5524–5532 Nucleic Acids Research, 2013, Vol. 41, No. 10 Published online 10 April 2013
doi:10.1093/nar/gkt198

# Structures of DNA duplexes containing $O^6$-carboxymethylguanine, a lesion associated with gastrointestinal cancer, reveal a mechanism for inducing pyrimidine transition mutations

**Fang Zhang[1], Masaru Tsunoda[2], Kaoru Suzuki[2], Yuji Kikuchi[1,2], Oliver Wilkinson[3], Christopher L. Millington[3], Geoffrey P. Margison[4], David M. Williams[3,*], Ella Czarina Morishita[5] and Akio Takénaka[1,2,5,*]**

[1]Graduate School of Science and Engineering, Iwaki-Meisei University, Iwaki 970-8551, Japan, [2]Faculty of Pharmacy, Iwaki-Meisei University, Iwaki 970-8551, Japan, [3]Department of Chemistry, Centre for Chemical Biology, Krebs Institute, University of Sheffield, Sheffield S3 7HF, UK, [4]Cancer Research-UK Carcinogenesis Group, Paterson Institute for Cancer Research, University of Manchester, Manchester M20 9BX, UK and [5]Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama 226-8501, Japan

Received January 29, 2013; Revised February 26, 2013; Accepted March 1, 2013

## ABSTRACT

***N*-nitrosation of glycine and its derivatives generates potent alkylating agents that can lead to the formation of $O^6$-carboxymethylguanine ($O^6$-CMG) in DNA. $O^6$-CMG has been identified in DNA derived from human colon tissue, and its occurrence has been linked to diets high in red and processed meats. By analogy to $O^6$-methylguanine, $O^6$-CMG is expected to be highly mutagenic, inducing G to A mutations during DNA replication that can increase the risk of gastrointestinal and other cancers. Two crystal structures of DNA dodecamers d(CGCG[$O^6$-CMG]ATTCGCG) and d(CGC[$O^6$-CMG]AATTCGCG) in complex with Hoechst33258 reveal that each can form a self-complementary duplex to retain the B-form conformation. Electron density maps clearly show that $O^6$-CMG forms a Watson–Crick–type pair with thymine similar to the canonical A:T pair, and it forms a reversed wobble pair with cytosine. *In situ* structural modeling suggests that a DNA polymerase can accept the Watson–Crick–type pair of $O^6$-CMG with thymine, but might also accept the reversed wobble pair of $O^6$-CMG with cytosine. Thus, $O^6$-CMG would permit the mis-incorporation of dTTP during DNA replication. Alternatively, the triphosphate that would be formed by carboxymethylation of the nucleotide triphosphate pool d[$O^6$-CMG]TP might compete with dATP incorporation opposite thymine in a DNA template.**

## INTRODUCTION

Diets high in red and especially processed meats are known to be risk factors of colorectal cancer, which is one of the most common cancers worldwide (Globocan 2008 Cancer fact sheet, http://globocan.iarc.fr/factsheets/cancers/colorectal.asp). Promutagenic lesions $O^6$-methylguanine ($O^6$-MeG) and $O^6$-carboxymethylguanine ($O^6$-CMG) (Figure 1a) are commonly found in colorectal DNA, and their frequency might be indicative of a risk factor of colorectal cancer (1–3). One route that can lead to the formation of such lesions involves the initial nitrosation of amino acids, such as glycine and derivatives thereof, e.g. *N*-glycyl-peptides and the bile acid conjugate glycocholic acid. Nitrosation derives from reaction at neutral or alkaline pH with dinitrogen trioxide ($N_2O_3$), which in turn is generated by the oxidation of NO (4), from dietary nitrite or after exposure to ionizing radiation (5).

*To whom correspondence should be addressed. Tel: +81 246 29 5354; Fax: +81 246 29 5354; Email: atakenak@sakura.email.ne.jp
Correspondence may also be addressed to David M. Williams. Tel: +44 114 222 9502; Fax: +44 114 222 9346; Email: d.m.williams@sheffield.ac.uk
Present addresses:
Oliver Wilkinson, Genome Damage and Stability Centre (GDSC), University of Sussex, Falmer, Brighton, East Sussex, BN1 9RQ, UK.
Christopher L. Millington, Institut de Génétique et Développement de Rennes, Université de Rennes, 35043 Rennes Cedex, France.
Geoffrey P. Margison, Centre for Occupational and Environmental Health, Faculty of Medical and Human Sciences, University of Manchester, Manchester M13 9PL, UK.
Ella Czarina Morishita, Institute of Molecular and Cellular Biosciences, The University of Tokyo, Tokyo 113-0032, Japan.
Akio Takénaka, Research Institute, Chiba Institute of Technology, Tsudanuma, Narashino 275-0016, Japan.

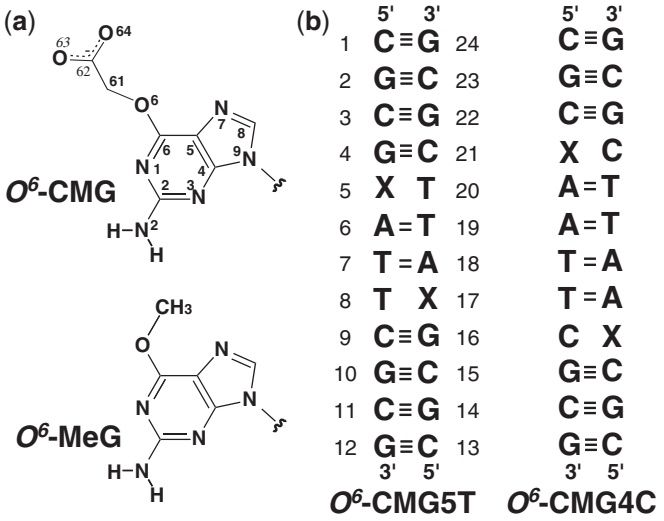


**Figure 1.** Chemical structures of $O^6$-CMG with the atomic numbering and CMG (**a**), and the sequences and numbering schemes of $O^6$-CMG-containing DNA duplexes (**b**). X indicates $O^6$-CMG.

*N*-Nitrosoglycine is converted into diazoacetate or α-lactone (6,7), potent mutagens that can alkylate guanine in DNA to form $O^6$-CMG and $O^6$-MeG (8). In humans, $O^6$-methylguanine-DNA methyltransferase (MGMT) repairs DNA containing a wide variety of different $O^6$-alkylguanine lesions by transferring the alkyl group to the thiolate side chain of the active site Cys (9). Recently, we have shown for the first time that DNA containing $O^6$-CMG is also a substrate for MGMT (10). *In vivo* and *in vitro* evidence suggests that $O^6$-CMG predominantly induces G:C→A:T transition mutations (3,11), implying that $O^6$-CMG within a DNA template not only directs the incorporation of complementary dCTP but also allows the mis-incorporation of non-complementary dTTP into the newly synthesized DNA. In general, DNA polymerases accept only the Watson–Crick–type pairs in the B-form conformation (12–14). To understand the mechanism of such mutations, it is necessary to reveal the pairing or interaction geometry of $O^6$-CMG with T and C bases.

Here, we describe the crystal structures of DNA duplexes containing $O^6$-CMG at residue positions that place the modified base opposite T or C in the palindromic B-form Dickerson–Drew sequence d(CGCGAATTCGCG) (15).

## MATERIALS AND METHODS

### Oligodeoxyribonucleotide synthesis and purification

Oligodeoxyribonucleotides (ODNs) with the sequences d(CGCG[$O^6$-CMG]ATTCGCG) and d(CGC[$O^6$-CMG]A ATTCGCG) were synthesized and purified by reversed-phase HPLC as described previously (16), and they were characterized by ESI–mass spectrometry. For crystallization, the samples in pure water were purified on an ÄKTAprime plus (GE Healthcare) using a Superdex 30 pg 16/60 column at flow rate of 0.5 ml/min with a gradient of 0–100% of pH 7.2 buffer solution (50 mM $NaH_2PO_4$ and 150 mM NaCl); the ODN-containing fractions monitored by a UV monitor were confirmed by PAGE analysis with TBE. Finally, the ODNs were desalted by a series of C18 (Waters Corp.), AG50W-X8 (BioRad Co.) and Chelex 100 (BioRad Co.) columns, in turn. The eluted solutions were dried *in vacuo* at room temperature to store the samples.

### Crystallization and data collection

Initial screenings of crystallization conditions were performed at 277 K by the hanging-drop vapor diffusion method using a kit for nucleic acids reported by Berger *et al.* (17). Two-microliter droplets were equilibrated against 700 μl of the reservoir solution. The optimized conditions for obtaining $O^6$-CMG5T and $O^6$-CMG4C crystals were as follows. For $O^6$-CMG5:T, a droplet of 40 mM sodium cacodylate buffer solution at pH 7.0 containing 1 mM ODN, 10% (v/v) 2-methyl-2,4-pentanediol (MPD), 12 mM spermine tetrahydrochloride, 80 mM sodium chloride, 12 mM potassium chloride, 20 mM magnesium chloride and 1 mM Hoechst33258 (2′-(4-hydroxyphenyl)-6-(4-methyl-1-piperazinyl)-2,6′-bi-*1H*-benzimidazole) was equilibrated against 35% (v/v) MPD. For $O^6$-CMG4C, a droplet of 40 mM sodium cacodylate buffer solution at pH 7.0 containing 1 mM ODN, 10% (v/v) MPD, 12 mM spermine tetrahydrochloride, 40 mM lithium chloride, 80 mM strontium chloride and 1 mM Hoechst 33258 was equilibrated against 35% (v/v) MPD.

$O^6$-CMG5T and $O^6$-CMG4C crystals suitable for X-ray data collections were picked up from their droplets with a nylon loop (Hampton Research) and transferred into liquid nitrogen. X-Ray diffraction experiments of these crystals were performed at 100 K with synchrotron radiation ($\lambda = 1.00$ Å at BL-5 A and 0.98 Å at BL-17A) of the Photon Factory in Tsukuba (Japan). Diffracted intensities were recorded on a CCD detector Quantum 315r positioned 200.0 and 155.4 mm from $O^6$-CMG5T and $O^6$-CMG4C crystals, respectively. A total of 180 frames of the patterns for one crystal were taken at 1° oscillation steps with 1 s exposure per frame. Raw diffraction images were indexed, and intensities around Bragg spots were integrated using the computer programs, *HKL2000* (18) for $O^6$-CMG5T data and *Mosflm*(19)-*Scala* (20) of the *CCP4* suite (21) for $O^6$-CMG4C data. To compensate for the overloaded reflections, the intensity data were merged with those collected at different exposure doses. The crystal data and statistics of data collection are summarized in Table 1.

### Structure determination and refinement

Using the program autoMR in the *CCP4* suite (21), the phases of the two data sets were separately estimated by the molecular replacement method with the unmodified ODN structure d(CGCGAATTCGCG) (22) as a probe. The atomic parameters were refined using the maximum-likelihood least-squares technique in *REFMAC5* (23) of *CCP4* and *CNS* (24). The crystal structures were constructed and modified by adding other molecules and ions using the program Coot (25) in *CCP4*. The resultant structures were validated by interpretation of OMIT maps at every nucleotide residue. Electron densities assignable to a magnesium ion and three strontium ions were found in $O^6$-CMG5T and $O^6$-CMG4C, respectively, and these

**Table 1.** Crystal data and statistics of data collection and structure refinement

| Crystal code | $O^6$-CMG5T | $O^6$-CMG4C |
|---|---|---|
| Crystal data | | |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ |
| Unit cell (Å) | | |
| *a* | 25.1 | 24.6 |
| *b* | 40.4 | 41.2 |
| *c* | 65.7 | 63.1 |
| Data collection | | |
| Resolution range (Å) | 50.0–1.94 | 25.1–1.54 |
| Outer shell (Å) | 1.97–1.94 | 1.62–1.54 |
| Observed reflections | 27 766 | 62 239 |
| Unique reflections | 4877 | 9559 |
| Completeness (%)[a] | 91.0 (82.9) | 95.8 (100) |
| $R_{merge}$ (%)[b] | 5.7 (52.2) | 8.9 (28.5) |
| $I/\sigma$ | 8.1 (0.7) | 14 (6.4) |
| Redundancy[c] | 5.7 (1.4) | 6.5 (6.7) |
| Structure refinement | | |
| Resolution range (Å) | 34.4–1.94 | 25.1–1.54 |
| Used reflections | 3842 | 9017 |
| $R_{factor}$ (%)[d] | 23.9 | 19.4 |
| $R_{free}$ (%)[e] | 26.7 | 23.6 |
| Rms deviation | | |
| Bond distances (Å) | 0.05 | 0.03 |
| Bond angles (degree) | 4.3 | 3.8 |
| No. of DNA duplex | 1 | 1 |
| No. of Hoechst33258 | 1 | 1 |
| No. of ions | $1Mg^{2+}$ | $3Sr^{2+}$ |
| No. of water molecules | 43 | 135 |

[a]Values in parentheses indicate those in the outer shell.
[b]$R_{merge} = 100 \times \Sigma_{hj}|I_{hj} - <I_h>| / \Sigma_{hj}I_{hj}$, where $I_{hj}$ is the *j*th measurement of the intensity of reflection **h** and $<I_h>$ is its mean value.
[c]Diffraction patterns of 1° oscillation ranges were collected in 180 frames of X5T and X4C.
[d]*R*-factor $= 100 \times \Sigma||F_o| - |F_c|| / \Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively.
[e]Calculated using a random set containing 5% observations that were not included during refinement (26).

cations were included in the subsequent refinements. The structural restraints applied initially on DNA and Hoechst33258 were released. The $R_{factor}$ and $R_{free}$ values converged with further rounds of the structure refinements.

The statistics of structure refinements are summarized in Table 1. The atomic parameters of $O^6$-CMG5T and $O^6$-CMG4C crystal structures were deposited in the protein data bank (PDB-IDs = 4ITD and 4IJ0). Figure 2 shows the $2F_o$-$F_c$ electron density maps of the modified nucleotide and its partner in the pair formation, together with the corresponding $F_o$-$F_c$ omit maps calculated without the pairs. These maps were depicted by the program *Dino* (27). All the residues were traced on the electron density maps. The $O^6$-CMG residues assigned on $F_o$-$F_c$ omit maps of $O^6$-CMG5T and $O^6$-CMG4C fit well into the $2F_o$-$F_c$ maps. All the global and local helical parameters, as well as the torsion angles and pseudo-rotation phase angles of sugar rings, were calculated using the program 3DNA (28). Some of them are given in Table 2.

## RESULTS AND DISCUSSION

### Overall structure of DNA duplexes

At the initial stage of crystallization condition survey, it was difficult to crystallize the two ODNs. However, by adding Hoechst33258 as a dye to stabilize (30) the duplex formation in ODN solutions, single crystals suitable for X-ray analyses were obtained. As shown in Figure 3, the two homododecamers in both $O^6$-CMG5T and $O^6$-CMG4C crystals are associated with each other to form a right-handed double helix. Their average local helical parameters (Table 2) are close to those of the high-resolution B-form DNA duplexes (22,30,31). However, superimposition onto the unmodified structures (Figure 3c) reveals the local variations in the backbone conformations of the duplexes, the root-mean-square (rms) deviations being 1.2 Å for $O^6$-CMG5T and 1.4 Å for $O^6$-CMG4C. Although the largest deviations occurs at the $O^6$-CMG residues in both $O^6$-CMG5T and $O^6$-CMG4C duplexes, their sugar puckers fluctuate around the $C^{2'}$-*endo* conformation, which is the conformation typically found in B-type DNA. These data indicate that carboxymethylation of guanine residues does not significantly affect the overall DNA conformation. The slightly large rms deviation of $O^6$-CMG4C may be related to the interaction geometry between $O^6$-CMG and C, which will be described later in detail. Figure 3 shows Hoechst33258 molecules bound in the minor grooves of $O^6$-CMG5T and $O^6$-CMG4C. They seem to stabilize the duplex structures with no drastic changes in the base pair geometry, as the modified sites are in the major groove. Similar examples are already found in other structures of DNA duplexes crystallized in the presence of this duplex-stabilizing dye (32,33). Only a slight change is found in the groove width; the central shortest *P*...*P* distance between the two ODN backbones is 8.9 Å in the absence of Hoechst33258 (23), whereas those of $O^6$-CMG5T and $O^6$-CMG4C are 9.4 and 9.9 Å, respectively. Such expansions ~1.0 Å are commonly found in the referred examples (32,33).

### Formation of $O^6$-CMG:T and $O^6$-CMG:C base pairs

All unmodified base pairs in the dodecamer sequences except those containing $O^6$-CMG are of the standard Watson–Crick type. $F_o$-$F_c$ omit and $2F_o$-$F_c$ electron density maps of the $O^6$-CMG5T duplex indicate that the $O^6$-CMG residues form base pairs with the two opposite T residues in the palindromic sequence (Figure 2a and b). The shapes of the densities of these base pairs except the carboxymethyl groups are similar to the canonical Watson–Crick A:T pair found in the original Dickerson–Drew dodecamer (16). On the other hand, although $F_o$-$F_c$ omit maps and $2F_o$-$F_c$ electron density maps of the $O^6$-CMG4C duplex (Figure 2c and d) show that the $O^6$-CMG residues form base pairs with the opposite C residues, the two geometries of the $O^6$-CMG:C pairs are different from the standard G:C pair geometry found in the unmodified dodecamer.

### Geometries of the $O^6$-CMG:T and $O^6$-CMG:C base pairs

The interaction geometries of base paired $O^6$-CMGs are shown in Figure 4 and summarized in Table 2. In the $O^6$-CMG5:T20 pair, the interatomic distances between the $N^1$ atom of $O^6$-CMG and the $N^3$ atom of T and between the $N^2$ atom of $O^6$-CMG and the $O^2$ atom of T are 3.3 and 3.1 Å, respectively. In the $O^6$-CMG17:T8 pair,

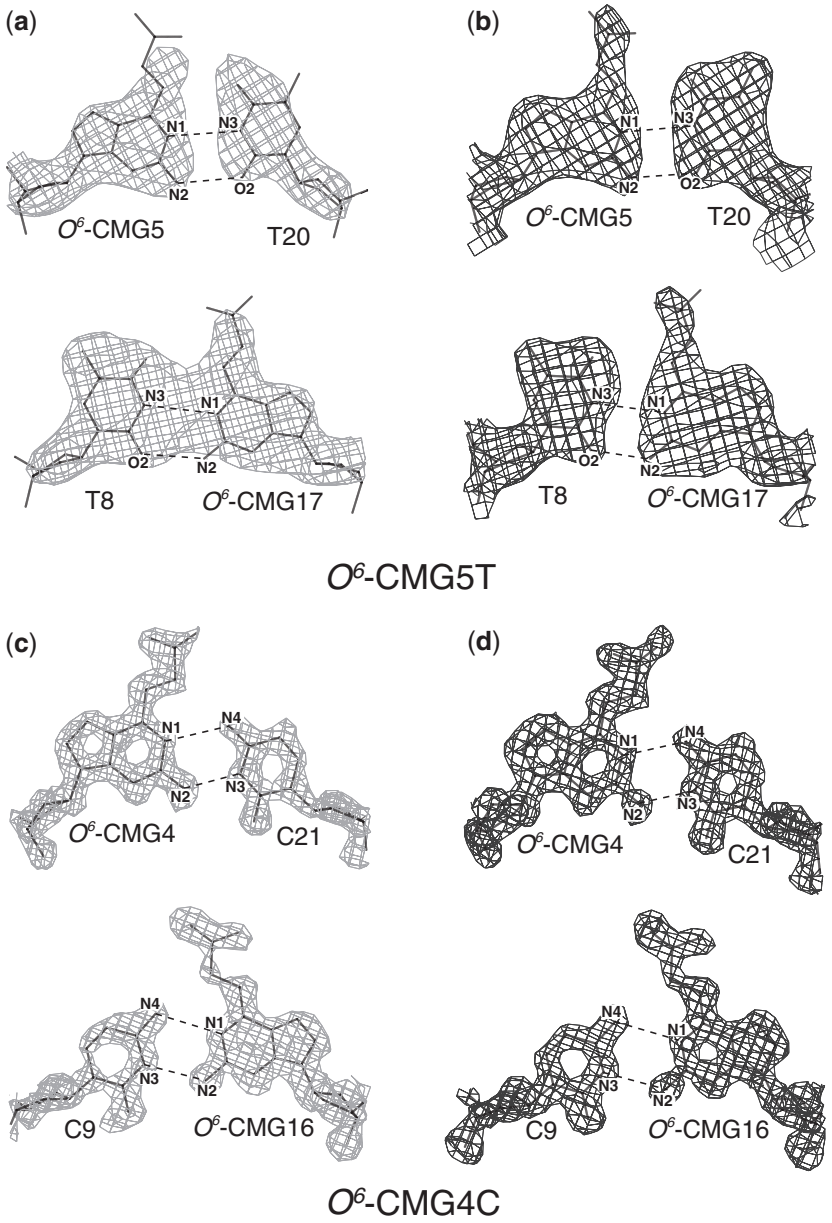


**Figure 2.** $F_o$-$F_c$ omit maps (in the left column) of $O^6$-CMG5:T20 pairs and T8:$O^6$-CMG17 pairs (**a**), and $2F_o$-$F_c$ maps (in the right column) of $O^6$-CMG5:T20 pairs and T8:$O^6$-CMG17 pairs (**b**). The corresponding maps of $O^6$-CMG4:C21 pairs and C9:$O^6$-CMG16 pairs (**c**) and those of $O^6$-CMG4:C21 pairs and C9:$O^6$-CMG16 pairs (**d**). Broken lines show possible hydrogen bonds. Electron densities are contoured at the 2.6, 3.0, 2.5 and 2.0 σ levels for the top-to-bottom pairs in the left column, and the corresponding $2F_o$-$F_c$ maps are contoured at the 0.8, 0.8, 1.0 and 1.0 σ levels in the right column.

the corresponding distances are 3.4 and 3.0 Å, respectively. These values suggest base pair formation between $O^6$-CMG and T. Thus, although the natural base G forms a wobble base pair with T (34), the $O^6$-CMG:T pair is similar to the canonical Watson–Crick A:T pair. In this $O^6$-CMG:T pair, although the electronegative $O^6$ atom of $O^6$-CMG is exposed to the $O^4$ atom of T, the ensuing repulsion between these two sites is reduced by propeller-twisting between the paired bases. The twisting angles are $-18^\circ$ on average at the two sites. The $N^1(O^6\text{-CMG})\ldots N^3(\text{T})$ distances look slightly longer, perhaps because of the $O^6(O^6\text{-CMG})\ldots O^4(\text{T})$ repulsion, which separates the $O^6$ and $O^4$ atoms at an average distance of 3.6 Å. The interaction geometries of $O^6$-CMG:T pairs are similar to those of $O^6$-MeG:T pairs found in the $O^6$-MeG-containing B-DNA (35). As the carboxymethyl groups of the $O^6$-CMG residues protrude into the major groove, they do not drastically alter the overall DNA conformation. However, as shown in Figure 2, the electron densities of the terminal carboxyl groups are not clear, suggesting that they are disordered in the solvent region.

In the $O^6$-CMG4:C21 pair, the interatomic distances between the $N^1$ atom of $O^6$-CMG and the $N^4$ atom of C

**Table 2.** Average local helical parameters and base pair parameters[a]

| | X-Displacement (Å) | Inclination (°) | Helical twist (°) | Helical rise (Å) |
|---|---|---|---|---|
| $O^6$-CMG5T | 0.57 | −0.26 | 36 | 3.3 |
| $O^6$-CMG4C | 0.67 | 0.73 | 36 | 3.3 |
| B-DNA[b] | 0.05 | 2.1 | 36.5 | 3.3 |
| A-DNA[b] | −4.2 | 15 | 32.5 | 2.8 |

| Base pair | Propeller twisting (°) | Opening angle (°) | Buckle angle (°) |
|---|---|---|---|
| $O^6$-CMG :T | −18 | 8.6 | 13 |
| $O^6$-CMG :C | −9.4 | 3.1 | 19 |
| Wobble pair | $C^{1'}\ldots C^{1'}$ distance (Å) | $\lambda_I$ angle (°) | $\lambda_{II}$ angle (°) |
| $O^6$-CMG4:C21 | 11 | 69 | 42 |
| C9:$O^6$-CMG16 | 11 | 44 | 69 |

[a]Calculated with the program 3DNA (29)
[b]High-resolution A- and B-form DNA structures by Olson *et al.* (30)

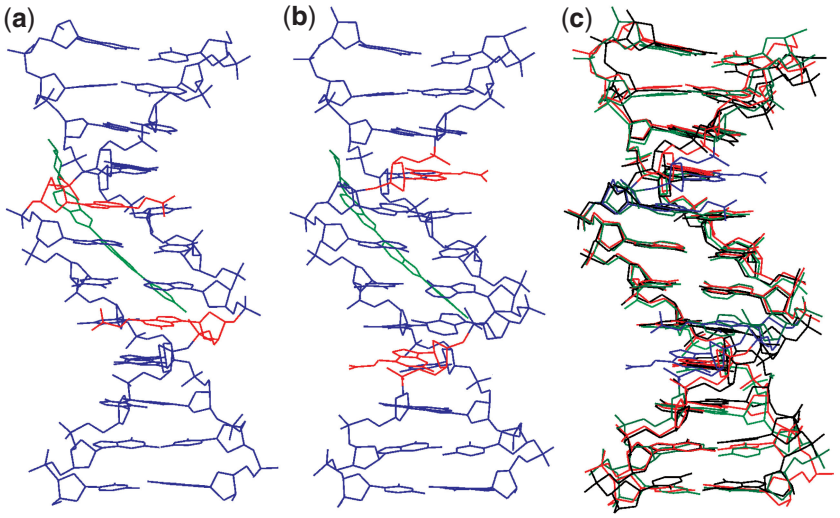


**Figure 3.** Overviews of $O^6$-CMG5T duplex (**a**) and $O^6$-CMG4C duplex (**b**) and their superimposition onto the unmodified duplex (black) (**c**). In (a) and (b), the $O^6$-CMG residues and Hoechst33258 molecules are colored in red and green, respectively. In (c), $O^6$-CMG5T, $O^6$-CMG4C and the unmodified duplexes are colored in red, green and black, respectively, and the modified residues are in blue.

and between the $N^2$ atom of $O^6$-CMG and the $N^3$ atom of C are 3.0 and 2.9 Å, respectively. In the $O^6$-CMG16:C9 pair, the corresponding distances are the same 3.1 and 2.9 Å, respectively. These values indicate that $O^6$-CMG can form a base pair with C through hydrogen bonds at these two sites. In both base pairs, the purine moiety of $O^6$-CMG moves towards the major groove side, whereas C remains by its original position (Figure 5). This deformation occurs at both modified sites.

Although this pairing mode has been referred to as a wobble base pair (36,37), we refer to the $O^6$-CMG:C pair as a reversed wobble pair, as in the typical wobble pair, G moves to the minor groove, and U or T shifts to the major groove. Such a reversed wobble pair has been found between G and 5-formyluracil, an analog that derives after oxidation of T with oxygen radicals (38). Another example is found in an ODN-containing $O^6$-ethylguanine ($O^6$-EtG) at the fourth position of the Dickerson sequence, where one of the two base pairs is of the reversed wobble type (39,40). In many cases, this wobbling makes the $C^{1'}\ldots C^{1'}$ distance longer by 1.0 Å and the $\lambda_I$ and $\lambda_{II}$ angles asymmetric (Table 2), as compared with those of the unmodified pairs.

The carboxyl groups of $O^6$-CMG (in its pairing with C) are clearly visible on the high resolution maps, as shown in Figure 2. They adopt a *syn* conformation against the $N^1$ atom of $O^6$-CMG. In contrast, the alkyl groups of $O^6$-EtG:C pairs (39,40) adopts an *anti* conformation and protrudes into the major groove. The longer carboxymethyl groups of $O^6$-CMG protrude into the major groove to make contacts with water molecules. The reversed wobble $O^6$-CMG:C pairs seem to be further stabilized by two additional water-mediated hydrogen bonds between an oxygen atom of the carboxyl group of $O^6$-CMG and the $N^4$ atom of C and between the $N^2$ atom of $O^6$-CMG and the $O^2$ atom of C. Three Sr atoms are all hydrated with eight water molecules and bound to the minor groove, major groove and

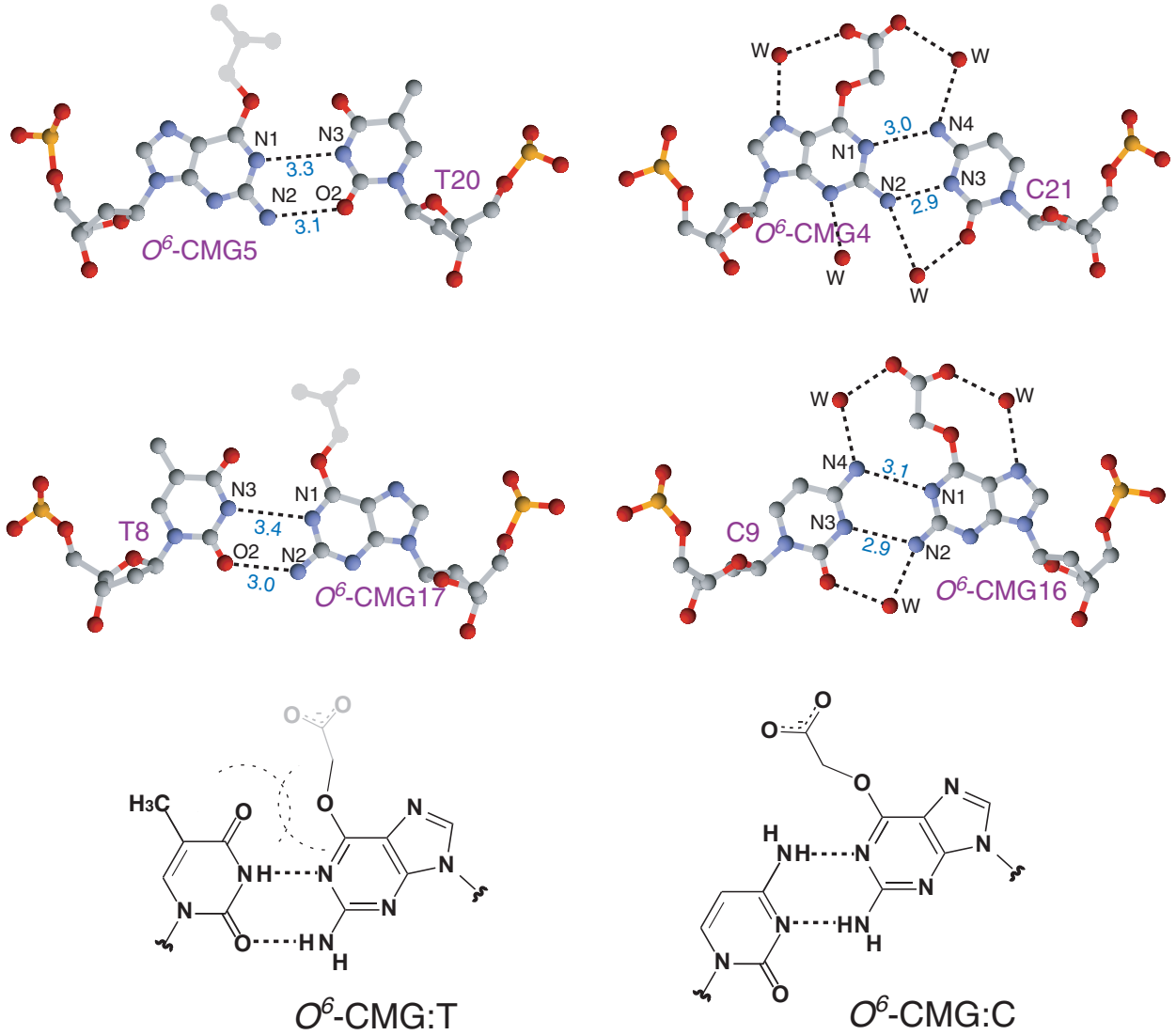


**Figure 4.** Watson–Crick–type pairs between $O^6$-CMG and T (left column) and reversed wobble pairs between $O^6$-CMG and C (right column). Broken lines suggest possible hydrogen bonds, and values indicate atomic distances in Å. The corresponding chemical structures of interacting $O^6$-CMGs are shown in the bottom. The carboxyl groups of $O^6$-CMG:T pairs are invisible because of disorder. Character W indicates a water oxygen atom.

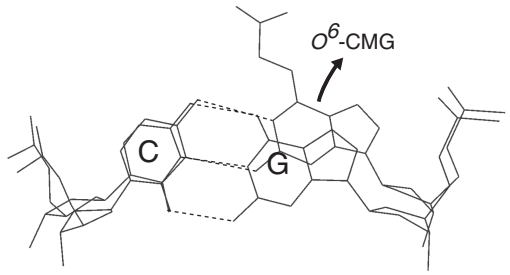


**Figure 5.** A comparison of paired-base positions when $O^6$-CMG4C and the unmodified duplexes are superimposed between the phosphate backbones. An arrow indicates that only the modified base wobbles towards the major groove side.

phosphate backbone, respectively, so that they do not influence the base pairs formed by $O^6$-CMG. The same preference is also seen for the modified base in the $O^6$-CMG:C base pair. The change in preference for the orientation of the alkyl group in $O^6$-CMG compared with $O^6$-MeG may result from additional interactions of the carboxyl group of the alkyl side chain with the $O^4$ and $N^4$ atoms of T and C, respectively, that are implied from the crystal structures.

## Biological implications

From the present study, it has been found that $O^6$-CMG can form base pairs with both thymine and cytosine, and the pairing modes are Watson–Crick type and reversed wobble type, respectively. To examine a possibility if the two pair formations of $O^6$-CMG:T and $O^6$-CMG:C are acceptable in DNA polymerase, these pairs were incorporated into the active sites of human DNA polymerase η in complex with DNA (41), using the X-ray structure of PDB-ID = 4ED8. The plausible models were energetically refined by the computer program *REFMAC5* using the X-ray intensity data of 4ED8 that were downloaded from

PDB and truncated at low resolution (5 Å). As shown in Figure 6 and Supplementary Data (see the next page), the *in silico* structural models suggest that the Watson–Crick–type pair of $O^6$-CMG:T can be accommodated in the template site, consistent with this damaged base being able to induce pyrimidine transition mutations. The reversed wobble pair of $O^6$-CMG:C can also potentially be incorporated into the active site but would require a slight rotation of the base pair.

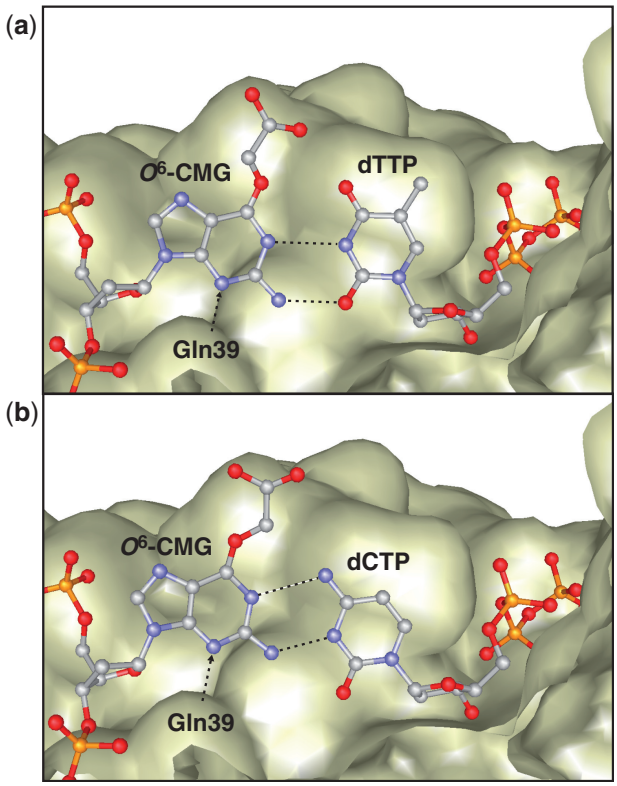


**Figure 6.** *In silico* structural models of human DNA polymerase η (41) in complex with B-DNA containing $O^6$-CMG paired with dTTP (**a**) and dCTP (**b**). In the minor groove, the hydrophilic Gln39 forms a hydrogen bond with the template bases. In addition, hydrophobic amino-acid residues (Phe18, Ile48, Leu89, Tyr92 and Ile114) are packed closely to form a pocket for the paired bases so that there is no space to accept any modification of the paired bases. In the major groove side, however, there is a widely opened space for modified bases. Broken lines indicate possible hydrogen bonds. The viewing directions are slightly different between (a) and (b).

DNA replication relies on cognate Watson–Crick–type base pair formation in the active site of a DNA polymerase (13–15). Typically, there is not enough space for a wobble type or other non-complementary base pairs. In addition, as the polymerase is bound in the minor groove of DNA, extrusion of the carboxymethyl group into the major groove is unlikely to interfere with binding to the DNA polymerase or with nucleotide incorporation opposite the damaged base. Taking the Watson–Crick–type $O^6$-CMG:T and the reversed wobbling-type $O^6$-CMG:C pairings into consideration, it is deduced that when $O^6$-CMG is in the template, it can accept a thymine and, to a much lesser extent, a cytosine residue in the newly synthesized DNA.

Based on these two cases of such mis-incorporations, three possible routes of pyrimidine transition at the modified G site could be proposed as shown in Figure 7. In the case that the template strand is damaged, the original G:C pair can be replaced with an A:T pair. In the first replication, a thymine residue is introduced in the daughter strand by accepting both dTTP and dCTP, and then the synthesized strand is used as a second template. In the second replication, dATP is bound against the mutated thymine residue. After two steps of replication, a pyrimidine transition mutation can be achieved.

Alternatively, the triphosphate that would be formed by carboxymethylation of the nucleotide triphosphate pool d[$O^6$-CMG]TP might compete with dATP incorporation opposite thymine in a DNA template. Once d[$O^6$-CMG]TP could pair with a T residue in a template strand, leading to the insertion of A in the opposite strand, not only C but also T will be introduced opposite the template $O^6$-CMG residue in the second replication. At the third replication, the incorporated C residue directs the insertion of G in the opposite strand. After three cycles of DNA replication at least, the pyrimidine transition mutation will be completed. Another case is when the

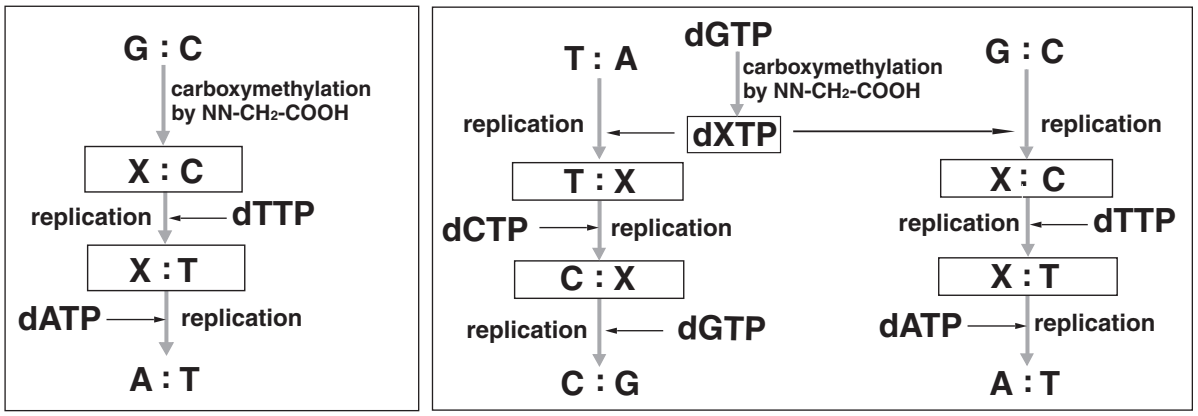


**Figure 7.** Three possible schemes of pyrimidine transition mutations. A template guanine base is $O^6$-carboxymethylated (X = $O^6$-CMG) in the left box. In the right box, dGTP is $O^6$-carboxymethylated (dXTP) to be incorporated opposite T and C residues, respectively.

d[$O^6$-CMG]TP residue is initially paired opposite C in a template strand and then the introduced $O^6$-CMG residue accepts a T residue. In the third cycle, an A is inserted opposite the T residue. Through the three cycles, the original G:C pair is converted to an A:T pair.

## CONCLUSION

In this study, we have determined the crystal structures of two $O^6$-CMG-containing DNA duplexes. The carboxymethylated guanine base can form a Watson–Crick–type pair with T (in the $O^6$-CMG5T crystal) and a reversed wobble pair with C (in the $O^6$-CMG4C crystal). *In silico* structural modeling suggests that both the Watson–Crick–type $O^6$-CMG:T and the reversed wobble-type $O^6$-CMG:C pairing modes, found in the present study, could be accepted by the DNA polymerase. In other words, $O^6$-CMG residues in a damaged DNA template would direct the incorporation not only of the complementary dCTP but also of the non-complementary dTTP into the newly synthesized DNA strand. Finally, we conclude that the G:C→A:T transition mutations, demonstrated by *in vivo* and *in vitro* experiments (3,12) as a factor in the etiology of gastrointestinal cancer, likely occur as a consequence of the Watson–Crick–type pairing of $O^6$-CMG with T.

## ACCESSION NUMBERS

4ITD and 4IJ0.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–2.

## ACKNOWLEDGEMENTS

The authors thank Y. Yamada, N. Matsugaki, N. Igarashi and S. Wakatsuki (Photon Factory, Tsukuba, Japan) for their assistance during data collection at the synchrotron facility. Figures 2, 4 and 6 are depicted with the programs *Coot-Dino* (27), *RASMOL* (42) and MolFeat (FiatLux Corporation, Tokyo), respectively. Figures 3 and 5 are produced with *PyMOL* (43).

## FUNDING

Biotechnology and Biological Sciences Research Council, UK (to O.J.W., D.M.W.) (in part); Cancer Research-UK (to G.P.M.) (in part); Engineering and Physical Sciences Research Council, UK (to C.L.M., D.M.W.) (in part).

*Conflict of interest statement*. None declared.

## REFERENCES

1. Margison,G.P., Santibanez-Koref,M.F. and Povey,A.C. (2002) Mechanisms of carcinogenicity /chemotherapy by $O^6$-methylguanine. *Mutagenesis*, **17**, 483–487.
2. Povey,A.C., Hall,C.N., Badawi,A.F., Cooper,D.P. and O'Connor,P.J. (2000) Elevated levels of the pro-carcinogenic adduct, O(6)-methylguanine, in normal DNA from the cancer prone regions of the large bowel. *Gut*, **47**, 362–365.
3. Lewin,M.H., Bailey,N., Bandaletova,T., Bowman,R., Cross,A.J., Pollock,J., Shuker,D.E. and Bingham,S.A. (2006) Red meat enhances the colonic formation of the DNA adduct $O^6$-carboxymethyl guanine: implications for colorectal cancer risk. *Cancer Res.*, **66**, 1859–1865.
4. Gladwin,M.T. (2004) Haldane, hot dogs, halitosis, and hypoxic vasodilation: the emerging biology of the nitrite anion. *J. Clin. Invest.*, **113**, 19–21.
5. Gisone,P., Dubner,D., Pérez,M.R., Michelin,S. and Puntarulo,S. (2004) The role of nitric oxide in the radiation-induced effects in the developing brain. *In Vivo*, **18**, 281–292.
6. Shephard,S.E. and Lutz,W.K. (1989) Nitrosation of dietary precursors. *Cancer Surv.*, **8**, 401–421.
7. Garcia-Santos Mdel,P., Calle,E. and Casado,J. (2001) Amino acid nitrosation products as alkylating agents. *J. Am. Chem. Soc.*, **123**, 7506–7510.
8. Shuker,D.E. and Margison,G.P. (1997) Nitrosated glycine derivatives as a potential source of $O^6$-methylguanine in DNA. *Cancer Res.*, **57**, 366–369.
9. Daniels,D.S., Woo,T.T., Luu,K.X., Noll,D.M., Clarke,N.D., Pegg,A.E. and Tainer,J.A. (2004) DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol.*, **11**, 714–720.
10. Senthong,P., Millington,C.L., Wilkinson,O.J., Marriott,A.S., Watson,A.J., Reamtong,O., Eyers,C.E., Williams,D.M., Margison,G.P. and Povey,A.C. (2013) The nitrosated bile acid DNA lesion, $O^6$-carboxymethylguanine, is a substrate for the human DNA repair protein, $O^6$-methylguanine DNA methyltransferase. *Nucleic Acids Res.*, **41**, 3047–3055.
11. Gottschalg,E., Scott,G.B., Burns,P.A. and Shuker,D.E. (2007) Potassium diazoacetate-induced p53 mutations in vitro in relation to formation of $O^6$-carboxymethyl- and $O^6$-methyl-2-deoxyguanosine DNA adducts: relevance for gastrointestinal cancer. *Carcinogenesis*, **28**, 356–362.
12. Kiefer,J.R., Mao,C., Braman,J.C. and Beese,L.S. (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature*, **391**, 304–307.
13. Harris,V.H., Smith,C.L., Cummins,W.J., Hamilton,A.L., Adams,H., Dickman,M., Hornby,D.P. and Williams,D.M. (2003) The Effect of tautomeric constant on the specificity of nucleotide incorporation during DNA replication: support for the rare tautomer hypothesis of substitution mutagenesis. *J. Mol. Biol.*, **326**, 1389–1401.
14. Wang,W., Hellinga,H.W. and Beese,L.S. (2011) Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc. Natl Acad. Sci. USA*, **108**, 17644–17648.
15. Dickerson,R.E. and Drew,H.R. (1981) Structure of a B-DNA dodecamer. Influence of base sequence on helix structure. *J. Mol. Biol.*, **149**, 761–786.
16. Millington,C.L., Watson,A.J., Marriott,A.S., Margison,G.P., Povey,A.C. and Williams,D.M. (2012) Convenient and efficient syntheses of oligodeoxyribonucleotides containing $O^6$-(carboxymethyl)guanine and $O^6$-(4-oxo-4-(3-pyridyl)butyl)guanine. *Nucleosides Nucleotides Nucleic Acids*, **31**, 328–338.
17. Berger,I., Kang,C., Sinha,H., Wolters,M. and Rich,A. (1996) A highly efficient 24-condition matrix for the crystallization of nucleic acid fragment. *Acta Crystallogr. D Biol. Crystallogr.*, **52**, 465–468.
18. Otwinowsky,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
19. Battye,T.G., Kontogiannis,L., Johnson,O., Powell,H.R. and Leslie,A.G. (2011) *iMOSFLM*: a new graphical interface for diffraction-image processing with *MOSFLM*. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 271–281.
20. Evans,P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 72–82.
21. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G., McCoy,A.

*et al*. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr*., **67**, 235–242.
22. Shui,X., McFail-Ison,L., Hu,G.G. and Williams,L.D. (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry*, **37**, 8341–8355.
23. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likehood method. *Acta Crystallogr. D Biol. Crystallogr*., **53**, 240–255.
24. Brünger,A.T., Adams,P.D., Clore,G.M., Delano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al*. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr*., **54**, 905–921.
25. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr*., **60**, 2126–2132.
26. Brünger,A.T. (1992) The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
27. Biasini,M., Mariani,V., Haas,J., Scheuber,S., Schenk,A.D., Schwede,T. and Philippsen,A. (2010) Open structure: a flexible software framework for computational structural biology. *Bioinformatics*, **26**, 2626–2628.
28. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res*., **31**, 5108–5121.
29. Olson,W.K., Banasal,M., Burley,S.K., Dickerson,R.E., Gerstein,M., Harvey,S.C., Heinemann,U., Lu,X.J., Neidle,S., Shakked,Z. *et al*. (2001) A standard reference frame for the despription of nucleic acid base-pair geometry. *J. Mol. Biol*., **313**, 229–237.
30. Wiederholt,K., Rajur,S.B. and McLaughlin,L.W. (1997) Oligonucleotides tethering Hoechst33258 derivatives: effect of the conjugation site on duplex stabilization and fluorescence properties. *Bioconjug. Chem*., **8**, 119–126.
31. Vlieghe,D., Turkenburg,J.P. and Van Meervelt,L. (1999) DNA decamer d(GGCCAATTGG) at atomic resolution (1. 15 A). *Acta Crystallogr. D Biol. Crystallogr*., **55**, 1495–1502.
32. Juan,E.C., Shimizu,S., Ma,X., Kurose,T., Haraguchi1,T., Zhang,F., Tsunoda,M., Ohkubo,A., Sekine,M., Shibata,T. *et al*. (2010) Insights into the DNA stabilizing contributions of a bicyclic cytosine analogue: crystal structures of DNA duplexes containing 7,8-dihydropyrido[2,3-*d*]pyrimidin-2-on. *Nucleic Acids Res*., **38**, 6737–6745.
33. Teng,M.K., Usman,N., Frederick,C.A. and Wang,A.H. (1988) The molecular structure of the complex of Hoechst 33258 and the DNA dodecamer d(CGCGAATTCGCG). *Nucleic Acids Res*., **16**, 2671–2690.
34. Robinson,H., Gao,Y.G., Bauer,C., Roberts,C., Switzer,C. and Wang,A.H. (1998) 2′-Deoxyisoguanosine adopts more than one tautomer to form base pairs with thymidine observed by high-resolution crystal structure analysis. *Biochemistry*, **37**, 10897–10905.
35. Leonard,G.A., Thomson,J., Watson,W.P. and Brown,T. (1990) High-resolution structure of a mutagenic lesion in DNA. *Proc. Natl Acad. Sci. USA*, **87**, 9573–9576.
36. Crick,F. (1966) Codon-anticodon pairing: the wobble hypothesis. *J. Mol. Biol*., **19**, 548–555.
37. Varani,G. and McClain,W. (2000) The G×U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep*., **1**, 18–23.
38. Tsunoda,M., Sakaue,T., Naito,S., Sunami,T., Abe,N., Ueno,Y., Matsuda,A. and Takénaka,A. (2010) Insights into the mutagenic structures of DNA damaged by hydroxyl radical, XI: crystal structures of DNA duplexes containing 5-formyluracil. *J. Nucleic Acids*, **2010**, Article ID 107289, 10 pages; doi:10.4061/2010/107289.
39. Sriram,M., van der Marel,G.A., Roelen,H.L., van Boom,J.H. and Wang,A.H. (1992) Structural consequences of a carcinogenic alkylation lesion on DNA: effect of $O^6$-ethylguanine on the molecular structure of the d(CGC[e$^6$G]AATTCGCG)-netropsin complex. *Biochemisrty*, **31**, 11823–11834.
40. Sriram,M., van der Marel,G.A., Roelen,H.L., van Boom,J.H. and Wang,A.H. (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$^6$G]AATTCGCG complexed with Hoechst 33258 or Hoechst 33342. *EMBO J*., **11**, 225–232.
41. Nakamura,T., Zhao,Y., Yamagata,Y., Hua,Y.J. and Yang,W. (2012) Watching DNA polymerase η make a phosphodiester bond. *Nature*, **487**, 196–201.
42. Style,R.A. and Milner-White,E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci*., **20**, 374–376.
43. Delano,W.L. (2008) *The PyMOL Molecular Graphics System*. DeLano Scientific LLC, Palo Alto, CA.