23.3. Nucleic acids

BY R. E. DICKERSON

23.3.1. Introduction

In 1953, James Watson and Francis Crick solved the structure of double-helical DNA (Watson & Crick, 1953; Crick & Watson, 1954). So what has a dedicated cadre of X-ray crystallographers been doing for the subsequent 45 years? That is the subject of this chapter: the advance of our knowledge of nucleic acid duplexes, primarily from single-crystal X-ray diffraction, and the biological implications of this new knowledge. The focus will be primarily on DNA because much more is known about it, but DNA/RNA hybrids and duplex RNA will also be considered. Because the emphasis is on the geometry of the nucleic acid double helix, exotic structures, such as quadruplexes, hammerhead ribozymes and aptamers, will be omitted, as will larger-scale structures such as tRNA.

Fibre diffraction showed that there were two basic forms of DNA duplex: the common B form and a more highly crystalline A form (Fig. 23.3.1.1) that, in some but not all sequences, could be produced by dehydrating the fibre (Franklin & Gosling, 1953; Langridge *et al.*, 1960; Arnott, 1970; Leslie *et al.*, 1980). A- and B-DNA are contrasted in Figs. 23.3.1.2 and 23.3.1.3. The high-humidity B form has base pairs sitting squarely on the helix axis and roughly perpendicular to that axis. In contrast, in the low-humidity A form, the base pairs are displaced off the helix axis by *ca* 4 Å and are inclined 10–20° away from perpendicularity to that axis. The two grooves in B-DNA are of comparable depth because base pairs sit *on* the helix axis, but the major groove is wider than the minor because of asymmetry of attachment of base pairs to the backbone chains. In A-DNA, the minor groove is broad and shallow, whereas the major groove is cavernously deep (all the way from the surface of the helix, to the helix axis, and beyond) but can be quite narrow.

Pohl and co-workers had shown in the 1970s that alternating poly(dC-dG) is special in that it undergoes a reversible salt- or alcohol-induced conformation change (Pohl & Jovin, 1972; Pohl, 1976). Hence, it was not surprising that when DNA synthesis methods advanced to the stage where oligonucleotide crystallization became feasible, two separate research groups – those of Alexander Rich at MIT and Richard Dickerson at Caltech – elected to synthesize, crystallize and solve a short, alternating C-G oligomer. The result was a third family of DNA duplexes, Z-DNA (Fig. 23.3.1.4), first as the hexamer C-G-C-G-C-G (Z1) and then the tetramer C-G-C-G (Z3). (References to A-, B- and Z-DNA structures are listed at the end of Tables A23.3.1.1, A23.3.1.2 and A23.3.1.3 in the Appendix, respectively. They are cited by numbers beginning with A, B or Z.) Single-crystal analyses of the traditional helix types soon followed: B-DNA as C-G-C-G-A-A-T-T-C-G-C-G (B1), and A-DNA as both C-C-G-G (A1) and G-G-T-A-T-A-C-C (A2).

23.3.2. Helix parameters

23.3.2.1. *Backbone geometry*

Before making detailed comparisons of the three helix types, one must define the parameters by which the helices are characterized. The fundamental feature of all varieties of nucleic acid double helices is two antiparallel sugar–phosphate backbone chains, bridged by paired bases like rungs in a ladder (Fig. 23.3.2.1). Using the convention that the positive direction of a backbone chain is from 5′ to 3′ within a nucleotide, the right-hand chain in Fig. 23.3.2.1 runs downward, while the left-hand chain runs upward. A- or B-DNA is then obtained by twisting the ladder into a right-handed helix. But Z-DNA cannot be obtained from Fig. 23.3.2.1 simply by giving it a left-handed twist; both backbone chains run in the wrong direction for Z-DNA. A more complex adjustment is required, and this will be addressed again later.

The conformation of the backbone chain along each nucleotide is described by six torsion angles, labelled α through ζ, as shown in Fig. 23.3.2.2. An earlier convention termed these same six angles as ω, φ, ψ, ψ', φ', ω' (Sundaralingam, 1975), but the alphabetical nomenclature is now generally employed. Torsion angles are defined in Fig. 23.3.2.3, which also shows three common configurations: *gauche*$^-$ (−60°), *trans* (180°) and *gauche*$^+$ (+60°). These three configurations are especially favoured with sp^3 hybridization or tetrahedral ligand geometry at the two ends of the bond in question, because their 'staggered' arrangement minimizes ligand–ligand interactions across the bond. An 'eclipsed' arrangement with ligands at −120°, 0° (*cis*), and 120° is unfavourable because it brings substituents at the two ends of the bond into opposition. Table 23.3.2.1 lists the mean values and standard deviations of all six main-chain torsion angles for A-, B- and Z-DNA, as recently observed in 96 oligonucleotide crystal structures (Schneider *et al.*, 1997).

23.3.2.2. *Sugar ring conformations*

The type of ligand–ligand clash just mentioned is an important element in ensuring that five-membered rings, such as ribose and deoxyribose, are not ordinarily planar, even though the internal bond angle of a regular pentagon, 108°, is close to the 109.5° of tetrahedral geometry. A stable compromise is for one of the four ring atoms to lie out of the plane defined by the other four, as in Fig. 23.3.2.4. This is termed an 'envelope' or E conformation, by analogy with a four-cornered envelope having a flap at an angle. Intermediate 'twist' or T forms are also possible, in which two adjacent atoms sit on either side of the plane defined by the other three, but this discussion will focus on the simple envelope conformations. In most cases, the accuracy of a nucleic acid crystal structure determination is such that it would be difficult to distinguish clearly between a given E form and its flanking T forms. For this reason, most structure reports consider only the E alternatives.

A convenient and intuitive nomenclature is to name the conformation after the out-of-plane atom and then specify whether it is out of plane on the same side as the C5′ atom (*endo*) or the opposite side (*exo*). Ten such conformations exist: five *endo* and

This chapter is dedicated to Irving Geis, who died on 22 July 1997 at the age of 88, just as the chapter was begun. Irv was a pioneer in the representation of protein and DNA structures, beginning with illustrations for *Scientific American* articles on myoglobin (Kendrew, 1961), lysozyme (Phillips, 1966), cytochrome *c* (Dickerson, 1972) and DNA (Dickerson, 1983). He was coauthor with the present writer of *Structure and Action of Proteins* (Dickerson & Geis, 1969) and two later textbooks (Dickerson & Geis, 1976, 1983) and contributed drawings and paintings to a great number of other books and articles, most notably Voet & Voet's *Biochemistry* (Voet & Voet, 1990, 1995), which is a veritable gallery of Irv's art. His meticulous and carefully thought-out diagrams and drawings of myoglobin and haemoglobins have never been matched. More information about his life, work and art may be found in three articles by the present author (Dickerson, 1997*a*,*b*,*c*). Irv saw his role as one of bringing an understanding of protein structure to life scientists and sometimes referred to himself half-humorously as 'the Andreas Vesalius of molecular anatomy'. In view of the formative influence that his art exerted on the first generation of protein crystallographers and molecular biologists, it is more appropriate to remember Irv as the Leonardo da Vinci of macromolecules. As of late 2000, nearly all of Irving Geis' work – paintings, drawings, illustrations and correspondence – is being preserved for study as the Geis Archives at the Howard Hughes Medical Institute, Washington DC.

references