Crystal structures of nucleic acids and their drug complexes

Stephen Neidle and Christine M. Nunn

The CRC Biomolecular Structure Unit, The Institute of Cancer Research, Sutton, Surrey, UK SM2 5NG

1 Introduction

The proposal for the structure of the DNA double helix in 1953[1] is generally considered to be the point at which genetic phenomena started to be understandable in molecular and chemical terms. Indeed Watson and Crick were themselves directly dependent for their success on the chemical knowledge of nucleotides from the work of Todd and his school, as well as relying heavily on X-ray crystallographic studies of nucleobases and nucleosides.[2]

The Watson–Crick structure is a model based on the fibre-diffraction data of Franklin and Wilkins. As such, it represents a low-resolution structure averaged over all sequences and conformations. More detailed descriptions of DNA structure, and of its wide variability, had to await the availability of pure quantities of defined oligonucleotide sequences in the 1970s. These have enabled single crystals of a wide range of DNA (and now RNA) structural types to be obtained and their structures solved, in a large part without recourse to any assumptions about particular models. In addition, the crystal structures of a large number of biologically-relevant drug–nucleic acid and protein–nucleic acid complexes have now been solved. The latter is an area of especially rapid growth, and will not be discussed here.[3] This review will survey the significant crystal structures of nucleic acids and their drug complexes published over the past few years, which have seen an exponential growth in the oligonucleotide structures reported in the literature. The attention of readers is drawn to the comprehensive and up-to-date compilation of all these nucleic acid single-crystal structures in the Nucleic Acid Database,[4] with 618 entries as of February 1997. Background information on nucleic acid structures, together with reviews on such topics as fibre diffraction from polynucleotides and NMR studies in solution, is available elsewhere.[5–8]

As with all crystal structures, considerations of accuracy and reliability are important when analysing and comparing structures. Nucleic acid structures are generally not at atomic resolution, with 2.0–2.5 Å being typical. This means that their crystallographic refinement requires the use of constraints and restraints on geometry and conformation. Refinement procedures themselves, as well as the individual geometric parameters used, have significantly changed over the past few years, which affects any detailed comparison of structures.[9] The X-PLOR program[10] is now almost universally the choice for all nucleic acid-containing structures, involving the use of an empirical force-field. This is generally considered to be well-parameterised for nucleic acid duplexes, but may not be adequate for the increasing number of structures with non-standard base–base interactions and folds, especially of RNA molecules. Extra parameters are invariably required for nucleic acid–drug complexes; however these are rarely available in the primary literature so it is not straightforward to assess the validity of such parameterisations.

2 DNA native duplexes

2.1 B-DNA structures—DNA close to the physiological state?

Fibre-diffraction studies of helical natural and synthetic polynucleotides have characterised their structures in terms of precisely-repetitious mono- or di-nucleotide units, with only small variations being observed with particular sequences. The structure of the first fully base-paired full turn of double helix to be determined by single-crystal methods,[11] using multiple isomorphous phasing, was that of the sequence d(CGCGAATTCGCG), at 1.9 Å resolution. This structure shows, by contrast, a number of sequence-dependent features, such as a narrow minor groove in the AT region, high propeller twists for the A·T base pairs and considerable variation in base and base-step helical parameters such as helical twist and roll. These features are sensitive to the protocols used in the crystallographic refinement,[12] which further underlines the inadvisability of deriving general conclusions about the detailed features of sequence-dependent DNA structure from one of even a few crystal structures. The AT region of the minor groove in d(CGCGAATTCGCG) was found to contain an ordered array of water molecules, the 'spine of hydration', whose existence has more recently been confirmed by NMR spectroscopic[13] and simulation[14] studies. There is an increasing realisation of the importance of structured water molecules, not only in stabilising particular aspects of DNA (and RNA) structure, but in acting as probes for nucleic acid recognition.[15]

A large number of variants[16] of this dodecamer sequence, mostly with small changes in the central 4–6 base pairs, have subsequently been analysed. Almost all crystallise in the same space group ($P2_12_12_1$), with interdigitation of adjacent helices in the crystal lattice and involvement of the terminal two base pairs at each end of an individual duplex. The fact that the central 6–8 base pair region in this packing arrangement is unaffected by adjacent molecules in the lattice, suggests that it is suitable for systematic studies of features occurring in this region, such as base mismatches and drug binding.[17] It can be argued that the dodecamer duplex is less suitable, at least on its own, for comparative studies of sequence-dependent features, since ideally one would wish to observe a particular sequence type in more than one crystallographic context in order for

lattice packing effects to be minimised.[18] An initial result of surveys which included decamer as well as dodecamer structures (see below) was the finding that there appears to be considerable variability in the geometries of many base steps and individual short sequences, and therefore that a search for sequence-dependent features is pursuing an illusory goal. It has even been suggested that the variations in structural features merely represent distortions arising from crystal packing forces. However a more realistic interpretation is that the variability observed for many features in these B-DNA structures, are real effects, and represent their degrees of flexibility. Analysis of structure and conformation for ensembles of structures *via* the Nucleic Acid Database is an effective approach. It has been used to show[19] that phosphodiester backbone torsion angles in the higher resolution sub-set of B-DNA oligonucleotide structures cluster in discrete regions. The majority of these have been previously noted in studies of individual structures, but several new clusters are now apparent, which suggest hitherto unsuspected structural correlations and modes of concerted flexibility between backbone torsion angles.

2.1.1 DNA bending

The dodecamer duplexes reported to date (apart from those involving mismatches or bound drug), are almost all of the type d(CGCX$_6$GCG), where X$_6$ is an AT-containing sequence, and where the overall sequence is self-complementary. Structures with runs of both alternating and non-alternating A·T base pairs have been reported. The structure[20] of the non-self-complementary sequence d(CGCGAAAAAACG), crystallised with the sequence d(CGTTTTTTCGCG), is of particular interest. It is non-isomorphous with almost all other dodecanucleotide structures, even though it is still in the same space group ($P2_12_12_1$). The extended AT tract has typically narrow minor groove and high propeller-twisted A·T base pairs. The helix has an overall bend of 30° in the major groove direction, which is somewhat distinct from the bending towards the major groove observed in the crystal structures of the various members of the d(CGCGAATTCGCG) duplex family. All of these crystal structures appear to differ from the phased bending of A-tracts in solution, usually within rather longer sequences, in which bending has been inferred to occur towards the minor groove direction.[21]

The definition of bending in molecular terms for DNA containing A-tracts, continues to be controversial. The contrast in interpretations from solution and crystallographic studies has been ascribed to the effects of the hydrophobic alcohols such as 2-methylpentane-2,4-diol (MPD) commonly used in the crystallisation of oligonucleotides.[22] It has been suggested that MPD has the ability to markedly decrease DNA curvature, in accord with its effect of reducing the anomalous behaviour of A-tract DNA in gel retardation experiments. This interpretation has been challenged in a comprehensive study[23] of all the crystal structures which show bending, where it is pointed out that, even though MPD does indeed have some effect on bending, it does not involve simple dehydration around the DNA, and moreover does not actually shed any light on the molecular basis of the phenomenon. There is no correlation between the degree of bending observed in the crystal, and the MPD concentration. These authors conclude that the crystallographic observations are consistent with bending occurring outside the A-tracts themselves, which remain straight. The details of the structural changes at these interface sequences have been suggested by several crystal structures,[24–26] although none with true phased A-tract sequences have as yet have been reported.

A comparative study[27] of four oligonucleotide crystal structures, combined with gel retardation experiments, has provided further support for A-tract models with a narrow minor groove, high propeller twist for the A·T base pairs, cross-strand major groove bifurcated hydrogen bonds. In all these structures, the A-tract itself is straight, with bending at the flanking sequences. It is clear that there is no unique direction of the bending, and DNA sequences will deform in ways that are dictated by environment and the nature of the molecules with which it interacts. This is shown on the one hand by the consistent bending towards major groove directions shown by the oligonucleotide crystal structures, indicating that this mode of bending is accessible to even the comparatively lattice forces in a crystalline environment. On the other hand, it is typical for the DNA in protein–DNA complexes[27] to bend in the minor groove direction, suggesting that this type of bending requires the greater enthalpy of interaction with a protein. An extreme example of DNA bending has been observed in oligonucleotide complexes with the TATA-box binding protein,[28] where the DNA is bent by *ca.* 80°, but now towards the major groove. This is accompanied by unwinding so that the DNA most affected assumes a non-B-DNA A-like conformation, which can be modelled by simple deformations involving local conformation changes.[29]

2.1.2 Sequence-dependent DNA structure

The relevance of crystal structures of native oligonucleotides containing gene-regulatory protein recognition sequences, to the structures of the protein–DNA complexes, has been demonstrated in several studies. The crystal structure[30] of the *trp* operator/repressor complex incorporates the six base-pair recognition site d(ACTAGT). The crystal structure[31] of the native decanucleotide duplex d(CCACTAGTGG), shows features of structure and hydration directly analogous to those in the protein complex. For example, the depth and contours of the major groove are similar for this sequence in both native and protein-bound states, as is the pattern of base hydration. Ten hydration sites in the major groove are conserved in the two states; the three which mediate critical protein–DNA contacts also have conserved hydrogen-bonding geometries. It is suggested that these features are intrinsic to this particular DNA sequence. On the other hand, a study[32] of a constituent of this sequence, the tetranucleotide d(CTAG), which occurs in both the *trp* and *met* repressor–operator complexes, has shown that when compared with its structure in the native sequence d(CTCTAGAG), its conformation is somewhat variable. This variability is most pronounced at the central TpA step.

The implication of these comparative studies is that the details of conformation, helical parameters and hydration at particular sequences, as seen in native oligonucleotide crystal structures, can have relevance to the situations in protein complexes. However the relationship between sequence and structure is complex, and still only imperfectly understood at the native DNA level, in spite of much effort over the past 15 years. There have been a large number of studies which have attempted to correlate sequence-dependent structural features in native oligonucleotide structures. This initially focused on the ten distinct dinucleotide steps, but as an increasing number of oligonucleotide structures have been determined, it has been realised that the essential minimal unit of structure description is the tetranucleotide sequence, of which there are 134 variants.[33, 34] Even at the dinucleotide level, it is apparent that only some XpY steps show consistent behaviour. This has been observed where a particular sequence has been crystallised in more than one packing arrangement. For example, the sequence d(CCAACITTGG) (where I is the non-natural nucleoside inosine) occurs in both monoclinic and trigonal space groups.[35] The helical twists at the CpA and TpG dinucleotide steps are nearly 15° greater for the monoclinic structure. The TpA step has also been found to have a variable geometry, as observed[36] in two structures [of d(CGCTAGCG) and d(CGCTCTAGAGCG)].

These observations of variability in structural features have been interpreted as being due to crystal packing effects, although there is little evidence to demonstrate this for

particular runs of sequence. Rather, these observations are showing the inherent flexibility of some dinucleotide and tetranucleotide steps, which when observed in different structures, are sampled at diverse points on their broad energy surfaces. This interpretation suggests that consideration of the available ensemble of structures, as has been done for backbone torsion angles,[18] could provide a fuller picture of the flexibility available to the different steps in an oligonucleotide. An examination[37, 38] of 60 native structures, on the basis solely of the ten dinucleotide steps, has derived an overall kinematic classification involving base-pair slide, roll and helical twist. This analysis has confirmed and extended predictions made on the basis of empirical force-field calculations.[39] These emphasise,[40] on the one hand, the importance of electrostatic contributions to those relative motions of bases (shift and slide) which retain them parallel to each other, and on the other, van der Waals interactions which primarily contribute to those motions (such as roll, tilt and slide) which do not retain base parallelism. A comparative analysis of dinucleotide steps is valid since, with the exception of the under-represented ApG step, the available database of structures is now sufficiently representative of all nine other steps for statistically meaningful conclusions to be drawn. However, it should be borne in mind that such an analysis necessarily ignores nearest-neighbour and longer-range effects, since few of the 134 tetranucleotide steps are represented in the oligonucleotide database. The available evidence is that these effects (for example, the influence of an adjacent A-tract) can sometimes dominate structure at a particular sequence locus.[41]

2.1.3 Non-standard B-DNA structures

The crystal packing motifs of many B-DNA oligonucleotides involves either the end-to-end stacking of helices (many decanucleotide structures) or the inter-digitating of the ends of helices (in almost all dodecanucleotide structures). A notable sub-class of the former has the end-to-end helices packed not in side-by-side parallel arrangements with respect to each other, but inclined at 40–60° to form crossed helices. This motif has been recognised as being a model for four-way junctions envisaged to occur during genetic recombination. It involves groove–backbone intermolecular close contacts[42] (Fig. 1), involving the major groove of one helix and phosphate groups from another.[43] Molecular modelling has been used to generate plausible four-way junctions models from these structures.[44] The sequence d(CGCAATTGCG) forms distinct crystal structures, dependent on environmental conditions. One is of a fully base-paired decamer duplex,[44] whereas the other has the 3′ and 5′ terminal nucleotides swung out from the helix and hydrogen bonding in the grooves of symmetry-related duplexes.[45] The crystal structure of the non-self-complementary decamer sequence d(CGACGATCGT) shows a helical octamer duplex stem formed by d(ACGATCGT) and its complement, together with a 5′-d(CG) sticky end.[46]

2.2 A-DNA structures

The A polymorph of double-helical DNA was identified in early fibre-diffraction studies, as being formed under conditions of low relative humidity. A-type structures are characterised by having base pairs inclined with respect to the helix axis and significantly displaced from it, thus changing the dimensions of both major and minor grooves compared to B-type structures. The biological significance of the A family of DNA structures continues to be controversial. The determination of the structure of the TATA-box protein–DNA complex, which has shown that deforming a region of a B-DNA structure into an A form produces a widened minor groove and an overall bend of the DNA.[28, 29] This suggests a role in protein-induced bending, and that A-type structures are most likely to exist when forced to by external factors such as crystal packing or protein binding. It is striking that no NMR studies of native DNA duplexes have revealed the existence

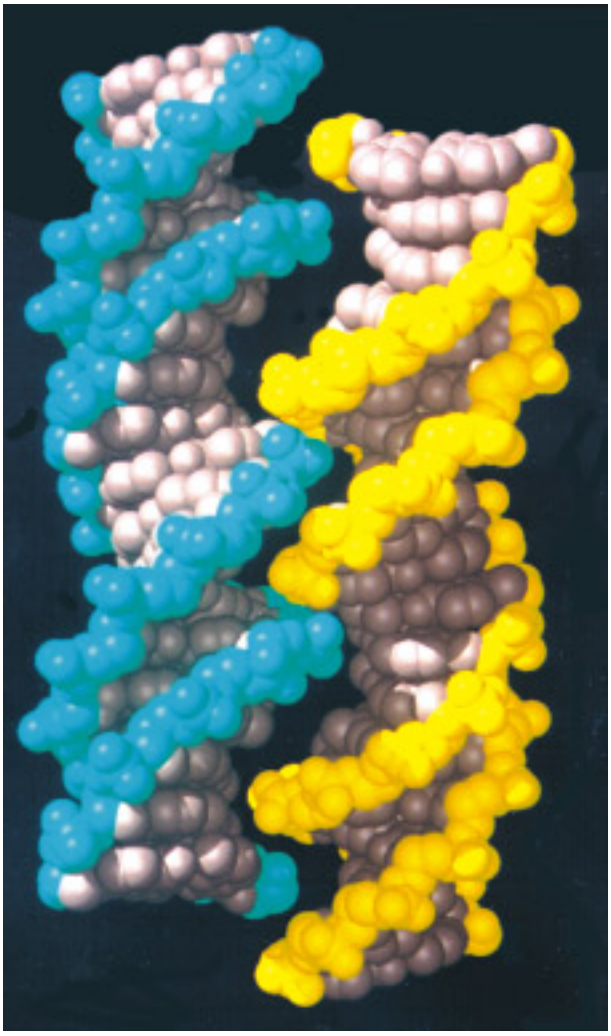

Fig. 1 Groove–backbone interactions in the crystal structure of the DNA decamer d(CGCAATTGCG) (ref. 44)

of the A form in solution, *i.e.* when removed from these factors.

The A family has been observed in numerous crystal structures of self-complementary octanucleotides, most likely as a consequence of specific crystal packing constraints,[47–49] as well as in a number of decanucleotides with a high proportion of C·G base pairs. Considerable variability, in particular, in groove width, has been observed in these structures. The packing motifs of A-DNA oligonucleotides invariably involve interdigitation of the terminal base pairs from one duplex into the minor groove of an adjacent one in the lattice. The importance of crystallographic studies on this family of oligonucleotides is that they provide insight into the range of structures, and hence the flexibility of the A-type structural class, which in turn illuminates the mechanisms whereby some gene-regulatory proteins deform DNA.

Several polymorphic crystal structures of d(CCGGGC CCGG)[50] and two cytosine-methylated analogues have been determined, all of which require the polycationic amine spermine for crystallisation, and the more compact show ordered spermine molecules in the crystal structures. In common with the A-DNA octamer crystal structures, conformations of most of these polymorphs differ significantly from that of classic fibre-diffraction A-DNA. For the three polymorphs of d(CCGGCC^{5}meCGG), the number of residues per complete turn of helix ranges from 10.7 to 11.6, and the average inclination of base pairs to the helix axis ranges from 10.7 to 18.2°. The hydration of A-DNA structures has been found to

be sequence-dependent,[51] and the superior hydration of CpG base pairs compared to TpA correlates with differences in helical twist and base-pair roll.[52]

The sequence d(AGGCATGCCT) forms an unprecedented structure,[53] with the central eight nucleotides base-pairing with a symmetry-related strand to form an A-form octamer duplex. The terminal bases are swung away from this short helix to form A·T base pairs with other symmetry-related molecules, forming an overall infinite chain-like arrangement.

3 DNA base-mismatched structures at the duplex level

DNA bases can be covalently modified by the alkylating effects of a wide range of chemicals. The resulting base lesions are mutagenic, and can ultimately produce cell transformation and cancer, if they are not correctly repaired by cellular repair enzymes. The recognition of lesions by these enzymes is likely to be based on the structural differences resulting from them. A number of crystal structures have examined DNA sequences with guanine methylated at the O6 position, a consequence of exposure to carcinogenic agents such as *N*-nitroso-*N*-methylurea and *N*-methyl-*N*-nitrosoguanidine.

(O6Me)Guanine···cytosine base pairing has been observed in a left-handed Z-DNA structure,[54] with standard Watson–Crick rather than wobble-type hydrogen bonding between the affected bases (Fig. 2). This arrangement can only occur if

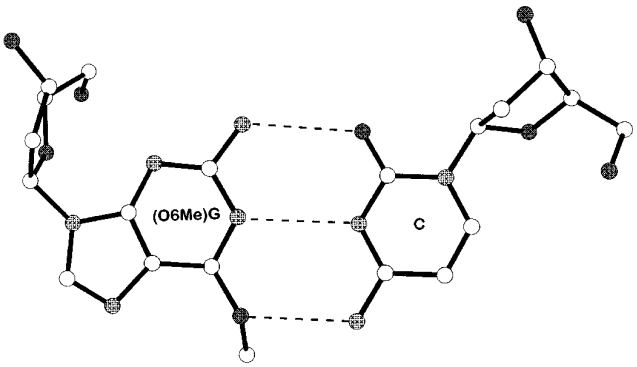

Fig. 2 The (O6Me)G·C base pair (ref. 54)

there is protonation of the guanine or the cytosine, or if one or other is in a non-standard tautomeric state. It is likely that this is a consequence of the particular base-stacking requirements in Z-DNA, which would not favour a wobble arrangement that would have the methylated guanine partially unstacked and protruding into the major groove.

The O6 methylation event results in GC→AT transition mutations, and thus the formation of a O6G···T base pair during replication. The structures of two dodecanucleotide duplexes containing this mismatch have been reported.[55, 56] The base pairing in both is analogous to the normal Watson–Crick arrangement (Fig. 3), whereas NMR solution studies[57] have suggested that the N1G···N3T hydrogen bond is absent or very weak. The origin of this difference is not clear, other than that it may reflect differences between crystal and solution environments.

4 Base triplets and triple-helical nucleic acids

It has long been known that a third nucleic acid strand comprising pyrimidines can associate with a stretch of duplex consisting of purines on one strand and pyrimidines on the other, to form a triple helix.[57] This parallel triple helix involves Hoogsteen hydrogen-bonded base triplets of the form T–A·T and C^+–G·C, and can be either inter- or intra-molecular. In

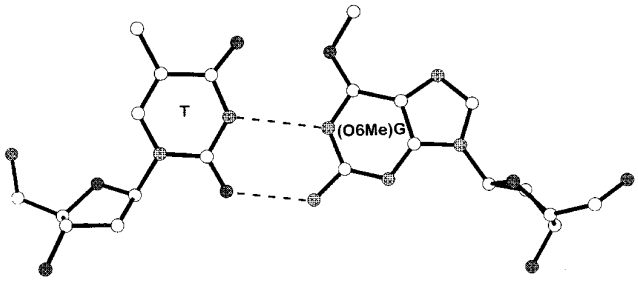

Fig. 3 The (O6Me)G·T base-pair in the crystal structure of the d[CGTGAATTC(O6Me)GCG]$_2$ duplex (ref. 55)

both instances, the third strand occupies the major groove of the initial duplex. The stringent sequence requirements of triple helix formation are currently being explored for the artificial down-regulation of the expression of particular genes, ultimately for therapeutic purposes.[58] The structural details of DNA triple helices remain elusive; fibre diffraction studies have only been at low resolution and have been interpreted in terms of both A- and B-type conformations. No single-crystal structure of a pure DNA triple helix is as yet available, in spite of much effort in a number of laboratories. The structure of a mixed DNA–peptide nucleic acid (PNA) has been reported.[59] It has a standard DNA purine strand together with PNA sequences forming the two pyrimidine strands having a hexapeptide linking them in a hairpin-like manner (Fig. 4). The

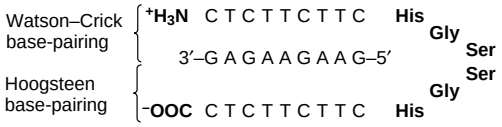

Fig. 4 Schematic view of the PNA:DNA triplex (ref. 59)

resulting triple helix (Fig. 5) has helical features of both A- and B-DNA, with bases approximately perpendicular to the helix axis (*i.e.* B form) yet significantly displaced from it (*i.e.* A form). It is not clear to what extent features such as the very wide major groove, reflect the non-nucleic acid nature of the PNA backbone. A short two-triplet stretch of C^+–G·C triplets have been observed in a drug–decanucleotide crystal structure where the terminal nucleosides do not form part of the duplex but interact with neighbouring duplexes in the lattice, to form this triplet.[60]

Anti-parallel triple helices can be formed with third-strand purines and involve, for example, G–G·C base triplets. Again, there is no crystal structure of such a triple helix. The triplet base arrangements have been observed in several crystal structures of duplexes with over-hanging ends [d(GCGAAT TCG)[61, 62] and d(GGCAATTGG)[63]], or with mismatches [d(GGCAATTGCG)].[64] In all instances the triplets are formed by intermolecular hydrogen bonding of bases.

5 DNA quadruplexes

The richness of possibilities for DNA structures is seen at its most striking for sequences containing either runs of either G,T or C,A nucleotides. These sequences are of considerable biological importance in view of their occurrence as repeated sequences, predominantly at the ends of chromosomes, forming so-called telomeres. Telomeric sequences can form a variety of four-stranded structures, all of which necessarily contain non-standard base-pairings.

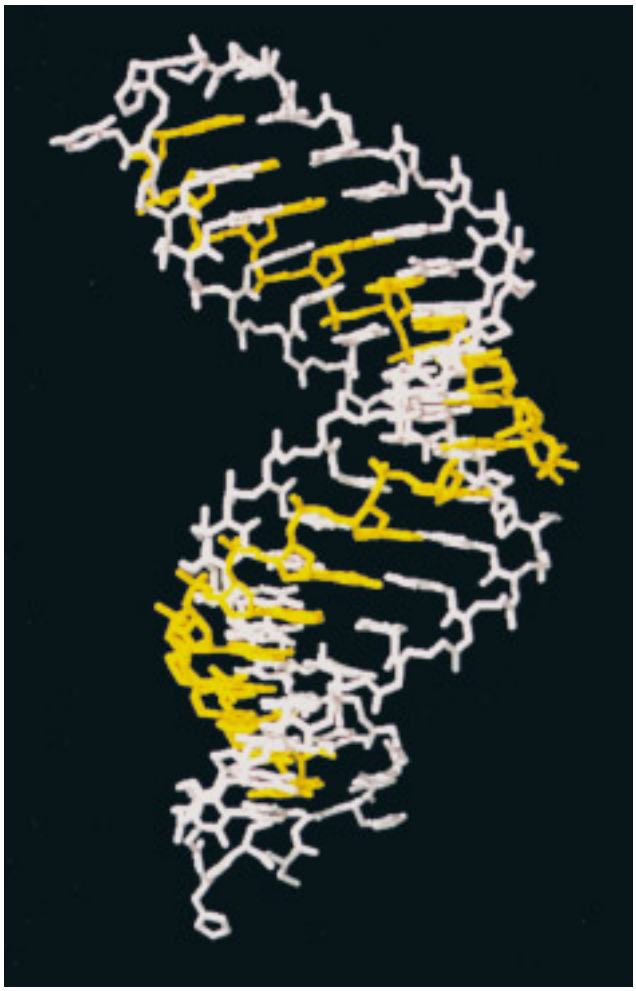

Fig. 5 The crystal structure of the PNA:DNA triplex (ref. 59)

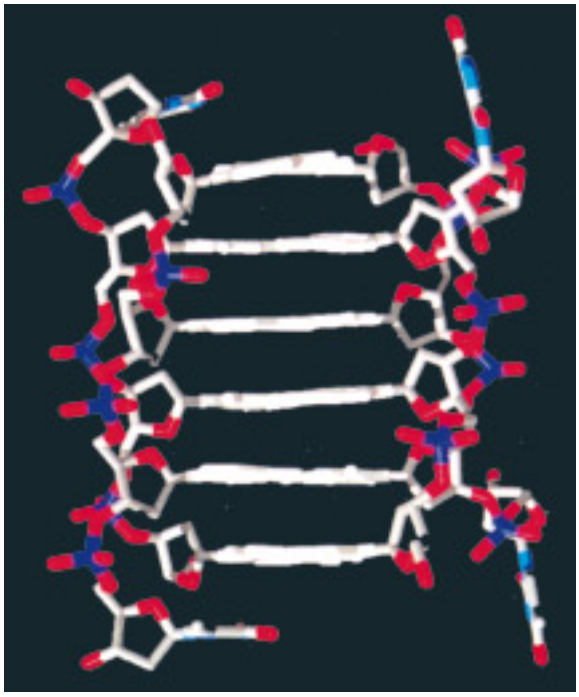

Fig. 6 The four-stranded intercalated structure formed by d(CCCT) (ref. 66)

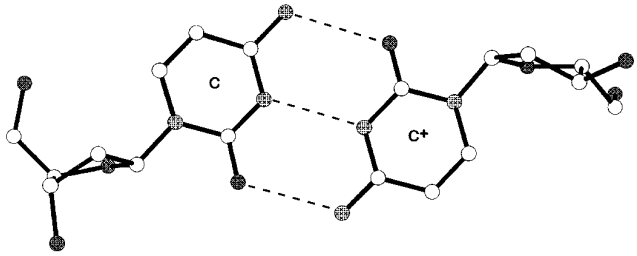

Fig. 7 A C^+·C base pair, of the type found within the C-rich quadruplex structures

The cytosine-rich sequences form four-stranded intercalated complexes.[65] That formed by the sequence d(CCCT) is typical, with two parallel-stranded duplexes intercalated into each other (Fig. 6), and the arrangement as a whole being formed by four d(CCCT) strands.[66] Each duplex has cytosine···cytosine base pairs (Fig. 7), with one of them required to be protonated (even though crystallisations were successful at pH 6 or even 7). The thymines do not actively participate in the duplexes, although some of them are stacked onto the ends. The cytosine rings are not directly stacked on top of one another. Instead, amino and carbonyl substituent groups form stacks, separated by 3.1 Å rather than the normal 3.4 Å in oligonucleotide helices. The same arrangement of intercalated cytosine duplexes has been observed in the structure[67] of the sequence d(TAACCC), corresponding to the human telomere repeat. The structure has four strands associating together so as to produce an intercalated four-stranded structure, still with parallel cytosine duplexes, but now with the 5′ terminal d(TAA) sequences forming intermolecular loops held together by A···T base pairings (Fig. 8). One of these has a Hoogsteen arrangement, whilst the other shows reverse Watson–Crick pairing. The crystal structure of the sequence d(CCCAAT) has the intercalation motif extended by adenine···adenine base pairs as part of the arrangement of parallel duplexes. The 3′ terminal thymines participate in a variety of intermolecular A·A·T base triplets, which serve to stabilise the crystal structure.

Although these C-rich sequences all appear to occur in telomeres within duplexes (in contrast to the G-rich repeats—see below), the fact that stable four-stranded structures can be formed by them with individual short sequences, suggests that if biological C-rich sequences became looped-out of the duplex they might then be stabilised by an intramolecular four-stranded arrangement. Such loops could occur as a result of negative supercoiling.

The G-rich single strand of telomeres has repetitive sequences of the type $d(TTTTGGGG)_n$ (in *Oxytricha*), $d(TGGG)_n$ (in budding yeast) and $d(TTAGGG)_n$ (in *Homo sapiens*). Four-stranded structures can be formed in a variety of ways, by intramolecular folding of several consecutive repeats, by intermolecular association of four strands or by a combination of both.[68] The crystal structure[69] of the sequence d(GGGGTTTTGGGG) shows a stack of sets of the G-quartet motif (Fig. 9), four guanine bases hydrogen-bonded together in one plane (Fig. 10). The four thymines form loops between the stacks, as shown. A subsequent NMR spectroscopy study[70] has reported a different arrangement for the four strands, whilst retaining the same arrangement of stacked G-quartets.

The structure of the sequence d(TGGGGT) has been determined at 0.95 Å resolution,[71] the highest resolution to date of any nucleic acid structure. It consists of four individual strands associated by means of a stack of four G-quartets. Sodium ions stabilise the structure, between and within the G-quartets. An unconventional four-stranded structure is formed[72] by the sequence d(GCATGCT), with two strands associating together

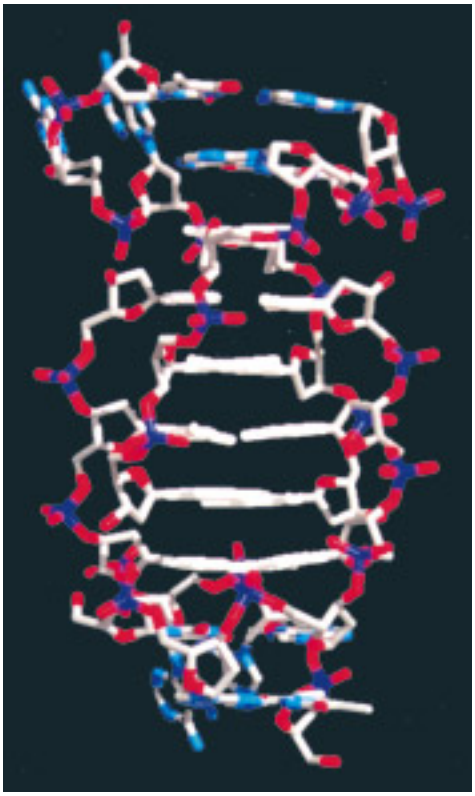

Fig. 8 The crystal structure of the four-stranded intercalated structure formed by d(TAACCC) (ref. 67)

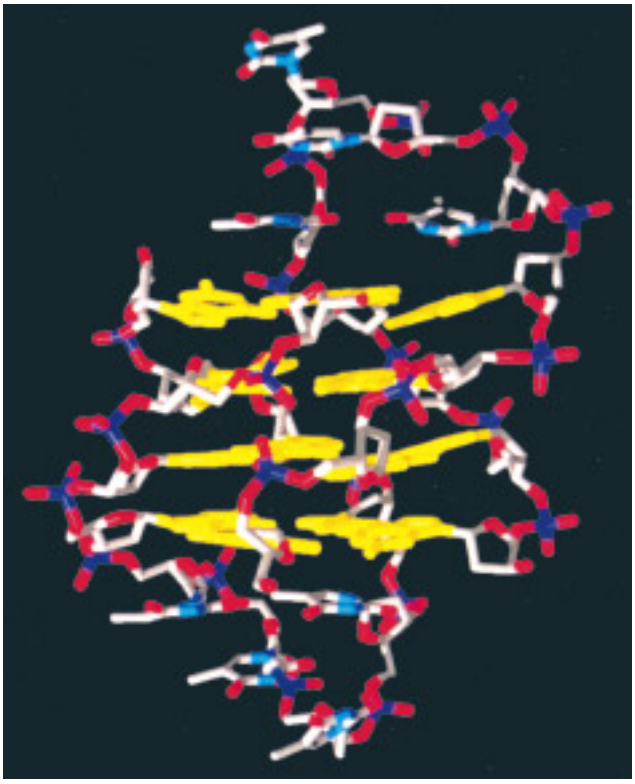

Fig. 9 The crystal structure of d(GGGGTTTTGGGG) (ref. 69)

(Fig. 11), each folding back so that two quartets of $(CG)_2$ are formed. This structure suggests that quartet-type structures may be more prevalent than hitherto supposed, since there is no longer a restriction on quartets to solely contain guanines.

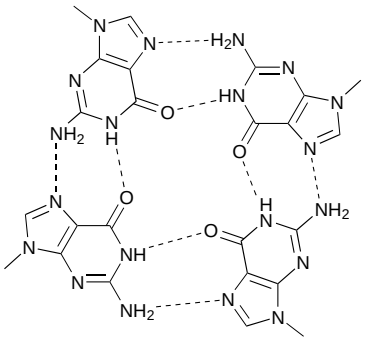

Fig. 10 The hydrogen-bonding arrangement in the guanine quadruplex

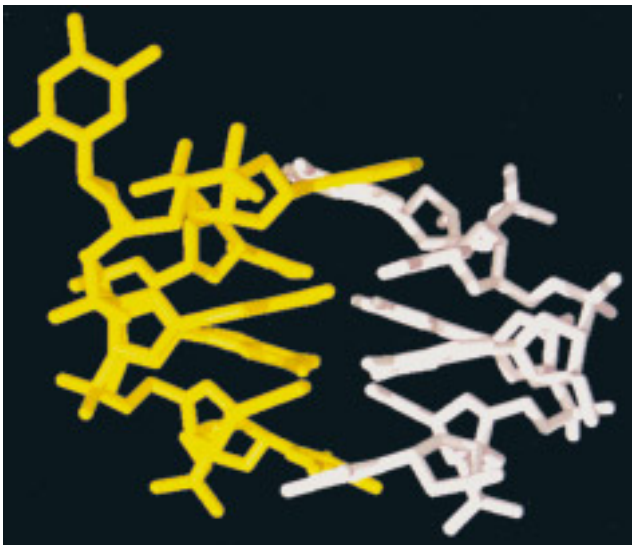

Fig. 11 The crystal structure of d(GCATGCT) (ref. 72)

6 RNA and DNA–RNA hybrid structures

Early X-ray crystallographic investigations into RNA structure in the 1970s were concerned, on the one hand, with transfer RNA[73] and on the other with dinucleoside monophosphate mini-helices both with and without bound drug molecules.[74] These short helices form right-handed antiparallel duplexes with an A-form conformation.

Investigations into RNA structure over the last few years has accelerated to produce a plethora of new and exciting structures. This results in a large part from recent developments in RNA synthesis, both chemical and enzymatic, which allows for the production of large quantities of high-quality RNA for crystallisation studies. Specific RNA sequences can now be synthesised, crystallised and their three-dimensional structures determined.[75] This section will survey recent X-ray crystallographic structural studies on them. In addition to pure RNA structures RNA–DNA hybrids will also be discussed.

6.1 RNA duplex structures

The octamer duplex r(CCCCGGGG) has been crystallised in both rhombohedral and hexagonal lattices.[76] There is little difference between the two different crystal forms. When compared with the structure of the DNA sequence d(CCCCG GGG)[77] the RNA is seen to be more extensively hydrated than the analogous DNA sequence with the ribose 2′-hydroxy groups propagating stable and conserved water networks in both grooves of the RNA duplex.

Crystallographic studies of ribosomal RNA are still in their infancy. The *E. coli* Shine–Dalgarno consensus ribosome binding site r(UAAGGAGGUGAU):r(AUCACCUCCUUA) has

been found to form a duplex structure with Watson–Crick base-pairing interactions along its length.[78] Two unique duplexes for this sequence show very similar conformations and both resemble calf thymus A-DNA as found from X-ray fibre diffraction studies. X-Ray crystallographic studies for the entire 5S rRNA from the thermophilic bacterium *Thermus flavus*[79] have only produced crystals to date which exhibit very poor resolution quality (8 Å). By subdividing the entire 5S ribosomal RNA into subunits it has been possible to obtain good-quality diffracting crystals. 5S Ribosomal RNA has been subdivided into five subdomains A–E. Subdomain A comprises two RNA strands which form a double helix involving Watson–Crick base-pairing interactions in addition to being stabilised by two U·G and G·U base-pairs.[80] Domain E of *Thermus flavus* 5S rRNA contains the very stable and highly conserved 5′-GCGA tetraloop. Crystallisation and preliminary diffraction studies have been reported[81] on it and structure determination is underway.

The structure of the self-complementary tetradecamer sequence r[U(UA)$_6$A] is an A-type duplex with two kinks along its length.[82] The kink angles are 8.5 and 13.8° with an overall angle between the two distal ends of the duplex of 21.7°.

The structure of r(UUCGCG)[83] has a central (CGCG)$_2$ duplex formed by two RNA strands with an overhang of the 5′-UU bases from each strand. These overhanging uracil bases interact with the overhang from a symmetry-related duplex within the unit cell to form novel Hoogsteen-like U·U base-pairs (Fig. 12). There is one hydrogen bond between atoms O4

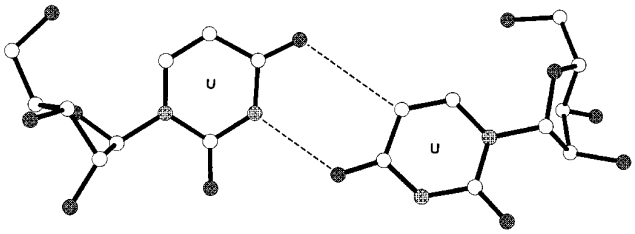

Fig. 12 U·U base-pair in the crystal structure of r(UUCGCG) (ref. 83)

and N3 of two uracil bases, while there is a second, less conventional hydrogen bond between C5–H and O4. This base-pairing arrangement results in a *trans* U·U pair on antiparallel strands in contrast to the usual *cis* base-pairs. It differs from that found in RNA dodecamer duplexes with non-Watson–Crick base-pairs. The potential importance of C–H···O hydrogen bonding interactions in nucleic acid structure has been discussed.[84]

6.2 Mismatches in RNA structures

In addition to the U·U base-pairing interactions which exist within the structure of r(UUCGCG),[83] base-pairing mismatches have been observed in a number of other RNA structures [Fig. 13(*a*)–(*e*)]. Mismatched regions in RNA structure are a common secondary structural motif and termed internal loops. These loop regions act as potential protein recognition sites. The structure of r(GGCGCUUGCGUC)[85] contains tandem U·U base-pairs within a dodecamer duplex in addition to two G·U mismatches. As a result, the duplex has an overall bend of 11–12°. The structure of r(GGACUUUGG UCC)[86] similarly contains tandem U·U base pairs. The *cis* U·U wobble pairs observed in both of these structures have two hydrogen bonds, with potential sites for the binding of water molecules in the major and minor grooves. Both also contain, in addition to U·U base pairs, two G·U pairs along the length of the dodecamer duplexes. In the case of r(GGACUUUGGUCC) there is a run of four mismatched base pairs.

The structure of r(GGACUUCGGUCC)[87] has G·U and C·U mismatches in the central third of the structure. The structure has a central two-fold symmetry axis and two unique mismatches. The G·U mismatch is stabilised by a solvent molecule in the minor groove which also interacts with the ribose hydroxy group. The structure of r(CGCGAAUUA GCG)[88] has two separated G·A mismatches within the dodecamer duplex structure. The structure of r(GGCCGAA AGGCC)[89] has an internal loop with G·A and A·A mismatches. Again the dodecamer duplex has a two-fold symmetry axis, with two unique mismatches. The G·A mismatches in this structure involve reverse Hoogsteen hydrogen bonding and have been termed sheared G·A base-pairs. Both G·A and A·A base-pairs are very common in internal loops of RNAs, including ribosomal RNA and ribozymes. This structure has a 34° end-to-end curvature for the helix and its diameter is narrowed by 24% in the internal loop.

The large number of structures of non-self-complementary RNA sequences forming duplex structures with mismatched base-pairs has resulted from attempts to crystallise RNA sequences which have the potential to form stem-loop structures. Studies using NMR spectroscopy for such sequences have shown them to form hairpin structures under the conditions used. However under the conditions required for the crystallisation of these sequences there is a preference for duplex RNA to be formed.

6.3 RNA as an enzyme

Certain RNA sequences, termed ribozymes, have the ability to either cleave other RNA molecules, or to self-cleave, by means of a catalytic mechanism. The best studied ones are the hammerhead ribozymes, which consist of three double helical regions joined by 15 highly conserved nucleotides. These 15 central nucleotides are essential for ribozyme activity and form a complex structure which mediates RNA folding and catalysis. Hammerhead ribozymes comprise two strands, one corresponding to the 'enzyme' and the other to the 'substrate'. A divalent metal ion such as Mg^{2+} is required by the ribozyme to mediate catalytic cleavage of an RNA species.

Two crystallographic studies for hammerhead ribozymes (Fig. 14) have been carried out. The first structure reported has an RNA 'enzyme' strand and a DNA 'substrate' strand.[90] The DNA strand was employed in order to prevent catalytic cleavage. The second hammerhead ribozyme structure (Fig. 15) is composed entirely of RNA with a single 2′-methoxy group modification at the active site to prevent cleavage.[91] The two structures have very similar structures for the catalytic core region and results from these two structural studies have led to proposals for the mechanism of RNA-cleavage catalysis.[92, 93]

In addition to the two hammerhead ribozyme structures the crystal structure of a self-splicing Group I ribozyme domain has been recently determined.[94] This 160 nucleotide P4–P6 domain from the *Tetrahymena* group I intron has been reported to a resolution of 2.8 Å. The structure consists of two extended helical regions that pack side by side with long-range interactions between them. These interactions involve an A-rich bulge of one helix and the minor groove of another, together with a GAAA tetraloop at the end of one helix and a tetraloop receptor in the minor groove of the other. This structure has close packing of the ribose–phosphate backbones, which is mediated by hydrated magnesium ions.

6.4 RNA–DNA hybrid structures

Short stretches of RNA–DNA hybrid structures are formed during both replication and transcription, and are important in antisense therapeutic applications where a deoxyoligonucleotide sequence is targeted to a mRNA. One RNA–DNA hybrid crystal structure determined to date is left-handed and of Z-type,[95] while the remainder are all right-handed. This hybrid consists of alternating purine and pyrimidine bases

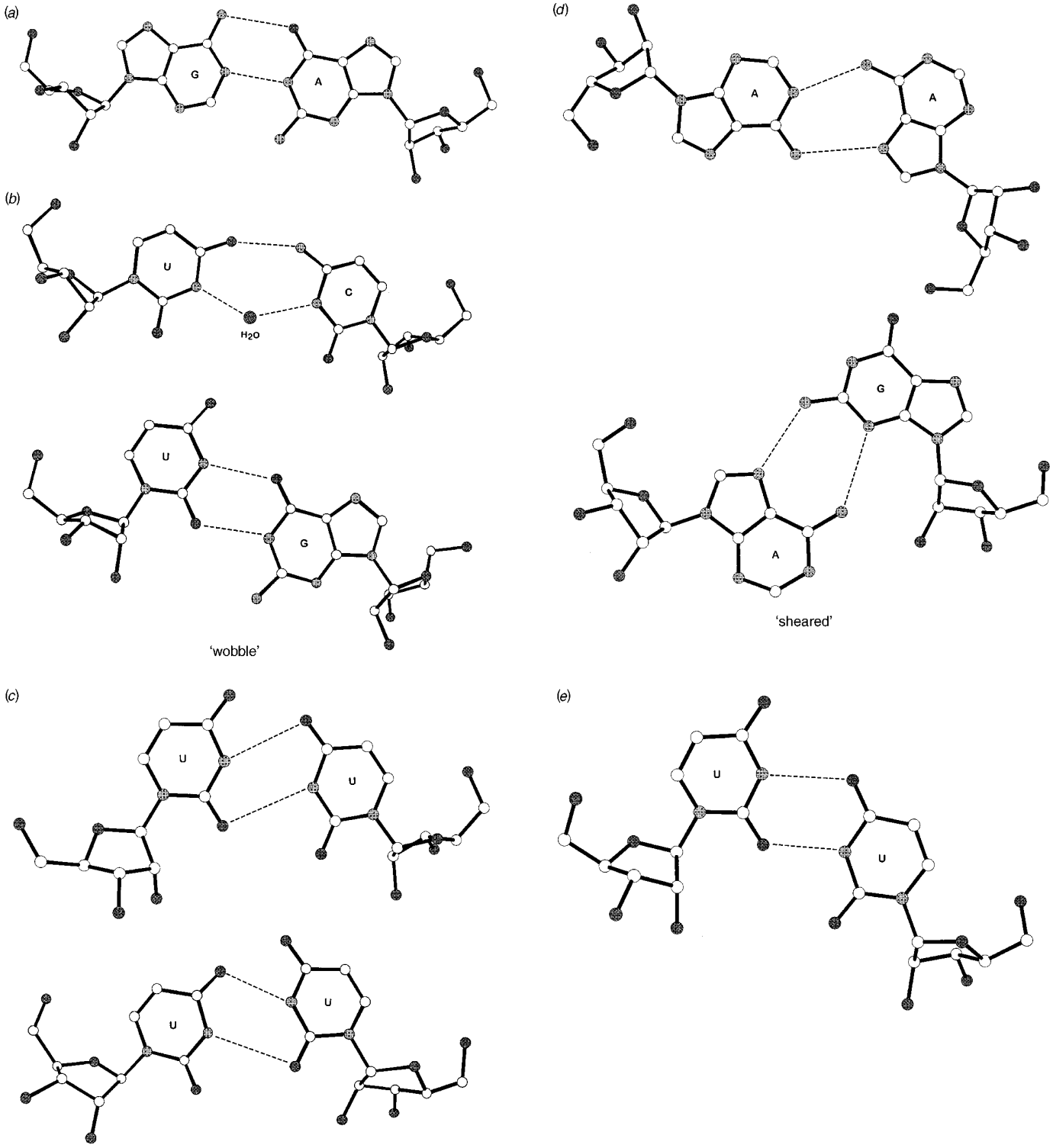

Fig. 13 Mismatched base-pairs in dodecamer duplex RNA structures: (*a*) r(CGCGAAUUAGCG), (*b*) r(GGACUUCGGUCC), (*c*) r(GGACUUUGGUCC), (*d*) r(GGCCGAAAGGCC) and (*e*) r(GGCGCUUGCGUC)

d(CG)r(CG)d(CG) and assumes the conformation seen for all DNA sequences of this type. The octameric sequence r(GUAUAUA)d(C)[96] has a 3′-terminal deoxycytidine residue at the end of a heptamer RNA sequence. This structure forms a right-handed A-form double helix. The structure of the two octamer sequences d(I)r(C)d(ICICIC) and d(I)r(C)d(I)r(C)d(ICIC) have both been determined bound to the minor groove binding drug distamycin A.[97] These structures adopt a B-form structure with two distamycin A molecules lying within the minor groove of the B-form duplex (see below for a further discussion of these two structures). These are the first examples of RNA duplexes adopting the B-form family of helices, which have been considered to be less stable for RNA compared to DNA. The presence of drug binding may induce the transition from A to B form in this structure which would suggest that upon protein binding both duplex RNA and RNA–DNA chimeric structures may be able to access a B form conformation if required.

The nonamer sequence r(GCUUCGGC)d^{Br}U forms an octamer duplex structure[98] with two different types of base

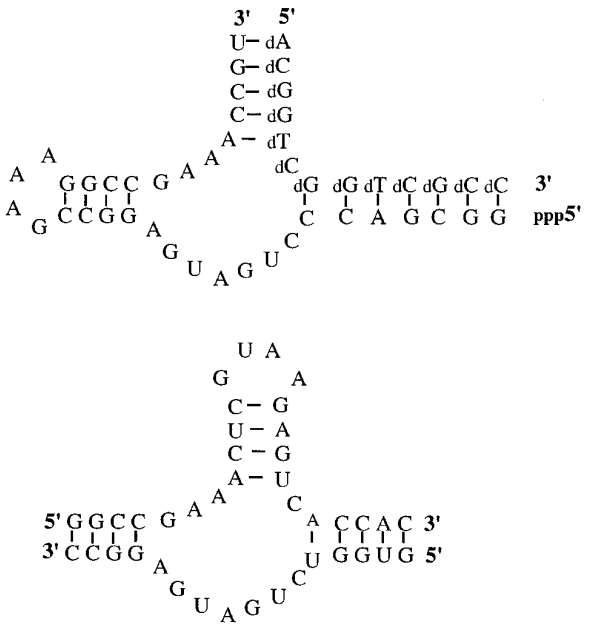

Fig. 14 Schematic representation of the two hammerhead ribozyme structures (refs. 90 and 91)

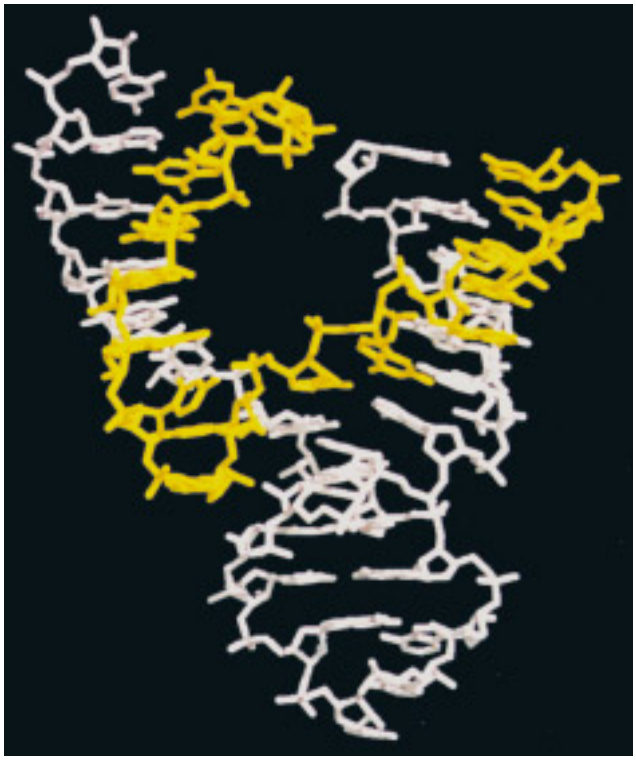

Fig. 15 The crystal structure of the all-RNA hammerhead ribozyme (ref. 91). The two RNA strands are shown in different colours

mismatches. It has a G·U base pair with wobble hydrogen bonding, while the C·U mismatch involves just one hydrogen bonded contact, in addition to a bridging water molecule. The terminal ^{Br}U-3′ bases pair with another ^{Br}U *via* interactions involving two hydrogen bonds.

A number of decamer chimeric (*i.e.* RNA and DNA residues in the same strand) structures have been reported. The structure of r(GCG)d(TATACGC) is of an A-type decamer duplex.[99] The three decamer structures d(GGGTATACGC)/r(GCG)d(TATACCC) (which is an Okazaki fragment), r(G)d(CGTATACGC) and d(GCGT)r(A)d(TACGC)[100] all assume A-type duplex conformations. For d(GGGTATACGC)/r(GCG)d(TATACCC) there is no difference in backbone conformation between the r(GCG)·d(CGC) portion of the structure and the d(TATACCC)·d(GGGTATA) all-DNA helical region. The three decamer structures d(CCGGC)r(G)d(CCGG),[101] r(C)d(CGGCGCCG)r(G)[102] and r(GC)d(GTATACGC),[103] with one or two ribose bases along the decamer length, all adopt an A-type conformation. One structure of a DNA duplex with a 2′-*O*-methylribonucleotide insert[104] has been reported.

The decamer sequence r(UUCGGGCGCC)/d(GGCGCCCGAA)[105] is specifically recognised by the ribonuclease H function of HIV reverse transcriptase. The structure of this sequence has neither an A- or B-type conformation but instead has characteristics of both. The structure of r(GCG)d(ATATA)r(CGC) has been determined in two different crystal forms.[106] This structure (Fig. 16) contains a single adenosine bulge—a common secondary structural motif in RNA.

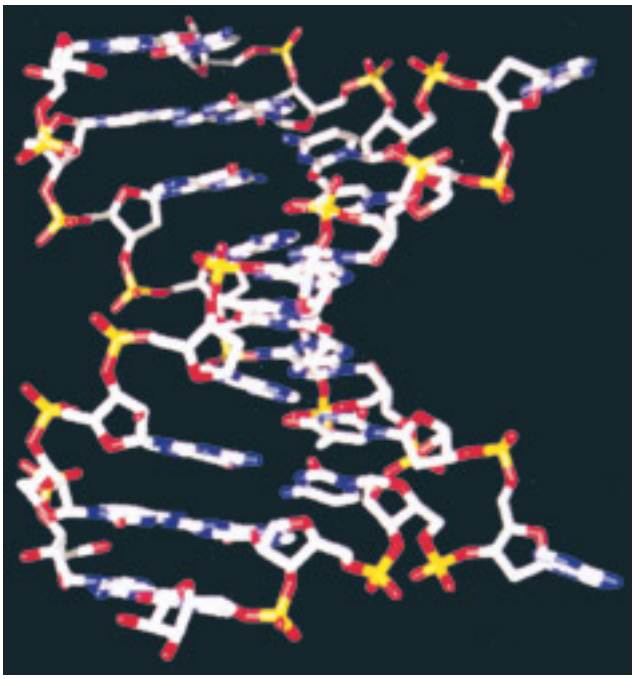

Fig. 16 The crystal structure of r(GCG)d(ATATA)r(CGC) (ref. 106)

7 Drug–nucleic acid structures

A wide range of drugs are known to exert their biological effects by means of interactions with cellular nucleic acids, especially with chromosomal DNA.[107–108] The majority are anticancer agents, but antibacterial, antiviral, antifungal and antiparasitic activities have also been found for some of them. The DNA sequence selectivities shown by most of these drugs is usually modest, and even those which covalently bind to particular bases have selectivity for only short runs of sequence. Thus the biological selectivities shown, for example by the clinically-useful anthracycline anticancer drugs, cannot be solely ascribed to DNA binding. There is increasing evidence that interference with particular protein–DNA interactions is critical for biological activity.[109–111] For example, the anthracyclines and related drugs have been shown to stabilise the cleavable complex between DNA and the enzyme DNA topoisomerase II. A number of covalent and non-covalent DNA minor-groove agents compete with transcription factors, and so interfere with gene regulation. There is as

yet no structural data on any drug–DNA–protein ternary complex, raising the question of the relevance of structural studies on drug–DNA complexes alone. It is reassuring that these have been at least partly successful in rationalising structure–activity behaviour, suggesting that the structures of the binary complexes have at least some relevance to those of ternary complexes.

Crystallographic studies have concentrated on the intercalative and minor-groove categories of complexes, with few exceptions. Very few covalent complexes have been reported, in spite of the biological and medicinal importance of drugs such as mitomycin and the nitrogen mustards. The problems associated with obtaining significant quantities of pure oligonucleotide adducts of these drugs have been largely overcome, as attested by the numerous NMR studies on them. The experience in this and other laboratories points to the problem being largely of difficulties in obtaining crystals suitable for studies even at medium (*ca.* 2.5 Å) resolution.

7.1 Intercalation complexes

The first drug–DNA crystal structures to be determined were of structurally-simple intercalating molecules, typified by the acridine proflavine,[74] bound to dinucleoside monophosphate mini-duplexes. These structures showed the planar drug chromophore sandwiched between the two base pairs of the dinucleoside duplex, but were unable to address issues of conformational change in a nucleic acid beyond the immediate intercalation site. This has to some extent been addressed by subsequent structural studies on a large number of complexes involving the clinically-important anthracycline anticancer drug daunomycin (Fig. 17) and many of its derivatives, all

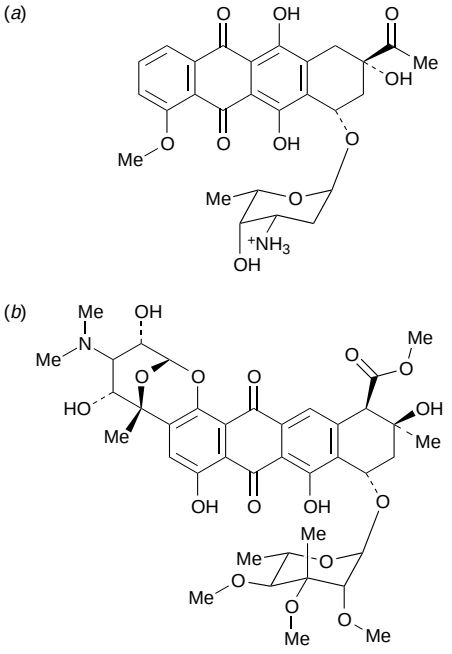

Fig. 17 (*a*) Daunomycin and (*b*) nogalamycin

complexed with hexanucleotide duplexes.[112] As of May 1997, there were 22 entries for anthracycline complexes in the Nucleic Acid Database.[4] Most crystallise in the tetragonal space group $P4_12_12$. In general, these structures have two drug molecules bound per hexamer duplex, one each at the terminal base-pair sites, and with the daunosamine sugars lying in the minor groove. The semi-synthetic derivative idarubicin co-crystallises with the sequence d(CGATCG) in the trigonal space group $P3_1$. The resulting structure is essentially identical to the tetragonal ones, with invariance in the orientation of bound drug chromophore and in the drug–DNA hydrogen-bond contacts being observed.[113] A novel *bis*-daunomycin (Fig. 18) has recently been designed and synthesised[114] using

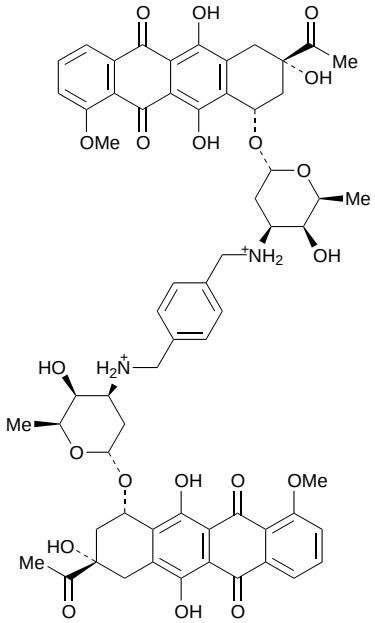

Fig. 18 The *bis*-daunomycin molecule (ref. 115)

the established anthracycline-hexamer crystal structures as a starting-point. The semi-rigid linker between the two anthracycline chromophores was chosen so as to preserve the position and orientation of each, as seen in these monomer structures. The new compound *bis*-intercalates into duplex DNA with high affinity and shows promising ability in cytotoxicity studies to circumvent multi-drug resistance in tumour cells. The crystal structure[115] of the *bis*-daunomycin compound complexed with the sequence d(CGTACG) (Fig. 19) shows that many of the predictions are realised, with the linker positioned in the minor groove of the hexamer duplex. It is surprising that the linker appears to have very few close van der Waals contacts with the groove surface, suggesting that modifications to the structure of the linker might result in further enhancements of DNA affinity.

The related antitumour antibiotic nogalamycin (Fig. 17) also preferentially intercalates at pyrimidine-3′,5′-purine sites, but *via* a threading mechanism. Crystallographic studies, again on hexanucleotide duplex complexes, have shown that this anthracycline binds with the nogalose and aminoglucose groups lying in the minor and major grooves, respectively (Fig. 20). A complex with the sequence d(TGTACA) has shown the drug bound in the two high-affinity TpG sites,[116] and highlights the contribution of solvent-mediated contacts to the observed sequence selectivity of this drug. The complex between nogalamycin and the sequence d(CCCGGG) is notable[117] in that it uniquely shows a 1:1 anthracycline–duplex complex, with the one drug molecule bound at the central CpG site (Fig. 20). The conformation of the unwound DNA has features of both A- and B-type helices. This structure is thus representative of anthracyclines bound in extended lengths of DNA sequence.

Intercalation at the centre of extended sequences has also been observed[118] for the antitumour antibiotic actinomycin (Fig. 21), showing this antitumour antibiotic bound at the central GpC site of the sequence d(GAAGCTTC). The two cyclic pentapeptide rings lie in the minor groove, and there is a predicted set of hydrogen bonds from threonine residues to N2 atoms of the two central guanines. The same DNA sequence

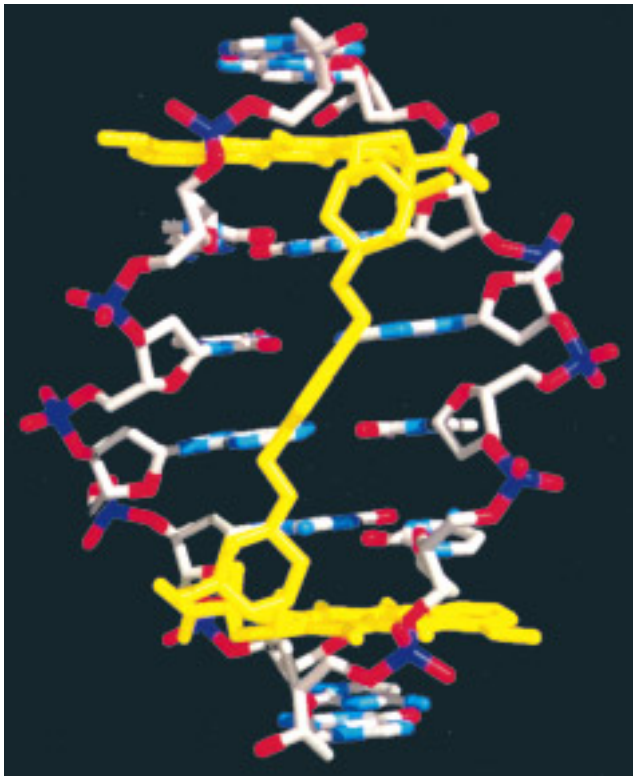

Fig. 19 Crystal structure of a *bis*-daunomycin–d(CGTACG) complex (ref. 115)

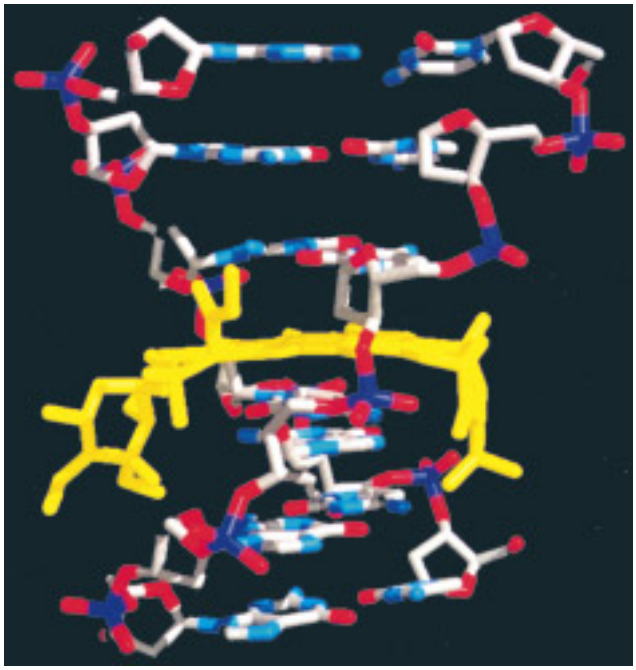

Fig. 20 Crystal structure of nogalamycin complexed with d(CCCGGG) (ref. 117)

has also been cocrystallised with N8-actinomycin, where the 8-position in the phenoxazone ring has been replaced by a nitrogen atom.[119]

The interactions of *meso*-substituted porphyrins with DNA have been much studied. The crystal structure of the tetrapyridyl porphyrin TMPy with the sequence d(CGTACG) shows[120] that intercalation of the central planar porphyrin ring system is accompanied by flipping-out of one nucleoside

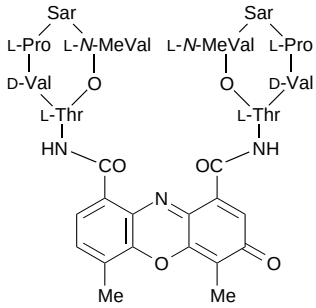

Fig. 21 Actinomycin

from the intercalation site. This is not unexpected, given the stringent steric requirements of the bulky porphyrin system with respect to intercalation.

7.2 Groove-bound complexes

A large group of drugs bind non-covalently in the minor groove of AT-rich regions of B form DNA duplexes. Structural studies have focused on the nature of this sequence selectivity, which is currently being extensively exploited in the design of analogues with selectivities for mixed-sequence DNA.[121–125] Particular use has been made of the self-complementary dodecamer duplex sequences d(CGCXAATTYGCG)$_2$, where X=A or G and Y=T or C, since the central regions are largely unaffected by intermolecular contacts in the crystal, other than with water molecules. The pattern of hydrogen bonding in these complexes has been systematically examined.[126]

The crystal structures of a number of complexes with drugs typified by pentamidine, berenil and their analogues, have been reported.[127–130] These have shown that hydrophobic interactions with hydrogen atoms from the nucleotide backbone (H1′, H4′ and H5′, which line the walls of the minor groove), are in large part responsible for the AT-selectivity of these drugs. The minor groove is narrow at AT sequences and its width is close to the cross-sectional diameter of these drugs. Hydrogen bonding to A·T base-pair edges is, by contrast, relatively weak and variable.

7.2.1 Complexes with Hoechst analogues

A second category[130] of non-covalent minor-groove drugs, typified by netropsin, distamycin and Hoechst 33258 (Fig. 22), tend to have fixed patterns of hydrogen bonding to base pair edges, to N3 of adenines and O2 of thymines. Thus the two benzimidazole groups in Hoechst 33258 and its analogues consistently form two bifurcated pairs of hydrogen bonds to base edges (Fig. 23), extending over three base pairs in total.[131–133] The analogue with three linked benzimidazole groups[134] follows the same pattern of hydrogen bonding, with now three pairs of hydrogen bonds, covering four A·T base pairs. This ligand has an overall binding site of *ca.* 7.5 bases, so that it covers three quarters of a complete turn of a B-DNA double helix. In order for the three benzimidazole groups to be in register (in phase) with the four successive A·T base pairs, one end of the ligand is not in close van der Waals contact with the floor of the minor groove, so that the ligand is overall not isohelical with the contour of the groove.

The isohelicity principle[135] is also seen to be violated in the crystal structure of an amidinium analogue of Hoechst 33258 bound to d(CGCGAATTCGCG)$_2$, where there is effective binding to DNA, yet with the drug not having a concave inner curvature to match minor groove isohelicity.[136] Here, the dominance of hydrogen bonds from one amidinium group is sufficient to overcome the usual requirement for effective van der Waals contacts along the length of the ligand. Altering AT selectivity for GC by simple replacement of a hydrogen bond

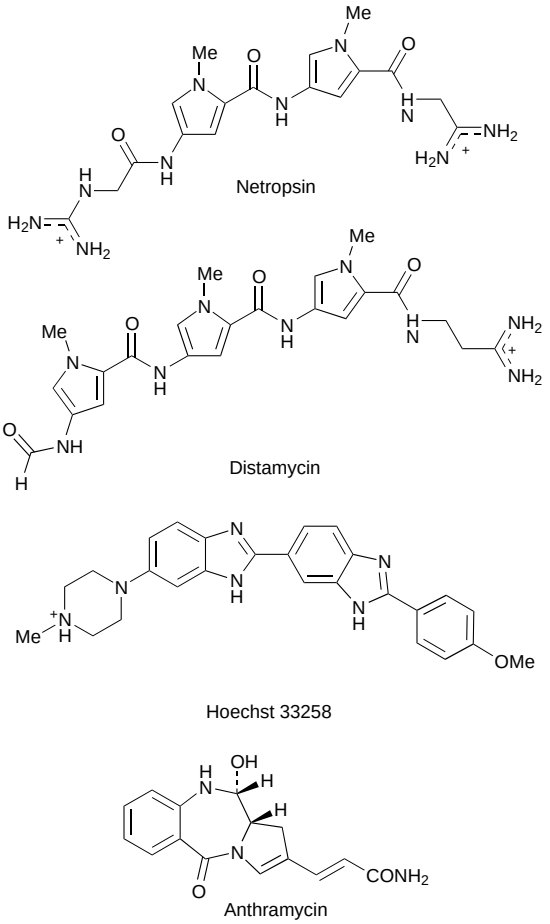

Fig. 22 Several groove-binding molecules

donor in the minor groove by an acceptor (for the N2 group of guanine), has been shown to be ineffective in the case of the *meta*-hydroxy analogue of Hoechst 33258, contrary to theoretical predictions.[137] The hydroxy group is positioned insufficiently deeply into the groove for such an interaction to occur, and the analogue is thus not able to actively recognise G·C base pairs.

7.2.2 Complexes with netropsin and analogues

The antibiotic netropsin (from *Streptomyces netropsis*) is in many ways the paradigm for minor groove binding drugs, although several crystal structures have shown it bound in subtly different ways within AT sequences. A careful study[138] has been made of optimal refinement and electron-density map interpretation for the complex with $d(CGCGAATTCGCG)_2$ has confirmed the original assignments of bifurcated hydrogen bonds to the AT base pairs, analogous to those formed by the benzimidazole drugs. That this result is not an artefact of dodecamers was shown by the crystal structure[60] of netropsin bound to the decamer sequence $d(CGCAATTGCG)_2$, where the same pattern of drug–DNA hydrogen bonding was observed as with the dodecamer complex. The lexitropsins have been developed as netropsin analogues, where methylpyrrole ring(s) have been replaced by methylimidazole rings in order to attempt to switch recognition from AT to GC sequences. As with the Hoechst analogue above, the lexitropsins have not fulfilled these predictions, and a crystal structure of a lexitropsin–dodecanucleotide complex[139] shows the methylimidazole ring to be lying in the AT region, as in netropsin itself.

An alternative way of recognising GC sequences has been developed, based on the observations from oligonucleotide

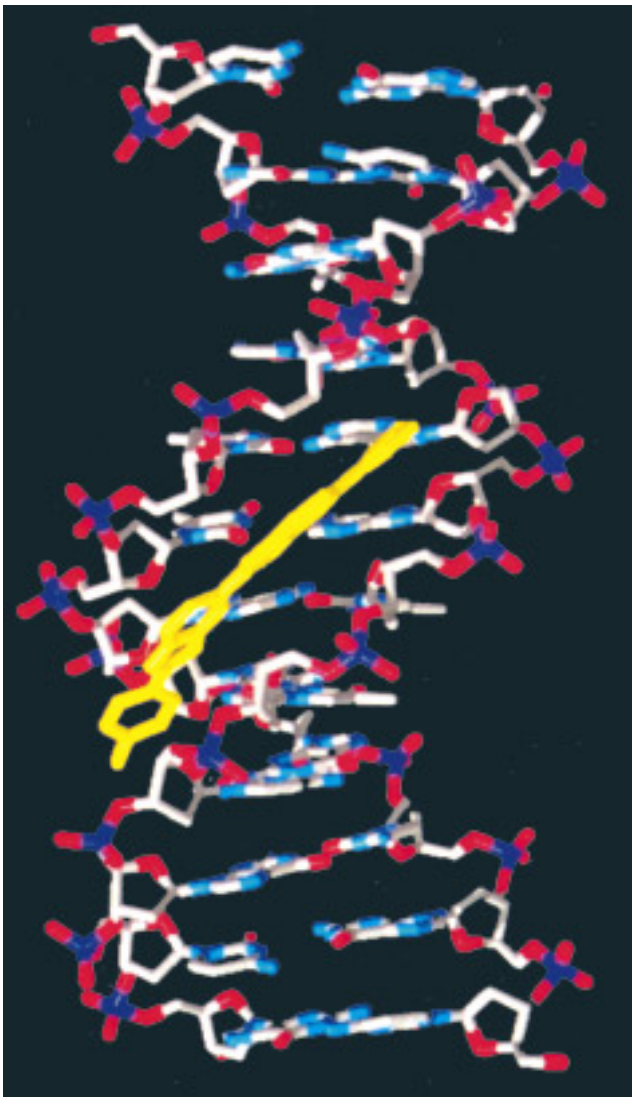

Fig. 23 Crystal structure of the complex between the DNA duplex d(CGCGAATTGCGC) and a *bis*-benzimidazole compound (ref. 133)

crystal structures, that such sequences tend to have a wide minor groove. The ability of the monocationic antibiotic distamycin to form side-by-side dimers in such regions, has been the basis of fruitful studies on the design of ligands which can select mixed DNA sequences with high affinity and specificity.[123, 125] Several crystal structures have been reported for distamycin complexes with sequences such as d(ICICICIC)[140] and d(ICITACIT),[141] as well as variants containing DNA/RNA chimers.[97] In all cases, two distamycin molecules are bound per duplex (Fig. 24), which assumes a B-form structure, with a minor groove widened to 7.8 Å to accommodate the two drug molecules. Each distamycin hydrogen bonds to the nearest strand, with backbone amide nitrogen atoms acting as donors to thymine O2 and adenine N3 atoms, in a manner analogous to the hydrogen seen in distamycin 1:2 complexes.

7.3 Covalent complexes

The crystal structures of only two covalently-linked drug–DNA complexes involving significant lengths of oligonucleotide, have been reported to date. That of the clinically-important antitumour drug *cis*-platinum has the platinum linked to N7 atoms of adjacent guanine bases in the sequence d(CCTCTG*G*TCTCC):d(GGAGACCAGAGG), where G* represents platinated guanine,[142, 143] *i.e.* with *cis*-platinum bound only on one strand. There are two duplexes in the asymmetric unit, and both show significant bending at

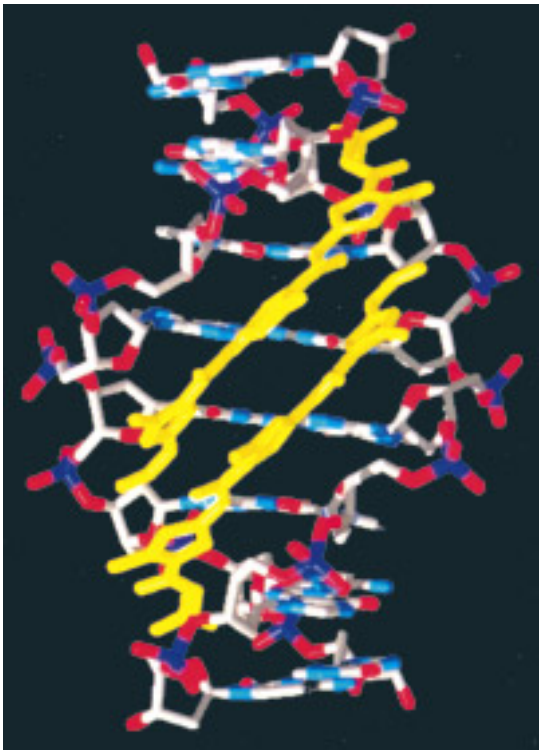

Fig. 24 Crystal structure of the 2:1 complex between the minor-groove binding drug distamycin and the DNA duplex d(ICICICIC) (ref. 140)

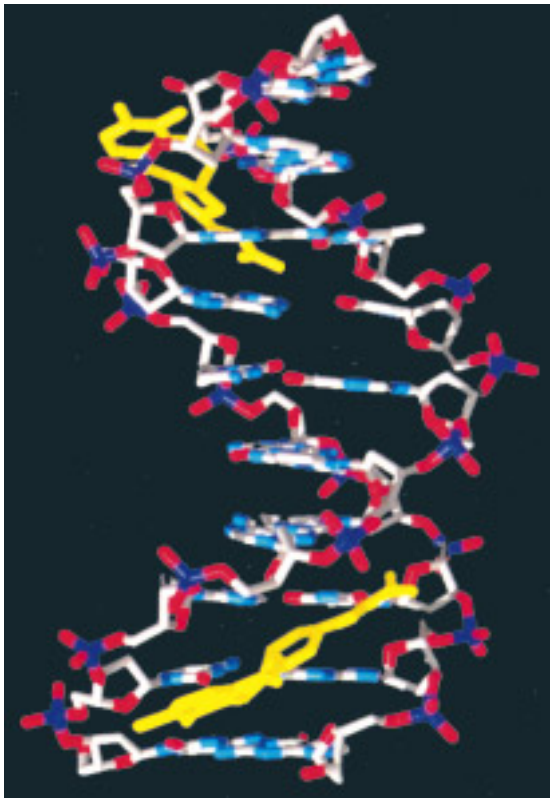

Fig. 25 The crystal structure of the anthramycin complex with the decamer duplex d(CCAACGTTG*G) (ref. 144)

the platination site, of 26° towards the major groove side. The minor groove is opened up, in a manner reminiscent of the DNA bending observed in the structures of TFIIIA–DNA complexes,[27–29] with backbone geometry having features of both A- and B-form DNA. It is suggested that the platinum-induced bending of DNA could provide a signal for protein binding, especially for those involved in DNA damage repair.

The antitumour antibiotic anthramycin (from *Streptomyces refuineus*) binds covalently *via* the benzodiazepine ring, to the N2 atom of guanine, with a preference for guanine flanked on both sides by a purine. The complex with the decamer d(CCAACGTTG*G) has two anthramycin molecules bound per duplex[144] (Fig. 25), with each lying in a narrowed region of the minor groove. The sequence specificity is considered to arise from lower twist angles at purine–purine steps, rather than from hydrogen bonding to particular bases.

8 Acknowledgements

We are grateful to the Cancer Research Campaign for continuing support of structural studies on nucleic acid structures (grant SP1384 to SN). John Trent is thanked for useful discussions, and for his help in surveying the literature.

9 References

1 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737.
2 J. D. Watson, *The Double Helix*, Weidenfeld and Nicholson, London, 1968.
3 See, for example, reviews in *Current Opinion in Structural Biology*.
4 H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan and B. Schneider, *Biophys. J.*, 1992, **63**, 751. The Nucleic Acid Database is available on the World Wide Web at http://www.NDB.rutgers.edu, with a mirror site in Europe at http://www.NDB.ebi.uk.ac.
5 W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, Berlin, 1984.
6 S. Neidle, *DNA Structure and Recognition*, Oxford University Press, 1994.
7 C. R. Calladine and H. R. Drew, *Understanding DNA*, 2nd Edn., Academic Press, San Diego, 1997.
8 *Oxford Handbook of Nucleic Acid Structure* (ed. S. Neidle), Oxford University Press, 1998, in the press.
9 L. Clowney, S. C. Jain, A. R. Srinivasan, J. Westbrook, W. K. Olson and H. M. Berman, *J. Am. Chem. Soc.*, 1996, **118**, 509; A. Gelbin, B. Schneider, S. C. Jain, S.-H. Hsieh, W. K. Olson and H. M. Berman, *J. Am. Chem. Soc.*, 1996, **118**, 519; G. Parkinson, J. Vojtechovsky, L. Clowney, A. T. Brünger and H. M. Berman, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1996, **52**, 57.
10 A. T. Brünger, J. Kuriyan and M. Karplus, *Science*, 1987, **235**, 458.
11 R. E. Dickerson and H. R. Drew, *J. Mol. Biol.*, 1981, **149**, 761.
12 M. Hahn and U. Heinemann, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1993, **49**, 468.
13 M. G. Kubinec and D. E. Wemmer, *J. Am. Chem. Soc.*, 1992, **114**, 8739.
14 M. A. Young, B. Jayaram and D. L. Beveridge, *J. Am. Chem. Soc.*, 1997, **119**, 59.
15 H. M. Berman, *Curr. Opin. Struct. Biol.*, 1994, **4**, 345.
16 R. E. Dickerson, *Methods Enzymol.*, 1992, **211**, 67.
17 R. E. Dickerson, D. E. Goodsell and S. Neidle, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 3579.
18 K. Grzeskowiak, *Chem. Biol.*, 1996, **3**, 785.
19 B. Schneider, S. Neidle and H. M. Berman, *Biopolymers*, 1997, **42**, 113.
20 A. D. DiGabriele and T. A. Steitz, *J. Mol. Biol.*, 1993, **231**, 1024.
21 J. G. Nadeau and D. M. Crothers, *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 2622.
22 D. Sprous, W. Zacharias, Z. A. Wood and S. C. Harvey, *Nucleic Acids Res.*, 1995, **23**, 1816.
23 R. E. Dickerson, D. Goodsell and M. L. Kopka, *J. Mol. Biol.*, 1996, **256**, 108.
24 D. Goodsell, M. L. Kopka, D. Cascio and R. E. Dickerson, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 2930.
25 K. Grzeskowiak, D. S. Goodsell, M. Kaczor-Grzeskowiak, D. Cascio and R. E. Dickerson, *Biochemistry*, 1993, **32**, 8923.
26 D. Goodsell, M. Kaczor-Grzeskowiak and R. E. Dickerson, *J. Mol. Biol.*, 1994, **239**, 79.
27 M. Shatzky-Schwartz, N. D. Arbuckle, M. Eisenstein, D. Rabinovich, A. Bareket-Samish, T. E. Haran, B. F. Luisi and Z. Shakked, *J. Mol. Biol.*, 1997, **267**, 595.

28 J. L. Kim and S. K. Burley, *Nat. Struct. Biol.*, 1994, **1**, 638, and references therein.
29 A. Lebrun, Z. Shakked and R. Lavery, *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 2993.
30 Z. Otwinowski, R. W. Schevitz, R.-G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi and P. B. Sigler, *Nature*, 1988, **335**, 321.
31 Z. Shakked, G. Guzikevich-Guerstein, F. Frolow, D. Rabinovich, A. Joachimiak and P. B. Sigler, *Nature*, 1994, **368**, 469.
32 L. Urpi, V. Tereshko, L. Malinina, T. Huynh-Dunh and J. A. Subirana, *Nat. Struct. Biol.*, 1996, **3**, 325.
33 G. G. Privé, K. Yanagi and R. E. Dickerson, *J. Mol. Biol.*, 1991, **217**, 177.
34 K. Yanagi, G. G. Privé and R. E. Dickerson, *J. Mol. Biol.*, 1991, **217**, 201.
35 A. Lipanov, M. L. Kopka, M. Kaczor-Grzeskowiak, J. Quintana and R. E. Dickerson, *Biochemistry*, 1993, **32**, 1373.
36 V. Tereshko, L. Urpi, L. Malinina, T. Huynh-Dinh and J. A. Subirana, *Biochemistry*, 1996, **35**, 11 589.
37 M. A. El Hassan and C. R. Calladine, *J. Mol. Biol.*, 1996, **259**, 95.
38 M. A. El Hassan and C. R. Calladine, *Philos. Trans. R. Soc. London, A*, 1997, **355**, 43.
39 C. A. Hunter, *J. Mol. Biol.*, 1993, **230**, 1025.
40 C. A. Hunter and X.-J. Lu, *J. Mol. Biol.*, 1997, **265**, 603.
41 S. L. Lam and S. C. F. Au-Yeung, *J. Mol. Biol.*, 1997, **266**, 745.
42 Y. Timsit and D. Moras, *J. Mol. Biol.*, 1995, **251**, 629.
43 D. S. Goodsell, K. Grzeskowiak and R. E. Dickerson, *Biochemistry*, 1995, **34**, 1022.
44 A. A. Wood, C. M. Nunn, J. O. Trent and S. Neidle, *J. Mol. Biol.*, 1997, **269**, 827.
45 N. Spink, C. M. Nunn, J. Vojtechovsky, H. M. Berman and S. Neidle, *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 10 767.
46 H. Qui, J. C. Dewan and N. C. Seeman, *J. Mol. Biol.*, 1997, **267**, 881.
47 M. C. Wahl and M. Sundaralingam, *Nucleic Acid Reviews*, 1997, in press.
48 M. Sundaralingam, in *Oxford Handbook of Nucleic Acid Structure* (ed. S. Neidle), Oxford University Press, 1998, in the press.
49 D. B. Tippin and M. Sundaralingam, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1996, **52**, 997.
50 D. B. Tippin and M. Sundaralingam, *J. Mol. Biol.*, 1997, **267**, 1171.
51 M. Eisenstein and Z. Shakked, *J. Mol. Biol.*, 1995, **248**, 662.
52 D. B. Tippin and M. Sundaralingam, *Biochemistry*, 1997, **36**, 536.
53 C. M. Nunn and S. Neidle, *J. Mol. Biol.*, 1996, **256**, 340.
54 S. L. Ginell, S. Kuzmich, R. A. Jones and H. M. Berman, *Biochemistry*, 1990, **29**, 10 461.
55 J. Vojtechovsky, M. D. Eaton, R. Jones and H. M. Berman, *Biochemistry*, 1995, **34**, 16 632.
56 G. A. Leonard, J. Thomson, W. P. Watson and T. Brown, *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 9573.
57 N. T. Thuong and C. Hélène, *Angew. Chem., Int. Ed. Eng.*, 1993, **32**, 666.
58 S. Neidle, *Anti-Cancer Drug Des.*, 1997, **12**, 433.
59 L. Betts, J. A. Josey, J. M. Veal and S. R. Jordan, *Science*, 1995, **270**, 1838.
60 C. M. Nunn, E. Garman and S. Neidle, *Biochemistry*, 1997, **36**, 4792.
61 L. van Meervelt, D. Vileghe, A. Dautant, B. Gallois, G. Précigoux and O. Kennard, *Nature*, 1995, **374**, 742.
62 D. Vileghe, L. van Meervelt, A. Dautant, B. Gallois, G. Précigoux and O. Kennard, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1996, **52**, 766.
63 D. Vileghe, L. van Meervelt, A. Dautant, B. Gallois, G. Précigoux and O. Kennard, *Science*, 1996, **273**, 1702.
64 C. M. Nunn and S. Neidle, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, in the press.
65 I. Berger, C. H. Kang, A. Fredian, R. Ratliff, R. Moyzis and A. Rich, *Nat. Struct. Biol.*, 1995, **2**, 416.
66 C. H. Kang, I. Berger, C. Lockshin, R. Ratliff, R. Moyzis and A. Rich, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 11 636.
67 C. H. Kang, I. Berger, C. Lockshin, R. Ratliff, R. Moyzis and A. Rich, *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 3874.
68 D. Rhodes and R. Giraldo, *Cur. Opin. Struct. Biol.*, 1995, **5**, 311.
69 C. H. Kang, X. Zhang, R. Ratliff, R. Moyzis and A. Rich, *Nature*, 1992, **356**, 126.
70 P. Schultze, F. W. Smith and J. Feigon, *Structure* (*London*), 1994, **2**, 221.
71 G. Laughlin, A. I. H. Murchie, D. G. Norman, M. H. Moore, P. C. E. Moody, D. M. J. Lilley and B. Luisi, *Science*, 1994, **265**, 520; K. Phillips, Z. Dauter, A. I. H. Murchie, D. M. J. Lilley and B. Luisi, *J. Mol. Biol.*, 1997, **273**, 171.
72 G. A. Leonard, S. Zhang, M. R. Peterson, S. J. Harrop, J. R. Helliwell, W. B. T. Cruse, B. Langlois d'Estaintot, O. Kennard, T. Brown and W. N. Hunter, *Structure*, 1995, **3**, 335.
73 See for example: J. L. Sussman, S. R. Holbrook, R. W. Warrant, G. M. Church and S.-H. Kim, *J. Mol. Biol.*, 1978, **123**, 607; B. E. Hingerty, R. S. Brown and A. Jack, *J. Mol. Biol.*, 1978, **124**, 523; M. B. Comarmond, R. Giege, J. C. Thierry, D. Moras and J. Fischer, *Acta Crystallogr., Sect. B: Struct. Sci.*, 1986, **42**, 272.
74 For example; J. M. Rosenberg, N. C. Seeman, R. O. Day and A. Rich, *J. Mol. Biol.*, 1976, **104**, 145; H. M. Berman, W. Stallings, H. L. Carrell, J. P. Glusker, S. Neidle, G. Taylor and A. Achari, *Biopolymers*, 1979, **18**, 2405.
75 S. R. Holbrook and S.-H. Kim, *Biopolymers*, 1997, **44**, 3.
76 S. Portmann, N. Usman and M. Egli, *Biochemistry*, 1995, **34**, 7569; M. Egli, S. Portmann and N. Usman, *Biochemistry*, 1996, **35**, 8489.
77 T. E. Haran, Z. Shakked, A. H.-J. Wang and A. Rich, *J. Biomol. Struct. Dynam.*, 1986, **5**, 199; M. Eisenstein and Z. Shakked, *J. Mol. Biol.*, 1995, **248**, 662.
78 H. Schindelin, M. Zhang, R. Bald, J.-P. Fuerste, V. A. Erdmann and U. Heinemann, *J. Mol. Biol.*, 1989, **209**, 459.
79 S. Lorenz, C. Betzel, E. Raderschall, Z. Dauter, K. S. Wilson and V. E. Erdmann, *J. Mol. Biol.*, 1991, **219**, 399.
80 C. Betzel, S. Lorenz, J. P. Furste, R. Bald, M. Zhang, T. R. Schneider, K. S. Wilson and V. A. Erdmann, *FEBS Lett.*, 1994, **351**, 159.
81 A. Nolte, S. Klussman, S. Lorenz, R. Bald, C. Betzel, Z. Dauter, K. Wilson, J. P. Furste and V. A. Erdmann, *FEBS Lett.*, 1995, **374**, 292.
82 A. C. Dock-Bregeon, B. Chevrier, A. Podjarny, J. Johnson, J. S. De Bear, G. R. Gough, P. T. Gilham and D. Moras, *J. Mol. Biol.*, 1989, **209**, 459.
83 M. C. Wahl, S. T. Rao and M. Sundaralingam, *Nat. Struct. Biol.*, 1996, **3**, 24.
84 G. A. Leonard, J. B. Thomson, W. P. Watson and W. N. Hunter, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1995, **51**, 136.
85 S. E. Lietzke, C. L. Barnes and C. E. Kundrot, *Structure* (*London*), 1996, **4**, 917.
86 K. J. Baeyens, H. L. De Bondt and S. R. Holbrook, *Nat. Struct. Biol.*, 1995, **2**, 56.
87 S. R. Holbrook, C. Cheong, I. Tinoco, Jr. and S.-H. Kim, *Nature*, 1991, **353**, 579.
88 G. A. Leonard, K. E. McAuley-Hecht, S. Ebel, D. M. Lough, T. Brown and W. N. Hunter, *Structure* (*London*), 1994, **2**, 483.
89 K. J. Baeyens, H. L. De Bondt, A. Pardi and S. R. Holbrook, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 12 851.
90 H. W. Pley, K. M. Flaherty and D. B. McKay, *Nature*, 1994, **372**, 68.
91 W. G. Scott, J. T. Finch and A. Klug, *Cell*, 1995, **81**, 991.
92 W. G. Scott, J. B. Murray, J. R. P. Arnold, B. L. Stoddard and A. Klug, *Science*, 1996, **274**, 2065.
93 J. A. Doudna, *Structure* (*London*), 1995, **3**, 747; W. G. Scott and A. Klug, *TIBS*, 1996, **21**, 220.
94 J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech and J. A. Doudna, *Science*, 1996, **273**, 1678.
95 M. Teng, Y.-C. Liaw, G. A. van der Marel, J. H. van Boom and A. H.-J. Wang, *Biochemistry*, 1989, **28**, 4923.
96 M. C. Wahl, C. Ban, C. Sekharudu, B. Ramakrishnan and M. Sundaralingam, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1996, **52**, 655.
97 X. Chen, B. Ramakrishnan and M. Sundaralingam, *Nat. Struct. Biol.*, 1995, **2**, 733.
98 W. B. T. Cruse, P. Saludjian, E. Biala, P. Strazewski, T. Prange and O. Kennard, *Proc Natl. Acad. Sci. USA.*, 1994, **91**, 4160.
99 A. H.-J. Wang, S. Fujii, J. H. van Boom, G. A. van der Marel, S. A. A. van Boeckel and A. Rich, *Nature*, 1982, **299**, 601.
100 M. Egli, N. Usman and A. Rich, *Biochemistry*, 1993, **32**, 3221.
101 C. Ban, B. Ramakrishnan and M. Sundaralingam, *J. Mol. Biol.*, 1994, **236**, 275.
102 C. Ban, B. Ramakrishnan and M. Sundaralingam, *Nucleic Acids Res.*, 1994, **22**, 5466.
103 Y.-G. Gao, H. Robinson, J. H. van Boom and A. H.-J. Wang, *Biophys. J.*, 1995, **69**, 559.
104 P. Lubini, W. Zurcher and M. Egli, *Chem. Biol.*, 1994, **1**, 39.
105 N. C. Horton and B. C. Finzel, *J. Mol. Biol.*, 1996, **264**, 521.
106 S. Portmann, S. Grimm, C. Workman, N. Usman and M. Egli, *Chem. Biol.*, 1996, **3**, 173.

107 *Molecular Aspects of Anticancer Drug–DNA interactions*, eds. S. Neidle and M. J. Waring, 1993, Macmillan Press, London.
108 L. H. Hurley, *J. Med. Chem.*, 1989, **32**, 2027.
109 D. Sun and L. H. Hurley, *Chem. Biol.*, 1995, **2**, 457.
110 S.-Y. Chang, J. J. Welch, F. J. Rauscher III and T. A. Beerman, *J. Biol. Chem.*, 1996, **271**, 23 999.
111 S.-Y. Chang, T. C. Bruice, J. C. Azizhan, L. Gawron and T. A. Beerman, *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 2811.
112 See, for example, C. A. Frederick, L. D. Williams, G. Ughetto, G. A. van der Marel, J. H. van Boom, A. Rich and A. H.-J. Wang, Biochemistry, 1990, **29**, 2538.
113 A. Dautant, B. Langlois d'Estaintot, B. Gallois, T. Brown and W. N. Hunter, *Nucleic Acids Res.*, 1995, **23**, 1710.
114 J. B. Chaires, F. Leng, T. Przewloka, L. Fokt, Y. H. Ling, R. Perez-Soler and W. Priebe, *J. Med. Chem.*, 1997, **40**, 261.
115 G. G. Hu, X. Shui, F. Leng, W. Priebe, J. B. Chaires and L. D. Williams, *Biochemistry*, 1997, **36**, 5940.
116 C. K. Smith, J. A. Brannigan and M. H. Moore, *J. Mol. Biol.*, 1996, **263**, 237.
117 W. B. T. Cruse, P. Saludjian, P. Y. Leroux, G. Léger, D. El Manouni and T. Prangé, *J. Biol. Chem.*, 1996, **271**, 15 558.
118 S. Kamitori and F. Takusagawa, *J. Am. Chem. Soc.*, 1994, **116**, 4154.
119 M. Shinomiya, W. Chu, R. G. Carlson, R. F. Weaver and F. Takusagawa, *Biochemistry*, 1995, **34**, 8481.
120 L. A. Lipscomb, F. X. Zhou, S. R. Prsenell, R. J. Woo, M. E. Peek, R. R. Plaskon and L. D. Williams, *Biochemistry*, 1996, **35**, 2818.
121 P. B. Dervan, *Science*, 1986, **232**, 464.
122 M. Mrksich and P. B. Dervan, *J. Am. Chem. Soc.*, 1995, **117**, 3325.
123 C. R. Watts, S. M. Kerwin, G. L. Kenyon, I. D. Kuntz and D. A. Kallick, *J. Am. Chem. Soc.*, 1995, **117**, 9941.
124 D. Sengupta, A. Blaskó and T. C. Bruice, *Bioorg. Med. Chem.*, 1996, **4**, 803.
125 J. M. Gottesfield, L. Neely, J. W. Trauger, E. E. Baird and P. B. Dervan, *Nature*, 1997, **387**, 202.
126 L. Tabernero, J. Bella and C. Alemán, *Nucleic Acids Res.*, 1996, **24**, 3458.
127 C. M. Nunn and S. Neidle, *J. Med. Chem.*, 1995, **38**, 2317.
128 C. A. Laughton, F. Tanious, C. M. Nunn, D. W. Boykin, W. D. Wilson and S. Neidle, *Biochemistry*, 1996, **35**, 5655.
129 J. O. Trent, G. R. Clark, A. Kumar, W. D. Wilson, D. W. Boykin, J. E. Hall, R. R. Tidwell, B. L. Blagburn and S. Neidle, *J. Med. Chem.*, 1996, **39**, 4554.
130 S. Neidle, *Biopolymers*, 1997, **44**, 105.
131 N. Spink, D. G. Brown, J. V. Skelly and S. Neidle, *Nucleic Acids Res.*, 1994, **22**, 1607.
132 M. C. Vega, I. G. Saez, J. Aymami, G. A. van der Marel, J. H. van Boom, A. Rich and M. Coll, *Eur. J. Biochem.*, 1994, **222**, 721.
133 A. A. Wood, C. M. Nunn, A. Czarny, D. W. Boykin and S. Neidle, *Nucleic Acids Res.*, 1995, **23**, 3678.
134 G. R. Clark, E. J. Gray, S. Neidle, Y.-H. Li and W. Leupin, *Biochemistry*, 1996, **35**, 13 745.
135 D. Goodsell and R. E. Dickerson, *J. Med. Chem.*, 1986, **29**, 727.
136 G. R. Clark, D. W. Boykin, A. Czarny and S. Neidle, *Nucleic Acids Res.*, 1997, **25**, 1510.
137 G. R. Clark, C. J. Squire, E. J. Gray, W. Leupin and S. Neidle, *Nucleic Acids Res.*, 1996, **24**, 4882.
138 D. S. Goodsell, M. L. Kopka and R. E. Dickerson, *Biochemistry*, 1995, **34**, 4983.
139 D. S. Goodsell, H. L. Ng, M. L. Kopka, J. W. Lown and R. E. Dickerson, *Biochemistry*, 1995, **34**, 16654.
140 X. Chen, B. Ramakrishnan, S. T. Rao and M. Sundaralingam, *Nat. Struct. Biol.*, 1994, **1**, 169.
141 X. Chen, B. Ramakrishnan and M. Sundaralingam, *J. Mol. Biol.*, 1997, **267**, 1157.
142 P. M. Takahara, A. C. Rosenzweig, C. A. Frederick and S. J. Lippard, *Nature*, 1995, **377**, 649.
143 P. M. Takahara, C. A. Frederick and S. J. Lippard, *J. Am. Chem. Soc.*, 1996, **118**, 12 309.
144 M. L. Kopka, D. S. Goodsell, I. Baikalov, K. Grzeskowiak, D. Cascio and R. E. Dickerson, *Biochemistry*, 1994, **33**, 13 593.