

# $O^6$ -carboxymethylguanine in DNA forms a sequence context dependent wobble base pair structure with thymine.

Fang Zhang<sup>1</sup>, Masaru Tsunoda<sup>2</sup>, Yuji Kikuchi<sup>1,2</sup>, Oliver Wilkinson<sup>3,a</sup>, Christopher L. Millington<sup>3,b</sup>, Geoffrey P. Margison<sup>4</sup>, David M. Williams<sup>3,\*</sup> and Akio Takénaka<sup>2,5,\*</sup>

<sup>1</sup>Graduate School of Science and Engineering, <sup>2</sup>Faculty of Pharmacy, Iwaki-Meisei University, Iwaki 970-8551, Japan, <sup>3</sup>Centre for Chemical Biology, Department of Chemistry, Krebs Institute, University of Sheffield, Sheffield S3 7HF, UK, <sup>4</sup>Centre for Occupational and Environmental Health, Faculty of Medical & Human Sciences, University of Manchester, Manchester M13 9PL, UK and <sup>5</sup>Research Institute, Chiba Institute of Technology, Tsudanuma, Narashino 275-0016, Japan

## **Supplementary data**

**Tables 1~3 and Figures 1-2.**

<sup>a</sup> Current address: DNA-Protein Interaction Unit, School of Biochemistry, University of Bristol, Bristol, BS8 1TD, UK

<sup>b</sup> Current address: Institut de Génétique et Développement de Rennes, Université de Rennes, 35043 Rennes Cedex, France

\*To whom correspondence should be addressed to: Email: [atakenak@sakura.email.ne.jp](mailto:atakenak@sakura.email.ne.jp)  
Correspondence may also be addressed to: Email: [d.m.williams@sheffield.ac.uk](mailto:d.m.williams@sheffield.ac.uk)

**Table 1.** Phosphate backbone conformation of the four duplexes, together with the unmodified duplex.

Base	$\alpha(^{\circ})$					$\beta(^{\circ})$					$\gamma(^{\circ})$					$\delta(^{\circ})$					$\epsilon(^{\circ})$					$\zeta(^{\circ})$				
	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
Strand I																														
C	-	-	-	-	-	-	-	-	-	-	-54	-60	-66	-46	-70	149	150	139	142	145	-148	-156	-168	-160	-172	-140	-140	-147	-144	-98
G	-57	-49	-28	-32	-70	171	170	147	156	-172	37	37	41	33	43	135	138	128	132	148	179	179	-173	-168	-151	-103	-106	-94	-106	-157
C	-60	-58	-62	-50	-39	164	164	167	161	131	50	49	51	45	50	89	94	97	97	93	-176	-174	177	179	-165	-93	-94	-103	-101	-81
G/CMG	-46	-50	-47	-58	-65	171	174	170	171	174	51	54	54	60	50	148	144	151	144	145	171	169	-178	188	-167	-99	-100	-96	-92	-145
A	-61	-62	-50	-61	-47	177	172	175	175	156	61	63	42	51	54	135	122	119	131	133	-169	-177	-178	-170	-178	-115	-102	-103	-115	-91
A	-56	-56	-48	-49	-64	172	178	179	162	-173	51	53	42	48	46	126	126	130	127	128	-177	-173	-177	-171	177	-91	-91	-86	-98	-96
T	-59	-60	-63	-55	-49	168	168	174	170	173	51	48	55	47	49	106	107	107	115	113	177	179	167	177	-177	-88	-87	-91	-109	-96
T	-53	-61	-53	-46	-54	179	177	-177	172	168	45	54	52	51	53	123	121	139	133	114	-176	-174	-170	-174	172	-103	-100	-109	-110	-95
C/T	-46	-50	-51	-44	-53	157	157	157	155	-173	49	51	51	46	51	95	93	92	92	138	-163	-166	-168	-167	-156	-75	-74	-76	-80	-96
G	-63	-65	-71	-69	-60	178	-176	179	-179	163	43	43	56	51	40	151	152	148	147	143	-116	-121	-113	-108	-100	161	168	157	157	146
C	-64	-63	-61	-60	-73	133	131	138	130	144	58	62	53	53	51	146	149	142	159	144	-146	-142	-171	-159	-164	-94	-102	-92	-105	-126
G	105	102	-72	-70	53	-167	-169	-177	179	144	177	176	45	42	-66	80	82	135	135	148	-	-	-	-	-	-	-	-	-	-
Strand II																														
G	-71	-50	-63	-60	-62	170	170	154	154	170	50	42	59	43	53	92	117	81	82	88	-	-	-	-	-	-	-	-	-	-
C	-52	-50	-58	-48	-57	149	142	145	137	130	41	45	34	33	51	133	122	137	134	84	-175	-173	-169	-162	-161	-100	-98	-124	-120	-78
G	-58	-57	-66	-55	-67	177	174	-179	173	177	41	46	49	45	45	144	137	139	139	144	-139	-141	-117	-124	-149	-176	-174	170	173	-173
C/T	-55	-53	-66	-61	-58	160	158	158	161	160	55	48	61	51	51	92	92	88	96	96	-169	-173	-169	-167	-170	-82	-80	-69	-79	-81
T	-52	-49	-44	-44	-45	177	176	179	172	172	52	52	51	52	43	130	129	136	135	135	-172	-172	-176	-169	-164	-102	-98	-98	-106	-107
T	-58	-63	-69	-71	-49	171	172	167	170	179	50	54	59	53	52	103	99	98	103	128	174	173	167	175	-173	-92	-90	-94	-104	-107
A	-60	-63	-50	-159	-62	178	-179	176	136	176	51	53	50	159	45	127	126	133	141	113	179	-180	-172	-156	166	-91	-90	-95	-81	-91
A	-57	-45	-61	-66	-56	-167	-169	-165	-165	-160	58	51	56	54	54	142	142	139	139	146	174	175	175	179	-178	-94	-94	-96	-76	-92
G/CMG	-56	-61	-56	-61	-62	166	169	169	171	176	61	64	57	61	62	140	142	146	142	141	151	156	161	159	151	-94	-102	-98	-89	-90
C	-53	-51	-49	-37	-57	162	165	153	155	160	56	52	58	46	57	98	104	91	97	82	-176	-171	-169	-176	-174	-94	-95	-95	-96	-85
G	-53	-47	-67	-64	-58	149	147	164	157	167	37	34	43	46	50	127	132	119	128	115	-178	-179	175	-179	178	-103	-105	-108	-117	-94
C	-	-	-	-	-	-	-	-	-	-	57	58	61	68	59	155	153	148	147	138	-129	-129	-147	-145	-169	176	172	-111	-124	-105

\* #1, O<sup>6</sup>-CMG4T-1; #2, O<sup>6</sup>-CMG4T-2; #3, O<sup>6</sup>-CMG4T-3; #4, O<sup>6</sup>-CMG4T-4; #5, unmodified duplex PDB-ID=355D

**Table 2.** Sugar puckers of the four duplexes, together with the unmodified duplex.

		$O^6$ -CMG4T-1	$O^6$ -CMG4T-2	$O^6$ -CMG4T-3	$O^6$ -CMG4T-4	PDB-ID=355D
#	Strand I					
1	C	C2'-endo	C2'-endo	C1'-exo	C2'-endo	C2'-endo
2	G	C2'-endo	C2'-endo	C2'-endo	C2'-endo	C2'-endo
3	C	C4'-exo	C4'-exo	C4'-exo	C4'-exo	C4'-exo
4	CMG/G	C2'-endo	C2'-endo	C3'-exo	C2'-endo	C2'-endo
5	A	C2'-endo	C1'-exo	C1'-exo	C1'-exo	C2'-endo
6	A	C1'-exo	C2'-endo	C1'-exo	C2'-endo	C2'-endo
7	T	O4'-endo	O4'-endo	C1'-exo	O4'-endo	C1'-exo
8	T	C1'-exo	C1'-exo	C2'-endo	C2'-endo	C1'-exo
9	T/C	C3'-endo	C3'-endo	C3'-endo	C3'-endo	C2'-endo
10	G	C2'-endo	C2'-endo	C2'-endo	C2'-endo	C2'-endo
11	C	C2'-endo	C2'-endo	C2'-endo	C2'-endo	C2'-endo
12	G	C3'-endo	C3'-endo	C2'-endo	C2'-endo	C3'-exo
#	Strand II					
24	G	C4'-exo	O4'-endo	C3'-endo	C3'-endo	C3'-endo
23	C	C1'-exo	C1'-exo	C2'-endo	C2'-endo	C3'-endo
22	G	C2'-endo	C2'-endo	C1'-exo	C2'-endo	C2'-endo
21	T/C	C4'-exo	C4'-exo	C4'-exo	C4'-exo	O4'-endo
20	T	C1'-exo	C1'-exo	C2'-endo	C2'-endo	C2'-endo
19	T	O4'-endo	O4'-endo	O4'-endo	O4'-endo	C1'-exo
18	A	C1'-exo	C1'-exo	C1'-exo	C3'-exo	C1'-exo
17	A	C2'-endo	C2'-endo	C2'-endo	C2'-endo	C2'-endo
16	CMG/G	C2'-endo	C2'-endo	C2'-endo	C2'-endo	C2'-endo
15	C	C4'-exo	C4'-exo	C4'-exo	C4'-exo	C4'-exo
14	G	C2'-endo	C2'-endo	C1'-exo	C1'-exo	C1'-exo
13	C	C2'-endo	C2'-endo	C3'-exo	C2'-endo	C2'-endo

# Residue number

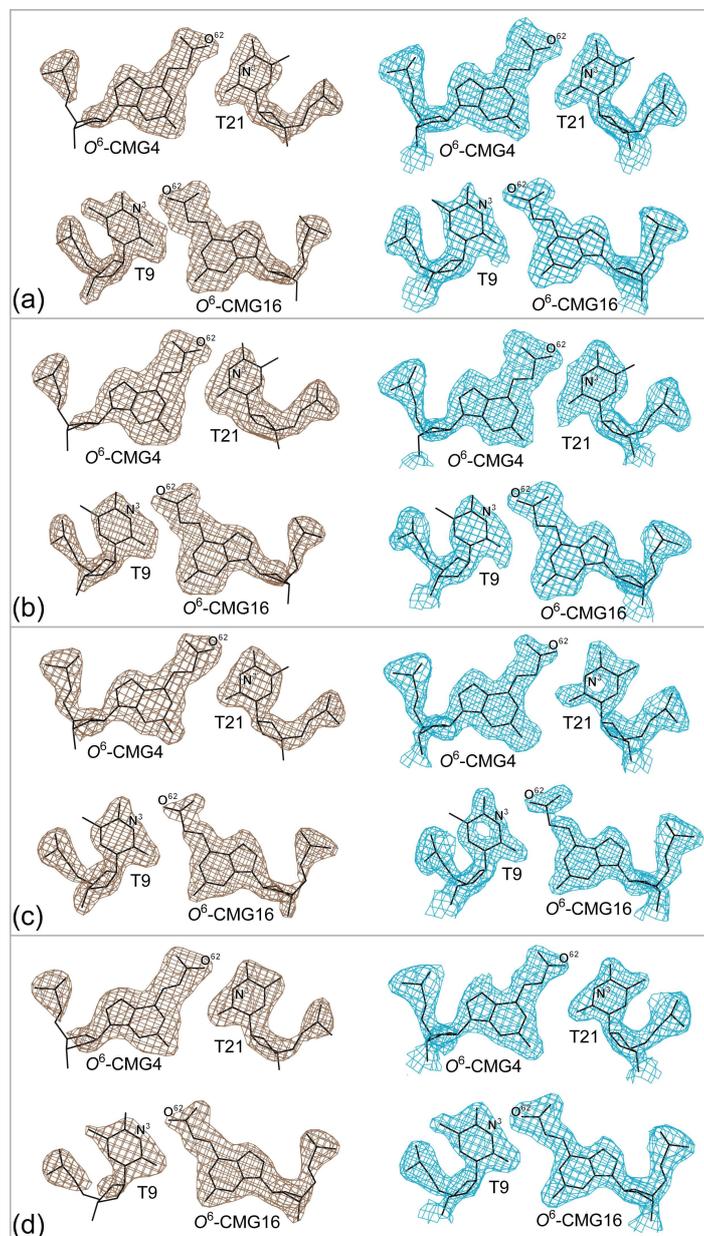
The three types of C<sup>2</sup>-endo and C<sup>3</sup>-endo families, and others are colored blue, red and green, respectively. The 3<sup>rd</sup> C and 15<sup>th</sup> C residues always adopt a C4'-exo pucker to form the G:G interaction between the two duplexes. Only the 9<sup>th</sup> and 19<sup>th</sup> T residues adopt a C<sup>3</sup>-endo pucker, different from the unmodified DNA.

**Table 3.** Two modes of Hoechst33258 binding in the minor groove.

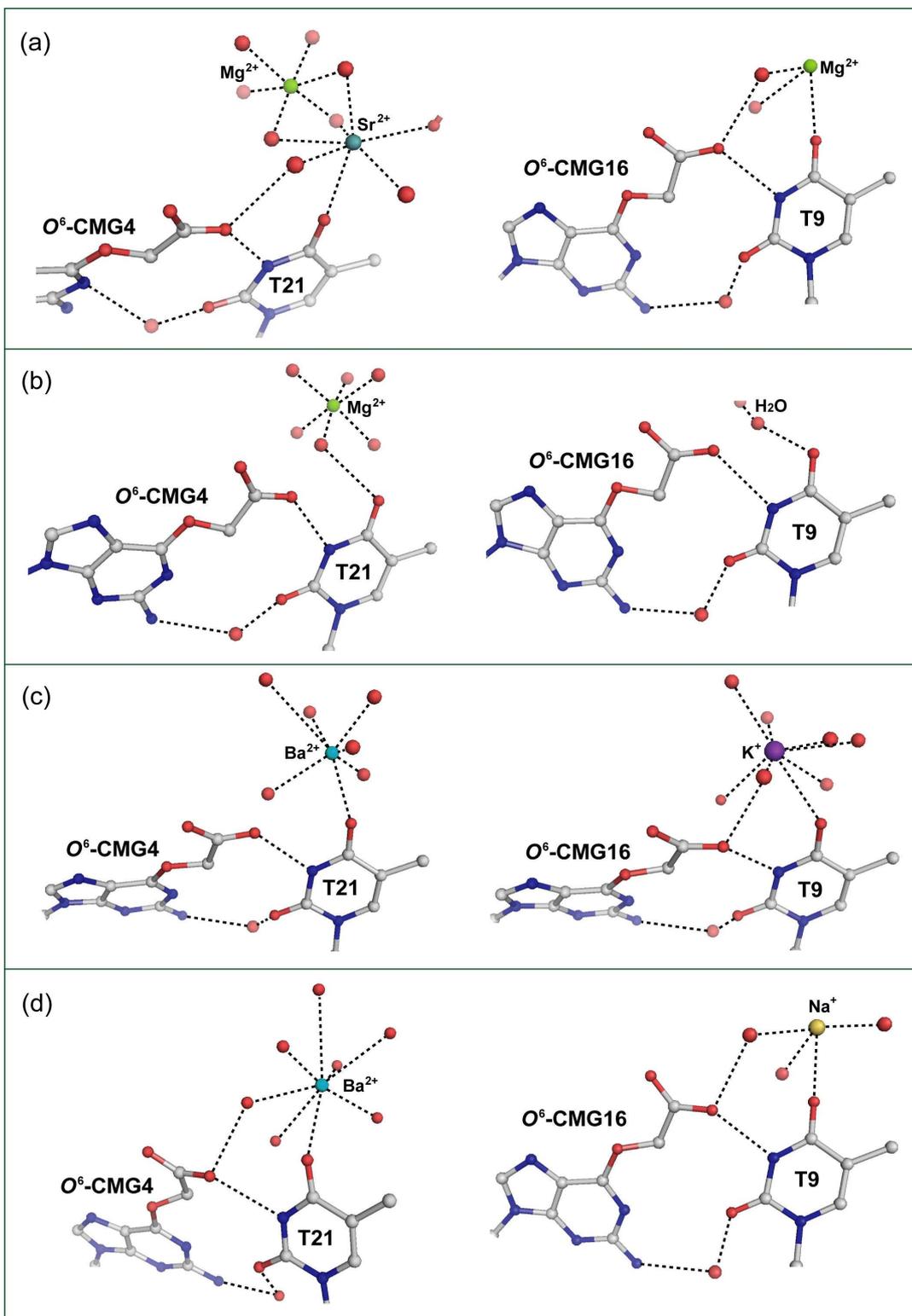
Hydrogen bond donor	N <sup>1</sup> -H		N <sup>3</sup> -H	
	Proximal	Distal	Proximal	Distal
Acceptor (mode A)*	O <sup>2</sup> (T19)	O <sup>2</sup> (T7)	N <sup>3</sup> (A6)	O <sup>2</sup> (T20)
Acceptor (mode B)*	O <sup>2</sup> (T20)	N <sup>3</sup> (A6)	O <sup>2</sup> (T7)	O <sup>2</sup> (T19)

\* Between the two modes, the orientation of Hoechst33258 molecule varies with 180° rotation at the molecular centre, but the acceptor atoms and their residues of the palindromic dodecamer are not equivalent (PDB ID 403D, 442D, 443D) (Robinson *et al.*, 1998; Squire *et al.*, 2000).

O<sup>2</sup>(T19) indicates the O<sup>2</sup> atom of the T19 residue.



**Figure 1.**  $F_0-F_c$  Omit maps (brown lines) of  $O^6$ -CMG4 and T21 residues and  $O^6$ -CMG16 and T9 residues (left), contoured at the  $3.0\sigma$  level, and their  $2F_0-F_c$  maps (cyan lines) contoured at the  $1.0\sigma$  level (right); (a) for  $O^6$ -CMG4T-1, (b) for  $O^6$ -CMG4T-2, (c) for  $O^6$ -CMG4T-3 and (d) for  $O^6$ -CMG4T-4. Cations and some water atoms are also omitted. These maps were depicted by the program Dino [Biasini, M., Mariani, V., Haas, J., Scheuber, S., Schenk, A.D., Schwede, T. & Philippsen, A. (2010) Open Structure: a flexible software framework for computational structural biology, *Bioinformatics*, **26**, 2626-2628].



**Figure 2.** Cations and water molecules contacting to the O<sup>4</sup> atoms of thymine residues protruded into solvent region by large wobbling, found in the O<sup>6</sup>-CMG4T-1(a), O<sup>6</sup>-CMG4T-2(b), O<sup>6</sup>-CMG4T-3(c) and O<sup>6</sup>-CMG4T-4(d) crystals.

## References

- Lu, X.J. & 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.
- 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., Sklenar, H., Suzuki, M., Tung, C.-S., Westhof, E., Wolberger, C. & Berman, H. (2001). A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, **313**, 229-237.
- Robinson, H., Gao, Y.G., Bauer, C., Roberts, C., Switzer, C. & Wang, A.H.-J. (1998). 2'-Deoxyisoguanosine adopts more than one tautomer to form base pairs with thymidine observed by high-resolution crystal structure analysis. *Biochem.*, **37**, 10897-10905.
- Squire, C.J., Baker, L.J., Clark, G.R., Martin, R.F. & White, J. (2000). Structures of *m*-iodo Hoechst-DNA complexes in crystals with reduced solvent content: Implications for minor groove binder drug design. *Nucleic Acids Res.*, **28**, 1252-1258.
- Shui, X., McFail-Isom, L., Sines, C., Derveer, v.D. & Williams, L.D. (1998). Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochem.*, **37**, 16877-16887.