Structure of a New Crystal Form of a DNA Dodecamer Containing T·(O^6Me)G Base Pairs[†,‡]

Jaroslav Vojtechovsky,[§] Mark D. Eaton,[∥] Barbara Gaffney, Roger Jones, and Helen M. Berman*

Department of Chemistry, Rutgers University, Piscataway, New Jersey 08855-0939

Received August 1, 1995; Revised Manuscript Received October 3, 1995[⊗]

ABSTRACT: The structure of the synthetic dodecamer {d[CGTGAATTC(O^6Me)GCG]}$_2$ has been determined to a resolution of 2.25 Å and refined to a final *R* factor of 16.7%. The volume of the unit cell is significantly smaller by 16% than the original Drew and Dickerson parent dodecamer [Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78,* 7318−7322]. The double helix is in a different position in the unit cell, rotated by −85.9°, and translated by 9.9 Å around the helical axis with respect to the parent structure. The intermolecular arrangement of helices, characterized by double hydrogen bonded guanine−guanine minor groove interactions, remains conserved. The molecular geometry exhibits several significant changes that are related to the changed position of the helix and the presence of two mismatched base pairs with O^6-methylguanine. Both mispairs are found in a symmetrical T(*anti*)·(O^6Me)G(*anti*) conformation, and the methyl groups are in proximal orientation. The hydration pattern of the structure is different and can be related to changes in the minor groove geometry. An incorrect model that was isomorphous to the parent dodecamer could be refined to a low *R* factor. Characteristics of the refinement and of the geometry that are indicative of incorrect structures have been analyzed.

DNA sequences containing O^6-alkylated guanines are of interest because of their mutagenic activity. The modified guanine can be involved in mispairing and result in several types of mutations (Balmain & Prangell, 1983; Eva & Aaronson, 1983). It is also a carcinogenic lesion as demonstrated by the occurrence of brain tumors in rats exposed to the alkylating agent *N*-ethyl-*N*-nitrosourea (Goth & Rajevsky, 1974). There is a preferred pairing of O^6-methylguanine to thymine although other pairings have been observed (Abbot & Saffhill, 1979; Kamiya et al., 1992). To induce mutations, the T·(O^6Me)G mispair must be able to escape the proofreading process with a high efficiency.

Significant attention has been focused on understanding postreplicative repair mechanisms. There are two known pathways for correcting T·G mispair in DNA. The "short patch" repair is initiated by a specific thymine-DNA glycosylase (Wiebauer & Jiricny, 1989). "Long patch" repair consists of sequence independent excision and resynthesis of a relatively long stretch of DNA (Brown & Jiricny, 1988; Holmes et al., 1990; Thomas et al., 1991). The failure of the later repair process is more likely involved in mediating the cytotoxicity of O^6-methylguanine (Griffin et al., 1994). In the same study, a reduced affinity of binding to T·(O^6-Me)G with respect to T·G and C·(O^6Me)G was observed, further supporting the importance of this type of mispairing.

In an effort to understand the structural basis for these observations, there have been several solution and crystallographic analyses of nucleosides and oligonucleotide sequences involving O^6-methylguanine mispairing.

Proton NMR experiments using lipophilic nucleoside derivatives in nonaqueous solvent (Williams & Shaw, 1987) found that O^6-methylguanine forms hydrogen bonds with thymine and with protonated cytosine, but no interaction with unprotonated cytosine was detected. The preference of T·(O^6Me)G pairing over C·(O^6Me)G in this study is consistent with the high occurrence of the G→A transition.

In the crystal structure of a Z-DNA hexamer {d[CGC(O^6Me)GCG]}$_2$, the base pairs adopt a Watson−Crick geometry (Ginell et al., 1990). This geometry is consistent with the protonation of one of the bases. A similar study of B-DNA structures of {d[CGC(O^6Et)GAATTCGCG]}$_2$ (Sriram et al., 1992) complexed to various minor groove binding drugs revealed two different conformations. One involves wobble pairing, while the other is an opened bifurcated pairing, with N2 of O^6-methylguanine bonded to both O2 and N3 of cytosine. An additional reverse wobble pairing scheme was observed in the crystal structure of the same sequence complexed to SN6999 (Gao et al., 1993). In solution, ^{15}N NMR spectroscopic results demonstrated wobble base pairing at pH >6; a protonated symmetrical Watson−Crick scheme occurred at lower pH (Gaffney et al., 1993).

Similar variability has been reported for A·(O^6Me)G mispairing in studies of {d[CGAGAATTC(O^6Me)GCG]}$_2$. The O^6-methylguanine pairs to adenine with an A(*syn*)·(O^6-Me)G(*anti*) conformation in the crystal structure (Ginell et al., 1994), while an A(*anti*)·(O^6Me)G(*anti*) conformation was observed in ^{1}H and ^{31}P NMR spectroscopic studies of the same sequence (Patel et al., 1986a). This geometry would require one of the bases to be protonated in order to form a two hydrogen bond base pair.

[†]This work was funded by NIH Grants GM21589 (H.M.B.) and GM31483 (R.J.).

[‡]Coordinates have been deposited with the Brookhaven Protein Data Bank under the file name 218D.

* To whom correspondence should be addressed.

[§] Current address: MPI for Molecular Physiology, Postfach 102664, 44026 Dortmund, Germany.

[∥] Current address: Department of Medicinal Chemistry, P.O. Box 908540, Virginia Commonwealth University, Richmond, VA 23298-0540.

[⊗] Abstract published in *Advance ACS Abstracts,* December 1, 1995.

0006-2960/95/0434-16632$09.00/0

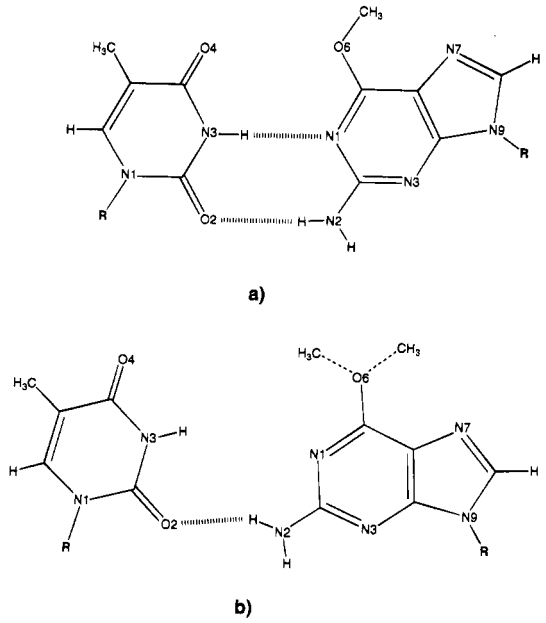

FIGURE 1: Hydrogen bonding schemes of T·(O^6Me)G: (a) Watson–Crick-like symmetrical T(*anti*)·(O^6Me)G(*anti*) conformation; (b) an opened T(*anti*)·(O^6Me)G(*anti*) conformation, as observed by the NMR spectroscopy of the title sequence (Patel et al., 1986b; Goswami et al., 1993). In this case, the methyl group is drawn in both possible orientations: proximal (the torsion angle C5-C6-O6-C6M ~0°) and distal (the torsion angle C5-C6-O6-C6M ~180°).

As shown in Figure 1a, T and (O^6Me)G can form a Watson–Crick-like symmetrical base pair without the protonation required for symmetrical C·(O^6Me)G base pairs. This type of geometry could cause misreplication and escape the proofreading process. One crystal structure of a B-DNA duplex containing two T·(O^6Me)G base pairs in a Watson–Crick orientation has been determined: {d[CGC(O^6Me)-GAATTTGCG]}$_2$ (Leonard et al., 1990). ^{1}H and ^{31}P NMR spectroscopy of the title sequence, {d[CGTGAATTC(O^6Me)-GCG]}$_2$, suggests, that, in solution, the hydrogen bond between N1 of O^6-methylguanine and N3 of thymine was either larger than usual or absent (Patel et al., 1986b) (Figure 1b). The same result was confirmed by a recent ^{15}N NMR study (Goswami et al., 1993). In order to bring further insight into the structural consequences of T·(O^6Me)G base pairing, we have determined the crystal structure of the title sequence.

EXPERIMENTAL PROCEDURES

Crystals were grown at 23 °C by the technique of vapor diffusion from hanging drops with a total volume of 10 μL. The drops, containing 1.0 mM DNA, 20 mM sodium cacodylate buffer, 20 mM $MgCl_2$, and 15% 2-methyl-2,4-pentanediol (MPD), were equilibrated against 30% MPD. The crystals were highly polymorphic. All crystals diffracted in orthorhombic space group $P2_12_12_1$. The cell dimensions ranged for *a* between 21.2 and 23.0 Å, for *b* between 35.0 and 40.3 Å, and for *c* between 63.5 and 65.8 Å. The individual cell dimensions were within the range seen for other dodecamer structures that had been determined to be isomorphous to the parent sequence {d[CGCGAATTCGCG]}$_2$ (Drew et al., 1981). The best data set was collected at room temperature, with an Enraf Nonius CAD4 diffractometer on an Enraf Nonius 571 rotating anode generator. A graphite monochromator was used. A total of 4218 reflections to the resolution 1.97 Å were measured using the omega-theta type scan, of which 2387 were observed [above $2\sigma(F)$]. The ratio of observed reflections dropped below 50% in the shell between 2.3 and 2.2 Å. Data up to 2.25 Å were used for the refinement. The crystal data and data collection statistics are given in Table 1.

Table 1: Crystal Data and Data Collection Statistics

Crystal Data	
contents of asymmetric unit	
strands of d(CGTGAATTC(O^6Me)GCG)	2
total molecular weight of DNA duplex (daltons)	7328.8
V_m (Å^3/dalton)	1.89
estimated hydration (%)	32
space group	$P2_12_12_1$
unit cell	
a (Å)	22.12(1)
b (Å)	38.81(2)
c (Å)	64.48(2)
$\alpha = \beta = \gamma = 90°$	
volume (Å^3)	55 354
Data Collection Statistics	
crystal size (mm)	0.30 × 0.20 × 0.15
temperature	−5 °C
crystal mounting	capillary
data collection device	Enraf Nonius CAD4
radiation	Cu Kα
unique data collected	4218
completeness	99.5%
reflections $F > 2\sigma(F)$	2387
resolution limit (Å)	1.97
effective resolution (Å)	2.25
% reflections with $F > 2\sigma(F)$ (<2.25 Å)	72.1%

The first attempt to determine the structure was done using the coordinates of the parent sequence {d[CGCGAAT-TCGCG]}$_2$ (Drew et al., 1981). The rotation results using the program MERLOT (Fitzgerald, 1988, 1991) were consistent with a helix in the same orientation. The translation results were inconclusive. It was decided to proceed with the refinement using the Drew and Dickerson model. The least-squares refinement, using the programs CORELS (Sussman et al., 1977) and NUCLSQ (Westhof et al., 1985), resulted in an *R* factor of 0.171 and a correlation factor of 0.94. The fit in the electron density maps was acceptable. However, the rms values were high, and data in the shell 6–10 Å exhibited poor statistics. In addition, some of the backbone torsion angles deviated significantly from expected values. Therefore, a complete molecular replacement search was carried out using the program ULTIMA (Rabinovich & Shakked, 1984), which incorporates packing criteria and *R* factor calculations in the multidimensional search. One outstanding solution was found which corresponds to a shift of the original model by 9.9 Å along the helix axis and rotation around this axis by −87°. The calculated *R* factor for the rigid helix model of the title structure was 0.35.

The C·G pairs at steps 3 and 10 were replaced by a model of T(*anti*)·(O^6Me)G(*anti*). The helix was broken into 36 rigid groups and the structure refined to an *R* factor 0.26. Several difference electron density maps, with the T·(O^6Me)G pairs omitted from the refinement and structure factor calculations, were used to determine the correct conformation. They clearly revealed that T(*anti*)·(O^6Me)G(*anti*) is the best conformation to fit the maps in both cases. The positional least squares refinement was done using the program *NUCLSQ*. After several cycles of refinement, the

Table 2: Final Refinement Statistics

structure factors information	
resolution range	10−2.25 Å
number of reflections [$F > 2\sigma(F)$]	2040
final R factor[a]	16.7%
correlation factor	0.950
σ on F_o map (eÅ^{-3})	0.55
highest $F_o - F_c$ peak (eÅ^{-3})	0.13
lowest $F_o - F_c$ peak (eÅ^{-3})	−0.13
σ on $F_o - F_c$ map (eÅ^{-3})	0.10
geometry restraints	rmsd
bond distances (Å)	0.013
bond angles (deg)	1.80
dihedral angles (deg)	19.6
improper angles (deg)	1.82

[a] $R = \Sigma|F_o - F_c|/\Sigma F_o$; correlation factor is $\Sigma[(F_o - <F_c>)(F_c - <F_c>)]/[\Sigma(F_o - <F_o>)2\Sigma(F_c - <F_c>)^2]^{1/2}$.

structure converged to a final R factor of 0.167 and a correlation factor of 0.95. At the very end of refinement, simulated annealing and positional refinement using the new set of parameters (Parkinson et al., 1996) and the program X-PLOR (Brünger et al., 1987) were used. The final R factor is 0.167, the overall bond distance rms deviation is 0.013 Å, and overall angle rms deviation is 1.80°. There are a total of 74 solvent molecules located in the structure, 12 of which are in the second hydration shell. Only those within the range 2.3−3.2 Å from the potential hydrogen bonding partner were accepted. A summary of the refinement statistics are listed in Table 2. The final coordinates and structure factors have been deposited in the Protein Data Bank.

After the structure completely converged, the refinements of both models were redone in parallel in order to ascertain benchmarks of correct and incorrect structures. A comparison of refinement statistics of the two models is given in Table 3. A discussion of these results is given later.

Molecular replacement and the preliminary refinement with three other datasets from polymorphic crystals were carried out. In spite of the large differences in the cell dimensions, the crystal arrangement is the same as presented here for the highest resolution data set.

RESULTS AND DISCUSSION

The structure of {d[CGTGAATTC(O^6Me)GCG]}$_2$ forms a duplex in a B-DNA conformation. The rms deviation between this structure and the parent dodecamer {d[CGCGAATTCGCG]}$_2$ (Drew et al., 1981) is 1.2 Å with the mismatched bases omitted from the calculations. The superposition of the helix and the parent molecule is shown on Figure 2. In general, the pattern of groove widths is similar to what has been seen in other dodecamers, although the minor groove is more narrow at one end of the duplex (Figure 3).

The base pair morphology calculated using the program NEWHELIX93 (Dickerson et al., 1984) reveals some differences from the parent structure. As will be discussed later, many of these changes are correlated with differences in packing and hydration. The values for roll, propeller twist, and X and Y displacement (Dickerson et al., 1989) are drawn in Figure 4. There are deviations from the parent dodecamer in the values for roll at steps 3, 4, 8, and 9 at or close to the T·(O^6Me)G mispairs. There is a higher propeller twist for both mismatched pairs. A comparable deviation occurs in the terminal base pair G12·C13. There are significant changes in Xdsp in base pairs C1·G24 and G12·C13. The Ydsp differs for base pairs C1·G24 and G2·C23.

The base pair openings of the mismatches calculated with respect to the reference standard C·G geometry (Babcock & Olson, 1992) for T3·(O^6Me)G22 and T15·(O^6Me)G10 are higher than in the corresponding C·G base pairs of the parent structure (see Figure 4). In addition, base pairs A5·T20, A6·T19 and T7·A18 have significantly different values for the openings than in the original dodecamer.

As was seen in the other crystal structure containing T·(O^6Me)G (Leonard et al., 1990), the base pairs T3·(O^6Me)G22 and T15·(O^6Me)G10 are in a "Watson−Crick-like" orientation. The final annealed $F_o - F_c$ omit maps as calculated using the program X-PLOR (Brünger et al., 1987) are drawn in Figure 5. The distances between hydrogen bonding heteroatoms are similar to the values observed in the structure of {d[CGC(O^6Me)GAATTTGCG]}$_2$ (Leonard

Table 3: A Comparison of the Refinement of the "Isomorphous" Model and the Correct Model[a]

refinement type	resolution (Å) F cut off	statistics R (%)	k	N	rmsd of bond and angle distances (Å): base−sugar bond distance	base−sugar angle distance	phosphate bond distance	phosphate angle distance	rmsd of nonbonded contacts (Å): single-torsion contacts	multiple-torsion contacts	hydrogen bonds
rigid helix	15−5	**53.0**	**0.27**	**0**							
CORELS	$F > 4\sigma(F)$	41.7	0.61	0							
36 groups	15−5	**41.2**	**0.57**	**1**	**0.39**	**0.20**	**0.40**	**0.63**			
CORELS	$F > 4\sigma(F)$	32.6	0.72	1	0.39	0.19	0.39	0.64			
36 groups	8−3	**39.9**	**0.57**	**11**	**0.47**	**0.31**	**0.50**	**0.69**			
CORELS	$F > 4\sigma(F)$	26.3	0.84	2	0.38	0.18	0.38	0.65			
positional	8−2.9	**35.2**	**0.69**	**15**	**0.021 (60)**	**0.047 (80)**	**0.055 (80)**	**0.068 (90)**	**0.16 (20)**	**0.15 (20)**	**0.18 (20)**
NUCLSQ	$F > 4\sigma(F)$	21.0	0.91	1	0.016 (60)	0.035 (80)	0.042 (80)	0.047 (90)	0.11 (20)	0.13 (20)	0.15 (20)
+B-factors	8−2.9	**29.3**	**0.81**	**107**	**0.035 (50)**	**0.074 (70)**	**0.080 (60)**	**0.115 (90)**	**0.16 (20)**	**0.19 (20)**	**0.25 (20)**
NUCLSQ	$F > 4\sigma(F)$	17.9	0.95	37	0.024 (50)	0.047 (70)	0.053 (60)	0.077 (90)	0.14 (20)	0.16 (20)	0.18 (20)
+base waters	8−2.9	**22.6**	**0.90**	**99**	**0.031 (50)**	**0.067 (70)**	**0.066 (60)**	**0.102 (90)**	**0.18 (20)**	**0.18 (20)**	**0.23 (20)**
NUCLSQ	$F > 4\sigma(F)$	16.1	0.96	19	0.023 (50)	0.043 (70)	0.050 (60)	0.071 (90)	0.13 (20)	0.16 (20)	0.18 (20)
+all waters	6−2.3	**17.1**	**0.94**	**124**	**0.022 (30)**	**0.056 (50)**	**0.063 (50)**	**0.086 (70)**	**0.19 (20)**	**0.20 (20)**	**0.20 (20)**
NUCLSQ	$F > 2\sigma(F)$	14.7	0.96	63	0.015 (30)	0.039 (50)	0.055 (50)	0.068 (70)	0.12 (20)	0.15 (20)	0.20 (20)
final model	10−2.25	**26.1**	**0.85**	**216**	**0.020 (20)**	**0.033 (25)**	**0.041 (25)**	**0.043 (30)**	**0.22 (25)**	**0.25 (25)**	**0.28 (25)**
NUCLSQ	$F > 2\sigma(F)$	16.7	0.95	78	0.014 (20)	0.026 (25)	0.020 (25)	0.028 (30)	0.12 (25)	0.17 (25)	0.20 (25)

[a] The R factor and correlation factor (k) were calculated from the same formulas as in Table 2. N is the number of bond and angle distances deviating from ideality by more than 2σ. The σ values used to constrain the bond distances, angle distances, and nonbonded contacts are in parentheses.

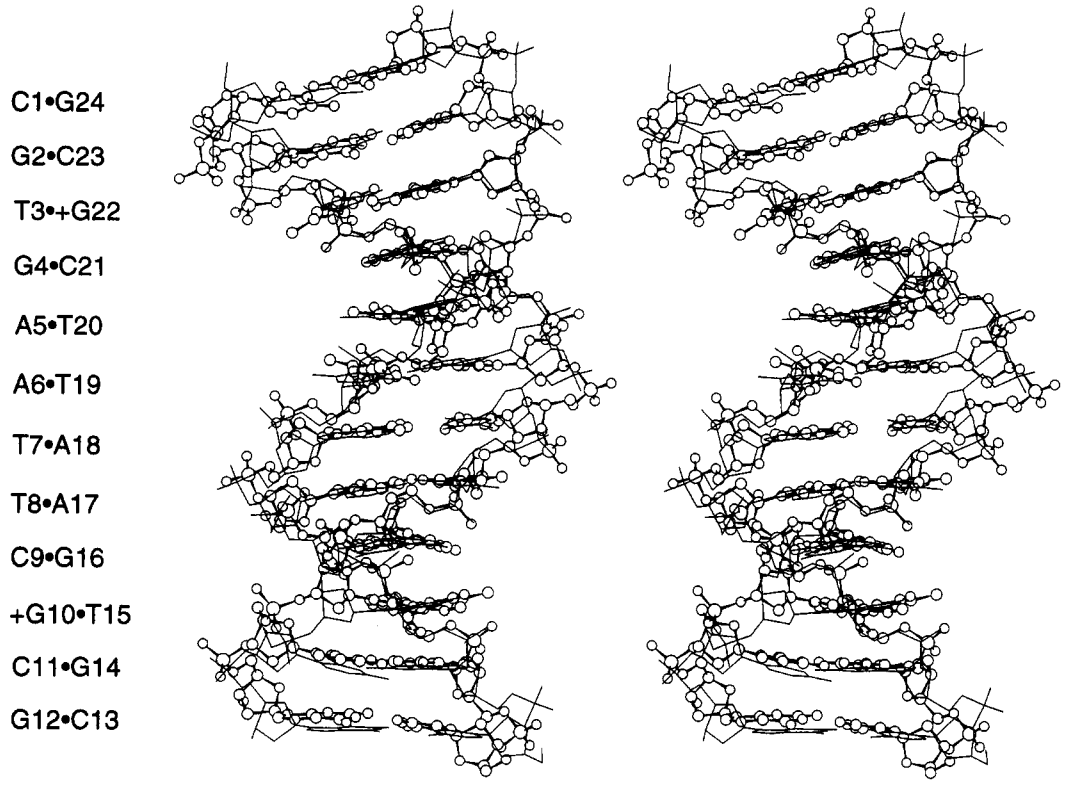

FIGURE 2: Superposition of the title (ball and stick) and the parent structure (solid line; Drew et al., 1981) B-DNA helix. The T·(O^6Me)G were omitted when calculating the least-squares fit. The residue numbering corresponding to the title structure is shown on the left. O^6-methylguanine is labeled as +G.

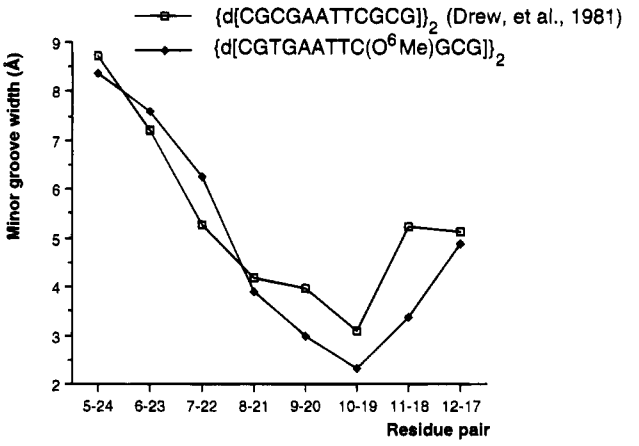

FIGURE 3: Graph of the minor groove width (Å), calculated as the interatomic P···P distance less the sum of the van der Walls radii (5.2 Å).

et al., 1990) in Table 5a. There are two normal hydrogen bonds in the base pairs as well as close interactions between the O6 of (O^6Me)G and O4 of T. The methyl groups of the two modified guanines are in the proximal orientation; the torsion angle, C5-C6-O6-C6M, is 20° for T3·(O^6Me)G22 and −11° for T15·(O^6Me)G10.

The helix is shifted by 9.9 Å and rotated around the *c* axis by −85.9° as calculated from the superposition of the final refined structure and the parent dodecamer (Drew et al., 1981). This shift corresponds to the rise of three base pairs and rotation around the helical axis of the dodecamer. Thus, the strongest reflections resulting from the base stacking of the semi-infinite helix exhibit a very similar diffraction pattern to the parent structure. There are three other examples in which the B-DNA dodecamers are not in the same position as the original dodecamer. Their crystal data and the shift and rotation of the helix with respect to its position in the parent structure are listed in Table 4. In the structure of {d[CGCGAATT^{Br}CGCG]}$_2$ (Fratini et al., 1982), rotation of the helix and reduced bending were observed after a change of the crystallization temperature. In the structure of the nicked dodecamer d[CGCGAAAA-CGCG] + d[CGCGTT] + d[TTCGCG] (Aymami et al., 1990) the helix is repositioned, but the major features of the molecular geometry are similar to the parent sequence. In the structure of the drug−DNA complex {d[CGC(O^6Et)-GAATTCGCG]}$_2$−SN6999 (Gao et al., 1993) that contains O^6-ethylated guanine, the bent DNA duplex could not adopt the same crystal packing as the parent structure. In contrast, the volume of the unit cell containing this sequence is significantly lower. In the structures shown in Table 4 the cell volume deviates by more than 6%. The title structure and the structure of d[CGCGAAAACGCG] + d[CGCGTT] + d[TTCGCG] (Aymami et al., 1990) deviate by more than the 10% threshold suggested in studies of isomorphism (Kitaigorodsky, 1984; Haget et al., 1984). Although there are several B-DNA structures in which the individual cell dimensions show large deviations from the parent, there are no other structures with comparable deviations of cell volume.

In spite of the shift and rotation, the packing of the helices has similar features as observed in other B-DNA dodecamers (Figure 6). The DNA molecules form a "semi-infinite helix" along *c* and join each other by double hydrogen bonded guanines in their minor grooves. These intermolecular contacts involve the molecules related by the two-fold screw along the *c* axis. In order to determine whether the intermolecular packing influences the helix geometry, the

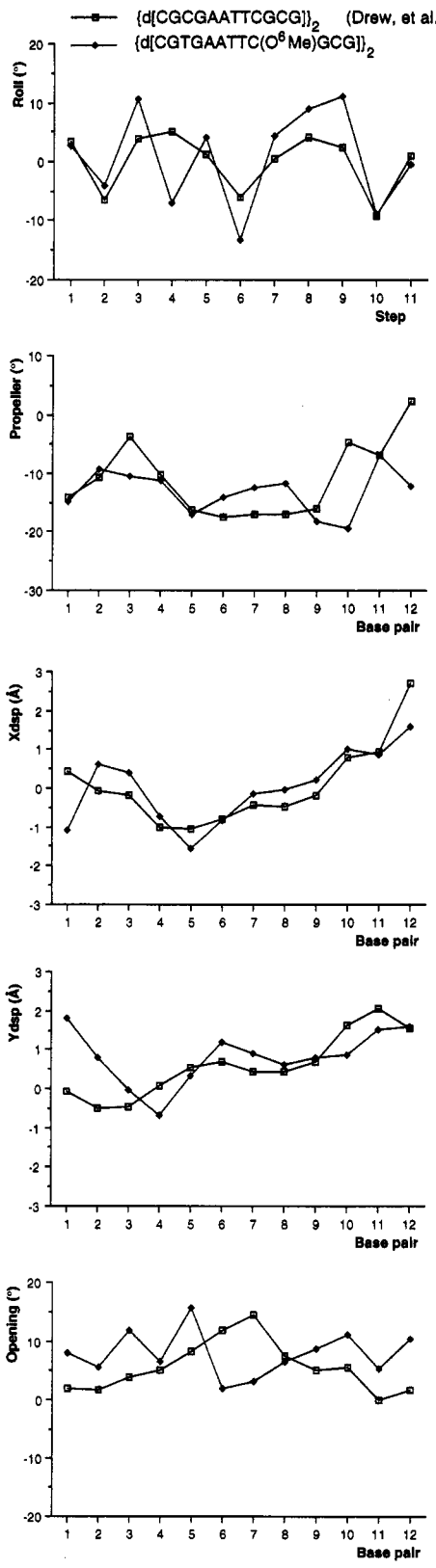

FIGURE 4: Helix parameters, as calculated by the program NEWHELIX93 (Dickerson et al., 1984). The steps are numbered from C1·G24:G2·C23 to C11·G14:G12·C13, and the base pairs are numbered from C1·G24 to G12·C13. The base pair openings were obtained by algorithm of Babcock and co-workers (Babcock & Olson, 1992). The values for T·(O^6Me)G were calculated with respect to the Watson–Crick C·G standard geometry.

parent structure (Drew et al., 1981) was superimposed on the title molecule in the new position. In the resulting model, the guanine–guanine minor groove contacts were elongated such that they do not form hydrogen bonds. The values for the *X* and *Y* displacements of the base pairs C1·G24, G2·C23, and G12·C13 in the title structure allow these hydrogen bonds to occur. The importance of maintaining this guanine–guanine interaction is also demonstrated in an unusual helix–helix junction in a DNA decamer in which bases are flipped out of the duplex and form intermolecular minor groove guanine–guanine hydrogen bonds (Spink et al., 1996).

The details of the intermolecular interactions in this structure are compared with the structures of {d[CGCGAATTCGCG]}$_2$ (Drew et al., 1981) and {d[CGC(O^6Me)GAATTTGCG]}$_2$ (Leonard et al., 1990) in Table 5b. The minor groove contacts involving the bases of G2 and G12 are maintained in all three structures. The contact involving the sugar O4′ is seen only in the T·(O^6Me)G structure reported here. The minor groove contacts involving the bases of G14 and G24 that were observed in the parent dodecamer are elongated for three of the four cases for both methylated structures. Both G16, which is methylated in the structure by Leonard et al., and G22, which is methylated in this structure, form intermolecular contacts with O3′ atoms in all three structures. Thus, despite differences in the position in the sequence in the two structures, these methylated residues are involved in similar types of interactions in the crystalline state.

There are two contacts that reflect the differences in crystal packing between the present structure and other dodecamers. N4 of base C1 and O2P of a symmetry related (O^6Me)G22 are within hydrogen bonding distance; in the parent dodecamer and in the structure by Leonard et al. they are 24 Å apart. The contact between O5′ of C1 and O1P of T8 observed in other dodecamers is not seen in the title structure. These differences are a reflection of the differences in lateral interactions between helices caused by the shift.

The spine of hydration within the minor groove has a different pattern in this structure as shown in Figure 7. Three solvent molecules are observed in the central -AATT- area, and there is a solvent bridge between a modified base (O^6-Me)G10 in one strand and G6 in the other. Two of the minor groove solvent molecules in the title structure are hydrogen bonded to one base in the -AATT- region, whereas the spine of hydration in the parent dodecamer contains three interstrand solvent bridges in the central area. This change in hydration geometry is correlated with the differences observed for the base pair openings of base pairs A5·T20, A6·T19, and T7·A18 and with the narrowing of the minor groove. The wide side of the minor groove does not have a bridge involving the modified (O^6Me)G22.

The solvent pattern on the major groove side of the methylated guanine is clearly influenced by the O^6-methyl group. Instead of two solvent molecules associated with O6 and N7 found in unmodified guanines (Schneider et al., 1993), there is one solvent molecule close to N7 and C6M of both O^6-methylguanines (see Figure 5). No solvent molecules were located close to O4 of mispaired thymines—an atom that is usually strongly hydrated.

Most of the 66 B-DNA dodecamer structures have been refined by direct replacement of the original {d[CGCGAATTCGCG]}$_2$ double helix (Drew et al., 1981). Similar isomorphism has been observed in other DNA systems studied by X-ray structure analysis. As our studies show, however, one cannot depend on this isomorphism in the structure determination of a new DNA structure, no matter

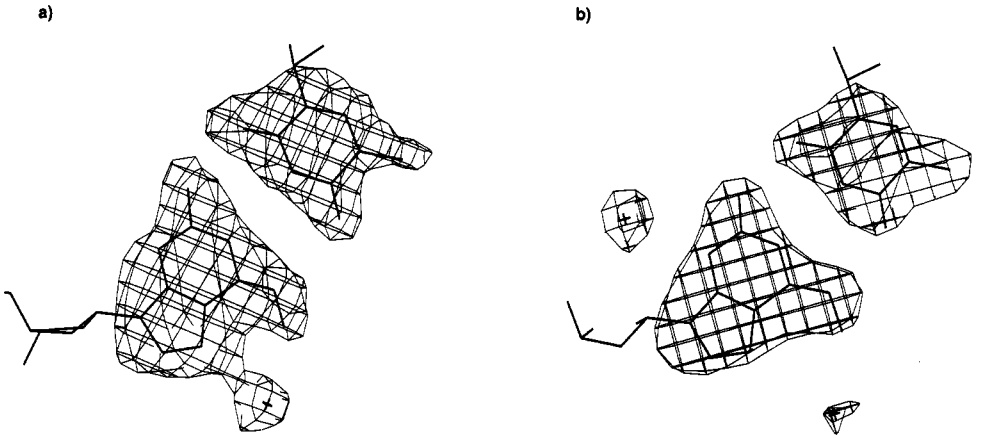

FIGURE 5: "Annealed omit" $F_o - F_c$ map calculated using the strategy developed in the program X-PLOR (Brünger, 1992) with T·(O^6-Me)G omitted from the refinement and calculation: (a) T3·(O^6Me)G22, (b) T15·(O^6Me)G10. Both maps are drawn at 2.5σ level. The crosses indicate the solvent molecules interacting with O^6-methylguanine.

Table 4: Dodecamer DNA Structures in Which the Helix Is in a Different Position Than in d(CGCGAATTCGCG)$_2$ (Drew et al., 1981)[a]

	cell dimensions (Å)				translation (Å)			rotation (deg)			
NDB code	*a*	*b*	*c*	volume (Å^3)	t_a	t_b	t_c	ϕ_a	ϕ_b	ϕ_c	reference
BDL001	24.8	40.3	66.2	66 163							Drew et al. (1981)
BDLB04	24.2	40.1	64.0	62 107	−0.6	0.5	−2.0	−5.3	2.4	0.0	Fratini et al. (1982)
BDL021	26.0	44.0	66.6	76 190	1.1	2.7	7.1	1.4	−0.7	−6.2	Aymami et al. (1990)
GDLB24	28.5	36.1	69.6	71 608	4.2	−3.0	37.4	1.2	0.0	15.4	Gao et al. (1993)
title structure	22.1	38.8	64.5	55 355	2.0	3.1	9.9	1.2	2.5	−85.9	this paper

[a] The position of the helix with respect to the position in the parent structure is given by the translation vector (Å) and by rotation angles around the crystallographic axes *a*, *b*, and *c* (deg).

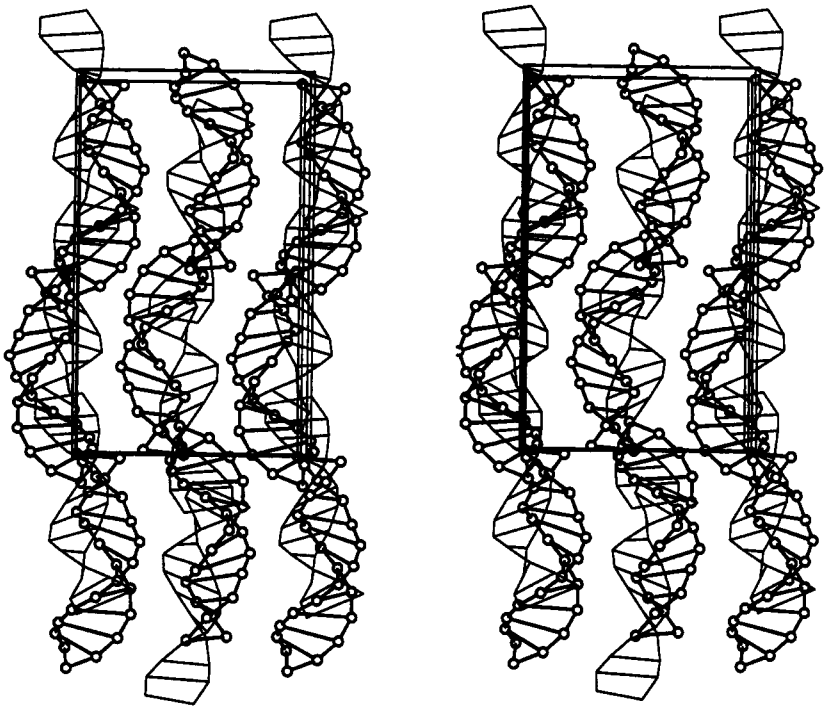

FIGURE 6: Crystal packing of B-DNA double helices of the title structure (ball and stick) as compared to the parent structure (solid line; Drew et al., 1981). Only the C1′ atoms are displayed.

how similar the individual cell dimensions may appear. The title structure was refined as "isomorphous" to a statistical *R* factor of 17.1% and a correlation factor of 0.94. The electron density was reasonable in most parts of the structure, including the mispaired O^6-methylated guanine. In order to provide benchmarks for refinement of correct and incorrect structures, the refinements were repeated in parallel. The statistics are given for each major refinement step in Table 3. In each case, the first line shows the statistics for the incorrect "isomorphous" model. The second line gives the statistics for the correct model.

For the beginning stages of refinement, virtually every indicator is better for the structure reported than for the "isomorphous" model. The *R* factor is significantly lower and the correlation factor much higher for the correct model. However, the differences begin to become smaller as *B*

Table 5: (a) Hydrogen Bonding Geometry of T·(O^6Me)G in the Title Structure (T·+G) and in {d[CGC(O^6Me)GAATTTGCG]}$_2$ (+G·T; Leonard et al., 1990), and (b) Intermolecular Contacts d(i − j) in the Title Structure (T·+G) As Compared to Equivalent Distances in the Structure of {d[CGC(O^6Me)GAATTTGCG]}$_2$ (+G·T; Leonard et al., 1990) and the Parent Structure (C·G; Drew et al., 1981)[a]

(a) Base Pair Hydrogen Bonding Geometry

i		*j*			*i*		*j*		
atom	residue	atom	residue	d($i-j$) (Å)	atom	residue	atom	residue	d_{ij} (Å)
				T·+G					
O2	T3	N2	+G22	2.8	O2	T15	N2	+G10	2.6
N3	T3	N1	+G22	3.2	N3	T15	N1	+G10	3.0
O4	T3	O6	+G22	3.3	O4	T15	O6	+G10	3.1
				+G·T					
O2	T21	N2	+G4	2.7	O2	T9	N2	+G16	2.9
N3	T21	N1	+G4	3.0	N3	T9	N1	+G16	2.9
O4	T21	O6	+G4	3.1	O4	T9	O6	+G16	2.8

(b) Intermolecular Crystalline Contacts

i		*j*		d($i-j$) (Å)			
atom	residue	atom	residue	T·+G	+G·T	C·G	symmetry operation
				Minor Groove Contacts			
N3	G2	N2	G12	3.0	3.1	2.8	i
N2	G2	N3	G12	2.8	3.1	2.9	i
N2	G2	O4′	G12	3.1	4.1	4.1	i
N3	G14	N2	G24	3.6	3.7	3.2	ii
N2	G14	N3	G24	3.1	3.8	3.0	ii
O3′	G12	N2	G22	2.6[b]	2.6	2.3	ii
O3′	G12	O4′	C23	3.2	4.3	6.0	ii
O3′	G12	O2	C23	2.8	3.7	3.6	ii
N3	G16	O3′	G24	2.8	3.1	3.0	ii
O4′	A17	O3′	G24	3.2	3.2	3.1	ii
				Other Contacts			
N4	C1	O2P	G22	3.0	24.7	24.2	iii
O5′	C1	O1P	T8	17.3	3.1	2.7	iv

[a] Contacts within 3.2 Å range in one of the structures containing T·(O^6Me)G are listed. The symmetry operations are (i) $-x+1.5$, $-y+1.0$, $z+0.5$, (ii) $-x+1.5$, $-y+1.0$, $z-0.5$, (iii) $x+0.5$, $-y+1.5$, $-z+1.0$, (iv) $-x+1.0$, $y+0.5$, $-z+0.5$. [b] Contacts involving a methylated base are underlined.

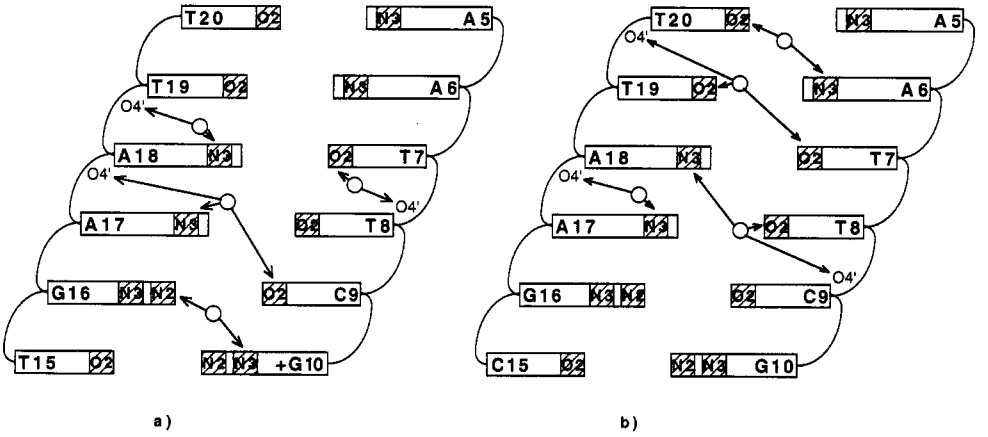

FIGURE 7: Schematic showing the minor groove hydration for (a) the title structure and (b) the parent structure (Drew et al., 1981). The contacts between solvent molecules and bases shorter than 3.2 Å are displayed as arrows. The solvent molecules are shown as circles. O^6-methylguanine is labeled as +G.

factors are refined and water molecules are added. Removal of the low resolution shell results in an acceptable R factor and correlation factor for incorrect structure (17.1% and 0.94 respectively). Although R_{free} (Brünger, 1992) was not monitored in the original analysis, the cross-validation technique was used for the purposes of comparing the two models. For a rigid body refinement with the model of the dodecamer split into 12 rigid base pairs, $R = 32.7\%$, $R_{free} = 43.4\%$ for the correct structure and $R = 37.3\%$, $R_{free} = 58.2\%$ for the "isomorphous" model for resolution range 15−4 Å. The incorrect structure had twice as many bond distances and angles deviating by more than 2σ from their target values, and the incorrect model does not converge below estimated σ values for the distances and angles. The last entry in Table 3 compares the two models with conditions the same as for the final structure presented here. In this case every indicator is significantly better, and there is no doubt that the initial isomorphous model is incorrect. Similar results have been reported for another moderate resolution structure of a B-DNA duplex, pointing out the importance of a careful analysis of the difference density maps, bad intermolecular contacts in the structure, and unusual backbone conformations (Heinemann & Alings, 1991). The analysis presented here emphasizes once again the impor-

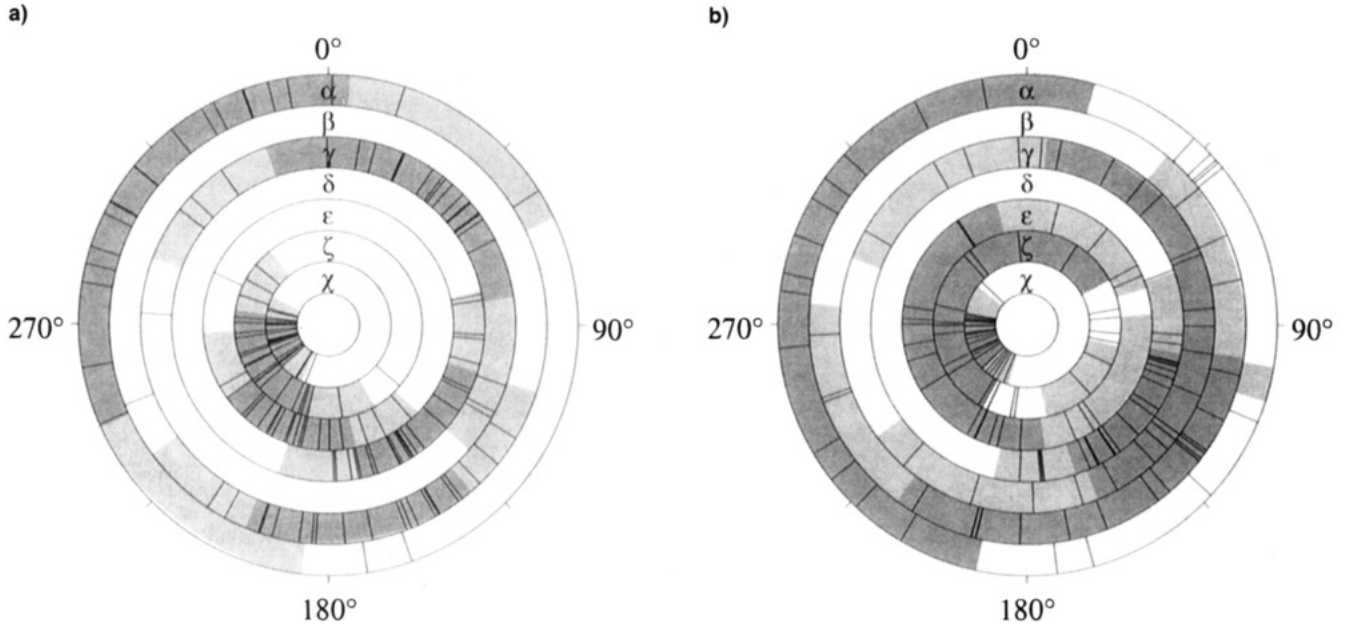

FIGURE 8: Backbone dihedral angles of (a) the correct model and (b) the incorrect isomorphous model.

tance of including the full resolution range for the later stages of refinement, the dangers of incorporating water molecules in early stages, and the importance of using realistic σ values.

In order to discover geometric criteria for detecting incorrect structures, the torsion angles for both models were compared as shown in Figure 8. In both cases, α, the torsion around P-O5′, and γ, the torsion around C5′-C4′, are largely scattered because these two torsion angles are correlated; small changes in one can be compensated by changes in the other angle. On the other hand, in the correct model, the distributions of the values of β (O5′-C5′), ϵ (C3′-O3′) and ζ (P-O3′) are much smaller and closer to the values seen in a survey of all oligonucleotides (B. Schneider, H. M. Berman, and S. Neidle, personal communication). This effect can be of a major importance for evaluation of the model.

The geometry observed in this structure is consistent with the conclusion based on studies of nucleosides in nonaqueous solvent (Williams & Shaw, 1987). As in the case of {d-[CGAGAATTC(O^6Me)GCG]}$_2$(Ginell et al., 1994), the crystal conformation appears to be different from the one observed with the title sequence in solution by NMR (Goswami et al., 1993). If the methyl groups were to be in the distal conformation seen in the crystal structure of a single nucleoside (Yamagata et al., 1988), the base pairs would be forced to open into the major groove such that there would be no hydrogen bond between the N3 of the thymine and the N1 of the (O^6Me)G. At least one conserved intermolecular contact (between O3′ G12 and N2 G22) would be disrupted. In the only other example of a B-DNA containing this base pair, {d[CGC(O^6Me)GAATTTGCG]}$_2$ (Leonard et al., 1990), the effect of changing the conformation of the methyl group would be to remove another conserved intermolecular contact (between N3 G16 and O3′ G24). Since there are no energetic reasons that force the methyl groups of the guanine residues to be in a proximal conformation, it is likely that the difference between the molecular structure reported here and the solution NMR structure reflects the differences between the solution and crystal environments.

ACKNOWLEDGMENT

We thank Professor W. K. Olson and Dr. A. Gorin for their valuable suggestions in the structural analysis of the helix geometry.

REFERENCES

Abbot, P. J., & Saffhill, R. (1979) *Biochim. Biophys. Acta 562*, 51−61.

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

Babcock, M. S., & Olson, W. K. (1992) in *Computation of Biomolecular Structures: Achievements, Problems, and Perspectives* (Soumpasis, D. M., & Jovin, T. M., Eds.) pp 65−85, Springer Verlag, Heidelberg.

Balmain, A., & Prangell, I. B. (1983) *Nature 303*, 72−74.

Brown, T., & Jiricny, J. (1988) *Cell 54*, 705−711.

Brünger, A. T. (1992a) *Nature 355*, 6359.

Brünger, A. T. (1992b) *X-PLOR Manual, version 3. 1*, Yale University Press, New Haven, CT.

Brünger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458−460.

Dickerson, R. E., Kopka, M. L., & Pjura, P. (1984) in *Biological Macromolecules and Assemblies: Nucleic Acids and Interactive Proteins* (Jurnak, F. A., & McPherson, A., Eds.) pp 38−126, John Wiley, New York.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., Wang, A. H.-J., & Zhurkin, V. B. (1989) *EMBO J. 8*, 1−4.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2179−2183.

Eva, A., & Aaronson, S. A. (1983) *Science 220*, 955−956.

Fitzgerald, P. M. D. (1988) *J. Appl. Crystallogr. 21*, 273−278.

Fitzgerald, P. M. D. (1991) *MERLOT—An Integrated Package of Computer Programs for the Determination of Crystal Structures by Molecular Replacement*, Version 2. 4, Merck Sharp & Dohme Research Laboratories, Rahway, NJ.

Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686−14707.

Gaffney, B. L., Goswami, B., & Jones, R. A. (1993) *J. Am. Chem. Soc. 115*, 12607−12608.

Gao, Y.-G., Sriram, M., Denny, W. A., & Wang, A. H.-J. (1993) *Biochemistry 32*, 9639−9648.

Ginell, S. L., Kuzmich, S., Jones, R. A., & Berman, H. M. (1990) *Biochemistry 29*, 10461−10465.
Ginell, S. L., Vojtechovsky, J., Gaffney, B., Jones, R., & Berman, H. M. (1994) *Biochemistry 33*, 3487−3493.
Goswami, B., Gaffney, B. L., & Jones, R. A. (1993) *J. Am. Chem. Soc. 115*, 3832−3833.
Goth, R., & Rajevsky, M. F. (1974) *Z. Krebsforsch. 82*, 37−64.
Griffin, S., Branch, P., Xu, Y., & Karran, P. (1994) *Biochemistry 33*, 4787−4793.
Haget, Y., Housty, J. R., Maiga, A., Bonpunt, L., Chanh, N.-B., Cuevas, M., & Estop, E. (1984) *J. Chim. Phys. 84*, 197−206.
Heinemann, U., & Alings, C. (1991) *EMBO J. 10*, 35−43.
Holmes, J., Clark S., & Modrich, P. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5837−5841.
Kamiya, H., Sakaguchi, T., Fujimoro, M., Miura, H., Ishikawa, H., Shimizu, M., Inoue, H., Matsukage, A., Masutani, C., Hanaoka, F., & Ohtsuka, E. (1992) *Chem. Pharm. Bull. 40*, 2792−2795.
Kitaigorodsky, A. I. (1984) in *Mixed Crystals*, pp 75−84, Springer-Verlag, Berlin.
Leonard, G. A., Thomson, J., Watson, W. P., & Brown, T. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9573−9576.
Parkinson, G., Vojtechovsky, J., Clowney, L., & Berman, H. M. (1996) *Acta Crystallogr. D* (in press).
Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., & Jones, R. A. (1986a) *J. Mol. Biol. 188*, 677−692.
Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., & Jones, R. A. (1986b) *Biochemistry 25*, 1036−1042.
Rabinovich, D., & Shakked, Z. (1984) *Acta Crystallogr. A40*, 195−200.
Schneider, B., Cohen, D. M., Schleifer, L., Srinivasan, A. R., Olson, W. K., & Berman, H. M. (1993) *Biophys. J. 65*, 2291−2303.
Spink, N., Nunn, C. M., Vojtechovsky, J., Berman, H. M., & Neidle, S. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 10767−10771.
Sriram, M., van der Marel, G. A., Roelen, L. P. F., van Boom, J. H., & Wang, A. H.-J. (1992) *EMBO J. 11*, 225−232.
Sussman, J. L., Holbrook, S. R., Church, G. M., & Kim, S.-H. (1977) *Acta Crystallogr. A33*, 800−804.
Thomas, D. C., Roberts, J. D., & Kunkel, T. A. (1991) *J. Biol. Chem. 266*, 3744−3751.
Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119−145.
Wiebauer, K., & Jiricny, J. (1989) *Nature 339*, 234−236.
Williams, L. D., & Shaw, B. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1779−1783.
Yamagata, Y., Kohda, K., & Tomita, K. (1988) *Nucleic Acids Res. 16*, 9307−9311.

BI951788E