

## Structure of the Nucleic Acids

There are two types of nucleic acids, **DNA**, deoxyribonucleic acids and **RNA**, ribonucleic acids.

The genetic information is usually encoded in **DNA** (RNA in many viruses). DNA is the repository of all the information needed to construct a biological organism. Ultimately all the development and functioning of an organism can be traced back to the information stored in the DNA. During cell division DNA directs its own replication by acting as a template for a new DNA strand.

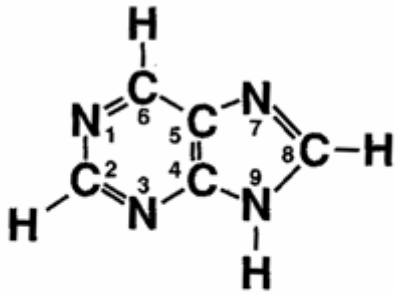
The base sequences in **RNA** molecules is copied from the DNA code and carries out other functions such as acting as a messenger or adapter between the “code” and the synthesis of proteins. RNA can also act as an enzyme and has been proposed to be the original “Urgene”.

In order to understand how DNA and RNA can carry out their functions, let's look at their structures.

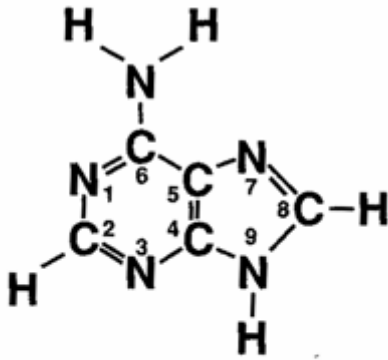
Nucleic acids are composed of aromatic bases (a purine or a pyrimidine ring), ribose sugars and phosphate groups.

## ***Bases***

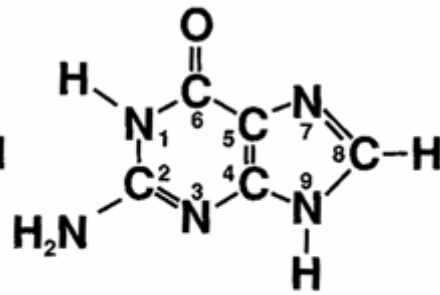
Purines: adenine (A) and guanine (G)



**Purine**

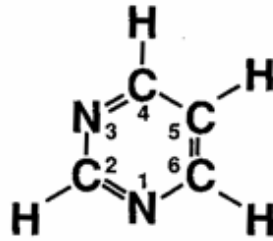


**Adenine  
(A)**

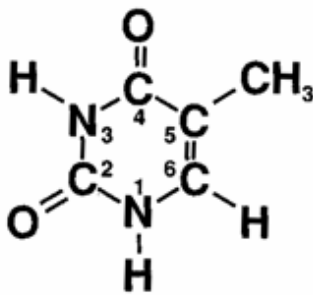


**Guanine  
(G)**

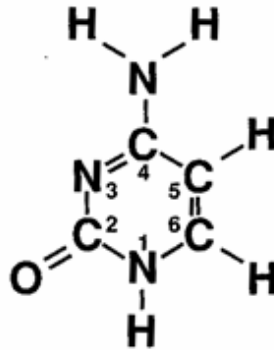
Pyrimidines: thymine (T), cytosine (C) and uracil (U –  
in RNA)



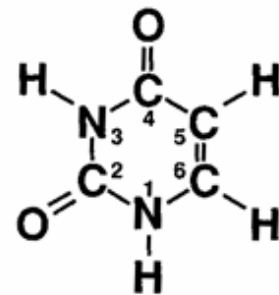
**Pyrimidine**



**Thymine  
(T)**



**Cytosine  
(C)**



**Uracil  
(U)**

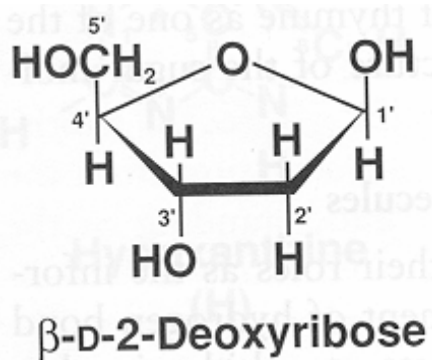
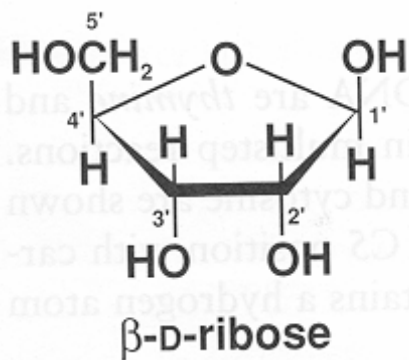
> differential placement of hydrogen bond donor and acceptor groups in the different bases makes it possible for the bases to serve as the genetic information

> planar molecules: important for organization of bases within the helix, for the stacking

## ***Sugars***

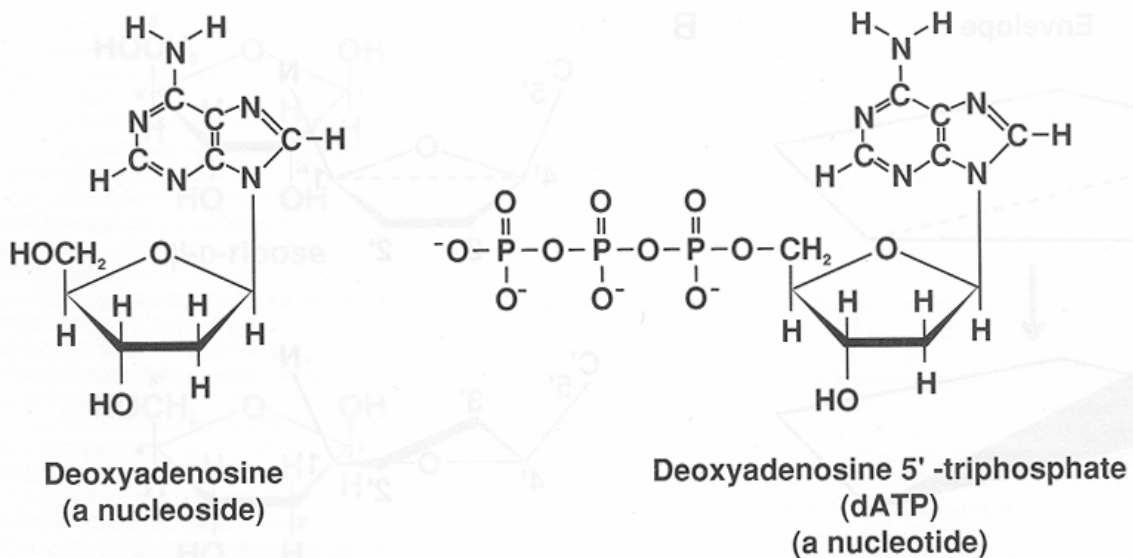
The sugars of DNA and RNA differ by the presence or absence of a hydroxyl group at the 2'-position of the ribose sugar:  $\beta$ -D-ribose in RNA  $\beta$ -D-2-deoxyribose in DNA

> flexible and dynamic parts of DNA or RNA



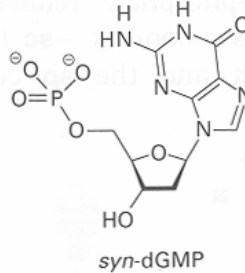
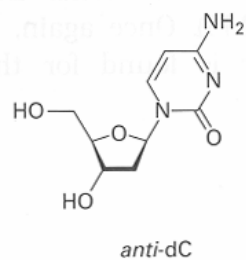
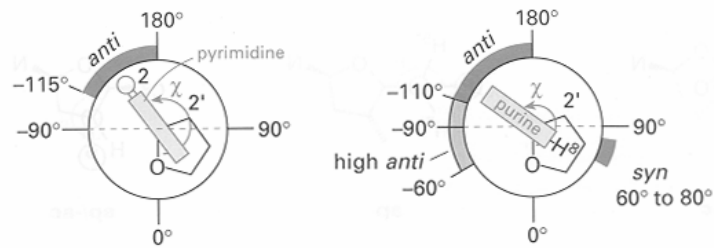
## ***Nucleosides and nucleotides***

The bases are attached to the sugars, forming a **nucleoside** (adenosine, guanosine, thymidine, cytidine, uridine). Phosphate groups are then attached to the sugar at the 5' carbon, forming a **nucleotide**. One, two or three **phosphate groups**, designated  $\alpha$ ,  $\beta$  and  $\gamma$  for the first, second and third respectively (right to left in figure below), can be attached, e.g. ATP = adenosine 5'-triphosphate

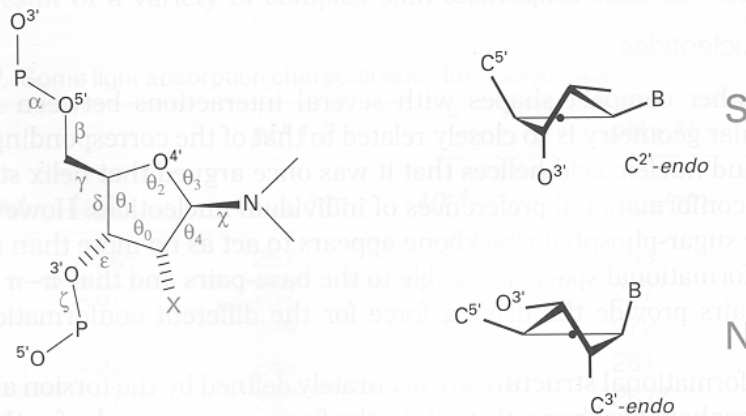


- The bond between the base and the sugar is called a **glycosidic bond**.
- The glycosidic bond is up with respect to the sugar ring ( **$\beta$ -conformation**) rather than down ( **$\alpha$ -conformation**).

- The plane of the base is almost perpendicular to the plane of the sugars and almost bisects the  $O^{4'}-C^{1'}-C^{2'}$  angle.
- The base is free to rotate around the glycosidic bond (torsion angle  $\chi$ ). Two main orientations are adopted, called ***syn*** and ***anti*** (see top figure page 4). In the *anti* conformation the 6-membered ring of the purines and the oxygens of the pyrimidines point away from the sugar group, instead the smaller H6 (pyrimidines) or H8 (purines) lie above the sugar ring. The *syn* conformer has the larger O2 (pyrimidines) or N3 (purines) above the sugar ring. Pyrimidines adopt a narrow range of *anti* conformations whereas purines have a wider range of *anti* conformations. The *anti* conformation is always preferred except guanine prefers the *syn* conformation in alternating oligomers e.g. d(CpGpCpG), in Z-DNA and in mononucleotides. The *syn* conformation can only be built into left-handed helices.
- The sugar furanose rings are twisted out of plane to minimize non-bonded interactions between their substituents. This is called **puckering** and is a displacement of the C2' and C3' from the median plane of C1'-O4'-C4' (see bottom figure on page 4). The C2'-endo pucker is preferred in standard B-DNA, the C3'-endo pucker occurs in A-RNA. When a purine base adopts the *syn* conformation with respect to the sugar, as in Z-DNA, the sugar adopts the C3'-endo conformation to avoid steric clashes.



**Fig. 2.12** *Anti* and *syn* conformational ranges for glycosidic bonds in pyrimidine (left) and purine (right) nucleosides and drawings of the *anti* conformation for deoxycytidine (lower left) and the *syn* conformation for deoxyguanosine 5'-phosphate (lower right).

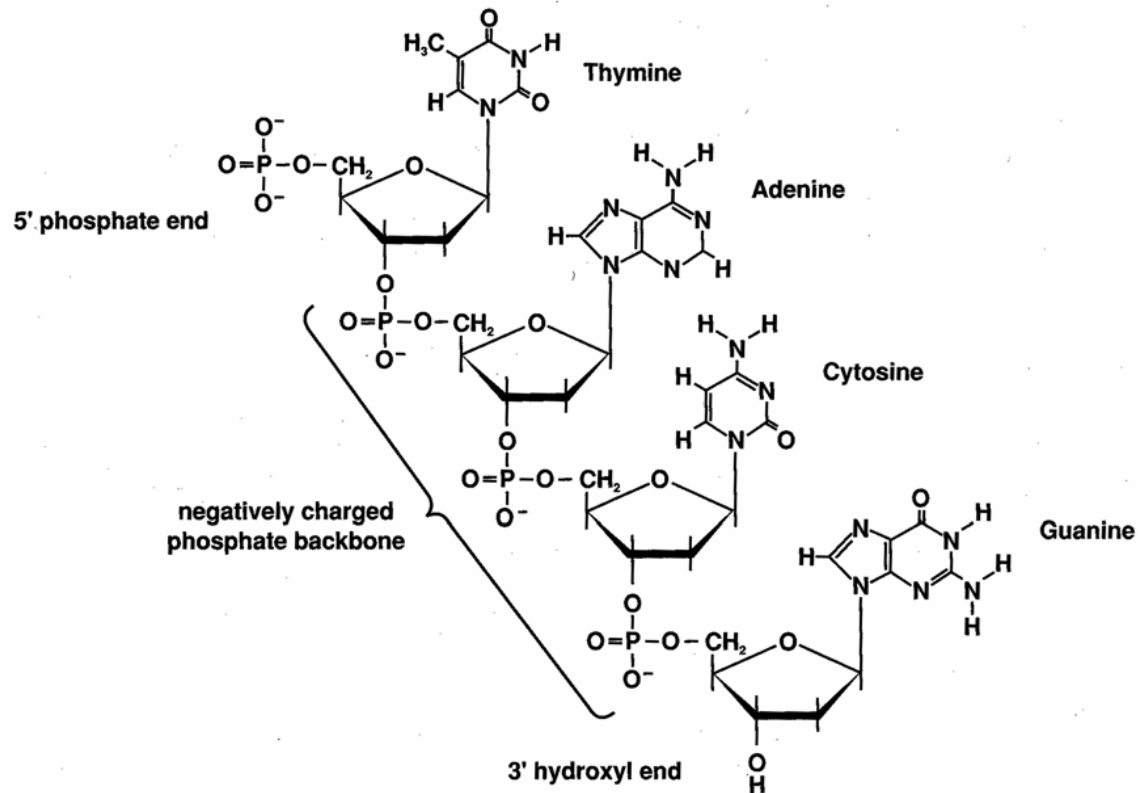


**Fig. 2.11** Torsion angle notation (IUPAC) for polynucleotide chains and structures for the  $C^{2'}$ -*endo* (**S**) and  $C^{3'}$ -*endo* (**N**) preferred sugar pucker.

## ***Phosphodiester bond***

In DNA and RNA the individual nucleotides are joined by **5'-3' phosphodiester bonds**. These are formed during their biochemical synthesis, by the enzyme DNA polymerase for DNA (in the process called **replication**) and by the enzyme RNA polymerase for RNA (in the process of **transcription**). Below is shown a single strand oligonucleotide with sequence TACG. To extend the polynucleotide an incoming deoxyribonucleotide 5'-triphosphate forms a phosphodiester bond to 3'OH position of the polynucleotide chain. The  $\beta$ - $\gamma$  pyrophosphate group is split off and hydrolyzed into individual phosphate molecules, making the reaction thermodynamically favorable.





**Figure 1.6 A single strand of DNA.** This chain of DNA is composed of the bases thymine, adenine, cytosine, and guanine. Note that the nucleotides are linked through the phosphate groups connected between the 5' carbon and the 3' carbon of adjacent deoxyribose sugar molecules. The chain of DNA has a negatively charged phosphate backbone. The DNA chain has two chemically distinct ends. At the 5' end (*top*) a phosphate group is attached to the 5' carbon of the deoxyribose sugar. The 3' end (*bottom*) has a hydroxyl group on the 3' carbon. This sequence of DNA would be written TACG since sequences are written in the 5' → 3' direction by convention.

The polynucleotide has two chemically distinct ends, a 5' and a 3' end, giving the DNA **polarity**. The sequence of a DNA or RNA molecule is always given from 5' to 3'.

Phosphate diesters are **tetrahedral** at the phosphorus (top figure page 6) and show antiperiplanar conformations for the C<sup>5'</sup>-O<sup>5'</sup> bond. Similarly the C<sup>3'</sup>-O<sup>3'</sup> bond lies in the antiperiplanar to

antichiral sector. This has led to the concept of **virtual bonds**: the chains  $P^{5'}-O^{5'}-C^{5'}-C^{4'}$  and  $P^{3'}-O^{3'}-C^{2'}-C^{4'}$  can be analyzed as rigid, planar units linked at the phosphorus and at  $C^{4'}$ .

The skewed conformation of the C-O-P-O-C system observed in X-ray structures of dinucleotides and polynucleotides has been explained by a **gauche-effect** (bottom figure page 6): there are favorable interactions between a non-bonding electron pair on  $O^{5'}$  ( $O^{3'}$ ) with the P-O<sup>3'</sup> (P-O<sup>5'</sup>) bond, making this conformation  $\sim 30$  kJ/mol more favorable than the extended W-conformation.

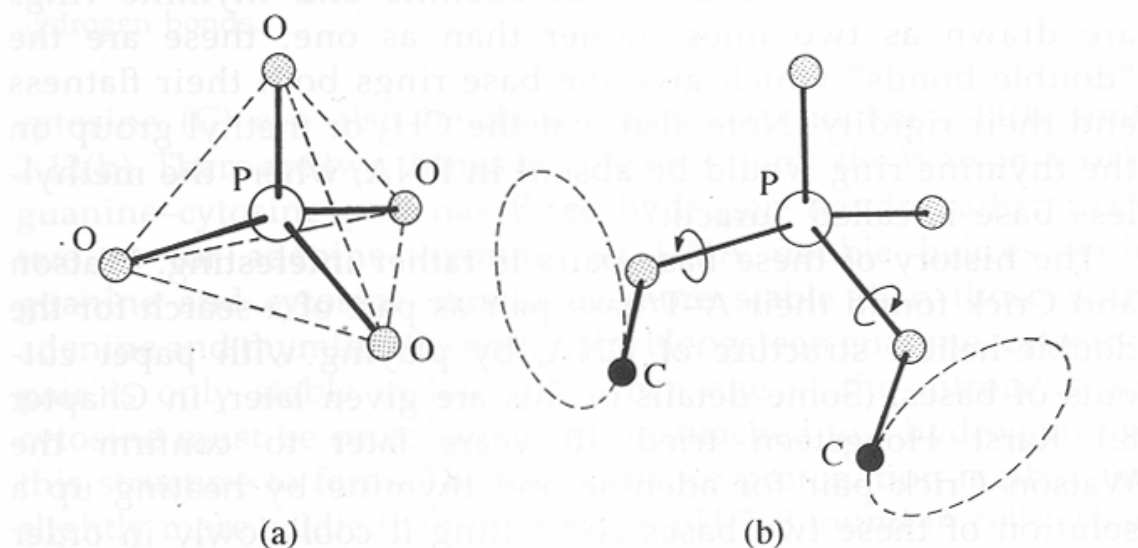
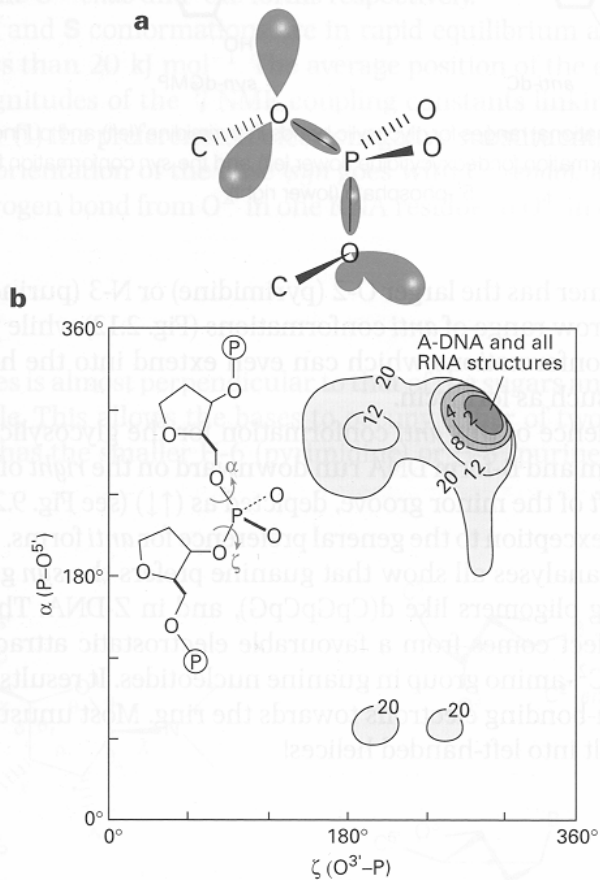


Figure 2.10 (a) Tetrahedral arrangement of phosphate group: atom types are phosphorus (P) and oxygen (O); (b) freedom of rotation for adjacent links of the chain, with atom type carbon (C).



**Fig. 2.14** (Upper) Gauche conformation for phosphate diesters showing the antiperiplanar alignment of an occupied non-bonding oxygen orbital with the adjacent P–O bond. (Lower) Contour map for P–O bond rotations calculated for diribose triphosphate (energies in  $\text{kJ mol}^{-1}$ ) (adapted from Govil, G. (1976). *Biopolymers*, **15**, 2303–7. Copyright (1976) John Wiley and Sons, Inc.).

## ***The double helix***

The phosphate and sugars are very soluble in water, but the bases are not. Their insolubility does place strong constraints on the overall conformation of a large DNA or RNA molecule in solution. To be stable in water at neutral pH, the bases have to tuck themselves into the very center of a folded structure so as to avoid the water, while the sugars and phosphates will have to be on the outside. This is exactly what happens. DNA forms a spiral or helix in solution. There is a wide variety of double-helical DNA conformations, and this structural polymorphism is denoted by the use of different letters (A, B, Z, C, D,...). The particular conformation adopted by a DNA or RNA molecule depends on factors as the humidity, the cations present, the base sequence.

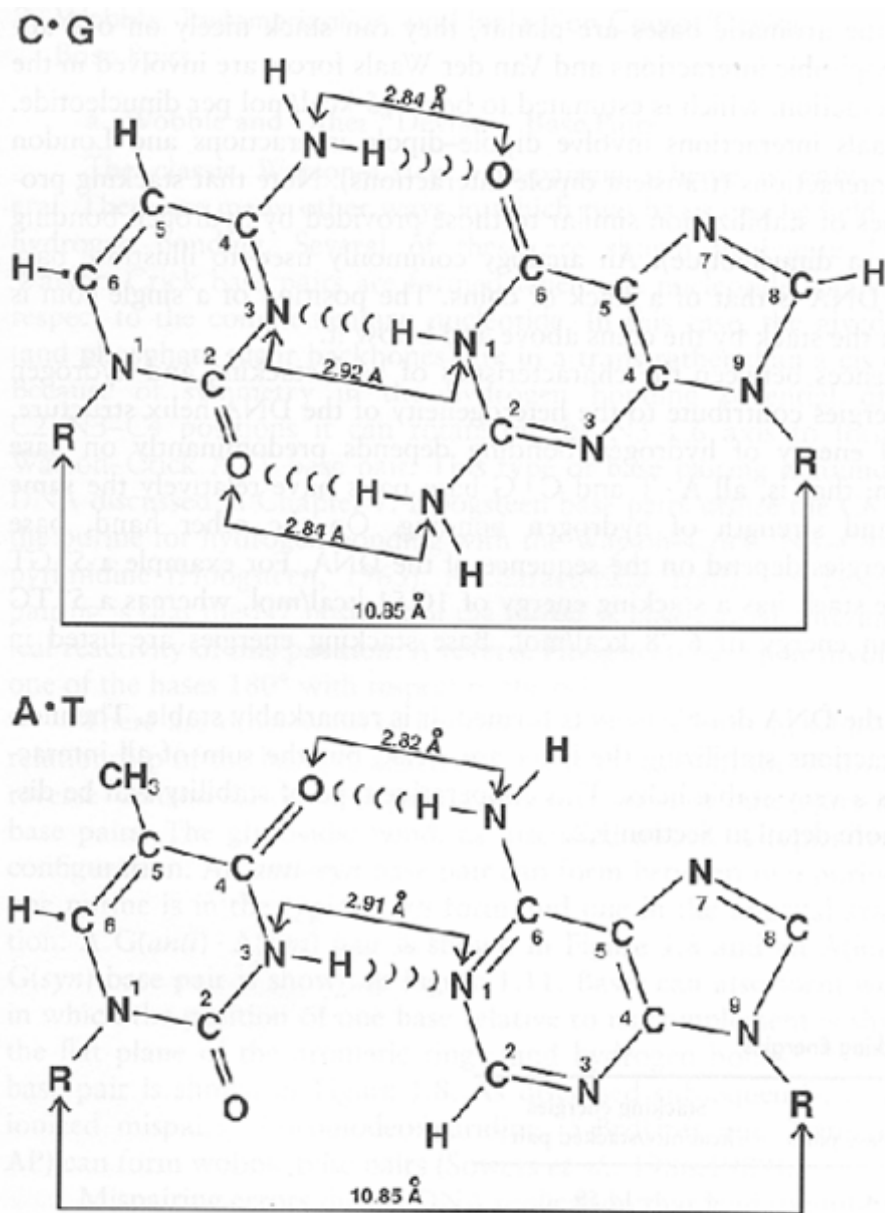
The “classical” Watson and Crick structure of **B-DNA** has the following major features:

- two polynucleotide strands wind about a common axis with a *right-handed* twist to form a  $\sim 20\text{\AA}$  diameter double helix. The strands are *antiparallel* (=run in opposite directions: one strand is oriented in the  $5' \rightarrow 3'$  direction and the other in the  $3' \rightarrow 5'$  direction. The bases occupy the core of the helix and the sugar-phosphate chains are coiled about its periphery.
- The bases are stacked predominantly above their neighbors in the same strand and are perpendicular to the helix axis. Each base is *hydrogen bonded* to

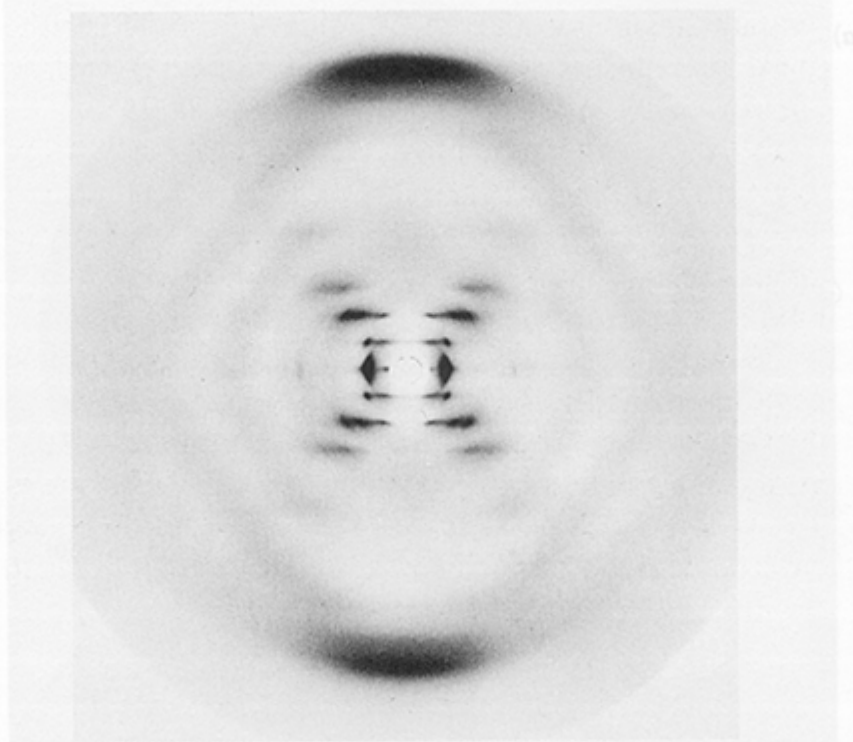
- a base on the opposite strand to form a *base pair*. These hydrogen bonding interactions result in the *specific* association of two chains of the double helix: each adenine residue pairs with a thymine residue and each guanine residue pairs with a cytosine residue and vice versa. The geometry of these *Watson-Crick base pairs* is shown on page 8.
- The two sugars to which a base-pair are attached lie closer to one side of the base-pair than the other, the edge which lies closer to an imaginary line drawn between the two sugars is called the minor-groove side, while the other edge is called the major groove side. (to remember: going along a strand from 5'→3' the major groove is ahead of you). The grooves have distinct sizes due to the structurally distinct base pair edges and the asymmetric deoxyribose residues. The two grooves provide very distinct surfaces with which proteins can interact. In B-form DNA, the base pairs sit directly on the helix axis so that the minor and major grooves are of similar depths.
  - The sugars have the  $C^{2'}$ -*endo* pucker and all the glycosides have the *anti* conformation.
  - The *ideal* B-DNA double helix has *10 base pairs per turn* (a helical twist of  $36^\circ$  per base pair), and a *pitch* (rise per turn) of  $34 \text{ \AA}$ .

Critical information used by Watson and Crick were **Chargaff's rule** and **X-ray diffraction patterns of DNA fibers**. Analyzing DNA hydrolysates Chargaff found that the amount of adenine always equaled the

amount of thymine and the amount of guanine always equaled the amount of cytosine. This led to the Watson-and-Crick **complementary base pairing scheme**, which, as was pointed out by Watson and Crick in their famous paper, can explain how DNA can direct its own replication and act as the carrier of the genetic information. The X-ray diffraction patterns (see figure later) indicated that the geometric shape of DNA is a right-handed helix.



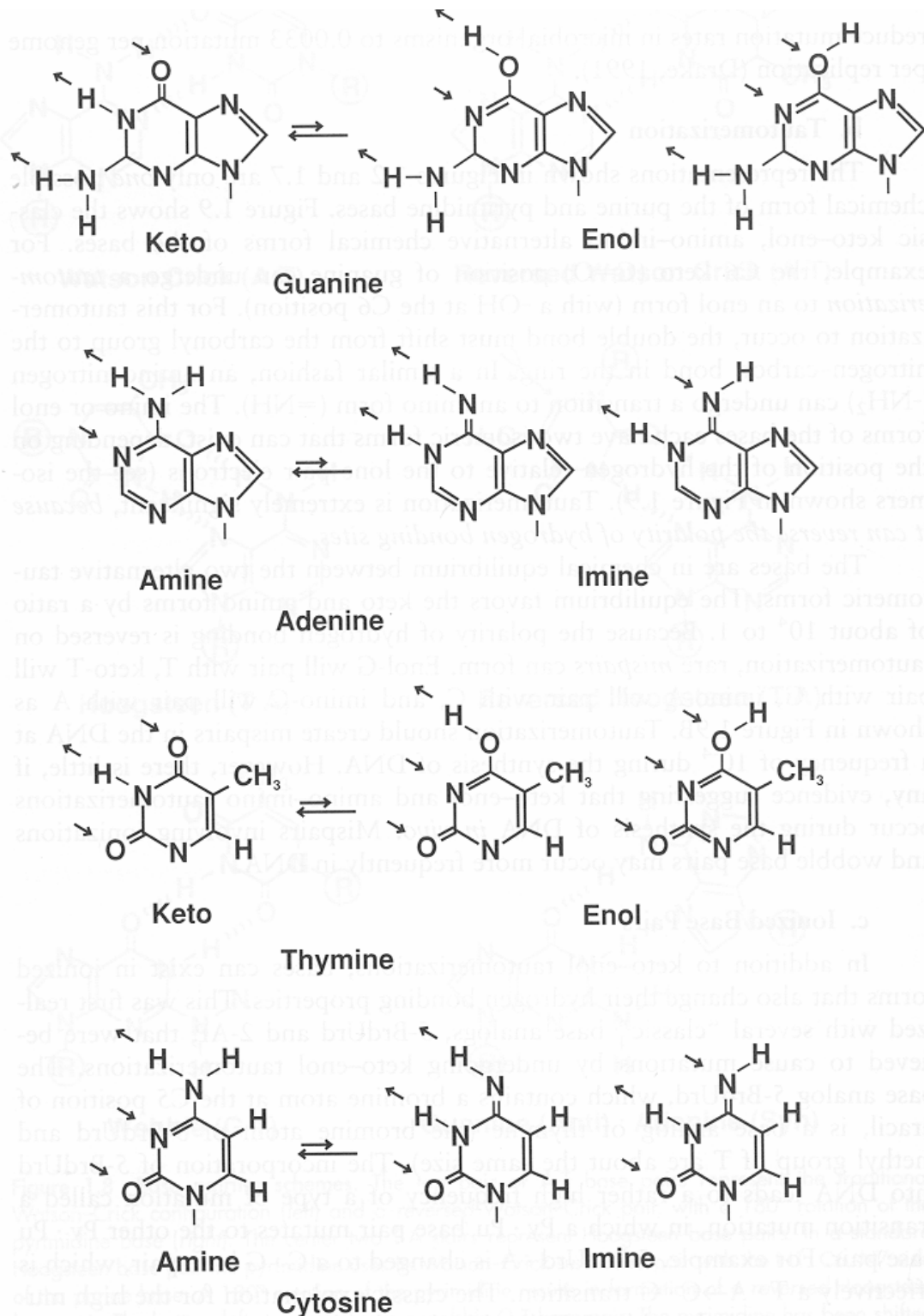
**Figure 1.7** Watson-Crick base pairs. The interatom hydrogen bond distances and distances between the C1' positions of the ribose sugars are indicated. The curved lines represent the hydrogen bonds. The curves are in the direction of the hydrogen bond acceptor (N or O atoms). Figure modified with permission from Arnott *et al.* (1965).



**FIGURE 28-4.** An X-ray diffraction photograph of a vertically oriented Na<sup>+</sup> DNA fiber in the B conformation taken by Rosalind Franklin. This is the photograph that provided key information for the elucidation of the Watson–Crick structure. The central X-shaped pattern of spots is indicative of a helix, whereas the heavy black arcs on the top and bottom of the diffraction pattern correspond to a distance of 3.4 Å and indicate that the DNA structure largely repeats every 3.4 Å along the fiber axis. [Courtesy of Maurice Wilkins, King's College, London.]

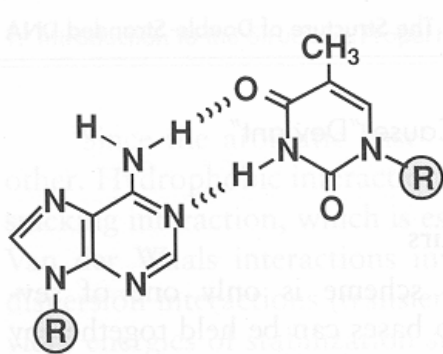
The elucidation of the B-DNA structure was not trivial. The bases can adopt different **tautomeric forms**, reversing the polarity of the hydrogen bonding (figure below). In addition to keto-enol tautomerizations, bases can exist in ionized forms that also change their hydrogen bonding properties.



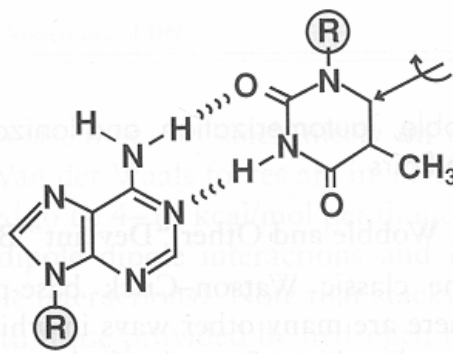


**Figure 1.9 Tautomerization of the bases.** (A) Base tautomerization. The bases can exist in different chemical isomeric forms. The top panel shows guanine which normally exists in the keto configuration (right) in which there is a carbonyl group at a C6 position of the ring. The hydrogen bonding properties of the bases are indicated by the arrows. The arrows pointing away from the hydrogens represent hydrogen bonding donors. The arrows pointing toward the negative centers

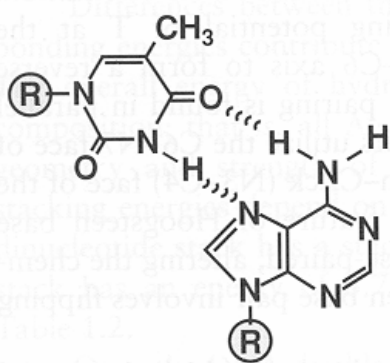
The classic Watson-Crick base pairing scheme is only one of several. There are many other ways in which two bases can be held together by hydrogen bonding.



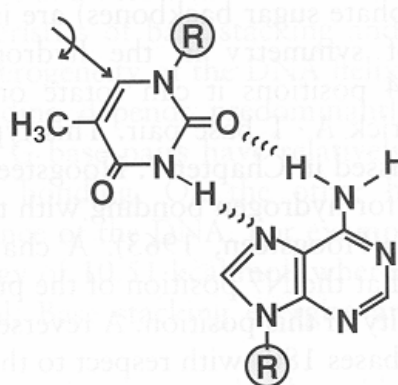
Watson Crick (A·T)



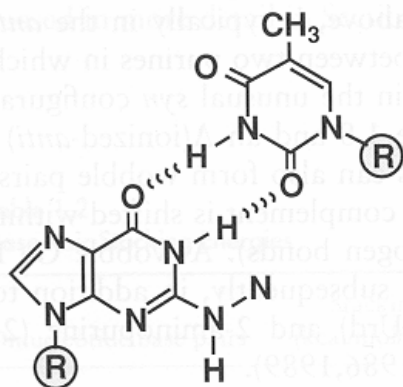
Reversed Watson Crick (A·T)



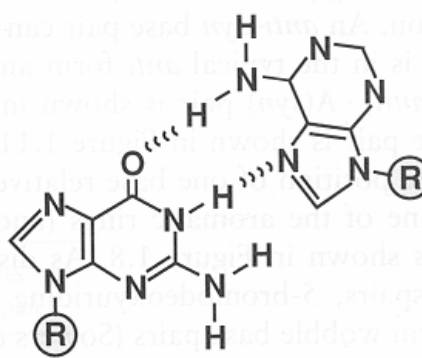
Hoogsteen (T·A)



Reversed Hoogsteen (T·A)



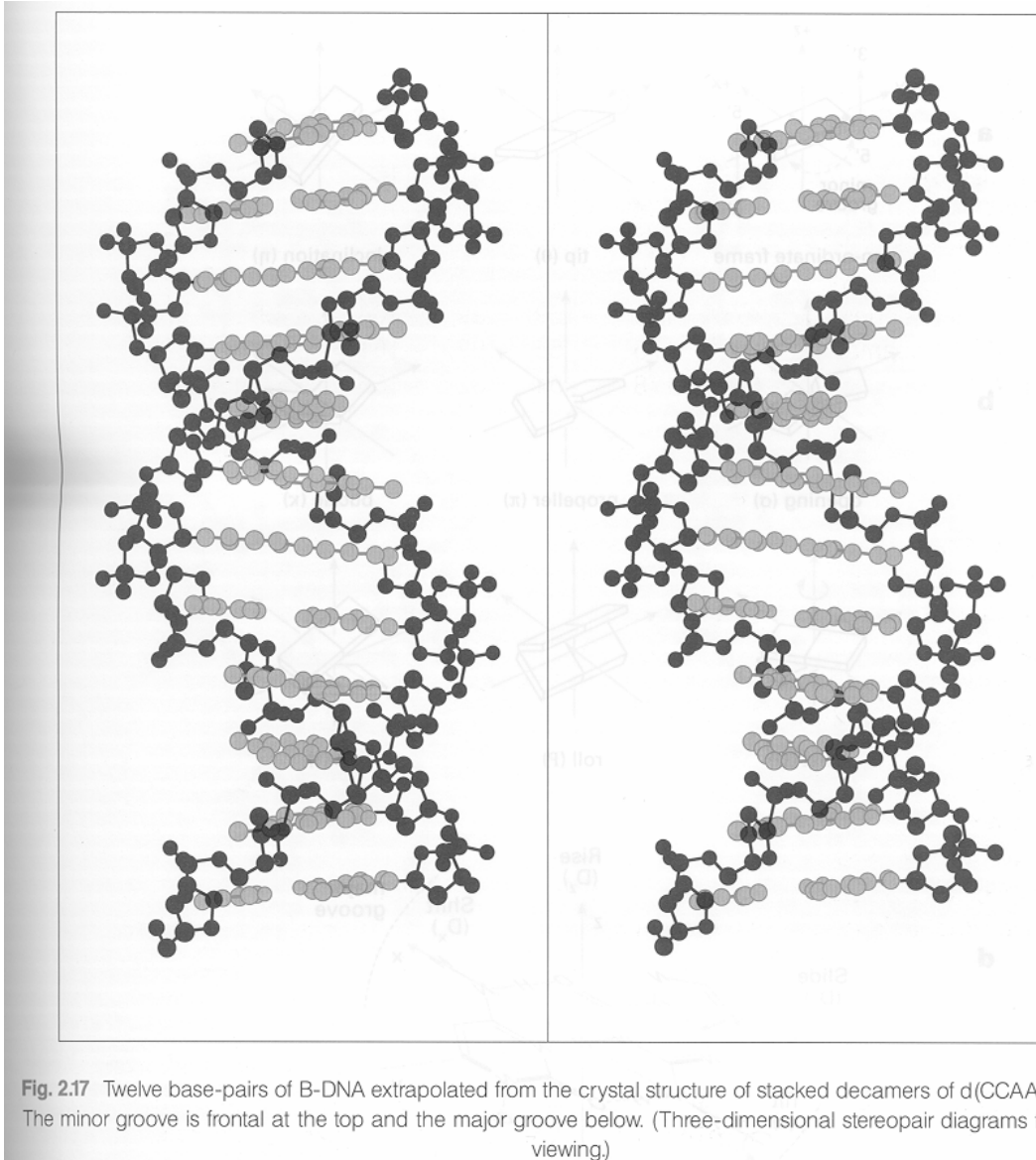
Wobble (G·T)



Guanine (Anti) · Adenine (Syn)

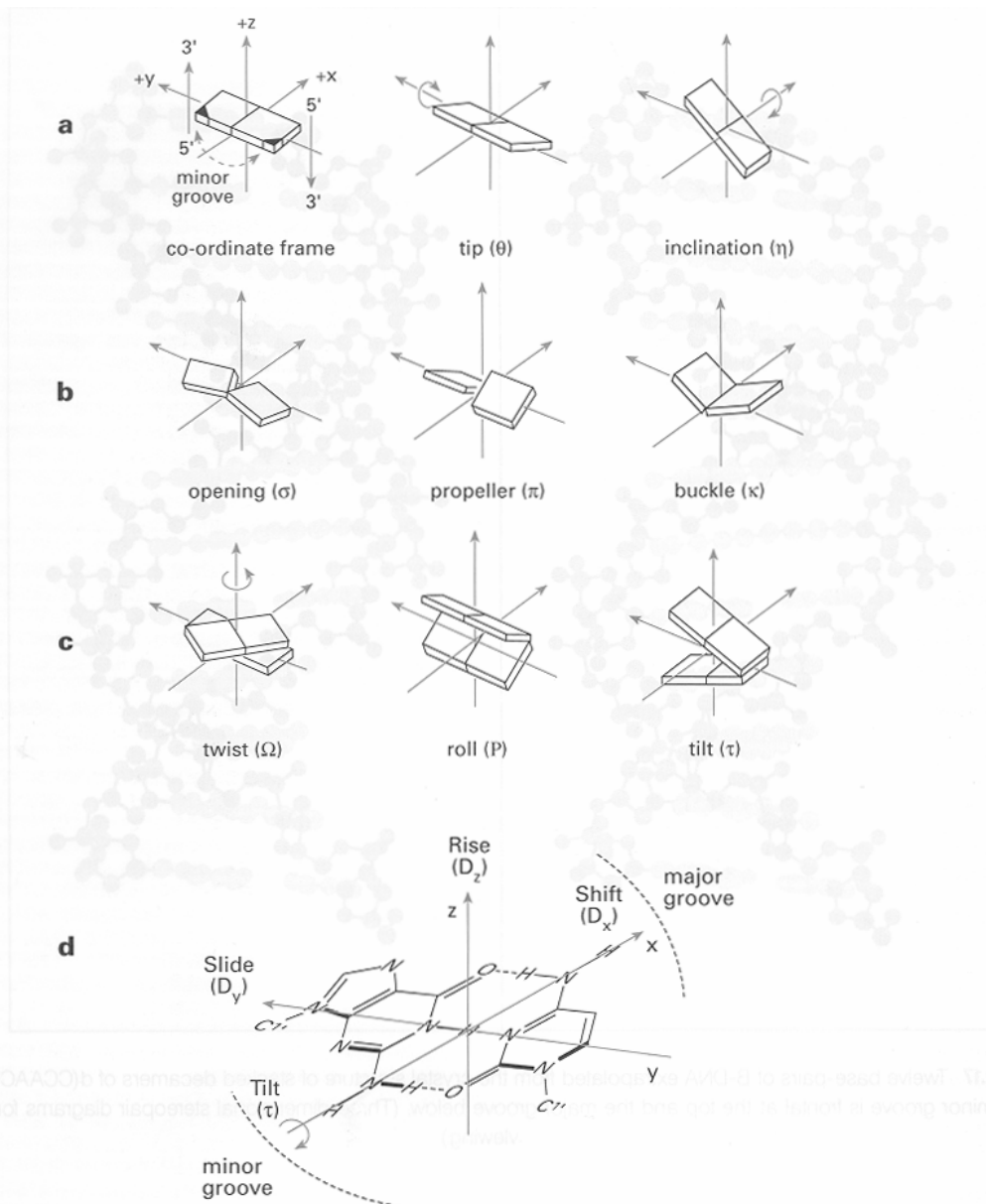
**Figure 1.8 Base pairing schemes.** The top pair of A·T base pairs represent the traditional Watson–Crick configuration (*left*) and a *reversed* Watson–Crick pair, with a 180° rotation of the pyrimidine base (*right*). The center two T·A pairs represent Hoogsteen base pairs. In a standard Hoogsteen base pair the pyrimidine uses its Watson–Crick surface to pair with the N1, C6, N7 side of the purine base. A 180° rotation of the pyrimidine results in formation of a *reversed* Hoogsteen base pair. The bottom left structure shows a wobble G·T base pair. The pyrimidine has been shifted up the length of about one C–C (or C–N) bond. The structure on the bottom right shows a G(anti) · A(syn) base pair with the glycosidic bond of the adenine in the syn configuration.

Below is shown a stereodiagram of a B-DNA duplex based on a high-resolution X-ray structure of a DNA decamer duplex.



X-ray and NMR studies of numerous double-helical DNA oligomers have confirmed the main structural features of the B-DNA helix as described by Watson and Crick, but their structure is clearly an idealized

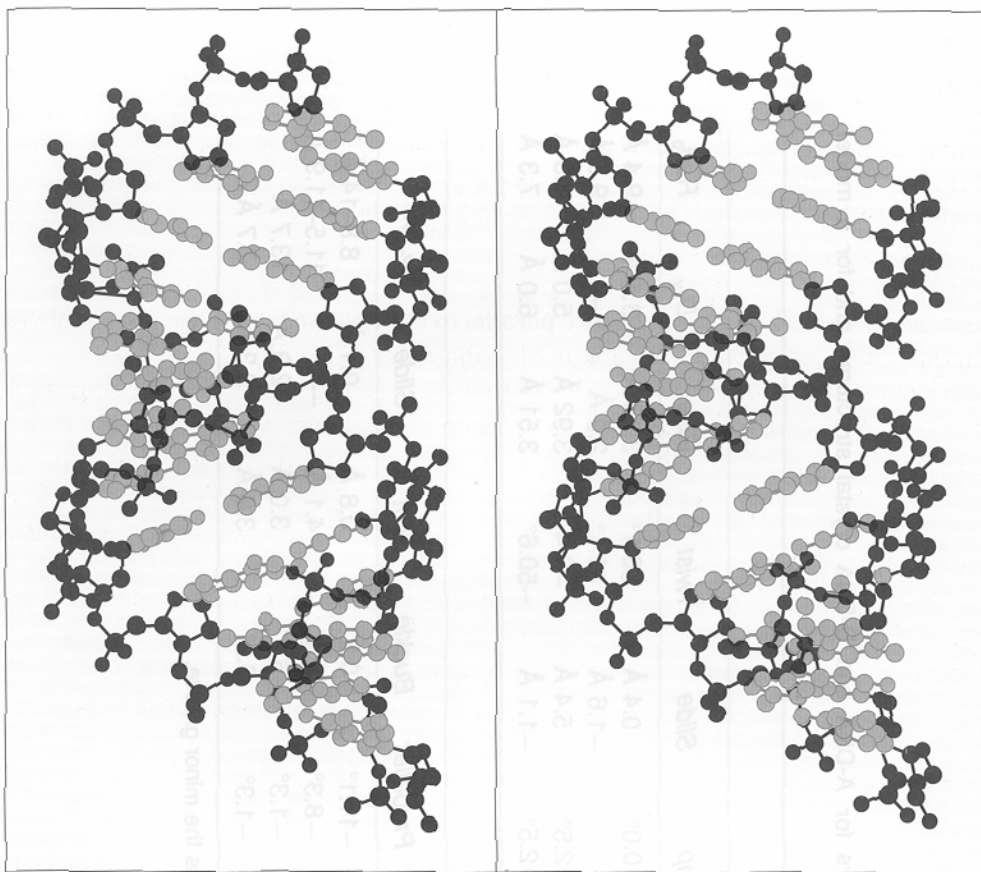
one. The conformation of DNA is very **irregular** in a **sequence-specific** manner. DNA does not exist as a monotonously uniform helix, rather there is a **continuum** of DNA helical structures and the helical parameters (figure page 11) vary quite a bit. The rules specifying how the sequence governs the conformation are far from elucidated, but certain trends have been recognized. We will discuss some of them later. These variations are important for specific interactions with other macromolecules and with drug molecules.



**Fig. 2.18** Diagrams illustrating movements of bases in sequence-dependent structures. Rows (a)–(c) show local rotational helix parameters (from the Cambridge DNA nomenclature accord, see R. E. Dickerson *et al.* (1989), *Nucleic Acids Research*, **17**, 1797–803). Within each of the three vertical columns, rotations are in the x, y, and z axes, from *right to left* respectively. (a) Bases of a pair moving in concert. (b) Bases of a pair moving in opposition. (c) Steps between two base-pairs. (d) Translocational movements of base-pairs relative to the helix axis and to the major and minor grooves.

X-ray diffraction analysis of DNA fibers at 75% relative humidity elucidated the structure of **A-DNA**. A-DNA forms a *wider* and flatter *right-handed* helix than does B-DNA. A-DNA has 11 bp per turn and a pitch of 28 Å, which gives the helix an *axial hole*. The

planes of the base pairs are *tilted* about  $20^\circ$  with respect to the helix axis. A-DNA has a very deep major groove and a very shallow minor groove. The sugar pucker is in the  $C_3'$ -endo (compared with  $C_2'$ -endo for B-DNA). Double-stranded *RNA* molecules form an A-like helix.

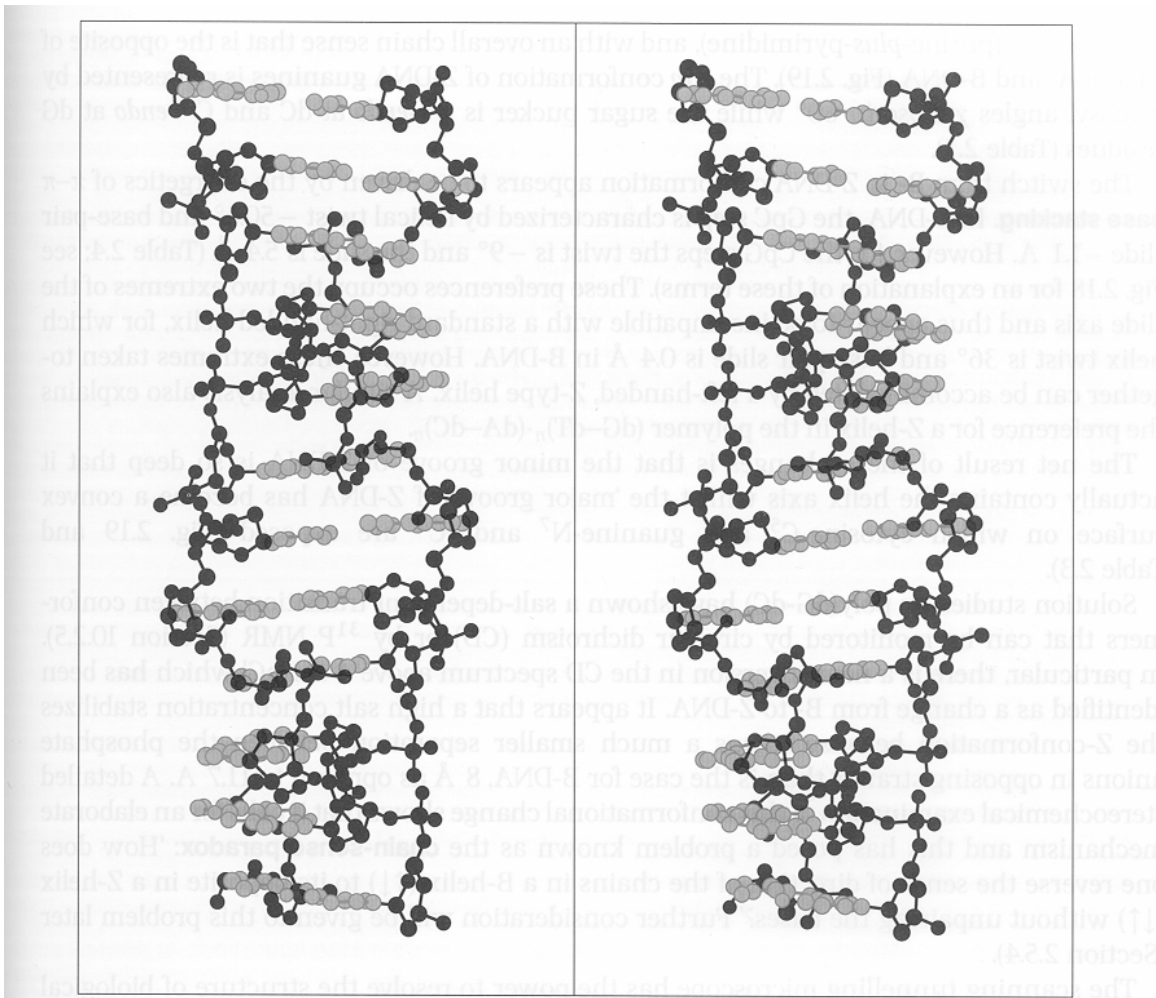


**Fig. 2.16** Twelve base-pairs of A-DNA, generated from the crystal structure of the octadeoxynucleotide, d(GGTATACC). The shallow minor groove is frontal at the top and the very deep major groove is frontal at the bottom of the helix. A significant feature of polymorphism in A-DNA structures is the width of opening of the major groove. (Three-dimensional stereo-pair diagram for parallel viewing.)

**Z-DNA** is a *left-handed anti-parallel* helix that is very different from right-handed DNA forms. It can form in

alternating purine-pyrimidine tract under certain conditions including high salt, the presence of certain divalent cations or DNA supercoiling. The phosphate backbone in Z-DNA follows a *zigzag path* in contrast to the smooth backbone in B-DNA. Only *one narrow and deep groove* is visible, structurally analogous to the minor groove in B-DNA. The Z-DNA helix (diameter 18 Å) is *narrower* than B-DNA (diameter 20 Å) (□ stabilized relative to B-DNA at high salt). The helix repeat is 12 base pairs per turn and the helix pitch is about 45 Å. The sugar pucker for dC is C2'-endo, but the pucker for dG is C3'-endo. The zigzag phosphate backbone is the result of the alternating C2'-endo conformation in pyrimidines and C3'-endo in purines. The C3'-endo sugar pucker reduces the distance between the phosphates attached to the C5' and C3' ribose positions and results in the displacement of the base pairs from the center of the helix towards the phosphate backbone. The purines have the *syn* conformation at the glycosidic bond, the pyrimidines the *anti* conformation. The biological relevance of Z-DNA is still under dispute.





**Fig. 2.19** Twelve base-pairs of left-handed, Z-DNA duplex generated from the crystal structure of the hexamer d(CGCGCG). The minor groove is frontal at the top of the helix while the major groove is seen as a convex surface in the front of the lower part of the figure. The zig-zag nature of the phosphate residues is clearly visible, ascending from right to left across the centre of the helix, as is the close approach of the guanosine 5'-phosphate groups across the minor groove. (Three-dimensional stereopair diagram for parallel viewing.)

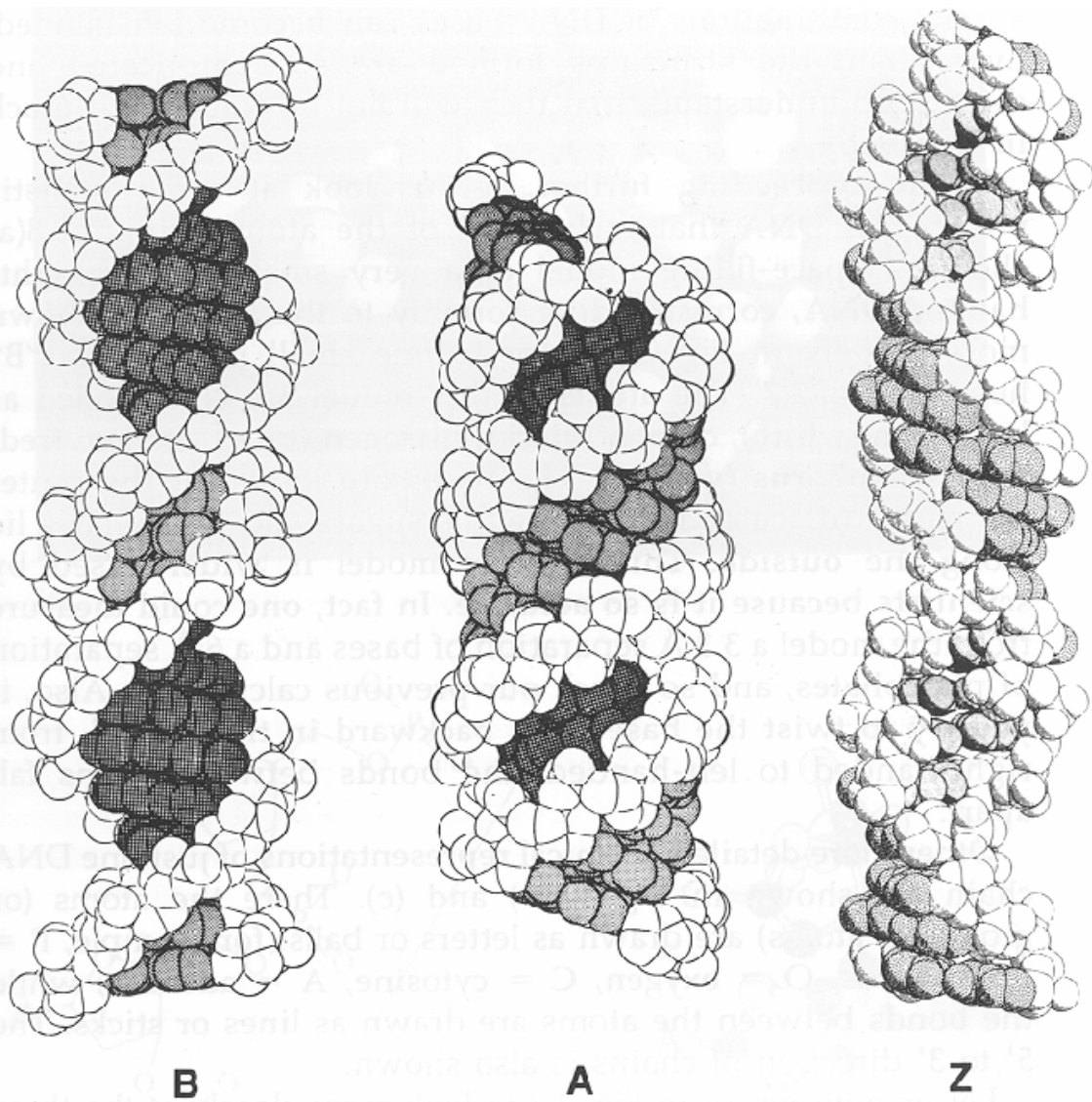


Figure 2.7 Three well-known (but highly idealized) forms of DNA: "B" and "A" are right handed with 10 and 11 phosphates per helical turn, respectively, while "Z" is left-handed with 12 phosphates per turn. Real right-handed DNA in solution averages about 10.5 phosphates per turn, or halfway between "B" and "A". Pictures of "A," "B" from C.J. Alden and S-II. Kim (1979) *Journal of Molecular Biology* 132, 411-34. Picture of "Z" from H.R. Drew and R.E. Dickerson (1981) *Journal of Molecular Biology* 152, 723-36 (with atoms shown somewhat smaller).

TABLE 28-1. STRUCTURAL FEATURES OF IDEAL A-, B-, AND Z-DNA

	A	B	Z
Helical sense	Right handed	Right handed	Left handed
Diameter	~26 Å	~20 Å	~18 Å
Base pairs per helical turn	11	10	12 (6 dimers)
Helical twist per base pair	33°	36°	60° (per dimer)
Helix pitch (rise per turn)	28Å	34Å	45Å
Helix rise per base pair	2.6 Å	3.4 Å	3.7 Å
Base tilt normal to the helix axis	20°	6°	7°
Major groove	Narrow and deep	Wide and Deep	Flat
Minor groove	Wide and shallow	Narrow and deep	Narrow and deep
Sugar pucker	C3'-endo	C2'-endo	C2'-endo for pyrimidines; C3'-endo for purines
Glycosidic bond	Anti	Anti	Anti for pyrimidines; syn for purines

### ***Further reading***

- **DNA structure and function** Sinden, R. R. (Academic Press, 1994)
- **Understanding DNA – The molecule and how it works** Calladine, C. R. and Drew, H. R. (Academic Press, 1997)
- **Unraveling DNA – The most important molecule of life** Frank-Kamenetskii, M. D. (New York, VCH Publishers, 1993)
- **Nucleic acids in Chemistry and Biology** Blackburn, G. M. and Gait, M. J. (Oxford University Press, 1996)