Variability in DNA minor groove width recognised by ligand binding: the crystal structure of a bis-benzimidazole compound bound to the DNA duplex d(CGCGAATTCGCG)₂

Alexis A. Wood, Christine M. Nunn, Agnieszka Czarny¹, David W. Boykin¹ and Stephen Neidle*

The CRC Biomolecular Structure Unit, Institute of Cancer Research, Cotswold Road, Sutton, Surrey SM2 5NG, UK and ¹Department of Chemistry and Center for Biotechnology and Drug Design, Georgia State University, Atlanta, GA 30303, USA

Received June 21, 1995; Revised and Accepted August 11, 1995

PDB identity code 109d

ABSTRACT

An analogue of the DNA-binding compound Hoechst 33258, in which the piperazine ring has been replaced by an imidazoline group, has been cocrystallized with the dodecanucleotide sequence d(CGCGAATTCGCG)₂. The structure has been solved by X-ray diffraction analysis and has been refined to an R-factor of 19.7% at a resolution of 2.0 Å. The ligand is found to bind in the minor groove, at the central four AATT base pairs of the B-DNA double helix, with the involvement of a number of van der Waals contacts and hydrogen bonds. There are significant differences in minor groove width for the two compounds, along much of the AATT region. In particular this structure shows a narrower groove at the 3' end of the binding site consistent with the narrower cross-section of the imidazole group compared with the piperazine ring of Hoechst 33258 and therefore a smaller perturbation in groove width. The higher binding affinity to DNA shown by this analogue compared with Hoechst 33258 itself, has been rationalised in terms of these differences.

INTRODUCTION

It is well established that drugs such as netropsin, distamycin, berenil, pentamidine and Hoechst 33258 bind non-covalently to AT regions of duplex DNA (1–3) and that their modes of biological action in terms of antiviral, antimicrobial or anticancer activity, involves DNA binding as the critical step, possibly followed by interference with the function of a DNA topoisomerase (4–9). These compounds and their analogues generally possess several common structural features; cationic charge and concave shape along one long edge together with coplanarity of aromatic groups when viewed along this edge.

Figure 1. The structures of Hoechst 33258 (1) and of 5-(2-(4-hydroxyphenyl)-5-benzimidazolyl)benzimidazole (2).

An extensive body of data from, in particular, X-ray crystallography (e.g. refs 10–15) as well as NMR studies (e.g. refs 16–18) have established that these molecules bind in the minor groove of B-type DNA double helices, largely in AT regions. Almost all of these studies have been performed on drug complexes with DNA dodecamer duplexes, using sequences of the type d(CGCXXXXXXGCG)₂, where XXXXXX represents a run of AT-rich base pairs. In the most commonly studied system, XXXXXX is GAATTC. These structural studies have shown that the drugs are stabilised in the minor groove by arrays of close non-bonded van der Waals interactions involving hydrogen atoms on the floor and walls of the groove (19). Direct hydrogen bonding between donor atoms on the drug and acceptor atoms on base pair edges has been frequently, though not invariably observed (e.g. see ref. 15). The consequent suggestion (20) that such drug–DNA hydrogen bonding is a secondary feature of the binding, rather than a determinant of sequence selectivity, has

* To whom correspondence should be addressed

Figure 2. The molecule of **2** superimposed on the electron density calculated from a $2F_o - F_c$ difference map, drawn at a contour level of 1σ .

given added credence by observations that patterns of hydrogen bonding for a given drug can markedly alter with even small changes in DNA sequence (14,21,22).

The design of molecules with altered sequence recognition properties, that is those that have the ability to actively recognise GC base pairs, was initially based on considerations of hydrogen bonding differences between AT and GC base pairs (23). These led to molecules that could tolerate GC base pairs, but were generally unable to show specificity to them. The more recent design of dimeric distamycin or peptide-like molecules has largely overcome this problem (24–27). Such molecules take advantage of the sequence dependence of groove width rather than of hydrogen bonding alone and in particular the enlarged width in GC regions compared to AT ones.

The bis-benzimidazole compound Hoechst 33258 (**1**; Fig. 1) has long been used in chromosome staining. It has been the starting point for a number of recent drug design studies aimed at developing analogues with, in particular, anticancer activity (5,7,28,29). Hoechst 33258 is an exception to the general category of non-covalent groove binders in that it can bind to sequences that do not solely contain AT base pairs; it can have a GC base pair at one end of its (4–5 bp) binding site, as shown by DNA footprinting (30,31). Nuclear magnetic resonance studies (32,33) have concurred with this, showing that the phenyl and benzimidazole rings of the drug are in the AT region of the minor groove and that the non-planar piperazine ring is in a GC region. X-ray crystallographic studies have been reported for Hoechst 33258 complexed to several dodeca-oligonucleotides. One study, with the sequence $d(\text{CGCGAATTCGCG})_2$, showed the drug to be bound at the sequence 5'-ATTC (34), whereas a second analysis (35) located it in the sequence 5'-AATT, that is those

displaced by 1 base pair, in agreement with a subsequent low temperature study (36). X-ray analysis of the drug complexed to the sequence $d(\text{CGCGATATCGCG})_2$ (37) found the drug at the site 5'-GATA (equivalent to TATC). X-ray analyses of Hoechst 33258 complexed with the sequence $d(\text{CGCAAATTTGCG})_2$ have located it in a unique orientation in the 5 bp sequence 5'-ATTTG, with the piperazine ring of the drug reaching the edge of the GC base pair, where the minor groove is sufficiently wide to accommodate this non-planar ring (22,38).

This paper reports the crystal structure of the complex formed between the self complementary dodecanucleotide sequence $d(\text{CGCGAATTCGCG})_2$ and an analogue (**2**) of Hoechst 33258 (Fig. 1), in which an imidazole ring replaces the piperazine group, in order to observe the structural consequences of this substitution and so gain further insight into the factors responsible for binding affinity and specificity. We have recently shown (39) that compound **2** binds to the dodecamer sequence with a ΔT_m of 23.9°C compared with 17.9°C for Hoechst 33258 itself and has a binding constant 5-fold greater than that of Hoechst.

MATERIALS AND METHODS

Synthesis and crystallisation

The DNA dodecamer $d(\text{CGCGAATTCGCG})_2$ was purchased from the Oswel DNA Service (University of Edinburgh) and annealed before use. Compound **2** was synthesised by standard methodology (39). Crystals of the complex were grown from 14 μl drops containing 3 μl of 3 mM dodecamer, 4 μl of 2-methylpentane-2,4-diol (MPD) at 35% v/v (MPD/30 mM sodium cacodylate buffer pH 7.0), 3 μl of 200 mM MgCl_2 , 2 μl

of 2.4 mM spermine HCl and 2 μ l of 9.8 mM compound **2** equilibrated against a 4 ml reservoir of 35% v/v MPD using the sitting drop method at 19°C. The crystals took 2 months to grow.

A single crystal (of dimensions $\sim 0.4 \times 0.3 \times 0.15$ mm) was mounted in a 0.5 mm Lindemann glass capillary along with a small amount of mother liquor. Diffraction data was collected to a maximum resolution of 2.0 Å using a Siemens–Xentronics multiwire area detector and an Enraf–Nonius rotating anode X-ray generator operated at 40 mA and 70 kV. A crystal–detector distance of 10 cm and a swing angle of 20° was used with χ fixed at 45° to obtain datasets for $\phi = 120^\circ$ and $\phi = 180^\circ$. The crystal was rotated through 100° in ω , collecting 180 s frames every 0.2°. Examination of the frames at regular intervals during the data collection revealed no detectable crystal decay. Data processing was performed using XENGEN 1.3. 11 176 reflections out of a possible 12 209 were collected to a resolution of 2.0 Å. Merging

of the data resulted in 4365 unique reflections out of a possible 5026 (87%) with a merging R-value of 3.1%.

The crystal was found to be in the space group $P2_12_12_1$ with cell dimensions $a = 24.59$, $b = 40.44$ and $c = 65.76$ Å. This indicated that the structure was isomorphous with the native dodecamer sequence and other drug-dodecamer structures. Accordingly, coordinates for the native structure (40) were used as a starting model for the refinement. This process was carried out with the program XPLOR v3.1 (41), beginning with rigid body refinement for data in the range 8–3.5 Å. The molecular graphics program 'O' (42) was used to check that the model fitted the difference electron density in all regions of the structure. Rigid group refinement continued with the DNA duplex divided into 24 groups. This eventually led to an R factor of 30.2% for all data to 2.0 Å. Further positional and temperature factor refinement

Figure 5. Superimposition of the present structure and the Hoechst LT one, demonstrating the overall similarity in position of the ligands in the minor groove and their differences of orientation.

followed and $F_o - F_c$ electron density maps were calculated that clearly indicated a large volume of continuous electron density in the minor groove of the DNA. An energy-minimised model of compound **2** was created using the program MacroModel (43) and was fitted into this density. In order that the conformation of the analogue would be dictated by the X-ray data and not the refinement procedure, no restraints were placed on rotation about bonds between the four rings of the ligand. To determine the orientation of compound **2** in the groove, positional refinement was carried out for both possibilities and clearly revealed the correct orientation by a significant difference in R factor (24.3 versus 27.0%). The correct orientation is shown in Figure 2. Partial charges for compound **2** were calculated using MOPAC/ESP (44). Positional and temperature factor refinement continued and a total of 90 water molecules and a magnesium ion were located in an $F_o - F_c$ electron density map and included in the structure. The final R factor was 19.7% for 4365 reflections with $F > 2\sigma F$ with mean rms deviations of 0.020 Å and 4.3° from ideality for bond lengths and angles, respectively. Final refined coordinates, together with observed and calculated structure factors have been deposited in the Brookhaven Protein Data Bank with the identity code 109d.

RESULTS AND DISCUSSION

The overall structure of the ligand/DNA complex

The ligand was observed to lie pseudo-symmetrically in the 5'-AATT region of the minor groove of the dodecamer duplex,

which adopts a B-DNA conformation (Fig. 3). The duplexes pack in the unit cell with the two terminal GC base pairs interacting with the next duplex via minor groove base pair interactions, as reported for the native sequence (40). The phenol group of compound **2** lies towards the C1•G24 end of the double helix. There are three hydrogen bonds between the inner-facing nitrogen atoms of the benzimidazole groups and the edges of the bases; two (2.7 and 3.1 Å) to the O2 atoms of Thy 20 and Thy 19 respectively and one (2.9 Å) to N3 of Ade 6 (Fig. 4). These hydrogen bonds were also found in the Hoechst 33258-dodecamer low-temperature (LT) structure (36). The Thy 7 O2 distance to ligand atom N3 is probably too long to be a true hydrogen bond (3.4 Å) though it is suggestive of a weak interaction. The three hydrogen bonds result in the ligand being held in the same four base pair binding site as in the low temperature Hoechst structure, and in a similar, though not identical orientation (Fig. 5).

There are a large number of close van der Waals contacts between the ligand and the floor and walls of the minor groove (Table 1). These reflect the close fit between the ligand and the minor groove along the whole of the binding site, which is maintained by rotations about the bonds connecting the four planar rings of **2**. The torsion angles between the rings (Table 2) of the ligand indicate that each successive planar ring is twisted so as to follow the right handed curvature of the minor groove of the helix, with a dihedral angle between the two terminal planar groups in the molecule of 46°. Comparison with values in other Hoechst structures shows that there is a significantly larger dihedral angle

Figure 6. Plot of minor groove widths calculated with inter-strand C1' atoms for the present structure, the Hoechst 33258-dodecamer LT structure and the native dodecamer.

between the two benzimidazole groups in the present structure. The plot of the minor groove widths of **2** and the low-temperature structure (Fig. 6) indicates that the former has a greater groove width in the region of the three phenol-benzimidazole-benzimidazole rings, implying that this is due to the larger twist in the bond connecting the two benzimidazole rings and thus to the conformation of the ligand. Further evidence for this can be seen by examining the minor groove width in the vicinity of the imidazole and piperazine groups. The narrower groove at the 3' end of the binding site for compound **2** (~1 Å) can be attributed to the slim imidazole group being considerably less 'bulky' than the saturated piperazine ring. The extra twist in the molecule is also believed to contribute to the slightly larger overall bend in the DNA helical axis of 11°, as compared with 8° in the Hoechst LT structure.

Helical parameters for this structure, Hoechst and the native dodecamer were computed using the CURVES program (45). A comparison of these shows that the most noticeable differences between the Hoechst LT and the present structure are for the base-base buckle for Ade 6•Thy 19 and Thy 7•Ade 18 and for propeller twist for the same base pairs (Fig. 7), where the values are markedly lower for the present structure, especially for the latter base pair. A•T base pairs are normally associated with high propeller twist and consequently a narrowing of minor groove width; in this region of the structure we observe the opening of the groove, due to the benzimidazole groups possessing a large dihedral angle between them, forcing the smaller propeller twist angle at Thy•Ade 18. The Ade 6•Thy 19 base pair is also involved in hydrogen bonding with the ligand, which may be an additional,

though probably secondary factor in producing the low propeller twists. There is a marked alternation in the values of interbase roll (Fig. 7) for base steps Thy 19-Ade 18 and Ade 18-Ade 17.

Table 1. Van der Waals contact distances (Å) between ligand **2** and the dodecamer

Ligand atom	DNA atom	Distance
H3	18 H4'	2.1
N2	9 H4'	2.6
H5	18 H4'	2.4
H2	9 O2	2.8
C2	18 H4'	3.0
H10	19 O2	2.5
H10	7 O2	2.6
N4	8 H4'	2.4
N5	7 H4'	2.8
H15	20 O2	2.4
H21	6 H2	2.0
H21	19 O2	2.6
H16	5 H2	2.0
H16	20 O2	2.5
H17	5 H2	2.3

Hydrogen atoms were not observed in the crystal structure. Instead their positions were generated using standard geometric considerations. These were used in the calculations of contact distances listed.

Table 2. Inter-ring torsion angles in °, defined in terms of atoms facing into the groove

Bond	Present structure	Hoechst 33258 LT structure ^a	Hoechst 33258 A3T3 structure ^b
Phenol-Benzimidazole	20	24	11
Benzimidazole-Benzimidazole	12	0	-2
Benzimidazole-Imidazole	12	-	-

^aQuintana *et al.*, 1991.

^bSpink *et al.*, 1994

Binding to DNA

The binding affinity shown by the ligand **2** for a given DNA sequence is determined by a number of factors, including hydrogen bonding, electrostatic interactions and ligand conformation. The present structure suggests that hydrogen bonding is not the dominating factor. Since the number (and type) of ligand-DNA hydrogen bonds is identical to that found for Hoechst 33258 in the low-temperature dodecamer structure and their geometric characteristics are similar, they cannot by themselves account for its experimentally-determined higher binding affinity (38). Instead the structure and conformation of the ligand is a key factor, with the slim imidazole group enables binding to the minor groove to

Figure 7. Plots of base buckle, propeller twist and roll angles for the present structure, calculated with CURVES.

take place with minimal perturbation of the groove width compared with Hoechst 33258 itself. The several structural studies of this ligand complexed to DNA all concur in finding that the piperazine ring is located in a region of wider groove width compared to that in the central region of the 5'-AATT sequence. The average values for the atomic temperature factor for the Hoechst 33258 and two molecules in the two complexes indicates the relative mobilities of the two ligands in their binding sites. The values for the two ligands are 30 and 26 Å², respectively, showing that compound 2 is less mobile than Hoechst 33258 in the AATT minor groove site, and by inference is more tightly bound, in accord with the biophysical data.

ACKNOWLEDGEMENTS

We are grateful to Peter Collins for his help with the data collection and Richard Lavery for supplying the CURVES program. This work has been supported by the Cancer Research Campaign (Grant SP1384 to SN) and NIH (grant AI-33363 to DWB and WD Wilson).

REFERENCES

- Zimmer, C. and Wähnert, U. (1986) *Prog. Biophys. Mol. Biol.*, **47**, 31–112.
- Dervan, P. B. (1986) *Science*, **232**, 464–471.
- Kopka, M. L. and Larsen, T. A. (1992) In Propst, C. L. and Perun, T. J. (eds) *Nucleic Acid Targeted Drug Design*, Marcel Dekker, Inc., New York, pp. 304–374.
- Cory, M., Tidwell, R. R. and Fairley, T. A. (1992) *J. Med. Chem.*, **35**, 431–438.
- Beerman, T. A., McHugh, M. M., Sigmund, R., Lown, J. W., Rao, K. E. and Bathini, Y. (1992) *Biochim. Biophys. Acta*, **1131**, 53–61.
- Chen, A. Y., Chiang, Y., Gatto, B. and Liu, L. F. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 8131–8135.
- Gravatt, G. L., Baguley, B. C., Wilson, W. R. and Denny, W. A. (1994) *J. Med. Chem.*, **37**, 4338–4345.
- Wong, S. S. C., Sturm, R. A., Michel, J., Zhang, X.-M., Danoy, P. A. C., McGregor, K., Jacobs, J. J., Kaushal, A., Dong, Y., Dunn, I. S. and Parsons, P. G. (1994) *Biochem. Pharmacol.*, **47**, 827–837.
- Boykin, D. W., Kumar, A., Spychala, J., Zhou, M., Lombardy, R. J., Wilson, W. D., Dykstra, C. C., Jones, S. K., Hall, J. E., Tidwell, R. R., Laughton, C., Nunn, C. M. and Neidle, S. (1995) *J. Med. Chem.*, **38**, 912–916.
- Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R. E. (1985) *J. Mol. Biol.*, **183**, 553–563.
- Coll, M., Frederick, C. A., Wang, A. H. -J. and Rich, A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8385–8389.
- Brown, D. G., Sanderson, M. R., Skelly, J. V., Jenkins, T. C., Brown, T., Garman, E., Stuart, D. I. and Neidle, S. (1990) *EMBO J.*, **9**, 1329–1334.
- Edwards, K. J., Jenkins, T. C. and Neidle, S. (1992) *Biochemistry*, **31**, 7104–7109.
- Tabernero, L., Verdager, N., Coll, M., Fita, I., van der Marel, G. A., van Boom, J. H., Rich, A. and Aymami, J. (1993) *Biochemistry*, **32**, 8403–8410.
- Nunn, C. M., Jenkins, T. C. and Neidle, S. (1993) *Biochemistry*, **32**, 1383–1384.
- Pelton, J. G. and Wemmer, D. E. (1990) *J. Biomol. Struct. Dynamics*, **8**, 81–97.
- Jenkins, T. C., Lane, A. N., Neidle, S. and Brown, D. G. (1993) *Eur. J. Biochem.*, **213**, 1175–1184.
- Blaskó, A., Browne, K. A. and Bruce, T. C. (1994) *J. Am. Chem. Soc.*, **116**, 3726–3737.
- Neidle, S. (1992) *FEBS Lett.*, **298**, 97–99.
- Nunn, C. M. and Neidle, S. (1995) *J. Med. Chem.*, **38**, 2317–2325.
- Brown, D. G., Sanderson, M. R., Garman, E. and Neidle, S. (1992) *J. Mol. Biol.*, **226**, 481–490.
- Spink, N., Brown, D. G., Skelly, J. V. and Neidle, S. (1994) *Nucleic Acids Res.*, **22**, 1607–1612.
- Lown, J. W. (1988) *Anti-Cancer Drug Des.*, **3**, 25–40.
- Dwyer, T. J., Geierstanger, B. H., Bathini, Y., Lown, J. W. and Wemmer, D. E. (1992) *J. Am. Chem. Soc.*, **114**, 5911–5919.
- Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E. and Dervan, P. B. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 7586–7590.
- Mrksich, M. and Dervan, P. B. (1993) *J. Am. Chem. Soc.*, **115**, 2572–2576.
- Geierstanger, B. H., Jacobsen, J. P., Mrksich, M., Dervan, P. B. and Wemmer, D. E. (1994) *Biochemistry*, **33**, 3055–3062.
- Sun, Q., Gatto, B., Yu, C., Liu, A., Liu, L. F. and LaVoie, E. J. (1994) *Bioorg. Med. Chem. Lett.*, **4**, 2871–2876.
- Gupta, R., Wang, H., Huang, L. and Lown, J. W. (1995) *Anti-Cancer Drug Des.*, **10**, 25–41.
- Harshman, K. D. and Dervan, P. B. (1985) *Nucleic Acids Res.*, **13**, 4825–4835.
- Portugal, J. and Waring, M. J. (1988) *Biochim. Biophys. Acta*, **949**, 158–168.
- Parkinson, J. A., Barber, J., Douglas, K. T., Rosamond, J. and Sharples, D. (1990) *Biochemistry*, **29**, 10181–10190.
- Fede, A., Labhardt, A., Bannwarth, W. and Leupin, W. (1991) *Biochemistry*, **30**, 11377–11388.
- Pjura, P. E., Grzeskowiak, K. and Dickerson, R. E. (1987) *J. Mol. Biol.*, **197**, 257–271.
- Teng, M., Usman, N., Frederick, C. A. and Wang, A. H.-J. (1988) *Nucleic Acids Res.*, **16**, 2671–2690.
- Quintana, J. R., Lipanov, A. A. and Dickerson, R. E. (1991) *Biochemistry*, **30**, 10294–10306.

- 37 Carrondo, M. A. A. F. de C. T., 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.
- 38 Vega, M. C., Saez, I. G., Aymami, J., van der Marel, G. A., van Boom, J.H., Rich, A. and Coll, M. (1994) *Eur. J. Biochem.*, **222**, 721–726.
- 39 Czamy, A., Boykin, D. W., Wood, A. A., Nunn, C. M., Neidle, S., Zhao, M. and Wilson, W. D. (1995) *J. Am. Chem. Soc.*, **117**, 4716–4717.
- 40 Dickerson, R. E. and Drew, H. R. (1981) *J. Mol. Biol.*, **151**, 761–768.
- 41 Brünger, A. T., Kuriyan, J. and Karplus M. (1987) *Science*, **235**, 458–460.
- 42 Jones, T. A., Zou, J. -Y., Cowan, S. W. and Kjeldgaard, M. (1991) *Acta Crystallogr.*, **A47**, 110–119.
- 43 Mohamadi, F., Richards, N. G. J., Guida, W. C., Liskamp, R., Caufield, C., Chang, G., Hendrickson, T. and Still, W. C. (1990) *J. Comput. Chem.*, **11**, 440.
- 44 Stewart, J. P. P. Mopac 6.0 (QCPE 1990) available from the Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN 47405, USA.
- 45 Lavery, R. and Sklenar, H. (1988) *J. Biomol. Struct. Dynamics*, **6**, 63–91.