2′-Deoxyisoguanosine Adopts More than One Tautomer To Form Base Pairs with Thymidine Observed by High-Resolution Crystal Structure Analysis†

Howard Robinson,‡ Yi-Gui Gao,‡ Cornelia Bauer,§ Christopher Roberts,|| Christopher Switzer,*,|| and Andrew H.-J. Wang*,‡,§

Department of Cell and Structural Biology and Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, and Department of Chemistry, University of California at Riverside, Riverside, California 92521

Received April 13, 1998; Revised Manuscript Received June 11, 1998

ABSTRACT: The questions of whether different tautomeric forms of nucleic acid bases exist to any significant extent in DNA, or what their possible roles in mutation may be, are under intense scrutiny. 2′-Deoxyisoguanosine (iG) has been suggested to have a propensity to adopt the enol form. Isoguanine (also called 2-hydroxyadenine) can be found in oxidatively damaged DNA generated from treating DNA with a Fenton-type reactive oxygen-generating system and is known to cause mutation. We have analyzed the three-dimensional structure of the DNA dodecamer d(CGC[iG]AATTTGCG) (denoted iG-DODE) by X-ray crystallography and NMR. The crystal structure of the iG-DODE complexed with the minor groove binder Hoechst 33342, refined to 1.4 Å resolution, showed that the two independent iG·T base pairs in the dodecamer duplex adopt different (one in Watson−Crick and the other in wobble) conformations. The high-resolution nature of the structure also affords unprecedented clear information about the conformation and interactions of the Hoechst drug. The Hoechst 33342 binds in the narrow minor groove at the iGAATT site, with the *N*-methylpiperazine ring near the iG4·T21 base pair. Three hydrogen bonds are found between the NH of the Hoechst ligand and T-O2 DNA atoms. In solution, the two iG·T base pairs in iG-DODE predominantly are in the wobble form at 2 °C. At higher temperatures, another duplex form (likely involving the enol form of iG) is in slow exchange with the keto form and becomes significantly populated, reaching ∼40% at 40 °C. Our data support the conclusion that iG pairs with T in a Watson−Crick configuration to a significant extent at physiological temperature (37 °C), which may explain the facile incorporation rate of T across from an iG during in vitro DNA replication.

It has long been debated whether different tautomeric forms of nucleic acid bases exist to any significant extent in DNA (*1*, *2*). Both guanine and thymine have been suggested to have the potential of forming an "enol" minor tautomeric form. Indeed, the mutation rate of $10^{-4}-10^{-6}$ associated with guanine or thymine during DNA replication has been taken to imply that their enol forms may transiently exist and form a Watson−Crick base pair with thymine or guanine, respectively, thereby causing a transition mutation. However, attempts to directly detect such tautomers have only been successful in the gas phase (*3*, *4*).

To address this question, we have studied a nonstandard nucleoside, 2′-deoxyisoguanosine (iG[1]), which has been suggested to have a propensity to adopt the enol form (Figure 1). Isoguanine is an interesting alternative nucleic acid base. Early on, Shugar and colleagues (*5*) examined the properties of iG using optical methods and concluded that iG favors the keto (N1-H) form in aqueous solution, whereas it favors the enol form in nonpolar solution and at high temperatures. Nonetheless, the evidence for the enol (O2-H) form was not unequivocal. Further, the existence of an iG N3-H tautomer has been proposed (*6*). More recently, this issue has been re-examined by Seela and colleagues (*7*), where the general findings made by Shugar and colleagues were confirmed on the basis of ^{13}C NMR spectroscopy.

It was also noted by Seela and colleagues that iG is a potent inducer for the formation of parallel DNA duplex structure (*8*, *9*). Significantly, it has been shown that iG- and iC-containing DNA oligonucleotides can form remarkably stable parallel-stranded duplexes with the complementary G- and C-containing DNA or RNA strands (*10*). The T_m of the PS duplex is comparable with that of the corresponding antiparallel duplex even for a 10mer (*10*). The structure of an iG- and iC-containing parallel-stranded duplex has been analyzed by NMR spectroscopy (*11*). More recently, DNA molecules containing several consecutive iG's

† This work was supported by NIH grants to A.H.-J.W. (GM41612) and to C.S. (GM47375) and a NASA grant (NAGW-4184) to C.S. The Varian VXR500 NMR spectrometer at the University of Illinois at Urbana-Champaign was supported in part by NIH Shared Instrumentation Grant 1S10RR06243.

* To whom correspondence should be addressed.

‡ Department of Cell and Structural Biology, University of Illinois at Urbana-Champaign.

§ Department of Chemistry, University of Illinois at Urbana-Champaign.

|| University of California at Riverside.

[1] Abbreviations: iG or iso-G, isodeoxyguanosine, sometimes called 2-hydroxy-2′-deoxyadenosine; e^6G, O^6-ethylguanine; NMR, nuclear magnetic resonance; 2D NOESY, two-dimensional nuclear Overhauser effect spectroscopy; H258, Hoechst 33258; H342, Hoechst 33342; iG-DODE−H342, CGC[iG]AATTCGCG−H342 complex.

S0006-2960(98)00818-6 CCC: $15.00
Published on Web 07/11/1998

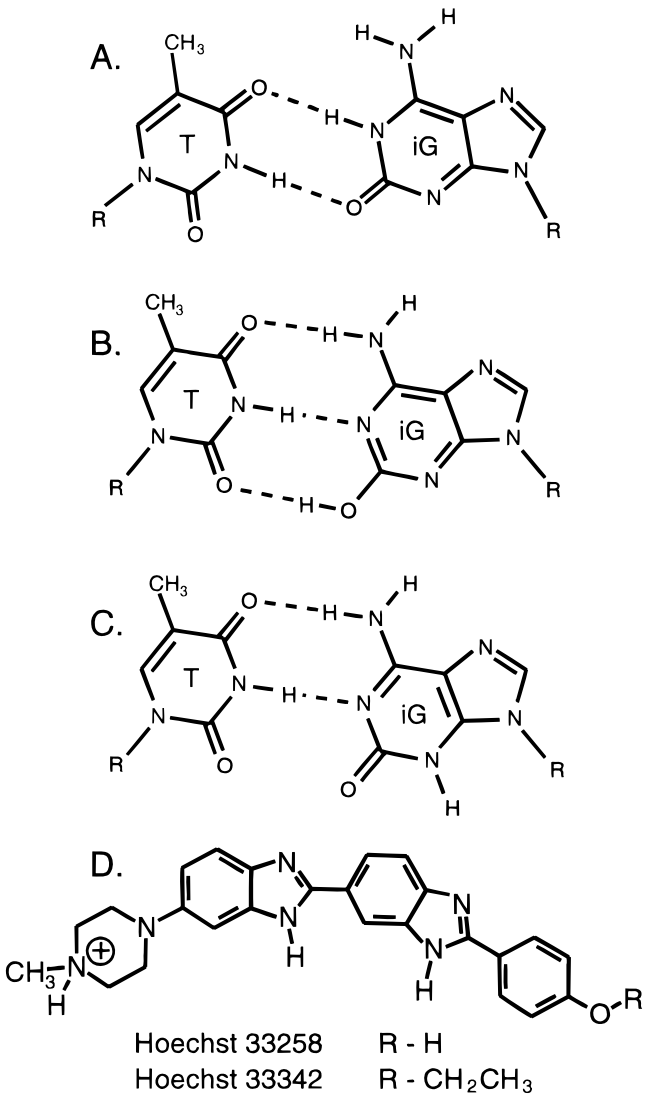

FIGURE 1: Possible configurations of iG·T base pairs with iG in different tautomeric forms: (A) wobble (iG in the N1-H or keto form), (B) Watson−Crick (iG in the O2-H or enol form), and (C) Watson−Crick (iG in the N3-H form) iG·T base pairs. The crystal structure of the keto form of iG has been solved using a N1-modified iG nucleoside (*29*). (D) Chemical structures of Hoechst 33258 and 33342.

have been demonstrated to form a very stable parallel tetraplex structure (*9*, *12*, *13*).

Interestingly, isoguanine (also called 2-hydroxyadenine) can be found in oxidatively damaged DNA generated from treating DNA with a Fenton-type reactive oxygen-generating system, e.g., Fe^{2+}−EDTA. The yield of the formation of iG in DNA is somewhat lower than that of the 8-hydroxyguanine. It is known that both 2-hydroxyadenine (i.e., iG) and 8-oxoguanine are mutagens. The repair mechanism of the lesions involving these two oxidized bases is under intense study (*14*−*16*). Incorporation of iG in DNA, or present as the nucleotide, results in the formation of incorrect base pairs during in vitro replication (*15*−*19*). Indeed, iG could be incorporated efficiently into a DNA duplex across from a T during in vitro DNA replication (*17*, *18*). Such data seemed to support the possibility that the enol tautomer of iG may form at physiological temperature and may pair with T in a Watson−Crick geometry (Figure 1B), thereby escaping the editing function of the polymerase.

Whereas the structural consequences due to the 8-oxoguanine incorporation have been studied (*20*, *21*), little is known about the structural effect of 2-hydroxyadenine (i.e., iG) in DNA. Therefore, it is important to understand the properties associated with the modified base 2-hydroxyadenine (i.e., isoguanine). To date, structural work on iG has dealt with this issue only at the nucleoside level. How isoguanosine behaves in a DNA duplex remains largely unknown.

In this paper, we have analyzed, by X-ray crystallography and NMR, the three-dimensional structure of an iG-containing DNA dodecamer d(CGC[iG]AATTTGCG) (denoted iG-DODE), wherein iG is base paired with T in the duplex. The refined crystal structure of the iG-DODE complexed with the minor groove binder Hoechst 33342, using data to 1.4 Å resolution, revealed that the two independent iG·T base pairs in the dodecamer duplexes adopt different (one in Watson−Crick and the other in wobble) conformations, as well as an unprecedented clear view of the binding interactions of the Hoechst ligand with DNA. Our results represent a direct visualization of a Watson−Crick DNA base pair incorporating a minor tautomeric form of a DNA base without exocyclic chemical modifications (e.g., methylation). The structural data provide a basis for understanding iG·T pairing during in vitro DNA replication and may have a bearing on mutagenesis in vivo involving minor tautomers of the common nucleobases.

MATERIALS AND METHODS

The synthesis of the i-dG nucleoside phosphoramidite followed the published procedure (*22*). The modified dodecamer was synthesized using an ABI 391-EP automated DNA synthesizer. The molecule was purified by reverse phase HPLC using a Hamilton PRP-1 apparatus at 60 °C, with a gradient from 5% CH_3CN/triethylammonium acetate (100 mM, pH 7) to 15% CH_3CN over the course of 20 min (trityl-off purification), and the appropriate fraction was dried on a Speed Vac apparatus at 20 °C to give iG-DODE as a fluffy colorless powder: laser desorption ionization time-of-flight MS *m*/*e* 3569 (MH^+) (requires 3568).

X-ray Crystallographic Analysis. Crystallization was performed (*23*) using a solution containing 4 mM DNA duplex (2.5 μL), 4 mM Hoechst 33342 (2.5 μL), 100 mM sodium cacodylate buffer (pH 6.5) (10 μL), 100 mM magnesium chloride (2.5 μL), 15 mM spermine tetrachloride (2.5 μL), and 30% 2-methyl-2,4-pentanediol (2-MPD) (5.0 μL) which was allowed to equilibrate with the reservoir (30 mL of 40% 2-MPD). Crystals appeared after 2 weeks at room temperature. The crystallization box was transferred to a cold room (4 °C) and left there for several months. Suitable crystals were chosen and mounted on nylon loops for data collection at −150 °C using a Rigaku R-AXIS IIc image plate system mounted on a RU-200 rotating-anode X-ray generator (50 kV and 80 mA) with CuKα radiation (1.5418 Å). The crystal is in space group $P2_12_12_1$ with the following unit cell dimensions: $a = 25.77$ Å, $b = 41.10$ Å, and $c = 64.30$ Å. In contrast, only poor quality crystals could be obtained from solutions without the Hoechst ligand.

The image plate to crystal distance was 90 mm with the 2θ angle set at −20°. Sixty frames of oscillation data (3.0°/20 min) were collected to cover a range of 180° of the ϕ angle. After the data were processed using the Molecular

Structure Corp. (Woodlands, TX) software package, the data set produced a 3-fold redundancy in the observed reflections. Care was taken to account for intensity saturation on the image plate for strong reflections. Some frames were recollected with a faster oscillation speed (4°/2 min) and reduced X-ray intensity so those strong (saturated) reflections could be recovered. There are 11 505 reflections collected [$>1.0\sigma(F)$] to 1.4 Å resolution with an R_{merge} of 5.4%, completeness of 81.5%, and a Wilson *B*-factor of 13.3 Å^2.

Initial attempts to refine the structure using the atomic coordinates from either the Hoechst 33258−d(CGCGAATTCGCG) crystal structures (*24*) or the complex of SN6999 with the e^{6}G-modified dodecamer d(CGC[e^{6}G]AATTCGCG) crystal structure (*25*) were unsuccessful, indicating that the locations of the these dodecamers in their respective crystal lattices are different. The structure was finally solved by the molecular replacement program ULTIMA (*26*) using the Hoechst 33258−d(CGCGAATTCGCG) complex (*24*) as a search model. The structure was refined by the simulated annealing procedure as set up in X-PLOR (*27*). The force field parameters for standard DNA nucleotides were from Parkinson et al. (*28*) as implemented for X-PLOR. Those for the Hoechst 33342 were obtained from comparison with related structures in the Cambridge Crystal Database. The force field parameters for all three iG tautomeric forms (N1-H, O2-H, and N3-H forms) were obtained using the crystal structure of 1-allylisoguanosine (*29*) and ab intio calculations (unpublished data of C. Switzer).

Water molecules were then located from subsequent Fourier ($2|F_o| - |F_c|$) maps and added to the refinement using the procedure implemented in X-PLOR (*27*). The simulated annealing procedure in X-PLOR consisted of slowly warming from 100 to 1000 K in 25 K increments and then cooling by 25 K increments to 300 K, followed by 120 steps of conjugate gradient minimization. At this stage, the refinement was continued using the program SHELX-97 (*30*) with individual isotropic temperature factors. The procedure involving the refinement of 12 anisotropic scaling parameter as suggested by Parkin et al. (*31*), implemented in SHELX-97, was applied. The default weighting scheme for structure factors in the refinement was used.

The refinement statistics are as follows: 8608 reflections [$>2.0\sigma(I)$ between 10 and 1.4 Å resolution], *R*-factor and *R*-free values (5% data) of 0.220 and 0.296, respectively, rmsds of bond distances and angle distances for DNA of 0.009 and 0.031 Å, respectively, one $Mg(H_2O)_6$, and 122 water molecules. For all data between 10 and 1.4 Å resolution (11 253 reflections), the *R*-factor and *R*-free values (5% data) are 0.245 and 0.315, respectively.

As a reference, we also carried out a parallel refinement using only 1.9 Å resolution data and obtained the following refinement statistics: 4987 reflections [$>2.0\sigma(I)$ between 10 and 1.9 Å resolution] and *R*-factor and *R*-free values (5% data) of 0.175 and 0.282, respectively. Those refinement statistics indicated that our refinement result is significantly improved when compared to those of previous structures of the dodecamer−drug complexes which were refined at ∼2 Å resolution (vide infra). The relatively high *R*-free value is not unexpected, considering the fact that less than half of the water molecules were located and included in the refinement. Moreover, only one hydrated magnesium ion was found, indicating that the remaining cations (sodium, magnesium, and spermine) are disordered. Nevertheless, the good quality of the structure is reflected by the fact that the electron density envelope of the entire DNA and Hoechst molecules is completely continuous at the 1.0σ level. The final atomic coordinates of the structure and other relevant information have been deposited in the Nucleic Acids Database (accession number BDD001 for atomic coordinates of the X-ray structure, structure factors, and force field parameters) and Brookhaven Protein Data Bank (accession numbers 1bhr and rlbhrmr, respectively, for the atomic coordinates of the NMR structure and NOE constraints).

NMR Spectroscopy. Solutions of the DNA oligomer were prepared as described earlier (*32*). Lyophilized powder (3.4 mg) was dissolved in 0.55 mL of H_2O containing 20 mM phosphate buffer at pH 7.0, resulting in a 0.77 mM duplex solution. NMR spectra were collected on a Varian VXR500 500 MHz spectrometer, and the data were processed with FELIX v1.1 (Hare Research, Woodinville, WA). The nonexchangeable two-dimensional (2D) NOE spectra were collected at 2 °C with a mixing time of 100 ms and a total recycle delay of 5.4 s where the average T1 relaxation was 1.8 s. The data were collected by the States−TPPI technique (*33*) with 512 t_1 increments and 2048 t_2 complex points each the average of 24 transients. TOCSY (total correlated spectroscopy) spectra were used together with the NOE spectra to derive the assignment using the standard sequential assignment procedure. The chemical shifts of all assigned resonances are listed in Table 1S of the Supporting Information.

Two starting models were built. In the first model, the refined crystal structure was used and a 2-fold symmetry operation was then applied to the duplex model which was energy-minimized by conjugate gradient minimization using X-PLOR (*27*). X-PLOR's all-atom force field for DNA was used with explicit hydrogen bond potentials. Refinement of the starting model was carried out by the SPEDREF procedures (*32*). Minimization of the residual errors is performed by conjugate gradient minimization with the NOE constraints within the program X-PLOR (*27*) employing 40 cycles of SPEDREF refinement. The isotropic correlation time (t_c) for each refinement was empirically determined to be 7 ns. Another starting model was built using canonical B-DNA and was similarly refined. A family of converged models were obtained from the refinement with the NMR *R*-factor in the range of ∼18%. The refined structures based on the two starting models have a root-mean-square deviation of 1.1 Å between them, suggesting a reasonable convergence of the refinement process.

2D NOESY spectra at 2 °C in 90% H_2O were collected with the 1-not-1 pulse sequence (*34*) as the read pulse of the NOESY. Twenty-four transients were averaged with a recycle delay of 2.9 s and a mixing time of 100 ms. The excitation offset was set to $^1/_4$ of the spectral bandwidth which was set to 12 000 Hz so that the imino resonances around 13 ppm were maximally excited.

RESULTS AND DISCUSSION

Structure of the Complex. The 1.4 Å resolution structure of the Hoechst 33342−d(CGC[iG]AATTTGCG) complex (denoted iG-DODE−H342) is shown in Figure 2. The

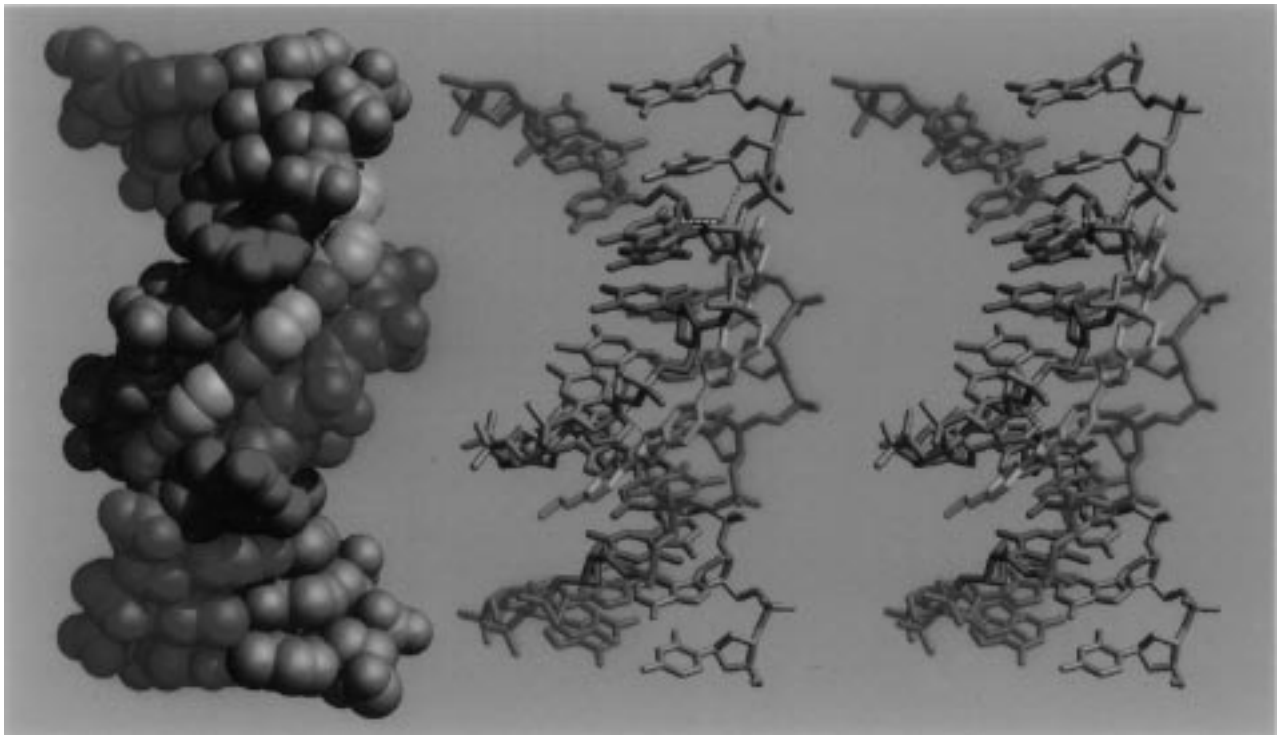

FIGURE 2: Crystal structure of the complex of Hoechst 33342 and iG-DODE. (Left) The van der Waals drawing. DNA is shown in green and light orange, except for the two iG·T base pairs shown in red, with the Hoechst 33342 (yellow) covering the central iGAATT base pairs. (Right) Thee stereoscopic skeletal drawing. The hydrogen bonds between the NH groups of the aromatic benzimidazoles of Hoechst 33342 and the AT base pairs are shown as white dotted lines. Note that the *N*-methylpiperazine of Hoechst 33342 is close to the iG4·T21 base pair with a bridging water molecule between them.

Table 1: Helical Parameters of DNA in the iG-DODE Complexed with Hoechst 33342[a]

base pair	roll	tilt	incl	propeller twist (ω)	buckle (κ)	helical twist (Ω)	rise (Å)
C1·G24	9.4	−1.6	0.1	−11.8	−0.7	35.9	3.61
G2·C23	−4.5	−1.5	−0.5	−18.0	−10.0	37.6	3.24
C3·G22	5.7	2.8	−1.5	−5.7	−8.9	29.5	4.00
iG4·T21	3.1	−3.6	0.1	−15.5	11.7	38.6	3.40
A5·T20	4.5	−3.5	−3.0	−20.4	4.8	34.9	3.36
A6·T19	2.2	2.4	−5.8	−21.6	−0.2	32.8	3.31
T7·A18	5.5	3.6	−4.7	−21.7	−4.0	34.9	3.44
T8·A17	−5.9	5.8	−3.3	−15.5	−8.8	34.7	3.18
T9·iG16	8.3	−2.0	0.5	−15.0	−9.6	35.1	3.46
G10·C15	−11.8	−2.2	0.0	−6.5	3.8	41.5	3.27
C11·G14	2.2	5.6	−1.5	−18.8	−3.8	36.8	3.34
G12·C13	–	–	1.9	−10.6	−3.8	–	–

[a] These parameters are calculated using the program Curves 3.0 (*64*).

Hoechst ligand is bound in the narrow minor groove of the B-DNA duplex at the iGAATT site as seen in another Hoechst−DNA complex (*35*), but different from those found in other Hoechst−DNA complexes (*24*, *36*).

The overall structure of the dodecamer DNAs in these two complexes is similar to related dodecamers (*24*, *37*−*40*). The helical parameters (including the base pair buckle and propeller twist angles) and individual torsion angles of the iG-DODE helix are listed in Table 1 and Table 2S of the Supporting Information, respectively. While most of the sugars in iG-DODE−H342 have the *S*-type pucker (with the pseudorotation angle *P* ranging from 92.0 to 188.4°), some have the *N*-type (C3, C11, and G12 with their *P* values of 41.1, 30.1, and 12.0°, respectively). All phosphates are in the B_I conformation (the $\epsilon-\zeta$ values being in the range of −33.7 to −98.7°), except for the G10, iG16, and G22 nucleotides which have B_{II} conformations. The duplex has a characteristic narrow minor groove at the AATT region. However, there are changes in the DNA conformation in these complexes relative to that of the canonical AATT dodecamer (*37*), distributed throughout the helix, presumably due to the insertion of the two iGs, the binding of the Hoechst drug, different crystal lattices, or a combination thereof. The root-mean-square deviation of the structures between iG-DODE−H342 and the canonical AATT dodecamer is 1.12 Å.

In most of the CGCGAATTCGCG and related dodecamer structures, the base pairs in the central AT region have been found to have large propeller twist angles which result in the bifurcated hydrogen bonds from the N6 amino group of an adenine simultaneously to the O4 atoms of two thymines in the opposite strand (one from the Watson−Crick mate and the other from its adjacent 5′ T) (*38*). In the iG-DODE−H342 structure, all four A·T base pairs maintain large propeller twists (average of −19.8°). In fact, most of the base pairs in the helix have either a buckle or a propeller twist angle greater than 10°, with the exception of G3·C22 and G10·C15 base pairs. The two G·C base pairs at both ends of the helix are involved in the interlocking lattice interactions using the G14·G24* and G12·G23* (* stands for a symmetry-related duplex) hydrogen bonding pairing in the minor groove (Figure 1S in the Supporting Information). As noted previously (*41*), this type of G·G pairing is associated with a large dihedral angle between the two guanines. This may impose conformational distortion in the participating (terminal and penultimate) base pairs. For example, two penultimate G2·C23 and C11·G14 base pairs

have large propeller twist angles ω (−18.0 and −18.8°, respectively), which are unusual for G·C base pairs.

The packing of the complexes also involves a hexahydrated magnesium ion (Figure 2S of the Supporting Information) which was found to be located between two symmetry-related helixes, with its coordinated water molecules hydrogen bonded to N7 and O6 of G2 and O6 of G22 from one helix and the O1P of T7 and O1P and O2P of A6 from a symmetry-related helix.

iG·T Base Pairing Conformations. The base pairing scheme of the two iG·T base pairs has been carefully examined. Initially, the force field parameters associated with the N1-H form of iG were used for the simulated annealing (SA) refinement, and the structure was refined to convergence. Interestingly, the iG16·T9 and iG4·T21 base pairs were found in the wobble and Watson−Crick configurations, respectively. The excellent fittings of the model in the electron density map are shown in Figure 3. The temperature factors of both iG·T base pairs are not unusual compared to those of other base pairs.

In addition, the question of whether the O2-H or N3-H form of iG is found in the Watson−Crick conformation of the iG4·T21 base pair was addressed. It is known that when a base ring nitrogen is protonated, the C−N−C bond angle is about 5° larger than that in the base associated with an unprotonated ring nitrogen. The SA refinement of the structure with the iG in either of the three (N1-H, O2-H, or N3-H) forms, using a 10-fold-reduced force field strength, did not yield an unambiguous preference for the geometry (especially bond angles) of iG. We concluded that the resolution (despite it being at 1.4 Å) of the structure remains insufficient to resolve the issue of tautomeric forms of iG in a de novo way. Nevertheless, our results of the two unequivocal iG·T base pair conformations seen in the high-resolution X-ray structure (shown in Figure 3) strongly support our contention that iG can adopt more than one tautomer: with iG16 adopting the N1-H form and iG4 adopting the O2-H or N3-H form.

The iG16·T9 base pair (Figure 3, bottom panel) adopts a wobble configuration of Figure 1A. The O4 and N3 atoms of the T9 base are within hydrogen bonding distances, respectively, of N1 (2.64 Å) and O2 (2.71 Å) of iG16. Three bridging water molecules between the two bases are found. Note that such a wobble base pair has an opposite effect on the positioning of the bases relative to that found in a wobble T·G base pair. In a wobble iG·T base pair, the iG base is moved toward the major groove, whereas in a T·G wobble base pair, the G base is moved toward the minor groove.

It is interesting to note that the iG4·T21 base pair is consistent with a Watson−Crick configuration of panel B or C of Figure 1. The Watson−Crick iG4·T21 base pair is essentially isostructural with respect to a normal G·C base pair. The *N*-methylpiperazine ring of the Hoechst 33342 molecule lies close to the edge of the iG4·T21 base pair in the minor groove (Figure 3, middle panel). The proximity of the piperazine ring to this pair likely serves to drive off the bridging water molecule which bonds to the O2 atoms of both T and iG in the wobble configuration, making the local environment more hydrophobic. In other words, the reduced water activity surrounding an iG·T base pair makes the iG favor the nonketo form, as noted by Shugar and colleagues (*5*). For this reason, we believe that the iG4·T21 base pair is encouraged to adopt the Watson−Crick configuration, which requires that the iG4 base is either the O2-H enol or the N3-H tautomer form.

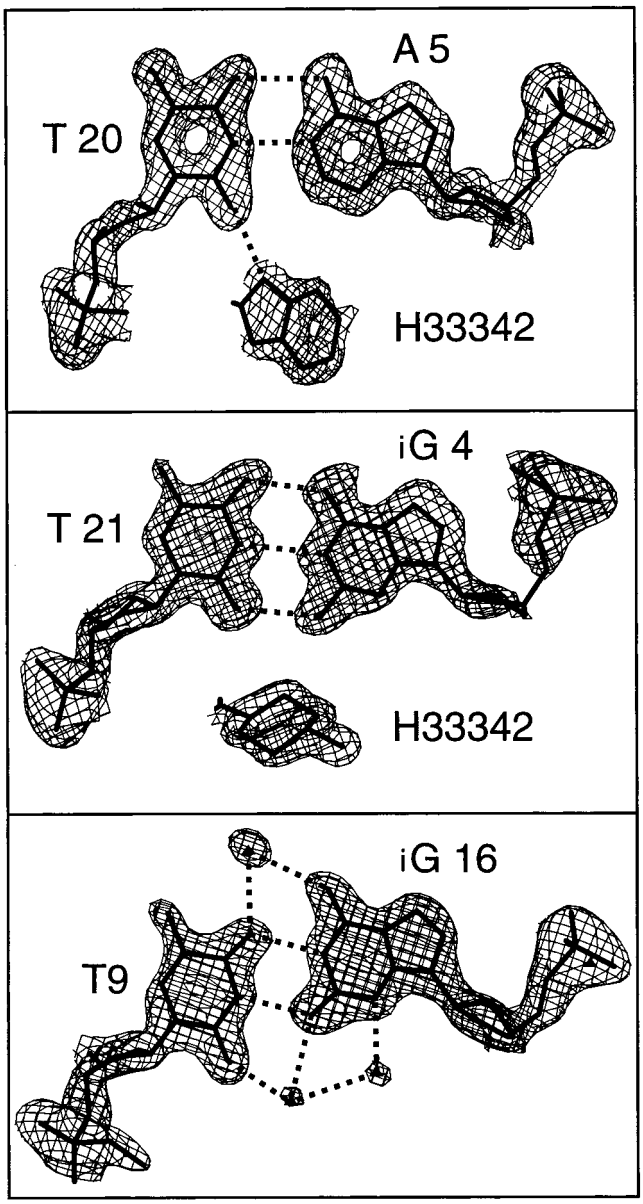

FIGURE 3: $2F_o - F_c$ electron density maps (contoured at the 1.8σ level) of the A5·T20, iG4·T21, and iG16·T9 base pairs. iG4 and iG16 are shown as the O2-H and the N1-H forms, respectively. The excellent fit of the model in the electron density suggests that a disorder between the Watson−Crick and wobble conformers is not likely. The *N*-methylpiperazine of Hoechst 33342 near the iG4·T21 base pair is shown. Three bridging water molecules are seen in the T9·iG16 wobble base pair.

Two related, but crystallographically independent, base pairs in a palindromic sequence that adopt different base pairing geometries have been observed previously in the crystal structures of the B-DNA d(CGC[e^{6}G]AATTCGCG) (*25*, *42*, *43*) and the Z-DNA d(CGCGPG) (P is a cytosine analogue) (*44*). However, to the best of our knowledge, this is the first example where this type of asymmetric pairing, involving the keto−enol tautomers of a nucleic acid base (without exocyclic chemical modification), has been directly visualized.

NMR Analysis of the iG-DODE Duplex. The question of whether a tautomeric equilibrium exists in solution has been

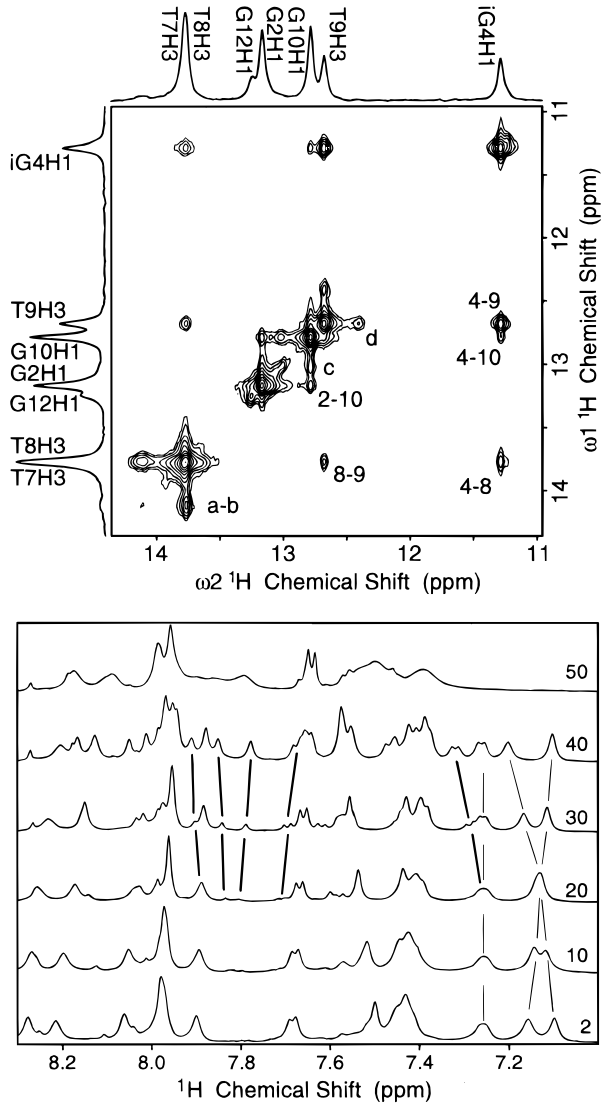

FIGURE 4: (Top) Proton 2D NOESY spectra of the imino−imino region of d(CGC[iG]AATTTGCG) at 2 °C. The minor species imino proton resonances (including three T imino protons) have clear exchange cross-peaks (labeled a−d) to the respective resonances of the major species. The data suggest that the minor species is not the hairpin form which would place the T residues in the loop region, causing the imino proton resonance to disappear in the 12−14 ppm region. The cross-peaks in the imino−amino region (data not shown) are consistent with the iG in the N1-H keto form, since its imino proton has a strong cross-peak to its own N6 amino protons (7.35 ppm). (Bottom) Temperature-dependent 1D NMR spectrum showing the increased population of the minor species at higher temperatures. Some resonances from the minor species are marked by connecting thick lines. The coexistence of the two species indicated that they are in slow exchange on the NMR time scale.

examined by NMR. The exchangeable proton 2D NMR spectra at 2 °C (Figure 4, top) shows seven resonances in the 11−14 ppm region whose identities have been assigned with 2D NOESY and TOCSY spectra (data not shown). The data are consistent with a self-complementary duplex. The chemical shifts of the imino protons of the iG4 (11.29 ppm) and T9 (12.68 ppm) bases are consistent with the two iG·T base pairs being in the wobble configuration. [The G·T wobble base pairs have their respective chemical shifts at ∼11 and ∼12 ppm (*45*)] Conclusive evidence for the wobble base pairs came from the observed NOE cross-peaks between iG4 H1 and T9 H3 (labeled 4-9, very strong, Figure 4, top panel), iG4 N6 amino (strong), A5 H2 (medium), and C3 N4 amino (weak) protons, and between T9 H3 and A5 H2 (strong), iG4 N6 amino (weak), and T9 H1′ protons (data not shown). It is interesting to note that most of the imino resonances have clear exchange cross-peaks (labeled a−d in Figure 4, top panel) to a set of minor species resonances. This minor species cannot be the hairpin form since all three T residues have their imino proton chemical shifts at 14.12 (T7 and T8) and 12.40 ppm (T9), indicative of Watson−Crick base pairs for T7 and T8 residues. Raising the temperature increased the population of the minor species (Figure 4, bottom panel). The coexistence of the two populations suggested that their exchange rate is slow on the NMR time scale. Recently, the conversion between tautomers of a base analogue in the gas state has been shown to be extremely fast using femtosecond molecular dynamics spectroscopy (*46*). However, the rate involving the equilibrium between the wobble and the Watson−Crick conformations appeared to be slow in the pairing between G and a modified cytosine (*47*−*49*), consistent with our study here.

The three-dimensional (3D) solution structure of iG-DODE with wobble iG·T base pairs was obtained by a combined SPEDREF (*32*) and NOE-constrained molecular dynamics refinement (*27*) with the chemical shifts and NMR refinement statistics listed in Tables 1S and 3S in the Supporting Information, respectively, and the experimental and simulated 2D NOESY spectra shown in Figure 3S of the Supporting Information. In the refined iG-DODE duplex, the iG4·T21 base pair adopts the wobble configuration and iG4 stacks significantly with the C3 base, causing the C3 H2′ and H2" protons to lie above the iG4 base (Figure 4S of the Supporting Information). This is consistent with the unusually ring current upfield-shifted C3 H2′ and H2″ resonances (to 1.60 and 2.15 ppm, respectively).

Structure and Interactions of the Hoechst Ligand. Figure 5 shows the $2F_o - F_c$ Fourier electron density maps of the Hoechst 33342 in the iG-DODE−H342 complex. The drug molecule can be seen to fit nicely in the caterpillar-shaped, very well-resolved, electron density envelope which allowed us to define the position, the polarity, and the conformation of the drug molecule in the duplex. The Hoechst 33342 molecule lies in the narrow minor groove of the B-DNA duplex in the iGAATT region (Figure 2). The drug covers nearly six base pairs with the *N*-methylpiperazine ring (in a chair conformation) touching the iG4·T21 on one end and the ethyl tail hanging near the iG16·T9. Its positively charged N4′ points toward DNA, with a bridging water molecule hydrogen bonded from N4′ to DNA. The polarity of the Hoechst drug binding mode is the same as in another complex crystallized in a similar lattice (*35*), but different from those in other complexes (*24*, *42*, *50*). The extra ethyl group in Hoechst 33342 makes steric clashes with the C1·G24 base pair from the neighboring 2_1-axis related duplex (Figure 1S of the Supporting Information) if the Hoechst 33342 is bound in the reverse polarity.

Figure 6 schematically summarizes the detailed interactions between the crescent-shaped Hoechst 33342 and DNA. The NH of the benzimidazole ring B forms bifurcated hydrogen bonds with T7 O2 (3.09 Å) and T19 O2 (2.97 Å) at the central A6pT7 step, and the NH of the benzimidazole ring C forms another hydrogen bond with T20 O2 (2.82 Å). A bridging water plays important roles in anchoring the

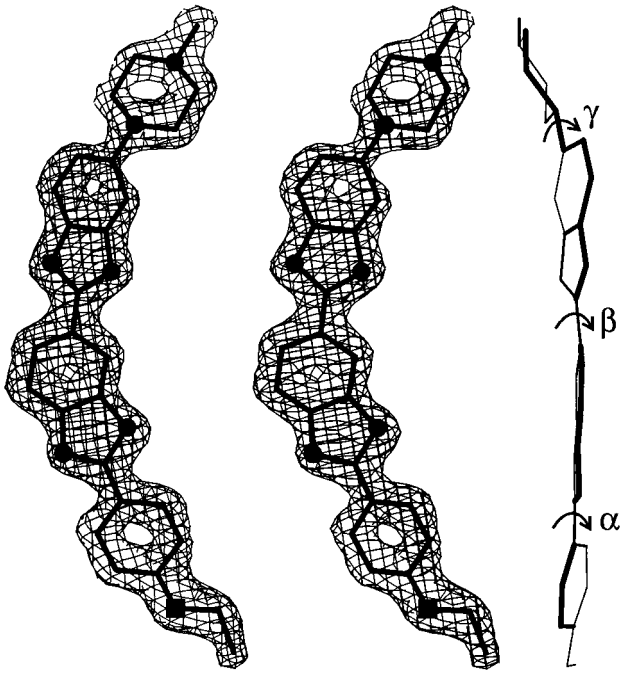

FIGURE 5: (Left and middle) Stereoscopic $2F_o - F_c$ electron density maps (contoured at the 1.8σ level) of the Hoechst 33342. (Right) The final refined Hoechst structure has a crescent shape with significant dihedral angles between the successive rings such that it can adapt to the curved contour surface of the B-DNA helix. The *N*-methylpiperazine ring is in a chair conformation.

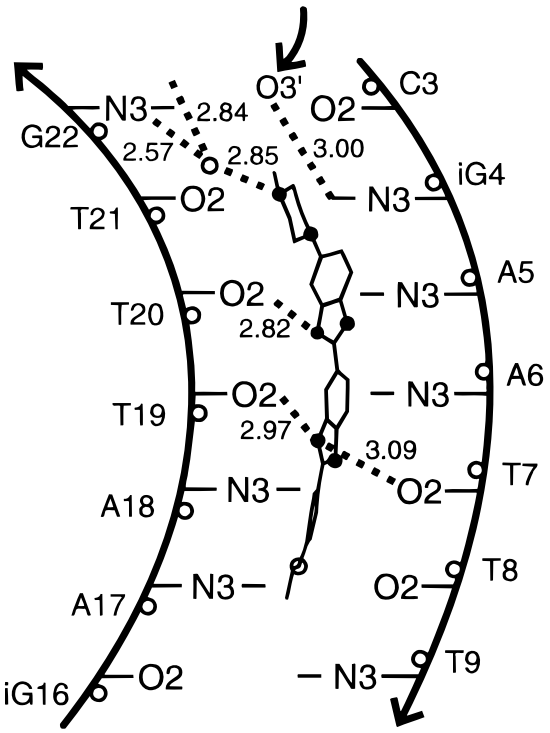

FIGURE 6: Schematic diagram showing the interactions between Hoechst 33342 and iG-DODE. The Hoechst molecule is sandwiched in the minor groove at the AATT site between the two antiparallel backbones of the DNA helix. Many van der Waals interactions, such as the dipole$-\pi$ interaction between the O4′ (e.g., from sugars of A6, A18, and T20) and the aromatic rings of the Hoechst drug, are used to stabilize the binding. Hydrogen bonds are shown as dotted lines. A bridging water is boxed between the *N*-methylpiperazine ring and DNA. There is a hydrogen bond between iG4 O2 and G*12 O3′ from an abutting duplex related by the 2_1-axis in the *c*-axis direction.

N-methylpiperazine end of the Hoechst ligand in its position. That water molecule forms three strong hydrogen bonds with G22 N3 (2.57 Å), C23 O4′ (2.84 Å), and the positively charged nitrogen atom of the *N*-methylpiperazine ring (2.85 Å) (Figure 2).

This complex of DNA and the Hoechst drug provided additional information on how Hoechst drug molecules adjust their conformation to adapt to the contour surface of the narrow minor groove in B-DNA (*24*, *35*, *40*, *50*−*53*). The three dihedral angles, α, β, and γ, in the iG-DODE−H342 complex are 21, 20, and −30°, respectively. It has been pointed out that in many drug−DNA complexes, both the drug and DNA molecules change their respective conformation to adapt to each other (*54*). This work is consistent with that concept. Superposition of the Hoechst drugs from several Hoechst drug−DNA complexes by fitting the ring B between them showed that they differ from one another in the overall curvature and the dihedral angles between successive rings. In general, H258 has a higher curvature than H342. The structural analysis of the complexes suggests that this is due to the additional ethyl group in H342 which would have a close contact with DNA if the H342 maintains the same curvature as H258. The variation in the drug conformation may have some relevance in the fluorescence quantum yield of the Hoechst drug (*55*) which may be related to the interplanar dihedral angle between the aromatic rings in the drug molecule.

Our structures reinforce the observations that the binding of the Hoechst molecule to DNA is stabilized by several types of forces, including electrostatic attraction between the positively charged drug and the negatively charged DNA, van der Waals interaction between the DNA sugar atoms along the two walls of the minor groove, and hydrogen bonds between the NH of the two benzimidazoles and the T O2 and A N3 of DNA (*35*, *50*−*53*). The preference for an AT sequence is aided by the natural tendency of the AT segment to have a narrow minor groove which provides a favorable surrounding for having the above-described interactions. Finally, the additional G N2 amino group presents a steric hindrance toward the drug binding. As the drug actually covers six base pairs, it requires at least four AT core sequence for tight binding (*56*).

Conclusion. Our X-ray crystallographic and NMR data taken together demonstrate that iG has a high propensity to adopt multiple tautomeric forms in a DNA duplex. Furthermore, high temperatures or low-dielectric medium was observed to shift the equilibrium toward the minor tautomeric species, consistent with those conditions found in earlier studies to favor the enol tautomer (*5*, *7*). On the basis of these observations, it is probable that the identity of the minor tautomeric form in iG-DODE is the O2-H form, although the N3-H tautomer cannot be categorically excluded. The binding of a hydrophobic drug molecule (Hoechst 33342) in the minor groove near the iG site may promote minor tautomer formation. A corollary is that the environment in the DNA replication apparatus (polymerases) is likely to have a low dielectric, thus being conducive to the minor tautomeric (enol) form.

It should be pointed out that iG (also called 2-hydroxyadenine) is an ionizing radiation damage product of adenine (*15*, *16*, *19*, *57*). While the repair of other radiation-damaged bases, e.g., 8-oxoadenine, has been extensively studied (*58*),

such a study on iG has not been carried out. On the basis of our results, we surmise that iG·T mismatches may escape repair to the extent that they assume a Watson−Crick base pairing geometry.

Last, structural studies of alternative base pairs here and elsewhere (*11*, *59*, *60*) or alternative backbones (3′−5′ vs 2′−5′) (*61*, *62*) in nucleic acids should be directly relevant in understanding the limiting parameters for early nucleic acid development (*63*). Why DNA and RNA were selected as carriers of genetic information is largely unknown. As concerns iG, a possible answer may be that it was eliminated from genetic material due to its facile equilibrium between the ambivalent keto−enol tautomers, which would be detrimental to the fidelity of information transfer.

SUPPORTING INFORMATION AVAILABLE

Four figures showing the crystal packing interactions, interaction of the hexahydated magnesium ion, experimental and simulated 2D-NOESY spectra, and a comparison of base pair stacking interactions and four tables listing the proton chemical shifts, NMR refinement statistics, completeness of diffraction data, and torsion angles of the iG-DODE duplex (9 pages). Ordering information is given on any current masthead page.

REFERENCES

1. Goodman, M. F. (1995) *Nature 378*, 237−238.
2. Topal, M. D., and Fresco, J. R. (1976) *Nature 263*, 285−287.
3. Sheina, G. G., Stephanian, S. G., Badchenko, E. D., and Blaoi, Y. P. (1987) *J. Mol. Struct. 158*, 275−292.
4. Szczepaniak, K., Szczesniak, M., and Person, W. B. (1988) *Chem. Phys. Lett. 153*, 39−44.
5. Sepiol, J., Kazimierczuk, Z., and Shugar, D. (1976) *Z. Naturforsch. 31C*, 361−370.
6. Pitsch, S., Krishnamurthy, R., Bolli, M., Wendeborn, S., Holzener, A., Minton, M., Lesueur, C., Schlonvogt, I., Jaun, B., and Eschenmoser, A. (1995) *Helv. Chim. Acta 78*, 1621−1635.
7. Seela, F., Wei, C., and Kazimierczuk, Z. (1995) *Helv. Chim. Acta 78*, 1843−1854.
8. Seela, F., Chen, Y., Melenewski, A., Rosemeyer, H., and Wei, C. (1996) *Acta Biochim. Pol. 43*, 45−52.
9. Seela, F., and Wei, C. F. (1997) *Helv. Chim. Acta 80*, 73−85.
10. Sugiyama, H., Ikeda, S., and Saito, I. (1996) *J. Am. Chem. Soc. 118*, 9994−9995.
11. Yang, X.-L., Sugiyama, H., Ikeda, S., Saito, I., and Wang, A. H.-J. (1998) *Biophys. J.* (in press).
12. Roberts, C., Chaput, J. C., and Switzer, C. (1997) *Chem. Biol. 4*, 899−908.
13. Roberts, C., Bandaru, R., and Switzer, C. (1997) *J. Am. Chem. Soc. 119*, 4640−4649.
14. Tchou, J., and Grollman, A. P. (1993) *Mutat. Res. 299*, 277−287.
15. Kamiya, H., and Kasai, H. (1996) *FEBS Lett. 391*, 113−116.
16. Kamiya, H., and Kasai, H. (1997) *Biochemistry 36*, 11125−11130.
17. Switzer, C., Moroney, S. E., and Benner, S. A. (1989) *J. Am. Chem. Soc. 111*, 8322−8323.
18. Switzer, C. Y., Moroney, S. E., and Benner, S. A. (1993) *Biochemistry 32*, 10489−10496.
19. Kamiya, H., and Kasai, H. (1995) *J. Biol. Chem. 270*, 2595−2600.
20. Grollman, A. P., Johnson, F., Tchou, J., and Eisenberg, M. (1994) *Ann. N.Y. Acad. Sci. 726*, 208−213.
21. Lipscomb, L. A., Peek, M. E., Morningstar, M. L., Verghis, S. M., Miller, E. M., Rich, A., Essigmann, J. M., and Williams, L. D. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 719−723.
22. Roberts, C., Bandaru, R., and Switzer, C. (1995) *Tetrahedron Lett. 36*, 3601−3604.
23. Wang, A. H.-J., and Gao, Y.-G. (1990) *Methods 1*, 91−99.
24. Teng, M.-k., Usaman, N., van Boom, J. H., and Wang, A. H.-J. (1988) *Nucleic Acids Res. 16*, 2671−2690.
25. Gao, Y.-G., Sriram, M., van der Marel, G. A., van Boom, J. H., and Wang, A. H.-J. (1993) *Biochemistry 32*, 9639−9648.
26. Rabinovich, D., and Shakked, Z. (1984) *Acta Crystallogr. A40*, 195−200.
27. Brünger, A. (1993) *X-PLOR*, version 3.1, The Howard Hughes Medical Institute and Yale University, New Haven, CT.
28. Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. T., and Berman, H. M. (1996) *Acta Crystallogr. 52*, 57−64.
29. Liaw, Y.-C., Chern, J.-W., Lin, G.-S., and Wang, A. H.-J. (1992) *FEBS Lett. 297*, 4−8.
30. Sheldrix, G. M. (1997) *SHELX-97*.
31. Parkin, S., Moezzi, B., and Hope, H. (1995) *J. Appl. Crystallogr. 28*, 53−56.
32. Robinson, H., and Wang, A. H.-J. (1992) *Biochemistry 31*, 3524−3533.
33. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.
34. Hore, P. J. (1983) *J. Magn. Reson. 54*, 539−542.
35. Vega, M. C., Saez, I. G., Aymami, J., Eritja, R., van der Marel, G. A., van Boom, J. H., Rich, A., and Coll, M. (1994) *Eur. J. Biochem. 222*, 721−726.
36. Kopka, M. L., and Larsen, T. A. (1992) in *Nucleic Acid Targeted Drug Design* (Propst, C. L., and Perun, T. J., Eds.) pp 303−374, Dekker, New York.
37. Drew, H. R., and Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535−556.
38. Coll, M., Frederick, C. A., Wang, A. H.-J., and Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385−8389.
39. Coll, M., Aymami, J., van der Marel, G. A., van Boom, J. H., Rich, A., and Wang, A. H.-J. (1989) *Biochemistry 28*, 310−320.
40. Carrondo, M., Coll, M., Aymami, J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., and Rich, A. (1989) *Biochemistry 28*, 7849−7859.
41. Coll, M., Sherman, S. E., Gibson, D., Lippard, S. J., and Wang, A. H.-J. (1990) *J. Biomol. Struct. Dyn. 8*, 315−330.
42. Sriram, M., van der Marel, G. A., Roelen, H. L. P. F., van Boom, J. H., and Wang, A. H.-J. (1992) *EMBO J. 11*, 225−232.
43. Sriram, M., van der Marel, G. A., Roelen, H. L. P. F., van Boom, J. H., and Wang, A. H.-J. (1992) *Biochemistry 31*, 11823−11834.
44. Moore, M., van Meervelt, L., Salisbury, S. A., Kong Thoo Lin, P., and Brown, D. M. (1995) *J. Mol. Biol. 251*, 665−673.
45. Patel, D. J., Kozlowski, S., Ikuta, S., and Itakura, K. (1984) *Fed. Proc. 43*, 2663−2670.
46. Douhal, A., Kim, S. K., and Zewail, A. H. (1995) *Nature 378*, 260−263.
47. Fazakerley, G. V., Gdaniec, Z., and Sowers, L. (1993) *J. Mol. Biol. 230*, 6−10.
48. Gdaniec, Z., Ban, B., Sowers, L., and Fazakerley, G. V. (1996) *Eur. J. Biochem. 242*, 271−279.
49. Nedderman, A., Stone, M. J., Williams, D. H., Kong Thoo Lin, P., and Brown, D. M. (1993) *J. Mol. Biol. 230*, 1068−1076.
50. Quintana, J. B., Lipanov, A. A., and Dickerson, R. E. (1991) *Biochemistry 30*, 10294−10306.
51. Wang, A. H.-J., and Teng, M.-K. (1990) in *Crystallographic and Modeling Methods in Molecular Design* (Bugg, C. E., and Ealick, S. E., Eds.) pp 123−150, Springer-Verlag, New York.
52. Clark, G. R., Gray, E. J., Neidle, S., and Leupin, W. (1996) *Biochemistry 35*, 13745−13752.
53. Clark, G. R., Squire, C. J., Gray, E. J., Leupin, W., and Neidle, S. (1996) *Nucleic Acids Res. 24*, 4882−4889.
54. Wang, A. H.-J., Liaw, Y. C., Robinson, H., and Gao, Y. G. (1990) in *Molecular basis of specificity in nucleic acid-drug interactions* (Pullman, B., and Jortner, J., Eds.) pp 1−21, Kluwer, Dordrecht, The Netherlands.

55. Loontiens, F. G., McLaughlin, L. W., Diekmann, S., and Clegg, R. M. (1991) *Biochemistry 30*, 182−189.
56. Zimmer, C., and Wahnert, U. (1986) *Prog. Biophys. Mol. Biol. 47*, 31−112.
57. Mori, T., and Dizdaroglu, M. (1994) *Radiat. Res. 140*, 85−90.
58. Wallace, S., and Kow, W., Eds. (1994) *DNA Damage: Effects on DNA Structure and Protein Recognition*, Annals of the New York Academy of Sciences, Vol. 726.
59. Bhat, B., Neelima, Leonard, N. J., Robinson, H., and Wang, A. H.-J. (1996) *J. Am. Chem. Soc. 118*, 3065−3066.
60. Bandaru, R., Hashimoto, H., and Switzer, C. (1995) *J. Org. Chem. 60*, 786−787.
61. Robinson, H., Wang, A. H.-J., Jung, K.-E., and Switzer, C. (1995) *J. Am. Chem. Soc. 117*, 837−838.
62. Hashimoto, H., and Switzer, C. Y. (1992) *J. Am. Chem. Soc. 114*, 6255−6256.
63. Joyce, G. F. (1989) *Nature 338*, 217−224.
64. Lavery, R., and Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 655−667.

BI980818L