Biochem. J. (1987) 243, 1–13 (Printed in Great Britain)

REVIEW ARTICLE

DNA structure and perturbation by drug binding

Stephen NEIDLE, Laurence H. PEARL and Jane V. SKELLY
Cancer Research Campaign Biomolecular Structure Unit, The Institute of Cancer Research, Clifton Avenue, Sutton, Surrey SM2 5PX, U.K.

Introduction

The selectivity and specificity of low-M_r non-covalently binding ligands and drugs (as well as proteins) for particular DNA sequences is increasingly seen to be related to the sequence-dependent properties of DNA itself. Advances in our knowledge of these details at the three-dimensional structural level have arisen in large part from the application of X-ray crystallography, n.m.r. and molecular modelling techniques to defined-sequence oligonucleotides. Studies using chemical and enzymic probes of DNA restriction fragments, with and without bound drugs, have revealed further aspects of specificity, for example, that are more apparent in longer sequences of DNA. This Review discusses these aspects of both DNA itself, and of selected DNA–drug complexes.

DNA structures

The classic picture of DNA structure as that of a precisely repetitive polynucleotide in the standard B-form conformation has for some years been known to be too simplistic a view. The extensive fibre diffraction studies of Arnott and his colleagues have demonstrated the high polymorphism inherent in double-helical random-sequence DNA when subjected to differing environmental effects (Arnott *et al.*, 1982, 1983*a*). These result in several major forms (A,B,C,D) together with a number of sub-classes (Mahendrasingam *et al.*, 1986). Defined-sequence polynucleotides show similar effects, although with these it is common for the structure and polymorphism shown by one sequence to be inaccessible to another. A noteworthy example of this is poly(dA)·poly(dT), whose fibre diffraction pattern has been interpreted in terms of a 'heteronomous' structure with quite distinct nucleotide conformations and sugar pucker for the two strands of the duplex (Arnott *et al.*, 1983*b*). The poly(dA) strand has an A-type conformation, and the poly(dT) has a B-type. Table 1 gives the major geometric features of a number of these polynucleotide helices; especially noteworthy are the marked differences in major and minor groove dimensions for the polymorphs. The question of groove size is of fundamental importance in relation to the interactions of both small and large molecules with DNAs, since internal DNA binding sites are necessarily accessed via either major or minor grooves, and thus will be addressed in detail in subsequent sections of this Review.

The study of short lengths of DNA by single-crystal X-ray methods has provided a complementary perspective to the polynucleotide results of fibre diffraction. It has moreover provided a greater level of detail about individual sequences, and the effects of base type and sequences. The discovery of left-handed Z-DNA as the high-salt form of the hexanucleotide $d(CpG)_3$ in the crystal (Wang *et al.*, 1979) suggested that alternating pyrimidine–purine sequences [and sometimes imperfectly alternating ones as well (Wang *et al.*, 1985)] can, under appropriate environmental conditions, undergo an extreme polymorphic right → left-handed helix transition. This has been confirmed to occur in solution by a number of techniques, including n.m.r. (Singleton, 1986), although the precise physiological role(s) played by such sequences in genomic DNA still remain relatively obscure. A number of crystal structures of oligonucleotides showing right-handed helical form have now been determined, although as yet none is longer than 12 nucleotides (just over one complete turn of DNA double helix). This dodecanucleotide is of the sequence d(CGCGAATTCGCG), and has an overall B-DNA structure, with the bases being approximately perpendicular to the global helix axis (Wing *et al.*, 1980; Dickerson & Drew, 1981*a*,*b*; Dickerson, 1983). In

Table 1. Groove dimensions and helical parameters in nucleic acids

	Minor groove		Major groove		Rotation per nucleotide (°)	Axial translation per nucleotide (nm)
	Depth (nm)§	Width (nm)	Depth (nm)	Width (nm)		
A helix*	0.28	11.0	1.35	0.27	32.7	0.256
B helix*	0.75	0.57	0.85	1.17	30.0	0.326
Dodecamer†						
at central AATT	–	0.32	–	1.27		
at end of ATT	–	0.70	–	1.06		
d(GGATGGGAG)‡		0.83–0.97		0.93		

* In classic monopolynucleotide duplex helices (Arnott, 1981).
† Dickerson *et al.* (1985). Groove depths are not given in this study.
‡ McCall *et al.* (1986).
§ 1 nm = 10 Å.

contrast with a fibrous polynucleotide having a mononucleotide or dinucleotide repeat and a unique helix axis, it is not possible to describe this structure exactly in such straightforward terms. Each residue has distinct conformational characteristics [although all residues are classified in the same overall B-DNA family (Neidle & Berman, 1983)], and local rather than global helical parameters must be used to describe fully the variation in helical properties. Variations in the propellor twist of bases in base pairs are observed to be sequence-dependent and correlate with helical twist, which varies over a range of some 12° in the crystal structures (Dickerson, 1983; Dickerson *et al.*, 1985). A consequence of these sequence-dependent differences is that the major and minor grooves vary in dimensions along the extent of the dodecamer, with the minor groove being narrow in the central AATT section and wider towards the ends of the helix. The major groove widths change in a compensatory manner (Table 1). Related to these observations is the network of ordered water molecules found concentrated in the A+T-rich region minor groove of the structure (Drew & Dickerson, 1981), and termed the 'spine of hydration'. This sequence-dependent hydration has been implicated as a factor in the B-form nature of this structure; it is also plausible that the hydrogen-bonding of water molecules to bases in the region is a factor in the narrowing of the minor groove, and even in the propellor twists of base pairs. Detailed X-ray crystallographic data on at least several other B-DNA oligomers is required for these issues to be fully resolved. Nevertheless, the 'Calladine rules' that have been proposed (Calladine, 1982) to account for the base pair propeller twists in terms of relief of steric clashes between, in particular, purines, provides overall good correspondence with the features observed in this crystal structure. These simple rules are based on the differences in stacking between different dinucleotide steps, and detail the consequent bistable behaviour of some steps.

By contrast with the sole single-crystal representative of the B-family, a number of A-type oligonucleotide structures have now been reported. These vary considerably in sequence type; for example, d(GGTATACC) (Shakked *et al.*, 1983), d(GGCCGGCC) (Wang *et al.*, 1982) and d(GGGGCCCC) (McCall *et al.*, 1985). All are characterized by broad shallow minor grooves and deep narrow major grooves, as determined earlier in A-form polynucleotide fibres. Sequence-dependency is a notable feature found in these structures; the greater number of A-type oligonucleotides compared with B-type is now enabling firmly-based generalizations to be proposed concerning the structural and helical properties of individual dinucleotide steps in them. Nevertheless, we are still far from an understanding of longer-range sequence dependency, which is strikingly illustrated in the case of A+T-rich tracts which have been found to undergo bending in solution (Koo *et al.*, 1986) that has a dependency on the helical screw repeat, and may also be a reflection of the atypical polynucleotide features proposed for 'heteronomous' poly(dA)·poly(dT) (Arnott *et al.*, 1983*b*). Other anomalous features of this polynucleotide that have been previously noted include its inability to produce nucleosome reconstitution (Simpson & Kunzler, 1979) and its non-B-DNA helical repeat of 10.1 base pairs per turn (Rhodes & Klug, 1981; Peck & Wang, 1981). The very recent crystal structure analysis (McCall *et al.*, 1986) of the DNA sequence corresponding to the recognition sequence of the eukaryotic transcription factor TFIIIA (Rhodes & Klug, 1986) has shown features that appear to be distinct yet again from those observed in other oligonucleotides, for example in the dimensions of the major and minor groove widths (Table 1). Overall, this sequence most closely resembles the A′-form of double-helical RNA previously observed in polynucleotide fibres, with a helical repeat of 12 base pairs per turn, as had been suggested (Rhodes & Klug, 1986) on the basis of an analysis of the periodicity of nuclease digestion data.

DNA structures in solution. The increasing body of X-ray crystallographic data on oligonucleotides is in part complemented by data on solution conformations. N.m.r. methods can monitor base-pair imino protons (Patel *et al.*, 1982*a*) in a variety of environmental conditions, the dodecamer d(CGCGAATTCGCG) being especially well studied in this regard (Patel *et al.*, 1982*b*). Overall, the pattern of chemical shifts of both protons and phosphorus resonances has been interpreted to be consistent with the sequence-dependent features of the X-ray structure with features such as the base-stacking and propeller twisting being unambiguously confirmed by nuclear Overhauser measurements (Patel *et al.*, 1983). Two-dimensional n.m.r. techniques are increasingly able to provide full assignments of proton spectra, and it is then relatively straightforward to differentiate between different helix types on the basis of distance criteria, as well as details of sugar pucker geometries and glycosidic angles. Thus, the octamer d(GGAATTCC) has been found to exist as a B-type right-handed helix (Patel *et al.*, 1986), as has the decamer d(GCATTAATGC) (Chazin *et al.*, 1986). The latter has considerable base-pair propeller twisting, in accord with the crystallographic data on the decamer d(CGCGAATTCGCG) (Dickerson & Drew, 1981*a*). Complete three-dimensional structure determination by n.m.r. has been performed to date on very few oligonucleotides (Clore & Gronenborn, 1985). The hexamer duplex of d(CGTACG) has been analysed to be of B-DNA type by both restrained least-squares (Clore *et al.*, 1985) and restrained molecular dynamics (Nilsson *et al.*, 1986) methods; as in the dodecamer B-helix crystal structure, significant sequence-dependent features of the 'averaged' dynamic structure are observed. It is probably premature to compare these features as found in crystal and solution, not least because the data base in both instances is so scanty. There are nevertheless some indications (Clore & Gronenborn, 1985) that the application of the crystallographically-based steric clash rules to predictions of local sequence variation (Calladine, 1982) may not always result in consistency between the two approaches. The apparent preference for B-type structures in solution appears to be in striking contrast to their infrequency in the crystal, and may suggest that the A form is preferred by crystal environment as much as by sequence. Theoretical studies of oligonucleotide sequences have frequently used molecular mechanics methods, which tend to emphasize the importance of steric hindrance and resultant local group movements that counter these. It is thus unsurprising that results from such analyses have supported experimental findings of base tilt and roll in B-form helices (Kollman *et al.*, 1981), and of significant local structure variation (Rao & Kollman, 1985). The conclusion of C-3′ *endo* sugar pucker for the thymine

strand in poly(dA)·poly(dT) and C-2′ *endo* pucker for the adenine strand (Rao & Kollman, 1985), is at variance with the X-ray fibre diffraction analysis (Arnott *et al.*, 1983*b*); the discrepancy may be due in part to the theoretical difficulty of incorporating solvent and counter-ion effects, as well as deficiencies in the united-atom approach to the calculation of molecular energies and conformations. Molecular dynamics methods, although computationally more costly, offer significant advantages in terms of the range of conformational space that can be explored. A molecular dynamics simulation of the short sequence d(CGCGA) in duplex form (Singh *et al.*, 1985; Seibel *et al.*, 1985) with the explicit inclusion of hydrated counter-ions, has found sequence-dependent flexibility in base pairs and sugar pucker that, in the former case, appears to be only partly in accord with the X-ray data. Again, comparisons may be premature and await more extensive data. An alternative approach to the understanding of sequence-dependent oligomer properties involves consideration of the electrostatic potential and steric accessibility (to solvent) at particular loci (Lavery & Pullman, 1981), and has importance for the study of sequence-specific DNA–ligand interactions (see below).

Sequence variation in 'biological' DNA. The various approaches to DNA structure outlined above have, in the main, focused attention on the local structure in short oligonucleotide sequences. There is an inherent problem with these, of end-effects, that could mask or even counter-balance true sequence-dependent features.

Probes of much longer (100–200 base pair) sequences that may have biological function are therefore highly significant tests of the relevance of the findings described above. Both enzymic (Drew, 1984) and small-molecule (Dervan, 1986) DNA-cleaving reagents have been employed. In the case of DNAase I, the cutting of the double-stranded nucleic acid backbone is most probably from the minor groove direction, as evidenced from both the cleavage data itself, and from the crystal structure of the enzyme (Suck & Oefner, 1986). The cleavage patterns of a 160 bp sequence of the *tyrT* promoter have been studied in detail (Drew & Travers, 1984). Base sequence in itself is less important as a cause of the observed sequence-dependent cleavage patterns than are variations in nucleic acid geometry. These dictate the accessibility of the two strands to each other through the minor groove, and thus render to the enzyme a measure of structure specificity. Long runs of A+T [especially of oligo(dA)·oligo(dT)] and G+C are quite resistant to DNAase I cleavage. The single-crystal structures of oligonucleotides (especially of the B-DNA dodecamer) have provided, in so far as their restricted number allows, an explanation for these findings in terms of the likely narrowness of the minor groove in A+T regions (Calladine, 1982; Dickerson, 1983; Shakked *et al.*, 1983), compared with the markedly wider minor groove in G+C regions. Thus, DNAase I has optimal activity with a minor groove having an intermediate width. The more complex effects on nuclease cutting as a consequence of differing flanking sequences (Drew & Travers, 1985) appear at present to be dependent on both sugar-phosphate (i.e. groove dimensionality) and more specific base hydrogen-bonding recognition. Clearly, crystal structures of more (and longer) sequences, in conjunction with their solution n.m.r. analyses, will do much to shed light on this much more complex problem. The detailed study of oligomers in the 15–25 base range will also be relevant to the understanding of the functioning of DNA regulatory sequences. For example, the binding of SV40 large T antigen to origin region 1 in SV40 produces promoter repression, and the essential 17 bp binding site is apparently required to be in a bent conformation (Ryder *et al.*, 1986) in order for recognition to take place.

DNA intercalation complexes

The general features of the interaction between structurally simple planar drug molecules (typified by the acridines) and double-helical DNA were set out by Lerman (1961), with base pairs separating from 0.34 nm to 0.68 nm so as to accommodate a planar molecule sandwiched equi-spacially between them. The concomitant unwinding in DNA has been the subject of much subsequent study (summarized in Gale *et al.*, 1981), with the unwinding produced in covalently closed-circular DNA being used as a diagnostic test for intercalation. The unwinding produced by an individual intercalant molecule may be quantified. Unwinding angles are relative to that of 26° for ethidium; mono-intercalators with a single chromophore have values ranging from that of 11° for daunomycin to 26° for actinomycin D, whilst bis-intercalators with two chromophores produce approximately double this degree of unwinding. The differing effects appear to be dependent on the nature of the drug, although extent of unwinding does not correlate with other measurements of the interaction, such as association constants, extent of stabilization of the double helix to thermal denaturation, or to the nucleic acid synthesis-inhibition, cytotoxic and other biological effects shown by many intercalators. Unwinding as measured in 'random-sequence' DNA is a bulk property, averaged over binding sites with differing sequences which, by analogy with current views of DNA microstructure, themselves have sequence-dependent helical twists that will unwind in a sequence-dependent manner. Unwinding, then, is in principle governed by rules analogous to Calladine's rules of unperturbed B-DNA (Calladine, 1982), that would take into account the steric properties of the bound intercalant.

The Lerman model does not provide information on the perturbations produced in DNA conformation or their sequence-dependency, or of whether and to what extent interactions themselves show sequence specificity. Until comparatively recently, binding studies with synthetic polynucleotides having simple repeats were the sole approach to this last problem. The introduction of DNA footprinting, which monitors the inhibitory effect of an intercalating drug on the (enzymic or chemical) scission of a defined-sequence DNA fragment (typically of $\sim$ 200 bp in length), has provided extensive data on this point. N.m.r. methods have to date provided less detailed information than on the unperturbed oligonucleotides, in large part on account of the multiplicity of oberved and overlapping binding sites, with a consequent very considerable increase in complexity of analysis, as well as the complication of line-broadening effects.

Consideration of the molecular factors involved in intercalative recognition provides further indications of

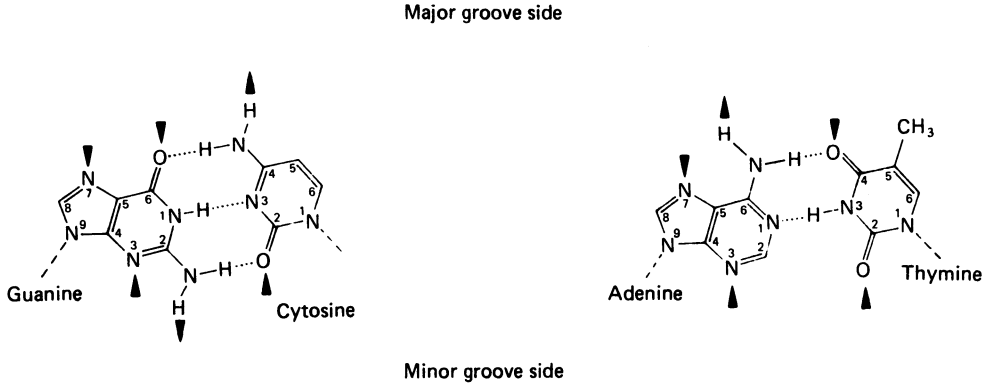


Fig. 1. Donor and acceptor groups in the grooves of A·T and G·C base pairs

Arrows leading into an atom designate acceptor sites, and those leading out designate donor sites.

the potential diversity in binding behaviour shown by intercalating molecules.

(i) Distinct intercalation sites, and differences in neighbouring sequences, imply steric and electronic differences that can be sensitive to both the planar chromophore of an incoming ligand, and its substituents/side chains.

(ii) Steric differences that can differentiate between major and minor DNA grooves can arise when an intercalant has a bulky side-chain or grouping attached.

(iii) The intercalating chromophore itself can have distinct orientations with respect to the base pairs (Islam *et al.*, 1985); extremes are the parallel and perpendicular orientations observed in the crystal structures of respectively proflavine and daunomycin complexed to short oligonucleotides (Berman *et al.*, 1979; Quigley *et al.*, 1980).

(iv) Side chains or groups with potential hydrogen-bonding ability can interact with nucleobase hydrogen-bond donors or acceptors in either major or minor grooves (Fig. 1), as has been suggested for nucleic acid–amino acid interactions (Seeman *et al.*, 1976).

(v) The existence of cationic charge, either on the chromophore itself (cf. the acridines), or on the attached group (cf. the anthracyclines), is an important contributor to the total stabilization of the drug complex (Islam & Neidle, 1984; Neidle *et al.*, 1986). This requirement is not an absolute one, since uncharged drugs with extensive side-chains, typically oligopeptide in nature, are effective DNA-binders; these are typified by actinomycin and echinomycin (Fig. 2). In these cases, extensive non-bonded attractive interactions involving the amino-acid residues and the grooves are important contributors to the binding energies, and thus broadly equate with the electrostatic component of drugs such as daunomycin (Nuss *et al.*, 1979; Chen *et al.*, 1986).

There are ten distinct dinucleoside intercalation sites (detailed in Gale *et al.*, 1981). Their differing strengths of interactions with drug chromophores reflect in part extent of base-chromophore 'overlap' interaction, and in part their mutual polarizibility. For structurally simple mono-intercalators such as the acridines, ethidium or daunomycin, n.m.r. evidence on the basis of chemical shift data with short-length oligonucleotides is strongly suggestive of a preference for pyrimidine–3′,5′–purine sequences (Patel & Canuel, 1976; Phillips & Roberts, 1980; Reinhardt & Krugh, 1978). This has received support from energy calculations on the basis of the relative ease of destacking this type of sequence compared to, in particular, the reversed purine–3′, 5′–pyrimidine type (Broyde & Hingerty, 1979; Ornstein & Rein, 1979; Nuss *et al.*, 1979). In accord with their greater polarizibility, G·C base pair sites have been found to be generally preferred for the intercalation of many drugs to DNA in solution, on the basis of chemical shift data (Feigon *et al.*, 1984), although some classes of intercalators are A·T-selective, probably on the basis of drug–base interactions in addition to stacking ones (Wilson *et al.*, 1985*a*). A recent extensive analysis has re-examined the question of specificity by means of full conformational energy minimization on ethidium with a number of different dinucleoside and hexanucleotide sequences (Lybrand & Kollman, 1985). This has confirmed the earlier findings of the pyrimidine–3′,5′–purine CpG site being preferred over its sequence isomer, by a few kJ/mol; for the hexamer complexes this

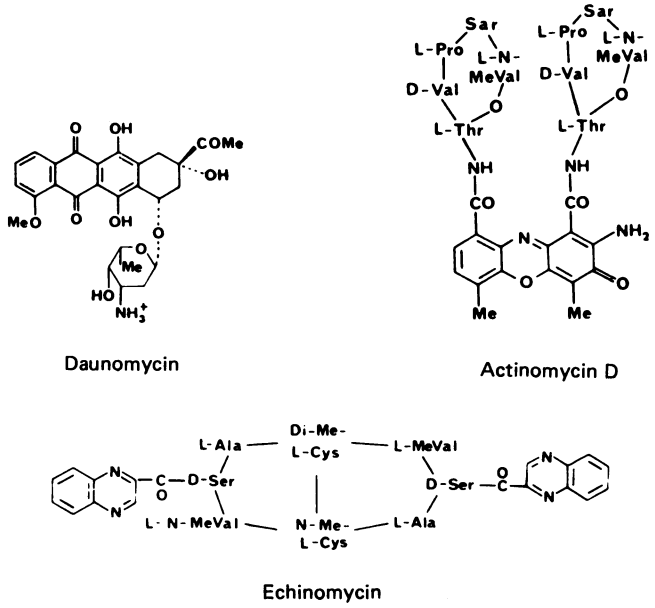


Fig. 2. Structural formulae of various intercalating drugs

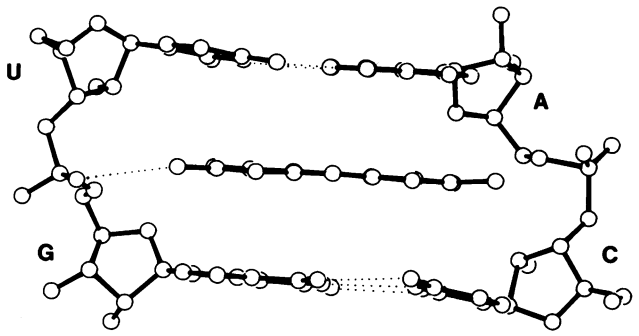


Fig. 3. Computer-drawn picture of the molecular structure of a UpG·CpA–proflavine intercalation complex (Aggarwal *et al.*, 1984)

sequence preference does not hold for A-T sites. For these, the calculations suggest that A–3′,5′–T and T–3′,5′–A intercalation sites are equivalent, and are significantly more favoured than homo-oligomer (i.e. A–3′,5′–A or T–3′,5′–T) sites. The abnormal behaviour of the homopolymer poly(dA)·poly(dT) with intercalators in solution (e.g. Bresloff & Crothers, 1981; Wilson *et al.*, 1985*b*) has long been known, and may be at least in part due to this polynucleotide, with homo-A·T tracts, having structural features, such as bending or heteronomous strands, that are very different from those of other sequences. As yet, structural details remain to be determined, and theoretical studies of drug binding have not examined any of these models.

Crystallographic studies on intercalation complexes have provided further support for these sequence preferences. For the simple mono-intercalators, all except one of the crystal structures of nucleic acid model complexes are with dinucleosides. Only when the dinucleoside is of pyrimidine–3′,5′–purine sequence does the complex have Watson–Crick base pairs, with two dinucleoside strands forming an anti-parallel duplex (Fig. 3). Such structures are formed with, for example, ethidium and (ribo)CpG (Jain *et al.*, 1977; Jain & Sobell, 1984), proflavine with a variety of dinucleosides: CpG (Berman *et al.*, 1979), d(CpG) (Shieh *et al.*, 1980), and in the ternary complex with CpA and UpG (Aggarwal *et al.*, 1984). The conformational disposition to form a duplex-like structure with the pyrimidine–3′,5′–purine sequence is such that in the proflavine–CpA complex (Westhof *et al.*, 1980), two CpA strands associate in an unprecedented parallel manner via C···C and A···A base pairs with an intercalated proflavine situated between them. Without exception, drug complexes with dinucleosides of different sequence type involve non-duplex, 'opened-out' structures; the pentapeptide antibiotic actinomycin has a near-absolute requirement for a GpX intercalation site, yet the crystal structure of its complex with d(GpC) still follows this pattern (Takusagawa *et al.*, 1982).

The sole representative of a crystal structure showing drug mono-intercalation into a sequence longer than a dinucleoside is that of daunomycin and the hexanucleotide d(CGTACG) (Quigley *et al.*, 1980) (Fig. 4*a*). In this, the drug is intercalated from the minor groove, with its sugar group (carrying a protonated amino group) resting into the side of the groove. The anthracycline chromophore is oriented approximately perpendicular to the base-pair long axis direction, in good agreement with n.m.r. chemical shift data (Patel *et al.*, 1981). The selection of the C·G pyrimidine–3′,5′–purine site is in part explained by the existence of a hydrogen bond from the drug to the exocyclic N-2 atom of the guanine, although theoretical calculations (Chen *et al.*, 1985, 1986) indicate a more subtle interplay of sequence-dependent and hydrogen-bonding forces.

Sequence specificity can be revealed by footprinting when (*a*) the drug has a significantly higher affinity for some sites, (*b*) when its dissociation from such sites is both slow enough to inhibit strand cleavage (thus competing effectively with the cleavage agent), and slower than from other sites, and possibly (*c*) when the groove sensed by the cleavage agent is the one actually blocked by ligand. In general, the structurally less complex intercalators have not been successfully footprinted, probably because of the second factor detailed above. By contrast, the more complex drug actinomycin shows specific binding to G+C-rich regions that necessarily contain G–3′,5′–C binding sites (Van Dyke *et al.*, 1982; Lane *et al.*, 1983; Fox & Waring, 1984*b*). This is in accord with earlier data on this drug's requirement for a guanine residue at the 5′ side of its intercalation site, and its exceptionally slow dissociation kinetics from DNA (Muller & Crothers, 1968). Footprinting has also shown that binding of the drug induces conformational changes in the DNA that lead to enhanced cleavage by DNAase I, mostly in A+T-rich runs that in the absence of drug are relatively resistant to the enzyme (Fox & Waring, 1984*b*). It is presumed that the drug binding has increased the minor groove width to a size better able to accommodate the active residues of the enzyme; 1.2–1.3 nm has been suggested as optimal (Suck & Oefner, 1986). These perturbations in DNA structure up- or down-stream from an intercalation site have not as yet been directly observed in crystal structures of drug–oligonucleotide complexes; none of these has a drug molecule with a significant number of nucleotide residues surrounding it in both 3′ and 5′ directions. The crystal structure of the daunomycin–hexamer complex (Quigley *et al.*, 1980) has base pair–base pair unwinding not at the intercalation site itself, but at adjacent residues, indicating that they are significantly affected by the binding. A number of molecular modelling studies have addressed this problem (reviewed in Neidle & Abraham, 1984). The majority have attempted to smooth out any structural perturbations; however, when the intercalation geometry is constrained to its crystallographically-observed form, adjacent residues are markedly altered from their A- or B-type structures (Berman & Neidle, 1979). As yet, molecular dynamics methods have not been successfully applied to intercalated systems; a study of the thermal motion observed in the proflavine–d(CpG) crystal structure has indicated increased flexibility compared with the non-drug sequence (Aggarwal & Neidle, 1985). This is in apparent contradiction to a molecular mechanics estimate of lack of flexibility in the drug complexes (Taylor & Olson, 1983); however, at the present time, the most difficult feature of such structures to parameterize meaningfully is the chromophore–base pair component, which is subject to the most thermal motion.

Echinomycin and analogues. The quinoxaline antibiotics are representative of a more complex class of

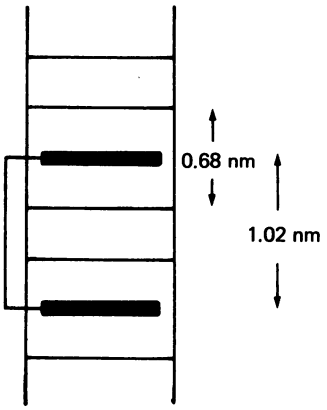


Fig. 5. Schematic representation of bis-intercalation into DNA

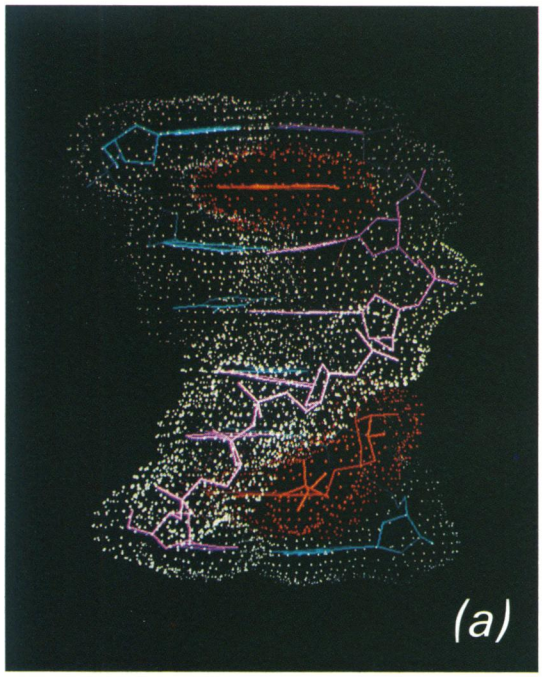
intercalating agents. These have two chromophores that can simultaneously intercalate, with a spacing of 1.02 nm (3 × 0.34 nm) between them (Fig. 5). The parent drug echinomycin (Fig. 2) has eight amino acids arranged in cyclic form. It inhibits DNA replication and DNA-directed RNA synthesis by specific binding to double-stranded DNA (Waring & Fox, 1983). Classic equilibrium-binding studies have shown that the drug binds more strongly to G+C-rich DNA, with a complex pattern of kinetic behaviour (Fox & Waring, 1985). Footprinting studies (Low *et al.*, 1984; Van Dyke & Dervan, 1984) have been made on this drug using respectively DNAase I, and the synthetic footprinting reagent MPE [methidiumpropyl–EDTA–iron(II)]. The latter DNA cleavage agent, being a much smaller molecule than DNAase I, has finer resolving powers for the determination of binding site size. Both methods concur in finding that protected sites are of the C·G type; the MPE footprinting has been able to define the tetranucleotide sequences TCGT, ACGT, TCGA and GCGG as strong binding sites, and that these represent the size of the binding site. The role of adjacent sequences is demonstrated by the observation (Low *et al.*, 1984) that the ability of some XCGY sequences to be protected from cleavage is dependent on their position in the fragment, as is the enhancement to cleavage found for sequences rich in oligo(dA)·oligo(dT) tracts. Presumably, drug binding, even at sites relatively remote from these sequences (typically five or six bases away), is able to propagate sufficient structural perturbation such that these normally DNAase-resistant sequences subsequently have a minor groove width that more precisely fits the active residues of the enzyme. Modelling studies on drug intercalation complexes (Berman & Neidle, 1979; H. M. Berman, S. Neidle & L. H. Pearl, unpublished work), are strongly suggestive of long-range perturbations extending over at least four or five bases.

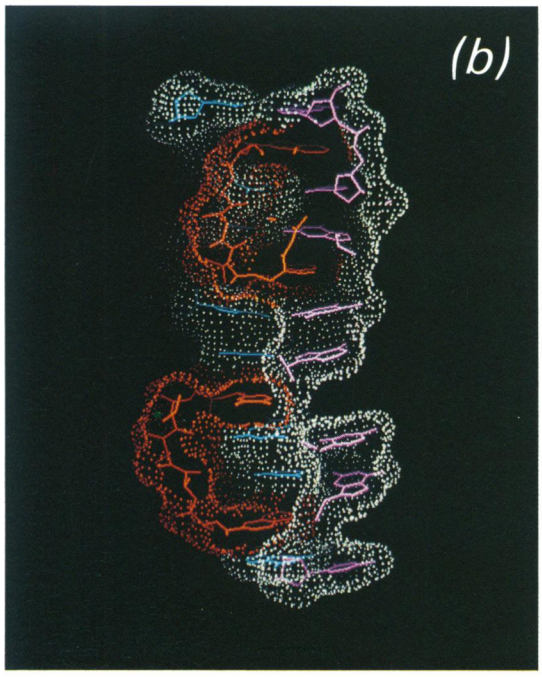
The crystal structures of echinomycin and the closely-related antibiotic triostin A (modified in the disulphide bridge), bound to the hexanucleotide sequence d(CGTACG) have been determined (Wang *et al.*, 1984; Ughetto *et al.*, 1985). The closely-related structures show the anticipated bis-intercalation, with the two quinozaline chromophores being separated by 1.02 nm and thus having two base pairs between them. These are two C·G ones, as found in the footprinting studies. The drugs are bound in the minor groove, with both hydrogen-bonding and van der Waals' interactions between amino acid residues and nucleotide backbone and bases clearly playing important stabilization roles that also impart specificity in several instances. In particular, one alanine residue hydrogen-bonds to the exocyclic N-2 atom of a guanine; both alanines hydrogen-bond to N-3 of both guanines. A remarkable feature of these hexanucleotide complexes is that the central A·T base pairs have Hoogsteen rather than Watson–Crick base-pairing, with the adenosines having to adopt *syn* conformations. This feature also accords with the footprinting data on the C·G binding sites being most frequently flanked by A·T base pairs; these are more easily able to form Hoogsteen base pairs than C·G ones, at least in neutral pH conditions. The analysis of a triostin A complex with the octanucleotide d(GCGTACGC) has confirmed and extended these findings (Quigley *et al.*, 1986). In this structure, two drug molecules each bis-intercalate into the double-stranded duplex (Fig. 4*b*), such that as in the hexamer structures the quinoxaline rings of each drug molecule sandwich between a C-G sequence. The cyclic peptide moiety of the drug similarly hydrogen bonds to the intercalated guanines via the alanine residues. The central A·T base pairs in the sequence are again in the Hoogsteen form; surprisingly, the outermost C·G base pairs at the 3′ and 5′ ends are also in this form, even though this necessitates the cytosine base in each to be protonated at N-3. The fact that the complex was crystallized at pH 6.5, which is higher than the normal pK_a of N-3 protonation, is suggestive of structural factors forcing Hoogsteen base pairing rather than conventional Watson–Crick ones. Quigley *et al.* (1986) suggest that the former maximize attractive van der Waals' interactions between the drug and aspects of the backbone such as sugar residues.

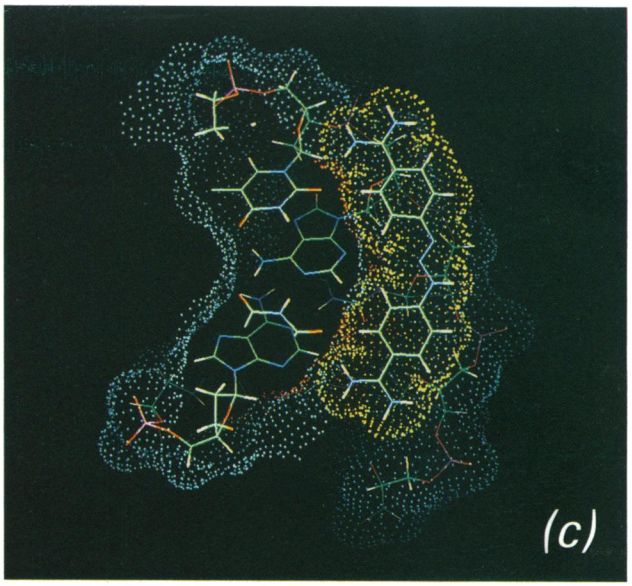
A clear resultant of the Hoogsteen base pairs is that DNA residues adjacent to the bis-intercalation sites would be perturbed from standard B-DNA geometry. To what extent, and in what manner, awaits structural and molecular modelling studies on longer sequences. It will also be of some interest to ascertain whether these features pertain in solution; as yet, n.m.r. studies are largely hampered by the low solubility of these drugs in aqueous systems.

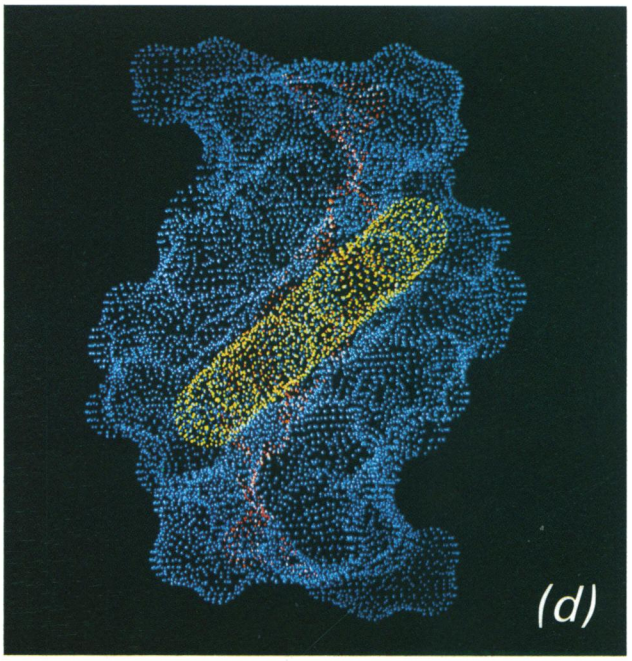
Fig. 4. Computer-drawn representations of drug–oligonucleotide complexes

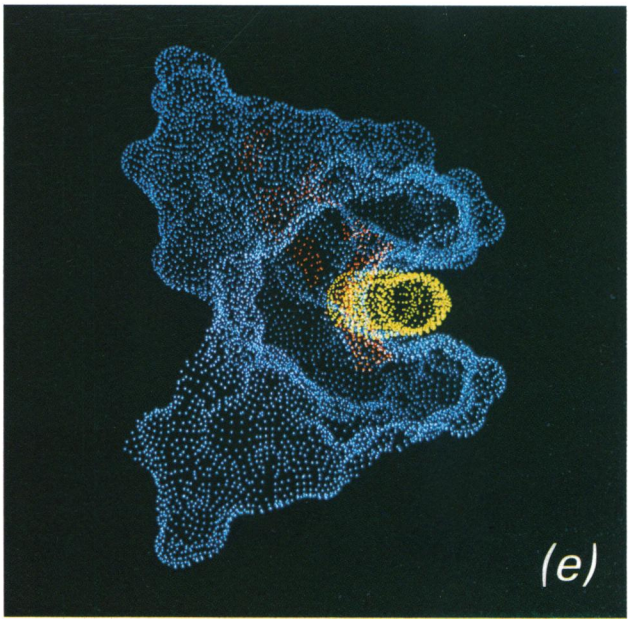
These were drawn with the MIDAS program on an IRIS workstation. (*a*) Daunomycin–hexamer complex (co-ordinates supplied by A. H.-J. Wang). The molecular solvent-accessible surfaces are shown, with daunomycin molecules in red. (*b*) Triostin A–octanucleotide complex (co-ordinates supplied by A. H.-J. Wang). Drug molecules are shown in red; the solvent-accessible surface of both drug and oligomer duplex are shown, sliced through in the Z-axis direction. (*c–d*) Berenil–oligo(dA–dT)·oligo(dA–dT) complex; (*c*), a view along the minor groove, showing the fit of the berenil molecule to groups in the groove; (*d*), a view of the complex as a whole, showing the molecule lying in the minor groove; (*e*), a view down the minor groove.


(a)


(b)


(c)


(d)


(e)

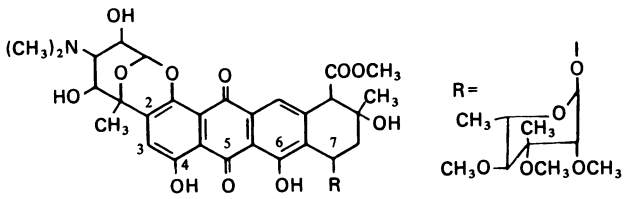


Fig. 6. Structural formula of nogalamycin

Nogalamycin. This anti-tumour antibiotic is a member of the anthracycline family (Brown, 1983). It is distinguished from simpler members such as daunomycin by having bulky sugar residues at both ends of the chromophore (Fig. 6). An X-ray crystallographic analysis of the drug (Arora, 1983), has shown that it has an overall boat-like conformation, with the sterically bulky sugars being prow and stern. In spite of this factor, classic DNA-binding experiments have shown it to be a true DNA intercalator. Molecular modelling (Collier *et al.*, 1984) has suggested that the intercalated complex is forced to have one sugar in each groove (Fig. 7). In energetic terms, each sugar ring does not differentiate especially significantly between major and minor grooves, when using the conformation of the intercalated dinucleoside duplex from the crystal structure of a proflavine complex (Shieh *et al.*, 1980) as a model for intercalated DNA. The nogalmycin complex has a specific hydrogen bond between adenine on the 3′ side of the binding site and an acceptor atom in the sugar group(s), thereby possibly imparting a measure of specificity. Consideration of potential mechanisms for the intercalation of this bulky drug leads at first sight to the paradox of the maximum base-pair separation of an intercalation site being $\sim$ 1.0 nm even with maximum staggering of the nucleotide backbone, compared with the width of this conformationally inflexible drug molecule being $\sim$ 1.2 nm; yet the bound drug state is definitely an intercalated one. The hypothesis has accordingly been advanced (Collier *et al.*, 1984; Neidle, 1986) that the drug initially interacts with non-base-paired, melted ('breathing') stretches of DNA. These would then close up to produce the intercalated state (Fig. 8). This mechanism implies a preference for A+T-rich regions in terms of the association steps. Conversely, dissociation would be expected to be significantly slower from G+C-rich sites, on account of the greater resistance to the disruption of G·C base pairs, which has been estimated to be some 25 times less (Teitelbaum & Englander, 1975).

Study of the association kinetics between nogalamycin and various polynucleotides has found a marked difference between them in terms of reaction rates (Fox & Waring, 1984*a*). The synthetic polymer poly(dA-dT)·poly(dA–dT) reacts fastest, with the absence of a slow component to the overall rate that is observed with calf thymus DNA. This slow component is the sole one found for poly(dG-dC)·poly(dG-dC). These results have been interpreted in a manner consistent with the above model, although it has been pointed out that time-dependent redistribution of drug on the polynucleotide is potentially an important factor. The microstructure of a particular region of DNA, especially with respect to groove width factors variation, is also presumed to play a role in these processes. The measured dissociation kinetics of nogalamycin from various polynucleotides (Fox *et al.*, 1985), are also in accord with the model, since dissociation is almost 200-fold slower from the alternating G-C polymer than the A-T one. The dissociation from natural, random-sequence DNA is more complex, indicating the presence of distinct types of binding site.

Footprinting patterns for nogalamycin with the 160 bp *TyrT* DNA sequence and DNAase I inhibition (Fox & Waring, 1986*a*) show time-dependent changes that are interpretable in terms of drug molecules shuffling between many sites until those of higher affinity/slower dissociation are arrived at. Clearly, this drug is not

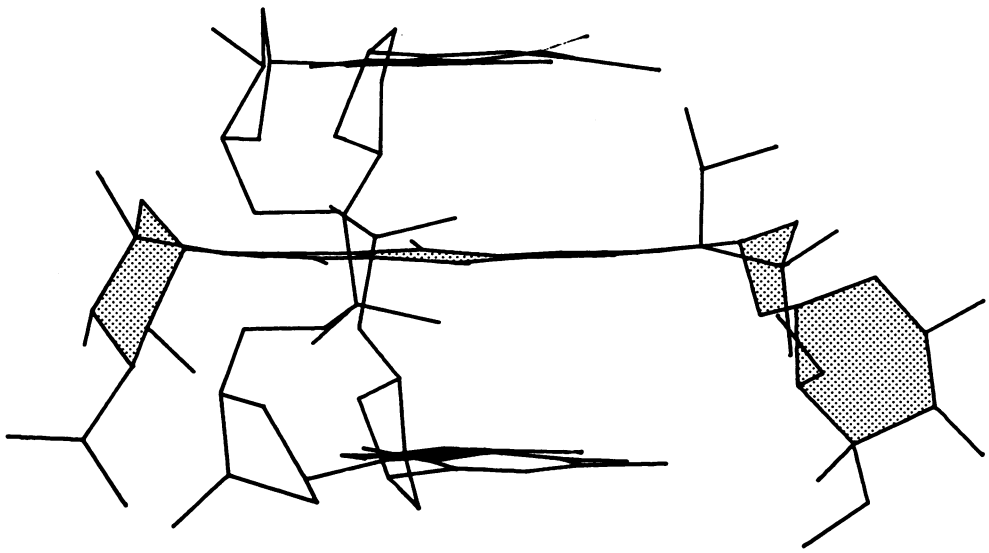


Fig. 7. Computer-drawn representation of a nogalamycin intercalation complex, as visualized by molecular modelling

The view is along the base pairs, looking approximately along the interstrand phosphorus–phosphorus vector; the two nucleic acid strands are almost superimposed. The nogalamycin molecule is shown shaded.

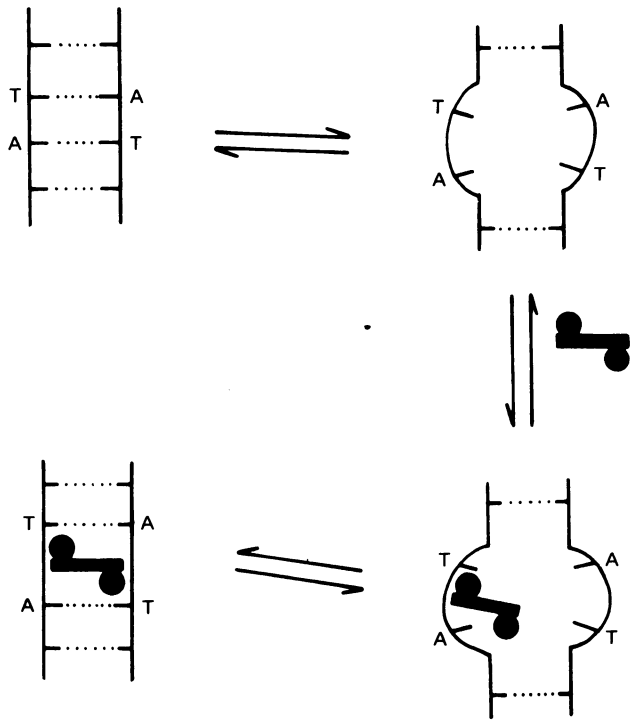


Fig. 8. Schematic model for the interaction of a nogalamycin molecule (shaded) with a region of DNA

sequence-specific in the same straightforward manner as echinomyin or actinomycin. A comprehensive survey of sequence preferences from several DNA restriction fragments (Fox & Waring, 1986*b*) shows that some (though not all) alternating pyrimidine–purine dinucleoside sites are protected from DNAase I digestion by drug binding. This resembles the situation found with a family of anthraquinone derivatives (Fox *et al.*, 1986), and is also probably related to the presumed differing minor and/or major groove dimensions at different loci along the sequences. The indication from the modelling studies of two equi-energetic 180°-related orientations for the bound drug is another factor capable of adding yet further complexity to this picture. Again, the conclusion is suggested that micro-structure, rather than precise sequence *per se*, is a critical factor. In view of this it is unsurprising to note that nogalamycin produces significant enhancements of cleavage, primarily at oligo(dA)·oligo(dT) stretches.

DNA groove binding

It has long been recognized that intercalative mechanisms are not the sole means whereby ligands can interact non-covalently with DNA. Purely non-specific electrostatic interaction can take place with phosphate groups, as shown by, for example, lithium and sodium ions. The major and minor grooves provide ready-made binding sites, and it is therefore unsurprising that a wide variety of molecules (Gale *et al.*, 1981), have been postulated to be groove binders, in that they do not bind intercalatively (as judged by standard criteria such as lack of unwinding of closed-circular DNA), yet bind with comparable association constants and with other than purely electrostatic association.

Binding in major and minor grooves in principle involves steric, electrostatic and hydrogen-bonding factors. The greater width of the major over the minor groove in B-DNA enables it to be the target for α-helices in DNA recognition by at least some prokaryotic gene regulatory proteins (Ohlendorf *et al.*, 1982); this feature is also observed in the crystal structure of an oligonucleotide complex with the restriction endonuclease *Eco*RI. In addition, these proteins are involved in numerous specific hydrogen-bond contacts with the base donors and acceptors available on their edges that form the floors of the grooves. It can be readily inferred (Seeman *et al.*, 1976) that these distinctive differences in donor/acceptor potential for A·T compared with G·C could form the basis of sequence-selective hydrogen-bonded recognition by both proteins and smaller molecules. Such specific interactions have been located in the sole protein–DNA complex structure available at suitable resolution, that of *Eco*RI; it is striking that these major groove ones involve only purine residues along the DNA.

The DNA major groove has greater recognition potential than the minor groove, in terms of both specific hydrogen bonds and the thymine methyl group to provide extra van der Waals' interaction. However, most, if not all, small molecule ligands identified to date as groove binders appear to be minor-groove-specific. This is in large part due to the width of this groove being a good match for the width of these ligands, which are thus stabilized by van der Waals' and hydrophobic interactions with the groove walls. Electrostatic interactions also play an important role (Lavery *et al.*, 1986; Caldwell & Kollman, 1986), which is unsurprising in view of the cationic nature of groove-binders. By no means all have hydrogen-bonding capability, yet A+T selectivity is invariably observed. We address this issue in the following sections.

In general, groove-binding ligands appear to induce only slight changes in gross DNA structure, which is unsurprising since they are fundamentally only displacing minor-groove solvent molecules. Any structural changes would be of 'induced-fit'. The structural feature necessary for this is a complementarity between the curvature of the groove surface and the surface of the drug, which must therefore be helically concave (Goodsell & Dickerson, 1986). The quantification of this concept is clearly of importance for the design of molecules that can bind to and recognize long DNA sequences by means of complementarity with the helicity of B-DNA. This analysis was performed on idealized, precisely repetitious, DNA; as noted in earlier sections, there are sequence-dependent structural aberrations which do complicate the issue, but which nonetheless need to be brought into a more extensive analysis. The idealized analysis shows that the dipyrrole-amide antibiotic netropsin (see below), although having a bowed conformation (Berman *et al.*, 1979), does not perfectly match the minor groove, and that it is unsurprisingly that longer pyrrole-amide repeating units which have been strung together bind less effectively to DNA than does netropsin itself (Dervan, 1986).

Netropsin. This antiviral and antitumour compound has peptide-like features, as well as two pyrrole rings. Its guanidinium groups, one at each end, impart basicity. The drug has long been established as having an A+T preference and to bind to B-form duplex DNA (Zimmer,

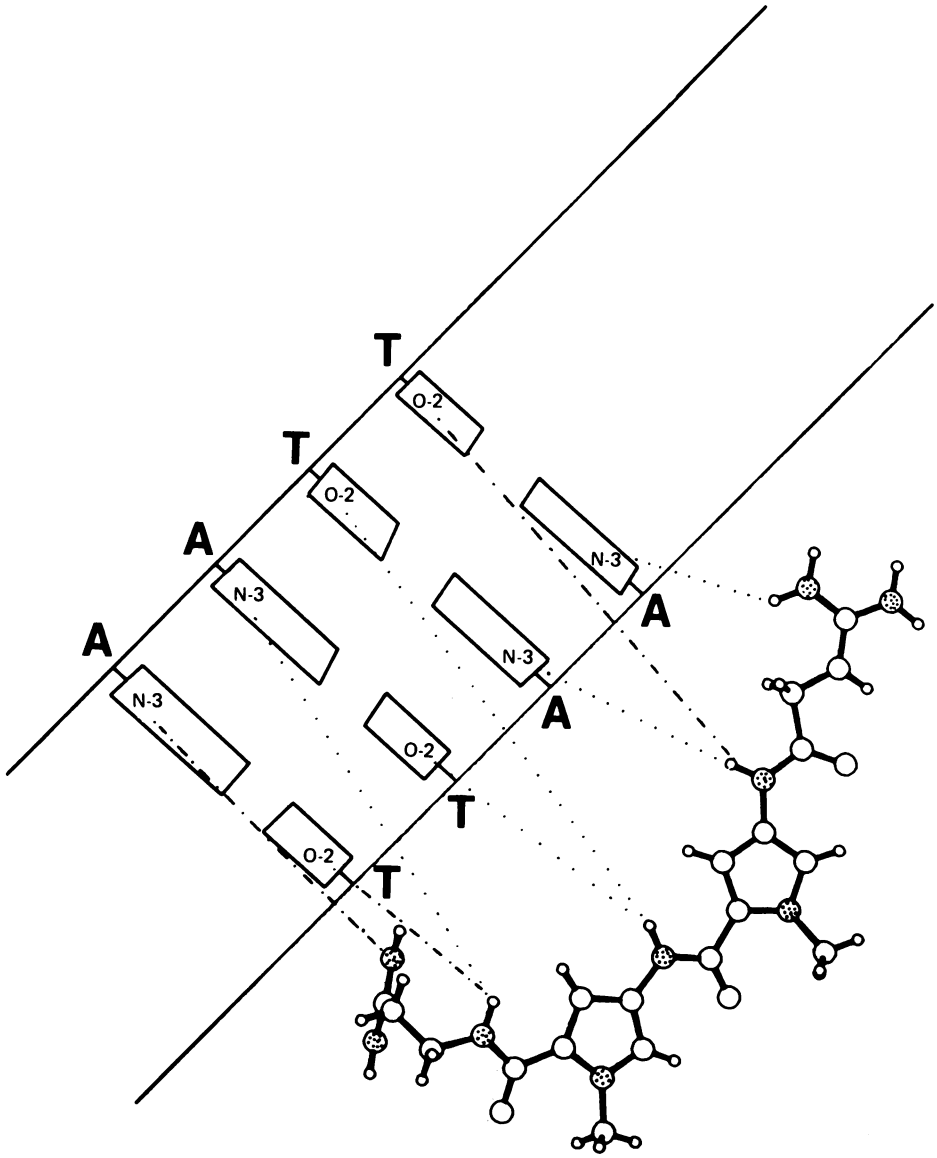


Fig. 9. Schematic representation of the interactions in the minor groove, as found in the crystal structure of the netropsin–dodecanucleotide complex (Kopka *et al.*, 1985*a*,*b*)

Broken lines represent 'good' hydrogen bonds; dotted lines are weaker ones.

1975); A+T homopolymers are preferred. Footprinting experiments on both netropsin and its tripyrrolic analogue distamycin have confirmed and extended this data (Fox & Waring, 1984*b*; Lane *et al.*, 1983; Van Dyke *et al.*, 1982; Harshman & Dervan, 1985), using both enzymic and chemical cleavage reagents. These studies show that the binding sites are A+T-rich, are 5 ± 1 bases long and tend to be non-alternating (for example, AAATT and AATAA). There is clearly some latitude in the precise sequence allowed, even to the extent of a G·C base pair being present in one weakly protected site. This suggests that there is some flexibility in the binding site; as with echinomycin, the DNAase I footprinting (Fox & Waring, 1984*b*) has shown some sequences normally poorly cleaved by enzyme that in the presence of drug (in this instance distamycin), show enhanced cleavage. These are G+C-rich ones.

The crystal structure of netropsin bound to the dodecamer sequence d(CGCGAATTCGCG) has been determined (Kopka *et al.*, 1985*a*,*b*). The native structure, which had earlier been determined as B-form (Dickerson, 1983), is only slightly perturbed by the drug. Netropsin is located symmetrically in the minor groove, where it has displaced the 'spine of hydration' observed in the native structure, so that it is centred in the AATT region. The drug molecule participates in numerous interactions with the DNA (Fig. 9) with the more hydrophobic convex face of the drug facing outwards into the groove. The hydrophilic concave face has exclusively hydrogen-bond donors on it. The shortest hydrogen bonds are to adenine N-3 and thymine O-2 at the 3′ end of the AATT sequence, and to the O-2 of the thymine four base pairs away at the 5′ side. There is an elaborate network of long hydrogen bonds, some of which are bifurcated, primarily involving the two central A·T base pairs. Their weakness suggests that replacement of this AATT sequence by

such as ATAT or AAAT would not make a significant difference to them. It is further suggested (Kopka *et al.*, 1985*a*,*b*) that this observed binding site could not, on steric grounds, accommodate C·G base pairs on account of the N-2 of guanine, and that close contacts of C–H hydrogen atoms on the drug to adenine H-2 atoms are primarily responsible for A·T recognition, with the hydrogen bonds playing a secondary positioning role. Changes in both dodecanucleotide and drug molecule structures compared with their unbound forms have been observed in the complex, and are small. The minor groove width opens up in the vicinity of the drug by 0.05–0.1 nm, and some aspects of the drug's conformation undergo small alterations. These changes aid the fitting of the drug into the binding site, and increase attractive close contacts.

N.m.r. studies have examined netropsin bound to several oligonucleotides at AATT and TATA sequences (Patel & Shapiro, 1985, 1986*a*,*b*), although these analyses have been hampered by the inability to record the chemical shifts of drug alone in solution, on account of its instability. Nuclear Overhauser measurements have confirmed non-bonded contacts involving adenosine H-2 protons of the central two A·T base pairs in d(GGAATTCC) of 0.20–0.25 nm. These contacts are to the amide protons of the drug, and to the pyrrolic protons lining its concave face, and are thus between the exchangeable and non-exchangeable protons of drug and oligonucleotide. They imply that the drug is positioned in a manner close to that found in the crystalline complex. Small changes are found in the conformations of the oligonucleotides; these, though, are indicative of the same type of changes as seen in the crystal structures of native versus drug-complexed oligomers. Overall, the correspondence between solution and crystal is close, although the n.m.r. data are unable to address the question of the relative importance to the binding of non-bonded compared with hydrogen-bonded interactions. More controversial is the conclusion from a qualitative one-dimensional nuclear Overhauser n.m.r. study (Sarma *et al.*, 1985) that netropsin binds asymmetrically to oligo(dA)·oligo(dT) with a widened minor groove such that the close interactions are solely with the dA strand. Hydrogen bonds between imino protons and N-3 adenine atoms are postulated, on the basis of nuclear Overhauser effects being observed only between the H-2 proton of adenine and netropsin pyrrolic and imino protons. This model has been refuted (Dickerson & Kopka, 1985) on the basis of both stereochemical arguments and a re-examination of the n.m.r. data presented by Sarma *et al.* (1985). They show that nuclear Overhauser effects between netropsin and thymine protons would not be expected on the basis of the crystal structure of the complex being close to its solution conformation (and indeed are not found), yet their absence does not imply that the drug is only bound to adenines. Even though final confirmation must await the crystal structure of netropsin bound to a longer oligo(dA)·oligo(dT) sequence than in the dodecamer complex, the majority of evidence appears to support its general relevance.

There have been a number of theoretical studies of netropsin binding to DNA sequences. A molecular mechanics approach (Zakrzewska et al., 1983) found that the drug has oligo(dA)·oligo(dT) minor groove preference, by some 188–209 kJ/mol over a G·C sequence, with in the former instance a number of hydrogen bonds to the drug. Not all of these hydrogen bonds are observed in the crystal structure of the complex. A further analysis (Lavery *et al.*, 1985), taking into account DNA flexibility and cations, concurs with the crystal structure in terms of an increased minor groove width, although in terms of the positioning of the netropsin molecule, an energetically stable state was found with some preference for adenosines. This is in partial accord with both the crystal structure and the alternative Sarma *et al.* (1985) model. A molecular mechanics study with several A+T sequences, as well as the dodecamer d(CGCGAATTCGCG), has been reported (Caldwell & Kollman, 1986). This also finds that oligo(dA)·oligo(dT) is a good binding sequence for the drug, and moreover that the predicted oligomer complex is close to that (subsequently) found in the crystal. The pattern of hydrogen bonds in the computed complex resembles the experimentally-determined one, apart from strong guanidinium–phosphate bonds in the former. These are most probably an artifact arising from the lack of solvent-shielding effects in the calculations.

In summary, the extensive studies on netropsin–nucleic acid interactions have provided an almost entirely consistent test case for the hypothesis that the structure of the complex in the crystalline state is a reliable indication of its solution structure. Such differences as have been found are in part due to deficiencies in methodology, but may also reflect aspects of the drug's mobility in its binding site.

Berenil. This diarylamidine (Fig. 10) possesses antiparasitic and DNA polymerase inhibitory activity, almost certainly on account of its DNA-binding ability (Newton, 1967). The crystal structure of the drug itself (L. H. Pearl, J. V. Skelly & S. Neidle, unpublished work), shows the crescent shape characteristic of groove binders, and footprinting studies (J. Portugal & M. J. Waring, personal communication) have indicated an A+T specificity similar to that of netropsin. Molecular modelling to the oligo(dA-dT)·oligo(dA-dT) shows that berenil fits tightly into the minor groove (Figs. 4*c*–4*e*), and that its molecular surface complements precisely that of the groove, in accord with the isohelical predictions of Goodsell & Dickerson (1986). This fitting is representative of many groove-binding ligands. The postulated hydrogen-bonding scheme of berenil with the polynucleotide (Fig. 11) differs from that of netropsin, in that only two adjacent A·T base pairs are involved, which are required to be alternating. In this way, hydrogen bonds to the O-2 atoms of the thymine can be formed with good geometry and therefore binding to the alternative central A-A sequence is less effective. The alternating requirement is in accord with other theoretical studies (Gresh & Pullman, 1984; Zakrzewska *et al.*, 1984).

The A+T selectivity of these and other groove binders have led several workers to attempts at designing molecules that have G+C recognition features in addition (Dervan, 1986; Dickerson *et al.*, 1986; Lown *et al.*, 1986). One postulated approach is to alter the

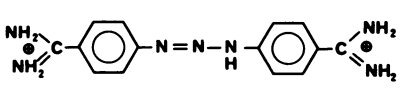


Fig. 10. Structural formula of berenil

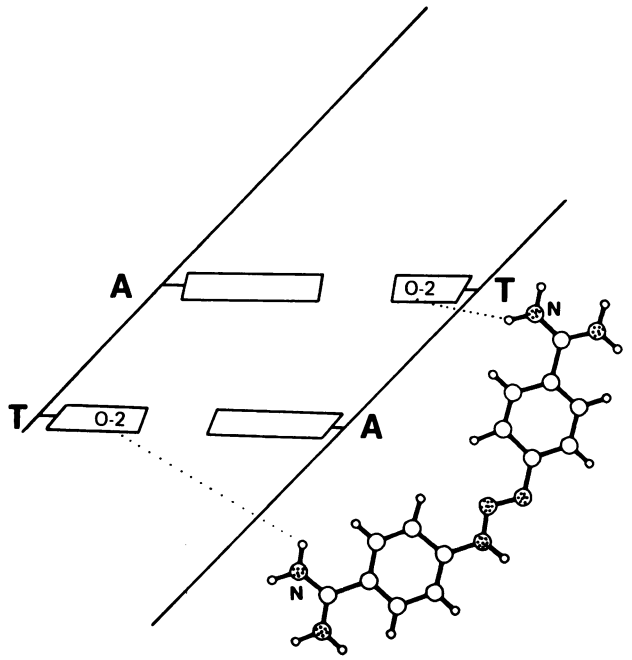


Fig. 11. Schematic representation of the hydrogen bonding between berenil and the O-2 atoms of thymine

hydrogen-bonding ability in part of the ligand from donor to acceptor by replacing a pyrrole group with an amidazole. Another has been (Dervan, 1986) to attach covalently two distamycin molecules to the intercalating phenoxazone chromophore of actinomycin (which would be expected to show a G+C preference). The results from this molecule have been less specific than hoped, which in retrospect is unsurprising in view of the current lack of knowledge concerning the structural perturbations produced by simultaneous intercalation and groove binding. More radical approaches, for example exploiting the hydrogen-bonding capabilities of the major groove, have not as yet been explored.

Concluding remarks

The variations in DNA structure, both locally and long-range, are due to the interdependent factors of external environment and sequence. We are now beginning to understand the structure and dynamics of short oligonucleotide sequences, from both X-ray crystallographic and high-field n.m.r. studies. At least some features, such as sequence-dependent groove dimensionality, are probably applicable to longer sequences, and possibly relate to DNA in nucleosomes, although the structural data for such generalizations is as yet relatively scanty. The more complex features of DNA flexibility are much less well understood, especially those concerned with longer-range effects involving the relationships between structural perturbations themselves separated by long stretches of nucleotide sequence.

Interactions of DNA with either low-M_r or macromolecular ligands invariably involves at least some degree of conformational change in the DNA, although remarkably little is known about many of them. Intercalation is probably the best-studied category of DNA–ligand interaction, which necessarily involves a large-scale local structural change in order to produce the binding site; the crystallographic data on intercalation complexes has enabled detailed understanding of the geometry of the site itself and of its immediate surroundings. Longer-range effects, as observed in nuclease digestion experiments, are not yet describable in such structural terms. Sequence-specific recognition of DNA by groove-binding ligands involves a number of factors: minor-groove and hydrophobic electrostatic interaction, hydrogen-bonding and displacement of bound water. The crystal structure of the netropsin–dodecanucleotide complex has shown that there are also a large number of small individual changes in DNA confirmation that open up the minor groove and maximize these interactions by a sequence-dependent induced fit. Again, there are long-range DNA structural ramifications remote from drug binding sites whose nature is as yet not understood. The elucidation of the rules governing all these changes will have to be taken in conjunction with advances in knowledge of the recognition processes themselves, before the goal of synthetic DNA-recognition molecules tailor-made to defined sequences can be reached.

We are grateful to Helen Berman, Peter Dervan, Keith Fox and Michael Waring for discussions, and to Marius Clore, Dick Dickerson, Dinshaw Patel and Andy Wang for information and data prior to publication. Bob Langridge is thanked for providing the MIDAS programs. Work in the authors' laboratory has been supported by the Cancer Research Campaign and the Medical Research Council.

References

Aggarwal, A. K. & Neidle S. (1985) Nucleic Acids Res. **13**, 5671–5684
Aggarwal, A., Islam, S. A., Kuroda, R. & Neidle, S. (1984) Biopolymers **23**, 1025–1041
Arnott, S. (1981) in Topics in Nucleic Acid Structure (Neidle, S., ed.), pp. 65–82, Macmillan, London
Arnott, S. Chandrasekaran, R., Hall, I. H., Puigjaner, L. C., Walker, J. K. & Wang, M. (1982) Cold Spring Harbor Symp. Quant. Biol. **47**, 53–65
Arnott, S., Chandrasekaran, R., Banerjee, A. K., He, R. & Walker, J. K. (1983*a*) J. Biomol. Struct. Dyn. **1**, 437–452
Arnott, S., Chandrasekaran, R., Hall, I. H. & Puigjaner, L. C. (1983*b*) Nucleic Acids Res. **11**, 4141–4155
Arora, S. K. (1983) J. Am. Chem. Soc. **105**, 1328–1332
Baguley, B. (1982) Mol. Cell. Biochem. **43**, 167–181
Berman, H. M. & Neidle, S. (1979) in Nucleic Acid Geometry and Dynamics (Sarma, R. H., ed.), pp. 325–340, Pergamon Press, New York
Berman, H. M., Neidle, S., Stallings, W., Taylor, G., Carrell H. L., Glusker, J. P. & Achari, A. (1979) Biopolymers **18**, 2405–2429
Berman, H. M., Neidle, S., Zimmer, C. & Thrum, H. (1979) Biochim. Biophys. Acta **561**, 124–131
Bresloff, J. J. & Crothers, D. M. (1981) Biochemistry **20**, 3537–3553
Broyde, S. & Hingerty, B. (1979) Biopolymers **18**, 2905–2910
Brown, J. R. (1983) in Molecular Aspects of Anti-cancer Drug Action (Neidle, S. & Waring, M. J., eds.), pp. 57–92, Macmillan, London
Caldwell, J. & Kollman, P. (1986) Biopolymers **25**, 249–266
Calladine, C. R. (1982) J. Mol. Biol. **161**, 343–352
Chazin, W. J., Wüthrich, K., Hyberts, S., Rance, M., Denny, W. A. & Leupin, W. (1986) J. Mol. Biol. **190**, 439–453
Chen, K.-X., Gresh, N. & Pullman, B. (1985) J. Biomol. Struct. Dyn. **3**, 445–466
Chen, K.-X., Gresh, N. & Pullman, B. (1986) Nucleic Acids Res. **14**, 2251–2267

Clore, G. M. & Gronenborn, A. M. (1985) in Biomolecular Stereodynamics IV (Sarma, R. H. & Sarma, M. H., eds.), pp. 139–155, Adenine Press, New York
Clore, G. M., Gronenborn, A. M., Moss, D. S. & Tickle, I. J. (1985) J. Mol. Biol. **185**, 219–226
Collier, D. A., Neidle, S. & Brown, J. R. (1984) Biochem. Pharmacol. **33**, 2877–2880
Dervan, P. B. (1986) Science **232**, 464–471
Dickerson, R. E. (1983) J. Mol. Biol. **166**, 419–441
Dickerson, R. E. & Drew, H. E. (1981*a*) J. Mol. Biol. **149**, 761–786
Dickerson, R. E. & Drew, H. R. (1981*b*) Proc. Natl. Acad. Sci. U.S.A. **78**, 7318–7322
Dickerson, R. E. & Kopka, M. L. (1985) J. Biomol. Struct. Dyn. **3**, 423–436
Dickerson, R. E., Kopka, M. L. & Pjura, P. (1985) in Biological Macromolecules and Assemblies (Jurnak, F. A. & McPherson, A., eds.), vol. 2, pp. 37–126, John Wiley, New York
Dickerson, R. E., Kopka, M. L. & Pjura, P. (1986) Chem. Scripta **26**, in the press
Drew, H. R. (1984) J. Mol. Biol. **176**, 535–557
Drew, H. R. & Dickerson, R. E. (1981) J. Mol. Biol. **151**, 535–556
Drew, H. R. & Travers, A. A. (1984) Cell **37**, 491–502
Drew, H. R. & Travers, A. A. (1985) Nucleic Acids Res. **13**, 4445–4467
Feigon, J., Denny, W. A., Leupin, W. & Kearns, D. R. (1984) J. Med. Chem. **27**, 450–465
Feuerstein, B. G., Pattabiraman, P. & Marton, L. J. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 5948–5952
Fox, K. R. & Waring, M. J. (1984*a*) Biochim. Biophys. Acta **802**, 162–168
Fox, K. R. & Waring, M. J. (1984*b*) Nucleic Acids Res. **12**, 9271–9285
Fox, K. R. & Waring, M. J. (1985) Nucleic Acids Res. **13**, 595–603
Fox, K. R. & Waring, M. J. (1986*a*) Biochemistry **25**, 4349–4356
Fox, K. R. & Waring, M. J. (1986*b*) Nucleic Acids Res. **14**, 2001–2014
Fox, K. R., Brassett, C. & Waring, M. J. (1985) Biochim. Biophys. Acta **840**, 383–392
Fox, K. R., Waring, M. J., Brown, J. R. & Neidle, S. (1986) FEBS Lett. **202**, 289–294
Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B.-C., Greene, P., Boyer H. W. & Rosenberg, J. M. (1984) Nature (London) **309**, 327–332
Gale, E. F., Cundliffe, E., Reynolds, P. E., Richmond, M. H. & Waring, M. J. (1981) The Molecular Basis of Antibiotic Action, 2nd edn., John Wiley, London
Goodsell, D. & Dickerson, R. E. (1986) J. Med. Chem. **29**, 727–733
Gresh, N. (1985) Biopolymers **24**, 1527–1542
Gresh, N. & Pullman, B. (1984) Mol. Pharmacol **25**, 452–458
Harshman, K. D. & Dervan, P. V. (1985) Nucleic Acids Res. **13**, 4825–4835
Islam, S. A. & Neidle, S. (1984) Acta Crystallogr. **40**, 424–429
Islam, S. A., Neidle, S., Gandecha, B. M., Partridge, M., Patterson, L. H. & Brown, J. R. (1985) J. Med. Chem. **28**, 857–864
Jain, S. C. & Sobell, H. M. (1984) J. Biomol. Struct. Dyn. **1**, 1179–1194
Jain, S. C., Tsai, C.-C. & Sobell, H. M. (1977) J. Mol. Biol. **114**, 319–331
Kollman, P. A., Weiner, P. K. & Dearing, A. (1981) Biopolymers **20**, 2583–2621
Koo, H.-S., Wu, H.-M. & Crothers, D. M. (1986) Nature (London) **320**, 501–506
Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E. (1985*a*) J. Mol. Biol. **183**, 553–563
Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E. (1985*b*) Proc. Natl. Acad. Sci. U.S.A. **82**, 1376–1380
Lane, M. J., Dabrowiak, J. C. & Vournakis, J. N. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3260–3264
Lavery, R. & Pullman, B. (1981) Nucleic Acids Res. **9**, 3765–3777
Lavery, R., Zakrzewska, K. & Pullman, B. (1986) J. Biomol. Struct. Dyn. **3**, 115–117
Lerman, L. S. (1961) J. Mol. Biol. **3**, 18–30
Low, C. M. L., Drew, H. R. & Waring, M. J. (1984) Nucleic Acids Res. **12**, 4865–4879
Lown, J. W., Kowicki, K., Balzarini, J. & De Clercq, E. (1986) J. Med. Chem. **29**, 1210–1214
Lybrand, T. & Kollman, P. (1985) Biopolymers **24**, 1863–1879
McCall, M., Brown, T., Hunter, W. N. & Kennard, O. (1986) Nature (London) **322**, 661–664
McCall, M., Brown, T. & Kennard, O. (1985) J. Mol. Biol. **183**, 385–396
Mahendrasingam, A., Forsyth, V. T., Hussain, R., Greenall, R. J., Pigram, W. J. & Fuller, W. (1986) Science **233**, 195–197
Muller, W. & Crothers, D. M. (1968) J. Mol. Biol. **35**, 251–290
Neidle, S. (1986) Drug Exp. Clin. Res. **12**, 455–462
Neidle, S. & Abraham, Z. (1984) CRC Crit. Rev. Biochem. **17**, 73–121
Neidle, S. & Berman, H. M. (1983) Prog. Biophys. Mol. Biol. **41**, 43–66
Neidle, S., Abraham, Z. H. L., Collier, D. A. & Islam, S. A. (1986) in New Avenues in Developmental Cancer Chemotherapy (Harrap, K. R. & Connors, T. A., eds.), Academic Press, in the press
Newton, B. (1967) Biochem. J. **105**, 50–51
Nilsson, L., Clore, G. M., Gronenborn, A. M., Brunger, A. T. & Karplus, M. (1986) J. Mol. Biol. **188**, 455–475
Nuss, M. E., Marsh, F. J. & Kollman, P. A. (1979) J. Am. Chem. Soc. **101**, 825–833
Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Matthews, B. W. (1982) Nature (London) **298**, 718–723
Ornstein, R. L. & Rein, R. (1979) Biopolymers **18**, 1277–1291
Patel, D. J. & Canuel, L. L. (1976) Proc. Natl. Acad. Sci. U.S.A. **73**, 3343–3347
Patel, D. J. & Shapiro, L. (1985) Biochimie **67**, 887–915
Patel, D. J. & Shapiro, L. (1986*a*) Biopolymers **25**, 707–727
Patel, D. J. & Shapiro, L. (1986*b*) J. Biol. Chem. **261**, 1230–1240
Patel, D. J., Kozlowski, S. & Rice, J. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 3333–3337.
Patel, D. J., Kozlowski, S. A., Marky, L. A., Broka, C., Rice, J. A., Itakura, K. & Breslauer, K. J. (1982*a*) Biochemistry **21**, 428–436
Patel, D. J., Pardi, A. & Itakura, K. (1982*b*) Science **216**, 581–590
Patel, D. J., Kozlowski, S. A. & Bhatt, R. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3908–3912
Patel, D. J., Shapiro, L. & Hare, D. (1986) J. Biol. Chem. **261**, 1223–1229
Peck, L. C. & Wang, J. C. (1981) Nature (London) **292**, 375–378
Phillips, D. R. & Roberts, G. C. K. (1980) Biochemistry **19**, 4795–4801
Quigley, G. J., Wang, A. H.-J., Ughetto, G., van der Marel, G., van Boom, J. H. & Rich, A. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 7204–7208
Quigley, G. J., Ughetto, G., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J. & Rich, A. (1986) Science **232**, 1255–1258
Rao, S. N. & Kollman, P. (1985) J. Am. Chem. Soc. **107**, 1611–1617
Reinhardt, C. G. & Krugh, T. R. (1978) Biochemistry **17**, 4845–4854
Rhodes, D. & Klug, A. (1981) Nature (London) **292**, 378–380
Rhodes, D. & Klug, A. (1986) Cell **46**, 123–132
Ryder, K., Silver, S., DeLucia, A. L., Fanning, E. & Tegtmeyer, P. (1986) Cell **44**, 719–725
Sarma, M. H., Gupta, G. & Sarma, R. H. (1985) J. Biomol. Struct. Dyn. **2**, 1085–1095

Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) Proc. Natl. Acad. Sci. U.S.A. **73**, 804–808
Seibel, G. L., Singh, U. C. & Kollman, P. A. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 6537–6540
Shakked, Z., Rabinovich, D., Kennard, O., Cruse, W. B. T., Salisbury, S. A. & Viswamitra, M. A. (1983) J. Mol. Biol. **166**, 183–201
Shieh, H.-S., Berman, H. M., Dabrow, M. & Neidle, S. (1980) Nucleic Acids Res. **8**, 85–97
Singh, U. C., Weiner, S. J. & Kollman, P. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 755–759
Singleton, C. K. (1986) in Topics in Nucleic Acid Structure (Neidle, S., ed), vol. 3, Macmillan, London, in the press
Simpson, R. T. & Kunzler, P. (1979) Nucleic Acid Res. **6**, 1387–1415
Suck, D. & Oefner, C. (1986) Nature (London) **321**, 620–625
Takusagawa, F., Dabrow, M., Neidle, S. & Berman, H. M. (1982) Nature (London) **296**, 466–469
Taylor, E. R. & Olson, W. K. (1983) Biopolymers **22**, 2667–2702
Teitelbaum, H. & Englander, S. W. (1975) J. Mol. Biol. **92**, 55–78
Ughetto, G., Wang, A. H.-J., Quigley, G. J., van der Marel, G. A., van Boom, J. H. & Rich, A. (1985) Nucleic Acids Res. **13**, 2305–2323
Van Dyke, M. M. & Dervan, P. B. (1984) Science **225**, 1122–1127
Van Dyke, M. W., Hertzberg, R. P. & Dervan, P. B. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 5470–5474
Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. & Rich, A. (1979) Nature (London) **282**, 680–686
Wang, A. H.-J., Fujii, S., van Boom, J. H. & Rich, A. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 3968–3972
Wang, A. H.-J., Ughetto, G., Quigley, G. J., Hakoshima, T., van der Marel, G. A., van Boom, J. H. & Rich, A. (1984) Science **225**, 1115–1121
Wang, A. H.-J., Gessner, R. V., van der Marel, G. A., van Boom, J. H. & Rich, A. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 3611–3615
Waring, M. J. & Fox, K. R. (1983) in Molecular Aspects of Anti-cancer Drug Action (Neidle, S. & Waring, M. J., eds.), pp. 127–156, Macmillan, London
Westhof, E., Rao, S. T. & Sundaralingam, M. (1980) J. Mol. Biol. **142**, 331–361
Wilson, W. D., Wang, Y.-H., Krishnamoorthy, C. R. & Smith, J. C. (1985*a*) Biochemistry **24**, 3991–3999
Wilson, W. D., Wang, Y.-H., Kusuma, S., Chandrasekaran, S., Yang, N. C. & Boykin, D. W. (1985*b*) J. Am. Chem. Soc. **107**, 4989–4995
Wing, R. M., Drew, H. R., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1980) Nature (London) **287**, 755–758
Zakrzewska, K., Lavery, R. & Pullman, B. (1983) Nucleic Acids Res. **11**, 8825–8839
Zakrzewska, K., Lavery, R. & Pullman, B. (1984) Nucleic Acids Res. **12**, 6559–6574
Zimmer, C. (1975) Prog. Nucleic Acids Res. Mol. Biol. **15**, 285–318