Nucleic Acids Research, 1995, Vol. 23, No. 17 **3385–3392**

DNA sequence preferences of several AT-selective minor groove binding ligands

Anita Abu-Daya, Philip M. Brown and Keith R. Fox*

Department of Physiology and Pharmacology, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, UK

Received June 20, 1995; Revised and Accepted August 3, 1995

ABSTRACT

We have examined the interaction of distamycin, netropsin, Hoechst 33258 and berenil, which are AT-selective minor groove-binding ligands, with synthetic DNA fragments containing different arrangements of AT base pairs by DNase I footprinting. For fragments which contain multiple blocks of $(A/T)_4$ quantitative DNase I footprinting reveals that AATT and AAAA are much better binding sites than TTAA and TATA. Hoechst 33258 shows the greatest discrimination between these sites with a 50-fold difference in affinity between AATT and TATA. Alone amongst these ligands, Hoechst 33258 binds to AATT better than AAAA. These differences in binding to the various AT-tracts are interpreted in terms of variations in DNA minor groove width and suggest that TpA steps within an AT-tract decrease the affinity of these ligands. The behaviour of each site also depends on the flanking sequences; adjacent pyrimidine–purine steps cause a decrease in affinity. The precise ranking order for the various binding sites is not the same for each ligand.

INTRODUCTION

A large number of compounds have been described which bind reversibly to the minor groove of AT-rich DNA regions. In recent years many attempts have been made to synthesize new compounds, with altered or enhanced sequence selectivity so as to permit accurate targeting of specific DNA sequences (1,2). However, before novel derivatives can be rationally designed it is necessary to understand the factors that influence their binding and specificity. In this paper we examine in detail the sequence selectivity of four minor groove binding ligands (Fig. 1) and attempt to correlate the results with known variations in DNA structure. All these compounds bind in the DNA minor groove where they interact with between three and five AT base pairs. Guanine bases are excluded from the binding sites as a result of steric clash with the 2-amino group.

The AT-preference of distamycin and netropsin has been well established by several footprinting studies (3–6). Netropsin requires a target site of four AT base pairs, while distamycin binds optimally to five contiguous AT base pairs, on account of its larger size. Footprinting studies have shown that distamycin is more tolerant of a GC base pair at the end of the binding site (5–7). Both distamycin and netropsin bind to $d(A)_n.d(T)_n$ better than $d(AT)_n$ (8) and, in common with other minor groove binders, the presence of a TpA (but not ApT) step reduces the binding affinity (9). The crystal structure of netrosin bound to CGCGAATTCGCG (10,11) shows the compound bound to the narrow minor groove of the central AATT region. The crescent-shaped compound is isohelical with DNA and binds with its three amides facing the floor of the groove, forming two sets of bifurcated hydrogen bonds with N3 of adenine and O2 of thymine. Van der Waals interactions with the wall of the narrow minor groove stabilize the complex. In contrast the crystal structure of netropsin bound to the dodecamer CGCGATATCGCG (12) shows that it is anchored by single hydrogen bonds between each amide NH group and the O2 of adenine and the N3 of thymine.

The AT specificity of berenil has been confirmed by footprinting studies (6,13). In a complex with CGCGAATTCGCG (14) the drug binds to the AAT segment, with one amidinium group forming a hydrogen bond to N3 of adenine. At the other amidine a water molecule forms a bridge to adenine N3 and the ring oxygen of the adjacent deoxyribose. In contrast, the structure of berenil with CGCAAATTTGCG shows the ligand symmetrically bound to the central four AT base pairs (15), with its amidinium groups hydrogen bonded to O2 of thymines at each end of the binding site. This difference between AAATTT and AATT emphasises the subtle role played by DNA structure in determining the precise location of these ligands. Molecular modelling studies suggest that the formation of hydrogen bonds to adenine N3 atoms is hindered by the adenine H2 atoms which come in close contact with the phenyl rings of berenil (16); this might explain the preference for binding to ApT over TpA, and is consistent with the observation that GCTTAAGC is a weaker binding site than GCAATTGC (17).

Footprinting studies indicate that Hoechst 33258 covers at least four AT-base pairs, though a GC can be accommodated at the end of the binding site, and that TpA steps weaken the binding (18,19). Two X-ray structures have been determined for Hoechst 33258 bound to CGCGAATTCGCG, which show the drug bound across either ATTC (20) or AATT (21,22). Similar differences are found in the structures of Hoechst 33258 bound to CGCAAATTTGCG, which show the ligand located either centrally across AATTT (23) or across ATTTG (25). The selectivity of this

* To whom correspondence should be addressed

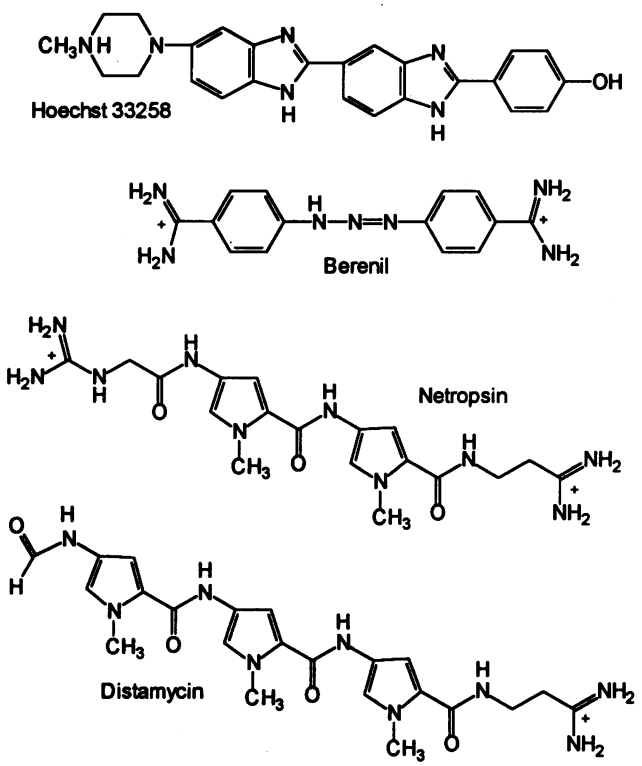

Figure 1. Structures of the four minor groove binding ligands.

ligand is emphasised in an NMR study with GGTAATTAC, in which it binds in a unique position across the central AATT, even though several other AT-binding modes are theoretically possible (25).

Despite the large number of structures which have been determined for these ligands bound to short DNA fragments, there have been few systematic studies comparing their interaction with different combinations of A and T residues. It is clear that binding should be best at those sequences which possess a narrow minor groove, forming good Van der Waals contacts with the walls of the groove. However, it is also clear that the contacts between the ligands and their binding sites, as well as the arrangement and number of hydrogen bonds, varies between different sequences as well as between different ligands. We have therefore used quantitative DNase I footprinting to assess the binding of distamycin, netropsin, Hoechst 33258 and berenil to different arrangements of A and T bases. To facilitate comparison of the binding to different AT-sites we have used cloned synthetic DNA fragments as footprinting substrates, each of which contains multiple $(A/T)_4$ sites located in identical sequence contexts.

MATERIALS AND METHODS

Chemicals and enzymes

Distamycin, netropsin, berenil and Hoechst 33258 were purchased from Sigma and stored as 2 mM solutions in 10 mM Tris–HCl pH 7.5 containing 10 mM NaCl at –20°C. Bovine pancreatic DNase I, was purchased from Sigma and stored at –20°C at a concentration of 7200 U/ml. AMV reverse transcriptase was purchased from Promega; restriction enzymes were purchased from Promega, Pharmacia or New England Biolabs. *Bam*H1 (or *Sma*I) cut, alkaline phosphatase treated pUC18 was purchased from Pharmacia.

DNA fragments

For plasmid pAAD1 (sequence shown in Table 1) the appropriate oligonucleotides were cloned into the *Bam*HI site of pUC18. The ligation mixtures were transformed into *Escherichia coli* TG2 and successful clones were picked in the usual way as white colonies from agar plates containing X-gal and IPTG. Plasmids pAT_4a, pAT_4b and pAT_4d were obtained from Dr P. E. Nielsen (Department of Biochemistry, Panum Institute, University of Copenhagen) and transformed into *E.coli* TG2. The sequences of all these inserts were confirmed using a T7 sequencing kit (Pharmacia) and are shown in Table 1. Labelled DNA fragments containing these inserts were obtained by cutting the plasmids with *Hin*dIII, labelling at the 3′ end using [α-^{32}P]dATP and AMV reverse transcriptase and cutting again with *Eco*RI or *Sac*I (for fragments containing an internal *Eco*RI site). Radiolabelled fragments were separated from the rest of the plasmid on non-denaturing 8% (w/v) polyacrylamide gels. The isolated DNA fragments were dissolved in 10 mM Tris–HCl containing 0.1 mM EDTA at ~10 nM strand concentration.

Table 1. Sequences of plasmid inserts for pAAD1, pAT_4a, pAT_4b and pAT_4d

pAAD1
GTACGCGTTAACGCGCGATATCGCGCGTAATCGCGCGTATACGCG CGAATTCGC
pAT_4a
GATCCGCGTTAACGCGAATTCGCGTATACGCGATATCGCGTAATC GCGATTACGCGTTTTCGCG
pAT_4b
GATCCGGCAAATCGGCAATACGGCATAACGGCTAAACGGCTTTA CGGCTTATCGGCTATTCGGCATTTCGGCAATTCGGC
pAT_4d
GATCGCCGAAAAGCCGTAATGCCGATTAGCCGATATGCCGTATAG CCGAATTGCCGTTAAGCCG

In each case the strand shown is the one seen by labelling the fragment at the 3′-end of the *Hin*dIII site.

DNase I footprinting

The radiolabelled DNA fragments (1.5 µl) containing the target sites were mixed with 1.5 µl of the minor groove-binding ligand, dissolved in 10 mM Tris–HCl pH 7.5 containing 10 mM NaCl. The complexes were allowed to equilibrate at room temperature for at least 30 min. The samples were then digested with 2 µl DNase I (0.01 U/ml dissolved in 1 mM $MgCl_2$, 1 mM $MnCl_2$, 20 mM NaCl). The digestion was stopped after 1 min by the addition of 3.5 µl DNase I stop solution (80% formamide containing 10 mM EDTA). Samples were heated at 100°C for 3 min before electrophoresis.

Gel electrophoresis

Products of the digestion were separated on 10% polyacrylamide gels, containing 8 M urea and run at 1500 V for ~2 h. For the

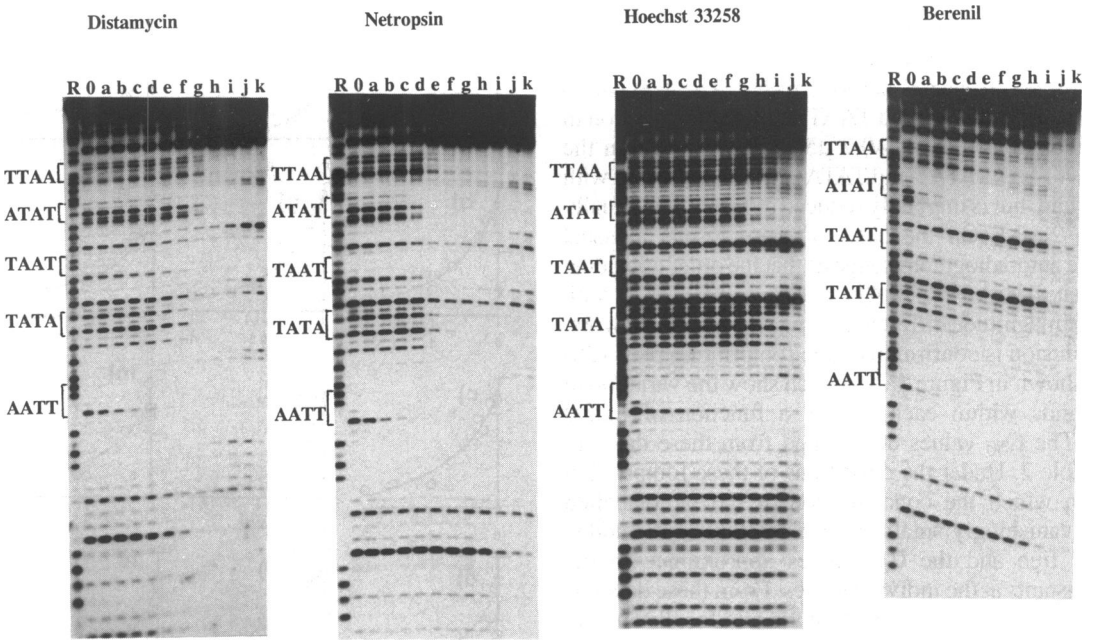

Figure 2. DNase I digestion of the fragment derived from pAAD1 in the presence of increasing concentrations of distamycin, netropsin, Hoechst 33258 and berenil. For distamycin and netropsin lanes a–k correspond to concentrations of 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1, 2, 5, 10 and 20 μM. For Hoechst 33258 and berenil lanes a–k correspond to concentrations of 0.1, 0.2, 0.3, 0.5, 0.7, 1, 2, 5, 10, 20 and 50 μM. Tracks labelled 'R' are Maxam–Gilbert formic acid–piperidine markers specific for G+A. Tracks labelled '0' are control lanes showing digestion in the absence of added ligands. The positions of the various $(A/T)_4$ blocks are indicated at the side of each panel.

fragment from pAAD1, which contains repeated blocks of eight consecutive GC residues, the gel also contained 20% formamide to prevent unacceptable band compressions in these regions. After running the gels were fixed in 10% (v/v) acetic acid before drying at 80°C and subjecting to autoradiography at –70°C using an intensifying screen. Bands were assigned by comparison with Maxam–Gilbert sequencing lanes specific for guanine or purines. The intensity of bands in the autoradiographs was estimated by scanning with a Hoefer GS365 microdensitometer and were quantified using the manufacturer's software. C_{50} values, indicating the ligand concentration reducing the intensity of a band by 50%, derived from these data, were estimated by fitting simple binding curves to plots of band intensity against ligand concentration. For this analysis we chose one band in each site which was well resolved and cut well in the control. The bands used for analysing the data from pAAD1 corresponded to cleavage of the following bonds: AATTpC, TATApC, TAATpC, ATApT and TTAApC. At some of the weaker binding sites there is an apparent increase in band intensity at the lower ligand concentrations. This is probably due to redistribution of the DNase I molecules away from the tightest binding sites, causing relatively more cleavage at these weaker sites which are not occupied by the ligand. Points on the rising portion of these curves were not included when calculating the C_{50} values.

RESULTS

Although the interaction of minor groove binding ligands has previously been studied by quantitative footprinting, there have been no reports which systematically examine their affinities for different arrangement of AT-residues. In the following studies we have used semi-quantitative DNase I footprinting on synthetic DNA fragments, each of which contains several different arrangements of $(A/T)_4$ residues located in identical sequence environments. Since several different $(A/T)_4$ sites are contained within each fragment, it is possible to estimate the relative affinity for each site by comparing the concentration dependence of the footprints. DNase I and hydroxyl radical cleavage patterns of several fragments, each containing a single $(A/T)_4$ site, has previously shown that the precise arrangement of AT-residues affects the width of the minor groove (35), a parameter which is likely to affect the interaction with these ligands.

Figure 2 presents DNase I footprinting patterns for distamycin, netropsin, Hoechst 33258 and berenil on the DNA fragment derived from pAAD1. This DNA fragment contains five blocks of AT-base pairs (AATT, TATA, TAAT, ATAT and TTAA) each separated by six GC base pairs (CGCGCG). These are arranged so that each AT-block is flanked by a 5′-G and 3′-C. As previously noted with shorter DNA fragments (35), the DNase I cleavage pattern is not even; cleavage of each dinucleotide step is not constant but depends on the surrounding bases. This is clearest for the ApT step, which is cut well in ATAT and TATA, but poorly in AATT and TAAT. As expected the ligands protect from DNase I cleavage around all these AT-rich sequences. Since DNase I is such a large molecule, which generates an uneven cleavage pattern, it is not a good probe for estimating the exact location and size of each binding site. Nonetheless it is an excellent tool for estimating the relative affinity of the ligands at each AT-block. Other probes, such as hydroxyl radicals, will be necessary to assign the exact position of the ligands at each site. Examination of Figure 2 reveals that the appearance of a footprint at each site has a different concentration dependence. Looking first at the data

for distamycin it can be seen that AATT is the best binding site; bands within this region show a marked reduction in intensity at concentrations as low as 0.2 μM (lane c), at which concentration cleavage at the other AT-sites is hardly affected. The next site to be affected by the ligand is around TAAT, showing a reduction in cleavage efficiency between 0.5–0.7 μM. Cleavage within the other three sites (TTAA, ATAT and TATA) is hardly affected with 0.7 μM distamycin, but is markedly reduced at 1 μM and virtually abolished with 2 μM. On the basis of these results a visual inspection of the autoradiographs suggests that the order of binding for these five sites is AATT>TAAT>TTAA=ATAT=TATA, with about an order of magnitude difference between the best and worst sites. This conclusion is confirmed by the footprinting plots (26) for these data, shown in Figure 3a–e, which show the variation in cleavage of bands within each site as a function of ligand concentration. The C_{50} values determined from these data are presented in Table 2. Under the conditions of these footprinting experiments (in which the concentration of the radiolabelled target DNA is vanishingly small) a large fraction of the added ligand remains free and the C_{50} values approximate to the dissociation constants at the individual sites. From these data it is clear that AATT represents the best binding site for distamycin with a C_{50} of ~0.15 μM compared with ~0.9 μM for ATAT, TATA, TAAT and TTAA, which are the same within experimental error.

Table 2. C_{50} values for the interaction of distamycin, netropsin, Hoechst 33258 and berenil with different $(A/T)_4$ sites in the fragment derived from pAAD1

	Distamycin	Netropsin	Hoechst 33258	Berenil
AATT	0.148±0.008	0.086±0.015	0.036±0.011	0.039±0.005
TAAT	0.71±0.37	0.31±0.07	1.97±0.81	0.103±0.026
ATAT	0.97±0.36	0.21±0.06	4.52±2.97	0.027±0.007
TATA	0.96±0.61	1.25±1.01	12.8±7.3	0.36±0.11
TTAA	0.89±0.30	0.31±0.09	15.1±11.6	0.45±0.07

C_{50} is the ligand concentration (μM) which reduces the intensity of cleavage by 50% and was determined as described in Materials and Methods.

Comparing these with the patterns obtained in the presence of netropsin, it can be seen that the ranking order of the sites is very similar. AATT seems to be the best site, since the footprint around this region appears at the lowest ligand concentration (0.2–0.3 μM). Bands within TAAT, TTAA and ATAT each disappear around 0.3 μM. In contrast the cleavage pattern at TATA is still evident with 0.3 μM ligand and requires higher ligand concentrations to abolish the pattern. This order of binding preference is confirmed by the footprinting plots, shown in Figure 3f–j, and the C_{50} values presented in Table 2. It therefore appears that the binding preference for netropsin is AATT>TAAT=TTAA=ATAT>TATA, though the difference in affinity between the best and worst sites is less pronounced than for distamycin.

Hoechst 33258 produces a similar pattern of selectivity between the five sites. However in this instance the preference for AATT is even more pronounced; the intensity of bands in this site is attenuated even at the lowest concentration (0.1 μM). In contrast TAAT and TATA are not affected until 2–5 μM, while TTAA and TATA require even higher concentrations of Hoechst

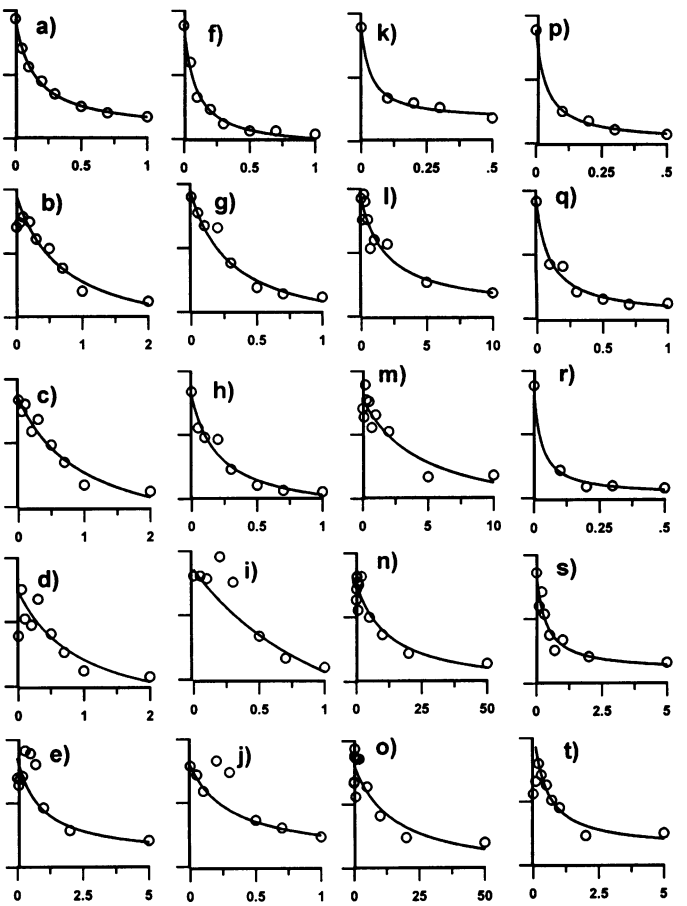

Figure 3. Footprinting plots showing intensity of bands within each AT-site as a function of ligand concentration (26). (a–e) distamycin; (f–j) netropsin; (k-o) Hoechst 33258; (p–t) berenil. (a,f,k,p) AATT; (b,g,l,q) TAAT; (c,h,m,r) ATAT; (d,i,n,s) TATA; (e,j,o,t) TTAA. In each case the ordinate shows the band intensity, expressed on an arbitrary scale. The abscissa indicates the ligand concentration (μM). Note that the scales are not the same for each plot.

33258 to produce clear footprints. This is confirmed by the footprinting plots (Fig. 3k–o) and the C_{50} values presented in Table 2. From these data it is clear that AATT is the best site, TAAT and ATAT are weaker sites which are the same within experimental error, while TATA and TTAA are weaker still; the order of binding for Hoechst 33258 is AATT>TAAT=ATAT>TTAA=TATA, similar to that for the other ligands with ~100-fold difference in affinity between the best (AATT) and worst (TATA and TTAA) sites.

In the presence of berenil, the footprints at AATT, ATAT and TAAT appear at similar low concentrations; bands in these sites are markedly reduced at the lowest ligand concentration (0.1 μM). Bands in TAAT persist to slightly higher concentrations than AATT and ATAT. In contrast the footprints at TATA and TTAA require somewhat higher ligand concentrations. This is shown in the footprinting plots (Fig. 3p–t) and the C_{50} values presented in Table 2, suggesting that the order of binding preference is AATT=ATAT>TAAT>TTAA=TATA.

These results reveal that the minor groove binding ligands do not bind to all AT-sites with the same affinity and suggest that the preferred order of A and T may be subtly different for each ligand. However it should be noted, amongst the five $(AT)_4$ blocks

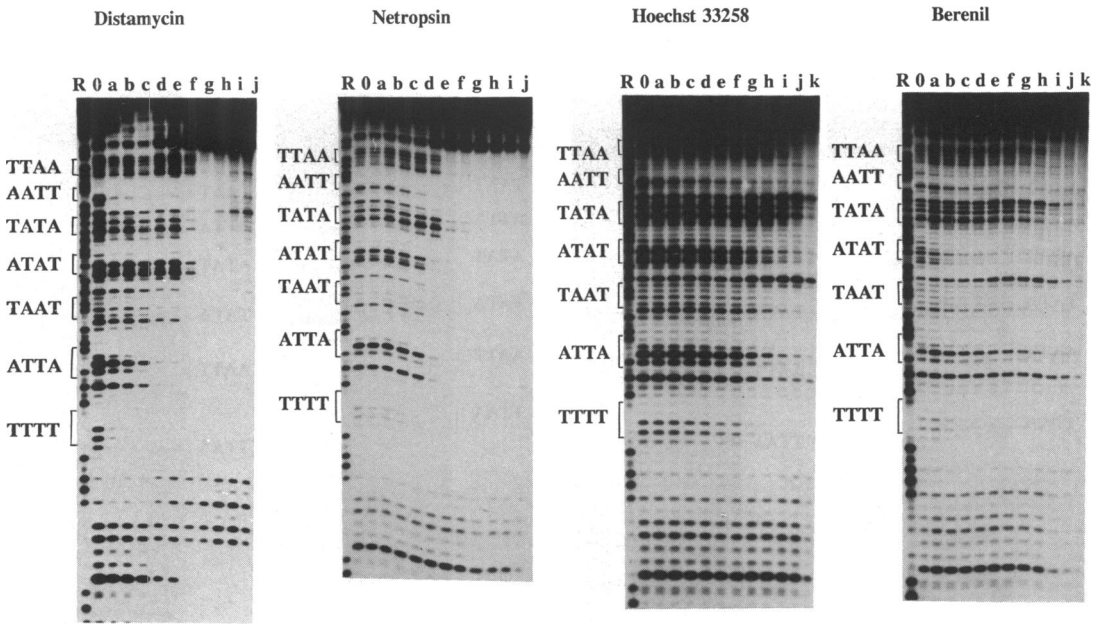

Figure 4. DNase I footprinting of the DNA fragment derived from pAT_4a in the presence of the four minor groove binding ligands. For distamycin lanes a–j correspond to concentrations of 0.1, 0.2, 0.3, 0.5, 0.7 , 1, 2, 5, 10 and 20 μM. For netropsin lanes a–k correspond to concentrations of 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1, 2, 5, 10 and 20 μM. For Hoechst 33258 and berenil lanes a–k correspond to concentrations of 0.1, 0.2, 0.3, 0.5, 0.7, 1, 2, 5, 10, 20 and 50 μM. Tracks labelled 'R' are Maxam–Gilbert formic acid–piperidine markers specific for G+A. Tracks labelled '0' are control lanes showing digestion in the absence of added ligands. The positions of the various $(A/T)_4$ blocks are indicated at the side of each panel.

present in this fragment, AATT is one of the best binding sites, while binding to TTAA and TATA is consistently much weaker. We have confirmed and extended these results by performing footprinting experiments on a similar DNA fragment, derived from pAT_4a, containing the same AT sites, together with TTTT, in which the AT-blocks are separated by only four GC base pairs (CGCG). The sequence of the relevant portion of this fragment is shown in Table 1. The DNase I digestion patterns are presented in Figure 4. Looking at the results for distamycin it can be seen that clear footprints are produced around TTTT and AATT at low ligand concentrations (0.1 μM). Footprints at the two ATTA/TAAT sites require slightly higher ligand concentrations (0.3–0.5 μM), while TATA and ATAT do not yield footprints until 1 μM ligand (lane f). TTAA can be seen to require the highest distamycin concentration since the footprint is not complete until 2 μM (lane g). Within this fragment the binding preference for distamycin is therefore AATT=TTTT>ATTA/TAAT>ATAT=TATA>TTAA. With netropsin the first footprints to appear are at AATT and TTTT (lane d); ATTA and ATAT require slightly higher concentrations (0.5 μM, lane e), while TTAA and TATA only produce footprints between 1 and 2 μM. With low concentrations of Hoechst 33258 (0.5–1 μM) clear footprints can be seen around TTTT and AATT; ATTA/TAAT and ATAT require slightly higher concentrations (2–10 μM). TTAA and TATA present poorer binding sites and only yield footprints at the highest concentrations (50 μM, lane k). A similar effect is seen with berenil for which footprints at TTAA and TATA require much higher concentrations than the other sites. The rank order of binding sites for these four ligands on this fragment is TTTT=AATT>ATTA/TAAT=ATAT>TATA=TTAA.

As expected the rank order of the various binding sites is very similar for the fragments derived from pAAD1 and pAT_4a, which only differ in the number of GC residues separating each $(A/T)_4$ site, and confirm that AATT and TTTT are good binding sites for all these minor groove binding ligands. For each ligand, footprints at TATA and TTAA require much higher concentrations, indicating that these are weaker binding sites. The discrimination between the best and worst sites appears to be greatest for Hoechst 33258 and berenil.

Since each of these AT blocks is located within the sequence context $G(A/T)_4C$ it is possible that the differences in binding to the various sites may also be influenced by the order of the flanking GC base pairs. We have tested this possibility by examining the interaction of these ligands with a fragment derived from pAT_4d (sequence shown in Table 1), in which the blocks of AT residues are contained within the sequence $G(A/T)_4G$, with each site separated by the sequence GCCG. The results of these footprinting experiments are presented in Figure 5. Looking first at the results for distamycin it can be seen that, on increasing the ligand concentration, footprints first appear at AAAA and AATT (0.1–0.2 μM), as noted for the previous fragments. The ATTA/TAAT sites yield footprints at slightly higher concentrations (0.3–0.5 μM), while TTAA, TATA and ATAT require the highest distamycin concentrations (0.7–2 μM). In contrast to the results described above for pAAD1 and pAT_4a, ATAT represents one of the weakest binding sites, suggesting that flanking sequences do affect the ability of the ligand to distinguish between various combinations of A and T bases. With netropsin the different footprints all appear at similar ligand concentrations (0.5–0.7 μM, lanes e–f), though the changes with AATT and AAAA appear one lane earlier. Within this sequence context it appears that netropsin exhibits far less sequence discrimination than distamycin. The results for Hoechst 33258 confirm that AATT is an exceptionally good site, showing reduced

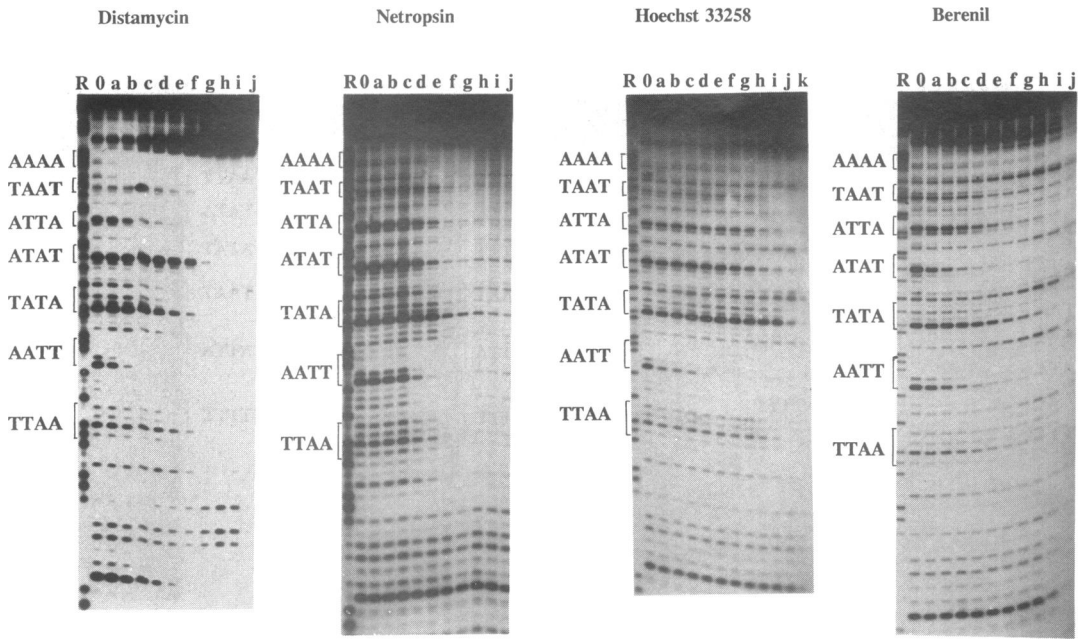

Figure 5. DNase I footprinting of the fragment derived from pAT_4d in the presence of the four minor groove binding ligands. For distamycin lanes a–j correspond to concentrations of 0.1, 0.2, 0.3, 0.5, 0.7, 1, 2, 5, 10 and 20 μM. For netropsin and berenil lanes a–j correspond to concentrations of 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1, 2, 5 and 10 μM. For Hoechst 33258 lanes a–k correspond to concentrations of 0.1, 0.2, 0.3, 0.5, 0.7, 1, 2, 5, 10, 20 and 50 μM. Tracks labelled 'R' are Maxam–Gilbert formic acid–piperidine markers specific for G+A. Tracks labelled '0' are control lanes showing digestion in the absence of added ligands. The positions of the various $(A/T)_4$ blocks are indicated at the side of each panel.

band intensity at the lowest ligand concentration (0.1 μM). AAAA also appears to be a good site with the footprint appearing between 0.7–1 μM. This preference of Hoechst for AATT over AAAA, is not observed with the other ligands and will be considered further in the Discussion. All the other sites produce footprints at concentrations between 5 and 10 μM except TATA which is only affected at the highest ligand concentration (50 μM, lane k). Once again Hoechst 33258 displays a large discrimination between the various sites, with a 50-fold concentration difference between the best (AATT) and worst (TATA) sites. The interaction of berenil with this sequence shows footprints appearing at ~0.2 μM (lane c) for AAAA, ATAT, AATT and TTAA. Higher concentrations are required at ATTA/TAAT (0.5 μM), while TATA represents the weakest binding site (2–5 μM). The rank order of binding of these ligands to the various sites on pAT_4d is summarized in Table 3.

Table 3. Rank order of binding preferences for the four ligands on the DNA fragment derived from pAT_4d ($GCCG(A/T)_4GCCG$)

Distamycin	AAAA=AATT>ATTA/TAAT>TTAA=TATA>ATAT
Netropsin	AAAA=AATT>ATTA/TAAT=TTAA=TATA=ATAT
Hoechst	AATT>AAAA>ATTA/TAAT=TTAA=ATAT>TATA
Berenil	AAAA=ATAT=AATT=TTAA>ATTA/TAAT>TATA

We have also examined the interaction of these ligands with other sites containing four AT base pairs, in asymmetrical arrangements, using the fragment derived from pAT_4b (sequence shown in Table 1). In this fragment the sites are separated by the sequence CGGC and is therefore similar to pAT_4d, in which the sites are separated by GCCG. This fragment also contains AATT as a good binding site against which the other sites can be compared. The results are shown in Figure 6. With distamycin AATT is by far the best site as expected, producing a footprint at ~0.3 μM; all the other sites require higher ligand concentrations (sites towards the top of the gel are not sufficiently well resolved to comment on their affinity). The sites in the centre of the fragment (TTAT, TTTA, TAAA and ATAA) require higher ligand concentrations than those towards the bottom (ATTT and TATT). A similar effect is evident with netropsin, Hoechst 33258 and berenil, for which AATT is the best site, while AAAT, TATT and ATTT produce footprints at lower concentrations than the remaining sites. With netropsin, for which bands at the top of the gel are better resolved, it can be seen that AAAT is also a good binding site, which produces at footprint at the same concentration as ATTT. However, AATA is amongst the poorer sites and requires higher netropsin concentrations than TTAT, located towards the bottom. The difference between AATA and TATT must arise from their difference sequence contexts (GAATAG CTATTC as opposed to CAATAC GTATTG) and will be considered further in the Discussion.

DISCUSSION

The results presented in this paper confirm that all these minor groove binding ligands interact with AT-rich sequences, but demonstrate that the affinity varies widely between different arrangements of AT base pairs, and that this is not the same for each ligand. For sites containing four consecutive AT residues AATT and AAAA consistently represent the best sites, while TATA and TTAA have a much lower binding affinity. In this

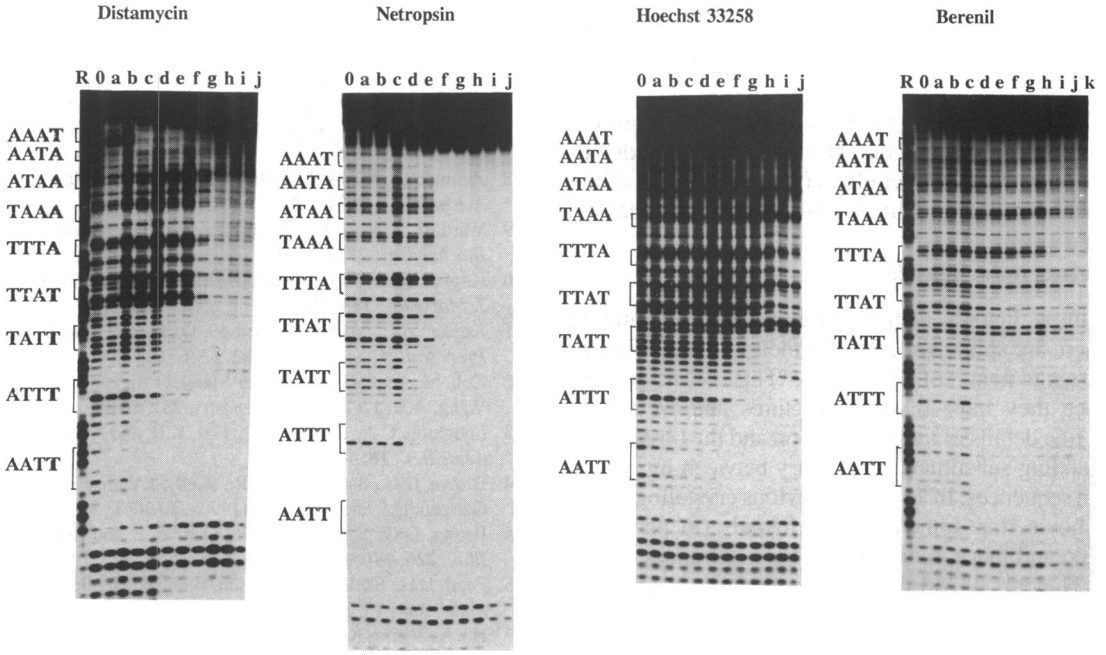

Figure 6. DNase I footprinting of the fragment derived from pAT_4b in the presence of the four minor groove binding ligands. For distamycin lanes a–j correspond to concentrations of 0.1, 0.2, 0.3, 0.5, 0.7, 1, 2, 5, 10 and 20 μM. For netropsin lanes a–j correspond to concentrations of 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1, 2, 5 and 10 μM. For Hoechst 33258 and berenil lanes a–k correspond to concentrations of 0.1, 0.2, 0.3, 0.5, 0.7, 1, 2, 5, 10, 20 and 50 μM. Tracks labelled 'R' are Maxam–Gilbert formic acid–piperidine markers specific for G+A. Tracks labelled '0' are control lanes showing digestion in the absence of added ligands. The positions of the various $(A/T)_4$ blocks are indicated at the side of each panel.

Discussion we will first consider differences between the individual sequences which are common to all the ligands, before discussing variations in the binding preferences of the individual compounds.

Comparison of binding sites with four consecutive AT base pairs

The results presented in this paper demonstrate that AAAA and AATT consistently represent the best binding sites for these minor groove binding ligands. Several crystal structures have shown that the minor groove narrows in the centre of the CGCGAATTCGCG (27,28). Amongst the other symmetrical sites TTAA and TATA represent poor binding sites. Although there is less structural information for sequences containing these arrangements of AT base pairs, TTAA possesses a wider minor groove (29) and does not contain an ordered array of bound water molecules (30). These differences can be attributed to the small twist angle at the TpA step, which has also been observed in sequences containing TATA and ATAT (31,32). The wider minor groove at these sequences will result in weaker van der Waals interaction thereby decreasing the enthalpy factor of the interaction. In addition, since these ligands replace a spine of bound water molecules, sequences which lack such an ordered array will have a lower entropy of binding and therefore have a lower free energy of interaction. This is consistent with thermodynamic studies which have shown that the interaction of netropsin with poly(dA) poly(dT) is entropy driven, whereas enthalpy is the more important factor for the interaction with poly[d(A–T)] (33,34). In general, ATAT is a better binding site than TTAA, in the fragments from pAAD1 and pAT_4a [i.e. $G(A/T)_4C$]. These sites each contain a single TpA step, which is centrally located, so that differences in their binding affinity must reflect variations in minor groove width, suggesting that ATAT is narrower than TTAA, as previously suggested (35). However, it has been suggested that tetranucleotides alone are not sufficient to define local DNA structure (29,36) and that the surrounding base steps may influence the binding. In this regard it should be noted that in these fragments ATAT is flanked by RR (YY) steps, whereas TTAA is flanked by on either side by RY. One might therefore suggest that flanking RR steps promote the formation of a narrower minor groove in the central region than RY. The effect of flanking sequence can also be seen for the interaction of netropsin with pAT_4d for which CTATTC presents a better binding site than CAATAC. These two sites possess the same arrangement of AT residues, but in opposite orientations. The former is flanked by YY steps whereas the latter has 5′-YR and 3′-RY. If a general property of YR steps (not just TpA) is to widen the minor groove, then the difference in binding can be adequately explained. A similar effect is apparent when considering the affinity of these ligands for ATAT, relative to the other sites. In pAT_4a and pAAD1 this site is flanked by RR (YY steps) and is a better binding site than TTAA and TATA. In contrast with pAT_4d, for which it is flanked by 5′-GA and 3′-TG, this is one of the worst binding sites for netropsin, distamycin and Hoechst. In addition, although GAATTG is the best binding site on this fragment, the discrimination in favour of this site is less marked than in pAAD1 and pAT_4a, probably because of the 3′-YR step.

It is also informative to consider the ranking order of the various asymmetric sites in pAT_4b [$C(AT)_4C$], for which ATTT and TATT are better than TTAT, TTTA, TAAA, ATAA and

AATA. Those tetranucleotides with an adenine at the 5′-end (ATTT, ATAA and AATA) are preceded by a YR step (CpA) which, according to the argument described above, should display weaker binding affinities. Since ATAA and AATA contain central TpA steps it is not surprising that these represent weaker binding sites. For ATTT the AT step is at the 5′-edge of the binding site and so has smaller effect on binding. Each of the other sites contain TpA step within the block of AT residues.

Differences between the ligands

Although all these ligands display similar sequence binding preferences, there are subtle variations in their relative affinities at different arrangements of AT bases. This ought not be surprising, since they have diverse structures and we would expect that the fine details of their interaction and the positioning of hydrogen bonding substituents must vary between ligands as well as between sequences. In addition, previous crystallographic studies have shown that some of these ligands (in particular Hoechst 33258) can bind in more than one configuration. In the present results Hoechst 33258 displays the greatest discrimination between different arrangements of AT residues, with a 50-fold difference in the concentration required to generate a footprint at AATT as compared with TATA and TTAA. It may be significant that this compound has a higher affinity for AATT than the other ligands yet requires higher concentrations to produce footprints at the other sites. It might therefore be argued that novel compounds with altered and enhanced sequence selectivity should be based on a bisbenzimidazole structure, rather than the polypyrrole nucleus which is more normally employed.

Although it is virtually impossible to assign accurate positions for the various ligands, based on DNase I footprinting data, the autoradiographs indicate that the ligands may not always bind to the same portion of the binding site. An example of this can be seen in the footprints produced at the TATA site in pAT_4d [G(A/T)$_4$C] (Fig. 5). Looking at the cleavage pattern in this region, consisting of three strong bands in the control, it can be seen that, at low concentrations, distamycin and netropsin reduce the intensity of the upper two, while berenil and Hoechst 33258 affect the lower two bands. This is consistent with the crystal structure data for CGCGATATCGCG which show netropsin bound centrally across the ATAT section (12), while Hoechst binds to GATA (37).

ACKNOWLEDGEMENTS

This work was supported by grants from the Cancer Research Campaign, the Medical Research Council, the Royal Society and the Nuffield Foundation.

REFERENCES

1 Kopka, M.L. and Larsen, T.A. (1992) In Propst, C.L. and Perun, T.J. (eds) *Nucleic Acids Targeted Drug Design.* Marcel Dekker Inc., New York. pp. 303–374.
2 Lown, J.W. (1993) In Neidle, S. and Waring, M.J. (eds) *Molecular Aspects of Anticancer Drug–DNA Interactions.* Macmillan, London, vol. 1. pp. 322–355.
3 Van Dyke, M.W., Hertzberg, R.P. and Dervan, P.B. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 5470–5474.
4 Fox, K.R. and Waring, M.J. (1984) *Nucleic Acids. Res.*, **12**, 9271–9285.
5 Portugal, J. and Waring, M.J. (1987) *FEBS Lett.*, **225**, 195–200.
6 Portugal, J. and Waring, M.J. (1987) *Eur. J. Biochem.*, **167**, 281–289.
7 Churchill, M.E.A., Hayes, J.J. and Tullius, T.D. (1990) *Biochemistry*, **29**, 6043–6050.
8 Zimmer, C., Marck, C., Schneider, C. and Guschlbauer, W.(1979) *Nucleic Acids Res.*, **6**, 2831–2837.
9 Ward, B. Rehfuss, R., Goodisman, J. and Dabrowiak, J.D. (1988) *Biochemistry*, **27**, 1198–1205.
10 Kopka, M.L., Yoon, D., Goodsell, D., Pjura, P. and Dickerson, R.E. (1985) *J. Mol. Biol.*, **183**, 553–563.
11 Kopka, M.L., Yoon, D., Goodsell, D., Pjura, P. and Dickerson, R.E. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1376–1380.
12 Coll, M., Aymami, J., Van der Marel, G.A., van Boom, J.H., Rich, A. and Wang, A.H.-J. (1989) *Biochemistry*, **28**, 310–320.
13 Laughton, C.A., Jenkins, T.C., Fox, K.R. and Neidle, S. (1990) *Nucleic Acids Res.*, **18**, 4479–4488.
14 Brown, D.G., Sanderson, M.R., Skelly, J.V., Jenkins, T.C., Brown, T., Garman, D.I. and Neidle, S. (1990) *EMBO J.*, **9**, 1329–1334.
15 Brown, D.G., Sanderson, M.R., Garman E. and Neidle S. (1992) *J. Mol. Biol.*, **226**, 481–490.
16 Pearl, L.H., Skelly, J.V., Hudson, B.D. and Neidle, S. (1987) *Nucleic Acids. Res.*, **15**, 3469–3477.
17 Hu, S., Weisz, K., James, T.L. and Shafer, R.H. (1992) *Eur. J. Biochem.*, **204**, 31–38.
18 Harshmann, K.D. and Dervan, P.B. (1985) *Nucleic Acids Res.*, **13**, 4825–4835.
19 Portugal, J. and Waring, M.J. (1988) *Biochim. Biophys. Acta*, **949**, 158–168.
20 Pjura, P.E., Grzeskowiak, K. and Dickerson, R.E. (1987) *J. Mol. Biol.*, **197**, 257–271.
21 Teng, M.-K., Usman, N., Frederick, C.A. and Wang, A.H.-J. (1988) *Nucleic Acids Res.*, **16**, 2671–2690.
22 Quintana, J.R., Lipanov, A.A. and Dickerson, R.E. (1991) *Biochemistry*, **30**, 10 294–10 306.
23 Vega, M.C., Garci Saez, I., Aymami, J., Eritja, R. van der Marel, G.A., van Boom, J. H., Rich, A. and Coll, M. (1994) *Eur. J. Biochem.*, **222**, 721–726.
24 Spink, N., Brown, D.G. Skelly, J.V. and Neidle, S. (1994) *Nucleic Acids Res.*, **22**, 1607–1612.
25 Embrey, K.J., Searle, M.S. and Craik, D.J. (1993) *Eur. J. Biochem.*, **211**, 437–447.
26 Dabrowiak, J.C. and Goodisman, J. (1989) In N.R. Kallenbach, (ed.) *Chemistry and Physics of DNA–Ligand Interactions.* Adenine Press, New York. pp.143–174.
27 Dickerson, R.E. and Drew, H.R. (1981) *J. Mol. Biol.*, **149**, 761–786.
28 Fratini, A.V., Kopka, M.L., Drew, H.R. and Dickerson, R. E. (1982) *J. Biol. Chem.*, **257**, 14 686–14 707.
29 Quintana, J.R., Grzeskowiak, K. Yanagi, K. and Dickerson, R.E. (1992) *J. Mol. Biol.*, **225**, 379–395.
30 Liepinsh, E., Leupin, W. and Otting, G. (1994) *Nucleic Acids Res.*, **22**, 2249–2254.
31 Cheng, J.-W., Chou, S.-H., Salazar, M. and Reid, B.R. (1992) *J. Mol. Biol.*, **228**, 118–137.
32 Yoon, C., Prive, G.P., Goodsell, D.S. and Dickerson, R.E. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6332–6336.
33 Breslauer, K.J., Remeta, D.P., Chou, W.-Y., Ferrante, R., Curry, J., Zaunczkowsi, D., Snyder, J.G. and Marky, L.A. (1987) *Proc. Natl. Acad. Sci. USA*, **24**, 8922–8926.
34 Marky, L.A. and Breslauer, K.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4359–4363.
35 Fox, K.R. (1992). *Nucleic Acids Res.*, **20**, 6487–6493.
36 Leonard, G.A. and Hunter, W.N. (1993) *J. Mol. Biol.*, **234**, 198–208.
37 Carrondo, M.A., Coll, M., Aymami, J., Wang, A.H.-J., van der Marel, G.A., van Boom, J.H. and Rich, A. (1989) *Biochemistry*, **28**, 7849–7859.