Volume 16 Number 6 1988 Nucleic Acids Research

The molecular structure of the complex of Hoechst 33258 and the DNA dodecamer d(CGCGAATTCGCG)

Mai-kun Teng, Nassim Usman, Christin A.Frederick and Andrew H.-J.Wang*

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received November 30, 1987; Revised and Accepted February 17, 1988

ABSTRACT

The crystal structure of the complex between the dodecamer d(CGCGAATTCGCG) and a synthetic dye molecule Hoechst 33258 was solved by X-ray diffraction analysis and refined to an R-factor of 15.7% at 2.25 Å resolution. The crescent-shaped Hoechst compound is found to bind to the central four AATT base pairs in the narrow minor groove of the B-DNA double helix. The piperazine ring of the drug has its flat face almost parallel to the aromatic bisbenzimidazole ring and lies sideways in the minor groove. No evidence of disordered structure of the drug is seen in the complex. The binding of Hoechst to DNA is stabilized by a combination of hydrogen bonding, van der Waals interaction and electrostatic interactions. The binding preference for AT base pairs by the drug is the result of the close contact between the Hoechst molecule and the C2 hydrogen atoms of adenine. The nature of these contacts precludes the binding of the drug to G-C base pairs due to the presence of N2 amino groups of guanines. The present crystal structural information agrees well with the data obtained from chemical footprinting experiments.

INTRODUCTION

Hoechst dye 33258 is a synthetic compound that contains two consecutive benzimidazole rings with a phenolic and an N-methyl-piperazine group at either end of the elongated molecule (Fig. 1). The compound has an anthelminthic activity, but no antitumor activity [1]. It is known to bind to double helical DNA, and as such it has been widely used as a fluorescent cytological stain for DNA. Several studies have shown that Hoechst 33258 binds preferentially at the AT-rich region of DNA with a minimum binding size of four consecutive AT base pairs [2]. This has prompted suggestions that the Hoechst compound, with its bisbenzimidazole size and shape, similar to the N-methylpyrrole-containing netropsin, is also a minor groove binding molecule. Indeed, recent comparative studies of the binding sites of Hoechst 33258, netropsin, and distamycin on plasmid DNA restriction fragments by methidium-

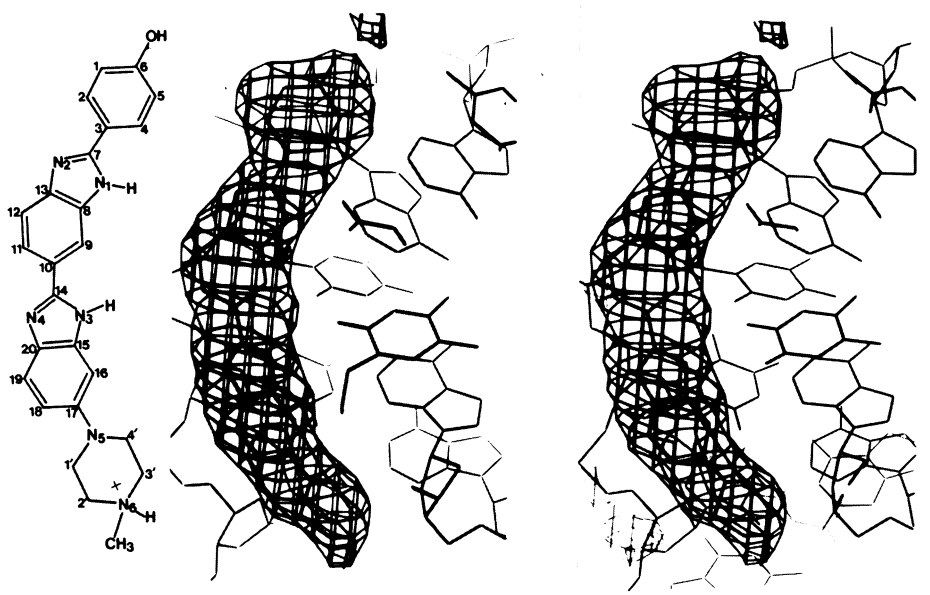


Figure 1: (A). Molecular formula of Hoechst 33258. There are two consecutive benzimidazole rings (bisbenzimidazole) in the molecule with a phenolic ring on one end and a positively charged piperazine at the other end. The pKa of the benzimidazole, a weak base, is 5.48. At neutral pH, the drug molecule has only one positive charge, on the N6 nitrogen atom, rather than the three positive charges as shown in reference [4]. (B). A portion of the difference Fourier electron density map based on the phases from DNA and solvent atoms only. The cresent-shaped Hoechst molecule fits the density very well and there is no ambiguity in the orientation of the drug molecule.

propyl EDTA.Fe(II) footprinting experiments showed that they share common AT binding sites of 4-5 base pairs in length [3].

More recently, a crystal structure of the complex between the Hoechst 33258 and a DNA dodecamer d(CGCGAATTCGCG) has been reported [4]. In this structure, the drug molecule is stated to lie in the minor groove near one end of the AATT region such that the piperazine ring protrudes into the wider groove corresponding to the GC region. This implies a prefered binding sequence of ATTC which is different from that of the netropsin molecule [5]. In addition, the piperazine ring appears to adopt two radically different conformations in roughly equal populations in the crystal structure. From these results, it has been suggested that piperazine rings provide possible GC specific binding elements. Therefore, there appear to be some

inconsistencies between the X-ray crystallographic studies [4] and footprinting experiments [3] with respect to the binding specificity of Hoechst 33258 to DNA.

We have carried out an independent crystal structure analysis on the same Hoechst-DNA complex at 2.25 Å resolution which shows a different binding mode of the drug from that reported earlier by Pjura *et. al.* [4]. In our structure, the Hoechst drug molecule binds tightly in the minor groove of the DNA double helix completly covering the four AATT base pairs. The piperazine ring has its flat face almost parallel to the bisbenzimidazole ring and there is no structural disorder associated with the rotation of this ring. The binding of Hoechst to DNA is stabilized by a combination of hydrogen bonds between the benzimidazole NH groups and the O2(T) and N3(A) atoms of the AT base pairs, van der Waals interactions between the benzimidazole and phenol aromatic π electron clouds with O4' atoms of the sugar along the side wall of the minor groove, and finally the electrostatic interactions between the positively charged Hoechst molecule and the negatively charged DNA molecule.

MATERIALS AND METHODS

The DNA dodecamer d(CGCGAATTCGCG) was synthesized on an Applied Biosystems 380B DNA synthesizer and purified as described earlier [6]. Hoechst dye 33258 was a gift from Dr. P. Dervan. The crystals were obtained by the vapor diffusion technique by equilibrating a solution containing 1.2 mM (single strand) DNA dodecamer, 20 mM sodium cacodylate (pH 6.5), 8 mM $MgCl_2$, 1 mM spermine tetrachloride, 1.2 mM Hoechst 33258 and 12% 2-methyl-2,4-pentanediol (2-MPD) against a reservoir of 50% 2-MPD at room temperature. After two months, crystals with a faint yellow color could be seen in the crystallization droplets. We also used the restriction endonuclease EcoRI to confirm the authenticity of the DNA molecule in the dips. To each crystallization dip was added 500μl of 10 mM Tris/0.1mM EDTA-pH8.0 buffer to dissolve the remaining crystals and precipitates. The solution was then extracted sequentially with 500μl *n*-BuOH, 500μl $CHCl_3$/*i*-amylalcohol (24/1) and, 500μl Et_2O. A small portion of the sample was compared with an authentic sample of d(CGCGAATTCGCG) on a 16% polyacrylamide sizing gel and found to be identical in terms of mobility. Once the size of the molecule was confirmed the remaining solution, along with the authentic d(CGCGAATTCGCG)

and a sample of a related dodecamer d(CGCAAATTTGCG), were passed through a C_{18}-Sep-Pak column (Millipore) to remove any traces of drug and salt, 200 pmol aliquots of the solution were dissolved in 50 mM NaCl/100mM Tris-pH7.5/5mM $MgCl_2$ and treated with EcoRI (New England Biolabs) at 20° for 16 hr [7]. Only the crystallized dodecamer and the authentic d(CGCGAATTCGCG) were cut, the related d(CGCAAATTTGCG) was completely resistant to cleavage, thereby confirming the sequence of the crystallized dodecamer.

The complex crystallized in the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of a=25.23, b=40.58 and c=66.08 Å A crystal with dimensions of 0.2 x 0.4 x 0.7 mm was mounted in a glass capillary and sealed with a droplet of mother liquor for data collection. The three-dimensional diffraction data were collected on a Nicolet P3 X-ray diffractometer at 15°C using an ω-scan mode with CuKα radiation to 2.25 Å resolution. Out of the total 3634 reflections collected, 2000 reflections (between 20 Åand 2.25 Åwere considered to be observed to at least the 2.0 σ(F) level and they were used in the refinement.

Due to the overall similarity between this crystal and other related dodecamer crystals, trial coordinates from the crystal structure of d(CGCGAm^{6}ATTCGCG) [8] were used for the initial model. The structure was refined using the Konnert-Hendrickson constrained refinement procedure [9]. Initial refinement was carried out without the inclusion of the Hoechst drug nor any solvent molecules to an R-factor of 29.5%. A difference Fourier map at this stage revealed clearly the drug molecule in the minor groove. However, at this time, we refrained from including the drug atoms in the refinement. Instead, solvent molecules found in other parts of the lattice were added in small increments. With the addition of 34 solvent molecules, the R-factor was reduced to 24.8%. At this time, the location of the Hoechst molecule was determined from the difference Fourier map displayed on an Evans-Sutherland PS340 graphics system.

Because Hoechst 33258 is not a symmetrical molecule, there are two possible orientations for the drug molecule in which it may interact with the DNA double helix. Both were examined for the best fit to the strong difference Fourier electron density in the minor groove and it was concluded that one had a clearly better fit than the other. The atoms of the drug molecule in this orientation as well as more solvent molecules were included in the subsequent refinement, reducing the R-factor

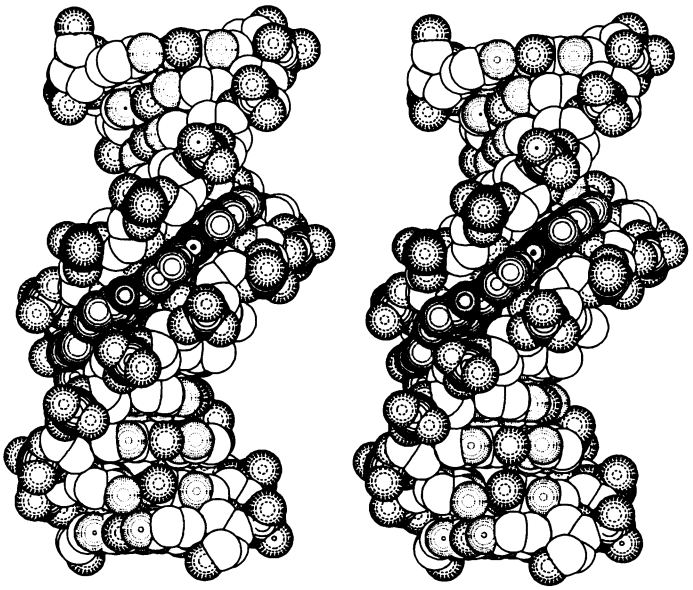


Figure 2: A stereoscopic van der Waals drawing of the complex of Hoechst 33258 and d(CGCGAATTCGCG) looking in the minor groove direction. In this orientation, it can be seen that the central minor groove region is completely occupied by the drug molecule (shaded slightly darker) with close van der Waals contacts between the atoms from the side walls of the DNA minor groove and the planar aromatic rings of the drug.

to 16.2%. In order to visualize and confirm the fitting of the Hoechst drug model to the electron density, a deletion difference Fourier map based on the calculated phases from all atoms, (DNA and solvent), except the drug molecule was calculated and is shown in Figure 1B. It can be seen that the drug model lies entirely within the electron density. A solvent water peak near the phenolic hydroxyl group at the top of the figure is also seen. This supports the correctness of the orientation of the drug molecule since the piperazine ring at the other end has a methyl group that cannot form hydrogen bonds with the solvent water molecules. No evidence of a disordered conformation of the piperazine ring is seen.

During the refinement, there were no hydrogen bonding distance or geometry constraints imposed between the Hoechst and DNA, although any unreasonable van der Waals close contacts between the two were carefully excluded. Bond length and angle constraints were held fairly tight such that the bond geometry of the complex

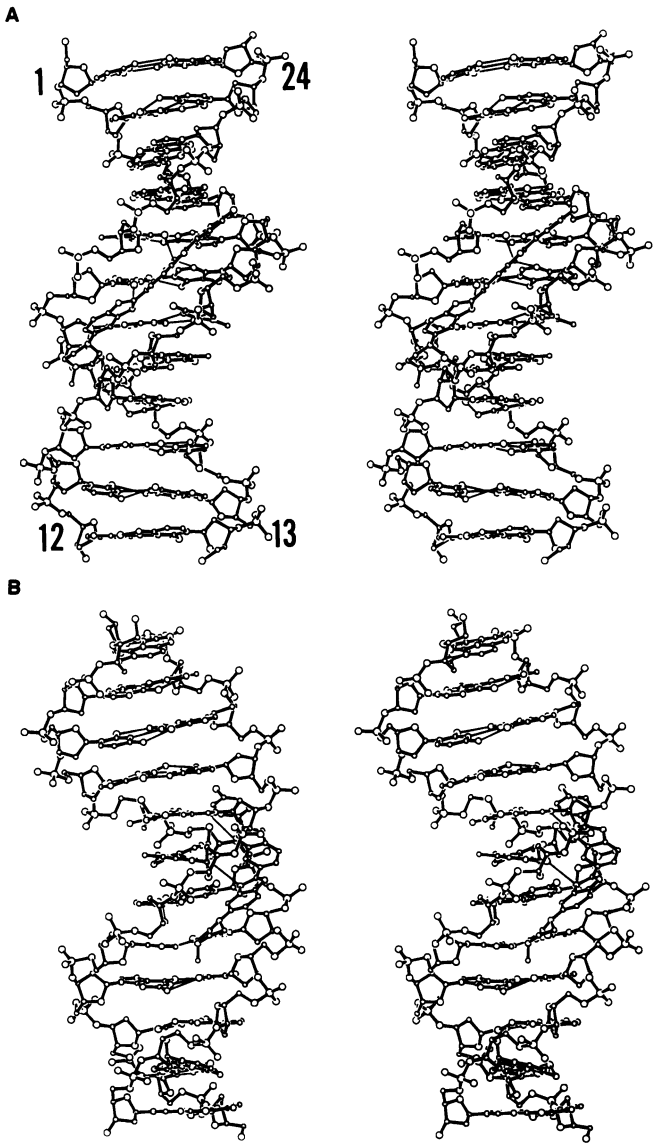


Figure 3: Stereoscopic skeletal drawings of the complex. (A). A view looking into the minor groove as in Figure 2. The DNA double helix is relatively straight with small propeller twist in most of the base pairs. The nucleotides are numbered at four ends of the dodecamer duplex. The Hoechst molecule has a twist in its structure (32° between two benzimidazoles) in order to follow the right-handed curvature of the groove and it covers the central AATT four base pairs completely. The slightly puckered piperazine ring lies sideways in the groove. (B). The complex rotated 90° relative to (A).

is quite reasonable with a root-mean-square deviation for the single bond lengths of 0.030 Åfrom the idealized value. The final R-factor for the structure of the complex with 126 solvent molecules is 15.7% using 2000 reflections at the 2 σ(F) level between 20 Å and 2.25 Å resolution. No spermine or metal ions could be identified unambiguously. The atomic coordinates of the complex will be deposited in the Brookhaven Protein Data Bank.

RESULTS AND DISCUSSION

Structure of the Complex

The overall structure of the complex between Hoechst 33258 and the d(CGCGAATTCGCG) dodecamer is illustrated with the van der Waals diagram in Figure 2. The Hoechst 33258 has an elongated, slightly curved, planar conformation such that it can fit into the narrow minor groove at the AATT region of the right-handed dodecamer double helix. The drug lies near the center of the double helix covering the entire AATT base pairs with the two ends of the molecule reaching the edges of the two adjacent GC base pairs. When the drug is bound to DNA in the minor groove, it is forced to follow the curved contour of the groove. This generates a significant twist in the drug molecule. It can be seen in Fig. 2 that the phenolic group is almost coplanar (dihedral angle = 8°) with the first benzimidazole ring (denoted BZI1). However, there is a much larger twist angle, 32°, between the two consecutive benzimidazole (BZI1 and BZI2) rings. The best plane of the slightly puckered piperazine ring is almost parallel to the adjacent second benzimidazole ring with a dihedral angle of 14°.

The fitting of the drug molecule to the DNA minor groove is very tight, as evidenced by the large number of close van der Waals interactions between them. This is clearly shown in Figure 3 by two views of the skeletal stereoscopic drawing of the complex. Figure 3a is viewed from the same orientation in the minor groove as in the van der Waals diagram of Figure 2. The inner side of the crescent-shaped Hoechst drug is making contacts with the bottom of the minor groove with the two nitrogen atoms (N1 and N3) of the benzimidazole rings forming hydrogen bonds to the AT base pairs. Specifically, the N1 atom of BZI1 ring is forming bifurcated hydrogen bonds to N3 of adenine A6 (3.16 Å) and O2 of thymine T20 (2.78 Å), while the N3 of BZI2 is hydrogen bonded only to O2 of thymine T19 (3.00 Å).

Table 1
Distances(Å) between Hoechst 33258 and DNA Base (< 3.5Å)

H25 N1	- A 6 C2	=	3.26
H25 N1	- A 6 N3	=	3.16*
H25 N1	- T20 O2	=	2.78*
H25 N3	- T19 O2	=	3.00*
H25 C3	- T20 O2	=	3.31
H25 C4	- A 5 C2	=	3.22
H25 C4	- A 6 N3	=	2.85
H25 C4	- T20 O2	=	2.82
H25 C5	- A 5 C2	=	2.97
H25 C5	- A 5 N3	=	3.15
H25 C7	- T20 O2	=	3.25
H25 C9	- T 7 O2	=	3.38
H25 C9	- T19 O2	=	3.16
H25 C3'	- A17 C2	=	3.31
H25 C3'	- A17 N3	=	3.44
H25 C4'	- T 8 O2	=	3.33

* Hydrogen bonds

No other direct hydrogen bond between the drug and DNA is found; rather, many van der Waals interactions are seen. There are 39 intermolecular distances which are less than 3.5 Å These are listed in Table 1. Some of them are of interest, as they may be considered to be weak attractive interactions. For example, the C4 atom of the BZI1 ring is equi-distant from N3 of A6 (2.85 Å) and O2 of T20 (2.82 Å) in the DNA. It has been pointed out that the hydrogen atom on an aromatic benzene ring has a small positive charge which can interact with a negatively charged π cloud of another aromatic ring [10]. Here the aromatic C-H hydrogen may be considered weakly hydrogen bonded to the electronegative O2 and N3 atoms of the DNA base pairs. A similar situation is found for the C9 atom of the BZI2 ring of the Hoechst drug, where it is very close to O2 of T7 (3.38 Å) and O2 of T19 (3.16 Å).

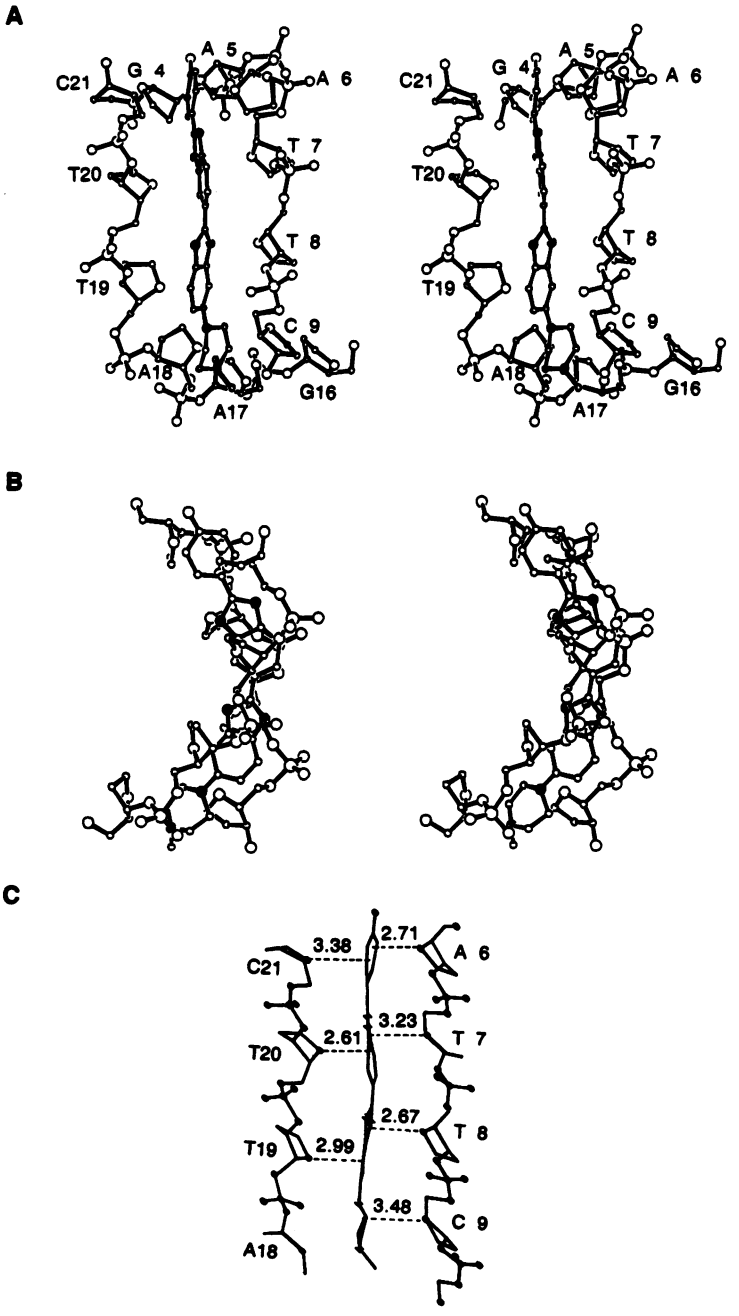
It should be pointed out that there are a number of close contacts between the drug atoms and the adenine C2 atoms in the AATT stretch (see Table 1). If those AT base pairs were replaced by GC base pairs, there would be severe clashes between the N2 amino groups of guanine and the Hoechst drug. This provides an explanation for the AT preference of these minor groove binding compounds.

Another predominant force in stabilizing the drug-DNA complex arises from the interaction between the planar aromatic rings of the Hoechst drug and the side wall

of the narrow DNA minor groove. This is illustrated in detail in Figure 4. In Figures 4a and 4b, the bases have been removed in order to clarify the illustrations. Figure 4a is a view looking directly into the minor groove with the Hoechst molecule in the edge-on orientation. It is apparent here that the two benzimidazole rings have a rather large dihedral angle (32°) between them thus enabling the drug molecule to follow the right-handed curvature of the minor groove of the double helix. A remarkable feature of these interactions is that the Hoechst drug is tucked into the groove so tightly that it fills the space of the groove completely. This is particularly obvious when the drug is shown face-on in Figure 4b where the whole length of the Hoechst molecule is totally sandwiched between the two sugar-phosphate backbones.

Where does the stabilizing interaction come from then? By examining Figure 4a and 4b carefully, one can see that the O4' atoms of the deoxyribose rings all point toward the middle of the minor groove. The positions of these sugar O4' atoms are regularly spaced along the two side walls of the minor groove such that they can interact with the π electron clouds of the aromatic rings of the Hoechst molecule. Figure 4c is a portion of the cylindrical plot of the complex near the drug molecule, labelled with the perpendicular distances between the O4' atoms and the plane of the aromatic rings of the drug. To a first approximation, some of those distances are unequivocally shorter than the sum (3.1 Å) of the van der Waals radius of the oxygen atom (1.4 Å) and the half thickness of an aromatic ring (1.7 Å). In fact, those short distances exist on both sides of the drug and there is no room for the drug to move about in either direction. In other words, the drug is trapped right at the center of the minor groove, in the AATT region.

This type of interaction between the sugar O4' atom and an aromatic planar ring has recently been observed in several nucleic acid structures. For example, in the CpG step of the left-handed Z-DNA structure, the O4' atom of the deoxycytidine residue lies directly over the six-membered ring of the *syn* guanine base [11]. The average value of the vertical O4'-guanine distances in the $d(CG)_3$ Z-DNA structure is 2.9 Å, again shorter than the 3.1 Å mentioned above. We believe that this type of interaction is important in the stabilization of the Z-DNA double helix and that the same general principle is also operating in the complex structure of this minor groove binding drug and DNA.


A
C21
G 4
A 5
A 6
T 7
T20
T 8
T19
C 9
A18
A17
G16
B
C
3.38
2.71
C21
A 6
3.23
2.61
T20
T 7
2.67
T 8
T19
2.99
3.48
C 9
A18

DNA Conformation

The conformation of the double helical B-DNA dodecamer, d(CGCGAATTCGCG), has been extensively examined both alone and complexed with a minor groove binding drug [5, 12]. The overall conformation of all these DNA molecules is fairly similar, which is not totally surprising since they all crystallized in a very similar crystal lattice with the space group $P2_12_12_1$. However, microheterogeneity in conformation associated with individual nucleotides in the double helix is common. For example, the AT base pairs tend to have larger propeller twist angles than the GC base pairs.

In this complex, the middle two AT base pairs have high propeller twist angles (A6--T19=27.7° and T7--A18=19.1°) in comparison with the average value (10.6°) of the other ten base pairs. There is one exception, the G13--C11 base pair, which is also found to have a high (20.8°) propeller twist. If we exclude this GC base pair, the remaining nine base pairs have an average twist angle of 9.8°. The high twist value associated with the C11--G13 base pair is likely to be related to the crystal packing as it has been observed in all the related dodecamers crystallized in this orthorhombic $P2_12_12_1$ crystal lattice.

In the structure of the complex of d(CGCAAATTTGCG) with distamycin, the whole A_3T_3 region adopts an unusual conformation associated with a large positive propeller twist (average=20°). In this conformation, the adenine N6 amino group forms bifurcated hydrogen bonds to two adjacent thymine O4 atoms of the opposite strand within the major groove. This conformation provides a building motif for constructing a modified B-DNA structure (called P-DNA due to the high propeller-twisted base pairs) which incorporates a continuous string of bifurcated hydrogen bonds along the major groove. It has been proposed that the P-DNA conformation may explain the unusual properties of poly(dA).poly(dT) [13, 14].

Figure 4: (A). A stereo view near the Hoechst binding site showing in detail the van der Waals interactions between drug and DNA. The drug is seen edge on and the large twist between the two benzimidazole rings is obvious. The oxygen O4' atoms of the sugar are all pointing toward the drug. The base pairs have been removed for clarity. (B). A view 90° from (A). This view clearly shows the manner in which the binding of the planar drug in the narrow minor groove. (C). A portion of the cylindrical plot of the complex indicates all the close van der Waals contacts between DNA sugar O4' atoms and the drug molecule.

Table 2
Backbone Torsion Angles(degrees) for the d(CGCGAATTCGCG)+Hoechst 33258 Complex*

Base	α	β	γ	δ	ε	ζ	χ	P#
C 1	----	----	-158	168	142	-77	-87	-137
G 2	-140	-165	100	135	-155	-119	-111	180
C 3	-11	164	1	163	-158	-140	-106	170
G 4	-79	172	63	153	-163	-133	-100	-174
A 5	-39	157	42	139	-161	-134	-111	155
A 6	28	-167	-58	180	176	-87	-102	-147
T 7	-59	-163	37	142	-152	-145	-79	158
T 8	40	-153	-85	165	-156	-97	-115	-158
C 9	-57	-179	32	109	-161	-120	-113	112
G10	-12	148	26	125	144	-75	-99	157
C11	-76	-158	56	158	-172	-111	-101	178
G12	-103	-173	76	140	----	----	-97	149
C13	----	----	143	177	-125	-153	-97	-158
G14	-60	128	47	95	151	-57	-128	98
C15	-133	-162	113	95	-136	-140	-133	84
G16	-25	127	47	127	149	-69	-99	144
A17	-74	-134	36	153	117	-39	-91	-162
A18	-122	-141	60	154	-119	-154	-88	161
T19	1	-124	-83	174	-177	-76	-114	-158
T20	-52	-110	-15	173	174	-104	-93	-147
C21	-21	164	2	140	-63	154	-91	159
G22	-72	176	6	171	-88	150	-69	180
C23	-92	170	33	144	-174	-122	-105	169
G24	-58	160	43	88	----	----	-147	0
AVERAGE	-55	-176	23	145	-165	-117	-103	169
S.D.	(48)	(32)	(67)	(27)	(43)	(44)	(17)	(36)

* The torsion angle is defined as

```
                α      β      γ      δ      ε      ζ
O3'-----P-----O5'-----C5'-----C4'-----C3'-----O3'
```

\# The pseudorotation angle.

We have compared the structure of the DNA in the Hoechst-dodecamer complex with that of the netropsin-dodecamer complex and found that basically they are quite similar with a root-mean-square difference in atomic positions of the base atoms only 0.79 Å However, there appear to be significant differences in the fine details, especially regarding the torsion angles of the DNA backbone. However, both structures were solved and refined at medium resolution (~2.2 Å). Inherent

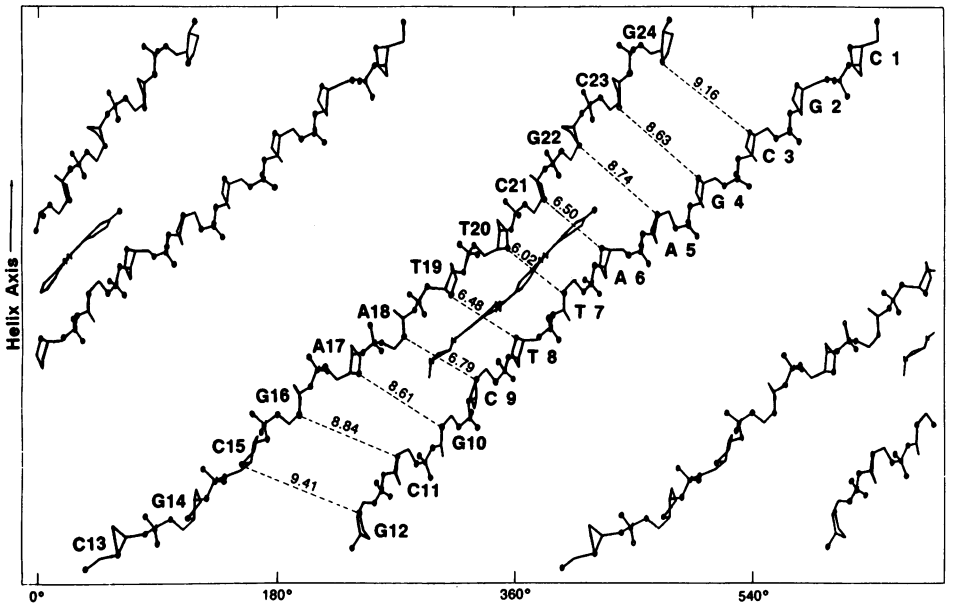


Figure 5: A cylindrical plot of the complex which reveals the narrow minor groove near the central AATT region.

ambiguities exist in the interpretation of the precise location of atoms in such a structure. This is well illustrated in Figure 1B, which shows the best fit of the drug molecule to the electron density of a refined structure at 2.25 Å resolution. It can be seen that the overall shape and orientation of the drug is clearly defined within the boundary of the electron density map, but the location of individual atoms is not very precise.

Table 2 lists all the backbone torsion angles in this structure. Some angles show fairly large variations. For example, the α angle centers around -60°, but a few individual positive values exist as well. Most of the deoxyribose rings adopt pseudorotation angles near 170°, which is associated with the pucker of a typical C2'-endo conformation. However, a number of them (C9, G14, C15) are closer to an O4'-endo conformation. Only the terminal guanine G24 residue adopts a C3'-endo conformation.

A common feature of the B-DNA dodecamers crystallized in this orthorhombic lattice, which include d(CGCGAATTCGCG) alone [12], methylated [8], and

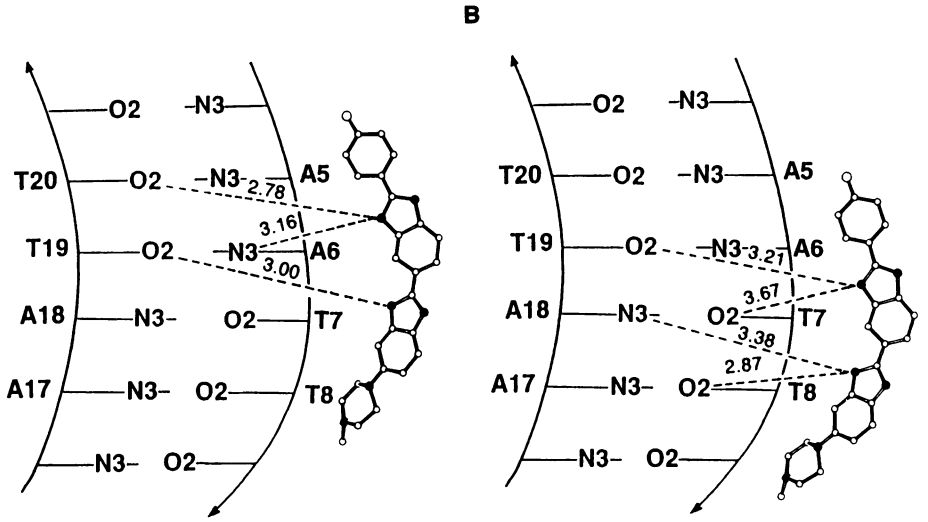


Figure 6: Schematic diagram showing the hydrogen bonding interactions between Hoechst and DNA. (A). This structure. (B). The structure by Pjura *et. al.*[4]. The site of interaction of the drug molecule along the DNA sequence is clearly different in these two structures. In Pjura's structure (B), the drug molecule covers the ATTC base pair region, with the lower half of the molecule existing in two conformations. In order to accommodate this position, and to avoid clashes between the piperazine ring and the N2 amino of the G16-C9 base pair, the piperazine ring is almost perpendicular to the second benzimidazole ring. As explained in the text, this is inconsistent with the anilinic nature of the NH_2 group and also leads to rather long hydrogen bond distances between the drug and the DNA. On the other hand, in the present struture (A), we see only one orientation for the drug molecule, covering the four central AATT base pairs. This requires no awkward ring orientations and leads to the three hydrogen bonds shown.

complexed with netropsin [5],d(CGCGATATCGCG) complexed with netropsin [15]), d(CGCAAATTTGCG) alone and its complex with distamycin [13] and the duplex d(CGCA_6GCG).d(CGCT_6GCG) [14], is the narrow minor groove in the region of the AT base pairs. This is shown clearly in the cylindrical plot of the complex in Figure 5. The width of the minor groove is estimated by the shortest distance between the O4' atoms across the groove, which usually occurs between the O4' atom of a nucleotide from one strand and the opposite O4' atom of the (n + 2) nucleotide on the complementary strand. It can be seen that the central four distances are significantly shorter with an averaged value of 6.45 Å compared to the outer six

distances (average = 8.90 Å). This observation reinforces the idea that the Hoechst drug binds effectively to the narrow minor groove partly due to the favorable environment this region provides for the non-covalent binding mode of the drug. It is interesting to point out that the average groove width of 6.45 Å at the AATT stretch is comparable to the 6.8 Å of the adjacent base pair separation at a base-stacked intercalation site.

This non-covalent binding mode in the minor groove of a B-DNA helix is distinct from the example found in the end-on stacking of the terminal base pairs onto the surface of the shallow but wide minor groove of A-DNA [16]. The latter interaction is used ubiquitously in all the A-DNA oligonucleotides' crystal packing, and it has been suggested as an alternate method, other than intercalation, for large planar aromatic compounds to bind, perhaps non-specifically, to the DNA double helix [16].

Comparison with Other Complexes

The crystal structures of several minor groove binding drugs complexed to DNA oligonucleotides have been solved and analyzed. These include netropsin/d(CGCGAATTCGCG) [5], distamycin/d(CGCA$_3$T$_3$GCG) [13], netropsin/d(CGCGATATCGCG) [15] and Hoechst 33258/d(CGCGAATTCGCG) complexes [4]. The three dimensional structural information has provided a reasonable explanation of the nucleotide sequence binding preference for AT base pairs in the B-DNA helix for minor groove binding drugs. Netropsin and Hoechst 33258 are similar to each other in their overall size and shape, while distamycin is slightly larger. They are all positively charged molecules, however, under conditions of neutral pH, netropsin is doubly charged, while distamycin and Hoechst 33258 are likely each to have only a single positive charge (see Figure 1). This strengthens the attraction of these drugs for the highly negatively charged DNA molecule. The minor groove of B-DNA in an AT-rich region has been suggested to have a large negative charge potential by theoretical analysis [17], thereby providing a favorable binding locus for the drug. This acts as a preliminary process in determining the preferred binding regions on the DNA helix for the drugs.

The further fine tuning of the AT specificity is achieved by a combination of the following three steps. First, the donating NH groups (amide NH from netropsin and

distamycin, benzimidazole NH from Hoechst 33258) form hydrogen bonds (often bifurcated) to the O2 of thymine and N3 of adenine on the concave surface of the minor groove. Second, other atoms of the molecule such as the pyrrole hydrogens from netropsin and distamycin, or phenolic hydrogen from Hoechst 33258, are in close van der Waals contact with the hydrogen atom on the C2 atom of adenine. The corresponding N2 amino group of a guanine base would push the drug away from DNA and dislodge it from the minor groove. Finally, the minor groove width associated with a stretch of AT base pairs appears to be significantly narrower than that of the mixed-sequence DNA region. As described above, the narrow minor groove of the AT region can support the binding of the drug using the stabilizing interactions coming from the polarizable O4' atoms of the deoxyribose along the side wall of the groove.

The present structure of the Hoechst 33258/d(CGCGAATTCGCG) clearly demonstrates that the Hoechst compound possesses all the properties of an AT-rich minor groove binding molecule similar to the naturally occurring drugs, netropsin and distamycin. The crystal structure of the same complex has been studied independently by Pjura *et al.* [4]. Overall, the two structures are similar in the way in which Hoechst binds to DNA. However, when they are compared in detail, significant differences are found as shown schematically in Figure 6. In our structure (Figure 6A), there are three hydrogen bonds between the drug and DNA bases with bifurcated (three-centered) hydrogen bonds from Hoechst N1 to O2 of T20 and N3 of A6, and a single hydrogen between Hoechst N3 and O2 of T19. The drug positions itself evenly in the central AATT stretch.

In contrast, in the structure of Pjura *et al.*(Figure 6B), the Hoechst molecule slides one base pair lower with the piperazine ring now sitting adjacent to the C19-G16 base pair. There are two sets of bifurcated hydrogen bonds, though two of the four bonds are somewhat long. A number of features in Pjura's structure are worth noting. It is stated that a remnant of the spine of hydration remains in the minor groove near the A5 base where the phenol ring in our structure is located. In Pjura's structure, the plane of the piperazine ring is almost perpendicular to the second benzimidazole, which is surprising as the N5 atom actually possesses some aromatic anilinic properties. It would cost extra energy to twist an aniline NH_2 group out of the plane of the benzene ring. Lines are drawn between N5 and N6 nitrogen atoms

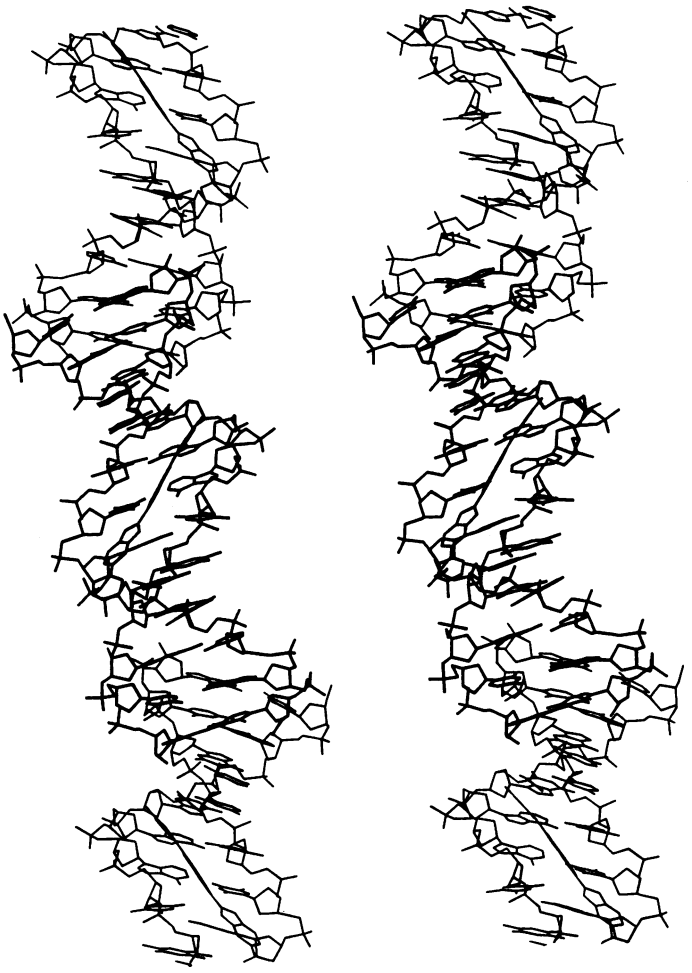


Figure 7: A stereoscopic skeletal drawing showing the crystal packing interactions along the c-axis. The Hoechst drug occupies the minor groove in the central AATT region displacing the spine of hydration. The rest of the minor groove is filled by the interlocking GC base pairs at both ends of the helix leaving very little room for the solvent molecules.

and DNA bases in Figures 2B and 4 of Pjura's paper suggesting possible hydrogen bonds. This does not seem to make chemical sense. The N5 atom is 3.58 Å away from the C9 O2 atom (Figure 5b in [4]) and the N6 atom is 3.96 Å from N2 of G16. Neither can be considered a hydrogen bond because they are too long and have incorrect pairs of donor-acceptor atoms. The piperazine ring is found to occupy two

different positions in Pjura's structure in contrast to the well-resolved fully occupied position in our structure.

We do not know exactly what the sources are of the different observations between these two structures of the same complex. It is possible that a slight variation of the crystallization conditions could result in a different binding mode, although an alternate interpretation of the structure due to experimental procedures (data collection, refinement, etc.) can not be completely ruled out. A full comparison of these structures will require extensive independant refinement using the x-ray structure factor data of the two crystal forms. The data for both structures have been deposited in the Brookhaven Protein Data Bank and therefore, a detailed study can be carried out in the future.

Our structure appears to explain satisfactorily the protection data of the Hoechst drug based on chemical footprinting experiments in which the Hoechst drug is shown to share common binding sites with netropsin and distamycin, though with minor variations. However, there is no need to invoke the wide minor groove associated with the GC base pair. A GC base pair outside the core AATT (or other AT-rich tetranucleotides, see Table 1 in [3]) sequence can be tolerated because Hoechst completely covers the AATT region, but only partially the outer two base pairs. No severe intermolecular clash would occur regardless of whether the outer two base pairs are GC or AT base pairs.

Crystal Packing Interactions

Several related DNA dodecamer molecules, both alone and complexed with minor groove binding drugs, have been crystallized in an isomorphous orthorhombic crystal lattice with space group $P2_12_12_1$. A detailed examination of their packing interactions reveals that the principal interactions involve the terminal two CpG base pair steps at both ends of the molecule. One end of the molecule abuts the other end of the dodecamer in their minor grooves as shown in Figure 7. The terminal GC base pair has its flat surface lying along the sugar-phosphate backbone of one strand of the DNA, much like that observed in the A-DNA crystal packing [16]. A similar situation exists at the other end. In addition, the two guanine bases (G12 and G14) of the last two CG steps of one molecule form tertiary G---G base pairs with two other guanines (G24 and G2 respectively) of the two-fold-screw symmetry-related

molecule along the direction of the crystallographic c-axis. Those G---G base pairs are formed by a pair of hydrogen bonds using the N2 amino group and the N3 atom of its base mate. By virtue of the interlocking tertiary G---G base pairing, the last three base pairs at both ends of the double helix are almost completely blocked from solvent in this minor groove. This is the primary reason why ordered solvent molecules are observed only in the central AT regions in these dodecamer crystal structures.

Similarly, due to the mutual occupation of the minor groove at the two ends, only the space near the central six base pairs in the minor groove is available for drug binding. For example, in this structure the Hoechst drug almost fills the entire space. In the structure of the complex between a longer binding drug distamycin and d(CGCAAATTTGCG), the two ends of distamycin are almost touching the two 3' sugar rings. Therefore it is obvious that any complex of a drug molecule binding site longer than 6 base pairs with a related dodecamer such as d(CGTAAATTTACG) would not be able to crystallize in this lattice.

In this structure, 126 solvent molecules were found. Many of them are fairly ordered and they showed up strongly in the electron density map. In general, they occupy the first hydration shell forming hydrogen bonds to hydrophilic atoms of the base pairs in the major groove or the oxygen atoms of the phosphate groups. In the minor groove, very few first layer hydration water molecules were found due to the binding of the drug as described above. There was one water molecule bridging the Hoechst phenolic hydroxyl group to N2 of G4 and O2 of C21 of DNA in the minor groove. The rest of the solvent molecules are located in the large solvent channel as part of the second shell hydration structure.

ACKNOWLEDGEMENT:

This research was supported by grants from NIH (GM37641) and Burroughs-Wellcome Co. to AHJW. NU was a Natural Sciences & Engineering Research Council (Canada) postdoctoral fellow and CAF was a fellow of the Leukemia Society. We thank Dr. J. Habener and C. Vo for the assistance in the DNA synthesis. We also thank Dr. A. Rich for his encouragement and support, Dr. M. Coll for useful discussions; and Dr. G. Quigley for computer plotting programs.

*To whom correspondence should be addressed

References

1. Lämmler, G., Herzog, H., Saupe, E. and Schutze, H. R. *WHO Bull.* **44**, 751-756(1971).
2. Zimmer, C. and Wähnert, U., *Progr. Biophys. Mol. Biol.* **47**, 31-112(1986).
3. Harshman, K. D. and Dervan, P. B., *Nucleic Acids Research* **13**, 4825-4835(1985).
4. Pjura, P. E.,Grzeskowiak, K. and Dickerson, R. E. *J. Mol. Biol.* **197**, 257-271(1987).
5. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. and Dickerson, R. E. *Proc. Natl. Acad. Sci.* **82**, 1376-1380(1985).
6. Habener, J. F., Vo, C. D., Le, D. B., Gryan, G. P., Ercolani, L. and Wang, A. H.-J. *Proc. Natl. Acad. Sci.*, **in press**,(1988).
7. Greene, P. J., Poonian, M. S., Nusbaum, A. L., Tobias, L., Garfin, D. E., Boyer, H. W. and Goodman, H. M. *J. Mol. Biol.* **99**, 237-261(1975).
8. Frederick, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J. and Rich, A. *J. Biol. Chem.* **submitted**,(1987).
9. Hendrickson, W. A. and Konnert, J. (1979) in Biomolecular Structure, Conformation, Function and Evolution, Srinvasan, R., Eds., pp. 43-57, Pergamon Press, Oxford.
10. Burley, S. K., Wang, A. H.-J. Votano, J. R. and Rich, A. *Biochemistry* **26**, 5091-5099(1987).
11. Wang, A. H.-J., Quigley, G. J., Kopka, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. A. and Rich, A. *Nature* **282**, 680-686(1979).
12. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R. E. *Nature* **287**, 755-758(1980).
13. Coll, M., Frederick, C. A., Wang, A. H.-J. and Rich, A. *Proc. Natl. Acad. Sci.* **84**, 8385-8389(1987).
14. Nelson, H. C. M., Finch, J. T., Luisi, B. F. and Klug, A. *Nature* **330**, 221-226(1987).
15. Coll, M., Aymami, J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J. and Rich, A. *Manuscript in preparation*, (1988).
16. Wang, A. H.-J., Fujii, S., van Boom, J. H. and Rich, A. *Proc. Natl. Acad. Sci.* **79**, 3968-3972(1982).
17. Zakrzewska, K., Lavery, R. and Pullman, B. *Nucleic Acids Res.* **11**, 8825-8839(1983).