Structural Bioinformatics (C3210)

DNA and RNA Structure
Importance of DNA/RNA 3D Structure

Nucleic acids are essential materials found in all living organisms. Their main function is to maintain and transmit the genetic code. This information is stored in the form of long polymer chains. Although the information they carry is one-dimensional, it is essential to understand the 3D structure of nucleic acids. This 3D structure dictates their organization, functions and interactions with proteins.
DNA 3D Structure

DNA is made of two helical strands of polynucleotides, one going up and the other going down.
Nucleotides are composed of three covalently bonded components: base, sugar and phosphate. The phosphate group is connected to carbon number 5 of the sugar. The sugar is connected to the base by a covalent single bond between carbon number 1 of the pentose ring and a nitrogen atom on the aromatic ring of the base. Nucleosides consist of the base and the sugar without the phosphate.
The Building Blocks

Three types of chemicals make up the building blocks for nucleic acids: an aromatic base, a sugar ring and a phosphate group. The sugar and the phosphate constitute the non-specific backbone of the DNA. The base constitutes the specific part which actually holds the information and accounts for most of the interactions with other molecules.
Base

The group that gives each nucleic acid unit its specificity is the organic base. DNA contains two purine bases (adenine and guanine) and two pyrimidine bases (cytosine and thymine).

In RNA the thymine base is replaced by uracil.

Polar atoms in the ring or attached to the ring are capable of creating hydrogen bonds with polar atoms of other bases.
Sugar

The sugars in DNA and RNA are pentoses (5 carbons). The ring of the sugar is flexible and non-planar (in contrast to the rings of the base), and can adopt several conformations.

![Sugar structure](image)
Phosphate

The inorganic acid $\text{H}_3\text{PO}_4$ (phosphoric acid) gives the nucleic acids an overall net negative charge.

Most of the interactions between the DNA and proteins which are not specific to the DNA sequence are with the phosphate groups.
Nomenclature of Nucleotides and Nucleosides

Nucleotides are the monomers of nucleic acids. They may have one, two or three phosphate groups; the sugar can be ribose (in RNA) or deoxy-ribose (in DNA) and the base can be a purine or a pyrimidine base. Mono-phosphate nucleotides are the structural units of nucleic acids.
DNA is the Genetic Material

DNA is the genetic material of living cells. The language of the genetic code contains only 4 letters (A,T,G,C). Information about protein synthesis, which is the main machinery in the cells, is carried as shown in the table. Each triplet of nucleotides ("codon") encodes one amino acid.
The Double Helix Structure

Each strand of DNA has directionality. In other words, one edge is not identical to the other and there is no symmetry along the strand. The molecule has two different edges, termed the 3-prime edge and the 5-prime edge. The DNA molecule is arranged such that the two strands are lying next to each other but in opposite directions, such that the 3' edge of one strand is adjacent to the 5' of the other strand and vice versa. A phosphate group is always found in the 5' edge and hydroxyl group is found in the 3'.

![Diagram of DNA strands with 5' and 3' edges labeled](image.png)
Hydrogen Bonding Pattern

The two strands of the DNA are connected by hydrogen bonds. These forces are non-covalent interactions but are relatively strong and keep the two DNA strands in a tight association. Adenine and Thymine interact with two hydrogen bonds. Cytosine and Guanine interact with three hydrogen bonds and therefore have a stronger association.
Aromatic Base Stacking

Another important interaction that stabilizes the DNA structure is termed stacking and refers to the interaction between adjacent parallel aromatic bases. This interaction results from sharing the electron clouds of the aromatic rings. The interaction is slightly stronger in purines, which possess a double ring.
Major and Minor Grooves

The DNA helix contains grooves which are preferred sites for proteins and other molecules to interact with the DNA. One significant groove in which the bases are highly exposed is termed the major groove. Most of the protein interactions occur at this site. The minor groove is located on the other side of the bases. It also serves for interactions with proteins, but less often.
Drugs Binding to DNA Helix

Molecules of some drugs bind to minor or major grooves or can be intercalated between base pairs.

Figure:
Double-helical DNA interacting with drug molecules.
(a) and (b) shows the fit of flat molecule (Hoechst 33258) into the minor groove.
(c) drug daunomycin intercalated between base pairs of DNA.
(d) cis-platinum covalently bound to guanine bases.
DNA forms

There are several forms of DNA double helices. The most common is the B-DNA. In every B-DNA cycle there are 10 base pairs; the distance between successive bases along the molecule axis is about 3.4 Å. In dehydrated environments, the DNA may appear as A-DNA. In every A-DNA cycle there are 11 base pairs so the molecule is more condensed, and accordingly the diameter of A-DNA is greater than that of B-DNA. Another uncommon form of DNA is the Z-DNA. The helix has an opposite rotation, composed of C and G nucleotides and is less symmetrical.

Figure:
A-, B- and Z-DNA double helices in the canonical conformations. All helices on the picture are formed by 22 base pairs.
DNA Packaging

Because of its length, the DNA molecules must be carefully packed and preserved. Many proteins in the cell are involved in these processes (histones).
G-Quadruplex Conformation

In addition to the duplex DNA conformations, certain DNA sequences can fold into structures that are four-stranded. In particular Guanine-rich nucleic acid sequences are capable of adopting this type of organization, which is called G-quadruplex. The G-quadruplex structure is stabilized by hydrogen bonds between the edges of the bases and chelation with a metal (e.g. potassium) located in the center. Below the 3D and 2D structure of a G-quadruplex is illustrated.
DNA versus RNA

There are several important differences between DNA and RNA:
1) The sugar of RNA is always a ribose instead of a deoxyribose (with an extra hydroxyl group at carbon number 2 of the sugar)
2) the RNA molecule does not contain thymine bases; instead they are replaced by uracil
3) the DNA molecule normally appears as a double strand, while the RNA molecule appears as a single strand.
3D Folds of RNA

Due to its single strand structure, the RNA molecule is less stable than DNA. Base pairing might occur between bases along the single strand giving rise to various three dimensional folds.

Figure on the left: The structure on tRNA molecule

Figure on the right: The structure of hairpin fold. Intra-molecular hydrogen bond interactions are formed between bases of the same single strand.
Conformations of DNA and RNA

The torsion angles for the sugar phosphate backbone are given as Greek letters $\alpha, \beta, \gamma, \delta, \varepsilon, \zeta$ and $\chi$. Graph of dependencies of these angles can be constructed.

Figure on the left: torsion angles in DNA nucleotide pair
Figure on the right: The backbone conformations between $\zeta$ and $\alpha$ for nucleotides of DNA (dark blue), DNA complexes with proteins or drugs (yellow) and RNA (red). Different helical forms are denoted by A, AII, BI, BII, ZI, ZII.
Conformations of the Sugar Ring

The five member deoxyribose sugar ring is nonplanar. This nonplanarity is termed **puckering** and specific ring conformations are called **puckers**. Because of steric and energetic reasons some puckers are preferred over the others.

Although there is a continuum of configurations for the sugar ring, the planar geometry is the least favorable.
**Configurations of Base Pair Rings**

The individual bases in a nucleic acid are flat aromatic rings, but base pairs bound together only by non-rigid hydrogen bonds can show considerable flexibility. Geometric relations between bases are described by number of rotational and translational parameters.