DRUG DISCOVERY

Integrative in silico analysis of *Pinus Roxburghii* phytochemicals for drug discovery

Nayankumar Prajapati*, **Nikunj Patel**

To Cite:
Prajapati N, Patel N. Integrative in silico analysis of *Pinus Roxburghii* phytochemicals for drug discovery. *Drug Discovery* 2023; 17: e31dd1946
doi: https://doi.org/10.54905/disssi.v17i40.e31dd1946

Author Affiliation:
Department of Microbiology, Faculty of Science & Humanities, Smt. S. S. Patel Nootan Science and Commerce College, Sankalchand Patel University, Visnagar-384 315, Gujarat, India

***Corresponding Author**
Department of Microbiology, Faculty of Science & Humanities, Smt. S. S. Patel Nootan Science and Commerce College, Sankalchand Patel University, Visnagar-384 315, Gujarat, India
Email: nayanprajapati200528@gmail.com

Peer-Review History
Received: 20 June 2023
Reviewed & Revised: 24/June/2023 to 21/August/2023
Accepted: 27 August 2023
Published: 31 August 2023

Peer-Review Model
External peer-review was done through double-blind method.

Drug Discovery
pISSN 2278–540X; eISSN 2278–5396

ABSTRACT

The aim of the present study is to investigate in silico analysis of *Pinus Roxburghii* plant's photo component for the disease of non-small-cell lung cancer (NSCLC). We observed that many people have problems like lung cancer, and they were treated with synthetic medicine, which is already made from chemical compounds, and so for this study, we are targeting the plant which is in INDIAN Tropical forests. That plant's bark contains various chemicals that have been used to prevent a disease like lung cancer. In this work, we use various In Silico tools for many testing we use PubChem Database to obtain the details of the chemical components of the plants like chemical structures, properties, and other relevant data for small molecules, etc., with the help of PubChem we download the ligand and protein of the disease. After that, we use I gem Dock for docking the ligand and protein interaction. Then, we use VEGA QSAR for mutagenicity, carcinogenicity, Toxicity, etc. Then, ADMET/ADME tools are used to predict compounds' absorption, distribution, metabolism, excretion, and toxicity. After that, we use the Lipinski rule of five. After performing all these methods, I found that the plant's Bark compound is highly able to interact with the Disease protein it is shown the inhibition is the same as Drugs that are available in the market.

Keywords: *Pinus Roxburghii*, In Silico, QSAR, ADMET, Docking.

1. INTRODUCTION

Cancer is a group of various Diseases characterized by the uncontrolled growth of cells and the spread of abnormality in the body. Those cells affect normal tissues and organs and can also spread to other parts of the body through the bloodstream or lymphatic system, a process known as metastasis (Hanahan and Weinberg, 2011). There are many different types of cancer, and they are generally classified according to the part of the body in which they start. Some of the most common types of cancer include Lung Cancer, Breast Cancer, Skin Cancer, Prostate Cancer, and Colorectal Cancer, Other less common types of cancer include pancreatic cancer, ovarian cancer, cervical cancer, bladder cancer, kidney cancer, and liver cancer, among others (Vogelstein et al., 2013; Soria et al., 2018). Non-small-cell lung cancer (NSCLC) is the type of Lung cancer that causes 85% of the share as compared to others.

According to the American Cancer Society, in 5 years, the survival rate for the NSCLC is about 25% it is likely that people without the disease to live for at least

DISCOVERY SCIENTIFIC SOCIETY

5 years after completing diagnosis. One of the mechanisms of resistance to treatment that has been identified in NSCLC is the "spared" mechanism. This mechanism refers to the ability of cancer cells to survive treatment by avoiding or resisting the effects of chemotherapy or targeted therapy drugs (Reck et al., 2016). Because of these chemical-based medicines and other drugs, the patient has reactions also. For the alternative approach, we go through the Natural Option for medicine after reading many research papers and reviews, I found many plants from the Indian Territorial places like Uttarakhand, Kashmir, Nepal etc (Futreal et al., 2004). From the reviews, I selected the plant *Pinus Roxburghii*, also known as Chir Pine or Long leaf Pine, is a species of pine. Which is located in the Himalayas and commonly found in India, Pakistan, and Nepal (Kaushik et al., 2012; Kumar et al., 2012).

It is a medicinal plant and as per reviews and further studies, it gives various many activities like anti-microbial, anti-bacterial, anti-carcinogenicity, etc. We choose the methods for our work in silico analysis where we select various tools like PubChem, IGEM Dock, VEGA QSAR, ADMET/ADME, etc. These tools are free to access for all. Each tool has different functions and the operative's methods will also be different. We check ligand vs. protein interaction, Toxicity, carcinogenicity, and Lipinski rule of five. Based on the computational tools, we combine all the work and go through Drug discovery, and we use this study to conclude that the taken sample or ligand of the plant, Standard drugs that are available in markets are compared with protein of Lung cancer (Reck et al., 2016).

2. MATERIAL AND METHODOLOGY

Literature search

I use databases like PubMed, Scopus, Web of Science, and Google Scholar to find scientific material. After the entire evaluation process has been done, select databases pertinent to the study topic and the numerous parameters that I have decided to include in my study.

Selection of phytocompounds

GC-MS is a Gas Chromatography-Mass Spectrometry a combined technique to separate and quantify the compounds from any sample for characterization purposes. This present study was carried out by using data from a GC-MS study that was performed by (Satyal et al., 2013; Bhardwaj et al., 2022; Thapa et al., 2018).

Ligand Library

An enormous amount of data about chemical nomenclature, chemical structures, identifiers, physical, chemical, and biological properties, patents, health, safety, toxicity data, and other descriptors can be found in the open chemical structure database known as PubChem®. The use of various programmatic access points to accomplish virtually automated screening of chemical compounds makes the PubChem® database valuable information in the drug development process. Additionally, this database enables users to obtain PubChem® data files in a variety of formats and upload them to local computing resources, allowing data integration between PubChem® and other resources like web browsing tools (Xie XQ, 2010).

The following information will be gathered from this database: The PubChem® ID, the molecular formula, the molecular weight, the CAS (Chemical Abstracts Service) no., the EC (European Community) no., and the canonical SMILE (Simplified Molecular-Input Line-Entry) structures. Using a translator program (https://cactus.nci.nih.gov/translate/), the .sdf file of each chosen phytocompound will be converted to a .pdb file, which will then be used as input while doing docking interaction analysis (Yu et al., 2020).

Lipinski Rules of Five, Toxicity, Carcinogenicity & Mutagenicity prediction

Lipinski's rule of five, also known as Lipinski's rule, is a set of guidelines used to determine the drug-likeness of a molecule. It was developed by Christopher Lipinski in 1997 and is based on the observation that most orally administered drugs have specific physicochemical properties that allow them to be absorbed and distributed throughout the body. According to Lipinski's rule, a molecule is likely to have good oral bioavailability and is a drug candidate if it meets the following criteria: Molecular weight (MW) ≤ 500, octanol-water partition coefficient (LogP) ≤ 5, hydrogen bond donors (HBD) ≤ 5, hydrogen bond acceptors (HBA) ≤ 10. These rules were derived from the analysis of more than 2000 drugs and are, in many cases, a good predictor of oral bioavailability.

It is important to note that Lipinski's rule is not a strict code, but rather a guideline. There are many examples of drugs that break one or more of these rules but are still effective. Therefore, Lipinski's rule should be used as a tool to help identify potential drug candidates rather than as a definitive decision tool (Lipinski, 2000). VEGA is a freely available web platform that includes a series of QSAR (quantitative structure-activity relationship) models that can be accessed to predict the toxicity of selected

phytocompounds (Benfenati et al., 2013). This tool is easily installed and can be used in any operating system supporting JAVA. Users can easily use this program as a series of different models after selecting an SMILE structure or adding a chemical structure as an input file (Kumar et al., 2019).

Six models (mutagenicity (Ames test) CONSENSUS model 1.0.3; carcinogenicity model (CAESAR) 2.1.9; developmental toxicity model (CAESAR) 2.1.7; Table 3) are selected to conduct this study, which accounts for different toxicities such as for example mutagenicity, carcinogenicity and as toxicity to select a potent non-toxic compound. These models are used to screen compounds for drug design/development in silico. Non-toxic, non-mutagenic and non-carcinogenic compounds were filtered from the Ligand Library and we also perform various parameters such as skin sensitization model (CEZARO) (version 2.1.6), skin sensitization model (IRFMN/JRC) (version 1.0.0), hepatotoxicity model (IRFMN) (version 1.0.0), Whole Body Elimination Half-Life (QSARINS) (version 1.0.0), Fish Acute (LC50) Toxicity classification (SarPy/IRFMN) (version 1.0.2), Fish Acute (LC50) Toxicity Model (KNN/Read-Across) (version 1.0.0), LogP Model (Meylan/Kowwin) (Version 1.1.4), LogP Model (MLogP) (Version 1.0.0), LogP Model (ALogP) (Version 1.0.0), Water soluble model (IRFMN)) (version 1.0.0), Skin Permeation (LogKp) Model (Potts and Guy) (Version 1.0.0), Skin Permeation (LogKp) Model (Ten Berge) (Version 1.0.0), (Computer Kernel-Version: 1.2. 8) (Benfenati et al., 2013).

Target selection for docking study

In the next experiment, a total of 2 non-small cell lung cancer (NSCLC) proteins were targeted to investigate the effectiveness of the phytoconstituents as its drug molecule. Different proteins are selected based on their virulence. File in PDB format of target proteins Receptor tyrosine-protein kinase erbB-4 (2L2T and 3BCE) downloaded from PDB (Protein Data Bank) database (Table 4). (https://www.rcsb.org/) [79,80, 81]

Selection of Standard Drugs

Docking interaction analysis is essential in drug discovery and predicting the binding affinity between a ligand molecule (.sdf file) and target proteins (.pdb file). This evaluation predicts the optimal orientations (ie positions) of ligand-protein binding affinity to predict the formation of a stable complex. Docking interactions were performed using the iGEMDOCK software in the same manner as for phytocompounds. Various standard drugs have been selected for NSCLC, which are very effective in the human body, such as osimertinib etc. These drugs were selected for their effective inhibitory functions against selected diseases. Molecular docking study

The iGEMDOCK program was used for a molecular docking study between selected phytocompounds, standard drug vs. target proteins in various proteins of NSCLC to identify potential therapeutic phytocompounds and predict ligand-protein interactions. For the iGEMDOCK study, target proteins were selected from the total protein data bank (PDB): For non-small cell lung cancer (NSCLC): 2L2T, 3BCE. iGEMDOCK software used .pdb files of target proteins and selected phytocompounds as input to predict docking interactions between ligands and proteins further. A molecular docking interaction study was performed between non-toxic phytocompounds (compounds found in toxicity study) and these target proteins (RCSB, 2018; Rose et al., 2017).

Evaluation of Pharmacokinetics Study

Docking interaction analysis is essential in drug discovery and predicting the binding affinity between a ligand molecule (.sdf file) and target proteins (.pdb file). This evaluation predicts the optimal orientations (ie positions) of ligand-protein binding affinity to predict the formation of a stable complex. Docking interactions were performed using the iGEMDOCK software in the same manner as for phytocompounds. Various standard drugs have been selected for NSCLC, which are very effective in human body such as osimertinib etc. These drugs were selected for their effective inhibitory functions against selected diseases.

Molecular docking study the iGEMDOCK program was used for a molecular docking study between selected phytocompounds, standard drug Vs target proteins in various proteins of NSCLC to identify potential therapeutic phytocompounds and predict ligand-protein interactions. For the iGEMDOCK study, target proteins were selected from the total protein data bank (PDB): For non-small cell lung cancer (NSCLC): 2L2T, 3BCE. iGEMDOCK software used .pdb files of target proteins and selected phytocompounds as input to predict docking interactions between ligands and proteins further. A molecular docking interaction study was performed between non-toxic phytocompounds.

3. RESULT

In this work, I selected the plant *Pinus Roxburghii* name from them. A total 17 chemical compounds were included in the study for in silico. (Table 1). The chemical compound of *Pinus Roxburghii* was downloaded from the PubChem Compound database of the Therapeutic Target Database (Table 2). First Stage QSAR Study of Lipinski's Rule of Five Entire library compounds were screened for Lipinski rule five by Swiss ADME software. Out of a total 31 chemical compounds, only 81% (26) compounds were found to meet Lipinski's Rule of Five. Second Stage QSAR Study for Mutagenicity, Carcinogenicity, and Toxicity Initial filtering of the entire compound for the Lipinski's Rule of Five then filter for Mutagenicity, Carcinogenicity and Toxicity prediction. In-silico Batch predictions for Mutagenicity by Mutagenicity (Ames test) CONSENSUS model – 1.0.3 method was carried out using VEGA QSAR software. Out of a total 26 compounds, only total 92% (24) compounds as non-Mutagenicity (Table 3).

Table 1 Classification of Plant

Sr. No.	Kingdom	Plantae
01	*Clade*	Tracheophytes
02	*Clad*	Gymnosperms
03	Diviion	Pinophyta
04	Class	Pinopsida
05	Order	Pinales
06	Family	Pinaceae
07	Genus	*Pinus*
08	Subgenus	*P. subg. Pinus*
09	Section	*P. sect. Pinus*
10	Subsection	*Pinus subsect. Pinaster*
11	Species	*P. roxburghii*

Table 2 Pub Chem study of Plants Phytocomponents

No.	Name of Compound	PubChem ID	Mol. Formula	Mol. Weight	SMILE Structure
1.	2-chloropropionyl chloride	111019	C3H4Cl2O	126.97	CC(C(=O)Cl)Cl
2.	Boric acid, trimethyl ester	8470	C3H9BO3	103.92	B(OC)(OC)OC
3.	1-chloro butane	8005	C4H9Cl	92.57	CCCCCl
4.	Benzoic acid, 4-ethoxy-, ethyl ester	90232	C11H14O3	194.23	CCOC1=CC=C(C=C1)C(=O)OCC
5.	Anthracene	8418	C14H10	178.23	C1=CC=C2C=C3C=CC=CC3=CC2=C1
6.	Phthalic acid, isobutyl octadecyl ester	6423451	C30H50O4	474.7	CCCCCCCCCCCCCCCCCCOC(=O)C1=CC=CC=C1C(=O)OCC(C)C
8	2,2-dibromocholestanone	22212696	C27H44Br2O	544.4	C[C@H](CCCC(C)C)[C@H]1CC[C@@H]2[C@@]1(CC[C@H]3[C@H]2CCC4[C@@]3(CC(C(=O)C4)(Br)Br)C)C
9	Terpinolene	11463	C10H16	136.23	CC1=CCC(=C(C)C)CC1
10	Linalool	6549	C10H18O	154.25	CC(=CCCC(C)(C=C)O)C

11	Isoborneol	6321405	C10H18O	154.25	C[C@@]12CC[C@@H](C1(C)C)C[C@H]2O
12	p-Mentha-1,5-dien-8-ol	519323	C10H16O	152.23	CC1=CCC(C=C1)C(C)(C)O
13	Terpinen-4-ol	11230	C10H18O	154.25	CC1=CCC(CC1)(C(C)C)O
14	m-Cymen-8-ol	255195	C10H14O	150.22	CC1=CC(=CC=C1)C(C)(C)O
15	p-Cymen-8-ol	14529	C10H14O	150.22	CC1=CC=C(C=C1)C(C)(C)O
16	Estragole (=Methyl chavicol)	66957732	C20H24O2	296.4	CC1=C(C=CC(=C1)CC=C)O.COC1=CC=C(C=C1)CC=C
17	Citronellol	8842	C10H20O	156.26	CC(CCC=C(C)C)CCO
18	Neral	643779	C10H16O	152.23	CC(=CCC/C(=C\C=O)/C)C
19	Geraniol	637566	C10H18O	154.25	CC(=CCC/C(=C/CO)/C)C
20	Geranial	638011	C10H16O	152.23	CC(=CCC/C(=C/C=O)/C)C
21	Isobornyl acetate	6950273	C12H20O2	196.29	CC(=O)O[C@H]1C[C@@H]2CC[C@]1(C2(C)C)C
22	Linalool propanoate	6431132	C14H24O2	224.34	CCCC(=O)OC(C)(CCCC(=C)C)C=C
23	Citronellyl acetate	9017	C12H22O2	198.3	CC(CCC=C(C)C)CCOC(=O)C
24	Eugenol	3314	C10H12O2	164.2	COC1=C(C=CC(=C1)CC=C)O
25	Neryl Acetate	1549025	C12H20O2	196.29	CC(=CCC/C(=C\COC(=O)C)/C)C
26	Geranyl acetate	1549026	C12H20O2	196.29	CC(=CCC/C(=C/COC(=O)C)/C)C
27	Longifolene (=Junipene)	1796220	C15H24	204.35	C[C@]12CCCC([C@@H]3[C@H]1CC[C@@H]3C2=C)(C)C
28	Methyl eugenol	7127	C11H14O2	178.23	COC1=C(C=C(C=C1)CC=C)OC
29	(E)-Caryophyllene	5281515	C15H24	204.35	C/C/1=C\CCC(=C)[C@H]2CC([C@@H]2CC1)(C)C
30	Precocene I (=6-Demethoxyageratochromene)	28619	C12H14O2	190.24	CC1(C=CC2=C(O1)C=C(C=C2)OC)C
31	(E)-Ethyl cinnamate	637758	C11H12O2	176.21	CCOC(=O)/C=C/C1=CC=CC=C1
32	n-Dodecanol	8193	C12H26O	186.33	CCCCCCCCCCCCO
33	(E)-Nerolidol	5284507	C15H26O	222.37	CC(=CCC/C(=C/CCC(C)(C=C)O)/C)C

Table 3 Toxicity Prediction by VEGA QSAR

No.	Name of compounds	SMILE	Carcinogenity model (CAESAR) 2.1.9	Developmental Toxicity model (CAESAR) 2.1.7	Non - Toxicant
1	1-chloro butane	CCCCCl	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
2	Benzoic acid, 4-ethoxy-, ethyl ester	CCOC1=CC=C(C=C1)C(=O)OCC	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
3	Terpinolene	CC1=CCC(=C(C)C)CC1	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
4	Linalool	CC(=CCCC(C)(C=C)O)C	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
5	Citronellol	CC(CCC=C(C)C)CCO	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
6	Neral	CC(=CCC/C(=C\C=O)/C)C	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
7	Geraniol	CC(=CCC/C(=C/CO)/C)C	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
8	Geranial	CC(=CCC/C(=C/C=O)/C)C	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
9	Isobornyl acetate	CC(=O)O[C@H]1C[C@@H]2CC[C@]1(C2(C)C)C	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
10	Linalool propanoate	CCCC(=O)OC(C)(CCCC(=C)C)C=C	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
11	Eugenol	COC1=C(C=CC(=C1)CC=C)O	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
12	Neryl Acetate	CC(=CCC/C(=C\COC(=O)C)/C)C	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
13	Geranyl acetate	CC(=CCC/C(=C/COC(=O)C)/C)C	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
14	Longifolene (=Junipene)	C[C@]12CCCC([C@@H]3[C@H]1CC[C@@H]3C2=C)(C)C	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
15	Methyl eugenol	COC1=C(C=C(C=C1)CC=C)OC	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
16	(E)-Ethyl cinnamate	CCOC(=O)/C=C/C1=CC=CC=C1	Non - Mutagenic	Non - Carcinogen	Non - Toxicant
17	(E)-Nerolidol	CC(=CCC/C(=C/CCC(C)(C=C)O)/C)C	Non - Mutagenic	Non - Carcinogen	Non - Toxicant

In-silico Batch predictions for Carcinogenicity-by-Carcinogenicity oral classification model (IRFMN) – 1.0.0 method was carried out using VEGA QSAR software. Out of total 24 compounds only a total 91% (22) compounds as non-Carcinogenicity (Table 3). In-silico Batch predictions for Developmental Toxicity by the Developmental Toxicity model (CAESAR) – 2.1.7 method was carried out using VEGA QSAR software. Out of total 22 compounds only total 77% (17) compounds as Developmental Non-Toxicant (Table 3). In compare total 53% (31) out of 17 compounds were selected as Non-Mutagenicity, Non-Carcinogenicity & Developmental NON-Toxicant. Those 53% (17) filtered compounds were selected for toxicity prediction.

VEGA QSAR calculations were obtained for three different models/analyses and subsequently used to predict whether a compound was either toxic or nontoxic. Among compounds selected plant therapeutic compound: PubChem ID Eugenol (3314),

Linalool (6549), Methyl eugenol (7127), 1-chloro butane (8005), Citronellol (8842), terpinolene (11463), Benzoic acid, 4-ethoxy-, ethyl ester (90232), Geraniol(637566), (E)-Ethyl cinnamate (637758), Geranial (638011), Neral (643779), Neryl Acetate (1549025), Geranyl acetate (1549026), Longifolene (=Junipene), (1796220), (E)-Nerolidol (5284507), Linalool propanoate (6431132), Isobornyl acetate (6950273) are non- Mutagenic, non-Carcinogen, non-Toxicant.

Selection of Target

There were six successful and two research targets were selected from the literature survey and TTD (Therapeutic Target Database). 3D structure of the protein was downloaded from the PDB (Table 4). The resulting receptor was saved into a *.pdb file format for further Docking study.

Table 4 Protein Data Bank

No.	PROTEIN ID	Description	Type of TARGATE
1	2L2T	Receptor tyrosine-protein kinase erbB-4	Successful Target
2	3BCE	Receptor tyrosine-protein kinase erbB-4	Successful Target
3	6LUD	Epidermal growth factor receptor	Successful Target

Molecular Docking studies

The 37% (6) compounds having a drug like properties were selected as ligands to carry out for molecular docking studies in iGMDOCK software against the receptors. iGEM dock data Linalool propanoate (-96.7688), Geranyl acetate (-94.5596), (E)-Nerolidol (-93.3832) possessed lowest binding energy with 2L2T and (E)-Nerolidol (-98.3249), Neryl Acetate (-88.9739), Geranyl acetate (-88.1579) possessed lowest binding energy with 3BCE. This lowest binding energy gives a more stable complex between drug and protein. Out of 17 compounds Geranyl acetate and (E)-Nerolidol had the most stable binding with both 3BCE and 2L2T proteins. And after performing docking with all the Ligands and drugs with Main Protein that use when the Drug will be Created then we found the results shown in (Figure 1, 2, 3).

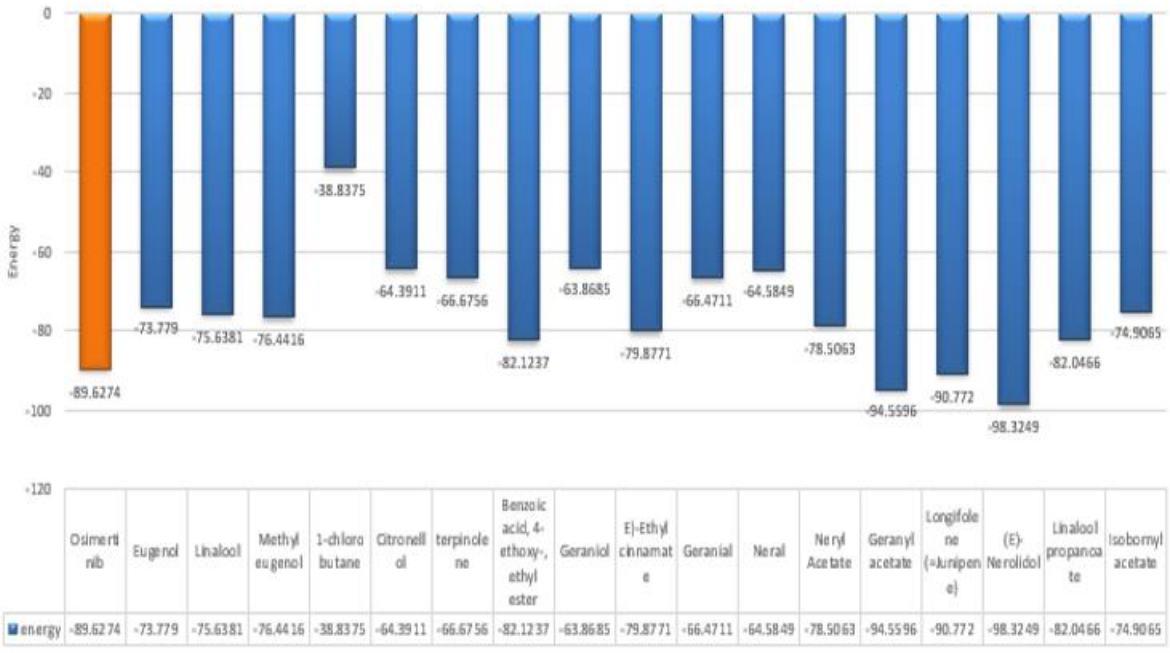

Figure 1 Binding energy values of the ligands with Protein 2L2T

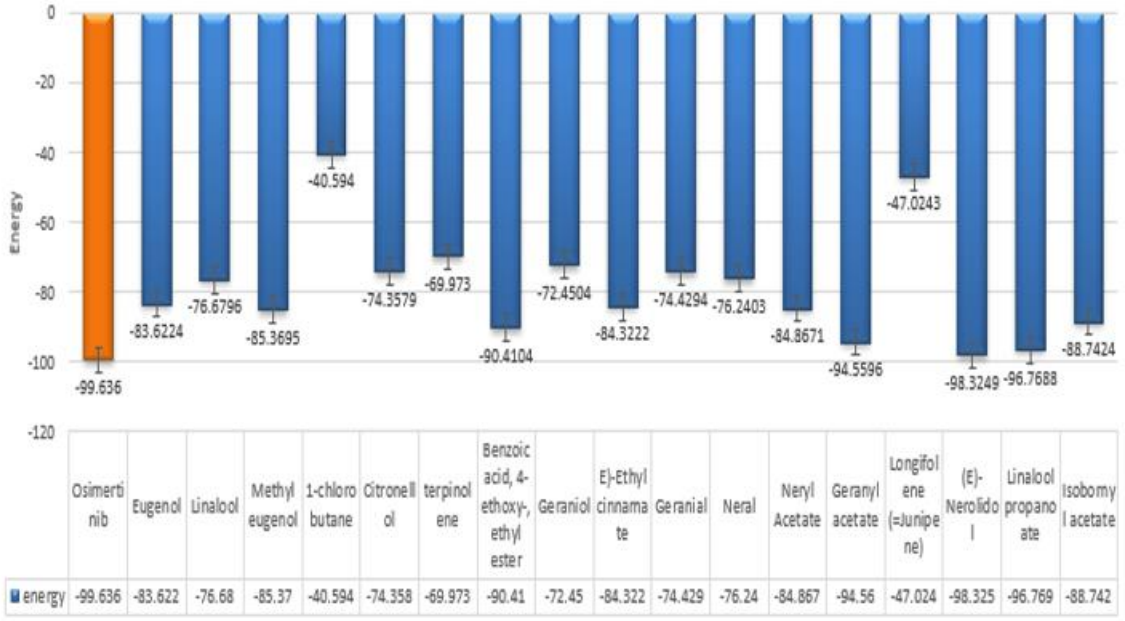

Figure 2 Binding energy values of the ligands with Protein 3u2p

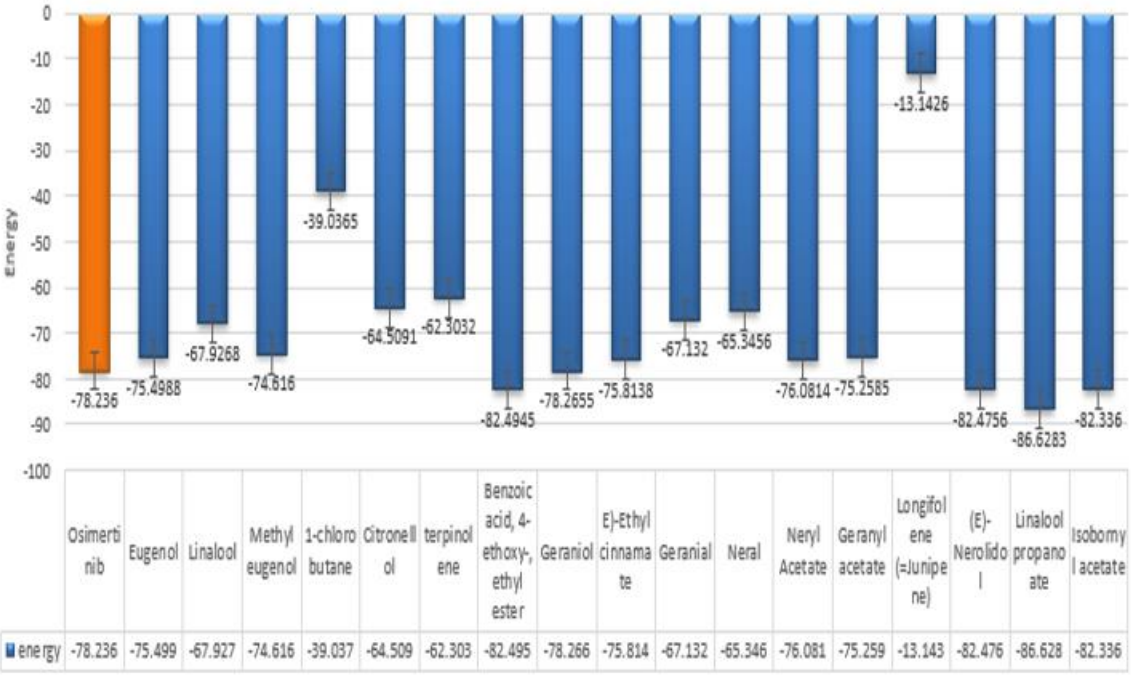

Figure 3 Binding energy values of the ligands with Protein 6LUD

Hydrogen Bond Interaction

The best score ligand was further analyzed for H-bond interaction. Ligand PubChem ID: 5284507 (E)-Nerolidol was found to have zero hydrogen bond with Receptor tyrosine-protein kinase erbB-4 (Figure 4). The best score ligand was further analyzed for H-bond interaction. Ligand PubChem ID: 5284507 (E)-Nerolidol was found to have zero hydrogen bond Receptor tyrosine-protein kinase erbB-4 (Figure 5). The best score ligand was further analyzed for H-bond interaction. Ligand PubChem ID: 71496458 Osimertinib was found to -8.5 hydrogen bond with the Epidermal growth factor receptor (6LUD) (Figure 6).

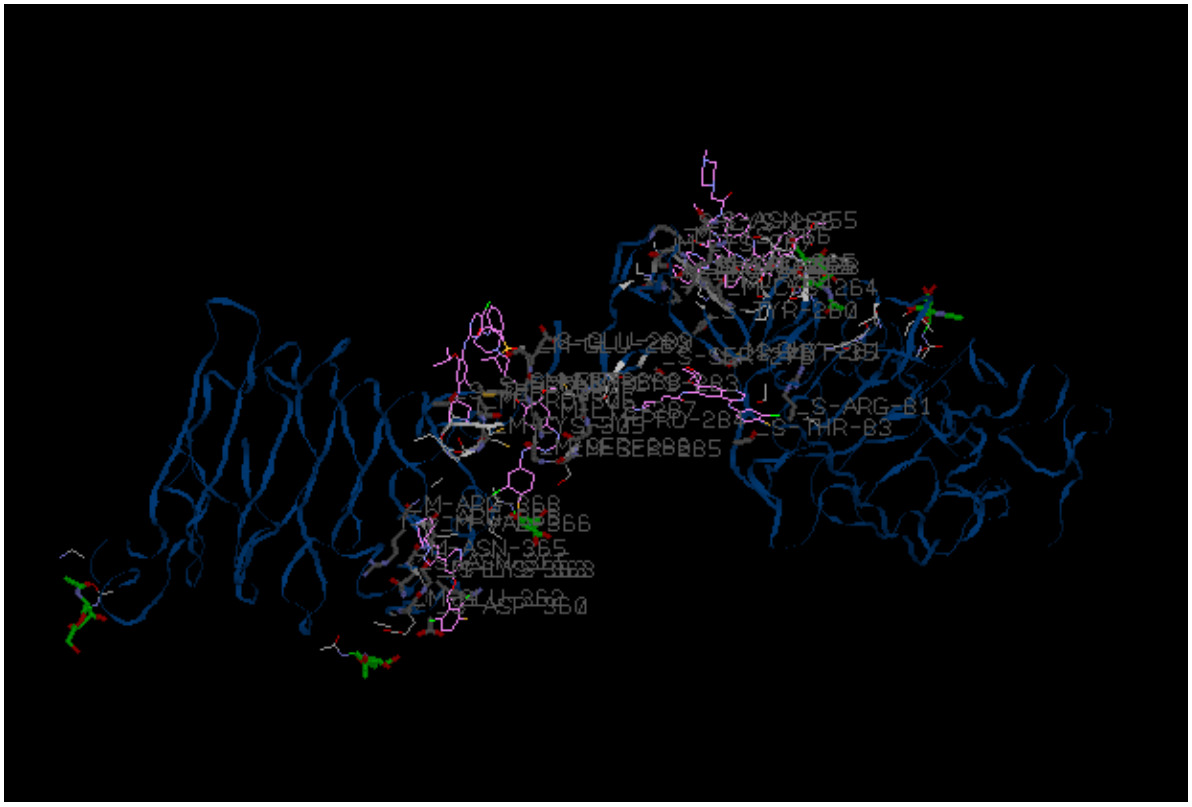

Figure 4 Hydrogen Bond Interaction with Receptor tyrosine-protein kinase erbB-4 (2L2T)

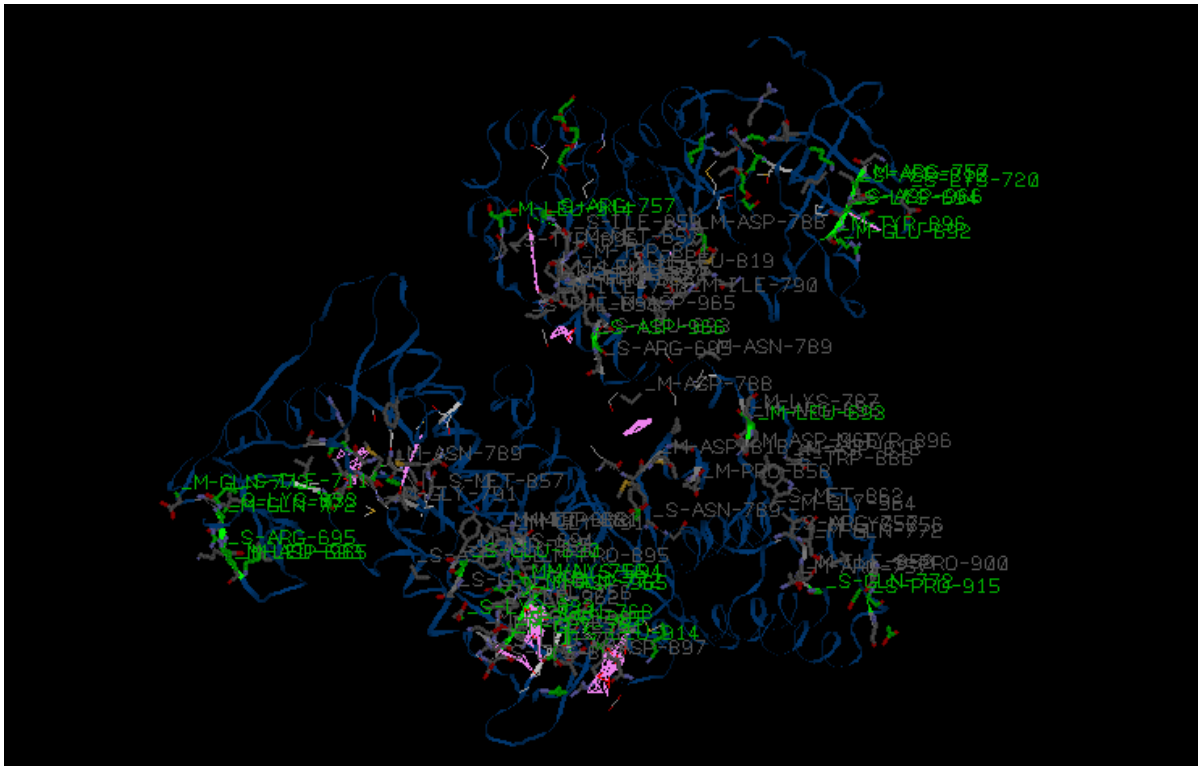

Figure 5 Hydrogen Bond Interaction with Receptor tyrosine-protein kinase erbB-4 (3BCE)

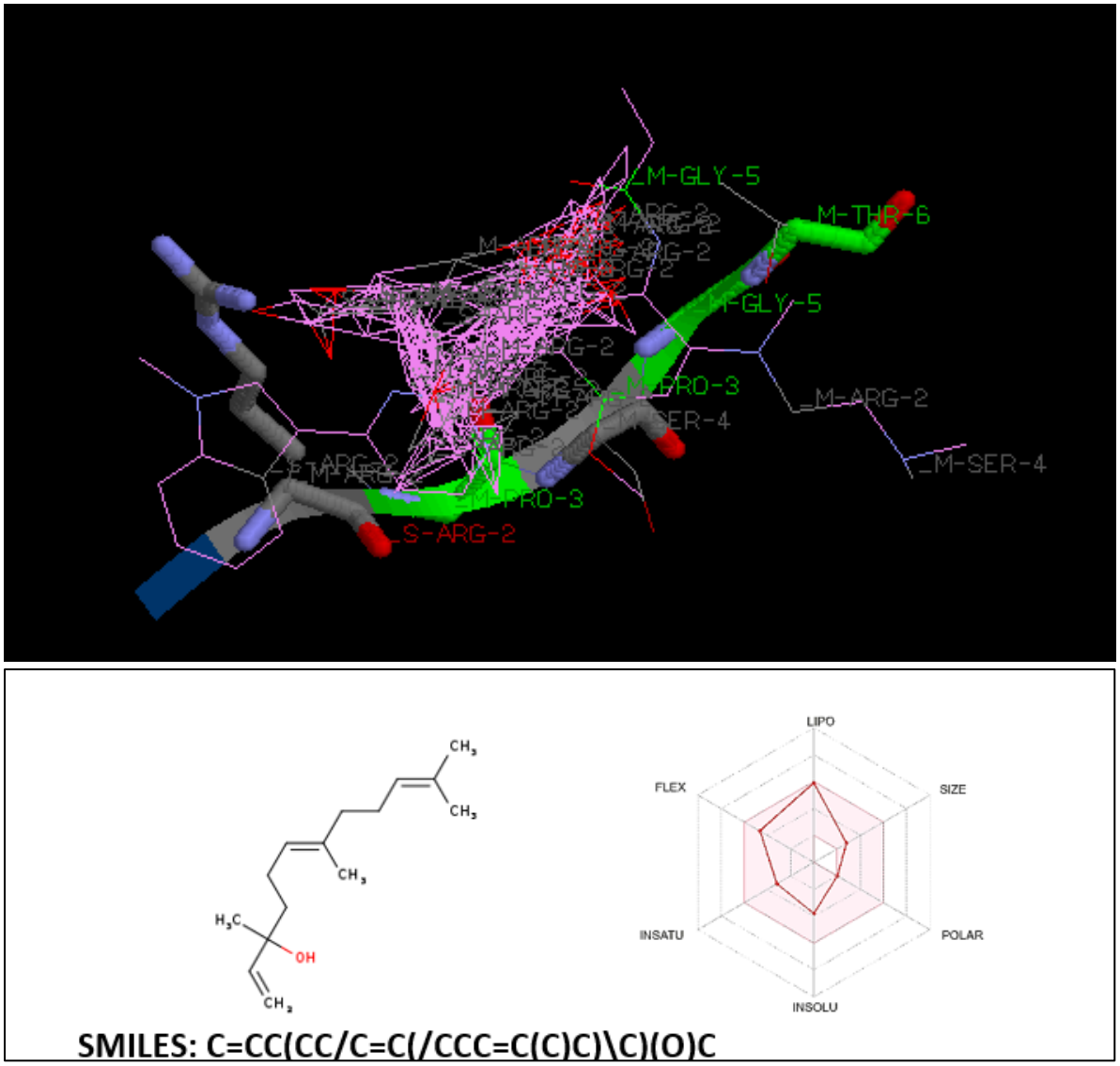

Figure 6 Hydrogen Bond Interaction with Epidermal growth factor receptor (6LUD)

Evaluation of Pharmacokinetics by Swiss ADME

The final set consisted of PubChem ID: 5284507 (E)-Nerolidol chemical compounds were selected for drug-like compounds. Predicts the value of physicochemical properties like Formula- C15H26O, Molecular weight-222.37 g/mol, Num. heavy atoms-16, Num. arom. heavy atoms-0, Fraction Csp3-0.6, Num. rotatable bonds-7 Num. H-bond acceptors-1, Num. H-bond donors-1, Molar Refractivity-74, TPSA-20.23 Å² (Table 5). Predicts the value of lipophilicity like Log Po/w (iLOGP), Log Po/w (XLOGP3), Log Po/w (WLOGP), Log Po/w (MLOGP), Log Po/w (SILICOS-IT), Consensus Log Po/w are followed as 3.64, 4.83, 4.4, 3.86, 4.21,4.19 (Table 6).

Table 5 Physicochemical Properties

1	Formula	C15H26O
2	Molecular weight	222.37 g/mol
3	Num. heavy atoms	16
4	Num. arom. heavy atoms	0
5	Fraction Csp3	0.6
6	Num. rotatable bonds	7
7	Num. H-bond acceptors	1
8	Num. H-bond donors	1
9	Molar Refractivity	74

10	TPSA	20.23 Å²

Table 6 Lipophilicity

1	Log Po/w (iLOGP)	3.64
2	Log Po/w (XLOGP3)	4.83
3	Log Po/w (WLOGP)	4.4
4	Log Po/w (MLOGP)	3.86
5	Log Po/w (SILICOS-IT)	4.21
6	Consensus Log Po/w	4.19

Table 7 Water Solubility

1	Log S (ESOL)	-3.8
2	Solubility	3.53e-02 mg/ml ; 1.59e-04 mol/l
3	Class	Soluble
4	Log S (Ali)	-4.99
5	Solubility	2.29e-03 mg/ml ; 1.03e-05 mol/l
6	Class	Moderately soluble
7	Log S (SILICOS-IT)	-3.15
8	Solubility	1.56e-01 mg/ml ; 7.00e-04 mol/l
9	Class	Soluble

Predicts the Water Solubility like Log S (ESOL), Solubility, Class, as -3.8 1. 3.53e-02 mg/ml; 1.59e-04 mol/l, Soluble. Log S (Ali), Solubility, Class, as -4.99, 2.29e-03 mg/ml; 1.03e-05 mol/l, moderately soluble. Log S (SILICOS-IT), Solubility, Class, as -3.15, 1.56e-01 mg/ml; 7.00e-04 mol/l, Soluble (Table 7). Predicts the medicinal chemistry like PAINS-0 alert, Brenk-1 alert: isolated alkene, Lead likeness- No; 2 violations: MW<250, XLOGP3>3.5, Synthetic accessibility 3.53 (Table 8). Predicts the Pharmacokinetics like GI absorption-High, BBB permeant-Yes, P-GP substrate-No, CYP1A2 inhibitor-Yes, CYP2C19 inhibitor-No, CYP2C9 inhibitor-Yes, CYP2D6 inhibitor-No, CYP3A4 inhibitor-No, Log Kp (skin permeation)- (-4.23 cm/s) (Table 9). Predicts the drug-likeness like Lipinski- Yes; 0 violation: MLOGP>4.15, Ghose- Yes, Veber- Yes, Egan- Yes, Muegge- No; 1 violation: Heteroatoms<2, Bioavailability Score 0.55 (Table 10). In total, highly predictive qualitative classification models were implemented.

Table 8 Medicinal Chemistry

1	PAINS	0 alert
2	Brenk	1 alert: isolated alkene
3	Lead likeness	No; 2 violations: MW<250, XLOGP3>3.5
4	Synthetic accessibility	3.53

Table 9 Pharmacokinetics

1	GI absorption	High
2	BBB permeant	Yes
3	P-gp substrate	No
4	CYP1A2 inhibitor	Yes
5	CYP2C19 inhibitor	No
6	CYP2C9 inhibitor	Yes
7	CYP2D6 inhibitor	No
8	CYP3A4 inhibitor	No
9	Log Kp (skin permeation)	-4.23 cm/s

These models include human intestinal absorption, blood-brain barrier penetration, Caco-2 permeability, P-glycoprotein inhibitor, CYP450 substrate and inhibitor (CYP1A2, 2C9, 2D6, 2C19, and 3A4), Human Ether-a-go-go-Related Gene inhibition inhibitor, AMES Mutagenicity, Carcinogenicity (binary), honeybee toxicity, and Tetrahymena Pyriformis toxicity (Table 11). The predictive values as human intestinal absorption (+), blood-brain barrier penetration (+), Caco-2 permeability (+), P-glycoprotein inhibitor (-), CYP1A2 inhibition (+), CYP2C19 inhibition (-), CYP2C9 inhibition (-), CYP2C9 substrate (+), CYP2D6 inhibition (-), CYP2D6 substrate (-), CYP3A4 inhibition (-), CYP3A4 substrate (-), Human Ether-a-go-go-Related Gene inhibition inhibitor (+), AMES Mutagenicity (-), Carcinogenicity (binary) (-), honeybee toxicity (-), Tetrahymena Pyriformis toxicity (0.014838654).

Table 10 Drug likeness

1	Lipinski	Yes; 0 violation
2	Ghose	Yes
3	Veber	Yes
4	Egan	Yes
5	Muegge	No; 1 violation: Heteroatoms<2
6	Bioavailability Score	0.55

Advances in computational tools and techniques played an important role in the drug design and discovery process. To reduce the demerits of drug discovery such as cost, time, and manpower etc, virtual screening procedures are routinely used. It utilizes docking and scoring of each phytocompounds from a dataset and predicts the binding interaction between ligands and target proteins. Molecular docking techniques have helped important proceedings in drug discovery for a prolonged time. It is helpful to study posing interaction as well as pose mode in the binding pocket of a target protein and to predict binding properties between them. All in all, these procedures will be led to further pharmacological evaluation.

Table 11 Prediction of admet SAR Properties

	Compound	5284507
1	Ames mutagenesis	-
2	Acute Oral Toxicity (c)	III
3	Androgen receptor binding	-
4	Aromatase binding	-
5	Avian toxicity	-
6	Blood Brain Barrier	+
7	BRCP inhibitior	-
8	Biodegradation	+
9	BSEP inhibitior	-
10	Caco-2	+
11	Carcinogenicity (binary)	-
12	Carcinogenicity (trinary)	Non-required
13	Crustacea aquatic toxicity	+
14	CYP1A2 inhibition	-
15	CYP2C19 inhibition	-
16	CYP2C9 inhibition	-
17	CYP2C9 substrate	-
18	CYP2D6 inhibition	-
19	CYP2D6 substrate	-
20	CYP3A4 inhibition	-
21	CYP3A4 substrate	-
22	CYP inhibitory promiscuity	-

23	Eye corrosion	-
24	Eye irritation	+
25	Estrogen receptor binding	-
26	Fish aquatic toxicity	+
27	Glucocorticoid receptor binding	-
28	Honey bee toxicity	-
29	Hepatotoxicity	-
30	Human Ether-a-go-go-Related Gene inhibition	-
31	Human Intestinal Absorption	+
32	Human oral bioavailability	-
33	MATE1 inhibitior	-
34	Mitochondrial toxicity	-
35	Micronuclear	-
36	Nephrotoxicity	+
37	Acute Oral Toxicity	1.418096423
38	OATP1B1 inhibitior	+
39	OATP1B3 inhibitior	+
40	OATP2B1 inhibitior	-
41	OCT1 inhibitior	-
42	OCT2 inhibitior	-
43	P-glycoprotein inhibitior	-
44	P-glycoprotein substrate	-
45	PPAR gamma	+
46	Plasma protein binding	0.641673327
47	Reproductive toxicity	-
48	Respiratory toxicity	-
49	Skin sensitisation	+
50	Subcellular localzation	Lysosomes
51	Tetrahymena pyriformis	0.014838654
52	Thyroid receptor binding	-
53	UGT catelyzed	+
54	Water solubility	-3.145595036

4. DISCUSSION

NSCLC or non-small cell lung cancer is a type of lung cancer that accounts for around 85% of all lung cancer cases. It is a complex disease that can be caused by a variety of factors, including smoking, exposure to air pollution, genetics, and certain occupational exposures. One of the most well-known risk factors for NSCLC is smoking. According to a study published in the Journal of Thoracic Oncology, smoking is responsible for up to 85% of lung cancer cases, and smokers are 15-30 times more likely to develop lung cancer than non-smokers (Sundbom et al., 2018). Other risk factors include exposure to radon, asbestos, and other chemicals found in the workplace, as well as a family history of lung cancer.

In terms of treatment, NSCLC can be treated in a variety of ways, including surgery, radiation therapy, chemotherapy, targeted therapy, and immunotherapy. The choice of treatment depends on the stage of the cancer, the overall health of the patient, and other factors. One recent study published in the Journal of Clinical Oncology compared the effectiveness of two different treatments for NSCLC: Chemotherapy and immunotherapy. The study found that immunotherapy was more effective in patients with advanced NSCLC who had high levels of a specific protein, called PD-L1, in their tumors (Herbst et al., 2016). Another study published in the New England Journal of Medicine compared the effectiveness of two different targeted therapies for NSCLC:

Osimertinib and gefitinib. The study found that Osimertinib was more effective than gefitinib in patients with NSCLC who had a specific genetic mutation, called EGFR (Soria et al., 2018).

NSCLC is a complex disease with many different causes and treatment options. Smoking is one of the most well-known risk factors for NSCLC, and a variety of treatment options exist, including surgery, radiation therapy, chemotherapy, targeted therapy, and immunotherapy. Recent studies have shown promising results for immunotherapy and targeted therapy in the treatment of NSCLC. Causes: The most significant risk factor for NSCLC is smoking tobacco. Other risk factors include exposure to radon gas, asbestos, air pollution, and genetic factors. Studies have shown that passive smoking, or exposure to second-hand smoke, can also increase the risk of NSCLC (Kalemkerian et al., 2018). The process of spading NSCLC typically involves the removal of the tumor along with a margin of healthy lung tissue. The extent of the spading depends on the size and location of the tumor, as well as the stage of the cancer.

In some cases, a lobe of the lung may need to be removed (lobectomy), while in others, a smaller section of the lung may be removed (wedge resection or segmentectomy). Several studies have investigated the effectiveness of spading in NSCLC. One study published in the Journal of Thoracic Oncology found that spading was associated with improved survival in patients with early-stage NSCLC. The study followed over 5,000 patients who underwent spading for stage I or II NSCLC and found that the 5-year survival rate was 73%. Another study published in the Annals of Thoracic Surgery compared different surgical approaches for spading NSCLC, including lobectomy, segmentectomy, and wedge resection. The study found that lobectomy was associated with the lowest risk of cancer recurrence and the highest overall survival rate, while wedge resection was associated with the highest risk of cancer recurrence.

While spading is a common treatment for NSCLC, it is not always appropriate for all patients. Factors such as the patient's age, overall health, and stage of cancer need to be taken into consideration when deciding on the best treatment approach. In addition to surgical resection, other treatment options for NSCLC include radiation therapy, chemotherapy, targeted therapy, and immunotherapy. The choice of treatment depends on several factors, including the stage and location of the cancer, as well as the patient's overall health. Treatments: There are several treatment options available for NSCLC, including surgery, radiation therapy, chemotherapy, targeted therapy, and immunotherapy. The choice of treatment depends on the stage and type of NSCLC, as well as the patient's overall health. (Reck et al., 2016) Surgery is the preferred treatment for early-stage NSCLC. It involves removing the tumor and surrounding tissue.

Radiation therapy and chemotherapy may also be used in combination with surgery to increase the chances of success. For advanced-stage NSCLC, targeted therapy, and immunotherapy are often used. Targeted therapy drugs are designed to target specific genes or proteins in cancer cells, while immunotherapy drugs stimulate the body's immune system to fight cancer cells. These treatments are usually less toxic than chemotherapy and may have fewer side effects. Herbst et al., (2016) Osimertinib is a small-molecule drug that is used to treat non-small cell lung cancer (NSCLC) with a specific mutation in the epidermal growth factor receptor (EGFR). It works by inhibiting the activity of the mutated EGFR protein, which slows down the growth and division of cancer cells. Osimertinib was first approved by the FDA in 2015 under the brand name Tagrisso.

Studies have shown that Osimertinib is effective in treating NSCLC with the T790M mutation, which is resistant to other EGFR inhibitors. In addition, Osimertinib has been shown to have fewer side effects compared to other EGFR inhibitors. While Oimertinib has been a significant advancement in the treatment of NSCLC, researchers are still working on developing new drugs to improve the effectiveness of treatment for this disease. One approach is to combine Oimertinib with other drugs that target different pathways involved in cancer growth and progression. For example, a phase II clinical trial is currently investigating the combination of Oimertinib with the drug bevacizumab, which targets the vascular endothelial growth factor (VEGF) pathway. Another approach is to develop new drugs that target other mutations in the EGFR pathway. For example, the drug lazertinib is currently being tested in clinical trials for its ability to treat NSCLC with the L858R mutation in the EGFR gene.

5. CONCLUSION

Our integrative in silico analysis of *Pinus Roxburghii* phytochemicals for drug discovery in non-small-cell lung cancer (NSCLC) has yielded promising results. By using various computational tools, including PubChem, Vega Qsar, Lipinski, iGEM dock, ADME, and ADMET, we were able to identify several potentially effective compounds for the treatment of NSCLC. Our analysis has revealed that several of the phytochemicals found in *Pinus Roxburghii* possess potent pharmacological activities. These findings suggest that *Pinus Roxburghii* phytochemicals may have significant potential as lead compounds for the development of novel drugs. Moreover, the integrative in silico analysis has allowed us to screen the compounds for various pharmacokinetic and pharmacodynamic properties, providing a comprehensive understanding of their suitability for drug development.

The results of our study can be used as a basis for further experimental investigations to validate the potential of these phytochemicals as drug candidates. our findings suggest that *Pinus Roxburghii* has a phytochemical name (E)-Nerolido, that has the potential to be effective treatment for non-small-cell lung cancer (NSCLC). However, further research is needed to confirm the efficacy of these compounds and to optimize their use in the treatment of this disease. In conclusion, the integrative in silico analysis of *Pinus Roxburghii* phytochemicals for drug discovery in non-small-cell lung cancer (NSCLC) represents a significant step forward in the search for effective treatments for this devastating disease, and I look forward to further research in this area.

Acknowledgement

I would like to express my sincere gratitude to my mentor for their invaluable support and guidance throughout this work. Their expertise and insights have been instrumental in shaping my ideas and bringing this project to fruition. I would also like to thank Mr Haresh S Kalasariya from Centre for Natural Products Discover, Liverpool John Moores University, UK. For their contribution and Providing resource for this work, which has been essential in achieving the goals of this work. I would like to extend my thanks to all those who have supported me along the way, including Faculties and Non- teaching staff of Department of Microbiology, Nootan Science and Commerce College, Visnagar for their encouragement, patience, and understanding and becoming help full whenever I need.

Author contribution

Nayankumar Prajapati designed and conducted the research and work of writing is performed. Dr Nikunj Patel is Mentoring this all work to be performed.

Informed consent

Not applicable.

Ethical approval

Not applicable.

Conflicts of interests

The authors declare that there are no conflicts of interests.

Funding

The study has not received any external funding.

Data and materials availability

All data associated with this study are present in the paper.

REFERENCES AND NOTES

1. Benfenati E, Manganaro A, Gini GC. VEGA-QSAR: AI inside a platform for predictive toxicology. CEUR Workshop Proc 2013; 1107:21–28.
2. Bhardwaj K, Sharma R, Cruz-Martins N, Valko M, Upadhyay NK, Kuča K, Bhardwaj P. Studies of phytochemicals, antioxidant, and antibacterial activities of Pinus gerardiana and Pinus Roxburghii seed extracts. Biomed Res Int 2022.
3. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. A census of human cancer genes. Nat Rev Cancer 2004; 4(3):177-183.
4. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144(5):646-74.
5. Herbst RS, Baas P, Kim DW. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet 2016; 387(10027):1540-1550.
6. Kalemkerian GP, Narula N, Kennedy EB, Biermann WA, Donington J, Leighl NB, Lew M, Pantelas J, Ramalingam SS, Reck M, Saqi A, Simoff M, Singh N, Sundaram B. Molecular Testing Guideline for the Selection of Patients with Lung Cancer for Treatment with Targeted Tyrosine Kinase Inhibitors: American Society of Clinical Oncology Endorsement of the College of American Pathologists/International Association for the Study of Lung Cancer/Association for Molecular Pathology Clinical Practice Guideline Update. J Clin Oncol 2018; 36(9):911-919. doi: 10.1200/JCO.2017.76.7293

7. Kaushik D, Kumar A, Kaushik P, Rana AC. Analgesic and Anti-Inflammatory Activity of Pinus Roxburghii Sarg. Adv pharmacol Sci 2012; 2012:245431.
8. Kumar KK, Yugandhar P, Devi BU, Kumar TS, Savithramma, N, Neeraja P. Applications of in silico methods to analyze the toxicity and estrogen receptor-mediated properties of plant-derived phytochemicals. Food Chem Toxicol 2019; 125:361-369.
9. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. J Pharmacol Toxicol Methods 2000; 44(1):235-249.
10. Kumar R, Kumar P, Upadhyay RS. Control of brown spot needle blight of Pinus Roxburghii by Trichoderma harzianum and Pseudomonas fluorescens. J Plant Prot Res 2012.
11. RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. Protein Sci 2018; 27:316?330. doi: 10.1002/pro.3331
12. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, Waterkamp D. Pembrolizumab versus chemotherapy for PD-L1–positive non–small-cell lung cancer. N Engl J Med 2016; 375(19):1823-1833.
13. Rose PW, Prlić A, Altunkaya A, Bi C, Bradley AR, Christie CH, Costanzo LD, Duarte JM, Dutta S, Feng Z, Green RK, Goodsell DS, Hudson B, Kalro T, Lowe R, Peisach E, Randle C, Rose AS, Shao C, Tao YP, Valasatava Y, Voigt M, Westbrook JD, Woo J, Yang H, Young JY, Zardecki C, Berman HM, Burley SK. The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. Nucleic Acids Res 2017; 45(D1):D271–D281.
14. Satyal P, Paudel P, Raut J, Deo A, Dosoky NS, Setzer WN. Volatile constituents of Pinus Roxburghii from Nepal. Pharmacog Res 2013; 5(1):43.
15. Soria JC, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, Lee KH, Dechaphunkul A, Imamura F, Nogami N, Kurata T, Okamoto I, Zhou C, Cho BC, Cheng Y, Cho EK, Voon PJ, Planchard D, Su WC, Gray JE, Lee SM, Hodge R, Marotti M, Rukazenkov Y, Ramalingam SS; FLAURA Investigators. Oimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. N Engl J Med 2018; 378(2):113-125.
16. Sundbom LT, Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. J Thorac Oncol 2018; 13(10):1353-1365.
17. Thapa R, Upreti A, Pandey BP. Chemical profiling and biological activity analysis of cone, bark, and needle of Pinus Roxburghii collected from Nepal. J Intercult Ethnopharmacol 2018; 1(1):66-75.
18. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science 2013; 339(6127):1546-1558.
19. Xie XQ. Exploiting PubChem for virtual screening. Expert Opin Drug Discov 2010; 5(12):1205-1220.
20. Yu R, Chen L, Lan R, Shen R, Li P. Computational screening of antagonists against the SARS-CoV-2 (COVID-19) coronavirus by molecular docking. Int J Antimicrob Agents 2020; 56(2):106012.