The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix

Hideharu Hashimoto[1], John R. Horton[1], Xing Zhang[1], Magnolia Bostick[2], Steven Jacobsen[2,3], and Xiaodong Cheng[1]

[1]Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA

[2]Department of Molecular Cell and Developmental Biology, University of California, Los Angeles, 621 Charles E. Young Dr. South, Los Angeles, CA, 90095, USA

[3]Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, CA, 90095, USA

Figure S1. Mapping the SRA domain within mouse UHRF1
a, Mouse UHRF1 (mUHRF1) is a 782-residue protein containing four recognizable domains including a ubiquitin-like domain, a plant homeodomain (PHD) that may be involved in histone H3 tail binding [1,2], a SET and RING associated (SRA) domain that preferentially binds to hemimethylated CpG sites [3], and a C-terminal Really Interesting New Gene (RING) domain may confer E3 ubiquitin ligase activity [1]. **b**, Proteolytic digestion, mass spectrometry, and deletion analyses identified the SRA domain boundaries of mUHRF1 as residues 419-628. Left panel: SDS-PAGE gel of the purified recombinant mUHRF1 full length and its trypsin digestion products. Purified hexahistidine-SUMO-tagged mUHRF1 full length protein (1.7 μg) in 20 mM Hepes 7.0, 400 mM NaCl, 5% glycerol, and 0.1% 2-mercaptoethanol was treated with 0, 0.5, 5, 50 ng of trypsin for 30 min and separated on a 13% SDS gel. Mass spectrometry determined molecular masses of individual fragments. Each fragment was constructed into a hexahistidine-SUMO tagged fusion protein and expressed in *Escherichia coli*. We also varied the starting and ending amino acids of fragments to reach maximum expression and solubility. Right panel: SDS-PAGE gel of the purified recombinant SRA domain (between the molecular weight markers of 20 and 26.6 kDa) used for crystallization.

References:
1. Citterio, E. et al. Np95 is a histone-binding protein endowed with ubiquitin ligase activity. Mol Cell Biol 24, 2526-35 (2004).
2. Karagianni, P., Amazit, L., Qin, J. & Wong, J. ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. Mol Cell Biol 28, 705-17 (2008).
3. Bostick, M. et al. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317, 1760-4 (2007).

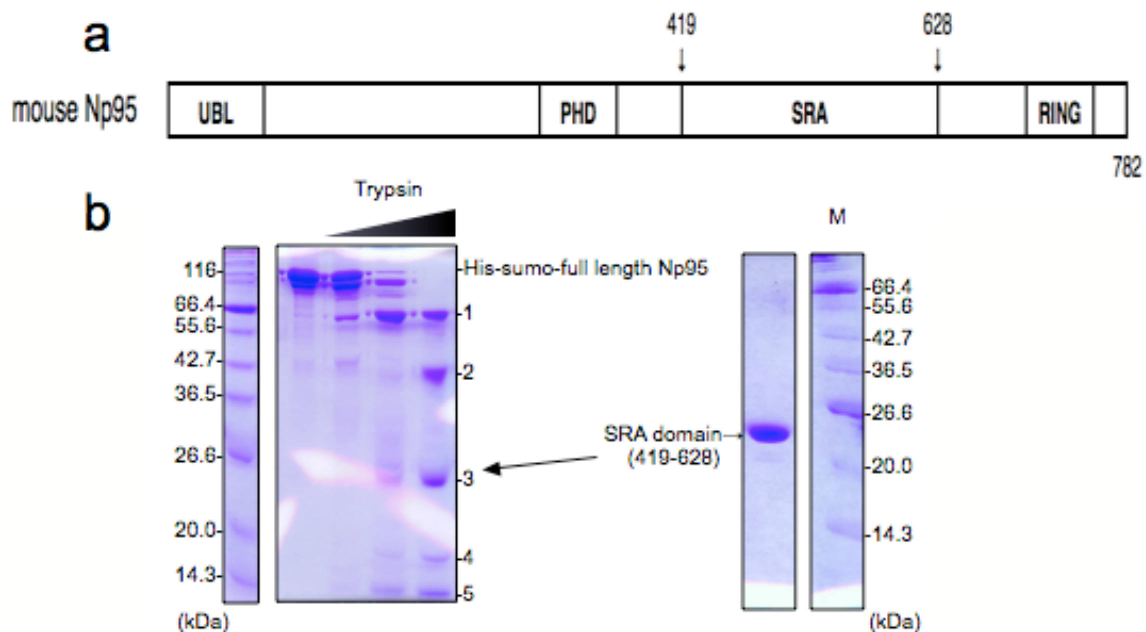


Figure S2. Sequence alignment of mammalian and plant SRA domains
Secondary structural elements are indicated in light blue. Numbering above the sequences corresponds to the mouse ortholog. White-on-black residues are invariant among the blocks of sequences examined, while gray-highlighted positions are conserved (with ≤1 substitution). Positions highlighted in * are responsible for various functions as indicated, and the red circles with a letter P are amino acids that interact with the DNA phosphate backbone. Residues from two parts of the polypeptide form the hydrophobic patch. Mammalian UHRF1 included are Mouse (AAH22167), Rat (NP_001008882), Human (ABQ59043), Chimpanzee (XP_001139655), Cow (NP_001096568), Dog (XP_868468), Chicken (XP_418269), and Zebrafish (NP_998242).

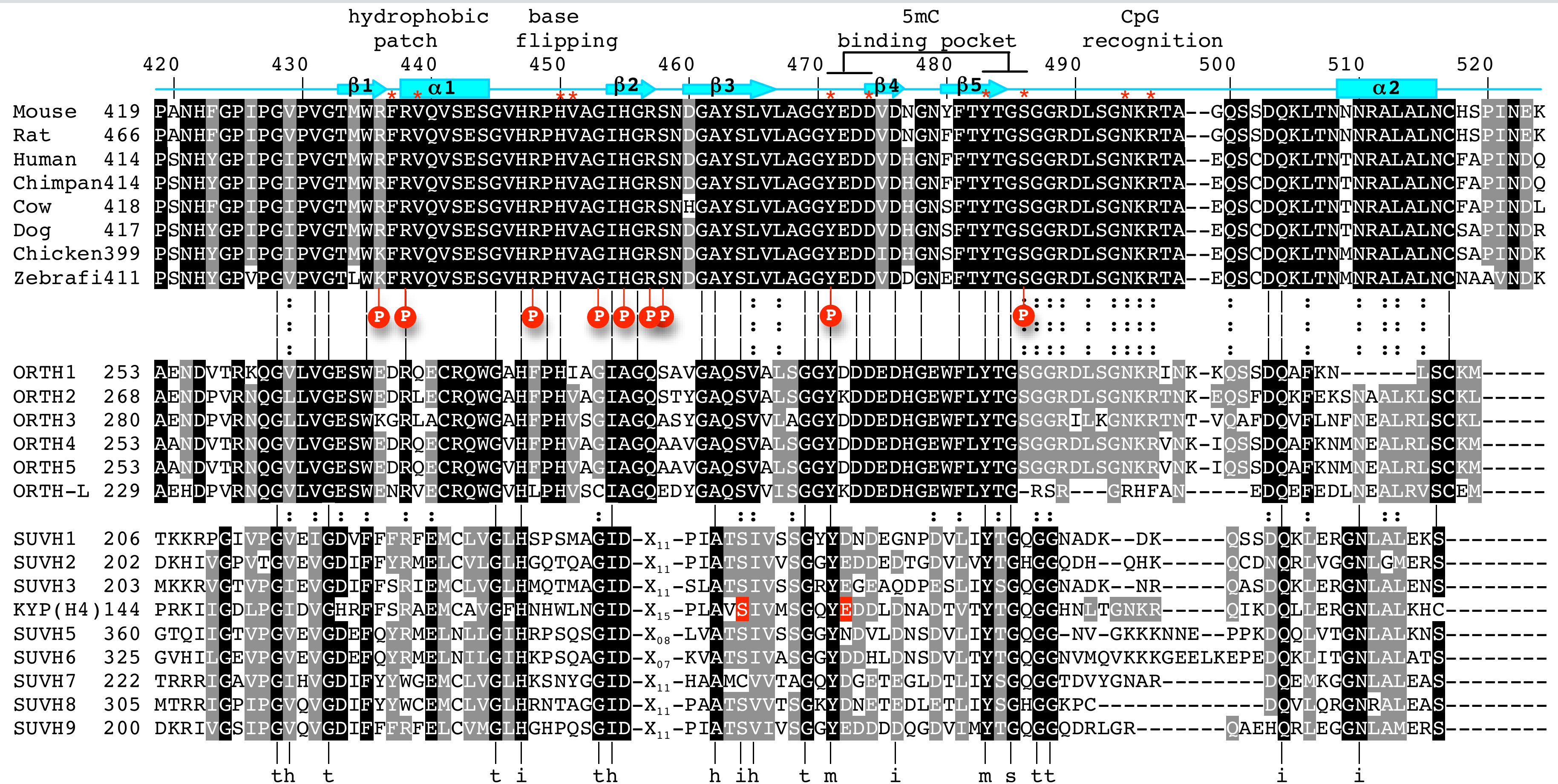
Arabidopsis homologs of UHRF1 included are ORTH1 (At5g39550), ORTH2/VIM1 (At1g57820), ORTH3 (At1g57800), ORTH4 (At1g66040), ORTH5 (At1g66050), and ORTH-like (At4g08590). Also included are the Arabidopsis SRA containing histone methyltransferase genes (SUVH genes) exemplified by the KRYPTONITE (KYP) protein involved in DNA methylation control [4]. X_n represents an insertion of variable (n) amino acids in the SUVH proteins.

The large majority of the invariant amino acids are involved in structural and intramolecular interactions. For example, conserved hydrophobic side chains intercalate with each other to form the hydrophobic core of the molecule. In addition, many of the invariant residues are polar or charged and are critically involved in stabilizing a network of polar interactions involving different parts of molecule. For example, H447 interacts S464 (Fig. 2e). These two residues are critical for stabilizing a network of polar interactions involving (1) the main chain carbonyl oxygen of A452, the amino acid next to the key residue (V451) important for base flipping, (2) R438 interacting with DNA phosphate, and (3) a water-mediated network involving main chain atom of R448, and side chains of H455 and R538. R448 and H455 interact with DNA phosphates (see Fig. 1a). R538 interacts with the main chain atoms of H422, G424, and V446 (not shown).

A network of interactions involves D476…R541…D560…R558 (Fig. 2f). D560 is part of the three consecutive invariant residues YDG initially used to name the domain [5]. Q504 interacts with the main chain atoms of G485, G487, and W580. N510 interacts with the main chain atoms of S486 and K505; E571 bridges between W569 and R581.

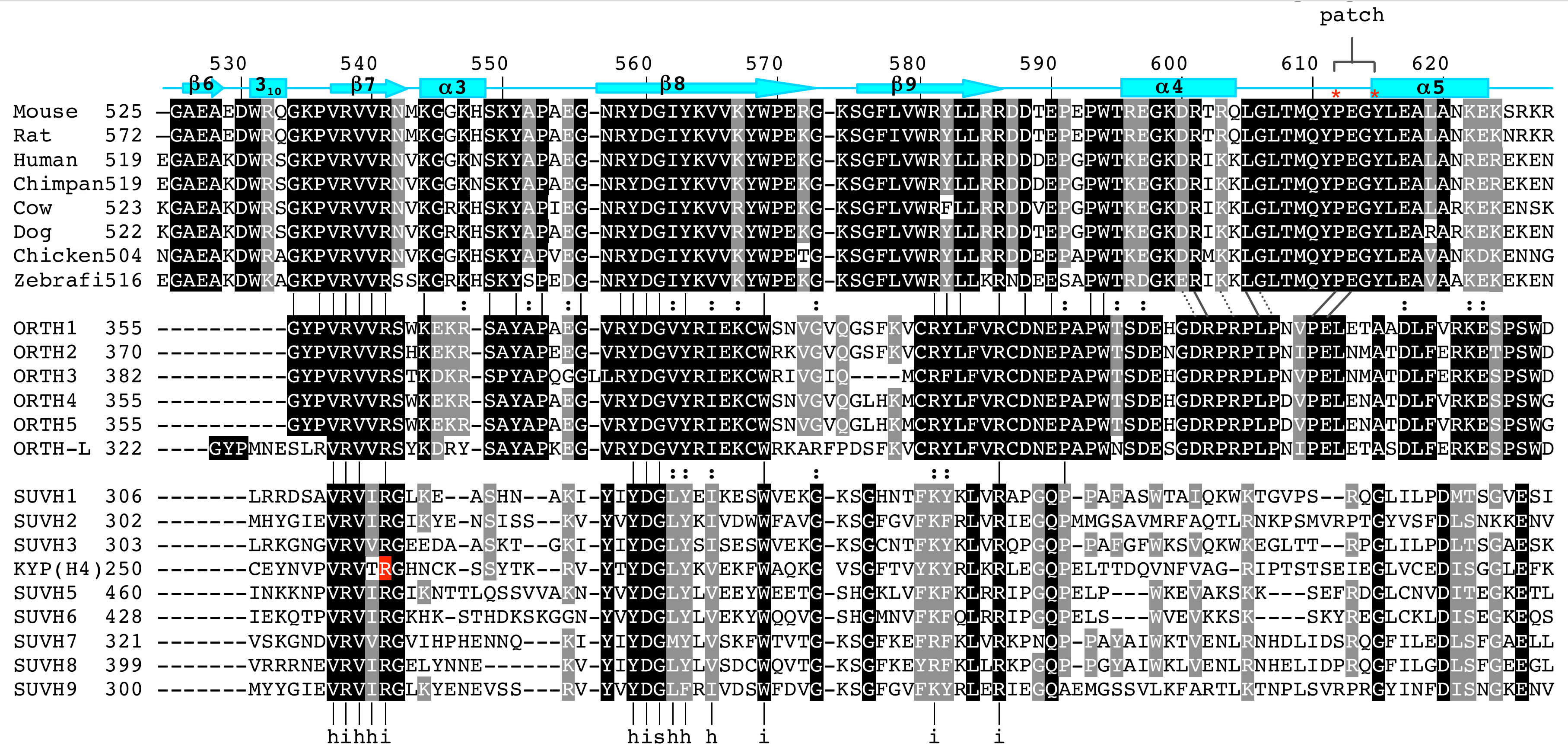
References

4. Johnson, L. M. et al. The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol 17, 379-84 (2007).
5. Baumbusch, L. O. et al. The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Res 29, 4319-33 (2001).



t=structural turn; h=hydrophobic core; i=intromolecular polar interaction; m=5mC binding; s=small space only for Gly

Supplementary Figure S2 (page 1/2)



h=hydrophobic core; i=intromolecular polar interaction; s=small space only for Gly

Supplementary Figure S2 (page 2/2)

Figure S3. mUHRF1 SRA domain binds oligonucleotide with an increased affinity for hemimethylated CpG site

Increasing amounts of GST-mUHRF1 SRA domain (amino acids 393-621) [3, 6] are incubated with a radiolabelled oligonucleotide with a single CG site (5'-A GG GA TG GG GT TT XG TT TT CT CT CT CT C-3' / 5'-G AG AG AG AG AA AA YG AA AC CC CA TC CC T-3') that is either unmethylated (X=Y=C; upper panel), hemi-methylated (X=5mC, Y=C; middle panel), or fully methylated (X=Y=5mC; bottom panel). No protein is present in lane 1, and the amount of DNA in per reaction was 85 picogram for unmethylated, 70.4 picogram for hemimethylated, and 83.5 picogram for fully methylated oligonucleotides. Protein concentration increases in each lane with 0.125 μg (lane 2), 0.25 μg (lane 3), 0.5 μg (lane 4), 1 μg (lane 5), 2 μg (lane 6), 3 μg (lane 7), 4 μg (lane 8), 5 μg (lane 9), and 6 μg (lane 10).

References

3. Bostick, M. et al. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317, 1760-4 (2007).
6. Sharif, J. et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450, 908-12 (2007).

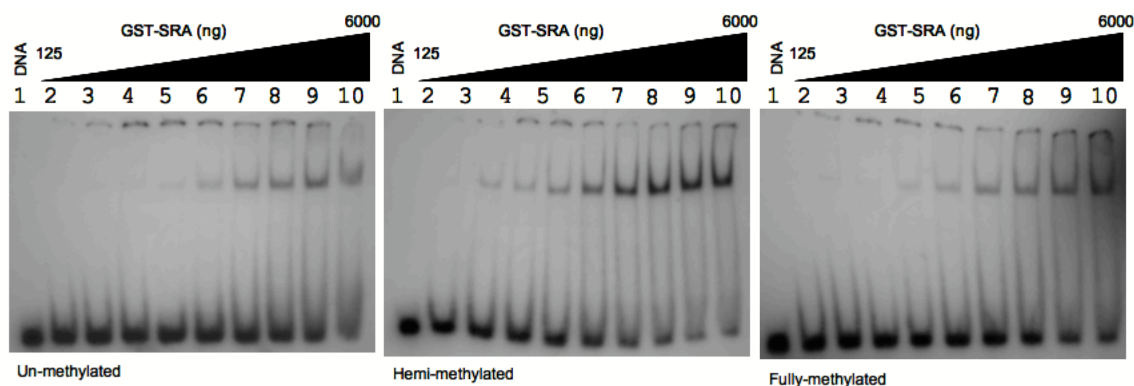


Figure S4. Flipping of Ade at the position 9

a, Ade9, two bases 3’ to the 5mC, adopts an extra-helical conformation, thus destabilizes the DNA duplex. **b**, F437 and Y615 of a surface hydrophobic patch stabilizes the extra-helical Ade9. F437 and Y615 are part of a hydrophobic surface patch formed by two stretches of residues, one from the SRA domain N-terminus (F437 and V439) and the other from its C-terminus (P612 and Y615) (Supplementary Fig. S2).

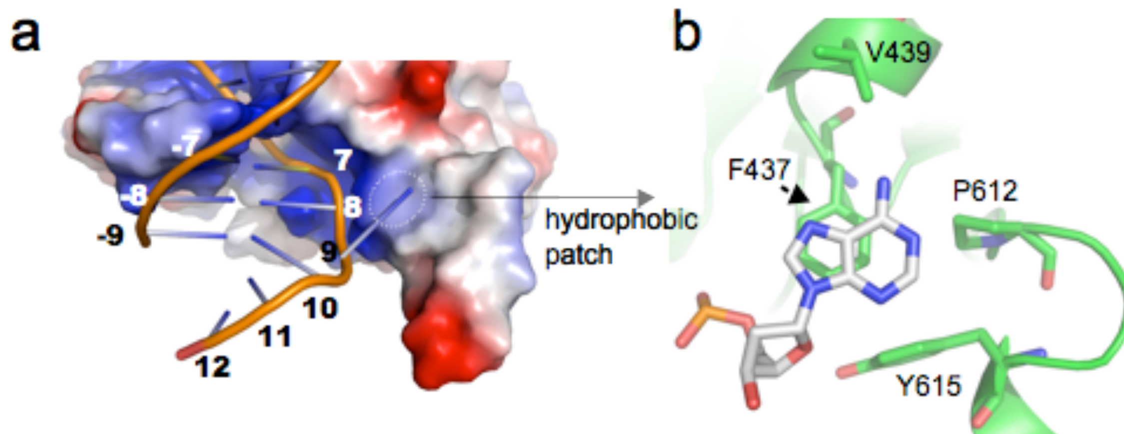


Figure S5. Schematic SRA-DNA interactions in the space group $P2_12_12_1$

The SRA residues that interact with DNA are marked in red; residues in blue belong to the symmetry-related SRA molecule.

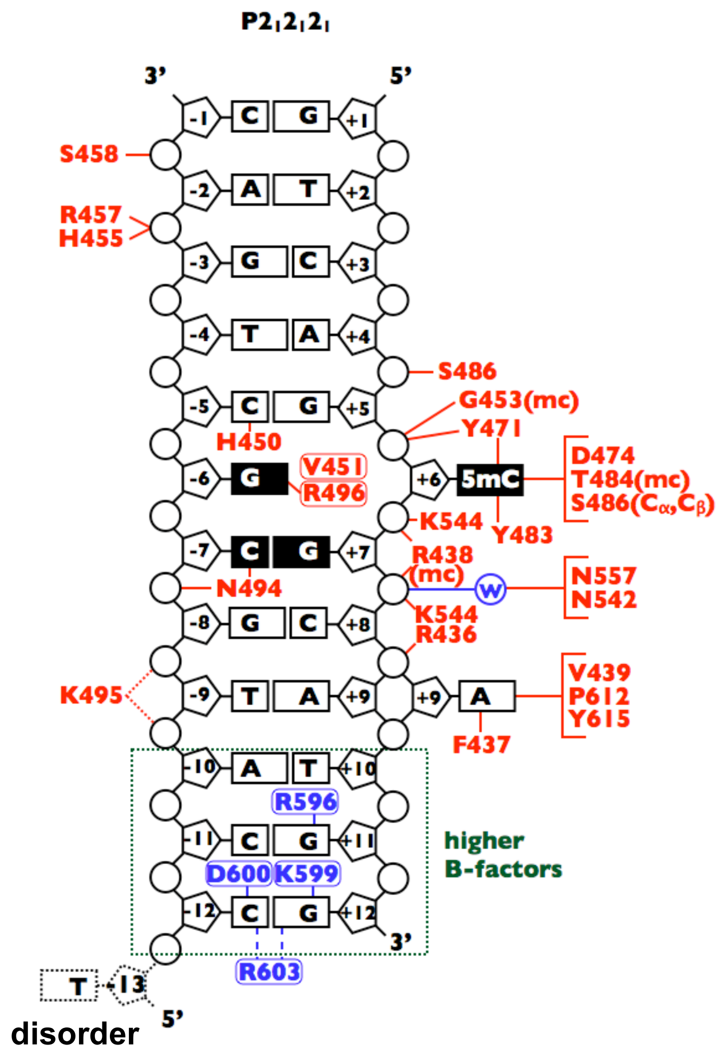


Figure S6. Comparison of the SRA with HhaI methyltransferase

a, DNA structure bound by SRA (left) and by HhaI (right). The intercalating amino acids are shown in each case. **b**, Structures of the two opposite-side DNA-approaching loops of SRA (left) and HhaI (right).

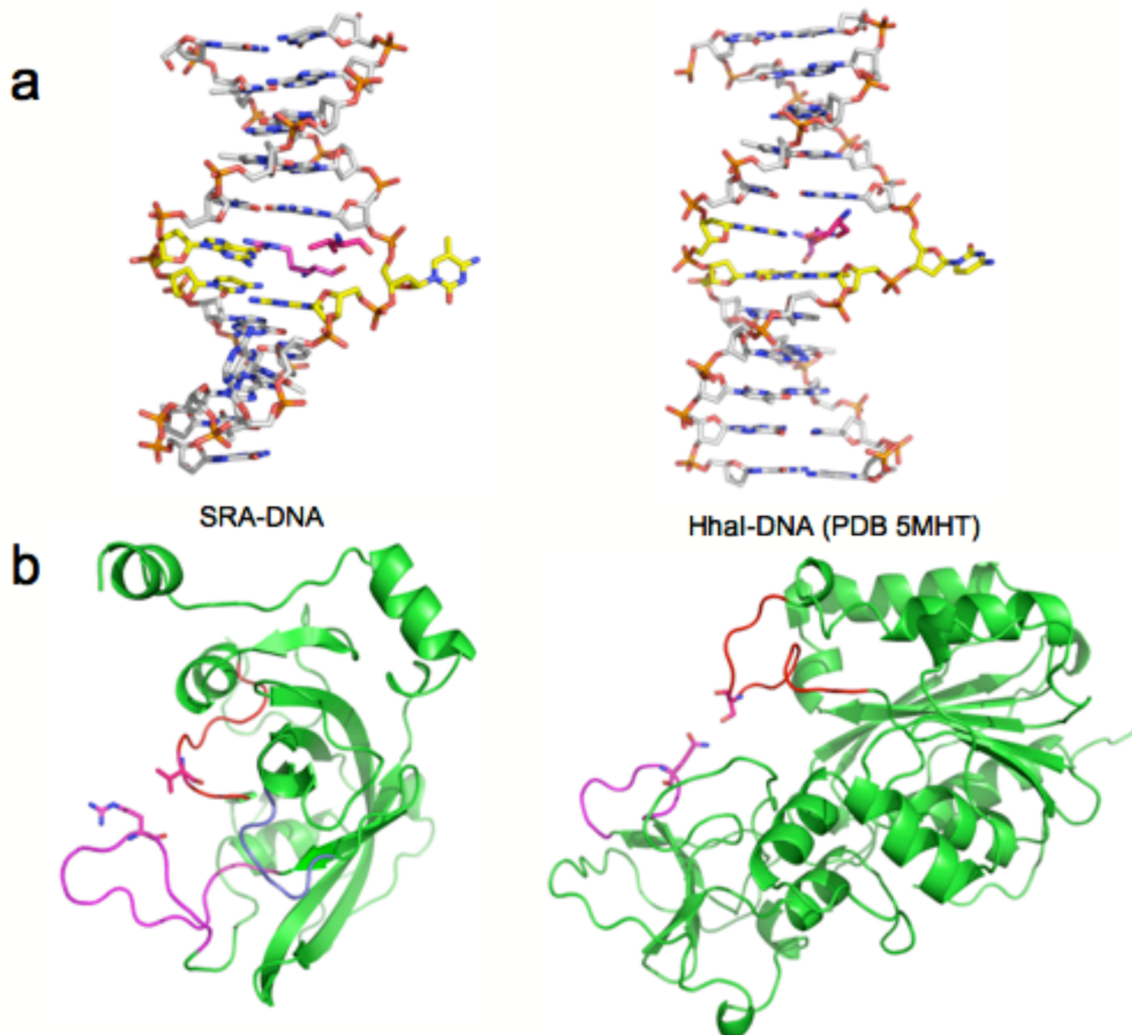
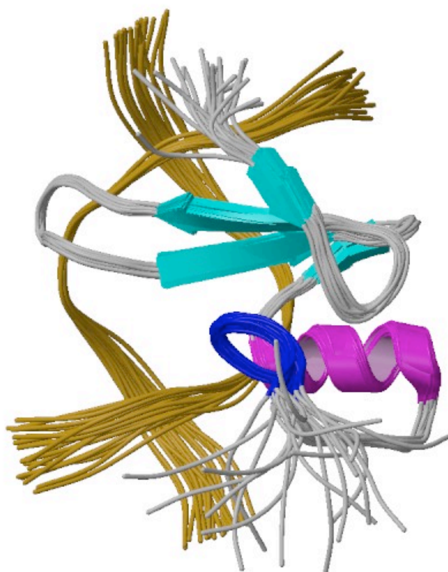


Figure S7. NMR structure of MBD1-DNA shows MBD domain inserts a beta-hairpin through the DNA major groove [7]

The methyl-binding domains of MBD1 [7] and MeCP2 [8], instead of using a base-flipping mechanism, recognize changes in hydration of the major groove of a fully methylated CpG rather than detecting methyl groups directly.

References

7. Ohki, I. et al. Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell 105, 487-97 (2001).
8. Ho, K. L. et al. MeCP2 Binding to DNA Depends upon Hydration at Methyl-CpG. Mol Cell 29, 525-31 (2008).



MBD of MBD1 (PDB 1ig4)

Supplementary Table T1. Data collection and refinement statistics (molecular replacement)

Data collection	Crystal 1	Crystal 2	Crystal 3
DNA	5′-GTCAGMGCATGG-3′ 3′-CAGTCGCGTACCT-5′		5'-AACTGCGCAGTT-3' 3′-TTGACGCGTCAA-5′
Space group	**$P2_12_12_1$**	**$P4_12_12$**	**$P6_122$**
Cell dimensions	(α=β=γ=90°)		(α=β=90°, γ=120°)
a (Å)	62.0	62.0	81.5
b (Å)	69.0	62.0	81.5
c (Å)	93.0	164.2	182.2
Beamline	APS 22-ID (SERCAT)		
Wavelength (Å)	1.00000	1.00000	0.97924
Resolution (Å) *	29.40-2.19 (2.27-2.19)	34.23-1.96 (2.03-1.96)	34.67 – 3.09 (3.20 – 3.09)
R_{sym} or R_{merge} *	0.073 (0.292)	0.076 (0.465)	0.111 (0.392)
I/σI *	21.8 (5.4)	17.4 (2.6)	13.6 (4.5)
Completeness (%) *	91.9 (63.7)	99.5 (98.1)	94.3 (93.7)
Redundancy *	10.4 (6.8)	14.2 (10.3)	7.9 (7.8)
Observed reflections	205,307	337,304	53,278
Unique reflections *	19,727 (1,343)	23,766 (2,276)	6,782 (638)
Refinement			
Resolution (Å)	2.19	1.96	3.09
No. reflections	18,627	22,899	6,424
R_{work} / R_{free}	0.217 / 0.253	0.221 / 0.246	0.232 / 0.291
No. of atoms			
protein	1,623	1,587	1,521
DNA	528	448	486
heterogen	-	8 (2 ethylene glycerol)	-
water	42	97	5
B-factors ($Å^2$)			
protein	69.6	40.6	35.5
DNA	86.3	61.4	107.3
heterogen	-	61.2	40.7
water	62.2	44.7	16.6
R.m.s. deviations			
Bond lengths (Å)	0.006	0.005	0.006
Bond angles (˚)	1.1	1.1	1.3
Dihedral angles (˚)	22.3	22.4	22.9
Improper angles (˚)	0.95	0.94	1.05

* Highest resolution shell is shown in parenthesis.