Peer-reviewed version available at Frontiers in Pharmacology 2018, 9, 1144; doi:10.3389/fphar.2018.01144

Review

Computational Approaches to Identify Natural Products as Inhibitors of DNA Methyltransferases

Fernanda I. Saldívar-González, Alejandro Gómez-García, Norberto Sánchez-Cruz, Javier Ruiz-Rios, B. Angélica Pilón-Jiménez and José L. Medina-Franco *

Department of Pharmacy, School of Chemistry, National Autonomous University of Mexico, Mexico City 04510, Mexico; felilang_12@hotmail.com (F.I.S-G.); alex.go.ga21@hotmail.com (A.G-G.); norbertosc90@gmail.com (N.S.-C.); javier_ruiz@chemist.com (J.R-R.); angiepilon96@gmail.com (B.A.P-J.)

* Correspondence: medinajl@unam.mx; Tel.: +52-55-5622-3899. Ext. 44458

Abstract: Naturally occurring small molecules include a large variety of natural products from different sources that have confirmed activity against epigenetic targets. In this work we review chemoinformatic, molecular modeling and other computational approaches that have been used to uncover natural products as inhibitors of DNA metiltransferases, a major family of epigenetic targets with significant potential for the treatment of cancer and several other diseases. Examples of these computational approaches include docking, similarity-based virtual screening, and pharmacophore modeling. It is also commented the chemoinformatic-based exploration of the chemical space of naturally occurring compounds as epigenetic modulators which may have significant implications in epigenetic drug discovery and nutriepigenetics.

Keywords: chemical space; chemoinformatics; data mining; databases; DNMT inhibitors; drug discovery; epi-informatics; molecular modeling; similarity searching; virtual screening

1. Introduction

Epigenetics has been defined as a change in phenotype without an underlying change in genotype [1]. Historically, in the 1940s Conrad Waddington coined the term 'epigenetics' trying to describe "the interactions of genes with their environment, which brings the phenotype into being" [2]. Alterations in epigenetic modifications have been associated with a number of diseases including cancer and other diseases such as diabetes, neurodegenerative disorders, and immune-mediated diseases [3] [4-6]. Moreover, epigenetic targets have also been recognized for the treatment of antiparasitic infections [7].

In epigenetic drug discovery epigenetic targets have been classified into three main groups [8]. 'Writers' are enzymes that catalyze the addition of group to a protein or nucleic acid; 'readers' are macromolecules that function as recognition units that are able to distinguish a native macromolecule vs. the modified one; and 'erasers' that are enzymes that aid in the removal of chemical modifications introduced by the writers. Thus far, several targets from these three major families have reached different stages of the drug discovery, ranging from lead discovery, preclinical development, clinical trials and approval. Currently, there are seven compounds approved for clinical use [8].

Among the 'writers', DNA metiltransferases (DNMTs) is a family of enzymes responsible for DNA methylation that is the addition of a methyl group at C5 position of cytosine. As surveyed in this work, since DNA methylation has an essential role for cell differentiation and development. Alterations in the function of DNMTs have been associated with cancer [9] and other diseases [10].

Several natural products have been identified as inhibitors of epigenetic targets including DNMTs. Most of these compounds have been uncovered from random approaches although, more recently, there are efforts to screen systematically natural products as DNMT inhibitors. The

Peer-reviewed version available at Frontiers in Pharmacology 2018, 9, 1144; doi:10.3389/fphar.2018.01144

vastness of the chemical space of natural products led to the hypothesis that many more active compounds could potentially been identified. Indeed, it has been estimated that more than 95% of the biodiversity in nature has not been explored yet for biological activity [11].

The aim of this work is to discuss a broad range of computational methods to identify novel inhibitors of DNMTs from natural products. The manuscript also discusses the chemical space of natural products as inhibitors of DNMTs. The review is organized into nine major sections. After this introduction, Section 2 reviews briefly the structure of DNMTs including different isoforms. The next section covers major aspects of the function of DNMTs including the mechanism of methylation. Section 4 reviews currently known inhibitors of DNMTs from natural sources including food chemicals. Section 5 discusses the epigenetic relevant chemical space of natural products comparing the chemical space of DNMT inhibitors from natural sources vs. other inhibitors. The next section discusses different computational strategies that are used to identify pharmacologically active natural compounds as epi-hits or epi-leads targeting DNMTs. Sections 7 and 8 presents Summary conclusions and Perspectives, respectively.

2. Structure of DNMTs

The human genome encodes DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L. While DNMT1, DNMT3A and DNMT3B have catalytic activity, DNMT2 and DNMT3L do not. Of note, DNMTs are also conserved in plants and DNMT4, DNMT5 and DNMT6 have been identified in algae and fungi [10].

DNMT1, a maintenance methyltransferase whose structure is shown in Figure 1A, is responsible for duplicating the pattern of DNA methylation during replication, it is essential for proper mammalian development, and it has been proposed as the more interesting target for experimental cancer therapies [3]. DNMT3A and DNMT3B are *de novo* methyltransferases. Human DNMT1 is a protein with 1616 amino acids whose structure can be divided into an N-terminal regulatory domain and a C-terminal catalytic domain [12,13]. The N-terminal domain contains a replication foci-targeting domain, a DNA-binding CXXC domain, and a pair of bromo-adjacent homology domains. The C-terminal catalytic domain has 10 amino acid motifs. The cofactor and substrate binding sites in the C-terminal catalytic domain are comprised of motif I and X and motif IV, VI, and VIII, respectively [14]. The target recognition domain which is maintained by motif IX and involved in DNA recognition is not conserved between the DNMT families. Figure 1A shows a three-dimensional (3D) model of a DNMT1 (PDB ID: 4WXX) [15].

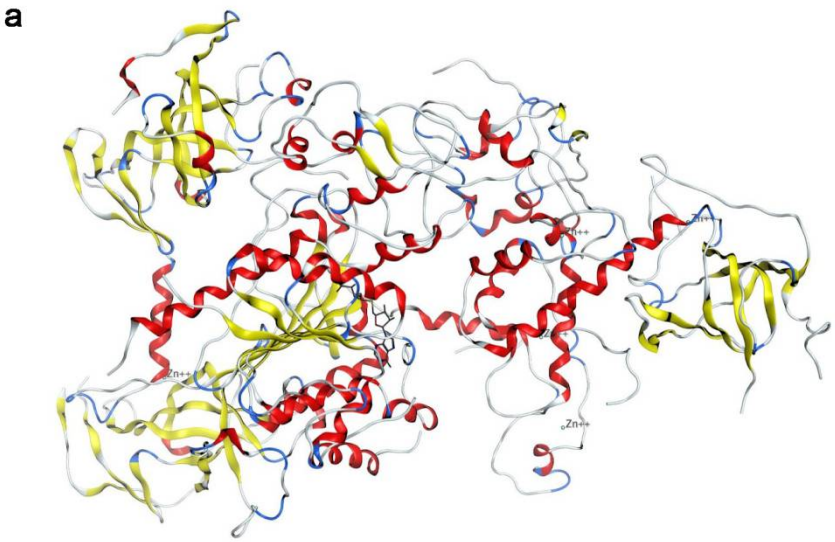
2.1. Isoforms

Two isoforms of DNMT3A have been identified, DNMT3A1 and DNMT3A2. At the N-terminal domain both isoforms have a PWWP (Pro-Trp-Trp-Pro) and an ADD (ATRX-DNMT3-DNMT3L) domains which is a zinc-finger domain. Both domains target DNMT3A to molecules of histone H3 [13]. The C-terminal domain is identical in the 2 isoforms; comprises the catalytic region and retains the motifs of the C-terminal region of DNMT1. The difference between both isoforms is that DNMT3A2 lacks 220 N-terminal amino acids of DNMT3A1. These 220 amino acids do not belong neither PWWP nor ADD domains [16].

There are more than 30 isoforms of DNMT3B, however, only DNMT3B1 and DNMT3B2 are catalytically active [17]. Similar to DNMT3A, DNMT3B1 and DNMT3B2 have a PWWP domain and an ADD domain at the N-terminal domain [10]. The rest of the isoforms are not catalytically active because the C-terminal domain is truncated. Some of these isoforms such as DNMT3B3, DNMT3B4 and DNMT3B7 are overexpressed in many tumor cell lines [18]. ΔDNMT3B has seven isoforms and lacks 200 amino acids from the N-terminal region of DNMT3B. Only ΔDNMT3B1 and ΔDNMT3B2 possess the complete PWWP domain [19]. ΔDNMT3B1-4 possess catalytic activity, whereas ΔDNMT3B5-7 lacks the catalytic domain [19]. ΔDNMT3B is mainly expressed in non-small cell lung cancer [17,19]. Figure 1B shows the identity matrix of 14 DNMTs isoforms. The identity matrix indicates that the amino acid sequence at the catalytic site of DNMT3A1 and DNMT3A2 isoforms is identical. In the same manner, the amino acid sequence at the C-terminal domain of the catalytically

Peer-reviewed version available at *Frontiers in Pharmacology* **2018**, *9*, 1144; doi:10.3389/fphar.2018.01144

active isoforms DNMT3B1, DNMT3B2 and ΔDNMT3B1-4 is identical. DNMT1, DNMT2 and DNMT3L show a significant difference in the sequence of the catalytic site with respect to the rest of the isoforms. Therefore, it can be anticipated that is possible to identity or design selective inhibitors for these isoforms.



b

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1:DNMT1		20.5	12.8	18.3	18.3	17.6	18.1	16.1	16.1	16.1	16.1	16.1	16.1	17.1
2:DNMT2:	17.4		14.3	19.0	19.0	20.9	22.2	19.4	19.4	19.4	19.4	19.4	19.4	21.8
3:DNMT3L	5.4	7.2		23.3	23.3	27.2	29.6	23.3	23.3	23.3	23.3	23.3	23.3	33.5
4:DNMT3A1	11.1	13.6	33.2		100.0	81.6	83.3	80.6	80.6	80.6	80.6	80.6	80.6	81.8
5:DNMT3A2	11.1	13.6	33.2	100.0		81.6	83.3	80.6	80.6	80.6	80.6	80.6	80.6	81.8
6:DNMT3B5	9.1	12.8	33.2	69.9	69.9		100.0	85.7	85.7	85.7	85.7	85.7	85.7	99.4
7:DNMT3B3	8.5	12.3	32.7	64.5	64.5	90.4		77.4	77.4	77.4	77.4	77.4	77.4	99.4
8:DNMT3B1	9.8	13.8	33.2	80.6	80.6	100.0	100.0		100.0	100.0	100.0	100.0	100.0	99.4
9:DNMT3B2	9.8	13.8	33.2	80.6	80.6	100.0	100.0	100.0		100.0	100.0	100.0	100.0	99.4
10:DNMTΔ3B1	9.8	13.8	33.2	80.6	80.6	100.0	100.0	100.0	100.0		100.0	100.0	100.0	99.4
11:DNMTΔ3B2	9.8	13.8	33.2	80.6	80.6	100.0	100.0	100.0	100.0	100.0		100.0	100.0	99.4
12:DNMTΔ3B3	9.8	13.8	33.2	80.6	80.6	100.0	100.0	100.0	100.0	100.0	100.0		100.0	99.4
13:DNMTΔ3B4	9.8	13.8	33.2	80.6	80.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0		99.4
14:DNMT3B4	6.3	9.5	29.1	49.8	49.8	70.7	78.2	60.6	60.6	60.6	60.6	60.6	60.6	

Figure 1. (**a**) Three-dimensional model of DNMT1, amino acid residues 351-1600. Figure rendered from the Protein Data Bank PDB ID: 4WXX. (**b**) Identity matrix of the catalytic site of 14 DNMTs isoforms. Note that there is a significant difference in the sequence of DNMT1, DNMT2 and DNMT3L.

3. Function and mechanism of DNMTs

As outlined in section 2, cytosine-5 DNMTs catalyze the addition of methylation marks to genomic DNA. All DNMTs employ a related catalytic mechanism that is featured by the formation of a covalent adduct intermediate between the enzyme and the substrate base. All DNMTs use *S*-adenosyl-*L*-methionine (SAM) as the donor of the methyl group. Figure 2 depicts the mechanism of DNA cytosine-C5 methylation [20,21]. DNMT forms a complex with DNA and the cytosine which

Peer-reviewed version available at Frontiers in Pharmacology 2018, 9, 1144; doi:10.3389/fphar.2018.01144

will be methylated flips out from the DNA [22]. A conserved cysteine residue in the PCQ motif or motif IV performs a nucleophilic attack to the 6-position of the target cytosine yielding a covalent intermediate. The 5-position of the cytosine is activated and conducts a nucleophilic attack on the cofactor SAM to form the 5-methyl covalent adduct and *S*-adenosyl-*L*-homocysteine (SAH). The attack on the 6-position is aided by a transient protonation of the cytosine ring at the endocyclic nitrogen atom N3, which can be stabilized by a glutamate residue. An arginine residue could assist in the stabilization of the intermediate making a hydrogen bonding interaction with the carbonyl oxygen of cytosine. The covalent complex between the methylated base and the DNA is resolved by deprotonation at the 5-position to generate the methylated cytosine and the free enzyme (Figure 2).

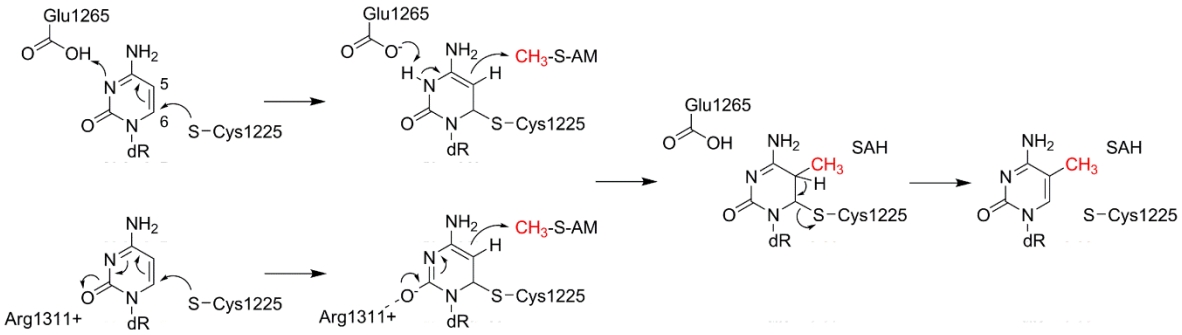


Figure 2. Suggested mechanism of DNA cytosine-C5 methylation. Amino acid residue numbers are based on human DNMT1.

4. Known inhibitors of DNMTs from natural sources

Thus far more than 500 compounds have been tested as inhibitors of DNMTs. The structural diversity and coverage in chemical space has been analyzed using chemoinformatic methods [23,24]. The chemical space of DNMT inhibitors have been compared with inhibitors of other epigenetic targets [25]. Furthermore, the structure-activity relationships (SAR) of DNMT inhibitors using the concept of activity landscape has been documented [26,27] and 3D activity cliffs have been analyzed at the molecular level using induced fit docking [28].

DNMT inhibitors have been obtained from a broad number of different strategies including organic synthesis, virtual screening and high-throughput screening [29]. Organic synthesis has been used in several instances for optimization of lead compounds [30-32]. Natural products and food chemicals have also been a major source of active molecules. Natural products that are known to act as DNMT inhibitors or demethylating agents have been extensively reviewed by Zwergel et al. [33]. These natural products are of the type polyphenols, flavonoids, antraquinones, and others classes. One of the first natural products described were curcumin, (-)-epigallocatechin-3-gallate (EGCG), mahanine, genistein, and quercetin. Other natural products that have described as inhibitors of DNMT or demethylating agents are silibinin, luteolin, kazinol Q, laccaic acid, hypericin, boswellic acid, and lycopene. Figure 3 shows the chemical structure of representative DNMT inhibitors with emphasis on compounds from natural origin.

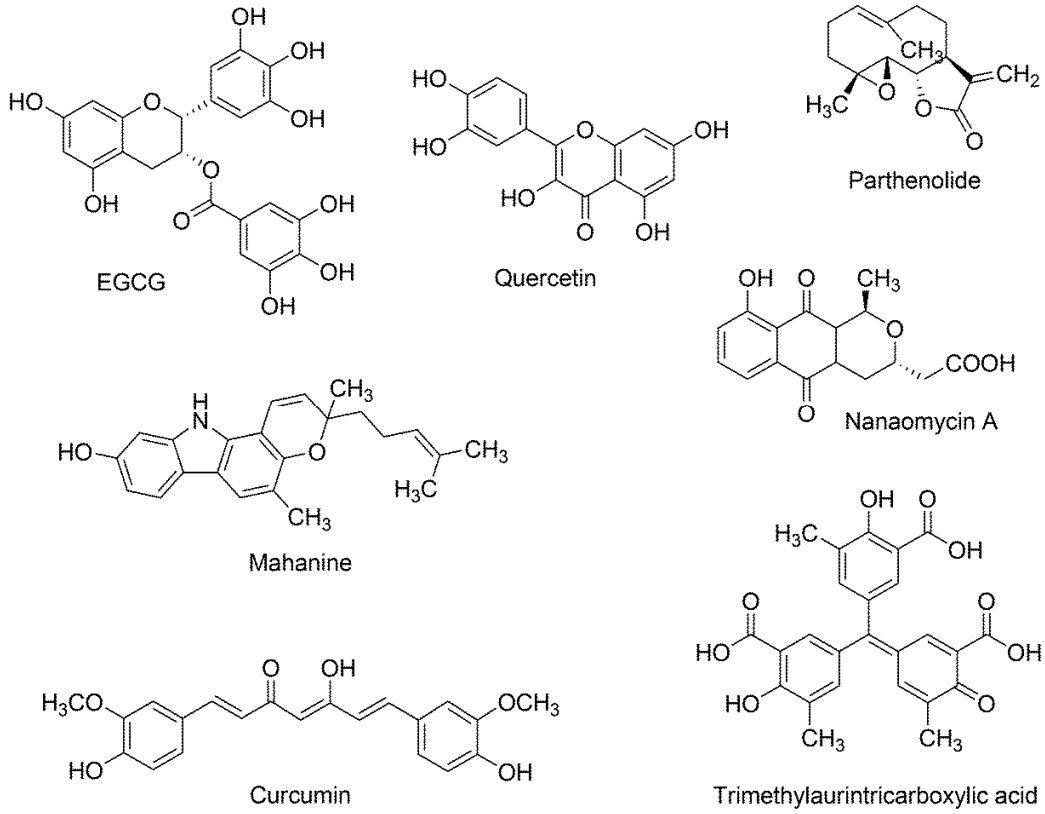


Figure 3. Chemical structures of representative inhibitors of DNMTs from natural sources.

The bioactivity profile and potency in enzymatic and/or cell-based assays of these natural products has been discussed in detail by Zwergel et al. [33]. It will be valuable if all natural products could have been screened under the same conditions. For few natural products the selectivity has been characterized with nanaomycin A as an exception (*vide infra*). Indeed, for about eight natural products the IC_{50} has been measured in enzymatic based assays. Despite the fact the potency of these natural products is not very high in enzymatic-based assays e.g., IC_{50} between 0.5 and 10 μM, several natural products have shown promising activity in cell based assays. Notably, natural products have distinct chemical scaffolds that could be used as a starting point in lead optimization efforts. Moreover, quercetin in combination with green tea extract has advanced into phase I clinical trials for the treatment of prostate cancer.

Most of the natural products with demethylating activity or ability to inhibit DNA methyltransferases in enzymatic assays have been identified fortuitously. However, as discussed in the following sections, there are efforts towards the identification of bioactive demethylating agents using systematic approaches such a virtual screening. Of note, the natural product nanaomycin A (Figure 3) was identified from a virtual screening campaign initially focused on the identification of inhibitors of DNMT1. This quinone-based antibiotic isolated from Streptomyces showed antiproliferative effects in three human tumor cell lines, HCT116, A549 and HL60 after 72 h of treatment. Moreover, the natural product showed reduced global methylation levels in all three cell lines when tested at concentrations ranging from 0.5 μM to 5 μM. Nanaomycin A reactivated transcription of the RASSF1A tumor suppressor gene inducing its expression up to 18-fold at 5 μM, higher than the reference drug 5- azacytidine (6-fold at 25 μM). In a enzymatic inhibitory assay, nanaomycin A showed enzymatic inhibitory activity selectively towards with an IC_{50} = 0.50 μM

4.1. Natural products and food chemicals

It is remarkable that a number of natural products are found as dietary sources such as curcumin, caffeic acid and chlorogenic acid found in Coffea arabica, genistein found in soybean, quercetin found in fruits, vegetables and beverages. Of course, there is a large overlap between the

chemical space of food chemicals and natural products [34]. This has given rise to systematically screen food chemical databases for potential regulators of epigenetic targets using the principles of food informatics [35,36].

5. Epigenetic relevant chemical space of natural products: focus on DNMT inhibitors

In drug discovery projects it is generally accepted that a major benefit of natural products vs. purely synthetic organic molecules is, overall, the feasibility of the former to exert a biological activity and increased chemical diversity [11]. The chemical space of natural products is vast, and its molecular diversity has been quantified in several studies over the years [37-41]. A major contribution to these studies has been the increasing availability of natural products collections in the public domain [42]. Examples of major compound collections are the Traditional Chinese Medicine [43], natural products from Brazil – NUBBE [44], AfroDB [45] or collections available for screening in a medium to high-throughput screening mode. These and other natural product databases are reviewed elsewhere [42]. The large importance of natural products in drug discovery has boosted the development of open access applications to mine these rich repositories. Few examples are ChemGPS-NP, TCMAnalyzer and other resources described elsewhere [46-49].

The chemical space of natural products from different sources has been compared to several other collections including the chemical space of drugs approved for clinical use and synthetic compounds [37-41]. These studies revels that certainly the chemical space of natural products is vast; there is a notable overlap with the chemical space of drugs but also natural products cover novel regions of the chemical space. The overlap with the chemical space of approved drugs is not that surprising since there are a large percentage of drugs from natural origin. Figure 4 shows a visual representation of the chemical space of 15 representative DNMT inhibitors from natural sources vs 4103 compounds for a commercial vendor library of natural products (MEGX www.analyticon.com), 206 fungi metabolites [39], and 6253 marine natural products [50]. The visual representation was generated with principal component analysis of six physicochemical properties of pharmaceutical relevance, namely molecular weight (MW), topological surface area (TPSA), number of hydrogen bond donors and acceptors (HBD/HBA), number of rotatable bonds (RB) and octanol/water partition coefficient (logP). The first two principal components capture about 90% of the total variance (meaning that the two-dimensional -2D- plot in Figure 4 is a good approximation of the chemical space). The visual presentation of the chemical space in Figure 4 indicates that the marine natural products (data points in blue) cover a broader area of the chemical space followed by natural products in the MEGX collection (data points in orange) and by fungi metabolites (data points in green). DNMT inhibitors from natural origin (purple) are, in general, inside the subspace of the DNMT1 inhibitors (data points in red). This visualization of the chemical space indicates that there would be expected to identify more DNMT1 inhibitors in the marine and MEGx collections, as well as in the data set of fungi metabolites.

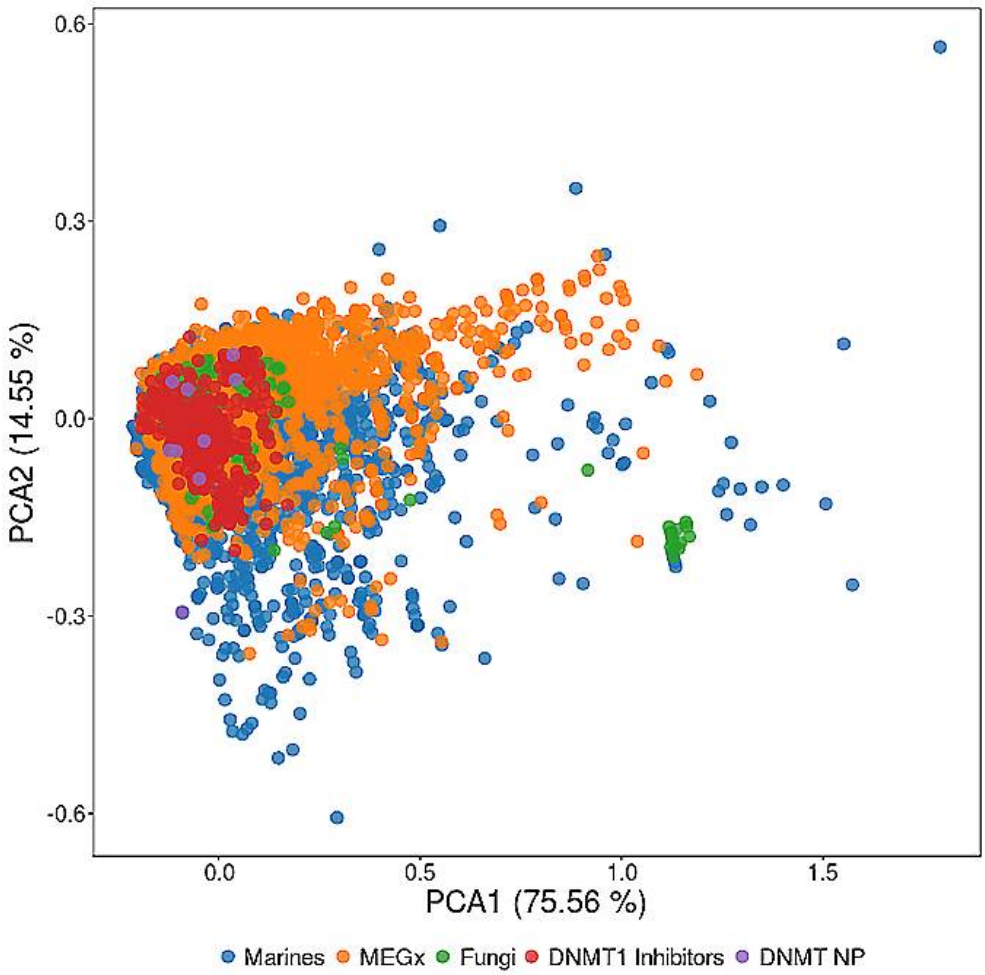


Figure 4. Visualization and comparison of the chemical space of DNMT inhibitors from natural sources (DNMT NP) vs DNMT1 inhibitors and different natural products data sets. The visual representation of the chemical space was based on principal component analysis of six physicochemical properties of pharmaceutical interest. The percentage of variance is shown on each axis of the plot.

6. Opportunities for searching for natural products as DNMT inhibitors

Most of the DNMT inhibitors from natural sources have been identified by serendipity. As discussed in section 5, the vast chemical space of natural products and food chemicals can be explored in a systematic manner using computational approaches. A classical and general approach is using virtual (or also called *in silico*) screening. The main aim of virtual screening is filtering compound data sets including large databases to select a reduced number of compounds with increased probability to show biological activity. Virtual screening has proven to be useful to identify hit compounds [51,52]. Table 1 summarizes representative case studies where virtual screening has led to the identification of active compounds with novel scaffolds. The table highlights the computational approach and the major conclusion of the study. In other published studies, *in silico* screening has uncovered compounds with potential activity but experimental validation still needs to be reported. Examples of representative virtual screening studies are further discussed in the following sections.

Peer-reviewed version available at Frontiers in Pharmacology 2018, 9, 1144; doi:10.3389/fphar.2018.01144

Table 1. Summary of virtual screening hits as inhibitors of DNMTs.

Study	*In silico* approach	Major outcome	Ref.
Structure-based screening of a lead-like subset of NP from ZINC	Cascade docking followed by a consensus approach	One computational had reported activity. Additional natural products were identified for screening.	[53]
Ligand- and structure-based screening of 800 NP	QSAR model based on linear discriminant analysis and consensus docking.	Six consensus hits were identified as potential inhibitors.	[54]
Structure-based screening of 111,121 molecules.	Docking-based screening of synthetic screening compounds.	Identification of a low micromolar hit with a novel scaffold. Further similarity searching led to the identification of two more potent hits.	[55]
Ligand-based screening of 500 compounds.	Pharmacophore-based virtual screening.	Identification of one inhibitor of DNMT1 with activity in the low micromolar range. The hit showed some selectivity vs. DNMT3B.	[56]
Structure- and ligand-based screening of 53,000 synthetic compounds.	Pharmacophore model, a Naïve Bayesian classification model, and ensemble docking.	Two compounds showed DNMT1 inhibitory activity at single but low concentration of 1 μM.	[50]

NP: natural products.

There are several published studies of virtual screening of natural products to identify DNMT inhibitors and/or demethylating agents. In an early work, Medina-Franco et al. reported the screening of a lead-like subset of natural products available in ZINC. Authors of that work implemented a multistep virtual screening approach selecting consensus hits identified from three different docking programs. One computational hit showed DNMT1 activity in a previous study. Other candidate compounds were identified for later experimental validation [53].

In a separate work, Maldonado-Rojas et al. developed a QSAR model based on linear discriminant analysis to screen 800 natural products. Hits selected were further docked with two crystallographic structures of human DNMT employing two docking programs. Six consensus hits were identified as potential inhibitors [54].

Virtual screening of synthetic libraries has also been reported to identify active compound with novel scaffolds, which are suitable for lead optimization. For instance, Chen et al. [55] reported a docking-based virtual screening of a commercial screening compound. The compound library SPECS had 111,121 compounds after filtering compounds with undesirable physicochemical properties. Results let to the identification of a compound with a novel scaffold with low micromolar IC_{50} (10.3 μM). Starting from the computational hit, similarity searching led to the identification of two more potent compounds.

Hassanzadeh et al. recently reported a pharmacophore-based virtual screening of a compound database with 500 compounds. The pharmacophore was generated using a ligand-based approach by superimposing a group of active nucleoside analogues. Selected hits, which are structurally related to the barbituric acid, were docked into the substrate binding site of DNMT1. One compound

was identified with a novel chemical scaffold that inhibits DNMT1 in the low micromolar range (IC_{50} = 4.1 μM). The compound also showed some selectivity on DNMT1 over DNMT3 enzymes [56].

Also recently Krishna et al. implemented a virtual screening protocol using several structure- and ligand-based approaches. Methods included a pharmacophore model, a Naïve Bayesian classification model, and ensemble docking. Three out of ten selected compounds from a commercial library of synthetic molecules e.g., Maybridge with 53,000 small drug-like compounds, showed DNMT1 inhibitory activity at compound concentration of 20 μM. Two of these molecules showed activity at 1 μM [50].

In addition to the studies discussed above and summarized in Table 1, the next subsections discuss other approaches that can be explored. Case studies for each strategy are outlined briefly.

6.1. Similarity-based virtual screening of NP

Similarity searching is a commonly used approach for identifying new hit compounds. Major goals are identifying starting points for later optimization or expand the SAR of analogue series. Since similarity searching is a fast approach it can be used as a first and fast approach to filter large chemical databases. Similarity searching can be conveniently used in combination with other computational approaches to refine the list of similarity searching hits, e.g., molecular docking.

Similarity searching involves two major components: a molecular representation and a similarity coefficient. In practice, one of the most common molecular representations are 2D fingerprints. Overall, a fingerprint is generally a bit vector of ceros and ones that denote the presence or absence of molecular features. In turn, one of the most common similarity coefficients is Tanimoto [57]. Full discussion of molecular representations and similarity coefficients are published elsewhere [58,59].

A novel approach to encode the chemical structures of data sets was recently developed in the so-called database fingerprint (DBFP) [60]. The rationale of DBFP is account for the most structural features encoded in bit positions of an entire data set. In principal, virtual any data set can be represented, for instance, it can be a small or large chemical database of screening compounds. Also, the data set can be a group of active compounds or molecules with a desired chemical property. DBFP has several applications including visual representation of the chemical space of large data sets [25] and similarity searching [60]. More recently, this approach was further refined into the so-called statistical based database fingerprint (SB-DFP) [61]. This approach has the same underlying idea and application of DBFP. A key improvement is the approach to account for the most relevant structural features that are derived from a statistical comparison between the structural features of a data set of interest vs. a (large) data base of reference. Further details of SB-DBFPs are provided elsewhere [61].

6.2. Pharmacophore-based

Thus far, several pharmacophore modeling studies have been conducted for inhibitors of DNMT1. Different approaches and input molecules have been used to develop these models. Most of the pharmacophore models have been employed to do a virtual screening of chemical databases and identify novel hit compounds.

In 2011 Yoo et al. reported one of the first pharmacophore models for inhibitors of DNMT1. The model was generated based on the docking poses of 14 known inhibitors available at that time. The docking was conducted with a homology model the catalytic domain of DNMT1. Of note, at the time of that study the crystallographic structure of human DNMT1 was not available. Part of the inhibitors used to develop the pharmacophore model included the natural products curcumin, parthenolide, EGCG and mahanine [62]. A year later was reported that trimethylaurintricarboxylic acid (Figure 3) showed a good agreement with this structure-based pharmacophore model. The trimethylaurintricarboxylic acid is a compound structurally related to 5,5-methylenedisalicylic acid that has an inhibition of DNMT1 in a low micromolar range (IC_{50} = 4.79 μM) [63,64].

More recently, as described in the first part of point 6, Hassanzadeh et al. developed a pharmacophore model based on a ligand-based approach by 3D superimposition of active

nucleoside analogues [56]. That model was used to do virtual screening (*vide supra*). In the same year, Krishna et al. developed, with the aid of the Hypogen module of the software DS4.1, a ligand-based pharmacophore model using the structures of 20 compounds obtained from the literature. The model was validated through the classification of an external set with known active and inactive compounds. The validated pharmacophore models were employed as part of a combined strategy to identify novel active molecules [50].

7. Conclusions

Epigenetic targets are attractive to develop therapeutic strategies. Among these targets, DNA methyltransferases is a major enzyme family that was one of the first studied epigenetic targets, in particular for the treatment of cancer. However, over the past few years, more therapeutic opportunities related to the modulation of DNMTs are emerging. Therefore, there is a growing interest in the scientific community to identify and develop small molecules that can be used as epi-drugs or epi-probes targeting DNMTs. Virtual screening is become more used in the recent years to uncover natural products as inhibitors of DNMTs and/or demethylating agents. To this end, well stablished structure- and ligand-based virtual screening approaches are being used, for example, automated docking, QSAR and similarity searching. Also, novel chemoinformatic approaches are being developed. An example of the latter is the statistical-based database fingerprint that is being used to screen collections of natural products. These and basically any other structure- and-ligand based approaches can be used in a combination to increase the probability of finding active compounds. Of course, the computational methods should be validated with rigorous *in vitro* and *in vivo* experiments to support their application.

Natural products have a well stablished history, not only in drug discovery in general but also as inhibitors of DNMTs and demethylating molecules. However, most of the active natural products have been identified by serendipity. The availability of 3D structures of DNMTs either as crystallographic molecules or homology models, in combination with *in silico* approaches, and better computational resources is boosting the systematic search for new active compounds. Moreover, the availability of natural product databases in the public domain or developed in house are of large benefit to further advance the natural product-based discovery of epi-drugs and epi-probes targeting DNMTs.

8. Perspectives

Natural products inside or even outside of the traditional drug-like chemical space represent a large promise to develop novel compounds with DNMT inhibitory activity or demethylating properties. This is because the traditional chemical space is highly represented by small molecules that over the past few years have not be very successful. A notable example in this direction is the reemergence of peptide-based drug discovery. Indeed, linear, cyclic peptides and peptidomimetics are regaining interest in drug discovery [65,66].

Other promising an emerging avenue are the modulators of protein-protein interactions (PPIs) [67]. DNMTs are known to be involved in several PPIs [67]. Modulation of such interactions can be conveniently achieved with natural products. This is because PPIs are “difficult targets” not easily addressed by small molecules from the traditional chemical space [68]). In other words, since PPIs have unique features these can be approached with novel chemical libraries but focused on a medicinally relevant chemical space. Natural products collections represent excellent candidates for this purpose [69].

Overall, it is anticipated an augmented hit and led identification based on natural products combining major technologies used in drug discovery. Such technologies involve experimental and computational approaches such as high-throughput screening, structure-, ligand-based *in silico* screening, structure-based optimization of active natural products, similarity searching of active natural products for more potent compounds or even synthetic molecules that resemble the natural product. The later strategy, that can be regarded as scaffold hopping [70], is convenient in case the core scaffolds of the natural products are too complex or too expensive to synthesize.

(for instance though scaffold hopping strategies) in case Also it is of outmost importance take into consideration the toxicity profile through the development of the natural products (and any other compound with potential therapeutic activity). Potency should not be the only criteria. Potency driven approaches are the most traditional but necessarily the most effective. Indeed, toxicity issues play a major part in the lack of success of drug discovery projects.

Author Contributions: Methodology and Formal Analysis, All; Data Curation, F.I.S-G, J.R-R, B.A.P-J.; Writing-Original Draft Preparation, A.G-G., F.I.S-G., J.L.M-F.; Writing-Review & Editing, A.G-G., F.I.S-G., J.L.M-F.; Visualization, A.G-G., F.I.S-G, BAP-J.; Project Administration, J.L.M-F.

Funding: This research was funded by *Consejo Nacional de Ciencia y Tecnologia* (CONACYT, Mexico) *Ciencia Básica* grant number 282785, by the *Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica* (PAPIIT) grant IA203718, UNAM, and by *Programa de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza* (PAPIME) grant PE200118, UNAM.

Acknowledgments: FS-G, and AG-G acknowledge *Consejo Nacional de Ciencia y Tecnologia* (CONACyT, México) for the graduate scholarships. JR-R thanks the *Programa de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza* (PAPIME) for the undergraduate scholarship. We also thank Chanachai Sae-Lee for providing the sequences used in Figure 1. Authors acknowledge all current and past members of the DIFACQUIM research group for their comments and discussions that enriched this manuscript.

Conflicts of Interest: The authors state no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

Two-dimensional	2D
Three-dimensional	3D
Database fingerprint	DBFP
DNA metiltransferases	DNMT
(-)-epigallocatechin-3-gallate	EGCG
Hydrogen bond acceptors	HDA
Hydrogen bond donors	HBD
Octanol/water partition coefficient	LogP
Molecular weight	MW
Protein-protein interactions	PPIs
Rotatable bonds	RB
S-adenosyl-*L*-homocysteine	SAH
S-adenosyl-*L*-methionine	SAM
Structure-activity relationships	SAR
Statistical based database fingerprint	SB-DFP
Topological surface area	TPSA

References

1. Berger, S.L.; Kouzarides, T.; Shiekhattar, R.; Shilatifard, A. An operational definition of epigenetics. *Genes Dev.* **2009**, *23*, 781-783.
2. Waddington, C.H. The epigenotype. *Int. J. Epidemiol.* **2012**, *41*, 10-13.
3. Dueñas-González, A.; Jesús Naveja, J.; Medina-Franco, J.L. Chapter 1 - introduction of epigenetic targets in drug discovery and current status of epi-drugs and epi-probes. In *Epi-informatics*, Academic Press: Boston, 2016; pp 1-20.
4. Lu, W.; Zhang, R.; Jiang, H.; Zhang, H.; Luo, C. Computer-aided drug design in epigenetics. *Front. Chem.* **2018**, *6*.

5. Hwang, J.-Y.; Aromolaran, K.A.; Zukin, R.S. The emerging field of epigenetics in neurodegeneration and neuroprotection. *Nat. Rev. Neurosci.* **2017**, *18*, 347.
6. Tough, D.F.; Tak, P.P.; Tarakhovsky, A.; Prinjha, R.K. Epigenetic drug discovery: Breaking through the immune barrier. *Nat. Rev. Drug Discov.* **2016**, *15*, 835.
7. Cortés-Ruiz, E.M.; Palomino-Hernández, O.; Rodríguez-Hernández, K.D.; Espinoza, B.; Medina-Franco, J.L. Computational methods to discover compounds for the treatment of chagas disease. In *Adv. Protein Chem. Struct. Biol.*, Academic Press: 2018. In press. DOI: 10.1016/bs.apcsb.2018.03.005
8. Ganesan, A. Epigenetic drug discovery: A success story for cofactor interference. *Philos. Trans. R. Soc., B* **2018**, *373*.
9. Castillo-Aguilera, O.; Depreux, P.; Halby, L.; Arimondo, P.; Goossens, L. DNA methylation targeting: The dnmt/hmt crosstalk challenge. *Biomolecules* **2017**, *7*, 3.
10. Lyko, F. The DNA methyltransferase family: A versatile toolkit for epigenetic regulation. *Nat. Rev. Genetics* **2017**, *19*, 81.
11. Ho, T.T.; Tran, Q.T.N.; Chai, C.L.L. The polypharmacology of natural products. *Future Med. Chem.* **2018**, *10*, 1361-1368.
12. Jeltsch, A. Beyond watson and crick: DNA methylation and molecular enzymology of DNA methyltransferases. *ChemBioChem* **2002**, *3*, 274-293.
13. Jurkowska, R.Z.; Jurkowski, T.P.; Jeltsch, A. Structure and function of mammalian DNA methyltransferases. *ChemBioChem* **2011**, *12*, 206-222.
14. Lan, J.; Hua, S.; He, X.N.; Zhang, Y. DNA methyltransferases and methyl-binding proteins of mammals. *Acta Biochim. Biophys. Sin.* **2010**, *42*, 243-252.
15. Zhang, Z.-M.; Liu, S.; Lin, K.; Luo, Y.; Perry, J.J.; Wang, Y.; Song, J. Crystal structure of human DNA methyltransferase 1. *J. Mol. Biol.* **2015**, *427*, 2520-2531.
16. Choi, S.H.; Heo, K.; Byun, H.-M.; An, W.; Lu, W.; Yang, A.S. Identification of preferential target sites for human DNA methyltransferases. *Nucleic Acids Res.* **2011**, *39*, 104-118.
17. Ostler, K.R.; Davis, E.M.; Payne, S.L.; Gosalia, B.B.; Expósito-Céspedes, J.; Beau, M.M.L.; Godley, L.A. Cancer cells express aberrant dnmt3b transcripts encoding truncated proteins. *Oncogene* **2007**, *26*, 5553.
18. Gordon, C.A.; Hartono, S.R.; Chédin, F. Inactive dnmt3b splice variants modulate de novo DNA methylation. *PLoS ONE* **2013**, *8*, e69486.
19. Wang, L., Wang, J., Sun, S., Rodriguez, M., Yue, P., Jang, S. J., Mao, L. A novel dnmt3b subfamily, ðdnmt3b, is the predominant form of dnmt3b in non-small cell lung cancer. *Int. J. Oncol.* **2006**, *29*, 201-207.
20. Vilkaitis, G.; Merkiene, E.; Serva, S.; Weinhold, E.; Klimasauskas, S. The mechanism of DNA cytosine-5 methylation - kinetic and mutational dissection of hhai methyltransferase. *J. Biol. Chem.* **2001**, *276*, 20924-20934.
21. Du, Q.; Wang, Z.; Schramm, V.L. Human dnmt1 transition state structure. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 2916-2921.
22. Klimasauskas, S.; Kumar, S.; Roberts, R.J.; Cheng, X.D. Hhal methyltransferase flips its target base out of the DNA helix. *Cell* **1994**, *76*, 357-369.
23. Fernandez-de Gortari, E.; Medina-Franco, J.L. Epigenetic relevant chemical space: A chemoinformatic characterization of inhibitors of DNA methyltransferases. *RSC Adv.* **2015**, *5*, 87465-87476.
24. Palomino-Hernández, O.; Medina-Franco, J.L. Comparative cheminformatic analysis of inhibitors of DNA methyltransferases. *Chem. Inf.* **2017**, *3*, 1-10.

25. Naveja, J.J.; Medina-Franco, J.L. Insights from pharmacological similarity of epigenetic targets in epipolypharmacology. *Drug Discovery Today* **2018**, *23*, 141-150.
26. Naveja, J.J.; Medina-Franco, J.L. Activity landscape sweeping: Insights into the mechanism of inhibition and optimization of dnmt1 inhibitors. *RSC Adv.* **2015**, *5*, 63882-63895.
27. Naveja, J.J.; Medina-Franco, J.L. Activity landscape of DNA methyltransferase inhibitors bridges chemoinformatics with epigenetic drug discovery. *Exp.Opin.Drug Discov.* **2015**, *10*, 1059-1070.
28. 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*, 3253-3261.
29. Medina-Franco, J.L.; Méndez-Lucio, O.; Yoo, J.; Dueñas, A. Discovery and development of DNA methyltransferase inhibitors using in silico approaches. *Drug Discovery Today* **2015**, *20*, 569-577.
30. 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*, 2321-2325.
31. Kabro, A.; Lachance, H.; Marcoux-Archambault, I.; Perrier, V.; Dore, V.; Gros, C.; Masson, V.; Gregoire, J.M.; Ausseil, F.; Cheishvili, D., *et al.* Preparation of phenylethylbenzamide derivatives as modulators of dnmt3 activity. *MedChemComm* **2013**, *4*, 1562-1570.
32. Davide, G.; Sandra, A.; Emily, B.; Mattia, C.; Marta, G.; Annalisa, C.; Livio, S.; Gianluca, M.; Chiara, C.; Eli, F.-d.G., *et al.* Design and synthesis of n-benzoyl amino acid derivatives as DNA methylation inhibitors. *Chem. Bio. Drug Des.* **2016**, *88*, 664-676.
33. Clemens, Z.; Sergio, V.; Antonello, M. DNA methyltransferases inhibitors from natural sources. *Curr. Top. Med. Chem.* **2016**, *16*, 680-696.
34. Naveja, J.J.; Rico-Hidalgo, M.P.; Medina-Franco, J.L. Analysis of a large food chemical database: Chemical space, diversity, and complexity. *F1000Research* **2018**, 7(Chem Inf Sci), 993.
35. Andrea Peña-Castillo; Oscar Méndez-Lucio; John R. Owen; Karina Martínez-Mayorga; Medina-Franco, J.L. Chemoinformatics in food science. In *Applied chemoinformatics*, Engel, T.; Gasteiger, J., Eds. Wiler: 2018.
36. Martinez-Mayorga, K.; Medina-Franco, J.L. *Foodinformatics: Applications of chemical information to food chemistry*. Springer: New York, 2014; p 306.
37. Singh, N.; Guha, R.; Giulianotti, M.A.; Pinilla, C.; Houghten, R.A.; Medina-Franco, J.L. Chemoinformatic analysis of combinatorial libraries, drugs, natural products, and molecular libraries small molecule repository. *J. Chem. Inf. Model.* **2009**, *49*, 1010-1024.
38. López-Vallejo, F.; Giulianotti, M.A.; Houghten, R.A.; Medina-Franco, J.L. Expanding the medicinally relevant chemical space with compound libraries. *Drug Discovery Today* **2012**, *17*, 718-726.
39. González-Medina, M.; Prieto-Martínez, F.D.; Naveja, J.J.; Méndez-Lucio, O.; El-Elimat, T.; Pearce, C.J.; Oberlies, N.H.; Figueroa, M.; Medina-Franco, J.L. Chemoinformatic expedition of the chemical space of fungal products. *Future Med. Chem.* **2016**, *8*, 1399-1412.
40. Olmedo, D.A.; González-Medina, M.; Gupta, M.P.; Medina-Franco, J.L. Cheminformatic characterization of natural products from panama. *Mol. Diversity* **2017**, *21*, 779-789.
41. Shang, J.; Hu, B.; Wang, J.; Zhu, F.; Kang, Y.; Li, D.; Sun, H.; Kong, D.-X.; Hou, T. Cheminformatic insight into the differences between terrestrial and marine originated natural products. *J. Chem. Inf. Model.* **2018**, *58*, 1182-1193.
42. Medina-Franco, J.L. Discovery and development of lead compounds from natural sources using computational approaches. In *Evidence-based validation of herbal medicine*, Mukherjee, P., Ed. Elsevier: 2015; pp 455-475.

43. Chen, C.Y.-C. Tcm database@taiwan: The world's largest traditional chinese medicine database for drug screening in silico. *PLoS One* **2011**, *6*, e15939.
44. Pilon, A.C.; Valli, M.; Dametto, A.C.; Pinto, M.E.F.; Freire, R.T.; Castro-Gamboa, I.; Andricopulo, A.D.; Bolzani, V.S. Nubbedb: An updated database to uncover chemical and biological information from brazilian biodiversity. *Sci. Rep.* **2017**, *7*, 7215.
45. Ntie-Kang, F.; Zofou, D.; Babiaka, S.B.; Meudom, R.; Scharfe, M.; Lifongo, L.L.; Mbah, J.A.; Mbaze, L.M.; Sippl, W.; Efange, S.M.N. Afrodb: A select highly potent and diverse natural product library from African medicinal plants. *PLoS One* **2013**, *8*, e78085.
46. Rosen, J.; Lovgren, A.; Kogej, T.; Muresan, S.; Gottfries, J.; Backlund, A. ChemGPS-NPweb: Chemical space navigation online. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 253-259.
47. Chen, Y.; de Bruyn Kops, C.; Kirchmair, J. Data resources for the computer-guided discovery of bioactive natural products. *J. Chem. Inf. Model.* **2017**, *57*, 2099-2111.
48. Liu, Z.; Du, J.; Yan, X.; Zhong, J.; Cui, L.; Lin, J.; Zeng, L.; Ding, P.; Chen, P.; Zhou, X., *et al.* Tcmanalyzer: A chemo- and bioinformatics web service for analyzing traditional chinese medicine. *J. Chem. Inf. Model.* **2018**, *58*, 550-555.
49. Gonzalez-Medina, M.; Naveja, J.J.; Sanchez-Cruz, N.; Medina-Franco, J.L. Open chemoinformatic resources to explore the structure, properties and chemical space of molecules. *RSC Adv.* **2017**, *7*, 54153-54163.
50. Krishna, S.; Shukla, S.; Lakra, A.D.; Meeran, S.M.; Siddiqi, M.I. Identification of potent inhibitors of DNA methyltransferase 1 (dnmt1) through a pharmacophore-based virtual screening approach. *J. Mol. Graphics Modell.* **2017**, *75*, 174-188.
51. Lavecchia, A.; Di Giovanni, C. Virtual screening strategies in drug discovery: A critical review. *Curr. Med. Chem.* **2013**, *20*, 2839-2860.
52. Clark, D.E. What has virtual screening ever done for drug discovery? *Expert Opin. Drug Discovery* **2008**, *3*, 841-851.
53. Medina-Franco, J.L.; López-Vallejo, F.; Kuck, D.; Lyko, F. Natural products as DNA methyltransferase inhibitors: A computer-aided discovery approach. *Mol. Diversity* **2011**, *15*, 293-304.
54. Maldonado-Rojas, W.; Olivero-Verbel, J.; Marrero-Ponce, Y. Computational fishing of new DNA methyltransferase inhibitorsfrom natural products. *J. Mol. Graphics Modell.* **2015**, *60*, 43-54.
55. Chen, S.J.; Wang, Y.L.; Zhou, W.; Li, S.S.; Peng, J.L.; Shi, Z.; Hu, J.C.; Liu, Y.C.; Ding, H.; Lin, Y.Y., *et al.* Identifying novel selective non-nucleoside DNA methyltransferase 1 inhibitors through docking-based virtual screening. *J. Med. Chem.* **2014**, *57*, 9028-9041.
56. Malihe, H.; Rustem, K.; Shabnam, M.; Mehdi, A.; Max, E.; Michael, D.; Pavel, B.; Albert, J.; Massoud, A. Discovery of novel and selective DNA methyltransferase 1 inhibitors by pharmacophore and docking-based virtual screening. *ChemistrySelect* **2017**, *2*, 8383-8392.
57. Bajusz, D.; Rácz, A.; Héberger, K. Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminf.* **2015**, *7*, 1-13.
58. Willett, P.; Barnard, J.; Downs, G. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*.
59. Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular similarity in medicinal chemistry. *J. Med. Chem.* **2014**, *57*, 3186-3204.
60. Fernández-de Gortari, E.; García-Jacas, C.R.; Martinez-Mayorga, K.; Medina-Franco, J.L. Database fingerprint (dfp): An approach to represent molecular databases. *J. Cheminf.* **2017**, *9*, 9.

61. Sánchez-Cruz, N.; Medina-Franco José, L. Statistical-based database fingerprint: Application in ligand-based virtiual screening. Cinf-153. In *256th ACS National Meeting, Boston, MA, United States*, American Chemical Society, Washington, D. C: Boston, MA, United States, 2018.
62. Yoo, J.; Medina-Franco, J.L. Homology modeling, docking, and structure-based pharmacophore of inhibitors of DNA methyltransferase. *J. Comp.-Aided Mol. Des.* **2011**, *25*, 555-567.
63. 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. *Adv. Protein Chem. Struct. Biol.* **2012**, *87*, 219-247.
64. Yoo, J.; Medina-Franco, J.L. Trimethylaurintricarboxylic acid inhibits human DNA methyltransferase 1: Insights from enzymatic and molecular modeling studies. *J. Mol. Model.* **2012**, *18*, 1583-1589.
65. Henninot, A.; Collins, J.C.; Nuss, J.M. The current state of peptide drug discovery: Back to the future? *J. Med. Chem.* **2018**, *61*, 1382-1414.
66. Fosgerau, K.; Hoffmann, T. Peptide therapeutics: Current status and future directions. *Drug Discovery Today* **2015**, *20*, 122-128.
67. Díaz-Eufracio, B.I.; Naveja, J.J.; Medina-Franco, J.L. Chapter three - protein–protein interaction modulators for epigenetic therapies. In *Adv. Protein Chem. Struct. Biol.*, Donev, R., Ed. Academic Press: 2018; Vol. 110, pp 65-84.
68. O., V.B.; A., K.M.; Jean-Luc, P.; Heriberto, B.-G.; Céline, L.; David, L.; Olivier, S.; A., M.M. Drug-like protein protein interaction modulators: Challenges and opportunities for drug discovery and chemical biology. *Mol. Inf.* **2014**, *33*, 414-437.
69. Medina-Franco, J.L.; Martinez-Mayorga, K.; Meurice, N. Balancing novelty with confined chemical space in modern drug discovery. *Expert Opin. Drug Discovery* **2014**, *9*, 151-165.
70. Sun, H.M.; Tawa, G.; Wallqvist, A. Classification of scaffold-hopping approaches. *Drug Discovery Today* **2012**, *17*, 310-324.