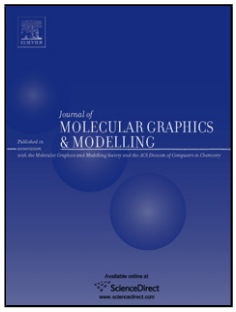


Title: COMPUTATIONAL FISHING OF NEW DNA METHYLTRANSFERASE INHIBITORSFROM NATURAL PRODUCTS

Author: Wilson Maldonado-Rojas Jesus Olivero-Verbel Yovani Marrero-Ponce

PII: S1093-3263(15)00076-5
DOI: http://dx.doi.org/doi:10.1016/j.jmgm.2015.04.010
Reference: JMG 6538

To appear in: *Journal of Molecular Graphics and Modelling*

Received date: 2-2-2015
Revised date: 28-3-2015
Accepted date: 22-4-2015

Please cite this article as: W. Maldonado-Rojas, J. Olivero-Verbel, Y. Marrero-Ponce, COMPUTATIONAL FISHING OF NEW DNA METHYLTRANSFERASE INHIBITORSFROM NATURAL PRODUCTS, *Journal of Molecular Graphics and Modelling* (2015), http://dx.doi.org/10.1016/j.jmgm.2015.04.010

Ms. Ref. No.: JMGM-D-15-00047

Title:

COMPUTATIONAL FISHING OF NEW DNA METHYLTRANSFERASE INHIBITORS FROM NATURAL PRODUCTS

Authors:

Wilson Maldonado-Rojas, Jesus Olivero-Verbel* and Yovani Marrero-Ponce [a]

[a] *W. Maldonado-Rojas, J. Olivero-Verbel and Y. Marrero-Ponce*
Environmental and Computational Chemistry Group. School of Pharmaceutical Sciences, Zaragocilla Campus, University of Cartagena, 130014, Cartagena, Colombia.

*. **Corresponding author:** joliverov@unicartagena.edu.co

Highlights

- A LDA-based QSAR model was built to identify new DNMT inhibitors.
- Virtual screening was performed on 800 NPs from NatProd Collection data base.
- AutoDock Vina and Surflex-Dock were used for molecular docking on DNMTs structures.
- Contact patterns and molecular diversity analysis were used to prioritize DNMTis.
- Six consensus NPs were identified as potential DNMTis.

ABSTRACT

DNA methyltransferase inhibitors (DNMTis) have become an alternative for cancer therapies. However, only two DNMTis have been approved as anticancer drugs, although with some restrictions. Natural products (NPs) are a promising source of drugs. In order to find NPs with novel chemotypes as DNMTis, 47 compounds with known activity against these enzymes were used to build a LDA-based QSAR model for active/inactive molecules (93% accuracy) based on molecular descriptors. This classifier was employed to identify potential DNMTis on 800 NPs from NatProd Collection. 447 selected compounds were docked on two human DNA methyltransferase (DNMTs) structures (PDB codes: 3SWR and 2QRV) using AutoDock Vina and Surflex-Dock, prioritizing according to their score values, contact patterns at 4 Å and molecular diversity. Six consensus NPs were identified as virtual hits against DNMTs, including 9,10-dihydro-12-hydroxygambogic, phloridzin, 2',4'-dihydroxychalcone 4'-glucoside, daunorubicin, pyrromycin and centaurein. This method is an innovative computational strategy for identifying DNMTis, useful in the identification of potent and selective anticancer drugs.

Keywords: Anticancer, screening, QSAR, Docking, Cluster.

1 INTRODUCTION

DNA methylation is a covalent biochemical modification defined as an epigenetic change important in the regulation of gene expression [1, 2]. The progression of DNA methylation involves a cycle of demethylation, *de novo* methylation, and methylation maintenance, catalyzed by family enzymes known as DNA methyltransferases (DNMTs, EC#: 2.1.1.37) [3-5]. These are responsible of transferring a methyl group from S-adenosyl-L-methionine (SAM) to the carbon-5 position of cytosine in DNA. This mechanism has been proposed for several authors (**Fig. 1**) [6-9] . Currently, three types of cytosine-5 DNMTs have been identified, including two *de novo* DNA methyltransferases; DNMT3A and DNMT3B, which establish the methylation patterns during embryonic development in mammals and in differentiated cells [10, 11]; and DNMT1, the most abundant and active of these enzymes responsible for copying the methylation pattern of DNA during cell division [12, 13]. It has been shown that DNMT1 plays an important role in carcinogenesis [12, 14], therefore, these targets are of particular interest to search for specific inhibitors [15-17].

Figure 1 may go here

The inhibition of DNMTs activity has been presented as a possible pathway to reactivate genes silenced by methylation of their promoters in different diseases, including cancer [1, 12, 18]. The relationship between the hypermethylation of promoter of tumor suppressor genes and cancer development has been clearly demonstrated [3, 4, 19], suggesting DNMTs as promising drug targets for the discovery of new and more potent/selective anticancer drugs [20]. To date, several DNMT inhibitors (DNMTis) of different structural classes have been published; basically categorized as

nucleoside DNMTis and non-nucleoside analogue DNMTis, **Fig. 2** [16, 21]. DNMTis nucleosides, cytosine analogs, have been studied in several cancer types [22, 23]. Most of these compounds that inhibit the activity of DNMT have been related with significant ¨off target¨ effects [24], a fact that has prevented their use with pharmacological purposes. However, two of them, azacitidine and decitabine, have been approved by FDA (Food and Drug Administration) for the treatment of myelodysplastic syndromes [25, 26]. Moreover, non-nucleoside DNMTis, such as RG-108 [27], which was identified via virtual screening methods; and SGI-1027, a quinoline derivative, have been proposed as DNMTis [28, 29]. However, the weak inhibitory activity of these compounds [30] indicates a need for the search of more effective inhibitors in the future.

Natural products (NPs) are a promissory source of drugs due to their molecular diversity and low toxicity. It is estimated that 60% of approved drugs are derived from natural sources [31-33]. To date, several NPs have been proposed as DNMT inhibitors [34, 35]. Some studies have shown epigallocathechin-3-gallate (EGCG) [36], curcumin [34], genistein [37], psammaplin A [38], mahanine [39], caffeic acid [40], laccaic acid [41], among others. However, they exhibit no significant inhibitory pharmacological activity.

Figure 2 may go here

Computational tools help to elucidate the basic structural requirements of the inhibitors for activity against DNMT. The increase in the number of available crystallographic DNMTs structures has prompted the use of computational molecular docking and structure-based approaches for search DNMTis [42-44]. Computational and experimental screening methodologies have served to identify promissory DNMTis, for instance NSC 14778. Recently, two new DNMTis (SW155246 and SID49645275) have been discovered using experimental high-throughput screening (HTS) and they have been proposed as molecular scaffold for discovering new DNMTis [43, 45]. However, they showed a lack of pharmacological information, specific of DNMTs. For this reason, the systematic searching of new promissory compounds with DNA hypomethylating activity by using NPs data sets is a great path and an urgent need to discover new and more effective DNMTis. Recently, Medina-Franco et al, shown that NPs are a rich source for demethylating agents that need to be rigorously characterized in theoretical and/or experimental studies [31].

The main aim of this report was to develop a virtual screening to find potential DNMTis on 800 NPs from NatProd Collection, MicroSource Discovery Systems

(www.msdiscovery.com/natprod.html). The protocol included a Linear Discriminant Analysis (LDA)-based QSAR model from a set of 47 active/inactive molecules against DNMTs, with further optimization of the promising compounds by molecular docking procedures, and finally, a selection based on molecular diversity by *k*-means cluster analysis. Following this multistep approach, six consensus NPs hits with new chemotypes as inhibitors of DNMTs were identified.

2 Materials and methods

2.1 LDA-based Discriminant Model. A set of 47 diverse compounds reported in the literature as active and not active against DNMTs were used to develop a discriminant model, employing 32 compounds as training set and the other 15 as test set as shown in Table 1 (Chemical structure of these compounds are shown in **Fig. S1** and **Fig. S2**. Molecules in each group were randomly selected. Compounds were downloaded from PubChem database (http://pubchem.ncbi.nlm.nih.gov/), and their 3D structures were optimized using molecular mechanic methods, with Tripos force field with Gasteiger charges, a gradient convergence of 0.05 kcal/mol and a maximum number of optimization iterations of 1000. All calculations were performed using SYBYL-X 2.0 package [46].

Table 1 may go here

A LDA–based QSAR model was developed for classification of compounds as active or inactive against DNMTs. A whole set of 0-3D MDs were calculated employing DRAGON 5.5 program [47]. LDA–based QSAR model was developed using forward stepwise procedure as feature selection method (wrapper strategy) in STATISTICA 8.0 software [48], and the principle of parsimony (Occam's Razor) was established selecting the best model. Statistical parameters such as, Wilks' λ (U-statistic), square Mahalanobis distance (D^2), Fisher ratio (F) with the corresponding *p*-level (*p*(F)), as well as the percentage of good classification (accuracy) in both training and test sets were used to assess the quality and performance of the model [49].

In order to increase the hit chance, the classification probabilities produced by LDA-based QSAR model were used for subsequent selection of compounds. A NP is classified as active if ΔP% > 50% or as inactive if ΔP% the otherwise, being ΔP% = [P (Active) − P (Inactive)] × 100, P (Active) and P (Inactive) are the probabilities to be active or inactive according to the classifier model.

In addition to the greedy-based wrapper method, we initially employed several unsupervised and supervised methods based on the computation of Shannon's entropy type measures for make a

drastic dimension reduction of all DRAGON MDs [50, 51]. This algorithm has been implemented in an *in house* computer program denominated IMMAN (acronym for Information Theory based CheMoMetricANalysis) [52]. This approach has been employed to compare proposed MDs and various families of indices reported in the literature, as well as MD computing software [53].

Six different types of MDs from different families were finally selected in the model, including: information index (SIC2), 3D-moRSE descriptor (Mor13m), 2D autocorrelation (GATS5m), Randić molecular profile (SHP2), topological descriptor (ZM2V), and atomic centered-fragment (H-047). Definitions for these descriptors can be seen in Table S1.

2.2 Molecular Docking with AutoDockVina and Surflex-Dock. The feasibility of NPs to work as inhibitors of DNMTs was evaluated with two docking programs frequently used for virtual screening, AutoDock Vina [54-56] and Surflex-Dock [57]; which are based on different score functions to evaluate the binding mode of a ligand onto a receptor.

2.2.1 Preparation and selection of crystallographic structures of DNMTs for molecular docking with NPs. Ten available mammalian DMNT structures (DMT1, PDB codes: 3AV4, 3AV5, 3AV6, 3EPZ, 3PT6, 3PT9, 3PTA, 3SWR, 4DA4 and DNMT3A, PDB code: 2QRV) were downloaded from Protein Data Bank (http://www.rcsb.org/pdb/home/) and prepared with SYBYL-X 2.0 package for molecular docking. This process consisted of removing water molecules and other ligands, with subsequent repairing and fixation of amide in side chains. Optimization protocols were performed using Powell conjugate gradient method; with dielectric constant value of 1.0, gradient convergence fixed to 0.005 kcal mol^{-1}, maximum number of iterations at 1000, and Kollman United/All-atoms force fields with AMBER charges.

A comparative structural analysis for DMNTs was performed with SYBYL-X 2.0 by multiple sequence alignments (% of identity, %ID) and spatial similarity (RMSD in Å) [58]. Details are shown in Table S2, S3 and S4. Based on these results, two representative DNMTs crystallographic structures were selected for protein-ligand docking protocols: DNMT1 (3SWR) and DNMT3A (2QRV).

2.2.2 Docking calculations of NPs on DNMTs. Docking calculations with AutoDock Vina program for NPs on selected DNMT structures (3SWR and 2QRV) were defined by establishing a

cube at the geometric center of the co-crystalized ligand present on each selected DNMTs (3SWR-sinefungin and 2QRV-SAH), with dimensions 30 × 30 × 30 Å, covering the catalytic site and cofactor binding site (SAM) in the enzyme, employing a grid point spacing of 1.0 Å. The x, y, and z coordinates for the center grid boxes on 3SWR were -4.93, -2.66 and 33.36, whereas for 2QRV were 106.61, 45.73 and -1.71, respectively. Three docking runs were performed for each ligand, and the pose with the highest absolute value of affinity (kcal/mol) was saved. Finally, the mean affinity value for best poses was taken as the value of the binding affinity for a particular complex [54, 59].

The NPs that presented the best affinities with AutoDock Vina were additionally evaluated with Surflex-Dock, following a protomol generation approach based on the proximal residues to the co-crystalized ligand present in the crystallographic of selected DNMT structures using a threshold of 0.5 Å and bloat of 0, with the others parameters set as default. All docking calculations for ligands were performed keeping the same parameters (20 poses each), using only the highest ranked pose by Surflex-Dock [57, 60].

In order to validate the results obtained with AutoDock Vina and Surflex-Dock, the experimental ligands, sinefungin and SAH, were docked to the crystal structure of human DNMT1 (3SWR) and human DNMT3A (2QRV) [45], respectively. In total, 100 docking runs were performed with AutoDock Vina and Surflex-Dock for each ligand, using the same parameters described above for docking of NPs on DNMTs.

2.2.3 Identification of main interaction of selected NPs on DNMTs binding site. Identification of interacting residues at distances less than to 4 Å was carried out using Pymol program [61] in order to identify those most likely involved in the inhibition mechanism of DNMTs [62]. The interaction nature of the residues that interact with selected NPs on DNMTs binding site was performed using LigandScout 3.1 program [63, 64], which utilizes simplified pharmacophores to detect the number and type of existing ligand-residue interactions on the protein active site for a particular protein-ligand complex. The most important DNMTs-NPs complexes were visualized with PyMol program [61] depicting the binding mode for selected NPs.

2.2.4 Molecular docking validation using biological information for DNMT1 Inhibitors. Compared to DNMT3A and DNMT3B, DNMT1 is most targeted to search for inhibitors as its high enzyme activity is the greatest among the DNMTs. This has increased the availability of biological

information regarding this protein. In order to validate the reliability of the results generated by docking protocols, 3D structures and biological data for forty-five DNMTis were obtained from PubChem chemical library (http://pubchem.ncbi.nlm.nih.gov/), and their binding affinities and total scores were calculated with AutoDock Vina and Surflex-Dock following the same parameters described for NPs on DNMT1 (PDB code: 3SWR). The biological information consisted of the half maximal inhibitory concentration (μM, IC50) values for DNMT1 activity, as reported for these chemicals in PubChem BioAssay (AID: 602386, Dose response confirmation of DNMT1 inhibitors in a Fluorescent Molecular Beacon assay). Correlation analysis to establish the relationships between binding scores on DNMT1, generated by AutoDock Vina and Surflex-Dock, and experimental biological data (LogIC50) [56, 60] for DNMTis were calculated using STATISTICA 8.0 software [48].

2.3 Selection of Promissory NPs as DNMTis by Cluster Analysis. A molecular diversity *k*-means cluster analysis for selected NPs and well-known inhibitors, active and inactive include in training and test sets, was made based on 28 MDs (including the six calculated for screening) with STATISTICA 8.0 software [48]. The number of members in each cluster and the standard deviation of the variables in the cluster (kept as low as possible) were taken into account, to have an acceptable statistical quality of data partitions into the clusters. The values of the standard deviation between and within clusters, the respective Fisher ratio and their *p*-level of significance, were also examined. In the final selection of molecular descriptors vector the unsupervised filters in IMMAN software [53, 65] and the Fisher ratio in cluster analysis were taken into consideration. The final selected MDs included molecular weight, sum of the electronegativities (Se), sum of the polarizabilities (Sp), hydrogen (nH), carbon (nC), oxygen (nO), ALOGP, Neoplastic-80, among others, have been useful in the identification of anti-cancer drugs [66-68].

2.4 Searching of Information for Selected NPs (Assertions from Literature). The selection of promissory NPs was performed utilizing, as a last filter, their available biological information. That is, an exhaustive searching of biological data reported for promissory NPs was performed using several search engines including; GoPubMed (http://www.gopubmed.com), PubReMiner (http://bioinfo.amc.uva.nl/human-genetics/pubreminer/), PubGraph (http://datamining.cs.ucla.edu/pubgraph/), among others.

2.5 Virtual Screening of NPs from NatProd Collection.

The searching of new NPs as DNMTis was conducted in three stages: i) selection of actives NPs against DNMTs by LDA-based QSAR model (**Eq. 1**, see below) from 800 NPs contained in NatProd Collection (www.msdiscovery.com/natprod.html). These compounds were optimized under the same parameters utilized for the 47 compounds included in the training and test sets. Subsequently, only six MDs required (SIC2, Mor13m, GATS5m, SHP2, ZM2V, H-047) were calculated for these NPs using DRAGON 5.5 program. These MDs calculated for 800 NPs were evaluated by the classifier; ii) Molecular docking using AutoDockVina and Surflex-Dock protocols for NPs classified as active by the LDA-based QSAR model on two DNMT structures (3SWR and 2QRV); iii) cluster analysis was carried out for selected NPs taking as reference training and test sets compounds (both actives and inactives) and searching for biological information from literature. Finally, six NPs were selected with new scaffolds as virtual DNMTi hits. That ensemble protocol is shown in Scheme1.

Scheme 1 may go here

3 RESULTS AND DISCUSSION

3.1 LDA-based QSAR Model for identification of DNMTis.

In order to perform a fast and reliable identification of NPs as DNMTis, a QSAR model was built (Eq. **1**) utilizing six DRAGON's molecular descriptors (MDs) [47]. QSAR model with LDA-statistical parameters are show below:

$$\textit{Class} = 56.44\times\mathbf{SIC2} + -9.66\times\mathbf{Mor13m} + 10.94\times\mathbf{GATS5m} + 130.55\times\mathbf{SHP2} + 0.05\times\mathbf{ZM2V} + 0.67\times\mathbf{H\text{-}047} - 135.41 \quad (\mathbf{1})$$

$$N = 47,\ \text{Wilks'}\ \lambda = 0.204,\ D^2 = 15.17,\ F\ (6,25) = 16.262,\ p < 0.00001$$

The accuracy values (Q_{total}) of **Eq. 1** for the training set was 96.9%, with a sensitivity and specificity of 94.7% and 100%, respectively, and the Matthews correlation coefficient (***C***) value obtained was 0.94 (see Table 2). In addition to this, similar values were obtained for test set, carrying out statistical parameters for accuracy of 93.3%, sensitivity of 88.9 %, specificity of 100%, and ***C*** value of 0.87. Moreover, both training and test sets presented a false positive rate of 0.0 %. Finally, statistical parameters obtained for this classifier are comparable with LDA-based QSAR computational studies focused in drug discovery [49, 69, 70], showing the reliability of the QSAR to the virtual discovery of active compounds against DNMTs. Details of these results can be seen in Tables S5 and S6.

Table 2 may go here

3.1.1 Application of LDA-based QSAR Model for Discovery of NPs as DNMTis.

The NatProd Collection database comprises a diverse group of NPs constituted in 75% by alkaloids, flavanoids, sterols, triterpenes, diterpenes, sesquiterpenes, benzophenones, chalcones, stilbenes, limonoids, quassinoids, chromones, coumarins; and the remainer 25% encompasses quinones, quinonemethides, benzofurans, benzopyrans, rotenoids, xanthones, carbohydrates, benztropolones, depsides, and depsidones (Table S7). Virtual screening studies have also been conducted using this database to search for NPs as inhibitors of proteins related to inflammatory processes, ulcerative colitis and cancer, such as Interleukin 6 (IL-6)[71] and mitogen-activated protein kinase Phosphatase-1, demonstrating its applicability in natural-based drugs discovery.

LDA-based QSAR model was performed for discriminating of active/inactive molecules as DNMTis from 800 NPs from NatProd Collection previously optimized by molecular mechanics. In total, 447 compounds contained in the NPs collection were classified as active against DNMTs with ΔP% greater than 50% with potential activity (ΔP% values for NPs can be seen in Table S8.

3.2 Searching of Promissory NPs as DNMTis by Molecular Docking Methods.

Docking protocols validation by re-docking of protein with co-crystalized ligand.

Molecular docking validation performed by re-docked of sinefungin and SAH on their complex with DNMTs (3SWR and 2QRV, respectively), revealed an optimal reproduction of the

predicted poses compared with experimental binding mode for these ligands (co-crystalized ligand), with satisfactory results with both AutDock Vina and Surflex-Dock protocols.

For illustrative purposes, the best poses obtained for each DNMT-ligand complexes with utilized molecular docking protocols are shown in **Fig. 3**, showing the best binding pose obtained according to experimental co-crystallized ligand. RMSD values, binding affinity (AutoDock Vina) and Total score (Surflex-Dock) are reported in Table S9.

Figure 3 may go here

The best conformation of pose replication expressed as the root mean standard deviation (RMSD) was 0.4997 Å. This value was obtained with AutoDock Vina for the 2QRV-SAH complex, where the RMSD value for 3SWR-sinefungin was 0.8756 Å. In the case of Surflex-Dock, the best pose for sinefungin on 3SWR binding site presented a RMSD value of 1.9170 Å, and SAH presented a RMSD value of 1.1934 Å on 2QRV binding site. Obtained results were in agreement with similar studies conducted for this purpose. Medina-Franco et al. 2014, reported a RMSD value for sinefungin on 3SWR binding site of 0.547 Å using Glide XP program [9, 45]. These results showed the capability of using molecular docking protocols to reproduce the co-crystalized conformation of sinefungin and SAH on their experimental complexes with DNMTs.

3.2.1 Molecular docking results using AutoDock Vina and Surflex-Dock. Molecular docking of NPs on human DNMT1 and DNMT3A were preceded by pose validation using a re-docking co-crystalized ligand approach. A total of 447 compounds predicted as active against DNMTs by the LDA-based QSAR model were evaluated by molecular docking with DNMT1 (3SWR) and DNMT3A (2QRV). Finally, 67 NPs were selected according to the obtained affinity values (kcal/mol)[54] selecting those NPs with values less than -9.0 kcal/mol for 3SWR and less than -7.0 kcal/mol for 2QRV. These cut-off values were established according to the results obtained by molecular docking for the compounds used in the training and test sets on 3SWR and 2QRV structures (See Table S10).

In order to optimize the selection of compounds, 67 NPs selected by affinity values criteria with AutoDock Vina were subsequently evaluated with Surflex-Dock (SYBYL-X 2.0) as a third filter. The selection criteria for NPs was a Total score value greater than 6.0 (-log(*Kd*)) for both

DNMT structures used, selecting the best 19 NPs as promising DNMTis, as observed in Table 3, which also show affinity values obtained with AutoDock Vina for these compounds.

Table 3 may go here

Several studies focused in DNMTis discovery have been developed using protein-ligand docking approaches. Yoo et al. (2012) reported aurintricarboxylic acid (ATA) as a novel inhibitor of DNMT1 and DNMT3A, employing the Glide program [9] to explain also the relevant implications in DNMTs inhibition [72]. Medina-Franco et al. (2014), conducted a structure-based rationalization of the activity of SW155246 and their structural analogues with human DNMT1 using the same program [45]. Although a number of studies for the discovery of DNMTis have been supported by methods of molecular docking, this report is the first where two different docking protocols, AutoDock Vina and Surflex-Dock are used to determine the interaction feasibility between compounds and DNMTs, applying the diversity of criteria for selection of NPs as new DNMTis.

3.2.2. Molecular docking validation with DNMTis biological data. The relationship between calculated binding scores (AutoDock Vina and Total score) and IC50 values (μM) could be taken as a measure of the likeliness of a particular compound to behave as a DNMTi. A group of forty-five DNMT1 inhibitors (AID: 602386) were docked to DNMT1 (PDB code: 3SWR), and their respective binding affinity and total score values were calculated. The relationship between the biological activity (logIC50) and the binding scores for these inhibitors on DNMT1 are shown in **Fig. 4**. The correlation analysis indicated the inhibition of DNMT1 activity follows a highly dependence with calculated binding scores for these compounds (r= 0.83, *p* value <0.0001; for binding affinity *vs* logIC50, r= 0.65, *p* value <0.0001; for Total score *vs* logIC50), with correlation coefficients comparable to those reported in other studies for these validations [56, 59, 60, 73] as shown in **Fig. 4a** and **Fig. 4b**. These relationships were mostly linear in nature. PubChem chemical structure identifier (CID), AutoDock Vina affinity value, Total score values, and biological activity (IC50 and logIC50) reported for DNMTis are presented in Table S11. The data revealed that NPs with affinity values lower than −9.0 kcal/mol and Total score values greater than 6.0 are likely to have IC50 values in the range of 4 to 25 μM, which suggest them as good candidates for DNMTs inhibition.

Figure 4 may go here

3.3. Final Selection of NPs as new DNMTis by Cluster Analysis, Contact Patterns and Biological Information Reported so far.

3.3.1 Cluster Analysis 19 NPs selected. Molecular diversity analysis by *k*-mean clustering for calculated 28 MDs of 19 NPs selected, and training and test set compounds; showed better clustering for compounds with *k* value = 11. Below are described the clusters containing the NPs selected: *Cluster 1*; comprised of dihydromunduletone, phloridzin, derrubone, pomiferin and 4,2'-Dihydroxychalcone 4-glucoside were placed in the same cluster with the active compounds ATA, EGCG and NCS97317 with centroid distance values between 0.31 to 0.51; *Cluster 7*: comprised of active compounds SGI-110, laccaic acid, pectolinarin, amygdalin, iridin, centaurein, daunorubicin, pyrromycin, naringin grouped with the centroid distance values between 0.19 to 0.63; suggesting their probability as DNMTis. Interesting results were obtained from *Cluster 11* comprised of 9,10-dihydro-12-hydroxygambogic acid, digoxin, reserpine, berbamine, paclitaxel, and dimethyl gambogate. In this case, some NPs were not structurally similar to well-known DNMTis, suggesting that these compounds could become in new scaffold for cancer therapies. Folic acid was grouped in *Cluster 2* with the active compounds aza-Adomet, adoHcy, chlorogenic acid and sinefungin with centroid distance value of 0.46. Considering the representativeness of each cluster, 12 NPs were selected from the last analysis; these NPs were: phloridzin, pomiferin, 2',4'-dihydroxychalcone 4-glucoside, centaurein, daunorubicin, pyrromycin, 9,10-dihydro-12-hydroxygambogic acid, digoxin, reserpine, paclitaxel, folic acid and amygdalin (See Table S12).

3.3.2 Contact Patterns at Distances Less Than 4 Å. A chemical environment analysis for these 12 selected NPs on the DNMTs binding site was conducted by counting and identifying the nearest-neighbor residues for each compound in order to determinate contact patterns at distances less than 4 Å on DNMTs structures (3SWR and 2QRV), and then comparing contact residues of selected NPs against those observed for known DNMTis, such as RG-108, sinefungin, SGI-110 and ATA (contacts for inactive compounds were also considered). Residues involved in the inhibition mechanism of DNMT1 and DNMT 3A [72, 74], such as, Phe-1145, Glu-1168, Glu-1169, Cys-1191, Glu-1266 and Val-1580 showed distances of less than 4 Å with selected NPs on 3SWR binding site. Furthermore, for 2QRV relevant residues were Phe-636, Ile-639, Glu-660, Val-661 and Asp-682, being these residues of interest in the searching for new DNMTis (Details of this analysis can be

seen in Tables S13 and S14). Finally, six NPs were selected as new and promising DNMTis (**Fig. 5**); which are structurally different to those DNMT is reported in the literature.

Figure 5 may go here

3.3.3 Biological Information of Selected NPs. Exhaustive searching of biological information for selected NPs was conducted. A summary of the most relevant information for each of these compounds is presented below.

Phloridzin is a dietary natural product specifically found in apples with well-known beneficial biological effects [75]. It has been used as a pharmaceutical and research tool for physiology for over 150 years. This compound has been related to inhibition of the sodium-glucose symporters [76], and commonly applied to the treatment of diseases such as diabetes [77], obesity [78], stress [79], inflammation [80], polycystic kidney disease [81], and cancer [82].

2',4'-Dihydroxychalcone 4'-glucoside is isolated from *Adhatoda vasica* flowers [83]. This chemical contains very little biological information in the literature; however, its aglycone 2', 4'-Dihydroxychalcone and its derivatives have been implicated in pharmacological properties including, antimicrobial [84], antidepressant [85], and antineoplastic effects [86, 87].

Centaurein, a flavonoid isolated from *Bidens pilosa*, has been associated with immunomodulatory [88], antifungal [89], and anti-oxidant properties [90]. The role of centaurein in IFN-gamma regulation has been reported by inducing activity of NFAT and NFkB enhancers in Jurkat cells [91].

Daunorubicin, an anthracycline produced by *Streptomyces peucetius* [92], has been extensively used for the treatment of acute promyelocytic leukemia by acting through epigenetic inactivation of genes in the INK4/CDK/RB cell cycle pathway [93, 94], and other cancer therapies [95, 96], as well as against unwanted intraocular proliferation [97]. However, its anticancer effects have been related to inhibition of DNA topoisomerase II [98].

Pyrromycin is also an anthracycline produced by *Streptomyces* [99] and its action modes has been mostly related to DNA topoisomerase II inhibition. This NP has shown several pharmacological activities, including antibiotic [100] and antitumor effects [101]. Also, this compound has been identified as an up-regulator of the ATP-binding cassette transporter A1 (ABCA1), a membrane transporter that directly contributes to high-density lipoprotein (HDL) biogenesis [102].

Finally, 9,10-dihydro-12-hydroxygambogic acid is a derivative of gambogic acid, found in *Garcinia hurburyi* tree [103]. Gambogic acid analogs have been proposed as caspases activators and apoptosis inducers [104]. Besides, it has been related to growth inhibition of distinct cancer cell lines, including hepatoma, breast, gastric, lung carcinoma, and also T-cell lymphoma cells [105]. Although it has been related with histone acetyltransferases (HAT) inhibitory properties, its anticancer mechanism is not clearly understood [106].

The discovery of inhibitors of DNMT1 is under permanent attention. A well-known bioassay, which uses a novel non-radioactive high-throughput process, has been employed to assay ligands as human DNMT1 inhibitors. The data are the most comprehensive worldwide (359521 compounds) and the results are publicly available at PubChem (http://pubchem.ncbi.nlm.nih.gov/) and ChEMBL (www.ebi.ac.uk/chembl/) databases, under the tittle: uHTS identification of DNMT1 inhibitors in a Fluorescent Molecular Beacon assay [Primary Screening] (AID: 588458). NPs digoxin, pomiferin, folic acid, paclitaxel, naringin, and amygdalin were classified as inactive for DNMT1 in this assay, and they were not included in the final group of selected NPs. However, classical DNMTis inhibitors, such as curcumin [43], genistein [107], caffeic acid and chlorogenic acid [40], among others, were also reported as inactive in that experimental assay. Although fluorescence-based high-throughput screening (HTS) has been proposed as an economical, rapid and robust technique [108], these false inactives could be explained considering factors such as low solubility of evaluated compounds and interference by stray excitation light, which may underestimate activities and reduce HTS-hit rates [109].

3.3.4 Binding Mode Analysis for NPs on DNMTs Active Site. The binding pose for two selected NPs with the best affinity value (2',4'-dihydroxychalcone 4'-glucoside) and Total score (9,10-dihydro-12-hydroxygambogic acid) were analyzed by characterizing the interactions on DNMTs binding site, as predicted using LigandScout 3.1 software [63]. This software uses simplified pharmacophores based on Molecular Operating Environment (MOE) to detect the number and nature of interactions existing between ligand-residue on the protein binding site. The binding mode of 9,10-dihydro-12-hydroxygambogic acid and 2',4'-dihydroxychalcone 4'-glucoside on DNMT1 and DNMT3A binding sites are shown in **Fig. 6a** and **Fig. 7a**, respectively.

Interacting residues for 9,10-dihydro-12-hydroxygambogic acid binding site on DNMT1 determined by LigandScout program are shown in 3D (**Fig. 6b**) and 2D (**Fig. 6c**) views. Glu-1168

and Gln-1127 presented H-bond interaction with hydroxyl and carboxyl group of 9,10-dihydro-12-hydroxygambogic acid, and hydrophobic interactions were observed for amino acids Phe-648, Met-1169 and Leu-1247 (**Fig. 6c**). On the other hand, 2',4'-dihydroxychalcone 4'-glucoside forms H-bonds through the hydroxyl groups present in the glycoside with the backbone of Phe-1145 and Ala-699, and the side chain of Glu-1266 in the ENV motif, which is a target of most DNMTis (motif IV: E/Glu-1266, N/Asn-1267, V/Val-1268). This last residue is an important residue because it participates directly in the mechanism of C5 cytosine methylation by stabilizing the substrate [9]. The residues Met-1169, Leu-1247, Ile-1167, and Phe-1145 showed hydrophobic interactions for this compound (**Fig. 6d** and **Fig. 6e**). Of interest, Met-1169 and Phe-1145, like Glu-1266 and Glu-1168, are present as interacting residues in the crystallographic structure of DNMT1-sinefungin complex (PDB code: 3SWR).

Figure 6 may go here

The best conformation for 9,10-dihydro-12-hydroxygambogic acid on DNMT3A binding site (**Fig. 7b** and **Fig. 7c**) forms a H-bond with Arg-887, and an ionizable negative interaction with the side-chain of Arg-684 through the carboxyl group. However, only hydrophobic interactions are observed with Val-661, Leu-726, Val-754 and Leu-884. The interactions with Arg-684 and Arg-887 residues have been reported for the inhibitor SGI-1027 on DNMT3A binding site [72]. The best conformations for 2',4'-dihydroxychalcone-4'-glucoside on DNMT3A binding site (**Fig. 7d** and **Fig. 7e**) showed H-bonds with the side-chain of Ser-704, Cys-706, Glu-752 through the hydroxyl groups of the glycoside fragment. For Phe-636, Val-661, Val-683 and Leu-726 residues only hydrophobic interactions were observed. It is important to highlight that the interaction with Phe-636 on DNMT3A binding site has been reported for known DNMTis, SGI-1027 RG108 and procainamide-conjugate (CBC12) [72]. Also, Phe-636 and Val-683 have been reported for the crystallographic DNMT3A-SAH complex (PDB code2QRV).

Binding mode analysis of NPs on DNMTs reveals that compounds such as 9,10-dihydro-12-hydroxygambogic acid and 2',4'-dihydroxychalcone 4'-glucoside have important interactions on DNMT1 and DNMT3A binding sites, according to reports based on the crystallographic structures of inhibitors with DNMT1 (PDB: 3SWR) and DNMT3A (PDB: 2QRV), as well as from molecular docking studies with known DNMTis [43, 45, 72]. These results suggest the potential application of these NPs in anticancer therapies through the reduction of DNMTs activity.

Figure 7 may go here

It is clear that some compounds reported here are clinical drugs, such as daunorubicin; which have pharmacokinetic and toxicological profiles as well as a well-known synthetic path. However, to date, these selected NPs have not been reported as DNMTis. Computational approaches proposed here may allow to make repurposing of these NPs in epigenetic studies. Furthermore, these compounds are structurally different from the well-known DNMTis reported so far, opening a new window to the design of more potent and specific compounds in the treatment of cancer and related diseases.

4 CONCLUSIONS

Proper identification of a lead molecule is the most critical component of the drug discovery process. To this purpose, this report has shown how computational approaches like QSAR, and molecular docking procedures, combined with statistical techniques analysis could be used as tools to ensure high hit rates expected in the discovery of new cancer drugs. LDA-based QSAR model of active/inactive molecules used in this virtual screening, showed satisfactory results, with an accuracy of 93%, molecular docking results using AutoDock Vina and Surflex-Dock, combined with *k*-means analysis cluster, with k=11, and contact patters of interacting residues up to 4 Å, suggest that several new structural chemotypes NPs (9,10-dihydro-12-hydroxygambogic acid, phloridzin, 2',4'-dihydroxychalcone 4'-glucoside, daunorubicin, pyrromycin and centaurin) can be potential DNMTis. It is also suggested that the therapeutic effect of NPs could be achieved not only by inhibition of DNMT1 but also of DNMT3A. A major perspective of this study is the experimental validation of the consensus hits, and in addition to this, the multistep strategy used here can be implemented to screen other larger natural products databases. Our results also reinforce the notion that compounds with inhibitory DNMT activity can be found in natural sources, including dietary products. The methodology proposed in this study is an innovative approach for targeted identification of novel and selective inhibitors of DNMTs as a first step prior to *in vitro* and *in vivo* evaluations useful in the search of anticancer drugs from NPs.

ACKNOWLEDGEMENTS

The authors wish to thank to the University of Cartagena, Cartagena (Colombia); as well as the program to support research groups, sponsored by the Vice-Rectory for Research of the University of Cartagena (2013-2014), CENIVAM (RC-0572-7-2012) and the National Program for Doctoral

Studies in Colombia (Colciencias, 528-2011). Technical help from Ricardo Hernandez Lambraño is highly appreciated. Marrero-Ponce, Y. thanks the Ph.D. Program in Environmental Toxicology at the University of Cartagena.

REFERENCES

[1] Ren, J., Singh, B.N., Huang, Q., Li, Z., Gao, Y., Mishra, P., et al. DNA hypermethylation as a chemotherapy target. Cell Signal. 2011, 23, 1082-93.
[2] Vogt, A., Tamewitz, A., J Skoko, RP Sikorski, Giuliano, K., Lazo, J. The Benzo[c]phenanthridine Alkaloid, Sanguinarine, Is a Selective, Cell-active Inhibitor of Mitogen-activated Protein Kinase Phosphatase-1. J Biol Chem 2005, 280, 19078.
[3] Venza, M., Visalli, M., Catalano, T., Fortunato, C., Oteri, R., Teti, D., et al. Impact of DNA methyltransferases on the epigenetic regulation of tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) receptor expression in malignant melanoma. Biochem Biophys Res Commun. 2013, 441, 743-50.
[4] Fernandez, A.F., Huidobro, C., Fraga, M.F. De novo DNA methyltransferases: oncogenes, tumor suppressors, or both? Trends Genet. 2012, 28, 474-9.
[5] Sun, X., He, Y., Huang, C., Ma, T.-T., Li, J. The epigenetic feedback loop between DNA methylation and microRNAs in fibrotic disease with an emphasis on DNA methyltransferases. Cell Signal. 2013, 25, 1870-6.
[6] Santi, D.V., Garrett, C.E., Barr, P.J. On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. Cell. 1983, 33, 9-10.
[7] Taylor, S.M., Jones, P.A. Mechanism of action of eukaryotic DNA methyltransferase: Use of 5-azacytosine-containing DNA. J Mol Biol. 1982, 162, 679-92.
[8] Yoder, J.A., Soman, N.S., Verdine, G.L., Bestor, T.H. DNA (cytosine-5)-methyltransferases in mouse cells and tissues. studies with a mechanism-based probe. J Mol Biol. 1997, 270, 385-95.
[9] Yoo, J., Kim, J.H., Robertson, K.D., Medina-Franco, J.L. Molecular Modeling of Inhibitors of Human DNA Methyltransferase with a Crystal Structure: Discovery of a Novel DNMT1 Inhibitor. In: Advances in Protein Chemistry and Structural Biology, Christo, C., Tatyana, K.-C. Eds., Academic Press, 2012, Vol. Volume 87, pp. 219-47.
[10] Okano, M., Bell, D.W., Haber, D.A., Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell. 1999, 99, 247-57.
[11] Li, J.-Y., Pu, M.-T., Hirasawa, R., Li, B.-Z., Huang, Y.-N., Zeng, R., et al. Synergistic Function of DNA Methyltransferases Dnmt3a and Dnmt3b in the Methylation of Oct4 and Nanog. Mol Cell Biol. 2007, 27, 8748-59.
[12] Dhawan, D., Ramos-Vara, J.A., Hahn, N.M., Waddell, J., Olbricht, G.R., Zheng, R., et al. DNMT1: An emerging target in the treatment of invasive urinary bladder cancer. Urol Oncol. 2013, 31, 1761-9.
[13] Cheng, P., Chen, H., Zhang, R.-P., Liu, S.-r., Zhou-Cun, A. Polymorphism in DNMT1 may modify the susceptibility to oligospermia. Reprod Biomed Online. 2014, 28, 644-9.
[14] Bian, E.-B., Huang, C., Wang, H., Chen, X.-X., Zhang, L., Lv, X.-W., et al. Repression of Smad7 mediated by DNMT1 determines hepatic stellate cell activation and liver fibrosis in rats. Toxicol Lett. 2014, 224, 175-85.
[15] Castellano, S., Kuck, D., Sala, M., Novellino, E., Lyko, F., Sbardella, G. Constrained Analogues of Procaine as Novel Small Molecule Inhibitors of DNA Methyltransferase-1. J Med Chem. 2008, 51, 2321-5.
[16] Asgatay, S., Champion, C., Marloie, G., Drujon, T., Senamaud-Beaufort, C., Ceccaldi, A., et al. Synthesis and Evaluation of Analogues of N-Phthaloyl-l-tryptophan (RG108) as Inhibitors of DNA Methyltransferase 1. J Med Chem. 2013, 57, 421-34.
[17] Suzuki, T., Tanaka, R., Hamada, S., Nakagawa, H., Miyata, N. Design, synthesis, inhibitory activity, and binding mode study of novel DNA methyltransferase 1 inhibitors. Bioorg Med Chem Lett. 2010, 20, 1124-7.
[18] Valente, S., Liu, Y., Schnekenburger, M., Zwergel, C., Cosconati, S., Gros, C., et al. Selective Non-nucleoside Inhibitors of Human DNA Methyltransferases Active in Cancer Including in Cancer Stem Cells. J Med Chem. 2014, 57, 701-13.
[19] Van De Voorde, L., Speeckaert, R., Van Gestel, D., Bracke, M., De Neve, W., Delanghe, J., et al. DNA methylation-based biomarkers in serum of patients with breast cancer. Mutat Res Rev Mutat Res. 2012, 751, 304-25.

[20] Mai, A. Targeting Epigenetics in Drug Discovery. ChemMedChem. 2014, 9, 415-7.
[21] Lyko, F., Brown, R. DNA Methyltransferase Inhibitors and the Development of Epigenetic Cancer Therapies. J Natl Cancer Inst. 2005, 97, 1498-506.
[22] Billam, M., Sobolewski, M., Davidson, N. Effects of a novel DNA methyltransferase inhibitor zebularine on human breast cancer cells. Breast Cancer Res Treat. 2010, 120, 581-92.
[23] Gray, S., Baird, A., O'Kelly, F., Nikolaidis, G., Almgren, M., Meunier, A., et al. Gemcitabine reactivates epigenetically silenced genes and functions as a DNA methyltransferase inhibitor. Int J Mol Med. 2012, 30 1505-11.
[24] Fandy, T.E., Jiemjit, A., Thakar, M., Rhoden, P., Suarez, L., Gore, S.D. Decitabine Induces Delayed Reactive Oxygen Species (ROS) Accumulation in Leukemia Cells and Induces the Expression of ROS Generating Enzymes. Clin Cancer Res. 2014, 20, 1249-58.
[25] Quintas-Cardama, A., Santos, F.P.S., Garcia-Manero, G. Therapy with azanucleosides for myelodysplastic syndromes. Nat Rev Clin Oncol. 2010, 7, 433-44.
[26] Méndez-Lucio, O., Tran, J., Medina-Franco, J.L., Meurice, N., Muller, M. Toward Drug Repurposing in Epigenetics: Olsalazine as a Hypomethylating Compound Active in a Cellular Context. ChemMedChem. 2014, 9, 560-5.
[27] Brueckner, B., Garcia Boy, R., Siedlecki, P., Musch, T., Kliem, H.C., Zielenkiewicz, P., et al. Epigenetic Reactivation of Tumor Suppressor Genes by a Novel Small-Molecule Inhibitor of Human DNA Methyltransferases. Cancer Res. 2005, 65, 6305-11.
[28] Datta, J., Ghoshal, K., Denny, W.A., Gamage, S.A., Brooke, D.G., Phiasivongsa, P., et al. A New Class of Quinoline-Based DNA Hypomethylating Agents Reactivates Tumor Suppressor Genes by Blocking DNA Methyltransferase 1 Activity and Inducing Its Degradation. Cancer Res. 2009, 69, 4277-85.
[29] Rilova, E., Erdmann, A., Gros, C., Masson, V., Aussagues, Y., Poughon-Cassabois, V., et al. Design, Synthesis and Biological Evaluation of 4-Amino-N-(4-aminophenyl)benzamide Analogues of Quinoline-Based SGI-1027 as Inhibitors of DNA Methylation. ChemMedChem. 2014, 9, 590-601.
[30] Sharma, S., Kelly, T.K., Jones, P.A. Epigenetics in cancer. Carcinogenesis. 2010, 31, 27-36.
[31] Medina-Franco, J.L., López-Vallejo, F., Kuck, D., Lyko, F. Natural products as DNA methyltransferase inhibitors: a computer-aided discovery approach. Mol Divers. 2011, 15, 293–304.
[32] Cragg, G.M., Newman, D.J., Snader, K.M. Natural Products in Drug Discovery and Development. J Nat Prod. 1997, 60, 52-60.
[33] Medina-Franco, J.L. ADVANCES IN COMPUTATIONAL APPROACHES FOR DRUG DISCOVERY BASED ON NATURAL PRODUCTS. Rev. Latinoamer. Quím. 2013, 41, 95-110.
[34] Liu, Z., Xie, Z., Jones, W., Pavlovicz, R.E., Liu, S., Yu, J., et al. Curcumin is a potent DNA hypomethylation agent. Bioorg Med Chem Lett. 2009, 19, 706-9.
[35] Vaid, M., Prasad, R., Singh, T., Jones, V., Katiyar, S.K. Grape seed proanthocyanidins reactivate silenced tumor suppressor genes in human skin cancer cells by targeting epigenetic regulators. Toxicol Appl Pharmacol. 2012, 263, 122-30.
[36] Fang, M.Z., Wang, Y., Ai, N., Hou, Z., Sun, Y., Lu, H., et al. Tea Polyphenol (−)-Epigallocatechin-3-Gallate Inhibits DNA Methyltransferase and Reactivates Methylation-Silenced Genes in Cancer Cell Lines. Cancer Res. 2003, 63, 7563-70.
[37] Xie, Q., Bai, Q., Zou, L.-Y., Zhang, Q.-Y., Zhou, Y., Chang, H., et al. Genistein inhibits DNA methylation and increases expression of tumor suppressor genes in human breast cancer cells. Genes Chromosomes Cancer. 2014, 53, 422-31.
[38] García, J., Franci, G., Pereira, R., Benedetti, R., Nebbioso, A., Rodríguez-Barrios, F., et al. Epigenetic profiling of the antitumor natural product psammaplin A and its analogues. Bioorg Med Chem. 2011, 19, 3637-49.
[39] Agarwal, S., Amin, K., Jagadeesh, S., Baishay, G., Rao, P., Barua, N., et al. Mahanine restores RASSF1A expression by down-regulating DNMT1 and DNMT3B in prostate cancer cells. Mol Cancer. 2013, 12, 99.
[40] Lee, W.J., Zhu, B.T. Inhibition of DNA methylation by caffeic acid and chlorogenic acid, two common catechol-containing coffee polyphenols. Carcinogenesis. 2006, 27, 269-77.
[41] Fagan, R.L., Cryderman, D.E., Kopelovich, L., Wallrath, L.L., Brenner, C. Laccaic Acid A Is a Direct, DNA-competitive Inhibitor of DNA Methyltransferase 1. J Biol Chem. 2013, 288, 23858-67.
[42] Siedlecki, P., Boy, R.G., Musch, T., Brueckner, B., Suhai, S., Lyko, F., et al. Discovery of Two Novel, Small-Molecule Inhibitors of DNA Methylation. J Med Chem. 2005, 49, 678-83.
[43] Medina-Franco, J., Yoo, J. Docking of a novel DNA methyltransferase inhibitor identified from high-throughput screening: insights to unveil inhibitors in chemical databases. Mol Divers. 2013, 17, 337-44.

[44] Singh, N., Dueñas-González, A., Lyko, F., Medina-Franco, J.L. Molecular Modeling and Molecular Dynamics Studies of Hydralazine with Human DNA Methyltransferase 1. ChemMedChem. 2009, 4, 792-9.
[45] Medina-Franco, J., Méndez-Lucio, O., Yoo, J. Rationalization of Activity Cliffs of a Sulfonamide Inhibitor of DNA Methyltransferases with Induced-Fit Docking. Int J Mol Sci. 2014, 15, 3253-61.
[46] Tripos. Sybyl X. In: 1699S Hanley Rd, St. Louis, MO, 2005.
[47] Mauri, A., Consonni, V., Pavan, M., Todeschini, R. DRAGON software: An easy approach to molecular descriptor calculations. Match-Commun Math Co. 2006, 56, 237-48.
[48] StatSoft. Statistica 8.0.360 In: Tulsa, OK, 2007.
[49] Meneses-Marcel, A., Marrero-Ponce, Y., Machado-Tugores, Y., Montero-Torres, A., Pereira, D.M., Escario, J.A., et al. A linear discrimination analysis based virtual screening of trichomonacidal lead-like compounds: Outcomes of in silico studies supported by experimental results. Bioorg Med Chem Lett. 2005, 15, 3838-43.
[50] J W Godden, F L Stahura, Bajorath, J. Variability of Molecular Descriptors in Compound Databases Revealed by Shannon Entropy Calculations. J Chem Inf Comput Sci. 2000, 40, 796-800.
[51] J W Godden, Bajorath, J. Chemical Descriptors with Distinct Levels of Information Content and Varying Sensitivity to Differences between Selected Compound Databases Identified by SE-DSE Analysis. J Chem Inf Comput Sci. 2002, 42, 87-93.
[52] Barigye, S.J., Pino Urias, R.W., Marrero-Ponce, Y. IMMAN (Information Theory based Chemometric Analysis). CAMD-BIR Unit, Santa Clara 2013.
[53] Barigye, S.J., Marrero-Ponce, Y., Martínez López, Y., Artiles Martínez, L.M., Pino-Urias, R.W., Martínez Santiago, O., et al. Relations Frequency Hypermatrices in Mutual, Conditional and Joint Entropy-Based Information Indices. J. Comput. Chem. 2013, 34, 259-74.
[54] Trott, O., Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010, 31, 455-61.
[55] Tuccinardi, T., Poli, G., Romboli, V., Giordano, A., Martinelli, A. Extensive consensus docking evaluation for ligand pose prediction and virtual screening studies. Journal of Chemical Information and Modeling. 2014, 54, 2980-6.
[56] Carrasco, M.P., Gut, J., Rodrigues, T., Ribeiro, M.H.L., Lopes, F., Rosenthal, P.J., et al. Exploring the Molecular Basis of Qobc1 Complex Inhibitors Activity to Find Novel Antimalarials Hits. Molecular Informatics. 2013, 32, 659-70.
[57] Jain, A. Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. J Comput Aided Mol Des. 2007, 21, 281-306.
[58] Arnatt, C.K., Zhang, Y. G Protein-Coupled Estrogen Receptor (GPER) Agonist Dual Binding Mode Analyses Toward Understanding of Its Activation Mechanism: A Comparative Homology Modeling Approach. Molecular Informatics. 2013, 32, 647-58.
[59] Maldonado-Rojas, W., Olivero-Verbel, J. Food-Related Compounds That Modulate Expression of Inducible Nitric Oxide Synthase May Act as Its Inhibitors. Molecules. 2012, 17, 8118-35.
[60] Maldonado-Rojas, W., Olivero-Verbel, J. Potential interaction of natural dietary bioactive compounds with COX-2. J Mol Graph Model. 2011, 30, 157-66.
[61] DeLano, W.L. The PyMOL Molecular Graphics System. LLC, D.S. Ed., San Carlos, CA, USA, 1998–2003.
[62] Nguyen, E.D., Norn, C., Frimurer, T.M., Meiler, J. Assessment and challenges of ligand docking into comparative models of G-protein coupled receptors. PloS one. 2013, 8, e67302.
[63] Wolber, G., Langer, T. LigandScout:□ 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. J Chem Inf Model. 2004, 45, 160-9.
[64] Vuorinen, A., Nashev, L.G., Odermatt, A., Rollinger, J.M., Schuster, D. Pharmacophore Model Refinement for 11β-Hydroxysteroid Dehydrogenase Inhibitors: Search for Modulators of Intracellular Glucocorticoid Concentrations. Molecular Informatics. 2014, 33, 15-25.
[65] Barigye, S.J., Pino Urias, R.W., Marrero-Ponce, Y., . IMMAN (Information Theory based Chemometric Analysis). CAMD-BIR Unit, Santa Clara 2013.
[66] Portugal, J. Evaluation of molecular descriptors for antitumor drugs with respect to noncovalent binding to DNA and antiproliferative activity. BMC Pharmacol. 2009, 9, 11.
[67] Shi, L.M., Fan, Y., Myers, T.G., O'Connor, P.M., Paull, K.D., Friend, S.H., et al. Mining the NCI Anticancer Drug Discovery Databases:□ Genetic Function Approximation for the QSAR Study of Anticancer Ellipticine Analogues. J Chem Inf Comput Sci. 1998, 38, 189-99.
[68] Zhang, S., Wei, L., Bastow, K., Zheng, W., Brossi, A., Lee, K.-H., et al. Antitumor Agents 252. Application of validated QSAR models to database mining: discovery of novel tylophorine derivatives as potential anticancer agents. J Comput Aided Mol Des. 2007, 21, 97-112.
[69] Montero-Torres, A., García-Sánchez, R.N., Marrero-Ponce, Y., Machado-Tugores, Y., Nogal-Ruiz, J.J., Martínez-Fernández, A.R., et al. Non-stochastic quadratic fingerprints and LDA-based

QSAR models in hit and lead generation through virtual screening: theoretical and experimental assessment of a promising method for the discovery of new antimalarial compounds. Eur J Med Chem. 2006, 41, 483-93.
[70] Casañola-Martin, G.M., Marrero-Ponce, Y., Khan, M.T.H., Khan, S.B., Torrens, F., Pérez-Jiménez, F., et al. Bond-Based 2D Quadratic Fingerprints in QSAR Studies: Virtual and In vitro Tyrosinase Inhibitory Activity Elucidation. Chem Biol Drug Des. 2010, 76, 538-45.
[71] Galvez-Llompart, M., Recio Iglesias, M., Gálvez, J., García-Domenech, R., DOI. Novel potential agents for ulcerative colitis by molecular topology: suppression of IL-6 production in Caco-2 and RAW 264.7 cell lines. Mol Divers. 2013, 17, 573–93.
[72] Yoo, J., Choi, S., Medina-Franco, J.L. Molecular modeling studies of the novel inhibitors of DNA methyltransferases SGI-1027 and CBC12: Implications for the mechanism of inhibition of DNMTs. PloS one. 2013, 8, e62152.
[73] Hare, A.A., Leng, L., Gandavadi, S., Du, X., Cournia, Z., Bucala, R., et al. Optimization of i N-benzyl-benzoxazol-2-ones as receptor antagonists of macrophage migration inhibitory factor (MIF). Bioorg Med Chem Lett. 2010, 20, 5811-4.
[74] Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., Cheng, X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature. 2007, 449, 248-51.
[75] Nair, S., Ziaullah, Z., Rupasinghe, H.P.V. Phloridzin fatty acid esters induce apoptosis and alters gene expression in human liver cancer cells (261.2). FASEB J. 2014, 28.
[76] Ehrenkranz, J.R., Lewis, N.G., Ronald Kahn, C., Roth, J. Phlorizin: a review. Diabetes Metab Res Rev. 2005, 21, 31-8.
[77] Zhao, H., Yakar, S., Gavrilova, O., Sun, H., Zhang, Y., Kim, H., et al. Phloridzin improves hyperglycemia but not hepatic insulin resistance in a transgenic mouse model of type 2 diabetes. Diabetes. 2004, 53, 2901-9.
[78] Najafian, M., Jahromi, M.Z., Nowroznejhad, M.J., Khajeaian, P., Kargar, M.M., Sadeghi, M., et al. Phloridzin reduces blood glucose levels and improves lipids metabolism in streptozotocin-induced diabetic rats. Mol Biol Rep. 2012, 39, 5299-306.
[79] Wang, G.-E., Li, Y.-F., Wu, Y.-P., Tsoi, B., Zhang, S.-J., Cao, L.-F., et al. Phloridzin improves lipoprotein lipase activity in stress-loaded mice via AMPK phosphorylation. Int J Food Sci Nutr. 2014, 0, 1-7.
[80] Huang, W.C., Chang, W.T., Wu, S.J., Xu, P.Y., Ting, N.C., Liou, C.J. Phloretin and phlorizin promote lipolysis and inhibit inflammation in mouse 3T3-L1 cells and in macrophage-adipocyte co-cultures. Mol Nutr Food Res. 2013, 57, 1803-13.
[81] Wang, X., Zhang, S., Liu, Y., Spichtig, D., Kapoor, S., Koepsell, H., et al. Targeting of sodium–glucose cotransporters with phlorizin inhibits polycystic kidney disease progression in Han: SPRD rats. Kidney Int. 2013.
[82] Leveen, H.H., Leveen, R.F., LeVeen, E.G. Inhibiting glucose transport. Google Patents, 1989.
[83] Bhartiya, H.P., Gupta, P.C. A chalcone glycoside from the flowers of Adhatoda vasica. Phytochemistry. 1982, 21, 247.
[84] Alvarez, M., Debattista, N., Pappano, N. Antimicrobial activity and synergism of some substituted flavonoids. Folia Microbiol (Praha). 2008, 53, 23-8.
[85] Guan, L.-P., Zhao, D.-H., Chang, Y., Wen, Z.-S., Tang, L.-M., Huang, F.-F. Synthesis of 2, 4-dihydroxychalcone derivatives as potential antidepressant effect. Drug Res (Stuttg). 2013, 63, 46-51.
[86] Xie, C., Sun, Y., Pan, C.-Y., Tang, L.-M., Guan, L.-P. 2, 4-Dihydroxychalcone derivatives as novel potent cell division cycle 25B phosphatase inhibitors and protein tyrosine phosphatase 1B inhibitors. Pharmazie. 2014, 69, 257-62.
[87] Yang, G.-M., Yan, R., Wang, Z.-X., Zhang, F.-F., Pan, Y., Cai, B.-C. Antitumor effects of two extracts from Oxytropis falcata on hepatocellular carcinoma in vitro and in vivo. Chin J Nat Med. 2013, 11, 519-24.
[88] Soidrou, S.H., Bousta, D., Lachkar, M., Hassane, S.O., El Youbi-Hamsas, A., El Mansouri, L., et al. Immunomodulatory Activity of Phenolic Fraction from Piper Borbonense and Cassytha Filiformis Growing in Comoros Islands. In: Chemistry: The Key to our Sustainable Future, Springer, 2014, pp. 105-12.
[89] Policegoudra, R., Chattopadhyay, P., Aradhya, S., Shivaswamy, R., Singh, L., Veer, V. Inhibitory effect of Tridax procumbens against human skin pathogens. J Herb Med. 2014, 4, 83-8.
[90] Bartolome, A.P., Villaseñor, I.M., Yang, W.-C. Bidens pilosa L.(Asteraceae): Botanical properties, traditional uses, phytochemistry, and pharmacology. Evid Based Complement Alternat Med. 2013, 2013.
[91] Chang, S.-L., Chiang, Y.-M., Chang, C.L.-T., Yeh, H.-H., Shyur, L.-F., Kuo, Y.-H., et al. Flavonoids, centaurein and centaureidin, from Bidens pilosa, stimulate IFN-γ expression. J Ethnopharmacol. 2007, 112, 232-6.

[92] Stutzman-Engwall, K.J., Otten, S., Hutchinson, C.R. Regulation of secondary metabolism in Streptomyces spp. and overproduction of daunorubicin in Streptomyces peucetius. J Bacteriol. 1992, 174, 144-54.
[93] Adès, L., Chevret, S., Raffoux, E., Guerci-Bresler, A., Pigneux, A., Vey, N., et al. Long-term follow-up of European APL 2000 trial, evaluating the role of cytarabine combined with ATRA and Daunorubicin in the treatment of nonelderly APL patients. Am J Hematol. 2013, 88, 556-9.
[94] Chim, C.S., Wong, A.S.Y., Kwong, Y.L. Epigenetic inactivation of INK4/CDK/RB cell cycle pathway in acute leukemias. Annals of Hematology. 2003, 82, 738-42.
[95] Schreier, V.N., Pethő, L., Orbán, E., Marquardt, A., Petre, B.A., Mező, G., et al. Protein Expression Profile of HT-29 Human Colon Cancer Cells after Treatment with a Cytotoxic Daunorubicin-GnRH-III Derivative Bioconjugate. PloS one. 2014, 9, e94041.
[96] Zanette, R.A., Kontoyiannis, D.P. Pre-exposure of Candida species to cytarabine and daunorubicin does not affect their in vitro antifungal susceptibility and virulence in flies. Virulence. 2013, 4, 344.
[97] Chhablani, J., Nieto, A., Hou, H., Wu, E.C., Freeman, W.R., Sailor, M.J., et al. Oxidized porous silicon particles covalently grafted with daunorubicin as a sustained intraocular drug delivery system. Invest Ophthalmol Vis Sci. 2013, 54, 1268-79.
[98] Hasinoff, B.B., Yalowich, J.C., Ling, Y., Buss, J.L. The effect of dexrazoxane (ICRF-187) on doxorubicin- and daunorubicin-mediated growth inhibition of Chinese hamster ovary cells. Anticancer Drugs. 1996, 7, 558-67.
[99] Igina, E., Tulemisova, K., Golubchikov, V., Nikitina, E. Comparison of antibiotic complexes formed by cultures of Streptomyces griseoruber. Antibiot Khimioter. 1989, 34, 13-6.
[100] Oki, T. New anthracycline antibiotics. Jpn J Antibiot. 1977, 30, 70-84.
[101] Nettleton Jr, D.E., Balitz, D.M., Doyle, T.W., Bradner, W.T., Johnson, D.L., O'Herron, F.A., et al. Antitumor agents from bohemic acid complex, III. The isolation of marcellomycin, musettamycin, rudolphomycin, mimimycin, collinemycin, alcindoromycin, and bohemamine. J Nat Prod. 1980, 43, 242-58.
[102] Gao, J., Xu, Y., Yang, Y., Yang, Y., Zheng, Z., Jiang, W., et al. Identification of upregulators of human ATP-binding cassette transporter A1 via high-throughput screening of a synthetic and natural compound library. J Biomol Screen. 2008.
[103] Zhang, H.-Z., Kasibhatla, S., Wang, Y., Herich, J., Guastella, J., Tseng, B., et al. Discovery, characterization and SAR of gambogic acid as a potent apoptosis inducer by a HTS assay. Bioorg Med Chem. 2004, 12, 309-17.
[104] Cai, S.X., Drewe, J.A., Kasibhatla, S., Tseng, B., Wang, Y., Zhang, H.Z. Gambogic acid, analogs and derivatives as activators of caspases and inducers of apoptosis. Google Patents, 2002.
[105] Li, R., Chen, Y., Zeng, L.-l., Shu, W.-x., Zhao, F., Wen, L., et al. Gambogic acid induces G0/G1 arrest and apoptosis involving inhibition of SRC-3 and inactivation of Akt pathway in K562 leukemia cells. Toxicology. 2009, 262, 98-105.
[106] Furdas, S.D., Kannan, S., Sippl, W., Jung, M. Small Molecule Inhibitors of Histone Acetyltransferases as Epigenetic Tools and Drug Candidates. Arch Pharm (Weinheim). 2012, 345, 7-21.
[107] Li, Y., Tollefsbol, T.O. Impact on DNA methylation in cancer prevention and therapy by bioactive dietary components. Curr Med Chem. 2010, 17, 2141.
[108] Ye, Y., Stivers, J.T. Fluorescence-based high-throughput assay for human DNA (cytosine-5)-methyltransferase 1. Anal Biochem. 2010, 401, 168-72.
[109] Di, L., Kerns, E.H. Biological assay challenges from compound solubility: strategies for bioassay optimization. Drug Discov Today. 2006, 11, 446-51.

Tables

Table 1. Compounds Tested Against DNMTs used for Building (training set) and Validating (test set) the LDA-based Discriminant Model

No.	Active (Inhibitor)[a]	No.	Inactive
	Training set		
1	RG108	20	NCS622444
2	5-azacytidine	21	NCS137546
3	5-Fluoro-2′-deoxycytidine	22	NSC319745
4	Zebularine	23	NCS106084
5	Deoxycytidine	24	NCS54162
6	*EGCG*	25	NCS158324
7	*Genistein*	26	NCS348926
8	*Curcumin*	27	CHEMBL1780248
9	*Caffeic acid*	28	NCS4092
10	Procaine	29	NSC21970
11	Hydralazine	30	NCS19555
12	*Nanaomycin A*	31	NCS27292
13	SGI-110	32	NCS303530
14	AdoHcy (SAH)		
15	*Parthenolide*		
16	*Mahanine*		
17	NCS97317		
18	*Chlorogenic acid*		
19	Δ^2Isoxazoline		
	Test set		

1	Aurintricarboxylic acid (ATA)	10	NCS408488
2	RG108-1	11	NCS56071
3	5-aza-2'-deoxycytidine	12	NCS154957
4	5,6-Dihydro-5-azacytidine	13	NCS138419
5	Aza-Adomet	14	NCS345763
6	*Laccaic acid A*	15	NCS27278
7	*Psammaplin A*		
8	SGI-1027		
9	Sinefungin		

[a] Compounds in this study have been previously evaluated against DNMTs activity, Natural DNMTis in *italics*. Chemical structures of both sets (training and test) are shown in **Fig. S1** and **Fig. S2**.

Table 2. Prediction performance of LDA-based QSAR model for classifying molecules as active or inactive against DNMTs.

	Actives	**Inactives**	**N_{total}**	***C*** [a]	**Q_{total}** [b] **(%)**	***Sensitivity rate* (%)**	***Specificity* (%)**	***False positive rate* (%)**
Training set [c]								
Actives	18	1	19					
Inactives	0	13	13	0.94	96.9	94.7	100.0	0.0
N_{total}	18	14	32					
Test Set [d]								
Actives	8	1	9					
Inactives	0	6	6	0.87	93.3	88.9	100.0	0.0
N_{total} [e]	8	7	15					

[a] Matthews correlation coefficient, [b] accuracy value, [c] 32 compounds (19 actives and 13 inactives against DNMTs), [d] 15 compounds (9 actives and 6 inactives against DNMTs), [e] number of compounds.

Table 3. Compounds selected from NatProd collection as DNMT inhibitors using QSAR and molecular docking criteria.

Compound	*Class* [a]	ΔP% [b]	3SWR kcal/mol AV [c]	2QRV kcal/mol AV	3SWR -log(*Kd*) [d] TS [e]	2QRV -log(*Kd*) TS
Pectolinarin	35.10	100.00	-10.4	-7.8	6.1	9.1
Daunorubicin	15.39	100.00	-9.3	-8.2	7.6	7.4
Derrubone	14.82	100.00	-9.1	-8.1	7.5	6.6
9,10-dihydro-12-hydroxygambogic acid	14.33	100.00	-9.5	-7.2	12.5	12.1
Digoxin	13.83	100.00	-11.4	-8.6	6.6	7.3
Pomiferin	13.76	100.00	-9.6	-9.1	7.5	6.4
Dihydromunduletone	13.33	100.00	-9.8	-7.3	7.0	8.1
Iridin	12.63	100.00	-9.4	-7.0	7.9	7.9
Phloridzin	11.03	100.00	-9.1	-8.3	9.2	8.5
Folicacid	8.42	99.96	-9.9	-8.0	8.7	8.1
Naringin	8.17	100.00	-9.8	-8.6	7.6	6.7
Centaurein	8.08	100.00	-9.7	-7.4	8.0	7.6
Amygdalin	7.78	99.99	-9.2	-8.7	9.6	8.1
Pyrromycin	6.79	100.00	-9.2	-7.4	8.9	9.8
4,2'-Dihydroxychalcone 4-glucoside	5.94	99.99	-9.7	-8.9	6.4	7.6
Reserpine	5.77	100.00	-9.2	-7.8	9.1	7.8
Dimethylgambogate	5.69	100.00	-9.2	-8.2	12	11.6
Berbamine	5.02	100.00	-9.4	-7.7	7.4	7.3
Paclitaxel	4.50	100.00	-9.6	-7.8	9.9	9.9

[a] *Class*: Classification obtained with LDA-based QSAR model, [b] ΔP%:[P (Active) − P (Inactive)] × 100, [c] AV: Affinity using AutoDockVina, [d] logarithm negative of dissociation constant. [e] Total score using Surflex-Dock.

Figures

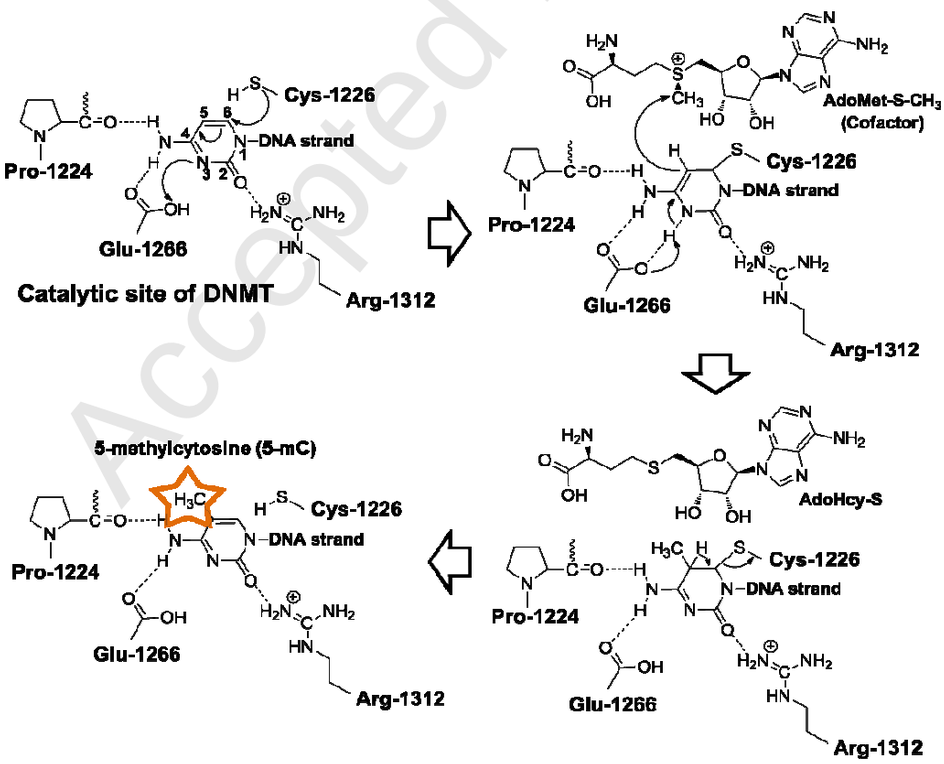


Fig. 1. Mechanism of cytosine DNA methylation catalyzed by DNMTs. DNMTs contain a conserved cysteine residue that attacks the C(6) atom of cytosine forming a covalent bond. A nucleophilic attack occurs on the methyl group of S-adenosyl-L-methionine (AdoMet), which is converted

to S-adenosyl-L-homocysteine (AdoHcy-S). The last step comprises β-elimination across the C(5)-C(6) bond, releasing the enzyme.

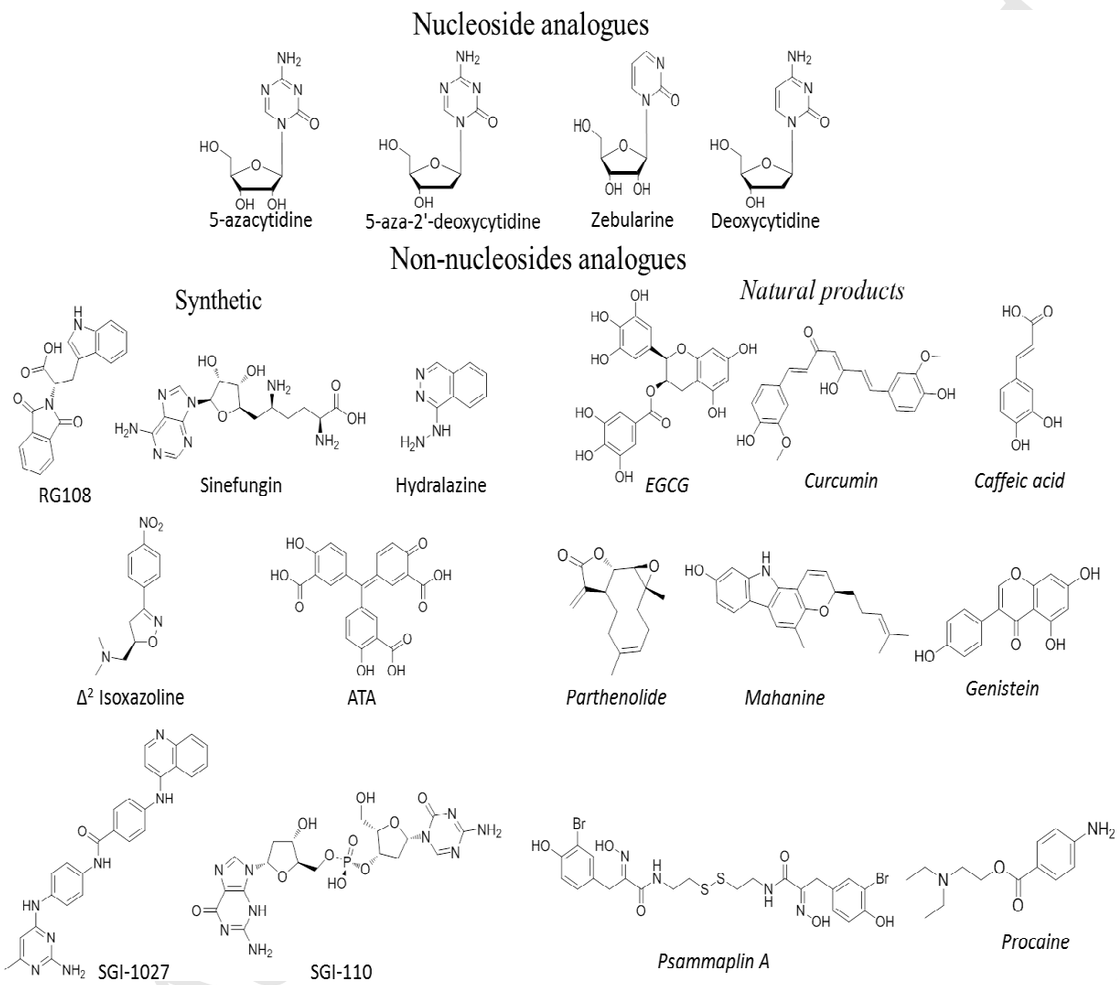


Fig. 2. Chemical structure of representative known nucleoside analogues and non-nucleoside DNMTis, classified as synthetic and natural products.

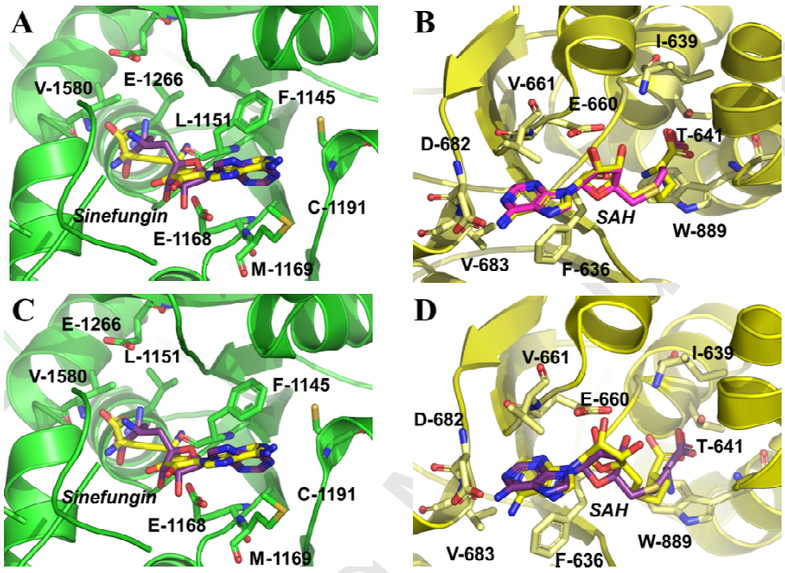


Fig. 3. Molecular superposition between predicted (yellow) and co-crystalized pose (purple) for sinefungin (3SWR) and SAH (2QRV) with AutoDock Vina (A,B) and Surflex-Dock (C,D) protocols, showing interacting residues on experimental complexes with DNMTs.

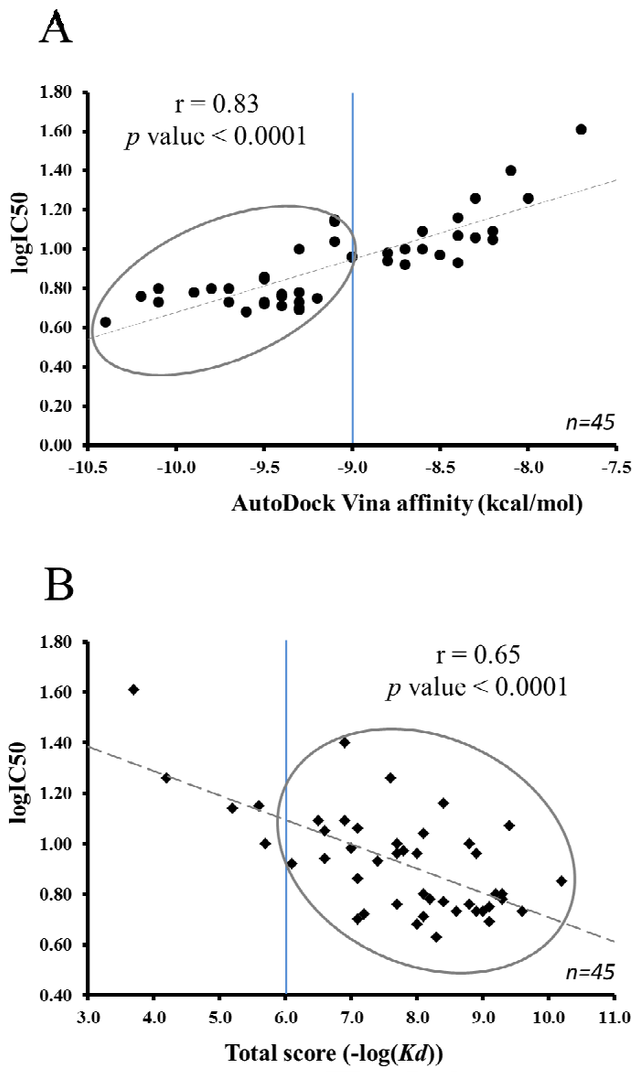


Fig. 4. Correlation between binding affinities calculated by AutoDock Vina (A) and Total score by Surflex-Dock (B) and their half maximal inhibitory concentration for forty-five DNMT1 inhibitors; μM (LogIC50). Circles show molecules with affinity values lower than −9.0 kcal/mol and Total score values greater than 6.0. Regression lines were added for illustrative purposes.

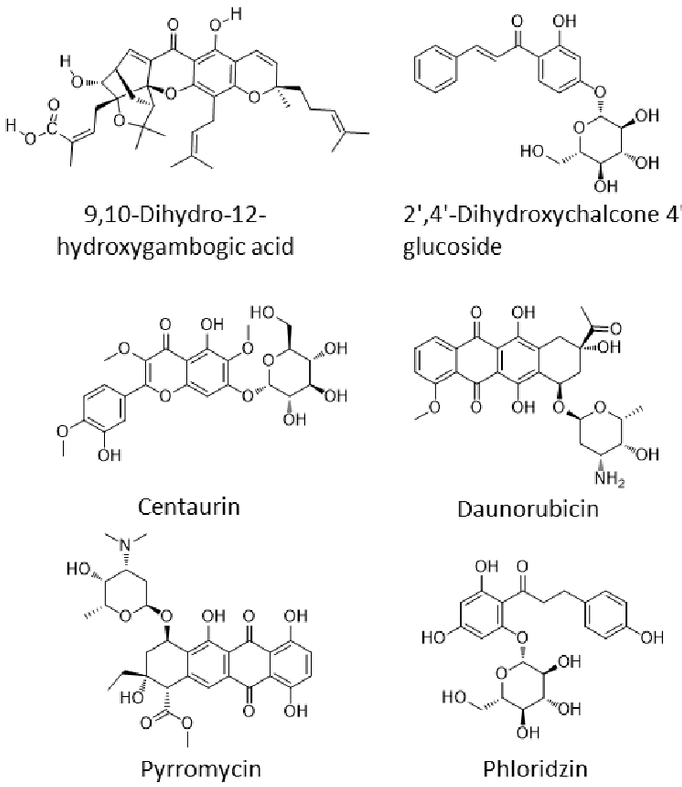


Fig. 5. Six NPs selected as promissory DNMTis.

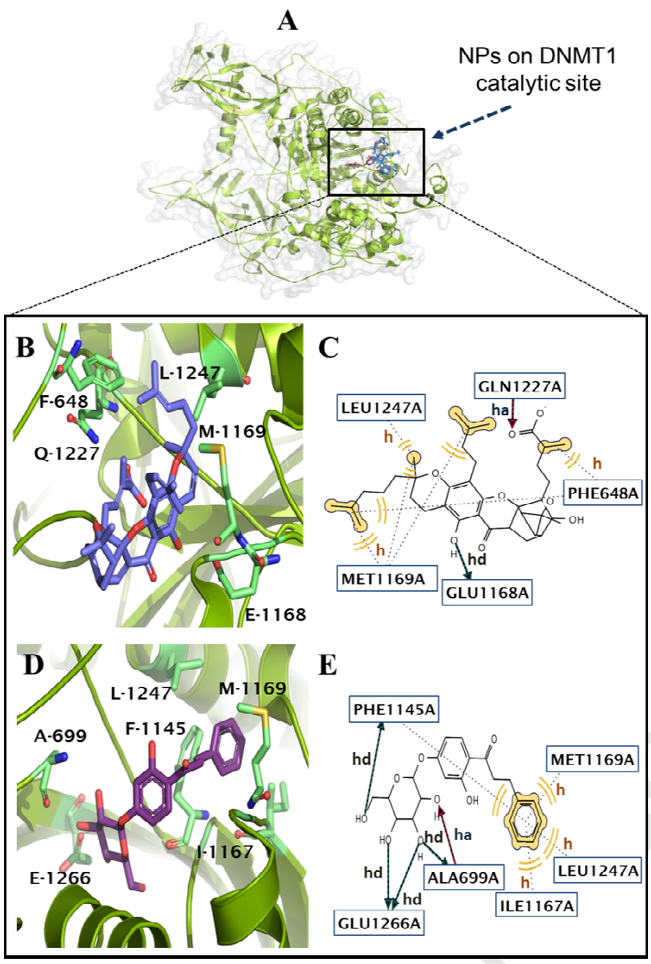


Fig. 6. DNMT1-NPs (9,10-dihydro-12-hydroxygambogic acid and 2',4'-dihydroxychalcone 4'-glucoside) complexes (A). Interacting residues for the best pose of 9,10-dihydro-12-hydroxygambogic acid (3D (B) and 2D (C) views) and 2',4'-dihydroxychalcone 4'-glucoside (3D (D) and 2D (E) views) on DNMT1 binding site as predicted by LigandScout 3.1. Interaction types: h=hydrophobic, ha = H-bond acceptor, hd = H-bond donor.

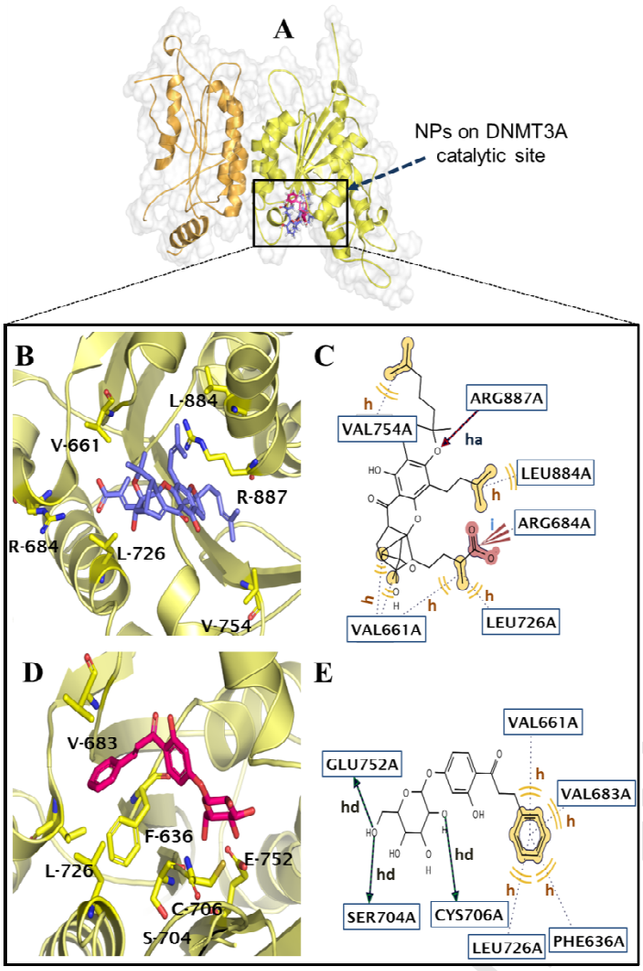
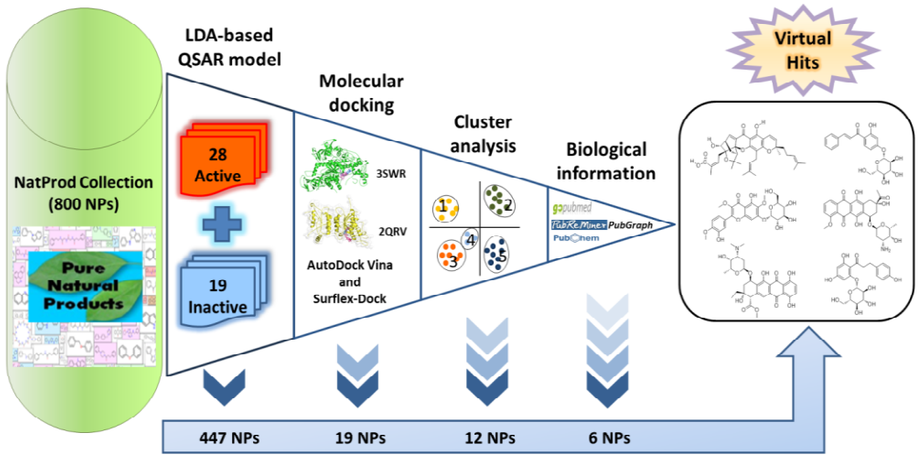


Fig. 7. DNMT3A-NPs (9,10-dihydro-12-hydroxygambogic acid and 2',4'-dihydroxychalcone 4'-glucoside) complex (A). Interacting residues for the best pose of 9,10-dihydro-12-hydroxygambogic acid (3D (B) and 2D (C) views) and 2',4'-dihydroxychalcone 4'-glucoside (3D (B) and 2D (C) views) on DNMT3A binding site as predicted by LigandScout 3.1. Interactions type: i = negative ionizable, h= hydrophobic, ha = H-bond acceptor, hd = H-bond donor.



Scheme 1. Pipeline of virtual screening for DNMTis discovery.