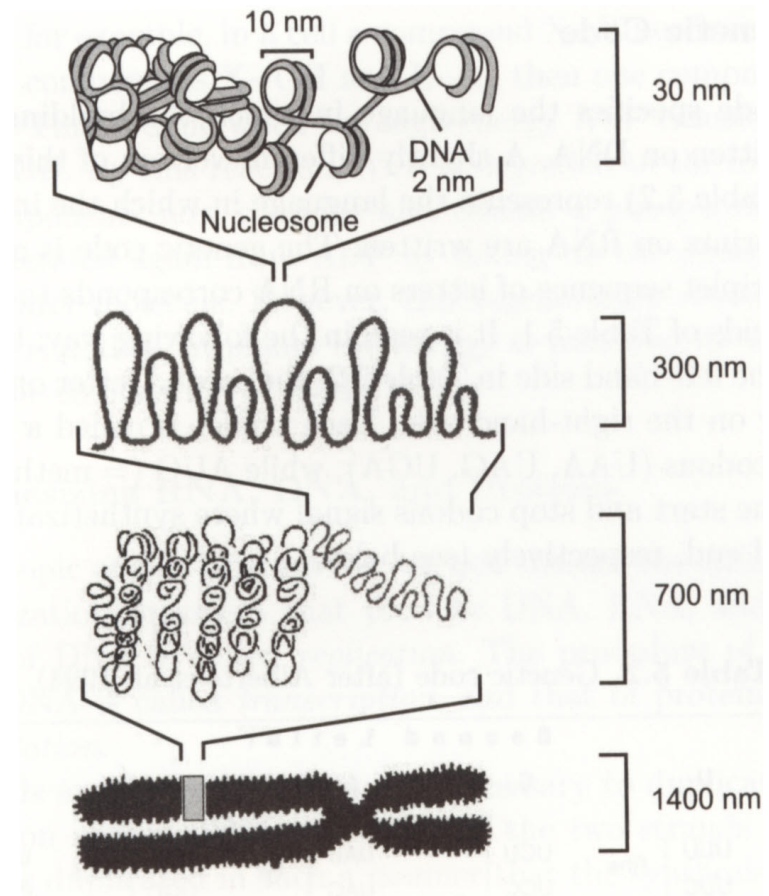
# Genes

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

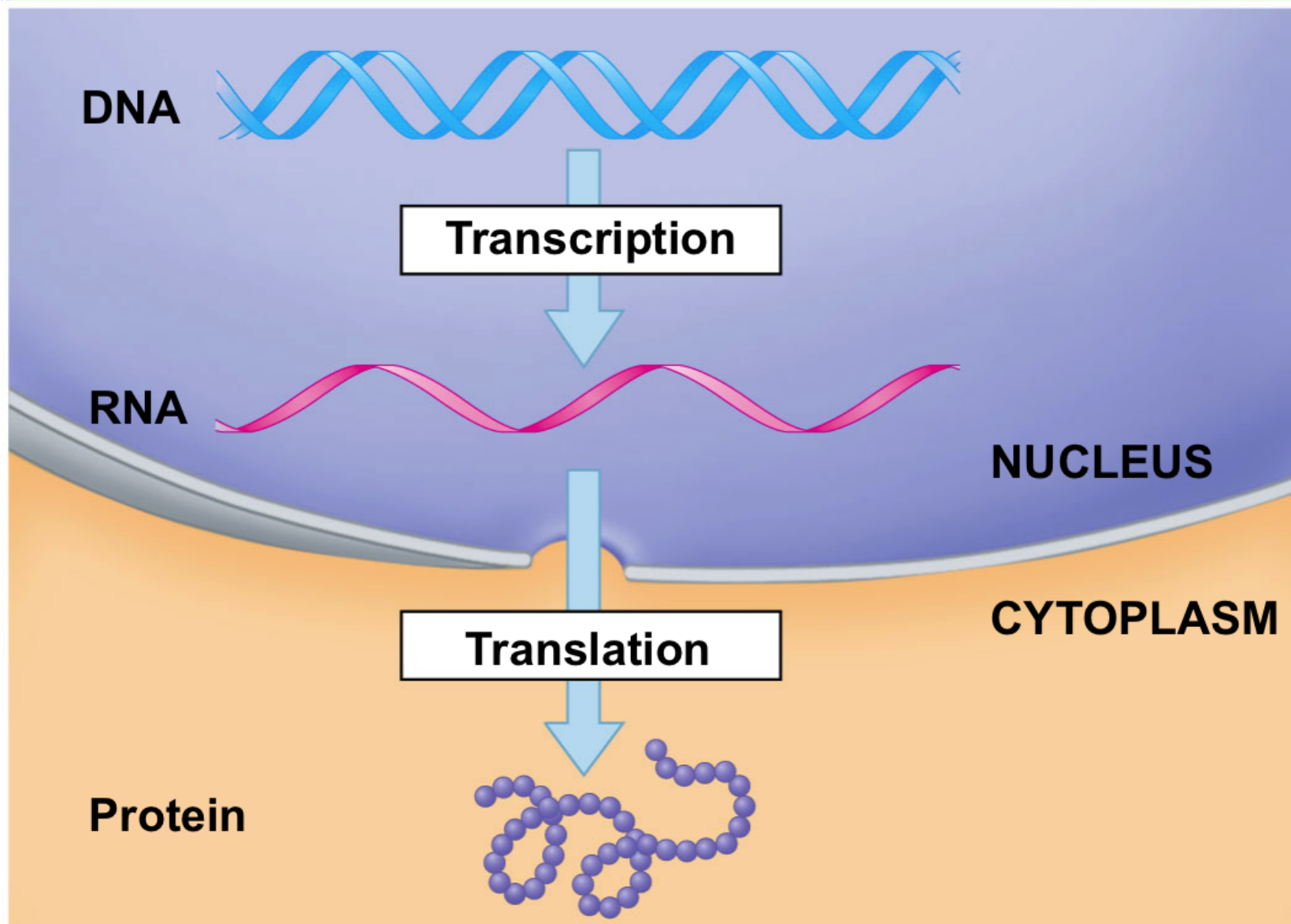
From the point of view of the information content, a gene is a sequence of instructions with a specific function

As an example, 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



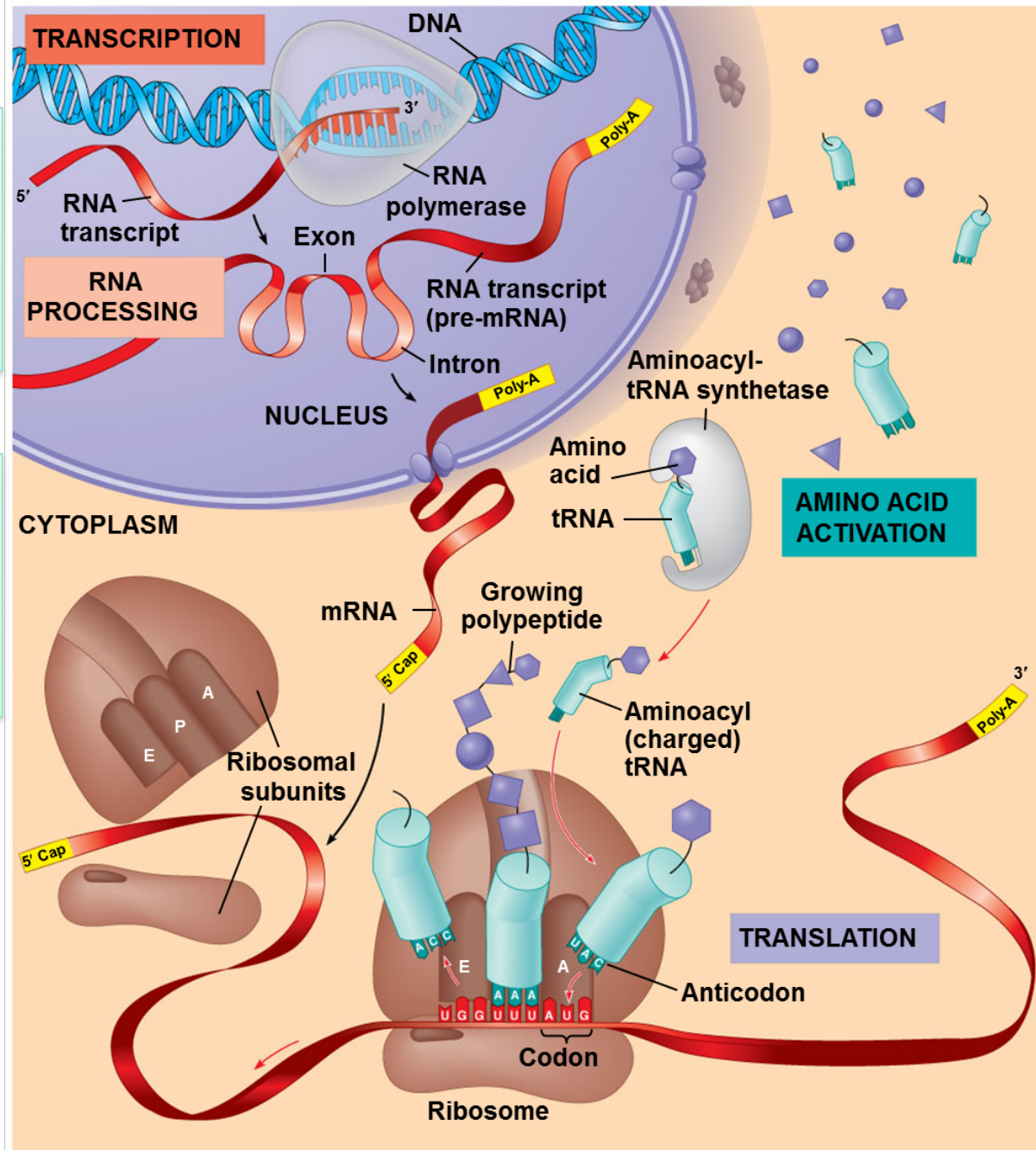
# The “central dogma”



# A closer look to the molecular machinery

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)



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  $V_{\text{codon}} \cong 3 d_n \pi r_h^2 \cong 3.2 \text{ nm}^3$
- Each codon has 6 bits of information ( $64$  combinations= $2^6$ ), corresponding to 0.75 bytes
- The density of information is therefore  $0.75 \text{ B}/(3.2 \text{ nm}^3) = 0.23 \text{ B}/\text{nm}^3 = 2.3 \times 10^5 \text{ TB}/\text{mm}^3$
- 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

