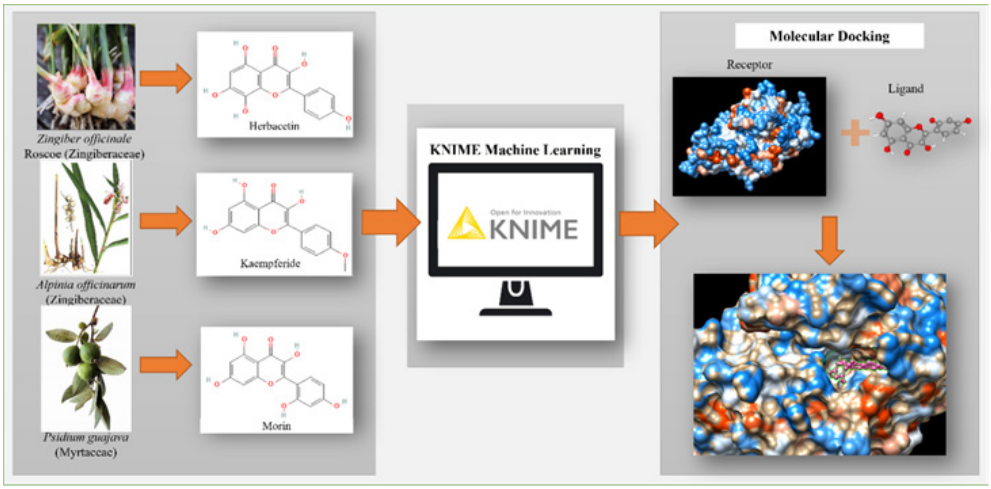
RESEARCH ARTICLE

Computer-aided drug design to discover DNMT inhibitors from phytochemicals

L. R. L. S. Kumari and W. R. P. Wijesinghe*



Highlights

- Epigenetic therapy of cancer faces limited drug options, with only few approved drugs, thus highlighting the need for novel drugs.
- The study utilizes modified TeachOpenCADD KNIME workflows to search plant molecules structurally similar to existing DNMT inhibitors.
- Through machine learning (ML) in KNIME, eight phytochemicals identified as potential DNMT inhibitors, and subsequent molecular docking reveals Herbacetin, Kaempferide, and Morin as promising compounds with anti-DNMT activity.

Ceylon Journal of Science 53 (2) 2024: 275-290
DOI: http://doi.org/10.4038/cjs.v53i2.8262

RESEARCH ARTICLE

Computer-aided drug design to discover DNMT inhibitors from phytochemicals

L. R. L. S. Kumari [1] and W. R. P. Wijesinghe [1,2*]

[1] *Department of Botany, Faculty of Science, University of Peradeniya, 20400, Sri Lanka*
[2] *Postgraduate Institute of Science, University of Peradeniya, 20400, Sri Lanka*

Received: 05.11.2023; Accepted: 18.05.2024

Abstract: Inhibitors of DNA methyltransferase (DNMTs) are now a major family of epigenetic targets with therapeutic interest. However, only two cytosine analogues 5-azacytosine (azacytidine) and 20-deoxy-5-azacytidine (decitabine), have been approved as the most cutting-edge medications for treating epigenetic cancer with some restrictions. In this context, computational methods that rely on quantitative structure-activity relationship (QSAR) play a crucial role allowing us to predict the biological activity of potential molecules based on the theoretically calculated physicochemical properties of these compounds. When coupled with machine learning (ML), QSAR approaches create an ideal platform for discovering potential drug candidates. In this study, three Machine Learning (ML) models; Random Forest, Support Vector Machine, and Artificial Neural Network, were trained using modified TeachOpenCADD KNIME workflows and applied it to the identification of plant molecules that are structurally similar to the active pharmaceuticals of current DNMT inhibitors. Then molecular docking simulations were performed using AutoDock Vina, employing two human DNMT structures (PDB codes: 4WXX and 2QRV) as target proteins and the predicted phytochemicals as ligands. Additionally, we focused on the R882H mutation hotspot in the catalytic domain of DNMT3A, which is associated with aberrant DNA methylation in acute myeloid leukemia (AML). Consequently, the structure of R882H DNMT3A (PDB code: 6W8J) was docked with the identified novel ligands. As a result of our computational analysis, eight phytochemicals were predicted as potential DNMT inhibitors through the ML approaches from KNIME. Subsequently, three of these phytochemicals, namely Herbacetin, Kaempferide, and Morin were identified as virtual hits against DNMTs following the molecular docking simulations. Overall, our study demonstrates the effectiveness of this computational strategy in identifying DNMT inhibitors. These findings hold promise for the discovery of potent and selective anticancer drugs targeting DNMTs.

***Keywords*:** DNMT inhibitors; Docking; KNIME**;** Machine Learning; Phytochemicals

INTRODUCTION

DNA methyltransferases (DNMTs) are epigenetic enzymes that methylate cytosine at the C5 position, playing a vital role in cell differentiation and development. (Cheng & Blumenthal, 2008; Goll & Bestor, 2005; Hermann et al., 2004). The three major DNMTs involved in this process are DNMT1, DNMT3A, and DNMT3B. DNMT1 is primarily considered a maintenance DNA methyltransferase responsible for maintaining CpG methylation patterns, playing important roles in embryonic development and the survival of somatic cells (Li et al., 2016). DNMT3A and DNMT3B are classified as *de novo* methyltransferases that play essential roles in establishing DNA methylation patterns during gametogenesis and early development, contributing to the establishment of embryonic methylation patterns (Okano et al., 1999). DNMTs are large multi-domain proteins, consisting of a catalytic C-terminal methyltransferase domain responsible for methylating DNA and a complex N-terminal part containing diverse targeting and regulatory functions (Cheng & Blumenthal, 2008). DNMTs recognize flipped-out cytosines within double-stranded DNA and operate via the nucleophilic attack mechanism (Jones & Liang, 2009). In this mechanism, the unstable methyl group from S-adenosylmethionine (SAM) is transferred to the C-5 atom of cytosine, leading to the formation of 5-methylcytosine (Pfaffeneder et al., 2011). De-regulation of the DNMTs has been shown in many types of cancer including the lung, breast, stomach, and colon, as well as in leukemia (Chik et al., 2011; Chik & Szyf, 2010; Gnyszka et al., 2013). In cancer cells, DNMTs can be overexpressed, leading to hypermethylation of tumor-suppressor genes, silencing their expression and promoting tumor growth. Conversely, DNMTs can also be downregulated, resulting in global hypomethylation and genomic instability (Delpu et al., 2013).

One of the benefits of epigenetic alterations, in contrast to genetic mutations, is their potential reversibility (Esteller, 2011). Certain drugs like azacytidine and decitabine have been developed as epigenetic therapies, targeting DNMTs to reverse abnormal DNA methylation patterns (Jones & Taylor, 1980). These drugs work by inhibiting DNMT activity, leading to the reactivation of silenced genes and restoring normal cellular functions. In addition to synthetic drugs, natural products, such as phytochemicals derived from plants, have also been identified as DNMT inhibitors (Saldívar-gonzález et al., 2018). These natural compounds offer the advantage of being relatively more readily available and often exhibit lower toxicity compared to synthetic compounds, making them promising candidates for developing DNMT-targeted therapies with potentially fewer side effects (Mund et al., 2016; Malongane et al., 2017). The discovery of such agents provides opportunities for developing novel epigenetic-based treatments and advancing the field of personalized medicine.

**Corresponding Author's Email: priyangaw@sci.pdn.ac.lk*
https://orcid.org/0000-0002-7736-5038

For decades, the process of natural product-based drug discovery involved a trial-and-error approach, where compounds from natural sources were isolated and tested *in vitro* and animal models to determine their efficacy against specific diseases, resulting in a lengthy and costly process to bring new drugs to market. A recent report showcased the development of a drug molecule using a computer-aided drug discovery approach, resulting in significant time savings of nearly a decade in the drug discovery process (Jarada et al., 2020). Such computational approaches when coupled with advances in genomics, proteomics, and metabolomics approaches as well as advanced cheminformatics applications such as theoretical quantitative structure-activity relationship (QSAR) calculations have the potential to revolutionize the pharmaceutical drug discovery sciences (Chakravarti & Alla, 2019).

In recent years advances in computational approaches and cheminformatics applications have resulted in large collections of molecular data being available in publicly accessible databases. Accessing and undertaking *in silico* experimentation with these vast libraries of data is now possible due to open-source cheminformatics libraries such as CDK, ChemmineR, RDKit, OpenChem, etc. However, the use of these software requires a significantly high level of coding knowledge in Python, R, C++, or Java. This creates a barrier to entry for experts highly knowledgeable in plant sciences, who do not have a computer science background. To overcome this limitation, the platform is referred to as Konstanz Information Miner (KNIME) allows the creation of modular and yet very powerful data mining workflows using a visual programming approach.

This study examines the effectiveness of phytochemicals in inhibiting DNMTs and employs computational methods to discover new potential inhibitors of DNMTs from a personally curated phytochemical database containing over 400 natural compounds. To the best of our knowledge, this is the first investigation to systematically screen a diverse collection of natural products for DNMT inhibitors using the KNIME analytic platform. Predicted drug-like compounds exhibiting favourable DNMTs binding traits and possessing diverse chemical scaffolds were identified through a docking-based virtual approach utilizing the AutoDock Vina program (Trott & Olson, 2010).

MATERIALS AND METHODS

Data collection, training and evaluation ML models from KNIME

KNIME Analytics Platform v4.5.2 (Fillbrunn et al., 2017) was downloaded and then cheminformatics extension (Roughley, 2018) was added to KNIME software.

Simplified Molecular Input Line Entry System structural data (SMILES notation) of 477 molecules from the ChEMBL database was retrieved in JSON format by using the 'GET Request' KNIME node. Of total drug molecules 30 drug molecules were used as DNMT inhibitors, 408 molecules are phytochemicals, and rest of the molecules, 39 drugs are rheumatoid arthritis as already known nonactive drugs for DNMT inhibitors.

Canonical SMILES of molecules were converted to RDKit molecule objects by the 'RDKit from Molecule' KNIME node. Afterward, pre-defined structural features of the FeatMorgan fingerprint type (1024-bits) were calculated for each drug molecule by the 'RDKit Fingerprint' KNIME node to prepare the data set for ML.

By following the modified TeachOpenCADD KNIME workflow (KNIME Hub: https://hub.knime.com/volkamerlab/space/TeachOpenCADD) (Sydow & Volkamer, 2019), three types of ML models, Artificial Neural Network (ANN: Multi-layer Feed Forward Network), Random Forest (RF), and Support Vector Machine (SVM) were trained on ChEMBL compounds to discriminate active and inactive compounds against DNMTs with 10 number of validations.

Performance of the trained ML algorithms was tested using 'Scorer' and 'ROC curve' KNIME node. Overall statistics and confusion matrix was calculated by the Scorer node.

Drug likeness test from KNIME

Predicted molecules were filtered with the help of Lipinski's rules of 5 to distinguish between drug-like and non-drug-like molecules with 18 KNIME nodes. It predicts a high probability of success or failure due to drug-likeness for molecules complying with 3 or more of the following rules by calculating molecular weight ≤ 500, number of hydrogen bond donors ≤ 5, number of hydrogen bond acceptors ≤ 10, molar refractivity (SMR) in between 40 to 130 and log P value ≤ 5 (Lipinski, 2004).

Docking-based virtual screening

Chimera v1.16 (Pettersen et al., 2004) along with Autodock vina v1.1.2 (Eberhardt et al., 2021) was used to dock compounds (Trott & Olson, 2010). 3D structures of 8 compounds that were predicted as DNMT inhibitors from self-prepared machine learning KNIME workflow, were optimized with Tripos force field with Gasteiger charges and docked on two human DNMT structures [PDB codes: 4WXX (Zhang et al., 2015) and 2QRV (Jia et al., 2007)] using AutoDock Vina, prioritizing according to their value of affinity (lower than - 6 kcal/mol), hydrogen bond count, root-mean-square deviation of atomic positions (RMSD, Å) and number of active torsions. Also, the R882H DNMT3A is a mutation hotspot in catalytic domain of DNMT3A causing aberrant DNA methylation in acute myeloid leukemia (AML) (Anteneh et al., 2020). Therefore, R882H DNMT3A structure [PDB code: 6W8J (Anteneh et al., 2020)] was docked with those found as novel ligands. All the protein structures were downloaded from Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/). The x, y, and z coordinates for the centre grid boxes on A chain of 4WXX were -47.0, -60.0 and 7.0, whereas for E chain of 2QRV were 63.0, -16.0 and -2.0, and A chain of 6W8J were 165.0, -146.0 and 16.0 respectively. There is one missing loop in 2QRV E chain and 5 missing regions in 4WXX, which are far away from the active site and no missing loop in 6W8J A chain. Three docking runs were performed for each ligand, and the pose with the highest absolute value of affinity was saved. Finally, the binding affinity value for a

specific complex was determined using the mean affinity value for the optimal pose. UCSF Chimera was used for the graphical visualisation (Pettersen et al., 2004).

RESULTS

The entire KNIME workflow for training and evaluating the models contains 60 KNIME nodes (Figure 1). The model is capable of identifying molecules that have similar structures to drugs that are used to treat as DNMT inhibitors with an accuracy of 95.81% in RF and 95.39% in ANN, and 94.13% in SVM. Eight phytochemicals (Allicin, Betaine, Citric acid, Dehydrocostuslactone, Herbacetin, Kaempferide, Morin, Pyrogallol) were predicted as DNMT inhibitor in ANN (Table S1a). Although the already known DNMT drugs were predicted from the SVM and RF models in the KNIME workflow, none of phytochemicals were identified as DNMT inhibitors.

Drug likeness test results can be observed bellow (Table 1, S1b). According to Lipinski's rules of 5, all phytochemicals were predicted as successful oral drugs by complying 3 or more rules of Lipinski's (Figure 2).

Molecular docking validation performed by re-docked of Azacitidine on their complex with DNMTs (4WXX, 2QRV, and 6W8J, respectively), revealed an optimal reproduction of the predicted poses compared with experimental binding mode for these ligands (co-crystalized ligand), with satisfactory results with AutoDock Vina. Obtained results were in agreement with similar studies conducted for this purpose (Robert et al., 2006), reported the best RMSD value >2-5 Å.

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) which are reported in supplementary file Table S2.

Molecular docking of 8 compounds predicted as active compounds against DNMTs by KNIME ML model, were evaluated by molecular docking with DNMT1 (4WXX) and DNMT3A (2QRV & 6W8J). As shown in table 2, three phytochemicals were selected according to the obtained affinity values less than -6.0 kcal/mol (Table 2, S2). Also, these 3 compounds showed > 2-5 Å RMSD values, and 1-3 hydrogen bonds.

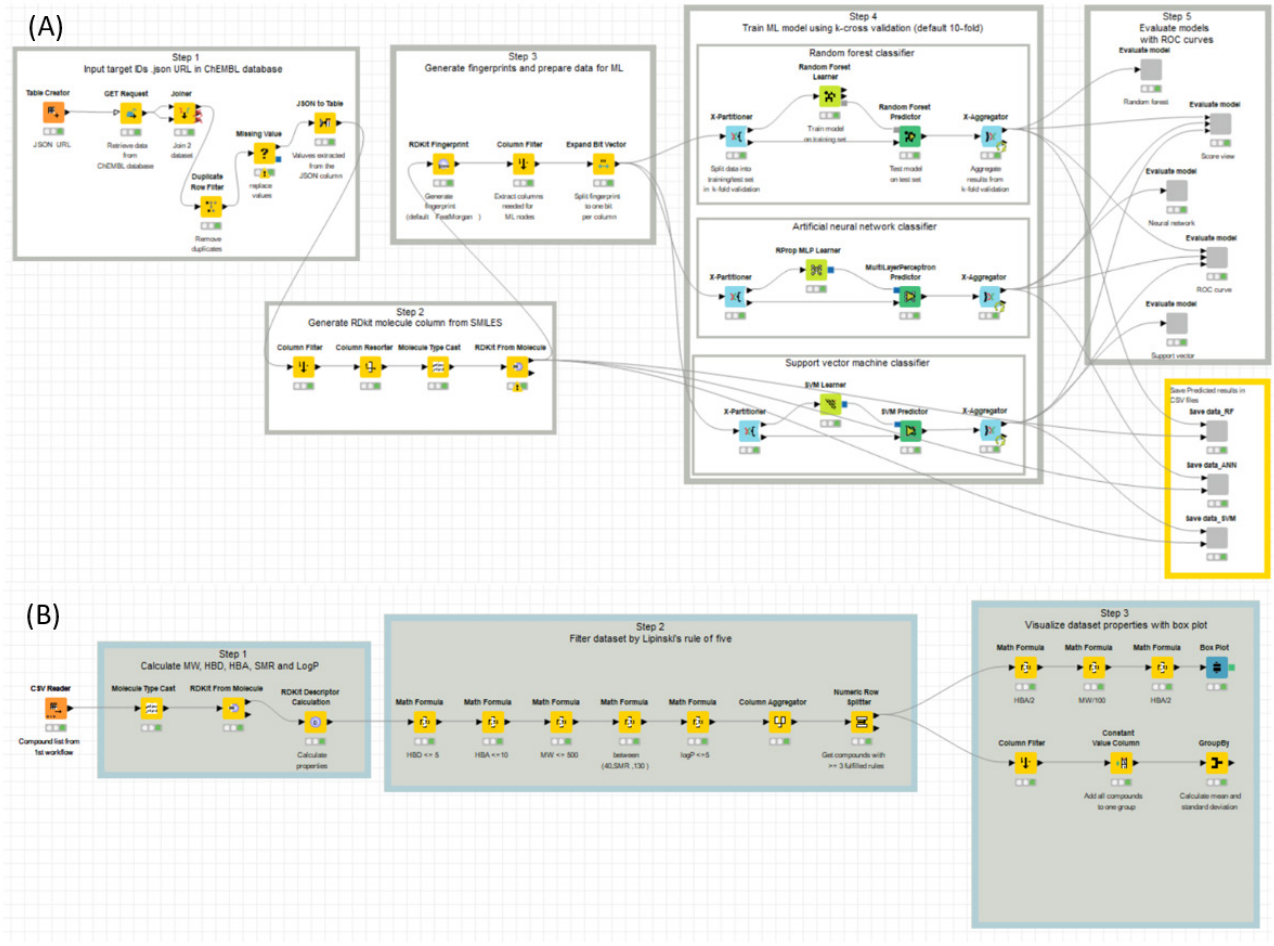


Figure 1: The graphical interface of KNIME, demonstrating structure-based screening of DNMT inhibitors from machine learning (A) and molecular filtering from Lipinski Rules of Five to show drug likeness test (B). Every node and metanode in the workflow are tagged with a concise topic description and outlines the primary steps within the workflow.

Table 1: The analysis results of Lipinski rules of five

Information	**High lipophilicity (Slog P)**	**Hydrogen bond donors (NumHBD)**	**Hydrogen bond acceptors (NumHBA)**	**Molecular Mass (ExactMW)**	**Molar refractivity (SMR)**
Rules	Less than five	Less than five	Less than ten	Less than 500	should be between 40-130
Allicin	1.7553	0	2	162.0173	45.8614
Betaine	-0.2228	1	1	118.0863	30.5352
Citric acid	-1.2485	4	4	192.027	37.0912
Dehydrocostus lactone	3.0166	0	2	230.1307	66.238
Herbacetin	1.988	5	7	302.0427	76.244
Kaempferide	2.5854	3	6	300.0634	79.4664
Morin	1.988	5	7	302.0427	76.244
Pyrogallol	0.8034	3	3	126.0317	31.4364

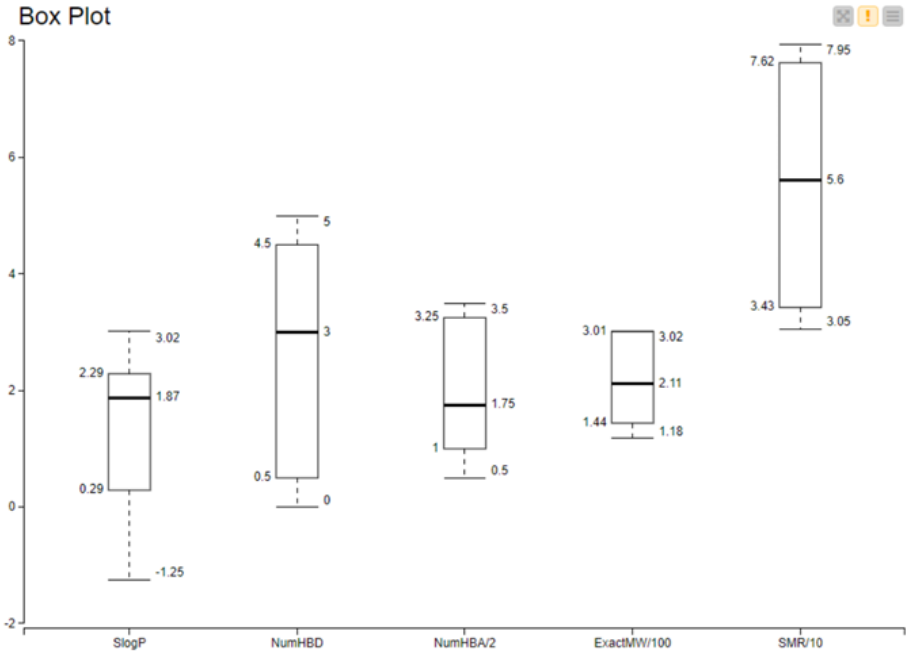


Figure 2: Visualization of dataset properties that adhere to three or more of the following five criteria in a drug likeness test: (1) log P value (Slog P) ≤ 5; (2) number of hydrogen bond donors (NumHBD) ≤ 5; (3) number of hydrogen bond acceptors (NumHBA) ≤ 10; (4) molecular weight (ExactMW) ≤ 500; (5) molar refractivity (SMR) in between 40 to 130. NumHBA, ExactMW and SMR were devided by 2, 100 and 10 respectively to get more distinct visual representation.

Table 2: Molecular docking results showing affinity value using AutoDock Vina

Compound	4WXX (DNMT1) (kcal/mol)	2QRV (E chain of DNMT3A) (kcal/mol)	6W8J (A chain of DNMT3A mutated) (kcal/mol)	Activity Status (threshold value -6 kcal/mol)
Allicin	0.0	-4.4	-4.1	Inactive
Betaine	0.0	-3.8	-3.5	Inactive
Citric acid	0.0	-6.1	-5.7	Inactive
Dehydrocostus lactone*	-6.7	-8.0	-4.7	Inactive
Herbacetin	-7.9	-7.8	-7.9	Active
Kaempferide	-8.4	-7.5	-7.2	Active
Morin	-8.6	-8.0	-9.0	Active
Pyrogallol	0.0	-5.9	-5.4	Inactive

*No active torsions

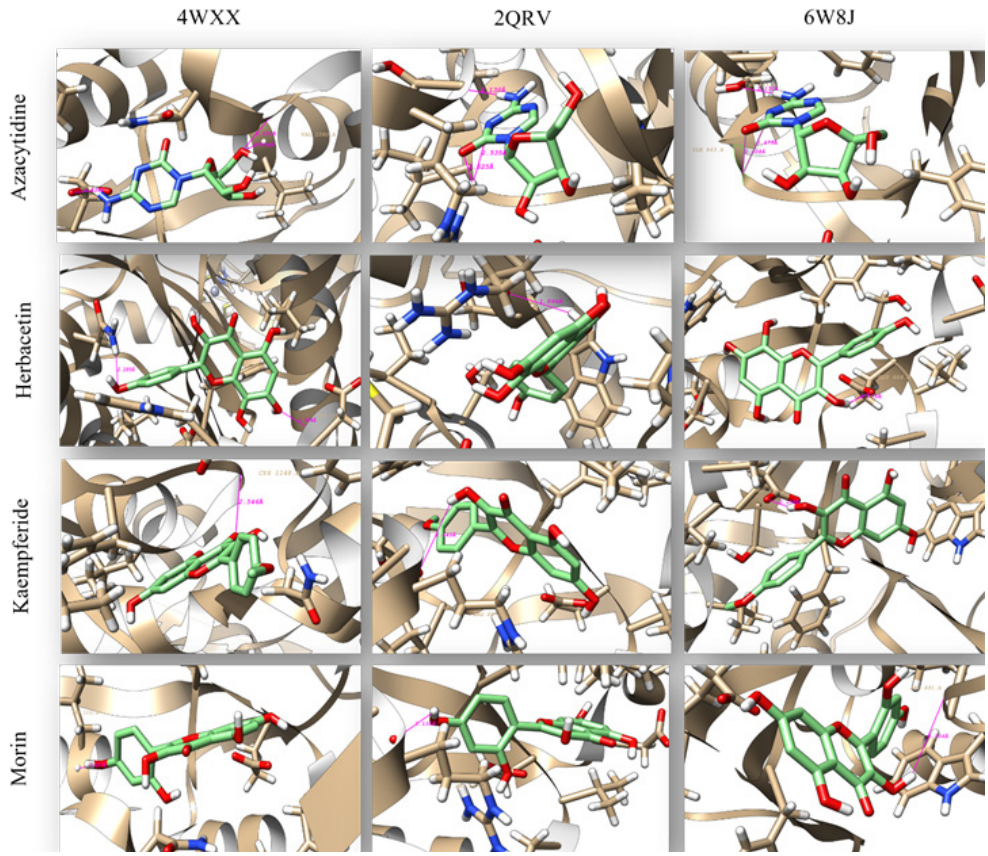


Figure 3: Molecular docking interactions involving the compounds Azacytidine, Herbacetin, Kaempferide, and Morin with three different protein structures: 4WXX (DNMT1), 2QRV (E chain of DNMT3A), and 6W8J (A chain of R882H DNMT3A). The ligand and residues in the active site are shown in stick representation. The purple sticks in the figure illustrate the presence of hydrogen bonds established between these ligands and specific amino acids within the catalytic pocket of the respective proteins.

DISCUSSION

In this study, we built a modified ML model with three types of ML algorithms and the model can predict the structural similarity with > 94% accuracy. This KNIME workflow is the initial step of our drug discovery pipeline to identify potential drug candidates from plant molecules. We harnessed the power of SMILES notation for representing molecular data, a widely accepted and efficient method. Our data sourcing involved retrieving information from the ChEMBL database, a high ranking resource, curated by experts. This database provides access to an extensive array of compounds with comprehensively documented biological activities, in line with the established standard procedure within the field for obtaining bioactivity data. (Gaulton et al., 2012). We methodically selected three distinct categories: DNMT inhibitors, phytochemicals, and rheumatoid arthritis drugs (those lacking DNMT inhibitory activity). This classification scheme played a crucial role in forming a well-balanced dataset, which is essential for the effectiveness of robust machine learning. FeatMorgan fingerprint generation is a reasonable choice for molecular feature extraction with high number of bits . Nevertheless, we acknowledge that alternative fingerprinting techniques such as MACCS or Morgan fingerprints are worth exploring in future studies (Barta, 2016).

ML approaches are a highly effective method to lower the cost and time in drug discovery methodology and the current study was based on identifying plant molecules that have similar structures to DNMT inhibitors available in the drug industry. ANN, Random Forest, and SVM are widely used for cheminformatics tasks (Tropsha, 2010). Also, 10-fold cross-validation is a common practice in machine learning model evaluation (Chicco, 2017). Evaluation metrics such as accuracy, precision, recall, and F1-score are commonly used in cheminformatics (Basha et al., 2019; Russo et al., 2018). Here we used scorer (JavaScript) node to show the evaluation matrix. ROC curve analysis provides a comprehensive view of classifier performance (Fawcett, 2006). Once these molecules were identified, another method was applied to check the drug's likeness. Lipinski's Rule of Five is a widely accepted guideline for drug-likeness (Lipinski, 2004).

Chimera is a widely used molecular visualization tool (Pettersen et al., 2004). AutoDock Vina is a popular docking software known for its accuracy and efficiency (Trott & Olson, 2010). DNMT1, DNMT3A, and DNMT3B, play essential roles in DNA methylation processes (Hermann et al., 2004). This study is explaining how the docking simulations with DNMT structures, including those with mutations, can shed light on potential novel ligands or

compounds that may target the active sites or allosteric regions of DNMTs. DNMT3A mutations, especially the R882H mutation, have gained considerable attention in the context of AML. This mutation is associated with abnormal DNA methylation patterns and is considered as a driver mutation in AML pathogenesis (Anteneh et al., 2020) and contributes to epigenetic dysregulation and are implicated in disease progression (Shih et al., 2012). To date, scientists have investigated the use of xenogeneic hematopoietic stem cell transplantation (Xu et al., 2015) and chemotherapy (Ayala et al., 2021; Döhner et al., 2018) as potential treatments for individuals with AML who have the DNMT3A R882 mutation. Certain research findings have indicated that a significant enhancement in survival rates among these patients can be achieved solely with a high dosage of daunorubicin (Luskin et al., 2016). Additionally, such elevated dosages have been shown to stimulate side effects (Lebaron et al., 1988).

Although a number of studies for the discovery of DNMTs have been supported by methods of molecular docking, this report is the first where two different approches, ML and AutoDock Vina are used to determine the structural similarity and interaction feasibility between compounds and DNMTs, applying the diversity of criteria for selection of phytochemicals as new DNMT inhibitors.

From chemical structures of predicted molecules which are depicted in Figure 4, Herbacetin is a naturally occurring flavonoid compound that can be found in sources like *Ephedrae herba* (Koyama et al., 2021) and various other plants. Herbacetin is particularly renowned for its potent antioxidant capabilities and its ability to combat tumors in breast, colon, and skin tissues (Kim et al., 2017; Kim et al., 2016).

Kaempferide, an O-methylated flavonol, is closely related to Kaempferol, which serves as the precursor for Kaempferide. The main distinction between them lies in a monomethoxy substitution on the 4th position of the B ring in Kaempferide. Kaempferide were widely found in many plants (Lai et al., 2007). Notably, Kaempferide has shown superior pharmacokinetic properties compared to various other flavonoids, including Kaempferol, as demonstrated in a study by Jiang et al., 2018. Researchers have investigated the anticancer potential of Kaempferide against a range of cancer types *in vitro*. These include cervical cancer (Nath et al., 2015), breast cancer (Yusuf et al., 2021), lung cancer (Li et al., 2020), and colon cancer (Chen et al., 2021).

Morin is a polyphenolic flavonol compound primarily found in plant families like Moraceae, Rosaceae, and Fagaceae, as indicated by Solairaja et al., in 2021. It has been shown to act as a potent cell proliferation inhibitor for human leukemia (Kuo et al., 2007), anti-tumor promotion effect by significantly inhibiting skin tumor promotion (Iwase et al., 2001). Morin also acts as a chemopreventive agent against oral carcinogenesis, *in vitro* and *in vivo* (Kawabata et al., 1999). To the best of our knowledge, while Herbacetin, Kaempferide, and Morin anticancer properties have been explored in a limited number of other cancer types, its impact on DNMT has not been documented yet.

A wide array of secondary metabolites, particularly flavonoids, are present in significant quantities in a variety of medicinal plants, fruits, and vegetables. These compounds offer numerous health benefits, such as acting as antioxidants and having the potential to combat inflammation and tumors (Grotewold, 2006). Flavonoids are among the most prevalent phenolic compounds in foods like fruits, vegetables, grains, spices, beverages, and medicinal plants. They are characterized by a chemical structure with three carbon rings (C6-C3-C6) and distinct A, B, and C rings (as depicted in Figure 5). Flavonoids can be further categorized into subclasses, including flavones, flavonols, flavanones, anthocyanidins, and isoflavonoids (Fan et al., 2019).

With reference to the Kanwal et al. (2016) *in silico* docking studies with some flavones, the presence of methylation sites in predicted phytochemicals, such as R4=OCH3 in Kempferide and R4=OH in Herbacetin and Morin in the B ring (Figure 5, Table 1) bind efficiently to the catalytic pocket of DNMT, and seems that these compounds may share similarities with well-known Food and Drug Administration (FDA)-approved drugs like azacitidine (5-azacytidine) and decitabine (5-aza-deoxy-cytidine). These drugs contain a nitrogen atom within their pyrimidine ring, which is not found in regular cytidine and is essential for their inhibitory effects on DNA methyltransferases (DNMTs). Similarly, the flavonoids identified in our docking study sugested that structural elements that could potentially exchange methyl group and subsequently inhibit DNMTs. In addition flavones offer some advantages over azacitidine as a novel non-nucleoside inhibitor of DNMT1 because they are known to rapidly intercalate with DNA (Kausar et al., 2009; Zhang et al., 2012).

Furthermore, methyltransferase activity can be directly

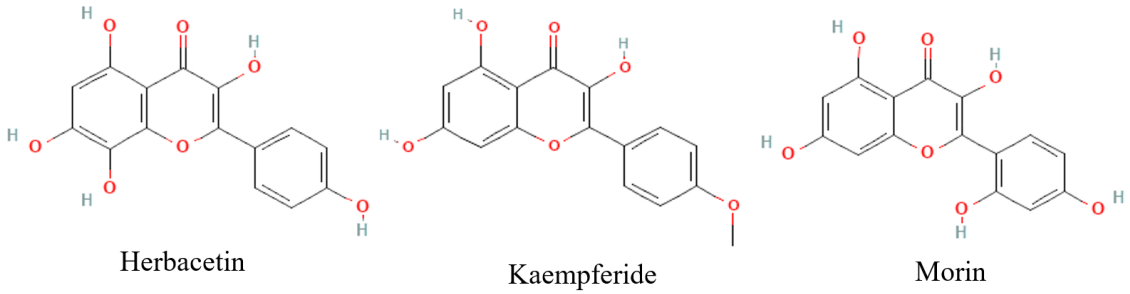


Figure 4: Phytochemical structures selected as DNMT inhibiotors

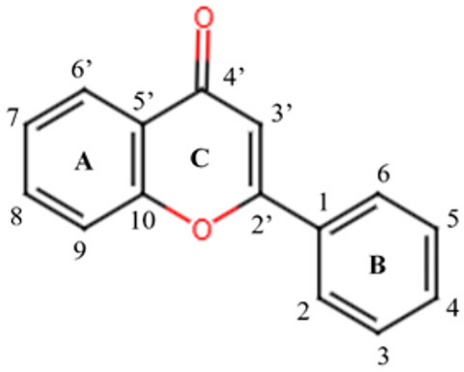


Figure 5: Core structure of Flavonoids

Table 3: Main differences in predicted phytochemicals

Compounds	Substructure
Herbacetin	R3', R6', R8, R9=OH
Kaempferide	R3', R6', R8=OH, R4=OCH3
Morin	R3', R6', R8', R2=OH, R4=OH

and competitively inhibited by various natural compounds, including EGCG from green tea, genistein from soybeans, apigenin and luteolin, and myricetin as a representative dietary flavonoid (Busch et al., 2015; Kanwal et al., 2016; Morris et al., 2016; Sanaei et al., 2018). This inhibition underscores the potential of flavonoids to interfere with the activity and stability of methyltransferases.

Computer-aided drug design is an important field in the discovery of new drugs as its utility is beyond estimation. Machine learning models and molecular docking simulations play a crucial role in identifying potential drug-like molecules from vast chemical libraries. However, these techniques encounter limitations. In developing machine learning models, it becomes necessary to consider the quantity of training data along with quality while selecting pertinent molecular properties that will give optimum performance (Badar, 2023).

The interaction between a protein and its ligand in docking simulations is usually simplified because factors such as the mobility of the protein or solvent effects, which would lead to unreliable results, are often ignored (Pujadas et al., 2008). As a result, laboratory experiments (*in vivo* and *in vitro*) play a significant role in the verification of computational techniques, while improved computational algorithms are needed for addressing such limitations. However, computational methodologies can significantly reduce cost and time in the drug discovery process.

CONCLUSION

The results presented in this study demonstrate the successful classification of structurally similar molecules, namely Herbacetin, Kaempferide, and Morin as potential inhibitors against DNMT using trained machine learning models. However, further *in-vitro* and *in-vivo* efficacy activity needs to be investigated for these phytochemicals. By combining ML, and molecular docking, researchers can efficiently screen a vast number of compounds, including those derived from natural sources, to identify potential lead molecules for further development as DNMT inhibitors. The integration of molecular visualization tools like Chimera and precise docking simulations via AutoDock Vina complemented our study. Significantly, our research sheds light on the interaction feasibility between compounds and DNMT structures, including those with mutations, offering valuable insights for drug discovery efforts targeting epigenetic regulators. These findings contribute to the ongoing efforts in developing novel epigenetic therapies, pave the way for future advancements in the field of DNMT inhibition for cancer treatment and structure-based drug discovery.

ACKNOWLEDGEMENT

We extend our gratitude to Dr. Dayanjan S. Wijesinghe, Associate Professor at the Department of Pharmacotherapy & Outcomes Science, Virginia Commonwealth University, USA for introducing the authors to the KNIME Analytics Platform. The authors would also like to extend their gratitude to the University of Peradeniya, Sri Lanka for providing necessary resources and facilities to conduct this study.

DECLARATION OF CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Anteneh, H., Fang, J., & Song, J. (2020). Structural basis for impairment of DNA methylation by the DNMT3A R882H mutation. *Nature communications*, *11*(1), 2294. https://doi.org/10.1038/s41467-020-16213-9

Ayala, R., Rapado, I., Onecha, E., Martínez-Cuadrón, D., Carreño-Tarragona, G., Bergua, J. M., Vives, S., Algarra, J. L., Tormo, M., Martinez, P., Serrano, J., Herrera, P., Ramos, F., Salamero, O., Lavilla, E., Gil, C., Lorenzo, J. L. L., Vidriales, M. B., Labrador, J., … Montesinos, P. (2021). The mutational landscape of acute myeloid leukaemia predicts responses and outcomes in elderly patients from the PETHEMA-FLUGAZA phase 3 clinical trial. *Cancers*, *13*(10), 2458. https://doi.org/10.3390/cancers13102458

Badar, M. S. (2023). A Guide to Applied Machine Learning for Biologists. *machine learning*, *1050*, 101. https://doi.org/10.1007/978-3-031-22206-1_11

Barta, G. (2016). Identifying biological pathway interrupting toxins using multi-tree ensembles. *Frontiers in Environmental Science*, *4*(AUG), 1–12.

Basha, S. M., Bagyalakshmi, K., Ramesh, C., Rahim, R., Manikandan, R., & Kumar, A. (2019). Comparative study on performance of document classification using supervised machine learning algorithms: KNIME. *International Journal on Emerging Technologies*, *10*(1), 148–153.

Busch, C., Burkard, M., Leischner, C., Lauer, U. M., Frank, J., & Venturelli, S. (2015). Epigenetic activities of flavonoids in the prevention and treatment of cancer.

Clinical Epigenetics, 1–18. https://doi.org/10.1186/s13148-015-0095-z

Chakravarti, S. K., & Alla, S. R. M. (2019). Descriptor free QSAR modeling using deep learning with long short-term memory neural networks. *Frontiers in artificial intelligence*, *2*, 17. https://doi.org/10.3389/frai.2019.00017

Chen, Y., Hao, E., Zhang, F., Du, Z., Xie, J., Chen, F., Yu, C., Hou, X., & Deng, J. (2021). Identifying Active Compounds and Mechanism of Camellia nitidissima Chi on Anti-Colon Cancer by Network Pharmacology and Experimental Validation. *Evidence-Based Complementary and Alternative Medicine*, *2021*. https://doi.org/10.1155/2021/7169211

Cheng, X., & Blumenthal, R. M. (2008). Mammalian DNA Methyltransferases: A Structural Perspective. *Structure*, *16*(3), 341–350. https://doi.org/10.1016/j.str.2008.01.004

Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *BioData mining*, *10*(1), 35. https://doi.org/10.1186/s13040-017-0155-3

Chik, F., & Szyf, M. (2010). Effects of specific DNMT gene depletion on cancer cell transformation and breast cancer cell invasion; toward selective DNMT inhibitors. *Carcinogenesis*, *32*(2), 224-232. https://doi.org/10.1093/carcin/bgq221

Chik, F., Szyf, M., & Rabbani, S. A. (2011). Role of epigenetics in cancer initiation and progression. *Advances in Experimental Medicine and Biology*, *720*, 91–104. https://doi.org/10.1007/978-1-4614-0254-1_8

Delpu, Y., Cordelier, P., Cho, W. C., & Torrisani, J. (2013). DNA methylation and cancer diagnosis. *International Journal of Molecular Sciences*, *14*(7), 15029–15058. https://doi.org/10.3390/ijms140715029

Döhner, H., Dolnik, A., Tang, L., Seymour, J. F., Minden, M. D., Stone, R. M., del Castillo, T. B., Al-Ali, H. K., Santini, V., Vyas, P., Beach, C. L., MacBeth, K. J., Skikne, B. S., Songer, S., Tu, N., Bullinger, L., & Dombret, H. (2018). Cytogenetics and gene mutations influence survival in older patients with acute myeloid leukemia treated with azacitidine or conventional care. *Leukemia*, *32*(12), 2546–2557. https://doi.org/10.1038/s41375-018-0257-z

Eberhardt, J., Santos-Martins, D., Tillack, A. F., & Forli, S. (2021). AutoDock Vina 1.2. 0: New docking methods, expanded force field, and python bindings. *Journal of chemical information and modeling*, *61*(8), 3891-3898. https://doi.org/10.1021/acs.jcim.1c00203

Esteller, M. (2011). Epigenetic changes in cancer. *F1000 Biology Reports*, *3*(1), 1039–1059. https://doi.org/10.3410/B3-9

Fan, M., Ding, H., Zhang, G., Hu, X., & Gong, D. (2019). Relationships of dietary flavonoid structure with its tyrosinase inhibitory activity and affinity. *Lwt*, *107*, 25–34. https://doi.org/10.1016/j.lwt.2019.02.076

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fillbrunn, A., Dietz, C., Pfeuffer, J., Rahn, R., Landrum, G. A., & Berthold, M. R. (2017). KNIME for reproducible cross-domain analysis of life science data. *Journal of Biotechnology*, *261*(February), 149–156. https://doi.org/10.1016/j.jbiotec.2017.07.028

Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., & Overington, J. P. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Research*, *40*(D1), 1100–1107. https://doi.org/10.1093/nar/gkr777

Gnyszka, A., JASTRZĘBSKI, Z., & Flis, S. (2013). DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer. *Anticancer research*, *33*(8), 2989-2996.

Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annual Review of Biochemistry*, *74*, 481–514. https://doi.org/10.1146/annurev.biochem.74.010904.153721

Grotewold, E. (2006). *The science of flavonoids.* In Springer (Issue May 2014). https://doi.org/10.1007/978-0-387-28822-2

Hermann, A., Gowher, H., & Jeltsch, A. (2004). Biochemistry and biology of mammalian DNA methyltransferases. *Cellular and Molecular Life Sciences*, *61*(19–20), 2571–2587. https://doi.org/10.1007/s00018-004-4201-1

Iwase, Y., Takemura, Y., Ju-ichi, M., Mukainaka, T., Ichiishi, E., Ito, C., Furukawa, H., Yano, M., Tokuda, H., & Nishino, H. (2001). Inhibitory effect of flavonoid derivatives on Epstein-Barr virus activation and two-stage carcinogenesis of skin tumors. *Cancer Letters*, *173*(2), 105–109. https://doi.org/10.1016/S0304-3835(01)00615-2

Jarada, T. N., Rokne, J. G., & Alhajj, R. (2020). A review of computational drug repositioning: Strategies, approaches, opportunities, challenges, and directions. *Journal of Cheminformatics*, *12*(1), 1–23. https://doi.org/10.1186/s13321-020-00450-7

Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., & Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, *449*(7159), 248-251. https://doi.org/10.1038/nature06146

Jiang, Z., Wang, J., Chen, X., Wang, X., Wang, T., Zhu, Z., & Pan, J. (2018). Simultaneous determination of kaempferide, kaempferol and isorhamnetin in rat plasma by ultra-high performance liquid chromatography-tandem mass spectrometry and its application to a pharmacokinetic study. *Journal of the Brazilian Chemical Society*, *29*(3), 535–542. https://doi.org/10.21577/0103-5053.20170166

Jones, P. A., & Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. *Nature Reviews Genetics, 10*(11), 805–811. https://doi.org/10.1038/nrg2651

Jones, P. A., & Taylor, S. M. (1980). Cellular differentiation, cytidine analogs and DNA methylation. *Cell*, *20*(1), 85–93. https://doi.org/10.1016/0092-8674(80)90237-8

Kanwal, R., Datt, M., Liu, X., & Gupta, S. (2016). Dietary flavones as dual inhibitors of DNA methyltransferases and histone methyltransferases. *PloS one*, *11*(9), e0162956. https://doi.org/10.1371/journal.pone.0162956

Kausar, N., Siddiqa, A., Yaqub, A., & Sabahat, S. (2009). Spectrophotometric analysis of flavonoid-DNA binding interactions at physiological conditions. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *74*, 1135–1137. https://doi.org/10.1016/j.saa.2009.09.022

Kawabata, K., Tanaka, T., Honjo, S., Kakumoto, M., Hara, A., Makita, H., Tatematsu, N., Ushida, J., Tsuda, H., & Mori, H. (1999). Chemopreventive effect of dietary flavonoid morin on chemically induced rat tongue carcinogenesis. *International Journal of Cancer*, *83*(3), 381–386. https://doi.org/10.1002/(SICI)1097-0215(19991029)83:3<381::AID-IJC14>3.0.CO;2-X

Kim, D. J., Lee, M. H., Liu, K. D., Lim, D. Y., Roh, E., Chen, H., Kim, S. H., Shim, J. H., Kim, M. O., Li, W., Ma, F., Fredimoses, M., Bode, A. M., & Dong, Z. (2017). Herbacetin suppresses cutaneous squamous cell carcinoma and melanoma cell growth by targeting AKT and ODC. *Carcinogenesis*, *38*(11), 1136–1146. https://doi.org/10.1093/carcin/bgx082

Kim, D. J., Roh, E., Lee, M. H., Oi, N., Lim, D. Y., Kim, M. O., ... & Dong, Z. (2016). Herbacetin is a novel allostceric inhibitor of ornithine decarboxylase with antitumor activity. *Cancer research*, *76*(5), 1146-1157. https://doi.org/10.1158/0008-5472.CAN-15-0442

Koyama, S., Kondo, K., Ueha, R., Kashiwadani, H., & Heinbockel, T. (2021). Possible use of phytochemicals for recovery from COVID-19-induced anosmia and ageusia. *International Journal of Molecular Sciences*, *22*(16), 8912. https://doi.org/10.3390/ijms22168912

Kuo, H. M., Chang, L. S., Lin, Y. L., Lu, H. F., Yang, J. S., Lee, J. H., & Chung, J. G. (2007). Morin inhibits the growth of human leukemia HL-60 cells via cell cycle arrest and induction of apoptosis through mitochondria dependent pathway. *Anticancer Research*, *27*(1 A), 395–406.

Lai, J. P., Lim, Y. H., Su, J., Shen, H. M., & Ong, C. N. (2007). Identification and characterization of major flavonoids and caffeoylquinic acids in three Compositae plants by LC/DAD-APCI/MS. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, *848*(2), 215–225. https://doi.org/10.1016/j.jchromb.2006.10.028

Lebaron, S., Zeltzer, L. K., Lebaron, C., Scott, S. E., & Zeltzer, P. M. (1988). Chemotherapy side effects in pediatric oncology patients: Drugs, age, and sex as risk factors. *Medical and Pediatric Oncology*, *16*(4), 263–268. https://doi.org/10.1002/mpo.2950160408

Li, H. L., Li, S. M., Luo, Y. H., Xu, W. T., Zhang, Y., Zhang, T., Zhang, D. J., & Jin, C. H. (2020). Kaempferide Induces G0/G1 Phase Arrest and Apoptosis via ROS-Mediated Signaling Pathways in A549 Human Lung Cancer Cells. *Natural Product Communications*, *15*(7), 1–13. https://doi.org/10.1177/1934578X20935226

Li, H., Li, W., Liu, S., Zong, S., Wang, W., Ren, J., Li, Q., Hou, F., & Shi, Q. (2016). DNMT1, DNMT3A and DNMT3B Polymorphisms Associated With Gastric Cancer Risk: A Systematic Review and Meta-analysis. *EBioMedicine*, *13*, 125–131. https://doi.org/10.1016/j.ebiom.2016.10.028

Lipinski, C. A. (2004). Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discovery Today: Technologies*, *1*(4), 337–341. https://doi.org/10.1016/j.ddtec.2004.11.007

Luskin, M. R., Lee, J. W., Fernandez, H. F., Abdel-Wahab, O., Bennett, J. M., Ketterling, R. P., ... & Luger, S. M. (2016). Benefit of high-dose daunorubicin in AML induction extends across cytogenetic and molecular groups. *Blood, The Journal of the American Society of Hematology*, *127*(12), 1551-1558. https://doi.org/10.1182/blood-2015-07-657403

Malongane, F., McGAW, L. J., & Mudau, F. N. (2017). The synergistic potential of various teas, herbs and therapeutic drugs in health improvement: a review. *Journal of the Science of Food and Agriculture*, *97*(14), 4679-4689. https://doi.org/10.1002/jsfa.8472

Morris, J., Moseley, V. R., Cabang, A. B., Coleman, K., Wei, W., Garrett-Mayer, E., & Wargovich, M. J. (2016). Reduction in promotor methylation utilizing EGCG (Epigallocatechin-3-gallate) restores RXRα expression in human colon cancer cells. *Oncotarget*, *7*(23), 35313–35326. https://doi.org/10.18632/oncotarget.9204

Mund, M. D., Alam, S., Khan, U. H., Tahir, U., Zubair, M. S., Younas, T., & Mustafa, B. E. (2016). Phytochemicals as complementary and alternative therapeutic formulations with potential pro-apoptotic effects on various cancerous cell lines: a literature survey. *Focus Sci*, *2*(2), 1-5. https://doi.org/10.20286/focsci-020220

Nath, L. R., Gorantla, J. N., Joseph, S. M., Antony, J., Thankachan, S., Menon, D. B., Sankar, S., Lankalapalli, R. S., & Anto, R. J. (2015). Kaempferide, the most active among the four flavonoids isolated and characterized from Chromolaena odorata, induces apoptosis in cervical cancer cells while being pharmacologically safe. *RSC Advances*, *5*(122), 100912–100922. https://doi.org/10.1039/c5ra19199h

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*(3), 247–257. https://doi.org/10.1016/S0092-8674(00)81656-6

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C. A., Hagemeier, C., & Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angewandte Chemie International Edition, 50*(31), 7008–7012.

Pujadas, G., Vaque, M., Ardevol, A., Blade, C., Salvado, M., Blay, M., Fernandez-Larrea, J., & Arola, L. (2008). Protein-ligand Docking: A Review of Recent Advances and Future Perspectives. *Current Pharmaceutical Analysis*, *4*(1), 1–19. https://doi.org/10.2174/157341208783497597

Robert, P., Jennifer, C. C., Katherine, G., Kelly, A. M., Viswanathan, R., & Peter, R. (2006). Comparison of Protein Active Site Structures for Functional Annotation

of Proteins and Drug Design Robert. *PROTEINS: Structure, Function, and Bioinformatics*, *65*, 124–135. https://doi.org/10.1002/prot.21092

Roughley, S. D. (2018). Five Years of the KNIME Vernalis Cheminformatics Community Contribution. *Current Medicinal Chemistry*, *27*(38), 6495–6522. https://doi.org/10.2174/0929867325666180904113616

Russo, D. P., Zorn, K. M., Clark, A. M., Zhu, H., & Ekins, S. (2018). Comparing multiple machine learning algorithms and metrics for estrogen receptor binding prediction. *Molecular pharmaceutics*, *15*(10), 4361-4370. https://doi.org/10.1021/acs.molpharmaceut.8b00546

Saldívar-González, F. I., Gómez-García, A., Chavez-Ponce de Leon, D. E., Sánchez-Cruz, N., Ruiz-Rios, J., Pilón-Jiménez, B. A., & Medina-Franco, J. L. (2018). Inhibitors of DNA methyltransferases from natural sources: A computational perspective. *Frontiers in pharmacology*, *9*, 418529. https://doi.org/10.3389/fphar.2018.01144

Sanaei, M., Kavoosi, F., Roustazadeh, A., & Golestan, F. (2018). Effect of genistein in comparison with trichostatin a on reactivation of dnmts genes in hepatocellular carcinoma. *Journal of Clinical and Translational Hepatology*, *6*(2), 141–146. https://doi.org/10.14218/JCTH.2018.00002

Shih, A. H., Abdel-Wahab, O., Patel, J. P., & Levine, R. L. (2012). The role of mutations in epigenetic regulators in myeloid malignancies. *Nature Reviews Cancer*, *12*(9), 599–612. https://doi.org/10.1038/nrc3343

Solairaja, S., Andrabi, M. Q., Dunna, N. R., & Venkatabalasubramanian, S. (2021). Overview of Morin and Its Complementary Role as an Adjuvant for Anticancer Agents. *Nutrition and Cancer*, *73*(6), 927–942. https://doi.org/10.1080/01635581.2020.1778747

Sydow, D., Wichmann, M., Rodríguez-Guerra, J., Goldmann, D., Landrum, G., & Volkamer, A. (2019). TeachOpenCADD-KNIME: a teaching platform for computer-aided drug design using KNIME workflows. *Journal of Chemical Information and Modeling*, *59*(10), 4083-4086. https://doi.org/10.1021/acs.jcim.9b00662

Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics*, *29*(6–7), 476–488. https://doi.org/10.1002/minf.201000061

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, *31*(2), 455-461. https://doi.org/10.1002/jcc.21334

Xu, Y., Sun, Y., Shen, H., Ding, L., Yang, Z., Qiu, H., Sun, A., Chen, S., & Wu, D. (2015). Allogeneic hematopoietic stem cell transplantation could improve survival of cytogenetically normal adult acute myeloid leukemia patients with DNMT3A mutations. *American Journal of Hematology*, *90*(11), 992–997. https://doi.org/10.1002/ajh.24135

Yusuf, H., Kamarlis, R. K., Yusni, Y., & Fahriani, M. (2021). The anticancer activity of ethanol extract of Chromolaena odorata leaves in 7 , 12-Dimethylbenz [a] anthracene in (DMBA) induced breast cancer Wistar rats (Rattus novergicus). *Pharmacia*, *68*(2), 493–499. https://doi.org/10.3897/pharmacia.68.e63956

Zhang, S., Ling, B., Qu, F., & Sun, X. (2012). Investigation on the interaction between luteolin and calf thymus DNA by spectroscopic techniques. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *97*, 521–525. https://doi.org/10.1016/j.saa.2012.06.040

Zhang, Z. M., Liu, S., Lin, K., Luo, Y., Perry, J. J., Wang, Y., & Song, J. (2015). Crystal structure of human DNA methyltransferase 1. *Journal of molecular biology*, *427*(15), 2520-2531. https://doi.org/10.1016/j.jmb.2015.06.001

Supplementary table S1a: Predicted active compounds list from machine learning workflow

Activity	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
P (Activity=1)	1	1	1	1	1	1	1	1
P (Activity=0)	0	0	0	0	2.21E-11	0	0	0
Prediction (Activity)	1	1	1	1	1	1	1	1
molecule_chembl_id	CHEMBL359965	CHEMBL95889	CHEMBL1261	CHEMBL88985	CHEMBL611029	CHEMBL40919	CHEMBL28626	CHEMBL307145
molecule_properties.full_molformula	C6H10OS2	C5H12NO2+	C6H8O7	C15H18O2	C15H10O7	C16H12O6	C15H10O7	C6H6O3
pref_name	ALLICIN	BETAINE	CITRIC ACID	DEHYDROCOSTUSLACTONE	HERBACETIN	KAEMPFERIDE	MORIN	PYROGALLOL
molecule_synonyms.0.molecule_synonym	Allicin	Abromine	Acidum citricum	Dehydrocostuslactone	Herbacetin	NSC-407294	NSC-19801	NSC-5035
molecule_type	Small molecule	Small molecule	Small molecule	Small molecule	Small molecule	Small molecule	Small molecule	Small molecule
molecule_properties.full_mwt	162.28	118.16	192.12	230.31	302.24	300.27	302.24	126.11
molecule_properties.hba	2	1	4	2	7	6	7	3
molecule_properties.hbd	0	1	4	0	5	3	5	3
molecule_properties.heavy_atoms	9	8	13	17	22	22	22	9
molecule_properties.molecular_species	NEUTRAL	ACID	ACID		NEUTRAL	ACID	ACID	NEUTRAL
Activity (right)	0	0	0	0	0	0	0	0
JSON URL	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL359965.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL95889.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL1261.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL88985.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL611029.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL40919.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL28626.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL307145.json
Plant name	Garlic	Trigonella foenum-graecum	Anabasis aphylla (Chenopodiaceae)	Geranium oil	Ginger (Zingiber officinale Roscoe)	Alpinia officinarum (Zingiberaceae)	Psidium guajava	C. verum
Compound name	Allicin	Betain	CITRIC-ACID	Dehydrocostus lactone	Herbacetin	KAEMPFERID	Morin	Pyrogallol
Activity (right) (right)	0	0	0	0	0	0	0	0
Plant name (right)	Garlic	Trigonella foenum-graecum	Anabasis aphylla (Chenopodiaceae)	Geranium oil	Ginger (Zingiber officinale Roscoe)	Alpinia officinarum (Zingiberaceae)	Psidium guajava	C. verum
Compound name (right)	Allicin	Betain	CITRIC-ACID	Dehydrocostus lactone	Herbacetin	KAEMPFERID	Morin	Pyrogallol
Status (right)	200	200	200	200	200	200	200	200
Content type (right)	application/json	application/json	application/json	application/json	application/json	application/json	application/json	application/json
chemical_probe	0	0	0	0	0	0	0	0
molecule_hierarchy.active_chembl_id	CHEMBL359965	CHEMBL95889	CHEMBL1261	CHEMBL88985	CHEMBL611029	CHEMBL40919	CHEMBL28626	CHEMBL307145
molecule_properties.np_likeness_score	-0.03	0.88	1.05	3	1.68	1.31	1.6	0.8
SmilesValue (RDKit Mol)	C=CCS[S+]([O-])CC=C	C[N+](C)(C)CC(=O)O	O=C(O)CC(O)(CC(=O)O)C(=O)O	C=C1CC[C@H]2C(=C)CC[C@H]3C(=C)C(=O)O[C@@H]3[C@@H]12	O=c1c(O)c(-c2ccc(O)cc2)oc2c(O)c(O)cc(O)c12	COc1ccc(-c2oc-3cc(O)cc(O)c3c(=O)c2O)cc1	O=c1c(O)c(-c2ccc(O)cc2O)oc2cc(O)cc(O)c12	Oc1cccc(O)c1O

Supplementary table S1b: Filtered compounds from Lipinski's rules

Activity	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Prediction (Activity)	1	1	1	1	1	1	1	1
molecule_chembl_id	CHEMBL359965	CHEMBL95889	CHEMBL1261	CHEMBL88985	CHEMBL611029	CHEMBL40919	CHEMBL28626	CHEMBL307145
molecule_properties.full_molformula	C6H10OS2	C5H12NO2+	C6H8O7	C15H18O2	C15H10O7	C16H12O6	C15H10O7	C6H6O3
pref_name	ALLICIN	BETAINE	CITRIC ACID	DEHYDROCOSTUS-LACTONE	HERBACETIN	KAEMPFERIDE	MORIN	PYROGALLOL
molecule_properties.full_mwt	162.28	118.16	192.12	230.31	302.24	300.27	302.24	126.11
molecule_properties.hba	2	1	4	2	7	6	7	3
molecule_properties.hbd	0	1	4	0	5	3	5	3
molecule_properties.heavy_atoms	9	8	13	17	22	22	22	9
Activity (right)	0	0	0	0	0	0	0	0
JSON URL	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL359965.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL95889.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL1261.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL88985.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL611029.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL40919.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL28626.json	http://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL307145.json
Plant name	Garlic	Trigonella foenum-graecum	Anabasis aphylla (Chenopodiaceae)	Geranium oil	Ginger (Zingiber officinale Roscoe)	Alpinia officinarum (Zingiberaceae)	Psidium guajava	C. verum
Compound name	Allicin	Betain	CITRIC-ACID	Dehydrocostus lactone	Herbacetin	KAEMPFERID	Morin	Pyrogallol
Activity (right) (right)	0	0	0	0	0	0	0	0
Plant name (right)	Garlic	Trigonella foenum-graecum	Anabasis aphylla (Chenopodiaceae)	Geranium oil	Ginger (Zingiber officinale Roscoe)	Alpinia officinarum (Zingiberaceae)	Psidium guajava	C. verum
Compound name (right)	Allicin	Betain	CITRIC-ACID	Dehydrocostus lactone	Herbacetin	KAEMPFERID	Morin	Pyrogallol
molecule_properties.np_likeness_score	-0.03	0.88	1.05	3	1.68	1.31	1.6	0.8
SmilesValue (RDKit Mol)	C=CCS[S+]([O-])CC=C	C[N+](C)(C)CC(=O)O	O=C(O)CC(O)(CC(=O)O)C(=O)O	C=C1CC[C@H]2C(=C)CC[C@H]3C(=C)C(=O)O[C@@H]3[C@@H]12	O=c1c(O)c(-c2ccc(O)cc2)oc2c(O)c(O)cc(O)c12	COc1ccc(-c2oc3cc(O)cc(O)c3c(=O)c2O)cc1	O=c1c(O)c(-c2ccc(O)cc2O)oc2cc(O)cc(O)c12	Oc1cccc(O)c1O
SmilesValue_(RDKit Mol)	C=CCS[S+]([O-])CC=C	C[N+](C)(C)CC(=O)O	O=C(O)CC(O)(CC(=O)O)C(=O)O	C=C1CC[C@H]2C(=C)CC[C@H]3C(=C)C(=O)O[C@@H]3[C@@H]12	O=c1c(O)c(-c2ccc(O)cc2)oc2c(O)c(O)cc(O)c12	COc1ccc(-c2oc3cc(O)cc(O)c3c(=O)c2O)cc1	O=c1c(O)c(-c2ccc(O)cc2O)oc2cc(O)cc(O)c12	Oc1cccc(O)c1O
SlogP	1.7553	-0.2228	-1.2485	3.0166	1.988	2.5854	1.988	0.8034
SMR	45.8614	30.5352	37.0912	66.238	76.244	79.4664	76.244	31.4364

ExactMW	162.0173	118.0863	192.027	230.1307	302.0427	300.0634	302.0427	126.0317
NumHBD	0	1	4	0	5	3	5	3
NumHBA	2	1	4	2	7	6	7	3
SMR_bool	1	0	0	1	1	1	1	0

Supplementary table S2: Docking results with AutoDock Vina

Azacitidine	**Score**	**RMSD l.b.**	**RMSD u.b.**	**Hbond (all)**	**Hbond Ligand Atoms**	**Hbond Receptor Atoms**	
4WXX (DNMT1) with Azacitidine (Compound) FDA approved drug	-6.5	0.0	0.0	1	1	1	REMARK 5 active torsions: REMARK status: ('A' for Active; 'I' for Inactive) REMARK I between atoms: C2_3 and N4_17 REMARK 1 A between atoms: C4_8 and N3_7 REMARK 2 A between atoms: C5_9 and O5_16 REMARK 3 A between atoms: C6_10 and O4_15 REMARK 4 A between atoms: C7_11 and C8_13 REMARK 5 A between atoms: C8_13 and O3_14
	-6.5	0.0	0.0	1	1	1	
	-6.5	0.0	0.0	2	2	2	
Average	-6.5	0.0	0.0	1	1	1	
2QRV (E chain DNMT3A) with Azacitidine (Compound) FDA approved drug	-7.2	0.0	0.0	3	3	2	
	-7.2	0.0	0.0	3	3	2	
	-7.2	0.0	0.0	5	5	4	
Average	-7.2	0	0	4	4	3	
6W8J (A chain DNMT3A mutated) with Azacitidine (Compound) FDA approved drug	-6.7	0.0	0.0	5	4	4	
	-6.7	0.0	0.0	5	4	4	
	-6.7	0.0	0.0	5	4	4	
Average	-6.7	0	0	5	4	4	
Dehydrocostus lactone							REMARK 0 active torsions: REMARK status: ('A' for Active; 'I' for Inactive)
4WXX (DNMT1)	-6.7	0.00	0.00	0	0	0	
	-6.7	0.00	0.00	0	0	0	
	-6.7	0.00	0.00	0	0	0	
Average	-6.7	0.00	0.00	0	0	0	
2QRV (E chain DNMT3A)	-8.0	0.00	0.00	0	0	0	
	-8.0	0.00	0.00	0	0	0	
	-8.0	0.00	0.00	0	0	0	
Average	-8.0	0.00	0.00	0	0	0	
6W8J (A chain DNMT3A mutated)	-5.0	2.17	4.62	2	1	2	
	-4.2	2.26	4.49	2	1	2	
	-5.0	2.17	4.62	2	1	2	
Average	-4.7	2.20	4.58	2	1	2	

Morin							REMARK 6 active torsions: REMARK status: ('A' for Active; 'I' for Inactive) REMARK 1 A between atoms: C3_3 and C7_9 REMARK 2 A between atoms: C4_4 and O2_8 REMARK 3 A between atoms: C6_6 and O1_7 REMARK 4 A between atoms: C8_10 and O7_22 REMARK 5 A between atoms: C11_14 and O6_21 REMARK 6 A between atoms: C13_16 and O5_20
4WXX (DNMT1)	-8.6	0.000	0.000	2	2	2	
	-8.6	0.000	0.000	2	2	2	
	-8.6	0.000	0.000	3	2	3	
Average	-8.6	0.000	0.000	2	2	2	
2QRV (E chain DNMT3A)	-8.2	0.000	0.000	1	1	1	
	-7.6	1.720	3.405	2	2	2	
	-8.2	0.000	0.000	2	2	2	
Average	-8.0	0.573	1.135	2	2	2	
6W8J (A chain DNMT3A mutated)	-9.0	0.000	0.000	2	2	2	
	-9.0	0.000	0.000	1	1	1	
	-9.0	0.000	0.000	1	1	1	
Average	-9.0	0.000	0.000	1	1	1	
Herbacetin							REMARK 6 active torsions: REMARK status: ('A' for Active; 'I' for Inactive) REMARK 1 A between atoms: C3_3 and O7_22 REMARK 2 A between atoms: C6_6 and C7_7 REMARK 3 A between atoms: C8_8 and O6_21 REMARK 4 A between atoms: C12_14 and O5_20 REMARK 5 A between atoms: C13_15 and O4_19 REMARK 6 A between atoms: C15_17 and O3_18
4WXX (DNMT1)	-7.9	0.000	0.000	2	2	2	
	-7.9	0.000	0.000	3	3	3	
	-7.9	0.000	0.000	2	2	2	
Average	-7.9	0.000	0.000	2	2	2	
2QRV (E chain DNMT3A)	-8.2	0.000	0.000	1	1	1	
	-7.9	0.000	0.000	1	1	1	
	-7.4	3.962	5.839	1	1	1	
Average	-7.8	1.321	1.946	1	1	1	
6W8J (A chain DNMT3A mutated)	-7.9	0.000	0.000	2	1	2	
	-7.9	0.000	0.000	3	2	3	
	-7.9	0.000	0.000	2	1	2	
Average	-7.9	0.000	0.000	2	1	2	
Kampferide							REMARK 5 active torsions: REMARK status: ('A' for Active; 'I' for Inactive) REMARK 1 A between atoms: C2_3 and O1_2 REMARK 2 A between atoms: C5_6 and C8_9 REMARK 3 A between atoms: C9_10 and O6_22 REMARK 4 A between atoms: C12_14 and O5_21 REMARK 5 A between atoms: C14_16 and O4_20
4WXX (DNMT1)	-8.4	0.0	0.0	2	2	2	
	-8.4	0.0	0.0	2	2	2	
	-8.4	0.0	0.0	1	1	1	
Average	-8.4	0.0	0.0	2	2	2	
2QRV (E chain DNMT3A)	-7.3	1.5	3.3	1	1	1	
	-7.3	1.5	3.3	1	1	1	
	-7.9	0.0	0.0	1	1	1	
Average	-7.5	1.0	2.2	1	1	1	
6W8J (A chain DNMT3A mutated)	-7.1	0.7	1.2	1	1	1	
	-7.7	2.0	7.5	1	1	1	
	-6.7	2.4	3.8	1	1	1	
Average	-7.2	1.7	4.2	1	1	1	

Allicin							REMARK 5 active torsions: REMARK status: ('A' for Active; 'I' for Inactive) REMARK 1 A between atoms: C2_2 and C3_3 REMARK 2 A between atoms: C3_3 and S1_4 REMARK 3 A between atoms: S1_4 and S2_5 REMARK 4 A between atoms: C4_7 and S2_5 REMARK 5 A between atoms: C4_7 and C5_8
4WXX (DNMT1)	0.0	0.0	0.0	0	0	0	
	0.0	0.0	0.0	0	0	0	
	0.0	0.0	0.0	0	0	0	
Average	0.0	0.0	0.0	0	0	0	
2QRV (E chain DNMT3A)	-4.4	0.0	0.0	2	2	1	
	-4.4	0.0	0.0	1	1	1	
	-4.3	0.0	0.0	1	1	1	
Average	-4.4	0.0	0.0	1	1	1	
6W8J (A chain DNMT3A mutated)	-4.1	0.0	0.0	0	0	0	
	-4.1	0.0	0.0	0	0	0	
	-4.1	0.0	0.0	0	0	0	
Average	-4.1	0.0	0.0	0	0	0	
Betaine							REMARK 2 active torsions: REMARK status: ('A' for Active; 'I' for Inactive) REMARK 1 A between atoms: C1_4 and N1_3 REMARK 2 A between atoms: C1_4 and C5_8
4WXX (DNMT1)	0.0	0.0	0.0	0	0	0	
	0.0	0.0	0.0	0	0	0	
	0.0	0.0	0.0	0	0	0	
Average	0.0	0.0	0.0	0	0	0	
2QRV (E chain DNMT3A)	-3.8	0.0	0.0	1	1	1	
	-3.8	0.0	0.0	1	1	1	
	-3.8	0.0	0.0	1	1	1	
Average	-3.8	0.0	0.0	1	1	1	
6W8J (A chain DNMT3A mutated)	-3.5	0.0	0.0	4	4	4	
	-3.5	0.0	0.0	3	3	3	
	-3.5	0.0	0.0	3	3	3	
Average	-3.5	0.0	0.0	3	3	3	
Citric acid							REMARK 9 active torsions: REMARK status: ('A' for Active; 'I' for Inactive) REMARK 1 A between atoms: C1_1 and C2_2 REMARK 2 A between atoms: C1_1 and C3_5 REMARK 3 A between atoms: C2_2 and O2_4 REMARK 4 A between atoms: C3_5 and C4_6 REMARK 5 A between atoms: C3_5 and C6_10 REMARK 6 A between atoms: C3_5 and O7_13 REMARK 7 A between atoms: C4_6 and C5_7 REMARK 8 A between atoms: C5_7 and O4_9 REMARK 9 A between atoms: C6_10 and O6_12
4WXX (DNMT1)	0.0	0.0	0.0	0.0	0.0	0.0	
	0.0	0.0	0.0	0.0	0.0	0.0	
	0.0	0.0	0.0	0.0	0.0	0.0	
Average	0.0	0.0	0.0	0	0	0	
2QRV (E chain DNMT3A)	-6.2	0.0	0.0	3	3	3	
	-6.1	0.0	0.0	3	3	3	
	-6.1	0.0	0.0	1	1	1	
Average	-6.1	0.0	0.0	2	2	2	

6W8J (A chain DNMT3A mutated)	-5.7	0.0	0.0	2	2	2	
	-5.7	0.0	0.0	4	3	3	
	-5.7	1.1	4.5	3	2	3	
Average	-5.7	0.4	1.5	3	2	3	
Pyrogallol							REMARK 3 active torsions:
4WXX (DNMT1)	0.0	0.0	0.0	0.0	0.0	0.0	REMARK status: ('A' for Active; 'I' for Inactive)
	0.0	0.0	0.0	0.0	0.0	0.0	REMARK 1 A between atoms: C3_3 and O3_9
	0.0	0.0	0.0	0.0	0.0	0.0	REMARK 2 A between atoms: C4_4 and O2_8
Average	0.0	0.0	0.0	0	0	0	REMARK 3 A between atoms: C5_5 and O1_7
2QRV (E chain DNMT3A)	-5.8	0.0	0.0	0	0	0	
	-5.9	0.0	0.0	0	0	0	
	-5.9	0.0	0.0	0	0	0	
Average	-5.9	0.0	0.0	0	0	0	
6W8J (A chain DNMT3A mutated)	-5.4	0.0	0.0	1	1	1	
	-5.4	0.0	0.0	1	1	1	
	-5.4	0.0	0.0	1	1	1	
Average	-5.4	0.0	0.0	1	1	1	