phosphates spiral forward, should they go clockwise or counter-clockwise? There are, in fact, known examples of both right-handed and left-handed double helices: the 'Z' form (Fig. 2.7) should perhaps be assigned −12 phosphates per turn, rather than +12, because it is left handed rather than right-handed. It turns out that most DNA double helices are right-handed because of certain details of the chemical structure.

These details can be seen clearly when we build accurate, space-filling models of DNA: the atoms do not easily fit together if we try to build left-handed versions. Only very special combinations of DNA bases can become left-handed. Even then, the structures formed are so complicated and difficult to understand that they will not be considered much in this book.

Before proceeding further, let us look closely at some realistic pictures of DNA that include all of the atoms. Figure 2.8(a) shows a space-filling model of a very small part of right-handed DNA, corresponding roughly to the fragment shown more crudely in Fig. 2.5(b), or to one small part of the 'B' helix in Fig. 2.7. The atoms in this model are color coded as hydrogen (white), carbon (black), nitrogen (blue), oxygen (red), and phosphorus (yellow). The bases are joined at the center in 'pairs' (discussed below), while sugars and phosphates lie along the outside. This kind of model is widely used by scientists because it is so accurate. In fact, one could measure from the model a 3.3 Å separation of bases and a 6 Å separation of phosphates, and so check our previous calculation. Also, if you try to twist the base pairs backward in this model, from right handed to left handed, the bonds between atoms fall apart.

Other more detailed chemical representations of just one DNA chain are shown in Fig. 2.8(b) and (c). There the atoms (or groups of atoms) are drawn as letters or balls: for example, P = phosphorus, O = oxygen, C = cytosine, and A = adenine; while the bonds between the atoms are drawn as lines or sticks. The 5′ to 3′ direction of chains is also shown.

Let us now return to Fig. 2.7 to look more closely at the three forms of DNA shown there. Because we now know something about the internal structure of DNA, these models make sense to us. In all cases we can see that the bases compactly fill the centers of the double helices, where they successfully escape from contact with the surrounding water, and that the sugars and phosphates spiral around the outside of each helix at a rate of 11 ± 1 phosphates per turn. The 'Z' form also shows some jagged features in its sugar–phosphate chain that we shall not analyze here; they come from irregularities in the packing of bases on the inside of the helix. Using Fig. 2.8 as a guide, you can attempt to identify single bases, sugars,

and phosphates in the pictures of Fig. 2.7, or even single atoms, if you have enough patience. It should be stressed that all of the pictures shown in Figs 2.7 and 2.8 are equally valid ways of representing the same thing; one kind of picture may be preferred over another, depending upon the level of detail at which you wish to visualize the DNA.

We have assumed throughout our exposition that the two chains of a DNA double helix run in opposite or 'antiparallel' directions

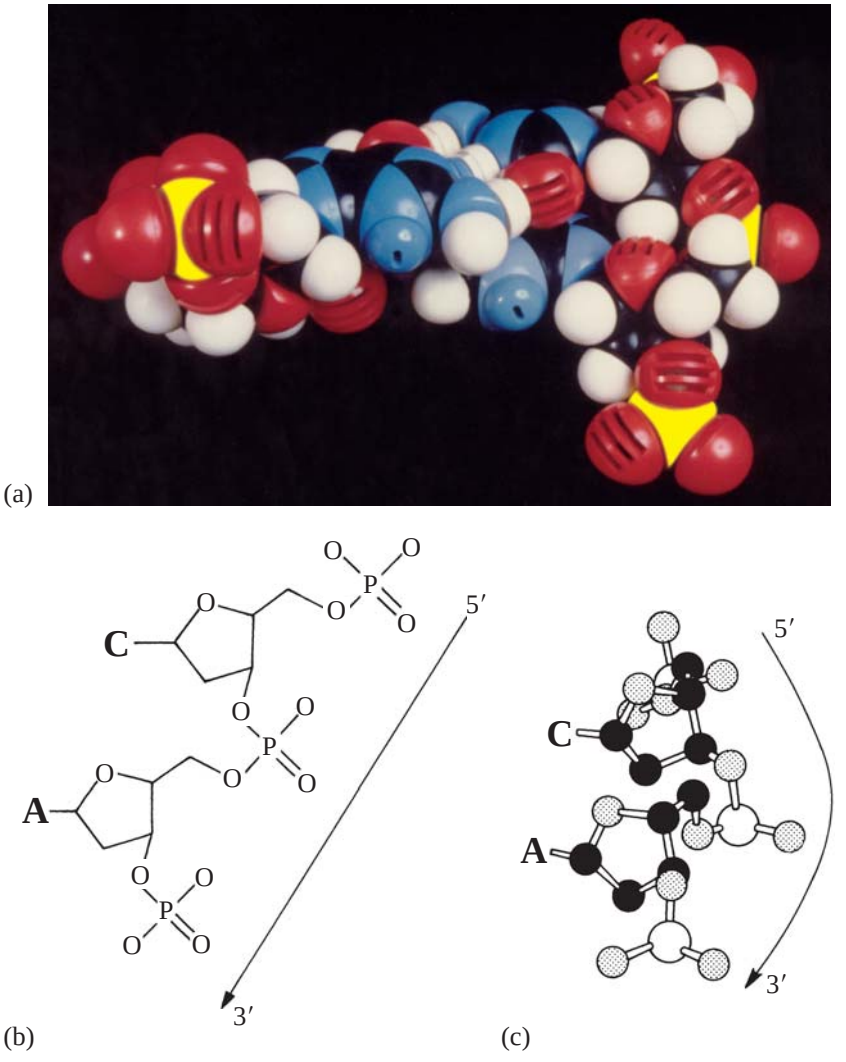

Figure 2.8 (a) Space-filling model of two stacked base-pairs, and their associated sugar–phosphate chains; (b) schematic, plane layout of the right-hand chain, with bases cytosine (C) and adenine (A) shown only as letters; (c) three-dimensional version of (b), in same configuration as (a).

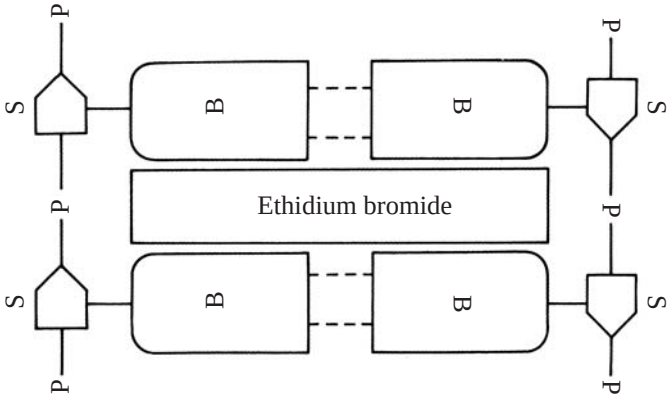

Figure 2.9 Ethidium bromide fills a gap between base pairs (cf. Fig. 2.3).

(Figs 1.7, 2.3, and 2.4), because 'parallel' double helices are not found in Nature. To a large extent, the interactions or pairings between bases at the core of the helix require that naturally occurring DNA be made from antiparallel chains. Some scientists have made short DNA molecules through chemical synthesis, in which the bases can only pair with one another when the chains run in a parallel sense: but these pairings are far less stable than when the chains run in the usual, antiparallel way, as found in Nature.

Another point before we go on: scientists sometimes add to DNA a substance called ethidium bromide, which fluoresces a bright red-pink when ultraviolet light shines on it, especially if it is bound to DNA. This provides a way of seeing exactly where the DNA is located in certain preparations, and so it constitutes an important tool for investigating DNA. How exactly does ethidium bromide stick to DNA? It is a greasy, mostly hydrophobic molecule, about the size of a base-pair. It can escape from contact with water by slipping between neighboring base-pairs along the chain, as shown in Fig. 2.9. Going back to Fig. 2.3, we must expect that the DNA will untwist locally to form an ordinary ladder before the ethidium bromide can fit in; and indeed it turns out that DNA plus ethidium bromide make a largely untwisted ladder. This feature, combined with the fluorescence mentioned above, makes ethidium bromide an important tool for scientists who investigate DNA.

It is the *flexibility* of sugar–phosphate chains which allows them to change from a spiral to a straight ladder when ethidium bromide is added. This flexibility is of an unusual and indirect sort which arises from features shown in Fig. 2.10. There we see in Fig. 2.10(a) that the phosphate group is essentially a rigid tetrahedron, having a phosphorus atom at its center and one oxygen at each vertex. Only when we go further along the chain from the phosphorus, as in Fig. 2.10(b),

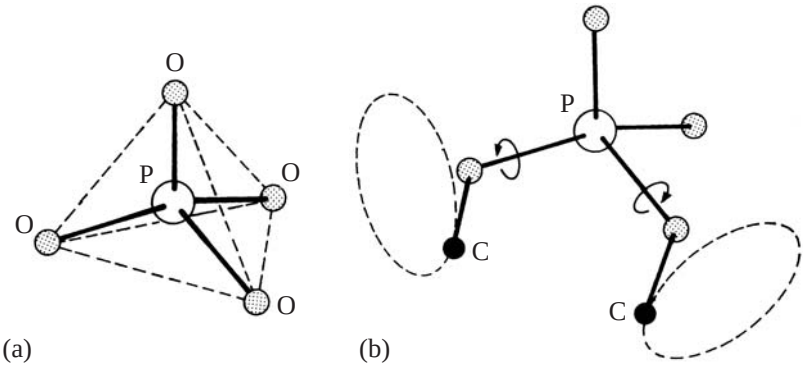

Figure 2.10 (a) Tetrahedral arrangement of phosphate group: atom types are phosphorus (P) and oxygen (O); (b) freedom of rotation for adjacent links of the chain with atom type carbon (C).

and attach two carbon atoms to two of the oxygens, can these carbon atoms swivel about the line of the phosphorus–oxygen bond. Imagine that your shoulder is the phosphorus, your elbow is the oxygen, and your hand is the carbon; obviously your hand and lower arm can swivel relative to your shoulder, but in DNA the angle at your elbow would be fixed. All parts of the DNA sugar– phosphate chain are rigid locally, but they have this kind of indirect rotational flexibility over several bonds.

We have almost finished our survey of the basic principles that determine the structure of DNA. All we have to do now is to learn how the bases adhere to one another in the central core of the double helix. James Watson and Francis Crick solved this problem in 1953, by putting forward a set of rules for base-pairing. They said that the most stable base pairs would be of the kind A–T or G–C, as shown in Fig. 2.11(a) and (b). One advantage of their scheme was that all four possible Watson–Crick base pairs, A–T, T–A, G–C, and C–G, were of the same size, and hence could fit easily into the framework of a regular double helix. Another advantage was that it explained how the genes in DNA could be duplicated (or stably inherited) on cell division. Whenever a cell divides, and needs to duplicate its DNA, it can do so simply by splitting the DNA into two separate strands; then certain enzymes will come along and use each of these old strands as a 'template' for the precise synthesis of a new strand, according to the Watson–Crick rules of base pairing: A with T and G with C. (More will be said about this in Chapter 4.)

Note that some of the interatomic connections within the A, T, G, and C rings are drawn as two lines, rather than as one: these are the 'double bonds', which give the base rings both their flatness and their rigidity. Also note that the CH_3 or methyl group on the

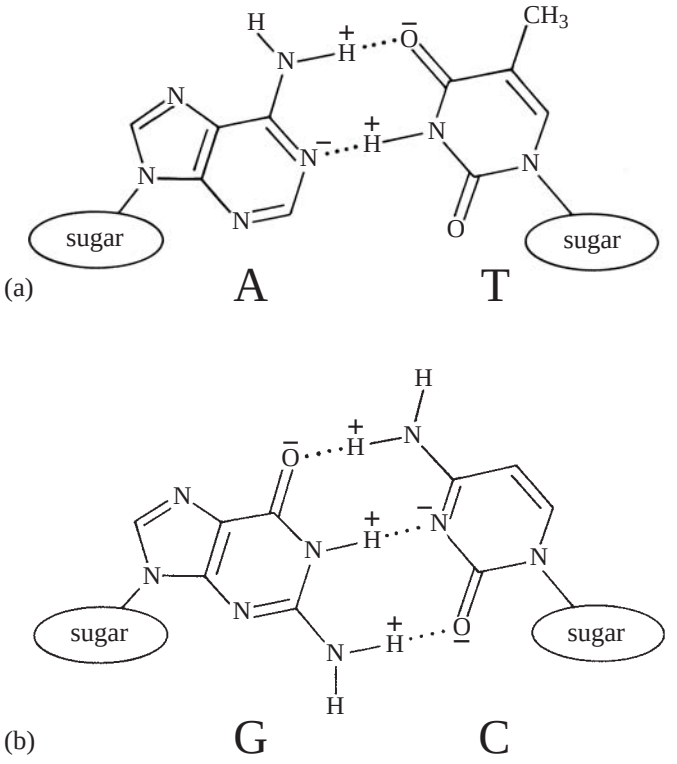

Figure 2.11 Watson–Crick base pairs showing hydrogen bonding: (a) A–T; (b) G–C. Atoms which are not labeled are carbon, while other atom types are hydrogen (H), nitrogen (N), and oxygen (O). Not all hydrogens are shown. Hydrogen bonds are represented by dotted lines. Symbols + and − here represent partial electric charges of about 1/3 electron or proton, which are typical for hydrogen bonds. A T–A pair is the same as an A–T pair, but turned over.

thymine ring would be absent in RNA, where the methyl-less base is called 'uracil' (see the right-hand part of Fig. 2.13 and its caption).

But what is the physical basis of these Watson–Crick rules of pairing? They are based on the simple fact that, within any DNA base, there is a small surplus of negative electric charge on nitrogen and oxygen atoms where they are *not* attached to hydrogen, while there is a small surplus of positive charge on these same atoms where they *are* attached to hydrogen. Thus, consider the base pair of adenine (A) and thymine (T), as shown in Fig. 2.11(a). Not counting the two nitrogens that are attached to sugars, there are three nitrogens on adenine and two oxygens on thymine that have a surplus of negative charge. On the other hand, one nitrogen on adenine and one nitrogen on thymine have a surplus positive charge. So all we have to do is to put the pluses and minuses together, thereby making the 'hydrogen bonds' which are shown here as dotted lines.

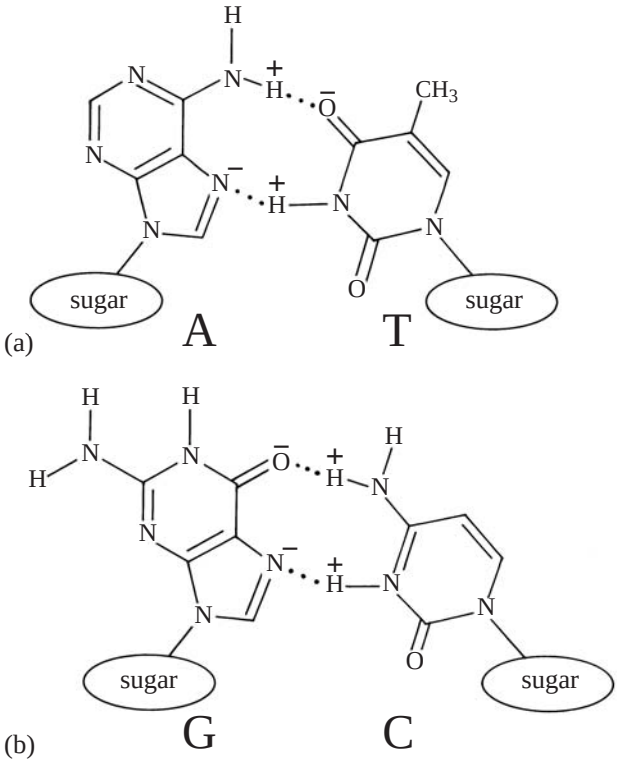

Figure 2.12 Hoogsteen base pairs, showing hydrogen bonding: (a) A–T; (b) G–C. Symbols + and − are partial charges, as in Fig. 2.11. The Hoogsteen G–C pair is stable only at mildly acidic pH (about 4–5), because it requires protonation of a cytosine nitrogen, that is, the adding of a hydrogen to it.

There are, in principle, two ways of doing this, consistent with making a double helix. These are called the Watson–Crick and Hoogsteen base pairs. The Watson–Crick A–T pair is shown in Fig. 2.11(a), while the Hoogsteen A–T pair is shown in Fig. 2.12(a). Both are roughly of equal stability; there are two hydrogen bonds which can be drawn between the bases in both cases. Note that the adenine base has to be rotated through 180° about the bond to the sugar in order to change between the two kinds of pairing – like rotating a tennis racquet from 'rough' to 'smooth'.

The history of these base-pairs is rather interesting. Watson and Crick found their A–T base-pair as part of a search for the double-helical structure of DNA by playing with paper cut-outs of bases. (Some details of this are given later, in Chapter 9.) Karst Hoogsteen tried 10 years later to confirm the Watson–Crick pair for adenine and thymine by heating up a solution of these two bases and letting it cool slowly in order to make a crystal; but he found instead a different kind of base-pair in his crystal.

Watson–Crick and Hoogsteen pairs for guanine (G) and cytosine (C) can also be drawn as shown in Figs 2.11(b) and 2.12(b). There are two things to note here. First, the Watson–Crick guanine–cytosine pair has three hydrogen bonds, rather than two as for adenine– thymine; so DNA double helices with guanine and cytosine should be more stable than those with adenine and thymine. Secondly, the Hoogsteen guanine–cytosine pair is only stable at low pH, since one of the nitrogens on cytosine must be protonated (i.e. be attached to a hydrogen) for this structure to form. The midpoint for protonation is pH 5, or more acidic than the normal pH 7 to 8 found in cells. That is the main reason why practically all DNA double helices contain Watson–Crick rather than Hoogsteen pairs; the Hoogsteen G–C pair is *not stable at neutral pH*. An additional reason is that a Watson– Crick G–C pair has more hydrogen bonds than a Hoogsteen G–C pair – three *versus* two. In principle, DNA containing only A and T bases should be relatively stable in either Watson–Crick or Hoogsteen forms, but few purely Hoogsteen-paired double helices have yet been detected with certainty. Also, it seems unlikely that both Watson– Crick and Hoogsteen pairs could be easily accommodated in close proximity within the same double helix, because the two kinds of base pair have different sugar-to-sugar distances.

Until recently it would have been enough to learn just about Watson–Crick base-pairs. But now it is important to learn about Hoogsteen pairs as well, because such pairs show up occasionally in complexes of DNA with anticancer drugs, and also in triple helices where a third strand of DNA joins the first two. The base 'triplet' then contains both Watson–Crick and Hoogsteen pairs: try to draw a triplet with one adenine and two thymines. Two scientists in California, Scott Strobel and Peter Dervan, have actually been able to cut a yeast chromosome cleanly in half by designing a third strand of DNA that binds only to one specific double-helical sequence within the whole chromosome, and then by attaching an iron atom to this third strand, so as to 'rust' (or oxidize) the DNA into pieces.

Why can't there be other stable base pairs, such as G with A, or C with T? Some of these are ruled out by the difficulty of making two or more hydrogen bonds. But others, such as guanine (G) with thymine (T) or uracil (U), as shown in Fig. 2.13, are not excluded for that reason. The hydrogen bonding produces a pair with similar overall shape to those in Fig. 2.11. In fact, a guanine–uracil pair is perfectly stable, and is used when transfer-RNA – the molecule that carries amino acids for protein synthesis – binds to messenger-RNA on the ribosome. You may recall from Chapter 1 that DNA makes a copy of itself in the form of RNA, and that this copy, called 'messenger-RNA', travels outside the nucleus to the protein-making machinery or

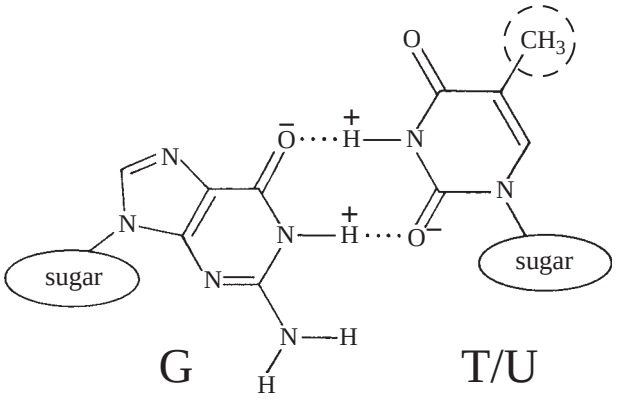

Figure 2.13 Pairing of G with T allows two good hydrogen bonds, plus an overall shape similar to that of the Watson–Crick pairs shown in Fig. 2.11(a) and (b). A closely related base-pair, that of G with U (uracil), is used routinely to specify amino acids for the synthesis of proteins, as shown in Fig. 2.14; as explained in Chapter 1, U is like T but without the CH_3 group here enclosed in a broken circle.

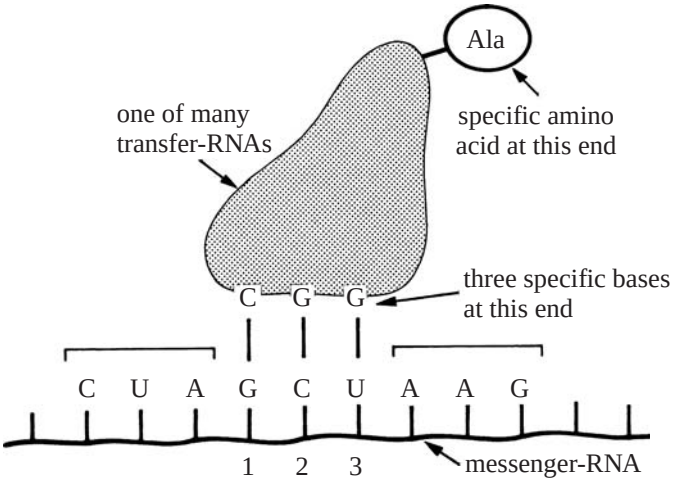

Figure 2.14 The transfer-RNA molecule that carries alanine can recognize its preferred triplet of bases on a messenger-RNA chain by using either a G–U or a G–C base-pair in position 3.

'ribosome'. There it becomes attached to a series of different transfer-RNA molecules, one of which is shown in Fig. 2.14. Each kind of transfer-RNA contains three specific bases at one end, and one specific amino acid at the other. If the three specific bases on transfer-RNA can form base-pairs with three neighboring bases in the messenger-RNA, then the amino acid being carried by the transfer-RNA is added to a growing protein chain.

Now, there are $4 \times 4 \times 4 = 64$ possible combinations of three bases in the messenger-RNA, and yet sometimes there are only 25 to 30 different kinds of transfer-RNA molecule in a cell. So several different sets of three bases in the messenger-RNA must have to 'share' a transfer-RNA in much the same way that two birds will share the same nest when there are not enough trees. In fact, the transfer-RNA is flexible, and will accept either of two possible base pairs in the third position. An example is shown in Fig. 2.14: here either an RNA sequence GCU with an 'unusual' G–U pair, or else a sequence GCC, with the 'normal' G–C pair, serves to determine one particular amino acid. A guanine–uracil pair is not the only unusual one to be used in this way; there are several others.

That brings us to a final question: if all of these different base pairings are stable enough to be used by transfer-RNA when it links with messenger-RNA, as in Fig. 2.14, why aren't they also found in DNA? If indeed they were used every time a cell divides, the sequence of bases in the DNA could change drastically. Yet this evidently does not happen, since it is now known that just a few G–C to G–T changes at critical places in the DNA program could cause cancer. In fact, Nature uses special proteins called 'proofreading enzymes' to prevent the occurrence of slight changes in sequence when DNA replicates. The enzymes that copy DNA to DNA, or DNA to RNA, are indeed very clever. They can sense at several stages during synthesis whether anything is going wrong; for example, if they have added or are about to add the wrong base, according to the Watson–Crick rules of pairing. Also, there are 'repair' enzymes that go around correcting occasional mistakes of copying or 'mismatches'. Thus, Nature goes to great lengths to avoid errors in the copying of DNA, even though the atoms in the DNA structure are actually quite tolerant of mismatch pairings. These enzymes are extremely efficient in doing their job, yet no one knows exactly how they work.

In summary, we have learned in this chapter how the insolubility of bases in water provides the driving force for DNA to form a double helix; and how the geometry at the core of the helix depends on subtle interactions between partial electrical charges on the bases. These subtle interactions alone are not sufficient for accurate copying of DNA from generation to generation, and so cells contain many enzymes that enhance the efficiency of copying. Yet unusual base pairs are a fact of life at other places in the cell, such as when proteins are made *via* the binding of transfer-RNA to messenger-RNA; so one has to be aware of all these subtle possibilities, in order to appreciate the internal structure of nucleic acids as they act in biology.

Further Reading

Abrescia, N.G., Thompson, A., Huynh-Dinh, T., and Subirana, J.A. (2002) Crystal structure of an antiparallel DNA fragment with Hoogsteen base pairing. *Proceedings of the National Academy of Sciences, USA* **99**, 2806–11. A short d(ATATAT) double helix which contains Hoogsteen A–T base pairs throughout.

Arnott, S. (1970) The geometry of nucleic acids. *Progress in Biophysics and Molecular Biology* **21**, 265–319. A summary of early X-ray studies on fibrous samples of DNA.

Chaires, J.B. and Waring, M.J., eds (2001) Drug-nucleic acid interactions. *Methods in Enzymology* **340**. An excellent volume which summarizes the latest techniques and results in drug-DNA work.

Dickerson, R.E., Drew, H.R., Conner, B.N., Wing, R.M., Fratini, A.V., and Kopka, M.L. (1982) The anatomy of A-, B- and Z-DNA. *Science* **216**, 475–85. A summary of the first work from X-ray studies of DNA in single crystals.

Matray, T.J. and Kool, E.T. (1999) A specific partner for abasic damage in DNA. *Nature* **399**, 704–7. The importance of correct size and shape for inclusion of bases within a double helix by polymerase enzymes.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York. A comprehensive review of the published literature on DNA structure up to the early 1980s.

Schwartz, T., Rould, M.A., Lowenhaupt, K., Herbert, A., and Rich, A. (1999) Crystal structure of the Zα domain of human editing-enzyme ADAR1 bound to left-handed Z-DNA. *Science* **284**, 1841–5. Possible biological role for left-handed DNA as a substrate for the RNA-editing enzyme adenosine deaminase.

Wan, C., Fiebig, T., Schiemann, O., Barton, J.K., and Zewail, A.H. (2000) Femtosecond direct observation of charge transfer between bases in DNA. *Proceedings of the National Academy of Sciences, USA* **97**, 14052–5. A study of how electrons move between adjoining base pairs in DNA over very short time scales.

Wittung, P., Nielsen, P.E., Buchardt, O., Egholm, M., and Norden, B. (1994) DNA-like double helix formed by peptide nucleic acid. *Nature* **368**, 561–3. The DNA bases form double-helical structures even if attached to a chain of peptides, rather than of sugars and phosphates.

Bibliography

Hoogsteen, K. (1959) Structure of a crystal containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Crystallographica* **12**, 822–3. Discovery of an alternative pairing-scheme for A with T.

Strobel, S.A. and Dervan, P.B. (1990) Site-specific cleavage of a yeast chromosome by oligonucleotide-directed triple-helix formation. *Science* **249**,