Supporting Information

Discovery of Potent Small-Molecule Inhibitors of MLL Methyltransferase

Ting-Rong Chern[+±†], *Liu Liu*[+], *Elyse Petrunak*[#], *Jeanne A. Stuckey*[#], *Mi Wang*[+], *Denzil Bernard*[+†], *Haibin Zhou*[±], *Shirley Lee*[^], *Yali Dou*[^], *Shaomeng Wang*[+§‡*]

Departments of [+]Medicinal Chemistry, [±]Internal Medicine, [§]Pharmacology, [^]Pathology, and [#]Life Sciences Institute,

University of Michigan, Ann Arbor, MI 48109, USA

* To whom correspondence should be addressed

Phone: 734-615-0362

Fax: (734) 764-2532 (Fax)

E-mail: shaomeng@umich.edu

Experimental section:

In vitro histone methyltransferase (HMT) assay. The HMT assay was performed as described previously.[1] Small molecule inhibitors were tested in reaction buffer consisting of 50 mM Tris–HCl, pH 8.0, 150 mM NaCl, and 10% glycerol. Equal stoichiometry of human MLL1 SET domain, WDR5, RbBP5 and ASH2L were mixed to form the active MLL1 complex at a final concentration of 125 nM. MLL1 complex were mixed with small molecules inhibitors at a final concentration of 10uM and incubated on ice. After 10 minutes the methyltransferase reaction was initiated by adding 1 μM of 3H-SAM and 4 μM of histone H3 (1–21) peptide. The reactions proceeded for 1 h at 22 °C and was terminated by adding 10 μL of reaction mixture onto Whatman P81 ion exchange filter paper. The filter paper was air dried and washed in three 10-min washes in 50 mM sodium bicarbonate pH 9.0 to remove excess SAM. Filter paper was heat-dried for 20 min and placed in 10 mL Ultima Gold scintillation fluid for 3H signal acquisition in the unit of counts per minute (c.p.m.).

Alpha LISA assay

The method for Alpah LISA assay developed by our group has been described.[2]

General Information: The purity of compounds was determined by Waters ACQUITY UPLC and all the final compounds were >95% pure.

Preparation of tert-butyl (S)-4-((((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)methyl)amino)-2-((tert-butoxycarbonyl)amino)butanoate (20)

Sodium triacetoxyborohydride (953.7 mg, 4.5 mmol) was added to a solution of 5'-amino-5'-deoxy-2',3'-o-(1-methylethylidene)-adenosine (918.9 mg mg, 3 mmol) and *t*-butyl (S)-2-[(tert-butoxycarbonyl)amino]-4-oxo-butanoate (819.9 mg, 3 mmol) in DCE and stirred at rt for 4 h. Then saturated sodium bicarbonate solution was added and stirring was continued for 20 min. The reaction mixture was extracted with DCM and the combined organic layers were dried over $MgSO_4$ and concentrated under reduced pressure. The compound was purified by preparative HPLC to give compound **20** (812.6 mg, 48 % yield). ^{1}H NMR (400 MHz, Methanol-d_4) δ 8.41 (s, 1H), 8.40 (s, 1H), 6.37 (d, *J* = 2.2 Hz, 1H), 5.42 (dd, *J* = 6.3, 2.2 Hz, 1H), 5.19 (dd, *J* = 6.3, 3.9 Hz, 1H), 4.55 (dt, *J* = 9.8, 3.5 Hz, 1H), 4.04 (dd, *J* = 9.3, 5.0 Hz, 1H), 3.61 (dd, *J* = 13.1, 9.8 Hz, 1H), 3.51 (dd, *J* = 13.1, 3.3 Hz, 1H), 3.08 (t, *J* = 7.7 Hz, 2H), 2.14 (dt, *J* = 13.3, 4.7 Hz, 1H), 1.96 – 1.83 (m, 1H), 1.64 (s, 3H), 1.46 (s, 10H), 1.43 (s, 8H), 1.42 – 1.40 (m, 3H). ESI-MS *m/z* $[M + H]^+$ calculated = 564.31, found = 564.22.

General method for preparation of compound 1-10. Sodium triacetoxyborohydride (1.5 eq) was added to a solution of compound **20** (1 eq) and the corresponding aldehyde (1.2 eq) in DCE and stirred for 2 h at rt. Then saturated sodium bicarbonate solution was added and stirring was continued for 20 min. The reaction mixture was extracted with DCM. The combined organic layers dried over $MgSO_4$ and concentrated under reduced pressure. Then the crude product was dissolved in water/trifluoroacetic acid solution (1:4) and stirred for 16 h. The solvent was removed *in vacuo* and the compounds were purified directly by preparative HPLC to yield the desired product.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(1-(3-cyclopentylpropyl)azetidin-3-yl)amino)butanoic acid (1)

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.36 (s, 1H), 8.34 (s, 1H), 6.02 (dd, J = 3.9, 1.4 Hz, 1H), 4.70 (t, J = 4.6 Hz, 1H), 4.29 (t, J = 5.9 Hz, 1H), 4.21 (s, 2H), 4.07 (t, J = 6.1 Hz, 1H), 3.93 (s, 3H), 3.16 – 3.08 (m, 3H), 2.83 (t, J = 6.6 Hz, 2H), 2.10 (dq, J = 31.9, 7.2 Hz, 2H), 1.77 (dt, J = 14.4, 8.8 Hz, 4H), 1.70 – 1.44 (m, 8H), 1.19 – 1.03 (m, 3H). ESI-MS *m/z* $[M + H]^+$ calculated = 533.31, found = 533.22.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(2-aminoethyl)amino)butanoic acid (2)

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.45 (s, 1H), 8.43 (s, 1H), 6.07 (d, *J* = 3.6 Hz, 1H), 4.65 (dd, *J* = 5.5, 3.7 Hz, 1H), 4.32 (t, *J* = 5.9 Hz, 1H), 4.26 (d, *J* = 7.4 Hz, 1H), 4.05 (t, *J* = 6.3 Hz, 1H), 3.07 (q, *J* = 4.6 Hz, 4H), 2.99 – 2.85 (m, 4H), 2.17 (td, *J* = 17.3, 14.8, 7.4 Hz, 1H), 2.05 (s, 1H). ESI-MS *m/z* $[M + H]^+$ calculated = 411.20, found = 411.12.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(3-((3-cyclopentylpropyl)amino)cyclobutyl)amino)butanoic acid (3)

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.38 (s, 1H), 8.37 (s, 1H), 6.08 (d, *J* = 3.6 Hz, 1H), 4.71 (dd, *J* = 5.4, 3.6 Hz, 1H), 4.44 (t, *J* = 5.8 Hz, 1H), 4.40 – 4.33 (m, 1H), 4.20 – 4.09 (m, 1H), 4.01 (dd, *J* = 7.8, 5.1 Hz, 1H), 3.80 (tdd, *J* = 7.4, 4.2, 1.7 Hz, 1H), 3.54 (dd, *J* = 14.5, 9.6 Hz, 1H), 3.43 (d, *J* = 14.1 Hz, 1H), 3.28 (t, *J* = 6.9 Hz, 2H), 2.95 – 2.88 (m, 2H), 2.84 – 2.73 (m, 2H), 2.66 – 2.48 (m, 2H), 2.35 – 2.27 (m, 1H), 2.19 – 2.09 (m, 1H), 1.87 – 1.78 (m, 3H), 1.76 – 1.57 (m, 6H), 1.41 (dt, *J* = 11.1, 6.3 Hz, 2H), 1.19 – 1.08 (m, 2H). ESI-MS *m/z* $[M + H]^+$ calculated = 547.33, found = 547.45.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-butylazetidin-3-yl)methyl)amino)butanoic acid (4)

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.45 (s, 1H), 8.42 (d, *J* = 5.4 Hz, 1H), 6.11 (d, *J* = 3.8 Hz, 1H), 4.79 (dt, *J* = 27.8, 4.9 Hz, 1H), 4.68 (t, *J* = 4.1 Hz, 1H), 4.42 (d, *J* = 5.2 Hz, 2H), 4.29 (ddd, *J* = 17.4, 9.5, 4.7 Hz, 1H), 4.20 – 4.08 (m, 1H), 4.03 (dt, *J* = 7.7, 4.0 Hz, 1H), 3.98 – 3.83 (m, 1H), 3.59 – 3.38 (m, 5H), 3.37 (d, *J* = 8.0 Hz, 1H), 3.29 (t, *J* = 8.7 Hz, 1H), 3.17 (d, *J* = 21.3 Hz, 2H), 2.37 (dq, *J* = 15.0, 7.5 Hz, 1H), 2.17 (dd, *J* = 14.5, 6.2 Hz, 1H), 1.58 – 1.45 (m, 2H), 1.38 (q, *J* = 7.9, 7.3 Hz, 2H), 0.97 (t, *J* = 7.2 Hz, 3H). ESI-MS *m/z* $[M + H]^+$ calculated = 493.28, found = 493.11.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(furan-3-ylmethyl)azetidin-3-yl)methyl)amino)butanoic acid (5)

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.36 (dd, *J* = 8.8, 3.8 Hz, 2H), 7.73 (s, 1H), 7.64 – 7.57 (m, 1H), 6.54 (d, *J* = 1.8 Hz, 1H), 6.07 (t, *J* = 3.7 Hz, 1H), 4.69 (dd, *J* = 5.2, 3.7 Hz, 1H), 4.39 (qd, *J* = 9.5, 8.5, 4.3 Hz, 3H), 4.23 (s, 3H), 4.16 (d, *J* = 25.5 Hz, 2H), 3.97 (dd, *J* = 7.8, 4.9 Hz, 3H), 3.27 – 3.14 (m, 3H), 2.28 (dd, *J* = 15.8, 8.2 Hz, 1H), 2.11 – 2.01 (m, 1H), 1.42 (d, *J* = 4.9 Hz, 2H). ESI-MS *m/z* $[M + H]^+$ calculated = 517.24, found = 517.15.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(3-cyclopentylpropyl)azetidin-3-yl)methyl)amino)butanoic acid (6)

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.37 (s, 1H), 8.35 (s, 1H), 6.08 (d, *J* = 3.9 Hz, 1H), 4.71 (t, *J* = 4.5 Hz, 1H), 4.38 (qd, *J* = 7.3, 4.8, 3.9 Hz, 2H), 4.26 (s, 1H), 4.17 – 4.03 (m, 1H), 3.96 (dd, *J* = 7.9, 4.8 Hz, 1H), 3.86 (d, *J* = 22.2 Hz, 1H), 3.36 (d, *J* = 9.1 Hz, 3H), 3.32 – 3.04 (m, 6H), 2.30 (h, *J* = 7.3 Hz, 1H), 2.06 (dq, *J* = 15.1, 5.2 Hz, 1H), 1.85 – 1.73 (m, 3H), 1.69 – 1.60 (m, 2H), 1.59 – 1.48 (m, 4H), 1.34 (q, *J* = 7.2 Hz, 2H), 1.18 – 1.05 (m, 2H). ESI-MS *m/z* $[M + H]^+$ calculated = 547.33, found = 547.45.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(3-phenylpropyl)azetidin-3-yl)methyl)amino)butanoic acid (7)

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.32 (d, J = 2.8 Hz, 2H), 7.34 – 7.28 (m, 2H), 7.24 – 7.19 (m, 3H), 6.08 – 6.03 (m, 1H), 4.71 (dd, J = 5.3, 3.7 Hz, 1H), 4.39 (t, J = 5.8 Hz, 1H), 4.33 (s, 2H), 4.13 – 4.02 (m, 1H), 3.95 (t, J = 6.3 Hz, 2H), 3.27 (s, 4H), 3.14 (s, 5H), 2.68 (t, J = 7.5 Hz, 2H), 2.25 (d, J = 14.3 Hz, 1H), 2.03 (s, 1H), 1.84 (q, J = 7.9 Hz, 2H), 1.44 – 1.29 (m, 1H). ESI-MS *m/z* $[M + H]^+$ calculated = 555.30, found = 555.22.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(3-((3-phenylpropyl)amino)propyl)amino)butanoic acid (8)

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.27 (d, J = 4.1 Hz, 2H), 7.32 – 7.26 (m, 2H), 7.19 (dd, J = 12.6, 7.1 Hz, 3H), 6.04 (d, J = 3.8 Hz, 1H), 4.78 – 4.73 (m, 1H), 4.46 – 4.42 (m, 1H), 3.85 (dd, J = 8.9, 3.8 Hz, 1H), 3.50 (dt, J = 3.3, 1.6 Hz, 1H), 3.42 – 3.39 (m, 1H), 3.30 – 3.25 (m, 1H), 3.15 (dt, J = 3.3, 1.6 Hz, 1H), 2.98 – 2.92 (m, 2H), 2.67 (t, J = 7.0 Hz, 2H), 2.23 – 2.19 (m, 1H), 2.09 – 1.95 (m, 5H), 1.67 (s, 5H). ESI-MS *m/z* $[M + H]^+$ calculated = 543.30, found = 543.21.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(3-(((3-phenylpropyl)amino)methyl)cyclobutyl)amino)butanoic acid (9)

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.29 (d, J = 2.7 Hz, 2H), 7.29 (t, J = 7.5 Hz, 2H), 7.25 – 7.14 (m, 3H), 6.06 (d, J = 3.2 Hz, 1H), 4.70 (t, J = 4.3 Hz, 1H), 4.51 (t, J = 5.9 Hz, 1H), 4.39 (t, J = 8.2 Hz, 1H), 3.85 (d, J = 6.4 Hz, 2H), 3.64 (t, J = 12.2 Hz, 2H), 3.09 (s, 1H), 3.07 (s, 1H), 3.02 – 2.93 (m, 2H), 2.70 (t, J = 7.2 Hz, 2H), 2.61 – 2.54 (m, 2H), 2.24 – 2.16 (m, 1H), 2.04 (ddt, J = 23.9, 19.0, 10.1 Hz, 4H), 1.71 (q, J = 5.4, 3.3 Hz, 4H). ESI-MS *m/z* $[M + H]^+$ calculated = 569.31, found = 569.49.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)(3-(((3-phenylpropyl)amino)methyl)cyclobutyl)amino)butanoic acid (10)

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.30 (s, 2H), 7.79 (s, 1H), 7.29 – 7.19 (m, 3H), 7.16 – 7.08 (m, 2H), 6.06 (d, *J* = 4.1 Hz, 1H), 4.76 – 4.71 (m, 1H), 4.71 – 4.62 (m, 2H), 4.47 – 4.41 (m, 1H), 4.34 (t, *J* = 5.4 Hz, 1H), 4.22 (s, 2H), 3.87 (t, *J* = 6.4 Hz, 1H), 3.27 (d, *J* = 11.6 Hz, 1H), 3.21 (t, *J* = 6.9 Hz, 3H), 3.17 – 3.13 (m, 1H), 2.35 – 2.25 (m, 1H), 2.10 – 2.02 (m, 1H), 1.42 – 1.36 (m, 1H). ESI-MS *m/z* $[M + H]^+$ calculated = 553.26, found = 553.19.

Preparation of tert-butyl (S)-4-((((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)methyl)(azetidin-3-ylmethyl)amino)-2-((tert-butoxycarbonyl)amino)butanoate (21)

Benzyl 3-formylazetidine-1-carboxylate (263.1 mg, 1.2 mmol) and sodium triacetoxyborohydride (317.91 mg, 1.5 mmol) in DCE were added to a solution of 5'-deoxy-5'-[[(3S)-4-(1,1-dimethylethoxy)-3-[[(1,1-dimethylethoxy)carbonyl]amino]-4-oxobutyl]amino]-2',3'-O-(1-methylethylidene)adenosine (563.7 mg, 1 mmol) and stirred for 3 h at rt. Then saturated sodium bicarbonate solution was added and stirring was continued for 20 min. The reaction mixture was extracted with DCM and the combined organic layers were dried over $MgSO_4$ and concentrated under reduced pressure. Then the crude product was dissolved in MeOH, Pd/C (0.05%) was added and the mixture was stirred under a hydrogen atmosphere. The black particles were filtered off and washed with MeOH several times. The filtrate was then concentrated under reduced pressure and purified by preparative HPLC to give compound **21** (266.2 mg, 42% yield). ^{1}H NMR (400 MHz, Methanol-d_4) δ 8.52 – 8.39 (m, 2H), 6.38 (d, *J* = 2.1 Hz, 1H), 5.36 (d, *J* = 6.3 Hz, 1H), 5.13 (d, *J*

= 5.5 Hz, 1H), 4.64 (s, 1H), 4.25 – 4.08 (m, 2H), 3.97 (s, 3H), 3.60 (s, 2H), 3.54 – 3.36 (m, 3H), 3.25 (s, 1H), 3.18 – 2.92 (m, 2H), 2.13 (d, *J* = 63.5 Hz, 2H), 1.81 (s, 1H), 1.64 (d, *J* = 6.2 Hz, 2H), 1.52 – 1.50 (m, 1H), 1.45 (d, *J* = 1.9 Hz, 15H), 1.40 (s, 3H), 1.32 (d, *J* = 9.8 Hz, 1H). ESI-MS *m/z* $[M + H]^+$ calculated = 633.36, found = 633.28.

General method for synthesis of compound 11-18. Sodium triacetoxyborohydride (1.5 eq) was added to a solution of compound **21** (1 eq) and the corresponding aldehyde (1.2 eq) in DCE and stirred for 2 h at rt. Then saturated sodium bicarbonate solution was added and stirring was continued for 20 min. The reaction mixture was extracted with DCM. The combined organic layers dried over $MgSO_4$ and concentrated under reduced pressure. Then the crude product was dissolved in water/trifluoroacetic acid solution (1:4) and stirred for 16 h. The solvent was removed *in vacuo* and the compounds were purified directly by preparative HPLC to yield the desired product.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-benzylazetidin-3-yl)methyl)amino)butanoic acid (11)

1H NMR (400 MHz, Methanol-d4) δ 8.34 – 8.19 (m, 2H), 7.57 – 7.37 (m, 5H), 6.01 (d, J = 3.8 Hz, 1H), 4.80 – 4.70 (m, 1H), 4.44 – 4.29 (m, 3H), 4.30 – 4.01 (m, 4H), 3.94 (dt, J = 24.7, 9.0 Hz, 2H), 3.80 (s, 1H), 3.22 – 3.13 (m, 1H), 3.11 – 2.83 (m, 5H), 2.23 – 2.11 (m, 1H), 1.97 (dt, J = 14.8, 6.0 Hz, 1H). ESI-MS *m/z* $[M + H]^+$ calculated = 527.27, found = 527.15.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(3-chlorobenzyl)azetidin-3-yl)methyl)amino)butanoic acid (12)

1H NMR (400 MHz, Methanol-d4) δ 8.43 (s, 1H), 8.40 (s, 1H), 7.54 – 7.36 (m, 4H), 6.10 (d, J = 3.7 Hz, 1H), 4.67 (dd, J = 4.8, 3.7 Hz, 1H), 4.44 – 4.34 (m, 4H), 4.29 – 4.14 (m, 2H), 4.12 – 3.96 (m, 3H), 3.44 (tdd, J = 20.2, 9.2, 6.4 Hz, 5H), 3.26 (dq, J = 13.2, 6.8, 6.2 Hz, 2H), 2.34 (dq, J =

14.9, 7.4 Hz, 1H), 2.13 (dq, J = 14.7, 5.8 Hz, 1H). ESI-MS *m/z* $[M + H]^+$ calculated = 561.23, found = 561.12.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(2-chlorobenzyl)azetidin-3-yl)methyl)amino)butanoic acid (13)

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.41 (s, 1H), 8.38 (s, 1H), 7.52 (ddd, *J* = 7.4, 3.0, 1.6 Hz, 2H), 7.42 (dtd, *J* = 23.7, 7.4, 1.5 Hz, 2H), 6.07 (d, *J* = 3.6 Hz, 1H), 4.62 (dd, *J* = 4.9, 3.5 Hz, 1H), 4.54 (s, 2H), 4.36 (dd, *J* = 5.7, 3.1 Hz, 2H), 4.32 – 4.27 (m, 1H), 4.22 (td, *J* = 8.8, 7.8, 4.9 Hz, 1H), 4.11 (ddd, *J* = 19.8, 10.6, 7.3 Hz, 2H), 3.99 (dd, *J* = 7.7, 5.2 Hz, 1H), 3.47 – 3.34 (m, 5H), 3.25 (dd, *J* = 13.8, 6.9 Hz, 2H), 2.30 (dt, *J* = 14.9, 7.4 Hz, 1H), 2.13 (ddd, *J* = 11.8, 9.7, 6.0 Hz, 1H). ESI-MS *m/z* $[M + H]^+$ calculated = 561.23, found = 561.12.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(4-chlorobenzyl)azetidin-3-yl)methyl)amino)butanoic acid (14)

^{1}H NMR (400 MHz, Methanol-d_4) δ 8.43 (s, 1H), 8.40 (s, 1H), 7.49 – 7.43 (m, 4H), 6.10 (d, *J* = 3.7 Hz, 1H), 4.67 (dd, *J* = 4.9, 3.6 Hz, 1H), 4.43 – 4.34 (m, 4H), 4.26 – 4.20 (m, 1H), 4.16 (td, *J* = 8.5, 7.6, 4.5 Hz, 1H), 4.10 – 3.98 (m, 3H), 3.42 (ddd, *J* = 20.3, 9.0, 4.5 Hz, 5H), 3.24 (dq, *J* = 13.3, 6.4 Hz, 2H), 2.33 (dd, *J* = 14.9, 7.4 Hz, 1H), 2.16 – 2.08 (m, 1H). ESI-MS *m/z* $[M + H]^+$ calculated = 561.23, found = 561.10.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(thiophen-2-ylmethyl)azetidin-3-yl)methyl)amino)butanoic acid (15)

1H NMR (400 MHz, Methanol-d4) δ 8.34 (s, 1H), 8.32 (s, 1H), 7.62 (s, 1H), 7.57 (dd, J = 5.0, 3.1 Hz, 1H), 7.16 (d, J = 5.2 Hz, 1H), 6.05 (d, J = 3.9 Hz, 1H), 4.70 (dd, J = 5.2, 3.5 Hz, 1H), 4.48 – 3.89 (m, 9H), 3.31 – 3.07 (m, 6H), 2.35 – 2.19 (m, 1H), 2.09 – 1.97 (m, 1H), 1.32 (s, 1H). ESI-MS *m/z* $[M + H]^+$ calculated = 533.22, found = 533.10.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(thiophen-2-ylmethyl)azetidin-3-yl)methyl)amino)butanoic acid (16)

1H NMR (400 MHz, Methanol-d4) δ 8.33 (d, J = 5.3 Hz, 2H), 7.60 (dd, J = 5.1, 1.2 Hz, 1H), 7.30 – 7.23 (m, 1H), 7.13 (dd, J = 5.1, 3.5 Hz, 1H), 6.05 (d, J = 3.7 Hz, 1H), 4.69 (dd, J = 5.4, 3.7 Hz, 1H), 4.57 (s, 2H), 4.38 (t, J = 5.9 Hz, 1H), 4.35 – 4.29 (m, 1H), 4.23 (t, J = 8.9 Hz, 1H), 4.14 (t, J = 9.3 Hz, 1H), 3.99 (dt, J = 22.6, 7.7 Hz, 3H), 3.30 – 3.08 (m, 7H), 2.25 (dd, J = 14.9, 7.3 Hz, 1H), 2.08 – 1.96 (m, 1H). ESI-MS *m/z* $[M + H]^+$ calculated = 533.22, found = 533.13.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(benzofuran-3-ylmethyl)azetidin-3-yl)methyl)amino)butanoic acid (17)

1H NMR (400 MHz, Methanol-d4) δ 8.34 (s, 1H), 8.32 (s, 1H), 7.66 (dt, J = 7.8, 1.0 Hz, 1H), 7.53 (dd, J = 8.3, 1.0 Hz, 1H), 7.39 (ddd, J = 8.5, 7.2, 1.4 Hz, 1H), 7.30 (td, J = 7.5, 1.0 Hz, 1H), 7.06 (s, 1H), 6.05 (d, J = 3.9 Hz, 1H), 4.70 (dd, J = 5.2, 3.8 Hz, 1H), 4.60 (s, 2H), 4.40 – 4.22 (m, 5H), 4.14 (dd, J = 10.8, 7.1 Hz, 1H), 4.08 (dd, J = 10.7, 7.0 Hz, 1H), 3.94 (dd, J = 7.6, 5.1 Hz, 1H), 3.31 – 3.20 (m, 4H), 3.17 – 3.11 (m, 2H), 2.25 (dt, J = 14.8, 7.5 Hz, 1H), 2.08 – 1.98 (m, 1H). ESI-MS *m/z* $[M + H]^+$ calculated = 567.26, found = 567.14.

(S)-2-amino-4-((((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl)((1-(3,3-diphenylpropyl)azetidin-3-yl)methyl)amino)butanoic acid (18)

1H NMR (400 MHz, Methanol-d4) δ 8.39 (s, 1H), 8.37 (s, 1H), 7.32 – 7.26 (m, 8H), 7.20 (ddd, J = 8.6, 5.6, 2.3 Hz, 2H), 6.07 (d, J = 3.7 Hz, 1H), 4.65 (dd, J = 4.9, 3.7 Hz, 1H), 4.42 – 3.71 (m, 9H), 3.41 – 3.33 (m, 3H), 3.27 – 3.06 (m, 5H), 2.28 (dq, J = 13.8, 7.5 Hz, 3H), 2.06 (dt, J = 15.3, 5.4 Hz, 1H). ESI-MS *m/z* $[M + H]^+$ calculated = 631.33, found = 631.20.

Cloning, Expression, Purification of MLL1 SET IL Mutant

All crystallization experiments were carried out with a construct of the MLL1 SET domain 3813-3969 containing N3861I and Q3867L mutations (MLL1 SET^{IL}) and expressed with an N-terminal 6xHis-SUMO tag. *Escherichia coli* Rosetta 2 (DE3) cells transformed with pET28a containing the gene for MLL1 SET^{IL} were grown in Terrific Broth supplemented with 30 mg/mL kanamycin and 100 µM zinc acetate. Expression was induced by addition of 0.4 mM IPTG once cells reached an OD_{600} of 1.0. Cells were incubated at 20 °C for 18 hours, harvested by centrifugation at 6700 xg and frozen at -80 °C.

MLL1 SET^{IL} was purified by resuspending cells from 1 L expression media in 40 mL lysis buffer containing 50 mM Tris-HCl pH 8.0, 400 mM NaCl, 10% glycerol, 10 mM $MgCl_2$, 2 µL Benzonase Nuclease (EMD Millipore), 10 µg/mL aprotinin, 1 µg/mL leupeptin, and 0.1% β-mercaptoethanol. Cells were lysed by sonication and lysate cleared by centrifugation at 34000 xg. Lysate was incubated with Ni-NTA (Qiagen) resin at 4 °C for 1 hour and washed with buffer containing 50 mM Tris-HCl pH 8.0, 400 mM NaCl, 10% glycerol, and 10 mM imidazole. Protein was eluted in buffer 50 mM Tris-HCl pH 8.0, 400 mM NaCl, 10% glycerol, 300 mM imidazole. Eluted protein was dialyzed in 50 mM Tris-HCl pH 8.0, 400 mM NaCl, 10% glycerol, 1 mM DTT, and the 6xHis-SUMO tag was cleaved by digestion with SUMO Protease ULP1 overnight, and removed by incubation with Ni-NTA resin. MLL1 SET^{IL} was further purified by size exclusion chromatography on a HiLoad 26/60 Superdex 75 prep grade column (GE Healthcare) pre-

equilibrated with 50 mM Tris-HCl pH 8.0, 300 mM NaCl, and 10% glycerol. Purified protein was concentrated to 15 mg/mL and stored at -80 °C.

Crystallization and Structure Determination of MLL1 SET^{IL} Mutant

Concentrated MLL1 SET^{IL} was incubated with 2 mM S-adenosyl-L-homocysteine at 4 °C for 1 hour and crystallized by sitting drop vapor diffusion at 4 °C. Crystals formed in drops containing a 1:1 ratio of protein to well solution containing 2.1-2.2 M sodium malonate pH 7.0. Crystals were transferred into a soak solution containing 2.1-2.2 M sodium malonate pH 7.0, 20% ethylene glycol, and 10 mM inhibitor. After 24 hours, crystals were removed from the soak solution and flash-frozen in liquid nitrogen. Diffraction data were collected on the Advanced Photon Source LS-CAT beamline 21-ID-G (Table S1) and processed with HKL2000[3]. Structures with inhibitors were solved by molecular replacement in Phaser[4] using the published structure of MLL SET^{IL} (PDB ID: 5F5E)[5] as a search model absent its ligand. Iterative model building and refinement were performed using COOT[6] and BUSTER[7], respectively. Ligand coordinates and restraints were generated in Grade[8].

The structure of MLL1 SET^{IL} with inhibitors TC-5109, TC-5139, TC-5140, and TC-5153 were solved in the space group P32 with two chains of protein per asymmetric unit. There was a high degree of similarity between the two polypeptide chains (average RMSD based on SSM superposition in COOT of 0.2 ± 0.1 Å) and between the structures themselves (average RMSD based on SSM superposition in COOT of 0.34 ± 0.07 Å).

Table S1: Crystallography Data Collection and Refinement Statistics

Data Collection	**TC-51**	**TC-5139**	**TC-5140**	**TC-5153**
PDBID	6U9M	6U9N	6U9R	6U9K
Space Group	P32	P32	P32	P32
Unit Cell (Å)	a = b = 54.810 c = 105.827 α = β = 90° γ = 120°	a = b = 54.629 c = 105.832 α = β = 90° γ = 120°	a = b = 54.692 c = 105.530 α = β = 90° γ = 120°	a = b= 54.949 c = 104.895 α = β = 90° γ = 120°
Wavelength (Å)	0.9786	0.9786	0.9786	0.9786
Resolution (Å)[1]	2.05 (2.09 - 2.05)	1.95 (1.98 - 1.95)	2.10 (2.14 - 2.10)	2.00 (2.03-2.00)
Rsym[2]	0.066 (0.687)	0.046 (0.574)	0.066 (0.819)	0.050 (0.559)
$\langle I/\sigma I \rangle$[3]	10 (2)	10 (2)	10 (2)	10 (2)
Completeness (%	100 (100)	100 (100)	100 (100)	100 (100)
Redundancy	7.8 (7.7)	7.8 (7.7)	8.7 (8.8)	7.8 (7.7)
Refinement				
Resolution (Å)	2.05	1.95	2.1	2.0
R-Factor [5]	22.69	22.69	21.45	20.69
Rfree [6]	24.81	24.81	25.17	24.41
Protein atoms	2293	2237	2264	2293
Ligands	2	2	2	2
Water Molecules	164	191	146	167
Unique Reflectior	22301	25710	20546	23928
R.m.s.d.[7]				
Bonds	0.008	0.010	0.010	0.010
Angles	1.00	1.05	1.04	1.02
MolProbity Score	1.67	1.63	1.83	1.59
Clash Score[8]	3.92	3.40	3.78	2.18

RSCC[9]	0.88/0.91	0.88/0.88	0.84/0.87	0.92/0.92
RSR[9]	0.17/0.19	0.17/0.17	0.20/0.17	0.14/0.15

[1]Statistics for highest resolution bin of reflections in parentheses.

[2]$R_{sym} = \Sigma_h \Sigma_j | I_{hj} - \langle I_h \rangle | / \Sigma_h \Sigma_j I_{hj}$, where I_{hj} is the intensity of observation j of reflection h and $\langle I_h \rangle$ is the mean intensity for multiply recorded reflections.

[3]Intensity signal-to-noise ratio.

[4]Completeness of the unique diffraction data.

[5]R-factor = $\Sigma_h | IF_oI - IF_cI | / \Sigma_h |F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes for reflection h.

[6]R_{free} is calculated against a 5% random sampling of the reflections that were removed before structure refinement.

[7]Root mean square deviation of bond lengths and bond angles.

[8]Chen et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica D66:12-21.

[9]wwPDB Validation Server.

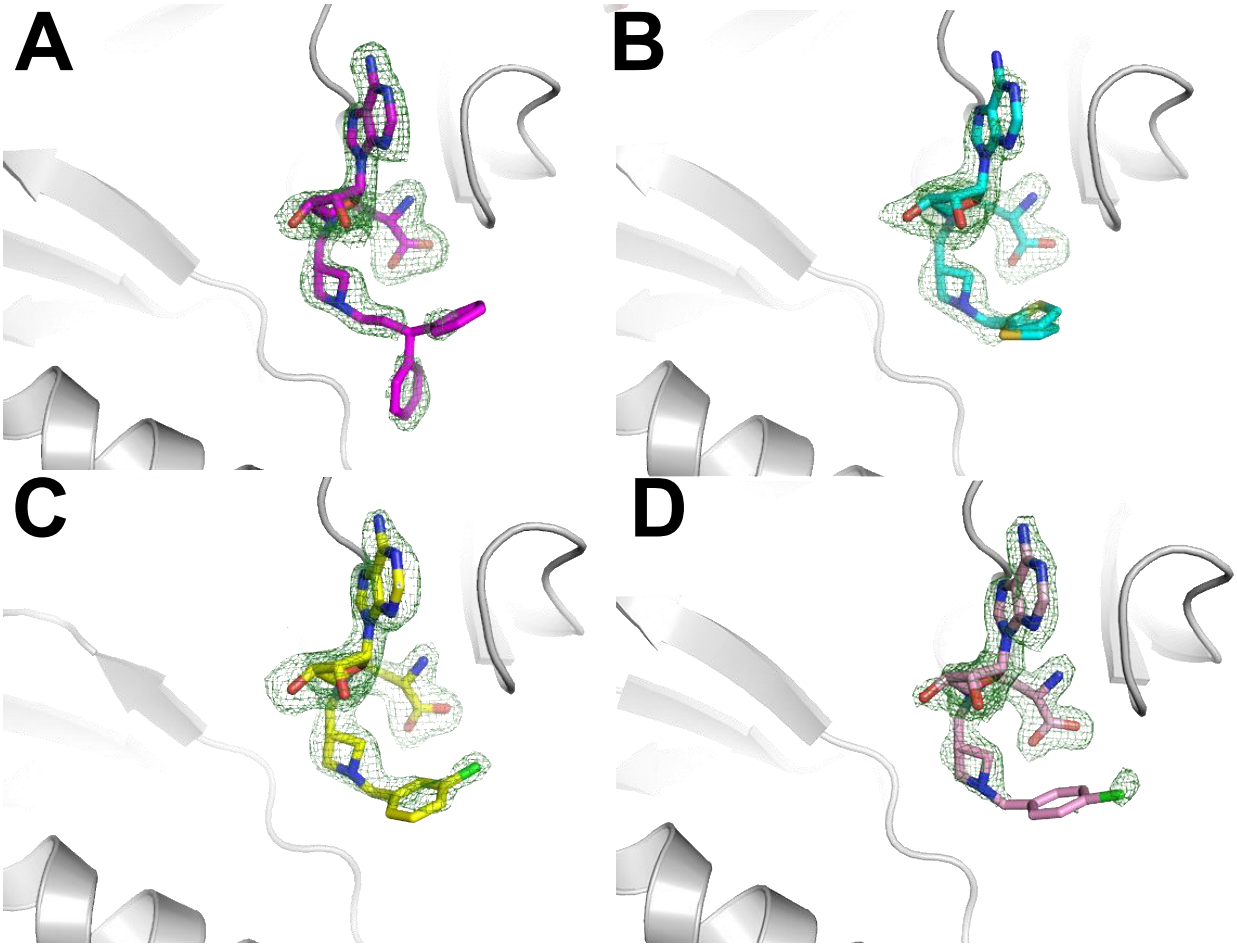

Figure S1: Ligand omit density (green mesh) for inhibitors TC-5153 (panel A, magenta sticks), TC-5109 (panel B, cyan sticks), TC-5140 (panel C, yellow sticks), and TC-5139 (panel D, pink sticks). Fo-Fc density shown is contoured to 3σ. Electron density associated with the thiophenyl substituent of TC-5109 does not confirm a single orientation of the thiophene ring, and it is thus shown modeled in two plausible orientations.

REFERENCE

(1) Dou, Y.; Milne, T. A.; Tackett, A. J.; Smith, E. R.; Fukuda, A.; Wysocka, J.; Allis, C. D.; Chait, B. T.; Hess, J. L.; Roeder, R. G. Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **2005**, *121*, 873-885.

(2) Karatas, H.; Li, Y.; Liu, L.; Ji, J.; Lee, S.; Chen, Y.; Yang, J.; Huang, L.; Bernard, D.; Xu, J.; Townsend, E. C.; Cao, F.; Ran, X.; Li, X.; Wen, B.; Sun, D.; Stuckey, J. A.; Lei, M.; Dou, Y.; Wang, S. Discovery of a Highly Potent, Cell-Permeable Macrocyclic Peptidomimetic (MM-589) Targeting the WD Repeat Domain 5 Protein (WDR5)-Mixed Lineage Leukemia (MLL) Protein-Protein Interaction. *J Med Chem* **2017**, *60*, 4818-4839.

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

(4) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J Appl Crystallogr* **2007**, *40*, 658-674.

(5) Li, Y.; Han, J.; Zhang, Y.; Cao, F.; Liu, Z.; Li, S.; Wu, J.; Hu, C.; Wang, Y.; Shuai, J.; Chen, J.; Cao, L.; Li, D.; Shi, P.; Tian, C.; Zhang, J.; Dou, Y.; Li, G.; Chen, Y.; Lei, M. Structural basis for activity regulation of MLL family methyltransferases. *Nature* **2016**, *530*, 447-452.

(6) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486-501.

(7) Bricogne, G.; Blanc, E.; Brandl, M.; Flensburg, C.; Keller, P.; Paciorek, W.; Roversi, P.; Sharff, A.; Smart, O. S.; Vonrhein, C.; Womack, T. O.: BUSTER. 2.10.3 ed.; Global Phasing Ltd.: Cambridge, United Kingdom, 2017.

(8) Smart, O. S.; Womack, T. O.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G.: GRADE. 1.2.13 ed.; Global Phasing Ltd.: Cambridge, United Kingdom, 2017.