International Tables for Crystallography (2006). Vol. F, Section 23.3.4.1.1, p. 603.

proposed by Seeman *et al.* (1976), and involves acceptors and donors as marked by *A* and *D* in Fig. 23.3.2.7. The wide major groove of B-DNA is read by several classes of control proteins that function by positioning an α-helix within the groove so that its amino-acid side chains can sense the pattern of hydrogen bonding. This category includes prokaryotic and eukaryotic helix-turn-helix or HTH proteins, zinc-finger and other zinc-binding proteins, basic leucine zippers and their basic helix-loop-helix cousins, and others (See Table I of Dickerson & Chiu, 1997). The narrower minor groove is a frequent target for long, planar drug molecules, such as netropsin and distamycin, as listed in Part II of Table A23.3.1.2.

In principle, this readout mechanism would work perfectly well with a regular, ideal, fibre-like B-DNA helix. But other control proteins that recognize the minor groove, such as TATA-binding protein (TBP) and integration host factor (IHF), depend not merely on passive hydrogen bonding to an ideally regular duplex, but on the *sequence-dependent deformability* of one region of the helix *versus* another. The remainder of this chapter will be concerned with this effect and its role in DNA recognition.

23.3.4.1. *Sequence-dependent deformability*

23.3.4.1.1. *Minor groove width*

The simplest and first-noticed sequence-dependent deformability of the B-DNA duplex was variation in minor groove width. The first B-DNA oligomer to be solved, C-G-C-G-A-A-T-T-C-G-C-G (B1–B6), had a narrow minor groove in the central A-A-T-T region, with only *ca* 3.5 Å of free space between opposing phosphates and sugar rings. (It has become conventional to define the free space between phosphates as the measured minimal P–P separation across the groove, less 5.8 Å to represent two phosphate-group radii. Similarly, the measured distance between sugar oxygens is decreased by 2.8 Å, representing two oxygen van der Waals radii.) The C-G-C-G ends of the helix had the 6–7 Å opening expected for ideal B-DNA, but the situation was clouded, because the outermost two base pairs at each end of the helix interlocked minor grooves with neighbours in the crystal. Hence, the wider ends could possibly be only an artifact of crystal packing.

After 1991, the situation was clarified by the structures of several decamers [Table A23.3.1.2, Part I(*c*)], which stack on top of one another without the interlocking of grooves. The normal minor groove opening is *ca* 7 Å. Regions of four or more AT base pairs can exhibit a significantly narrowed minor groove, although such narrowing is not mandatory. This behaviour is seen with the B-DNA decamer, C-A-A-A-G-A-A-A-A-G, in Fig. 23.3.4.1. The narrowing arises mainly from the larger allowable propeller twist in AT base pairs, which displaces C1′ atoms at opposite ends of the pair in different directions, and moves the backbone chains in such a way as to partially close the groove (Fig. 23.3.4.2).

This is an excellent example of the concept of *sequence-dependent helix deformability*, rather than simple deformation. The two hydrogen bonds of an AT base pair allow a larger propeller twist but do not require it. Hence, AT regions of helix permit a narrowing of the minor groove but do not demand it. Indeed, this lesson was brought home in the most dramatic way when Pelton & Wemmer (1989, 1990) showed *via* NMR that a 2:1 complex of distamycin with C-G-C-A-A-A-T-T-G-G-C or C-G-C-A-A-A-T-T-T-G-C-G could exist, in which two drug molecules sat side-by-side within an enlarged central minor groove. Fig. 23.3.4.3 shows a narrow minor groove with a single netropsin molecule, and Fig. 23.3.4.4 shows a wide minor groove enclosing two di-imidazole lexitropsins side-by-side. In summary, an AT-rich region of minor groove is capable of narrowing but is not inevitably narrow, in contrast to GC-rich regions where the third hydrogen bond tends to keep the base pairs flat and the minor groove wide. The AT minor groove is potentially *deformable* without being inevitably *deformed*.

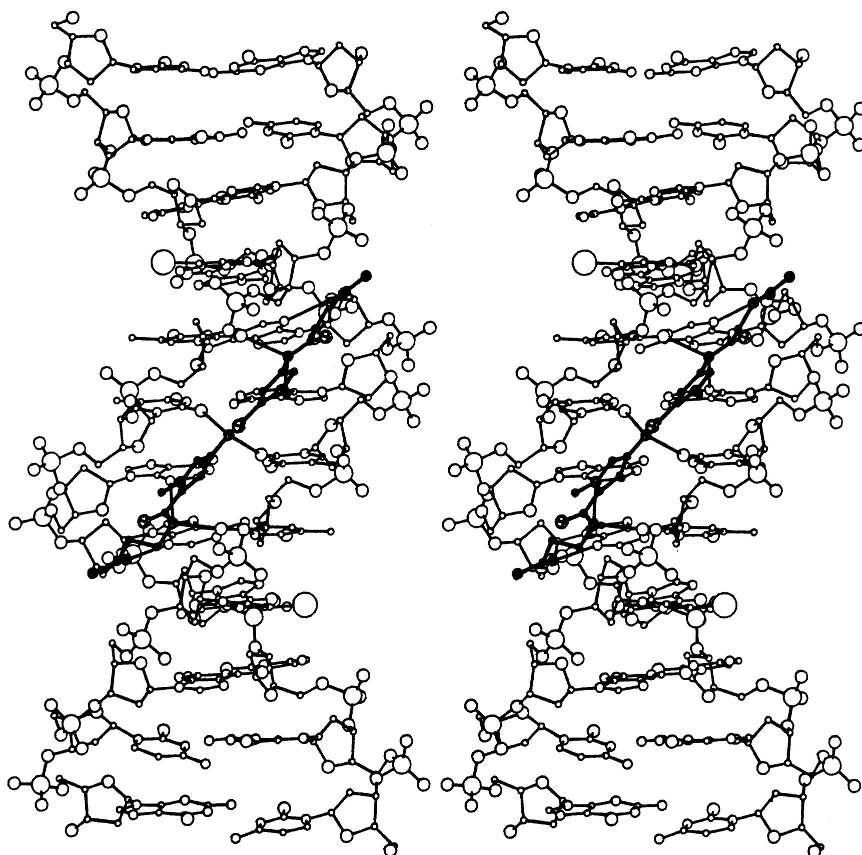

Fig. 23.3.4.3. Structure of the 1:1 complex of netropsin with C-G-C-G-A-A-T-T-C-G-C-G (B11, B12, B87). The drug binds to the central -A-A-T-T- region of the minor groove, which is barely wide enough to enclose the nearly planar polyamide molecule. The netropsin structure can be represented by

$$+(NH_2)_2C—NH—CH_2—CONH—Py—CONH—Py—CONH—CH_2—CH_2—C(NH_2)_2^+$$

where Py is a five-membered methylpyrrole ring. An even more compact representation, useful when comparing other polyamide netropsin analogues or lexitropsins, is $^+{=}Py{=}Py{=}^+$, where the common cationic tails are indicated only by a plus sign, and = represents a —CONH— amide.

23.3.4.1.2. *Helix bending*

Sequence-dependent bendability has been reviewed recently by Dickerson (1988*a*,*b*,*c*) and Dickerson & Chiu (1997). The relative bendability of different regions of B-DNA sequence is an important aspect of recognition, one that is used by countless control proteins that must bind to a particular region of double helix. Catabolite activator protein or CAP (Schultz *et al.*, 1991; Parkinson *et al.*, 1996), *lacI* (Lewis *et al.*, 1996) and *purR* (Schumacher *et al.*, 1994) repressors, $\gamma\delta$-

Copyright © 2006 International Union of Crystallography