Minor Groove Binding of SN6999 to an Alkylated DNA: Molecular Structure of d(CGC[e^{6}G]AATTCGCG)–SN6999 Complex†

Yi-Gui Gao,‡ M. Sriram,‡ William A. Denny,§ and Andrew H.-J. Wang*,‡

Division of Biophysics and Department of Cell & Structural Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, and Cancer Research Laboratory, Auckland Medical School, University of Auckland, Private Bag 92019, Auckland, New Zealand

Received April 8, 1993; Revised Manuscript Received July 1, 1993®

ABSTRACT: The interaction between a potent synthetic antitumor and antiviral minor groove binding drug 1-methyl-4-[4-[4-(4-(1-methylquinolinium)amino)benzamido]anilino]pyridinium dichloride (SN6999) and an alkylated DNA d(CGC[e^{6}G]AATTCGCG) dodecamer has been studied by X-ray crystallography. The complex forms a *new* crystal lattice in the space group $P2_12_12_1$ with unit cell dimensions of a = 28.48 Å, b = 36.11 Å, and c = 69.60 Å. The structure has been solved by the molecular replacement method and refined to an R-factor of 17.0% at ~2.5 Å resolution using 1618 reflections. In the complex, the SN6999 covers almost six base pairs in the narrow minor groove with the 1-methylquinolinium (Q) ring near T8–A17 and the 1-methylpyridinium (P) ring near the C3–G22 base pair. The central benzamido (BQ) and anilino (BP) rings are essentially coplanar, with the Q and P rings having large dihedral angles of 38° and 39°, respectively, to the plane of BQ/BP. There is only one direct hydrogen bond between the amide NH of SN6999 to T20O^2 of DNA. The drug–DNA interaction is stabilized by stacking interaction of sugar oxygens from T20O$^{4'}$ to BQ and C21O$^{4'}$ to BP. There is charge-induced dipole interaction between the positively charged nitrogen atom of 1-methylquinolinium with C9O$^{4'}$ and that of 1-methylpyridinium with G22O$^{4'}$. The crystal structure of the complex can be used to explain the NMR results. SN6999 lacks the crescent shape observed in other minor groove binding drugs and distorts the DNA duplex upon binding. The complex packs in the lattice using the G–N^2:G–N^3 interlocking base pairs at both ends of the helix. As in earlier cases, the two independent e^{6}G:C base pairs adopt different base pairing schemes. The e^{6}G16:C9 base pair adopts a previously observed bifurcated configuration involving three-centered hydrogen bonds and is similar to a Watson–Crick pairing. In contrast, the e^{6}G4:C21 base pair adopts a novel "reverse wobble" configuration with C21 being pushed toward the major groove side. The ethyl group is in the *proximal* orientation (to N^7) in both base pairs. Taken together with the observations found in the same DNA complexed to Hoechst 33258, Hoechst 33342, and netropsin from different crystal lattices, the results suggest that the e^{6}G:C base pairing is weak and polymorphic when compared to a normal G:C base pair and the DNA duplex containing this lesion is readily distorted.

Design of sequence-specific DNA binding molecules has been the goal of a number of intensive studies. Minor groove binding antibiotics (e.g., netropsin and distamycin) have been shown to be excellent examples of AT-recognition molecules.[1] These elongated and crescent-shaped molecules, consisting of planar pyrroles linked by amide bonds, bind to B-DNA narrow minor groove associated with AT-sequences through a combination of electrostatic, van der Waals, and hydrogen bonding interactions (Wang & Teng, 1990). Several synthetic compounds (e.g., Hoechst 33258, DAPI) have been shown to possess similar minor groove binding characteristics. The molecular basis of the interactions has been probed by various biophysical techniques including X-ray crystallography, NMR, and modeling studies (Wang & Robinson, 1992; Kopka & Larsen, 1992; Remers *et al.*, 1992). In particular, the complexes of a series of related dodecamers with minor groove binding drugs have been extensively studied by X-ray crystallography (Coll *et al.*, 1987, 1989; Carrondo *et al.*, 1989; Teng *et al.*, 1988; Wang & Teng, 1990; Sriram *et al.*, 1992a,b; Kopka & Larsen 1992).

More recently, it has been discovered that distamycin may bind to B-DNA minor groove with two distamycins "stacked" side-by-side in a head-to-tail fashion (Pelton & Wemmer, 1989, 1990). This observation stimulated the syntheses of compounds that bind to AT/GC mixed sequences (Dervan, 1986; Mrksich *et al.*, 1992). These results offer hope in the design of more general sequence-specific DNA recognizing molecules. However, a fuller understanding of all the forces involved in the recognition process is needed to reach the point of rationally and reliably designing sequence-specific DNA binding molecules.

1-Methyl-4-[4-[4-(4-(1-methylquinolinium)amino)benzamido]anilino]pyridinium dichloride (SN6999) (Figure 1) is a synthetic compound belonging to a family of quinoline-containing DNA binders, which have been shown to possess potent antitumor and antiviral activities both *in vivo* and *in vitro* (Cain *et al.*, 1969; Robertson & Baguley, 1982). A

† This work was supported by NIH Grants GM-41612 and CA-52506 (A.H.-J.W.).

* To whom correspondence should be addressed.

‡ University of Illinois at Urbana-Champaign.

§ University of Auckland.

® Abstract published in *Advance ACS Abstracts*, September 1, 1993.

[1] Abbreviations: A, T, G, C, adenine, thymine, guanine, cytosine or their respective nucleotide in DNA; m^{6}G, O^6-methylguanine or O^6-methyldeoxyguanosine; e^{6}G, O^6-ethylguanine or O^6-ethyldeoxyguanosine; H258, Hoechst 33258; H342, Hoechst 33342; e^{6}G_SN6999, d(CGC[e^{6}G]AATTCGCG)–SN6999 complex; e^{6}G_DH258, d(CGC[e^{6}G]AATTCGCG)–Hoechst 33258 complex; e^{6}G_DH342, d(CGC[e^{6}G]AATTCGCG)–Hoechst 33342 complex; e^{6}G_DNET, d(CGC[e^{6}G]AATTCGCG)–netropsin complex; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect crosspeak; RMSD, root mean square deviation; SN6999, 1-methyl-4-[4-[4-(4-(1-methylquinolinium)amino)benzamido]anilino]pyridinium dichloride.

0006-2960/93/0432-9639$04.00/0

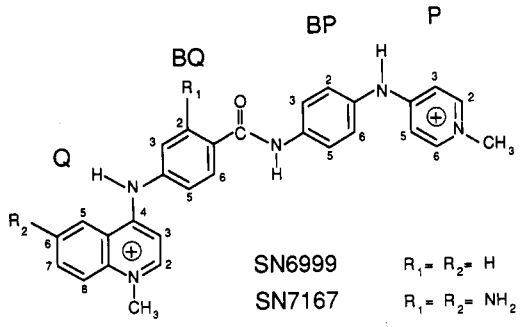

FIGURE 1: Molecular formula of 1-methyl-4-[4-[4-(4-(1-methylquinolinium)amino)benzamido]anilino]pyridinium dichloride (SN6999) and a derivative SN7167. The atom numbering scheme for each ring Q, BQ, BP, and P is as indicated and methyl carbons of Q and P are called CM and the peptide atoms are called CP#, OP#, and NP#.

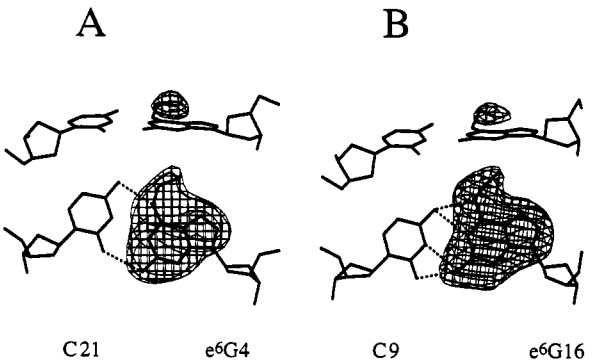

FIGURE 2: Four ($|F_o| - |F_c|$) difference Fourier maps showing the two e^6G–C base pairs of the d(CGC[e^6G]AATTCGCG)–SN6999 complex where either the e^6G ethyl group (top) or the entire e^6G base (bottom) is removed from the phase calculation. (A) The e^6G4:C21 base pair adopts the reverse wobble configuration. (B) The e^6G16:C9 base pair adopts the bifurcated configuration. Both ethyl groups are in the *proximal* (to N7) orientation and out of the plane of guanine base.

derivative of SN6999 (NSC 176 319) underwent preclinical and toxicological testing in preparation for clinical trials but was not evaluated due to severe side effects (Plowman & Adamson, 1978). NMR studies have shown that SN6999 binds to the minor groove along AT-stretches of DNA on the basis of the combined information from intermolecular NOEs, exchange-based line broadening effects, and changes in chemical shifts induced by drug binding (Leupin *et al.* 1986; Chen *et al.*, 1992). However, no detailed three-dimensional structure of the complex was available.

Recently, we have shown that the structure of O^6-alkylated guanine-containing DNA may be stabilized by minor groove binding compounds (netropsin, Hoechst 33258, and Hoechst 33342) and crystallized for structural analyses (Sriram *et al.*, 1992a,b). The structures of those complexes have allowed us to visualize the polymorphism of the e^6G:C base pair configurations. In NMR studies of DNA with e^6G:C or m^6G:C a wobble type of base pairing was found (Patel *et al.*, 1986; Kalnik *et al.*, 1989a,b). In the Z-DNA crystal structure of d(CGC[m^6G]CG), a G:C^+ Watson–Crick type was found (Ginell *et al.*, 1990). In the structures of the complexes of d(CGC[e^6G]AATTCGCG) with Hoechst 33258 (H258), Hoechst 33342 (H342) (Sriram *et al.*, 1992a), and netropsin (Sriram *et al.*, 1992b), either the wobble or the bifurcated (near Watson–Crick type with either one or two sets of three-centered hydrogen bonds) base pair was found.

In this paper we solved a new crystal structure of the complex of d(CGC[e^6G]AATTCGCG) with SN6999 to address two issues at the same time. First, it provided us an unambiguous view of how SN6999 binds to DNA. The comparison of SN6999 binding in the present structure with the available NMR results (Chen *et al.*, 1992) offered an understanding of the static and dynamic structural aspects of the drug binding. Second, it expands our study on the influence of O^6-alkylguanine on the B-DNA conformation. One of two e^6G:C base pairs in this new structure adopts a heretofore unseen "reverse wobble" base-pairing scheme. These structural clues reinforce the hypothesis on the possible functional consequences due to the alkylation of DNA. The e^6G:C base pair may adopt multiple conformations causing recognition ambiguity during replication/transcription. Furthermore, this inherent "weakness" of the e^6G:C base pair is observed to destabilize the DNA duplex, making it readily deformable.

MATERIALS AND METHODS

The preparation of d(CGC[e^6G]AATTCGCG) has been described by Roelen *et al.* (1990) and the synthesis of SN6999 by Cain *et al.* (1969). Crystallization experiments followed the procedure of Wang and Gao (1990). We were able to obtain crystals of the complex between d(CGC[e^6G]-AATTCGCG) and SN6999 from a solution containing 0.5 mM dodecamer (single strand concentration), 0.5 mM SN6999, 30 mM cacodylate buffer at pH 7.0, 8 mM $MgCl_2$, 1.2 mM spermine, 60 mM ammonium acetate at pH 7.0, and 2.5% 2-methyl-2,4-pentanediol (2-MPD), equilibrated against 50% 2-MPD by vapor diffusion at room temperature. One large crystal with somewhat irregular shape was obtained by repeatedly feeding fresh solution to an existing seed crystal over a period of 8 weeks. The crystal was mounted in a thin-walled glass capillary and sealed with a droplet of the crystallization mother liquor for data collection. The peak profile showed that the crystal was slightly twinned, but reasonable quality data were obtained as judged by the refined structure discussed later. It was found in the space group $P2_12_12_1$ by the index program of TEXSAN of Molecular Structure Corp. However, it is not isomorphous to any of the existing $P2_12_12_1$ crystal forms of several related dodecamers. The new crystal has unit cell dimensions of a = 28.48 Å, b = 36.11 Å, and c = 69.60 Å (V = 71 578 Å^3), in comparison to the dimensions of $a \approx 25$ Å, $b \approx 41$ Å, and $c \approx 66$ Å ($V \approx 70\,000$ Å^3) from the previous forms. The diffraction data set was collected at room temperature on a Rigaku AFC5R rotating-anode diffractometer, equipped with a copper anode and a graphite monochromator, at a power of 50 kV and 180 mA. An ω-scan mode was used for data collection with Cu Kα radiation (1.5406 Å), and the data set was collected to a resolution of 2.25 Å. Lorentz polarization, absorption, and decay corrections were applied and data reduction was done to obtain the structure factor magnitudes. A total of 1618 reflections were considered observable at $2\sigma(F_o)$ level, comprising of ~50% of possible reflections to 2.25 Å resolution. The number of observed reflections above $2\sigma(F_o)$ fell sharply beyond 2.5 Å resolution.

The B-DNA dodecamer without drug or ethyl group from the structure of e^6G_DNET (PDB Ent# 1D85; Sriram *et al.*, 1992b) was used as the starting model for the molecular replacement search using the ULTIMA program (Rabinovich & Shakked, 1984). The best solution (with the minimum R-factor of 45% using 25–7-Å resolution data after the rigid body refinement using the LOOP program in ULTIMA) was found at a new position in the unit cell. The location for the center of the gravity of the model was found at ϕ = 58°, ψ = 81°, θ = 88° and u = 0.143, v = 0.500, w = 0.675. This new

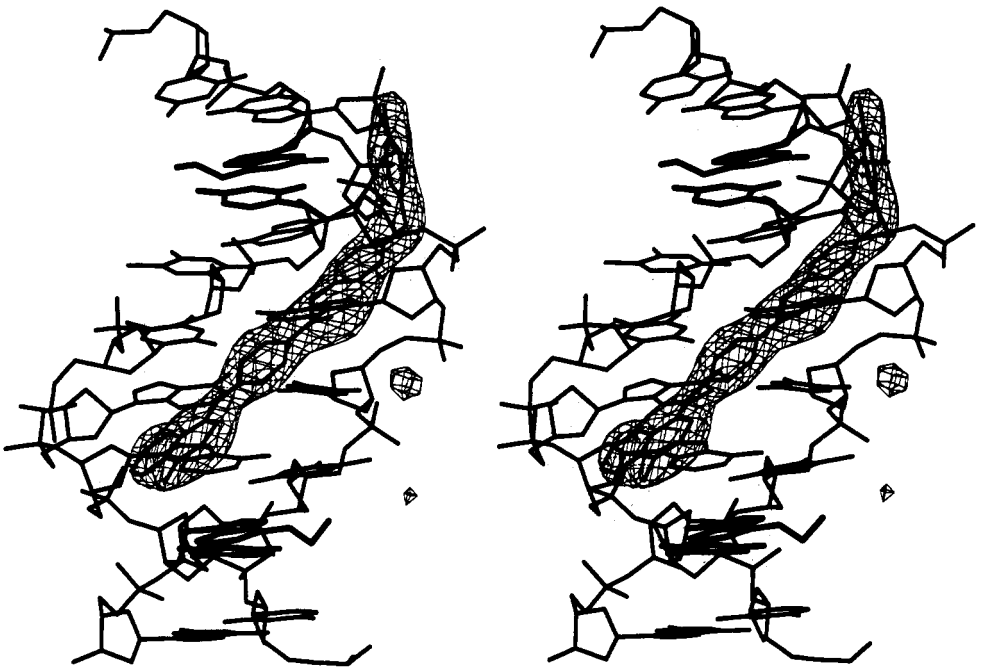

FIGURE 3: Stereoscopic ($|F_o| - |F_c|$) difference Fourier maps displaying in detail the six base pairs –C[e^{6}G]AATT– of the SN6999 complex where the drug is removed from the phase contribution. The density is well resolved, allowing for the position and the polarity of the drug to be determined.

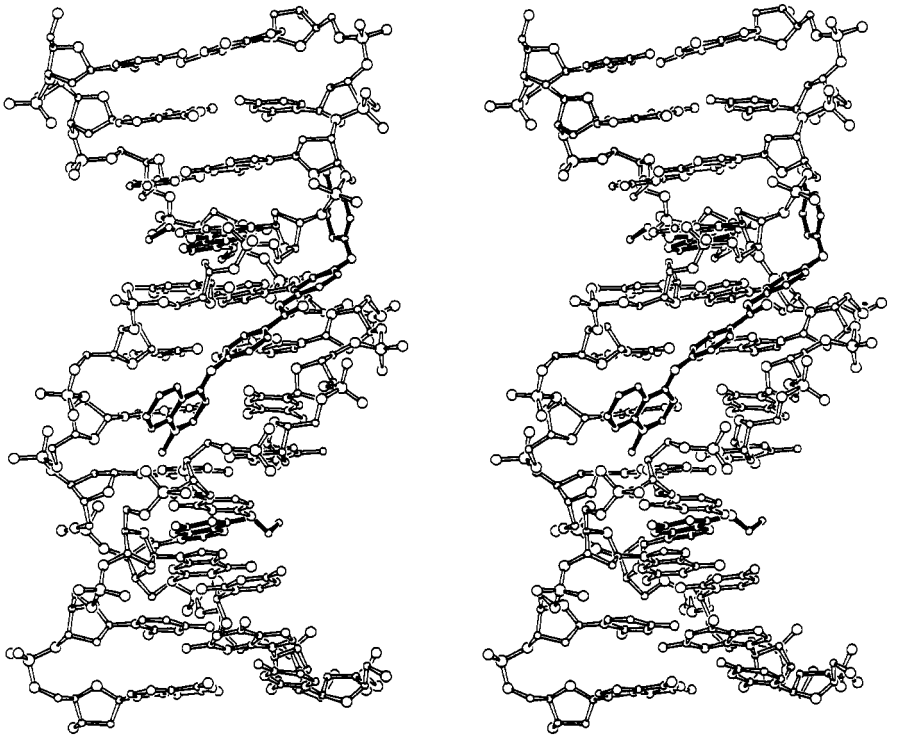

FIGURE 4: Stereoscopic skeletal drawing of the structure of the d(CGC[e^{6}G]AATTCGCG)–SN6999 complex. The SN6999 molecule is drawn with filled bonds binds in the minor groove of the B-DNA duplex. The two e^{6}G's are drawn with filled bonds, and their ethyl groups are located in the major groove.

position is different from that found in related e^{6}G dodecamers (Sriram *et al.*, 1992a,b) or in the d(CGCGAAAACGCG) + d(CGCGTT) + d(TTCGCG) gapped dodecamer (Aymami *et al.*, 1990). The position of the dodecamer in the new structure is related to the canonical dodecamer lattice by an x translation of ~6 Å, a y translation of ~3 Å, a z translation of ~–37 Å, and a rotation about the z axis of ~13°. The entire dodecamer duplex is in the asymmetric unit, so that the conformation of the two strands in the duplex are not identical. The model (after LOOP refinement) was refined using the Konnert–Hendrickson constrained refinement procedure (Hendrickson and Konnert, 1979; Westhof *et al.*, 1985). After many cycles of refinement with all available data, the R-factor converged to ~25.3% at 2.5 Å resolution. A difference Fourier ($|F_o| - |F_c|$) map revealed a clear continuous residual density in the minor groove of the DNA near the central AATT site which could be unambiguously assigned to a SN6999 molecule. The coordinates of the drug model built on QUANTA (Polygen Inc.) were used to locate the drug in the difference Fourier ($|F_o| - |F_c|$) map using the

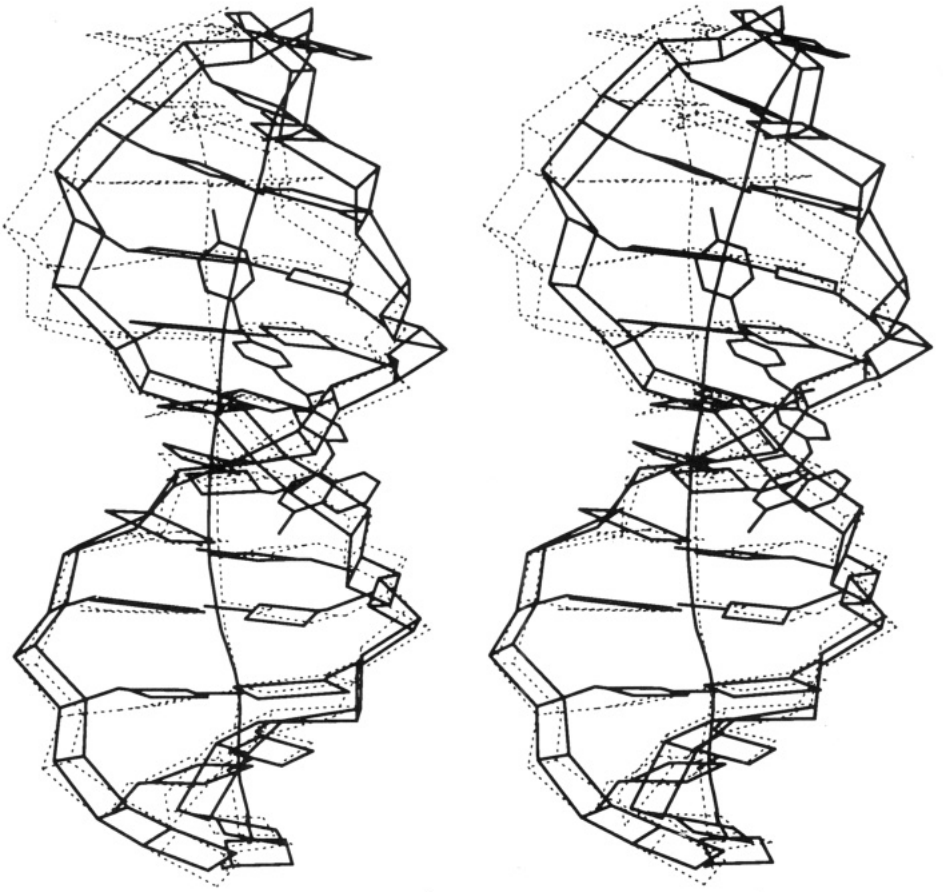

FIGURE 5: Least squares fit between the DNA duplexes from e^{6}G–SN6999 and e^{6}G–DNET complexes. The atoms of half the helix, i.e., base pairs T7:A18 to G12:C13, were used for least squares fit, and the RMSD is 0.90 Å. The DNA schematic and the helix axis shown are those calculated by CURVE (Lavery & Sklenar, 1989). The difference in helix curvature due to SN6999 binding is obvious.

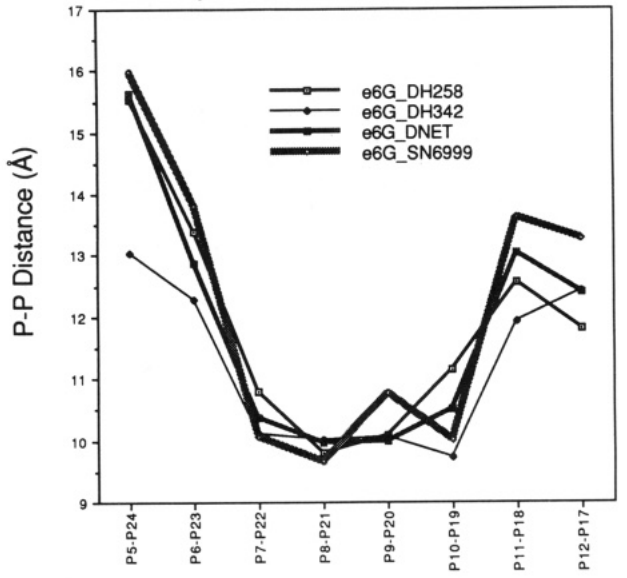

FIGURE 6: Comparison of the minor groove width of the complexes between d(CGC[e^{6}G]AATTCGCG) and four different minor groove binding drugs. Phosphate to phosphate distances in Å are shown for the two closest phosphates along the minor groove.

program FRODO/TOM (Jones, 1978) and included in the refinement. Inclusion of well-ordered solvent molecules located from the $2|F_o| - |F_c|$ Fourier maps in the subsequent refinement cycles gradually reduced the R-factor to about ~18%. Finally, a difference Fourier ($|F_o| - |F_c|$) map was used to locate the position of the ethyl groups of the e^{6}G bases. Throughout the refinement the base pair configuration of two independent e^{6}G:C base pairs was not constrained. The final two base pair configurations are shown in Figure 2.

The difference Fourier map, calculated by removing SN6999 from the phase contribution, clearly showed that the drug molecule fits well in the caterpillar-shaped residual electron density envelope with one end bulkier than the other end (Figure 3). This suggests that SN6999 binds only in one orientation in the crystal. No magnesium or spermine ions could be explicitly identified. The final R-factor is 17.0% using 1618 reflections ($F_o > 2\sigma(F_o)$) between 10 and 2.25 Å resolution. The root mean square deviation of bond distance from ideal value is 0.019 Å. The final model includes one e^{6}G-dodecamer duplex, one SN6999 molecule, and 68 water molecules in the asymmetric unit. The atomic coordinates of the structure have been deposited at Brookhaven Protein Databank.

RESULTS AND DISCUSSION

Structure of the Complex. The overall three-dimensional structure of the e^{6}G–SN6999 complex (Figure 4) is similar to those of the other related dodecamer-drug complexes (Coll *et al.*, 1987, 1989; Carrondo *et al.*, 1989; Teng *et al.*, 1988, Sriram *et al.*, 1992a,b; Kopka & Larsen, 1992). The SN6999 drug binds in the narrow minor groove of B-DNA, covering the central AATT sequence, but off-center toward one end of the helix. Since the complex forms a different crystal lattice, it is of interest to ask whether the general features associated with the existing dodecamer structures, like the 19° bend of the helix axis, are conserved. Calculation of the helical parameter of the e^{6}G–SN6999 complex by CURVE program

Table I: Selected Helical Parameters[a] of SN6999 d(CGC[e⁶G]AATTCGCG) Complex

base	τ_M	P	sugar pucker	Ω	κ	ω
C1	61	175	C2′-endo	30	8	−4
C13	34	149	C2′-endo			
G2	39	139	C1′-exo	42	−5	−17
G14	32	136	C1′-exo			
C3	49	158	C2′-endo	29	7	−7
C15	27	96	C4′-endo			
e⁶G4	53	162	C2′-endo	37	−11	−2
e⁶G16	43	149	C2′-endo			
A5	29	146	C2′-endo	35	1	−13
A17	33	140	C1′-exo			
A6	46	120	C1′-exo	35	1	−10
A18	61	130	C1′-exo			
T7	55	175	C2′-endo	33	−11	−18
T19	38	136	C1′-exo			
T8	55	158	C2′-endo	41	13	−2
T20	34	151	C2′-endo			
C9	47	154	C2′-endo	30	5	−8
C21	54	166	C2′-endo			
G10	35	116	C1′-exo	36	4	−4
G22	50	134	C1′-exo			
C11	43	157	C2′-endo	33	5	−14
C23	50	60	C4′-exo			
G12	17	124	C1′-exo		−15	−10
G24	26	133	C1′-exo			

[a] Nomenclature of the helical parameters follows that of Dickerson *et al.* (1990). Ω is the helical twist angle, and κ is the base pair buckle. ω stands for propeller twist angle. τ_M and P are the pseudorotation amplitude and angle, respectively.

(Lavery & Sklenar, 1989) indicated that its helix axis is bent by 25° but in a different manner. The helix axis bend is readily seen by a least-squares fitting of DNA in this new e⁶G–SN6999 complex with that of e⁶G–DNET complex (Figure 5).

The overall RMSD between all DNA atoms of e⁶G–SN6999 and e⁶G–DNET is 1.35 Å. In comparison, the RMSD of the DNA duplexes amongst the three e⁶G–DNA–drug complexes of the old lattice range from 0.77 to 0.89 Å. If one-half of the helix duplex (base pairs T7–A18 to G12–C13) is used for the least-squares fit (Figure 5), a large displacement is seen on the other end of the helix. Here the RMSD of the superimposed hexanucleotide fragment is 0.90 Å. The top end of the e⁶G–SN6999 complex is significantly tilted toward the right to accommodate the relatively rigid and straight drug moiety in the minor groove. Some of the atoms are now displaced by as much as 4.5 Å between e⁶G–SN6999 and e⁶G–DNET (Figure 5).

The influence of the binding of a new minor groove-binder and of a new crystal lattice on local structures (e.g., base pair buckle and propeller twist angles) and global features (e.g., P–P distance across the minor groove) is evident. Figure 6 shows a comparison of the minor groove width among four different e⁶G-containing dodecamers. It is interesting to see that despite the new crystal lattice the DNA duplex in e⁶G–SN6999 maintains a narrow minor groove near the center of the helix. Even the asymmetric shape of the curve is preserved. A list of selected helical parameters of the e⁶G–SN6999 complex is shown in Table I. The average helical twist angle (Ω) of 34.6° is similar to other related B-DNA dodecamers. Table II compares the buckle (κ) and the propeller twist (ω) angles among the four e⁶G related complexes {e⁶G–SN6999, e⁶G–DNET, e⁶G–DH258, e⁶G–DH342}. It can be seen that nearly every base pair in the three previous e⁶G-duplexes has one of those two angles (κ and ω) greater than 10°. But in general the e⁶G–SN6999 complex has lower values, with five out of the 12 base pairs having both κ and ω lower than 10°.

Interestingly, the two G:C base pairs at both ends of the helix again are involved in the interlocking lattice interactions using the G14:G24# and G12:G2# (# is from a symmetry-related duplex) hydrogen bonding pairing in the minor groove. We have noted before that this type of G–N^2:G–N^3 pairing is associated with a high dihedral angle between the two guanines (Coll *et al.*, 1990). In the e⁶G–SN6999 crystal, the dihedral angles of the G14:G24# and G12:G2# base pairs are 45° and 45°, respectively. These values are slightly larger than those from the other $P2_12_12_1$ crystal forms. This clearly has an effect on the conformation of the participating base pairs with the terminal G12:C13 having high buckle (−15°), whereas both penultimate G2:C23 and C11:G14 base pairs have a high propeller twist angle ω of −14° and −17°, respectively. The central AT base pairs in e⁶G–SN6999 have significant propeller twist angles, but not nearly as large as other dodecamers. In the present structure, only the T7:A18 maintains a high propeller twist of −18°.

Binding and Structure of SN6999. The position of SN6999 spans across the –C[e⁶G]AATT– site from the edge of C3 to the edge of T8 base pairs. The bulkier 1-methylquinolinium (Q) ring is near the middle of the helix, while the 1-methylpyridinium (P) ring is near the end of the helix. The central two rings (BQ and BP) are nearly coplanar, joined by a trans amide bond. The amide NH points toward the base pair with a hydrogen bond to T20O^2 (3.11 Å). Both the Q and P rings are joined to the BQ–BP plane through amine groups and are twisted sharply relative to it.

There are many close contacts (46 non-hydrogen atom contacts are shorter than 3.5 Å) between SN6999 and DNA

Table II: Comparison of Buckle and Propeller Twist Angles in the DNA of Four Drug–DNA Complexes[a]

	buckle (deg)				propeller twist (deg)			
base pair	e⁶G–SN6999	e⁶G–H258	e⁶G–H342	e⁶G–NET	e⁶G–SN6999	e⁶G–H258	e⁶G–H342	e⁶G–NET
C1–G24	8	−7	−15	−17	−4	−10	4	4
G2–C23	−5	−17	−9	−4	−17	−9	−19	−9
C3–G22	7	−4	−3	2	−6	12	−6	−4
G4–C21	6	−18	−22	−20	−2	6	2	−3
A5–T20	−11	−14	−1	3	−13	−12	−4	−6
A6–T19	1	−20	−11	−10	−10	−17	−26	−17
T7–A18	−11	0	6	7	−18	−3	3	−15
T8–A17	13	−2	4	4	−2	−6	−5	−6
C9–G16	5	12	24	−9	−8	−21	−24	−8
G10–C15	4	−5	−12	−4	−4	−10	−9	−11
C11–G14	5	−8	−5	6	−14	−15	4	−7
G12–C13	−15	6	−9	6	−10	−4	2	−9

[a] Key: e⁶G–SN6999, CGC[e⁶G]AATTCGCG + SN6999 complex; e⁶G–H258, CGC[e⁶G]AATTCGCG + Hoechst33258 complex; e⁶G–H342, CGC[e⁶G]AATTCGCG + Hoechst33342 complex; e⁶G–NET, CGC[e⁶G]AATTCGCG + netropsin complex.

Table III: Distances (Å) between SN6999 and DNA (<3.5 Å)[a]

SN6999	DNA	distance (Å)	SN6999	DNA	distance (Å)
P ring			BQ ring		
CM	G 22 O4′	3.24	C1	T 07 O4′	3.50
CM	C 23 O4′	3.17	C1	T 20 O4′	3.29
CM	G 12 O3′	3.45	C4	T 20 O4′	3.35
N1	G 22 O4′	3.09	C5	T 19 O2	2.44
C2	G 22 C5′	2.89	C5	T 20 O4′	2.88
C2	G 22 C4′	2.69	C6	T 19 O2	2.93
C2	G 22 O4′	2.53	C6	T 20 O4′	2.84
C3	G 22 C5′	3.06	Q ring		
C3	G 22 C4′	3.47	CM	C 09 O4′	2.97
C3	G 22 O4′	3.31	CM	C 09 O2	3.17
C5	G 04 N2	3.21	N1	C 09 O4′	3.19
C6	G 04 N2	2.78	C2	T 08 O2	2.95
BP ring			C2	A 18 C2′	3.09
C2	C 21 O4′	3.32	C2	A 18 N3	3.15
C3	C 21 C4′	3.04	C3	A 18 C2	3.45
C3	C 21 O4′	2.86	C7	C 09 C4′	3.30
C4	C 21 O4′	2.92	C8	C 09 C5′	3.33
C5	A 05 C2	3.31	C8	C 09 C4′	2.94
C5	A 06 N3	3.10	C8	C 09 O4′	2.95
C5	C 21 O4′	3.43	C9	C 09 C5′	3.27
C6	A 06 O4′	3.48	C9	C 09 C4′	3.39
peptide			C9	C 09 O4′	3.18
CP#	C 21 C5′	3.47	C10	C 09 C5′	3.41
OP#	C 21 C4′	3.30			
NP#	T 20 O2	3.11*			

[a] Key: *, hydrogen bond; CM, methyl carbon atom; # peptide atom.

(Table III). Three of the four rings of SN6999, Q, BQ, and BP have their edges facing toward the bottom of the minor groove. Therefore, the *N*-methyl group is near C9O$^{4'}$ and C9O^{2}, and the H^{2}, H^{3} hydrogen atoms of ring Q are near the A18H$^{2'}$, T8O^{2}, and A18C^{2} atoms, respectively. Similarly, the H^{5}/H^{6} atoms of both BQ and BP rings approach the A5-T20 and A6-T19 base pairs. Interestingly, ring P turns sharply so that the face of the ring approaches the edge of the e^{6}G4–C21 base pair. The only direct hydrogen bond involves the amide NH group with O^{2} of T20 (3.11 Å). The two amine linkers (between rings BP and P and rings Q and BQ), which are presumably protonated and positively-charged at neutral pH, do not participate in any direct hydrogen bond. Finally the two quaternary nitrogen atoms on rings Q and P are not near any of the negatively-charged phosphate atoms. Instead, they are close to the O$^{4'}$ atom of C9 and G22 deoxyriboses, respectively, possibly having charge-induced dipole interactions.

Figure 7 summarizes the interactions between SN6999 and DNA in the minor groove. As previously observed in the complexes of other minor groove binding drugs with DNAs (Wang & Teng, 1990), the binding of the SN6999 to DNA is stabilized by same kinds of forces including electrostatic attraction between the positively-charged quaternary ammonium drug and the negatively charged DNA, van der Waals interactions between nonpolar groups on the drug and the DNA sugar atoms along the two walls of the minor groove, hydrogen bond between the drug amide NH and the T–O^{2} of DNA, and stacking interaction between the drug aromatic rings and sugar O$^{4'}$ of DNA. The specificity toward the AT sequence is enhanced by the natural tendency of AT sequences having a narrow minor groove which provides a favorable environment to have the above-described interactions. Finally, the exocyclic N^{2} amino group in guanine presents a steric hindrance toward the drug and prevents the drug from adhering to the floor of the minor groove, and this would diminish the binding interactions substantially. Therefore, SN6999, like other minor groove binders, binds preferentially to the AT

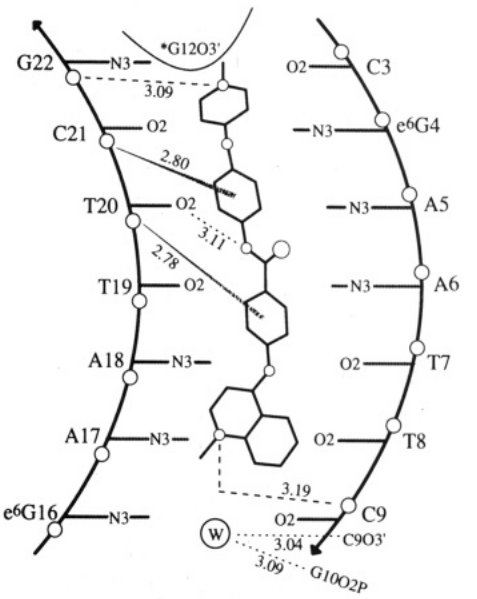

FIGURE 7: Schematic diagram showing the interactions between the drug and DNA in the e^{6}G–SN6999 structure. The drug molecule is sandwiched in the minor groove at the C[e^{6}G]AATT site between the two antiparallel backbones of the B-DNA. Van der Waals interactions (shown as elongated shaded triangles), such as the dipole-π interaction between the O$^{4'}$ (e.g., from sugars of T20 and C21 in the complex) and the aromatic rings of the drug are used to stabilize the binding along with hydrogen bonds (shown as dotted lines) and charge-induced dipole interactions (shown as dashed lines). The O$^{4'}$are shown as open circles on the DNA strand.

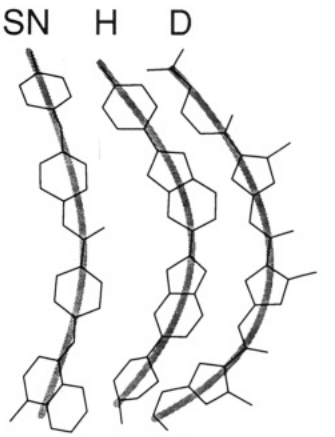

FIGURE 8: DNA-bound conformation (derived from crystal structures) of three minor groove binding drugs, SN6999 (SN), Hoechst 33258 (H), and distamycin (D) are shown. The three drugs were visually aligned, and the marked reduction in curvature in SN6999 from the crescent shape found in Hoechst 33258 and distamycin is obvious. The broad shaded line emphasizes the curvature of each drug.

sequence over GC sequence. While the SN6999 drug actually covers six base pairs, at least four contiguous AT sequence are needed for tight binding by the triply-linked Q, BQ, and BP rings. The positively charged P ring may extend, without steric hindrance, into the neighboring G–C base pairs which have a somewhat wider minor groove width. As the overall curvature of SN6999 is much less than that seen in the crescent shapes of other minor groove binding drugs like Hoechst 33258 or distamycin (Figure 8), and also due to the rather constrained conformation of the drug, the DNA duplex has to bend to provide a suitable minor groove pocket for drug binding. Figure 9 clearly shows the bend in helix axis caused by the straightening of the minor groove to accommodate SN6999 relative to that seen for binding Hoechst 33258 which is

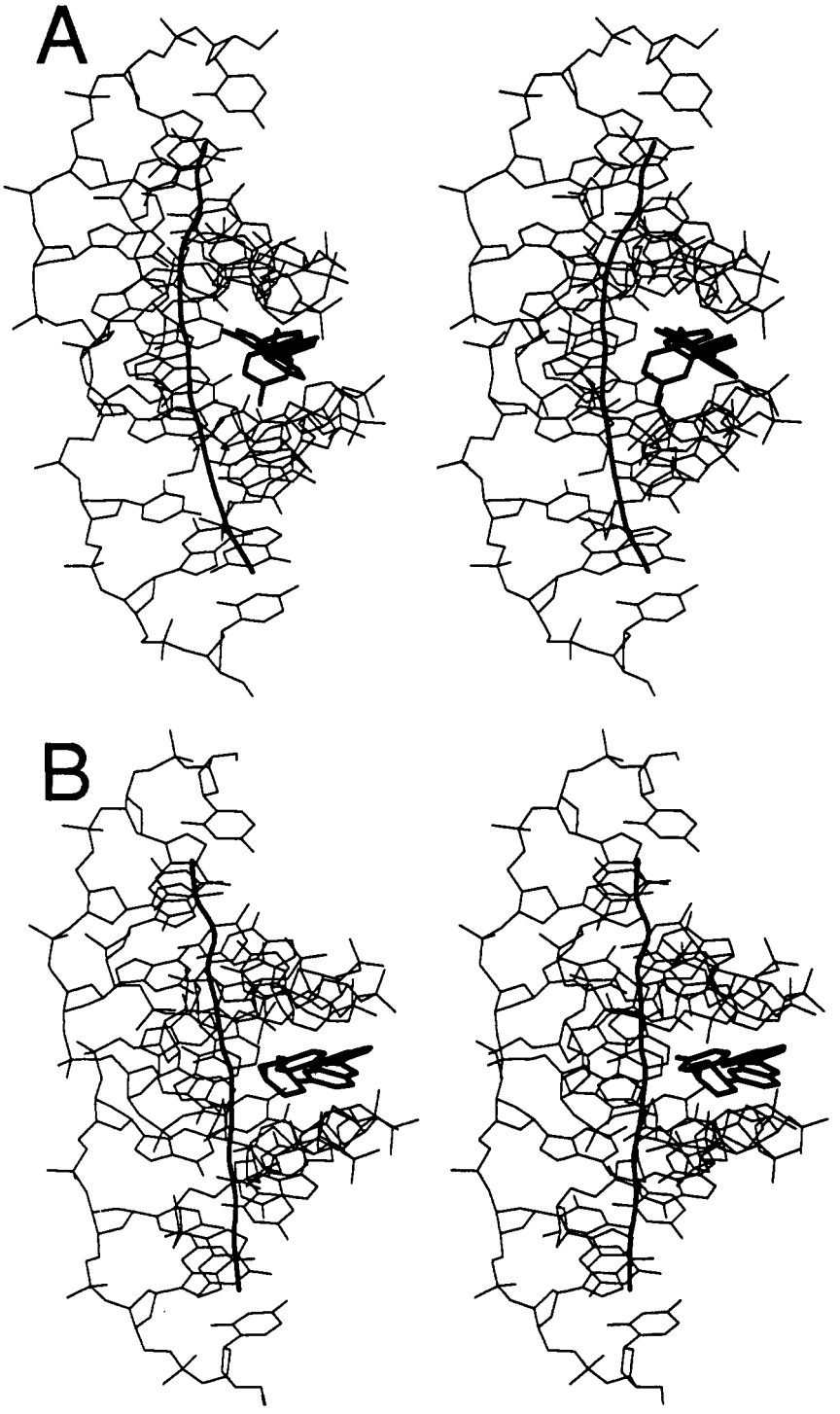

FIGURE 9: View of DNA–drug complexes down the minor groove. The helix axis as calculated by CURVE (Lavery & Sklenar, 1989) and the drug are shown in thick lines. (A) CGC[e^6G]AATTCGCG–SN6999 complex. (B) CGCGAATTCGCG–Hoechst 33528 complex (Teng *et al.*, 1988). The reduced curvature of SN6999 seems to strain the DNA as it compensates to adjust its minor groove, and an overall curvature of the helix results. In another view (not shown), at 90° to the present one, the two helix axes look very similar, and this is preserved to allow the packing in the crystal lattice using the G–N^2:G–N^3 interlocking base pairs at both ends of the helix.

crescent shaped and lodges in the minor groove without the DNA having to obviously adjust conformation.

The NMR studies of SN6999 with the self complementary duplex d(GCATTAATGC)$_2$ by Leupin *et al.* (1986) have established, using intermolecular NOE's between SN6999 and DNA, that the drug binds in the minor groove at the AT region. They observed fast exchange for most protons over a temperature range from 277–313 K, implying that the lifetime of drug in the DNA binding site is short relative to the NMR time scale. They proposed a model where SN6999 binds to two equivalent sites of ~5 AT base pairs, and an exchange exists between these two binding sites. In a more recent NMR study by Chen *et al.* (1992), a non-self-complementary DNA sequence of d(GGTTAATGCGGT)–d(ACCGCATTAACC) complexed 1:1 with SN6999 was scrutinized. They observed that a six base pair region with a 3′ overhang in either direction, specifically the d(T_3T_4-$A_5A_6T_7G_8$)–d($A_{18}T_{19}T_{20}A_{21}A_{22}C_{23}$) region of the DNA duplex, interacts with SN6999. They could unambiguously assign some intermolecular NOEs, including those from the SN6999 H^2 proton of Q ring to $A_{22}H^2$ as well as $A_{18}H^2$ of DNA and from the P ring's $H^{2,6}$ to both $A_{22}H^2$ as well as

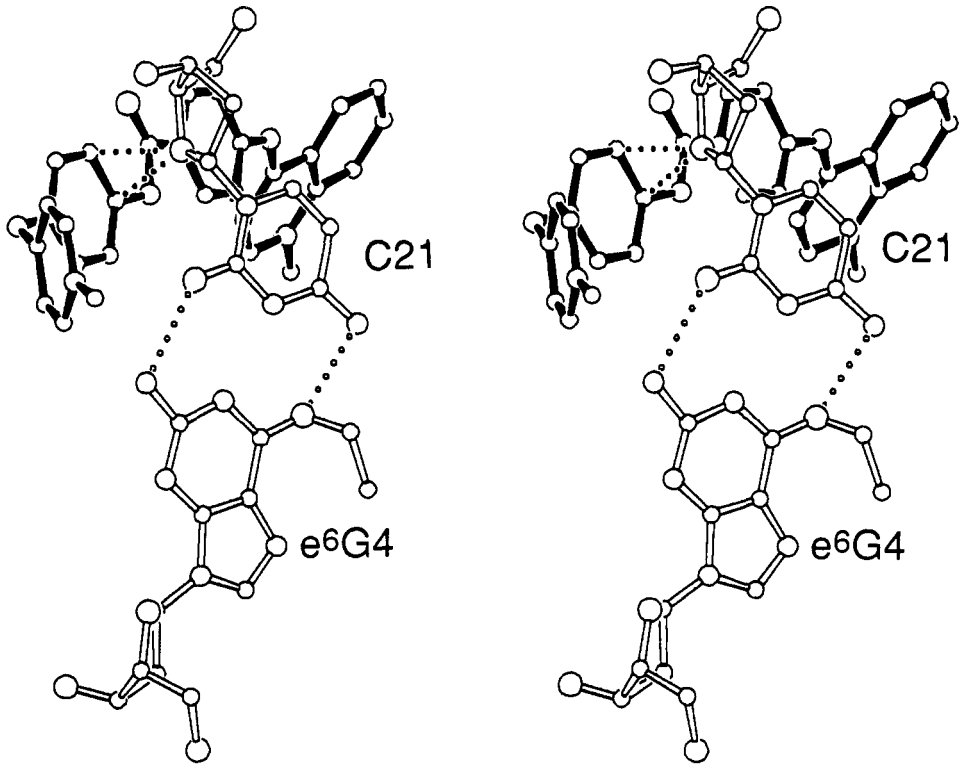

FIGURE 10: Detailed drawing in stereo, showing the local interaction of e^{6}G4:C21 base pair with SN6999. The hydrogen bonds of e^{6}G4:C21 are shown as large dots. The stacking intercation of C21O$^{4'}$ onto the BP ring of SN6999 is shown as small dots from C21O$^{4'}$ to C^4 and C^5 atoms of BP ring which are at a distance of 2.92 Å and 2.86 Å respectively.

$A_{18}H^2$ of DNA. Again, this intermolecular NOE set could be explained by a model where the drug binds along the minor groove of DNA in two orientations and rapidly exchanges between the two orientations. The present crystal structure supports and enhances the NMR studies and shows that a core of at least four AT base pairs is necessary for drug binding and that the drug spans a total of six base pairs. The crystal structure shows one orientation for the drug, and this could be due to a DNA alkylation/sequence effect. It would be interesting to know whether in solution the drug binds in two orientations to the AT site on alkylated DNA, as has been the case with two other nonalkylated DNA sequences studied by NMR.

Other derivatives related to SN6999 have also been synthesized and used in crystallization experiments. For example, SN7167 has two additional amino groups, one each on the Q and BP rings respectively (Figure 1). This compound is expected to bind to DNA in a manner very similar to that of SN6999, since the two amino groups would be pointing away from DNA. In fact, the amino groups may improve the aqueous solubility of the drug, and as a SN7167–DNA complex the drug amino groups may act as sites of recognition for protein binding along the minor groove.

e^{6}G:C Base Pairs. In our previous studies (Sriram *et al.*, 1992a,b), we have shown that an e^{6}G:C base pair may adopt different configurations such as wobble base pair or bifurcated base pair with two/one set(s) of three-centered hydrogen bonds, depending on the environments (see Figure 1 of Sriram *et al.* (1992b) for the definition of various base pair configurations). In the previous three complexes, the e^{6}G4:C21 base pair adopts a bifurcated configuration (Figure 1B of Sriram *et al.* (1992b)). In the present structure, this base pair adopts a new reverse wobble configuration (Figure 2). The C21 base is pushed away from the minor groove by the SN6999, and only two hydrogen bonds remain. The N^4 of C21 is 2.59 Å from the O^6 of e^{6}G4, and the N^2 of e^{6}G4 is 2.87 Å from the O^2 of C21. There is no hydrogen bond between N^3 of C21 and N^1 of e^{6}G4, despite the distance of 2.70 Å, because neither nitrogen atom has a hydrogen atom attached. The geometry of the base pair suggests that it is an unstable configuration in terms of hydrogen bonding but is stabilized by good stacking with neighboring base pairs. This unusual base pair is due to the close contact between the deoxyribose ring of C21 and the BP ring of SN6999 which is more clearly shown in Figure 10. The distance of C21O$^{4'}$ to the plane of ring BP is 2.80 Å. If the base pair were in the wobble or bifurcated configuration, there would be severe van der Waals clashes between the C21 residue (sugar and base) with the BP and P rings of SN6999. The inherently less stable e^{6}G:C base pair may allow the base pair to more easily accommodate the distortion due to SN6999 binding.

In the earlier three complexes (e^{6}G_DNET, e^{6}G_DH258, e^{6}G_DH342), the e^{6}G16:C9 base pair adopts the "wobble" pairing configuration, but in e^{6}G_SN6999 it adopts the bifurcated base pairing configuration. The N^4 of C9 is 2.69 and 2.63 Å from the N^1 and O^6 of e^{6}G16, respectively, and the N^2 of e^{6}G16 is 2.31 and 2.87 Å from the N^3 and O^2 of C9, respectively. One of these hydrogen bonds is too short, and this is probably due to the lack of hydrogen bond constraint during refinement.

The ethyl groups of the two e^{6}Gs lie in the major groove of the helix, and they are both out of the plane (by an average of ~0.63 Å) from the respective guanine bases (Figures 2 and 3). They point toward the opposite ends of the helix and make contact with the neighboring N^4/C^5 of C3 and C15 cytosines, respectively. Both e^{6}G bases have their ethyl groups in the *proximal* orientation (to N^7), which effectively blocks access of solvent, metal, or protein to the N^7 position of e^{6}G. The distance between the N^7 atom and the C_β atom of e^{6}G (C_α is the methylene carbon and C_β is the methyl carbon of

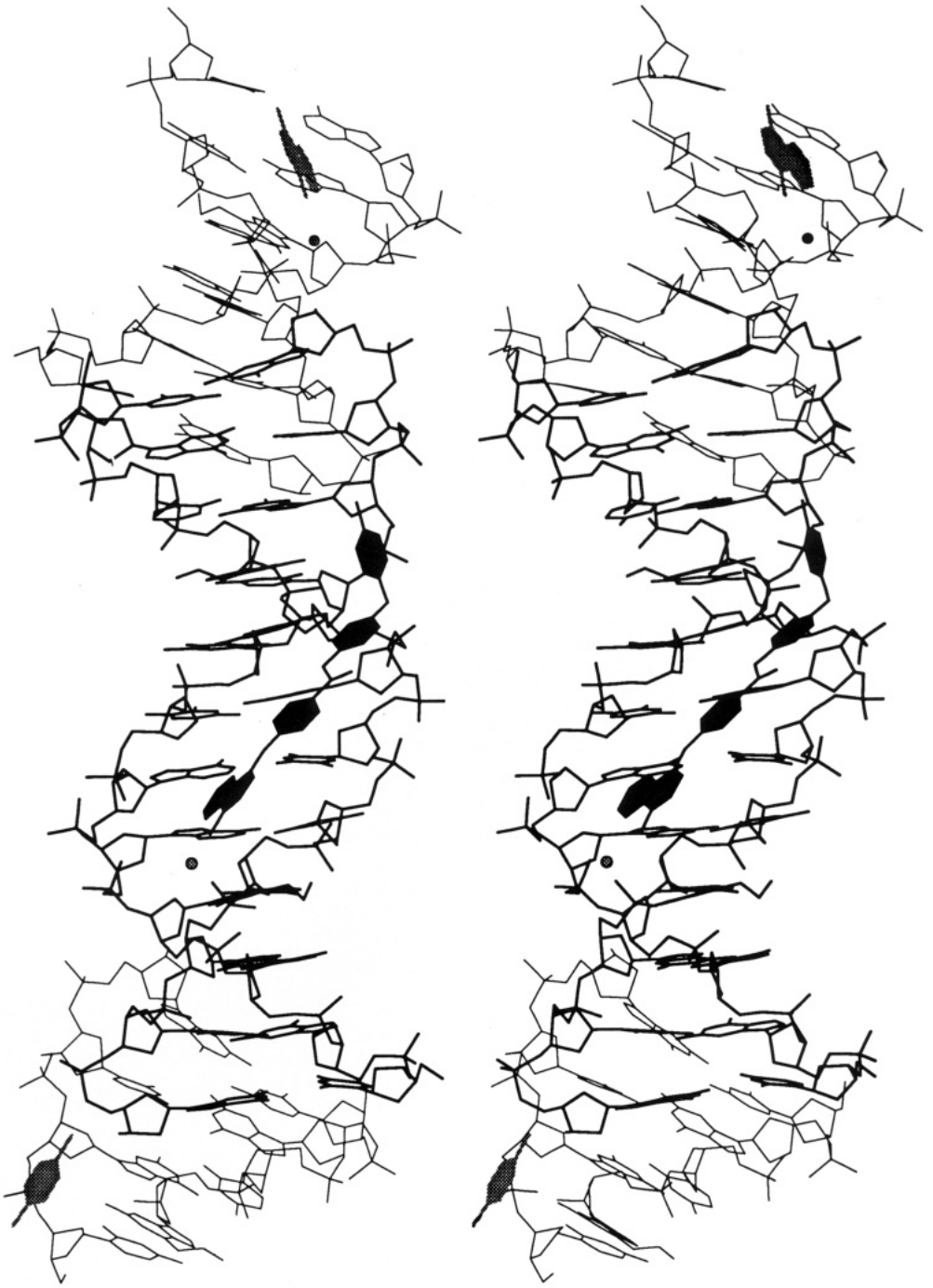

FIGURE 11: Stereoscopic skeletal illustration showing the extent of minor groove available for drug binding. The G–N^2:G–N^3 interlocking base pairs at both ends of the helix occlude a good portion of the minor groove of the dodecamer. Only approximately seven base-pair steps are accessible along the minor groove. The drug and one water molecule located in the minor groove occupy six of the seven base pairs. The symmetry-related complex fragments are shown in thin lines and gray shading.

the ethyl group) is 2.75 Å for e^6G4 and 3.13 Å for e^6G16. The torsion angle about the O^6–C_α bond is in the *synclinical*$^+$ conformation (70°) for e^6G4 and in the *anticlinical*$^+$ conformation (106°) for e^6G16.

Our results suggest that e^6G:C base pair configuration is polymorphic. In B-DNA it may adopt the wobble configuration (Figure 1B in Sriram *et al.* (1992b)) or the bifurcated pairing configuration (Figure 1A in Sriram *et al.* (1992b)) in the earlier structures of e^6G DNET, e^6G H258, and e^6G H342. In e^6G SN6999, the bifurcated pairing is again seen for e^6G16:C9. However, a new reverse wobble pairing is found for e^6G4:C21 in which the cytosine is pushed toward the major groove side. Both distorted conformations, the wobble as well as the reverse wobble, associated with the e^6G:C base pair may be recognized by the appropriate repair enzymes. However, it has been noted that among the polymorphic configurations the bifurcated pairing (Figure 1A in Sriram *et al.* (1992b)) is not very different from the normal Watson–Crick G:C configuration, making it more difficult for the repair enzymes to recognize the lesion site. This could explain the relative persistence of the e^6G lesion as well as why C can still be incorporated in the daughter strand opposite to the e^6G lesion during replication, in which either C or T may be inserted in the daughter DNA strand across the e^6G site. However, thymine can pair with e^6G in only one way, with a configuration similar to a regular Watson–Crick G:C base pair (Figure 1C in Sriram *et al.* (1992b); Leonard *et al.*, 1990). This may be a plausible explanation of why thymine is found prefer-

entially incorporated across the e^6G/m^6G lesion site during replication (Loechler *et al.*, 1984).

Three Crystal Lattices. The e^6G SN6999 complex crystallized in a new crystal lattice. The reason for adopting a new lattice is because the e^6G SN6999 complex cannot pack in the previous $P2_12_12_1$ lattice due to a relatively bent DNA duplex (Figure 5). Interestingly, the new molecule uses a nearly identical helix–helix interlocking mechanism, as found in numerous dodecamers, to form an infinitely long column of complexes along the *c* axis direction. This is shown in Figure 11. It can be seen that the terminal two base pairs of the neighboring molecules intrude into the minor grooves of the helix, leaving only a space of ~7 base pairs for the binding of SN6999. In fact, the methyl group of the P ring is in close contact with the $G12O^{3'}$ of the neighboring helices.

This type of interlocking interactions has now been observed in three different crystal lattices, all in the space group $P2_12_12_1$. These include many d(CGCGAATTCGCG)-related dodecamers (Kopka & Larsen, 1992; Wang & Teng, 1990), a gapped dodecamer d(CGCGAAAACGCG) + d(CGCGTT) + d(TTCGCG) (Aymami *et al.*, 1990), and the present structure. The ubiquitous occurrence of this G:G interlocking interaction suggests that it may occur elsewhere. In fact, if one focuses only on the last two base pairs of the neighboring helix, which occupy the minor groove of the parent helix, one can see that it would serve as a model for the stacked binding of the minor groove binder, not unlike that of distamycin by Pelton & Wemmer (1989, 1990). The specific donor (N^2) and acceptor (N^3) disposition of a guanine base may be used for the synthesis of minor groove binding drugs that will recognize a guanine base with a mode similar to the guanine–guanine interlocking interaction.

CONCLUSION

We have analyzed four complexes of a DNA containing e^6G lesions with the minor groove binding drugs (SN6999, Hoechst 33258, Hoechst 33342, and netropsin) (this work and Sriram *et al.*, 1992a,b) which have shown the multiplicity of the manner in which the carcinogenic e^6G pairs with cytosine. The base pairing scheme adopted can be the wobble, the bifurcated, or the reverse wobble hydrogen bond pairing, depending on the local environment. It is possible that for e^6G:C base pair there is a dynamic equilibrium among the various conformers in solution. This may be addressed by careful measurement of the relaxation time of the protons associated with the e^6G:C base pair by nuclear magnetic resonance. Finally, the helix with e^6G lesions next to AATT sequence is stabilized by the minor groove binding drug SN6999, offering the first visualization of how SN6999 interacts with B-DNA at the molecular level. These results should help us understand how other SN-series of quaternary nitrogen compounds may bind to DNA.

ACKNOWLEDGMENT

We thank Drs. G. A. van der Marel and J. H. van Boom for the e^6G-DNA molecule.

REFERENCES

Aymami, J., Coll, M., van der Marel, G. A., van Boom, J. H., Rich, A., & Wang, A. H.-J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2526–2530.

Cain, B. F., Atwell, G. J., & Seelye, R. N. (1969) *J. Med. Chem. 121*, 199–206.

Chen, S.-M., Leupin, W., Rance, M., & Chazin, W. J. (1992) *Biochemistry 31*, 4406–4413.

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.

Coll, M., Sherman, S. E., Gibson, D., Lippard, S. J., & Wang, A. H.-J. (1990) *J. Biomol. Struct. Dyn. 8*, 315–330

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

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

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekman, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., Wang, A. H.-J., & Zhurkin, V. B. (1989) *Nucleic Acids Res. 17*, 1797–1803.

Ginell, S. L., Kuzmich, S., Jones, R. A., & Berman, H. M. (1990) *Biochemistry 29*, 10461–10465.

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

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

Kalnik, M. W., Li, B. F. L., Swann, P. F., & Patel, D. J. (1989a) *Biochemistry 28*, 6170–6181.

Kalnik, M. W., Li, B. F. L., Swann, P. F., & Patel, D. J. (1989b) *Biochemistry 28*, 6182–6192.

Kopka, M. L., & Larsen, T. A. (1992) In *Nucleic Acid Targeted Drug Design* (Propst, C. L., & Perun, T. J., Eds.) pp. 303–374, Marcel Dekker, Inc., New York.

Lavery, R., & Sklenar, H. (1986) *J. Biomol. Struct. Dyn. 6*, 655–667.

Leonard, G. A., Thomson, J., Watson, W. P., & Brown, T. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9573–9576.

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

Loechler, E. L., Green, C. L., & Essigmann, J. M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 6271–6275.

Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E., & Dervan, P. B. (1992) *Proc. Natl. Acad. Sci. USA 89*, 7586–7590.

Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., & Jones, R. A. (1986) *Biochemistry 25*, 1027–1036.

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. Am. Chem. Soc. 112*, 1393–1399.

Plowman, J., & Adamson, R. H. (1978) *Pharmacology 17*, 61–68.

Rabinowich, D., & Shakked, Z. (1984) *Acta Crystallogr. A40*, 195–200.

Remers, W. A., Barkley, M. D., Hurley, L. H. (1992) In *Nucleic Acid Targeted Drug Design* (Propst, C. L., & Perun, T. J., Eds.) pp. 375–422, Marcel Dekker, Inc., New York.

Robertson, I. G. C., & Baguley, B. C. (1982) *Eur. J. Cancer Clin. Oncol. 18*, 271–279.

Roelen, H. C. P. F., Brugghe, H. F., van der Elst, H., Klien, J. C., van der Marel, G. A., van Boom, J. H., & Wang A. H.-J. (1992) *Recl. Trav. Chim. Pays-Bas 8*, 227–234.

Sriram, M., van der Marel, G. A., Roelen, H. C. P. F., van Boom, J. H., & Wang, A. H.-J. (1992a) *EMBO J. 11*, 225–232.

Sriram, M., van der Marel, G. A., Roelen, H. C. P. F., van Boom, J. H., & Wang, A. H.-J. (1992b) *Biochemistry, 31*, 11823–11834.

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

Wang, A. H.-J., & Gao, Y.-G. (1990) *Methods 1*, 91–99.

Wang, A. H.-J., & Robinson, H. (1992) In *Nucleic Acid Targeted Drug Design* (Propst, C. L., & Perun, T. J., Eds.) pp. 17–64, Marcel Dekker, Inc., New York.

Wang, A. H.-J., & Teng, M.-K. (1990) In *Crystallographic & Modeling Methods in Molecular Design* (Bugg, C. E., & Ealick, S. E., Eds.) pp. 123–150, Springer-Verlag, New York.

Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.