[145 pages; 25,753 words; 163,306 characters (with spaces); 90 refs.; 10 tables; 46 figs.]

CHAPTER 2

"DNA and RNA structure"

Martin Egli

Dept. of Biochemistry, Vanderbilt University, School of Medicine, Nashville, TN 37232, U.S.A. (martin.egli@bvanderbilt.edu)

In: *Nucleic Acids in Chemistry and Biology* (G. M. Blackburn, M. J. Gait, J. D. Loakes & D. M. Williams, Eds.), 3rd edition, Royal Society of Chemistry, Cambridge, UK, 2006, pp. 13-75.

Red: Second level heading
Green: Keywords
Blue: References to other chapters and sections (in part based on 2nd ed.)
Purple: Copyrights required

Contents

2.1 Structures of Components

Nucleic acids are very long, thread-like polymers, made up of a linear array of monomers called **nucleotides**. Different nucleic acids can have from around 80 nucleotides, as in tRNA, to over 10^8 nucleotide pairs in a single eukaryotic chromosome. The unit of size of a nucleic acid is the base-pair (for double-stranded species) or base (for single-stranded species). The abbreviation† bp is generally used, as are the larger units Mbp (million base-pairs) and kbp (thousand base-pairs). The chromosome in *E. coli* has $4x10^6$ base-pairs, 4 Mbp, which gives it a molecular mass of $3x10^9$ Da and a length of 1.5 mm. The size of the (haploid) fruit fly genome is 180 Mbp which, shared between four chromosomes, gives a total length of 56 mm. The genomic DNA of a single human cell has 3900 Mbp and is 990 mm long. How are these extraordinarily long molecules constructed?

2.1.1 Nucleosides and Nucleotides

Nucleotides are the phosphate esters of nucleosides and these are components of both ribonucleic acid (RNA) and deoxyribonucleic acid (DNA). RNA is made up of ribonucleotides while the monomers of DNA are 2'-deoxyribonucleotides.

All nucleotides are constructed from three components: a nitrogen heterocyclic **base**, a pentose **sugar**, and a **phosphate** residue. The major bases are monocyclic **pyrimidines** or bicyclic **purines** (some species of tRNA have tricyclic minor bases such as the Wye (Chapter 3, Figure 3.17). The major purines are **adenine (A)** and **guanine**

† A useful source for IUPAC nomenclature of nucleic acids can be found at http://www.chem.qmul.ac.uk/iupac/misc/naabb.html and for polynucleotide conformation at http://www.chem.qmul.ac.uk/iupac/misc/pnuc2.html#300

(G) and are found in both DNA and RNA. The major pyrimidines are **cytosine (C)**, **thymine (T)**, and **uracil (U)** (**Figure 2.1**).

[Figure 2.1 here]

In **nucleosides**, the purine or pyrimidine base is joined from a ring nitrogen to carbon-1 of a pentose sugar. In ribonucleic acid, the pentose is **D-ribose** which is locked into a five-membered **furanose** ring by the bond from C-1 of the sugar to N-1 of C or U or to N-9 of A or G. This bond is on the same side of the sugar ring as the C-5 hydroxymethyl group and is defined as a β-glycosylic linkage (**Figure 2.2**).

[Figure 2.2 here]

In DNA, the pentose is 2-deoxy-D-ribose and the four nucleosides are **deoxyadenosine**, **deoxyguanosine**, **deoxycytidine** and **deoxythymidine** (**Figure 2.3**). In DNA, the methylated pyrimidine base thymine takes the place of uracil in RNA, and its nucleoside with deoxyribose is still commonly called thymidine. However, since the discovery of **ribothymidine** as a regular component of tRNA species (Section 6.4.1) it has been preferable to use the name deoxythymidine rather than thymidine. Unless indicated otherwise, it is assumed that nucleosides, nucleotides and oligonucleotides are derived from D-pentofuranose sugars.

[Figure 2.3 here]

The **phosphate** esters of nucleosides are **nucleotides**, and the simplest of them have one of the hydroxyl groups of the pentose esterified by a single phosphate monoester function. Adenosine 5'-phosphate is a **5'-ribonucleotide** also called adenylic acid and abbreviated to AMP (**Figure 2.4**). Similarly, deoxycytidine 3'-phosphate is a **3'-deoxynucleotide**, identified as 3'-dCMP. Nucleotides which have two phosphate monoesters on the same sugar are called **nucleoside bisphosphates** while nucleoside monoesters of pyrophosphoric acid are **nucleoside diphosphates**. By extension, nucleoside esters of tripolyphosphoric acid are **nucleoside triphosphates**, of which the classic example is adenosine 5'-triphosphate (ATP) (Section 3.3.2). Finally, cyclic nucleotides are nucleosides which have two neighbouring hydroxyl groups on the same pentose esterified by a single phosphate as a diester. The most important of these is adenosine 3',5'-cyclic phosphate (cAMP).

[Figure 2.4 here]

In the most abbreviated nomenclature currently employed, **pN** stands for 5'-nucleotide, **Np** for a 3'-nucleotide and **dNP** for a 3'-deoxynucleotide (to be precise, a 2'-deoxynucleoside 3'-phosphate). This shorthand notation is based on the convention that an oligonucleotide chain is drawn horizontally with its 5'-hydroxyl group at the left and its 3'-hydroxyl group at the right-hand end. Thus, pppGpp is the shorthand representation of the ‘magic spot' nucleotide, guanosine 3'-diphosphate 5'-triphosphate, while ApG is

short for adenylyl-(3'→5')-guanosine, whose 3'→5' internucleotide linkage runs from the nucleoside on the left to that on the right of the phosphate.

2.1.2 Physical Properties of Nucleosides and Nucleotides

Because of their polyionic character, nucleic acids are soluble in water up to about 1 percent w/v according to size and are precipitated by the addition of alcohol. Their solutions are quite viscous and the long nucleic acid molecules are easily sheared by stirring or by passage through a fine nozzle such as a hypodermic needle or a fine pipette.

Ionization

The acid-base behaviour of a nucleotide is its most important physical characteristic. It determines its charge, its tautomeric structure, and thus its ability to donate and accept hydrogen bonds, which is the key feature of the base : base recognition. The pK_a values for the five bases in the major nucleosides and nucleotides are listed in **Table 2.1**.

[Table 2.1 here]

It is clear that all of the bases are uncharged in the physiological range 5 < pH < 9. The same is true for the pentoses, where the ribose 2',3'-diol only loses a proton above pH 12 while isolated hydroxyl groups ionise only above pH 15. The nucleotide phosphates lose one proton at pH 1 and a second proton (in the case of monoesters) at pH 7. This pattern of proton equilibria is shown for AMP across the whole pH range (**Figure 2.5**).

[Figure 2.5 here]

The three amino bases, A, C and G, each become protonated on one of the ring nitrogens rather than on the exocyclic amino group since this does not interfere with delocalisation of the NH_2 electron lone pair into the aromatic system. The C-NH_2 bonds of A, C and G are about 1.34 Å long which means that they have 40-50 percent double bond order, while the C=O bonds of C, G, T and U have some 85-90 percent double bond order. It is also noteworthy that the proximity of negative charge of the phosphate residues has a secondary effect, making the ring nitrogens more basic ($\Delta pK_a \approx +0.4$) and the amine protons less acidic ($\Delta pK_a \approx +0.6$).

Tautomerism

A tautomeric equilibrium involves alternative structures that differ only in the location of hydrogen atoms. The choices available to nucleic acid bases are illustrated by the **keto-enol** equilibrium between 2-pyridone and 2-hydroxypyridine and the **amine-imine** equilibrium for 2-aminopyridine (**Figure 2.6**). Ultraviolet, NMR, and IR spectroscopies have established that the five major bases exist overwhelmingly (> 99.99 %) in the **amino**- and **keto**-tautomeric forms at physiological pH (see **Figure 2.1**) and not in the benzene-like **enol** tautomers, in common use before 1950 (Chapter 1, Figure 1.3).

[Figure 2.6 here]

Hydrogen Bonding

The mutual recognition of A by T and of C by G uses hydrogen bonds to establish the fidelity of DNA transcription and translation. The NH groups of the bases are good hydrogen bond donors (**d**), while the sp^2-hybridized electron pairs on the oxygens of the base C=O groups and on the ring nitrogens are much better hydrogen bond acceptors (**a**) than are the oxygens of either the phosphate or the pentose. The **a•d** hydrogen bonds so formed are largely electrostatic in character, with a charge of about +0.2*e* on the hydrogens and about –0.2*e* on the oxygens and nitrogens, and they seem to have an average strength of 6-10 kJ mol^{-1}.

The predominant amino-keto tautomer for cytosine has a pattern of hydrogen bond acceptor and donor sites for which O^2•N^3•N^4 can be expressed as **a•a•d** (**Figure 2.7**). Its minor tautomer has a very different pattern: **a•d•a**. In the same way we can establish that the corresponding pattern for the dominant tautomer of dT is **a•d•a** while the pattern for N^2•N^1•O^6 of dG is **d•d•a** (**Figure 2.7**) and that for dA is **(-)•a•d.**

[Figure 2.7 here]

When Jim Watson was engaged in DNA model-building studies in 1952 (Section 1.4) he recognised that the hydrogen bonding capability of an A·T base-pair uses complementarity of **(-)•a•d** to **a•d•a** while a C·G pair uses the complementarity of **a•a•d** to **d•d•a.** This base-pairing pattern rapidly became known as **Watson-Crick pairing** (**Figure 2.8**). There are two hydrogen bonds in an A•T pair and three in a C•G pair. The

geometry of the pairs has been fully analysed in many structures from dinucleoside phosphates through oligonucleotides to tRNA species, both by the use of X-ray crystallography and, more recently, by NMR spectroscopy.

[Figure 2.8 here]

In planar base-pairs, the hydrogen bonds join nitrogen and oxygen atoms that are 2.8 Å to 2.95 Å apart. This geometry gives a C-1'···C-1' distance of 10.60 ± 0.15 Å with an angle of 68 ± 2° between the two glycosylic bonds for both the A•T and the C•G base-pairs. As a result of this **isomorphous geometry**, the four base-pair combinations A•T, T•A, C•G and G•C can all be built into the same regular framework of the DNA duplex.

While Watson-Crick base-pairing is the dominant pattern, other pairings have been suggested of which the most significant to have been identified so far are **Hoogsteen pairs** and Crick **'wobble' pairs**. Hoogsteen pairs, illustrated for A•T are not isomorphous with Watson-Crick pairs because they have an 80° angle between the glycosylic bonds and an 8.6 Å separation of the anomeric carbons (**Figure 2.8**). In the case of reverse Hoogsteen pairs and reverse Watson-Crick pairs (not shown), one base is rotated through 180° relative to the other.

Francis Crick proposed the existence of 'wobble' base-pairings to explain the degeneracy of the genetic code (Section 7.3.1). This phenomenon calls for a single base in the 5'-anticodon position of tRNA to be able to recognise either of the pyrimidines or, alternatively, either of the purines as its 3'-codon base partner. Thus a G•U 'wobble' pair has two hydrogen bonds, G-N1-H···O2-U and G-O6···H-N3-U and this requires a

sideways shift of one base relative to its positions in the regular Watson-Crick geometry (**Figure 2.9**). The resulting loss of a hydrogen bond leads to reduced stability which can be offset in part by the improved **base-stacking** (Section 2.3.1) that results from such sideways base displacement.

[Figure 2.9 here]

Base-pairings of these and other non-Watson-Crick patterns is significant in three structural situations. First, the compact structures of RNAs maximise both base-pairing and base-stacking wherever possible. This has lead to the identification of a considerable variety of reverse Hoogsteen and 'wobble' base-pairs as well as of tertiary base-pairs (or base-triplets) (Section 7.1.2). Secondly, where there are triple-stranded helices for DNA and RNA, such as (poly(dA)•2poly(dT)) and (poly(rG)•2poly(rC)), the second pyrimidine chain binds to the purine in the major groove by Hoogsteen hydrogen bonds and runs parallel to the purine chain (Sections 2.3.6 and 2.4.5). Thirdly, mismatched base-pairs are necessarily identified with anomalous hydrogen bonding and many such patterns have been revealed by X-ray studies on synthetic oligodeoxyribonucleotides (Section 2.3.2). They are also targets for some DNA repair enzymes (Section 8.11).

2.1.3 Spectroscopic Properties of Nucleosides and Nucleotides

Neither the pentose nor the phosphate components of nucleotides show any significant UV absorption above 230 nm. This means that both nucleosides and nucleotides have UV absorption profiles rather similar to those of their constituent bases and absorb

strongly with λ_{max} values close to 260 nm and molar extinction coefficients of around 10^4 (**Table 2.2**)

[Table 2.2 here]

The light absorptions of isolated nucleoside bases given above are measured in solution in high dilution. They undergo marked changes when they are in close proximity to neighbouring bases, as usually shown in ordered secondary structures of **oligo-** and polynucleotides. In such ordered structures, the bases can stack face-to-face and thus share π-π electron interactions that profoundly affect the transition dipoles of the bases. Typically such changes are manifest in a marked reduction in the intensity of UV absorption (by up to 30 percent), which is known as **hypochromicity** (Section 5.5.1). This phenomenon is reversed on unstacking of the bases.

There are two important applications of this phenomenon. First it is used in the determination of temperature-dependent and pH-dependent changes in base-stacking. Secondly, it permits the monitoring of changes in the asymmetric environment of the bases by circular dichroism (CD), or by optical rotatory dispersion (ORD) effects. Both of these techniques are especially valuable for studying helix-coil transitions (Section 11.1.3).

Infrared analysis of nucleic acid components has been less widely used, but the availability of laser Raman and Fourier transform IR methods is making a growing contribution (Section 11.1.4).

[Figure 2.10 here]

Nuclear magnetic resonance has had a dramatic effect on studies of oligonucleotides largely as a result of a variety of complex spin techniques such as NOESY and COSY for proton spectra, the use of ^{17}O, ^{18}O and sulfur substituent effects of ^{31}P NMR, and the analysis of nuclear Overhauser effects (NOE). These provide a useful measure of internuclear distances and with computational analysis can provide solution conformations of oligonucleotides (Section 2.2). Nucleosides, nucleotides and their analogues have relatively simple ^{1}H NMR spectra. The aromatic protons of the pyrimidines and purines resonate at low field (δ7.6 to δ8.3 with C5-H close to δ5.9). The anomeric hydrogen is a doublet for ribonucleosides and a double-doublet for 2'-deoxynucleosides at δ5.8-6.4. The pentoses provide a multi-spin system which generally moves from low to high field in the series: H-2', H-3', H-4', H-5' and H-5" in the region δ4.3 to 3.7. Lastly, 2'-deoxynucleosides have H-2' and H-2" as an ABMX system near δ2.5. The 400 MHz spectrum of a simple nucleoside, cytidine (**Figure 2.10**), shows why two-dimensional spin techniques are required for the complete analysis of the spectrum in a large oligomer, which may be equivalent to a dozen such monomer spectra superimposed.

2.1.4 Shapes of Nucleotides

Nucleotides have rather compact shapes with several interactions between non-bonded atoms. Their molecular geometry is so closely related to that of the corresponding nucleotide units in oligomers and nucleic acid helices that it was once argued that helix

structure is a consequence of the conformational preferences of individual nucleotides. However, the current view is that sugar-phosphate backbone appears to act as no more than a constraint on the range of conformational space accessible to the base-pairs and that **π-π interactions** between the base-pairs provide the driving force for the different conformations of DNA (Section 2.3.1).

[Figure 2.11 here]

The details of conformational structure are accurately defined by the torsion angles α, β, γ, δ, ε and ζ in the phosphate backbone, θ_0 to θ_4 in the furanose ring, and χ for the glycosylic bond (**Figure 2.11**). Because many of these torsional angles are interdependent, we can more simply describe the shapes of nucleotides in terms of four parameters: the sugar pucker, the *syn-anti* conformation of the glycosylic bond, the orientation of C4'-C5', and the shape of the phosphate ester bonds.

Sugar Pucker

The furanose rings are twisted out of plane in order to minimize non-bonded interactions between their substituents. This 'puckering' is described by identifying the major displacement of carbons-2' and -3' from the median plane of C1'-O4'-C4'. Thus, if the *endo* displacement of C-2' is greater than the *exo* displacement of C-3' the conformation is called C2'-*endo* and so on (**Figure 2.11**). The *endo* face of the furanose is on the same side as C5' and the base; the *exo* face is on the opposite face to the base. These sugar puckers are located in the north (***N***) and south (***S***) domains of the **pseudorotation cycle**

of the furanose ring and so spectroscopists frequently use ***N*** and ***S*** designations, which also fortuitously reflect the relative shapes of the C-C-C-C bonds in the C2'-*endo* and -*exo* forms, respectively.[1]

In solution, the *N* and *S* conformations are in rapid equilibrium and are separated by an energy barrier of less than 20 kJ mol^{-1}. The average position of the equilibrium can be estimated from the magnitudes of the ^{3}J NMR coupling constants linking H1'-H2' and H3'-H4'. This is influenced by (1) the preference of electronegative substituents at C2' and C3' for axial orientation, (2) the orientation of the base (*syn* goes with C2'-*endo*) and (3) the formation of an intra-strand hydrogen bond from O2' in one RNA residue to O4' in the next which favours C3'-*endo* pucker. However, in RNA helical regions, this latter hydrogen bond is not often observed and an axial C-H···O interaction between the C2'-H2'(*n*) group and the O4'(*n+1*) atom appears to make a more important contribution to the stability of RNA helices.

***Syn-anti* Conformation**

The plane of the bases is almost perpendicular to that of the sugars and approximately bisects the O4'-C1'-C2' angle. This allows the bases to occupy either of two principal orientations. The *anti* conformer has the smaller H-6 (pyrimidine) or H-8 (purine) atom above the sugar ring, while the *syn* conformer has the larger O-2 (pyrimidine) or N-3 (purine) in that position. Pyrimidines occupy a narrow range of *anti* conformations (**Figure 2.12**) while purines are found in a wider range of *anti* conformations that can even extend into the high-*anti* range for 8-azapurine nucleosides such as formycin.

[Figure 2.12 here]

One inevitable consequence of this *anti* conformation for the glycosylic bonds is that the backbone chains for A-form and B-form DNA run downward on the right of the minor groove and run upward on the left of the minor groove, depicted as (↑↓).

There is one important exception to the general preference for *anti* forms. Nuclear magnetic resonance, CD and X-ray analyses all show that guanine prefers the *syn* glycoside in mono-nucleotides, in alternating oligomers like d(CpGpCpG) and in Z-DNA. Theoretical calculations suggest that this effect comes from a favourable electrostatic attraction between the phosphate anion and the C2-amino group in guanine nucleotides. It results from polarization of one of the nitrogen non-bonding electrons towards the ring. Most unusually, this *syn* conformation can only be built into left-handed helices.

C4'-C5' Orientation

The conformation of the exocyclic C4'-C5' bond determines the position of the 5'-phosphate relative to the sugar ring. The three favoured conformers for this bond are the classical **synclinal (sc)** and **antiperiplanar (ap)** rotamers. For pyrimidine nucleosides, **+sc** is preferred while for purine nucleosides **+sc** and **ap** are equally populated. However, in the nucleotides, the 5'-phosphate reduces the conformational freedom and the dominant conformer for this γ-bond is **+sc** (**Figure 2.13**). Once again, the demands of Z-DNA have a major effect and the **ap** conformer is found for the *syn* guanine deoxynucleotides.

[Figure 2.13 here]

C-O and P-O Ester Bonds

Phosphate diesters are tetrahedral at phosphorus and show antiperiplanar conformations for the C5'-O5' bond. Similarly, the C3'-O3' bond lies in the **antiperiplanar** to **anticlinal** sector. This conformational uniformity has led to the use of the **virtual bond concept** in which the chains P5'-O5'-C5'-C4' and P3'-O3'-C3'-C4' can be analysed as rigid, planar units linked at phosphorus and at C4'. Such a simplification has been used to speed up initial calculations of some complex polymeric structures.

[Figure 2.14 here]

Our knowledge of P-O bond conformations comes largely from X-ray structures of tRNA and DNA oligomers. In general, H4'-C4'-C5'-O5'-P adopts an extended W-conformation in these structures. A skewed conformation for the C-O-P-O-C system has been observed in structures of simple phosphate diesters such as dimethyl phosphate and also for polynucleotides. This has been described as an **anomeric** effect and attributed to the favourable interactions of a non-bonding electron pair on O5' with the P-O3' bond, and vice versa for the P-O5' bond (**Figure 2.14**). This may arise from interaction of the electron lone pair with either phosphorus d orbitals or, more likely, with the P-O antibonding σ orbital. The interaction has been calculated at 30 kJ mol^{-1} more favourable than the extended W-conformation for the C-O-P-O-C system. Other non-bonded

interactions dictate that α and ζ both have values close to +300° in helical structures though values of +60° are seen in some dinucleoside phosphate structures.

Other P-O conformations have been observed in non-helical nucleotides while left-handed helices also require changed P-O conformations. These changes take place largely in the rotamers for α. In Z-DNA these are +sc for guanines but broadly **antiperiplanar** for the cytosines while ζ is +sc for cytosines but broadly **synperiplanar** for guanines (Section 2.2.2).

2.2 Standard DNA Structures

Structural studies on DNA began with the nature of the primary structure of DNA. The classical analysis, completed in mid 20thcentury, is easily taken for granted today when we have machines for DNA oligomer synthesis that presuppose the integrity of the 3'-to-5' phosphate diester linkage. Nonetheless, the classical analysis was the essential key that opened the door to later studies on the regular secondary structure of double-stranded DNA and thereby primed the modern revolution known as molecular biology. **Standard structures** for DNA have generally been determined on heterogeneous duplex material and are thus independent of sequence and apply only to Watson-Crick base-pairing.

2.2.1 Primary Structure of DNA

Klein and Thannhauser's work (Section 1.4) established that the primary structure of DNA has each nucleoside joined by a phosphate diester from its 5'-hydroxyl group to the 3'-hydroxyl group of one neighbour and by a second phosphate diester from its 3'-hydroxyl group to the 5'-hydroxyl of its other neighbour. There are no 5'-5' or 3'-3'

linkages in the regular DNA primary structure (**Figure 2.15**). This means that the uniqueness of a given DNA primary structure resides solely in the sequence of its bases.

[Figure 2.15 here]

2.2.2 Secondary Structure of DNA

In the first phase of investigation of DNA secondary structure, diffraction studies on heterogeneous DNA fibres identified two distinct conformations for the DNA double helix.[2] At **low humidity** (and high salt) the favoured form is the highly crystalline **A-DNA** while at **high humidity** (and low salt) the dominant structure is **B-DNA**. We now recognise that there is a wide variety of right-handed double helical DNA conformations and this **structural polymorphism** is denoted by the use of the letters A to T as illustrated by A, A', B, α-B', β-B', C, C', D, E and T forms of DNA. In broad terms, all of these can be classified in two generically different DNA families: A and B. These are associated with the sugar pucker C3'-*endo* for the A-family and C2'-*endo* (or the equivalent C3'-*exo*) for the B-family. However, as we shall see later it is the energetics of base-stacking which determines the conformation of the helix and sugar pucker is largely consequential. We shall also see that in B-form DNA the base-pairs sit directly on the helix axis and are nearly perpendicular to it. In A-form DNA the base-pairs are displaced off-axis towards the minor groove and are inclined.

The unexpected discovery by Wang, Rich and coworkers in 1979 that the hexamer d(CGCGCG) adopts a left-handed helical structure, now named **Z-DNA**, was one of the first dramatic results to stem from the synthesis of oligonucleotides in

sufficient quantity for crystallization and X-ray diffraction analysis.[3] Since then, over 100 different oligodeoxynucleotide structures have been solved and these have provided the details on which standard DNA structures are now based.[4,5] The main features of A-, B- and Z-DNA are shown in **Figures 2.16-2.19** and structural parameters are provided for a range of standard helices in **Tables 2.3** and **2.4**.

[Figure 2.16 here]

As more highly resolved structures have become available the idea that these three families of DNA conformations are restricted to standard structures has been whittled away.[6,7] We now accept that there are local, sequence-dependent modulations of structures that are primarily associated with the changes in the orientation of bases. Such changes seek to minimise non-bonded interactions between adjacent bases and maximise base-stacking. They are generally tolerated by the relatively flexible sugar-phosphate backbone. Other studies have explored perturbations in regular helices which result from deliberate mismatching of base-pairs and of lesions caused by chemical modification of bases, such as **base methylation** and **thymine photodimers** (Section 8.8.1). In all of these areas, the results derived from X-ray crystallography have been carried into solution phase by high-resolution NMR analysis, and rationalized by molecular modelling.

Finally our knowledge of higher-order structures, which began with Vinograd's work on DNA **supercoiling** in 1965, has been extended to studies on DNA cruciform structures to ‘bent' DNA and to other unusual features of DNA structures.

Regular DNA structures are described by a range of characteristic features.[8,9] The global parameters of **average rise D_z** and **helix rotation Ω** per base-pair define the pitch of the helix. Sideways tilting of the base-pairs through a **tilt angle τ** permits the separation of the bases along the **helix axis** D_z to be smaller than the van der Waals distance, 3.4 Å and so gives a shorter, fatter cylindrical envelope for DNA. The angle τ is positive for A-DNA (positive means a clockwise rotation of the base-pair when viewed end-on and towards the helix axis) but is smaller and negative for B-DNA helices. At the same time, the base-pairs are displaced laterally from the helix axis by a distance D_a. This parameter together with the groove width defines the depth of the **major groove** and the **minor groove** (**Table 2.3**).

[Tables 2.3 and 2.4 here]

2.2.3 A-DNA

Among the first synthetic oligonucleotides to be crystallized in the late 1970s were d(GGTATACC), an iodinated-d(CCGG) and d(GGCCGGCC). They all proved to have A-type DNA structures, similar to the classical A-DNA deduced from fibre analysis at low resolution. Several other oligomers, mostly octamers, also form crystals of the A-structure, but NMR studies suggest that some of these may have the B-form in solution. It is conceivable that crystal packing might especially favour A-DNA for octanucleotides.

The general anatomy of A-DNA follows the Watson-Crick model with antiparallel, right-handed double helices. The sugar rings are parallel to the helix axis

and the phosphate backbone is on the outside of a cylinder of about 24 Å diameter (**Figure 2.16**).

X-ray diffraction at atomic resolution shows that the bases are displaced 4.5 Å away from the helix axis and this creates a hollow core down the axis around 3 Å in diameter. There are 11 bases in each turn of 28 Å, which gives a vertical rise of 2.56 Å per base-pair. In order to maintain the normal van der Waals separation of 3.4 Å, the stacked bases are tilted sideways through 20°. The sugar backbone has skewed phosphate ester bonds, and antiperiplanar conformations for the adjacent C-O ester bonds. Finally, the furanose ring has a C3'-*endo* pucker and the glycosylic bond is in the *anti* conformation (**Table 2.3**). As a result of these features, the major groove of A-DNA is cavernously deep and the minor groove is extremely shallow, as can be appreciated from the three-dimensional picture of the helix (**Figure 2.16**). This is further characterised by an approximate 5.4 Å P···P separation between adjacent intra-strand phosphorus atoms.

2.2.4 The B-DNA Family

The general features of the B-type structure, obtained from DNA fibres at high relative humidity (95% R.H.) were first put into sharper focus by X-ray studies on the dodecamer d(CGCGAATTCGCG) and its C-5 bromo-derivative at cytosine-9. The structure of the so-called Dickerson-Drew dodecamer has now been revealed at atomic resolution.[10] The B-conformation has been observed in crystals of numerous oligomers and initial standard parameters were averaged from structures of ten isomorphous oligodeoxynucleotides (**Figure 2.17**).

In B-form DNA, the base-pairs sit directly on the helix axis so that the major and minor grooves are of similar depth (**Table 2.3**). Its bases are stacked predominantly

above their neighbours in the same strand and are perpendicular to the helix axis (**Table 2.4**). The sugars have the C2'-*endo* pucker (with some displaying puckers in the neighbouring ranges of the pseudororotation phase cycle, such as C1'-*exo* or O4'-*endo*), all the glycosides have the *anti* conformation, and most of the other rotamers have normal populations (**Table 2.5**). Adjacent phosphates in the same chain are further apart, P···P = 6.7 Å, than in A-DNA (**Table 2.4**).

[Figure 2.17 here]

[Table 2.5 here]

The interaction of water molecules around a DNA double helix can be very important in stabilising helix structure,[11] to the extent that **hydration** has sometimes been described as the 'fourth component' of DNA structure, after bases, sugars, and phosphates. Just how many water molecules per base-pair can be seen in an X-ray structure depends on the quality of structure resolution. In the best structures, up to 14 unique waters per base-pair have been resolved. For B-DNA, whose stability is closely linked to high humidity (Section 2.3.1), highly ordered water molecules can be seen in both major and minor grooves. The broad major groove is 'coated' by a unimolecular layer of water molecules that interact with exposed C=O, N, and NH functions and also extensively solvate the phosphate backbone. The narrow minor groove contains an inner and an outer zig-zag chain of water molecules that form four regular planar hexagons in the central A•T region of the Dickerson-Drew dodecamer (**Figure 2.18**).[12] The inner **spine of hydration** consists of alternating water molecules that are buried at the floor of

the groove, directly contacting the bases, and located in the second-shell, above and between first-shell water molecules and closer to the periphery of the groove, respectively.

[Figure 2.18 here]

To a first approximation, the differences between the A, B and other polymorphs of DNA can be described in terms of just two coordinates: slide (D_y) and roll (σ). Clearly A-DNA has high roll and negative slide while B-DNA has little roll and small positive slide. These and other movements of base-pairs are illustrated in **Figure 2.19** a-c and values of the parameters given in **Table 2.4**. This results in a greater hydrophobic surface area of the bases being exposed in A-DNA per base-pair. From this, it has been argued that B-DNA will have the lesser energy of solvation, explaining its greater stability at high humidity (95%) and that this hydrophobic effect may well tip the balance between the A-form and B-form helices.

[Figure 2.19 here]

Other B-DNA structures have much lower significance. C-DNA is obtained from the lithium salt of natural DNA at rather low humidity.[2] It has 28 bases and three full turns of the helix. D-DNA is observed for alternating A•T regions of DNA and has an overwound helix compared to B-DNA with 8 bp per turn. In phage T2 DNA, where cytosine bases have been replaced by glucosylated 5-hydroxymethylcytosines, the B-conformation observed at high humidity changes into a T-DNA form at low humidity

(<60% relative humidity), which also has eightfold symmetry around the helix (see **Table 2.3**).

2.2.5 Z-DNA

Two of the earliest crystalline oligodeoxyribonucleotides, d(CGCGCG) and d(CGCG), provided structures of a new type of DNA conformer, the left-handed Z-DNA, which has also been found for d(CGCATGCG). Initially it was thought that left-handed DNA had a strict requirement for alternating purine-pyrimidine sequences. We now know that this condition is neither necessary nor sufficient since left-handed structures have been found for crystals of d(CGATCG) in which cytosines have been modified by C-5 bromination or methylation and have been identified for GTTTG and GACTG sequences by supercoil relaxation studies (Section 2.3.4)

The Z-helix is also an antiparallel duplex but is a radical departure from the A- and B-forms of DNA. It is best typified by an alternating $(dG\text{-}dC)_n$ polymer. Its two backbone strands run downwards at the left of the minor groove and upwards at the right (↓↑), and this is the opposite from those of A- and B-DNA (↑↓) (NB: the forward direction is defined as the sequence O3'→P→O5'). In an idealized left-handed duplex, such reversed chain directions would require all the nucleosides to have the *syn*-conformation for their glycosylic bonds. However, this is not possible for the pyrimidines because of the clash between O2 of the pyrimidine and the sugar furanose ring (Section 2.1.4). So the cytosines take the *anti* conformation and the guanines the *syn* conformation. The name Z-DNA results from this *anti-syn* feature of the glycosylic bonds that **alternates regularly along the backbone** (**Figure 2.20**). It causes a local

chain reversal that generates a **zig-zag backbone** path and produces a helical repeat consisting of two successive bases (purine-*plus*-pyrimidine) and with an overall chain sense that is the opposite of that of A- and B-DNA. The *syn* conformation of Z-DNA guanines is represented by glycosyl angles χ close to 60^{o} while the sugar pucker is C2'-*endo* at dC and C3'-*endo* at dG residues (**Table 2.3**).[13]

[Figure 2.20 here]

The switch from B- to Z-DNA conformation appears to be driven by the energetics of π-π **base-stacking**. In Z-DNA the GpC step is characterised by helical twist of -50.6^{o} and a base-pair slide of –1.1 Å. However, for the CpG steps the twist is -9^{o} and the slide is 5.4 Å (**Table 2.4**; see **Figure 2.19** for an explanation of these terms). These preferences occupy the two extremes of the slide axis and thus appear to be incompatible with a standard right-handed helix, for which helix twist is 36^{o} and base-pair slide is 0.4 Å in B-DNA. However, these extremes taken together can be accommodated by a left-handed, Z-type helix. A similar analysis also explains the preference for a Z-helix in the polymer $(dG\text{-}dT)_n \bullet (dA\text{-}dC)_n$.

The net result of these changes is that the minor groove of Z-DNA is so deep that it actually contains the helix axis whilst the ‘major groove' of Z-DNA has become a convex surface on which cytosine-C5 and guanine-N7 and -C8 are exposed (**Figure 2.20** and **Table 2.3**).

Solution studies on poly(dG-dC) have shown a salt-dependent transition between conformers that can be monitored by circular dichroism (CD) or by ^{31}P NMR (Section

11.2). In particular, there is a near inversion in the CD spectrum above 4 M NaCl which has been identified as a change from B- to Z-DNA. It appears that a high salt concentration stabilises the Z-conformation because it has a much smaller separation between the phosphate anions in opposing strands than is the case for B-DNA, 8 Å as opposed to 11.7 Å. A detailed stereochemical examination of this conformational change shows that it calls for an elaborate mechanism and this has posed a problem known as the **chain-sense paradox**: 'How does one reverse the sense of direction of the chains in a B-helix (↑↓) to its opposite in a Z-helix (↓↑) without unpairing the bases?' Further consideration to this problem will be given later (Section 2.5.4).

The scanning tunnelling microscope has the power to resolve the structure of biological molecules with atomic detail (Section 11.5.2). Much progress has been made with dried samples of duplex DNA, in recording images of DNA under water, and in revealing details of single-stranded poly(dA). Such STM microscopy has provided images of poly(dG-me^{5}dC)•poly(dG-me^{5}dC) in the Z-form. Both the general appearance of the fibres and measurements of helical parameters are in good agreement with models derived from X-ray diffraction data.

2.3 Real DNA Structures

2.3.1 Sequence-dependent Modulation of DNA Structure

So far we have emphasised the importance of hydrogen bonds in base-pairing and DNA structure and have said little about base-stacking. We shall see later that both these two features are important for the energetics and dynamics of DNA helices (Section 2.5), but it is now time to look at the major part played by base-stacking in real DNA structures.

Two particular hallmarks of B-DNA, in contrast to the A- and Z-forms, are its flexibility and its capacity to make small adjustments in local helix structure in response to particular base sequences.[14]

Different base sequences have their own characteristic signature: they influence groove width, helical twist, curvature, mechanical rigidity and resistance to bending. It seems probable that these features help proteins to read and recognise one base sequence in preference to another (Chapter 10), possibly only through changes in the positions of the phosphates in the backbone. What do we know about these sequence-dependent structural features?

One surprise to emerge from single-crystal structure analyses of synthetic DNA oligomers has been the breadth of variation of local helix parameters relative to the mean values broadly derived from fibre diffraction analysis and used for the standard A- and B-form DNA structures described earlier. Dickerson has compared eight dodecamer and three decamer B-DNA structures.[15] The mean value of the **helical twist** angle between neighbouring base-pairs is 36.1° but the standard deviation (SD) is 5.9° and the range is from 24° to 51°. Likewise, the mean **helical rise** per base-pair is 3.36 Å with a SD of 0.46 Å but with a range from 2.5 to 4.4 Å. (NB: because rise is a parameter measured between the C-1' atoms of adjacent base-pairs, it can be smaller than the thickness of a base-pair if the ends of the two base-pairs bow towards each other. Such **bowing** is also defined as 'positive cup'). Roll angles between successive base-pairs average +0.6° but with a SD of 6.0° and a range from –18° to +16°. These variations in twist and roll have the effect of substantially re-orienting the potential hydrogen bond acceptors and donors at the edges of the bases along the floor of the DNA grooves, so they may well be a

significant component of the sequence-recognition process used by drugs and proteins (Chapters 9 and 10). These and other modes of local changes in the geometry of base-pairs are illustrated in **Figure 2.19**.

The major irregularities in the positions of the bases in real DNA structures contrast with only secondary, small conformational changes in their sugar-phosphate backbones. The main characteristic of these sequence-dependent modulations is **propeller twist**. This results when the bases rotate by some 5^{o} to 25^{o} relative to their hydrogen-bonded partner around the long axis through C-8 of the purine and C-6 of the pyrimidine (**Figure 2.19b**, centre). Sections of oligonucleotides with consecutive A residues, as in d(CGCAAAAAAGCG)•d(CGCTTTTTTGCG) have unusually high propeller twist (approximately 25^{o}) and these permit the formation of a three-centred hydrogen bonding network in the major groove between adenine-N6 and two thymine-O4 residues, the first being the W-C base-pair partner and the second being its 3'- neighbour, both in the opposing strand. This network of hydrogen bonds gives added rigidity to the duplex and may explain why long runs of adenines are not found in the more sharply curved tracts of chromosomes (Section 2.6.2) yet are found at the end of nucleosomal DNA with decreased supercoiling.

Why should the bases twist in this way?[16] The advantage of propeller twist is that it gives improved face-to-face contact between adjacent bases in the same strand and this leads to increased stacking stability in the double-helix. However, there is a penalty! The larger purine bases occupy the centre of the helix so that in alternating purine-pyrimidine sequences they overlap with neighbouring purines in the opposite strand. Consequently propeller twist causes a clash between such pairs in adjacent purines in

opposite strands. For pyrimidine (3→5')-purine steps, these purine-purine clashes take place in the minor groove where they involve guanine-N3 and –N2 and adenine-N3 atoms. For purine-(3'-5')-pyrimidine steps, they take place in the major groove between guanine-O^6 and adenine-N6 atoms (**Figure 2.21**). There are no such clashes for purine-purine and pyrimidine-pyrimidine sequences.

One of the consequences of these effects is that bends may occur at junctions between polyA tracts and mixed-sequence DNA as a result of propeller twist, base-pair inclination and base-stacking differences on two sides of the junction (see below).

[Figure 2.21 here]

Electrostatic Interactions Between Bases

There are two principal types of base-base interaction that drive the local variations in helix parameters described above and in **Figure 2.19a-c**. Firstly there are repulsive steric interactions between proximate bases and sugars. They are associated with steric interactions between thymine methyl groups, the guanine amino group, and the configuration of the step pyrimidine-purine (described as YR), purine-pyrimidine (described as RY) and RR/YY. Secondly there are π-π stacking interactions that are determined by the distribution of π-electron density above and below the planar bases.

Chris Hunter has identified four principal contributions to the energy of π-π interactions between DNA base-pairs.[17]

(1) van der Waals interactions (designated *vdW* and vary as r^{-6}).

(2) Electrostatic interactions between partial atomic charges (designated *atom-atom* and vary as r^{-1}).

(3) Electrostatic interactions between the charge distributions associated with the π-electron density above and below the plane of the bases (designated $\pi\sigma$-$\pi\sigma$ and vary approximately as r^{-5}).

(4) Electrostatic interactions between the charge distributions associated with the π-electron density and the partial atomic charges (designated as *atom-*$\pi\sigma$, this is the cross-term of (2) and (3) and varies as r^{-4}).

He has used these components to calculate the π-π interaction energies between pairs of stacked bases and applied the results to interpret the source of slide, roll and helical twist, of propeller twist, and of a range of other conformational preferences that are sequence-dependent. In addition, his calculations correlate very well with experimental observations on polymorphic forms of DNA. The main conclusions can broadly be summarised a follows:

- vdW-steric interactions are seen cross-strand at pyrimidine-purine (YR) and CX/XG steps and can be diminished by reducing propeller twist, reducing helical twist, or by positive slide or positive roll. They are seen as same-strand clashes between the thymine methyl group and the neighbouring 5'-sugar in AX/XT steps which are avoided by introducing negative propeller twist, reducing helical twist, or generating negative slide coupled with negative roll.

- Electrostatic interactions cause positive or negative slide with the sole exception of AA/TT. These slide effects are opposed by the hydrophobic effect which tends to force maximum base overlap and favours a zero-slide B-type conformation.
- Atom-atom interactions are most important for C•G base-pairs where there are large regions of charge and lead to strong conformational preferences for positive slide in CG steps and negative slide in GC steps (see **Table 2.4**). This leads poly(dCG) to adopt the Z-form left-handed duplex.
- Atom-πσ interactions lead to sequence-dependent effects which are repulsive in AX/XT, TX/XA and CX/XG steps where they can be reduced by negative propeller twist, by positive or negative slide, or by introducing buckle.
- πσ-πσ electrostatic interactions tend to be swamped by other effects and play a relatively minor role in sequence-dependent conformations.

In sequence-dependent structures, propeller twist is most marked for purines on opposing strands in successive base-pairs. The 'purine-purine' clash is much more pronounced for YR steps, where the clash is in the minor groove (**Figure 2.21a**), than for RY steps, where the clash is seen in the major groove (**Figure 2.21b**). While its origin was at first thought to result solely from van der Waals interactions, it seems now to be better explained by the total electrostatic interaction picture (see above).

Taken together, these sequence-dependent features suggest that DNA should most easily be unwound and/or unpaired in A-T rich sequences, which have only two hydrogen bonds per base-pair, and in pyrimidine-purine steps. It is noteworthy that the dinucleotide TpA satisfies both of these requirements and has been identified as the base

step that serves as a nucleus for DNA unwinding in many enzymatic reactions requiring strand separation.

Calladine's Rules

Notwithstanding the apparent success of the above calculations, the evidence from analyses of X-ray structures suggests that base step conformations are influenced by the nature of neighbouring steps. It follows that a better sequence-structure correlation is likely to emerge from examining each step in the context of its flankers: three successive base steps, or a tetrad of four successive base-pairs. However, until a majority of the 136 possible triads has been sampled by analysis of real structures, a set of empirical rules enunciated by Chris Calladine in 1982 will remain useful.[9]

Calladine observed that B-DNA structures respond to minimise the problems of sequence-dependent base clashes in four ways, which he articulated as follows:

- flatten the propeller twist locally for either or both base-pairs
- roll the base-pairs away from their clashing edges
- slide one or both of the base-pairs along their axis to push the purine away from the helix axis
- unwind the helix axis locally to diminish interstrand purine-purine overlap.

The relative motions required to achieve these effects are described by six parameters of which the most significant are ρ for roll, D_y for slide, and Ω for helix twist. These motions are illustrated for neighbouring G•C base-pairs (**Figure 2.19**).

In practice, the structures of crystalline oligomers have exhibited the following six types of conformational modulation which are sequence-dependent and which support these rules:

- the B-DNA helix axis need not be straight but can curve with a radius of 112 Å
- the twist angle, Ω is not constant at 36^{o} but can vary from 28^{o} to 43^{o}
- propeller twist averages -11^{o} for C•G pairs and $+17^{o}$ for A•T pairs
- base-pairs ‘roll' along their long axes to reduce clashing
- sugar pucker varies from C3'-*exo* to O4'-*endo* to C2'-*endo*
- there can be local improved overlap of bases by slide, as in d(TCG) where C-2 moves towards the helix axis to increase stacking with G-3.

The Calladine model is incomplete because it ignores such important factors as electrostatic interactions, hydrogen bonding, and hydration. For example, a major stabilizing influence proposed for the high propeller twist in sequences with consecutive adenines is the existence of cross-strand hydrogen bonding between adenine N-6 in one strand with thymine O-4 of the next base-pair in the opposite strand (see above).

Modulations of B-DNA structure which have been observed in the solid state have to some extent been mirrored by the results of solution studies for d(GCATGC) and d(CTGGATCCAG) obtained by a combination of NMR analysis and restrained molecular dynamics calculations. These oligomers have B-type structures which show clear, sequence-dependent variations in torsion angles and helix parameters. There is a strong curvature to the helix axis of the hexamer which results from large positive roll

angles at the pyrimidine-purine steps. The decamer has a straight central core but there are bends in the helix axis at the second (TpG) and eighth (CpA) steps which result from positive roll angles and large slide values.

Taken together, these X-ray and NMR analyses give good support for the general conclusion that minor groove clashes at pyrimidine-purine steps are twice as severe as major groove clashes at purine-pyrimidine steps. As a result, it is possible to calculate the behaviour of the helix twist angle, Ω, using sequence data only.

The Continuum of Right-handed DNA Conformations

The simple concept that the standard conformations for right-handed DNA represent discontinuous states, only stable in very different environments, has undergone marked revision. In addition to the range of conformations seen in crystal structures, CD and NMR analyses of solution structures have also undermined that naïve picture. In particular, CD studies have shown that there is a continuum of helix conformations in solution that is sequence-dependent while both CD analysis of the complete TFIIIA binding site of 54 bp and the crystal structure of a nonameric fragment from it have identified a conformer that is intermediate between the canonical A- and B-DNA forms. Crystallographic analysis of complexes between TATA-box binding protein (TBP) and DNA fragments containing TATA boxes has revealed a DNA structure that shares features of A- and B-DNA. In addition, A- and B-DNA polymorphs can coexist, as seen in the crystal structure of d(GGBrUABrUACC), and stable intermediates between the A- and B-DNA forms have been trapped in crystal structures.[18] By use of 13 separate structures of the hexamer duplex $[d(GGCGCC)]_2$ in different crystallographic

environments, P. Shing Ho and collaborators were able to map the transition from B-DNA to A-DNA.[19] Their analysis demonstrated that little correlation exists between helix type and base-pair inclination and that the single parameter with which to follow the B→A transition appears to be x-displacement (**Figure 2.19**).

Bending at helix junctions

Bent DNA was first identified as a result of modelling the junction between an A-type and a B-type helix. The best solution to this problem requires a bend of 26° in the helix axis in order to maintain full stacking of the bases. Bent DNA has gained support not only from NMR and CD studies on a DNA•RNA hybrid [poly(dG)•$(rC)_{11}$-$(dC)_{16}$], but also from studies on regular homopolymers which contain $(dA)_5$•$(dT)_5$ sections occurring in phase in each turn of a 10-fold or 11-fold helix. Moreover, bent DNA containing such dA•dT repeats has been investigated from a variety of natural sources.

It appears that bending of this sort happens at junctions between the stiff [dA•dT] helix and the regular B-helix (see above). In situations where such junctions occur every five bases and in an alternating sense, the net result is a progression of bends, which is equivalent to a continuous curve in the DNA.

2.3.2 Mismatched Base-pairs

The fidelity of transmission of the genetic code rests on the specific pairings of A•T and C•G bases. Consequently, if changes in shape result from base mismatches, such as A•G, they must be recognised and be repaired by enzymes with high efficiency (Section 8.11.6).

X-ray analysis of DNA fragments with potential mispairs cannot give any information about the transient occurrence of rare tautomeric forms at the instant of replication. However, it can define the structure of a DNA duplex which incorporates mismatched base-pairs and provide details of the hydrogen bonding scheme, the response of the duplex to the mismatch, and the influence of neighbouring sequence on the structure and stability of the mismatch, and the effect of global conformation.[20] All these are intended to provide clues about the ways in which mismatches might be recognised by the proteins that constitute repair systems. High-resolution NMR studies have extended the picture to solution conformations. The different types of base-pair mismatch can be grouped into **transition mismatches** which pair a purine with the wrong pyrimidine, and **transversion mismatches** which pair either two purines or two pyrimidines.

Transition Mismatches

The G•T base-pair has been observed in crystal structures for A-, B-, and Z-conformations of oligonucleotides. In every case it has been found to be a typical 'wobble' pair having *anti-anti* glycosylic bonds. The structure of the dodecamer, d(CGCGAATTTGCG), which has two G•T^9 mismatches, can be superimposed on that of the regular dodecamer and shows excellent correspondence of backbone atomic positions.

The A•C pair has been examined in the dodecamer d(CGCAAATTCGCG) and once again the two A^4•C mismatches are typical 'wobble' pairs, achieved by the protonation of adenine-N1 (**Figure 2.22**). It is notable that there is no significant

worsening of base-stacking and little perturbation of the helix conformation. However, it appears that no water molecules are bonded to these bases in the minor groove.

[Figure 2.22 here]

Transversion Mismatches

The G•A mismatch is the most thoroughly studied in solution and in the solid state and two different patterns have been found. Crystals of the dodecamer d(CGCGAATTAGCG) have an (*anti*)G•A(*syn*) mismatch with hydrogen bonds from Ade-N7 to Gua-N1 and from Ade-N6 to Gua-O6 (**Figure 2.23**) A similar (*anti*)I•A(*syn*) mismatch has been identified in a related dodecamer structure. Calculations on both of these mismatches suggest that they can be accommodated into a regular B-helix with minimal perturbation.

This work contrasts with both NMR and X-ray studies on d(CGAA**GA**TTGG) and NMR work on d(CG**A**GAATTC**G**CG) which have identified (*anti*)G•A(*anti*) pairings with two hydrogen bonds. The X-ray analysis of the dodecamer shows a typical B-helix with a broader minor groove and a changed pattern of hydration. This arises in part because the two mismatched G•A pairs are 2.0 Å wider (from $C^{1'}$ to $C^{1'}$) than a conventional Watson-Crick pair.

[Figure 2.23 here]

Insertion-deletion mispairs.

When one DNA strand has one nucleotide more than the other, the extra residue can either be accommodated in an intrastrand position or be forced into an extra-strand location. Tridecanucleotides containing an extra A, C or T residue have been examined in the crystalline solid and solution states. In one case, an extra A has been accommodated into the helix stack while in others a C or A is see to be extruded into an extrahelical, unstacked location.

In addition to such work on mismatched base-pairs, related investigations have made good progress into structural changes caused by covalent modification of DNA. On the one hand, crystal structures of DNA adducts with cisplatin have characterised its monofunctional linking to guanine sites in a B-DNA helix (Section 8.5.4) and, on the other, NMR studies of O^4-methylthymine residues and of thymine photodimers and psoralen:DNA photoproducts are advancing our understanding of the modifications to DNA structure that result from such lesions (Section 8.8.2). It seems likely that the range of patterns of recognition of structural abnormalities may be as wide as the range of enzymes available to repair them!

2.3.3 Unusual DNA structures

Since 1980, there has been a rapid expansion in our awareness of the heterogeneity of DNA structures which has resulted from a widening use of new analytical techniques notably structure-dependent nuclease action, structure-dependent chemical modification and physical analysis.[21] Unusual structures are generally sequence-specific, as we have already described for the A-B helix junction (Section 2.3.1). Some of them are also dependent on DNA supercoiling which provides the necessary driving energy for their

formation due to the release of torsional strain, as is particularly well defined for cruciform DNA. Consequently much use has been made of synthetic DNA both in short oligonucleotides and cloned into circular DNA plasmids where the effect of DNA supercoiling can be explored.

Curved DNA

The axial flexibility of DNA is one of the significant factors in DNA-protein interactions (Chapter 10).[22,23] DNA duplexes up to 150 bp long behave in solution as stiff, although not necessarily straight, rods. By contrast, many large DNA-protein complexes have DNA that is tightly bent. One of the best examples is the bending of DNA in the eukaryotic chromosome where 146 bp of DNA are wrapped around a protein core of histones (Section 10.6.1) to form nearly two complete turns on a left-handed superhelix with a radius of curvature of 43 Å. To achieve this, the major and minor grooves are compressed on the inside of the curve and stretched on its outside. At the same time, the helix axis must change direction.

DNA curvature has also been examined in kinetoplast DNA from trypanosomatids. It provides a source of **open** DNA minicircles whose curvature is sequence dependent rather than being enforced by covalent closure of the circles. Such circles can be examined by electron microscopy and have 360^{o} curvature for about 200 bp. Such kinetoplast DNA has short adenine tracts spaced at 10 bp intervals by general sequence. This fact led to solution studies on synthetic oligomers with repeated sets of four $CA_{5\text{-}6}T$ sequences spaced by 2-3 bp. These behave as though they have a 20^{o}-25^{o} bend for each repeat which led to the simple idea that DNA bending is an inherent

property of poly(dA) tracts (**Figure 2.24a**). In conflict with this idea, poly(dA) tracts in the crystal structures of several oligonucleotides are seen to be straight. What then is the real origin of DNA curvature?[24,25]

[Figure 2.24 here]

Richard Dickerson has examined helix bending in a range of B-form crystal structures of oligonucleotides containing poly(dA) tracts and has concluded that poly(dA) tracts are straight and not bent and that regions of A•T base-pairs exhibit a narrow minor groove, large propeller twist, and a spine of hydration in the minor groove. He argues that DNA curvature results from the direct combination of two general features of DNA structure:

- general sequence DNA writhes
- poly(dA) tracts are straight

Studies on the hydrodynamic properties of DNA show that general-sequence DNA migrates through gels more slowly than expected which is because the DNA helix occupies a cylindrical volume that has a larger diameter than that of a simple B-helix (Section 11.4.3). This phenomenon is a result of **DNA writhing**, which involves a continuously curved distortion of the helix axis to generate a spiral form and is nicely illustrated by the extension of a coiled telephone wire (**Figure 2.24b**). It follows that the repeated alternation of straight A-tracts with short sections of general sequence, each having half of a writhing turn, will generate curved DNA (**Figure 2.24c**). A detailed structural analysis of this explanation says that curvature of B-DNA involves rolling of

base-pairs, compresses the major groove (which corresponds to positive roll), has a sequence-determined continuum in the bending behaviour, and shows anisotropy of flexible bending.

DNA bending

Such intrinsic, sequence-dependent curvature must be distinguished from the bending of DNA which results from the application of an external force. Dickerson has also examined the bending of the DNA helix that occurs in many crystal structures of the B-form. It is associated with the step from a G•C to an A•T base-pair and results from rolling one base-pair over the next along their long axes in a direction that compresses the major groove (**Figure 2.18c**). He suggests that this junction is a flexible hinge that is capable of bending or not bending. Such 'facultative bending' responds to the influence of local forces, typically interactions with other macromolecules, for example control proteins or a nucleosome core. By contrast, poly(dA) tracts are known to resist bending in nucleosome reconstitution experiments. It can thus be seen that sequence-dependent variation in **DNA bendability** is an important factor in DNA recognition by proteins

One important conclusion emerges: DNA has evolved conformationally to interact with other macromolecules. A free, linear DNA helix in solution may, in fact, be the least biologically relevant state of all.[24]

Slipped structures have been postulated to occur at direct repeat sequences, and they have been found upstream of important regulatory sites. The structures described (**Figure 2.24a**) are consistent with the pattern of cleavage by single-strand nucleases but otherwise are not well characterized.

Purine-pyrimidine tracts manifest an unusual structure at low temperature with a long-range, sequence-dependent single base shift in base-pairing in the major groove. For the dodecamer d(ACCGGCGCCACA)•d(TGTGGCGCCGGT) the bases in the $d(CA)_n$ tract have high propeller twist (-32^o) and are so strongly tilted in the 3'-direction that there is disruption of Watson-Crick pairing in the major groove and formation of interactions with the 5'-neighbour of the complementary base. This alteration propagates along the B-form helix for at least half a turn with a domino-like motion. As a result, the DNA structure is normal when viewed from the minor groove and mismatched when seen from the major groove. Since $(CA)_n$ tracts are involved both in recombination and in transcription, this new recognition pattern has to be considered in the analysis of the various processes involved with reading of genetic information.

Anisomorphic DNA is the description given to DNA conformations associated with direct repair, DR2, sequences at 'joint regions' in viral DNA, which are known to have unusual chemical and physical properties. The two complementary strands have different structures and this leads to structural aberrations at the centre of the tandem sequences that can be seen under conditions of torsional stress induced by negative supercoiling.

Hairpin loops are formed by oligonucleotide single strands which have a segment of inverted complementary sequence. For example, the 16-mer d(CGCGCGTTTTCGCGCG) has a hexamer repeat and its crystal structure shows a hairpin with a loop of four Ts and a Z-DNA hexamer stem (**Figure 2.25a**). When such inverted sequences are located in a DNA duplex, the conditions exist for formation of a cruciform.

Cruciforms involve intra-strand base-pairing and generate two stems and two hairpin loops from a single unwound duplex region.[26] The inverted sequence repeats are known as **palindromes**, which have a given DNA duplex sequence followed after a short break by the same duplex sequence in the opposite direction. This is illustrated for a segment of the bacterial plasmid pBR 322 (**Figure 2.25b**), where a palindrome of two undecamer sequences exists.

X-ray, NMR and sedimentation studies of such stem-loop structures show that the four arms are aligned in pairs to give an oblique X structure with continuity of base-stacking and helical axes across the junctions (Section 6.8.1). Also, the loops have an optimum size of from four to six bases. Residues in the loops are sensitive to single-strand nucleases, such as S1 and P1, and especially to chemical reagents such as bromoacetaldehyde, osmium tetroxide, bisulfite, and glyoxal (Chapter 8). In addition, the junctions are cleavage sites for yeast resolvase and for T4 endonuclease VII.

[Figure 2.25 here]

David Lilley has shown that the formation of two such loops requires the unpairing and unstacking of three or more base-pairs and so will be thermodynamically unstable compared to the corresponding single helix.[27] While there can be some stacking of bases in the loops, the adverse energy of formation of a single cruciform has been calculated to be some 75 kJ mol^{-1}. In experiments on cruciforms using closed circular superhelical DNA, this energy can be provided by the release of strain energy in the form of negative supercoiling (next Section) and is directly related to the length of the arms of the cruciform: the formation of an arm of 10.5 bp unwinds the supercoil by a single turn.

There is also a kinetic barrier to cruciform formation and Lilley has suggested two mechanisms that have clearly distinct physical parameters and may be sequence-dependent. The faster process for cruciform formation, the S-pathway, has $\Delta G^{\ddagger}$ of about 100 kJ mol^{-1} with a small positive entropy of activation. This more common pathway is typified by the behaviour of plasmid pIRbke8. Following the formation of a relatively small unpaired region, a proto-cruciform intermediate is produced which then grows to equilibrium size by branch migration through the four-way junction (**Figure 2.26**). The slower mechanism, the C-pathway, involves the formation of a large bubble followed by its condensation to give the fully-developed cruciform. This behaviour explains the data for the pColl315 plasmid whose cruciform kinetics show $\Delta G^{\ddagger}$ about 180 kJ mol^{-1} with a large entropy of activation.

[Figure 2.26 here]

Such extrusion of cruciforms provides the most complete example of the characterisation of unusual DNA structures by combined chemical, enzymatic, kinetic and spectroscopic techniques. However, it is not clear whether cruciforms have any role *in vivo*. One reason may simply be that intracellular superhelical densities may be too low to cause extrusion of inverted repeat sequences. Equally, the kinetics of the process may also be too slow to be of physiological significance. However, cruciforms are formally equivalent to Holliday junctions and these four-way junctions involve two DNA duplexes that are formed during homologous recombination (Section 6.8).

Several X-ray crystallographic studies have provided a detailed picture of the three-dimensional structure of the Holliday junction.[28] Interestingly, DNA decamers with sequences CCGGGACCGG, CCGGTACCGG, and TCGGTACCGA fold into four-way junctions instead of adopting the expected B-form double helical geometry (**Figure 2.27**). The trinucleotide ACC (underlined) forms the core of the junction and its 3'-C•G base-pairs helps to stabilize the arrangement by engaging in direct and water-mediated hydrogen bonds to phosphate groups at the strand crossover. The four strands exhibit a stacked-X conformation whereby the two interconnected duplexes form coaxially stacked arms that cross at an angle of ca. 40°. Stable Holliday junctions were also observed with DNA decamers that featured an AC(Me^5C) trinucleotide core and the trinucleotide AGC when covalently intercalated by psoralen.

[Figure 2.27 here]

Role of Metal Ions

NMR in solution, X-ray crystallography and computational simulations (molecular dynamics, MD) have all shed light on the locations of metal cations surrounding nucleic acid molecules.[29] But whilst ions are often visible in structures of nucleic acids, it is not straightforward to determine how they affect the structure. The question of whether cations can assume specific roles in the control of DNA duplex conformation has stirred up controversy in recent years. Some in the field, notably Nicholas Hud and Loren Williams, believe that cation localization within the grooves of DNA represents a

significant factor in sequence-specific helical structure. By contrast, probably a majority of those studying the structure of DNA is of the opinion that the specific sequence dictates local DNA conformation and thus binding of metal cations. According to this second view, metal ions can bind to DNA in a sequence-specific manner and in turn modulate the local structure, but ions should not be considered the single most significant driving force of a number of DNA conformational phenomena.

To study the possible effects of metal ions on DNA conformation, all sequences can be divided into three principal groups: **A-tracts**, **G-tracts** and generic DNA, the latter representing the vast majority of DNA sequences.[30] A-tracts have an unusually narrow minor groove, are straight and have high base-pair propeller twist (see also **Figures 2.18, 2.19**). G-tracts have a propensity to undergo the B-form → A-form transition at increased ionic strength. The proponents of the 'ions are dominant' model believe that the DNA grooves are flexible ionophores and that DNA duplex structure is modulated by a tug of war between the two grooves for cation localization. They argue that the duplex geometry adopted by A-tracts (referred to as B*-DNA, **Figure 2.28a**) is due to ion localization in the minor groove as a result of the highly negative electrostatic potentials there. Conversely, G-tract DNA exhibits a highly negative electrostatic potential in the major groove (**Figure 2.28b**), leading to preferred localization of cations there and consequently a collapse of the DNA around the ions. Generic DNA on the other hand would have a more balanced occupation of its major and minor grooves by cations, consistent with a more or less canonical B-form geometry. By contrast, those who emphasize the dominating role of sequence in the control of DNA conformation

argue that it is the sequence that shapes the DNA in the first place and that the narrow minor groove of A-tract or B*-DNA is narrow even before ions settle in the groove.

[Figure 2.28 here]

Therefore, it is difficult to settle the issue of the relative importance of sequence and ions in governing DNA duplex conformation, and no single experimental or, certainly, theoretical method alone will provide a definitive answer. Although the 'ions first' hypothesis has a number of attractive features – i.e. it provides a link between sequence-specific cation localization and sequence-directed curvature of DNA (Section 2.3.3) – it cannot be overlooked that high-resolution crystal structures of oligodeoxynucleotides containing A-tracts have shown no variation of groove width as a consequence of different types and concentrations of alkali metal ions present in the crystallizations. Moreover, MD simulations of A-tract DNA in the presence of different classes and varying localizations of metal cations have not provided a picture that is consistent with a crucial role for metal ions with regard to the structure of duplex DNA. Thus, there will undoubtedly be more studies directed at a refined understanding of the relative importance of sequence and cation coordination in governing the structure of double helical DNA.

2.3.4 B-Z Junctions and B-Z Transitions

Segments of left-handed Z-DNA can exist in a single duplex in continuity with segments of right-handed B-DNA. This phenomenon has been observed both *in vitro* and *in vivo*. Because the backbone chains of these polymorphs run in opposite directions, (↓↑ and ↑↓) respectively (Section 2.2.5), there has to be a transitional region between two such

segments, and this boundary is known as a **B-Z junction**. Such structures are polymorphic and sequence-specific and six features had been described:

- B-Z junction can be as small as 3 bp
- at least one base-pair has neither the B- nor the Z-conformation
- hydrogen bonds between the base-pairs are intact below 50^{o}C
- chemical reagents specific for single-stranded DNA (chloroacetaldehyde, bromoacetaldehyde, and glyoxal (Section 8.5.3) show high reactivity with the junction bases
- junctions are sites for enhanced intercalation for psoralens (Section 8.8.2)
- junctions are neither strongly bent nor particularly flexible.

This conformational B-Z transition between the right- and left-handed helices has a high energy of activation (about 90 kJ mol^{-1}) but is practically independent of temperature (ΔG^{o} about 0 kJ mol^{-1}) (Section 2.5.3). Thus the B-Z transition is co-operative and propagates readily along the helix chains.

In the absence of structural data at high resolution, two different models had been suggested to explain the conformational switch that has to occur as a B-Z junction migrates, rather like a bubble, along a double helix. In the first, the bases unpair, guanine flips into the *syn* conformation, the entire deoxycytidine undergoes a conformational switch, and the base-pairs reform their hydrogen bonds. This model appears to be at variance with NMR studies that suggest the bases remain paired because their imino-protons do not become free to exchange with solvent water. In the second model the

backbone is stretched until one base-pair has sufficient room to rotate 180^{o} about its glycosyl bonds (**tip**, as shown in **Figure 2.19a**), and the bases re-stack. However, one might expect this 'expand-rotate-collapse' process to be impeded by linking bulky molecules to the edge of the base-pairs. Yet bonding *N*-acetoxy-*N*-acetyl-2-aminofluorene to guanine actually facilitates the B-Z transition. Thus, the dynamics of the B-Z transition poses a major conformational problem, and this has sometimes been called the chain-sense paradox.

In addition Ansevin and Wang suggested an alternative zig-zag model for the left-handed double-helical form of DNA that avoids this paradox and is accessible from B-DNA by simple untwisting. Their **W-DNA** has a Watson-Crick chain sense (↑↓) like B-DNA but similar glycosyl geometry to that of Z-DNA. It has reversed sugar puckers, C3'-*endo* at cytosine and C2'-*endo* at guanine, while in both W- and Z-DNA the minor groove is deep and the major groove broad and very shallow. In addition, this W-model explains (1) the incompatibility of poly(dA-dT)•poly(dA-dT) with a left-handed state, (2) the very slow rate of exchange of hydrogens in the 2-NH_2 group of guanine in left-handed DNA, and (3) the incompatibility of left-handed helix with replacement of O_R oxygens by a methyl group. They argue that Z-DNA has a lower energy than W-DNA and so is adopted in crystals of short oligonucleotides but it may be conformationally inaccessible to longer stretches of DNA in solution.

Finally, more than 25 years after the discovery of Z-DNA, a crystal structure of a B-Z junction has solved the mystery of how DNA switches from the right-handed low energy to the left-handed high energy form.[31] The junction was trapped in the crystal by stabilizing the Z-DNA portion at one end of a 15-base-pair segment with a Z-DNA

binding protein, with the rest of the DNA assuming the B-form geometry. Continuous stacking of bases between B-DNA and Z-DNA is found with the breaking of one base pair at the junction and the two bases extruded from the helix on either side. A sharp turn accommodates the reversal in the backbone direction and at the junction the DNA is bent by ca. 10° and the helical axes of the B- and Z-form duplexes are displaced from each other by ca. 5 Å.

2.3.5 Circular DNA and Supercoiling

The replicative form of bacteriophage ϕX174 DNA was found to be a double-stranded closed circle. It was later shown that bacterial DNA exists as closed circular duplexes, that DNA viruses have either single - or double-helical circular DNA, and that RNA viroids have circular single-stranded RNA as their genomic material. Plasmid DNAs also exist as small, closed circular duplexes.

Topologically unconstrained dsDNA in its linear, relaxed state is either biologically inactive or displays reduced activity in key processes such as recombination, replication, or transcription. It follows that topological changes associated with the constraints of circularisation of dsDNA have a profound biological significance.[32] While such circularisation can be achieved directly by covalent closure, the same effect can be achieved for eukaryotic DNA as a result of holding DNA loops together by means of a protein scaffold.

The molecular topology of closed circular DNA was described by Vinograd in 1965 and is especially associated with the phenomenon of superhelical DNA, which is also called supercoiled or supertwisted DNA. Vinograd's basic observation was that

when a planar, relaxed circle of DNA is strained by changing the pitch of its helical turns, it relieves this torsional strain by winding around itself to form a superhelix whose axis is a diameter of the original circle.

This behaviour is most directly observed by following the sedimentation of negatively supercoiled DNA as the pitch of its helix is changed by intercalating a drug, typically ethidium bromide (Section 9.6). Intercalation is the process of slotting the planar drug molecules between adjacent base-pairs in the helix. For each ethidium molecule intercalated into the helix there is an increase in rise of about 3.4 Å and a linked decrease of about 36° in twist (**Figure 2.19c,d**). The DNA helix responds first by reducing the number of negative, right-handed supercoils until it is fully relaxed and then by increasing the number of positive, left-handed supercoils. As this happens, the sedimentation coefficient of the DNA first decreases, reaching a minimum when fully relaxed, and then increases as it becomes positively supercoiled. As a control process, the same circular DNA can be nicked in one strand to make it fully relaxed. The result is that it now shows a low sedimentation coefficient at all concentrations of the intercalator species (**Figure 2.29**) (Section 11.4.1).

Vinograd showed that the topological state of these covalently closed circles can be defined by three parameters and that the fundamental topological property is **linkage**. The **topological winding number**, T_w, is the number of right-handed helical turns in the relaxed, planar DNA circle and the **writhing number**, W_r, gives the number of left-handed crossovers in the supercoil. The sum of these two is the **linking number**, L_k, which is the number of times one strand of the helix winds around the other (clockwise is

+ve) when the circle is constrained to lie in a plane. The simple equation is $L_k = T_w + W_r$.

[Figure 2.29 here]

Such behaviour can be illustrated simply (**Figure 2.30**) for a relaxed closed circle with 20 helical turns $T_w = 20$, $L_k = 20$, $W_r = 0$. One strand is now cut, unwound two turns, and resealed to give $L_k = T_w = 18$. This circle is thus underwound by two turns. To restore fully the normal B-DNA base-pairing and base-stacking, the circle needs to gain two right-handed helical turns, $\Delta T_w = +2$ to give $T_w = 20$. Since the DNA circles have remained closed and the linking number stays at 18, the formation of the right-handed helical turns is balanced by the creation of one right-handed supercoil, making $W_r = -2$.

[Figure 2.30 here]

The behaviour of a supercoil can be modelled using a length of rubber tubing. The ends are first held together to form a relaxed closed circle. If the end in your right hand is given one turn clockwise (right-handed twist) and the other end is given one turn in the opposite sense, the tube will relieve this strain by forming one left-handed supercoil. This is equivalent to unwinding the DNA helix by two turns, which generates one positive supercoil (four turns generate two supercoils, and so on). This model shows the relationship: two turns equals one supercoil.

In practice it is sometimes useful to describe the degree of supercoiling using the **super-helical density**, $\sigma = W_r/T_w$, which is close to the number of superhelical turns per 10 bp and is typically around 0.06 for superhelical DNA from cells and virions. The **energy of supercoiling** is a quadratic function of the density of supercoils as described by the equation:

$$\Delta G_s = 1050 \times \frac{RT}{N} \times \Delta L_k^2 \quad \text{kJ mol}^{-1}$$

where R is the gas constant, T is the absolute temperature, and N is the number of base-pairs.

B-Z transitions are especially important for supercoiling since the conversion of one right-handed B-turn into a left-handed Z-turn causes a change in T_w of –2. This must be complemented by $\Delta W_r + 2$ through the formation of one left-handed superturn.

Enzymology of DNA Supercoiling

DNA topoisomers are circular molecules which have identical sequences and differ only in their linking number. A group of enzymes, discovered by Jim Wang, can change that linking number.[33] They fall into two classes: Class I topoisomerases effect integral changes in the linking number, $\Delta L_k = n$, while Class II enzymes interconvert topoisomers with a step rise of $\Delta L_k = \pm 2n$. **Topoisomerase** I enzymes use a 'nick-swivel-close' mechanism to operate on supercoiled DNA. They break a phosphate diester linkage, hold its ends, and reseal them after allowing exothermic (i.e. passive) free rotation of the other strand. Such enzymes from eukaryotes can operate on either left- or right-handed

supercoils while prokaryotic enzymes only work on negative supercoils. The products of topoisomerase I action on plasmid DNA can be observed by gel electrophoresis, and show a ladder of bands, each corresponding to unit change in W_r as the supercoils are unwound, half at a time (**Figure 2.31**).

[Figure 2.31 here]

By contrast, Class II topoisomerases use a 'double-strand passage' mechanism to effect unit change in the number of supercoils, $\Delta W_r = \pm 2$, and such prokaryotic enzymes can drive the endothermic supercoiling of DNA by coupling the reaction to hydrolysis of ATP. These topoisomerases cleave two phosphate esters to produce an enzyme-bridged gap in both strands. The other DNA duplex is passed through the gap (using energy provided by hydrolysis of ATP), and the gap is resealed.[34] DNA gyrase from *E. coli* is a special example of the Class II enzyme. It is an A_2B_2 tetramer with the energy-free topoisomerase activity of the A subunit being inhibited by quinolone antibiotics such as nalidixic acid. The energy-transducing activity of the B subunit can be inhibited by novobiocin and other coumarin antibiotics. We should point out that such topiosomerases also operate on linear DNA that is torsionally stressed by other processes, most notably at the replication fork in eukaryotic DNA.

Supercoiling is important for a growing range of enzymes as illustrated by two examples. RNA polymerase *in vitro* appears to work ten times faster on supercoiled DNA $\sigma = 0.06$, than on relaxed DNA, and this phenomenon appears to be related to the enhanced binding of the polymerase to the promoter sequence. Secondly, the tyrT

promoter in *E. coli* is expressed *in vitro* at least 100 times stronger for supercoiled than for relaxed DNA and this behaviour seems linked to 'preactivation' of the DNA promoter region by negative supercoiling.[35]

Catenated and knotted DNA circles

While Type II topoisomerases usually only effect passage of a duplex from the same molecule through the separated double strands, they can also manipulate a duplex from a second molecule. As a result, two different DNA circles can be interlinked with the formation of a **catenane** (**Figure 2.32**). Such catenanes have been identified by electron microscopy and can be artificially generated in high yield from mammalian mitochondria. **Knotted DNA** circles are another unusual topoisomer species which are also formed by intramolecular double-strand passage from an incompletely unwound duplex (**Figure 2.32**).

[Figure 2.32 here]

2.3.6 Triple-stranded DNA

Triple helices were first observed for oligoribonucleotides in 1957. A decade later, the same phenomenon was observed for poly(dCT) binding to poly(dGA)•poly(dCT) and for poly(dG) binding to poly(dG)•poly(dC). Oligonucleotides can bind in the major groove of B-form DNA by forming Hoogsteen or reversed Hoogsteen hydrogen bonds using N-7 of the purine bases of the Watson-Crick base-pairs (**Figures 2.33 and 2.34**).[36] The resulting base-triplets form the core of a triple helix. In theory, G can form a base-triple

with a C•G pair and A with a T•A pair (**Figure 2.33b**), but the only combinations that have isomorphous location of their C-1' atoms are the two triplets TxA•T and C•GxC$^+$ (**Figure 2.33a**), where C$^+$ is the N-3 protonated form of cytosine. This means that the three strands of triple-helical DNA are normally two homopyrimidines and one homopurine (**Figure 2.33c**). However, despite the backbone distortion that must result from the heteromorphism of other base triplet combinations, oligonucleotides containing G and T, G and A, or G, T and C have been shown to form helices.

Intermolecular triple helices are now well characterised for short oligonucleotides binding in the groove of a longer DNA duplex,[37,38] H-DNA provides an example of an intramolecular triple helix because it has a mirror-repeat sequence relating homopurine and homopyrimidine tracts in a circular double-stranded DNA molecule and triplex formation is driven by supercoiling (**Figure 2.33d**).

[Figure 2.33 here]

Several studies on third-strand binding to a homopurine·homopyrimidine duplex have established the following features:

- a third homopyrimidine strand binds parallel to the homopurine strand using Hoogsteen hydrogen bonds (i.e. the homopyrimidine strands are antiparallel)
- a third homopurine strand binds antiparallel to the original homopurine strand using reversed Hoogsteen hydrogen bonds
- the bases in the third strand have a regular *anti* conformation of the glycosylic bond

- synthetic oligodeoxynucleotides having an α-glycosylic linkage also bind as a third strand, parallel for poly(d-α-T) and antiparallel for poly(d-α-TC).

[Figure 2.34 here]

Triple helices are less stable than duplexes. Thermodynamic parameters have been obtained from melting curves, from kinetics, and from the use of differential scanning calorimetry (DSC) (Section 11.4.4). With this last technique, values of $\Delta H^o = 22 \pm 2$ kJ mol^{-1} and $\Delta S^o = 70 \pm 7$ J mol^{-1} K^{-1} have been found for d($C^+TTC^+C^+TC^+C^+TC^+T$). The pK_a value of cytidine in isolation is 4.3 but it is higher in oligonucleotides because of their polyanionic phosphate backbone. So it is to be expected that the stability of triple helices is seen to decrease as the pH rises above 5.

Triple helix stability can be enhanced by the use of modified nucleotides. 5-Methylcytosine increases stability at neutral pH, probably by a hydrophobic effect, and 5-bromouracil can usefully replace thymine. Oligoribonucleotides bind more strongly than do deoxyribonucleotides and 2'-*O*-methylribonucleotides bind even better. Finally, Hélène has shown that at the attachment of an intercalating agent to the 5'- (or 3'-end) of the third strand can greatly enhance the stability of the triple-stranded helix.

The major application of triple helices relates to the specificity of the interaction between the single strand and a much larger DNA duplex. This is because homopyrimidines have been identified as potential vehicles for the sequence-specific delivery of agents that can modify DNA and thereby control genes. The DNA of the bacterium *E. coli* has 4.5 Mbp, so the minimum number of base-pairs needed to define a

unique sequence in its genome is 11 bp (i.e. 4^{11} = 4 194 304 assuming a statistically random distribution of the four bases). The corresponding number for the human genome is about 17 bp. Thus a synthetic 17-mer could be expected to identify and bind to a unique human DNA target and thus deliver a lethal agent to a specific sequence of DNA. In practice, the energetics of mismatched base-triples is complex and depends on nearest neighbours, metal ions, and other parameters. However, a value of about 1.5 kJ mol^{-1} per mismatch seems to fit much of the data and suggests that the specificity of triple helices is at least as good as that of double-helical complexes.

H-DNA

A new polymorph of DNA was discovered in 1985 within a sequence of $d(A\text{-}G)_{16}$ in the polypurine strand of a recombinant plasmid pEJ4. Its requirement for protons led to the name H-DNA (half of its C residues are protonated, so the transition depends on acid pH as well as on a degree of negative supercoiling). Probes for single-stranded regions of DNA (especially osmium tetroxide:pyridine (Section 8.3) and nuclease P1 cleavage) were used to identify specific sites and provide experimental support for the model advanced earlier for a triple helical H-DNA (**Figure 2.33d**). This has a Watson-Crick duplex which extends to the centre of the $(dT\text{-}dC)_n$• $(dG\text{-}dA)_n$ tract and the second half of the homopyrimidine tract then folds back on itself, antiparallel to the first half and winding down the major groove of the helix. The second half of the polypurine tract also folds back, probably in an unstructured single-stranded form. The energetics of nucleation of H-DNA suggests that it requires at least 15 bp for stability and the consequent loss of twist makes H-DNA favoured by negative supercoiling.

Although antibodies have been raised to detect triple-stranded structures, no evidence has yet been found for their natural existence in cells *in vivo*.

2.3.7 Other Non-canonical DNA Structures

Four-stranded Motifs

Both G-rich and C-rich DNA sequences have been found to adopt four-stranded motifs, also called **tetraplexes** or **quadruplexes** (see also Section 9.10.2). Sequences containing G-rich and C-rich strands are found at the telomeric ends of chromosomes (see Section 6.4.5), and such sequences are of fundamental importance in protecting the cell from recombination and degradation. It is known that when DNA containing a palindromic sequence of bases is subjected to supercoiling stress (see Section 2.3.5), a cruciform can be extruded (**Figure 2.26**). If the tips of the cruciform extrusions contain C residues in one limb and G residues in the other, G-rich and C-rich quadruplexes can be formed by combining two cruciforms of this type. So it is possible that formation of four-stranded G-rich motifs and C-rich motifs provides the physical basis for identical DNA sequences to bind together, i.e. during meiosis when identical chromosomes line up with each other. Proteins that bind to G-rich quadruplexes have been identified and it is unlikely that the C-rich motif is stable at neutral pH without also binding a protein factor, since the motif is held together by hemiprotonated C•C^+ base-pairs.

The structures of G-rich and C-rich quadruplexes are fundamentally different. In the case of four-stranded G-rich motifs, guanines join together via cyclic hydrogen bonding that involves four guanines at each level (often called a **G-tetrad** or **G-quartet**) (**Figure 2.35**). Each G base is engaged in four hydrogen bonds via its Hoogsteen and

Watson-Crick faces, such that guanines are related by a four-fold rotation axis and are nearly co-planar. In this way, each guanine directs its O-6 carbonyl oxygen into the central core of the tetrad. Although it had been found as early as 1910 that concentrated solutions of guanylic acid were unusually viscous and formed a clear gel upon cooling (see Section 1.3) and Gellert and Davies had described a four-stranded helix for guanylic acid based on fibre diffraction experiments more than forty years ago, detailed three-dimensional structures for G-tetrads have only emerged recently. Dinshaw Patel and coworkers determined the structure of the tetraplex adopted by the human telomeric repeat $d(AG_3(T_2AG_3)_3)$ in the presence of sodium ions by solution NMR. Under these conditions the 22mer adopts a four-stranded motif with three G_4 layers and lateral and diagonal loops. The four strands alternate between parallel and antiparallel orientations and G residues in adjacent layers alternate between the *syn* and *anti* conformations. A similar arrangement had also been found for an intramolecular quadruplex formed by the G_4T_4-repeat sequence from *Oxytricha nova*.

[Figure 2.35 here]

On the other hand, by use of X-ray crystallography it was found that the DNA sequences TG_4T (4x; Na^+ form), TAGGGTTAGGGT (2x; K^+ form) and the above 22mer $AG_3(T_2AG_3)_3$ (intramolecular; K^+ form, **Figure 2.35**) all adopted quadruplexes with parallel orientations of strands.[40] Most of the deoxyguanosine sugars exhibit the C2'-*endo* pucker. However, in contrast to the above parallel/antiparallel-type quadruplex, the glycosylic bonds in the four-stranded motifs with parallel orientation of strands adopt

exclusively the *anti* conformation. The local G-quartet rise is about 3.13 Å and the four strands writhe in a right-handed fashion with an average twist of around 30° between adjacent layers. Therefore, the G-rich quadruplex is extensively stabilized by π–stacking interactions between layers of guanines. Potassium ions are trapped in the core between stacked G-quartets, spaced at ca. 2.7 Å from each of a total of eight O-6 carbonyl groups (**Figure 9.2**).

A major difference between the four-stranded motifs formed by G-rich and C-rich sequences is that the former are stable at neutral pH whereas the latter require protonation of half the cytosine residues and hence are stable only at lower values of pH. Maurice Guéron's laboratory provided the initial structure of the C-rich motif a decade ago, determined by solution NMR methods.[41] They termed it intercalation or **i-motif** for the peculiar four-stranded arrangement involving two parallel intercalating duplexes, each held together by C•C$^+$ base-pairs (**Figure 2.36**). The two duplexes are intercalated with opposite polarity and a gentle right-handed helical twist (12-20°, the rise is 6.2 Å) between covalently linked residues gives the C-rich quadruplex a quasi two-dimensional form. The structure has two broad and two narrow grooves. In the latter, the antiparallel backbone pairs within intercalated duplexes are in van der Waals contact.

[Figure 2.36 here]

Several crystal structures of C-rich sequences, such as d(CCCC), d(CCCT), d(TCCCCC), d(TCC), d(CCCAAT) and d(TAACCC), have provided details of the conformation, stabilization and hydration of the i-motif.[42] One surprising finding is the absence of effective stacking between the cytosine rings of adjacent hemiprotonated C•C$^+$

base-pairs from intercalated duplexes, an obvious difference to the above G-rich quadruplex. However, a systematic base-on-deoxyribose stacking pattern as well as intracytidine C-H···O hydrogen bonds may partially compensate for the lack of effective base-base-stacking between layers of cytosines. The most unusual feature of the i-motif is a systematic, potentially stabilizing C–H···O hydrogen bonding network between the C2'-*endo* puckered deoxyribose sugars of antiparallel backbones.

The Hoogsteen duplex

Virtually all nucleic acid duplexes studied in the last twenty years contain either all G•C base-pairs or A•T base-pairs flanked by G•C base-pairs. Very little 3D-structural work has been carried out on AT-rich sequences, although the functional relevance of such sequences is well known. For example, the promoters of many eukaryotic structural genes contain stretches composed exclusively of A•T base-pairs. Further, coding sequences in the yeast genome tend to be clustered with AT-rich sequences separating them and AT-rich sequences are common in transposable elements (see Section 6.8.3). Crystal structures of TATA boxes bound to the TATA-box binding protein revealed highly distorted B-form conformations of the DNA. Fibre diffraction studies of AT-rich sequences provided indications for considerable structural polymorphism. By contrast, the mostly canonical geometries observed for A•T paired regions in the structures of oligonucleotide duplexes may have resulted from the constraints exerted by the G•C base-pair clamps at both ends.

A new crystal structure determined by Juan Subirana and coworkers of the alternating hexamer d(ATATAT) raises interesting questions with regard to the existence

of double-stranded DNA species that lack base-pairing of the Watson-Crick type and the possible biological relevance of such alternative DNA conformations.[43] In the crystal structure, but apparently not in solution, two hexameric fragments adopt antiparallel orientation with Hoogsteen pairing between adenine and thymine (**Figure 2.37**). The Hoogsteen duplex features an average of 10.6 base-pairs per turn, similar to B-DNA, and all sugars adopt the C2'-*endo* pucker. The diameter of the Hoogsteen duplex is also similar to B-form DNA and the minor groove widths of the two duplexes differ only marginally. A unique characteristic of the Hoogsteen duplex is the *syn* conformation of purine nucleosides. This arrangement generates a pattern of hydrogen bond donors and acceptors in the major and minor grooves that differs between B-DNA and Hoogsteen DNA. It also confers on the latter a less electronegative environment in the minor groove that may lead to preferred interactions with relatively hydrophobic groups at that site. Hoogsteen DNA also differs clearly from triple-stranded arrangements (Section 2.3.6). Thus, in the TxA•T triplex, adenine is always in the *anti* conformation. Moreover, T and A form the Hoogsteen pair in a parallel orientation while in the case of an antiparallel orientation (**Figure 2.33a**), base-pairing between A and T is of the reverse Hoogsteen type (**Figure 2.33b**). Furthermore, in the Ax(A•T) triplex, third-strand adenines always base-pair with adenine of the Watson-Crick base-pair (**Figure 2.33b**). Therefore, the antiparallel Hoogsteen DNA duplex found for d(ATATAT) is not simply a component of the triple helical motifs.

[Figure 2.37 here]

2.4 Structures of RNA Species

As with DNA, studies on RNA structure began with its primary structure. This quest was pursued in parallel with that of DNA, but had to deal with the extra complexity of the 2'-hydroxyl group in ribonucleosides. Today, we recognise also that RNA has greater structural versatility than DNA in the variety of its species, in its diversity of conformations, and in its chemical reactivity. Different natural RNAs can either form long, double-stranded structures or adopt a globular shape composed of short duplex domains connected by single-stranded segments. Watson-Crick base-pairing seems to be the norm, though tRNA structures have provided a rich source of unusual base-pairs and base-triplets (Section 7.1.2). In general, it is now possible to predict double-helical sections by computer analysis of primary sequence data, and this technique has been used extensively to identify secondary structural components of ribosomal RNA and viral RNA species. In this section, we shall focus attention mainly on regular RNA secondary structure.

2.4.1 Primary Structure of RNA

The first degradation studies of RNA using mild alkaline hydrolysis gave a mixture of mononucleotides, originally thought to have only four components – one for each base, A, C, G and U. However, Waldo Cohn used ion-exchange chromatography to separate each of these four into pairs of isomers, which were identified as the ribonucleoside 2'- and 3'-phosphates. This duplicity was overcome by Dan Brown's use of a phosphate diesterase isolated from spleen tissue which digests RNA from its 5'-end to give the four 3'-phosphates Ap, Cp, Gp and Up, whilst an internal diesterase (snake venom phosphate diesterase was used later) cleaved RNA to the four 5'-phosphates, pA, pC, pG and pU. It

follows that RNA chains are made up of nucleotides that have 3'→5'-phosphate diester linkages just like DNA (**Figure 2.38**).

[Figure 2.38 here]

The 3'→5' linkage in RNA is, in fact, thermodynamically less stable than the 'unnatural' 2'→5' linkage, which might therefore have had an evolutionary role. A rare example of such a polymer is produced in vertebrate cells in response to viral infection. Such cells make a glycoprotein called **interferon**, which stimulates the production of an oligonucleotide synthetase. This polymerizes ATP to give oligoadenylates with 2'→5' phosphate diester linkages and from 3 to 8 nucleotides long. Such (2'→5') $(A)_n$ (**Figure 2.39**) then activates an interferon-induced ribonuclease, RNase L, whose function seems to be to break down the viral messenger RNA (Note also the 2'→5' ester linkage is a key feature of self-splicing RNA (Section 7.2.2).

[Figure 2.39 here]

2.4.2 Secondary Structure of RNA: A-RNA and A'-RNA

Two varieties of A-type helices have been observed for fibres of RNA species such as poly(rA)•poly(rU). At low ionic strength, **A-RNA** has 11 bp per turn in a right-handed, anti-parallel double-helix. The sugars adopt a C3'-*endo* pucker and the other geometric parameters are all very similar to those for A-DNA (see **Tables 2.3 and 2.4**). If the salt concentration is raised above 20 per cent, an A'-RNA form is observed which has 12 bp per turn of the duplex. Both structures have typical Watson-Crick base-pairs, which are

displaced 4.4 Å from the helix axis and so form a very deep major groove and a rather shallow minor groove.

These features were confirmed by the analysis of the first single crystal structure of an RNA oligonucleotide, the 14mer r(UUAUAUAUAUAUAA).[44] This 14-mer can be treated as three segments of A-helix separated by kinks in the sugar-phosphate backbone, which perturb the major groove dimensions. It is noteworthy that the 2'-hydroxyl groups are prominent at the edges of the relatively open minor groove.[45] They are extensively hydrated and can be recognised by proteins (**Figure 2.40**). Many more crystal structures of oligoribonucleotides have now been determined, some to atomic resolution. These structures have provided a wealth of information regarding canonical RNA duplexes, the effects on conformation by mismatched base-pairs, hydration and cation coordination.

In one of the first NMR studies of a RNA duplex, Gronenborn and Clore combined two-dimensional NOE analysis[46] (Section 11.2) with molecular dynamics to identify an A-RNA solution structure for the hexaribonucleotide, 5'-r(GCAUGC)$_2$.[47] It shows sequence-dependent variations in helix parameters, particularly in helix twist and in base-pair roll, slide, and propeller twist (**Figure 2.19**). The extent of variation from base to base is much less than for the corresponding DNA hexanucleotide and seems to be dominated by the need of the structure to achieve very nearly optimal base-stacking. This picture supports experimental studies that indicate that base-stacking and hydrogen bonding are equally important as determinants of RNA helix stability.

[Figure 2.40 here]

Antisense RNA is defined as a short RNA transcript that lacks coding capacity, but has a high degree of complementarity to another RNA which enables the two to hybridise.[48] The consequence is that such antisense, or complementary, RNA can act as a repressor of the normal function or expression of the targeted RNA. Such species have been detected in prokaryotic cells with suggested functions concerning RNA-primed replication of plasmid DNA, transcription of bacterial genes, and messenger translation in bacteria and bacteriophages. Quite clearly, such regulation of gene expression depends on the integrity of RNA duplexes.

A crucial cellular 'security' machinery that also depends on double-stranded RNA is **RNA interference** or **RNAi** (see Section 5.7.2).[49-51] This mechanism has evolved to protect cells from hostile genes as well as to regulate the activity of normal genes during growth and development. Tiny RNAs that are termed short interfering RNAs (**siRNAs**) or micro RNAs (**miRNAs**), depending on their origin, are capable of downregulating gene expression by binding to complementary mRNAs, resulting either in mRNA elimination or arrest of translation. Although only discovered some 13 years ago in plants, RNA interference has now been found to be ubiquitous in all eukaryotes. The extraordinary specificity of RNAi and the simplicity of administering double-stranded RNAs to organisms with fully sequenced genomes (i.e. *C. elegans*, *D. melanogaster* and *X. lavis*) render RNAi a method of choice for functional genomics. As with potential applications of the antisense strategy for therapeutic purposes, the success of RNAi as a drug will depend on breakthroughs in cellular uptake and delivery.

2.4.3 RNA·DNA Duplexes

Helices which have one strand of RNA and one of DNA are very important species in biology.

- They are formed when reverse transcriptase makes a DNA complement to the viral RNA.
- They occur when RNA polymerase transcribes DNA into complementary messenger RNA.
- They are a feature in DNA replication of the short primer sequences in Okazaki fragments (Section 6.6.4).
- Antisense DNA is a single-stranded oligodeoxynucleotide designed to bind to a short complementary segment of a target nucleic acid (RNA or single-stranded DNA) with the potential for regulation of gene expression (Section 5.7.1).

Such hybrids are formed *in vitro* by annealing together two strands with complementary sequences, such as poly(rA)•poly(dT) and poly(rI)•poly(dC). These two heteroduplexes adopt the A-conformation common to RNA and DNA, the former giving an 11-fold helix typical of A-RNA and the latter a 12-fold helix characteristic of A'-RNA.

A self-complementary decamer r(GCG)d(TATACGC), also generates a hybrid duplex with Watson-Crick base-pairs. It has a helix rotation of 330° with a step-rise of 2.6 Å and $C^{3'}$-*endo* sugar pucker typical of A-DNA and A-RNA (see **Table 2.1**). The thermodynamic stability of RNA•DNA hybrids relative to the corresponding DNA•DNA duplexes is a function of the deoxypyrimidine content in the DNA strands of the former.[52] Hybrids with DNA strands containing 70-100% pyrimidines are more stable to thermal denaturation than

their DNA•DNA counterparts whereas those with less than 30% deoxypyrimidines are less stable than the DNAs. A pyrimidine content of ca. 50% is the 'break-even' point.

The greater stability in some cases of RNA•DNA heteroduplexes over DNA•DNA homoduplexes is the basis of the construction of **antisense DNA oligomers**.[53-56] These are intended to enter the cell where they can pair with, and so inactivate, complementary mRNA sequences. Additional desirable features such as membrane permeability and resistance to enzymatic degradation have focused attention on oligonucleotides with chemical modifications in the phosphate, sugar or base moieties (Section 4.4). In some cases, the resulting heteroduplexes have proved to have higher association constants than the natural DNA•RNA duplexes and the oligonucleotide analogues exhibit increased resistance to phosphate diesterase action (**Table 2.6**).

[Table 2.6 here]

The subtle differences in conformation between an RNA•DNA hybrid duplex and either DNA•DNA or RNA•RNA duplexes have significance for enzyme action and also for antisense therapy. The therapeutic objective of antisense oligodeoxynucleotides very much depends on their ability to create a duplex with the target RNA and thus make it a substrate for ribonuclease H (**Table 2.6**). Because RNase H cleaves DNA•RNA hybrids but does not cleave the corresponding RNA•RNA duplexes, it can be induced to degrade an endogeneous mRNA species through hybridisation with a synthetic antisense oligodeoxyribonucleotide.

X-ray structures of crystals of duplexes having DNA and RNA residues in both strands showed them to have pure A-form geometry (see above). Duplexes between RNA and DNA also adopt A-form geometry in the solid state and in two cases it has been shown that self-complementary DNA decamers with a single incorporated ribonucleotide are in the A-form in the crystal although the all-DNA sequences prefer the B-form in the crystal and in solution. It is possible that crystal lattice forces and crystallization kinetics play a role in the preference of the A-form geometry observed for all crystal structures of DNA•RNA duplexes.

By contrast, the hybrid duplexes d(GTCACATG)•r(CAUGUGAC) and d(GTGAACCTT)•r(AAGUUCAC) have been analysed by two-dimensional NOE NMR in solution and shown to have neither pure A-form nor pure B-form structure.[57] The sugars of the RNA strands have the regular C3'-*endo* conformation but those in the DNA strand have a novel, intermediate C4'-*endo* conformation. Glycosylic torsion angles in the DNA chain are typical of B-form (near -120^{o}) but those in the RNA chain are typically A-form values (near -140^{o}). Overall the global structure is that of an A-form helix in which the base-pairs have the small rise and positive inclination typical of an A-form duplex (**Figure 2.19a,d**). However the width of the minor groove appears to be intermediate between A-form and B-form duplexes and such structures have been modelled into the active site of RNase H. The results suggest that additional interactions of the protein with the DNA strand are possible only for this intermediate hybrid duplex conformation but not for an RNA•RNA duplex. So, it seems possible that these subtle changes in nucleotide conformation may explain the selectivity of RNase H for hybrid DNA•RNA duplexes.[58] Indeed, crystal structures for complexes between a bacterial RNase H[59] and the RNase H

domain from HIV-1 reverse transcriptase[60] and hybrid duplexes have revealed that the RNA adopts a standard A-form geometry whereas the DNA exhibits B-form sugar puckers. However, the DNA•RNA hybrid at the active site of the reverse transcriptase domain assumes a canonical A-form[60] and it is important to note that DNA•RNA hybrids at the active sites of enzymes can assume a range of conformations.

2.2.4 RNA Bulges, Hairpins and Loops

The functional diversity of RNA species is reflected in the diversity of their three-dimensional structures. Several structural elements have been identified that make up folded RNA and their thermodynamic stabilities relative to the unfolded single strand have been evaluated.[61] The folded conformations are largely stabilised by antiparallel double-stranded helical regions, in which **intrastrand** and **interstrand** base-stacking and hydrogen bonding provide most of the stabilization. Base-paired regions are separated by regions of unpaired bases, either as various types of loops or as single strands, as illustrated for a 55-nucleotide fragment from R17 virus (**Figure 2.41**). Recent years have brought a flurry of new crystal structures of ever larger RNA molecules, featuring many different non-canonical secondary and tertiary structural motifs and culminating in the high-resolution crystal structure analyses of the large and small ribosomal subunits (Sections 7.3.3 and 10.9.4).

Hairpin loops were first identified as components of tRNA structures (Section 7.1.4) where they contain many bases. In the secondary structure deduced for 16S ribosomal RNA, most of the loops have four unpaired bases and these are known as **tetra-loops** (Section 7.1.4). Smaller tri-loops of three bases can also be formed.

Nuclear magnetic resonance and crystallographic studies on such stable tetra-loop hairpins show that their stems have A-form geometry while the loops have additional, unusual hydrogen bonding and base-pair interactions (Section 7.1.2). For example, the GAAA loop has the unusual G•A base-pair and UUYG loops have a reverse wobble U•G base-pair. As a result, simple models appear to be inadequate to describe RNA hairpin stem-loop structures. The nonanucleotide r(CGCUUUGCG) forms a stable tri-loop hairpin whose thermodynamic stability has been determined by analysis of T_m curves to be –101 kJ mol^{-1} for ΔH^o and is close to the calculated value (–90 kJ mol^{-1}) for this RNA helix. Nuclear magnetic resonance analysis shows that the loop has an A-form stem and the chain reversal appears between residues U5 and U6. The three uridine residues on the tri-loop have the C2'-*endo* conformation and show partial base-stacking, notably involving the first U on the 5'-side of the loop. These very high-resolution NMR results give a structure different from those structures computed by restrained molecular dynamics (Section 11.7.2), indicating that further refinement of the computational model is needed.

[Figure 2.41 here]

The hairpin loop is not only an important and stable component of secondary structure, it is a key functional element in a number of well characterized RNA systems. For example, it is required in the RNA TAR region of HIV (human immunodeficiency virus) for *trans*-activation by the Tat protein, and several viral coat proteins bind to specific hairpin loop structures.

Bulges are formed when there is an excess of residues on one side of a duplex. For single base bulges, the extra base can either stack into the duplex, as in the case of an

adenine bulge in the coat protein binding site of R17 phage, or be looped out, as shown by NMR studies in uracil bulges in duplexes. Such bulges can provide high affinity sites for intercalators such as ethidium bromide (Section 9.6). In general, it appears that a bulge of one or two nucleotides has four effects on structure: (1) it distorts the stacking of bases in the duplex, (2) induces a bend in RNA, (3) reduces the stability of the helix and (4) increases the major groove accessibility at base-pairs flanking the bulge.

Internal loops occur where there are non-Watson-Crick mismatched bases. They can involve either one or two base-pairs with pyrimidine-pyrimidine opposition (as in **Figure 2.41**) or mismatched purine-purine or pyrimidine-pyrimidine pairs of which G•A pairs can form a mismatched base-pair compatible with an A-form helix. There are also many examples of larger internal lops. Some of those that are rich in purine residues have been implicated as protein recognition sites. Many of these larger loops show marked resistance to chemical reagents specific for single-strand residues and this, in combination with structural data, suggests that there is probably a high level of order in such loops, notably of base-stacking and base-triples. One general opinion is that the major differences between loop and stem regions are dynamic rather than structural.

Junctions are regions that connect three or more stems (the connecting region for two stems is an internal loop) and are a common feature of computer-generated secondary structures for large RNAs. A prime example is the four-stem junction in the cloverleaf structure of tRNAs in which stacking continuity between the acceptor and the T stems and between the anticodon and D stems is maintained (Section 7.1.4).[62] A junction of three stems forms the hammerhead structure of self-cleaving RNA (Section 7.6.2) and junctions of up to five stems have been observed for 16S RNA.

Thermodynamics of Secondary Structure Elements

The free energy of an RNA conformation has to take into account the contributions of interactions between bases, sugars, phosphates, ions and solvent. The most reliable parameters are those derived experimentally from the T_m profiles (Section 2.5.1) of double-helical regions of RNA and data for each of the 10 nearest-neighbour sequences are given in **Table 2.7**. They are accurate enough to predict the expected thermodynamic behaviour of any RNA duplex to within about 10 percent of its experimental value.[63]

Other structural features are less easy to predict. It is clear that stacking interactions are more important than base-pairings so that an odd purine nucleotide 'dangling' at the 3'-end of a stem can contribute some –4 kJ mol^{-1} to the stability of the adjacent duplex. The energies for mispairs or loops are rather less accurate, but always destabilizing and change with the size of the loop (**Table 2.8**). Energies of these irregular secondary structures also depend on base composition, for example a single base bulge for uridine costs about +8 kJ mol^{-1} and for guanosine about +14 kJ mol^{-1}.

By use of such data, the prediction of secondary structure is a conceptually simple task that can be handled by a modest computer while the more advanced programmes search suboptimal structures as well as that of lowest free energy.

[Table 2.7 here]

Interactions between separate regions of secondary structure are defined as tertiary interactions. One example is that of **pseudoknots**, which involve base-pairing between

one strand of an internal loop and a distant single-strand region (Section 7.6.3, Figure 7.41). Pseudoknots can also involve base-pairing between components of two separate hairpin loops and examples with 3-8 bp have been described as a result of both NMR and X-ray analysis. However, the computer prediction of tertiary interactions and base-triples appears to be still beyond the scope of present methodology.

[Table 2.8 here]

2.4.5 Triple-stranded RNAs

The first triple-stranded nucleic acid was described in 1957 when poly(rU)•poly(rA) was found to form a stable 2:1 complex in the presence of magnesium chloride. The extra poly(rU) strand is parallel to the poly(rA) strand and forms Hoogsteen base-triples in the major groove of an A-form Watson-Crick helix. Triplexes of 2poly(rA)•poly(rU) can also be formed while poly (rC) can form a triplex with poly(rG) at pH 6 which has two cytidines per guanine, one of them being protonated to give the C^{+}xG•C) base-triple also seen for triple-helical DNA (**Figure 2.33**). Base triples are also a very common feature of tRNA structure (Section 7.1.4).[62]

The importance of added cations to overcome the repulsion between the anionic chains of the Watson-Crick duplex and the polypyrimidine third strand is an essential feature of triple-helix formation. $Co^{3+}(NH_3)_6$ and spermine are also effective counter-ions as well as the more usual Mg^{2+}.

Poly(rG) as well as guanosine and GMP, can form structures with four equivalent hydrogen-bonded bases in a plane, with all four strands parallel. It is not clear whether this structure has any relevance to RNA folding.

2.5 Dynamics of Nucleic Acid Structures

Any over-emphasis on the stable structures of nucleic acids runs the risk of playing down the dynamic activity of nucleic acids that is intrinsic to their function. Pairing and unpairing, breathing and winding are integral features of the behaviour of these species.[64]

Established studies on structural transitions of nucleic acids have for a long time used classical physical methods, which include light absorption, NMR spectroscopy, ultracentrifugation, viscometry and X-ray diffraction (Chapter 11). More recently, these techniques have been augmented by a range of powerful computational methods (Section 11.7). In each case, the choice of experiment is linked to the time-scale and amplitude of the molecular motion under investigation.

2.5.1 Helix-coil Transitions of duplexes

Double helices have a lower molecular absorptivity for UV light than would be predicted from the sum of their constituent bases. This **hypochromicity** is usually measured at 256 nm while C•G base-pairs can also be monitored at 280 nm. It results from coupling of the transition dipoles between neighbouring stacked bases and is larger in amplitude for A•U and A•T pairs than for C•G pairs. As a result, the UV absorption of a DNA duplex *increases* typically by 20-30 percent when it is denatured. This transition from a helix to

an unstacked, strand-separated coil has a strong entropic component and so is temperature dependent. The midpoint of this thermal transition is known as the **melting temperature** (T_m).

Such dissociation of nucleic acid helices in solution to give single-stranded DNA is a function of base composition, sequence and chain length as well as of temperature, salt concentration and pH of the solvent. In particular, early observations of the relationship between T_m and base composition for different DNAs showed that A•T pairs are less stable than C•G pairs, a fact which is now expressed in a linear correlation between T_m and the gross composition of a DNA polymer by the equation:

$$T_m = X + 0.41\ [\%(C + G)]\ \ /^oC$$

The constant X is dependent on salt concentration and pH and has a value of 69.3°C for 0.3 M sodium ions at pH 7 (**Figure 2.42**).

[Figure 2.42 here]

A second consequence is that the steepness of the transition also depends on base sequence. Thus, melting curves for homopolymers have much sharper transitions than those for random-sequence polymers. This is because A•T rich regions melt first to give unpaired regions, which then extend gradually with rising temperature until, finally, even the pure C•G regions have melted (**Figure 2.42 and 2.43**). In some cases, the shape of the melting curve can be analysed to identify several components of defined composition melting in series.

[Figure 2.43 here]

Because of end-effects, short homo-oligomers melt at lower temperatures and with broader transitions than longer homopolymers. For example for poly(rA)$_n$•poly(rU)$_m$, the octamer melts at 9^{o}C, the undecamer at 20^{o}C, and long oligomers at 49^{o}C in the same sodium cacodylate buffer at pH 6.9. Consequently, in the design of synthetic, self-complementary duplexes for crystallization and X-ray structure determination, C•G pairs are often places at the ends of hexamers and octamers to stop them 'fraying'. Lastly, the marked dependence of T_m on salt concentration is seen for DNA from *Diplococcus pneumoniae* whose T_m rises from 70^{o}C at 0.01 M KCl to 87^{o}C for 0.1 M KCl and to 98^{o}C at 1.0 M KCl.

Data from many melting profiles have been analysed to give a **stability matrix** for nearest neighbour stacking (**Table 2.9**). This can be used to predict T_m for a B-DNA polymer of known sequence with a general accuracy of 2-3^{o}C.[65]

The converse of melting is the **renaturation** of two separated complementary strands to form a correctly paired duplex. In practice, the melting curve for denaturation of DNA is reversible only for relatively short oligomers, where the rate determining process is the formation of a nucleation site of about 3 bp followed by rapid zipping-up of the strands and where there is no competition from other impeding processes.

[Table 2.9 here]

When solutions of unpaired, complementary large nucleic acids are incubated at 10-20^{o}C below their T_m, renaturation takes place over a period of time. For short DNAs of up to several hundred base-pairs, nucleation is rate-limiting at low concentrations and each duplex zips to completion almost instantly (>1000 bp s^{-1}). The nucleation process is bimolecular, so renaturation is concentration dependent with a rate constant around 10^6 M^{-1} s^{-1}.[66] It is also dependent on the complexity of the single strands. Thus, for the simplest cases of homopolymers and of short heterogeneous oligonucleotides, nucleation sites will usually be fully extended by rapid zipping-up. This gives us an 'all-or-none' model for duplex formation. By contrast, for bacterial DNA each nucleation sequence is present only in very low concentration and the process of finding its correct complement will be slow. Lastly, in the case of eukaryotic DNA the existence of repeated sequences means that locally viable nucleation sites will form and can be propagated to give relatively stable structures. These will not usually have the two strands in their correct overall register. Because such pairings become more stable as the temperature falls, complete renaturation may take an infinitely long time.

Longer nucleic acid strands are able to generate intrastrand hairpin loops, which optimally have about six bases in the loop and paired sections of variable length. They are formed by rapid, unimolecular processes which can be 100 times faster than the corresponding bimolecular pairing process. Although such hairpins are thermodynamically less stable than a correctly paired duplex, their existence retards the rate of renaturation, so that propagation of the duplex is now the rate-limiting process (**Figure 2.44**). One notable manifestation of this phenomenon is seen when a hot solution of melted DNA is quickly quenched to +4^{o}C to give stable denatured DNA.

With longer DNA species, Britten and Kohne have shown that the rate of recombination, which is monitored by UV hypochromicity, can be used to estimate the size of DNA in a homogenous sample. The time t for renaturation at a given temperature for DNA of single-strand concentration C and total concentration C_0 is related to the rate constant *k* for the process by an equation which in its simplest form is:

$$C/C_0 = (1 + kC_0t)^{-1}$$

In practice where C/C_0 is 0.5 the value of C_0t is closely related to the complexity of the DNA under investigation.

This **annealing** of two complementary strands has found many applications. For DNA oligomers it provided a key component of Khorana's chemical synthesis of a gene (Section 5.4.1). It is now an integral feature of the insertion of chemically synthesized DNA into vectors. For RNA•DNA duplexes, it has provided a tool of fundamental importance for gene identification (Section 5.5) and is being explored in the applications of antisense DNA (Section 5.7.1).

[Figure 2.44 here]

2.5.2 DNA Breathing

Complete separation of two nucleic acid strands in the melting process is a relatively slow, long-range process that is not easily reversible. By contrast, the hydrogen bonds between base-pairs can be disrupted at temperatures well below the melting temperature to give

local, short-range separation of the strands. This readily reversible process is known as **breathing**.

The evidence for such dynamic motion comes from chemical reactions which take place at atoms that are completely blocked by normal base-pairings. Those used include tritium exchange studies in hydrogen-bonded protons in base-pairs, the reactivity of formaldehyde with base NH groups, and NMR studies of imino-proton exchange with solvent water. This last technique can be used on a time scale from minutes down to 10 ms. It shows that in linear DNA the base-pairs open singly and transiently with a lifetime around 10 ms at 15^{o}C.

Because NMR can distinguish between imino- and amino-proton exchange it can also be used to identify breathing in specific sequences.[67] Some of the most detailed work of this sort has come from studies on tRNA molecules, which show that, with increasing temperature, base-triplets (**Figure 2.33**) are destabilised first followed by the ribothymidine helix and then the dihydrouridine helix. Finally the acceptor helix ‘melts' after the anticodon helix (Section 7.1.4).

Another possible motion that might be important for the creation of intercalation sites is known as ‘soliton excitation'. The concept here is of a stretching vibration of the DNA chain which travels like a wave along the helix axis until, given sufficient energy, it leads to local unstacking of adjacent bases with associated deformation of sugar pucker and other bond conformations.

Such pre-melting behaviour may well relate to the process of drug intercalation, to the association of single-strand specific DNA binding proteins (Section 10.3.8), and to the

reaction of small electrophilic reagents with imino and amino groups such as cytosine-N3 (Section 8.5)

2.5.3 Energetics of the B-Z Transition

The isomerization equilibrium between the right-handed B-form and the left-handed Z-form of DNA is determined by three factors:

- chemical structure of the polynucleotide (sequence, modified bases)
- environmental conditions (solvent, pH, temperature, etc.)
- degree of topological stress (supercoiling, cruciform formation).

Many quantitative data have been obtained from spectroscopic, hydrodynamic and calorimetric studies and linked to theoretical calculations. While these have not yet defined the kinetics or complex mechanisms of the B-Z transition, it is evident that the small transition enthalpies involved lie within the range of the thermal energies available from the environment. So, for example, the intrinsic free energy difference between Z- and B-forms is close to 2 kJ mol^{-1} for poly-d(G•C) base-pairs, only 1 kJ mol^{-1} for poly-d(G•m^5C) base-pairs, and greater than 5 kJ mol^{-1} for poly-d(A•T) base-pairs. It thus appears that local structural fluctuations may be key elements in the mediation of biological regulatory functions through the B-Z transition.[68,69]

2.5.4 Rapid DNA Motions

Rotations of single bonds, either alone or in combination, are responsible for a range of very rapid DNA motions with time-scales down to fractions of a nanosecond. For example, the twisting of base-pairs around the helix axis has a lifetime around 10^{-8} s while crankshaft rotations of the β, α, ζ and ε C-O-P-O-C bonds (see **Figure 2.11**) lead to an oscillation in the position of the phosphorus atoms on a millisecond time-scale. Various calculations on the interconversion of C3'-*endo* and C2'-*endo* sugar pucker have given low activation energy barriers for their interconversion, in the range 3 to 20 kJ mol^{-1}, showing that the conformers are in rapid, although weighted, equilibrium at 37°C. Lastly, rapid fluctuations in propeller twist can result from oscillations of the glycosylic bond.

2.6 Higher-order DNA Structures

The way in which eukaryotic DNA is packaged in the cell nucleus is one of the wonders of the macromolecular structure. In general, higher organisms have more DNA than lower ones (**Table 2.10**) and this calls for correspondingly greater condensation of the double helix. Human cells contain a total of 7.8×10^9 bp, which corresponds to an extended length of about 2 m. The DNA is packed into 46 cylindrical chromosomes of total length 200 μm, which gives a net packaging ratio of about 10^4 for such metaphase human chromosomes (see also Section 6.4). The overall process has been broken down into two stages: the formation of nucleosomes and the condensation of nucleosomes into chromatin.[70]

[Table 2.10 here]

2.6.1 Nucleosome Structure

The first stage in the condensation of DNA is the **nucleosome**, whose core has been crystallised by Aaron Klug and John Finch and analysed by X-ray diffraction. The DNA duplex is wrapped around a block of eight histone proteins to give 1.75 turns of a left-handed super-helix (Section 10.6, Figure 10.23).[71,72] This process achieves a packing ratio of 7. The number of base-pairs involved in nucleosome structures varies from species to species, being 165 bp for yeast, 183 bp for HeLa cells, 196 pb for rat liver and 241 bp for sea urchin sperm. Such nucleosomes are joined by linker DNA whose length ranges from 0 bp in neurons to 80 bp in sea urchin sperm but usually averages 30-40 bp. The details of packaging the histone proteins are discussed later (Section 10.6.1).[73]

As the DNA winds around the nucleosome core, the major and minor grooves are compressed on the inside with complementary widening of the grooves on the outside of the curved duplex. Runs of A•T base-pairs, which have an intrinsically narrow minor groove should be most favourably placed on the inside of the curved segment while runs of G•C base-pairs should be more favourably aligned with minor grooves facing outwards, where they are more accessible to enzyme cleavage. In practice, Drew and Travers measured the periodicities of A•T and G•C base-pairs by cleavage with DNase I and found them to be exactly out of phase and having a periodicity of 10.17 ± 0.5 bp.[74,75] This result was later confirmed by hydroxyl radical cleavage, which avoids the steric constraints of DNase I.

2.6.2 Chromatin Structure

Chromatin is too large and heterogeneous to yield its secrets to X-ray analysis, so electron microscopy is the chosen experimental probe (Section 11.5.1). At intermediate salt

concentration (~ 1 mM NaCl) the nucleosomes are revealed as ‘beads on a string'. Spherical nucleosomes can be seen with a diameter of 7-10 nm joined by variable-length filaments, often about 14 nm long. If the salt concentration is increased to 0.1 M NaCl, the spacing filaments get shorter and a zig-zag arrangement of nucleosomes is seen in a fibre 10-11 nm wide (Section 10.6, Figure 10.23). At even higher salt concentration and in the presence of magnesium, these condense into a 30 nm diameter fibre, called a **solenoid**, which is thought to be either a right-handed or a left-handed helix made up of close-packed nucleosomes with a packing ratio of around 40.

For the further stages in DNA condensation, one of the models proposed suggests that loops of these 30 nm fibres, each containing about 50 solenoid turns and possibly wound in a supercoil, are attached to a central protein core from which they radiate outwards.[76] Organisation of these loops around a cylindrical scaffold could give rise to the observed miniband structure of chromosomes, which is some 0.84 μm in diameter and 30 nm in thickness. A continuous helix of loops would then constitute the **chromosome**. These ideas are illustrated in a possible scheme (**Figure 2.45**).

[Figure 2.45 here]

It is clear from all of the relevant biological experiments that the single DNA duplex has to be continuously accessible despite all this condensed structure in order for replication to take place. Some of the most exciting electron micrographs of DNA have been obtained from samples where the histones have been digested away leaving only the DNA as a

tangled network of interwound superhelices radiating from a central nuclear region where the scaffold proteins remain intact (**Figure 2.46**).

Even then, the most condensed packing of nucleic acid is found in the sperm cell. Here a series of arginine-rich proteins called protamines bind to DNA, probably with their α-helices in the major groove of the DNA where they neutralise the phosphate charge and so enable very tight packing of DNA duplexes.

[Figure 2.46 here]

Bacterial DNA is also condensed into a highly organised state (Section 10.6.2). In *E. coli* the genome has 4400 kbp in a closed circle which is negatively supercoiled. It is condensed around histone-like proteins, HU and HI, to form a **nucleoid** and achieve a compaction of 1000-fold which is followed by further condensation into supercoil domains. Unlike chromosomal DNA in eukaryotes, there is some additional negative supercoiling in prokaryotes that is not accounted for by protein binding.[77] This is probably a consequence of the activity of the bacterial DNA gyrase, which is capable of actively introducing further negative supercoiling, driven by hydrolysis of ATP. This whole process differs in several respects from assembly of chromatin in eukaryotes:

- there is no apparent regular repeating structure equivalent to the eukaryotic nucleosome although short DNA segments of 60-129 bp are organised by means of their interaction with abundant DNA-binding proteins
- there is no prokaryotic equivalent to the solenoid structure

- bacterial DNA seems to be torsionally strained *in vivo* and organised into independently supercoiled domains of about 100 kpb.

The establishment of DNA architecture in the bacterial chromosome has progressed through the analysis of two types of structure. First, the interaction of a dimer of the HU protein from *B. stearothermophilus* shows loops of antiparallel β-sheets inserted into the DNA minor groove (Section 10.6.2, Figure 10.25C) in non-sequence specific binding. Secondly, a large nucleoprotein complex in bacteria is involved in integration of phage DNA into the host chromosome and is called an **intasome**. This has the phage DNA wrapped as a left-handed supercoil around a complex of proteins including several copies of two DNA-binding proteins, the phase-coded integrase, and the IHP protein (integration host factor). The IHP binds to a specific DNA sequence. These developments suggest that structural analysis of the bacterial chromosome may well overtake that of eukaryotic systems.

Tremendous progress has been made in the characterisation of nucleic acid structure during the past three decades. The ability to chemically synthesize oligodeoxynucleotides paved the way to a characterization of DNA structure in atomic detail. Following the focus of early studies on (a) the conformations of the double helical families, (b) the sequence dependence of their structures, and (c) interactions between DNA and small molecule drugs, attention has progressively shifted to new tertiary structural motifs, such as junctions and four-stranded motifs, some of which were discussed in this chapter, and the interactions between DNA and proteins. Considerable numbers of DNA structures are now deposited in public data bases every year, many of them revealing surprises and shedding light on

familiar but hitherto relatively poorly characterised phenomena such as conformational transitions.[78] 'Is there anything then that we still don't know about the structure of DNA?' the reader may ask. Of course: new exploitations of DNA's chemical and conformational versatility warrant structural characterisations. Supramolecular assemblies and nanostructures constructed from DNA are one example.[79,80] Many questions regarding the interactions between proteins and DNA and the important role that DNA plays in them remain to be answered. Suffice it to mention replication and the need for understanding the nature of nucleotide incorporation by high-fidelity and trans-lesion polymerases opposite native and lesioned DNA templates.[81] And studies directed at a chemical etiology of nucleic acid structure based on the creation and characterisation of dozens of artificial pairing systems have created a further need for structural data.[82] With regard to RNA, the last decade has witnessed an explosion in the analysis of its structure and function. In 1994, the only RNA molecule whose relatively complex tertiary structure had been revealed was transfer RNA. Then came the structures of ribozymes and those of numerous oligoncucleotide-sized fragments, offering a glimpse at the repertoire of RNA's tertiary structural motifs, and a flurry of protein-RNA complexes, and – at last – atomic resolution structures of ribosomal subunits and whole ribosomes. Much remains to be discovered in terms of the mechanism of translation,[83,84] but the availability of ribosome structures and oligonucleotide fragments mimicking portions thereof has reinvigorated the interest in RNA as a drug target.[85,86] On the functional side, in vitro selection and the emergence of a flurry of so-called aptamers, RNA molecules with the capacity to recognize and tightly bind small and large molecules have given a boost to the RNA-world hypothesis, and have subsequently led to the identification of natural control elements in messenger RNAs, 'ribsoswitches'

(Section 5.7.2), that regulate gene expression.[87,88] And as if this were not enough, the advent of RNA interference (RNAi) has further underscored the importance of RNA in the mediation and control of biological information transfer. Perhaps it does not come as a surprise then that well over one third of the human genes appear to be conserved micro RNA (miRNA) targets.[89] New functions come with their structural underpinnings and the structural biology of RNA-mediated gene silencing has already yielded first insights into novel RNA-protein interactions.[90] A little over 50 years after Watson's and Crick's model of the DNA double helix there is no end in sight in the quest for the structural analysis of DNA and RNA.

References

1. Altona, C., and Sundaralingam, M. (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Amer. Chem. Soc*., **94**, 8205-12.

2. Chandrasekaran, R., and Arnott, S. (1989). *The structures of DNA and RNA in oriented fibres*. Springer Verlag, Berlin.

3. Rich, A., and Zhang, S. (2003) Z-DNA: the long road to biological function. *Nature Reviews Genetics*, **4**, 566-72.

4. Kennard, O., and Hunter, W. N. (1991) Single-crystal X-ray diffraction studies of oligonucleotides and oligonucleotide-drug complexes. *Angew. Chem. Intl. Ed. Engl.*, **30**, 1254-77.

5. Dickerson, R. E. (2001). Nucleic acids. In *Crystallography of Biological Macromolecules*. (ed. M. G. Rossmann and E. Arnold). International Tables of Crystallography, Vol. F. Kluwer Academic Publishers, Dordrecht, pp. 588-622.

6. Jurmak, F. A. and McPherson, A. (ed.) (1984). *Biological macromolecules and assemblies*, Vol. 1. John Wiley, New York.

7. Hartmann, B., and Lavery, R. (1996) DNA structural forms. *Quart. Reviews in Biophyscis*, **29**, 309-68.

8. Saenger, W. (1984). *Principles of nucleic acid structure*. Springer Verlag, New York.

9. Calladine, C. R., and Drew, H. R. (1997). *Understanding DNA. The molecule and how it works*. Academic Press Ltd., London.

10. Minasov, G., Tereshko, V., and Egli, M. (1999). Atomic-resolution crystal structures of B-DNA reveal specific influences of divalent metal ions on conformation and packing. *J. Mol. Biol*., **291**, 83-99.

11. Egli, M., Tereshko, V., Teplova, M., Minasov, G., Joachimiak, A., Sanishvili, R., Weeks, C. M., Miller, R., Maier, M. A., An, H., Cook P. D., and Manoharan, M. (2000). X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution. *Biopolymers (Nucleic Acid Sciences)*, **48**, 234-52.

12. Tereshko, V., Minasov, G., and Egli, M. (1999). A "hydrat-ion" spine in a B-DNA minor groove. *J. Am. Chem. Soc*., **121**, 3590-5.

13. Rich, A. (1995) The nucleic acids. A backward glance. In: *DNA: the double helix. Annals of the New York Academy of Sciences*, Vol. 758, 97-142.

14. Gorin, A. A., Zhurkin, V. B., and Olson, W. K. (1995). B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol*., **247**, 34-48.

15. Yanagi, K., Privé, G. G., and Dickerson, R. E. (1991). Analysis of the local helix geometry in three B-DNA decamers and eight dodecamers. *J. Mol. Biol*., **217**, 201-14.

16. El Hassan, M. A., and Calladine, C. R. (1996). Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J. Mol. Biol*., **259**, 95-103.

17. Hunter, C. A. (1993). Sequence-dependent DNA structure, the role of base-stacking interactions. *J. Mol. Biol.,* **230**, 1025-54.

18. Ng, H.-L., Kopka, M. L., and Dickerson, R. E. (2000). The structure of a stable intermediate in the A ⇔ B DNA helix transition. *Proc. Natl. Acad. Sci. U.S.A*., **97**, 2035-9.

19. Vargason, J. M., Henderson, K., and Ho, P.S. (2001). A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl. Acad. Sci. U.S.A*., **98**, 7265-70.

20. Brown, T., and Kennard, O. (1992). Structural basis of DNA mutagenesis**.** *Curr.*

Opin. Struct. Biol., **2**, 354-60.

21. Wells, R. D. (1988). Unusual DNA structures. *J. Biol. Chem.*, **263**, 1095-8.

22. Calladine, C. R., Drew, H. R., and McCall, M. J. (1988). The intrinsic curvature of DNA in solution. *J. Mol. Biol.*, **210**, 127-37.

23. Crothers, D. M., Gartenberg, M. R., and Shrader, T. R. (1992). DNA bending in protein-DNA complexes. *Progress in Nucleic Acids and Molecular Biology*, **208**, 118-45.

24. Dickerson, R. E., Goodsell, D. S. and Neidle, S. A., (1994). "…The tyranny of the lattice…" *Proc. Natl Acad. Sci. USA.*, **91**, 3579-83.

25. Dickerson, R. E. (1998). DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.*, **26**, 1906-26.

26. Timsit, Y., and Moras, D. (1996). Cruciform structures and functions. *Quarterly Reviews in Biophysics*, **29**, 279-307.

27. Lilley, D. M., Sullivan, K. M., Murchie, A. I. H., and Furlong, J. C. (1988). Cruciform extrusion in supercoiled DNA-mechanisms and contextual influence. In

Unusual DNA structures (ed. R.D. Wells and S.C. Harvey), Springer Verlag, Heidelberg, pp.55-72.

28. Ortiz-Lombardía, M., González, A., Eritja, R., Aymamí, J., Azorín, F., and Coll, M. (1999). Crystal structure of a DNA Holliday junction. *Nature Struct. Biol*., **6**, 913-7.

29. Egli, M. (2002). DNA-cation interactions: quo vadis? *Chemistry & Biology*, **9**, 277-86.

30. Hud, N. V., and Plavec, J. (2003). A unified model for the origin of DNA sequence-directed curvature. *Biopolymers*, **69**, 144-59.

31. Ha, S. C., Lowenhaupt, K., Rich, A., Kim, Y.-G., and Kim, K. K. (2005). Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature*, **437**, 1183-86.

32. Palacek, P. (1991). Local supercoil-stabilized DNA structures. *Crit. Rev. Biochem. Molec. Biol.*, **26**, 151-226.

33. Wang, J. C. (1985). DNA topoisomerases. *Ann. Rev. Biochem.,* **54**, 665-97.

34. Liv, L. F. (1989). DNA topoisomerase poisons as antitumour drugs. *Ann. Rev. Biochem.,* **58**, 351-75.

35. Drew, H. R., and Travers, A. A. (1984). DNA structural variations in the *E. coli tyrT* promoter. *Cell*, **37**, 491-502.

36. Soyfer, V. N., and Potaman, V. N. (1996). Triple-helical nucleic acids. Springer Verlag, New York.

37. Moser, H. E., and Dervan, P. B. (1987). Sequence-specific cleavage of double helical DNA by triple-helix formation. *Science*, **238**, 645-50.

38. Thuong, N. T. and Hélène, C. (1993). Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Agnew. Chem. Int. Ed. Engl.*, **32**, 666-90.

39. Rhee, S., Han, Z.-J., Liu, K., Miles, T., and Davies, D. R. (1999). Structure of a triplex helical DNA with a triplex-duplex junction. *Biochemistry*, **38**, 16810-5.

40. Parkinson, G. N., Lee, M. P. H., and Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876-80.

41. Gehring, K., Leroy, J.-L., and Guéron, M. (1993). A tetrameric DNA structure with protonated cytosine•cytosine bas pairs. *Nature*, **363**, 561-65.

42. Chen, L., Cai, L., Zhang, X., and Rich, A. (1994). Crystal structure of a four-stranded intercalated DNA: d(C_4). *Biochemistry*, **33**, 13540-6.

43. Abrescia, N. G. A., Thompson, A., Hyunh-Dinh, T., and Subirana, J. A. (2002). Crystal structure of an antiparallel DNA fragment with Hoogsteen base-pairing. *Proc. Natl. Acad. Sci. U.S.A*., **99**, 2806-11.

44. Dock-Bregeon, A. C., Chevrier, B, Podjarny, A., Moras, D., deBear, J. S., Gough, G. R., Gilham, P. T., and Johnson, J. E. (1988). High resolution structure of the RNA duplex [U(U-A)$_6$A]$_2$. *Nature*, **335**, 375-78.

45. Egli, M., Portmann, S., and Usman, N. (1996). RNA hydration: a detailed look. *Biochemistry*, **35**, 8489-94.

46. Neuhaus, D., and Williamson, M.P. (1989). The nuclear Overhauser effect in structural and conformation analysis, Chapters 5 and 12. VCH, Weinheim.

47. Happ, S. C., Happ, E., Nilges, M., Gronenborn, A. M., and Clore, G. M. (1988). Refinement of the solution structure of the ribonucleotide 5'r(GCAUGC)$_2$. *Biochemistry*, **27**, 1735-43.

48. Houba-He'rin, N., and Inouye, M. (1987). Antisense RNA. In *Nucleic acids and molecular biology* (ed. F. Eckstein and D. M. J. Lilley), Vol.1. Springer Verlag, Berlin, pp. 210-21.

49. Hannon, G. J. (2002). RNA interference. *Nature*, **418**, 244-51.

50. Carrington, J. C., and Ambros, V. (2003). Role of microRNAs in plant and animal development. *Science*, **301**, 336-38.

51. Lau, N. C., and Bartel, D. P. (2003). Censors of the genome. *Scientific American*, 34-41.

52. Lesnik, E. A., and Freier, S. M. (1995). Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry*, **34**, 10807-15.

53. Cohen, J. S. (ed.) (1989). *Oligodeoxynucleotides, antisense inhibitors of gene expression*. Macmillan, London.

54. Uhlmann, E., and Peyman, A. (1990). Antisense oligonucleotides, a new therapeutic principle. *Chem. Rev.*, **90**, 543-84.

55. Sanghvi, Y. S., and Cook, P. D. (Eds.) (1994). *Carbohydrate modifications in antisense research*. *ACS Symp. Ser*., Vol. 580. Am. Chem. Soc., Washington, DC.

56. Nielsen, P. E. (Ed.) (1999). *Oligonucleotide antisense*. *Biochim. Biophys. Acta*, **1489**.

57. Lane, A. M., Ebel, S. and Brown, T. (1993). NMR assignments and solution conformation of the DNA:RNA hybrid d(GCGAACTT).r(AAGUUCAC). *Eur. J. Biochem.*, **213**, 297-306.

58. Fedoroff, O. Y., Salazar, M., and Reid, B. R. (1993). Structure of a DNA:RNA hybrid duplex, Why RNase H does not cleave pure RNA. *J. Mol. Biol.*, **233**, 509-23.

59. Nowotny, M., Gaidamakov, S. A., Crouch, R. J., and Yang, W. (2005). Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell*, **121**, 1005-16.

60. Sarafianos, S. G., Das, K., Tantillo, C., Clark Jr., A. D., Ding, J., Whitcomb, J. M., Boyer, P. L., Hughes, S. H., and Arnold, E. (2001). Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA, *EMBO J.*, **20**, 1449-61.

61. Jaeger, J. A., SantaLucia, J., and Tinoco, I. (1993). Determination of RNA structure and thermodynamics. *Ann. Rev. Biochem*., **62**, 255-87.

62. Schimmel, P. R., Söll, D., and Abelson, J. N. (1979). *Transfer RNA: Structure and dynamics of RNA*. NATO ASI Series. Plenum, New York.

63. Ereler, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D.H. (1986). Improved free energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA*, **83**, 9373-7.

64. McCammon, J. A., and Harvey, S. C. (1987). *Dynamics of proteins and nucleic acids*. Cambridge University Press, Cambridge.

65. Breslauer, K. J., Frank, R., Blöcker, H., and Marky, L.A. (1986). Predicting DNA duplex stability from base sequence. *Proc. Natl. Acad. Sci. USA*, **83**, 3746-50.

66. Wetmur, J. G. (1976). Hybridization and renaturation kinetics of nucleic acids. *Annu. Rev. Biophys. Bioeng.*, **5**, 337-61.

67. James, T. L. (1984). Relaxation behaviour of nucleic acids. In *Phosphorus-31 NMR* (ed. D.G. Gorenstein). Academic Press, New York, pp. 349-400.

68. Soumpasis, D. M., and Jovin, T. M. (1987). Energetics of the B-Z transition. In *Nucleic acids and molecular biology* (ed. F. Eckstein and D. M. J. Lilley), Vol. 1. Springer Verlag, Heidelberg, pp. 85-111.

69. Guéron, M., and Demaret, J.-P. (1992). A simple explanation of the electrostatics of the B-to-Z transition of DNA. *Proc. Natl . Acad. Sci. U.S.A*., **89**, 5740-3.

70. Cold Spring Harbor Symposia (1978). Chromatin. *Cold Spring Harbor Symp. Quant. Biol.*, Vol. 42.

71. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251-60

72. Richmond, T. J., and Davey, C. A. (2003). The structure of DNA in the nucleosome core. *Nature*, **423**, 145-50.

73. Pederson, D. S., Thorma, F., and Simpson, R. T. (1986). Core particles, fibre and transcriptionally active chromatin structure. *Ann. Rev. Cell Biol.*, **2**, 117-47.

74. Travers, A. A., and Klug, A. (1987). The bending of DNA in nucleosomes and its wider implications. *Phil. Trans. Roy. Soc. Lond. B*, **317**, 537-61.

75. Travers, A. A. (1989). DNA conformation and protein binding. *Ann. Rev. Biochem.*, **58**, 427-52.

76. Selker, E. U. (1990). DNA methylation and chromatin structure: a view from below. *Trends Biol. Sci.*, **15**, 103-7.

77. Schmid, M. B. (1988). Structure and function of the bacterial chromosome. *Trends Biol. Sci.*, **13**, 131-5.

78. Egli, M. (2004). Nucleic acid crystallography: current progress. *Curr. Opin. Chem. Biol.*, **8**, 580-91.

79. Paukstelis, P. J., Nowakowski, J. , Birkoft, J. J., and Seeman, N. C. (2004). Crystal structure of a continuous three-dimensional DNA lattice. *Chemistry & Biology*, **11**, 1119-26.

80. Egli, M. (2004). "Deoxyribo nanonucleic acid": antiparallel, parallel and unparalleled. *Chemistry & Biology*, **11**, 1027-29.

81. Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L., and Aggarwal, A. K. (2004). Replication by human DNA polymerase-ι occurs by Hoogsteen base-pairing. *Nature*, **430**, 377-80.

82. Eschenmoser, A. (1999). Chemical etiology of nucleic acid structure. *Science*, **284**, 2118-24.

83. Ferbitz, L., Maier, T., Pratzelt, H., Bukau, B., Deuerling, E., and Ban, N. (2004). Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature*, **431**, 590-96.

84. Takyar, S., Hickerson, R. P., and Noller, H. F. (2005). mRNA helicase activity of the ribosome. *Cell*, **120**, 49-58.

85. Carter, A. P., Clemons, W. M., Broderson, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, **407**, 340-48.

86. Vourloumis, D., Winters, G. C., Simonsen, K. B., Takahashi, M., Ayida, B. K., Shandrick, S., Zhao, Q., Han, Q., and Hermann, T. (2005). Aminoglycoside-hybrid ligands targeting the ribosomal decoding site. *Chem. BioChem.*, **6**, 58-65.

87. Winkler, W., Nahvi, A., and Breaker, R. R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952-56.

88. Batey, R. T., Gilbert, S. D., and Montange, R. K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, **432**, 411-15.

89. Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15-20.

90. Ma, J.-B., Ye, K., and Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, **429**, 318-22.

Tables

Table 2.1. pK_a values for bases in nucleosides and nucleotides

Bases	(site of protonation)	Nucleoside	3'-Nucleotide	5'-Nucleotide
Adenine	(N-1)	3.63	3.74	3.74
Cytosine	(N-3)	4.11	4.30	4.56
Guanine	(N-7)	2.20	2.30	2.40
Guanine	(N-1)	9.50	9.36	9.40
Thymine	(N-3)	9.80	----	10.00
Uracil	(N-3)	9.25	9.43	9.50

These data approximate to 20˚ C and zero salt concentration. They correspond to *loss* of a proton for pK_a >9 and *capture* of a proton for pK_a >5.

Table 2.2. Some light absorption characteristics for nucleotides

		pH 1-2		pH > 11	
Compound	$[\alpha_D]$*	λ_{MAX} /nm	10^{-4} x ε	λ_{MAX} /nm	10^{-4} x ε
Ado 5'-P	-26˚	257	1.5	259	1.54
Guo 3'-P	-57˚	257	1.22	257	1.13
Cyd 3'-P	+27˚	279	1.3	272	0.89
Urd 2'-P	+22˚	262	0.99	261	0.73
Thd 5'-P	+7.3˚	267	1.0	267[a]	1.0
3',5'-cAMP	-51.3˚	256	1.45	260[b]	1.5

[a]pH 7.0; [b]pH 6.0; * specific molar rotation.

Table 2.3. Average helix parameters for the major DNA conformations

Structure Type	Helix sense	Residues per turn	Twist per bp Ω°	Displacement bp D/Å	Rise per bp /Å	Base tilt τ°	Sugar pucker	Groove Minor	Width/Å Major	Groove Minor	Depth/ Majo
A-DNA	R	11	32.7	4.5	2.56	20	C3'-*endo*	11.0	2.7	2.8	13.5
dGGCCGGCC	R	11	32.6	3.6	3.03	12	C3'-*endo*	9.6	7.9	--	--
B-DNA	R	10	36	-0.2 to –1.8	3.3 – 3.4	-6	C2'-*endo*	5.7	11.7	7.5	8.8
dCGCGAATTCGCG	R	9.7	37.1	--	3.34	-1.2	C2'-*endo*	3.8	11.7	--	--
C-DNA	R	9.33	38.5	-1.0	3.31	-8	C3'-*exo*	4.8	10.5	7.9	7.5
D-DNA	R	8	45	-1.8	3.03	-16	C3'-*exo*	1.3	8.9	6.7	5.8
T-DNA	R	8	45	-1.43	3.4	-6	C2'-*endo*	narrow	wide	deep	shallo
Z-DNA	L	12	-9, -51	-2 to -3	3.7	-7	C3'-*endo (syn)*	2.0	8.8	13.8	3.7
A-RNA	R	11	32.7	4.4	2.8	16-19	C3'-*endo*				
A'-RNA	R	12	30	4.4	3.0	10	C3'-*endo*				

Table 2.4. Comparison of helix parameters for A-DNA and B-DNA crystal structures and for a model Z-DNA helix

1. Base step parameters									
Helix	Step	Roll	Tilt	Cup	Slide	Twist	Rise	D_{xy}	Rad_p
B	All	0.6˚	0.0˚	10.0˚	0.4 Å	36.1˚	3.36 Å	3.5 Å	9.4 Å
A	All	6.3˚	----	----	-1.6 Å	31.1˚	2.6 Å	---	9.5 Å
Z	C-G	-5.8˚	0.0˚	12.5˚	5.4 Å	-9.4˚	3.92 Å	5.0 Å	6.3 Å
Z	C-G	5.8˚	0.0˚	-12.5˚	-1.1 Å	-50.6˚	3.51 Å	6.0 Å	7.3 Å
2. Base-pair parameters									
	Base	Tip	Inclination	Propeller	Buckle	Shift	Slide	P-P[a]	
B	All	0.0˚	2.4˚	-11.1˚	-0.2˚	0.8 Å	0.1 Å	8.8-14 Å	
A	All	11.0˚	12.0˚	-8.3˚	-2.4˚	-4.1 Å	---	11.5-11.9 Å	
Z	C	2.9˚	-6.2˚	-1.3˚	-6.2˚	3.0 Å	-2.3 Å	13.7 Å	
Z	G	-2.9˚	-6.2˚	-1.3˚	6.2˚	3 .0 Å	2.3 Å	7.7 Å	

[a]P-P is the shortest interstrand distance across the minor groove.

Table 2.5. Average torsion angles (˚) for DNA helices

Structure Type	α	β	γ	δ	ε	ζ	χ
A-DNA[a]	-50	172	41	79	-146	-78	-154
GGCCGGCC	-75	185	56	91	-166	-75	-149
B-DNA[a]	-41	136	38	139	-133	-157	-102
CGCGAATTCGCG	-63	171	54	123	-169	-108	-117
Z-DNA (C residues)	-137	-139	56	138	-95	80	-159
Z-DNA (G residues)	47	179	-169	99	-104	-69	68
DNA-RNA decamer	-69	175	55	82	-151	-75	-162
A-RNA	-68	178	54	82	-153	-71	-158

[a] Fibres

Table 2.6. Properties of antisense oligonucleotides and 1st and 2nd generation analogues

Oligonucleotide type	Duplex stability[a]	Nuclease resistance[b]	RNase H activation[c]
Oligodeoxyribonucleotide (PO_2)	Par[a]	--	Yes
Oligodeoxyribonucleotide phosphorothioate	-	++	Yes
Oligodeoxyribonucleotide methylphosphonate	-	+++	No
Oligodeoxyribonucleotide phosphoramidate	+	++	No
Oligoribonucleotide (PO_2)	+	--	No
Oligo(2'-*O*-Me)ribonucleotide (PO_2)	++	+	No
Oligo(2'-*O*-(2-methoxyethyl)ribonucleotide[d] (PO_2)	++	+	No
Oligo(2'-*O*-(3-aminopropyl)ribonucleotide[e] (PO_2)	+	+++	No
Oligo(2'-*O*-(N,N-dimethylaminooxyethyl)[f] (PO_2)	++	+++	No
Oligo(2',4'-methylene …)[g] (PO_2)	+++	+	No
Oligo(2'-fluoroarabinonucleotide)[h] (PO_2)	+	+	Yes
Peptide Nucleic Acids[i]	+	+++	No
Oligodeoxy(5-propyne-cytidine) (PO_2)	++	--	Yes

[a] Compared to DNA-RNA stability under physiological conditions.
[b] Compared to DNA (phosphate diesterase digestion)
[c] Activation of RNase H by the duplex formed between the oligonucleotide and RNA
[d] 2'-*O*-MOE
[e] 2'-*O*-AP
[f] 2'-*O*-DMAEOE
[g] Locked nucleic acid (LNA)
[h] FANA
[i] PNA

Table 2.7. Thermondynamic parameters for RNA helix initiation and propagation in 1 M NaCl

Propagation Sequence	ΔH° kJ mol^{-1}	ΔS° JK^{-1} mol^{-1}	ΔG° kJ mol^{-1}	Propagation Sequence	ΔH° KJ mol^{-1}	ΔS° JK^{-1} mol^{-1}	ΔG° kJ mol^{-1}
↑ A•U A•U ↓	-27.7	-77.3	-3.8	↑ A•U G•C ↓	-55.8	-149	-9.6
↑ U•A A•U ↓	-23.9	-65.1	-3.8	↑ U•A G•C ↓	-42.8	-110	-8.8
↑ A•U U•A ↓	-34.1	-94.9	-4.6	↑ G•C C•G ↓	-33.6	-81.5	-8.4
↑ A•U C•G ↓	-44.1	-117	-7.6	↑ C•G G•C ↓	-59.6	-147	-14.7
↑ U•A C•G ↓	-31.9	-80.6	-7.1	↑ G•C C•G ↓	-51.2	-125	-12.4
Initiation	(0)	-45.4	14.6				
Symmetry correction (self-complementary)	0	-5.9	1.7	Symmetry correction (non-self-complementary)	0	0	0

Arrows point in a 5'-3' direction to designate the stacking of adjacent base-pairs
The enthalpy change for helix initiation is assumed to be zero.

Table 2.8. Free energy increments for loops (kJ mol^{-1} in 1 M NaCl, 37˚C)

Loop Size	Internal Loop	Bulge Loop	Hairpin Loop
1	---	+14	---
2	+4	+22	---
3	+5.4	+25	+31
4	+7.1	+28	+25
5	+8.8	+31	+18.5
6	+10.5	+34.5	+18

Table 2.9. Thermal stability matrix for nearest-neighbour stacking in base-paired dinucleotide fragments with B-DNA geometry

	3'-neighbour			
5'-neighbour	A	C	G	T
A	54.50	97.73	58.42	57.02
C	54.71	85.97	72.55	58.42
G	86.44	136.12	85.97	97.73
T	36.73	86.44	54.71	54.50

Numbers give T_m values in ˚C at 19.5 mM Na^+

Table 2.10. Cellular DNA content of various species

Organism	Numbers of base-pairs	DNA length mm	Number of chromosomes
Escherichia coli	4×10^6	1.4	1
Yeast (*Saccharomyces cerevisiae*)	1.4×10^6	4.6	16
Fruit fly (*Drosophila melanogaster*)	1.7×10^7	56.0	4
Humans (*Homo sapiens*)	3.9×10^9	990.0	23

Values are provided for haploid genomes

Figure Legends

Figure 2.1 Structures of the five major purine and pyrimidine bases of nucleic acids in their dominant tautomeric forms and with the IUPAC numbering systems for purines and pyrimidines.

Figure 2.2 Structures of the four ribonucleosides. The bases retain the same numbering system and the pentose carbons are numbered 1' through 5'. By convention, the furanose ring is drawn with its ring oxygen at the back and C-2' and C-3' at the front. Hydrogen atoms are usually omitted for clarity.

Figure 2.3 Structures of the four major deoxynucleosides. By convention, only hydrogens bonded to oxygen or nitrogen are depicted.

Figure 2.4 Structures of some common nucleotides. All are presented as their sodium salts in the state of ionization observed at neutral pH.

Figure 2.5 States of protonation of adenosine 5'-phosphate (AMP) from strongly acidic solution (left) to strongly alkaline solution (right).

Figure 2.6 Keto-enol tautomers for 2-pyridone:2-hydroxypyridine (left) and amine-imine tautomerism for 2-aminopyridine (right).

Figure 2.7 Tautomeric equilibria for deoxycytidine showing hydrogen-bond acceptor **a** and donor **d** sites as used in nucleic acid base-pairing. The major tautomer for deoxyguanosine is drawn to show its characteristic **d·d·a** hydrogen-bond donor-acceptor capacity.

Figure 2.8 Watson-Crick base-pairing for C•G (left) and T•A (centre). Hoogsteen base-pairing for A:T (right).

Figure 2.9 'Wobble' pairings for U•G (left), U•I (centre) and A•I (right).

Figure 2.10 Proton NMR spectrum for cytidine (run in D_2O at 400 MHz).

Figure 2.11 **a**. Torsion angle notation (IUPAC) for polynucleotide chains and structures for the C2′-*endo***(S)** and C3′-*endo***(N)** preferred sugar puckers. **b**. Schematic of the pseudorotation phase angle (*P*) cycle with the angle ranges of selected pucker types indicated.

Figure 2.12 *Anti* and *syn* conformational ranges for glycosylic 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).

Figure 2.13 Preferred nucleotide conformations: **+sc** for $C^{4'}$-$C^{5'}$ (left); **ap** for $C^{5'}$-$O^{5'}$ (centre); and **ap/-ac** for $C^{3'}$-$O^{3'}$ (right).

Figure 2.14 (Upper) Gauche conformation for phosphate diesters showing the antiperiplaner 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 kJ mol^{-1}) (adapted from Govil, G. (1976). *Biopolymers*, **15**, 2303-7 (1976) © John Wiley and Sons, Inc.).

Figure 2.15 The primary structure of DNA (left) and three of the common shorthand notations: 'Fischer' (upper right), linear alphabetic (centre right) and condensed alphabetic (lower right).

Figure 2.16 Van der Waals representation of 10 base-pairs of A-form DNA. The view is across the major (bottom) and minor grooves (top). Atoms of the sugar-phosphate backbones of strands are coloured in red and green, respectively, and the corresponding nucleoside bases are coloured in pink and blue, respectively. Phosphorus atoms are highlighted in black.

Figure 2.17 Van der Waals representation of 10 base-pairs of B-form DNA. The view is across the major (top) and minor grooves (bottom). The colour code is identical to that in Figure 2.16.

Figure 2.18 The minor groove hydration 'ribbon' in the dodecamer d(CGCGAATTCGCG). The inner and outer water (1' to 9') spines define four fused hexagons that dissect the minor groove. Only 10 base-pairs are shown and terminal residues are numbered (from Tereshko, V. *et al*. (1999). *J. Amer. Chem. Soc.*, **121**, 3590-95; © American Chemical Society).

Figure 2.19 Diagrams illustrating movements of bases in sequence-dependent structures. Rows (a) to (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 around 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.

Figure 2.20 Van der Waals representation of 10 base-pairs of Z-form DNA. The drawing illustrates the narrow minor groove, visible in the centre of the top half of the duplex, and the lack of an effective major groove that takes on the shape of a convex surface instead, visible on the left-hand side of the bottom half of the duplex. The colour code is identical to that in Figure 2.16.

Figure 2.21 Diagrams illustrating (**a**) clockwise propeller twist for a C-(3'→5')-G clash between guanines in the **minor** groove and (**b**) clockwise propeller twist for a G-(3'→5')-C sequence showing purines clashing in the **major** groove.

Figure 2.22 'Wobble' pairs for transition mismatches G•T (left) and A•C (right).

Figure 2.23 Mismatched G•A pairings (a) for the decamer d(CCAAGATTGG) with (*anti*)G•A(*anti*) and (b) for the dodecamer d(CGAGAATTCGCG) with (*anti*)G•A(*syn*) conformation.

Figure 2.24 Curved DNA illustrated by straight poly(dA) tracts (upper), consecutive tracts of writhing DNA of general sequence (lower), and curved DNA (right) of alternating segments of linear poly(dA) and curved tracts of general sequence.

Figure 2.25 **a**. The hairpin loop formed by d(CGCGCGTTTTCGCGCG). **b**. Formation of a cruciform from an inverted repeat sequence of the bacterial plasmid pBR322. The inverted palindromic regions, each of 11 bp, are shown in colour.

Figure 2.26 Structures of a cruciform and alternative pathways for its formation (base-paired sections are helical throughout).

Figure 2.27 The Holliday junction adopted by four DNA decamers with sequence TCGGTACCGA (PDB: 1M6G). The view illustrates the side-by-side arrangement of the

two duplex portions, with the A nucleotides (red) C nucleotides (pink) of two decamers that form the core of the junction visible near the centre. The positions of phosphate groups in the backbones of individual oligonucleotides are traced with ribbons.

Figure 2.28 Graphical representations of electrostatic surface potentials (ESPs) calculated at the solvent accessible surfaces of model (**a**) A-tract and (**b**) G-tract DNA duplexes, each in two helical forms. The model for the A-tract is the duplex $(dA)_{12}$•$(dT)_{12}$ and the model for the G-tract is the duplex $(dG)_{12}$• $(dC)_{12}$. Colours of the DNA surfaces range from red, -8 kT/e, to blue, +3 kT/e, with increasing electrostatic potential (from Hud, N. V. and Plavec, J. (2003), *Biopolymers*, **69**, 144-59. © (2003) Wiley Periodicals, Inc.).

Figure 2.29 Sedimentation velocity for SV40 DNA as a function of bound ethidium (a) for closed circular DNA (–+–+–+–) and (b) for nicked circular DNA (–•–•–•–) showing the transition from a negative supercoil (left) through a relaxed circle (centre) to a positive supercoil (right) (adapted from Bauer, W. and Vinograd, J. (1968). *J. Mol. Biol.*, **33**, 141, © Academic Press).

Figure 2.30 Supercoil formation in closed circular DNA. **a**. Closed circle of 20 duplex turns (alternate turns in colour). **b**. Circle nicked, underwound two turns, and resealed. **c**. Base-pairing and stacking forces result in the formation of B-helix with two new right-handed helix turns and one compensating right-handed supercoil.

Figure 2.31 Topoisomers of plasmid pAT153 after incubation with topoisomerase I to produce partial relaxation. Electrophoresis in a 1 per cent agarose gel: Track 1 shows native supercoiled pAT153 (S1), and supercoiled dimer (S2), and nicked circular DNA (N); Track 2 shows products of topoisomerase I where ΔL_K up to 14 can be seen clearly (from Lilley, D. M. J. (1986). *Symp. Soc. Gen. Microbiol.*, **39**, 105-17. © Society of General Microbiology).

Figure 2.32 Action of topoisomerase II (red) on singly supercoiled DNA: (i) double-strand opening; (ii) double-strand passage; (iii) resealing to give (a) relaxed circle, (b) knotted circle, and (c) catenated DNA circles.

Figure 2.33 Base triples formed by (**a**) Hoogsteen bonding for T•AxT and C•GxC$^+$ and (**b**) reversed Hoogsteen binding for T•AxA and C•GxG. **c**. Model of triple-helical DNA based on fibre diffraction of poly(dA) •2poly(dT) (kindly provided by the late Professor Claude Hélène). **d**. A schematic representation of H-DNA showing loci for attack by single-strand specific reagents.

Figure 2.34 The triple helical structure formed in the crystal structure of the 1: 1 complex between the DNA 12mer d(CTCCTCCGCGCC) and the 9mer d(CGCGCGGAG) (PDB: 1D3R; http://www.rcsb.org).[39] The triplex segment consists of two 5'-halves of the 12mer (underlined, top and bottom left) and two 3'-terminal trimers GAG (underlined; top and bottom right) from two 9mers that are stacked tail-to-tail in the crystal: 5'-…-GAG-3'|3'-GAG-…-5' (visible near the centre, left of drawing). DNA bases

are coloured orange, yellow, cyan and blue for G, A, C and T, respectively, and the directions of the sugar-phosphate backbones are traced by ribbons.

Figure 2.35 **a**. The parallel-stranded G-tetrad motif formed by two molecules d(TAGGGTTAGGGT) (PDB: 1K8P). The directions of the sugar-phosphate backbones are traced by ribbons. The view is from the side, illustrating the disc-like shape with the three G-tetrad layers on the inside and the TTA 'propellers' on the outside. The 5'-termini of both strands are pointing downwards. **b**. Schematic line diagram illustrating the relative orientation of the two strands and the formation of three layers of G-quartets.

Figure 2.36 **a**. The i-motif adopted by four molecules of $d(C)_4$ (PDB:190D). Atoms are coloured green, red, blue and magenta for carbon, oxygen, nitrogen and phosphorus, respectively. Positions of phosphorus atoms from the two intercalated $C{\bullet}C^+$-paired duplexes are traced by orange (5'-ends at the top and 3'-ends at the bottom) and grey (5'-ends at the bottom and 3'-ends at the top) ribbons, respectively, to highlight their antiparallel relative orientation. The absence of significant overlap between cytosine planes from adjacent hemiprotonated base-pairs and two wide grooves and two narrow grooves with van der Waals contacts between antiparallel strands from two duplexes across the latter grooves are hallmarks of the i-motif. **b**. Schematic line diagram of the i-motif, illustrating the formation of intercalated parallel-stranded $C{\bullet}C^+$-paired duplexes with opposite polarities.

Figure 2.37 Structure of the antiparallel duplex observed in the crystal structure of d(ATATAT) which displays Hoogsteen pairing between adenine and thymine bases (PDB: 1GQU). The view is across the major (right) and minor (left) grooves. Atoms are coloured green, red, blue and magenta for carbon, oxygen, nitrogen and phosphorus, respectively, and base planes of A and T are filled yellow and blue, respectively.

Figure 2.38 The primary structure of RNA (left) and cleavage patterns with spleen (centre) and snake venom (right) phosphate diesterases.

Figure 2.39 Structure and formation of interferon-induced (2'→5')$(A)_n$.

Figure 2.40 Van der Waals representation of the RNA duplex $[r(UUAUAUAUAUAUAA)]_2$ (PDB code 1RNA). The view is into the narrow major groove of the central part of the duplex, with the minor groove visible near the top and bottom. Atoms are coloured grey, red, blue and magenta for carbon, oxygen, nitrogen and phosphorus, respectively. 2'-Oxygen atoms lining the minor groove are highlighted in cyan.

Figure 2.41 A possible secondary structure for a 55-nucleotide fragment from R17 virus which illustrates hairpin loop, interior loop and bulge structures. The free energy of this structure has been calculated to have a net ΔG° of –90 kJ mol^{-1} using appropriate values for base-pairs (Table 2.6) and for loops and bulges.

Figure 2.42 Thermal denaturation of DNAs as a function of base composition (per cent G•C) for three species of bacteria: **a**. *Pneumococcus* (38 per cent G•C). **b**. *E. coli* (52 per cent G•C). **c**. *M. phlei* (66 per cent G•C) (adapted from Marmur, J. and Doty, P. (1959), *Nature*, **183**, 1427-9. © (1959) Macmillian Magazines Ltd.).

Figure 2.43 Scheme illustrating the melting of A•T-rich regions (colour) followed by mixed regions, then by C•G-rich regions (black) with rise in temperature (left → right).

Figure 2.44 Renaturation processes (**a**) for short oligonucleotide and longer homopolymers and (**b**) for natural DNA strands.

Figure 2.45 Schematic drawing to illustrate the gradual organization of DNA into highly condensed chromatin. (1) DNA fixed to the protein scaffold; (2) DNA complexed with all histones except H1; (3) aggregation into 100 Å fibre; (4) formation of 'superbeads'; and (5) contraction into a 600 Å knob (from Rindt, K.-P. and Nover, L. (1980). *Biol. Zentralblat*. **99**, 641-73).

Figure 2.46 Electron micrograph of a histone-depleted chromosome showing that the DNA is attached to the scaffold in loops (from Paulson, J. K. and Laemmli, U. K. (1977). *Cell*, **12**, 817-28. © (1977) Cell Press).

Tiff-Format Figures

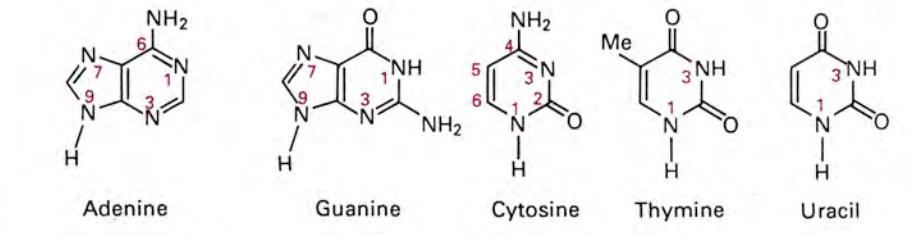


Fig. 2.1

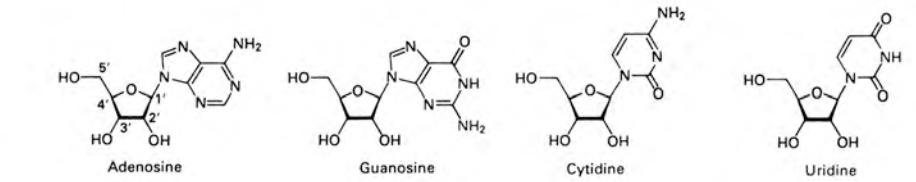


Fig. 2.2

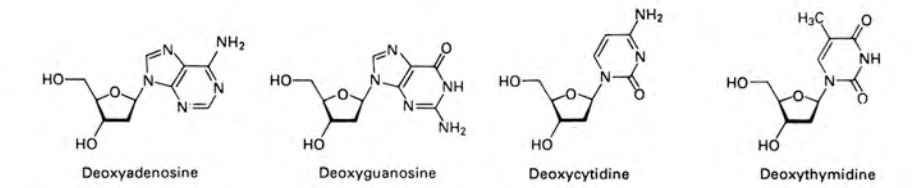


Fig. 2.3

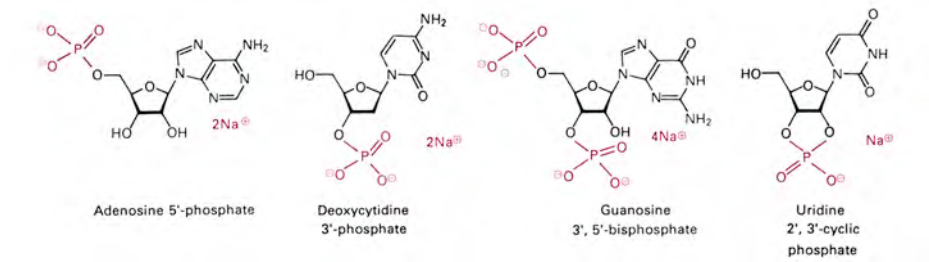


Fig. 2.4

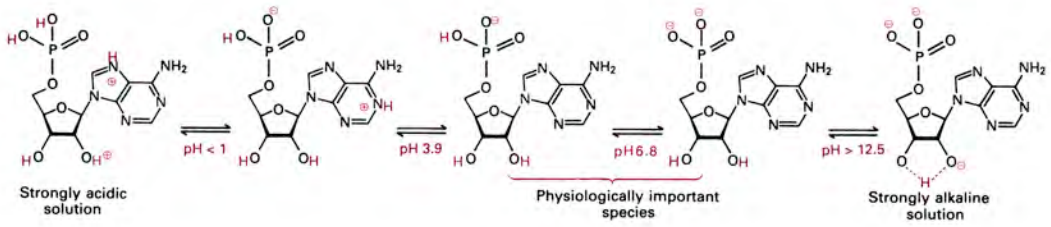


Fig. 2.5

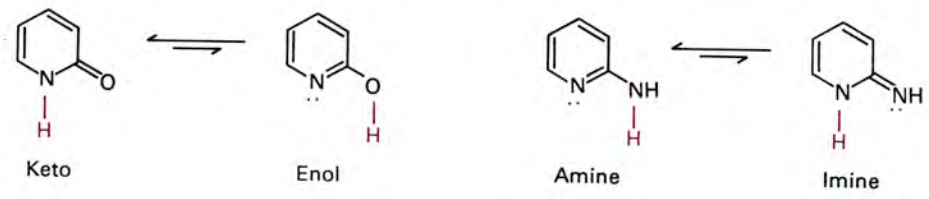


Fig. 2.6

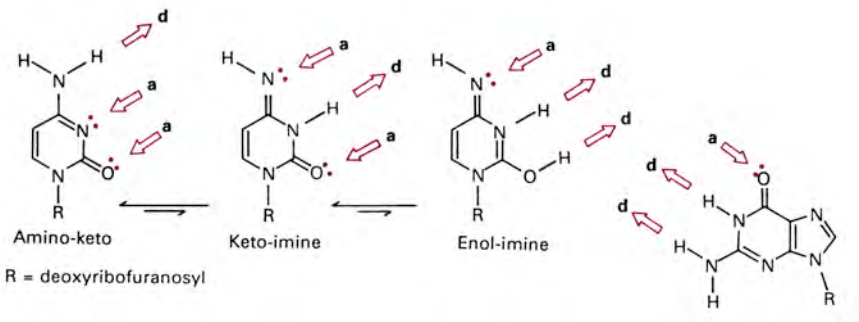


Fig. 2.7

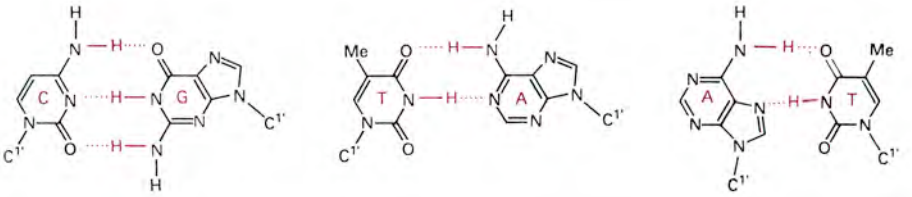


Fig. 2.8

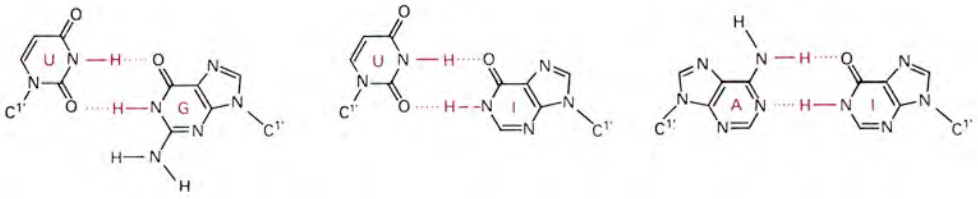


Fig. 2.9

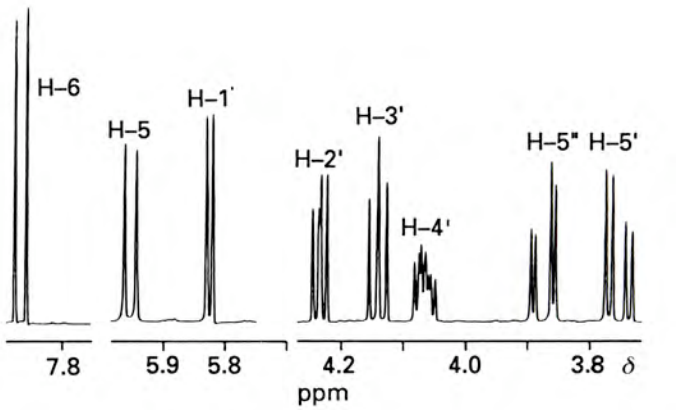
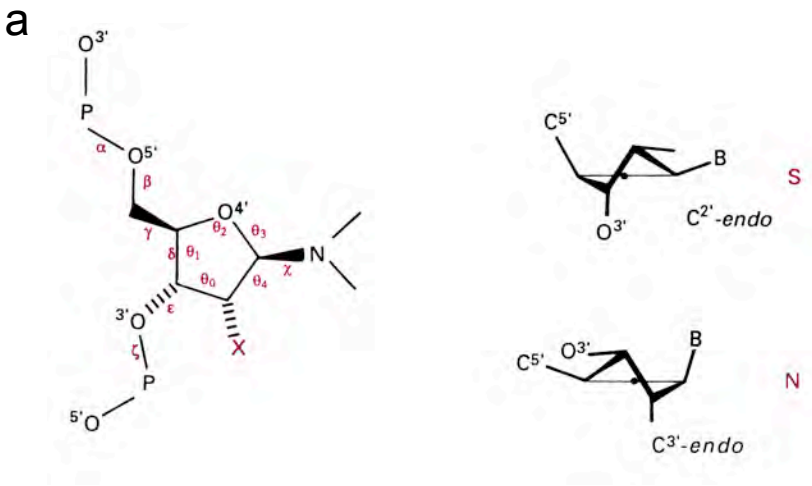


Fig. 2.10



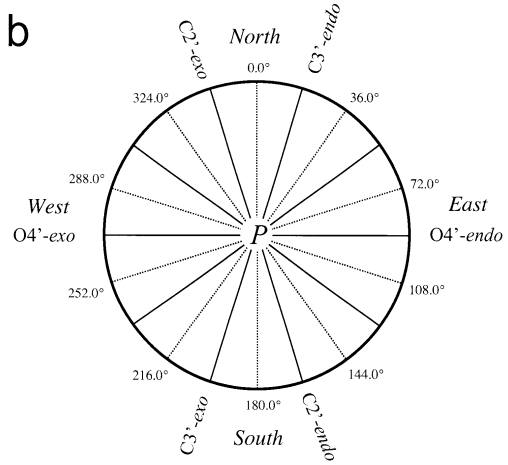


Fig. 2.11

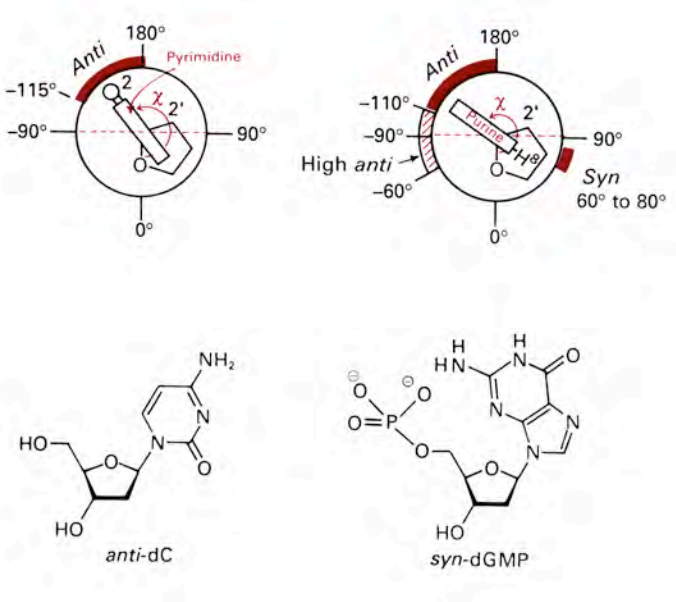


Fig. 2.12

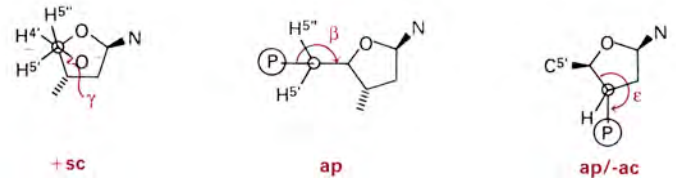


Fig. 2.13

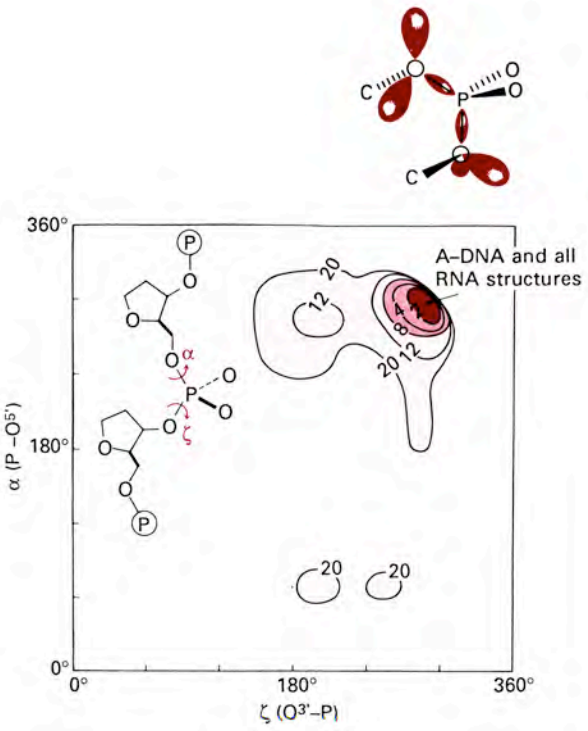


Fig. 2.14

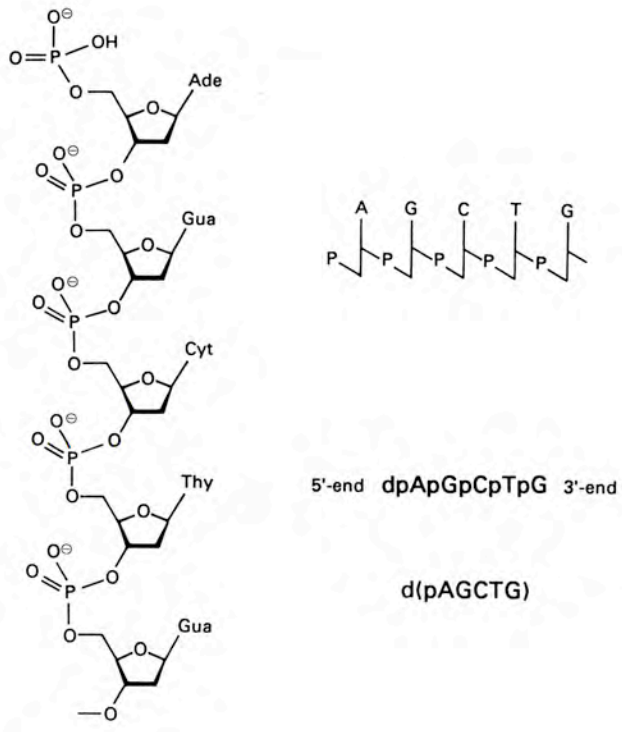


Fig. 2.15

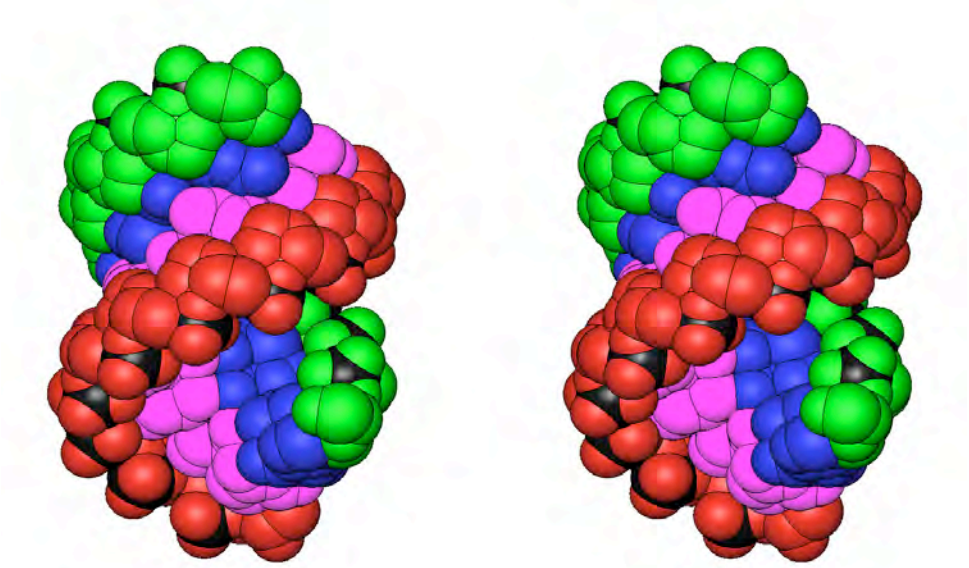


Fig. 2.16 (Stereo)

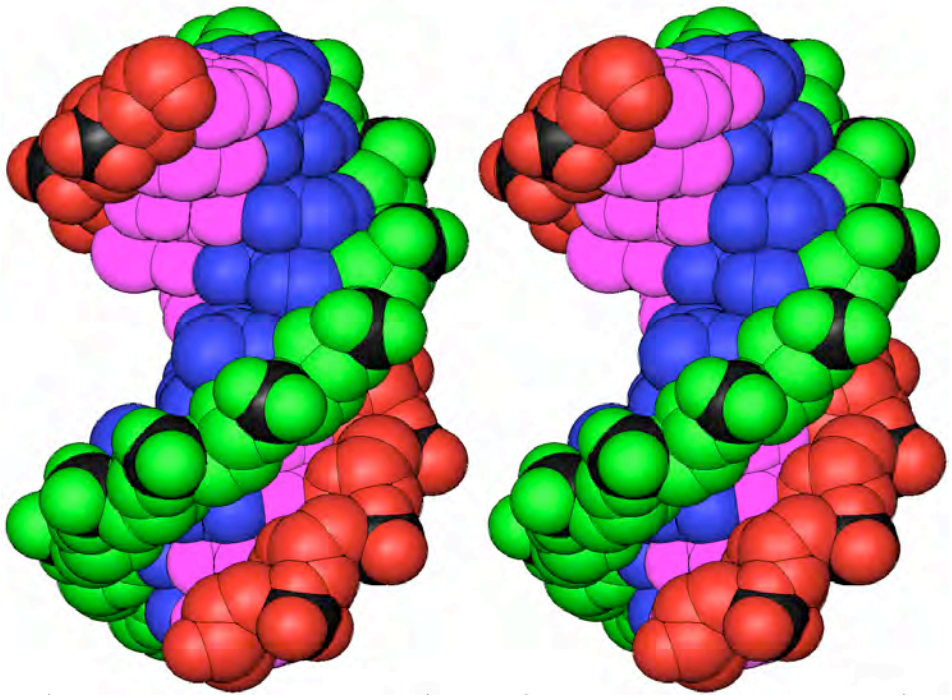


Fig. 2.17 (Stereo)

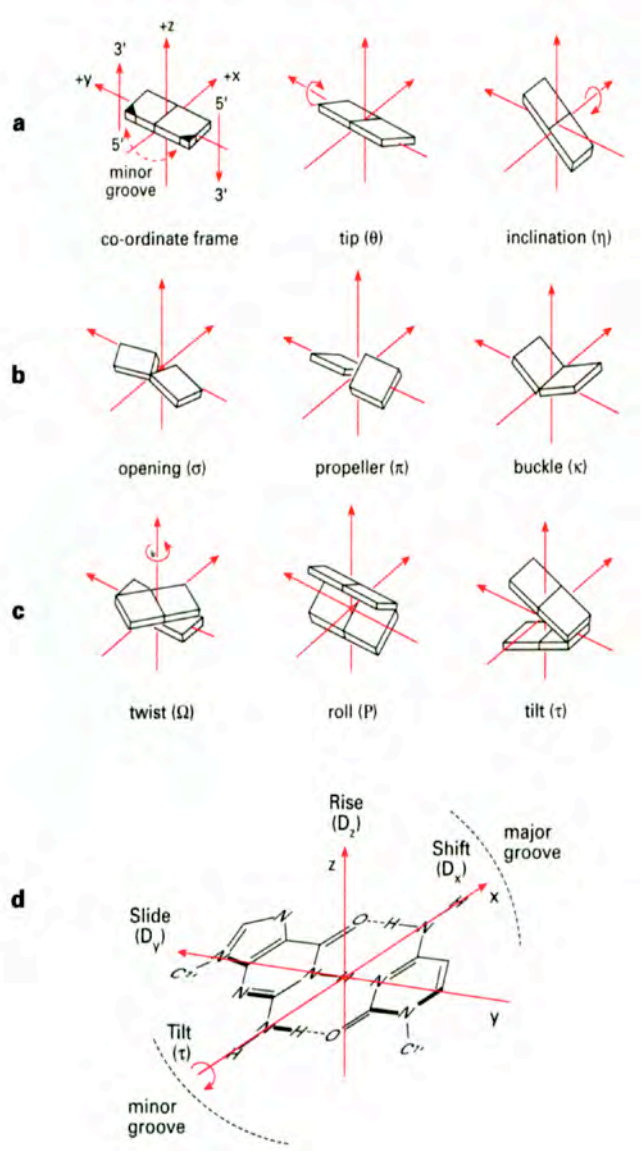

a
+z
3'
+y
+x
5'
5'
minor
groove
3'
co-ordinate frame
tip (θ)
inclination (η)
b
opening (σ)
propeller (π)
buckle (κ)
c
twist (Ω)
roll (P)
tilt (τ)
d
Rise
(D_z)
z
Shift
(D_x)
major
groove
x
Slide
(D_y)
y
Tilt
(τ)
minor
groove

Fig. 2.18

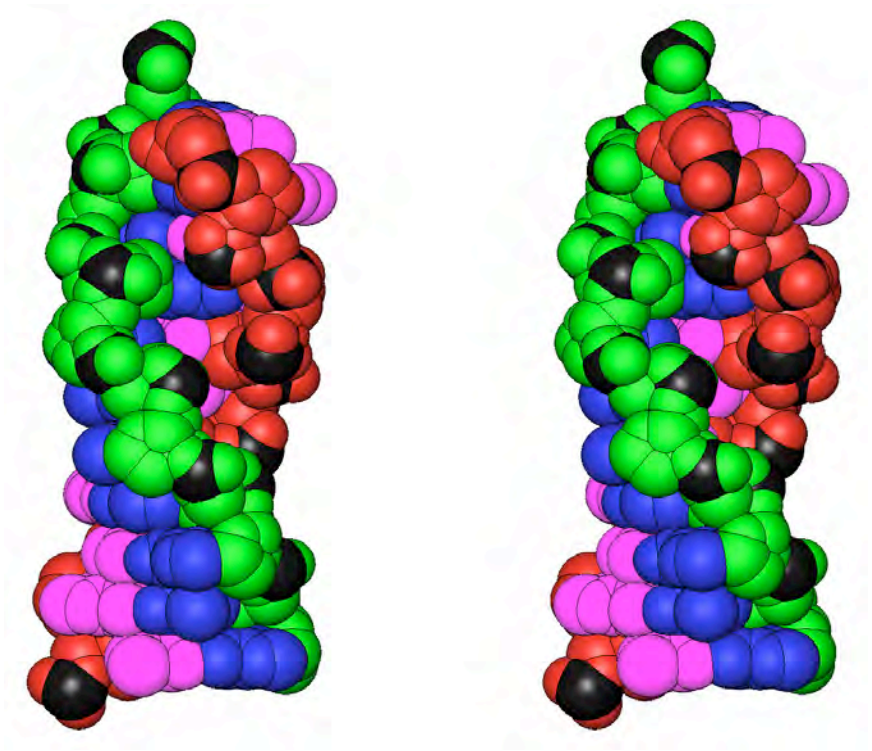


Fig. 2.19 (Stereo)

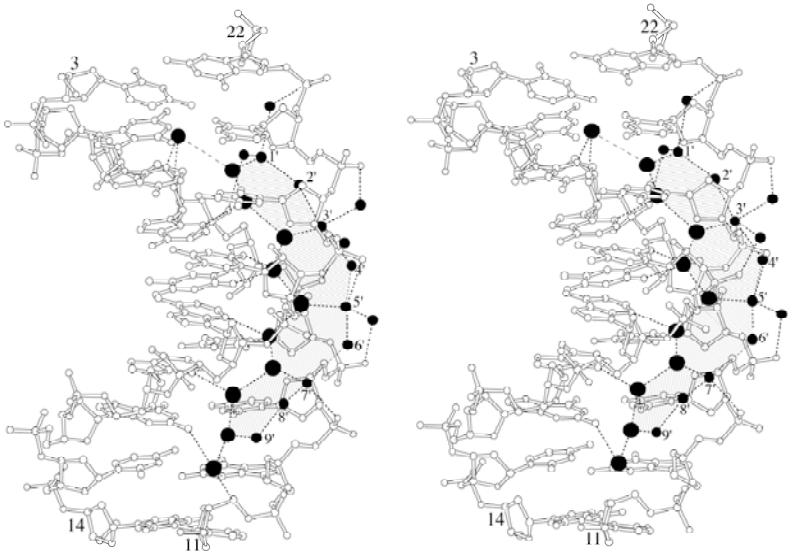


Fig. 2.20 (Stereo)

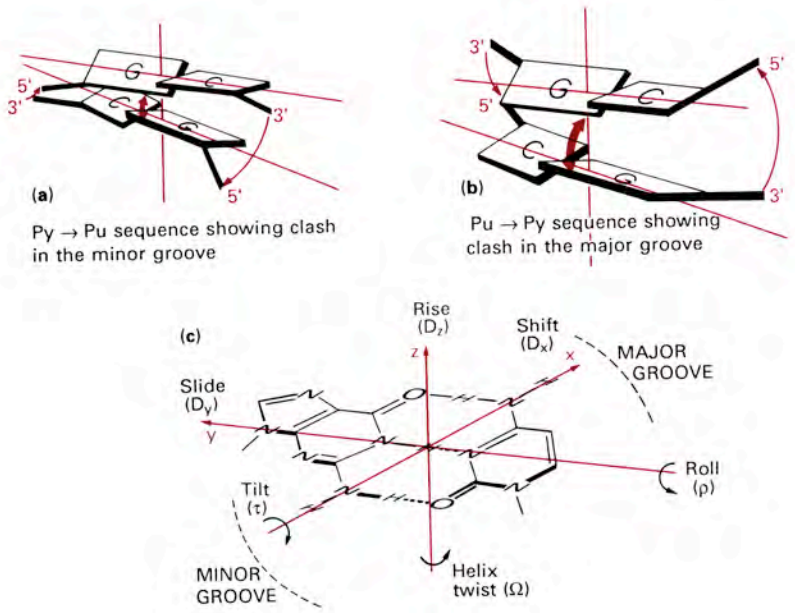


Fig. 2.21

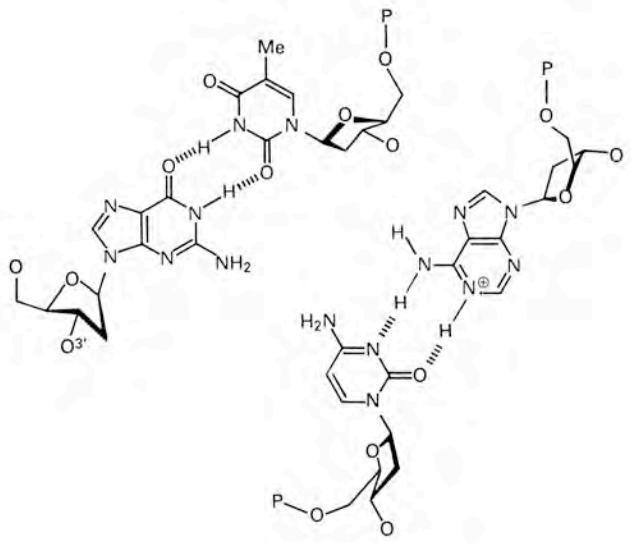


Fig. 2.22

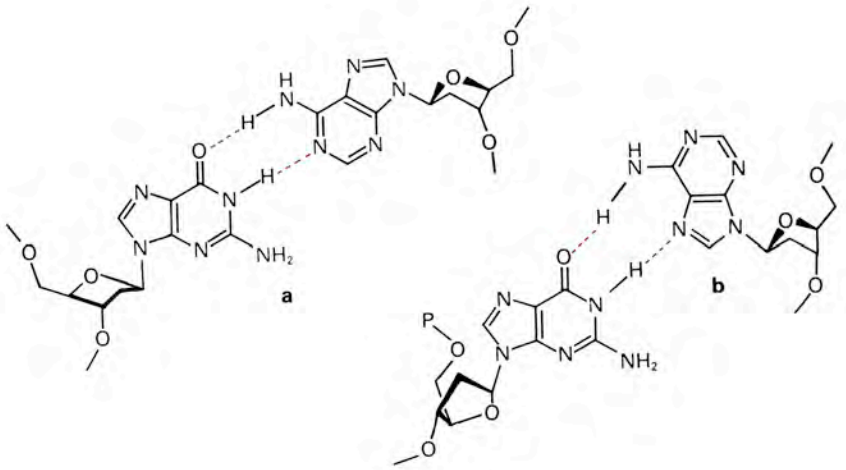


Fig. 2.23

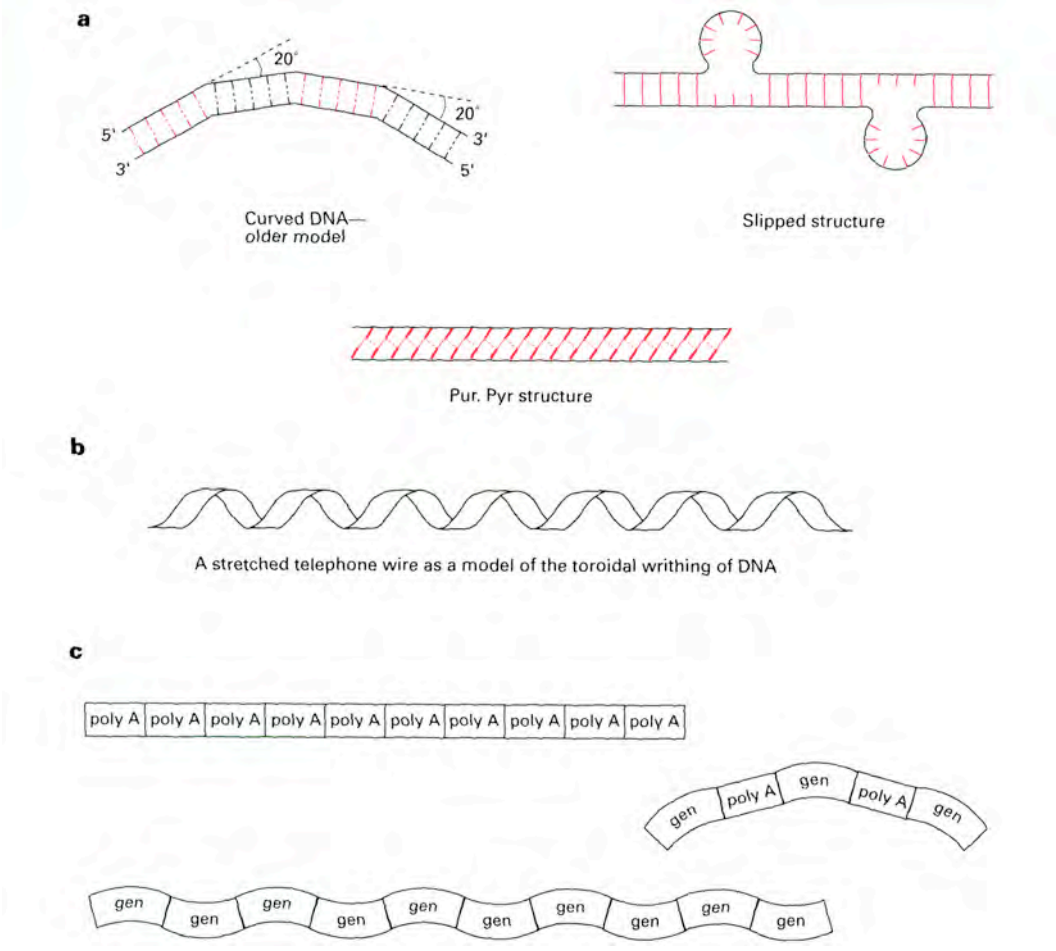


Fig. 2.24

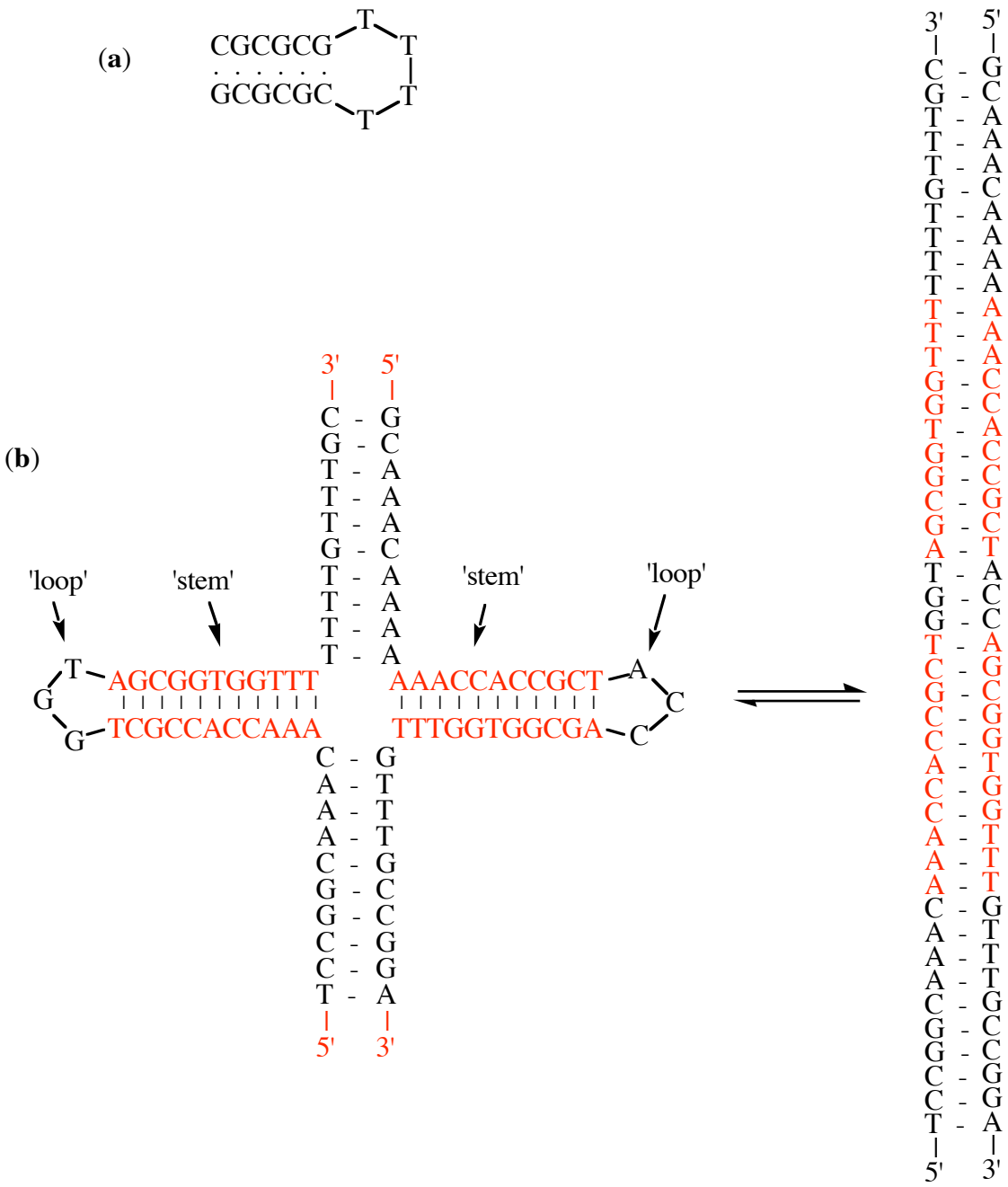

(a)
CGCGCG
GCGCGC
T
T
T
T
(b)
'loop'
'stem'
'stem'
'loop'
AGCGGTGGTTT
TCGCCACCAAA
AAACCACCGCT
TTTGGTGGCGA
3'
5'

Fig. 2.25

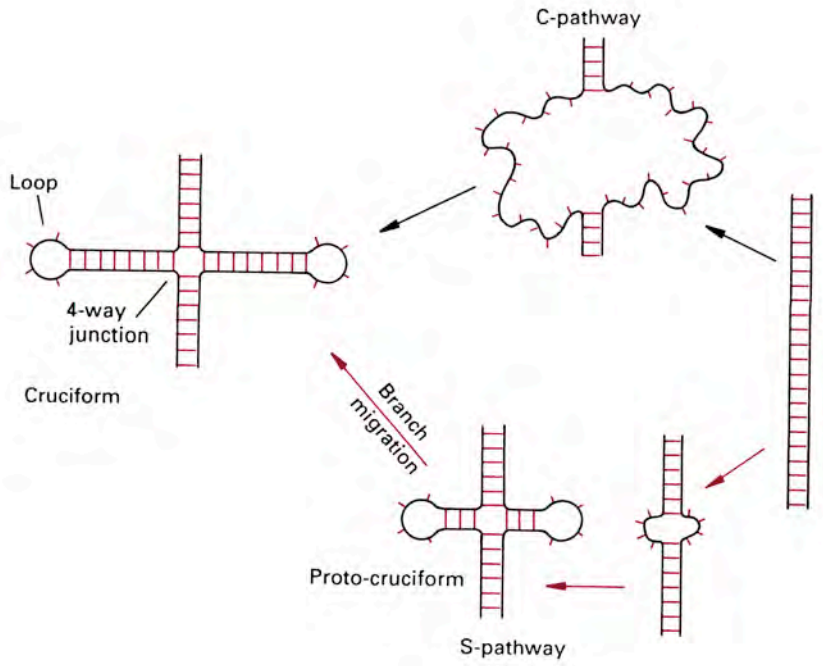


Fig. 2.26

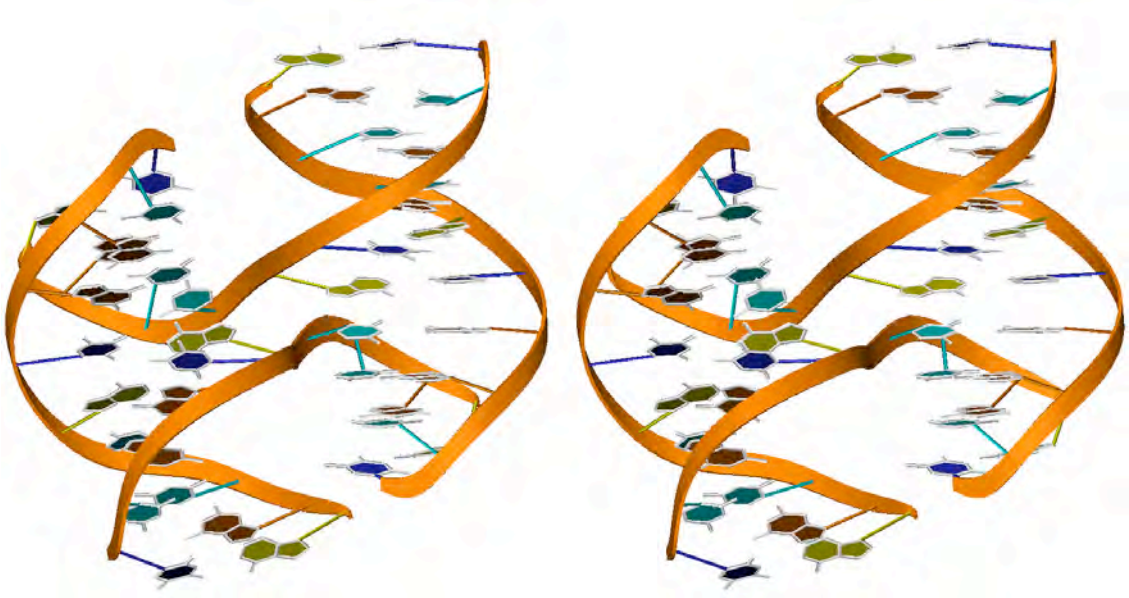


Fig. 2.27 (Stereo)

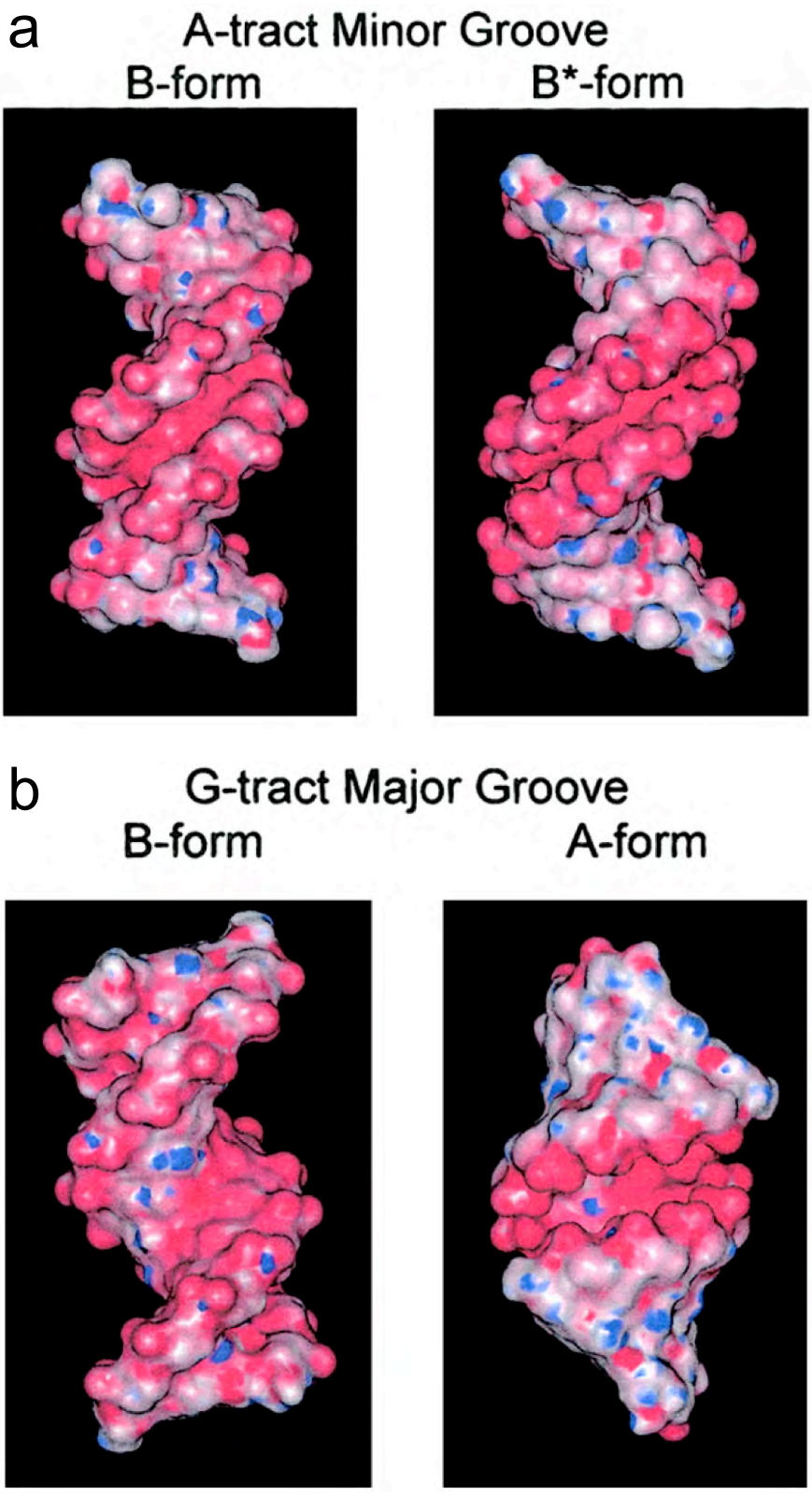

a
A-tract Minor Groove
B-form
B*-form
b
G-tract Major Groove
B-form
A-form

Fig. 2.28

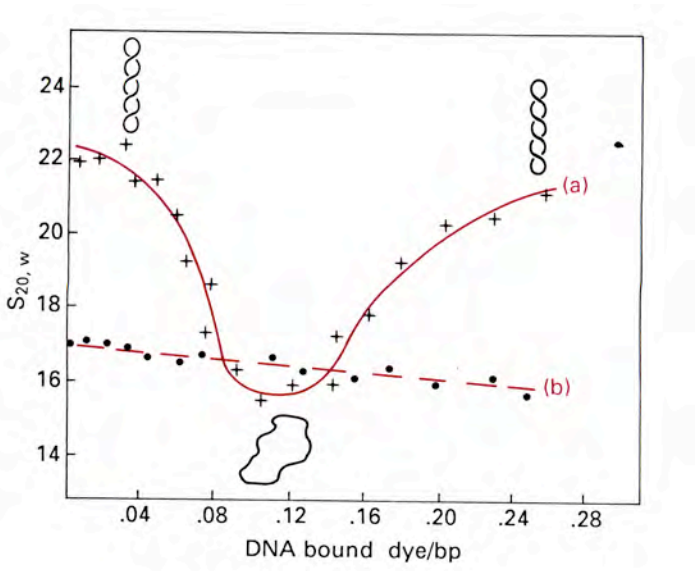


Fig. 2.29

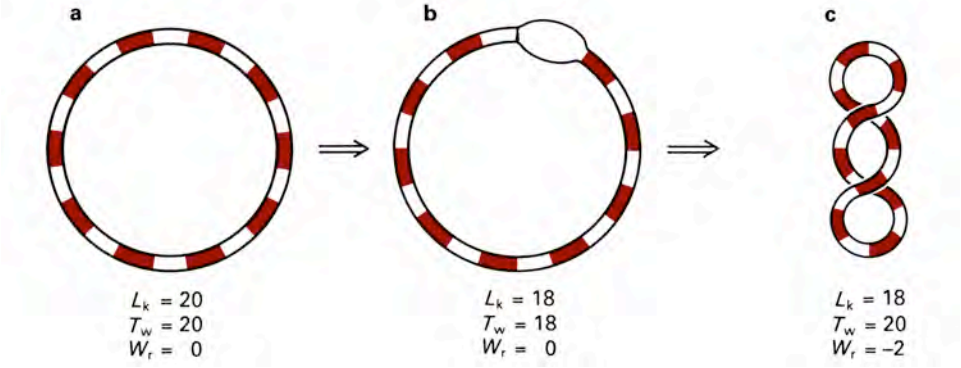


Fig. 2.30

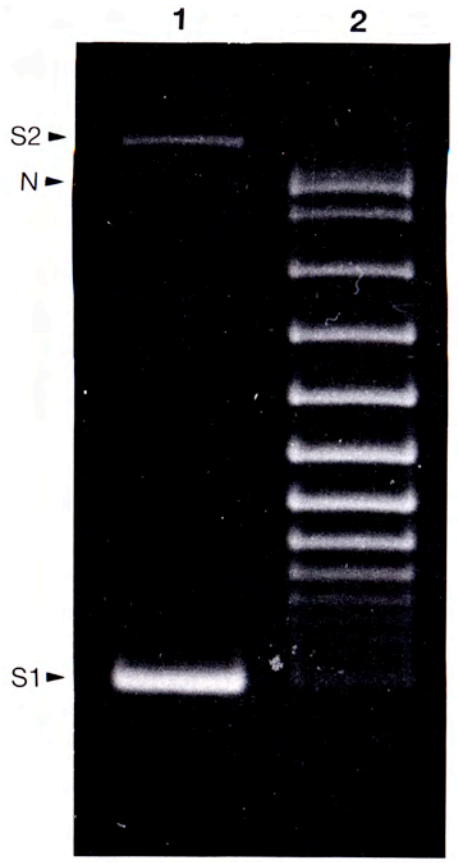


Fig. 2.31

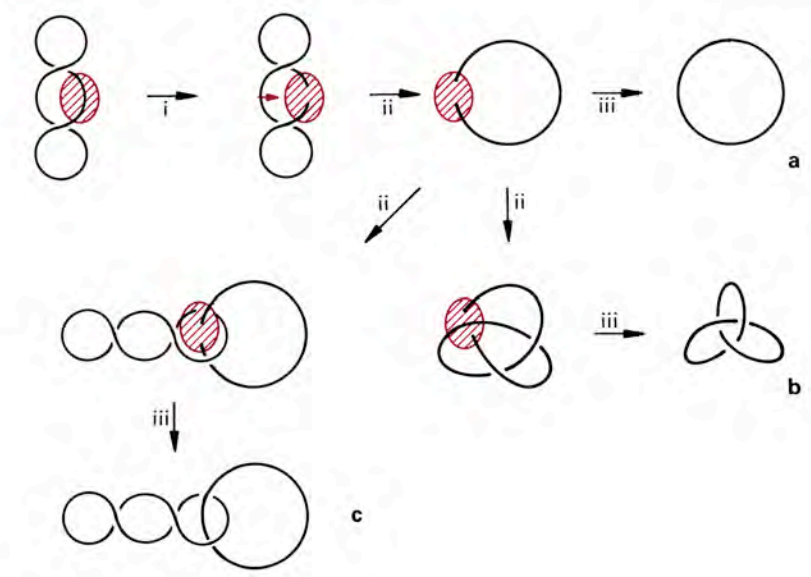


Fig. 2.32

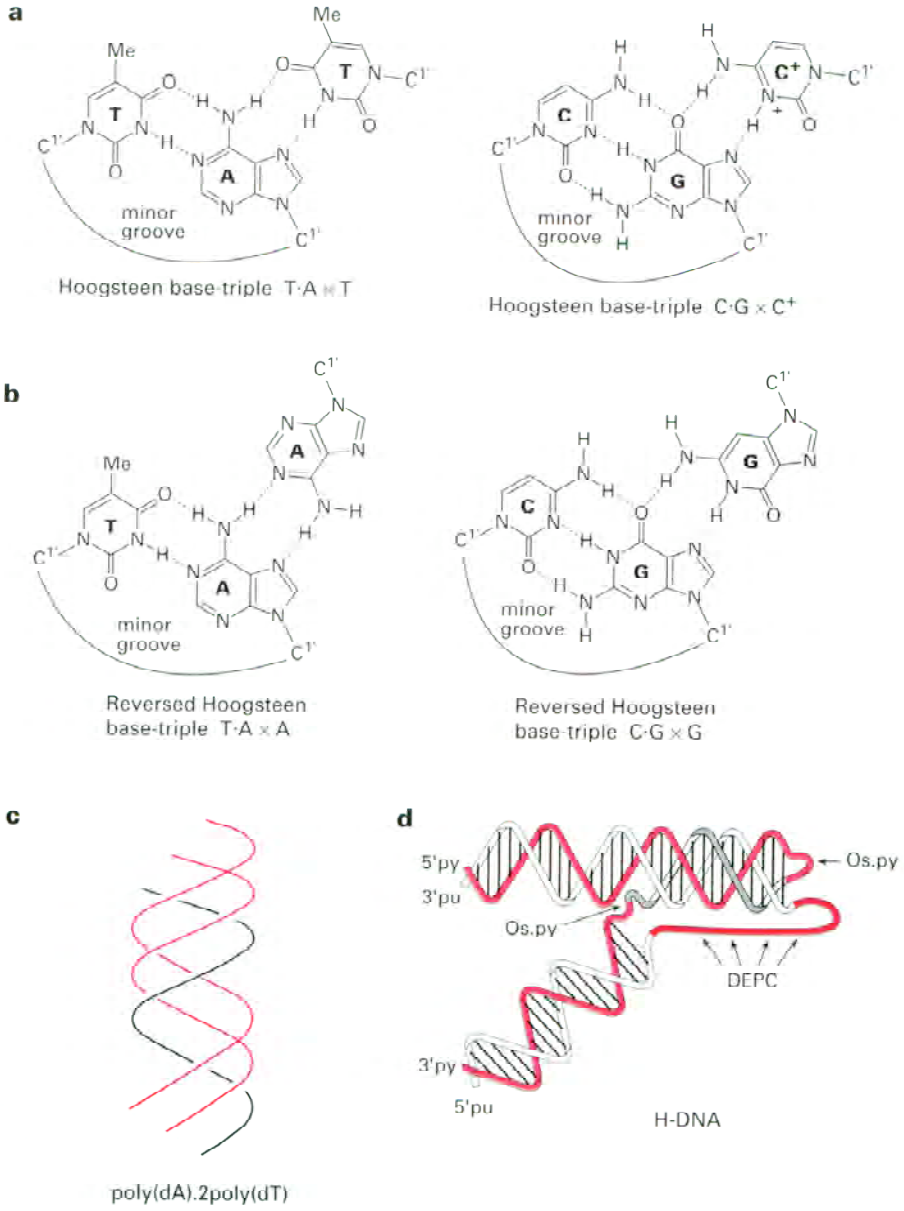

a
Me
T
A
minor groove
Hoogsteen base-triple T·A × T
C
G
C+
minor groove
Hoogsteen base-triple C·G × C+
b
A
T
A
minor groove
Reversed Hoogsteen base-triple T·A × A
C
G
G
minor groove
Reversed Hoogsteen base-triple C·G × G
c
poly(dA).2poly(dT)
d
5'py
3'pu
Os.py
Os.py
DEPC
3'py
5'pu
H-DNA

Fig. 2.33

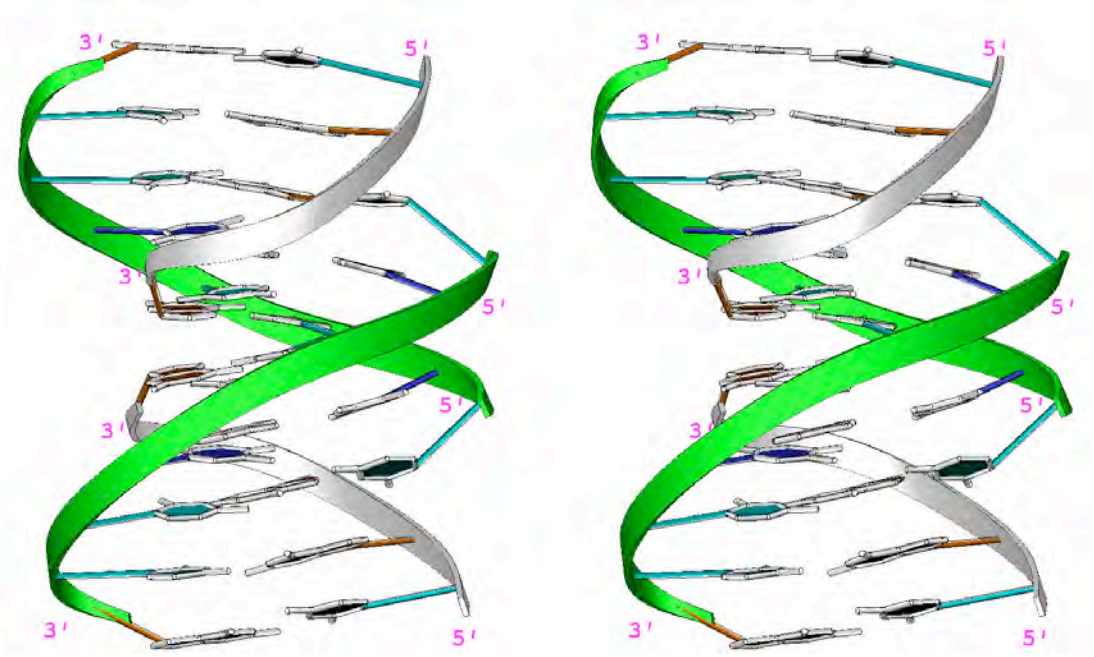


Fig. 2.34

Fig. 2.35 (Stereo)

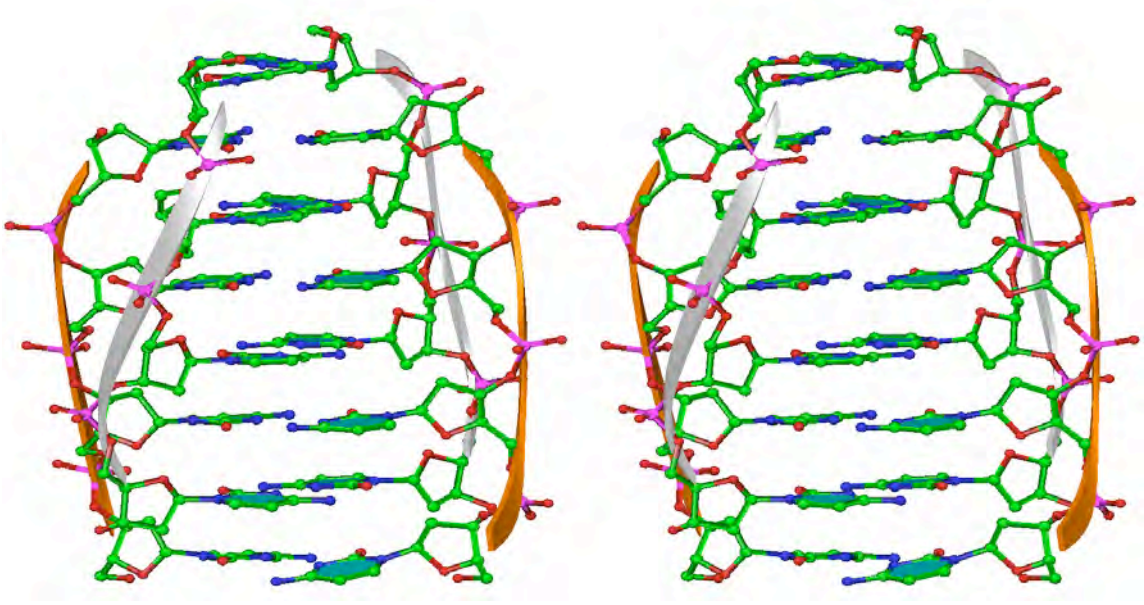


Fig. 2.36 (Stereo)

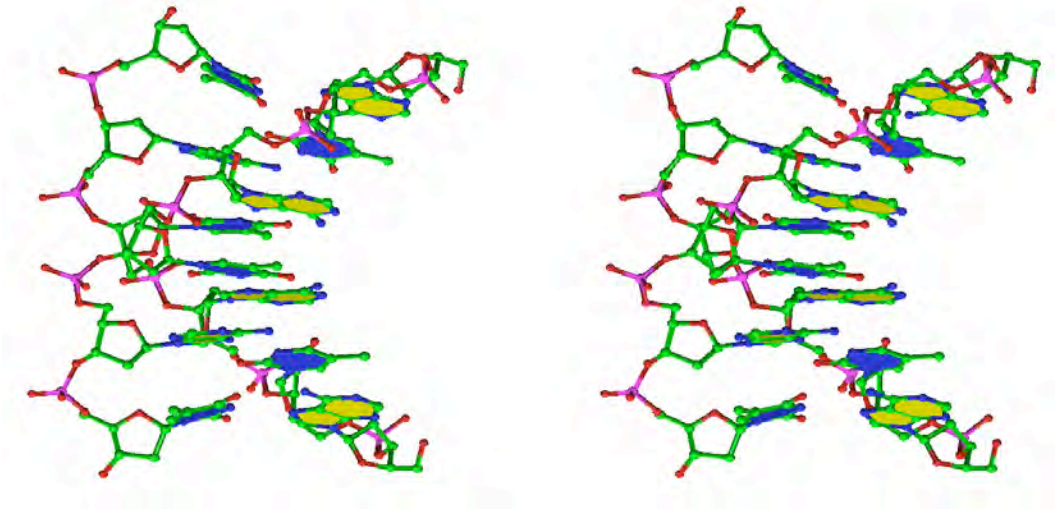


Fig. 2.38 (Stereo)

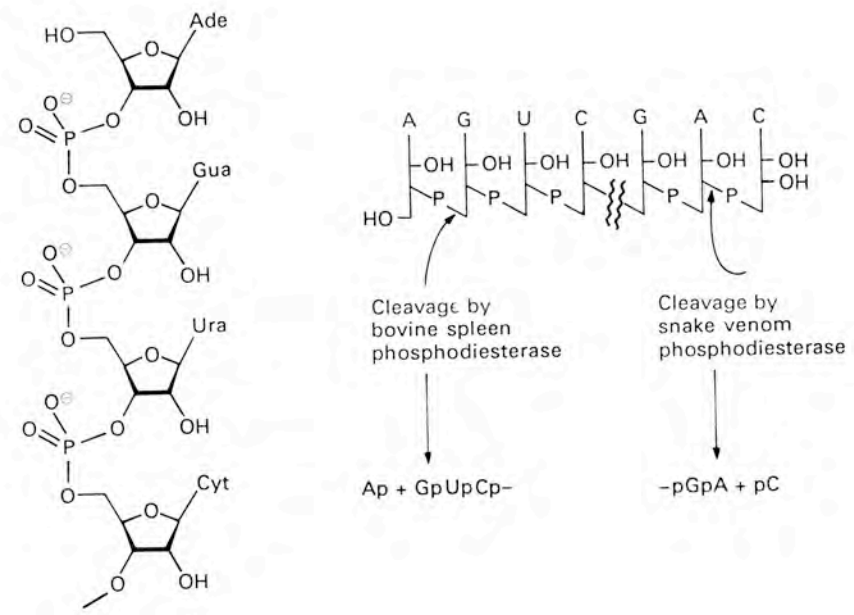


Fig. 2.38

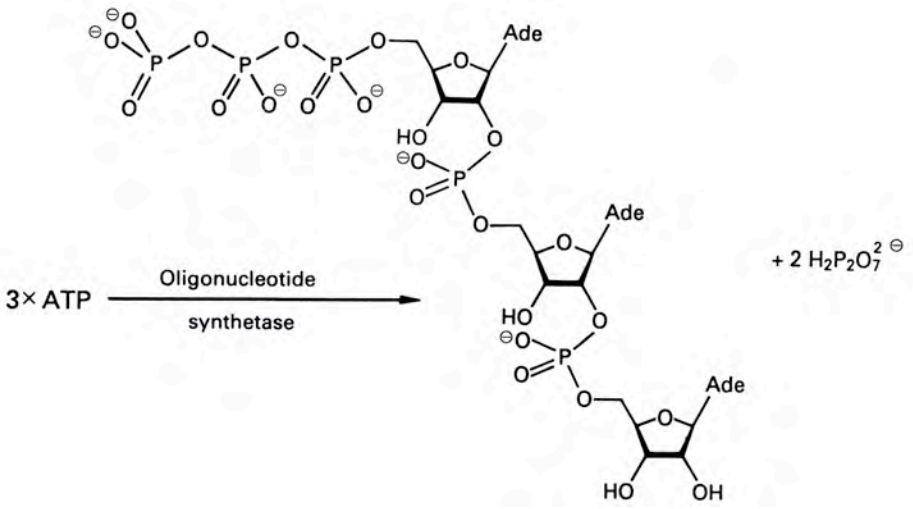


Fig. 2.39

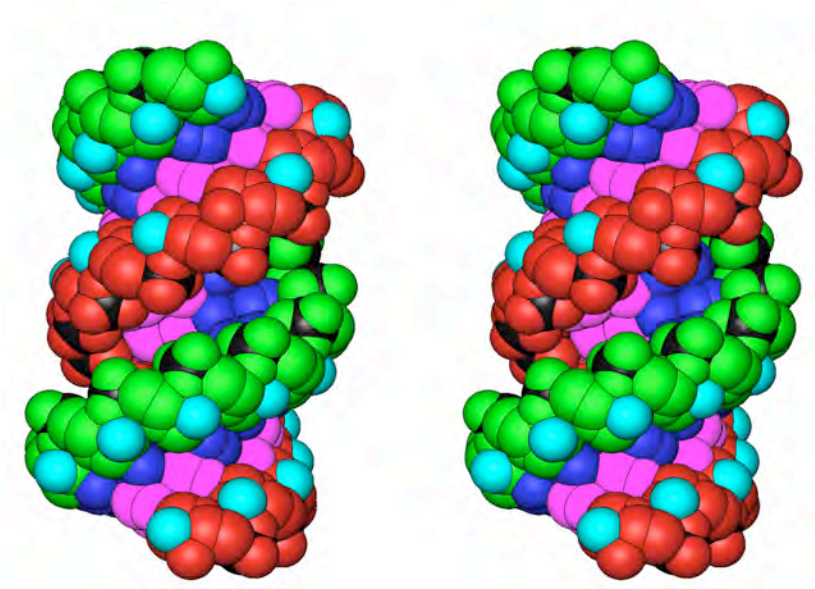


Fig. 2.40

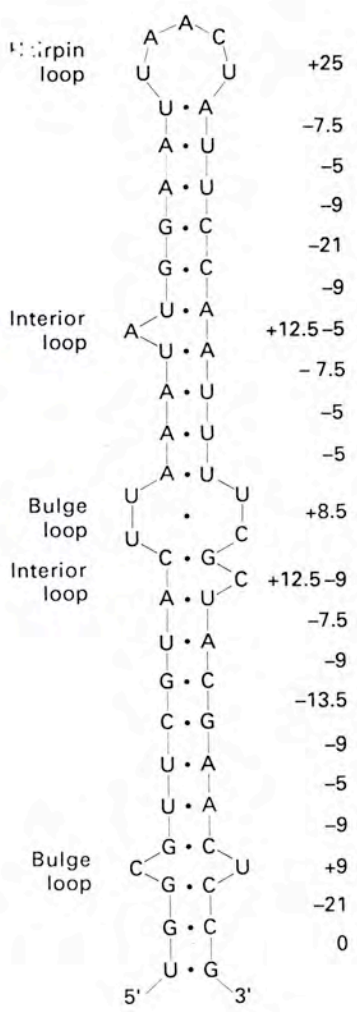

Hairpin loop
Interior loop
Bulge loop
Interior loop
Bulge loop
+25
–7.5
–5
–9
–21
–9
+12.5
–5
– 7.5
–5
–5
+8.5
+12.5
–9
–7.5
–9
–13.5
–9
–5
–9
+9
–21
0
5'
3'

Fig. 2.41

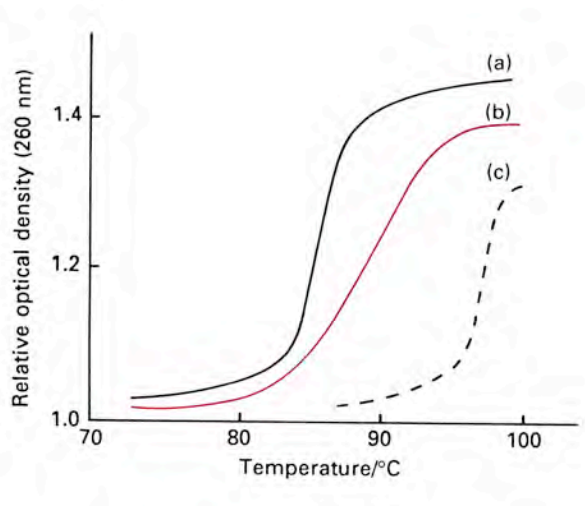


Fig. 2.42

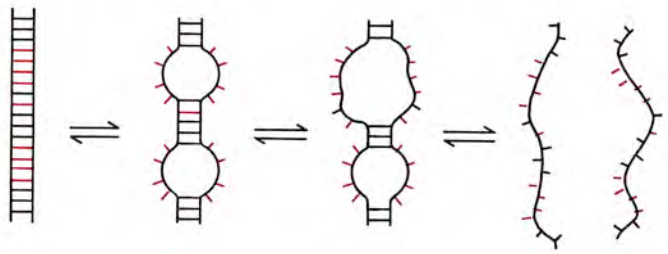


Fig. 2.43

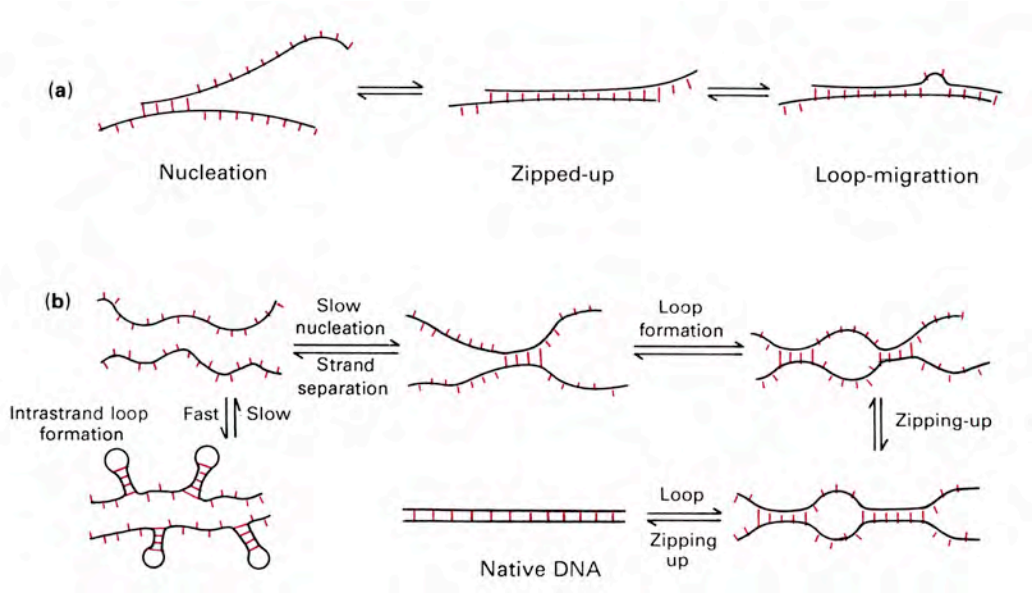


Fig. 2.44

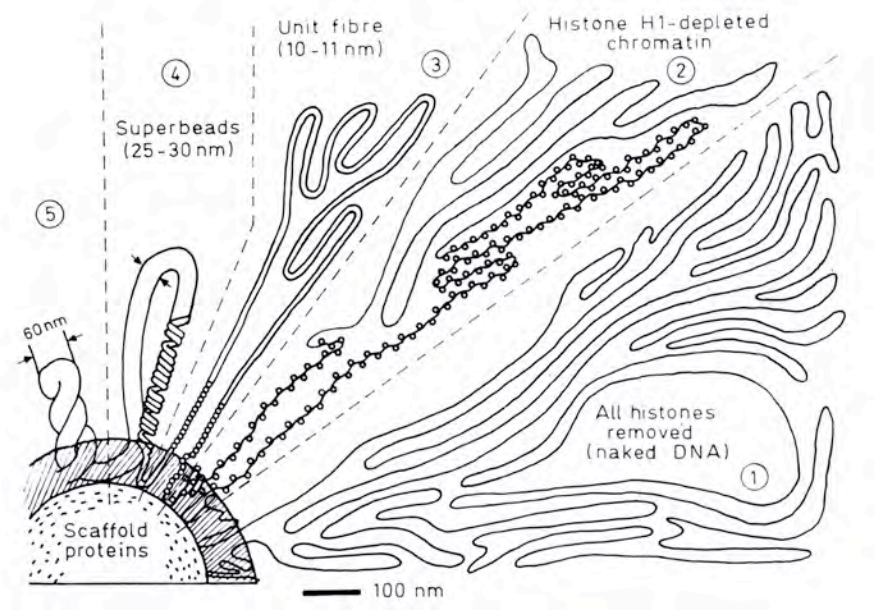


Fig. 2.45

Fig. 2.46