Supporting Information

CH---O Hydrogen Bonds Mediate Highly Specific Recognition of Methylated CpG sites by the Zinc Finger Protein Kaiso

Evgenia N. Nikolova†, Robyn L. Stanfield, H. Jane Dyson and Peter E. Wright*

Department of Integrative Structural and Computational Biology and Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California, 92037

* Corresponding Author Peter E. Wright, email wright@scripps.edu

† Present Address: Department of Biophysics, Johns Hopkins University, 3400 North Charles Street, Baltimore MD 21218

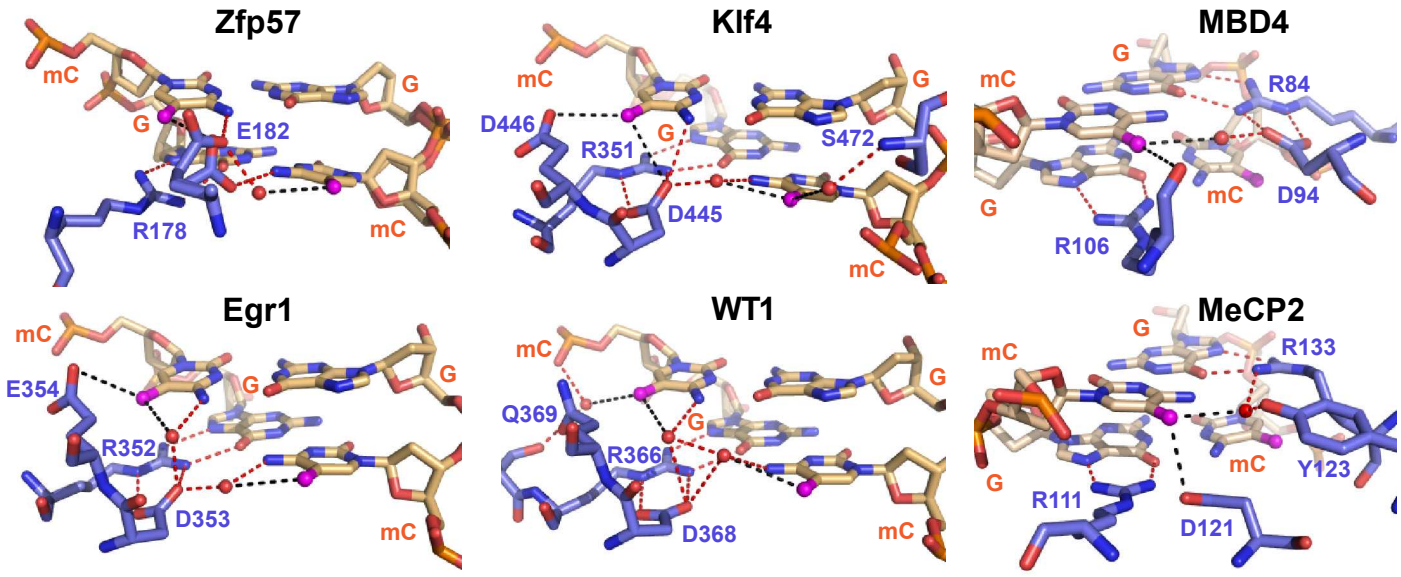

Figure S1. Common mode of mCpG recognition by methyl-CpG binding proteins (MBPs). Shown are specific major-groove interactions between protein residues (blue) and a DNA mCpG step (wheat) in crystal structures of zinc finger protein Zfp57 (PDB: 4GZN), Krüppel-like factor 4 Klf4 (PDB: 4M9E), methyl-CpG binding domain protein 4 (MBD4, PDB: 3VXX), Early growth response protein 1 (Egr1, PDB: 4R2A), Wilms tumor protein 1 (WT1, PDB: 4R2E), and methyl-CpG binding protein 2 (MeCP2, PDB: 3C2I) bound to methylated DNA (Kaiso is shown in Figure 1). mC methyl groups and water molecules are displayed as magenta and red spheres, respectively. Proposed direct and water-mediated canonical (red) and CH---O (black) hydrogen bonds are shown as dashed lines. The two alternative conformations of the E182 side chain are both shown for Zfp57.

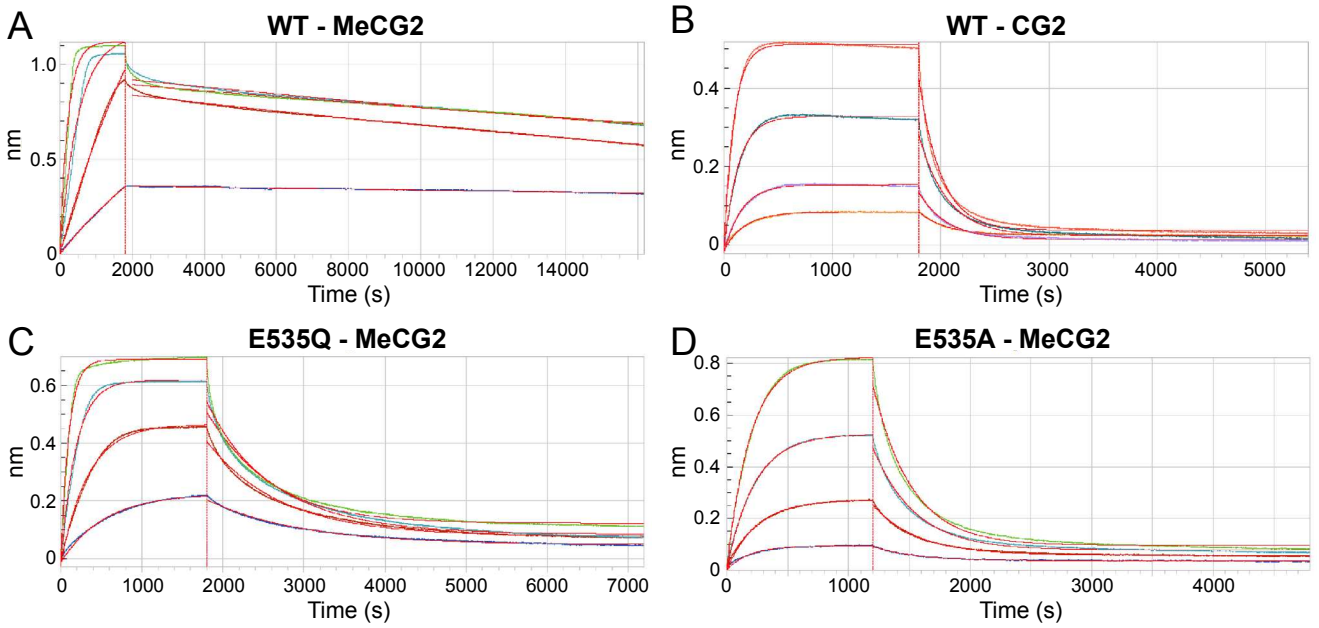

Figure S2. Representative biolayer interferometry binding data of (A) WT Kaiso and MeCG2, (B) WT Kaiso and CG2, (C) Kaiso E535Q and MeCG2, and (D) Kaiso E535A and MeCG2. Shown are association and dissociation curves as a function of time at four different protein concentrations (12.5, 25, 50, and 100 nM). Best fits to a two-state binding model are colored red (see Materials and Methods for details).

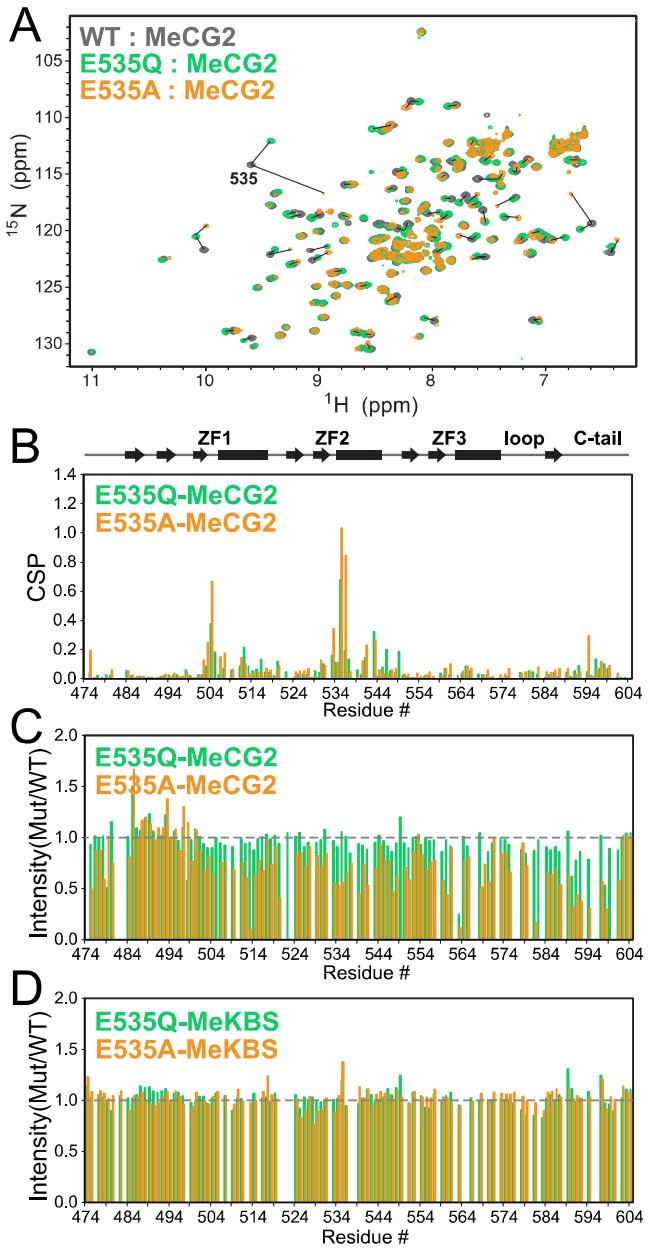

Figure S3. Effect of Kaiso E535A and E535Q mutations on Kaiso-MeCG2 complex conformation and dynamics. (A) Overlay of ^{1}H-^{15}N-HSQC spectra of WT, E535Q, and E535A Kaiso bound to MeCG2. (B) Corresponding weighted average CSPs plotted as a function of residue number. (C) Comparison of the normalized intensity ratios of the MeCG2 complexes of Kaiso E535Q (green) and E535A (orange) relative to WT showing extensive line broadening, particularly for E535A-MeCG2, that becomes progressively larger from ZF1 towards the C-terminal tail. (D) Comparison of the normalized intensity ratios of Kaiso E535Q and E535A relative to WT complexes with MeKBS.

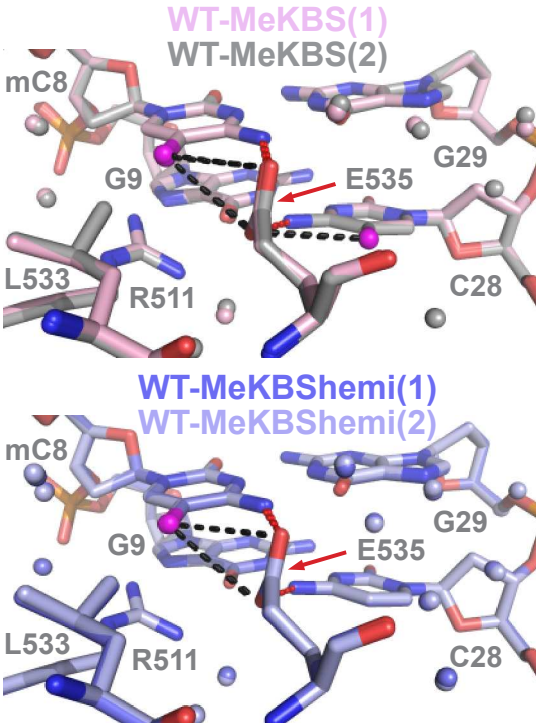

Figure S4. Structural agreement between the two Kaiso WT-MeKBS (1 and 2, top) and Kaiso WT-MeKBShemi (1 and 2, bottom) complexes, obtained under the same crystallization conditions. The local structure near the 5' mCpG steps of DNA are shown.

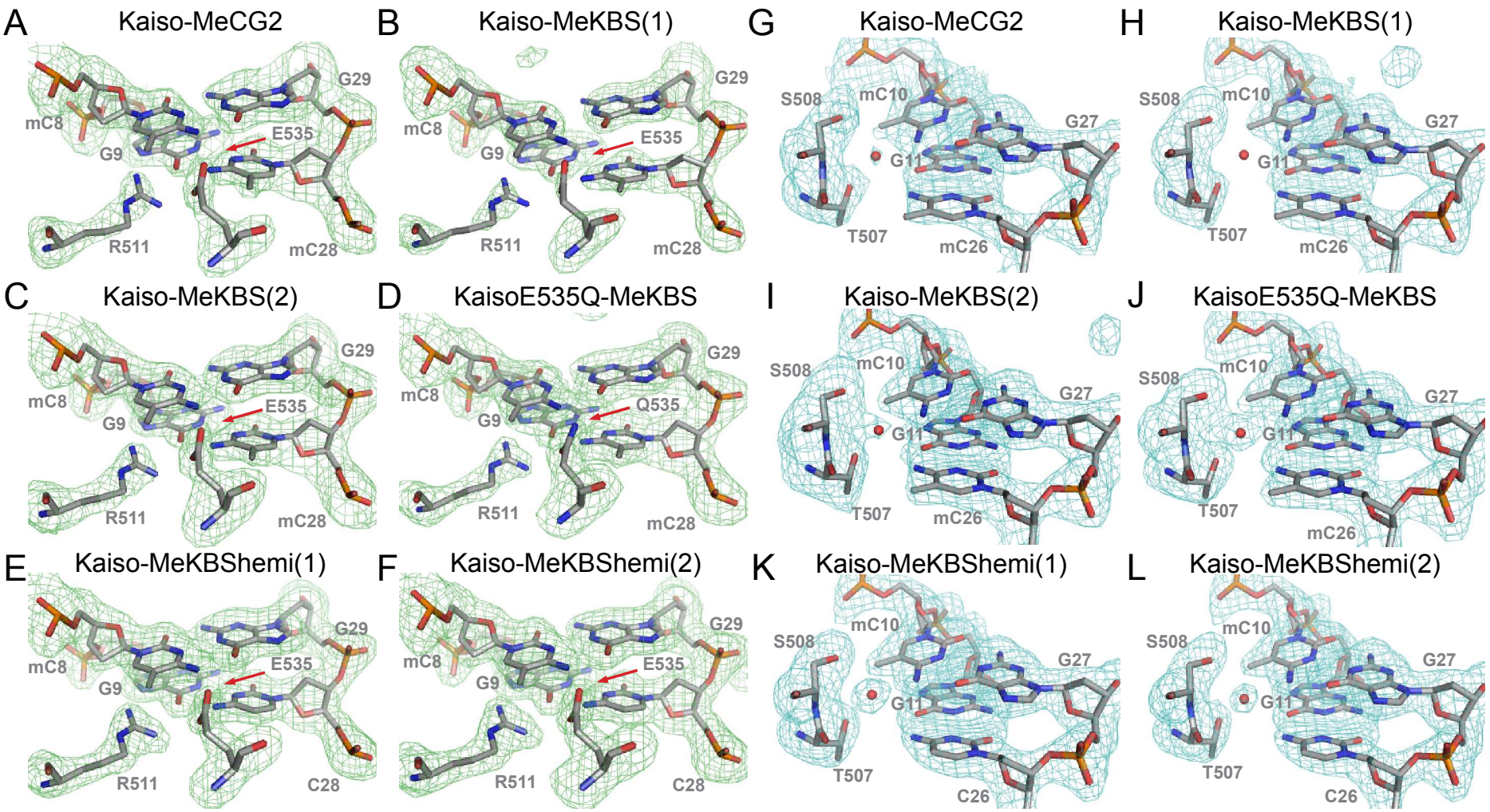

Figure S5. Polder omit maps showing interactions between Kaiso E/Q535 and R511 and the 5' mCpG site (A-F) and between Kaiso T507 and S508 and the 3' mCpG site (G-L). The maps were created in Phenix and contoured at 2.5 σ (A-F) or 2.0 σ (G-L).

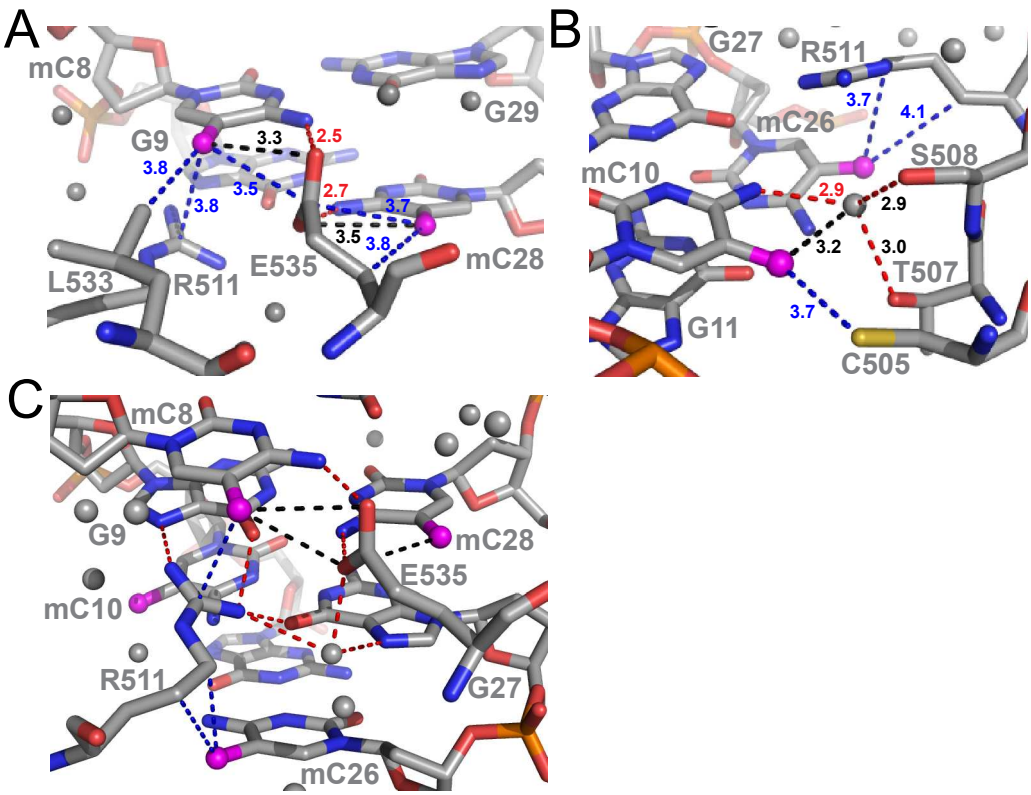

Figure S6. Summary of polar and van der Waals interactions between Kaiso and mC methyl groups in the Kaiso-MeKBS complex (structure (1) from Table S2). (A) Interactions between Kaiso R511, L533, and E535 and the 5' mCpG step (mC8 and mC28). (B) Interactions between Kaiso C505, T507, S508, and R511 and the 3' mCpG step (mC10 and mC26). (C) Extensive network of direct and water-mediated interactions connecting R511 with both the 5' and 3' mCpG step in MeKBS. Dashed lines indicate canonical H-bonds (red), CH---O H-bonds (black), and van der Waals contacts (blue), distances are labeled in A and B. The cytosine methyl groups are shown as magenta spheres and gray spheres indicate bound water molecules.

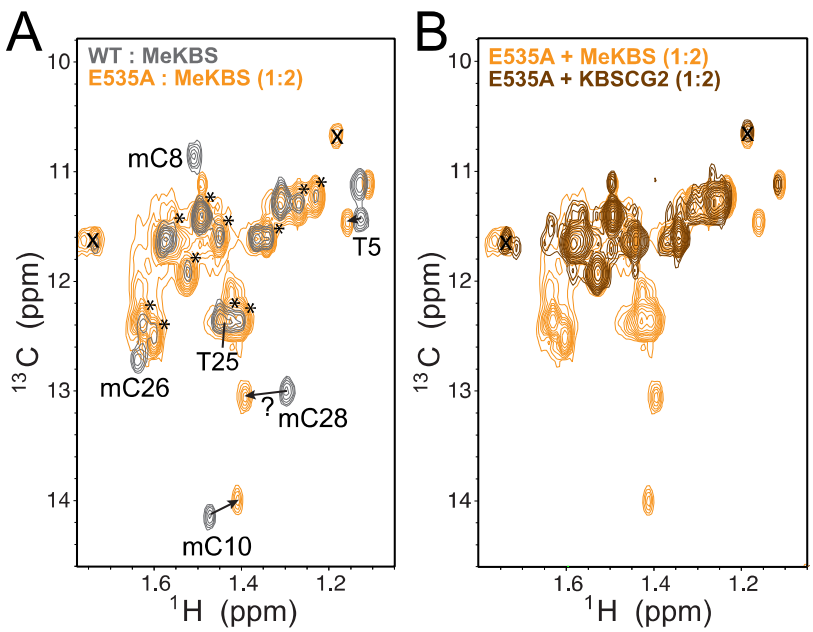

Figure S7. Effect of Kaiso E535A mutant on methylated DNA mC methyl groups. (A) Overlay of the natural abundance ^{1}H-^{13}C-HMQC methyl spectra for MeKBS in complex with WT Kaiso (gray) and E535A mutant (orange). Peaks resulting from excess free DNA in E535A complexes are indicated by asterisk (*), non-DNA peaks are indicated with an "X". DNA mC methyl crosspeaks are shifted and/or broadened in E535A mutant complexes. (B) Overlay of the natural abundance ^{1}H-^{13}C-HMQC methyl spectra for MeKBS (orange) and KBSCG2 (brown) in complex with Kaiso E535A, showing the absence of mC methyl peaks in the latter spectrum due to free and bound DNA.

Table S1. DNA constructs.

DNA construct	Top strand	Bottom strand
NMR and crystallography		
MeKBS	5’ TGCTTCC**C**G**C**GAATAACG*	5’ CGTTATT**C**G**C**GGGAAGCA
MeKBSsemi	5’ TGCTTCC**C**GCGAATAACG	5’ CGTTATTCG**C**GGGAAGCA
MeKBShemi	5’ TGCTTCC**C**G**C**GAATAACG	5’ CGTTATTCGCGGGAAGCA
KBSCG2	5’ TGCTTCCCGCGAATAACG	5’ CGTTATTCGCGGGAAGCA
MeCG2	5’ TGCTTCT**C**G**C**GAGAAGCA	5’ TGCTTCT**C**G**C**GAGAAGCA
MeCG2semi	5’ TGCTTCT**C**GCGAGAAGCA	5’ TGCTTCTCG**C**GAGAAGCA
MeCG2hemi	5’ TGCTTCT**C**G**C**GAGAAGCA	5’ TGCTTCTCGCGAGAAGCA
CG2	5’ TGCTTCTCGCGAGAAGCA	5’ TGCTTCTCGCGAGAAGCA
Biolayer Interferometry		
MeKBS	5’/Biotin/GGGCTTCC**C**G**C**GAATAACG	5’ CGTTATT**C**G**C**GGGAAGCCC
MeCG2	5’/Biotin/GGGCTTCT**C**G**C**GAGAAGCG	5’ CGCTTCT**C**G**C**GAGAAGCCC
MeCG2semi	5’/Biotin/GGGCTTCT**C**GCGAGAAGCG	5’ CGCTTCTCG**C**GAGAAGCCC
MeCG2hemi	5’/Biotin/GGGCTTCT**C**G**C**GAGAAGCG	5’ CGCTTCTCGCGAGAAGCCC
CG2	5’/Biotin/GGGCTTCTCGCGAGAAGCG	5’ CGCTTCTCGCGAGAAGCCC

*5-methylcytosine is colored red and underlined.

Table S2. Crystallographic statistics for Kaiso-DNA complexes.

Statistics	Kaiso-MeCG2	Kaiso-MeKBS(1)	Kaiso-MeKBS(2)	Kaiso-MeKBShemi(1)	Kaiso-MeKBShemi(2)	Kaiso(E535Q)-MeKBS
Data collection						
Beam line	SSRL (BL11-1)	SSRL (BL12-2)	SSRL (BL12-2)	SSRL (BL12-2)	SSRL (BL12-2)	SSRL (BL9-2)
Wavelength (Å)	0.9795	0.9795	0.9795	0.9795	0.9795	0.9795
Resolution range (Å)	45.44 - 2.31 (2.40 - 2.31)	42.34 - 2.35 (2.43 - 2.35)	45.67 - 2.40 (2.48 - 2.40)	43.10 - 2.05 (2.12 - 2.05)	43.04 - 2.00 (2.07 - 2.00)	45.24 - 2.32 (2.40 - 2.32)
Space group	C 2 2 2	C 2 2 2	C 2 2 2	C 2 2 2	C 2 2 2	C 2 2 2
Cell dimensions						
a b c (Å)	44.22 183.72 104.57	43.99 185.01 105.11	44.34 184.15 105.19	44.34 183.98 105.07	44.28 183.61 105.03	44.41 183.49 104.00
α β γ (°)	90 90 90	90 90 90	90 90 90	90 90 90	90 90 90	90 90 90
Total reflections	104329 (3445)	189922 (10420)	163193 (6448)	279038 (14444)	327293 (17828)	121542 (9016)
Unique reflections	17383 (946)	17827 (1609)	17193 (1540)	27026 (2370)	28976 (2521)	18669 (1629)
Multiplicity	6.0 (3.6)	10.4 (6.5)	9.5 (4.2)	10.3 (6.1)	11.3 (7.1)	6.5 (5.5)
Completeness (%)	90.8 (51.2)	96.2 (67.8)	98.4 (88.8)	98.0 (87.8)	98.4 (88.2)	98.4 (87.8)
Average I/σ(I)	13.1 (4.3)	11.4 (2.5)	12.8 (2.0)	12.3 (3.9)	12.6 (4.1)	18.9 (4.5)
Wilson B-factor (Å^2)	35.4	49.2	47.9	36.5	36.2	37.7
R-merge (%)	8.6 (27.7)	11.9 (56.2)	14.8 (63.4)	8.91 (47.3)	7.4 (46.4)	6.3 (30.9)
R-meas. (%)	9.4	12.5	15.6	9.4	7.8	6.9
CC1/2 (%)	99.6 (94.5)	99.4 (92.6)	99.7 (85.8)	99.8 (94.1)	99.8 (97.0)	99.8 (97.0)
Refinement						
Resolution (Å)	2.4	2.4	2.4	2.1	2.0	2.3
Reflections used for R-free	10.0%	10.0%	10.0%	10.0%	10.0%	10.0%
R-work (%)	20.5 (22.6)	17.9 (32.1)	17.7 (31.3)	16.2 (21.6)	16.4 (22.0)	16.7 (22.1)
R-free (%)	24.5 (27.4)	21.6 (36.8)	21.7 (34.9)	18.8 (22.5)	18.8 (24.5)	19.9 (28.2)
Number of atoms	2010	1841	1849	1917	1892	1909
macromolecules	1930	1810	1805	1812	1793	1811
ligands[†]	4	4	4	4	4	14
water	76	27	40	101	95	84
Protein residues	155	160	160	158	160	161
RMS(bonds)	0.009	0.009	0.008	0.009	0.008	0.007
RMS(angles)	1.09	1.11	1.15	1.13	1.12	1.10
Ramachandran favored (%)	96.5	96.8	95.2	98.0	97.6	96.0
Ramachandran allowed (%)	3.5	3.2	4.8	2.0	2.4	4.0
Clashscore	5.8	5.2	3.7	5.5	3.7	3.4
Average B-factor (Å^2)	67.9	78.6	75.2	60.4	61.9	60.3
macromolecules (Å^2)	68.7	79.0	75.8	61.1	62.5	61.2
ligands (Å^2)[†]	39.2	56.3	53.3	39.8	40.3	40.9
solvent (Å^2)	46.9	53.0	52.2	49.8	50.9	45.2

*Numbers in parenthesis are for the highest-resolution shell

[†] Refers to zinc, chloride, sodium and tartrate ions

Table S3. Hydrogen bonding distances between E535/Q535 and the 5' (hemi-)mCpG step in Kaiso-DNA crystal structures.

Bonded atoms	Kaiso-MeCG2	Kaiso-MeKBS(1)	Kaiso-MeKBS(2)	Kaiso-MeKBShemi(1)	Kaiso-MeKBShemi(2)	Kaiso(E535Q)-MeKBS
N---O						
535(Oε2/Nε2)-mC8(N4)[†] (Å)	2.72	2.49	2.77	2.66	2.69	2.98*
535(Oε1)-mC28(N4) (Å)	2.89	2.73	2.78	2.84	2.85	2.78
C---O/N						
535(Oε2/Nε2)-mC8(C5A) (Å)	3.44	3.31	3.54	3.61	3.66	4.38*
535(Oε1)-mC8(C5A) (Å)	3.51	3.39	3.38	3.56	3.58	3.66
535(Oε2/Nε2)-mC28(C5A) (Å)	3.69	4.01	4.05	-	-	3.89*
535(Oε1)-mC28(C5A) (Å)	3.42	3.47	3.52	-	-	3.48

[†] mC denotes 5-methylcytosine

* Distance between Q535(Nε2) and mC8(C5A), mC8(N4), and mC28(C5A) in the E535Q-MeKBS complex are shown for comparison.