Introduction to Nucleic Acids:
Structural Properties of Nucleic Acid Building Blocks

Function of DNA and RNA

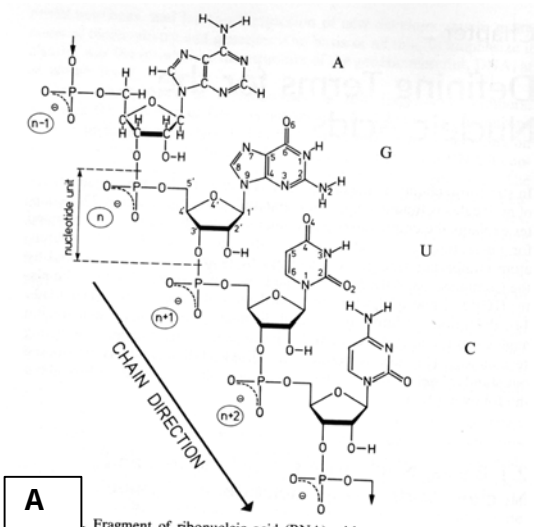
DNA and RNA are chainlike macromolecules that function in the storage and transfer of genetic information. They are major components of all cells ~15% of the cells dry weight. Just as the amino acids are building blocks of proteins, the nucleotides are the monomeric unit of nucleic acids.

In this section of the course, we will first examine the structure and chemistry of the nucleotides and polynucleotides. Subsequently, we will examine the non-covalent forces that form help nucleic acid helices and the types of helices that are possible. We will then move into an examination of higher order nucleic acid structure and its interactions with proteins

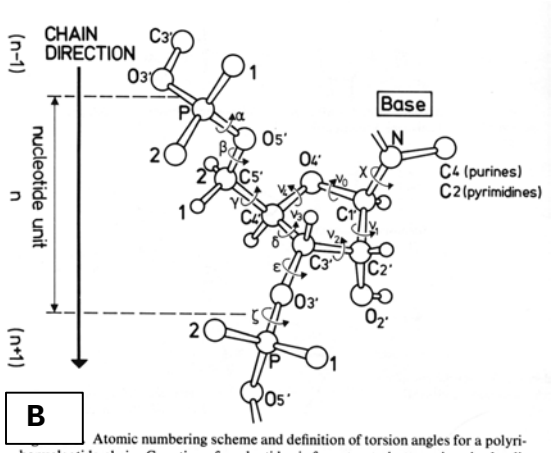
The components of Nucleic acids-Sugars and Bases.
The primary structure of a polynucleotide has some analogy with the primary structure of proteins. Like proteins nucleic acid polymers are chains of monomers, in the nucleic acids these monomer units are called nucleotides. In both RNA and DNA, each nucleotide repeating unit consists of three characteristic components:

a) a nitrogeneous heterocyclic base, which is either the derivative of a pyrimidine or a purine
b) a pentose sugar (the type differs between RNA and DNA)
c) a molecule of phosphoric acid.

Polynucleotides are polymers of this basic unit **(FIGURE 1)**. This figure illustrates the massive number of adjustable angles in the backbone and highlights the conformational variablity in the nucleotide chain. We will concern ourselves in this section of the course with the structural bases and physical forces that determine the conformation of polynucleotides. We will then examine how these factors affect the biological function of polynucleotides.



Fragment of ribonucleic acid (RNA) with sequence adenosine (A), guanosine (G), uridine (U), cytidine (C) linked by 3′,5′-phosphodiester bonds. Chain direction is from 5′- to 3′-end as shown by arrow. Atom numbering scheme is indicated in one framed nucleotide unit, 5′-GMP. All hydrogen atoms drawn in A and only functional hydrogens in other nucleotides. In short notation, this fragment would be pApGpUpCp or pAGUCp. In deoxyribonucleic acid (DNA), the hydroxyl attached to $C_{2'}$ is replaced by hydrogen and uracil, by thymine.



Atomic numbering scheme and definition of torsion angles for a polyribonucleotide chain. Counting of nucleotides is from top to bottom, i.e., in the direction $O_{3'} \rightarrow O_{3'}$. Hydrogens at $C_{3'}$ and oxygens at P are differentiated by 1 and 2 according to the rule given in the text. In deoxyribose, the hydrogen replacing $O_{2'}$ is labeled 1, the other one, 2. For a full description of torsion angles, see Table 2-2.

FIGURE 1

Nucleic Acid Bases

The common nucleotides come in 5 different flavors. Adenine, guanine, cytosine and thymine in DNA, with uracil substituting for thymine in RNA. In strict analogy with the amino acids, nucleotide bases have different functional groups and these differences determine the polynucleotide's structure and function **(FIGURE 2)**. Note different shapes, different functional groups. Purines-two ring system, pyrimidines single rings. Both pu and py have prononced aromatic character. Hence the conformation of the bases is fixed! You are responsible for the structure of the bases!!! The precise 3D structure of the bases is known with high precision. The pyrimidines are planar molecules; purines are nearly planar with a slight pucker. The angles between the atoms in the rings of the 5 bases are very close to 180^{o} and do not change. The important differences to note in the functional groups of the bases are their capacities to donate or accept hydrogen bonds and their ability or lack thereof to have hydrophobic interactions.

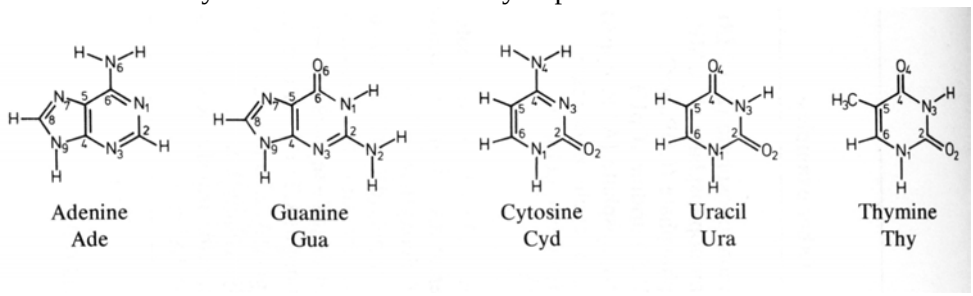


Figure 2

These common bases can and often are derivatized by the cell for various structural and functional uses. The most common derivatives are those which are methylated **(FIGURE 3)**. A list of some other types of modified bases is given in the **TABLE 1**. Most of these bizarre bases are found in tRNA. We will discuss them later.

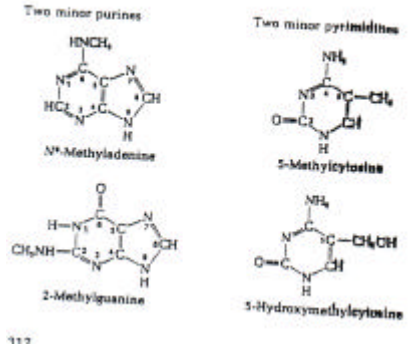


Figure 3

Table 1 Other rare bases found in nucleic acids

5,6-Dihydrouracil
1-Methyluracil
3-Methyluracil
5-Hydroxymethyluracil
2-Thiouracil
N⁴-Acetylcytosine
3-Methylcytosine
5-Methylcytosine
5-Hydroxymethylcytosine
1-Methyladenine
2-Methyladenine
7-Methyladenine
N⁶-Methyladenine
N⁶,N⁶-Dimethyladenine
N⁶-(Δ²-Isopentenyl)adenine
1-Methylguanine
7-Methylguanine
N²-Methylguanine
N²,N²-Dimethylguanine

Table 1

One useful property of the nucleic acid bases is their strong UV absorption in the range from 250-280nm (Figure 4). Having determined the extinction coefficient(s) [e] of the bases, this property allows one to determine the concentration of the nucleic acid in solution.

$$OD = c\, e\, l$$

The knowledge of the ratio of the average extinction coefficients of the bases at two UV wavelengths, commonly 260 vs. 280 allows one to assess the purity of the nucleic acid solution. Additionally, the OD of a particular concentration of nucleic acid bases depends on the structure into which they are assembled. Hence absorbance can be used as a structural probe of nucleic acids (see below).

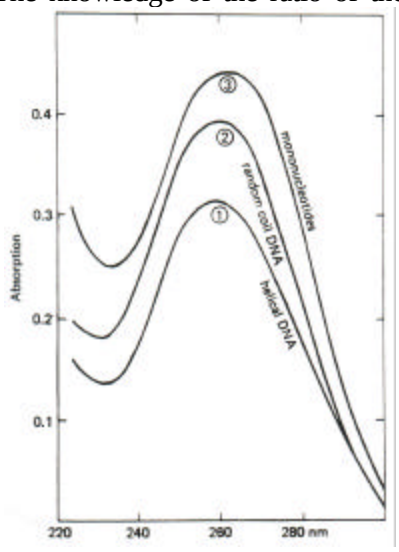
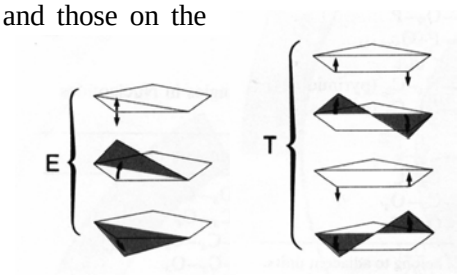


Figure 4

Sugars.

Both RNA and DNA nucleotides contain a cylic furanose-type sugar; ß-D-ribose in RNA and ß-D-deoxyribose in DNA. The difference in chemical structure of the sugar occurs at the C2' position. In DNA the -OH group is replaced by an H. The five-membered sugar ring is generally non-planar. It can be puckered in an <u>envelope</u> (**E**) form-four of the five atoms lie in a plane and the fifth atom lies out of plane by ~0.5 angstrom; or in a <u>twist</u> (**T**) conformation where two adjacent atoms are displaced on opposite sides of a plane through the three other atoms **(Figure 5).** Conformations in which the atoms are displaced from these three or four atom planes are on the same side as the C5' are called *endo* and those on the opposite side are called *exo*.



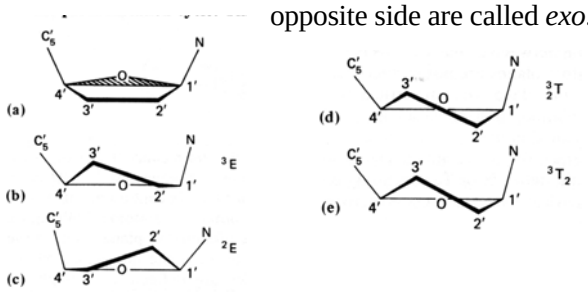


Figure 5

In an unsubstituted furanose ring, the conformational changes do not proceed via planar intermediate, but the maximum pucker rotates virtually without potential energy barriers, giving rise to a potentially infinite number of conformations. This is illustrated by the <u>pseudorotation cycle</u> **(FIGURE 6)**. This point will become important as we learn about the extraordinary flexibility of the DNA backbone.

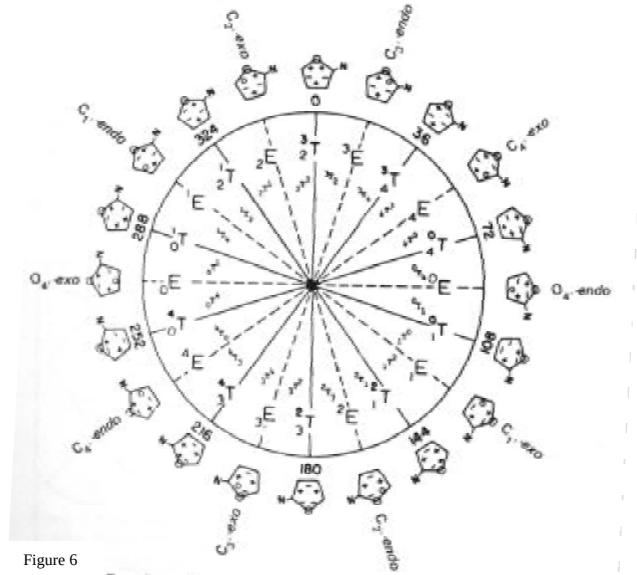


Figure 6

Pseudorotation cycle of the furanose ring in nucleosides. Values of phase angles given in multiples of 36°. Envelope E and twist T forms alternate every 18°. After rotation by 180° the mirror image of the starting position is found. On the periphery of the cycle, riboses with signs of endocyclic torsion angles are indicated. (+) Positive, (−) Negative, (0) angle at 0°. From (31).

Since the furanose ring in polynucleotides is unsymmetrically substituted, they do display a preferred puckering modes conformations. The pseudorotation angles determined from crystallographic investigations are <u>not</u> distributed evenly over the pseudorotation cycle. Instead they cluster in two domains, centered at C3'-*endo* (found in A-type helices) and C2'-*endo* (found in B-type helices). These two puckering modes are favored because they best alleviate steric clashes of the substituents of the sugar by placing them in a staggered, not eclipsed, conformation (Figure 7 & 8).

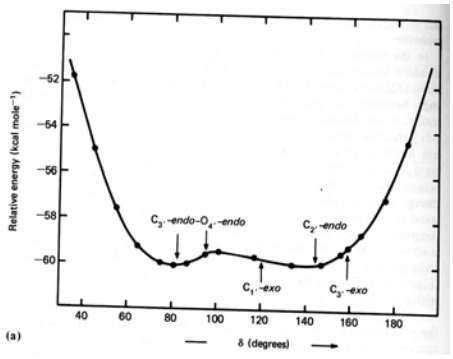


Figure 7

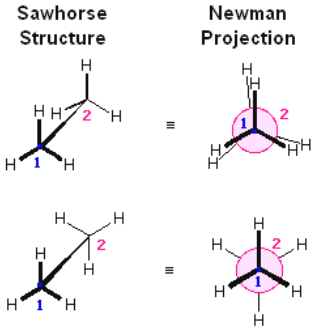
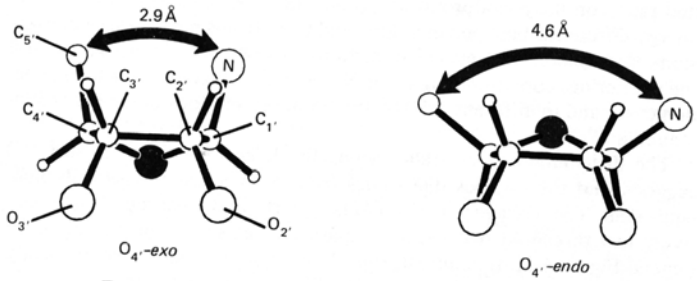


Figure 8



Equatorial and axial positioning of substituents N (base) and $C_{5'}$ in sugar puckering modes $O_{4'}$-*exo* (left) and $O_{4'}$-*endo* (right). In $O_{4'}$-*exo*, $C_{5'}$ and N interfere sterically and therefore the $C_{2'}$-*endo* $\rightleftharpoons$ $C_{3'}$-*endo* interchange occurs via $O_{4'}$-*endo* as intermediate.

Figure 9

The interconversion between these two forms can occur by two paths, via an O4'-endo or O4'-exo intermediate. The route taken by most interconversions in via the O4'-endo intermediate. Consideration of why this is true illustrates the kinds of constraints on the backbone of a polynucleotide **(Figure 9)**.

The O4'-exo intermediate of a substituted furanose ring causes the C5' and N-base to clash, whereas the O4'-endo allows adequate spacing between them. This serves as a paradigm for thinking about restrictions on the mobility of the backbone polynucleotides. THAT IS: even in single stranded polynucleotides there are restrictions on chain conformations!!!

So now we have talked about the bases and the sugar, now let's put them together to form a nucleoside and nucleotide.

Define: Nucleoside- base + sugar

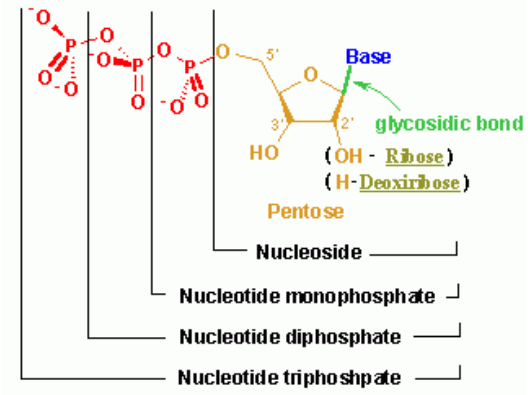


Figure 10

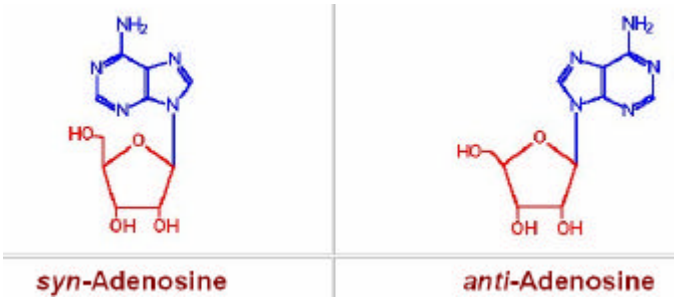


Figure 11

Nucleotide- base + sugar, the sugar is phosphorylated at one of the free sugar hydroxyls. The most common nucleotides are those which are phosphorylated at the C5', thus, *5' nucleotide monophosphate* **(Figure 10)**.

The formation of nucleotides adds another angle that we must be concerned about. This angle is that which the base pivots in relation to the nucleotide. This angle χ (*chi*), defines the glycosyl bond. Relative to the sugar moiety, the base can adopt two main orientations about the C1'-N link, called either *syn* or *anti*. In the *anti* configuration, the bulky part of the base is pointing away from the sugar. The bulky part of the purine base is obvious; the six membered ring is pointed away from the sugar. For the pyrimidines the O2 on the C2 carbon is the bulky group; in *anti*, this group points away from the base. The definitions of *syn* and *anti* are somewhat non-specific and range as shown in **FIGURE 11)**.

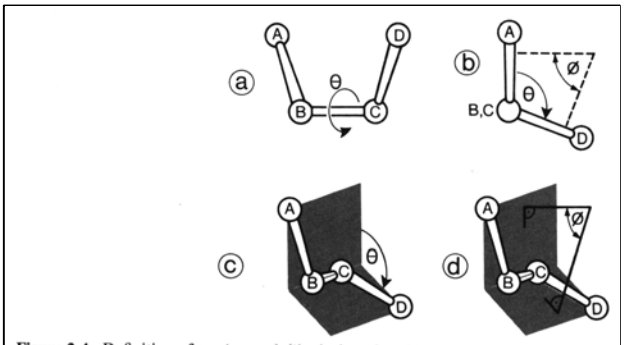


Figure 2-4. Definition of *torsion* and *dihedral angles*. (a) *Torsion angle* θ (A–B–C–D) describing orientations of bonds A–B and C–D with respect to the central bond B–C. (b) View along B→C. θ is the torsion angle between the projected bonds A–B and C–D; the complement ϕ is called the *dihedral angle*. If A–B and C–D are *cis*-planar (coinciding in projection), angles θ and ϕ are 0°; they are counted positive if the far bond C–D rotates clockwise with respect to the near bond A–B. (c) θ is defined as the angle between planes A–B–C and B–C–D. (d) The *dihedral angle* ϕ represents the angle between normals to these planes.

Figure 12

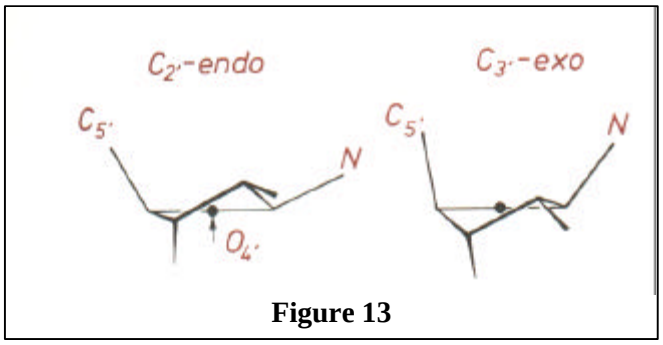


Figure 13

Figure 12 gives a definition of torsion angles. The gross value of the χ angle is correlated with the major observed sugar pucker. That is the steric hinderance in the *syn* conformation can be somewhat alleviated by placing the sugar in the C2'*endo* pucker. This is because the base and the C5' are then placed in an equatorial position and moved apart **(FIGURE 13)**. Another extremely relevant exception to the suggestion that *anti* configurations are preferred by the nucleotides is that of guanine. Both as a single nucleotide and in the alternating co-polymer G-C-G-C, G is in the *syn* configuration. This is probably due to the H-bonding of the N2 amino group of the guanine to its 5'phosphate.

Polynucleotides.

The condensation reaction between multiple 5' mononucleotides at the 3'-OH group forms a chain. In analogy with protein primary structure, 5'-3' chains have direction, a **polarity**, they have distinct and different ends. The features of the backbone repeat in sequence 5'-P, C5', C4', attached to sugar and base, C3', O3' and so on **(FIGURE 1)**. As you might expect from the rest of this lecture, this linkage of nucleotides adds more angles to be concerned about. We will not discuss them in detail now, but will cover them later in double helical structure of DNA.