RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

Activity landscape sweeping: Insights into the mechanism of inhibition and optimization of DNMT1 inhibitors

J. Jesús Naveja[a,b] and José L. Medina-Franco[a*]

[a]Facultad de Química, Departamento de Farmacia, Universidad Nacional Autónoma de México, Avenida Universidad 3000, México, D.F. 04510, México

[b]Facultad de Medicina, PECEM, Universidad Nacional Autónoma de México, Avenida Universidad 3000, México, D.F. 04510, México

[*]Address correspondence to this author: E-mails: medinajl@unam.mx; jose.medina.franco@gmail.com Tel. +5255-5622-3899. Ext. 44458

ABSTRACT

The interest in developing inhibitors of DNA methyltransferases (IDNMT) as modifiers of epigenetic features for the treatment of several chronic diseases is rapidly increasing. Herein, we present insights of a chemoinformatic characterization of IDNMT focused on the analysis of the chemical space and structure-activity relationships (SAR) using activity landscape modeling (ALM). Analysis of the chemical space revealed two main groups of compounds whose chemical structures are associated with either cofactor analogs or non-nucleoside compounds. The ALM showed that non-nucleoside compounds have a continuous SAR while cofactor analogs have a rough SAR with several deep activity cliffs. Molecular modeling helped to explain the structural basis of the activity cliffs. The significance of the results is threefold: 1) the combined analysis of chemical space with activity landscape gave rise to a novel 'activity landscape sweeping' strategy that enabled a better structure-based interpretation of the SAR; 2) it is feasible –and advisable– to develop predictive models for non-nucleoside IDNMT studied in this work, and 3) structure-based interpretation of the SAR gave clear insights into the molecular mechanism of inhibition of novel IDNMT suggesting specific strategies to optimize the activity of leads compounds.

Keywords: activity cliffs generators, cancer, chemical space, density SAS maps, epigenetics, epi-informatics, protein-ligand interaction fingerprints, structure-activity relationships.

List of abbreviations: 3D, three-dimensional; ALM, activity landscape modeling; DNMT, DNA methyltransferases; ECFP, Extended Connectivity Fingerprints; HTS, high-throughput screening; IDNMT, inhibitor of DNA methyltransferase; MDS, myelodysplastic syndrome; MOE, Molecular Operating Environment; PCA, principal component analysis; PDB, Protein Data Bank; PLIF, Protein Ligand Interaction Fingerprint; QSAR, quantitative structure-activity relationships; RMSD, root-mean-square deviation; SAH, *S*-Adenosyl-*L*-homocysteine; SAM, *S*-Adenosyl-*L*-methionine; SAR, structure-activity relationships; SAS maps, structure-activity-similarity maps.

INTRODUCTION

The term 'Epigenetics' was initially defined as "the interactions of genes with their environment, which brings the phenotype into being".[1] Epigenetic drug discovery is an attractive research area in oncology and for the treatment of other chronic diseases associated with epigenetic alterations, particularly those influenced by the environment. There are several epigenetic targets which are broadly classified in three major groups, namely; readers, writers and erasers of the epigenetic information.[2]

DNA methylation is a major epigenetic change that regulates gene expression in the genome of organisms that range from viruses to humans.[3] DNA methylation is regulated by the family of enzymes DNA methyltransferases (DNMTs). DNMTs are responsible for the covalent addition of a methyl group from the cofactor *S*-adenosyl-L-methionine (SAM or AdoMet) (Figure 1) to the carbon atom 5 of cytosine, preferably within CpG dinucleotides. Also, as a product of the methylation mechanism, *S*-Adenosyl-*L*-homocysteine (SAH) is generated.[4] In mammals, four DNMT enzymes have been identified: DNMT1 (the most abundant, it is a maintenance methyltransferase that acts on hemimethylated DNA); DNMT3A and DNMT3B (*de novo* methyltransferases that are capable of generating new methylation patterns in DNA), and DNMT3L that is associated with DNMT3A and DNMT3B, enhancing their activity.

<Insert Figure 1 here>

The structure of DNMTs can be organized into a C-terminal catalytic domain and an N-terminal regulatory domain. The catalytic domain of all DNMTs shares a common structure called "AdoMet (SAM)-dependent Mtase fold". The N-terminal domain is involved in distinguishing hemi- and unmethylated DNA. There are several three-dimensional (3D) structures of different domains of DNMTs, including the catalytic one.[5]

The role of DNMTs in carcinogenesis has been subject of intense research during the last ten years. Currently, there are two inhibitors of DNMT (IDNMT) in clinical use: 5-azacytidine and decitabine (Figure 1) both approved by the United States Food and Drug Administration –FDA– for the treatment of myelodysplastic syndrome (MDS).[6] However, these two drugs are cytosine analogues that are incorporated into DNA, which implies they are unspecific and have high toxicity due to their mutagenic effects that may occur in somatic cells. Many therapies involving IDNMT are under investigation, mainly as sensitizers to therapy, since epigenetic changes may be involved in rapid adaptation of cancerous cells to therapy. In addition to cancer, DNMTs are attractive targets for the treatment of other chronic and degenerative diseases such as Alzheimer's and psychiatric conditions. Also, DNA methylation has been involved in autoimmune diseases and inherited disorders.[7]

The low specificity and high toxicity of 5-azacytidine and decitabine has prompted the search for novel and specific IDNMTs. Currently there is a relatively large number of IDNMT and/or DNA demethylating compounds that have been obtained from different sources such as natural products, synthetic compounds, drugs approved for therapeutic indications other than cancer and high-throughput screening (HTS). As part of these efforts, computational analyses have been successfully implemented to model IDNMT and to identify novel inhibitors.[8]

Over the past few years, the structure and activity of compounds tested as IDNMT have been collected in public repositories such as ChEMBL.[9] The increasing amount of structure-activity data of IDNMT opens up the possibility to conduct systematic structure-activity relationships (SAR) studies, such as quantitative SAR (QSAR). Nevertheless, it has been recognized that typical QSAR studies

usually assume that compounds with similar structures have similar activity *i.e.,* a 'smooth' SARs. It is well known that compounds with high structural similarity but low activity similarity *i.e.*, 'activity cliffs',[10] reduce the predictive ability of QSAR models.[11, 12] Therefore, the early detection of activity cliffs is a convenient step before attempting to develop models such as QSAR.[13] Similarly, it is advisable to conduct detailed descriptive analysis to understand the SAR before developing predictive models.[14] Thus far, limited studies have been reported to navigate and describe the SAR of a large set of IDNMT in a systematic manner.

In this work, we report a chemoinformatic-based characterization of the SAR of a dataset of 280 compounds tested as IDNMT1 and deposited in ChEMBL. The analysis had three specific aims: a) characterization of the structural diversity and distribution in chemical space of the data set; b) descriptive SAR analysis using the concept of activity landscape modeling and c) structure-based interpretation of the activity cliffs. To the best of our knowledge this work represents one of the first activity landscape studies of IDNMT1. Indeed, it has been recently recognized that activity landscape modeling (ALM) is a convenient approach to explore systematically the SAR of screening data sets focused on epigenetic targets.[15] The characterization of the chemical space distinguished two major types of chemical structures with different activity landscape. As part of the first aim it was developed a novel 'activity landscape sweeping' approach, that is, a dissection of the global activity landscape (global SAR) into smaller but more structural interpretable local landscapes (local SARs). The structure-based interpretation of the SAR of the activity cliffs gave key insights into the molecular mechanism of inhibition of active molecules. This analysis also prompted for structural modifications to lead compounds to continue developing IDNMT as potential epi-drugs or epi-probes.

METHODS

Dataset

A data set of 280 compounds with different (no duplicate) chemical structures and activity against DNMT1 was obtained from ChEMBL (version 20)[9] and recent literature. Only compounds with

reported IC_{50} values obtained in enzymatic inhibitory assay were included in the analysis. The activity range for the compounds in the dataset was 18.6-1,600,000 nM (pIC_{50} range 7.73-2.80). Molecules were pre-processed with the 'washing' workflow implemented in Molecular Operating Environment (MOE) (version 2010.10,[16]). During the washing procedure, only the largest molecular structure was retained; counter ions, if present, were removed and protonation states were set to neutral. Visualization of the chemical structures was performed with MOE and Data Warrior (version 4.1.1).[17]

Structural similarity

In order to measure the structural similarity for each pair of compounds in the dataset (39,060 pair-wise comparisons) we employed two structural fingerprints of different design, namely, Molecular Access System (MACCS) 166-bits (dictionary based fingerprints)[16] as implemented in MOE, and Extended Connectivity Fingerprints (ECFP; radial based fingerprints),[18] with neighborhood radius of 2 as implemented in MayaChemTools (http://www.mayachemtools.org). The structural similarity was computed with the Tanimoto coefficient.[19, 20]

Data fusion

In order to explore the effect of data fusion in this study, two approaches were implemented to combine the similarity values computed with MACCS keys and ECFP: a) fusion mean *i.e.*, calculation of the mean values[21] and b) Z-fusion *i.e.*, addition of the Z-transformed values of both fingerprints.[15]

Visual representation of chemical space

To obtain a visual representation of the chemical space[22] we conducted a principal components analysis (PCA) on the similarity matrices computed for the 280 molecules with the two fingerprints and the two fusion approaches. This method has been broadly used to obtain visual representations of the chemical space.[20, 23] The PCA was performed with the FactoMineR R package version 1.29. For visualization, the ggplot2 R package was used (http://www.R-project.org/).[24] K-means method was

also conducted with R using in-house scripts to perform clustering of the PCA's output. Further details of the PCA and K-means analysis employed are in the Supporting Information.

SAS maps

The activity landscape of IDNMT was explored using Structure-Activity Similarity (SAS) maps.[25] SAS and related 2D- and 3D-SAS maps have been extensively employed to describe the SAR of a large number of data sets.[26-28] Features of SAS maps, including their advantages and disadvantages, are elaborated elsewhere.[29] Briefly, a typical SAS map is a 2D plot of the structural similarity vs. the potency difference of all possible pairs of compounds in a data set. The structural similarity can be computed with any similarity method. Aggregation of similarity values using data fusion may be implemented.[26, 30] To facilitate the visual interpretation of the SAS maps, 'density SAS maps' were used in this work. A density SAS map represents the frequency of data points usually with a continuous color scale.[15] Density SAS maps were generated for the entire data set (*e.g.*, analysis of the '*global* activity landscape') and for subsets of compounds that emerged from the analysis of the compounds in chemical space (*e.g.*, analysis of the '*local* activity landscape').

Activity cliffs generators

'Activity cliffs generators'[31] were defined as active compounds recurrent (frequency > 1) in the activity cliff region of the activity landscape. In turn, the 'activity cliff' region of the landscape was defined as the quadrant in the SAS map that contains pairs of compounds with high structure similarity and high potency difference. A quantitative definition of 'high' structure similarity is not straightforward. Herein, we considered high values those with two standard deviations above the mean similarity of the entire data set. Two values to define 'very high' and 'high' potency difference were used to distinguish 'deep' and 'shallow' activity cliffs, respectively *i.e.*, ΔpIC_{50} > 2 standard deviations above the mean (2SD) and 2SD >ΔpIC_{50} > 1.

Structure-based interpretation of activity cliffs

In order to provide a structure-based rationalization of the activity cliffs that emerged from the activity landscape analysis, we conducted computational studies with the crystallographic structure of the catalytic domain of human DNMT1 co-crystallized with SAH (PDB ID: 3PTA).[32] Notably, the conformation of the catalytic domain of human DNMT1 (Protein Data Bank (PDB) ID: 3PTA) shows the prevention of the *de novo* methylation mechanism by an auto-inhibitory linker that blocks DNA to reach the catalytic site. For docking studies we employed MOE 2010 using default parameters. The binding cavity was defined differently for nucleoside and non-nucleoside structures (*vide infra*). The docking protocol was validated by re-docking the co-crystal structure with a root-mean-square deviation (RMSD) of: 1.22 $Å^2$ for the best scored pose. The docking poses were further analyzed using Protein Ligand Interaction Fingerprints (PLIFs)[33] implemented in MOE as detailed below. PLIFs, also called structural interaction fingerprints, capture key 3D interactions between a ligand and a protein in 1D. PLIFs ware recently used in activity landscape studies.[34]

Molecular modeling of nucleoside activity cliffs

In order to propose a structure-based explanation of the activity cliffs of the cofactor analogues, we worked under the hypothesis that these compounds bind in the cofactor binding site. We also assumed that, in general, the compounds have a binding orientation comparable to that of SAH. Therefore, we conducted docking using pharmacophoric constraints that were obtained from the crystallographic binding mode of SAH. The pharmacophore had four points: hydrogen bond donor with Cys1191, hydrogen bond acceptor with Met1169, anion and hydrogen bond acceptor with Gly1150 and aromatic ring interacting with Phe1145. During docking in MOE it was enabled the partial homology criterion with the formation of at least three of the pharmacophoric constraints, the rest of the options remained as default (Figure S2 in the Supporting Information). The docked poses were post-processed with

PLIFs available in MOE to identify the most relevant interactions of the activity cliffs *i.e.*, potential hot spots.[35]

Molecular modeling of non-nucleoside activity cliffs

Since there is no experimental evidence of the binding site for most of the non-nucleoside IDNMT1, we searched for potential binding sites in the catalytic domain of DNMT1 using Site Finder in MOE with default parameters. Then, the most active compound forming activity cliffs was docked with MOE in the absence of the co-factor considering all putative binding sites. The geometry of the docking pose with the best docking score was minimized with the cofactor present using the MMFF94x force field as implemented in MOE. To conduct the minimization, the ligand and nearby residues of the binding pocket (with atoms within a radius of 4.5 Å) were selected. Default parameters were used. In order to detect putative 'interaction cliffs' (*i.e.*, ligand-target complexes with high structural and interaction similarity but a large potency difference of the ligands),[34] the optimized docked pose of the most active compound was used as a template to conduct flexible alignment of the other cliff forming compounds with which it formed activity cliffs. The flexible alignment was done in MOE using default parameters. For comparison, both regular and pharmacophore-constrained docking (see pharmacophore in Figure S3) were performed in the binding site proposed for the most active molecule (with the energy-minimized conformation of the protein).

RESULTS AND DISCUSSION

Structural diversity analysis

A structural diversity analysis of the 280 compounds was conducted using two molecular fingerprints of different design.[30] The distribution of the similarity values (Figure S1 in the Supporting Information) showed that, in general, this is a relatively diverse set with, for example, mean Tanimoto

similarity values of 0.63 (MACCS keys) and 0.11 (ECFP). This diversity is comparable to that reported for other sets of compounds tested for other therapeutic indications.[36]

Visualization of chemical space

Activity landscapes have been defined as methods that find the association between structure similarity and activity similarity.[37] Therefore, the next step in this work was to explore the distribution of the 280 compounds in chemical space. Figure 2 shows a visual representation of the chemical space obtained with PCA of the similarity matrix computed with ECFP and the Tanimoto coefficient. Data points are colored by the pIC_{50} values using a continuous color scale from red (more active) to gray (less active).

<Insert Figure 2 here>

Figure 2 shows two major clusters in chemical space herein labeled as cluster A (45 compounds) and cluster B (235 compounds), respectively. Both groups of compounds have active and inactive molecules *e.g.*, red and gray points. Furthermore, the active compounds in each cluster are not further grouped suggesting that they are structurally diverse.

Visual inspection of all compounds in each cluster revealed that all the chemical structures in cluster A have a purine ring in their structure and are structurally related to the co-factor SAM. In contrast, molecules in cluster B are non-nucleoside. Representative structures from each cluster are depicted in Figure 1 and are mapped into the visual representation of the chemical space of Figure 2. The visual representation of the chemical space in Figure 2 also suggested that molecules in cluster B (non-nucleoside) are structurally more diverse than the molecules in cluster A. Not surprisingly, the distribution of the similarity values (Figure S1 in the Supporting Information) confirmed that the non-nucleoside set has a higher structural diversity than the SAM-related compounds. This is because no further distinction is made on the type of chemical structures. In contrast, all compounds in cluster A are chemically related to SAM.

It is possible to further divide the non-nucleosides in smaller sub-sets chemically related. For instance, K-means clustering shows that 3-6 subgroups would provide an efficient clustering in terms

of number of clusters and within group's sums of squares (see the Supporting Information for a detailed explanation on K-means methodology followed). However, clustering in two groups already diminished by more than 40% of the within groups sums of squares (see Figure S4). Herein, we analyze the activity landscape of two clusters to discuss local SAR as general as possible. Undoubtedly, additional studies can be extended to analyze smaller clusters and provide information of more local SARs.

Equivalent clusters A and B were identified in the PCA of the combined ECFPs and MACCS keys similarity matrices using the fusion approaches detailed in the Methods section (Figures S5 in the Supporting Information). Interestingly, MACCS keys alone did not lead to the identification of the two clusters (Figure S5a); this can be attributed to the low resolution of this fingerprint.[30]

Overview of activity landscape

Global activity landscape (global SAR)

Figure 3a shows a density SAS map generated with ECFP and Tanimoto for the entire data set with 280 compounds. The four major quadrants (I-IV) are distinguished in the figure. The activity cliff zone (region IV) is further divided in two sub-regions (IVa and IVb) that distinguish the shallow from the deep activity cliffs, depending on the potency difference (1 vs. 2 log units; see the Methods section). The amount of data points in each different region of the SAS map is visually represented with a continuous color scale from red (more data points) to gray (fewer). Table 1 summarizes the fraction of data points in each region (I-IV) of the SAS map.

<Insert Table 1>

<Insert Figure 3 here>

Figure 3a and Table 1 indicate that, overall, IDNMT1 have a heterogeneous SAR with data points in the continuous and discontinuous regions of the SAR (zones III and IV).[15] Noteworthy, the scaffold hop, more recently called 'similarity cliffs'[38] region has the highest density of data points (92.6%). This indicates that there are quite different chemical structures with similar activity. Note however that

both compounds in the pair may be either active or inactive. Figure 3a and Table 1 also shows the presence of shallow and deep activity cliffs with a relatively small fraction of the entire data set (0.79 and 0.16%, respectively). The overall low frequency of activity cliffs is in agreement with the low frequency of activity cliffs observed for data sets for other molecular targets.[26-28, 30]

The high density of data points in the similarity cliff region of the SAS maps and the two main clusters of compounds distinguished in the chemical space analysis, prompted us to conduct analysis of local activity landscapes of clusters A and B. As discussed in the next section, the chemical structures of compounds in each cluster, plus the knowledge of the mechanism of DNA methylation, led to an interpretable SAR.

Local activity landscapes (local SARs)

Figure 3b and 3c show the density SAS maps generated for the 235 non-nucleosides and 45 SAM analogues identified in the analysis of the chemical space (clusters B and A in Figure 2, respectively). Table 1 summarizes the number and percentage of data points in the four major regions of the local SAS map. The number and fraction of the deep and shallow cliffs (regions IVa and IVb, respectively) are also summarized in the same table.

The lower fraction of similarity cliffs for SAM-related analogues (4.6%) vs. the fraction of similarity cliffs for the non-nucleoside analogues (94.8%) is in agreement with the type of structures and molecular diversity in each cluster. Indeed, the visual representation of the chemical space (Figure 2) and distribution of ECFP/Tanimoto similarity values for the compounds in each cluster (Figure S1) yield consistent results. Similarly, the higher percentage of compounds in the smooth SAR region (III) for SAM analogues (62%) as compared to the percentage of compounds for non-nucleoside analogues (1.2%) (Table 1) is in line with the structural diversity of the chemical structures of each type of compounds.

As mentioned above, the distribution of data points in the similarity cliff and smooth regions of the SAS maps are expected from the type of chemical structures. But surprisingly, for SAM related

analogues there is a larger fraction of deep and shallow activity cliffs as compared to the fraction of cliffs in the entire data set (4.9% and 28.2% vs. 0.16% and 0.79%, respectively; Table 1). In sharp contrast, the fraction of activity cliffs for the non-nucleosides is lower (0.05% and 0.11%, respectively, Table 1). These results indicate that SAM related analogues may be enriched with activity cliff generators.[31] The next sections discuss the activity landscapes of each set of compounds, *i.e.*, local activity landscapes. A brief analysis of the activity landscape of SAM-related compound is mentioned first followed by a more extensive discussion of the landscape of the non-nucleosides. We elaborated more on the non-nucleosides since they are currently more attractive as IDNMT1.[39]

Activity landscape of SAM-related compounds

As discussed above, SAM-related compounds have a discontinuous SAR with several (nearly 5%) of activity cliffs. For comparison, the non-nucleosides have 0.05% of activity cliffs. Despite the fact these proportions are dependent of the current contents of ChEMBL *i.e.*, the numbers may change as more activity data is published, this is a clear indication of the roughed nature of the SAR of SAM-related compounds. This observation highlights the challenge to conduct lead optimization of IDNMT1 using SAM-related compounds besides the risk of hitting other methyltransferases. Deep activity cliffs generators of SAM-related compounds (cluster A) are shown in Figure 1 (N1-N4). The compound pairs with whom they form activity cliffs are illustrated in Figures S7-S10 in the Supporting Information.

Activity landscape of non-nucleoside compounds

The activity landscape of the non-nucleoside compounds is more continuous than the landscape of SAM-related molecules. The landscape of the non-nucleosides is characterized by a small fraction of activity cliffs of which a small number are deep activity cliffs (Table 1).

As discussed in the literature, activity cliffs are rich in SAR information since they point to specific structural changes that have a large impact in the biological activity. In an activity landscape study

based on structural fingerprints, the interpretability of the activity cliffs is a key component.[37] In other words, the SAR of the activity cliffs should be easily translated in terms of specific structural changes. In the local activity landscape of non-nucleoside molecules we identified two major types of compounds with high ECFP/Tanimoto similarity whose chemical structures are structurally related, namely: compounds identified by HTS and structures related to SGI-1027.[40] All pairs of compounds from HTS are shallow cliffs and are shown in Figure 4 and Figure S11 of the Supporting Information. From the 30 shallow activity cliffs found in the SAS map for non-nucleoside compounds, 16 (53%) compounds were found to be from HTS assays (Figure 4 and Table 2). A considerable number of screenings and confirmatory assays were performed for these compounds, as found in PubChem.

<Insert Figure 4 here>

<Insert Table 2 here>

In the activity landscape of non-nucleoside molecules the deepest activity cliffs as well as the most relevant in medicinal chemistry were the structures related to the quinolone-based inhibitor SGI-1027. This compound is one of the most promising IDNMT1 that has been recently subject of a medicinal chemistry optimization program (*vide infra*). Therefore, in the next section we describe studies focused on the interpretation at a molecular level of activity cliffs related to SGI-1027.

Deep and shallow cliffs for compounds related to SGI-1027

Systematic analysis of all pairwise comparisons of the structure and activity of the 235 non-nucleoside compounds (39,060 comparisons), readily uncovered that analogues of SGI-1027 are the compounds with the most dramatic changes in activity associated with a small change in the structure. In fact, 14/30 (47%) of the shallow and 15/15 (100%) of the deep activity cliffs found in the non-nucleoside database were found to be related to the compounds recently synthesized by Valente *et al.*[41] The chemical structures of the nine activity cliff-forming compounds are presented in Figure 5. The enzymatic inhibitory activity of the nine compounds vs. DNMT1 was recently reported using the same assay conditions.[41] These molecules were synthesized as part of a hit-to-lead optimization program of

SGI-1027 which showed high potency in enzyme and cell assays.[40] Compounds in Figure 5 are regioisomers of SGI-1027.

<Insert Figure 5 here>

In order to describe the analogues of the lead compound, Valente et al. considered that SGI-1027 is composed of four fragments (4-aminoquinoline + 4-aminobenzoic acid + 1,4-phenylenediamine + 2,4-diamino-6-methylpyrimidine) linked in sequence with *para*/*para* orientation.[41] The most active compound in this series was CHEMBL3126646 which can be regarded as the *meta*/*meta* regioisomer of SGI-1027 (CHEMBL2336409).[41] Table 3 summarizes the deep activity cliffs that form CHEMBL3126646. It must be noted that this compound is the most important activity cliff generator in the database *i.e.*, it is the most prevalent compound within the activity cliff region of the SAS map.[31] The deepest activity cliffs of the *meta/meta* regioisomer are formed with *ortho* regioisomers such as CHEMBL3126644, 3126647, 3126648, 3126649 with potency differences of two or more logarithmic units (Figure 5 and Table 3).

<Insert Table 3 here>

Valente et al. reported docking models of CHEMBL3126646 with crystallographic structures of DNMT1. It was concluded from that studies that this molecule could interact with the CXXC auto-inhibitory domain of DNMT1 and be close to SAM but without making interactions with the cofactor or competing with any of the interactions that SAM makes.[41] However, no structure-based explanation of the large potency difference of the significantly less active SGI-1027 analogues (*e.g.*, *ortho* regioisomers) was explored. A structure-based interpretation of the activity cliffs that form the most active compound is elaborated in the next section.

Structural interpretation of representative activity cliffs

Structure-based interpretation of the activity cliffs can help to understand the SAR of data sets at the molecular level and provide insights to optimize the activity.[31, 35, 42] In this study, the availability of structure information of the 3D coordinates of DNMT1 enabled a structure-based interpretation of the

activity cliffs using molecular modeling. Of note, despite the fact docking studies of all compounds reported in ChEMBL as IDNMT1 is warranted, this is out of the scope of this work. Herein, we focus on the structure-based analysis of the most representative activity cliffs. As detailed in the Methods section, we employed different modeling strategies to study the activity cliffs related to SAM analogues and to CHEMBL3126646 based on the structural information available for each type of compounds.

SAM-related activity cliffs

The experimental co-crystal structure of SAH bound in the co-factor site of DNMT1 was the starting point of the structure-based studies of relevant activity cliffs related to SAM. As described in the Methods section, we worked under the assumption that SAM-related activity cliffs bind in the co-factor binding site. We conducted docking studies using pharmacophoric constrains of the compounds forming activity cliffs with the four most prominent activity cliff generators related to SAM-analogues: CHEMBL557902, CHEMBL560106, CHEMBL552309, and CHEMBL559283 (labeled N1-4 in Figures 1 and 2). The binding poses were analyzed using PLIFs.

Results of the PLIFs for the activity cliff generator CHEMBL557902 plus 11 related (paired) compounds are shown in Figure 6. The chemical structures are shown in Figure S8 of the Supporting Information. The data matrix in Figure 6a summarizes the protein-ligand contacts between the best two poses of 12 docked molecules and DNMT1. In this matrix, the rows represent the docked poses of the 12 molecules. The columns are the fingerprint bits indicating the amino acid residues that make at least one contact with one of the compounds. A black cell in the matrix indicates that a contact is present between the intersecting compound and amino acid residue *i.e.*, fingerprint bit turned ‘on’. In contrast, a white cell means that there is no contact *i.e.*, fingerprint bit turned ‘off”. Figure 6 revealed that interactions with Gly1223 (backbone hydrogen bond donor), Glu1266 (ionic attraction) and Arg1312 (both side chain hydrogen bond acceptor and ionic attraction) were found in the active SAM-analogue (CHEMBL557902) but not in the compounds with much lower pIC_{50} values. Similar analyses were

performed with the three remaining activity cliff generators related to SAM (Figures S12-S14). It was concluded that that the loss of a hydrogen bond donor that could interact with Asp1190 is generating cliffs for CHEMBL557902, CHEMBL560106, and CHEMBL559283.

<Insert Figure 6 here>

Non-nucleoside activity cliffs

As stated above, in this study we focused on the structure-based interpretation of the most significant activity cliffs of the non-nucleoside molecules, *i.e.*, structural analogues of SGI-1027. In particular, we focus on the analysis of the deep activity cliffs formed with CHEMBL3126646 (Figure 5). As explained above, these cliffs have large potency differences (>2 standard deviations above the dataset's mean) and the high ECFP/Tanimoto similarity (ranging from 0.38 to 0.75, see Table 3) of these activity cliffs is structurally interpretable.

There is no co-crystallized structure available for the most active compound CHEMBL3126646 with DNMT1 (this is the case for every non-nucleoside IDNMT1). Therefore, its precise binding region is unknown. In order to explore the putative binding zone, before docking all activity cliff forming compounds, CHEMBL3126646 was docked with DNMT1 as detailed in the Methods section. Results were compared with the experimental biochemical results and docking studies recently published for this molecule. Figure 7 shows the optimized docking model. In this model, CHEMBL3126646 is close to but does not occupy the binding region of the co-cofactor (as predicted for other type of IDNMT1[43, 44]). Remarkably, a potential hydrogen bond interaction was found between the carbonyl oxygen of CHEMBL3126646 and the O_2' oxygen atom of the co-crystal SAH. The molecule is able to make hydrogen bond contacts with the backbone of Ala647, and π-π interactions (T-shape) with the side chain of Phe648 of the CXXC domain. In addition, CHEMBL3126646 makes hydrophobic interactions with the side chains of Met696, Glu698, and Ala699 of the CXXC domain (See Figures 7 and Figure S15 in the Supporting Information for a 3D and 2D ligand interactions representation, respectively). The possibility of this inhibitor or making 'sandwich' interactions with both the CXXC domain and the co-factor in DNMT1 is in agreement with

the docking study reported by Valente et al.[41] Therefore, it is plausible that CHEMBL3126646 inhibits DNMT1 by a mechanism we previously proposed for SGI-1027 *i.e.*, stabilization of the autoinhibitory linker.[45] This hypothesis is further supported by the experimental evidence that CHEMBL3126646 seems to do not compete with the co-factor.

<Insert Figure 7 here>

The binding mode for the most active compound proposed herein also explains the activity cliffs to a large extent. The most pronounced *e.g.*, deepest activity cliffs with compound CHEMBL3126646 (Table 3) are regioisomers with at least one *ortho* substitution: CHEMBL3126644, 3126647, 3126648, 3126649 and *ortho*/*ortho* regioisomer (compound 9, as numbered in the Valente et al.[41] paper). In agreement with Valente et al., [41] the shape of the *ortho* regiosomers may not adopt the extended conformation required to stabilize inhibitory linker domain. Flexible alignment of SGI-1027 analogues with the most active compound CHEMBL3126646 (Figure 8) clearly shows the very different shape of the more active *meta*/*meta* and other non-*ortho* regioisomers (Figure 8A) as compared to the inactive *ortho* regioisomers of SGI-1027 (Figure 8B). Docking of the *ortho* containing compounds with DNMT1 (data not shown) showed the loss of the interaction with the co-factor also highlighting this key interaction of CHEMBL3126646.

<Insert Figure 8 here>

Preliminary regular and pharmacophore-constrained docking studies of the eight compounds related to CHEMBL3126646 (Figure 5) were conducted with a crystallographic structure of DNMT1. The docking poses were post-processed with PLIFs as detailed in the Methods section. Results are summarized in Figure S16a. In order to explore the protein-ligand contacts that may differentiate 'active' from 'inactive' compounds, the significance analysis implemented in MOE was performed. For this analysis we considered as "active" a compound with $pIC_{50} > 5$. Figure S16b shows that there are not statistically significant differences that might distinguish active from inactive molecules. This reflects the fact that *ortho* regioisomers are not unable of stretching to the required extent, but the

energy necessary to do so is higher, mainly due to their intermolecular interactions. Further computational analyses are required to test this hypothesis (see below section of Future directions).

Insights into the structure-based optimization of lead molecules

The structure-based interpretation of the activity cliff generators associated with CHEMBL3126646 leads to strategies to further optimize the affinity with DNMT1 and possible the biological activity. For example, addition of cationic moieties at both sides of the molecule would provide the structure with stronger ionic interactions. In addition, a hydrogen bond may be more easily formed with Asp701 in the CXXC domain if a further small elongation of the molecule is produced by adding a carbon or an aromatic ring into the structure. It remains to conduct additional molecular modeling analysis of the designed structures to further guide the structure-based optimization of quinolone-based inhibitors.

CONCLUSIONS

Analysis of the distribution in chemical space of 280 compounds tested as IDNMT1 readily revealed two well-defined groups of structures: SAM-analogues and non-nucleoside compounds. Local SAR analysis showed that the two clusters have different activity landscapes. Molecules similar to the cofactor SAM have a heterogeneous landscape with the presence of deep activity cliffs *i.e.*, similar molecules with large potency difference. In sharp contrast, non-nucleoside compounds have a smoother SAR with few shallow activity cliffs and fewer deep activity cliffs. The significance of this observation is that, at least in principle, almost any active small non-SAM-like molecule in this data set can be used as a query in similarity-based virtual screening. Also, in general, the non-nucleoside data set can be the starting point to develop predictive models. Of course, these conclusions depend on the current contents of ChEMBL. As the coverage of the chemical space of non-SAM-like compounds increases the corresponding landscape may change and more activity cliffs may emerge.

The structural interpretation of the activity cliffs indicated that SAM-related analogues contain several pharmacophoric interactions that are substantial to determining its potency. Therefore, even

small changes in its structure may produce deep activity cliffs. Hence, SAM-analogues may not be suitable for classical predictive approaches that assume linear relationships.

Structure-based analysis of the most relevant non-nucleoside activity cliff generator, a regioisomer of SGI-1027 developed recently, supported the hypothesis that this type of molecules may act through a stabilization of the auto-inhibitory linker domain of DNMT1. Results of the docking model are in agreement with the SAR of the deepest activity cliffs involving CHEMBL3126646. Results are also in agreement with the biochemical analysis showing that CHEMBL3126646 is not a competitive inhibitor of the co-factor.

During the course of this work we concluded that density SAS maps are convenient graphical representations that enhance the interpretation of the SAS maps. It was also highlighted the convenience of performing 'activity landscape sweeping' before the analysis of the activity landscape of a data set. The activity landscape sweeping presented in this work led to the exploration of local activity landscapes that provided interpretable SAR results and provided insights for the structure-based optimization of lead compounds as IDNMT1.

Future directions

As part of this chemoinformatics work, we focused on the initial docking and molecular modeling of active compounds forming the most representative activity cliffs. A next logical step of this study is to conduct the molecular modeling of all the active molecules including those with a smooth SAR. Similarly, comprehensive molecular modeling studies should be conducted to explain, at the molecular level, other activity cliffs (*e.g.*, non-quinolone based) identified in this work. As part of these studies, induced-fit docking and/or other methods that consider protein flexibility should be used. These studies are ongoing in our group and will be reported in due course. It remains to explore the similarity cliffs (scaffold hops) that emerged from this work. Finally, other perspective is to develop predictive models (such as QSAR) for non-nucleoside compounds. During writing of this manuscript, a paper reporting QSAR models of IDNMT1 was published.[46]

ACKNOWLEDGMENTS

This work was funded by the National Autonomous University of Mexico (UNAM), grant PAIP 5000-9163 to JLMF.

REFERENCES

1. C. H. Waddington, *Int. J. Epidemiol.* 2012, **41**, 10-13.
2. S. Knapp and H. Weinmann, *ChemMedChem*, 2013, **8**, 1885-1891.
3. K. D. Robertson, *Oncogene*, 2001, **20**, 3139-3155.
4. A. Jeltsch, *ChemBioChem*, 2002, **3**, 274-293.
5. J. L. Medina-Franco, J. Yoo and A. Dueñas-Gonzalez, in *Epigenetic Technological Applications*, Ed. Y. G. Zheng, Elsevier, 2015, Ch. 13, pp. 265-290.
6. E. J. B. Derissen, J. H. Beijnen and J. H. M. Schellens, *The Oncologist*, 2013, **18**, 619-624.
7. C. Gros, J. Fahy, L. Halby, I. Dufau, A. Erdmann, J.-M. Gregoire, F. Ausseil, S. Vispé and P. B. Arimondo, *Biochimie*, 2012, **94**, 2280-2296.
8. J. L. Medina-Franco, O. Méndez-Lucio, J. Yoo and A. Dueñas, *Drug Discovery Today*, 2015, **20**, 569-577.
9. A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani and J. P. Overington, *Nucleic Acids Res.*, 2012, **40**, D1100-D1107.
10. G. M. Maggiora, *J. Chem. Inf. Model.*, 2006, **46**, 1535.
11. R. Guha and J. H. Van Drie, *J. Chem. Inf. Model.*, 2008, **48**, 1716-1728.
12. A. Golbraikh, E. Muratov, D. Fourches and A. Tropsha, *J. Chem. Inf. Model.*, 2014, **54**, 1-4.
13. M. Cruz-Monteagudo, J. L. Medina-Franco, Y. Pérez-Castillo, O. Nicolotti, M. N. D. S. Cordeiro and F. Borges, *Drug Discovery Today*, 2014, **19**, 1069-1080.

14. J. Medina-Franco, G. Navarrete-Vázquez and O. Méndez-Lucio, *Fut. Med. Chem.*, 2015, **7**, 1197-1211.

15. J. J. Naveja and J. L. Medina-Franco, *Exp.Opin.Drug Discov.*, 2015, in press. DOI:10.1517/17460441.2015.1073257

16. Molecular Operating Environment (MOE), version 2010.10, Chemical Computing Group Inc., Montreal, Quebec, Canada. Available at http://www.chemcomp.com.

17. T. Sander, J. Freyss, M. von Korff and C. Rufener, *J. Chem. Inf. Model.*, 2015, **55**, 460-473.

18. D. Rogers, R. D. Brown and M. Hahn, *J. Biomol. Screening*, 2005, **10**, 682-686.

19. P. Jaccard, *Bull. Soc. Vaudoise Sci. Nat.*, 1901, **37**, 547-579.

20. J. L. Medina-Franco and G. M. Maggiora, in *Chemoinformatics for Drug Discovery*, Ed. J. Bajorath, John Wiley & Sons, Inc., 2014, Ch. 15, pp. 343-399.

21. P. Willett, *J. Chem. Inf. Mod.*, 2013, **53**, 1-10.

22. J. L. Medina-Franco, K. Martínez-Mayorga, M. A. Giulianotti, R. A. Houghten and C. Pinilla, *Curr. Comput.-Aided Drug Des.*, 2008, **4**, 322-333.

23. R. Kraft, A. Kahn, J. L. Medina-Franco, M. L. Orlowski, C. Baynes, F. Lopez-Vallejo, K. Barnard, G. M. Maggiora and L. L. Restifo, *Dis. Models & Mech.*, 2013, **6**, 217-235.

24. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, Springer New York, 2009.

25. V. Shanmugasundaram and G. M. Maggiora, presented in part at the 222nd ACS National Meeting, Chicago, IL, United States, Chicago, IL, United States, August 26-30, 2001.

26. J. Pérez-Villanueva, R. Santos, A. Hernández-Campos, M. A. Giulianotti, R. Castillo and J. L. Medina-Franco, *Bioorg. Med. Chem.*, 2010, **18**, 7380-7391.

27. A. Yongye, K. Byler, R. Santos, K. Martínez-Mayorga, G. M. Maggiora and J. L. Medina-Franco, *J. Chem. Inf. Model.*, 2011, **51**, 1259-1270.

28. F. Renee, L. Travis M., R. G. Santos, A. Morales, A. Nefzi, G. S. Welmaker, J. L. Medina-Franco, M. A. Giulianotti, R. A. Houghten and L. N. Shaw, *J. Med. Chem.*, 2015, **58**, 3340-3355.

29. J. L. Medina-Franco, *J. Chem. Inf. Model.*, 2012, **52**, 2485-2493.

30. J. L. Medina-Franco, K. Martínez-Mayorga, A. Bender, R. M. Marín, M. A. Giulianotti, C. Pinilla and R. A. Houghten, *J. Chem. Inf. Model.*, 2009, **49**, 477-491.

31. O. Mendez-Lucio, J. Perez-Villanueva, R. Castillo and J. L. Medina-Franco, *Mol. Inf.*, 2012, **31**, 837-846.

32. J. Song, O. Rechkoblit, T. H. Bestor and D. J. Patel, *Science*, 2011, **331**, 1036-1040.

33. S. C. Brewerton, *Curr. Opin. Drug Discovery Dev.*, 2008, **11**, 356-364.

34. O. Méndez-Lucio, A. J. Kooistra, C. d. Graaf, A. Bender and J. L. Medina-Franco, *J. Chem. Inf. Mod.*, 2015, **55**, 251-262.

35. B. Seebeck, M. Wagener and M. Rarey, *ChemMedChem*, 2011, **6**, 1630-1639.

36. J. L. Medina-Franco, K. Martínez-Mayorga, A. Bender and T. Scior, *QSAR Comb. Sci.*, 2009, **28**, 1551-1560.

37. D. Stumpfe, Y. Hu, D. Dimova and J. Bajorath, *J. Med. Chem.*, 2014, **57**, 18-28.

38. P. Iyer, D. Stumpfe, M. Vogt, J. Bajorath and G. M. Maggiora, *Mol. Inf.*, 2013, **32**, 421-430.

39. A. Erdmann, L. Halby, J. Fahy and P. B. Arimondo, *J. Med. Chem.*, 2014, **58**, 2569–2583.

40. J. Datta, K. Ghoshal, W. A. Denny, S. A. Gamage, D. G. Brooke, P. Phiasivongsa, S. Redkar and S. T. Jacob, *Cancer Res.*, 2009, **69**, 4277-4285.

41. S. Valente, Y. W. Liu, M. Schnekenburger, C. Zwergel, S. Cosconati, C. Gros, M. Tardugno, D. Labella, C. Florean, S. Minden, H. Hashimoto, Y. Q. Chan, X. Zhang, G. Kirsch, E. Novellino, P. B. Arimondo, E. Miele, E. Ferretti, A. Gulino, M. Diederich, X. D. Cheng and A. Mai, *J. Med. Chem.*, 2014, **57**, 701-713.

42. J. Husby, G. Bottegoni, I. Kufareva, R. Abagyan and A. Cavalli, *J. Chem Inf. Model.*, 2015, **55**, 1062-1076.

43. A. Kabro, H. Lachance, I. Marcoux-Archambault, V. Perrier, V. Dore, C. Gros, V. Masson, J. M. Gregoire, F. Ausseil, D. Cheishvili, N. B. Laulan, Y. St-Pierre, M. Szyf, P. B. Arimondo and A. Gagnon, *MedChemComm*, 2013, **4**, 1562-1570.

44. S. Castellano, D. Kuck, M. Viviano, J. Yoo, F. López-Vallejo, P. Conti, L. Tamborini, A. Pinto, J. L. Medina-Franco and G. Sbardella, *J. Med. Chem.*, 2011, **54**, 7663-7677.

45. J. Yoo, S. Choi and J. L. Medina-Franco, *PLoS One*, 2013, **8**, e62152.

46. W. Maldonado-Rojas, J. Olivero-Verbel and Y. Marrero-Ponce, *J. Mol. Graphics Modell.*, 2015, **60**, 43-54.

TABLES

Table 1. Number and proportion of pairs of compounds into the four different regions of the global and two local SAS maps.

Quadrant	Region	Entire dataset[a]	SAM analogues (cluster A)[b]	Non-nucleosides (cluster B)[c]
I	Uncertainty	1571 (4.02%)	2 (0.20%)	1066 (3.88%)
II	Similarity cliff (scaffold hop)	36177 (92.61%)	46 (4.64%)	26,059 (94.78%)
III	Smooth SAR	939 (2.40%)	614 (62.02%)	325 (1.18%)
IVa	Deep activity cliffs	64 (0.16%)	49 (4.95%)	15 (0.05%)
IVb	Shallow activity cliffs	309 (0.79%)	279 (28.18%)	30 (0.11%)
Total		39,060 (100%)	990 (100%)	27,495 (100%)

[a]280 compounds; [b]45 compounds in cluster A of figure 2: [c]235 compounds in cluster B of figure 2.

Table 2. Shallow activity cliffs formed by non-nucleoside compounds that are not SGI-1027 regioisomers.

Compound pair	Activity of most active compound in the pair (pIC_{50})	ΔpIC_{50}	ECFP/Tanimoto
CHEMBL115145, CHEMBL1503050	5.17	1.14	0.28
CHEMBL1302528, CHEMBL1377441	5.34	1.34	0.3
CHEMBL1302528, CHEMBL1443718	5.17	1.17	0.3
CHEMBL1302528, CHEMBL1558192	5.37	1.37	0.28
CHEMBL1302528, CHEMBL256098	5.04	1.03	0.3
CHEMBL1303509, CHEMBL1332402	5.99	1.04	0.27
CHEMBL1328733, CHEMBL1332506	6.09	1.1	0.27
CHEMBL1328733, CHEMBL1411673	6.09	1.18	0.37
CHEMBL1379120, CHEMBL592316	5.91	1.9	0.28
CHEMBL1403497, CHEMBL2063048	5.8	1.14	0.28
CHEMBL1564869, CHEMBL3109084	4.7	1.29	0.39
CHEMBL1607517, CHEMBL1704614	5.87	1.49	0.36
CHEMBL1607517, CHEMBL1988862	5.99	1.6	0.46
CHEMBL1916517, CHEMBL1916672	3.82	1.03	0.55
CHEMBL1978925, CHEMBL1990599	5.27	1.26	0.32
CHEMBL1983083, CHEMBL1990599	5.07	1.07	0.38

Table 3. Activity cliffs formed by CHEMBL3126646 (*meta/meta* SGI-1027 regioisomer)

Compound	ΔpIC_{50}	ECFP/Tanimoto
CHEMBL3126647	3.16	0.75
Ortho/ortho SGI-1027 regioisomer	3.16	0.53
CHEMBL3126654	3.16	0.38
CHEMBL3126649	2.51	0.69
CHEMBL3126644	2.46	0.60
CHEMBL3126653	2.37	0.42
CHEMBL3126648	2.01	0.57

FIGURES

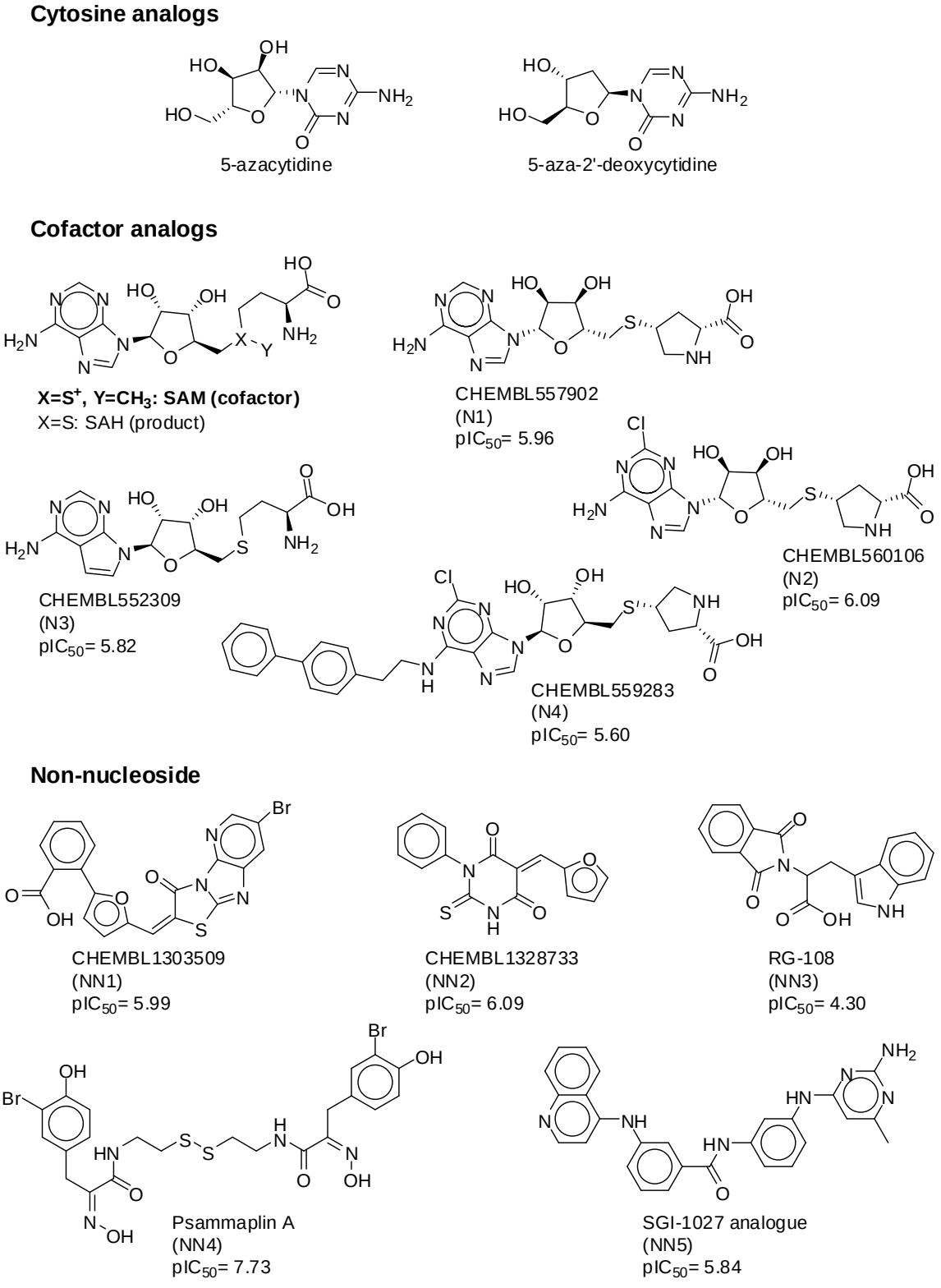

Figure 1. Structures of representative IDNMT1. The relative position in chemical space of selected compounds: 4 SAM analogues (N) and 5 non-nucleoside (NN) compounds is shown in Figure 2.

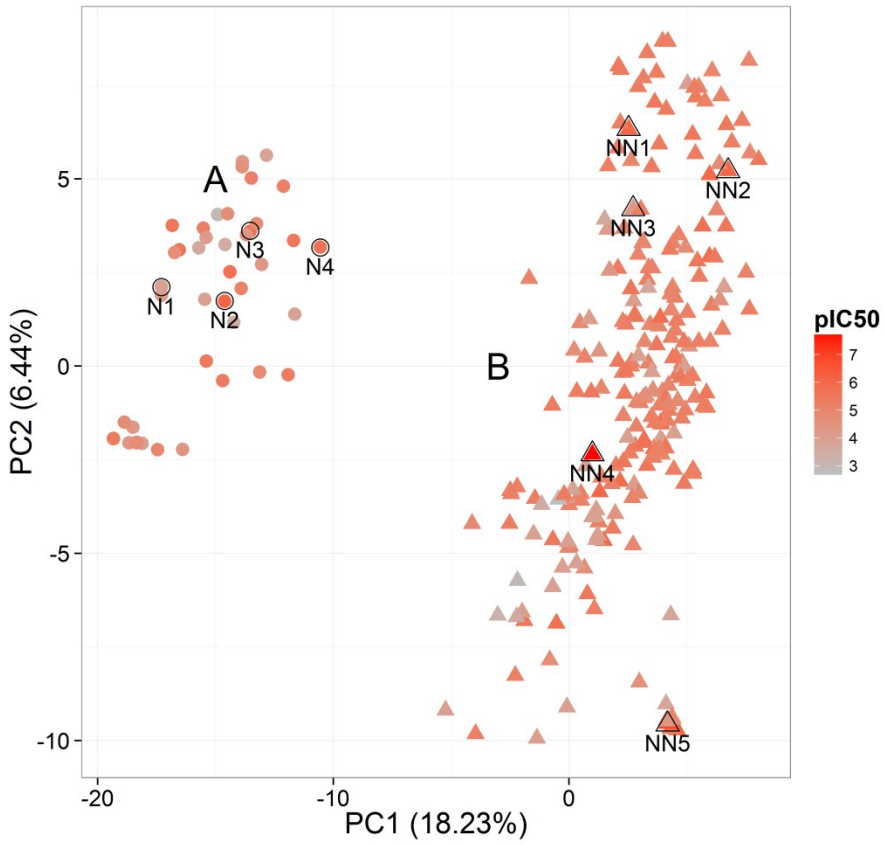

Figure 2. Visual representation of the chemical space of the 280 compounds in the data set. The visualization was obtained by principal component analysis of the similarity matrices computed with ECFP. The percentage of variance explained by each PC is indicated in the corresponding axis. Data points are colored by the pIC_{50} values in a continuous scale. Two main clusters (A: circles, B: triangles) are readily distinguished. Nine selected compounds are identified as SAM analogues (N) and non-nuclosides (NN) compounds. The chemical structures are shown in Figure 1.

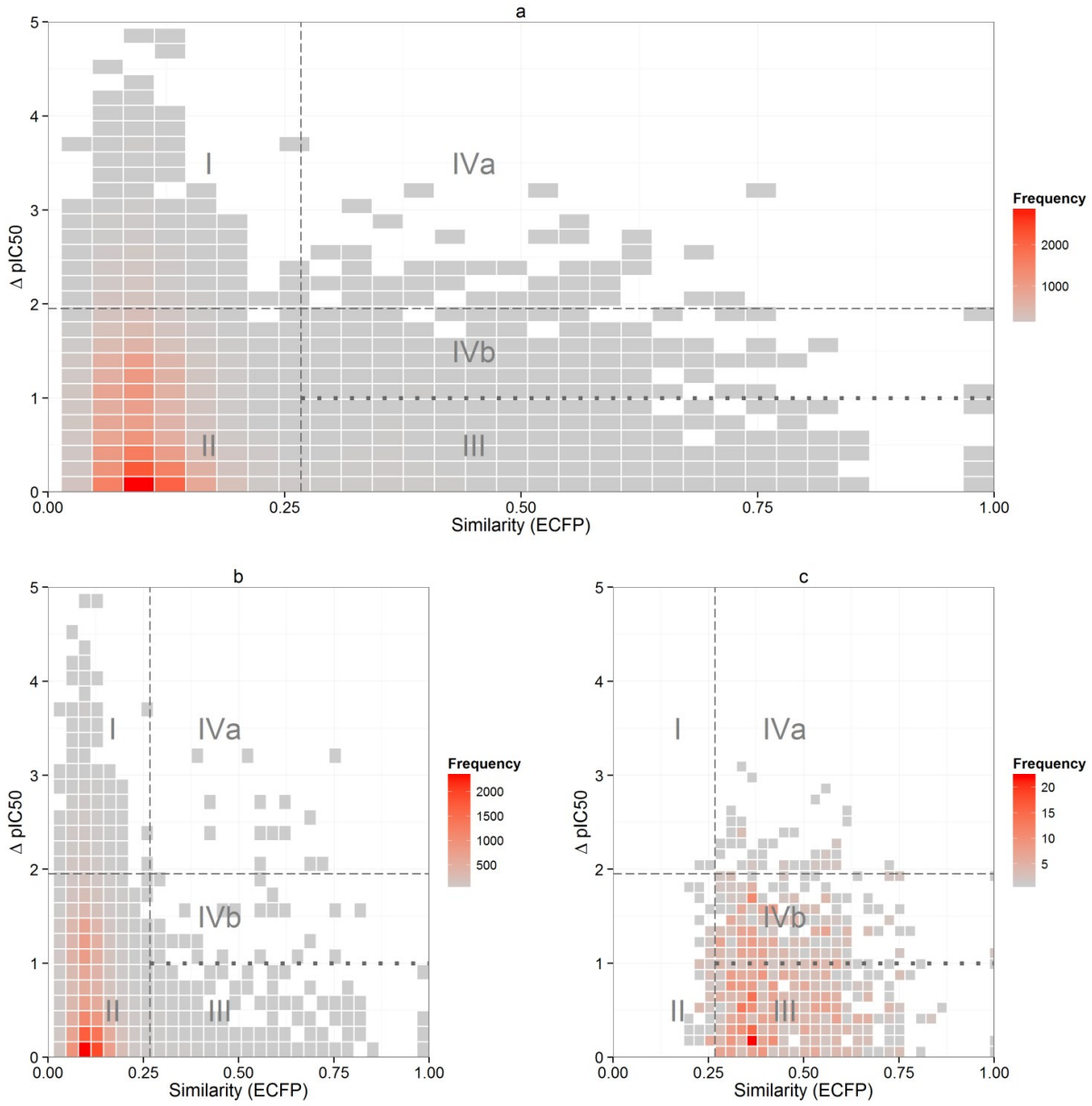

Figure 3. Density SAS maps of the global and local activity landscapes. The 2D plots are colored by the frequency of data points in the coordinates given. Dashed lines divide the maps into the four quadrants labeled I-IV. The dotted line further divides the activity cliff quadrant (IV) in two regions (IVa and IVb) to distinguish shallow and deep cliffs (see text for details). (a) IDNMT1 SAS map for the entire set with 280 compounds (39,060 pairwise comparisons). (b) SAS map for 235 non-nucleoside compounds (cluster B in Figure 2) (27,495 pairwise comparisons). c) SAS map for 45 SAM analogues (cluster A in Figure 2) (990 pairwise comparisons).

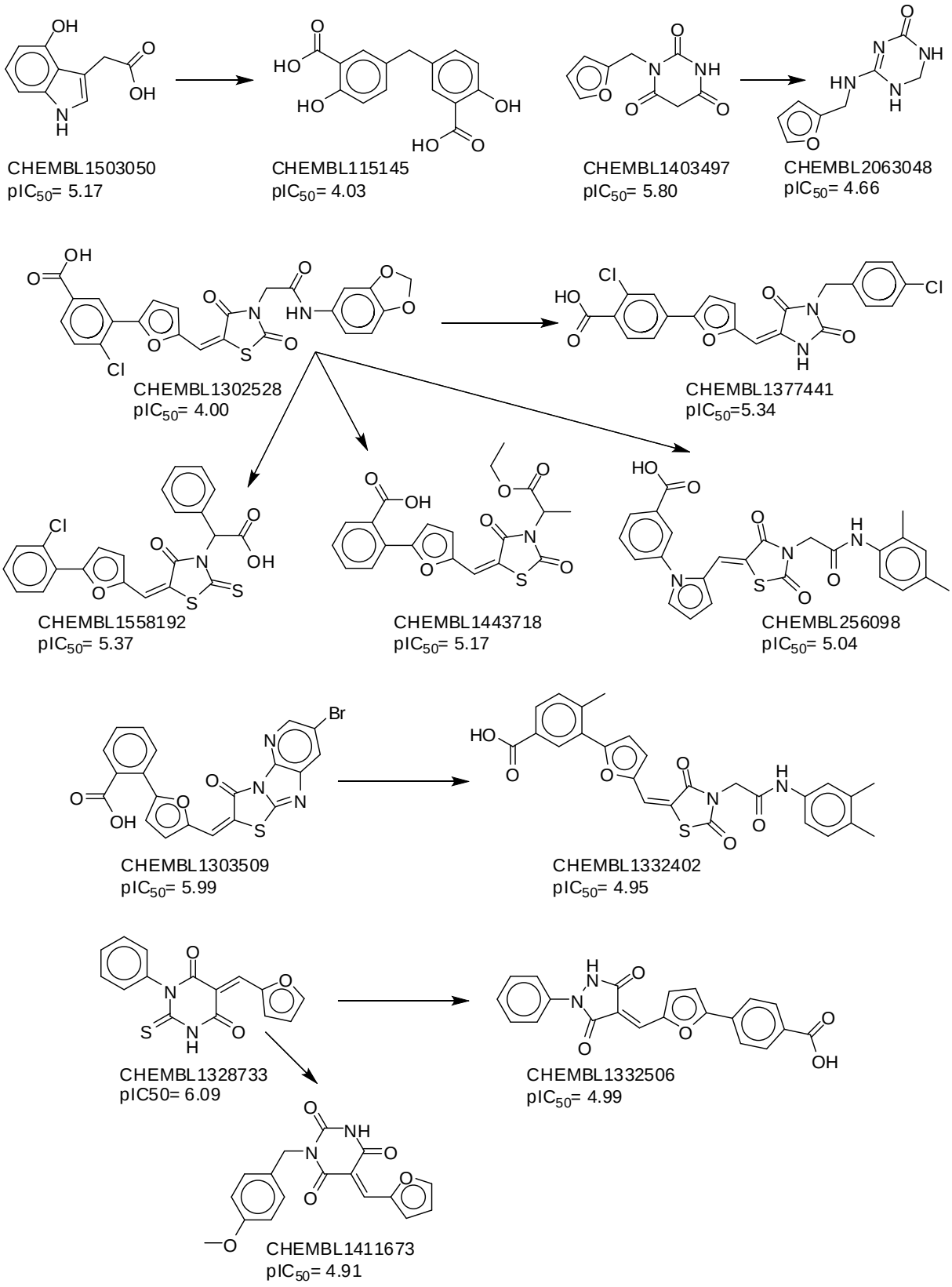

Figure 4. Structures of activity cliffs of non-nucleoside compounds identified by high-throughput screening. Table 2 summarizes the potency difference and structure similarity for each compound pair associated with an arrow.

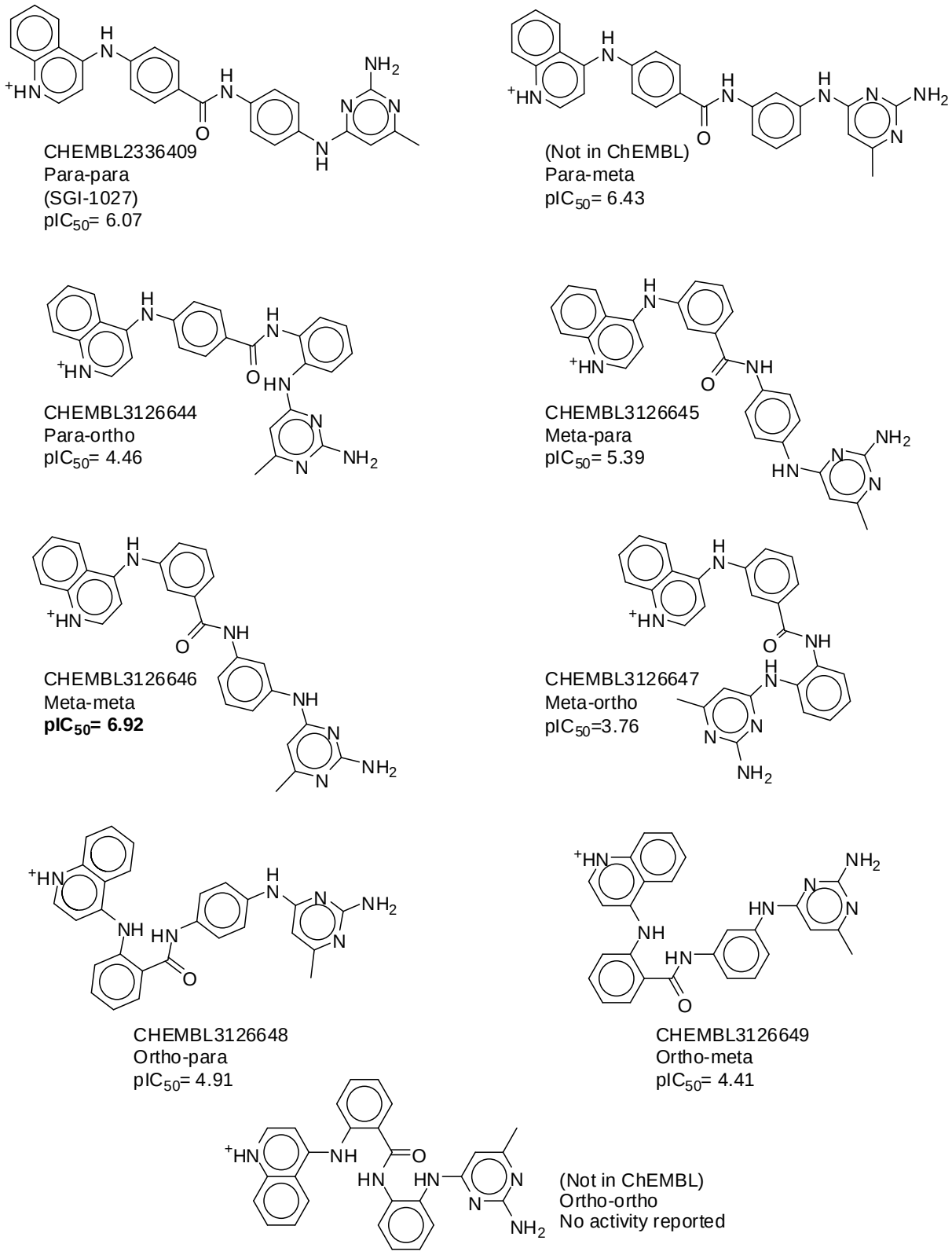

Figure 5. Chemical structures of non-nucleoside activity cliffs related to regioisomers of SGI-1027. Table 3 summarizes the potency difference and structure similarity for each compound in this figure and the lead molecule CHEMBL3126646.

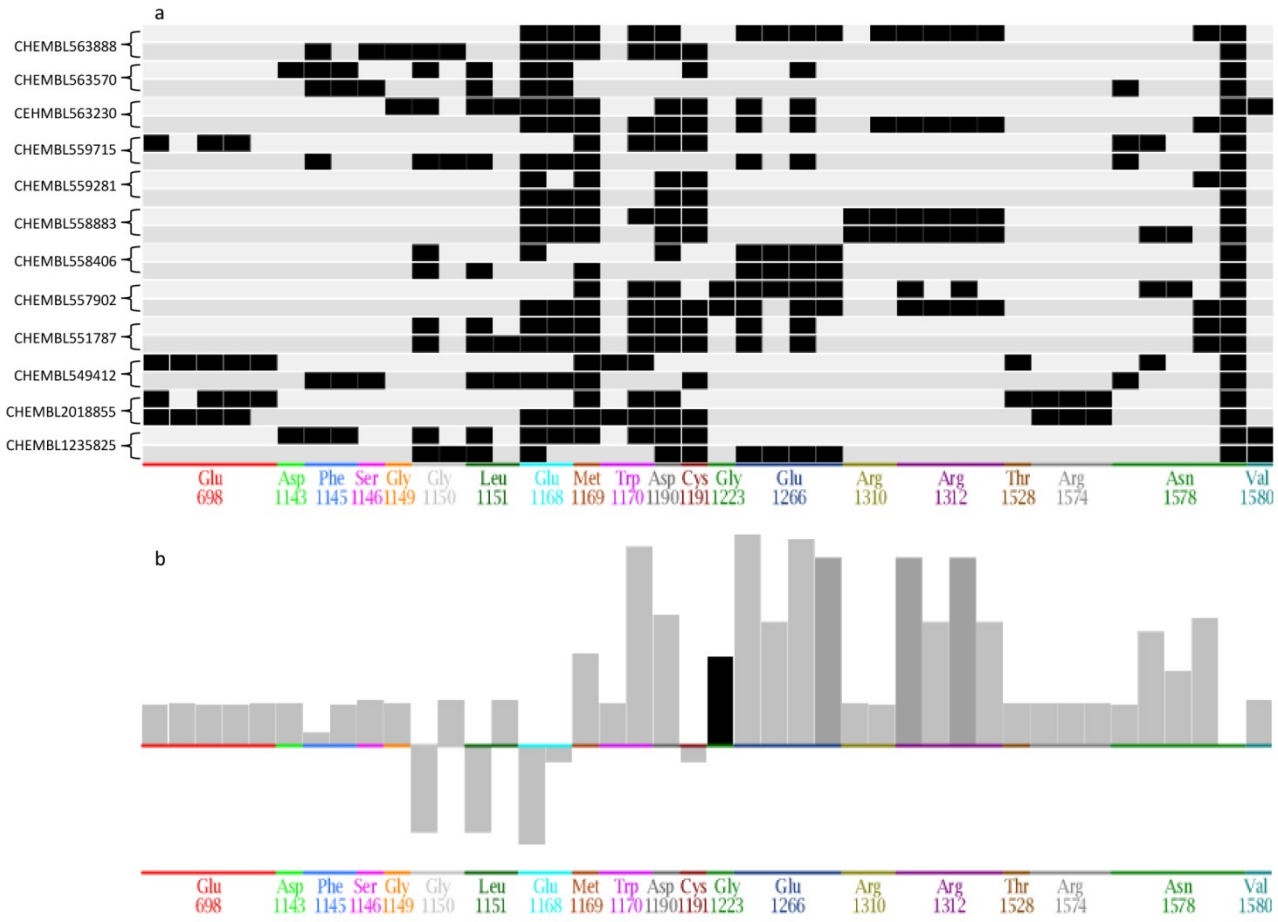

Figure 6. Summary of protein-ligand interaction fingerprint (PLIFs) analysis of the activity cliff generator CHEMBL552309 (compound N3 in Figure1) and 11 SAM-analogues that form activity cliffs with this compound (the chemical structures of the 11 molecules are shown in Figure S8 of the Supporting Information). For each compound the best two docking poses are represented. (a) Data matrix summarizing the protein-ligand contacts between the best two poses of 12 docked molecules and DNMT1. In this matrix, the rows represent the docked poses. The columns are the fingerprint bits indicating the amino acid residues that make at least one contact with one of the compounds. A black cell in the matrix indicates that a contact is present between the intersecting compound and amino acid residue *i.e.*, fingerprint bit turned 'on'. In contrast, a white cell means that there is no contact *i.e.*, fingerprint bit turned 'off'. (b) The statistically more significant PLIFs. A darker color means that the interaction is more associated to the active compound.

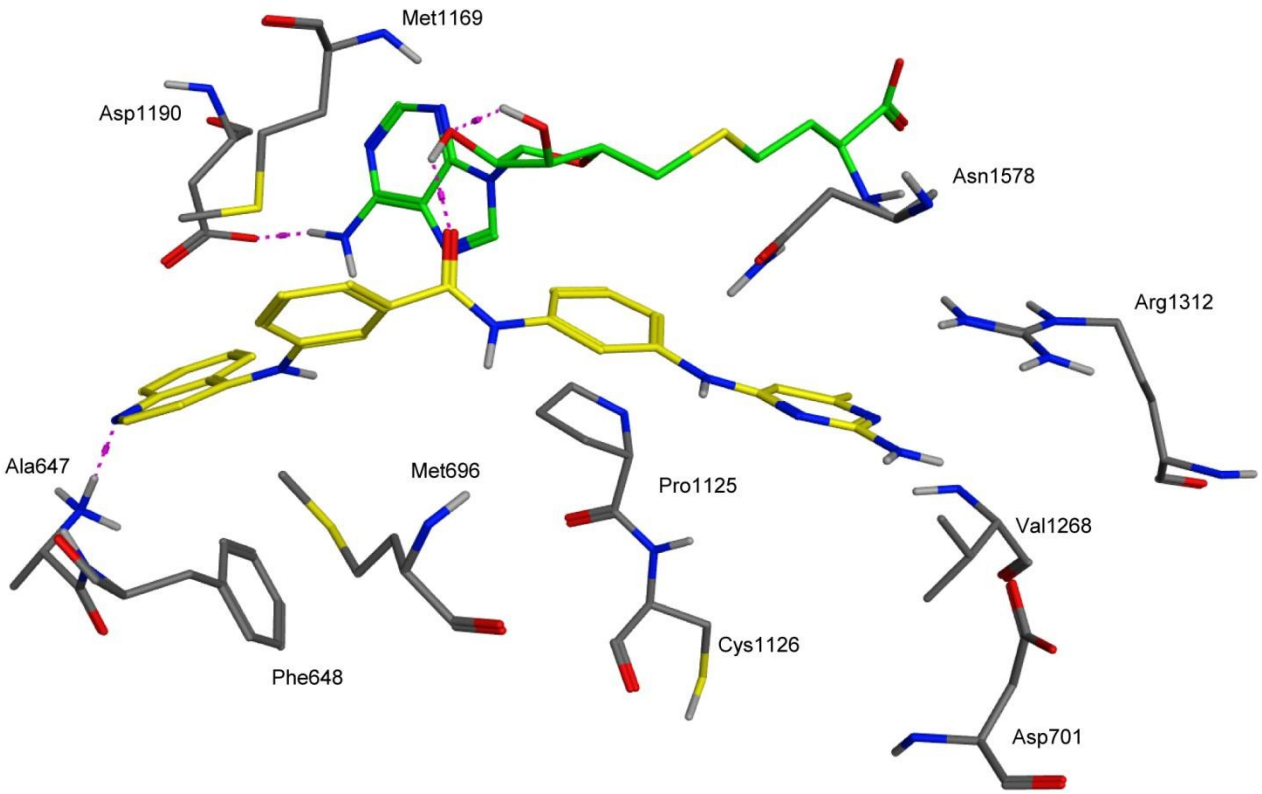

Figure 7. Docking model of CHEMBL3126646 (carbon atoms in yellow) with DNMT1. The position of the co-crystal SAH is displayed (carbon atoms in green). Selected residues of the binding pocket are shown. Hydrogen bond interactions are in dashed lines. Note the predicted hydrogen bond interaction between the carbonyl oxygen of CHEMBL3126646 and the O2' oxygen of SAH. Non-polar hydrogens are hidden for clarity.

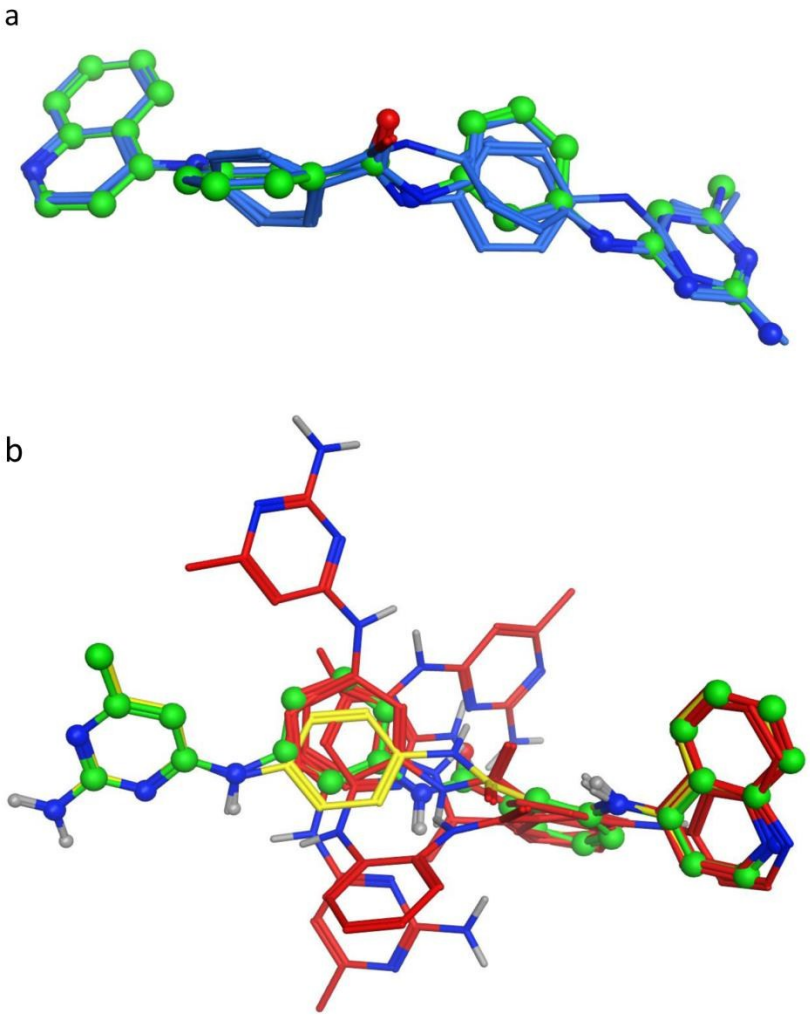

Figure 8. Flexible alignment of regioisomers of SGI-1027 (chemical structures are shown in Figure 5) with the best docked pose of CHEMBL3126646 (balls and sticks and carbon atoms in green). (a) Non-*ortho* regioisomers (carbon atoms in blue). (b) *Ortho* regioisomers [CHEMBL3126644, CHEMBL3126647, CHEMBL3126649 and *ortho/ortho* SGI-1027 analog (not registered in ChEMBL) (carbon atoms in red) and CHEMBL3126648 (*ortho/para*) (carbon atoms in yellow). Note the alignment different in the red molecules and the different orientation of the carbonyl oxygen in the both the red and the yellow molecules, which is not the case in (a).

For Tables of Contents

Inhibitors of DNA methyltransferases have distinct structure-activity relationships as revealed by the activity landscape sweeping study discussed in this work.