Guanine–1,N^6-Ethenoadenine Base Pairs in the Crystal Structure of d(CGCGAATT(ϵdA)GCG)[†,Δ]

Gordon A. Leonard,[‡] Katherine E. McAuley-Hecht,[§] Neil J. Gibson,[§] Tom Brown,[§] William P. Watson,[∥] and William N. Hunter*[,‡]

Department of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, U.K., Centre for Molecular Recognition, Department of Chemistry, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JJ, U.K., and Shell Research Ltd., Sittingbourne Research Centre, Sittingbourne, Kent ME9 8AG, U.K.

Received December 15, 1993; Revised Manuscript Received February 3, 1994[⊗]

ABSTRACT: A single-crystal X-ray analysis of the synthetic oligomer d(CGCGAATT(ϵdA)GCG) (ϵdA = 1,N^6-ethenoadenosine) has been carried out. The B-form duplex crystallizes in the orthorhombic space group $P2_12_12_1$ with unit cell dimensions a = 24.31 Å, b = 39.65 Å, and c = 63.05 Å. Refinement has converged with R = 0.182 for 2837 reflections in the resolution range 7.0–2.25 Å for a model consisting of the duplex, one Mg^{2+} ion, and 127 water molecules. The structure contains two G·ϵdA base pairings which adopt a G(*anti*)·ϵdA(*syn*) conformation. The geometry of the two mispairs suggests that the G·ϵdA pairings are held together by three interbase hydrogen bonds. These are N2(G)–H···N1(ϵdA), N1(G)–H···N9(ϵdA), and O6(G)···H–C8(ϵdA). The last interaction serves to alleviate the destabilizing effect that would occur due to the presence of an unfulfilled hydrogen bond acceptor. A superposition of the G(4)·ϵdA-(21) base pair found in this structure and the Watson–Crick G(4)·C(21) base pair observed in the native dodecamer d(CGCGAATTCGCG) indicates a significant difference in the sugar/phosphate backbone. However, the overall conformations of the two duplexes remain similar, suggesting that the modified base pairs are accommodated into the double helix mainly by alterations of the backbone conformation. Such structural rearrangement of the backbone, upon incorporation of ϵdA, may provide a signal to the 3-methyladenine–DNA glycosylase that repairs such lesions.

Vinyl chloride is a known carcinogen. A mechanism whereby this chemical may react with DNA is depicted in Figure 1. Epoxidation and subsequent rearrangement result in the formation of chloroacetaldehyde which reacts with DNA to form base adducts. It is also possible for the chlorooxirane to react directly with DNA bases (Guengerich et al., 1993). This type of reaction with DNA results in alterations to the availability of functional groups used in hydrogen-bonding patterns and the pairing of bases which are not complementary in the Watson–Crick sense. Reaction with deoxyadenosine and deoxycytosine yields 1,N^6-ethenodeoxyadenosine (ϵdA) and 3,N^4-ethenodeoxycytosine, respectively (Barrio et al., 1972; Kusmierek & Singer, 1982; Leonard, 1984) while reaction with deoxyguanosine gives both 1,N^2- and 3,N^2-ethenodeoxyguanosines (Sattsangi et al., 1977; Singer et al., 1987).

Of these base adducts, the most widely studied has been ϵdA (Figure 2). Although the role of ϵdA in carcinogenesis remains unclear, it has been found (Barbin et al., 1981) that, during *in vitro* replication by *Escherichia coli* DNA polymerase I, G was misincorporated opposite chloroacetaldehyde-treated poly(dA) and also that both G and C are preferentially inserted in DNA that has been treated with the same reagent

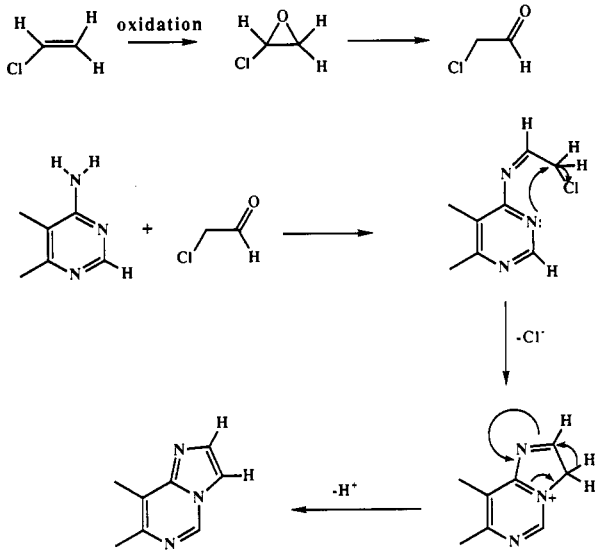


FIGURE 1: Mechanism by which vinyl chloride can be transformed into chloroacetaldehyde which then reacts with DNA to form base adducts.

(Hall et al., 1981). Singer and co-workers (Singer et al., 1984; Singer & Spengler, 1986) confirm that G is misinserted in DNA containing ϵdA but at a much lower frequency than was initially thought. They also suggested that thymine can be incorporated opposite ϵdA but that the resulting base pair behaved as a bulky lesion which could be bypassed during DNA replication, yielding frame-shift mutations. This hypothesis was given credence by the results of an NMR study (Kouchakdjian et al., 1991) which showed that while an ϵdA·T base pair could be incorporated into a DNA double helix without disrupting the adjacent base pairs, there is a non-

[†] Supported by the Wellcome Trust and the Science and Engineering Research Council (U.K.).

[Δ] The coordinates have been deposited in the Brookhaven Protein Data Bank under the code 150D.

* Author to whom correspondence should be addressed (phone, 061-275-4712; Fax, 061-275-4598; Bitnet, WNH@UK.AC.MANCHESTER.CHEMISTRY.V5).

[‡] Department of Chemistry, University of Manchester.

[§] Centre for Molecular Recognition, Department of Chemistry, University of Edinburgh.

[∥] Shell Research Ltd.

[⊗] Abstract published in *Advance ACS Abstracts*, April 1, 1994.

0006-2960/94/0433-4755$04.50/0

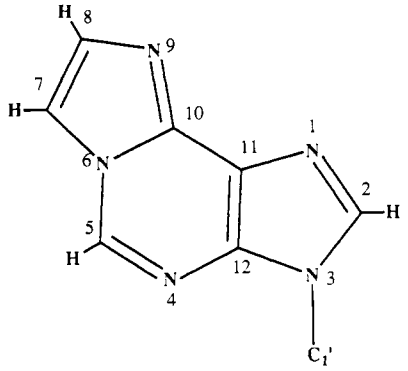


FIGURE 2: Representation of 1,N^6-ethenoadenosine with the atom numbering scheme.

coplanarity of the two bases which might cause a problem during DNA replication. Another NMR study (de los Santos et al., 1991) has indicated that the G·ϵdA base pair in the synthetic duplex d(CATGGGTAC)·d(GTAC(ϵdA)CATG) adopts a dG(*anti*)·ϵdA(*syn*) conformation analogous to the dG(*anti*)·dA(*syn*) geometry found for the unmodified non-Watson–Crick base pair (Brown et al., 1986). Our interest in the conformations adopted by base pairs involving chemically modified bases (Brown et al., 1993; Leonard et al., 1990a, 1992a) has led us to study the three-dimensional structure of the synthetic deoxydodecamer d(CGCGAATT(ϵdA)GCG), which contains two G·ϵdA base pairs. We present the structure of this duplex, as determined by single-crystal X-ray diffraction techniques, with particular reference to the conformation of the G·ϵdA mispair and the associated hydrogen-bonding pattern.

EXPERIMENTAL PROCEDURES

Synthesis. 2′-Deoxyadenosine monohydrate was treated with an excess (20 equiv) of 2 M aqueous chloroacetaldehyde solution at pH 7 and 38 °C for 16 h. The resulting acid solution was neutralized by the addition of sodium hydrogen carbonate. 2′-Deoxy-1,N^6-ethenoadenosine was isolated in a 74% yield by flash column chromatography [silica, 10% methanol (MeOH) in dichloromethane]. The nucleoside was protected as the 5′-dimethoxytrityl ether and converted to the 2-cyanoethyl *N*,*N*-diisopropylphosphoramidite prior to DNA synthesis. The dodecanucleotide was synthesized on an ABI 380B synthesizer using standard phosphoramidite methodology, and the modified base was coupled to the growing DNA chain with a 98.5% efficiency as calculated from the colorimetric analysis of the trityl cation released during synthesis. The 1,N^6-ethenodeoxyadenosine ring is degraded by aqueous base (Yip & Tsou, 1973), and consequently, deprotection of the synthetic dodecamer was achieved by soaking the solid support in a solution of 10% 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in dry MeOH for 5 days. The DBU/MeOH solution was then neutralized with Dowex (PyH^+) in water prior to purification of the dodecamer using reversed-phase high-pressure liquid chromatography (HPLC). Once purified, a sample of the DNA was subjected to digestion by snake venom phophodiesterase and alkaline phosphatase. HPLC analysis of the resulting nucleosides confirmed the successful incorporation of ϵdA into the oligomer.

Ultraviolet Melting Studies. Melting curves of synthetic oligonucleotide duplexes were measured on a Perkin-Elmer λ15 ultraviolet visible spectrometer controlled by an IBM PS2 microcomputer and equipped with a peltier block. All samples were heated at a rate of 1 °C/min, and the melting process was monitored at 265 nm. Data were collected and processed using the PECSS2 software package (Perkin-Elmer). The melting temperature (T_m) was defined as the temperature at which the first derivative of the melting curve was at its maximum. All samples contained 8.5 μM DNA dissolved in an aqueous buffer consisting of 100 mM sodium chloride, 10 mM sodium phosphate, and 1 mM EDTA adjusted to pH 7.0 by the addition of sodium hydroxide.

Crystallization, Data Collection, and Processing. Crystals were grown at 4 °C from 20-μL sitting drops containing d(CGCGAATT(ϵdA)GCG) (1 mM), sodium cacodylate buffer (12 mM, pH 6.5), magnesium chloride (19–24 mM), spermine tetrahydrochloride (0.5 mM), and 2-methyl-2,4-pentanediol (MPD) (12% v/v) equilibrated against an external reservoir of MPD. Diffraction-quality crystals grew over a period of days. A needle of dimensions 1.0 × 0.15 × 0.15 mm was mounted in a thin-walled glass capillary and used for data collection on an R-AXIS IIc image plate system mounted on a Rigaku RU200 rotating anode generator operating at 50 kV and 140 mA. Data were measured at 24 °C with graphite monochromated Cu Kα radiation (λ = 1.5418 Å, 0.5-mm focal spot). The image plate was set at a distance of 115 mm from the crystal, corresponding to a data collection limit of 2.25 Å, and data were collected on 30 3° oscillation frames with an exposure time of 60 min per frame. Unit cell determination and the crystal orientation were achieved by autoindexing reflections on three still exposures taken at 45° intervals. Data reduction was carried out using the software package PROCESS (Higashi, 1990). A total of 12 357 reflections were measured of which 9932 had $I/\sigma(I) \geq 0.5$. These were then used in scaling, post refinement, and intensity averaging procedures and yielded 3072 unique reflections with $R_{sym} = 0.032$ [$R_{sym} = \Sigma|I(k) - \langle I\rangle|/\Sigma I(k)$, where $I(k)$ and $\langle I\rangle$ represent the intensity measurements for individual measurements and the corresponding mean values]. This represents 94% of the theoretical data available to 2.25-Å resolution, 88% in the range 2.5–2.25 Å.

Structure Solution and Refinement. Orthorhombic unit cell dimensions (after post refinement) of $a = 24.311(2)$ Å, $b = 39.648(2)$ Å, and $c = 63.050(2)$ Å, space group $P2_12_12_1$, indicated that, like other similar sequences containing mismatch base pairs, the structure of d(CGCGAATT(ϵdA)GCG) is isomorphous to that of the native sequence d(CGCGAATTCGCG) (Drew et al., 1980), crystals of which have the unit cell dimensions $a = 24.87$ Å, $b = 40.39$ Å, and $c = 66.20$ Å. However, given that the difference in the c dimension is rather large, we used a full molecular replacement protocol to position the duplex in the unit cell rather than the rigid body approach we use for structures with cells closer in size to that of the native structure (Leonard et al., 1990, 1992a,b). Rotation and translation searches were carried out using the structure of d(CGTGAATTCACG) (Larsen et al., 1991) as a search model with the X-PLOR software package (Brünger, 1990) using Patterson correlation coefficients both to refine the orientation and to carry out the translation search. A Patterson radius of 15 Å was used in conjunction with data between 15- and 4-Å resolution. After Patterson correlation refinement the rotational parameters were optimized to Eulerian angles, $\theta_1 = 333.1°$, $\theta_2 = 7.5°$, and $\theta_3 = 26.1°$. The refined angles were input to a translation search again, just utilizing the data between 15- and 4-Å resolution. A 4σ peak in the translation function provided a solution for the positional parameters. The packing of the duplexes in the unit cell was confirmed as being very similar to that of the native structure, the hydrogen bond patterns, between symmetry-related

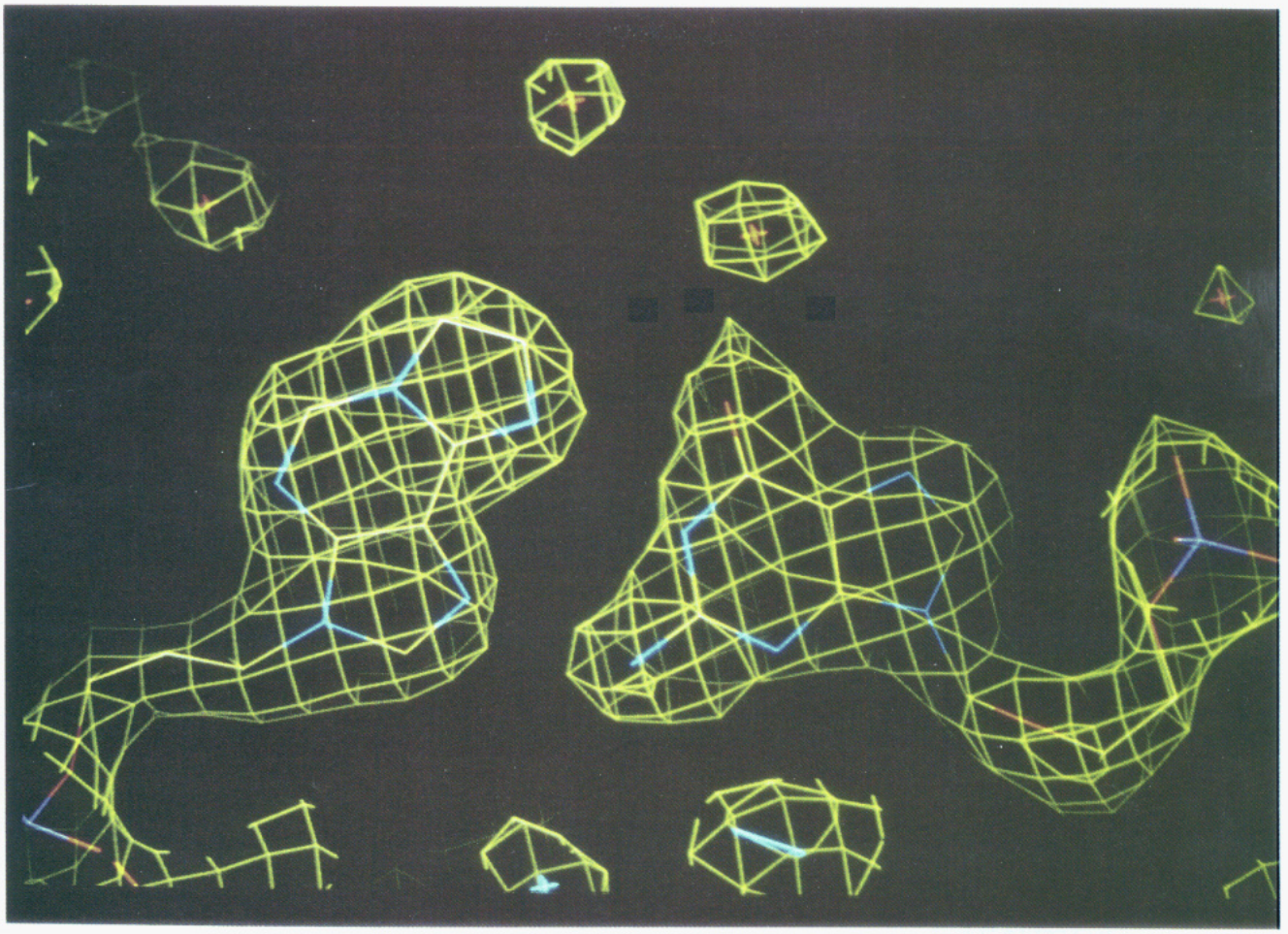


FIGURE 3: G4·ϵdA21 base pair superimposed on a $2F_o - F_c$ electron density map (green chicken wire) calculated using the coordinates of the final refined model. Also shown are some water molecules (orange crosses) found in the vicinity of the base pair. Those atoms from a symmetry-related asymmetric unit are shown in cyan. The map is contoured at approximately one rms deviation from the mean density observed in the unit cell.

duplexes, which link the minor grooves at the ends of the double helix being conserved. Rigid body refinement resulted in a model with $R = 0.50$ for 1268 reflections with $F \geq 2\sigma(F)$ to a resolution limit of 3.0 Å ($R = \Sigma|F_o - F_c|/\Sigma F_o$, where F_o and F_c are the observed and calculated structure factors, respectively), and further rigid body refinement of this model using 2111 $2\sigma(F)$ data in the range 7.0–2.5 Å resulted in $R = 0.475$.

All $2\sigma(F)$ data to 2.25-Å resolution were included, and refinement of the model continued with the restrained least-squares methods of Konnert and Hendrickson (1981) in the program NUCLSQ (Westhof et al., 1985). Thirteen cycles of positional refinement reduced R to 0.39, after which a further five cycles were carried out with the base residues coincident with the modified base-pair sites removed from the structure factor calculations. Electron density ($2F_o - F_c$) and difference density ($F_o - F_c$) maps were calculated and examined on an Evans and Sutherland ESV30 graphics workstation using FRODO (Jones, 1978). These Fourier syntheses indicated that the conformations of the modified base pairs are dG(*anti*)·ϵdA(*syn*), and alterations were made to the model. Additionally, the two terminal base pairs were adjusted to improve the fit to the electron density. Refinement was continued using NUCLSQ. The dictionary of restraints for the etheno ring of the ϵdA moieties was constructed on the basis of the single-crystal structure of the nucleoside (Jaskolski, 1982). Further refinement involved the inclusion of individual, though restrained, temperature factors, solvent molecules, and the periodic examination of electron density and difference density maps as we have described elsewhere for similar analyses (Leonard et al., 1990a, 1992a,b).

The refinement converged with $R = 0.182$ for 2837 $2\sigma(F)$ data in the resolution range 7.0–2.25 Å for a model consisting of the B-DNA duplex (494 atoms), one Mg^{2+} ion, and 127 water molecules modeled as oxygen atoms. The final geometry of the model is excellent with root mean square (rms) deviations from ideality of 0.009 and 0.013 Å for sugar/base and phosphate distances, respectively. For angle distances the rms deviations are 0.023 and 0.020 Å. The fit of the final model to the electron density is excellent in all respects, and an example is provided in Figure 3. Refined coordinates, with thermal parameters, and structure factors have been deposited with the Brookhaven Protein Data Bank (Abola et al., 1987).

RESULTS AND DISCUSSION

Overall Conformation of the Double Helix. In the final model the nucleotides are labeled C1 to G12 in the 5′ to 3′ direction on strand 1 and C13 to G24 in the same direction on strand 2. The Mg^{2+} ion is labeled I25, and the 127 water molecules are labeled HOH26 to HOH153 The B-DNA duplex contains 10 standard Watson–Crick base pairs and two G(*anti*)·ϵdA(*syn*) mispairs at positions G4·ϵdA21 and G16·ϵdA9. Figure 4a shows a superposition of a G(*anti*)·A(*syn*) base pair observed in the structure of d(CGCGAATTAGCG) (Brown et al., 1986) and the equivalent Watson–Crick G·C base pair in the native structure (Drew et al., 1980).

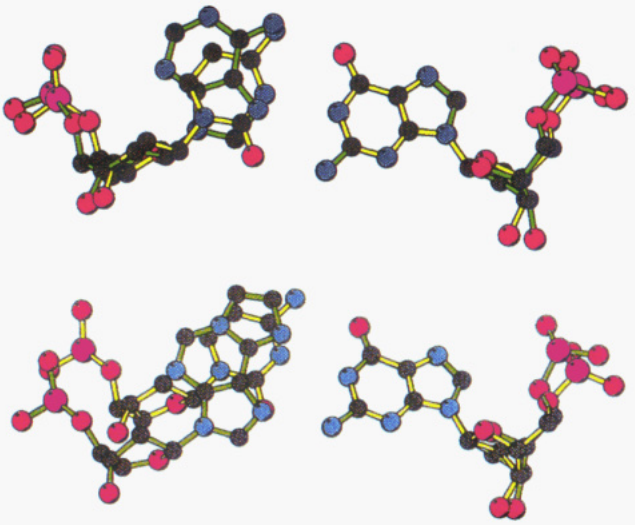


FIGURE 4: Superpositions of (a, top) the G4·A21 base pair found in the structure of d(CGCGAATTAGCG) and the equivalent Watson–Crick G·C base pair in the native structure and (b, bottom) the G4·ϵdA21 base pair found in the structure of d(CGCGAATT(ϵdA)GCG) and the Watson–Crick G·C base pair in the native structure. In both structures the native G·C base pair is shown with yellow bonds. The superpositions were achieved by a least-squares fit of those atoms in the guanine bases only. This diagram was obtained using MOLSCRIPT (Kraulis, 1991).

Two important points to note are that the *syn* adenine base is oriented so that the functional groups involved in pairing are in positions similar to those observed for the cytosine in the native base pair and that the sugar/phosphate backbones of both base pairs overlap closely. Figure 4b shows a similar superposition of the G4·ϵdA21 base pair found in the structure of d(CGCGAATT(ϵdA)GCG) and the equivalent base pair in the native structure. As observed for the G(*anti*)·A(*syn*) base pair, there is good overlap of the functional groups in the modified base and the native cytidine base. However, unlike the G(*anti*)·A(*syn*) base pair, there is not a good overlap of the sugar/phosphate backbone of the two base pairs. This distortion to the overall shape of the base pair appears to be dictated by a requirement to align the functional groups of the G(*anti*)·ϵdA(*syn*) mispairs and the native G·C base pairs. The only other base pair which causes a noticeable perturbation to the sugar/phosphate backbone is the G(*anti*)·A(*anti*) mismatch (Privé et al., 1987).

The global conformation of the d(CGCGAATT(ϵdA)GCG) duplex is very similar to that of the native structure. The sugar/phosphate torsion angles fall within the range usually observed for B-DNA dodecamers (Table 1; Saenger, 1984). It would appear that the G(*anti*)·ϵdA(*syn*) pairs are accommodated into the duplex by adjustments in the conformation of the sugar/phosphate backbone which do not disrupt the global conformation of the double helix. Additionally, the geometry of the base pairs and base steps adjacent to the G(*anti*)·ϵdA(*syn*) base-pair sites remains similar to that observed in the structure of the native duplex (Table 2).

We note that the thermal parameters of the mispairs are comparable to others in the structure and do not reflect any instability of the mispair in the crystal structure. However, in solution the situation changes. It is evident from ultraviolet melting experiments (Table 3) that incorporation of ϵdA into a DNA double helix destabilizes the duplex in comparison to sequences that contain only Watson–Crick base pairs. The melting temperature (T_m) of the sequence d(CGCAAATTTGCG) is 327 K while that for the duplex d(CGC(ϵdA)AATTGGCG) is only 307 K. It is noteworthy, however, that the latter sequence melts some 15 K higher than d(CGCAAATTGGCG), which has been shown to contain two A(*anti*)·G(*syn*) base pairs (Leonard et al., 1990b). So while incorporation of the chemically modified base in a mispair destabilizes relative to the Watson–Crick duplex, it is

Table 1: Sugar/Phosphate Backbone and Glycosyl Torsion Angles (deg) for d(CGCGAATT(ϵdA)GCG)[a]

residue	χ	α	β	γ	δ	ξ	ζ
C1	−106			161	113	−154	−96
G2	−88	−56	178	31	146	−153	−141
C3	−124	−52	144	50	89	−173	−95
A4	−108	−76	165	73	145	−176	−118
A5	−101	−59	175	52	150	−171	−112
A6	−104	−49	178	35	129	−186	−93
T7	−115	−50	169	45	99	−207	−66
T8	−112	−101	194	75	139	−146	−152
ϵdA9	71	−5	157	8	122	−172	−87
G10	−98	−58	164	41	140	−120	−200
C11	−122	−43	131	33	117	−176	−94
G12	−119	58	180	53	115		
C13	−119			77	134	−100	−196
G14	−125	−51	129	36	121	−160	−110
C15	−133	−49	151	52	98	−150	−118
A16	−94	−56	150	65	164	−143	−174
A17	−107	−20	152	9	151	−171	−118
A18	−98	−31	172	25	138	−175	−107
T19	−121	−40	158	49	117	−168	−126
T20	−117	−19	149	37	120	−174	−101
ϵdA21	69	−44	183	26	127	−173	−93
G22	−100	−58	168	55	142	−92	−194
C23	−112	−73	146	37	124	−156	−106
G24	−116	−50	156	38	89		

[a] Main-chain torsion angles are defined by O3′–Pα–O5′–β–C5′–γ–C4′–δ–C3–μ–ξ–O3′–ζ–P–O5′. The glycosyl torsion angle χ is defined by O4′–C1′–N1–C2 for pyrimidines and O4′–C1′–N9–C4 for purines.

Table 2: Values of Geometric Parameters Describing the Base Steps in the d(CGCGAATT(ϵdA)GCG) Duplex[a]

base step		twist (deg)	rise (Å)	cup (deg)	roll (deg)
1	C1pG2	34 (39)	3.9 (3.1)	−4 (6)	6 (1.0)
2	G2pC3	39 (38)	2.9 (3.0)	0 (2)	−5 (−9)
3	C3pG4	28 (28)	4.0 (4.1)	−18 (−14)	2 (2)
4	G4pA5	41 (37)	3.0 (3.0)	6 (3)	−4 (4)
5	A5pA6	39 (37)	3.1 (3.3)	6 (2)	3 (0)
6	A6pT7	31 (32)	3.2 (3.1)	6 (1)	−2 (−6)
7	T7pT8	38 (36)	3.1 (3.2)	5 (2)	2 (1)
8	T8p(ϵdA)9	43 (41)	2.8 (2.9)	10 (9)	−3 (5)
9	(ϵdA)9pG10	25 (31)	4.1 (4.1)	−23 (−12)	1 (4)
10	G10pC11	37 (40)	3.1 (3.0)	0 (5)	−6 (−6)
11	C11pG12	38 (37)	3.5 (3.8)	0 (−9)	−3 (3)

[a] All helical parameters and torsion angles quoted throughout this paper were calculated using the NEWHEL92 program distributed by R. E. Dickerson through the Brookhaven Protein Data Bank. Data in parentheses are for the native d(CGCGAATTCGCG) (Drew et al., 1980).

Table 3: Ultraviolet Melting Temperatures [T_m (K)] of Several ϵdA-Containing Self-Complementary Dodecamers and of Some Corresponding Unmodified Sequences[a]

sequence	T_m (K)
d(CGCGAATTCGCG)	331.9
d(CGCAAATTTGCG)	326.9
d(CGC(**ϵdA**)AATTTGCG)	288.7
d(CGC(**ϵdA**)AATT**G**GCG)	307.0
d(CGC(**ϵdA**)AATT**C**GCG)	283.5
d(CGC(**ϵdA**)AATT**A**GCG)	289.4
d(CGCGAATT(**ϵdA**)GCG)	298.3
d(CGCGAATT**A**GCG)	biphasic
d(CGCAAATT**G**GCG)	292.3

[a] All solutions were 100 mM NaCl, 10 mM phosphate, 1 mM EDTA, and 8.5 μM DNA at pH 7.00. The bases involved in non-Watson–Crick base pairing are shown in bold type.

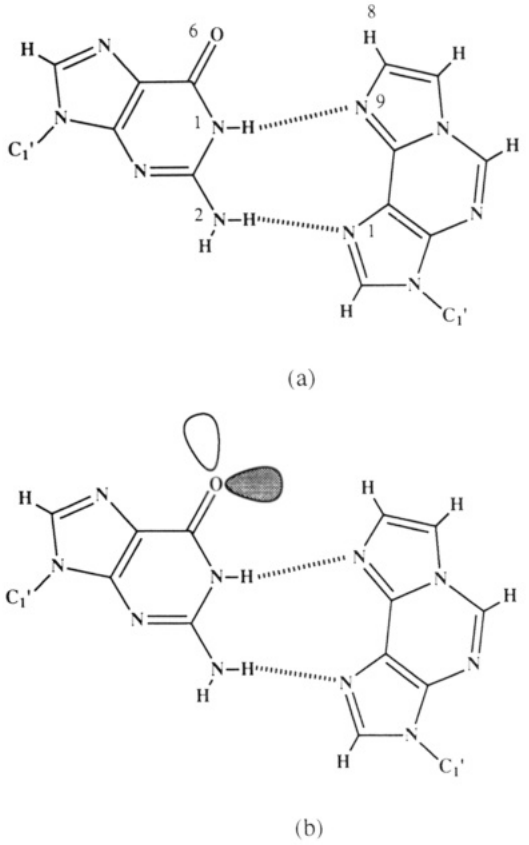


FIGURE 5: (a) Two obvious interbase hydrogen bonds in the G(*anti*)·ϵdA(*syn*) base pairs. (b) Carbonyl oxygen lone pair (shaded) which would not be able to form a hydrogen bond if not involved in C–H···O hydrogen bonding.

significantly more stable than a structure containing A(*anti*)·G(*syn*) base pairs. This enhanced stability, in comparison to the A·G mispair, infers that the presence of ϵdA could increase the incidence of A → C transversions if duplex instability is important in mismatch recognition and repair.

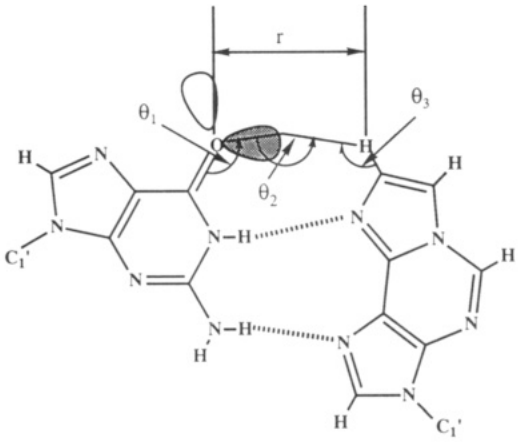
G·ϵdA Base Pairs. The conformation of the G·ϵdA mispairs in the crystal structure of d(CGCGAATT(ϵdA)GCG) is G(*anti*)·ϵdA(*syn*) and is similar to that proposed on the basis of NMR studies on the sequence d(CATGGGTAC)·d(GTAC-(ϵdA)CATG) (de los Santos et al., 1991). As with all purine–purine *anti–syn* base pairs and purine–pyrimidine wobble base pairs, there is significant asymmetry in the angles between the glycosidic bonds and the C1′–C1′ vector. These are designated λ_1 and λ_2 [for example, see Figure 3 of Leonard et al. (1992b)]. In Watson–Crick base pairs λ_1 and λ_2 have similar values which fall within a narrow range (52–62°) and give these base pairs an element of pseudosymmetry. For the G(*anti*)·ϵdA(*syn*) base pairs the values of λ_1 and λ_2 are 64° and 34° for the 4·21 pair and 63° and 33° for the 16·9 pair.

The G(*anti*)·ϵdA(*syn*) base pairs have two interbase distances that are clearly indicative of hydrogen bonding (Figure 5a). These are N2(G)–N1(ϵdA) (2.8 Å for the 4·21 pair and 2.7 Å for the 16·9 pair) and N1(G)–N9(ϵdA) (3.0 and 2.8 Å, respectively). We propose, however, that there are in fact three interbase hydrogen bonds. The main interbase interactions are attributable to the two N–H···N hydrogen bonds which orient the bases with respect to each other, the third hydrogen bond being provided by a weaker C–H···O=C interaction. Our justification for invoking such an interaction follows.

Theoretical and experimental charge density studies indicate that the lone-pair electron density of sp^2 oxygen atoms is concentrated into two lobes (Stevens & Coppens, 1980; Stevens, 1980; Coppens et al., 1981; Howard et al., 1992). The appearance of this density is not unlike the conventional textbook depiction of lone pairs. In addition, an extensive survey of N–H···O=C hydrogen bond geometry in accurate small molecule structures (Taylor et al., 1983) has clearly indicated that these hydrogen bonds form in the directions implied by lone-pair orientation. We therefore feel justified in depicting the G(*anti*)·ϵdA(*syn*) base pair as shown in Figure 5b. The *anti–syn* configuration of the G·ϵdA base pairs results in one of the sp^2 lone pairs on guanine O6 (shaded in Figure 5b) being unable to form a hydrogen bond to solvent as a result of steric hindrance by 1,N^6-ethenoadenine. It has long been recognized (Crick, 1966) that base pairs containing unfulfilled hydrogen bond donors or acceptors would be destabilized. With this in mind we proposed a reverse three-center hydrogen-bonding scheme which allows alleviation of a similar destabilizing effect in both the G(*anti*)·A(*syn*) and G(*anti*)·O8A(*syn*) base pairs (O8A = 8-oxoadenine) [see Figures 5 and 7 of Leonard et al. (1992a)]. In the G(*anti*)·A(*anti*) base pair (Privé et al., 1987) similar potential destabilization of the base pair is reduced when the lone pair in question participates in a cross-strand three-center hydrogen bond. However, in G(*anti*)·ϵdA(*syn*) base pairs, no form of three-center hydrogen bonding involving the guanine O6 carbonyl oxygen is possible.

We propose that the potential destabilization of G(*anti*)·ϵdA(*syn*) base pairs by a non-hydrogen bonded carbonyl oxygen lone pair is alleviated by the formation of an interbase C–H···O hydrogen bond involving the C8 hydrogen atom of the ϵdA moiety and one of the guanine O6 carbonyl lone-pair orbitals. Although C–H···O hydrogen bonds have been noted in the crystal structures of nucleosides [for example, see Sundaralingam (1966)], we are unaware of any publications which suggest they might contribute to the stability of base-pair formation in duplex structures.

The existence of C–H···O hydrogen bonds has been the subject of controversy (Donohue, 1968; Desiraju, 1991), but it is now well established that they do exist (Taylor & Kennard, 1982; Jeffery & Muluzynska, 1982; Willberg et al., 1991; Jeffery & Saenger, 1991; Steiner & Saenger, 1993). The electrostatic nature of hydrogen bonding means that the attractive force decreases only slowly as the atomic separation increases (Umeyama & Morokuma, 1977), and so the H···O distance need not be limited to the sum of the van der Waals radii of carbon and hydrogen. Given that the hydrogen-bonding directions of the guanine carbonyl O6 groups are as shown in Figure 5b (Taylor et al., 1983), we can postulate the existence of inter-base-pair C–H···O hydrogen bonds if we have H···O distances which are reasonable (around 3.0 Å or less) and a H–lone pair–O angle near 180°. The QUANTA software package (QUANTA Version 3.3, Molecular Simulations Inc., Waltham, MA) was used to generate idealized positions for the hydrogen atoms in our structure. We also [as in Leonard et al. (1992a)] generated ideal positions for the maximum electron density of the carbonyl O6 oxygen lone pairs by assuming that the oxygen–lone-pair distance would be 0.75 Å, half the van der Waals radius of an oxygen atom. This detail is important as it can be used to give a indication of the directionality of the C–H···lone-pair interaction. Relevant angles and distances were then measured (Figure 6). Both the H···O distances and angles are within the range observed for C–H···O hydrogen bond formation observed in numerous small-molecule structure analyses (Taylor et al., 1983). We therefore propose that the



Base-pair	θ_1(°)	θ_2(°)	θ_3(°)	r(Å)
G(4)•ɛdA(21)	139	155	99	3.1
G(16)•ɛdA(9)	142	150	103	2.7

FIGURE 6: Analysis of the geometry of the O6(G)···H–C8(ɛdA) hydrogen bonds in G(*anti*)•ɛdA(*syn*) base pairs. θ_1 is the C=O···H angle, θ_2 is the angle subtended at the lone-pair position between the carbonyl oxygen and the hydrogen, and θ_3 is the C–H···lone pair angle.

G(*anti*)•ɛdA(*syn*) base pairs in the structure of d(CGC-GAATT(ɛdA)GCG) are held together by three hydrogen bonds. We have no direct experimental evidence that the C–H···O hydrogen bonds proposed actually form as X-ray crystallography only provides geometrical information about the systems under investigation. However, the circumstantial evidence of nearly ideal H···O distances and O-lone pair–H angles coupled with the intrinsic destabilization of the base pairs that would occur if these C–H···O hydrogen bonds do not occur represents compelling evidence for us to confidently predict their formation.

There is a slight difference in the hydration of the two G•ɛdA base pairs. On the major-groove side of G4 there are two water molecules which link O6 and N7. The solvent interacting with O6 is depicted in Figure 3. The other solvent is not shown; it has been clipped out of the picture in removing electron density associated with adjacent bases so that a clear view of the mispair is given. G16 has a single solvent interacting with these functional groups. On the minor-groove side the N2 amino groups of each G•ɛdA base pair form a hydrogen bond with the terminal O3′ hydroxyl group of a symmetry-related duplex.

CONCLUSIONS

Our analysis of the crystal structure of the synthetic dodecanucleotide d(CGCGAATT(ɛdA)GCG) has shown that the two G•ɛdA base pairs in the duplex adopt an *anti–syn* conformation. The mispairs are incorporated into the double helix with no large alteration to the overall conformation of the duplex. A comparison of the G•ɛdA base pair with a G•C pair suggests that there is a significant alteration to the positioning of the sugar/phosphate backbone of the modified nucleotide. This alteration may serve as a signal to the specific enzyme (3-methyladenine–DNA glycosylase; Singer et al., 1992) that recognizes and removes this lesion in duplex DNA. The G•ɛdA base pair itself is stabilized by three hydrogen bonds, one of which [O6(G)···H–C8(ɛdA)] serves to alleviate any destabilization to the base pair that would result from the presence of an unfulfilled O6(G) hydrogen bond acceptor at the interface of the two bases. Both this study and a solution study of d(CATGGGTAC)•d(GTAC(ɛdA)CATG) (de los Santos et al., 1991) indicate that the G(*anti*)•ɛdA(*syn*) base pair is stable enough to allow duplex formation and thus suggests a mechanism for the transverse mutation of an A•T base pair to a C•G base pair if the lesion is not recognized and subsequently repaired.

REFERENCES

Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F., & Weng, J. (1987) in *Crystallographic Databases—Information Content, Software Systems, Scientific Applications* (Allen, F. H., Berghoff, G., & Seivers, R., Eds.) pp 107–132, Data Commission of the International Union of Crystallography, Cambridge, U.K.

Barbin, A., Bartsch, H., Leconte, P., & Radman, M. (1981) *Nucleic Acids Res. 9*, 375–387.

Barrio, J. R., Secrist, J. A., & Leonard, N. J. (1972) *Biochem. Biophys. Res. Commun. 46*, 579–604.

Brown, T., Hunter, W. N., Kneale, G. G., & Kennard, O. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2402–2406.

Brown, T., Hunter, W. N., & Leonard, G. A. (1993) *Chem. Br. 29*, 484–488.

Brünger, A. T. (1990) *X-PLOR Manual. Version 2.2*, Howard Hughes Medical Institute, Yale University, New Haven, CT.

Coppens, P., Boehme, R., Price, P. F., & Stevens, E. D. (1981) *Acta Crystallogr. A37*, 857–863.

Crick, F. H. C. (1966) *J. Mol. Biol. 19*, 548–555.

de los Santos, C., Kouchakdjian, M., Yarema, K., Basu, A., Essigman, J., & Patel, D. J. (1991) *Biochemistry 30*, 1828–1835.

Desiraju, G. R. (1991) *Acc. Chem. Res. 24*, 290–296.

Donohue J. (1968) in *Structural Chemistry and Molecular Biology* (Rich, A., & Davidson, N., Eds.) pp 459–463, W. H. Freeman, San Francisco.

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

Guengerich, F. P., Persmark, M., & Humphreys, W. G. (1993) *Chem. Res. Toxicol. 6*, 635–648.

Hall, J. A., Saffhill, R., Green, T., & Hathway, D. E. (1981) *Carcinogenesis 2*, 141–146.

Hendrickson, W. A., & Konnert, J. H. (1981) in *Biomolecular Structure, Conformation, Function and Evolution* (Srinivasan, R., Ed.) Vol. 1, pp 43–51, Pergamon Press, Oxford, U.K.

Higashi, T. (1990) *PROCESS: A program for indexing and processing R-AXIS II image plate data*, Rigaku Corp., Tokyo, Japan.

Howard, S. T., Hursthouse, M. B., Lehmann, C. W., Mallinson, P. R., & Frampton, C. S. (1992) *J. Chem. Phys. 97*, 5616–5630.

Jaskolski, M. (1982) *Acta Crystallogr. B34*, 3171–3174.

Jeffery, G. A., & Muluzynska, H. (1982) *Int. J. Biol. Macromol. 4*, 173–185.

Jeffery, G. A., & Saenger, W. (1991) *Hydrogen Bonding in Macromolecular Structures*, Springer, Berlin.

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

Kouchakdjian, M., Eisenberg, M., Yarema, K., Basu, A., Essigman, J., & Patel, D. J. (1991) *Biochemistry 30*, 1820–1828.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.

Kusmierek, J. T., & Singer, B. (1982) *Biochemistry 21*, 5717–5722.

Larsen, T. A., Kopka, M. L., & Dickerson, R. E. (1991) *Biochemistry 30*, 4443–4449.

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

Leonard, G. A., Booth, E. D., & Brown, T. (1990b) *Nucleic Acids Res. 18*, 5617–5623.

Leonard, G. A., Guy, A., Brown, T., Téoule, R., & Hunter, W. N. (1992a) *Biochemistry 31*, 8415–8420.

Leonard, G. A., Booth, E. D., Hunter, W. N., & Brown, T. (1992b) *Nucleic Acids Res. 20*, 4753–4759.

Leonard, N. J. (1984) *CRC Crit. Rev. Biochem. 15*, 125–199.

Prive, G. G., Heinemann, U., Kan, L. S., Chandrasegaran, S., & Dickerson, R. E. (1987) *Science 238*, 498–504.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Sattsangi, P. D., Leonard, N. J., & Frihart, C. R. (1977) *J. Org. Chem. 42*, 3292–3296.
Singer, B., & Spengler, S. J. (1986) in *The Role of Cyclic Nucleic Acid Adducts in Carcinogenesis and Mutagenesis* (Singer, B., & Bartch, H., Eds.) pp 359–371, IARC Scientific Publications 70, Lyon, France.
Singer, B., Abbot, L. J., & Spengler, S. J. (1984) *Carcinogenesis 5*, 1165–1171.
Singer, B., Spengler, S. J., Chavez, F., & Kusmierek, J. T. (1987) *Carcinogenesis 8*, 745–747.
Singer, B., Antoccia, A., Basu, A. K., Dosanjh, M. K., Fraenkel-Conrat, H., Gallagher, P. E., Kusmierek, J. T., Qiu, Z.-H., & Rydberg, B. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9386–9390.
Steiner, T., & Saenger, W. (1993) *J. Am. Chem. Soc. 115*, 4540–4547.
Stevens, E. D. (1980) *Acta Crystallogr. B36*, 1876–1886.
Stevens, E. D., & Coppens, P. (1980) *Acta Crystallogr. B36*, 1864–1876.
Sundaralingam, M. (1966) *Acta Crystallogr. 21*, 495–506.
Taylor, R., & Kennard, O. (1982) *J. Am. Chem. Soc. 104*, 5063–5070.
Taylor, R., Kennard, O., & Verishel, W. (1983) *J. Am. Chem. Soc. 105*, 5761–5766.
Umeyama, H., & Morokuma, K. J. (1977) *J. Am. Chem. Soc. 99*, 1316–1332.
Westhof, E., Dumas, P., & Moras, D. J. (1985) *J. Mol. Biol. 184*, 119–145.
Willberg, K. B., Waldron, R. F., Shulte, G., & Saunders, M. (1991) *J. Am. Chem. Soc. 113*, 971–977.
Yip, K. F., & Tsou, K. C. (1973) *Tetrahedron Lett. 33*, 3087–3090.