From: "Structure Correlation" (H.-B. Bürgi & J.D. Dunitz, Eds.) Vol. 2, (1994) VCH Publishers, Inc., Weinheim, Germany, pp. 705-749.

18 Structural Patterns in Nucleic Acids

Martin Egli

18.1 Introduction

Over the last ten years, our understanding of the three-dimensional structure of DNA has greatly improved. The most important reason for the accumulation of structural data is the automatization of oligonucleotide synthesis, which has resulted in the availability of milligramm amounts of pure material at relatively low costs. The subsequent crystallization of DNA fragments and the determination of their structures by single-crystal X-ray diffraction has allowed us to visualize the double-helical conformations in great detail and has led to a refined understanding of DNA structure, previously based on the results of fiber diffraction, crystal structure analysis of mono- and dinucleotides, spectroscopic methods, and various solution techniques. However, the crystallization and structure solution of oligonucleotides is still a difficult and slow business and, unlike the small molecule field, where some 80,000 single crystal structures are deposited in the Cambridge Structural Database, only a few dozen oligonucleotide crystal structures have been solved so far. Although about 25 crystal structures have been reported for each of the conformational families of double helical oligonucleotides, the number of distinct crystal lattices, in which these fragments crystallize, is much smaller. The often unsuccessful, frustrating search for crystallization conditions has led to a strategy, where one attempts to maintain certain features of a fragment already known to crystallize (e.g. similar length, sequence etc.), rather than trying to crystallize a completely new fragment. The hope is that the new fragment will crystallize in the same lattice as the previous one. This strategy has been successfully applied in more than twenty cases with the so-called Dickerson B-DNA dodecamer sequence. Although they have the same lattice, these crystal structures have had an important impact on our understanding of the sequence dependence of the helix geometry (e.g. narrowing of the minor groove in AT-rich stretches), and have provided information about the binding mode of minor-groove binders, the geometry of mismatched base pairs, and other phenomena. Similarly, the crystallization of the left-handed hexamer duplex $[d(CGCGCG)]_2$ under various conditions in almost always the same lattice has allowed the investigation of the binding mode of ions to DNA and their influence on the Z-DNA conformation.

The relative success in the elucidation of DNA structure by X-ray crystallographic methods is not paralleled by results from structural work on RNA. Whereas improved synthetic methods have paved the way to detailed knowledge of DNA structure, the difficulties encountered during RNA synthesis are mainly responsible for the lack of more information about RNA secondary structure. Although, in contrast to DNA, the conformation of double-helical RNA is insensitive to base composition and sequence, the additional 2′-hydroxyl group in RNA leads to a rich variety in secondary structure as, for example, in the ribozymes. At present, the high resolution crystal structures of a few transfer RNAs are the only source of exact information on secondary structure of complex RNA molecules. Improvements in RNA synthesis, making large amounts of pure RNA available, are one important step on the way to a greater number of secondary structure determinations of biologically interesting RNA fragments.

The following sections will give a summary of some recent insights into the conformation of the nucleic acids. Obviously, it is impossible to give a full account of the subject; the interested reader may consult various review articles cited in this contribution. Much like a whale, taking in great mouthfuls and straining the intake through its baleen, I have plunged through many publications and data and retained only what I judge to be relevant for a book on structure correlation. The chosen examples are mostly from structures of antiparallel double-helical oligonucleotides, and almost all data used originate from X-ray crystallographic studies.

18.2 Right-handed and Left-handed Double Helices

The variety of conformations adopted by double-helical nucleic acids has been the focus of a number of review articles, describing the results from both fiber-diffraction analyses and single-crystal X-ray crystallography [1 – 5]. Table 18.1 lists the geometrical parameters characteristic for each of the three conformational duplex families now commonly differentiated. This table gives the mean values found in each of the three families and does not fully account for the conformational variety found within each of them. The properties listed are the ones which are observed most frequently; in the case of the sugar puckers, for example, one also finds other modes which are conformationally and energetically similar to the listed ones. Figure 18.1 shows van der Waals models of A-DNA (A) and B-DNA (B), and Figure 18.2 shows a stereo picture of Z-DNA. The most striking difference between A- and B-DNA and Z-DNA is that the former two are right-handed double helices, whereas the Z double helix is left-handed. The B-DNA helix is slightly slimmer than the A-DNA helix and has a diameter of about 20 Å. In B-DNA, the arrangement of the sugar-phosphate backbones generates a wide or major groove and a narrow or minor groove, winding around the helix axis. An axial projection reveals that the bases are stacked on the inside of the duplex, in close proximity to the helical axis, leaving

Table 18.1. Selected structural parameters for A-, B- and Z-type duplexes

Parameter	A	B	Z
helical sense	right-handed	right-handed	left-handed
repeated unit	1 base pair	1 base pair	2 base pairs
repeats per turn (av.)	11	10	12
axial rise per repeat (av.)	2.4 – 3.3 Å	3.3 – 3.5 Å	3.5 – 4.0 Å
helical twist angle (av.)	33°	36°	−30° −9° (CpG step) −51° (GpC step)
propeller twist angle (av.)	10 – 18°	8 – 16°	2 – 5°
glycosidic angle	*anti*	*anti*	*syn* (G) *anti* (C)
sugar pucker	C(3′)-*endo*	C(2′)-*endo*	C(3′)-*endo* (G) C(2′)-*endo* (C)
references	[6 – 13]	[14, 15]	[16 – 18]

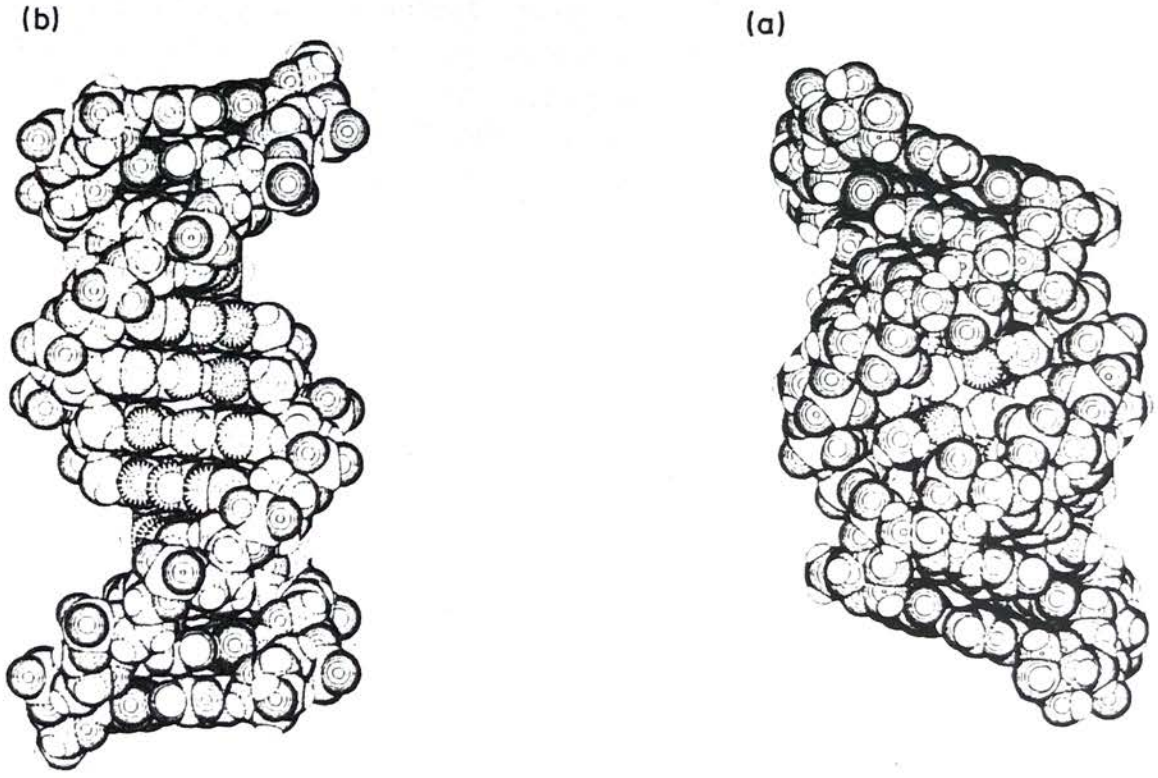

Fig. 18.1. Van der Waals diagrams, a: A-DNA; b: B-DNA. Both duplexes are dodecamer fragments with the sequence $[d(CGCGAATTCGCG)]_2$ and idealized conformations. The drawings illustrate the differences between the geometries of the major grooves (base atoms dashed) and the minor grooves of the two conformation families. Whereas the base pairs are tilted with respect to the helix axis in A-DNA, they are practically normal to the helix axis in B-DNA. Carbon atoms are black, nitrogen atoms are blue, oxygen atoms are red, phosphorus atoms are yellow and hydrogen atoms are white

no empty space on the inside. The phosphate groups are prominently exposed on the surface of the duplex. On the other hand, A-DNA shows a coil-like arrangement, hollow on the inside. Again one finds two different grooves, but the major groove

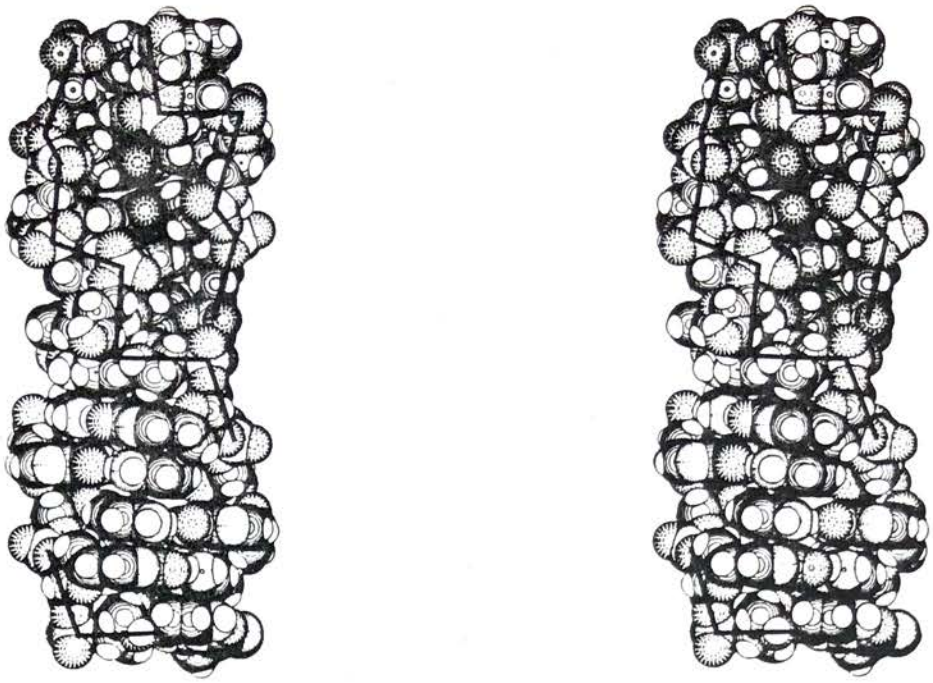

Fig. 18.2. Stereo drawing of left-handed Z-DNA. The dodecamer consists of two stacked duplexes with sequence d(CGCGCG). The base atoms in the deep minor groove are shaded in grey. The convex surface replacing the major groove in B-DNA runs from the upper left to the lower right with some base atoms lying on the surface of the duplex. Phosphate groups (including those which are missing at the ends of the hexamers) are traced with black lines to emphasize the dinucleotide repeat of the backbone in Z-DNA. Carbon atoms are concentric circles, nitrogen atoms and phosphorus atoms are dotted, oxygen atoms are dashed and calculated hydrogen positions are blank

in A-DNA is narrower than the one in B-DNA and also much deeper. The major groove is about three times as deep as the minor groove, which is now shallow and widened up compared with B-DNA. In both duplex types, the functional groups of the bases lie at the floor of the grooves. Molecules can readily interact with those groups via the major groove of B-DNA or the minor groove of A-DNA, but interactions via the minor groove of B-DNA or the major groove of A-DNA are very limited; this is because in the former case the cleft, bordered by the phosphate groups, is only about 4 to 6 Å wide, and in the latter because the groove is too deep for interacting molecules to form contacts to base atoms. Another characteristic difference between A-DNA and B-DNA is the orientation of the stacked bases with respect to the helix axis. In A-DNA the bases are tilted relative to the axis, whereas in B-DNA they are more or less normal to it. In both duplex types, the planes of base pairs are often rotated relative to each other around an axis parallel to their hydrogen-bond direction, a distortion referred to as propeller twist. In B-DNA, it is possible to connect the phosphorus atoms by a smooth curve and the same is more or less true for A-DNA as well. The conformation of the backbone of an oligonucleotide is described by six torsion angles α to ζ for each of its nucleotides, defined as follows: O(3′)−P−(α)−O(5′)−β−C(5′)−(γ)−C(4′)−(δ)−C(3′)−(ε)−O(3′)−ζ−P−O(5′) (Figure 18.3a). Often, one uses a qualitative description, based on the conformational ranges *gauche*$^+$ (g^+, also $+sc$), *gauche*$^-$ (g^-, also $-sc$) and

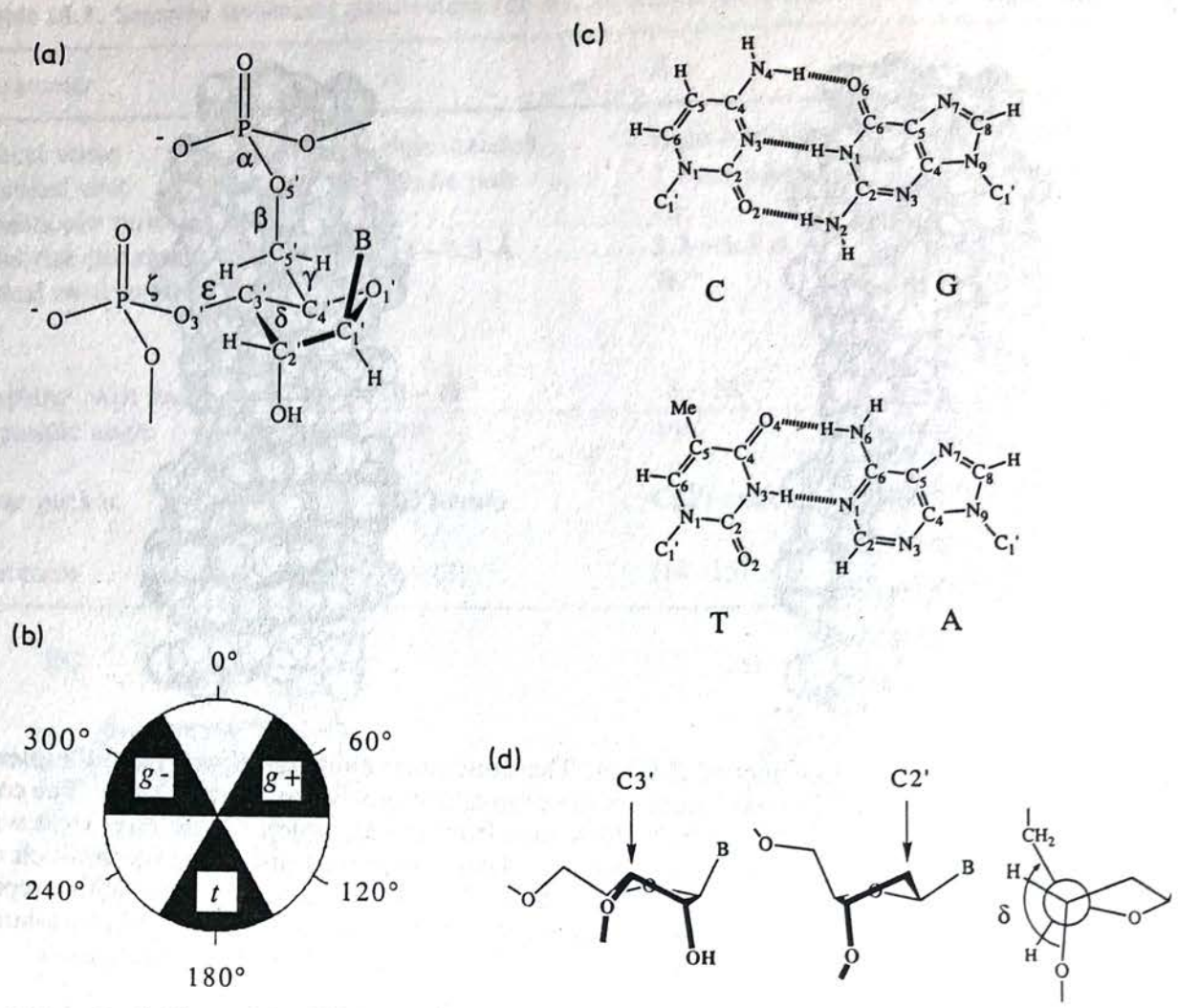

Fig. 18.3. a: Definition of backbone torsion angles α to ζ and standard numbering for ribose atoms. The ribose shown adopts a C(3′)-*endo* pucker; purine or pyrimidine bases are designated with a capital B. Hydrogens at carbon atoms C(3′) and C(4′) have been omitted for clarity.
b: Conformational ranges of torsion angles.
c: Standard atomic numbering for purines adenine and guanine, as well as pyrimidines thymine and cytosine, shown as the classical Watson-Crick G·C and A·T base pairs. Uracil lacks the methyl group at the thymine C(5) position.
d: C(3′)-*endo* and C(2′)-*endo* sugar puckers (left [ribose] and middle [deoxyribose]), and Newman projection of the sugar in a C(3′)-*endo*-conformation along the C(3′)–C(4′) bond, demonstrating the restriction of torsion angle δ to a maximum value of around 160° as a consequence of ring flattening in furanoses (right)

trans (*t*, also *ap*) of the individual torsion angles (Figure 18.3 B). Somewhat simplified, the most common backbone genus in the case of A-DNA has the conformation ($g^- tg^+ g^+ tg^-$), and in the case of B-DNA the backbone conformation is ($g^- tg^+ ttg^-$) (see Section 18.5.1 for a more detailed description). This is quite different from the situation in left-handed Z-DNA, where the corresponding angles are different for purine and pyrimidine residues. Because of this, and due to the dinucleotide repeat unit in Z-DNA, the backbone conformation cannot be described by an overall formula as in the other cases. As illustrated in Figure 18.2, the phosphate groups of Z-DNA can be connected with a zigzag curve, a characteristic which gave the left-handed form its name. With a diameter of about 18 Å, the Z duplex

is slimmer than the right-handed duplexes. The molecule looks like a cylinder (Figure 18.2) and has an internal symmetry which is different from A- or B-DNA, since the helix has six units per turn, each unit consisting of two base pairs (Table 18.1). In Z-DNA, the minor groove is considerably deeper than in B-DNA. Whereas the major groove in B-DNA is a concave surface with the base pairs lying at the bottom, the analogous part of Z-DNA is no longer a groove but a convex surface with the guanine bases lying close to the surface (Figure 18.2). A projection along the helical axis would reveal the guanine residues in Z-DNA to be quite distant from the helical axis, with the N(7) and C(8) atoms of the imidazole part of guanine exposed and accessible to molecules or ions interacting with the DNA (see Figure 18.3C for the atomic numbering of bases). Although B-DNA is probably the dominant form in biological systems, detailed structural results from X-ray diffraction studies at high resolution were not available until recently [15, 19]. On the other hand, crystals of left-handed fragments often diffract X-rays to atomic resolution, revealing all the details of the duplex geometry as well as the arrangement of ions and solvent molecules.

18.3 Variations in the Duplex Geometry

The differences in geometry between individual double-helical fragments adopting the same overall type of conformation (A, B or Z) provide information on the flexibility of that duplex and on the influence of specific parameters on the helix geometry (e.g. sequence, counter-ions, crystal packing). Because the number of crystal structures is still quite small, only a limited analysis of the influence of the sequence on the duplex conformation is possible. A complete understanding of the sequence dependence of the structure would require the investigation of different sequences in a similar or analogous crystallographic environment (partially fulfilled for A- and B-type DNA), as well as the study of the same base stretch in different crystallographic environments (structural data not available). The B-DNA dodecamers with sequence $CGC(x)_6GCG$ [14, 20–24], not including structures with base-pair mismatches or DNA-drug complexes, and decamers with sequence $CC(x)_6GG$ [15, 25, 26] represent groups of structures which partly satisfy the first requirement ($(x)_6$ are various self-complementary sequences). For A-DNA, the situation is somewhat similar. There, the largest group of crystal structures comprises octamer duplexes with various sequences that crystallize in essentially two distinct crystallographic environments. DNA sequences with purine and pyrimidine residues alternating (5′ to 3′ direction) seem to prefer the B conformation ([1]; for an example, see [27]), and most of the DNA sequences which show stretches of two or more purines alternating with stretches of two or more pyrimidines (independent of whether the 5′-end residue is purine or pyrimidine) adopt the A conformation in crystal structures. However, in right-handed helices which display a large variety of

conformational patterns, the adopted conformation does not always appear to be related to the sequence in a simple way. In the case of left-handed Z-DNA, most of the crystallographic studies used alternating pyrimidine-purine sequences of the type 5′-d$(CG)_n$ (in some cases TA steps were inserted within the $(CG)_n$ stretches). This lack of sequence variety is probably the main reason for the relatively small deviations of the duplex geometry between different crystal structures of left-handed fragments.

18.3.1 The Correlations between Backbone Torsion Angle δ and Glycosidic Torsion Angle χ in A-, B- and Z-DNA

In nucleotides and oligonucleotides, the torsion angle δ is defined by C(5′)–C(4′)–C(3′)–O(3′) (Figure 18.3a) and thus provides information about the ribose pucker mode. It is g^+ (or lying within that conformation range) for a C(3′)-*endo* conformation of the sugar (with C(3′) lying out of the plane formed by the other ring atoms, on the same side as C(5′) and the base, Figure 18.3D). It is *t* (or lying within that conformation range) for a C(2′)-*endo* conformation (with C(2′) lying out of the plane formed by the other ring atoms, on the same side as C(5′) and the base). These two ring conformations are the most frequently observed ones in oligonucleotides. The others are geometrically closely related and have similar stabilities (e.g. conformations C(3′)-*endo* and C(4′)-*exo* or C(2′)-*endo* and C(1′)-*exo*; for the relative stability of different sugar conformations, see [28, 29 (pp. 61–64)]). However, duplex RNA adopts exclusively the A conformation, which is associated with a C(3′)-*endo* pucker of the riboses [1, 31, 32, 33]. The puckering preference may be a consequence of the steric hindrance caused by the additional O(2′) oxygen in RNA [29 (pp. 64–68)]. The A conformation is also found in homo DNA-homo RNA hybrids [34], and self-complementary DNA-RNA hybrids with different numbers of RNA residues and varying positions of the RNA residues within the chimeric RNA-DNA strands [35, 36]. Finally, a DNA-RNA hybrid composed of the DNA strand d(GGGTATACGC) and the chimeric strand r(GCG)d(TATACCC) was also found to adopt the A conformation [37]. Such segments of hybrid RNA/DNA plus duplex DNA are referred to as Okazaki fragments. They are a consequence of the inability of DNA polymerases to initiate new chains de novo. During replication, one of the strands, the so-called lagging strand, has to be replicated in a discontinuous way. Okazaki fragments are formed when RNA primers, synthesized by RNA polymerase, are laid down by a primase and then elongated by the DNA polymerase, which requires a 3′-OH terminus of a primer oligo- or polynucleotide to initiate chain growth.

The other important torsion angle is χ around the glycosidic bond between the C(1′) sugar carbon and the nitrogen atom of the attached base. The glycosidic angle χ is defined by O(1′)–C(1′)–N(9)–C(4) in purines, and by O(1′)–C(1′)–N(1)–C(2) in pyrimidines, respectively Figure 18.3c. In the *syn* conformation, the sugar is rotated into the minor groove and in the *anti* conformation, it is rotated away. For purines,

anti is slightly favored over *syn*, which leads to close contacts between base nitrogen N(3) and the sugar atoms. However, for pyrimidines, the *syn* arrangement is disfavored because of steric hindrance between base oxygen O(2) and the sugar. Whereas the observed values of χ for *syn* conformations are in the range between +50 and +80°, they cover a large range of values between −80 and −180° for *anti* conformations.

Dickerson and colleagues investigated the correlation of δ and χ for duplex DNA (reviewed in [5]). Although only a few structures were available at the time, some of the observations are still valid. For the B-DNA dodecamer, a linear correlation between δ and χ was found to hold over a fairly broad distribution of conformations with purines displaying systematically higher δ and χ values than pyrimidines. On the contrary, the plot for the A-DNA tetramer showed clustering around the point defined by the ideal A structure from fiber diffraction. Although subsequent analyses of octamers show significant differences among each other [7, 10], they all resemble the ideal A-DNA structure derived from fiber diffraction experiments. The same is true for the A-DNA decamer [11], and the recent crystal structure of a complete turn of A-DNA [13]. However, Z-DNA shows quite different correlation of δ and χ. The most striking difference to the right-handed duplexes is the alteration of the glycosidic torsion, between *syn* in guanines and *anti* in cytosines (see Table 18.1). This alteration produces a concomitant rotation of adjacent bases into and away from the minor groove (see Section 18.3.4), resulting in the characteristic zigzag pattern found in Z-DNA. The *syn* conformation of guanine has also been found in the crystal structure of the G-rich strands of telomers which constitute the ends of chromosomes [30]. Here, the conformation of guanines in the four-stranded helix, adopted by two d(GGGGTTTTGGGG) strands, alters between *syn* and *anti*.

18.3.2 Influence of the Sequence on the Conformation of B-DNA

Local variations in the geometry of duplexes can be described with a number of helical parameters (for a summary see [38]). Some of these, like backbone torsion angles, have already been introduced and others are mentioned in Table 18.1. Some characterize the relative positions and orientations of bases within base pairs (shear, buckle, propeller twist etc., see Figure 18.4a) and others deal with the positions and orientations of the base pairs themselves (rise, displacements, roll, twist etc., see Figure 18.4b). Unaveraged parameters for individual base-pair steps (e.g. propeller twist) only became accessible through X-ray single-crystal structures. These data also allowed an investigation of the influence of the base sequence on the local geometry. One of the earliest interpretations of the observed variations of the helix geometry in the DNA dodecamer $[d(CGCGAATTCGCG)]_2$ was made by Calladine [39]. His qualitative model could explain individual observed twists, torsion angles δ and χ, as well as base pair roll angles, in terms of steric hindrance caused by adjacent purine bases of opposite strands. This clash is the consequence of the propeller twist between bases in base pairs which leads to improved stacking interactions. In so-called Y–R steps (Y = pyrimidine, R = purine) the steric hindrance occurs in the minor

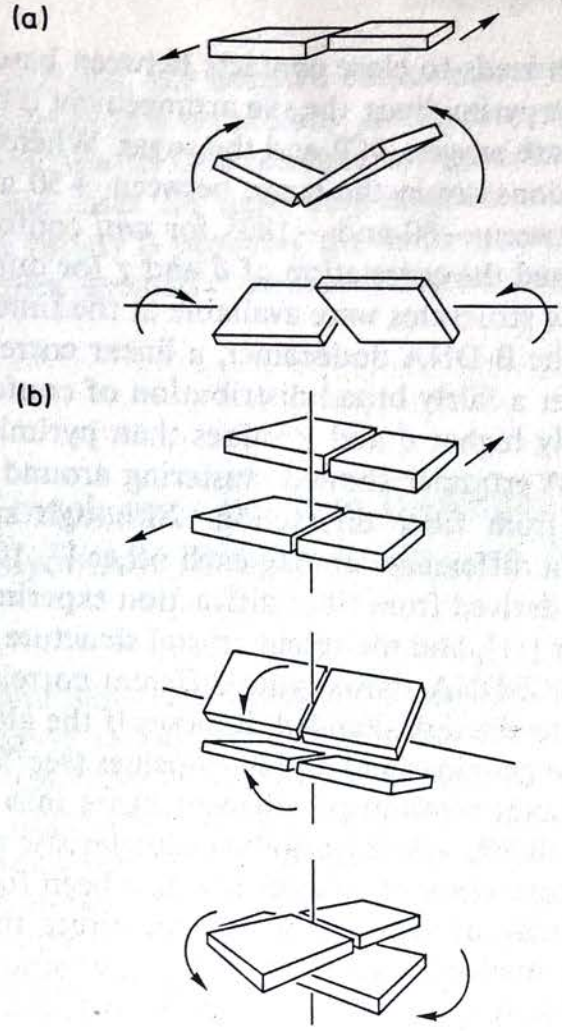

Fig. 18.4. a: Schematic representations of selected helical parameters, characterizing the relative positions and orientations of bases within base pairs. From top: Shear, buckle and propeller twist. b: Schematic representations of selected helical parameters, characterizing the relative positions and orientations of base pairs. From top: Displacement, roll and helical twist

groove, in R–Y steps it occurs in the major groove, and in R–R or Y–Y steps no hindrance is present. Calladine also assumed that the above hindrance is more severe in the minor groove than in the major groove. Dickerson then quantified the above model, characterizing the steric hindrances with simple additive terms, and fitting the resulting sum functions to the actual experimental parameters in A- and B-DNA [40]. Figure 18.5a depicts the agreement between observed helix-twist angles and those expected from the sum function $\sum 1$ for the above dodecamer. Figure 18.5B gives the same information for observed roll angles and those expected from the corresponding sum function $\sum 2$ (for details see figure caption). Although such approaches produce satisfactory explanations for a number of helical parameters observed in the B-DNA dodecamers and some A-DNA sequences, they cannot be considered as structural algorithms with a general meaning. Moreover, their application was restricted to sequences with several purine-pyrimidine or pyrimidine-purine steps, so that variations in the local geometry of a stretch of purines and a stretch of pyrimidines were left unexplained. A recent analysis of three B-DNA decamers

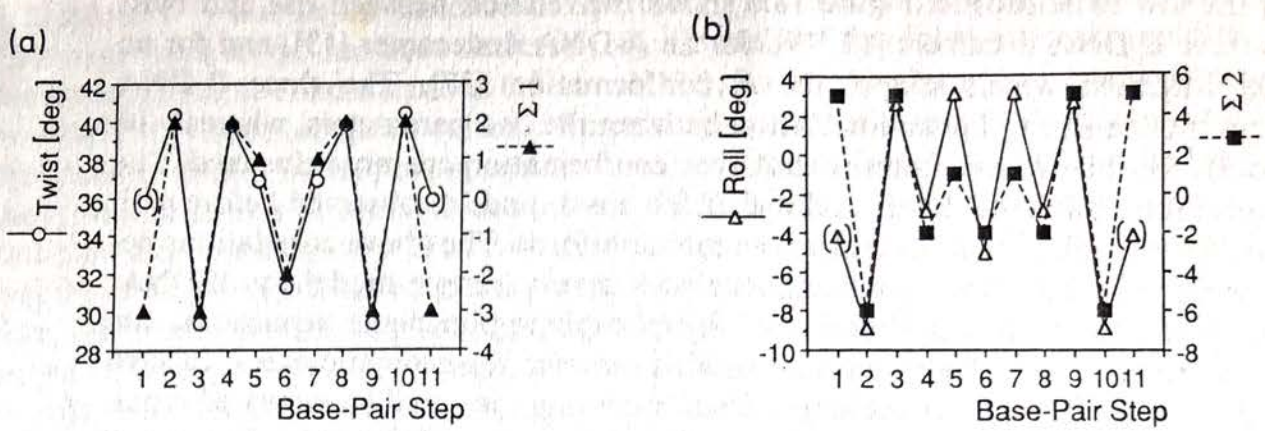

Fig. 18.5. Correlation between observed helix-twist angles and sum function $\sum 1$ (a) and correlation between observed roll angles and sum function $\sum 2$ (b) for the dodecamer $[d(CGCGAATTCGCG)]_2$. A decrease in helical twist or an opening of the roll angle on the side of the base pair where a clash occurs will relieve the steric hindrance between adjacent purines of opposite strands. For example, the clash at the first G–C base pair step (R–Y, step 2) takes place in the major groove. To relieve the steric hindrance, the twist angle has to be decreased, and at the same time the twist at each of the two neighboring steps is increased by half that amount. Thus, the contributions to $\sum 1$ for the three steps 1, 2 and 3 are +1, −2, +1. If the clash occurs in the minor groove (e.g. with the neighboring C–G step, step 3), each of the contributions has to be doubled. Thus, the contributions to $\sum 1$ for the three steps 2, 3 and 4 are +2, −4, +2 etc. (for details, see [39, 40]). The correlation coefficients were 0.94 (a) and 0.92 (b). For the correlations, the contributions of terminal base-pair steps (1 and 11) were omitted to avoid the influence of end effects

has shown that the local helix structure should be studied in the context of four base pairs rather than just two (Y–R and R–Y, see above) [19]. The authors concluded that it is probably too early to develop a reliable algorithm to explain and predict the influence of the sequence on the local helix geometry, because there are 256 different four-base steps and structural data is available only for 33 of them so far.

18.3.3 Other Correlations

The above analysis of various B-DNA dodecamer and the high-resolution B-DNA decamer structures also showed the importance of the geometrical constraints due to the sugar-phosphate backbone [19]. The three parameters twist, rise and cup are strongly correlated, a fourth parameter, roll, only weakly. (The parameter cup is defined as the change in buckle between two successive base pairs. Thus, if cup is positive, or if a negative buckle follows a positive one, then the two base pairs are bent toward one another like two cupped hands; compare Figure 18.4a.) Two extreme patterns exist. In one of these, the so-called high twist profile (HTP), high twist is combined with low rise, high cup and low roll. In the other, the so-called low twist profile (LTP), low twist is combined with high rise, low cup and high roll. The former class includes GpC and GpA base steps, whereas CpG, GpG and ApG base steps

display the low twist profile. Figure 18.6 shows the relation between rise and twist for the three B-DNA decamers [15, 19], for an A-DNA dodecamer [13], and for an Okazaki fragment which adopts the A conformation [37]. The three B-DNA decamers show a strong linear correlation between the two parameters, whereas the data points for the two structures with A type conformation are more clustered. The A conformation allows a closer stacking of the bases, and, as observed before with the δ-χ correlations, A-type structures are more uniform. The above correlations between helical rise and twist are consistent with a very simple model for B-DNA, whereby electrostatic forces (dominated by phosphate-phosphate repulsions) are competing with attractive stacking interactions between the aromatic bases. In such a model, the helical twists are prevented from becoming too small by increased repulsions between phosphates. These would eventually be relieved through adjustments in the backbone torsion angles, leading to bigger helical rises.

18.3.4 Variations in the Conformation of Z-DNA

Among the crystal structures of the left-handed hexamer duplex $d[(CGCGCG)]_2$, one can differentiate between three crystal forms. Two of them require crystal growth in the presence of varying concentrations of polyvalent inorganic cations (e.g. Mg^{2+}). The "mixed magnesium/spermine form" (the original Z-DNA hexamer) is obtained by crystallizing d(CGCGCG) in the presence of magnesium chloride and spermine tetrachloride (both around 20 $mmol \cdot l^{-1}$) [16]. The "magnesium form" crystallizes in the presence of high magnesium chloride concentrations

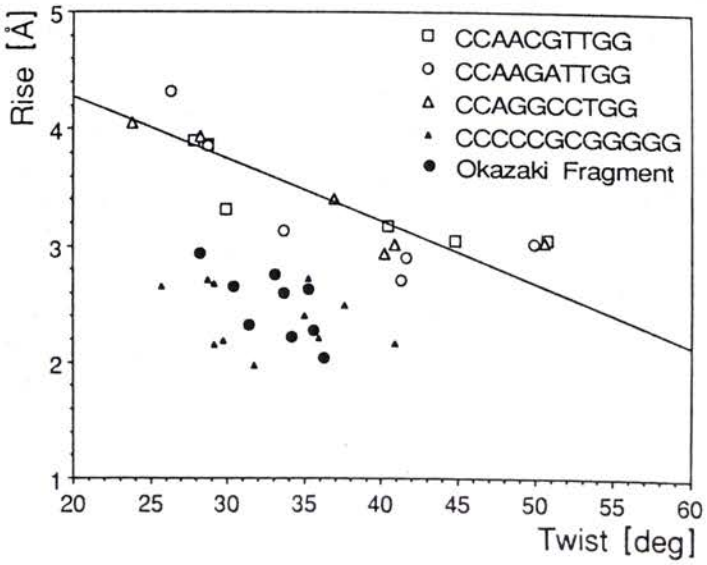

Fig. 18.6. Correlation between the helical parameters rise and twist. The correlation coefficient for the three B-DNA decamers (open symbols) is −0.83 [19]. The dodecamer and the Okazaki fragment both adopt the A conformation. The Okazaki fragment consists of the DNA sequence d(GGGTATACGC) and the chimeric RNA-DNA sequence r(GCG)d(TATACCC)

(120 mmol·l^{-1}) but in the absence of spermine [17]. Recently, a third form, the "pure-spermine form" has been detected [18]. Crystals of this form were grown with moderate spermine concentrations (15 mmol·l^{-1}), but in the absence of magnesium or other polyvalent inorganic cations. Surprisingly, the position and orientation of the hexamer duplex in the pure-spermine form is different from those determined in the crystal structures of over a dozen other Z-DNA hexamers with varying sequences and chemical modifications. However, the pure-spermine form is unique in the sense that the crystals were obtained in the absence of di- or polyvalent metal ions. In addition, the crystal morphology of the pure-spermine form differs from the one observed with $[d(CGCGCG)]_2$ crystals grown in the presence of polyvalent inorganic cations. Whereas mixed magnesium/spermine-form or magnesium form-crystals are typically of quasi-hexagonal shape, crystals of the pure-spermine form grow as rectangular plates.

The conformation of the Z-DNA duplex in the pure-spermine form is different from the conformations in the magnesium-containing forms, which – as mentioned in the introduction to Section 18.3 – do not vary much from one crystal structure to another. The pure-spermine form thus suggests that the left-handed form of DNA is conformationally not as rigid as had been assumed before. A comparison with the other hexamers shows that the DNA in the pure-spermine form is compressed along the helical axis [18]. Thus, a nominally complete turn of Z-DNA based on the average helical rise of the pure-spermine form would be 1.43 Å and 2.42 Å shorter than the complete turns based on the average rises of the magnesium form and the mixed magnesium/spermine form, respectively. The helical rise in the pure-spermine form varies much less than in the two other forms. Compared with the magnesium-containing forms, the DNA in the pure-spermine form is slightly overwound, the average twist angles are thus somewhat bigger. Another difference to the conformations in the magnesium-containing forms is the shift of base pairs into the major groove, away from the helical axis. This results in an altered position of the helix axis, which is now in the minor groove, instead of passing through the O(2) oxygens of cytosines. The different shifts into the major groove of individual base pairs account for altered stacking interactions between the bases in the pure-spermine form and might be the reason for the more compact conformation of the duplex. Figure 18.7 shows selected phosphorus-phosphorus distances for the three different forms of the Z-DNA hexamer. The intrastrand distances shown in chart A are distances between the phosphorus atoms of guanidine residues. In the drawing of left-handed DNA (Figure 18.2), they are the vertical components of the zigzag lines connecting the phosphate groups. Thus, these distances are a measure for variations in the axial dimensions of the left-handed duplexes. The more compact conformation of the pure-spermine form is indicated by the shorter distances between the phosphorus atoms in that form (distance P(10)–P(12) being the exception), compared with those found in the magnesium form and the mixed magnesium/spermine form. The average intrastrand phosphorus-phosphorus distance for the two strands in the pure-spermine form is 0.58 Å shorter than the corresponding average distances in the magnesium-containing forms. Similarly, the interstrand phosphorus-phosphorus distances in the pure-spermine form duplex are consistently shorter than the corresponding distances in the other two forms (Figure 18.7b). These distances are a measure for the width of

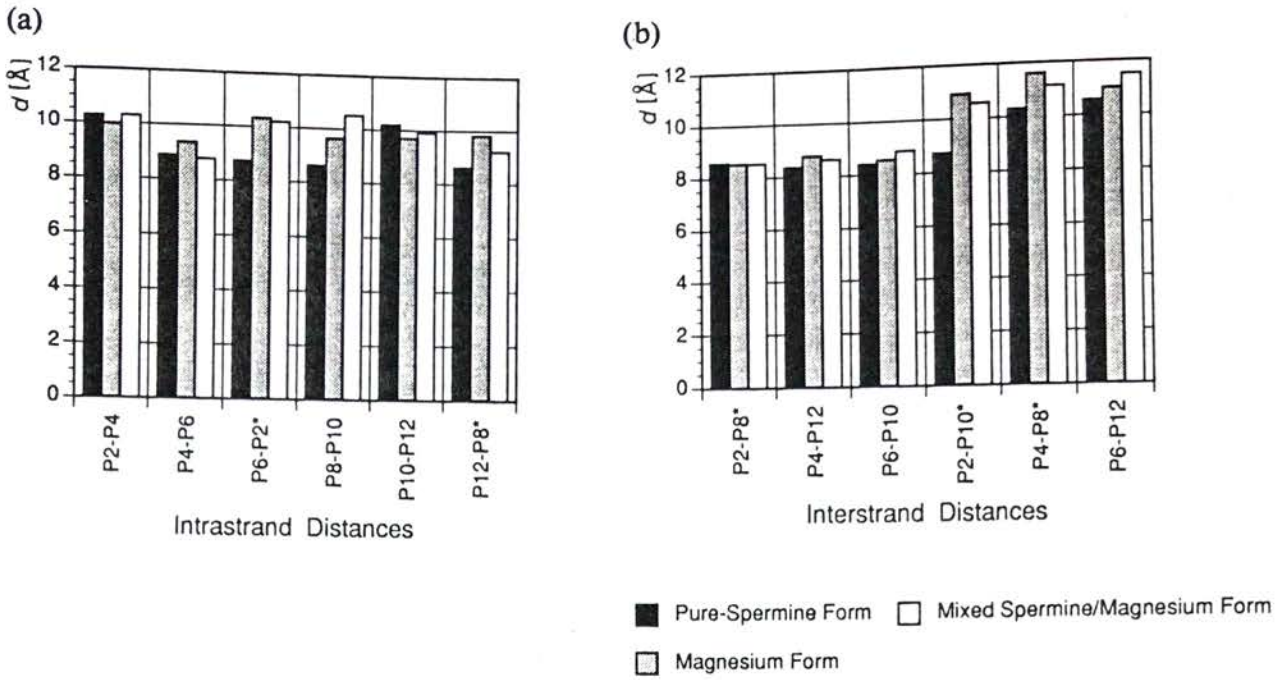

Fig. 18.7. Distances between phosphorus atoms of adjacent guanine residues within each of the two strands (a) and between phosphorus atoms of guanine residues in different strands, defining the width of the minor groove (b), in three crystal forms of the left-handed hexamer duplex $[d(CGCGCG)]_2$. The phosphorus atoms of one strand are numbered 2 to 6 (starting with P(2) of residue G2), and the phosphorus atoms of the complementary strand are numbered 8 to 12 (starting with P(8) of residue G8). Symmetry equivalent atoms are marked with asterisks. In the three crystal forms, the stacking of duplexes in a 3′–5′ and 5′–3′ manner leads to the formation of infinite helices with a continuous minor groove

the minor groove, which is narrower than in the other two forms (the average difference between the interstrand distances is 0.70–0.75 Å).

Near the floor of the minor grooves in Z-DNA duplexes, water molecules connect the O(2) oxygens of cytosine residues from the two strands [41, 42]. This partially buried spine of hydration is also observed in the room-temperature structure of the pure-spermine form, but an analysis of the structure at low temperature revealed the presence of a second spermine molecule in the minor groove [43]. This spermine molecule forms hydrogen bonds to the O(2) oxygens of four cytosines, replacing and displacing some of the water molecules normally bound to these atoms in that groove. Although the spermine molecule was not observed in the room-temperature structure, it is nevertheless present. Presumably, it does not occupy a definite position. Because of the crucial influence of hydration on DNA structure in general, e.g. [44] and [45 (pp. 500–504)], and specifically in the case of Z-DNA with its extended network of water molecules in the minor groove, one expects the duplex geometry to be altered upon binding of a spermine molecule in that groove. Both terminal ammonium nitrogens of the spermine molecule are hydrogen bonded to backbone phosphate oxygens from the inside of the minor groove. It is therefore not surprising that the minor groove of the pure-spermine form is slightly narrower than those of the other two forms.

In Z-DNA, all the CpG steps show very similar conformations of the backbone. This may be associated with the shearing of the two bases relative to each other at these steps, resulting in stacking interactions between adjacent cytosines of opposite strands. On the other hand, at GpC steps the backbone can adopt two conformations, referred to as Z_I and Z_{II} [46]. Figure 18.8 depicts two backbone segments of the Z-DNA hexamer in the magnesium form [17]. The phosphodiester linkage adopts a $-sc, ap$ conformation in Z_I (Figure 18.8a) and a $+sc, ap$ conformation in Z_{II} (Figure 18.8b). The torsion angles that are different in the two conformations are ε and ζ. In the Z_I conformation they are $-sc$, $-sc$ (residue G2 in Figure 18.8a), and in the Z_{II} conformation they are ap, $+sc$ (residue G4 in Figure 18.8b). In Figure 18.8, the views are into the minor groove, and a comparison between the two conformations shows that the phosphate group in the Z_{II} conformation is turned away from the minor groove. The line connecting the phosphorus atoms in a Z-DNA duplex with all the GpC steps adopting the Z_{II} conformation would therefore show an accentuated zigzag pattern [46]. The altered backbone conformation may be associated with the binding of a magnesium ion to the N(7) nitrogen of residue G6 (Figure 18.8b). One of its coordinated water molecules is hydrogen-bonded to the phosphate oxygen of P(5) and could thus stabilize the Z_{II} conformation at the G4pC5 step. This arrangement is found in both the mixed magnesium/spermine form and the magnesium form of $[d(CGCGCG)]_2$, as well as in the structure of the Z-DNA hexamer in the presence of cobalt hexammine [47]. In that structure, the

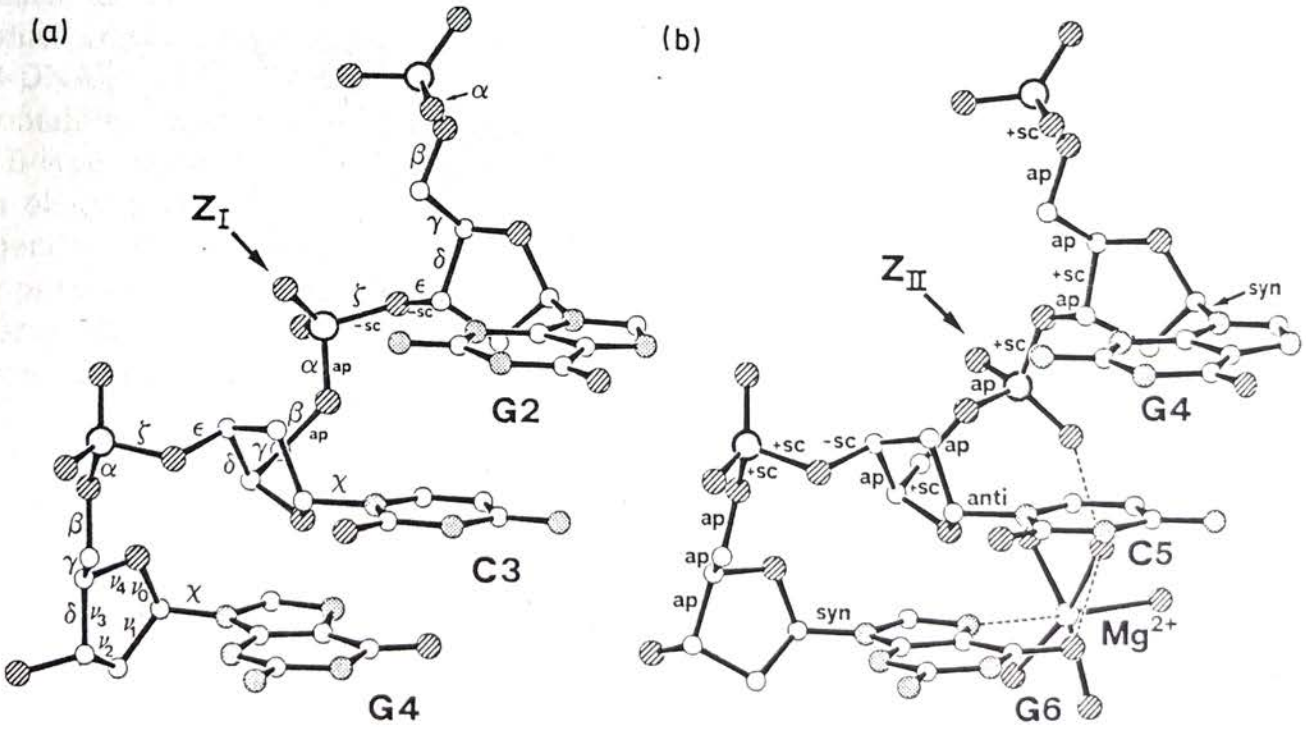

Fig. 18.8. a: Z_I conformation of the backbone in the magnesium form of the left-handed hexamer $[d(CGCGCG)]_2$ at the G2pC3 step.
b: Z_{II} conformation of the backbone in the same structure at the step G4pC5. In the drawings, the symbols and ranges for the backbone torsion angles as well as those for the glycosidic torsion angles have been included. The adoption of the Z_{II} conformation at the G4pC5 step is associated with the binding of a penta-hydrated magnesium ion to nitrogen N(7) of residue G6

$[Co(NH_3)_6]^{3+}$ ion replaces a second magnesium hexahydrate ion, bound to the above magnesium ion via its first solvent sphere (not included in Figure 18.8b). The cobalt ion makes five contacts to the DNA and thus presumably stabilizes the left-handed conformation. Although the presence of metal ions and the associated hydrogen bonds to the DNA backbone might be crucial for the alteration of the backbone conformation, the Z_{II} conformation at the G4pC5 step was also found in the presence of barium ions and spermine, where no ion is coordinated to residue G6 [46]. Thus, ion stabilization can be only one factor contributing to the stabilization of the Z_{II} conformation. This finding is supported by the backbone disorder around phosphorus P(9) in the pure-spermine form [18]. There, the backbone occupies partially the Z_I and partially the Z_{II} conformations (Figure 18.9), so that even in the absence of divalent inorganic cations, the stability of the Z_{II} conformation can approach that of the Z_I conformation.

18.3.5 Structure of DNA Duplexes with Mismatched Base Pairs

Faithful transmission of genetic information depends on the correct insertion of bases during replication, proofreading by DNA polymerases and the excision of misinserted bases. In addition to mutations induced by mismatches during replication, chemical agents, heat and UV radiation can damage DNA. Heat can cause

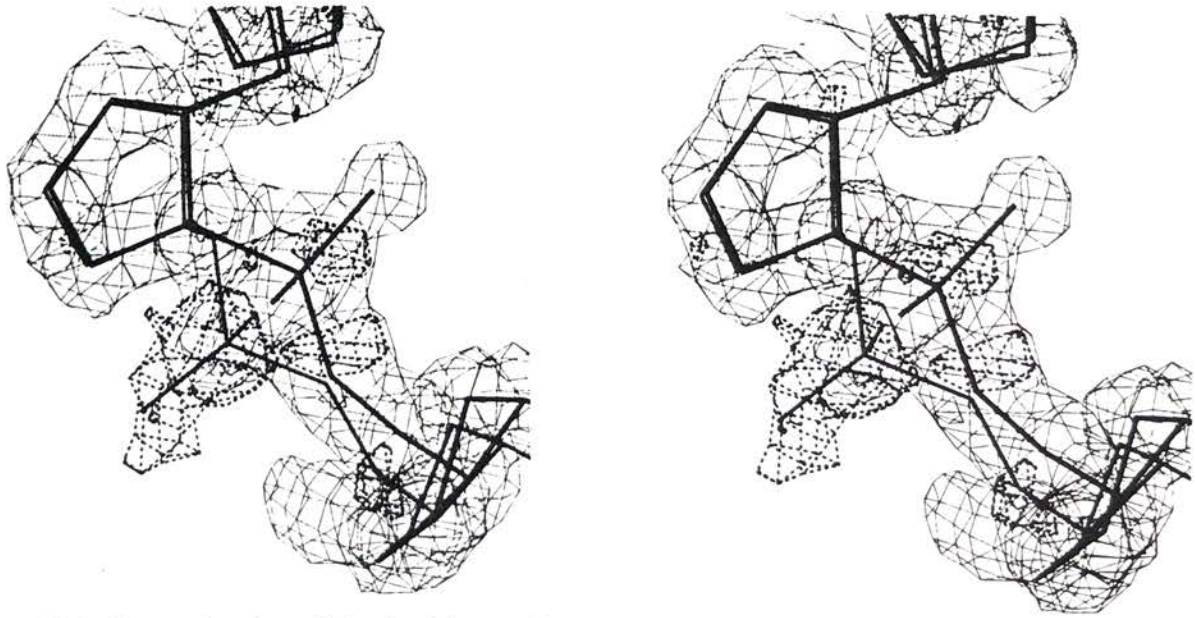

Fig. 18.9. Stereo drawing of the backbone of residues G8 and C9 around phosphorus P(9) in the pure-spermine form of $[d(CGCGCG)]_2$ in the Z_I conformation (thick lines). Sum $(2F_{obs} - F_{calc})$ and difference $(F_{obs} - F_{calc})$ electron density maps are represented as nets of thin lines (sum map) and thick dashed lines (difference map). The partial occupancy of the Z_{II} conformation at this site is demonstrated by superimposing the backbone of the magnesium form at the G4pC5 step (thin lines) on the sugars of the G8 and C9 residues of the pure-spermine form. The atoms of the phosphodiester linkage of the magnesium-form backbone fit well into the difference density

deamination of bases and loss of bases by glycosylic hydrolysis. UV radiation generates pyrimidine dimers and 6–4 photoadducts. Chemicals altering the genetic material include alkylating agents, polycyclic aromatic hydrocarbons, inorganic and organic electrophiles and many more ([48] and references cited therein). Most progress in the study of enzymes involved in DNA repair processes has been achieved in work with *E. coli*, although some eukaryotic repair enzymes have also been isolated and studied in recent years (see reviews [49, 50]). In *E. coli*, the two major regulatory networks activated by DNA damage are the SOS network and the adaptive response network. The former controls the synthesis of proteins which deal with excision repair, daughter-strand gap repair and double-strand break repair, and the latter is activated for removal of methyl and ethyl groups from DNA and excision repair of alkylated bases. There may be an additional repair circuit in *E. coli* which is responsible for repairing damage due to oxidation.

Detection of DNA damage or misinserted bases is crucial for the repair process, and structural alterations either of the DNA backbone or the groove topology are very likely to be an important element for repair activation. It is therefore interesting to incorporate altered bases and mismatches into oligonucleotides and study their structure with physical-chemical methods. Most of what we know about the structural changes induced by mispaired or chemically altered bases stems from crystal structures and solution studies of mispaired oligonucleotides (reviewed in [51]). There are now several crystal structures of oligonucleotides incorporating G·T mismatches in A-DNA [52, 53], in B-DNA [54] and in Z-DNA [55–57] (the mismatch in reference [57] being a G·^{Br}U pair). The other two types of mismatch for which crystal structures exist are the C·A pair in B-DNA [58] and the G·A pair in B-DNA [25, 59, 60]. Furthermore, there are two crystal structures of oligonucleotides containing mismatches with inosine (guanine lacking the N(2) aminogroup), I·A in a B-type duplex [61], and I·T in an A-type duplex [51]. Mismatched base pairs found in oligonucleotide crystal structures are shown in Figure 18.10. The arrangement found for the G·T mismatch is the same in all cases and corresponds to the originally proposed wobble base pair [62]. The adenine in the C·A base pair established by X-ray analysis (Figure 18.10C) is very likely protonated, although a contribution from a tautomer incorporating a protonated C (an example is shown in Figure 18.10B) cannot be ruled out completely. For the G·A pair, two alternative arrangements were found. In the B-DNA decamer d(CCAAGATTGG) [25] (mispaired bases underlined), both bases are in the *anti* conformation (Figure 18.10D). Alternatively, in the B-DNA dodecamers d(CGCGAATTAGCG) [59] and d(CGCAAGCTGGCG) [60], guanine is *anti* and adenine is *syn* (Figure 18.10C). From these structures we learn that although mispairs show a variety of hydrogen-bonding patterns, they almost always involve major tautomers. There are only small variations in the backbone conformation, and the DNA can thus accommodate mismatches without altering its overall shape. The G·T mismatch displays similar geometries in different environments and is thus not affected by the global conformation. In all cases, the mismatches are stabilized by an extensive hydrogen bonding network to the solvent.

The finding that the global conformations of duplexes with mismatched base pairs are practically identical with those of normal ones limits the structural differences

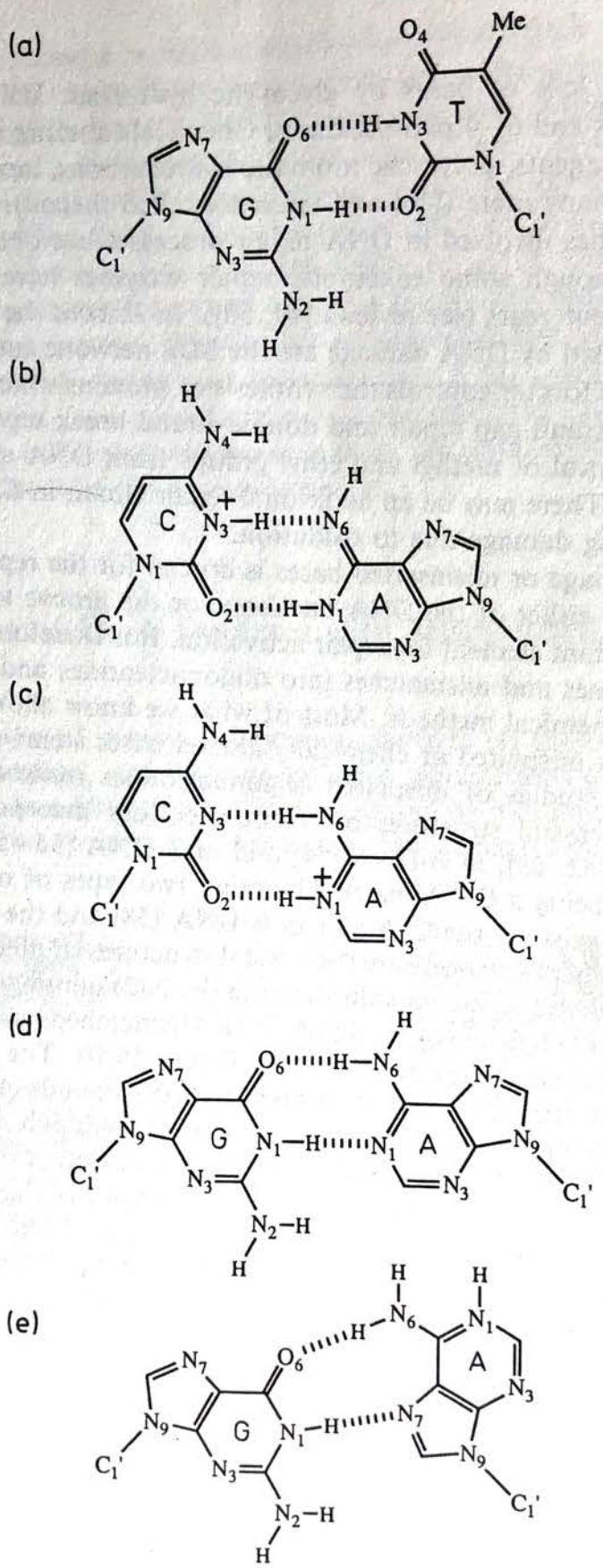

Fig. 18.10. Base pair mismatches found in crystal structures of mispaired oligonucleotides. The G·T wobble base pair (a), the C·A mismatch with cytosine protonated (b), the C·A mismatch with adenine protonated (c), the G·A mismatch with *anti* orientation of both bases (d), and the G·A mismatch with *syn* orientation of adenine (e)

which can be probed by enzymes to detect altered functional groups in the grooves at mismatch sites. Also, there seems to be no correlation between the repair efficiency and the thermodynamic stability of specific mismatches [63], even allowing for the influence of different neighboring base sequences on the stability of the mismatch [64, 65]. It has been shown by in-vitro experiments that the efficiencies by which mistakes are detected by DNA polymerases I and III during proofreading vary considerably for different types of mismatches [66]. G·T pairs were detected most efficiently, A·C pairs to an intermediate extent, and G·A with the lowest efficiency. These variations in the recognition of mismatches have been attributed to different extents of conformational alterations in the grooves as a result of the mispairing [55]. For the G·T base pair, the movements of O(4) and the methyl group of thymine into the major groove might result in steric hindrance at the active site of the polymerases or repair enzymes (see Figure 18.10a). Similarly, the N(4) of cytosine in the C·A pair is shifted further into the major groove (Figure 18.10b, c) and may be an important recognition element for repair enzymes [58]. The low efficiency in detecting the G·A pair could be explained by the arrangement found in the *anti-anti* conformation (Figure 18.10d), which is nearly indistinguishable from a regular G·C base pair with respect to the information content in the grooves. However, the arrangement with adenine in the *syn* conformation (Figure 18.10e) would provide very different steric disposition of functional groups in both the major and the minor groove. The argument that mismatches with larger structural deviations are more efficiently repaired than those resembling standard base pairs is supported by the results of the crystal structure of the mismatched dodecamer d(CGCm^{6}GAATTTGCG) ([67], mismatched bases underlined, m^{6}G is O^6-methylguanine). The m^{6}G·T mispair was found to be geometrically very similar to the standard G·C base pair, which might explain the failure of repair enzymes to recognize and remove this mutagenic lesion. The structure of a m^{6}G·C pair has also been investigated in a mispaired Z-hexamer, and the cytosine was found to be protonated [68]. Although the available results of structural studies and biochemical experiments are in partial accordance, there are not yet enough detailed structures of DNAs incorporating chemically altered bases. Moreover, definitive statements about the recognition of chemical alterations or lesions and about the mechanism of repair processes will have to await the purification and structural characterization of repair enzymes.

18.4 Binding of Drug Molecules to the Minor Groove of B-DNA

The series of crystal structures of the B-DNA dodecamer with sequence d(CGCGAATTCGCG) as well as of a 5-bromocytosine analogue with sequence d(CGCGAATT^{Br}CGCG) revealed that the central six base pairs show relatively high propeller twists with values between +17° and +28°, and that they are bracketed

by base pairs showing much lower propeller twists with values between +5° and +9° [14, 69]. The propeller twists in the central AATT region are 6–10° higher than the propeller twists for the CGCG ends. It is these high propeller twists in the central part which cause the narrowing of the minor groove in this region. The inherently smaller propeller twists of G·C base pairs as compared with A·T base pairs are not a consequence of the third hydrogen bond in the former base pair type, but can be attributed to interstrand base overlaps [70]. Large propeller twists in CpG or GpC steps would produce steric clashes between the N(2) nitrogens of adjacent guanine bases across the two strands. This is supported by the large propeller twists of the base pairs G4·C21 and G16·C9 at GpA steps in the above dodecamer duplexes. The adenines A5 and A17 lack N(2) nitrogens, thus reducing steric hindrance across the strands, and hence propeller twists are high only for G·C base pairs at these steps. Propeller twisting of bases moves the adenine N(3) nitrogens and thymine O(2) oxygens of adjacent residues in opposite strands closer together and allows the bridging of these atoms by water molecules. This leads to the formation of a water spine in the central GAATTC part of the minor groove [71]. The first-shell water molecules are arranged in a zigzag manner and are themselves bridged by second-shell water molecules, resulting in a tetrahedral coordination of the water molecules forming the spine. The regular arrangement of water molecules in this part of the minor groove is partly based on the narrowing of the groove associated with propeller twisting of the base pairs in this region. The water spine itself contributes to the stability of the duplex. An additional contribution to the stabilization of the water spine is made by the sugar O(1′) oxygens, which are directed into the minor groove and form an important part of the polar environment at the floor of the minor groove [72]. Away from the central region of the DNA, the water spine is interrupted, since the wider minor groove at the three terminal G·C base pairs at both ends does not allow the bridging of base acceptor atoms across the strands. In addition, the continuation of the chain of water molecules is disfavored by the presence of guanine N(2) nitrogens (donor) in the minor groove.

The narrowness of the minor groove in the GAATTC region of the B-DNA dodecamer favors the binding of a class of drug molecules referred to as minor-groove binders. These are chain-like molecules with flat aromatic rings, hydrogen-bond forming N–H groups and an overall positive charge. Examples are netropsin, distamycin, Hoechst 33258 or berenil (see formulas in Figure 18.11), which bind tightly to the minor groove of B-DNA via hydrogen bonds and van der Waals contacts. Most of the minor-groove binders possess a natural twist, which allows a snug fit of the molecules in the groove. Because of the displacement of the hydration spine upon drug binding, the entropy contribution is of crucial importance to the stability of these DNA-drug complexes. Figure 18.12 shows the minor-groove widths for the native palindromic B-DNA dodecamer $[d(CGCGAATTCGCG)]_2$ [14], a mispaired B-DNA dodecamer with sequence d(CGCAAGCTGGCG) ([60], mispaired bases underlined), and several dodecamers complexed with minor-groove binders. These are the crystal structures of complexes between (a) $[d(CGCGAATTCGCG)]_2$ and netropsin [73], (b) the same B-DNA dodecamer and Hoechst 33258 [74], (c) $[d(CGCAAATTTGCG)]_2$ and distamycin [75] and (d) $[d(CGCGAATTCGCG)]_2$ and berenil [76]. In addition to these complexes (the first to be published with respect

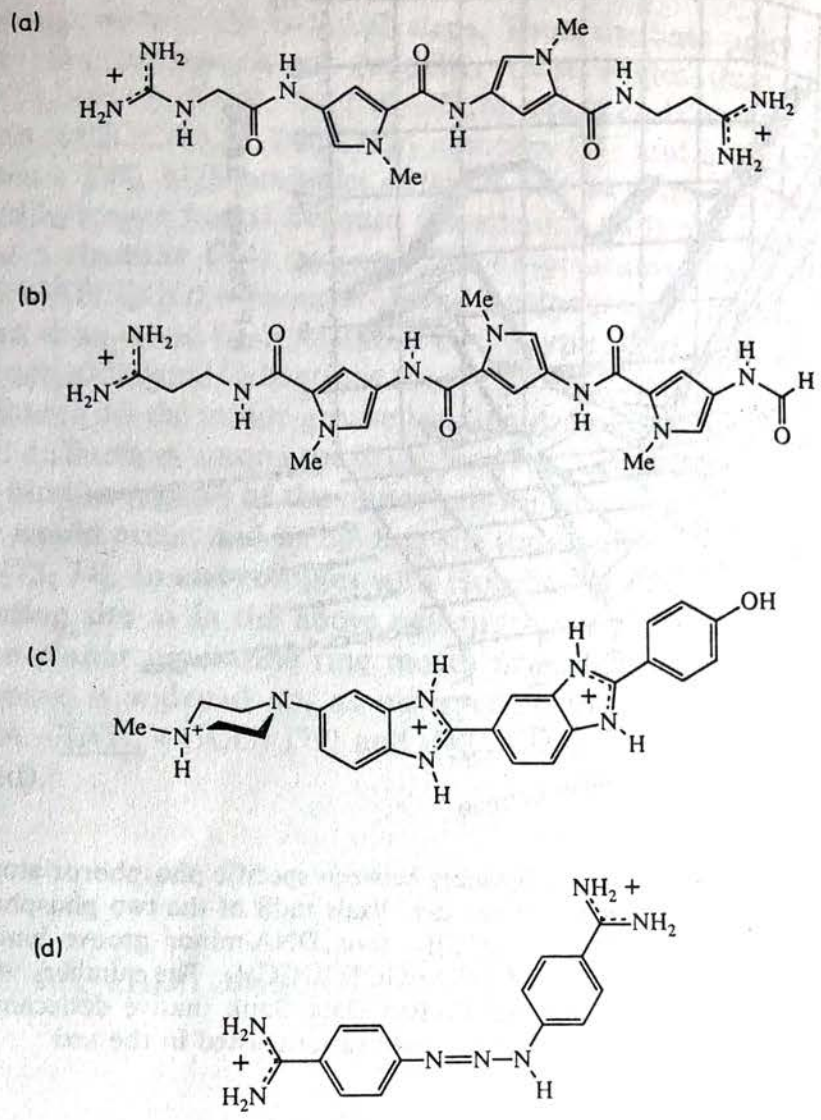

Fig. 18.11. Chemical formulae of minor groove binding drugs: a: Netropsin, b: distamycin, c: Hoechst 33258, and d: berenil

to each drug), several other crystal structures of such complexes are known [77–80]. In the uncomplexed B-DNA dodecamer with sequence d(CGCGAATTCGCG), the minor groove narrows continually and is only 3.2 Å wide at its narrowest (Figure 18.12). In regions where the minor groove is wide (around 7 Å), the major groove is relatively narrow (around 10 Å), and the major groove reaches its maximum width (12.7 Å) where the minor groove is narrowest (3.2 Å) [69]. Thus, the sum of adjacent major and minor groove widths stays quite constant at around 16.5 Å. The binding of a drug molecule in the minor groove of B-DNA results in subtle changes of the groove widths. Whereas the minor groove of the native dodecamer is narrowest near the 10–19 phosphorus pair, it is narrowest near the 8–21 phosphorus pair in the complexes (Figure 18.12). In some complexes, e.g. of netropsin or Hoechst 33258 with the above dodecamer, the minor groove widens upon binding of the drug molecule (in the netropsin complex by 0.5 to 2 Å). In other complexes, such as be-

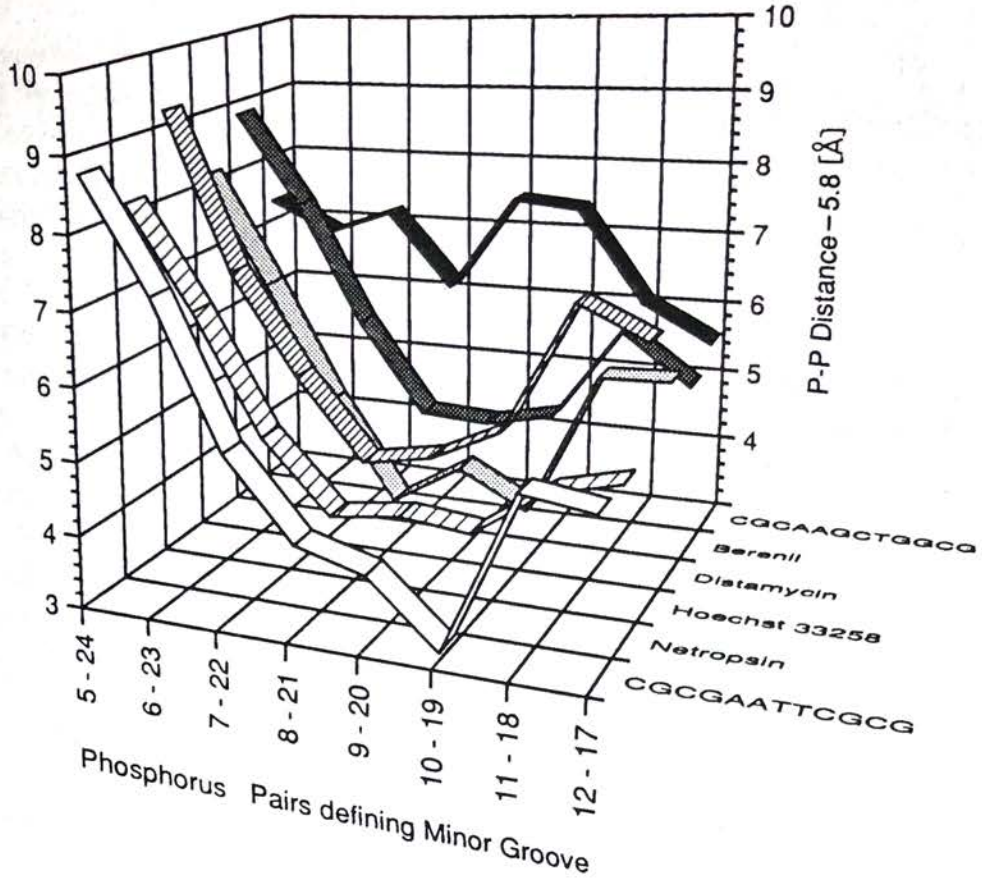

Fig. 18.12. Comparison of the minor groove widths (distances between specific phosphorus atom-pairs of opposite strands minus 5.8 Å, the sum of the van der Waals radii of the two phosphate groups) in the B-DNA dodecamer $[d(CGCGAATTCGCG)]_2$, four DNA-minor groove binder complexes and the G·A mispaired B-DNA duplex $[d(CGCAAGCTGGCG)]_2$. The numbers were obtained from crystallographic data in the Brookhaven Protein Data Bank (native dodecamer, netropsin and Hoechst 33258 complexes) or taken from the publications listed in the text

tween distamycin and $[d(CGCAAATTTGCG)]_2$, the minor groove collapses upon binding of the drug molecule. In general, the width of the minor grooves in the DNA-drug complexes is more regular than in the native dodecamer (Figure 18.12). The sequence dependence of the minor-groove width is manifested by a comparison of the groove widths in the above dodecamer and its complexes with those found in the structure of the A·G mispaired oligonucleotide duplex $[d(CGCAAGCTGGCG)]_2$. There, the absence of the AATT stretch probably leads to a more regular minor groove width with an average of 6.8 Å and a minimal value of 5.6 Å at the 12–17 residue [60].

Comparison of the DNA-drug complexes reveals some characteristic differences among them, both with respect to the geometry of the DNA and the binding mode of the drug molecule. As mentioned above, the high propeller twist of bases in the GAATTC section of the DNA is associated with a narrow minor groove. In the crystal structure of the netropsin-$[d(CGCGATATCGCG)]_2$ complex, the two outer A·T base pairs show very small propeller twist angles of 6 and 8°, resp. [78]. This was attributed to short distances between N(6) nitrogens of adjacent adenine bases

of opposite strands in the two ApT steps. Thus, the base pairs in an alternating AT sequence tend to have lower propeller twist angles than the base pairs in a $d(A)_n \cdot d(T)_n$ stretch. There, the N(6)–N(6) repulsion is absent, but, as found in the distamycin-$[d(CGCAAATTTGCG)]_2$ complex [75] and a dodecamer duplex with a $d(A)_6$ stretch [20], high propeller twists in the stretches result in the formation of bifurcated hydrogen bonds between the adenine N(6) nitrogens from adjacent base pairs and a thymine O(4) oxygen. The observations that neither the dodecamer d(CGCGATATCGCG) without drug nor the dodecamer d(CGCTTTAAAGCG) with or without drug could be crystallized in the typical orthorhombic dodecamer structure type are additional indications for the geometrical changes induced by TpA steps [75]. Although all the minor-groove binding drugs show a specificity for AT regions, structural differences among the drugs lead to changes in base preferences. Netropsin does not bind to regions of the minor groove where a steric clash with guanine N(2) nitrogens would occur, and its binding site (underlined) is therefore d(CGCG–AATT–GCG) [73, 74]. In one complex with Hoechst 33258, the drug molecule adopts the same binding site as in the above netropsin complex [77, 80]. In others, the more bulky non-planar piperazine ring moves towards the inner G·C region, where the minor groove is widened up, as observed in the complexes with the dodecamers d(CGCGA–TATC–GCG) [79] and d(CGCGA–ATTC–GCG) [74] (binding sites underlined).

18.5 DNA Intercalation

18.5.1 Conformational Changes in the Sugar-Phosphate Backbone

DNA intercalation is the insertion of a planar molecule or planar group (e.g. an aromatic ring) between base pairs without interruption of the hydrogen bonds between the bases. The intercalated DNA adopts an overall B conformation with some characteristic conformational alterations, the most conspicuous being the doubled distance of about 7 Å between the two base pairs wrapped around the intercalator. Moreover, intercalation unwinds DNA at certain base pair steps, whereby the site of unwinding depends on the nature of the intercalator, and the DNA backbone undergoes subtle conformational changes, with some of the torsion angles lying outside the normal ranges.

As a preparation for the structural description of drug intercalation some details of the native B-DNA structure are given here. The structures of several oligonucleotide duplexes have yielded some rules concerning the geometry of the backbone in the B-DNA conformation [27, 69]. Torsion angle α is always g^-, and torsion angle γ is always g^+. The only correlated pairs of torsion angles are δ and ε with ζ, and ε and ζ with β of the following residue. These correlations reflect two

conformational states of B-DNA: The "main" conformation (see Section 18.2) derived from fiber diffraction analyses [81], and an alternative "minor" conformation, initially proposed on theoretical grounds [82]. Both conformations occur in the structures of the B-DNA dodecamers, and have been termed B_I (main conformation) and B_{II} (minor conformation) [83]. In the main conformation, the ε and ζ angles are *t*, g^- and in the minor conformation, they are g^-, *t*. Changing torsion angle ζ from g^- to *t* widens torsion angle δ. The value of the latter is restricted by the geometry of the five-membered ring (see Figure 18.3 d right). Opening of ζ to *t* results in the transition from B_I to B_{II}. Change in δ is associated with alterations of the sugar pucker (Figure 18.3 d) and, as mentioned before, the sugar conformation is flexible with a number of different states besides the C(2′)-*endo* conformation. The "elbow" C(3′)–O(3′)–P, which points at the minor groove in the B_I conformation, swings inward in the transition to B_{II} and then points toward the helix axis. This results in an important property of the B_{II} conformation, namely unstacking of the bases which lie on either side of the phosphate group associated with the above conformation [15, 27]. In the phosphorothioate analogue of $[d(GC)_3]_2$, both strands more or less alternate between the B_I and B_{II} conformation, so that B_I conformations are opposite B_{II} conformations within base pair steps. On the other hand, in the B-DNA decamers, B_{II} conformations are found on both sides of some base pairs. This has no influence on the width of the minor groove, which does, however, tend to be widened at so-called step-diagonal pairs, where B_{II} conformations occur at adjacent steps across the strands. Crystal packing, by affecting the width of the minor groove, may indirectly induce the B_I to B_{II} conformational change. The most important point to bear in mind with respect to an analysis of the backbone geometries in DNA-intercalator complexes is the finding that the B_{II} conformation requires the bases to be unstacked, whereas the B_I conformation does not.

Because of their therapeutic application as anticancer drugs, the anthracycline family of antibiotics has been of particular interest, both in terms of their mode of function and of the structural properties of their complexes with DNA. Crystal structures of complexes between anthracycline intercalators and DNA fragments are now available for complexes between: (a) daunomycin (D) and $[d(CGTACG)]_2$ [84] (designated as CGTACGD in Figure 18.13), (b) both daunomycin and adriamycin (A) with $[d(CGATCG)]_2$ [85] (CGATCGD and CGATCGA), (c) 11-deoxydaunomycin (11) and $[d(CGTACG)]_2$ [86] (CGTACG11), (d) nogalamycin (N) and various derivatives of $[d(CGTACG)]_2$ [87–90] (CGTsACGh N, CGTsACGo Na, CGTsACGo Nb, see caption to Figure 18.13) and (e) 4′-epiadriamycin and $[d(CGTACG)]_2$ [91] (not included in the following data analysis). In all these complexes, the aromatic system of the drug is intercalated between the CpG steps of the DNA, and thus there are two drug molecules per DNA duplex. An analysis of the backbone torsion angles in these complexes reveals that α and γ are mostly in the g^- and g^+ ranges, respectively, similar to standard B-DNA (Figure 18.13 a). However, in the 11-deoxydaunomycin and nogalamycin complexes, some of the α and γ angles lie in quite different ranges. In two cases (11-deoxydaunomycin complex, $\alpha = +9°$ and $\gamma = -18°$ for residue G6, and nogalamycin complex, $\alpha = +74°$ and $\gamma = -165°$ for the same residue), the altered conformations may be attributed to the terminal position of

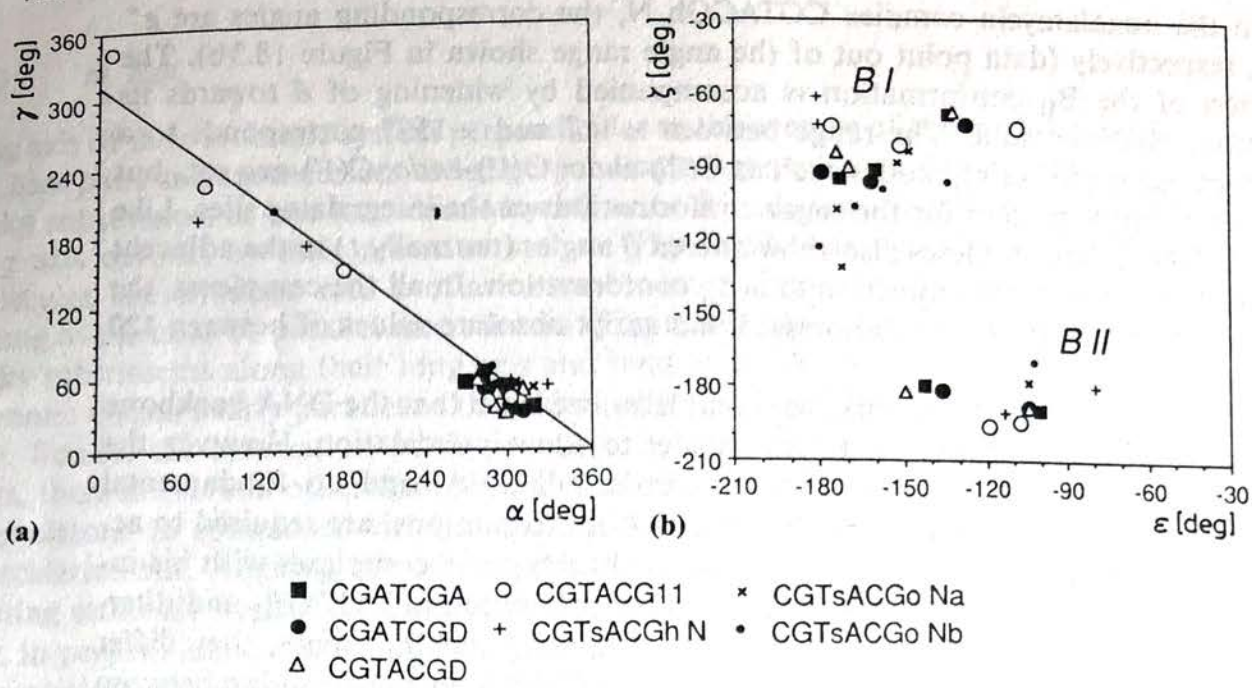

Fig. 18.13. a: Correlation between backbone torsion angles α and γ from six different crystal structures of DNA-intercalator complexes.
b: The B_I and B_{II} conformational families in a correlation between torsion angles ε and ζ from six different DNA-intercalator complexes. References for the crystal structures are given in the text. All the complexes except for the nogalamycin complex CGTsACGo N adopt crystallographic twofold symmetry. CGTsACGo Na and CGTsACGo Nb designate the two independent strands (a, b) in the orthorhombic (o) nogalamycin complex. In this complex as well as in the hexagonal (h) nogalamycin complex CGTsACGh N, the DNA incorporates a phosphorothioate group (s) at the TpA step

the corresponding residues. The linear correlation coefficient for α and γ is -0.86 (Figure 18.13 a). The backbone of DNA can thus adopt a number of different conformations for α and γ (e.g. t, t and g^+, t) to accommodate intercalation. As in uncomplexed B-DNA, the backbone in the DNA-intercalator complexes can adopt the alternative B_I and B_{II} conformations (Figure 18.13 b). With the exception of one of the nogalamycin complexes (CGTACGo Na, b; here a and b designate the two strands of the duplex; all other complexes have crystallographic twofold symmetry), residues G2 and C5 (and the symmetry related residues G8 and C11 of the opposite strand) assume the B_{II} conformation. Thus, the B_{II} conformations of the two strands do not occur at the same step (intercalation step), but rather each intercalation step is spanned by a B_I conformation in one and a B_{II} conformation in the other sugar-phosphate backbone. The adoption here of the B_{II} conformation reinforces the view that formation of B_{II} goes along with the unstacking of bases. The combination of B_I and B_{II} conformations at the intercalation step could contribute to the stacking interactions between the aromatic system of the intercalated drug and the DNA bases. Some residues have ε and ζ angles which do not fall into the t, g^- or g^-, t ranges (Figure 18.13 b). In most of the complexes, the backbone at the G2pT3 step (and the symmetry equivalent G8pT9 step) falls into the t, t range. In the 11-deoxydaunomycin complex, the ε and ζ angles in the terminal residue C1 are both

g^-. In the nogalamycin complex CGTACGh N, the corresponding angles are g^+ and t, respectively (data point out of the angle range shown in Figure 18.3 b). The adoption of the B_{II} conformation is accompanied by widening of δ towards its maximum possible value. The range between +117 and +159° corresponds to a variety of sugar puckering modes such as C(2′)-*endo*, O(1′)-*endo*, C(1′)-*exo* etc., but there is no simple pattern for the sugar conformations at the intercalator sites. Like native B-DNA, the complexes also show altered β angles (normally t) in the adjacent residues as a result of the transition to a B_{II} conformation. In all the complexes, the β angles of such residues are compressed and adopt absolute values of between 120 and 148°.

From the analysis of the intercalator complexes, we learn that the DNA backbone can extend the normal ranges of torsion angles to allow intercalation. However, the backbone is generally similar to that of standard B-DNA, and no fundamental changes, such as those between the A-, B- and Z-conformations, are required to accommodate intercalation. The alterations of the backbone in complexes with bis-intercalators are more severe. Whereas in the complex between $[d(CGCG)]_2$ and dicercalinium [92] the angles ε and ζ fall into the above B_I and B_{II} ranges, they differ considerably from any of those found in native B-DNA in the complex between triostin A and $[d(GCGTACGC)]_2$ [93]. Here, the ε, ζ combinations for residues G3; T4; A5 and C6 are t, t; g^-, g^-; g^+, g^+ and t, t.

18.5.2 Parallel and Perpendicular Intercalators, DNA Unwinding and Sequence Specificity of Intercalation

Somewhat surprisingly, the intrastrand distances between adjacent phosphorus atoms in DNA-intercalator complexes are quite regular [88, 94]. Considering the shifting of base pairs at intercalation sites, one might have expected distances between phosphorus atoms of backbone segments bridging these steps to be larger than those at unperturbed steps. The average P···P distances in the anthracycline-DNA complexes are between 6.6 and 6.8 Å with standard deviations of around 0.2 Å, and are thus similar to the average distance of 6.7 Å observed in the B-DNA dodecamer d(CGCGAATTCGCG). This finding and the unwinding of DNA at specific sites upon intercalation have led to the proposal that the helical twist in DNA is the result of two opposing interactions [94]. According to this model, the electrostatic repulsions between adjacent phosphate groups tend to wind the helix, whereas base-base stacking interactions tend to unwind it. It seems that the distance between adjacent phosphate groups is kept close to an optimal value, and that upon intercalation the DNA adapts to this by lengthening and unwinding. The location and degree of unwinding depend on the nature of the intercalator.

One way of classifying intercalators is by the orientation of their aromatic systems relative to the DNA base pairs, and one can thus differentiate between "perpendicular" and "parallel" intercalators [94]. The anthracyclines intercalate with the

long axis of their aromatic system perpendicular to the average of those of the flanking base pairs and are therefore called perpendicular intercalators. They usually have bulky substituents at one (e.g. daunomycin) or both ends (e.g. nogalamycin) of the long axis defined by their aromatic system (Figure 18.14a). On the other hand, molecules like ethidium and proflavine or bis-intercalators such as ditercalinium belong to the class of parallel intercalators (Figure 18.14b). These intercalators lack bulky substituents along their long axis and bind to DNA with the long axis of the aromatic system nearly parallel to those defined by the flanking base pairs (e.g. [95, 96]). Besides the orientation of the aromatic system relative to the flanking base pairs, there are several other features which distinguish parallel from perpendicular intercalators. In complexes with parallel intercalators, the DNA is unwound at the intercalation site. Aligning the two base pairs and the aromatic system of the intercalating molecule results in extensive stacking interactions in such complexes. However, in perpendicular type complexes the DNA is unwound at steps adjacent to the intercalation site. Here, unwinding at the intercalation step itself would only slightly improve the stacking. In perpendicular type complexes, base pairs flanking the intercalator tend to buckle (e.g. [87, 88]), whereas in parallel type complexes they are relatively planar. Such adaptions of base pairs at the intercalation site contribute to the stability of perpendicular type complexes via van der Waals contacts.

In all the crystal structures of anthracycline-DNA complexes, the drug molecule is intercalated at the CpG steps. A common feature of the more simple anthracycline antibiotics such as daunomycin and adriamycin is the aglycone aromatic system,

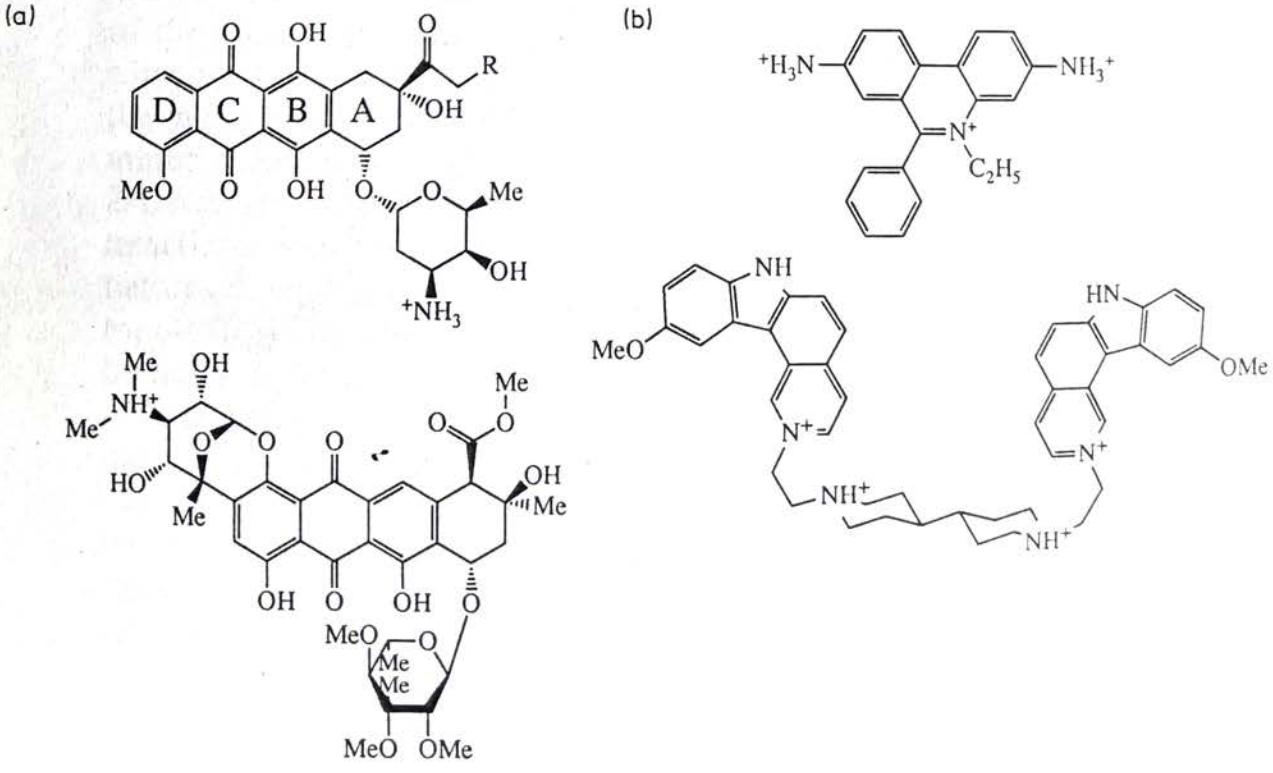

Fig. 18.14. Structures of DNA intercalators.
a: Perpendicular intercalators daunomycin (R = H), adriamycin (R = OH) and nogalamycin (from top);
b: parallel intercalators ethidium and ditercalinium (from top)

ontaining four fused rings, and a positively charged amino sugar attached to yclohexene ring A (Figure 18.14a). The cyclohexene ring and the amino sugar are laced in the minor groove, and the hydroxyl group at position 9 of ring A interacts a hydrogen bonds with nitrogens N(2) and N(3) of the adjacent guanine base. hese hydrogen bonds provide the specificity of the above anthracyclines, and erivatives lacking the O(9)H hydroxyl group are biologically inactive [97]. logalamycin is larger than the daunomycin-family anthracyclines and is substituted t both ends of its aromatic system (Figure 18.14a). The positively charged amino ugar of daunomycin is replaced by a bulky, uncharged nogalose sugar and a methyl ster group in nogalamycin. In addition, a bicyclic amino sugar with a positively harged dimethylamino group is fused to the other end of the aromatic system. In he complex with DNA, the nogalose rests in the minor groove, the carbonyl oxygen of an ester at the C(10) position of ring A is hydrogen-bonded to a guanine N(2) itrogen, and the two hydroxyl groups of the amino sugar at the opposite end form pecific hydrogen-bond contacts to adjacent bases in the major groove [87–90] Figure 18.15). On the basis of the hydrogen-bond pattern between nogalamycin and DNA, backed by modeling studies and footprinting data, it was concluded that

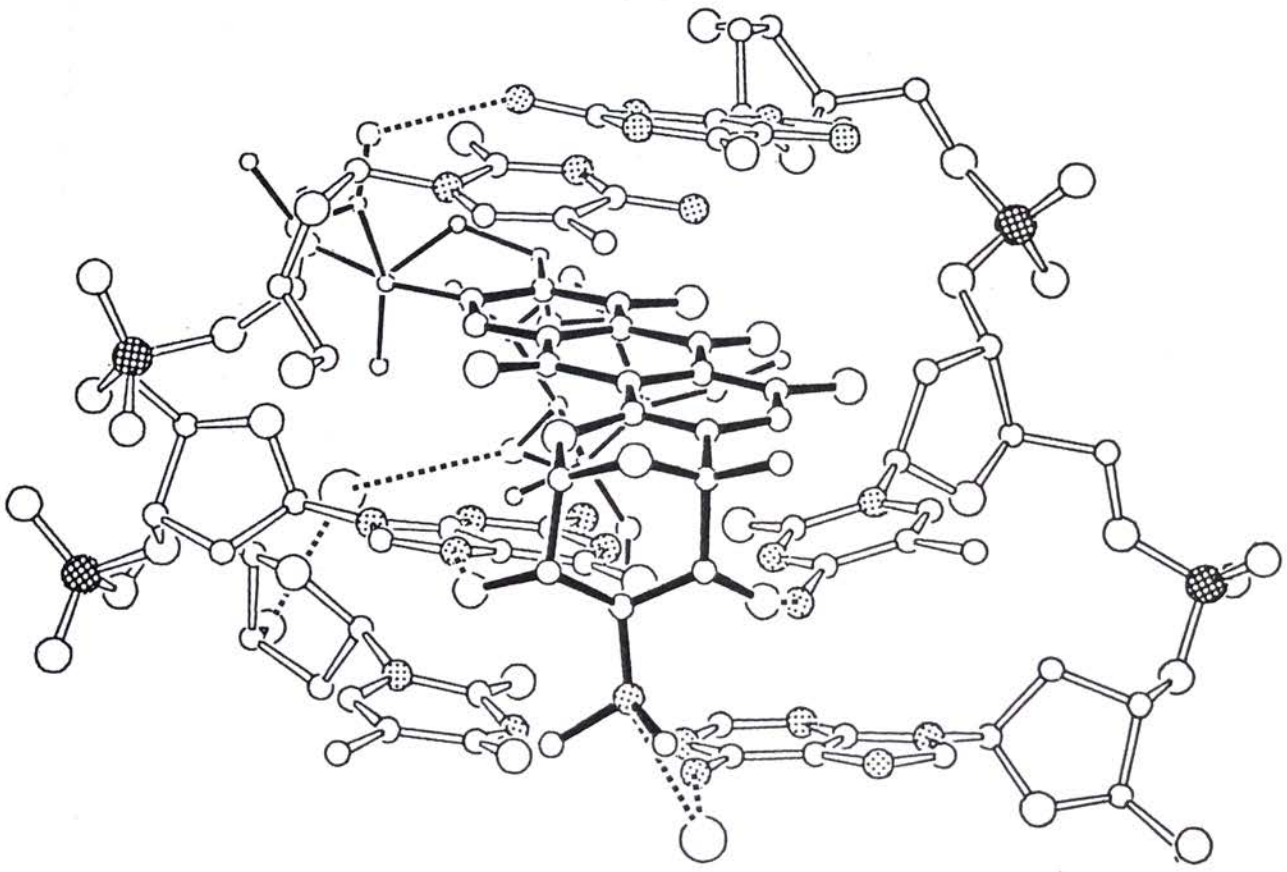

Fig. 18.15. The intercalation site of the $[d(^{me5}CGTsA^{me5}CG)]_2$ nogalamycin complex viewed from the major groove. The DNA contains cytosines methylated at the 5-positions and a phosphorothioate linkage, TsA. The bicyclic amino sugar is in the foreground, the DNA is drawn with open bonds and nogalamycin with solid bonds. Phosphorus atoms are stippled in black, nitrogen atoms are stippled in grey, and water molecules are drawn with larger circles. Water mediated hydrogen bonds and hydrogen bonds between nogalamycin and DNA are dashed lines

nogalamycin intercalates preferably at the 5′-side of guanine (e.g. between NpG steps) and at the 3′-side of cytosine (e.g. between CpN steps), the sugars facing the G·C base pair [90]. In general, the stability of complexes between anthracyclines and DNA is based on stacking interactions between the aromatic system of the intercalator and the flanking bases, with additional contributions from van der Waals interactions between drug substituents and the DNA in the grooves, as well as from water molecules and hydrated metal ions which bridge functional groups of the drug and the DNA. Atoms of the DNA backbone, including the negatively charged phosphate groups, usually do not form any contacts to the intercalated drug. The sequence specificity of the anthracyclines is based on various hydrogen-bonding interactions to acceptors and donors of flanking DNA bases.

18.6 Interactions between Nucleic Acids and Proteins

18.6.1 Sequence-specific Recognition of Double Helical Nucleic Acids

Since the base pairs are stacked, the only elements by which they can be recognized are their outer edges in the two grooves of the duplex. The main topological features of the three duplex families were outlined in Section 18.2. From these features, we can conclude that interactions between proteins and B-DNA must occur mainly via the major groove, and only to a limited extent via the minor groove. In A-DNA, the minor groove is accessible whereas the major groove is largely hidden. However, in Z-DNA, the convex surface, the equivalent of the major groove, is accessible to interactions with proteins while the minor groove is mostly hidden. As mentioned before, double-stranded RNA adopts the A conformation, and, from the above topological differences, one would expect the mechanisms for recognizing RNA to be quite different from those for B-DNA.

In an early attempt to identify the factors involved in discriminating between base pairs, use was made of the crystal structures of dinucleotide mini-helices, and potential recognition sites within the six different base-pair steps defined by the four possible Watson-Crick pairings were investigated [98]. The criterion for selecting a specific recognition site was its accessibility to an approaching probe. Six potential recognition sites were identified in the major groove. One is close to the imidazole N(7) nitrogen of purines or the C(5) carbon of pyrimidines (see Figures 18.16a and 18.16b for orientation). The second one is defined by the O(4) oxygen (thymine) or the N(4) nitrogen (cytosine) of pyrimidines, and is close to the third site which is defined by the O(6) oxygen (guanine) or the N(6) nitrogen (adenine) of purines. The other three sites are related to these three by a dyad axis which relates the two riboses

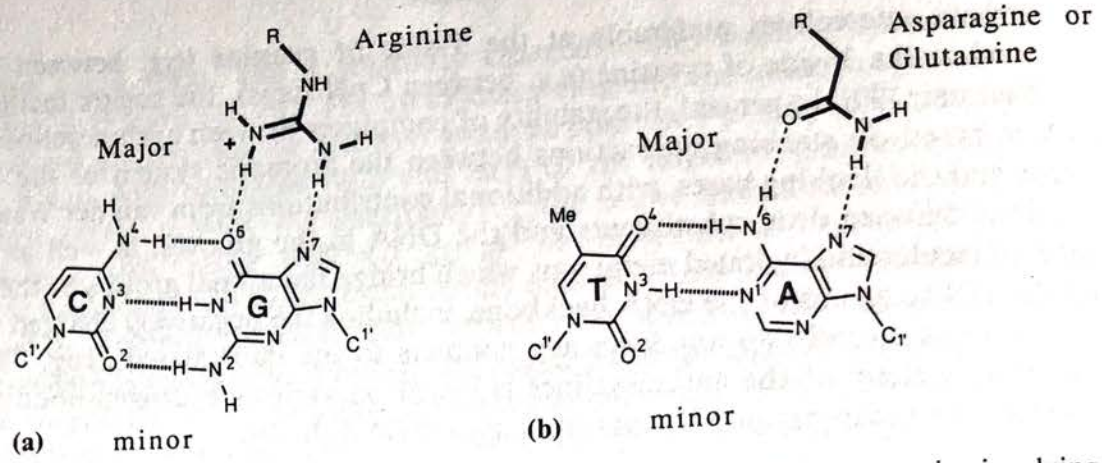

Figure 18.16. Sequence-specific binding of amino acid side chains to DNA base pairs, involving two hydrogen bonds.
a: Recognition of a G·C base pair by arginine in the major groove.
b: Recognition of an A·T base pair by asparagine or glutamine in the major groove.
The DNA bases are designated with capital letters and functional groups of the bases involved in Watson-Crick hydrogen bonds or in hydrogen bonds to the amino acid side chains are numbered. DNA bases are drawn with thin lines, amino acid side chains are bold and hydrogen bonds are dashed. To visualize the recognition sites in the major and the minor grooves, mentioned in the text, one can imagine the two base pairs (G·C and A·T) being shifted on top of each other and rotated around their centers by 36° (the average helical twist angle in B-DNA). The dyad axis relating the two riboses (not included in the drawing) is then lying in the paper plane, running from top to bottom

to each other. Of these six sites, only four are found in any of the four Watson-Crick base pairs. A closer inspection of the various recognition sites shows that a probe interacting via a single hydrogen bond at the outer major groove could discriminate, for example, a purine-pyrimidine from a pyrimidine-purine base pair. Similarly, a probe interacting at the center of the major groove could discriminate base pairs with an adenine or cytosine from those with a guanine or thymine at that place. Note that the presence or absence of a methyl group in the C(5) position of thymine, uracil and cytidine provides another recognition site in the outer major groove. Although this methyl group allows differentiation between the pyrimidines in A·T and G·C base pairs, the purines would still be degenerate. Thus, a single site probe in the major groove could differentiate among several base pairs, but not among all.

In the minor groove, three recognition sites are found. Two are at the outside and are occupied either by the N(3) nitrogen of purines or the O(2) oxygen of pyrimidines (see Figures 18.16a and 18.16b). These sites are probably not very important for the discrimination process, since both atoms are acceptors and a superposition of base pairs shows that they are only 0.5 Å apart. The third site is positioned in the center of the minor groove close to the dyad axis relating the two riboses. In the case of A·T or T·A base pairs, the site is occupied by the C(2) carbon of adenine and there is thus no apparent mechanism for a protein to differentiate between the above base pairs via either of the two recognition sites in the minor groove. However, with C·G or G·C base pairs, the site close to the dyad axis is occupied by the N(2) nitrogen of guanine. This allows a somewhat limited differentiation between a C·G and a

G·C base pair. However, a probe interacting with the center of the minor groove could discriminate G·C and C·G base pairs from A·T and T·A base pairs. Again, as in the case of the major groove, a single hydrogen-bond probe in the minor groove cannot differentiate unambiguously among all base pairs.

The above analysis suggested that discrimination among all base pairs must involve the formation of at least two hydrogen bonds with the base recognition sites [98]. A rigid secondary structure is one of the requirements for precise positioning of side chains relative to recognition sites in the grooves. This still leaves the problem of mobility of the protein probe with respect to the functional groups of the bases. A protein must therefore use the backbone of the double helix as a sort of frame of reference. There are several ways by which DNA binding proteins achieve this (see Section 18.6.2). In principle, the probed sites could lie far apart or could be located in different grooves. On the other hand, a protein could use adjacent recognition sites of only one base to discriminate between base pairs. Although a protein could form two hydrogen-bonds using two amino acid side chains, the formation of two hydrogen bonds by a single amino acid allows more precise recognition [98]. Possible schemes for this are depicted in Figure 18.16. Analogous arrangements are found in highly specific polynucleotide interactions involving a third nucleotide in the major groove. They always use two hydrogen-bonding recognition sites. Although Figure 18.16 shows recognition of base pairs by amino acid side chains via the major groove, recognition through formation of two hydrogen bonds in the minor groove is also possible. For example, asparagine or glutamine could form hydrogen bonds to N(2) and N(3) of guanine in the minor groove to recognize a G·C base pair (see Figure 18.16a).

Examples of such interactions of single amino acid side chains with functional groups of DNA bases have been found in a number of X-ray crystal structures of protein-DNA complexes. In the structure of the DNA-Eco RI endonuclease recognition complex, Arg 200 is hydrogen bonded to O(6) and N(7) of guanine [99, 100]. In the λ repressor-DNA complex, Gln 44 forms hydrogen bonds to N(6) and N(7) of adenine [101], and in the complex between λ cro repressor and DNA, Gln 27 forms two hydrogen bonds to an adenine and Arg 38 forms two hydrogen bonds to a guanine [102]. Contacts between single amino acid side chains and bases are also found in the DNA complexes of the 434 repressor and of the 434 cro repressor [103 – 105]. In both complexes, Gln 28 is hydrogen bonded to an adenine. In the trp repressor-operator complex, the interaction between Arg 69 and a guanine is the only direct hydrogen-bonding contact of the protein in the major groove of the DNA [106], and in the engrailed homeodomain-DNA complex, Asn 51 forms hydrogen bonds to an adenine [107]. Several single amino acid side chain interactions with bases are found in the zinc-finger DNA complex [108]: Arg 18 and 24 of finger 1, Arg 46 of finger 2, and Arg 74 and 80 of finger 3 each form two hydrogen bonds of the postulated type to guanine bases. Proteins, of course, do not exclusively rely on such contacts between single amino acids and bases for recognizing specific sites of the DNA. The specificity is based rather on a variety of contacts, including simultaneous interactions of amino acid side chains with adjacent DNA bases (e.g. in the Eco RI complex) or contacts between more than one amino acid and a DNA base, as observed in the zinc-finger DNA complex (the Asp-Arg-guanine interaction is present in all three fingers).

18.6.2 Architecture of DNA-binding Domains in Proteins

This section briefly summarizes the characteristics of DNA-binding proteins from a structural point of view. Here, we are interested mainly in those aspects responsible for the binding to DNA. Several review articles and book chapters describe the architecture of such proteins and the three-dimensional structures of protein-DNA complexes in some detail [109–112]. Table 18.2 lists the protein-DNA complexes for which X-ray crystal structures have been determined. In all of these DNA is found in the B-conformation.

DNA-binding proteins use a variety of tertiary structural elements (see Chapter 15) to fix their positions relative to DNA. The known secondary structural motifs such as α helices and parallel, as well as antiparallel, β sheets are extensively used in the binding domains. However, the motifs are combined in a specific way, found so far only in binding proteins. An important class of prokaryotic transcriptional regulatory proteins makes use of the so-called *helix-turn-helix* motif (Figure 18.17a). This element consists of about 20 amino acids which form two α helices, crossing at angles of around 100°, and which are joined by a short linker. In the repressors of the λ and 434 phages, this motif is formed by helices 2 and 3 (see references in

Table 18.2. Published crystal structures of protein-DNA complexes (April 1992)

Binding motif and protein	Resolution (Å)	Reference
Helix-turn-helix		
eukaryotic:		
engrailed homeodomain	2.8	[107]
prokaryotic:		
434 repressor	4.5 – 3.2	[103]
434 repressor	2.5	[104]
λ repressor	2.5	[101]
434 cro	4.0	[105]
λ cro	3.9	[102]
E. coli trp repressor	2.4	[106]
E. coli CAP	3.0	[113]
Zinc binding		
class 1:		
Zif268	2.1	[108]
class 2:		
glucocorticoid receptor	2.9	[114]
class 3:		
yeast Gal4	2.7	[148]
β ribbon		
E. coli met repressor	2.8	[112, 115]
Endonucleases		
E. coli Eco RI	2.8	[99, 100]

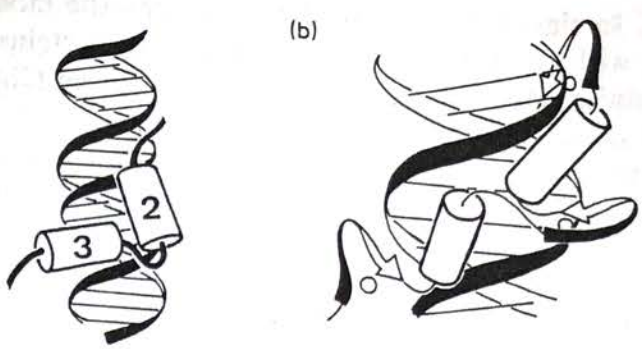

Fig. 18.17. a: Schematic drawing of the helix-turn-helix motif observed in the complexes between DNA and various prokaryotic and eukaryotic binding proteins. Helical regions are represented by cylinders and coil regions are drawn with thick lines. The diameter of an α helix with side chains is about 12 Å and it fits thus into the major groove of B-DNA. In the above protein-DNA complexes, the recognition helix in the groove is referred to as helix 3. Four to six DNA base pairs can be contacted by this helix, depending on its arrangement relative to the local direction of the groove. Helix 2 lies above the groove and helps anchoring the unit relative to the groove. The length and orientation of this helix are somewhat different in the cases of prokaryotic repressors and eukaryotic homeodomains. A number of DNA-binding proteins show sequence homologies in their helix-turn-helix region. For example, the position of a glycine is highly conserved in the linker between the two helices. This has been attributed to the particular backbone conformation of this amino acid, facilitating a tight turn.
b: Schematic drawing of the repeating structure of a three zinc finger segment bound to DNA. Each finger consists of an α helix (cylinder), packed against an antiparallel β sheet (wound arrow). The zinc ion (open circle) is liganded by two cysteines, flanking the turn in the β hairpin and two histidines from the α helix. The α helices are slightly tilted with respect to the major groove of B-DNA, and contacts to the bases are established by residues at one end of the helices. The β sheets are turned away from the DNA and shifted to one side of the major groove, where one of the strands per finger interacts with the DNA backbone. Only the antiparallel β sheet portion of the third finger has been included in the drawing

Table 18.2). Together with helix 1, they form a triangular unit, which is sealed off from the hydrophobic core of the protein by a fourth helix. Helix 3 is referred to as the recognition helix, since it is positioned in the major groove, where its hydrophilic side chains form numerous contacts to DNA bases and backbone. Helix 2 helps to anchor the unit with respect to the DNA and keeps the recognition helix from rolling into the groove [109]. In addition to the more rigid secondary structural elements, λ repressor uses a flexible arm at its N-terminal end to wrap around the DNA, resulting in a number of interactions between protein and DNA base and backbone atoms [116]. Another class of binding proteins is represented by the eukaryotic homeodomains. Whereas most prokaryotic regulators bind as dimers, the homeodomains can bind to DNA as monomers. The motif in these proteins is a single three-helix bundle with two helices lying parallel to one another [107]. Although there are some similarities in the overall helix-turn-helix motifs of the above prokaryotic regulators and the homeodomains, the arrangement of the anchoring helix 2 varies somewhat between the two protein classes [107, 111, 117].

Another major structural motif found in DNA-binding proteins are zinc-mediated loops. So far, three classes of zinc modules have been recognized [111]. The common feature of these classes is the use of a zinc ion as a crosslink to stabilize a small

globular protein. In class 1, the so-called zinc fingers, the module consists of about 30 amino acids, with the zinc ion complexed by two cysteines and two histidines. Class 2 zinc modules occur in receptors for steroids (see Chapter 14) and similar molecules and consist of approximately 70 amino acids with two zinc ions, each complexed by four cysteines. In class 3, found in yeast transcription activators, two proximal zinc ions share six cysteines. Three-dimensional structures of the DNA complexes of all three classes have now been determined ([108, 114, 148], see Table 18.2). In the DNA complex of the class 1 zinc-finger protein Zif268 from mouse, three fingers bind in essentially the same manner, each wrapping partway around the double helix and making direct contacts to a three base-pair subsite in the major groove. Each of these zinc fingers consists of an α helix packed against a β hairpin, with two cysteines from the antiparallel β-sheet region and two histidines from the α helix coordinated to the zinc ion (Figure 18.17b). The α helices fit into the major groove, and their N-terminal portions contact the edges of bases in the groove. However, one of the β strands on the back of the helix does not make any contacts with the DNA, and the second strand contacts the DNA only along one backbone. Although the structures of the DNA-binding domains are quite different for the three classes, they share the feature of exposing an α helix to the major groove of DNA. Zinc coordination occurs either at the amino or carboxy termini of these recognition helices. Class 1 zinc-finger-proteins often show a concatenation of fingers, each recognizing a sequence triplet. Consequently, these proteins interact with single-finger target sites and wrap around the DNA, following the major groove continuously. On the other hand, the DNA-binding mode of the class 2 glucocorticoid receptor and related steroid/nuclear receptors resembles the one exhibited by regulatory proteins of prokaryotes and phages. Here, homo-dimers bind palindromic sequences which lie on one side of the DNA. The recognition sequences of class 3 transcriptional activators such as Gal4 are also palindromic and the DNA-binding domains dimerize symmetrically on the DNA. However, the contact base triplets lie one-and-one-half helical turns apart and are therefore located on different sides of the DNA. The dimerization elements face the minor groove, and, in the Gal4 complex with DNA, the linker between the dimerization element and the DNA-binding zinc module makes several contacts to phosphate groups of the backbone.

Another class of DNA-binding proteins are the restriction endonucleases which cleave DNA at specific sites. The recognition motif found in the complex between DNA and Eco RI features α helices as well, but the interaction mode differs considerably from those found in the helix-turn-helix and zinc-finger proteins. In the Eco RI complex, a parallel bundle of four α helices penetrates the major groove, whereby only the ends of the helices interact with the DNA [99, 100]. In addition, an extension of the polypeptide chain of one of these helices wraps around the DNA and makes several contacts to bases.

A further class of DNA-binding proteins which use the α helix as a recognition element are the leucine zippers, but detailed structural information is so far only available for the protein itself, not for its DNA complex [118]. In still another recognition motif, the met repressor binds to DNA in the form of two highly intertwined monomers. The groove is contacted by an antiparallel β sheet with one strand of the sheet coming from each monomer [112, 115].

No detailed structure information on the recognition of the third form of double-helical DNA, the left-handed Z-DNA, by proteins is available so far. Although several Z-DNA binding proteins and antibodies have been isolated [119, 120], no protein has yet been expressed and purified to the point where X-ray crystallographic or nuclear magnetic resonance methods would allow detailed studies of protein-DNA interactions.

18.6.3 Protein-RNA Interactions

In the protein-DNA complexes described in the last section, the DNA molecules adopt essentially B-type conformation. However, double-stranded RNA adopts the A conformation, and, more important, RNA serves many more functions than DNA and therefore requires a conformational flexibility similar to that of protein folding. The secondary structural motifs in RNA include stem (double-stranded) regions, hairpin loops, single- or multiple-base bulges, internal loops and multiple-way junctions (see reviews [121, 122] and the description of tRNA secondary structure below, which includes some of the mentioned motifs). The conformational flexibility of RNA suggests that other types of recognition are operative in protein-RNA complexes than in complexes with DNA.

At present the crystal structure of transfer RNA (tRNA) provides the most detailed information on the folding of a complex RNA molecule [123, 124]. This nucleic acid plays an important role in the translation of the genetic code into the amino acid sequence of proteins. The first step in the translation of the DNA nucleotide sequence into the protein chain is the transcription of the DNA sequence into a complementary sequence of messenger RNA (mRNA). With four kinds of nucleotide bases present in mRNA, there are a total of 64 (4^3) possible three-base-pair codons, the relation between the codons and the amino acids specifying the genetic code. All but three of the 64 possible codons specify individual amino acids, and many amino acids have more than one codon. The translation of the mRNA sequence into protein then proceeds in two steps. In the first, an amino acid molecule is attached to its particular tRNA, a reaction which is catalyzed by the so-called aminoacyl-tRNA synthetases. For each of the twenty amino acids found in proteins, there is a specific synthetase. The second step takes place in a cellular organelle called ribosome. This complicated machinery consists of more than fifty different proteins and three RNA molecules. The tRNA with its attached amino acid binds to a specific site within the ribosome. There, the three codon bases of the mRNA strand that specify an amino acid are hydrogen bonded to the anticodon bases of the tRNA molecule. The growing polypeptide chain is transferred from a tRNA molecule at a neighboring site to the above tRNA. The empty tRNA then leaves its site and the ribosome moves along the mRNA by one codon, with the tRNA being shifted to a neighboring site, and the former site ready for an incoming charged tRNA [125].

Transfer RNAs are relatively constant in size and consist of between 74 to 93 nucleotides which can be organized into a secondary structure with the shape of a

cloverleaf (Figure 18.18a). The conventional numbering considers the molecule to have 76 fixed positions. Of those, 15 are occupied by nucleotides that are conserved among all tRNAs. Most of the tRNAs are type I species with four to five nucleotides in the variable loop. A few others are larger and are designated as type II tRNAs, with 13 to 21 nucleotides in the variable loop. The polynucleotide chains of different tRNAs fold in a similar way and adopt a globular L-shaped structure (Figure 18.18b). The L consists mainly of two double helices oriented at an angle of approximately 90°. One duplex (12 base pairs) is formed by the acceptor stem and the T stem, and the other (9 base pairs) by the anticodon stem and the D stem. Thus, the anticodon loop, the single-stranded segment which interacts with the mRNA, and the acceptor end, where the amino acid is attached, lie about 75 Å apart. The acceptor stem ends with the single-stranded tetranucleotide XCCA-O(3′)H (X is a variable base), containing the amino acid attachment base A76. Several bases from D loop and D stem, variable loop and TΨC loop are engaged in three-base interactions, which contribute to the stabilization of the tRNA conformation.

As was stated above, recognition of nucleic acids in the A conformation must necessarily be different from that of B-type DNA because of the narrow and deep major groove in the A form, and the lack of information in the minor groove. Also, with the conformationally more flexible RNA, other forms of recognition are likely to be used. These could involve single-stranded regions or the termini of double-helical segments, where functional groups lying in the major groove are accessible. Mismatched base-pairs such as G·U, which occur quite frequently in RNA, could also act as specific recognition sites [126]. The wobble base pairing of that base pair alters the conformation of the double helix by shifting nitrogen N(2) of guanine further into the minor groove, where it would be accessible to amino acid side chains (Figure 18.19). In addition, the discrimination of a U·A base pair from a G·C base pair could also depend on its smaller free energy of disruption.

A large portion of the tRNA molecule is composed of double-stranded regions, and one wonders therefore how the aminoacyl-tRNA synthetases recognize their cognate tRNAs. Synthetases can be divided into two classes, I and II, whereby the classification is based on differences in sequence elements associated with specific three-dimensional structural patterns. Each class includes ten members, or half the existing aminoacyl synthetases. They can be further divided in subgroups of extended homologies. The crystal structures of two tRNA-synthetase complexes allow us to address the question of recognition in some detail [127, 128]. The *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP [127] is a class I synthetase, and the yeast aspartyl-tRNA synthetase complexed with $tRNA^{Asp}$ [128] is a class II synthetase. In both complexes, the synthetase binds along the inside of the L-shaped tRNA molecule (contacting the tRNA in Figure 18.18B from the right). A major difference between the complexes is found in the relative positions of tRNA and synthetase. Gln-tRNA synthetase and Asp-tRNA synthetase interact with opposite sides of their tRNAs. Whereas $tRNA^{Gln}$ contacts its synthetase from the D-loop side and the minor groove of the acceptor stem, $tRNA^{Asp}$ makes contacts to the synthetase from the variable loop side and the major groove of the acceptor stem.

Earlier biochemical and genetic studies had determined some of the so-called identity elements, tRNA nucleotides that ensure correct recognition by synthetases

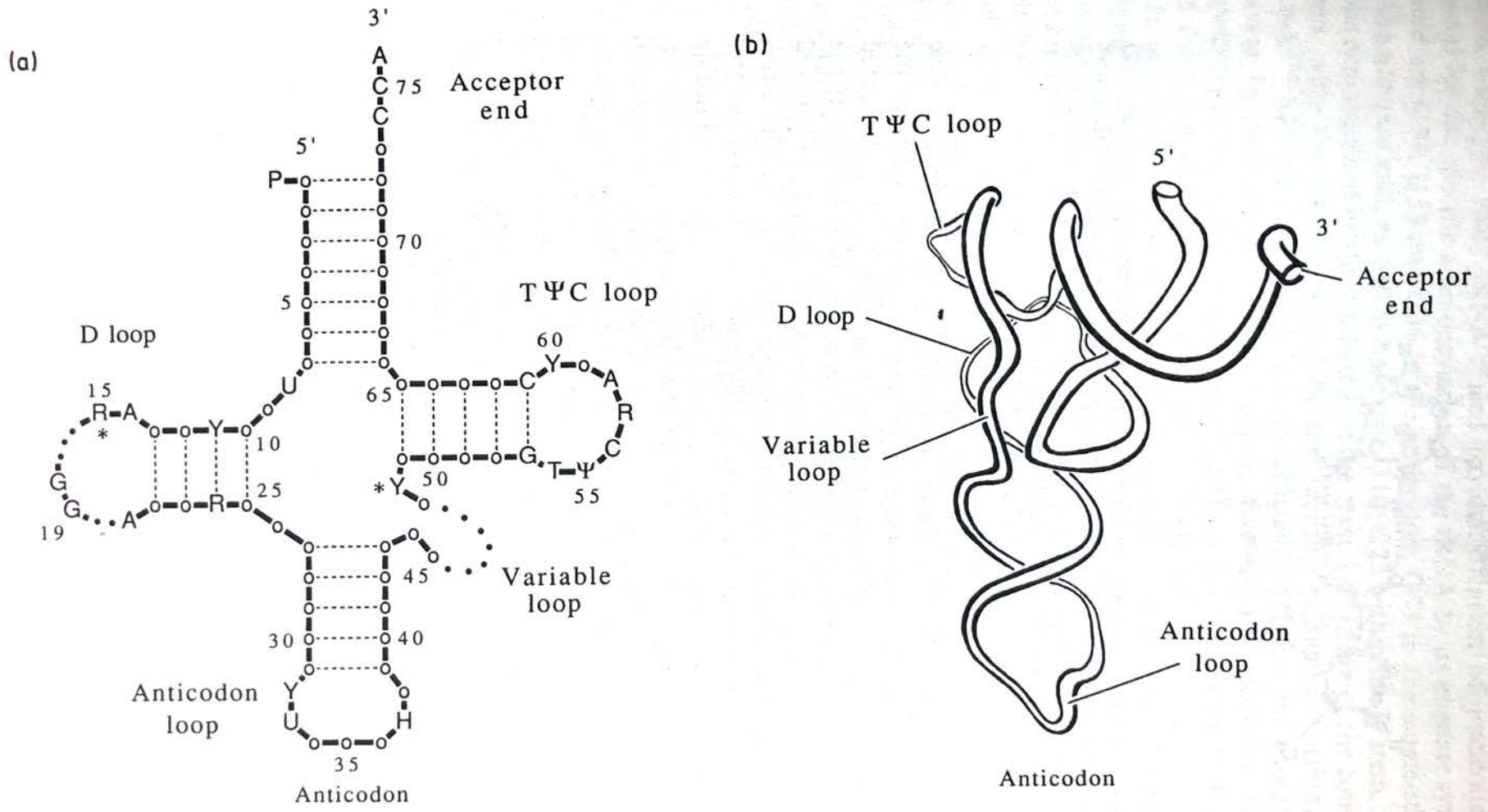

Fig. 18.18. a: Schematic cloverleaf representation of tRNA secondary structure (excluding initiator-tRNAs), showing invariant and semi-invariant bases. Conserved positions between various tRNAs are represented with standard base abbreviations, and variable positions with circles. Dots in the D and TΨC loops designate regions with differing numbers of nucleotides between various tRNAs. Ψ is pseudouridine [C5(H) of U is replaced by N(H)], H is a hypermodified purine, R designates purines, and Y pyrimidines. The starred bases R15 and Y48 are usually complementary, and the base numbering refers to $tRNA^{Phe}$.

b: Schematic representation of the three-dimensional folding of the $tRNA^{Phe}$ backbone as determined by X-ray crystallography. Specific portions of the molecule, such as acceptor stem, anticodon and loops, are designated with the same terminology as used in A. Thicker regions, held in bold, are closer to the reader

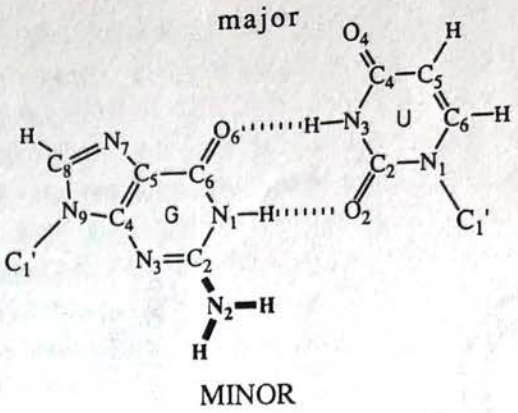

Fig. 18.19. Diagram of a G·U base pair, demonstrating the prominent location of the guanine N(2) amino group (bold) in the minor groove of duplex RNA. The capital letters used for MINOR groove are to remind the reader that this groove is wide and shallow in the A-form duplex, whereas the major groove is narrow and deep, and therefore largely inaccessible

(e.g. [129]). In the case of $tRNA^{Gln}$, these were U35 in the anticodon (see Figure 18.18a for numbering), G73 and base pair 1·72 in the acceptor stem. In the crystal structure of the $tRNA^{Gln}$-synthetase complex, base pair 1·72 is disrupted and the CCA 3′-end of tRNA forms a hairpin which points in the direction of the anticodon. Obviously, the specificity in recognizing the U·A base pair 1·72 is based on the smaller free energy cost of its disruption relative to that of G·C. The bases A76, C75 and G73 are stacked on top of each other, whereas C74 is looped out and interacts with a complementary pocket in the protein that also contains the binding site for the other two substrates, ATP and glutamine. Other extensive contacts between the tRNA and the protein occur at the anticodon. There, bases C34 and G36 are unstacked and base U35 is stacked underneath A37, the stabilizing interactions with the protein supposedly compensating for the loss of stacking energy. Thus, in the $tRNA^{Gln}$ complex, there are specific interactions with bases in single-stranded regions and the minor groove of the acceptor stem. In addition, the crystal structure suggests that base pairs G2·C71 and G3·C70 are important recognition elements. In the complex of $tRNA^{Asp}$ with its synthetase, the conformation of the single-stranded stretch at the acceptor end is quite different from that in the $tRNA^{Gln}$ complex. The fragment GCCA–O(3′)H maintains the helical conformation of the acceptor stem when it is bound to the active site of the protein. In the aspartyl system, the first U·A base pair in the acceptor stem is not disrupted, and the protein contacts this base pair through the major groove at the end of the duplex. Whereas the second and even the third base pair of the acceptor stem are in contact with the synthetase in the $tRNA^{Gln}$ complex, they are not important identity elements in the $tRNA^{Asp}$ complex. The anticodon loop in the aspartyl system undergoes a drastic conformational change. The protein approaches the end of the stem part of tRNA at this site from the major groove. Upon binding of the protein, the anticodon loop moves as much as 20 Å (distance between the phosphate groups of residue U35 of complexed and free tRNA). This movement and unstacking of bases in the anticodon results in extensive protein-RNA interactions. Mutations at specific sites

within wild-type yeast $tRNA^{Asp}$ led to identification of nucleotides which are crucial for the specific aminoacylation of the tRNA by its cognate synthetase [130]. These are G34, U35 and C36 in the anticodon, G73 in the single-stranded stretch at the acceptor end, as well as base pair G10·C25 within the D stem. All these positions are contacted by the synthetase in the crystal structure of the complex, in agreement with the determinations of the identity elements. Furthermore, a yeast $tRNA^{Phe}$ has been converted into an efficient substrate for aspartyl-tRNA synthetase by appropriate exchange of the five identity elements. In general, the specificity of aminoacylation is determined by a few nucleotides and can be transferred from one amino acid to another.

In principle, synthetases could distinguish their cognate tRNA from all others by relying essentially on interactions with the anticodons. Although that site is an important recognition element for a number of synthetases, there are some that do not need interactions with the anticodon for specific aminoacylation. For example, RNA minihelices that correspond to a portion of the acceptor stem are aminoacylated by their cognate alanine and histidine synthetases [131]. The major identity determinant in the charging of short RNA duplexes with alanine is base pair G3·U70 [132]. The experiments showed that the N(2) in the minor groove of the duplex marks the molecule for aminoacylation with alanine (see Figure 18.19).

The key role of particular secondary structural elements of RNA for protein-RNA recognition is also evident from the TAR-Tat system [133]. Binding of the viral Tat protein to its TAR target sequence (transactivation response RNA) is required for replication of the human immunodeficiency virus. TAR includes extensive duplex structure, a trinucleotide pyrimidine bulge and a hairpin with a six-base loop (Figure 18.20a). Both in vitro and in vivo studies indicate that the basic domain of the HIV-1 Tat protein binds specifically at the pyrimidine bulge and the adjacent duplex of TAR [134]. Mutation of the wild-type TAR sequence and the resulting effects on the peptide binding affinity show that base pair G25–C38 is important for maintenance of the secondary structure at the bulge and sequence-specific peptide-RNA interaction [135]. Only RNAs containing two- or three-nucleotide bulges with a U at position 22 (Figure 18.20a) are bound sequence-specifically by Tat-derived peptides. The strong correlation between major-groove accessibility and the affinity and specificity of RNA-interaction with these peptides is consistent with a widening of the major groove at the site of the bulge. Preliminary results of an NMR analysis of TAR show a much wider major groove at the bulge (Figure 18.20b), the edges of base pairs in that groove lying exposed to interacting groups [136]. Presumably, a bulge with just one base does not affect the width of the major groove enough to allow probing of the functional groups of bases, and hence the lack of sequence-specificity in the interaction of peptides with such sequences. The above analysis also revealed the formation of a base pair between C29 and G33 in the six-base loop (dashed line in Figure 18.20a) with the adenine base protruding into space. The structural characteristics of RNA bulges in the TAR-Tat system demonstrate yet another way to get around the restricted accessibility of the major groove in double-stranded RNA.

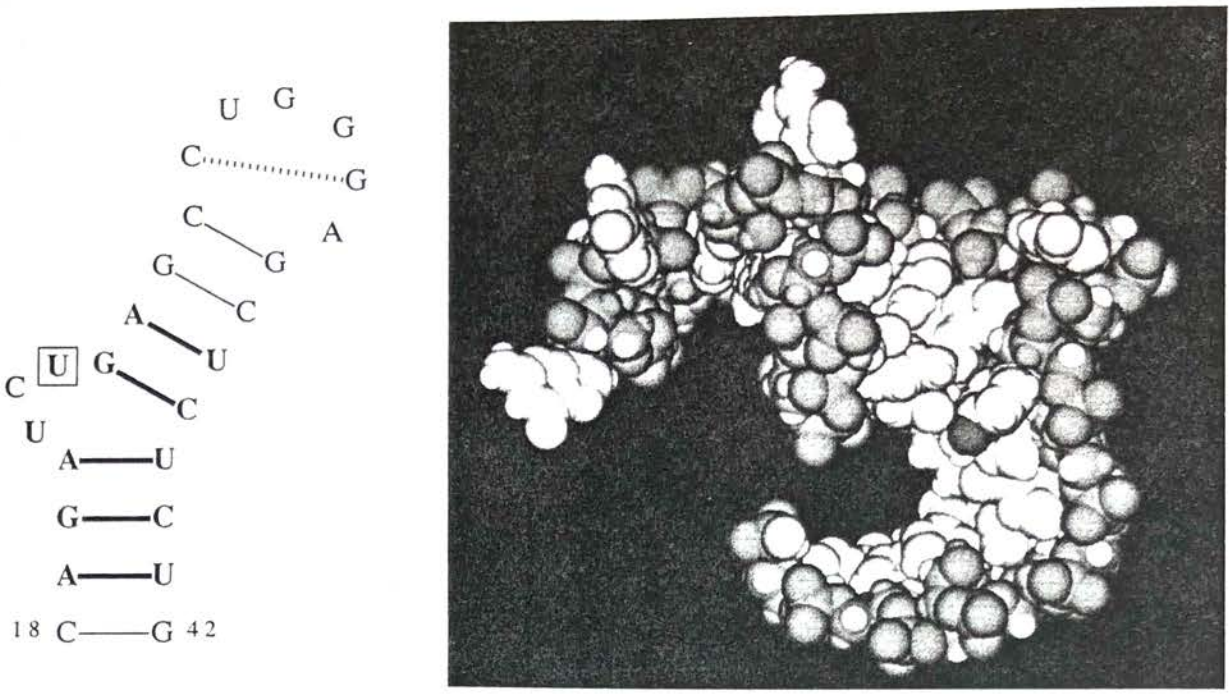

Fig. 18.20. Left: In vitro binding site of the basic subdomain of human immunodeficiency virus type-1 (HIV-1) transactivator protein (Tat) in the transactivation response (TAR) RNA sequence. Only a short portion of the stem region of TAR is shown. Sites where residue modification interferes with Tat binding (the putative binding site) are bold. They comprise part of the pyrimidine bulge and the adjacent duplex, but not the hairpin loop, shown at the top. The boxed base U24 enhances binding when eliminated, whereas modification of C23 does not interfere with peptide binding. The dashed line connecting C and G in the six-base loop indicates the possible formation of a base pair.
Right: Space-filling representation of a portion of the TAR structure, determined by NMR spectroscopy. The deep and narrow major groove of the stem region is located in the lower middle, the three-base bulge is at the upper right, and the six-base loop is at the left. Backbone atoms are dark blue and base atoms are drawn with lighter blue. Residue U22 of the UCU bulge is highlighted in yellow, with a red O(4) atom. The extruded A residue of the loop is also shaded yellow. At present, the position of U24 in the bulge (lying over the phosphate backbone in this view) is not well-defined by NMR. The widening of the major groove at the bulge is apparent from the increased distance between backbone atoms at this site and the acceptor and donor groups of bases lying relatively close to the surface

18.7 Conclusions and Future Prospects

The structures discussed in this review illustrate the progress in our understanding of the double-stranded conformations of DNA and details of interactions between proteins and B-DNA. Many details of the double-helical structure only became accessible through the determination of the three-dimensional structures of synthetic oligonucleotides by X-ray crystallography. Topics of current interest include the exact geometry of the right- and left-handed conformations, the topology of the grooves,

several rules regarding the relation between sequence and conformation, the interactions of ions with DNA, the hydration of DNA and its role in structure stabilization, and the origins of sequence-specificity in drug-DNA interactions. Moreover, crystallographic analyses allow statements to be made about the dynamic properties of molecules in the crystal. Because of the intrinsic rigidity of the double-helical conformation as well as other factors, the changes in geometry of DNA duplexes related to sequence alterations or complex formation are rather subtle, and, in order to observe them, high-resolution data are needed. Because of the small number of sufficiently precise structures and the restricted sequence variation among them, we have at present only a very limited understanding of the sequence dependence of the duplex geometry. In general, DNA has a higher potential for structural variations than implied by the crystallographic work on the double-helical oligomers. Thus, there is almost no structural information on transitions between the various double-helical conformations (e.g. the A-DNA B-DNA junction), as well as on several other conformational states of DNA, such as triple strands (e.g. [137, 138]), quartet formation with G-rich strands [30, 139, 140], and DNAs with altered backbone [141, 142] or base structures [143].

The rich secondary structure present in RNA molecules constitutes an even greater challenge for the crystallographer. Partly because of the lack of sufficient amounts of material, and perhaps also due to somewhat limited crystallization methods, the number of X-ray crystal structures of RNA molecules is frustratingly small. More than fifteen years after its initial structure determination, tRNA remains the only biologically relevant RNA of which we possess a high-resolution crystal structure. The solution structures of an RNA hairpin and an RNA pseudoknot have been studied with NMR [144, 145], but since NMR and X-ray crystallography are complementary approaches, there remains an imminent need for crystal structures of such molecules. To the extent to which NMR assignments and three-dimensional structure determination by NMR are possible without reference to corresponding crystal structures, new topologies can be obtained with NMR methods. NMR also allows a dynamic characterization of molecules under various solution conditions. The assignment problems posed by the very crowded proton spectra of large RNA molecules may be overcome by incorporation of specifically labeled nucleotides into RNA (e.g. ^{15}N). The information gained from simplified spectra of various labeled species can then be combined to determine the structure of the native molecule [136, 146].

The common feature in the crystal structures of DNA complexed to DNA-recognizing proteins is the presence of α helices in the recognition motif. In both the helix-turn-helix motif and the zinc modules, α helices are positioned in the major groove, contacting several bases via amino acid side chains. A recent determination of the met repressor complexed with DNA shows that the β ribbon can also be used for probing the major groove in B-DNA ([147] and references cited therein). The next years will show whether nature uses additional motifs to recognize DNA. At the moment, it is also unclear whether each of these motifs recognizes specific classes of DNA sequences. Characterization of these motifs is a crucial step on the way to the design of sequence-specific probes with applications in many areas.

Acknowledgements

I thank Ms. Lynn Albers for help with the torsion angle analysis of the DNA-intercalator complexes, and Drs. Reinhard V. Gessner and Garry C. King for allowing me to use their figures. Drs. Sohee Park and Shuguang Zhang carefully read the manuscript and made helpful comments. I am grateful to Ms. Alice Colby for her kind help throughout my stay at M.I.T. and to Dr. Alexander Rich for financial support.

References

[1] Arnott, S., Chandrasekaran, R., Selsing, E., in: *Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interaction*, Sundaralingam, M., Rao, S. T., (eds.), University Park Baltimore, **1975**, pp. 577–596
[2] Dickerson, R. E., Drew, H. R., Conner, B., in: *Biomolecular Stereodynamics*, Sarma, R. H. (ed.), Adenine, New York, **1981**, Vol. 1., pp. 1–34
[3] Rich, A., Quigley, G. J., Wang, A. H.-J., in: *Biomolecular Stereodynamics*, Sarma, R. H. (ed.), Adenine, New York, **1981**, Vol. 1, pp. 35–52
[4] Kennard, O.,, Hunter, W. N., *Angew. Chem.*, **1991**, *103*, 1280–1304; *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 1254–1277.
[5] Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, R. M., Fratini, A. V., Kopka, M. L., *Science* **1982**, *216*, 475–485
[6] Conner, B. N., Takano, T., Tanaka, S., Itakura, K., Dickerson, R. E., *Nature* **1982**, *295*, 294–299
[7] Wang, A. H.-J., Fujii, S., van Boom, J. H., Rich, A., *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 3968–3972
[8] Shakked, Z., Rabinovich, D., Kennard, O., Cruse, W. B. T., Salisbury, S. A., Viswamitra, M. A., *J. Mol. Biol.* **1983**, *166*, 183–201
[9] McCall, M., Brown, T., Kennard, O., *J. Mol. Biol.* **1985**, *183*, 385–396
[10] Haran, T. E., Shakked, Z., Wang, A. H.-J. Rich, A., *J. Biomol. Struct. Dynam.* **1987**, *5*, 199–217
[11] Frederick, C. A., Teng, M. K., Quigley, G. J., Coll, M., van der Marel, G. A., van Boom, J. H., Rich, A., Wang, A. H.-J., *Europ. J. Biochem.* **1989**, *181*, 295–307
[12] Jain, S., Zon, G., Sundaralingam, M., *Biochemistry* **1989**, *28*, 2360–2364
[13] Verdaguer, N., Aymami, J., Fernández-Forner, D., Fita, I., Coll, M., Huynh-Dinh, T., Igolen, J., Subirana, J. A., *J. Mol. Biol.* **1991**, *221*, 623–635
[14] Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., Dickerson, R. E., *Nature* **1980**, *287*, 755–758
[15] Privé, G. G., Yanagi, K., Dickerson, R. E., *J. Mol. Biol.* **1991**, *217*, 177–199
[16] Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. A., Rich, A., *Nature* **1979**, *282*, 680–686
[17] Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A., Wang, A. H.-J., *J. Biol. Chem.* **1989**, *264*, 7921–7935
[18] Egli, M., Williams, L. D., Gao, Q., Rich, A., *Biochemistry* **1991**, *30*, 11388–11402
[19] Yanagi, K., Privé, G. G., Dickerson, R. E., *J. Mol. Biol.* **1991**, *217*, 201–214
[20] Nelson, H. C. M., Finch, J. T., Luisi, B. F., Klug, A., *Nature* **1987**, *330*, 221–226

[21] Coll, M., Frederick, C.A., Wang, A.H.-J., Rich, A., *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385–8389
[22] Yoon, C., Privé, G.G., Goodsell, D.S., Dickerson, R.E., *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 6332–6336
[23] DiGabrieli, A.D., Sanderson, M.R., Steitz, T.A., *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 1816–1820
[24] Shakked, Z., Guerstein-Guzikevich, G., Zaytzev, A., Eisenstein, M., Frolow, F., Rabinovich, D., in: *Structure and Methods*, Sarma, R.H., Sarma, M.H. (eds.), Adenine, Schenectady, NY, **1990**, Vol. 3, 55–72
[25] Privé, G.G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M.L., Dickerson, R.E., *Science* **1987**, *238*, 498–504
[26] Heinemann, U., Alings, C., *J. Mol. Biol.* **1989**, *210*, 369–381
[27] Cruse, W.B.T., Salisbury, S.A., Brown, T., Cosstick, R., Eckstein, F., Kennard, O., *J. Mol. Biol.* **1986**, *192*, 891–905
[28] Levitt, M., Warshel, A., *J. Am. Chem. Soc.* **1978**, *100*, 2607–2613
[29] Saenger, W., *Principles of Nucleic Acid Structure*, Springer, Berlin, **1984**
[30] Kang, C.H., Zhang, X., Ratliff, R., Moyzis, R., Rich, A., *Nature* **1992**, *356*, 126–131
[31] Arnott, S., Hukins, D.W.L., Dover, S.D., Fuller, W., Hodgson, A.R., *J. Mol. Biol.* **1973**, *81*, 107–122
[32] Dock-Bregeon, A.C., Chevrier, B., Podjarny, A., Moras, D., deBear, J.S., Gough, G.R., Gilham, P.T., Johnson, J.E., *Nature* **1988**, *335*, 375–378
[33] Holbrook, S.R., Cheong, C., Tinoco Jr., I., Kim, S.-H., *Nature* **1991**, *353*, 579–581
[34] Milman, B., Langridge, R., Chamberlin, M.J., *Proc. Natl. Acad. Sci. USA* **1967**, *57*, 1804–1810
[35] Wang, A.H.-J., Fujii, S., van Boom, J.H., van der Marel, G.A., van Boeckel, S.A.A., Rich, A., *Nature* **1982**, *299*, 601–604
[36] Egli, M., Usman, N., Rich, A., *Biochemistry*, **1993**, *32*, 3221–3237
[37] Egli, M., Usman, N., Zhang, S., Rich, A., *Proc. Natl. Acad. Sci. USA* **1991**, *89*, 534–538
[38] Dickerson, R.E. et al., *Nucleic Acids Res.* **1989**, *17*, 1797–1803
[39] Calladine, C.R., *J. Mol. Biol.* **1982**, *161*, 343–352
[40] Dickerson, R.E., *J. Mol. Biol.* **1983**, *166*, 419–441
[41] Chevrier, B., Dock, A.C., Hartmann, B., Leng, M., Moras, D., Thuong, M.T., Westhof, E., *J. Mol. Biol.* **1986**, *188*, 707–719
[42] Gessner, R.V., *Dissertation*, Freie Universität Berlin, **1989**, 157–166
[43] Bancroft, D., Williams, L.D., Rich, A., Egli, M., *Biochemistry*, in press
[44] Saenger, W., Hunter, W., Kennard, O., *Nature* **1986**, *324*, 385–388
[45] Jeffrey, G.A., Saenger, W., *Hydrogen Bonding in Biological Structures*, Springer, Berlin, **1991**
[46] Wang, A.H.-J., Quigley, G.J., Kolpak, F.J., van der Marel, G.A., van Boom, J.H., Rich, A., *Science* **1981**, *211*, 171–176
[47] Gessner, R.V., Quigley, G.J., Wang, A.H.-J., van der Marel, G.A., van Boom, J.H., Rich, A., *Biochemistry* **1985**, *24*, 237–240
[48] Basu, A.K., Essigmann, J.M., *Chem. Res. Toxicol.* **1988**, *1*, 1–18
[49] Walker, G.C., *Ann. Rev. Biochem.* **1985**, *54*, 425–457
[50] Sancar, A., Sancar, G.B., *Ann. Rev. Biochem.* **1988**, *57*, 29–67
[51] Kennard, O., in: *Structure and Expression*, Sarma, R.H., Sarma, M.H. (eds.), Adenine, New York, **1988**, Vol. 2, 1–25
[52] Brown, T., Kennard, O., Kneale, G., Rabinovich, D., *Nature* **1985**, *315*, 604–606
[53] Kneale, G., Brown, T., Kennard, O., Rabinovich, D., *J. Mol. Biol.* **1986**, *186*, 805–814
[54] Hunter, W.N., Brown, T., Kneale, G., Anand, N.N., Rabinovich, D., Kennard, O., *J. Biol. Chem.* **1987**, *262*, 9962–9970
[55] Kennard, O., *J. Biomol. Struct. Dynam.* **1985**, *3*, 205–225
[56] Ho, P.S., Frederick, C.A., Quigley, G.J., van der Marel, G.A., van Boom, J.H., Wang, A.H.-J., Rich, A., *EMBO J.* **1985**, *4*, 3617–3623
[57] Brown, T., Kneale, G., Hunter, W.M., Kennard, O., *Nucleic Acids Res.* **1986**, *14*, 1801–1809
[58] Hunter, W.N., Brown, T., Anand, N.N., Kennard, O., *Nature* **1986**, *320*, 552–555
[59] Hunter, W.N., Brown, T., Kennard, O., *J. Biomol. Struct. Dynam.* **1986**, *4*, 173–191

[60] Webster, G. D., Sanderson, M. R., Skelly, J. V., Neidle, S., Swann, P. F., Li, B. F., Tickle, I. J., *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 6693–6697
[61] Corfield, W. R., Hunter, W. N., Brown, T., Robinson, P., Kennard, O., *Nucleic Acids Res.* **1987**, *15*, 7935–7948
[62] Crick, F. H. C., *J. Mol. Biol.* **1966**, *19*, 548–555
[63] Aboul-ela, F., Koh, D., Tinoco Jr., I., Martin, F. H., *Nucleic Acids Res.* **1985**, *13*, 4811–4824
[64] Werntges, H., Steger, G., Riesner, D., Fritz, H.-J., *Nucleic Acids Res.* **1986**, *14*, 3773–3790
[65] Kramer, B., Kramer, W., Fritz, H.-J., *Cell* **1984**, *38*, 879–887
[66] Fersht, A. R., Knill-Jones, J. W., Tsui, W.-C., *J. Mol. Biol.* **1982**, *156*, 37–51
[67] Leonard, G. A., Thomson, J., Watson, W. P., Brown, T., *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 9573–9576
[68] Ginell, L., Narendra, N., Jones, R. A., Berman, H. M., *Biophys. J.* **1990**, *57*, 452
[69] Fratini, A. V., Kopka, M. L., Drew, H. R., Dickerson, R. E., *J. Biol. Chem.* **1982**, *257*, 14686–14707
[70] Dickerson, R. E., Drew, H. R., *J. Mol. Biol.* **1981**, *149*, 761–786
[71] Drew, H. R., Dickerson, R. E., *J. Mol. Biol.* **1981**, *151*, 535–556
[72] Kopka, M. L., Fratini, A. V., Drew, H. R., Dickerson, R. E., *J. Mol. Biol.* **1983**, *163*, 129–146
[73] Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., Dickerson, R. E., *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376–1380
[74] Pjura, P. E., Grzeskowiak, K., Dickerson, R. E., *J. Mol. Biol.* **1987**, *197*, 257–271
[75] Coll, M., Frederick, C. A., Wang, A. H.-J., Rich, A., *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385–8389
[76] Brown, D. G., Sanderson, M. R., Skelly, J. V., Jenkins, T. C., Brown, T., Garman, E., Stuart, D. I., Neidle, S., *EMBO J.* **1990**, *9*, 1329–1334
[77] Teng, M., Usman, N., Frederick, C. A., Wang, A. H.-J., *Nucleic Acids Res.* **1988**, *16*, 2671–2690
[78] Coll, M., Aymami, J., van der Marel, G. A., van Boom, J. H., Rich, A., Wang, A. H.-J., *Biochemistry* **1989**, *28*, 310–320
[79] de C. T. Carrondo, M. A. A. F., Coll, M., Aymami, J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., Rich, A., *Biochemistry* **1989**, *28*, 7849–7859
[80] Quintana, J. R., Lipanov, A. A., Dickerson, R. E., *Biochemistry* **1991**, *30*, 10294–10306
[81] Arnott, S., Hukins, D. W. I., *Biochem. Biophys. Res. Comm.* **1972**, *47*, 1504–1509
[82] Gupta, G., Bansal, M., Sasisekharan, V., *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 6486–6490
[83] Dickerson, R. E., Kopka, M. L., Drew, H. R., in: *Conformation in Biology*, Srinivasan, R., Sarma, R. H. (eds.), Adenine, New York, **1982**, pp. 227–257
[84] Quigley, G. J., Wang, A. H.-J., Ughetto, G., van der Marel, G. A., van Boom, J. H., Rich, A., *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 7204–7208
[85] Frederick, C. A., Williams, L. D., Ughetto, G., van der Marel, G. A., van Boom, J. H., Rich, A., Wang, A. H.-J., *Biochemistry* **1990**, *29*, 2538–2549
[86] Williams, L. D., Egli, M., Ughetto, G., van der Marel, G. A., van Boom, J. H., Quigley, G. J., Wang, A. H.-J., Rich, A., Frederick, C. A., *J. Mol. Biol.* **1990**, *215*, 313–320
[87] Williams, L. D., Egli, M., Gao, Q., Bash, P., van der Marel, G. A., van Boom, J. H., Rich, A., Frederick, C. A., *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 2225–2229
[88] Egli, M., Williams, L. D., Frederick, C. A., Rich, A., *Biochemistry* **1991**, *30*, 1364–1372
[89] Liaw, Y.-C., Gao, Y.-G., Robinson, H., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., *Biochemistry* **1989**, *28*, 9913–9918
[90] Gao, Y.-G., Liaw, Y.-C., Robinson, H., Wang, A. H.-J., *Biochemistry* **1990**, *29*, 10307–10316
[91] Williams, L. D., Frederick, C. A., Ughetto, G., Rich, A., *Nucleic Acids Res.* **1990**, *18*, 5533–5541
[92] Gao, Q., Williams, L. D., Egli, M., Rabinovich, D., Chen, S.-H., Quigley, G. J., Rich, A., *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2422–2426
[93] Quigley, G. J., Ughetto, G., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., Rich, A., *Science* **1986**, *232*, 1255–1258
[94] Williams, L. D., Egli, M., Gao, Q., Rich, A.: in: *Structure & Function: Proceedings of the Seventh Conversation in Biomolecular Stereodynamics*, Sarma, R. H., Sarma, M. H. (eds.), Adenine, New York, **1992**, pp. 107–126

[95] Shieh, H.-S., Berman, H.M., Dabrow, M., Neidle, S., *Nucleic Acids Res.* **1980**, *8*, 85–97
[96] Jain, S.C., Sobell, H.M., *J. Biomol. Struct. Dynam.* **1984**, *1*, 1179–1194
[97] Wang, A.H.-J., Liaw, Y.-C., Robinson, H., Gao, Y.-G., in: *Molecular Basis of Specificity in Nucleic Acid-Drug Interactions*, Pullman, B., Jortner, J. (eds.), Kluwer Academic, Dordrecht, **1990**, 1–21
[98] Seeman, N.C., Rosenberg, J.M., Rich, A., *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 804–808
[99] McClarin, J.A., Frederick, C.A., Wang, B.-C., Greene, P., Boyer, H.W., Grable, J., Rosenberg, J.M., *Science* **1986**, *234*, 1526–1541
[100] Kim, Y., Grable, J.C., Love, R., Greene, P., Rosenberg, J.M., *Science* **1990**, *249*, 1307–1309
[101] Jordan, S.R., Pabo, C.O., *Science* **1988**, *242*, 893–899
[102] Brennan, R.G., Roderick, S.L., Yoshinori, T., Matthews, B.W., *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8165–8169
[103] Anderson, J.E., Ptashne, M., Harrison, S.C., *Nature* **1987**, *326*, 846–852
[104] Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M., Harrison, S.C., *Science* **1988**, *242*, 899–907
[105] Wolberger, C., Dong, Y., Ptashne, M., Harrison, S.C., *Nature* **1988**, *335*, 789–795
[106] Otwinowski, Z., Schevitz, R.W., Zhang, R.-G., Lawson, C.L., Joachimiak, A., Marmorstein, R.Q., Luisi, B.F., Sigler, P.B., *Nature* **1988**, *335*, 321–329
[107] Kissinger, C.R., Liu, B., Martin-Bianco, E., Kornberg, T.B., Pabo, C.O., *Cell* **1990**, *63*, 579–590
[108] Pavletich, N.P., Pabo, C.O., *Science* **1991**, *252*, 809–817
[109] Pabo, C.O., Sauer, R.T., *Ann. Rev. Biochem.* **1984**, *53*, 293–321
[110] Aggarwal, A.K., Harrison, S.C., *Ann. Rev. Biochem.* **1990**, *59*, 933–969
[111] Harrison, S.C., *Nature* **1991**, *353*, 715–719
[112] Branden, C., Tooze, J., *Introduction to Protein Structure*, Garland, New York, **1991**
[113] Schultz, S.C., Shields, G.C., Steitz, T.A., *Science* **1991**, *253*, 1001–1007
[114] Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R., Sigler, P.B., *Nature* **1991**, *352*, 497–505
[115] Rafferty, J.B., Somers, W.S., Saint-Girons, I., Phillips, S.E.V., *Nature* **1989**, *341*, 705–710
[116] Clarke, N.D., Beamer, L.J., Goldberg, H.R., Berkower, C., Pabo, C.O., *Science* **1991**, *254*, 267–270
[117] Laughon, A., *Biochemistry* **1991**, *30*, 11357–11366
[118] O'Shea, E.K., Klemm, J., Kim, P.S., Alber, T., *Science* **1991**, *254*, 539–544
[119] Rich, A., in: *Proceedings of the Robert A. Welch Foundation Conferences on Chemical Research XXIX. Genetic Chemistry: The Molecular Basis of Heredity*, Houston, **1985**, pp. 11–34
[120] Nordheim, A., Herrera, R.E., Rich, A., *Nucleic Acids Res.* **1987**, *15*, 1661–1677
[121] Wyatt, J.R., Puglisi, J.D., Tinoco Jr., I., *BioEssays* **1989**, *11*, 100–106
[122] Chastain, M., Tinoco Jr., I., in: *Progress in Nucleic Acid Research and Molecular Biology*, Cohn, W.E., Moldave, K. (eds.), Academic Press, San Diego, **1991**, Vol. 41, pp. 131–178
[123] Rich, A., RajBhandary, U.L., *Ann. Rev. Biochem.* **1976**, *45*, 805–860
[124] Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F.C., Klug, A., *Nature* **1974**, *250*, 546–551
[125] Rich, A., Kim, S.-H., *Sci. Am.* **1978**, *238*, 52–62
[126] Holbrook, S.R., Cheong, C., Tinoco Jr., I., Kim, S.-H., *Nature* **1991**, *353*, 579–581
[127] Rould, M.A., Perona, J.J., Söll, D., Steitz, T.A., *Science* **1989**, *246*, 1135–1142
[128] Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C., Moras, D., *Science* **1991**, *252*, 1682–1689
[129] Schulman, L.H., Pelka, H., *Biochemistry* **1985**, *24*, 7309–7314
[130] Pütz, J., Puglisi, J.D., Florentz, C., Giegé, R., *Science* **1991**, *252*, 1696–1699
[131] Schimmel, P., *FASEB J.* **1991**, *5*, 2180–2187
[132] Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R., Schimmel, P., *Science* **1991**, *253*, 784–786
[133] Frankel, A.D., Mattaj, I.D., Rio, D.C., *Cell* **1991**, *67*, 1041–1046
[134] Weeks, K.M., Ampe, C., Schultz, S.C., Steitz, T.A., Crothers, D.M., *Science* **1990**, *249*, 1281–1285

[135] Weeks, K.M., Crothers, D.M., *Cell* **1991**, *66*, 577–588
[136] King, G.C., Rice University, Houston, personal communication
[137] Moser, H.E., Dervan, P.B., *Science* **1987**, *238*, 645–650
[138] Pilch, D.S., Levenson, C., Shafer, R.H., *Biochemistry* **1991**, *30*, 6081–6087
[139] Blackburn, E., *Nature* **1991**, *350*, 569–572
[140] Smith, F.W., Feigon, J., *Nature* **1992**, *356*, 164–168
[141] Schneider, K.C., Benner, S.A., *Tetrahedron Lett.* **1990**, *31*, 335–338
[142] Eschenmoser, A., Dobler, M., *Helv. Chim. Acta* **1992**, *75*, 218–259
[143] Piccirilli, J.A., Krauch, T., Moroney, S.E., Benner, S.A., *Nature* **1990**, *343*, 33–37
[144] Cheong, C., Varani, G., Tinoco Jr., I., *Nature* **1990**, *346*, 680–682
[145] Puglisi, J.D., Wyatt, J.R., Tinoco Jr., I., *J. Mol. Biol.* **1990**, *214*, 437–453
[146] Nikonowicz, E.P., Pardi, A., *Nature* **1992**, *355*, 184–186
[147] Kim, S.-H., *Science* **1992**, *255*, 1217–1218
[148] Marmorstein, R., Carey, M., Ptashne, M., Harrison, S.C., *Nature* **1992**, *356*, 408–414