*International Tables for Crystallography* (2006). Vol. F, Chapter 23.3, pp. 588–622.

# 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 $\alpha$ through $\zeta$, as shown in Fig. 23.3.2.2. An earlier convention termed these same six angles as $\omega$, $\varphi$, $\psi$, $\psi'$, $\varphi'$, $\omega'$ (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.

five *exo*. In Fig. 23.3.2.4 (top), pushing the C3′ atom of the C3′-*endo* conformation into the plane of the ring would tend to push C2′ below the ring, passing through a T state and creating a C2′-*exo* conformation. C2′ can, in turn, be returned to the ring plane if C1′ is pushed above the ring, forming C1′-*endo*, and so on, around the ring. In this way, a contiguous series of alternating *endo/exo* conformations is produced, as listed in Table 23.3.2.2.

This ten-conformation *endo/exo* cycle can be generalized to a continuous distribution of intermediate conformations, characterized by a pseudorotation angle, *P* (Altona *et al.*, 1968; Altona & Sundaralingam, 1972), with the ten *endo/exo* conformations spaced 36° apart (Table 23.3.2.2). Fig. 23.3.2.5 shows the calculated potential energy of conformations around the pseudorotation cycle (Levitt & Warshel, 1978). Note that C2′-*endo* and C3′-*endo* are most stable, that the pathway between them along the right half of the circle remains one of low energy, but that a large 6 kcal mol$^{-1}$ potential energy barrier (1 kcal mol$^{-1}$ = 4.184 kJ mol$^{-1}$) effectively forbids conformations around the left half of the circle.

As Fig. 23.3.2.4 indicates, the main-chain torsion angle, $\delta$, is sensitive to ring conformation, because the C5′—C4′ and C3′—O3′ bonds that define the angle shift as ring puckering changes. The idealized relationship between torsion angle, $\delta$, and pseudorotation angle, *P* (Saenger, 1984), is

$$\delta = 40^\circ \cos(P + 144^\circ) + 120^\circ.$$

Fig. 23.3.2.6 shows the observed torsion angles, $\delta$, and pseudorotation angles, *P*, from X-ray crystal structure analyses of synthetic DNA oligonucleotides: 296 examples from A-DNA and 280 from B-DNA. The most striking aspect of this plot is the radically different behaviour of A- and B-DNA. The prototypical sugar conformation for A-DNA obtained from fibre diffraction modelling, C3′-*endo*, is, in fact, adhered to quite closely in A-DNA crystal structures.

However, B-DNA shows a quite different behaviour. Although earlier fibre diffraction led one to expect C2′-*endo* sugars, the actual experimental distribution is quite broad, extending up the right-hand side of the pseudorotation circle of Fig. 23.3.2.5, through C1′-*exo*, O1′-*endo* and C4′-*exo*, in some cases all the way to C3′-*endo* itself. Indeed, the mean value of $\delta$ observed in B-DNA oligomer crystal structures is 128° rather than 144° (Table 23.3.2.1), making C1′-*exo* a better description of sugar conformation in B-DNA than C2′-*endo*. Old habits die hard, however, and the B-DNA sugar conformation is still colloquially termed C2′-*endo*, a designation of historical significance but of little practical value. The apparent greater malleability of the B helix compared to A may indeed be one feature that makes B-DNA particularly suitable for expressing its base sequence to drugs and control proteins *via* local helix structure changes.

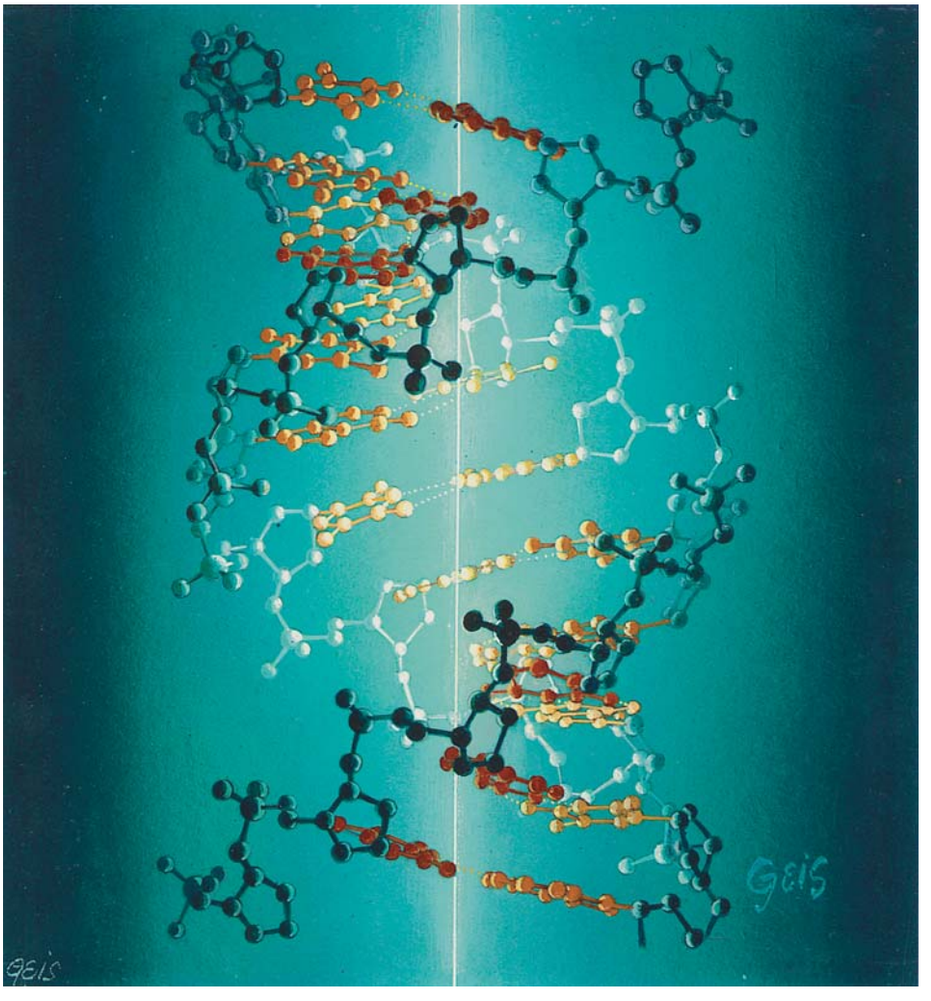


Fig. 23.3.1.1. 'Hot wire' painting of A-DNA by Irving Geis. Geis produced two dramatic paintings of horse-heart cytochrome *c*, in which the sole light source was the central iron atom within the haem, producing a glowing 'molecular lantern' effect. One painting showed this central luminous haem surrounded by hydrophobic side chains; the other featured the polar side chains extending out from the surface. These are to be seen today on the front and back covers of Voet & Voet's *Biochemistry* (Voet & Voet, 1990, 1995). In the present A-DNA painting, Geis chose the imaginary central axis of the helix as a monofilament light source, thereby reversing the conventional illumination: atoms lining the deep major groove glow brightly, whereas the outer surface of the helix is in dark silhouette. Geis struggled with the B helix as an artistic subject, but was never satisfied with the results. Hence, this glowing A-DNA helix represents his nucleic acid artistic legacy. Reprinted courtesy of the estate of Irving Geis. Rights owned by Howard Hughes Medical Institute.

### 23.3.2.3. *Base pairing*

The key to the biological role of DNA is that one of the two purines can pair with only one of the pyrimidines: A with T, and G with C. Hence, genetic information present in one strand is passed on to the complementary strand. The standard two-base pairs are shown in Fig. 23.3.2.7 along with the conventional numbering of the atoms. Backbone sugar and phosphate atoms are primed while base atoms are unprimed, as, for example, C1′ and N9 at opposite ends of a purine glycosydic bond. The G·C base pair is held together by three hydrogen bonds, whereas an A·T pair has only two. This means that A·T pairs show less resistance to propeller twisting (counter-rotation of the two bases about their common long axis), and this will have an effect on minor groove width, as seen later. The patterns of hydrogen-bond acceptors (*A*) and donors (*D*) on the major and minor groove edges of base pairs are important elements in recognition of base sequence by drugs and control proteins.

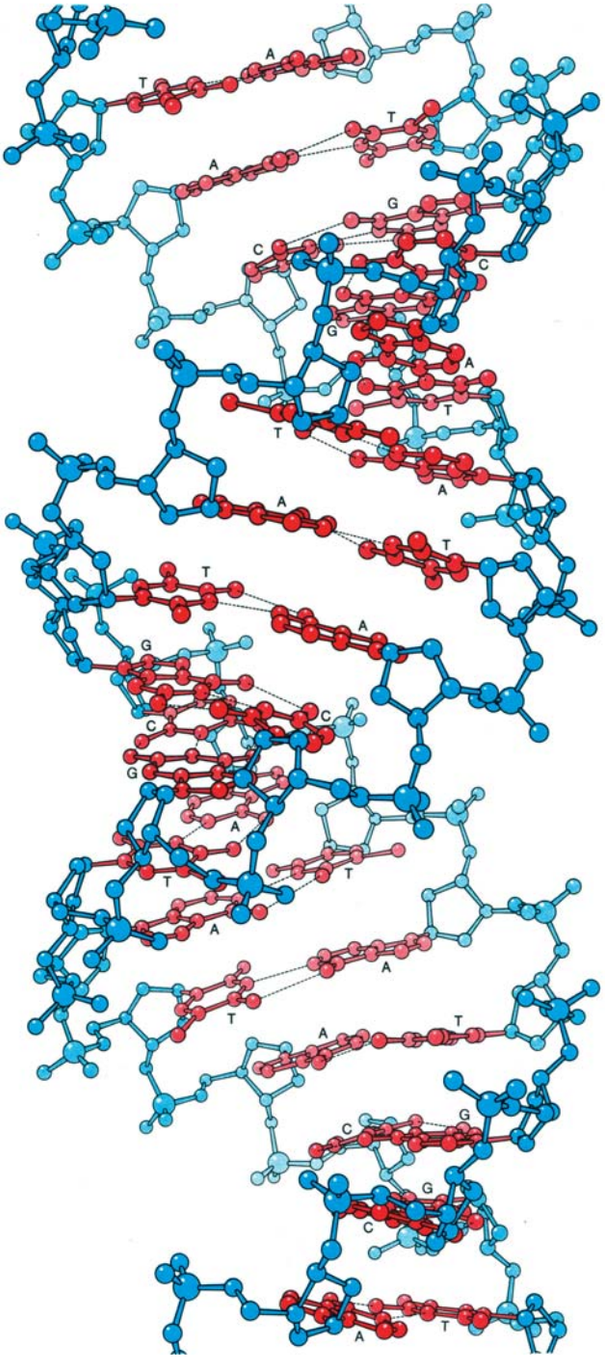

Fig. 23.3.1.2. Infinite A-DNA helix, generated from the X-ray crystal structure of the hexamer G-G-T-A-T-A-C-C (references A2 and A7 in Table A23.3.1.1) by deleting the outer base pair from each end and stacking images of the resulting truncated hexamer so their outer phosphate groups overlapped. This generates an endless helix that exhibits the local structural features of the X-ray crystal structure. Note the degree to which the A helix resembles an antiparallel double-stranded ribbon wound around an invisible helical core (the 'hot wire' axis of Fig. 23.3.1.1). (From Dickerson, 1983.) Reprinted courtesy of the estate of Irving Geis. Rights owned by Howard Hughes Medical Institute.

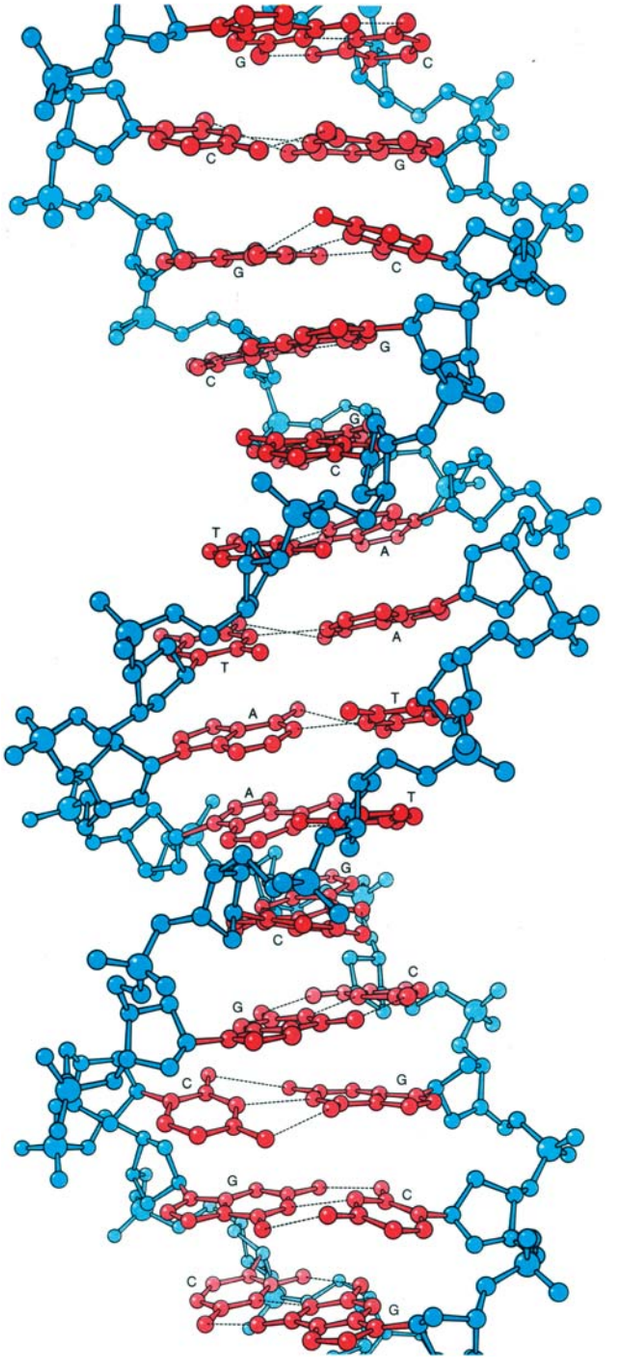

Fig. 23.3.1.3. Infinite B-DNA helix, generated in a similar manner to Fig. 23.3.1.2 from the central ten base pairs of the dodecamer C-G-C-G-A-A-T-T-C-G-C-G (B1–B5). Note that the minor groove is narrow in the AT region facing the viewer at the centre, but appreciably wider in the GC regions on the back side of the helix at top and bottom. Propeller twisting, or deviations of bases from coplanarity within one pair, is one sequence-dependent aspect of DNA that was not suspected from the averaged structures obtained from fibres. (From Dickerson, 1983.) Reprinted courtesy of the estate of Irving Geis. Rights owned by Howard Hughes Medical Institute.

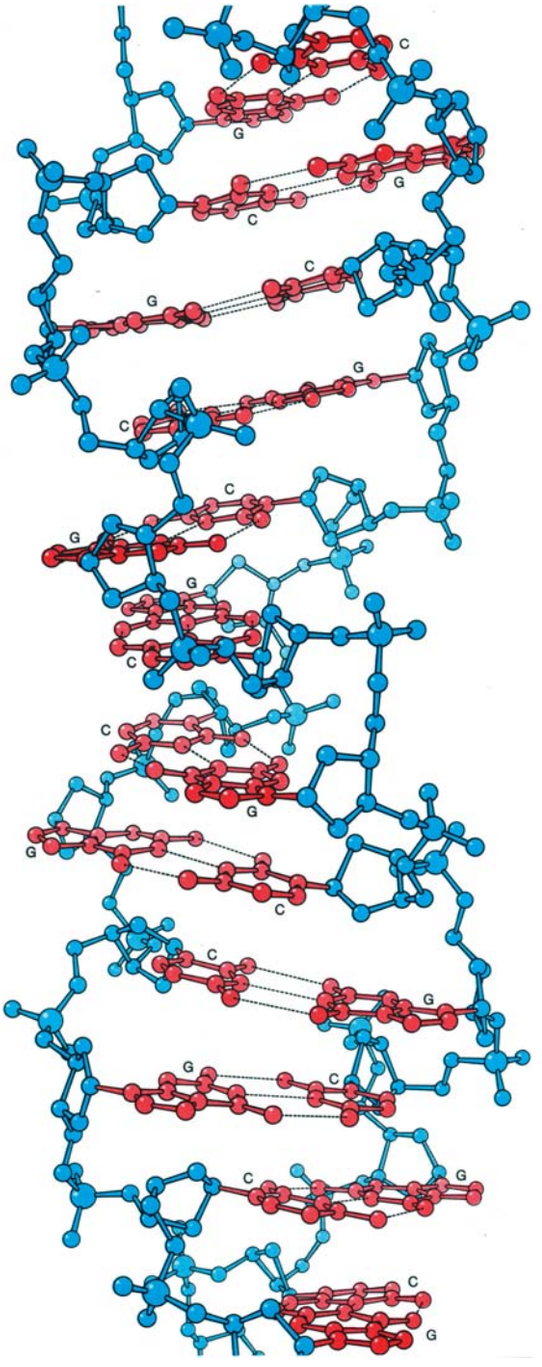

Fig. 23.3.1.4. Infinite Z-DNA helix, generated as before from the central four base pairs of the hexamer C-G-C-G-C-G (Z1). G and C bases alternate along each chain. The sugar–phosphate backbone adopts a pronounced zigzag pathway, rising vertically past each guanine, but travelling horizontally across the helix at cytosines. Hence, the formal helix repeat is two base pairs, G followed by C, rather than a single base pair, as in the A and B helices. Note that the structures of Z-DNA and A-DNA are in many ways the inverse of one another. The Z helix is left-handed, tall and slim, with a deep minor groove, a flattened major groove and small propeller twist. The A helix is right handed, short and broad, with a deep major groove, a shallow minor groove and large propeller twist. (From Dickerson, 1983.) Reprinted courtesy of the estate of Irving Geis. Rights owned by Howard Hughes Medical Institute.

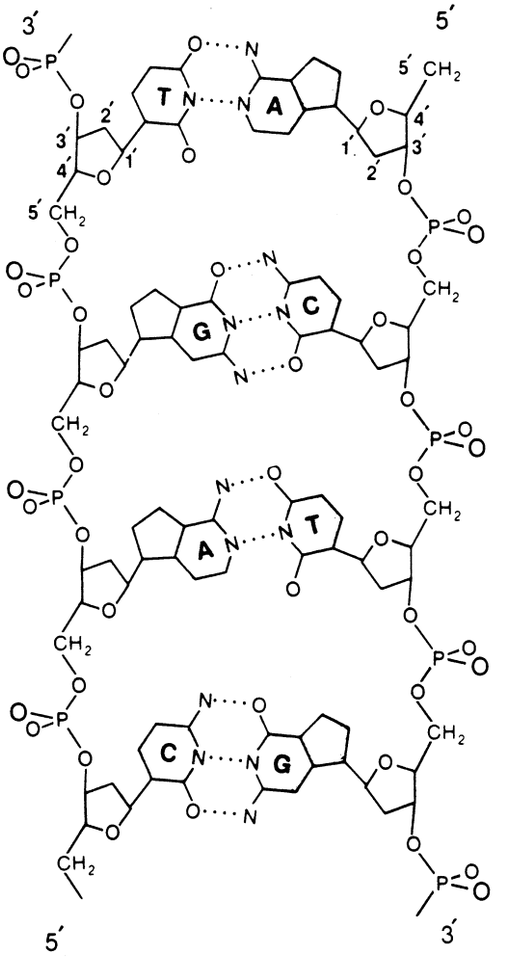

Fig. 23.3.2.1. Unrolled schematic of A- or B-DNA, viewed into the minor groove. Paired bases are attached to backbone chains that run in opposite directions: downward on the right and upward on the left. Z-DNA differs from A- and B-DNA in that the two backbone chains run in opposite directions from those shown here. Hence, Z-DNA cannot be obtained from A- or B-DNA by simple twisting around the helix axis.

Other related but nonstandard base pairs are compared in Fig. 23.3.2.8. Inosine (I) is useful in studying properties of DNA in that, when paired with cytosine (C), it creates a G·C-family base pair having overall similarity to A·T. Similarly, diaminopurine (DAP) [also known as 2-aminoadenine (2aA)], when paired with thymine (T), creates a G·C-like pair from A·T-family bases. Hence, in a given experimental situation, one can unscramble the relative significance of number of hydrogen bonds *versus* identity and location of exocyclic groups.

The conventional Watson–Crick base pairing of Fig. 23.3.2.7 uses the hexamer 'end' of the purine base. A different type of base pairing was proposed many years ago by Hoogsteen (1963), in which the upper edge of the purine was used: N7 and N6/O6. Hoogsteen base pairing is shown between the left-hand two bases in each part of Fig. 23.3.2.9. Note that in Hoogsteen base pairing of A and T, each ring provides both a hydrogen-bond donor and an acceptor. Guanine cannot do this, since both its N7 and O6 positions are acceptors. As a consequence, in a G·C pair, C must supply both of the hydrogen-bond donors. It can only form a Hoogsteen base pair with G when the cytosine ring is protonated. This would lead one to expect triplex formation only at low pH. However, the stability of a triplex can, to a certain extent, alter the p$K_a$ of the N—H proton itself. (Recall the shift in p$K_a$ of buried Asp and His

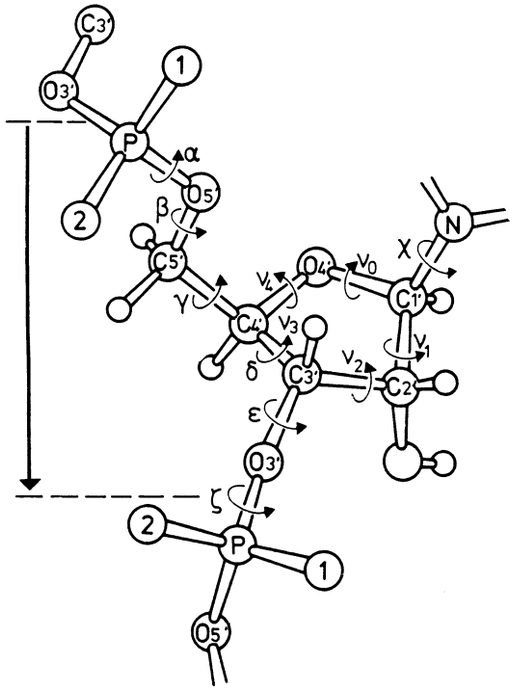


Fig. 23.3.2.2. Sugar–phosphate backbone of RNA and DNA polynucleotides. One nucleotide begins at a phosphorus atom and extends just short of the phosphorus atom of the following nucleotide, with the conventional positive direction being P→O5′—C5′—C4′—C3′—O3′→P, as indicated by the arrows. Main-chain torsion angles are designated $\alpha$ through $\zeta$, and torsion angles about the five bonds of the ribose or deoxyribose ring are $\nu_0$ through $\nu_4$, as shown. If one imagines atoms O3′—P—O5′ as a hump-backed bridge, as one crosses the bridge in a positive chain direction, oxygen atom O1 is to the left and O2 is to the right. These oxygens, accordingly, are sometimes designated $O_L$ and $O_R$. The —OH group attached to the C2′ atom of the ribose ring in RNA shown here is replaced by —H in the deoxyribose ring of DNA. Atom N to the right is part of the base attached to the sugar ring: N1 in pyrimidines and N9 in purines. Torsion angle $\chi$ is defined by O4′—C1′—N1—C2 in pyrimidines and O4′—C1′—N9—C4 in purines.

groups in the active sites of enzymes.) Hence, with a single-chain DNA, G-A-G-A-G-A-A-C-C-C-C-T-T-C-T-C-T-T-C-T-T-T-C-T-C-T-C-T-T, that folds back upon itself twice to build a triplex, NMR experiments indicate a significant amount of triplex remaining even at pH 8.0 (Sklenár & Feigon, 1990; Feigon, 1996).

### 23.3.2.4. *Helix parameters*

An important advantage of single-crystal oligonucleotide structures over fibre-based models is that one can actually observe local sequence-based departures from ideal helix geometry. B-DNA fibre models indicated a mean twist of *ca* 36° per step, or ten base pairs per turn, whereas A-DNA fibre patterns indicated less winding: *ca* 33° per step or 11 base pairs per turn. Twist, rise per base pair along the helix axis, horizontal displacement of base pairs off that axis, and inclination of base pairs away from perpendicularity to the axis are all intuitively obvious parameters. But when single-crystal structures began appearing in great numbers in the mid-1980s, it became imperative that uniform names and definitions be used for these and for less obvious, but increasingly significant, local helix parameters.

An EMBO workshop on DNA curvature and bending, held at Churchill College, Cambridge, in September 1988, led to an agreement on definitions and conventions that was published

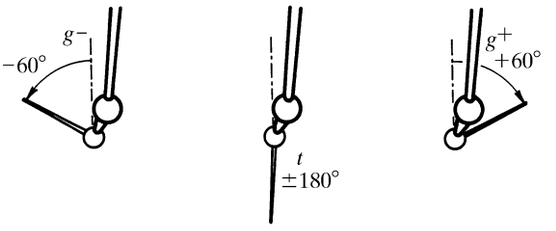


Fig. 23.3.2.3. Definition of torsion angles. A positive angle results from clockwise rotation of the farther bond, holding the nearer bond fixed. Torsion angle +60° is designated as *gauche*$^+$ or $g^+$, angle 180° is *trans* or *t* and angle −60° is *gauche*$^-$ or $g^-$.

simultaneously in four journals (Dickerson *et al.*, 1989). Fig. 23.3.2.10 shows the reference frames for two successive base pairs, and Figs. 23.3.2.11 and 23.3.2.12 illustrate local helix parameters involving rotation and translation, respectively. Subsequent experience has shown the most useful parameters to be inclination, propeller, twist and roll among the rotations, and *x* displacement, rise and slide among the translations. As mentioned at the beginning of this chapter, inclination and *x* displacement are the two properties that best differentiate A- from B-DNA. The four most widely used computer programs for calculation of local helix parameters are

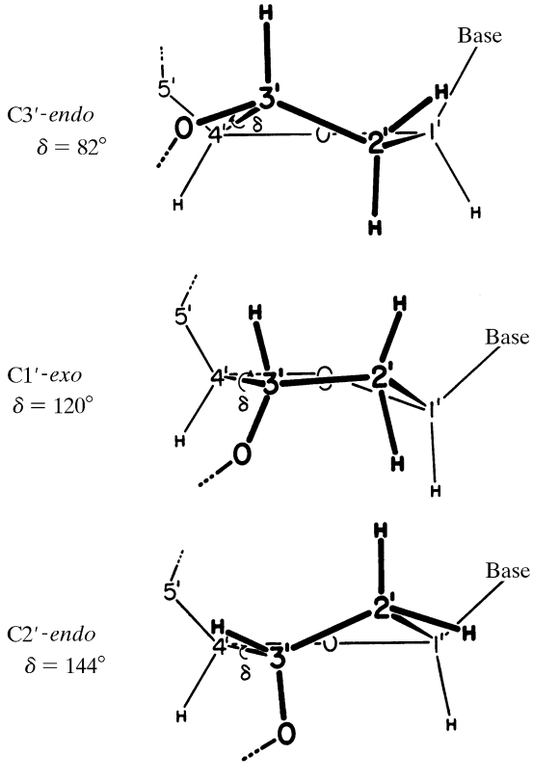


Fig. 23.3.2.4. The three most common furanose ring geometries. The planar form of the five-membered ribose or deoxyribose ring is unstable because of steric hindrance from side groups; one of the five atoms prefers to pucker out-of-plane on one side of the ring or the other. Puckering toward the same side of the ring as the C5′ atom is termed *endo*, and puckering toward the opposite 'outside' surface is termed *exo*. The main-chain torsion angle $\delta$ is related to sugar ring conformation because of the motion undergone by the C3′—O3′ bond during changes in puckering.

Table 23.3.2.1. *Average torsion-angle properties of A-, B- and Z-DNA (°)*

Values listed are mean torsion angles, with standard deviations in parentheses. Conformations are only approximate; — indicates a non-*gauche/trans* conformation. $B_{II}$ and $Z_{II}$ are less common variants. For $\delta$, the sugar ring geometry is quoted in place of *gauche/trans*. $\chi$ for B-DNA combines pyrimidines and purines. Values were obtained from a sample of 30 A-DNAs, 34 B-DNAs, 22 Z-DNAs and ten nonstandard DNAs in the Nucleic Acid Database. From Schneider *et al.* (1997).

| | $\alpha$ | $\beta$ | $\gamma$ | $\delta$ | $\varepsilon$ | $\zeta$ | $\chi$ |
|---|---|---|---|---|---|---|---|
| A-DNA | 293 (17) | 174 (14) | 56 (14) | 81 (7) | 203 (12) | 289 (12) | 199 (8) |
| Conformation | $g^-$ | *t* | $g^+$ | C3′-*endo* | *t* | $g^-$ | *t* |
| B-DNA | 298 (15) | 176 (9) | 48 (11) | 128 (13) | 184 (11) | 265 (10) | 249 (16) |
| Conformation | $g^-$ | *t* | $g^+$ | C1′-*exo* | *t* | $g^-$ | $g^-$ |
| $B_{II}$-DNA | | 146 (8) | | 144 (7) | 246 (15) | 174 (14) | 271 (8) |
| Conformation | | — | | C2′-*endo* | $g^-$ | *t* | $g^-$ |
| $Z_I$-DNA – purines | 71 (13) | 183 (9) | 179 (9) | 95 (8) | 95 (8) | 301 (16) | 63 (5) |
| Conformation | $g^+$ | *t* | *t* | O4′-*endo* | $g^+$ | $g^-$ | $g^+$ |
| $Z_{II}$-DNA – purines | | | | | 189 (12) | 52 (14) | 58 (5) |
| Conformation | | | | | *t* | $g^+$ | $g^+$ |
| $Z_I$-DNA – pyrimidines | 201 (20) | 225 (16) | 54 (13) | 141 (8) | 267 (9) | 75 (9) | 204 (98) |
| Conformation | *t* | — | $g^+$ | C2′-*endo* | $g^-$ | $g^+$ | *t* |
| $Z_{II}$-DNA – pyrimidines | 168 (16) | 166 (14) | | | | | |
| Conformation | *t* | *t* | | | | | |

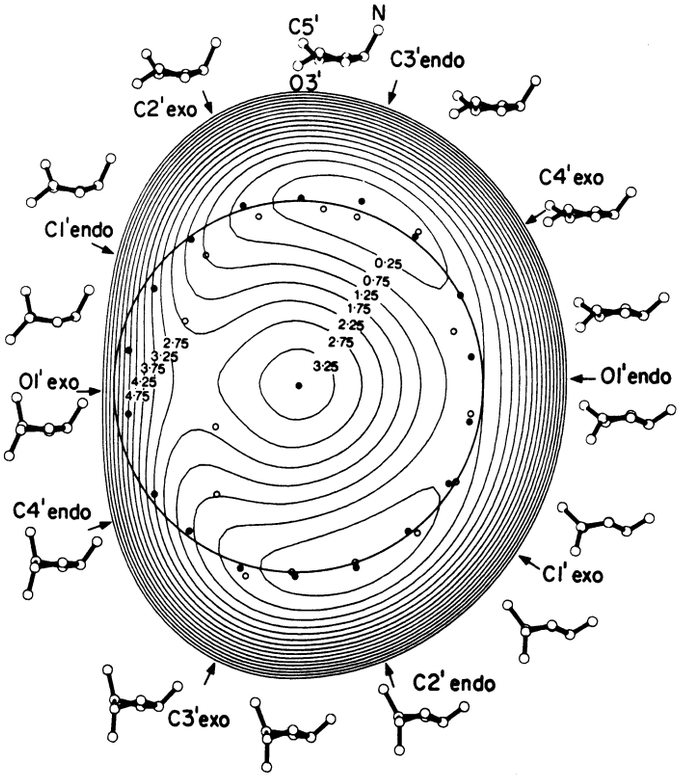


Fig. 23.3.2.5. Potential plot of all furanose ring conformations. Energies are in kcal $mol^{-1}$. The distance from the central point gives the maximum displacement of the out-of-plane atom from the plane of the other four. The circle is a constant-displacement trajectory chosen to pass through the potential minima on the right three-quarters of the plot. C2′-*endo* and C3′-*endo* are especially favoured, whereas O1′-*exo* on the left is highly disfavoured. The path from C2′-*endo* through C1′-*exo*, O1′-*endo* and C4′-*exo* to C3′-*endo* is a low-energy path, and many examples all along this path are known in B-DNA helices. Reprinted with permission from Levitt & Warshel (1978). Copyright (1978) American Chemical Society.

Table 23.3.2.2. *Sugar ring conformations, pseudorotation angles and torsion angle δ*

| Ring conformation | Pseudorotation angle (°) | Torsion angle $\delta$ (°) |
|---|---|---|
| C3′-*endo* | 18 | 82 |
| C4′-*exo* | 54 | 82 |
| O4′-*endo* | 90 | 96 |
| C1′-*exo* | 126 | 120 |
| C2′-*endo* | 162 | 144 |
| C3′-*exo* | 198 | 158 |
| C4′-*endo* | 234 | 158 |
| O4′-*exo* | 270 | 144 |
| C1′-*endo* | 306 | 120 |
| C2′-*exo* | 342 | 96 |

*NEWHELIX* by Dickerson (B7, B46), *CURVES* by Lavery & Sklenar (1988, 1989), *BABCOCK* by Babcock & Olson (Babcock *et al.*, 1993, 1994; Babcock & Olson, 1994) and *FREEHELIX* (Dickerson, 1998*c*). *NEWHELIX* was the earliest of these, but it performs all calculations relative to a best overall helix axis. This is satisfactory for single-crystal DNA structures, but makes the program unusable for the 180° bending observed in some protein–DNA complexes. *CURVES* is especially convenient for mapping the axis of a bent or curved helix. *FREEHELIX*, which evolved from *NEWHELIX*, calculates all parameters relative to local base-pair geometry, without assuming an overall axis, and permits display of normal vector plots that are especially useful in analysing bending in DNA–protein complexes (Dickerson & Chiu, 1997).

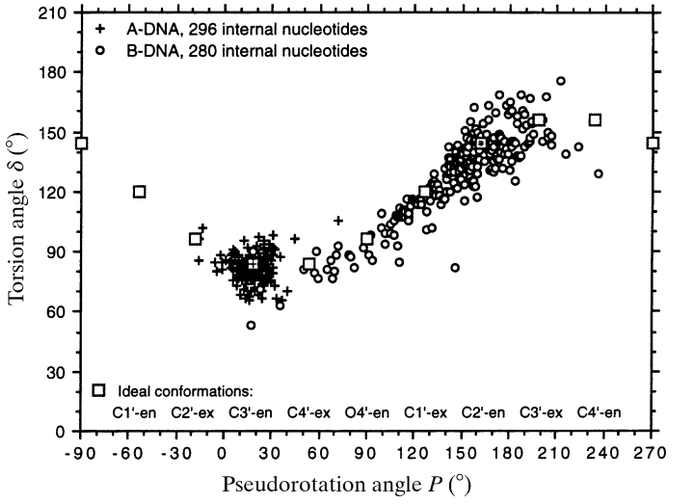


Fig. 23.3.2.6. Plot of observed sugar conformations in 296 nucleotides of A-DNA (crosses) and 280 of B-DNA (open circles). Open squares mark ideal relationships between torsion angle $\delta$ (vertical axis) and pseudorotation angle $P$ (horizontal axis) from the expression $\delta = 40° \cos(P + 144°) + 120°$. Deviations from this ideal curve for real helices arise, because the amplitude of pseudorotation (or displacement of one atom from the mean plane of the others) varies from one ring to another. Note the tight clustering of A-DNA points around C3′-*endo* and the broader distribution of B-DNA conformations.

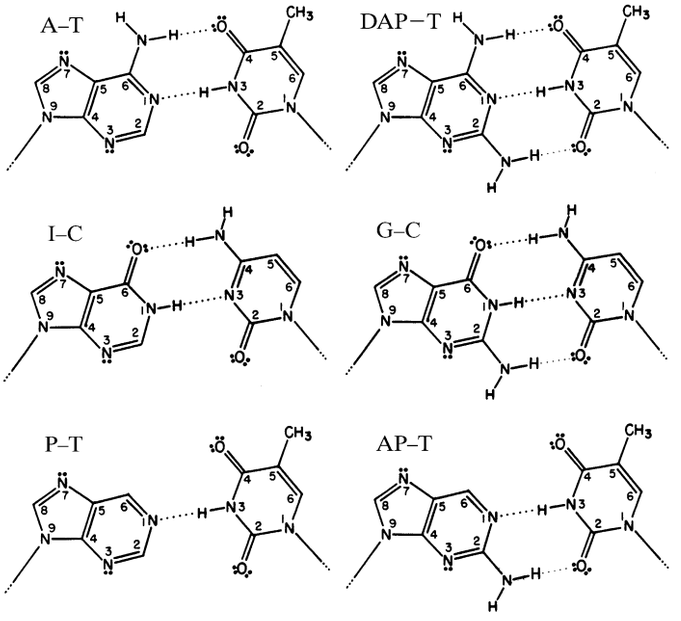


Fig. 23.3.2.8. Alternative purines and pyrimidines, and possible base pairings. Purines: P = purine; AP = 2-aminopurine; A = adenine or 6-aminopurine; DAP = 2,6-diaminopurine (also known as 2aA = 2-aminoadenine); G = guanine; I = inosine. Pyrimidines: T = thymine (uracil if methyl group is absent); C = cytosine. DAP–T is a nonstandard AT-family analogue of G–C, and I–C is a nonstandard GC-family analogue of A–T.

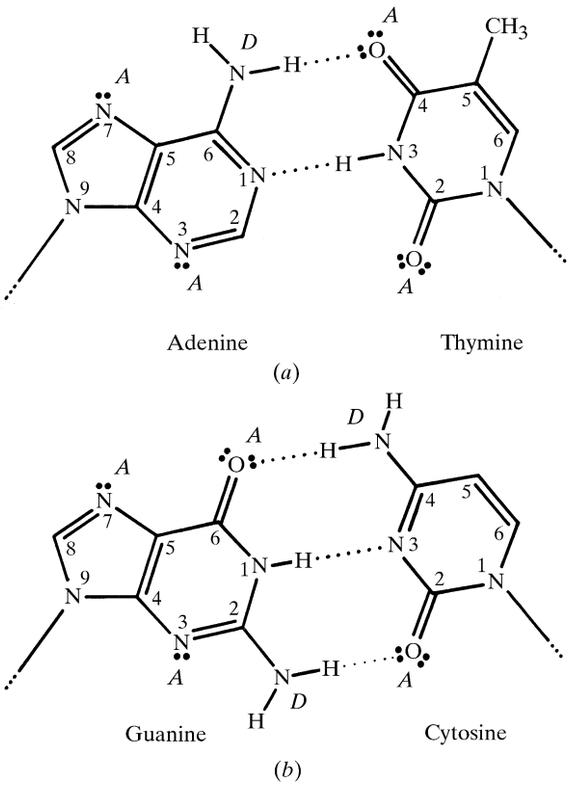


Fig. 23.3.2.7. A·T and G·C base pairs with minor groove edge below and major groove edge above. *A* is a hydrogen-bond acceptor, *D* is as hydrogen-bond donor.

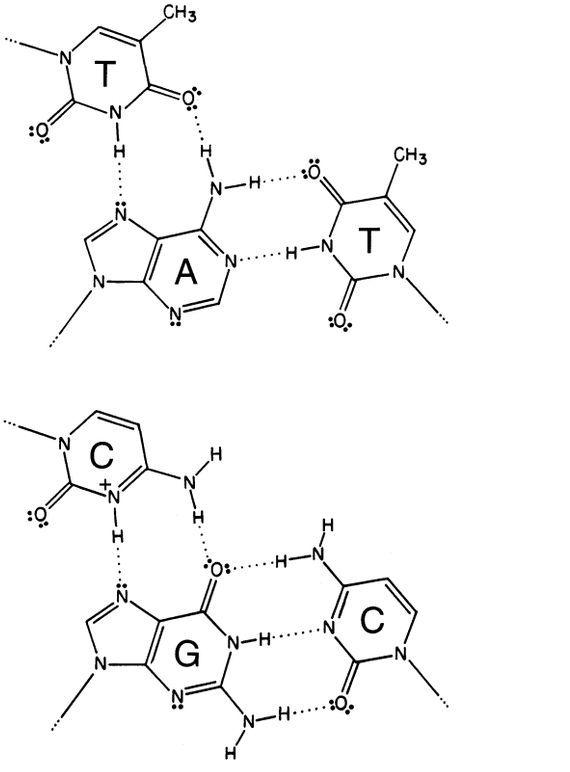


Fig. 23.3.2.9. Watson–Crick pairing of a purine (A or G) with a pyrimidine to its right (T or C), and Hoogsteen pairing of the same purine with a pyrimidine above it. This combination of Watson–Crick and Hoogsteen pairing is found in triple helices or triplexes. Note that Hoogsteen pairing of G and C can only occur at a pH at which C is protonated, because the extra proton is essential for the second hydrogen bond.

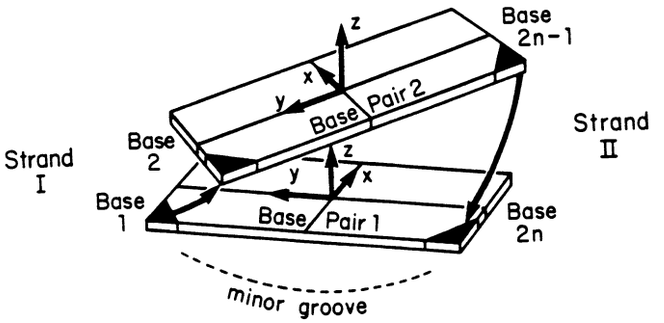


Fig. 23.3.2.10. Definitions of local reference axes ($x$, $y$, $z$) at the first two base pairs of an $n$-base-pair double helix. Base 1 is paired with base $2n$, base 2 with base $2n-1$ *etc*. Shaded corners represent attachment points to sugar rings. Curved arrows denote 5′-to-3′ 'positive' directions of each backbone chain. Note that when looking into the minor groove, as here, the two strands illustrate a clockwise rotation, upwards on the left and downwards on the right. This is true for A- and B-DNA, but for Z-DNA, the sense of the two backbone strands is reversed.

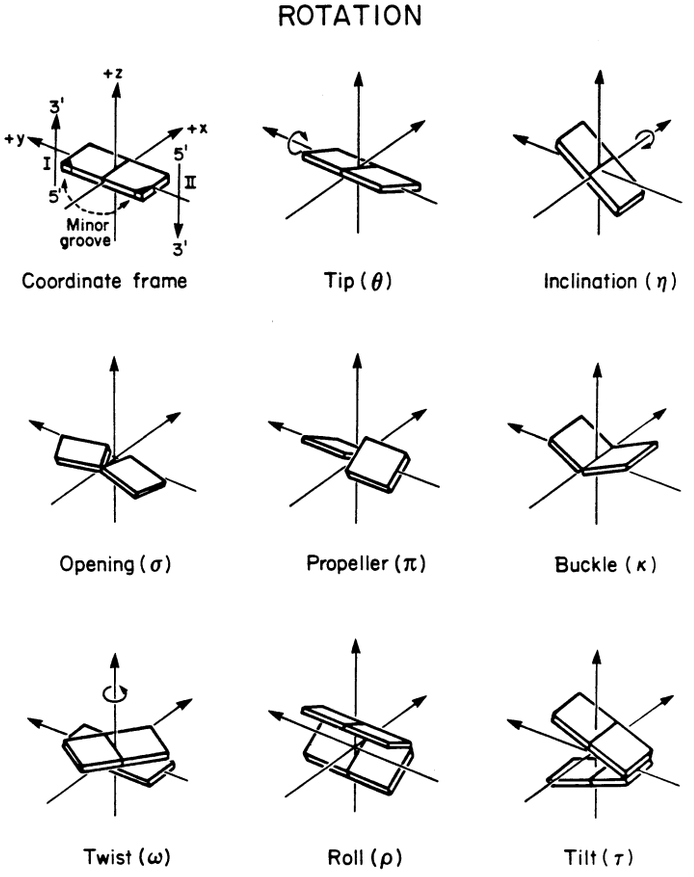


Fig. 23.3.2.11. Local helix parameters involving rotations. Tip and inclination describe the orientation of a base pair relative to the helix axis, produced by rotation about the base-pair long axis or short axis, respectively. Opening, propeller and buckle describe rotations of the two bases of a pair relative to one another. Twist, roll and tilt describe changes of orientation from one base pair to the next, *via* rotations about the $z$, $y$ and $x$ axes, respectively.

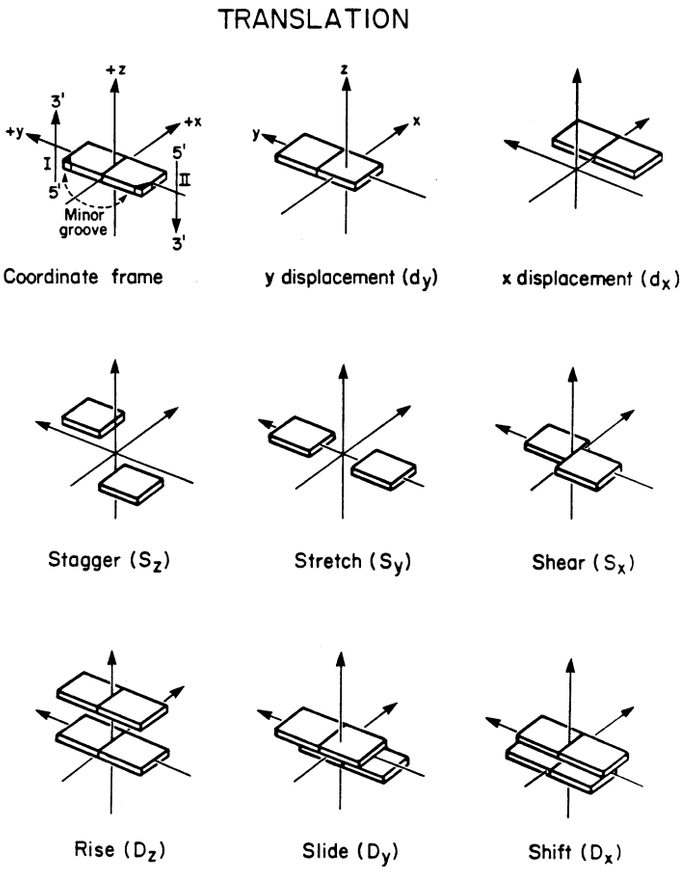


Fig. 23.3.2.12. Local helix parameters involving translations. $y$ and $x$ displacements describe shifts of a lone base pair along its long or short axis, respectively. Stagger, stretch and shear describe displacements of the two bases of a pair relative to one another. Rise, slide and shift describe displacements from one base pair to the next, *via* translations along the $z$, $y$ and $x$ axes, respectively.

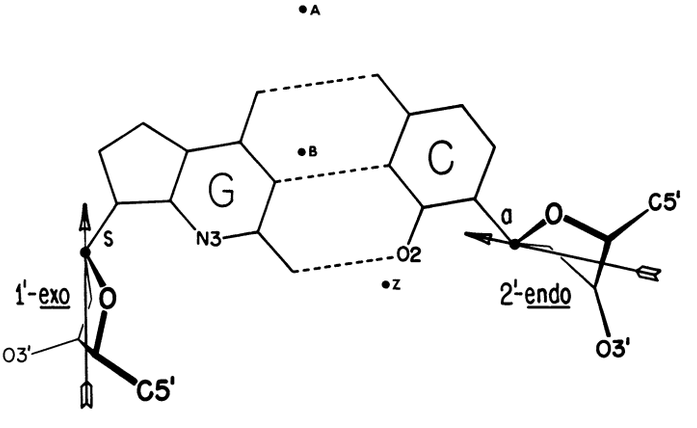


Fig. 23.3.2.13. *Syn versus anti* orientation about the glycosyl bond connecting sugar and base. Right: *anti* conformation, with $\chi$ *ca* 210°. Left: *syn* conformation, with $\chi$ around 60°. Both A- and B-DNA only employ the *anti* geometry; Z-DNA uses *anti* for pyrimidines and *syn* for purines, as shown here. Note that the 5′-to-3′ direction in both rings is down into the paper. Hence, antiparallel backbone chains can be achieved only by a zigzag chain geometry with local chain reversals, as shown later in Fig. 23.3.3.4. Black dots labelled A, B and Z indicate the position of the helix axis relative to the base pairs in A-, B- and Z-DNA.

23.3.2.5. *Syn/anti glycosyl bond geometry*

The glycosyl bond angle, $\chi$, about the bond connecting a sugar ring to a base is a special case of torsion angle, and is defined by O4′—C1′—N1—C2 for pyrimidines and O4′—C1′—N9—C4 for purines. In A- and B-DNA, the normal range of $\chi$ is 160 to 300°. This is known as the *anti* conformation (right-hand side of Fig. 23.3.2.13) and swings the sugar ring out away from the minor groove edge of the base pair. In Z-DNA, pyrimidines also exhibit the *anti* glycosyl bond conformation, but purines adopt the *syn* geometry shown on the left-hand side of Fig. 23.3.2.13. Now the sugar ring is rotated so that it intrudes into the minor groove, and $\chi$ lies in the range 50 to 90°.

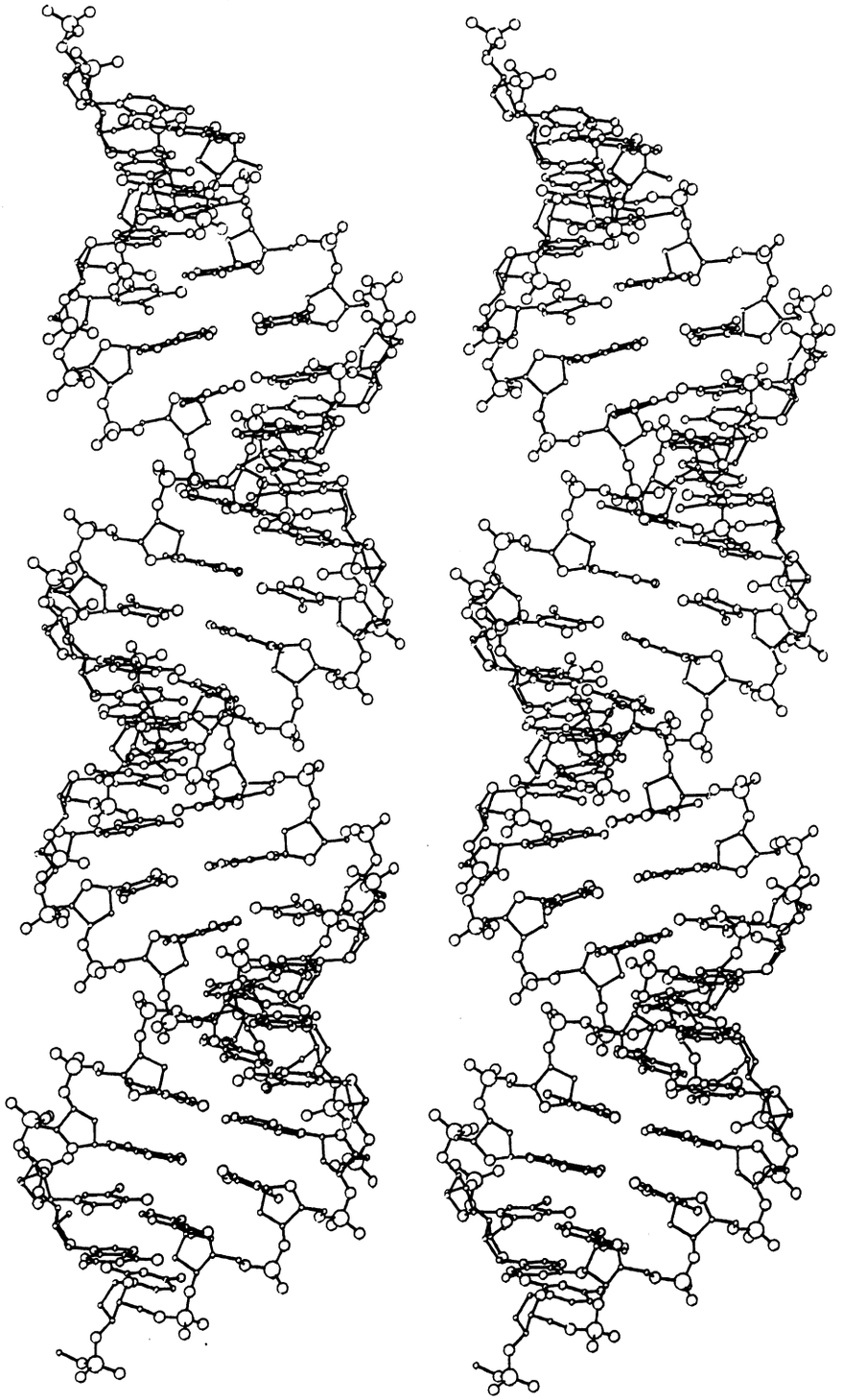

Fig. 23.3.3.1. The A-DNA stereo pair drawing from which Fig. 23.3.1.2 was derived, with repeating sequence -(G-T-A-T-A-C)$_n$-. The impression of the A helix as a ribbon wrapped around an imaginary core is even more strongly developed in this stereo. (From Dickerson, 1983.)

## 23.3.3. Comparison of A, B and Z helices

Figs. 23.3.3.1–23.3.3.3 show the original stereo pairs that were re-drawn by Irving Geis in preparing Figs. 23.3.1.2–23.3.1.4. These stereo pairs were constructed from X-ray structures of A-, B- and Z-DNA oligomers by deleting the outermost base pair from each end, eliminating the backbone as far as the first phosphate group, and then stacking these trimmed-down helices on top of one another, with phosphate groups overlapping, to create an infinite helix. They are improvements over the idealized infinite helices generated from fibre diffraction in that they display local variation in helix parameters that only single-crystal analyses can reveal. In the present context, they are good subjects for discussion of the differences between the three helix types.

23.3.3.1. *x displacement and groove depth*

A-DNA (Wahl & Sundaralingam, 1996, 1998), B-DNA (Berman, 1996; Dickerson, 1998*b*) and Z-DNA (Ho & Mooers, 1996; Basham *et al.*, 1998) have each been the subject of recent reviews, to which the reader is referred for details that cannot be covered here. The distinctive properties of the three helices are listed in Table 23.3.3.1. The most obvious distinction is handedness: A and B are right-handed helices, whereas Z is left-handed. Moreover, the position of each base pair relative to the helix axis is quite different. As noted in Fig. 23.3.2.13, the helix axis passes through base pairs in B-DNA, lies on the minor groove side of base pairs in Z-DNA, and on the major groove side in A-DNA. In terms of the helix parameters of Fig. 23.3.2.12, A-DNA has a typical *x* displacement of $d_x = +3$ to $+5$ Å, B-DNA has $d_x = -1$ to 0 Å, and Z-DNA has $d_x = -3$ to $-4$ Å. There is virtually no overlap between these three ranges; *x* displacement, $d_x$, in fact, is a better criterion for differentiating the three classes of helix than is sugar ring conformation.

A direct consequence of these *x* displacement values is great differences in depths of major and minor grooves. Both grooves are of equivalent depth in B-DNA because base pairs sit on the helix axis. In A-DNA, a base pair is pushed off-axis so

that its minor edge approaches the helix surface, making the minor groove very shallow and the major groove cavernously deep. In Z-DNA, it is the major edge of each base pair that is pushed toward the surface, so that the minor groove is deep and the major groove is so shallow as hardly to be characterized as a groove at all. It is sometimes stated that 'Z-DNA has no major groove', but space-filling stereos, such as Fig. 1 of reference Z6 or Fig. 3 of Z23 reveal the shallowest of major grooves running around the helix cylinder, flanked by very slightly higher phosphate backbones.

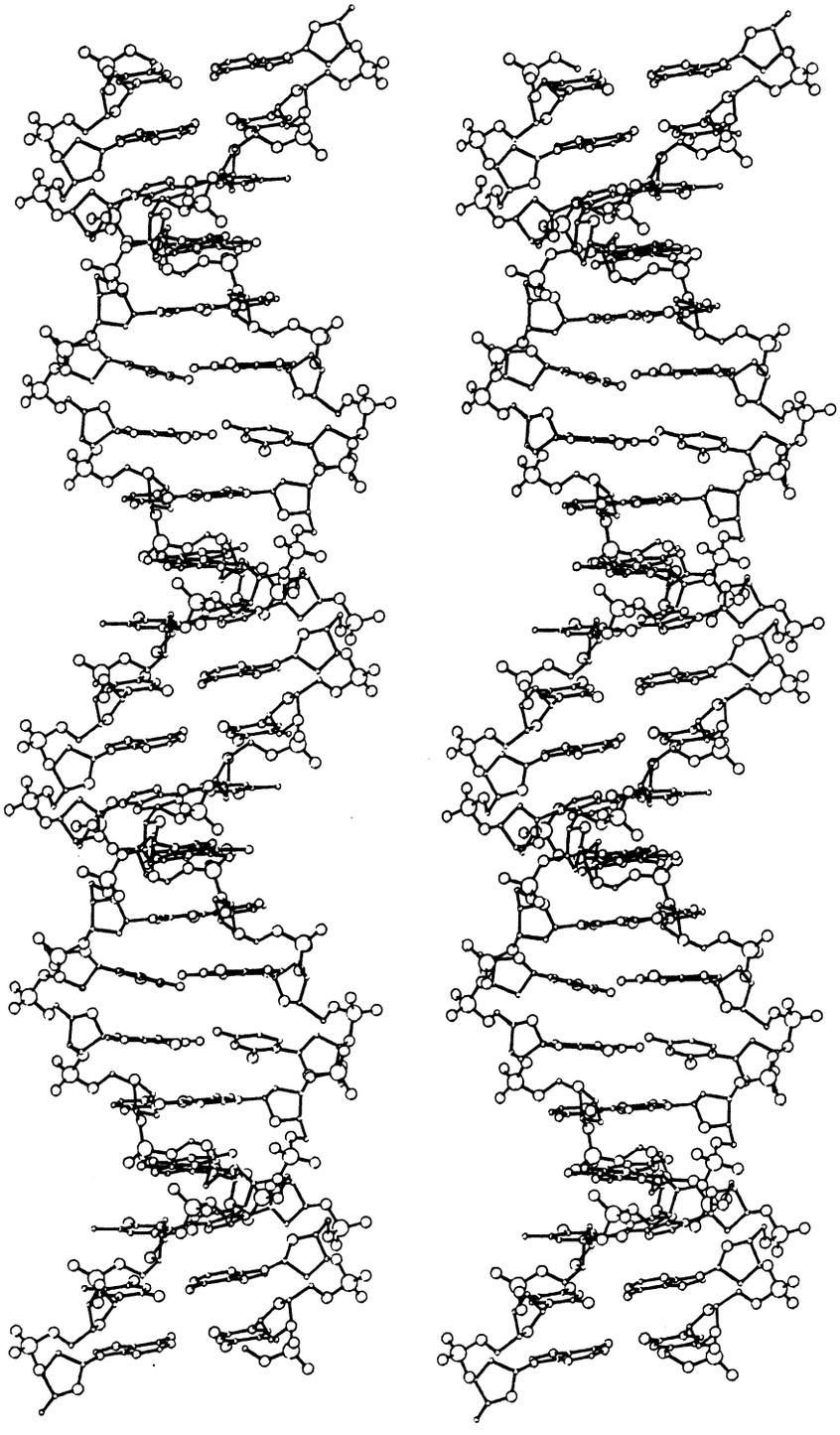

Fig. 23.3.3.2. The B-DNA stereo pair drawing from which Fig. 23.3.1.3 was derived, with repeating sequence -(G-C-G-A-A-T-T-C-G-C)$_n$-. The variation of minor groove widths on the front and back sides of the helix is striking. (From Dickerson, 1983.)

### 23.3.3.2. *Glycosyl bond geometry*

In both A- and B-DNA, all glycosydic bonds are *anti*, with sugar rings swung to either side away from the minor groove, as in Fig. 23.3.3.4(*a*). As mentioned earlier, when viewed into the minor groove, the backbone chains describe a clockwise rotation, with the chain on the right running downward, and that on the left upward, as in Fig. 23.3.2.1. In Z-DNA, both chains run in the *opposite* direction, leading to a counterclockwise rotation sense viewed into the minor groove. But Z-DNA has yet another striking (and defining) feature. Purines and pyrimidines alternate along each chain. G and C are most strongly favoured by far, but A and T can substitute intermittently at a price in stability. Breaking the strict alternation of purines and pyrimidines is even more unfavourable and is rarely encountered in crystal structures (Table A23.3.1.3). At each purine base, the glycosyl bond is rotated into the minor groove to the *syn* position, as in Fig. 23.3.3.4(*c*). This causes the local backbone directions, defined by sugar ring atoms C4′ and C3′, to be parallel in the two strands. Z-DNA avoids becoming a parallel-chain helix by performing a local chain reversal at each pyrimidine. In Fig. 23.3.3.4(*c*), although the local C4′–C3′ chain direction at the cytosine sugar is downward, the double loop in backbone chain gives it a net upward orientation. In stereo Fig. 23.3.3.3, the ascending backbone chain rises smoothly past each guanine, with a chain path parallel to the helix axis. However, the chain bends abruptly at right angles when passing a cytosine, in a direction tangential to the helix cylinder. Guanine sugar rings point their O4′ oxygen atoms in the backward chain direction (as is also true for all bases in A- and B-DNA), but cytosine sugars point their oxygens in the forward direction. This 'up at G, across at C' pathway and inversion of sugar rings is what produces the zigzag backbone pathway that leads to the name Z-DNA. The O4′ atom of each cytosine sugar is stacked on top of the guanine ring of the subsequent nucleotide, and this stacking of a polar O (or N) on top of a polarizable aromatic ring contributes to the stability of the Z helix, as it does to many other base–base interactions to be discussed later (Bugg *et al.*, 1971; Thomas *et al.*, 1982; B32).

### 23.3.3.3. *Sugar ring conformations*

Sugar ring conformations in A- and B-DNA have a logical structural basis. The B-DNA backbone is more extended than the A-DNA backbone, with P–P distances of *ca* 6.6 Å along one chain, compared with *ca* 5.5 Å in A-DNA. In turn, C2′-*endo* is a more extended ring conformation than C3′-endo, demonstrable in Fig. 23.3.2.4 by a greater distance between C5′ and O3′ atoms. Hence, it is logical that

the more extended ring conformation should be associated with the more extended backbone chain. In Z-DNA, the extended C2′-*endo* form is adopted at cytosine, where a zigzag double chain reversal must be accommodated, while the more compact C3′-*endo* occurs at the straight backbone segment running past a guanine.

The cramped *syn* glycosyl conformation is strongly disfavoured, although not absolutely forbidden, at pyrimidines, most probably because of steric clash between the pyrimidine O2 and the *syn* ring (Haschmeyer & Rich, 1967; Davies, 1978; Ho & Mooers, 1996; Basham *et al.*, 1998). Hence, the Z-DNA helix is effectively limited to alternating pyrimidine/purine sequences, with a price that must be paid for intermittent substitution of A and T for G and C, and an even higher price paid for breaking the pyrimidine/purine alternation. This is reflected in the X-ray crystal structures listed in Table A23.3.1.3. Only one non-alternating sequence has been completely solved and published: *C-G-G-G-*C-G (Z40), where adoption of the Z form has been forced by 5-methylation of cytosines (*C). A second non-alternating sequence that includes AT base pairs, *C-G-A-T-*C-G (Z13), was solved in 1985, but its coordinates have never been made public. It, too, required methylation of cytosines to induce the Z form. A third sequence, C-C-G-C-G-G (Z42), opens its terminal base pairs to make intermolecular base pairs with crystal neighbours. The 52 remaining Z-DNA structures in Table A23.3.1.3 all have strict alternation of pyrimidines and purines.

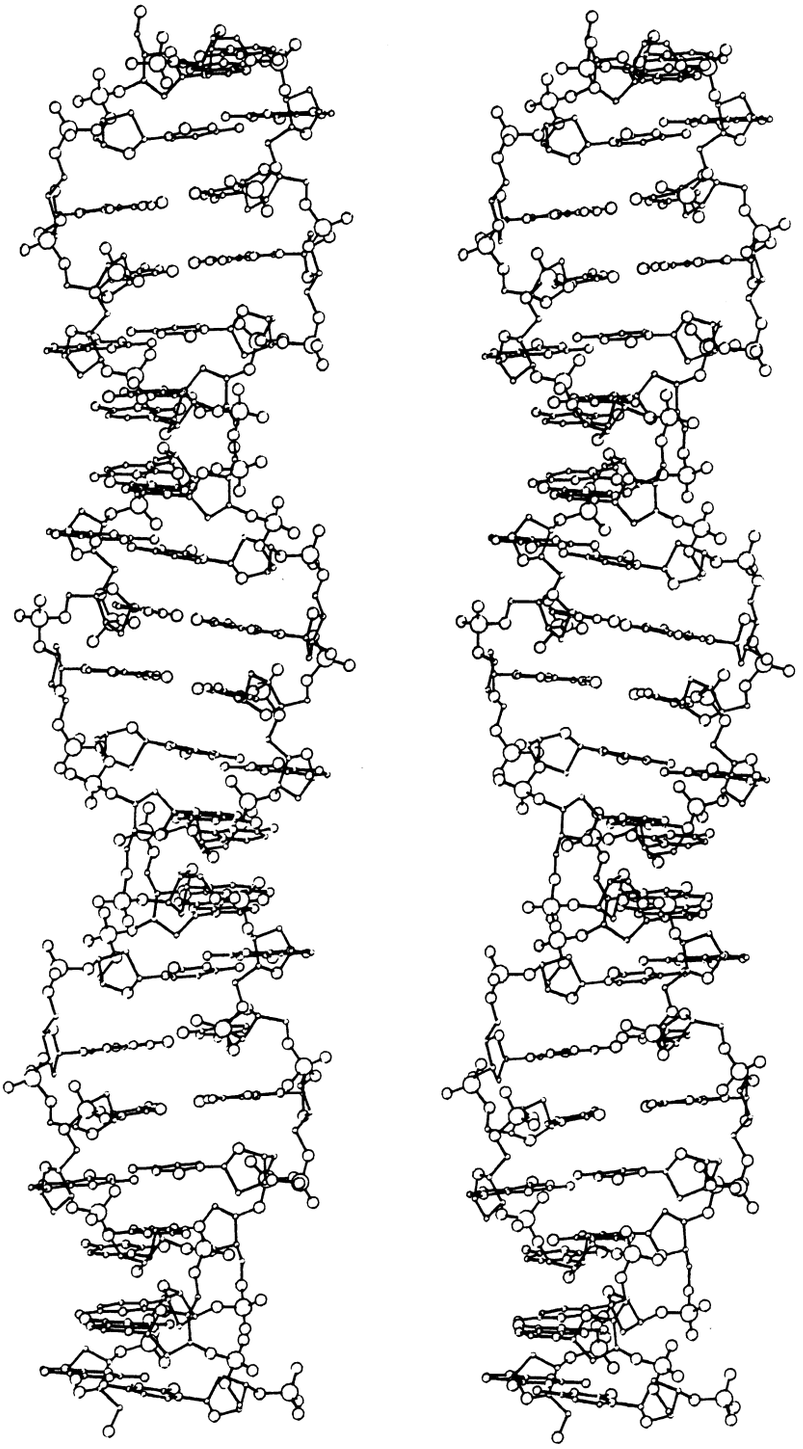

Fig. 23.3.3.3. The Z-DNA stereo pair drawing from which Fig. 23.3.1.4 was derived, with repeating sequence -(G-C-G-C)$_n$-. Note the left-handed zigzag path of the sugar–phosphate backbone, which led to its designation as the Z helix. (From Dickerson, 1983.)

### 23.3.3.4. *Helical twist and rise, and propeller twist*

The helical repeat unit in Z-DNA is therefore two successive base pairs, rather than the single base pair of A- and B-DNA. Ho & Mooers (1996) propose that the C-G or $^{5'}$pyrimidine-P-purine$^{3'}$ step be considered the fundamental unit of the Z-helical structure, because of the tight overlap between the two base pairs. As can be seen in Fig. 23.3.3.3, in a C-G step the pyrimidine rings from the two base pairs actually stack over one another, whereas the purine rings are packed against neighbouring sugar O4′ atoms. Helix-axis rotation at this step is only −8°, whereas the preceding and following G-C steps have a mammoth −52° twist. Hence, although Z-DNA has 12 base pairs per turn, it technically is not a dodecamer helix, but a hexamer with a two-base-pair repeating unit and a total rotation of −60° per unit.

This virtual restriction to purine/pyrimidine alternation means that Z-DNA cannot be involved in the coding of genetic information. A and B helices have no such restriction; their structures can accommodate a random sequence of bases. Average twist angles are as shown in Table 23.3.3.1, although extreme variation in twist is observed at individual steps in single-crystal structure analyses, from as little as 16° to as much as 55°. Base-sequence preferences for local helix parameters are discussed below.

In both B and Z helices, base pairs are very nearly perpendicular to the helix axis, whereas in the winding double ribbon of A-DNA, the long axis of each base pair is inclined by 10 to 20° away from perpendicularity to the axis. Hence, the rise per base pair for all B-helical steps and for G-C steps of Z-DNA is equal to the thickness of a base pair, 3.4 Å. The rise at a C-G step of

Table 23.3.3.1. *Comparison of structures of A, B and Z helices*

| | A | B | Z |
|---|---|---|---|
| Handedness | Right | Right | Left |
| Helix axis relative to base pairs | Major groove side | Through centre of base pair | Minor groove side |
| Major groove | Very deep and narrow | Wide, same depth as minor | Very shallow and broad |
| Minor groove | Shallow and broad | Variable, same depth as major | Very deep and narrow |
| Glycosydic bonds | *anti* | *anti* | C: *anti*<br>G: *syn* |
| Minor groove backbone chain sense * | Clockwise | Clockwise | Counterclockwise |
| Sugar conformation | C3′-*endo* (narrow range) | C1′-*exo*/C2′-*endo* (broad range) | C: C2′-*endo*<br>G: C3′-*endo* |
| Base pairs per helix repeat | 1 | 1 | 2 |
| Base sequence limitations | None | None | Alternating $(C\text{-}G)_n$ or close variants |
| Rise per base pair (average) | 2.9 Å | 3.4 Å | C-G: 4.1 Å<br>G-C: 3.5 Å |
| Base pair inclination | 10–20° | *ca* 0° | *ca* 0° |
| Mean twist angle | 30–33° | 34–36° | C-G: −8°<br>G-C: −52° |
| Helix repeats per turn | 11–12 | 10–10.5 | 6 (2 base pairs) |
| Propeller twist | Often substantial, 0–25° | Often substantial, 0–25° | Usually small |
| Common biological occurrence | RNA | DNA | None? |

* Relative 5′-to-3′ directions of the two backbone chains, when viewed into the minor groove.

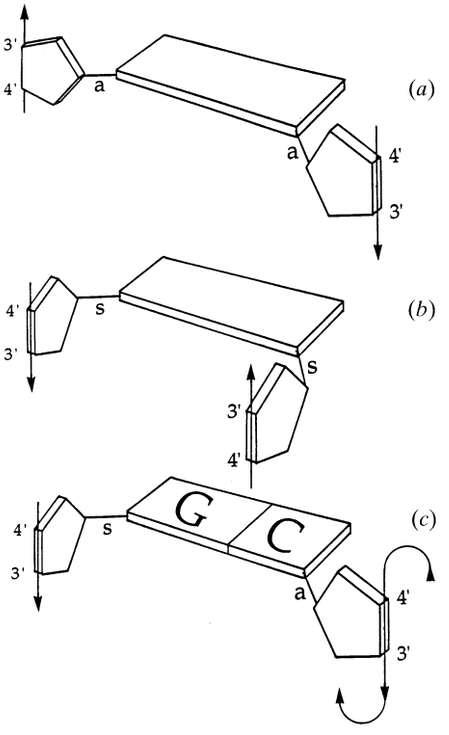


Fig. 23.3.3.4. Glycosyl conformation and chain sense. (*a*) Glycosyl conformations *anti/anti*, backbone chains antiparallel, with clockwise sense when viewed into the minor groove, as here. This is typical for A- and B-DNA. (*b*) Glycosyl conformation *syn/syn*, backbone chains antiparallel, with counterclockwise sense viewed into minor groove. This is not known for any nucleic acid duplex. (*c*) Glycosyl conformation *syn* at G and *anti* at C, with the C4′—C3′ edge of the sugar pointing downward in both strands, which would seem to imply a parallel-stranded helix. However, in Z-DNA, antiparallel strands are achieved by a local reversal of chain direction at each C, as shown here. This produces the zigzag backbone pathway that is characteristic of the Z helix, visible in Fig. 23.3.3.3.

Z-DNA is larger because it involves stacking of a sugar oxygen on each purine ring, not ring stacked on ring. For A-DNA, the rise along the helix axis can actually be less than the thickness of a base pair, because adjacent base pairs are stacked at an incline. The perpendicular distance from one base pair to the next in A-DNA is still 3.4 Å. Both A- and B-DNA exhibit considerable base pair propeller twist, especially at A·T pairs with only two hydrogen bonds rather than three. In contrast, Z-DNA, with predominately G·C pairs, shows only a small propeller twist.

The stacking of base pairs has immediate consequences for crystal growth. For Z-DNA, four base pairs are one-third of a helical turn, and six base pairs are a half turn. Hexamers are the most common crystal form in Table A23.3.1.3 by a large majority. In contrast, octamers and decamers are not simple fractions of a turn, and they stack in a disordered manner. One would predict that dodecamers of Z-DNA might crystallize well if the oligomers were not so long as to fall prey to cylindrical disorder.

By the same principles, B-DNA decamers stack easily and well to build pseudo-infinite helices through the crystal, with ordered cylindrical rods packed in six different space groups. The other common crystallization mode for B-DNA, the dodecamer, has a two-base-pair overlap of ends that both stablizes the crystals and yields a functional ten-base-pair repeat. (See Fig. 2 of Dickerson *et al.*, 1987.) Because the dodecamers are held by their outer two base pairs, the central eight pairs are unobstructed and accessible in the crystal, making dodecamers particularly good subjects for the study of minor-groove binding drugs.

A-RNA duplexes [Table A23.3.1.1, part (*k*)] also stack end-for-end in a manner simulating an infinite A helix, even though the end base pairs are inclined and are not perpendicular to the helix axis. This behaviour has been seen for octamers with roughly two-thirds of a helical turn, for nonamers, and for dodecamers with roughly a full turn.

In contrast, crystals of A-DNA behave quite differently. Regardless of chain length, A-DNA helices crystallize with the outer base pair of one helix packed against one wall of the broad, open and relatively hydrophobic minor groove of another helix.

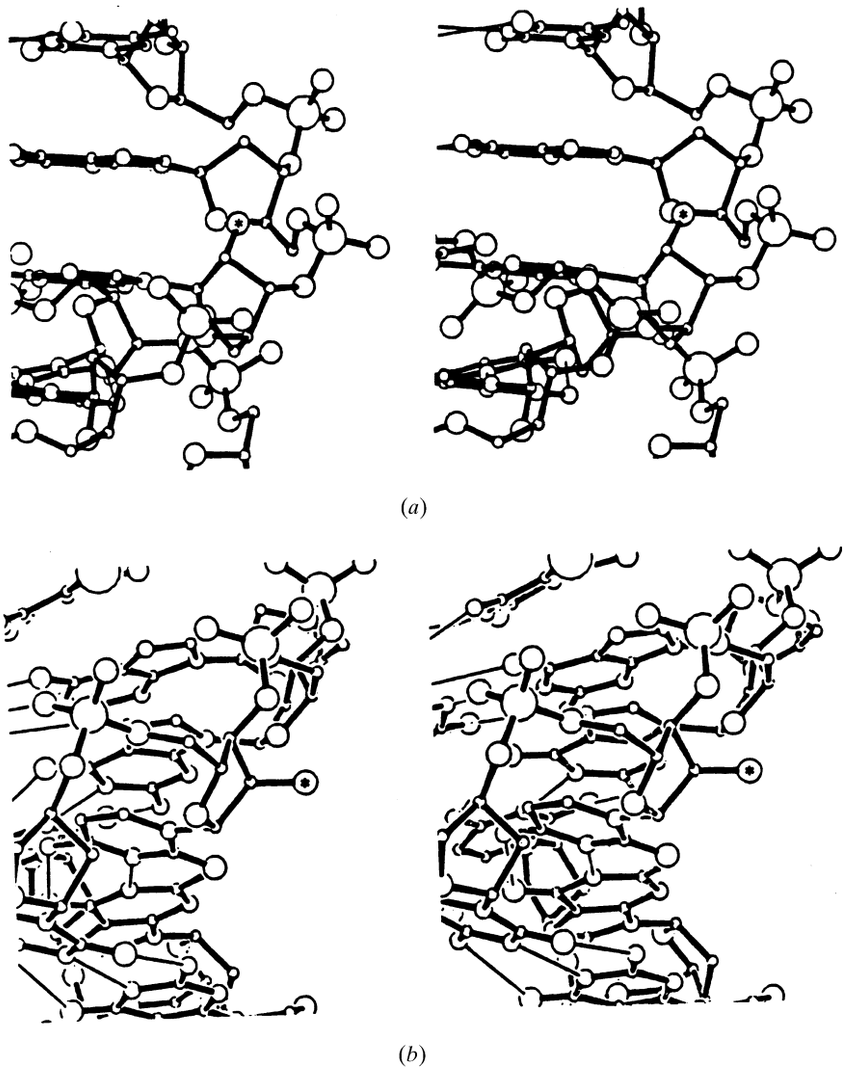

Fig. 23.3.3.5. The role of the C2′-OH in RNA helix geometry. (*a*) Addition of a C2′-OH group (*) to the B-DNA helix leads to close contacts and unallowable steric hindrance with the following O5′ and O4′ atoms and, to a lesser degree, the subsequent base itself. (*b*) A C2′-OH group added to the A-DNA helix extends outward radially from the helix cylinder surface and produces no steric clashes. Hence, A-RNA is quite possible, whereas B-RNA is disallowed.

This packing mode is sufficiently adaptable to accommodate duplexes of lengths four, six, eight, nine, ten and 12 base pairs. Hence, A-DNA does not simulate infinite helices through the crystal lattice, as A-RNA and B- and Z-DNA do.

### 23.3.3.5. *Allowable RNA helices*

So far this discussion has only been concerned with DNA. Which of the three helix types can be adopted by RNA? Fig. 23.3.3.5 shows that addition of a 2′-OH group to a B-DNA helix [part (*a*)] creates severe steric clash with the phosphate group and sugar ring of the following nucleotide, whereas in an A helix [part (*b*)], the added hydroxyl group extends radially outward from the helix cylinder and causes no steric problems. Hence, the natural helical form for RNA is the A helix, not the B helix. Table A23.3.1.1 shows several single-crystal analyses of A-RNA and RNA/DNA hybrids; Table A23.3.1.2 shows no B-RNA structures. One RNA/DNA hybrid is known as a Z helix: C-G-c-g-C-G (Z24), in which the two central nucleotides are RNA. If one mentally adds an —OH to each C2′ atom in Fig. 23.3.3.3, on the same side of the ring as O3′, it is apparent that the C2′-OH is not inherently incompatible with the Z helix, as it is with the B helix. At guanine sugars, the C2-OH points out and away from the helix, while at cytosine sugars it points away from the base into the spacious minor groove.

### 23.3.3.6. *Biological applications of A, B and Z helices*

The B helix is the biologically relevant structure for DNA. The A form might logically be adopted at the stage of transient DNA–RNA duplexes during transcription, but elsewhere the B form holds sway. It was once thought that binding of DNA to a protein surface, most particularly nucleosomal winding, might constitute a sufficient dehydration of bound water molecules from the DNA duplex to shift it to the A form. This proved to be false; nucleosomal DNA clearly retains the B conformation. The closest that one comes to biological A-DNA is local deformations upon binding of B-DNA to a few proteins that have been described as ‘A-like distortions’. On the other hand, the A helix has been found repeatedly in RNA duplexes, including tRNA and ribozymes.

The situation is even more restrictive with the Z helix. Although its alternating purine/pyrimidine sequence makes it unusable for genetic *coding*, the suggestion has been made on many occasions that Z-DNA might be an important element in genetic *control* by being involved in negative supercoiling (Herbert & Rich, 1996). It has been shown that a left-handed DNA conformation can be induced by negative superhelical stress, but it is not absolutely clear that this induced, left-handed conformation is the same as the Z helix seen in crystal structures of small oligomers. As noted by Herbert & Rich (1996), after nearly twenty years of enquiry, it is still far from certain that Z-DNA itself has any demonstrable biological role.*

A major stumbling block is the cumbersome mechanism that must be invoked to explain a B-to-Z interconversion. As mentioned previously, a simple twisting of the helix from right to left is not sufficient, because the backbone chains run in opposite directions in the two forms. Fig. 23.3.3.6 demonstrates the steps that must still be undertaken after both B and Z helices have been unwound so as to remove all of their helical character. Note the opposite sense of the backbone strands in B [part (*a*)] and Z [part (*e*)]. In order to accomplish the interconversion, base pairs of B-DNA must be

* Rich and co-workers (Schwartz *et al.*, 1999) have recently solved the crystal structure of the Z$\alpha$ domain of the human editing enzyme ADAR1 in a complex with a six-base-pair Z-DNA helix of sequence CGCGCG. This left-handed hexamer may suffer from the same length *versus* conformation uncertainty mentioned later in this chapter in connection with oligonucleotide crystals, especially since protein–DNA contacts in the Z$\alpha$ complex occur only with the zigzag phosphate backbone, which is not that dissimilar in Z-DNA (Fig. 23.3.3.3) and Z(WC)-DNA (Fig. 23.3.3.7). Nevertheless, it is encouraging to see a short segment of Z actually making contact with its protein in a presumably biologically relevant context.

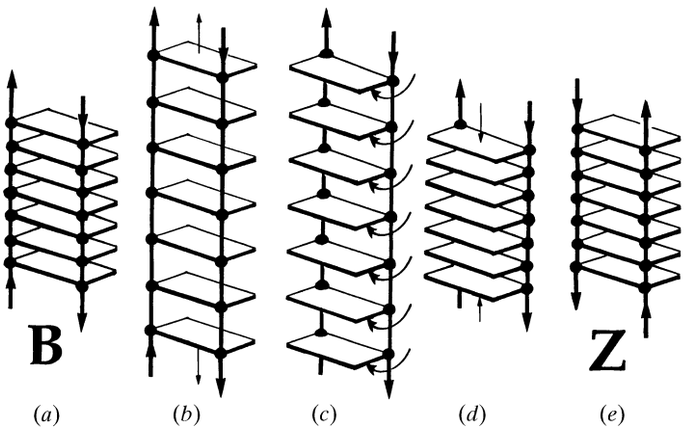


Fig. 23.3.3.6. Interconversion of a B to a Z helix. Because the strands have opposite directions in B (*a*) and Z (*e*), interconversion must involve opening up the helix (*b*), flipping each base pair to the other side (*c*), and re-stacking base pairs (*d*). (*d*) and (*e*) are identical upon rotation about a vertical axis.

pulled apart, as in part (*b*), and each base pair swung around to the opposite side of the backbone 'ladder' [part (*c*)]. This would automatically lead to *syn* conformations at both ends of the base pair, as drawn in Fig. 23.3.3.4(*b*). Returning pyrimidines to an *anti* conformation would create the zigzag backbone chain (Fig. 23.3.3.4*c*). Base pairs can then be re-stacked, as in parts (*d*) and (*e*) in Fig. 23.3.3.6 (which differ only by rotation of the entire helix about the vertical), to yield the backbone geometry of a Z helix. This is the simplest interconversion and one which was recognized and proposed in the very first Z-DNA structure paper (Z1). Other alternatives have been suggested, involving breaking individual base pairs, swinging the bases independently around their backbone chains, and re-forming the pairs. But one kind of special mechanism or another must be invoked if a B-to-Z interconversion is to be achieved.

### 23.3.3.7. *'Watson–Crick' Z-DNA*

Ansevin & Wang (1990) have proposed an alternative left-handed double helix, with many of the properties of Z-DNA, but possessing the same backbone chain orientations as A- and B-DNA. With such a helix, a B-to-Z conversion would require only a twisting of the duplex about its axis – no separation of bases or unpairing, and no pulling apart of the stack. Ansevin & Wang did not challenge the X-ray crystal structure analyses of short Z-DNA oligomers. Instead, they suggested that Z-DNA was globally the most stable form, adopted in short oligomers where chain unravelling and rearrangement is easy, but that their 'Watson–Crick' Z-DNA or Z(WC)-DNA was the structure that was actually produced by *in vitro* or *in vivo* manipulations of long DNA duplexes. They noted that most solution measurements focus on only two characteristics of the DNA: left-handedness and a dinucleotide repeat, both shared by Z-DNA and Z(WC)-DNA.

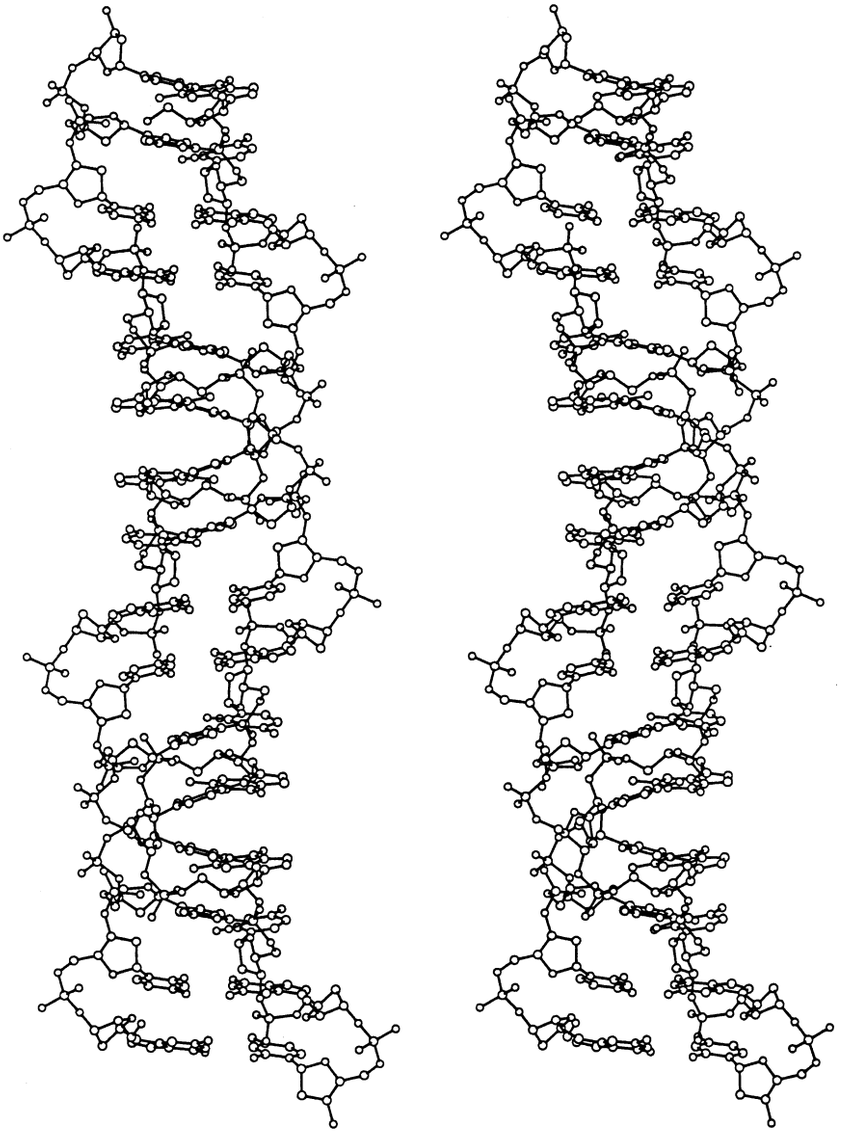

Fig. 23.3.3.7. Z(WC)-DNA, or 'Watson–Crick Z-DNA', a proposed left-handed, zigzag, alternating purine/pyrimidine helix with many of the properties of Z-DNA, but with the backbone chain sense found in A- and B-DNA (Ansevin & Wang, 1990). Coordinates courtesy of Allen T. Ansevin.

The Z(WC) helix is shown in Fig. 23.3.3.7, and a different stereo view appears as Fig. 7 of Dickerson (1992). Like Z-DNA, it is left-handed, with a deep minor groove and shallow major groove. Cytosines with *anti* glycosyl bonds and guanines with *syn* bonds alternate along each backbone strand. However, sugar puckering is reversed: cytosines are C3′-*endo*, while guanines are C2′-*endo*. In Z-DNA, the backbone chain runs parallel to the helix axis past G, and at right angles to the axis past C. In Z(WC)-DNA, this is reversed: parallel to the helix past C, and at right angles past G. Because of efficient stacking of base pairs, the logical two-base-pair structural unit in Z-DNA is $^{5'}$C–G$^{3'}$; in Z(WC)-DNA it is $^{5'}$G–C$^{3'}$. One such unit is clearly visible in the centre of Fig. 23.3.3.7. This behaviour is reflected in local twist angles:

| Helix | C–G | G–C | Sum |
|---|---|---|---|
| Z-DNA | −8° | −52° | −60° |
| Z(WC)-DNA | −70° | +10° | −60° |

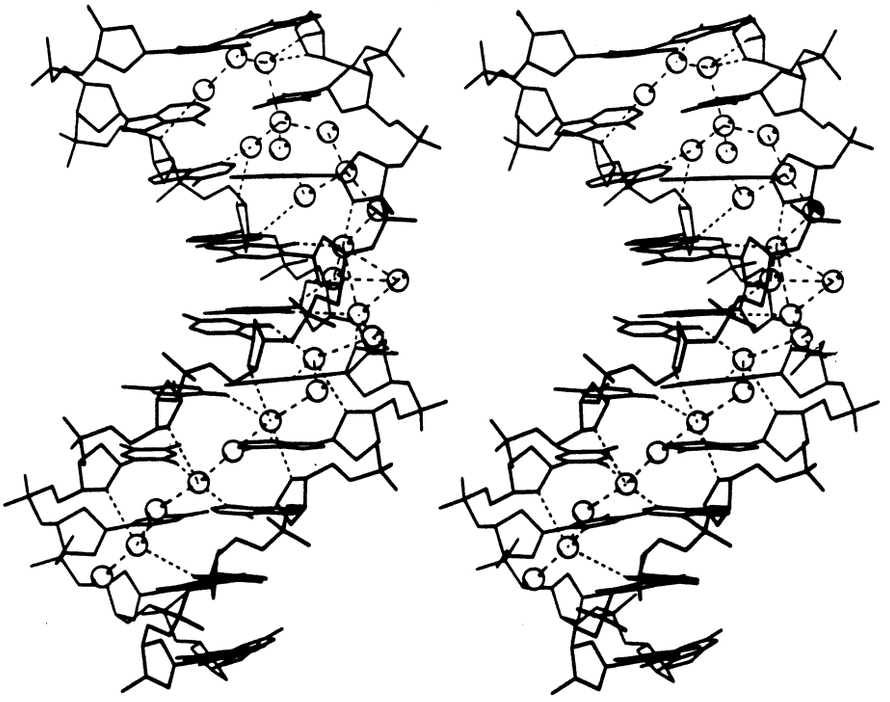

Fig. 23.3.4.1. Structure of C-A-A-A-G-A-A-A-A-G (B107). The lower half of the helix, with -A-A-A-A-G, exhibits the narrow minor groove commonly associated with the AT region of the helix and a single zigzag spine of hydration, as was first seen in C-G-C-G-A-A-T-T-C-G-C-G (B1–B6). The upper half, with C-A-A-A-G-, has the wider minor groove of general-sequence B-DNA and two separate rows of hydrating water molecules along the two walls of the wider groove.

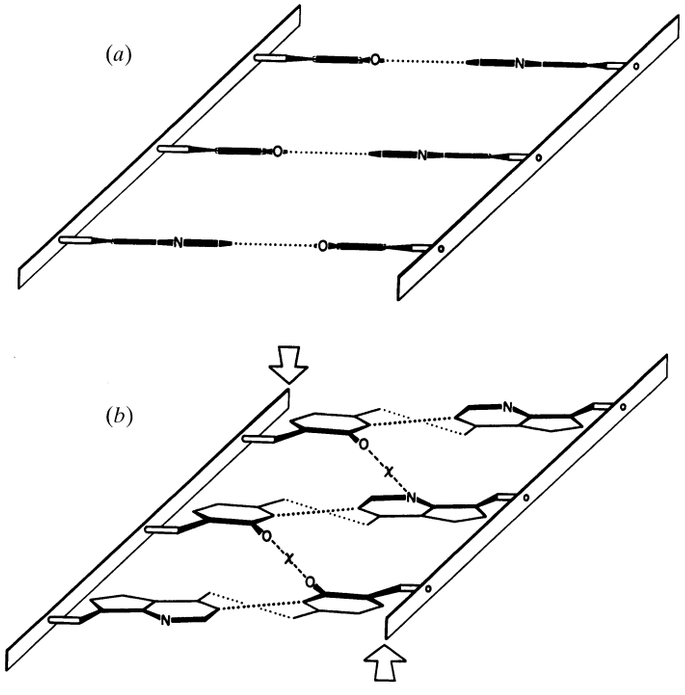


Fig. 23.3.4.2. Relationship between minor groove width and propeller twist. (*a*) View into the minor groove of B-DNA, with base pairs seen on edge and with the sugar–phosphate backbones shown schematically as inclined ladder uprights. (*b*) Consequences of propeller twisting the base pairs. Glycosyl bonds connected to sugar C1′ atoms are all displaced upward in the right strand and downward in the left strand. This shifts the backbone chains as indicated by the arrows. Hence, the gap between the chains is decreased, and the minor groove is narrowed.

The Ansevin–Wang helix has been sedulously ignored since its publication in 1990, especially by crystallographers. The Science Citation Index lists an average of *one* citation of their paper per year since publication, most commonly by spectroscopists. Ho & Mooers (1996) are almost alone among crystallographers in coupling the B-to-Z interconversion dilemma to the possible existence of a different kind of left-handed structure in long polynucleotides. Of course the Z(WC)-DNA structure, as presented here, is only a model; it could be far from the true structure in many respects. But its interest lies in the fact that a left-handed alternating helix with 'standard' backbone dirctions *can* be built with reasonable bond geometries and with properties that fit the various physical measurements as well as Z-DNA. It calls into question not the correctness of the Z-DNA structure obtained from short oligomers with free helix ends, but the relevance of that structure to the production of left-handed regions in longer duplexes with constrained ends.

### 23.3.4. Sequence–structure relationships in B-DNA

Two channels of information exist in B-DNA by which base sequence is expressed to the outside world. One of these is the Watson–Crick base pairing of A with T and G with C that is used in the storage of genetic information and in replication and transcription. The other channel, used in control and regulation of the expression of this genetic information, involves the hydrogen-bonding patterns of base-pair edges along the floors of the grooves and any systematic deformations of local helix structure that result explicitly from the base sequence.

The simplest and most direct expression of this second channel is the passive reading of hydrogen-bonding patterns along the floor of the major and minor grooves. This readout mechanism was first

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 $\alpha$-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*.

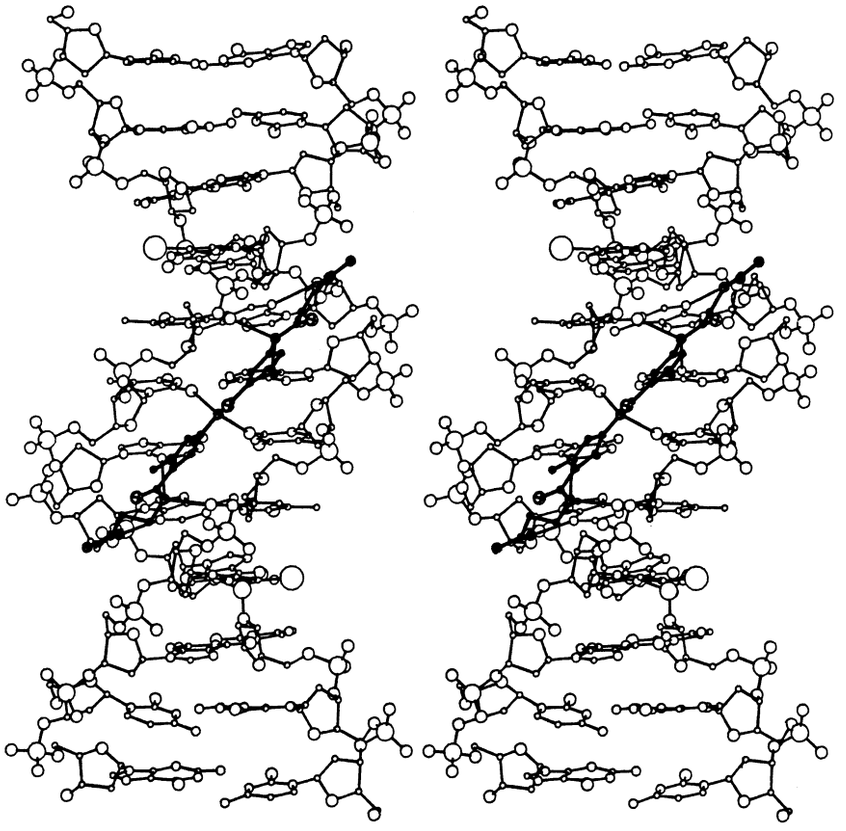

Fig. 23.3.4.3. Structure of the 1:1 complex of netropsin with C-G-C-G-A-A-T-T-C-G-C-G (B11, B12, B87). The drug binds to the central -A-A-T-T- region of the minor groove, which is barely wide enough to enclose the nearly planar polyamide molecule. The netropsin structure can be represented by

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

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, $\gamma\delta$-

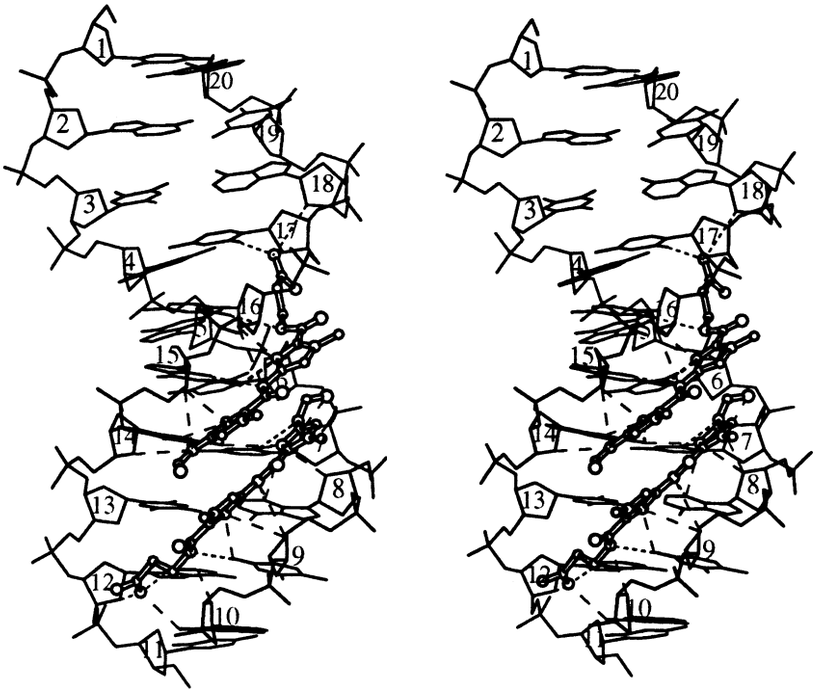


Fig. 23.3.4.4. Structure of the 2:1 complex of a di-imidazole lexitropsin with C-A-T-G-G-C-C-A-T-G (B108). The drug now is represented by

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

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.

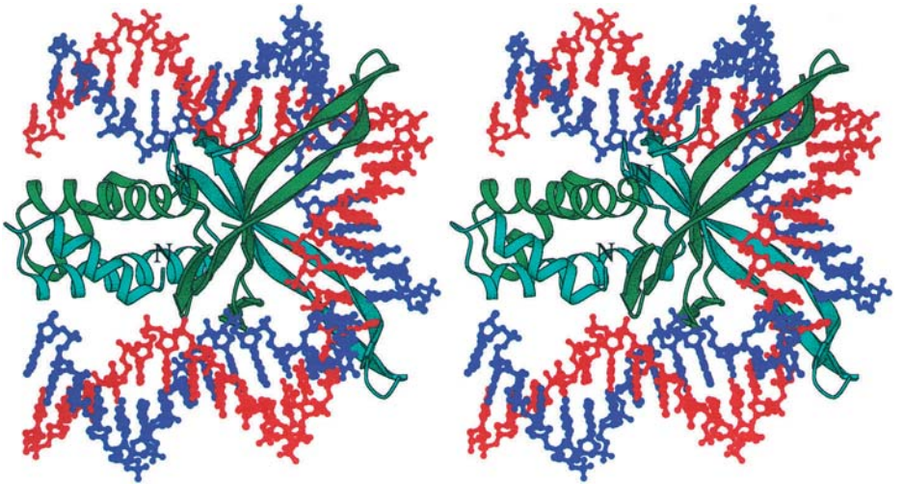

Fig. 23.3.4.5. DNA duplex (red and blue strands) looped around IHF or integration host factor. The two subunits of the IHF duplex are green and turquoise. Two antiparallel loops of protein chain, one from each subunit, insert into the minor groove of B-DNA at the sequence C-A-A-T/A-T-T-G and produce 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

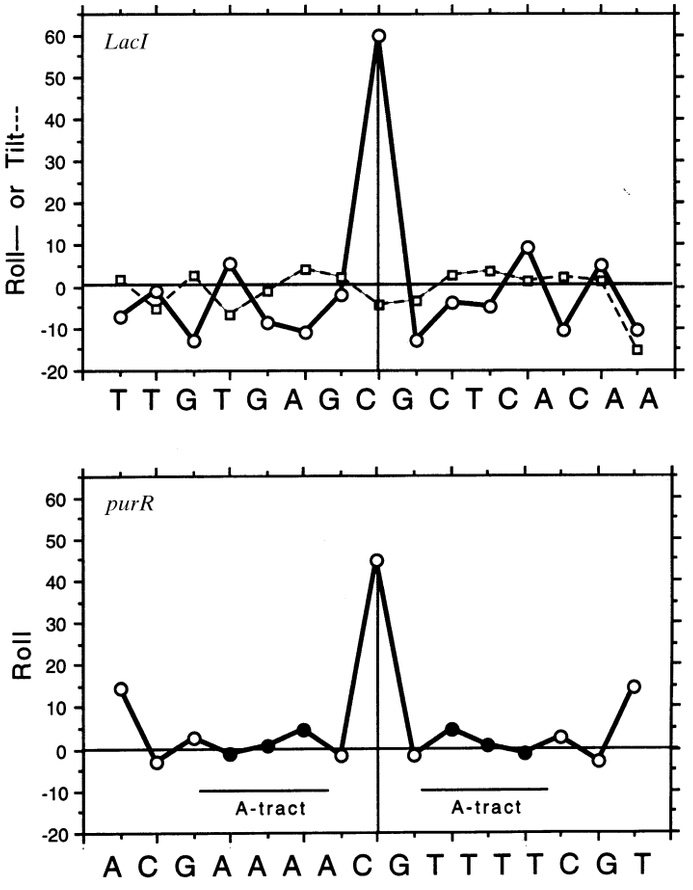


Fig. 23.3.4.6. Roll-angle plots for sequence-specific DNA–protein complexes with *lacI* (top) and *purR* (bottom). In each case, bending occurs *via* localized roll at a C-G step. Other steps of the sequence have random rolls of *ca* 10° or less. Note that, as with IHF, A-tracts are especially straight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

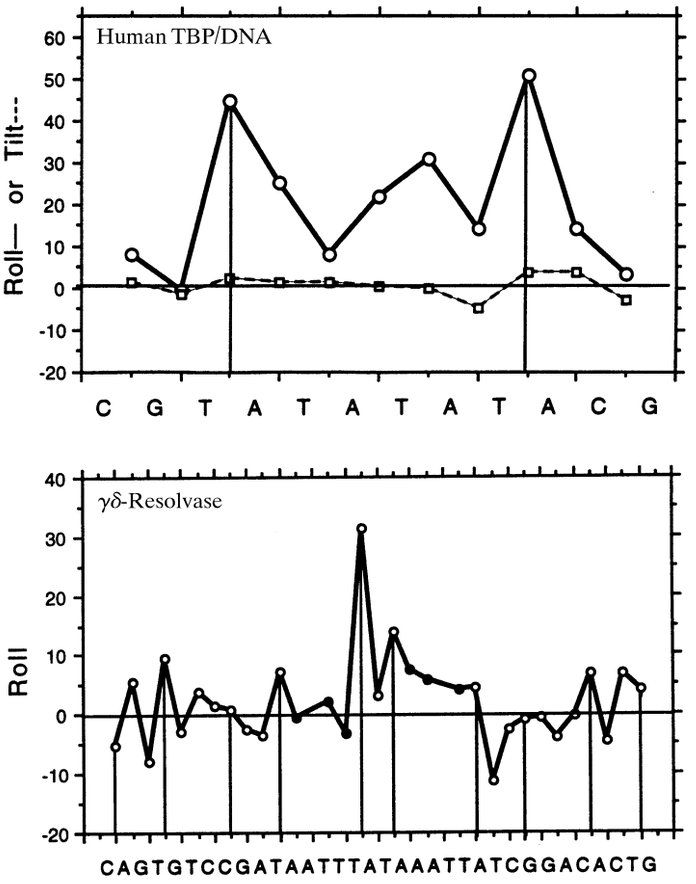


Fig. 23.3.4.7. Bending *via* roll at T-A steps in TBP or the TATA-binding protein (top) and in $\gamma\delta$-resolvase (bottom). Note that not every T-A step in TBP or $\gamma\delta$-resolvase is necessarily bent. Note also in $\gamma\delta$-resolvase that C-A = T-G steps, which in proteins such as 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.

### 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.

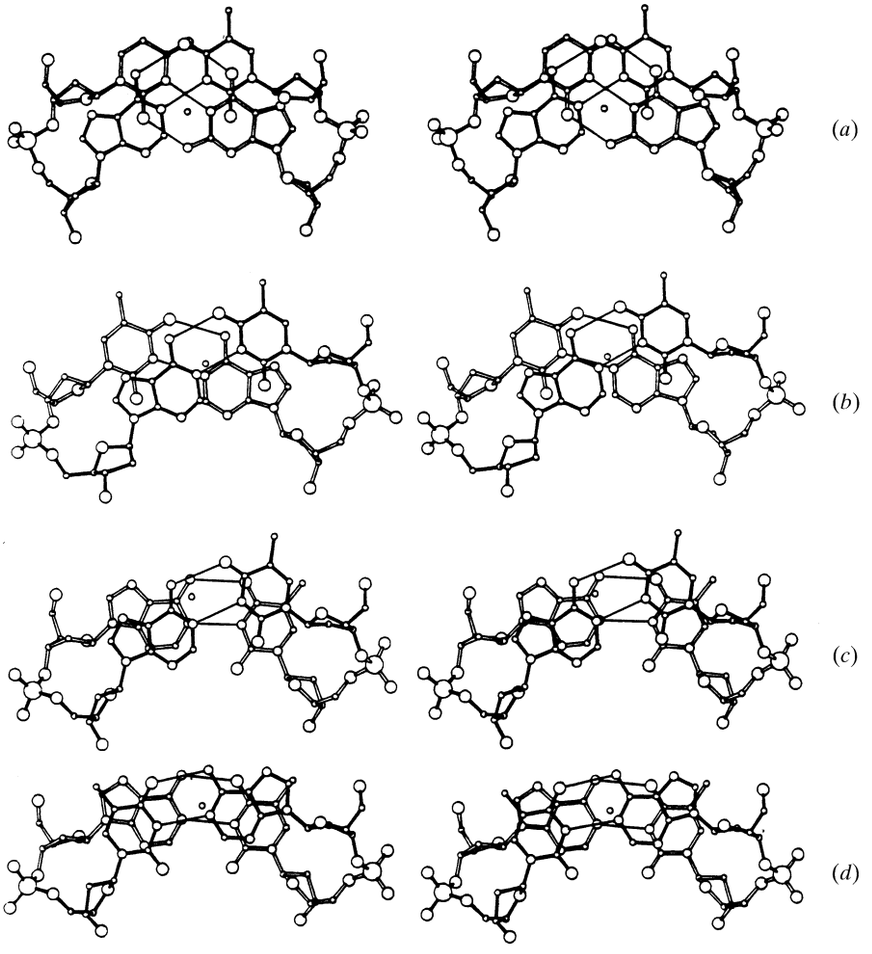


Fig. 23.3.4.8. Representative base-pair steps from B-DNA single-crystal X-ray analyses. (*a*) Pyrimidine-purine C-A step from C-C-A-A-G-A-T-T-G-G (B22, B46) (roll/slide/twist = $-7.4°/2.6$ Å$/49.9°$). Note the lack of ring-on-ring stacking, replaced by the stacking of pyrimidine O2 and purine N6 or O6, 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.

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-

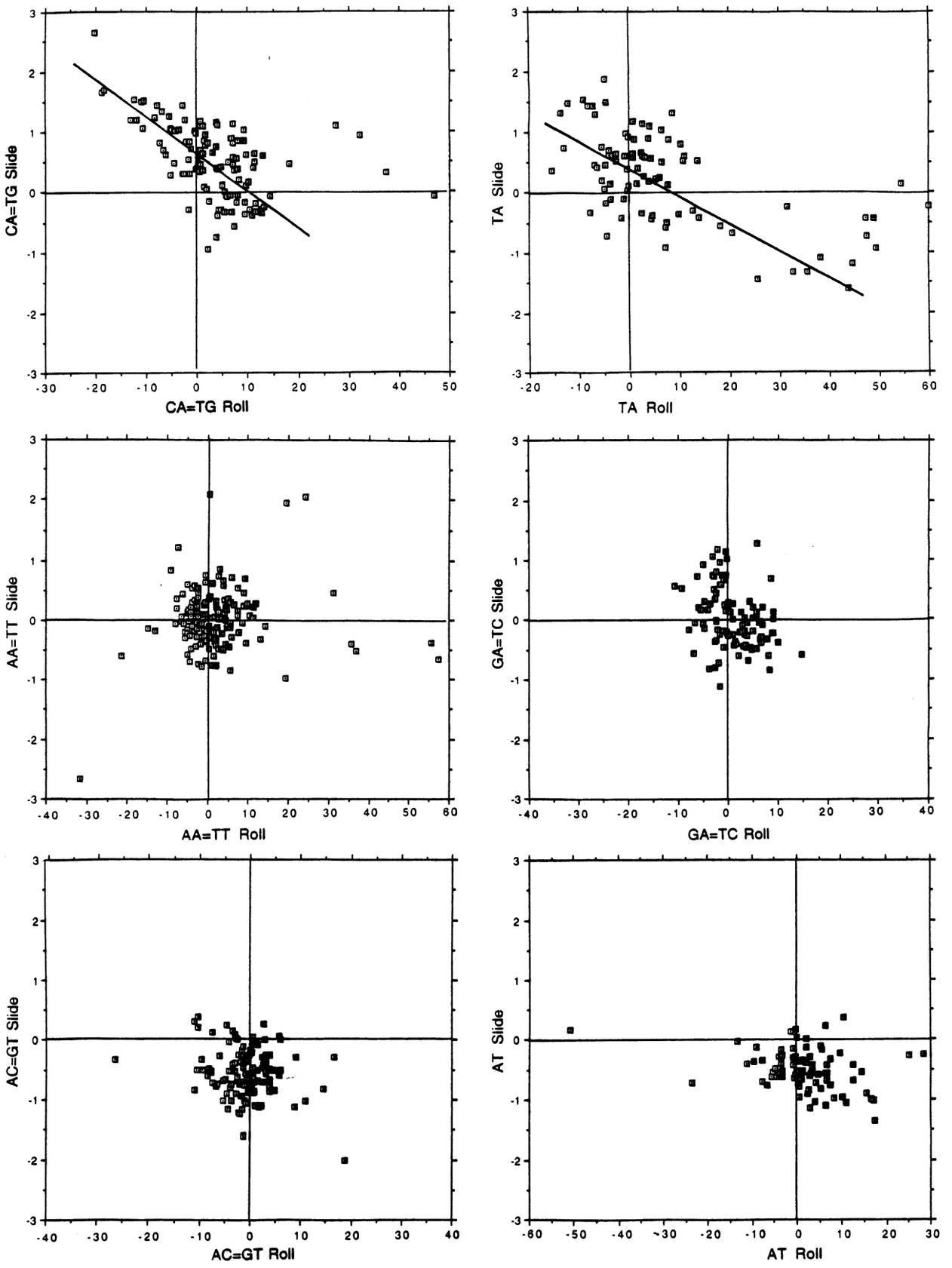


Fig. 23.3.4.9. Slide *versus* roll plots for six of the ten possible base-pair steps. Data points are from 971 steps in crystal structure analyses of 63 sequence-specific protein–DNA complexes. A complete set of 30 plots for slide/roll, twist/roll and slide/twist at all ten steps is to be found in Dickerson (1998*b*), and equivalent plots for DNA alone are given by El Hassan & Calladine (1997). Y-R steps exhibit a broad range of roll, slide and twist values, with roughly linear correlations between pairs of variables. Points for A-A and other R-R steps cluster tightly around the origin, showing little tendency toward roll bending. Curiously enough, points for R-Y steps tend to favour negative values of slide and twist, and, hence, to concentrate in the lower left quadrant of a slide/tilt plot.

TAAAACGTTTTAAAACCCCCC is much less curved because the roll kink at CG is counterbalanced by roll kinks in the opposite direction at the two flanking TA steps. In both cases, A-tracts are straight and completely unbent. (Note that both roll kinks can involve compression of the major groove, as expected, because the kink sites are a half turn of helix apart.)

This similarity of behaviour of DNA in crystals and in protein–DNA complexes should come as no surprise, since the local molecular environments – close intermolecular contacts, partial dehydration, low water activity, low local dielectric constant, high ionic strength, presence of divalent cations – are similar in these two cases and quite different from that of free DNA in dilute aqueous solution. Far from being unwanted 'crystal deformations', the local changes in structure resulting from intermolecular contacts in DNA crystals provide positive information about sequence-dependent deformability that is relevant to the protein recognition process. With regard specifically to A-tract behaviour, Occam's Razor would argue in favour of model (3) above for the behaviour of A-tracts in solution. The situation in dilute aqueous solution becomes of secondary importance if what is wanted is an understanding of A-tract B-DNA behaviour in protein–DNA complexes. Here, the answer is unambiguous: A-tracts in their biological setting are inherently rigid structural elements, chosen by natural selection when bending should be avoided.

### 23.3.5. Summary

Three families of nucleic acid double helix have been found – A, B and Z – with widely different structures and usages. The A and B helices are right-handed and have no limitations on base sequence. Z is left-handed and effectively limited to alternating purines and pyrimidines, with G and C overwhelmingly favoured. B is the biologically significant helix for DNA and is used in genetic coding. A is the helix of preference for RNA because it can accommodate the C2′-OH group of ribose, which produces steric clash in the B helix. The Z helix has, as yet, no well established biological function. A left-handed DNA configuration can be induced in longer DNA segments by negative supercoiling in solution, but it is not clear that this left-handed configuration is identical to the Z-DNA seen in short crystalline oligomers, because of the reversed orientation of backbone strands in Z-DNA.

B-DNA is an inherently malleable or deformable duplex. Its sugar ring conformations are much more variable than those of A-DNA. The base sequence of B-DNA is expressed directly *via* hydrogen bonds between bases of a pair, and indirectly *via* hydrogen-bond donors and acceptors along the floor of the major and minor groove. Sequence is also expressed as a *differential deformability* of different regions of the duplex. The two most obvious parameters affected by base sequence are minor groove width and helix bendability. Certain sequences of B-DNA are not statically bent, but are more bendable under stress than are other sequences. Bending occurs *via* roll, usually in the direction that compresses the broad major groove. Pyrimidine-purine or Y-R steps are most conducive to roll bending, and purine-purine steps are least bendable, particularly A-tracts of four or more AT base pairs without the weak T-A step. Natural selection has engineered Y-R steps into a DNA sequence where a sharp roll bend is wanted, and short A-tracts into a sequence where bending is not desired.

## Appendix 23.3.1.
## X-ray analyses of A, B and Z helices

Table A23.3.1.1. *X-ray analyses of A helices, DNA and RNA*

This table and the two that follow are intended as a historical background and a focus on the geometry of the intact double helix. References are current as of late 1997; sequences marked 'to be published' in 1997 that still are unpublished two years later have been deleted. Also omitted are sequences with fewer than four base pairs in the asymmetric unit, complexes with intercalating drugs, helices with bulges or looped-out bases, unusual structures such as quadruplexes, hammerhead ribozymes and tRNA. For information on these and for more recent results, consult the Nucleic Acid Database (NDB) at http://ndbserver.rutgers.edu/. An NDB number in parentheses indicates that the authors have never made coordinates available to the public. These structures are of little scientific value, but have been included for historical reasons.

Notes: Overhanging, unpaired bases are double underlined. Single underlining calls attention to mismatched bases or other interesting or relevant sequence aspects. $Z$ = number of asymmetric units per cell. Ubp = number of base pairs per asymmetric unit. NDB No. = Nucleic Acid Database serial number. Abbreviations: 2am = 2-amino; 5br = 5-bromo; 6ame = 6′-α-methyl; 4mo = 4-methoxy; 5me = 5-methyl; 6aOH = 6′-α-hydroxyl; 6mo = 6-methoxy; 8oxo = 8-oxo; 6et = 6-ethyl; ara = arabinosyl; ps = phosphorothioate; (P) = leading phosphate; A, T, G, C = DNA; a, u, g, c = RNA; Py = pyrrole; Im = imidazole.

(*a*) Dodecamers

| Sequence | Space group | $Z$ | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| CCCCCGCGGGGG | $P3_221$ | 6 | 12 | 1991, Barcelona | ADL025 | (A38) |
| CCGTACGTACGG | $P6_122$ | 12 | 6 | 1992, Ohio State | ADL045 | (A41) |
| GCGTACGTACGC | $P6_122$ | 12 | 6 | 1992, Ohio State | ADL046 | (A39) |

(*b*) Decamers

| Sequence | Space group | $Z$ | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| GCGGGCCCGC | $P6_122$ | 12 | 5 | 1993, Ohio State | ADJ051 | (A46) |
| GCACGCGTGC | $P6_122$ | 12 | 5 | 1996, Ohio State | ADJ075 | (A60) |
| ACCGGCCGGT | $P6_122$ | 12 | 5 | 1989, MIT | ADJ022 | (A26) |
| ACCGGCCGGT | $P6_122$ | 12 | 5 | 1995, MIT | ADJ065 | (A55) |
| ACCCGCGGGT | $P6_122$ | 12 | 5 | 1995, MIT | ADJ066 | (A55) |
| CCCGGCCGGG | $P2_12_12_1$ | 4 | 10 | 1993, Ohio State | ADJ049 | (A47) |
| CCIGGCC$^{5me}$CGG | $P2_12_12_1$ | 4 | 10 | 1995, Ohio State | ADJB61 | (A58) |

Table A23.3.1.1. *X-ray analyses of A helices, DNA and RNA* (*cont.*)

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| GCGGGCCCGC | $P2_12_12_1$ | 4 | 10 | 1993, Ohio State | ADJ050 | (A46) |
| ACCGGCCGGT | $P2_12_12_1$ | 4 | 10 | 1995, MIT | ADJ067 | (A55) |
| CCGGGCCGCG | $P2_12_12_1$ | 4 | 10 | 1997, Ohio State | ADJ081,2 | (A71) |
| $C^{5me}$CGGGCCCGG | $P2_12_12_1$ | 4 | 10 | 1997, Ohio State | ADJB87 | (A71) |
| CCGGG$^{5br}$CCCGG | $P2_12_12_1$ | 4 | 10 | 1997, Ohio State | ADJB80 | (A71) |
| CCGGGCC$^{5me}$CGG | $P2_12_12_1$ | 4 | 10 | 1997, Ohio State | ADJB84,5 | (A71) |
| $C^{5me}$CGGGCCCGG | $P6_1$ | 6 | 10 | 1997, Ohio State | ADJB86 | (A71) |
| CCGGGCC$^{5br}$CGG | $P6_1$ | 6 | 10 | 1997, Ohio State | ADJB79 | (A71) |
| CCGGGCC$^{5me}$CGG | $P6_1$ | 6 | 10 | 1997, Ohio State | ADJB83 | (A71) |

(*c*) Nonamers

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| GGATGGGAG | $P4_3$ | 4 | 9 | 1986, Cambridge | ADI009 | (A14) |

(*d*) Octamers, space group $P4_32_12$

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| CCCCGGGG | 8 | 4 | 1987, Weizmann/MIT | ADH012 | (A16) |
| CCCCGGGG, 298 K | 8 | 4 | 1995, Weizmann | ADH056 | (A54) |
| CCCGCGGG | 8 | 4 | 1997, Moscow | ADH0106 | (A69) |
| CCCTAGGG | 8 | 4 | 1996, Ohio State | ADH078 | (A64) |
| GCCCGGGC | 8 | 4 | 1987, Berlin | ADH008 | (A17) |
| GCCC*GGGC (*methylenephosphonate) | 8 | 4 | 1991, Berlin | ADHP36 | (A36) |
| GGCCGGCC | 8 | 4 | 1982, MIT | ADH013,098 | (A4,5) |
| GGCCGGCC, 288 K | 8 | 4 | 1995, Weizmann | ADH058 | (A54) |
| GG$^{5me}$CCGGCC | 8 | 4 | 1987, MIT | (ADHB21) | (A15) |
| GGGCGCCC, 293 K | 8 | 4 | 1988, Weizmann | ADH026 | (A22, A34) |
| GGGCGCCC, 115 K | 8 | 4 | 1988, Weizmann | ADH027 | (A20, A34) |
| GGGCGCCC, 115 K, re-refinement | 8 | 4 | 1995, Weizmann | ADH057 | (A54) |
| GTGCGCAC | 8 | 4 | 1992, Ohio State | ADH047 | (A40) |
| GTGTACAC/spermine | 8 | 4 | 1987, Wisconsin | ADH014 | (A18, A29) |
| CTCTAGAG | 8 | 4 | 1989, Cambridge | ADH020 | (A27) |
| GTACGTAC | 8 | 4 | 1990, Kansas | ADH024 | (A35) |
| GTACGTAC | 8 | 4 | 1990, Bordeaux | ADH023 | (A32) |
| GTCTAGAC | 8 | 4 | 1992, Manchester | ADH041 | (A42) |
| ATGCGCAT | 8 | 4 | 1990, Institute of Cancer Research | (ADH032) | (A31) |
| ATGCGCAT/spermine | 8 | 4 | 1990, Institute of Cancer Research | ADH033 | (A31) |
| ACGTACGT | 8 | 4 | 1996, Trinity, Dublin | ADH070 | (A66) |

(*e*) Octamers, space group $P2_12_12_1$

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| CCCGCGGG | 4 | 8 | 1997, Moscow | ADH0102–5 | (A69) |

(*f*) Octamers, space group $P6_1$

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| GGGGCCCC | 6 | 8 | 1985, Cambridge | ADH006 | (A11) |
| GGGATCCC | 6 | 8 | 1988, Berlin | ADH007 | (A21) |
| GGGCGCCC, 293 K | 6 | 8 | 1989, Weizmann | (ADH028) | (A30, A34) |
| GGGCGCCC, 100 K | 6 | 8 | 1989, Weizmann | ADH029 | (A30, A34) |
| GGGTACCC, 293 K | 6 | 8 | 1990, Weizmann | ADH030 | (A33) |
| GGGTACCC, 100 K | 6 | 8 | 1990, Weizmann | ADH031 | (A33) |
| GGGTGCCC | 6 | 8 | 1988, Weizmann | ADH016 | (A22) |
| GGTATACC | 6 | 8 | 1981, Weizmann/Cambridge | ADH010 | (A2, A7) |
| GG$^{5br}$UA$^{5br}$UACC | 6 | 8 | 1981, Weizmann/Cambridge | ADHB11 | (A2, A7, A13) |

Table A23.3.1.1. *X-ray analyses of A helices, DNA and RNA* (*cont.*)

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| GGCATGCC | 6 | 8 | 1997, Institute of Cancer Research | ADH076 | (A70) |
| GGIGCTCC | 6 | 8 | 1989, Cambridge | ADHB17 | (A24) |
| GGGGCTCC mismatch | 6 | 8 | 1985, Cambridge/Weizmann | ADH019 | (A9, A12) |
| GGGGTCCC mismatch | 6 | 8 | 1985, Cambridge/Weizmann | ADH018 | (A10) |
| GGGTGCCC mismatch | 6 | 8 | 1988, Weizmann | ADH016 | (A22) |

(*g*) Octamers, space group $P6_122$

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| GTGTACAC | 12 | 4 | 1989, Wisconsin | ADH034 | (A28) |
| GTGTACAC/spermine | 12 | 4 | 1993, Ohio State | ADH038 | (A48) |
| GTGTACAC/spermidine | 12 | 4 | 1993, Ohio State | ADH039 | (A48) |

(*h*) Octamers, space group $P2_12_12$

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| GTACGTAC | 4 | 8 | 1993, Bordeaux | ADH059 | (A44) |

(*i*) Hexamers, space group $C222_1$

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| GCCGGC | 8 | 6 | 1995, Oregon State | ADF073 | (A56) |
| $G^{5me}CG^{5me}CGC$ | 8 | 6 | 1995, Oregon State | ADFB62 | (A56) |
| $G^{5me}CCGGC$ | 8 | 6 | 1995, Oregon State | ADFB63 | (A56) |
| $G^{5me}CGCGC$ | 8 | 6 | 1995, Oregon State | ADFB72 | (A56) |

(*j*) Tetramers

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| $^{5i}$CCGG | $P4_32_12$ | 8 | 4 | 1981, UCLA (CIT) | ADDB01 | (A1, A3, A8) |

(*k*) RNA/DNA and RNA/RNA (lower case = RNA)

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| CCGGC g CCGG | $P2_12_12_1$ | 4 | 10 | 1994, Ohio State | AHJ052 | (A49) |
| c CGGCGCCGg | $P2_12_12_1$ | 4 | 10 | 1994, Ohio State | AHJ060 | (A50) |
| g CGTATACGC | $P2_12_12_1$ | 4 | 10 | 1993, MIT | AHJ043 | (A45) |
| GCGTaTACGC | $P2_12_12_1$ | 4 | 10 | 1993, MIT | AHJ044 | (A45) |
| $GCGT^{me}aTACGC$ | $P2_12_12_1$ | 4 | 10 | 1994, ETH Zürich | AHJS55 | (A53) |
| g c GTATACGC | $P2_12_12_1$ | 4 | 10 | 1995, MIT | AHJ068 | (A55) |
| g c g TATACGC | $P2_12_12_1$ | 4 | 10 | 1982, MIT | AHJ015 | (A4, A6) |
| g c g TATACCC\ \GGGTATACGC | $P2_12_12_1$ | 4 | 10 | 1992, MIT | AHJ040 | (A43) |
| u u c g g g c g c c\ \GGCGCCCGAA | $P4_322$ | 8 | 10 | 1996, Upjohn | UHJ055 | (A62) |
| c c c c g g g g | $P6_122$ | 12 | 4 | 1995, ETH Zürich | ARH063 | (A57) |
| c c c c g g g g | $R32$ | 18 | 8 | 1995, ETH Zürich | ARH064 | (A57) |
| c c c c g g g g | $R32$ | 18 | 8 | 1996, Northwestern | ARH074 | (A61) |
| g u a u a u a C | $R3$ | 9 | 8 | 1996, Ohio State | AHH071 | (A65) |
| g u a u g u a C | $R3$ | 9 | 8 | 1997, Ohio State | AHH077 | (A68) |
| g u g u g u a C | $R3$ | 9 | 8 | 1997, Ohio State | AHH089 | (A67) |
| g c u u c g g c $^{br}$U | $C2$ | 4 | 9 | 1994, Cambridge | AHIB53 | (A51) |
| (P)g g a c u u c g g u c c | $C2$ | 4 | 6 | 1991, Berkeley | ARL037 | (A37) |
| c g c g a a t t a g c g | $P2_1$ | 2 | 12 | 1994, Manchester | ARL048 | (A52) |
| u a a g g a g g u g a u | $P1$ | 1 | 24 | 1995, Berlin | ARL062 | (A59) |
| g g c g c u u g c g u c | $P1$ | 1 | 24 | 1996, Colorado | URL050 | (A63) |
| u u a u a u a u a u a u a a | $P2_12_12_1$ | 4 | 4 | 1988, Strasbourg | ARN035 | (A19, A25) |

Table A23.3.1.1. *X-ray analyses of A helices, DNA and RNA* (*cont.*)

References (numbered chronologically by year and alphabetically by first author within each year)


Year Reference

1981 (A1) R. E. Dickerson, H. R. Drew & B. N. Conner (1981). *Biomolecular stereodynamics*, Vol. 1, edited by R. H. Sarma, pp. 1–34. New York: Adenine Press.

(A2) Z. Shakked, D. Rabinovich, W. B. T. Cruse, E. Egert, O. Kennard, G. Sala, S. A. Salisbury & M. A. Viswamitra (1981). *Proc. R. Soc. London Ser. B*, **213**, 479–487.

1982 (A3) B. N. Conner, T. Takano, S. Tanaka, K. Itakura & R. E. Dickerson (1982). *Nature (London)*, **295**, 294–299.

(A4) S. Fujii, A. H.-J. Wang, J. van Boom & A. Rich (1982). *Nucleic Acids Res. Symp. Ser.* **11**, 109–112.

(A5) A. H.-J. Wang, S. Fujii, J. H. van Boom & A. Rich (1982). *Proc. Natl Acad. Sci. USA*, **79**, 3968–3972.

(A6) A. H.-J. Wang, S. Fujii, J. H. van Boom, G. A. van der Marel, S. A. A. van Boeckel & A. Rich (1982). *Nature (London)*, **299**, 601–604.

1983 (A7) Z. Shakked, D. Rabinovich, O. Kennard, W. B. T. Cruse, S. A. Salisbury & M. A. Viswamitra (1983). *J. Mol. Biol.* **166**, 183–201.

1984 (A8) B. N. Conner, C. Yoon, J. L. Dickerson & R. E. Dickerson (1984). *J. Mol. Biol.* **174**, 663–695.

1985 (A9) T. Brown, O. Kennard, G. Kneale & D. Rabinovich (1985). *Nature (London)*, **315**, 604–606.

(A10) G. Kneale, T. Brown, O. Kennard & D. Rabinovich (1985). *J. Mol. Biol.* **186**, 805–814.

(A11) M. McCall, T. Brown & O. Kennard (1985). *J. Mol. Biol.* **183**, 385–396.

1986 (A12) W. N. Hunter, G. Kneale, T. Brown, D. Rabinovich & O. Kennard (1986). *J. Mol. Biol.* **190**, 605–618.

(A13) O. Kennard, W. B. T. Cruse, J. Nachman, T. Prange, Z. Shakked & D. Rabinovich (1986). *J. Biomol. Struct. Dyn.* **3**, 623–647.

(A14) M. McCall, T. Brown, W. N. Hunter & O. Kennard (1986). *Nature (London)*, **322**, 661–664.

1987 (A15) C. A. Frederick, D. Saal, G. A. van der Marel, J. H. van Boom, A. H.-J. Wang & A. Rich (1987). *Biopolymers*, **26**, S145–S160.

(A16) T. E. Haran, Z. Shakked, A. H.-J. Wang & A. Rich (1987). *J. Biomol. Struct. Dyn.* **5**, 199–217.

(A17) U. Heinemann, H. Lauble, R. Frank & H. Bloeker (1987). *Nucleic Acids Res.* **15**, 9531–9550.

(A18) S. Jain, G. Zon & M. Sundaralingam (1987). *J. Mol. Biol.* **197**, 141–145.

1988 (A19) A. C. Dock-Bregeon, B. Chevrier, A. Podjarny, D. Moras, J. S. de Bear, G. R. Gough, P. T. Gilham & J. E. Johnson (1988). *Nature (London)*, **335**, 375–378.

(A20) M. Eisenstein, H. Hope, T. E. Haran, F. Frolow, Z. Shakked & D. Rabinovich (1988). *Acta Cryst.* B**44**, 625–628.

(A21) H. Lauble, R. Frank, H. Bloecker & U. Heinemann (1988). *Nucleic Acids Res.* **16**, 7799–7816.

(A22) D. Rabinovich, T. Haran, M. Eisenstein & Z. Shakked (1988). *J. Mol. Biol.* **200**, 151–161.

1989 (A23) C. A. Bingman, S. Jain, D. Jebaratnam & M. Sundaralingam (1989). Sixth Conversation in Biomolecular Stereodynamics, Albany, NY, Abstracts p. 28.

(A24) W. B. T. Cruse, J. Aymani, O. Kennard, T. Brown, A. G. C. Jack & G. A. Leonard (1989). *Nucleic Acids Res.* **17**, 55–72.

(A25) A. C. Dock-Bregeon, B. Chevrier, A. Podjarny, J. Johnson, J. S. de Bear, G. R. Gough, P. T. Gilham & D. Moras. (1989). *J. Mol. Biol.* **209**, 459–474.

(A26) C. A. Frederick, G. J. Quigley, M.-K. Teng, M. Coll, G. A. van der Marel, J. H. van Boom, A. Rich & A. H.-J. Wang (1989). *Eur. J. Biochem.* **181**, 295–307.

(A27) W. N. Hunter, B. L. D'Estaintot & O. Kennard (1989). *Biochemisty*, **28**, 2444–2451.

(A28) S. Jain & M. Sundaralingam (1989). *J. Biol. Chem.* **264**, 12780–12784.

(A29) S. Jain, G. Zon & M. Sundaralingam (1989). *Biochemistry*, **28**, 2360–2364.

(A30) Z. Shakked, G. Guerstein-Guzikevich, F. Frolow & D. Rabinovich (1989). *Nature (London)*, **342**, 456–460.

1990 (A31) G. R. Clark, D. G. Brown, M. R. Sanderson, T. Chwalinski, S. Neidle, J. M. Veal, R. L. Jones, W. D. Wilson, G. Zon, E. Garman & D. I. Stuart (1990). *Nucleic Acids Res.* **18**, 5521–5528.

(A32) C. Courseille, A. Dautant, M. Hospital, B. Langlois d'Estaintot, G. Precigoux, D. Molko & R. Teoule (1990). *Acta Cryst.* A**46**, FC9–FC12.

(A33) M. Eisenstein, F. Frolow, Z. Shakked & D. Rabinovich (1990). *Nucleic Acids Res.* **18**, 3185–3194.

(A34) Z. Shakked, G. Guerstein-Guzikevich, A. Zaytzev, M. Eisenstein, F. Frolow & D. Rabinovich (1990). In *Structure and methods*, Vol. 3. *DNA and RNA*, edited by R. H. Sarma & M. H. Sarma, pp. 55–72. Schenectady, NY: Adenine Press.

(A35) F. Takusagawa (1990). *J. Biomol. Struct. Dyn.* **7**, 795–809.

1991 (A36) U. Heinemann, L.-N. Rudolph, C. Alings, M. Morr, W. Heikens, R. Frank & H. Bloecker (1991). *Nucleic Acids Res.* **19**, 427–433.

(A37) S. R. Holbrook, C. Cheong, I. Tinoco Jr & S.-H. Kim (1991). *Nature (London)*, **353**, 579–581.

(A38) N. Verdaguer, J. Aymami, D. Fernandez-Forner, I. Fita, M. Coll, T. Huynh-Dinh, J. Igolen & J. A. Subirana (1991). *J. Mol. Biol.* **221**, 623–635.

1992 (A39) C. Bingman, S. Jain, G. Zon & M. Sundaralingam (1992). *Nucleic Acids Res.* **20**, 6637–6647.

(A40) C. A. Bingman, X. Li, G. Zon & M. Sundaralingam (1992). *Biochemistry*, **31**, 12803–12812.

(A41) C. A. Bingman, G. Zon & M. Sundaralingam (1992). *J. Mol. Biol.* **227**, 738–756.

(A42) A. R. Cervi, B. Langlois d'Estaintot & W. N. Hunter (1992). *Acta Cryst.* B**48**, 714–719.

(A43) M. Egli, N. Usman, S. Zhang & A. Rich (1992). *Proc. Natl Acad. Sci. USA*, **89**, 534–538.

1993 (A44) B. Langlois D'Estaintot, A. Dautant, C. Courseille & G. Precigoux (1993). *Eur. J. Biochem.* **213**, 673–682.

(A45) M. Egli, N. Usman & A. Rich (1993). *Biochemistry*, **32**, 3221–3237.

(A46) B. Ramakrishnan & M. Sundaralingam (1993). *Biochemistry*, **32**, 11458–11468.

(A47) B. Ramakrishnan & M. Sundaralingam (1993). *J. Mol. Biol.* **231**, 431–444.

(A48) N. Thota, X. H. Li, C. Bingman & M. Sundaralingam (1993). *Acta Cryst.* D**49**, 282–291.

Table A23.3.1.1. *X-ray analyses of A helices, DNA and RNA* (*cont.*)


Year Reference

1994 (A49) C. Ban, B. Ramakrishnan & M. Sundaralingam (1994). *J. Mol. Biol.* **236**, 275–285.
(A50) C. Ban, B. Ramakrishnan & M. Sundaralingam (1994). *Nucleic Acids Res.* **22**, 5466–5476.
(A51) W. Cruse, P. Saludjian, E. Biala, P. Strazewski, T. Prange & O. Kennard (1994). *Proc. Natl Acad. Sci. USA*, **91**, 4160–4164.
(A52) G. A. Leonard, K. E. McAuley-Hecht, S. Ebel, D. M. Lough, T. Brown & W. N. Hunter (1994). *Structure*, **2**, 483–494.
(A53) P. Lubini, W. Zuercher & M. Egli (1994). *Chem. Biol.* **1**, 39–45.

1995 (A54) M. Eisenstein & Z. Shakked (1995). *J. Mol. Biol.* **248**, 662–678.
(A55) Y.-G. Gao, H. Robinson, J. H. van Boom & A. H.-J. Wang (1995). *Biophys. J.* **69**, 559–568.
(A56) B. H. Mooers, G. P. Schroth, W. W. Baxter & P. S. Ho (1995). *J. Mol. Biol.* **249**, 772–784.
(A57) S. Portmann, N. Usman & M. Egli (1995). *Biochemistry*, **34**, 7569–7575.
(A58) B. Ramakrishnan & M. Sundaralingam (1995). *Biophys. J.* **69**, 553–558.
(A59) H. Schindelin, M. Zhang, R. Bald, J.-P. Fuerste, V. A. Erdmann & U. Heinemann (1995). *J. Mol. Biol.* **249**, 595–603.

1996 (A60) C. Ban & M. Sundaralingam (1996). *Biophys. J.* **71**, 1222–1227.
(A61) M. Egli, S. Portmann & N. Usman (1996). *Biochemistry*, **35**, 8489–8494.
(A62) N. C. Horton & B. C. Finzel (1996). *J. Mol. Biol.* **264**, 521–533.
(A63) S. E. Lietzke, C. L. Barnes & C. E. Kundrot (1996). *Structure*, **4**, 917–930.
(A64) D. B. Tippin & M. Sundaralingam (1996). *Acta Cryst.* D**52**, 997–1003.
(A65) M. C. Wahl, C. Ban, C. Sekharudu, B. Ramakrishnan & M. Sundaralingam (1996). *Acta Cryst.* D**52**, 655–667.
(A66) D. J. Wilcock, A. Adams, C. J. Cardin & L. P. G. Wakelin (1996). *Acta Cryst.* D**52**, 481–485.

1997 (A67) R. Biswas & M. Sundaralingam (1997). *J. Mol. Biol.* **270**, 511–519.
(A68) R. Biswas, M. C. Wahl, C. Ban & M. Sundaralingam (1997). *J. Mol. Biol.* **267**, 1149–1156.
(A69) L. G. Fernandez, J. A. Subirana, N. Verdagauer, D. Pyshni, L. Campos & L. Malinina (1997). *J. Biomol. Struct. Dyn.* **15**, 151–163.
(A70) C. M. Nunn & S. Neidle (1997). *Acta Cryst.* D**53**, 269–273.
(A71) D. B. Tippin & M. Sundaralingam (1997). *J. Mol. Biol.* **267**, 1171–1185.


Table A23.3.1.2. *X-ray analyses of B-DNA helices and their complexes with minor-groove-binding drug molecules*

See introductory notes to Table A23.3.1.1. Space group $P2_12_12_1$ unless specified otherwise.

Notes: (triplet) = external triplet formed from overhanging bases. Overhanging, unpaired bases are double underlined. Single underlining calls attention to interesting or relevant sequence aspects. Other notes as in Table A23.3.1.1.

I. DNA duplexes without bound drugs

(*a*) Dodecamers, space group $P2_12_12_1$

(1) Oligonucleotides without mismatches

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| CGCGAATTCGCG, 290 K | 4 | 12 | 1980, UCLA (CIT) | BDL001 | (B1–5, B75) |
| CGCGAATTCGCG, 16 K | 4 | 12 | 1982, UCLA (CIT) | BDL002 | (B6) |
| CGCGAATTCGCG, re-refinement | 4 | 12 | 1987, Strasbourg | BDL020 | (B23) |
| CGCGAATTCGCG, anisotropic temperature-factor refinement | 4 | 12 | 1985, Berkeley | BDL005 | (B10) |
| CGCGAATT$^{5br}$CGCG, 293 K | 4 | 12 | 1982, UCLA (CIT) | BDLB03 | (B7, B8) |
| CGCGAATT$^{5br}$CGCG, 280 K | 4 | 12 | 1982, UCLA (CIT) | BDLB04 | (B7, B8, B75) |
| CGCGA$^{6me}$ATTCGCG | 4 | 12 | 1988, MIT | BDLB13 | (B24) |
| CGCGAA$^{6ame}$T$^{6ame}$TCGCG | 4 | 12 | 1997, Northwestern | BDLS79 | (B111) |
| CGCGAA$^{6aOH}$T$^{6aOH}$TCGCG | 4 | 12 | 1997, Northwestern | BDLS80 | (B111) |
| CGCGAASSCGCG | 4 | 12 | 1996, Manchester | BDLS67 | (B97) |
| CGCAIAT$^{5me}$CTGCG | 4 | 12 | 1997, Weizmann | BDLB82 | (B113) |
| CGCAAAAAAGCG | 4 | 12 | 1987, Cambridge | BDL006 | (B20, B75) |
| CGCAAAAATGCG | 4 | 12 | 1989, Yale | BDL015 | (B31, B75) |
| CGCAAATTTGCG | 4 | 12 | 1987, MIT | BDL016 | (B17) |
| CGCAAATTTGCG | 4 | 12 | 1992, Institute of Cancer Research | BDL038 | (B52, B75) |
| CGCATATATGCG | 4 | 12 | 1988, UCLA | BDL007 | (B27) |
| CGCGTTAACGCG | 4 | 12 | 1991, Ohio State | BDL059 | (B40, B86) |
| CGCGATATCGCG | 4 | 12 | 1997, Weizmann | BDL078 | (B113) |
| CGCAIAT$^{5me}$CTGCG | 4 | 12 | 1997, Weizmann | BDLB76 | (B113) |
| CGTGAATTCACG | 4 | 12 | 1991, UCLA | BDL029 | (B44, B75) |
| CGTGAATTCACG | 4 | 12 | 1991, Rutgers | BDL028 | (B45) |

Table A23.3.1.2. *X-ray analyses of B-DNA helices and their complexes with minor-groove-binding drug molecules* (*cont.*)

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| CGCGAAAACGCG/ CGCGTT/TTCGCG (nicked strand) | 4 | 12 | 1990, MIT | BDL021,32 | (B35) |

(2) Mismatch oligonucleotides (mismatches underlined)

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| CGCGAATTGGCG | 4 | 12 | 1993, Institute of Cancer Research | BDL046 | (B72) |
| CGCGAATTAGCG | 4 | 12 | 1986, Cambridge | BDL012 | (B13, B15) |
| CGCGAATT$^{6et}$AGCG | 4 | 12 | 1994, Manchester | BDLB54 | (B79) |
| CGCGAATT$^{8oxo}$AGCG | 4 | 12 | 1992, Manchester | BDLB33 | (B57) |
| CGCGAATTTGCG | 4 | 12 | 1985, Cambridge | BDL009 | (B19) |
| CGC$^{6me}$GAATTTGCG | 4 | 12 | 1990, Edinburgh | BDLB26 | (B38) |
| CGCAATTGGCG | 4 | 12 | 1989, Manchester | BDL014 | (B28, B37) |
| CGCAAGCTGGCG | 4 | 12 | 1990, Institute of Cancer Research | BDL022 | (B39, B75) |
| CGCAAATT$^{8oxo}$GGCG | 4 | 12 | 1994, Edinburgh | BDLB56 | (B80) |
| CGCAAATTCGCG | 4 | 12 | 1986, Cambridge | BDL011 | (B16) |
| CGCAAATTIGCG | 4 | 12 | 1992, Edinburgh | BDLB41 | (B56) |
| CGCIAATTAGCG | 4 | 12 | 1987, Cambridge | BDLB10 | (B18) |
| CGCIAATTCGCG | 4 | 12 | 1992, Thomas Jefferson | BDLB40 | (B61) |
| CGAGAATTC$^{6me}$GCG | 4 | 12 | 1994, Rutgers | BDLB53 | (B76) |
| CGTGAATTC$^{6me}$GCG | 4 | 12 | 1995, Rutgers | BDLB58 | (B95) |

(*b*) Dodecamers: other space groups

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| CGCTCTAGAGCG | $P2_1$ | 2 | 24 | 1996, Barcelona | BDL070 | (B102) |
| CGTAGATCTACG | *C*2 | 4 | 12 | 1993, Manchester | BDL042 | (B69, B75) |
| CGCGAAAAAACG | $P2_12_12$ | 4 | 24 | 1993, Yale | BDL047 | (B64, B75) |
| ACCGGCGCCACA | *R*3 | 9 | 12 | 1989, Strasbourg | BDL018 | (B34, B48, B49) |
| ACCGCCGGCGCC | *R*3 | 9 | 12 | 1989, Strasbourg | BDL035 | (B48, B49) |
| ACCGC$^{5me}$CGGCGCC | *R*3 | 9 | 12 | 1997, Strasbourg | BDLB83 | (B109) |
| ACCGGCGCCACA | *R*3 | 9 | 12 | 1991, Strasbourg | BDL034 | (B48) |

(*c*) Decamers

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| CCAAGATTGG mismatch | *C*2 | 4 | 5 | 1987, UCLA | BDJ008 | (B22, B25) |
| CCAACGTTGG, Mg | *C*2 | 4 | 5 | 1991, UCLA | BDJ019 | (B46, B50) |
| CCAACITTGG, Ca | *C*2 | 4 | 5 | 1992, UCLA | BDJB44 | (B70) |
| CCAGGCCTGG | *C*2 | 4 | 5 | 1989, Berlin | BDJ017 | (B32) |
| CCAGGC$^{ara}$CTGG | *C*2 | 4 | 5 | 1991, MIT | BDJS30 | (B41) |
| CCA$^{8oxo}$GCGCTGG | *C*2 | 4 | 5 | 1995, MIT | BDJB57 | (B91) |
| CTCTCGAGAG | *C*2 | 4 | 10 | 1994, UCLA | BDJ060 | (B89) |
| CGCAATTGCG | *C*2 | 4 | 10 | 1997, Institute of Cancer Research | BDJ069 | (B114) |
| CAAAGAAAAG | *C*2 | 4 | 20 | 1997, UCLA | BDJ081 | (B107) |
| CGACGATCGT TGCTAGCAGC | $P2_1$ | 2 | 10 | 1997, NYU | UDJ060 | (B112) |
| GGCCAATTGG GGTTAACCGG | $P2_12_12_1$ | 4 | 10 | 1996, Cambridge | UDJ049 | (B103) |
| CGATCGATCG, Mg | $P2_12_12_1$ | 4 | 10 | 1991, UCLA | BDJ025 | (B42) |
| CGATTAATCG, Mg | $P2_12_12_1$ | 4 | 10 | 1992, UCLA | BDJ031 | (B58) |
| CGATATATCG, Mg | $P2_12_12_1$ | 4 | 10 | 1992, UCLA | BDJ037 | (B62) |
| CGATATATCG, Ca | $P2_12_12_1$ | 4 | 10 | 1992, UCLA | BDJ036 | (B62) |
| CATGGCCATG, Ca | $P2_12_12_1$ | 4 | 10 | 1993, UCLA | BDJ051 | (B66) |
| CGATCG$^{6me}$ATCG | $P3_221$ | 6 | 10 | 1992, UCLA | BDJB48 | (B63) |
| CCAACITTGG, Mg | $P3_221$ | 6 | 10 | 1992, UCLA | BDJB43 | (B70) |
| CCATTAATGG, Mg | $P3_221$ | 6 | 10 | 1994, UCLA | BDJ055 | (B77) |
| CCACTAGTGG | $P3_221$ | 6 | 10 | 1994, Weizmann | BDJ061 | (B82) |
| CCAGGC$^{5me}$CTGG | *P*6 | 6 | 10 | 1992, Berlin | BDJB27 | (B43, B54) |

Table A23.3.1.2. *X-ray analyses of B-DNA helices and their complexes with minor-groove-binding drug molecules* (*cont.*)

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| CCAGGC$^{5me}$CTGG | *P*6 | 6 | 10 | 1993, Berlin | BDJB49 | (B68) |
| CCAGGC$^{5me}$CTGG | *P*6 | 6 | 10 | 1993, Berlin | BDJB50 | (B68) |
| CCAAGCTTGG | *P*6 | 6 | 10 | 1993, UCLA | BDJ052 | (B67) |
| CCGGCGCCGG | *R*3 | 9 | 10 | 1992, Berlin | BDJ039 | (B55) |
| CCGCCGGCGG | *R*3 | 9 | 10 | 1994, Strasbourg | BD0015 | (B85) |
| CCIIICCCGG | $P3_1$ | 3 | 10 | 1997, Weizmann | BDJB77 | (B113) |

(*d*) Other oligonucleotide lengths

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| GCGAATTCG (triplet)<br>GCTTAAGCG | $P2_12_12_1$ | 4 | 8 | 1996, Cambridge | UDI030 | (B94) |
| CGCTAGCG | $P2_12_12_1$ | 4 | 16 | 1996, Barcelona | BDH071 | (B102) |
| CGGTGG<br>CCACCG | $P6_122$ | 12 | 6 | 1995, Manitoba | BDF062 | (B93) |
| CTCGAG | $P6_222$ | 12 | 3 | 1996, Ohio State | BDF068 | (B104) |
| GpsCGpsCGpsC | $P2_12_12_1$ | 4 | 6 | 1987, Cambridge | BDFP24 | (B14) |

II. DNA complexes with minor-groove-binding drugs

(*a*) Netropsin family of polyamides

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| Netropsin: $^{+}$Py-Py$^{+}$ | | | | | | |
| CGCGAATT$^{5br}$CGCG/N | | 4 | 12 | 1985, UCLA | GDLB05 | (B11, B12) |
| CGCGAATT$^{5br}$CGCG/N | | 4 | 12 | 1995, UCLA | GDLB31 | (B88) |
| CGCGAATTCGCG/N | | 4 | 12 | 1992, Illinois | GDL018 | (B59) |
| CGC$^{6et}$GAATTCGCG/N | | 4 | 12 | 1992, Illinois | GDLB17 | (B59) |
| CGCAAATTTGCG/N | | 4 | 12 | 1993, MIT | GDL014 | (B73) |
| CGCGATATCGCG/N | | 4 | 12 | 1989, MIT | GDL001,4 | (B30) |
| CGCGTTAACGCG/N | | 4 | 12 | 1995, Ohio State | GDL030 | (B86) |
| CGCAATTGCG/N | | 4 | 12 | 1997, Institute of Cancer Research | GDJ046 | (B110) |
| Lexitropsin: $^{+}$Im-Py$^{+}$ | | | | | | |
| CGCGAATTCGCG/1L | | 4 | 12 | 1995, UCLA | GDL037,8 | (B90) |
| 2:1 Di-imidazole Lexitropsin: $^{0}$Im-Im$^{+}$ | | | | | | |
| CATGGCCATG/2D | | 4 | 10 | 1997, UCLA | GDJ054 | (B108) |
| Distamycin: $^{0}$Py-Py-Py$^{+}$ | | | | | | |
| CGCAAATTTGCG/1D | | 4 | 12 | 1987, MIT | GDL003 | (B17) |
| ICICICIC/2D | $P4_122$ | 8 | 4 | 1994, Ohio State | GDHB25 | (B74) |
| IcICICIC/2D | $P4_122$ | 8 | 4 | 1995, Ohio State | GHHB34 | (B87) |
| IcIcICIC/2D | $P4_122$ | 8 | 4 | 1995, Ohio State | GHHB35 | (B87) |
| ICATATIC | $P4_122$ | 8 | 4 | 1997, Ohio State | GHHB50 | (B105) |
| ICITACIC | $P4_122$ | 8 | 4 | 1997, Ohio State | GHHB51 | (B105) |
| ICATATIC | *C*2 | 4 | 4 | 1997, Ohio State | GDLB49 | (B105) |

(*b*) Hoechst family

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| Hoechst 33258 (*para* -OH on phenyl ring A) | | | | | |
| CGCGAATTCGCG/H | 4 | 12 | 1987, UCLA | GDL006 | (B21) |
| CGCGAATTCGCG/H | 4 | 12 | 1988, MIT | GDL002 | (B26) |
| CGCGAATTCGCG/H, 273 K | 4 | 12 | 1991, UCLA | GDL010,11 | (B47) |
| CGCGAATTCGCG/H, 248 K | 4 | 12 | 1991, UCLA | GDL012 | (B47) |
| CGCGAATTCGCG/H, 173 K | 4 | 12 | 1991, UCLA | GDL013 | (B47) |
| CGCGATATCGCG/H | 4 | 12 | 1989, MIT | GDL007 | (B29) |

Table A23.3.1.2. *X-ray analyses of B-DNA helices and their complexes with minor-groove-binding drug molecules* (*cont.*)

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| CGCAAATTTGCG/H | 4 | 12 | 1994, Institute of Cancer Research | GDL028 | (B83) |
| CGCAAATTTGCG/H | 4 | 12 | 1994, MIT | GDL026 | (B84) |
| CGCGAATTCGCG/H | 4 | 12 | 1992, Illinois | GDL022 | (B60) |
| CGC$^{6et}$GAATTCGCG/H | 4 | 12 | 1992, Illinois | GDLB19 | (B60) |
| Meta-OH(N) Hoechst 33258 (*meta* -OH on ring A) | | | | | |
| CGCGAATTCGCG/H 'in' | 4 | 12 | 1996, Institute of Cancer Research | GDL047 | (B99) |
| CGCGAATTCGCG/H 'out' | 4 | 12 | 1996, Institute of Cancer Research | GDL048 | (B99) |
| Hoechst 33342 (*para* -OEt on ring A) | | | | | |
| CGCGAATTCGCG/H | 4 | 12 | 1992, Illinois | GDLB20 | (B60) |
| CGC$^{6et}$GAATTCGCG/H | 4 | 12 | 1992, Illinois | GDLB20 | (B60) |
| Bis-benzimidazole compound (imidazole for piperazine on Hoechst 33258) | | | | | |
| CGCGAATTCGCG/B | 4 | 12 | 1995, Institute of Cancer Research | GDL033 | (B96) |
| Tribiz or Tris-benzimidazole (extended Hoechst 33258 analogue) | | | | | |
| CGCAAATTTGCG/T | 4 | 12 | 1996, Institute of Cancer Research | GDL039 | (B98) |
| Bis-amidinium derivative of Hoechst 33258 | | | | | |
| CGCGAATTCGCG | 4 | 12 | 1997, Institute of Cancer Research | GDL052 | (B106) |

(*c*) Berenil family

| Sequence | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|
| Berenil | | | | | |
| CGCGAATTCGCG/B | 4 | 12 | 1990, Institute of Cancer Research | GDL009 | (B36) |
| CGCGAATTCGCG/B | 4 | 12 | 1992, Institute of Cancer Research | GDL016 | (B51) |
| 2,5-Bis(4-guanylphenyl)furan (berenil analogue) | | | | | |
| CGCGAATTCGCG/F | 4 | 12 | 1996, Institute of Cancer Research | GDL036 | (B100) |
| 2,5-Bis{[4-(*N*-isopropyl)amidino]phenyl}furan (berenil analogue) | | | | | |
| CGCGAATTCGCG/F | 4 | 12 | 1996, Institute of Cancer Research | GDL044 | (B101) |
| 2,5-Bis{[4-(*N*-cyclopropyl)amidino]phenyl}furan (berenil analogue) | | | | | |
| CGCGAATTCGCG/F | 4 | 12 | 1997, Institute of Cancer Research | GDL045 | (B101) |

(*d*) Other minor-groove binders

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| DAPI | | | | | | |
| CGCGAATTCGCG/D | | 4 | 12 | 1989, UCLA | GDL008 | (B33) |
| Pentamidine | | | | | | |
| CGCGAATTCGCG/P | | 4 | 12 | 1992, Institute of Cancer Research | GDL015 | (B53) |
| $\gamma$-Oxapentamidine | | | | | | |
| CGCGAATTCGCG/P | | 4 | 12 | 1994, Institute of Cancer Research | GDL027 | (B81) |
| Propamidine | | | | | | |
| CGCGAATTCGCG/P | | 4 | 12 | 1993, Institute of Cancer Research | GDL023 | (B71) |
| CGCGAATTCGCG/P | | 4 | 12 | 1995, Institute of Cancer Research | GDL032 | (B92) |

Table A23.3.1.2. *X-ray analyses of B-DNA helices and their complexes with minor-groove-binding drug molecules* (*cont.*)

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| SN6999<br>CGC$^{6et}$GAATTCGCG/S | | 4 | 12 | 1993, Illinois | GDLB24 | (B65) |
| Anthramycin<br>CCAACGTTGG/A | $P3_221$ | 6 | 5 | 1993, UCLA | GDJB29 | (B78) |

References (numbered chronologically by year and alphabetically by first author within each year)


Year Reference

1980 (B1) R. M. Wing, H. R. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura & R. E. Dickerson (1980). *Nature (London)*, **287**, 755–758.

1981 (B2) R. E. Dickerson & H. R. Drew (1981). *J. Mol. Biol.* **149**, 761–786.

(B3) R. E. Dickerson, H. R. Drew & B. N. Conner (1981). *Biomolecular stereodynamics*, Vol. 1, edited by R. H. Sarma, pp. 1–34. New York: Adenine Press.

(B4) H. R. Drew & R. E. Dickerson (1981). *J. Mol. Biol.* **151**, 535–556.

(B5) H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura & R. E. Dickerson (1981). *Proc. Natl Acad. Sci. USA*, **78**, 2179–2183.

1982 (B6) H. R. Drew, S. Samson & R. E. Dickerson (1982). *Proc. Natl Acad. Sci. USA*, **79**, 4040–4044.

(B7) A. V. Fratini, M. L. Kopka, H. R. Drew & R. E. Dickerson (1982). *J. Biol. Chem.* **257**, 14686–14707.

1983 (B8) M. L. Kopka, A. V. Fratini, H. R. Drew & R. E. Dickerson (1983). *J. Mol. Biol.* **163**, 129–146.

1984 (B9) R. M. Wing, P. Pjura, H. R. Drew & R. E. Dickerson (1984). *EMBO J.* **3**, 1201–1206.

1985 (B10) S. R. Holbrook, R. E. Dickerson & S.-H. Kim (1985). *Acta Cryst.* B**41**, 255–262.

(B11) M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura & R. E. Dickerson (1985). *Proc. Natl Acad. Sci. USA*, **82**, 1376–1380.

(B12) M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura & R. Dickerson (1985). *J. Mol. Biol.* **183**, 553–563.

1986 (B13) T. Brown, W. N. Hunter, G. Kneale & O. Kennard (1986). *Proc. Natl Acad. Sci. USA*, **83**, 2402–2406.

(B14) W. B. T. Cruse, S. A. Salisbury, T. Brown, R. Cosstick, F. Eckstein & O. Kennard (1986). *J. Mol. Biol.* **192**, 891–905.

(B15) W. N. Hunter, T. Brown & O. Kennard (1986). *J. Biomol. Struct. Dyn.* **4**, 173–191.

(B16) W. N. Hunter, T. Brown, N. N. Anand & O. Kennard (1986). *Nature (London)*, **320**, 552–555.

1987 (B17) M. Coll, C. A. Frederick, A. H.-J. Wang & A. Rich (1987). *Proc. Natl Acad. Sci. USA*, **84**, 8385–8389.

(B18) P. W. R. Corfield, W. N. Hunter, T. Brown, P. Robinson & O. Kennard (1987). *Nucleic Acids Res.* **15**, 7935–7949.

(B19) W. N. Hunter, T. Brown, G. Kneale, N. N. Anand, D. Rabinovich & O. Kennard (1987). *J. Biol. Chem.* **21**, 9962–9970.

(B20) H. C. M. Nelson, J. T. Finch, B. F. Luisi & A. Klug (1987). *Nature (London)*, **330**, 221–226.

(B21) P. E. Pjura, K. Grzeskowiak & R. E. Dickerson (1987). *J. Mol. Biol.* **197**, 257–271.

(B22) G. G. Privé, U. Heinemann, S. Chandrasegaran, L.-S. Kan, M. L. Kopka & R. E. Dickerson (1987). *Science*, **238**, 498–504.

(B23) E. Westhof (1987). *J. Biomol. Struct. Dyn.* **5**, 581–600.

1988 (B24) C. A. Frederick, G. J. Quigley, G. A. van der Marel, J. H. van Boom, A. H.-J. Wang & A. Rich (1988). *J. Biol. Chem.* **263**, 17872–17879.

(B25) G. G. Privé, U. Heinemann, S. Chandrasegaran, L.-S. Kan, M. L. Kopka & R. E. Dickerson (1988). *Structure and expression*, Vol. 2. *DNA and its drug complexes*, edited by R. H. Sarma & M. H. Sarma, pp. 27–47. Schenectady, NY: Adenine Press.

(B26) M. Teng, N. Usman, C. A. Frederick & A. H.-J. Wang (1988). *Nucleic Acids Res.* **16**, 2671–2690.

(B27) C. Yoon, G. G. Privé, D. S. Goodsell & R. E. Dickerson (1988). *Proc. Natl Acad. Sci. USA*, **85**, 6332–6336.

1989 (B28) T. Brown, G. A. Leonard, E. D. Booth & J. Chambers (1989). *J. Mol. Biol.* **207**, 455–457.

(B29) M. A. A. F. de C. T. Carrondo, M. Coll, J. Aymami, A. H.-J. Wang, G. A. van der Marel, J. H. van Boom & A. Rich (1989). *Biochemistry*, **28**, 7849–7859.

(B30) M. Coll, J. Aymami, G. A. van der Marel, J. H. van Boom, A. Rich & A. H.-J. Wang (1989). *Biochemistry*, **28**, 310–320.

(B31) A. D. DiGabriele, M. R. Sanderson & T. A. Steitz (1989). *Proc. Natl Acad. Sci. USA*, **86**, 1816–1820.

(B32) U. Heinemann & C. Alings (1989). *J. Mol. Biol.* **210**, 369–381.

(B33) T. A. Larsen, D. S. Goodsell, D. Cascio, K. Grzeskowiak & R. E. Dickerson (1989). *J. Biomol. Struct. Dyn.* **7**, 477–491.

(B34) Y. Timsit, E. Westhof, R. P. P. Fuchs & D. Moras (1989). *Nature (London)*, **341**, 459–462.

1990 (B35) J. Aymami, M. Coll, G. A. van der Marel, J. H. van Boom, A. H.-J. Wang & A. Rich (1990). *Proc. Natl Acad. Sci. USA*, **87**, 2526–2530.

(B36) D. G. Brown, M. R. Sanderson, J. V. Skelly, T. C. Jenkins, T. Brown, E. Garman, D. I. Stuart & S. Neidle (1990). *EMBO J.* **9**, 1329–1334.

(B37) G. A. Leonard, E. D. Booth & T. Brown (1990). *Nucleic Acids Res.* **18**, 5617–5623.

(B38) G. A. Leonard, J. Thomson, W. P. Watson & T. Brown (1990). *Proc. Natl Acad. Sci. USA*, **87**, 9573–9576.

(B39) G. D. Webster, M. R. Sanderson, J. V. Skelly, S. Neidle, P. F. Swann, B. F. Li & I. J. Tickle (1990). *Proc. Natl Acad. Sci. USA*, **87**, 6693–6697.

1991 (B40) J. Balendrian & M. Sundaralingam (1991). *J. Biomol. Struct. Dyn.* **9**, 511–516.

(B41) Y.-G. Gao, G. A. van der Marel, J. H. van Boom & A. H.-J. Wang (1991). *Biochemistry*, **30**, 9922–9931.

(B42) K. Grzeskowiak, K. Yanagi, G. G. Privé & R. E. Dickerson (1991). *J. Biol. Chem.* **266**, 8861–8883.

Table A23.3.1.2. *X-ray analyses of B-DNA helices and their complexes with minor-groove-binding drug molecules* (*cont.*)


Year Reference

(B43) U. Heinemann & C. Alings (1991). *EMBO J.* **10**, 35–43.

(B44) T. A. Larsen, M. L. Kopka & R. E. Dickerson (1991). *Biochemistry*, **30**, 4443–4449.

(B45) N. Narayana, S. L. Ginell, I. M. Russu & H. M. Berman (1991). *Biochemistry*, **30**, 4450–4455.

(B46) G. G. Privé, K. Yanagi & R. E. Dickerson (1991). *J. Mol. Biol.* **217**, 177–199.

(B47) J. R. Quintana, A. A. Lipanov & R. E. Dickerson (1991). *Biochemistry*, **30**, 10294–10306.

(B48) Y. Timsit, E. Vilbois & D. Moras (1991). *Nature (London)*, **354**, 167–170.

(B49) Y. Timsit & D. Moras (1991). *J. Mol. Biol.* **221**, 919–940.

(B50) K. Yanagi, G. D. Privé & R. E. Dickerson (1991). *J. Mol. Biol.* **217**, 201–214.

1992 (B51) D. G. Brown, M. R. Sanderson, E. Garman & S. Neidle (1992). *J. Mol. Biol.* **226**, 481–490.

(B52) K. J. Edwards, D. G. Brown, N. Spink, J. V. Skelly & S. Neidle (1992). *J. Mol. Biol.* **226**, 1161–1173.

(B53) K. J. Edwards, T. C. Jenkins & S. Neidle (1992). *Biochemistry*, **31**, 7104–7109.

(B54) U. Heinemann & M. Hahn (1992). *J. Biol. Chem.* **267**, 7332–7341.

(B55) U. Heinemann, C. Alings & M. Bansal (1992). *EMBO J.* **11**, 1931–1939.

(B56) G. A. Leonard, E. D. Booth, W. N. Hunter & T. Brown (1992). *Nucleic Acids Res.* **20**, 4753–4759.

(B57) G. A. Leonard, A. Guy, T. Brown, R. Teoule & W. N. Hunter (1992). *Biochemistry*, **31**, 8415–8420.

(B58) J. R. Quintana, K. Grzeskowiak, K. Yanagi & R. E. Dickerson (1992). *J. Mol. Biol.* **225**, 379–395.

(B59) M. Sriram, G. A. van der Marel, H. L. P. F. Roelen, J. H. van Boom & A. H.-J. Wang (1992). *Biochemistry*, **31**, 11823–11834.

(B60) M. Sriram, G. A. van der Marel, H. L. P. F. Roelen, J. H. van Boom & A. H.-J. Wang (1992). *EMBO J.* **11**, 225–232.

(B61) J.-C. Xuan & I. T. Weber (1992). *Nucleic Acids Res.* **20**, 5457–5464.

(B62) H. Yuan, J. R. Quintana & R. E. Dickerson (1992). *Biochemistry*, **31**, 8009–8021.

1993 (B63) I. Baikalov, K. Grzeskowiak, K. Yanagi, J. Quintana & R. E. Dickerson (1993). *J. Mol. Biol.* **231**, 768–784.

(B64) A. D. DiGabriele & T. A. Steitz (1993). *J. Mol. Biol.* **231**, 1024–1029.

(B65) Y.-G. Gao, M. Sriram, W. A. Denny & A. H.-J. Wang (1993). *Biochemistry*, **32**, 9693–9648.

(B66) D. S. Goodsell, M. L. Kopka, D. Cascio & R. E. Dickerson (1993). *Proc. Natl Acad. Sci. USA*, **90**, 2930–2934.

(B67) K. Grzeskowiak, D. S. Goodsell, M. Kaczor-Grzeskowiak, D. Cascio & R. E. Dickerson (1993). *Biochemistry*, **32**, 8923–8931.

(B68) M. Hahn & U. Heinemann (1993). *Acta Cryst.* D**49**, 468–477.

(B69) G. A. Leonard & W. N. Hunter (1993). *J. Mol. Biol.* **234**, 198–208.

(B70) A. Lipanov, M. L. Kopka, M. Kaczor-Grzeskowiak, J. Quintana & R. E. Dickerson (1993). *Biochemistry*, **32**, 1373–1389.

(B71) C. M. Nunn, T. C. Jenkins & S. Neidle (1993). *Biochemistry*, **32**, 13838–13842.

(B72) J. V. Skelly, K. J. Edwards, T. C. Jenkins & S. Neidle (1993). *Proc. Natl Acad. Sci. USA*, **90**, 804–808.

(B73) L. Tabernero, N. Verdaguer, M. Coll, I. Fita, G. A. van der Marel, J. H. van Boom, A. Rich & J. Aymami (1993). *Biochemistry*, **32**, 8403–8410.

1994 (B74) X. Chen, B. Ramakrishnan, S. T. Rao & M. Sundaralingam (1994). *Nature Struct. Biol.* **1**, 169–170.

(B75) R. E. Dickerson, D. S. Goodsell & S. A. Neidle (1994). *Proc. Natl Acad. Sci. USA*, **91**, 3579–3583.

(B76) S. L. Ginnell, J. Vojtechovsky, B. Gaffney, R. Jones & H. M. Berman (1994). *Biochemistry*, **33**, 3487–3493.

(B77) D. S. Goodsell, M. Kaczor-Grzeskowiak & R. E. Dickerson (1994). *J. Mol. Biol.* **239**, 79–96.

(B78) M. L. Kopka, D. S. Goodsell, K. Grzeskowiak, I. Baikalov, D. Cascio & R. E. Dickerson (1994). *Biochemistry*, **33**, 13593–13610.

(B79) G. A. Leonard, K. E. McAuley-Hecht, N. J. Gibson, T. Brown, W. P. Watson & W. N. Hunter (1994). *Biochemistry*, **33**, 4755–4761.

(B80) K. E. McAuley-Hecht, G. A. Leonard, N. J. Gibson, J. B. Thomson, W. P. Watson, W. N. Hunter & T. Brown (1994). *Biochemistry*, **33**, 10266–10270.

(B81) C. M. Nunn, T. C. Jenkins & S. Neidle (1994). *Eur. J. Biochem.* **226**, 953–961.

(B82) Z. Shakked, G. Guzlkevich-Guerstein, F. Frolow, D. Rabinovich, A. Joachimiak & P. B. Sigler (1994). *Nature (London)*, **368**, 469–473.

(B83) N. Spink, D. G. Brown, J. V. Skelly & S. Neidle (1994). *Nucleic Acids Res.* **22**, 1607–1612.

(B84) M. C. Vega, I. Garcia-Saez, J. Aymami, R. Eritja, G. A. van der Marel, J. H. van Boom, A. Rich & M. Coll (1994). *Eur. J. Biochem.* **222**, 721–726.

(B85) Y. Timsit & D. Moras (1994). *EMBO J.* **13**, 2737–2746.

1995 (B86) K. Balendiran, S. T. Rao, C. Y. Sekharudu, G. Zon & M. Sundaralingam (1995). *Acta Cryst.* D**51**, 190–198.

(B87) X. Chen, B. Ramakrishnan & M. Sundaralingam (1995). *Nature Struct. Biol.* **2**, 733–735.

(B88) D. S. Goodsell, M. L. Kopka & R. E. Dickerson (1995). *Biochemistry*, **34**, 4983–4993.

(B89) D. S. Goodsell, K. Grzeskowiak & R. E. Dickerson (1995). *Biochemistry*, **34**, 1022–1029.

(B90) D. S. Goodsell, H. L. Ng, M. L. Kopka, J. W. Lown & R. E. Dickerson (1995). *Biochemistry*, **34**, 16654–16661.

(B91) L. A. Lipscomb, M. E. Peek, M. L. Morningstar, S. M. Verghis, E. M. Miller, A. Rich, J. M. Essigmann & L. D. Williams (1995). *Proc. Natl Acad. Sci. USA*, **92**, 719–723.

(B92) C. M. Nunn & S. Neidle (1995). *J. Med. Chem.* **38**, 2317–2325.

(B93) L. W. Tari & A. S. Secco (1995). *Nucleic Acids Res.* **23**, 2065–2073.

(B94) L. Van Meervelt, D. Vlieghe, A. Dautant, B. Gallois, G. Precigoux & O. Kennard (1995). *Nature (London)*, **374**, 742–744.

(B95) J. Vojtechovsky, M. D. Eaton, B. Gaffney, R. Jones & H. M. Berman (1995). *Biochemistry*, **34**, 16632–16640.

(B96) A. A. Wood, C. M. Nunn, A. Czarny, D. W. Boykin & S. Neidle (1995). *Nucleic Acids Res.* **23**, 3678–3684.

Table A23.3.1.2. *X-ray analyses of B-DNA helices and their complexes with minor-groove-binding drug molecules* (*cont.*)


Year Reference

1996 (B97) T. J. Boggon, E. L. Hancox, K. E. McAuley-Hecht, B. A. Connolly, W. N. Hunter, T. Brown, R. T. Walker & G. A. Leonard (1996). *Nucleic Acids Res.* **24**, 951–961.

(B98) G. R. Clark, E. J. Gray, S. Neidle, Y.-H. Li & W. Leupin (1996). *Biochemistry* **35**, 13745–13752.

(B99) G. R. Clark, C. J. Squire, E. J. Gray, W. Leupin & S. Neidle (1996). *Nucleic Acids Res.* **24**, 4882–4889.

(B100) C. A. Laughton, F. Tanious, C. M. Nunn, D. W. Boykin, W. D. Wilson & S. Neidle (1996). *Biochemistry*, **35**, 5655–5661.

(B101) J. O. Trent, G. R. Clark, A. Kumar, W. D. Wilson, D. W. Boykin, J. E. Hall, R. R. Tidwell, B. L. Blagburn & S. Neidle (1996). *J. Med. Chem.* **39**, 4554–4562.

(B102) L. Urpi, V. Tereshko, L. Malinina, T. Huynh-Dinh & J. A. Subirana (1996). *Nature Struct. Biol.* **3**, 325–328.

(B103) D. Vlieghe, L. Van Meervelt, A. Dautant, B. Gallois, G. Precigoux & O. Kennard (1996). *Science*, **273**, 1702–1705.

(B104) M. C. Wahl, S. T. Rao & M. Sundaralingam (1996). *Biophys. J.* **70**, 2857–2866.

1997 (B105) X. Chen, B. Ramakrishnan & M. Sundaralingam (1997). *J. Mol. Biol.* **267**, 1157–1170.

(B106) G. R. Clark, D. W. Boykin, A. Czarny & S. Neidle (1997). *Nucleic Acids Res.* **25**, 1510–1515.

(B107) G.-W. Han, M. L. Kopka, D. Cascio, K. Grzeskowiak & R. E. Dickerson (1997). *J. Mol. Biol.* **269**, 811–826.

(B108) M. L. Kopka, D. S. Goodsell, G. W. Han, T. K. Chiu, J. W. Lown & R. E. Dickerson (1997). *Structure*, **5**, 1033–1046.

(B109) C. Mayer-Jung, D. Moras & Y. Timsit (1997). *J. Mol. Biol.* **270**, 328–335.

(B110) C. M. Nunn, E. Garman & S. Neidle (1997). *Biochemistry*, **36**, 4792–4799.

(B111) S. Portmann, K.-H. Altmann, N. Reynes & M. Egli (1997). *J. Am. Chem. Soc.* **119**, 2396–2403.

(B112) H. Qiu, J. C. Dewan & N. C. Seeman (1997). *J. Mol. Biol.* **267**, 881–898.

(B113) M. Shatzky-Schwartz, N. D. Arbuckle, M. Eisenstein, D. Rabinovich, A. Bareket-Samish, T. E. Haran, B. F. Luisi & Z. Shakked (1997). *J. Mol. Biol.* **267**, 565–623.

(B114) A. A. Wood, C. M. Nunn, J. O. Trent & S. Neidle (1997). *J. Mol. Biol.* **269**, 827–841.


## Table A23.3.1.3. *X-ray analyses of Z helices*

See introductory notes to Table A23.3.1.1. odm = 6*H*,8*H*-3,4-dihydropyrimido[4,5*c*][1,2]oxazin-7-one.

(*a*) Hexadecamers

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| CGCGCGTTTTCGCGCG (hairpin) | *C*2 | 4 | 8 | 1988, UCLA | UDP011 | (Z20, Z25) |

(*b*) Decamers (disordered)

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| GCGCGCGCGC | $P6_5$ | 6 | 2 | 1996, Ohio State | ZDJ050 | (Z46) |

(*c*) Octamers

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| CGCICICG | $P6_5$ | 6 | 8 | 1992, Thomas Jefferson | ZDH030 | (Z32) |
| CGCGCGCG | $P6_5$ | 6 | 8 | 1985, MIT | (ZDH017) | (Z10) |
| CGCATGCG | $P6_5$ | 6 | 8 | 1985, MIT | (ZDH016) | (Z10) |

(*d*) Hepamers (overhanging 5′ bases)

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| GCGCGCG | $P2_12_12_1$ | 4 | 6 | 1997, Oregon State | ZDG054 | (Z50) |
| $G^{5me}$CGCGCG | $P2_12_12_1$ | 4 | 6 | 1997, Oregon State | ZDG055 | (Z50) |
| GCGCGCG/ GCGCGCT | $P2_12_12_1$ | 4 | 6 | 1997, Oregon State | ZDG056 | (Z50) |
| GCGCGCG | $P2_12_12_1$ | 4 | 6 | 1997, Ohio State | ZDG057 | (Z51) |

(*e*) Hexamers

(1) Alternating CG: Pu-Py alternation retained

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| CGCGCG, Mg | $P2_12_12_1$ | 4 | 6 | 1989, MIT | ZDF002 | (Z23) |
| CGCGCG, DL racemate | $P\bar{1}$ | 2 | 6 | 1993, Osaka | ZDF040 | (Z36) |

Table A23.3.1.3. *X-ray analyses of Z helices* (*cont.*)

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| CGCGCG/spermine | $P2_12_12_1$ | 4 | 6 | 1991, MIT | ZDF029 | (Z29) |
| CGCGCG/spermine, 163 K | $P2_12_12_1$ | 4 | 6 | 1994, MIT | ZDF035 | (Z41) |
| CGCGCG/spermine, Mg | $P2_12_12_1$ | 4 | 6 | 1979, MIT | ZDF001 | (Z1, Z23) |
| CGCGCG/spermidine | $P2_12_12_1$ | 4 | 6 | 1996, MIT | ZDF052 | (Z47) |
| CGCGCG/thermospermidine | $P2_12_12_1$ | 4 | 6 | 1996, MIT | ZDF053 | (Z48) |
| CGCGCG, Co, Mg | $P2_12_12_1$ | 4 | 6 | 1985, MIT | (ZDF019) | (Z11) |
| CGCGCG, Co, Mg | $P2_12_12_1$ | 4 | 6 | 1993, Illinois | (ZDF044) | (Z37) |
| CGCGCG/spermine, Co | $P2_12_12_1$ | 4 | 6 | 1993, Illinois | (ZDF045) | (Z37) |
| CGCGCG, Ru | $P2_12_12_1$ | 4 | 6 | 1987, MIT | (ZDF007) | (Z18) |
| CG c g CG | $P2_12_12_1$ | 4 | 6 | 1989, MIT | (ZHF026) | (Z24) |

(2) Alternating CG: Pu-Py alternation broken

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| CCGCGG | $C222_1$ | 8 | 6 | 1994, Moscow | UDF025 | (Z42) |

(3) Modified CG bases: Pu-Py alternation retained

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| CG$^{ara}$CGCG | $P2_12_12_1$ | 4 | 6 | 1989, MIT | (ZDFS27) | (Z24) |
| CGC$^{6mo}$GCG | $P2_12_12_1$ | 4 | 6 | 1990, Rutgers | ZDFB21 | (Z26) |
| CGCG$^{4mo}$CG | $P2_12_12_1$ | 4 | 6 | 1990, Cambridge | ZDFB25 | (Z27) |
| CGCG$^{4mo}$CG | $P2_12_12_1$ | 4 | 6 | 1993, Manchester | ZDFB36 | (Z35) |
| CGCG$^{5br}$CG | $P2_12_12_1$ | 4 | 6 | 1996, Manchester | ZDFB51 | (Z49) |
| CGCG$^{odm}$CG | $P2_12_12_1$ | 4 | 6 | 1995, Cambridge | ZDFB43 | (Z43) |
| $^{5me}$CG$^{5me}$CG$^{5me}$CG | $P2_12_12_1$ | 4 | 6 | 1982, MIT | ZDFB03 | (Z6, Z7) |
| $^{5br}$CG$^{5br}$CG$^{5br}$CG, 291 K | $P2_12_12_1$ | 4 | 6 | 1986, Strasbourg | ZDFB04 | (Z16, Z19) |
| $^{5br}$CG$^{5br}$CG$^{5br}$CG, 310 K | $P2_12_12_1$ | 4 | 6 | 1986, Strasbourg | ZDFB05 | (Z16, Z19) |
| Aminohexyl-CG$^{5br}$CGCG | $C2$ | 4 | 6 | 1993, Illinois | (ZDFA32) | (Z38) |
| $^{ara}$CG$^{ara}$CG$^{ara}$CG (disordered) | $P6_522$ | 12 | 2 | 1992, Illinois | ZDFS33 | (Z34) |

(4) Modified CG bases: Pu-Py alternation broken

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| $^{5me}$CGGG$^{5me}$CG | $P2_12_12_1$ | 4 | 6 | 1993, Oregon State | ZDFB37 | (Z40) |

(5) With A, T, U, I bases: Pu-Py alternation retained

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| $^{5me}$CGTA$^{5me}$CG | $P2_12_12_1$ | 4 | 6 | 1984, MIT | ZDFB06 | (Z8) |
| CGT$^{2am}$ACG | $P3_221$ | 6 | 3 | 1995, Rutgers | ZDFB41 | (Z44) |
| CGT$^{2am}$ACG, Pt | $P3_221$ | 6 | 3 | 1995, Rutgers | ZDFB42 | (Z44) |
| CGU$^{2am}$ACG | $P2_12_12_1$ | 4 | 6 | 1992, Rutgers | ZDFB31 | (Z33) |
| $^{5me}$CGUA$^{5me}$CG | $P2_12_12_1$ | 4 | 6 | 1990, Oregon State | ZDFB24 | (Z28) |
| $^{5me}$CGUA$^{5me}$CG, Cu | $P2_12_12_1$ | 4 | 5 | 1991, Oregon State | ZDFB10 | (Z30) |
| CACGTG | $P2_12_12_1$ | 4 | 6 | 1988, MIT | (ZDF008) | (Z21) |
| C$^{2am}$ACGTG | $P2_12_12_1$ | 4 | 6 | 1986, MIT | ZDFB11 | (Z17) |
| CGCICG | $P2_12_12_1$ | 4 | 6 | 1993, Thomas Jefferson | (ZDFB34) | (Z39) |
| CACGCG/CGCGTG | $P2_12_12_1$ | 4 | 6 | 1995, Madras | ZDF039 | (Z45) |
| CGCACG/CGTGCG | $P2_1$ | 2 | 6 | 1995, Madras | ZDF038 | (Z45) |

(6) With A, T, U, I bases: Pu-Py alternation broken

| Sequence | Space group | *Z* | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| $^{5br}$CGAT$^{5br}$CG | $P2_12_12_1$ | 4 | 6 | 1985, MIT | (ZDFB09) | (Z13) |

Table A23.3.1.3. *X-ray analyses of Z helices* (*cont.*)

(7) With mismatches (underlined)

| Sequence | Space group | $Z$ | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| C<u>G</u>CG<u>T</u>G | $P2_12_12_1$ | 4 | 6 | 1985, MIT | ZDF013 | (Z12) |
| C<u>G</u>CG$^{5fl}$<u>U</u>G | $P2_12_12_1$ | 4 | 6 | 1989, MIT | ZDFB12 | (Z22) |
| $^{5br}$<u>U</u>GCGC<u>G</u> | $P2_12_12_1$ | 4 | 6 | 1986, Cambridge | ZDFB14 | (Z15) |
| C<u>G</u>CG<u>T</u>G, Co, Mg | $P2_12_12_1$ | 4 | 6 | 1993, Illinois | ZDF046 | (Z37) |
| C<u>G</u>CG<u>T</u>G, Cu, Mg | $P2_12_12_1$ | 4 | 6 | 1993, Illinois | ZDF047 | (Z37) |
| $^{5me}$C<u>G</u>$^{5me}$CG<u>T</u>G, Ba | $P2_12_12_1$ | 4 | 6 | 1993, Illinois | (ZDFB48) | (Z37) |

($f$) Tetramers

| Sequence | Space group | $Z$ | Ubp | Date, institution | NDB No. | Reference |
|---|---|---|---|---|---|---|
| CGCG | $C222_1$ | 8 | 4 | 1980, UCLA (CIT) | ZDD015 | (Z3, Z4, Z5) |
| CGCG (disordered) | $P6_5$ | 6 | 6 | 1980, MIT | ZDD023 | (Z2) |

References (numbered chronologically by year and alphabetically by first author within each year)


Year Reference

1979 (Z1) A. H.-J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel & A. Rich (1979). *Nature (London)*, **282**, 680–686.

1980 (Z2) J. L. Crawford, F. J. Kolpak, A. H.-J. Wang, G. J. Quigley, J. H. van Boom, G. van der Marel & A. Rich (1980). *Proc. Natl Acad. Sci. USA*, **77**, 4016–4020.

(Z3) H. R. Drew, T. Takano, S. Tanaka, K. Itakura & R. E. Dickerson (1980). *Nature (London)*, **286**, 567–573.

1981 (Z4) R. E. Dickerson, H. R. Drew & B. N. Conner (1981). *Biomolecular stereodynamics*, Vol. 1, edited by R. H. Sarma, pp. 1–34. New York: Adenine Press.

(Z5) H. R. Drew & R. E. Dickerson (1981). *J. Mol. Biol.* **152**, 723–736.

1982 (Z6) S. Fujii, A. H.-J. Wang, G. van der Marel, J. H. van Boom & A. Rich (1982). *Nucleic Acids Res.* **10**, 7879–7892.

(Z7) S. Fujii, A. H.-J. Wang, J. van Boom & A. Rich (1982). *Nucleic Acids Res. Symp. Ser.* **11**, 109–112.

1984 (Z8) A. H.-J. Wang, T. Hakoshima, G. van der Marel, J. H. van Boom & A. Rich (1984). *Cell*, **37**, 321–331.

1985 (Z9) R. G. Brennan & M. Sundaralingam (1985). *J. Mol. Biol.* **181**, 561–563.

(Z10) S. Fujii, A. H.-J. Wang, G. J. Quigley, H. Westerink, G. van der Marel, J. H. van Boom & A. Rich (1985). *Biopolymers*, **24**, 243–250.

(Z11) R. V. Gessner, G. J. Quigley, A. H.-J. Wang, G. A. van der Marel, J. H. van Boom & A. Rich (1985). *Biochemistry*, **24**, 237–240.

(Z12) P. S. Ho, C. A. Frederick, G. J. Quigley, G. A. van der Marel, J. H. van Boom, A. H.-J. Wang & A. Rich (1985). *EMBO J.* **4**, 3617–3623.

(Z13) A. H.-J. Wang, R. V. Gessner, G. A. van der Marel, J. H. van Boom & A. Rich (1985). *Proc. Natl Acad. Sci. USA*, **82**, 3611–3615.

1986 (Z14) R. G. Brennan, E. Westhof & M. Sundaralingam (1986). *J. Biomol. Struct. Dyn.* **3**, 649–665.

(Z15) T. Brown, G. Kneale, W. N. Hunter & O. Kennard (1986). *Nucleic Acids Res.* **14**, 1801–1809.

(Z16) B. Chevrier, A. C. Dock, B. Hartmann, M. Leng, D. Moras, M. T. Thuong & E. Westhof (1986). *J. Mol. Biol.* **188**, 707–719.

(Z17) M. Coll, A. H.-J. Wang, G. A. van der Marel, J. H. van Boom & A. Rich (1986). *J. Biomol. Struct. Dyn.* **4**, 157–172.

1987 (Z18) P. S. Ho, C. A. Frederick, D. Saal, A. H.-J. Wang & A. Rich (1987). *J. Biomol. Struct. Dyn.* **4**, 521–534.

(Z19) E. Westhof (1987). *J. Biomol. Struct. Dyn.* **5**, 581–600.

1988 (Z20) R. Chattopadhyaya, S. Ikuta, K. Grzeskowiak & R. E. Dickerson (1988). *Nature (London)*, **334**, 175–179.

(Z21) M. Coll, I. Fita, J. Lloveras, J. A. Subirana, F. Bardella, T. Huynh-Dinh & J. Igolen (1988). *Nucleic Acids Res.* **16**, 8695–8705.

1989 (Z22) M. Coll, D. Saal, C. A. Frederick, J. Aymami, A. Rich & A. H.-J. Wang (1989). *Nucleic Acids Res.* **17**, 911–923.

(Z23) R. V. Gessner, C. A. Frederick, G. J. Quigley, A. Rich & A. H.-J. Wang (1989). *J. Biol. Chem.* **264**, 7921–7935.

(Z24) M.-K. Teng, Y.-C. Liaw, G. A. van der Marel, J. H. van Boom & A. H.-J. Wang (1989). *Biochemistry*, **28**, 4923–4928.

1990 (Z25) R. Chattopadhyaya, K. Grzeskowiak & R. E. Dickerson (1990). *J. Mol. Biol.* **211**, 189–210.

(Z26) S. L. Ginell, S. Kuzmich, R. A. Jones & H. M. Berman (1990). *Biochemistry*, **29**, 10461–10465.

(Z27) L. Van Meervelt, M. H. Moore, P. K. T. Lin, D. M. Brown & O. Kennard (1990). *J. Mol. Biol.* **216**, 773–781.

(Z28) G. Zhou & P. S. Ho (1990). *Biochemistry*, **29**, 7229–7236.

1991 (Z29) M. Egli, L. D. Williams, Q. Gao & A. Rich (1991). *Biochemistry*, **30**, 11388–11402.

(Z30) B. H. Geierstanger, T. F. Kagawa, S.-L. Chen, G. J. Quigley & P. S. Ho (1991). *J. Biol. Chem.* **266**, 20185–20191.

1992 (Z32) V. D. Kumar, R. W. Harrison, L. C. Andrews & I. T. Weber (1992). *Biochemistry*, **31**, 1541–1550.

(Z33) B. Schneider, S. L. Ginnell, R. Jones, B. Gaffney & H. M. Berman (1992). *Biochemistry*, **21**, 9622–9628.

(Z34) H. Zhang, G. A. van der Marel, J. H. van Boom & A. H.-J. Wang (1992). *Biopolymers*, **32**, 1559–1569.

1993 (Z35) A. R. Cervi, A. Guy, G. A. Leonard, R. Teoule & W. N. Hunter (1993). *Nucleic Acids Res.* **21**, 5623–5629.

(Z36) M. Doi, M. Inoue, K. Tomoo, T. Ishida, Y. Ueda, M. Akagi & H. Urata (1993). *J. Am. Chem. Soc.* **115**, 10432–10433.

(Z37) Y.-G. Gao, K. Sriram & A. H.-J. Wang (1993). *Nucleic Acids Res.* **21**, 4093–4101.

(Z38) Y.-C. Jean, Y.-G. Gao & A. H.-J. Wang (1993). *Biochemistry*, **32**, 381–388.

Table A23.3.1.3. *X-ray analyses of Z helices* (*cont.*)

Year Reference

(Z39) V. D. Kumar & I. T. Weber (1993). *Nucleic Acids Res.* **9**, 2201–2208.

(Z40) G. P. Schroth, T. F. Kagawa & P. Shing Ho (1993). *Biochemistry*, **32**, 13381–13392.

1994 (Z41) D. Bancroft, L. D. Williams, A. Rich & M. Egli (1994). *Biochemistry*, **33**, 1073–1086.

(Z42) L. Malinina, L. Urpi, X. Salas, T. Huynh-Dinh & J. A. Subirana (1994). *J. Mol. Biol.* **243**, 484–493.

1995 (Z43) M. H. Moore, L. Van Meervelt, S. A. Salisbury, P. Kong Thoo Lin & D. M. Brown (1995). *J. Mol. Biol.* **251**, 665–673.

(Z44) G. N. Parkinson, G. M. Arvantis, L. Lessinger, S. L. Ginnell, R. Jones, B. Gaffney & H. M. Berman (1995). *Biochemistry*, **34**, 15487–15495.

(Z45) C. Sadasivan & N. Gautham (1995). *J. Mol. Biol.* **248**, 918–930.

1996 (Z46) C. Ban, B. Ramakrishnan & M. Sundaralingam (1996). *Biophys. J.* **71**, 1215–1221.

(Z47) H. Ohishi, I. Nakanishi, K. Inubushi, G. A. van der Marel, J. H. van Boom, A. Rich, A. H.-J. Wang, T. Hakoshima & K. Tomita (1996). *FEBS Lett.* **391**, 143–156.

(Z48) H. Ohishi, N. Terasoma, I. Nakanishi, G. A. van der Marel, J. H. van Boom, A. Rich, A. H.-J. Wang, T. Hakoshima & K. Tomita (1996). *FEBS Lett.* **398**, 291–296.

(Z49) M. R. Peterson, S. J. Harrop, S. M. McSweeney, G. A. Leonard, A. W. Thompson, W. N. Hunter & J. R. Helliwell (1996). *J. Synchrotron Rad.* **3**, 24–34.

1997 (Z50) B. H. M. Mooers, B. F. Eichman & P. S. Ho (1997). *J. Mol. Biol.* **269**, 796–810.

(Z51) B. Pan, C. Ban, M. Wahl & M. Sundaralingam (1997). *Biophys. J.* **83**, 1553–1561.

### 23.2 (*cont.*)


Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). *Crystal structure of trp repressor/operator complex at atomic resolution*. *Nature (London)*, **335**, 321–329; erratum (1988), **335**, 837.

Pabo, C. O. & Sauer, R. T. (1992). *Transcription factors: structural families and principles of DNA recognition*. *Annu. Rev. Biochem.* **61**, 1053–1095.

Pardee, A. B. (1966). *Purification and properties of a sulfate-binding protein from Salmonella typhimurium*. *J. Biol. Chem.* **241**, 5886–5892.

Pavletich, N. P. & Pabo, C. O. (1991). *Zinc finger–DNA recognition: crystal structure of a Zif268–DNA complex at 2.1 Å*. *Science*, **252**, 809–817.

Pflugrath, J. W. & Quiocho, F. A. (1985). *Sulphate sequestered in the sulphate-binding protein of Salmonella typhimurium is bound solely by hydrogen bonds*. *Nature (London)*, **314**, 257–260.

Quiocho, F. A. (1986). *Carbohydrate-binding proteins: tertiary structures and protein–sugar interactions*. *Annu. Rev. Biochem.* **55**, 287–315.

Quiocho, F. A., Sack, J. S. & Vyas, N. K. (1989). *Substrate specificity and affinity of a protein modulated by bound water molecules*. *Nature (London)*, **340**, 404–407.

Quiocho, F. A., Wilson, D. K. & Vyas, N. K. (1987). *Stabilization of charges on isolated charged groups sequestered in proteins by polarized peptide units*. *Nature (London)*, **329**, 561–564.

Rademacher, T. W., Parekh, R. B. & Dwek, R. A. (1988). *Glycobiology*. *Annu. Rev. Biochem.* **57**, 785–838.

Rould, M. A., Perona, J. J. & Steitz, T. A. (1991). *Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase*. *Nature (London)*, **352**, 213–218.

Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). *Sequence-specific recognition of double helical nucleic acids by proteins*. *Proc. Natl Acad. Sci. USA*, **73**, 804–808.

Shimon, L. J. & Harrison, S. C. (1993). *The phage 434 OR2/R1-69 complex at 2.5 Å resolution*. *J. Mol. Biol.* **232**, 826–838.

Steitz, T. A. (1990). *Structural studies of protein–nucleic acid interaction: the sources of sequence-specific binding*. *Q. Rev. Biophys.* **23**, 205–280.

Sutton, R. B., Fasshauer, D., Jan, R. & Brunger, A. T. (1998). *Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution*. *Nature (London)*, **395**, 347–353.

Tucker, P. W., Hazen, E. E. Jr & Cotton, F. A. (1979). *Staphylococcal nuclease reviewed: a prototypic study in contemporary enzymology. III. Correlation of the three-dimensional structure with the mechanisms of enzymatic action*. *Mol. Cell. Biochem.* **23**, 67–86.

Ueda, H., Iyo, H., Doi, M., Inoue, M. & Ishida, T. (1991). *Cooperative stacking and hydrogen bond pairing interactions of fragment peptide in cap binding protein with mRNA cap structure*. *Biochim. Biophys. Acta*, **1075**, 181–186.

Ueda, H., Iyo, H., Doi, M., Inoue, M., Ishida, T., Morioka, H., Tanaka, T., Nishikawa, S. & Uesugi, S. (1991). *Combination of Trp and Glu residues for recognition of mRNA cap structure. Analysis of m7G base recognition site of human cap binding protein (IF-4E) by site-directed mutagenesis*. *FEBS Lett.* **280**, 207–210.

Varani, G. (1997). *A cap for all occasions. Structure*, **5**, 855–858.

Volbeda, A., Fontecilla-Camps, J. C. & Frey, M. (1996). *Novel metal sites in protein structures. Curr. Opin. Struct. Biol.* **6**, 804–812.

Vyas, N. K. (1991). *Atomic features of protein–carbohydrate interactions. Curr. Opin. Struct. Biol.* **1**, 732–740.

Vyas, N. K., Vyas, M. N. & Quiocho, F. A. (1988). *Sugar and signal-transducer binding sites of the Escherichia coli galactose chemoreceptor protein. Science*, **242**, 1290–1295.

Wang, Z., Luecke, H., Yao, N. & Quiocho, F. A. (1997). *Nature Struct. Biol.* **4**, 519–522.

Weis, W. I. & Drickamer, K. (1996). *Structural basis of lectin–carbohydrate recognition. Annu. Rev. Biochem.* **65**, 441–473.

Werner, M. H., Gronenborn, A. M. & Clore, G. M. (1996). *Intercalation, DNA kinking, and the control of transcription. Science*, **271**, 778–784; erratum (1996), **272**, 19.

Worm, S. H. van den, Stonehouse, N. J., Valegard, K., Murray, J. B., Walton, C., Fridborg, K., Stockley, P. G. & Liljas, L. (1998). *Crystal structures of MS2 coat protein mutants in complex with wild-type RNA operator fragments*. *Nucleic Acids Res.* **26**, 1345–1351.

Yao, N., Ledvina, P. S., Choudhary, A. & Quiocho, F. A. (1996). *Modulation of a salt link does not affect binding of phosphate to its specific active transport receptor. Biochemistry*, **35**, 2079–2085.


### 23.3


Altona, C., Geise, H. J. & Romers, C. (1968). *Conformation of non-aromatic ring compounds, XXIV. On the geometry of the perhydrophenanthrene skeleton in some steroids. Tetrahedron*, **24**, 13–32.

Altona, C. & Sundaralingam, M. (1972). *Conformational analysis of the sugar ring in nucleosides and nucleotides. J. Am. Chem. Soc.* **94**, 8205–8212.

Ansevin, A. T. & Wang, A. H. (1990). *Evidence for a new Z-type left-handed DNA helix. Nucleic Acids Res.* **18**, 6119–6126.

Arnott, S. (1970). *The geometry of nucleic acids. Prog. Biophys. Mol. Biol.* **21**, 265–319.

Babcock, M. S. & Olson, W. K. (1994). *The effect of mathematics and coordinate system on comparability and 'dependencies' of nucleic acid structure parameters. J. Mol. Biol.* **237**, 98–124.

Babcock, M. S., Pednault, E. & Olson, W. (1993). *Nucleic acid structure analysis: a users guide to a collection of new analysis programs. J. Biomol. Struct. Dyn.* **11**, 597–628.

Babcock, M. S., Pednault, E. & Olson, W. (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.

Basham, B., Eichman, B. F. & Ho, P. S. (1998). *The single-crystal structures of Z-DNA*. In *Oxford handbook of nucleic acid structure*, edited by S. Neidle, ch. 7, pp. 200–252. Oxford University Press.

Berman, H. M. (1996). *Crystal studies of B-DNA: the answers and the questions. Biopolymers Nucleic Acid Sci.* **44**, 23–44.

Bugg, C. E., Thomas, J. M., Sundaralingam, M. & Rao, S. T. (1971). *Stereochemistry of nucleic acids and their constituents. X. Solid-state base-stacking patterns in nucleic acid consituents and polynucleotides. Biopolymers*, **10**, 175–219.

Crick, F. H. C. & Watson, J. D. (1954). *The complementary structure of deoxyribonucleic acid. Proc. R. Soc. London Ser. A*, **223**, 80–96.

Crothers, D. M. & Drak, J. (1992). *Global features of DNA structure by comparative gel electrophoresis. Methods Enzymol.* **212**, 46–71.

Crothers, D. M., Haran, T. E. & Nadeau, J. G. (1990). *Intrinsically bent DNA. J. Biol. Chem.* **265**, 7093–7096.

Davies, D. B. (1978). *Conformations of nucleosides and nucleotides. Prog. Nucl. Magn. Reson. Spectros.* **12**, 135–186.

Dickerson, R. E. (1972). *The structure and history of an ancient protein. Sci. Am.* **226** (April), 58–72.

Dickerson, R. E. (1983). *The DNA helix and how it is read. Sci. Am.* **249** (December), 94–111.

Dickerson, R. E. (1985). *Helix polymorphism and information flow in DNA*. In *Proceedings of the Robert A. Welch Foundation conferences on chemical research XXIX: Genetic chemistry; the molecular basis of heredity*, pp. 38–79.

Dickerson, R. E. (1992). *DNA structure from A to Z. Methods Enzymol.* **211**, 67–111.

Dickerson, R. E. (1997*a*). *Obituary: Irving Geis, 1908–1997. Structure*, **5**, 1247–1249.

Dickerson, R. E. (1997*b*). *Irving Geis, molecular artist, 1908–1997. Protein Sci.* **6**, 2843–2844.

Dickerson, R. E. (1997*c*). *Biology in pictures: molecular artistry. Curr. Biol.* **7**, R720–R741.

**23.3 (*cont.*)**


Dickerson, R. E. (1998*a*). *Sequence-dependent B-DNA conformation in crystals and in protein complexes*. In *Structure, motion, interaction and expression of biological macromolecules*, edited by R. H. Sarma & M. H. Sarma, pp. 17–36. New York: Adenine Press.

Dickerson, R. E. (1998*b*). *Helix structure and molecular recognition by B-DNA*. In *Oxford handbook of nucleic acid structure*, edited by S. Neidle, ch. 7, pp. 145–197. Oxford University Press.

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

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., von Kitzing, E., Wang, A. H.-J. & Zhurkin, V. B. (1989). *Definitions and nomenclature of nucleic acid structure components. EMBO J.* **8**, 1–4; *J. Biomol. Struct. Dyn.* **6**, 627–634; *Nucleic Acids Res.* **17**, 1797–1803; *J. Mol. Biol.* **206**, 787–791.

Dickerson, R. E. & Chiu, T. K. (1997). *Helix bending as a factor in protein/DNA recognition. Biopolymers Nucleic Acid Sci.* **44**, 361–403.

Dickerson, R. E. & Geis, I. (1969). *The structure and action of proteins*. New York: Harper & Row and Menlo Park: W. A. Benjamin Co.

Dickerson, R. E. & Geis, I. (1976). *Chemistry, matter and the universe*. Menlo Park: Benjamin/Cummings Co.

Dickerson, R. E. & Geis, I. (1983). *Hemoglobin: structure, function, evolution, and pathology*. Menlo Park: Benjamin/Cummings Co.

Dickerson, R. E., Goodsell, D. & Kopka, M. L. (1996). *MPD and DNA bending in crystals and in solution. J. Mol. Biol.* **256**, 108–125.

Dickerson, R. E., Goodsell, D. S., Kopka, M. L. & Pjura, P. E. (1987). *The effect of crystal packing on oligonucleotide double helix structure. J. Biomol. Struct. Dyn.* **5**, 557–579.

Dickerson, R. E., Goodsell, D. S. & Neidle, S. (1994). . . . *the tyranny of the lattice*. . .. *Proc. Natl Acad. Sci. USA*, **91**, 3579–3583.

El Hassan, M. A. & Calladine, C. R. (1997). *Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. Philos. Trans. R. Soc. London A*, **355**, 43–100.

Feigon, J. (1996). *DNA triplexes, quadruplexes & aptamers*. In *Encyclopedia of nuclear magnetic resonance*, edited by D. M. Grant & R. K. Harris, pp. 1726–1731. New York: Wiley.

Franklin, R. E. & Gosling, R. G. (1953). *The structure of sodium thymonucleate fibres. I. The influence of water content. Acta Cryst.* **6**, 673–677.

Haschmeyer, A. E. V. & Rich, A. (1967). *Nucleoside conformation: an analysis of steric barriers to rotation about the glycosidic bond. J. Mol. Biol.* **27**, 369–384.

Herbert, A. & Rich, A. (1996). *The biology of left-handed Z-DNA. J. Biol. Chem.* **271**, 11595–11598.

Ho, P. S. & Mooers, B. H. M. (1996). *Z-DNA crystallography. Biopolymers Nucleic Acid Sci.* **44**, 65–90.

Hoogsteen, K. (1963). *The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. Acta Cryst.* **16**, 907–916.

Hunter, C. A. & Sanders, J. K. M. (1990). *The nature of $\pi$–$\pi$ interactions. J. Am. Chem. Soc.* **112**, 5525–5534.

Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J. & Dickerson, R. E. (1996). *How proteins recognize the TATA box. J. Mol. Biol.* **261**, 239–254.

Kendrew, J. C. (1961). *The three-dimensional structure of a protein molecule. Sci. Am.* **205** (December), 96–110.

Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). *Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature (London)*, **365**, 520–527.

Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). *Crystal structure of a yeast TBP/TATA-box complex. Nature (London)*, **365**, 512–520.

Koo, H.-S., Drak, J., Rice, J. A. & Crothers, D. M. (1990). *Determination of the extent of DNA bending by an adenine-thymine tract. Biochemistry*, **29**, 4227–4234.

Koo, H.-S., Wu, H.-M. & Crothers, D. M. (1986). *DNA bending at adenine-thymine tracts. Nature (London)*, **320**, 501–506.

Kostrewa, D. & Winkler, F. K. (1995). *$Mg^{2+}$ binding to the active site of EcoRV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 Å resolution. Biochemistry*, **34**, 683–696.

Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, M. H. F. & Hamilton, L. D. (1960). *The molecular configurations of deoxyribonucleic acid. II. Molecular models and their Fourier transforms. J. Mol. Biol.* **2**, 38–64.

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

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

Leslie, A. G. W., Arnott, S., Chandrasekaran, R. & Ratliff, R. L. (1980). *Polymorphism of DNA double helices. J. Mol. Biol.* **143**, 49–72.

Levitt, M. & Warshel, A. (1978). *Extreme conformational flexibility of the furanose ring in DNA and RNA. J. Am. Chem. Soc.* **100**, 2607–2613.

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G. & Lu, P. (1996). *Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science*, **271**, 1247–1254.

Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. (1982). *Bent helical structures in kinetoplast DNA. Proc. Natl Acad. Sci. USA*, **79**, 7664–7668.

Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G. & Burley, S. K. (1996). *Crystal structure of a human TATA box-binding protein/TATA element complex. Proc. Natl Acad. Sci. USA*, **93**, 4862–4867.

Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. H. & Berman, H. M. (1996). *Structure of the CAP–DNA complex at 2.5 angstroms resolution: a complete picture of the protein–DNA interface. J. Mol. Biol.* **260**, 395–408.

Pelton, J. G. & Wemmer, D. E. (1989). *Structural characterization of a 2:1 distamycin A/d(CGCAAATTGGC) complex by two-dimensional NMR. Proc. Natl Acad. Sci. USA*, **86**, 5723–5727.

Pelton, J. G. & Wemmer, D. E. (1990). *Binding modes of distamycin-A with d(CGCAAATTTGCG)2 determined by two-dimensional NMR. J. Am. Chem. Soc.* **112**, 1393–1399.

Phillips, D. C. (1966). *The three-dimensional structure of an enzyme molecule. Sci. Am.* **215** (November), 78–90.

Pohl, F. M. (1976). *Polymorphism of a synthetic DNA in solution. Nature (London)*, **260**, 365–366.

Pohl, F. M. & Jovin, T. M. (1972). *Salt-induced co-operative conformational change of a syhnthetic DNA: equlibrium and kinetic studies with poly(dG-dC). J. Mol. Biol.* **67**, 375–396.

Rice, P. A., Yang, S.-W., Mizuuchi, K. & Nash, H. A. (1996). *Crystal structure of an IHF–DNA complex: a protein-induced DNA U-turn. Cell*, **87**, 1295–1306.

Saenger, W. (1984). *Principles of nucleic acid structure*. New York, Berlin, Heidelberg and Tokyo: Springer-Verlag.

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

Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). *Crystal structure of a CAP–DNA complex: the DNA is bent by 90 degrees. Science*, **253**, 1001–1007.

Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. (1994). *Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. Science*, **266**, 763–770.

Schwartz, T., Rould, M. A., Lowenjaupt, K., Herbert, A. & Rich, A. (1999). *Crystal structure of the Z$\alpha$ domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. Science*, **284**, 1841–1845.

### 23.3 (*cont.*)


Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). *Sequence-specific recognition of double helical nucleic acids by proteins. Proc. Natl Acad. Sci. USA*, **73**, 804–808.

Sklenár, V. & Feigon, J. (1990). *Formation of a stable triplex from a single DNA strand. Nature (London)*, **345**, 836–838.

Sprous, D., Young, M. A. & Beveridge, D. L. (1999). *Molecular dynamics studies of axis bending in d(G$_5$-(GA$_4$T$_4$C)$_2$-C$_5$) and d(G$_5$-(GT$_4$A$_4$C)$_2$-C$_5$): effects of sequence polarity on DNA curvature. J. Mol. Biol.* **285**, 1623–1632.

Sprous, D., Zacharias, W., Wood, Z. A. & Harvey, S. C. (1995). *Dehydrating agents sharply reduce curvature in DNAs containing A-tracts. Nucleic Acids Res.* **23**, 1816–1821.

Sundaralingam, M. (1975). *Principles governing nucleic acid and polynucleotide conformations*. In *Structure and conformation of nucleic acids and protein–nucleic acid interactions*, edited by M. Sundaralingam & S. T. Rao, pp. 487–524. Baltimore: University Park Press.

Thomas, K. A., Smith, G. M., Thomas, T. B. & Feldmann, R. J. (1982). *Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. Proc. Natl Acad. Sci. USA*, **79**, 4843–4847.

Voet, D. & Voet, J. G. (1990). *Biochemistry*. New York: John Wiley & Sons.

Voet, D. & Voet, J. G. (1995). *Biochemistry*, 2nd edition. New York: John Wiley & Sons.

Wahl, M. C. & Sundaralingam, M. (1996). *Crystal structures of A-DNA duplexes. Biopolymers Nucleic Acid Sci.* **44**, 45–63.

Wahl, M. C. & Sundaralingam, M. (1998). *A-DNA duplexes in the crystal*. In *Oxford handbook of nucleic acid structure*, edited by S. Neidle, ch. 5, pp. 117–144. Oxford University Press.

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

Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K. & Wilson, K. S. (1993). *The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. EMBO J.* **12**, 1781–1795.

Wu, H.-M. & Crothers, D. M. (1984). *The locus of sequence-directed and protein-induced DNA bending. Nature (London)*, **308**, 509–513.

Yang, W. & Steitz, T. A. (1995). *Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. Cell*, **82**, 193–207.


### 23.4


Allen, K. N., Bellamacina, C. R., Ding, X., Jeffery, C. J., Mattos, C., Petsko, G. A. & Ringe, D. (1996). *An experimental approach to mapping the binding surfaces of crystalline proteins. J. Phys. Chem.* **100**, 2605–2611.

Badger, J. (1993). *Multiple hydration layers in cubic insulin crystals. Biophys. J.* **65**, 1656–1659.

Baker, E. N. & Hubbard, R. E. (1984). *Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol.* **44**, 97–179.

Beglov, D. & Roux, B. (1997). *An integral equation to describe the solvation of polar molecules in liquid water. J. Phys. Chem.* **101**, 7821–7826.

Bellamacina, C., Mattos, C., Griffith, D., Ivanov, D., Stanton, M., Petsko, G. A. & Ringe, D. (1999). Unpublished results.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *The Protein Data Bank. Nucleic Acids Res.* **28**, 235–242.

Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dall'Acqua, W., Souchon, H., Schwarz, F. P., Maiuzza, R. A. & Poljak, R. J. (1994). *Bound water molecules and conformational stabilization help mediate an antigen–antibody association. Proc. Natl Acad. Sci. USA*, **91**, 1089–1093.

Blake, C. C. F., Pulford, W. C. A. & Artymiuk, P. J. (1983). *X-ray studies of water in crystals of lysozyme. J. Mol. Biol.* **167**, 693–723.

Brooks, C. L. & Karplus, M. (1989). *Solvent effects on protein motion and protein effects on solvent motion. Dynamics of the active site region of lysozyme. J. Mol. Biol.* **208**, 159–181.

Bryant, R. G. (1996). *The dynamics of water–protein interactions. Annu. Rev. Biophys. Biomol. Struct.* **25**, 29–53.

Chervenak, M. C. & Toone, E. J. (1994). *A direct measure of the contribution of solvent reorganization to the enthalpy of ligand binding. J. Am. Chem. Soc.* **116**, 10533–10539.

Clackson, T. & Wells, J. T. (1995). *A hot spot of binding energy in a hormone–receptor interface. Science*, **267**, 383–386.

Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. (1993). *Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature (London)*, **364**, 412–420.

Clore, G. M., Bax, A., Omichinski, J. G. & Gronenborn, A. M. (1994). *Localization of bound water in the solution structure of a complex of the erythroid transcription factor GATA-1 with DNA. Curr. Biol.* **2**, 89–94.

Condon, P. & Royer, W. (1994). *Crystal structure of oxygenated Scapharca dimeric hemoglobin at 1.7 Å resolution. J. Biol. Chem.* **269**, 25259–25267.

Deisenhofer, J. & Steigemann, W. (1975). *Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. Acta Cryst.* B**31**, 238–250.

Edsall, J. T. & McKenzie, H. A. (1978). *Water and proteins. I. The significance and structure of water; its interaction with electrolytes and non-electrolytes. Adv. Biophys.* **10**, 137–207.

Edsall, J. T. & McKenzie, H. A. (1983). *Water and proteins. II. The location and dynamics of water in protein systems and its relation to their stability and properties. Adv. Biophys.* **16**, 53–183.

Gunsteren, W. F. van, Luque, F. J., Timms, D. & Torda, A. E. (1994). *Molecular mechanics in biology: from structure to function, taking account of solvation. Annu. Rev. Biophys. Biomol. Struct.* **23**, 847–863.

Hayward, S., Kitao, A., Hirata, F. & Go, N. (1993). *Effect of solvent on collective motions in globular protein. J. Mol. Biol.* **234**, 1207–1217.

Hendrickson, W. A. & Teeter, M. M. (1981). *Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. Nature (London)*, **290**, 107–113.

Hendsch, Z. S., Jonsson, T., Sauer, R. T. & Tidor, B. (1996). *Protein stabilization by removal of unsatisfied polar groups: computational approaches and experimental tests. Biochemistry*, **35**, 7621–7625.

Hendsch, Z. S. & Tidor, B. (1994). *Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Sci.* **3**, 211–226.

Herron, J. N., Terry, A. H., Johnston, S., He, S.-M., Guddat, L. W., Voss, E. W. & Edmundson, A. B. (1994). *High resolution structures of the 4-4-20 Fab–fluorescein complex in two solvent systems: effects of solvent on structure and antigen-binding affinity. Biophys. J.* **67**, 2167–2175.

Holdgate, G., Tunnicliffe, A., Ward, W. H. J., Weston, S. A., Rosenbrock, G., Barth, P. T., Taylor, I. W. F., Pauptit, R. A. & Timms, D. (1997). *The entropic penalty of ordered water accounts for weaker binding of the antibiotic Novobiocin to a resistant mutant of DNA gyrase: a thermodynamic and crystallographic study. Biochemistry*, **36**, 9663–9673.

Hubbard, S. J., Gross, K.-H. & Argos, P. (1994). *Intramolecular cavities in globular proteins. Protein Eng.* **7**, 613–626.

Jiang, J.-S. & Brünger, A. (1994). *Protein hydration observed by X-ray diffraction. Solvation properties of penicillopepsin and neuraminidase crystal structures. J. Mol. Biol.* **243**, 100–115.

Karplus, P. A. & Faerman, C. (1994). *Ordered water in macromolecular structure. Curr. Opin. Struct. Biol.* **4**, 770–776.

Kauzmann, W. (1959). *Some factors in the interpretation of protein denaturation. Adv. Protein Chem.* **14**, 1–63.

Kendrew, J. C. (1963). *Myoglobin and the structure of proteins. Science*, **139**, 1259–1266.

Komives, E. A., Lougheed, J. C., Liu, K., Sugio, S., Zhang, Z., Petsko, G. A. & Ringe, D. (1995). *The structural basis for pseudoreversion of the E165D lesion by the secondary S96P*