# Refinement of Netropsin Bound to DNA: Bias and Feedback in Electron Density Map Interpretation[†,‡]

David S. Goodsell,[§] Mary L. Kopka, and Richard E. Dickerson*

*Molecular Biology Institute, Department of Chemistry & Biochemistry, and Institute of Geophysics & Planetary Physics, University of California, Los Angeles, California 90095*

*Received December 16, 1994; Revised Manuscript Received February 10, 1995*[⊗]

ABSTRACT: The X-ray crystal structure of the complex of the B-DNA dodecamer CGCGAATTCGCG with the antitumor drug netropsin has been reexamined to locate the drug accurately for computer-based drug design. The optimum solution is with the drug centered in the AATT region of the minor groove, making three good bifurcated hydrogen bonds with adenine N3 and thymine O2 atoms along the floor of the groove. Pyrrole rings of netropsin are packed against the C2 positions of adenines, leaving no room for the amine group of guanine and, hence, providing a structural rationale for the A·T specificity of netropsin. An alternative positioning in which the drug is shifted along the minor groove by ca. one-half base pair step is rejected on the basis of free *R* factor calculations and the appearance of the original drug-free difference maps. Final omit maps, although of more pleasing appearance, are not a dependable means of discriminating between right and wrong structures. The shifted alternative drug position ignores potential hydrogen bonding along the floor of the groove, provides no explanation for netropsin's observed A·T specificity, and is contradicted by NMR results [Patel, D. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79,* 6424].

The first crystal structure of an antitumor drug bound within the minor groove of B-DNA, netropsin (Figure 1) complexed with CGCGAATT$^{5Br}$CGCG (Kopka et al., 1985a−c), and the nuclear magnetic resonance analysis of netropsin with CGCGAATTCGCG (Patel, 1982) successfully explained both the groove-binding properties of the drug and its sequence preference for regions of A·T base pairs (Van Dyke et al., 1982). This A·T preference can be attributed to three factors: (a) a deeper minor groove in A·T than G·C regions because of the absence of the N2 amine group of guanine; (b) a narrower minor groove because its walls are brought closer together by the greater propeller twist of A·T base pairs [see Figure 9 of Fratini et al. (1982)]; and (c) a deeper electrostatic potential well within A·T regions of the minor groove. The first factor was proposed by Patel from his NMR studies and confirmed by X-ray analysis, the second followed from the X-ray analysis, and the third factor came from theoretical studies (Pullman & Pullman, 1981; Pullman, 1983; Zakrzewska et al., 1987).

This netropsin/DNA structure and its explanation for A·T specificity have been the basis for design and synthesis of "lexitropsins", analogues of netropsin capable of recognizing and interacting with longer binding sites and of reading G·C base pairs at specific loci along the site. The synthesis and testing of lexitropsins is a subject of active research (Lown et al., 1986a,b, 1989; Kissinger et al., 1987; Younquist & Dervan, 1987; Burckhardt et al., 1989; Wang & Lown, 1992). It has been found that sequence-specific drug binding to a longer specificity region is best obtained by connecting two

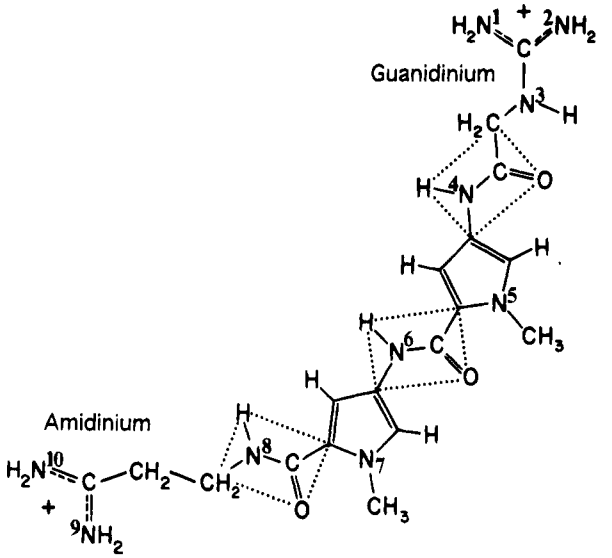


FIGURE 1: Structure of netropsin, with nitrogen atoms numbered. Amide NH groups N4, N6, and N8 are potential hydrogen-bonding agents to the floor of the minor groove in complexes with B-DNA. During the refinement described in this paper, the three amide groups, the two methylpyrrole rings, and the guanidinium and amidinium tails were individually restrained to planar geometry, as demanded by their bonding structure. In particular, no rotation is permitted about bonds: C−N3, C−N4, C−N6, or C−N8. Guanidinium torsion angles N1−C−N3−C and N2−C−N3−C should be 0° and 180°, respectively.

netropsins or lexitropsins together via a short connecting chemical group or linker. Initial design of these linkers involved trial and error and chemical intuition, but for a systematic, computer-based design of linker groups such as that discussed by Walker et al. (1995), one must know the precise positioning of drug monomers on the DNA duplex. Even a slight mispositioning of drug will lead to erroneous geometry for the linkers to be synthesized.

[†] This research was carried out with the assistance of American Cancer Society Grant DHP-22F and NIH Grant GM31299.

[‡] Final refined coordinates are available from the Brookhaven Protein Data Bank under file name 101D.

[§] Current address: Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037.

[⊗] Abstract published in *Advance ACS Abstracts,* April 1, 1995.

0006-2960/95/0434-4983$09.00/0 

Table 1: Structures of B-DNA Complexes of Netropsin and Distamycin[a]

| | number of reflections $F_o > 2\sigma(F)$ | resolution (Å) | final $R$ factor | number waters located | reference |
|---|---|---|---|---|---|
| | | Class I | | | |
| Nt/CGCGAATT$^{5Br}$CGCG | 2528 | 2.2 | 0.211 | 75 | Kopka et al., 1985a–c |
| Nt/CGCGAATT$^{5Br}$CGCG | 2430 | 2.25 | 0.163 | 37 | this work |
| Nt/CGCAAATTTGCG | 2348 | 2.2 | 0.198 | 72 | Tabernero et al., 1993 |
| Ds/CGCAAATTTGCG | 2369 | 2.2 | 0.196 | 70 | Coll et al., 1987 |
| | | Class II | | | |
| Nt/CGCGATATCGCG | 1848 | 2.4 | 0.200 | 60 | Coll et al., 1989 |
| Nt/CGCGAATTCGCG | 2056 | 2.2 | 0.164 | 74 | Sriram et al., 1992 |
| Nt/CGC$^{6Et}$GAATTCGCG | 1514 | 2.5 | 0.156 | 89 | Sriram et al., 1992 |

[a] Nt, Netropsin; Ds, distamycin.

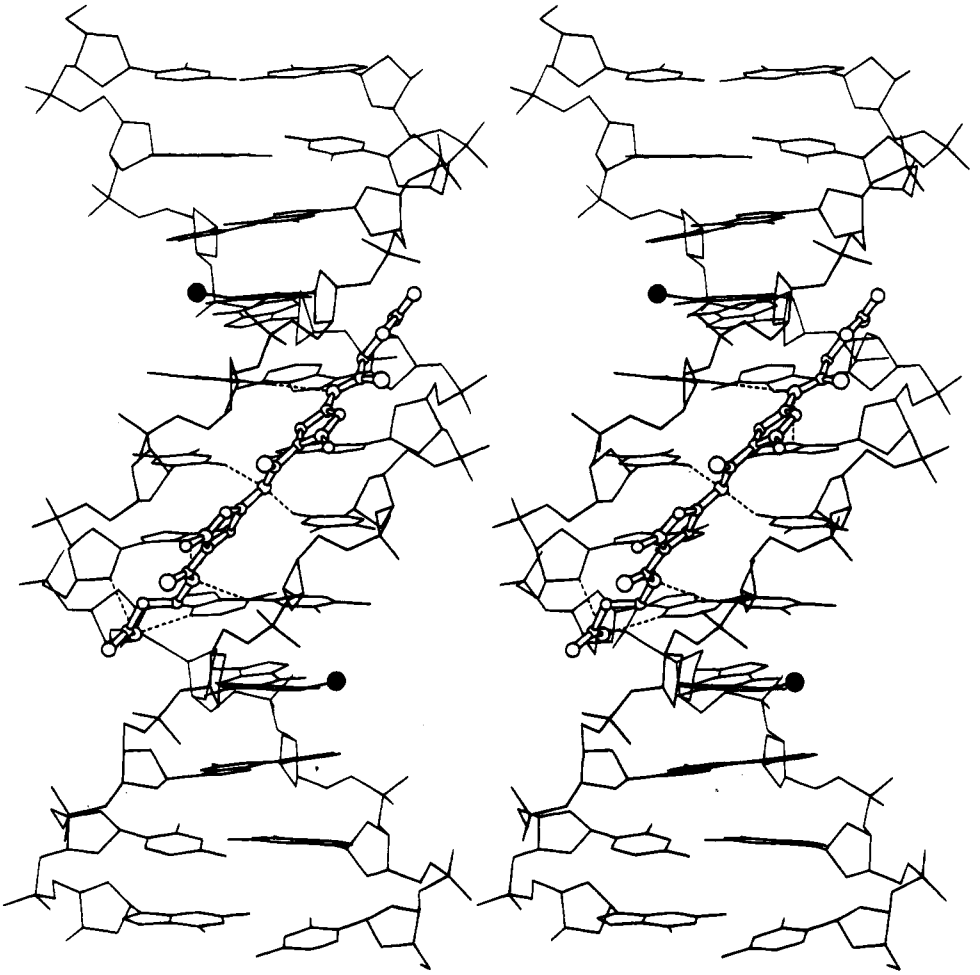

FIGURE 2: Structure of netropsin bound to CGCGAATT$^{5Br}$CGCG, as re-refined in this paper. Base pair C1•G24 is at the top of the helix as drawn here, and the guanidinium group is at the upper end of the drug molecule. Dark spheres are the two bromine atoms on 5-bromocytosines. Hydrogen bonds of 3.5 Å or less are dotted, with distances given in Figure 3. The three amide nitrogens make bifurcated hydrogen bonds with thymine O2 and adenine N3 as schematized in Figure 3a. The amidinium end is hydrogen-bonded to an adenine N3 (3.06 Å) and a sugar O4′ (2.67 Å). This re-refined structure is virtually identical to that originally published by Kopka et al. (1985a–c).

Several 1:1 complexes of netropsin bound to various dodecanucleotides have been reported since the original structure of Kopka et al. of 1985.[1] These, listed in Table 1, fall into two general structural models or classes, differing in the location of drug along the DNA double helix. Class I includes the complexes of netropsin with CGCGAATT$^{5Br}$CGCG and with CGCAAATTTGCG and of the related distamycin with the latter sequence. In these, each amide group of netropsin is positioned approximately midway between two successive base pairs (Figure 2). The amide NH groups make bifurcated hydrogen bonds to adenine or thymine bases on opposite strands and on two adjacent base pairs, exactly like the water bridges observed in the minor groove spine of hydration in the unliganded DNA (Drew & Dickerson, 1981; Kopka et al., 1983). This pattern of bifurcated hydrogen bonds is shown in Figure 3 for netropsin/CGCGAATT$^{5Br}$CGCG, for netropsin/CGCAAATTTGCG, and distamycin/CGCAAATTTGCG.

[1] This paper deals only with complexes in which a single drug molecule sits along the minor groove. Side-by-side complexes, such as have been reported by Wemmer et al. (Pelton & Wemmer, 1989, 1990; Mrksich et al., 1992; Geierstanger et al., 1993, 1994a,b) and by Sundaralingam et al. (Chen et al., 1994), pose a quite different DNA recognition problem.

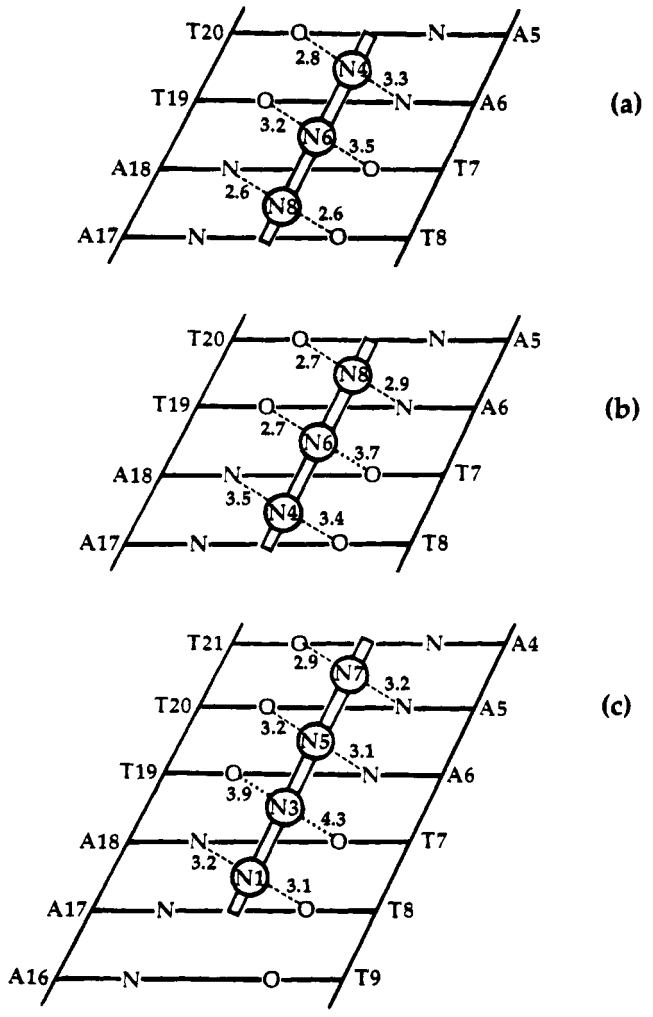


FIGURE 3: Bifurcated hydrogen bonding between amide NH on the drug molecule and DNA minor groove N and O atoms in three class I structures (Table 1), for which only a single drug orientation is reported, and for which the pyrrole rings sit opposed to individual base pairs along the helix axis: (a) netropsin complexed with CGCGAATT$^{5Br}$CGCG, (b) netropsin complexed with CGCAAATTTGCG, and (c) distamycin complexed with CGCAAATTTGCG. Successive amide NH groups along the drug molecule are numbered N4–N6–N8 for netropsin and N1–N3–N5–N7 for distamycin. N–N or N–O distances are given in angstroms. Distances of 3.5 Å or less are dashed and may be interpreted as hydrogen bonds. Longer distances are dotted and shown for reference, but their interpretation as bonds is problematic. Straight segments of the drug backbone between numbered amide circles mark the locations of pyrroles.

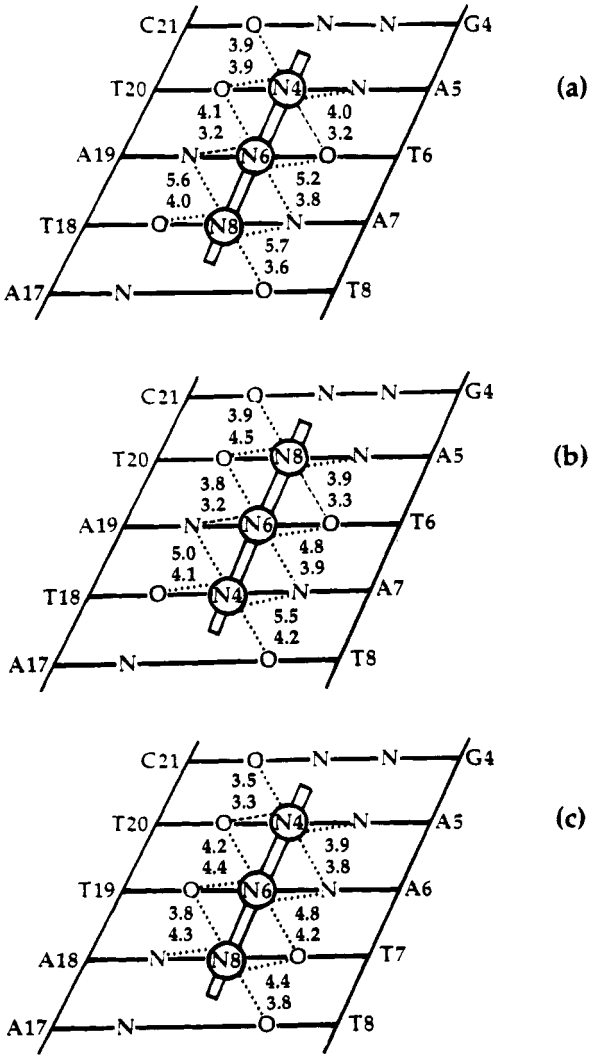


FIGURE 4: Distances between drug amide nitrogens and base pair O and N atoms along the floor of the minor groove, for class II structures (Table 1). Pyrrole rings now sit between base pairs and offer no explanation for the observed A·T base specificity of netropsin. The orderly pattern of hydrogen bonding of Figure 3 is gone, and only a few distances are short enough to be regarded as possible hydrogen bonds at all. (a) Netropsin complexed with CGCGATATCGCG in the down orientation. (b) Netropsin complexed with CGCGATATCGCG in the up orientation. (c) Netropsin complexed with CGCGAATTCGCG in the down orientation. In each case, an absolute orientation of the drug molecule in the crystal could not be determined, and the structure was refined as though it had end-for-end disorder. However, coordinates for the up orientation of the second structure (c) were not deposited with the Protein Data Bank. Only 3 of the 36 distances shown here are short enough to be interpreted as hydrogen bonds; the other 33 are reduced to puzzling long range dipole/dipole interactions in a generally hydrophobic environment.

In these class I complexes, each pyrrole ring nests against the adenine C2 of one A·T base pair. Indeed, this particular positioning of the drug against DNA is the key to its A·T specificity: if a G·C base pair were present instead of A·T, its N2 amine would push the pyrrole ring away. Moreover, this positioning of drug on DNA also is central to the original idea for making a G·C-reading lexitropsin, since an imidazole substituted for pyrrole would be in the correct position to accept a hydrogen bond from a guanine amine, thereby strengthening the drug/DNA association.

A second mode of drug binding has been proposed in three recent complexes of netropsin with CGCGATATCGCG, CGCGAATTCGCG, and CGC$^{6Et}$GAATTCGCG (Table 1). In these class II models, the drug molecule is interpreted as exhibiting end-for-end disorder: either the drug is truly disordered as it sits in the crystal or the resolution of the X-ray analyses is insufficient to permit the assignment of a single definite orientation. Furthermore, in these class II interpretations, the entire drug molecule is shifted down the minor groove by approximately one-half a base pair step. Amide nitrogens now are positioned roughly in the plane of the bases, and pyrrole rings are located between base pairs. No regular pattern of bifurcated hydrogen bonds now exists, and as Figure 4 indicates, distances between amide nitrogens of the drug and base pair N and O atoms are only occasionally short enough to be described as hydrogen bonds at all. Interactions between the drug molecule and the floor of the minor groove appear to be completely unsystematic,

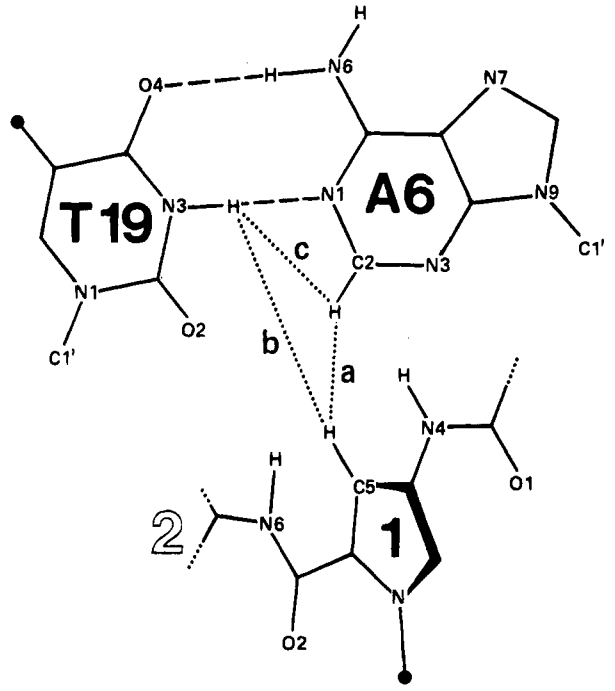


FIGURE 5: One of the two close H−H contacts in the class I positioning of netropsin against CGCGAATTCGCG. Distance a is less than 2.0 Å and represents a van der Waals contact of hydrogen atoms. Introduction of an amine on adenine C2 would push the drug molecule away. Corresponding distances in class II positioning are 3.5−4.4 Å (Table 2), offering no stereochemical basis for the observed A·T specificity of netropsin [from Dickerson and Kopka (1985)].

and no reason exists for the drug to be located where it is rather than being shifted up or down the groove at random. This positioning of the drug molecule also eliminates the previously mentioned explanation for the A·T specificity of netropsin and its cousin distamycin, and no alternative explanation of sequence specificity is offered.

The class II positioning of netropsin not only disagrees with the original X-ray analysis but it also disagrees with nuclear Overhauser NMR studies of netropsin/CGCGAATTCGCG carried out by Patel (1982). Patel analyzed NOE shifts involving thymine imino protons, adenine C2 protons, and netropsin pyrrole ring protons on the concave (H-3) and convex (H-5) edges of the netropsin molecule (Figure 1). He concluded: "The NOE experiments convincingly establish two contacts between the antibiotic and the DNA, each involving a pyrrole H-3 proton of netropsin and an adenosine H-2 of dA·dT base pair 6 of the dodecanucleotide." (Base pair 6 in Patel's nomenclature is either of the two central A·T base pairs: our A6·T19 or T7·A18.) "Thus, a third contact is established between the antibiotic and the dodecanucleotide involving the adenosine H-2 of dA·dT base pair 5 and the $CH_2$ segment at one end of the netropsin molecule." (Base pair 5 is either of the outer A·T pairs: our A5·T20 or T8·A17.) The proton−proton distance involved is shown by a in Figure 5. In the class I structure where pyrrole rings sit opposite A·T base pairs, the measured H−H distances are 1.39 and 1.75 Å (Table 2), well within the limits for NOE energy transfer. But in the class II structure where pyrrole rings are positioned between two successive base pairs, adenine C2-to-pyrrole H−H distances are all longer: 3.5−4.4 Å. No adenine-to-pyrrole H−H distances close enough to produce the observed NOE intensities exist in the class II positioning of drug. Hence, this model is contradicted by NMR evidence.

Table 2: Proton−Proton Distances in a Class I and a Class II Netropsin/DNA Complex[a]

| | | H−H distances (Å) | |
|---|---|---|---|
| from adenine C2-H | to pyrrole C-H | class I | class II |
| A5 | ring 1 | 4.65 | 4.22 |
| A6 | ring 1 | 1.75[b] | 3.49 |
| A6 | ring 2 | 4.84 | 4.20 |
| A18 | ring 2 | 1.39 | 4.39 |

[a] Ideal H atom positions were generated for the two structures on the basis of coordinates of the heavier atoms, C, N, and O. Class I: netropsin/CGCGAATT$^{5Br}$CGCG (this work). Class II: netropsin/CGCGAATTCGCG (Sriram et al., 1992). [b] Distance a in Figure 5.

One crucial distinction between the class I and class II structure analyses may be the number of experimental observations above $2\sigma$ in $F$. All of the structures in Table 1 are isomorphous and have essentially the same cell dimensions and unit cell volumes. Hence, a direct comparison of the number of reflections stronger than $2\sigma$ is meaningful. This number is 2348−2528 for the class I structure determinations, but 1514−2056 for the class II analyses. The class I determinations have, on average, 34% more information content than the class II structures. Hence, many of the problems discussed in this paper may stem from a lack of resolution of the data sets.

Because the precise alignment of the netropsin molecule on DNA is so critical to systematic drug design, especially the synthesis of bis-linked dimers (Walker et al., 1995), we have reexamined the issue of class I vs class II positioning of the netropsin molecule in its DNA complex. Using the original netropsin/CGCGAATT$^{5Br}$CGCG data, we have re-refined netropsin both in the class I position presented by Kopka et al. (1985a−c) and in the class II alternative reported more recently by Sriram et al. (1992). As will be seen in the course of this paper, both models may be refined to an acceptable residual error or $R$ value, and both models yield a reasonably convincing final omit map at the conclusion of refinement. However, the class I position best accounts for the difference electron density that is observed in completely unbiased difference maps calculated *before* drug atoms are incorporated into the model, and free $R$ factor calculations give a small but definite preference to the class I positioning of netropsin on DNA. Furthermore, the class I structure offers a chemical reason for being where it is, whereas the class II structure appears to be positioned at random.

To anticipate the conclusion of our reanalysis, these trials illustrate the fact that a subsequent course of crystallographic refinement is highly biased by any molecular model that is built into it. Unless they are grossly wrong, atoms once placed into difference density will tend to feed back to yield reasonable electron density and omit map difference density. As Heinemann and Alings have noted (1991), at the intermediate resolution of most drug/DNA analyses, an incorrect structure can yield an apparently acceptable residual error or $R$ factor. Neither a pleasing final omit map nor a reasonable $R$ factor is a sure guarantee of the correctness of drug positioning. Experience described in this paper suggests that the only dependable route to correct drug positioning is careful placement of the drug framework within *the original difference electron density map*, after preliminary DNA refinement but before any assumptions about drug location

Table 3: Conventional and Free $R$ Factor Refinement of Class I and Class II Models

| | $R_C$ factor with 90% data used for refinement | | $R_F$ factor with 10% data sequestered from refinment | |
|---|---|---|---|---|
| stage of analysis | class I | class II | class I | class II |
| (1) refinement of DNA prior to drug modeling | 0.207 | 0.207 | | |
| (2) after initial positional and $B$ refinement | 0.186 | 0.191 | 0.248 | 0.254 |
| (3) initial solvent peak adjustments | 0.177 | 0.183 | 0.247 | 0.260 |
| (4) final, with 36 solvent ions and 1 magnesium ion | 0.163 | 0.167 | 0.252 | 0.262 |

have been made. This first, completely unbiased difference map should always be published, as it offers the best evidence of drug positioning. The traditionally published final omit map density, although aesthetically more pleasing, is devoid of critical information.

## MATERIALS AND METHODS

Netropsin bound to CGCGAATT$^{5Br}$CGCG was refined in two positions: the class I structure of Kopka et al. (1985a–c), with pyrroles aligned with base pairs as in Figure 3, and the shifted class II position of Sriram et al. (1992), with pyrroles situated between base pairs as in Figure 4. X-ray structure factors, $F_o$, were taken from PDB listing R6BNASF, and a $2\sigma$ cutoff on $F$ was applied, as in the original report of the netropsin structure. (All $R$ values are calculated with this $2\sigma$ cutoff.) The DNA dodecamer was positioned by molecular replacement (Brunger, 1993), starting with native CGCGAATTCGCG coordinates from PDB listing 1BNA. NUCLSQ refinement of atomic positions and temperature factors (Hendrickson & Konnert, 1980), using data from 8.0 to 2.25 Å, resulted in a model with a residual error of $R$ = 0.247.

Solvent peaks then were added in groups of 5–10, carefully avoiding the large, continuous drug difference density visible in the minor groove. Water molecules were assigned only to peaks that were simultaneously (a) greater than $3\sigma$ in $(F_o - F_c)$ difference density; (b) greater than $1\sigma$ in $(2F_o - F_c)$ density; and (c) within proper H-bonding distance of DNA atoms or other solvent peaks. Positional and temperature factor refinement followed each round of solvent addition. Solvent postions were checked after each round of refinement by calculation of an $(F_o - F_c)$ difference map, omitting all solvent positions from phasing. Positions that did not return density greater than $2.5\sigma$ in this omit map were deleted. The requirement that a prospective solvent peak must appear in *both* $(F_o - F_c)$ and $(2F_o - F_c)$ density maps, and return in $(F_o - F_c)$ solvent omit maps, led to a stable and dependable water structure.

After the addition of 30 solvent positions, but prior to the introduction of a drug molecule, the residual error stood at $R$ = 0.207 for the 2430 reflections above $2\sigma$ in $F$ at a resolution of 2.25 Å. At this point a difference map was calculated in which DNA and solvent atoms were subtracted out. The remaining unexplained electron density is shown in Figure 6. This difference density then was fitted with a netropsin molecule in both the class I and class II positions, shown by the black skeletons in Figure 6a,b, respectively. Class I coordinates were taken directly from Protein Data Bank coordinate file 6BNA (Kopka et al., 1985a–c). Netropsin coordinates in the shifted class II position were obtained by first least-squares fitting the central four AATT base pairs of the Sriram et al. (1992) structure (PDB coordinate file 1D86) to the central four base pairs of our DNA under refinement and then transferring drug coordinates to our structure. The RMS difference in DNA atoms between the two structure determinations for these four bases was 0.55 Å. For both class I and class II models, the temperature factors of all drug atoms initially were set to 20. Figure 6 compares the degree to which the two possible netropsin positions fit the original, unbiased drug density, before the refinement of drug was begun. As will be seen in the following, the comparison of parts a and b of Figure 6 is critical to a final judgement on the relative merits of the two netropsin positions.

At this point in each refinement, 10% of the data chosen at random was sequestered for a free $R$ factor or $R_F$ calculation over the drug refinement process (Brunger, 1993). The $R_F$ factor is a technique for detecting feedback errors by isolating a subset of the data from subsequent refinement. The remainder of the data is used for refinement in the normal manner. At the conclusion of refinement, a conventional $R$ factor or $R_C$ calculated with the other 90% of the data may have a low value either (a) because the structure is correct or (b) because a high ratio of refinement parameters to intensity data has permitted the adjustment of enough parameters to fit the observed data with a poor structure. With (a), the $R_F$ factor calculated from the sequestered 10% of the data should also be low, because the refinement was based on a real structure. This real structure will be reflected in the sequestered $F_o$ data, even though they played no part in the refinement process. But if an incorrect but overparametrized model has been fitted successfully to the 90% data set, there is no reason why fitting should be equally good for the sequestered 10%. Hence, the $R_F$ factor, calculated with the 10% of the data that was not used in refinement, is a strong check on the correctness of that refinement.

The course of the refinement of class I and class II models is listed in Table 3 and displayed in Figure 7. After the initial refinement of positions and temperature factors of the drug, $R_C$ factors for the two models were only 0.005 apart, and $R_F$ factors differed only by 0.006 (Table 3, stage 2). Five solvent peaks in class I and six in class II faded from subsequent difference maps and were deleted. These solvent peaks presumably filled noise ripples caused by the absence of drug atoms from the earlier model. Five new solvent positions were added to each model, and refinement was continued. Positional and temperature factor refinement kept $R_C$ values for the two models to within 0.006 of one another, but the separation of $R_F$ values widened to 0.013 (Table 3, stage 3).

Refinement continued by the addition of solvent peaks, followed by positional and temperature factor refinement. All refinement parameters were held at the same values in the two parallel processes. Refinement statistics are given in Table 4. The final models (Table 3, stage 4) include 36

Table 4: Refinement Parameters

| | RMS deviations from ideality | | |
|---|---|---|---|
| | class I | class II | target $\sigma$ |
| Geometric Restraints | | | |
| bonds (B + S + N)[1] | 0.024 | 0.024 | 0.030 |
| angle distances (B + S + N) | 0.040 | 0.041 | 0.040 |
| bonds (P) | 0.026 | 0.027 | 0.025 |
| angle distances (P) | 0.057 | 0.055 | 0.050 |
| chiral volumes (S) | 0.161 | 0.153 | 0.150 |
| planes (B+N) | 0.014 | 0.014 | 0.020 |
| single torsion contact | 0.093 | 0.093 | 0.100 |
| multiple torsion contact | 0.097 | 0.098 | 0.100 |
| hydrogen bonds | not restrained | | |
| Isotropic Thermal Restraints | | | |
| across bonds (B + S + N) | 4.282 | 4.468 | 6.000 |
| across angles (B + S + N) | 4.990 | 5.125 | 6.000 |
| across bonds (P) | 5.693 | 5.670 | 6.000 |
| across angles (P) | 5.227 | 5.126 | 6.000 |

[a] B, base atoms; S, sugar atoms; P, phosphate atoms; N, netropsin atoms.

solvent positions and one magnesium ion (located by its unusually low *B* value and by coordination of solvent positions). Final refined coordinates are available from the Brookhaven Protein Data Bank as PDB File 101D. It should be noted that the other analyses listed in Table 1, including the previous analysis of the present data set, ultimately included 60−89 water molecules, but the 36 solvent peaks in the present analysis are those in whose reality we have the most confidence. We felt that, at an *R* factor of 0.163, the point about refinement strategy had been made and that no further purpose would be served by continuing to add more solvent peaks merely to lower the *R* factor.

## RESULTS AND DISCUSSION

*Refinement Results.* Refinement of DNA structure with NUCLSQ has been well characterized in numerous structure analyses. Hahn and Heinemann (1993) have shown NUCLSQ to compare favorably with, or even to be superior to, TNT or XPLOR for DNA structure refinement. Standard weighting schemes were employed in the present study, and identical weighting schemes were used for models I and II. Hence, their results may be compared directly in terms of R factors and standard deviations from ideal geometry.

For class I and class II models, Figure 6a,b compares the initial ($F_o - F_c$) difference electron density calculated just before the addition of drug molecules. Figure 8a,b compares the final ($F_o - F_c$) omit electron density maps, calculated by deleting the fully refined drug molecule from the phase calculation at the very end of the analysis. Only the drug molecule and any as yet unexplained solvent peaks should appear in either map; the DNA itself has been subtracted out. The concave edge at the left of the drug molecule fits along the floor of the minor groove. The convex edge of the difference density at the right extends toward the opening of the groove. In the initial difference map of Figure 6, the difference density is distinguished by three central protruberances, with smaller spacing between the upper pair than the lower pair. The upper end of the density is well defined, and a large spherical peak sits immediately above and to left of the continuous density. It should be noted that this spherical peak *does not lie in the plane* of the continuous drug density; it is closer to the viewer, as seen in Figure 6. At the bottom, the drug density tapers off in a weaker tail.

Class I and class II structures differ in how these experimental features of the unbiased difference map are interpreted. The netropsin molecule itself has five projections along its convex edge, as labeled in Figures 6 and 8: carbonyl groups a, c, and e and pyrrole methyl groups b and d. Projecting features b and c are close together, as are d and e. In model I (Figure 6a), the central density protruberance is assigned to amide carbonyl c, fitting pyrrole methyls b and d into flanking density bulges. The relative spacing of features b−d is correct if the drug molecule is considered to run from guanidinium at the top to propylamidinium at the bottom. The upper difference density then is precisely the proper length for a guanidinium tail. The inherently more flexible amidinium end fits loosely into weaker density at the bottom. Carbonyls a and e are both contained within difference density, and the flat region of drug density in the vicinity of protruberance b is assigned to a pyrrole ring. The disconnected peak at the upper left is interpreted as a solvent position, within correct hydrogen-bonding distance and geometry to one of the guanidinium nitrogens.

In contrast, in model II (Figure 6b) the central density protrusion is assigned to pyrrole methyl group d and the two flanking peaks to amide carbonyls c and e. This shifts the entire drug upward by about 2 Å. It places the amidinium tail at the bottom in strong density over its entire length and beyond, but pushes both pyrrole methyl b and the guanidinium end out of density. Sriram et al. (1992) twisted the guanidinium group out of planarity in order to extend it into the disconnected peak at the top left. Torsion angles $N_1$−C−$N_3$−C and $N_2$−C−$N_3$−C, which in a planar guanidinium would be 0° and 180°, respectively (Figure 1), are 75° and 140° instead in their structure. In the present reanalysis, we chose not to break the known planarity of the guanidinium group and, hence, could not fit the group into the disconnected peak as snugly as is seen at lower left of Figure 3 of Sriram et al. (1992).

In the class II interpretation the most conspicuous flattened disk of difference density is occupied, not by a pyrrole ring but by the peptide bond involving carbonyl group c. Sriram et al. (1992) reported that the drug skeleton could be fitted equally well either as in Figure 6b or inverted end-for-end, with amidinium at the top and guanidinium at the bottom. Presumably, in that case, the amidinium tail then would have to be fitted into the unconnected spherical peak at the upper left of Figure 6. We have refined only the guanidinium-up model, as only its coordinates were available from the Protein Data Bank. It should be noted that, if the netropsin molecule is fitted in two different end-for-end orientations in an off-center density region nearer to one end of the helix than the other, then this corresponds to two chemically distinct binding modes in the DNA/drug complex when free in solution—one with the guanidinium nearest to an end of the self-complementary double helix and one with the amidinium end nearest. It is unlikely that such dual binding modes should exist in solution. Hence, any end-for-end disorder of this type reported in the crystal probably should be interpreted as a resolution-based inability to distinguish the ends of the drug image, not a true molecular disorder.

The final refined structure omit maps in Figure 8a,b both show reasonably acceptable density, illustrating the fact that such maps have little value in discriminating between structures. All of the distinctive features of the netropsin

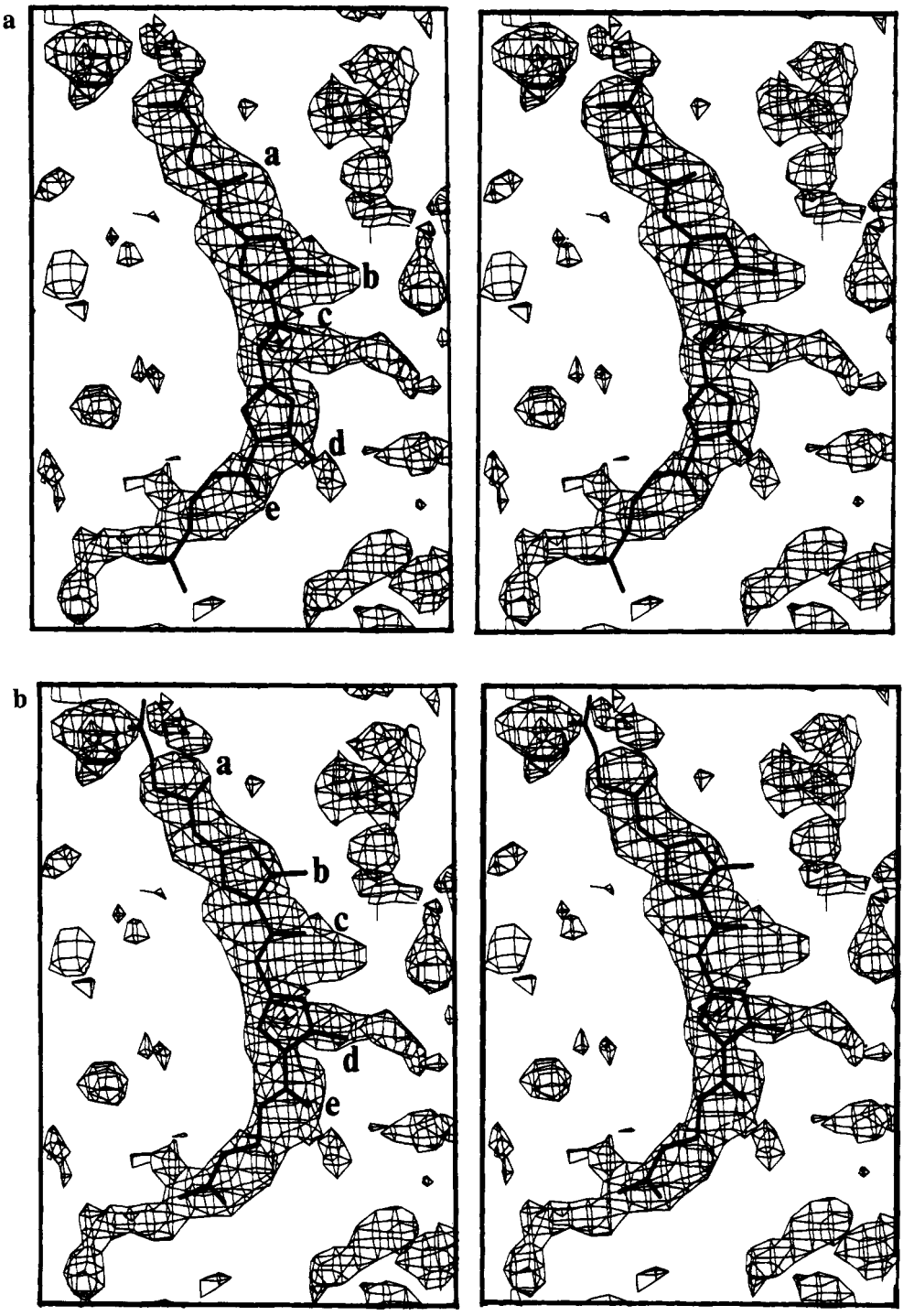


FIGURE 6: Initial, unbiased difference electron density maps for netropsin complexed with CGCGAATT$^{5Br}$CGCG and with the drug skeleton in its (a) class I or (b) class II position. Contour level is 1.5$\sigma$. DNA base pair C1·G24 would be at the top of the stereo drawing and G12·C13 is at bottom. Letters a–e on the netropsin skeleton indicate protruding features used to fit drug to density: a, carbonyl 1; b, pyrrole methyl 1; c, carbonyl 2; d, pyrrole methyl 2; e, carbonyl 3. Note that features b and c are close together, as are d and e. (a) In the class I model, projections b, c, and d fit the density well and have the proper relative spacing. Carbonyl groups a and e at least lie within density, as does the guanidinum at the top. Only the flexible amidinium tail extends out of difference density. (b) With the drug in a class II position, features c, d, and e are fitted into density, although their spacings require that c be pushed to one extreme of a very large side peak that is otherwise unused. Peak b violates the density, and the guanidinium tail extends out of the drug density toward an out-of-plane solvent peak. The guanidinium is held planar in this re-refinement, although in the earlier class II refinement (Sriram et al., 1992) the planarity of the guanidinium was broken in order to extend it farther into the solvent peak. A continuation of the drug density at lower left is unaccounted for in this class II placement.

model are visible in the class I omit density (Figure 8a): forked tails at both ends, protrusions for both pyrrole methyls (b, d), and all three carbonyls (a, c, e). Density for the class II model is also generally good, showing problems only at carbonyl b and the guanidinium tail. Carbonyl b was not well fit into density in the original unbiased map (Figure 6b), and subsequent refinement was unable to supply a protruberance for this feature. The guanidinium group fits badly into what is an obvious solvent peak even after refinement. Figure 8b should be compared with Figure 3

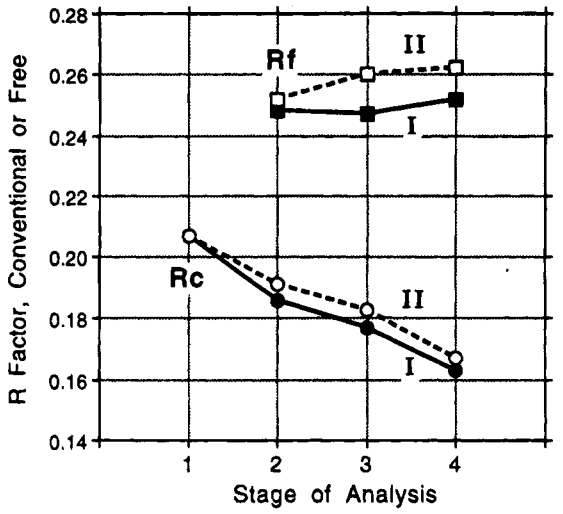


FIGURE 7: Course of $R_F$ factor (top) and $R_C$ factor refinement (bottom) for class I and class II models of the netropsin/DNA complex. The four stages of analysis are described in Table 3.

of Sriram et al. (1992), which is also a final refined omit map. The points of similarity between the two maps indicate the robustness of the refined difference density image, as the two maps were calculated with independently collected X-ray data sets from a brominated and an unbrominated decamer.

*Comparison of Final Refined Class I and Class II Models.* Classical $R_C$ values for the two final models differ by only 0.004. Either model alone would have been accepted by this criterion, and one is barely justified in making a choice between them on the basis of classical $R$ factors, even with both models to compare. But the $R_F$ factors differ by 0.010, and this, combined with the appearance of the unbiased difference maps of Figure 6, constitutes evidence for the correctness of the class I model.

The final DNA structures from class I and class II refinements are virtually identical, with an RMS deviation of 0.16 Å on all DNA atoms. This RMS value was calculated directly from crystallographic coordinates, without least-squares fitting of one helix onto the other. Hence, the two DNA helices are identical in structure and occupy identical positions in the unit cell. Water structures also are very similar. Twenty-seven of the thirty-six solvent positions are identical to less than 1 Å. Twenty-five of these common positions were established prior to fitting of the drug, and only two appeared after introduction of the drug. The nine remaining solvent peaks, which do not agree within 1 Å, are the product of later refinement with the drug in place.

Netropsin in its refined class II position differs slightly from the structure of Sriram et al. (1992), with an RMS deviation of 1.02 Å over atoms of the drug molecule. This difference is localized primarily in the guanidinium tail and arises because we imposed planarity of the $NHC(NH_2)_2^+$ group as a condition of refinement. As has been mentioned, in the earlier analysis, the planarity of the guanidinium group was broken in order to fit the end of the molecule into the large off-planar solvent peak at one end of the drug molecule.

Netropsin in its refined class I position is identical to the original structure of Kopka et al. (1985a–c) with an RMS deviation of 0.60 Å over all drug atoms. The force field used in the original EREF refinement of netropsin bound to CGC-GAATT$^{6Br}$CGCG led to an unstable refinement, particularly when modeling the bromine atoms. That structure, therefore, had several stereochemical problems and yielded an $R$ value at the high end of those reported for similar dodecamer structures (Table 1). The present refinement is more stable, building on the experience of a subsequent 10 years of dodecamer and decamer structural refinements. The major difference between the structure of Kopka et al. and the current refined model, apart from better overall stereochemistry, is that all three amide nitrogen atoms now make bifurcated hydrogen bonds with base edges (Figure 3a). In the 1985 reports, interactions with amide N6 were slightly too long to be convincingly assigned as hydrogen bonds, like the interactions with amide N3 in the distamycin complex diagrammed in Figure 3c. The regularity of the present model strengthens the arguments for a steric basis of the A·T specificity of netropsin: bifurcated hydrogen bonds position the drug on the DNA, forcing the pyrrole rings into close contacts with base edges, which result in A·T specificity.

*Choosing the Proper Model.* This comparison of class I and class II positioning of netropsin on DNA is perhaps a little unfair. As noted earlier, the analyses that yielded class I locations for the drug have, on average, 34% more data than those that led to class II positions. They also permitted the establishment of drug orientation, whereas the lower resolution analyses could only invoke end-for-end disorder. Even with 2348 reflections greater than $2\sigma$, Tabernero et al. (1993) comment that end-for-end disorder was considered in their class I structure. However, one orientation seemed to fit better than the other, even though the difference in the final $R$ factor was only 0.002. On the issue of the position of the drug molecule (centered class I vs off-center class II), they conclude: "There was no evidence of disorder in the position of the netropsin molecules along the A·T stretch."

Our analysis agrees with that of Tabernero et al. (1993). The X-ray evidence for class I positioning of the drug molecule, centered in the AATT region with pyrrole rings opposed to base pairs, is modest but ultimately convincing. It is corroborated by the fact that the NMR evidence of Patel (1982) flatly contradicts a class II drug positioning. The strongest X-ray evidence is the appearance of the original difference density map, before any assumptions about drug location had been added to the analysis (Figure 6). The $R_F$ factor evidence of Figure 7 points in the same direction, and even the final omit maps (Figure 8) are most consistent with the correctness of the class I model, although the refinement of each individual model has blurred the distinction present in the unbiased maps of Figure 6. But most disturbingly, from both final difference maps and conventional R factors, each of the two positions for the drug molecule probably would have been regarded as acceptable, *if examined alone without comparison with alternates.* This comparative analysis confirms the original centered positioning of netropsin on B-DNA of Kopka et al. from 1985, but also illustrates the great care that must be taken to consider all possibilities in an intermediate resolution structure analysis such as this.

As Tabernero et al. (1993), so we also find a small preference for one end-for-end orientation over the other or over an assumed end-for-end disorder. Conventional or even free $R$ factors are too close to call. Ultimately, one is driven to rely on the appearance of the original, unbiased difference

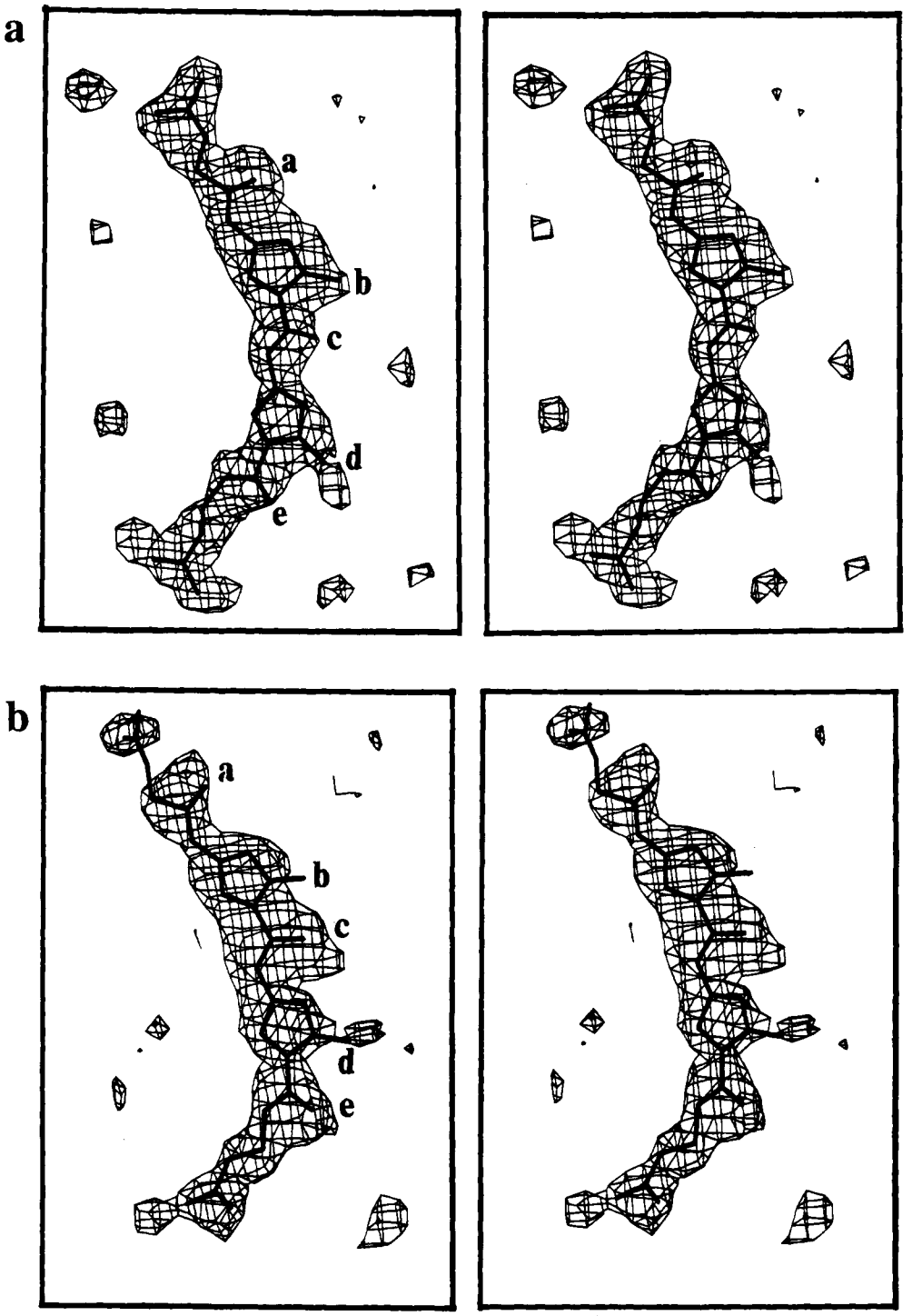


FIGURE 8: Final omit maps, in the same orientation as Figure 6. These are calculated by deleting the drug from the fully refined DNA/drug complex and computing a difference map with phases from DNA and solvent molecules only. Contour level is 2.0$\sigma$. These final omit maps are much cleaner than the initial maps of Figure 6, which is why people like to publish these rather than the earlier maps. But the maps of Figure 6 contain real information, with which decisions can be made; these maps of Figure 8 are merely ornamental. Although the intent of deleting the drug is to confirm its correctness by examining the difference map image that reappears, information on the refined drug position is in fact implicit in the rest of the structure. Hence, the omit map drug image is biased by assumptions regarding the refined drug position. (a) Class I model: the upper pyrrole ring is well-centered within a broad, flat region of density, and the lower pyrrole is acceptable, with a suitably positioned peak for its methyl group at d. Both ends of the drug molecule are well-fitted. (b) Class II model: the flat region now is used only for a peptide bond, pyrrole methyl group b lies out of density, and the guanidinium tail at the top extends through an empty region to end in what is most probably a solvent peak.

map (Figure 6a). Both of the pyrrole methyl groups, marked b and d, extend in a well-defined direction from the flat regions of density that are interpreted as being the pyrrole rings. The spacing of density protruberances corresponding to features b, c, and d is also supportive of the guanidinium--up orientation shown in Figure 6a. Hence, only this orientation has been reported. Unlike an off-center class II model, end-for-end reversal of a centered class I model would leave unchanged the observed pattern of three bifurcated hydrogen bonds to base pair edges.

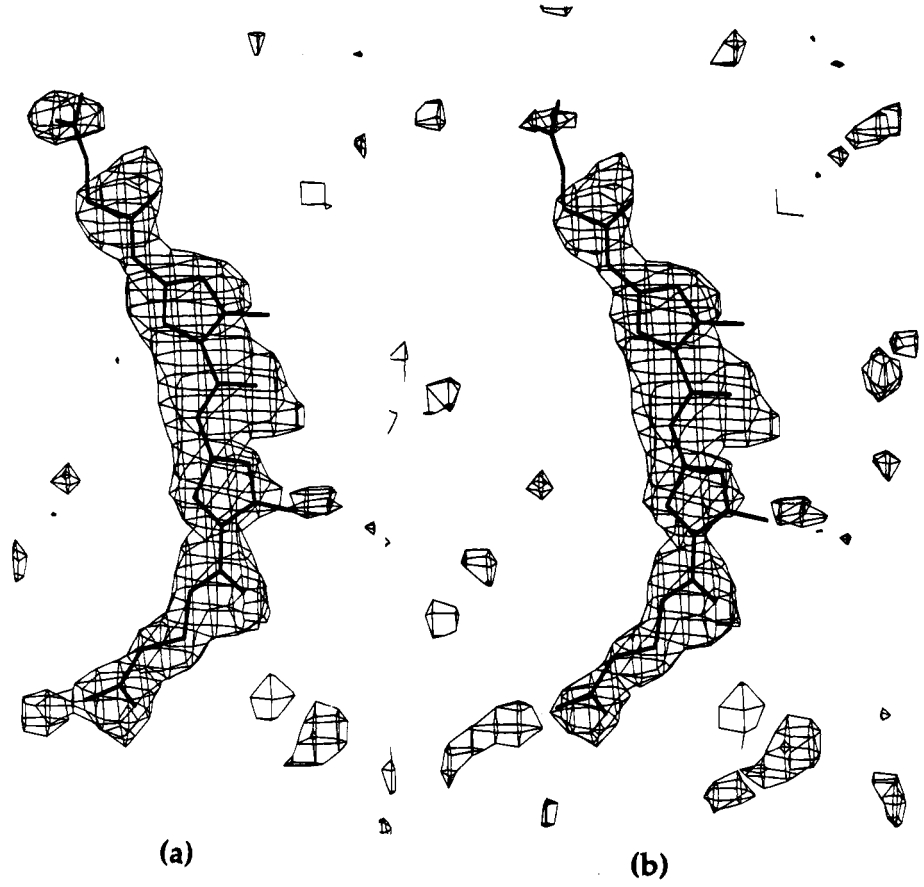


FIGURE 9: Effect of refinement on omit electron density maps: (a) omit density from the final model refined in the class II position, identical with Figure 8b; (b) omit map after the removal of drug atoms from the model and five cycles of NUCLSQ refinement. Both maps are contoured at 2.0$\sigma$. Refinement after removal of the drug produces virtually no change in the omit map and therefore does not remove any bias that had been present.

The idea that a final omit map derived from a fully refined structure should contain a bias toward whatever assumptions had been made during refinement is not a new idea, nor original to us. In an attempt to remove this bias, it is sometimes the practice to delete the drug image from the final refined structure and then give the DNA a few cycles of additional refinement without the drug. However, this does not solve the problem: the "memory" of an assumed drug position remains even after the physical removal of its coordinates from the final structure. Figure 9 compares omit map density for the class II positioning of netropsin: (a) with simple removal of drug coordinates and (b) with five cycles of subsequent positional and temperature factor refinement. The two maps are virtually identical; the added cycles of refinement failed to produce their intended result. If a final omit map is indeed improperly biased, a few cycles of last stage refinement will not necessarily remove that bias.

As was mentioned earlier, the fundamental problem in these class I vs class II refinements may be one of the quantity of experimental data in each analysis. In Table 1, the nominal resolution in angstroms, or the outer limit of the data sphere, is the same in five of the seven structures reported. But the actual number of reflections measured at a significant intensity varies considerably, and hence the *effective* resolution is different. As a parallel example, the first X-ray crystal structure analysis of a B-DNA dodecamer showed an ordered, zig-zag spine of hydration running down the narrow portion of the minor groove. In subsequent dodecamer structures with narrow minor groove regions, a comparably well-ordered spine of hydration was not always observed, and the generality of the spine of hydration was called into question. But then a comparison of several structure analyses, like that in the present paper, showed that an ordered spine in a narrow region of the minor groove was always observed, *provided that the resolution of the X-ray analysis was sufficiently high.* The problem had been not the variability in minor groove hydration but the information content of a particular structure determination. By the same token we can predict that netropsin or distamycin, when bound singly to B-DNA (i.e., not in the side-by-side binding mode), will be found to sit with its pyrrole rings opposed to base pairs, and with a bifurcated pattern of hydrogen bonding, provided only that the quantity of data (the effective resolution) is sufficient. This item of structural information about the geometry of binding of the netropsin/distamycin family is critical in the design of new monomeric and bis-linked variants for the recognition of specific base sequences.

## REFERENCES

Brunger, A. T. (1993) *Acta Crystallogr. D49*, 24−36.
Burckhardt, G., Luck, G., Zimmer, C., Storl, J., Krowicki, K., & Lown, J. W. (1989) *Biochim. Biophys. Acta 1009*, 11−18.
Chen, X., Ramakrishnan, B., Rao, S. T., & Sundaralingam, M. (1994) *Struct. Biol. 1*, 169−175.
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., & Kopka, M. L. (1985) *J. Biomol. Struct. Dyn. 3*, 423−431.
Drew, H. R., & Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535−556.
Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686−14707.
Geierstanger, B. H., Dwyer, T. J., Bathini, Y., Lown, J. W., & Wemmer, D. E. (1993) *J. Am. Chem. Soc. 115*, 4474−4482.
Geierstanger, B. H., Jacobsen, J. P., Mrksich, M., Dervan, P. B., & Wemmer, D. E. (1994a) *Biochemistry 33*, 3055−3062.
Geierstanger, B. H., Mrksich, M., Dervan, P. B., & Wemmer, D. E. (1994b) *Science 266*, 646−650.
Hahn, M., & Heinemann, U. (1993) *Acta Crystallogr. D49*, 468−477.
Heinemann, U., & Alings, C. (1991) *EMBO J. 10*, 35−43.
Hendrickson, W. A., & Konnert, J. H. (1980) in *Computing in Crystallography* (Diamond, R., Ramaseshan, S., & Venkatesan, K., Eds.) pp 13.01−13.23, Indian Academy of Sciences, Bangalore, India.
Kissinger, K., Krowicki, K., Dabrowiak. J. C., & Lown, J. W. (1987) *Biochemistry 26*, 5590−5595.
Kopka, M. L., Fratini, A. V., Drew, H. R., & Dickerson, R. E. (1983) *J. Mol. Biol. 163*, 129−146.
Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985a) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1376−1380.
Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985b) *J. Mol. Biol. 183*, 553−563.
Kopka, M. L., Pjura, P. E., Yoon, H.-C., Goodsell, D. S., & Dickerson, R. E. (1985c) in *Structure and Motion: Membranes, Nucleic Acids and Proteins* (Clementi, E., Corongiu, G., Sarma, M. H., & Sarma, R., Eds.) pp 461−483, Adenine Press, New York.
Lown, J. W., Krowicki, K., Bhat, U. G., Skorobogaty, A., Ward, B., & Dabrowiak, J. C. (1986a) *Biochemistry 25*, 7408−7416.
Lown, J. W., Krowicki, K., Balzarini, J., & De Clerq, E. (1986b) *J. Med. Chem. 29*, 1210−1214.
Lown, J. W., Krowicki, K., Balzarini, J., Newman, R. A., & De Clerq, E. (1989) *J. Med. Chem. 32*, 2368−2375.
Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. G., Wemmer, D. E., & Dervan, P. B. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7586−7590.
Patel, D. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6424−6428.
Pelton, J. G., & Wemmer, D. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5723−5727.
Pelton, J. G., & Wemmer, D. E. (1990) *J. Biomol. Struct. Dyn. 8*, 81−97.
Pullman, B. (1983) *J. Biomol. Struct. Dyn. 1*, 773−794.
Pullman, B., & Pullman, A., (1981) *Stud. Biophys. 86*, 95−102.
Sriram, M., van der Marel, G. A., Roelen, H. L. P. F., van Boom, J. H., & Wang, A. H.-J. (1992) *Biochemistry 31*, 11823−11834.
Tabernero, L., Verdaguer, N., Coll, M., Fita, I., van der Marel, G. A., van Boom, J. H., Rich, A., & Aymami, J. (1993) *Biochemistry 32*, 8403−8410.
Van Dyke, M. W., Hertzberg, R. P., & Dervan, P. B. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5470−5474.
Walker, W. L., Kopka, M. L., Filipowsky, M. E., Dickerson, R. E., & Goodsell, D. S. (1995) *Biopolymers* (in press).
Wang, W., & Lown, J. W. (1992) *J. Med. Chem. 35*, 2890−21897.
Youngquist, R. S., & Dervan, P. B. (1987) *J. Am. Chem. Soc. 109*, 7564−7566.
Zakrzewska, K., Lavery, R., & Pullman, B. (1987) *J. Biomol. Struct. Dyn. 4*, 833−843.

BI942892J