Sequence-dependent effects in drug – DNA interaction: the crystal structure of Hoechst 33258 bound to the d(CGCAAATTTGCG)₂ duplex

N.Spink, D.G.Brown⁺, J.V.Skelly and S.Neidle*

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

Received February 9, 1994; Revised and Accepted March 28, 1994

ABSTRACT

The bis-benzimidazole drug Hoechst 33258 has been co-crystallized with the dodecanucleotide sequence d(CGCAAATTTGCG)₂. The structure has been solved by molecular replacement and refined to an R factor of 18.5% for 2125 reflections collected on a Xentronics area detector. The drug is bound in the minor groove, at the five base-pair site 5'-ATTG and is in a unique orientation. This is displaced by one base pair in the 5' direction compared to previously-determined structures of this drug with the sequence d(CGCGAATTCGCG)₂. Reasons for this difference in behaviour are discussed in terms of several sequence-dependent structural features of the DNA, with particular reference to differences in propeller twist and minor-groove width.

INTRODUCTION

The study of interactions of low molecular weight natural and synthetic molecules with double-stranded DNA is currently the subject of considerable activity. Attention has increasingly focussed on those compounds that can recognize particular sequences of DNA, especially in the minor groove of AT regions [1–3]. There is now a large body of experimental and theoretical data on such systems which has been useful in the design of new molecules with altered sequence recognition properties [4–7].

The synthetic bis-(benzimidazole) derivative Hoechst 33258 [8] (Fig. 1) is widely used in chromosome staining and also possesses antihelmintic properties as well as modest activity against L1210 and P388 tumours in mice. The drug binds to AT-rich regions in DNA, in common with drugs such as netropsin, berenil and distamycin [1–3]. Footprinting studies [9,10] have shown that the Hoechst 33258 binding site in an AT region is at least four base pairs long, that a GC base pair can be accommodated at the 5' end of the binding site, and that the presence of a 5'-TpA step in the site is rare.

There have been a number of crystal structure analyses and NMR studies of Hoechst 33258 complexed to various oligonucleotide duplexes containing stretches of AT base pairs [11–17,30]. These have consistently shown that the drug binds in the minor grooves of the duplexes. The crystal structures of complexes with the duplex sequence d(CGCGAATTCGCG)₂ show that the drug is located in the central four AT base pair region such that the phenyl ring of the drug is at the 5' end of the site and the piperazine ring is close to the GC base pair at the 3' end of the site [11–13]. The location of the drug along this sequence and the pattern of drug–DNA hydrogen bonding is in general accord with an NMR study in solution [14], as well as with analyses of other sequences containing the AATT motif [15–17], although there are differences in detail. In particular, the analysis of the complex with the sequence d(GTGGGAATTCAC)₂ has revealed the coexistence of two families of bound drug conformers, with the terminal phenol and piperazine rings having significant mobility.

The present study reports the crystal structure of the complex formed between Hoechst 33258 and the self complementary dodecanucleotide duplex d(CGCAAATTTGCG)₂, in order to gain structural information on the site specificity of the drug in a sequence context distinct from yet closely related to the earlier studies. This sequence has several potential high-affinity binding sites, each with the binding site size requirement of four AT base pairs. We have previously reported on the crystal structures of dodecamer complexes for the sequences d(CGCGAATTCGCG)₂ and d(CGCAAATTTGCG)₂ with the bis-(phenyl-amidinium) groove binding drug berenil [18,19]. These suggested that there are inherent subtle differences in the structural features of the AT tracts in the two structures such that berenil is forced to be bound in the minor grooves of the two sequences in quite distinctive ways. The present study has been undertaken to ascertain whether there are such differences in the binding of Hoechst 33258, and thus how this drug interacts with differing AT tracts.

*To whom correspondence should be addressed

⁺ Present address: The Randall Institute, King's College London, London WC2B 5RL, UK

Figure 1. Structure of Hoechst 33258.

Figure 2. Two views showing the position of the bound drug in the molecular structure of the $d(CGCAAATTTGCG)_2$ -Hoechst 33258 complex, produced with the GRASP program (Nicholls, A and Honig, B. Columbia University). In (a), the solvent-accessible surface of the drug molecule is shown, with the DNA in schematic ribbon-and-box form. (b) Shows the solvent-accessible surface of the DNA with now the drug in skeletal representation.

METHODS

Crystallisation

The dodecanucleotide $d(CGCAAATTTGCG)_2$ was synthesised by solid phase phosphoramidite chemistry on an automated Applied Biosystems B91EP Synthesiser, and purified by anion-exchange and reverse-phase hplc. Hoechst 33258 (Aldrich Chemical Co.) was used as the trihydrochloride pentahydrate salt. Crystals were grown from 22 μ l droplets containing 5 μ l of 20mg/ml DNA [$d(CGCAAATTTGCG)_2$], 10 μ l of 2-methylpentan-2,4-diol (MPD) at 23% vol/vol (MPD/30mM sodium cacodylate buffer pH 7.0 containing 0.02% azide) and 2 μ l of 100mM $MgCl_2$, against a reservoir of 23% vol/vol MPD/30mM sodium cacodylate buffer pH 7.0 by the hanging droplet vapour diffusion method. Crystals were grown at 15°C; suitably-sized ones were obtained over a period of 18–24 days.

Figure 3. A view of the difference ($F_o - F_c$) map showing the initial difference electron density found in the minor groove of the $d(CGCAAATTTGCG)_2$ complex; this density corresponds to the Hoechst 33258 molecule, which is shown in skeletal representation.

Crystallographic data collection

A single crystal with approximate dimensions $0.5 \times 0.15 \times 0.15$ mm sealed in a quartz capillary was used to collect data to 2.25Å resolution on a Siemens Xentronics X-1000 area detector mounted on a Enraf-Nonius FR571 rotating anode X-ray generator, operated at 40kV and 60mA. The crystal-to-detector distance was set at 8cm and the 2θ swing angle at 10° . A frame collection time of 180 seconds was used. The cell constants $a=25.27$ $b=41.32$ and $c=65.11$ Å, and the space group $P2_12_12_1$ were obtained by auto-indexing using the XENGEN v1.4 software package. Two 100° scans in ϕ were carried out with a ω rotation of 60° between them, in order to obtain a near-complete data set. A total of 11,799 out of a possible 13,852 reflections were collected. These were scaled and merged using the XENGEN software to give 3,114 unique reflections out of a possible 3,437 to a resolution of 2.1Å, with an unweighted merging R-factor of 6.1%.

Structure solution and refinement

The crystals are isomorphous with those of the same sequence complexed with the previously studied minor groove drug berenil [19], as well as the native sequence [20] and so the dodecamer from these structures was used as a starting model. Five cycles of rigid body restrained/constrained refinement was carried out using the program CORELS [21] for reflection data between 10–4Å (662 unique reflections), giving an R-factor of 36.0%. At this stage a Fourier sum ($2F_o - F_c$) map was calculated using the PROTEIN suite of programs [22] and displayed on a Silicon Graphics Iris 3130 workstation using the TOM display package

Table 1. Conformational angles for the d(CGCAAATTTGCG)₂–Hoechst 33258 complex (boldface) compared to those of the native d(CGCAAATTTGCG)₂ structure [28]

Residue	Glycosidic angle, °	Main-chain torsion angle, °					
	χ	α	β	γ	δ	ϵ	ζ
C1	274 228			126 131	160 105	180 174	248 282
G2	267 277	252 277	202 192	59 45	146 156	219 219	216 203
C3	247 222	291 301	152 148	50 47	137 89	221 198	190 261
A4	252 258	314 309	134 160	50 65	156 129	249 187	193 256
A5	274 262	315 320	163 179	10 24	161 147	206 189	234 254
A6	270 251	309 296	193 177	13 53	161 127	204 175	233 272
T7	228 244	312 293	138 175	58 54	103 100	190 168	268 276
T8	224 240	246 282	158 175	106 64	117 125	203 196	237 231
T9	265 257	38 339	161 149	314 26	146 104	195 179	203 281
G10	279 274	45 307	129 190	329 23	164 170	185 252	211 156
C11	240 262	304 304	160 147	64 35	140 145	179 206	279 253
G12	270 277	251 293	212 182	57 28	131 139		
C13	223 243			164 356	160 126	187 322	238 132
G14	270 237	46 244	154 147	307 56	169 132	180 198	247 234
C15	219 227	296 309	158 144	66 54	85 87	185 164	274 284
A16	262 262	293 290	168 175	62 66	150 144	190 183	258 232
A17	261 261	297 339	207 162	32 26	159 147	188 179	255 246
A18	256 258	287 317	175 167	60 33	152 134	215 178	229 249
T19	217 234	294 325	136 146	72 40	102 107	211 190	240 250
T20	248 261	302 321	161 177	43 25	151 148	240 189	196 240
T21	234 253	162 189	304 328	165 159	45 30	147 138	265 210
C23	255 261	311 329	161 150	358 21	164 146	187 206	223 228
G24	222 236	299 311	165 150	48 34	82 91		

[23], to check that the model was essentially correct. Data in the range 8–3 Å was then used for five more cycles of rigid body refinement, and the R factor was found not to have changed significantly, indicating the general correctness of the model. The structure was then divided into 48 rigid groups, of individual phosphates, sugars and nucleotides, and subjected to 10 cycles of positional refinement followed by 5 cycles of temperature factor refinement, reducing the R-factor to 25.6%. Refinement using the least squares refinement program NUCLSQ [24] for a data range between 7–2.25 Å gave an R-factor of 27.6%. The location of water molecules followed by further refinement reduced the R factor to 21.3%. A subsequent difference electron density map showed the ligand, but in an ambiguous orientation and so the XPLOR package [25] was used for further refinement. An initial 40 cycles of positional refinement using data in the range 8–2.25 Å gave an R-factor of 27.9%. Solvent refitting was carried out with the aid of the water search protocol from the CCP4 package [26]. In this, Fourier sum and difference maps were averaged and the structure density masked prior to a peak search above 2.5 σ . The resulting solvent peaks were checked graphically for geometry constraints using the 'O' molecular display program [27] on a Silicon Graphics Iris Indigo XS24 workstation. 44 initial water molecules were added and refined to give an R-factor of 22.1%. A difference map was then calculated which unambiguously showed electron density for the drug, and that it was in a particular 'piperazine-down' orientation in the minor groove. Rigid body refinement on the drug alone, using electrostatic and van der Waals energy constraints, was carried out both on the fitted model and the alternative 'piperazine-up' one to verify the assignment of orientation. The 'piperazine-down' model gave a R-factor of 19.4% as compared to the 'piperazine-up' R-factor of 20.5%. Fourier sum and difference maps also showed the former orientation to be correct. This d(CGCAAATTTGCG)₂–Hoechst 33258 complex structure was further refined to a final R-factor of 18.5%, when all

observed data between the range of 8–2.25 Å (2125 reflections) with $I > 2.0\sigma(I)$ were included, together with 61 solvent molecules. Final atomic coordinates and structure factors have been deposited in the Brookhaven Protein Database.

RESULTS AND DISCUSSION

Drug–DNA interactions in the complex

The overall structure of the complex between Hoechst 33258 and the DNA dodecamer d(CGCAAATTTGCG)₂ is illustrated in Fig. 2. The drug is bound in the AT region of the minor groove of the duplex, with no evidence of disorder or mobility along the groove, as shown by the discrete electron density for the drug (Fig. 3). The drug molecule adopts a slightly twisted, non-planar conformation with a dihedral angle of 19° between the two benzimidazole rings. The drug lies in a unique, non-disordered position fully covering the lower four A·T base pairs and the adjacent G·C one (ie the sequence 5'-ATTTG), with the N-methyl piperazine ring of the drug reaching the edge of the lower adjacent G·C base pair C11·G14 and at the other end the hydroxyl group extending almost to the A·T base pair A5·T20 (Fig. 4a). This is the 'piperazine-down' orientation. The non-bonded contacts at and around van der Waals separations between the inner-facing atoms C21 and C22 of the piperazine group and these base pairs, are mostly with one side of the groove, especially with sugar ring O4' atoms of residues G10 and C11. There is also a (3.4 Å) contact between N3 of G10 and atom C21.

A number of hydrogen-bonded contacts between drug and the dodecamer are apparent. The N1 and N3 nitrogen atoms of the benzimidazole rings both form bifurcated hydrogen bonds to the A·T base pairs; the N1 atom of one benzimidazole group forms hydrogen bonds to O2 of thymine T7 and O2 of thymine T19 and the N3 atom of the other benzimidazole forms hydrogen bonds to O2 of thymine T8 and N3 of adenine A18. The phenol terminal oxygen atom of Hoechst 33258 forms a hydrogen bond

Figure 4. Schematic representations of hydrogen-bond interactions in the minor grooves of the Hoechst 33258 complexes with (a) d(CGCAAATTTGCG)₂ and (b) d(CGCGAATTCGCG)₂.

to a water molecule, which is part of a short solvent chain. This chain is itself hydrogen-bonded to several 5' end bases of one strand and accordingly is asymmetrically situated in the minor groove.

Previous crystal structures of the complex between Hoechst 33258 and the sequence d(CGCGAATTCGCG)₂ have shown the drug to be more symmetrically bound with respect to the two fold axis of the self-complementary sequence (Fig. 4b), such that it is situated in the 5'-AATT site [12,13]. NMR studies on this complex and with other sequences containing the 5'-AATT motif generally concur with this assignment of binding site. In both of the previous crystallographic studies with the 5'-AATT sequence the drug has a 'piperazine-down' orientation, as found in the present analysis, although the study [13] at both low and ambient temperatures found evidence for a 'piperazine-up'

orientation as well in the ambient-temperature structure. Fig. 4 shows, by contrast, that in the present structure the drug molecule has its binding site (of now 5'-ATTTG) displaced by one complete base pair in the 5' direction, even though the patterns of hydrogen-bonded drug-DNA contacts are equivalent for the two sequences d(CGCGAATTCGCG)₂ and d(CGCAAATTTGCG)₂. The differences in sequences that form the binding sites, results in N1 from the 3'-end benzimidazole ring contacting a pair of thymine O2 atoms in the latter sequence compared to, in the former structure, atoms O2 and N3 from thymine T20 and adenine A6.

The fact that the bound drug extends into the GC region of the sequence here is fully in accord with the footprinting studies [9] on Hoechst 33258, which also indicate a binding site size of five base pairs, as indeed is found in the present study.

Figure 5. Histograms of (a) propeller twist and (b) roll angles calculated with the NEWHEL92 program, for the d(CGCGAATTCGCG)₂ [19] and d(CGCAAATTTGCG)₂ [28] duplex structures, with and without bound Hoechst 33258 [13]. Coordinates were taken from the Brookhaven Databank.

The conformation of the DNA

There are few changes of significance to DNA backbone torsion angles and sugar puckers compared (Table 1) to those in the crystal structures of the native d(CGCGAATTCGCG)₂ [20] and d(CGCAAATTTGCG)₂ [28] sequences. All values fall within normal B-DNA ranges. Similarly, the parameters of base and base pair morphology (calculated with the NEWHEL92 program) are not significantly distinct from those in the native sequence, although the values of propeller twist and (especially) roll at some points along the sequence do differ when comparing the native and drug-bound 5'-AAATTT structures with the 5'-AATT one (Fig. 5). In particular, there are significant differences in roll, of ca 10° at base-pair steps 3, 4 and 6 (ie at CG/AT, AT/AT and AT/TA reading from the 3' end). There is a difference of about 10° in the propeller twists of base pair 9, which is possibly a consequence of being A·T in the present structure compared to C·G in the 5'-AATT complex, in accord with the general tendency for A·T base pairs to have a higher propeller twist.

The width of the minor groove in the present Hoechst 33258 complex is shown in Fig. 6, compared to that found in the crystal

Figure 6. Plot of minor groove widths in the d(CGCAAATTTGCG)₂ native [28] and Hoechst-bound structures, together with values for the d(CGCGAATTCGCG)₂-Hoechst complex [13]. Values on the vertical axis are the inter-strand P - P distances, less 5.8 Å. The horizontal axis indicates the P - P pairs involved in each instance.

structures of the native sequence [28] and the d(CGCGAATTCGCG)₂-Hoechst complex [13]. In overall terms, all three structures have a narrowed AT region, although there are differences in detail. It is notable that the native and Hoechst 5'-AAATTT structures have an almost identical pattern of groove narrowing, with a minimum width of ~4.5 Å in the region of bases 8-21 and 10-19, ie centred around base pair 9. By contrast, in the Hoechst-d(CGCGAATTCGCG)₂ complex the narrowest point in the groove, of ~4.0 Å, is around base pair 8, yet at base pair 10 it is ~1 Å wider than in the two 5'-AAATTT structures. The greater narrowness of the minor groove in the vicinity of base pairs 8 and 9 in the 5'-AAATTT Hoechst complex (and the native sequence) are likely to be a consequence of the differences in propeller twist detailed above, in turn a result of the extended AT tract in the sequence. This is in accord with observations elsewhere that runs of AT base pairs tend to have high propeller twists and to result in narrow minor grooves [for example, 29].

This crystal-structure analysis has shown that the minor-groove binding drug Hoechst 33258 has distinct binding sites with an AT region, depending on the nature of the flanking bases. When confronted with the sequence ..GAATTC.. as in the dodecamer d(CGCGAATTCGCG)₂ and in the sequences studied by NMR [14-17] it exhibits a strong preference for the central symmetric 5'-AATT site. This preference is not sustained in the present structure where the central six base-pair sequence 5'-AAATTT could still allow 5'-AATT to be the binding site. We have previously noted that there are small but significant differences in base-pair roll, propeller twist and minor groove width between the two native dodecamer sequences. It was observed that these differences were also apparent in their complexes with the drug berenil [18,19] that resulted in this drug having differing binding site sizes and hydrogen bonding features to the bases in the two sequences. The present Hoechst complex confirms and extends these findings, that the central sequence 5'-AAATTT in a

dodecamer imposes small though significant structural differences compared to that with 5'-AATT, which force the drug to have a differing binding site, with the narrowed minor groove in the 3' direction of the former sequence presumably favouring drug binding to be displaced to this direction. Thus in none of its crystal structures with DNA duplexes are berenil or Hoechst 33258 able to appreciably distort the structures so as to produce a single type of binding site for each drug. In other words, the DNA sequence imposes a particular structure that is not significantly altered by the drug. Hoechst 33258 has also been co-crystallized as a complex with the sequence d(CGCGATATCGCG)₂, with the central alternating 5'-ATAT sequence having marked alternations in several of its structural features [30]. In this case, the observed differences in binding characteristics from the d(CGCGAATT-CGCG)₂ complex are less surprising.

ACKNOWLEDGEMENTS

We are grateful to the Cancer Research Campaign for support and Research Studentships (to DGB and NS), and to Mark Sanderson for much useful discussion. We also thank Werner Leupin (Hoffman-la-Roche, Basle) for discussion and for providing information on his NMR study in advance of publication. Aspects of this work have been presented elsewhere (Brown, D.G., Ph.D. thesis, University of London, 1992); an independent 2.4 Å study has subsequently been briefly reported by Coll *et al.* at the International Congress of Biophysics (Budapest, 1993) and at the Instituto Juan March Workshop on DNA-drug Interactions (Madrid, 1993).

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 *Nucleic Acid Targeted Drug Design* (Propst, C.L. and Perun, T.J. eds.) pp 304–374, Marcel Dekker, Inc., New York, USA.
- Goodsell, D. and Dickerson, R.E. (1986). *J. Med. Chem.*, **29**, 727–733.
- Bruice, T.C., Mei, H.-Y., He, G.-X. and van Lopez (1992). *Proc. Natl. Acad. Sci. USA*, **89**, 1700–1704.
- Dwyer, T.J., Geierstanger, B.H., Bathini, Y., Lown, J.W. and Wemmer, D.E. (1992). *J. Amer. Chem. Soc.*, **114**, 5911–5919.
- Wade, W.S., Mrksich, M. and Dervan, P.B. (1992). *J. Amer. Chem. Soc.*, **114**, 8783–8794.
- Loewe, H. and Urbanietz, J. (1974). *Arzneim.-Forsch.*, **24**, 1927–1933.
- Harshman, K.D. and Dervan, P.B. (1985). *Nucleic Acids Res.*, **13**, 4825–4835.
- Murray, V. and Martin, R.F. (1988). *J. Mol. Biol.*, **203**, 63–73.
- Pjura, P.E., Grzeskowiak, K. and Dickerson, R.E. (1987). *J. Mol. Biol.*, **197**, 257–271.
- Teng, M.-K., 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.
- 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.
- Embrey, K.J., Searle, M.S. and Craik, D.J. (1993). *Eur. J. Biochem.*, **211**, 437–447.
- Fede, A., Billeter, M., Leupin, W. and Wüthrich, K. (1993). *Structure*, **1**, 177–186.
- Brown, D.G., Sanderson, M., Skelly, J.V., Jenkins, T.C., Brown, T., Garman, E., Stuart D.I. and Neidle, S. (1990). *EMBO J.*, **9**, 1329–1334.
- Brown, D.G., Sanderson, M.R., Garman, E. and Neidle, S. (1992). *J. Mol. Biol.*, **226**, 481–490.
- Dickerson, R.E. and Drew, H.R. (1981). *J. Mol. Biol.*, **151**, 761–768.
- Sussman, J.L. (1984) in *Methods and Applications in Crystallographic Computing* (Hall, S.R. and Ashida, T., eds.) pp 206–237, Clarendon Press, Oxford, UK.
- Steigemann, W. (1974). Dissertation, Technisch Universität, München.
- Cambilleau, C. (1988). TOM, a version of FRODO amended to run on a Silicon Graphics workstation.
- Westhof, E., Dumas, P. and Moras, D. (1985). *J. Mol. Biol.*, **184**, 119–145.
- Brünger, A.T., Kuriyan, J. and Karplus, M. (1987). *Science*, **235**, 458–460.
- CCP4 Package, Science and Engineering Research Council, Daresbury Laboratory, UK.
- Sriram, M., van der Marel, G.A., Roelen, H.L.P.F., van Boom, J.H. and Wang, A.H.-J. (1992). *EMBO J.*, **11**, 225–232.
- Edwards, K.J., Brown, D.G., Spink, N., Skelly, J.V. and Neidle, S. (1992). *J. Mol. Biol.*, **226**, 1161–1173.
- Privé, G.G., Yanagi, K. and Dickerson, R.E. (1991). *J. Mol. Biol.*, **217**, 177–199.
- de C.T. Carrondo, M.A.A.F., 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.