The EMBO Journal vol.11 no.1 pp.225-232, 1992

Conformation of B-DNA containing O^6-ethyl-G-C base pairs stabilized by minor groove binding drugs: molecular structure of d(CGC[e^6G]AATTCGCG complexed with Hoechst 33258 or Hoechst 33342

M.Sriram, Gijs A.van der Marel[1], Harlof L.P.F.Roelen[1], Jacques H.van Boom[1] and Andrew H.-J.Wang

Department of Physiology and Biophysics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA and [1]Gorlaeus Laboratories, Leiden State University, 2300RA Leiden, The Netherlands

Communicated by W.G.J.Hol

O^6-ethyl-G (e^6G) is an important DNA lesion, caused by the exposure of cells to alkylating agents such as *N*-ethyl-*N*-nitrosourea. A strong correlation exists between persistence of e^6G lesion and subsequent carcinogenic conversion. We have determined the three-dimensional structure of a DNA molecule incorporating the e^6G lesion by X-ray crystallography. The DNA dodecamer d(CGC[e^6G]AATTCGCG), complexed to minor groove binding drugs Hoechst 33258 or Hoechst 33342, has been crystallized in the space group $P2_12_12_1$, isomorphous to other related dodecamer DNA crystals. In addition, the native dodecamer d(CGCGAATTCGCG) was crystallized with Hoechst 33342. All three new structures were solved by the molecular replacement method and refined by the constrained least squares procedure to R-factors of ~16% at ~2.0 Å resolution. In the structure of three Hoechst drug-dodecamer complexes in addition to the one published earlier [Teng *et al.* (1988) *Nucleic Acids Res.*, 16, 2671-2690], the Hoechst molecule lies squarely at the central AATT site with the ends approaching the G4-C21 and the G16-C9 base pairs, consistent with other spectroscopic data, but not with another crystal structure reported [Pjura *et al.* (1987) *J. Mol. Biol.*, 197, 257-271]. The two independent e^6G-C base pairs in the DNA duplex adopt different base pairing schemes. The e^6G4-C21 base pair has a configuration similar to a normal Watson-Crick base pair, except with bifurcated hydrogen bonds between e^6G4 and C21, and the ethyl group is in the proximal orientation. In contrast, the e^6G16-C9 base pair adopts a wobble configuration and the ethyl group is in the distal orientation. There may be a dynamic equilibrium between these two configurations for the e^6G-C base pair, which presents an ambiguous signal to the cellular replication and repair mechanisms. In contrast, thymine can pair with e^6G in only one way, albeit imperfect, mimicking a Watson-Crick base pair. This may be a plausible explanation of why thymine is found preferentially incorporated across the e^6G during replication.

Key words: DNA conformation/DNA lesion/drug-DNA interaction/X-ray diffraction

Introduction

Many chemical carcinogens act by forming covalent adducts with DNA (Singer and Gruberger, 1983). Among them, alkylating agents such as *N*-methyl-*N*-nitrosourea (MNU) and *N*-ethyl-*N*-nitrosourea (EtNU) constitute an important class of DNA modifiers. The chemical and biological consequence of the reaction between these alkylating agents with DNA has been the subject of intensive studies (Singer *et al.*, 1978; Thomale *et al.*, 1990). It has been shown that many nucleophilic sites on DNA bases (e.g. N^7 and N^3 of purine) are readily attacked by the methyl or ethyl cation which can be generated via a non-enzymatic heterolytic reaction by MNU and EtNU (Singer *et al.*, 1978). While the alkylation on those sites is potentially deleterious to the function of DNA, cells have developed efficient repair systems to remove those lesions. However, the alkylation at the O^6 position of guanine results in the formation of O^6-methyl-guanine (m^6G) and O^6-ethyl-guanine (e^6G), which is repaired by a different mechanism, using the suicide enzyme O^6-alkylguanine alkyltransferase (AGT; EC 2.1.1.63) to remove the alkyl group and to regenerate an intact G (Lindahl *et al.*, 1988; Thomale *et al.*, 1990). In some cells, this repair system is deficient and the m^6G/e^6G lesion remains persistently in DNA which results in the misincorporation of T opposite to the lesion site (Loechler *et al.*, 1984). The relative repair capacity of cells, for the repair of O^6-alkyl-guanine, is a critical determinant for the risk of malignant conversion by *N*-nitroso carcinogens (Jurgen *et al.*, 1990; Leonard *et al.*, 1990).

From a structural point of view, when the O^6 of guanine is alkylated, the modified base changes its tautomeric form such that its N^1 no longer has a proton and hence cannot be a hydrogen bond donor. Furthermore, the O^6-alkyl group may adopt two possible orientations, proximal and distal, as shown in Figure 1 (top). The distal configuration is energetically more favorable as predicted by theoretical calculations (Pedersen *et al.*, 1990). In fact, the crystal structure of the free nucleoside O^6-methyl-guanosine showed unequivocally that it adopted the distal configuration (Parthasarathy and Fridey, 1986). However, when a m^6G/e^6G is incorporated into the DNA double helix and participates in the base pairing interaction, the methyl/ethyl group in the distal orientation is expected to hinder the hydrogen bond formation with the opposing base, causing unusual base pairing schemes which may destabilize the helix.

This critical issue related to the type of base pairing schemes that m^6G/e^6G may form with other bases in DNA remains unresolved. Figure 2A-D shows some of the possibilities for e^6G-C and e^6G-T base pairs. In the Z-DNA crystal structure of d(CGC[m^6G]CG), the type in Figure 2D, which requires a protonated C, was found (Ginell *et al.*, 1990). In the B-DNA structure of d(CGC[m^6G]AATTTGCG), the type in Figure 2C was

found (Leonard *et al.*, 1990). However, there is still no definitive conclusion about which m^6G-C base pairing scheme exists in B-DNA under physiological conditions. On the basis of NMR studies of the DNA dodecamer d(CGCGAATTC[m^6G]CG) in solution, Patel *et al.* (1986) tentatively proposed that the wobble type of Figure 2B is adopted in the helix. Clearly, more definitive structural studies are needed to answer this question.

Thus far, no double helical structure of any kind (A, B or Z) incorporating e^6G lesions has been determined by X-ray crystallography. In this work, we present the crystal structure of the DNA dodecamer d(CGC[e^6G]AATTCGCG) in the presence of the minor groove binding drugs Hoechst 33258 (H258) or Hoechst 33342 (H342) [Figure 1 (bottom)]. Crystals of d(CGC[e^6G]AATTCGCG) duplex could be obtained only in the presence of minor groove binding drugs and these drugs seem to play a role in stabilizing the DNA duplex containing the e^6G lesion (see below). The two independent e^6G-C base pairs in the B-DNA double helix adopt different base pairing schemes in which the O^6-ethyl group plays an important role in influencing the conformation of the base pair. These O^6-ethylated drug–DNA complexes are compared with the corresponding native complexes.

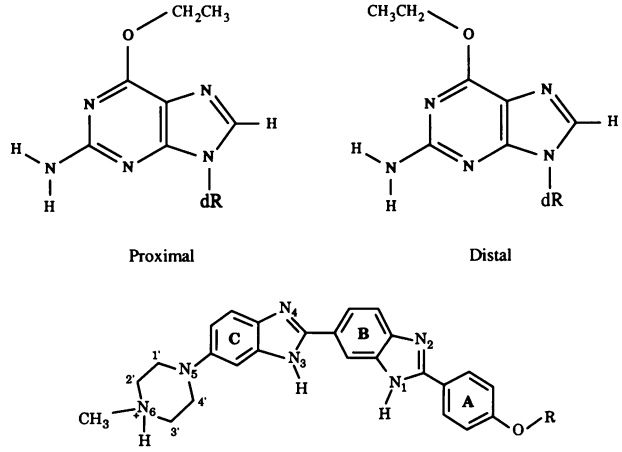

Fig. 1. Top: the molecular formula of O^6-ethyl-G (e^6G), showing the two possible orientations (proximal and distal to N^7) of the ethyl group. Bottom: the molecular formula of Hoechst 33258 and Hoechst 33342.

Results and discussion

Structure of complexes

The difference Fourier electron density maps of the d(CGC[e^6G]AATTCGCG)–Hoechst 33342 complex (e^6G-DODE/H342) structure was calculated by removing the H342 from the phase contribution (Sriram *et al.*, 1991). The drug molecule was seen to fit nicely in the residual caterpillar-shaped electron density envelope. The density was sufficiently well resolved to allow us to define the position and polarity of the drug molecule in the duplex. The Hoechst 33342 molecule lies in the narrow minor groove of the B-DNA duplex in the AATT region (Figure 3). The drug actually covers 6 bp with the *N*-methyl-piperazine ring approaching the G4-C21 on one end and the ethyl tail hanging near the G16-C9. This polarity of Hoechst drug binding mode is the same in all four complexes {e^6G-DODE/H342, d(CGC[e^6G]AATTCGCG)–Hoechst 33258 (e^6G-DODE/H258), d(CGCGAATTCGCG)–Hoechst 33342 (DODE/H342) and d(CGCGAATTCGCG)–Hoechst 33258 (DODE/H258) (Teng *et al.*, 1988)}.

The overall structure of the dodecamer DNAs in these four

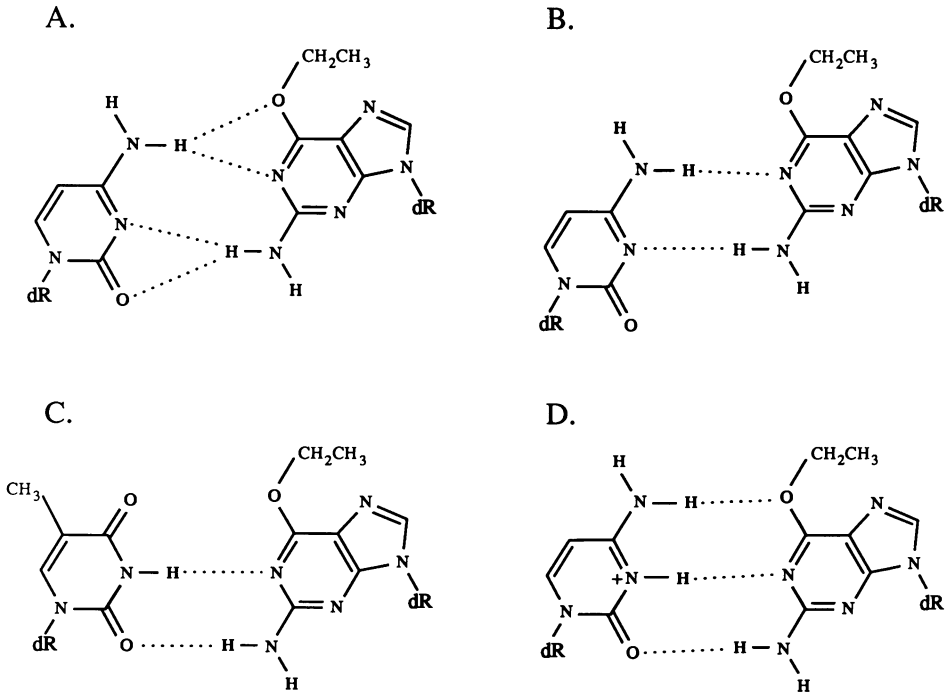

Fig. 2. Possible hydrogen bonding configurations of e^6G-C and e^6G-T base pairs. (**A**) Bifurcated e^6G-C base pair with two sets of bifurcated hydrogen bonds. (**B**) Wobble e^6G-C base pair with two hydrogen bonds. (**C**) e^6G-T base pair with a shape similar to a Watson–Crick base pair. (**D**) Watson–Crick e^6G-C^+ base pair. The cytosine is protonated.

complexes is similar to other related dodecamers (Drew and Dickerson, 1981; Coll *et al.*, 1987, 1989; Teng *et al.*, 1988; Carrondo *et al.*, 1989). It has a characteristic narrow minor groove at the AATT region. However, the changes in the DNA conformation in these four complexes relative to that of the canonical AATT dodecamer are numerous and distributed throughout the helix, presumably due to the insertion of the two e^6Gs, the binding of the Hoechst drug, or both. The root mean square deviations of the structures among the four DNA duplexes range from 0.651 Å to 0.854 Å (Table I). Instead of comparing the individual torsion angles between various DNA structures, we focused on the base pair buckle and propeller twist angles of these four complexes listed in Table II. It can be seen that nearly every base pair in the helix has either the buckle or the propeller twist angle greater than 10°, with the exception of C3-G22 and G12-C13 base pairs. The two G-C base pairs at both ends of the helix are involved in the interlocking lattice interactions using the G14-G24# and G12-G2# (# stands for a symmetry-related duplex) hydrogen bonding pairing in the minor groove. As noted before (Coll *et al.*, 1990), this type of G-G pairing is associated with a high dihedral angle between the two guanines. This may impose conformational distortion in the participating (terminal and penultimate) base pairs. Therefore, the terminal C1-G24 has high buckle (average −12°), whereas the penultimate G2-C23 has higher propeller twist angle ω (average −9° and −18° in d(CGCGGAATTCGCG)–Hoechst 33258 and d(CGC[e^6G]AATTCGCG)–Hoechst 33258 complexes (H258C) and d(CGCGAATTCGCG)–Hoechst 33342 and d(CGC[e^6G]AATTCGCG)–Hoechst 33342 complexes (H342C) respectively) and so does the penultimate C11-G14 ω with average −15° for all except e^6G-DODE/H342 (4°).

In most of the dodecamer structures, the base pairs in the central AT region have been found to have high propeller twist angles which result in the bifurcated hydrogen bonds from the N^6 amino group of an adenine simultaneously to the O^4 atoms of two thymines in the opposite strand (one from the Watson–Crick mate and the other from its adjacent 5′ T) (Coll *et al.*, 1987; Nelson *et al.*, 1987). This has been suggested as a possible reason for the unusual property (e.g. bent DNA) associated with the $A_n \cdot T_n$ sequence (for a recent review, see Crothers *et al.*, 1990). In the present structures, only the A6-T19 maintains a very high propeller twist (average −24°). The average distance in these four structures between N^6 of A5 and O^4 of T20 or O^4 of T19 is 2.93 Å and 3.05 Å respectively, satisfying the condition of the interbase bifurcated hydrogen bond. Interestingly, the propeller twist of T7-A18 decreases in the e^6G complexes relative to the regular complexes from average −15° (in DODE/H258 and DODE/H342) to −3° in the e^6G-DODE/H258 and 3° in e^6G-DODE/H342. This may be related to the compensatory increase of the buckle of adjacent A6-T19 from average −3° in the normal to average −15° in the e^6G complexes.

The ethyl groups of the two e^6Gs are in the major groove of the helix and they are both out of the plane from the guanine base (Figures 3 and 4). They point toward the opposite ends of the helix and make contact with the neighboring cytosines. The ethyl of e^6G4 and e^6G15 is close to the N^4/C^5 of C3 and C15 cytosines, respectively. These close contacts seem to push the e^6Gs away from the cytosines, possibly inducing a substantial conformational distortion in the e^6G-C base pairs as discussed below. We cannot say for certain whether the orientations adopted by the ethyl groups are their natural position, i.e. can the ethyl groups swing the other way so that they point toward the center of the helix?

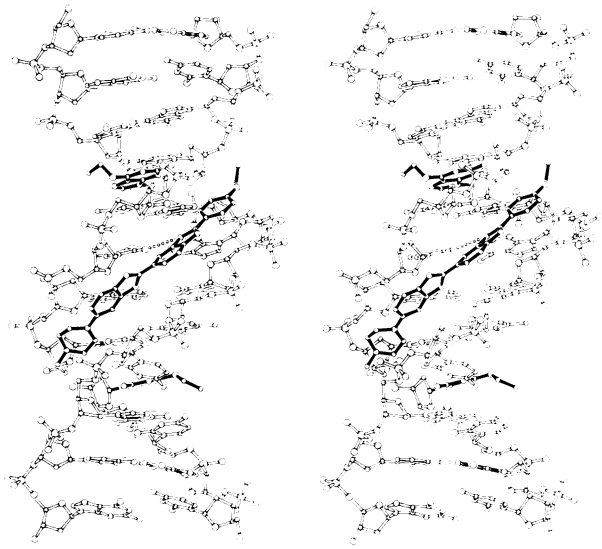

Fig. 3. The stereoscopic skeletal drawing of the structure of the d(CGC[e^6G]AATTCGCG)–Hoechst 33342 complex. The Hoechst 33342 molecule binds in the minor groove of the dodecamer B-DNA duplex. The drug molecule has a large dihedral angle between the successive rings such that it can adapt to the curved contour surface of the B-DNA helix. The two e^6Gs are drawn with filled bonds and their ethyl groups are located in the major groove.

e^6G-C base pairs

Figure 4 displays the ($F_o - F_c$) difference Fourier electron density (ED) of the ethyl groups of e^6Gs (they were not included in the phase contribution) and shows the structure of the two independent e^6G-C base pairs in the dodecamer duplex of the e^6G-DODE/H342 complex. The e^6G-DODE/H258 complex has similar results. It is interesting to note that they adopt different base pairing schemes. In order to ascertain that we have a reliable interpretation of the configuration associated with these modified base pairs, we examined the ED map very carefully, especially in and around the base pairs. It can be seen that there are very clear EDs for these ethyl groups.

The ED for the e^6G4 base indicates that the ethyl group is in the proximal orientation and is out of the best plane of guanine by 1.12 Å. The distance between the N^7 atom and the C_β atom (C_α is the methylene carbon and C_β is the

Table I. Root mean square deviation (Å) of the least-squares fit of drug (above diagonal) and DNA molecules (below diagonal) in four drug–DNA complexes

	DH258	D*H258	DH342	D*H342
DH258		0.482	0.429	0.357
D*H258	0.836		0.584	0.482
DH342	0.651	0.793		0.350
D*H342	0.854	0.768	0.682	

DH258, CGCGAATTCGCG–Hoechst 33258 complex;
D*H258, CGC[e^6G]AATTCGCG–Hoechst 33258 complex;
DH342, CGCGAATTCGCG–Hoechst 33342 complex;
D*H342, CGC[e^6G]GAATTCGCG–Hoechst 33342 complex.

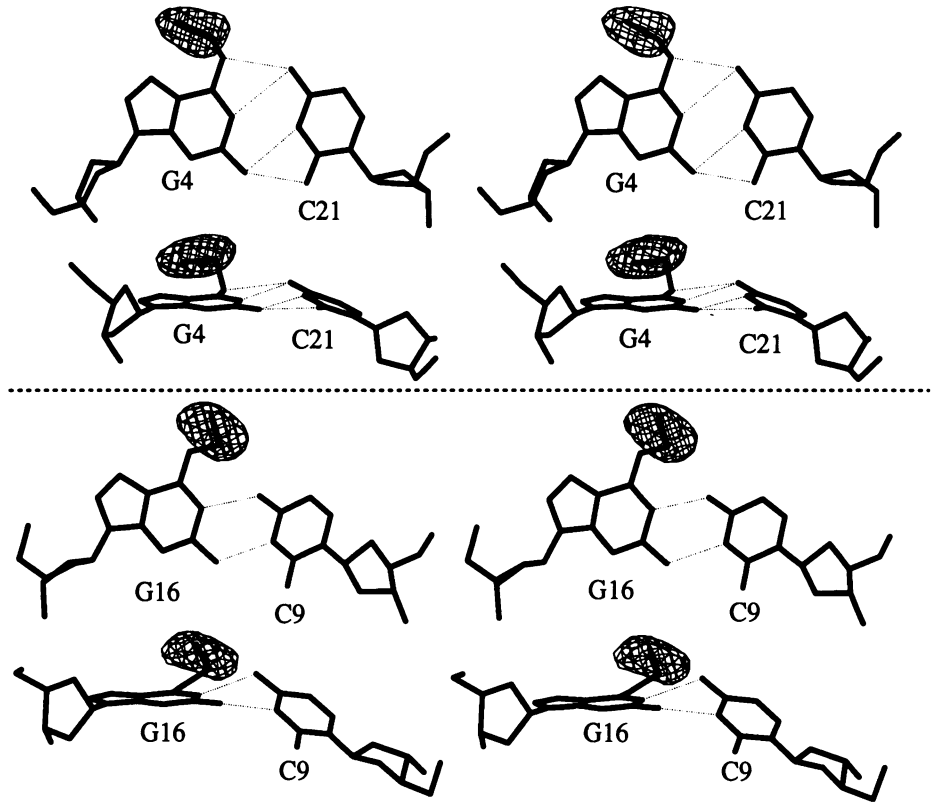

Fig. 4. Stereoscopic diagram of the detailed geometry of the two e^6G-C base pairs. The difference Fourier (F_o-F_c) electron density map with the O^6-ethyl groups removed from the phase contribution is displayed. Top: the e^6G4-C21 base pair adopts the bifurcated configuration with the ethyl group in the proximal orientation. Bottom: e^6G16-C9 base pair adopts the wobble configuration. The difference Fourier electron density is close to the N^4 amino group of C9. It is possible to fit the ethyl group in the density envelope with two conformations. In both base pairs, the ethyl group is out of the plane of guanine base.

methyl carbon of the ethyl group) is 2.8 Å. The proximal orientation of the ethyl group effectively blocks any access of solvent, metal or protein to the N^7 position of e^6G. The torsion angle about the O^6-C_α bond is in the anticlinical$^+$ conformation (91°). The e^6G4-C21 base pair has a configuration similar to a normal Watson–Crick base pair, but a close inspection of it suggests that it may be close to the type shown in Figure 2A. The N^4 of C21 is 2.97 Å from the O^6 and 3.16 Å from the N^1 of G4, and the N^2 of G4 is 2.63 Å from the O^2 and 2.90 Å from the N^3 of C21. Therefore, both amino groups (N^4 of C21 and N^2 of G4) appear to participate in bifurcated hydrogen bonding interactions. Notice that the base pair is quite distorted with a buckle of −22° (−18°) and propeller twist of 6° (2°) for the e^6G-DODE/H342 (e^6G-DODE/H258) complexes.

In contrast, the e^6G16-C9 base pair adopts a wobble configuration of the type shown in Figure 2B. The N^4 of C9 is 2.60 Å from the N^1 of G16 and the N^2 of G16 is 2.61 Å from the N^3 of the C9. This is due to the fact that the ethyl group is in the distal orientation, as clearly shown in the difference Fourier ED for the ethyl group. We noted that its ED envelope, which is a little more diffuse than that of the e^6G4-C21 base pair, may accommodate the ethyl group in two different orientations. In either orientation, the ethyl group also is out of the best plane of guanine (by 1.02 Å) as in the e^6G4-C21 base pair. One orientation is shown in Figure 4. Here the C^6-O^6-C_α-C_β torsion angle is in the anticlinical$^+$ conformation (134°) and the N^1-C^6-O^6-C_α torsion angle is in the gauche$^+$ conformation (30°). The distance between the C_α atom and the N^4 of C9 is 2.7 Å. The other ethyl orientation (not shown) is by rotating the C^6-O^6 bond so that the C^6-O^6-C_α-C_β torsion angle is now in the anticlinical$^-$ conformation (−106°) and the N^1-C^6-O^6-C_α torsion angle is in the gauche$^+$ conformation (84°). Here the distance between the C_β atom and the N^4 of C9 is 3.4 Å. In either orientation, the ethyl group approaches the N^4 of C9, forcing the base pair to adopt the wobble configuration and causing a significant conformational distortion in the base pair with a buckle of 24° and propeller twist of −24° for the e^6G-DODE/H342 complex. Interestingly, the conformation (buckle) of the same base pair in the e^6G-DODE/H258 complex is normal, as compared with the unmodified DODE/H258 complex. However, the propeller twist is −11° higher in e^6G-DODE/H258 complex (Table II). In the e^6G-DODE/H258 complex conformational changes occur mainly in the AT region. For example, the A6-T19 and A5-T20 base pairs in the e^6G-DODE/H258 complex have large buckles of −20° and −14°, respectively.

Our results here for the first time show unambiguously that e^6G-C base pair may adopt the wobble configuration (Figure 2B) or the bifurcated pairing configuration (Figure 2A) near physiological neutral pH conditions. Both types have been suggested as possible base pairing schemes by theoretical calculations (Pedersen *et al.*, 1990). A common feature associated with the e^6G-C base pair at neutral pH is the distorted conformation (high buckle and propeller twist angles) in and around the lesion site. This may be more easily

Table II. Selected deformation parameters[a] in the DNA of the four drug–DNA complexes

Base pair	Buckle (degrees)				Propeller twist (degrees)			
	DH258	D*H258	DH342	D*H342	DH258	D*H258	DH342	D*H342
C1-G24	−15	−7	−11	−15	−5	−10	−14	4
G2-C23	0	−17	−4	−9	−8	−9	−16	−19
C3-G22	2	−4	5	−3	−8	12	−6	−6
G4-C21	−8	−18	−19	−22	−2	6	−1	2
A5-T20	−4	−14	−8	−1	−9	−12	−10	−4
A6-T19	−1	−20	−4	−11	−27	−17	−24	−26
T7-A18	−9	0	−4	6	−17	−3	−13	3
T8-A17	11	−2	2	4	−7	−6	−8	−5
C9-G16	11	12	20	24	−10	−21	−13	−24
G10-C15	−7	−5	−14	−12	−8	−10	0	−9
C11-G14	0	−8	2	−5	−19	−15	−12	4
G12-C13	9	6	0	−9	−6	−4	0	2

DH258, CGCGAATTCGCG–Hoechst 33258 complex; D*H258, CGC[e^{6}G]AATTCGCG–Hoechst 33258 complex; DH342, CGCGAATTCGCG–Hoechst 33342 complex; D*H342, CGC[e^{6}G]GAATTCGCG–Hoechst 33342 complex.
[a]Dickerson *et al.*, 1989.

recognized by the appropriate repair enzymes. However, it is clear that the bifurcated pairing configuration (Figure 2A) is not very different from the normal Watson–Crick G-C configuration. This could explain why C can still be incorporated in the daughter strand opposite to the e^{6}G lesion during replication.

There have been studies which proposed that e^{6}G does not pair with a neutral C, instead it only pairs with a protonated C^+ (Williams and Shaw, 1987). This proposal is neither consistent with our observation here, nor with the results from the NMR study at neutral pH of d(CGCGAATTC[m^{6}G]CG) in which a wobble m^{6}G-C base pair was proposed (Patel *et al.*, 1986). A pH-dependent melting study has been carried out on d(CGC[e^{6}G]-AATTCGCG), the same sequence used in this work, which showed a slightly biphasic melting curve with the highest T_m of 298 K at pH 5.0 (Leonard *et al.*, 1990). This was interpreted as the result of the formation of the protonated Watson–Crick configuration (Figure 2D) in the duplex. Curiously, we have not been able to obtain a suitable crystal of d(CGC[e^{6}G]AATTCGCG) under low pH condition (<6.0) with or without drug. While the protonation of C^+ may stabilize a e^{6}G-C^+ base pair in acidic condition, it destabilizes the normal G-C base pair elsewhere in the molecule. Under the physiological condition, it is unlikely that DNA is protonated to any significant extent. As noted above, the bifurcated pairing configuration (Figure 2A) is similar to the normal Watson–Crick G-C configuration. As a result, we believe there is no need to invoke protonated C^+ to explain the biological consequence due to the e^{6}G lesion.

Our findings here in regard to the orientation of the ethyl group suggest that both proximal and distal (Figure 1) are probable in B-DNA helix. The results from other modified DNAs are consistent with these observations. As mentioned above, in the Z-DNA crystal structure of d(CGC[m^{6}G]CG) (Ginell *et al.*, 1990) and in the B-DNA structure of d(CGC[m^{6}G]AATTTGCG) (Leonard *et al.*, 1990), the methyl group of the e^{6}G adopts a proximal orientation, as does the methyl group of the N^6-methyl-A in d(CGCGA-[m^{6}A]TTCGCG) (Frederick *et al.*, 1988). But the methoxy group of the N^4-methoxy-C in d(CG[N^4-methoxy-C]GCG) adopts a distal orientation (Van Meervelt *et al.*, 1990).

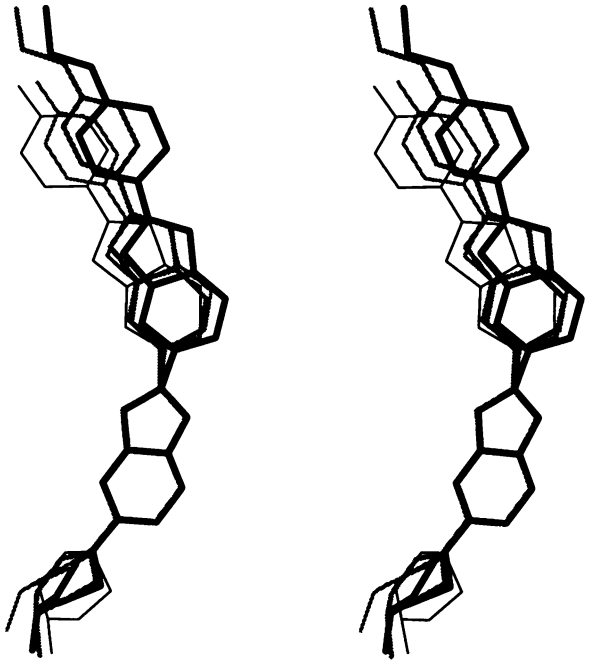

Fig. 5. A stereoscopic view of the superposition of Hoechst drug molecules from four different complexes by fitting their benzimidazole ring C together. All four Hoechst molecules have slightly different conformations, reflecting the drug molecule's ability to adapt to different local minor groove environments of the dodecamer helixes.

Clearly, the conformation of those exocyclic modifications depends on the local environment.

Drug–DNA interactions

We have determined four different complexes of DNA and Hoechst drugs which provide us with an additional wealth of information on how Hoechst drug complexes adjust their conformation to adapt to the contour surface of the narrow minor groove in B-DNA (Teng *et al.*, 1988; Carrondo *et al.*, 1989; Wang and Teng, 1990). Figure 5 shows the

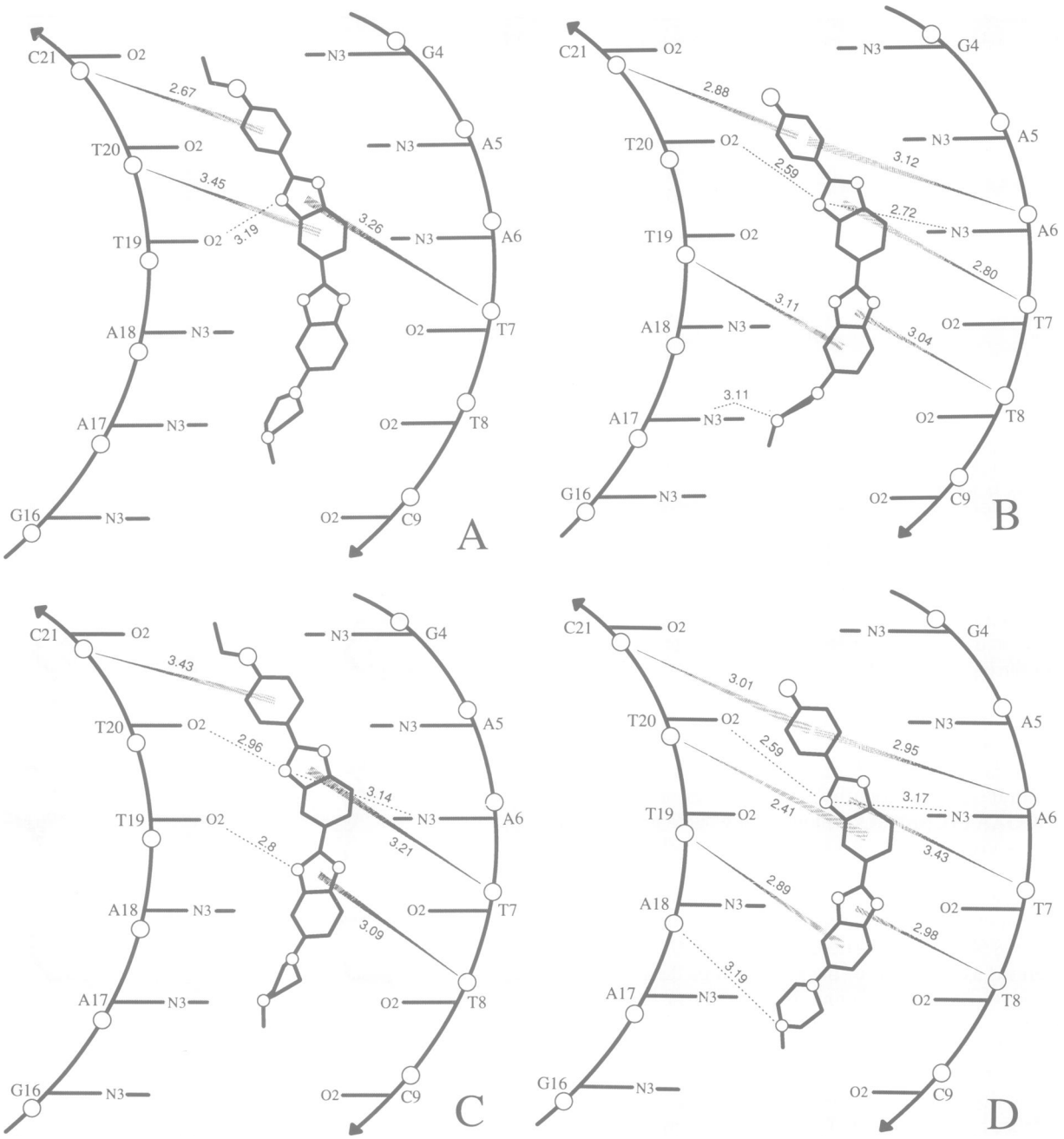

Fig. 6. A schematic diagram showing the interactions between the Hoechst drug molecule in the four different structures. (**A**) e^6G-DODE/H342, (**B**) e^6G-DODE/H258, (**C**) DODE/H342 and (**D**) DODE/H258. In all four structures, the Hoechst molecule is sandwiched in the minor groove at the AATT site between the two anti-parallel backbones of the DNA helix. However, they differ somewhat in details. Many van der Waals interactions (shown as elongated shaded triangles), such as the dipole−π interaction between the $O^{4'}$ (e.g. from sugars of T7, T20 and C21 in the e^6G-DODE/H342 complex) and the aromatic rings of the Hoechst drug, are used to stabilize the binding along with hydrogen bonds (shown as dotted lines). The $O^{4'}$ are shown as open circles on the DNA strand.

superposition of the four Hoechst drugs (two H258s and two H342s) by fitting the ring C between them. They differ from one another in the overall curvature and the dihedral angles between successive rings. In general, H258 has a higher curvature than H342. The structural analysis of the complexes suggests that this is due to the additional ethyl group in H342 which would have a close contact with DNA if the H342 maintains the same curvature as H258. The *N*-methyl-piperazine ring has a different orientation relative to ring C and in the complex its positively charged $N^{4'}$ nitrogen points toward the sugar $O^{4'}$ (and in DODE/H258 and DODE/H342 complexes there is a hydrogen bond between them). The variation in the drug conformation is interesting since it may have some relevance in the fluorescence quantum yield of the Hoechst drug (Loontiens *et al.*, 1991) which may be related to the dihedral angle between the aromatic rings in the drug molecule.

Figure 6 schematically summarizes the detailed interactions between the crescent-shaped Hoechst drug with DNA by comparing four different complexes. The H342 or H258 molecule lies squarely at the central AATT site with the ends approaching the G4-C21 and the G16-C9 base pairs. Other spectroscopic data, including the fluorescent (Loontiens *et al.*, 1991) and NMR (Parkinson *et al.*, 1990)

Table III. Relevant crystal data and final refinement parameters

Complex	Space group	Unit cell parameters Å	Resolution Å	R-factor %	No. of reflections	RMSD[a] Å
d(CGCGAATTCGCG) + Hoechst 33342	$P2_12_12_1$	a = 25.69 b = 41.07 c = 66.42	~2.25	16.8	1735(2σ)	0.020
d(CGC[e^6G]AATTCGCG) + Hoechst 33342	$P2_12_12_1$	a = 25.71 b = 41.32 c = 67.08	~2.5	15.7	1194(2σ)	0.019
d(CGCGAATTCGCG) + Hoechst 33258	$P2_12_12_1$	a = 25.23 b = 40.58 c = 66.08	~2.0	17.2	2000(2σ)	0.020
d(CGC[e^6G]AATTCGCG) + Hoechst 33258	$P2_12_12_1$	a = 25.64 b = 41.31 c = 66.99	~2.5	14.5	1225(3σ)	0.020

[a]RMSD is the root mean square deviation of bond distances.

studies, supported our previous observations (Teng *et al.*, 1988) and our present results further provided conclusive data that the binding site of Hoechst drug for the sequence of 5′-GAATTC- is 5′-AATT, not 5′-ATTC. This is in contrast to that reported by Pjura *et al.* (1987). Interestingly, H258 binds to the 5′-ATAC site in the d(CGCGTATA-CGCG) dodecamer (Carrondo *et al.*, 1989), suggesting a sequence-dependent binding specificity.

The binding of the Hoechst molecule to DNA is stabilized by several types of forces. All four structures reinforce the observations made previously regarding the molecular basis of the mode of actions of those drugs. As already discussed in the work of the DODE/H258 complex (Teng *et al.*, 1988; Wang and Teng, 1990), this may be summarized here as: (i) electrostatic attraction between the positively charged drug and the negatively charged DNA, (ii) van der Waals interaction between the DNA sugar atoms along the two walls of the minor groove, (iii) hydrogen bonds between the NH of the benzimidazole or the piperazine and the T-O^2, A-N^3 or the $O^{4'}$ of the sugar of DNA. The specificity toward the AT sequence is aided by the natural tendency of the AT segment to have a narrow minor groove which provides a favorable surrounding to have the above described interaction. Last, but not least, is that the additional N^2 amino group in guanine presents a severe hindrance toward the drug and pushes the drug away from the floor of the minor groove and this would diminish the binding interactions substantially. Therefore Hoechst drugs bind preferentially to AT sequence over GC sequence. As the drug actually covers 6 bp, it requires at least four AT core sequences for tight binding. Recently, we pointed out that in drug–DNA complexes, both the drug and DNA molecules change their respective conformation to adapt to each other (Wang *et al.*, 1990). The present work is fully consistent with that concept.

Conclusion

The structural analyses of the complexes of the DNA containing e^6G lesions with the minor groove binding drugs (Hoechst 33258 and 33342) provided important information on how carcinogen-modified e^6G pairs with cytosine. Our data suggest that the base pairing scheme adopted can be the wobble or the bifurcated hydrogen bond pairing, depending on the local environment. The latter pairing configuration is similar to a normal G-C base pair. This structural similarity may allow the e^6G in DNA to escape the repair system. During replication, either C or T may be inserted in the daughter DNA strand across the e^6G site. However, there may be a dynamic equilibrium between the two configurations of the e^6G-C base pair, which presents an ambiguous signal to the polymerase and is subsequently edited out. In contrast, thymine can pair with e^6G in only one way (with a configuration similar to a regular Watson–Crick G-C base pair), albeit imperfect. This may be a plausible explanation of why thymine is found preferentially incorporated across the e^6G lesion site during replication (Loechler *et al.*, 1984). Finally, the helix with e^6G lesions next to AATT sequence is stabilized by minor groove binding drugs. These results suggest that other lesions including mismatch base pairs may be similarly stabilized.

More analyses like the present work, e.g. by inserting e^6G in different nucleotide sequences, would enable us to understand more fully the sequence-dependent structural perturbations caused by O^6-alkylated lesions in DNA. Our goal is to assemble sufficient new information regarding the structural consequences of various types of DNA lesions which may lead us to a better understanding of chemical carcinogenesis.

Materials and methods

The synthesis of the O^6-ethyl-deoxyguanosine followed the procedure of Roelen *et al.* (unpublished results). The nucleoside was then converted into the phosphoramidite precursor and incorporated into the oligonucleotides on a Pharmacia DNA synthesizer. The sequence d(CGC[e^6G]AATTCGCG) was selected as we believed we could coerce the molecule into the lattice of the native AATT crystal (Drew and Dickerson, 1981) as in many other minor groove binding drug–dodecamer complexes (Coll *et al.*, 1987, 1989; Teng *et al.*, 1988; Carrondo *et al.*, 1989). Crystallization experiments using the procedure described previously (Wang and Gao, 1990) were carried out. To our dismay, no crystal could be obtained after numerous attempts. At that point, we recalled that the 'gapped' DNA duplex d(CGCGAAAACGCG) + d(CGCGTT) + d(TTCGCG) could only be crystallized in the presence of a minor groove binding drug like netropsin or Hoechst 33258 (Aymami *et al.*, 1990). We employed the same strategy and were able to obtain useful crystals using Hoechst 33258, Hoechst 33342 and netropsin. The crystallization solution in general contained 0.8 mM dodecamer (single strand concentration), 31 mM cacodylate buffer at pH 6.0, 4 mM $MgCl_2$, 1 mM spermine, 0.8 mM drug and 2% 2-methyl-2,4-pentanediol (2-MPD) and it was equilibrated against 50% 2-MPD by the vapor diffusion technique at room temperature. Large crystals with somewhat irregular shape appeared after 4 weeks. We have also crystallized the complex of Hoechst 33342–d(CGCGAATTCGCG) and determined its structure for comparisons. The complexes of other minor groove binding drugs with a series of related dodecamers have been studied by a number of investigators and reviewed elsewhere (Coll *et al.*, 1987, 1989; Carrondo *et al.*, 1989; Teng *et al.*, 1988; Larsen *et al.*, 1989; Wang and Teng, 1990).

In this paper, we focused on three new structures, e^6G-DODE/H342,

e^6G-DODE/H258 and DODE/H342 and compared them with the DODE/H258 structure. Crystal of each complex was mounted in a thin-walled capillary and sealed with a droplet of the crystallization mother liquor for data collection. All of them are in the isomorphous orthorhombic space group $P2_12_12_1$ and have unit cell dimensions of $a \sim 26$ Å, $b \sim 42$ Å and $c \sim 64$ Å. The diffraction data were collected at room temperature on a Rigaku AFC-5R rotating-anode diffractometer, using a ω-scan mode at 20°C with CuK_α radiation (1.5406 Å with graphite monochromater) at a power of 50 kV and 180 mA, to 2.0 Å resolution. Lorentz polarization, absorption and decay corrections were applied to the data before using it in the refinement.

As evident from the unit cell dimensions, this crystal form is closely related to crystal lattices of other B-DNA dodecamer–drug complexes crystal lattices (Wang and Teng, 1990). A B-DNA dodecamer without drug or solvent molecules was used as the starting model for refinement. The model was placed in the same position as that in the DODE/H258 crystal and it was refined using the Konnert–Hendrickson constrained refinement procedure (Hendrickson and Konnert, 1979; Westhof *et al.*, 1985). The entire dodecamer duplex was in the asymmetric unit, therefore the two strands of the duplex are not identical in their conformation. In order not to bias the base pairing scheme of the two e^6G-C base pairs, no hydrogen bonding distance constraints were imposed on them. After many cycles of refinement, the R-factor was ~33% at 2.0 Å resolution. The Hoechst drugs were located from the (F_o-F_c) difference Fourier map using the program FRODO/TOM (Jones, 1978) and included in the refinement. Solvent molecules located from the $(2F_o-F_c)$ Fourier maps, excluding those in the minor groove or near the e^6G4 or e^6G16 residues, were gradually added in the subsequent refinement cycles. At this stage, the R-factor was ~18%. For the e^6G-containing complexes, the (F_o-F_c) difference Fourier map was then used to locate the position of the ethyl groups. DODE/H258 structure too was re-refined to a comparable bond constraint. The relevant crystal data along with the final refined parameters are listed in Table III. The final atomic coordinates of these four structures have been deposited in the Brookhaven Protein Databank.

Acknowledgements

This work was supported by NIH GM-41612 and CA-52054 (A.H.-J.W.). M.S. was supported in part by a grant from DCCA of the Illinois state to the Beckman Institute at UIUC.

References

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

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

Coll,M., Frederick,C.A., Wang,A.H.-J. and Rich,A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8385–8389.

Coll,M., Aymami,J., van der Marel,G.A., van Boom,J.H., Rich,A. and Wang,A.H.-J. (1989) *Biochemistry*, **28**, 310–320.

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

Crothers,D.M., Haran,T.E. and Nadeau,J.G. (1990) *J. Biol. Chem.*, **265**, 7093–7096.

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

Drew,H.R. and Dickerson,R.E. (1981) *J. Mol. Biol.*, **151**, 535–556.

Frederick,C.A., Quigley,G.J., van der Marel,G.A., van Boom,J.H., Wang,A.H.-J. and Rich,A. (1988) *J. Biol. Chem.*, **263**, 17872–17879.

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

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

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

Jurgen,T., Huh,N.-H., Nehls,P., Eberle,G. and Rajewsky,M.F. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 9883–9887.

Larsen,T.A., Goodsell,D.S., Cascio,D., Grzeskowiak,K. and Dickerson, R.E. (1989) *J. Biomol. Struct. Dyn.*, **7**, 477–491.

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

Lindahl,T., Sedgwick,B., Sekiguchi,M. and Nakabeppu,Y. (1988) *Annu. Rev. Biochem.*, **57**, 133–158.

Loechler,E.L., Green,C.L. and Essigman,J.M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6271–6275.

Loontiens,F.G., McLaughlin,L.W., Diekmann,S. and Clegg,R.M. (1991) *Biochemistry*, **30**, 182–189.

Nelson,H.C.M., Finch,J.T., Luisi,B.F. and Klug,A. (1987) *Nature*, **330**, 221–226.

Parkinson,J.A., Barber,J., Douglas,K.T., Rosamond,J. and Sharples,D. (1990) *Biochemistry*, **29**, 10181–10190.

Parthasarathy,R. and Fridey,S.M. (1986) *Carcinogenesis*, **7**, 221–227.

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

Pedersen,L.G., Darden,T.A., Deerfield,D.W., II, Anderson,M.W. and Hoel,D.G. (1990) *Carcinogenesis*, **9**, 1553–1562.

Pjura,P.E., Grzeskopwiak,K. and Dickerson,R.E. (1987) *J. Mol. Biol.*, **197**, 257–291.

Rich,A., Nordheim,A. and Wang,A.H.-J. (1984) *Annu. Rev. Biochem.*, **53**, 791–846.

Singer,B. and Grunberger,D. (1983) *Molecular Biology of Carcinogens and Mutagens.* Plenum Press, New York.

Singer,B., Bodell,W.L., Cleaver,J.E., Thomas,G.H. and Rajewsky,M.F. (1978) *Nature*, **276**, 85–88.

Sriram,M., van der Marel,G.A., van Boom,J.H. and Wang,A.H.-J. (1991) *Seventh Conversation in Biomolecular Stereodynamics.* Adenine Press, New York, p. a211.

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

Thomale,J., Huh,N.-H., Nehls,P., Eberle,G. and Rajewsky,M.F. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 9883–9887.

Van Meervelt,L., Moore,H.M., Lin,P.K.T., Brown,D.M. and Kennard,O. (1990) *J. Mol. Biol.*, **216**, 773–781.

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

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

Wang,A.H.-J., Liaw,Y.-C., Robinson,H. and Giao,Y.-G. (1990) In Pullman,B. and Jortner,J. (eds), *23rd Jerusalem Symposium in Quantum Chemistry and Biochemistry.* Kluwer Academic Press, Dordrecht, Netherlands, pp. 1–21.

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

Williams,L.D. and Shaw,B.R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1779–1783.

Received on September 5, 1991; revised on October 8, 1991