Recognition and cleavage of 5-methylcytosine DNA by bacterial SRA-HNH proteins

Tiesheng Han[1], Megumu Yamada-Mabuchi[2], Gong Zhao[1], Li Li[3], Guang liu[1], Hong-Yu Ou[1], Zixin Deng[1], Yu Zheng[2,*] and Xinyi He[1,*]

[1] State Key Laboratory of Microbial Metabolism, and School of Life Sciences & Biotechnology Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai, 200030, China

[2] New England BioLabs, Inc., 240 County Road, Ipswich, MA, 01938, USA.

[3] Engineering Research Center of Industrial Microbiology (Ministry of Education), and College of Life Sciences, Fujian Normal University, Fuzhou, Fujian 350108 (China)

*For Correspondence:

Xinyi He Tel: (86)-2162932943-2131(EXT) E-mail: xyhe@sjtu.edu.cn

Yu Zheng Tel: (1)-617-417-1366 E-mail: yu.zhengyu@gmail.com

SUPPLEMENTARY MATERIAL

Supplementary Tables

Supplementary Figure S1-19

Supplementary methods

Supplementary Tables

Table S1. Strains and plasmids used in this study.

STRAINS	CHARACTERISTICS	REFERENCE
Escherichia coli		
DH10B	F- *mcr*A Δ*(mrr- hsd*RMS- *mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara*, *leu*)7697 *gal*U *gal*K λ- *rps*L *nup*G	GIBCO BRL
JTU006	DH10B derivative with a *dcm* gene replaced by *aad*A	(1)
BL21(DE3)/pLysS	F- *omp*T *hsd*S(r_B- m_B-) *dcm* gal λ(DE3) [pLysS Cam^r]	Novagen
ER2566	F- λ- fhuA2 [*lon*] *omp*T *lac*Z::T7 gene 1 *gal sul*A11 Δ(*mcr*C-*mrr*)114::IS10 R(*mcr*-73::miniTn10-TetS)2 R(zgb-210::Tn10)(TetS) *end*A1 [*dcm*]	NEB
ER2984	*F' proA+B+ lacIq* ∆*lacZM15 / fhuA2* ∆*(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)TetS endA1 thi-1* ∆*(hsdS-mcrB)5*	NEB
T7 Express	fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10	NEB
PLASMIDS	**CHARACTERISTICS**	**REFERENCE**
pET44b	Expression vector with 6XHis-tag. Amp^r	Novagen
pJTU4356	pET44b derivative for heterologous expression of Sco5333.	This study
pJTU4356*	pJTU4356 with a spontaneous amino acid change of E42G.	This study
pJTU4356M	pJTU4356 harbouring triple amino acid changes at HNH domain, namely His228Ala, Asn244Ala and His253Ala	This study
pJTU4381	pJTU4356 derivative with a amino acid change at Gly32Ala	This study
pJTU4382	pJTU4356 derivative with a amino acid change at Tyr50Ala	This study
pJTU4383	pJTU4356 derivative with a amino acid change at His228Ala	This study
pJTU4384	pJTU4356 derivative with a amino acid change at His253Ala	This study
pJTU4386	pJTU4356 derivative with a amino acid change at Cys252Asp	This study
pUC18	2686 bp, *bla*, *lac*Zα, *ori*^{pMB1}	(2)
pBR322	4361bp, *bla*, *tet*, *ori*^{pMB1}	(3)
pTXB1	Protein expression vector with *Mxe GyrA* intein-CBD tag. Amp^r	(4-5)
pTbis1	pTXB1 derivative for over-expression of Tbis1	This study
pET15b	Protein expression vector with 6×His-tag. Amp^r	Novagen
pJTU4357	pET15b derivative for over-expression and purification of Dcm	This study
pRSFDuet1	Protein co-expression vector. 3829bp, Kan^r, *ori*^{RSF}	Novagen
pJTU4385	pRSFDuet1 derivative for heterologous expression of Dcm.	This study

Table S2. Primers that are used in this study

PRIMERS	SEQUENCE	USE
5333EX-F	gatccatatgattggtgagatccctg	to clone *sco5333* with NdeI site
5333EX-R	gatcctcgagctggcgccaacgttcacggt	to clone *sco5333* with XhoI site
219bp-F	ccctgacgagcatcacaa	to amplify 219bp
219bp-R	cgaacgacctacaccgaact	to amplify 219bp
dcmEX-F	aattcatatgatgcaggaaaatatatcagt	to clone *dcm* with NdeI site
dcmEX-R	agaattcttatcgtgaacgtcggcc	to clone *dcm* with EcoRI site
5333M-F	gtacggacgccatgctgtgtctctgcccgaactgtgccgtgatgctcg	to clone pJTU4356M

5333M-R	ggtccgggccgttgtgggggcgccccagagcctgaatagcggcgccttcg	to clone pJTU4356M
G32A-F	ccctccaggcagccatctgcggtacg	to clone pJTU4381
G32A-R	ggcgatggaccccggcttcgtgggc	to clone pJTU4381
Y50A-F	tatccggcggagccaaggacgacg	to clone pJTU4382
Y50A-R	ccacgatggactctgccccagtcgc	to clone pJTU4382
H228A-F	gaaggcgccgctattcaggctctggg	to clone pJTU4383
H228A-R	gctatagaaaccgccagggacttcg	to clone pJTU4383
H253A-F	gcccgaactgtgccgtgatgctcg	to clone pJTU4384
H253A-R	agagacacagcatgttgtccgtacgg	to clone pJTU4384
Tbis1-F	ccccccccatatgcaccatcaccatcaccataaggcgcacaatacgctgaaattc	to clone pTbis1
Tbis1-R	ccccccgaattctcagcgaggctgggttactgc	to clone pTbis1

Table S3. DNA Oligos that are used in this study

OLIGOS	SEQUENCE
55nt-1	5-FAM-gaaacccgacaggactataaagatac^{m}**C**aggcgtttcccctggaagctccctcgt
55nt-2	5-FAM-gaaacccgacaggactataaagataccaggcgtttcccctggaagctccctcgt
55nt-3	acgagggagcttccaggggaaacgc^{m}**C**tggtatctttatagtcctgtcgggtttc
55nt-4	acgagggagcttccaggggaaacgcctggtatctttatagtcctgtcgggtttc
55nt-5	5-FAM-actataaagataccaggcgtttcccc^{m}**C**tggaagctccctcgtgcgctctcctgtt
55nt-6	5-FAM-actataaagataccaggcgtttcccctggaagctccctcgtgcgctctcctgtt
55nt-7	aacaggagagcgcacgagggagcttc^{m}**C**aggggaaacgcctggtatctttatagt
55nt-8	aacaggagagcgcacgagggagcttccaggggaaacgcctggtatctttatagt
55nt-9	5-FAM-gaaacccgacaggactataaagatag^{m}**C**aggcgtttcccctggaagctccctcgt
55nt-10	5-FAM-gaaacccgacaggactataaagatagcaggcgtttcccctggaagctccctcgt
55nt-11	acgagggagcttccaggggaaacgc^{m}**C**tgctatctttatagtcctgtcgggtttc
55nt-12	acgagggagcttccaggggaaacgcctgctatctttatagtcctgtcgggtttc
55nt-13	5-FAM-gaaacccgacaggactataaagataa^{m}**C**aggcgtttcccctggaagctccctcgt
55nt-14	5-FAM-gaaacccgacaggactataaagataacaggcgtttcccctggaagctccctcgt
55nt-15	acgagggagcttccaggggaaacgc^{m}**C**tgttatctttatagtcctgtcgggtttc
55nt-16	acgagggagcttccaggggaaacgcctgttatctttatagtcctgtcgggtttc
55nt-17	5-FAM-gaaacccgacaggactataaagatat^{m}**C**aggcgtttcccctggaagctccctcgt
55nt-18	5-FAM-gaaacccgacaggactataaagatatcaggcgtttcccctggaagctccctcgt
55nt-19	acgagggagcttccaggggaaacgc^{m}**C**tgatatctttatagtcctgtcgggtttc
55nt-20	acgagggagcttccaggggaaacgcctgatatctttatagtcctgtcgggtttc
55nt-21	5-FAM-gaaacccgacaggactataaagatac^{m}**C**tggcgtttcccctggaagctccctcgt
55nt-22	5-FAM-gaaacccgacaggactataaagatacctggcgtttcccctggaagctccctcgt
55nt-23	acgagggagcttccaggggaaacgc^{m}**C**aggtatctttatagtcctgtcgggtttc
55nt-24	acgagggagcttccaggggaaacgccaggtatctttatagtcctgtcgggtttc
55nt-25	5-FAM-gaaacccgacaggactataaagatac^{m}**C**cggcgtttcccctggaagctccctcgt
55nt-26	5-FAM-gaaacccgacaggactataaagatacccggcgtttcccctggaagctccctcgt
55nt-27	acgagggagcttccaggggaaacgc^{m}**C**gggtatctttatagtcctgtcgggtttc
55nt-28	acgagggagcttccaggggaaacgccgggtatctttatagtcctgtcgggtttc

55nt-29	5-FAM-gaaacccgacaggactataaagatac^{m}**C**gggcgtttcccccctggaagctccctcgt
55nt-30	5-FAM-gaaacccgacaggactataaagataccgggcgtttcccccctggaagctccctcgt
55nt-31	acgagggagcttccaggggggaaacgc^{m}**C**cggtatctttatagtcctgtcgggtttc
55nt-32	acgagggagcttccaggggggaaacgcccggtatctttatagtcctgtcgggtttc
54nt-1	gaaacccgacaggactataaagatac^{m}**C**ggcgtttcccccctggaagctccctcgt
54nt-2	gaaacccgacaggactataaagataccggcgtttcccccctggaagctccctcgt
54nt-3	acgagggagcttccaggggggaaacgc^{m}**C**ggtatctttatagtcctgtcgggtttc
54nt-4	acgagggagcttccaggggggaaacgccggtatctttatagtcctgtcgggtttc
56nt-1	cggcgtttccgggttccataggctccgctc^{m}**C**ggcctctgatgaccagggcatcaca
56nt-2	cggcgtttccgggttccataggctccgctccggcctctgatgaccagggcatcaca
56nt-3	tgtgatgccctggtcatcagaggc^{m}**C**ggagcggagcctatggaacccggaaacgccg
56nt-4	tgtgatgccctggtcatcagaggccggagcggagcctatggaacccggaaacgccg
56nt-5	cggcgtttccgggttccataggctccgctc^{hm}**C**ggcctctgatgaccagggcatcaca
56nt-6	tgtgatgccctggtcatcagaggc^{hm}**C**ggagcggagcctatggaacccggaaacgccg
66nt-1	aaaaacggcgtttccgggttccataggctccgcac^{m}**C**ggtctctgatgaccagggcatcacattttt-5-FAM
66nt-2	aaaaacggcgtttccgggttccataggctccgcaccggtctctgatgaccagggcatcacattttt-5-FAM
66nt-3	5-FAM-aaaaatgtgatgccctggtcatcagagac^{m}**C**ggtgcggagcctatggaacccggaaacgccgttttt
66nt-4	5-FAM-aaaaatgtgatgccctggtcatcagagaccggtgcggagcctatggaacccggaaacgccgttttt

REFERENCE

1. Zhou, H., Wang, Y., Yu, Y., Bai, T., Chen, L., Liu, P., Guo, H., Zhu, C., Tao, M. and Deng, Z. (2012) A non-restricting and non-methylating Escherichia coli strain for DNA cloning and high-throughput conjugation to Streptomyces coelicolor. *Current microbiology*, **64**, 185-190.
2. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103-119.
3. Sutcliffe, J.G. (1978) Nucleotide sequence of the ampicillin resistance gene of Escherichia coli plasmid pBR322. *Proceedings of the National Academy of Sciences of the United States of America*, **75**, 3737-3741.
4. Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A. *et al.* (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, **192**, 271-281.
5. Evans, T.C., Jr., Benner, J. and Xu, M.Q. (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein science : a publication of the Protein Society*, **7**, 2256-2264.

Supplementary Figures:

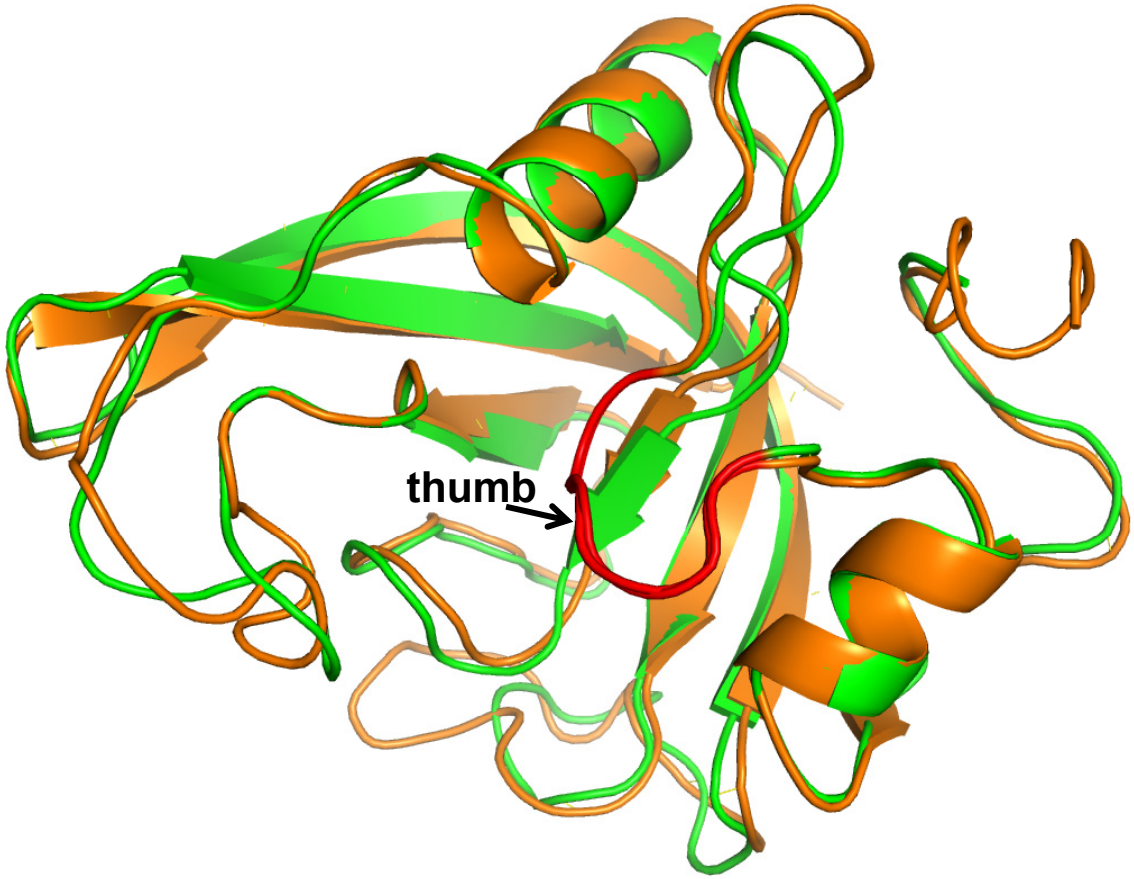


Figure S1. Superposition of SRA_{Tbis1} (Orange) and SRA_{5333} (Green)

The predicted tertiary structure of SRA_{Tbis1} (in Orange) and SRA_{5333} (in green) are almost identical. The thumb of two complexes is colored in red.

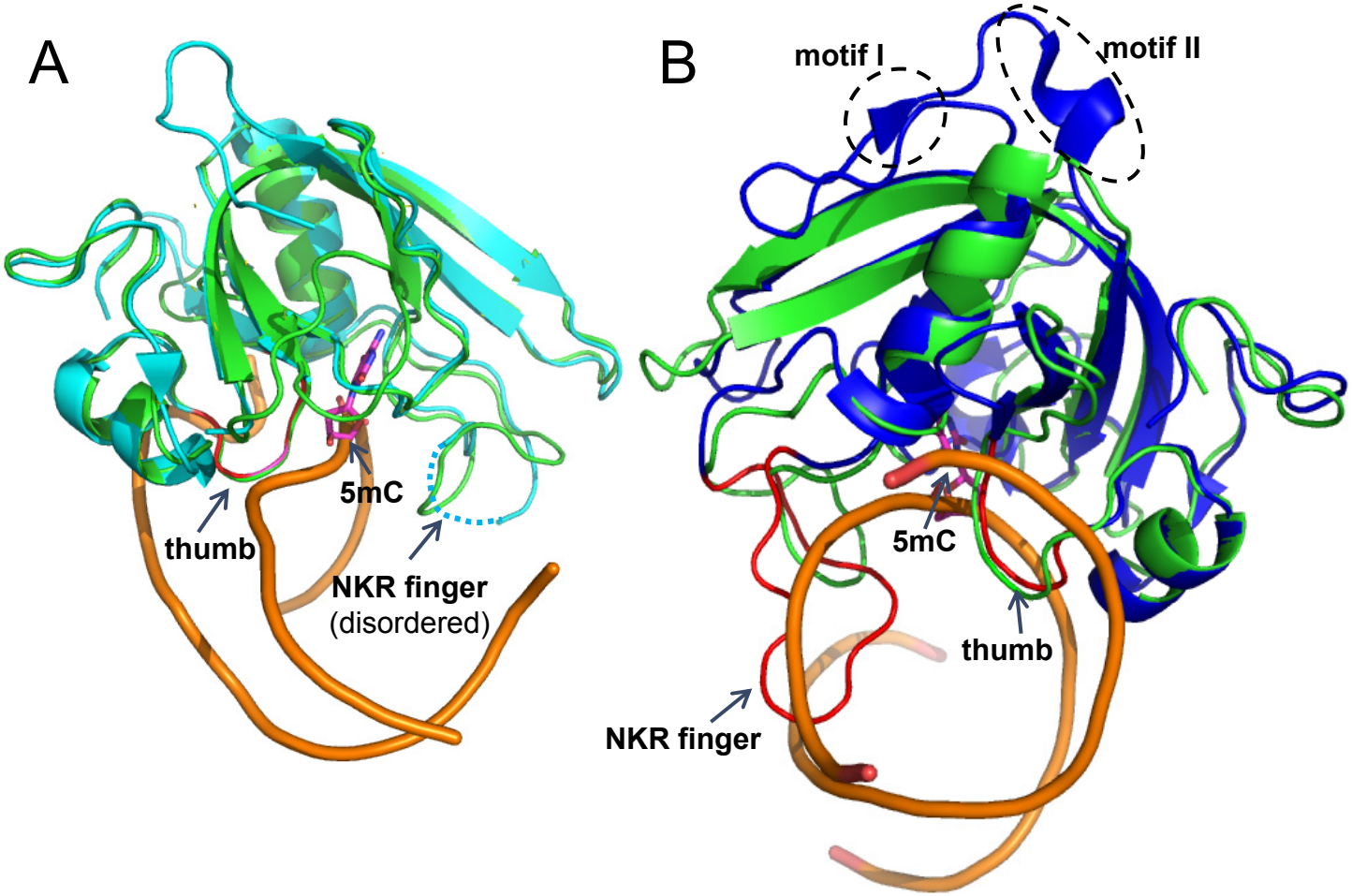


Figure S2. Superposition of SRA_{5333} (Green) to SRA_{SUVH5} (Cyan) and SRA_{UHRF1} (Blue)

A) Superposition of SRA_{5333} (Green) to SRA_{SUVH5} (Cyan). The thumb of SRA_{SUVH5} is colored in red, and the dotted cyan line represents the disordered NKR finger of SRA_{SUVH5}.

B) Superposition of SRA_{5333} (Green) to SRA_{UHRF1} (Blue). The thumb and NKR finger of SRA_{UHRF1} is colored in red. The two dotted ellipse designate motif I and motif II of figure 1B.

SRA_{5333} is more like SRA_{SUVH5} than SRA_{UHRF1} according to the much shorter NKR finger and lacking of motif I, II, and III (motif III not shown).

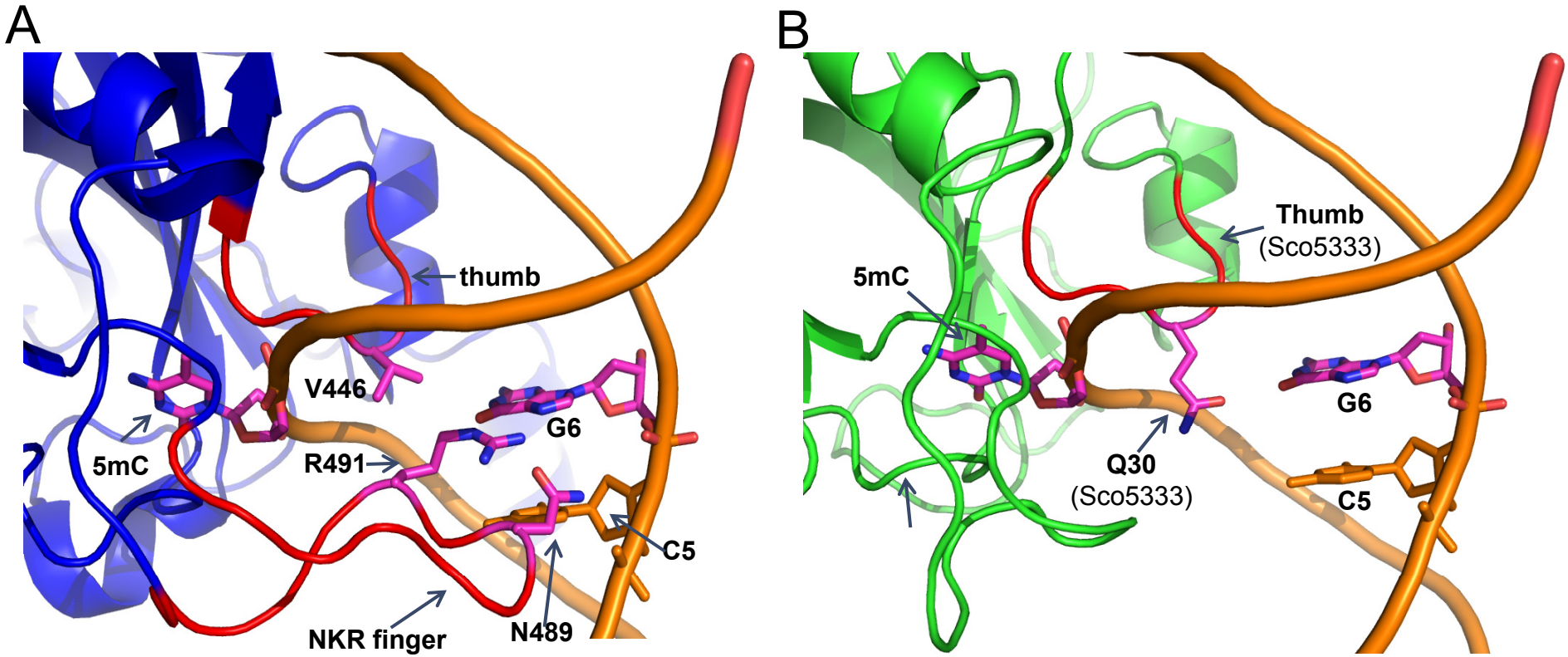


Figure S3. Thumb and NKR finger comparison between SRA_{5333} (green) and SRA_{UHRF1} (blue).

A) Crystal structure of SRA_{UHRF1} (blue) bound to hemi-methylated DNA.
The side chain of R491, which inserts into the DNA duplex, forms intermolecular hydrogen bonds with the Hoogsteen edge of the G6 base. The side chain of N489 prevents potential flipping out of the un-methylated C6 base on the partner strand. V446, which corresponds in sequence to Q392 in the SUVH5 SRA, interacts with the hydrophobic side chain of R491.
B) Predicted tertiary structure of SRA_{5333} (green) after superposition with SRA_{UHRF1} (blue) bound to hemi-methylated DNA.
SRA_{5333} does not contain a long-enough NKR finger to insert into the major groove of the DNA.

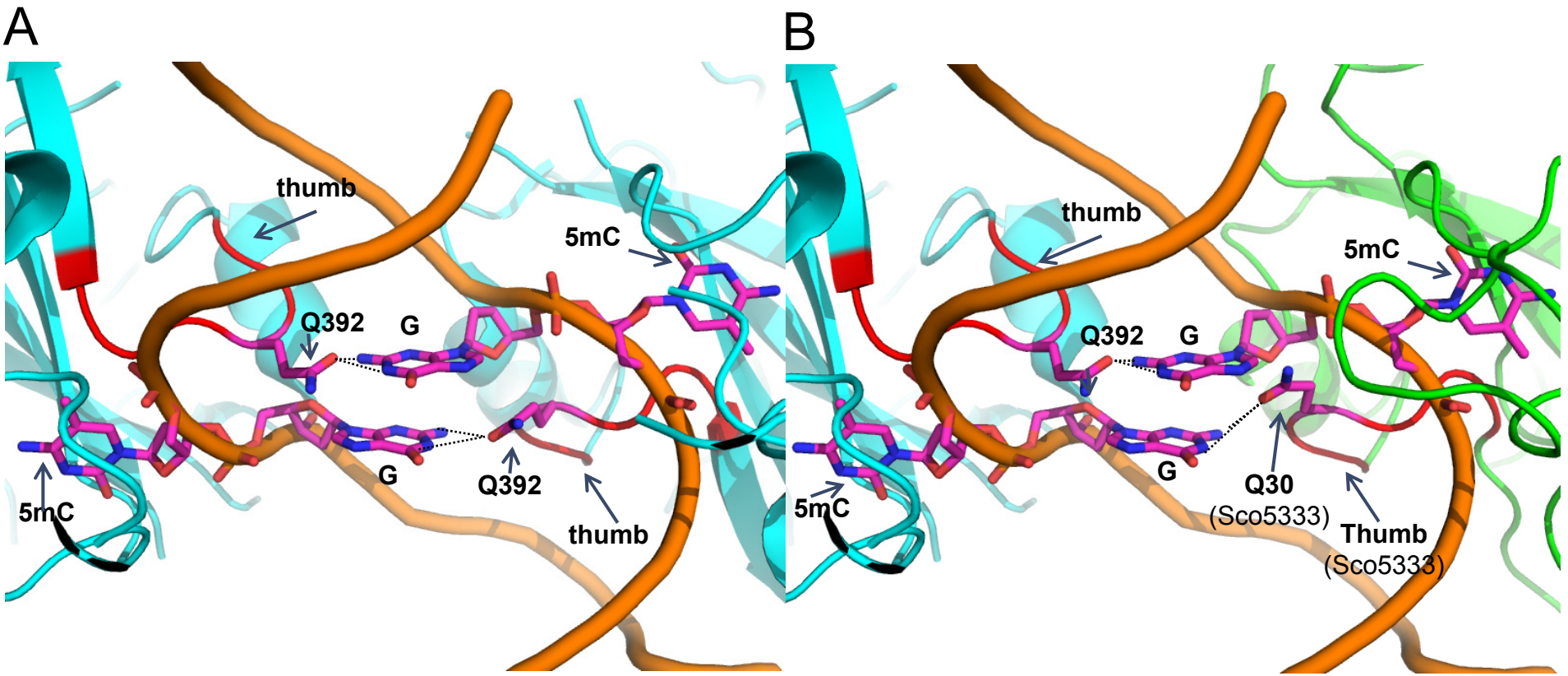
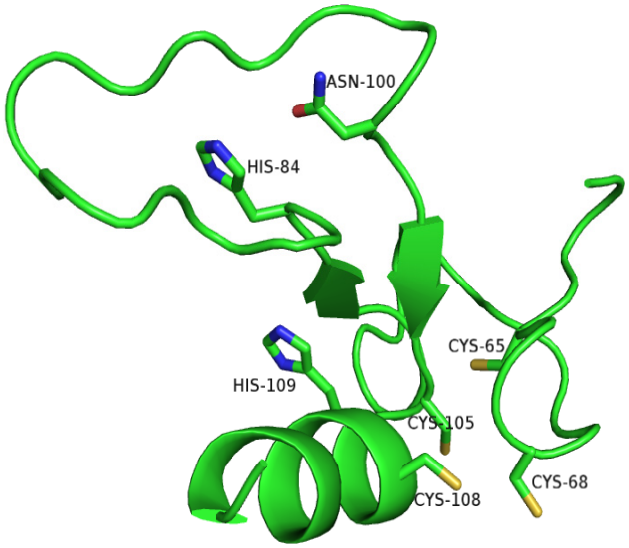


Figure S4. Thumb comparison between SRA_{5333} (green) and SRA_{SUVH5} (cyan).

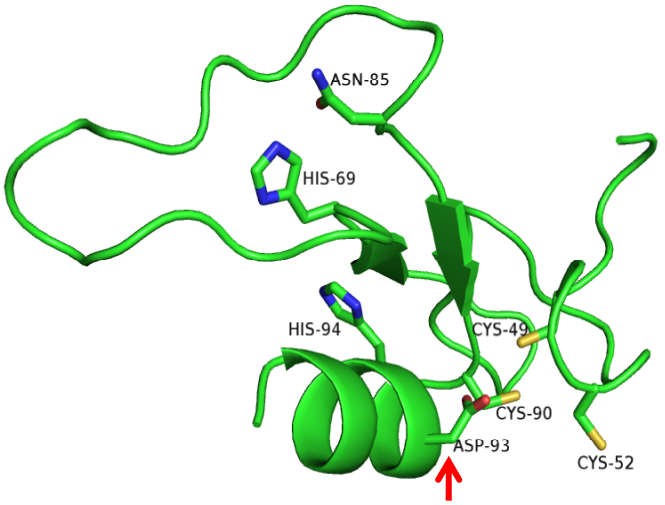
A) Crystal structure of SRA_{SUVH5} (cyan) bound to fully-methylated DNA.
The SUVH5 SRA domain recognizes the minor groove of the DNA whereas the UHRF1 SRA domain mainly recognizes the major groove of the DNA. The base that inserts into the DNA duplex and replaces the flipped out 5mC is Q392 of the thumb loop in the SUVH5 SRA-DNA complex and R491 of the NKR finger in the UHRF1 SRA-DNA complex.
B) Predicted tertiary structure of SRA_{5333} (green) after superposition with SRA_{SUVH5} (cyan) bound to fully-methylated DNA.
SRA_{5333} takes the same strategy to flip out 5mC.

Figure S5. Comparison of predicted tertiary structure of HNH_{5333} and HNH_{Tbis1} to crystal structure of HNH motif of AnaCas9, Gme and T4eVII.

Sco5333_HNH and AnaCas9_HNH as well as Gme_HNH and T4eVII contain a HNH motif hybridized with a CCCC type zinc finger, whereas Tbis1_HNH harbors a ruined CCCD zinc finger, as the red arrow labeled.

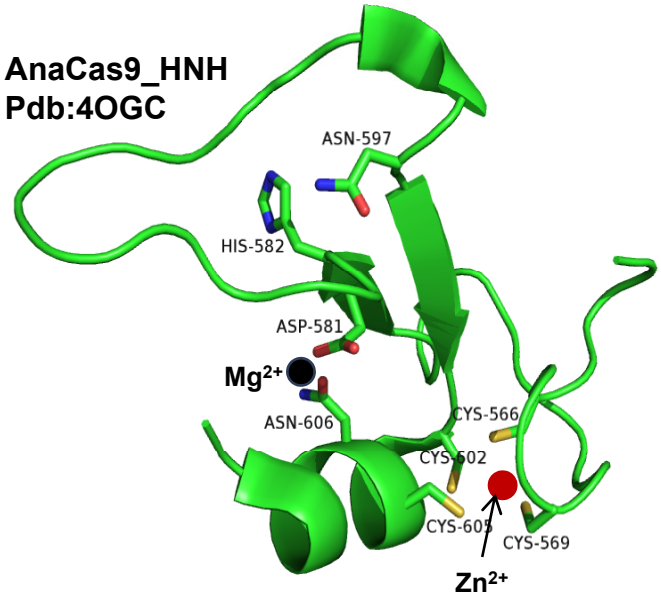


Sco5333_HNH
Predicted using A145 to Q301

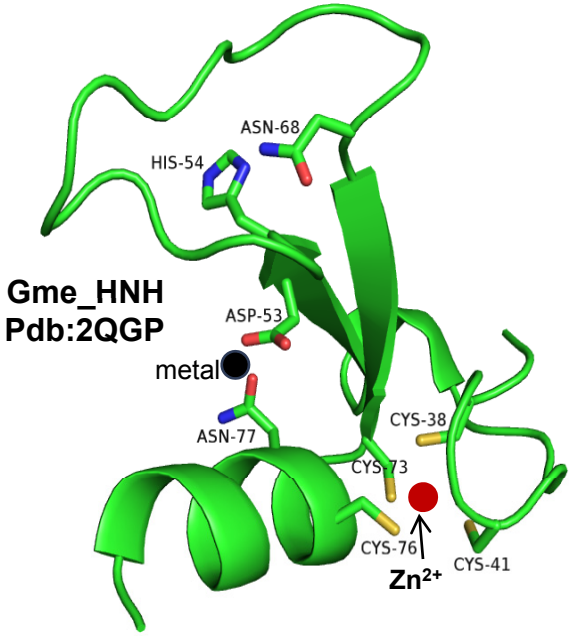


Tbis1_HNH
Predicted using D161 to R309

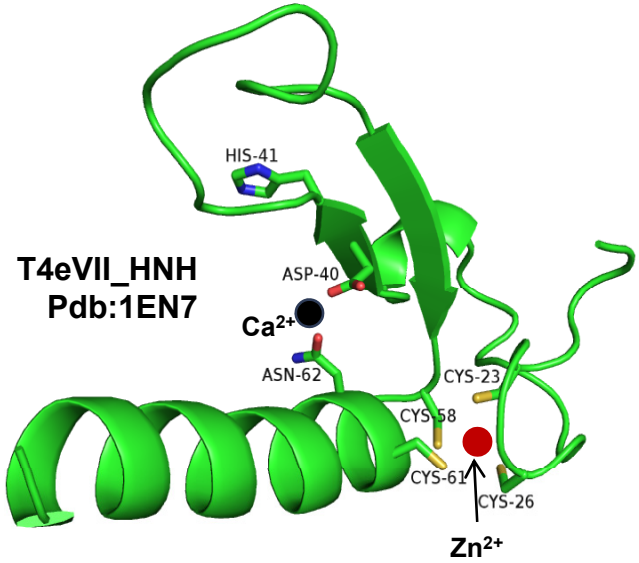
AnaCas9_HNH
Pdb:4OGC



Gme_HNH
Pdb:2QGP



T4eVII_HNH
Pdb:1EN7



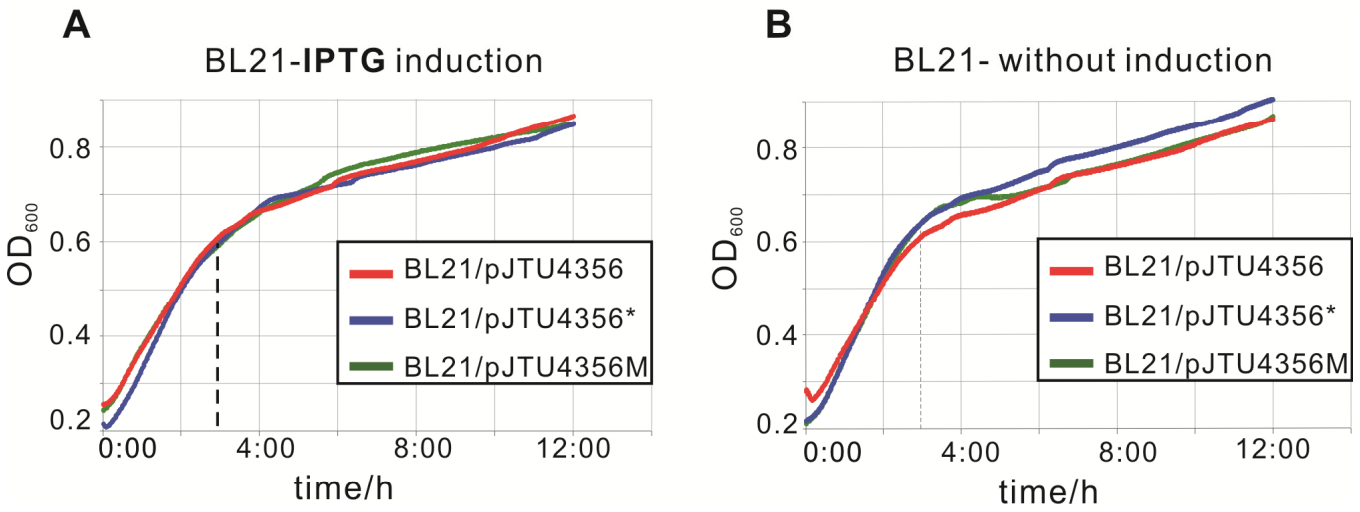


Figure S6. Growth curve measurement of plasmids encoding Sco5333 and its point-mutants in BL21(DE3)plysS.

pJTU4356 encoding Sco5333, pJTU4356* encoding $Sco5333_{E42G}$ and pJTU4356M encoding $Sco5333_{His228Ala,Asn244Ala,His253Ala}$ are introduced into BL21(DE3)plysS respectively and measured the growth curve using microplate reader under 30 ℃. The cell growth shows no distinguishable difference between these three strains whether with IPTG induction or not.

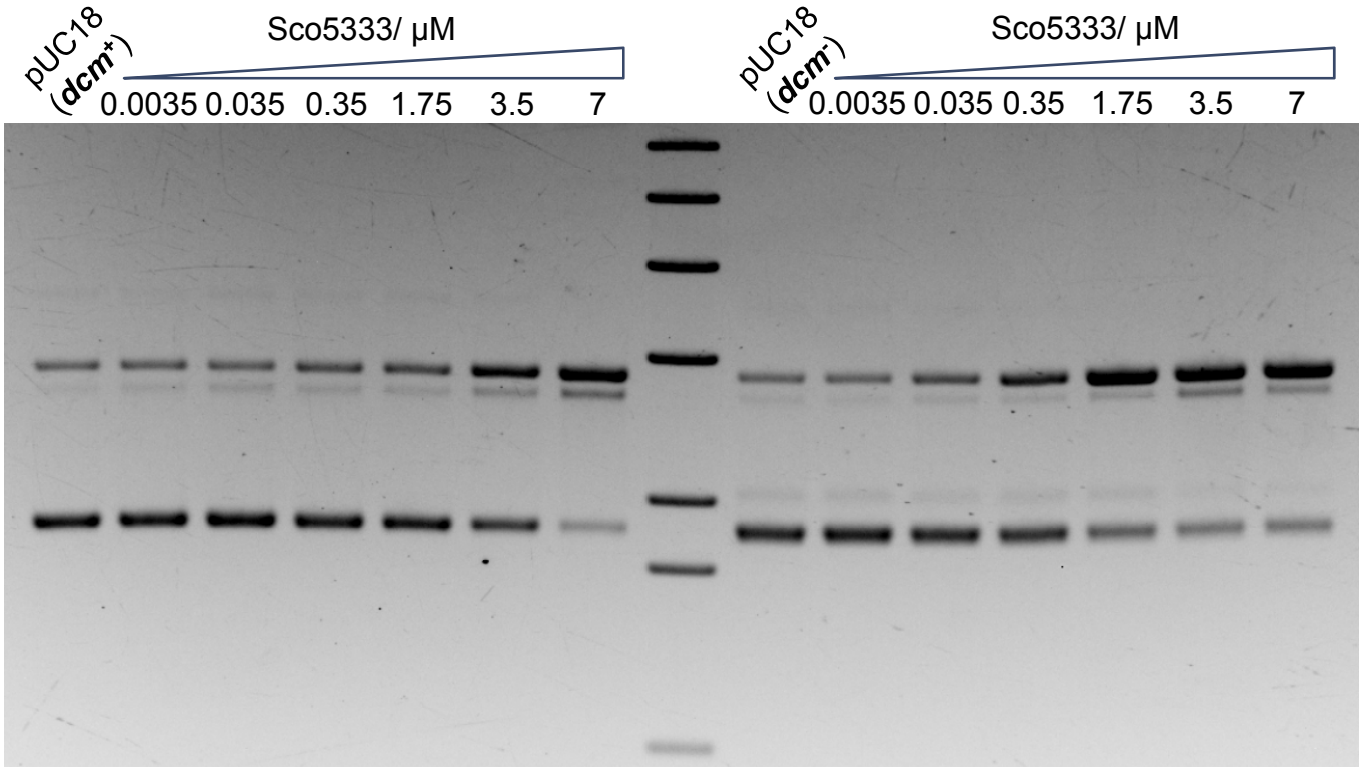


Figure S7. *In vitro* cleavage activity assays of Sco5333 to Dcm-methylated and non-Dcm methylated pUC18 plasmid DNA in presence of 2mM Mg^{2+}.

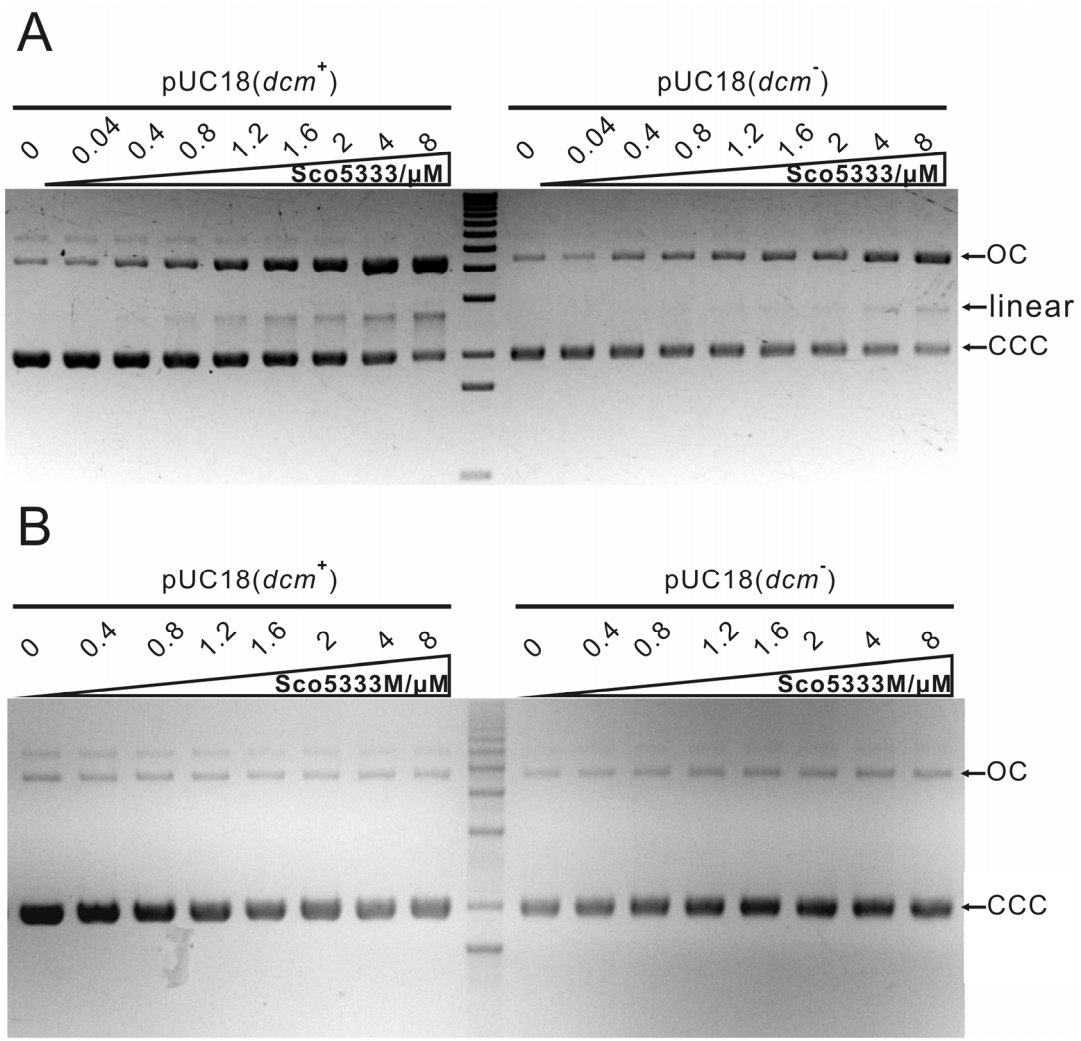


Figure S8. *In vitro* cleavage activity assays of Sco5333 and Sco5333M to Dcm-methylated and non-Dcm methylated pUC18 plasmid DNA in presence of 2mM Mg^{2+}.

A) 0 to 8 μM Sco5333 was added into the reaction system respectively. Weak cleavage activity was observed as concentration of Sco5333 increases.

B) 0 to 8 μM Sco5333M(Sco5333 with triple amino acid changes H228A, N244A and H253A) was added into the reaction system respectively. No cleavage activity was showed to both types of DNA.

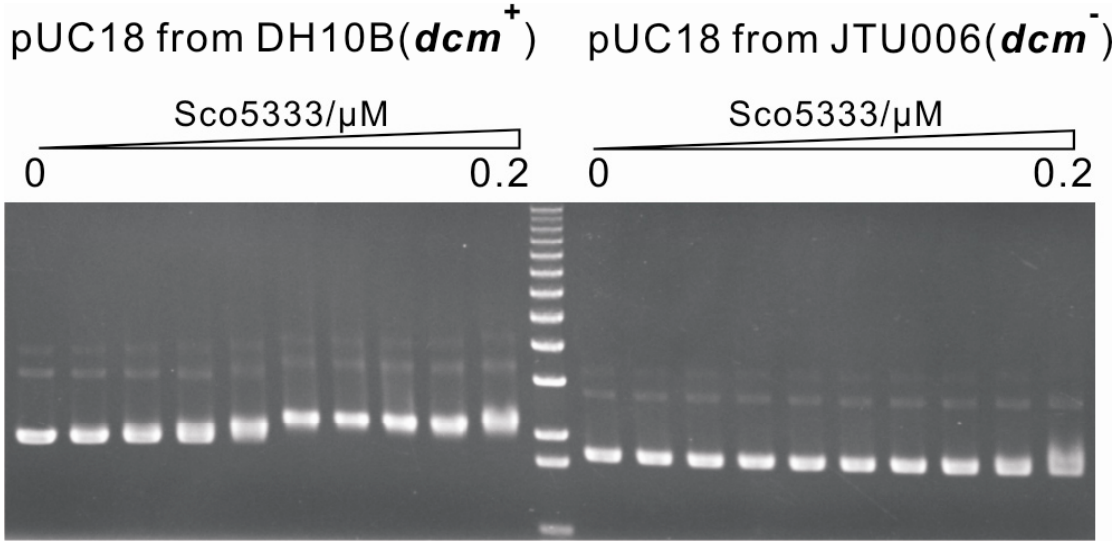


Figure S9. Titration of Sco5333 to pUC18.

pUC18 isolated from DH10B (*dcm*$^+$) or JTU006 (*dcm*$^-$) are performed as the substrate for gradient EMSA of Sco5333(2-folds interval), and Sco5333 shows high specificity to methylated DNA.

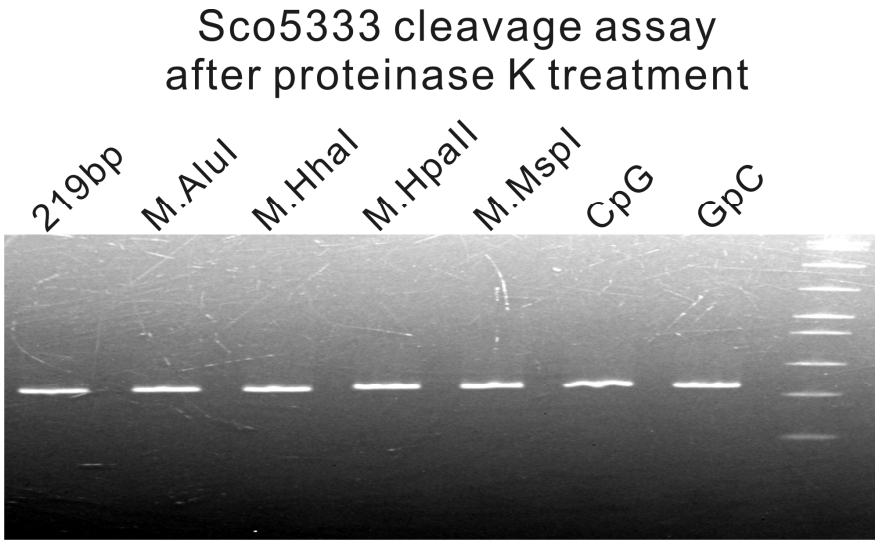


Figure S10. ***In vitro*** **cleavage assay of Sco5333 for 219bp modified by six commercial MTases.**

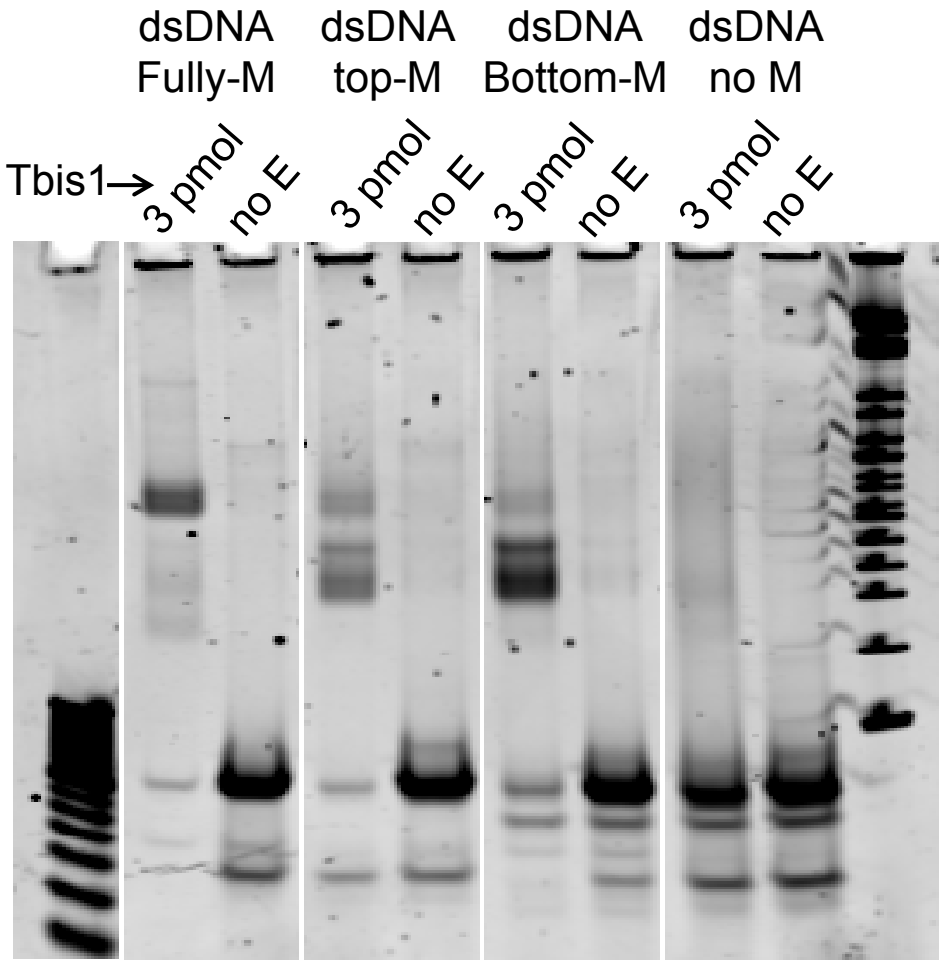


Figure S11. EMSA of Tbis1 to fully-, top-, bottom- and non-methylated oligos.

0.6 pmol of fully-, top-, bottom- and non-methylated 56nt double-strand oligos were mixed with 3 pmol Tbis1 in EMSA system respectively.

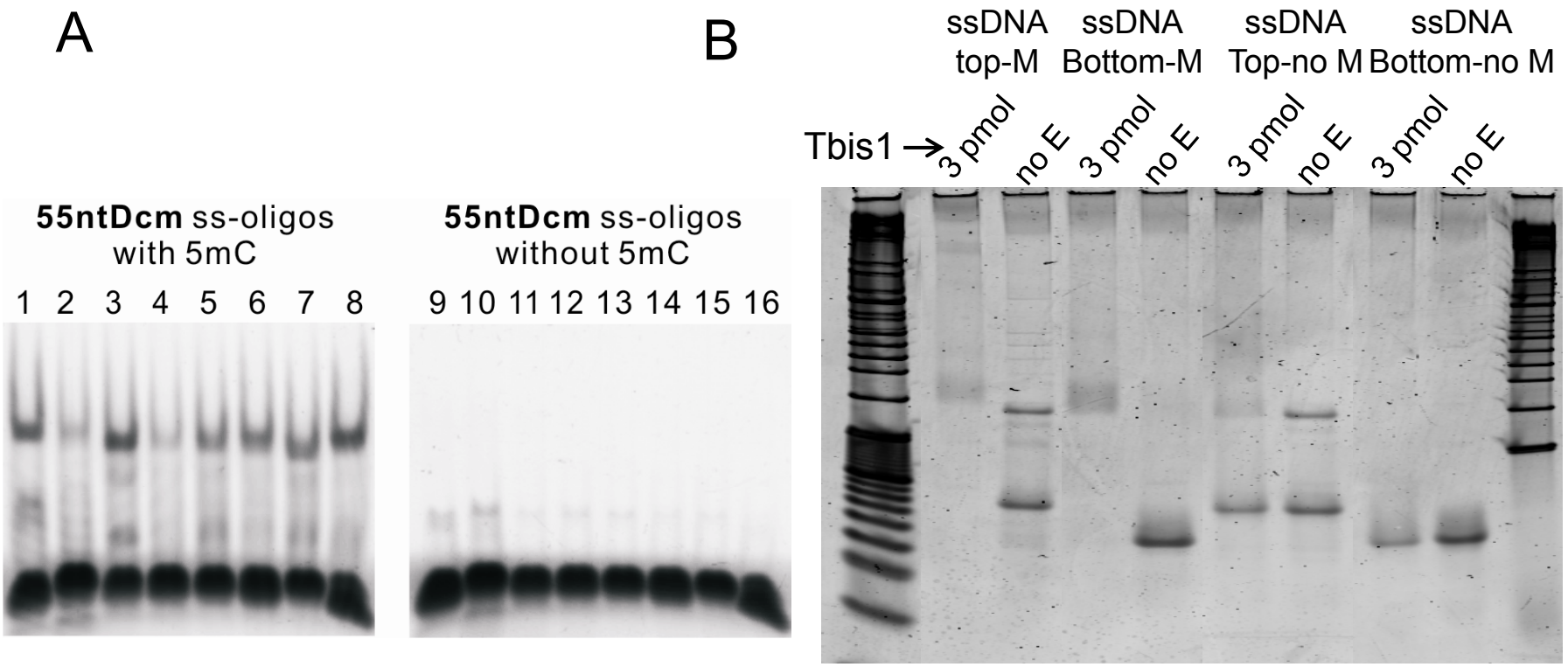


Figure S12. EMSA of Sco5333 and Tbis1 to single-strand DNA.

A) The single-strand 5-FAM labelled 55nt oligos within or without 5mC were performed as substrates for EMSA of Sco5333. Lane 1 to 16 are: 55nt-1, 55nt-5, 55nt-9, 55nt-13, 55nt-17, 55nt-21, 55nt-25, 55nt-29, 55nt-2, 55nt-6, 55nt-10, 55nt-14, 55nt-18, 55nt-22, 55nt-26, 55nt-30**(see table S3)**

B) The single-strand 56nt oligos within or without 5mC were performed as substrates for EMSA of Tbis1. ssDNA top-M, bottom-M, top-no M, bottom-no M were listed in table S3 as 56nt-1, 56nt-3, 56nt-2 and 56nt-4, respectively.

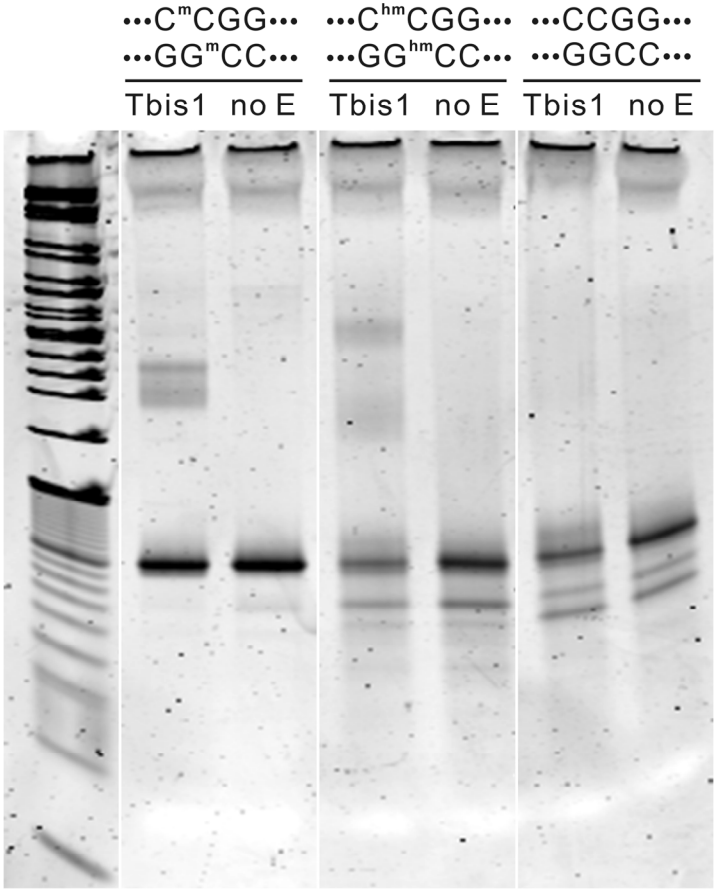


Figure S13. ***In vitro*** **binding affinity assay of Tbis1 to methylated-, hydroxymethylated and non-methylated ds-DNA.**

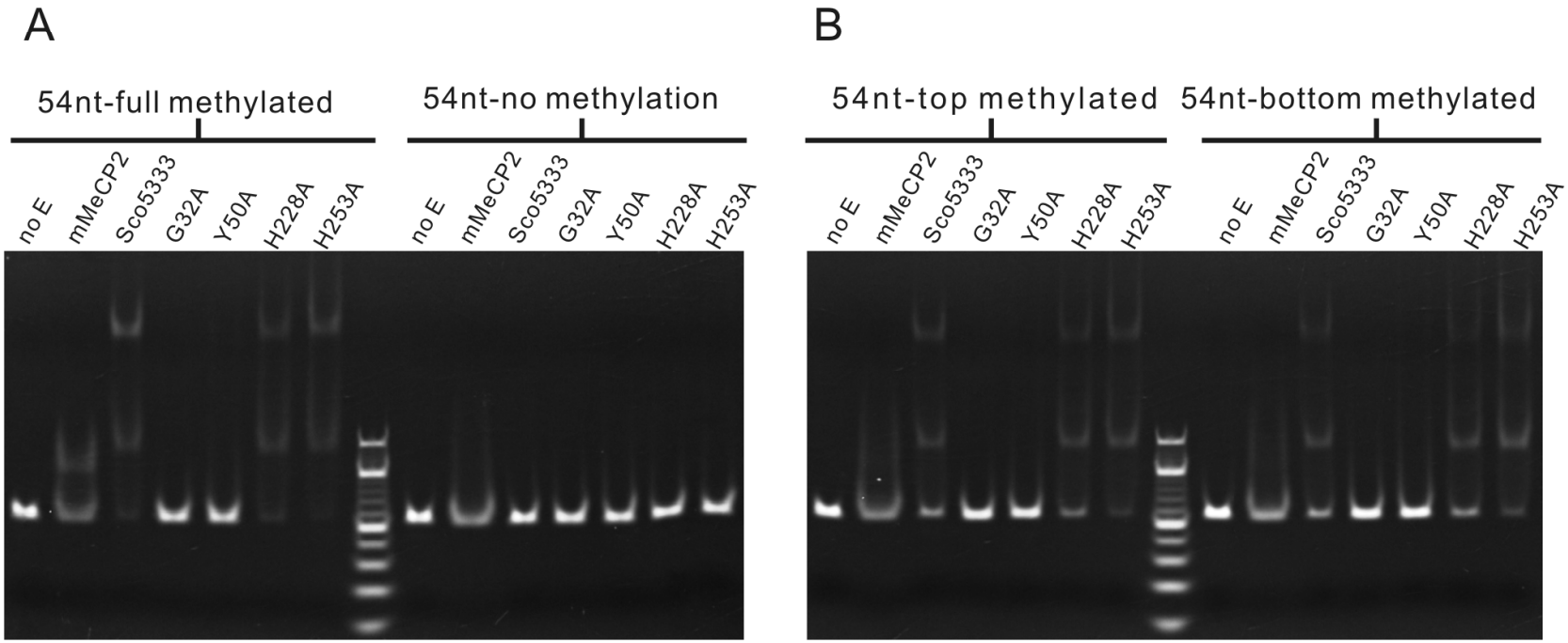


Figure S14. EMSA of Sco5333 and its single point mutants to 54nt fragments.

Wild type Sco5333 and its single amino acid change in SRA domain (G32A and Y50A) as well as amino acid change in HNH motif (H228A and H253A) are tested the binding activity to fully-, top-, bottom- and non methylated 54nt fragment which is centred by a C^{m}CGG pattern. The MBD domain of mouse MeCP2 is tested in parallel as a control.
A) EMSA of MBD_{mMeCP2}, Sco5333, G32A, Y50A, H228A and H253A to fully- and non-methylated 54nt fragment.
B) EMSA of MBD_{mMeCP2}, Sco5333, G32A, Y50A, H228A and H253A to top- and bottom-methylated 54nt fragment.
MeCP2 shows slightly non-specificity to non-methylated 54nt fragment. amino acid changes in SRA domain of Sco5333 lack the DNA binding activity, while amino acid changes in HNH motif retain the binding activity, indicating that its SRA domain that govern the 5mC DNA discrimination and binding.

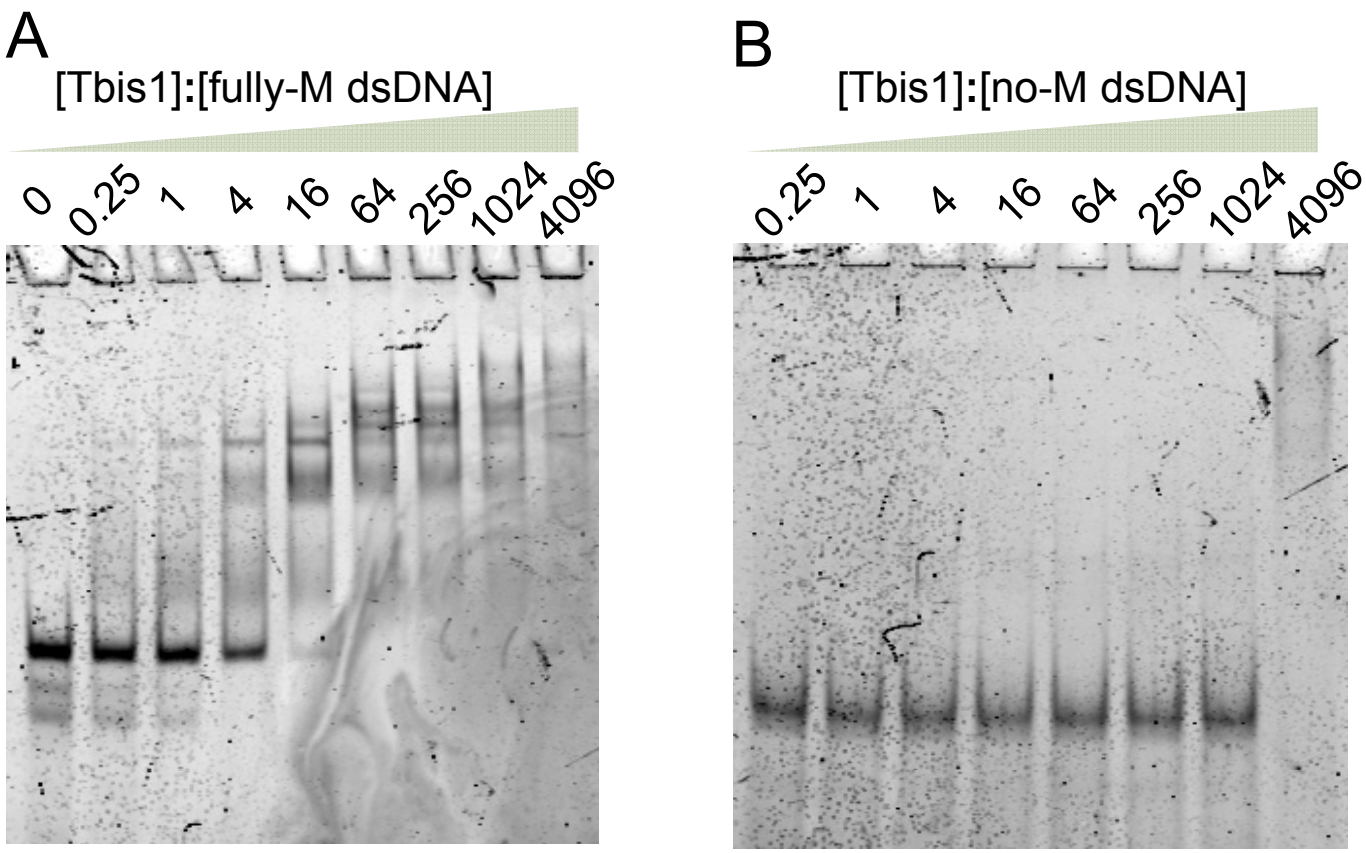


Figure S15. Titration of Tbis1 to fully- and non-methylated DNA

The protein:DNA molar ratio is from 0.25 to 4096. In **A)** the fully-methylated showed clear shift bands at protein:DNA molar ratio 1 and 4, while in **B)** is about 4096 for non-methylated DNA, meaning that full length Tbis1 discriminates fully-methylated DNA from non-methylated DNA by >100 folds.

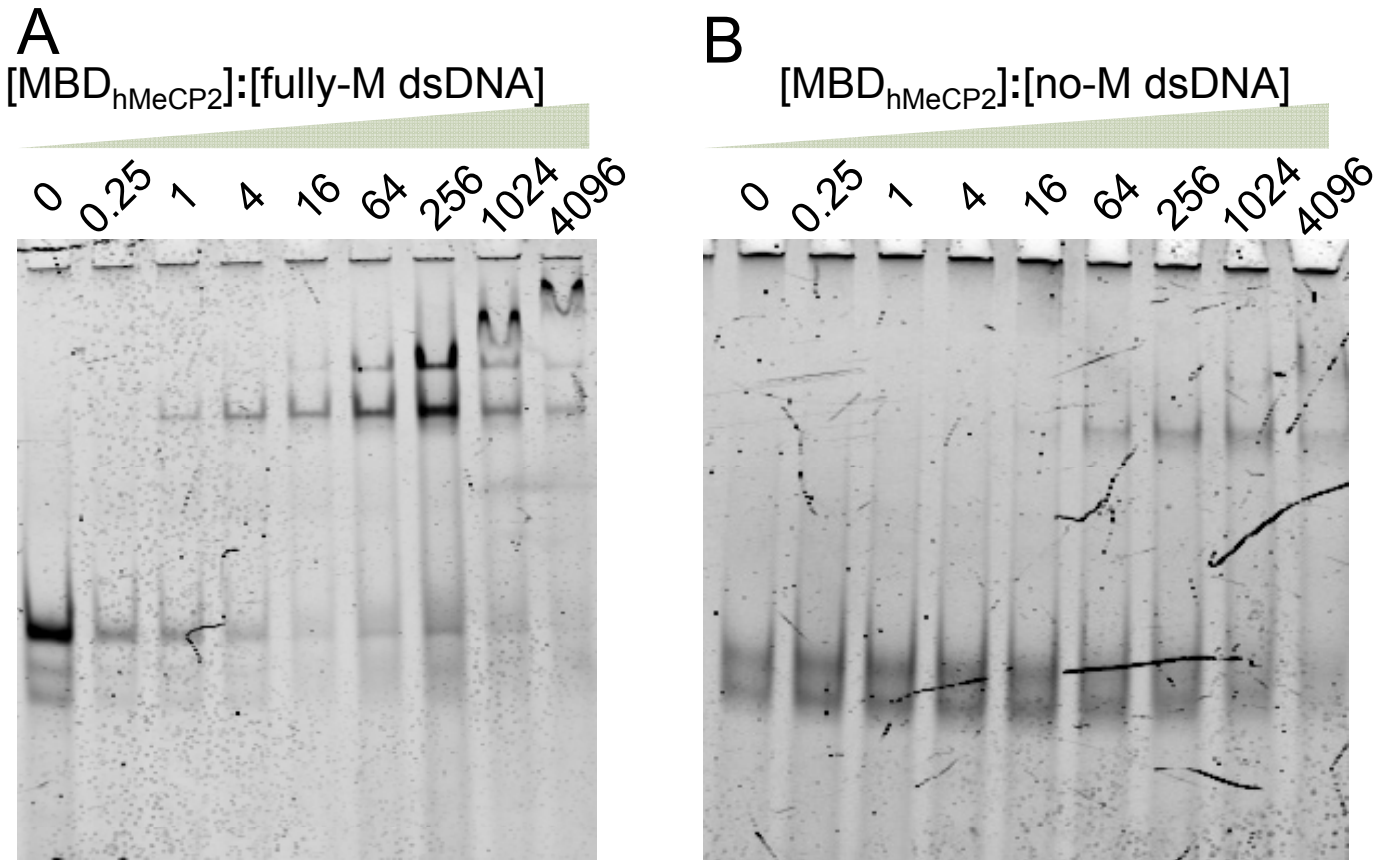


Figure S16. Titration of MBD$_{hMeCP2}$ to fully- and non-methylated DNA

The protein:DNA molar ratio is from 0.25 to 4096. In **A)** the fully-methlayted showed clear shift bands at protein:DNA molar ratio 1, and in **B)** is 64 for non-methylated DNA, which means that MBD$_{hMeCP2}$ discriminates fully-methylated DNA from non-methylated DNA by 64 folds.

In comparison with figure S12, the binding affinity for fully-methylated DNA by Tbis1 is about 2-4 folds higher than by MBD$_{hMeCP2}$.

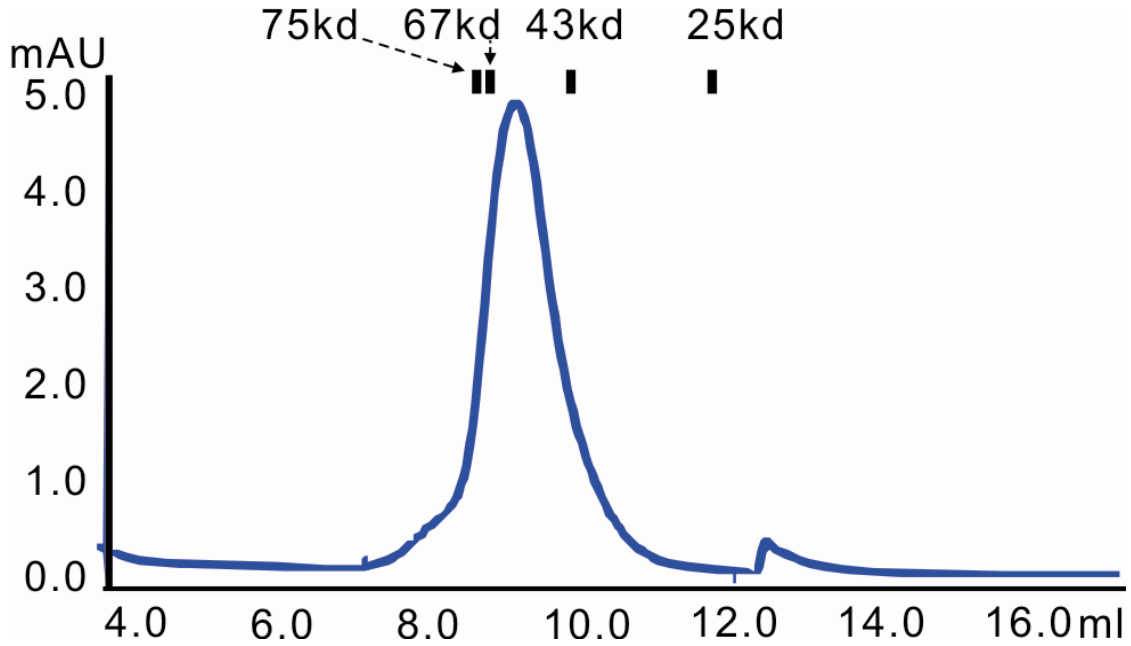


Figure S17. Determination of Sco5333 absolute molecular weight in free state via Gel filtration

100 µl 0.1 mg/ml Sco5333 or 100 µl 0.1 mg/ml Chymotrypsinogen A (25Kd), Ovalbumin (43Kd), Albumin (67Kd), Conalbumin (75Kd) are loaded into the Superdex75 10/300 GL column respectively.
Chromatogram shows that the MW of Sco5333 is ca 2 folds than the monomer MW of Sco5333 which is 34 KD, indicating Sco5333 is a dimer in free state.

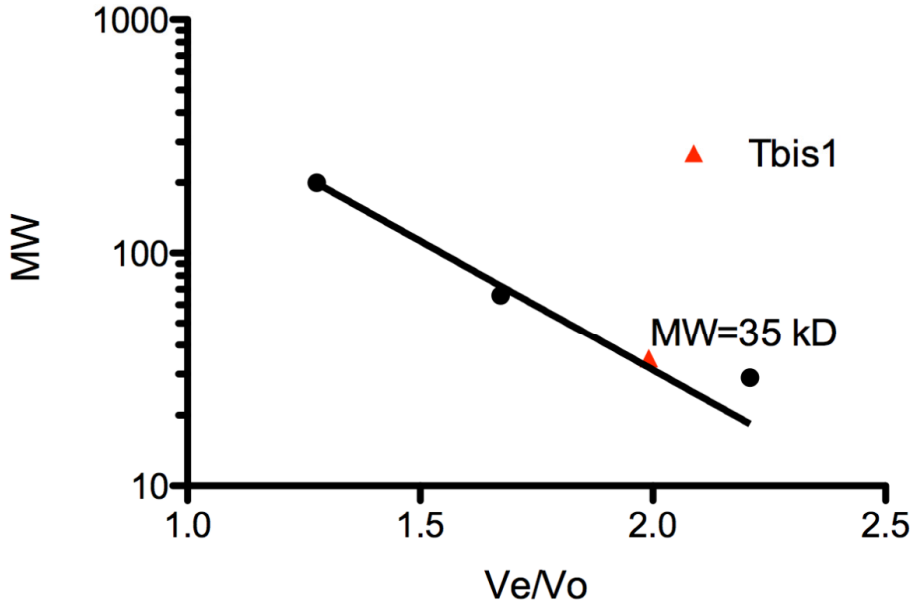


Figure S18. Determination of Tbis1 absolute molecular weight in free state via Gel filtration

Curve simulation shows that the MW of Tbis1 in free state is the same as the monomer MW which is 35 KD, indicating Tbis1 is a monomer in free state.

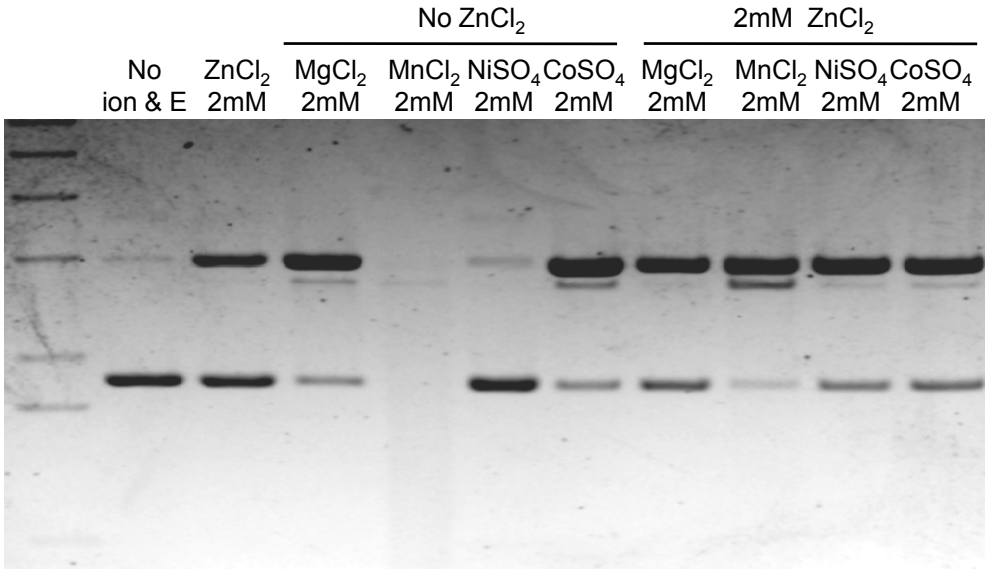


Figure S19. Single amino acid change of Sco5333$_{C252D}$ decreased the suppression of DNA cleavage activity by Zn^{2+}

Sco5333$_{C252D}$ which contains a defect CCCD type Zinc Finger(Figure 1C) displays higher cleavage activity than Sco5333 in presence of Mn^{2+}, and equal molar of Zn^{2+} cannot suppress the activated cleavage activity.

Supplementary Methods

Protein Sequence alignment and structure analysis

The protein sequence of SRA domains was aligned by ClustalW software, and the sequence alignment of HNH motifs was achieved by Cobalt (http://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi). The SRA domain and HNH motif tertiary structure of Sco5333 and Tbis1 were predicted using an on-line server iTASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) respectively.

The secondary structure of SRA domains were drawn based on the crystal structure of SRA_{UHRF1} (pdb:3CLZ) and SRA_{SUVH5} (pdb:3Q0C), and the predicted tertiary structure of $SRA_{Sco5333}$ and SRA_{Tbis1}. Modelling of protein tertiary structure and superposition were accomplished using the PyMOL software.

The measurement of growth curve of *E. coli* hosts expressing Sco5333 and its mutants

pJTU4356* encoding a Sco5333E42G is a spontaneous mutant obtained by introduction of pJTU4356 into the DH10B. The amino acid change of E42G occurred within the SRA domain. *E. coli* BL21/plysS strains harbouring pJTU4356, pJTU4356* or pJTU4356M were cultured overnight and measured for OD_{600} value of the cell density by the Microplate Reader (BioTeK Syngery2). Same amount of cells of each strain was inoculated in 200 μL LB at a ratio of V/V=1:100 in the well of microplate supplied with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol, and were incubated at 30°C in the Reader that shakes once per minute till its OD_{600} value reaches or slightly over 0.6. IPTG was then added at a concentration of 0.4 mM, followed by incubation at 30°C for another 9 hours to induce protein expression.

Electrophoresis mobility shift assay (EMSA) of methylated DNA by Tbis1 and MeCP2

The binding affinity assay for Tbis1 is same as that for DNA cleavage except for exclusion of divalent metal ion in the reaction systems. The reaction mixture was examined by electrophoresis in 1.2% agarose gel for plasmid DNA or native PAGE for oligo DNA. The following 56nt primers were synthesized:

56nt-1: 5'-cggcgtttccgggttccataggctccgctc^{m}Cggcctctgatgaccagggcatcaca-3'

56nt-2: 5'- cggcgtttccgggttccataggctccgctccggcctctgatgaccagggcatcaca -3'

56nt-3: 5'-tgtgatgccctggtcatcagagg^{m}Ccggagcggagcctatggaacccggaaacgccg-3'

56nt-4: 5'-tgtgatgccctggtcatcagaggccggagcggagcctatggaacccggaaacgccg-3'

56nt-5: 5'-cggcgtttccgggttccataggctccgctc^{hm}Cggcctctgatgaccagggcatcaca-3'

56nt-6: 5'-tgtgatgccctggtcatcagaggc^{hm}Cggagcggagcctatggaacccggaaacgccg-3'

Pairs of 56nt-1&3, 1&4, 2&3 and 2&4 were mixed in 1:1 molar ratio and annealed to generate Fully-, Top-, Bottom- and non-methylated double-strand DNA oligos, which were substrates of Tbis1 EMSA assay.

In order to investigate Tbis1 affinity to 5hmC bearing DNA, 56nt-5&6 were mixed in 1:1 molar ratio and annealed to generate Fully-hydroxymethylated double-strand DNA.

For each EMSA reaction, 0.6 pmol double-strand or single-strand oligo DNA and 3pmol Tbis1 were added into the 20 μl binding system mentioned above.

In order to compare the binding affinity of Tbis1 and MBD_{hMeCP2} (h stands for human) to fully-methylated and non-methylated DNA, four 66nt primers were synthesized as below:

66nt-1: 5'-aaaaacggcgtttccgggttccataggctccgcac^{m}Cggtctctgatgaccagggcatcacattttt-FAM-3'

66nt-2: 5'-aaaaacggcgtttccgggttccataggctccgcaccggtctctgatgaccagggcatcacattttt-FAM-3'

66nt-3: 5'-FAM-aaaaatgtgatgccctggtcatcagagac^{m}Cggtgcggagcctatggaacccggaaacgccgttttt-3'

66nt-4: 5'-FAM-aaaaatgtgatgccctggtcatcagagaccggtgcggagcctatggaacccggaaacgccgttttt-3'

Pairs of 66nt-1&3 and 2&4 were mixed in 1:1 molar ratio and annealed to generate Fully- and non methylated double-strand DNA oligos, which were substrates of MBD_{hMeCP2} and Tbis1 EMSA assay. For each EMSA reaction, 0.6 pmol oligo DNA and 3pmol Tbis1 were added into the 20 μl binding system mentioned above.

EMSA of MBD_{mMeCP2}, Sco5333 and its mutants on the 54 nt DNA with one $C^{m}CGG$ site

In order to compare the binding affinity of Sco5333 and MBD_{mMeCP2} (m stands for mouse) to fully-methylated, hemi-methylated and non-methylated DNA, the 54nt DNA oligos (5'-GAAACCCGACAGGACTATAAAGATAC**^{m}C**GGCGTTTCCCCCTGGAAGCTCCCTCGT-3') with replacement of $C^{m}CAGG$ into $C^{m}CGG$ of 55ntDcm1 was synthesized following the 55ntDcm1 annealing strategy. Four pairs of fully-, hemi- or non-methylated duplexes were generated (Table S3, 54nt-1 to 4) and used for EMSA for Sco5333 and $MBD_{mMeCP2.}$ EMSA conditions of Sco5333 to the 54nt fragment were the same as that for 55ntDcm1.

Gel filtration to determine the absolute molecular weight of Sco5333 and Tbis1

For Sco5333 gel filtration, the Superdex 75 10/300 GL column (GE healthcare) was equilibrated with 20mM Tris, pH8.0, 150mM NaCl at 0.4 ml/min through an ÄKTA FPLC (GE Healthcare). 100 µl of 0.1mg/ml Sco5333 and 0.1 mg/ml Chymotrypsinogen A (25 KD), Ovalbumin (43 KD), Albumin (67 KD), Conalbumin (75 KD) were loaded into the column respectively, prior to elution at a flow rate of 0.4ml/min.

For Tbis1, the Superdex 200 10/300GL column (GE healthcare) was equilibrated with 20mM Tris, pH8.0, 150mM NaCl, 1mM EDTA, 1mM DTT and 10% glycerol at 0.5ml/min through an ÄKTA FPLC (GE Healthcare). 250 µl of 0.1mg/ml Tbis1 was loaded into the column prior to the elution at flow rate of 0.5ml/min. Size of the eluted protein peaks were determined by the size exclusion chromatogram of Gel Filtration Markers Kit (Sigma).