Raghunathan, R., Sarai, A., Ting, K.-L., & Jernigan, R. L. (1989) *Biochemistry* (submitted for publication).
Rhodes, D. (1979) *Nucleic Acids Res. 6*, 1805–1816.
Sarai, A., Mazur, J., Nussinov, R., & Jernigan, R. L. (1988a) in *Structure and Expression* (Olson, W. K., Sarma, M. H., Sarma, R., & Sundaralingam, M., Eds.) Vol. 3, pp 213–223, Alan Liss, New York.
Sarai, A., Mazur, J., Nussinov, R., & Jernigan, R. L. (1988b) *Biochemistry 27*, 8498–8502.
Satchwell, S., Drew, H. R., & Travers, A. A. (1986) *J. Mol. Biol. 191*, 659–675.
Schleif, R. (1988) *Science 240*, 127–128.
Simpson, R. T., & Kunzler, P. (1979) *Nucleic Acids Res. 6*, 1387–1415.
Srinivasan, A. R., Torres, R., Clark, W., & Olson, W. K. (1987) *J. Biomol. Struct. Dyn. 5*, 459–496.
Tung, C.-S., & Harvey, S. C. (1986) *J. Biol. Chem. 261*, 3700–3709.
Ulyanov, N. B., & Zhurkin, V. B. (1984) *J. Biomol. Struct. Dyn. 2*, 361–385.
Yoon, C., Privé, G. G., Goodsell, D. S., & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332–6336.
Zhurkin, V. B., Poltev, V. I., & Florentiev, V. I. (1981) *Mol. Biol. (Moscow) 14*, 887–895.
Zhurkin, V. B., Lysov, Y. P., Florentiev, V. I., & Ivanov, V. I. (1982) *Nucleic Acids Res. 10*, 1181–1830.
Zhurkin, V. B., Ulyanov, N. B., & Ivanov, V. I. (1988) in *Structure and Expression* (Olson, W. K., Sarma, M. H., Sarma, R., & Sundaralingam, M., Eds.) Vol. 3, pp 169–190, Alan Liss, New York.

Binding of a Hoechst Dye to d(CGCGATATCGCG) and Its Influence on the Conformation of the DNA Fragment†

Maria A. A. F. de C. T. Carrondo,‡,§ Miquel Coll,‡ Joan Aymami,‡ Andrew H.-J. Wang,‡,∥ Gijs A. van der Marel,⊥ Jacques H. van Boom,⊥ and Alexander Rich*,‡

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and Gorlaeus Laboratory, Leiden State University, 2300 RA Leiden, The Netherlands

Received January 25, 1989

ABSTRACT: Hoechst dye 33258 is a planar drug molecule that binds to the minor groove of DNA, especially where there are a number of A·T base pairs. We have solved the structure of the Hoechst dye bound to the DNA dodecamer d(CGCGATATCGCG) at 2.3 Å. This structure is compared to that of the same dodecamer with the minor-groove-binding drug netropsin bound to it, as well as to structures that have been solved for this Hoechst dye bound to a DNA dodecamer containing the central four base pairs with the sequence AATT. We find that the position of the Hoechst drug in this dodecamer is quite different from that found in the other dodecamer since it has an opposite orientation compared to the other two structures. The drug covers three of the four A·T base pairs and extends its piperazine ring to the first G·C base pair adjacent to the alternating AT segment. Furthermore, the drug binding has modified the structure of the DNA dodecamer. Other DNA dodecamers with alternating AT sequences show an alternation in the size of the helical twist between the ApT step (small twist) and the TpA step (large twist). In this structure the alternation is reversed with larger twists in the ApT steps than in the TpA step. In addition, there is a rotation of one of the thymine bases in the DNA dodecamer that is associated with hydrogen bonding to the Hoechst drug. This structure illustrates the considerable plasticity found in the DNA molecule when it binds to different planar molecules inserted into the minor groove.

There are a number of planar drug molecules that bind to DNA in the minor groove (Zimmer & Wahnert, 1986). These molecules, which include netropsin, distamycin, and the Hoechst dye 33258, all have the property of being formed of repetitions of a structural motif that produce an arc conformation of the planar molecule. This arc matches somewhat the turn of the DNA double helix, and these drugs fit snugly into the double helix in regions in which there are predominantly no G·C base pairs. The N2 amino group of guanine effectively blocks the binding of these drugs both by its presence and by its effect in widening the minor groove of DNA. In the absence of an N2 amino group on guanine, the minor groove has a narrow form that permits these planar molecules to bind using both hydrogen-bonding and van der Waals interactions.

Hoechst dye 33258 (Figure 1) is widely used as a generalized DNA stain, but it is also active an an antihelminthic agent (Lammler et al., 1971). This drug is known to have both a low-level tight-binding mode and a high-level low-affinity binding that are differentiated by their fluorescent characteristics. The interaction of this drug with DNA has been studied extensively. Early work has shown that the Hoechst dye binds preferentially to A·T-rich DNA (Latt & Wohlleb,

† This work was supported by grants from NIH, NSF, NASA, and ONR to A.R. M.A.A.F.C.T.C. acknowledges support from the Junta Nacional de Investigação Científica e Tecnológica. J.v.B. was supported by the Netherlands Organization for the Advancement of Pure Research (ZWO). M.C. was a Fulbright Fellow. J.A. was a NATO Fellow.
* To whom correspondence should be addressed.
‡ Massachusetts Institute of Technology.
§ Present address: Centro de Química Estructural, Instituto Superior Técnico, 1096 Lisboa Codex, Portugal.
∥ Present address: Department of Physiology and Biophysics, University of Illinois at Urbana–Champaign, Urbana, IL 61801.
⊥ Leiden State University.

0006-2960/89/0428-7849$01.50/0

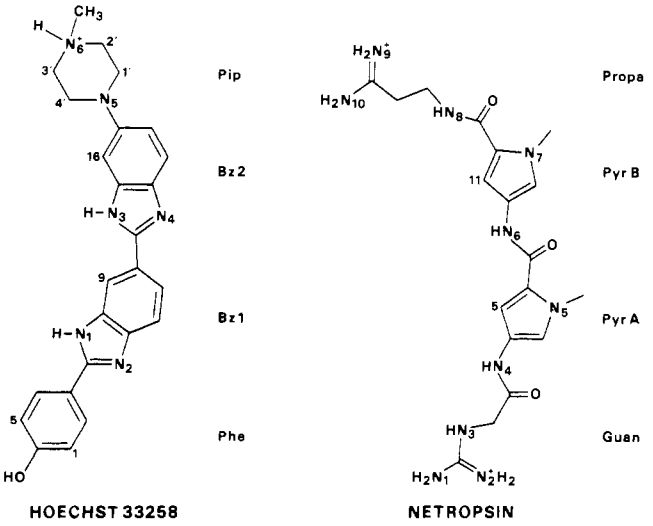

FIGURE 1: Molecular structures of Hoechst dye 33258 and netropsin. Abbreviations: Phe, phenol; Bz1, benzimidazole 1; Bz2, benzimidazole 2; Pip, piperazine; Guan, guanadinium; PyrA, pyrrole A; PyrB, pyrrole B; Propa, propylamidium. Both molecules are drawn in the arc configuration they adopt in DNA complexes. The numbering systems of the Hoechst dye and the drug netropsin are shown with a few of the carbon atom numbers illustrated. The numbering is from bottom to top.

1975; Holmquist, 1975), and a model for a minor-groove-binding mode was proposed (Mikhailov et al., 1981). Footprinting of this Hoechst dye shows that it spans three to five base pairs that predominantly involve A·T base pairs, but there are occasional G·C base pairs at the end of the footprinting sites (Harshman & Dervan, 1985). Similar results have been obtained by using ^{125}I-labeled Hoechst dye 33258, which causes cleavage of the DNA molecule at the binding site (Martin & Holmes, 1983; Murray & Martin, 1988). There have been two X-ray crystallographic studies of the binding of this dye to a DNA dodecamer with the sequence d-(CGCGAATTCGCG) (Pjura et al., 1987; Teng et al., 1988), and they reveal somewhat similar but nonetheless distinctly different binding modes. The dye either straddles the four AATT residues in the center of the molecule symmetrically or is displaced by one base pair so that the piperazine ring is located in the segment in which the minor groove has broadened out because of the presence of the first G·C base pair. We have studied its binding to a closely related dodecamer in which the central AATT sequence is replaced by an ATAT sequence. In the structure of Hoechst dye 33258 complexed to the DNA dodecamer d(CGCGATATCGCG) we have found that the molecule has adopted a conformation distinct from the two that have been seen previously. The DNA dodecamer in all three structures does not have a 2-fold rotation axis perpendicular to the double helix even though the sequence has this symmetry. This is because the molecule has conformational differences at the two ends, largely due to the lattice packing. In the present structure, the drug molecule is bound into the minor groove of the DNA in an orientation opposite to that found by Pjura et al. (1987) and Teng et al. (1988). However, it does not cover the four central bases as found in the structure by Teng et al. (1988) but instead is displaced by one base pair so that the piperazine ring is now located in the broadened minor groove near the first G·C base pair. In addition, the binding of the drug has reversed the alternating twist seen in the DNA molecule previously. The same DNA dodecamer, d(CGCGATATCGCG), has been studied complexed to the drug netropsin (Coll et al., 1989), and it shows an alternation of helical twist between the ApT and TpA base pairs so that there is a succession of smaller and larger twists between the base pairs. The same effect was found in the alternating AT region of the dodecamer d-(CGCATATATGCG) (Yoon et al., 1988). However, in the present structure, the step that formerly had the largest twist angle now has the smaller twist angle and vice versa. This graphically demonstrates the plasticity of the B-DNA conformation when small molecules or drugs are bound to it.

Materials and Methods

The DNA dodecamer d(CGCGATATCGCG) was synthesized by the phosphotriester method and purified by Sephadex G-50 column chromatography. The purity of the final product was checked by HPLC and was found to be greater than 95%. The Hoechst dye 33258 was a gift from Dr. P. Dervan. Crystals were obtained from a solution containing 1 mM DNA dodecamer (single strand), 40 mM sodium cacodylate (pH 7.5), 12.5 mM $MgCl_2$, 4.5 mM spermine tetrachloride, and 2 mM Hoechst dye 33258; 20% 2-methyl-4,4-pentanediol (2MPD) was put into the solution, and it was equilibrated against a reservoir containing 50% 2MPD. After the solution stood at room temperature for 4 weeks, crystals appeared in the solution that had a pale yellow color. The crystals were clear and translucent. Some of the thicker crystals had a darker yellow color. The complex crystallized in the orthorhombic space group $P2_12_12_1$. The unit cell dimensions were a = 25.59, b = 40.56, and c = 67.10 Å. A crystal with the dimensions 0.2 × 0.4 × 0.5 mm was mounted in a glass capillary with a droplet of mother liquor and then sealed. Three-dimensional diffraction data were collected on a Nicolet P3 X-ray diffractometer at 15 °C using an ω-scan mode with copper Kα radiation. Data were collected out to a resolution of 2.3 Å; 3636 reflections were measured, and 1573 reflections between 20 and 2.3 Å were observed above the 2.0σ(F) level and used in the refinement. Decay, Lorentz polarization, and empirical absorption corrections were applied. The unit cell dimensions were very similar to those obtained with related DNA dodecamer crystals, and accordingly trial coordinates of the DNA dodecamer were used from the molecular structure of the netropsin–d(CGCGATATCGCG) complex (Coll et al., 1989). The structure was refined by using the Konnert–Hendrickson constrained refinement procedure (Hendrickson & Konnert, 1981; Westhof et al., 1985). The initial refinement was carried out without the use of the Hoechst drug or any solvent molecules, and an R value of 25% at 2.3-Å resolution was obtained. Some solvent molecules were then added that lowered the R factor to 18% before the drug was added. A difference Fourier map at this stage revealed a large crescent-shaped region of electron density in the minor groove of the DNA dodecamer as shown in Figure 2. The location of the Hoechst dye was determined from the difference Fourier map by using an Evans and Sutherland PS390 graphic system.

In view of the different positions that were found in the previous studies of the Hoechst dye binding, careful attention was addressed to the problem of its orientation. As can be seen in Figure 2, the difference Fourier showed an essentially flat and narrow molecule with a somewhat bulbous end at the upper part. The only nonplanar segment of the molecule is the piperazine ring. It can be seen that it fits the electron density clearly. Attempts were made to fit the drug in an inverted orientation, and independent refinements were carried out for the two different conformations. The conformation shown in Figure 2 refined rapidly to an R value of 15%. Addition of further solvent molecules lowered the R factor to a final value of 13.7%. In the opposite orientation from that

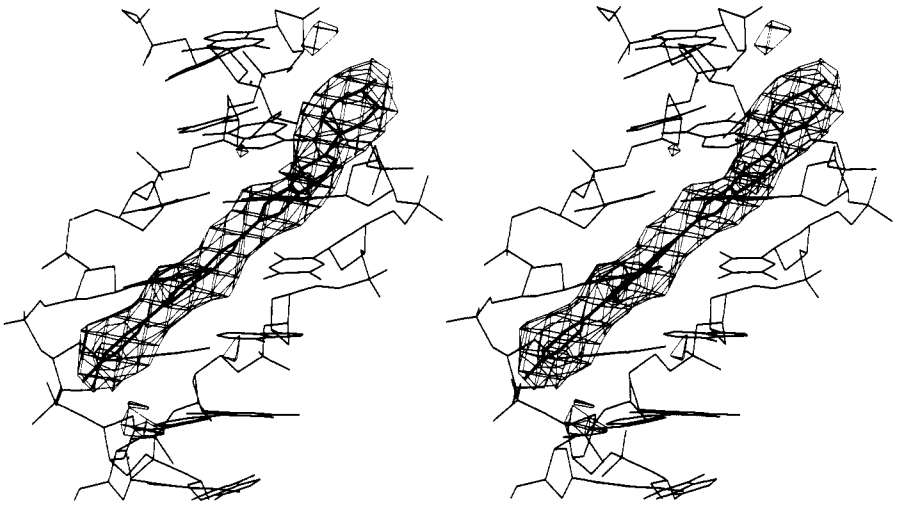

FIGURE 2: Stereo diagram of a final omit map of the region where the drug was located. The drug atoms were omitted, and five cycles of refinement were run before the difference Fourier calculation. The residual electron density map is shown as a meshwork. The position of the Hoechst dye is drawn within that map. In this orientation the helical axis is vertical.

shown in Figure 2, the molecule refined only very slowly to an R value of 15.2%. However, in that orientation it was clear that the NH groups on the imidazole residues failed to make satisfactory hydrogen bonds to electronegative atoms in the base of the minor groove. On the other hand, in the orientation shown in Figure 2, all of the imidazole NH groups made hydrogen bonds. In addition, in this orientation the hydroxyl group in the phenol end was making a hydrogen bond with a water molecule, which also suggested that it was the correct conformation. Furthermore, there was no evidence of disorder in the position of the Hoechst drug once it had been refined, judging from the good fit with the electron density map. Until the final cycles of the refinement there were no constraints imposed between the Hoechst dye and the DNA on either hydrogen-bonding distances or geometry. Only when the hydrogen bonds between the DNA and the drug were well established by the refinement itself were very weak H-bond constraints applied. The DNA had only weak constraints for the hydrogen bonding between the bases. However, as usual, bond length and angle constraints were held fairly tightly so that the geometry of the molecules is quite reasonable. In the final refined structure the root mean square deviation for the bond lengths was 0.018 Å from idealized values and the deviation for the angles was 0.028 Å. The final R factor for the structure of the complex with 112 solvent molecules was 13.7% from only the 1573 reflections at the $2\sigma(F)$ level between 20 and 2.3 Å. No metal ions or spermine were located in the crystal lattice. Atomic coordinates of the complex have been deposited in the Brookhaven protein data bank.

Table I: Backbone[a] and Glycosyl Torsion Angles (in Degrees) for d(CGCGATATCGCG)

residue	α	β	γ	δ	ϵ	ζ	χ	P^b
C1			60	167	−178	−117	−103	190
G2	−42	−139	21	155	151	−93	−95	217
C3	144	−153	−169	89	−138	−94	−161	45
G4	2	−168	−20	174	128	−69	−81	218
A5	−81	−127	43	179	−85	176	−80	193
T6	−107	135	86	89	−166	−110	−146	94
A7	36	176	−39	161	−180	−110	−88	199
T8	−19	−156	2	147	141	−101	−81	195
C9	58	161	−66	132	176	−103	−108	216
G10	40	143	−19	178	−98	137	−61	197
C11	−86	170	43	127	56	115	−117	163
G12	−70	−124	−104	146			−111	221
C13			1	159	−156	−84	−91	186
G14	−56	−47	−104	163	169	−34	−143	254
C15	−120	170	86	92	−133	−126	−136	57
G16	−55	105	88	79	129	4	−148	40
A17	−102	−148	30	163	−105	−167	−100	181
T18	−93	−180	41	140	−169	−120	−103	168
A19	−53	133	97	90	−103	−73	179	86
T20	−23	−54	−114	154	−173	−86	−147	213
C21	84	160	−114	155	−158	−133	−130	206
G22	39	143	−48	162	175	−128	−82	187
C23	−23	−177	15	145	−172	−97	−97	183
G24	−117	156	98	124			−146	134

[a] The backbone torsion angles are defined as O3′—P$\underline{\alpha}$O5′$\underline{\beta}$C5′$\underline{\gamma}$C4′$\underline{\delta}$C3′$\underline{\epsilon}$O3′$\underline{\zeta}$P—O5′. [b] Sugar pseudorotation angle.

Results

The structure of the complex is shown in a skeletal stereo drawing in Figure 3, and the backbone and glycosyl torsion angles are listed in Table I. It can be seen that the drug shown in black is bound to the minor groove of DNA with the two benzimidazole rings almost coplanar. A slight torsion is found in the position of the phenol ring relative to the imidazole rings, and a larger torsion out of the imidazole ring plane is found for the piperazine ring, which is seen in a chair conformation. The phenol ring and the two benzimidazole rings cover three of the four A·T base pairs in the molecule. The piperazine ring is found next to the C·G base pair in the upper part of the diagram. Figure 4 shows approximately the same view with van der Waals models, but now all the hydrogen atoms have been included at their calculated positions. In this diagram the hydrogen atoms of the Hoechst drug are white; all other hydrogen atoms are stippled. This diagram shows the extent to which the Hoechst drug fits tightly into the bottom of the minor groove with a number of van der Waals interactions. The minor groove is seen to be quite narrow in the region in which the phenol and two benzimidazole rings are found. However, the minor groove widens near the top where the positively charged piperazine ring is found.

A clearer view of the overlap between the sugar–phosphate backbone forming the walls of the narrow minor groove and the drug is found in the stereo diagrams of Figure 5, in which only the sugar–phosphate backbone and the drug are shown. Figure 5A is a view down to the plane of the drug. It shows the extent to which the drug molecule is located between the two sugar–phosphate backbones. Figure 5B shows the view perpendicular to the plane of the drug. There are a number of van der Waals contacts between the sugar–phosphate backbone and the drug, primarily involving those atoms of the

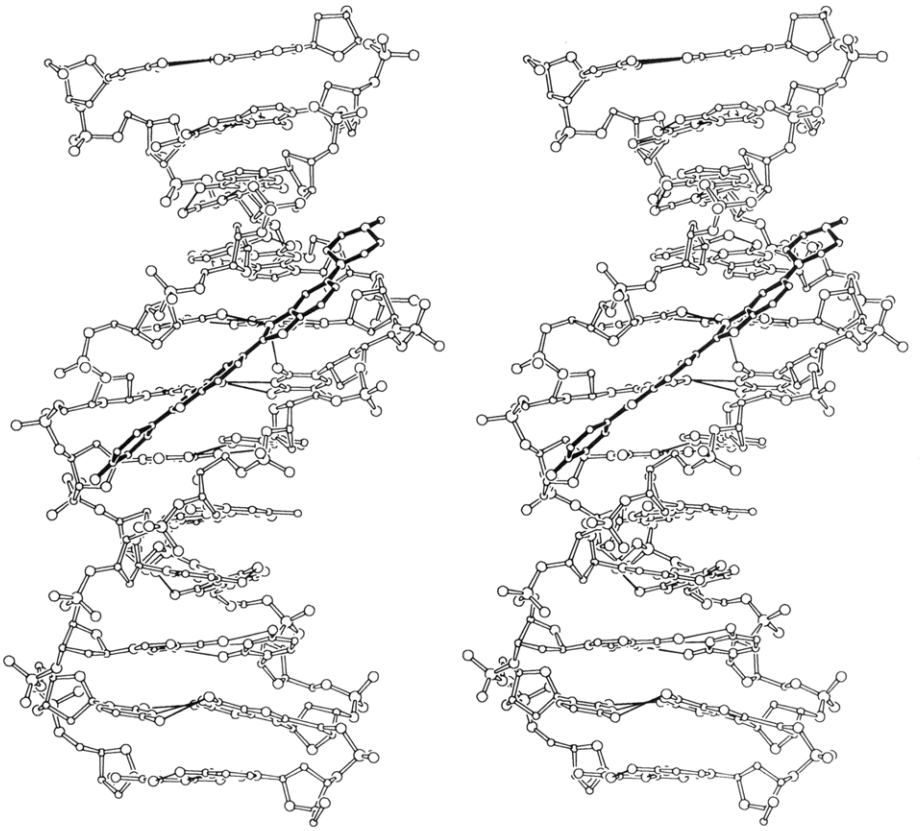

FIGURE 3: Stereo diagram drawn with ball and stick showing the DNA dodecamer and the Hoechst drug with the bonds shaded in black. Hydrogen bonds are shown as single lines. It can be seen that a thymine residue (T6) appears to be rotated somewhat out of the plane of its base pair to form a hydrogen bond with the drug. The drug covers one G·C base pair and three A·T base pairs.

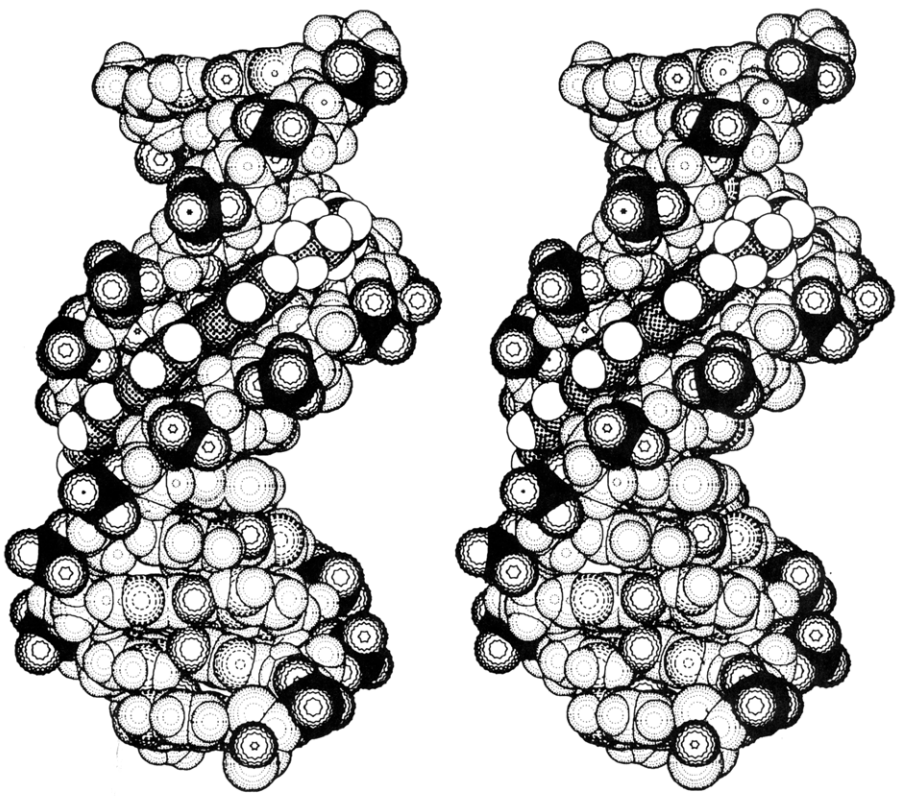

FIGURE 4: Stereo van der Waals drawing of the DNA dodecamer with the Hoechst drug bound in the minor groove. Hydrogen atoms in the Hoechst dye are unshaded, while hydrogen atoms on the DNA dodecamer are lightly stippled. All hydrogen atoms were added at positions calculated geometrically. The phosphorus atoms are shaded black. A denser level of shading is used for the non-hydrogen atoms of the Hoechst dye. It should be noted that the piperazine ring of the drug is at the upper part of the molecule where the minor groove can be seen to widen. The remainder of the Hoechst dye is securely bound by van der Waals interactions associated with the narrow minor groove.

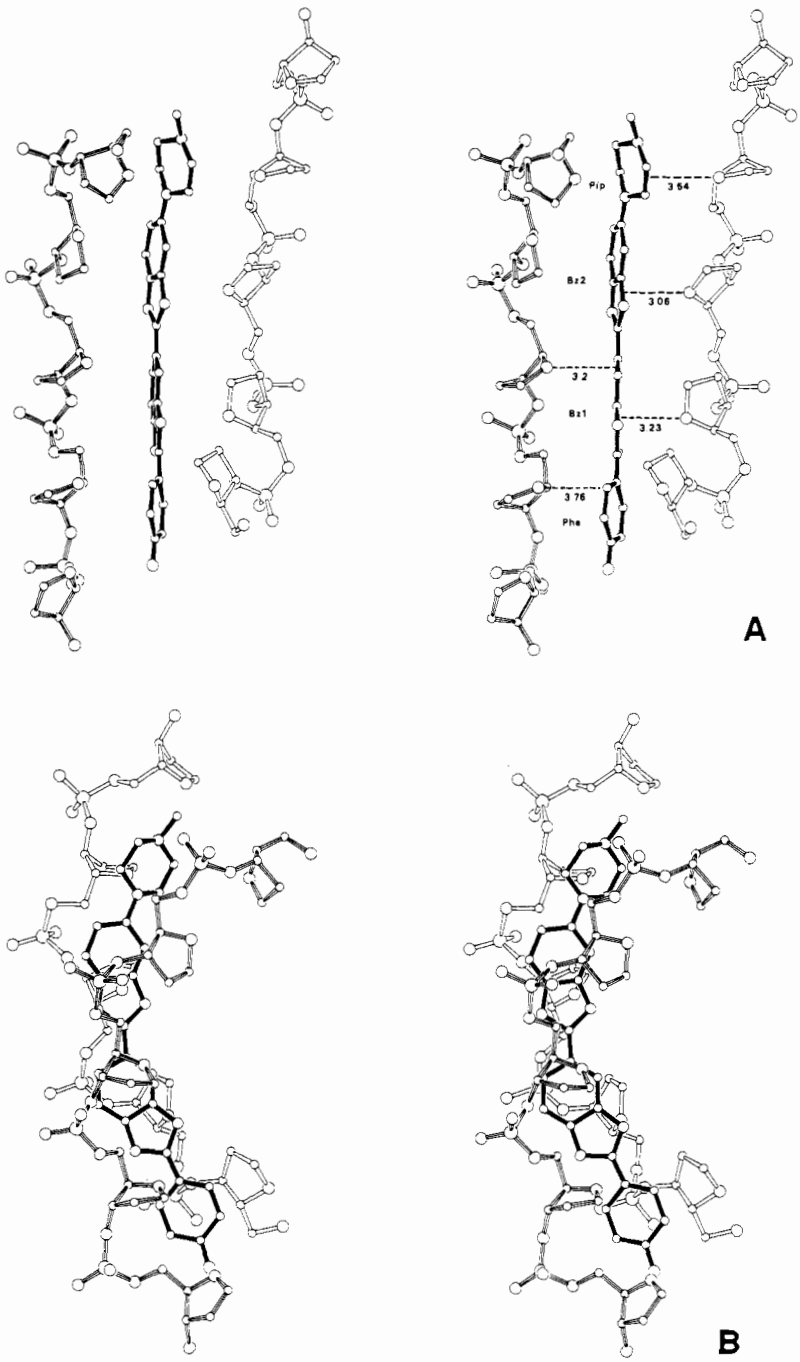

FIGURE 5: Stereo diagrams showing the interaction of the Hoechst drug with the sugar–phosphate DNA backbone in two different orientations. The bonds of the Hoechst dye are shaded in black. (A) In this orientation the view is down into the minor groove with the Hoechst dye seen end-on. The distances indicated are between the O4′ atoms of the sugar residues and the mean plane of the nearby rings. (B) Side view of the same interactions between the sugar–phosphate backbone of the DNA and the Hoechst dye.

drug that are closer to the base pairs rather than the outer atoms removed from the base pair. The major interactions involve the O4′ oxygens of the ribose rings that are in contact with the unsaturated π electrons of the benzimidazole and phenol rings as shown in Figure 5A. This type of stabilizing interaction has been seen in a number of crystal structures, including Z-DNA molecules.

The netropsin molecule, like the Hoechst dye, binds to a similar portion of the minor groove (Coll et al., 1989). Since both of these drugs have been studied bound to the same DNA dodecamer, it is enlightening to compare the modes of binding. Figure 6 shows a comparison of the base stacking and hydrogen-bond interactions for both Hoechst dye 33258 and netropsin in the minor groove. The views are taken down the helical axis, visualizing two base pairs at a time and neighboring segments of the drug. There is a close van der Waals contact between the piperazine ring and the DNA (Figure 6A and Table II), i.e. between C4′ of piperazine and O2 of C21 (3.19 Å). The two hydrogens of the piperazine C4′ are effectively straddling the C4′–O2 close contact. The distance and orientation of this interaction makes it a suitable candidate for bifurcated hydrogen bonds between the CH_2 group and the O2 of C21. Several close contacts between CH groups and electronegative atoms have been found that are belived to involve weak hydrogen bonding (Taylor & Kennard, 1982; Jeffrey et al., 1985). Further down the DNA in Figure 6B, close interactions are found in the junction between benzimidazole 1 and 2. One of these involves a hydrogen bond between N3 of benzimidazole 2 and O2 of T20. As shown in Figure 6C, N3 of benzimidazole 1 is making another hy-

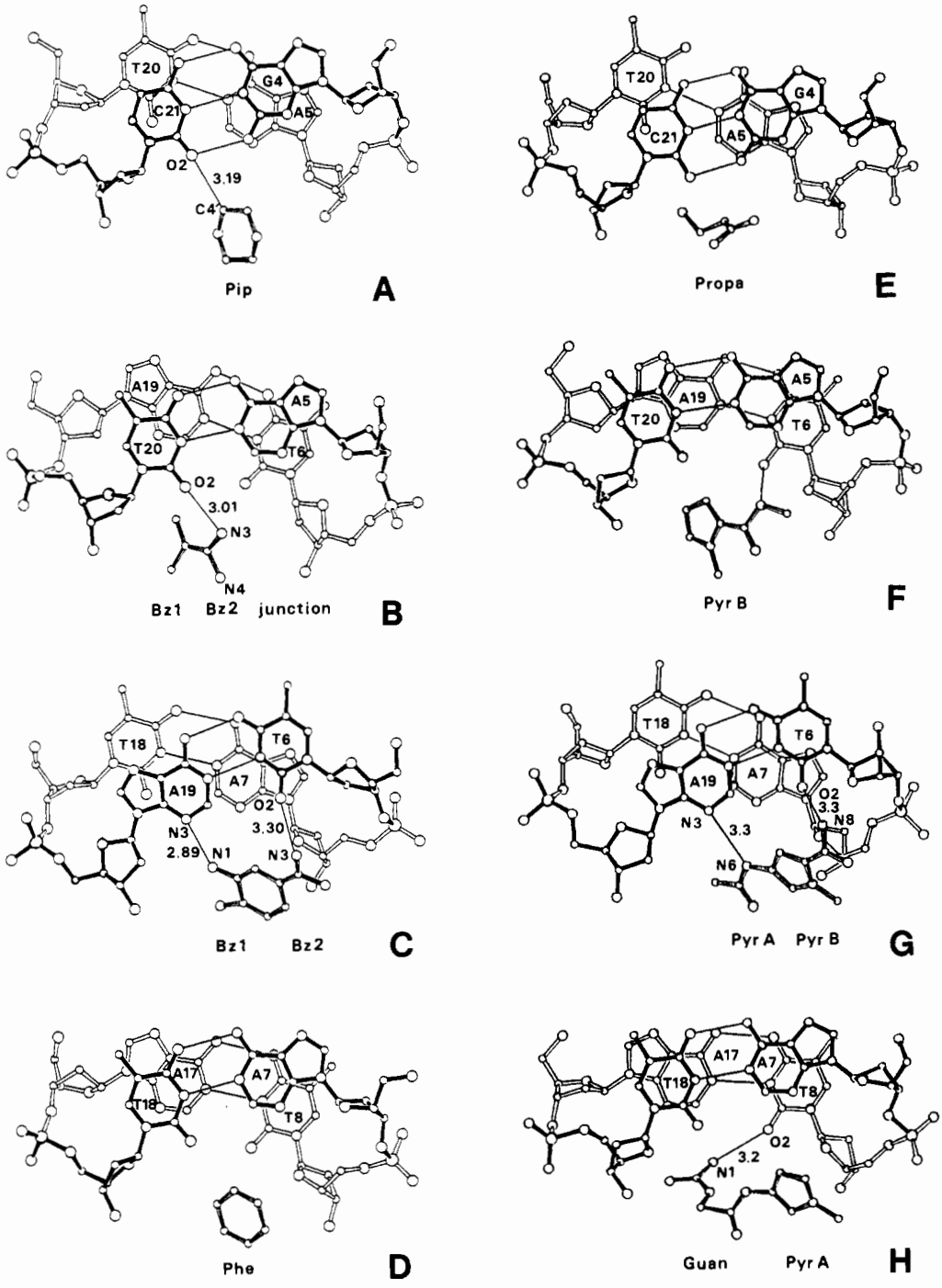

FIGURE 6: Comparison of the base stacking and the interactions between the Hoechst dye and the DNA dodecamer d(CGCGATATCGCG) (A–D) compared with netropsin bound to the same dodecamer (E–H) (Coll et al., 1989). This diagram shows the manner in which the conformation of the DNA molecule is modified by two different groove-binding molecules. Each diagram shows two stacked base pairs in a view down the helix axis. The base pair with solid bonds is situated above the pair with open bonds. At the same time, the segment of the drug interacting with the base pairs is drawn in the minor groove. The same base pairs are illustrated in each horizontal segment of the diagram. Thus the base pairs in (A) and (E) (B) and (F), etc., are identical. There are significant changes in the stacking of the base pairs that are due to the modifications in the DNA structure associated with different molecules binding in the minor groove. Compare, for example, (B) and (F), in which the overlap of the bases has changed considerably. The helical twist in these equivalent ApT steps is 39° in (B) and 33° in (F). Hydrogen-bonding interactions between the drug and the DNA are shown by single lines. The abbreviations for the drugs are those shown in Figure 1.

drogen bond to the O2 of thymine 6, this and the previous one representing a three-centered hydrogen bond. N1 of benzimidazole 1 is also making two close contacts with adenine A19, one of which is a hydrogen-bonding interaction. Finally, the phenol ring also has close interactions with the O2 of T18 with two close van der Waals contacts (Table II).

Figure 6E–H shows the same stacking diagrams for netropsin binding to the same sequence. Careful inspection of the diagrams in Figure 6 reveals some interesting differences in the helical twist between corresponding base pairs. For example, it might be noted that the twist is greater in Figure 6B than in Figure 6F. Likewise, the twist is smaller in Figure 6C compared to that in Figure 6G. These differences in helical twist will be discussed more fully under Discussion.

Figure 7 presents a schematic comparison of the hydrogen-bonding interactions that are found for both Hoechst dye 33258 and netropsin bound to the same DNA dodecamer sequence d(CGCGATATCGCG) (Figure 7A,B) and Hoechst

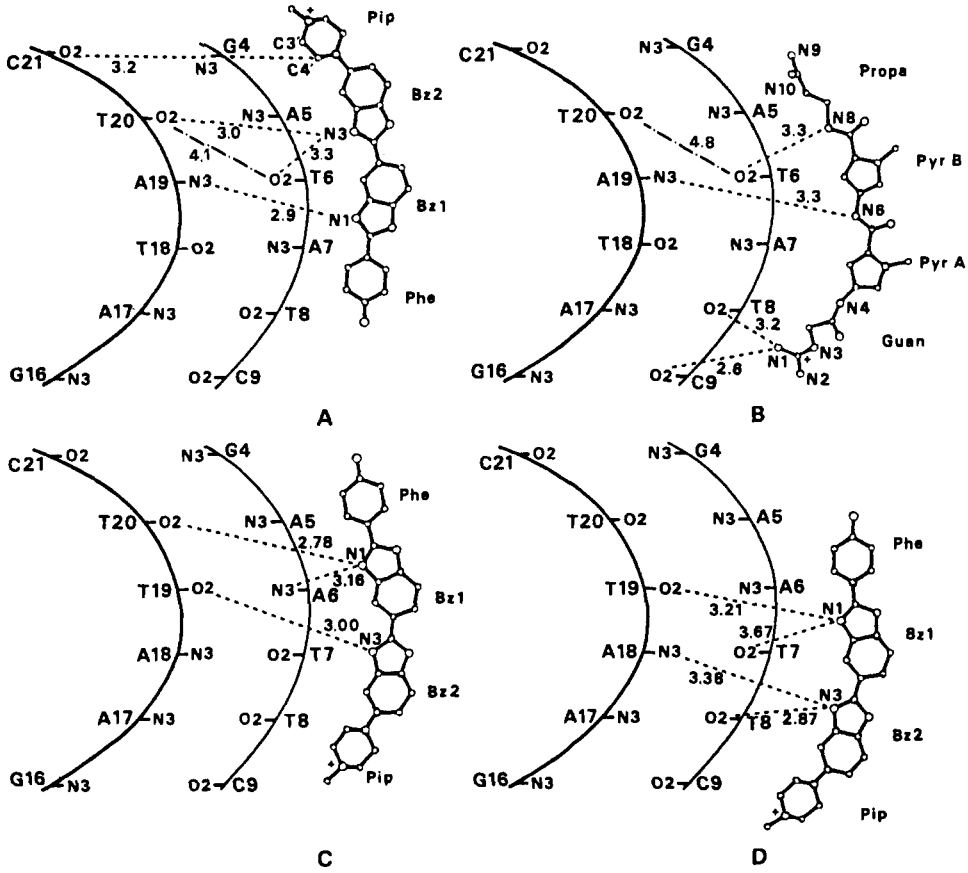

FIGURE 7: Diagrams illustrating the interactions between minor-groove-binding drugs and DNA dodecamers. (A) and (B) show respectively the interaction between the Hoechst dye and netropsin interacting with the same DNA dodecamer d(CGCGATATCGCG). It should be noted that in the Hoechst dye N3 is forming a three-centered hydrogen bond with the O2 of T6 and the O2 of T20. The thymine T6 has twisted around so that the distance between O2 of T6 and O2 of T20 is 4.1 Å. This is in contrast to the interaction between N8 of netropsin (B), which makes only a single hydrogen bond with O2 of T6. Because that T6 is not twisted, the distance between O2 of T6 and O2 of T20 is 4.8 Å. (C) and (D) show the interaction of the Hoechst dye with the dodecamer d(CGCGAATTCGCG). (C) is the structure by Teng et al. (1988) and (D) that by Pjura et al. (1987). It should be noted that the orientation of the drug in these two cases is the opposite of that seen in (A). The hydrogen-bonding interactions between the drug and the DNA dodecamers are shown by dotted lines. All three Hoechst structures [(A), (C), and (D)] have the dye covering four base pairs. However, in (A) the drug covers one G·C base pair and three A·T base pairs, since the piperazine ring is covering the G4·C21 base pair. In (C) the Hoechst molecule covers the four central A·T base pairs, while in (D) it has shifted down compared to (C) and covers three A·T base pairs and the C9·G16 base pair with its piperazine ring.

Table II: Distances for Drug–DNA of 3.5 Å or Less

C3(H25)–O3′(A19)	3.5	C10(H25)–C4′(A7)	3.5
C4(H25)–O2(A18)	3.4	C10(H25)–O4′(TA7)	3.3
C4(H25)–C4′(A19)	3.5	C15(H25)–O4′(C21)	3.2
C4(H25)–C2′(A19)	3.5	N1(H25)–C2(A19)	3.4
C4(H25)–C1′(A19)	3.0	N1(H25)–N3(T19)	2.9[a]
C5(H25)–O2(T18)	3.4	N1(H25)–O4′(T20)	3.1
C5(H25)–C4′(A19)	3.3	N3(H25)–O2(T6)	3.3[b]
C5(H25)–C1′(A19)	3.5	N3(H25)–O2(T20)	3.0[b]
C8(H25)–O4′(T20)	3.3	N3(H25)–O4′(C21)	3.1
C9(H25)–O4′(A7)	3.3	C3′(H25)–N2(G4)	3.5
C9(H25)–C1′(T20)	3.4	C4′(H25)–O2(C21)	3.2[c]
C9(H25)–O2(T20)	3.4		

[a] Single hydrogen bond (two-centered). [b] Bifurcated hydrogen bond (three-centered). [c] CH_2–O bifurcated hydrogen bond.

dye 33258 bound to the d(CGCGAATTCGCG) sequence (Figure 7C,D). In the binding of the Hoechst molecule to d(CGCGATATCGCG) (Figure 7A), it can be seen that there are three hydrogen bonds involving NH···O interactions including the N1 and N3 residues of the two benzimidazole rings. In addition, the diagram includes the CH_2···O2 interaction involving the piperazine ring discussed above. In contrast, the netropsin molecule has four NH···O or NH···N hydrogen bonds (Figure 7B), two involving the nitrogen atoms of the peptide bond N6 and N8 and two from the amino group of the guanidinium group N1 binding to O2 atoms of pyrimidine residues (Coll et al., 1989). The Hoechst molecule has one tight hydrogen bond (2.9 Å) from N1 of the benzimidazole 1 and N3 of A19. In contrast, netropsin has one very tight hydrogen bond between N1 of the guanidinium group and O2 of C9. The other hydrogen bonds are all somewhat weaker with bond lengths generally over 3 Å. It is reasonable to believe that the detailed positioning of the two drugs in the minor groove is mediated to a considerable extent by the geometry and position of the various hydrogen-bond donors. As discussed below, the differences in this detailed positioning account for an interesting conformational modification in the DNA associated with a sharp propeller twisting of thymine 6 in the Hoechst complex.

For comparative purposes, we have included in parts C and D of Figure 7 two different results that were obtained for the binding of the Hoechst drug to the sequence d-(CGCGAATTCGCG). Both structures have been solved to a resolution comparable to that seen in the present structure, and they differ in the detailed positioning of the drug. In the structure by Teng et al. (1988) the drug is seen straddling the AATT segment with the piperazine ring interacting with the A17·T8 base pair (Figure 7C). In contrast, the structure of Pjura et al. (1987) has the drug positioned one base pair down so that the piperazine ring is now covering the base pair C9·G16 (Figure 7D). In the structure by Teng et al. (1988), the benzimidazole N1 makes bifurcated hydrogen bonds with O2 of T20 and N3 of A6. In contrast, Pjura et al. (1987) find

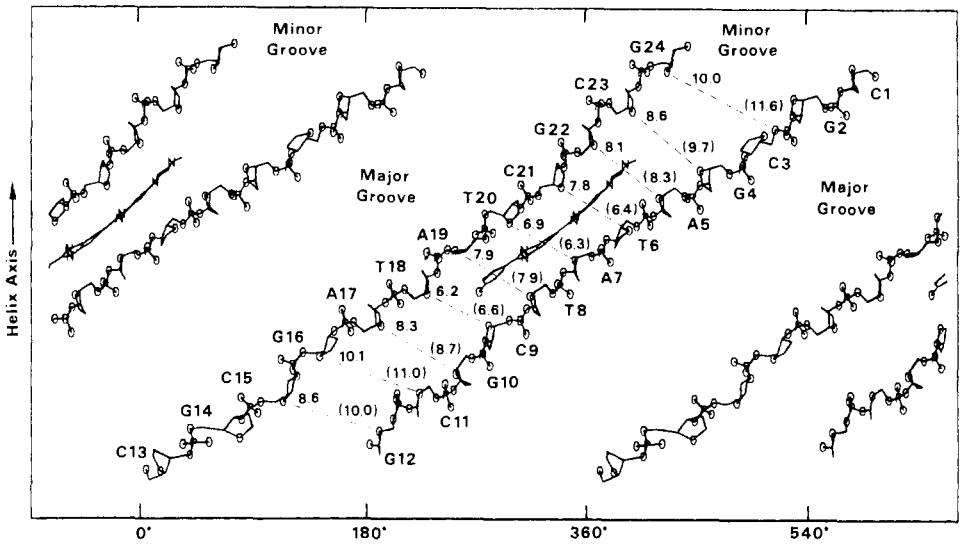

FIGURE 8: Cylindrical projection of the DNA dodecamer–Hoechst dye complex. This diagram is made by constructing a cylinder around the molecule at a radius of 8 Å. All atoms are projected onto the cylinder, and the cylinder is then cut parallel to the axis and flattened out. The right-handed helix can be seen because the molecule goes from the lower left to the upper right with the major groove and minor grooves indicated. The distance across the minor groove is indicated by the dotted line, and it represents the distance from O4′ of one sugar atom to O4′ of the sugar across the minor groove. The upper numbers represent the distances in angstroms in the present structure. The lower numbers in parentheses represent the distances found in this molecule when complexed to netropsin. It can be seen in both molecules that there is a considerable decrease in the O4′–O4′ distance in the region of the molecule in which there are A·T base pairs and the narrow groove becomes considerably wider when G·C base pairs are found. The position of the Hoechst drug is found in the narrowest part of the minor groove. The piperazine part of the drug is at the top, where the minor groove is beginning to widen out.

N1 making a hydrogen bond with O2 of T19. In a similar fashion, in the Teng structure N3 of benzimidazole 2 forms a single hydrogen bond with O2 of T19, while in the Pjura structure N3 forms bifurcated hydrogen bonds with N3 of A18 and O2 of T8. A comparison of the hydrogen-bonding potential of the three Hoechst complexes shows that the predominant interactions consist of hydrogen bonds from the NH groups of the benzimidazole residues to the electronegative atoms at the floor of the minor groove, either N3 of adenine or O2 of thymine in the minor groove. These may either be bifurcated or single hydrogen bonds depending on the detailed positioning of the drug and the adaptation of the DNA to the interactions with the drug. What is immediately clear in this figure is that, from a comparison of the three Hoechst structures, the two interactions found by the Hoechst drug to the same DNA sequence by both Teng and Pjura involve the drugs with the same orientation but with a different relative position. However, in the present structure, we see that the drug has an inverted orientation relative to the sequence in that the piperazine ring in Figure 7A is at the top of the diagram, whereas in parts C and D of Figure 7 it is at the bottom. In our case, as in the case of the structure by Pjura et al., the piperazine ring is in the C·G base pair region immediately outside the central four A·T base-pair region.

Another way of visualizing the preference of groove-binding molecules for a particular DNA stretch is a cylindrical plot in which the atomic positions are projected onto a cylinder at a particular radius. The cylinder is then cut, and the positions of these molecules then appear as diagonals running across the sheet. An example of this is shown in Figure 8. In it, the Hoechst drug is shown lying in the middle of the minor groove. The O4′–O4′ distances across the minor groove are indicated for the same sequence bound to either the Hoechst drug or netropsin (in parentheses). It can be seen that the distance is shorter in the segments of the minor groove where A·T base pairs are present. This narrowing of the minor groove is found in the segments containing A·T base pairs of all DNA dodecamers solved to the present irrespective of the sequence. Thus, it is equally present in the sequence $d(A)_n$·$d(T)_n$ (Wing et al., 1980; Coll et al., 1987; Nelson et al., 1987; DiGabriele et al., 1989) or in $d(AT)_n$ (Coll et al., 1989; Yoon et al., 1988). Figure 8 clearly shows that the positively charged piperazine is located near base pair C21·G4 where the minor groove starts to widen.

We might ask whether the difference in the disposition of the drug molecules and its positive charges at all influences the distance between neighboring phosphate groups. Accordingly, we have calculated the closest distance between phosphate groups on either side of the minor groove. This calculation has been made for the same sequence d-(CGCGATATCGCG) involving both the Hoechst drug and netropsin complexes (Figure 9A) and the two d-(CGCGAATTCGCG) Hoechst complexes (Figure 9B). It can be seen that the shape of the curve generally is similar for all the complexes, that is, with a shorter phosphorus to phosphorus distance across the minor groove in the central region where the A·T base pairs are found and the drugs are situated. A detailed inspection reveals some differences. For example, the P–P distance seems to be shorter when a positively charged group of the drug is present in that particular region. In Figure 9A, residues 10 and 19 and residues 11 and 18 are about 1 Å closer when the guanidinium group is between them. This is in contrast with the diagram of Figure 9B, where the P–P distance is wider for residues 11 and 18 when the piperazine is located between them. However, in the latter structure, the piperazine ring is not coplanar with the rest of the drug rings. Therefore, different steric interactions may be involved in this case.

Discussion

In this study we have had the opportunity to analyze in some detail the manner in which two similar minor-groove-binding drugs, Hoechst dye 33258 and netropsin, can bind to the same DNA sequence and bring about changes in the DNA due to this interaction. Some of these changes are modest in size but some are rather large. Most of them can be understood by a detailed analysis of the differences between the positions of the various components of the drugs relative to the DNA

Table III: Propeller Twist (in Degrees)[a]

base pair	ATAT·Hoe[b]	ATAT·net[c]	AATT[d]	AAATTT·dis[e]	AAAAAA[f]
1–24	5	7	13	11	19
2–23	16	14	12	14	12
3–22	19	11	6	18	8
4–21	19	12	12	15	15
5–20	10	6	17	16	23
6–19	21	16	18	21	26
7–18	10	16	17	19	23
8–17	9	8	17	19	18
9–16	23	10	17	25	19
10–15	14	14	5	11	11
11–14	19	20	18	12	15
12–13	5	3	4	14	6
av	14	11	13	16	16
av AT	13	11	17	19	20

[a] The boxes include the AT stretches. [b] d(CGCGATATCGCG)·Hoechst dye (this work). [c] d(CGCGATATCGCG)·netropsin (Coll et al., 1989). [d] d(CGCGAATTCGCG) (Drew et al., 1981). [e] d(CGCAAATTTGCG)·distamycin (Coll et al., 1987). [f] d(CGCAAAAAAGCG)·d(CGCTTTTTTGCG) (Nelson et al., 1987).

sequence. The interaction has three major components. First, the detailed position of hydrogen bonds that are often bifurcated (three-centered) because of the quasi twofold axis of symmetry between O2 of thymine and N3 of adenine in the floor of the minor groove. The second component is van der Waals interactions, which include close contacts to both the base pairs in the floor of the minor groove and the sugar–phosphate walls. In particular, short contacts are present between the O4′ atoms of the deoxyriboses and the planar rings of the drug molecules often involving an interaction with the π electron system. The third component is electrostatic, associated with positive charges in the drug molecule.

All of these interactions can be seen to modify in subtle ways the conformation of the DNA. One example of this modification is shown in Figure 7A. In the Hoechst drug, N3 of benzimidazole 2 makes two hydrogen bonds and forms a three-centered or bifurcated hydrogen bond with O2 of T20 and O2 of T6 (Figure 7A). In contrast, the NH group on netropsin at N8 forms a hydrogen bond only with O2 of T6; it is too far away from O2 of T20 to form a hydrogen bond (Figure 7B). Since the N8–O2 and N3–O3 distances are the same (3.3 Å), we may ask why there is not a secondary bond between N8 of the netropsin molecule and O2 of T20. The answer is found in Table III, which lists the propeller twists between base pairs for a number of DNA dodecamers either by themselves or with drugs bound to them. It is interesting to compare the propeller twists in the AT region for the Hoechst dye and the netropsin drugs bound to the same sequence. Comparison of the two lists of propeller twist shows that one base pair, T6·A19, has a very high propeller twist of 21° compared to the significantly lower values of the other three base pairs in that sequence. Significantly, the propeller twists for the netropsin complex do not show any twists in the range near 20°. As shown in the last two columns of Table III, DNA containing A-tracts or oligo-A sequences has a high propeller twist in the A·T base pairs (Coll et al., 1987; Nelson et al., 1987; DiGabriele et al., 1989; Aymami et al., 1989). However, one might ask how it is possible to get a propeller twist of 21° in a sequence involving alternating ATAT, since it is known that there are considerable constraints involving clashes, especially with twists of the adenine residue (Coll et al., 1989; Yoon et al., 1988). The answer to this is shown in Figure 3, where the bifurcated hydrogen bonds from N3 of benzimidazole 2 are shown to both O2 of T20 and O2 of T6. It can be seen that the thymine residue T6 is turned sharply away from the plane of its Watson–Crick paired adenine 19. Because of this twist of the thymine residue, the N3 of the benzimidazole ring is able to make a hydrogen bond both to it and to O2 of thymine 20. A comparison of the O2(T6)···O2(T20) distance in the Hoechst and netropsin structures (Figure 7A,B) shows that in the Hoechst complex these two atoms are pulled together (4.1 Å), allowing the formation of the three-centered H-bond which was not possible in the netropsin complex (4.8 Å). In this case it is reasonable to believe that the other hydrogen-bonding interactions and van der Waals contacts drive the Hoechst drug into its present position and, as a consequence of having N3 of benzimidazole 2 in its vicinity, some stabilization energy is gained by the rotation of thymine 6 out of the plane. From Figure 3 it is clear that only the thymine residue twists out of the plane to produce the high propeller twist since the adenine residue is largely coplanar with bases flanking it. In the case of the netropsin complex, the drug is positioned somewhat lower in the helix.

Table IV: Helical Twist in Alternating AT (in Degrees)

step	ATAT·Hoe[a]	ATAT·net[b]	ATATAT[c]	model[d]
ApT	39, 36	33, 27	33, 31, 33	33
TpA	31	38	39, 42	39

[a] d(CGCGATATCGCG)·Hoechst dye (this work). [b] d(CGCGATATCGCG)·netropsin (Coll et al., 1989). [c] d(CGCATATATGCG) (Yoon et al., 1988). [d] Poly(AT) model (Klug et al., 1979).

Other conformational changes in the DNA are illustrated by changes in the helical twist (Table IV). As mentioned in the description of Figure 6, there are significant differences seen in the helical twist from one base pair to the other in comparing the netropsin and Hoechst complexes. Some time ago it was suggested that the sequence TpA would adopt a larger twist near 40°, while the sequence ApT would have a smaller twist of near 30° (Klug et al., 1979). It can be seen that this alternating AT conformation is adopted by a DNA dodecamer that has $(AT)_3$ in the middle of it (Yoon et al., 1988) and is also adopted for the central region of the netropsin complex (Coll et al., 1989). However, it is interesting that the presence of the Hoechst dye has resulted in a near reversal of these twists, i.e., larger twist in the ApT step than in the TpA step. Thus, the sequences A5pT6 and A7pT8 now have helical twists of 39° and 36°, respectively, while the sequence T6pA7 has a helical twist of 31°. This illustrates the relative plasticity of the DNA molecule in accommodating conformational changes associated with the binding of a small molecule. In the netropsin case, the binding fits the alternating conformation that is found in the native molecule, while binding of the Hoechst molecule produces an inversion of that alternation. The reason for this is likely to be associated with the differences in the detailed van der Waals interactions of the two drug molecules that have different shapes and, accordingly, different close contacts with the sugar–phosphate backbones. A significant part of this may be associated with the fact that the Hoechst molecule is relatively inflexible, that is, it consists of four fairly rigid units connected by three bonds that can rotate only between them. In contrast, the netropsin molecule has two smaller rings connected by a peptide bond and flanked by two flexible segments that probably can adapt themselves better to the changes in DNA twist associated with the native molecule.

It is interesting to note here that the three different positions found for the Hoechst drug bound to DNA as described in Figure 7 are associated with small changes in the dihedral angle between the planes of the four rigid units making up the Hoechst dye. Thus, the dihedral angle between the phenol

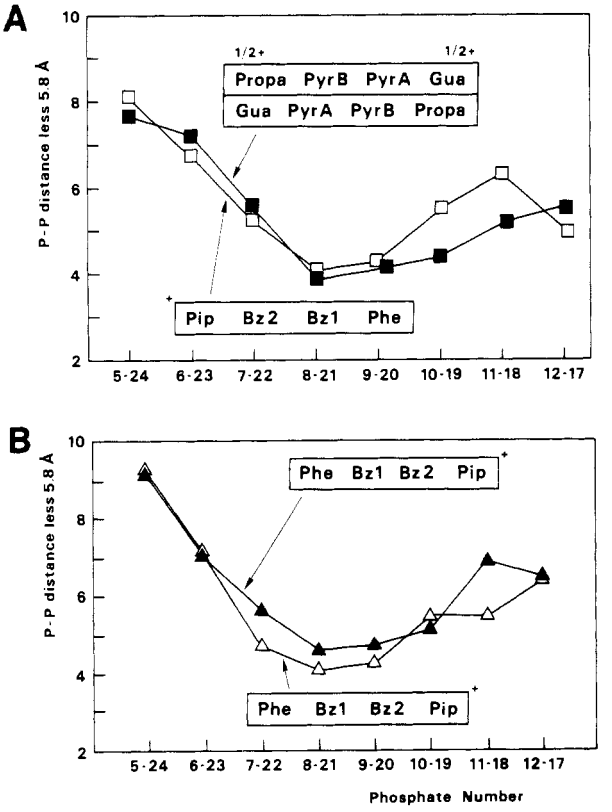

FIGURE 9: (A) Distance between nearest phosphorus atoms across the minor groove plotted for the two dodecamer structures with the alternating ATAT segment and the Hoechst and netropsin drugs, respectively. This distance is plotted minus 5.8 Å, which represents the sum of the van der Waals radii for the two phosphate groups. The position of the Hoechst drug is shown at the bottom of the diagram, and the open squares correspond to the phosphorus–phosphorus distances in the present structure. The position of the disordered netropsin molecule is shown in the upper part of the diagram, while the phosphorus–phosphorus distances for the corresponding dodecamer are plotted as solid squares. It can be seen that the distances are fairly similar except at the right side of the diagram, where there is a significantly shorter distance between phosphates 10 and 19 and phosphates 11 and 18 in the dodecamer with the netropsin molecule compared to the same distances with the Hoechst drug. It is possible that this close distance is associated with the half-positive charge found on the guanidinium group of the netropsin molecule that is located in this region, whereas nothing is found for the Hoechst dye in that region. (B) Phosphorus–phosphorus distances for the dodecamer with the AATT sequence for the two different Hoechst complexes solved. The upper box shows the position of the Hoechst drug in the structure by Pjura et al. (1987), while the bottom box shows the position of the drug found in the structure by Teng et al. (1988). The distances between the phosphate groups appear to be similar with the exception of the distance between phosphates 11 and 18, where the Teng structure has a somewhat smaller distance.

group and benzimidazole 1 is 18° in the present structure, 8° in the structure reported by Teng et al. (1988), and near zero in the Pjura et al. (1987) structure. Alternatively, the dihedral angle between benzimidazoles 1 and 2 is 13° in the present structure and 32° and 36° reported by Teng et al. (1988) and Pjura et al. (1987), respectively. Finally, the dihedral angle between benzimidazole 2 and the mean plane of the piperazine ring is 21° in the present structure, while Teng et al. (1988) report 14°, and nearly 60° is found in the structure reported by Pjura. The differences in the latter figures are readily understood by considering the position of the piperazine ring. In the structure reported by Teng et al. (1988) the piperazine ring is sitting in the narrow groove near the A·T base pairs; consequently, it lies in a position almost coplanar with benzimidazole 2. In contrast, in the present structure as well as in the structure reported by Pjura et al. (1987) the piperazine ring is located in the wider groove near the adjacent C·G base pair. In the wider groove, the ring can rotate away from the plane of the benzimidazole ring. That rotation is moderate in the present structure and large in the structure reported by Pjura et al.

The differences in the orientation of these four planar units in the Hoechst dye in the three different structures illustrate the fact that the manner in which the drug interacts with the DNA depends critically on the detailed positioning of the drug relative to the sequence. Small changes in position are associated with significant alterations in the dihedral angles between the components of the Hoechst drug.

This study has provided us with a detailed view of how DNA dodecamers with the same sequence binding to different drugs are modified by the interaction with the drug. Although the overall stabilizing features are the same in these interactions, the details of the conformation differ considerably from one structure to the other. This shows the plasticity that is inherent in the structure of B-DNA; that is, it can undergo various local structural modifications when it interacts with other molecules.

REFERENCES

Aymami, J., Coll, M., Frederick, C. A., Wang, A. H.-J., & Rich, A. (1989) *Nucleic Acids Res.* (in press).

Coll, M., Frederick, C. A., Wang, A. H.-J., & Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385–8389.

Coll, M., Aymami, J., van der Marel, G. A., van Boom, J. H., Rich, A., & Wang, A. H.-J. (1989) *Biochemistry 28*, 310–320.

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761–786.

DiGabriele, A. D., Sanderson, M. R., & Steitz, T. A. (1989) *Proc. Natl. Acad. Sci. U.S.A.* (in press).

Harshman, K. D., & Dervan, P. (1985) *Nucleic Acids Res. 13*, 4825–4835.

Hendrickson, W. A., & Konnert, J. H. (1981) in *Biomolecular Structure, Conformation, Function and Evolution* (Srinvasan, R., Subramanian, E., & Yathindra, N., Eds.) Vol. 1, pp 43–47, Pergamon, New York.

Holmquist, G. (1975) *Chromosoma 49*, 333–356.

Jeffrey, G. A., Maluszynska, H., & Mitra, J. (1985) *Int. J. Biol. Macromol. 7*, 336–348.

Klug, A., Jack, A., Viswamitra, M. A., Kennard, O., Shakked, Z., & Steitz, T. A. (1979) *J. Mol. Biol. 131*, 669–680.

Lammler, G., Herzog, H., Saupe, E., & Schutze, H. R. (1971) *WHO Bull. 44*, 4751–756.

Latt, S. A., & Wohlleb, J. C. (1975) *Chromosoma 52*, 297–316.

Martin, R. F., & Holmes, N. (1983) *Nature 302*, 452–454.

Mikhailov, M. V., Zasedatelev, A. S., Krylou, A. S., & Gurskii, G. U. (1981) *Mol. Biol. 15*, 690–705.

Murray, V., & Martin, R. F. (1988) *J. Mol. Biol. 201*, 437–442.

Nelson, H. C. M., Finch, J., Luisi, B. F., & Klug, A. (1987) *Nature 330*, 221–226.

Pjura, P. E., Grzeskowiak, K., & Dickerson, R. E. (1987) *J. Mol. Biol. 197*, 257–271.

Taylor, R., & Kennard, O. (1982) *J. Am. Chem. Soc. 104*, 5063–5070.

Teng, M., Usman, N., Frederick, C. A., & Wang, A. H.-J. (1988) *Nucleic Acids Res. 16*, 2671–2690.

Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.
Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. (1980) *Nature 287*, 755–758.
Yoon, C., Prive, G. G., Goodsell, D. S., & Dickerson, E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332–6336.
Zimmer, C., & Wahnert, U. (1986) *Prog. Biophys. Mol. Biol. 47*, 31–112.

NMR Studies of Triple-Strand Formation from the Homopurine–Homopyrimidine Deoxyribonucleotides $d(GA)_4$ and $d(TC)_4$†

Ponni Rajagopal and Juli Feigon*

Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Los Angeles, California 90024

Received March 9, 1989; Revised Manuscript Received May 25, 1989

ABSTRACT: The complexes formed by the homopurine and homopyrimidine deoxyribonucleotides $d(GA)_4$ and $d(TC)_4$ have been investigated by one- and two-dimensional ^1H NMR. Under appropriate conditions [low pH, excess $d(TC)_4$ strand] the oligonucleotides form a triplex containing one $d(GA)_4$ and two $d(TC)_4$ strands. The homopurine and one of the homopyrimidine strands are Watson–Crick base paired, and the second homopyrimidine strand is Hoogsteen base paired in the major groove to the $d(GA)_4$ strand. Hoogsteen base pairing in GC base pairs requires hemiprotonation of C; we report direct observation of the C^+ imino proton in these base pairs. Both homopyrimidine strands have C3′-endo sugar conformations, but the purine strand does not. The major triplex formed appears to have four TAT and three CGC^+ triplets formed by binding of the second $d(TC)_4$ strand parallel to the $d(GA)_4$ strand with a 3′ dangling end. In addition to the triplexes formed, at least one other heterocomplex is observed under some conditions.

Stretches of homopurine–homopyrimidine sequences are a widely dispersed component of eukaryotic genomes. Several lines of evidence have indicated that such sequences may adopt alternative (noncanonical B-DNA) structures and that these structures may act as regulatory signals in gene expression [for a review, see Wells et al. (1988)]. It has also been suggested that homopurine–homopyrimidine sequences may play a role in telomer formation (Blackburn, 1984), chromosome folding (Johnson & Morgan, 1978), and recombination (Hoffman-Liebermann et al., 1986; Konopka et al., 1988; Wohlrab et al., 1987). Homopurine–homopyrimidine sequences frequently flank transcribed eukaryotic genes (Wells et al. 1988). Evidence for an unusual structure for these sequences includes a pH and supercoil density dependent hypersensitivity to DNase I and single-strand-specific nucleases (Elgin, 1984; Cantor & Efstratiadis, 1984). On the basis of this and other evidence, Pulleyblank and co-workers proposed a model for $d(TC)_n$·$d(GA)_n$ in supercoiled plasmids in which Watson–Crick AT base pairs alternate with Hoogsteen $^{syn}GC^+$ base pairs (Pulleyblank et al., 1985). A currently more widely accepted model is that of Frank-Kamenetskii and co-workers, in which mirror repeat homopurine–homopyrimidine sequences loop out to form a pyr–pur–pyr (homopyrimidine–homopurine–homopyrimidine) triplex and a single-stranded homopurine stretch, termed H-DNA (Mirkin et al., 1987). Several recent studies support this model (Johnson, 1988; Htun & Dahlberg, 1988). Christophe et al. (1985) have also proposed an intramolecular triplex model.

Triplex formation has also been demonstrated in vitro by Moser and Dervan (1987) and by Praseuth et al. (1988). The former synthesized homopyrimidine oligodeoxyribonucleotides with attached EDTA–Fe cleaving agents and found binding to corresponding homopurine–homopyrimidine tracts in DNA oligonucleotides and a restriction fragment. This work has generated great interest because of the potential applications of such sequences as highly sequence-specific probes for use in chromosome mapping (Strobel et al., 1988) and possible medical applications as antisense DNA.

Evidence for triplex formation in RNA and DNA homopolymers has been known since the 1950s (Felsenfeld et al., 1957), but such structures were generally assumed to be biologically irrelevant (Wells, 1988). Models of triplexes were proposed in which the second homopyrimidine strand was bound in the major groove parallel to the homopurine strand by Hoogsteen (1959) base pairing, and those models were supported by fiber diffraction data on RNA and DNA triplexes composed of two T and one A strand (Arnott & Selsing, 1974). The more recent chemical studies of Moser and Dervan (1987) and Praseuth et al. (1988) have also supported the early models and confirmed the parallel orientation of the second homopyrimidine strand relative to the purine strand which had been proposed on the basis of steric considerations. More recent physical chemical studies (CD, melting) of $d(GA)_n$ and $d(TC)_n$ polymers have indicated that pyr–pur–pyr triplexes as well as some other hetero- and homocomplexes are formed under various conditions (Lee et al., 1979; Antao & Gray, 1988).

In addition to pyr–pur–pyr triplexes, triplexes composed of two *homopurine* and one homopyrimidine strand have also been reported in polymers (Broitman et al., 1987; Letai et al., 1988). Recently, an in vitro triplex composed of a 27-base

† This research was supported by National Institutes of Health Grant R01 GM 37254-01 and Office of Naval Research Contract N00014-88-K-0180.

* Author to whom correspondence should be addressed.

0006-2960/89/0428-7859$01.50/0