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 diimidazole 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*.

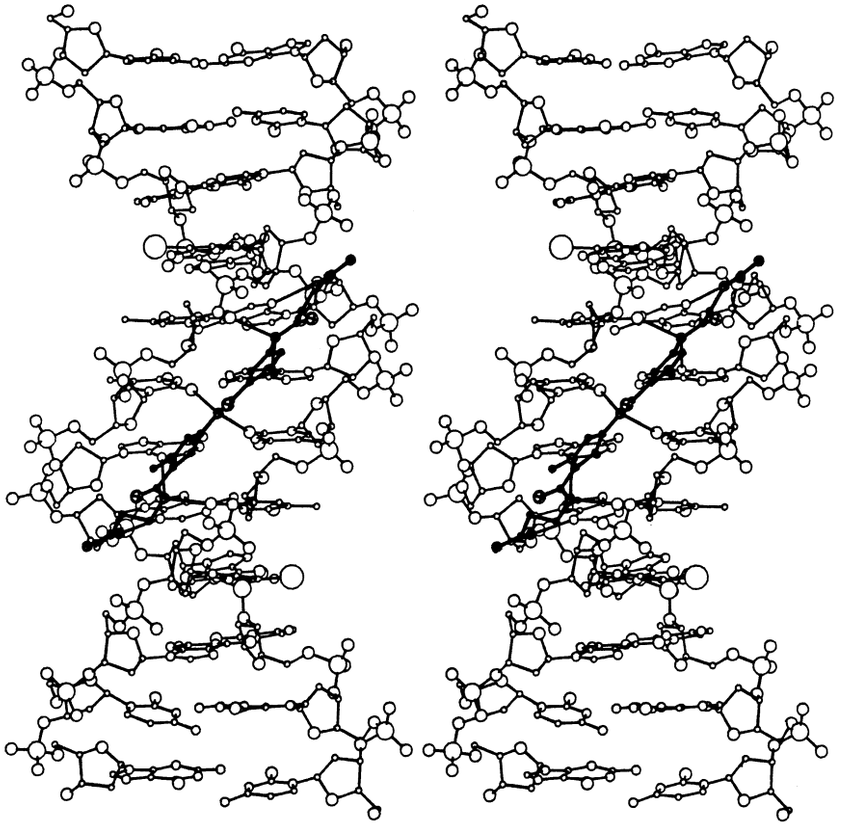

Fig. 23.3.4.3. Structure of the 1:1 complex of netropsin with C-G-C-G-A-A-T-T-C-G-C-G (B11, B12, B87). The drug binds to the central -A-A-T-T- region of the minor groove, which is barely wide enough to enclose the nearly planar polyamide molecule. The netropsin structure can be represented by

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

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, $\gamma\delta$-

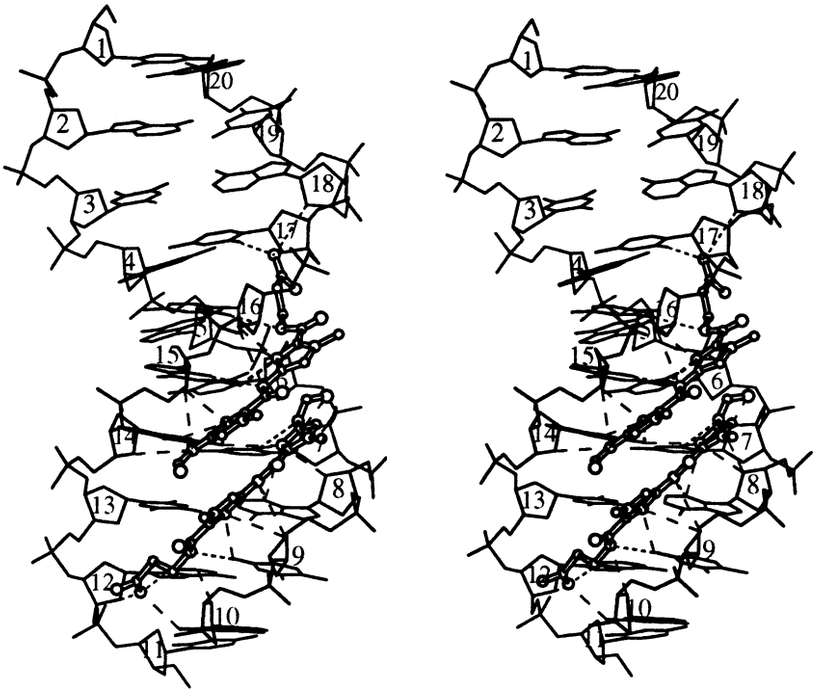

Fig. 23.3.4.4. Structure of the 2:1 complex of a di-imidazole lexitropsin with C-A-T-G-G-C-C-A-T-G (B108). The drug now is represented by

$$H—CONH—Im—CONH—Im—CONH—CH_2—CH_2—C(NH_2)_2^+$$

where Im is a five-membered imidazole ring, or again more compactly by $^0{=}Im{=}Im{=}^+$. The uncharged leading amide group, characteristic of distamycins, is identified by ^{0}. Distamycin itself would be represented in this shorthand notation by $^0{=}Py{=}Py{=}Py{=}^+$. Reprinted from B108, copyright (1977), with permission from Excerpta Medica Inc.

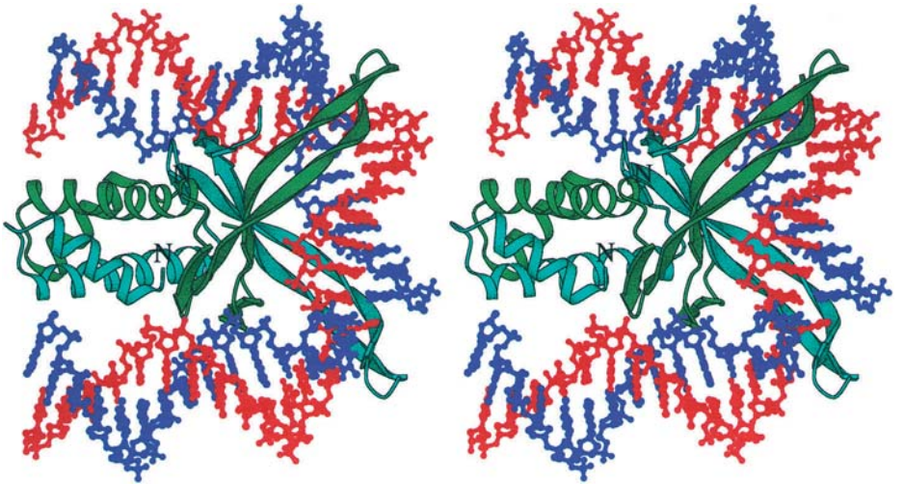

Fig. 23.3.4.5. DNA duplex (red and blue strands) looped around IHF or integration host factor. The two subunits of the IHF duplex are green and turquoise. Two antiparallel loops of protein chain, one from each subunit, insert into the minor groove of B-DNA at the sequence C-A-A-T/A-T-T-G and produce abrupt bends *via* local roll angles of 60°. The two localized bends are additive because they occur one helical turn apart. All other steps have roll angles of 5° or less. The two flanking helix segments pack against the IHF dimer and must be kept straight and unbent. This is accomplished in one of the two segments by an A-tract of sequence C-A-A-A-A-A-A-G. From Dickerson & Chiu (1997). Coordinates courtesy of P. Rice.

Table 23.3.4.1. *Sequence-dependent differential deformability in B-DNA. I. The Major Canon*

See Dickerson (1998*a*,*b*,*c*) and Dickerson & Chiu (1997).

(1) *Structural basis for helix bending in B-DNA*

Bending is nearly always the result of roll between successive base pairs, seldom tilt.

Positive roll, compressing the wide major groove, is more common than negative roll, in which the narrower minor groove is compressed.

Observed bends in B-DNA are of three main types: (*a*) *localized kinks* (large positive roll at one or two discrete base steps), (*b*) *three-dimensional writhe* (positive roll at a series of successive steps), or (*c*) *smooth curvature* (alternation of positive and negative roll every half turn, with side-to-side zigzagging at intermediate positions). (*a*) and (*b*) are easier to accomplish than (*c*), and hence are more common.

Local writhe in a DNA helix produces macroscopic curvature only when the extent of writhe does not match the natural rotational periodicity of the helix. Endless writhe results in a straight helix, and indeed A-DNA can be regarded as a continuously writhed variant of the B form. Conversely, the bending effect of writhe can be amplified if it is repeated with the periodicity of the helix itself – that is, repeated alternation of writhed and unwrithed segments every ten base pairs, as with A-tract B-DNA.

(2) *Pyrimidine-purine (Y-R) steps: C-A = T-G, T-A and C-G*

Little ring–ring stacking overlap.

Polar N or O stacked over polarizable aromatic rings.

Y-R steps are natural fracture points for the helix. They can show (but are not required in every case to show) large twist and slide deformations, and bending mainly *via* positive roll, compressing the major groove.

(3) *Purine-purine (R-R) steps: A-A = T-T, A-G = C-T, G-A = T-C and G-G = C-C*

Extensive ring–ring overlap.

Base pairs tend to pivot about stacked purines as a hinge, with greater ring–ring separation at pyrimidine ends.

Tight stacking, with only minor roll, slide and twist deformations.

(4) *Purine-pyrimidine (R-Y) steps: A-C = G-T, A-T and C-G*

Behaviour in general like R-R steps, with extensive ring–ring overlap and tight stacking, with again only minor roll, slide and twist deformations.

(5) *A-A and A-T steps, as contrasted with T-A*

Especially resistant to roll bending, probably because of sawhorse interlocking of highly propellered base pairs, supplemented by inter-base-pair hydrogen bonds within grooves. In contrast, T-A is particularly weak and subject to roll bending.

A-tracts, defined as four or more consecutive AT base pairs without the disruptive T-A step, are especially straight and resistant to bending. Natural selection has apparently chosen short A-tracts for regions of protein–DNA contacts where bending is not wanted.

resolvase (Yang & Steitz, 1995), *Eco*RV restriction enzyme (Winkler *et al.*, 1993; Kostrewa & Winkler, 1995), integration host factor or IHF (Rice *et al.*, 1996), and TBP or TATA-binding protein (Kim, Gerger *et al.*, 1993; Kim, Nikolov & Burley, 1993; Nikolov *et al.*, 1996; Juo *et al.*, 1996) are all sequence-specific DNA-binding proteins that bend or deform the nucleic acid duplex severely during the recognition process. IHF in Fig. 23.3.4.5 may be taken as representative of this class of DNA-binding proteins. The bend is produced by two localized rolls of *ca* 60° in a direction compressing the major groove and are additive, because they are spaced nine base pairs, or roughly one turn of helix, apart. In IHF, the two helix segments flanking the bend should be straight and unbent, and this is accomplished in one segment *via* a six-adenine A-tract: -C-A-A-A-A-A-A-G-.

The bending locus in IHF is C-A-A-T/A-T-T-G . It is C-G in *lacI* and *purR* repressors (Fig. 23.3.4.6), C-A = T-G in CAP (Fig. 10 of Dickerson, 1998*b*), and T-A in *Eco*RV, $\gamma\delta$-resolvase and TBP (Fig. 23.3.4.7). Pyrimidine-purine or Y-R steps appear to be especially suitable loci for roll bending. The dashed lines in Figs. 23.3.4.6 and 23.3.4.7 plot tilt, and demonstrate its insignificance in bending, compared with roll. (This is intuitively obvious. Imagine yourself standing near a tall stack of wooden planks in a lumberyard during an earthquake. Where would you prefer to stand: alongside the stack, or at one end?)

In summary, bending of the B-DNA helix nearly always involves roll, not tilt. The easier direction of bending is that which compresses the broad major groove, although examples of roll compression of the minor groove are known. Y-R steps are especially prone to roll bending. Again, the phenomenon is one of sequence-induced bendability, not mandatory bending. No one imagines that the IHF binding sequence of Fig. 23.3.4.5 is permanently kinked at its two C-A-A-T/A-T-T-G steps, wandering deformed through the nucleus, looking for an IHF molecule to bind to. Instead, this sequence has a potential bendability that other sequences, such as A-A-A-A-A-A, lack.

Table 23.3.4.1 summarizes the observed behaviour of Y-R, R-R and R-Y steps from a great many X-ray crystal structure analyses, with and without bound DNA. In the present context, these rules are termed the 'Major Canon', since they are well established and generally well understood. Some understanding of the proneness of Y-R steps to bend can be obtained by looking at stereo pairs of two successive base pairs viewed down the helix axis. Fig. 23.3.4.8 gives a few representative examples; many more can be found in Figs. 4–6 of Dickerson (1988*b*) and in the original literature. In brief, Y-R steps, especially C-A and T-A, tend to orient so that polar exocyclic N and O atoms stack against polarizable rings of the other base pair. This is the same type of polar-on-polarizable stacking stabilization mentioned earlier in connection with O4′ and guanine in Z-DNA (Bugg *et al.*, 1971; Thomas *et al.*, 1982; Hunter & Sanders, 1990; B32). Base pairs in T-A steps tend not to slide over one another along their long axes, keeping pyrimidine O2 stacked over the purine five-membered ring (Fig. 23.3.4.8*b*). C-A steps can adopt this same stacking, or the base pairs can slide until the pyrimidine O2 sits over the purine six-membered ring instead (Fig. 23.3.4.8*a*).

Purine-purine or R-R steps behave quite differently (Fig. 23.3.4.8*c*). They stack ring-on-ring, usually with greater overlap on the purine end than the pyrimidine. The net effect is that the pivot

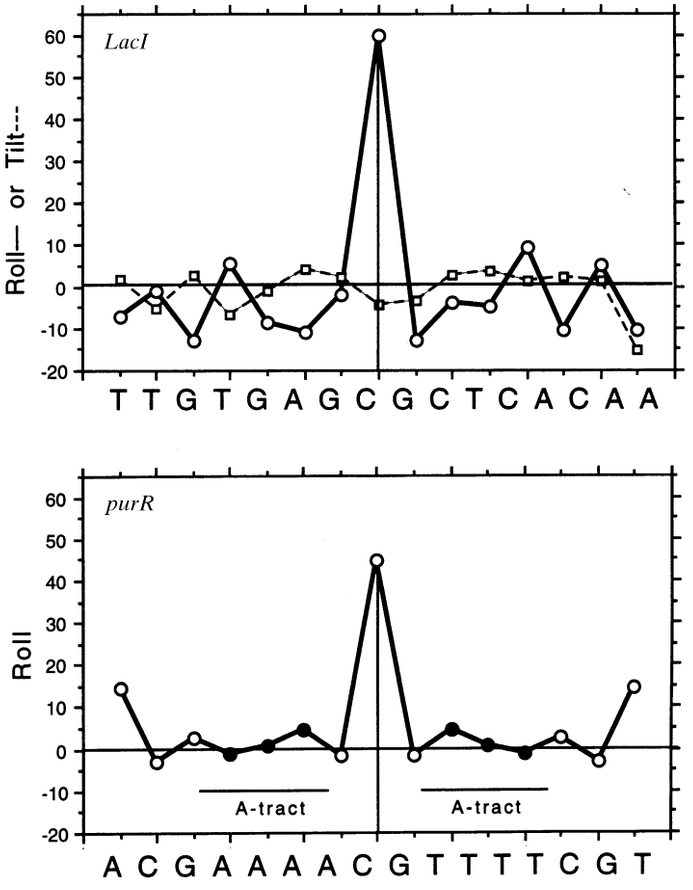

Fig. 23.3.4.6. Roll-angle plots for sequence-specific DNA–protein complexes with *lacI* (top) and *purR* (bottom). In each case, bending occurs *via* localized roll at a C-G step. Other steps of the sequence have random rolls of *ca* 10° or less. Note that, as with IHF, A-tracts are especially straight and unbent. Dashed lines in the *lacI* plot demonstrate the unimportance of tilt in production of helix bending.

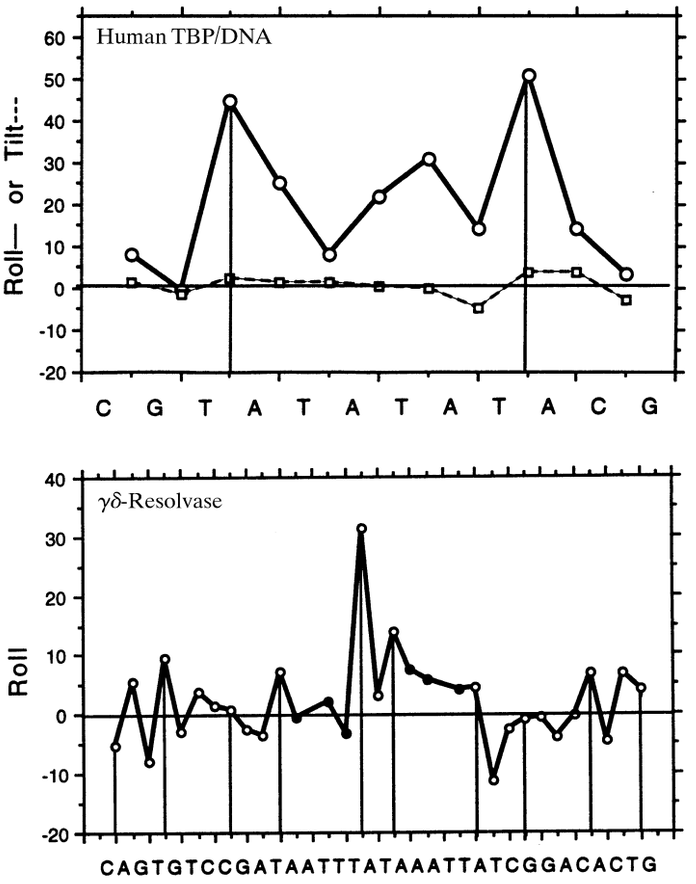

Fig. 23.3.4.7. Bending *via* roll at T-A steps in TBP or the TATA-binding protein (top) and in $\gamma\delta$-resolvase (bottom). Note that not every T-A step in TBP or $\gamma\delta$-resolvase is necessarily bent. Note also in $\gamma\delta$-resolvase that C-A = T-G steps, which in proteins such as CAP are used to generate sharp roll bends, here, frequently, are local roll maxima, even though they contribute little to the overall bending. They have a bending *potential* that is not used in this particular setting.

Table 23.3.4.2. *Sequence-dependent differential deformability in B-DNA. II. The Minor Canon*

These generalizations are illustrated by Fig. 23.3.4.9, and are justified at greater length by El Hassan & Calladine (1997) and Dickerson (1998*a*,*b*,*c*).

(6) *Heterogeneous steps ending in A: C-A, T-A and G-A*

Steps ending in adenine, aside from A-A, tend to display (*a*) negative correlation between slide and roll, and between twist and roll, and (*b*) positive correlation between slide and twist.

(7) *Purine-pyrimidine steps*

R-Y steps display, on average, a systematic preference for negative slide and for twist below 36°.

(8) *Relative step frequencies in sequence-specific protein–DNA complexes*

Step A-A is the most common of all, and in 55% of the cases it occurs within A-tracts.

Steps containing only GC base pairs are least common, and seemingly are less compatible with formation of sequence-specific protein complexes.

(9) *Local environment and DNA behaviour*

Sequence-dependent local helix deformations are quite similar in DNA crystals and in protein–DNA complexes. DNA molecules packed against proteins in their normal biological environment appear to have more in common with DNA packed against other DNA helices in the crystal than with free DNA in solution.

appears to pass through or near the purines, while pyrimidines at the other end of the pairs stack O2-on-ring as with Y-R steps. R-Y steps tend to stack ring-on-ring, with little contribution from exocyclic atoms.

El Hassan & Calladine (1997) have recently examined roll, slide and twist behaviour at 400 different steps observed in crystal structures of 24 A- and 36 B-DNA oligomers. The author has carried out a similar analysis of 1137 steps from 86 sequence-specific protein–DNA complexes (Dickerson, 1998*a*,*c*; Dickerson & Chiu, 1997). A striking feature is that trends in local parameters are just the same in DNA crystals and in protein–DNA complexes. The frequently invoked nightmare of 'crystal packing deformations' appears to be of only minor significance. In both studies (El Hassan & Calladine, 1997; Dickerson, 1998*b*), roll *versus* slide, slide *versus* twist and twist *versus* roll plots are presented for all ten possible base-pair steps. Fig. 23.3.4.9 illustrates roll *versus* slide plots for two Y-R, two R-R and two R-Y steps.

Table 23.3.4.2 summarizes observations from these roll/slide/twist plots. These are labelled the 'Minor Canon' since they are recent, approximate and not well understood. However, they provide goals for future investigations of helix behaviour.

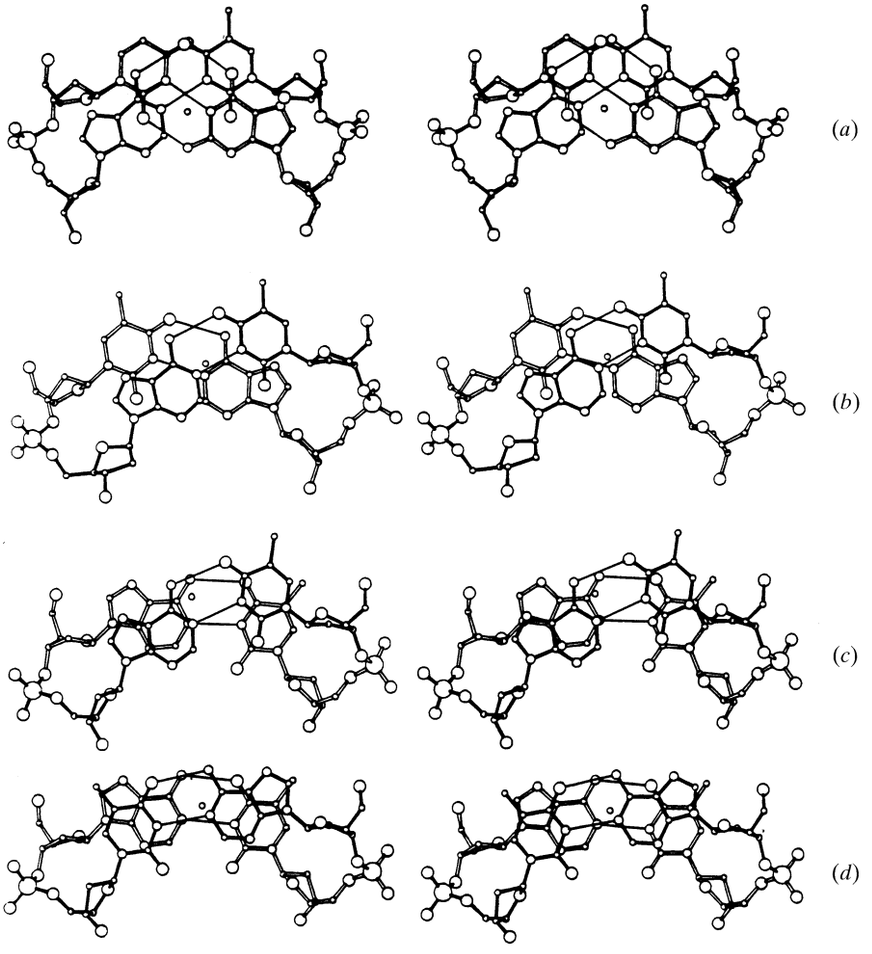

Fig. 23.3.4.8. Representative base-pair steps from B-DNA single-crystal X-ray analyses. (*a*) Pyrimidine-purine C-A step from C-C-A-A-G-A-T-T-G-G (B22, B46) (roll/slide/twist = $-7.4°/2.6$ Å$/49.9°$). Note the lack of ring-on-ring stacking, replaced by the stacking of pyrimidine O2 and purine N6 or O6, on aromatic rings of the adjacent base pair. This stacking opens up the twist angle to an unusual 50°. Note also the large +2.6 Å slide, which positions pyrimidine O2 over the six-membered rings of the neighbouring purines. (*b*) Pyrimidine-purine T-A step from C-G-A-T-A-T-A-T-C-G (B62) (roll/side/twist = $3.8°/-0.2$ Å$/39.5°$). The stacking is similar to C-A, except that a near-zero slide positions pyrimidine O2 over the five-membered rings of purines. (*c*) Purine-purine A-A step from C-C-A-A-C-G-T-T-G-G (B46, B50) (roll/slide/twist = $8.8°/0.5$ Å$/28.7°$). Ring-on-ring overlap now predominates, with consequently lowered twist angle and essentially zero slide. Note that purines are more extensively stacked than pyrimidines, which appear to be approaching the O2-on-ring stacking of Y-R steps. (*d*) Purine-pyrimidine A-T step from C-G-A-T-A-T-A-T-C-G (B62) (roll/slide/twist = $5.2°/0.0$ Å$/25.2°$). Ring-on-ring stacking again lowers the twist angle and keeps slide around zero. Now there is no stacking of exocyclic N or O on neighbouring rings.

23.3.4.2. *A-tract bending*

It has long been known that introduction of short A-tracts into general-sequence B-DNA in phase with the natural 10–10.5 base-pair repeat produced overall curvature that could be detected *via* eletrophoretic gel retardation, ring-cyclization kinetics and other physical measurements in solution (Marini *et al.*, 1982; Wu & Crothers, 1984; Koo *et al.*, 1986; Crothers & Drak, 1992). However, the microscopic source of the observed macroscopic curvature remained unclear. Solution measurements alone cannot discriminate between three alternative curvature models: (1) local bending within the A-tracts themselves; (2) bending at junctions between A-tract B-DNA and general-sequence B-DNA; or (3) inherently straight and unbent A-tracts, with curvature resulting from removal of the normal writhe expected in general-sequence B-DNA (Koo *et al.*, 1990; Crothers *et al.*, 1990). The three curvature models are compared schematically in Fig. 10 of reference B77.

X-ray crystallographic results for DNA oligomers come down unequivocally in favour of model (3) above. Short A-tracts of four to six base pairs are straight and unbent in C-G-C-G-A-A-T-T-C-G-C-G (B1–B6), C-G-C-A-A-A-A-A-A-G-C-G (B20), C-G-C-A-A-A-A-A-T-G-C-G (B31), C-G-C-A-A-A-T-T-T-G-C-G (B17, B52), C-G-C-G-A-A-A-A-A-A-G-C (B64) and C-A-A-A-G-A-A-A-A-G (B105) (A-tracts are double-underlined). It has been claimed (Sprous *et al.*, 1995) and disputed (Dickerson *et al.*, 1994, 1996) that the observed straightness of crystalline A-tracts was only an artifact of crystal packing, or of the high levels of methyl-2,4-pentanediol (MPD) used in the crystallization. This concern now is put to rest by the observation that B-DNA packed against a protein molecule in its biological working environment behaves exactly the same as B-DNA packed against other DNA molecules in the crystal, as borne out by the roll/slide/twist studies of El Hassan & Calladine (1997) for DNA and of Dickerson (1998*a*,*b*,*c*) and Dickerson & Chiu (1997) for protein–DNA complexes. Added support has come from recent molecular-dynamics simulations by Beveridge and co-workers (Sprous *et al.*, 1999), who have demonstrated that the duplex of sequence GGGGGGAA-AATTTTCGAAAATTTTCCCCCC is severely curved because of a roll kink at the double-underlined central CG step, whereas the duplex GGGGGGTTT-