Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 9573–9576, December 1990
Biochemistry

High-resolution structure of a mutagenic lesion in DNA

(x-ray diffraction/ultraviolet melting/guanine methylation)

G. A. LEONARD*, J. THOMSON*, W. P. WATSON†, AND T. BROWN*‡

*Department of Chemistry, Edinburgh University, Kings Buildings, EH9 3JJ Scotland, United Kingdom; and †Shell Research Ltd, Sittingbourne Research Centre, Kent ME9 8AG, United Kingdom

Communicated by John Cornforth, September 4, 1990 (received for review June 26, 1990)

ABSTRACT **The self-complementary dodecanucleotide d[CGC(m⁶G)AATTTGCG]₂ (where m⁶G is *O*⁶-methylguanine), which contains two m⁶G·T base pairs, has been analyzed by x-ray diffraction methods and the structure has been refined to a residual error of *R* = 0.185 at 2.0-Å resolution. The m⁶G·T mispair closely resembles a Watson–Crick base pair and there are very few structural differences between the m⁶G·T duplex and the native analogue. The similarity between the m⁶G·T base pair and a normal G·C base pair explains the failure of mismatch repair enzymes to recognize and remove this mutagenic lesion. A series of ultraviolet melting studies over a wide pH range on a related dodecamer indicate that the m⁶G·C mispair can exist in two conformations; one is a wobble pair and the other is a protonated Watson–Crick pair. The former, which predominates at physiological pH, will be removed by normal proofreading and repair enzymes, whereas the latter is likely to escape detection. Hence, the occasional occurrence of the protonated m⁶G·C base pair may explain why the presence of m⁶G in genomic DNA does not always give rise to a mutation.**

The initial stages of chemical carcinogenesis frequently involve the interaction of genotoxic agents with DNA to produce covalent modifications in the form of DNA adducts (1–3). An important example of this is the alkylation of the O⁶ position of guanine residues in DNA resulting from exposure to methylating agents such as *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (4) and methyl methanesulfonate and *N*-methyl-*N*-nitrosourea (5). The presence of *O*⁶-methylguanine (m⁶G) constitutes a mutagenic lesion that is known to specifically induce G·C to A·T transition mutations (6) and it has been established that protooncogenes can be converted to oncogenes by such a process (7). Hence, the formation of the m⁶G·T base pair during replication can give rise to a carcinogenic lesion (8, 9). In recent years, the biochemical processes involved in chemically induced carcinogenesis have been studied in considerable depth. However, to understand further the mechanisms of mutagenesis, it is necessary to analyze precisely the molecular details of the lesions produced when genotoxic agents interact with DNA. With this overall objective in mind, we have determined the structure of such a lesion, the m⁶G·T base pair in a B-DNA duplex.§

MATERIALS AND METHODS

All oligonucleotides were synthesized by the solid-phase method on an ABI model 380B DNA synthesizer using cyanoethyl phosphoramidite monomers. For those containing m⁶G, the following protocol was observed: The 5'-dimethoxytrityl-*N*²-isobutyryl-*O*⁶-methyldeoxyguanosine 3'-cyanoethyl phosphoramidite monomer was utilized to introduce *O*⁶-methyldeoxyguanosine and the fully assembled oligonucleotide was cleaved from the solid support and deprotected in a 5% solution of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in anhydrous methanol for 2 weeks at ambient temperature in an atmosphere of nitrogen (10). At no time was the oligonucleotide exposed to ammonia, as this can lead to the slow conversion of m⁶G to 2,6-diaminopurine (11). DBU was removed with Dowex-50 cation-exchange resin (Na^+ form) and purification was accomplished by reversed-phase HPLC (octyl), eluting with a linear gradient of acetonitrile in 0.1 M ammonium acetate (0%–20%, 30 min) to give a major and a minor product in approximate ratios of 9:1. These were easily separated due to the much greater retention time of the latter. A sample of each was digested to the free nucleosides with snake venom phosphodiesterase and alkaline phosphatase and the base composition was analyzed by reversed-phase HPLC (10). In all cases, this confirmed that the oligonucleotide containing *O*⁶-methyldeoxyguanosine was the major product. The minor product was found to contain *N*²-isobutyryl-*O*⁶-methylguanine, and it could be converted to the desired product by further treatment with DBU in methanol. Before carrying out pH-dependent ultraviolet melting studies, the stability of m⁶G at low pH was determined in the following way: The oligonucleotide d[CGC(m⁶G)AATTCGCT] was dissolved in 0.1M sodium phosphate (pH 5.0) and, after 2 weeks, reversed-phase HPLC analysis showed that there was no significant degradation. Mixed injections with the native sequence d(CGCGAATTCGCG), which elutes much earlier on reversed-phase HPLC, further confirmed that the m⁶G-containing oligonucleotide had not undergone demethylation.

The self-complementary dodecamer d(CGCm⁶GAATTTGCG)₂ crystallized isomorphously with the native dodecamer (12). Crystals were grown at 277 K and pH 6.3 from 25-μl drops containing 1.0 mM oligonucleotide, 33 mM $MgCl_2$, 16.7% (vol/vol) hexylene glycol, 1.25 mM spermine, and 5 mM sodium cacodylate. A single crystal of dimensions 1.3 × 0.3 × 0.3 mm was mounted in a sealed glass capillary and x-ray data were collected at 277 K on a Rigaku RU200 HB diffractometer equipped with a rotating anode x-ray generator and a Nicolet–Siemens X100 area detector system. Data were processed using the XDS software package (13). A total of 8543 reflections were merged to give 4481 unique reflections (*R* merge = 0.03), representing 88% of the total data to a resolution of 2.0 Å. A second data set was collected on a crystal of similar dimensions using a conventional Stoe–Siemens AED2 four circle diffractometer. The resultant 3735 reflections were corrected for absorption, decomposition, Lorentz and polarization effects and merged (*R* merge = 0.06) to yield 1953 unique reflections to 2.5 Å. The two data sets were scaled and merged (*R* merge = 0.03) to give a total of 4557 unique reflections, 91% of the total data to 2.0 Å. The starting model for the refinement was

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: m⁶G, *O*⁶-methylguanine.

‡To whom reprint requests should be addressed.

§The atomic coordinates and structure factors have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 1D27, R1D27SF).

obtained from the coordinates of the quasiisomorphous d(CGCGAATTCGCG)$_2$ (12) after idealizing the geometry to allow for the slightly different unit cell (a = 25.44 Å; b = 40.70 Å; c = 65.99 Å; $\alpha = \beta = \gamma = 90°$; space group $P2_12_12_1$). The initial stepwise rigid body refinement, using a modified version of SHELX (14) on all data from 10.0 Å to 2.5 Å, converged at R = 0.36. Subsequent Konnert–Hendrickson refinement (15) to 2.0 Å, using NUCLSQ (16), including all reflections with $F > 2\sigma(F)$ and omitting the base pairs G(4)–C(21) and C(9)–G(16) from the structure factor calculations converged at R = 0.44. Examination of difference Fourier and $2F_o - F_c$ maps on an Evans and Sutherland PS300 system using FRODO (17) clearly indicated that the mispairs had Watson–Crick geometry. The bases in positions 4–21 and 9–16 were then included as G·T base pairs to avoid biasing the position of the methyl groups of the m^{6}G bases and refinement was carried out using all data with $F > 3\sigma(F)$ in the range 7.0 Å to 2.0 Å. Positional and thermal parameters were included in the refinement that converged at R = 0.190 with the inclusion of 69 solvent molecules, all of which displayed good spherical density and hydrogen-bonding geometry. The resultant $F_o - F_c$ difference maps showed substantial density in the plane of the mispairs, in a position expected for the proximal methyl group of m^{6}G. Further refinement with the inclusion of the m^{6}G base at positions 4 and 16 of the duplex converged to give R = 0.185 for 3118 reflections in the range 7.0 Å to 2.0 Å.

RESULTS AND DISCUSSION

The x-ray structure analysis of the self-complementary dodecanucleotide d[CGC(m^{6}G)AATTTGCG]$_2$ shows that in overall shape the m^{6}G·T mispair is remarkably similar to a Watson–Crick base pair. The bases are directly opposite each other, the glycosyl linkages are related by a pseudodyad, and in the minor groove the base pair is indistinguishable from a G·C base pair. Overall, there are very few structural differences between the m^{6}G·T duplex and the well-studied native G·C duplex, d(CGCGAATTCGCG)$_2$ (12). The helical parameters, torsion angles, and hydration pattern are almost identical and the presence of the mutagenic base pair does not distort the sugar-phosphate backbone. Despite these considerable similarities, ultraviolet melting studies indicate that the m^{6}G·T duplex is much less stable than the native G·C duplex ($\Delta\Delta G°$ = 46.8 kJ/mol) (Table 1). Hence, the m^{6}G·T base pair has a powerful destabilizing effect in B-DNA.

The two m^{6}G·T base pairs in the dodecamer duplex are essentially identical and the $2F_o - F_c$ map of the m^{6}G(9)·T (16) pair is shown in Fig. 1*a*. There are three close contacts between the m^{6}G and T bases; one of 2.9 Å from purine-O^6 to pyrimidine-O^4, a second of 2.9 Å from purine-N^1 to pyrimidine-N^3, and a third of 2.8 Å between purine-N^2 and pyrimidine-O^2. At this resolution, we cannot observe hydrogen atoms and the most likely base pair is shown in Fig. 2, with each base as its most common tautomer. There is unlikely to be a hydrogen bond between purine-O^6 and pyrimidine-O^4, as neither has an attached hydrogen and neither functional group is sufficiently basic to be protonated at pH 6.3. In addition, there is no evidence in the pH-dependent ultraviolet melting profile of the duplex (Fig. 3*a*) for a protonated base pair. Hence, formation of the double strand from the fully hydrated single strands will lead to the loss of a hydrogen bond between each of these oxygen atoms and neighboring solvent molecules. As these are not replaced by interbase hydrogen bonds, the effect will be to destabilize the duplex. The interaction between purine-O^6 and pyrimidine-O^4 will be destabilizing for the additional reason that the two electronegative oxygen atoms are forced together by the two adjacent strong hydrogen bonds in the base pair (20), one from pyrimidine-N^3 to purine-N^1 and a second from purine-N^2 to pyrimidine-O^2. It is possible to postulate other forms of the m^{6}G·T base pair by invoking minor tautomers but there is no direct experimental evidence for their existence. The base pair in Fig. 2 has two hydrogen bonds that are adjacent to each other and the resultant cooperativity allows the formation of a stable base pair. The m^{6}G·T base pairing found in the present x-ray structure is identical to that postulated from molecular orbital and molecular mechanical calculations and from an NMR study of a mixture of ribonucleosides of m^{6}G and thymine in chloroform solution (21). However, it differs in important details from that proposed in an oligonucleotide NMR study, although interpretation of the spectra was limited by the lack of direct information on the relative orientation of the two bases (22).

The most striking differences between the m^{6}G·T mispair and a G·C base pair are in the major groove, due to the presence of the methyl group attached to the purine-O^6 atom, located proximal to the N^7 atom in the plane of the purine ring (Fig. 1*b*). This will give rise to steric repulsion between the guanine base and the attached methyl group. Although the distal conformation is preferred in the free nucleoside (23, 24), it is likely to be very unstable in the m^{6}G·T base pair (25) as it will prevent the formation of interbase hydrogen bonds. In the proximal orientation, the methyl group presents a steric barrier to any regulatory or repair enzyme that might otherwise interact with the guanine-O^6 or -N^7 atom. The appearance of the m^{6}G·T base pair in the major groove is different from that of a Watson–Crick A·T or G·C base pair in two additional details: (*i*) there are two methyl groups, one attached to each base, and (*ii*) there is no heteroatom with a capacity to donate hydrogen bonds.

When a guanine base in genomic DNA is converted to m^{6}G by a chemical mutagen, the modified base normally codes for thymine instead of cytosine (26), resulting in an *in vivo* mutation frequency of between 15% (27) and 75% (28). The *in vitro* misinsertion frequency is >95% (29) and the only efficient form of repair involves demethylation of the O^6 atom of guanine by the enzyme m^{6}G methyltransferase to regenerate guanine (30, 31). Thus, the m^{6}G·T base pair is recognized as being more similar to a Watson–Crick base pair than is the m^{6}G·C base pair. The reasons for this are unlikely to be thermodynamic in origin (18, 19), as ultraviolet melting

Table 1. Thermodynamic parameters for d(CGCXAATTYGCG)$_2$ duplexes at pH 7.0 in 1.0 M NaCl/10 mM sodium phosphate/1 mM EDTA

Base pair	$\Delta H°$, kJ·mol^{-1}	$\Delta S°$, J·mol^{-1}·K^{-1}	$\Delta G°$, kJ·mol^{-1}	t_m, K
G·C	−430.1	−1164	−83.2	344.6
A·T	−399.0	−1085	−75.7	341.6
G·T	−331.7	−937.6	−52.3	324.4
m^{6}G·C	−213.7	−592.3	−37.2	317.1
m^{6}G·T	−208.4	−577.3	−36.4	315.2

Thermodynamic parameters were determined from the concentration dependence of ultraviolet melting by standard methods (18, 19). Each point on the curve was measured in triplicate. t_m, melting temperature at 40 μM oligonucleotide; X, any nucleotide; Y, any nucleotide except that represented by X.

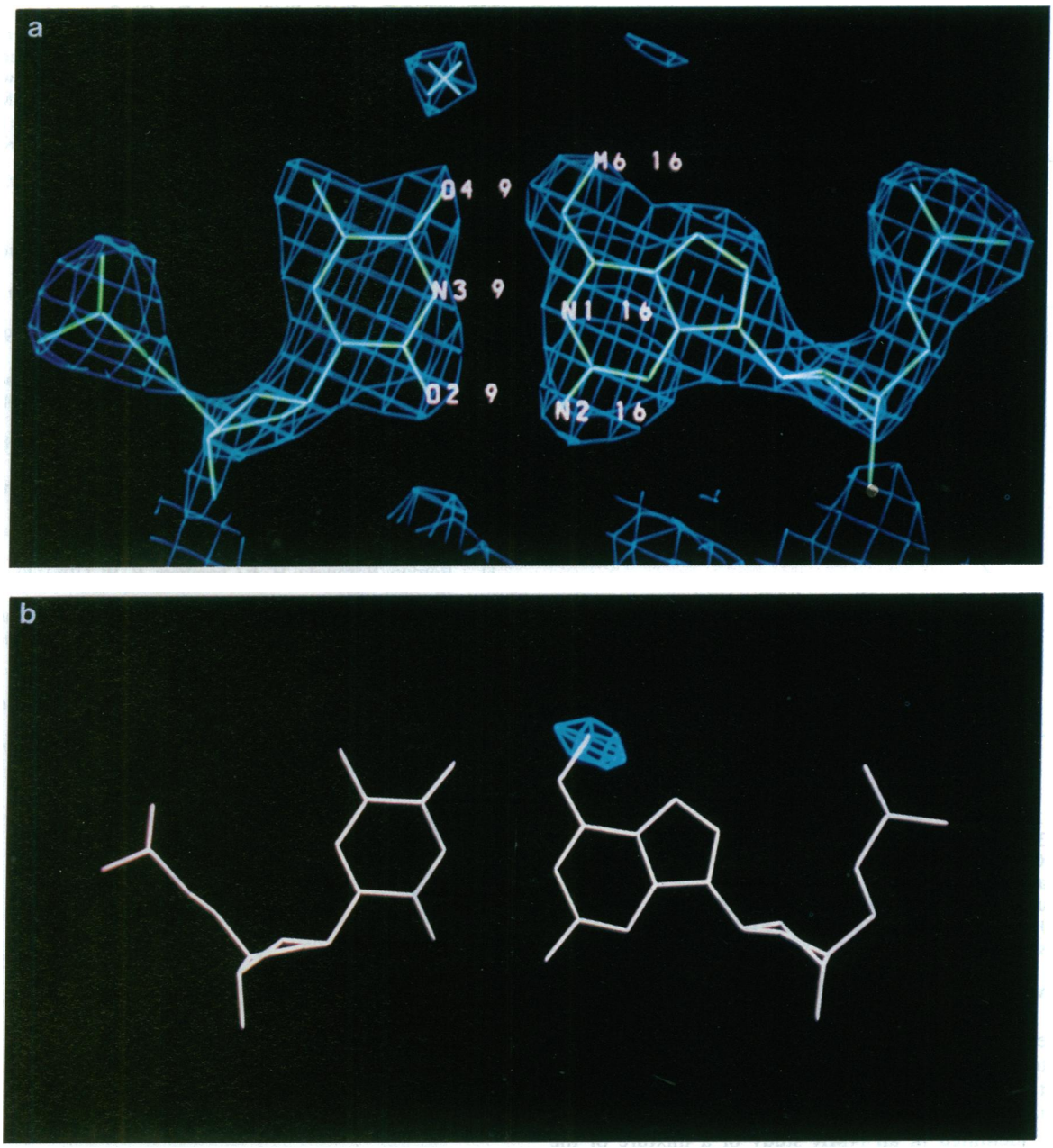

FIG. 1. (*a*) $2F_o - F_c$ map of the m^{6}G(9)·T(16) base pair. Some key atoms are labeled and a solvent molecule is highlighted in the major groove, within hydrogen bonding distance of thymine-O^4. (*b*) $F_o - F_c$ difference density map of the m^{6}G(9)·T(16) base pair with the guanine methyl group removed from the structure factor calculations. The methyl group is located in the proximal conformation in the plane of the purine ring.

studies show that the duplex containing the m^{6}G·T base pair is slightly less stable than the corresponding m^{6}G·C duplex at neutral pH ($\Delta\Delta G^\circ$ = 0.8 kJ/mol) (Table 1). Moreover, the m^{6}G·T duplex is less stable than the duplex containing the G·T wobble base pair mismatch, which is rarely incorporated during replication due to efficient proofreading ($\Delta\Delta G^\circ$ = 15.9 kJ/mol).

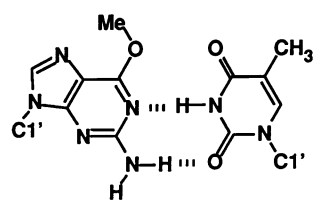

FIG. 2. m^{6}G·T base pair with each base as its major tautomer.

The incorporation of the m^{6}G·T base pair in preference to the m^{6}G·C base pair in genomic DNA can be rationalized in structural terms. The similarity in shape between the m^{6}G·T mispair and a Watson–Crick base pair, particularly in the minor groove is striking, whereas the m^{6}G·C base pair has been postulated on the basis of theoretical and NMR studies to be a reverse wobble base pair as in Fig. 4*a* (32, 25). This would, by analogy with mismatch base pairs, be removed by proofreading. Thus the enzymes in the cell nucleus responsible for DNA synthesis and repair discriminate in favor of the mutagenic lesion. We have studied the stability profile of the DNA duplex containing the m^{6}G·C base pair over a wide pH range (Fig. 3*b*) and we have observed that the melting temperature falls slightly from pH 8.5 to pH 6.5, as would be expected for the unprotonated wobble base pair 4a. At lower pH however, there is a clear indication of duplex stabilization, probably due to protonation of cytosine-N^3 of the

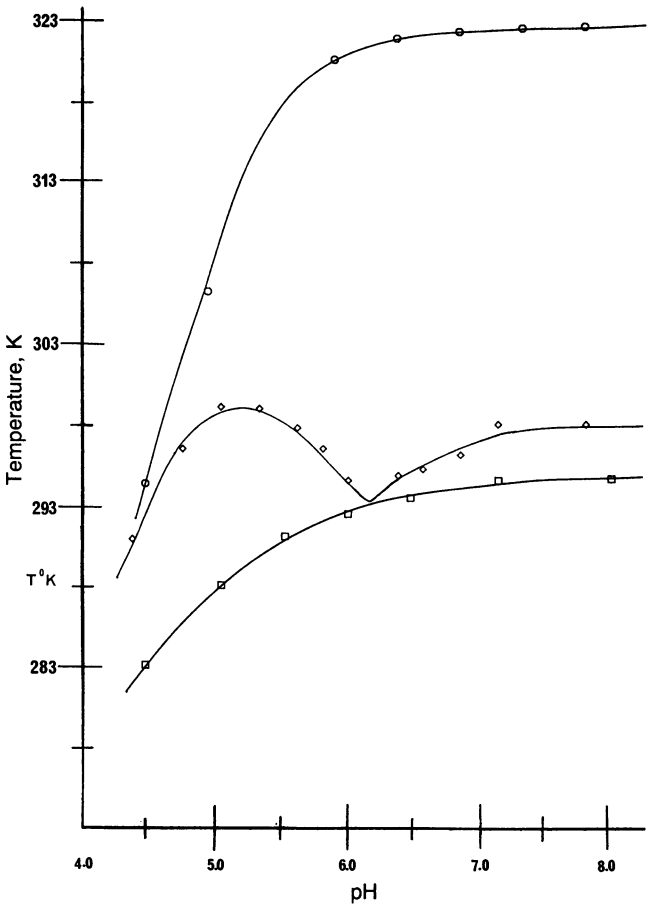

FIG. 3. pH-dependence of ultraviolet melting for d(CGCm^{6}GAATTTGCG) (squares), d(CGCm^{6}GAATTCGCG) (diamonds), and d(CGCGAATTCGCG) control (circles). Ultraviolet melting curves were measured over a wide pH range at 264 nm in aqueous 0.1 M sodium phosphate/1 mM EDTA. Each point on the curve was measured in triplicate.

mispair, which would lead to the formation of a Watson–Crick-like base pair, a resonance form of which is shown in Fig. 4*b*. Thus, the m^{6}G·C base pair displays conformational flexibility. The protonated base pair 4b, which is probably present to some extent at neutral pH, would be expected on structural grounds to be incorporated during replication. Such a base pair has been identified in an NMR study of a mixture of the nucleosides in nonaqueous solvents (21) and in a preliminary report of an x-ray structure of a Z-DNA duplex (33). The occasional occurrence of 4b might explain why the presence of m^{6}G in genomic DNA does not always lead to a mutation.

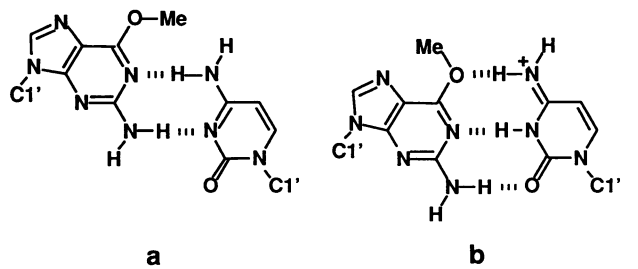

FIG. 4. (*a*) m^{6}G·C wobble base pair. (*b*) Protonated m^{6}G·C$^+$ base pair.

We thank Dr. S. V. Phillips and Dr. W. Summers for valuable guidance in the use of the Rigaku RU200 HB diffractometer, rotating anode, and Nicolet–Siemens X100 area detector system in the Department of Biophysics at Leeds University. This work was funded by a Science and Engineering Research Council Molecular Recognition Initiative grant and a Science and Engineering Research Council CASE Studentship with Shell Research Ltd, U.K.

1. Miller, E. C. & Miller, J. A. (1981) *Cancer* **47**, 2327–2345.
2. Searle, C. E. (1986) *Chem. Br.* **29**, 211–220.
3. Ames, B. N. (1979) *Science* **204**, 587–593.
4. Karran, P. & Marinus, M. G. (1982) *Nature (London)* **296**, 868–869.
5. Saffhill, R., Margison, G. P. & O'Connor, P. (1985) *Biochim. Biophys. Acta* **823**, 111–145.
6. Loechler, E. L., Green, C. L. & Essigman, J. M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6271–6275.
7. Mitra, G., Pauly, G. T., Kumar, R., Pei, G. K., Hughes, S. H., Moschel, R. C. & Barbacid, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8650–8654.
8. Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. (1982) *Nature (London)* **300**, 149–152.
9. Zarbl, H., Sukumar, S., Arthur, A. V., Martin-Zanka, D. & Barbacid, M. (1985) *Nature (London)* **315**, 382–385.
10. Kuzmich, S., Marky, L. A. & Jones, R. A. (1983) *Nucleic Acids Res.* **11**, 3393–3404.
11. Borowy-Borowski, H. & Chambers, R. W. (1987) *Biochemistry* **26**, 2465–2471.
12. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1980) *Nature (London)* **287**, 755–758.
13. Kabsch, W. (1988) *J. Applied Cryst.* **21**, 916–924.
14. Sheldrick, G. M. (1976) *Shelx76 System of Computing Programs* (Univ. Cambridge, U.K.).
15. Hendrickson, W. A. & Konnert, J. H. (1981) in *Biomolecular Structure, Conformation, Function and Evolution*, ed. Srinivasan, R. (Pergamon, Oxford), Vol. 1, 43–57.
16. Westhof, E., Dumas, P. & Moras, D. (1985) *J. Mol. Biol.* **184**, 119–145.
17. Jones, T. A. (1978) *J. Applied Cryst.* **11**, 268–272.
18. Gaffney, B. L. & Jones, R. A. (1989) *Biochemistry* **28**, 5881–5889.
19. Gaffney, B. L., Marky, L. A. & Jones, R. A. (1984) *Biochemistry* **23**, 5686–5691.
20. Nagata, C. & Aida, M. (1988) *J. Mol. Struct. (THEOCHEM)* **179**, 451–466.
21. Williams, L. D. & Shaw, B. R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1779–1783.
22. Patel, D. J., Shapiro, L., Kozlowski, S., Gaffney, B. L. & Jones, R. A. (1988) *Biochemistry* **25**, 1036–1042.
23. Yamagata, Y., Kohda, K. & Tomita, K. (1988) *Nucleic Acids Res.* **16**, 9307–9321.
24. Parthasarathy, R. & Fridey, S. M. (1986) *Carcinogenesis* **7**, 221–227.
25. Pedersen, L. G., Darden, T. A., Deerfield, D. W., II, Anderson, M. W. & Hoel, D. G. (1988) *Carcinogenesis* **9**, 1553–1562.
26. Snow, E. T., Foote, R. S. & Mitra, S. (1984) *J. Biol. Chem.* **259**, 8095–8100.
27. Chambers, R. W., Sledziewska-Gojska, E., Hirani-Hojatti, S. & Borowy-Borowski, H. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7173–7177.
28. Bhanot, O. S. & Ray, A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7348–7352.
29. Toorchen, D. & Topal, M. D. (1983) *Carcinogenesis* **4**, 1591–1597.
30. Olsson, M. & Lindahl, T. (1980) *J. Biol. Chem.* **255**, 10569–10571.
31. Demple, B., Jacobsson, A., Olsson, M., Robins, P. & Lindahl, T. (1982) *J. Biol. Chem.* **257**, 13776–13780.
32. Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L. & Jones, R. A. (1986) *Biochemistry* **25**, 1027–1036.
33. Ginell, L., Narendra, N., Jones, R. A. & Berman, H. M. (1990) *Biophys. J.* **57**, 452 (abstr.).