Structural Chemistry of the Histone Methyltransferases Cofactor Binding Site

Valérie Campagna-Slater[†,$], Man Wai Mok[†], Kong T. Nguyen[†], Miklos Feher[‡], Rafael Najmanovich[‖], Matthieu Schapira[†,⊥,]*

[†] Structural Genomics Consortium, University of Toronto, MaRS Centre, South Tower, 7th floor, 101 College Street, Toronto, Ontario, Canada, M5G 1L7
[‡] Campbell Family Institute for Breast Cancer Research, University Health Network, Toronto, ON, Canada, M5G 1L7
[‖] Department of Biochemistry, University of Sherbrooke, 3001, 12e Avenue Nord, Sherbrooke, Québec, Canada, J1H 5N4
[⊥] Department of Pharmacology and Toxicology, University of Toronto, Medical Sciences Building, 1 King's College Circle, Toronto, Ontario, Canada, M5S 1A8
* matthieu.schapira@utoronto.ca

TITLE RUNNING HEAD: Structure of HMT cofactor site

[$] Present address: Department of Chemistry, McGill University, Otto Maass Chemistry Building, 801 Sherbrooke Street West, Montréal, Québec, Canada, H3A 2K6

ABSTRACT

Histone methyltransferases (HMTs) transfer a methyl group from the cofactor S-adenosyl methionine to lysine or arginine residues on histone tails, thereby regulating chromatin compaction, binding of effector proteins and gene transcription. HMTs constitute an emerging target class in diverse disease areas, and selective chemical probes are necessary for target validation. Potent and selective competitors of the substrate peptide have been reported, but the chemical tractability of the cofactor binding site is poorly understood. Here, a systematic analysis of this site across structures of fourteen human HMTs or close homologues was conducted. The druggability, interaction hotspots and diversity of the cofactor binding pocket were dissected. This analysis strongly suggests that this site is chemically tractable. General principles underlying tight binding and specific guidelines to achieve selective inhibition are presented.

KEYWORDS

Histone methyltransferases, S-adenosyl-methionine, cofactor, structure, druggability, chemical tractability

INTRODUCTION

The systematic "chemogenomic" coverage of a protein family with potent and selective chemical probes is a paradigm that is gaining popularity in drug discovery.[1-5] Chemical probes are potent compounds specific for a single target or for a very limited set of closely homologous targets, and constitute a first step towards the rational design of drugs with a pre-defined polypharmacology. They can help dissect the biology of gene families, prioritize targets, and inspire medicinal chemistry. When co-crystallized to their target, these molecular tools also highlight preferred contact points lining the receptor pocket that can be exploited to derive potency or selectivity. For instance, a large number of diverse compounds inhibit the human kinome, and in-depth structural knowledge of the inhibitor binding site enabled the rational design of molecules with improved selectivity profiles.[6,7] Systematically analyzing the structural chemistry of a gene family is therefore an important component of a knowledge-base approach to drug discovery.

The rapidly evolving field of epigenetics is expected to uncover novel therapeutic opportunities.[8] For instance, histone deacetylase inhibitors were recently approved in oncology.[9] Similar to histone deacetylases, histone methyltransferases (HMTs) generally antagonize the action of acetyltransferases, and constitute another emerging target class.[10,11] Through the deposition of methyl marks on histone tails, these enzymes control the compaction state of chromatin and play an important role in the epigenetic regulation of gene transcription.[12,13] They are divided into two groups: histone lysine methyltransferases (HKMTs)[14,15] and protein arginine methyltransferases (PRMTs).[16] The structural mechanism is well understood, and relies on the recruitment of a substrate peptide and a methyl donating cofactor – S-adenosyl methionine (SAM) – at two distinct, but adjacent binding sites. In this respect, an analogy can be made with kinases, where a phosphate group is transferred from a small molecule cofactor to a substrate peptide. A major difference is that numerous compounds with varied positions in the drug discovery pipeline are known that bind at the cofactor binding site of kinases, while very few are known for HMTs.[10,11] To date, the only small-molecule inhibitors co-crystallized with HMTs are BIX-01294 (PDB: 3FPD), an EHMT1/GLP and EHMT2/G9a dual inhibitor, and its derivatives UNC0224 (PDB: 3K5K), and UNC0638 (PDB: 3NNI), neither of which bind at the cofactor site,[17-19] as well as Sinefungin (e.g. SETD7 – PDB: 3CBP), a close analogue of the cofactor which likely

lacks specificity. Another compound acting as a SAM competitive inhibitor is the fungal mycotoxin chaetocin, which shows selectivity for HKMTs targeting lysine 9 of histone 3 (H3K9).[20]

Clearly, HMTs constitute a promising target class, and targeting the cofactor site seems a relevant strategy. However, the chemical tractability of this site has not been fully validated, and its structural diversity has been discussed but not thoroughly analyzed.[10] In this work, the chemical properties of the cofactor site are systematically analyzed across structures representing 14 human HMTs (10 SET domain, and 4 non-SET domain enzymes). The relative druggability of each site is evaluated, conserved interactions with the natural substrate are compared with predicted interaction hotspots (i.e. regions expected to be major contributors to binding free energy or selectivity[21]), and points of structural diversity are outlined. This study defines structural features that could be exploited to design potent and selective inhibitors.

METHODS

Multiple sequence alignment

Human HMTs containing a SET domain (defined as a130 amino-acid sequence carrying the signature motifs ELxF/YDY and NHS/CxxPN[22]) were extracted from the SMART database.[23] 62 SET domain containing proteins were identified. After manually removing redundancies (e.g. ESPN00000373354 is a variant of Q53H47), fragments, and predicted genes, replacing obsolete entries (e.g. PRDM2's ID is Q13029, not Q5VUM0), and adding missing entries from the literature (e.g. SETD3) the final list contained 51 SET domain HMTs. The non-SET domain human HMTs DOT1L (the only known non-SET domain human lysine HMT), and PRMT1 to PRMT9 (9 human arginine HMTs known to possess methyltransferase activity - PRMT4 is later referred to as CARM1) were added to the list.[24,25] A multiple sequence alignment was generated with ClustalW, using the following sequences: the Superfamily domain SSF82199 of HKMTs when available (SSF82199 includes both SET and post-SET), SET domain of PRDMs (for which SSF82199 is not defined), full length arginine methyltransferases (which have no SET domain), and the sequence of the crystallized catalytic domain of DOT1L (PDB: 1NW3). The program PhyloDraw was used to derive a phylogenetic tree from the multiple sequence alignment (Figure 1).

Druggability

The percentage of the cofactor's molecular surface accessible to solvent was used as an indicator of druggability. It should be noted that this first approximation ignores other factors, such as hydrophobicity and flexibility. However, within a single protein family that binds the same cofactor, a binding site that makes more contacts with a ligand may offer more possibilities to design specific interactions relative to other members of the family. The surface area of the cofactor was calculated with ICM (Molsoft LLC, San Diego, CA). The complex structure was first converted to icm format (which includes addition of hydrogen atoms), the surface energy term was activated (with the command: set term "sf"), the surface of the cofactor was calculated (with the command: show surface area a_x//* a_x//*, where x is the name of the cofactor molecule), water molecules were deleted, and the solvent accessible surface of the cofactor was calculated (with the command: show surface area a_x//*). The program SiteMap (Schrodinger, New York, NY) was used as a complementary method to evaluate druggability.[26] Protein structures were protonated with the protein preparation wizard of Schrodinger's Maestro interface, and the binding sites were defined by selecting the bound cofactor in SiteMap. The druggability score (DScore) was automatically computed by SiteMap as $0.094\ n^{1/2} + 0.6\ e - 0.324\ p$, where n, e and p are measures of the volume, enclosure, and hydrophilicity of the pocket.[26]

Mapping interaction hotspots

Unlike druggability, which is a variable that defines an entire pocket and indicates whether small molecules can bind with high affinity to this pocket, interaction potentials are indicating the specific atomic positions along the pocket surface where strong interactions can take place. A pocket may be void of strong interaction potentials, but still be sufficiently large, enclosed, and hydrophobic to be druggable. Alternatively, an undruggable pocket may include a few strong interaction potentials scattered along a rather shallow surface. Interaction potential maps (GRID maps) were applied to gauge

similarities and differences in the distribution of interaction hot-spots between different HMT structures. These potentials attempt to estimate where different chemical probes have favourable interactions with a molecular surface. The GRID maps applied here are based on the work of Goodford et al.[27] and Boobyer et al.[28] and were implemented in the Molecular Operating Environment (MOE) modeling suite (Chemical Computing Group Inc., Montreal, Canada). A variety of probes are available, spanning a range of different combinations of size, charge, and hydrogen bond donor/acceptor properties. In this work, the results with only two probes are described for simplicity: a water probe to find hydrogen bond acceptor / donor hot-spots and a dry probe to identify favourable hydrophobic interactions. In order to compare these maps among all studied proteins, they were displayed at a constant level (-7.5 kcal/mol for the water map and -2.5 kcal/mol for the DRY map), however, in order to ensure that no interactions are 'lost' as a result of slight differences in the electrostatics of these proteins, these levels were varied slightly (up to 10%) to identify further potential cusps in the surface. The obtained potential surfaces were then displayed in a schematic manner, manually indicating whether such surfaces are at or near particular atoms of the 2D structure. The spots shown in these figures are not indicative of the magnitude and shape of the 3D surfaces, merely their locations are indicated to pinpoint differences among the proteins.

Structure-based clustering of the cofactor binding sites

Amino acid selection was as follows: residues within 6 Å of the bound cofactor or bound analogue, and within 2 Å of the molecular surface of the protein were selected in all HMT structures. All SET-domain structures were superimposed, and for each amino-acid selected, residues located at the same position in other structures were identified and added to the selection if they had originally been missed. The procedure was exactly identical for the non-SET domain HMTs, and for the kinase set. The corresponding selections are shown in Supporting Information Table 2 and Table 3.

The structure-based clustering of cofactor binding sites was performed using the subset of solvent accessible atoms belonging to the amino acid selections above. Solvent accessible areas were calculated analytically using a constrained Voronoi procedure.[29] All-against-all cofactor binding site 3D atomic correspondences (pairwise subsets of atoms of equivalent atom types in equivalent relative geometric positions) were detected using the graph-matching-based IsoCleft method with default parameters using the exact Bron-Kerbrosh procedure.[30] A number of measures can be used to quantify similarity; while none is perfect, in our experience the Local Tanimoto Score of 3D similarity (LTS_{3D}) provides the best measure of chemical similarity.[30] The Local Tanimoto Score of 3D similarity is defined as follows:

$$LTS_{3D} = \frac{C}{(N_A + N_B - C)}$$

Where C represents the number of atoms in common between the two clefts and N_A, N_B represent the total number of atoms in the two clefts under comparison. This measure of similarity varies from 0 (when there are no atoms in common) to 1 (when all atoms are common). It is important to notice that this measure of similarity is very sensitive to the size of the clefts under comparison. The resulting similarity matrices are shown in the first tab in Supporting Information Table 4 and Table 5 respectively for the HMT and Kinase datasets. The second tab in Supporting Information Table 4 and Table 5 shows the number of atoms in common (C) detected by IsoCleft as well as the total number of atoms used in the comparisons (N_i, diagonal elements) with which LTS_{3D} can be calculated as described above. The average group agglomerative hierarchical clustering method was used to cluster the two datasets as this method produced the largest correlation between the cophenetic matrix (a matrix with the minimal cluster tree heights) and the original IsoCleft distance data. The larger the cophenetic correlation, the better the clustering tree represents the structure of the data in the similarity matrix used to perform the clustering. We obtained cophenetic correlation coefficients of 0.730 and 0.950 respectively for the HMT and Kinase datasets. Dendrograms based on the clustering of LTS_{3D} dissimilarity matrices ($1 - LTS_{3D}$) were produced with the open source software package R with the following command:

SIMT<-read.table("similarity_matrix.dat")

plot(as.dendrogram(hclust(as.dist(1-SIMT),method="average")),main="data1", xlab="", horiz=FALSE, ylim=c(0,1.))

RESULTS

Phylogenetic Classification

Human HMTs were extracted from the SMART database[23] and the published literature, and aligned with ClustalW (see Methods section for details). The program PhyloDraw was used to derive a phylogenetic tree from the multiple sequence alignment (Figure 1). A representative structure was selected for the catalytic domain of each gene, when available. More complete, and higher resolution structures were prioritized when multiple instances were available. These included three PRMTs (CARM1 [PDB: 3B3F, from rat – 100% sequence identity with human gene], PRMT1 [PDB: 1OR8, from rat - 100% sequence identity with human gene], PRMT3 [PDB: 2FYT]), ten SET-domain HKMTs (EHMT1/GLP [PDB: 2RFI], EHMT2/G9a [PDB: 3K5K], SUV39H2 [PDB: 2R3A], SETMAR [PDB: 3BO5], SETD8 [PDB: 1ZKK], SETD2 [PDB: 3H6L], SETD7 [PDB: 1O9S], MLL [PDB: 2W5Z], SMYD1 [PDB: 3N71, from mouse – 94% sequence identity with human gene], SMYD3 [PDB: 3MEK]), and the non-SET domain HKMT DOT1L [PDB:1NW3] (Figure 1). A number of human PRDM structures have been solved, but none are in complex with SAM or an analogue, and the cofactor binding site appears incomplete in all cases. These structures were excluded from the analysis.

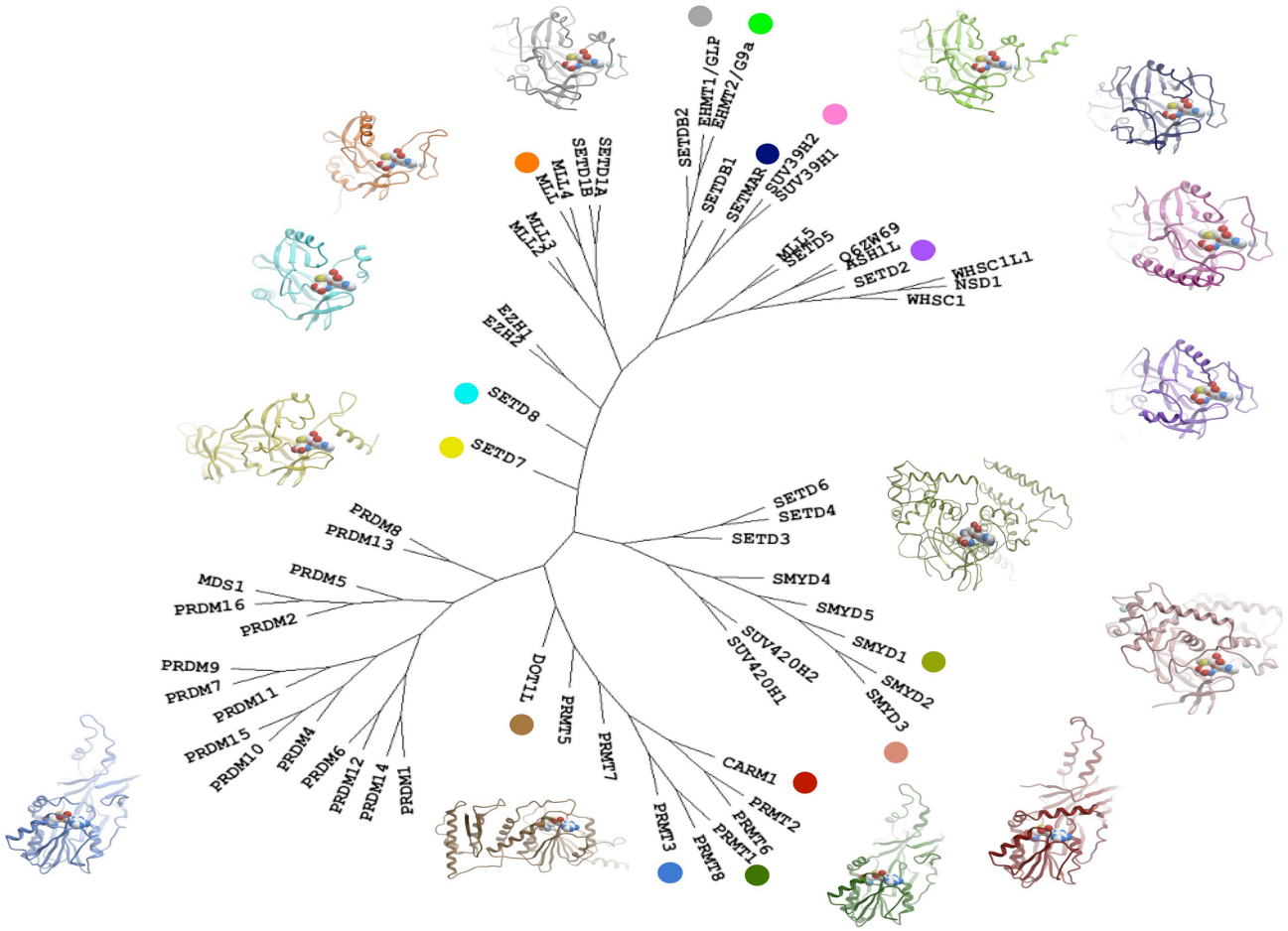

Figure 1: Structural coverage of human HMTs. Representative experimental structures used in this study are mapped on a phylogenetic tree of the protein family (drawn using phyloDraw): red – CARM1 (PDB: 3B3F, from rat); brown – DOT1L (PDB: 1NW3); bright green – EHMT2/G9a (PDB: 3K5K); grey – EHMT1/GLP (PDB: 2RFI); orange – MLL (PDB: 2W5Z); dark green – PRMT1 (PDB: 1OR8, from rat); royal blue – PRMT3 (PDB: 2FYT); purple – SETD2 (PDB: 3H6L); yellow – SETD7 (PDB:

1O9S) cyan – SETD8 (PDB: 1ZKK), dark blue – SETMAR (PDB: 3BO5); salmon – SMYD3 (PDB: 3MEK); olive green – SMYD1 (PDB: 3N71, from mouse); pink – SUV39H2 (PDB: 2R3A).

Relative Druggability of the Cofactor Site

Once the phylogenetic landscape and structural coverage of the protein family established, we evaluated the druggability of the cofactor binding site across the representative structures. The fact that SAM binds to HMTs does not mean that high affinity ligands can be found for the corresponding sites. For instance, the affinity of SAM for G9a is around 5 μM (unpublished data); whether more potent compounds can be identified is unclear. Desolvation is generally the major force driving protein-ligand interactions,[31, 32] and it is commonly accepted that the druggability of a binding pocket increases with its ability to shield small molecules from the solvent, and decreases with its hydrophilicity (other factors, such as flexibility, can also play a role).[26, 33, 34] Since all representative structures are in complex with SAM, the reaction product S-adenosyl homocysteine (SAH) or the very close analogue Sinefungin in the case of SMYD1, the solvated fraction of the bound molecule was used as a first approximation of druggability (Figure 2). This revealed a significantly higher enclosure for non-SET domain proteins (PRMTs and DOT1L), where only a small portion of the nucleotide ring of the cofactor remains accessible to solvent (the cofactor appears less buried in the PRMT1 complex because the N-terminal extremity of the protein, that is expected to close on the ligand, is missing from the structure). The SETD2 pocket appeared also among the most enclosed, with less than 3% of the cofactor surface accessible to solvent. This is due to the shielding of the ribose ring by His1603, which is part of an insertion specific to this enzyme. Next were the two SMYD proteins as well as SETMAR and SUV39H2, where the cofactor remains largely enclosed in the binding site, but the pocket opens-up along both the adenosyl, and methionine moiety, resulting in around 7.5% of cofactor surface accessible to solvent. The G9a and GLP pockets are slightly more open, leaving 11% to 12% of the cofactor solvated (over 50% more than in SMYD complexes). The fact that chaetocin, a SAM-competitive natural product, inhibits human SUV39H and mouse G9a with IC50s of 0.8 and 2.5 μM respectively suggests that these levels of desolvation are compatible with at least low micromolar affinity for substrate or inhibitors.[20] Finally, the cofactor lies in a similar orientation in the SETD7 and SETD8 complexes, but the pockets are significantly more open. The distribution of HMTs based on the relative enclosure of their SAM pocket does not match the evolutionary classification of the HMT domain shown Figure 1. Indeed, two domains with overall homologous sequences can display large variations locally. This observation was previously made for kinases, where phylogenetic classification based on residues lining the ATP site is different from a phylogeny derived from the sequence alignment of the entire catalytic domain.[35]

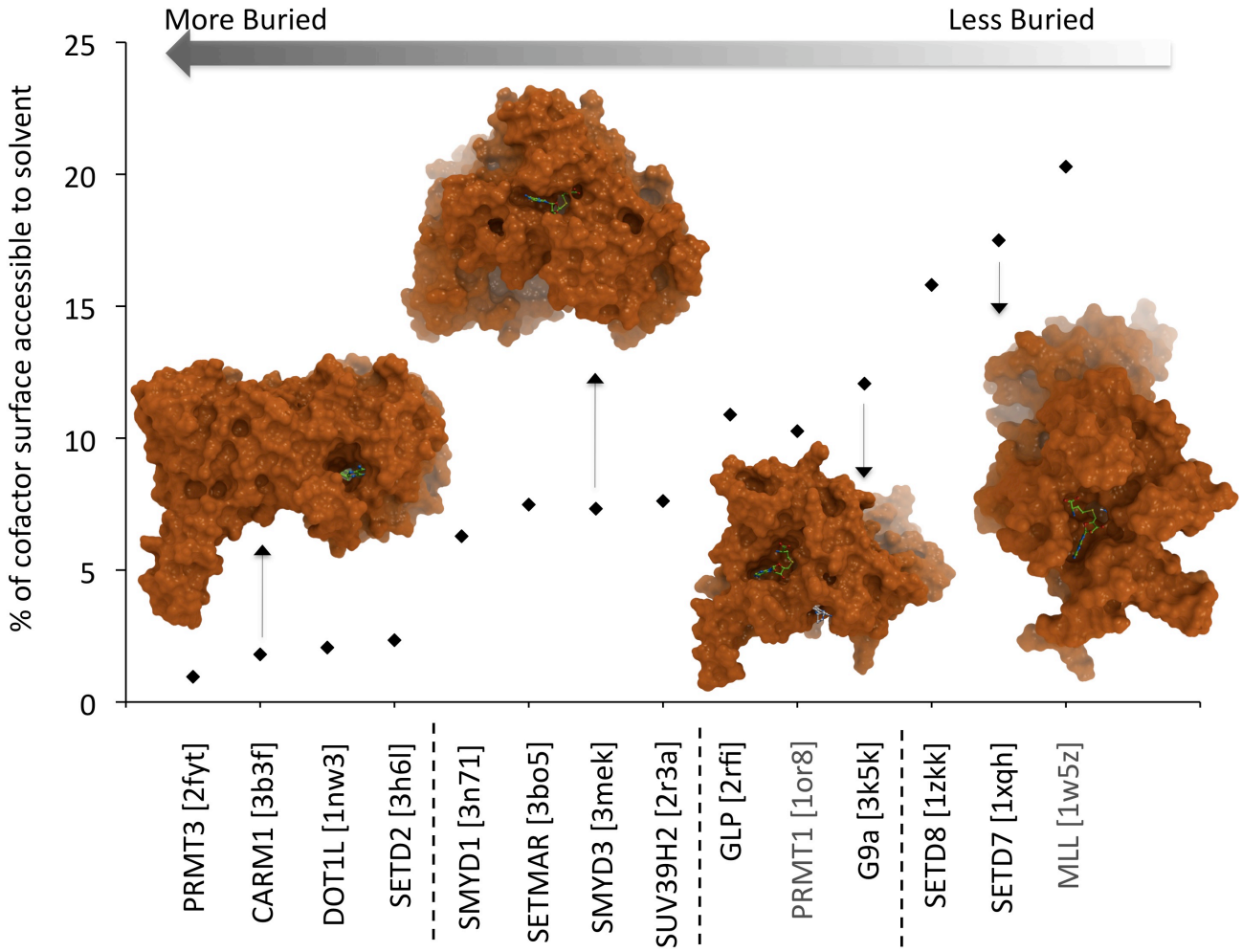

Figure 2: Enclosure of HMT cofactor sites. The surface of the cofactor accessible to the solvent reflects to which extent it is buried when bound to an HMT, and is used as an indicator of the cofactor site's druggability. The structures can be divided into four groups from most to less buried cofactor. (1) non-SET domain HMTs; (2) SMYDs, SUV39H2, SETMAR; (3) G9a, GLP; and (4) SETD7 and SETD8. (The PRMT1 pocket is expected to be more enclosed - i.e. druggable - than shown here as C-terminal residues predicted to close the pocket as seen in CARM1 and PRMT3 are missing from the structure. The MLL structure captures a state of the enzyme isolated from its obligate interaction partners and may not reflect accurately the enclosure of its cofactor site).

To complement this analysis, we used SiteMap to calculate the druggability index of the cofactor sites.[26] This program not only accounts for the geometry of pockets, but also for their hydrophobicity. A large validation study indicated that the average druggability scores of undruggable, difficult, and druggable pockets were 0.827, 0.995, and 1.091 respectively.[26] The calculated druggability scores were between 0.96 and 1.13 for all SAM binding sites, which suggests that they are all druggable (Supporting Information Table 1). The SETMAR pocket was an exception, with a score of 0.92, which could be attributed to its highly hydrophilic nature. In some cases the predicted druggability varied significantly from one protein chain to another within the same structure, based on small conformational variations, but always remained in the druggable range.

Together, these results show that the HMT cofactor site is druggable and that its druggability may vary across human HMTs, with PRMTs and DOT1L having the most favourable geometries.

Conformation of the cofactor.

Superimposing all available structures representing the 10 SET domain human HMTs reveals a high conservation of the cofactor conformation with a maximum root-mean-squared deviations (RMSDs) of 0.49 Å (Figure 3A). Similarly, the bound conformation of the cofactor is highly conserved among the PRMTs and the DOT1L structures with a maximum RMSD of 0.5 Å (Figure 3B). On the other hand, the conformation of the cofactor differs significantly between human SET and non-SET HMTs, with a minimum RMSD of 1.53 Å (Figure 3C).

To compare the conformation of SAH/SAM in the different HMT structures available as well as in other SAM binding proteins, all structures co-crystallized with SAH and SAM were extracted from the Protein Data Bank (PDB). The RMSDs of the cofactors were calculated against SAH co-crystallized to EHMT1/GLP (PDB: 2RFI) and to CARM1 (PDB: 3B3F), as representative structures for the SET- and non-SET domain enzymes, respectively (Figure 3D). Interestingly cofactor molecules bound to SET-domain HMTs (with the exception of the viral protein vSET) define a distinct conformational cluster. On the other hand, the cofactor conformation space defined by non-SET domain PRMTs and DOT1L overlaps with that of non-histone methyltransferases. This should not come as a surprise as the SAM-MT fold observed in both the PRMTs and DOT1L/Dot1p but not in SET-domain HMTs is common to many methyltransferases.[36] These results suggest that, from a ligand-based standpoint, ligands binding at the cofactor site of SET-domain HMTs are less likely to be promiscuous toward other SAM-binding target classes.

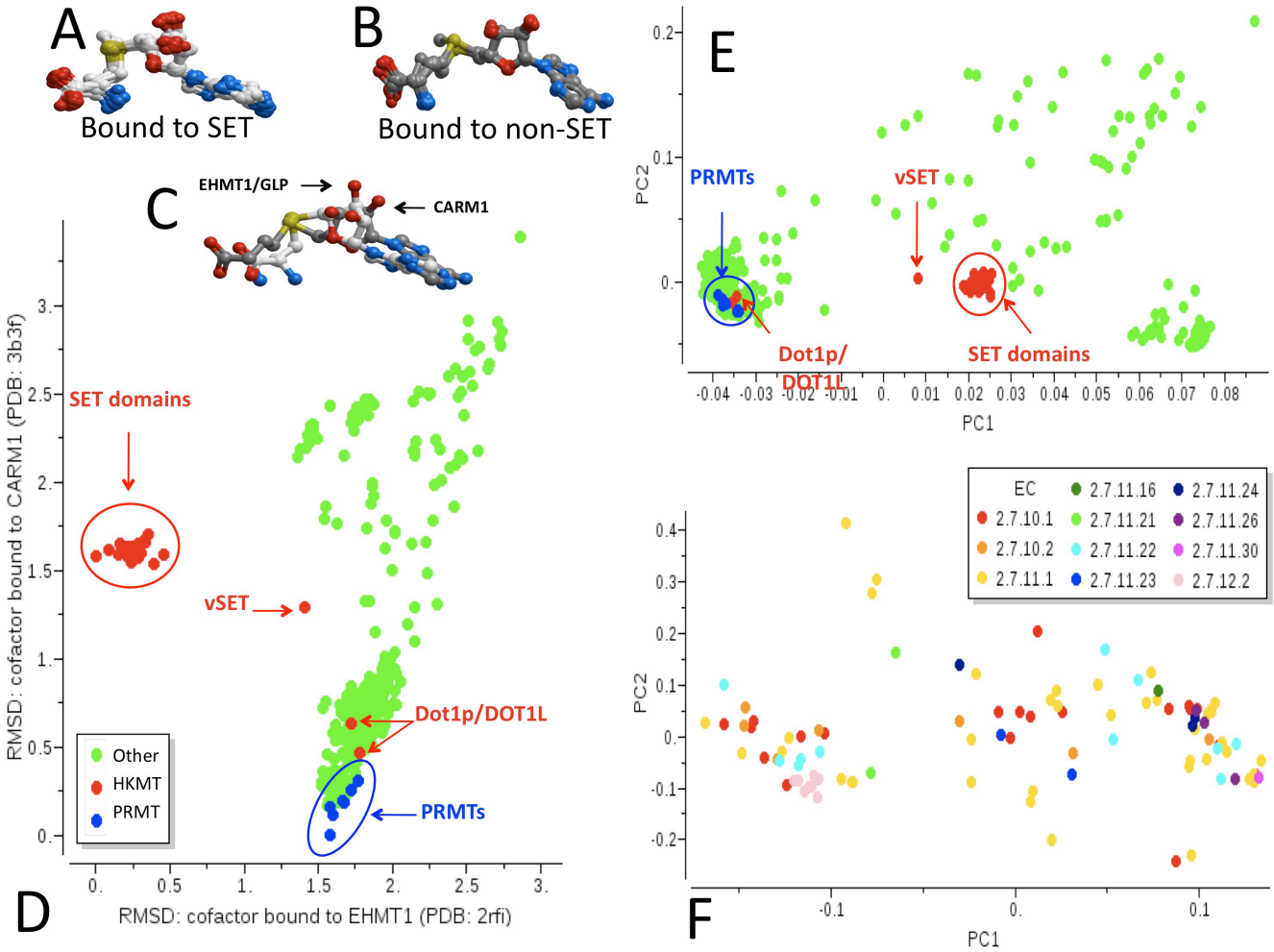

Figure 3: Conformational diversity of SAM/SAH in available PDB structures. The bound conformation is absolutely conserved in mammalian SET-domain HKMTs (A, 26 structures covering 10 proteins), and non-SET domain HMTs (B, 8 structures covering 4 proteins), but the two conformations are distinct (C, EHMT1/GLP complex in white versus CARM1 complex in gray). Plotting RMS deviations from EHMT1/GLP- and CARM1-bound conformations (D), or plotting conformational states based on the two principal components of their distribution (E) show that the conformation bound to SET domain

HMTs is conserved and distinct from that bound to other protein or non-protein methyltransferases. Using a similar representation as in E, the conformational diversity of all ATP/ADP/AMP/ANP molecules in complex with human kinases available in the PDB does not relate bound conformation to enzyme class (F, color-code based on enzyme class number).

Next, a matrix of pairwise RMSDs between all SAH/SAM conformations observed in the PDB was calculated. The values of the two principal components extracted from this matrix were plotted to cluster the enzymes in a 2D space (Figure 3E). Using this unbiased approach, the cofactor conformational space extracted from SET-domain structures was again the only one not overlapping with other targets. To compare the conformational diversity observed for methyltransferase cofactors with that of kinase cofactors, a similar clustering analysis was performed with 102 ATP/ADP/AMP/ANP structures co-crystallized to human kinases (Figure 3F – EC numbers are used as labels for the kinases: protein-Tyr kinases 2.7.10.x, protein-Ser/Thr kinases 2.7.11.x, and dual-specificity kinases [Tyr & Ser/Thr] 2.7.12.x). The only kinase subclass showing a distinct cofactor conformation (as observed for SET-domain HKMTs) was the mitogen-activated protein kinases (MAP2K – EC 2.7.12.2), however, in this case, all structures were from the same enzyme, MAP2K1. These results indicate that, unlike HMTs for which a correlation is observed between cofactor conformation and protein function, the bound conformations of kinases cofactor do not cluster according to the biology of their targets (defined by EC numbers). Furthermore, the cofactor conformation is not always conserved for a single protein, as demonstrated, for instance, by the various available structures of CDK2 (Figure 3F, only representative of the EC 2.7.11.22 class).

Three conclusions can be drawn from this ligand-based analysis. First, there is room to develop SAM competitors selective for SET domain HMTs. Second, PRMT inhibitors could be selective versus SET domain enzymes, but achieving selectivity against other non HMT SAM-binding proteins may be more challenging, at least as can be judged by ligand shape alone. Third, the conformational diversity of cofactors bound to HMTs is low, which is the opposite of what is observed for kinases. This could reflect an absence of chemical diversity at the HMT binding pocket, or the presence of a limited set of conserved interactions that impose the observed cofactor structure, within a structurally diverse pocket.

Cofactor-enzyme interactions

To address this question, we compared the network of hydrogen bonds formed with the cofactor across human SET-domain HMTs (lysine methyltransferases), non SET domain PRMTs (arginine methyltransferases) and in the non-SET domain lysine HMT DOT1L (Figure 4). While hydrophobic interactions are generally most contributing to binding enthalpy, lack of polar complementarity comes at an energetic cost, and hydrogen-bonds are driving specificity. While the number of hydrogen bonds with SET-domain HMTs varies from 7 for GLP and G9a to 10 for SETD2, a conserved network of six hydrogen bonds is observed in all structures (Figure 4A). These interactions are clustering at two distinct structural elements of the SET domain. First the backbone nitrogen and carbonyl oxygen of a histidine residue (H1170 in GLP) located at the canonical pseudo-knot motif of SET domain HMTs is forming two hydrogen bonds with the adenine ring of the cofactor. One residue upstream of this histidine is a conserved asparagine (N1169 in GLP), the side-chain of which extends towards and forms a hydrogen-bond with the amine nitrogen of the cofactor's methionine moiety. This conserved set of interactions is mapping at the pseudo-knot, a structural hallmark of SET-domain HMTs.[22] Second, a cluster of three backbone atoms located at the N-terminal extremity of the SET domain (M1105 and W1107 in GLP) are engaged in three conserved hydrogen bonds with the carboxylate and amine of the cofactor's methionine. This set of six hydrogen bonds observed in all representative structures of human SET-domain HMTs is distributed at both ends of SAM and dictates the conserved cofactor conformation. Interestingly, 5 of these 6 interactions are formed with HMT backbone atoms, and do not depend on the nature of the side-chains. This is probably accompanied by a weak evolutionary pressure for side-chain

conservation, which in turns translates into significant side-chain variability at the cofactor site (vide infra).

As observed for SET-domain HMTs, the hydrogen-bonding pattern observed between cofactor and PRMTs evolves around a conserved network of seven direct, and two water-mediated interactions (Figure 4B). First, two nitrogens of the purine are interacting with the conserved side-chain of E331 (PRMT3 nomenclature) and the backbone nitrogen of I330. Importantly, this pair is distinct from that observed in SET domain HMTs. Second, the carboxyl group of Asp302 forms a conserved pair of hydrogen bonds with the two hydroxyl groups of the ribose ring, a feature not observed in SET-domain HMTs. Third, both amine and carboxylate of the cofactor's methionine end are forming conserved hydrogen bonds with the enzymes. This pattern of interactions is conserved in DOT1L (Figure 4C), but a notable difference is the loss of a direct electrostatic contact between the cofactor's carboxylate and the guanidinium group of an arginine, conserved in PRMTs only (Supporting Information Figure 1). As previously observed, this extended array of interactions is spread along the entire cofactor molecule, and leaves little room for flexibility.

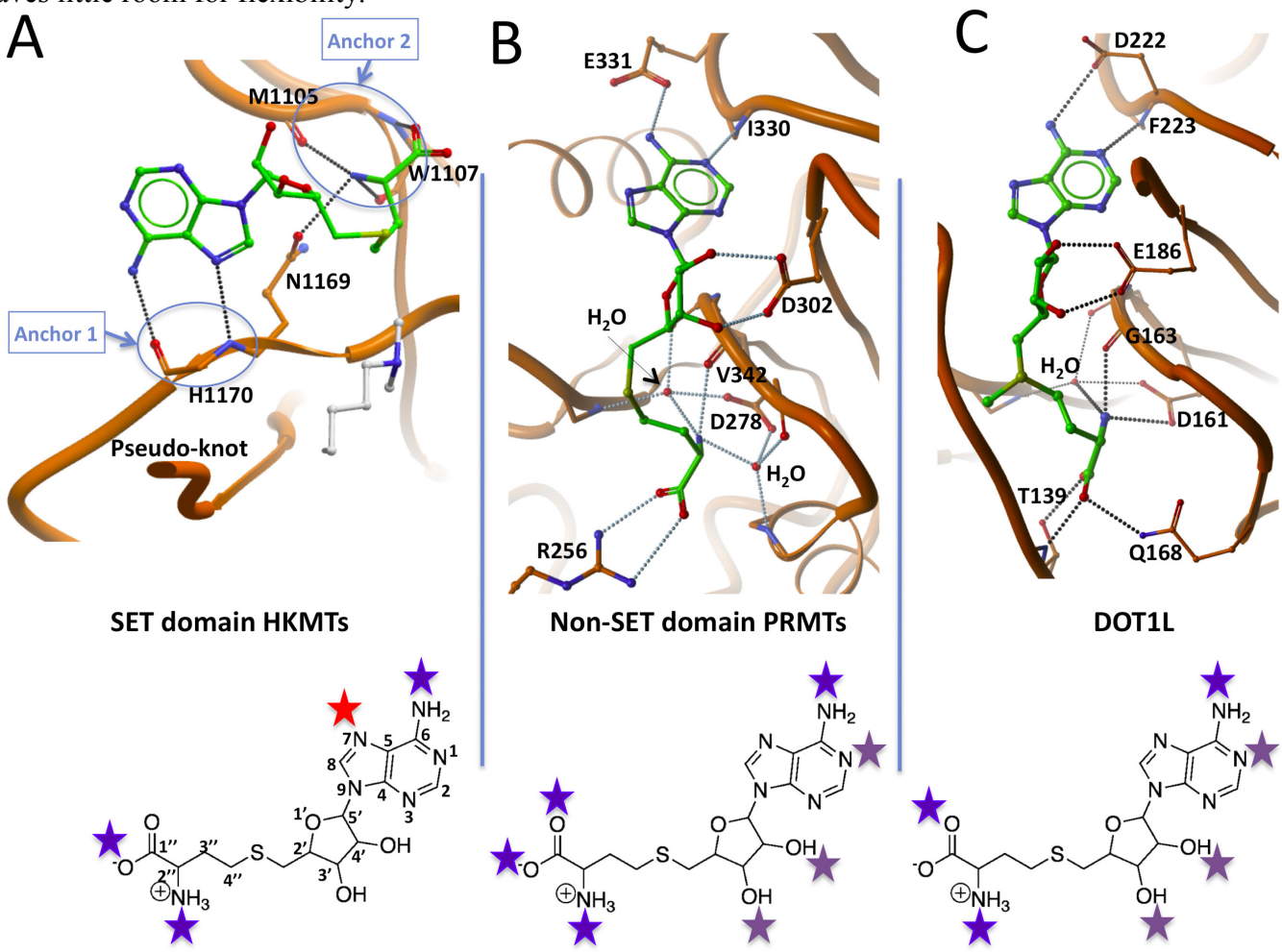

Figure 4: Conserved hydrogen bonds. A- Six hydrogen-bonds are conserved in all 10 human SET domain HMTs co-crystallized with cofactor, and are clustered either at the N-terminus or the pseudo-knot of the SET domain. The EHMT1/GLP structure is shown here (PDB: 2RFI). All hydrogen-bonds are with backbone atoms, with the exception of the amide side-chain carbonyl of a conserved asparagine. B- 9 hydrogen-bonds are conserved in structures of 3 PRMTs co-crystallized with cofactor (PRMT3, PDB: 2FYT, is shown). These interactions are mediated by conserved side-chains, backbone and water molecules. C- While the network of interaction at the adenosyl end observed in DOT1L (PDB: 1NW3) is conserved with PRMTs, the environment of the methionine moiety of SAM diverges significantly. Side-chain not engaged in a hydrogen-bond with the cofactor are not shown for clarity; green: SAM; white: di-methylated substrate lysine. Schematic 2D summaries of conserved hydrogen-

bonds are shown at the bottom (blue star: conserved hydrogen bond; red star: hydrogen-bond exclusively conserved in SET domain HMTs; purple star: hydrogen bonds exclusively conserved in non-SET domain HMTs). Atom naming conventions are indicated on the bottom left 2D sketch.

This analysis defines distinct hydrogen bonding patterns for SET and non-SET domain HMTs, which suggests that simple cofactor mimetics missing hydrogen bonding atoms that are systematically exploited in one family but not in the other may display interesting selectivity profiles.

Structural diversity of the cofactor pocket

To further investigate structural features that may be exploited to design selective SAM competitors, the chemical profile of the SAM binding pocket was interrogated with the GRID technology as implemented in the computational chemistry platform MOE (Chemical Computing Group Inc.).[27,28] Only residues within 4.5 Å of the bound ligand (i.e. residues delimiting the SAM binding pocket) were considered. We note here that this definition is restrictive, as ligands that only partially occupy the SAM binding pocket, and extend into the substrate lysine channel could also act as SAM competitors. Briefly, a water probe and a dry probe were scanned along the molecular surface of the cofactor pocket to identify hydrogen bond and hydrophobic interaction hotspots, respectively. This approach only considers the protein and ignores the bound small molecule (here the cofactor), the only role of which is to determine the site of interest. The resulting interaction field potentials, exclusively derived from the receptor, were mapped on a 2D projection of the cofactor (Figure 5). As previously mentioned, it should be noted that portion of a N-terminal helix expected to cover the ligand is missing from the PRMT1 structure, which may affect the result. Also, due to the difficulties of displaying 3D surfaces on a 2D structure, fields adjacent to the ligand atoms but not including them (R>0.5Å) were ignored in Figure 5.

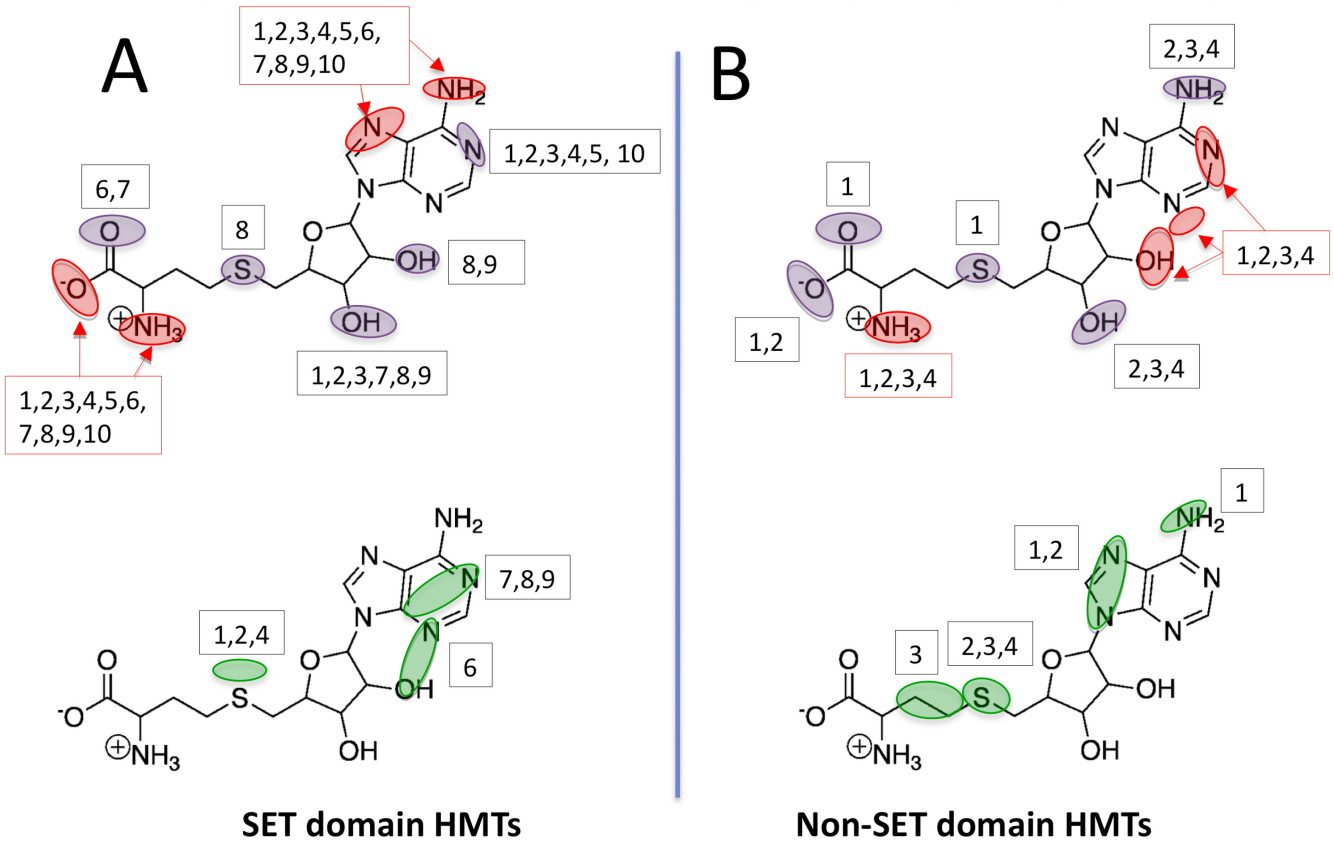

Figure 5: Receptor-based interaction hotspots. Hydrogen bonding (top) and hydrophobic (bottom) interaction potentials were derived from the structure of the target, and projected on a 2D representations of the cofactor. These interaction hotspots indicate the preferred physical properties of

small molecules binding at the cofactor site. Each hotspot is labeled with the identity of the HMTs where it is present. A- Numbering scheme for SET domain HMTs: 1. SETMAR [PDB: 3BO5]; 2. SETD2 [PDB: RH6L]; 3. G9A/EHMT2 [PDB: 3K5K]; 4. GLP/EHMT1 [PDB: 2RFI]; 5. SUV39H2 [PDB: 2R3A]; 6. SETD7 [PDB: 1XQH]; 7. SETD8 [PDB: 1ZKK]; 8. SMYD3 [PDB: 3MEK]; 9. SMYD1 [PDB: 3N71]; 10. MLL [PDB: 2W5Z]. B- Numbering scheme for non-SET domain HMTs: 1. DOT1L [PDB: 1NW3]; 2. PRMT3 [PDB: 2FYT]; 3. PRMT1 [PDB: 1OR8]; 4. CARM1 [PDB: 3B3F]. Hotspots observed in all structures are highlighted in red.

A first observation is that conserved hydrogen bonding hotspots (Figure 5, red boxes) are generally at the site of conserved hydrogen bond formation between HMTs and cofactor (Figure 4). This further suggests that interactions at these positions are important for binding, and may be necessary to achieve significant binding potency. On the other hand, none of the observed hydrophobic hot spots are conserved, which reflects the extremely polar nature of the pocket.

Interestingly, a hydrogen bonding hotspot at the position occupied by the N7 nitrogen of the purine ring (atom naming convention is shown Figure 4) is present in all SET domain HMTs, and absent from all non-SET domain enzymes; inversely, a polar hotspot next to the purine N3 nitrogen is observed in all non-SET domain structures, and absent from all SET domain HMTs (Figure 5). These preferred interaction sites could be exploited to design subfamily-specific cofactor mimetics. The interaction potentials also suggest avenues to engineer more refined selectivity profiles. For instance, hydrogen-bonds are favourable at the hydroxyl group on the 4' carbon of the ribose ring only in SMYD structures among SET domain enzymes (label "8,9", Figure 5A, top). SETD7 is unique in favouring a hydrophobic group at this position (label "6", Figure 5A, bottom). Another example is a polar interaction hotspot mapping at one of the carboxylate oxygens of the cofactor in SETD7 and SETD8, but no other SET domain HMT (label "6,7", Figure 5A). Surprisingly, the N1 nitrogen of the purine ring is located at a hydrogen bonding hotspot in all HMTs but SETD8, SMYD1 and SMYD3, where a hydrophobic group is preferred. Equally surprising is the absence of a polar hotspot and the preferred hydrophobicity mapping at the primary amine of the puring ring in the DOT1L structure (label "1", Figure 5B, bottom), while a polar group is more favourable in other HMTs. Although in GRID-type analyses it is quite usual to see a given position being part of polar and hydrophobic hot spots simultaneously, reflecting the proximity and favourable orientation of both types of groups, in this particular case, the suboptimal conformation of the D222 side-chain in the DOT1L structure may be responsible for the absence of the polar interaction hotspot.

This analysis indicates that, while the bound conformation of the cofactor is highly conserved in the two HMT families, the chemical environment of the cofactor varies.

Structural clustering of the cofactor pocket

We next evaluated whether this diversity is sufficient to design selective SAM competitors, and what inhibition selectivity profile may be achieved. We used IsoCleft to detect 3D atomic similarities between pockets, and calculate structural distances between all SAM binding sites.[30] An HMT similarity matrix was consequently generated (Supporting Information Table 4), and a corresponding dendogram was produced (Figure 6, top – see Methods section for details).

A general agreement was observed between the phylogenetic classification derived from multiple sequence alignment of the enzymes' catalytic domain (Figure 1), and the atom based classification derived from the structural chemistry of the SAM pocket (Figure 6). (1) PRMTs and HMTs were separated into distinct clusters. (2) Local clusters of close homologues were sometimes conserved. For instance GLP is the closest homologue of G9a (78% sequence identity), followed by SUV39H2 (36% sequence identity) in the domain-based tree. A similar classification can be seen in the pocket-based dendogram, where GLP is the closest neighbour of G9a (3D structural distance 0.43), followed by SUV39H2 (3D structural distance 0.62). (3) The sequences and structures of the catalytic domains of SMYD1 and SMYD3 are very distinct from all other HMTs (Figure 1, PDB: 3N71 and 3MEK resp., [37]),

which is recapitulated by the very distinct clustering of their cofactor site (Figure 6). The closest pocket is that of DOT1L, but at 85% structural distance, it should be considered as unrelated.

While the HMT dendogram is indicative of the relative structural similarity of the cofactor binding sites, it remains unclear how this correlates in terms of opportunity to develop selective SAM competitors. One way to address this question is to evaluate whether the closest pockets are sufficiently remote structurally to design selective inhibitors. If the case, it would be reasonable to infer that selective inhibition is chemically tractable at all sites. A detailed comparison of the two closest pairs indentified, PRMT3-CARM1, and GLP-G9a (distance of just over 0.4, corresponding to an IsoCleft similarity $LTS_{3D} = 0.573$) is depicted Figure 6. In each case, the few divergent side-chains lining the cofactor pocket are shown. The binding sites are highly conserved: clearly it would be very challenging to design selective inhibitors. However, the presence of a single divergent side-chain in a favourable orientation, such as R1226 in GLP (Q1169 in G9a), or Q303 in PRMT3 (A216 in CARM1) may be sufficient to achieve selective inhibition (Figure 6, top).

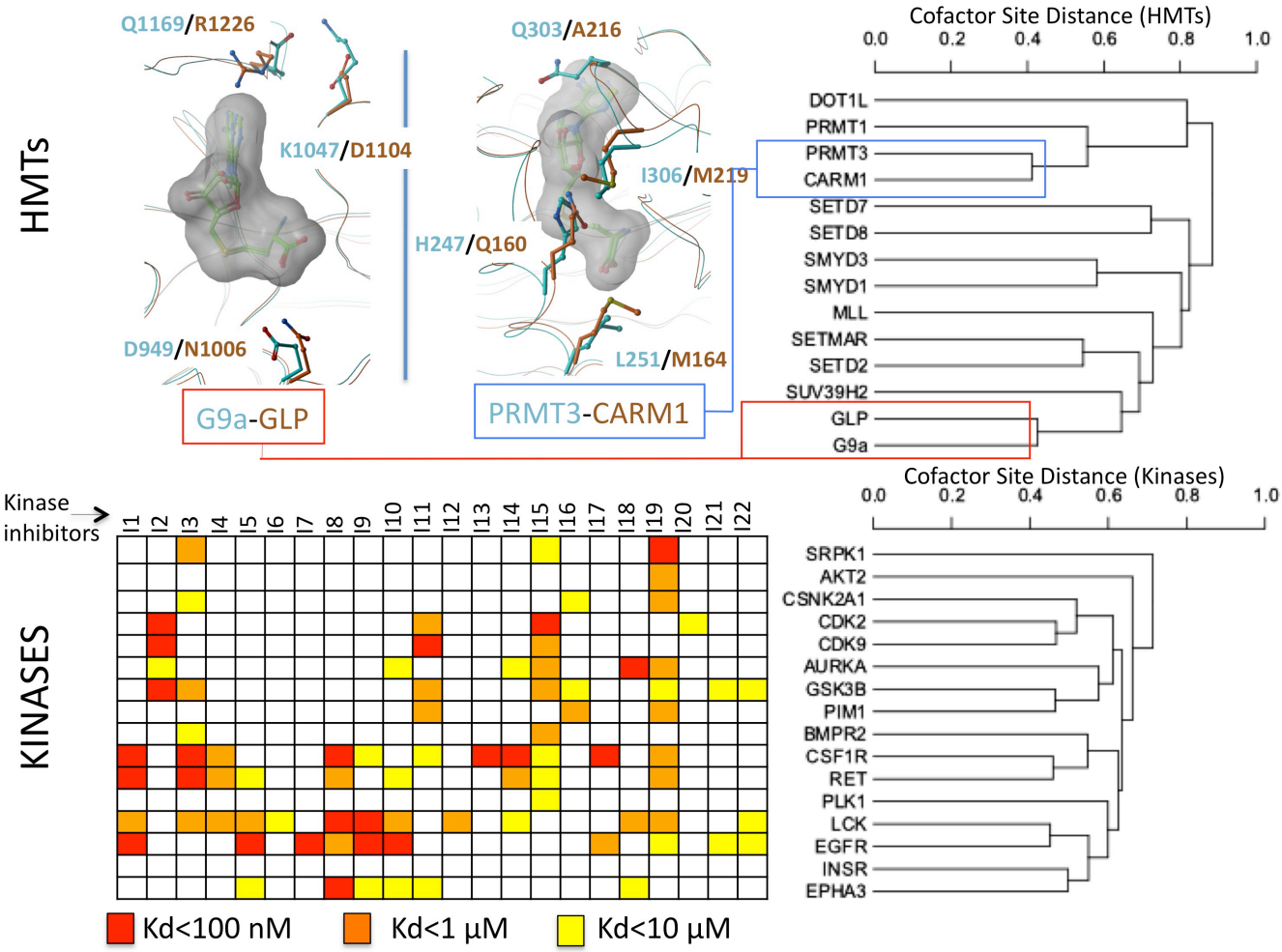

Figure 6: 3D, atom-based comparison of cofactor sites. The program IsoCleft (Najmanovich et al.[30]) was used to calculate structural distances between cofactor sites of HMTs (top dendogram) and kinases (bottom dendogram), based on their atomic composition, irrespective of amino-acid sequence. Corresponding structural divergences are shown for the closest HMT pairs (top left). For the kinase set, the observed distances are compatible with development of selective inhibitors (I1 to I32). Red: K_d<100 nM. Orange: K_d< 1 µM. Yellow: K_d<10 µM. White: K_d> 10 µM (Ambit panel, Karaman et al.[7], bottom left).

Another way to evaluate the structural diversity of HMT cofactor sites is to compare them with that of kinase cofactor sites. A vast body of chemogenomics data is available for kinases, and IsoCleft distances can be compared to experimental selectivity profiles of inhibitors. All structures of human kinases in complex with AMP/ADP/ATP/ANP were extracted from the PDB. For each gene, a representative structure was selected if it was in the DFG-in conformation and there was no missing residue in the vicinity of the cofactor pocket. This selection process resulted in 16 structures (Supporting Information Table 3). For both kinases and HMTs, the binding pocket was defined as residues within 6 Å of the cofactor and 2 Å of the surface (see Methods section for details). ATP and SAM are close in size, and so are their binding pockets. The kinase and HMT sets are therefore comparable both in number of structures and in the composition of the binding sites investigated. A matrix of IsoCleft distances was generated for the 16 ATP sites (Supporting Information Table 5), and the corresponding dendograms produced (Figure 6, bottom).

The closest pairs of ATP sites identified were separated by IsoCleft distances slightly above 0.4, which is comparable with distances separating the closest SAM sites. To judge whether such distances are compatible with selective pharmacological inhibition, we extracted from the Ambit kinase set all Type I inhibitors (i.e. ATP competitors targeting the DFG-in conformation) having some activity against at least one of our 16 kinases (Supporting Information Table 6).[7] The dissociation constant of these compounds for each kinase is indicated in the heat map Figure 6 (blank cells indicate that the compound was tested but was inactive or had a K_d > 10 μM). Some correlation can be observed between the dendogram derived from the structure of the ATP-bound site, and the selectivity profile of the inhibitors. For instance, the CDK2 and CDK9 sites are closely related, and so are the arrays of their inhibitors. The same observation can be made for the LCK-EGFR pair, or the CSF1R-RET pair. Selective inhibitors able to dissociate two structurally similar ATP sites are rare. Notable exceptions include inhibitor I7 (CP-724714), which has a K_d <100 nM for EGFR but > 10 μM for the structurally close ATP site of LCK (Figure 6 and Supporting Information Table 5). Another example is inhibitor I17 (MLN-518) with a K_d < 100 nM for CSF1R and > 10 μM for RET. No inhibitor is shown that substantially dissociates the two CDK sites.

This analysis suggests that pharmacological dissociation of binding sites separated by structural distances less than 0.4 is very challenging. However, when distances are 0.5 or higher, numerous selective ligands can be found (compare, among many examples, CSNK2A1 and CDK, or INSR and EPHA3 inhibitors). This is in agreement with the structural superimposition suggesting that selective inhibition of GLP versus G9a (at 0.4 structural distance) may be very challenging (Figure 6 top). This also suggests that the numerous HMT sites separated by structural distances greater than 0.5 (Figure 6, top) may be selectively targeted by SAM competitors. The fact that pharmacological dissociation can be achieved between ATP sites separated by IsoCleft distances of 0.5 does not prove that the same rule holds for HMT cofactor sites. The kinase data presented here should only be used as a reference to better judge the diversity distribution of HMTs. In the light of the kinase data, it is reasonable to assume that selective SAM competitors can be developed. In the end, this hypothesis can only be verified by the discovery of such molecules.

Together, these results indicate that, in spite of the poor variability of bound cofactor conformation across HMTs, the cofactor binding site displays sufficient structural diversity to develop small molecule SAM competitors with refined selectivity profiles.

While we believe that the physico-chemical analysis presented here is conceptually more appropriate, we understand that a more classical comparison of the binding pockets relying on sequence alignment can appear more meaningful, or complementary to some. Therefore, we provide as Supporting Information an analysis of the pocket diversity in terms of amino-acid composition (Supporting Information Figure 1 to 5).

DISCUSSION

In this work, the structural chemistry of the cofactor binding site of human HMTs was systematically analyzed across representative structures of 14 genes. The druggability of this site was predicted to vary between enzymes, with potent binders expected for PRMTs and DOT1L, but less so for SETD7. Comparison of the bound conformation of the cofactor or analogues suggested that SAM competitors targeting SET domain proteins are less likely to inhibit non-HMT SAM binding proteins. We showed that the poor conformational diversity of the bound cofactor relies on the presence of distinct patterns of hydrogen-bonds acting as anchor points specific to SET or non-SET domain HMTs. However, the structural diversity surrounding these anchor points is significant within both subfamilies, and may be exploited to design selective ligands.

While druggability predictions of the SAM pocket were based exclusively on the residues surrounding the cofactor, all HMTs have a channel, generally narrow, that allows transfer of the methyl group from the donating cofactor to the accepting nitrogen of the lysine or arginine substrate. It is therefore conceivable that weak inhibitors binding at a cofactor site with mediocre druggability (as is predicted for SETD7 for instance) could be significantly enhanced by adding a tail that would insert into the lysine/arginine channel. A similar strategy was recently proposed to develop non-reactive amine analogues of SAM bearing a guanidinylated tail against PRMTs.[38]

A surprising finding of this study was the large difference in conformational distributions of kinase and HMT cofactors: HMT-bound SAM is restricted to only two possible conformations, kinase-bound ATP can adopt a diverse range of conformations (Figure 3). Both the ATP site of kinases and the SAM site of HMTs are known to have dynamic structures.[39,40] Consequently, one may expect variable conformations of the cofactor for both gene families. This however does not apply to SET domain HMTs, as the dynamics of the cofactor binding site relies primarily on the motion of the post-SET domain, which folds onto SAM in an induced-fit mechanism where both protein and ligand are locked in place.[40] SAM does not bind to HMTs if the post-SET domain is unfolded, as illustrated by numerous PRDM structures (e.g. PRDM2, PDB: 2QPW). Additionally, the purine ring is anchored to the backbone of the pseudoknot histidine of SET-domain HMTs, while the opposite extremity of the molecule is forming a conserved hydrogen bond with the asparagine immediately juxtaposed to the histidine: both extremities of the cofactor are tied to adjacent and conserved residues, which leaves little room for flexibility (Figure 4A). This does not apply to non-SET domain HMTs. In this case, the fact that SAM fully occupies a structurally conserved pocket is probably at the origin of the lack of conformational diversity. This is in contrast with the rather flexible ATP binding site of kinases (both the Glycin-rich and activation loops are mobile).

The dissected mechanism of SAM binding can inform inhibitor design. General guidelines for both HMT sub-families can be drawn: potent ligands probably need to recapitulate hydrogen bonds at the two anchor points conserved in SET-domain enzymes (Figure 4A). In the case of non-SET domain proteins, a conserved acidic residue forming hydrogen bonds with the ribose hydroxyl groups (Figure 4B,C) is creating a strong polar interaction potential that cannot be ignored by inhibitors (Figures 5B). Another important polar interaction hotspot present in all structures is at the amine group of the cofactor's methionine moiety. Other lessons, more specific to individual targets, can also guide medicinal chemistry efforts. Examples include the hydrophobic potentials surrounding the cofactor nitrogen N1 in SMYDs, N3 in SETD7, and N7 in DOT1L and PRMT3.

CONCLUSION

Following recent clinical successes, protein families involved in epigenetic signaling are becoming the center of considerable attention in the drug discovery community.[8,9] Among these, histone methyltransferases play a critical role in the epigenetic control of gene expression, and it is expected that HMT inhibitors will reach the clinic.[10] The development of peptide competitors was recently established experimentally as a valid approach to inhibit HMTs with high potency and selectivity.[18,19] Using cofactor-bound structures of 14 human HMTs or close homologues, we analyzed systematically the structural chemistry of the SAM binding site, highlighting commonalities and divergences in shape,

druggability and chemical environment that could be exploited towards the development of inhibitors. Importantly, our analysis suggests that, as demonstrated for kinases, the structural diversity observed at the cofactor site is probably sufficient to achieve selective inhibition.

ACKNOWLEDGMENTS

V.C.-S. acknowledges the Natural Sciences and Engineering Research Council of Canada for funding. This work was supported by the Structural Genomics Consortium. The SGC is a registered charity (number 1097737) that receives funds from the Canadian Institutes for Health Research, the Canadian Foundation for Innovation, Genome Canada through the Ontario Genomics Institute, GlaxoSmithKline, Karolinska Institutet, the Knut and Alice Wallenberg Foundation, the Ontario Innovation Trust, the Ontario Ministry for Research and Innovation, Merck & Co., Inc., the Novartis Research Foundation, the Swedish Agency for Innovation Systems, the Swedish Foundation for Strategic Research and the Wellcome Trust.

Supporting Information Available. Additional tables, and figures are presented as Supporting Information. Table 1: Druggability scores , Table 2 and 3: residues used for the IsoCleft diversity analysis, Table 4 and 5: matrices of IsoCleft similarities, Table 6: K_d for 22 kinase inhibitors. Residues lining the binding pocket of CARM1 (Figure S1), EHMT2/G9a (Figure S2), SETD2 (Figure S3), MLL (Figure S4), SETD7 (Figure S5). This material is available free of charge at http://www.pubs.acs.org

REFERENCES

1. Frye, S. V. Structure-activity relationship homology (SARAH): a conceptual framework for drug discovery in the genomic era. *Chem. Biol.* **1999,** *6*, R3-R7.
2. Bredel, M.; Jacoby, E. Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat. Rev Genet.* **2004,** *5*, 262-275.
3. Jacoby, E.; Schuffenhauer, A.; Floersheim, P. Chemogenomics knowledge-based strategies in drug discovery. *Drug News Perspect.* **2003,** *16*, 93-102.
4. Klabunde, T. Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. *Brit. J. Pharmacol.* **2007,** *152*, 5-7.
5. Harris, C. J.; Stevens, A. P. Chemogenomics: structuring the drug discovery process to gene families. *Drug Discov. Today* **2006,** *11*, 880-888.
6. Fernandez, A.; Sanguino, A.; Peng, Z.; Crespo, A.; Ozturk, E.; Zhang, X.; Wang, S.; Bornmann, W.; Lopez-Berestein, G. Rational drug redesign to overcome drug resistance in cancer therapy: imatinib moving target. *Cancer Res.* **2007,** *67*, 4028-4033.
7. Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P. A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2008,** *26*, 127-132.
8. Copeland, R. A.; Olhava, E. J.; Scott, M. P. Targeting epigenetic enzymes for drug discovery. *Curr. Opin. Chem. Biol.* **2010,** *14*, 505-510.
9. Xu, W. S.; Parmigiani, R. B.; Marks, P. A. Histone deacetylase inhibitors: molecular mechanisms of action. *Oncogene* **2007,** *26*, 5541-5552.
10. Copeland, R. A.; Solomon, M. E.; Richon, V. M. Protein methyltransferases as a target class for drug discovery. *Nat. Rev. Drug Discov.* **2009,** *8*, 724-732.
11. Spannhoff, A.; Hauser, A. T.; Heinke, R.; Sippl, W.; Jung, M. The emerging therapeutic potential of histone methyltransferase and demethylase inhibitors. *ChemMedChem* **2009,** *4*, 1568-1582.
12. Zhang, Y.; Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Gene. Dev.* **2001,** *15*, 2343-2360.

13. Kouzarides, T. Histone methylation in transcriptional control. *Curr. Opin. Genet. Dev.* **2002,** *12*, 198-209.
14. Völkel, P.; Angrand, P.-O. The control of histone lysine methylation in epigenetic regulation. *Biochimie* **2007,** *89*, 1-20.
15. Martin, C.; Zhang, Y. The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Bio.* **2005,** *6*, 838-849.
16. Bedford, M. T.; Richard, S. Arginine Methylation: An Emerging Regulator of Protein Function. *Mol. Cell* **2005,** *18*, 263-272.
17. Chang, Y.; Zhang, X.; Horton, J. R.; Upadhyay, A. K.; Spannhoff, A.; Liu, J.; Snyder, J. P.; Bedford, M. T.; Cheng, X. Structural basis for G9a-like protein lysine methyltransferase inhibition by BIX-01294. *Nat. Struct. Mol. Biol.* **2009,** *16*, 312-317.
18. Liu, F.; Chen, X.; Allali-Hassani, A.; Quinn, A. M.; Wasney, G. A.; Dong, A.; Barsyte, D.; Kozieradzki, I.; Senisterra, G.; Chau, I.; Siarheyeva, A.; Kireev, D. B.; Jadhav, A.; Herold, J. M.; Frye, S. V.; Arrowsmith, C. H.; Brown, P. J.; Simeonov, A.; Vedadi, M.; Jin, J. Discovery of a 2,4-diamino-7-aminoalkoxyquinazoline as a potent and selective inhibitor of histone lysine methyltransferase G9a. *J. Med. Chem.* **2009,** *52*, 7950-7953.
19. Liu, F.; Chen, X.; Allali-Hassani, A.; Quinn, A. M.; Wigle, T. J.; Wasney, G. A.; Dong, A.; Senisterra, G.; Chau, I.; Siarheyeva, A.; Norris, J. L.; Kireev, D. B.; Jadhav, A.; Herold, J. M.; Janzen, W. P.; Arrowsmith, C. H.; Frye, S. V.; Brown, P. J.; Simeonov, A.; Vedadi, M.; Jin, J. Protein lysine methyltransferase G9a inhibitors: design, synthesis, and structure activity relationships of 2,4-diamino-7-aminoalkoxy-quinazolines. *J. Med. Chem.* **2010,** *53*, 5844-5857.
20. Greiner, D.; Bonaldi, T.; Eskeland, R.; Roemer, E.; Imhof, A. Identification of a specific inhibitor of the histone methyltransferase SU(VAR)3-9. *Nat. Chem. Biol.* **2005,** *1*, 143-145.
21. DeLano, W. L. Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* **2002,** *12*, 14-20.
22. Qian, C.; Zhou, M. M. SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell. Mol. Life Sci.* **2006,** *63*, 2755-2763.
23. Schultz, J.; Milpetz, F.; Bork, P.; Ponting, C. P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U S A* **1998,** *95*, 5857-5864.
24. Krause, C. D.; Yang, Z. H.; Kim, Y. S.; Lee, J. H.; Cook, J. R.; Pestka, S. Protein arginine methyltransferases: evolution and assessment of their pharmacological and therapeutic potential. *Pharmacol. Therapeut.* **2007,** *113*, 50-87.
25. Feng, Q.; Wang, H.; Ng, H. H.; Erdjument-Bromage, H.; Tempst, P.; Struhl, K.; Zhang, Y. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **2002,** *12*, 1052-1058.
26. Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009,** *49*, 377-389.
27. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985,** *28*, 849-857.
28. Boobbyer, D. N.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.* **1989,** *32*, 1083-1094.
29. McConkey, B. J.; Sobolev, V.; Edelman, M. Quantification of protein surfaces, volumes and atom-atom contacts using a constrained Voronoi procedure. *Bioinformatics* **2002,** *18*, 1365-1373.
30. Najmanovich, R.; Kurbatova, N.; Thornton, J. Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites. *Bioinformatics* **2008,** *24*, i105-i111.
31. Schapira, M.; Totrov, M.; Abagyan, R. Prediction of the binding energy for small molecules, peptides and proteins. *J. Mol. Recognit.* **1999,** *12*, 177-190.
32. Janin, J.; Chothia, C. Role of hydrophobicity in the binding of coenzymes. Appendix. Translational and rotational contribution to the free energy of dissociation. *Biochemistry* **1978,** *17*, 2943-2948.

33. Henrich, S.; Salo-Ahen, O. M.; Huang, B.; Rippmann, F. F.; Cruciani, G.; Wade, R. C. Computational approaches to identifying and characterizing protein binding sites for ligand design. *J. Mol. Recognit.* **2010,** *23*, 209-219.
34. Sheridan, R. P.; Maiorov, V. N.; Holloway, M. K.; Cornell, W. D.; Gao, Y. D. Drug-like density: a method of quantifying the "bindability" of a protein target based on a very large set of pockets and drug-like ligands from the Protein Data Bank. *J. Chem. Inf. Model.* **2010,** *50*, 2029-2040.
35. Fedorov, O.; Marsden, B.; Pogacic, V.; Rellos, P.; Muller, S.; Bullock, A. N.; Schwaller, J.; Sundstrom, M.; Knapp, S. A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases. *Proc. Natl. Acad. Sci. U S A* **2007,** *104*, 20523-20528.
36. Martin, J. L.; McMillan, F. M. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **2002,** *12*, 783-793.
37. Sirinupong, N.; Brunzelle, J.; Ye, J.; Pirzada, A.; Nico, L.; Yang, Z. Crystal structure of cardiac specific histone methyltransferase SmyD1 reveals unusual active site architecture. *J. Biol. Chem.* **2010,** *285,* 40635-40644.
38. Dowden, J.; Hong, W.; Parry, R. V.; Pike, R. A.; Ward, S. G. Toward the development of potent and selective bisubstrate inhibitors of protein arginine methyltransferases. *Bioorg. Med. Chem. Lett.* **2010,** *20*, 2103-2105.
39. Nolen, B.; Taylor, S.; Ghosh, G. Regulation of protein kinases; controlling activity through activation segment conformation. *Mol. Cell.* **2004,** *15*, 661-675.
40. Wu, H.; Min, J.; Lunin, V. V.; Antoshenko, T.; Dombrovski, L.; Zeng, H.; Allali-Hassani, A.; Campagna-Slater, V.; Vedadi, M.; Arrowsmith, C. H.; Plotnikov, A. N.; Schapira, M. Structural biology of human H3K9 methyltransferases. *PLoS One* **2010,** *5*, e8570.

TOC GRAPHIC: