Structural Biology of Human H3K9 Methyltransferases

Hong Wu[1], Jinrong Min[1], Vladimir V. Lunin[1], Tatiana Antoshenko[1], Ludmila Dombrovski[1], Hong Zeng[1], Abdellah Allali-Hassani[1], Valérie Campagna-Slater[1], Masoud Vedadi[1], Cheryl H. Arrowsmith[1,2], Alexander N. Plotnikov[1¤]*, Matthieu Schapira[1,3]*

1 Structural Genomics Consortium, University of Toronto, Toronto, Ontario, Canada, 2 Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada, 3 Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada

Abstract

SET domain methyltransferases deposit methyl marks on specific histone tail lysine residues and play a major role in epigenetic regulation of gene transcription. We solved the structures of the catalytic domains of GLP, G9a, Suv39H2 and PRDM2, four of the eight known human H3K9 methyltransferases in their apo conformation or in complex with the methyl donating cofactor, and peptide substrates. We analyzed the structural determinants for methylation state specificity, and designed a G9a mutant able to tri-methylate H3K9. We show that the I-SET domain acts as a rigid docking platform, while induced-fit of the Post-SET domain is necessary to achieve a catalytically competent conformation. We also propose a model where long-range electrostatics bring enzyme and histone substrate together, while the presence of an arginine upstream of the target lysine is critical for binding and specificity.

Enhanced version: This article can also be viewed as an enhanced version in which the text of the article is integrated with interactive 3D representations and animated transitions. Please note that a web plugin is required to access this enhanced functionality. Instructions for the installation and use of the web plugin are available in Text S1.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: matthieu.schapira@utoronto.ca (MS); alexander.plotnikov@mssm.edu (ANP)

¤ Current address: Department of Structural and Chemical Biology, Icahn Medical Institute, Mount Sinai School of Medicine, New York, New York, United States of America

Introduction

Post-translational modifications of histone proteins regulate chromatin compaction, mediate epigenetic regulation of transcription, and control cellular differentiation in health and disease [1,2]. Methylation of histone tails is one of the fundamental events of epigenetic signaling [3]. Tri-methylation of lysine 9 of histone 3 (H3K9) mediates chromatin recruitment of HP1, heterochromatin condensation and gene silencing [4,5]. Similarly, methylation of H3K27 and H4K20 are associated with a repressed state of chromatin, whereas expressed genes are methylated at H3K4, H3K36 and H3K79 ([3,6] for review). Histone methyltransferases are divided into protein arginine methyltransferases (PRMTs) and histone lysine methyltransferases (HKMTs). HKMTs catalyze the transfer of a methyl group from the co-factor S-adenosyl-L-methionine (SAM) to a substrate lysine and, with the exception of DOT1L, are all organized around a canonical SET domain [7,8]. The structures of a number of HKMTs have been reported, including ternary complexes of human orthologs with co-factor and substrate peptides (SETD7-H3K4, SETD8-H4K20 and MLL1-H3K4 [9,10,11,12]), as well as *N. crassa* Dim-5 in complex with a H3K9 peptide [13] and a viral protein complexed to

H3K27 [14] (Figure 1). These structures collectively highlighted a remarkable plasticity of the peptide-binding site and a lack of clear structural motifs that correlate with sequence selectivity [8,15].

Methylation of H3K9 in humans relies mostly on members of the Suv39 family, namely EHMT1/GLP, EHMT2/G9a, SUV39H1, SUV39H2, SETDB1 and SETDB2, as well as then non-Suv39 enzymes PRDM2 and ASH1L [6] (Figure 1). Here we report the high-resolution crystal structures of the methyltransferase domains of GLP, G9a, SUV39H2 and PRDM2, and propose a structural mechanism for substrate recognition. Our data also provide important insight to guide the development of potent and selective inhibitors of HKMTs, which are likely to have applications in a variety of diseases including regenerative medicine, oncology and inflammation [16,17,18,19,20] ([21,22] for review).

Results and Discussion

Overall Structure

We have solved the crystal structures of the catalytic domain of four H3K9 methyltransferases: (1) the complexes of GLP/EHMT1 co-crystallized with the co-factor product S-adenosyl-L-
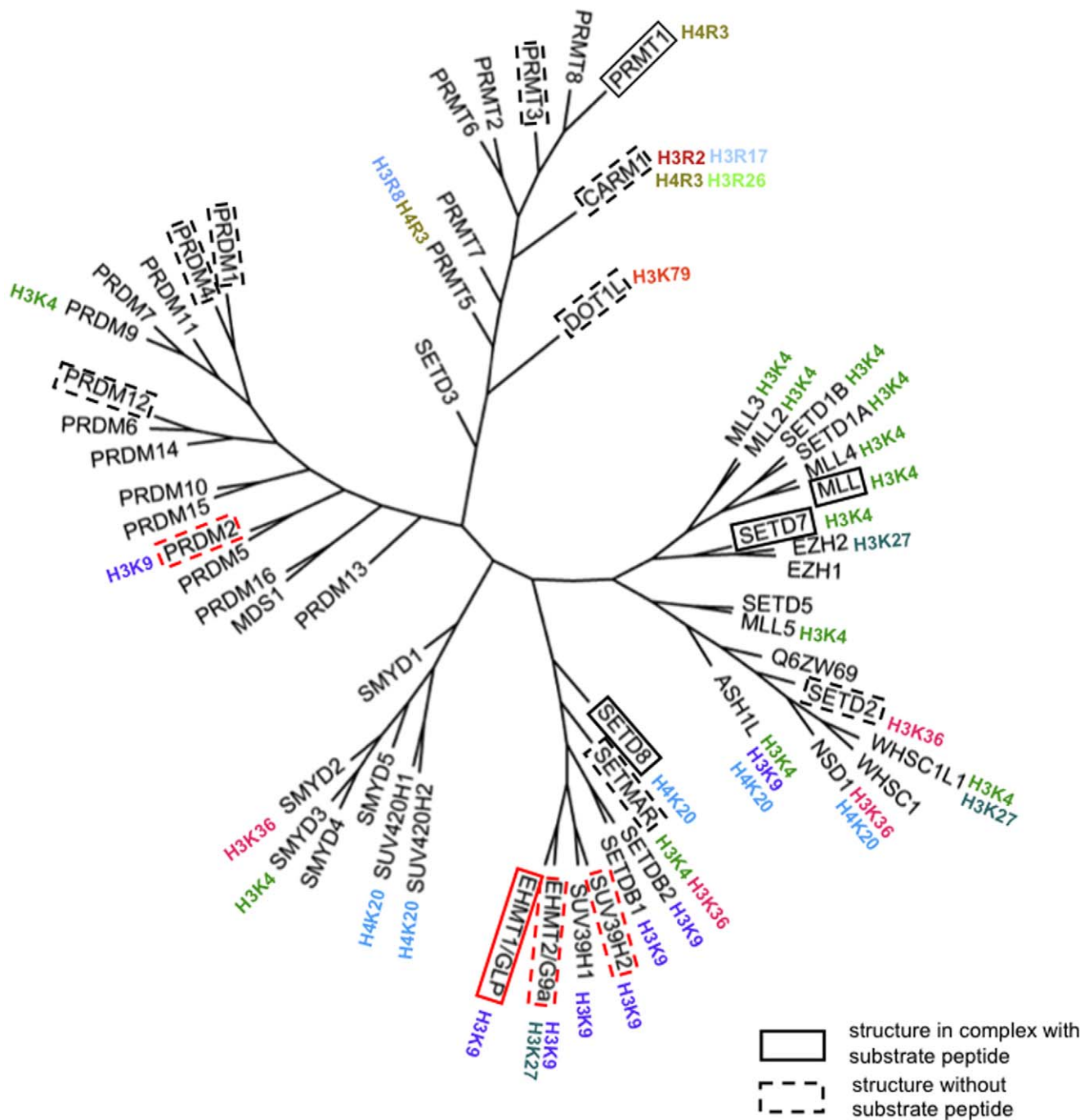
Figure 1. Phylogenetic tree of human histone methyltransferases. Phylogeny is based on a multiple sequence alignment of the methyltransferase domain including the N-SET, Pre_SET, SET, I-SET, and Post-SET motifs. Substrate selectivity was extracted from Kouzarides [1]. Enzymes with solved structure are highlighted by a frame, dotted if no peptide substrate complex is available. Structures solved in the present work are framed in red.

doi:10.1371/journal.pone.0008570.g001

homocysteine (SAH) alone (accession code 2IGQ) and with an H3K9me (accession code 3HNA) or H3K9me2 peptide (accession code 2RFI), (2) G9a/EHMT2 (accession code 2O8J) and SUV39H2 (accession code 2R3A) in complex with SAH and SAM respectively, and (3) the non Suv39 protein PRDM2 (accession code 2QPW). The Suv39 structures G9a, GLP and SUV39H2 adopt a typical fold composed of a conserved SET domain and variable I-SET insert, flanked by Pre- and Post-SET

regions, and characterized by canonical features such as a pseudo-knot next to the catalytic site, distinct co-factor and substrate binding areas meeting at the site of methyl transfer, and a narrow substrate lysine docking channel [8,23,24]. The four structures are also characterized by the presence of an N-SET region located N-terminal to the Pre-SET, that wraps around the core SET domain (Figure 2A–D). The H3K9 peptide lies in a groove formed by the I-SET and Post-SET domains (Figure 2A), as previously observed
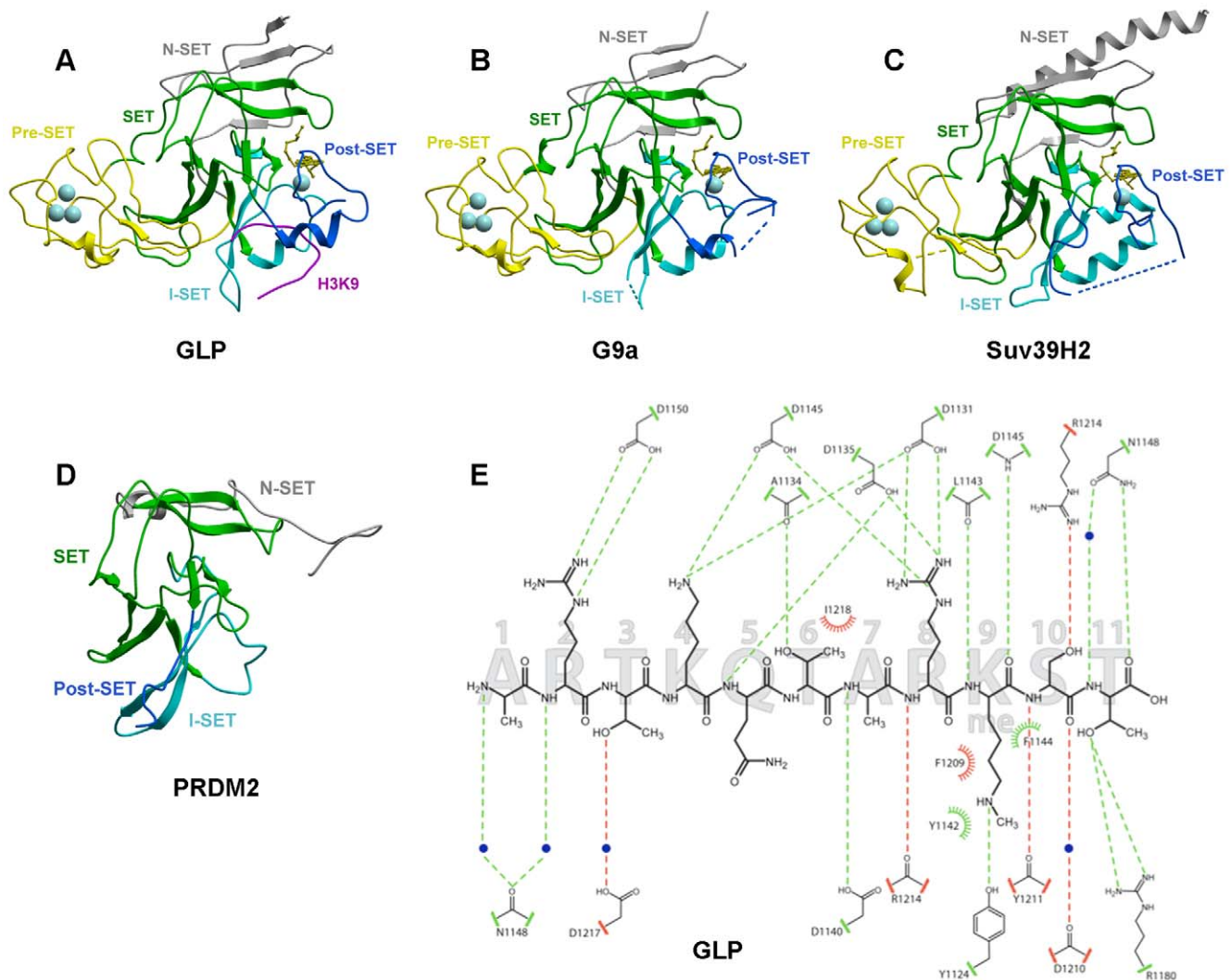
Figure 2. Structures of four human H3K9 HKMTs. (A) a ternary complex of GLP with SAH and a H3K9Me substrate peptide, (B) G9a in complex with SAH, (C) SUV39H2 in complex with SAM and (D) PRDM2, highlighting Pre-Set, SET, I-SET, Post-SET domains and the conserved presence of an N-SET domain. The co-factor is shown as yellow sticks. Residues flanking un-resolved regions are connected by dotted lines. (E) The detail of the interactions between GLP and an H3K9Me peptide.
doi:10.1371/journal.pone.0008570.g002

in other HKMTs ([8] for review), and makes extensive contact with the enzyme through both backbone and side-chain interactions (Figure 2E).
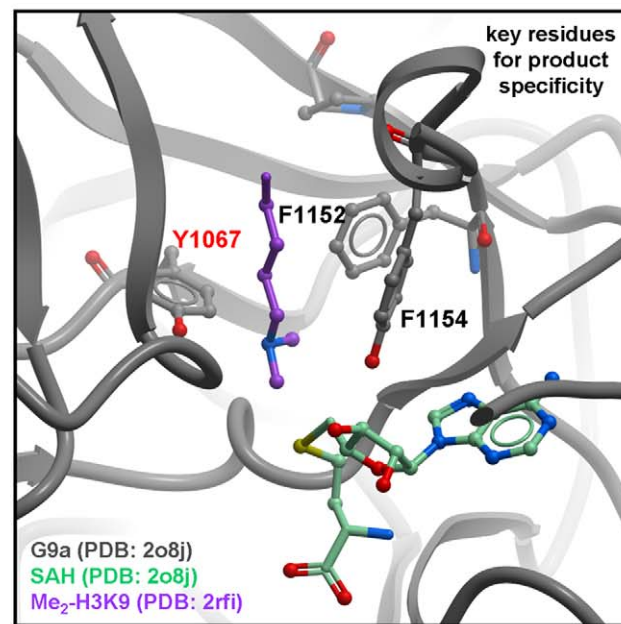
Methylation State Specificity

Mono-, di- or tri- methylation of H3K9 constitute distinct biochemical signals and are established by distinct histone methyltransferases; G9a and GLP are mono- and di-methylases, and SUV39H2 di- and tri-methylates a mono-methylated substrate. The specificity of PRDM2 is unknown [25]. Several aromatic residues line the GLP channel occupied by the substrate lysine leading to the catalytic site and contribute to methylation state specificity as previously noted for several other HKMTs [9,12,26,27]. Two residues are of particular importance for catalysis. First, a conserved tyrosine residue of the post-SET domain (Y1211 in GLP and Y1154 in G9a) is a major component of the lysine binding channel while its hydroxyl group participates in catalysis. This residue cannot be mutated without losing catalytic activity (Figure 3) [8,28]. Second, the hydroxyl group of

GLP's Y1124 (Y1067 in G9a) hydrogen-bonds to the methyl-accepting nitrogen, thereby inhibiting an orientation of the dimethyl-amine that would favor transfer of a methyl from SAM, as was previously shown for SETD7 ([12,28] for review). To confirm this model, we showed that, unlike wild-type G9a, the Y1067F mutant is able to tri-methylate H3K9 (Figure 3). Similarly, it was shown that the F1152Y G9a mutant can only mono-methylate H3K9 [29]. Our structures show that this residue is perfectly superimposed with GLP's F1209 (0.1 Å RMSD), which, if mutated to Tyr, would hydrogen-bond with the ε-amine nitrogen of H3K9 and impair the alignment of the accepting amine's lone pair with the methyl-sulfur bond of SAM (Figure 3). Thus, the methylation state selectivity appears to be inversely proportional to the number of tyrosine residues surrounding the methyl accepting nitrogen.

I-SET Is a Rigid Peptide-Docking Platform

The structure of the I-SET domain appears relatively conserved, whether in the apo, co-factor-, or peptide-bound form,

Figure 3. Structural determinants of G9a mono/di-methylation specificity. A model of substrate lysine-bound G9a (Top panel) was generated from the superposition of the active sites of the GPL ternary complex and GLP. Y1067 of G9a stabilizes the di-methylamine end of the substrate lysine in an orientation where the lone-pair is not facing the co-factor, thereby disfavoring transfer of a third methyl group. The Y1067F mutant loses this restriction and can tri-methylate its substrate, as indicated in the table. Previous work had shown that the F1152Y G9a mutant can only mono-methylate H3K9 [29].

doi:10.1371/journal.pone.0008570.g003

peptide	G9a-Y1067F			G9a		
	Km (µM)	kcat (h⁻¹)	kcat/Km (h⁻¹.mM⁻¹)	Km (µM)	kcat (h⁻¹)	kcat/Km (h⁻¹.mM⁻¹)
H3K9	115 ± 9	55 ± 4	480	5 ± 0.1	65 ± 5	13000
K3K9Me1	240 ± 15	53 ± 2	220	9 ± 1	56 ± 4	6200
K3K9Me2	75 ± 3	90 ± 3	1200	NA	NA	NA

and is composed of a helix followed by a two-stranded anti-parallel β-sheet, linked by loops of variable lengths (Figure 4). Superimposition of our six new H3K9 methyltransferase structures (see Materials and Methods) shows that in all cases the first β-strand is in a conformation which would preserve the pair of backbone hydrogen-bonds observed between Lys-9 substrate and strand-1 of I-SET in the GLP-peptide complex (Figure 4). Furthermore, a systematic comparison of the ternary structures presented here for GLP-H3K9 and previously published for Dim-5-H3K9, SETD8-H4K20, SETD7-H3K4, SETD7-TAF10, SETD7-P53, vSET-H3K27 reveals that this pair of hydrogen bonds between the backbone of a single substrate peptide residue and the first strand of I-SET is observed (1) in all HKMT ternary structures to date and (2) always and only at the substrate lysine (Figure 5A–E - SETD7-TAF10 and SETD7-P53 complexes not shown). This suggests that an evolutionary pressure enforces conservation of this "double hydrogen-bond", which likely plays an important role in the binding mechanism, probably by imposing the proper orientation of the peptide when lysine inserts into the active site; flipping the substrate by 180° in its groove would not allow formation of the double hydrogen bond.

Our structures single-out residue R-1 of H3 (the methylated lysine is used as reference position 0 for peptide residue numbering throughout the text) as the major contributor to the interaction after the substrate lysine itself, with four direct hydrogen-bonds between the arginine guanidinium group and GLP (Figure 2E). This is in agreement with recent mutational analysis showing that no substitution at position -1 is tolerated by G9a [30]. This critical interaction takes place exclusively with I-SET residues (Figure 5F). GLP residues that contribute to substrate binding are conserved in G9a (Figure S1), and it is reasonable to assume that the peptide binding mode observed in GLP holds for G9a. On the other hand, mutation of R-1 to alanine only mildly affects peptide binding to Dim-5, a H3K9 methyltransferase in *N. crassa* [31], indicating that the selectivity mechanism observed for human GLP and G9a is not universal.

Intriguingly, our structures of GLP in complex with mono- or di-methylated susbtrate peptides reveal that residue H3K4 is making 2 hydrogen-bonds with the D1131 and D1145 side-chains of the I-SET domain (Figure 6), which suggests that H3K4 methylation may lower binding affinity, and reduce H3K9 methylation efficiency. We tested this hypothesis, and observed a
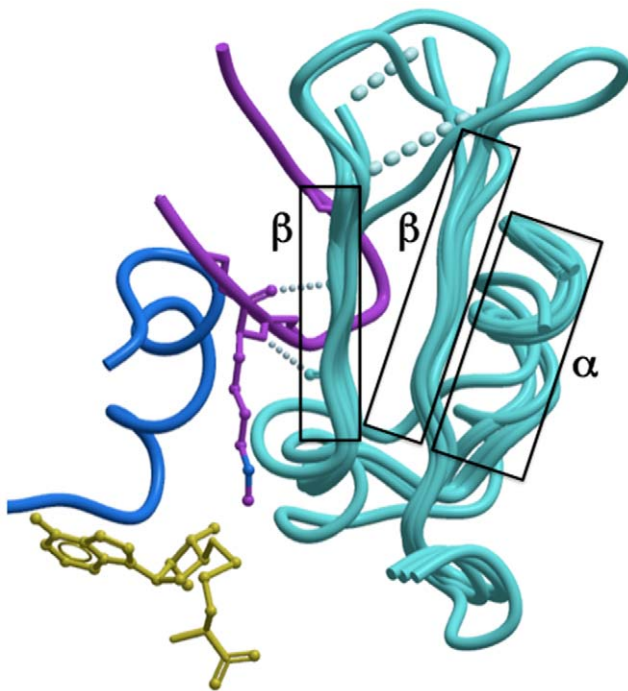
Figure 4. The I-SET domain is relatively rigid and structurally conserved. Structural superimposition of the ternary GLP structure with G9a or Suv39H2 in complex with co-factor and with the apo-structure of PRDM2 shows that the I-SET (cyan) conformation is conserved. The backbone atoms engaged in a double hydrogen-bond with the substrate lysine observed in all available HKMT-peptide complexes are already positioned in the absence of peptide or co-factor.
doi:10.1371/journal.pone.0008570.g004

mild decrease of 43% in affinity of GLP for a H3K9 peptide trimethylated at lysine 4, and similar reduction in enzymatic efficiency, while Kcat was unaffected. Mono and di-methylation of H3K4 had no or very limited effect (data not shown). It is not clear whether this variation is biologically significant.

These results suggest a model in which a mostly pre-formed I-SET domain acts as a receiving platform for the histone 3 tail. Binding includes a conserved pair of hydrogen-bonds with the backbone of the substrate lysine, and critical contacts with a basic side-chain upstream of the methyl acceptor.

Mobile Post-SET Domain Closes onto the Peptide Substrate

The Post-SET domains of G9a, GLP, SUV39H2, but not PRDM2 include a ZnCys motif previously observed in the structures of Dim-5 [13] and the H3K4 methyltransferase MLL1 [10]. The Post-SET domains of G9a and GLP present an α-helix that contributes to peptide binding where other HKMTs have a loop. Unlike I-SET, Post-SET is absent from the PRDM2 structure, which lacks co-crystallized SAM or SAH (Figure 2D). It is partially folded in the structures of G9a and SUV39H2 in complex with SAH and SAM respectively (Figure 2B–C), and fully ordered in the ternary complexes of GLP with SAH and H3K9 peptide (Figure 2A). As previously observed with other HKMTs ([9,11,8] for review), the co-factor contributes to the formation of a hydrophobic, mostly aromatic cluster (composed of Post-SET Y1211/Y1154/Y261, F1215/F1158, W1216/W1159/L298, F1223/F1166/T285 and SET H1170/H1113/H220 in GLP/G9a/SUV39H2) necessary for partial folding of the Post-SET domain. Surprisingly, in our structure of SUV39H2, Post-SET

Lys-264 is inserted into the partially formed substrate lysine binding channel, which may represent some form of auto-inhibitory mechanism (Figure 7). Post-SET is fully structured only when bound to the substrate peptide (Figure 2A), or to a small molecule inhibitor [32], but the density is incomplete otherwise (Figure 2B–D). This implies that Post-SET is naturally flexible, which may be important for peptide turn-over, as recently proposed [10].

Altogether, these results suggest the presence of three conformational states for the Post-SET domain. (1) A flexible or even disordered state when no co-factor or peptide is bound. (2) A loose conformation when SAM or SAH, but no peptide is bound, which may also accommodate non-specific sequences. (3) A more rigid conformation in which the I-SET domain closes onto the substrate peptide. These states are probably in a dynamic equilibrium, with the co-factor and substrate shifting the equilibrium toward conformation 3.

Long Range Electrostatics Attract Histone Peptides to the HKMT Binding Groove

Histone tails are rich in lysines and arginines and consequently are electropositive (Figure 8A). Mapping the electrostatic potential along the molecular surface of GLP, G9a, SUV39H2 and PRDM2 shows that the peptide-binding groove is consistently electronegative (Figure 8B–E). This feature is also conserved in the structure of the N. crassa H3K9 methyltransferase Dim-5 (Figure 8F). This suggests that non-specific long-range electrostatic attractions play an evolutionarily conserved role in guiding the substrate-binding groove towards histone tails.

Based on the four HKMT structures presented here, we propose a mechanism for selective lysine H3K9 methylation, in which (1) long-range electrostatics attract the enzyme onto basic histone tails, (2) a pre-formed I-SET domain carries structural determinants necessary for specific interactions with the substrate peptide, and (3) catalytically competent conformation is achieved by subsequent closing of the Post-SET domain on the substrate. Considering the electronegative potential of the binding groove, our analysis suggests that HKMT inhibitors should be rather basic. To achieve selectivity, inhibitors should bind sites with clear interaction field potential occupied by residues distal to the substrate lysine. A recent co-crystal structure of the first specific HKMT inhibitor supports these general concepts [32].

Materials and Methods

Cloning, Expression and Purification of Human GLP

DNA fragments encoding the methyltransferase domain of human GLP (residues 951-1235), G9a (residues 913–1193), SUV39H2 (residues 112–410) and PRDM2 (residues 2–148) were amplified by PCR and sub-cloned into the pET28a-LIC vector (http://www.sgc.utoronto.ca/SGC-WebPages/toronto-vectors.php), downstream of the poly-histidine coding region. The methyltransferase domains of the above proteins were expressed in *E.coli* BL21 (DE3) codon plus RIL strain (Stratagene) by addition of 1 mM isopropyl-1-thio-D-galactopyranoside and incubated overnight at 15°C. The proteins were purified as following: harvested cells were resuspended in phosphate-buffered-saline (pH 7.4) supplemented with 250 mM NaCl, 2 mM β-mercaptoethanol, 5% glycerol, 0.1% Igepal and 1 mM phenylmethyl sulfonyl fluoride. The cells were lysed by passing through a microfluidizer (Microfluidics Corp.). The lysate was loaded onto a 5 ml HiTrap Chelating column (GE Health Care), charged with Ni^{2+}. The column was washed with 10 column volume of 20 mM Tris-HCl (pH 8.0), containing 250 mM NaCl and 50 mM imidazole, 5%
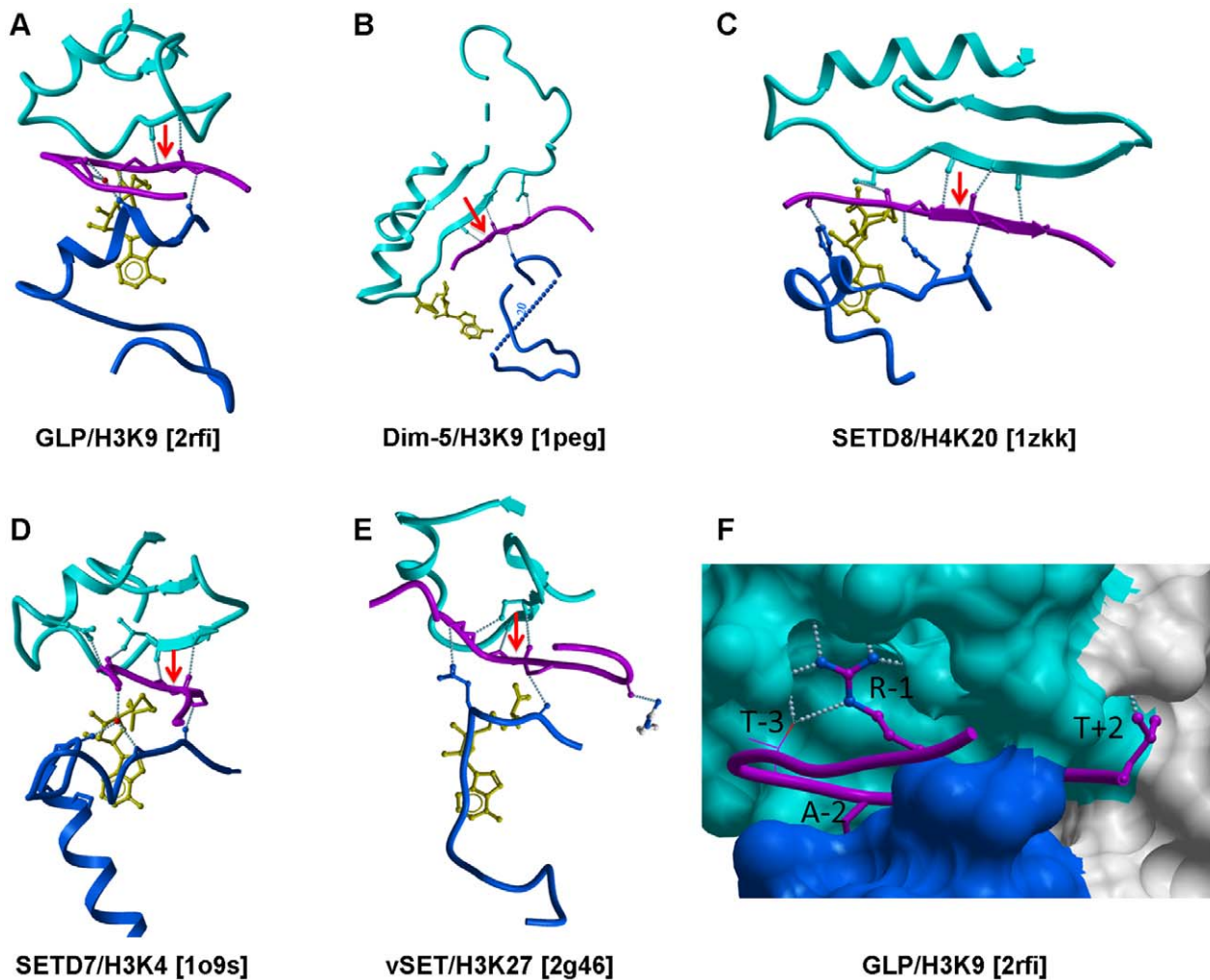
Figure 5. Backbone and side-chain contributions to peptide binding. A–E: Both Post-SET (blue) and I-SET (cyan) backbone atoms are engaged in a network of hydrogen-bonds with the peptide main-chain (magenta). A pair of hydrogen-bonds between backbone atoms of the I-SET and substrate lysine are conserved in all available HKMTs ternary complexes to date (dotted lines flanking red arrow). F: the substrate peptide sits in a groove formed by the I-SET (cyan) and the Post-SET (blue) domains. Peptide side-chains contributing most to the interaction are shown (magenta sticks). The guanidinium group of H3R8 (R-1) makes extensive contacts with the I-SET domain.
doi:10.1371/journal.pone.0008570.g005

glycerol, and the protein was eluted with elution buffer (20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 250 mM imidazole, 5% glycerol). The protein was loaded onto a Superdex 200 column (GE Health Care) equilibrated with 20 mM Tris-HCl (pH 8.0) and 150 mM NaCl. Thrombin (Sigma) was added to combined fractions containing the target proteins to remove the His_6- tag. The protein was further purified to homogeneity by ion-exchange chromatography.

Crystallization

Purified GLP, G9a and SUV39H2 proteins were crystallized in the presence of S-adenosyl-L-homocysteine (Sigma) or S-adenosyl-L-methionine (Sigma) using the hanging drop vapor diffusion method at 20°C by mixing equal volume of the protein solution with the reservoir solution. The GLP-SAH binary complex (protein/SAH molar ratio of 1:5) was crystallized in 20% PEG 4,000, 10% isopropanol, 0.1 M HEPES (pH 7.5). The G9a-SAH complex (protein/SAH molar ratio of 1:5) was crystallized in 20% PEG 3,350, 0.2 M NaF, 0.1 M Bis-Tris propane (pH 7.5), 5% ethylene glycol. The SUV39H2-SAM complex (protein/SAM molar
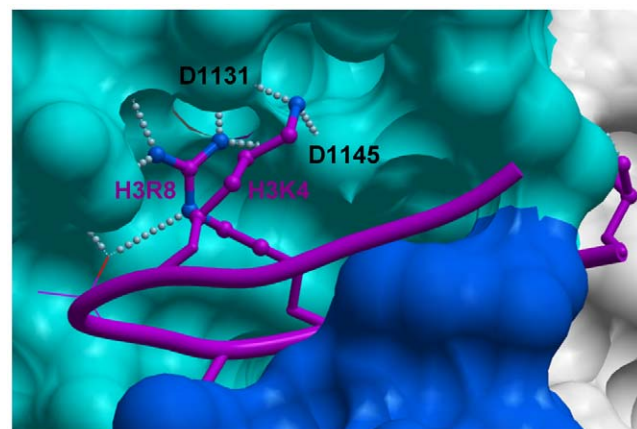
Figure 6. Contribution of H3K4 to H3K9 binding. Our structures of GLP in complex with H3K9me or H3K9me2 show that H3K4 folds on top of H3R8, making polar interactions with D1131 and D1145 of GLP.
doi:10.1371/journal.pone.0008570.g006
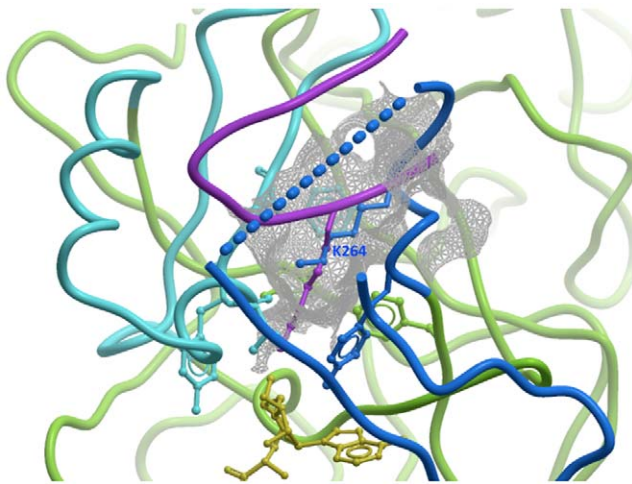
**Figure 7. Auto-inhibitory conformation of SUV39H2.** In our structure of SUV39H2, the C-terminus of the Post-Set domain (blue) adopts a conformation that positions its K264 side-chain (blue sticks) half-way into the substrate lysine channel (gray mesh). The H3K9me peptide (magenta) from a superimposed GLP-H3K9me structure is shown as a reference. SET and Post-SET of SUV39H2 are colored green and blue respectively.
doi:10.1371/journal.pone.0008570.g007

ratio of 1:10) was crystallized in 20% PEG 10, 000, 0.1 M HEPES (pH 7.5). The GLP-SAH-H3K9me and GLP-SAH-H3K9me2 complexes (protein/SAH/peptide molar ratio of 1:5:10) were crystallized in 16% PEG 4,000, 10% isopropanol, 0.1 M HEPES (pH 7.5). Purified PRDM2 was crystallized using hanging drop vapor diffusion method at 20°C by mixing 1.5 μl of the protein solution with 1.5 μl of the reservoir solution containing 22% PEG 5,000 MME, 0.2 M ammonium sulfate, 0.1 M MES (pH 7.0). All the crystals were soaked in the corresponding mother liquor supplemented with 20% glycerol as cryoprotectant before freezing in liquid nitrogen, except the G9a-SAH complex crystals, for which paraton-N was used as cryoprotectant.

Data Collection and Structure Determination

X-ray diffraction data were collected at 100 K at beamline 17ID of Advanced Photon Source (APS) at Argonne National Laboratory, beamline X25 of the National Synchrotron Light Source, beamline A1 of Cornell High Energy Synchrotron Source (CHESS), Cornell University, and a Rigaku FR-E home source. Data were processed using the HKL-2000 software suite [33]. The structures of methyltransferase domain of GLP, G9a, and SUV39H2 were solved by molecular replacement using the program MOLREP [34]. ARP/wARP [35] was used for automatic model building. Graphics program COOT [36] was used for model building and
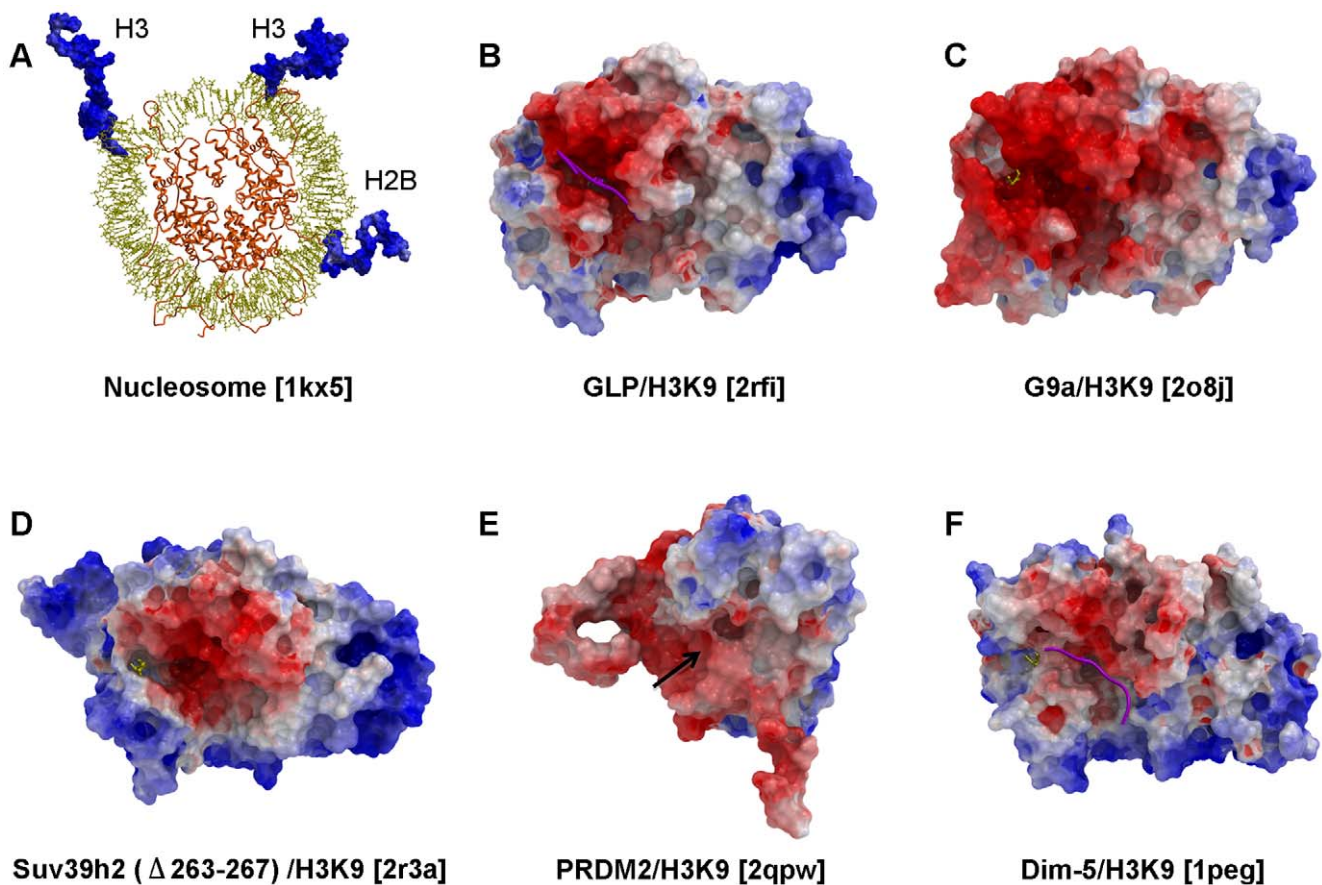
Figure 8. Electrostatic component to H3K9 peptide binding. While the overall electrostatic profile of available H3K9 methyltransferases structures varies, the peptide-binding groove is consistently electronegative (B–E: this work, F: *N. crassa* methyltransferase Dim-5), in contrast with the largely positive electrostatic potential of histone tails (A). When present, the substrate peptide is shown in magenta. Residues 264–267 of SUV39H2 were partially occupying the binding site and were removed. The Post-SET domain of PRDM2 is entirely disordered and the position of the substrate lysine binding channel is indicated with a black arrow.
doi:10.1371/journal.pone.0008570.g008

Table 1. Summary of X-ray diffraction data.

	EHMT1+AdoHcy	EHMT1+AdoHcy+H3K9me2	EHMT1+AdoHcy+H3K9me	EHMT2+AdoHcy	SUV39H2+AdoMet	PRDM2
PDB Code	2IGQ	2RFI	3HNA	2O8J	2R3A	2QPW
Data collection						
X-ray source	NSLS-X25	APS-17ID	APS-17ID	APS-17ID	Rigaku FR-E	CHESS-A1
Space group	$P2_1 2_1 2_1$	$P2_1 2_1 2_1$	$P2_1 2_1 2_1$	$P1$	$P1\,2_1\,1$	$P2_1 2_1 2_1$
Cell dimensions						
a, b, c (Å)	74.52, 95.86, 102.12	84.59, 85.63, 95.67	83.53, 83.37, 95.13	56.59, 70.96, 83.05	45.75, 63.92, 64.53	36.46, 51.20, 70.89
$\alpha,\ \beta,\ \gamma$ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	81.11, 75.13, 88.39	90.00, 109.33, 90.00	90.00, 90.00, 90.00
Resolution (Å) (highest resolution shell)	69.84-2.00 (2.07-2.00)	42.80-1.60 (1.64-1.60)	62.75-1.50(1.58-1.50)	79.30-1.80 (1.86-1.80)	50.00-2.00 (2.08-2.00)	50.00-1.79 (1.86-1.79)
Unique reflections	44160	89676	106827	100323	22956	23999
R_{sym} (%)	9.6	8.0	9.8	4.8	8.4	6.3
$I/\sigma I$	8.8(1.3)	10.8 (1.4)	15.2(4.3)	13.7(3.0)	15.0(3.4)	34.3 (6.0)
Completeness (%)	92.6(83.2)	96.5 (67.2)	100.0(100.0)	92.3(72.7)	96.6 (92.9)	99.7 (99.3)
Redundancy	3.7(2.0)	7.4 (3.0)	8.2(8.2)	3.2(3.2)	2.4 (2.3)	7.2 (6.4)
Refinement						
Resolution (Å)	42.11-2.00	42.80-1.59	62.75-1.50	79.31-1.80	35.38-2.00	41.49-1.79
No. reflections (test set)	44160 (2365)	85102 (4475)	101405 (5324)	100323 (5287)	21763 (1182)	13205 (841)
R_{work}/R_{free} (%)	19.1/23.7	19.4/22.0	16.7/19.3	18.4/22.1	17.9/21.3	17.1/22.2
No. atoms						
Protein	4084	4322	4550	8331	2154	1192
Cofactor	60	60	60	120	31	N/A
Water	566	681	888	1780	254	145
B-factors (Å^2)						
Protein	31.7	24.8	15.1	24.2	29.2	21.8
Cofactor	44.5	33.2	13.1	24.3	25.7	N/A
Water	40.4	32.6	31.9	32.8	38.3	36.2
R.m.s deviations						
Bond lengths (Å)	0.015	0.009	0.009	0.013	0.017	0.019
Bond angles (°)	1.53	1.17	1.28	1.35	1.55	1.51
Ramachandran plot % residues						
Favored	82.9	85.0	85.1	86.4	86.7	87.8
Additionally allowed	16.2	14.1	14.3	13.2	13.3	12.2
Generously allowed	0.5	0.4	0.2	0	0	0
Disallowed	0.5	0.4	0.4	0.4	0	0

visualization. PRDM2 structure was solved by single-wavelength anomalous diffraction (SAD) at low resolution, using a seleno-methionine derivative crystal with the program SHELXD [37], and the phasing was performed using SHELXE [38]. The low resolution structure was used as model to solve the native structure at higher resolution. Crystal diffraction data and refinement statistics for the structure are displayed in Table 1. One residue is in a disallowed area of the Ramachandran plot in our GLP (M1049) and G9a (I992) structures. This strained residue maps at a conserved location, remote from the peptide and cofactor binding sites.

Histone Methyltransferase Assay

The SAHH-coupled assay described by Collazo et al. [39] was optimized and employed to assay the activity of G9a. This assay utilizes S-adenosylhomocysteine hydrolase (SAHH) to hydrolyze the methyltransfer product S-adenosylhomocysteine to homocys-teine and adenosine in the presence of adenosine deaminase which converts adenosine to inosine. The homocysteine concentration is then determined through conjugation of its free sulfhydryl moiety to a thiol-sensitive fluorophore, ThioGlo (Calbiochem). Assays were performed at room temperature in 25 mM potassium phosphate buffer pH 8, 1 mM EDTA, 2 mM $MgCl_2$ and 0.01% Tween 20. Series of control experiments were conducted to establish the optimum assay condition for each methyltransferase and the optimum conditions were used to determine the kinetic parameters for GLP. Assay cocktails were prepared with 5 μM SAHH to avoid any SAH accumulation while produced from the methyltransferase reaction, 3 U/ml of adenosine deaminase from Sigma, 70 μM SAM, and GLP. The peptide concentrations were varied over the range of 2 μM to 4 mM. Assays were initiated by the addition of peptide and immediately after starting the reaction, 2x volume of 20 μM ThioGlo solution was added to each well. The methylation reaction was followed by monitoring the increase in fluorescence using Biotek Synergy2 plate reader with 360/40 nm excitation filter and 528/20 nm emission filter for 20 min in 384 well-plate format. Homocysteine generated in the assay was quantified using standard curves. Activity values were corrected by subtracting background caused by the peptide or the protein. K_m and k_{cat} values were calculated using the Michaelis-Menten equation and Sigmaplot 9.0. Standard deviations were calculated from two independent experiments.

Structure Superimposition and Electrostatic Potential Coloring

Optimal structure superimpositions were identified with ICM (Molsoft LLC). Briefly, the algorithm uses an iterative procedure to find the best "alignable" main chain core in both structures based on seed alignments of 15 residues as follows: (1) start with the most reliable seed alignment of 15 residues; (2) set all weights to 1; (3) perform weighted superposition and evaluate RMSD; (4) calculate the deviation Di for each backbone atom pair; (5) sort the deviations and find the deviation D50 corresponding to 50-percentile of the deviation array; (6) calculate weights W according to the formula $Wi = exp(-D50^2/Di^2)$; (7) go back to step 3 unless a limit of 10 iterations is reached. The electrostatic potential was calculated with ICM using a boundary element solution of the Poisson equation. Color saturation was set to calculated values of $+/- 5$ kcal/electron units ($+5 =$ blue $-5 =$ red) when the electrostatic potential was projected on molecular surfaces.

Supporting Information

Figure S1 Sequence alignment of the methyltransferase domain of H3K9 HKMTs. Residues within 4Å of bound H3K9 peptide in our GLP-H3K9me complex are highlighted in red. The three aspartate making polar interactions with arginine H3R8 are colored blue. The sequence of PRDM2 is too divergent from other H3K9 HKMTs and was not included in the alignment. Large inserts present in SETDB1 and SETDB2 sequences are not shown for clarity.
Found at: doi:10.1371/journal.pone.0008570.s001 (1.11 MB TIF)

Datapack S1 Standalone iSee datapack - contains the enhanced version of this article for use offline. This file can be opened using free software available for download at http://www.molsoft.com/icm_browser.html.
Found at: doi:10.1371/journal.pone.0008570.s002 (ICB)

Text S1 Instructions for installation and use of the required web plugin (to access the online enhanced version of this article).
Found at: doi:10.1371/journal.pone.0008570.s003 (0.75 MB PDF)

Acknowledgments

We thank Peter Loppnau, He Ren, and Yong Zhao for their excellent technical support, and Stephen Frye for comments on the manuscript.

Author Contributions

Conceived and designed the experiments: HW MV AP. Performed the experiments: HW JM VVL TA LD HZ AAH VCS. Analyzed the data: HW CHA AP MS. Wrote the paper: MS.

References

1. Kouzarides T (2007) Chromatin modifications and their function. Cell 128: 693–705.
2. Grewal SI, Moazed D (2003) Heterochromatin and epigenetic control of gene expression. Science 301: 798–802.
3. Martin C, Zhang Y (2005) The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol 6: 838–849.
4. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, et al. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410: 120–124.
5. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature 410: 116–120.
6. Volkel P, Angrand PO (2007) The control of histone lysine methylation in epigenetic regulation. Biochimie 89: 1–20.
7. Dillon SC, Zhang X, Trievel RC, Cheng X (2005) The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol 6: 227.
8. Qian C, Zhou MM (2006) SET domain protein lysine methyltransferases: Structure, specificity and catalysis. Cell Mol Life Sci 63: 2755–2763.
9. Xiao B, Jing C, Kelly G, Walker PA, Muskett FW, et al. (2005) Specificity and mechanism of the histone methyltransferase Pr-Set7. Genes Dev 19: 1444–1454.
10. Southall SM, Wong PS, Odho Z, Roe SM, Wilson JR (2009) Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. Mol Cell 33: 181–191.
11. Couture JF, Collazo E, Brunzelle JS, Trievel RC (2005) Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. Genes Dev 19: 1455–1465.
12. Xiao B, Jing C, Wilson JR, Walker PA, Vasisht N, et al. (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9. Nature 421: 652–656.
13. Zhang X, Tamaru H, Khan SI, Horton JR, Keefe LJ, et al. (2002) Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltrans-ferase. Cell 111: 117–127.
14. Qian C, Wang X, Manzur K, Sachchidanand, Farooq A, et al. (2006) Structural insights of the specificity and catalysis of a viral histone H3 lysine 27 methyltransferase. J Mol Biol 359: 86–96.
15. Marmorstein R (2003) Structure of SET domain proteins: a new twist on histone methylation. Trends Biochem Sci 28: 59–62.
16. Gonzalez ME, Li X, Toy K, DuPrie M, Ventura AC, et al. (2009) Downregulation of EZH2 decreases growth of estrogen receptor-negative invasive breast carcinoma and requires BRCA1. Oncogene 28: 843–853.

17. Krivtsov AV, Feng Z, Lemieux ME, Faber J, Vempati S, et al. (2008) H3K79 methylation profiles define murine and human MLL-AF4 leukemias. Cancer Cell 14: 355–368.
18. Schneider R, Bannister AJ, Kouzarides T (2002) Unsafe SETs: histone lysine methyltransferases and cancer. Trends Biochem Sci 27: 396–402.
19. Li Y, Reddy MA, Miao F, Shanmugam N, Yee JK, et al. (2008) Role of the histone H3 lysine 4 methyltransferase, SET7/9, in the regulation of NF-kappaB-dependent inflammatory genes. Relevance to diabetes and inflammation. J Biol Chem 283: 26771–26781.
20. Shi Y, Desponts C, Do JT, Hahm HS, Scholer HR, et al. (2008) Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. Cell Stem Cell 3: 568–574.
21. Copeland RA, Solomon ME, Richon VM (2009) Protein methyltransferases as a target class for drug discovery. Nat Rev Drug Discov 8: 724–732.
22. Spannhoff A, Hauser AT, Heinke R, Sippl W, Jung M (2009) The Emerging Therapeutic Potential of Histone Methyltransferase and Demethylase Inhibitors. ChemMedChem 4: 1568–1582.
23. Cheng X, Collins RE, Zhang X (2005) Structural and sequence motifs of protein (histone) methylation enzymes. Annu Rev Biophys Biomol Struct 34: 267–294.
24. Xiao B, Wilson JR, Gamblin SJ (2003) SET domains and histone methylation. Curr Opin Struct Biol 13: 699–705.
25. Peters AH, Kubicek S, Mechtler K, O'Sullivan RJ, Derijck AA, et al. (2003) Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell 12: 1577–1589.
26. Couture JF, Dirk LM, Brunzelle JS, Houtz RL, Trievel RC (2008) Structural origins for the product specificity of SET domain protein methyltransferases. Proc Natl Acad Sci U S A 105: 20659–20664.
27. Zhang X, Yang Z, Khan SI, Horton JR, Tamaru H, et al. (2003) Structural basis for the product specificity of histone lysine methyltransferases. Mol Cell 12: 177–185.
28. Smith BC, Denu JM (2009) Chemical mechanisms of histone lysine and arginine modifications. Biochim Biophys Acta 1789: 45–57.
29. Collins RE, Tachibana M, Tamaru H, Smith KM, Jia D, et al. (2005) In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. J Biol Chem 280: 5563–5570.
30. Rathert P, Dhayalan A, Murakami M, Zhang X, Tamas R, et al. (2008) Protein lysine methyltransferase G9a acts on non-histone targets. Nat Chem Biol 4: 344–346.
31. Rathert P, Zhang X, Freund C, Cheng X, Jeltsch A (2008) Analysis of the substrate specificity of the Dim-5 histone lysine methyltransferase using peptide arrays. Chem Biol 15: 5–11.
32. Chang Y, Zhang X, Horton JR, Upadhyay AK, Spannhoff A, et al. (2009) Structural basis for G9a-like protein lysine methyltransferase inhibition by BIX-01294. Nat Struct Mol Biol 16: 312–317.
33. Otwinowski Z, Borek D, Majewski W, Minor W (2003) Multiparametric scaling of diffraction intensities. Acta Crystallogr A 59: 228–234.
34. Vagin A, Teplyakov A (2000) An approach to multi-copy search in molecular replacement. Acta Crystallogr D Biol Crystallogr 56: 1622–1624.
35. Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS (2001) ARP/wARP and molecular replacement. Acta Crystallogr D Biol Crystallogr 57: 1445–1450.
36. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126–2132.
37. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 58: 1772–1779.
38. Sheldrick GM (2008) A short history of SHELX. Acta Crystallogr A 64: 112–122.
39. Collazo E, Couture JF, Bulfer S, Trievel RC (2005) A coupled fluorescent assay for histone methyltransferases. Anal Biochem 342: 86–92.