MOLECULAR DOCKING AND MACHINE LEARNING APPROACH TO ELUCIDATE Mentha aquatica PHYTOCHEMICALS ROLE IN BREAST CANCER TREATMENT

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE
OF

MASTER OF SCIENCE
IN **BIOTECHNOLOGY**

Submitted by:

RUCHI TIRKEY
2K21/MSCBIO/36

Under the supervision of
PROF. YASHA HASIJA

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

MAY, 2023

MOLECULAR DOCKING AND MACHINE LEARNING APPROACH TO ELUCIDATE Mentha aquatica PHYTOCHEMICALS ROLE IN BREAST CANCER TREATMENT

MASTER OF SCIENCE (BIOTECHNOLOGY)

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE
OF

MASTER OF SCIENCE
IN **BIOTECHNOLOGY**

Submitted by:

RUCHI TIRKEY
2K21/MSCBIO/36

Under the supervision of
PROF. YASHA HASIJA

[RUCHI TIRKEY]

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

MAY, 2023

2023

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi -110042

CANDIDATE'S DECLARATION

I, Ruchi Tirkey, Roll No. 2K21/MSCBIO/36 of MSc. Biotechnology, hereby declare that the project Dissertation titled "***MOLECULAR DOCKING AND MACHINE LEARNING APPROACH TO ELUCIDATE Mentha aquatica PHYTOCHEMICALS ROLE IN BREAST CANCER TREATMENT***" which is submitted by me to the Department of Biotechnology, Delhi Technological University, New Delhi in partial fulfillment of the requirement for the award of the degree of Master of Science, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any degree, Diploma Associateship, Fellowship or other similar title or recognition.

Title of the paper: In-silico approach to DNA codon optimization of Zaire Ebola virus genes for vaccine development.
Author's Names: Tirkey, Ruchi; Joshi, Soniya; Hasija, Yasha
Name of Conference: IEEE 2023 8th International Conference for Convergence in Technology (I2CT)
Conference Date and Venue: 7-9th April at Hotel Vivanta Pune (TAJ), Hinjawadi, Pune, Maharashtra, India
Registration: Done
Status of Paper Publication: Waiting
Date of Paper Communication: 28th February 2023
Date of Paper Acceptance: 12th March 2023

PLACE: Delhi **RUCHI TIRKEY**

Date: **2K21/MSCBIO/36**

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of
Engineering)Bawana Road, Delhi-
110042

CERTIFICATE

I hereby certify that the Dissertation titled "***MOLECULAR DOCKING AND MACHINE LEARNING APPROACH TO ELUCIDATE Mentha aquatica PHYTOCHEMICALS ROLE IN BREAST CANCER TREATMENT***" which is submitted by Ruchi Tirkey, 2K21/MSCBIO/36, Department of Biotechnology, Delhi Technological University, New Delhi is in partial fulfillment of the requirement for the award of the degree of Master of Science, is a record of the work carried out by the student under my supervision. To the best of my knowledge, this work has not been submitted in part or full for any Degree or Diploma to thisUniversity or elsewhere.

Place: Delhi
Date:

PROF. YASHA HASIJA
(SUPERVISOR)

Professor
Department of Biotechnology
Delhi technological university

PROF. PRAVIR KUMAR
HEAD OF DEPARTMENT (HoD)

Department of Biotechnology
Delhi Technological University

ACKNOWLEDGEMENT

I, **Ruchi Tirkey**, am honored to present the dissertation project as a part of the M.Sc. Biotechnology curriculum. Without the help and direction of everyone who helped me along the way, I would not have been able to finish this project.

I want to express my appreciation to my supervisor, **Prof. Yasha Hasija** of Delhi Technological University, for allowing me to conduct the research and for her ongoing guidance, support, and encouragement that enabled the project's successful completion.

I also want to express my gratitude to the **PhD Scholars** from **Delhi Technological University**, **Mr. Rajkumar Chakraborty**, **Mrs. Jaishree Meena**, and **Ms. Neha Kumari**, for giving me the technical know-how, support, and direction I needed to complete the project. I also want to express my gratitude to everyone who, in some way, contributed to the success of this project.

RUCHI TIRKEY

2K21/MSCBIO/36

ABSTRACT

With rising cases of breast cancer and developing resistance to available drugs for treatment, the need for novel drug inhibitors that produce an effective response becomes crucial. Phytochemicals are currently being studied for their therapeutic effects on a wide range of diseases. The present study focuses on targeting and inhibiting CDK6, HER2 and ER proteins that are upregulated in breast cancer by phytochemicals present in the plant *Mentha aquatica* commonly referred to as the water mint. It has an abundance of phytochemicals that have wonderful healing properties and are used to alleviate stress, reduce gastrointestinal discomfort, reduce joint and muscle aches, etc. The present study examines the phytochemical profile of *Mentha aquatica* and its possible role in treatment for breast cancer via in-silico studies. Molecular docking of 84 compounds was done to analyse their binding affinities towards the target proteins using AutoDock Vina. The results are validated using the PLIP-Protein ligand interaction profiler, and validation and classification by machine learning algorithms on the basis of IC50 values finally gave us 11 active compounds with binding affinities similar to or slightly better than the standard inhibitors selected for the study. Furthermore, ADMET analyses were done to predict the drug likeness via Lipinski's rule of 5, bioavailability, toxicity, and absorption of these compounds for predicting their efficacy against breast cancer as medicative agents.

Keywords—*Breast cancer, CDK6, HER2, ER, docking, machine learning, phytochemicals, M. aquatica*

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CDK6	Cyclin dependent kinase 6
HER2	Human epidermal growth factor 2
ER	Estrogen receptor
ERα+	Estrogen receptor alpha receptor
HR	Hormone receptor
PLIP	Protein ligand interaction profiler
ADMET	Absorption Distribution Metabolism Excretion Toxicology
TNBC	Triple negative breast cancer
Rb	Retinoblastoma proteins
EGFR	Epidermal growth factor receptor
NF-κB	Nuclear factor kappa B
GLOBOCAN	Global cancer observatory
BRCA – 1/2	Breast cancer gene 1/2
HRT	Hormone replacement therapy
CEM	Contrast enhanced mammography
DBT	Digital breast tomosynthesis
MRI	Magnetic resonance imaging
DFS	Disease free survival
PI3KCA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
TP53	Tumour protein 53
SERM	Selective estrogen receptor modulators
SERD	Selective estrogen receptor down regulators
AI	Aromatase inhibitor
NACT	Neo-adjuvant chemotherapy
CTLA4	Cytotoxic T-lymphocyte–associated antigen 4
PD-1	Programmed cell death protein 1
PD-L1	Programmed cell death ligand 1
FDA	U.S Food and Drug Administration
IgG	Immunoglobulin G
MAPK	Mitogen-activated protein kinase
PI3K	phosphatidylinositol 3-kinase

AIB1	Amplified in breast cancer 1
MAO	Monoamine oxidase
GABA-A	γ-Aminobutyric acid type A
G2	Growth 2 phase
M	Mitotic phase
RCSB	Research Collaboratory for Structural Bioinformatics
PDB	Protein Data Bank
SMILE	Scheme of Support for Marginalized Individuals for Livelihood and Enterprise
IC50	Inhibitory concentration 50
ML	Machine learning
HIA	Human intestinal absorption
CL	clearance
PPB	Plasma protein binding
CYCP450	Cytochrome P 450
h-ERG	human ether-a-go-go related gene
IMPPAT	Indian Medicinal Plants, Phytochemistry and Therapeutics
QASR	Quality structure activity relationship
LQTS	Long QT syndrome

CHAPTER 1

INTRODUCTION

1.1 Breast cancer

Breast cancer has been observed as a frequent malignancy among women globally and even with the rapid advancements in the field of therapeutics, good percentage of patients have exhibited resistance to the ongoing treatment methods. Based on biomarkers profiling, breast cancer has several molecular subtypes – Luminal A, Luminal B, human epidermal growth factor 2 (HER2) and TNBC (Triple negative breast cancer). Prognosis and treatment methods differ depending on these subtypes. Therapeutic strategies for cancer treatment mainly involve surgery, endocrine therapy, adjuvant therapy, chemotherapy, radiotherapy or immunotherapy with monoclonal antibodies [1]. Few examples of target proteins which play a direct or an indirect role in cell division and cell metabolism are CDKs (cyclin-dependent kinases), HER2 (Human epidermal growth factor receptor 2) and Erα+ (Estrogen receptor alpha positive). Despite the development of different drug inhibitors, cases have been observed where the patients show unresponsiveness to the administered drug or after a certain period of time and develop resistance, hence a recent trend of using phytoconstituents in treatment of cancers has been observed [1],[2]. Plant derived formulations pose reduced side effects and target several pathways that help in controlling cell growth. Derived from various parts of plants, these phytochemicals have been identified as inhibitors of protein kinases, EGFRs and ER while showing antioxidant, anti-cancerous and anti-inflammatory properties [3], [4], [5]. The thesis mainly focuses on identifying phytochemicals of *Mentha aquatica* as potential inhibitors of CDK6, HER2 and Erα+ proteins through molecular docking followed by incorporation of machine learning algorithm to validate the activity of docked phytochemicals in inhibiting breast cancer progression.

1.2 Statistics

With being the most prevalent among women, breast cancer has now become most diagnosed cancer after surpassing lung cancer. An estimate of 2.3 million cases were reported in 2020 and on the basis of data acquired from GLOBOCAN 2020 database, a collection of data on cancer estimates from 185 different countries, it has been predicted that if the recent trends of population growth continue the number of cases may rise to 3 million with 1 million/year deaths [6]. Research study conducted showed that breast cancer observed in pre-menopausal women showed tumours of higher histological grade and proliferation rate along with decreased probability of overall survival, which were much more aggressive in nature when compared to breast cancers in post-menopausal women. Although early breast cancer might be curable, metastatic breast cancer cases have seen high mortality [7]. Prevalence of male breast cancer is rare but do occur and account for <1% of the cases. Lack of awareness and rarity of the disease have led approximately 81.5% cases to be diagnosed at tumour stage. Due to limited data, presently there has been no separate research on therapeutic approaches specific to male breast cancer [8].

CHAPTER 2

LITERATURE REVIEW

2.1 Risk factors

2.1.1 Age

Most frequently it has been diagnosed in women below 40 years of age and less than 64 years of age. Almost 5% these breast cancer cases accounts for women under 40 years of age. With age the risk of cancer increases, although pre-menopausal breast cancers have found to have decreased survival rate [9].

2.1.2 Genetics

Probability of developing breast cancer increases 1.75 times higher when a close relative is diagnosed with breast cancer. This risk factor increases with number of family members having breast cancer [10]. Inheritability of breast cancer has been linked with autosomal dominant mutations of BRCA-1 and BRCA-2 [9].

2.1.3 Reproductive factors

Early menarche causes early ovulatory cycles increases the risk of breast cancer. Cancer risk reduces by 5% as menarche is delayed by 1 year [9], [10]. Breast cancer risks have found to be relatively high in patients with late menopause. With each 1 year delay the risk factor increased by 1.029 [9]. Studies have related late age at first birth may also act as a risk factor. Also with each additional live birth the risk factor for ER+/PR+ cancer reduces by 11% [9].

2.1.4 Estrogen

Uncontrolled estrogen levels may pose as a high risk factor of breast cancer. It is produced endogenously by ovary in pre-menopausal women and exogenously by the use of oral contraceptives such as HRT. Relative risk factor between HRT users and non-users was

found to 1.66. However, the risk associated with HRT decreases within 2 years of withdrawal.

2.1.5 Lifestyle

Lifestyle habits such as excess intake of alcohol and dietary fat increase the chances of breast cancer. Excessive alcohol may result in estrogen related hormones which further activate estrogen receptors. High fat has been related to high mortality risks and poor prognosis [9].

2.2 Diagnosis

2.2.1 Mammography

Mammography is a preferable technique to analyse breast cancer. Contrast-enhanced mammography (CEM) and digital breast tomosynthesis (DBT) are the 2 strategies that are presently being used and have shown better accuracy and higher specificity, respectively. Mammography is an inexpensive method with high accuracy and screening rate, but use of contrast agents and X-rays makes it unsuitable for use and it is generally not recommended for patient <40 years of age [11].

2.2.2 Ultrasonography

It is used for analysing morphological variations and locating lesions. It can be used as an alternative for mammography for all age groups as it involves less use of contrast agents and harmful radiations. But this method poses certain drawbacks such as low resolution and requires professional expertise [11].

2.2.3 Magnetic resonance imaging (MRI)

MRI helps in early diagnosis of breast cancer but it has certain limitations such as long imaging time, expensive and cannot be used in patients with metal implants [11].

2.2.4 Biopsy

It involves obtaining and analysing tumour tissues through histopathology and one strategy being currently used is core needle biopsy. It has advantages such as convenience, small trauma area and single puncture can obtain large samples. One major drawback of this method is that may cause tumour transfer leading to neoplastic seeding [11].

2.3 Molecular sub-types of breast cancer

Breast cancer expresses different biological markers – estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and Ki67, a prognostic marker that helps in cell division. Based on genetic profiling of tumours and biological markers expression, breast cancer can be grouped into 4 molecular subtypes - Luminal-A, Luminal-B, human epidermal growth factor 2 (HER2) and triple negative type [1], [12].

2.3.1 Luminal A

At transcriptomic and proteomic levels, Luminal A and B differ in genes regulating cell cycle and hormonal pathways as Luminal A has high ER, and low Ki67 (<14%) expression which is related to cell cycle progression. Mutations at genomic level are comparatively lesser in Luminal A breast cancers [13]. Luminal A tumours with >20% PR have shown better diseases free survival (DFS). PIK3CA E545K mutation has been unique to Luminal A breast cancers and is currently under study for its potential to target luminal A subtype. Compared to other subtypes, luminal A breast cancer patients have shown better survival rates with delayed risk of mortality [14].

2.3.2 Luminal B

Luminal B showcases high ER and Ki67 (>20%) expression whereas Luminal B mutations at genomic level are comparatively more than in Luminal A breast cancers [13]. ZNF703 is one of the oncogenes that stimulates cell proliferation and repair in luminal B tumours. They show lower PIK3CA mutations (29%) and higher (29%) TP53 mutations when compared to luminal A subtype (PIK3CA - 45%, TP53 – 12%). Comparatively this subtype

has much more aggressive clinical outcome with an increased rate of relapses after 5 years of diagnosis [15].

2.3.3 HER2

HER-2 subtype showcases a higher expression of HER-2 receptor and proliferation related genes, moderate to low expression of luminal and basal related genes at proteomic level. And do not express ER and PR receptors. At DNA level, HER-2 tumours show high mutations in TP53 and PIK3CA [13].

2.3.4 Triple negative breast cancer (TNBC)/ Basal type

In TNBC, ER, PR and HER-2 receptors are not expressed [12], [13]. Overexpression of basal related genes is observed and minimal expression of luminal related genes at proteomic level whereas at genomic level they show second highest mutations in form of hypomethylation in genes TP53 and PIK3CA [13].

2.4 Therapy

Therapeutic strategies for cancer treatment mainly involve surgery, endocrine (hormonal) therapy, adjuvant therapy, chemotherapy, radiotherapy or immunotherapy with monoclonal antibodies which target proteins that play a direct or indirect role in cell division and cell metabolism [1]. Often surgeries, chemotherapies and radiotherapies are combined with systemic therapies that involves the use of drugs depending on its effectiveness against a particular molecular subtype of breast cancer. Apart from this, adjuvant therapies (post-surgery) and neo-adjuvant therapies (pre-surgery) also help in increasing the efficacy of the treatments and increasing the chances of diseased free survival. Adjuvant therapies help in eliminating the latent metastatic tumours post-surgery and neo-adjuvant therapies help in increasing the rate and chances of breast conserving surgeries [12].

2.4.1 Surgery

Two main breast cancer surgeries include mastectomy and lumpectomy which involves breast removal and tumour lump surgeries, respectively. During mastectomy, entire breast including the surrounding tissue is removed, whereas in lumpectomy only the tumorous lumps are removed and can help in conserving majority of breast tissues during surgery. Lumpectomy is often an option for early breast cancer cases only, followed by radiation as adjuvant therapy [16]. Surgeries for dissecting lymph nodes helps in removing cancer cell migrated to axillary lymph nodes (underarm).

2.4.2 Endocrine therapy

Endocrine therapy involves blocking of hormones that promote tumorigenesis. Hormonal therapy has found to be highly effective against ER+ breast cancers [12]. This therapy works by blocking estrogen receptors or estrogen production. There are various classes of endocrine therapy drugs. Tamoxifen is a selective estrogen receptor modulator (SERM) that block estrogen expression by inhibiting the dimerization of ER. AI (Aromatase inhibitor) such as letrozole block estrogen by aromatase inhibition [17].

2.4.3 Chemotherapy

Chemotherapies involve used of drugs that are administered as neo-adjuvants, to decrease the size of tumours and increase chances of breast conservation during surgery. Neo-adjuvant chemotherapies (NACT) have shown higher pathological response to HER2 subtype when compared with ER+ subtype. ER+ gave 18% and 31% response before and after administration trastuzumab. HER2 gave 30% and 50% response before and after trastuzumab treatment. With pertuzumab, HER2 response was high as 80% [18].

2.4.4 Radiotherapy

This method uses X-rays which have high energy and are capable of killing cancer cells. Patients undergoing mastectomy and lumpectomy also undergo radiotherapy if the tumour size was big to remove any remaining cancer cells in the surrounding lymph nodes and skin tissues [19]. Radiation therapy can also be administered pre-surgically for tumour size reduction hence acting as a neo-adjuvant or it may be opted as an alternative if the tumour

cannot be removed surgically. However, exposure of breast tissue to radiations also affect the heart, and studies have noted increased risks of cardiovascular diseases after radiotherapies [20].

2.4.5 Immunotherapy

It is a class of treatments that administer medicines that enhances the immune system to detect and kill cancer cells. CTLA4 present on activated T cells when binds to antigen presenting cells via B7 receptor thus blocking T cell activation. To combat these, monoclonal antibodies such as Tremelimumab and Ipilimumab are administered for CTLA4 receptors on activated T cells. PD-1 are receptors upregulated in active T cells and binds to PD-L1. Cancer cells inhibhit T cell activation by producing PD-L1 receptors on its surface which bind to PD-1 of active T cells, leading to its inactivation. Prembrolizumab is a monoclonal antibody that acts as a checkpoint inhibitor of PD-1.

Breast cancer subtype	**Type of therapy**	**Therapy**	**Reference**
Premenopausal ER/PR positive	Hormonal	Tamoxifen (10 years)	[12]
Postmenopausal ER/PR positive	Hormonal	Aromatase inhibitor (5 years)	[12]
Axillary lymph node metastases	Chemotherapy	Anthracycline (A) regimen, TC regimen (docetaxel and cyclophosphamide)	[12]
HER-2 positive	Anti Her-2	Chemotherapy, Trastuzumab, Pertuzumab (1 year)	[12]

Table 2.1 Adjuvant therapies available for breast cancer.

Breast cancer subtype	Type of therapy	Therapy	Reference
All	Chemotherapy	AC regimen (Adriamycin and cyclophosphamide) + taxane regimen (docetaxel or paclitaxel)	[12]
HER-2 positive	Anti Her-2	Chemotherapy, Trastuzumab, Pertuzumab (1 year)	[12]
Triple negative	Chemotherapy	Carboplatin + AC regimen (Adriamycin and cyclophosphamide) + taxane regimen (docetaxel or paclitaxel)	[12]
Postmenopausal ER/PR positive	Hormonal	Aromatase inhibitor (5 years)	[12]

Table 2.2 Neo-adjuvant therapies available for breast cancer.

CHAPTER 3

BREAST CANCER BIOMARKERS

As mentioned, the difference in expression level of biomarkers such as ER and HER2, helps in categorizing breast cancer into its molecular sub-types. Apart from the receptors exclusive to breast cancer, certain other proteins can be targeted that are normally upregulated in other types of cancer. Protein kinases are one such protein that regulate cell cycle, transcription, DNA repair, and metabolism of proteins in different signalling pathways [3].

3.1 Cyclin dependent kinas 6 (CDK6)

Uncontrolled cell growth is a hallmark of cancerous cells and hence cyclin-dependent kinases (CDKs) can be targeted and have shown to be a promising target for novel drug inhibitors [3]. CDKs belong to the serine/threonine protein kinases and are responsible for regulated cell cycle progression along with CDK inhibitors (CDI) [3]. Non-selective CDK inhibitors failed to show positive response during the clinical trials and thus CDK6 is being targeted as a for slowing down the progression and growth of tumorigenic cells [2]. CDK6 mediates transition of G1 phase to S phase and in association with cyclin D proteins phosphorylates the retinoblastoma proteins (Rb). This causes disruption of E2F-Rb complex. Subsequently, the E2F transcription factor enters the nucleus and initiates transcription of genes that produce cyclin E1 and cyclin E2, which help in transition to S phase of the cell cycle and synthesizing DNA [2]. Both CDK6 and CDK4 share homology and share closely related functioning of cell cycle progression but studies have found role of only CDK6 in tumour progression and regulation of oncogene expression [21].

3.1.1 FDA approved CDK6 inhibitors

Palbociclib is a CDK6 inhibitor belonging to the class of pyridopyrimidines and is being used for the treatment of HR+, HER2 –breast cancers and has shown an effective response in combination with letrozole [22]. Ribociclib along with aromatase inhibitor has been found to be effective in post-menopausal women. Abemaciclib and antiestrogen are used for treating HR+, HER2- that metastatis and are administered after endocrine therapy [23]. All three drugs have shown good efficacy and have prolonged survival from 12-14 months to >25 months. Even after being used in combinational therapies, patients eventually developed resistance to these drugs [24]. Fisetin a natural flavonoid found in apple, cucumber etc. have shown antioxidant and anti-inflammatory activities and heped in slowing down of cancer cell progression. For instance, fisetin inhibits heat shock factor 1 thus blocking transcription which eventually leads to apoptosis, in gastric cancer it causes apoptosis induced by mitochondria and in pancreatic cancer it blocks the NF-κB pathway, again which leads to apoptosis [25].

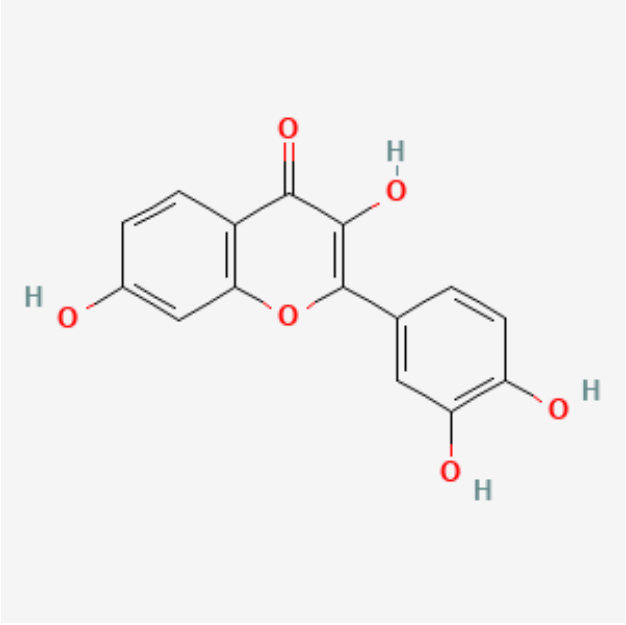


Figure 3.1 Structure of Fisetin [PubChem].

3.2 Human epidermal growth factor receptor 2 (HER2)

HER2 (Human epidermal growth factor receptor 2) are membrane glycoproteins which is a type of epidermal growth factor receptor (EGFR/ErbB) with 20% breast cancer cases showing overexpression and hence it holds potential to act as a target for drug inhibitors [26],[27]. HER proteins are basically tyrosine kinases which when bound to ligand undergo conformational changes. HER2 does not have a ligand so it undergoes phosphorylation and dimerization either with its self or with other HER proteins. Phosphorylated tyrosine kinases bind with intracellular signalling molecules and activate downstream signalling pathways which further activate transcription factors that regulate cell proliferation, angiogenesis and metastasis thus promoting tumour progression [28].

3.2.1 FDA approved HER2 inhibitor

Trastuzumab is a humanised monoclonal antibody that has been given clearance for the treatment of HER2+ type of breast cancers. It can be used in combination with taxane and chemotherapies with an increased survival advantage. Presently it is administered for 1 year in early breast cancer cases as adjuvant therapy [29]. Trastuzumab directly inhibits HER2 by downregulation and indirectly by activating immune response through ADCC (Antibody dependent cytotoxic cellular) pathway. After proteolytic cleavage of HER2, a truncated receptor with no extracellular binding part for Trastuzumab may retain and activate downstream signalling thus acquiring resistant to it [30]. Pertuzumab, an anti-HER2 IgG1 monoclonal antibody that prevents dimerization of HER2 [31]. Kinase inhibitors such as Lapitinib can target both HER2 and EGFR. Research study conducted used a pyrrolo [3,2-d] pyrimidine-based potent inhibitor which was highly selective to tyrosine kinase residues of HER2A chain [32]. Nernatinib another HER2 inhibitor is able to increase disease free survival rates when given after chemotherapy and trastuzumab therapy [31].

3.3 Estrogen receptor alpha positive (Erα+)

ERα+ (Estrogen receptor alpha positive) is most common subtype with a 70% occurrence rate. These receptors are present in the mammary epithelium and normally help in development of mammary glands. But when overexpressed they are responsible for cancer progression [33]. ER consists of α and β domains, of which α domain binds to estrogen thus activating the ER receptor. The activated ERα binds to the promoter region of gene and initiates transcription. Thus, overexpression of ERα+ may cause unregulated transcription and promote tumour progression. Co-regulatory proteins of ERα such as A1BI are overexpressed and promote metastasis. ERα also plays an important role in extracellular signalling by activating various pathways which contribute in metastasis of breast cancer [33].

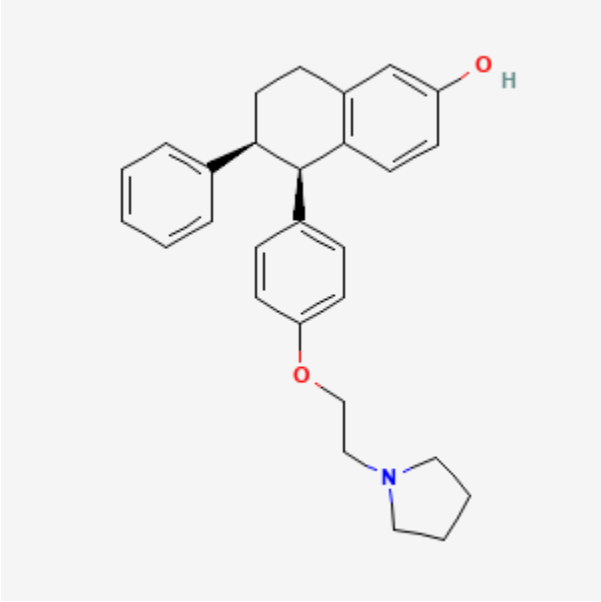


Figure 3.2 Structure of Lasofoxifene [PubChem].

3.3.1 FDA approved ERα+ inhibitors

Tamoxifen is an ER agonist that comes under the category of SERM (selective estrogen receptor modulators) and shows effective response to ER+ sub-type. Although the efficacy rate of Tamoxifen has been high, certain side effects such as increased risk of cancer, irregular menstrual cycle, uterine cancer and thromboembolic phenomena has also been

observed. It is administered for only 5 years after which other therapeutics are administered depending on patient's health [34]. Lasofoxifene is also a SERM belonging to the class of tetralins that binds strongly to the ER α and β domains and was originally planned to be used for treating osteoporosis but was not. It has not been administered as a direct therapy for breast cancer but studies conducted found its antagonist properties remained unaffected with ERα mutations and it promoted better antitumor activities in pre-clinical models of breast cancer [35].

CHAPTER 4

PHYTOCHEMICAL

4.1 Phytochemicals as a replacement for drug inhibitors

Prediction of effective and novel treatment methods differ with the variability of subtypes. Even with the rapid advancements in the field of therapeutics, a large proportion of patients acquire resistance to the ongoing therapies. Due to heterogenic expression of biomarkers of these subtypes and extreme variability in physical health among patients, it becomes difficult to predict clinical outcomes of a particular therapy and thus need for precision therapy specifically modelled in account with patients' health becomes important. For instance, aggressively proliferating tumours may initially respond to chemotherapies but then eventually metastasize and show drug resistance [36], [37]. Plants contain a variety of phytochemicals that have shown therapeutic and anti-cancer properties previously and are therefore being utilized for drug discovery. Although phytochemicals usually show low water solubility and are required in high doses to exert an effective inhibitory action on oncogenic signalling pathways, several phytochemical analogs with improved physiochemical and anti-cancer properties are being developed for improving its efficacy [36]. There has been a recent trend of studying these bioactive phytochemicals from various medicinal plants and comparing them with the conventional chemically synthesized drug inhibitors. Several research studies have also proposed combinational therapies using phytochemicals with other therapeutic methods, that not only increase the efficacy but also decrease the adverse side effects of chemical drug inhibitors such as muscle stiffness, risk of cardiovascular diseases, sleep disturbance, fatigue etc., to some extent [38], [39], [40].

4.2 Therapeutic properties of *Mentha aquatica*

Mentha aquatica also known as water mint, is a flowering plant mostly found in moist conditions and popularly used for treating gastrointestinal and pulmonary disorders, along with

mental health disorders like epilepsy, depression. Extracts like triterpenoids, phenolic compounds, and flavonoids have been obtained from *M. aquatica* and have shown inhibitory properties against monoamine oxidase (MAO), GABA-A receptor, pancreatic lipase and α-amilase, along with antimicrobial and anti-inflammatory properties. Inhibitory effects of some extracts on cell lines of cervical cancer, breast cancer, prostate cancer and melanoma have also been studied [41]. Research conducted on anticancer activity of various dietary plants found *M. aquatica* comparatively contained the highest content of phenolic compounds (337 mg/g). Phenolic compounds inhibit cell proliferation and contribute in tumour cell killing. While checking growth inhibition properties on different cancer cell lines, *M. aquatica* showed excellent proliferation inhibition of breast cancer cell line (MCF-7) [42]. Although literature on *Mentha aquatica* 's direct therapeutic effect on breast cancer is limited, it has been researched for skin carcinogenesis. Essential oils derived from *M. aquatica* were able to induce G2/M arrest and suppress MAPK activation in cancer cell lines, thus inhibiting cancer cell proliferation [43]. The thesis focuses on utilizing the phytochemicals of *Mentha aquatica* and checking their binding affinities with the target proteins.

CHAPTER 5

METHDOLOGY

5.1 Acquisition of target proteins and phytochemicals

5.1.1 CDK6

As discussed, CDK-6 was selected as one of the target proteins for this research. RCSB PDB (https://www.rcsb.org/) was used to extract 3D PDB file of cyclin CDK-6 (PDB ID- 1XO2) with a flavanol like inhibitor called Fisetin.

5.1.2 HER2

RCSB PDB was used to extract 3D PDB file of HER-2 Kinase Domain (PDB ID- 3PP0) which was present in complex with inhibitor SYR127063 and TAK285.

5.1.3 ERα

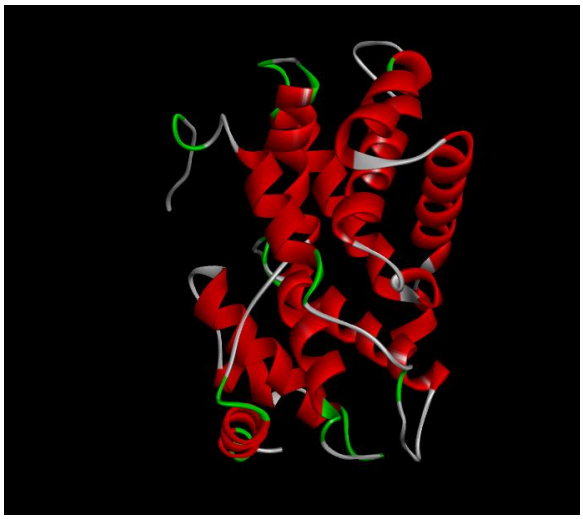
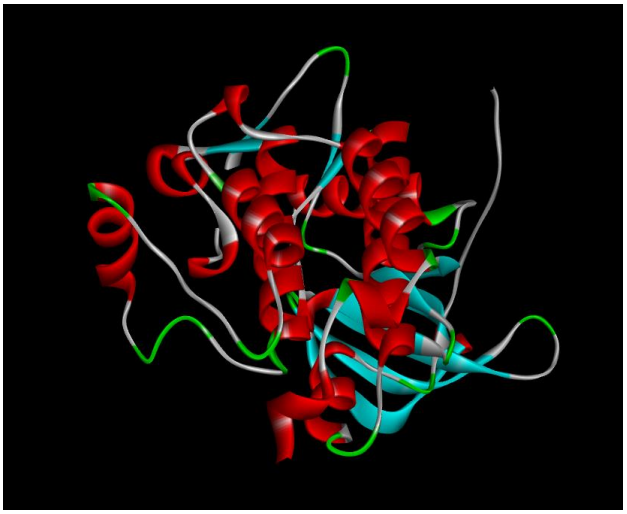
3DPDB file of ERα complexed with Lasofoxifene was retrieved (PDB ID- 2OUZ).

5.1.4 Phytochemical extraction

Out of 220 Phytochemicals of *Mentha aquatica* or water mint from IMPPAT, 84 phytochemicals were acquired in 3D PDB format after removing eliminating redundant compounds [44], [45]. All the phytochemicals/ ligand files were modified to pdbqt format using OpenBabel software to prepare them for docking.

5.2 Preparation of target proteins

PDB files of target proteins were prepared in Biovia Discovery studio. Water molecules were removed, ligand sites were selected and then ligands were also removed. For CDK6, chain A was removed. For HER2, chain B and TAK285 were removed. After this, polar hydrogens were added and the prepared proteins were saved as a pdb file and proceed to Autodock Tool software to add Kollman charges and the grid.

a)  b) 

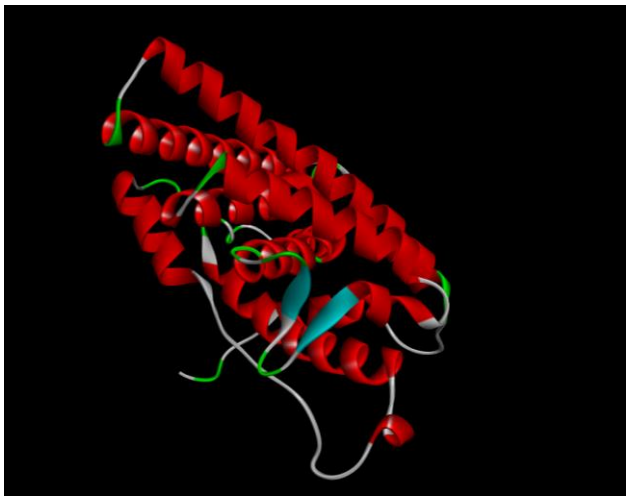
c) 

Figure 5.1 Proteins prepared using Discovery studios **a)** CDK6 B chain **b)** HER2 kinase domain **c)** ERα receptor.

5.3 Molecular docking using AutoDock Vina

All the pdbqt files of ligands and the prepared protein files were kept in a single folder. Perl and Autodock Vina [46] were downloaded. A configuration text file with grid dimensions of ligand binding site was also prepared and saved in the same folder. Finally, docking was performed for each of the 84 ligands that are potential inhibitors with each of the three proteins.

5.4 Visualisation using PyMOL and PLIP

PyMOL helps in visualising if the chosen ligands interact with active binding site of the target protein. The interacting amino acids can also be identified here only. But to further justify results, each ligand and protein complex was downloaded and analysed using PLIP (https://plip-tool.biotec.tu-dresden.de/plip-web/plip/index). PLIP is a well-known Protein-Ligand interaction profiler that tells which amino acid of protein is interacting with the ligand [47]. According to binding affinity values and PLIP interaction results that gave maximum number of interacting amino acids, ligands were selected. Their SMILE sequences were acquired and recorded in an excel file. This excel file was used to generate a descriptor file to proceed for machine learning validation.

5.5 Validation by Machine Learning

Random forest algorithm was prepared to train the data where 80% of it is used for training and the rest 20% for testing the accuracy. Based on standard IC50 values of the data taken from the library, the ligands were categorized as 'Active', 'Intermediate' and 'Inactive'. This helped in predicting whether or not the ligand, showing good binding affinity and number of interacting amino acids similar to our standard inhibitor, is actually a potent drug.

5.5.1 Preparing Descriptors

Webresource_client library was used from CHEMBL database to gather data.

```
[7]
    ! pip install chembl_webresource_client

[8] import pandas as pd
    from chembl_webresource_client.new_client import new_client
    pd.set_option('display.max_columns', 500)
```

Figure 5.2 Installing websource_client from CHEMBL database.

PaDELPy was used to prepare descriptors of selected phytochemicals. PaDELPy is an open-source package that offers a molecular descriptor calculation program and a Python wrapper for the PaDEL-Descriptor. While working with scientific data, the PaDEL-Descriptor can be used to determine the molecular fingerprint of particular molecules that are then utilized to create machine learning models in the field of science.

```
[15] from padelpy import padeldescriptor
     fingerprint = 'Substructure'
     fingerprint_output_file = ''.join([fingerprint,'.csv'])
     fingerprint_descriptortypes = fp[fingerprint]

     padeldescriptor(mol_dir='mol.smi', d_file=fingerprint_output_file,
                     descriptortypes= fingerprint_descriptortypes,
                     detectaromaticity=True,
                     standardizenitro=True,
                     standardizetautomers=True,
                     threads=2,
                     removesalt=True,
                     log=True,
                     fingerprints=True)

     descriptors = pd.read_csv(fingerprint_output_file)
```

Figure 5.3 Preparing descriptor file using PaDELPy.

The descriptor file was downloaded and uploaded in the Machine learning algorithm.

a)

descriptors

	Name	SubFP1	SubFP2	SubFP3	SubFP4	SubFP5	SubFP6	SubFP7	SubFP8	SubFP9	...	SubFP298	SubFP299	SubFP300	SubFP301	SubFP302	SubFP303	SubFP304	SubFP305	Subf
0	alphaterpineol	1	1	1	0	1	0	0	0	0	...	0	0	1	1	1	0	0	0	
1	caryophylleneoxide	1	1	1	1	1	0	0	0	0	...	0	0	1	1	0	0	0	0	
2	cymene	1	0	1	0	0	0	0	0	0	...	0	0	1	1	1	0	0	0	
3	dborneol	1	1	1	1	0	0	0	0	0	...	0	0	1	1	0	0	0	0	
4	humeleneepoxidell	1	1	0	1	1	0	0	0	0	...	0	0	1	1	0	0	0	0	
5	isomenthone	1	1	1	0	0	0	0	0	0	...	0	0	1	1	1	0	0	0	
6	isomintlactone	1	1	1	0	1	0	0	0	0	...	0	0	1	1	0	1	0	0	
7	methylenecyclodecaciene	1	1	1	0	1	0	0	0	0	...	0	0	1	1	1	0	0	0	

8 rows × 308 columns

b)

```
[15] descriptors
```

	Name	SubFP1	SubFP2	SubFP3	SubFP4	SubFP5	SubFP6	SubFP7	SubFP8	SubFP9	...	SubFP298	SubFP299	SubFP300	SubFP301	SubFP302
0	alloaromadendrene	1	1	1	1	1	0	0	0	0	...	0	0	1	1	0
1	alphacadinol	1	1	1	0	1	0	0	0	0	...	0	0	1	1	1
2	alphacopaene	1	1	1	1	1	0	0	0	0	...	0	0	1	1	1
3	dihydrocarvylacetate	1	1	1	0	1	0	0	0	0	...	0	0	1	1	1
4	isomintlactone	1	1	1	0	1	0	0	0	0	...	0	0	1	1	0

5 rows × 308 columns

c)

```
descriptors
```

	Name	SubFP1	SubFP2	SubFP3	SubFP4	SubFP5	SubFP6	SubFP7	SubFP8	SubFP9	...	SubFP298	SubFP299	SubFP300	SubFP301	SubFP
0	transalphabergamotene	1	1	1	1	1	0	0	0	0	...	0	0	1	1	
1	humuleneepoxidell_	1	1	0	1	1	0	0	0	0	...	0	0	1	1	
2	gammaterpinene	1	1	1	0	1	0	0	0	0	...	0	0	1	1	
3	carvone	1	1	1	0	1	0	0	0	0	...	0	0	1	1	
4	betapinene	1	1	1	1	1	0	0	0	0	...	0	0	1	1	
5	betabourbonene	1	1	1	1	1	0	0	0	0	...	0	0	1	1	
6	alphaguaiene	1	1	1	0	1	0	0	0	0	...	0	0	1	1	
7	alphacurcumene	1	1	1	0	1	0	0	0	0	...	0	0	1	1	
8	4carvomenthenol	1	1	1	0	1	0	0	0	0	...	0	0	1	1	

9 rows × 308 columns

Figure 5.4 Descriptors prepared for a) CDK6, b) HER2 and c) ERα using PaDELPy.

5.5.2 Machine Learning algorithm

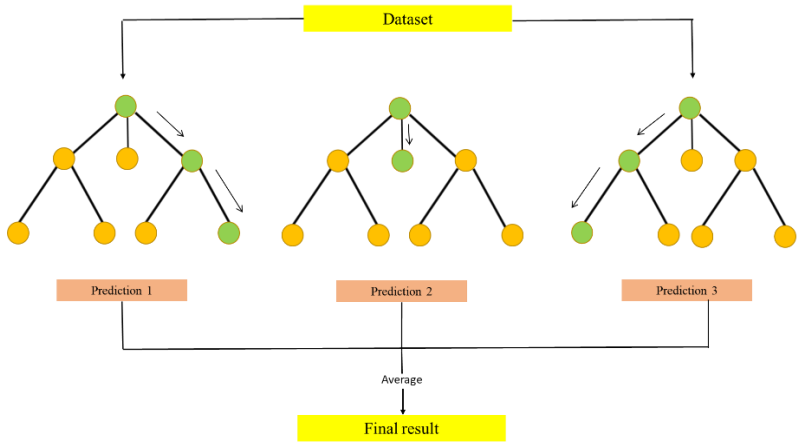


Figure 5.5 Random Forest classifier model.

Random Forest Classifier model was used and training/testing data was acquired from CHEMBL database. Based on their standard IC50 values stored in the library, the ligands

were categorized as ‘Active’, ‘Intermediate’ and ‘Inactive’. This helped in predicting whether or not the ligand, showing good binding affinity and number of interacting amino acids similar to our standard inhibitor, is actually a potent drug.

5.6 Physicochemical and pharmacokinetic analysis

ADMETlab 2.0 and admetSAR 2.0 [48], [49] were used to perform the ADMET profiling that is Absorption, Distribution, Metabolism, Excretion, Toxicity was estimated based on different parameters. Along with this ‘Lipinski rule of five’ was also tested based on physiochemical properties of the phytochemicals. Finally, a conclusion is drawn to predict if any of the phytochemicals were suitable for use in humans.

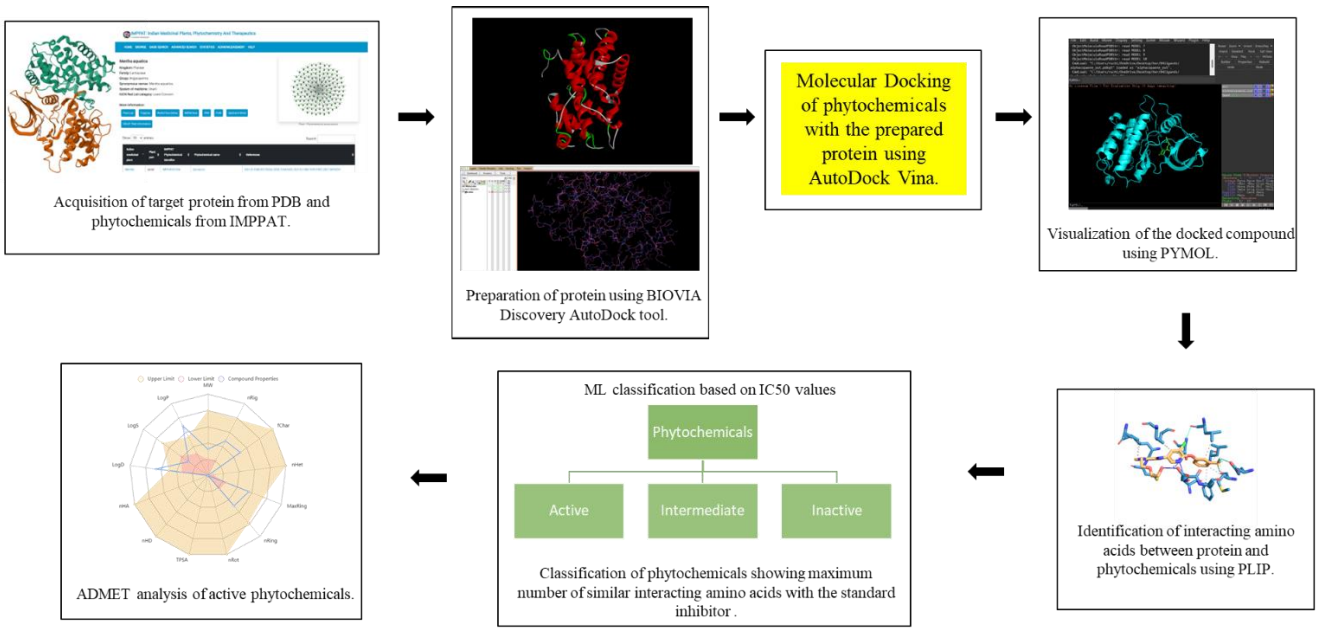


Figure 5.6 Basic steps for methodology.

CHAPTER 6

RESULTS

6.1 Docking analysis

6.1.1 CDK6

Out of 84 phytochemicals extracted from water mint (*M. aquatica*), 46 showed docking scores in the range of standard inhibitor Fisetin that is -6.5 Kcal/mol with CDK6 protein.

6.1.2 HER2

Out of 84 phytochemicals, 28 phytochemicals showed docking scores higher than the inhibitor SYR127063 that is -7.3 Kcal/mol

6.1.3 ERα+

Out of 84 phytochemicals, none of them showed docking score higher than the inhibitor lasofoxifene (-8.6 Kcal/mol), but 9 of the phytochemicals docking score were in close proximity with the standard inhibitor's docking score.

6.2 PLIP analysis

PLIP analysis revealed that only 8 ligands showed a good similarity between the interacting amino acids of test ligands and the standard ligand with the target protein CDK-6 as shown in table 6.1.

Only ligands showing a good similarity between the interacting amino acids of test ligands and the standard ligand with the target protein HER2 as shown in table 6.2.

Since the obtained ligands showing comparable docking scores with the standard inhibitor were less in number, eliminating ligands on the basis of interacting amino acids, was skipped for ERα protein and all the 9 phytochemicals were checked for its activity using machine learning algorithm.

S.no.	Phytochemicals	Binding affinity	Number of amino acids similar to standard protein-ligand binding
1.	Alpha terpineol	-7.4	5
2.	Caryophyllene oxide	-6.7	5
3.	Cymene	-6.6	5
4.	Iso mint lactone	-6.8	6
5.	Iso menthone	-6.7	5
6.	Methylene cyclodecadiene	-6.8	5
7.	D-borneol	-6.9	5
8.	Humulene epoxide II	-7.2	5

Table 6.1 Ligands selected after PLIP analysis that exhibited similar interacting amino acids for CDK6 protein.

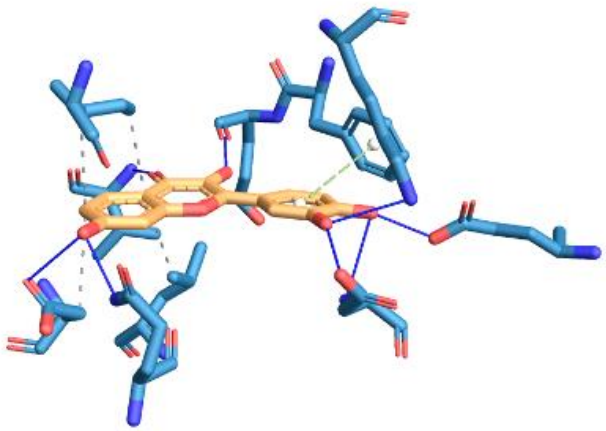


Figure 6.1 Protein ligand interaction between target protein CDK6 B chain and standard inhibitor Fisetin,

S.no.	Phytochemicals	Binding affinity	Number of amino acids similar to standard protein-ligand binding
1.	Allo-Aromadendrene	-7.3	6
2.	Alpha-Cadinol	-8	8
3.	Alpha-Copaene	-8.5	8
4.	Dihydrocarvyl acetate	-7.6	5
6.	Isomint lactone	-7.1	5

Table 6.2 Ligands selected after PLIP analysis that exhibited similar interacting amino acids with HER2 protein.

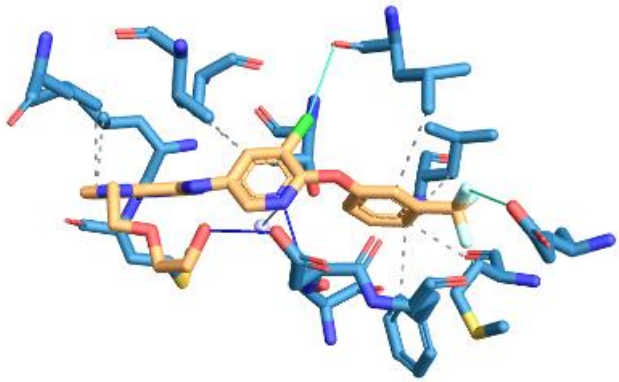


Figure 6.2 Protein ligand interaction between target protein HER2 A chain and standard inhibitor SYR127063.

S.no.	Phytochemicals	Binding affinity
1.	Trans-alpha-Bergamotene	-8.1
2.	Humuleneepoxidell_	-7.9
3.	Gamma-Terpinene	-8
4.	Carvone	-8.1
5.	Beta-pinene	-8
6.	Beta-bourbonene	-8
7.	Alpha-guaiene	-7.9
8.	Alpha-curcumene	-8
9.	4-Carvomenthenol	-8.2

Table 6.3 Docking scores of ligands with ERα protein.

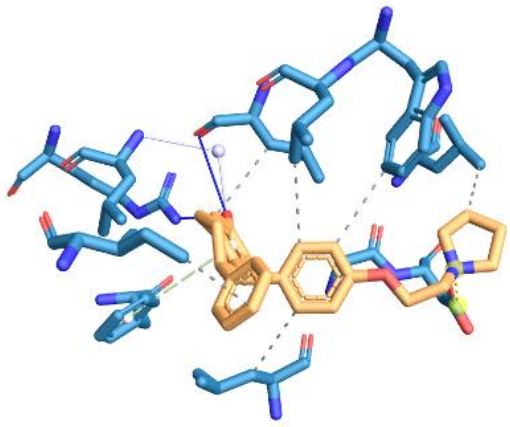


Figure 6.3 Protein ligand interaction between target protein ERα and standard drug inhibitor lasofoxifene.

6.3 ML validation

These 8 ligands were then validated for their potential as a drug using a Machine learning algorithm called RandomForest Classifier. With an accuracy of 90% for CDK6, 80% for HER2 and ERα, the model classified the phytochemicals into 'active', 'inactive' and 'intermediate' class based on their IC50 values. Compounds with the standard value/IC50 greater than/equal to 10000 were classified as inactive and below or equal to 1000 were classified as active. Anything between was categorized as intermediate.

a)

```
y_test_pred
```

```
array(['active', 'active', 'inactive', 'active', 'active', 'inactive',
       'intermediate', 'inactive'], dtype=object)
```

b)

```
y_test_pred
```

```
array(['active', 'active', 'active', 'active', 'inactive'], dtype=object)
```

c)

```
[43] y_test_pred
```

```
array(['intermediate', 'inactive', 'intermediate', 'active',
       'intermediate', 'intermediate', 'intermediate', 'active',
       'intermediate'], dtype=object)
```

Figure 6.4 ML algorithm classifying phytochemicals based on their IC50 values for **a)** CDK6, **b)** HER2, **c)** ERα.

6.4 ADMET profiling

ML validated potent and active drug candidates targeting,

- CDK6 - Alpha-terpineol, Caryophyllene oxide, D-Borneol and Humulene epoxide II.
- HER2 – Allo-aromadendrene, Alpha-cadinol, Alpha-copaene, Dihydrocarvylacetate, Isomintlactone
- ERα – Carvone, Alpha-curcumene

The SMILE Sequences of these compounds were uploaded to ADMETlab 2.0 for pharmacokinetic and physicochemical analysis. And based on this, Lipinski;s rule of 5 was calculated to check drug like properties of these phytochemicals. The final results are summarised in table.

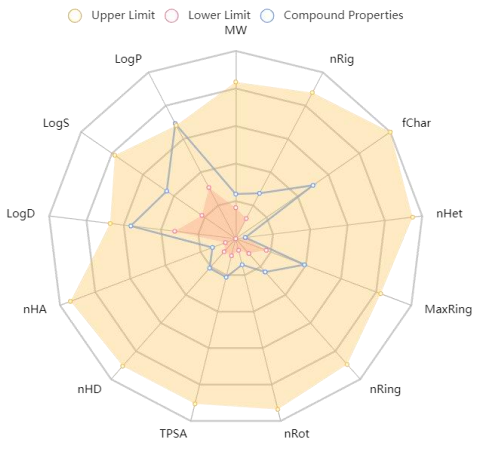
6.4.1 Physiochemical properties

At most violation of 2 rules is allowed to fall under Lipinski's rule. All of the phytochemicals followed the Lipinski's rule with no violations except Humelene epoxide II, Alpha-copaene and Alpha-curcuemene, as their estimated logP values an indicative of lipophilicity, were higher than the standard limit. Through the radial graphs obtained from ADMETlab 2.0, physiochemical properties could be analysed visually as well. If the properties of the compound were within the shaded region it indicated high oral bioavailability.

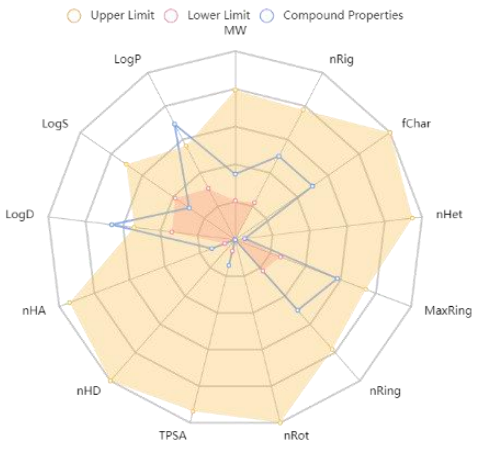
Active compounds	Lipinski's rule of 5					Bioavailability
	Molecular weight (g/mol)	H-bond donor	H- bond acceptor	Lipophilicity (logP)	No. of violations	
CDK6						
Alpha-terpineol	154.14	1	1	3.08	0	-0.51
Caryophyllene oxide	220.18	0	1	4.54	0	+0.74
d-Borneol	154.14	1	1	2.8	0	-0.54
Humulene epoxide II	220.18	0	1	5.23	1	+0.57

HER2						
Allo-aromadendrene	204.19	0	0	5.0	0	+0.57
Alpha-cadinol	222.2	1	1	4.52	0	+0.64
Alpha-copaene	204.19	0	0	5.43	1	+0.68
Dihydro-carvylacetate	196.15	0	2	3.22	0	+0.67
Isomintlactone	166.1	1	2	3.14	0	+0.57
ERα						
Carvone	150.1	0	1	2.13	0	+0.57
Alpha-curcumene	202.17	0	0	5.9	1	+0.64

Table 6.4 Physiochemical analysis of active phytochemicals through ADMETlab 2.0 and admetSAR 2.0.



a)



b)

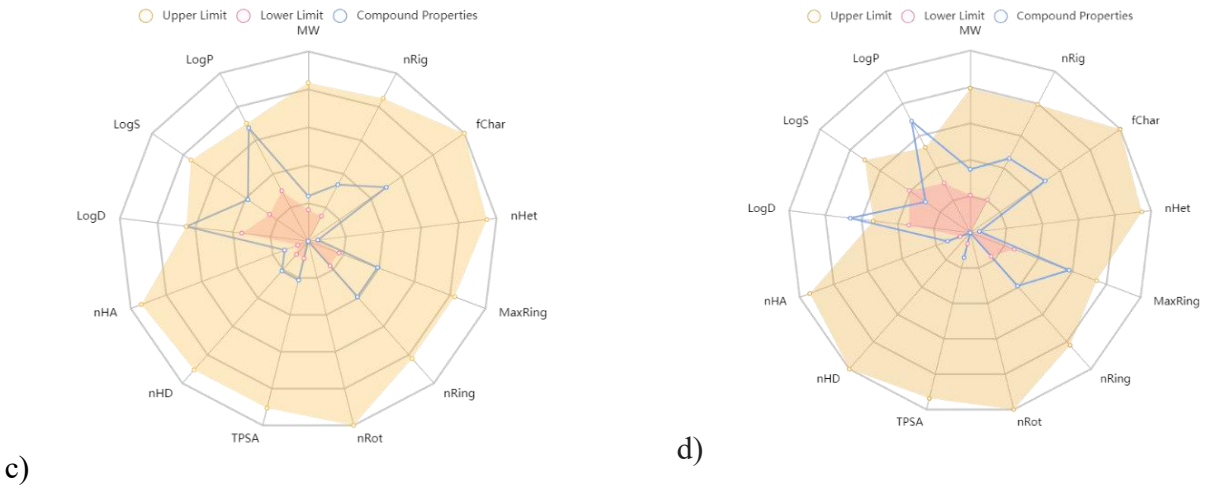
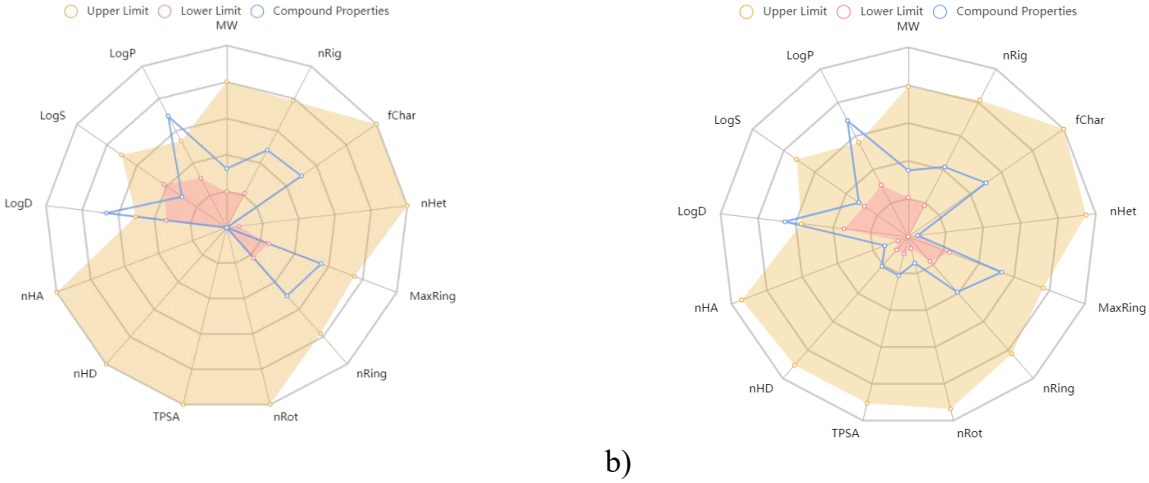
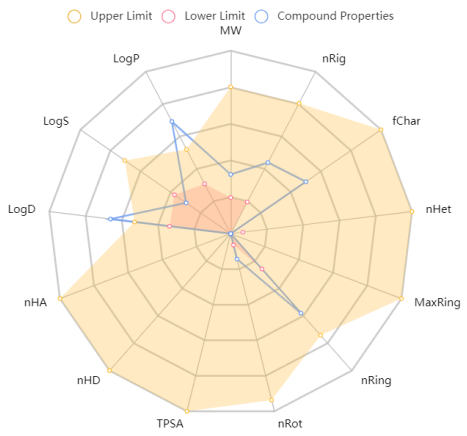
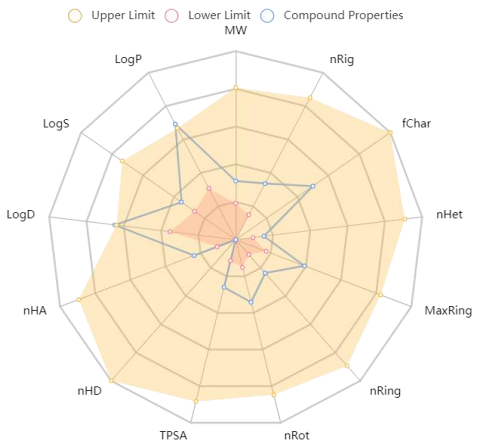


Figure 6.5 Bio availability radar of **a)** Alpha-terpineol **b)** Caryophylleneoxide **c)** d-Borneol **d)** Humulene epoxide II using swissADME predictor.

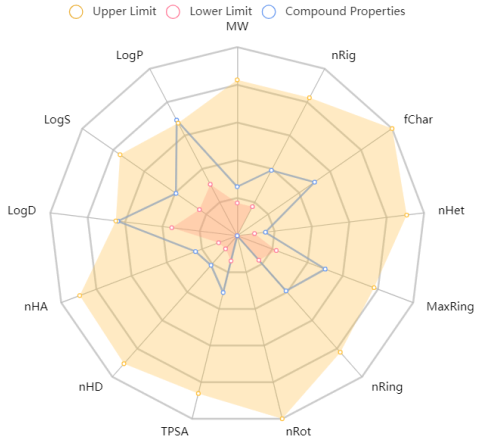




c)



d)



e)

Figure 6.6 Bio availability radar of **a)** Allo-aromadendrene **b)** Alpha-cadinol **c)** Alpha-copaene **d)** Dihydrocarvylacetate **e)** Isomintlactone using swissADME predictor.

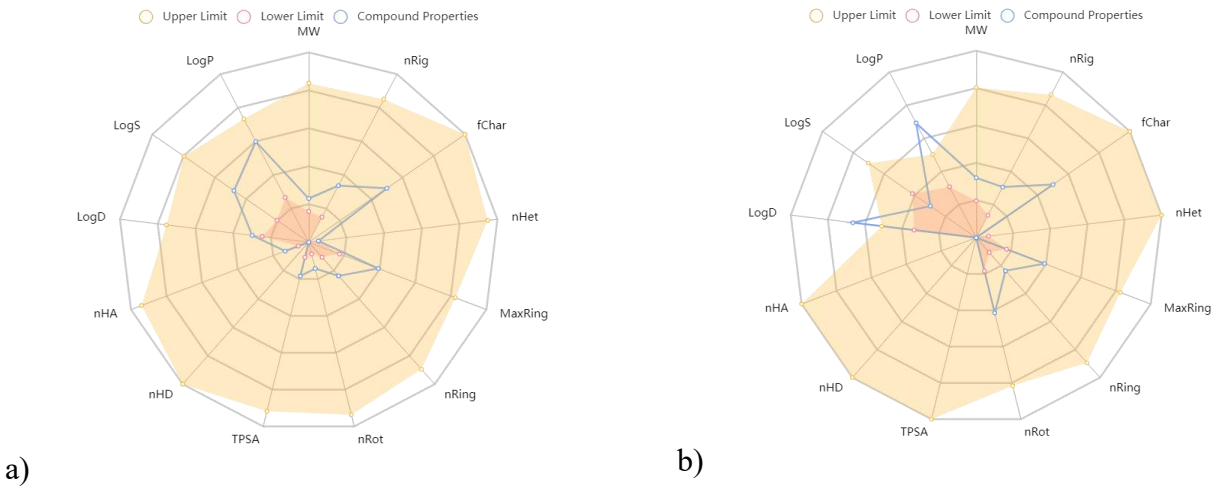


Figure 6.7 Bio availability radar of **a)** Carvone **b)** Alpha-curcuemene using swissADME predictor.

6.4.2 Pharmacokinetic properties

All 11 phytochemicals showed excellent probability scores for HIA, PPB and low probability scores for carcinogenecity. P450, a human cytochrome that has 5 isozymes which are responsible for metabolizing drugs, inhibition and substrate activity with all 11 phytochemicals was checked and all of them showed substrate activity, except Alpha-copaene, Isomintlactone and Alpha-curcumene that showed inhbhition activity with few isozymes. High CL (clearance rate of drug) values for Allo-aromadendrene, Alpha-cadinol, Alpha-copaene, Isomintlactone, Carvone and Alpha-curcumene was recorded. h-ERG (human ether-a-go-go related gene) blocking was observed for Alpha-cadinol and Alpha-curcumene, whereas rest of the phytochemicals showed negative probability scores for blocking of h-ERG gene.

Active compounds	**HIA**	**PPB**	**P450**	**CL (ml/min/Kg)**	**hERG**	**Carcinogenicity**
CDK6						
Alpha-terpineol	+0.99	+0.85	Substrate (2C19, 2C9)	8.9	-0.44	-0.9
Caryophyllene oxide	+0.99	+0.89	Substrate (1A2, 2C19, 2D6)	6.45	-0.46	-0.76

d-Borneol	+0.99	+0.64	Substrate (1A2, 2C19, 2C9, D6)	7.92	-0.6	-0.77
Humulene epoxide II	+1	+0.41	Substrate (2C19, 2C9, 2D6)	4.5	-0.56	-0.97
HER2						
Allo-aromadendrene	+0.99	+0.6	Substrate (1A2, 2C19, 2D6)	12.3	-0.54	-0.6
Alpha-cadinol	+0.99	+0.82	Substrate (2C19, 3A4)	13.4	+0.66	-0.59
Alpha-copaene	+0.99	+0.66	1A2 Inhibitor, 2C19 Substrate	20.2	-0.51	-0.78
Dihydro-carvylacetate	+0.98	+0.51	Substrate (2C19, 2C9, 2D6)	5.9	-0.45	-0.66
Isomintlactone	+1.0	+0.41	1A2 Inhibitor, 1A2, 2C19, 2C9, 2D6 Substrate	13.15	-0.56	-0.97
ERα						
Carvone	+0.99	+0.61	Substrate (1A2, 2C19, 2D6)	13.06	-0.54	-0.63
Alpha-curcumene	+0.99	+0.82	1A2, 2C19, 2C9, 3A4 Inhibitor, 1A2, 2C19, 2C9, 2D6, 3A4 Substrate	12.78	+0.66	-0.59

Table 6.5 Pharmacokinetic analysis of active phytochemicals through ADMETlab 2.0 and admetSAR 2.0. [HIA (Human intestinal absorption), PPB (Plasma protein binding), P450 isozymes CYP 1A2 / 2C19 / 2C9 / 2D6 / 3A4, h-ERG (human ether-a-go-go related gene), CL (clearance rate of drug)].

CHAPTER 7

DISCUSSION

For the present study, Water Mint / *Mentha aquatica* was selected as the source for phytochemicals and performed molecular docking of 84 phytochemicals to find out best fit compounds that can work as inhibitors of CDK6, HER2 and ER. The standard inhibitors were Fisetin, SYR127063 and lasofoxifene. The present study extracted 3D structures of target protein from RCSB PDB and prepared it using Biovia software. The phytochemicals were retrieved in 3D PDB format from IMPPAT. Docking was performed using Autodock Vina and the results were validated using Machine learning algorithm. Random Forest Classifier model was used and training/testing data was acquired from CHEMBL database using webresource_client library. Descriptor files were generated which are basically the end results of mathematical execution for conversion of chemical structure of a molecule into numerical values. PaDEL-Descriptor is a software which calculates the molecular descriptor of a compound to produce molecular fingerprint which helps in building QSAR (Quality structure activity relationship) models and can further be utilized for analysing biological activities of novel compounds. While working with scientific data, the PaDEL-Descriptor can be used to determine the molecular fingerprint of particular molecules that are then utilized to create machine learning models in the field of science. But PaDEL-Descriptor requires JAVA to execute calculations, so an alternative was to use an open-source package PaDELPy that calculates molecular fingerprints using Python [50].

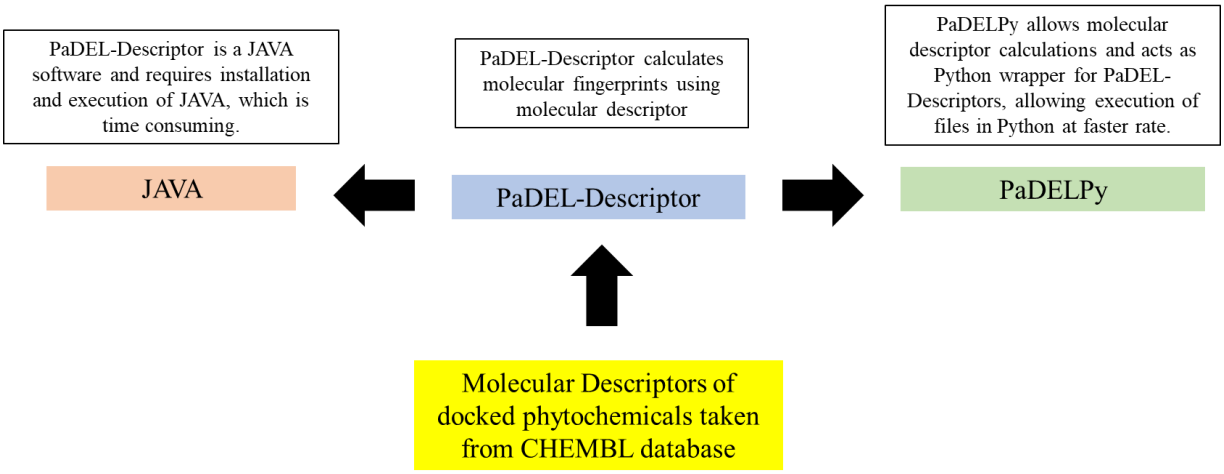


Figure 7.1 Flow chart for generating Descriptor files using PaDELPY

So, after PaDELPy generated the descriptor file for the phytochemicals, it was downloaded and uploaded in the Machine learning algorithm. With an accuracy of 90% for CDK6 and 80% for HER2 and ER, the model was capable of classifying the best docked compounds as active or inactive drugs on the basis of their IC50 values. An IC50 or Inhibitory Concentration 50 value is drug concentration required during in-vitro analysis to show 50% inhibition. Hence smaller IC50 values are indicative of minimum amount of drug required to produce an effective response and is preferable while designing a drug. After machine learning, a total of 11 active phytochemicals were acquired for the three target proteins that are Alpha-terpineol, Caryophyllene oxide, d-Borneol, and Humulene epoxide II for CDK6; Allo-aromadendrene, Alpha-cadinol, Alpha-copaene, Dihydro-carvylacetate and Isomintlactone for HER2; Carvone and Alpha-curcumene for ER. Their binding affinities as mentioned were similar to but better than the inhibitor except for the case of ERα. Lasofoxifene the standard inhibitor selected for ER α showed higher binding affinity and hence phytochemicals that showed binding affinity in ranges (a difference of 0.4-0.6) of lasofoxinene were selected for further analysis. Before concluding that the drug is fit for use, it is important to perform ADMET/ Lipinsk's rule of 5 tests. It is an important set of parameters to evaluate the drug-like properties of a compound and to check if the chosen compound can act as a suitable orally active drug for humans. The rule of five depends on following principles - Molecular mass must be less than 500 D, Lipophilicity should be high (indicated by logP<=5), hydrogen bond donors should be <5, Hydrogen bond acceptors should <10. During ADMET

analysis absorption is measured by HIA (Human intestinal absorption) which has been closely related to oral bio-availability. It is calculated by comparing the dose of drug orally administered and dose of drug that has reached the portal vein [51]. All 11 phytochemicals showed HIA probability scores in the range 0.98-1.0, hence indicating high absorption. PPB (Protein plasma binding) is a measure that helps in analysing distribution of a compound in body with the help of its binding affinities to serum proteins which help in its uptake and distribution. PPB values <90% are considered good, as high value maybe an indication of low therapeutic index. PPB values were obtained in the range of 0.41-0.89, with lowest being that of Humulene epoxide II and Isomintlactone and highest being of Caryophyllene oxide. Cytochrome P450 are mainly present in liver and are responsible for metabolizing various drugs. CYP450 has 5 isozymes CYP1A2, CYP2C19, CYP 2C9, CYP2D6, CYP3A4 and depending on whether the target compound acts as an inducer or inhibitor for each of these isozymes, toxicity and therapeutic efficacy can be estimated [51]. Alpha-copaene and Isomintlactone showed inhibition to CYP1A2 and Alpha-curcumene that showed inhbhition activity against CYP 1A2/2C19/2C9/3A4. CL is a measure of clearance of a drug from body and gives an idea about half-life of the drug thus help in in estimating an effective dosing drug concentration. Its unit is ml/min/Kg; >15 is considered as high clearance, 5-15 moderate and <5 as poor.

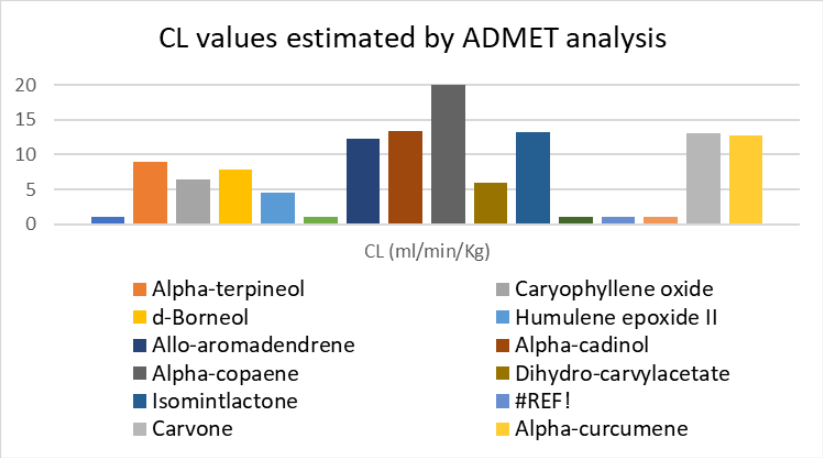


Figure 7.2 Comparison of CL values of all 11 phytochemicals.

Humulene epoxide II shows the poor clearance (4.5 ml/min/Kg) and Alpha-copaene shows the highest CL (20.2 ml/min/Kg), rest of the phytochemicals show values in the moderate range.

Toxicity profile is checked by estimating the inhibition activity of phytocompounds with h-ERG (human ether-a-go-go related gene). This gene synthesizes a K^+ channel that regulates cardiac potential during polarization and repolarization. Blocking of this may result in increased duration of repolarization which may cause long QT syndrome (LQTS), arrhythmia and severe heart failure [51]. All phytochemicals showed inhibition to h-ERG except Alpha-cadinol and Alpha-curcumene, which showed 66% probability of inhibiting h-ERG. Carcinogenicity of all the phytochemicals were obtained in negative probability scores indicating non-carcinogenecity. According to the ADMET analysis, Alpha-terpineol, d-borneol can be used for synthesizing potent CDK6 inhibitors; Alpha-aromadendrene, Alpha-copaene, Dihyrocarvylacetate and Isomintlactone can be used for synthesizing potent HER2 inhibitors; And Carvone was the only one phytochemical that showed best ADMET analysis and could be utilized in synthesizing ERα inhibitors.

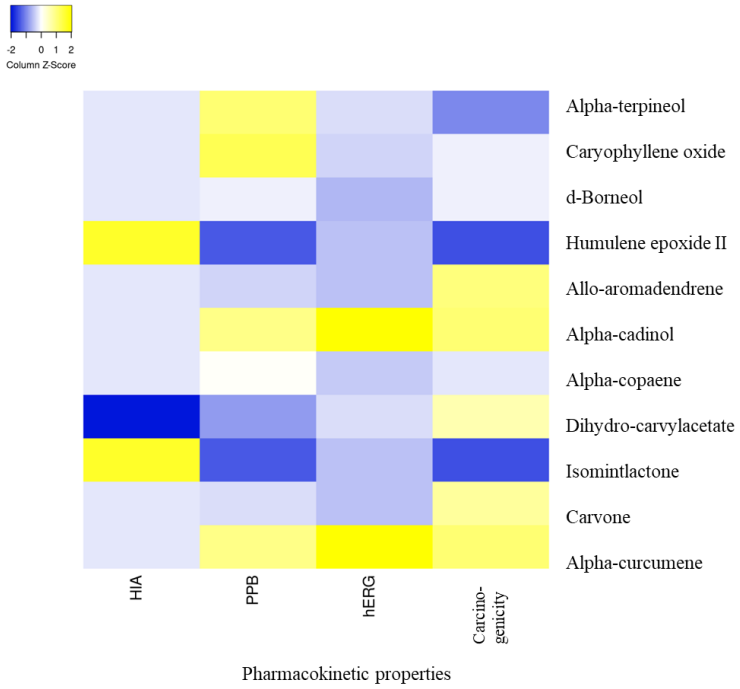


Figure 7.3 Heatmap analysis of the probability scores of HIA, PPB, h-ERG and carcinogenicity.

CHAPTER 8

CONCLUSION

Being the most commonly diagnosed cancer among women, breast cancer has shown rapid advancements in the field of therapeutics partially due to patients acquiring resistance to the ongoing treatment methods. Based on biomarkers profiling, breast cancer has several upregulated proteins such as CDK6, HER2 and ER. Among these HER2 and ER help in classifying breast cancer into its sub-types. Roles of cyclin D subunits that are CDK4/6 along with HER2 and ER are prominent in breast cancer. CDK4/6 regulate cell cycle progression but if present in abundance, can lead to abnormal growth of cells called cancer. HER2 activates cell proliferation via homo- or hetero-dimerization and ER in association with estrogen binds to promoter regions of genes and initiates transcription. 50% of the cases have shown overexpressed Cyclin-D levels, with enhanced HER2 and ER levels in 20% and 70% of the cases. Treatments for breast cancer nowadays involve targeting these upregulated proteins using drug inhibitors. Due to several side effects of these chemically synthesized drug inhibitors, phytochemicals of herbal plants are being studied and can present us with an alternative and better treatment plan. phytochemicals usually show low water solubility and are required in high doses to exert an effective inhibitory action on oncogenic signalling pathways, several phytochemical analogs with improved physiochemical and anti-cancer properties are being developed for improving its efficacy. Thus, it can be concluded by our results that phytochemicals Alpha-terpineol, d-borneol can be used for synthesizing potent CDK6 inhibitors; Alpha-aromadendrene, Alpha-copaene, Dihyrocarvylacetate and Isomintlactone can be used for synthesizing potent HER2 inhibitors; And Carvone was the only one phytochemical that showed best ADMET analysis and could be utilized in synthesizing ERα inhibitors. Incorporation of these active phytocompounds of *M. aquatica* into studies can help in exploring new insights in breast cancer therapy.

APPENDICES

Appendix 1 Comparison of data obtained through ADMET analysis of all phytochemicals.

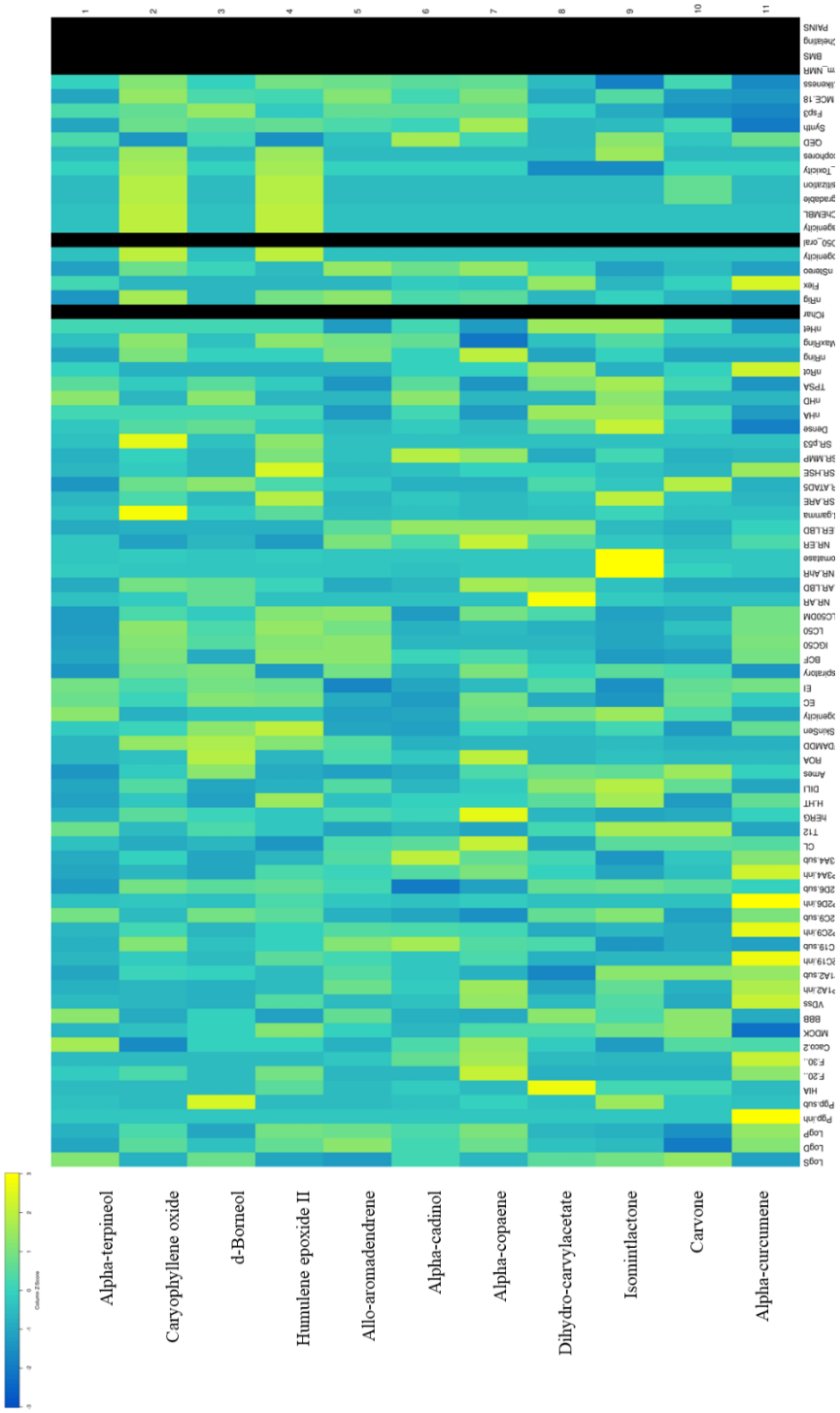


Figure A 1.1 Heatmap analysis of the probability scores of various parameters obtained through ADMET analysis from ADMETlab 2.0. [52].

REFERENCES

[1] F. A. Fisusi and E. O. Akala, "Drug Combinations in Breast Cancer Therapy," *Pharm. Nanotechnol.*, vol. 7, no. 1, pp. 3–23, 2019, doi: 10.2174/2211738507666190122111224.

[2] K. Pandey *et al.*, "Molecular mechanisms of resistance to CDK4/6 inhibitors in breast cancer: A review," *Int. J. Cancer*, vol. 145, no. 5, pp. 1179–1188, Sep. 2019, doi: 10.1002/ijc.32020.

[3] M. Yousuf *et al.*, "Ellagic Acid Controls Cell Proliferation and Induces Apoptosis in Breast Cancer Cells via Inhibition of Cyclin-Dependent Kinase 6," *Int. J. Mol. Sci.*, vol. 21, no. 10, p. 3526, May 2020, doi: 10.3390/ijms21103526.

[4] H. Prabhavathi, K. R. Dasegowda, K. H. Renukananda, P. Karunakar, K. Lingaraju, and H. Raja Naika, "Molecular docking and dynamic simulation to identify potential phytocompound inhibitors for EGFR and HER2 as anti-breast cancer agents," *J. Biomol. Struct. Dyn.*, vol. 40, no. 10, pp. 4713–4724, Jul. 2022, doi: 10.1080/07391102.2020.1861982.

[5] M. A. Moga *et al.*, "Pharmacological and Therapeutic Properties of Punica granatum Phytochemicals: Possible Roles in Breast Cancer," *Molecules*, vol. 26, no. 4, p. 1054, Feb. 2021, doi: 10.3390/molecules26041054.

[6] M. Arnold *et al.*, "Current and future burden of breast cancer: Global statistics for 2020 and 2040," *The Breast*, vol. 66, pp. 15–23, Dec. 2022, doi: 10.1016/j.breast.2022.08.010.

[7] I. Erić, "Breast Cancer in Young Women: Pathologic and Immunohistochemical Features," *Acta Clin. Croat.*, 2018, doi: 10.20471/acc.2018.57.03.13.

[8] M. Methamem, I. Ghadhab, S. Hidar, and R. Briki, "Breast cancer in men: a series of 45 cases and literature review," *Pan Afr. Med. J.*, vol. 36, 2020, doi: 10.11604/pamj.2020.36.183.22574.

[9] K. ROJAS and A. STUCKEY, "Breast Cancer Epidemiology and Risk Factors," *Clin. Obstet. Gynecol.*, vol. 59, no. 4, pp. 651–672, Dec. 2016, doi: 10.1097/GRF.0000000000000239.

[10] Y.-S. Sun *et al.*, "Risk Factors and Preventions of Breast Cancer," *Int. J. Biol. Sci.*, vol. 13, no. 11, pp. 1387–1397, 2017, doi: 10.7150/ijbs.21635.

[11] Z. He *et al.*, "A review on methods for diagnosis of breast cancer cells and tissues," *Cell Prolif.*, vol. 53, no. 7, 2020, doi: 10.1111/cpr.12822.

[12] T. Shien and H. Iwata, "Adjuvant and neoadjuvant therapy for breast cancer," *Jpn. J. Clin. Oncol.*, vol. 50, no. 3, pp. 225–229, Mar. 2020, doi: 10.1093/jjco/hyz213.

[13] A. Prat *et al.*, "Clinical implications of the intrinsic molecular subtypes of breast cancer," *Breast*, vol. 24, pp. S26–S35, 2015, doi: 10.1016/j.breast.2015.07.008.

[14] J. J. Gao and S. M. Swain, "Luminal A Breast Cancer and Molecular Assays: A Review," *Oncologist*, vol. 23, no. 5, pp. 556–565, May 2018, doi: 10.1634/theoncologist.2017-0535.

[15] F. Ades *et al.*, "Luminal B Breast Cancer: Molecular Characterization, Clinical Management, and Future Perspectives," *J. Clin. Oncol.*, vol. 32, no. 25, pp. 2794–2803, Sep. 2014, doi: 10.1200/JCO.2013.54.1870.

[16] D. F. Roses and A. E. Giuliano, "Surgery for Breast Cancer," *Breast Cancer*, pp. 401–459, 2005, doi: 10.1016/B978-0-443-06634-4.50027-3.

[17] R. E. Reinbolt *et al.*, "Endocrine therapy in breast cancer: The neoadjuvant, adjuvant, and metastatic approach," *Semin. Oncol. Nurs.*, vol. 31, no. 2, pp. 146–155, 2015, doi: 10.1016/j.soncn.2015.02.002.

[18] E. Provenzano, "Neoadjuvant Chemotherapy for Breast Cancer: Moving Beyond Pathological Complete Response in the Molecular Age," *Acta Med. Acad.*, vol. 50, no. 1, p. 88, May 2021, doi: 10.5644/ama2006-124.328.

[19] E. Senkus-Konefka, M. Wełnicka-Jaśkiewicz, J. Jaśkiewicz, and J. Jassem, "Radiotherapy for breast cancer in patients undergoing breast reconstruction or augmentation," *Cancer Treat. Rev.*, vol. 30, no. 8, pp. 671–682, 2004, doi: 10.1016/j.ctrv.2004.06.008.

[20] C. W. Taylor and A. M. Kirby, "Cardiac Side-effects From Breast Cancer Radiotherapy," *Clin. Oncol.*, vol. 27, no. 11, pp. 621–629, Nov. 2015, doi: 10.1016/j.clon.2015.06.007.

[21] W. Liu *et al.*, "CDK6 Is a Potential Prognostic Biomarker in Acute Myeloid Leukemia," *Front. Genet.*, vol. 11, 2021, doi: 10.3389/fgene.2020.600227.

[22] N. C. Turner *et al.*, “Palbociclib in Hormone-Receptor–Positive Advanced Breast Cancer,” *N. Engl. J. Med.*, vol. 373, no. 3, pp. 209–219, Jul. 2015, doi: 10.1056/NEJMoa1505270.

[23] M. Piezzo *et al.*, “Targeting Cell Cycle in Breast Cancer: CDK4/6 Inhibitors,” *Int. J. Mol. Sci.*, vol. 21, no. 18, p. 6479, Sep. 2020, doi: 10.3390/ijms21186479.

[24] K. Pandey *et al.*, “Molecular mechanisms of resistance to CDK4/6 inhibitors in breast cancer: A review,” *Int. J. Cancer*, vol. 145, no. 5, pp. 1179–1188, 2019, doi: 10.1002/ijc.32020.

[25] S. Jia, X. Xu, S. Zhou, Y. Chen, G. Ding, and L. Cao, “Fisetin induces autophagy in pancreatic cancer cells via endoplasmic reticulum stress- and mitochondrial stress-dependent pathways,” *Cell Death Dis.*, vol. 10, no. 2, 2019, doi: 10.1038/s41419-019-1366-y.

[26] L. N. Klapper *et al.*, “The ErbB-2/HER2 oncoprotein of human carcinomas may function solely as a shared coreceptor for multiple stroma-derived growth factors,” *Proc. Natl. Acad. Sci.*, vol. 96, no. 9, pp. 4995–5000, Apr. 1999, doi: 10.1073/pnas.96.9.4995.

[27] S. Kunte, J. Abraham, and A. J. Montero, “Novel HER2–targeted therapies for HER2–positive metastatic breast cancer,” *Cancer*, vol. 126, no. 19, pp. 4278–4288, Oct. 2020, doi: 10.1002/cncr.33102.

[28] C. Gutierrez and R. Schiff, “HER2: biology, detection, and clinical implications.,” *Arch. Pathol. Lab. Med.*, vol. 135, no. 1, pp. 55–62, Jan. 2011, doi: 10.5858/2010-0454-RAR.1.

[29] G. Valabrega, F. Montemurro, and M. Aglietta, “Trastuzumab: mechanism of action, resistance and future perspectives in HER2-overexpressing breast cancer,” *Ann. Oncol.*, vol. 18, no. 6, pp. 977–984, Jun. 2007, doi: 10.1093/annonc/mdl475.

[30] A. Derakhshani *et al.*, “Overcoming trastuzumab resistance in HER2-positive breast cancer using combination therapy,” *J. Cell. Physiol.*, vol. 235, no. 4, pp. 3142–3156, Apr. 2020, doi: 10.1002/jcp.29216.

[31] D.-Y. Oh and Y.-J. Bang, “HER2-targeted therapies — a role beyond breast cancer,” *Nat. Rev. Clin. Oncol.*, vol. 17, no. 1, pp. 33–48, Jan. 2020, doi: 10.1038/s41571-019-0268-3.

[32] K. Aertgeerts *et al.*, “Structural Analysis of the Mechanism of Inhibition and Allosteric

Activation of the Kinase Domain of HER2 Protein," *J. Biol. Chem.*, vol. 286, no. 21, pp. 18756–18765, May 2011, doi: 10.1074/jbc.M110.206193.

[33] S. Saha Roy and R. K. Vadlamudi, "Role of Estrogen Receptor Signaling in Breast Cancer Metastasis," *Int. J. Breast Cancer*, vol. 2012, pp. 1–8, 2012, doi: 10.1155/2012/654698.

[34] Shagufta and I. Ahmad, "Tamoxifen a pioneering drug: An update on the therapeutic potential of tamoxifen derivatives," *Eur. J. Med. Chem.*, vol. 143, pp. 515–531, 2018, doi: 10.1016/j.ejmech.2017.11.056.

[35] M. Lainé *et al.*, "Lasofoxifene as a potential treatment for therapy-resistant ER-positive metastatic breast cancer," *Breast Cancer Res.*, vol. 23, no. 1, p. 54, Dec. 2021, doi: 10.1186/s13058-021-01431-w.

[36] P. R. Dandawate, D. Subramaniam, R. A. Jensen, and S. Anant, "Targeting cancer stem cells and signaling pathways by phytochemicals: Novel approach for breast cancer therapy," *Semin. Cancer Biol.*, vol. 40–41, pp. 192–208, Oct. 2016, doi: 10.1016/j.semcancer.2016.09.001.

[37] E. R. M. Zunica, C. L. Axelrod, and J. P. Kirwan, "Phytochemical Targeting of Mitochondria for Breast Cancer Chemoprevention, Therapy, and Sensitization," *Int. J. Mol. Sci.*, vol. 23, no. 22, p. 14152, Nov. 2022, doi: 10.3390/ijms232214152.

[38] M. Z. Aumeeruddy and M. F. Mahomoodally, "Combating breast cancer using combination therapy with 3 phytochemicals: Piperine, sulforaphane, and thymoquinone," *Cancer*, vol. 125, no. 10, pp. 1600–1611, May 2019, doi: 10.1002/cncr.32022.

[39] Z. Raptopoulos and C. Constantinou, "The Effect of Exercise on the Alleviation of Side Effects Induced by Aromatase Inhibitors in Postmenopausal Breast Cancer Patients," *Curr. Oncol. Rep.*, vol. 22, no. 11, p. 110, Nov. 2020, doi: 10.1007/s11912-020-00971-2.

[40] L. Fleming, S. Agnew, N. Peddie, M. Crawford, D. Dixon, and I. MacPherson, "The impact of medication side effects on adherence and persistence to hormone therapy in breast cancer survivors: A quantitative systematic review," *The Breast*, vol. 64, pp. 63–84, Aug. 2022, doi: 10.1016/j.breast.2022.04.010.

[41] L. B. de Souza *et al.*, "Mentha aquatica L. aerial parts: in vitro anti-proliferative evaluation on human tumour and non-tumour cell lines," *Nat. Prod. Res.*, vol. 36, no. 12,

pp. 3117–3123, Jun. 2022, doi: 10.1080/14786419.2021.1937152.

[42] F. Conforti, G. Ioele, G. A. Statti, M. Marrelli, G. Ragno, and F. Menichini, "Antiproliferative activity against human tumor cell lines and toxicity test on Mediterranean dietary plants," *Food Chem. Toxicol.*, vol. 46, no. 10, pp. 3325–3332, Oct. 2008, doi: 10.1016/j.fct.2008.08.004.

[43] C. T. Chang, W. N. Soo, Y. H. Chen, and L. F. Shyur, "Essential Oil of Mentha aquatica var. Kenting water mint suppresses two-stage skin carcinogenesis accelerated by BRAF inhibitor vemurafenib," *Molecules*, vol. 24, no. 12, 2019, doi: 10.3390/molecules24122344.

[44] K. Mohanraj *et al.*, "IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics," *Sci. Rep.*, vol. 8, no. 1, p. 4329, Mar. 2018, doi: 10.1038/s41598-018-22631-z.

[45] R. P. Vivek-Ananth, K. Mohanraj, A. K. Sahoo, and A. Samal, "IMPPAT 2.0: An Enhanced and Expanded Phytochemical Atlas of Indian Medicinal Plants," *ACS Omega*, vol. 8, no. 9, pp. 8827–8845, Mar. 2023, doi: 10.1021/acsomega.3c00156.

[46] A. J. (2010). Trott, O., & Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry, 31(2), 455-461."

[47] M. F. Adasme *et al.*, "PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA," *Nucleic Acids Res.*, vol. 49, no. W1, pp. W530–W534, Jul. 2021, doi: 10.1093/nar/gkab294.

[48] G. Xiong *et al.*, "ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties," *Nucleic Acids Res.*, vol. 49, no. W1, pp. W5–W14, Jul. 2021, doi: 10.1093/nar/gkab255.

[49] H. Yang *et al.*, "admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties," *Bioinformatics*, vol. 35, no. 6, pp. 1067–1069, Mar. 2019, doi: 10.1093/bioinformatics/bty707.

[50] C. W. Yap, "PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints," *J. Comput. Chem.*, vol. 32, no. 7, pp. 1466–1474, May 2011,

doi: 10.1002/jcc.21707.

[51] N. A. Durán-Iturbide, B. I. Díaz-Eufracio, and J. L. Medina-Franco, “In Silico ADME/Tox Profiling of Natural Products: A Focus on BIOFACQUIM,” *ACS Omega*, vol. 5, no. 26, pp. 16076–16084, 2020, doi: 10.1021/acsomega.0c01581.

[52] S. Babicki *et al.*, “Heatmapper: web-enabled heat mapping for all,” *Nucleic Acids Res.*, vol. 44, no. W1, pp. W147–W153, Jul. 2016, doi: 10.1093/nar/gkw419.

8th International Conference for Convergence in Technology (I2CT)

IEEE BOMBAY SECTION

7th – 9th April 2023

Certificate

This is to certify that Dr./Prof./Mr./Ms. Ruchi Tirkey has presented paper entitled In-silico approach to DNA codon optimization of Zaire Ebola virus genes for vaccine development in 8th International Conference for Convergence in Technology (I2CT) during 7th & 9th April 2023.

Dr. Chanakya Kumar Jha
General Chair - I2CT

Tirkey, Ruchi; Joshi, Soniya; Hasija, Yasha In-silico approach to DNA codon optimization of Zaire Ebola virus genes

6% Overall Similarity

Top sources found in the following databases:

- 4% Internet database
- Crossref database
- 5% Submitted Works database
- 3% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	researchsquare.com Internet	<1%
2	researchgate.net Internet	<1%
3	mdpi.com Internet	<1%
4	academic.oup.com Internet	<1%
5	Multimedia University on 2021-06-08 Submitted works	<1%
6	University College London on 2011-03-30 Submitted works	<1%
7	University of Nottingham on 2022-12-14 Submitted works	<1%
8	University of East London on 2020-03-19 Submitted works	<1%

9	krex.k-state.edu Internet	<1%
10	biomedcentral.com Internet	<1%
11	frontiersin.org Internet	<1%
12	University of Kent at Canterbury on 2019-04-03 Submitted works	<1%
13	University of York on 2022-08-15 Submitted works	<1%
14	archive.interconf.center Internet	<1%
15	ncbi.nlm.nih.gov Internet	<1%
16	Foote, Kevin M., Kevin Blades, Anna Cronin, Shaun Fillery, Sylvie S. Gui... Crossref	<1%
17	Liverpool John Moores University on 2022-11-23 Submitted works	<1%
18	docplayer.net Internet	<1%
19	hbsra.org Internet	<1%
20	jfd.i3l.ac.id Internet	<1%

21	rcastoragev2.blob.core.windows.net Internet	<1%
22	87040 on 2014-11-20 Submitted works	<1%
23	Imperial College of Science, Technology and Medicine on 2019-08-28 Submitted works	<1%
24	Kathleen Aertgeerts, Robert Skene, Jason Yano, Bi-Ching Sang et al. "S... Crossref	<1%
25	University College London on 2020-05-04 Submitted works	<1%
26	University of Sheffield on 2019-03-18 Submitted works	<1%
27	Vels University on 2017-11-24 Submitted works	<1%
28	Wei Lv, Jinzhong Liu, Todd C. Skaar, Elizaveta O'Neill, Ge Yu, David A. F... Crossref	<1%
29	journals.sagepub.com Internet	<1%
30	mdpi-res.com Internet	<1%
31	ris.utwente.nl Internet	<1%
32	medrxiv.org Internet	<1%

33	sabcs.org Internet	<1%
34	science.gov Internet	<1%
35	Bryant College on 2022-12-06 Submitted works	<1%
36	Rajiv Gandhi Centre for Biotechnology on 2022-06-16 Submitted works	<1%
37	Paula Cabello Navarro. "Epigenetic regulation of resistance to treatme... Crossref posted content	<1%
38	University of Southampton on 2020-02-12 Submitted works	<1%

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi -110042

CANDIDATE'S DECLARATION

I, Ruchi Tirkey, Roll No. 2K21/MSCBIO/36 of MSc. Biotechnology, hereby declare that the project Dissertation titled "***MOLECULAR DOCKING AND MACHINE LEARNING APPROACH TO ELUCIDATE Mentha aquatica PHYTOCHEMICALS ROLE IN BREAST CANCER TREATMENT***" which is submitted by me to the Department of Biotechnology, Delhi Technological University, New Delhi in partial fulfillment of the requirement for the award of the degree of Master of Science, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any degree, Diploma Associateship, Fellowship or other similar title or recognition.

Title of the paper: In-silico approach to DNA codon optimization of Zaire Ebola virus genes for vaccine development.
Author's Names: Tirkey, Ruchi; Joshi, Soniya; Hasija, Yasha
Name of Conference: IEEE 2023 8th International Conference for Convergence in Technology (I2CT)
Conference Date and Venue: 7-9th April at Hotel Vivanta Pune (TAJ), Hinjawadi, Pune, Maharashtra, India
Registration: Done
Status of Paper Publication: Waiting
Date of Paper Communication: 28th February 2023
Date of Paper Acceptance: 12th March 2023

PLACE: Delhi

Date: 30-05-23

Ruchi

RUCHI TIRKEY

2K21/MSCBIO/36

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of
Engineering)Bawana Road, Delhi-
110042

CERTIFICATE

I hereby certify that the Dissertation titled "***MOLECULAR DOCKING AND MACHINE LEARNING APPROACH TO ELUCIDATE Mentha aquatica PHYTOCHEMICALS ROLE IN BREAST CANCER TREATMENT***" which is submitted by Ruchi Tirkey, 2K21/MSCBIO/36, Department of Biotechnology, Delhi Technological University, New Delhi is in partial fulfillment of the requirement for the award of the degree of Master of Science, is a record of the work carried out by the student under my supervision. To the best of my knowledge, this work has not been submitted in part or full for any Degree or Diploma to thisUniversity or elsewhere.

Place: Delhi

Date: 30-05-23

30.05.23

PROF. YASHA HASIJA
(SUPERVISOR)

Professor
Department of Biotechnology
Delhi technological university

30/05/2023

PROF. PRAVIR KUMAR
HEAD OF DEPARTMENT (HoD)

Department of Biotechnology
Delhi Technological University