2023
10(2), 537

Chemical Biology
LETTERS

In silico identification of small natural inhibitors against DNA methyl transferase 3-like protein by integrative molecular docking and molecular dynamics approach

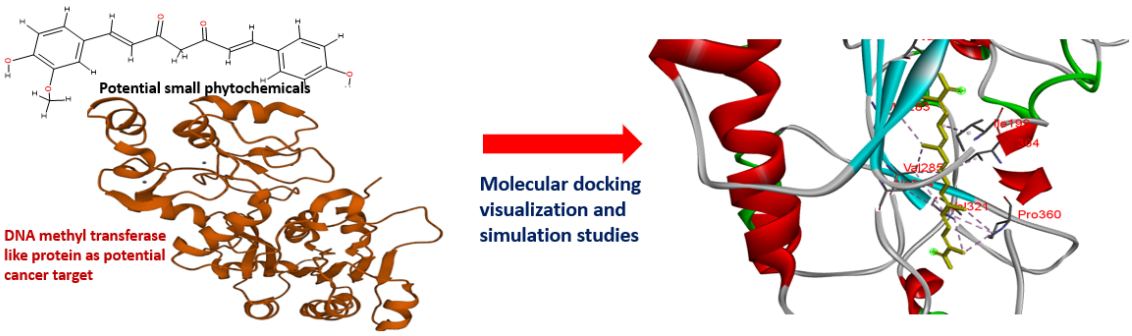
Paratpar Sarkar[1], Niraj Niraj[2] and Vivek Srivastava[1*]

[1]*Department of Chemistry & Biochemistry, Sharda School of Basic Sciences & Research, Sharda University, Greater Noida-201308, UP, India,* [2]*Dr. R.P.Centre, All India Institute of Medical Sciences, Delhi, India.*

Submitted on: 22-Sept-2022, Accepted and Published on: 27-Jan-2023

Article

ABSTRACT



Tiny phytochemicals found in different types of spices may play a significant role as anticancer drugs, particularly by focusing on a specific cancer receptor to modulate the cancer signaling cascade. We performed an *in silico* ADME and PASS analysis of twelve structurally different phytochemicals. Further, an integrative molecular docking and structural dynamics approach was carried out against a potential cancer target: DNA methyl transferase 3-like protein. Adopting an amalgamation of docking, simulation study, and MM-GBSA binding energy approach, we found carnosic acid (Ca) and crocetin (Cr) to be the best lead molecules. Hence, Ca and Cr may be proposed as strong potential anti-cancerous compounds against DNA methyl transferase3-like proteins.

Keywords: DNA methyl transferase 3-like protein, PASS, Molecular Docking, Molecular Dynamics, ADMET

INTRODUCTION

The cancer is the leading cause of death worldwide.[1] Despite various progress in the emergence of new drug therapeutics and various combinations, none of these serve as suitable options during the advanced stage of cancer prevention. The targeted approach somewhat leads to the stopping of cell signaling approach.[2] DNA methyl transferases catalyze the methylation of cytosine at the 5-position of the nitrogenous base: cytosine. In mammals, it is DNMT1 that plays an important role in epigenetic processes and mainly methylates newly synthesized DNA. During cancer, the level of expression of some of the types of DNA methyl transferases like DNMT3B, DNMT3A, and DNMT3L increases and are responsible for hypermethylation of tumor suppressor genes of promoter region rich in CpG.[2,3] Various studies have reported that the inhibition of DNA methyl transferase through various drugs has benefited the cancer patient by reducing tumor size. This has also allowed more expression of tumor suppressor genes in those patients and thus DNA methyl transferase has been a potential cancer target for various anti-cancer drugs.[3] Apart from many synthetic drugs, phytochemicals from dietary origins have tremendous potential to inhibit DNA methyl transferases in a cancer cell and thus can be a very good option for cancer. The role of DNA methyltransferase-like protein (DNMT3L) is very minimal. This enzyme act as a stimulatory factor and its function is to modulate DNMT3a activity.[3,4]

Bioactive compounds in various plant species play a significant role in human beings. Compounds with high molecular weight and low biological activities such as tannins, saponins, and quinones, and the others are those of low molecular weight and high biological activities such as coumarins, flavonoids, terpenoids, and alkaloids.[5-7] Various research suggests that many new drugs have been generated from natural products, particularly from plant origin and these small molecules play a significant role in different properties linked with biochemical activities. These molecules are applied as

*Corresponding Author: Dr. Vivek Srivastava, Ph.D.
Tel: +91-9873950422
Email: vivek.srivastava1@sharda.ac.in

URN:NBN:sciencein.cbl.2023.v10.537

ISSN: 2347–9825
https://pubs.thesciencein.org/cbl

therapeutics for the cure of different diseases including cancer. Even small molecules have been targeted for activating DNA methylation repressed genes.[8,9]

Studying different properties like absorption, distribution, metabolism, elimination, and toxicity of drug-like molecules before experimenting in the laboratory gives researchers a wide option to screen hundreds and thousands of molecules through computer-based in-silico studies. These studies have a lot of contribution in determining the viability of drugs screened particularly for their oral behavior and their role in the distribution and the final destination of reaching and binding with the target. Every year, nearly 50% of the drugs failed in clinical trials because of their incomplete in silico screening of absorption, distribution, metabolism, elimination, and toxicity (ADMET) properties.[10] In this regard, Swiss ADME and ADMET SAR provide detailed and extensive physicochemical profiles, medicinal properties, and ADMET properties of any drug molecule.[11] The absorption and distribution characteristics of any drug are greatly affected by the solubility of that drug in aqueous media. As 75 percent of our body is composed of water, so drugs soluble in water are more easily transported to reach targets. Even the last marketing of drugs failed because of poor pharmacokinetic properties like stability, solubility, absorption, and toxicity.[12]

Lipinski has proposed the very first drug likeliness feature which is widely known as the rule-of-five (MWT $\leq$ 500, log P $\leq$ 5, H-bond donors $\leq$ 5, H-bond acceptors $\leq$ 10).[13] A combination of drug-likeness, cLog P, log S, molecular weight, and toxicity risks makes the overall drug score for any particular drugs proposed as oral inhibitors. Topological polar surface area (TPSA)[14] calculates the area of any drug which has polar characteristics necessary for binding with a polar region of the protein target. This property of the drug ensures proper absorption of the drugs through the plasma membrane if it is going to bind with the protein target inside the cytoplasm.[15,16] The study is equally important for the study of another important property of drugs: cross the blood-brain barrier (BBB).[17]

To keep away from harmful substances reaching the brain there are arrays of closely packed blood vessels surrounding the brain. Water, oxygen, carbon dioxide, and general anesthetics can easily pass into the brain through the blood-brain barrier but this barrier avoids other bigger molecules. Some of the small molecule drugs can cross the barrier and have tremendous potential to serve as anti-tumor drugs to treat brain tumors.[18] Drugs administered through oral routes should have a property of absorption through the human intestine. Many new drugs are screened for better absorption through computational approaches. Specific models for human intestine absorption can greatly contribute towards screening potential compounds. Acute oral toxicity is an important parameter to evaluate the efficacy of any drug. It determines how quickly any drug shows its side effects after consuming it for a single dose or multiple doses. The level of toxicity is measured within 24 hours after consumption of the drugs. It refers to some adverse effects that our body produces in response to the drugs and if there is any effect, the drugs are treated acutely toxic.[19,20] TP 53 genes produce a protein called tumor suppressor protein p53. TP53 acts as a marker gene for cancer diagnosis and their ability to interact with small phytochemicals will ensure different behavior of cancer cells.[21]

Hence our study deals with many investigations (molecular docking, molecular dynamics simulation, molecular mechanics generalized born surface area calculations, ADMET and pass prediction) of small phytochemicals against one of the potential cancer targets: DNA methyl transferase 3-like proteins (DNMT3L).

METHODOLOGY

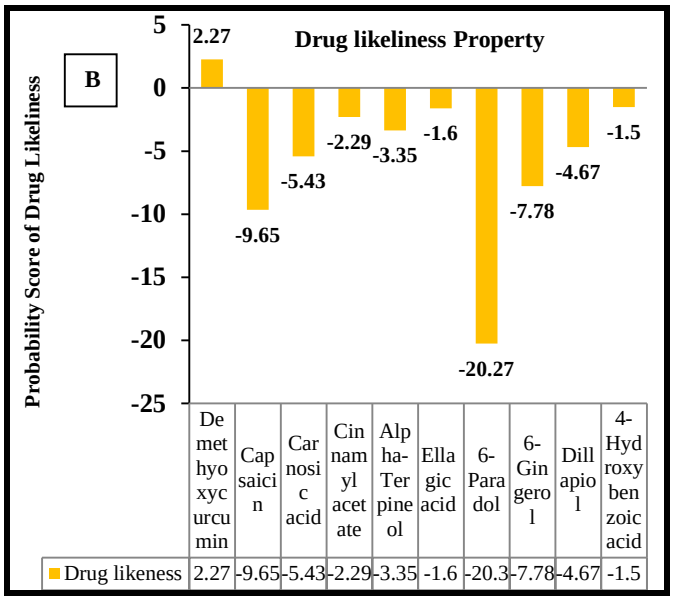
ADMET study

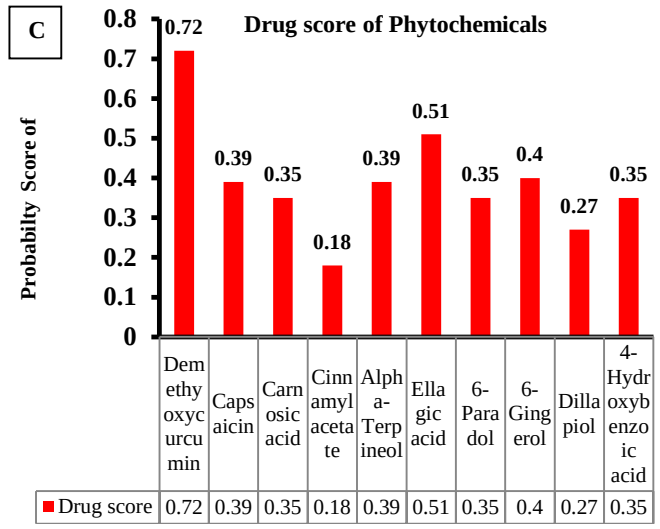
The online software, SWISS ADME and ADMET SAR were chosen to predict ADME and pharmacokinetics descriptors based on physiological parameters, drug-like nature, and medicinal chemistry.[11] All the studied molecules were downloaded from PubChem, and then their canonical SMILES were used as the initial information for the ADMET study using SWISS ADME and ADMET SAR (Table 1). A group of canonical SMILES was fed as input files into this software and then using the set algorithm, various parameters like Clog P, drug likeliness, drug score, TPSA, BBB, HIA, $CaCO_2$, AMES toxicity, carcinogens, and acute oral toxicity have been selected for the AMET study. For parameters like CaCO2, AMES Toxicity, carcinogens, and acute oral toxicity ADMET SAR has been used for all the predictions. All the results were depicted in the form of an excel sheet which shows the probable score of all the phytochemicals against all the set parameters. Later detailed analyses were done by drawing bar graphs against all the parameters (figure 1).

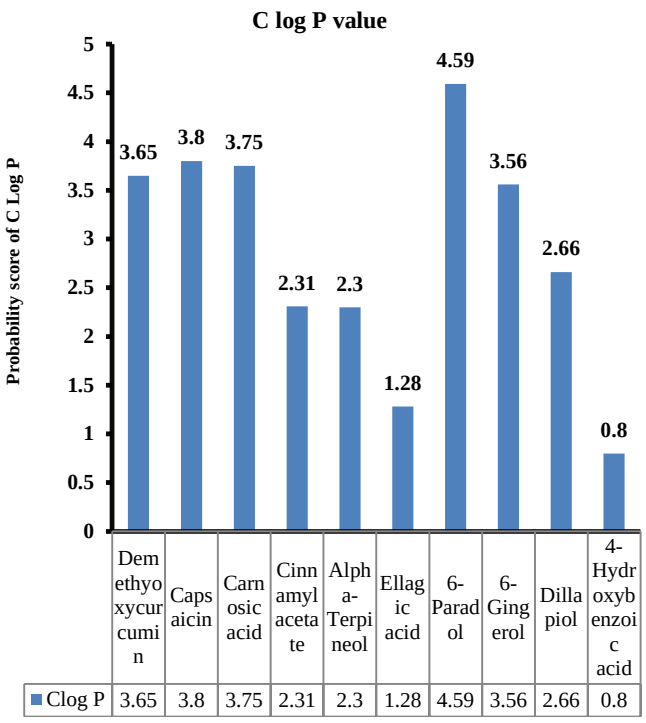
Table 1: Small phytochemicals from different plant species with their SMILE notation

Small phytochemicals	Comp (CID)	SMILE notation	Sources
Ferulic acid	445858	CC(C)(C)[Si](C)(C)OC1=C(C=C(C=C1)C=CC(=O)O[Si](C)(C)C(C)(C)C)OC	*Syzygium Aomaticum*
Crocetin	5281232	CC(=CC=CC=C(C)C=CC=C(C)C(=O)O)C=CC=C(C)C(=O)O	*Crocus sativus*
Cinnamic acid	444539	C1=CC=C(C=C1)C=CC(=O)O	*Cinnamon*
Eugenol	3314	COC1=C(C=CC(=C1)CC=C)O	*Cinnamon*
Cinnamaldehyde	637511	C1=CC=C(C=C1)C=CC=O	*Cinnamon*
Allicin	65036	C=CCSS(=O)CC=C	*Allium sativum*
Alpha tumerone	558173	CC1=CC=C(CC1)C(C)CC(=O)C=C(C)C	*Curcumin Longa*
Curcumin	10134135	CC1=C(C=C(C=C1)C=CC(=O)CC(=O)C=CC2=CC(=C(C=C2)C)[N+](=O)[O-])[N+](=O)[O-]	*Curcumin Longa*
Estragole	8815	COC1=CC=C(C=C1)CC=C	*Ocimum basilicum*

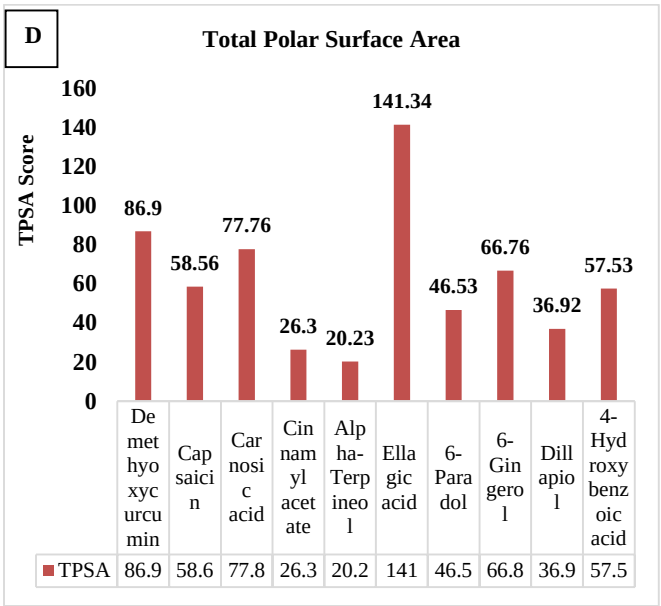
Shogaol	5281794	CCCCCC=CC(=O)CCC1=CC(=C(C=C1)O)OC	*Ginger officinale*
Demethoxy curcumin	5469424	COC1=C(C=CC(=C1)C=CC(=O)CC(=O)C=CC2=CC=C(C=C2)O)O	*Curcumin longa*
Capsaicin	1548943	CC(C)/C=C/CCCCC(=O)NCC1=CC(=C(C=C1)O)OC	*Piper nigrum*
Cinnamyl acetate	5282110	CC(=O)OC/C=C/C1=CC=CC=C1	*Curcumin Longa*
Alpha terpineol	17100	CC1=CCC(CC1)C(C)(C)O	*Elettaria cardamomum*
Ellagic acid	5281855	C1=C2C3=C(C(=C1O)O)OC(=O)C4=CC(=C(C(=C43)OC2=O)O)O	*Syzygium aromaticum*
6-Paradol	94378	CCCCCCCC(=O)CCC1=CC(=C(C=C1)O)OC	*Ginger officinale*
6-Gingerol	442793	CCCCCC(CC(=O)CCC1=CC(=C(C=C1)O)OC)O	*Ginger officinale*
Dillapiol(Di)	10231	COC1=C(C2=C(C=C1CC=C)OCO2)OC	*Foeniculum vulgare*
4-Hydroxybenzoic acid	135	C1=CC(=CC=C1C(=O)O)O	*Coriandrum sativum plant*
Carnosic acid	65126	CC(C)C1=C(C(=C2C(=C1)CCC3C2(CCCC3(C)C)C(=O)O)O)O	*Rosmarinus officinalis*

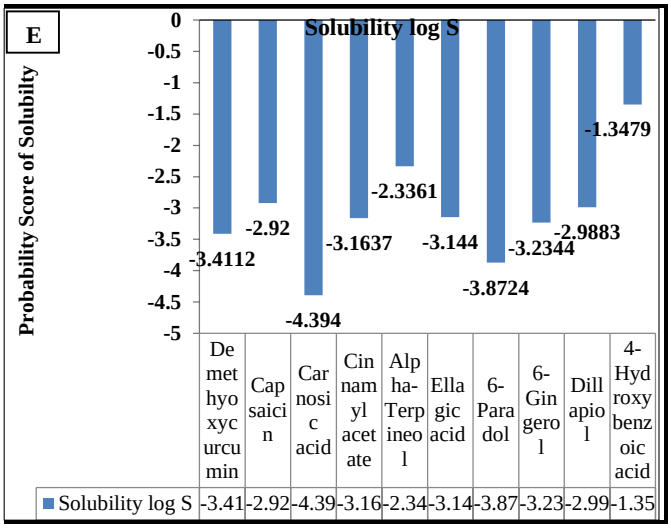


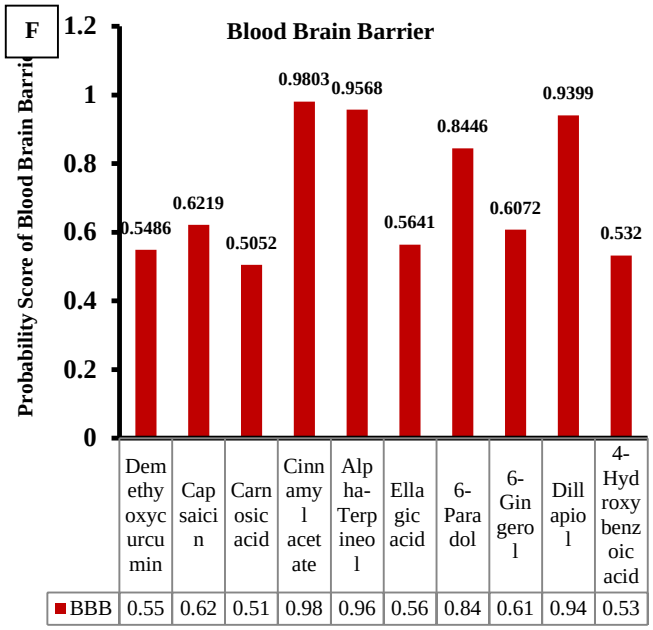


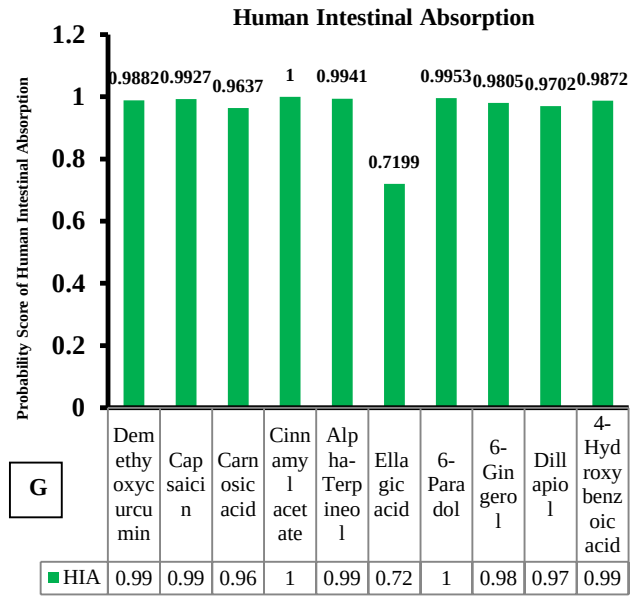


A

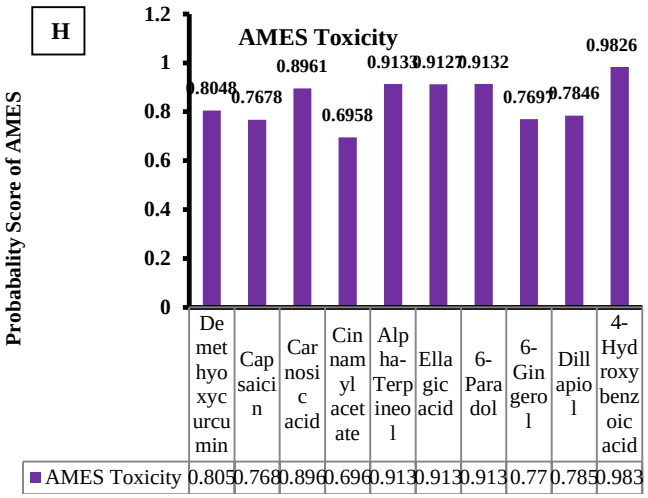


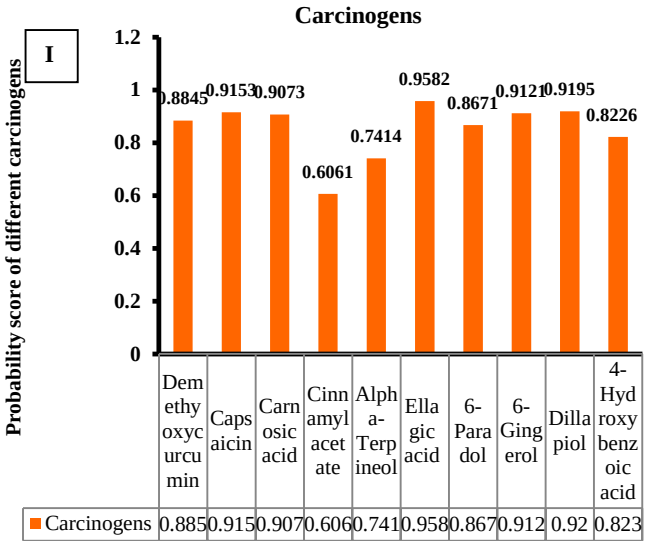






G





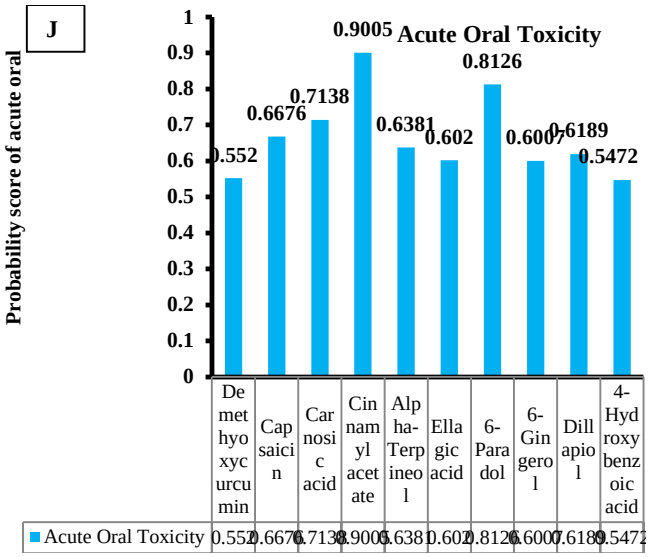


Figure 1: a to j SWISS ADMET analysis of the studied phytochemicals using SWISS ADME software **a.** Calculated Log P prediction **b.** drug likeliness studies **c.** drug score **d.** total polar surface area **e.** aqueous solubility Log S **f.** Blood-brain barrier **g.** human intestinal absorption **h** AMES toxicity assay **i.** carcinogens **j.** acute oral toxicity

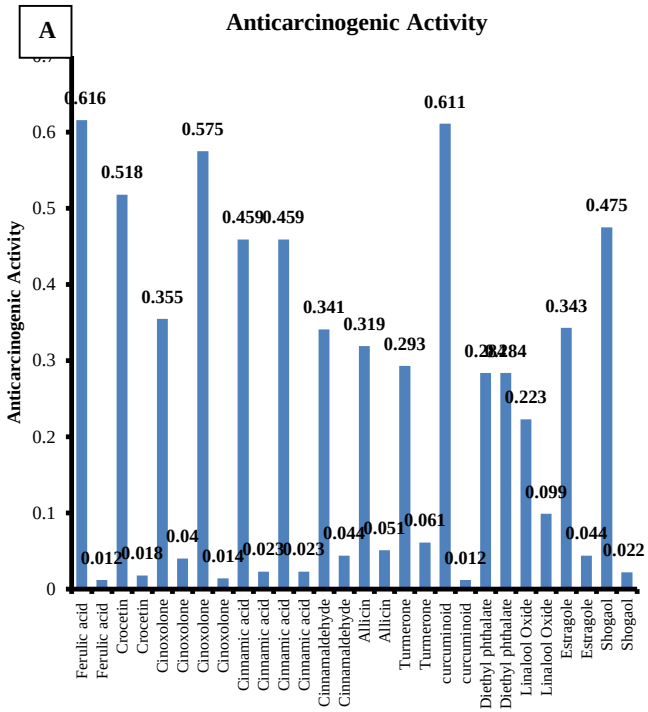
PASS Analysis

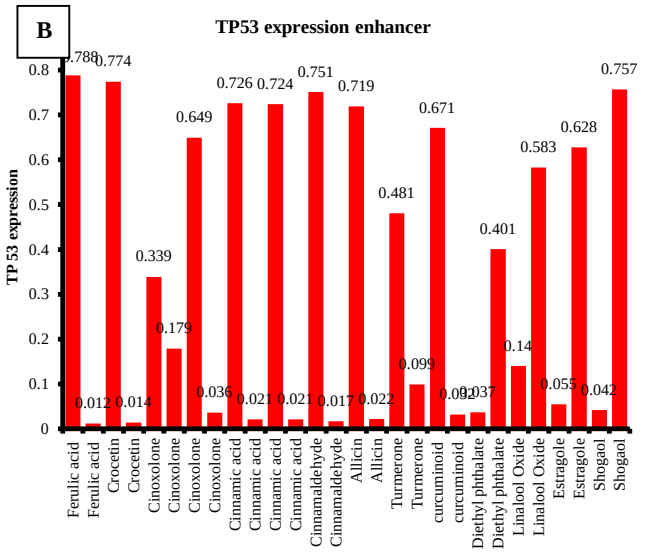
Activity Spectra Prediction for Substances also known as (PASS) is used for the prediction of different physicochemical activities in cancer-defined cells and is based on the molecular formula of organic and biochemical compounds. PASS is a web-based web tool (http://www.way2drug.com/) that has a training set of more than 200000 compounds demonstrating 3750 and more biological activities.[27] The resulting spectrum of PASS is designated by Probable activity (Pa) and Probable inactivity (Pi) scores.[29]

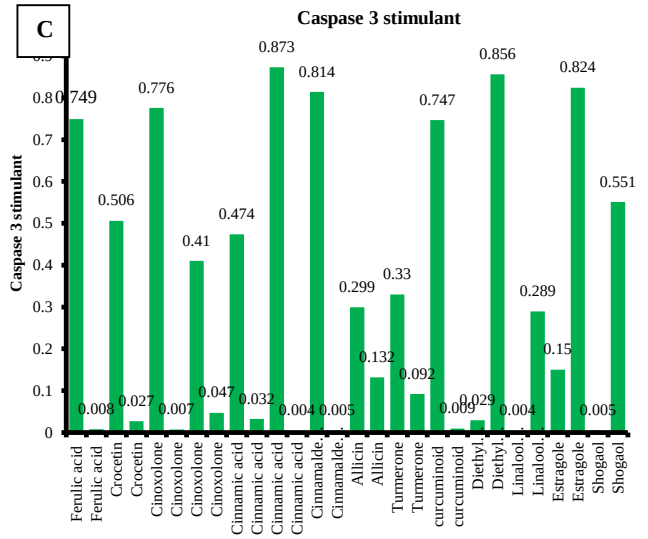
On the basis of the evaluation of different descriptors for potentially active and physiologically non-active compounds from those set values that undergo a defined training, two different possibilities are calculated for every biological activity: *Pa* - the probability of the compound being active and *Pi* - the possibility of whether the molecules are potentially inactive. Molecules and their information are fed one by one into the PASS software where based on the descriptors and parameters out files were generated depicting both Pa and Pi shown in figure 2.

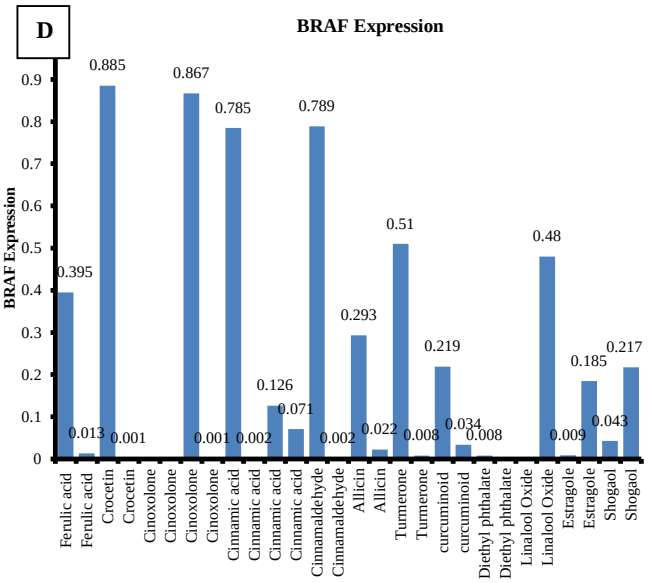
Molecular Docking study

Patch Dock is a molecular docking algorithm and is based on the shape complementarity principle[25] was used to undergo molecular docking procedures of different molecules. Docking procedures allow us to understand the potential binding abilities of the phytochemicals with desired cancer target: DNMT3L. The enzyme DNMT3L was selected as a drug target for cancer due to its beneficial role in DNA methylation. The 3D structure of the enzyme was taken from the RSCB protein data bank (PDB) with ID: 2PVO and with a resolution of 3.30 Å (https://www.rcsb.org/structure/2PV0). All the compounds (small phytochemicals) were downloaded from the PubChem









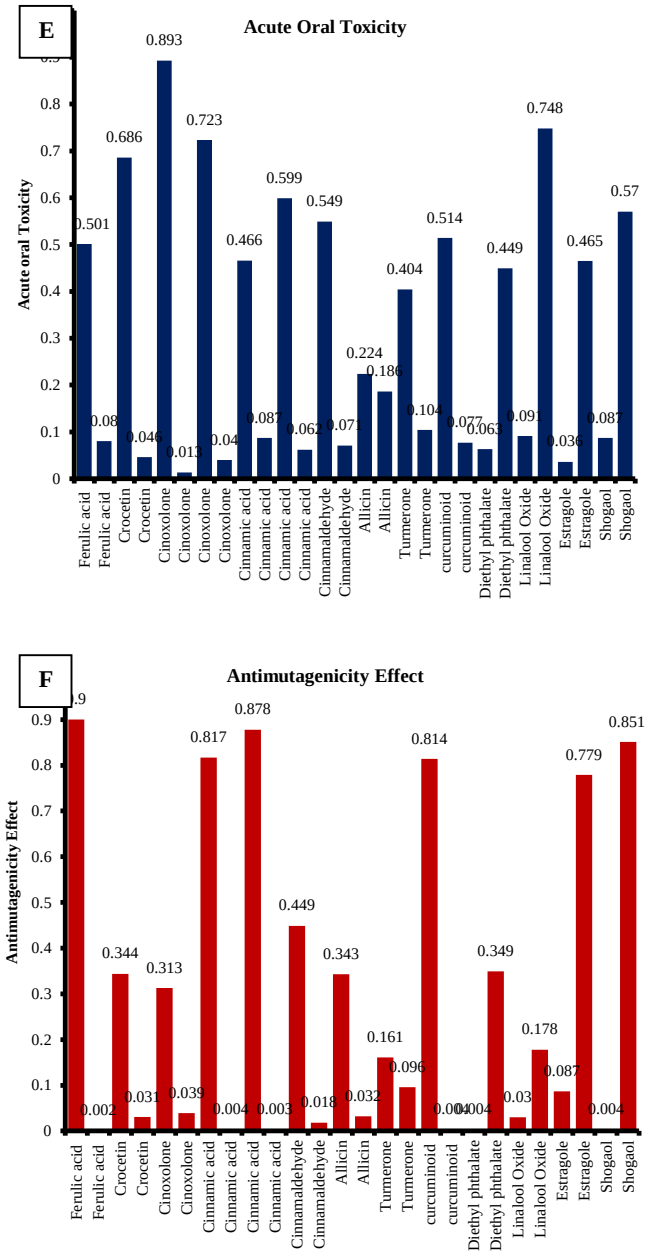


Figure 2 a-f: PASS (Prediction of Activity Spectra for Substances) analysis of studied phytochemicals **a.** Anti-carcinogenic parameters **b.** TP53 expression pattern **c.** caspase 3 stimulant **d.** BRAF expression inhibitor **e.** toxicity studies **f.** antimutagenic studies

Table 2: Small phytochemicals with different binding energy released during interaction with 2PV0

Small phytochemicals	Pub Chem Comp ID (CID)	Binding energy (Global energy) after docking with 2PV0 (kcal/mol)	Sources of these phytochemicals
Ferulic acid	445858	-37.44	*Syzygium aomaticum*
Crocetin	5281232	**-49.27**	*Crocus sativus*
Cinnamic acid	444539	-30.99	*Cinnamon*
Eugenol	3314	-29.72	*Cinnamon*
Cinnamaldehyde	637511	-27.60	*Cinnamon*
Allicin	65036	-33.27	*Allium sativum*
Curcumin	101341351	-38.19	*Curcumin longa*
Estragole	8815	-29.69	*Ocimum basilicum*
Demethoxy curcumin	5469424	**-48.27**	*Curcumin longa*
Capsaicin	1548943	**-42.85**	*Piper nigrum*
Cinnamyl acetate	5282110	-36.53	*Curcumin longa*
Alpha terpineol	17100	-27.65	*Elettaria cardamomum*
Carnosic acid	65126	**-44.72**	*Rosmarinus officinalis*

database (https://pubchem.ncbi.nlm.nih.gov/) in SDF format. Later using an online SMILES translator, SDF format was converted to PDB format which was then used for the docking study against the cancer target. Water and other heteroatoms were removed to avoid unnecessary contact with the phytochemicals. Patch Dock molecular docking allows researchers to screen potential noncompetitive and uncompetitive inhibitors. The output file generated after the docking process was in the form of different interactions between the compound and the target and depicts the binding energy of interactions along different non-covalent associations. Similarly different docking and binding pose of the selected phytochemicals against DNMT3L were visualized first in 2D conformation and then later by 3D conformation. BIOVIA Discovery Studio Visualizer was used in the molecular docking visualization study.[26] Best-docked complexes (DNMLT with phytochemicals) with the highest binding affinities were selected for molecular dynamics (MD) simulations as well as Molecular Mechanics Generalized Born Surface Area (MM GBSA study).

Molecular dynamics (MD) simulation

MD simulation of ligand-protein complexes were carried out using the Desmond module to confirm the binding mode of the ligand.[32] An orthorhombic simulation box of size 10 x 10 x 10 Å3 was used to solvate the system. The 3-site transferable intermolecular potential (TIP3P) water model 33 with a minimum separation of 10 Å was specified between the box wall and the ligand-protein complexes. Additionally, the system was neutralized by adding counter ions (Na^+ or Cl^-), and a physiologically appropriate isosmotic environment was created by supplying 0.15 M NaCl. With a maximum of 2000 iterations

and a convergence requirement of 1 kcal/mol, the constructed system was then subjected to energy minimization utilizing the Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithms with the steepest descent and limited memory.[34] The minimized system was relaxed at 300 K temperature and 1.013 bars atmospheric pressure with a relaxation time of 1 ps and 2 ps, respectively, prior to the production run using the Berendsen NVT and NPT ensemble. Throughout the relaxation technique, the isothermal isobaric ensemble (NPT) was maintained. The Nose-hoover thermostat and the Martyne-Tobias-Klein barostat techniques were used to maintain constant temperature and pressure throughout the whole production run.[35] The reversible reference system propagator algorithms (RESPA) integrator was used throughout the molecular dynamics simulation. The bonding interactions for a time step of 2 fs were calculated in the final production run. The Particle Mesh Ewald (PME) algorithm was used to determine the long-series electrostatic interactions during the simulation.[36] The same protocol was employed for all complexes' molecular dynamic studies. All simulation runs used the OPLS4 all-atom force field with the default settings. A near-credible atomistic simulation of intricate biomolecular systems is made possible by the use of the very precise and consistent molecular mechanics force field known as OPLS4.[37] The simulation interaction diagram was created utilizing the trajectory data that was obtained following the simulation run, such as interactions during simulation time, protein and ligand deviations in the form of root mean square deviation (RMSD) and protein chain fluctuation (RMSF), and the outcomes were examined using several MD simulation protocols. To assess the reproducibility and dependability of the results, a production run of 100 ns with identical parameters was performed for each complex.

Binding free energy calculation (MM-GBSA)

The MM-GBSA (Molecular Mechanics Generalized Born Surface Area) is a tool to compute the ligand binding energy within the continuum solvation model. The conformations of protein-ligand complexes, which spanned for 70 to 100 ns MD run, were used to determine the binding energy (DG bind) by prime modules 38, 39 using the thermal_mmgbsa.py script. The components of DE Molecular Mechanics are estimated based on a classical OPLS molecular mechanics force field. The calculation of free energy of binding occurs as:

$\Delta G_{binding} = \Delta G_{complex} - (\Delta G_{protein} + \Delta G_{ligand})$

Where, $\Delta G_{complex}$, $\Delta G_{protein}$ and ΔG_{ligand} represent the total free energies of the complex, the protein and the ligand respectively.

RESULTS AND DISCUSSION

In silico pharmacodynamics analysis: ADMET analysis

Calculated Log P predicts the highest probability score for 6-paradol, whereas demethoxycurcumin, capsaicin, carnosic acid, and 6-gingerol were showing similar results. The log P value of a compound, s log of its partition coefficient between n-octanol and water log (coctanol/cwater) that measures the hydrophilicity of any compound. Poor absorption or permeation is due to fewer hydrophilicity of the compounds and high Log P values.[12,13] With these contexts, compounds like ellagic acid, 4-hydroxybenzoic acid, cinnamyl acetate, and alpha-terpineol were showing low probability scores and hence has the highest ability to cross the membrane (figure 1a). Drug likeliness study revealed that cinnamyl acetate, alpha-terpineol, ellagic acid, and hydroxybenzoic acid are better drugs than the other compounds. Also carnosic acid and dillapiol are next to these drugs. Although from the parameters like Log P and drug likeliness smallest compound with the least molecular weight is acting as a better drug. The result is shown in figure 1b. Overall drug score proves that demethoxycurcumin has the highest value of probability score far more than any other compound. The next compound is ellagic acid followed by all the other remaining compounds (figure 1c). This study shows that all the selected compounds have a better ability to behave as oral drugs. TPSA is an important parameter that predicts different transportation properties of drug-like molecules. Polar surface area is defined as a summation of surfaces with atoms of polar nature (usually oxygen, nitrogen, and attached hydrogen) in a molecule.[14] This parameter has been shown to very well match with absorption through the human intestine and barrier of blood-brain penetration. Except for alpha terpineol, all the other compounds have values lesser than 140 angstroms squared (Å2), so these compounds have the ability to transport as oral inhibitors so that they can reach and bind with the cancer target. The probability score for TPSA has been shown in figure 1d Parameters like aqueous solubility[15,16] predict that all the compounds show good solubility in a non-polar solvent and are partially soluble in a polar solvent. Figure 1e shows that cinnamyl acetate is highly insoluble in water, whereas 4-hydroxy benzoic acid is partially soluble in water, ether, and acetone. The probability of predicting features like blood-brain barrier[17,18] of compounds like cinnamyl acetate, alpha-terpineol, 6-paradol, and dillapiol is quite high, which shows their ability to cross the barrier of brain blood and may be useful drug candidates for the treatment of neurological and brain disorders (Figure 1f). Results from figure 1g show the ADMET probability score of human intestinal absorption and the result predicts that all the studied have the desired ability to absorb through the human intestine. AMES toxicity[19] revealed from figure 1h shows that most of the compounds showed do not toxicity towards the AMES mutagen test as their values lie below the predicted value of 1 only those compounds whose value lies above 1 will toxicity towards AMES mutagenicity test. AMES mutagenicity test had been predicted using ADMET SAR, the result is for the test for carcinogen shown in figure 1i against all these studied phytochemicals. Predicted values lie below 1 is an indication that these are non-carcinogens. Even the acute oral toxicity results show that these compounds do not show acute oral toxicity as the predictive values were less than one. This result has been depicted in figure 1j.

Prediction of Activity Spectra for Substances (PASS) analysis

This in-silico approach explores new biological activities of selected phytochemicals and thus can contribute to elucidating various mechanisms and related side-effects associated with these molecules. The application of PASS prediction with natural

products has also been demonstrated in some research.[27, 28] The current version of PASS predicts around 3750 different biological modes of action with an accuracy of 95%. 29 Twelve bioactive phytochemicals were chosen for PASS prediction analysis and the result from the anti-carcinogenic study shown in figure 2a revealed that ferulic acid (0.616), cinoxolone (0.575), curcuminoid (0.611) have a very high anti-carcinogenic activity which is followed by crocetin, cinnamic acid and shogaol (values less than 0.5). TP53 activity shown in figure 2b elucidated that except turmerone and diethyl phthalate (values less than 0.5), the rest all the bioactive compounds were showing TP53 values more than 0.5. A high value of TP53 expression is an indication that some of these compounds have a greater ability to activate the expression pattern of TP53, which is shown in the form of probability value. As TP53 acts as tumor suppressor genes that play a vital role in preventing normal cells to become cancer, so activating these genes would lead to the stopping of a normal cell to become cancerous.[24]

Other parameters like caspase 3 stimulant shown in figure 2c elucidated that cinoxolone, cinnamic acid, allicin, turmerone, and linalool oxide are having probability score values less than 0.5. A higher caspase 3 value suggests the effective role of these phytochemicals to activate this enzyme for undergoing apoptosis of cancer cells. Caspase 3 stimulant is one of the essential features in Pass Prediction software that initiates the process of cell apoptosis in the cancer-affected region.[25]

The result of BRAF depicted in figure 2d suggested that out of all phytochemicals under study, ferulic acid, cinnamic acid, allicin, curcuminoid, diethyl phthalate, linalool oxide, estragole, and shogaol were having the least potential to behave as inhibitors against BRAF. BRAF plays an important role in cancer signaling and is now regarded as a potential biomarker of cancer.[26]

Many phytochemicals from the chosen plant species for our study reveal that these can also be slightly toxic as far as cancer cells are concerned. Cinoxolone, crocetin, cinnamic acid, cinnamaldehyde. Circuminoid and linalool oxide had fewer probability values (less than 0.5), suggesting their behavior toward a toxic role (Figure 2e). Many studied phytochemicals like crocetin, cinoxolone, allicin, turmerone, diethyl ether, and linalool oxide were having pass prediction values less than 0.5 and thus suggest their role as non-anti-mutagenic substances shown in figure 2f, whereas those having higher probability score were regarded as anti-mutagenic.

Molecular docking analysis

In order to understand the binding and inhibition capability of all the proposed phytochemicals given in table 1, DNMT3L was chosen with PDB id: 2PV0. Patch Dock web server[30,35] was chosen for the docking study. Molecular docking study reveals their strong binding ability with the target. Functional groups like the oxygen and other electronegative atoms present in the phytochemicals are making a hydrogen bond with the hydrogen atom of the amino acid residues present on the target enzyme. Apart from the hydrogen bonding interaction, these compounds also make other non-covalent interactions (NCI) with the target (figure S3 a-m (Supplementary file). The result of binding energies of the molecular docking interaction between small phytochemicals and DNMT3L have been shown in Table 2. Allicins make hydrogen bonding with GLY 236 and VAL 232 apart from showing other non-covalent interactions (NCI) with LEU 193, ARG 191, LEU266, and TRP 235 (Figure S3a).

Cinnamaldehyde makes only NCI other than hydrogen bonding with PHE 238, LEU 193, and LYS 219. During this process of interaction of the compound with the target, binding of -33.27 kcal/mol and -27.0 kcal/mol are released (Shown in figure S3a and S3b) from these compounds respectively. Even cinnamic acid, which is a derivative of cinnamaldehyde makes one hydrogen bond with VAL 184 and ARG 161 and NCI with LEU 144. The result is depicted in figure S3d. Cinnamyl acetate another derivative of the same compound makes only NCI with LEU 240, ILE 327, and MET 283 (figure S3f). This compound does not show any hydrogen bonding with the target. The binding energies associated with these two compounds are -30.99 and -36.53 kcal/mol respectively. Different isoforms of DNMT3L have been used as a cancer target to showcase the docking experiment. Similar kinds of molecular docking procedures have been performed for Alzheimer's disease.[32-34]

Molecular docking with Alpha terpeniol (figure S3c) from the plant species Elettaria cardamomum makes only NCI with LEU 144, TRP 185, PRO 183, TYP 160, CYS 113, and CYS 145. This compound makes hydrogen bonding with LEU 144. LEU 144 is involved in various types of interaction with the phytochemical. The binding energy released during the process is -27.65 kcal/mol.

Capsaicin another useful bioactive ingredient found in the Piper nigram interacts with ALA 83, CYS 100, LEU 67, LEU 164, and PRO 66 through NCI. A binding energy of -42.85 kcal/mol is released during this process. The visualization result has been shown in figure S3e. This amount of binding energy is considered very high and proves its ability to act as a strong inhibitor against the target. In a similar pattern of high binding energy, other compounds like crocetin with a binding energy of -49.27 kcal/mol (figure S3m), demethoxycurcumin (-48.27 kcal/mol) shown in figure S3j, carnosic acid (-44.72 kcal/mol) depicted in figure S3g have been depicted through the molecular docking interaction. All these phytochemicals show greater binding energy above (-40 kcal/mol) which shows their strength as strong inhibitors against DNMT3L

Crocetin (figure S3m) is making NCI with VAL 192, MET 283, ALA 244, VAL 285, PRO 360, VAL 364, and VAL 321. Valine, proline alanine, and methionine all non-polar amino acids form a small pocket in the interior of DNA methyl transferase 3-like protein. The number of more interactions (seven) with the crocetin and also the total binding energy released during the docking interaction is very high (-49.27 kcal/mol). Thus the structure of crocetin allows it to fit properly into the groove of the target protein and might act as a potential inhibitor against DNMT3L. Similarly, demethoxycurcumin (figure S3j) also makes NCI with LEU 371, PHE 208, LEU 240, ILE 327, GLU 197, and ALA 244.

Carnosic acid interacts with PHE 208, LEU 203, ILE 199, MET 283, VAL 285, and VAL 364 (figure S3g) through various

non-covalent interactions. All the interacting amino acids are non-polar and reside within the core of the target protein and form a pocket. Thus from the docking study, it may be concluded that carnosic acid goes deep inside the target to form a non-covalent association and releases a very binding energy of 44.72 kcal/mol. In a similar line with crocetin, the structure of carnosic acid allows the molecule to fit properly into the pocket of the target protein. Estragole interacts with ARG 161, PRO 183, CYS 145, and TRP 160.[33] Estragole (shown in figure S3h) makes special van der Waals interaction with ARG 161 and apart from that non-covalent interaction with PRO 183, CYS 145, TRP 160, and curcumin is shown in figure S3i makes non-covalent interaction with CYS 145, LEU 144, TRP 160, ARG 115, CYS 113, ARG 161 and LYS 159. Eugenol present in the plant species cinnamon makes two hydrogen bonds with VAL 364, and PHE 368 of the target and NCI with CYS 367, VAL 323, MET 283, and VAL 285(figure S3k), similarly, ferulic acid also makes four hydrogen bonding with the target through amino acids, MET 283, GLY 243, VAL 285 and VAL 364. This compound also makes various NCI with CYS 367 and ILE 199 (figure S3l). Based on the molecular docking analysis, four molecules (carnosic acid, capsaicin, crocetin, and demethoxycurcumin) with better binding affinities have been selected for further analysis. These molecules were taken for molecular dynamics analysis for 100 ns in order to understand the stability of docked complexes at the binding pocket of DNMT3L.

MD simulation analysis

Proteins are dynamic in nature, and their motion is usually essential for their function. The study of their dynamic behavior supports a deeper understanding of their structural and functional role. A more realistic and detailed analysis of protein motion requires molecular dynamics (MD) simulations. Moreover, MD simulations also permit an accurate interpretation of the protein-ligand complex's mode of binding and stability. Further, MD simulation studies of 100 ns time were done to assess the strength and dynamic behavior of the selected docked complexes (capsaicin, crocetin, demethoxycurcumin, and carnosic acid) in the active site of an enzyme. After the production run, the generated trajectories were examined using the simulation interaction diagram to analyze the root mean square deviation (RMSD), root mean square fluctuations (RMSF), and protein-ligand interaction behavior of each docked complex.

The RMSD calculated by initially aligning all protein frames on the reference backbone frame indicated that the backbone Cα atoms were stable during the simulation time period with a deviation range of around 2-3 Å (Figure S4-a, b, c, and d). The MD simulation results were also employed to investigate the structural changes of all the protein-ligand complexes during the MD simulation. To reveal the structural changes, we extracted the structure at 100 ns simulation time of each dock complex.

The RMSD behavior of the carnosic acid complex maintained a stable behavior during the entire simulation with a mean value of 2 Å. Initially, the complex displayed deviations up to 20ns, and thereafter, the trajectories stabilized. In the crocetin complex, a stable behavior was observed after 40 ns throughout the simulation time. Other complexes (capsaicin and demethoxycurcumin) showed unestablished behavior throughout the simulation. Thus, RMSD trajectory analysis of all the simulations reveals that amongst all the screened compounds, the trajectory for compound carnosic acid was the most stable with variations within the acceptable range followed by crocetin complex. Some studies using potential phytochemicals were also observed to show average RMSD values within the 1- 3 Å range. Thus, lower RMSD values of our protein-ligand trajectory also validated the stability of molecular docking poses. 39 The estimated RMSD value after superimposition was observed for carnosic acid (0.5 Å), crocetin (2 Å), capsaicin (2.5-3.5 Å) and demethoxycurcumin (3.0-3.5 Å) respectively. The compounds at 100 ns, carnosic acid, crocetin, capsaicin, and demethoxycurcumin on superimposition did not reveal any significant conformational changes in the protein. Thus, no remarkable structural changes were observed in all the docked complexes

We further performed root mean square fluctuations (RMSF) analysis of the identified hit complexes (carnosic acid, crocetin, demethoxycurcumin, and capsaicin,) to assess the oscillation of each residue from its average mean position. This indicates the extent of atomic variations related to the structural constancy of molecular interaction during simulation. Generally, a higher value of RMSF is observed for loops, and the terminal residues are considered flexible, while lower values designate a more rigid conformation.[38,39] The complex did not reveal any significant RMSF changes compared to the other complexes' trajectories except residue range 170-180 (figure S5 a, b, c, and d). The RMSF plot displayed the slightest fluctuations at the binding pocket residue range 170-180 for all docked complexes with a flexible range of 1.5-4.0 Å.

These results suggest that the complexes formed with carnosic acid and crocetin were robust and hence, these have been considered as a la lead molecule in the protein-carnosic acid complex, the ligand is stabilized by a hydrogen bond formed by residues TYR377, and hydrophobic interactions formed by ILE 199 and LEU371 between the ligand and the protein. In the protein-crocetin complex, the ligand is stabilized by two hydrogen bonds formed by residues ASN287, and LYS358 hydrophobic interactions formed by ILE 199 and VAL285, LYS354, ALA356, and ALA357 of the protein. This suggests that the carnosic acid and crocetin were tightly bound to the protein in the binding pocket. The interactions between protein and ligand execute a significant role in governing ligand stability. In a carnosic acid complex and crocetin complexes the effective interactions of the compound, with the catalytic residues were retained during the whole simulation time. These involved hydrogen-bonded and non-covalent interactions between the ligand and the protein residues (Figure S6 a, b, c, d, e, and f). The protein-ligand interaction plot delivers a well-designed conformation of the spatially tightly bound ligand in the docked complex at the protein binding site. The protein complexes with the other two ligands, capsaicin, and demethoxycurcumin did not showcase good stable interactions within the binding pocket as indicated by their higher RMSF and RMSD and diminished significant interactions. Thus, these were excluded from further

analysis. This finding also correlates with our docking and MD simulation results where the screened compounds were found to interact with key negatively charged residues of the binding site. All these four molecules were further taken for MM-GBSA analysis for further confirmation of the MD simulation study.

Binding free energy analysis

The binding free energy (DG) was assessed using the prime MM-GBSA method to reveal each ligand's relative binding-free energy (DG bind) in each docked complex. Prime MM-GBSA estimates the energy of enhanced free protein structure, free ligand, and docked protein-ligand complex (ref). The MM-GBSA-based free binding energy of the top four docked complexes was calculated from the trajectories spanning 70 to 100 ns (Table 3S). The compound carnosic acid demonstrated free binding energy of -64.69 kcal/ mol, whereas compounds, crocetin capsaicin, and demethoxycurcumin showed binding energy of -51.85, -48.93, and -41.95 kcal/mol respectively, suggesting that compound carnosic acid has the highest negative binding affinity followed by compound crocetin. These two molecules were screened (carnosic acid and crocetin) from the selected four molecules (carnosic acid, crocetin, capsaicin, and demethoxycurcumin). During the interaction with DNMT3L (100 ns MD simulation and MM-GBSA analysis) carnosic acid and crocetin form the most stable complexes with the non-polar amino acids of the target. The study also confirms that the chosen target has potential sites to accommodate these compounds. These interactions might alter the cell's signaling process during cancer initiation.

CONCLUSION

In order to suggest small phytochemicals from different plant species as anti-cancerous compounds against DNA methyl transferase 3-like protein, we performed molecular docking, molecular dynamics, free binding energy calculations, ADMET, and PASS analysis. Finally, based on the results of the 100 ns molecular dynamics simulation and the MM-GBSA binding energy analysis, we have filtered out carnosic acid and crocetin as the lead molecules against the selected cancer target. Consequently, it is clear that the enzyme possesses a potential binding pocket to accommodate smaller phytochemicals. Carnosic acid and crocetin both have good solubility scores as well as they can pass through the blood-brain barrier and cell membrane. Hence both these compounds seem to be proposed as better anti-cancerous compounds than the twelve other studied phytochemicals. Both being tiny natural molecules with diversities in the scaffolds, do not have any on the living cell, designing of synthetic drugs can be taken forward.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

ACKNOWLEDGMENT

The authors would like to thank Sharda School of Basic Sciences & Research, Sharda University, Greater Noida, UP, India, and also like to thank Dr. Mukesh Kumar, AIIMS, New Delhi for his valuable suggestions and comments on the manuscript

SUPPLEMENTARY INFORMATION FILE

Figure 1 (enlarge view), Figure 2 (enlarged view), Figure 3-6 and Table 3 are provided as supplementary file and can be obtained from article page.

REFERENCES

1. B.S. Chhikara, K. Parang. Global Cancer Statistics 2022: the trends projection analysis. *Chem. Biol. Lett.* **2023**, 10 (1), 451.
2. A. Gnyska, Z. Jastrazebski, S. Flis. DNA Methyltransferase Inhibitors and Their Emerging Role in Epigenetic Therapy of Cancer. Anticancer Res. **2013**, 33, 2989-2996.
3. Cheng, C. He, M. Wang, X. Ma, F. Mo, S. Yang, J. Han, X. Wei. Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. Signal Transduct. Target. Ther. **2019**, 4, 62.
4. Chedin , M.R. Lieber, C.L. Hsie. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc Natl Acad Sci. **2002**, 24, 99, 26.
5. Mujeeb, P. Bajpai, N. Pathak. Phytochemical Evaluation, Antimicrobial Activity, and Determination of Bioactive Components from Leaves of Aegle marmelos, Biomed Res Int. **2014**, 497606
6. Hadacek, G. Bachmann. Low-molecular-weight metabolite systems chemistry. Front. Environ. Sci. **2015**, 3, 12.
7. S. Teoh. Secondary Metabolites of Plants. Medicinal Orchids of Asia. **2015**, 5, 59–73.
8. Zhong, Y. Li, L. Xiong, W. Wang, M. Wu, T. Yuan, W. Yang, C. Tian, Z. Miao, T. Wang, S. Yang. Small molecules in targeted cancer therapy: advances, challenges, and future perspectives. Signal Transduct. Target. Ther. **2021**, 6:201.
9. X. Li, E. Shang, Q. Dong, Y. Li, J. Zhang, S. Xu, Z. Zhao, W. Shao, C. Lu, Y. Zheng, H. Wang, X. Lei, B. Zhu, Z. Zhang. Small molecules capable of activating DNA methylation– repressed genes targeted by the p38 mitogen-activated protein kinase pathway. J. Biol. Chem. **2018**, 293, 19, 7423–7436.
10. Daina, O. Michielin, V. Zoete1. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Nature Scientific Reports. **2017**, 7, 42717.
11. A D. Masi. Trends in Drug Development Costs. Times, and Risks. **1995**, 29, 2, 375-384.
12. T. Savjani, A. K. Gajjar, J. K. Savjani. Drug Solubility: Importance and Enhancement Techniques. ISRN Pharm. **2012**, 195727.
13. A Lipinski. Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol. **2004**, 1, 4, 337-41
14. Prasanna , R. J. Doerksen. Topological Polar Surface Area: A Useful Descriptor in 2D-QSAR. Curr Med Chem. **2009**, 16,1, 21–41.
15. N. Wang, C. Huang, J. Dong, Z.J. Yao, M. F. Zhu, Z.K. Deng, A. F. Chenac, D. S. Cao. Predicting human intestinal absorption with modified random forest approach: a comprehensive evaluation of molecular representation, unbalanced data, and applicability domain issues. RSC Adv. **2017**, 7, 19007.
16. Hou. ADME Evaluation in Drug Discovery and the Prediction of Human Intestinal Absorption by a Support Vector Machine. J. Chem. Inf. Model. **2007**, 47, 2408-2415.
17. Daneman, A. Prat. The Blood–Brain Barrier. Cold Spring Harb Perspect Biol. **2015**.
18. M. Larsen, D. R Martin, M. E Byrne. Recent advances in delivery through the blood-brain barrier. Curr Top Med Chem. **2014**, 14, 9, 1148-60.
19. Walum. Acute Oral Toxicity. Environ. Health Perspect. 1998, 106, 2.
20. Saleema, S. Aminb , B. Ahmadb , H. Azeemb , F. Anwarb , S. Maryb. Acute oral toxicity evaluation of aqueous ethanolic extract of Saccharum munja Roxb roots in albino mice as per OECD 425 TG, Toxicology Reports. **2017**, 4, 580-585.
21. Wang, Q. Sun. TP53 mutations, expression and interaction networks in human cancers. Oncotarget. **2017**, 8, 1, 624-643

22. Huang, F. Li, X. Liu, W. Li, W. Shi, F.F. Liu, B. O. Sullivan, Z. He, Y. Peng, A. C. Tan, L. Zhou, J. Shen, G. Han, X.J. Wang, J. Thorburn, A. Thorburn, A. Jimeno, D. Raben, J. S. Bedford, C.Y. Li. Caspase 3-mediated stimulation of tumor cell repopulation during cancer radiotherapy. Nat Med. **2015**, 17, 7, 860–866.
23. Proietti, N. Skroza, S. Michelini, A. Mambrin, V. Balduzzi, N. Bernardini, A. Marchesiello, E. Tolino, S.Volpe, P. Maddalena, M. D. Fraia, G. Mangino, G. Romeo, C. Potenza. BRAF Inhibitors: Molecular Targeting and Immunomodulatory Actions, Cancers. **2020**, 12, 1823.
24. A Filimonov, A.A. Lagunin, T.A. Gloriozova, A.V.Rudik, D.S. Druzhilovskii, P.V. Pogodin, V.V. Poroikov. Prediction of the biological activity spectra of organic compounds using the PASS online web resource. Chem. heterocycl. **2014**, 50, 3, 444-457.
25. S. Duhovny, Y. Inbar, R. Nussinov, H. J. Wolfson. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. **2005**, 1; 33.
26. BIOVIA, Dassault Systèmes, BIOVIA Workbook, Release **2020**; BIOVIA Pipeline Pilot, Release 2020, San Diego: Dassault Systems
27.Subramaniam, R. Thombre, A. Dhar , S. Anant. DNA methyltransferases: a novel target for prevention and therapy. Front. Oncol. **2014**, 4, 1-13.
28. Sharadchandra, K. Vrushali, S. Tambe, A. Dnyaneshwar, W. Prajakta, B. Kothawade. A Comparative Molecular Docking Study of Crocetin with Multiple Receptors for the Treatment of Alzheimer's Disease. Biomedical and Biotechnology Research Journal. **2022**.
29. Kanwal, M. Datt4, X. Liu, S. Gupta. Dietary Flavones as Dual Inhibitors of DNA Methyltransferases and Histone Methyltransferases. PLOS one. **2016** 0162956
30. Z. Sibuh, S. Khanna, P. Taneja, P. Sarkar, N. K. Taneja. Molecular docking, synthesis and anticancer activity of thiosemicarbazone derivatives against MCF-7 human breast cancer cell line. Life Sciences. **2021**, 273, 119305.
31. Cheng, W. Li, Y. Zhou, J. Shen, Z. Wu, G. Liu, P. W. Lee, Y. Tang, admetSAR: A Comprehensive Source and Free Tool for Assessment of Chemical ADMET Properties, J. Chem. Inf. Model. **2012**, 52, 11, 3099–3105
32.Shaw, D. Desmond Molecular Dynamics System; V 3.0. Schrödinger: New York, NY, USA, **2011**
33.Price, D.J.; Brooks, C.L. A Modified TIP3P Water Potential for Simulation with Ewald Summation. J. Chem. Phys. **2004**, 121, 10096–10103. https://doi.org/10.1063/1.1808117.
34. Kaczor, A. A., Targowska-Duda, K. M., Patel, J. Z., Laitinen, T., Parkkari, T., Adams, Y., Nevalainen, T. J., & Poso, A.. Comparative molecular field analysis and molecular dynamics studies of α/β hydrolase domain containing 6 (ABHD6) inhibitors. Journal of molecular modeling, **2015**, 21(10), 250. 2789-8
35. Kumar, M.; Roy, A.; Rawat, R.S.; Alok, A.; Tetala, K.K.R.; Biswas, N.R.; Kaur, P.; Kumar, S. Identification and Structural Studiesof Natural Inhibitors against SARS-CoV -2 Viral RNA Methyltransferase (NSP16). *J.* Biomol. Struct. Dyn.
36. Kratz, E. G., Duke, R. E., & Cisneros, G. A. Long-range electrostatic corrections in multipolar/polarizable QM/MM simulations. Theoretical Chemistry Accounts, **2016**, 135(7), 166.
38. Lu, C., Wu, C., Ghoreishi, D., Chen, W., Wang, L., Damm, W., Ross, G. A., Dahlgren, M. K., Russell, E., Von Bargen, C. D., Abel, R., Friesner, R. A., & Harder, E. D, OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. Journal of chemical theory and computation, **2021**, 17(7), 4291–4300.
39.Kollman, P.A. ChemInform Abstract: Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. ChemInform, **2010**, 33.
40. Kumar, M., Tripathi, M. K., Gupta, D., Kumar, S., Biswas, N. R., Ethayathulla, A. S., & Kaur, P, N-acetylglucosamine-phosphatidylinositol de-N-acetylase as a novel target for probing potential inhibitor against Leishmania donovani. Journal of biomolecular structure & dynamics, **2022**, 1–15.