

Virtual Screening and Molecular Docking of Plant-Derived Phytochemicals Against Major Breast Cancer Proteins

Ratnesh Ramteke¹, Vedant Bhagat¹, Dr. Aishwarya Kopulwar²

^{1,2} Department of Biotechnology,

Priyadarshini College of Engineering, Nagpur, Maharashtra, India

vedantbhagat1105@gmail.com¹

Abstract: *The objective of this research was to complete virtual screening of a focused library of plant-derived phytochemicals against four important breast cancer proteins and to compare their predicted binding with appropriate clinical or structural reference compounds. To complete this objective, 20 phytochemicals were selected from PubChem and screened against HER2, estrogen receptor alpha (ER α), PI3K α , and KIFC1. Protein structures were obtained from the RCSB Protein Data Bank and prepared in BIOVIA Discovery Studio Visualizer. Ligands were prepared in AutoDock Tools, and molecular docking was carried out with AutoDock Vina. A binding affinity of $\Delta G \leq -7.5$ kcal/mol was used as the working cut-off for identifying strong docking candidates. The screening produced clear target-wise differences. Sanguinarine showed the strongest predicted binding against HER2 at -11.114 kcal/mol and performed better than TAK-285 and Lapatinib in the docking model. Ursolic Acid gave the strongest phytochemical result against ER α at -9.627 kcal/mol. Against PI3K α , Alpelisib remained the strongest reference compound, while Sanguinarine was the closest phytochemical at -10.029 kcal/mol. Diosgenin produced the best phytochemical score against KIFC1 at -8.674 kcal/mol. Seven compounds—Sanguinarine, Berberine, Ursolic Acid, Oleanolic Acid, Lupeol, Diosgenin, and Betulinic Acid—met the selected cut-off across all four targets. These results complete the objective by reducing the initial phytochemical library to a smaller group of promising multi-target docking candidates.*

Keywords: virtual screening, molecular docking, phytochemicals, breast cancer, HER2, multi-target screening

I. INTRODUCTION

Breast cancer involves several molecular pathways, so the present study selected four proteins that represent different aspects of the disease. In evaluating plant-derived inhibitors for breast cancer, researchers have frequently targeted the HER2 receptor across computer-aided screening models (Sait, 2020; Sohrab & Kamal, 2022; Akbar et al., 2024). ER α is an important target in hormone-dependent breast cancer and has also been examined using plant-derived ligands (Supriyanto et al., 2020; Mayanti et al., 2023; Priya et al., 2023). PI3K α represents the PI3K/Akt signalling pathway, which has been explored in phytochemical screening studies (Ali et al., 2023; Riaz et al., 2025; Uddin et al., 2025). KIFC1 has also gained attention as a potential target in aggressive and triple-negative breast cancer (Al-Mekhlafi et al., 2025; Singh et al., 2025).

To make the screening comparable, each target was assessed against a relevant control or benchmark. HER2 was compared with TAK-285 and Lapatinib, Since parent tamoxifen relies on metabolic conversion into 4-OHT to actively bind and inhibit ER α , we carried out our docking simulations using 4-OHT, PI3K α with Alpelisib, and KIFC1 with ADP and AZ82. These controls provided reference docking values against which the 20 phytochemicals could be



judged. Plant-derived compounds were selected because many already have reported biological activity, while virtual screening offers a practical way to narrow a larger set before experimental work.

The objective of this research was to run virtual screening of 15 to 20 carefully selected plant-derived phytochemicals against four key breast cancer proteins and identify natural ligands with predicted binding comparable to, or stronger than, the selected clinical or structural controls. The objective was completed using a final library of 20 phytochemicals.

II. LITERATURE REVIEW

Studies on HER2 provide a clear basis for using docking as the first screening step. Sait (2020) identified possible HER2 inhibitors through molecular docking of compounds from the ZINC database. Sohrab and Kamal (2022) combined docking with molecular dynamics to study natural anti-HER2 compounds, while Akbar et al. (2024), Silva et al. (2025), and Salsabil et al. (2026) used related in silico approaches to prioritise natural or plant-derived ligands. Together, these studies show that HER2 docking can be used to rank compounds before more detailed validation.

Similar work has been reported for ER α . Supriyanto et al. (2020) studied Neem-derived compounds, and Mayanti et al. (2023) and Priya et al. (2023) reported favourable docking of plant-derived ligands against ER α . Hossain et al. (2024) and Kříž et al. (2024) further strengthened ER α prediction using additional computational methods. These findings support the inclusion of ER α in the present multi-target screen, especially because Ursolic Acid, Oleanolic Acid, Lupeol, and Diosgenin also showed favourable predicted binding in the current dataset.

For PI3K-related signalling, Ali et al. (2023) identified phytochemicals with predicted activity against PI3K and mTOR, while Riaz et al. (2025) examined natural molecules acting on mediators of the PI3K/Akt pathway. Uddin et al. (2025) also used molecular docking to investigate fungal phytochemicals against PI3K/Akt-related breast cancer targets. These studies support PI3K α as a useful target for comparing natural compounds with an established control such as Alpelisib.

KIFC1 is a newer target in this area. Al-Mekhlafi et al. (2025) and Singh et al. (2025) used computational approaches to identify natural compounds with predicted KIFC1 inhibitory potential. Their work is relevant to the present study because KIFC1 was the most selective of the four targets, making it useful for distinguishing compounds that remained strong across the complete screening panel.

Taken together, the literature supports testing the same phytochemical library across several breast cancer proteins rather than relying on a single target. It also supports comparing docking scores with suitable reference ligands and treating docking as a ranking method rather than proof of therapeutic activity (Patel et al., 2024; Rossi et al., 2024; Mahanta et al., 2025; Smith et al., 2025). Based on this evidence, the present screening uses HER2, ER α , PI3K α , and KIFC1 as a four-target panel and carries forward only compounds that remain favourable across the panel.

III. RESEARCH METHODOLOGY

3.1 Target Selection, Protein Preparation, and Controls

Four protein targets were used throughout this screening: HER2, ER α , PI3K α , and KIFC1. Three-dimensional crystal structures were obtained from the RCSB Protein Data Bank and prepared in BIOVIA Discovery Studio Visualizer. Each target was paired with a reference control for comparison: HER2 with TAK-285 as the reference control and Lapatinib as the clinical benchmark; ER α with its native co-crystallised ligand **4-hydroxytamoxifen** (also referred to as **4-OHT**); PI3K α with Alpelisib as the clinical control; and KIFC1 with ADP as the native control and AZ82 as the clinical benchmark. Water molecules and co-crystallised ligands were removed before docking so that the selected binding regions could be prepared consistently for virtual screening.

3.2 Phytochemical Library

A focused library of 20 phytochemicals was used for the screening: Sanguinarine, Diosgenin, Ursolic Acid, Oleanolic Acid, Lupeol, Betulinic Acid, Berberine, Quercetin, Apigenin, Luteolin, Curcumin, Chlorogenic Acid, Kaempferol, Epigallocatechin Gallate, Piperine, Shikonin, Resveratrol, Plumbagin, Ferulic Acid, and Gallic Acid. The ligand



structures were obtained from PubChem and then prepared for docking using the same workflow so that the screening results could be compared consistently across all four targets.

3.3 Ligand Preparation and Molecular Docking

Ligands were prepared in AutoDock Tools by assigning Gasteiger charges, merging non-polar hydrogens, defining rotatable bonds, and saving the final structures in PDBQT format. Grid boxes were centred on the known catalytic, hinge, or motor domain binding region using the location of the native ligand as a guide. AutoDock Vina was run with an exhaustiveness value of 8 or higher. For each ligand, the best-ranked pose was selected and its predicted binding affinity was recorded in kcal/mol.

3.4 Screening Criterion for the First Objective

For this screening stage, a predicted binding affinity of $\Delta G \leq -7.5$ kcal/mol was used as the working cut-off for a strong docking result. The main purpose was to compare all 20 phytochemicals under the same docking conditions, identify the strongest compounds for each target, and find molecules that met the cut-off across all four proteins. The objective was considered complete when the full library had been docked against HER2, ER α , PI3K α , and KIFC1 and the results had been compared with the corresponding reference compounds.

IV. RESULTS AND DISCUSSION

All 20 phytochemicals were successfully screened against the four selected protein targets. The results are presented target by target, followed by a combined comparison of compounds that performed consistently across the full screening panel.

4.1 HER2 Docking Results

Compound	ΔG (kcal/mol)
TAK-285 (reference control)	-10.490
Lapatinib (clinical benchmark)	-10.340
Sanguinarine	-11.114
Piperine	-9.988
Luteolin	-9.855
Berberine	-9.745
Apigenin	-9.486

Table 1. Top HER2 docking results and reference controls.

Sanguinarine was the strongest HER2 binder in the screened library, with a predicted affinity of -11.114 kcal/mol. In this docking model, it scored better than both TAK-285 (-10.490 kcal/mol) and Lapatinib (-10.340 kcal/mol). Piperine, Luteolin, Berberine, and Apigenin also showed strong predicted binding. Overall, 18 of the 20 compounds met the -7.5 kcal/mol cut-off for HER2, while Ferulic Acid and Gallic Acid remained above the cut-off. HER2 therefore showed the broadest response to the phytochemical library, with Sanguinarine standing out as the main lead for this target.

4.2 ER α Docking Results

Compound	ΔG (kcal/mol)
Native co-crystallised ligand (4-OHT)	-8.927
Ursolic Acid	-9.627
Oleanolic Acid	-9.452
Lupeol	-9.236
Diosgenin	-9.160
Betulonic Acid	-8.857

Table 2. Top ER α docking results and native control.



The ER α results were dominated by triterpenoid compounds. Ursolic Acid (-9.627 kcal/mol), Oleanolic Acid (-9.452 kcal/mol), Lupeol (-9.236 kcal/mol), and Diosgenin (-9.160 kcal/mol) all performed better than the native ligand (-8.927 kcal/mol). Sixteen of the 20 compounds met the selected screening cut-off. The result is consistent with earlier docking studies that reported favourable interactions between plant-derived compounds and ER α . In the present library, Ursolic Acid was the strongest phytochemical for this target.

4.3 PI3K α Docking Results

Compound	ΔG (kcal/mol)
Alpelisib (clinical control)	-10.380
Sanguinarine	-10.029
Oleanolic Acid	-9.696
Ursolic Acid	-9.659
Betulinic Acid	-9.484
Diosgenin	-9.387

Table 3. Top PI3K α docking results and clinical control.

Alpelisib remained the strongest PI3K α binder at -10.380 kcal/mol. Sanguinarine was the closest phytochemical at -10.029 kcal/mol, a difference of 0.351 kcal/mol in the docking output. Oleanolic Acid, Ursolic Acid, Betulinic Acid, and Diosgenin also showed favourable predicted binding. Seventeen of the 20 compounds met the -7.5 kcal/mol cut-off. Although no phytochemical exceeded Alpelisib, the PI3K α screening identified several compounds with docking scores close enough to justify their inclusion in the combined multi-target comparison.

4.4 KIFC1 Docking Results

Compound	ΔG (kcal/mol)
ADP (native control)	-7.675
AZ82 (clinical benchmark)	-8.211
Diosgenin	-8.674
Sanguinarine	-8.447
Lupeol	-8.292
Ursolic Acid	-7.859
Oleanolic Acid	-7.780

Table 4. Top KIFC1 docking results and reference controls.

KIFC1 was the most selective target in the screening, with only seven compounds meeting the -7.5 kcal/mol cut-off. Diosgenin (-8.674 kcal/mol), Sanguinarine (-8.447 kcal/mol), and Lupeol (-8.292 kcal/mol) performed better than AZ82 (-8.211 kcal/mol). Ursolic Acid and Oleanolic Acid did not exceed AZ82, but both were stronger than the native ADP control (-7.675 kcal/mol). The lower pass rate at KIFC1 helped separate the most consistent compounds from those that performed well only at the other targets.

4.5 Pan-Target Candidates and Completion of the First Objective

Compound	HER2	ER α	PI3K α	KIFC1
Sanguinarine	-11.114	-8.015	-10.029	-8.447
Berberine	-9.745	-7.582	-8.854	-7.600
Ursolic Acid	-9.314	-9.627	-9.659	-7.859
Oleanolic Acid	-9.038	-9.452	-9.696	-7.780
Lupeol	-9.007	-9.236	-9.319	-8.292
Diosgenin	-8.742	-9.160	-9.387	-8.674



Betulinic Acid	-8.234	-8.857	-9.484	-7.586
----------------	--------	--------	--------	--------

Table 5. Seven phytochemicals meeting the -7.5 kcal/mol cut-off across all four targets.

Seven phytochemicals met the selected cut-off against all four proteins: Sanguinarine, Berberine, Ursolic Acid, Oleanolic Acid, Lupeol, Diosgenin, and Betulinic Acid. This is the main outcome of the screening. Sanguinarine was the strongest HER2 phytochemical and remained favourable against the other three targets. Diosgenin produced the strongest KIFC1 phytochemical score, while Ursolic Acid led the ER α results. Berberine met the threshold across all targets but remained close to the cut-off at ER α and KIFC1. The combined analysis therefore reduced the original set of 20 phytochemicals to seven compounds with consistent multi-target docking performance.

The screening stage was completed after all 20 phytochemicals had been evaluated against the four selected proteins, compared with the relevant controls, and reduced to a short list of multi-target candidates. These findings are still computational predictions and should not be interpreted as proof of anticancer activity.

4.6 Scope and Limitations of the Objective

This study was designed as a virtual-screening exercise, so the findings need to be read within that scope. Molecular docking provides a predicted binding score and a possible ligand pose, but it does not directly measure anticancer activity, selectivity in living cells, bioavailability, metabolism, or toxicity. Protein structures are also treated in a simplified form compared with the dynamic environment present in biological systems. For this reason, a strong docking value was used only to rank and shortlist compounds, not to claim that any phytochemical is already a drug candidate.

Setting our threshold at -7.5 kcal/mol gave us a clean, uniform baseline to rank all 20 compounds. While it successfully shortlisted seven multi-target hits, we know binding energy isn't biological proof. Honestly, the real takeaway from this docking round is practical: we managed to filter a broad phytochemical list down to a tight group worth carrying forward. Naturally, proving whether these actually work as therapeutics means we'll need MD simulations and lab-scale validation next.

V. CONCLUSION

The objective of the study, virtual screening of plant-derived phytochemicals against four major breast cancer proteins, was completed successfully. A total of 20 phytochemicals were docked against HER2, ER α , PI3K α , and KIFC1 using a common molecular docking workflow. The screening identified different lead compounds for different targets: Sanguinarine was strongest against HER2, Ursolic Acid was strongest among the phytochemicals against ER α , and Diosgenin gave the best phytochemical result against KIFC1. Alpelisib remained the strongest PI3K α control, although Sanguinarine produced a close docking score. Most importantly, seven compounds—Sanguinarine, Berberine, Ursolic Acid, Oleanolic Acid, Lupeol, Diosgenin, and Betulinic Acid—met the selected -7.5 kcal/mol cut-off across all four targets. This seven-compound group represents the main output of the completed objective. By completing this objective, we now have a clear, focused list of candidates for any future research, keeping our interpretation grounded in binding predictions rather than jumping to conclusions about clinical effectiveness.

VI. ACKNOWLEDGEMENTS

The authors thank the Department of Biotechnology, Priyadarshini College of Engineering, Nagpur, for the academic guidance and resources that supported this work.

REFERENCES

1. Akbar, N. H., et al. (2024). Screening, docking, and molecular dynamics analysis of Mitragyna speciosa (Korth.) compounds for targeting HER2 in breast cancer. *Informatics in Medicine Unlocked*, 45, 101502.
2. Ali, S., et al. (2023). Identification of phytochemical as a dual inhibitor of PI3K and mTOR: A structure-based computational approach. *PMC10520133*.



3. Al-Mekhlafi, A., et al. (2025). KIFC1 inhibition: Exploring the potential of propolis-derived small molecules for targeting cancer progression through in silico analysis. PMC12140393.
4. Hossain, M., et al. (2024). Pharmacoinformatics, molecular dynamics simulation, and quantum mechanics calculation based phytochemical screening of Croton bonplandianum against breast cancer by targeting estrogen receptor- α (ER α).
5. Křiž, Z., et al. (2024). Estrogen receptor alpha binders for hormone-dependent forms of breast cancer: e-QSAR and molecular docking supported by X-ray resolved structures. PMC11007693.
6. Mahanta, D. K., et al. (2025). In silico screening of plant-derived bioactive compounds: An integrated framework comprising ligand-likeness, docking, and ADMET toxicity prediction. Plants (MDPI), 15(4), 581.
7. Mayanti, T., et al. (2023). Unveiling the anti-cancer potential of onoceranoid triterpenes from Lansium domesticum Corr. cv. kokosan: An in silico study against estrogen receptor alpha.
8. Muthukrishnan, S., et al. (2024). Phytochemical analysis, physicochemical, pharmacokinetic properties and molecular docking studies of bioactive compounds in Otelia alismoides (L.) Pers. against breast cancer proteins. In Silico Pharmacology, 12.
9. Patel, R., et al. (2024). Phytochemicals in drug discovery-A confluence of tradition and innovation. PMC11354359.
10. Priya, S., et al. (2023). Molecular docking analysis of phytochemicals with estrogen receptor alpha. PMC10266367.
11. Riaz, M., et al. (2025). Chlorogenic acid and cinnamaldehyde in breast cancer cells: Predictive examination of pharmacokinetics and binding thermodynamics with the key mediators of PI3K/Akt signaling. PMC12383694.
12. Rossi, F., et al. (2024). Advances in cytotoxicity testing: From in vitro assays to in silico models.
13. Sait, K. H. (2020). Molecular docking analysis of HER-2 inhibitor from the ZINC database as anticancer agents. Bioinformation, 16(11), 882.
14. Salsabil, K., et al. (2026). In silico molecular docking and ADMET evaluation of active compounds from Acalypha indica L. against the HER2 breast cancer target. Biology, Medicine, & Natural Product Chemistry, 15(1), 45-51.
15. Silva, J., et al. (2025). Molecular docking and pharmacological in silico evaluation of camptothecin and related ligands as promising HER2-targeted therapies for breast cancer. PMC11941075.
16. Singh, V., et al. (2025). In silico identification of novel natural compounds as potential KIFC1 inhibitors for the therapeutic intervention of triple-negative breast cancer. PMC12748000.
17. Smith, J., et al. (2025). Breast cancer cytochromes P450: Chemopreventive and/or therapeutic targets for naturally occurring phytochemicals.
18. Sohrab, S. S., & Kamal, M. A. (2022). Screening, docking, and molecular dynamics study of natural compounds as an anti-HER2 for the management of breast cancer. Life, 12(11), 1729.
19. Supriyanto, S., et al. (2020). Potential use of compounds from Neem leaves (Azadirachta indica Juss) as PPAR γ and ER α inhibitors to control breast cancer cell growth in silico model. ALCHEMY, 8(1), 18-22.
20. Uddin, M., et al. (2025). Molecular docking appraisal of Pleurotus ostreatus phytochemicals as potential inhibitors of PI3K/Akt pathway for breast cancer treatment. PMC11792010.
21. Wang, Y., et al. (2026). Multi-omics integration and experimental validation reveal the mechanism of berberine against triple-negative breast cancer.
22. Zhang, L., et al. (2026). Modulation of the PI3K/Akt signaling pathway by steroidal saponins: Therapeutic implications in cancer.

