Supporting Information

Selective Inhibitors of Histone Methyltransferase DOT1L: Design, Synthesis and Crystallographic Studies

Yuan Yao,[†,§] Pinhong Chen,[†,§] Jiasheng Diao,[†] Gang Cheng,[†] Lisheng Deng,[†] Justin L. Anglin,[†] B. V. Venkataram Prasad,[#] Yongcheng Song[†]*

[†]Department of Pharmacology, [#]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030

Table of Contents:

Figure S1	Page S2
Figure S2	Page S3
Table S1	Page S4
Figure S3	Page S5
Figure S4	Page S6
Experimental Section	Page S6
Full author list of Ref. 16	Page S13
References for S.I.	Page S14

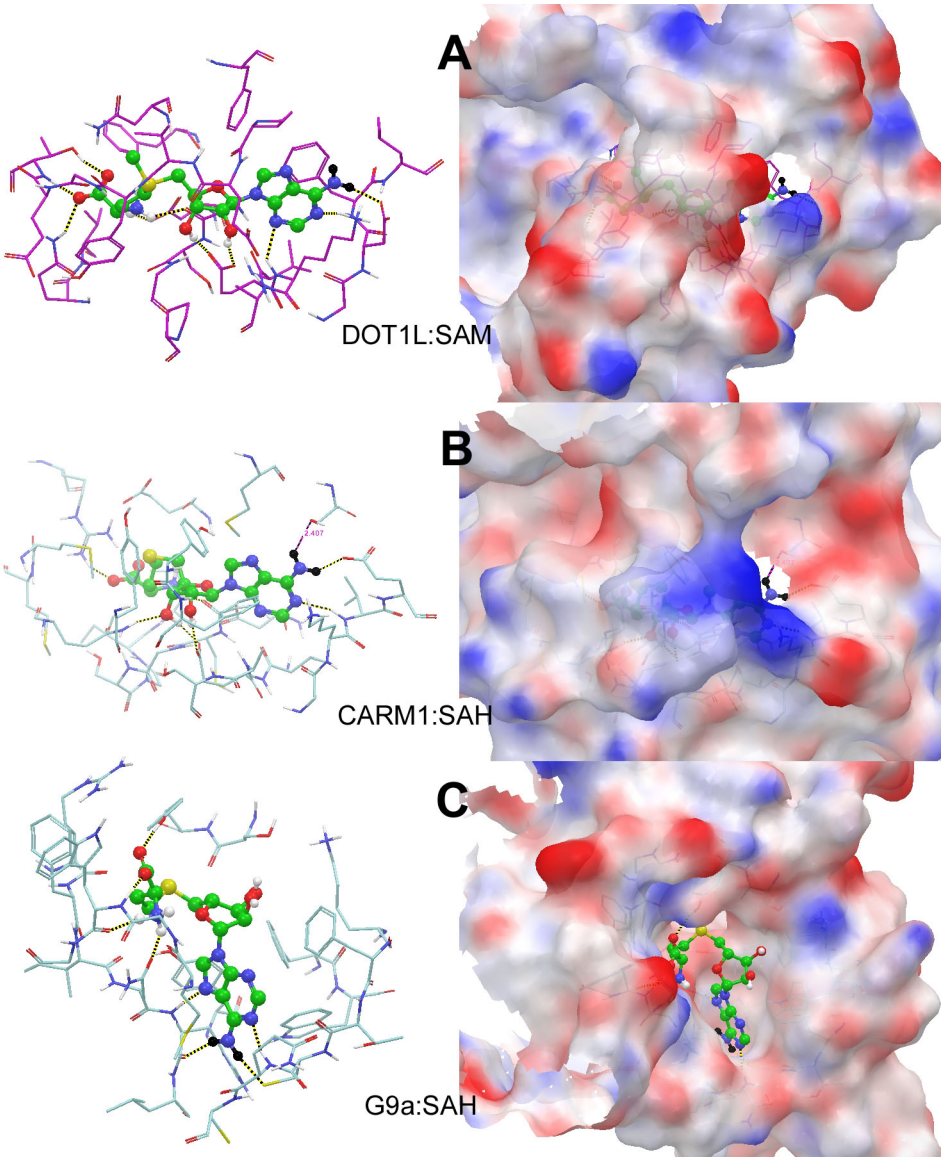

Figure S1. Crystal structures of (**A**) DOT1L:SAM (PDB: 1NW3); (**B**) CARM1:SAH (PDB: 2V74); and (**C**) G9a:SAH (PDB: 3K5K), showing the 6-NH_2 (the two H atoms highlighted in black) of SAM forming only one H-bond with DOT1L (**A**, left) with a mainly hydrophobic cavity nearby (**A**, right). The 6-NH_2 group of SAH forms two H-bonds with CARM1 and G9a (**B** and **C**). Electrostatic potential molecular surfaces of the proteins are shown with 20% transparency.

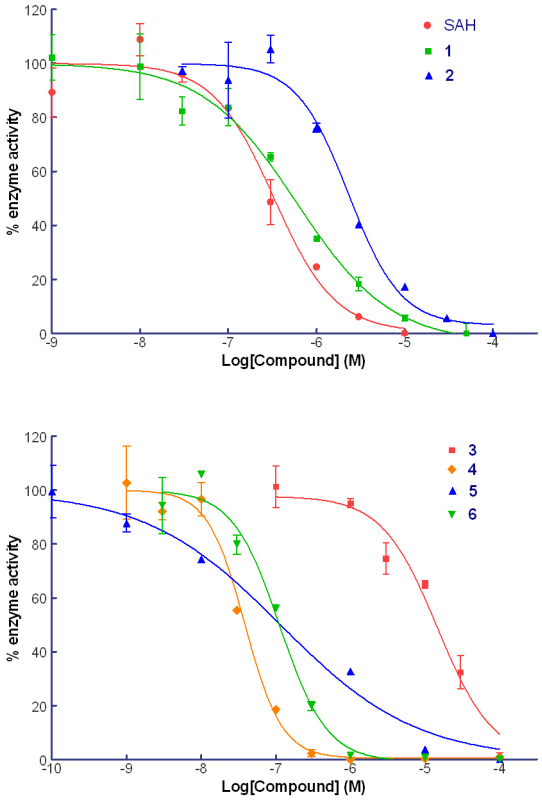

Figure S2. Dose response curves for DOT1L inhibited by compounds SAH, **1** – **6**.

Table S1. Data processing and refinement statistics

A. Data processing	
Wavelength (Å)	1.542
Space group	$P6_5$
Unit cell dimensions	
a, b, c (Å)	152.75, 152.75, 50.89
α, β, γ (º)	90.0, 90.0, 120.0
Resolution (Å)	100-2.5(2.54-2.5)
Unique reflections	23040(1232)
Completeness (%)	96.9(100)
Redundancy	12.9(13.0)
R_{sym} (%)	14.0(86.3)
$I/\sigma(I)$[b]	18.0(3.0)
B. Refinement	
Resolution (Å)	30.55-2.5(2.54-2.5)
Number of reflections used in working set	21865(1101)
Number of reflections for R_{free} calculation	1154(50)
R_{work} (%)	23.4(34.4)
R_{free} (%)[a]	27.4(39.2)
Number of all non-hydrogen atoms	2805
Number of solvent waters	91
Mean B-factor from Wilson plot (Å^2)	63.0
Mean B-factor, protein atoms (Å^2)	63.7
Mean B-factor, solvent atoms (Å^2)	61.9
Mean B-factor, inhibitors (Å^2)	48.2
Root mean square deviations from ideality	
Bond length (Å)	0.007
Bond Angle (º)	1.3
Dihedral (º)	21.9
Improper (º)	0.81
Ramachandran plot[b]	
Residues in most favored regions	89.8%
Residues in additional allowed regions	10.2%
Residues in generously allowed regions	0.0%
Residues in disallowed regions	0.0%

[a]A subset of the data (5%) was excluded from the refinement and used to calculate R_{free}.
[b]Ramachandran plot is generated by Procheck[1].

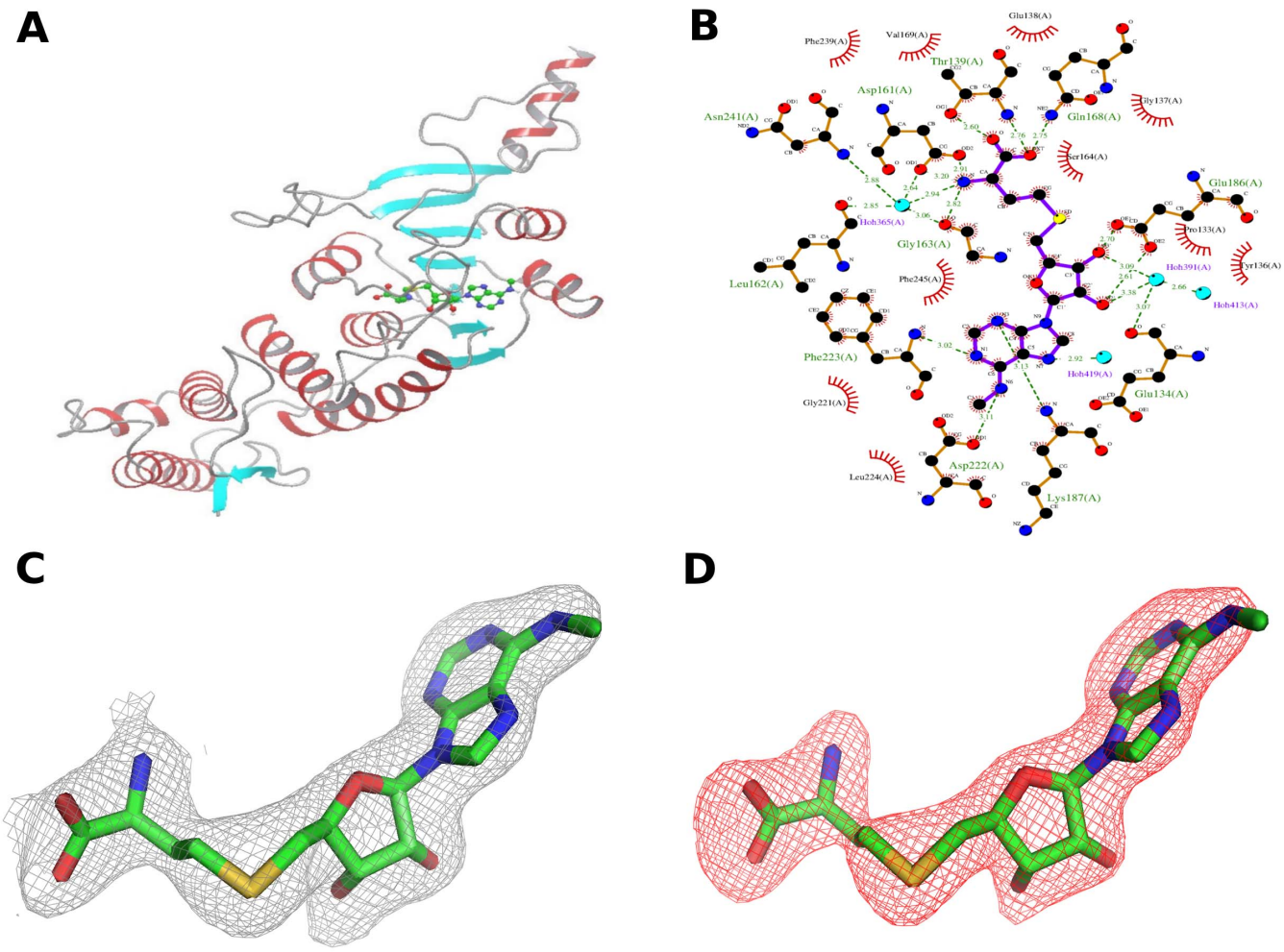

Figure S3. (**A**) The overall structure of human DOT1L in complex with compound 1; (**B**) Protein-ligand interaction diagram for **1** in DOT1L; (**C**) The $2F_o$-F_c electron density map of **1**, contoured at 1σ; (**D**) The F_o-F_c omit map of **1**, contoured at 3σ.

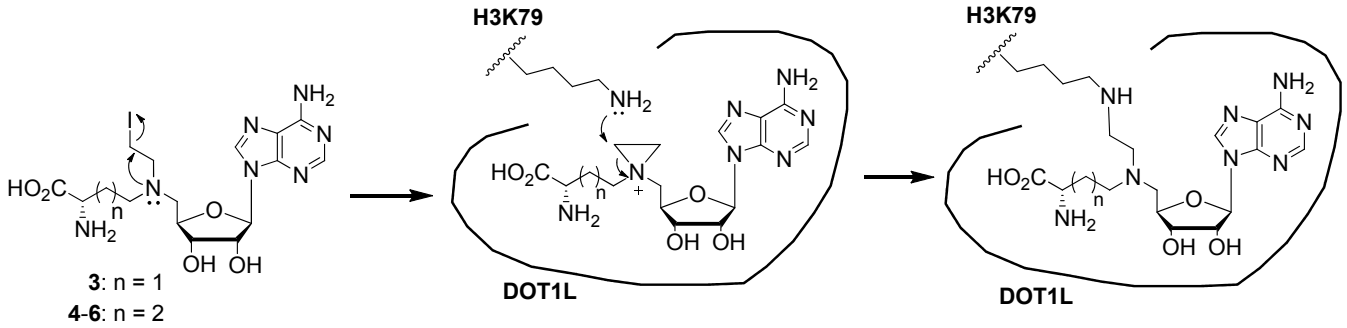

Figure S4. Mechanism of action of compounds **3** – **6**.

Experimental Section

All reagents were purchased from Alfa Aesar (Ward Hill, MA) or Aldrich (Milwaukee, WI). Compounds were characterized by ^{1}H NMR on a Varian (Palo Alto, CA) 400-MR spectrometer and the purities monitored by a Shimadzu Prominence HPLC with a Zorbax C18 or C8 column (4.6 x 250 mm) or using ^{1}H (at 400 MHz) absolute spin-count quantitative NMR analysis with imidazole as an internal standard. Identities of all new compounds were confirmed with high resolution mass spectra (HRMS) using a ThermoFisher LTQ-Orbitrap mass spectrometer.

S-(N^6-Methyl-adenosyl)-*L*-homocysteine (1). It was prepared according to a literature method,[2] using N^6-methyl-adenosine (562 mg, 2.0 mmol) as the starting compound, giving **1** as a white powder (270 mg, 34% overall yield). ^{1}H NMR (400 MHz, D_2O): δ 8.25 (s, 1 H), 8.18 (s, 1 H), 6.02 (d, *J* = 4.4 Hz, 1 H), 4.86 (m, 1 H), 4.36 (m, 1 H), 4.28 (m, 1 H), 3.72 (m, 1 H), 3.03 (s, 3 H), 2.98-2.88 (m, 2 H), 2.63 (t, *J* =7.6 Hz, 2 H), 2.12-1.96 (m, 2 H).

S-(N^6-Benzyl-adenosyl)-*L*-homocysteine (2). It was prepared similarly as **1** using N^6-benzyl-adenosine (714 mg, 2.0 mmol) as the starting compound, giving **2** as a white powder (245 mg, 26% overall yield). ^{1}H NMR (400 MHz, D_2O): δ 8.28 (s, 1 H), 8.14 (s, 1 H), 7.33-7.15 (m, 5 H), 6.02 (d, *J* = 4.4 Hz, 1 H), 4.80 (m, 1 H), 4.36 (m, 1 H), 4.28 (m, 1 H), 3.72 (m, 1 H), 3.30 (s, 2 H), 2.98-2.88 (m, 2 H), 2.63 (t, *J* =7.6 Hz, 2 H), 2.12-1.96 (m, 2 H).

N^γ-(5'-Adenosyl)-N^γ-(2-iodoethyl)-(*S*)-2,4-diaminobutyric acid hydrochloride (3). It was prepared according to a published procedure.[3] ^{1}H NMR (400 MHz, D_2O): δ 8.46 (s, 1 H), 8.45 (s, 1 H), 6.17 (d, *J* = 4.4 Hz, 1 H), 4.86 (t, *J* = 4.8 Hz, 1 H), 4.80 (m, 1 H), 4.51 (m, 2 H), 4.10-3.46 (m, 6 H), 3.38 (t, *J* = 8.0 Hz, 2 H), 2.41-2.11 (m, 2 H).

N^δ-(5'-Adenosyl)-N^δ-(2-iodoethyl)-(*S*)-2,5-diaminopentanoic acid hydrochloride (**4**). It was prepared according to Scheme 1. To a suspension of adenosine (8.03 g, 30 mmol) in 100 mL dry acetone was added trimethyl orthoformate (2.4 mL), followed by $SOCl_2$ (6.75 mL, 90 mmol) dropwise. After stirring

overnight, the solid was filtered, dissolved in saturated $NaHCO_3$, and neutralized to pH ~ 7. The solid was collected by filtration, washed with ether (20 mL), and dried *in vacuo* to give 2',3'-isopropylidene-N^6-methyl-adenosine, to which (3.07 g, 10 mmol) in dry THF (20 mL) were added phthalimide (1.62 g, 11 mmol) and PPh_3 (2.88 g, 11 mmol), followed by diisopropyl azodicarboxylate (DIAD, 1.08 g, 11 mmol). After 2.5 h, the white solid was filtered, washed with 20 mL of cold Et_2O. The crude Mitsunobu product and hydrazine hydrate (2.4 mL, 50 mmol) were refluxed overnight in ethanol (20 mL). After cooling, the reaction mixture was filtered and the filtrate evaporated to dryness. To the product (612 mg, 2.0 mmol) in THF (5 mL) were added Et_3N (0.9 mL, 6.4 mmol) and ethyl bromoacetate (0.28 mL, 2.5 mmol). After stirring overnight, the resulting mixture was filtered and evaporated to dryness. The residue oil thus obtained was dissolved in THF (10 mL) and cooled to -20 °C, followed by addition of $LiAlH_4$ (166 mg, 4.4 mmol). The reaction was allowed to warm to room temperature over 2.5 h, quenched with saturated $NaHCO_3$, and the product was extracted with 50 mL of EtOAc, washed successively with saturated $NaHCO_3$, water, and saturated NaCl. The organic layer was dried over sodium sulfate, evaporated, and purified with a flash column chromatography (silica gel, CH_2Cl_2/MeOH 95:5) to give **7** as a white solid (525 mg, 75%). Compound **7** (211 mg, 0.6 mmol) was added to a solution of aldehyde **8**[4] (310 mg, 0.8 mmol) in anhydrous MeOH (3 mL). $NaBH_3CN$ (67 mg, 1.05 mmol) and HCl (0.2 mL, 2.5 M in ethanol) were added to the stirring solution. The reaction mixture was stirred at room temperature overnight before being diluted with EtOAc (20 mL) and $NaHCO_3$. The organic layer was washed with $NaHCO_3$, dried over Na_2SO4, and evaporated. Purification with a column chromatography (silica gel, CH_2Cl_2/MeOH 90:10) gave **9** as a white solid (240 mg, 55%). To a solution of triphenylphosphine (77 mg, 0.30 mmol) and imidazole (20 mg, 0.30 mmol) in CH_2Cl_2 (1 mL) was added I_2 (79 mg, 0.31 mmol) at 0 ^oC, followed by addition of **9** (140 mg, 0.19 mmol) in CH_2Cl_2 (1 mL). The reaction mixture was stirred for 2 h, diluted with ice-chilled CH_2Cl_2 (20 mL) and H_2O (5 mL). The organic layer was washed with H_2O before being cooled to 0 ^oC, to which was added HCl in dioxane (1.5 mL, 4 N). After 1 h, the solvent was removed *in vacuo* and the residue was triturated with

diethylether (20 mL) to give **4** as a white powder (85 mg, 75% overall yield from **9**). ^{1}H NMR (400 MHz, D_2O): δ 8.46 (s, 1 H), 8.44 (s, 1 H), 6.15 (d, *J* = 3.6 Hz, 1 H), 4.86 (m, 1 H) 4.80 (m, 1 H), 4.57-4.46 (m, 2 H), 3.94-3.66 (m, 4 H), 3.41-3.29 (m, 4 H), 1.99-1.79 (m, 4 H). HRMS (ESI) $[M+H]^+$ Calcd for $C_{17}H_{27}N_7O_5I^+$: 536.1118, Found: 536.1111.

N^{δ}-[5'-(N^6-Methyl-adenosyl)]-N^{δ}-(2-iodoethyl)-(*S*)-2,5-diaminopentanoic acid hydrochloride (**5**). It was prepared from N^6-methyl-adenosine (562 mg, 2.0 mmol) following the above general procedure as a white powder (150 mg, 22% overall yield). ^{1}H NMR (400 MHz, D_2O): δ 8.43 (s, 1 H), 8.39 (s, 1 H), 6.17 (d, *J* = 3.6 Hz, 1 H), 4.80 (m, 2 H), 4.55-4.49 (m, 2 H), 3.90 (t, *J* = 4.8 Hz, 2 H), 3.84-3.75 (m, 2 H), 3.47-3.35 (m, 4 H), 3.21 (s, 3 H), 1.96-1.83 (m, 4 H). HRMS (ESI) $[M+H]^+$ Calcd for $C_{18}H_{29}N_7O_5I^+$: 550.1275, Found: 550.1255.

N^{δ}-[5'-(N^6-Benzyl-adenosyl)]-N^{δ}-(2-iodoethyl)-(*S*)-2,5-diaminopentanoic acid hydrochloride (**6**). It was prepared from N^6-benzyl-adenosine (803mg, 2.25 mmol) following the above general procedure as a white powder (170 mg, 22% overall yield). ^{1}H NMR (400 MHz, d_6-DMSO): δ 8.44 (s, 1 H), 8.28 (s, 1 H), 7.35-7.20 (m, 5 H), 5.98 (d, *J* = 4.0 Hz, 1 H), 4.71-4.62 (m, 2 H), 4.40 (m, 1 H), 4.23 (m, 1 H), 3.91 (m, 2 H), 3.70-3.59 (m, 2 H), 3.54-3.36 (m, 4 H), 3.20 (s, 2 H), 1.84-1.65 (m, 4 H). HRMS (ESI) $[M+H]^+$ Calcd for $C_{24}H_{33}N_7O_5I^+$: 626.1588, Found: 626.1576.

Expression and purification of human DOT1L. Human DOT1L(1-472) was expressed and purified as described in the literature.[5,6] In brief, BL21-CodonPlus strain (Agilent) was transformed with pGEX-KG-hDOT1L(1-472) plasmid and cultured at 37 °C in LB medium containing ampicillin (50 μg/mL) and chloramphenicol (34 μg/mL). Upon reaching an optical density of ~1.3 at 600 nm, DOT1L expression was induced by adding 0.2 mM isopropylthiogalactoside at 16 °C for 20 hours. Cells were harvested, lysed, centrifuged at 20,000 rpm for 20 min and the supernatant was collected and subjected to an affinity column chromatography using the glutathione sepharose resin (GE Healthcare). The GST-hDOT1L fusion protein was eluted with 10 mM of glutathione solution, and after desalting (HiTrap, GE

Healthcare), the GST tag was removed by thrombin digestion overnight at 4 °C. DOT1L was purified by chromatography using a glutathione sepharose column and a Superdex 75 gel filtration column with ~80% purity (SDS-PAGE).

hDOT1L(1-351), which is used for crystallization, was sub-cloned from pGEX-KG-hDOT1L(1-472) using 5'-TGGTGGAATTCACATGGGGGAGAAGCTGG-3' and 5'-GACACTCGAGTCAGCTCTTGCTCTCGCGCTG-3' as forward and reverse primers, respectively, and inserted into pGEX-KG vector. The correctness of insert was verified by sequencing. The expression and purification of hDOT1L(1-351) were similarly performed as described above.

Enzyme inhibition assays

PRMT1 and SUV39H1 were purchased from BPS Biosciences (San Diego, CA) and G9a from New England Biolabs (Ipswich, MA). The expression plasmid (pGEX-KG-CARM1) for human CARM1 was obtained from Dr. Qin Feng (Baylor College of Medicine). The expression and purification of CARM1 were similarly carried out as those of hDOT1L.

Determination of K_m values of SAM and substrate. To determine K_m values of SAM, a methyltransferase (minimal amount to produce sufficient activity, ranging from 50 to 100 nM), a saturated concentration of its substrate and an increasing concentration of SAM (ranging from 0.01 to 50 μM) in 20 μL of 20 mM Tris buffer (containing 1 mM EDTA, 0.5 mM DTT and 50 μg/mL BSA, pH = 8.0) were incubated at 30 ºC for 10 min. The reaction was stopped by adding SAH to a final concentration of 100 μM. 15 μL of reaction mixture was then transferred to a small piece of P81 filter paper (Whatman) that binds the substrate, washed three times with 50 mM $NaHCO_3$, dried, and transferred into a scintillation vial containing 2 mL of scintillation cocktail. Radioactivity on the filter paper that corresponds to the amount of ^{3}H-methyl transferred to the substrate was measured using a Beckman LS-6500 scintillation counter. K_m value was obtained by fitting the triplicate experimental data to Michaelis-Menten model in Prism (version 5.0, GraphPad Software, Inc., La Jolla, CA). The

determination of K_m values of the substrates was done in a similar manner by varying the concentration of the substrate. Oligo-nucleosome (from chicken erythrocytes) was used as the substrate for DOT1L and CARM1, histone H4 (New England Biolabs) for PRMT1, and histone H3 peptide (1-21) (Abcam, UK) for G9a and SUV39H1.

Using this method, the K_m values of SAM and nucleosome for DOT1L were determined to be 0.76 and <0.05 μM, respectively, which are similar to those reported in a recent publication (K_m: 0.65 and 0.0086 μM).[7] The K_m values of SAM for CARM1, G9a and SUV39H1 were determined to be 1.6, 14.1 and 27.9 μM, respectively. The K_m value (6 μM) of SAM for PRMT1 is available from the literature.[8]

DOT1L inhibition assay. Human DOT1L(1-472) enzyme assay was performed using 100 nM enzyme, 0.76 μM ^{3}H-SAM (10 Ci/mM; Perkin-Elmer), 1.5 μM oligo-nucleosome in 20 μL of 20 mM Tris buffer (containing 1 mM EDTA, 0.5 mM DTT and 50 μg/mL BSA, pH = 8.0). For inhibition assay, compounds with concentrations ranging from 1 nM to 100 μM were incubated with the enzyme for 10 min before adding [^{3}H]-SAM to initiate the reaction. After 30 min at 30 ºC, the reaction was stopped by adding SAH to a final concentration of 100 μM. 15 μL of reaction mixture was then transferred to a small piece of P81 filter paper that binds histone H3 protein, washed three times with 50 mM $NaHCO_3$, dried, and transferred into a scintillation vial containing 2 mL of scintillation cocktail. Radioactivity on the filter paper was measured using a Beckman LS-6500 scintillation counter. IC_{50} values were obtained by using a dose response curve fitting in Prism (version 5.0). Figure S2 shows representative dose response curves of inhibitors SAH, **1** – **6**. The reported IC_{50}s were the mean values from at least three experiments. K_i values for competitive inhibitors SAH, **1** and **2** were calculated using the Cheng-Prusoff equation $K_i = IC_{50}/(1+[SAM]/K_m)$.

Inhibition assays for other methyltransferases. Enzyme inhibition assays for all other histone methyltransferases were performed similarly as described above, using 50 – 100 nM enzyme, K_m of ^{3}H-SAM, saturated concentration of the substrate (≥10× K_m) in 20 μL of 20 mM Tris buffer (containing 1

mM EDTA, 0.5 mM DTT and 50 μg/mL BSA, pH = 8.0). Data collection and processing were carried out similarly to determine the IC_{50} and/or K_i values, using Prism.

Crystallization and structure determination. The crystallization of human DOT1L(1-351) was carried out as described in the literature.[6] hDot1L(1-351) (25 mg/mL) containing 5 mM of compound **1** was crystallized under the condition of 1.25-1.7 M $(NH_4)_2SO_4$, 0.1 M NaAc (pH 5.3). Data were collected to 2.5 Å using a Rigaku FR-E+ SuperBright X-ray source at Baylor College of Medicine and processed using the program HKL2000.[9] The initial structure was obtained by the program Phaser[10] using the coordinates of 1NW3 as a target. The refinement was carried out using the program CNS,[11] starting with a simulated annealing routine. The final refinement statistics were summarized in Table S1 and the coordinates were deposited into Protein Data Bank as entry 3SR4. Figures 2, S1 and S3 were generated using Maestro,[12] except for Figure S3b using Ligplot.[13]

Protein structural analysis. Protein structure analysis and visualization were performed using Maestro[12] (version 9.1) in Schrödinger suite 2010.[14] PDB files of the crystal structures of histone methyltransferases were imported and prepared using the module "protein preparation wizard" with default settings: water molecules (>3.0 Å away from a ligand) were removed, hydrogen atoms added, ligands (substrate or inhibitor) remained in the protein structure. H-bonds were then optimized and the protein was energy-minimized using OPLS-2005 force field with all heavy atoms fixed.

Complete list of authors for Reference 16

Daigle, S. R.; Olhava, E. J.; Therkelsen, C. A.; Majer, C. R.; Sneeringer, C. J.; Song, J.; Johnston, L.D.; Scott, M.P.; Smith, J. J.; Xiao, Y.; Jin, L.; Kuntz, K. W.; Chesworth, R.; Moyer, M. P.; Bernt, K. M.; Tseng, J. C.; Kung, A. L.; Armstrong, S. A.; Copeland, R. A.; Richon, V. M.; Pollock, R. M. *Cancer Cell*, **2011**, *20*, 53.

References for Supporting Information

(1) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. App. Cryst.* **1993**, *26*, 283.

(2) David, B. L.; Amal, W. *Tetrahedron Lett.*, **2009**, *50*, 3939.

(3) Weller, R. L.; Rajski, S. R. *ChemBioChem* **2006**, *7*, 243.

(4) Adamczyk, M.; Johnson, D. D.; Reddy, R. E. *Tetrahedron: Asymmetry* **1999**, *10*, 775.

(5) Feng, Q.; Wang, H.; Ng, H. H.; Erdjument-Bromage, H.; Tempst, P.; Struhl, K.; Zhang, Y. *Curr. Biol.* **2002,** *12*, 1052.

(6) Min, J.; Feng, Q.; Li, Z.; Zhang, Y.; Xu, R. M. *Cell* **2003,** *112*, 711.

(7) Richon, V. M.; Johnston, D.; Sneeringer, C. J.; Jin, L.; Majer, C. R.; Elliston, K.; Jerva, L. F.; Scott, M. P.; Copeland, R. A. *Chem. Biol. Drug Des.* **2011**, *78*, 199.

(8) Rust, H. L.; Zurita-Lopez, C. I.; Clarke, S.; Thompson, P. R. *Biochemistry* **2011**, *50*, 3332.

(9) Otwinowski Z.; Minor W. In *Macromolecular Crystallography*, Part A, Method Enzymology, Volume 276; Carter, C. W. Jr., Sweet, R. M., Eds.; Academic Press: New York, 1997; pp 307-326.

(10) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Cryst.* **2007**, *40*, 658.

(11) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr. D Biol. Crystallogr.* **1998**, *54*, 905.

(12) Maestro, version 9.1, Schrödinger, LLC, New York, NY, 2010.

(13) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. *Protein Eng.* **1996**, *8*, 127.

(14) Schrödinger suite 2010, Schrödinger, LLC, New York, NY, 2010.