Hare, D. R., Shapiro, L., & Patel, D. J. (1986) *Biochemistry 25*, 7445-7456.
Haran, T. E., Shakked, Z., Wang, A. H.-J., & Rich, A. (1987) *J. Biomol. Struct. Dyn. 5*, 199-217.
Ho, P. S., Frederic, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., & Rich, A. (1985) *EMBO J. 4*, 3617-3623.
Hunter, W. N., Brown, T., Anand, N. N., & Kennard, O. (1986) *Nature (London) 320*, 552-555.
Kalnik, M. W., Kouchakdjian, M., Li, B. F. L., Swann, P. F., & Patel, D. J. (1988a) *Biochemistry 27*, 100-108.
Kalnik, M. W., Kouchakdjian, M., Li, B. F. L., Swann, P. F., & Patel, D. J. (1988b) *Biochemistry 27*, 108-115.
Kennard, O. (1985) *J. Biomol. Struct. Dyn. 3*, 205-226.
Kubasek, W. L., Wang, Y., Thomas, G. A., Patapoff, T. W., Schoenwaelder, K.-H., van der Sande, J. H., & Peticolas, W. L. (1986) *Biochemistry 25*, 7440-7445.
Letellier, R., Ghomi, M., & Taillandier, E. (1987) *Eur. Biophys. J. 14*, 243-252.
Lord, R. C., & Thomas, G. J., Jr. (1967) *Spectrochim. Acta, Part A 23A*, 2551-2591.
Nishimura, Y., Tsuboi, M., Sato, T., & Aoki, K. (1986) *J. Mol. Struct. 146*, 123-153.
Prescott, B., Steinmetz, W., & Thomas, G. J., Jr. (1984) *Biopolymers 23*, 235-256.
Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., & Dickerson, R. E. (1987) *Science (Washington, D.C.) 238*, 498-504.
Quignard, E., Fazakerely, G. V., van der Marel, G., van Boom, J. H., & Guschlbauer, W. (1987) *Nucleic Acids Res. 15*, 3397-3409.
Thamann, T., Lord, R. C., Wang, A. H. J., & Rich, A. (1981) *Nucleic Acids Res. 9*, 5443-5457.
Thomas, G. J., Jr. (1986) *Adv. Spectrosc. 13*, 233-309.
Thomas, G. J., Jr., & Agard, D. A. (1984) *Biophys. J. 46*, 763-768.
Thomas, G. J., Jr., & Benevides, J. M. (1985) *Biopolymers 24*, 1101-1105.
Thomas, G. J., Jr., & Wang, A. H.-J. (1988) in *Nucleic Acids and Molecular Biology* (Eckstein, F., & Lilley, D. M. J., Eds.) Vol. 2, Springer-Verlag, Berlin.
Thomas, G. J., Jr., Prescott, B., & Benevides, J. M. (1986) *Biomol. Stereodyn. 4*, 227-254.
Thomas, G. J., Jr., Prescott, B., Opella, S. J., & Day, L. A. (1988) *Biochemistry 27*, 4350-4357.
van der Marel, G. A., van Boeckel, C. A. A., Wille, G., & van Boom, J. H. (1981) *Tetrahedron Lett.* 3887.
Wang, A. H.-J., Fujii, S., van Boom, J. H., van der Marel, G. A., van Boeckel, S. A., & Rich, A. (1982) *Nature (London) 299*, 601-604.
Wartell, R. M., & Harrell, J. T. (1986) *Biochemistry 25*, 2664-2671.

Molecular Structure of the Netropsin-d(CGCGATATCGCG) Complex: DNA Conformation in an Alternating AT Segment†

Miquel Coll,‡ Juan Aymami,‡ Gijs A. van der Marel,§ Jacques H. van Boom,§ Alexander Rich,‡ and Andrew H.-J. Wang*,‡

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

Received May 23, 1988; Revised Manuscript Received August 18, 1988

ABSTRACT: The molecular structure of the complex between a minor groove binding drug (netropsin) and the DNA dodecamer d(CGCGATATCGCG) has been solved and refined by single-crystal X-ray diffraction analysis to a final R factor of 20.0% to 2.4-Å resolution. The crystal is similar to that of the other related dodecamers with unit cell dimensions of a = 25.48 Å, b = 41.26 Å, and c = 66.88 Å in the space group $P2_12_12_1$. In the complex, netropsin binds to the central ATAT tetranucleotide segment in the narrow minor groove of the dodecamer B-DNA double helix as expected. However, in the structural refinement the drug is found to fit the electron density in two orientations equally well, suggesting the disordered model. This agrees with the results from solution studies (chemical footprinting and NMR) of the interactions between minor groove binding drugs (e.g., netropsin and distamycin A) and DNA. The stabilizing forces between drug and DNA are provided by a combination of ionic, van der Waals, and hydrogen-bonding interactions. No bifurcated hydrogen bond is found between netropsin and DNA in this complex due to the unique dispositions of the hydrogen-bond acceptors (N3 of adenine and O2 of thymine) on the floor of the DNA minor groove. Two of the four AT base pairs in the ATAT stretch have low propeller twist angles, even though the DNA has a narrow minor groove. Alternating helical twist angles are observed in the ATAT stretch with lower twist in the ApT steps than in the TpA step.

The conformational diversity of DNA has been under intense study in recent years with a variety of techniques. One of the intriguing questions related to DNA conformation is its relation to nucleotide base sequence and different environmental factors that affect the equilibrium between alternative conformers. For example, the left-handed Z-DNA double helix is favored by alternating CG sequence (Wang et al., 1979). The interconversion between Z-DNA and the righ-handed B-DNA is influenced by metal ions, ionic strength, supercoiling, and Z-DNA binding proteins as reviewed elsewhere (Rich et al., 1984). More recently, another conformational

† This work was supported by grants from NIH, NSF, NASA, and ONR to A.R. and from NSF (DMB 8612286) and Burroughs-Wellcome to A.H.-J.W. M.C. is a Fulbright Fellow. J.A. is a NATO Fellow.

* Address correspondence to this author at the Department of Physiology and Biophysics, University of Illinois at Urbana-Champaign, Urbana, IL 61801.
‡ Massachusetts Institute of Technology.
§ Leiden State University.

0006-2960/89/0428-0310$01.50/0 © 1989 American Chemical Society

state of DNA, namely, bent DNA, has received a great deal of attention due to its potential role in gene regulation and nucleosome phasing (Bossi & Smith, 1984; Widom, 1985). Bent DNA is apparently sequence directed and is strongly favored by oligo(dA)$_n$ (n = 4–6) stretches when they are appropriately and repeatedly spaced along DNA (Diekmann, 1987). However, some subtle idiosyncrasies exist. For example, DNA made of repeats of a (GAAAATTTTC) motif exhibits a strong bent DNA tendency but not that built of (GTTTTAAAAT) (Burkhoff & Tullius, 1988). The reason for this may be related to the dinucleotide steps in the center, TpA vs ApT. We have addressed this question by attempting to crystallize a number of DNA dodecamers with sequences closely related to d(CGCGAATTCGCG), whose structure has been analyzed in great detail (Wing et al., 1980; Drew & Dickerson, 1981). In particular, we were interested in determining the structure of molecules with a TpA step in the central AT segment in the hope that it might reveal a unique property associated with this step.

It was found that many of those DNA molecules could not be crystallized alone. We sought to create a DNA–ligand complex which might crystallize, but we also tried not to perturb the basic architecture of the molecule by using ligands which interact with DNA double helix through noncovalent molecular forces. Another consideration was the binding specificity. For example, a simple intercalator molecule such as ethidium binds to DNA by inserting its planar aromatic moiety between two adjacent base pairs. This type of interaction in general affords little nucleotide sequence binding specificity, as no functional groups are found in ethidium to provide specific interactions such as hydrogen bonds to the DNA double helix. As the complexity of DNA binding compounds increases, so does their binding specificity. For example, quinoxaline antibiotics (triostine A and echyinomycin) are bis-intercalating molecules which have a strong affinity to tetranucleotide sequences with a CpG step in the middle of the sequences. The specificity for CpG sequence is provided by the hydrogen bonds between the two alanine residues in the octadepsipeptide backbone of the quinoxaline antibiotics and the NH_2 amino group of guanine in the minor groove of the highly distorted right-handed double helix (Wang et al., 1984).

Netropsin and distamycin (Figure 1) are DNA binding compounds which have a binding preference to stretches of AT-rich sequences (Zimmer & Wahnert, 1986). The molecular basis of the interactions between them have recently been under intensive study by a number of methods including chemical footprinting experiments (Taylor et al., 1984; Dervan, 1987), NMR (Patel, 1979; Klevit et al., 1986; Leupin et al., 1986), and single-crystal X-ray diffraction studies (Kopka et al., 1985; Coll et al., 1987). The picture emerging from these studies is that this class of compounds binds in the narrow minor groove of the B-DNA double helix by use of a combination of interactions including hydrogen bonds, ionic charge attractions, and van der Waals interactions. These elongated molecules contain aromatic rings (*N*-methylpyrrole) linked together by amide groups which donate their NH hydrogen atoms to form bifurcated hydrogen bonds to acceptor atoms (N3 of A and O2 of T) in the AT base pairs at the floor of the minor groove. The pyrrole rings fit snugly in the narrow minor groove which is characteristic of base pairs in an AT stretch. There are important van der Waals interactions to the sugar–phosphate backbone at the side wall of the groove. The AT binding preference arises from the clashes that would occur between the NH_2 group of guanine on C2 and the hydrogen atom HC2 of the pyrrole ring.

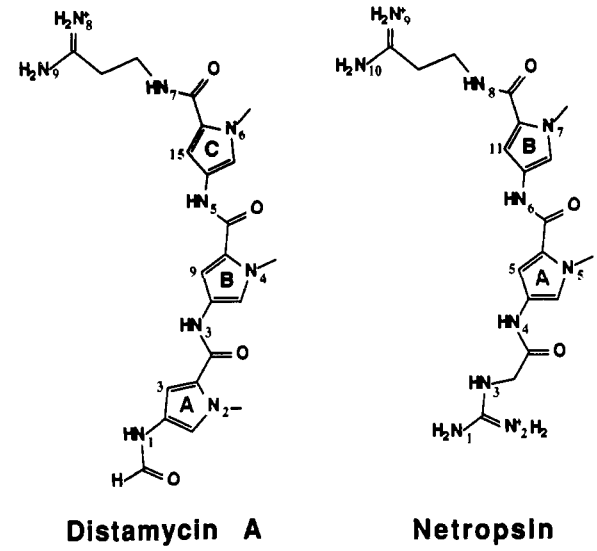

FIGURE 1: Chemical formula of netropsin and distamycin A. Note that the molecules are asymmetric without a 2-fold rotation axis. The lines on the pyrrole ring nitrogen atoms represent methyl groups.

When we survey the results from the chemical footprinting and temperature melting experiments of netropsin and distamycin on DNA restriction fragments and synthetic polymers, it is evident that there exists a gradation of binding affinity of these drugs to various AT-rich sequences (Zimmer & Wahnert, 1986). For example, netropsin binds better to an oligo[d(A)]·oligo[d(T)] stretch than to an oligo(dA-dT) stretch. On the basis of the structural information available so far, it is not immediately apparent why such sequence binding microheterogeneity exists.

We have approached these problems by solving the crystal structure of the complex between netropsin and a DNA dodecamer, d(CGCGATATCGCG). In this structure, it was found that the netropsin molecule binds in the minor groove at the central ATAT region as expected. Furthermore, the drug seems to bind in two different orientations equally well in the crystal lattice by use of single hydrogen bonds rather than bifurcating bonds. Although the minor groove is narrow in the ATAT stretch, two of the four AT base pairs in this structure have very low propeller twists in contrast to those found in other dodecamers containing nonalternating AT sequences.

Materials and Methods

The DNA dodecamer d(CGCGATATCGCG) was synthesized by an improved phosphotriester method and purified by Sephadex G-50 column chromatography (van der Marel et al., 1982). The purity of the final product was checked by HPLC and judged to be greater than 95%. Netropsin was a generous gift from Drs. H. Fritzsche and H. Thrum of ZIMET Jena, Academy of Sciences, GDR. The complex of netropsin and the dodecamer was crystallized from a solution containing 1 mM DNA dodecamer (single-strand concentration), 40 mM sodium cacodylate buffer at pH 6.5, 7.0 mM magnesium chloride, 2.0 mM netropsin, 4.5 mM spermine tetrachloride, and 5% 2-methyl-2,4-pentanediol (2-MPD), equilibrated against 50% 2-MPD by the vapor diffusion technique. Crystals could be grown either at room temperature or in the cold room (10 °C). The crystals obtained at low temperature were of significantly better quality, yet they were still too small for X-ray diffraction data collection. We improved the quality of these crystals by seeding a small clean crystal in a crys-

tallization solution. Some crystals with the shape of rectangular thin plates having a size up to 0.2 × 0.6 × 1.1 mm could be obtained in this manner. Attempts were made to crystallize the dodecamer alone under conditions similar to those used for other DNA oligonucleotides, but they were unsuccessful.

A crystal was mounted in the thin-wall capillary sealed with a droplet of the crystallization mother liquor for data collection. The crystal was found to be in the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of a = 25.48 Å, b = 41.26 Å, and c = 66.88 Å. The diffraction data were collected on a Nicolet P3 diffractometer using an ω-scan mode at 10 °C to 2.4-Å resolution. A total of 5696 reflections from two octants of the reciprocal lattice were collected. After the symmetry-related pairs of the reflections were averaged, 1848 were considered to be observable at the 2.0 $\sigma(F)$ level above background. Lp, empirical absorption, and decay corrections were applied.

The overall unit cell dimensions and diffraction pattern resemble those of the other closely related dodecamers, suggesting a very similar crystal lattice. A model of the dodecamer d(CGCGATATCGCG) was generated by replacing the atomic coordinates of the central two AT base pairs of the d(CGCGAATTCGCG) dodecamer (Wing et al., 1980) from an ApT step to a TpA step. This model was refined according to the Konnert–Hendrickson refinement procedure (Hendrickson & Konnert, 1979) to an R factor of 28%. At this stage, solvent molecules located from the difference Fourier map were gradually included in the refinement. However, we refrained from adding any solvent molecules in the minor groove region of the double helix to avoid any interference in the interpretation of the difference Fourier map there. After the inclusion of 29 water molecules in addition to the dodecamer duplex, the R factor was 24%, and the difference Fourier map clearly showed distinct residual electron density in the minor groove at the ATAT region. By close examination of the map, it was evident that there were two lobes of electron density connected by weaker density, each about the size of half of a netropsin molecule. We could fit the netropsin molecule to the residual density in the minor groove of the DNA double helix in two orientations using the Evans & Sutherland PS-390 graphic system with the program FRODO (Jones, 1978). Although in one of the two orientations one of the pyrrole rings fits better than the other orientation, we could not decide which one is correct. This is due in large part to the quasi 2-fold symmetry of the netropsin molecule. Consequently, both models were refined independently with identical parameters. After several cycles of constrained refinement, the final R factors were 20.1% and 20.2%, respectively, from the 1848 reflections between 20.0- and 2.4-Å resolution. Similar temperature factors for the netropsin molecule were obtained from either refinement. As a result, we could not unambiguously determine which of the two orientations is better. This is clearly illustrated in Figure 2 where the netropsin molecules in either orientation are seen to fit the final "omit" Fourier map very well. In fact, both could exist simultaneously in the crystal (i.e., a disordered structure), which is consistent with the results from solution studies including chemical footprinting and NMR experiments as will be discussed later. We did not attempt to refine a disordered model including partial occupancies for the two orientations by the Konnert–Hendrickson method as it would increase substantially the number of parameters to be refined. In view of the already low ratio of observations (reflections) to parameters, the refinement incorporating disordered netropsins deemed to be unwarranted.

In the final structure, the asymmetric unit of the crystal lattice included one complete dodecamer duplex, one netropsin (with two possible orientations), and 60 water molecules. The electron density for the netropsin molecule in the final Fourier map had a continuous envelope covering the whole length of the molecule (Figure 2). No magnesium or spermine ions could be identified unambiguously. A reasonably strong constraint was imposed on the geometrical parameters in the DNA and drug molecule. The final root mean square deviation of bond lengths from the ideal values was 0.024 Å. Care was taken to avoid any bad van der Waals contacts between the drug and DNA molecules. The final atomic coordinates of the complex have been deposited in the Brookhaven Protein Databank.

Results and Discussion

Structure of the Netropsin–DNA Complex. The overall structure of the netropsin–DNA complex is shown with a van der Waals diagram looking into the minor groove of the right-handed B-DNA double helix in Figure 3. This figure is shown with the netropsin in one of the two orientations. The structure of the complex involving the netropsin in the other orientation is very similar as a whole, although they differ in some details as will be described later. It can be seen that the elongated netropsin molecule fits tightly in the narrow minor groove at the central ATAT region covering slightly over four base pairs. The drug molecule is sandwiched by the side wall of the minor groove which is made of the two antiparallel sugar–phosphate backbones. Many atoms from DNA have close van der Waals contacts to both faces of the flat netropsin molecule.

A more detailed depiction of the structure is illustrated by the two stereoscopic skeletal views which have the intermolecular hydrogen bonds shown as thin lines. Figure 4A is a view with the same orientation as in the van der Waals diagram of Figure 3. Here the netropsin molecule has a dihedral angle of 23° between the two pyrrole rings. In the other orientation of this angle is 26°. These values are similar to the corresponding values found in the netropsin–d(CGCGAATTCGCG) complex (Kopka et al., 1985) and in the crystal structure of the netropsin molecule alone (Berman et al., 1979).

There are four hydrogen bonds between the drug and the nucleophilic oxygen and nitrogen atoms of DNA at the floor of the minor groove. In this case, no bifurcated hydrogen bonds from the amide nitrogen NH groups to DNA are found as in other structures; instead, only single hydrogen bonds are seen. This observation may be associated with the unique disposition of N3 and O2 atoms of ATAT sequence at the floor of the minor groove as will be discussed later.

A different view of the complex is displayed in Figure 4B where the molecule is rotated 90° relative to that in Figure 4A. Here it is clear how close the crescent-shaped drug hugs the DNA double helix with the pyrrole HC5 and HC11 hydrogens approaching the HC2 hydrogens of adenine bases. For example, the HC2 atom of pyrrole A is 2.6 Å away from the HC2 atom of adenine A7, and the HC11 atom of pyrrole B is 2.6 Å from the HC2 atom of adenine A19 (all hydrogen atom positions were calculated at their theoretical positions). These close van der Waals contacts are the principal interactions that determine the binding preference for AT base pairs. As suggested previously, the N2 amino group of a guanine base in the minor groove presents a major hindrance for the entry of the drugs into the minor groove, thereby providing the discrimination for a more favored binding toward AT base pairs (Kopka et al., 1985). The netropsin in this

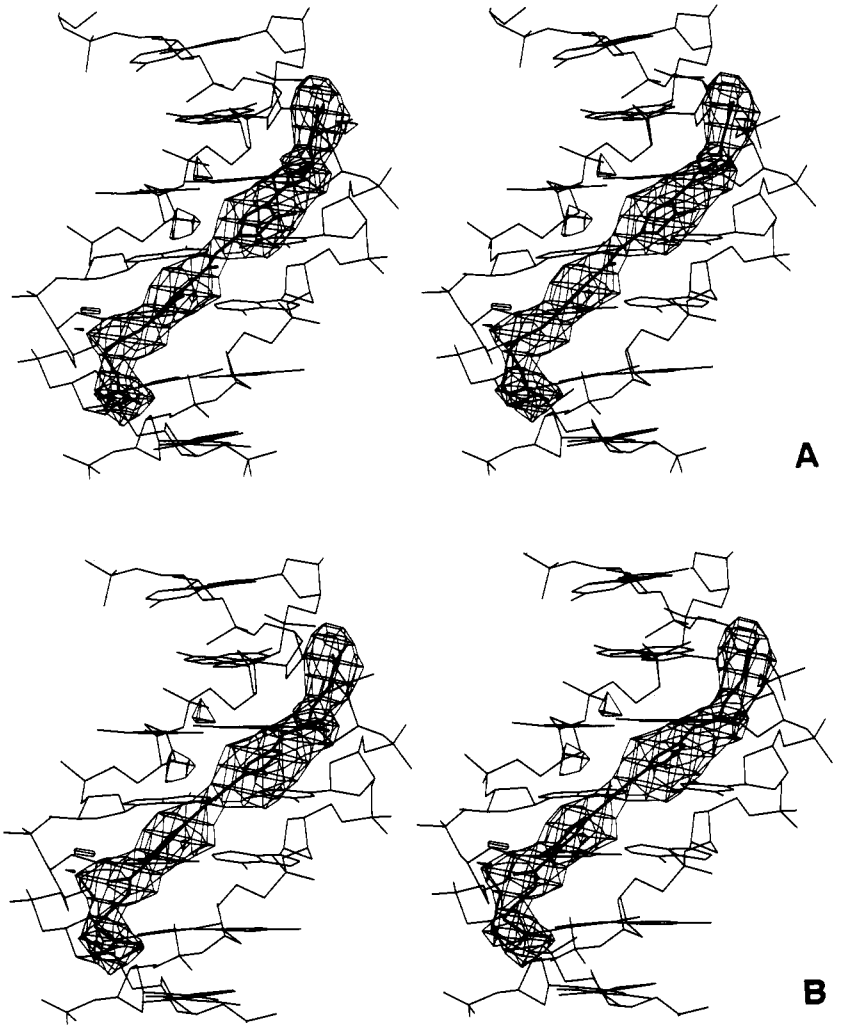

FIGURE 2: Final omit difference Fourier electron density map associated with the netropsin molecule in the crystal of the netropsin–d(CGCGATATCGCG) complex displayed by the FRODO program on an Evans & Sutherland PS-390 graphic system. Phases were calculated on the basis of all the atoms in the structure except netropsin, after three cycles of refinement to remove the possible bias contributed from netropsin. Two major lobes of electron densities connected by somewhat weaker density can be seen in which the netropsin molecule may be fitted in two opposite orientations (top and bottom panels). The electron density level shown is 0.32 e/Å^2.

complex completely covers the ATAT segment with its linear guanidinium and amidinium ends touching the two GC outer base pairs (G4-C21 and C9-G16). Again, this agrees well with the previous observations made by the footprinting experiments (Taylor et al., 1984).

DNA Conformation. The overall conformation of this DNA dodecamer duplex seen in Figures 3 and 4 is not unlike those found in other closely related dodecamer sequences such as d(CGCGAATTCGCG) (Wing et al., 1980; Kopka et al., 1985), d(CGCAAATTTGCG) (Coll et al., 1987), and d(CGCA$_6$GCG)·d(CGCT$_6$GCG) (Nelson et al., 1987). The backbone torsion angles in this dodecamer are listed in Table I. All of the averaged torsion angles are very similar to those of the canonical d(CGCGAATTCGCG) double helix, despite the fact that each individual value has quite large standard deviations ranging from 35° to 79°. This is probably related to the way in which the constrained refinement procedure is carried out in a medium-resolution structure. The dodecamer molecules mentioned above are all in the righ-handed B-DNA family, but with a distinct narrow minor groove at the central AT region. This is easily seen by comparing the minor groove width of all the known dodecamer duplexes, diagrammatically shown in Figure 5. A consistent trend is obvious in these composite curves where they all reach the lowest value (~4 Å) near the 8/21 residue, except for the AATT dodecamer without a bound drug. This is also clearly illustrated by the radial projection of the drug–DNA complex in Figure 6 where the distances between two O4′ atoms across the minor groove are listed. Small numbers ranging from 6.3 to 7.9 Å are seen near the drug molecule, whereas the corresponding distances further removed from the drug are larger (8.3–11.6 Å). Some atoms in the sugar rings are in fact in close van der Waals contacts with the netropsin molecule. For example, the O4′ of the T20 residue is only 3.2 Å away from the plane of the netropsin pyrrole ring, close to the sum of the van der Waals radius of oxygen (1.4 Å) and the half-thickness of an aromatic pyrrole ring. As suggested previously, this sandwiching of the planar drug by the sugar–phosphate backbones from the side walls of the minor groove of the B-DNA double helix strongly stabilizes the complex and may be considered as a pseudointercalation (Teng et al., 1988).

It has been suggested that the narrow minor groove is associated with the high propeller twist of AT base pairs which can be stabilized by the spine of hydration or drug molecules (Wing et al., 1980; Kopka et al., 1985). Therefore, it is somewhat surprising to find that the mean propeller twist angle for the central ATAT base pairs is only 11°, a low value compared to those found in other dodecamers and about the same as that found in the GC stretch. Table II lists all the propeller twist angles of four different dodecamers. The

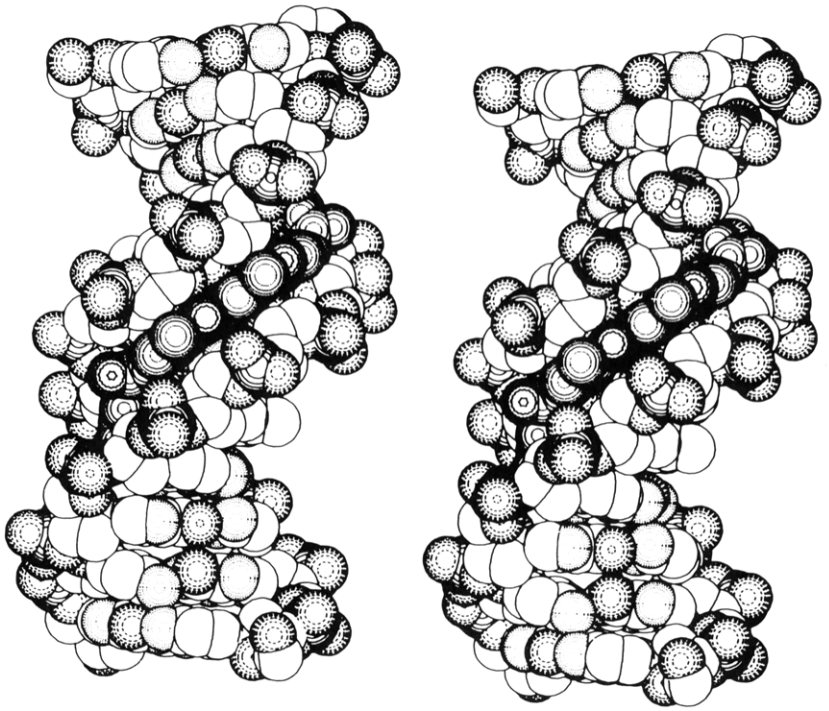

FIGURE 3: Stereoscopic van der Waals diagram of the complex looking into the minor groove along the direction of the molecular diad axis. The elongated flat netropsin fits snuggly in the narrow minor groove with many van der Waals contacts to the side walls of the groove. Only one of the 2-fold disordered orientations is shown.

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

residue	α	β	γ	δ	ϵ	ζ	χ	P^b
C1			148	89	158	−28	−160	76
G2	−98	−97	7	171	90	−26	−94	−125
C3	−154	−163	128	81	−162	−72	−170	74
G4	−7	−135	−26	178	111	−27	−83	−124
A5	−98	−102	3	149	67	18	−88	−133
T6	−121	−127	41	170	150	−110	−65	−159
A7	−70	−167	58	167	−125	−140	−86	168
T8	−73	172	32	144	−74	156	−92	158
C9	−13	−163	−70	−171	−129	−113	−108	194
G10	−125	160	77	148	−32	−173	−89	173
C11	−136	83	105	82	−122	−59	−175	79
G12	−149	107	139	134			−153	−96
C13			48	145	−150	−92	−130	178
G14	−141	144	132	136	−141	−88	−129	169
C15	−52	148	51	116	−118	−162	−136	114
G16	−69	111	92	108	−170	−80	−140	102
A17	−106	167	97	137	−180	−64	−132	162
T18	−58	−108	9	174	−179	−118	−75	−152
A19	−96	138	122	90	−168	−40	−162	81
T20	−74	−154	29	166	−161	−129	−80	193
C21	−70	−162	40	140	−167	−98	−101	162
G22	−71	−171	54	165	−148	−149	−86	179
C23	−63	134	81	74	−109	−125	−166	71
G24	−34	97	87	90			−163	87
av	−85	176	62	135	−129	−95	−119	159
SD	40	51	54	35	79	54	35	58

[a] The backbone torsion angles are defined as O3′—$\overset{\alpha}{P}$—O5′—$\overset{\beta}{C}$5′—$\overset{\gamma}{C}$4′—$\overset{\delta}{C}$3′—$\overset{\epsilon}{O}$3′—$\overset{\zeta}{P}$—O5′. [b] Sugar pseudorotation angle.

present dodecamer has the lowest average value both in the overall value throughout the whole molecule and the in the central AT stretch. For example, the mean propeller twist angle of the AT stretch is 20° and 17° for d-(CGCAAATTTGCG) (Coll et al., 1987) and d-(CGCGAATTCGCG) (Dickerson & Drew, 1981), respectively. In the former case the extremely high propeller twist angles are associated, in some steps, with a novel bifurcated hydrogen bond in the major groove between the N6 amino group of an adenine and the carbonyl O4 groups of two adjacent thymines of the opposite strand. Using this nucleotide structural motif incorporating bifurcated hydrogen bonds in the major groove, we were able to construct a modified B-DNA model of poly(dA)·poly(dT), which has a very narrow minor groove (Coll et al., 1987). Our current structure here, with its low propeller twist angle in the AT region, specifically in the first and last AT base pairs of the stretch, which is shown in detail in Figure 7A, has lead us to ponder what is the principal factor responsible for the narrow minor groove.

There is a possibility that the low propeller twist angle of the ATAT base pairs is the consequence of an average image of the two disordered structures in the crystal lattice. However,

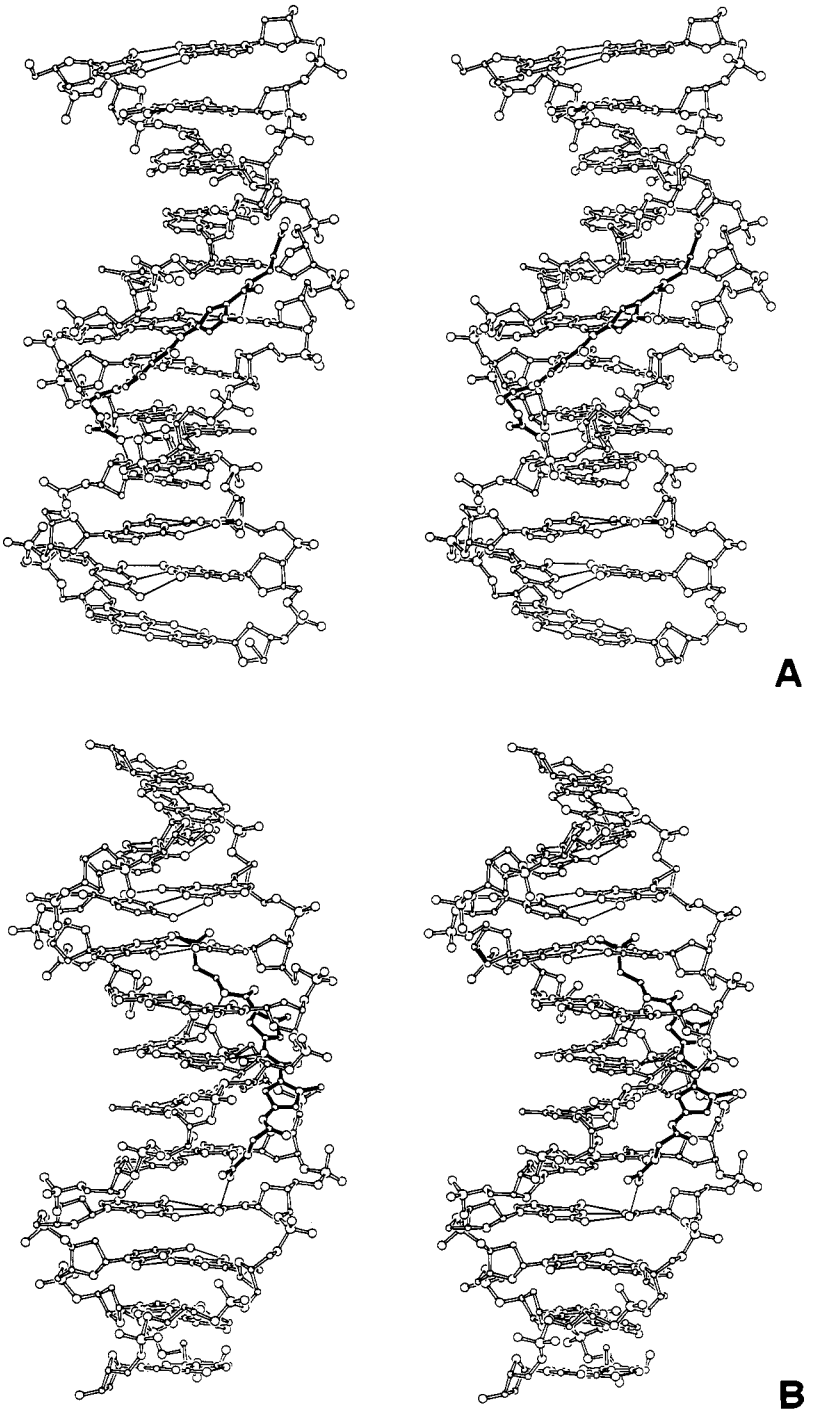

FIGURE 4: Stereoscopic skeletal diagrams of the complex with DNA in open bonds and netropsin in filled bonds. The elongated netropsin molecule lies in the narrow minor groove of the double helix with three hydrogen bonds to bases of the DNA. The two views in (A) and (B) are 90° apart around the helix axis of DNA.

this seems unlikely for two reasons. First, the electron density map associated with these four AT base pairs reveals no evidence of diffused and weakened pattern as shown in Figure 7B. The bases are completely enclosed in the well-resolved and flat-shaped envelopes even at this resolution (2.4 Å). The temperature factors for these bases are normal (with an averaged thermal factor of B = 12.3 Å^2) compared with other bases in the molecule averaged for C and G bases with B = 14.3 Å^2). Second, the sense (or direction) of the propeller twist is usually positive [for the definition of the propeller twist, see Dickerson et al. (1982)]. Therefore, when a self-complementary duplex molecule is disordered in the crystal lattice due to the molecular 2-fold symmetry, the 2-fold related base pairs (e.g., A5-T20 vs A17-T8) will have the same sign of the propeller twist regardless which orientation the molecule is viewed. Therefore, it is reasonable that the propeller twist of a base pair in the molecule is not likely to be averaged to a lower value due to two large propeller twist angles of opposite sign from the 2-fold related base pairs. From these arguments, we conclude that the low AT propeller twist in this dodecamer is not an artifact of a disordered structure.

Furthermore, a continuous high propeller twist is unlikely to occur in an alternating AT stretch because of the repulsion of the N6 atoms of consecutive adenines from opposite strands in the major groove of DNA. This is evident in the stacking diagram of the three central consecutive AT steps (A5/T6,

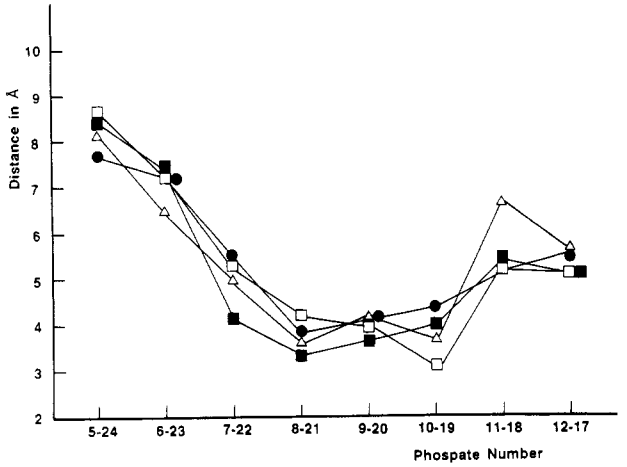

FIGURE 5: A plot of the phosphate to phosphate distances (P–P distance minus 5.8 Å) across the minor groove with the shortest distance near P8/P21 and P9/P20. This is consistent with the results shown in Figure 6. Four different DNA dodecamers are shown here: (●) d(CGCGATATCGCG) + netropsin (this work); (△) d-(CGCAAATTTCGCG) + distamycin (Coll et al., 1987); (□) d-(CGCGAATTCGCG) (Wing et al., 1980); (■) d(CGCA$_6$GCG)·d-(CGCT$_6$GCG) (Nelson et al., 1987).

Table II: Propeller Twist (deg)[a]

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

[a] The boxes include the AT stretches. [b] d(CGCGATATCGCG–netropsin, this work. [c] d(CGCGAATT^{Br}CGCG)–netropsin, Kopka et al. (1985). [d] d(CGCGAATTCGCG), Drew et al. (1981). [e] d-(CGCAAATTTGCG)–distamycin, Coll et al. (1987). [f] d-(CGCAAAAAAGCG)·d(CGCTTTTTTGCG), Nelson et al. (1987).

T6/A7, and A7/T8) shown in Figure 8. It is interesting to note that there is very little base overlap (both inter- and intrastrand) in the middle T6pA7 step as might be expected for the pyrimidine(3′–5′)purine step in B-DNA. Somewhat surprising is that these two base pairs both have relatively high propeller twist (16°, see Table II), resulting in a very close A7C2 to A19C2 distance (2.7 Å) in the minor groove. We did not attempt to remove this van der Waals close contact by forcing these two atoms apart in the refinement. Moreover, it is likely that 16° is very close to the limit for the propeller twist angle associated with a TpA step and it might be stabilized by the presence of netropsin.

The two outer ApT steps are not the same. In the A5pT6 step (Figure 8A), a somewhat diminished base stacking is seen, compared to that of the A7pT8 step (Figure 8C). On the right side of the figure, the thymine T6 base is shifted toward the lower right corner of base A5 by about 1 Å (relative to T8 in Figure 8C), resulting a smaller ring–ring overlap. An opposite movement is seen on the left side where T20 is shifted by about 1.5 Å toward the lower left corner. This nonequivalent stacking pattern at these two steps has also been similarly observed in other related dodecamer molecules previously (Drew & Dickerson, 1981).

The two outer base pairs (A5-T20 and T8-A17) have very small propeller twist angles (6° and 8°, respectively), which are associated with the close adenine N6–adenine N6 distances. The A5N6 to A19N6 distance is 3.3 Å in the A5pT6 step, while the equivalent distance A7N6 to A17N6 is 3.1 Å in the A7pT8 step. From these data, it can be understood why an alternating AT sequence tends to have low propeller twist angles. On the other hand, in an oligo(dA) stretch, the N6–N6 repulsion does not exist between adjacent base pairs; instead, a bifurcated hydrogen bond may be formed from the amino N6 of an adenine to O4 of thymine in an adjacent base pair

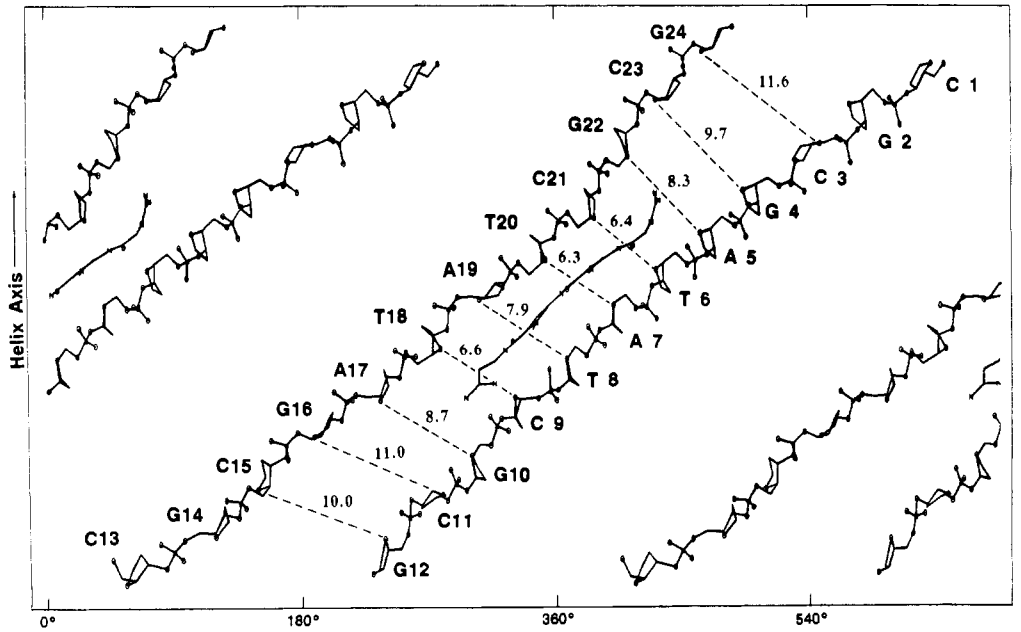

FIGURE 6: Cylindrical projection of the netropsin–d(CGCGATATCGCG) complex. The bases have been removed for clarity. The shortest distances between two sugar O4′ atoms across the minor groove are marked in angstroms, and it can be seen that the central AT stretch is significantly narrower than the outer GC stretch. The aromatic pyrrole rings of the netropsin fit tightly in the narrow minor groove of the AT segment.

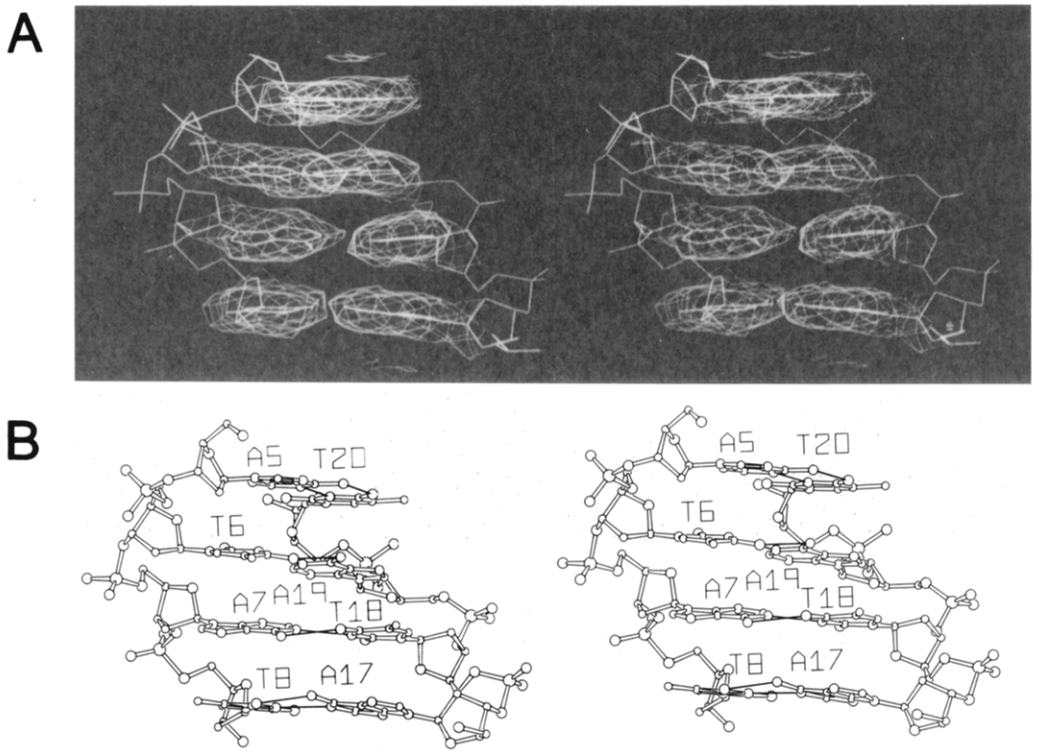

FIGURE 7: (A) Difference Fourier electron density map of the central ATAT base pairs. (B) Stereoscopic skeletal diagrams of the same ATAT base pairs. The low propeller twist angle in the A5/T20 and T8/A17 base pairs is evident.

if there ia a high propeller twist in the stretch (Coll et al., 1987; Nelson et al., 1987).

This raises an interesting question on the relation between the groove width and base pair propeller twist angle. It is clear that the narrow minor groove observed in this structure is not the result of the high propeller twist, unlike those seen in other structures. We are not certain what dictates the groove width of a helix. It seems that the characteristics of the minor groove of B-DNA helix are dependent on the base composition and sequence. The groove may be more easily distorted (e.g., compressed) for AT-rich sequences. It is conceivable that many conformational states, including those associated with normal and narrow minor grooves, are in constant equilibrium. Therefore, the crystal packing forces may allow only those conformers with a narrow minor groove of many related dodecamer molecules to pack in this orthorhombic lattice. Only the molecules having a narrow groove can be accommodated in this lattice, because there are certain close intermolecular contacts between two phosphate groups (P2/P7 of 7.4 Å and P10/P18 of 8.1 Å) from two symmetry-related decamers. It is worth noting that both P7 and P18 are in the ATAT stretch. Those close P–P contacts may restrict the DNA conformation to a very defined range in order to pack them together. In other words, molecules with a wide minor groove would have severe crowding between those phosphate groups.

It is interesting to note that the narrow minor groove may be stabilized by groove binding drugs like netropsin. Keep in mind that the DNA dodecamer d(CGCGATATCGCG) without any drug could not be crystallized in this orthorhombic lattice, supporting the notion that perhaps the ATAT tract has a wider minor groove without a bound drug. On the other hand, dodecamers with $(dA)_n$ tracts in the middle crystallize readily in this orthorhombic lattice with and without groove binding drugs, suggesting that the narrow minor groove already exists in solution stabilized by the high propeller twist of the AT base pairs. It is also interesting to note that another related dodecamer sequence, d(CGCTTTAAAGCG), did not crystallize either with or without any drug. This may be related to the unique TpA step in those sequences which has been suggested to be disruptive to DNA bending (Diekmann, 1987; Burkhoff & Tullius, 1988).

This is consistent with the result in which all dodecamers crystallized in this lattice possess quite similar helical parameters. For example, when we inspect the roll angles of six dodecamers crystallized in this lattice, while they vary somewhat from molecule to molecule, it is abundantly clear that the overall trend is the same in all structures. Foremost are the two steps with a large roll angle, namely, the C3-G4 step with an average of +8° and the G10-C11 step with an average of –9°. It is likely that they are more the consequences of the lattice packing interactions rather than intrinsic properties of a particular dinucleotide step.

Another interesting property of the ATAT stretch in this structure is the alternating helical twist angles. As can be seen in Figure 8, the helical twist is lower in the two ApT steps (33° and 27°, panels A and C of Figure 8, respectively) than in the TpA step (38°, Figure 8B). This agrees with the model proposed for poly(dA-dT) by Klug et al. (1979). However, we have observed an opposite trend, namely, high helical twist in the ApT step than in the TpA step, in the structure of the complex of d(CGCGATATCGCG) with the dye Hoechst 33258 (unpublished results).

Sequence Preference of Netropsin. It is known that netropsin and distamycin bind preferentially to the minor groove of the B-DNA helix at the AT-rich regions (Zimmer & Wahnert, 1986). From the high-resolution chemical footprinting experiments, it can be further shown that those drugs bind better to an oligo(dA) stretch than to an alternating AT stretch (Taylor et al., 1984; Dervan, 1987). The molecular basis for this difference may be understood from the dispositions of the N3 of adenine and O2 of thymine on the floor of the minor groove for various AT sequences. In the structure of the complex of distamycin and d(CGCAAATTTGCG) (Coll et al., 1987), the distances between the successive N3

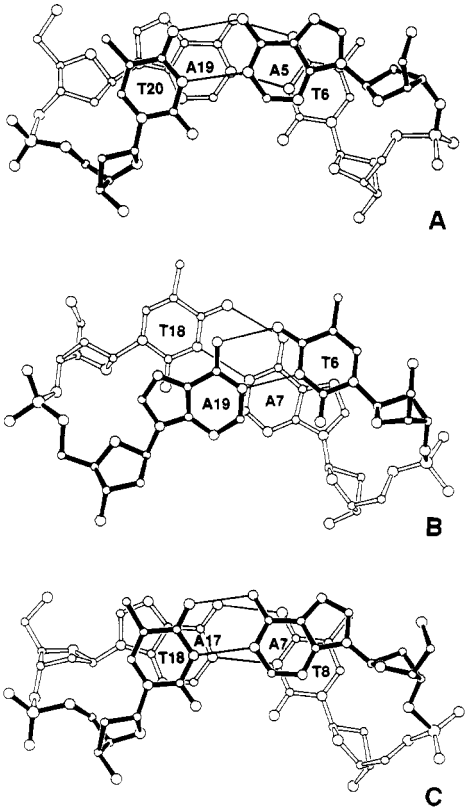

FIGURE 8: Stacking diagram of the three consecutive AT base pairs in the center of the d(CGCGATATCGCG) dodecamer duplex. The view is from the direction perpendicular to the mean plane of the lower base pair (open bonds). Notice the outer two ApT steps (A and C) do not have the same stacking pattern, but both have very close N6 to N6 distances between two adenines in adjacent base pairs. The middle TpA step (B) has almost no base–base overlap, and the two base pairs have relatively large propeller twist angles (16°) resulting a very close C2 to C2 close contacts in the minor groove between two neighboring adenines. The helical twist is lower in the ApT steps [33° and 27° in (A) and (C), respectively] than in the TpA step [38° in (B)].

and O2 (or O2 to O2) of two neighboring AT base pairs range from 3.4 to 4.4 Å. Those separations are ideally suited for a three-centered (bifurcated) hydrogen bond by placing a hydrogen-bond donor (usually an NH group from the amide linkage) halfway between the two acceptor atoms near the bottom of the groove. Water molecules can also fulfill this role, thereby connecting a run of A's through a spin of hydration network. The result of this is that several hydrogen bonds can be formed between the drug and DNA. For example, there are five hydrogen bonds in the netropsin–d(CGCGAATTCGCG) complex (Kopka et al., 1985), while there are seven in the distamycin–d(CGCAAATTTGCG) complex (Coll et al., 1987).

In the alternating AT sequence, the separations between two adjacent N3 and O2 atoms became highly uneven. In the current structure, the O2 to O2 distance for two pairs of thymines are 4.8 and 4.7 Å, respectively, for T6–T20 and T8–T18, whereas the N3–N3 distance is much shorter (3.3 Å) for A7–A19 bases. It becomes difficult for the hydrogen-bond donor NH group to form a bifurcated hydrogen bond for either the O2–O2 pair or the N3–N3 pair. In the former case, the NH group would have to get quite close to both acceptor atoms in order to attain the necessary distance (2.7–3.1 Å) for the hydrogen bond. This would lead to close van der Waals clashes from other parts of the drug molecule. A converse effect would be in place for the latter situation. There the shorter N3–N3 distance would cause the drug molecule to fall away from the DNA, thereby destabilizing the binding of drug molecule. The net effect is that the netropsin adjusts itself to form single, rather than three centered bifurcated, hydrogen bonds between NH of netropsin and O2 (and N3) of DNA. However, this adjustment perturbs the precise registration of the successive NH groups, which are optimally spaced in the slightly curved netropsin molecule for bifurcated interactions, to their corresponding acceptors at the floor of the minor groove. As a consequence, fewer hydrogen bonds can be formed between netropsin (and other minor groove binding compounds) and alternating AT DNA sequences as can be seen clearly in Figure 9. This provides a satisfactory explanation at the molecular level for the different AT binding affinity of netropsin to DNA observed by the chemical footprinting experiments.

Dynamics of Drug Binding. Our refinement results suggested a plausible model of a disordered netropsin in the crystal lattice which reinforces the experimental interpretation of chemical affinity cleaving studies on restriction DNA fragments by synthetic distamycin-like molecules which have an EDTA–Fe(II) cleaving function attached at one end. From the cutting pattern on the high-resolution gel, it was suggested that distamycin binds to DNA sites in two orientations with an asymmetric extent of cleavages at either end of the binding site (Taylor et al., 1984). This is also in complete accord with several NMR studies in which the minor binding compounds are shown to interact with DNA double helix in a "flip-flop" fashion on the NMR time scale (Patel, 1979; Klevit et al., 1986; Leupin et al., 1986). Therefore, one can envision a dynamic process existing in solution where the asymmetric netropsin hops on and off the self-complementary double helix to create a nonsymmetric complex which can deposit itself onto the crystal surface of the nucleating crystals in the $P2_12_12_1$ lattice in either one of the two orientations. This would generate a statistically 2-fold disordered structure.

This type of 2-fold disordered structure is not uncommon in oligonucleotide crystals as shown by the nanomer d(GGATGGGAG)·d(CTCCCATCC) in the A-DNA conformation (McCall et al., 1986) and the B-DNA dodecamer d(CGCAAAAATGCG)·d(CGCTTTTTAGCG) duplex (T. Steitz, personal communication). Another smaller synthetic compound, DAPI, has also recently been shown to bind to the minor groove of a d(CGCGAATTCGCG) double helix with four different disordered positions (Goodsell et al., 1988). This highly disordered binding mode has been attributed to the small size of DAPI, which can only cover three base pairs.

It is interesting to note that four other complexes, netropsin–d(CGCGAATTCGCG) (Kopka et al., 1985), distamycin–d(CGCAAATTTGCG) (Coll et al., 1987), Hoechst 33258–d(CGCGAATTCGCG) (Pjura et al., 1987; Teng et al., 1988), and Hoechst 33258–d(CGCGATATCGCG) (unpublished results) crystallize in the isomorphous lattice in only one unique orientation. In the case of the distamycin complex, this may be due to the hydrogen bond between the guanidinium end of distamycin and the O3′ of G24 which may predispose a slight preference for this particular alignment. Similar small but significant interactions may be in operation for the other two complexes. Those structures also support the notion that the orientations in the crystal are energetically very similar. If no specific lattice interactions could be established to bias the stabilization of a particular orientation, then a disordered structure could easily be formed.

Generalizations. The present structure of the complex between netropsin and d(CGCGATATCGCG) not only reaffirms previous observations on the manner in which a minor

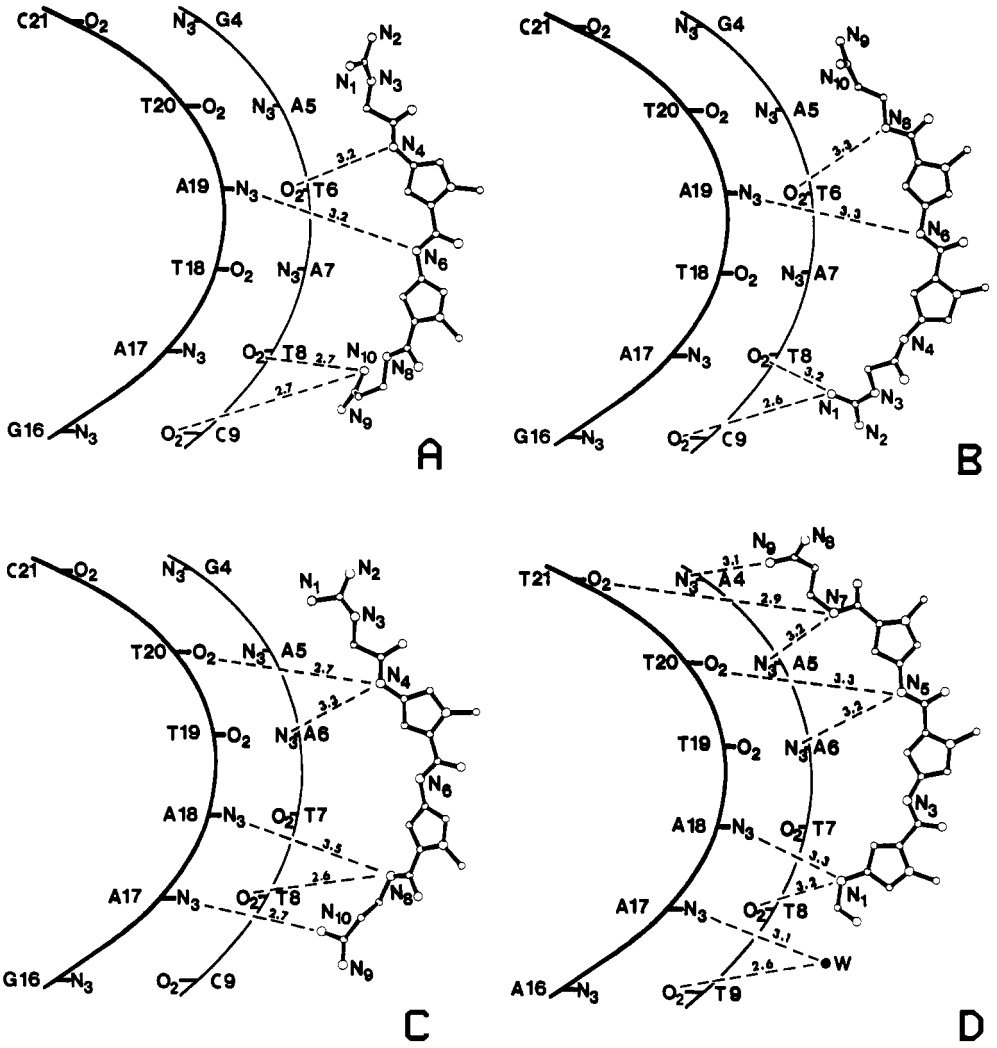

FIGURE 9: Comparative illustration of different complexes of DNA with netropsin/distamycin. (A and B) The netropsin–d(CGCGATATCGCG) complexes with netropsin in two possible orientations. Four single hydrogen bonds are found (this work). (C) Netropsin–d(CGCGAATTCGCG) complex. Five hydrogen bonds exist between netropsin and DNA, of which four are bifurcated hydrogen bonds from the amide nitrogen atoms N4 and N8 to base O2 and N3 atoms (Kopka et al., 1985). (D) Distamycin–d(CGCAAATTTGCG) complex. Three sets of bifurcated hydrogen bonds are found (Coll et al., 1987). Notice that in (C) and (D) the orientation of the drug with respect to the DNA double helix is reversed.

groove binding drug interacts with a B-DNA double helix but also provides us with new insight on sequence-dependent conformations of DNA. As summarized earlier (Teng et al., 1988), four major factors are responsible for the AT binding preference for netropsin, distamycin, and related compounds. All of them are positively charged molecules, thereby enhancing the nonspecific attraction between the drug and DNA. In addition, it has been suggested that the deepest negative charge potential of a DNA double helix residues near the bottom of the minor groove at the AT region. The drug molecule may be driven into the vicinity of the AT region by such potential. The third factor is the AT sequence may be associated with a higher tendency to adopt a narrow minor groove which would provide many van der Waals stabilizing interactions with the sugar–phosphate side walls of the groove. The dipole interactions between the O4′ atoms and the aromatic pyrrole π-electron clouds immobilize the drug in place so that the amide NH group can form hydrogen bonds to N3 of adenine and O2 of thymine on the floor of the minor groove. Finally, the amino NH2 group of guanine bases presents a prominent steric hindrance to the entry of the drug into the groove, hence establishing a definite AT binding preference.

The current structure extends the role of a drug like netropsin beyond being merely a passive ligand attracted to the binding sites. Much like an intercalator which extends the helix length by 3.4 Å, netropsin is capable of helping to collapse a "normal" minor groove into a "narrow" groove as evident from the present structure. The ATAT sequence in the dodecamer d(CGCGATATCGCG) alone may have a low intrinsic propeller twist angle which may disfavor the formation of a narrow minor groove. The binding of netropsin to the ATAT segment will allow the sugar–phosphate backbones to come in close van der Waals contacts with netropsin to create a narrow groove, while still maintaining the low propeller twist in the AT base pairs. The readjustment of the backbone would cost slightly extra energy which may account partly for the low affinity of netropsin to alternating AT sequence, in comparison to that for an oligo(A) sequence. This kind of small but significant conformational rearrangement undoubtly can also be achieved by proteins interacting with DNA.

Finally, the TpA step may destabilize a run of high propeller twist base pairs in the AT-rich tract. Its presence would tend to disrupt the propagation of a base pair with high propeller twist which would facilitate narrowing of the minor groove. One prediction is that the alternating poly(dA-dT)·poly(dA-dT) double helix would resemble the mixed-sequenc B-DNA structure with a normal minor groove width. Thus, it would be less likely to accommodate molecules like netropsin due to

the TpA steps (50%) in the polymer.

This structure allows us to visualize a new property associated with a particular dinucleotide step, in this case low propeller twist with TpA. We have also noticed other unique qualities associated with a CpG step in the A-DNA conformation which has been studied in three different lattices for the following three molecules, respectively: d(GGCCGGCC) (Wang et al., 1982), d(CCCCGGGG) (Haran et al., 1987), and d(ACCGGCCGGT) (Frederick et al., 1988b). There, the CpG step has an unusually low helical twist angle (~25°). There are likely to be other examples such as these. More structures need to be solved in order to compile a comprehensive dictionary of all the building blocks of DNA, so that the detailed conformation of DNA can be accurately predicted on the basis of its nucleotide sequence.

ACKNOWLEDGMENTS

We thank Dr. C. Frederick for useful suggestions and Dr. G. J. Quigley for providing various graphic plotting programs.

Registry No. Netropsin–d(CGCGATATCGCG), 117269-76-8.

REFERENCES

Berman, H. M., Neidle, S., Zimmer, C., & Thrum, H. (1979) *Biochim. Biophys. Acta 561*, 124–131.

Burkhoff, A. M., & Tullius, T. D. (1988) *Nature (London) 331*, 455–457.

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

Dervan, P. B. (1986) *Science (Washington, D.C.) 232*, 464–471.

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

Diekmann, S. (1987) *Nucleic Acids Mol. Biol. 1*, 138–156.

Frederick, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., & Rich, A. (1988a) *J. Biol. Chem.* (in press).

Frederick, C. A., Teng, M.-K., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Rich, A., & Wang, A. H.-J. (1988b) (submitted for publication).

Goodsell, D. S., Larsen, T., Cascio, D., Grzeskowiak, K., & Dickerson, R. E. (1988) (submitted for publication).

Haran, T. E., Shakked, Z., Wang, A. H.-J., & Rich, A. (1987) *J. Biomol. Struct. Dyn. 5*, 199–217.

Hendrickson, W. A., & Konnert, J. (1979) in *Biomolecular Structure, Conformation, Function and Evolution* (Srinvasan, R., Ed.) pp 43–57, Pergamon, Oxford.

Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.

Klevit, R. E., Wemmer, D. E., & Reid, B. R. (1986) *Biochemistry 25*, 3296–3303.

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

Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1376–1380.

Leupin, W., Chazin, W. J., Hyberts, S., Denny, W. A., & Wuthrich, K. (1986) *Biochemistry 25*, 5902–5908.

McCall, M., Brown, T., Hunter, W. N., & Kennard, O. (1986) *Nature (London) 322*, 661–664.

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

Patel, D. J. (1978) *Eur. J. Biochem. 99*, 369–379.

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

Rich, A., Nordheim, A., & Wang, A. H.-J. (1984) *Annu. Rev. Biochem. 53*, 791–846.

Taylor, J. S., Schultz, P. G., & Dervan, P. B. (1984) *Tetrahedron 40*, 457–465.

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

van der Marel, G. A., van Boeckel, C. A. A., Wille, G., & van Boom, J. H. (1981) *Tetrahedron Lett. 22*, 3887–3888.

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. A., & Rich, A. (1979) *Nature (London) 282*, 680–686.

Wang, A. H.-J., Fujii, S., van Boom, J. H., & Rich, A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3968–3972.

Wang, A. H.-J., Ughetto, G., Quigley, G. J., Hakoshima, T., van der Marel, G. A., van Boom, J. H., & Rich, A. (1984) *Science (Washington, D.C.) 225*, 1115–1121.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature (London) 287*, 755–758.

Zakrzewska, K., Lavery, R., & Pullman, B. (1983) *Nucleic Acids Res. 11*, 8825–8839.

Zimmer, C., & Wahnert, U. (1986) *Prog. Biophys. Mol. Biol. 47*, 31–112.