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

Fang Zhang, Masaru Tsunoda, Kaoru Suzuki, Yuji Kikuchi, Oliver Wilkinson, Christopher L. Millington, Geoffrey P. Margison, David M. Williams, Ella Czarina Morishita and Akio Takenaka

1Graduate School of Science and Engineering, Iwaki-Meisei University, Iwaki 970-8551, Japan, 2Faculty of Pharmacy, Iwaki-Meisei University, Iwaki 970-8551, Japan, 3Department of Chemistry, Centre for Chemical Biology, Krebs Institute, University of Sheffield, Sheffield S3 7HF, UK, 4Cancer Research-UK Carcinogenesis Group, Paterson Institute for Cancer Research, University of Manchester, Manchester M20 9BX, UK and 5Graduate 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).
N-Nitrosoglycine is converted into diazoacetate or α-lactone (6, 7), potent mutagens that can alkylate guanine in DNA to form O6-CMG and O6-MeG (8). In humans, O6-methylguanine-DNA methyltransferase (MGMT) repairs DNA containing a wide variety of different 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 O6-CMG is also a substrate for MGMT (10). In vivo and in vitro evidence suggests that O6-CMG predominantly induces G:C→A:T transition mutations (3, 11), implying that O6-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 interaction geometry of O6-CMG with T and C bases.

Here, we describe the crystal structures of DNA duplexes containing O6-CMG at residue positions that place the modified base opposite T or C in the palindromic B-form Dickerson–Drew sequence d(CGCGAATTCCGCG) (15).

MATERIALS AND METHODS

Oligodeoxyribonucleotide synthesis and purification

Oligodeoxyribonucleotides (ODNs) with the sequences d(CGCG[O6-CMG]AATTCCGCG) and d(CGCG[O6-CMG]ATTTCCGCG) 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 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 μl of the reservoir solution. The optimized conditions for obtaining O6-CMG5T and O6-CMG4C crystals were as follows. For O6-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)-2,6-bis-1H-benzimidazole) was equilibrated against 35% (v/v) MPD. For O6-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.

O6-CMG5T and O6-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 (λ = 1.00 Å at BL-5A 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 O6-CMG5T and O6-CMG4C crystals, respectively. A total of 180 frames of the patterns for one crystal were taken at 1 s 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 O6-CMG5T data and Mosflm(19)-Scala (20) of the CCP4 suite (21) for O6-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(CGCGAATTCCGCG) (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 O6-CMG5T and O6-CMG4C, respectively, and these
Table 1. Crystal data and statistics of data collection and structure refinement

<table>
<thead>
<tr>
<th>Crystal code</th>
<th>O&lt;sup&gt;6&lt;/sup&gt;-CMG5T</th>
<th>O&lt;sup&gt;6&lt;/sup&gt;-CMG4C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crystal data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P&lt;sub&gt;2&lt;/sub&gt;1&lt;sup&gt;2&lt;/sup&gt;1</td>
<td>P&lt;sub&gt;2&lt;/sub&gt;1&lt;sup&gt;2&lt;/sup&gt;1</td>
</tr>
<tr>
<td>Unit cell (Å)</td>
<td>25.1</td>
<td>24.6</td>
</tr>
<tr>
<td>a</td>
<td>40.4</td>
<td>41.2</td>
</tr>
<tr>
<td>b</td>
<td>65.7</td>
<td>63.1</td>
</tr>
<tr>
<td>c</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>34.4–1.94</td>
<td>25.1–1.54</td>
</tr>
<tr>
<td>Outer shell (Å)</td>
<td>1.97–1.94</td>
<td>1.62–1.54</td>
</tr>
<tr>
<td>Observed reflections</td>
<td>27 766</td>
<td>62 239</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>4877</td>
<td>9559</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>91.0 (82.9)</td>
<td>95.8 (100)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt; (%)</td>
<td>5.7 (52.2)</td>
<td>8.9 (28.5)</td>
</tr>
<tr>
<td>I/σ</td>
<td>8.1 (0.7)</td>
<td>14 (6.4)</td>
</tr>
<tr>
<td>Redundancy (%)</td>
<td>5.7 (1.4)</td>
<td>6.5 (6.7)</td>
</tr>
<tr>
<td><strong>Structure refinement</strong></td>
<td>34.4–1.94</td>
<td>25.1–1.54</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>34.4–1.94</td>
<td>25.1–1.54</td>
</tr>
<tr>
<td>Used reflections</td>
<td>3842</td>
<td>9017</td>
</tr>
<tr>
<td>R&lt;sub&gt;factor&lt;/sub&gt; (%)</td>
<td>23.9</td>
<td>19.4</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>26.7</td>
<td>23.6</td>
</tr>
<tr>
<td>Rms deviation</td>
<td>Bond distances (Å)</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Bond angles (degree)</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>No. of DNA duplex</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>No. of Hoechst33258</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>No. of ions</td>
<td>1Mg&lt;sup&gt;2+&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>No. of water molecules</td>
<td>43</td>
</tr>
</tbody>
</table>

<sup>a</sup>Values in parentheses indicate those in the outer shell.
<sup>b</sup>R<sub>merge</sub> = 100 × Σ<sub>i</sub>Σ<sub>j</sub>|<i>F<sub>i</sub>o</i> − |<i>F<sub>i</i>−j</i>|| / Σ<sub>i</sub>Σ<sub>j</sub> |<i>F<sub>i</sub>o</i>||, where <i>k</i> is the jth measurement of the intensity of reflection k and |<i>F<sub>i</sub>o</i>−j|| is its mean value.
<sup>c</sup>Diffraction patterns of 1<sup>st</sup> oscillation ranges were collected in 180 frames of XST and X4C.
<sup>d</sup>R-factor = 100 × Σ|<i>F<sub>i</i>o</i> − |<i>F<sub>i</i>|| / Σ|<i>F<sub>i</i>||, where |<i>F<sub>i</i>o</i>| and |<i>F<sub>i</i>|| are the observed and calculated structure factor amplitudes, respectively.
<sup>e</sup>Calculated using a random set containing 5% observations that were not included during refinement (26).

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<sup>6</sup>-CMG5T and O<sup>6</sup>-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<sup>6</sup>-CMG5T and 1.4 Å for O<sup>6</sup>-CMG4C. Although the largest deviations occur at the O<sup>6</sup>-CMG residues in both O<sup>6</sup>-CMG5T and O<sup>6</sup>-CMG4C duplexes, their sugar pucker states fluctuate around the C<sup>2</sup>-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<sup>6</sup>-CMG4C may be related to the interaction geometry between O<sup>6</sup>-CMG and C, which will be described later in detail. Figure 3 shows Hoechst33258 molecules bound in the minor grooves of O<sup>6</sup>-CMG5T and O<sup>6</sup>-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<sup>6</sup>-CMG5T and O<sup>6</sup>-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<sup>6</sup>-CMG:T and O<sup>6</sup>-CMG:C base pairs

All unmodified base pairs in the dodecamer sequences except those containing O<sup>6</sup>-CMG are of the standard Watson–Crick type. F<sub>T</sub>=F<sub>C</sub> and 2F<sub>T</sub>=F<sub>C</sub> electron density maps of the O<sup>6</sup>-CMG5T duplex indicate that the O<sup>6</sup>-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<sub>T</sub>=F<sub>C</sub> and 2F<sub>T</sub>=F<sub>C</sub> electron density maps of the O<sup>6</sup>-CMG4C duplex (Figure 2c and d) show that the O<sup>6</sup>-CMG residues form base pairs with the opposite C residues, the two geometries of the O<sup>6</sup>-CMG:C pairs are different from the standard G:C pair geometry found in the unmodified dodecamer.

Geometries of the O<sup>6</sup>-CMG:T and O<sup>6</sup>-CMG:C base pairs

The interaction geometries of base paired O<sup>6</sup>-CMGs are shown in Figure 4 and summarized in Table 2. In the O<sup>6</sup>-CMG5:T20 pair, the interatomic distances between the N<sup>1</sup> atom of O<sup>6</sup>-CMG and the N<sup>3</sup> atom of T and between the N<sup>2</sup> atom of O<sup>6</sup>-CMG and the O<sup>2</sup> atom of T are 3.3 and 3.1 Å, respectively. In the O<sup>6</sup>-CMG17:T8 pair,
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 $\pm 18^\circ$ 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 carboxylic 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...
and between the N² atom of O⁶-CMG and the N³ atom of C are 3.0 and 2.9 Å, respectively. In the O⁶-CMG16:C9 pair, the corresponding distances are the same 3.1 and 2.9 Å, respectively. These values indicate that O⁶-CMG can form a base pair with C through hydrogen bonds at these two sites. In both base pairs, the purine moiety of O⁶-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 O⁶-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 O⁶-ethylguanine (O⁶-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 C¹ . . . C¹ distance longer by 1.0 Å and the λ₁ and λ₁₁ angles asymmetric (Table 2), as compared with those of the unmodified pairs.

The carboxyl groups of O⁶-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 N¹ atom of O⁶-CMG. In contrast, the alkyl groups of O⁶-EtG:C pairs (39,40) adopts an anti conformation and protrudes into the major groove. The longer carboxymethyl groups of O⁶-CMG protrude into the major groove to make contacts with water molecules. The reversed wobble O⁶-CMG:C pairs seem to be further stabilized by two additional water-mediated hydrogen bonds between an oxygen atom of the carboxyl group of O⁶-CMG and the N³ atom of C and between the N² atom of O⁶-CMG and the O² atom of C. Three Sr atoms are all hydrated with eight water molecules and bound to the minor groove, major groove and

---

**Table 2. Average local helical parameters and base pair parameters**

<table>
<thead>
<tr>
<th></th>
<th>X-Displacement (Å)</th>
<th>Inclination (°)</th>
<th>Helical twist (°)</th>
<th>Helical rise (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O⁶-CMG5T</td>
<td>0.57</td>
<td>−0.26</td>
<td>36</td>
<td>3.3</td>
</tr>
<tr>
<td>O⁶-CMG4C</td>
<td>0.67</td>
<td>0.73</td>
<td>36</td>
<td>3.3</td>
</tr>
<tr>
<td>B-DNA</td>
<td>0.05</td>
<td>2.1</td>
<td>36.5</td>
<td>3.3</td>
</tr>
<tr>
<td>A-DNA</td>
<td>−4.2</td>
<td>15</td>
<td>32.5</td>
<td>2.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Base pair</th>
<th>Propeller twisting (°)</th>
<th>Opening angle (°)</th>
<th>Buckle angle (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O⁶-CMG :T</td>
<td>−18</td>
<td>8.6</td>
<td>13</td>
</tr>
<tr>
<td>O⁶-CMG :C</td>
<td>−9.4</td>
<td>3.1</td>
<td>19</td>
</tr>
<tr>
<td>Wobble pair</td>
<td>C¹ . . . C¹ distance (Å)</td>
<td>λ₁ angle (°)</td>
<td>λ₁₁ angle (°)</td>
</tr>
<tr>
<td>O²-CMG4:C21</td>
<td>11</td>
<td>69</td>
<td>42</td>
</tr>
<tr>
<td>C9:O⁶-CMG16</td>
<td>11</td>
<td>44</td>
<td>69</td>
</tr>
</tbody>
</table>

---

*Calculated with the program 3DNA (29)

*High-resolution A- and B-form DNA structures by Olson et al. (30)

---

**Figure 3.** Overviews of O⁶-CMG5T duplex (a) and O⁶-CMG4C duplex (b) and their superimposition onto the unmodified duplex (black) (c). In (a) and (b), the O⁶-CMG residues and Hoechst33258 molecules are colored in red and green, respectively. In (c), O⁶-CMG5T, O⁶-CMG4C and the unmodified duplexes are colored in red, green and black, respectively, and the modified residues are in blue.
phosphate backbone, respectively, so that they do not influence the base pairs formed by O6-CMG. The same preference is also seen for the modified base in the O6-CMG:C base pair. The change in preference for the orientation of the alkyl group in O6-CMG compared with O6-MeG may result from additional interactions of the carboxyl group of the alkyl side chain with the O4 and N4 atoms of T and C, respectively, that are implied from the crystal structures.

**Biological implications**

From the present study, it has been found that O6-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 O6-CMG:T and O6-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 O6-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 O6-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 O6-CMG:T and the reversed wobbling-type O6-CMG:C pairings into consideration, it is deduced that when O6-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[O6-CMG]TP might compete with dATP incorporation opposite thymine in a DNA template. Once d[O6-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 O6-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 6.** In silico structural models of human DNA polymerase η (41) in complex with B-DNA containing O6-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).

**Figure 7.** Three possible schemes of pyrimidine transition mutations. A template guanine base is O6-carboxymethylated (X = O6-CMG) in the left box. In the right box, dGTP is O6-carboxymethylated (dXTP) to be incorporated opposite T and C residues, respectively.
d[O6-CMG]TP residue is initially paired opposite C in a template strand and then the introduced O6-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 O6-CMG-containing DNA duplexes. The carboxymethylated guanine base can form a Watson–Crick–type pair with T (in the O6-CMG5T crystal) and a reversed wobble pair with C (in the O6-CMG4C crystal). In silico structural modeling suggests that both the Watson–Crick–type O6-CMG:T and the reversed wobble-type O6-CMG:C pairing modes, found in the present study, could be accepted by the DNA polymerase. In other words, O6-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, is likely to occur as a consequence of the Watson–Crick–type pairing of O6-CMG with T.

ACCESSION NUMBERS
4ITD and 4IJO.

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 RASMOL (42) and MolFeat (FiatLux Corporation, Tokyo), respectively. Figures 3 and 5 are produced with PyMOL (43).

FUNDING

Conflict of interest statement. None declared.

REFERENCES
34. Robinson, H., Gao, Y.G., Baurer, 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.