PNAS
www.pnas.org

Supplementary Information for

Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context biased non-CG DNA methylation

Xueqin Li[1,6,8], C. Jake Harris[2,8], Zhenhui Zhong[2,3], Wei Chen[1], Rui Liu[1], Bei Jia[1], Zonghua Wang[3,5], Sisi Li[4], Steven E. Jacobsen[2,7,9], Jiamu Du[1,6,9]

[1]National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Center for Plant Stress Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201602, China.

[2]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095, USA.

[3]State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou, 350002, China.

[4]Department of Biology, Southern University of Science and Technology of China, Shenzhen, Guangdong 518055, China.

[5]Institute of Oceanography, Minjiang University, Fuzhou, 350108, China.

[6]University of Chinese Academy of Sciences, Beijing 100049, China.

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

[8]Co-first authors.

[9]Co-corresponding authors: Steven E. Jacobsen (jacobsen@ucla.edu), Jiamu Du (jmdu@sibs.ac.cn)

This PDF file includes:

Supplementary text
Figs. S1 to S4
Tables S1 to S4
References for SI reference citations

Supporting Information

SI Materials and Methods

Protein expression and purification

An N-terminal deletion construct of *Arabidopsis thaliana* SUVH6 (residues 265-790) was cloned in to a self-modified pET-SUMO vector and transformed into *E. coli* strain BL21(DE3) RIL. The cell was cultured in LB medium at 37 °C until OD 600 reached 0.6. The cells were then cooled to 20 °C and IPTG was added into the cell culture to a final concentration of 0.25 mM to induce protein expression overnight. The recombinant expressed protein was purified with a nickel affinity column (GE Healthcare). The His-SUMO tag was digested by ulp1 protease and removed by a second step nickel column. The protein was further purified by a Heparin column and a Superdex G200 column (GE Healthcare). The Se-Met labelled protein was expressed in Se-Met containing M9 medium and purified using the same protocol as native protein. All the mutations were generated by a PCR based method and expressed and purified using the same protocol as wild type protein. The functional fragments of KYP (residues 93–624) and SUVH5 (residues 271-794) were cloned, expressed, and purified using the same protocol as SUVH6. The mDNA and the peptide were ordered from Shanghai Generay Biotech Company and Shanghai GL Biochem Company, respectively. The chemicals were ordered from Sigma-Aldrich. The oligos used in this research are listed in **Table S2.**

Crystallization and data collection

The purified SUVH6 protein was concentrated to 10 mg/ml for crystallization. For obtaining cofactor bound state, SUVH6 protein was incubated with SAM at a molar ratio of 1:3 at 4 °C for 1 hour and then used for crystal screening. The Se-Met labelled SUVH6-SAM was crystallized under two different conditions. One is 0.2 M potassium sodium tartrate and 20% PEG3350 and the other one is 0.2 M di-Sodium tartrate and 20% PEG3350. For obtaining the methylated DNA bound form of SUVH6, the SUVH6 protein was incubated with various mCHG DNA with molar ratio of 1:1.5 at 4 °C for 1 hour. Finally, SUVH6 was crystallized with a 13 bp mCHG DNA with a 5' G/C overhang (**Table S2**) in a condition of 20% ethanol and 10% w/v glycerol. All the crystals were cryo-protected into the reservoir solution supplemented with 15% glycerol and flash cooled into liquid nitrogen. The diffraction data were collected at beamlines BL17U1 and BL19U1 of the National Center for Protein Sciences Shanghai (NCPSS) at the Shanghai Synchrotron Radiation Facility (SSRF) and were processed with HKL2000/3000 suite (1, 2). A summary of the statistics of the diffraction data is listed in **Table S1**.

Structure determination

The structure of Se-SUVH6-SAM complex was determined using single-wavelength anomalous dispersion method as implemented in the program Phenix (3). The model building and structure refinement were carried out using the programs Coot and Phenix (3, 4), respectively. Throughout the refinement, the geometry of the model was monitored using the program Molprobity (5). All the remaining structures were determined using the molecular replacement method

using the program Phenix and refined using the same protocol as Se-SUVH6-SAM complex structure (3). A summary of the statistics of the refinement and structure models is shown in **Table S1.**

Microscale thermophoresis (MST) based binding assay

SUVH proteins were labeled with a Monolith NT Protein Labeling Kit (NanoTemper Technologies) according to the supplied protocol. All the labeled samples were diluted in a buffer of 20 mM HEPES pH 7.5 and 1% Tween-20. MST experiments were performed on a Monolith NT.115 instrument (NanoTemper Technologies) with blue/red filters. The optimal fluorescence is situated between 200 and 1600 units of fluorescence. DNA oligos (**Table S2**) were annealed together and half-and-half diluted in 16 steps with ddH2O, covering the range from 100 μM to 2 nM. After mixing the labeled protein and DNA, the samples were loaded into Monolith standard-treated capillaries (NanoTemper Technologies). Laser power was set to 20% using 30 seconds on-time. All experiments were performed at 25 °C with three repeats for each measurement. Data analyses were performed using the NanoTemper analysis software (NanoTemper Technologies).

***In vitro* histone MTase assay**

The purified proteins were dissolved into 150 mM NaCl, 5 mM DTT, 20 mM Tris pH 7.5. Methylation reactions were initiated by adding 10 μM unmodified H3(1-15) peptide substrate into a 50 μL reaction system of 50 mM glycine pH 9.8, 10 mM DTT, 1 mM SAM, and 0.4 μg protein. For measuring the effects of DNA binding on

enzyme activity, the methylated or unmethylated CHG DNA were added into the reaction system with equal molar concentration with the protein. After incubation at 25 °C for the indicated times, the reaction was stopped by 98 °C heating for 1 min. The reaction mixture was further desalted by ZipTip (Millipore). The MALDI-TOF mass-spectrum analysis was carried out using a similar protocol as described previously (6). In brief, a 5800 MALDI-TOF/TOF mass spectrometer (ABsciex, Foster City, CA) was used to measure the peptide mass. For quantification, triplicate reactions were applied with each spectrum acquired from a total of 1250 laser shots. The resulting mass spectra were analyzed by the Data Explorer software. To make sure the proper comparison, the protein samples used in the same experimental batch were expressed and purified simultaneously with the same protocol and the peptide substrate and chemicals used in the same experimental batch were from the same ordering batches, too.

Plant materials

Arabidopsis thaliana Col-0 WT ecotype and T-DNA mutants used; *KYP* (aka *SUVH4*, AT5G13960, SALK_041474), *SUVH5* (AT2G35160, GK-263C05), *SUVH6* (AT2G22740, SAIL_1244_F04), the *kyp/suvh5/6* triple mutant consists of the same T-DNA mutants as the singles above. Plants were grown under constant light at 22 °C.

ChIP-seq

Starting material for the ChIPs was 4g of 10-12 day old seedling tissue. ChIPs were performed as described previously (7) with minor modifications. Samples were crosslinked *in vitro*, sheared using a Bioruptor Plus (Diagenode), chromatin was then split for each sample and abcam1220 (Abcam) antibody was used for H3K9me2 and abcam1791 (Abcam) antibody was used for H3. Libraries generated with NuGEN Ovation Ultra Low System V2 kit, according to manufacturer's instructions, proceeding from step F (DNA purification) and were sequenced on an Illumina HiSeq 4000 instrument.

BS-seq analysis

Raw bisulfite sequencing reads of WT, *kyp*, *suvh5*, *suvh6*, *kyp/suvh5/6* were downloaded from GEO (8). BS-seq reads were mapped to TAIR10 reference genome by Bismark (v0.18.2) (9) with default settings. Reads with three or more consecutive CHH sites were considered as low converted reads and have been removed. DNA methylation levels were calculated by #C/ (#C + #T). DMC (Differentially Methylated Cytosine) were called by methdiff.py in Bsmap (v2.90) (10) with $p < 0.01$ and a methylation difference cutoff, between mutant and WT, for CG, CHG, and CHH levels, of at least 0.4, 0.2, and 0.1, respectively.

ChIP-seq analysis

ChIP-seq fastq reads were aligned to the TAIR10 reference genome with Bowtie (v1.1.2) (11), allowing only uniquely mapping reads with 0 mismatches. Duplicated reads were removed. ChIP-seq peaks in WT (Fig. 5B) were called by callpeak

function in MACS2 (v2.1.1.) (12). Differential peaks for WT and *kyp/suvh5/6* (Fig. 5C) were compared by bdgdiff function in MACS2. ChIP-seq data metaplots were plotted by deeptools (v2.5.1) (13).

SI References

1. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276:307-326.
2. Wang QS*, et al.* (2015) The macromolecular crystallography beamline of SSRF. *Nucl. Sci. Tech.* 26(1):12-17.
3. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
4. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.
5. Chen VB*, et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12-21.
6. Yang Z*, et al.* (2018) Structure of the Arabidopsis JMJ14-H3K4me3 Complex Provides Insight into the Substrate Specificity of KDM5 Subfamily Histone Demethylases. *Plant Cell* 30(1):167-177.
7. Zhong X*, et al.* (2012) DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. *Nat Struct Mol Biol* 19(9):870-875.
8. Stroud H, Greenberg MV, Feng S, Bernatavichute YV, & Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* 152(1-2):352-364.
9. Krueger F & Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27(11):1571-1572.
10. Xi Y & Li W (2009) BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* 10:232.
11. Langmead B, Trapnell C, Pop M, & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
12. Zhang Y*, et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137.
13. Ramirez F, Dundar F, Diehl S, Gruning BA, & Manke T (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42(Web Server issue):W187-191.

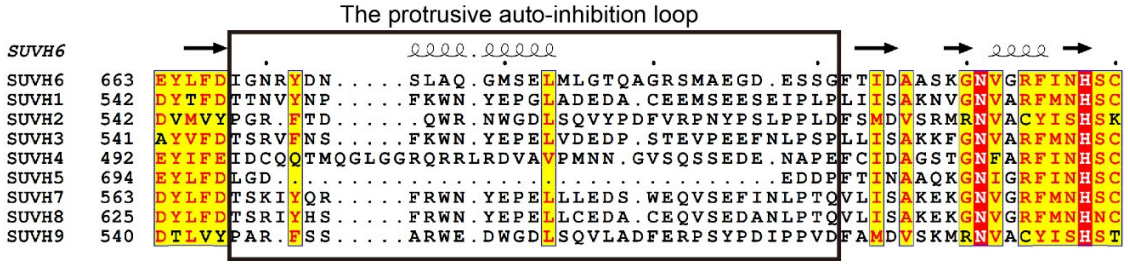

Fig S1. A structure-based sequence alignment of the protrusive auto-inhibition loop of the SUVH6 SET domain. The insertion exists in all the SUVH proteins except SUVH5. The sequences of this region from different SUVH proteins are not conserved. The sequence used for alignment can be found in Arabidopsis Genome Initiative with locus identifiers: SUVH1 (At5g04940), SUVH2 (At2g33290), SUVH3 (At1g73100), SUVH4/KYP (At5g13960), SUVH5 (At2g35160), SUVH6 (At2g22740), SUVH7 (At1g17770), SUVH8 (At2g24740), and SUVH9 (At4g13460).

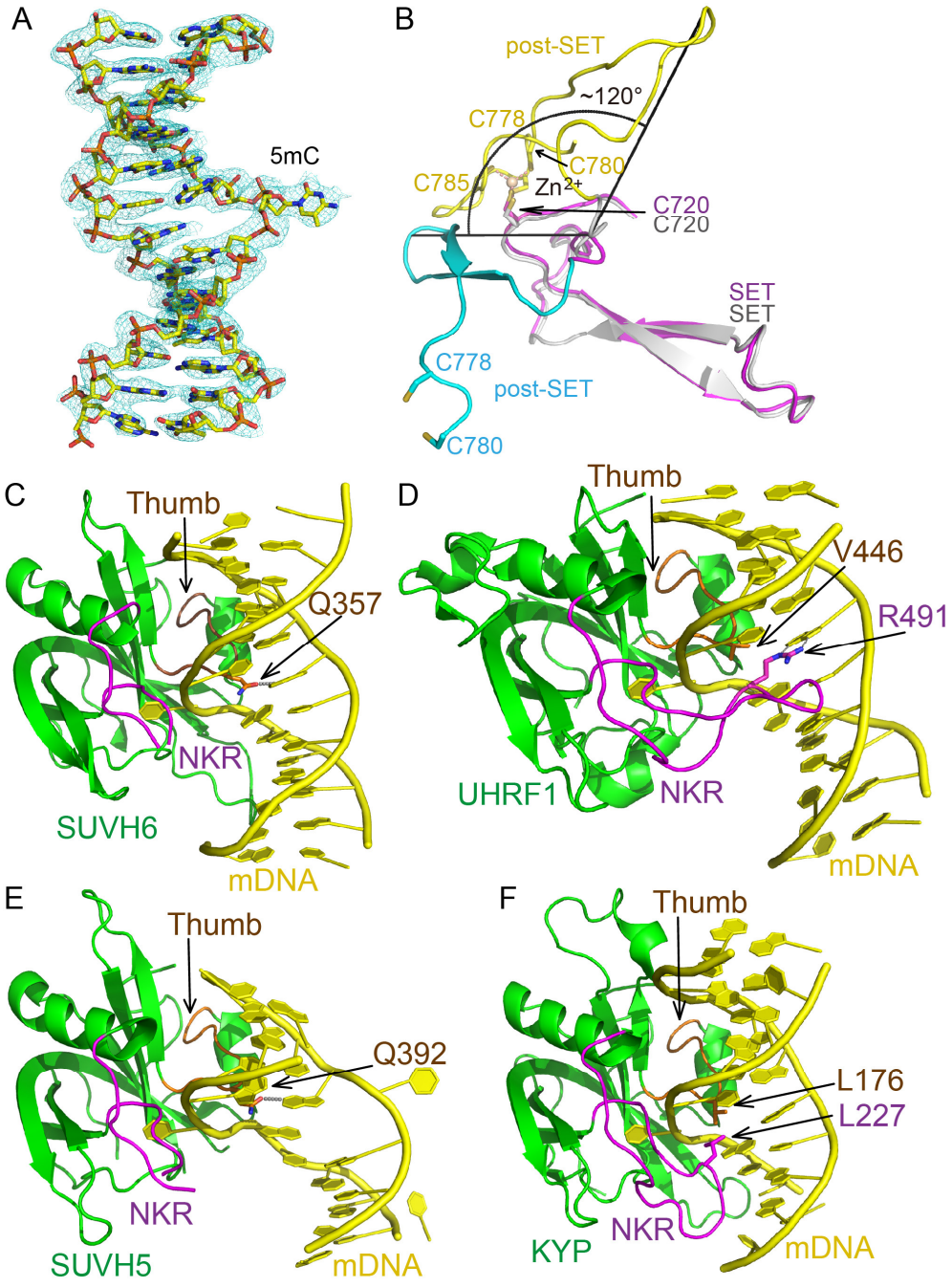

Fig S2. Structure of mDNA bound SUVH6.

(A) The SIGMAA weighted 2Fo-Fc electron density map of the methylated DNA. **(B)** An enlarged view of the post-SET domain from the SUVH6-SAM complex (SET domain in magenta and post-SET domain in yellow) and the SUVH6-mDNA

complex (SET domain in silver and post-SET domain in cyan). The conserved Zn^{2+} coordinating Cys residues from the SET and post-SET domains are highlighted in stick representation. The Zn^{2+} is shown in wheat ball. The post-SET domain undergoes ~ 120° conformational change in the SUVH6-mDNA complex, resulting in a disruption of the Zn^{2+} coordination.

(C-F) A comparison of the mDNA recognition mechanism by SUVH6 (C), UHRF1 (D, PDB code: 3CLZ), SUVH5 (E, PDB code: 3Q0C), and KYP (F, PDB code: 4QEO). The protein and mDNA are colored in green and yellow, respectively. The thumb loop and NKR finger element of the SRA domain are highlighted in brown and magenta, respectively. The residues interacting with the orphaned guanidine or occupying the gap by 5mC flipping are highlighted in stick. While UHRF1 use both the thumb loop and NKR finger residues to occupy the 5mC leaving gap with NKR finger more important, the SUVH5/6/KYP use the thumb loop to achieve the same molecular function, indicating a different mechanism other than the UHRF1.

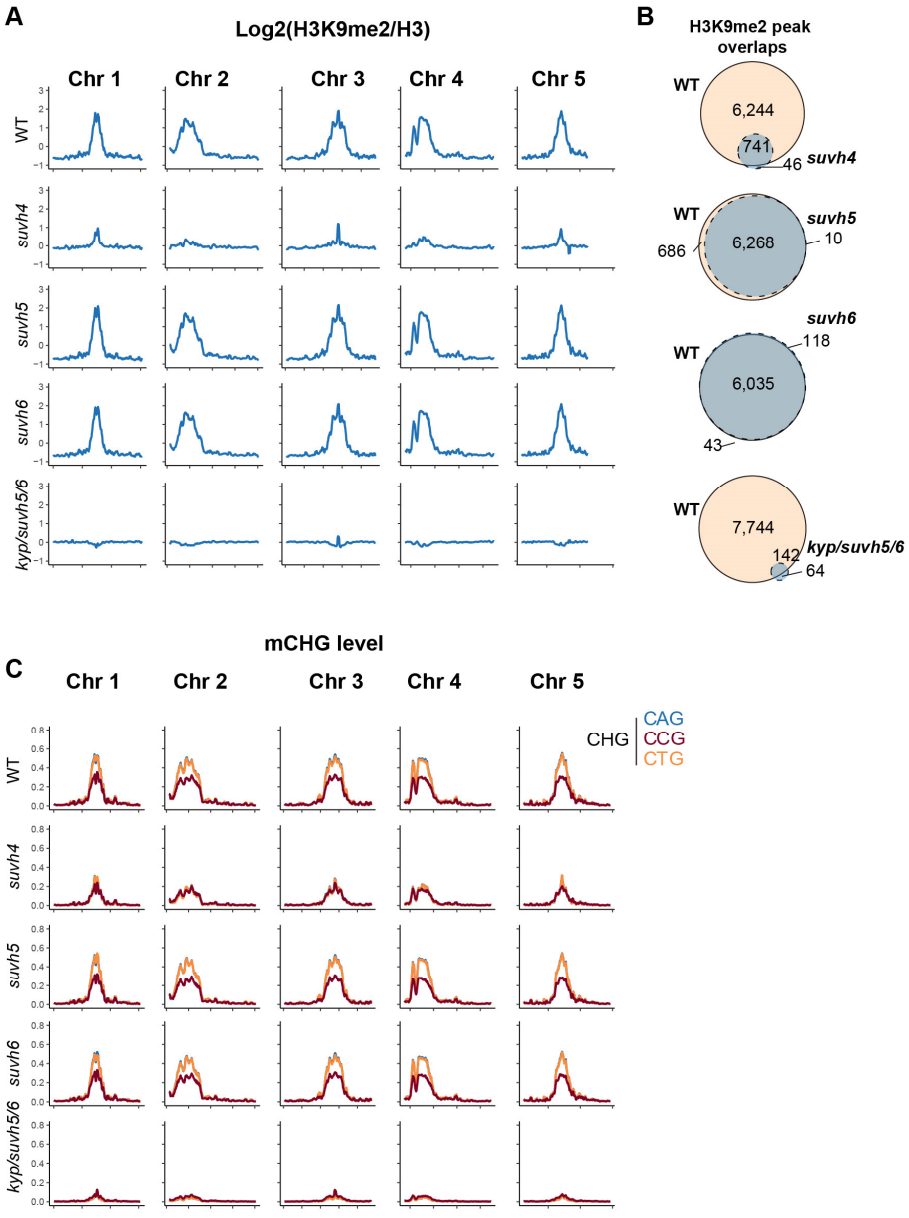

Fig S3. Genome wide effects of loss of SUVH MTases on H3K9me2 and CHG methylation.

(A) Chromosomal distribution of log2 ratio of H3K9me2/H3 in WT, *kyp*, *suvh5*, *suvh6* and *kyp/suvh5/6* (bins=100 Kb).

(B) Overlap of H3K9me2/H3 peaks in WT vs. *kyp*, WT vs. *suvh5*, WT vs. *suvh6* and WT vs. *kyp/suvh5/6*.

(C) Chromosomal distribution of CHG (mCAG, mCGG, mCTG) methylation in WT, *kyp*, *suvh5*, *suvh6* and *kyp/suvh5/6* (bin 100 Kb).

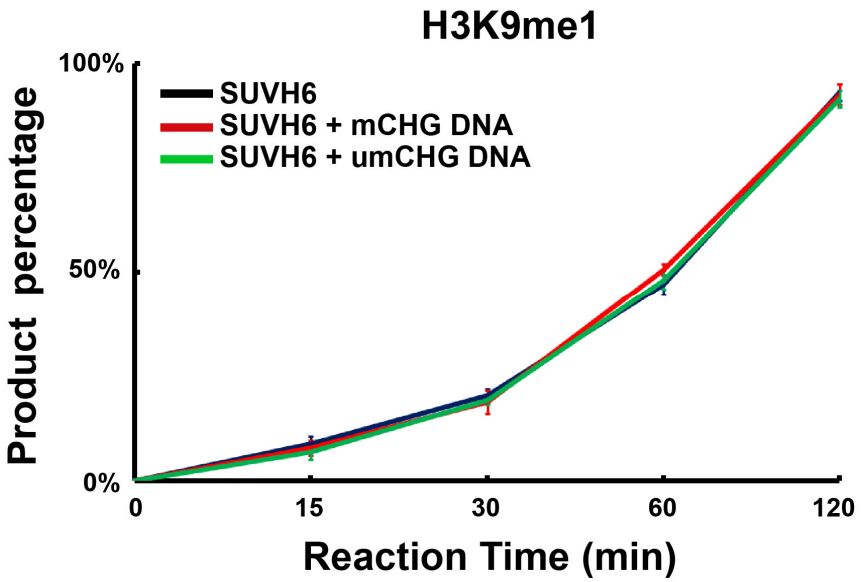

Fig. S4. The *in vitro* activity assay of SUVH6 in DNA-free (in black), mCHG bounded (in red), and unmethylated CHG bounded (in green) conditions indicating that the binding of DNA has no obvious effect on the activity of SUVH6. The percentage of the product H3K9me1 are plotted as means ± s.d. (n = 3). Reactions were stopped after incubation at the times indicated. The activity curve of the DNA-free SUVH6 is not totally identical as the activity curve of wild-type SUVH6 in Fig. 1F, which is probably due to the different batches of reactions with different batches of protein samples, SAM, and peptides.

Table S1. Data collection and refinement statistics.

	Se-SUVH6+SAM-1	Se-SUVH6+SAM-2	SUVH6+13bp DNA
Data collection			
PDB code	6A5K	6A5M	6A5N
Beamline	SSRF-BL17U1	SSRF-BL17U1	SSRF-BL19U1
Space group	$P2_12_12$	$P2_1$	$P2_12_12_1$
Wavelength (Å)	1.2827	0.9777	0.9792
Cell dimensions			
a, *b*, *c* (Å)	77.1, 122.2, 56.3	57.5, 77.1, 66.8	74.1, 79.4, 107.8
α, β, γ (°)	90, 90, 90	90, 110.1, 90	90, 90, 90
Resolution (Å)	50.0-1.9 (1.93-1.90)*	50.0-2.3 (2.38-2.30)	50.0-2.4 (2.49-2.40)
R_{merge}	0.071 (0.874)	0.117 (0.525)	0.081 (0.715)
$I / \sigma I$	41.6 (2.4)	13.6 (2.3)	32.3 (3.4)
Completeness (%)	99.9 (99.8)	98.7 (98.2)	97.9 (99.7)
Redundancy	16.9 (10.9)	5.5 (5.1)	5.4 (5.8)
Refinement			
R_{work} / R_{free}	0.184 / 0.214	0.178 / 0.223	0.230 / 0.279
No. atoms	4,115	4,078	4,515
Protein / DNA	3,681 / -	3,854 / -	3,938 / 534
SAM	27	27	-
Water / Zn^{2+}	403 / 4	193 / 4	40 / 3
B-factors (Å^2)	28.6	42.7	78.1
Protein / DNA	27.9 / -	42.9 / -	75.2 / 100.5
SAM	17.5	36.0	-
Water / Zn^{2+}	35.2 / 30.2	40.5 / 36.5	70.0 / 61.8
R.m.s. deviations			
Bond lengths (Å)	0.005	0.007	0.009
Bond angles (°)	0.945	1.123	1.314

*Highest-resolution shell is shown in parentheses.

Table S2. DNA oligos used in this research.

Name/Abbreviation	Oligonucleotide sequence	Design purpose
TCG/AGC (C = 5mC)	5'-GAGTACTCGTCAGTTC-3' 3'-CTCATGAGCAGTCAAG-5'	MST assay
TCG/AGC	5'-GAGTACTCGTCAGTTC-3' 3'-CTCATGAGCAGTCAAG -5'	MST assay
CAG/GTC	5'-GAGTACTCAGCAGTTC-3' 3'-CTCATGAGTCGTCAAG-5'	MST assay
CAT/GTA	5'-GAGTACTCATCAGTTC-3' 3'-CTCATGAGTAGTCAAG-5'	MST assay
CCG/GGC	5'-GAGTACTCCGCAGTTC-3' 3'-CTCATGAGGCGTCAAG-5'	MST assay
CCG/GGC	5'-GAGTACTCCGCAGTTC-3' 3'-CTCATGAGGCGTCAAG-5'	MST assay
Unmethylated CG, CAG, CAT containing DNA	5'- GAGTACGCATCAGTTCAT -3' 3'- CTCATGCGTAGTCAAGTA -5'	MST assay
mCHG-13bp	5'-GAGTACTCAGCAGT-3' 3'-TCATGAGTCGTCAC-5'	Crystallization, activity assay
Unmethylated CHG-13bp	5'-GAGTACTCAGCAGT-3' 3'-TCATGAGTCGTCAC-5'	Activity assay
Su4-93-F	5'-ATAGGATCC AATGGTAAGGACGTGAACTTGG-3'	Primer
Su4-624-R	5'-TATGCGGCCGCTCAGTAAAGGCGTTTCCTACAATTTAG-3'	Primer
Su5-271-F	5'-CGGGATCCATGAGGAAGAATAGTGAGAGG-3'	Primer
Su5-794-R	5'-CCCTCGAGTTAGTAGAGCCTACCACTACAC-3'	Primer
Su6-265-F	5'- GGATCCAGCGGTGATAGTAGTCGGAAC -3'	Primer
Su6-790-R	5'-GCGGCCGCTCAATAGAGCCTACGCCTAC-3'	Primer
Su6-762-790-dele-F	5'-GGATCCAGCGGTGATAGTAGTCGGAAC-3'	Primer
Su6-762-790-dele-R	5'-GCGGCCGCTCAGTAATTGTAGTCGTAACAGAGTTC-3'	Primer
Su6-674-692-dele-F	5'-GGTTCTATGGCAGAAGGTGATGAGTCGA-3'	Primer
Su6-674-692-dele-R	5'-CGATCCATCATATCTGTTACCAATATC-3'	Primer

Table S3. Median CHG methylation levels at *kyp/suvh5/suvh6* H3K9me2 peaks.

Type	Sample	Average	Reduction*
CCG	WT	0.364	-
	kyp	0.208	42.8%
	suvh5	0.351	3.4%
	suvh6	0.348	4.3%
	kyp/suvh5/6	0.048	86.9%
CTG	WT	0.564	-
	kyp	0.263	53.5%
	suvh5	0.561	0.5%
	suvh6	0.538	4.5%
	kyp/suvh5/6	0.026	95.3%
CAG	WT	0.578	-
	kyp	0.261	54.9%
	suvh5	0.571	1.2%
	suvh6	0.551	4.7%
	kyp/suvh5/6	0.026	95.6%

***Reduction = (WT-mutant)/WT. This table corresponds to the boxplots presented in Figure 5E.**

Table S4. Median CHH methylation levels at *kyp/suvh5/suvh6* H3K9me2 peaks.

Type	Sample	Average	Reduction*
CAA	WT	0.395	-
	kyp	0.203	48.6%
	suvh5	0.353	10.6%
	suvh6	0.366	7.5%
	kyp/suvh5/6	0.065	83.6%
CTA	WT	0.378	-
	kyp	0.194	48.6%
	suvh5	0.345	8.8%
	suvh6	0.351	7.3%
	kyp/suvh5/6	0.061	84.0%
CTT	WT	0.088	-
	kyp	0.038	56.3%
	suvh5	0.078	11.7%
	suvh6	0.081	7.6%
	kyp/suvh5/6	0.013	85.6%
CTC	WT	0.096	-
	kyp	0.041	57.6%
	suvh5	0.087	9.6%
	suvh6	0.088	8.2%
	kyp/suvh5/6	0.013	86.3%
CAT	WT	0.065	-
	kyp	0.024	63.0%
	suvh5	0.053	17.2%
	suvh6	0.057	11.3%
	kyp/suvh5/6	0.009	86.6%
CAC	WT	0.065	-
	kyp	0.026	59.9%
	suvh5	0.057	12.3%
	suvh6	0.059	9.0%
	kyp/suvh5/6	0.008	87.2%
CCA	WT	0.045	-
	kyp	0.018	59.6%
	suvh5	0.041	9.0%
	suvh6	0.041	9.0%
	kyp/suvh5/6	0.006	87.7%
CCC	WT	-	-
	kyp	-	-
	suvh5	-	-
	suvh6	-	-

	kyp/suvh5/6	-	-
CCT	WT	-	-
	kyp	-	-
	suvh5	-	-
	suvh6	-	-
	kyp/suvh5/6	-	-

***Reduction = (WT-mutant)/WT. This table corresponds to the boxplots presented in Figure 6A.**