Original Research Article

Integrative Computational Identification of Non-Toxic Flavonoid Derivatives as MMP-1 Inhibitors in Breast Cancer

Abstract

Aims: This study aims to identify non-toxic, flavonoid-derived phytochemicals as potential inhibitors of Matrix Metalloproteinase-1 (MMP-1), a key enzyme involved in breast cancer invasion and metastasis.

Study Design: An integrative in silico approach was adopted, incorporating structure-based drug design (SBDD), network pharmacology, ADMET profiling, and toxicity prediction to discover phytochemical-based MMP-1 inhibitors.

Methodology: A total of 140 phytochemicals were selected based on their reported anti-cancer and anti-metastatic properties from Dr. Duke's and PubChem databases. Biological activity prediction was performed using PASS software, while Molinspiration and Lipinski's Rule of Five were used to evaluate drug-likeness. Toxicity prediction was carried out using DeepToxLab. Gene-disease association and protein interaction networks were constructed via DisGeNET, STRING, and Cytoscape. Molecular docking studies were performed using CB-Dock2, and ligand-receptor interactions were visualized in Discovery Studio.

Results: PASS analysis identified several flavonoids with high potential for anti-cancer and anti-MMP activity (Pa > 0.5). Docking studies revealed that Isohamnetin showed the strongest binding affinity to MMP-1 with a Vina score of -9.0 kcal/mol, followed by Kaempferol and Diosmetin (both -8.5 kcal/mol), and Hispidulin (-8.4 kcal/mol). Geneistein showed the weakest binding at -7.8 kcal/mol. Visualization confirmed that all compounds bound within the active site of MMP-1, forming stable interactions. These lead compounds also exhibited favorable ADMET profiles and low carcinogenicity scores (<0.2), with high prediction confidence from DeepToxLab.

Conclusion: The findings suggest that certain flavonoid compounds—especially Isohamnetin, Kaempferol, and Diosmetin—have strong potential as non-toxic MMP-1 inhibitors. This study provides a computational basis for further preclinical development of these phytochemicals as novel agents against metastatic breast cancer.

Keywords*: Flavonoids, Matrix Metalloproteinase-1 (MMP-1), Breast Cancer Metastasis, In Silico Drug Design, Network Pharmacology, Phytochemicals, Molecular Docking, ADMET Profiling*

I INTRODUCTION

Breast cancer is a prevalent malignant tumor globally, with 2.3 million new cases in 2022. It ranks fourth in mortality rate and is a heterogeneous disease with various molecular subtypes. HER2-positive and triple-negative breast cancer (TNBC) are highly prone to metastasis and high recurrence rates. Treatment strategies have evolved from single surgical resection to comprehensive therapies including chemotherapy, radiotherapy, targeted therapy, and immunotherapy. Despite advancements, recurrence and metastasis remain the primary causes of treatment failure (Jiang et al., 2025).The World Health Organization (WHO) has performed a study based on the GLOBOCAN database to examine the worldwide burden of female breast cancer. According to the study, in 2022, 2.3 million new cases and 670,000 deaths occurred globally. Annual rates were elevated by 1–5% in half of the countries reviewed, with 29 high Human Development Index (HDI) countries seeing mortality rates fall. Seven nations, such as Denmark and Belgium, are already fulfilling the Global Breast Cancer Initiative target of a 2.5% annual decline. In 2050, new instances and fatalities will grow by 38% and 68%, respectively (Kim *et al.,* 2025). Breast cancer is determined by environmental, lifestyle, and genetic factors worldwide. An increase in population and adoption of a western lifestyle are contributing to rising cases of breast cancer in developing nations. Death rates will decrease with enhanced access to prevention services, early diagnosis, and medical care (Xiong *et al.,* 2025).

Cancer is the fifth most common cause of death in India, and it accounts for 5.7% of total deaths. The age-specific incidence rate of breast cancer differs, with 32.8% in the age group 25-49 years, 27.9% in the age group 50-69 years, and 23.4% in the age group 70 years and above. In 2020, breast cancer was found in 25.8 per 100,000 women and had a mortality rate of 12.7 per 100,000 women. The southern and northern parts of the country had the highest burden at 685.5 and 677.6 DALYs per 100,000 women, respectively (Pillai *et al.,* 2025).

Metalloproteinases (MPs) play important roles in cancer progression, hemostasis, and development. MP inhibitors have limited therapeutic action owing to their multifunctionality in various cell types and normal conditions. Epigenetic regulators of MP expression have emerged as indirect targets, and epigenetics-targeting drugs are more developed than MP inhibitors (Seehawer & Polyak, 2025). Matrix metalloproteinases (MMPs) are a group of zinc-binding endopeptidases that play an important role in breast cancer development by remodeling the extracellular matrix (ECM) and basement membrane, thereby promoting tumor invasion, angiogenesis, and metastasis. Various MMPs, such as MMP-1, MMP-2, MMP-7, MMP-9, and MMP-14, have been found to be involved in breast cancer, with each of them playing different but overlapping roles. Among these, MMP-1, also referred to as interstitial collagenase, holds special significance. MMP-1 directly degrades interstitial collagens—types I, II, and III—important structural elements of connective tissue. Through the degradation of these collagens, MMP-1 facilitates cancer cell invasion through tissue barriers and into adjacent stroma, thus enhancing metastatic dissemination to distant organs like the lungs and bones. Besides its proteolytic activity, MMP-1 is also able to bind to the protease-activated receptor-1 (PAR-1), activating downstream signaling pathways leading to enhanced tumor cell migration, proliferation, angiogenesis, and survival. Clinical reports have shown that increased MMP-1 expression is associated with more aggressive tumor phenotypes, higher metastatic capabilities, and worse patient prognosis in breast cancer (Duffy *et al.,* 2007) (Overall & Kleifeld, 2006) (Gong *et al.,* 2014). Recent studies also emphasize MMP-1 as a good therapeutic target, with approaches aimed at blocking its activity demonstrating effectiveness in preclinical models of breast cancer (Winer *et al.,* 2018).

This research will identify new and non-toxic phytochemical derivatives, more specifically flavonoid-derived compounds, as possible chemotherapeutic agents against Matrix Metalloproteinase-1 (MMP-1) in breast cancer. MMP-1 is important in tumor invasion and metastasis by breaking down interstitial collagens and hence an appealing molecular target

for anticancer therapy. The research utilizes structure-based drug design (SBDD) methods to screen thousands of ligands and estimate their binding affinity against chosen disease targets. Structure-based virtual screening is well known as an effective tool to identify potential hits efficiently while removing non-complementary compounds from large chemical libraries. Here, the emphasis is placed on the design and screening of new flavonoid derivatives, the identification of high-binding-score ligands, and their ADMET profiles to ascertain drug-likeness and safety. Besides virtual screening, a network pharmacology approach is incorporated to elucidate the multi-target interaction of the hit compounds, charting their activity across major signaling pathways and biological processes implicated in breast cancer development. This dual approach maximizes the discovery of candidate drugs by factoring both molecular binding affinity and systemic therapeutic effect. Overall, the study presents new findings and a starting point for the development of new anticancer agents with improved efficacy and fewer side effects.

2 Methodology

2.1 Collection of phytochemicals used in the study

Phytochemicals for the study were chosen from Dr.Dukes Botanical database (https://phytochem.nal.usda.gov/phytochem/search), review and research articles using the key words anti-breast cancer property, anti-metastatic property and anti-matrix metalloproteinase property. The structure of these compounds were downloaded from National Institute of Health's PubChem online database that are freely available for virtual screening (Hanessian et al., 2001). Totally 140 compounds including terpenes, flavonoids, flavones, flavanones, chalcones, isoflavones alkaloids, aldehydes and anthocyanidins were identified and further screened using the PASS online software.

2.2 Prediction of Ligand Activity

The PASS (Prediction of Activity Spectra for Substances), an online software predicts the result based on structure, which is equal to its activity and is used to compare the function of unknown biologically active compounds (Filimonov and Poroikov 2008). It estimates the probability of a particular substance's characteristic to the active and in-active sub-sets of a substance from the Structure Activity Relationship Base (SAR Base). Using the Bayesian approach implemented in computer program PASS compared the function of selected structures (Geronikaki et al., 2002). SMILES (Simplified Molecular-Input Line-Entry System) format for the phytochemical's structure was given as input. Four activities – anti-neoplastic, anti-breast cancer, anti-metastatic and anti-MMP expression were considered. For each ligand, Pa (Probable activity) and Pi (Probable inactivity) values for all the four activities were obtained. Only the activities with Pa>Pi and Pa>0.5 were considered as experimentally possible for a particular compound (Poroikov et al., 2002).

2.3 DEG Analysis and Construction of the PPI Network

DisGeNET is a discovery platform integrating information on gene-disease associations (GDAs) from several public data sources and the literature (Piñero et al., 2016). Online tool STRING (http://string-db.org) was used to identify the potential interaction networks of protein products of these genes (Szklarczyk et al., 2016).Further the protein-protein interactions (PPI) network was constructed and visualized by Cytoscape version 3.0.1 (Institute of Systems Biology, Seattle, WA, USA) (Majeed & Mukhtar, 2023). Nodes were screened out in the PPI network with degree ≥ 1, and 'degree' represented the connections with other nodes. According to the top pathways with their corresponding targets, the diseases, and the co-expressed protein network was constructed using Cytoscape version 3.8.0 (Guo et al., 2017).

2.4 ADME Prediction

Molinspiration offers a variety of software tools for drug design, such as molecule manipulation, normalization, tautomer generation, fragmentation of molecules, and molecular modeling. The tools facilitate substructure and similarity searching, fragment-based virtual screening, prediction of bioactivity, and data visualization. Molinspiration's Java-based tools are available on any computer platform, and this improves the comprehension of molecular properties (*Calculation of Molecular Properties and Bioactivity Score*, n.d.). Lipinski's rule of five requires that a drug orally active should not contain more than five hydrogen bond acceptors and donors, molecular weight less than 500 g/mol, partition coefficient log P five, and fewer than four violations. The two drug-like molecules were analyzed by Molinspiration a Physicochemical Properties Calculator, which determines partition coefficient, molecular weight, number of heavy atoms, hydrogen donors, acceptors, and violations (Zashumo *et al.,* 2022).

2.5 Toxicity Prediction

The DeepToxLab platform utilizes a deep learning-based predictive model to assess the toxicity potential of small molecules, particularly focusing on carcinogenicity. It integrates chemical structure input, typically provided in SMILES format, and converts it into graph-based and physicochemical descriptors that feed into pre-trained deep neural networks. These networks have been trained on extensive public datasets such as Tox21 and PubChem BioAssay, enabling robust quantitative structure-activity relationship (QSAR) modeling. The model assigns a toxicity probability score between 0 and 1, where lower values indicate reduced toxic potential. A key feature of DeepToxLab is its inclusion of a confidence label for each prediction—classified as either high or low. High-confidence predictions, denoted by a blue "H" icon, signify strong model reliability and are particularly valuable during early-stage drug development. In this study, only compounds with both low carcinogenicity scores (below 0.2) and high-confidence predictions were considered for further analysis, aligning with the platform’s guidelines for reliable toxicity interpretation (DeepToxLab, 2025).

2.6 Docking and visualization

CB-Dock2 is a blind docking server that makes predictions of ligand binding sites and conducts molecular docking without knowing the binding pocket in advance. It uploads three-dimensional structures of the receptor protein and the ligand in PDB format, detects possible binding cavities, ranks the top five, and establishes suitable docking boxes. Molecular docking is conducted by an integrated AutoDock Vina engine, producing binding poses and Vina scores. The optimal receptor-ligand complexes are additionally examined using Discovery Studio Visualizer, which investigates important molecular interactions such as hydrogen bonds, hydrophobic contacts, pi-pi stacking, and electrostatic interactions. Visualization of 2D interaction plots and 3D complex structure is possible through the software, giving further insight into the binding between protein active site residues and the ligand.

3 Result

3.1 Selection of phytochemicals

For the present study, 140 phytochemical compounds were taken. The phytocompounds were chosen based on intricate search in research and review articles which satisfied any one of the following conditions: (1) anti-breast cancer (2) anti-metastatic properties. Since, choosing all phytocompounds from all plants with the specified properties was not practically possible, the number of phytocompounds taken for the study was restricted to 150. Consequently after removal of duplicate entries (owing to more than one common name for a single flavonoid), the final number of phytcompounds was 140. Phytochemicals list with the above said properties was obtained from Dr.Dukes Botanical database (https://phytochem.nal.usda.gov/phytochem/search) (Hanessian et al., 2001) and grouped into flavonoids and non-flavonoids.

The two dimensional unique structures of 140 phytocompound ligands were obtained from National Institute of Health's PubChem online database (http://pubchem.ncbi.nlm.nih.gov/). The list of phytocompounds with their PubChem ID used in the study is provided in Appendix 1 (Table A2.1)

3.2 Classification of chosen Phytochemicals

Based upon the structure, phytochemicals chosen for the study were classified precisely. Table 3.1 shows the distribution of the 140 established phytochemicals with respect to the number of constituent phytocompounds in the chosen data set. The 140 ligands were classified into several classes (flavonoids and non-flavonoids) of phytochemicals as shown in Figure 3.1.

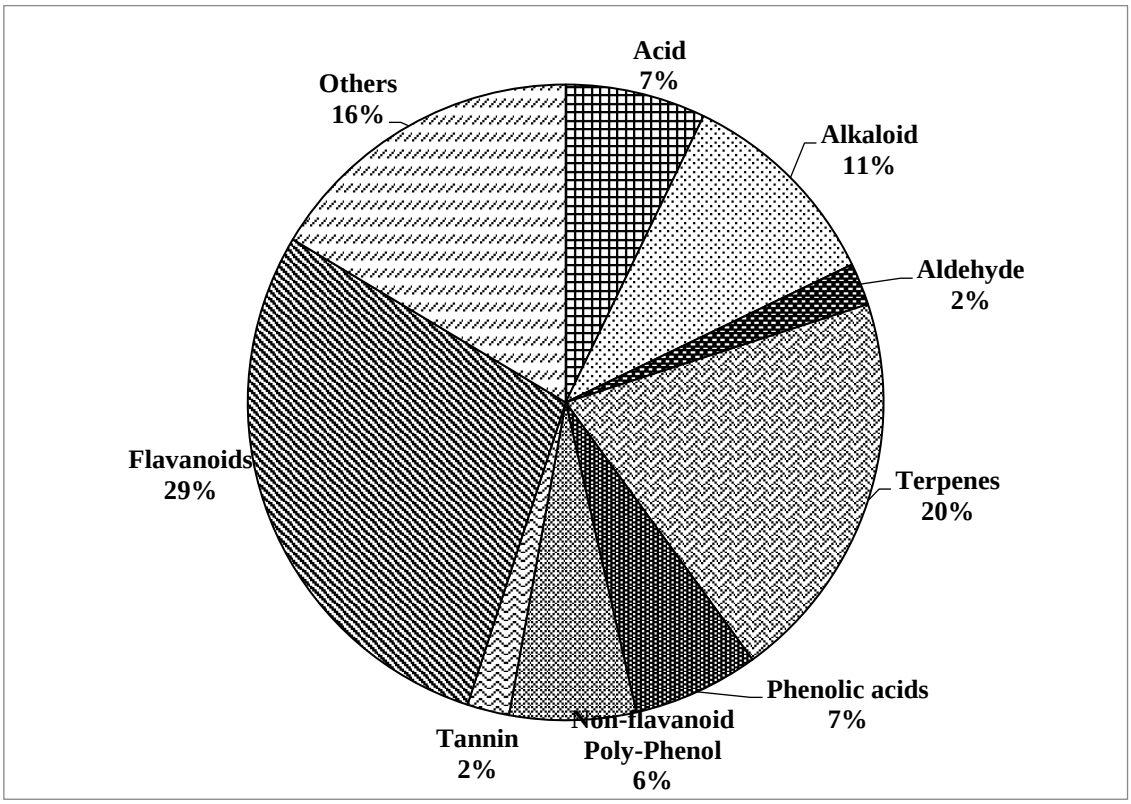

Figure 3.1 Pie Chart showing percentage distribution of phytochemicals used in this study

From the results, major percentage of phytochemicals was represented by (i) non-flavonoids with 71% distribution and (ii) flavonoids, with 29%. In the non-flavonoids category, terpenes were in higher distribution with 20% representation. All the acids including, benzoic acid, cinnamic acid, carboxylic acid, caffeic acid were grouped into one category as acids and their distribution is 7%., also phenolic acids share 7% distribution and poly-phenols other than flavonoids have 6% distribution. Alkaloids distribution was 11%. Other groups such as tannin, aldehyde, polysaccharides, phytosteroid, organosulfur compounds demonstrated very low distribution.

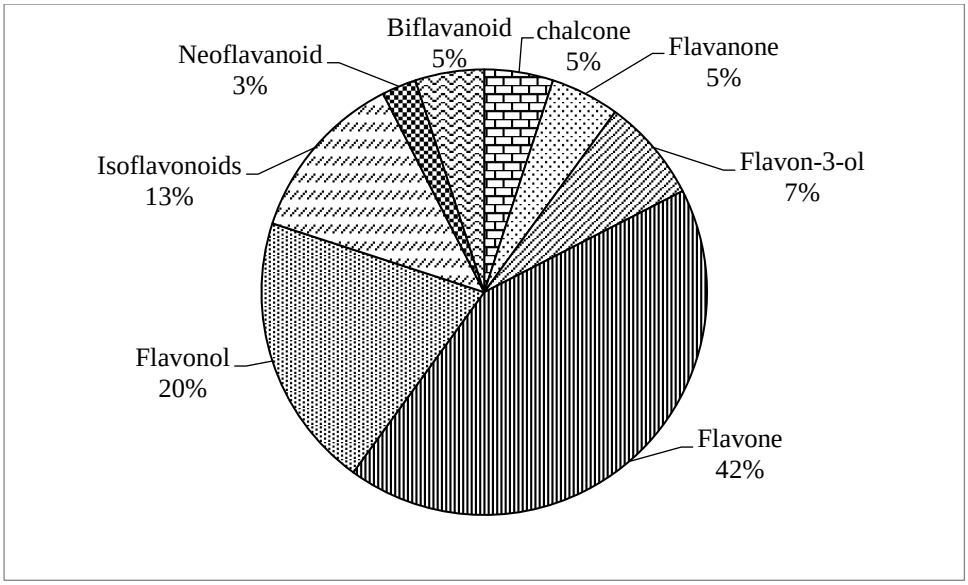

Figure 3.2 Pie Chart showing percentage distribution of flavonoids used in this study

The flavonoids (as represented in Figure 4.1) were further sub-divided into flavones (42%), flavonols (20%), flavan-ols (7%), flavanones (5%), isoflavonoids (13%), neoflavonoids (3%), bi-flavonoids (5%) and chalcones (5%).

Earlier studies on MMP-1 inhibition by phytochemicals are only very few in number. Those studies suggested that, the inhibition of MMP-1 is frequently restricted to only a few compounds, either plant specific or compound specific ligands (Maity et al., 2011) (Lim & Kim, 2007). Hence, in this study, we have attempted to investigate the inhibition activity of MMP-1 through more than hundred phytocompounds based upon their classification. More assortment groups used here are considered important, as it will increase the probability of eventually identifying the best compound that has more possibility to inhibit MMP-1.

3.3. Prediction of Ligand Activity

All the 140 phytocompounds were tested for their anti-cancer related activity in the PASS tool. Based upon the combined properties such as MMP inhibition, anti-metastatic, anti-neoplastic and anti-breast cancer property, the compounds predicted with more than one property with Pa>Pi were only considered. The filtered 56 compounds (out of 140) based on Pa and Pi values are given in Table 3.1, and they have been grouped into flavonoids (41 compounds) and non-flavonoids (25 compounds).

Table 3.1 : Flavonoids short listed from PASS Online Tool

Sl.No	Phytocompounds	Class	Pa	Pi	Activity
1.	Apigenin	Flavone	0.780	0.004	MMP expression inhibitor
			0.775	0.015	Anti-neoplastic
			0.669	0.007	Anti-neoplastic (breast cancer)
2.	Baicalein	Flavone	0.789	0.013	Anti-neoplastic
			0.588	0.011	Anti-neoplastic (breast cancer)
			0.681	0.008	MMP expression inhibitor

3.	Chrysin	Flavone	0.654	0.007	Anti-neoplastic (breast cancer)
			0.766	0.004	MMP expression inhibitor
4.	Diosmetin	Flavone	0.682	0.006	Anti-neoplastic (breast cancer)
			0.814	0.003	MMP expression inhibitor
5.	Eupatilin	Flavone	0.698	0.005	Anti-neoplastic (breast cancer)
			0.724	0.005	MMP expression inhibitor
6.	Genkvannin	Flavone	0.798	0.003	MMP expression inhibitor
			0.634	0.008	Anti-neoplastic (breast cancer)
7.	Hispidulin	Flavone	0.752	0.004	MMP expression inhibitor
			0.694	0.005	Anti-neoplastic (breast cancer)
8.	Isothymonin	Flavone	0.622	0.009	Anti-neoplastic (breast cancer)
			0.673	0.009	MMP expression inhibitor
9.	Luteolin	Flavone	0.672	0.007	Anti-neoplastic (breast cancer)
			0.777	0.004	MMP expression inhibitor
10.	Nobiletin	Flavone	0.579	0.018	MMP expression inhibitor
			0.698	0.005	Anti-neoplastic (breast cancer)
11.	Rhamnazin	Flavone	0.767	0.004	MMP expression inhibitor
			0.586	0.012	Anti-neoplastic (breast cancer)
12.	Rhoifolin	Flavone	0.53	0.014	Anti-metastatic
			0.504	0.019	Anti-neoplastic (breast cancer)
13.	Tangeretin	Flavone	0.604	0.015	MMP expression inhibitor
			0.679	0.006	Anti-neoplastic (breast cancer)
14.	Fisetin	Flavonol	0.514	0.018	Anti-neoplastic (breast cancer)
			0.638	0.012	MMP expression inhibitor
15.	Galangin	Flavonol	0.722	0.005	MMP expression inhibitor
			0.550	0.014	Anti-neoplastic (breast cancer)
16.	Isohamnetin	Flavonol	0.777	0.004	MMP expression inhibitor
			0.599	0.010	Anti-neoplastic (breast cancer)
17.	Kaempferol	Flavonol	0.738	0.005	MMP expression inhibitor
			0.565	0.013	Anti-neoplastic (breast cancer)
18.	Myricetin	Flavonol	0.578	0.012	Anti-neoplastic (breast cancer)
			0.842	0.008	Anti-neoplastic
			0.696	0.007	MMP expression inhibitor

19.	Quercetin	Flavonol	0.576	0.012	Anti-neoplastic (breast cancer)
			0.734	0.005	MMP expression inhibitor
20.	Rutin	Flavonol	0.536	0.016	Anti-neoplastic (breast cancer)
			0.522	0.015	Anti-metastatic
21.	Diadzein	Isoflavonoid	0.864	0.002	MMP expression inhibitor
			0.506	0.018	Anti-neoplastic (breast cancer)
22.	Formononetin	Isoflavonoid	0.887	0.002	MMP expression inhibitor
			0.526	0.016	Anti-neoplastic (breast cancer)
23.	Geneistein	Isoflavonoid	0.571	0.013	Anti-neoplastic (breast cancer)
			0.908	0.001	MMP expression inhibitor
24.	Glycitein	Isoflavonoid	0.564	0.013	Anti-neoplastic (breast cancer)
			0.888	0.002	MMP expression inhibitor
25.	Epigallocatechin	Flavan-3-ol	0.722	0.005	MMP expression inhibitor
			0.500	0.020	Anti-neoplastic (breast cancer)
26.	Epigallocatechin-3-Gallate	Flavan-3-ol	0.562	0.020	MMP expression inhibitor
			0.540	0.015	Anti-neoplastic (breast cancer)
27.	Hesperitin	Flavanone	0.656	0.007	Anti-neoplastic (breast cancer)
			0.801	0.003	MMP expression inhibitor
28.	Naringenin	Flavanone	0.639	0.008	Anti-neoplastic (breast cancer)
			0.764	0.004	MMP expression inhibitor
29.	Dalbergin	Neoflavonoid	0.599	0.010	Anti-neoplastic (breast cancer)
			0.745	0.004	MMP expression inhibitor
30.	Ginkgetin	Biflavonoid	0.666	0.007	Anti-neoplastic (breast cancer)
			0.750	0.004	MMP expression inhibitor
31.	Butein	Chalcone	0.591	0.011	Anti-neoplastic (breast cancer)
			0.770	0.004	MMP expression inhibitor

From 100 non-flavonoids initially chosen for the study, 25 compounds have been filtered through PASS online tool and the list is given in Table 3.1 (b).

3.4 Network Pharmacology

Network pharmacology approach was used to narrow down the MMP inhibitor as a useful target for arresting metastasis stage of breast cancer because, prediction of potential proteins is of more importance towards the analysis of therapeutic identification in biological systems (Vyas et al., 2023). The constructed network system can predict the main active

protein and their corresponding proteins by ‘network targets’. In the current study, Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database and cytoscape (version 10.0) was applied to investigate target expression data in the framework of biological network information. Cytoscape has successfully given the valuable clue that exposes the interactions of MMP-1 in metastasis of breast cancer.

3.5 Predicting protein targets

The functional interactions between MMP-1 and other key proteins can provide framework in molecular mechanism of cellular processing of metastasis. When the relationships between protein interactions are significantly high, then the protein/gene is declared as differentially expressed genes (DEGs). These encoding genes have direct interactions of their respective proteins and will finally affect the expression level of protein. The significant related proteins to breast cancer were extracted from String data base as per the protocol of (Cai et al., 2017). In our study, total of 438 DEGs that correlated to breast cancer metastasis in association with MMP-1 were screened by protein query of STRING database. The DEGs were screened by weighted correlation analysis with $|\log FC| > 0.58$ and $P < 0.05$. The top 20 DEGs and their functional enrichment pathway are shown in Table 3.2. As can be seen, IL-6 was the upregulated DEG with the highest fold change of 0.999 in the current study. Besides, the DEG of MMP family MMP-2 and MMP-9 were also upregulated with the combined score of 0.993 and 0.997 respectively.

Table 3.2 : Most significant DEGs in association with breast cancer

Symbol	geneid	Symbol	Combined Score	Pathway
MMP1	3569	IL6	0.999	Cellular responses to stress; Immune System; Signal Transduction
MMP1	4318	MMP9	0.997	Developmental Biology; Extracellular matrix organization; Signal Transduction
MMP1	7422	VEGFA	0.995	Cellular responses to stress; Developmental Biology; Hemostasis; Signal Transduction
MMP1	7040	TGFB1	0.994	Developmental Biology; Extracellular matrix organization; Hemostasis; Signal Transduction
MMP1	7157	TP53	0.994	Cell Cycle; Cellular responses to stress; DNA Repair; Gene Expression; Hemostasis; Programmed Cell Death; Signal Transduction
MMP1	5743	PTGS2	0.993	Metabolism
MMP1	4313	MMP2	0.993	Developmental Biology; Extracellular matrix organization; Metabolism of proteins
MMP1	3553	IL1B	0.992	Immune System
MMP1	3576	CXCL8	0.991	Cellular responses to stress; Metabolism of proteins; Signal Transduction
MMP1	4314	MMP3	0.991	Extracellular matrix organization; Signal Transduction
MMP1	596	BCL2	0.99	Immune System; Programmed Cell Death
MMP1	6347	CCL2	0.99	Metabolism of proteins; Signal Transduction
MMP1	1029	CDKN2A	0.99	Cell Cycle; Cellular responses to stress; Metabolism of proteins
MMP1	3458	IFNG	0.989	Immune System

MMP1	4524	MTHFR	0.989	Metabolism
MMP1	3552	IL1A	0.989	Cellular responses to stress; Immune System
MMP1	1636	ACE	0.987	Metabolism of proteins
MMP1	1956	EGFR	0.986	Developmental Biology; Immune System; Signal Transduction
MMP1	7099	TLR4	0.985	Immune System; Programmed Cell Death

Table 3.2 also shows the co-expression of MMP-1 with BCL-2 and VEGF-A proteins. Based on the STRING database, protein network was constructed.

Analysis of these data revealed that, out of 438 proteins that interacted with MMP-1 during metastasis of breast cancer, 161 proteins were upregulated.

3.6 PPI Network Construction

A network is made up of nodes (the points of communication or redistribution of proteins/genes) and edges (the lines of communication or relation joining the proteins/genes) (Raman, 2010). The interaction network of 161 MMP-1 related upregulated proteins were analyzed using the STRING database. Most of the upregulated proteins derived using this tool are over-expressed in breast cancer cells and involved in survival of those cells. The co-expression network of MMP-1 regulated genes, was established by setting the Pearson correlation efficient at >0.7. There were 483 edges in PPI network for 161 DEGs (Figure 3.3). Degrees >10 was set as the cutoff criterion and the average node degree was found to be 14.2. Among the 161 nodes, BCL2, p53, EGFR, VEGF-A, IL-6, IL-10, TNF, MMP-2 and MMP-9 were identified as hub genes with higher degrees. Moreover, it clearly shows that MMP-1 had interactions with proteins that are significantly correlated with metastasis of BC. Interestingly, our study is in correlation with previous study, which showed a relationship between MMP-1 and aggressiveness of breast cancer cells (Chen et al., 2015). It also depicts the ability of MMP-1 to instruct immune cells to modify the expression of other proteins, such as MMP-2 and MMP-9.

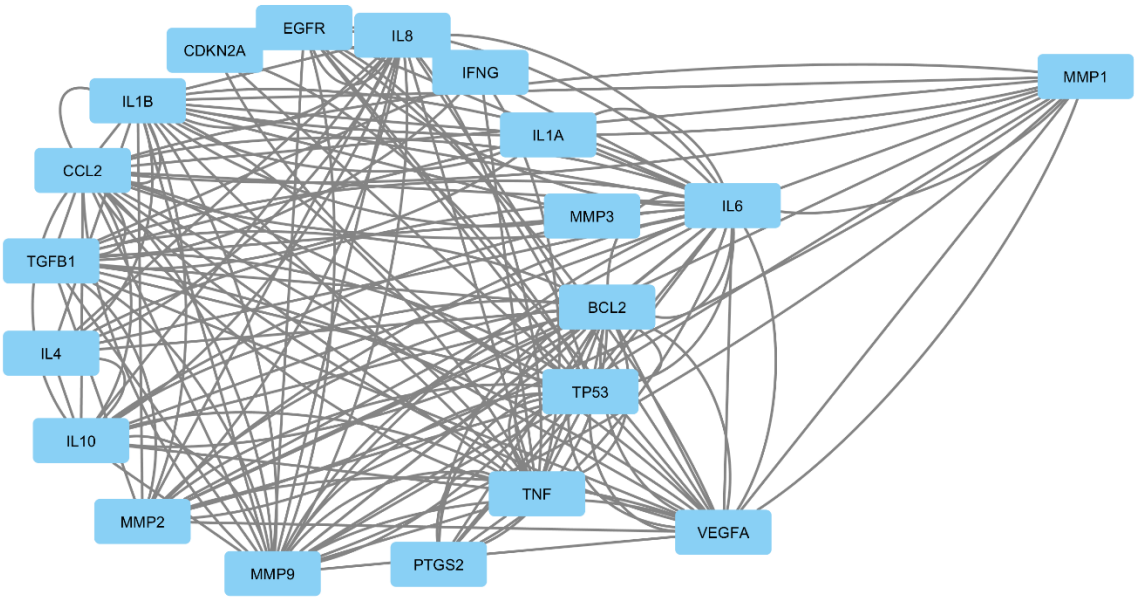

Figure 3.3 Cluster-1 Protein-protein interaction network of MMP-1 and its co-related proteins in association with breast cancer

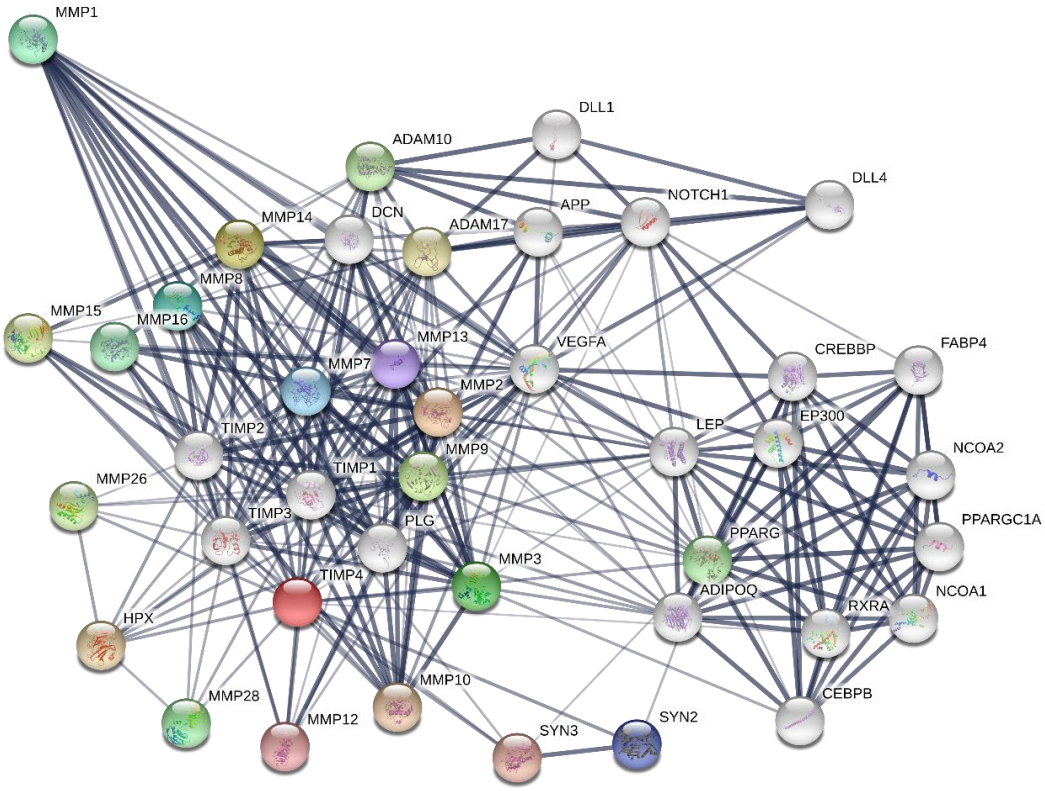

Figure 3.4 STRING cytoscape (version 10.0) analysis showed the protein–protein interaction (PPI) network of 161 proteins

In the sub-network of MMP-1 that were involved in metastasis process, BCL-2 (B Cell Lymphoma-2), TN53 (Tumour protein), TNF (Tumour necrosis factor), MMP-2 (Matrix metalloproteinase -2), MMP-9 (Matrix metalloproteinase-9), IL-1 (Interleukin -1), IL-6 (Interleukin -6), VEGF-A (Vascular endothelial growth factor- A) and EGFR (Epidermal growth factor) genes were identified to exhibit a higher degree of interaction in breast cancer (Figure 3.4).

Proteins that co-expressed with MMP-1 are important in breast cancer survival. BCL-2 inhibits apoptosis in different cell systems, promoting survival. MMP-1 interacts with MMP-2 and MMP-9, which are important regulators of degrading extracellular matrix proteins and tissue repair. TGFB1 signaling in breast cancer is associated with seeding lung metastasis by interfering with cell-cell interaction. Cytokines such as IL-1A, IL-1B, IL-4, IL-6, IL-8, and IL-10 are involved in inflammation, initiation of tumors, promotion of tumors, angiogenesis, and metastasis. IL-1 is implicated in poor prognosis of breast tumors and causes metastasis. IL-6 and IL-10 enhance tumor invasion and angiogenesis. CCL2 initiates a chemokine cascade and enhances metastatic sites of breast cancer cells. VEGF-A is postulated to be the regulator of signaling tumor cells, controlling cancer cells. These proteins are more connected than 0.9wddx with MMP-1 protein.

3.7 ADME Prediction

In the present investigation, a set of 24 phytochemicals, mostly flavonoids and structurally related polyphenols, were screened for their drug-likeness-relevant physicochemical attributes. Some of the crucial parameters like molecular weight (Mol Wt), lipophilicity (Log P), aqueous solubility (Log S), polar surface area (MolPSA), molecular volume (Mol Vol), drug-likeness score, hydrogen bond acceptors (HBA), and donors (HBD) were considered.

With the exception of baicalein (446.08 Da) and epigallocatechin-3-gallate (458.08 Da), most ligands had molecular weights in the optimal range (<500 Da) and were according to Lipinski's Rule of Five (Lipinski et al., 2001). All compounds expressed moderate to very good drug-likeness values (>0.1), except for naringenin (1.13), which had the highest value along with quercetin (0.93) and was indicated to have considerable potential for the development of a drug.

Lipophilicity, as indicated by Log P values, varied from 0.13 (baicalein) to 3.45 (formononetin), implying differential membrane permeability for the ligands. Substances such as formononetin, apigenin, and chrysin, with Log P values >3.0, may exhibit greater cellular uptake but at the potential cost of poor solubility or off-target activity.

Water solubility (Log S) throughout the ligands was generally moderate (-3 to -4 range), with camptothecin as an outlier at a Log S of -4.84, indicating its lower aqueous solubility. In drug development, such low solubility molecules usually need special delivery systems.

MolPSA values also validated the permeability profiles, with all but two of the compounds being within optimal PSA ranges (<140 Å²). Epigallocatechin-3-gallate (158.72 Å²) and baicalein (143.74 Å²) were higher than this, indicating possible problems in passive diffusion, especially through the blood-brain barrier.

Hydrogen bond donors and acceptors (HBD/HBA) were in agreement with reported structural characteristics of flavonoids. Ligands like epigallocatechin-3-gallate and baicalein contained a high donor and acceptor count, which can improve targeted interactions with biological targets but at the same time compromise membrane permeability if in excess.

Amongst the candidates, naringenin, quercetin, hesperetin, and epigallocatechin were identified as lead scaffolds for drug development, with a balance of drug-likeness, solubility, and permeability attributes. Highly hydrophilic or large molecules such as epigallocatechin-3-gallate, on the other hand, would need formulation approaches to enhance bioavailability.

Therefore, the physicochemical profiling indicates that most of the flavonoids studied have good drug-like characteristics, warranting their further assessment in in-silico docking

Table 3.3: Filtered ligands using Molinspiration Software

Sl.No	Ligands	Mol Wt	Log P	Log S	MolPSA	Mol Vol	Drug likeness	HBA	HBD
1.	Naringenin	272.07	2.30	-3.69	71.98	251.12	1.13	5	3
2.	Quercetin	302.04	2.11	-3.87	102.61	281.71	0.93	7	5
3.	Epigallocatechin	290.08	1.88	-3.24	90.45	261.13	0.92	6	5
4.	Hesperitin	302.08	2.27	-3.81	78.55	283.84	0.88	6	3
5.	Luteolin	286.05	2.68	-4.07	89.05	272.86	0.86	6	4
6.	Butein	272.07	2.60	-3.67	80.48	265.25	0.82	5	4
7.	Kaempferol	286.05	2.49	-3.96	87.13	268.99	0.77	6	4
8.	Apigenin	270.05	3.06	-4.16	73.57	260.14	0.77	5	3
9.	Fisetin	286.05	2.49	-3.93	86.06	271.53	0.76	6	4
10.	Baicalein	446.08	0.13	-3.15	143.74	397.35	0.73	11	6
11.	Hispidulin	300.06	2.91	-3.47	79.15	291.88	0.72	6	3

12.	Geneistein	270.05	2.72	-3.89	72.64	260.69	0.71	5	3
13.	Formononetin	268.07	3.45	-4.36	46.02	271.80	0.70	4	1
14.	Eupatilin	344.09	3.23	-3.79	76.79	344.60	0.69	7	2
15.	Isohamnetin	316.06	2.46	-3.98	93.69	301.71	0.67	7	4
16.	Rhamnazin	330.07	2.81	-4.20	83.62	323.01	0.65	7	3
17.	Diadzein	254.06	3.10	-3.95	56.09	250.50	0.56	4	2
18.	Glycitein	284.07	3.07	-4.01	62.66	283.22	0.47	5	2
19.	Diosmetin	300.06	3.03	-4.28	80.13	292.86	0.42	6	3
20.	Epigallocatechin-3-gallate	458.08	2.58	-3.76	158.72	404.49	0.39	11	8
21.	Isothymonin	360.08	2.73	-3.00	92.04	354.36	0.33	8	3
22.	Chrysin	254.06	3.32	-4.30	55.95	249.59	0.18	4	2
23.	Galangin	270.05	2.75	-4.10	69.51	258.44	0.13	5	3
24.	Camptothecin	348.11	1.58	-4.84	62.92	387.50	0.11	5	1

3.8 Toxicity Prediction

Table 3.4 ; Non-Carcinogenic Compounds Identified by DeepToxLab with High-Confidence Predictions

Compounds	Carcinogenicity	Carcinogenicity _uncertainty	developmental_toxicity	developmental_toxicity_ uncertainty
Kaempferol	0.160004	High-confidence	0.050185	High-confidence
Hispidulin	0.063692	High-confidence	0.790232	Low-confidence
Geneistein	0.127951	High-confidence	0.279911	Low-confidence
Isohamnetin	0.101123	High-confidence	0.112059	Low-confidence
Rhamnazin	0.14471	High-confidence	0.192321	Low-confidence
Diosmetin	0.161894	High-confidence	0.60431	Low-confidence
Isothymonin	0.038897	High-confidence	0.094264	High-confidence

Among seven compounds identified as non-carcinogenic based on an initial screen of 24 compounds on the DeepToxLab platform, backed by low scores for carcinogenicity (all ≤0.2) and high-confidence predictive output, Kaempferol, Hispidulin, Genistein, Isohamnetin, Rhamnazin, Diosmetin, and Isothymonin were all recognized as potential candidates. The predictive model singled out Isothymonin and Hispidulin as particularly compelling with scores of 0.039 and 0.064, respectively. All seven compounds are flavonoid derivatives, a structurally diverse family of compounds well characterized for their anti-inflammatory, antioxidant, and chemoprotective activities. Their safe toxicity profiles, as reliably predicted in silico, concur with past literature reports of the low carcinogenic risk posed by naturally occurring flavonoids. DeepToxLab's high-confidence predictions provide a good quality early-screening option, enabling prioritization of lead molecules toward development. But as is the case with any computational screen, these results deserve experimental confirmation using in-vitro and in-vivo toxicity studies prior to considering them viable drug or therapeutic agents.

3.9 Docking

Table 3.5 ; Docking scores of flavonoids with MMP-1

COMPOUNDS	DOCKING SCORE
Kaempferol	-8.5
Hispidulin	-8.4
Geneistein	-7.8
Isohamnetin	-9.0
Rhamnazin	-8.0
Diosmetin	-8.5
Isothymonin	-8.1

Molecular docking experiments were performed to assess the binding efficiencies of several flavonoid molecules with the active site of matrix metalloproteinase-1 (MMP-1). Docking scores are presented in Table 1.

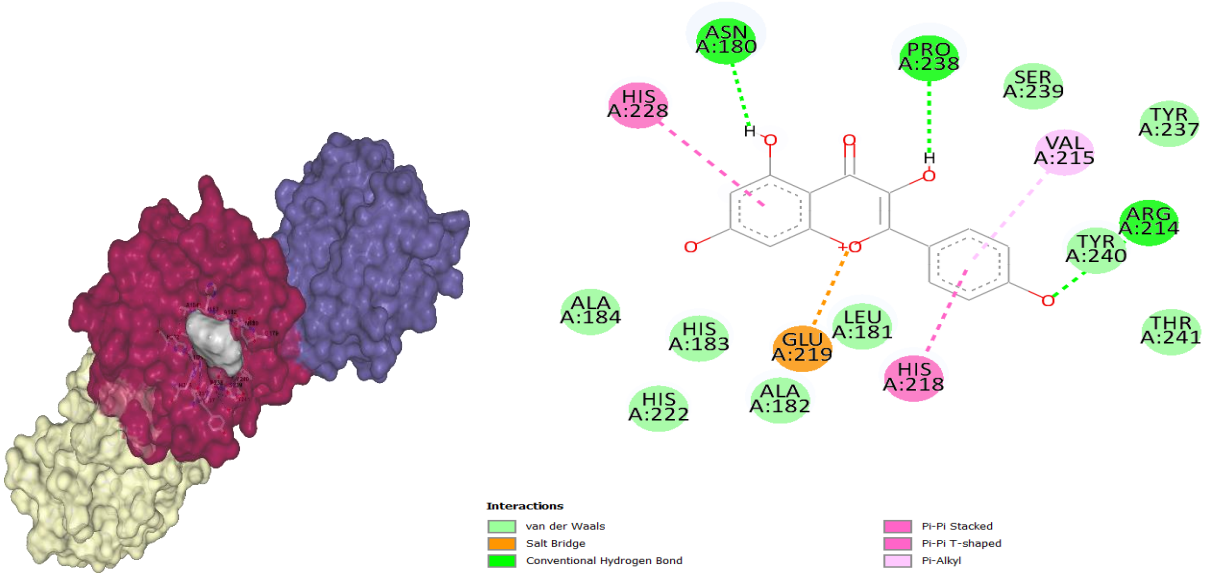

Figure 3.5 ; Kaempferol binds with the active site of MMP-1

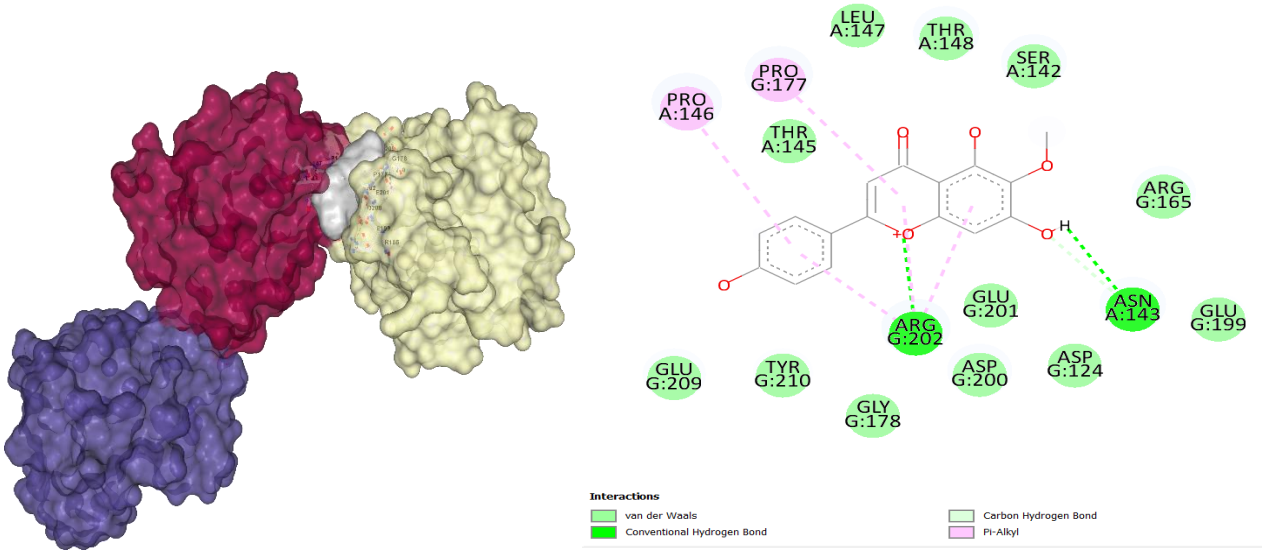

Figure 3.6 ; Hispidulin binds with the active site of MMP-1

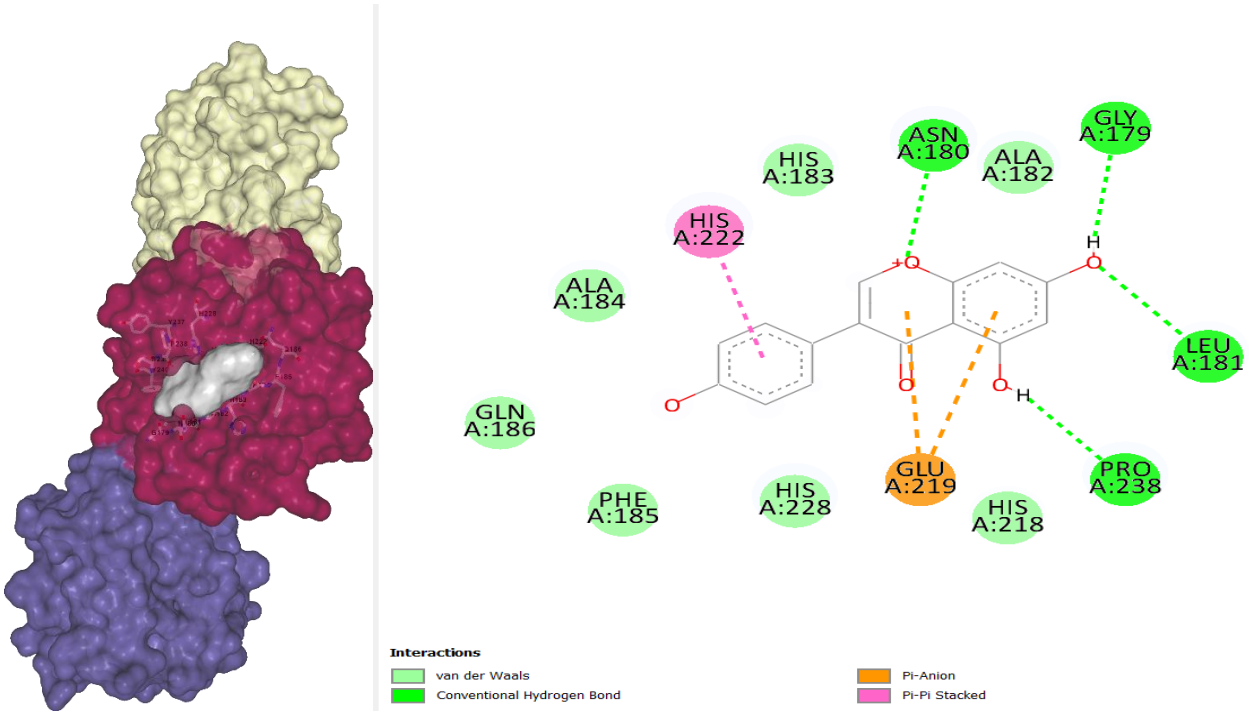

Figure 3.7 ; Geneistein binds with the active site of MMP-1

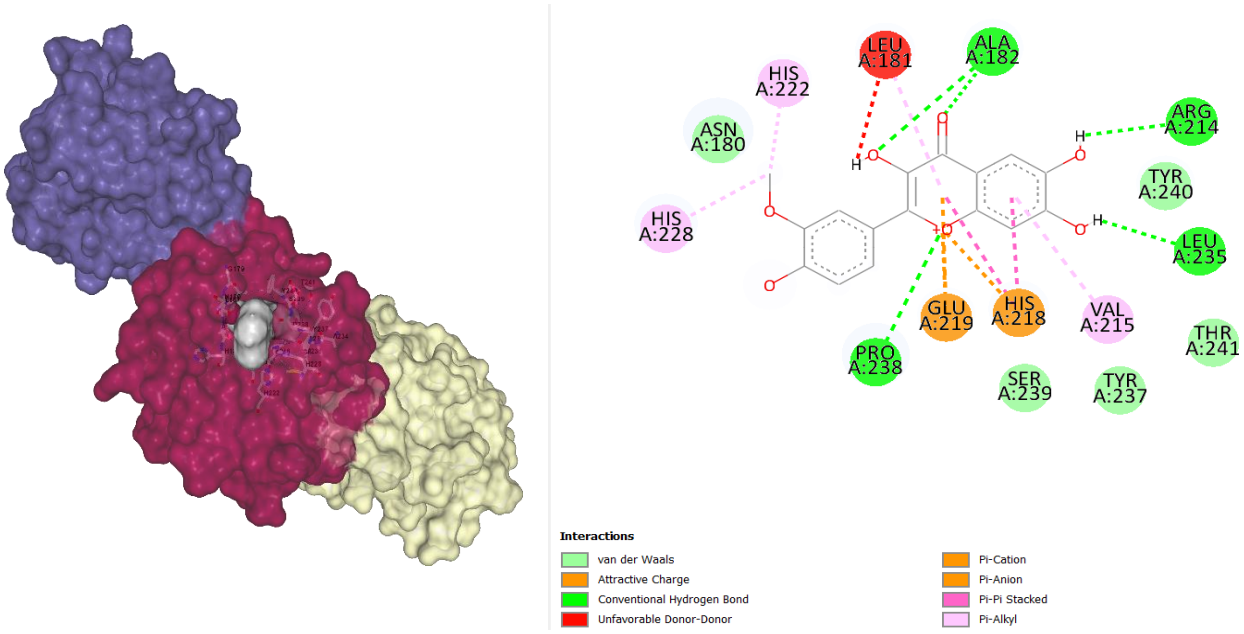

Figure 3.8 ; Isohamnetin binds with the active site of MMP-1

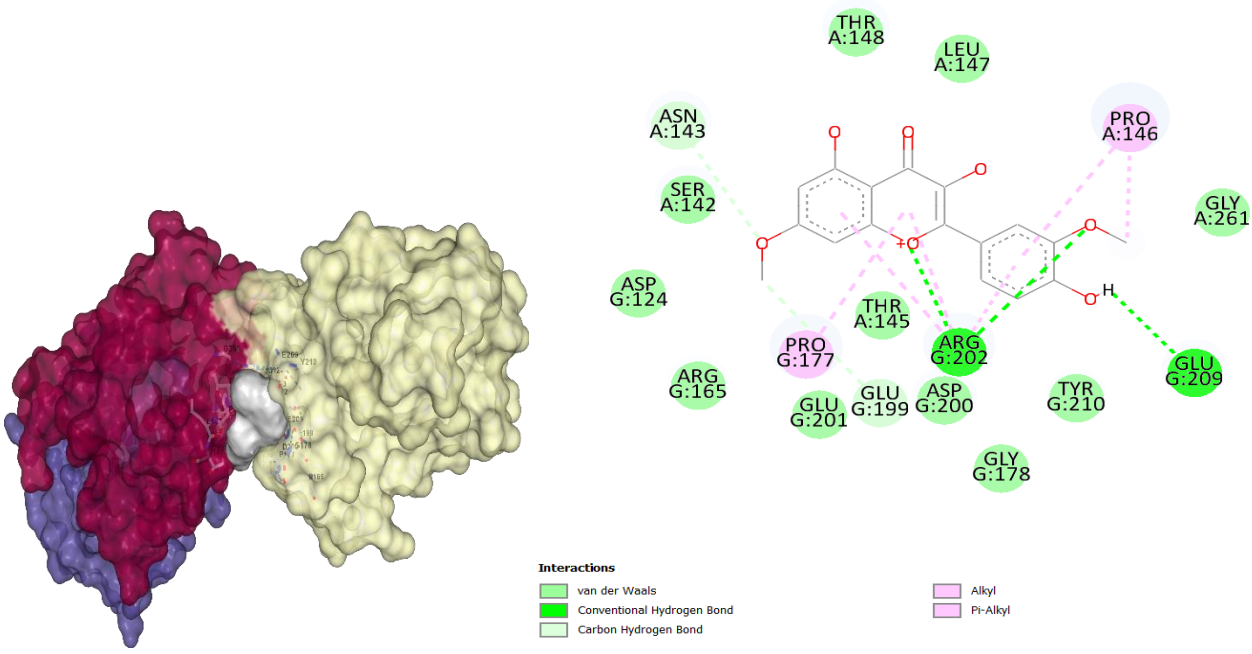

Figure3.9 ; Rhamnazin binds with the active site of MMP-1

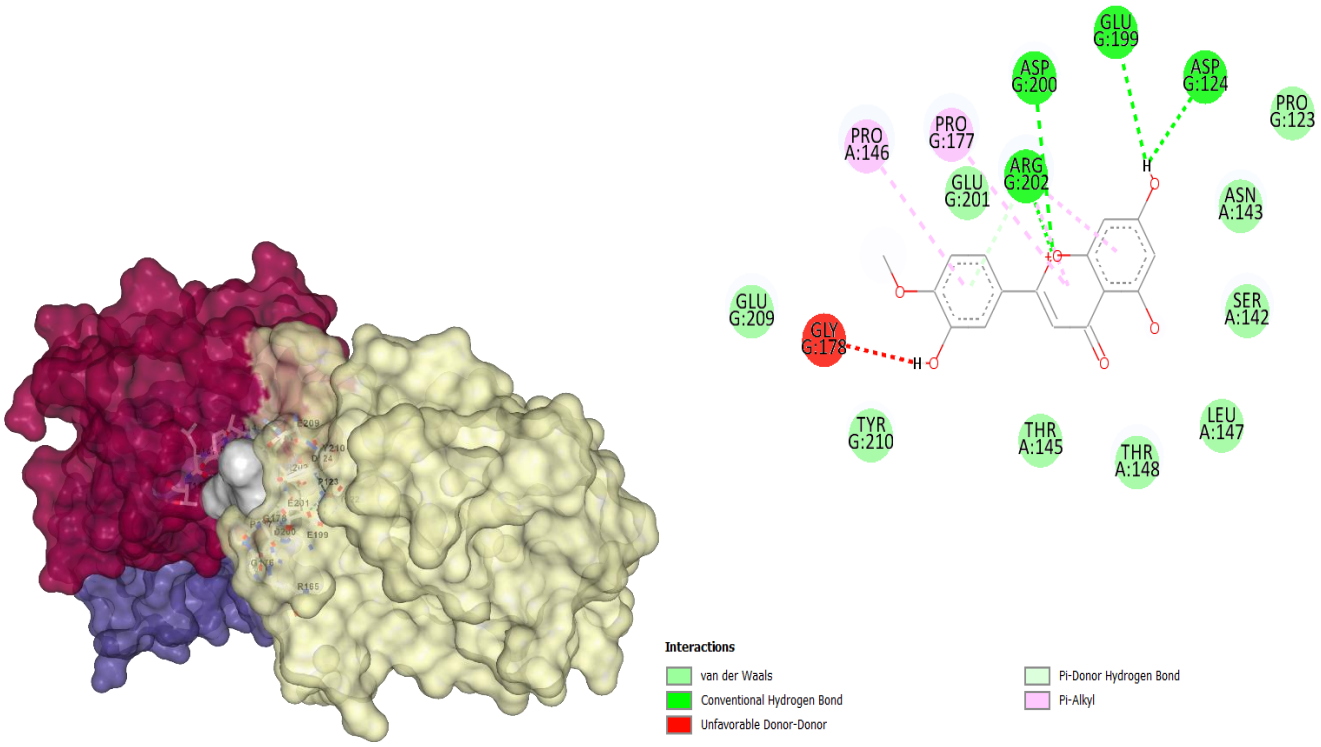

Figure 3.10 ; Diosmetin binds with the active site of MMP-1

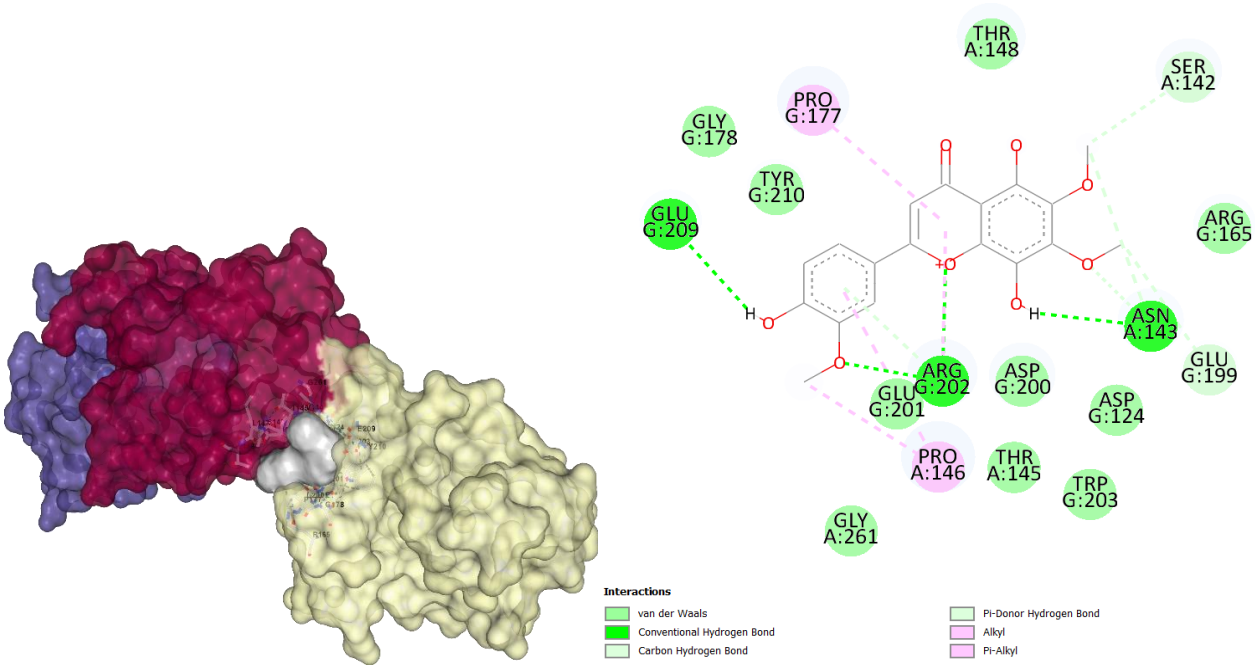

Figure 3.11 ; Isothymonin binds with the active site of MMP-1

Among the compounds tested, Isohamnetin had the highest binding activity to MMP-1 with a docking score of -9.0 kcal/mol, suggesting it as a highly potent MMP-1 inhibitor. Kaempferol and Diosmetin had high binding activities as well, both with docking scores of -8.5 kcal/mol. Geneistein, on the other hand, had the poorest interaction with a docking score of -7.8 kcal/mol.

Visualization of the docking poses indicated that all compounds were able to bind within the active site of MMP-1 and form multiple interactions likely to be responsible for their binding stability. The results indicate that some flavonoids, notably Isohamnetin, can be good lead compounds for further development as MMP-1 inhibitors.

4 Discussion

Better understanding of proteins involved in breast cancer pathogenesis is warranted to develop novel therapeutic methods for treating this disease (Gao et al., 2016). In this end, we selected top upregulated candidate proteins by STRING database and then performed visualization using Cytoscape tool. Molecular interaction network of proteins can be proficiently studied using network visualization software. Cytoscape can make a reputed protein-protein interaction network for target protein. Its vital arrangement principle is a network diagram, with biological entity (genes/proteins) represented as nodes and biological interactions represented as edges (Doncheva et al., 2012).

Visualization results illustrated that most of the top ranked candidate genes involved in metastasis of breast cancer are directly connected to either breast cancer or known breast cancer-associated proteins. The results of the current study indicated that MMP-1 may have key functions in the progression of breast cancer by interaction networks.

So far, prognostic targets were screened based on co-expression modules analysis for prostate cancer (Li et al., 2008), pancreatic cancer (Bahadorimonfared et al., 2024) and breast cancer (Place et al., 2011). However, our study is the first of its kind to validate the involvement of MMP-1 and its co-expressed proteins in breast cancer. Thus based on the literature (Nagel, 2019) (Wang et al., 2017). indicating the role of MMP-1 in the metastasis of breast cancer and based on network analysis indicating the interactions of various proteins with MMP-1 in breast cancer, MMP-1 was chosen as the target molecule and its inhibition by phytocompounds using *in-silico* studies were done.

5 Conclusion

A pre-filtering screening of 140 phytochemicals with the PASS online tool for 56 candidate compounds with satisfactory multi-target potential against breast cancer, metastasis, and MMP expression was noted. The cutoff criteria favored the selection of candidate compounds with more likelihood of activity in at least two relevant pharmacological actions. The enrichment in flavonoids in this filter set is congruent with prevailing literature on the varied biological activity of flavonoids, which includes anticancer and anti-MMP activities. This preliminary virtual screening offers a solid basis for the rest of the research. Combining DEG analysis with the building of PPI networks will position MMP-1 within the molecular context of breast cancer and determine lead compounds' possible synergic effects on related pathways. Admet and toxicity profiling will evaluate drug-likeness and safety so that only those with good pharmacokinetic and toxicological profiles are progressed to molecular docking studies. The combination of structure-based drug design and network pharmacology will enable the creation of new, potentially more effective, and safer phytochemical-derived therapies against breast cancer.

Conflict of Interest Statement

The authors declare that there is no conflict of interest regarding the publication of this manuscript.

REFERENCES

1. Duffy, M., McGowan, P., & Gallagher, W. (2007). Cancer invasion and metastasis: changing views. *The Journal of Pathology*, *214*(3), 283–293. https://doi.org/10.1002/path.2282
2. Gong, Y., Chippada-Venkata, U. D., & Oh, W. K. (2014). Roles of matrix metalloproteinases and their natural inhibitors in prostate cancer progression. *Cancers*, *6*(3), 1298–1327. https://doi.org/10.3390/cancers6031298
3. Kim, J., Harper, A., McCormack, V., Sung, H., Houssami, N., Morgan, E., Mutebi, M., Garvey, G., Soerjomataram, I., & Fidler-Benaoudia, M. M. (2025). Global patterns and trends in breast cancer incidence and mortality across 185 countries. *Nature Medicine*. https://doi.org/10.1038/s41591-025-03502-3
4. Overall, C. M., & Kleifeld, O. (2006). Validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nature Reviews. Cancer*, *6*(3), 227–239. https://doi.org/10.1038/nrc1821
5. Pillai, R. N., Alex, A., MS, N., Verma, V., Shaji, A., Pavithran, K., Vijaykumar, D. K., & John, D. (2025). Economic burden of breast cancer in India, 2000–2021 and forecast to 2030. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-024-83896-1
6. Seehawer, M., & Polyak, K. (2025). Epigenetic drivers of metalloproteinases and metastasis. *Trends in Cell Biology*. https://doi.org/10.1016/j.tcb.2025.02.010
7. Jiang, D., Wen, P., Zhang, S., Zhang, N., Shao, Q., Wang, G., Wang, L., Li, S., Qin, Y., Qu, F., & Zeng, X. (2025). Knowledge mapping of tumor microenvironment for breast cancer: a bibliometric analysis from 2014 to 2023. *Frontiers in Immunology*, *16*. https://doi.org/10.3389/fimmu.2025.1550988
8. Winer, A., Adams, S., & Mignatti, P. (2018). Matrix metalloproteinase inhibitors in cancer therapy: Turning past failures into future successes. *Molecular Cancer Therapeutics*, *17*(6), 1147–1155. https://doi.org/10.1158/1535-7163.mct-17-0646
9. Xiong, X., Zheng, L., Ding, Y., Chen, Y., Cai, Y., Wang, L., Huang, L., Liu, C., Shao, Z., & Yu, K. (2025). Breast cancer: pathogenesis and treatments. *Signal Transduction and Targeted Therapy*, *10*(1). https://doi.org/10.1038/s41392-024-02108-4
10. *Calculation of molecular properties and bioactivity score*. (n.d.). https://www.molinspiration.com/cgi/properties
11. *DeepToxLab*. (n.d.). https://deeptoxlab.scbdd.com/
12. Jejurikar, B. L., & Rohane, S. H. (2021). Drug designing in Discovery Studio. *Asian Journal of Research in Chemistry*, *14*(2), 135–138. https://doi.org/10.5958/0974-4150.2021.00025.0

13. Khan, T., Dixit, S., Ahmad, R., Raza, S., Azad, I., Joshi, S., & Khan, A. R. (2017). Molecular docking, PASS analysis, bioactivity score prediction, synthesis, characterization and biological activity evaluation of a functionalized 2-butanone thiosemicarbazone ligand and its complexes. *Journal of Chemical Biology*, *10*(3), 91–104. https://doi.org/10.1007/s12154-017-0167-y
14. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings 1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3–25. 1. *Advanced Drug Delivery Reviews*, *46*(1–3), 3–26. https://doi.org/10.1016/s0169-409x(00)00129-0
15. Liu, Y., Yang, X., Gan, J., Chen, S., Xiao, Z., & Cao, Y. (2022). CB-Dock2: improved protein–ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic Acids Research*, *50*(W1), W159–W164. https://doi.org/10.1093/nar/gkac394
16. Yang, X., Liu, Y., Gan, J., Xiao, Z., & Cao, Y. (2022). FitDock: protein–ligand docking by template fitting. *Briefings in Bioinformatics*, *23*(3). https://doi.org/10.1093/bib/bbac087
17. Zashumo, K. J., Mathivanan, V., Grace, H., & Leelavathi, D. (2022). In silico Toxicity Prediction on Novel Antiobesity Drug Using Cheminformatics Approaches. *International Journal of Pharmaceutical Investigation*, *12*(4), 418–422. https://doi.org/10.5530/ijpi.2022.4.72
18. Bahadorimonfared, A., Farahani, M., Tavirani, M. R., Razzaghi, Z., Arjmand, B., Rezaei, M., Nikzamir, A., Ardakani, M. J. E., & Mansouri, V. (2024). Stage analysis of pancreatic ductal adenocarcinoma via network analysis. *PubMed*, *17*(3), 297–3030. https://doi.org/10.22037/ghfbb.v17i3.2887
19. Cai, Y., Zhang, Q., Zhang, Y., Chen, L., & Huang, T. (2017). Identification of Genes Associated with Breast Cancer Metastasis to Bone on a Protein–Protein Interaction Network with a Shortest Path Algorithm. *Journal of Proteome Research*, *16*(2), 1027–1038. https://doi.org/10.1021/acs.jproteome.6b00950
20. Chen, X., Zhou, J., Luo, L., Han, B., Li, F., Chen, J., Zhu, Y., Chen, W., & Yu, X. (2015). Black rice anthocyanins suppress metastasis of breast cancer cells by targeting RAS/RAF/MAPK pathway. *BioMed Research International*, *2015*, 1–11. https://doi.org/10.1155/2015/414250
21. Doncheva, N. T., Assenov, Y., Domingues, F. S., & Albrecht, M. (2012). Topological analysis and interactive visualization of biological networks and protein structures. *Nature Protocols*, *7*(4), 670–685. https://doi.org/10.1038/nprot.2012.004
22. Gao, F., Nan, F., Feng, J., Lv, J., Liu, Q., & Xie, S. (2016). Identification of conserved and novel microRNAs in Porphyridium purpureum via deep sequencing and bioinformatics. *BMC Genomics*, *17*(1). https://doi.org/10.1186/s12864-016-2985-7
23. Geronikaki, A., Lagunin, A., Poroikov, V., Filimonov, D., Hadjipavlou-Litina, D., & Vicini, P. (2002). Computer aided prediction of biological activity spectra: Evaluating versus known and predicting of new activities for thiazole derivatives. *SAR And QSAR in Environmental Research*, *13*(3–4), 457–471. https://doi.org/10.1080/10629360290014322
24. Guo, L., Zhang, K., & Bing, Z. (2017). Application of a co-expression network for the analysis of aggressive and non-aggressive breast cancer cell lines to predict the clinical outcome of patients. *Molecular Medicine Reports*, *16*(6), 7967–7978. https://doi.org/10.3892/mmr.2017.7608
25. Hanessian, S., Moitessier, N., & Therrien, E. (2001). A comparative docking study and the design of potentially selective MMP inhibitors. *Journal of Computer-Aided Molecular Design*, *15*(10), 873–881. https://doi.org/10.1023/a:1014356529909
26. Hu, Z., Zhang, X., Yang, H., Qin, S., Liu, Y., Xiong, W., Yuan, B., Li, L., Yao, W., & Wu, D. (2017). Alterations in expression levels of genes in p53-related pathways

determined using RNA-Seq analysis in patients with breast cancer following CIK therapy. *Oncology Letters*. https://doi.org/10.3892/ol.2017.7205

27. Li, D., Zheng, Q., Fang, X., Ji, H., Yang, J., & Zhang, H. (2008). Theoretical study on potency and selectivity of novel non-peptide inhibitors of matrix metalloproteinases MMP-1 and MMP-3. *Polymer*, *49*(15), 3346–3351. https://doi.org/10.1016/j.polymer.2008.05.026
28. Lim, H., & Kim, H. (2007). Inhibition of mammalian collagenase, matrix metalloproteinase-1, by Naturally-Occurring flavonoids. *Planta Medica*, *73*(12), 1267–1274. https://doi.org/10.1055/s-2007-990220
29. Maity, N., Nema, N. K., Abedy, M. K., Sarkar, B. K., & Mukherjee, P. K. (2011). Exploring Tagetes erecta Linn flower for the elastase, hyaluronidase and MMP-1 inhibitory activity. *Journal of Ethnopharmacology*, *137*(3), 1300–1305. https://doi.org/10.1016/j.jep.2011.07.064
30. Majeed, A., & Mukhtar, S. (2023). Protein–Protein interaction network exploration using Cytoscape. *Methods in Molecular Biology*, 419–427. https://doi.org/10.1007/978-1-0716-3327-4_32
31. Nagel, J. (2019). Exploring the role of chemotherapy-induced anastasis in triple-negative breast cancer. *University of British Columbia*. https://doi.org/10.14288/1.0380841
32. Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F., & Furlong, L. I. (2016). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research*, *45*(D1), D833–D839. https://doi.org/10.1093/nar/gkw943
33. Place, A. E., Huh, S. J., & Polyak, K. (2011). The microenvironment in breast cancer progression: biology and implications for treatment. *Breast Cancer Research*, *13*(6). https://doi.org/10.1186/bcr2912
34. Poroikov, V. V., Filimonov, D. A., Ihlenfeldt, W., Gloriozova, T. A., Lagunin, A. A., Borodina, Y. V., Stepanchikova, A. V., & Nicklaus, M. C. (2002). PASS Biological Activity spectrum predictions in the enhanced Open NCI Database browser. *Journal of Chemical Information and Computer Sciences*, *43*(1), 228–236. https://doi.org/10.1021/ci020048r
35. Raman, K. (2010). Construction and analysis of protein–protein interaction networks. *Automated Experimentation*, *2*(1), 2. https://doi.org/10.1186/1759-4499-2-2
36. Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., Bork, P., Jensen, L. J., & Von Mering, C. (2016). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research*, *45*(D1), D362–D368. https://doi.org/10.1093/nar/gkw937
37. Vyas, B., Kumar, S., Bhowmik, R., & Akhter, M. (2023). Predicting the molecular mechanism-driven progression of breast cancer through comprehensive network pharmacology and molecular docking approach. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-40684-7
38. Wang, J., Ye, C., Lu, D., Chen, Y., Jia, Y., Ying, X., Xiong, H., Zhao, W., Zhou, J., & Wang, L. (2017). Matrix metalloproteinase-1 expression in breast carcinoma: a marker for unfavorable prognosis. *Oncotarget*, *8*(53), 91379–91390. https://doi.org/10.18632/oncotarget.20557