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'-ATTTG 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(CGCG-AATTCGCG)$_2$. 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 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(GTGGAATT-CCAC)$_2$ 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(CGCGAATT-CGCG)$_2$ 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
METHODS

Crystallisation

The dodecanucleotide d(CGCAAAATTTGC)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 μl droplets containing 5 μl of 20mg/ml DNA [d(CGCAAAATTTGC)₂], 10 μl 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 μl of 100mM MgCl₂, 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.

Crystallographic data collection

A single crystal with approximate dimensions 0.5×0.15×0.15mm sealed in a quartz capillary was used to collect data to 2.25Å resolution on a Siemens Xentronics X-1000 area detector mounted on an 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=65.11Å, and the space group P2₁₂₁ 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 (2Fₒ-Fᵡ) map was calculated using the PROTEIN suite of programs [22] and displayed on a Silicon Graphics Iris 3130 workstation using the TOM display package.
difference maps also showed the former orientation to be correct. Using electrostatic and van der Waals energy constraints, was carried out both on the fitted model and the alternative drug, and that it was in a particular 'piperazine-down' orientation to give an R-factor of 22.1%. A difference map was then carried out with the aid of the water search protocol from the range 8—2.25Å. A subsequent difference electron density map showed the ligand, but in an ambiguous orientation. Rigid body refinement of the drug alone, with I >2.0σ(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 C1 terminal oxygen atom of Hoechst 33258 forms a hydrogen bond to the AT base pairs; the N1 atom of one benzimidazole group forms hydrogen bonds to O2 of thymine T8 and N3 of adenine A18. The phenol of the drug binds to the inner-facing atoms C21 and C22 of the piperazine group and with sugar ring 04' atoms of residues G10 and C11. There is also a (3.4Å) contact between N3 of G10 and atom C21.

An initial 40 cycles of positional refinement using data in the range 8—3Å was then used for five more cycles of rigid body refinement, reducing the R-factor to 25.6%. Refinement of the hydration shell was further refined to a final R-factor of 18.5%, when all observed data between the range of 8—2.25Å 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 phosphate, sugar, and nucleotide, and subjected to 10 cycles of positional refinement following 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—2.25Å 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.

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 N1 terminal oxygen atom of Hoechst 33258 forms a hydrogen bond to the AT base pairs; the N1 atom of one benzimidazole group forms hydrogen bonds to O2 of thymine T8 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 of the drug binds to the inner-facing atoms C21 and C22 of the piperazine group and with sugar ring 04' atoms of residues G10 and C11. There is also a (3.4Å) 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

<table>
<thead>
<tr>
<th>Residue</th>
<th>Glycosidic angle,°</th>
<th>Main-chain torsion angle,°</th>
<th>χ</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>δ</th>
<th>ε</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>274 228</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>126 131</td>
<td>160 105</td>
<td>180 174</td>
<td>248 282</td>
</tr>
<tr>
<td>G2</td>
<td>265 277</td>
<td></td>
<td>220 192</td>
<td>59 45</td>
<td>146 156</td>
<td>219 219</td>
<td>216 203</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>247 222</td>
<td></td>
<td>291 301</td>
<td>50 47</td>
<td>137 89</td>
<td>221 198</td>
<td>265 234</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>252 258</td>
<td></td>
<td>314 309</td>
<td>50 65</td>
<td>156 129</td>
<td>249 187</td>
<td>219 266</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>274 262</td>
<td></td>
<td>315 320</td>
<td>10 24</td>
<td>161 147</td>
<td>216 175</td>
<td>233 232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>270 251</td>
<td></td>
<td>309 296</td>
<td>13 53</td>
<td>161 127</td>
<td>204 175</td>
<td>233 223</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T7</td>
<td>228 244</td>
<td></td>
<td>312 293</td>
<td>38 54</td>
<td>103 100</td>
<td>190 168</td>
<td>268 276</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T8</td>
<td>224 240</td>
<td></td>
<td>246 282</td>
<td>158 175</td>
<td>106 64</td>
<td>117 125</td>
<td>203 196</td>
<td>237 231</td>
<td></td>
</tr>
<tr>
<td>T9</td>
<td>265 257</td>
<td></td>
<td>38 339</td>
<td>161 149</td>
<td>314 26</td>
<td>146 104</td>
<td>195 179</td>
<td>203 281</td>
<td></td>
</tr>
<tr>
<td>G10</td>
<td>279 274</td>
<td></td>
<td>45 307</td>
<td>129 190</td>
<td>329 23</td>
<td>164 170</td>
<td>185 252</td>
<td>211 156</td>
<td></td>
</tr>
<tr>
<td>C11</td>
<td>240 262</td>
<td></td>
<td>304 304</td>
<td>160 147</td>
<td>64 35</td>
<td>140 145</td>
<td>179 206</td>
<td>279 253</td>
<td></td>
</tr>
<tr>
<td>G12</td>
<td>270 277</td>
<td></td>
<td>281 293</td>
<td>212 182</td>
<td>77 28</td>
<td>131 139</td>
<td>187 322</td>
<td>238 132</td>
<td></td>
</tr>
<tr>
<td>C13</td>
<td>223 343</td>
<td></td>
<td>46 244</td>
<td>154 147</td>
<td>307 56</td>
<td>169 132</td>
<td>198 180</td>
<td>247 243</td>
<td></td>
</tr>
<tr>
<td>G14</td>
<td>270 237</td>
<td></td>
<td>296 309</td>
<td>158 144</td>
<td>66 54</td>
<td>85 87</td>
<td>185 164</td>
<td>274 284</td>
<td></td>
</tr>
<tr>
<td>C15</td>
<td>219 227</td>
<td></td>
<td>293 290</td>
<td>168 175</td>
<td>62 66</td>
<td>150 144</td>
<td>190 183</td>
<td>258 232</td>
<td></td>
</tr>
<tr>
<td>A16</td>
<td>262 262</td>
<td></td>
<td>297 339</td>
<td>207 162</td>
<td>32 26</td>
<td>159 147</td>
<td>188 179</td>
<td>255 246</td>
<td></td>
</tr>
<tr>
<td>A17</td>
<td>261 261</td>
<td></td>
<td>287 317</td>
<td>175 167</td>
<td>60 33</td>
<td>152 134</td>
<td>215 178</td>
<td>229 249</td>
<td></td>
</tr>
<tr>
<td>A18</td>
<td>256 238</td>
<td></td>
<td>294 325</td>
<td>136 146</td>
<td>72 40</td>
<td>102 107</td>
<td>211 190</td>
<td>240 250</td>
<td></td>
</tr>
<tr>
<td>T19</td>
<td>217 234</td>
<td></td>
<td>302 321</td>
<td>161 177</td>
<td>43 25</td>
<td>151 148</td>
<td>240 189</td>
<td>196 240</td>
<td></td>
</tr>
<tr>
<td>T20</td>
<td>248 261</td>
<td></td>
<td>162 189</td>
<td>304 328</td>
<td>165 159</td>
<td>45 30</td>
<td>147 138</td>
<td>265 210</td>
<td></td>
</tr>
<tr>
<td>T21</td>
<td>234 253</td>
<td></td>
<td>311 329</td>
<td>161 150</td>
<td>358 21</td>
<td>164 146</td>
<td>187 206</td>
<td>223 228</td>
<td></td>
</tr>
<tr>
<td>C23</td>
<td>255 261</td>
<td></td>
<td>299 311</td>
<td>165 150</td>
<td>48 34</td>
<td>82 91</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Conformational angles for the d(CGCAAATTTGCG)2-Hoechst 33258 complex (boldface) compared to those of the native d(CGCAAATTTGCG)2 structure [28]
Figure 4. Schematic representations of hydrogen-bond interactions in the minor grooves of the Hoechst 33258 complexes with (a) d(CGCGAATTCGCG)\textsubscript{2} and (b) d(CGCAAATTTGCG)\textsubscript{2}.

Previous crystal structures of the complex between Hoechst 33258 and the sequence d(CGCGAATTCGCG)\textsubscript{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)\textsubscript{2} and d(CGCAAATTTGCG)\textsubscript{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.
The conformation of the DNA

The are few changes of significance to DNA backbone torsion angles and sugar pucker 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 structures of the native sequence [28] and the d(CGCGAATTTCGCG)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 ~4.5Å 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 ~4.0Å, is around base pair 8, yet at base pair 10 it is ~1Å 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 2.4Å 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