Structures of DNA duplexes containing O\textsuperscript{6}-carboxymethylguanine, a lesion associated with gastrointestinal cancer, reveal a mechanism for inducing pyrimidine transition mutations

Fang Zhang\textsuperscript{1}, Masaru Tsunoda\textsuperscript{2}, Kaoru Suzuki\textsuperscript{2}, Yuji Kikuchi\textsuperscript{1,2}, Oliver Wilkinson\textsuperscript{3}, Christopher L. Millington\textsuperscript{3}, Geoffrey P. Margison\textsuperscript{4}, David M. Williams\textsuperscript{3,*}, Ella Czarina Morishita\textsuperscript{5} and Akio Take\textsuperscript{na\textsuperscript{1,2,5,*}}

\textsuperscript{1}Graduate School of Science and Engineering, Iwaki-Meisei University, Iwaki 970-8551, Japan, \textsuperscript{2}Faculty of Pharmacy, Iwaki-Meisei University, Iwaki 970-8551, Japan, \textsuperscript{3}Department of Chemistry, Centre for Chemical Biology, Krebs Institute, University of Sheffield, Sheffield S3 7HF, UK, \textsuperscript{4}Cancer Research-UK Carcinogenesis Group, Paterson Institute for Cancer Research, University of Manchester, Manchester M20 9BX, UK and \textsuperscript{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\textsuperscript{6}-carboxymethylguanine (O\textsuperscript{6}-CMG) in DNA. O\textsuperscript{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\textsuperscript{6}-methylguanine, O\textsuperscript{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\textsuperscript{6}-CMG]ATTCGCG) and d(CGC[O\textsuperscript{6}-CMG]ATTCGCG) 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\textsuperscript{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\textsuperscript{6}-CMG with thymine, but might also accept the reversed wobble pair of O\textsuperscript{6}-CMG with cytosine.

Thus, O\textsuperscript{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\textsuperscript{6}-CMG]TP might compete with dATP incorporation opposite thymine in a DNA template.

INTRODUCTION

Diet 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\textsuperscript{6}-methylguanine (O\textsuperscript{6}-MeG) and O\textsuperscript{6}-carboxymethylguanine (O\textsuperscript{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\textsubscript{2}O\textsubscript{3}), which in turn is generated by the oxidation of NO (4), from dietary nitrate or after exposure to ionizing radiation (5).
N-Nitrosoglycine is converted into diazoacetate or \( \alpha \)-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).

\( O^6 \)-CMG predominantly induces \( G:C \rightarrow A:T \) transition mutations (3,11), implying that \( O^6 \)-CMG repairs DNA in 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). 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][ATTCCGCG]) and d(CGCG[\( O^6 \)-CMG][ATTTCGCG]) were synthesized and purified by reverse-phase HPLC according to methods previously described (16), and they were characterized by ESI–mass spectrometry. For crystallization, the samples in pure water were purified on an AKTAprime 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 \( \alpha \)-lactone buffer solution (50 mM NaH₂PO₄ 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 \( \mu \)l of the reservoir solution. The optimized conditions for obtaining \( O^6 \)-CMG5T and \( O^6 \)-CMG4C crystals were as follows. For \( O^6 \)-CMG5T, 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′-bi-\( \alpha \)-phenanthridine) 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 (\( \lambda = 1.00 \)\( \AA \) at BL-5A and 0.98\( \AA \) at BL-17A) from 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)-Scal (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...
observed and calculated structure factor amplitudes, respectively. The intensity of reflection \( I(h) \) from the crystal is calculated from the structure factor amplitudes using the following equation:

\[
I(h) = h \cdot F(h) - h_0 \cdot F(h)
\]

where \( F(h) \) is the structure factor amplitude, \( h_0 \) is the background intensity, and \( h \) is the index of the reflection. The observed reflections were collected in 180 frames of XST and X4C.

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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 electron density maps of the modified nucleotides and their partner in the pair formation, together with the corresponding omit maps. These maps were depicted by the program \( 3 \)DNA (28). The 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 corresponding distances are 3.4 and 3.0 Å, respectively. These values suggest base pair formation between O₆-CMG and T. Thus, although the natural base G forms a wobble base pair with T (34), the O₆-CMG:T pair is similar to the canonical Watson–Crick A:T pair. In this O₆-CMG:T pair, although the electronegative O₆ atom of O₆-CMG is exposed to the O₄ 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 $N^1(O₆-CMG)\ldots N^3(T)$ distances look slightly longer, perhaps because of the $O^6(O₆-CMG)\ldots O^4(T)$ repulsion, which separates the $O^6$ and $O^4$ atoms at an average distance of 3.6 Å. The interaction geometries of O₆-CMG:T pairs are similar to those of O₆-MeG:T pairs found in the O₆-MeG-containing B-DNA (35). As the carboxymethyl groups of the O₆-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 O₆-CMG4:C21 pair, the interatomic distances between the $N^1$ atom of O₆-CMG and the $N^4$ atom of C

---

**Figure 2.** $F_o-F_c$ omit maps (in the left column) of O₆-CMG5:T20 pairs and T8:O₆-CMG17 pairs (a), and 2$F_o-F_c$ maps (in the right column) of O₆-CMG5:T20 pairs and T8:O₆-CMG17 pairs (b). The corresponding maps of O₆-CMG4:C21 pairs and C9:O₆-CMG16 pairs (c) and those of O₆-CMG4:C21 pairs and C9:O₆-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 2$F_o-F_c$ maps are contoured at the 0.8, 0.8, 1.0 and 1.0 σ levels in the right column.
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 $C^i \ldots C^j$ distance longer by 1.0 Å and the $\lambda_1$ 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, 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 $C^i \ldots C^j$ distance longer by 1.0 Å and the $\lambda_1$ 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, 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 $C^i \ldots C^j$ distance longer by 1.0 Å and the $\lambda_1$ 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, 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 $C^i \ldots C^j$ distance longer by 1.0 Å and the $\lambda_1$ 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, 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 $C^i \ldots C^j$ distance longer by 1.0 Å and the $\lambda_1$ and $\lambda_{II}$ angles asymmetric (Table 2), as compared with those of the unmodified pairs.
phosphate backbone, respectively, so that they do not influence the base pairs formed by O\textsuperscript{6}-CMG. The same preference is also seen for the modified base in the O\textsuperscript{6}-CMG:C base pair. The change in preference for the orientation of the alkyl group in O\textsuperscript{6}-CMG compared with O\textsuperscript{6}-MeG may result from additional interactions of the carboxyl group of the alkyl side chain with the O\textsuperscript{4} and N\textsuperscript{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\textsuperscript{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\textsuperscript{6}-CMG:T and O\textsuperscript{6}-CMG:C are acceptable in DNA polymerase, these pairs were incorporated into the active sites of human DNA polymerase \(\eta\) 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

---

**Figure 4.** Watson–Crick–type pairs between O\textsuperscript{6}-CMG and T (left column) and reversed wobble pairs between O\textsuperscript{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\textsuperscript{6}-CMGs are shown in the bottom. The carboxyl groups of O\textsuperscript{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\textsuperscript{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.
PDB and truncated at low resolution (5 Å). As shown in Figure 6 and Supplementary Data (see the next page), the \textit{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.

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

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6.png}
\caption{\textit{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).}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.png}
\caption{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.}
\end{figure}
$d(O^6\text{-CMG})TP$ residue is initially paired opposite C in a template strand and then the introduced $O^6\text{-CMG}$ 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 $O^6\text{-CMG}$-containing DNA duplexes. The carboxymethylated guanine base can form a Watson–Crick–type pair with T (in the $O^6\text{-CMG}5\text{GT}$ crystal) and a reversed wobble pair with C (in the $O^6\text{-CMG}4\text{C}$ crystal). In *silico* structural modeling suggests that both the Watson–Crick–type $O^6\text{-CMG}:T$ and the reversed wobble-type $O^6\text{-CMG}:C$ pairing modes, found in the present study, could be accepted by the DNA polymerase. In other words, $O^6\text{-CMG}$ 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$\rightarrow$A:T transition mutations, demonstrated by in *vivo* and in *vivo* experiments (3,12) as a factor in the etiology of gastrointestinal cancer, likely occur as a consequence of the Watson–Crick–type pairing of $O^6\text{-CMG}$ 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 $O^6\text{-methylguanine}$. *Mutagene*sis, 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., Bandaloeva,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 $O^6\text{-carboxymethyl guanine}$: 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., Perez,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 nitrosation products as alkylating agents. *J. Am. Chem. Soc.*, 123, 7506–7510.

7. Shuker,D.E. and Margison,G.P. (1997) Nitrosated glycine derivatives as a potential source of $O^6\text{-methylguanine}$ in DNA. *Cancer Res.*, 57, 366–369.

8. 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.

9. Senthong,P., Millington,C.L., Wilkinson,O.J., Marriott,A.S., Watson,A.J., Reamtiong,O., Eyers,C.E., Williams,D.M., Margison,G.P. and Povey,A.C. (2013) The nitrosated bide acid DNA lesion, $O^6\text{-carboxymethylguanine}$, is a substrate for the human DNA repair protein, $O^6\text{-methylguanine}$ DNA methyltransferase. *Nucleic Acids Res.*, 41, 3047–3055.

10. 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 $O^6\text{-carboxymethyl}$- and $O^6\text{-methyl}-2\text{-deoxyguanosine}$ DNA adducts: relevance for gastrointestinal cancer. *Carcinogenesis*, 28, 356–362.

11. Kiefer,J.R., Mao,C., Braman,J.C. and Beebe,L.S. (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature*, 391, 304–307.

12. 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.*, 328, 1389–1401.

13. Wang,W., Helllinga,H.W. and Beele,L.S. (2011) Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc. Natl Acad. Sci. USA.*, 108, 17644–17648.

14. 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.

15. 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 $O^4\text{(carboxymethylguanine)}$ and $O^4\text{(4\text{-oxo-4\text{-3\text{-pyridyl}}butyl)guanine}$. *Nucleosides Nucleotides Nucleic Acids*, 31, 328–338.

16. Berger,J., Kang,C., Sinha,H., Wolters,M. and Rich,A. (1996) A highly efficient 24-condition matrix for the crystallization of a reversed wobble pair with C (in the $6\text{-carboxymethyl guanine}$: implications for colorectal cancer risk. *Cancer Res.*, 66, 1859–1865.

17. Berger,J., Kang,C., Sinha,H., Wolters,M. and Rich,A. (1996) A highly efficient 24-condition matrix for the crystallization of a reversed wobble pair with C (in the $6\text{-carboxymethyl guanine}$: implications for colorectal cancer risk. *Cancer Res.*, 66, 1859–1865.

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) Scanning and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 72–80.
et al. (2011) Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr., 67, 235–242.
22. Shi,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-likelihood 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, 903–921.
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., Scheuer,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 description 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(CGCGAATTCCGG). 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. Varami,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 Takenaka,A. (2010) Insights into the mutagenic structures of DNA damaged by hydroxyl radical, XI: crystal structures of DNA duplexes containing 5-formyluracil. Nucleic Acids Res., 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 O6-ethylguanine on the molecular structure of the d(CG[e6G]AATTCCGG)-netropsin complex. Biochemistry, 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 O6-ethyl-G-C base pairs stabilized by minor groove binding drugs: molecular structure of d(CG[e6G]AATTCCGG) 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.