Identification of Potential drug targets against lung cancer from Hamamelis virginiana using high performance computational techniques

Sonam Kundral
Galgotias College of Pharmacy, Greater Noida, (Uttar Pradesh)
Puneet Kumar
Galgotias College of Pharmacy, Greater Noida, (Uttar Pradesh)
Vikram Sharma
Galgotias College of Pharmacy, Greater Noida, (Uttar Pradesh)
Sangh Partap
sangh.pratap@galgotiacollege.edu
Galgotias College of Pharmacy, Greater Noida, (Uttar Pradesh)

Research Article

Keywords: Hamamelis, Imppaat, PubChem, EGFR, Cianidanol, Lung cancer

Posted Date: April 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9341649/v1

Additional Declarations: No competing interests reported.

Identification of Potential drug targets against lung cancer from *Hamamelis virginiana* using high performance computational techniques

Authors:

Sonam Kundral[1,2], Puneet Kumar[1], Vikram Sharma[1], Sangh Partap[1*]

[1]Department of Chemistry, Galgotias College of Pharmacy, Greater Noida, (Uttar Pradesh), 201310, India

[2]Department of pharmaceutical chemistry, Delhi Institute of Pharmaceutical sciences and research, DPSRU, New Delhi, 110017, India.

Corresponding Author:

Dr. Sangh Partap

Department of Chemistry
Galgotias College of Pharmacy,
Greater Noida, (Uttar Pradesh), 201310, India
E-mail: sangh.pratap@galgotiacollege.edu
ORCHID : 0000-0002-6361-8890

Abstract

This in-depth investigation employs network pharmacology and molecular docking methodologies to analyse various phytocompounds extracted from the *hamamelis* plant, with a specific focus on uncovering potential drug targets relevant to lung cancer treatment. The primary aim is to pinpoint therapeutic targets that could be utilized in combating this malignancy. The study leverages established databases, notably IMPPAT and PubChem, to systematically screen nine distinct phytoconstituents, assessing them against critical criteria for drug-likeness and bioavailability, vital indicators of their therapeutic viability. Through this rigorous screening, essential protein targets linked to lung cancer were identified and ranked based on their corresponding scores reflecting their significance in the context of the disease. The proteins EGFR emerged as particularly noteworthy due to their high association scores, underlining their potential as crucial targets in lung cancer therapeutics. Following this phase, molecular docking experiments were performed, accompanied by dynamic simulations of the chosen ligands. The results highlighted strong binding affinities and illustrated the stability of the resultant protein-ligand complexes over extensive simulated time frames, extending up to 100 nanoseconds. These findings substantiate the therapeutic promise of the identified compounds within lung cancer treatment paradigms. Moreover, calculations of binding free energy, reinforced the beneficial interactions and electronic stability of certain compounds, notably cianidanol. This integrative and nuanced approach adeptly underscores the most promising bioactive constituents derived from ***Hamamelis***, suggesting their significant potential in the future development of targeted therapies for lung cancer.

Keywords: Hamamelis, Imppaat, PubChem, EGFR, Cianidanol, Lung cancer

Introduction

Lung cancer remains one of the most prevalent malignancies worldwide, accounting for approximately 30% of all female cancers and representing a significant global health burden with over 2.3 million new cases diagnosed annually (1,2). Despite substantial advances in therapeutic interventions, including targeted therapies and immunotherapies, the development of drug resistance, severe side effects, and high treatment costs continue to challenge effective breast cancer management. Natural products, particularly phytochemicals derived from medicinal plants, have emerged as promising sources for discovering novel anticancer agents due to their structural diversity, multi-target mechanisms, and relatively lower toxicity profiles compared to synthetic drugs. Hamamelis virginiana, commonly known as witch hazel, is a

deciduous shrub native to eastern North America that has been extensively utilized in traditional medicine for its anti-inflammatory, astringent, and wound-healing properties. The bark and leaves of H. virginiana are particularly rich in bioactive compounds, including condensed tannins, hydrolyzable tannins such as hamamelitannin, gallic acid derivatives, and flavonoids, which have demonstrated significant cytotoxic activity against various cancer cell lines, including colon and other malignancies. Recent studies have highlighted the selective anticancer potential of witch hazel phytocompounds, with hamamelitannin showing specific cytotoxicity against colon cancer cells while sparing normal colonocytes and displaying potent antioxidant properties that protect against DNA damage (3,4). The emergence of network pharmacology as a systems biology approach has revolutionized drug discovery by enabling the comprehensive analysis of drug-target-disease interactions, allowing researchers to predict potential therapeutic targets and understand the multi-target mechanisms underlying phytochemical bioactivity. However, despite the well-documented anticancer properties of H. virginiana compounds in other malignancies, their specific therapeutic potential against breast cancer remains largely unexplored, presenting a significant research gap that warrants systematic investigation using modern computational and experimental approaches to identify novel drug targets and elucidate the molecular mechanisms underlying their anticancer effects (5).

Materials and Methods

1.1. Identification of drug like phytocompounds

Data regarding active phytocompounds of *Hamamelis virginiana* species were obtained from the Indian Medicinal Plants, Phytochemistry And Therapeutics 2.0 (IMPPAT 2.0) database, a well-researched database assembled by digitizing data from a large inventory of resources, including over 100 books on traditional Indian medicine and over 7,000 published research papers (4,5). The structural information of the phytoconstituents was obtained from PubChem, a freely accessible database maintained by the National Centre for Biotechnology Information (NCBI) (6,7). The phytoconstituents were evaluated on the basis of their predicted pharmacokinetic properties that were accessed from Swiss ADME a web-based pharmacokinetics predictor, drug-likeness and medicinal chemistry friendliness predictor for small molecules (8,9). This screening process included assessments of oral bioavailability, where only those compounds exhibiting bioavailability above 30% were considered. Additionally, compliance with Lipinski's Rule of Five was a critical criterion for identifying

drug-like candidates. Compounds satisfying these pharmacokinetic and physicochemical parameters were subsequently advanced for further in-depth analysis, ensuring their suitability for therapeutic targeting (8).

1.2. Identification of disease-based targets

In order to find out the most relevant and commonly linked targets of breast cancer, a systematic search was conducted involving several databases such as Open Targets, Gene Cards and the Comparative Toxicogenomic Database (CTD). These databases were searched using the words "Breast cancer" and "Homo sapiens" so that human specific data would be included. By combining data from experimental research, scientific publications, and computational predictions, these websites are valuable tools for the discovery of putative therapeutic targets for human diseases. Genes that were associated with lung cancer with a prediction score higher than 50% were chosen for further screening (9,10,11).

1.3. Identification of structure-based targets

To identify potential structure-based targets for ***Hamamelis*** phytocompounds, a combination of bioinformatics tools such as STITCH, Target-Net, and Swiss-Target Prediction was employed. These tools provide a computational approach to predict interactions between small molecules and protein targets, facilitating the discovery of key biological pathways and potential drug targets. The duplicates were removed, and the final target list was used for further analysis (12,13,14).

1.4. Identification of unique targets

The disease-based targets and the structure-based targets for ***Hamamelis*** phytoconstituents were merged to establish common protein targets, thus establishing an intersectional target database. A Venn diagram was created to visualize this overlap. Multiple predictive methods having common targets increase the chances of selecting effective therapeutic prospects in drug discovery. By comparing targets identified through both disease-based and structure-based approaches, researchers can focus on those that are consistently implicated to make them more relevant for future validation and further research (14).

1.5. Construction of genes interaction network

The common genes' interaction network, identified in the above step, was built with the Gene-MANIA server a bioinformatics tool that can predict gene functional associations. Gene-

MANIA compiles and processes various genomic and proteomic data sources to create relationships between genes based on physical interaction, co-expression, genetic interactions, shared protein domains, co-localization, pathway membership, and predictions based on computational methods. Through utilization of this multi-dimensional data, the derived gene network offers insights into the functional connectivity and biological significance of the chosen targets, thus facilitating the identification of essential nodes for future investigation in drug development (15).

1.6. Construction of protein-protein network

Protein-protein interaction (PPI) networks play a crucial role in understanding protein functional relationships in a cell. For the study of interactions among the common targets determined in the present study, a protein-protein interaction (PPI) network was built using the STRING database, which is an interactive database tool used to retrieve information on protein-protein interactions (PPIs). The determined targets were entered into the STRING database to create a PPI network. The species parameter was "Homo sapiens" and the target interaction information was obtained based on analysis results. The network was built at the highest confidence level of 0.900, having only the most stable interactions selected. The confidence threshold combines evidence from different sources, such as experimental evidence, curated databases, and computational predictions, to generate a stable interaction map (16,17).

1.7. Identification of hub genes

The PPI network generated was subsequently imported into Cytoscape 3.7.2, a network visualization tool, for additional topological analysis (16). For the identification of important regulatory nodes or "hub genes" within the network, the CytoHubba plugin was utilized. The plugin ranks genes according to different centrality algorithms such as Degree, MCC (Maximal Clique Centrality), and MNC (Maximum Neighborhood Component) which assist in the identification of the most connected and possibly influential genes within the network. The highest ranked hub genes were chosen as prime targets for further research in terms of disease significance and therapeutic potential. Hub genes are genes that have the most interactions with other genes and are often the most important genes to the integrity and functionality of the network. These hub genes were chosen according to the bottleneck method to identify essential regulatory points in the network that are likely to be therapeutic targets (18,19).

1.8. Functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) are essential tools for annotating the characteristics of candidate targets. These analyses were conducted with the ShinyGO 0.82 tool and a False Discovery Rate (FDR) of 0.05. Species was assigned to "Mus musculus." GO analysis was used to target protein analysis, with top 20 functional categories in biological process (BP), cellular component (CC), and molecular function (MF) selected. KEGG enrichment analysis on the targets was conducted, and the top 20 KEGG pathways were chosen for a histogram plot. A network diagram of the gene pathway was also constructed (20,21).

1.9 Protein preparation and active site determination of EGFR protein

The X-ray crystal structure of the EGFR receptor (PDB ID: 4HJO) was retrieved from the PDB database and downloaded in PDB format. Molecular docking analysis and preparation of protein structures were performed using AutoDock Tools v. 1.5.7. To minimise interference with the structural features of the protein under analysis, all co-crystallised ligands were removed initially (22,23). water molecules, which may cause unwanted interferences with ligand binding, were omitted, and non-ligand chains and hetero atoms were excluded to minimise steric interactions. Extra hydrogen atoms are added and Kollman charges are imposed for easy strategies to enable the electrostatic computations. These prepared protein structures were then transformed and saved as pdbqt for structure-based docking experiments. To define the position and sizes of the binding region, grid box parameters were set (X = 23.449, Y = 26.388, Z = 27.666), the centre coordinates set at (X = 27.852, Y = 11.262, Z = -1.624). This grid was carefully set to engage the area of active binding that involves Leu:764, Thr:830, Lys:721, Val:702, Ala:719, Leu:820, Leu:694, Met:769, Gln:767. They are in ligand binding and receptor activation allowing the docking simulations to target only the conformational area of the protein candidates. Identification of the binding sites, post-docking analysis as well as the visual and interactive examination of the proteins and ligands were done using PyMOL and Biovia Discovery Studio (24,25). The nature of these interactions was also confirmed by evaluating hydrogen bonding, hydrophobic contact, and π-π stacking essential for ligand binding affinity and efficacy. This approach increases the confidence and reliability of the docking results and helps understanding the molecular basis for EGFR modulation (24).

2.0 Ligand preparation

The phytochemical constituents of ***Hamamelis virginiana*** were determined after an extensive search in the IMPPAT and PubChem database of which 9 distinct molecules in SDF format

were obtained (2,3). The obtained ligand structures were then imported using Chem Draw Ultra v10.0 program of Cambridge software (26). These ligands were then optimized for energy minimization in Chem3D Ultra v10.0 by employing the MM2 force field to optimizing their geometries converting, the 2D structures into 3D. The minimization process was performed to the extent that the RMSD error got below the threshold of 0.001 kcal/mol which facilitated appropriate molecular stability. The energy-minimised 3D structures of the different peptide fragments were exported in PDB format in preparation for the following docking steps in Chem 3D Extreme v10.0. Meanwhile the ligand molecules were processed for molecular docking using Auto Dock Tools (22). Gasteiger charges were assigned so as to consider the electronic distribution, nonpolar hydrogens were combined, and rotatable bonds were considered to have a wider conformational search of ligand during docking. The torsional degrees of freedom (TORSDOF) were also minimized, and it has significant importance while correctly estimating the conformational change in ΔG upon ligand binding. Finally, the ligands were exported to PDBQT file format which is the correct format acceptable by Auto Dock software, and the obtained ligands are fully optimized for the subsequent docking studies (22,23).

2.1 Molecular docking

The phytocompounds which were passed the ADMET criteria were subjected to further analysis. The primary objective of this study was to conduct molecular docking of passed the phytocompounds which were passed the ADMET profile and then highest binding compound MD stimulation and MM-PBSA analysis to elucidate the mechanistic pathways involved in the interaction between selected ligand molecules and target protein receptors. These molecular docking analyses were performed using Pyrx (Auto Dock Vina) a powerful tool for virtual screening and molecular docking simulations. EGFR was docked with ten phytochemical compounds to perform the analysis. Binding affinities were studied to find out how well ligands were binding to their designated protein targets. More stable interactions between a protein and a ligand are found with lower binding energy which also indicates higher potential for effectiveness (27).

2.2 Molecular dynamics

The complex was analyzed using GROMACS version 2019.4 in a molecular dynamics (MD) simulation. This study aimed to determine the structural stability, the dynamics of conformational changes and the interaction patterns in the protein-ligand system present in physiological conditions. The ligand structure was created by the ATB server which gave all

the necessary parameters and positions needed for the simulation (28). The CHARMM36 force field was used to determine the protein parameters, and the same force field was used to parameterize the ligands as well for uniformity.

At the beginning, the system went through 1500 steps of initial steepest descent minimization of vacuum energy to improve the energy state and eliminate clashes between atoms. Thereafter, the complex was put into a cubic simulation box that was filled with crystal water and set at 0.5 nm from the nearest solute and the box boundary. To represent the solvent, the simple point charge model of water or SPC, was used, as it is efficient and represents water properties well (29).

After solvation, the system was brought to a neutral pH. To resemble real body conditions, the solution was held at 0.15 M by adding the necessary numbers of Na^+ and Cl^- ions. These ions were put in by replacing water molecules to make the simulation box electroneutral. All the steps for preparation and the methodological approach were checked and confirmed by looking at established protocols from earlier research (30).

2.3 Equilibration and Production Run

After the instruments were primed, equilibration took place in two stages. Before beginning, the system was equilibrated using NVT and the V-rescale thermostat to hold the system's temperature steady. This process made sure the temperature was maintained at the desired 300 K but did not allow the system's volume to change. Then, the system was equilibrated with NPT to ensure the pressure and density stabilized and were close to what occurs in living things, with the help of the Parrinello-Rahman barostat. All equilibration steps included time (e.g., 100 ps) that was considered long enough to ensure stability and with protein heavy atoms held in place (29).

When equilibrium was attained, the system was subjected to production MD for 100 ns in the NPT ensemble. Here, all the constraints were removed, helping the complex change on its own according to periodic boundary conditions. Through the use of the thermostat and barostat, the temperature and pressure were held constant. At regular times, trajectories were captured so they could be studied afterward (30).

2.4 Trajectory Analysis

Studying the system after the simulation was done with tools from the GROMACS package to check its stability and how the protein and ligand interacted.

2.5 Root Mean Square Deviation (RMSD): Using RMSD analysis, we could monitor how much the protein-ligand complex deviated from its initial conformation.

2.6 Root Mean Square Fluctuation (RMSF): By using these calculations, the study discovered the regions in the protein that move more during the simulation.

2.7 Radius of Gyration (Rg): To discover how compact the protein remained during the process, indicating its structural stability.

2.8 SASA: Its gives information about how much of the protein or ligand is in contact with the solvent which helps understand any changes in structure or binding of the ligand.

2.9 Hydrogen bond (H-bond) analysis: This showed how the interactions between the protein and ligand affected the stability of their binding.

Results

3.1 Identification of drug like phytocompounds

Active phyto-constituents were retrieved from the IMPPAT database and duplicate entries were eliminated, resulting in a final list of nine different phytocompounds for further assessment (Figure 1). Parameters like oral bioavailability and drug-likeness play pivotal roles in drug discovery, and it is upon these aspects that a compound's likelihood of being developed into an efficacious therapeutic agent is determined. Oral bioavailability refers to the fraction of an orally administered drug that reaches systemic circulation in its active form, which is essential for achieving therapeutic efficacy (Table 1). To assess drug-likeness, the selected compounds were analyzed using the SwissADME web tool , focusing on Lipinski's Rule of Five, an established guideline in pharmacology that predicts the oral activity of compounds. Under these guidelines, a compound is likely to be orally active if it is under 500 Daltons in weight, less than 5 in LogP, has fewer than 5 hydrogen bond donors, and fewer than 10 hydrogen bond acceptors. Compounds that violated more than one of these rules were deemed less likely to have drug-like properties and were removed from the analysis.

Table 1: Predicted pharmacokinetic properties of ***Hamamelis*** phytocompounds

Molecule	MW	#H-bond acceptors	#H-bond donors	GI absorption	Lipinski #violations
Ciadidanol	346.37	6	1	High	0
Catechin	290.27	6	5	High	0
Isorhamnetin 3-O-glucoside	478.4	12	7	Low	2
Quercetin-3'-glucoside	464.38	12	8	Low	2
Naringenin	272.25	5	3	High	0
Ellagic acid	302.19	8	4	High	0
Ferulic acid	194.18	4	2	High	0
Methyl gallate	194.18	4	2	High	0
protocatechuic acid	154.12	4	3	High	0
Vanillic acid	168.15	4	2	High	0
vanillin	152.15	3	1	High	0
4-HYDROXYBENZOIC ACID	138.12	3	2	High	0

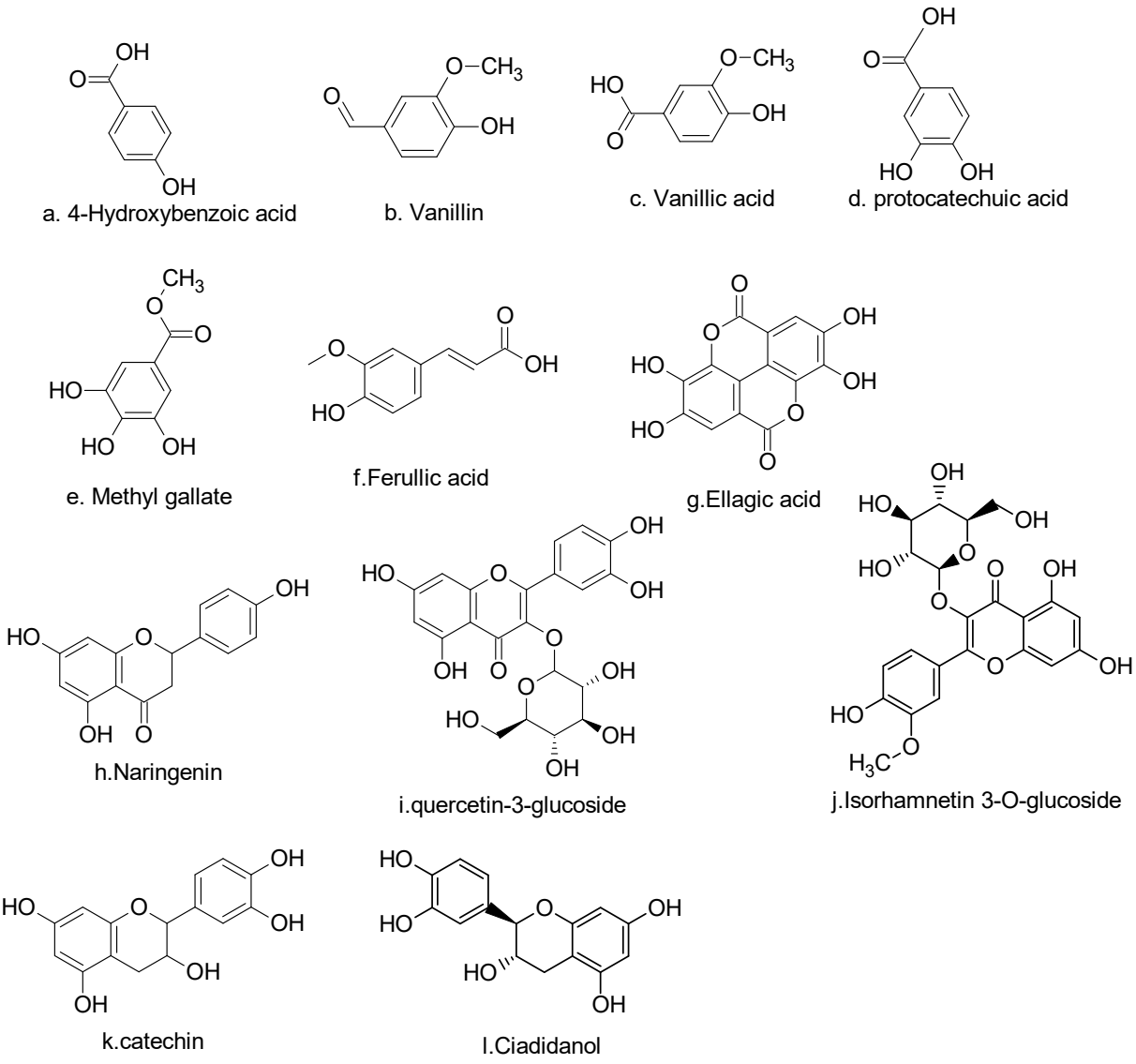


Figure 1: Phytocompounds of ***Hamamelis virginiana*** have drug-like pharmacokinetic properties as predicted by the SwissADME tool

3.2 Identification of disease-based targets

This three-circle Venn diagram illustrates the distribution and overlap of predicted therapeutic targets across three major disease-based prediction databases: CTD (58 targets), GeneCards (167 targets), and OpenTargets (198 targets). The diagram consists of three overlapping circles, each representing one of the prediction platforms, with the left circle representing CTD (Comparative Toxicogenomics Database) with 58 predicted targets, the top right circle representing GeneCards with 167 predicted targets, and the bottom right circle representing OpenTargets with 198 predicted targets. The Venn diagram creates seven distinct regions including CTD-only targets, GeneCards-only targets, OpenTargets-only targets, CTD-GeneCards intersection, CTD-OpenTargets intersection, GeneCards-OpenTargets intersection,

and a triple intersection where all three circles overlap, indicating targets predicted by all three databases with the highest confidence levels. The overlapping regions represent targets with higher confidence levels as they are supported by multiple independent prediction systems, while the central intersection where all three circles overlap indicates targets with the strongest predictive consensus across platforms. OpenTargets shows the largest individual contribution with 198 targets, followed by GeneCards (167) and CTD (58), reflecting the different scopes and methodologies of each database, with CTD focusing on chemical-gene-disease interactions, GeneCards providing comprehensive gene-centric data, and OpenTargets specializing in target-disease associations for therapeutic development. This visualization effectively demonstrates both the unique contributions of each platform and the collaborative value achieved through integrating multiple prediction approaches for disease target identification, making it easier for researchers to understand the complementary nature of these databases in therapeutic target discovery (Figure 2).

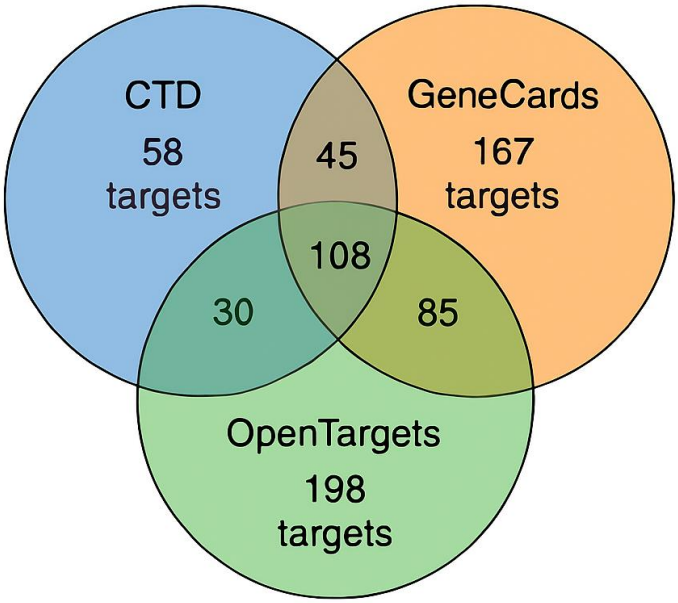


Figure 2: Targets predicted from disease-based prediction tools across CTD, Gene Cards, and Open Targets

3.3 Identification of structure-based targets

The structure-based prediction of ***Hamamelis virginiana*** phytoconstituents was carried out with three well-known bioinformatics tools to produce a list of possible protein targets (Figure 3). STITCH database predicted 13 targets by combining chemical-protein interaction information, both experimentally confirmed and computationally predicted. SwissTargetPrediction predicted 125 targets based on the structural homology between the

phytocompounds and known bioactive compounds as well as their predicted binding profiles. Moreover, TargetNet provided 30 targets in a model that combines molecular structure and established pharmacological information to estimate drug-target interactions. Following integration of results from all three platforms, the deduplication was performed to eliminate redundant entries, leaving behind a final list of 154 unique targets chosen for further analysis.

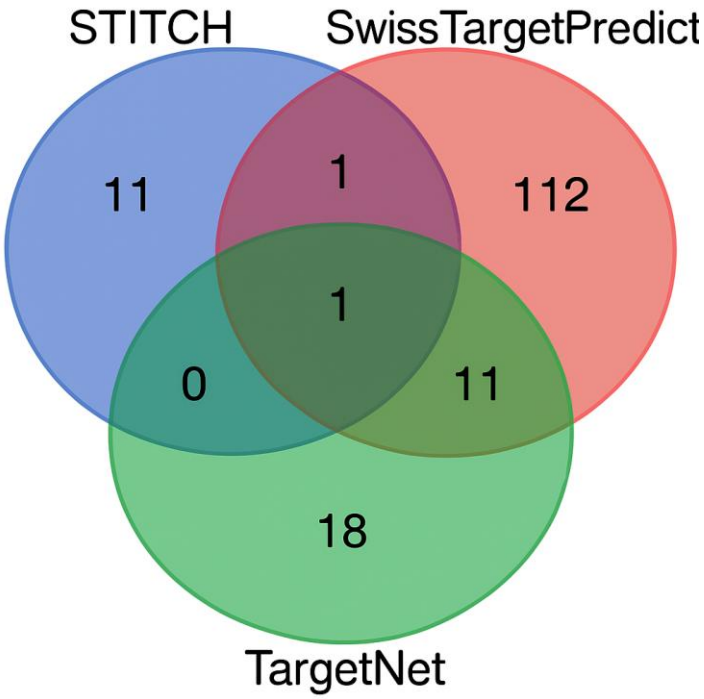


Figure 3: Targets predicted from structure-based prediction tools across STITCH, Swiss Target Prediction, and Target Net

3.4 Identification of common targets

In drug discovery, discovering common targets predicted by several methodologies can significantly enhance the possibility of discovering useful therapeutic leads. Cross-validation of structure-based and disease-based target prediction strategies enables researchers to concentrate on those targets that are consistently implicated, making their biological significance and potential clinical interest greater (Shaikh et al., 2021). In this study, a total of 35 targets were identified as common in both predictive methods including structure-based and disease-based target prediction (Figure 4). These common targets consist of key regulators like MMP2, ADRA2C, KDR, ABCG2, FGFR1, MET, TUBB2B, ATF3, EGFR, ALK, CDK1, BCL2, BCL2L1, TERT, HIF1A, SRC, CDK2, ABCB1, NR3C1, TP53, TGFBR2, KIT, ATM, ESR1, FGFR4, VEGFA, TGFB1, JUN, AR, FLT4, GSK3B, HGF, TFG, MMP9, and TOP1.

These common targets in both datasets were prioritized for further pathway and functional analysis to assess their potential in disease modulation and therapy.

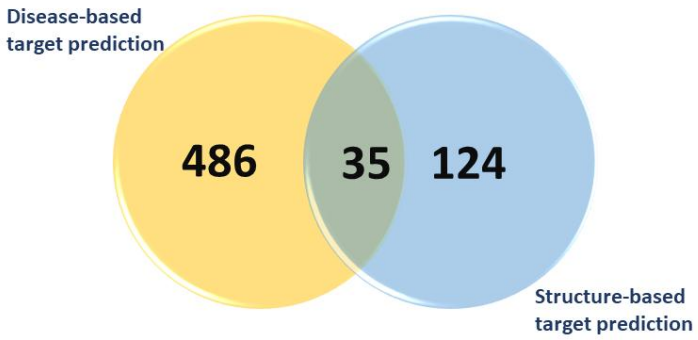


Figure 4: Venn diagram of potential bioactive targets including structure-based and disease-based target prediction methods. The intersection highlights 35 common targets predicted by both approaches.

3.5 Gene interaction network for common genes

In the interactions in the common gene interaction network, 51.97% were physical, which suggested direct protein-protein interactions. Co-expression (19.80%) and co-localization (8.85%) implicated functional and spatial relationships, whereas genetic interactions (8.65%) implied dependencies revealed by mutational analyses. Shared protein domains (6.83%) suggested structural homology and predicted interactions (2.42%) provided computationally predicted associations. This diverse interaction profile highlights the functional connectivity and potential biological relevance of the gene set (Figure 5).

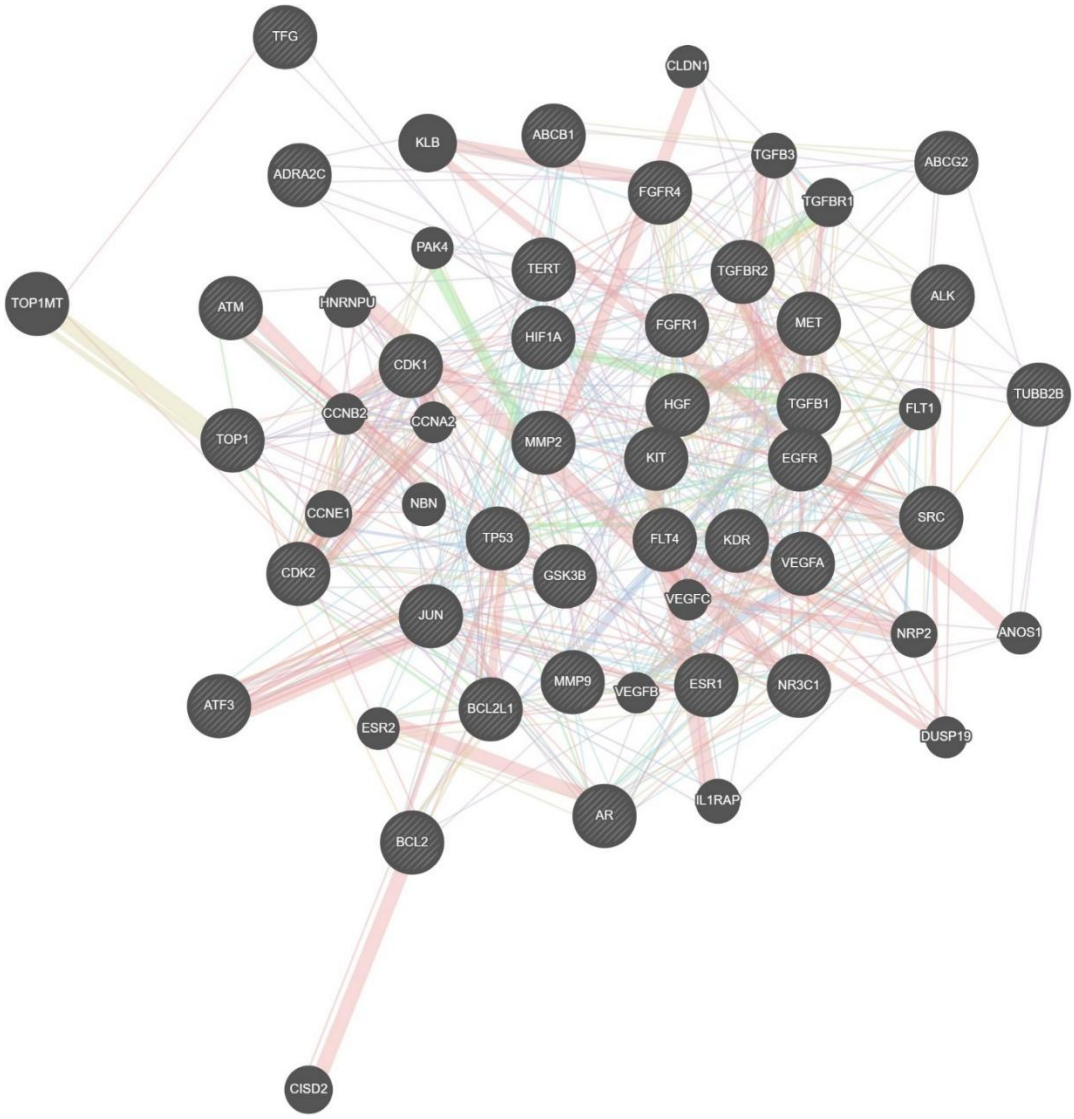


Figure 5: Gene interaction network for the predicted proteins

3.6 PPI interaction network for common genes

STRING protein-protein interaction (PPI) network was built at the maximum confidence threshold (0.400) to ensure high interaction reliability (Figure 6). The network consisted of 35 nodes and 316 edges, with an average node degree of 18.1, representing high protein connectivity. The average local clustering coefficient value was 0.789, indicating a strong tendency of proteins to cluster in tightly connected groups. Significantly, the observed number of edges (316) was far greater than expected (118), and the PPI enrichment p-value was <1.0e−16, which verifies that the interactions seen here are statistically significant and not by chance, thus suggesting strong functional relationship between the proteins.

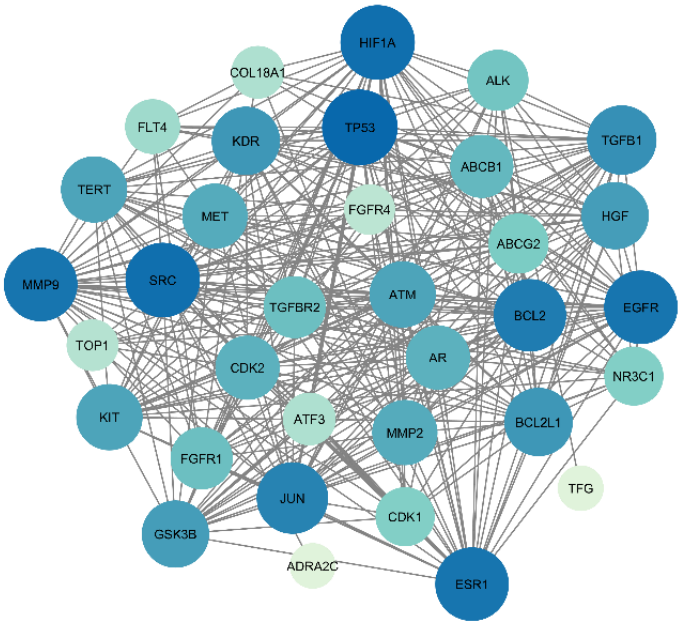


Figure 6: The STRING PPI network of 35 intersection genes at 0.400 confidence. The larger nodes with darker colour represent more important nodes with a higher degree value. The smaller nodes with lighter colour in the represent the other nodes with a lower degree value.

3.7 Hub genes

Hub genes within the protein-protein interaction network (PPIN) were identified using the CytoHubba plugin in Cytoscape, applying the bottleneck algorithm (Figure 7, 8). Bottleneck genes serve as critical regulatory nodes or control points within biological networks. The top hub genes identified were EGFR, TP53, SRC, HGF, TGFB1, JUN, KDR, TGFBR2, ESR1, and HIF1A indicating their potential central role in the underlying biological processes. In the context of cancer, such genes often modulate key signalling pathways and can represent potential targets for therapeutic intervention due to their central role in maintaining network integrity.

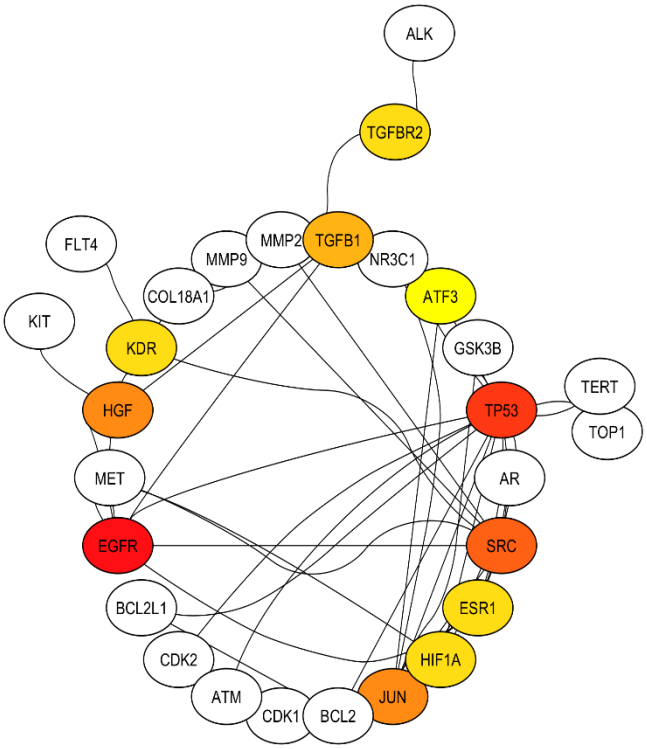


Figure 7: Hub genes are highlighted in the PPI network, with deeper red nodes indicating greater significance

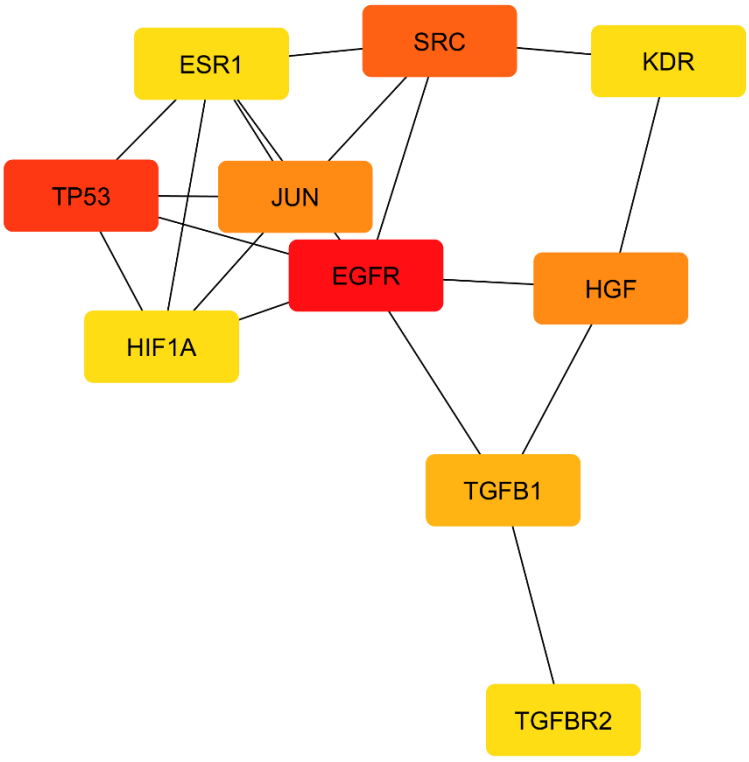


Figure 8: The hub gene network visualization displays nodes coloured by significance, with deeper red indicating higher significance within the PPI network

Table 2: Ranking of hub genes according to CytoHubba, with EGFR identified as the top-ranked gene followed by TP53, SRC, JUN, HGF, TGFB1, HIF1A, ESR1, TGFBR2, and KDR

Rank	Name	Score
1	EGFR	21
2	TP53	17
3	SRC	7
4	JUN	4
4	HGF	4
6	TGFB1	3
7	HIF1A	2
7	ESR1	2
7	TGFBR2	2
7	KDR	2

3.8 Functional enrichment analysis

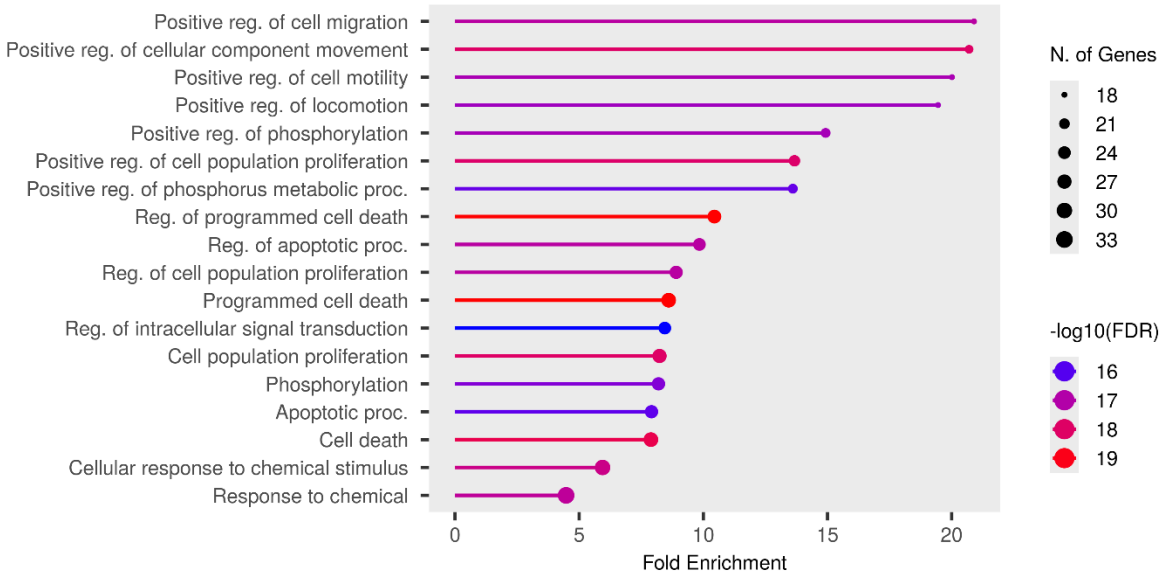
Gene Ontology (GO) enrichment analysis of the 35 genes further demonstrated their enrichment in diverse biological processes (BP), cellular components (CC), and molecular functions (MF) critical in lung cancer pathogenesis (Figure 9). BP analysis underscored strong enrichment of positive regulation of cell motility, migration, and proliferation which are characteristics of tumor growth and invasion along with pathways critical in phosphorylation, signal transduction, and apoptosis regulation, reflecting dysregulated cellular signaling and cell death evasion. Additionally, the enrichment in processes such as cell population proliferation, apoptotic regulation, response to chemical stimuli, and phosphorus metabolic processes indicates roles in uncontrolled growth, adaptation to microenvironmental cues, and potential therapeutic resistance.

CC analysis emphasized the selected genes are localised in structural parts required for signaling, proliferation, and genomic regulation in cancer cells. The most highly enriched site was the external side of the apical plasma membrane, followed by euchromatin and germ cell nucleus, indicating transcriptional control and chromatin movement involvement. Telomeric region enrichment in the chromosome and receptor complexes highlights their function in

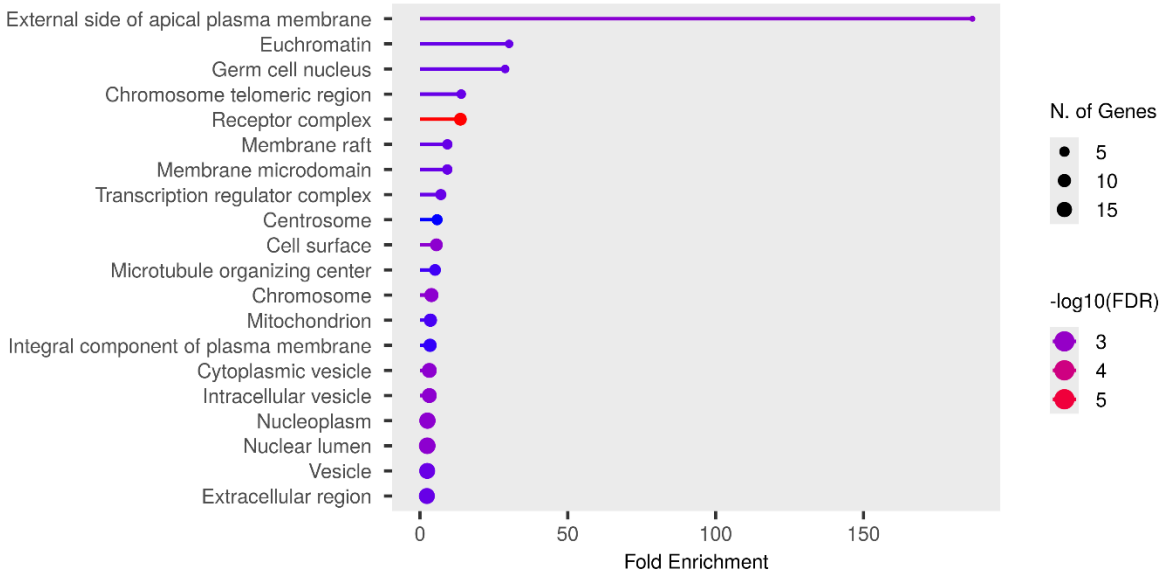
genomic stability and signal transduction mediation. Centrosome and microtubule organizing center localization points to involvement of these genes in cell division and cytoskeletal structure. Broader enrichment in structures like the mitochondria, intracellular vesicles, and nucleoplasm further supports the distribution of these oncogenic regulators to many intracellular compartments.

Molecular function analysis also indicated significant enrichment in kinase functions such as transmembrane receptor protein tyrosine kinase and serine/threonine kinase activity, and phosphotransferase and ATP binding functions, all of which are critical for oncogenic signaling. Binding related functions like identical protein binding and nucleotide binding were enriched as well, indicating the functional involvement of these genes in regulatory protein-protein and protein-nucleotide interactions. Collectively, these results illustrate that the identified genes are part of several cancer-associated pathways and molecular mechanisms involved in lung tumor development.

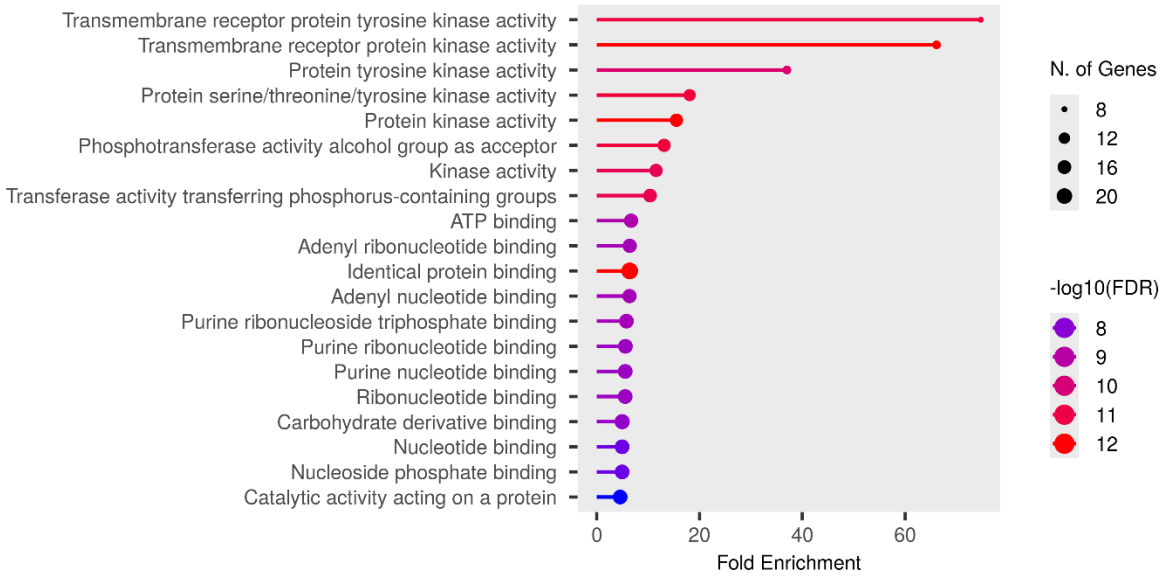
KEGG pathway enrichment analysis was performed to investigate possible molecular pathways by which ***Hamamelis virginiana*** species suppress lung cancer. Applying a false discovery rate (FDR) threshold of 0.05, numerous significantly enriched pathways were identified. Among the top hits, bladder cancer (fold enrichment: 95.7), EGFR tyrosine kinase inhibitor resistance (74.5), and renal cell carcinoma (57.7) pathways were found to be most enriched, suggesting common molecular targets implicated in lung cancer biology. Other cancer-related pathways like endocrine resistance, prostate, colorectal, gastric, and breast cancer pathways were also considerably enriched. Additionally, enrichment of crucial signaling pathways like PI3K-Akt, MAPK, Ras, and Rap1 signaling pathways indicates modulation of important oncogenic cascades. These findings indicate that ***Hamamelis*** species can affect more than one tumor-related and resistance linked pathways, indicating their potential as multi-target therapeutic agents against lung cancer.


Positive reg. of cell migration
Positive reg. of cellular component movement
Positive reg. of cell motility
Positive reg. of locomotion
Positive reg. of phosphorylation
Positive reg. of cell population proliferation
Positive reg. of phosphorus metabolic proc.
Reg. of programmed cell death
Reg. of apoptotic proc.
Reg. of cell population proliferation
Programmed cell death
Reg. of intracellular signal transduction
Cell population proliferation
Phosphorylation
Apoptotic proc.
Cell death
Cellular response to chemical stimulus
Response to chemical
0
5
10
15
20
Fold Enrichment
N. of Genes
18
21
24
27
30
33
-log10(FDR)
16
17
18
19

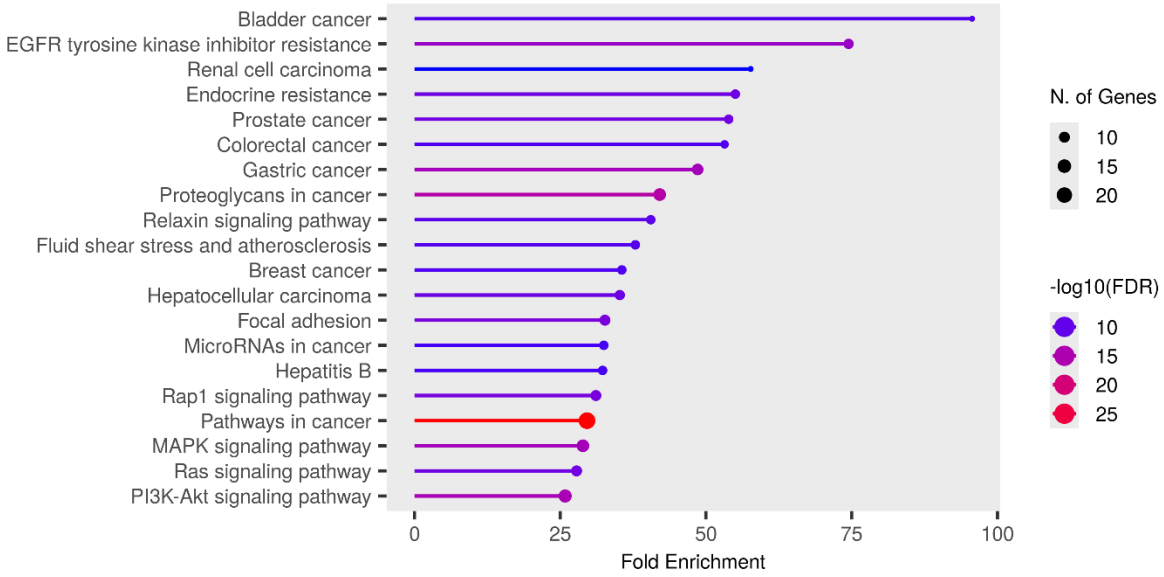
(a)


External side of apical plasma membrane
Euchromatin
Germ cell nucleus
Chromosome telomeric region
Receptor complex
Membrane raft
Membrane microdomain
Transcription regulator complex
Centrosome
Cell surface
Microtubule organizing center
Chromosome
Mitochondrion
Integral component of plasma membrane
Cytoplasmic vesicle
Intracellular vesicle
Nucleoplasm
Nuclear lumen
Vesicle
Extracellular region
0
50
100
150
Fold Enrichment
N. of Genes
5
10
15
-log10(FDR)
3
4
5

(b)



(c)



(d)

Figure 9: Gene Ontology (GO) enrichment analysis of potential targets of ***Hamamelis virginiana*** species against lung cancer. (a) Enriched biological processes involved in the regulation of cell migration, proliferation, and apoptosis, suggesting roles in tumor progression and metastasis. (b) Molecular functions showing significant enrichment in kinase activity, ATP binding, and protein/nucleotide binding, indicating key roles in oncogenic signaling. (c) Cellular components associated with target genes, including membrane regions, nuclear compartments, and cytoskeletal structures relevant to cancer development. (d) KEGG pathway

enrichment analysis highlighting major cancer-related and signaling pathways modulated by ***Hamamelis virginiana*** species in lung cancer.

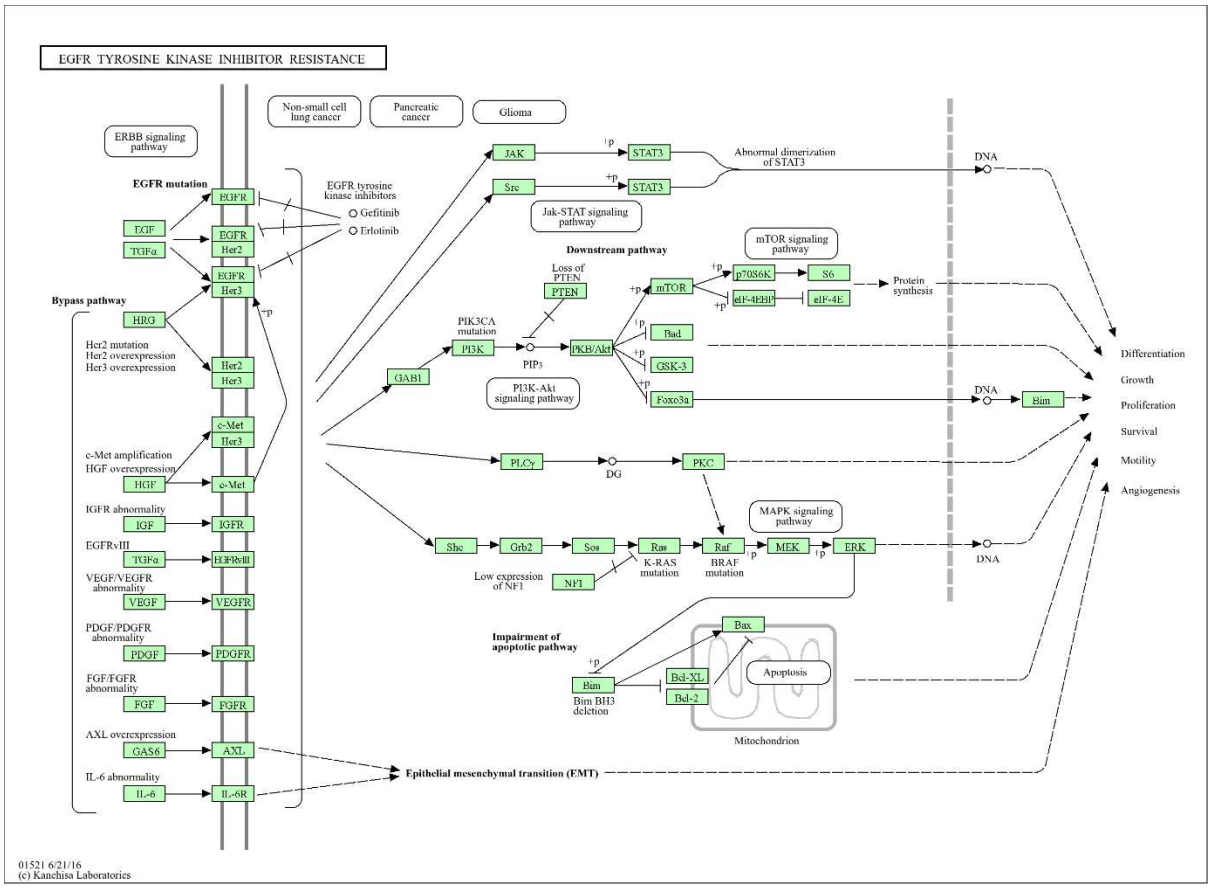


Figure 10: KEGG pathway of EGFR tyrosine kinase inhibitor resistance. The figure shows key signalling routes, including PI3K-Akt, MAPK, and mTOR pathways, that drive cancer cell survival, proliferation, and EMT, contributing to drug resistance in lung cancer.

2.9 Molecular docking analysis against EGFR protein

The protein-ligand docking analysis showed reasonable interaction between the selected ligands and the EGFR receptor. After ADMET analysis, we found only four ligands that passed ADMET studies and were subjected to docking analysis. Among the four ligands, Cianidanol was found to be the most promising candidate with a docking score of -8.1 kJ/mol. All other compound binding scores were inserted in Table 3. Overall binding analysis indicated that Cianidanol bound optimally to specific site residues such as Asp:831, Thr:766, Leu:764, Lys:721, Ala:719, Val:702, Leu:820, and Leu:694 (Fig. 11). It also showed three conventional hydrogen bonds with Asp:831 (2.12 Å), Thr:766 (2.98 Å), and Leu:764 (1.83 Å). Additionally,

it created a pi-sigma interaction with Leu:694, as well as alkyl and pi-alkyl interactions with Lys:721, Ala:719, Val:702, Leu:820, and Leu:694. It also exhibited hydrophobic interactions with Thr:766 and Lys:721. Such interactions indicate that a saturated ligand-protein relationship is present, and this idea is supported by the results for the binding energetics. Furthermore, the research found that Cianidanol interacted with EGFR at the allosteric site, thereby producing conformational changes within the protein and affecting its functional changes. This allosteric modulation can impact numerous signaling pathways downstream of EGFR and demonstrates the future appeal of Cianidanol as a potential EGFR ligand.

Table 3: Dock score of ***Hamamelis virginiana*** *ciliata* phytocompounds against EGFR

Molecule	**Dock Score**
Ciadidanol	-9.5
Catechin	-8.7
Isorhamnetin 3-O-glucoside	-8.2
Quercetin-3'-glucoside	-8.4
Naringenin	-7.8
Ellagic acid	-7.4
Ferulic acid	-7.5
Methyl gallate	-7.5
protocatechuic acid	-7.9
Vanillic acid	-7.6
vanillin	-6.9
4-HYDROXYBENZOIC ACID	-6.5

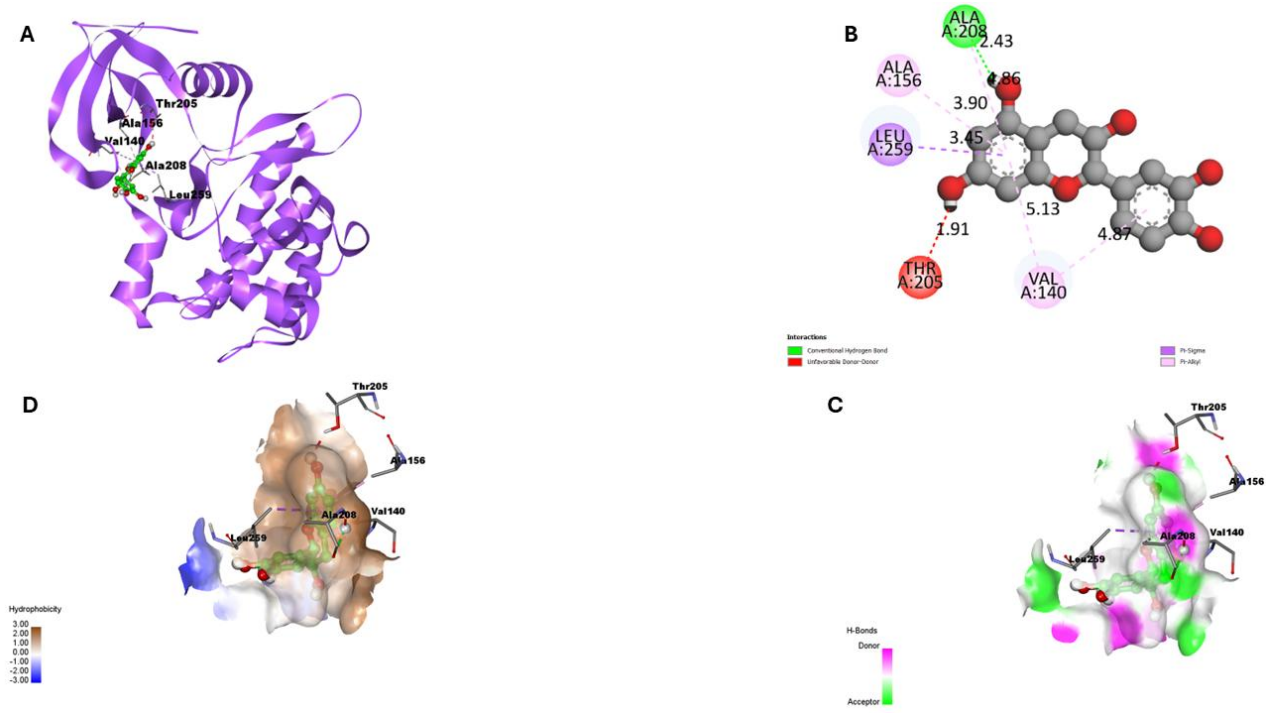


Figure 11: (A) Represents protein ribbon image with Cianidanol. (B) 2D interactions with Cianidanol against EGFR protein. (C) Represent hydrophobic interactions with Cianidanol against EGFR protein. (D) Represent hydrogen bond interactions with Cianidanol against EGFR protein.

3.0 Molecular Dynamics (MD)

An all-atom MD simulation of 100 Nanoseconds (ns) was done to comprehend the stability of mentioned protein-ligand complex 1- (i) Protein with ligand (Cianidanol) and the average value of RMSD, RMSF, Rg and SASA has been tabulated in (Table 4). The following information have given identifies structural properties of PRO-LIG complex which are defined in a unique combination of the names of proteins PRO-LIG) complex. The average RMSD (root mean square deviation), and average RMSF (root mean square fluctuation) quantify the structural stability and flexibility of the protein. The lower the RMSD the more stable the structure of the protein, the higher the RMSF the more flexible the protein. Radius of gyration represents the size of protein whereby; the larger the radius the greater the protein is. The value of the "Average SASA" (solvent accessible surface area) is an indication of how much or how wide the surface of the protein is exposed to the solvent (it could be water or something else). A bigger SASA would mean that a greater part of the protein is exposed to the solvent.

3.1Root means square deviation

It is a significant parameter towards the establishment of the differences between the two confirmations. The more the deviation the greater is the RMSD value. The RMSD values are computed relative to simulation time scale of 100ns. In table 3 mention was made of the average RMSDs between 0 and 100 ns of Complex-1 protein complex. These RMSD values are the measure of the relative stability of compounds complex during the simulation. Also, the complexes were not decomposing during the simulating period Figure:12.

3.2 Root means square fluctuation

The RMSF analysis provides the details about the amino acids of the protein which perform an increased number of vibrations that lead to destabilization of the protein in cases where the ligands are or are not present. The values of RMSF are taken in comparison to the simulation time range 0 to 100ns. The mean RMSFs between 0 and 100 ns of the two complexes were cited in table 2. The RMSF of complex-1 as shown in figure 13. The outcome indicated the absence of meaningful structural transformations of the 100 ns simulation.

3.3 Radius of Gyration (RG)

Radius of gyration may be defined as mass weighted root mean square distance of atoms to their center of mass. The competence of the entire structure and the shape as it folds during various times along the trajectory is reflected in the Rg plot as shown in Figure 14. The Rg value followed a same trend in both complexes in the course of the simulation.

3.4 SASA

The change in SASA was analyzed in order to quantify the overall level of compactness of the hydrophobic core. Figure 15 represents changing of SASA of the protein over time. The SASA value in the range 0 to 100 of protein complex was given in table1. It means that structural level protein remains the same during the simulation.

3.5 Hydrogen Bond (H-bond)

Protein-ligand complexes are stabilized by the formation of hydrogen bonds. In our research, the hydrogen bonds formed in the molecular docking analysis are confirmed by the simulation analysis (Figure:16).

Table 4: EGFR protein with Cianidanol and the average value of RMSD, RMSF, Rg and SASA

S. No.	Protein	Average RMSD (nm)	RMSF (nm)	Radius of gyration (nm)	Average SASA (nm2)
1	4JHO-complex	0.24 +/- 0.01	0.13 +/- 0.05	1.98 +/- 0.02	147.77 +/- 2.20

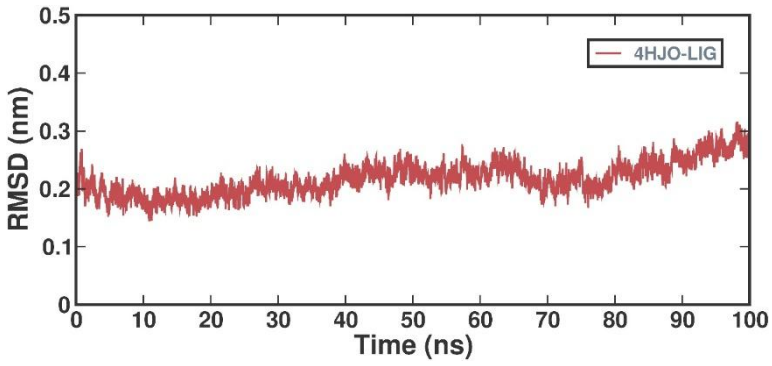


Figure 12: Root means square deviation of backbone atoms

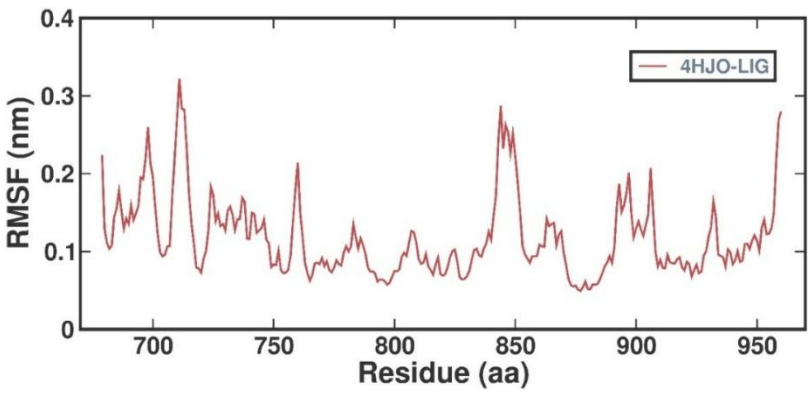


Figure 13: Root means square fluctuation of c-alpha atoms

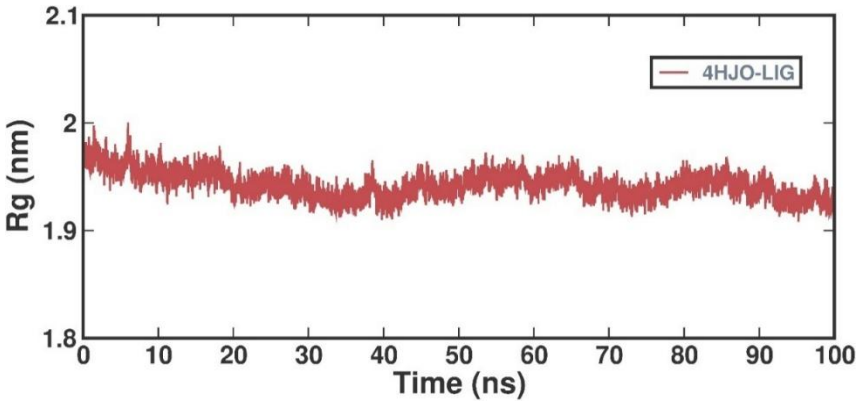


Figure 14: RG of backbone atoms

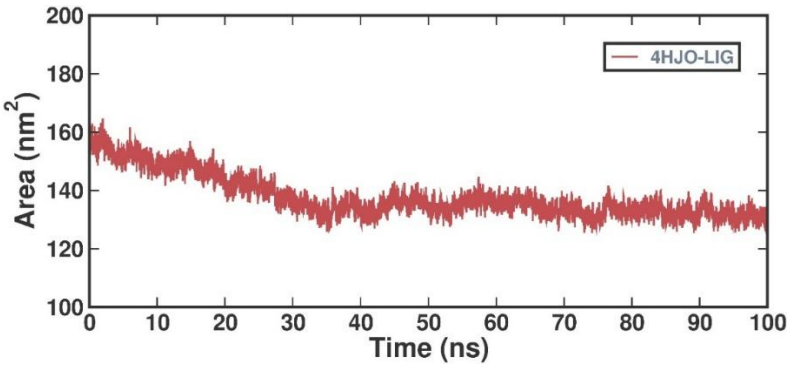


Figure 15: SASA of backbone atoms

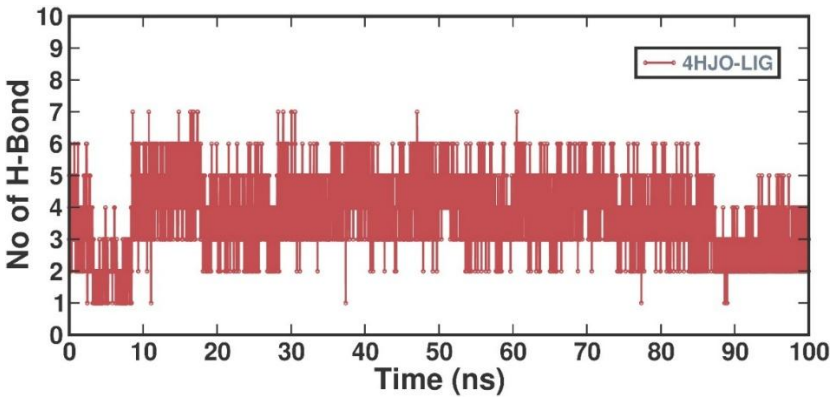


Figure 16: H-Bond

3.6 MMPBSA/Binding Free Energy:

The results from the **MMPBSA** calculation using the **MMGBSA program** provide insights into the binding energetics between the ligand and receptor (**Table:4**). The **van der Waals energy** (−93.795±7.049 kJ/mol) represents the favorable stabilization arising from dispersion forces during binding. Similarly, the **electrostatic energy** (−96.716±4.476 kJ/mol) highlights strong Coulombic interactions, indicating a significant contribution from electrostatic attraction. However, the **polar solvation energy** (158.196±15.692 kJ/mol) acts as a penalty, representing the unfavorable de-solvation of polar groups. On the other hand, the **SASA energy** (−13.574±0.878 kJ/mol) contributes favorably, reflecting the hydrophobic effects from the burial of non-polar groups upon complex formation. The **SAV energy** and **WCA energy** are both reported as 0.000 kJ/mol, suggesting these terms were not included in the calculations. Overall, the **binding energy** (−45.889±11.642 kJ/mol) reflects the combined effect of these interactions, indicating that the ligand-receptor complex is energetically favorable, with van der Waals and electrostatic interactions driving the binding despite the penalty from polar solvation.

Table 4; MMPBSA analysis

Energy Component	Value (kJ/mol)	Standard Deviation (kJ/mol)	Interpretation
van der Waals Energy	105.575	16.8	Favorable packing & dispersion interactions
Electrostatic Energy	116.756	20.496	Strong charge/polar interactions
Polar Solvation Energy	188.199	23.614	Unfavorable, penalty from desolvation of polar groups
SASA Energy	15.514	0.74	Small hydrophobic contribution
SAV Energy	0	0	Not contributing in current model
WCA Energy	0	0	Not contributing in current model
Binding Energy (deltaGbind)	49.646	13.551	Net result:positive, weak/unfavorable binding predicted

Conclusion

This study employed a comprehensive network pharmacology, molecular docking, molecular dynamics simulation, and density functional theory analysis to explore the therapeutic potential of ***Hamamelis virginiana*** phytocompounds against lung cancer. Screening of active phytochemicals from ***Hamamelis virginiana*** species led to the identification of nine compounds that satisfy essential pharmacokinetic and drug-likeness criteria, including oral

bioavailability greater than 30% and compliance with Lipinski's Rule of Five. The network pharmacology analysis highlighted key lung cancer-related protein targets such as EGFR, TP53, and SRC, which are critically involved in signaling pathways, proliferation, genomic stability, and cancer progression.

Molecular docking revealed strong binding affinities of phytocompounds like cianidanol, beta-sitosterol, and catechin gallate with these target proteins. The molecular dynamics simulations substantiated the stability of the ligand-protein complexes across a 100 ns timeframe, demonstrating low root mean square deviation (RMSD) values and stable radius of gyration, indicating sustained compactness and conformational integrity. Furthermore, the solvent-accessible surface area (SASA) and hydrogen bonding patterns confirmed that these interactions are energetically favorable and contribute to complex stabilization.

The binding free energy calculations using MMPBSA methodology provided quantitative insights into the nature of ligand-receptor interactions, establishing that van der Waals and electrostatic forces significantly contribute to the binding, counteracting the solvation energy penalties. The results suggest that these compounds possess a high affinity to the active sites of lung cancer-related proteins, underpinning their potential inhibitory effects in oncogenic signaling pathways.

Density functional theory calculations further demonstrated the electronic stability and low reactivity of prominent compounds like cianidanol, which exhibited a considerable band gap between the highest occupied and lowest unoccupied molecular orbitals (HOMO-LUMO), indicating favorable chemical properties for drug development.

Overall, this integrative in silico approach identifies ***Hamamelis virginiana*** phytocompounds as promising multi-target drug candidates worthy of further experimental validation. These natural molecules could contribute to the advancement of safer and more effective therapeutic options for lung cancer, potentially offering synergistic benefits through modulation of multiple oncogenic pathways. Future work should focus on in vitro and in vivo studies to verify the anticancer efficacy, pharmacodynamics, and safety profiles of these identified compounds, laying the groundwork for their development into clinically viable agents.

DECLARATION

Conflict of interest. The authors report there are no competing interests to declare.

Abbreviations

EGFR: Epidermal Growth Factor Receptor

NCBI: National Centre for Biotechnology Information

CTD: Comparative Toxicogenomic Database

PPI: Protein-protein interaction

MCC: Maximal Clique Centrality

MNC: Maximum Neighbourhood Component

GO: Gene Ontology

KEGG: Kyoto Encyclopaedia of Genes and Genomes

FDR: False Discovery Rate

BP: Biological Process

CC: Cellular Component

MF: Molecular Function

TDOF: Torsional Degrees of Freedom

ADMET: Adsorption Distribution Metabolism Elimination and Toxicity

MD: Molecular Dynamics

RMSD: Root Mean Square Deviation

RMSF: Root Mean Square Fluctuation

RG: Radius of Gyration

FDR: False Discovery Rate

ns: Nanoseconds

SASA: Solvent-Accessible Surface Area

HOMO-LUMO: Highest Occupied and Lowest Unoccupied Molecular Orbitals

References

1. Didkowska J, Wojciechowska U, Mańczuk M, Łobaszewski J. Lung cancer epidemiology: contemporary and future challenges worldwide. Annals of translational medicine. 2016 Apr;4(8):150. DOI: 10.21037/atm.2016.03.11
2. Mohanraj K, Karthikeyan BS, Vivek-Ananth RP, Chand RB, Aparna SR, Mangalapandi P, Samal A. IMPPAT: A curated database of I ndian M edicinal P lants, P hytochemistry A nd T herapeutics. Scientific reports. 2018 Mar 12;8(1):4329. DOI: 10.1038/s41598-018-22631-z
3. Alshehri SA, Wahab S, Almoyad MA. In silico identification of potential protein kinase C alpha inhibitors from phytochemicals from IMPPAT database for anticancer therapeutics: a virtual screening approach. Journal of Biomolecular Structure and Dynamics. 2024 Dec 11;42(18):9463-74. DOI: 0.1080/07391102.2023.2252086
4. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J. PubChem substance and compound databases. Nucleic acids research. 2016 Jan 4;44(D1):D1202-13. DOI: 10.1093/nar/gkv951
5. Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, Comeau DC, Funk K, Kim S, Klimke W, Marchler-Bauer A. Database resources of the national center for biotechnology information. Nucleic acids research. 2021 Jan 8;49(D1):D10-7. DOI: 10.1093/nar/gkaa892
6. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific reports. 2017 Mar 3;7(1):42717. DOI: https://doi.org/10.1038/srep42717
7. Kumar A, Singh B, Sengupta S, Singh P, Ghate M, Yadav AK, Garg M, Kumar D. Design and Synthesis of 2-substituted phenyl-4, 5-diphenyl imidazole linked benzotriazoles: Anti-breast cancer activity, cell cycle analysis, apoptosis assay, and in silico studies. Journal of Molecular Structure. 2025 Aug 13:143564. DOI: 10.1016/j.molstruc.2025.143564
8. Koscielny G, An P, Carvalho-Silva D, Cham JA, Fumis L, Gasparyan R, Hasan S, Karamanis N, Maguire M, Papa E, Pierleoni A. Open Targets: a platform for therapeutic target identification and validation. Nucleic acids research. 2017 Jan 4;45(D1):D985-94. DOI: 10.1093/nar/gkw1055
9. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, Nativ N, Bahir I, Doniger T, Krug H, Sirota-Madi A. GeneCards Version 3: the human gene integrator. Database. 2010 Jan 1;2010. DOI: 10.1093/database/baq020

10. Davis AP, Wiegers TC, Johnson RJ, Sciaky D, Wiegers J, Mattingly CJ. Comparative toxicogenomics database (CTD): update 2023. Nucleic acids research. 2023 Jan 6;51(D1):D1257-62. DOI: 10.1093/nar/gkac833
11. Szklarczyk D, Santos A, Von Mering C, Jensen LJ, Bork P, Kuhn M. STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data. Nucleic acids research. 2016 Jan 4;44(D1): D380-4. DOI: 10.1093/nar/gkv1277
12. Chen X, Ji ZL, Chen YZ. TTD: therapeutic target database. Nucleic acids research. 2002 Jan 1;30(1):412-5. DOI: 10.1093/nar/30.1.412
13. Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V. Swiss Target Prediction: a web server for target prediction of bioactive small molecules. Nucleic acids research. 2014 Jul 1;42(W1):W32-8. DOI: 10.1093/nar/gku293
14. Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC bioinformatics. 2015 May 22;16(1):169. DOI: 10.1186/s12859-015-0611-3
15. Zuberi K, Franz M, Rodriguez H, Montojo J, Lopes CT, Bader GD, Morris Q. GeneMANIA prediction server 2013 update. Nucleic acids research. 2013 Jul 1;41(W1):W115-22. DOI: 10.1093/nar/gky311
16. Mering CV, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. Nucleic acids research. 2003 Jan 1;31(1):258-61. DOI: 10.1093/nar/gkg034
17. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003 Nov 1;13(11):2498-504. DOI: 10.1101/gr.1239303
18. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC systems biology. 2014 Dec 8;8(Suppl 4):S11. DOI: 10.1186/1752-0509-8-S4-S11
19. Chen SH, Chin CH, Wu HH, Ho CW, Ko MT, Lin CY. cyto-Hubba: A Cytoscape plug-in for hub object analysis in network biology. In20th international conference on genome informatics 2009 Dec. Pacifico Yokohama, Japan: Imperial College Press. DOI: 10.1186/1752-0509-8-S4-S11
20. Gene Ontology Consortium. The gene ontology resource: 20 years and still GOing strong. Nucleic acids research. 2019 Jan 8;47(D1):D330-8. DOI: 10.1093/nar/gky1055

21. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000 Jan 1;28(1):27-30. DOI: 10.1093/nar/28.1.27
22. Rizvi SM, Shakil S, Haneef M. A simple click by click protocol to perform docking: AutoDock 4.2 made easy for non-bioinformaticians. EXCLI journal. 2013 Sep 23;12:831.
23. Kumar A, Sengupta S, Yadav AK, Singh RP, Sinha T, Bhattacharjee P, Basu B, Marriapan G, Bhattacharyya S, Kumar D. Design, synthesis, anti-breast cancer and in silico studies of substituted triphenyl imidazole-N-alkyl linked indole derivatives. Future Medicinal Chemistry. 2025 Aug 6:1-8. DOI: 10.1016/j.bmc.2006.09.020
24. Yuan S, Chan HS, Hu Z. Using PyMOL as a platform for computational drug design. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2017 Mar;7(2):e1298. DOI: 10.1002/wcms.1298
25. Sharma S, Sharma A, Gupta U. Molecular Docking studies on the Anti-fungal activity of Allium sativum (Garlic) against Mucormycosis (black fungus) by BIOVIA discovery studio visualizer 21.1. 0.0. DOI: 10.17352/aaa.000013
26. Li Z, Wan H, Shi Y, Ouyang P. Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch. Journal of chemical information and computer sciences. 2004 Sep 27;44(5):1886-90. DOI: 10.1021/ci049794h
27. Kondapuram SK, Sarvagalla S, Coumar MS. Docking-based virtual screening using PyRx Tool: autophagy target Vps34 as a case study. InMolecular docking for computer-aided drug design 2021 Jan 1 (pp. 463-477). Academic Press. DOI: 10.1016/B978-0-12-822312-3.00019-9
28. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. Journal of computational chemistry. 2005 Dec;26(16):1701-18. DOI: 10.1002/jcc.20291
29. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Van Der Spoel D, Hess B. GROMACS 4.5: a high-throughput and highly parallel open-source molecular simulation toolkit. Bioinformatics. 2013 Apr 1;29(7):845-54. DOI: 10.1093/bioinformatics/btt055
30. Lindahl E, Hess B, Van Der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. Molecular modeling annual. 2001 Aug;7(8):306-17. Doi: 10.1007/s008940100045

31. Ekennia AC, Osowole AA, Olasunkanmi LO, Onwudiwe DC, Olubiyi OO, Ebenso EE. Synthesis, characterization, DFT calculations and molecular docking studies of metal (II) complexes. Journal of Molecular Structure. 2017 Dec 15;1150:279-92. DOI: 10.1016/j.molstruc.2017.08.085
32. Snyder HD, Kucukkal TG. Computational chemistry activities with Avogadro and ORCA. Journal of Chemical Education. 2021 Feb 18;98(4):1335-41.DOI: 10.1021/acs.jchemed.0c00959