Conformation of Guanine·8-Oxoadenine Base Pairs in the Crystal Structure of d(CGCGAATT(O8A)GCG)†

Gordon A. Leonard,‡ André Guy,§ Tom Brown,‡ Robert Téoule,§ and William N. Hunter*,‖

Department of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, U.K., Department of Chemistry, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JJ, U.K., and Service d'Etudes des Systèmes et Architectures Moleculaires, Département de Recherche Fondamentale de la Matière Condensée, Centre d'Etudes Nucléaires de Grenoble, BP 85X, F 38041 Grenoble Cedex, France

Received April 2, 1992; Revised Manuscript Received June 1, 1992

ABSTRACT: The structure of the synthetic deoxydodecamer d(CGCGAATT(O8A)GCG)$_2$ (O8A = 8-oxoadenine) has been determined by single-crystal X-ray diffraction techniques. The oligonucleotide crystallizes in the orthorhombic space group $P2_12_12_1$ with cell dimensions of a = 25.48 Å, b = 41.84 Å, and c = 64.91 Å. The refinement has converged with an R-factor of 0.151 for 1119 reflections in the resolution range 8.0–2.25 Å. Sixty-seven solvent molecules were located during the course of the refinement. The B-DNA helix consists of ten Watson–Crick base pairs and two guanine·8-oxoadenine (G·O8A) base pairs. In order to achieve hydrogen-bonding complementarity between the two bases, an unusual G(anti)·O8A(syn) wobble conformation is adopted. It is proposed that the G·O8A mispairs are held together by a network of four interbase hydrogen bonds which are the result of the formation of two reverse three-center hydrogen-bonding systems. These involve one carbonyl oxygen lone pair interacting with two hydrogen atoms. In a departure from previous observations of the characteristics of purine·purine anti·syn base pairs, λ_1 and λ_2, the angles between the glycosidic bonds and the C1′–C1′ vector, are symmetric. A reassessment of the other purine·purine mispairs suggests that similar three-center hydrogen bonds may occur and make a contribution to stabilizing other base pairings.

7,8-Dihydro-8-oxoadenine (O8A) is a major product resulting from the degradation of adenine under the action of ionizing radiation (Conley, 1963; van Hemmen et al., 1971; Téoule, 1987). The modification arises due to hydroxyl radical attack on the adenine base. The C8 oxidation of the purine DNA bases has important biological implications, and hence there has been interest in the structure and function of both O8A and the corresponding guanine analogue 7,8-dihydro-8-oxoguanine (O8G).

These bases adopt several tautomeric forms (Cho & Evans, 1981; Culp et al., 1989), but under physiological conditions the 6,8-diketo species predominates for O8G (Aida & Nishimura, 1987) and the 6-amino-8-keto form predominates for O8A (Figure 1). This is also the case when O8A and O8G are found in oligonucleotides (Oda et al., 1991; Guy et al., 1988). C8 oxidation of adenine and guanine therefore modifies the hydrogen-bonding characteristics of these bases only when they adopt the syn conformation. As purine·purine base pairs generally have one base in the syn conformation (Hunter et al., 1986a; Brown et al., 1989; Leonard et al., 1990), C8 oxidation of purine bases is therefore likely to affect primarily their ability to form base pairs with other purine bases, and this, in part, will determine the mutagenic and carcinogenic effects of O8A and O8G when incorporated into DNA.

It has been shown that DNA polymerase allows DNA synthesis to proceed past both O8A and O8G lesions in DNA and that when O8G is present, A or C is inserted opposite the modified base with the ratio of A to C dependent on the polymerase used (Shibutani et al., 1991). When O8A is present, for the large majority of events, thymine is incorporated opposite this lesion (Guschlbauer et al., 1991). An NMR analysis (Kouchakdjian et al., 1991) of a dodecanucleotide containing O8G·A mispairs has suggested that these mismatches adopt an O8G(syn)·A(anti) conformation directly analogous to that found for G(syn)·A(anti) mispairs (Brown et al., 1990; Lenoard et al., 1990) (Figure 2). However, in the former base pair there is no requirement for the protonation of the N1 of adenine, and therefore at physiological pH this base pair is likely to be more stable than the unmodified G(syn)·A(anti) base pair. This may be the reason why, when O8G is present in DNA, DNA polymerase directs

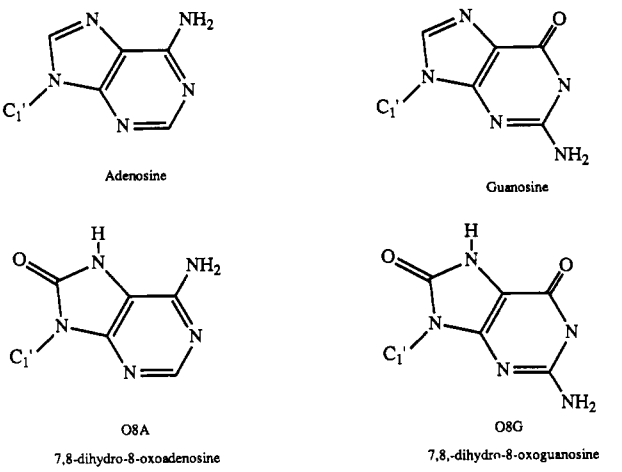

FIGURE 1: (Top) Molecular formulas of adenosine and guanosine. (Bottom) The 6-amino-8-keto form of 8-oxoadenosine and the 6,8-diketo form of 8-oxoguanine. These are the major tautomeric forms of the bases present at physiological pH. This figure and Figures 2, 6, 7, and 8 were produced using the ChemDraw software from Cambridge Scientific Computing, Inc., Cambridge, MA.

† Supported by The Wellcome Trust, the Science and Engineering Council, U.K., and Shell U.K. Ltd.

* Correspondence should be addressed to this author.

‡ Department of Chemistry, University of Edinburgh.

§ Département de Recherche Fondamentale de la Matière Condensée, Centre d'Etudes Nucléaires de Grenoble.

‖ Department of Chemistry, University of Manchester.

0006-2960/92/0431-8415$03.00/0

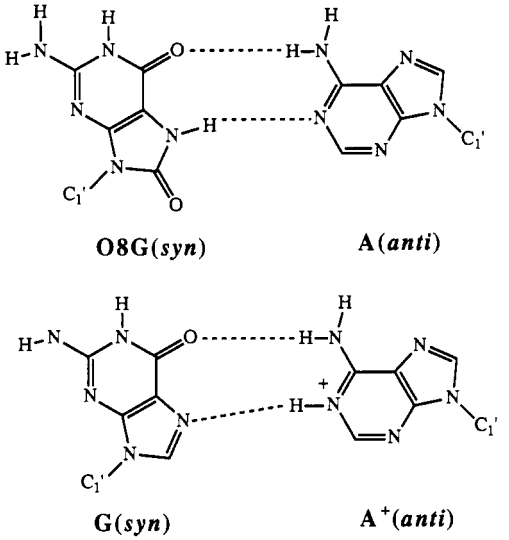

FIGURE 2: (Top) Structure of the A(anti)·O8G(syn) base pair as determined from an NMR analysis (Kouchakdjian et al., 1991). (Bottom) Analogous unmodified base pair as revealed by single-crystal X-ray diffraction techniques (Brown et al., 1990; Leonard et al., 1990). In this latter base pair there is a requirement for protonation of the N1 of the adenine in order that the two interbase hydrogen bonds may form.

the misinsertion of significant amounts of A opposite this lesion. Thymine is almost exclusively incorporated opposite O8A. We are elucidating the structures of base pairs involving O8A to discover why this should be so, given the altered functional groups present on this modified base. We present here the results of a crystallographic study of the synthetic DNA dodecamer d(CGCGAATT(O8A)GCG)$_2$—a duplex containing two G·O8A base pairs.

EXPERIMENTAL PROCEDURES

Synthesis and Crystallization. The monomer 5′-*O*-(4,4′-dimethoxytrityl)-7,8-dihydro-N^6-(phenoxyacetyl)-2′-deoxy-8-oxoadenosine 3′-*O*-(2-cyanoethyl) *N*,*N*-diisopropylphosphoramidite was synthesized according to Guy et al. (1988). The dodecamer synthesis was performed using standard cyanoethyl diisopropylphosphoramidite methodology with 10 μmol of dG^{iBu} linked on an aminopropylated controlled pore glass support. Cleavage from the solid support and deprotection of the oligodeoxynucleotide were then carried out with (4 × 500 μL) aqueous ammonia solution for 6 h at room temperature. A further 24 h at 60 °C was necessary to fully remove the isobutyryl protecting groups. The oligodeoxynucleotide was purified by high-pressure liquid chromatography on an anion-exchange column (0.75 × 30 cm, Partisil SAX, 10 μm) using a linear gradient of 0.3 M potassium dihydrogen orthophosphate buffer (pH 6.8) with 30% acetonitrile over 50 min. The pure product (400 A_{260} units) was desalted by dialysis.

Crystals of the dodecamer d(CGCGAATT(O8A)GCG) were obtained from sitting drops containing approximately 20 mM sodium cacodylate buffer, pH 7.4, 1 mM dodecanucleotide, 12 mM magnesium acetate, 12 mM spermine tetrahydrochloride, and 9% (v/v) 2-methyl-2,4-pentanediol at 4 °C. Colorless needles were observed to grow in a period of weeks. Despite several attempts to improve the crystal size, we were unable to grow specimens with a thickness greater than 0.1 mm.

Data Collection and Processing. The largest crystal obtained, of dimensions 0.1 × 0.1 × 0.8 mm, was mounted in a thin-walled glass capillary for characterization and data collection on a Rigaku AFC5 diffractometer and rotating anode operating at 50 kV and 140 mA. The instrument is configured with a crystal to detector distance of 400 mm, a continuously evacuated beam tunnel, and an incident collimator. Data were measured at 24 °C with graphite-monochromated Cu Kα radiation (λ = 1.54 Å, 0.5-mm focal spot). Software for the diffractometer and data processing was provided by the Molecular Structure Corp., Houston, TX. Intensities were measured with 1° ω-scans using the method of Lehman and Larsen (1974). Those reflections with $F_o < 10\sigma(F_o)$, where F_o is the observed structure factor, were measured in triplicate to improve counting statistics. Three standard reflections were recorded every 150 measurements. Intensities were measured at 4°/min for *hkl* and *–h,–k,–l* for all reflections to 2.2-Å resolution. A total of 5624 reflections were measured, and these reduced to 3485 independent reflections with *R*-merge = 0.24. *R*-merge is defined as $\sum|I(k) - \langle I\rangle|/\sum|I(k)|$, where $I(k)$ and $\langle I\rangle$ represent the intensity values of individual reflections and the corresponding mean values. The summation is over all multiple measurements. For those reflections with $F_o > 5\sigma(F_o)$ *R*-merge was 11%. Although data to a resolution of 2.25 Å were recorded, we noted that reflections beyond 3.0 Å were weak; see below. This is mainly a consequence of the small crystal size used in this analysis. We attribute the high *R*-merge values as primarily due to the large number of weak reflections that were measured and secondly due to the noticable absorption effects; see below. Data were corrected for Lorentz and polarization factors as well as for a 6% crystal decay during data collection. An empirical absorption correction was also applied (North et al., 1968) with a minimum transmission factor of 0.62. This value reflects the crystal orientation with the long axis tilted at approximately 40° to the length of the capillary.

Structure Refinement. Orthorhombic unit cell dimensions of a = 25.48 Å, b = 41.84 Å, and c = 64.91 Å, space group $P2_12_12_1$, indicated quasi-isomorphism with the native dodecanucleotide d(CGCGAATTCGCG) (Drew et al., 1980), and coordinates for this structure were used as a starting model in the refinement procedure. These coordinates were obtained from the Brookhaven Protein Data Bank (Abola et al., 1987), entry 1BNA. Initially the starting model was refined as a rigid body using a modified version of SHELX (Sheldrick, 1976). This part of the refinement converged at R = 0.47 for 1428 reflections wth $F > 0$ in the region 10.0–3.0 Å. *R*, the crystallographic residual, is defined as $\sum|F_o - F_c|/\sum|F_o|$, where F_o and F_c are the observed and calculated structure factors. The bases G4, C9, G16, and C21 were then removed from the structure factor calculations, and the refinement was continued using the methods of restrained least squares (Hendrickson & Konnert, 1981) with the program NUCLSQ (Westhof et al., 1985) and $2\sigma(F)$ data in the region 7.0–2.5 Å. Seven cycles of positional refinement led to a reduction in *R* from 0.42 to 0.34 for the 1000 reflections in the resolution range. Electron density ($2F_o - F_c$) and difference density ($F_o - F_c$) maps were then calculated using software provided with the CCP4 suite of crystallographic programs (*CCP4*, 1986). These were then examined on an Evans and Sutherland graphics workstation using FRODO (Jones, 1978; P. R. Evans, personal communication). The difference maps clearly showed that the 4·21 mispair was a G(anti)·O8A(syn) wobble base pair and although the quality of the electron density for the 9·16 base pair was not as good, it was indicative of the same base pair conformation. The modifications to the structure were carried out and all bases included in the structure factor calculations. Refinement was continued with data in the same resolution

range with R being reduced from 0.35 to 0.27 after eight cycles. Inclusion of restrained isotropic temperature factors into the refinement procedure further reduced R to 0.24. The resolution was then extended to include the data to 2.25 Å. Solvent molecules were then introduced into the model and electron density maps inspected at regular intervals. The solvent molecules were included into the model on the basis of spherical density on difference maps and reasonable hydrogen-bonding geometry with respect to the functional groups on the bases, the sugar/phosphate backbone, and other solvent positions.

In the initial stages of the refinement procedure it was possible to obtain a better fit of a number of base pairs (including the mispairs) to the electron density by manual manipulation than had been achieved by automatic methods. In order to maintain this fit, the restraints applied to the interbase distances derived from the manual procedure were considerably strengthened by treating them in the same fashion as angle distances in the phosphate groups. In the later stages of the refinement the restraints on the interbase distances for the Watson–Crick base pairs were relaxed considerably while those governing the interbase distances in the G·O8A mispairs were removed completely. When restrained isotropic temperature factors were included in the refinement, a few became nonpositive definite, and to counter this, the minimum value for B was set at 4 Å^2.

The final model consisting of the B-DNA duplex (492 atoms) and 67 water molecules was refined to give a final crystallographic residual of 0.151 for 1119 reflections with $F \geq 2\sigma(F)$ in the resolution range 7.0–2.25 Å. This represents approximately 60% of the data in the range 7.0–3.0 Å, 50% of the data from 7.0 to 2.75 Å, and 13% of the data from 2.75 to 2.25 Å. This falloff of strong reflections at around 3-Å resolution is typical of crystals of oligonucleotides with sequences based on that of the native dodecamer d(CGCGAATTCGCG) (Larsen et al., 1991). We judge the effective resolution of this analysis to be about 2.8 Å. The number of weak reflections measured and the R-merge value for all data suggested that a $2\sigma(F_o)$ cutoff for data to be used in the refinement would be judicious. In conjunction with this we maintained tight restraints on the geometry of the model. The geometry of the final model is thus excellent with average deviations from ideality of 0.009 Å for sugar/base distances and 0.006 Å for phosphate distances. For angle distances the average deviations from ideality are 0.022 and 0.016 Å, respectively. The refined coordinates have been deposited with the Brookhaven database (Abola et al., 1987). Structure factors are available from the authors.

RESULTS

Conformation of the Double Helix. The dodecamer is quasi-isomorphous with the native dodecamer (Drew et al., 1980) and similar mismatch dodecamers (Hunter et al., 1986a,b; Leonard et al., 1990; Corfield et al., 1987) and adopts a double-helical structure consisting of ten Watson–Crick base pairs and two G(syn)·O8A(anti) wobble base pairs. There are approximately ten base pairs per turn. The residues on strand 1 are labeled C1 to G12 (5′ to 3′ direction), and those on strand 2 are labeled C13 to G24 (5′ and 3′). The mispairs are G(4)·O8A(21) and G(16)·O8A(9). The 67 solvent molecules are labeled W25 to W91. As with other mismatch dodecamer structures the mispairs have been incorporated into the B-DNA duplex with only a slight perturbation of the sugar/phosphate backbone, and the usual configuration of the sugar/phosphate backbone for a B-DNA helix is maintained for the large majority of the residues. A further indication that there is minimal disturbance of the backbone is that the average distance between adjacent phosphorus atoms along each strand is 6.7 Å, the same as observed in the native dodecamer. Likewise, an average C1′–C1′ distance of 10.44 Å in this mismatch dodecamer compares well with a value of 10.49 Å found in the native structure.

Table I: N···O and NH···O Distances (Å) and C=O···H Angles (deg) for the G(anti)·O8A(syn) Base Pairs in the Structure of d(CGCGAATT(O8A)GCG)$_2$

	base pair			
	G(4)·O8A(21)		G(16)·O8A(9)	
	N···O	H···O	N···O	H···O
distance				
N6H···O6	2.8	2.0	3.5	2.5
N7H···O6	2.4	1.5	2.8	1.6
N1H···O8	3.2	2.1	3.0	2.1
N2H···O8	3.6	2.6	3.3	2.5
angle				
C6=O6···N6H	169		169	
C6=O6···N7H	100		109	
C8=O8···N1H	88		90	
C8=O8···N1H	138		142	

The G·O8A Base Pairs. The electron density maps for the mispairs are unambiguous (Figure 3) and show that both mispairs adopt a G(anti)·O8A(syn) wobble configuration. A least-squares fit of the coordinates of the unmodified G·A mismatch (Brown et al., 1986; Hunter et al., 1986a) to those of the structure under discussion here shows that the wobble nature of the mispair is a result of the guanine moving approximately 1.65 Å into the minor groove of the double helix while the position of the 8-oxoadenine is unchanged when compared to that of the adenine in the native base pair. The fit was achieved by superimposing the positions of the N9 atoms and those atoms in the sugar/phosphate backbone of the purine nucleotides at positions 4 and 21 in both structures.

In both of the G·O8A base pairs there are four interbase distances indicative of the formation of hydrogen bonds (Table I). These are O6(G)–N7(O8A), O6(G)–N6(O8A), O8(O8A)–N1(G), and O8(O8A)–N2(G). If all these potential hydrogen bonds are to exist, then both of the carbonyl oxygen acceptor groups in the base pair must interact with two different donor groups on the other base of the pair to give a hydrogen-bonding arrangement as shown in Figure 4.

N2(G4) and O8(O8A21) are linked by a bridge of two water molecules. Here the N2(G)–W37 distance is 2.4 Å, the O8(O8A)–W70 distance is 3.1 Å, and the W37–W70 distance is 3.3 Å. Both the $2F_o - F_c$ and $F_o - F_c$ electron density maps show some evidence of a solvent-mediated linkage of N2(G16) and O8(O8A9); however, the density around the positions of the proposed solvent molecules did not fulfill the strict requirements that we have used in the placement of solvent molecules, and therefore these atoms were not included into our model. Likewise, evidence of solvent-mediated linkages of N8(O8A) to thymines in adjacent base pairs was also not considered to be strong enough to introduce the linkages into our model. There is, however, a reasonably well-ordered solvent molecule (W76) that links N2(G16) to both O3′(G24) of a symmetry-related molecule and O4′(A17).

Although there is minimal distortion of the sugar/phosphate backbone of the B–DNA helix upon incorporation of the two G·O8A mispairs and the geometrical properties of the Watson–Crick base pairs are within the range usually found in B-DNA, there is a major difference between the properties of the G·O8A mispairs and those of other mismatch base pairs found in

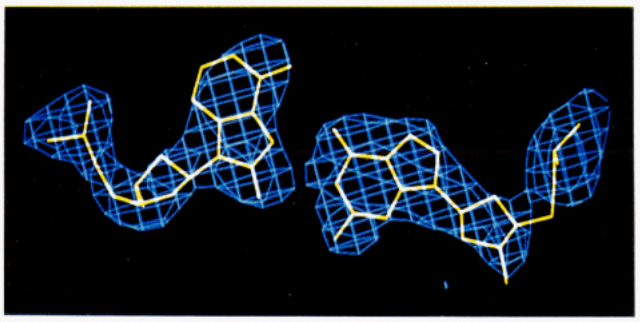

FIGURE 3: Fragment $F_o - F_c$ electron density map (blue chicken wire) for the G(4)·O8A(21) base pair with the final refined structure of the base pair superimposed. The map was obtained by removing all the atoms of G(4) and O8A(21) from the structure factor calculations and calculating a difference Fourier synthesis. The map contour level is 3.0 times the rms value of the difference density in the unit cell.

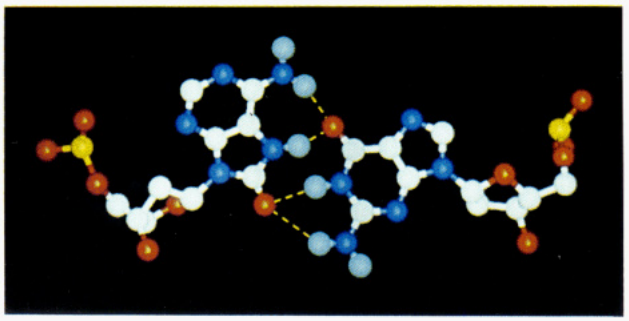

FIGURE 4: Ball and stick representation of the G·O8A pairing. The possible interbase hydrogen bonds in the G(anti)·O8A(syn) base pair are depicted as dashed yellow lines. Oxygen atoms are red, nitrogen atoms blue, carbon atoms white, phosphorus atoms yellow, and hydrogen atoms light blue.

B-DNA duplexes. This concerns the values of λ_1 and λ_2—the angles between the glycosidic bonds and the C1′–C1′ vector. In the native dodecamer (Drew et al., 1980) the values of λ_1 and λ_2 at the positions of interest are 53° and 56° (4·21) and 54° and 54° (16·9). These values are typical of Watson–Crick base pairs (Seeman et al., 1976; Rosenberg et al., 1976). In the GA mismatch dodecamer (Brown et al., 1986; Hunter et al., 1986a) the values are 56°, 41° and 62°, 40°, respectively. This asymmetry between λ_1 and λ_2 has been noted for all purine·purine anti·syn base pairs and all purine·pyrimidine wobble base pairs so far analyzed. However, in the structure under discussion here, λ_1 and λ_2 are highly symmetric with values of 39°, 43° (4·21) and 45°, 49° (16·9).

DISCUSSION

Hydrogen Bonding in the Mispairs. Do all four interbase hydrogen bonds in the G·O8A mispair really exist? A survey of the geometry of the hydrogen bonds formed by carbonyl oxygens in the structures of small molecules (Taylor et al., 1983) has shown that these functional groups can and do form hydrogen bonds by interacting with two, or even three, different hydrogen bond donor groups. These hydrogen bonds are thought to be real if the H- - -O distances are short enough and the C=O- - -H angle is ≥90°.

Idealized positions for the hydrogen atoms in the G·O8A base pairs in our structure were generated with the SYBYL software package (SYBYL Software, Tripos Associates, St. Louis, MO) using N–H distances of 1.03 Å. The NH- - -O and C=O- - -H distances and angles were then measured, and the results of this analysis are shown in Table I. As can be seen, given the accuracy of our structure, all the relevant distances and angles are within the accepted range for hydrogen

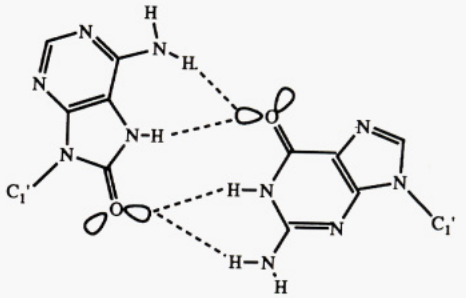

FIGURE 5: Schematic representation of the possible reverse three-center hydrogen-bonding systems for the G(anti)·O8A(syn) base pairs. Note that the two carbonyl oxygen sp^2 lone pairs not involved in the hydrogen bonding point away from the other base in the base pair and are available for hydrogen bonding to solvent molecules.

bond formation. Although there may be no intrinsic electronic factors governing sp^2 lone pair directionality, studies of NH- - -O=C hydrogen bonds in small molecules (Taylor et al., 1983) have shown that there is a tendency for these hydrogen bonds to form along the conventionally viewed directions of sp^2 lone pairs. Bearing this in mind, it is apparent that only one lone pair from each carbonyl oxygen could be involved in interbase hydrogen bond formation as the geometry of the G·O8A base pairs results in one of the lone pairs on each carbonyl oxygen pointing away from the other base in the pair. For all four potential interbase hydrogen bonds to be formed, each carbonyl oxygen would then be involved in an interbase three-centered hydrogen bond in which the lone pair electrons and not, as is usual, a hydrogen atom is at the center of the hydrogen-bonding system (Figure 5).

To test this hypothesis, we then generated ideal positions for the maximum electron density of the carbonyl oxygen lone pairs by assuming that the oxygen lone pair distance would be 0.75 Å, half the value of the van der Waals radius of an oxygen atom. The geometry of the potential three-centered systems were then analyzed (Figure 6) and were found to be remarkably similar to conventional three-centered hydrogen-bonding systems (Table II). For both the G·O8A base pairs the sum of θ_1, θ_2, and α for the arrangements centered on both the O6(G) and the O8(O8A) carbonyl oxygen lone pairs is close to the values required for conventional three-center hydrogen bond formation (Taylor et al., 1984; Jeffrey, 1989; Fritsch & Westhof, 1991). Given this evidence, we therefore propose that the G·O8A base pairs in our structure are held together by four hydrogen bonds which are the result of the formation of two three-center hydrogen bonds centered on one lone pair each of O6(G) and O8(O8A). We have termed these systems reverse three-center hydrogen bonds. It should be noted that the two lone pairs not involved in the reverse three-center hydrogen bonding are free to form hydrogen bonds to solvent.

Another factor involved in proposing this type of hydrogen-bonding system for the G·O8A base pairs is the resulting stabilization of the N^2-amino group. The geometry of the base pairs under discussion here means that, in the absence of the reverse three-centered system that we propose, one of the guanine N^2-amino group hydrogens cannot form any sort of hydrogen bond. The presence of an unfulfilled hydrogen bond donor would have a large destabilizing influence on the base pair. Indeed, it had been thought that such a situation could prevent base pairs from forming (Crick, 1966). However, in the system that we propose, this hydrogen atom is involved in hydrogen bonding by sharing a carbonyl oxygen lone pair, and although the energetics of the situation are likely to fall short of a full hydrogen bond, it will considerably

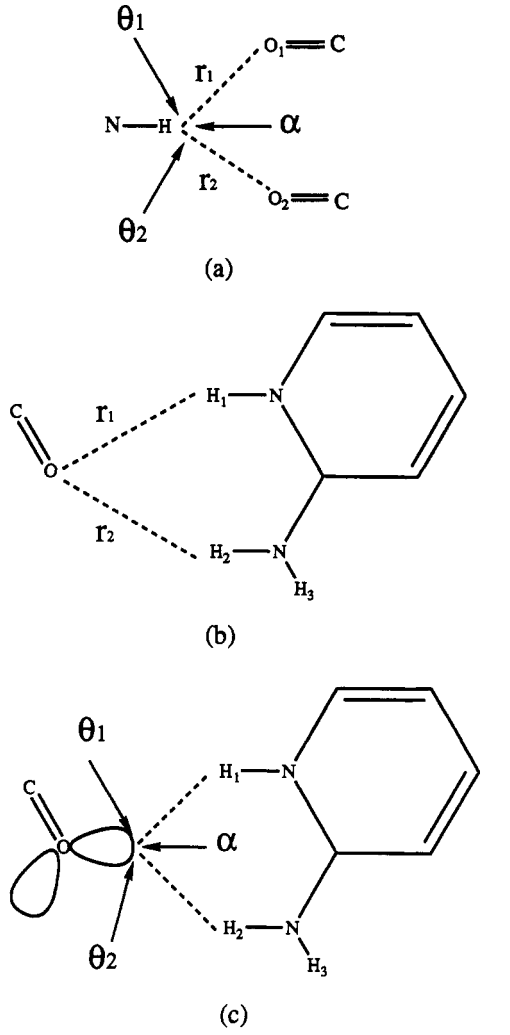

FIGURE 6: (a) Geometrical parameters involved in describing conventional three-centered hydrogen bonding (Taylor et al., 1984; Fritcsh & Westhof, 1991; Jeffrey, 1989). (b and c) Similar parameters applied to a reverse three-centered system. In all cases the geometry should be $r_1 < r_2 < 3$ Å and $\theta_1 + \theta_2 + \alpha \approx 360°$.

Table II: Geometrical Parameters for the Reverse Three-Center Hydrogen-Bonding Systems in the G(anti)·O8A(syn) Base Pairs in the Structure of d(CGCGAATT(O8A)GCG)$_2$[a]

	base pair	
carbonyl oxygen	G(4)·O8A(21)	G(16)·O8A(9)
O6(G)		
r_1 (Å)	1.5	1.7
r_2 (Å)	1.9	2.5
θ_1 (deg)	139	159
θ_2 (deg)	85	108
α (deg)	135	92
$\theta_1 + \theta_2 + \alpha$ (deg)	359	360
O8(O8A)		
r_1 (Å)	2.2	2.1
r_2 (Å)	2.8	2.4
θ_1 (deg)	128	144
θ_2 (deg)	139	137
α (deg)	69	80
$\theta_1 + \theta_2 + \alpha$ (deg)	336	351

[a] The parameters are defined in Figure 7.

alleviate the destabilization of the base pair that is likely to be caused by the presence of a completely unfulfilled hydrogen bond donor.

There are two other non-Watson–Crick base pairs for which the crystal structures are known and where the presence of this reverse three-centered hydrogen-bonding system may help

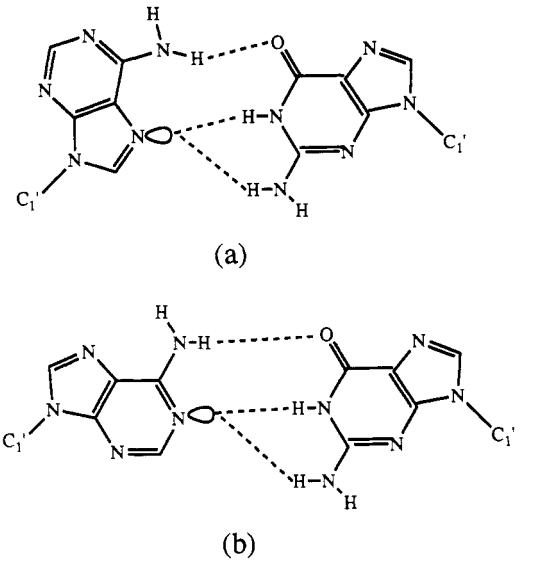

FIGURE 7: Possible reverse three-centered hydrogen-bonding systems for (a) the G(anti)·A(syn) base pair and (b) the G(anti)·A(anti) base pair. Geometrical analysis shows that this system would not exist in the crystal structure of the latter base pair. Refined coordinates for the analysis of the G(anti)·A(anti) base pair were obtained from the Brookhaven database (Abola et al., 1987).

to stabilize the base pair. These are the G(anti)·A(syn) (Brown et al., 1986; Hunter et al., 1986a) and the G(anti)·A(anti) (Privé et al., 1987) base pairs.

In the G(anti)·A(syn) base pair the lone pair at the center of the hydrogen-bonding system would have to belong to the N7 atom of the adenine base (Figure 7a). Our analysis of this base pair suggests that the reverse three-centered hydrogen-bonding system may be present and that, once again, the geometry of the arrangement is close to that expected for a conventional three-centered hydrogen bond with N1H(G)–N7(A) (r_1) and N2H(G)–N7(A) (r_2) distances of 1.9 and 2.3 Å, respectively, and values of $\theta_1 = 157°$, $\theta_2 = 106°$, and $\alpha = 87°$. Therefore, in this base pair also, the potential destabilizing effect of an unfulfilled hydrogen bond donor is again alleviated by the fact that the hydrogen atom in question is involved in a reverse three-centered hydrogen-bonding arrangement and the base pair may be held together by three hydrogen bond interactions and not the two initially proposed from the crystal structure analysis (Brown et al., 1986; Hunter et al., 1986a).

For the G(anti)·A(anti) base pair any reverse three-centered arrangement would have to center on the lone pair of the N1 of adenine (Figure 7b). Our analysis of this base pair showed that while the N1H(G)–N1(A) (r_1) and N2H(G)–N1(A) (r_2) distances of 1.9 and 2.9 Å are within the range for acceptable hydrogen bond formation, the sum of θ_1 (149°), θ_2 (107°), and α (62°) is much less than would be acceptable for three-center hydrogen bond formation (Fritcsh & Westhof, 1991). However, in this base pair there is no need for the formation of a reverse three-centered system as the potential destabilization of the base pair due to an unfulfilled hydrogen bond donor is already alleviated by the fact that the base pairs have a large propeller twist which allows the hydrogen atom in question to form a hydrogen bond with the O2 of the thymine of an adjacent A·T base pair.

A recent NMR study of the G·A mismatch in the DNA duplex d(GCCACAAGCTC)·d(GAGCTGGTGGC) revealed an unusual chemical shift of the imino proton of the guanine involved in G(anti)·A(anti) base pairing (Carbonnaux et al., 1991). This was interpreted as suggesting that, in the

sequence studied, the G(anti)·A(anti) mispair took up a conformation which is similar to that shown in Figure 7b). It was also suggested that this resulted in the formation of interbase bifurcated hydrogen bonds. However, in this system it is the lone pair of N1 of adenine which interacts with two hydrogen bond donors, and therefore the base pairing would have to be the result of the formation of a reverse three-centered hydrogen-bonding system similar to that we have found in our analysis of the crystal structures of the G(anti)·O8A(syn) and G(anti)·A(syn) base pairs.

A common factor of our examination of the G·O8A, G(anti)·A(syn), and G(anti)·A(anti) base pairs and the NMR analysis of the G(anti)·A(anti) in the sequence d(GCCACAAGCTC)·d(GAGCTGGTGGC) is that in all four cases a potential destabilization of the base pairs due to the presence of an unfulfilled hydrogen bond donor has been alleviated either by the involvement of the hydrogen atom in question in a reverse three-centered hydrogen bond or because the hydrogen atom forms a hydrogen bond to a base in an adjacent base pair. This suggests that the alleviation of this potential destabilization has been an important factor in determining the exact conformation of these base pairs.

ACKNOWLEDGMENT

We thank Drs. R. Taylor and F. Allen for their helpful discussions.

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 Cystallography, Cambridge, U.K.

Aida, M., & Nishimura, S. (1987) *Mutat. Res. 192*, 83–89.

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

Brown, T., Leonard, G. A., Booth, E. D., & Chambers, J. (1990) *J. Mol. Biol. 212*, 437–440.

Carbonnaux, C., van der Marel, G. A., van Boom, J. H., Gushlbauer, W., & Fazakerley, G. V. (1991) *Biochemistry 30*, 5449–5458.

CCP4—The Cooperative Computing Project in Crystallography (1986) SERC, Daresbury Laboratory, Warrington, U.K.

Cho, B. P., & Evans, F. E. (1991) *Nucleic Acids Res. 19*, 1041–1047.

Conley, J. J. (1963) *Nature 197*, 555.

Corfield, P. W. R., Hunter., W. N., Brown, T., Robinson, P., & Kennard, O. (1987) *Nucleic Acids Res. 15*, 7935–7949.

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

Culp, S. J., Cho., B. P., Kadlubar, F. F., & Evans, F. E. (1989) *Chem. Res. Toxicol. 2*, 416–422.

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

Fritcsh, V., & Westhof, E. (1991) *J. Am. Chem. Soc. 113*, 8271–8277.

Guschlbauer, W., Duplaa, A.-M., Guy, A., Teoule, R., & Fazakerley, G. V. (1991) *Nucleic Acids Res. 19*, 1753–1758.

Guy, A., Duplaa, A.-M., Harel, P., & Téoule, R. (1988) *Helv. Chim. Acta 71*, 1566–1572.

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.

Hunter, W. N., Brown, T., & Kennard, O. (1986a) *J. Biomol. Struct. Dyn. 4*, 173–191.

Hunter, W. N., Brown, T., Anand, N. N., & Kennard, O. (1986b) *Nature 320*, 552–555.

Jeffrey, G. A. (1989) *Landolt–Bornstein, Numerical Data and Functional Relationships in Science and Technology, Group VII, Vol. 1, Nucleic* Acids, *Subvol. b, Crystallographic and Structural Data II* (Saenger, W., Ed.) pp 277–342, Springer-Verlag, Berlin-Heidleberg.

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

Kouchakdjian, M., Bodepudi, V., Shibutani, S., Eisenberg, M., Johnson, F., Grollman, P., & Patel, D. J. (1991) *Biochemistry 30*, 1403–1412.

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

Lehman, M. S., & Larsen, F. K. (1974) *Acta Crystallogr. A30*, 580–584.

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

North, A. C. T., Phillips, D., & Matthews, F. S. (1968) *Acta. Crystallogr. A24*, 580–584.

Oda, Y., Uesegi, S., Ikehara, M., Nishimura, S., Kawase, Y., Ishikawa, K., Inoue, H., & Ohtsuka, E. (1991) *Nucleic Acids Res. 19*, 1407–1412.

Privé, G. G., Heinemann, U., Kan, L. S., Chandrasegaran, S., & Dickerson, R. E. (1987) *Science 238*, 498–504.

Rosenberg, J. M., Seeman, N. C., Day, R. O., & Rich, A. (1976) *J. Mol. Biol. 104*, 145–167.

Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Kim, J. J. P., & Rich, A. (1976) *J. Mol. Biol. 104*, 109–144.

Sheldrick, G. M. (1976) SHELX76 System of Computer Programs, University of Cambridge, U.K.

Shibutani, S., Takeshita, M., & Grollman, P. (1991) *Nature 349*, 431–434.

Taylor, R., Kennard, O., & Verishel, W. (1983) *J. Am. Chem. Soc. 105*, 5761–5766.

Taylor, R., Kennard, O., & Verishel, W. (1984) *J. Am. Chem. Soc. 106*, 244–248.

Téoule, R. (1987) *Int. Radiat. Biol. 51*, 573–589.

Van Hemmen, J. J., & Bleichrodt, J. J. (1971) *Radiat. Res. 46*, 444–456.

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