Supplemental Material:

Molecular Basis for Histone N-terminal Methylation by NRMT1

Ruoxi Wu[1], Yuan Yue[1], Xiangdong Zheng[1,2], and Haitao Li[1,3]

[1]MOE Key Laboratory of Protein Sciences, Center for Structural Biology, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing 100084, P.R. China.

[2]Tsinghua-Peking Center for Life Sciences, Beijing, 100084, China

[3]Collaborative Innovation Center for Biotherapy, West China Hospital, Sichuan University, Chengdu 610041, P.R. China

To whom correspondence should be addressed:

Haitao Li

Medical Science Building C228

Tsinghua University

Beijing 100084, P.R.China

Tel: (+) 86-10-6277-1392

E-mail: lht@tsinghua.edu.cn

Supplemental Materials and Methods

Protein expression and purification

The full-length gene of human NRMT1 (residues1-223, GenBank ID: *28989*) was a gift from Dr. Jiahuai Han of Xiamen University. Wild type and mutant NRMT1 were cloned into the pET28b vector to produce N-terminal His-tagged protein. All NRMT1 mutants were generated by two-step overlap PCR and verified by gene sequencing. All NRMT1 proteins were overexpressed in E.*coli* strain BL21 (DE3) at 16°C for 16 hours and induced by 0.2mM isopropyl-β-D-thiogalactopyranoside (IPTG) at a cell density of 0.8 (OD_{600nm}). Cells were harvested, re-suspended in lysis buffer (200 mM NaCl, 20 mM Tris-Na, pH 8.0 and 5% Glycerol) and lysed by EmulsiFlex-C3 high pressure homogenizer (AVESTIN). After centrifugation, the 6xHis full-length NRMT1 and its mutants were purified to homogeneity over successive HisTrap, anion exchange Q, and Superdex G75 columns (GE Healthcare). Protein purity was analyzed by SDS-PAGE. Proteins concentrations were determined by UV spectroscopic measurement at 280 nm.

Crystallization, data collection and structure determination

The full-length NRMT1 protein was concentrated to 18 mg ml^{-1} in a buffer containing 150 mM NaCl, 20 mM Tris-Na, pH 8.0 and 2 mM DTT. NRMT1 was incubated with SAH and α-N-dimethylated CENP-A N-terminal peptide 1-9 ($_1G_{me2}PRRRSRPK_9$) in a molar ratio of 1:5:1.5 for about two hours at 4 °C before crystallization. Complex crystals were grown at 18 °C by mixing an equal volume of the protein and reservoir solution using the sitting-drop vapor diffusion method. NRMT1 crystals appeared after one week in crystallization buffer

consisting of 0.2 M ammonium acetate, 0.1 M sodium citrate tribasic dihydrate, pH 5.6 and 30% PEG4000.

For DmH2B complex crystallization, unmodified DmH2B 1-9 ($_1$PPKTSGKAA$_9$) peptide was incubated with SAM and NRMT1 in a 1.5:5:1 ratio for two hours at 4°C. After optimization, crystal was grown in a reservoir condition of 0.2 M ammonium acetate, 0.1 M sodium citrate tribasic dihydrate, pH 5.7 and 28% PEG4000.

Before data collection, crystals were flash frozen in liquid nitrogen under cryoprotectant conditions (reservoir solution supplemented with 12% PEG400). Diffraction data were collected at beamline BL17U of the Shanghai Synchrotron Radiation Facility (SSRF) and integrated and scaled with HKL2000 suite (Otwinowski and Minor 1997). All structures were determined by molecular replacement using MOLREP in CCP4 (Vagin and Teplyakov 2010) with the peptide-free NRMT1 structure (PDB code: 2EX4) as the search model. Refinement and model building were carried out with programs PHENIX (Adams et al. 2010) and COOT (Emsley and Cowtan 2004), respectively. The refined structures were validated by PROCHECK (Laskowski et al. 1993). Structural analysis and figure preparation were mostly performed using the program PyMol (http://www.pymol.org). Electrostatic potential surfaces were calculated by the APBS tool (Baker et al. 2001) within PyMol.

Radiometric filter methyltransferase assay

N-terminal CENP-A peptides (1-20) were combined with wild type or mutant NRMT1 to generate a 23 µl reaction mixture in buffer containing 20 mM Tris-Na, pH 7.9, 4 mM EDTA, 0.5 mM DTT. Wide type and mutant enzymes were kept at a final concentration of 0.04 µM

after optimization. Final concentrations of CENP-A peptide and radio-labelled SAM were 17 μM and 0.56 μM, respectively. To initiate the reaction, 2 μl ^{3}H-S-adenosyl methionine (SAM) (55mCi/ml, PerkinElmer Life Sciences) was added to the reaction mixture. After 1 hour incubation at 30°C, 5 μl of reaction mixture was immediately spotted on P81 phosphocellulose paper (Millipore), washed 3 x 10 min with sodium carbonate, pH 9.0 and air-dried for scintillation counting. Scintillation Cocktail (PerkinElemer) was added to filter paper and counted by MicroBeta Jet (PerkinElemer). All assays were prepared in triplicate. Controls were blank-filter paper only, ^{3}H-SAM only and enzyme free samples, which showed a background C.P.M within the range of 20-80.

MALDI-TOF-based enzymatic assay

For NMRT1 methylation processivity assay, the reaction mixtures were prepared in 10 μl containing 2 μM NRMT1, 2 μM CENP-A (1-20) and 100 μM SAM, and incubated at 25 °C at different time courses (0.5 h, 2 h, 5h). The reaction was terminated by 10% formic acid and diluted 10 times by Millipore water for MALDI-TOF mass spectrometry analysis. Similar procedures were adopted for histone profiling and mutant CENP-A methylation assays. An incubation time of 5 hours was used in order to maximize the MS signal. All peptides were synthesized by the Scilight-Peptide company (http://www.scilight-peptide.com).

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) assay was conducted at 25°C with a MicroCal iTC200 instrument (GE Healthcare). Full length NRMT1 was exchanged to ITC buffer containing

100 mM NaCl and 20 mM Hepes-Na, pH 7.5 by gel filtration. Lyophilized peptide was directly dissolved in ITC buffer for ITC titration. Concentrations of NRMT1 and CENP-A peptide used for titration were 86 μM and 860 μM, respectively. Protein concentration was determined by its 280 nm absorbance. Peptides were quantified by weighing on a large scale. Acquired ITC data were analyzed by the program Origin 8.0 (GE Healthcare) using the “One Set of Binding Sites” fitting model.

Fluorescence-based thermal shift assay

The fluorescence-based thermal shift assay (TSA) was performed with a CFX96^{TM} real-time PCR instrument (Bio-Rad). A typical TSA solution is composed of 1 mg/mL NRMT1, 180 μM SAH, 360 μM histone peptides and 2xSypro Orange (Invitrogen). All solutions were prepared in 25 μl under the buffer: 200 mM NaCl, 20 mM Tris-Na, pH 8.0. During TSA assays, all samples were heated from 25°C to 90°C at an increase rate of 0.5°C per minute. Protein denaturation was monitored by increased fluorescence signal of Sypro Orange, which captures exposed hydrophobic residues during thermal unfolding. The recorded curves were analyzed by the software CFX-Manager (Bio-Rad). The temperature corresponding to the inflection point was defined as Tm.

Supplemental Table S1 Data collection and refinement statistics

	NRMT1-CENP-A -SAH	NRMT1-DmH2B -SAH
Data collection		
Space group	C2	C2
Cell dimensions		
a, b, c (Å)	114.8, 66.2, 69.0	114.8, 66.3, 68.7
α, *β*, *γ* (°)	90, 106.7, 90	90, 106.4, 90
Resolution (Å)	50-1.3 (1.32-1.30)*	50-1.5 (1.53-1.50)
R_{sym} or R_{merge}	10.8 (63.1)	9.7 (81.7)
I / σ*I*	16.4 (2.9)	30.6 (3.0)
Completeness (%)	98.3 (96.7)	99.2 (98.0)
Redundancy	4.7 (4.6)	7.3 (7.3)
Refinement (F>0)		
Resolution (Å)	30-1.3	32-1.5
No. reflections	119377	78797
R_{work} / R_{free}	13.2/16.0	15.3/18.0
No. atoms		
Protein	3716	3674
Peptide/SAH	123/52	98/52
Water	734	754
B-factors		
Protein	11.5	15.8
Peptide/SAH	16.5/6.7	21.6/10.3
Water	25.2	28.6
R.m.s. deviations		
Bond lengths (Å)	0.006	0.005
Bond angles (°)	1.16	1.05

* Values in parentheses are for highest-resolution shell.

Supplemental Table S2 Histone methylation profiling by human NRMT1

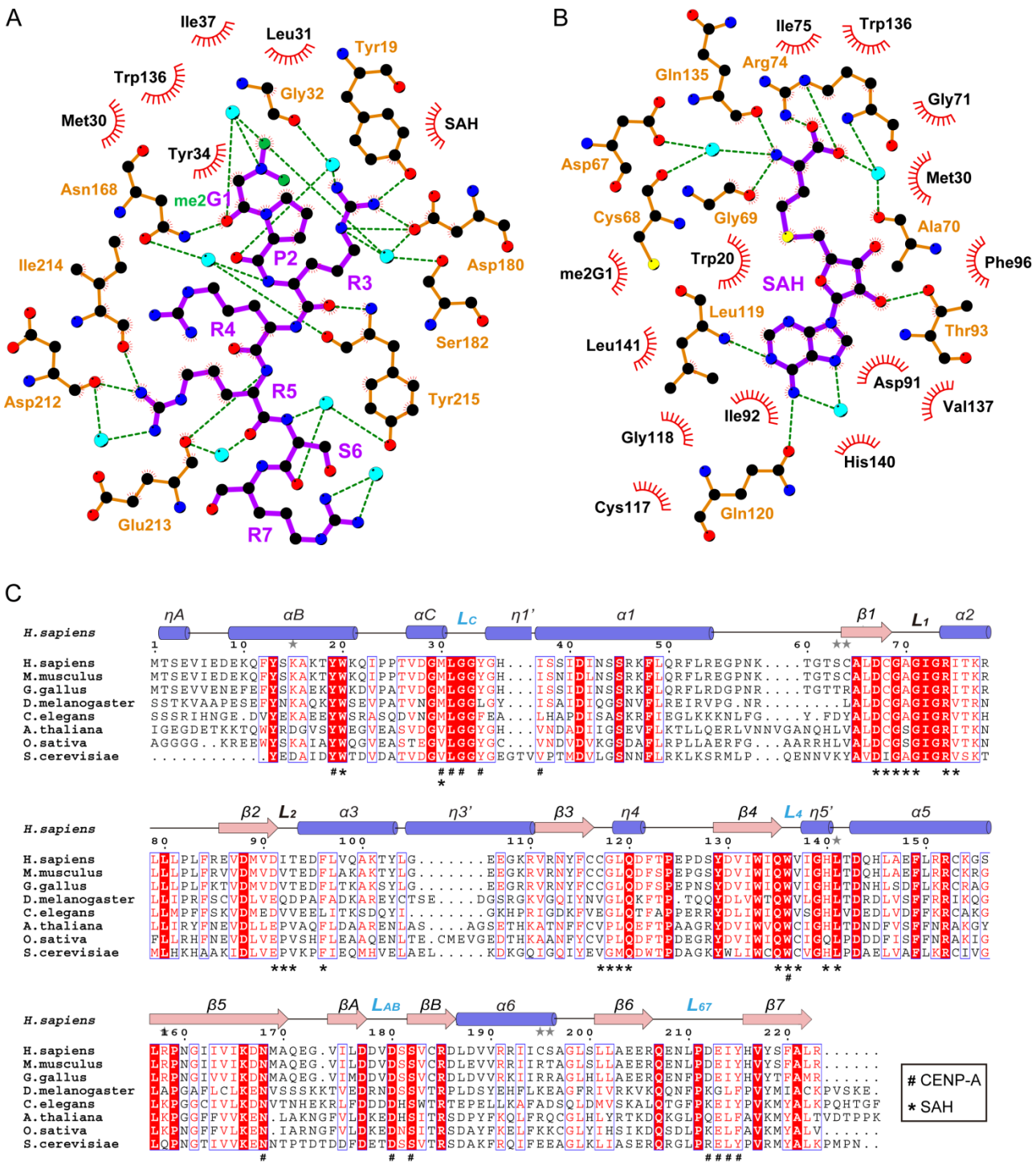
Histones	Sequence (1-9)	Methylated? Yes (▲) No (○)	Histones	Sequence (1-9)	Methylated? Yes (▲) No (○)
H1.1	SETAPAAPA	○	H2B 3-B	PDPSKSAPA	○
H1.3	SETVPAAPA	○	H2B 1B/1K	PEPSKSAPA	○
H1t	SETVPLAPT	○	H2B 1D	PEPTKSAPA	○
H1.5	SETAPAETA	○	H2B 2-C	PRRRRRRGS	○
H1.0	TENSTSAPA	○	H2B 2-D	PEPAKFAPA	○
H1x	SVELEEALP	○	H2B FS/2E	PEPAKSAPA	○
H1oo	APGSVTSDI	○	H2B 1-K	PEPAKSAPA	○
macroH2A.1	SSRGGKKKS	○	H2B 1-L	PELAKSAPA	○
MacroH2A.2	SGRSGKKKM	○	H2B 1-M	PEPVKSAPV	○
H2A.V/H2A.Z	AGGKAGKDS	○	H2B F-M	KCHAGSLSA	○
H2A 1/1A/1B/ 1C/1D/1J/1H/ 2A/2B/2C/J	SGRGKQGGK	○	H2B W-T	LRTEVPRLP	○
			H2B 1-J	PEPAKSAPA	○
			H2B 1-C/E/F/G/I	PEPAKSAPA	○
H2A.X	SGRGKTGGK	○	DmH2B (fruit fly)	PPKTSGKAA	▲, Me2
H2A-Bbd-1/2/3	PRRRRRRGS	○	H3.1/3.1t/3.2/3.3/3.3C	ARTKQTARK	○
H4	SGRGKGGKG	○	CENP-A	GPRRRSRKP	▲, Me3

§ MALDI-TOF mass spectrometry was applied to detect the methylation products.

Supplemental Table S3 Co-evolution analysis of NRMT1 recognition motif

	N-terminal Sequence	"Xaa-P-K/R" motif?		N-terminal Sequence	"Xaa-P-K/R" motif?
H.sapiens			*D.melanogaster*		
CENP-A	GPRRRSRKP	√	CID (CENP-A homolog)	PRHSRAKRA	×
CENP-B	GPKRRQLTF	√	RCC1	PRRKALTNN	×
RCC1	SPKRIAKRR	√	H2B	PPKTSGKAA	√
H2B	PEPAKSAPA	×			
M.musculus			*T.thermophila*		
RCC1	PPKRIAKRR	√	CNA1 (CENP-A homolog)	ARKAYQPKR	×
CENP-B	GPKRRQLTF	√	RCC1	IVNAQQQQF	×
CENP-A	GPRRKPQTP	√	H2B	APKKAPAAA	√
H2B	PEVAVKGAT	×			
G.Gallus			*S.cerevisiae*		
CENP-A	PRPKPRSPR	×	Cse4p (CENP-A homolog)	SSKQQWVSS	×
RCC1	SGKRAAKKS	×	HTB2 (H2B homolog)	SSAAEKKPA	×
H2B 7	PEPAKSAPA	×	Srm1p (RCC1 homolog)	VKRTVATNG	×

§ CENP-B homolog sequences are unavailable in current protein database in the organisms of *G.Gallus, D. melanogaster, T. thermophile* and *S. cerevisiae*.



Supplemental Figure S1 Sequence alignment and LigPlot analyses of CENP-A and SAH recognition by NRMT1

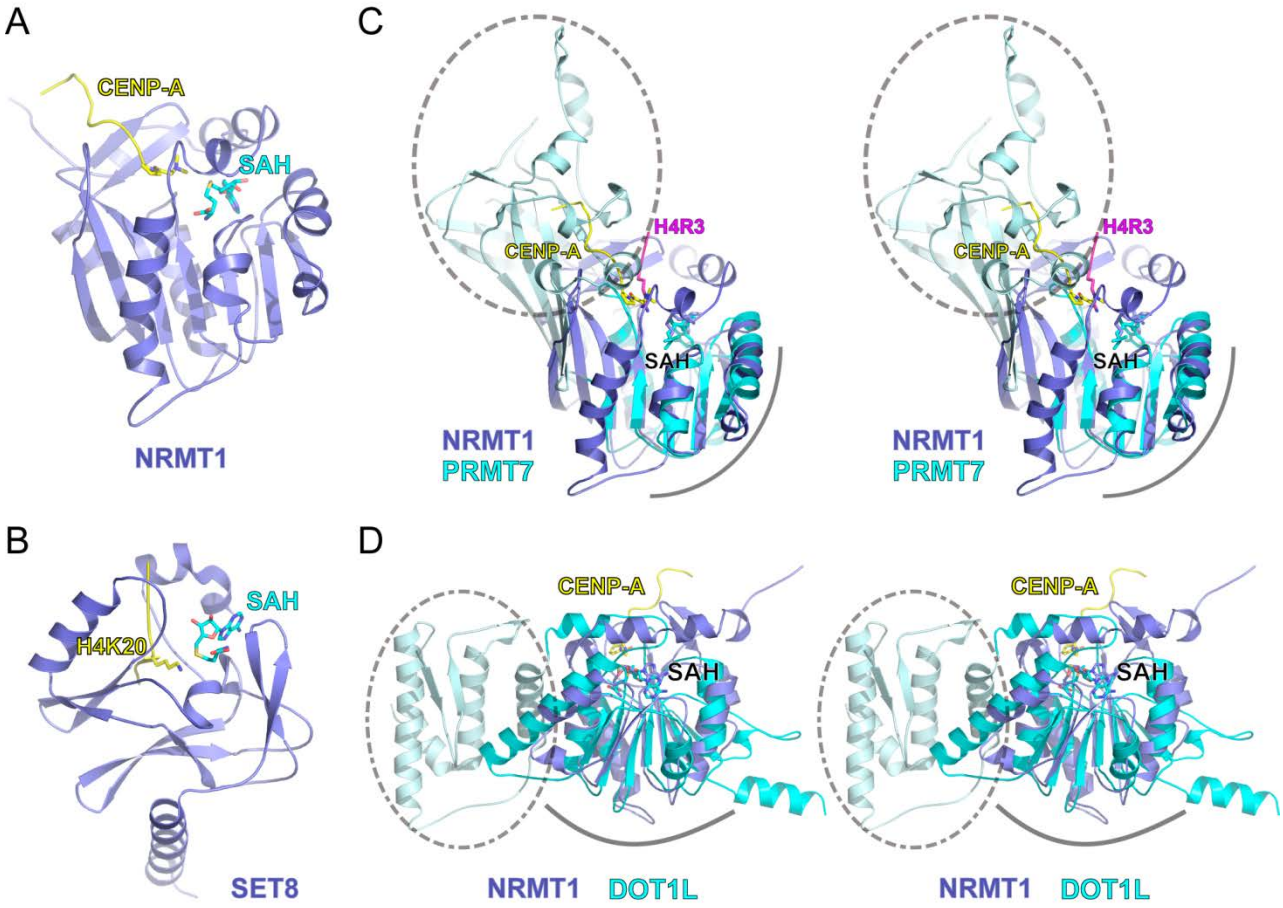
(A) Binding details around dimethylated CENP-A_{1-7} peptide. CENP-A peptide is colored in

purple; hydrophobic interactions between CENP-A and NRMT1 are illustrated by red eyelash marks; hydrogen bonds are labeled as green dot lines; waters are depicted as cyan balls.

(B) Binding details around SAH. SAH is colored in purple. The meanings of the other symbols in the plot are the same as described in panel A.

(C) Sequence alignment among NRMT1 homologs in human (H. *sapiens*), mouse (M. *musculus*), Chicken (G. *gullus*), fruit fly (D. *melanogaster*), nematode (C. *elegans*), Arabidopsis thaliana (A. *thaliana*), Alfalfa (O. *sativa*) and baker's yeast (S. *cerevisiae*). SAH and CENP-A Interacting regions are highly conserved shaded as red. Residues that participate in CENP-A and SAH interaction are highlighted by symbols "#" and "*", respectively.

Wu_FigS2

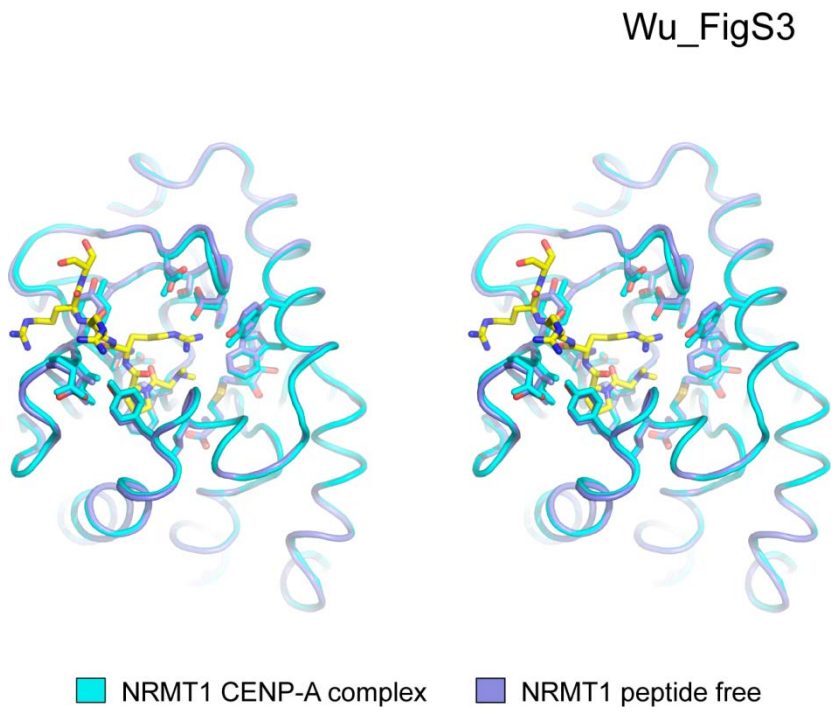


Supplemental Figure S2 Structural comparison among NRMT1, SET8, PRMT7, and DOT1L

(A) Ribbon view of NRMT1 ternary complex bound to CENP-A and SAH.

(B) Ribbon view of SET8 ternary complex bound to H4K20 peptide and SAH. Coordinates were taken from PDB entry 1ZKK. SET8 adopts a typical SET domain methyltransferase fold, which is distinct from the SAM-MTase fold of NRMT1.

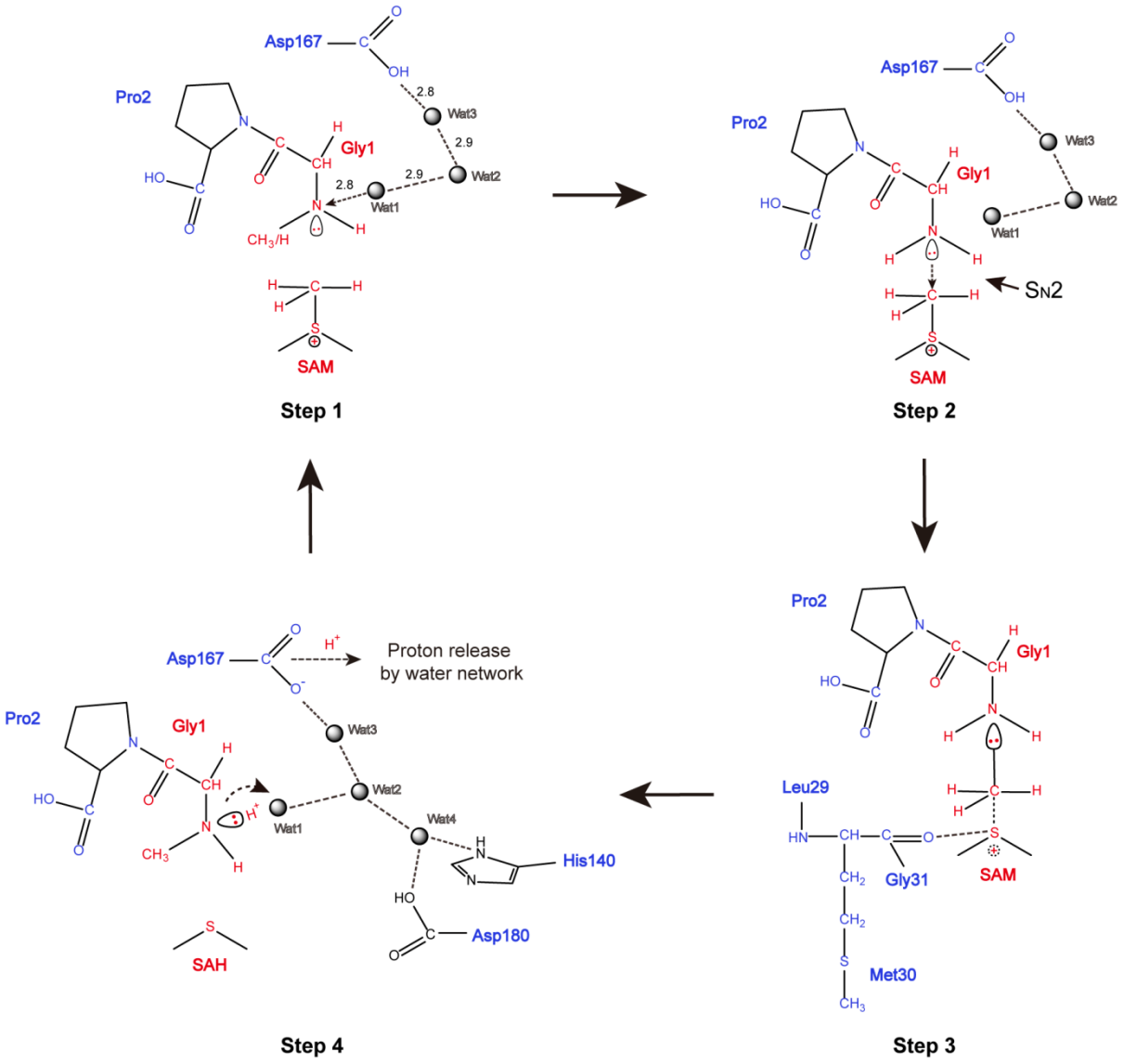
(C) and (D) Structure superimposition of NRMT1 with *T. brucei* PRMT7 (PDB: 4M38) and human DOT1L (PDB: 3QOW). Note the well aligned SAM-MTase core domains that are responsible for SAM binding, and misaligned extra N- or C-terminal segments that participate in recognition of the methyl acceptor substrate.



Supplemental Figure S3 Structural alignment of NRMT1 ternary and binary complexes with and without bound CENP-A peptide.

Figure is shown in stereo view. The ternary complex is colored cyan with CENP-A peptide shown as yellow sticks. The binary complex is colored slate blue. Little conformational change was observed upon CENP-A peptide binding.

Wu_FigS4



Supplemental Figure S4 Step-wise reaction cycle catalyzed by NRMT1

Supplemental References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW et al. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D-Biol Crystallogr* 66: 213-221.

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* 98: 10037-10041.

Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D-Biol Crystallogr* 60: 2126-2132.

Laskowski RA, Moss DS, Thornton JM. 1993. MAIN-CHAIN BOND LENGTHS AND BOND ANGLES IN PROTEIN STRUCTURES. *J Mol Biol* 231: 1049-1067.

Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276: 307-326.

Vagin A, Teplyakov A. 2010. Molecular replacement with MOLREP. *Acta Crystallogr Sect D-Biol Crystallogr* 66: 22-25.