Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

O^6-Carboxymethylguanine in DNA forms a sequence context-dependent wobble base-pair structure with thymine

Fang Zhang,[a] Masaru Tsunoda,[b] Yuji Kikuchi,[a,b] Oliver Wilkinson,[c]‡ Christopher L. Millington,[c]§ Geoffrey P. Margison,[d] David M. Williams[c]* and Akio Takénaka[b,e]*

[a]Graduate School of Science and Engineering, Iwaki-Meisei University, Iwaki 970-8551, Japan, [b]Faculty of Pharmacy, Iwaki-Meisei University, Iwaki 970-8551, Japan, [c]Centre for Chemical Biology, Department of Chemistry, Krebs Institute, University of Sheffield, Sheffield S3 7HF, England, [d]Centre for Occupational and Environmental Health, Faculty of Medical and Human Sciences, University of Manchester, Manchester M13 9PL, England, and [e]Research Institute, Chiba Institute of Technology, Tsudanuma, Narashino 275-0016, Japan

‡ Current address: DNA–Protein Interaction Unit, School of Biochemistry, University of Bristol, Bristol BS8 1TD, England.
§ Current address: Institut de Génétique et Développement de Rennes, Université de Rennes, 35043 Rennes CEDEX, France.

Correspondence e-mail: d.m.williams@sheffield.ac.uk, atakenak@sakura.email.ne.jp

Received 24 January 2014
Accepted 19 March 2014

PDB references:
O^6-CMG4T-1, 4o5w;
O^6-CMG4T-2, 4o5x;
O^6-CMG4T-3, 4o5y;
O^6-CMG4T-4, 4o5z

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, implying an association with the induction of colorectal cancer. By analogy to O^6-methylguanine, O^6-CMG is expected to be mutagenic, inducing G-to-A mutations that may be the molecular basis of increased cancer risk. Previously, the crystal structure of the DNA dodecamer d(CGCG[O^6-CMG]ATTCGCG) has been reported, in which O^6-CMG forms a Watson–Crick-type pair with thymine similar to the canonical A:T pair. In order to further investigate the versatility of O^6-CMG in base-pair formation, the structure of the DNA dodecamer d(CGC[O^6-CMG]AATTTGCG) containing O^6-CMG at a different position has been determined by X-ray crystallography using four crystal forms obtained under conditions containing different solvent ions (Sr^{2+}, Ba^{2+}, Mg^{2+}, K^+ or Na^+) with and without Hoechst 33258. The most striking finding is that the pairing modes of O^6-CMG with T are quite different from those previously reported. In the present dodecamer, the T bases are displaced (wobbled) into the major groove to form a hydrogen bond between the thymine N^3 N—H and the carboxyl group of O^6-CMG. In addition, a water molecule is bridged through two hydrogen bonds between the thymine O^2 atom and the 2-amino group of O^6-CMG to stabilize the pairing. These interaction modes commonly occur in the four crystal forms, regardless of the differences in crystallization conditions. The previous and the present results show that O^6-CMG can form a base pair with T in two alternative modes: the Watson–Crick type and a high-wobble type, the nature of which may depend on the DNA-sequence context.

1. Introduction

Diets high in red and processed meats are known to be risk factors for colorectal cancer (CRC), which is one of the most common cancers worldwide. The promutagenic DNA lesions O^6-methylguanine (O^6-MeG) and O^6-carboxymethylguanine (O^6-CMG[1]) (Fig. 1*a*) are commonly found in human colorectal DNA (Margison *et al.*, 2002; Povey *et al.*, 2000; Lewin *et al.*,

[1] Abbreviations: O^6-CMG refers to the O^6-carboxymethylguanine residue; O^6-CMG*n* and T*n* indicate O^6-CMG and T residues at the *n*th position, respectively; O^6-CMG4T refers to an oligodeoxyribonucleotide with the fourth and ninth residues replaced with O^6-CMG and T, respectively, in the Dickerson–Drew sequence; O^6-CMG4T-*n* refer to the four crystals (*n* = 1, 2, 3, 4) obtained under different conditions; O^6(O^6-CMG) and N^3(T9) indicate the O^6 atom of the O^6-CMG residue and the N^3 atom of the T residue at the ninth position, respectively; O^6-CMG4:T21 means base-pair formation between an O^6-CMG residue at the fourth position and a T residue at the 21st position.

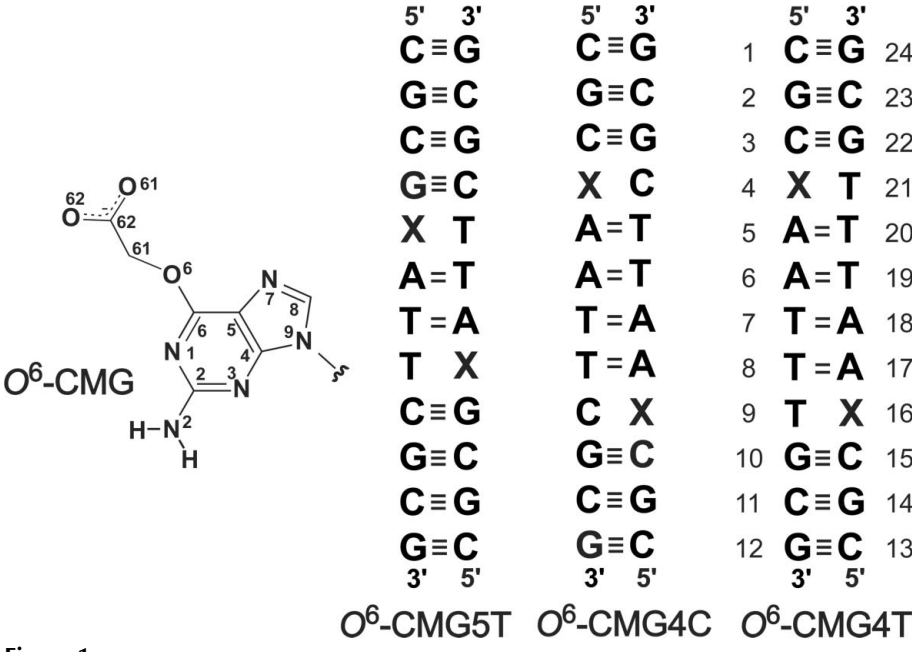

Figure 1
Chemical structure and atomic numbering of O^6-CMG (*a*) and the ODN sequences (*b*) considered in the present report. X indicates the O^6-CMG residue. The numbering system of the residues is shown for O^6-CMG4T.

2006). O^6-Alkylguanine lesions can be processed by the DNA-repair protein O^6-methylguanine-DNA methyltransferase (MGMT) by transferring the alkyl group to the thiolate side chain of the active-site Cys (Daniels *et al.*, 2004). Recently, we have shown for the first time that DNA containing O^6-CMG is also a substrate for MGMT (Senthong *et al.*, 2013). Previous reports have shown that methylation of the promoter of the MGMT gene, which would be expected to result in down-regulation of its expression, is another risk factor in CRC (Esteller *et al.*, 2000). Together, these observations implicate O^6-MeG and O^6-CMG as causative lesions in human colorectal carcinogenesis (Margison *et al.*, 2002; Povey *et al.*, 2000; Lewin *et al.*, 2006).

One mechanism that can lead to the formation of O^6-MeG and O^6-CMG involves the initial nitrosation of amino acids such as glycine and derivatives thereof, for example *N*-glycyl peptides and the bile acid conjugate glycocholic acid. Nitrosation derives from the reaction at neutral or alkaline pH with dinitrogen trioxide (N_2O_3), which in turn is generated by the oxidation of NO (Gladwin, 2004) from dietary nitrite or after exposure to ionizing radiation (Gisone *et al.*, 2004). *N*-Nitrosoglycine is converted into diazoacetate or α-lactone (Shephard & Lutz, 1989; García-Santos *et al.*, 2001), which are potent mutagens that can alkylate guanine in DNA to form O^6-MeG and O^6-CMG (Shuker & Margison, 1997). *In vivo* and *in vitro* evidence suggests that O^6-CMG predominantly induces G:C→A:T transition mutations (Gottschalg *et al.*, 2007), implying by analogy with O^6-MeG 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 only accept Watson–Crick-type pairs in the B-form conformation (Kiefer *et al.*, 1998; Harris *et al.*, 2003; Wang *et al.*, 2011).

To understand the mechanism of such mutations, we previously determined the crystal structures of two DNA duplexes (O^6-CMG5T and O^6-CMG4C; Fig. 1) 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) (Dickerson & Drew, 1981). This revealed that O^6-CMG forms a Watson–Crick-type pair with T similar to the canonical A:T pair and forms a reversed-wobble pair with C (Zhang *et al.*, 2013). These interaction geometries are those expected from the chemical structure of O^6-CMG. Based on these observations, we proposed possible mechanisms of G:C→A:T transition mutations when such a modified DNA is used as a template strand for replication. In the present study, to confirm and extend additional possibilities for base pairs involving O^6-CMG, we have performed X-ray analyses of dodecamers containing O^6-CMG at different positions and found an alternative base-pairing mode with T. We describe this site-specific interaction geometry of O^6-CMG with T and consider its significance and biological implications.

2. Materials and methods

2.1. Oligodeoxyribonucleotide synthesis and purification

An oligodeoxyribonucleotide (ODN) with the sequence d(CGC[O^6-CMG]AATTTGCG) (O^6-CMG4T) was synthesized and purified by the method described previously (Millington *et al.*, 2012) and characterized by ESI-mass spectrometry. For crystallization, the sample of ODN in pure water was further purified on an ÄKTAprime plus (GE Healthcare) using a Superdex 30 pg 16/60 column at a flow rate of 0.5 ml min^{-1} with a gradient of 0–100% buffer solution at pH 7.2 (50 m*M* NaH_2PO_4, 150 m*M* NaCl); the ODN-containing fractions monitored using a UV monitor were confirmed by PAGE analysis with TBE. Finally, the ODN was desalted by a series of C18 (Waters), AG50W-X8 (Bio-Rad) and Chelex 100 (Bio-Rad) columns in turn. The eluted solution was dried *in vacuo* at room temperature to store the samples.

2.2. Crystallization and data collection

Screenings of crystallization conditions were performed at 277 K by the hanging-drop vapour-diffusion method using a kit for nucleic acids as described by Berger *et al.* (1996). 2 µl droplets containing the sample and screening reagents were equilibrated against 700 µl reservoir solution containing

35%(*v*/*v*) 2-methyl-2,4-pentanediol (MPD). Table 1 shows the initial conditions of the crystallization drops for the four different crystals obtained. The two crystal forms O^6-CMG4T-1 and O^6-CMG4T-2 appeared using conditions containing 0.5 m*M* Hoechst 33258 [2′-(4-hydroxyphenyl)-6-(4-methyl-1-piperazinyl)-2,6′-bi-1*H*-benzimidazole] that differed slightly in pH and in the presence of ions (Li^+ and Sr^{2+} or Na^+ and K^+). Crystals of two additional forms, O^6-CMG4T-3 and O^6-CMG4T-4, were obtained without the assistance of Hoechst 33258 and the mother liquors differed in pH and in the additive ion (K^+ or Na^+).

Crystals suitable for X-ray data collection 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 the Photon Factory, Tsukuba, Japan. Diffraction intensities from the O^6-CMG4T-1, O^6-CMG4T-2, O^6-CMG4T-3 and O^6-CMG4T-4 crystals were recorded on two ADSC CCD detectors: a Quantum 210r positioned at 117.1 and 142.1 mm for the former two crystals on beamline AR-NW12A and a Quantum 315r positioned at 166.4 mm for the latter two crystals on BL-5A. A total of 180 frames of the patterns from a crystal were taken at 1° oscillation steps with 10, 10, 2 and 5 s exposure times per frame for the four respective crystals. Raw diffraction images were indexed, and intensities around Bragg spots were integrated using *HKL*-2000 (Otwinowski & Minor, 1997). 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. Although some diffraction spots were observed above the significant level of $I/\sigma(I)$ in the outer resolution shells, it was difficult to integrate their intensities because their spot shapes were deformed owing to shock on cooling the crystal specimens.

Table 1
Crystallization conditions, crystal data and statistics of data collection and structure refinement.

Values in parentheses are for the outer shell.

Crystal	O^6-CMG4T-1	O^6-CMG4T-2	O^6-CMG4T-3	O^6-CMG4T-4
PDB code	4o5w	4o5x	4o5y	4o5z
Crystallization condition				
Sodium cacodylate (m*M*)	20	20	20	20
pH	7.0	6.0	6.0	7.0
ODN (m*M*)	0.5	0.5	0.5	0.5
MPD† [%(*v*/*v*)]	5	5	5	5
Spermine.4HCl (m*M*)	6	6	6	6
LiCl (m*M*)	20	—	—	—
NaCl (m*M*)	—	40	—	40
KCl (m*M*)	—	6	40	—
$MgCl_2$ (m*M*)	10	10	—	—
$SrCl_2$ (m*M*)	40	—	—	—
$BaCl_2$ (m*M*)	—	—	10	10
Hoechst 33258† (m*M*)	0.5	0.5	—	—
CHAPS† [%(*w*/*w*)]	0.1	—	—	—
Crystal data				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)				
a	25.9	25.9	25.9	25.9
b	40.3	40.4	41.5	41.8
c	65.3	65.4	64.7	64.6
Data collection				
Resolution range (Å)	50.0–1.60 (1.63–1.60)	50.0–1.60 (1.63–1.60)	50.0–1.75 (1.78–1.75)	50.0–1.75 (1.78–1.75)
Observed reflections	91890	130788	85101	68560
Unique reflections	9517	9637	7543	7529
Completeness (%)	96.9 (90.2)	98.6 (95.1)	97.9 (94.1)	91.1 (71.7)
R_{merge}‡ (%)	4.4 (24.9)	4.7 (32.7)	8.1 (48.4)	4.4 (61.2)
$\langle I/\sigma(I)\rangle$	15.5 (7.8)	16.3 (8.4)	15.6 (8.5)	10.5 (1.8)
Multiplicity§	9.7 (7.0)	13.6 (13)	11.5 (10.1)	9.6 (2.1)
Structure refinement				
Resolution range (Å)	34.3–1.60	34.4–1.60	34.9–1.75	35.1–1.75
Reflections used	8977	9093	7022	6734
R factor¶ (%)	22.5	24.8	22.9	22.8
R_{free}†† (%)	25.5	26.9	26.9	26.3
R.m.s. deviations				
Bond distances (Å)	0.03	0.03	0.02	0.02
Bond angles (°)	3.1	3.1	2.4	2.6
No. of DNA duplexes	1	1	1	1
No. of Hoechst 33258 molecules	1	1	—	—
No. of ions	1 Sr^{2+}, 2 Mg^{2+}	1 Mg^{2+}	1 Ba^{2+}, 1 K^+	1 Ba^{2+}, 1 Na^+
No. of water molecules	109	81	138	125

† MPD, 2-methyl-2,4-pentanediol; Hoechst 33258, 2′-(4-hydroxyphenyl)-5-(4-methyl-1-piperazinyl)-2,5′-bi-1*H*-benzimidazole trihydrochloride; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]propanesulfonate. ‡ $R_{merge} = 100 \times \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th measurement of the intensity of reflection *hkl* and $\langle I(hkl)\rangle$ is its mean value. § Diffraction patterns of 1° oscillation ranges were collected as a total of 180 frames for O^6-CMG4T-1, O^6-CMG4T-2, O^6-CMG4T-3 and O^6-CMG4T-4. ¶ *R* factor $= 100 \times \sum_{hkl} \big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl} |F_{obs}|$, where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure-factor amplitudes, respectively. †† Calculated using a random set containing 5% of the observations that were not included during refinement (Brünger, 1992).

2.3. Structure determination and refinement

Using *AutoMR* in the *CCP*4 suite (Winn *et al.*, 2011), the phases of the four data sets were separately estimated by the molecular-replacement method with the unmodified ODN structure d(CGCGAATTCGCG) (Shui, McFail-Isom, Hu *et al.*, 1998) as a probe. The atomic parameters were refined using the maximum-likelihood least-squares technique in *REFMAC*5 (Murshudov *et al.*, 2011) in *CCP*4 and *CNS* (Brünger *et al.*, 1998). The crystal structures were constructed and modified by adding other molecules and ions using *Coot* (Emsley & Cowtan, 2004) in *CCP*4. The resultant structures were validated by interpretation of OMIT maps at every nucleotide residue. Cations assignable to electron densities were included in the subsequent refinements. Conformational restraints were applied to DNA and Hoechst 33258. The *R*-factor and R_{free} values converged with further rounds of structure refinement. The values are slightly high, especially for the O^6-CMG4T-2 crystal, owing to the spot-shape problem.

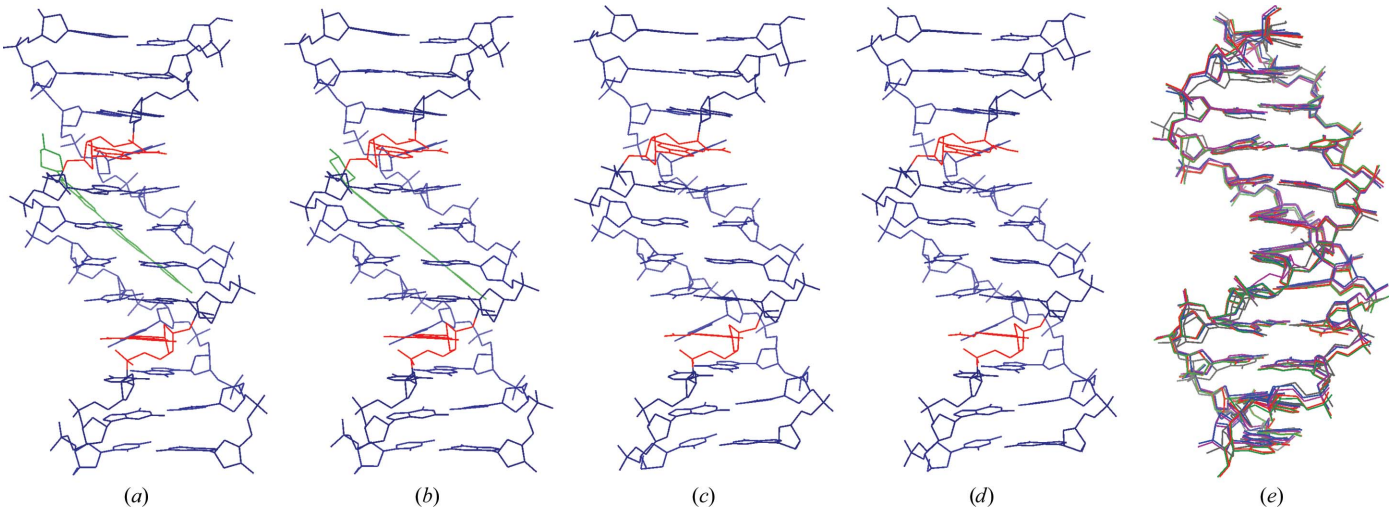

Figure 2
X-ray structures of four O^6-CMG4T duplexes found in the O^6-CMG4T-1 (*a*), O^6-CMG4T-2 (*b*), O^6-CMG4T-3 (*c*) and O^6-CMG4T-4 (*d*) crystals. DNA duplexes, O^6-CMG residues and Hoechst 33258 molecules are coloured blue, red and green, respectively. The four structures (red, green, blue and purple lines) are superimposed (*e*) on that of the unmodified duplex (grey line).

The statistics of the structure refinements are summarized in Table 1. The atomic parameters for all four crystal structures have been deposited in the PDB (PDB entries 4o5w, 4o5x, 4o5y and 4o5z). Supplementary Fig. S1[2] shows the $2F_o - F_c$ electron-density maps of the modified nucleotide and its partner in pair formation, together with the corresponding $F_o - F_c$ OMIT maps calculated without the pairs. All of the residues were traced in the electron-density maps. The O^6-CMG residues assigned in the $F_o - F_c$ OMIT maps of O^6-CMG4T-1, O^6-CMG4T-2, O^6-CMG4T-3 and O^6-CMG4T-4 fit well to the $2F_o - F_c$ maps, as shown in Supplementary Fig. S1. All of the global and local helical parameters, as well as the torsion angles and pseudo-rotation phase angles of the sugar rings, were calculated using 3*DNA* (Lu & Olson, 2003). Some of these are given in Supplementary Tables S1, S2 and S3. Figs. 3 and 4(*b*) were depicted with *RASMOL* (Sayle & Milner-White, 1995). Figs. 2, 4(*a*), 4(*c*), 5 and 7 were produced with *PyMOL* (DeLano *et al.*, 2008).

3. Results and discussion

3.1. Overall structure of DNA duplexes

Previously, we reported the structures of the ODNs d(CGCG[O^6-CMG]ATTCGCG) (O^6-CMG5T) and d(CGC[O^6-CMG]AATTCGCG) (O^6-CMG4C) (Zhang *et al.*, 2013), which required the addition of Hoechst 33258 to facilitate their crystallization. In contrast, the present ODN d(CGC[O^6-CMG]AATTTGCG) (O^6-CMG4T) was easily crystallized: four crystal forms were obtained under different conditions, two of which were obtained in the absence of Hoechst 33258. In addition, the four crystals diffracted well to 1.60 or 1.75 Å resolution. As shown in Fig. 2, the two homo-dodecamers are associated with each other to form a right-handed double helix. Their average local helical parameters (Table 2) are close to those of high-resolution B-form DNA duplexes (Shui, McFail-Isom, Hu *et al.*, 1998; Wiederholt *et al.*, 1997; Vlieghe *et al.*, 1999). Their superimpositions onto the unmodified structure (PDB entry 355d; Fig. 2*e*; Shui, McFail-Isom, Hu *et al.*, 1998), however, show local variations in the backbone conformations of the duplexes, with root-mean-square (r.m.s.) deviations between the corresponding P atoms of 0.91, 0.93, 0.77 and 0.83 Å for O^6-CMG4T-1, O^6-CMG4T-2, O^6-CMG4T-3 and O^6-CMG4T-4, respectively. Although the largest deviations occur at the O^6-CMG residues in the four duplexes, their sugar puckers fluctuate around the C$^{2'}$-*endo* conformation, except for the T9[3] and T21 residues (Table 3 and Supplementary Table S2), which is the conformation typically found in B-type DNA. This indicates that O^6-carboxymethylation of guanine residues does not significantly affect the overall DNA conformation, as found in previous DNA duplexes containing this modification (Zhang *et al.*, 2013). However, in each of the four crystal structures reported here significant r.m.s. deviations are found at the T residues that partner the O^6-CMG residues and in the complementary strand of the duplex. The detailed geometry of the O^6-CMG:T pairs is described in the next section. Fig. 2 shows Hoechst 33258 molecules bound in the minor grooves of O^6-CMG4T-1 and O^6-CMG4T-2. 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 have previously been found in other structures of DNA duplexes crystallized in the presence of this duplex-stabilizing dye (Juan *et al.*, 2010; Teng *et al.*, 1988).

[2] Supporting information has been deposited in the IUCr electronic archive (Reference: BE5256).

[3] The residue numbers are assigned from 1 to 12 in one of the two strands and from 13 to 24 for the other strand of the duplex, so that T21 indicates the ninth thymine residue of the second strand.

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

Calculated with the program *3DNA* (Lu & Olson, 2003).

(*a*) Helical parameters.

Duplex	O^6-CMG4T-1	O^6-CMG4T-2	O^6-CMG4T-3	O^6-CMG4T-4	B-DNA†	A-DNA†
x displacement (Å)	−1.1	−1.1	−1.0	−1.0	0.05	−4.2
Inclination (°)	8.3	9.0	7.5	7.6	2.1	15
Helical twist (°)	37	37	37	37	36.5	32.5
Helical rise (Å)	3.1	3.1	3.0	3.0	3.3	2.8

(*b*) Base-pair parameters.

	O^6-CMG:T‡								A:T	G:C
Duplex code	O^6-CMG4T-1		O^6-CMG4T-2		O^6-CMG4T-3		O^6-CMG4T-4		Unmodified§	
Propeller twist (°)	−21	−26	−22	−23	−28	−27	−27	−24	−16	−8.7
Opening angle (°)	32	32	30	32	32	32	32	34	4.2	−0.33
Buckle angle (°)	14	−15	14	−13	18	−15	16	−13	0.84	2.4
$C^{1'}\cdots C^{1'}$ distance (Å)	11.1	11.0	11.1	11.0	11.1	10.9	11.1	10.9	10	11
λ_I angle (°)	49	88	46	90	47	92	46	90	56	56
λ_{II} angle (°)	86	47	88	46	88	49	88	48	56	55
Canonical wobble G:T pair (PDB entry 113d) λ_I/λ_{II} angles (°)							43/77		72/45	

† High-resolution A-form and B-form DNA structures from Olson *et al.* (2001). ‡ The values are for two pairs, O^6-CMG4:T21 (left column) and T9:O^6-CMG16 (right column), in each duplex. § Averaged values in the AATT and the CGCG tracts of the Dickerson–Drew dodecamer (PDB entry 355d), respectively.

Table 3
Sugar puckers of O^6-CMG:T pairs.

$C^{3'}$-*exo* and $C^{4'}$-*exo* belong to the $C^{2'}$-*endo* and $C^{3'}$-*endo* families, respectively. For further details of other residues, see Supplementary Table S2.

	Site 1		Site 2	
Crystal	O^6-CMG4	T21	O^6-CMG16	T9
O^6-CMG4T-1	$C^{2'}$-*endo*	$C^{4'}$-*exo*	$C^{2'}$-*endo*	$C^{3'}$-*endo*
O^6-CMG4T-2	$C^{2'}$-*endo*	$C^{4'}$-*exo*	$C^{2'}$-*endo*	$C^{3'}$-*endo*
O^6-CMG4T-3	$C^{3'}$-*exo*	$C^{4'}$-*exo*	$C^{2'}$-*endo*	$C^{3'}$-*endo*
O^6-CMG4T-4	$C^{2'}$-*endo*	$C^{4'}$-*exo*	$C^{2'}$-*endo*	$C^{3'}$-*endo*

3.2. Formation of O^6-CMG:T base pairs

All unmodified base pairs in the ODN sequence except for those containing O^6-CMG are of the standard Watson–Crick type. $F_o - F_c$ OMIT (Supplementary Fig. S1) and $2F_o - F_c$ electron-density maps of the four crystals indicate that the O^6-CMG residues can interact with the opposite T residues in the palindromic sequence. In all four crystal structures the densities of the two bases of the O^6-CMG:T pair are unusually deformed, showing that the T residues of these pairs move largely toward the major-groove side. This is quite different from the Watson–Crick-type base pair found in the previous ODN (O^6-CMG5T; Zhang *et al.*, 2013), in which O^6-CMG is introduced at the fifth position in the Dickerson–Drew sequence.

3.3. Geometries of the O^6-CMG4:T base pairs

All of the interaction geometries of O^6-CMGs paired with Ts observed in the four crystal structures are shown in Fig. 3. In the O^6-CMG4T-1 crystal the atomic distance between the O^{62} atom of the carboxyl group of the O^6-CMG4 residue and the N^3 atom of the T21 residue suggests the formation of a hydrogen bond between these two sites. In addition, a water molecule bridges between the N^2 atom of O^6-CMG4 and the O^2 atom of T21 through two hydrogen bonds to stabilize the base-pair formation. Although these hydrogen bonds are not direct interactions between the two base moieties, the two bases appear to be held together through the mediation of this water molecule. Furthermore, the methyl C atom (C^{61}) contacts the O^2 atom of T21 at 3.3 Å through double C—H$\cdots$O interactions. The same pairing features are also observed between the O^6-CMG16 and T9 residues. Furthermore, the same pairing mode occurs in all four crystals: O^6-CMG4T-1 and O^6-CMG4T-2 in the presence of Hoechst 33258 and O^6-CMG4T-3 and O^6-CMG4T-4 in the absence of Hoechst 33258 (Fig. 3).

Superimposition of the four duplex structures on the unmodified duplex (PDB entry 355d) shows an interesting feature of the high-wobble pair found in the present structures, as highlighted in Fig. 4(*a*). The purine moieties of the O^6-CMG residues are very near the position of the natural guanine base in the Watson–Crick pair in the B-form duplex. In contrast, however, the T residues opposite are significantly displaced towards the major groove such that the O^4 atoms of the T9 and T21 residues are exposed to the solvent region. The large displacements (wobbling) of Ts in O^6-CMG:T pairs are characterized by the λ_I and λ_{II} angles of the two bases, with the average values being 47 and 88°, respectively, which are substantially different from those of the canonical G:T wobble pair (44 and 74°) and the normal G:C pair (56 and 56°), as shown in Fig. 4(*a*) and Table 2. This high wobble contrasts with a standard G:T wobble pair in that the displacement of T is larger such that the N^3 proton is aligned with the carboxylate side chain of O^6-CMG.

Figs. 5(*a*) and 5(*b*) show the two cases of base-pair stacking between the O^6-CMG:T pair and the adjacent pairs (above and below), which differ in that a purine is stacked on another purine and a pyrimidine is stacked on another pyrimidine (homo case) or the two purine–pyrimidine stacks are crossed (hetero case). In the homo case, the surface areas (van der Waals and water-accessible) vary depending on the pairing types (Watson–Crick type, wobble type and high-wobble type in turn), as shown in Table 4. Comparing the Watson–Crick and the wobble types in the hetero case, the stacked area is decreased in the latter as expected, with the area of the latter in the hetero case being extremely small. It is interesting to note that in the homo case the corresponding area of the high-wobble type is almost the same as that of the wobble type, while in the hetero case the area is increased from that of the wobble type. This comes from the additional carboxymethyl group extended from the O^6 atom, as shown in Fig. 5(*c*), because the area is largely decreased to a comparable value to the wobble type when the carboxymethyl group is removed

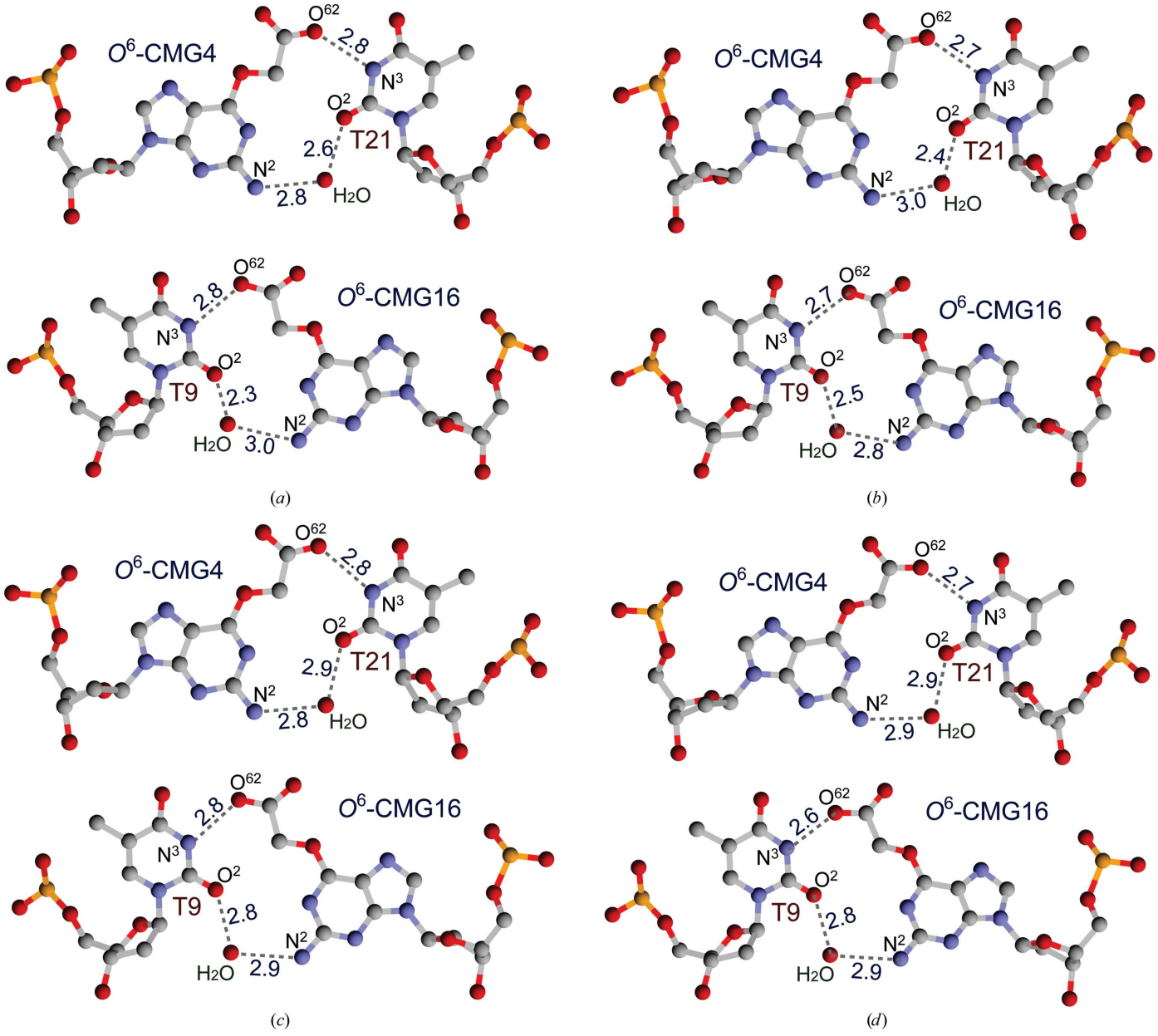

Figure 3
High-wobble O^6-CMG:T pairs found in the O^6-CMG4T-1 (*a*), O^6-CMG4T-2 (*b*), O^6-CMG4T-3 (*c*) and O^6-CMG4T-4 (*d*) crystals. Broken lines show possible hydrogen bonds and values indicate atomic distances (Å).

(Table 4). Therefore, it is deduced that the high-wobble pair is stabilized by this additional effect.

The exposed O^4 atoms of the O^6-CMG:T pairs are out of the base–base stacking and interact with the variety of cations used for crystallization (see Table 1) or water molecules. In the O^6-CMG4T-1 duplex, an Sr^{2+}–Mg^{2+} ion cluster (Supplementary Fig. S2) is bound to the O^4 atom of T21, while another Mg^{2+} ion[4] is bound to the O^4 atom of T9. In O^6-CMG4T-2, an Mg^{2+} ion is bound to the O^4 atom of T21 and a water molecule is bound to the O^4 atom of T9. In both O^6-CMG4T-3 and O^6-CMG4T-4, however, a Ba^{2+} ion is bound to the O^4 atom of T21 and another O^4 of T9 interacts with K^+ in O^6-CMG4T-3 and with Na^+ in O^6-CMG4T-4. The crystallization conditions for O^6-CMG4T-3 and O^6-CMG4T-4 crystals differ in the presence or the absence of potassium ions, but it is difficult to distinguish them from water molecules (Shui, McFail-Isom, Sines *et al.*, 1998; Sines *et al.*, 2000). Therefore, it looks likely that the T bases are not pulled out to interact with the surrounding cations, but that these cations cover the exposed O^4 atoms to stabilize them electrostatically. It could be concluded that the high-wobble structure of the O^6-CMG:T pairs are an intrinsic property of the interaction of O^6-CMG with T. Fig. 4(*b*) shows the averaged geometry of the O^6-CMG:T pairs. The N^1 atom of the O^6-CMG residue, which lacks the N^1 proton of guanine, cannot participate in hydrogen bonding inside the O^6-CMG:T

[4] Temporally assigned but water molecules bound to the cation are not well resolved, perhaps owing to disorder.

Table 4
Surface area (Å^2) covered by bases stacking with each other in the three different pairing types.

Surface area was calculated as the difference of the two molecular surface areas of the two pairs before and after stacking using *NACCESS* (Hubbard & Thornton, 1993). Values are calculated as the van der Waals surface and values in parentheses are those within the water-accessible radius (1.4 Å). The values given in the table are averaged for each type.

Pairing type	Homo case†	Hetero case†
Watson–Crick‡	7.3 (247)	4.6 (239)
Wobble‡	5.9 (259)	0.8 (235)
High-wobble	5.5 (269)	3.4 (264)
High-wobble§	4.6 (251)	0.9 (246)

† The homo case contains purine–purine and pyrimidine–pyrimidine stacks and the hetero case contains purine–pyrimidine crossed stacking (see Fig. 5). ‡ Using the atomic parameters from PDB entries 355d and 113d. § Without the carboxymethyl group.

pair. The electronegative N^1 and O^2 atoms of *O*^{6}-CMG and T, respectively, are separated by 3.9 Å, suggesting that the N^1 atom is not protonated. To form the high-wobble pair, the carboxymethyl group adopts a *syn* (or proximal) conformation relative to the N^1 atoms of the eight observed *O*^{6}-CMG residues, although there is a slight fluctuation from different surrounding effects, as shown in Fig. 4(*a*). This conformation is similar to that found in the *O*^{6}-CMG4C duplex (Zhang *et al.*, 2013), in contrast to the *anti* conformation found in the *O*^{6}-CMG5T duplex.

It is interesting to compare the two structures of the present duplexes with the *O*^{6}-CMG4C duplex (Zhang *et al.*, 2013) because they both contain modified guanines at the same fourth position. Fig. 4(*c*) shows the superimposed structures of *O*^{6}-CMG4:T and *O*^{6}-CMG4:C pairs. In the reversely wobbled pair *O*^{6}-CMG:C, the *O*^{6}-CMG base moves slightly toward the major-groove side to form two hydrogen bonds. In contrast, in the *O*^{6}-CMG4T duplex the T base largely moves towards the major-groove side in the *O*^{6}-CMG:T pairs. To facilitate this movement, the sugar puckering of the T residue adopts a local C$^{3'}$-*endo* conformation (Table 3 and Supplementary Table S2).

From the chemical structure of *O*^{6}-CMG, three modes (I, II and III) of interactions can be expected between *O*^{6}-CMG and T, as shown in Fig. 6. In the Watson–Crick-type *O*^{6}-CMG:T pair found in the *O*^{6}-CMG5T duplex (mode I), the two hydrogen bonds stabilize base-pair formation, but a large propeller twist of 18° is required in order to decrease the repulsion between the O^6(*O*^{6}-CMG) and O^2(T) atoms. In contrast, the corresponding two bases of the *O*^{6}-CMG4T

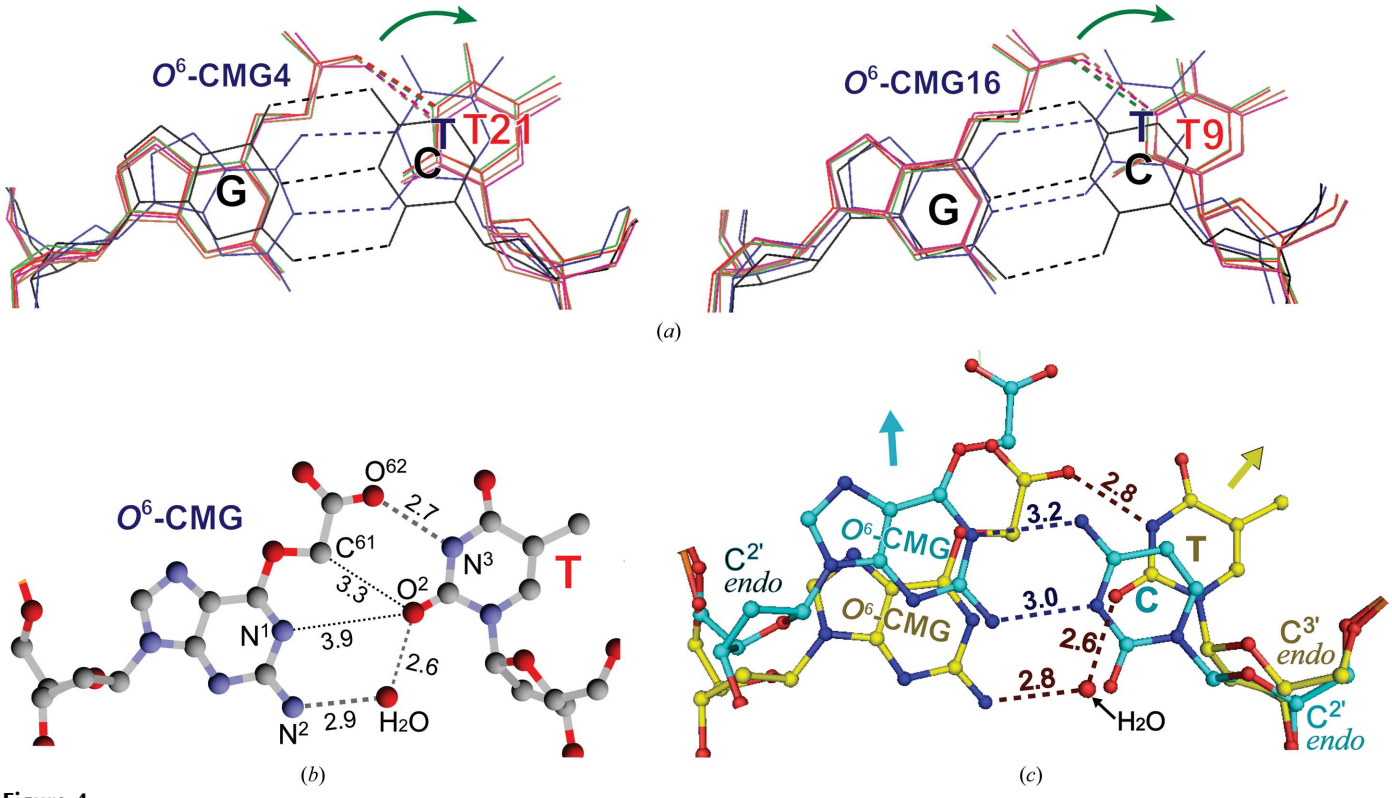

Figure 4
Geometric features of *O*^{6}-CMG:T base pairs. Arrows indicate residue movements, broken lines show possible hydrogen bonds and values are atomic distances in Å. (*a*) *O*^{6}-CMG:T pairs of the *O*^{6}-CMG4T-1, *O*^{6}-CMG4T-2, *O*^{6}-CMG4T-3 and *O*^{6}-CMG4T-4 duplexes (red, green, brown and purple lines, respectively) together with the G:T wobble pair-containing duplex (blue line; PDB entry 113d; Hunter *et al.*, 1987) superimposed on the unmodified duplex (black lines; PDB entry 355d; Shui, McFail-Isom, Hu *et al.*, 1998). The guanine moieties of *O*^{6}-CMG stay near the position of the Watson–Crick G:C pair, while only the T bases move, largely toward the major-groove side. (*b*) Averaged atomic distances of the eight *O*^{6}-CMG:T pairs. (*c*) The high-wobble *O*^{6}-CMG4:T21 pair and the reversed wobble *O*^{6}-CMG4:C21 pair, when superimposed between *O*^{6}-CMG4T-1 (yellow) and *O*^{6}-CMG4C (cyan). The *O*^{6}-CMG residue of *O*^{6}-CMG4C moves toward the major-groove side. The ribose puckers of C9 and C21 residues are C$^{2'}$-*endo* in the *O*^{6}-CMG4C duplex. In contrast, those of T9 and T21 in the four *O*^{6}-CMG4T duplexes fluctuate within the C$^{3'}$-*endo* family, while the opposite *O*^{6}-CMG residues are always in the C$^{2'}$-*endo* family (see Table 3 and Supplementary Table S2).

duplexes form the high-wobble O^6-CMG:T pair (mode II) using alternative hydrogen bonds. In this mode, the N^1(O^6-CMG) and O^2(T) atoms are sufficiently separated from one another to reduce their repulsive effects. This suggests that mode I is less stable than mode II, consistent with the fact that it was difficult to crystallize the O^6-CMG5T duplex in the absence of Hoechst 33258, as discussed previously (Zhang *et al.*, 2013).

There is a question as to why the pairing modes between O^6-CMG and T are different at the (fourth and fifth residue positions). The O^6-CMG4 and O^6-CMG16 residues of the O^6-CMG4T duplex form mode II wobble base pairs with the

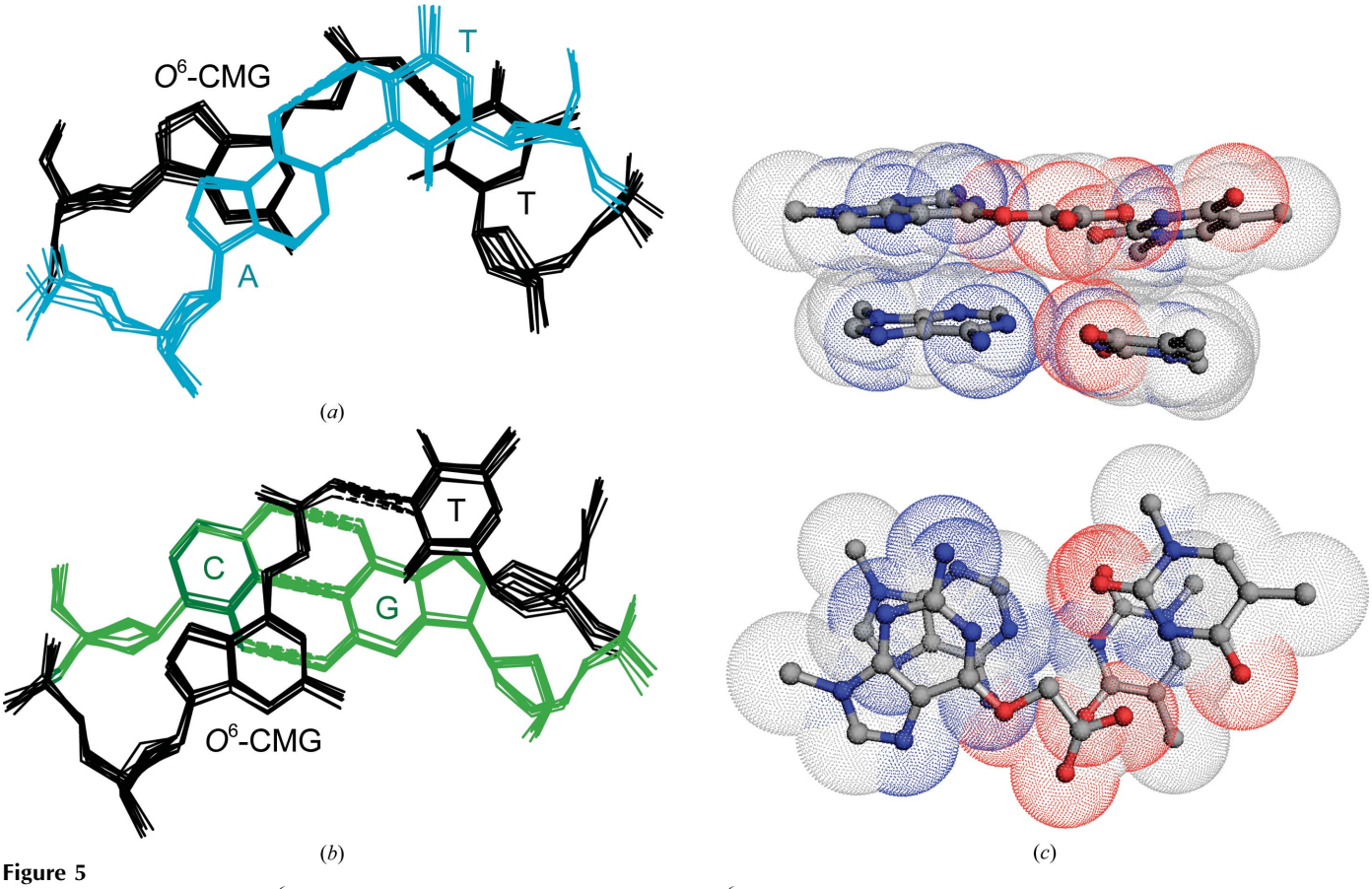

Figure 5
Stacking of the high-wobble O^6-CMG:T pair (black) on the adjacent pairs. (*a*) O^6-CMG:T and A:T (blue) for purine–purine and pyrimidine–pyrimidine stacking (homo case) and (*b*) O^6-CMG:T and C:G (green) for purine–pyrimidine crossed stacking (hetero case). The eight O^6-CMG:T pairs are superimposed on each other in both cases. (*c*) The van der Waals surface in the hetero case is drawn as side and top views using a CPK representation in a ball-and-stick style.

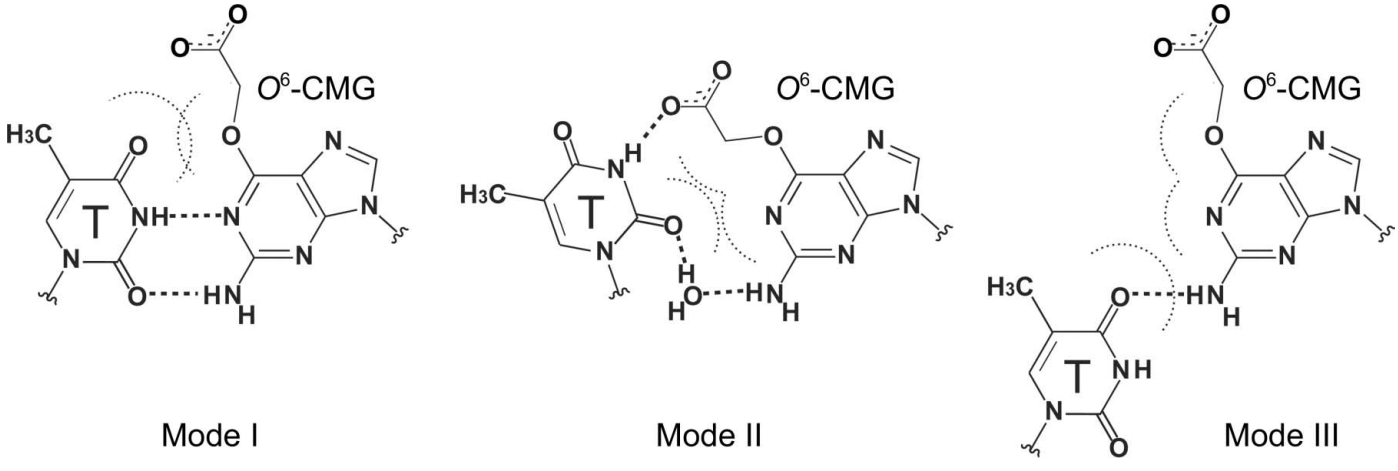

Figure 6
Three possible modes of hydrogen bonding between O^6-CMG and T bases.

respective T21 and T9 residues opposite. In the O^6-CMG5T duplex, however, both the O^6-CMG5 and the O^6-CMG17 residues form mode I pairs with the T20 and T8 residues, respectively. Although this duplex is palindromic, a Hoechst 33528 molecule is asymmetrically bound in the minor groove through four hydrogen bonds, as shown in Fig. 7. This binding scheme of Hoechst 33528 is the same as those of other Dickerson–Drew-type duplexes (Robinson *et al.*, 1998; see also PDB entry 403d). One of the four hydrogen bonds is formed between N^1 of Hoechst 33258 and O^2(T20)[5]. It is known that the Hoechst 33528 molecule binds asymmetrically to the central part of Dickerson–Drew DNA in two ways, as shown in Supplementary Table S3. The N^1 and N^3 donor atoms can form two sets of bifurcated hydrogen bonds (proximal and distal). The acceptor residues differ between the crystals (Robinson *et al.*, 1998; Squire *et al.*, 2000) through rotation of Hoechst 33528 by 180° at the molecular centre. However, disorder of the Hoechst 33528 molecule owing to such rotation is not observed in these crystal structures.

It is possible to consider that in the case of the O^6-CMG5T duplex this hydrogen bond could suppress movement of the T20 residue for large wobbling and results in the formation of a mode I pair with O^6-CMG5. On the other hand, the T8 residue corresponding to T20 in the same O^6-CMG5T duplex is free from such a constraint by Hoechst 33528. Nevertheless, in fact T8 forms a mode I pair, suggesting an intrinsic preference. Contrarily, however, in the O^6-CMG4T-1 and O^6-CMG4T-2 duplexes, even though the T21 and T9 residues are completely free from bound Hoechst 33528, these thymine bases form mode II pairs with O^6-CMGs. To explain this inconsistency, it is necessary to examine the possibility of a positional effect in the sequence.

The Dickerson–Drew sequence is composed of three parts: the central AATT tract and two rigid CGCG tetramers at both ends. The AATT tract is well known to be rather flexible, with the A:T pairs being allowed to adopt large propeller twists (Crothers & Shakked, 1999). Therefore, it is possible to speculate that the effect of the surroundings on the formation of a base pair is different between the two kinds of tracts. In fact, the Watson–Crick-type O^6-CMG:T pairs require a large propeller twist of −18° on average (Zhang *et al.*, 2013), comparable to that of the A:T pair in the AATT tract of the unmodified duplex (the propeller twist angle is −16°; for other cases, see Crothers & Shakked, 1999) and in contrast to the G:C pair (−8.7°) in the CGCG tract. In mode II, the T base positioned out of the base–base stacking is allowed to adopt a large propeller twist (−25° on average; for details, see Table 2). Therefore, it could be speculated that O^6-CMG forms a pair with T by interconversion between the two modes depending on its sequence context: presumably, these pairing modes have similar thermodynamic stabilities. Two similar mode I pairs were found in the crystal structure of the Dickerson–Drew dodecamer containing O^6-MeG at the fourth position (Leonard *et al.*, 1990). In this case, however, it seems unlikely that O^6-MeG can form a stable pair using the mode II motif because it lacks the carboxyl group essential for the motif. Therefore, the mode II motif will only occur in the case of a carboxymethylated guanine base.

4. Biological implications

In a previous paper (Zhang *et al.*, 2013), we suggested that the canonical wobble pair (Crick, 1966; Varani & McClain, 2000) is rejected by sieving the shape of the pairing mode using *in silico* structural modelling of DNA polymerase η in complex with B-DNA containing O^6-CMG. This rejection may also be applicable to the high-wobble pair of O^6-CMG:T. In the case of the O^6-CMG:T pair, however, the bound T base could interconvert between mode I and mode II (see Fig. 6). Since the base-pair binding pocket of the DNA polymerase would be flexible to some extent for accepting Watson–Crick-type pairing between four kinds of template bases and between four kinds of dNTPs (2′-deoxynucleoside triphosphate), in order to adapt to the requisites of the polymerase the two ligands between the O^6-CMG residue and dTTP (2′-deoxythymidine triphosphate) or between the T residue and dO^6-CMGTP (2′-deoxy-O^6-carboxymethylguanosine triphosphate) could adopt mode I pair formation after local separation to two single strands, even if the mode II pair is preferred in a rigid track of DNA duplex.

The mutagenicity of O^6-MeG has long been ascribed to the preference of high-fidelity DNA polymerases to mis-insert T rather than C opposite the damaged base during replication. The basis for this derives from structural information on damaged DNA in the absence of polymerase, which reveals

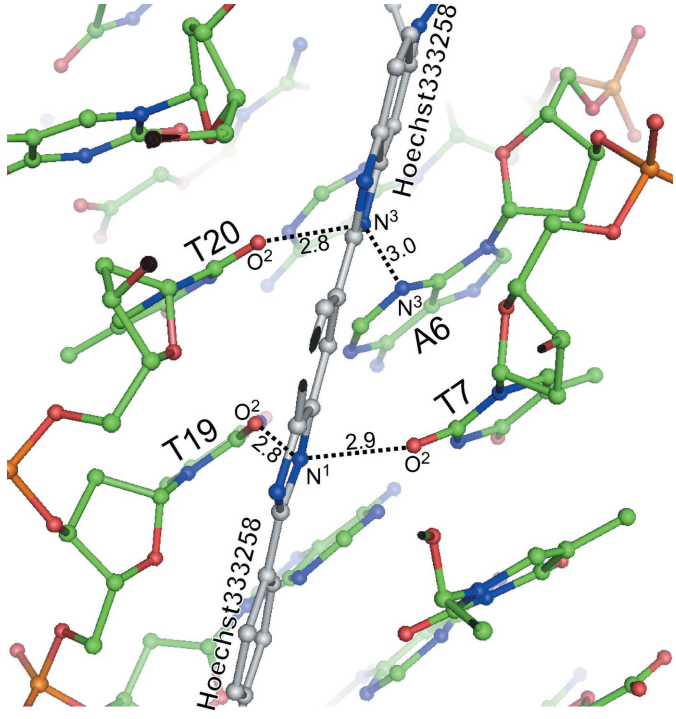

Figure 7
Asymmetric binding of Hoechst 33258 in the minor groove of the palindromic O^6-CMG4T-1 duplex. Dotted lines indicate possible hydrogen bonds with their lengths in Å.

[5] For the numbering system of the residues, see Fig. 1(*b*).

the O^6-MeG:T pair to be being largely isosteric to the natural G:C and A:T Watson–Crick pairs (Leonard *et al.*, 1990). In contrast, the alkylated base forms a reversed wobble pair with cytosine (Leonard *et al.*, 1990) similar to that found in our O^6-CMG4C duplex (Zhang *et al.*, 2013). More recently, information has become available from kinetic and, significantly, structural studies of complexes involving DNA polymerase, O^6-MeG-containing template and the nucleotides dCTP (2′-deoxycytidine triphosphate) or dTTP. The high-fidelity DNA polymerase I large fragment from *Bacillus stearothermophilus* (BF) incorporates dTTP about tenfold more efficiently than dCTP opposite O^6-MeG (Warren *et al.*, 2006). Interestingly, when bound within the active site, both C and T form base pairs with O^6-MeG with Watson–Crick-type geometry. However, whether the C:O^6-MeG pair is protonated is not confirmed. In the O^6-MeG:T pair, unusually, the methyl group of the alkylated base is directed towards the O^4 atom of T (*syn* conformation), with which a weak hydrogen-bond electrostatic interaction is proposed (Warren *et al.*, 2006). It is suggested that forcing the O^6-MeG:C pair into a Watson–Crick-type base geometry incurs an energetic penalty that results in a slight preference for the insertion of T since the O^6-MeG:T pair already has the requisite shape for nucleotide insertion. Replication past O^6-MeG within DNA templates is generally relatively inefficient and many O^6-alkylG lesions may instead be replicated by translesion Y-family polymerases such as pol δ, η, ι and κ. Kinetic data examining the insertion of dCTP or dTTP opposite a variety of O^6-alkylG lesions by these polymerases has shown that for pol δ, η and κ the efficiency of incorporation of dCTP and dTTP was similar, whilst pol ι showed a preference for inserting dTTP (Choi *et al.*, 2006). The Y-family polymerase Dpo4 from *Sulfolobus solfataricus* has also recently been crystallized in complex with an O^6-MeG-containing DNA template and dTTP or dCTP (Eoff *et al.*, 2007). In this complex, a reversed wobble structure similar to the O^6-CMG:C pair (Zhang *et al.*, 2013) was observed for the C:O^6-MeG pair. In contrast, a normal Watson–Crick-type pair was observed between T and O^6-MeG, also similar to the mode I O^6-CMG:T pair (Zhang *et al.*, 2013). These structural features explain the fact that this polymerase shows a preference for the correct insertion of C opposite the lesion since the preferred reversed wobble pairing is now possible.

We have already proposed two possible mechanisms for purine transition mutations during DNA replication when the template G residues are O^6-carboxymethylated and when dGTP (2′-deoxyguanosine triphosphate) in the nucleoside triphosphate pool is O^6-carboxymethylated (Zhang *et al.*, 2013). The previous *in silico* structural modelling of DNA polymerase η in complex with B-DNA containing O^6-CMG suggested that both a Watson–Crick-type O^6-CMG:T pair and the reversed wobble O^6-CMG:C pair could be accommodated within the polymerase active site. In the present study, the stabilization of O^6-CMG residues in double-stranded DNA by using the two types of pairing modes depending on the tract context has been discussed. In the DNA replication step, however, as DNA is split into two single strands locally, when the free O^6-CMG is incorporated in the polymerase active site the mode I pair could be formed with dTTP. In the same manner, dO^6-CMGTP also can form a mode I pair with the template T residue because polymerase should allow an A:T pair with large propeller twist to accept mode I pairs. Recently, we have shown that O^6-CMG is repaired by the human protein O^6-methylguanine-DNA methyltransferase (Senthong *et al.*, 2013). However, there is previous evidence that the nucleotide-excision repair (NER) pathway is also able to process this damaged base (O'Driscoll *et al.*, 1999). The type of O^6-CMG:T high-wobble base pair observed here has to our knowledge not been identified in other O^6-alkylG:T base pairs and may well be a unique feature of carboxymethylguanine. In this regard, it is intriguing to speculate that this type of structure may well facilitate recognition and repair by the NER pathway.

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. This work was supported in part by grants from the Biotechnology and Biological Sciences Research Council, UK (to OJW and DMW), Cancer Research UK (to GPM) and the Engineering and Physical Sciences Research Council, UK (to CLM and DMW).

References

Berger, I., Kang, C., Sinha, N., Wolters, M. & Rich, A. (1996). *Acta Cryst.* D**52**, 465–468.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
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., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Choi, J.-Y., Chowdhury, G., Zang, H., Angel, K. C., Vu, C. C., Peterson, L. A. & Guengerich, F. P. (2006). *J. Biol. Chem.* **281**, 38244–38256.
Crick, F. H. (1966). *J. Mol. Biol.* **19**, 548–555.
Crothers, D. M. & Shakked, Z. (1999). *Oxford Handbook of Nucleic Acid Structure*, edited by S. Neidle, pp. 455–470. Oxford University Press.
Daniels, D. S., Woo, T. T., Luu, K. X., Noll, D. M., Clarke, N. D., Pegg, A. E. & Tainer, J. A. (2004). *Nature Struct. Mol. Biol.* **11**, 714–720.
DeLano, W. L. (2008). *PyMOL*. http://www.pymol.org.
Dickerson, R. E. & Drew, H. R. (1981). *J. Mol. Biol.* **149**, 761–786.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Eoff, R. L., Irimia, A., Egli, M. & Guengerich, F. P. (2007). *J. Biol. Chem.* **282**, 1456–1467.
Esteller, M., Toyota, M., Sanchez-Cespedes, M., Capella, G., Peinado, M. A., Watkins, D. N., Issa, J. P., Sidransky, D., Baylin, S. B. & Herman, J. G. (2000). *Cancer Res.* **60**, 2368–2371.
García-Santos, M. P., Calle, E. & Casado, J. (2001). *J. Am. Chem. Soc.* **123**, 7506–7510.
Gisone, P., Dubner, D., Del Pérez Rosario, M., Michelin, S. & Puntarulo, S. (2004). *In Vivo*, **18**, 281–292.
Gladwin, M. T. (2004). *J. Clin. Invest.* **113**, 19–21.
Gottschalg, E., Scott, G. B., Burns, P. A. & Shuker, D. E. G. (2007). *Carcinogenesis*, **28**, 356–362.
Harris, V. H., Smith, C. L., Cummins, W. J., Hamilton, A. L., Adams, H., Dickman, M., Hornby, D. P. & Williams, D. M. (2003). *J. Mol. Biol.* **326**, 1389–1401.
Hubbard, S. J. & Thornton, J. M. (1993). *NACCESS*. Department of Biochemistry and Molecular Biology, University College London.

Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D. & Kennard, O. (1987). *J. Biol. Chem.* **262**, 9962–9970.
Juan, E. C. M., Shimizu, S., Ma, X., Kurose, T., Haraguchi, T., Zhang, F., Tsunoda, M., Ohkubo, A., Sekine, M., Shibata, T., Millington, C. L., Williams, D. M. & Takénaka, A. (2010). *Nucleic Acids Res.* **38**, 6737–6745.
Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998). *Nature (London)*, **391**, 304–307.
Leonard, G. A., Thomson, J., Watson, W. P. & Brown, T. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 9573–9576.
Lewin, M. H., Bailey, N., Bandaletova, T., Bowman, R., Cross, A. J., Pollock, J., Shuker, D. E. & Bingham, S. A. (2006). *Cancer Res.* **66**, 1859–1865.
Lu, X.-J. & Olson, W. K. (2003). *Nucleic Acids Res.* **31**, 5108–5121.
Margison, G. P., Santibáñez Koref, M. F. & Povey, A. C. (2002). *Mutagenesis*, **17**, 483–487.
Millington, C. L., Watson, A. J., Marriott, A. S., Margison, G. P., Povey, A. C. & Williams, D. M. (2012). *Nucleosides Nucleotides Nucleic Acids*, **31**, 328–338.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
O'Driscoll, M., Macpherson, P., Xu, Y.-Z. & Karran, P. (1999). *Carcinogenesis*, **20**, 1855–1862.
Olson, W. K. *et al.* (2001). *J. Mol. Biol.* **313**, 229–237.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Povey, A. C., Hall, C. N., Badawi, A. F., Cooper, D. P. & O'Connor, P. J. (2000). *Gut*, **47**, 362–365.
Robinson, H., Gao, Y.-G., Bauer, C., Roberts, C., Switzer, C. & Wang, A. H.-J. (1998). *Biochemistry*, **37**, 10897–10905.
Sayle, R. A. & Milner-White, E. J. (1995). *Trends Biochem. Sci.* **20**, 374–376.
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. & Povey, A. C. (2013). *Nucleic Acids Res.* **41**, 3047–3055.
Shephard, S. E. & Lutz, W. K. (1989). *Cancer Surv.* **8**, 401–421.
Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. (1998). *Biochemistry*, **37**, 8341–8355.
Shui, X., McFail-Isom, L., Sines, C., VanDerveer, D. & Williams, L. D. (1998). *Biochemistry*, **37**, 16877–16887.
Shuker, D. E. G. & Margison, G. P. (1997). *Cancer Res.* **57**, 366–369.
Sines, C. C., McFail-Isom, L., Howerton, S. B., VanDerveer, D. & Williams, L. D. (2000). *J. Am. Chem. Soc.* **122**, 11048–11056.
Squire, C. J., Baker, L. J., Clark, G. R., Martin, R. F. & White, J. (2000). *Nucleic Acids Res.* **28**, 1252–1258.
Teng, M.-K., Usman, N., Frederick, C. A. & Wang, A. H.-J. (1988). *Nucleic Acids Res.* **16**, 2671–2690.
Varani, G. & McClain, W. H. (2000). *EMBO Rep.* **1**, 18–23.
Vlieghe, D., Turkenburg, J. P. & Van Meervelt, L. (1999). *Acta Cryst.* D**55**, 1495–1502.
Wang, W., Hellinga, H. W. & Beese, L. S. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 17644–17648.
Warren, J. J., Forsberg, L. J. & Beese, L. S. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 19701–19706.
Wiederholt, K., Rajur, S. B. & McLaughlin, L. W. (1997). *Bioconjug. Chem.* **8**, 119–126.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Zhang, F., Tsunoda, M., Suzuki, K., Kikuchi, Y., Wilkinson, O., Millington, C. L., Margison, G. P., Williams, D. M., Morishita, E. C. & Takénaka, A. (2013). *Nucleic Acids Res.* **41**, 5524–5532.