2

The Building-Blocks of DNA and RNA

2.1 Introduction

Chemical degradation studies in the early years of the twenty-first century on material extracted from cell nuclei established that the high molecular-weight "nucleic acid" was actually composed of individual acid units, termed nucleotides. Four distinct types were isolated – guanylic, adenylic, cytidylic and thymidylic acids. These could be further cleaved to phosphate groups and four distinct nucleosides. The latter were subsequently identified as consisting of a deoxypentose sugar and one of four nitrogen-containing heterocyclic bases. Thus, each repeating unit in a nucleic acid polymer comprises these three units linked together – a phosphate group, a sugar, and one of the four bases.

The bases are planar aromatic heterocyclic molecules and are divided into two groups – the pyrimidine bases thymine and cytosine and the purine bases adenine and guanine. Their major tautomeric forms are shown in Fig. 2.1. Thymine is replaced by uracil in ribonucleic acids, which also have an extra hydroxyl group at the 2′ position of their (ribose) sugar groups. The standard nomenclature for the atoms in nucleic acids, as approved by the International Union of Biochemistry, is shown in Figs. 2.1. and 2.2. Accurate bond length and angle geometries for all bases, nucleosides and nucleotides have been well established by X-ray crystallographic analyses. Structural surveys (Clowney *et al.*, 1996; Gelbin *et al.*, 1996) have calculated mean values for these parameters (which define their equilibrium values) from the most reliable structures in the Cambridge and Nucleic Acid Databases. These have been incorporated in several implementations of the AMBER and CHARMM force fields widely used in molecular mechanics and dynamics modelling, and in a number of computer packages for both crystallographic and NMR structural analyses (Parkinson *et al.*, 1996). Accurate crystallographic analyses, at very high resolution, can also directly yield quantitative information on the electron-density distribution in a molecule, and hence on individual partial atomic charges. These charges for nucleosides have hitherto been obtained by ab initio quantum mechanical calculations, but are also available experimentally for all four DNA nucleosides (Pearlman and Kim, 1990).

Individual nucleoside units are joined together in a nucleic acid in a linear manner, through phosphate groups attached to the 3′ and 5′ positions of the sugars (Fig. 2.2). Hence the full repeating unit in a nucleic acid is a 3′,5′-nucleotide.

Nucleic acid and oligonucleotide sequences use single-letter codes for the five unit nucleotides – A, T, G, C and U. The two classes of bases can be abbreviated as

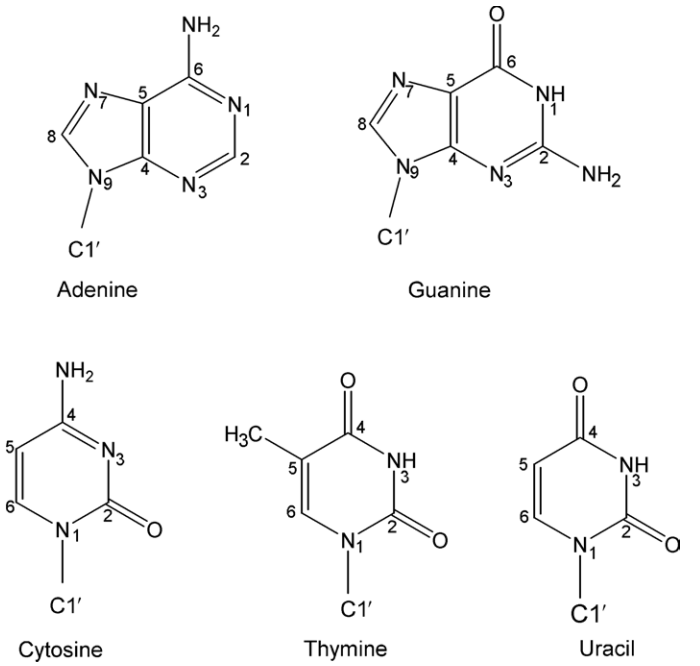

Figure 2.1 The five bases of DNA and RNA.

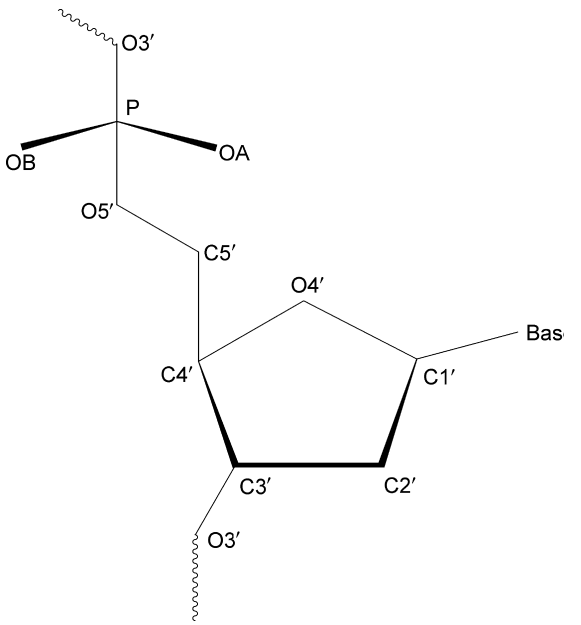

Figure 2.2 The organization of repeating units in a polynucleotide chain.

Y (pyrimidine) and R (purine). Phosphate groups are usually designated as p. A single oligonucleotide chain is conventionally numbered from the 5′ end, for example, ApGpCpTpTpG has the 5′ terminal adenosine nucleoside, with a free hydroxyl at its 5′ position and thus the 3′ end guanosine has a free 3′ terminal hydroxyl group. The letter p, to denote intervening phosphate groups, is sometimes omitted when a sequence is written down. Chain direction is sometimes emphasized with 5′ and 3′ labels. Thus an anti-parallel double-helical sequence can be written as:

$^{5'}$CpGpCpGpApApTpTpCpGpCpG
$^{3'}$GpCpGpCpTpTpApApGpCpGpC
or simply as
CGCGAATTCGCG

Structural publications on DNA usually prefix a sequence with "d", as in d(CGAT), to emphasize that the oligonucleotide is a deoxyribose one rather than being an oligoribonucleotide. The prefix "r" to denote ribonucleosides, ribonucleotides and their oligomers, is often used.

The bond between sugar and base is known as the glycosidic bond. Its stereochemistry is important. In natural nucleic acids the glycosidic bond is always β, that is the base is above the plane of the sugar when viewed onto the plane and therefore on the same face of the plane as the 5′ hydroxyl substituent (Fig. 2.3a). The absolute stereochemistry of other substituent groups on the deoxyribose sugar ring of DNA is defined such that when viewed end-on with the sugar ring oxygen atom O4′ at the rear (Fig. 2.4a), the hydroxyl group at the 3′ position is below the ring and the hydroxymethyl group at the 4′ position is above it. A unit nucleotide can have its

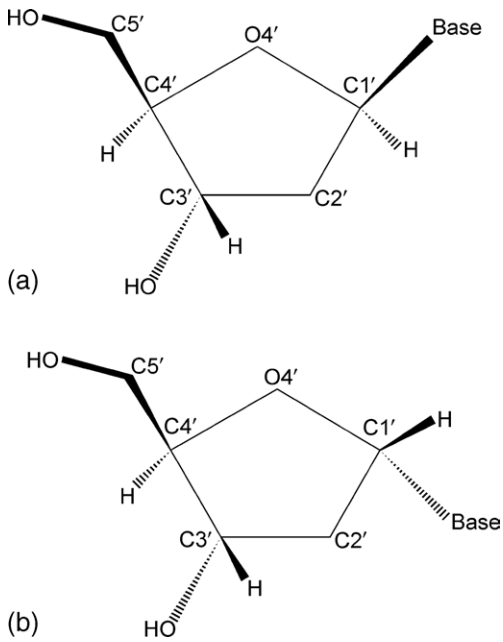

Figure 2.3 (a) The stereochemistry of a natural β-nucleoside. Solid bonds are coming out of the plane of the page, toward the reader. Dashed bonds are going away from the reader. (b) The stereochemistry of an α-nucleoside.

phosphate group attached either at the 3′ or 5′ ends, and is thus termed either a 3′- or a 5′- nucleotide. It is chemically possible to construct α-nucleosides and from them α-oligonucleosides, which have their bases in the "below" configuration relative to the sugar rings and their other substituents (Fig. 2.3b). These are much more resistant to nuclease attack than standard natural β-oligomers and have been used as antisense oligomers to mRNAs on account of their superior intracellular stability.

2.2 Base Pairing

The realization that the planar bases can associate in particular ways by means of hydrogen bonding was a crucial step in the elucidation of the structure of DNA. The important early experimental data of Chargaff showed that the molar ratios of adenine:thymine and cytosine:guanine in DNA were both unity. This led to the proposal by Watson and Crick that in each of these pairs the purine and pyrimidine bases are held together by specific hydrogen bonds, to form planar base pairs. In native, double-helical DNA the two bases in a base pair necessarily arise from two separate strands of DNA (with intermolecular hydrogen bonds) and so hold the DNA double helix together (Watson and Crick, 1953).

The adenine:thymine (A•T) base pair has two hydrogen bonds compared to the three in a guanine:cytosine (G•C) one (Fig. 2.4). (The initial Watson–Crick model assigned just two hydrogen bonds to the G•C pair, and the third hydrogen bond emerged some three years later (Wain-Hobson, 2006)). Fundamental to the Watson–Crick arrangement is that the sugar groups are both attached to the bases on the same side of the base pair. As will be seen in Chap. 3, this defines the mutual positions of the two sugar-phosphate strands in DNA itself. The two base pairs are required to be almost identical in dimensions by the Watson–Crick model. High resolution (0.8–0.9 Å) X-ray crystallographic analyses of the ribodinucleoside monophosphate duplexes r(GpC) and r(ApU) by A. Rich and colleagues in the early 1970s (Seeman *et al.*, 1976; Rosenberg *et al.*, 1976) established accurate geometries for these A•T and G•C base pairs (Table 2.1). These structure determinations showed that there are only small differences in size between the two types of base pairings, as indicated by the

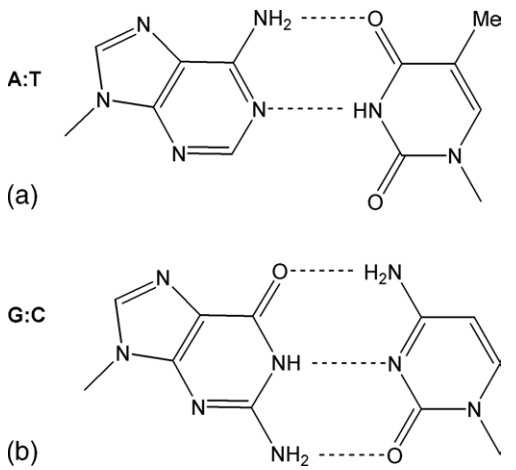

Figure 2.4 (a) A•T and (b) G•C base pairs, showing Watson–Crick hydrogen bonding.

Table 2.1 Hydrogen-Bond Distances (in Å) in Watson–Crick Base Pairs (Seeman *et al.*, 1976; Rosenberg *et al.*, 1976), in the Crystalline State. Estimated Standard Deviations are in Parentheses

U•A:	N3-H.....N1	2.835(8)
	O4.....H-N6	2.940(8)
C•G:	O2.....H-N2	2.86(1)
	N3.....H-N1	2.95(1)
	N4-H.....O6	2.91(1)

distance between glycosidic carbon atoms in a base pair. The C1′–C1′ distance in the G•C base pair structure is 10.67 Å, and 10.48 Å in the A•U-containing dinucleoside.

Theoretical studies on base pairing have enabled estimates to be made of the extra stability conferred on a G•C base pair by the extra hydrogen bond. Molecular dynamics simulations of both base pairs in an aqueous environment have given the free energies of A•T and G•C base pairs as –4.3 and –5.8 kcal/mol (Stofer, Chipot, and Lavery, 1999). An alternative, reversed Watson–Crick A•T base pairs involving a 180° rotation of one base, had similar stability. A more recent estimate, using ab initio quantum mechanical methods (Sponer, Jurecka and Hobza, 2004), has shown broad agreement with values calculated from the AMBER force field. This reinforces the reliability of the force-field approaches that are widely used and which emphasize van der Waals and electrostatic point-centered charge contributions, even though they do not take account of polarization and charge transfer effects. The interaction enthalpies of hydrated A•T and G•C base pairs have been calculated as 14.0 kcal/mol and 27.0 kcal/mol respectively (i.e. 7–9 kcal/mol per hydrogen bond), using AMBER together with density functional theory calculations (Liu, Wyttenbach, and Bowers, 2006). Other A•T base pair arrangements have been predicted by theory (Gould and Kollman, 1994) to be more stable than Watson–Crick pairing. However these alternatives would require significant changes in backbone conformation in order to be accommodated within a duplex. It is striking that these alternatives have not been observed in normal duplex DNA, suggesting that the requirements for optimal base stacking and minimal backbone distortions greatly favor the standard Watson–Crick arrangement.

2.3 Base and Base Pair Flexibility

The individual bases in a nucleic acid are flat, but base pairs (and consecutive bases on an individual strand), can show considerable flexibility. This flexibility is to some extent dependent on the nature of the bases and base pairs themselves, but is more related to their base-stacking environments. Thus descriptions of base morphology have become important in describing and understanding many sequence-dependent features and deformations of nucleic acids. The former features are often considered primarily at the dinucleoside local level, whereas longer-range effects such as helix bending, can also be analyzed at a more global level.

A number of rotational and translational parameters have been devised to describe these geometric relations between bases and base pairs, which were originally defined (10) in 1989 (the “Cambridge Accord”). These definitions, together with the

Cambridge Accord sign conventions, are given later. Confusingly, two distinct types of approaches have been reported in the literature to calculate these parameters (Lu, Babcock, and Olson, 1999) – the Cambridge Accord did not define a single unambiguous convention for their calculation. In one approach, the parameters are defined with respect to a global helical axis, which need not be linear. The other uses a set of local axes, one per dinucleotide step. Also, a variety of definitions of local and global axes have been used. The overall effect for most undistorted structures is fortunately that only a minority of parameters appear to have very different values depending on the method of calculation, using a number of the widely available programs (see later). Before detailing these, we introduce the parameters themselves. The major and minor grooves are the indentations in nucleic acid double helices formed as a consequence of the asymmetry of base pairs. The C1′–N9 (purine) and C1′–N1 (pyrimidine) base-sugar bonds are by convention on the minor-groove side, so that the C6/N7 (purine) and C4 (pyrimidine) base atoms and their substituents are on the major-groove side. The grooves and their characteristics are described further in Chap. 3.

(i) For individual base pairs:
   (a) **Propeller twist** (ω) between bases is the dihedral angle between normals to the bases, when viewed along the long axis of the base pair (Fig. 2.5). The angle has a negative sign under normal circumstances, with a clockwise rotation of the nearer base when viewed down the long axis. The long axis for a purine–pyrimidine base pair is defined as the vector between the C8 atom of the purine and the C6 of a pyrimidine in a Watson–Crick base pair. Analogous definitions can be applied to other nonstandard base pairings in a duplex including purine–purine and pyrimidine–pyrimidine ones.
   (b) **Buckle** (κ) is the dihedral angle between bases, along their short axis, after propeller twist has been set to 0° (Fig. 2.6). The sign of buckle is defined as

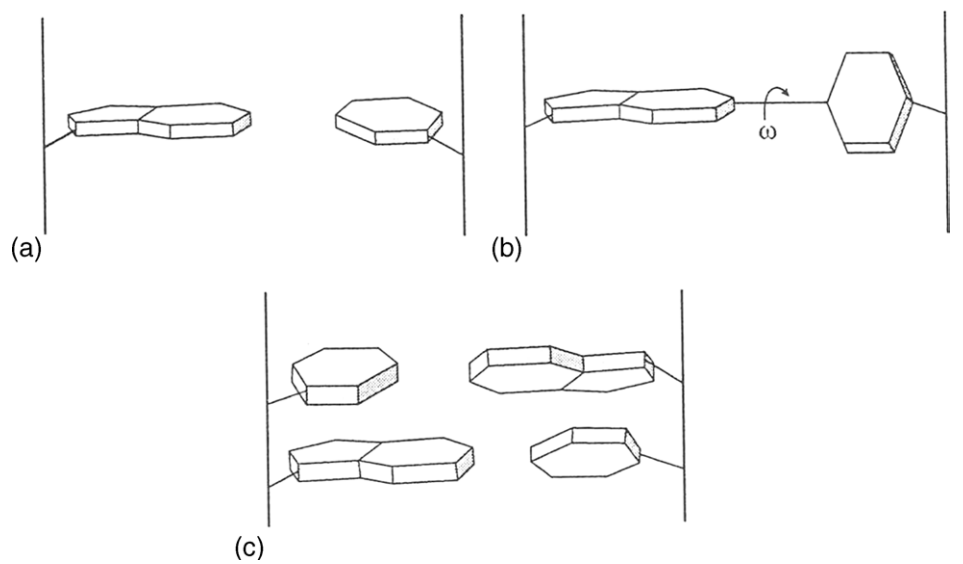

Figure 2.5 Schematic views of a base pair with (a) zero and (b) high propeller twist. Panel (c) shows the effects of propeller twist on two successive base pairs.

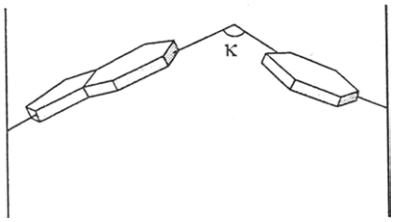

Figure 2.6 A schematic view of base-pair buckle.

positive if the distortion is convex in the direction 5′→3′ of strand 1. The change in buckle for succeeding steps, termed **cup**, has been found to be a useful measure of changes along a sequence. Cup is defined as the difference between the buckle at a given step, and that of the preceding one.

(c) **Inclination** (η) is the angle between the long axis of a base pair and a plane perpendicular to the helix axis. This angle is defined as positive for right-handed rotation about a vector from the helix axis toward the major groove.

(d) **X** and **Y displacements** define translations of a base pair within its mean plane in terms of the distance of the midpoint of the base pair long axis from the helix axis. X displacement is toward the major-groove direction, when it has a positive value. Y displacement is orthogonal to this, and is positive if toward the first nucleic acid strand of the duplex.

(ii) For base pair steps:

(e) **Helical twist** (Ω) is the angle between successive base pairs, measured as the change in orientation of the C1′–C1′ vectors on going from one base pair to the next, projected down the helix axis (Fig. 2.7). For an exactly repetitious double helix, helical twist is 360°/n, where n is the unit repeat defined above.

(f) **Roll** (ρ) is the dihedral angle for rotation of one base pair with respect to its neighbor, about the long axis of the base pair. A positive roll angle opens up a base pair step toward the minor groove (Fig. 2.8). **Tilt** (τ) is the corresponding dihedral angle along the short (i.e. *x*-axis) of the base pair.

(g) **Slide** is the relative displacement of one base pair compared to another, in the direction of nucleic acid strand one (i.e. the Y displacement), measured between the midpoints of each C6–C8 base pair long axis.

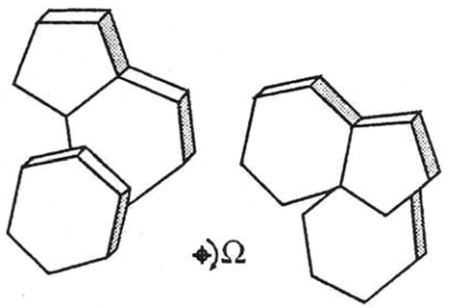

Figure 2.7 View down two successive base pairs, showing the helical twist angle between them.

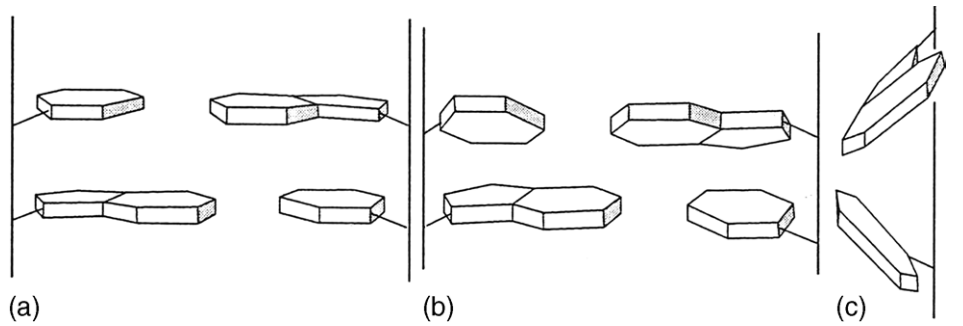

Figure 2.8 Views of two successive base pairs, (a) with 0° roll angle between them, and (b) with a positive roll angle. Panel (c) shows a view of positive roll along the long axis of the base pairs.

A standardized coordinate reference frame for the calculation of these parameters has been proposed (Lu and Olson, 1999), and have been endorsed by the successor to the Cambridge Accord, the 1999 Tsukuba Accord (Olson *et al.*, 2001). Details of the reference frame are also available on the Nucleic Acid Database web site at http://ndbserver.rutgers.edu/standards/supl_inf.html.

This reference frame is unambiguous and has the advantage of being able to produce values for the majority of local base-pair and base-step parameters which are almost independent of the algorithm used. A notable exception is rise, which is especially sensitive to the definition of origin and to small changes in buckle and roll. The right-handed reference frame used is shown in Fig. 2.9. It has the *x*-axis directed toward the major groove along the pseudo two-fold axis of an idealized Watson–Crick base pair (shown as •). The *y*-axis is along the long axis of the base pair, parallel to the C1′–C1′ vector. The position of the origin is clearly dependent on the geometry of the bases and the base pair. These have been taken from the published compilations (Clowney *et al.*, 1996; Gelbin *et al.*, 1996).

The most widely-used computer programs for base and base-pair morphology calculations, all of which use the above parameter definitions, are:

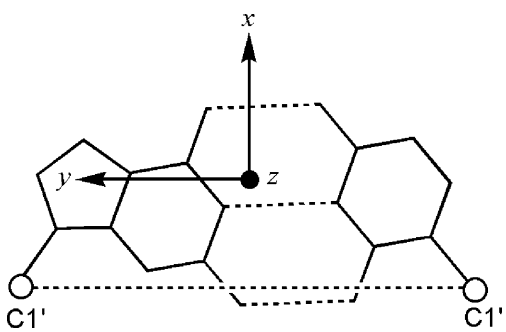

Figure 2.9 Reference frame for idealized helical DNA, showing the axis origin at (•), along the pseudo two-fold axis between two successive base pairs.

- *CURVES* (Lavery and Sklenar, 1988, 1989), which calculates an optimized global helix axis which can be (and invariably is), curved. This is especially useful for irregular structures. The reference frame is defined at the base rather than the base pair level. Local parameters can also be obtained with this program, and irregular nonstandard base-pairing arrangements can be accommodated.
- *FREEHELIX,* the successor to the original helical parameter program NEWHELIX, calculates an optimum linear helical axis, so that the derived parameters are basically global ones (Dickerson, 1998). Local bending can also be calculated (Goodsell and Dickerson, 1994).
- *RNA* (Babcock, Pednault, and Olson, 1994) uses the reference frame subsequently adopted by the Tsukuba Accord, and calculates local helical parameters. It can also be used for nonstandard bases.
- *CEHS* (Lu, El Hassan, and Hunter, 1997) also focuses on local parameters, and like FREEHELIX, uses the C6–C8 base-pair vector as the y-axis for the reference frame. Roll and tilt are commuted into a single variable RollTilt (Γ).

Readers of crystallographic or NMR structure analysis literature should be aware of whether local or global axis definitions have been used in parameter calculation. Ambiguities can be clarified by use of the NDB, which can produce tables of base helical parameters calculated with all of these methods.

2.4 Sugar Puckers

The five-membered deoxyribose sugar ring in DNA is inherently nonplanar. This nonplanarity is termed puckering**.** The precise conformation of a deoxyribose ring can be completely specified by the five endocyclic torsion angles within it (Fig. 2.10). The ring puckering arises from the effect of nonbonded interactions between substituents at the four ring carbon atoms – the energetically most stable conformation for the ring has all substituents as far apart as possible. Thus different substituent atoms would be expected to produce differing types of puckering. The puckering can be described by either

(i) A simple qualitative description of the conformation in terms of atoms deviating from ring coplanarity; or
(ii) Precise descriptions in terms of the ring internal torsion angles.

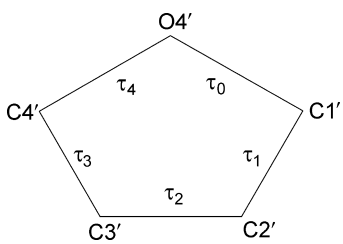

Figure 2.10 The five internal torsion angles in a ribose ring.

In principle, there is a continuum of interconvertible puckers, separated by energy barriers. These various puckers are produced by systematic changes in the ring torsion angles. The puckers can be succinctly defined by the parameters P and τ_m (Altona and Sundaralingam, 1972). The value of P, the phase angle of pseudorotation, indicates the type of pucker since P is defined in terms of the five torsion angles τ_0–τ_4:

$$\tan P = \frac{(\tau_4 + \tau_1) - (\tau_3 + \tau_0)}{2 * \tau_2 * (\sin 36° + \sin 72°)}$$

and the maximum degree of pucker, τ_m, by

$$\tau_m = \frac{\tau_2}{\cos P}$$

The pseudorotation phase angle can take any value between 0° and 360°. If τ_2 has a negative value, then 180° is added to the value of P. The pseudorotation phase angle is commonly represented by the pseudorotation wheel, which indicates the continuum of ring puckers (Fig. 2.11). Values of τ_m indicate the degree of puckering of the ring; typical experimental values from crystallographic studies on mononucleosides are in the range 25–45°. The five internal torsion angles are not independent of each other, and so to a good approximation any one angle τ_j can be represented in terms of just two variables:

$$\tau_j = \tau_m \cos[P + 0.8\pi(j - 2)]$$

An on-line interactive facility (PROSIT) is now available to calculate pseudorotation parameters (Sun *et al.*, 2004) at http://www.cactus.nci.nih.gov/prosit/.

A large number of distinct deoxyribose ring pucker geometries have been observed experimentally, by X-ray crystallography and NMR techniques. When one ring atom is out of the plane of the other four, the pucker type is an envelope one. More commonly, two atoms deviate from the plane of the other three, with these two either

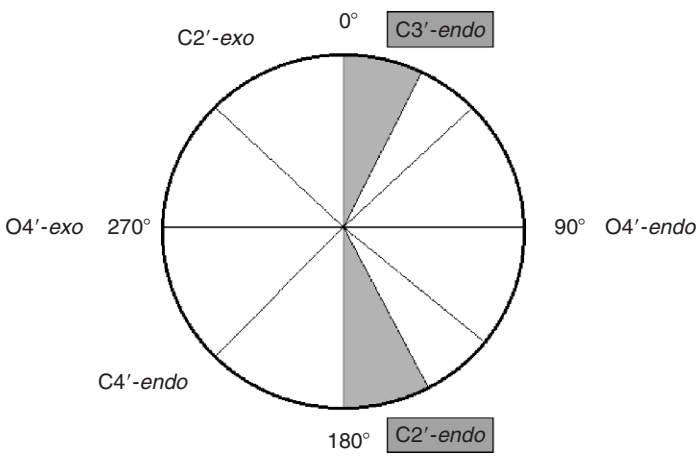

Figure 2.11 The pseudorotation wheel for a deoxyribose sugar. The shaded areas indicate the preferred ranges of the pseudorotation angle for the two principal sugar conformations.

side of the plane. It is usual for one of the two atoms to have a larger deviation from the plane than the other, resulting in a twist conformation. The direction of atomic displacement from the plane is important. If the major deviation is on the same side as the base and C4′–C5′ bond, then the atom involved is termed *endo*. If it is on the opposite side, it is called *exo*. The most commonly observed puckers in crystal structures of isolated nucleosides and nucleotides are either close to C2′-*endo* or C3′-*endo* types (Fig. 2.12a,b). In practice, these pure envelope forms are rarely observed,

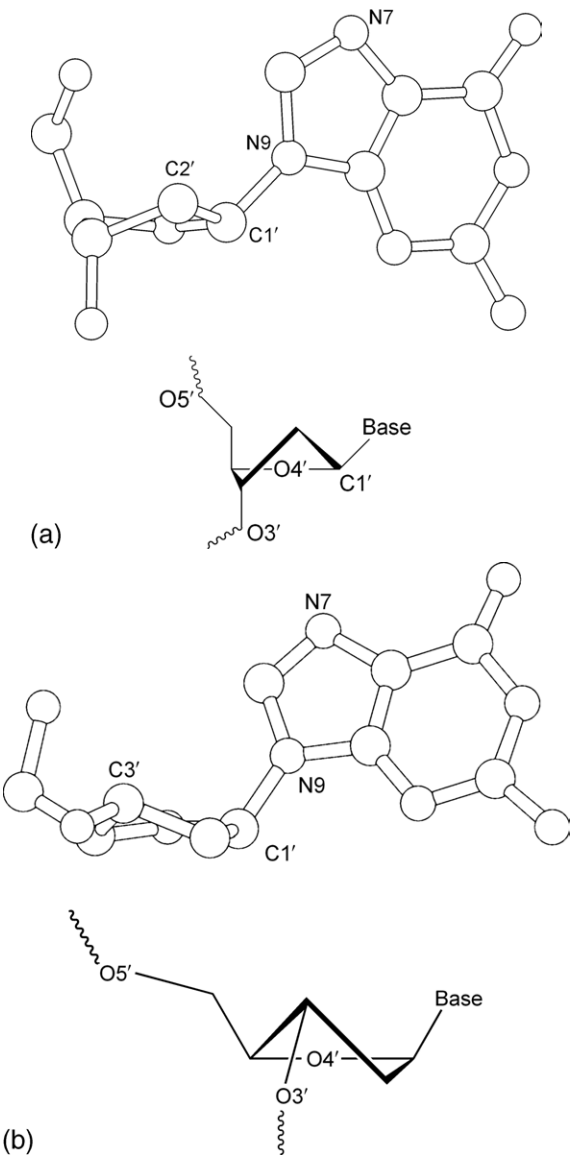

Figure 2.12 (a) C2′-*endo* sugar puckering for a guanosine nucleoside, viewed along the plane of the sugar ring. (b) C3′-*endo* sugar puckering for a guanosine nucleoside, with the sugar viewed in the same direction as in panel (a).

largely because of the differing substituents on the ring. Consequently the puckers are then best described in terms of twist conformations. When the major out-of-plane deviation is on the *endo* side, there is a minor deviation on the opposite, *exo* side. The convention used for describing a twist deoxyribose conformation is that the major out-of-plane deviation is followed by the minor one, for example C2′-*endo*, C3′-*exo*. The C2′-*endo* family of puckers have P values in the range of 140–185°; in view of their position on the pseudorotation wheel, they are sometimes termed S (south) conformations. The C3′-*endo* domain has P values in the range of –10 to +40°, and its conformation is termed N (north). These more geographical terms are mostly used by NMR spectroscopists.

The pseudorotation wheel implies that deoxyribose puckers are free to interconvert. In practice, there are energy barriers between major forms. The exact size of these barriers has been the subject of considerable study (Olson, 1982a; Olson and Sussman, 1982). The consensus is that the barrier height is dependent on the route around the pseudorotation wheel. For interconversion of C2′-*endo* to C3′-*endo* the preferred pathway is via the O4′-*endo* state, with a barrier of 2–5 kcal/mole found from an analysis of a large body of experimental data (Olson and Sussman, 1982), and a somewhat smaller (potential energy) value of 1.5 kcal/mole from a molecular dynamics study (Harvey and Prabhakaran, 1986). The former value, being an experimental one, represents the total free energy for interconversion. These values also broadly concur with those found by application of various ab initio methods, which have given values of 2.5–4.0 kcal/mole (Foloppe, Nilsson, and MacKerell, 2001).

Relative populations of puckers can be monitored directly by NMR measurements of the ratio of coupling constants between H1′–H2′ and H3′–H4′ protons. These show that in contrast to the "frozen-out" puckers found in the solid state structures of nucleosides and nucleotides, there is rapid interconversion in solution. Nonetheless, the relative populations of the major puckers are dependent on the type of base attached. Purines show a preference for the C2′-*endo* pucker conformational type whereas pyrimidines favor C3′-*endo*. Deoxyribose nucleosides are primarily (>60%) in the C2′-*endo* form and ribonucleosides favor C3′-*endo*. The latter are significantly more restricted in their mobility; this has significance for the structures of oligoribonucleotides (see Chap. 6). These differences in puckering equilibria and hence in their relative populations in solution and in molecular dynamics simulations, are reflected in the patterns of puckers found in surveys of crystal structures (Murray-Rust and Motherwell, 1978). Again, this is a demonstration of the complementarity of information provided by the different structural techniques. Sugar pucker preferences have their origin in the nonbonded interactions between substituents on the sugar ring, and to some extent on their electronic characteristics. For example, the C3′-*endo* pucker (Fig. 2.7b) would have hydroxyl substituents at the 2′ and 3′ positions further apart than with C2′-*endo* pucker; hence the preference of the former by ribonucleosides.

Numerous crystallographic and NMR studies have found correlations between sugar pucker and several backbone conformational variables, both in isolated nucleosides/nucleotides and in oligonucleotide structures. These are discussed later in this chapter. Changes in sugar pucker are important determinants of oligo- and polynucleotide structure because they can alter the orientation of C1′, C3′ and C4′ substituents, resulting in major changes in backbone conformation and overall structure, as indeed is found (Chap. 3).

2.5 Conformations About the Glycosidic Bond

The glycosidic bond links a deoxyribose sugar and a base, being the C1′–N9 bond for purines and the C1′–N1 bond for pyrimidines. The torsion angle χ around this single bond can in principle adopt a wide range of values, although as will be seen, structural constraints result in marked preferences being observed. Glycosidic torsion angles are defined in terms of the four atoms:

O4′–C1′–N9–C4 for purines
O4′–C1′–N1–C2 for pyrimidines

Theory has predicted two principal low-energy domains for the glycosidic angle, in accord with experimental findings for a large number of nucleosides and nucleotides. The *anti* conformation has the N1, C2 face of purines and the C2, N3 face of pyrimidines directed away from the sugar ring (Fig. 2.13a) so that the hydrogen atoms attached to C8 of purines and C6 of pyrimidines, are lying over the sugar ring. Thus, the Watson–Crick hydrogen-bonding groups of the bases are directed away from the sugar ring. These orientations are reversed for the *syn* conformation, with these hydrogen-bonding groups now oriented toward the sugar and especially its O5′ atom (Fig. 2.13b). A number of crystal structures of *syn* purine nucleosides have found hydrogen bonding between the O5′ atom

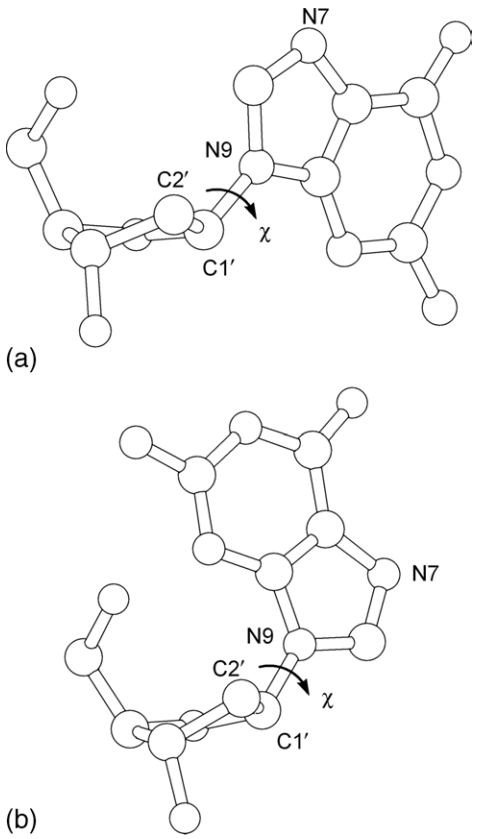

Figure 2.13 (a) A guanosine nucleoside with the glycosidic angle χ set in an *anti* conformation. (b) A guanosine nucleoside in the *syn* conformation.

Table 2.2 Energy Differences (in kcal/mole) Between Glycosidic Angle Conformers for B-form DNA Nucleotides, Calculated with the AMBER Program

Nucleotides	Energy difference
A(*anti*) > A(*syn*)	0.3
T(*anti*) > T(*syn*)	1.7
G(*syn*) > G(*anti*)	3.3
C(*anti*) > C(*syn*)	1.8

and the N3 base atom, which would stabilize this conformation. Otherwise, for purines, the *syn* conformation is slightly less preferred than the *anti*, on the basis of fewer nonbonded steric clashes in the latter case. The principal exceptions to this rule are guanosine-containing nucleotides, which have a small preference for the *syn* form because of favorable electrostatic interactions between the exocyclic N2 amino group of guanine and the 5′ phosphate atom. For pyrimidine nucleotides, the *anti* conformation is preferred over the *syn*, because of unfavorable contacts between the O2 oxygen atom of the base and the 5′-phosphate group. The results of molecular-mechanics energy minimizations on all four DNA nucleotides in both *syn* and *anti* forms (using the AMBER all-atom force field), are fully in accord with these observations (Table 2.2).

The sterically preferred ranges for the two domains of glycosidic angles are:

Anti: $-120 > \chi > 180°$
Syn: $0 < \chi < 90°$

Values of χ in the region of about −90° are often described as "high *anti*". There are pronounced correlations between sugar pucker and glycosidic angle, which reflect the changes in non-bonded clashes produced by C2′-*endo* versus C3′-*endo* puckers. Thus, *syn* glycosidic angles are not found with C3′-*endo* puckers due to steric clashes between the base and the H3′ atom, which points toward the base in this pucker mode.

2.6 The Backbone Torsion Angles and Correlated Flexibility

The phosphodiester backbone of an oligonucleotide has six variable torsion angles (Fig. 2.14), designated $\alpha \ldots \zeta$, in addition to the five internal sugar torsions τ_0–τ_4 and the glycosidic angle χ. As will be seen, a number of these have highly correlated values (and therefore correlated motions in a solution environment). Steric considerations alone dictate that the backbone angles are restricted to discrete ranges (Sundaralingam, 1969; Olson, 1982b) (Fig. 2.15), and are accordingly not free to adopt any value between 0° and 360°. Fig. 2.15 uses a conformational wheel to show these preferred values, which are directly readable from their positions around the wheel. The fact that angles α, β, γ, and ζ each have three allowed ranges, together with the broad range for angle ε that includes two staggered regions, leads to a large number of possible low-energy conformations for the unit nucleotide, especially when glycosidic angle and sugar pucker flexibility is taken into account. In reality, only a small number of DNA oligonucleotide and polynucleotide structural classes have actually been observed out of this large range of possibilities; this is doubtless in large part due to the restraints imposed by Watson–Crick base pairing on the backbone conformations when two DNA strands are intertwined together.

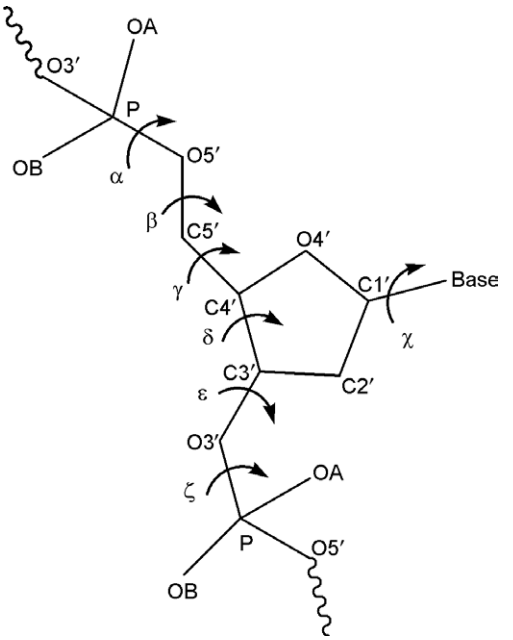

Figure 2.14 The backbone torsion angles in a unit nucleotide. Each rotatable bond is indicated by a curved arrow.

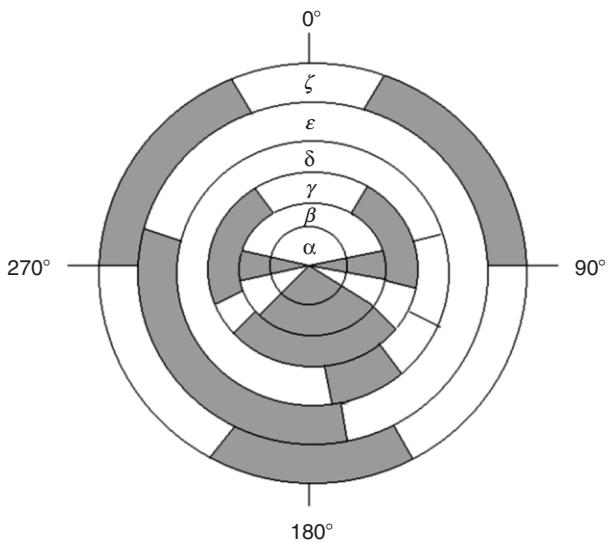

Figure 2.15 Conformational wheel showing the allowed ranges of backbone torsion angles (shaded) in nucleosides, nucleotides, deoxyoligonucleotides, and deoxypolynucleotides.

In contrast, crystallographic and NMR studies on a large number of standard and modified mononucleosides and nucleotides have shown their considerably greater conformational diversity, in accord with the possibilities indicated in Fig. 2.15. For these, backbone conformations in the solid state and in solution are not always in agreement; the requirements for efficient packing in the crystal can often overcome

the modest energy barriers between different values for a particular torsion angle. Many large RNA molecules are characterized by a wide range of base–base interactions, which can therefore adopt a wide variety of backbone conformations (Schneider, Moravek, and Berman, 2004) (see Chap. 6 for a detailed discussion of RNA conformational variability).

A common convention for describing these backbone angles is to term values of ~60° as *gauche*$^+$(g^+), –60° as *gauche* $^-$ (g^-) and ~180° as *trans* (t). Thus, for example, angles α (about the P–O5′ bond) and γ (the exocyclic angle about the C4′–C5′ bond), can be in the g^+, g^-, or t conformations. The two torsion angles around the phosphate group itself, α and ζ, have been found to show a high degree of flexibility in various dinucleoside crystal structures, with the tg^-, g^-g^-, and g^+g^+ conformations all having been observed (Kim *et al.,* 1973). As will be described in Chap. 3, only the g^-g^- conformation can place successive nucleotide units in arrangements that have their bases in potential hydrogen-bonding positions with respect to a second nucleotide strand. Thus, this is the phosphate conformation for DNA and RNA double helices. The torsion angle β, about the O5′-C5′ bond, is almost always found to be *trans*. All three possibilities for the γ angle have been observed in nucleoside crystal structures, although the g^+ conformation predominates in right-handed oligo- and polynucleotide double helices. The *trans* conformation for γ places the 5′-phosphate group in quite a distinct position with respect to the deoxyribose ring. The torsion angle δ around the C4′–C3′ bond adopts values that relate to the pucker of the sugar ring, since the internal ring torsion angle τ_3, (also around this bond), has a value of about 35° for C2′-*endo* and about 40° for C3′-*endo* puckers. δ is about 75° for C3′-*endo* and about 150° for C2′-*endo* puckers.

An alternative nomenclature for torsion angle ranges used by some workers in the nucleic acid field is that common in organic chemistry (the Klyne-Prelog system). In this, the *syn*(s) designation is given to angles clustered around 0°, and *anti* (a) for those around 180°. Intermediate angles are defined as ± *synclinal* (± *sc*) for around ± 60°, and ± *anticlinal* (± *ac*) for around ± 120°.

There are a number of well-established correlations involving pairs of these backbone torsion angles, as well as sugar pucker and glycosidic angle. These have been found (Sundaralingam, 1969) in mononucleosides and nucleotides (which are inherently more flexible in solution as well as being more subject to packing forces in the crystal), and more recently, in oligonucleotides (Schneider, Neidle, and Berman, 1997; Packer and Hunter, 1998; Varnai *et al.*, 2002). The existence of such correlations is important: it means that the atomic motions in oligo- and polynucleotides follow concerted patterns of inter-dependence. In general, these correlations are due to the diminution of non-bonded contacts that occur with particular conformations. Some of the more important correlations that have been observed in mononucleosides and nucleotides are:

- Between sugar pucker and glycosidic angle χ, especially for pyrimidine nucleosides. C3′-*endo* pucker is usually associated with median-value *anti* glycosidic angles, whereas C2′-*endo* puckers are commonly found with high *anti* χ angles. *Syn* glycosidic angle conformations show a marked preference for C2′-*endo* sugar puckers.
- The C4′–C5′ torsion angle γ is correlated with the glycosidic angle and to some extent with sugar pucker and backbone angle α (Schneider, Neidle, and Berman, 1997; Varnai et al., 2002). *Anti* glycosidic angles tend to correlate with g^+ conformations for γ.

References

(1989). Definitions and nomenclature of nucleic acid structure parameters. *EMBO J.*, **8**, 1–4.

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**, 8205–8212.

Babcock, M. S., Pednault, E. P. D., and Olson, W. K. (1994). Nucleic acid structure analysis. Mathematics for local Cartesian and helical structure parameters that are truly comparable between structures. *J. Mol. Biol.*, **237**, 125–156.

Clowney, L., Jain, S. C., Srinivasan, A. R., Westbrook, J., Olson, W. K. and Berman, H.M. (1996). Geometric parameters in nucleic acids: nitrogenous bases. *J. Amer. Chem. Soc.*, **118**, 509–518.

Dickerson, R. E. (1998). DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.*, **26**, 1906–1926.

Foloppe, N., Nilsson, L., and MacKerell, A. B., Jr. (2001). Ab initio conformational analysis of nucleic acid components: intrinsic energetic contributions to nucleic acid structure and dynamics. *Biopolymers*, **61**, 61–76.

Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.-H., Olson, W. K., and Berman, H. M. (1996). Geometric parameters in nucleic acids: sugar and phosphate constituents. *J. Amer. Chem. Soc.*, **118**, 519–529.

Goodsell, D. S. and Dickerson, R. E. (1994). Bending and curvature calculations in B-DNA. *Nucleic Acids Res.*, **22**, 5497–5503.

Gould, I. R. and Kollman, P. A. (1994). Theoretical investigation of the hydrogen bond strengths in guanine-cytosine and adenine-thymine base pairs. *J. Amer. Chem. Soc.*, **116**, 2493–2499.

Harvey, S. C. and Prabhakaran, M. (1986). Ribose puckering: structure, dynamics, energetics, and the pseudorotation cycle. *J. Amer. Chem. Soc.*, **108**, 6128–6136.

Kim, S.-H., Berman, H. M., Seeman, N. C., and Newton, M. D. (1973). Seven basic conformations of nucleic acid structural units. *Acta Crystallogr.*, **B29**, 703–710.

Lavery, R. and Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids *J. Biomol. Struct. Dynamics*, **6**, 63–91.

Lavery, R. and Sklenar, H. (1989). Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dynamics*, **6**, 655–667.

Liu, D., Wyttenbach, T., and Bowers, M. T. (2006). Hydration of mononucleotides. *J. Amer. Chem. Soc.*, **128**, 15155–15163.

Lu, X.-J. and Olson, W. K. (1999). Resolving the discrepancies among nucleic acid conformational analyses. *J. Mol. Biol.,* **285**, 1563–1575.

Lu, X.-J., Babcock, M. S., and Olson, W. K. (1999). Overview of nucleic acid analysis programs. *J. Biomol. Struct. Dynamics*, **16**, 833–843.

Lu, X.-J., El Hassan, M. A., and Hunter, C. A. (1997). Structure and conformation of helical nucleic acids: analysis program (SCHNAaP). *J. Mol. Biol.,* **273**, 668–680.

Murray-Rust, P. and Motherwell, S. (1978). Computer retrieval and analysis of molecular geometry. III. Geometry of the β-1′-aminofuranoside fragment. *Acta Crystallogr.*, **B34**, 2534–2546.

Olson, W. K. (1982a). How flexible is the furanose ring? 2. An updated potential energy estimate. *J. Amer. Chem. Soc.*, **104**, 278–286.

Olson, W. K. (1982b) In *Topics in Nucleic Acid Structure*, Part 2 (Ed. Neidle, S.), pp. 1–79. Macmillan Press, London.

Olson, W. K. and Sussman, J. L. (1982). How flexible is the furanose ring? 1. A comparison of experimental and theoretical studies. *J. Amer. Chem. Soc.*, **104**, 270–278.

Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X.J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C.-S., Westhof, E., Wolberger, C., and Berman, H. M. (2001). A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, **313**, 229–237.

Packer, M. J. and Hunter, C. A. (1998). Sequence-dependent DNA structure: the role of the sugar-phosphate backbone. *J. Mol. Biol.*, **280**, 407–420.

Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T., and Berman, H. M. (1996). New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr.*, **D52**, 57–64.

Pearlman, D. A. and Kim, S.-H. (1990). Atomic charges for DNA constituents derived from single-crystal X-ray diffraction data. *J. Mol. Biol.*, **211**, 171–187.

Rosenberg, J. M., Seeman, N. C., Day, R. O. and Rich, A. (1976). RNA double-helical fragments at atomic resolution. II. The crystal structure of sodium guanylyl-3′,5′-cytidine nonahydrate. *J. Mol. Biol.*, **104**, 145–167.

Schneider, B., Moravek, Z., and Berman, H. M. (2004). RNA conformational classes. *Nucleic Acids Res.*, **32**, 1666–1677.

Schneider, B., Neidle, S., and Berman, H. M. (1997). Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers*, **42**, 113–124.

Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Kim, J. P., and Rich, A. (1976). RNA double-helical fragments at atomic resolution. I. The crystal and molecular structure of sodium adenylyl-3′,5′-uridine hexahydrate. *J. Mol. Biol.*, **104**, 109–144.

Sponer, J., Jurecka, P., and Hobza, P. (2004). Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *J. Amer. Chem. Soc.*, **126**, 10142–10151.

Stofer, E., Chipot, C., and Lavery, R. (1999). Free energy calculations of Watson–Crick base pairing in aqueous solution. *J. Amer. Chem. Soc.*, **121**, 9503–9508.

Sun, G., Voigt, J. H., Filippov, I. V., Marquez, V. E., and Nicklaus, M. C. (2004). PROSIT: Pseudo-rotational online service and interactive tool, applied to a conformational survey of nucleosides and nucleotides. *J. Chem. Inf. Comput. Sci.*, **44**, 1752–1762.

Sundaralingam, M. (1969). Stereochemistry of nucleic acids and their constituents. IV. Allowed and preferred conformations of nucleosides, nucleoside mono-, di-, tri-, tetraphosphates, nucleic acids and polynucleotides. *Biopolymers*, **7**, 821–860.

Varnai, P., Djuranovic, D., Lavery, R., and Hartmann, B. (2002). Alpha/gamma transitions in the B-DNA backbone. *Nucleic Acids Res.*, **30**, 5398–5406.

Wain-Hobson, S. (2006). The third bond. *Nature*, **439**, 539.

Watson, J. D. and Crick, F. H. C. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.

Further Reading

Jeffrey, G.A. and Saenger, W. (1991). *Hydrogen Bonding in Biological Structures*. Chaps. 1**5**, 16. Springer, Berlin.

Sponer, J., Leszczynski, J., and Hobza, P. (2001). Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers*, **61**, 3–31.

Voet, D. and Rich, A. (1970). The crystal structures of purines, pyrimidines and their intermolecular complexes. *Prog. Nucleic Acid Res. Mol. Biol.,* **10**, 183–265.