Holmes, D. S., & Quigley, M. (1981) *Anal. Biochem. 114*, 193-197.
Imai, Y. (1987) *J. Biochem. 101*, 1129-1139.
Jantzen, H., Strahle, U., Gloss, B., Steward, F., Schmid, W., Boshart, M., Milescek, R., & Schutz, G. (1987) *Cell 49*, 29-38.
Jones, K. W., & Whitlock, J. P., Jr. (1990) *Mol. Cell. Biol. 10*, 5098-5105.
Leighton, J. K., & Kemper, B. (1984) *J. Biol. Chem. 259*, 11165-11168.
Leighton, J. K., DeBrunner-Vossbrink, B. A., & Kemper, B. (1984) *Biochemistry 23*, 204-210.
Maniatis, T., Fitsch, G. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
McGhee, J. D., Wood, W. I., Dolan, M., Engel, J. D., & Felsenfeld, G. (1981) *Cell 27*, 45-55.
Mead, D. A., Szczesna-Skorupa, E., & Kemper, B. (1986) *Protein Eng. 1*, 67-74.
Nebert, D. W., & Gonzalez, F. J. (1987) *Annu. Rev. Biochem. 56*, 945-993.
Nebert, D. W., & Jones, J. E. (1989) *Int. J. Biochem. 21*, 243-252.
Nebert, D. W., Nelson, D. R., Coon, M. J., Estrabrook, R. W., Feyereisen, R., Fujii-Kuriyama, Y., Gonzalez, F. J., Guengerich, F. P., Gunsalus, I. C., Johnson, E. F., Loper, J. C., Sato, R., Waterman, M. R., & Waxman, D. J. (1991) *DNA Cell Biol. 10*, 1-14.
Neuhold, L. A., Shirayoshi, Y., Ozato, K., Jones, J. E., & Nebert, D. W. (1989) *Mol. Cell. Biol. 9*, 2378-2386.
Peterlin, B. M., Hardy, K. J., & Larsen, A. S. (1987) *Mol. Cell. Biol. 7*, 1967-1972.
Petersen, D. D., Magnuson, M. A., & Granner, D. K. (1988) *Mol. Cell. Biol. 8*, 96-104.
Pike, S. F., Shephard, E. A., Rabin, B. R., & Philips, I. R. (1985) *Biochem. Pharmacol. 34*, 2489-2494.
Rangarajan, P. N., & Padmanaban, G. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 3963-3967.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503-517.
Stadler, J., Larsen, A., Engel, J. S., Dolan, M., Groudine, M., & Weintraub, H. (1980) *Cell 20*, 451-460.
Tuan, D. Y. H., Solomon, W. B., London, I. M., & Lee, D. P. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2554-2558.
Tukey, R. H., Okino, S., Barnes, H., Griffin, K. J., & Johnson, E. F. (1985) *J. Biol. Chem. 260*, 13347-13354.
Weintraub, H., Lasen, A., & Groudine, M. (1981) *Cell 24*, 333-344.
Whitlock, J. P., Jr., Denison, M. S., Fisher, J. M., & Shen, E. S. (1989) *Mol. Biol. Med. 6*, 169-178.
Yanagida, A., Sogawa, K., Yasumoto, K.-I., & Fujii-Kuriyama, Y. (1990) *Mol. Cell. Biol. 10*, 1470-1475.
Zhao, J., Leighton, J. K., & Kemper, B. (1987) *Biochem. Biophys. Res. Commun. 146*, 224-231.
Zhao, J., Chan, G., Govind, S., Bell, P., & Kemper, B. (1990) *DNA Cell Biol. 9*, 37-48.

Low-Temperature Crystallographic Analyses of the Binding of Hoechst 33258 to the Double-Helical DNA Dodecamer C-G-C-G-A-A-T-T-C-G-C-G†

Jordi R. Quintana, Andrei A. Lipanov,‡ and Richard E. Dickerson*
Molecular Biology Institute, Department of Chemistry and Biochemistry, and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024-1570

Received May 29, 1991; Revised Manuscript Received July 29, 1991

ABSTRACT: The crystal structure of the complex of Hoechst 33258 and the DNA dodecamer C-G-C-G-A-A-T-T-C-G-C-G has been solved from X-ray data collected at three different low temperatures (0, −25, and −100 °C). Such temperatures have permitted collection of higher resolution data (2.0, 1.9, and 2.0 Å, respectively) than with previous X-ray studies of the same complex. In all three cases, the drug is located in the narrow central A-A-T-T region of the minor groove. Data analyses at −25 and −100 °C (each with a 1:1 drug/DNA ratio in the crystallizing solution) suggest a unique orientation for the drug. In contrast, two orientations of the drug were found equally possible at 0 °C with a 2:1 drug/DNA ratio in solution. Dihedral angles between the rings of Hoechst 33258 appear to change in a temperature-dependent manner. The drug/DNA complex is stabilized by single or bifurcated hydrogen bonds between the two N–H hydrogen-bond donors in the benzimidazole rings of Hoechst and adenine N3 and thymine O2 acceptors in the minor groove. A general preference for AT regions is conferred by electrostatic potential and by narrowing of the walls of the groove. Local point-by-point AT specificity follows from close van der Waals contacts between ring hydrogen atoms in Hoechst 33258 and the C2 hydrogens of adenines. Replacement of one benzimidazole ring by purine in a longer chain analogue of Hoechst 33258 could make that particular site GC tolerant in the manner observed at imidazole substitution for pyrrole in lexitropsins.

Noncovalent DNA minor groove binding drugs are a subject of interest because they interact with the DNA in a sequence-selective manner (Zimmer & Wähnert, 1986). The subject of this analysis, Hoechst 33258 (Figure 1), shares many structural and DNA-binding characteristics with drugs such as netropsin (Kopka et al., 1985), distamycin (Coll et al., 1987), berenil (Brown et al., 1990), and DAPI (Larsen et al., 1989): (1) They all are crescent-shaped assemblies of flat aromatic rings or amide groups, connected by more or less

† This work was supported by NIH Program Project Grant GM-31299 and American Cancer Society Grant CH-4720.
* Corresponding author.
‡ Permanent address: Institute of Molecular Genetics, USSR Academy of Sciences, Moscow 123182.

0006-2960/91/0430-10294$02.50/0

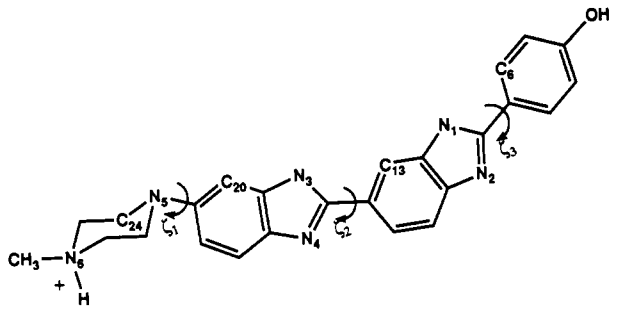

FIGURE 1: Molecular structure of Hoechst 33258. From right to left; phenol–benzimidazole 1–benzimidazole 2–piperazine, or phe–bz1–bz2–pip. Free rotation is possible about the connecting bonds ζ_1, ζ_2, and ζ_3.

flexible links; (2) they are positively charged and have hydrogen-bond donors or acceptors on their concave edge; (3) they bind preferably to AT-rich regions of DNA; and (4) they fit tightly between the phosphate–sugar walls of the minor groove, spanning 3–5 base pairs (Pjura et al., 1987; Teng et al. 1988). The detailed structural knowledge of these drug–DNA interactions obtained by X-ray crystallography can be an important factor in the design of new effectors of DNA expression (Lown, 1988). Modification of the chemical structure of currently known minor-groove binders using three-dimensional information is a strategy for the design of new drugs potentially capable of even more specific recognition of DNA sequence (Goodsell & Dickerson, 1986; Kumar et al., 1990).

Hoechst 33258 (hereafter simply "Hoechst") is a fluorochrome widely used in chromosome staining (Latt, 1976), that also possesses antihelmintic properties (Lämmler et al., 1971). It is built from four rings linked in the order phenol–benzimidazole–benzimidazole–piperazine (Figure 1). (For brevity these will be designated as phe–bzl–bz2–pip.) The binding of Hoechst to the minor groove of DNA has been studied by a variety of biophysical methods, including UV, CD, and fluorescence spectroscopy (Jorgenson et al., 1988; Kubota, 1990), footprinting (Harshman & Dervan, 1985; Portugal & Waring, 1988; Murray & Martin, 1988), NMR (Parkinson et al., 1990; Searle et al., 1990), and X-ray crystallography (Pjura et al., 1987; Teng et al., 1988; Carrondo et al., 1989). The common conclusion from footprinting studies is a preferential binding of Hoechst to AT-rich regions of DNA rather than to GC-rich regions, although GC base pairs are not completely excluded at the ends of a binding site (Harshman & Dervan, 1985; Portugal & Waring, 1988). Polymorphism in Hoechst–DNA binding has been observed using fluorescence and absorption spectroscopy (Loontiens et al., 1990). At low drug/DNA ratios a sequence-specific tight-binding mode occurs that has been studied by X-ray crystal structure analyses. As the drug concentration is increased, several other binding modes become possible, including a nonspecific salt-dependent binding that is mediated by electrostatic attraction between the charged drug and the negatively charged DNA.

Two X-ray studies have been published of crystals of the complex of Hoechst with the DNA dodecamer C-G-C-G-A-A-T-T-C-G-C-G at 2.2-Å resolution, using data collected at room temperature (Pjura et al., 1987) and at 15 °C (Teng et al., 1988). Both studies confirm the binding of Hoechst to the minor groove of DNA with the drug covering four base pairs and with the phenol ring closer to the C1·G24 base pair. [Although the two ends of the self-complementary double helix would be entirely equivalent in solution, in the crystal the local environments of the two ends differ. The "upper" or C1·G24 end of the dodecamer has a slight bend in helix axis that is not seen in the "lower" or G12·C13 end (Drew et al., 1981). This distinction will be made throughout between upper and lower ends of the helix in the crystal, even though it should be kept in mind that such a distinction may be meaningless in solution.] The Pjura and Teng structures differ, however, in the specific location of the drug binding site: an off-centered A-T-T-C site in Pjura et al. (1987), displaced by one base pair toward the lower end of the helix with the piperazine ring extended into the GC region, or a centered A-A-T-T site in Teng et al. (1988). Another 2.2-Å crystal structure analysis of Hoechst bound to the related dodecamer C-G-C-G-A-T-A-T-C-G-C-G (Carrondo et al., 1989) positions the drug at the off-center site G-A-T-A (= T-A-T-C), displaced by one base pair toward the upper end of the helix and with its phenol ring oriented toward the lower end. In solution, where the distinction between upper and lower ends of the helix is lost, this off-center binding locus with the piperazine extended into the GC region of the minor groove would be identical to that reported by Pjura et al. (1987).

Collecting X-ray crystallographic data at low temperatures has many advantages (Hope, 1988), including reduction of the dynamic disorder of the atoms in the crystal and lessening of the X-ray intensity decay caused by radiation damage. Both factors can lead to higher resolution in the diffraction pattern and better definition of the resultant structure. The low-temperature data collection method has been applied to crystals of B-DNA (Drew et al., 1982) and A-DNA (Shakked et al., 1989; Eisenstein et al., 1990). In this paper we present a low-temperature crystallographic study of the complex of Hoechst with the same C-G-C-G-A-A-T-T-C-G-C-G dodecamer that was studied previously at room temperature. The two main objectives of this study are to analyze the polymorphism of Hoechst binding to DNA under different crystallization conditions and different data-collection temperatures and to obtain higher resolution X-ray data the better to define the molecular structure of the Hoechst/DNA complex.

Materials and Methods

Synthesis, Crystallization, and Data Collection. The DNA dodecamer C-G-C-G-A-A-T-T-C-G-C-G was synthesized by solid-phase phosphoramidite methods from monomeric nucleotides (American Bionetics). After detritylation, the oligomer was purified by anionic exchange chromatography using DE-52 Whatman cellulose in 7 M urea/50 mM Tris·HCl buffer (pH 7.5) and a KCl gradient. The urea and KCl were removed later, and the counterion was changed to triethylammonium via a triethylammonium bicarbonate buffer (pH 7.5) over a second cellulose chromatography column. Hoechst 33258 was obtained from Boehringer Mannheim and used without further purification. The drug is light-sensitive, and hence all crystallization experiments were protected from light. The drug solutions and the solutions used for crystallization were stable in the dark, as monitored by UV and pH measurements.

Crystals of the Hoechst/C-G-C-G-A-A-T-T-C-G-C-G complex were obtained in a variety of conditions by changing the drug/DNA, cation/phosphate, and spermine/DNA ratios, as well as varying the salt used in crystallization. One crystal from each set of successful conditions was examined by X-ray precession photography. In all cases the resolution of the X-ray photographs extended at least to 2.5 Å. The pattern of intensities was the same regardless of crystallizing conditions and matched that from earlier photographs of Hoechst/C-G-C-G-A-A-T-T-C-G-C-G crystals taken by Pjura et al. (1987). It differed significantly from that of the DNA dodecamer alone.

Table I: Crystal Parameters for X-ray Data Sets of the Complex of Hoechst 33258 and the DNA Dodecamer C-G-C-G-A-A-T-T-C-G-C-G at Different Temperatures

	data collection temp				
	−100 °C	−25 °C	0 °C	15 °C[a]	RT[b]
drug/DNA ratio	1:1	1:1	2:1	2:1	1:1
space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a (Å)	23.93	23.89	23.95	25.23	25.04
b (Å)	39.05	38.52	38.99	40.58	40.33
c (Å)	65.15	66.34	65.18	66.08	65.85
V (Å^3)	60880	61049	60866	67655	66500
V/base pair (Å^3)	1268	1272	1268	1409	1385
resolution of data used (Å)	8.0–2.0	8.0–1.9	8.0–2.0	20.0–2.2	8.0–2.2
no. reflections $F > 2\sigma$ used in refinement	2072	2873	2540	2000	1561

[a] Teng et al. (1988). [b] Pjura et al. (1987).

Crystals used for data collection at −100 and −25 °C (ca. 0.7 × 0.2 × 0.1 mm) were grown by vapor diffusion at 4 °C from sitting drops initially containing 0.22 mM DNA duplex, 0.22 mM Hoechst (1:1 drug/DNA ratio in solution), 5.0 mM $Mg(OAc)_2$, 0.5 mM spermine·HCl (pH 6.9), and 15% (v/v) MPD (2-methyl-2,4-pentanediol), that were equilibrated slowly with a 32.5% (v/v) MPD solution in the reservoir. The 0.6 × 0.2 × 0.1 mm crystal used for data collection at 0 °C was grown by vapor diffusion at 4 °C from a sitting drop initially containing 0.22 mM DNA duplex, 0.44 mM Hoechst (2:1 drug/DNA ratio in solution), 5.0 mM $Mg(OAc)_2$, 0.5 mM spermine·HCl (pH 6.9), and 15% MPD, which was slowly equilibrated with a 32.5% (v/v) MPD solution in the reservoir.

All X-ray data were collected on a Rigaku AFC-5R diffractometer equipped with a graphite monochromator, using Cu Kα radiation from a rotating anode generator operated at 50 kV and 170 mA. The intensity decay was monitored by three standard check reflections that were measured periodically during data collection. X-ray intensities were corrected for Lorentz and polarization factors, time-dependent radiation damage, and empirical absorption factors. A flexible hose allowed the flow onto the capillary or crystal of cooled dried air (for the 0 and −25 °C data sets) or of dry nitrogen gas cooled with liquid nitrogen (for the −100 °C data set). In the latter case, the crystal was mounted on the tip of a glass fiber and protected from dehydration with a 50:50 mixture of Panatone N (Exxon) and mineral oil. For the −100 °C data set, a shielded box was built enclosing the diffractometer, and the air inside the box was dried continuously by a dehumidifier (Cargocaire Inc., Irvine, CA) to prevent ice formation. The temperature at the crystal was monitored and observed to remain stable during the different data collections.

Table I summarizes parameters of the three low-temperature data collections and compares them with the corresponding parameters reported for the two previous X-ray studies of the same complex. The −100 and −25 °C crystals were grown from solutions with a 1:1 drug/DNA ratio, as the earlier Pjura data had been. Crystals measured at 0 °C were grown from solution with a 2:1 drug/DNA ratio, as with the Teng et al. (1988) analysis. [For reference, the Carrondo et al. (1989) crystals with C-G-C-G-A-T-A-T-C-G-C-G were grown from a 4:1 drug/DNA solution.] Note that, although cell dimensions and cell volume shrink for all three of the low-temperature forms, the −100 °C/1:1 and 0 °C/2:1 sets most closely resemble one another. This similarity will persist through structure refinement.

The experience of many structure groups has shown that data for all B-DNA dodecamer crystals isomorphous with the original C-G-C-G-A-A-T-T-C-G-C-G exhibit a significant falloff in intensity at a resolution higher than 3.0 Å (sin $\theta/\lambda \geq 0.17$). This falloff may reflect packing difficulties with the overlapping-end manner in which these DNA dodecamer helices are arranged in the crystal (Dickerson et al., 1987). An informative way of displaying X-ray data is a *scatter plot*, in which actual observed structure factors, F_o, are plotted vertically along a baseline that gives the distance out in reciprocal space, sin θ/λ, as a measure of resolution. Figure S1 in the Supplementary Material (see note at end of paper) shows scatter plots for all five structure analyses listed in Table I. Of course observed values of the structure factor amplitudes (F_o) will depend on many variables such as crystal size and the device and conditions used for data collection. But it is apparent from Figure S1 that data collection at low temperature increased both the relative intensity and the quantity of $F \geq 2\sigma$ data lying beyond 3.0 Å. This in turn led directly to improved definition and clarity of the drug image in the difference electron density maps of Figure 2.

Table II: Comparison of Percent Differences between Scaled Data Sets and of RMS Differences in Atomic Position of Final Refined Structures

A. R Factors between Exponentially Scaled Data Sets, for the 911 Reflections with $F > 2\sigma$ Common to All of the Data Sets Compared

	data set		
data set	−25 °C	0 °C	RT
−100 °C	31.2%	16.1%	23.1%
−25 °C		36.3%	34.5%
0 °C			22.1%

B. RMS Differences (Å) in Atomic Positions Relative to Unit Cell Axes, 486 DNA Atoms Only

	data set			
data set	−100 °C	−25 °C	0 °C	RT
Drew native	1.47	1.68	1.54	0.94
−100 °C		1.17	0.57	1.14
−25 °C			1.23	1.71
0 °C				1.12

C. RMS Differences (Å) in Atomic Positions after Least-Squares Fitting of 486 DNA Atoms Only

	data set			
data set	−100 °C	−25 °C	0 °C	RT
Drew native	0.93	0.86	0.81	0.65
−100 °C		0.67	0.52	0.76
−25 °C			0.54	0.71
0 °C				0.59

Different low-temperature data sets were compared with one another and with the earlier Pjura room temperature data set, after exponential scaling using the relationship $F_1 = F_2 A e^{-B(\sin\theta/\lambda)^2}$. R factors between data sets, defined by the expression $R_{data} = \sum|F_1 - F_2|/\sum F_1$, are listed in Table IIA. For comparison purposes, scaling for this table was carried out using only the 911 reflections that were present and above the 2σ level in all four data sets: −100, −25, and 0 °C and RT. As with cell dimensions, observed intensities of the −100

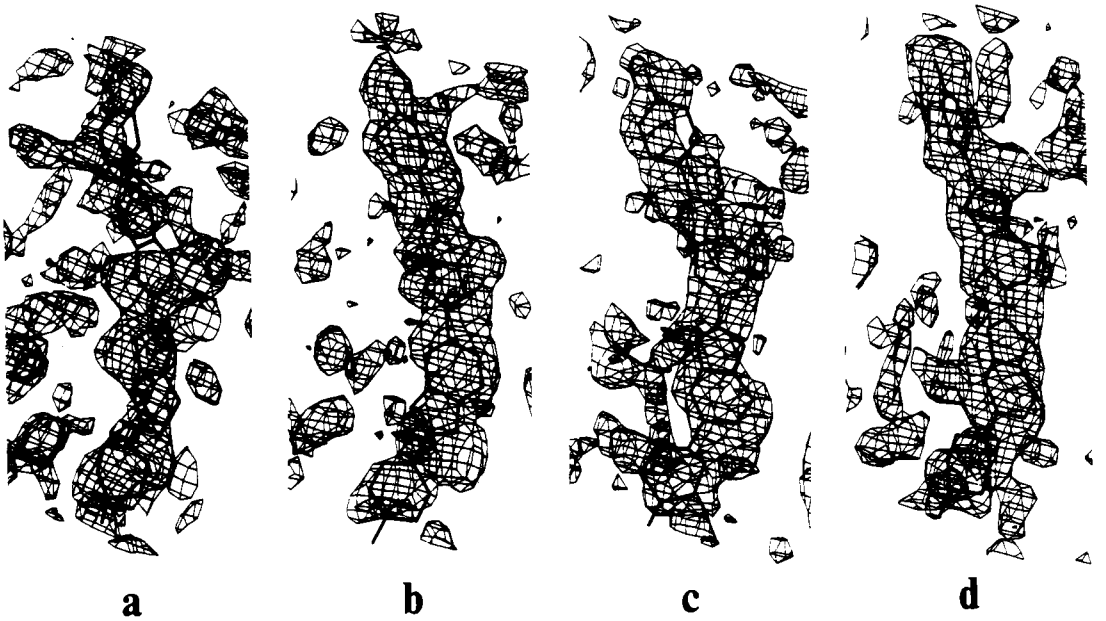

FIGURE 2: $F_o - F_c$ difference electron density maps of the central minor groove region in various DNA/Hoechst studies, showing the initial difference density image of the drug molecule that ultimately led to establishment of its position as indicated by the superimposed drug skeleton. All drug skeletons are as finally refined and are shown in the pip-down orientation except where noted. Phasing for these maps was obtained from a partially refined DNA double helix plus several solvent peaks in regions removed from the minor groove, but without inclusion of the drug molecule. In all cases, the drug image showed up in the central A-A-T-T region of the minor groove. (a) Reanalysis of the room temperature data set previously collected by Pjura et al. (1987). Map calculated at the 1.2σ level. $R = 17\%$ with 31 assigned waters (Table III). Portions of the phe and bz1 rings lie outside difference density, which probably is one reason why Pjura et al. elected to position the drug molecule one base pair lower down and fill the upper region with water molecules. (b) Map with 0 °C data set, calculated at the 1.7σ level. $R = 19\%$ with 41 waters. The drug is shown in the pip-up orientation, but the reverse orientation gave as good a fit to the density. The drug skeleton now lies completely within difference density except for one corner of the bz2 ring. (c) Map from the −25 °C data set, calculated at the 1.5σ level. $R = 22\%$ with 35 waters. Only a small part of the phe ring sits outside difference density. (d) Map from the −100 °C data set calculated at the 2.0σ level, with $R = 20\%$ and 45 assigned waters. The twist of the phe ring comes from the refinement process and not from the density image as seen here. Stereo versions of these four plots are available in the Supplementary Material.

and 0 °C data sets were more similar ($R = 16\%$) than were any other two sets.

The R factors between data sets are so large that it obviously would be improper to attempt to scale and merge any two of the individual data sets. In view of the wide variance in quality of data sets evidenced by the scatter plots of Figure S1, one must ask whether the scaling differences in Table IIA arise from real structural differences at the various temperatures or simply from data measurement errors. The differences could have one or more of three origins: (1) Experimental errors in data measurement. (2) Changes in position and orientation of the helices within the unit cell, including sampling of the molecular transform at different loci in reciprocal space because of changes in unit cell lengths. (3) Temperature-dependent changes in the fine structure of the double helix itself or of the kind and extent of interaction of Hoechst with the double helix. To anticipate the conclusions of the following section, factors 1 and 3 appear to be small compared with factor 2.

Structure Analyses and Refinement. A uniform strategy was followed for the independent refinement of each of the three low-temperature data sets in the first three columns of Table I (Figure S1c–e). In order to obtain an internally consistent comparison, the same refinement strategy also was employed in a rerefinement of data collected earlier at room temperature (RT) by Pjura et al. (1987) (right column of Table I and Figure S1a). The Teng et al. (1988) 15 °C data (Figure S1b) were not reanalyzed. The number of reflections with $F_o > 2\sigma$ used in each refinement is shown at the bottom of Table I. All crystals were nearly isomorphous with those of the native C-G-C-G-A-A-T-T-C-G-C-G (Wing et al., 1980; Drew et al., 1981), and hence the coordinates of that DNA dodecamer were used as a starting model for refinement.

Rigid body refinement of data with $F_o > 2\sigma$ from 8- to 3.0-Å resolution was carried out initially using version 1.5 of the X-PLOR program package (Brünger et al., 1987). In each of the four refinements, the dodecamer was begun as a single rigid body and gradually subdivided into 6, 12, 24, 48, and 70 rigid groups. It was found by trial and error that this rigid body refinement step was essential to obtaining the best structure solution. The remaining 2σ data from 3.0 Å out to the resolution limit were added gradually, and positions of individual atoms of the model were refined using X-PLOR. Positional refinement then was continued using NUCLSQ, the Hendrickson-Konnert (1980) least-squares refinement program adapted for nucleic acids (Westhof et al., 1985). Individual atomic temperature factors (B) were subsequently refined with NUCLSQ. Throughout this refinement procedure, slow-cooling molecular dynamics with the X-PLOR simulated-annealing program was used to check that the correct solution had been found.

Solvent peaks were located from $(F_o - F_c)$ and $(2F_o - F_c)$ difference electron density maps by an automated peak search procedure. The peaks were examined with the program FRODO (Jones, 1978) on an Evans and Sutherland PS390 graphics terminal. The highest difference density in early maps was observed in the central region of the minor groove and consisted of an extended lobe of density spanning the four central A·T base pairs. Although even at this stage a model of the Hoechst molecule could have been fitted within the limits of the difference density, this early drug image was ignored while solvent peaks were located and added to regions outside the minor groove, generally within the major groove or near phosphate oxygens. Solvent peaks were accepted only when they appeared simultaneously above the 3σ level in the $(F_o - F_c)$ maps and above the 1σ level in the $(2F_o - F_c)$ maps, and

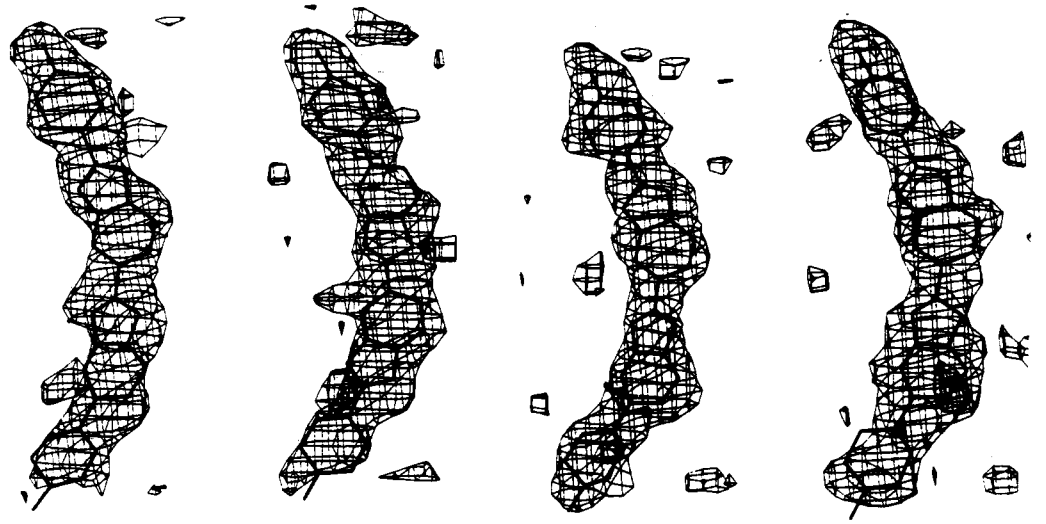

FIGURE 3: Final deletion difference electron density maps ($F_o - F_c$) for the three low-temperature studies, in which the drug molecule has been deleted from the final refined model. Several additional cycles of NUCLSQ refinement then were performed before calculating phases for these maps. The refined drug position is superimposed on the difference density for reference. (a) 0 °C data set, refined at the 2.2σ level with the pip-down orientation. (b) 0 °C data set, refined at the 2.2σ level with the pip-up orientation. (c) −25 °C data set, calculated at the 2.4σ level, pip-down. (d) −100 °C data set, calculated at the 2.2σ level, pip-down. The twist of the phe ring is no different than that in Figure 2d but is not as apparent because the drug image lies in the plane of the page here. Postrefinement difference maps such as in this figure make attractive journal illustrations but are not as informative as to the correctness of positioning of the drug as are the prerefinement difference maps of Figure 2. Stereo versions of these four plots are available in the Supplementary Material.

in addition were located within good hydrogen-bonding distances of chemically compatible atoms. The positions and temperature factors of the DNA atoms and the oxygens representing water molecules then were refined using NUCLSQ.

At this point in each analysis, a ($F_o - F_c$) difference map was calculated from the refined partial structure, in order to look for the drug molecule. These difference maps (Figure 2) show the first unbiased images of drug electron density within the minor groove for the RT, 0, −25, and −100 °C data analyses. In each map, the final refined drug skeleton has been added for reference. A "standard" Hoechst molecule was obtained from coordinates of the drug in the refined crystal structure of Pjura et al. (1987). The molecule was made planar by setting to zero the dihedral angles between the four rings: phe–bz1–bz2–pip. The planar drug model was fitted manually into density within the minor groove region of the difference map displayed via FRODO on a PS390 terminal. In the reanalysis of the Pjura data shown in Figure 2a, the drug density is broken in the regions of phe and bzl, a feature that also was present in Pjura's maps and that probably induced him to position the drug one base pair farther down the helix. But we saw no better density below this image that could be ascribed to a piperazine. On balance, the central A-A-T-T drug position seems preferable, even for the Pjura data. But the weakness of the RT data set makes any such decision tentative. Rerefinement of the old Pjura data was not pursued further.

In principle, two different orientations of the drug always were possible: with the piperazine ring closer to the top (C1·G24) of the helix or the bottom (G12·C13). These will be designated as pip-up and pip-down, respectively. For each of the three low-temperature data sets, both orientations were refined in parallel through several cycles of positional and temperature factor refinement using NUCLSQ. A choice then was made of that drug orientation that (a) best fitted the difference electron density maps, (b) converged to a lower R factor in the refinement, and (c) gave the best drug–DNA contacts. R factors for the partially refined pip-up and pip-down models are listed in Table III, but it must be remembered that these numbers constituted only one component of the decision-making process. For the −25 and −100 °C data sets, the pip-down orientation was superior in that it avoided steric clashes with the DNA, fitted the difference density slightly better, had a marginally lower R factor, and refined to consistently lower temperature factors, B. In contrast, both drug orientations were found to be equivalent in the 0 °C analysis, and there was no valid basis for discriminating between them.

NUCLSQ refinement of atomic positions and temperature factors was completed by gradual addition of more water molecules to the model using the criteria described earlier. Table III summarizes the course of the refinement for the different data sets analyzed and the makeup of the final independently refined models. The standard error in refined atomic positions, as estimated by Luzzati's method (1953), is 0.13 Å.

For each of the three new low-temperature analyses, Figure 3 shows the ultimate fit of drug to density in a deletion difference map at the conclusion of structure refinement, with the refined Hoechst molecule deleted from the phase calculation. As with all structures from this laboratory, both the original X-ray intensity data and the final refined coordinates for the low-temperature analyses have been deposited in the Brookhaven Protein Data Bank and are currently available upon request.

Results and Discussion

Variation of Structure with Temperature. In all three of the low-temperature structure analyses of complexes with C-G-C-G-A-A-T-T-C-G-C-G, Hoechst is observed to bind in the narrow central A-A-T-T region of the minor groove. The overall structure of the drug/DNA complex (Figures 4 and 5) is similar at all temperatures. Table IIB displays rms (root-mean-square) differences in atomic positions of DNA atoms among the three refined low-temperature structures and the room temperature structure as originally refined by Pjura et al. (1987). Table IIC shows rms differences after the

Table III: R Factor[a] Evolution of the Different Refinements of X-ray Data for Crystals of Hoechst 33258/C-G-C-G-A-A-T-T-C-G-C-G

refinement stage[b]	−100 °C	−25 °C	0 °C	RT[c]
starting model (8–3.0 Å)	51%	49%	50%	43%
rigid body + *XYZ* refinement (X-PLOR)	31%	27%	29%	27%
XYZ + *B* refinement (NUCLSQ)	25%	25%	24%	22%
no. waters and *R* before drug added	45	35	41	31
	20%	22%	19%	17%
partially refined *R*, pip-up vs pip-down	18.9%	21.2%	17.8%	refinement
	18.7%	20.4%	17.8%	not continued
final *R* factor	14.9%	15.2%	15.7%	
final no. of solvent molecules	91 H_2O plus $Mg(H_2O)_6^{2+}$	94 H_2O plus $Mg(H_2O)_6^{2+}$	65 H_2O plus $Mg(H_2O)_6^{2+}$	

[a] $R = \sum|F_o - F_c|/\sum|F_o|$. [b] *XYZ* = positional refinement; *B* = temperature factor refinement. [c] Reanalysis of previously collected room temperature data (Pjura et al., 1987).

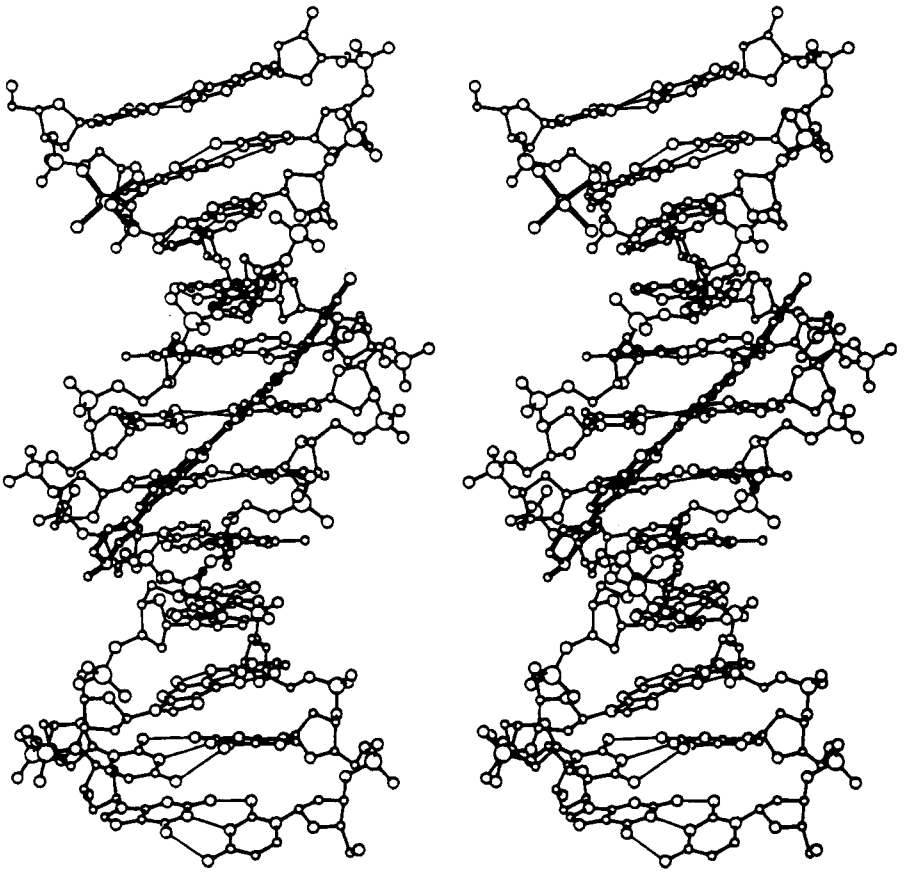

FIGURE 4: Stereoview of the final refined −25 °C structure of the complex of Hoechst with C-G-C-G-A-A-T-T-C-G-C-G, looking directly into the minor groove at the center of the helix. The C1·G24 base pair is at the top; G12·C13 is at the bottom. The drug spans the four central A-A-T-T base pairs in the minor groove, with the phenol ring closer to the top of the B-DNA dodecamer and the piperazine ring nearer the bottom. One ordered $Mg(H_2O)_6^{2+}$ ion is found near the G2 base at the top of the major groove. The reversed propeller of base pair G12·C13 at bottom is typical of nearly all of the dodecamer structures and is a consequence of crystal packing (Fratini et al., 1982). The top one-third of the helix is bent to the left as in most other dodecamer structures. The drug also causes the helix to bend backward at its center by ca. 8°. This can be detected by noticing that one views the underside of the third base pair from the top of the helix, but the upper surface of the base pair three from the bottom.

additional step of fitting one helix onto another by least-squares analysis.

The information in Table II allows one to decide between the three previously mentioned explanations for the differences in data sets: (1) data errors, (2) shifts of rigid helices within the cell, or (3) changes in helix structure. A strong linear correlation of $R = 0.80$ exists between the R_{data} factors relating pairs of data sets (Table IIA) and rms differences between final structures relative to their own unit cell axes (Table IIB). If the helices are removed from their respective cells and superimposed by least squares (Table IIC), then the rms differences all become smaller, of course. But they also become more uniform and no longer are correlated significantly with R_{data} differences in the X-ray data (linear correlation coefficient = 0.21). Hence the R_{data} differences in Table IIA arise mainly because the helices are positioned somewhat differently within their unit cells at different temperatures and because changes in cell dimensions (Table I) lead to sampling of the molecular transform at different loci, not because the helices themselves have a significant dependence of internal structure on temperature.

As has been mentioned earlier, the −100 and 0 °C data sets are most similar to one another (R_{data} = 16.1%), and their refined structures also are most alike (d_{rms} = 0.57 Å). But if individual helices are removed from their cells and superimposed, then the rms difference in atomic positions between 0 and −100 °C structures (0.52 Å) is not strikingly smaller than that between either of these and the −25 °C structure

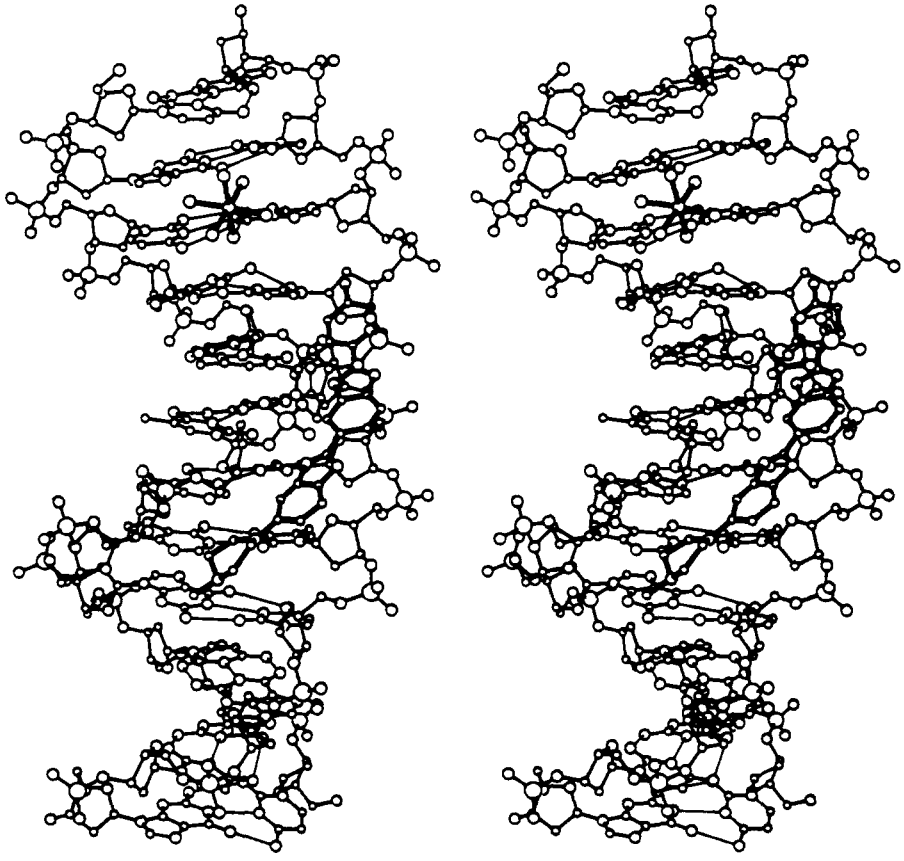

FIGURE 5: −25 °C Hoechst complex as in Figure 4 but rotated 30° about a vertical axis. The widening of the minor groove in going from an AT region to a GC region is clearly visible from right center to lower left. The $Mg(H_2O)_6^{2+}$ ion sits in the middle of the major groove at top.

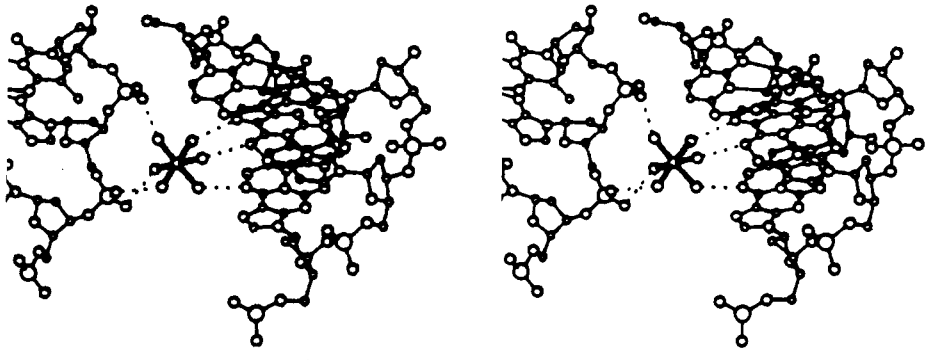

FIGURE 6: Intermolecular bridging of the $Mg(H_2O)_6^{2+}$ ion between G2 in the major groove of one helix at right and phosphates P6 and P7 of a neighbor at left. The two DNA helices are related by (x, y, z) and $(1 - x, y + 1/2, -z + 1/2)$. Base pairs from top to bottom at right are C1·G24, G2·C23, and C3·G22. At left, the two phosphates interacting with magnesium from bottom to top are P6 and P7. Dotted distances from magnesium cluster oxygens to DNA atoms in the −25 °C structure are to G2–N7, 2.92 Å; G2–O6, 2.94 Å; G22–O6, 2.80 Å; P6–O1P (O_L, toward major groove), 2.68 Å; P6–O2P (O_R, toward minor groove), 2.88 Å; P7–O1P (O_L), 2.98 Å.

(0.54 and 0.67 Å).

Structure of the Drug/DNA Complex. In each of the three low-temperature studies (and in the RT reexamination also), the optimum site for the Hoechst molecule is within the central A-A-T-T region of the minor groove (Figures 4 and 5). In addition, a well-defined octahedral mass of density is observed at the same location in $(F_o - F_c)$ difference electron density maps of all three low-temperature refinements: close to the G2 base in the major groove (upper left of Figure 5). It was added to the model as a magnesium hexahydrate cation, $Mg(H_2O)_6^{2+}$. When Westhof (1987) reanalyzed the data that had been collected by Drew et al. (1981) for the DNA dodecamer alone, they reported a peak near our Mg position that they suggested could be an ion, although they did not characterize it further. The persistence and clarity of the hydrated magnesium cation image in the present analyses may be ascribed in part to the improved definition afforded by higher intensity and higher resolution X-ray data. But shrinkage of cell dimensions in the low-temperature crystals (Table I) also may indicate a tighter intermolecular packing that helps hold the magnesium cation in place. Analogous temperature-dependent cell shrinkage also has been observed with crystals of this DNA dodecamer in the absence of drugs (Drew et al., 1982). Figure 6 shows the way in which the $Mg(H_2O)_6^{2+}$ ion bridges two symmetry-related dodecamers. Distances from magnesium-coordinated water oxygens to the N7 and O6 atoms of G2 in one helix or A6 and T7 phosphate oxygens in the other helix lie between 2.7 and 3.0 Å.

The −25 and −100 °C analyses favor a pip-down orientation of the Hoechst molecule (Figure 3c,d), but the 0 °C analysis could not discriminate between the two orientations (Figure 3a,b). The narrowness of the electron density one-quarter of

Table IV: Average Temperature Factors (B) of the Different Atom Groups in the Low-Temperature Analyses of Hoechst/C-G-C-G-A-A-T-T-C-G-C-G Complexes

data collection temp (°C)	phosphates	sugars	bases	Hoechst
0	21	14	11	38
−25	11	8	5.5	17
−100	9	6	3.5	30

Table V: Average Temperature Factors for Rings of Hoechst (phe–bz1–bz2–pip) from Different Low-Temperature Refinements of Hoechst/Dodecamer Complexes

data set	phe	bz1	bz2	pip
0 °C, pip-up	46	39	36	24
0 °C, pip-down	33	37	39	44
−25 °C, pip-down	15	15	17	20
−100 °C, pip-down	29	29	29	32

the way down from the top in Figure 3a,b might suggest that a five-membered ring is a better fit at that location (Figure 3a) than is a six-membered ring (Figure 3b), marginally favoring the pip-down alternative. The pip-down orientation also was reported at room temperature (Pjura et al., 1987) and at 15 °C (Teng et al., 1988). In contrast, the pip-up orientation has been reported in the crystal structure of Hoechst bound to C-G-C-G-A-T-A-T-C-G-C-G (Carrondo et al., 1989). It is not clear whether the ambiguity in orientation at 0 °C arises because the drug/DNA complex really has two equilibrium orientations within the crystal at that temperature, or whether the data collected at that temperature are simply insufficient in number and quality to establish a unique orientation with confidence. (One referee has suggested that the higher 2:1 drug/DNA ratio could itself be responsible for orientational disorder within the crystal. While such a situation is conceptually possible, it is difficult to see why such a relationship would come about.)

Mean temperature factors or B factors of the DNA atoms in the three low-temperature analyses decrease in the order: phosphates > sugars > bases, as has been observed for other DNA structures. Mean B's also decrease as the temperature of data collection is lowered from 0 to −100 °C (Table IV). Mean temperature factors of individual rings in the Hoechst molecule depend on its orientation (Table V). Whether pip-up or pip-down, the phe or pip ring nearest the bottom of the helix (G12·C13) always has a slightly larger B value than the ring at the other end of the drug molecule. In fact a uniform pattern exists, of increasing mean B of drug atoms in going from the upper or C1·G24 end of the helix to the G12·C13 end. This probably is a real effect that arises because, for the entire series of isomorphous dodecamers, the upper end of the double helix is pinned in place more tightly by neighboring molecules than is the lower end (Dickerson et al., 1987).

Drug-induced minor-groove widening also is greater at the lower end of the helix than at the upper, or at the right of Figure 7. This effect was seen with netropsin as well (Kopka et al., 1985), and, like the increase in B values, may arise because the lower half of the dodecamer is less constrained by its neighbors.

As was reported for the complex of this dodecamer with netropsin by Kopka et al. (1985), and the previous RT Hoechst analysis by Pjura et al. (1987), the double helix appears to be bent backward slightly at the site of drug binding (Figure 8). The drift of points from left to right in the region of base pairs 2–3–4–5–6 reflects a bend in overall helix axis in the upper half of the molecule, concave to the left as viewed in Figure 4. This bend has been a characteristic of every do-

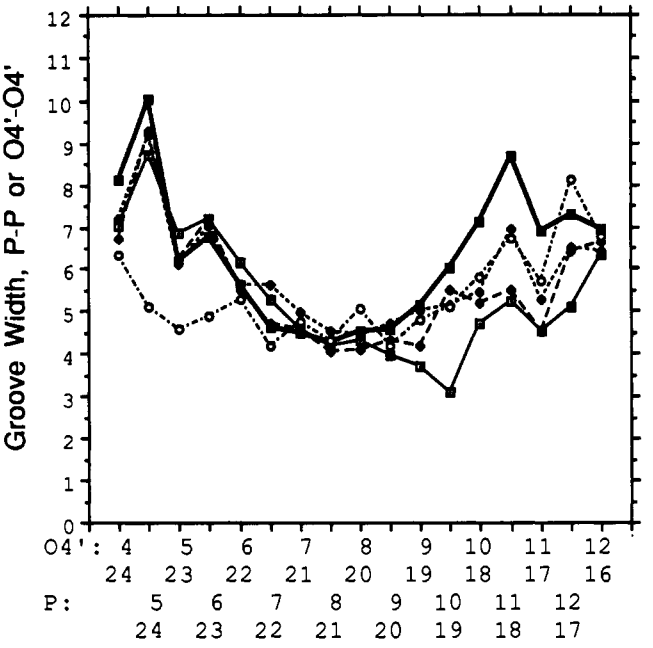

FIGURE 7: Minor groove width as measured by shortest P–P distances less 5.8 Å and by shortest O4′–O4′ distances less 2.8 Å. Oxygen and phosphorus atoms are numbered along the horizontal axis. The C1·G24 end of the helix is at left, and the G12·C13 end is at right. Symbols: open squares and thin solid lines, DNA dodecamer without drug; solid diamonds and short dashed lines, Pjura RT Hoechst complex; open diamonds and long dashed lines, Teng 15 °C Hoechst complex; solid squares and heavy lines, this −25 °C Hoechst complex; open circles and chain lines, netropsin complex. The large values at left arise from the bend in helix axis, which is not present in the netropsin complex with C-G-C-G-A-A-T-T-^{br}C-G-C-G.

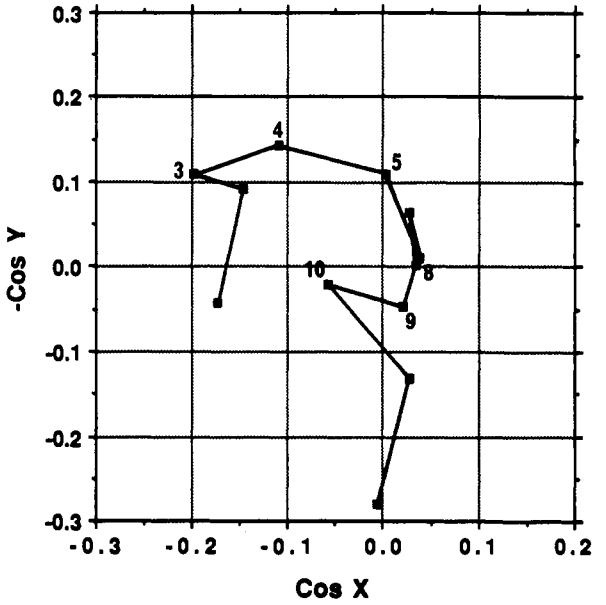

FIGURE 8: Normal vector plot for the 12 base pairs of the −25 °C complex of Hoechst with C-G-C-G-A-A-T-T-C-G-C-G. Cos X and cos Y are the X and Y direction cosines of normal vectors to the base pairs, as calculated by NEWHEL91. Each point represents the tip of a unit-length vector normal to the base pair in question, all vectors being plotted from a common origin. Drifts in these vectors can reveal bends in the helix axis, as discussed in the text.

decamer structure that has been solved, regardless of sequence, except for those that possess bromines in their major groove via 9-bromocytosines (Fratini et al., 1982; Dickerson, 1991). More importantly in the present context, the drift in normal vectors from top to bottom of the helix, from base pairs 3–4–5 compared with 8–9–10, indicates a bending backward of the

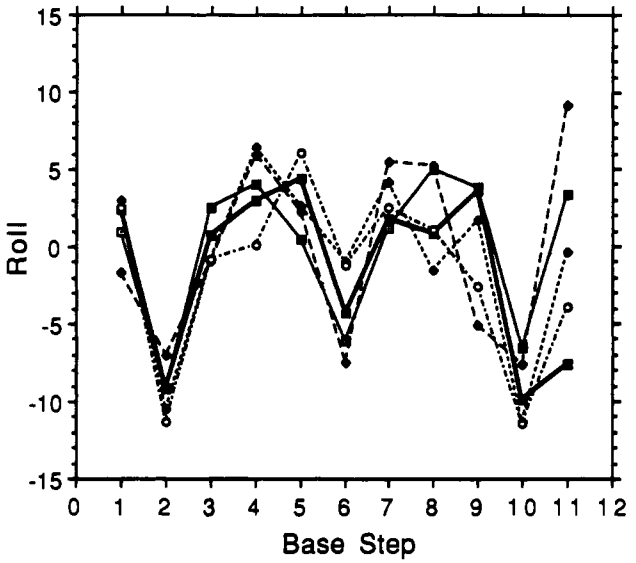

FIGURE 9: Roll angles between adjacent base pairs. Helices are identified as in Figure 7. The most extreme negative roll (opening toward major groove) occurs at purine–pyrimidine steps: G2–C3, A5–T6, and G10–C11. Binding of drug has little effect on roll angles.

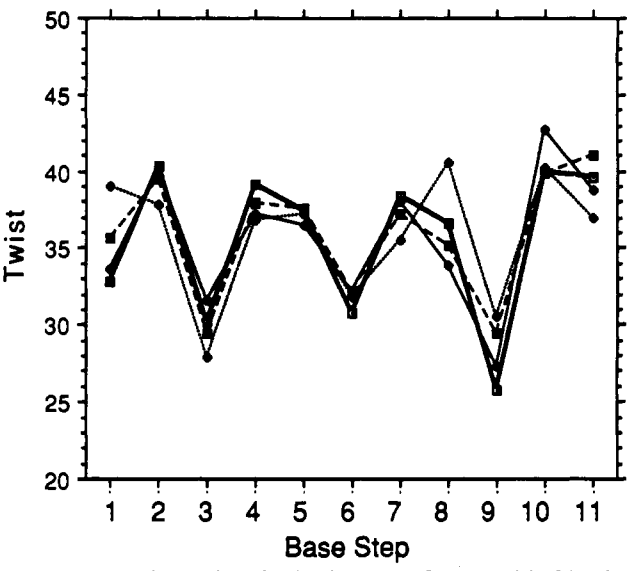

FIGURE 10: Twist angle behavior in parent DNA and in Hoechst complexes at three temperatures. Symbols: open diamonds and dotted lines, DNA dodecamer without bound drug; solid squares and dashed lines, Hoechst complex at 0 °C; solid diamonds and thin lines, Hoechst complex at −25 °C; open squares and thick lines, Hoechst complex at −100 °C. Twist values tend to be slightly more extreme at lower temperature.

axis as viewed in Figure 4, in a direction that opens up the minor groove slightly and compresses the major groove. As a consequence, the view in Figure 4 is of the undersurface of base pair C3·G22 and the upper surface of base pairs C9·G16 and G10·C15.

This backward bending is not encountered in the native dodecamer structure and presumably is a consequence of drug binding. (Normal vectors for base pairs 1 and 11–12 reflect end effects and intermolecular contacts within the crystal and contain little information about helix structure other than its ready deformability.) The difference of 0.140 between mean cos Y values for clusters 3–4–5 and 8–9–10 indicates an axial bending of arcsin 0.140 = 8°. This is identical to the bending angle obtained with netropsin (Kopka et al., 1985) and is slightly larger than the 6° bend reported for Hoechst by Pjura et al. (1987).

Helical parameters of the DNA in the three low-temperature drug/DNA complexes were calculated using the NEWHEL91 program. (A copy of this program is available by computer mail from one of the authors, R.E.D., at the address RED@UCLAUE.) These parameters for the −25 °C analysis are listed in Tables A1–A3 of the Supplementary Material. As Figure 9 shows, binding of either drug molecule, Hoechst or netropsin, has little effect on roll angles. The angle between adjacent base pairs continues to open toward the major groove (negative roll) at purine–pyrimidine steps such as G–C and A–T (steps 2, 6, and 10 in Figure 9). Twist angles (Figure 10) likewise tend to follow the pattern seen in the DNA dodecamer without bound drug. But as the various temperature plots in Figure 10 suggest, a lower temperature seems to exaggerate slightly the maxima and minima of twist.

The plot of buckle in Figure 11 is revealing. For both Hoechst and netropsin, the increased slope of the buckle curve between base pairs 4 and 9 by comparison with the drug-free or native helix indicates that drug binding causes the base pairs to bow convexly away to either side of the binding site, as though the centers of the four middle base pairs were being dragged toward the two ends of the helix by association with the drug molecule. This is a reasonable explanation, since Goodsell and Dickerson (1986) calculated that the natural repeat in both Hoechst and netropsin is approximately 20% longer than the distance from one base pair to the next along

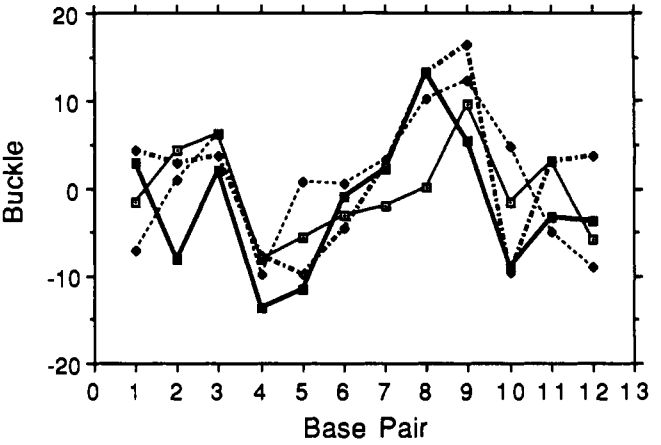

FIGURE 11: Buckle of base pairs in DNA dodecamer and in three Hoechst and netropsin complexes. Helices are identified as in Figure 7. The greater slope of the curve for the drug complexes at base pairs 4–8 indicates that the four central base pairs all dome outward from the center of the molecule. (Parameters for the Teng structure are omitted because they appear to be more erratic in this region of the molecule.)

the floor of the minor groove. A forced accommodation of the DNA to the longer repeat drug molecule would lead to just such a bowing or buckling.

Hence the binding of drug to DNA introduces significant local perturbation in overall helix geometry, without major alteration in local parameter values. The base pairs are bowed or buckled away from the binding site, the helix axis is bent backward by 8°, and the minor groove is widened in those regions where intermolecular constraints are least. The greater freedom in crystal packing in the lower half of the molecule, remarked in connection with minor groove width, also is reflected in slide (Figure 12), inclination, and X and Y displacement.

Drug Geometry. A planar Hoechst model was built initially into the difference density. No restraints were placed on rotation about bonds connecting rings, so that final refined torsion angles would be dictated by the X-ray data rather than the refinement model. Table VI lists the calculated dihedral

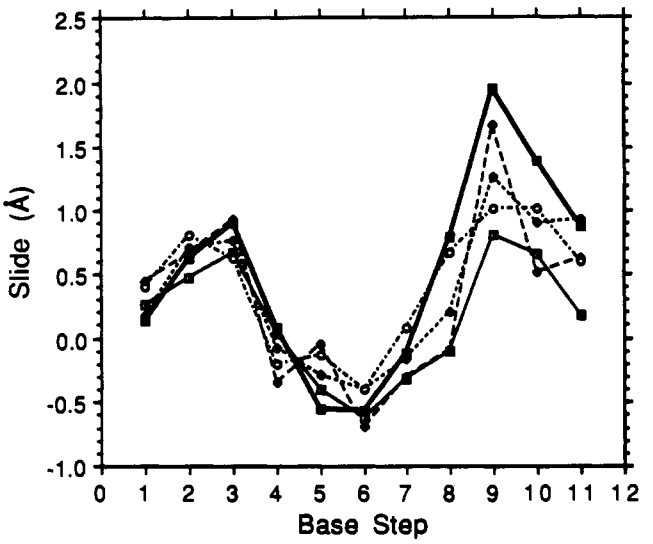

FIGURE 12: Slide of one base pair relative to the next. Helices are identified as in Figure 7. The greater freedom from intermolecular crystal packing at the bottom of the helix (right of plot) allows the drug molecule to induce a greater slide.

Table VI: Dihedral Angles between the Different Rings of Hoechst 33258 (phe–bz1–bz2–pip)[a]

data set	phe–bz1 ζ_3 (°)	bz1–bz2 ζ_2 (°)	bz2–pip ζ_1 (°)	ref
RT	0	37	60	Pjura et al. (1987)
15 °C	8	32	14	Teng et al. (1988)
0 °C (pip-up)	8	20	9	this work
0 °C (pip-down)	8	15	3	this work
−25 °C	10	12	8	this work
−100 °C	30	2	6	this work

[a] As reported in the text, standard deviations in these dihedral angles are estimated at 4°–6°.

angles between adjacent planar rings along the Hoechst molecule. If the standard deviation in atomic positions is 0.13 Å for all atoms involved in a dihedral angle between two planar rings, and if the distance across a ring is ca. 2.42 Å, then eq 17, p 331, Vol. II of the International Tables for Crystallography (Kasper & Lansdale, 1959) yields a standard deviation in dihedral angles ranging from 6.2° at dihedral angle zero to 4.6° at a dihedral angle of 30°.

As Table VI shows, a significant and temperature-dependent variation in dihedral angles is observed: The two benzimidazole rings become progressively more coplanar as the temperature is lowered. Larsen et al. (1989) have observed that two opposing forces act on these dihedral angles. Resonance or valence bond delocalization between aromatic rings would tend to make adjacent rings coplanar, but then they could not follow the spiraling walls of the minor groove. Close packing contacts between each planar ring and its portion of the minor groove would favor a twist at each connecting bond. Table VI suggests that packing against the groove walls predominates at higher temperature. But at lower temperature the molecule has insufficient energy to oppose the barrier to rotation imposed by resonance delocalization between rings, and so the two benzimidazoles approach planarity. The smaller phenol ring seems to be less affected by groove wall packing and maintains a nearly constant dihedral angle with bz1 except for the anomalously large 30° at the lowest temperature. This value is clearly outside the range of expected errors and furthermore has been checked by setting the dihedral angle to zero and observing that repeated refinement carries the angle back to a large value once again, whereas the other two dihedral angles remain essentially constant. We have no explanation for this large angle between phe and bz1 at −100 °C, other than the suggestion that, as bz1 and bz2 become coplanar, groove wall contacts become so significant for the small terminal phenol that the resonance bond is broken between phe and bz1.

Binding Specificity. Crystal structures of noncovalent minor groove binding drugs such as netropsin, distamycin, DAPI, berenil, and Hoechst complexed with DNA dodecamers have shown that several different types of interaction contribute to the observed AT specificity (Kopka et al., 1985; Larsen et al., 1989). But there has been a surprising tendency, even among theoreticians, to confuse the quite different concepts of binding *stability* vs binding *specificity*. Stability refers to the tendency of a drug to bind to the DNA in preference to remaining free in solution. Specificity, in the present context, refers to the tendency of the drug to bind to AT regions more strongly than to GC regions. The specificity can be *regional*, meaning that the drug binds preferentially to a broad zone of exclusively or predominately A·T base pairs. Specificity also can have a *local* component—a demand for an A·T base pair (or a G·C pair) at a particular position along the drug molecule.

As an illustration, the opposite charges of netropsin and DNA and the lower dielectric constant in the minor groove (Jin & Breslauer, 1988) both confer stability on the complex, but at first glance one would not expect them to contribute to specificity. Yet Pullman (1983) and co-workers have calculated that the electrostatic potential within the minor groove of B-DNA should be lower in AT-rich regions than in GC-rich regions. This could arise because the adenines of AT-rich zones do not have a string of electropositive N2 amine groups along the floor of the groove, as GC regions do. As a consequence, a cationic drug molecule could find a preferred resting place in AT-rich zones of base sequence. This is an example of long-range or regional specificity. Another regional component is the tendency of some AT-rich zones of DNA to have a narrower minor groove, between whose walls the flat planar drug molecules insert in a kind of "quasi-intercalation" (Larsen et al., 1989).

Both of these components of specificity have to do with long regions of DNA sequence of a given character, not with base-pair-by-base-pair sequence changes. In contrast, close van der Waals contacts between the drug molecule and the C2 hydrogen atoms of adenines provide local specificity. In netropsin (Figure 13b) the three hydrogen-bonding amides are surrounded by four components of the drug molecule that pack closely against the adenines: two pyrrole C–H in the center and two methylene $-CH_2-$ groups toward the ends. Guanines are unacceptable at these four positions along the drug because the pyrroles and methylenes leave no room for the guanine NH_2 group (dotted close interactions in Figure 13b). Kopka et al. (1985) suggested that guanine might be accommodated at a specific locus along the drug by replacing the pyrrole C–H by N or by O and named these modified sequence-reading netropsin analogues "lexitropsins". Hence base sequence specificity in netropsin has regional components from electrostatics and groove width and local components from close van der Waals contacts. The bifurcated hydrogen bonds between drug atoms and acceptors on the floor of the groove help to stabilize the complex, but contribute little to AT-specificity since they would be equally possible with A·T and G·C base pairs, were the guanine amine groups not present.

Hoechst resembles netropsin in that it also has a series of bifurcated hydrogen bonds to the floor of the minor groove conferring stability on the complex, flanked by close van der Waals contacts with adenine C2 hydrogens that confer spe-

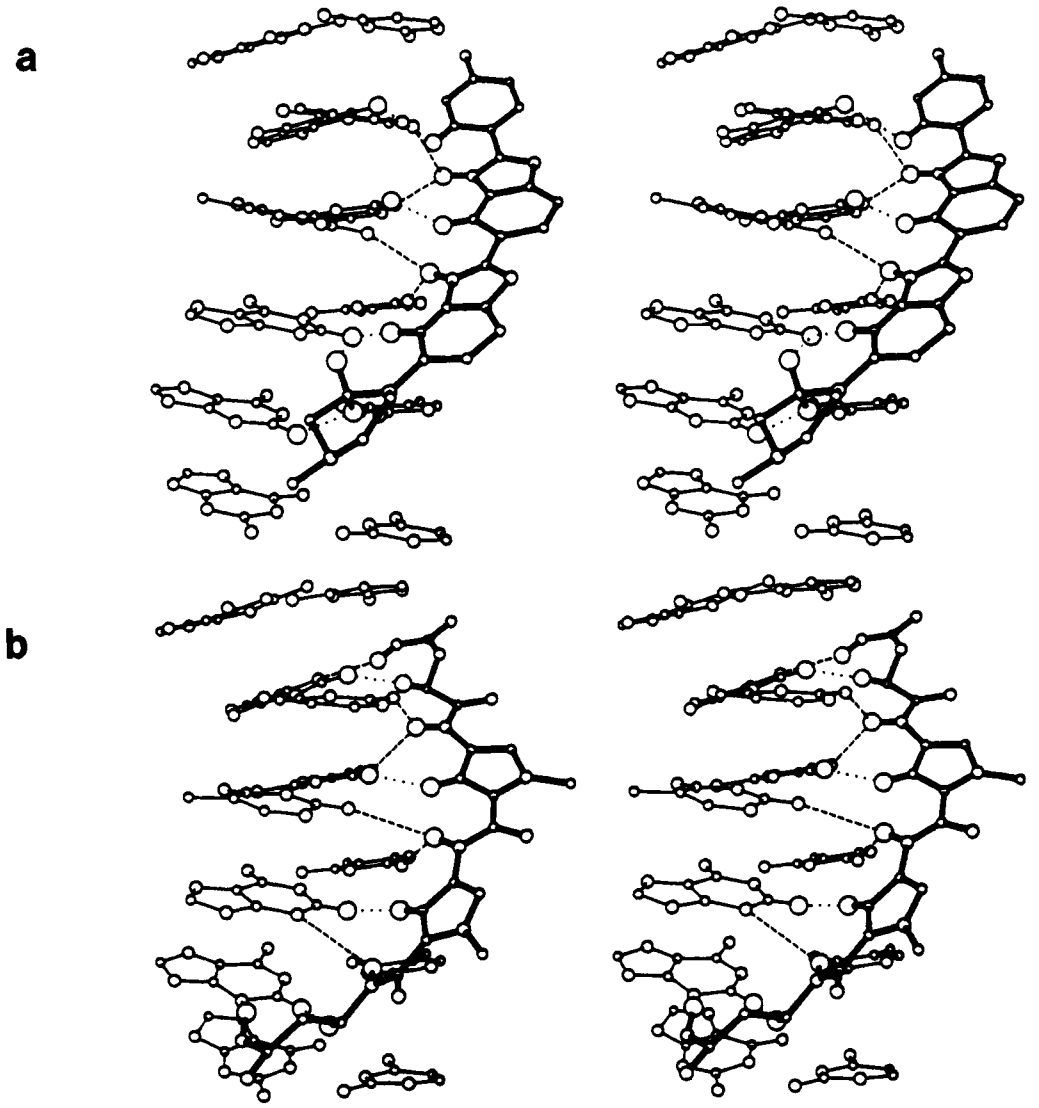

FIGURE 13: Stereo views of the interactions between Hoechst or netropsin and the central four base pairs of C-G-C-G-A-A-T-T-C-G-C-G. Only the base pairs of the DNA are shown. (a) Hoechst/dodecamer complex at −25 °C. (b) netropsin complexed with C-G-C-G-A-A-T-T-^{br}C-G-C-G (Kopka et al., 1985). Key hydrogen atoms are emphasized by large open circles. Dashed lines are hydrogen bonds between the drug and adenine N3 or thymine O2 atoms of DNA bases. Dotted lines indicate close van der Waals contacts between hydrogen atoms in the drug and those of adenine C2, which contribute to the AT specificity of the drug molecule.

Table VII: Distances Less Than 3.5 Å between Hoechst 33258 Atoms and Atoms of the Minor Groove of C-G-C-G-A-A-T-T-C-G-C-G

Hoechst (ring/atom)	DNA (base/atom)	distance (Å)			class
		−100 °C	−25 °C	0 °C	
phe/HC6	A5/HC2	2.25	2.67	2.96	van der Waals
bz1/N1	A6/N3	3.29	2.93	2.94	hydrogen bond
	T20/O2	3.19	3.20	3.14	
bz1/HC13	A6/HC2	2.15	1.95	1.58	van der Waals
bz2/N3	T7/O2	3.70	2.70	2.67	hydrogen bond
	T19/O2	2.89	3.09	2.70	
bz2/HC20	A18/HC2	3.08	1.76	1.59	van der Waals
pip/HC24	A18/HC2	2.94	2.61	2.17	van der Waals

cificity. However, Hoechst differs from netropsin in having a smaller recognition region. Netropsin has three bifurcated hydrogen bonds from amine NH, flanked by four van der Waals contact groups. Since each bifurcated hydrogen bond bridges two base pairs, the three bonds cover four base pairs, the minimal binding site for netropsin. But as Figure 13a shows, Hoechst has only two hydrogen-bonding elements (benzimidazole N–H), flanked by three van der Waals contact groups: the phenol ring and the two six-membered rings of the benzimidazoles. If the nonplanar piperazine ring were not present, the natural binding site for the planar aromatic components of Hoechst would be only three base pairs.

Table VII summarizes the observed contacts between Hoechst and DNA in the pip-down configurations of the three low-temperature complexes. Bifurcated hydrogen bonds occur from the N1 hydrogen of Hoechst ring bz1 to the N3 of adenine A6 and O2 of thymine T20 and from the N3 hydrogen of bz2 to the O2 atoms of thymines T7 and T19 (Figure 13a). The slight differences in the length of these hydrogen bonds may be related to the observed variation in the dihedral angles between the Hoechst rings. Although hydrogen atoms are not seen directly in the X-ray structure analysis, ideal hydrogen

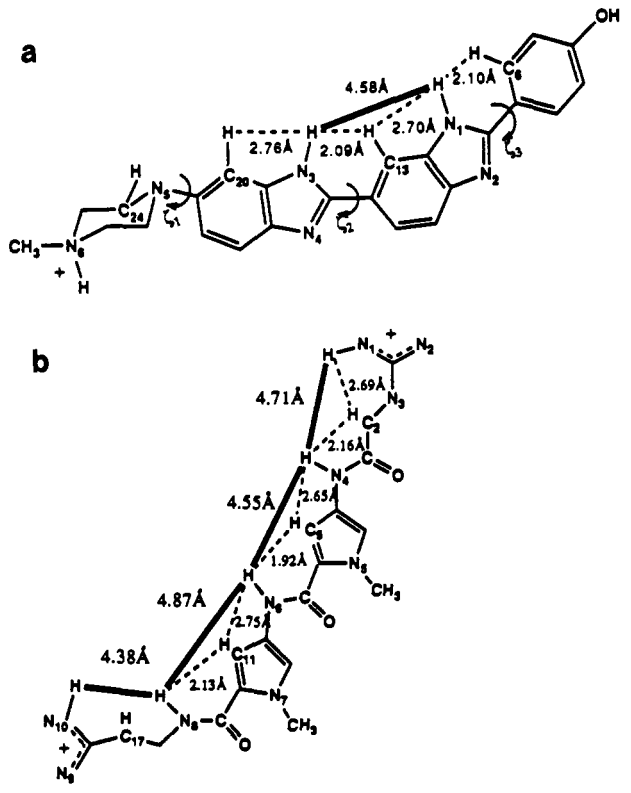

FIGURE 14: Calculated distances between hydrogen atoms along the concave edges of two drug molecules. Hydrogen atoms were calculated using the program ADDH by Gilbert G. Privé. (a) Hoechst/C-G-C-G-A-A-T-T-C-G-C-G at −25 °C. (b) Netropsin/C-G-C-G-A-A-T-T-^{Br}C-G-C-G. Dashed lines connect CH hydrogens with NH hydrogens, and heavy solid bars interconnect NH hydrogens. If one of the Hoechst benzimidazoles were to be replaced by purine, then C_{13}–H or C_{20}–H would become a nitrogen atom. This should permit acceptance of a GC base pair at that point along the drug, much as replacing a netropsin pyrrole by imidazole in a lexitropsin makes that site GC tolerant.

atom positions were generated for both DNA and Hoechst with Gilbert G. Prive's ADDH program, in order to observe close van der Waals contacts between the two molecules. As expected, the benzimidazole HC13 and HC20 hydrogens (Figure 1) are packed tightly against the C2 hydrogens of adenine A6 and A18, respectively. Toward the ends of the drug molecule, the phenol HC6 and piperazine HC24 atoms are slightly farther from C2 hydrogens of adenines A5 and A18, respectively. The absence of hydrogen bonds between both phe and pip and the DNA may tie them less closely to the helix.

Netropsin bridges all four of the central A·T base pairs in the dodecamer sequence C-G-C-G-A-A-T-T-C-G-C-G and hence fills up the entire AT region and has no freedom of binding (Figure 13b). In contrast, the smaller phe–bz1–bz2 core of Hoechst alone could bridge only the first three central A·T base pairs, C-G-C-G-A-A-T-T-C-G-C-G, or the last three, C-G-C-G-A-A-T-T-C-G-C-G. Depending on the orientation of the drug molecule, the nonplanar piperazine ring could sit in the remaining part of the AT center or could hang over into the adjacent GC zone. The former case was observed with C-G-C-G-A-A-T-T-C-G-C-G by Teng et al. (1988) and in the present low temperature analyses. The latter mode was observed by Carrondo et al. (1989) with C-G-C-G-A-T-A-T-C-G-C-G and had been reported with C-G-C-G-A-A-T-T-C-G-C-G at room temperature by Pjura et al. (1987), although the quality of the data in the latter analysis is such as to leave the issue doubtful. Still shorter groove-binding drugs such as berenil and DAPI fill the extra space in the central A-A-T-T site with a water molecule (Brown et al., 1990; Larsen et al., 1989).

The geometry of the groups interacting with DNA is surprisingly similar in Hoechst and in netropsin (Figure 14), even though the drug backbones are quite different chemically. In each case the separation between successive hydrogen-bond donors along the concave edge of the drug is roughly 4.7 Å. Because the hydrogen bonds are bifurcated, these donors will be positioned halfway between base edges along the minor groove. Directly facing the base edges are the C-H groups that make close contact with adenine C2 hydrogens: from phenol and benzimidazole in Hoechst or from pyrrole and methylene in netropsin. In both drugs these contact groups exhibit an asymmetry relative to their flanking hydrogen-bond donors, with proton–proton distances that alternate between 2.7 and 2.1 Å. Such an asymmetry might suggest a directionality of drug binding in a DNA sequence that was itself asymmetrical, but the issue cannot be decided with the current self-complementary sequences.

The central A-A-T-T sequence simulates a region of poly(dA)·poly(dT) in that it does not exhibit the alternation of twist that is seen in A-T-A-T (Coll et al., 1989) and longer alternating regions. This structural alternation may make drug binding difficult, which could explain the reported preference of netropsin for nonalternating over alternating AT sequences (Zimmer & Wähnert, 1986). An equivalent preference of Hoechst for nonalternating AT regions has not been reported, but the small size of the recognition site in Hoechst may make the drug less sensitive to an alternating DNA geometry. Longer analogues of Hoechst 33258, with a greater number of benzimidazole subunits, could begin to show sequence preferences analogous to those of netropsin. Indeed, replacement of one particular benzimidazole by purine could even permit acceptance of a GC base pair at that locus, in a Hoechst analogue of the lexitropsins. Kumar et al. (1990) have shown with NMR that a derivative of Hoechst having such a replacement exhibits a higher acceptance of GC base pairs. Longer polymers of linked purine and benzimidazole groups might actually be better candidates for reading specific base sequences than are the lexitropsins.

ACKNOWLEDGMENTS

We thank Kazimierz Grzeskowiak for DNA synthesis and Mary L. Kopka for discussions of drug binding.

SUPPLEMENTARY MATERIAL AVAILABLE

Figures S1–S3, showing the distribution of F_o's $\geq 2\sigma$ as a function of resolution, $F_o - F_c$ difference electron density maps of the central minor groove region in various DNA/Hoechst studies, and final deletion difference electron density maps ($F_o - F_c$) for the three low-temperature studies, and Tables A1–A3, giving helix parameters for the −25 °C Hoechst 33258 complex (8 pages). Ordering information is given on any current masthead page.

Registry No. CGCGAATTCGCG, 77889-82-8.

REFERENCES

Brown, D. G., Sanderson, M. R., Skelly, J. V., Jenkins, T. C., Brown, T., Garman, E., Stuart, D. I., & Neidle, S. (1990) *EMBO J. 9*, 1329–1334.

Brünger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458–460.

Carrondo, M. A., Coll, M., Aymami, J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1989) *Biochemistry 28*, 7849–7859.

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. (1991) *Methods Enzymol.* (in press).
Dickerson, R. E., Goodsell, D. S., Kopka, M. L., & Pjura, P. E. (1987) *J. Biomol. Struct. Dyn. 5*, 557–579.
Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2179–2183.
Drew, H. R., Samson, S., & Dickerson, R. E. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4040–4044.
Eisenstein, M., Frolow, F., Shakked, Z., & Rabinovich, D. (1990) *Nucleic Acids Res. 18*, 3185–3194.
Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686–14707.
Goodsell, D., & Dickerson, R. E. (1986) *J. Med. Chem. 29*, 727–733.
Grzeskowiak, K., Yanagi, K., Privê, G. G., & Dickerson, R. E. (1991) *J. Biol. Chem. 266*, 8861–8883.
Harshman, K. D., & Dervan, P. B. (1985) *Nucleic Acids Res. 13*, 4825–4835.
Hendrickson, W. A., & Konnert, J. H. (1980) in *Computing in Crystallography* (Diamond, R., Ramaseshan, S., & Venkatesan, K., Eds.) pp 13.01–13.23, The Indian Academy of Sciences, Bangalore, India.
Hope, H. (1988) *Acta Crystallogr. B44*, 22–26.
Jin, R., & Breslauer, K. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8030–8042.
Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.
Jorgenson, K. F., Varshney, U., & van de Sande, J. H. (1988) *J. Biomol. Struct. Dyn. 5*, 1005–1023.
Kasper, J. S., & Lansdale, K., Eds. (1959) *International Tables for X-ray Crystallography*, Vol. II, p 331, eq 17, Kynoch Press, Birmingham, AL.
Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985) *J. Mol. Biol. 183*, 553–563.
Kubota, Y. (1990) *Bull. Chem. Soc. Jpn. 63*, 758–764.
Kumar, S., Yadagiri, B., Zimmermann, J., Pon, R. T., & Lown, J. W. (1990) *J. Biomol. Struct. Dyn. 8*, 331–357.
Lämmler, G., Herzog, H., Saupe, E., & Schutze, H. R. (1971) *WHO Bull. 44*, 751–756.
Larsen, T. A., Goodsell, D. S., Cascio, D., Grzeskowiak, K., & Dickerson, R. E. (1989) *J. Biomol. Struct. Dyn. 7*, 477–491.
Latt, S. A. (1976) *Annu. Rev. Biophys. Bioeng. 5*, 1–37.
Loontiens, F. G., Regenfuss, P., Zechel, A., Dumortier, L., & Clegg, R. M. (1990) *Biochemistry 29*, 9029–9039.
Lown, J. W. (1988) *Anti-Cancer Drug Des. 3*, 25–40.
Luzzati, V. (1953) *Acta Crystallogr. 5*, 802–810.
Murray, V., & Martin, R. F. (1988) *J. Mol. Biol. 203*, 63–73.
Parkinson, J. A., Barber, J., Douglas, K. T., Rosamond, J., & Sharples, D. (1990) *Biochemistry 29*, 10181–10190.
Pjura, P. E., Grzeskowiak, K., & Dickerson, R. E. (1987) *J. Mol. Biol. 197*, 257–271.
Portugal, J., & Waring, M. J. (1988) *Biochim. Biophys. Acta 949*, 158–168.
Privê, G. G., Yanagi, K., & Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 177–199.
Pullman, B. (1983) *J. Biomol. Struct. Dyn. 1*, 773–794.
Searle, M. S., & Embrey, K. J. (1990) *Nucleic Acids Res. 18*, 3753–3762.
Shakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolow, F., & Rabinovich, D. (1989) *Nature 342*, 456–460.
Teng, M., Usman, N., Frederick, C. A., & Wang, A. H.-J. (1988) *Nucleic Acids Res. 16*, 2671–2690.
Westhof, E. (1987) *J. Biomol. Struct. Dyn. 5*, 581–600.
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. E. (1980) *Nature 287*, 755–758.
Zimmer, C., & Wähnert, U. (1986) *Prog. Biophys. Mol. Biol. 47*, 31–112.