A The basis of the organization of living matter

A.1 Nucleic Acids
The nucleic acids have the function of storing the genetic information, replicating and transmitting it. DNA is mainly involved in storage. RNA retains the storage function in some viruses, but, in general, it is mainly involved in the transcription-replication of DNA (RNA) and in the translation from DNA (RNA) to proteins. Additionally, the pure enzymatic function of RNA was recently discovered (ribozymes), leading to the fascinating hypothesis of an RNA-only early phase of the life on earth (the RNA-World hypothesis [1]).

![Diagram of nucleic acids and their components]

Concerning the structure, nucleic acids are hetero-polymers whose monomers (the nucleotides) are composed of a phosphate group, a sugar (ribose in RNA or deoxyribose in DNA) and a Purine (adenine A, guanine G) or Pyrimidine (cytosine C, thymine T in DNA, uracil U in RNA) nitrogenous base. The chain forms through a complex reaction finally resulting in a dehydration and formation of a covalent bond between the O3’ of the sugar and the phosphate [2,3]. The coding function is contained in the mere sequence of the four bases in the polynucleotide chain, also called the primary structure (Fig A.1).

The secondary structure (Fig A.2) forms through the coupling of Purine bases with Pyrimidine bases through specific and rather strong hydrogen bonds, also called the Watson-Crick (WC) pairing: C-G and T-A. In RNA U takes the place of T, however Uracil is also able to form a "wobble" base pairing with G. This fact, together with the different electrostatics of the ribose with respect to the deoxyribose, turns out in a profound difference in the secondary and tertiary
structure of RNA with respect to DNA (and in their function): DNA is basically present exclusively in standard WC double strands, allowing only for transitions between different structural forms (A, B and Z DNA, see Fig A.3) that depends on hydration conditions, on the interaction with proteins and on the sequence. This structure and pairing conservation is the basis of the accurate storage and transmission of the genetic information. Conversely, RNA allows both for WC and non-WC pairing, can be present both in single and in double strands, and allows a great variety of different secondary and tertiary structures, that correspond to its different functions (Fig A.3).

(a) WC base pairing in DNA. (b) Base pairing for uracil: WC (AU) and "wobble" (GU). (c) Base pairing forms the anti-parallel double strand. (d) DNA double strand conformation. (e) A possible single strand conformation of RNA.

Fig A.2 The secondary structure of nucleic acids. (a) WC base pairing in DNA. (b) Base pairing for uracil: WC (AU) and "wobble" (GU). (c) Base pairing forms the anti-parallel double strand. (d) DNA double strand conformation. (e) A possible single strand conformation of RNA.

WC pairing (CG and AU) is exploited by RNA in the transcription phase, when, after denaturation, i.e. separation of the double helix in single strands, DNA is copied into the messenger RNA (mRNA). Here it is important to notice that the two single strands are not identical, rather they are complementary in the WC sense, and they are called the (+)-sense and (−)-sense (or anti-sense) strands. Only one of the two, the (+)-sense, is copied in a coding mRNA. mRNA subsequently migrates to the ribosome, a complex machinery composed of ~50 between proteins and rRNA (ribosomal RNA) chains with different functions (Fig A.3 (c,d,e)), where it is translated into a polypeptide chain. In this process the transfer RNA (tRNA, see Fig A.3 (b)) takes a fundamental role: each different tRNA recognizes a specific triplet of bases encoding for a specific amino-acid (bound to it), so that the subsequent coupling of tRNAs to the mRNA (occurring in the ribosome) results in the correct alignment of the amino-acids to form the polypeptide chain. This process is based on the genetic code, i.e. the association of specific triplets of bases to amino acids. This association is not one-to-one: the same amino acid can be coded by more than one base triplet.

The ribosome is one example of complex macromolecular aggregates where nucleic acids are associated. Another example is the nucleosome (Fig A.3 (f)), i.e. the basic unit of the DNA
packing in the cell nucleus. It consists of two turns of the DNA double helix wrapped around a core of histone proteins to form a sort of “bead”. These beads are then packed in chromatin.

![Fig A.3](image)

(a) The three structural form of DNA. A is favored in low hydration conditions and when DNA interacts with proteins or with other nucleic acids. Additionally, it is the standard structural form for the RNA double helix or of DNA-RNA double helices. B is the standard DNA structural form in normal or high hydration conditions. Z is a highly distorted left handed structure that can form when the sequence has a long alternation of purine pyrimidine (e.g. CGCGCG). (b) The secondary and tertiary structure of the tRNA (transfer RNA), that is involved in the translation of the mRNA (messenger RNA) into the polypeptide chain. (c) Secondary and tertiary structure of a ribozyme (i.e. RNA with enzymatic function). (d) Secondary and tertiary structure of a part of a ribosome. (e) The S70 bacterial ribosome, with mRNA and tRNA evidenced. The other RNA chains are in blue, the proteins in pink. (f) A nucleosome, The DNA is in cyan, the histone proteins in colors.

A.2 Proteins
Apart from the already mentioned functions of the nucleic acids, all the other functions in cells are performed by proteins. Every biochemical process is initiated, regulated, stopped, or catalyzed by proteins, proteins complexes (or proteins nucleic-acids complexes). The catalytic function is the most frequent, thus most of the proteins are enzymes.

Proteins are polypeptides, i.e. hetero-polymers whose monomers are the amino acids (see Fig A.4). These are basically constituted of an acidic part (the carboxy-terminus) and a basic part (the amino-terminus) bridged by a tetrameric carbon (also called alpha carbon, Cα), which is also bond
to an hydrogen and to a variable chemical group, the residue or side chain. Twenty different amino acid residues occur in nature, with different sizes, hydrophobicity and electrostatic properties (Fig. A.5). Due to the acidity-basicity of the termini, an amino acid assumes a zwitterionic form in water, with the negative charge localized on the carboxy-terminus and the positive one on the amino-terminus.

Fig A.4 (a) Amino acid chemical structure. (b) The peptide bond formation. (c) The geometry of the peptide bond and of the polypeptide backbone.

The polymerization reaction consists in the condensation of one amino-terminus with one carboxy-terminus with the elimination of a water molecule and the formation of the C-N peptide bond (Fig A.4 (b)). This reaction occurs in the ribosome and is the final step of the translation process. The geometry of the peptide group is planar and rigid. The C$_\alpha$-N-C$_\alpha$ conformation is almost always trans, with some rare exceptions. Conversely, the dihedrals angles $\phi$ and $\psi$ defining the relative orientations of the peptide groups are highly variable, since the single bonds C$_\alpha$-C and N-C$_\alpha$ are rotable. Their local value is mainly determined by the interactions of the side chains of the adjacent amino acids, i.e., ultimately, by the amino acid sequence (the primary structure). In turn, the $\phi$ and $\psi$ values are a measure of how the protein backbone configures itself, i.e. they measure the secondary structure. For a given protein or set of protein, the density distribution on the $\phi$-$\psi$ plane is called the Ramachandran map (see Fig A.6). Some region of the map are forbidden due to the sterical hindrance of the side chains, and at each allowed region correspond a specific secondary structure: (H, $\pi$, 3-10, left handed) helix or (B) strands. Less defined secondary structure configurations (loops, or globular structures) occupy the intermediate regions of the Ramachandran map. The secondary structures are stabilized by local hydrogen bonds (helices) and/or local VdW interactions (helices and sheets).

The tertiary structure relies on the relative positioning and inter-configuration of secondary structures, and is stabilized by hydrogen bonds (sheets), VdW and electrostatic and hydrophobic interactions (sheets, helices bundles). Finally, a protein can have a quaternary structure, when more chains already folded in the tertiary structures are associated in a single protein.
<table>
<thead>
<tr>
<th>Ala</th>
<th>Arg</th>
<th>Asn</th>
<th>Asp</th>
<th>Cys</th>
<th>Gin</th>
<th>Glu</th>
<th>Gly</th>
<th>His</th>
<th>Ile</th>
<th>Leu</th>
<th>Lys</th>
<th>Met</th>
<th>Phe</th>
<th>Pro</th>
<th>Ser</th>
<th>Thr</th>
<th>Trp</th>
<th>Tyr</th>
<th>Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>R</td>
<td>N</td>
<td>D</td>
<td>C</td>
<td>Q</td>
<td>E</td>
<td>G</td>
<td>H</td>
<td>I</td>
<td>L</td>
<td>K</td>
<td>M</td>
<td>F</td>
<td>P</td>
<td>S</td>
<td>T</td>
<td>W</td>
<td>Y</td>
<td>V</td>
</tr>
</tbody>
</table>

**Fig A.5** The 20 natural amino acids. The three letter and one letter notation is reported at the bottom.
A.6 The hierarchical organization of the protein structure. (b) The Ramachandran map for a generic protein. The green region corresponds to right handed helices, the red to left handed helices, the blue to sheets-strands or other extended structures. The contours of weakly allowed regions are in cyan and yellow. All the rest are forbidden regions.

A.3 Cells, Viruses and other self replicating entities

Cells can be classified in prokaryotic and eukaryotic. Prokaryotic cells can only be found in some simple unicellular organisms (bacteria and Archaea) and are characterized by the absence of the cell nucleus (with genetic material floating in the cytoplasm) and other cell structures. Conversely, the more evolved unicellular organisms and all the multi-cellular organisms are formed by eukaryotic cells, whose main characteristic is the presence of a nuclear membrane separating the genetic material (DNA organized in chromatin or in chromosomes, and some structures necessary for DNA replication and transcription) from the cytoplasm.

A.7 (a) Prokaryotic cell (b) Eukaryotic cell (c) Nucleus of a prokaryotic cell.
Additionally, in the cytoplasm of the eukaryotic cells many complex and diverse organelles and structure are present (Fig A.7). Their description and function is reported in Table A.1.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Description</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytoskeleton</td>
<td>Network of protein filaments</td>
<td>Structural support; cell movement</td>
</tr>
<tr>
<td>Flagella (cilia, microvilli)</td>
<td>Cellular extensions</td>
<td>Motility or moving fluids over surfaces</td>
</tr>
<tr>
<td>Centrioles</td>
<td>Hollow microtubules</td>
<td>Moving chromosomes during cell division</td>
</tr>
<tr>
<td>Plasma membrane</td>
<td>Lipid bilayer in which proteins are embedded</td>
<td>Regulates what passes into and out of cell; cell-to-cell communication</td>
</tr>
<tr>
<td>Endoplasmic reticulum</td>
<td>Network of internal membranes; forms compartments and vesicles</td>
<td>Rough: type processes proteins for secretion and synthesizes phospholipids; smooth: type synthesize fats and steroids</td>
</tr>
<tr>
<td>Nucleus</td>
<td>Structure bounded by double membrane; contains chromosomes or chromatin</td>
<td>Control center of cell; directs protein synthesis and cell reproduction</td>
</tr>
<tr>
<td>Golgi complex</td>
<td>Stacks of flattened vesicles</td>
<td>Modifies and packages proteins for export from cell; forms secretory vesicles</td>
</tr>
<tr>
<td>Lysosomes</td>
<td>Vesicles derived from Golgi complex that contain hydrolytic digestive enzymes</td>
<td>Digest worn-out mitochondria and cell debris; play role in cell death</td>
</tr>
<tr>
<td>Mitochondria</td>
<td>Bacteria-like elements with inner membrane</td>
<td>Power plant of the cell; site of oxidative metabolism; synthesis of ATP</td>
</tr>
<tr>
<td>Chromosomes</td>
<td>Long threads of DNA that form a complex with protein</td>
<td>Contain hereditary information</td>
</tr>
<tr>
<td>Nucleolus</td>
<td>Site of rRNA synthesis</td>
<td>Assembles ribosomes</td>
</tr>
<tr>
<td>Ribosomes</td>
<td>Small, complex assemblies of protein</td>
<td>Site of protein synthesis</td>
</tr>
</tbody>
</table>

Table A.1 Procaricotic cell structures and organelles, and their function.

Viruses are a very interesting example of self-replicating organisms. They consist of a protein capsule (capsid) containing DNA or RNA (1000-200000 base pair) with all the information necessary for their replication. The replication, however, needs a host cell that dies afterwards, making viruses parasites. The discussion about their being living organisms or not is still open, as well as the question about their origin. There are several hypotheses: escaping from free-living organisms (like bacteria), or co-evolution with more complex organisms. In any case, the large variety of existing viruses, exploiting different infection mechanisms and biochemistry and adapted to infect any living organisms, suggest their very early appearance in the story of life evolution.

The virus classification can be based on different criteria. The hierarchical virus classification system was proposed in 1962 Lwoff, R. W. Horne, and P. Tournier (see Fig. A.8 (a)). It is based on the following properties of the viral particle: (1) Nature of the nucleic acid: RNA or DNA, and genome architecture (double strand – ds – or single strand – ss – positive sense (+) or anti-sense (-) direction of the strand) and the number and structure of chains); (2) Symmetry of the capsid; (3) Presence or absence of an envelope, a membrane coating the capsid probably a modification of the host cell membrane; (4) Dimensions of the virion and capsid.

The Baltimore system of virus classification is alternative and complementary to the hierarchical classification, an provides a useful guide with regard to the various mechanisms of viral genome replication. The central theme here is that all viruses must generate (+) strand mRNAs from their genomes, in order to use the cell machinery to produce proteins and replicate themselves. Thus, different genome architectures (ds or +/- ss) must follow different routes for replication, and seven different classes can be distinguished according to this (see Fig. A.8(b)).

As an example, we consider the HIV (human Immunodeficiency Virus). It is a an enveloped, icosahedral (+)ss (diploid) retro-virus. Its replication cycle is among the most complex ones (Fig. A.9, [4]). After binding to the cell membrane with the aid of the envelope proteins, the viral RNA enters the cell. Exceptionally, in retro-viruses the viral RNA is (+)sense, but it cannot be directly used by the cell as mRNA. It must be first reversely-transcribed into DNA through the reverse
transcriptase, an enzyme that the virus itself inject into the cell (this is why it is called retrovirus).

From this process, a double strand viral DNA is obtained, which is integrated in the host cell genome through another viral enzyme, the HIV-integrase. After this, the cell machinery follows the usual DNA replication steps: transcription in mRNA and translation of the viral DNA in the viral poly-proteins. These are cut in pieces by the HIV-protease (also injected together with viral RNA) to generate all the set of the viral proteins: the capsid and envelope proteins, the HIV-reverse transcriptase, integrase and protease. All of these are subsequently assembled in many new viral particles that exit the cell.

The current anti-AIDS therapies target the enzymes involved in any of the steps of the HIV replication: anti-AIDS drug are either HIV-reverse transcriptase, or HIV-integrase or HIV-protease inhibitors. However these drugs have usually a therapeutic efficiency limited in time, because the virus is able to mutate very rapidly reducing the affinity of its enzymes to drugs, that become
soon inefficient, so that new, more efficient drugs, must be constantly designed.

![Image](https://via.placeholder.com/150)

*Fig A.9* (a) A pictorial representation of HIV. (b) The HIV replication cycle.

Out of the conventional virus classification, we mention the sub-viral agents, i.e. viroids (small infectious nucleic acids fragments without capsid that cannot code for any protein, only replicate themselves) and satellites (small virus or nucleic acids fragment that can only co-infect a cell together a master-virus).

Finally, the hypothesis of the viral origin of the eukaryotic cell nucleus is worth mentioning: the nucleus may have evolved from a persisting large DNA virus that made a permanent home within prokaryotes. Some support for this idea comes from sequence data showing that the DNA polymerases (a DNA copying enzyme) of eukaryotes and bacteria are more closely related to similar enzymes found in viruses than they are to each other. Additionally, some researchers now believe that viruses have been instrumental in assembling the various molecular components that define the cell types. This would indicate that the virus is among the first organisms appearing on earth.

![Image](https://via.placeholder.com/150)

*FigA.10* (a) Model of the structural conversion of a prion. (b) Model of a generic amyloid fibril

The life-or-not-life nature of the prions is still more questionable than that of viruses. Prions are protein infectious agents. These proteins are normally residing in an non infected organism. The disease takes place when a portion of the protein mis-folds, usually a helical portion becomes a strand (Fig A.10 (a)). This causes either a wrong behavior of the protein, or an aggregation with the formation of fibers and plaques that accumulate in tissues[5,6]. The aggregation mechanism is not clear, however it is believed to be the amyloid fibril formation that exploit the tendency of strands to aggregate in sheets (Fig A.10 (b)). This is for instance the pathogenesis mechanism of the Transmissible Spongiform Encephalopathy (TSE, mad cow disease). In this case, the amyloid aggregates deposit in neurons and in the brain tissue, causing the known symptomatology and dead. The prion multiplication and transmission mechanism is particularly interesting: the mis-
folded form induces mis-folding of the normal form (Fig A.11), i.e. in some sense the prion is able to replicate itself. The mis-folding might be initiated by a mutation, and be initially sporadic and quiescent. But when a certain level is reached the propagation of the infectious form can be very fast. This is a noticeable example of protein-only replication mechanism. Additionally, the prion is transmissible, either through the contamination of infected tissues (for instance, through food), or genetically, through the transmission of the mutation.

![Diagram of PrP and PrP^sc](image)

**Fig A.11** The mechanism of the replication and transmission of the TSE prion.

References

[3] A list of Biology and biochemistry on-line textbooks

  http://web.mit.edu/esgbio/www/
  http://ftp.bio.umass.edu/
  http://courses.cm.utexas.edu/emarcotte/ch339k/fall2005/Lecture-Ch4.html
  http://universe-review.ca/F11-monocell.htm