Supramolecular DNA recognition†‡

Michael J. Hannon

Received 6th September 2006
First published as an Advance Article on the web 17th November 2006
DOI: 10.1039/b606046n

Non-covalent DNA-recognition by synthetic agents is surveyed in this *tutorial review*, and contrasted with biomolecular DNA-recognition. The principles and forces involved in DNA recognition are similar to those seen elsewhere in the wider field of supramolecular chemistry, although the size, surface dimensions and nature of DNA introduce new possibilities and challenges. Recent discoveries of new binding motifs, and new biological structural and genomic information from bioscience, are affording new opportunities for supramolecular chemistry, where shape, fit and orientation play such an important role.

School of Chemistry, University of Birmingham, Edgbaston, Birmingham, UK B15 2TT. E-mail: m.j.hannon@bham.ac.uk; Fax: + 44 121 414 7871

† Electronic supplementary information (ESI) available: links to PDB visualizations are included in the online supplementary data. See DOI: 10.1039/b606046n

‡ Dedicated to Professor Jean-Marie Lehn, still providing inspiration to the field 20 years after the Nobel Prize.

Michael J. Hannon

Mike Hannon studied Natural Sciences at the University of Cambridge (87–90) and remained there to undertake his PhD (90–93) in Inorganic Chemistry as the Ciba-Geigy PhD Scholar, under the supervision of Professor Ed Constable. He then moved to Strasbourg on a Royal Society ESEP Fellowship to undertake postdoctoral researcher with Professor Jean-Marie Lehn for what was an incredibly stimulating and enjoyable 12 month period. In October 1994 he took up a Lectureship at the University of Warwick where he was successively promoted to Senior Lecturer, Reader and then Professor. In the autumn of 2005 he moved to the Chair of Chemical Biology in the School of Chemistry at the University of Birmingham where he is currently the Director of the Molecular Synthesis and Chemical Biology research theme. His research interests span across supramolecular and bioinorganic chemistry. In particular his group have developed the use of supramolecular assemblies as agents for DNA recognition. His research in this area has led to awards from the Royal Society of Chemistry of the Sir Edward Frankland Fellowship Prize Award 2004–5 and the 2002 Bob Hay lectureship Prize Award. His research has benefited from collaborations with a number of other research groups across Europe and he has recently been elected as the Chairman of the EU COST Action D39 on metallo-drugs that aims to coordinate, network and enhance European activity in this area.

Introduction

DNA is the biomolecule in which the genetic data of most organisms is encoded. Its sequence defines many features ranging from organism type through physical traits to disease susceptibility. The information encoded in the DNA sequence is put into practice principally through the action of proteins. The DNA sequence is copied onto RNA molecules, which are then used in protein synthesis to encode a specific protein sequence. Inherent in this process is a dramatic information amplification of the genetic information: the single master copy of information in the DNA is used to create multiple RNA copies which in turn can each be used to create multiple proteins. The fact that expression of the DNA information is regulated by proteins, which bind to DNA, allows the information expression to respond to the environment. Protein binding to DNA is normally reversible and non-covalent in nature.

With the genetic sequence of many organisms now known (and particularly the human genome) attention is turning to establishing ways to control specific gene expression. This could be achieved by creating agents which can bind selectively to specific genes and turn their expression either on or off. The ability to turn a gene on or off is important in trying to elucidate the complex and intertwined biological pathways in the cell and also in medical treatment. Many diseases (most notably cancers) are expressed through protein action but actually originate at the DNA level, in the code or in its processing. The information amplification from DNA to proteins, makes DNA a particularly attractive medicinal target, since in principle a single drug molecule per cell could change the expression of a gene on DNA; multiple drug molecules per cell are required to act on the multiple copies of proteins. Both small synthetic molecules, or larger modified biomolecules have potential as such drugs, although small synthetic agents are more readily transported inside cells.

This field is one which supramolecular chemists are ideally placed to exploit and where supramolecular interactions dominate. This review is intended as an introduction that highlights key facets of the fascinating area of DNA recognition and explains why supramolecular chemistry has so much to offer the field. It focuses particularly on non-covalent (supramolecular) recognition of DNA although

 This journal is © The Royal Society of Chemistry 2007

covalent approaches are touched on briefly in highlighting their supramolecular context. As will be seen, the principles and forces involved in DNA recognition are very similar to those seen elsewhere in the wider field of supramolecular chemistry, although the size, surface dimensions and nature of the DNA biomacromolecule introduce new possibilities and challenges. Particular attention is paid to DNA-recognition by synthetic agents. Although protein–DNA recognition is multi-faceted and often complex, nature remains the most accomplished of supramolecular chemists, and aspects of biomolecular DNA recognition are highlighted. The subject matter is inherently visual and where possible views of X-ray crystal structures are selected to illustrate the various binding modes. PDB reference codes are given for each figure for ease of location of the data in the protein data bank or nucleic acids databank should readers wish to view and explore structures in more detail.†

B-DNA structure and recognition

DNA is comprised of deoxyribose sugars linked together in a chain through phosphate groups, and with a nucleotide-base attached to each sugar (Fig. 1).[1] Four different bases are present (guanine, G; adenine, A; cytosine, C; thymine, T) two of which are purines (G, A) and two pyrimidines (C, T). The double-stranded double-helical structure of DNA, elucidated by Watson and Crick over fifty years ago, is one of the scientific icons of the 20th Century.[2] This structure, termed B-DNA, has the two anionic sugar phosphate backbones wrapped around each other in a right-handed double-helix, with the bases hydrogen-bonded together in pairs (A with T and G with C) in the heart of the helix (Fig. 2). The hydrophilic sugar-phosphate units point out into solution while the more hydrophobic bases are in the core. The bases are perpendicular to the helical axis and are stacked in a parallel fashion upon each other (face–face π–π interactions) with a regular inter-planar separation of 3.5 Å. The attachment

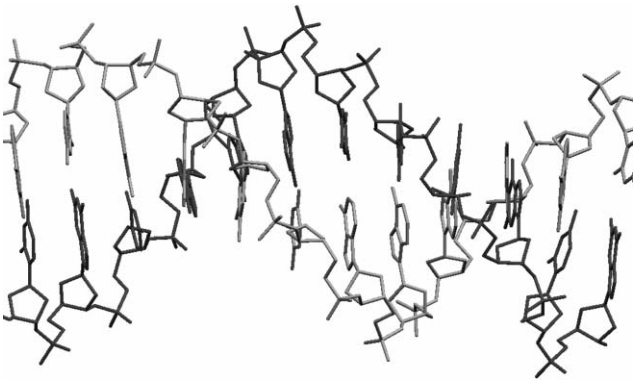

Fig. 2 The (Watson–Crick) double-helical structure of B-DNA.

points of the bases to the sugar-phosphate backbone are offset with respect to the hydrogen bonds between the bases. This leads to two distinct grooves in the double-helical structure, termed minor and major. The sequence of the DNA bases along the polymeric sugar-phosphate chain encodes the genetic information and the Watson–Crick hydrogen bond recognition of G with C and of A with T is key to the accurate reading of and replication of the genetic information. B-DNA is believed to be the most prevalent form of DNA in biological systems although other double-helical forms such as the left-handed Z-DNA and right-handed A-DNA (shorter and fatter than B-DNA) have been observed crystallographically and may have a more limited biological relevance.

Molecular recognition of B-DNA can take place in 5 distinct ways: major groove recognition; minor groove recognition; sugar-phosphate backbone binding; intercalation between the base pairs; covalent binding or metal-coordination to the bases. These DNA recognition modes were recognised in the 1960s and have dominated the field of DNA-recognition since, although very recently two new modes have been reported which offer new opportunities and impetus to the field.

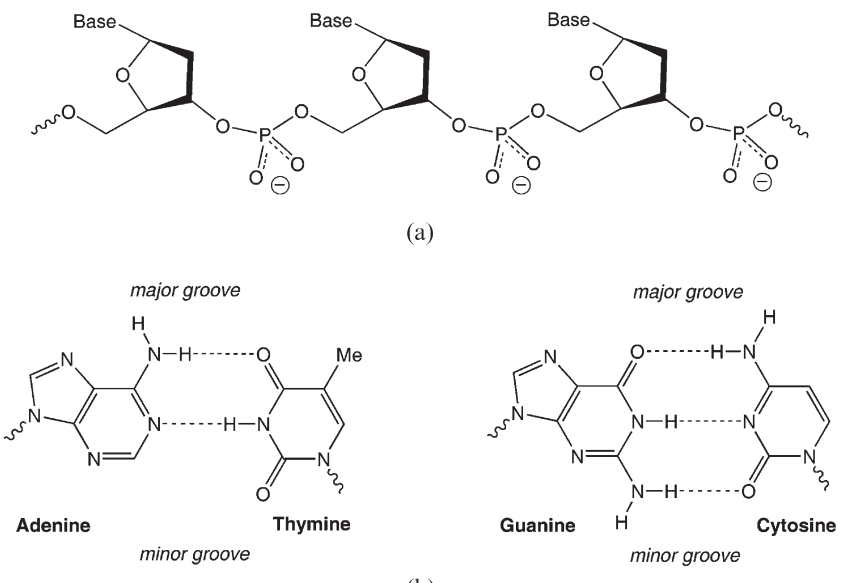

Fig. 1 The components of DNA: (a) The deoxyribose phosphate backbone. (b) The DNA bases and their Watson–Crick hydrogen-bond pattern.

Major groove recognition

Major groove recognition by proteins. Proteins frequently achieve DNA sequence recognition by binding in or around the major groove and forming specific hydrogen-bond contacts to the edges of the base pairs. This is because this groove not only shows the greater variation in size and shape with base sequence but perhaps more importantly has a greater number of, and variation in the pattern of, hydrogen-bond donor and acceptor units to which the protein can bind. Major groove recognition usually involves cylindrical binding units based on alpha helices, and the units are just the right size and shape to fit snugly into the major groove and too large for the minor groove. Motifs employed include helix-turn-helix structures, zinc fingers (Fig. 3) and leucine zippers.[1]

Amino acid residues employed on the outside of the alpha-helical units to form hydrogen-bonds to the DNA bases include arginine (guanadinium group), histidine (imidazole NH) lysine (NH_2), serine (OH), aspargine, glutamine (both $CONH_2$) and glutamate (carboxylate). There is, however, no simple one-to-one correlation between a certain amino acid side-chain and a particular DNA base; amino acids are found binding to different bases in different structures and can even bridge between two stacked bases forming hydrogen bonds to both. The matter is further complicated by the fact that the H-bond interactions may on occasions be relayed through bridging water molecules incorporated between the protein surface and the DNA groove.[3] The level of complexity makes it almost impossible to design *de novo* synthetic peptides which will recognize a specific DNA sequence. Some success can be achieved by modifying the residues of an established protein unit to affect its specificity or by creating peptides that combine different units whose individual sequence specificity is known. More usefully, libraries of peptides may be prepared, with rounds of selection and amplification used to identify the best binder for a given sequence. Particular successes have been achieved with synthetic peptides containing zinc finger motifs.[4]

Major groove recognition by oligonucleotides. Oligonucleotides (synthetic or natural) are the second class of agents which are able to recognise the major groove of DNA.[6] They do so by forming hydrogen-bonds to the major groove edges of the purine nucleobases, forming pairing motifs termed Hoogsteen or reverse-Hoogsteen (Fig. 4). T or A can recognise an A base, while G or a protonated C can recognise a G base. Such recognition leads to a triple-stranded DNA known as triplex DNA (Fig. 5) and can only be formed when the double-stranded DNA has an extended (and uninterrupted) sequence of purines (A or G). The recognition is sequence-specific and binding of oligonucleotides to DNA to form triplexes has been used to regulate gene expression. Oligonucleotides, however, are not selective for DNA and can also bind to complementary RNAs to form RNA–DNA duplexes.

Since both DNA and RNA binding can regulate protein synthesis, a variety of oligonucleotide anologues have entered clinical trials and one agent, Vitravene, that is believed to act by binding to viral DNA, has entered clinical use.[6] Vitravene is a phosphorothioate oligonucleotide in which a phosphate–oxygen bond has been replaced by a phosphate–sulfur bond in the nucleic acid backbone. This synthetic modification is one of a number that have been shown to confer greater resistance to nuclease degradation.

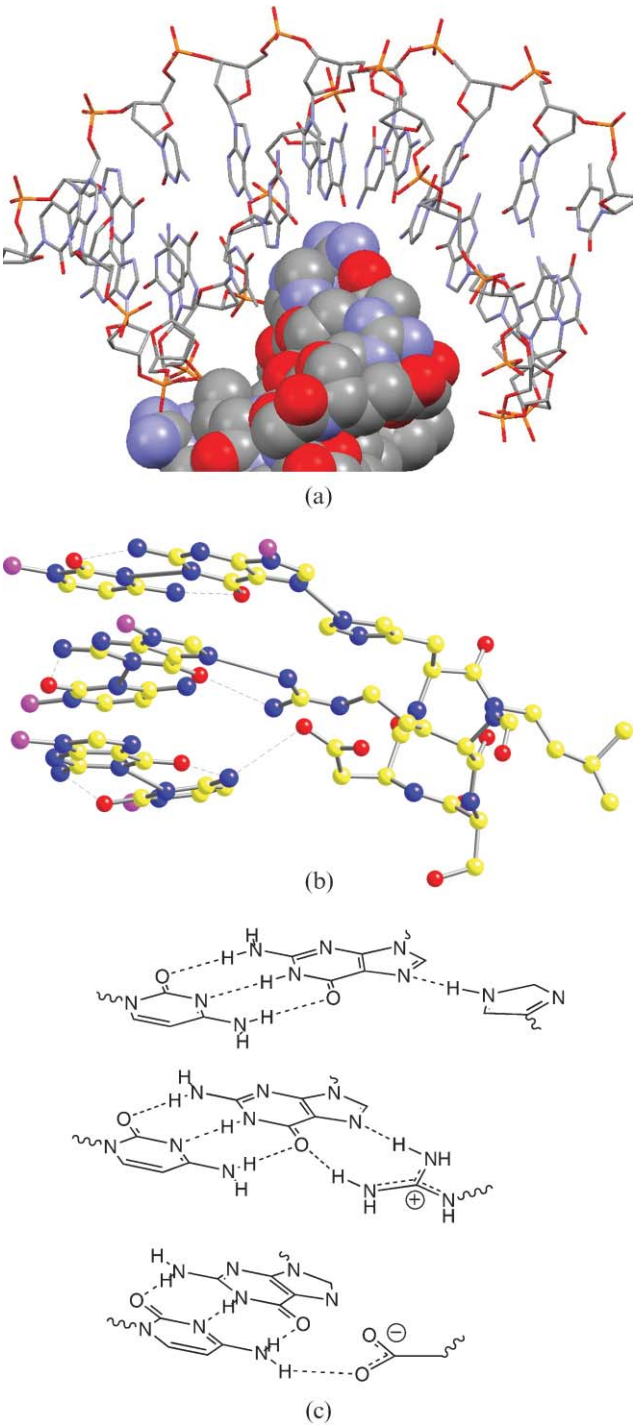

Fig. 3 (a) Illustration of the fit of a zinc finger unit of a polypeptide (space-filling) into the major groove of DNA (stick). (b) Close up of the H-bonding interactions of three protein residues with bases in the major groove in this structure. Top: Histidine NH – Guanine N(7). Middle: Arginine NH, NH – Guanine N(7), O(6). Bottom: Aspartate O – Cytosine N(4)H. (c) Schematic of the H-bond interactions (PDB ref. 1A1J†).[5]

This journal is © The Royal Society of Chemistry 2007

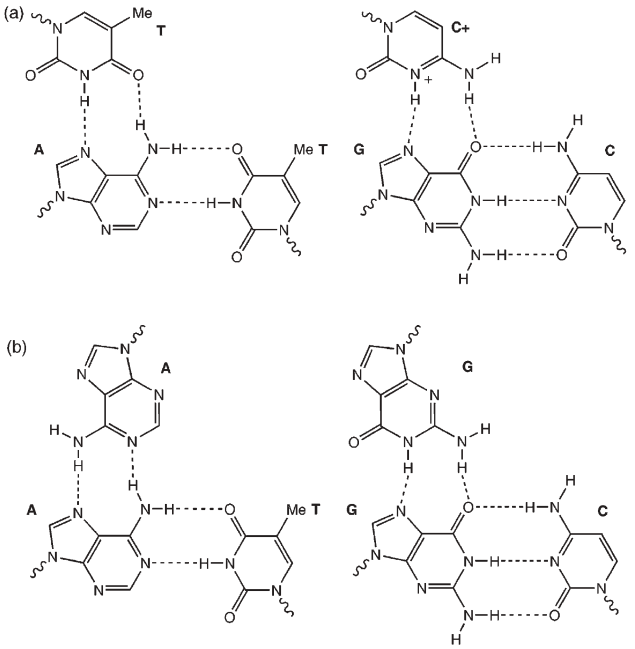

Fig. 4 Hoogsteen (a) and reverse-Hoogsteen (b) hydrogen bonding in the major groove enables triplex formation.

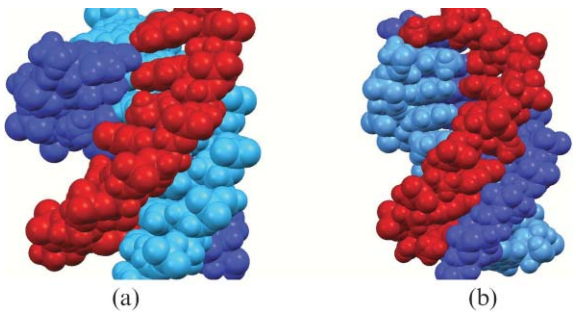

Fig. 5 NMR structures of Hoogsteen, pyrimidine.purine.pyrimidine (left) and reverse-Hoogsteen, purine.purine.pyrimidine (right) DNA triplexes. In each case the Watson–Crick strands are shaded blue and the additional strand in the major groove red. (PDB refs. 149D and 134D†).[7]

Triple-stranded DNA structures may also be naturally present in the DNAs of organisms, perhaps related to DNA repair or regulation, although precise roles have yet to be fully elucidated. Although the hydrogen-bonding capabilities of oligonucleotides makes them effective for sequence-specific recognition of polypurine tracts, their size and shape means that they do not fill the major groove as effectively as the protein DNA recognition units (compare Figs. 3 and 5) and nature seems not to use oligonucleotides (RNAs or DNAs) for duplex DNA recognition.

DNA recognition by PNA

An important issue for triplex formation, is the unfavourable electrostatic term that will be associated with using a polyanionic oligonucleotide to recognise the DNA polyanion (particularly for Hoogsteen recognition which places two sugar-phosphate backbones close together: Fig. 5). With this in mind Nielsen set out to design a neutral oligonucleotide analogue in which the sugar phosphate backbone was replaced with a polyamide (Fig. 6).[8] He termed these analogues peptide nucleic acids (PNAs). Although they are capable of forming analogous $(DNA)_2(PNA)$ triplexes with double-stranded DNA, it soon became apparent that their more common way of binding was to displace one DNA strand to form $(DNA)(PNA)_2$ triplexes. Thus the DNA double-helix is opened up and two PNA strands invade, with one binding in a Watson–Crick fashion to one of the DNA strands and the other binding to the major groove of that PNA–DNA double-helix (Fig. 7).

It is striking that the classes of agents (peptides and oligonucleotides) that target the DNA major groove are natural or semi-synthetic biomolecules or analogues. Synthetic molecules that target the DNA major groove are rare. Rather, because of their smaller size, synthetic molecules tend to bind in the minor groove or intercalate between the base pairs. These binding modes are considered in the following sections.

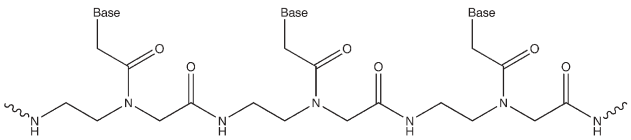

Fig. 6 General structure of PNA.

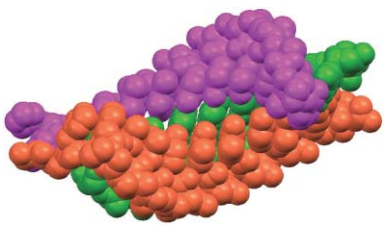

Fig. 7 Recognition of DNA (green) by two PNA units. In this structure the Watson–Crick (pink) and Hoogsteen (orange) strands are incorporated into a single molecular strand. (PDB ref. 1PNN†).[9]

Minor groove recognition

Minor groove recognition by synthetic agents. Considerable research effort has focused on the design of synthetic minor groove binders and such agents have found clinical application in treating cancers and protozoal diseases and as anti-viral and anti-bacterial agents.[10] Minor groove binding agents (Fig. 8) include the natural molecule distamycin A, and synthetic diarylamidines (such as DAPI, berenil, and pentamidine) and bis-benzimidazoles (such as Hoechst 33258).

DAPI was investigated in the clinic as an anti-parasitic agent although its use is limited by side effects. Its more common use is as a blue-fluorescent stain for DNA. It is readily transported across membranes into cells and has proved popular in microscopy studies. It binds to AT-rich regions of double-stranded DNA in the minor groove and inhibits DNA and RNA polymerase. AT regions of B-DNA have a narrower minor groove than GC regions and this allows a snug fit of the drug against the walls of the groove (Fig. 9). The absence of

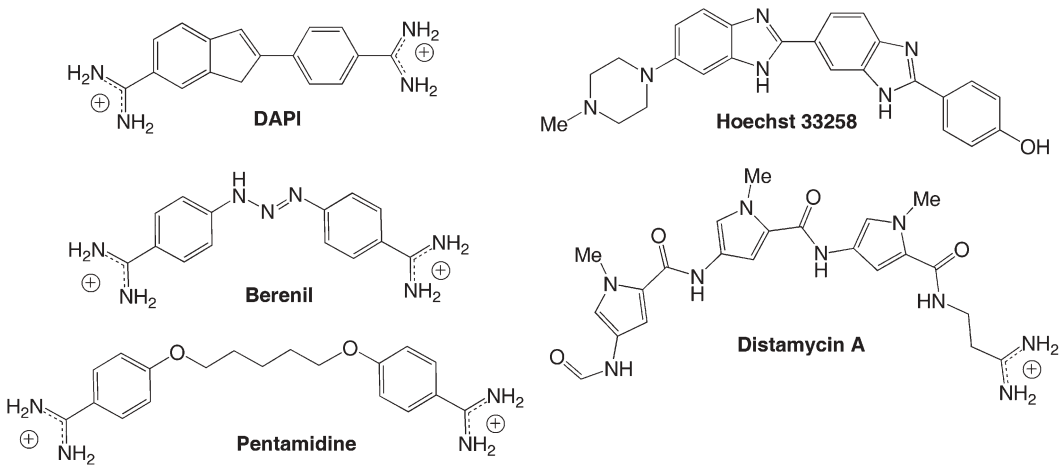

Fig. 8 Minor groove binding agents.

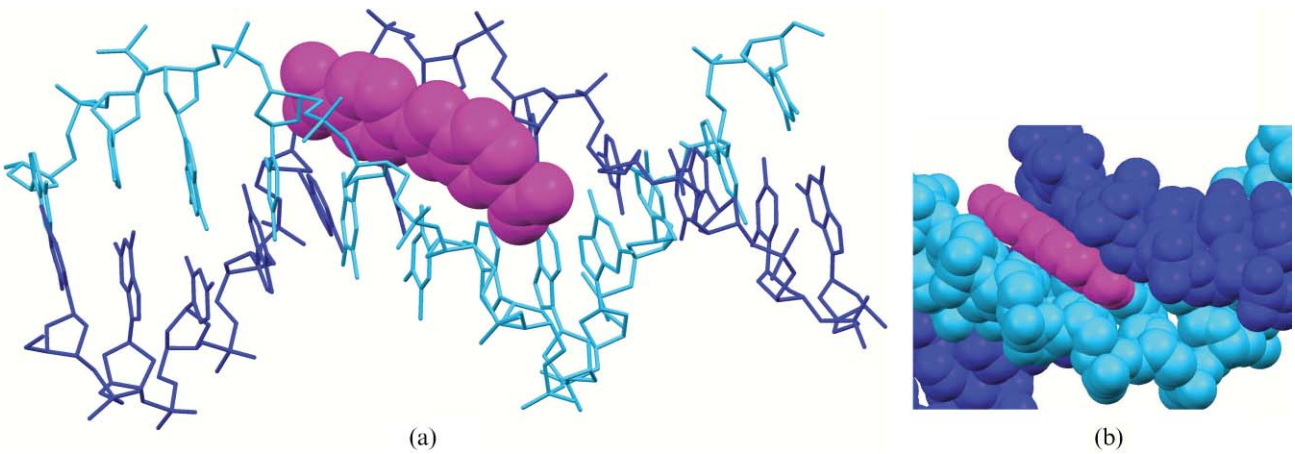

Fig. 9 (a) View of DAPI (pink) binding in the minor groove of DNA. (b) Close-up space-filling view showing the snug fit in the narrow and deep AT-rich minor groove. (PDB ref. 1D30†).[11]

the guanine-NH_2 group in the minor groove of AT regions, makes the groove slightly deeper and the drug is thus able to insert more fully. The phenyl and indole rings of DAPI lie parallel to the groove walls, stretching across a three base-pair sequence (ATT). The indole nitrogen forms a bifurcated hydrogen bond with the O2 atoms of two thymine bases. The cationic amidine groups at the ends of DAPI form hydrogen-bonds to adenine N3 and the electrostatic attraction of the cation for the DNA polyanion contributes significantly to the binding strength.[11] Indeed, although hydrogen-bonds or π-stacking interactions are important in determining the location of a synthetic agent on DNA, synthetic DNA binders are frequently cationic and electrostatic attraction is a major contribution to the strength of their binding.

Berenil is used in veterinary medicine for the treatment of the trypanosomiasis class of parasitic diseases (*e.g.* sleeping sickness). Its action is believed to involve binding to mitochondrial DNA in the kinetoplasts found at the base of the flagellum of the parasite. Berenil (like DAPI) has a particular affinity for AT-rich DNA sequences (Fig. 10). Pentamidine is also a diarylamidine anti-parasitic agent that shows preference for AT-rich DNA. It is used in the clinic to treat sleeping sickness and also pneumonia, particularly in HIV-positive patients.

The bisbenzimidazole Hoechst-33258 drug and its derivatives are a second class of minor groove binders. They readily enter cells, bind strongly in the DNA minor groove in AT-rich regions, spanning 4–5 base-pairs and inhibit transcription of

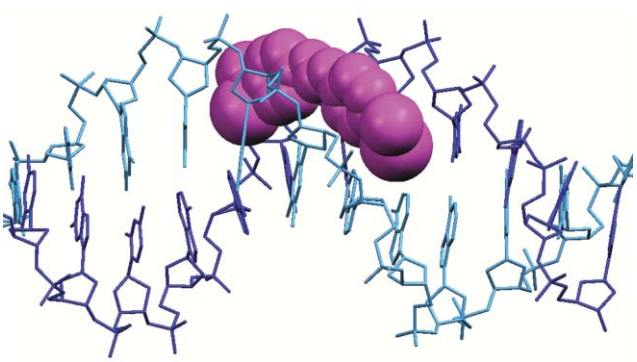

Fig. 10 Structure of berenil (pink) bound in the DNA minor-groove (PDB ref. 1D63†).[12]

This journal is © The Royal Society of Chemistry 2007

specific genes. They become strongly fluorescent on DNA-binding and like DAPI are widely used as stains in fluorescence microscopy. Their basic structure can be extended to allow longer DNA sequences to be recognized.

Distamycin A, a tri-(*N*-methylpyrrole) peptide, is a naturally occurring antibiotic with antibacterial and antiviral activity. Like the synthetic DNA-binders, distamycin is a cationic polyaromatic which binds to AT-rich DNA, forming hydrogen bonds from the amide NH to the N3 of A and O2 of T, supported by van der Waals contacts and electrostatic interactions. It binds to AT tracts of at least 4 base-pairs length; extending the structure increases the preferred length of AT tract. The basic distamycin structure has been elaborated in a number of studies which have led to the ambitious and remarkable studies from Dervan's laboratories which have established a set of codes by which synthetic agents based on modified distamycin-type structures can recognise the DNA sequence *via* the minor groove.[13]

The first experimental step in this process was simply to replace a methylpyrrole with a methylimidazole group; as anticipated the new nitrogen in the ring system is able to form hydrogen-bonds to the NH_2 of the guanine base in the minor groove. The key breakthrough was the discovery that, at high drug loading, two pyrrole/imidazole peptide strands could sit side-by-side in the minor groove (Fig. 11). At first sight, this is quite a surprise, given the snug fit of DAPI and berenil in their 1 : 1 complexes (Figs. 9 and 10). It reflects the fact that DNA is not a rigid structure and AT tracts are particularly flexible: indeed to accommodate two peptide strands, the minor groove must double in width. The two polyamide strands lie head-to-tail in the groove such that the cationic guanadinium ends are placed away from each other. (By contrast, in the case of diarylamidines the presence of guanadinium groups at each end of the groove binder would disfavour this 2 : 1 binding mode.) The pyrrole/imidazole rings of one strand are π-stacked with the amide carbonyls of the other (Fig. 11b). The importance of this structure is that each polyamide strand interacts with the bases of one strand of the DNA. Thus, using two strands, both bases in a base pair can be recognized. For this reason, hairpin linkers are often used to incorporate the two strands within one defined molecule. The hydrogen-bond base recognition patterns developed are shown in Fig. 12. The resulting arrays have been shown to be able to affect transcription of specific genes and have been combined with a variety of other DNA binding agents and with cytotoxics.

Minor groove recognition by proteins. While protein recognition of DNA (and especially sequence recognition) is more commonly achieved in the major groove, minor groove binding is also well established.[1] Some of the specific contacts, such as arginine binding to O2 of T, are similar to those seen with synthetic DNA binders. The protein structures used for such binding are most commonly β-sheet structures (α-helices being too large to fit in this smaller groove) and the surfaces more hydrophobic than those used in the major groove. Minor-groove protein recognition is also often linked to DNA bending (Fig. 21) and this is discussed in more detail below.

Intercalation

One of the most common ways that small aromatic molecules recognise DNA is through intercalation. Originally proposed

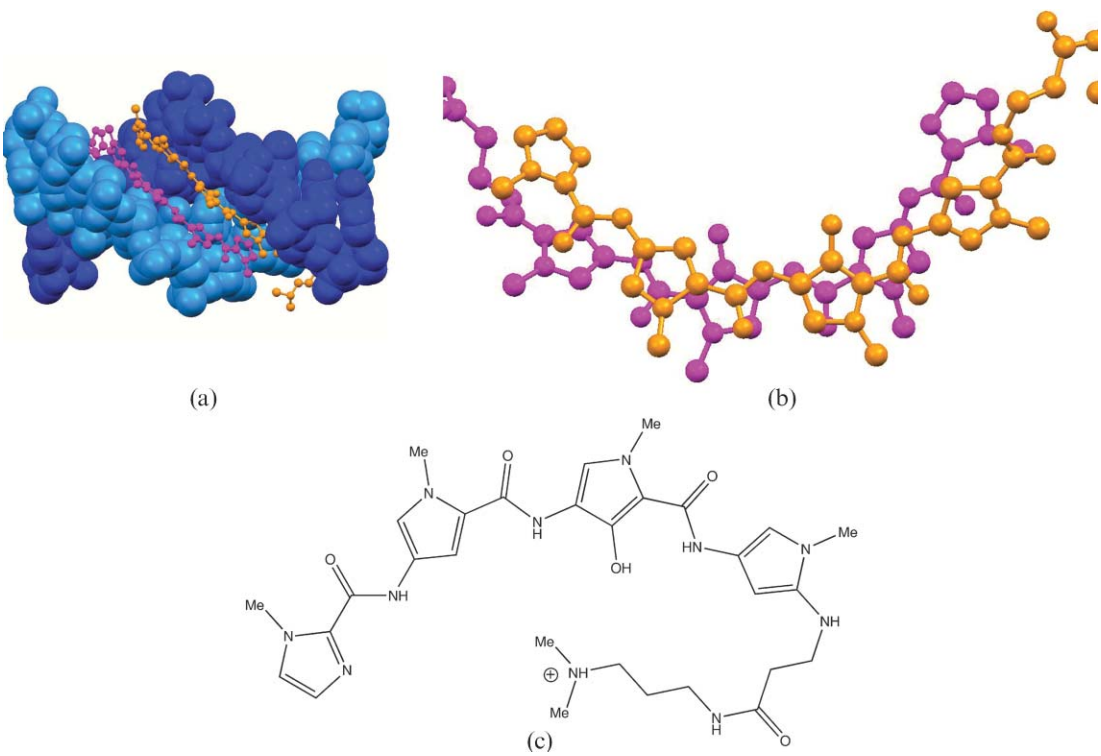

Fig. 11 (a) Two polyamide strands (shown in pink and orange) side-by-side in the DNA minor groove (b) view showing the relative orientations of the strands (c) The structure of the polyamide strand (PDB ref. 1CVX†).[14]

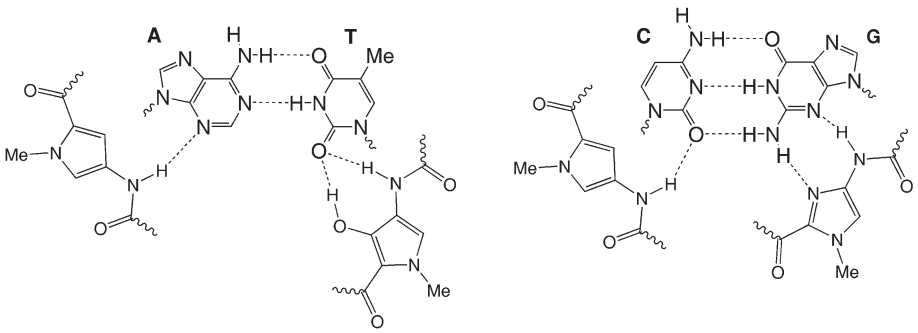

Fig. 12 Base-pair recognition patterns used in Dervan's polyamide minor-groove binders.

by Lerner in the 1960s,[15] this involves the gap between the stacked base-pairs being opened up and a planar aromatic molecule inserting into the resulting hydrophobic pocket, forming face–face π–π interactions with the bases above and below (Fig. 13). As with the synthetic minor-groove binders, electrostatics make an important contribution to the binding energy, and synthetic intercalators are frequently positively charged (Fig. 14). Intercalators can insert between the base pairs from either the major or the minor groove. The opening up of the inter-base-pair separation causes an unwinding of the DNA twist. The DNA backbone is not sufficiently flexible to allow a drug to be inserted into every gap between the bases; rather once one gap is filled, the adjacent ones must remain unfilled, giving a maximum loading of one drug per two gaps. This is referred to as neighbour-group exclusion principle. Intercalators are used as DNA stains (most notably ethidium) and also as anti-cancer agents. An early example of an anti-cancer drug that acts by intercalation is doxorubicin (trade name 'Adriamycin' or 'Rubex'). This drug, launched in the 1960s, gave much impetus to the field and has been followed by a number of variants. Doxorubicin and its analogues are classified by clinicians as an "anthracycline antiobiotics" And such intercalators remain important tools in the clinic for cancer treatment.[16]

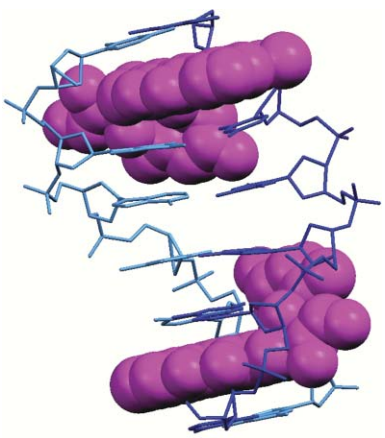

Fig. 13 Structure of a DNA oligonucleotide with two doxorubicin drugs intercalated between the base pairs, with insertion from the minor groove side (PDB ref. 1D12†).[17]

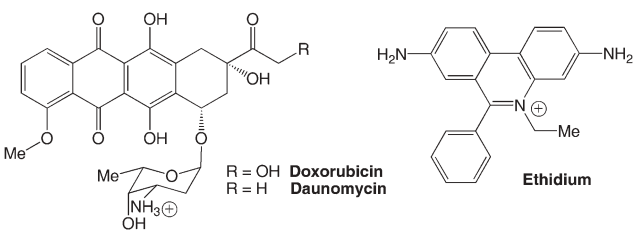

Fig. 14 Examples of organic intercalators.

Metallo-intercalators

An attractive way of imparting positive charge into an intercalator design is to incorporate a transition metal. The first example was Lippard's use of a planar platinum(II) terpyridine unit (Fig. 15). The advantage of platinum is its fairly inert nature and preference for square-planar geometry. Moreover it is a key component of the active anti-cancer drug cis-platin (See below).

However, it was soon realised that metals could be used not only to impart charge, but also other features such as luminescence (particularly ruthenium(II) polypyridine centres) and DNA cleavage (particularly related rhodium(III) complexes acting through photoinduced oxidation).[19] Since ruthenium and rhodium are octahedral metal centres, the molecular designs used tend to be based on a large planar ligand structure (usually a bidentate ligand) that can insert into the DNA and to which the metal can be attached. While the planar unit inserts, the metal and additional co-ligands (used to complete the metal coordination sphere) are left to reside in one of the DNA grooves. Depending on the precise complex this can be either the major or the minor groove.[19,20] Structurally similar complexes have been reported to occupy different grooves, complicating prediction in design.

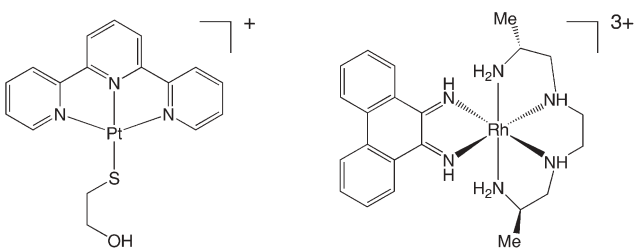

Fig. 15 Examples of metallo-intercalators: $[Pt(tpy)(SCH_2CH_2OH)]^+$ developed by Lippard[18] and $[Rh(phi)(Me_2trien)]^{3+}$ developed by Barton.[19]

This journal is © The Royal Society of Chemistry 2007

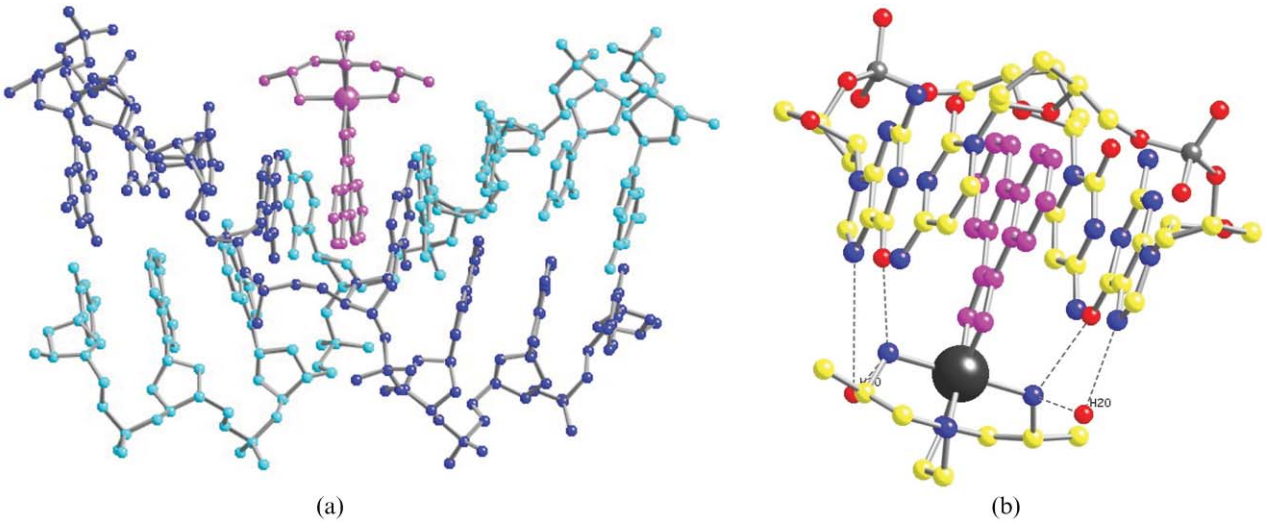

Fig. 16 (a) Binding of $[Rh(phi)(Me_2trien)]^{3+}$ (pink) to DNA by intercalation. (b) Close-up illustrating the π-stacking and H-bonding. (PDB ref. 454D†).[21]

Simple intercalators show only limited sequence preference, their π-surfaces tending to prefer either AT pairs or GC pairs and by their nature interacting with just two pairs. For octahedral metallo-intercalators, the presence of additional co-ligands located in a groove offers opportunities to introduce groups which could make specific contacts with bases above and below the intercalation site thus affording some sequence preference to the binding. An example of an intercalating metal complex is the rhodium phi complex of Barton for which the DNA-binding has been structurally characterised by X-ray crystallography.[21] Despite the intense activity in this area, this remains the only metallo-intercalator:DNA structure for which crystallographic data is available.[22]

The structure (Fig. 16) shows the aromatic ligand inserted between the base-pairs (GC and CG) and the metal and co-ligand placed in the DNA major groove. The amines of the trien co-ligand form hydrogen-bonds to the guanine O, with further water-mediated H-bonds to the guanine N7 also proposed.

Bulge recognition

The phi ligand is an appropriate size to slide in between and stack with the DNA base pairs. However, if the aromatic surface is extended then it will become too big and no longer be able to bind to a normal Watson–Crick duplex. Barton has used this to create larger agents (based on chrysi and phzi ligands; Fig. 17) that will not intercalate into normal DNA, but can insert when there is a base-pair mismatch (such as at a single nucleotide polymorphism) where the absence of Watson–Crick pairing will permit a larger intercalator to be accommodated. This approach has been used to develop cleavage agents based on rhodium and luminescent agents based on ruthenium which can be used for footprinting or fluorescence detection of mismatches.[23]

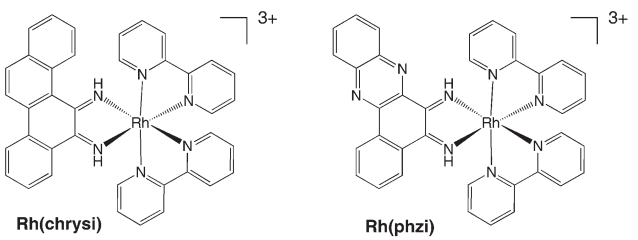

Fig. 17 Extended intercalators to recognise DNA bulges.

Bis-intercalators

Molecules with two intercalators linked together have also been prepared. If the linking chain is flexible and sufficiently long then the molecule can wrap and the two intercalators can intercalate into two sites on the same DNA duplex. If however a rigid spacer is used then this will not be possible and so the agent might target areas where two duplexes are in close proximity, as when DNA is packaged (*vide infra*). An example of this approach is the work of Lowe who linked two platinum terpyridyl intercalators (Fig. 18) or two organic acridine intercalators and used them to explore the spatial organization of DNA.[24]

An alternative bis-intercalator design which also employs platinum-terpyridyl units has been reported by Pikramenou.[25] In this design two of these intercalators are attached *via* the platinum onto aromatic thiolates attached to a neodinium aminocarboxylate complex (Fig. 19). The result is a hairpin-style complex in which the two intercalators are oriented to bind to the same DNA double-strand and with a separation of about 10.5Å. This is an ideal separation for bis-intercalation,

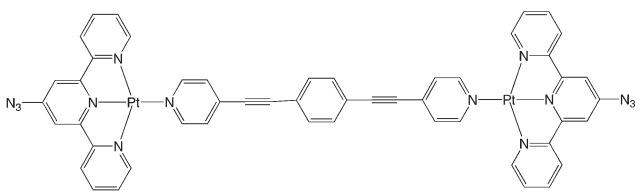

Fig. 18 Example of a rigid tetracationic bis-intercalator based on Pt(terpy) units.[24]

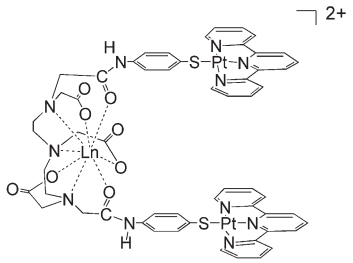

Fig. 19 Pikramenou's lanthanide-containing hairpin-shaped bis-intercalator.

allowing a gap to be left between intercalation units as required by the neighbour-group exclusion principle. The agent is assembled from its components in a programmed supramolecular assembly and binds strongly to calf-thymus DNA. The platinum-terpyridyl units not only act as the DNA recognition units but are able to harvest light energy which is used to drive the NIR emission of the neodinium.

Threading intercalators

Lincoln and Nordén have reported bis-intercalation for the complex $[\Delta,\Delta$-μ-(bidppz)-$(phen)_4Ru_2]^{4+}$ (Fig. 20).[26] The complex contains two linked $[Ru(phen)_2(dppz)]^{2+}$ motifs. That motif binds to DNA in a similar way to the rhodium phi complexes, inserting the dppz between the base pairs, and showing a dramatic enhancement of luminescence on binding.[19,27] This linked bis-intercalator is different from others studied, because the intercalating units are linked at the point through which they would usually insert into the DNA. Therefore to intercalate part of the molecule must thread through the DNA. This is indeed what happens, with an initial groove-bound non-luminescent state rearranging slowly (weeks at room temperature!) to form an intercalated emissive state in which one of the bulky $Ru(phen)_2$ moieties has been threaded through the DNA base stack.[26]

For related agents in which the dppz groups are linked through aliphatic chains, bis threading occurs with the dppz groups of the molecule intercalating from one groove and the bridging chain lying in the opposite groove.[28]

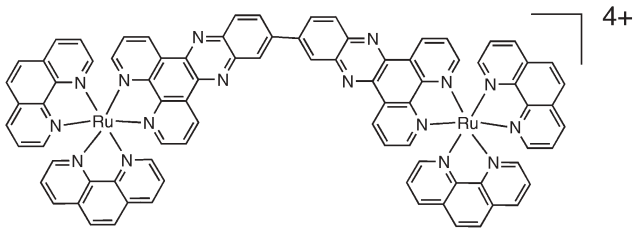

Fig. 20 The threading bis-intercalator $[\Delta,\Delta$-μ-(bidppz)-$(phen)_4Ru_2]^{4+}$.

Protein intercalation

While there are clear analogies between the ways that proteins and synthetic agents bind in the major groove and minor groove, this is not true for intercalation. The natural amino acids with available π-surfaces are phenylalanine, tyrosine, tryptophan and histidine. Of these three possess only single ring systems with the other (tryptophan) containing two fused rings (5 and 6 membered). Thus the surfaces are significantly smaller than those used in synthetic intercalators. For this reason an analogous intercalation is not usually observed in protein structures. Rather a different type of intercalation can be observed in which these smaller residues are partially inserted with the edges of the base pairs moving apart at the site of insertion.[1] This twists the base pairs away from their coplanar stacking and the DNA thus bends. This partial intercalation tends to be associated with regions of proteins that recognise the DNA minor groove. An example is the TATA box binding protein which contains two phenylalanine groups, one at either end of the DNA recognition region, which insert and bend the DNA (Fig. 21).[29] An even more dramatic example, where a minor groove binding HMG protein recognises platinated DNA (Fig. 23c) is discussed in the following section. While face-face π-interactions dominate the binding with synthetic intercalators, CH…π interactions appear important in protein partial intercalation. Indeed such binding is not restricted solely to the aromatic amino acids, with groups such as methionine also observed to partially intercalate.

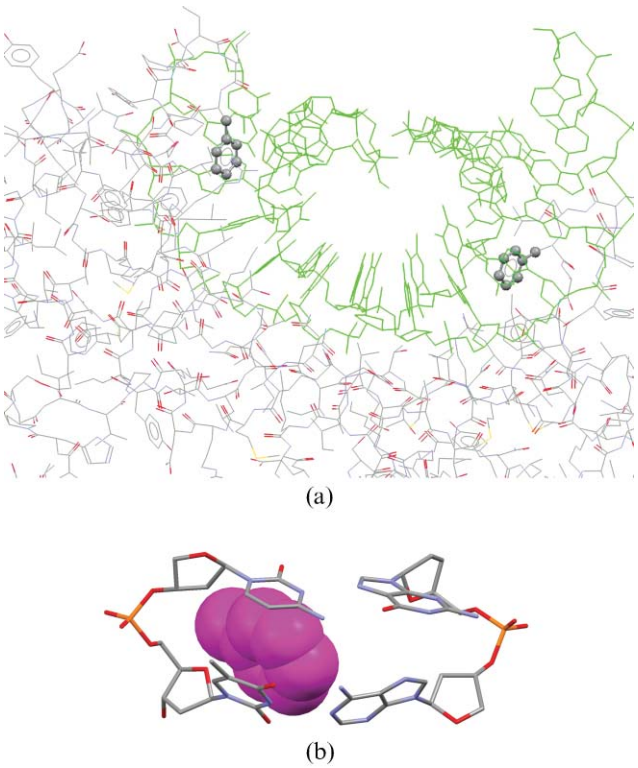

Fig. 21 (a) Interaction of the TATA box binding protein with DNA illustrating the two aryl rings which are intercalated at either end of the bent DNA. (b) Close-up showing the partial intercalation of a phenylalanine (pink) at one of these sites. (PDB ref. 1QNA†).[29]

Binding to the DNA bases

The other major site for DNA recognition used by clinical drugs is direct binding to the DNA bases, most commonly to the N7 atoms of G and A bases, located in the DNA major-groove. In the case of the nitrogen mustards (such as the clinical drug chlorambucil: Fig. 22) this involves alkylation of

 This journal is © The Royal Society of Chemistry 2007

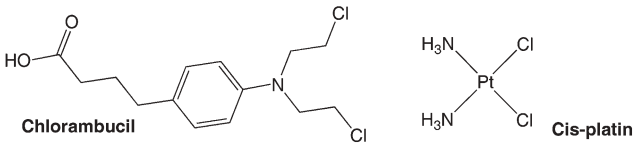

Fig. 22 Clinical 'alkylating' agents: Drugs that bind directly to the DNA bases.

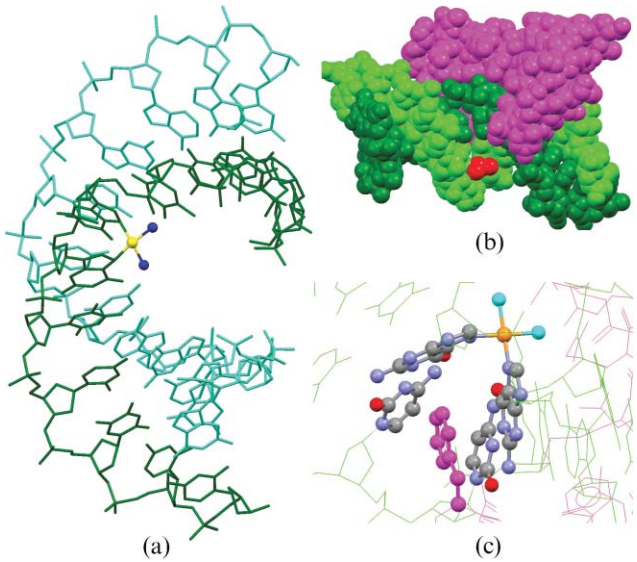

Fig. 23 (a) Structure of DNA with cis-platin bound illustrating the kink caused in the DNA structure. (PDB ref. 1AIO)[31] (b) Structure of HMG (pink) recognising DNA (green) bound to cis-platin (red). (PDB ref. 1CKT)[32] (c) Zoom view from the minor groove side, illustrating how a phenylalanine (pink) from the protein intercalates into the cavity formed between the DNA base-pairs forming face–face and face–edge π-interactions.

the nitrogen and as such is a covalent rather than non-covalent bond. Among the most successful of the clinical 'alkylating' agents used to treat cancers are cis-platin (Fig. 22) and its second generation derivatives oxaliplatin, nedaplatin and carboplatin.[30] Cis-platin binds to DNA with displacement of two chlorides and formation of two metal-coordination bonds to N7 of two adjacent purine DNA bases on the same strand. There is a strong preference for G over A with over 70% of the lesions being found at GG sites and around 20% at GA. The bifunctional binding causes a bend (kink) in the DNA of around 45 degrees, with the bend being towards the site of platination (Fig. 23a). This bent DNA structure is then recognized by nuclear HMG proteins which bind in the major groove and are believed to protect the lesion from DNA repair (Fig. 23b). A crystal structure of the DNA-binding domain of a HMG protein bound to platinated DNA reveals a phenylalanine residue from the protein inserting into the cavity created at the kink, forming a face–face π–π interaction with one platinated guanine and a face–edge (CH···π) interaction with the other (Fig. 23c).

Backbone binding

The final site that has traditionally been recognized as a potential DNA binding-site is the sugar-phosphate backbone. As a hard oxygen-rich polyanion surface, interaction with group I and II alkali metal cations is expected. Other than this there are few reports of synthetic agents that specifically target this area of the DNA. Protein interactions with the backbone are common, most usually by cationic residues (*e.g.* arginine) that can form hydrogen bonds with the phosphate oxygens, although these interactions are usually deployed alongside major and or minor groove recognition motifs.[1]

Farrell has recently explored the DNA binding of the trinuclear platinum(II) compound shown in Fig. 24. This agent is related to the polynuclear trans-platinum drug BBR3464 but has no reactive Pt–Cl groups and thus can bind to DNA only through hydrogen bonding and electrostatic interactions. Co-binding of minor-groove agents, such as Hoechst 33258, was shown to be cooperative, while intercalation by ethidium bromide was inhibited. Interaction in or around the minor groove was identified as a possible binding mode, and a very recent X-ray crystallographic study has revealed that the compound can form multiple hydrogen bonds with the phosphate oxygens. The complex can both track along the phosphate backbone, in a mode of binding termed "Backbone Tracking" or stretch across the minor groove making contacts with the phosphate backbones on either side. The formation of H-bonds from two cis amines to one phosphate oxygen is a repeated motif in the structure.[33]

Beyond B-DNA: Recognition of other DNA structures

While B-DNA is the predominant form of DNA in biology, this basic structure has higher levels of ordering and other structures are also possible. These offer alternative or additional sites for recognition. In cells DNA is packaged by being wrapped around protein units. For eukaryotic nuclear DNA, the first step of this packaging process is the winding of DNA twice around a histone (consisting of 8 protein units) to give a cylindrical structure known as a nucleosome, about 11 nm in diameter. This undergoes further levels of packaging, ultimately creating a highly packaged chromosome. Aside from this packaging there are a number of other structures which are potential recognition targets: The DNA at the ends of the strand are packed into telomere structures which contain a tetraplex DNA; DNA triplexes have been implicated

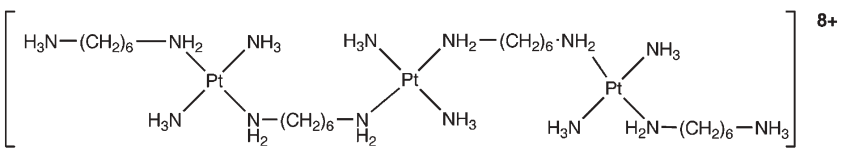

Fig. 24 Farrell's octacationic backbone binder.

in some DNA regions; DNA junctions can be present along the DNA and are formed transiently in replication and repair.

Recognising nucleosomal DNA

Dervan has demonstrated that his minor-groove polyamide DNA-binding agents can bind not only to free duplex DNA but that binding to the minor groove can also be achieved when the DNA has been wrapped onto the histone proteins to form the nucleosome. The DNA wrapping in the nucleosome places two sections of DNA adjacent to each other (Fig. 25). By linking two minor groove polyamide motifs, Dervan has demonstrated that this structure may be recognised by binding one motif to its appropriate sequence in the minor groove of one stretch of the DNA and the other motif to the sequence in the adjacent DNA duplex. This is shown in Fig. 25.

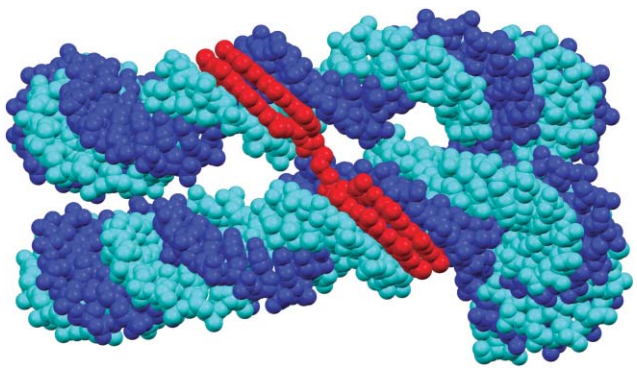

Fig. 25 View of the DNA in a nucleosome structure with a bis-minor groove binder recognising two adjacent minor grooves. The histone proteins about which the DNA wraps are excluded for clarity. (PDB ref. 1S32†).[34]

Recognition of triplex DNA structures

The ability to bind and stabilise DNA triplex structures is important both for enhancing the efficacy of oligonucleotide therapeutics and for designing probes that will allow us to probe the role of these structures in cells in more detail. Many intercalators and minor groove binders designed to bind to B-DNA can also bind to triplex DNA. Some intercalators have been demonstrated to have a strong preference for triplex over duplex DNA. Hélène was the first to show this with the cationic benzopyridoindole ligand BePI (Fig. 26).[35] He went on to use this observation and the known structures of triplexes to rationally design agents such as BfPQ and BQQ (Fig. 26) which show a strong preference for Hoogsteen triplexes over duplex DNA. The size of their π-surfaces means that they would have to insert into duplexes with their long axis perpendicular to the base pair. This would leave parts of the hydrophobic surface protruding and exposed to the aqueous solvent environment. By contrast, intercalation into triplex structures would allow all of the π-surface to engage in base-stacking, with little or none exposed to the solvent.[36] Intriguingly the BfPQ surfaces are not dissimilar to those used by Barton to recognise duplex DNA at base-pair mismatches and this highlights the complexity of creating structures which recognize only one of the many possible structures that could be present in genomic DNA.

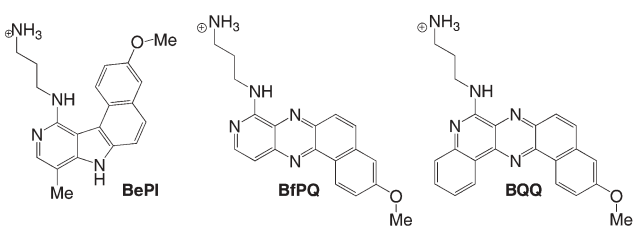

Fig. 26 Triplex-binding intercalators.

Recognition of quadruplex DNA structures

Quadruplex DNA structures, based on guanine quartets (Fig. 27) can be found in G rich sequences particularly in telomeres.[1] Telomeres are stretches of DNA with highly repetitive sequence that are located at the ends of chromosomes. These stretches of the DNA are not fully replicated and are consequently shortened in each replication cycle, leading to ageing. In cancer cells these stretches of DNA are elongated by the enzyme telomerase (a reverse transcriptase) thereby conferring immortality to the cancerous cell line. For these reasons agents that can bind specifically to quadruplex DNA have been a focus of recent attention. Quadruplexes can be formed by four discrete strands, by two strands which each double-back on themselves, or a single strand that folds back on itself three times. As such, a wide variety of different guanine quadruplex structures are possible which differ in the directionality of the strands and different numbers and connectivities of loop regions at their ends. Other quadruplex structures such as the i-Motif in which $(C.C^{+})_2$ units intercalate between each other are also possible.

As with triplexes, a variety of molecules that bind to duplex DNA can also bind to tetraplex DNA. In particular, some duplex intercalators (such as substituted anthraquinones) have been shown to also bind tetraplex DNA. However, in most structurally characterised cases, these intercalators bind not between base quartets in a typical intercalation

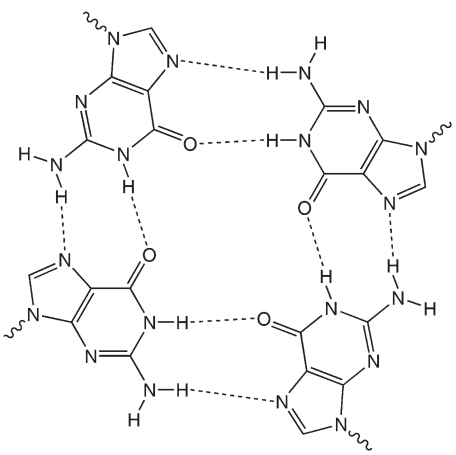

Fig. 27 Guanine tetramers, the basis for quadruplex formation.

This journal is © The Royal Society of Chemistry 2007

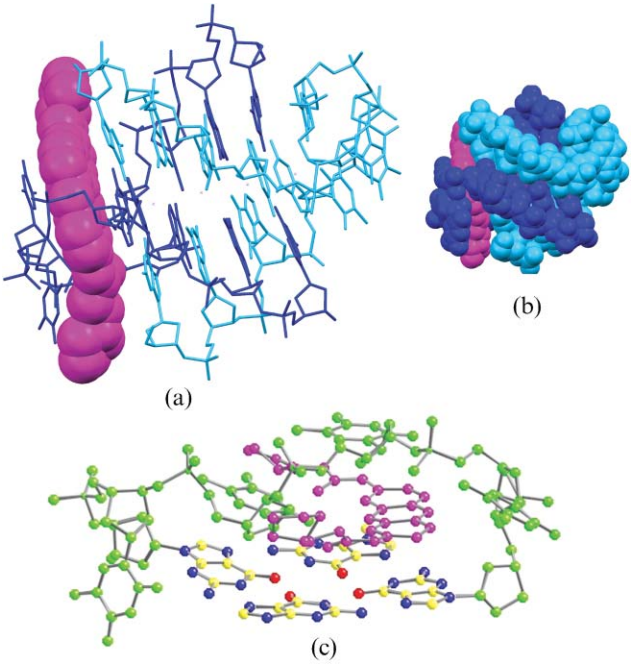

Fig. 28 (a) and (b) Two views of a crystallographically characterised complex between a dimeric antiparallel G-quadruplex from a telomeric DNA sequence with a di-substituted aminoalkylamido acridine compound (pink). (c) Close-up of the binding site. The loop region (which contains 3 C residues, one of which stacks with the intercalator) is shown in green and the intercalator in pink. (PDB ref. 1L1H†).[37]

fashion but rather at one end of the quartet stack stacking on the upper-most π-surface of the quadruplex structure (Figs. 28, 29), between the base quartet and the loop(s) that connect the strands of the quartet together.[37,38] Bases in the loop may be also π-stacked with the intercalator and the structure may thus have π-stacking above and below the drug as in conventional B-DNA intercalation. This binding mode (stacking on the π-surface of the uppermost bases) while not identical to intercalation, is clearly very closely related and is a binding mode through which classical B-DNA intercalators can bind to quadruplex DNA.

Recognition of DNA junction structures

DNA 4-way junctions (Holliday junctions) are formed when two duplexes come together and exchange strands. They are X-shaped branched structures, with four duplex 'arms' emerging from the branch point. In complex with proteins they are often in an open X-shaped form, while when free in solution they can fold up into a stacked conformation in which two arms are placed next to each other to give a more H-shaped structure. They are important in homologous genetic recombination, which is important in DNA repair and restart of failed replication forks and which also allows viral integration.[1]

Studies of agents that will specifically recognise such a 4-way junction structure are in their infancy and have focused on designing bifunctional agents that recognise two B-DNA arms of the structure. Lowe has employed his rigid bis-intercalator design. For one particular bis-acridine compound (Fig. 30) he

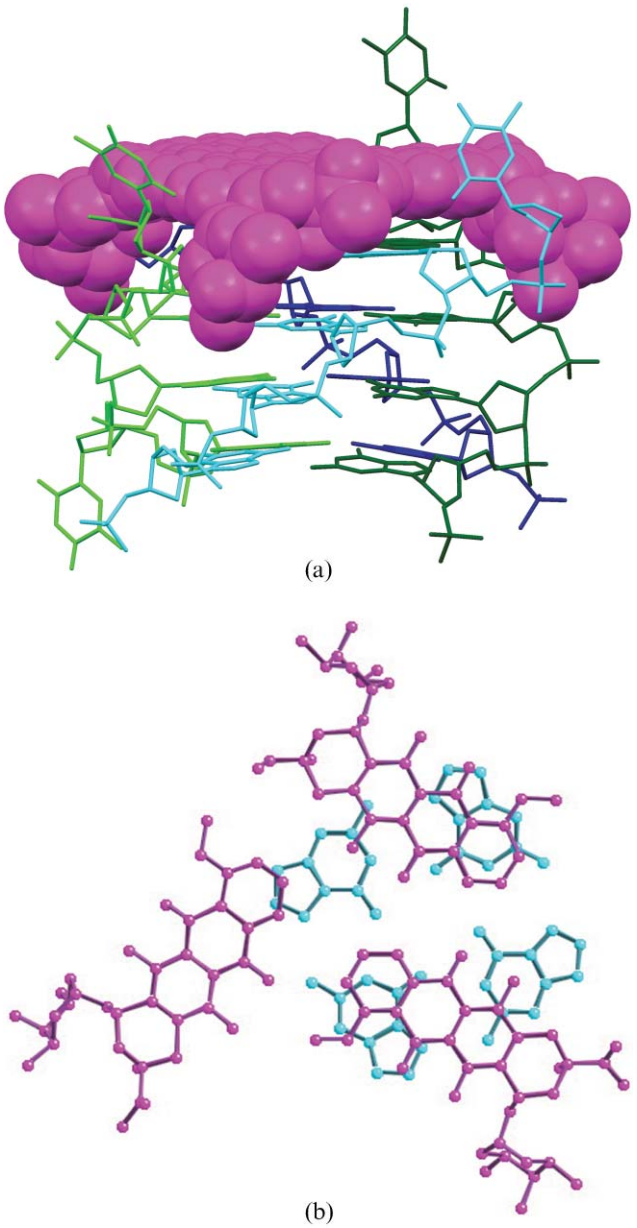

Fig. 29 (a) View of three dauncomycin molecules binding to the quartet π-surface at the end of a parallel quadruplex comprised of four separate strands. The daunomycin sugar units reside in three of the four grooves and form hydrogen bonds (some of which are water-mediated) and van der Waals contacts to the groove and the phosphate backbones. (b) Close-up showing how the three intercalators (pink) stack on the quartet surface (blue). (PDB ref. 1O0K†).[38]

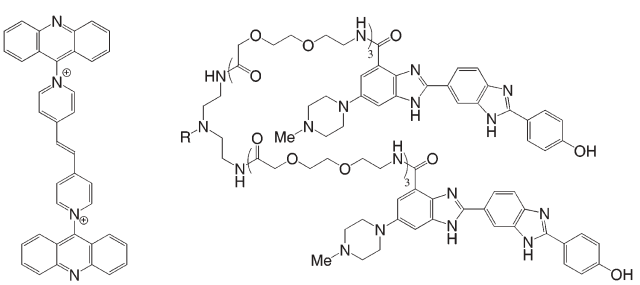

Fig. 30 Lowe's bis-acridines and Sasaki's bis-Hoechst agents use for junction recognition.

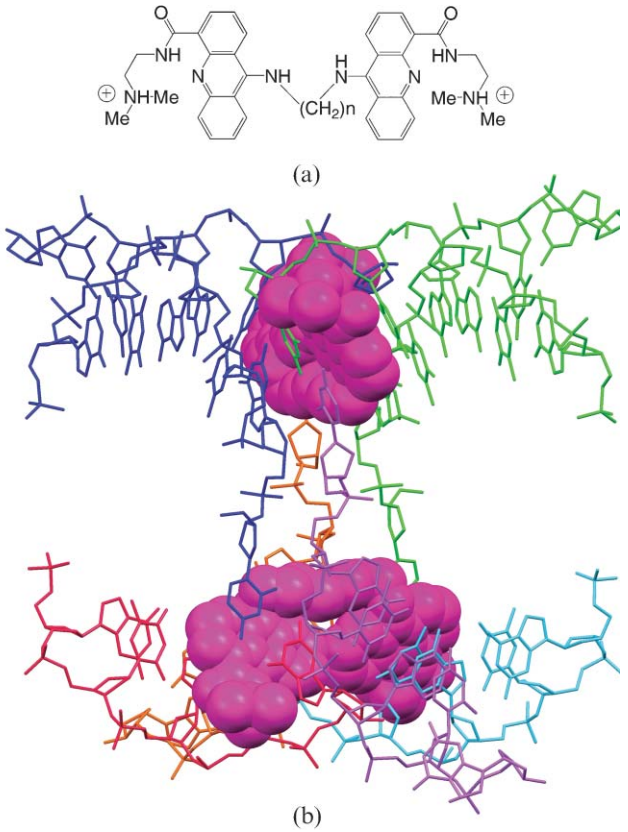

Fig. 31 (a) Cardin's bis-acridine intercalator. (b) Structure of two of these bis-intercalators (pink) with 8 oligonucleotides. (PDB ref. 1K2L†).[41]

noted increased access of a DNAase I to the phosphates at a CpT sequence some nine bases away from the centre of a 4-way (but not 3-way) junctions, implying that agents that target 4-way junctions by such an approach may indeed be feasible.[39] More recently Sasaki has developed a bis-minor groove binder based on two linked Hoechst 33258 units for recognition of DNA with two A3T3 motifs.[40] The flexible nature of the link means that the molecule does bind to duplex DNA possessing two A3T3 motifs at suitable separation, however it has also been shown to bind to a 3-way junction structure in which there are individual A3T3 motifs in two different arms.

No crystal structure of a 4-way junction with a bis intercalator (or bis-groove binder) has been reported. However, Cardin has crystallised a hexameric oligonucleotide with a bis-acridine intercalator and observed a complex structure in which eight oligonucleotides combine with two intercalators. Although the DNA structure is not that of a 4-way junction, the bis intercalators are located at positions where six oligonucleotide strands are coming together and four base pairs are placed in close proximity. The bis-intercalator is threaded into the structure such that each intercalating section sandwiches between two of the base-pairs. It therefore may be possible to thread bis-intercalators into DNA junctions at the junction point rather than simply designing agents to recognise adjacent B-DNA arms.

Recognition of a DNA 3-way junction by a supramolecular cylinder

My own research team has focused on DNA recognition by metallo-supramolecular agents.[42,43] The concept underpinning our work was our realisation that because supramolecular chemistry allows much larger structures to be designed than traditional covalent synthetic chemistry, it would allow us to create structures that more closely mimicked the dimensions of protein DNA recognition motifs. In particular we hoped to be able to move beyond intercalation and minor groove binding to create agents that were entirely synthetic yet would bind in the major groove. We designed a tetracationic triple-stranded cylinder of $\sim$2 nm length and $\sim$1 nm diameter (Fig. 32). These dimensions are similar to those of the alpha-helical DNA recognition unit of zinc fingers and the size is too great to fit into the minor groove of DNA. The agent does indeed appear to bind in the major groove and in addition unexpectedly and quite dramatically causes intramolecular DNA coiling, giving rise to small coils of DNA (Fig. 33).[42] As we had hoped, by stepping up to the size-scale of nature, quite remarkable and new effects were being observed.

To attempt to probe the relationship between the molecular-level groove recognition and the observed macromolecular coiling, we have explored both molecular dynamics simulations and the effects of synthetic modifications to the structure of the cylinder. The molecular dynamics simulations reproduce the coiling and imply that it is the precise size and shape of the

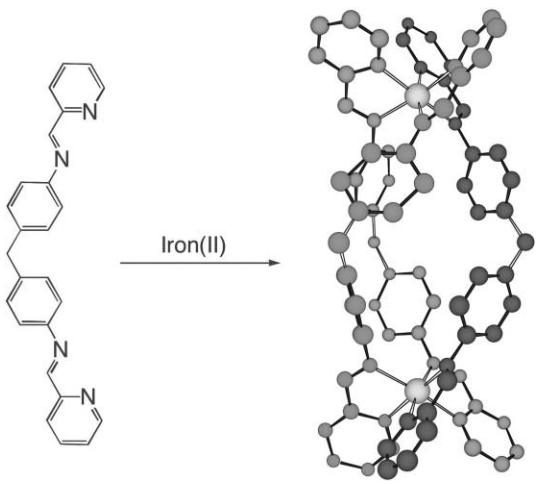

Fig. 32 Design of the tetracationic triple-stranded supramolecular cylinder.

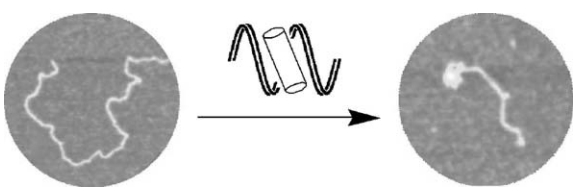

Fig. 33 Intramolecular DNA coiling induced by the supramolecular cylinder.

This journal is © The Royal Society of Chemistry 2007

cylinder (rather than its charge) that are important. Consistent with that, we have shown synthetically that increasing the cylinder dimensions (length and diameter) by around 10% led to a dramatic reduction in DNA coiling.[42]

Together with Miquel Coll, we were able to crystallise the cylinder with a palindromic hexanucleotide.[43] To our surprise it was not a simple duplex that crystallised but rather a DNA 3-way junction, in the heart of which resides the cylinder (Fig. 34). This is possible because in a palindromic sequence, Watson–Crick hydrogen bonding can be satisfied through a duplex structure, a 3-way junction or higher-order junction structures (*e.g.* 4WJ, 5WJ…). The cylinder has selected the 3-way junction from this dynamic combinatorial library of possibilities. The cylinder in the structure overlays almost perfectly with those in the crystal structures of the free cylinder, and the DNA at the junction similarly has the same structure as DNA in 3-way junctions crystallised with proteins. This confirms that neither structure is significantly perturbed by binding, rather that they are an almost perfectly matched pair, like a hand and a glove.

Aside from the electrostatic attraction, the aromatic surfaces of the cylinder play an important role in the binding, with the six phenyl rings at the heart of the cylinder face-face π-stacked onto the six DNA bases that are placed at the junction point as the three double-stranded arms come together. The two phenyl rings from one strand of the cylinder stack with the two bases of one of the strands of the DNA and this is shown in Fig. 35. One end of the cylinder protrudes into the region where the three minor grooves of the DNA join together (Fig. 34). The three pyridine rings at this end of the cylinder point down the three minor grooves and fit snugly there (Fig. 34b) as is seen for the aromatic rings of minor groove binding drugs such as berenil. The outer surface of the drug is a hydrocarbon framework, however the presence of the coordinated metal will make the pyridine and imine protons slightly acidic, with the effect expected to be greatest for the imine. Consistent with this, the imine hydrogen adjacent to the minor groove end of the cylinder forms a CH···X hydrogen bond with the adenine ring (N3) at the junction.

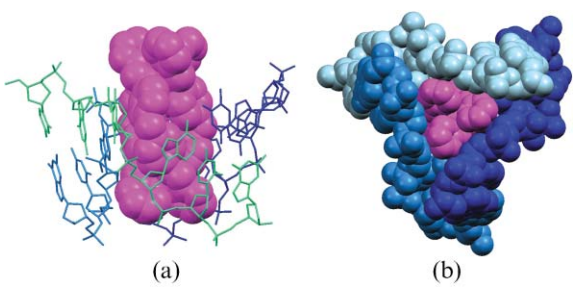

Fig. 34 (a) A DNA hexanucleotide 3-way junction with a supramolecular cylinder (pink) bound in the centre of the junction. (b) Space-filling view from the minor groove side. (PDB ref. 2ET0†).[43]

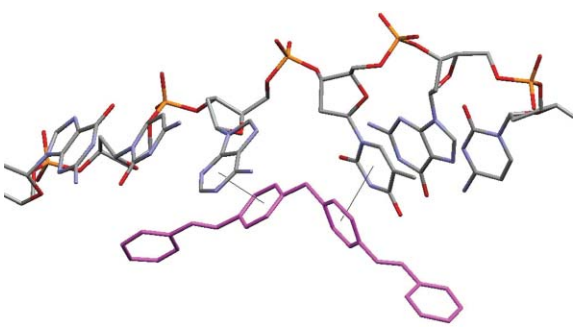

Fig. 35 Close up of the 3-way junction structure showing just one DNA strand and one ligand strand of the cylinder (pink). The stacking of the DNA bases at the junction point on the phenyl rings of the ligand strand is highlighted. (PDB ref. 2ET0†).[43]

While 3-way Y-shaped junction structures are much less well studied than 4-way X-shaped junctions they are particularly exciting because the DNA replication fork is a form of Y-shaped junction. 3-Way junctions have also been found to be present in diseases, such as Huntington's disease and myotonic dystrophy, and in certain viral genomes.

This binding mode, a perfect shape-fit in the heart of a DNA junction, is a quite different binding mode from those which, since the 1960s, have dominated research in the field of DNA recognition. The mode does have some resemblance to intercalation, in that an agent inserts through the DNA and stacks with the bases. Yet this agent is not a planar aromatic, does not have an extended flat π-surface and cannot intercalate into B-DNA structures. There does not seem to be a recognised analogue in the way that proteins or other biomolecules bind to DNA. The structure both opens up new ways of thinking about recognising DNA, and especially DNA junctions, and indicates that supramolecular chemistry may present a powerful addition to the design kit for DNA recognition.

The supramolecular cylinder is a triple-helicate[44] but differs from other helicate structures in its cylindrical shape and its rigidity; the strands are held together not only by metal–ligand interactions at each end but by sets of face–edge π-interactions in the centre of the structure. Lehn was the first to explore interaction of a helicate with DNA, describing the interaction with a cuprous supramolecular double-helicate, based on an oligo-bipyridine ligand with flexible alkyl-ether spacers.[45] The mode of binding does appear to be groove-binding although which groove could not be unambiguously established. Mononuclear copper(I) bis-phenanthroline complexes are known to bind to DNA and to effect oxidative DNA strand cleavage *via* action in the minor-groove.[46] The Lehn helicate is similarly able to cleave DNA. We have studied the DNA binding of (and cleavage by) a cuprous double-stranded cylinder based on our pyridylimine ligand design.[47] In contrast to the triple-stranded cylinder, this double-stranded cylinder does not bend the DNA. It also appears to bind to DNA in a slightly different orientation. Preliminary molecular dynamics simulations, coupled with experimental information about the orientation of binding, suggest that this cylinder could lie either in the major groove or just outside the minor groove.

Outlook

The need for probes that influence gene expression, and the desire for individual personalised medicines provide timely opportunities for the field of DNA recognition, and ones that the skills and techniques of supramolecular chemistry could

underpin. The key to both of these applications will be agents that act in a specific, rather than a generic way, on the DNA. Traditionally this has meant sequence-specific recognition. That has been primarily achieved by biomolecular motifs—modified peptides or oligonucleotides—with only Dervan's minor-groove binders providing real synthetic alternatives. A synthetic system that targets the major groove and is able to recognise its sequence remains a challenge, but one for which supramolecular chemistry may be able to provide sufficiently large structures and surfaces. However, recent studies have ably demonstrated that precise sequence-recognition may not be the only solution to this problem; rather, recognition of a particular unusual DNA structure may be a powerful alternative. In this the concepts of supramolecular chemistry, where shape, fit and orientation play such an important role, has much to offer.

The complexity will lie in selecting one DNA structure-type while avoiding binding to simple duplex DNA and the plethora of other possible DNA structures. For binders to triplex and quadruplex DNA, recognising the different grooves, as well as the different π-surfaces, is likely to be important. For junctions, the unanticipated discovery that a nano-sized cylinder can bind in the heart of the junction will change the way we think about recognising such structures. Farrell's phosphate clamps seem likely to attract people to recognition of the DNA backbone along with other regions (perhaps in concert as proteins so often do). This exciting period for the field is further fuelled by the results emerging from structural biology which are throwing up new structures as challenges for DNA-recognition chemists. For example, a structure of a DNA containing a B–Z junction has recently been reported, in which the DNA twists from a right-handed B-DNA helix through to a section of left-handed Z-DNA.[48] What better, complex new target for recognition could a supramolecular chemist want?

The initial phase of supramolecular chemistry has been one of aesthetics, where basic principles have been established and many extraordinary and beautiful structures have been described. It has also been a phase in which the chemistry has been inspired by observations in biology. During this phase the field has moved from being an exotic and esoteric topic to the point where its ideas and concepts are now infused throughout the mainstream of chemistry and nanoscience. This is a remarkable tribute to the vision and enthusiasm of its founding fathers and practitioners (particularly its inspirational champion Jean-Marie Lehn who shared the Nobel Prize 20 years ago). The challenge now is to move from the aesthetic phase into a phase of genuine and important applications that affect the quality of human life. DNA recognition is an area where this challenge can be addressed, where biology can now be inspired by observations in (supramolecular) chemistry and genuine benefits to society may accrue.

Acknowledgements

I am grateful to Professor Nick Farrell (Virginia Commonwealth University) for making available a copy of his manuscript on backbone-binding polynuclear platinum compounds ahead of publication.

References

1 See for example: J. M. Berg, J. L. Tymoczko and L. Stryer, *Biochemistry*, Freeman, New York, 5th edn, 2002; C. Branden and J. Tooze, *Introduction to Protein Structure*, Garland, New York, 2nd edn, 1999; S. Neidle, *Nucleic Acid Structure and Function*, OUP, Oxford, 2002.
2 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737.
3 A. Sarai and H. Kono, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **34**, 379–98 and refs therein.
4 D. Jantz, B. T. Amann, G. J. Gatto and J. M. Berg, *Chem. Rev.*, 2004, **104**, 789–799 and refs therein.
5 M. Elrod-Erickson, T. E. Benson and C. O. Pabo, *Structure*, 1998, **6**, 451–464.
6 T. Da Ros, G. Spalluto, M. Prato, T. Saison-Behmoaras, A. Boutorine and B. Cacciari, *Curr. Med. Chem.*, 2005, **12**, 71–88; N. T. Thuong and C. Helene, *Angew. Chem., Int. Ed. Engl.*, 1993, **32**, 666–690.
7 PDB 149D:† I. Radhakrishnan and D. J. Patel, *Structure*, 1994, **2**, 17; PDB 134D:† I. Radhakrishnan and D. J. Patel, *Structure*, 1993, **1**, 135–152.
8 P. E. Nielsen, *Curr. Med. Chem.*, 2001, **8**, 545–50.
9 L. Betts, J. A. Josey, J. M. Veal and S. R. Jordan, *Science*, 1995, **270**, 1838–41.
10 P. G. Baraldi, A. Bovero, F. Fruttarolo, D. Preti, M. A. Tabrizi, M. G. Pavani and R. Romagnoli, *Med. Res. Rev.*, 2004, **24**, 475–528.
11 T. A. Larsen, D. S. Goodsell, D. Cascio, K. Grzeskowiak and R. E. Dickerson, *J. Biomol. Struct. Dyn.*, 1989, **7**, 477–91.
12 D. G. Brown, M. R. Sanderson, E. Garman and S. Neidle, *J. Mol. Biol.*, 1992, **226**, 481–490.
13 P. B. Dervan, *Bioorg. Med. Chem.*, 2001, **9**, 2215–2235.
14 C. L. Kielkopf, R. E. Bremer, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan and D. C. Rees, *J. Mol. Biol.*, 2000, **295**, 557–567.
15 L. S. Lerner, *J. Mol. Biol.*, 1961, **3**, 18.
16 R. Martínez and L. Chacón-García, *Curr. Med. Chem.*, 2005, **12**, 127–151.
17 C. A. Frederick, L. D. Williams, G. Ughetto, G. A. van der Marel, J. H. van Boom, A. Rich and A. H. Wang, *Biochemistry*, 1990, **29**, 2538–2549.
18 K. W. Jennette, S. J. Lippard, G. A. Vassiliades and W. R. Bauer, *Proc. Natl. Acad. Sci. U. S. A.*, 1974, **71**, 3839–3843; M. Howe-Grant and S. J. Lippard, *Biochemistry*, 1979, **18**, 5762–5769; M. Howe-Grant, K. C. Wu, W. R. Bauer and S. J. Lippard, *Biochemistry*, 1976, **15**, 4339–4346.
19 K. E. Erkkila, D. T. Odom and J. K. Barton, *Chem. Rev.*, 1999, **99**, 2777–2295.
20 I. Greguric, J. R. Aldrich-Wright and J. G. Grant Collins, *J. Am. Chem. Soc.*, 1997, **119**, 3621–3622; J. G. Collins, A.-D. Sleeman, J. R. Aldrich-Wright, I. Greguric and T. W. Hambley, *Inorg. Chem.*, 1998, **37**, 3133–3141.
21 K. E. Erkkila, B. P. Hudson, J. K. Barton and D. C. Rees, *Nat. Struct. Biol.*, 2000, **7**, 117–121.
22 An early structure of a platinum–terpyridine complex with a dinucleotide was described in A. H. J. Wang, J. Nathans, G. Van der Marel, J. H. Van Boom and A. Rich, *Nature*, 1978, **276**, 471–474. Unfortunately no data is available.
23 J. R. Hart, M. D. Johnson and J. K. Barton, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 14040–14044; E. Ruba, J. R. Hart and J. K. Barton, *Inorg. Chem.*, 2004, **43**, 4570–4578.
24 G. Lowe, A. S. Droz, J. J. Park and G. W. Weaver, *Bioorg. Chem.*, 1999, **27**, 477–486; N. K. Annan, P. R. Cook, S. T. Mullins and G. Lowe, *Nucleic Acids Res.*, 1992, **20**, 983–990.
25 P. B. Glover, P. R. Ashton, L. J. Childs, A. Rodger, M. Kercher, R. M. Williams, L. De Cola and Z. Pikramenou, *J. Am. Chem. Soc.*, 2003, **125**, 9918–9919.
26 F. Westerlund, L. M. Wilhelmsson, B. Nordén and P. Lincoln, *J. Phys. Chem. B*, 2005, **109**, 21140–21144 and refs therein.
27 C. Hiort, P. Lincoln and B. Nordén, *J. Am. Chem. Soc.*, 1993, **115**, 3448–3454.
28 B. Önfelt, P. Lincoln and B. Nordén, *J. Am. Chem. Soc.*, 2001, **123**, 3630–3637.
29 G. A. Patikoglou, J. L. Kim, L. Sun, S.-H. Yang, T. Kodadek and S. K. Burley, *Genes Dev.*, 1999, **13**, 3217.

 This journal is © The Royal Society of Chemistry 2007

30 *Cisplatin, Chemistry and Biochemistry of a Leading Anti-Cancer Drug*, ed. B. Lippert, Wiley-VCH, Weinheim, Germany, 1999; J. Reedijk, *Chem. Commun.*, 1996, 801–806; Z. J. Guo and P. J. Sadler, *Adv. Inorg. Chem.*, 2000, **49**, 183–306 and refs therein.
31 P. M. Takahara, A. C. Rosenzweig, C. A. Frederick and S. J. Lippard, *Nature*, 1995, **377**, 649–652.
32 U. M. Ohndorf, M. A. Rould, Q. He, C. O. Pabo and S. J. Lippard, *Nature*, 1999, **399**, 708–712.
33 A. Harris, Y. Qu and N. Farrell, *Inorg. Chem.*, 2005, **44**, 1196–1198; S. Komeda, T. Moulaei, K. K. Woods, M. Chikuma, N. P. Farrell and L. D. Williams, *J. Am. Chem. Soc.*, in press.
34 R. S. Edayathumangalam, P. Weyermann, J. M. Gottesfeld, P. B. Dervan and K. Luger, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 6864–6869; R. K. Suto, R. S. Edayathumangalam, C. L. White, C. Melander, J. M. Gottesfeld, P. B. Dervan and K. Luger, *J. Mol. Biol.*, 2003, **326**, 371–380.
35 J. L. Mergny, G. Duval-Valentin, C. H. Nguyen, L. Perrouault, B. Faucon, M. Rougée, T. Montenay-Garestier, E. Bisagni and C. Hélène, *Science*, 1992, **256**, 1681.
36 C. Escudé, C. H. Nguyen, S. Kukreti, Y. Janin, J.-S. Sun, E. Bisagni, T. Garestier and C. Hélène, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 3591–3596.
37 S. M. Haider, G. N. Parkinson and S. Neidle, *J. Mol. Biol.*, 2003, **326**, 117–25.
38 G. R. Clark, P. D. Pytel, C. J. Squire and S. Neidle, *J. Am. Chem. Soc.*, 2003, **125**, 4066–4067.
39 M. L. Carpenter, G. Lowe and P. R. Cook, *Nucleic Acids Res.*, 1996, **24**, 1594–1601.
40 M. Tanada, S. Tsujita and S. Sasaki, *J. Org. Chem.*, 2006, **71**, 125–134.
41 S. C. M. Teixeira, J. H. Thorpe, A. K. Todd, H. R. Powell, A. Adams, L. P. G. Wakelin, W. A. Denny and C. J. Cardin, *J. Mol. Biol.*, 2002, **323**, 167–171.
42 M. J. Hannon, V. Moreno, M. J. Prieto, E. Moldrheim, E. Sletten, I. Meistermann, C. J. Isaac, K. J. Sanders and A. Rodger, *Angew. Chem., Int. Ed.*, 2001, **40**, 880–884; I. Meistermann, V. Moreno, M. J. Prieto, E. Moldrheim, E. Sletten, S. Khalid, M. Rodger, J. Peberdy, C. J. Isaac, A. Rodger and M. J. Hannon, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 5069–5074; C. Uerpmann, J. Malina, M. Pascu, G. J. Clarkson, V. Moreno, A. Rodger, A. Grandas and M. J. Hannon, *Chem.–Eur. J.*, 2005, **11**, 1750–6; S. Khalid, M. J. Hannon, A. Rodger and P. M. Rodger, *Chem.–Eur. J.*, 2006, **12**, 3493–3506.
43 A. Oleski, A. G. Blanco, R. Boer, I. Usón, J. Aymami, A. Rodger, M. J. Hannon and M. Coll, *Angew. Chem., Int. Ed.*, 2006, **45**, 1227–1231.
44 J.-M. Lehn, *Supramolecular Chemistry: Concepts and Perspectives*, VCH, Weinheim, 1995; C. Piguet, G. Bernardinelli and G. Hopfgartner, *Chem. Rev.*, 1997, **97**, 2005–2062; M. Albrecht, *Chem. Rev.*, 2001, **101**, 3457–3497; M. J. Hannon and L. J. Childs, *Supramol. Chem.*, 2004, **16**, 7–22 and refs therein.
45 B. Schoentjes and J.-M. Lehn, *Helv. Chim. Acta*, 1995, **78**, 1–12.
46 D. S. Sigman, T. W. Bruice, A. Mazumder and C. L. Sutton, *Acc. Chem. Res.*, 1993, **26**, 98–104; M. Pitie, C. Boldron, H. Gornitzka, C. Hemmert, B. Donnadieu and B. Meunier, *Eur. J. Inorg. Chem.*, 2003, 528–540 and refs therein.
47 L. J. Childs, J. Malina, B. E. Rolfsnes, M. Pascu, M. J. Prieto, M. J. Broome, P. M. Rodger, E. Sletten, V. Moreno, A. Rodger and M. J. Hannon, *Chem.–Eur. J.*, 2006, **12**, 4919–4927.
48 S. C. Ha, K. Lowenhaupt, A. Rich, Y. G. Kim and K. K. Kim, *Nature*, 2005, **437**, 1183–1186.