

COMPUTATIONAL INVESTIGATION OF NATURAL COMPOUNDS AS POTENTIAL THERAPEUTIC LEADS AGAINST BREAST CANCER: MOLECULAR DOCKING, MOLECULAR DYNAMICS AND ADMET ANALYSIS

Mahnoor Gul^{*1}, Dr Rashid Mahmood²

^{*1,2}Faculty of Sciences, Superior University of Lahore, Lahore

^{*1}mahnoorgul khan5124@gmail.com, ²rashid.mahmood.sgd@superior.edu.pk

DOI: <https://doi.org/10.5281/zenodo.22244099>

Keywords

Breast cancer, Natural products, EGCG, HER2, Molecular docking, Molecular dynamics, ADMET, Drug discovery.

Article History

Received: 14 July 2026

Accepted: 18 August 2026

Published: 31 August 2026

Copyright @Author

Corresponding Author: *

Mahnoor Gul

Abstract

Breast cancer is a diverse disease with molecular pathways related to DNA repair, tumor suppression, cell proliferation and hormonal signaling. This study performed analysis of the 10 selected natural compounds against the five breast cancer-associated protein targets by using an integrated in silico workflow of molecular docking, protein–ligand interaction analysis, docking validation, 200-ns molecular dynamics (MD) simulation and ADMET/drug-likeness prediction. The targets included: BRCA1, BRCA2, TP53, HER2 and estrogen receptor alpha (ER α). The highest predicted binding among the target proteins was seen with EGCG via docking score with the proteins BRCA1, BRCA2, TP53, HER2 and ER α of -9.6 , -9.2 , -8.8 , -10.1 and -9.4 kcal/mol respectively. HER2-EGCG complex was especially good as it contained six hydrogen bonds in total with the residues ASP863, THR862 and LYS753. The RMSD values obtained in the docking procedure were ranging from 1.23 to 1.67 Å for the docking validation. The average RMSD (1.85 Å), average RMSF (1.12 Å), average hydrogen-bond occupancy (6.4) and radius of gyration (21.45 Å) were the lowest for HER2–EGCG in MD simulations. Review of the predicted profiles by ADMET analysis was generally favorable, with EGCG being one violation of the Lipinski rule and showing low predicted hepatotoxicity. Integrated ranking was used and EGCG came out on top followed by Luteolin and Quercetin. The results show that EGCG (specifically when bound with HER2) represents the strongest computational lead in this data set but that experimental validation is necessary.

1. Introduction

Breast cancer is an example of a complex disease in which many molecular changes are involved and is an important area of therapeutic discovery. Computational methods can be used in drug discovery to prioritize candidates to be tested experimentally, which is a long process. Structurally diverse scaffolds have been found in nature and have been the focus of significant

interest as sources for potential therapeutic agents [3–5]. The molecular docking approach is well established for the estimation of binding modes and relative binding affinity of a protein–ligand complex, whereas molecular dynamics simulations give information about the temporal stability and flexibility of the predicted complexes [6,7]. Additionally, compounds can be ranked further based on their ADMET and

drug-likeness properties using ADMET and drug-likeness prediction. Thus, the present study tried to integrate these methods for screening natural compounds against five breast cancer-related proteins.

2. Materials and Methods

2.1 Protein targets and natural compounds

The rationale for choosing these five targets was based on their established roles in breast cancer-

associated DNA repair, tumor suppression, proliferation and regulation of hormones. These compounds were investigated: Quercetin, Curcumin, Resveratrol, epigallocatechin gallate (EGCG), Genistein, Luteolin, Kaempferol, Berberine, Apigenin and Naringenin.

Table 1. Selected breast cancer-associated protein targets.

Protein target	PDB ID	Resolution (Å)	Biological function
BRCA1	1T15	2.80	DNA repair
BRCA2	1MIU	2.70	DNA repair
TP53	3TS8	2.20	Tumor suppression
HER2	3PP0	2.25	Cell proliferation
ER α	3ERT	1.90	Hormonal regulation

Major Molecular Targets in Breast Cancer

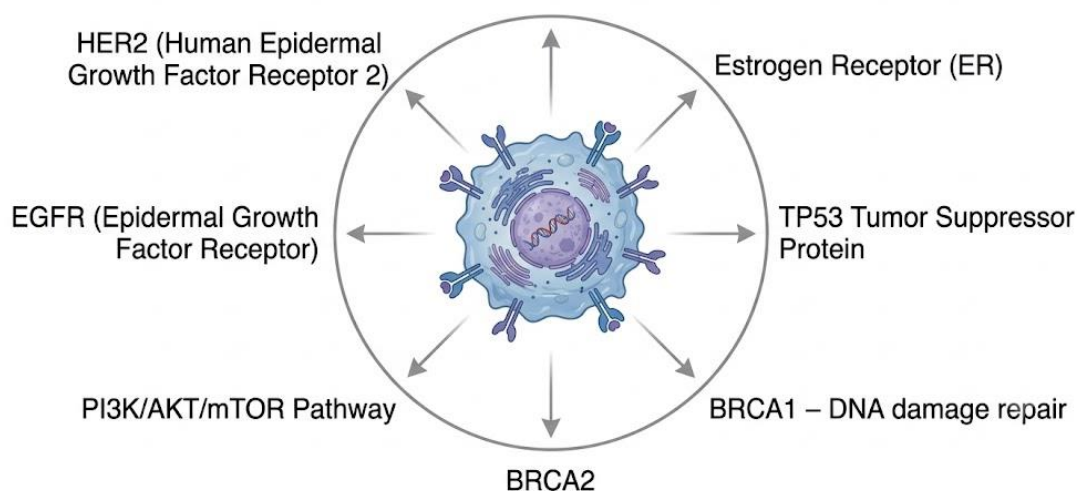


Figure 1. Major molecular targets associated with breast cancer, reproduced from the thesis source figure.

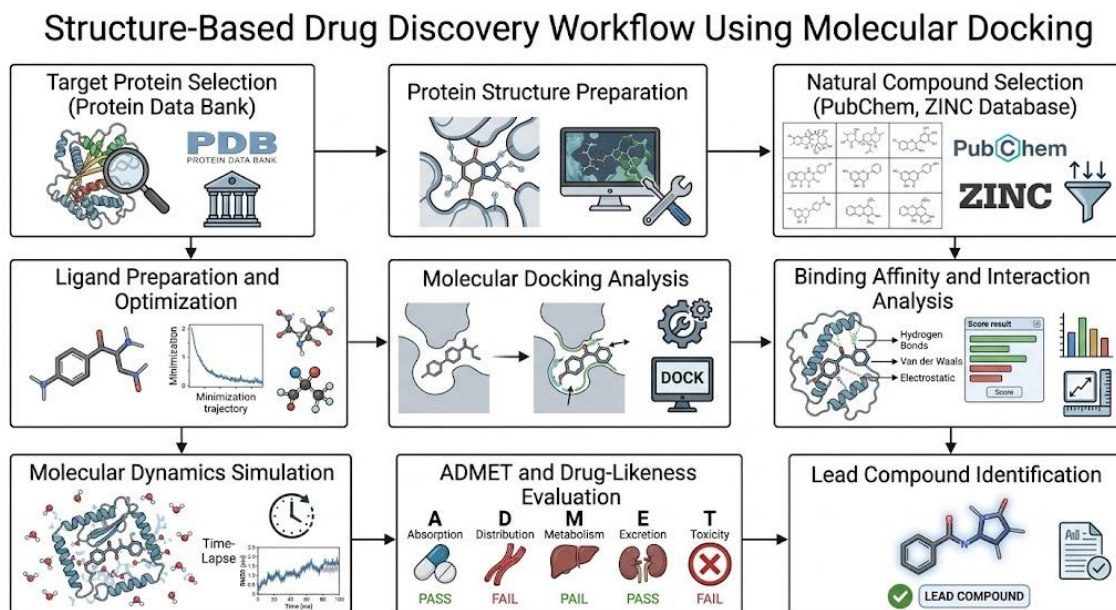


Figure 2. Structure-based drug-discovery and molecular-docking workflow used in the study.

The 3D structures of ligands have been retrieved from chemical databases in SDF format and prepared for docking. The physicochemical and the pharmacokinetic properties were evaluated using the SwissADME server and toxicity prediction was evaluated using the pkCSM server according to the methodology described in the thesis. Lipinski's Rule of Five used for Drug-likeness Assessment.

2.2 Molecular docking and interaction analysis

The selected protein targets were then docked with the prepared ligands. The more negative the docking score, the more that was predicted to dock. The identification of the hydrogen bonds, hydrophobic contacts and key residues were then performed by means of protein-ligand interaction analysis.

The reference ligands were re-docked to test the validation of docking and the comparison of the

experimental and predicted docked poses was achieved by RMSD. A value of less than 2.0 Å was deemed to be satisfactory. These 200-ns MD simulations were then carried out for the strongest complexes. RMSD, RMSF, hydrogen-bond occupancy and radius of gyration were used to evaluate complex stability. A drug-likeness and an ADMET analysis were then added in the process of final lead prioritization.

3. Results

3.1 Molecular docking and binding affinity

EGCG was found to have the highest predicted binding in all the five protein targets. HER2-EGCG (−10.1 kcal/mol), BRCA1-EGCG (−9.6 kcal/mol) and ERα-EGCG (−9.4 kcal/mol) showed the best score in the data set. Luteolin, Berberine, Curcumin and Quercetin revealed fairly good binding

Table 2. Binding affinity scores (kcal/mol) of the ten natural compounds against the five protein targets.

Compound	BRCA1	BRCA2	TP53	HER2	ERα
Quercetin	−8.4	−8.0	−7.9	−8.6	−8.2
Curcumin	−9.1	−8.8	−8.6	−9.5	−8.9
Resveratrol	−7.8	−7.6	−7.3	−8.0	−7.7
EGCG	−9.6	−9.2	−8.8	−10.1	−9.4
Genistein	−8.3	−8.1	−7.8	−8.5	−8.9
Luteolin	−8.7	−8.4	−8.2	−9.0	−8.6

Kaempferol	-8.1	-7.9	-7.6	-8.4	-8.0
Berberine	-8.9	-8.7	-8.4	-9.2	-8.7
Apigenin	-8.2	-7.8	-7.5	-8.3	-8.1
Naringenin	-7.9	-7.5	-7.2	-8.1	-7.8

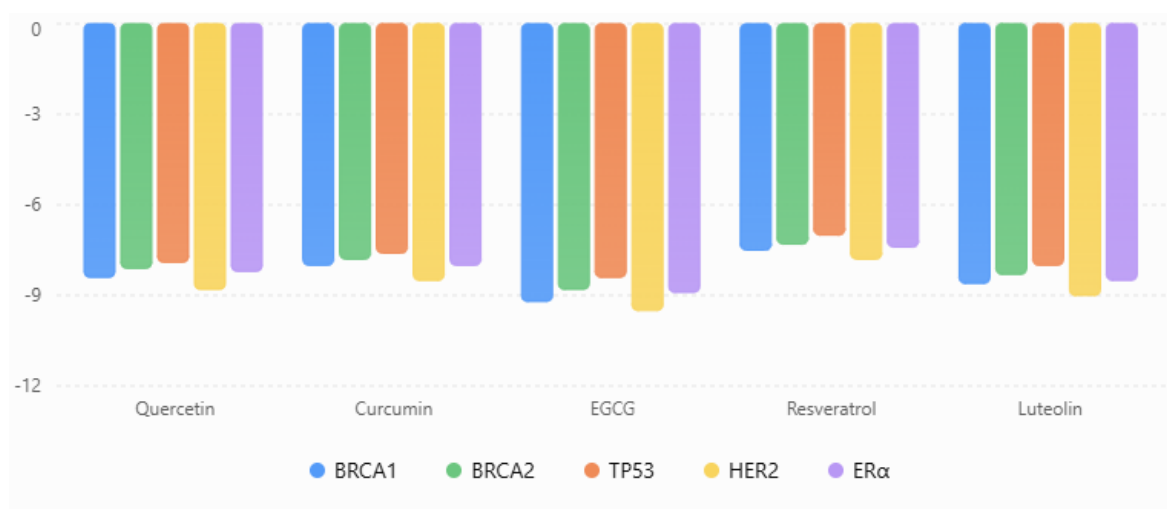


Figure 3. Binding affinities of selected natural compounds against breast cancer protein targets.

Figure 4.3 Heat Map of Docking Scores

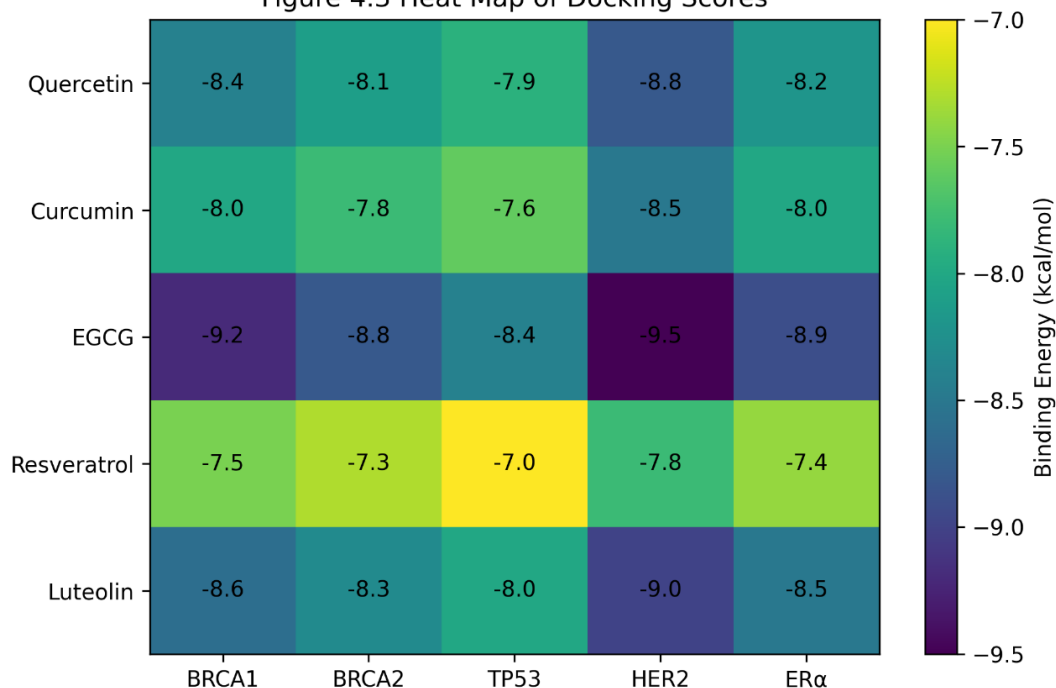


Figure 4. Heat map of docking scores across the selected targets.

3.2 Protein–ligand interaction and structural analysis

The HER2-EGCG complex had the largest interaction profile. Six hydrogen bonds were

reported: ASP863, THR862 and LYS753. Other complexes that benefited were HER2–Curcumin, BRCA1-EGCG, ERα-EGCG and HER2–Berberine.

Table 3. Top protein-ligand interactions identified in the docking analysis.

Complex	Binding energy (kcal/mol)	H-bonds	Key residues
HER2-EGCG	-10.1	6	ASP863, THR862, LYS753
HER2-Curcumin	-9.5	4	MET801, ASP808
BRCA1-EGCG	-9.6	5	ARG1699, SER1655
ER α -EGCG	-9.4	5	GLU353, ARG394
HER2-Berberine	-9.2	3	LEU726, VAL734

Figure 4.4. 2D Interaction Diagram of HER2-EGCG Complex

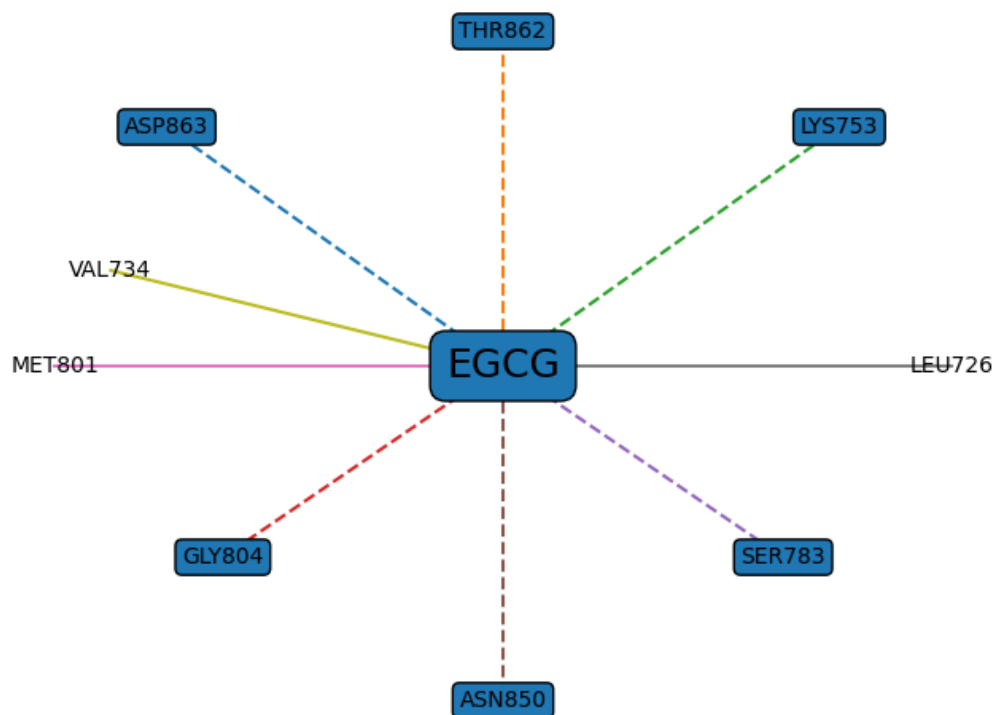


Figure 5. Two-dimensional interaction diagram of the HER2-EGCG complex.

Figure 4.5. 3D Binding Pose of EGCG within HER2 Active Site

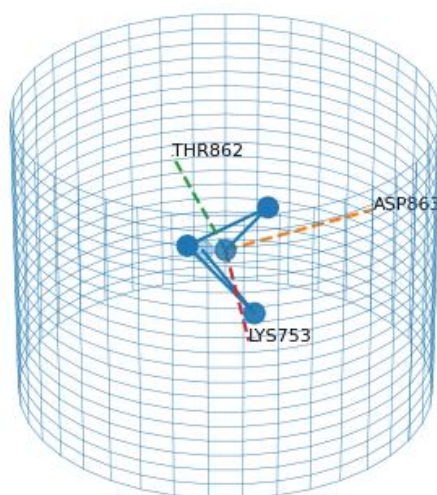


Figure 6. Three-dimensional binding pose of EGCG within the HER2 active site.

3.3 Docking validation

Re-docking validation yielded RMSD values ranging from 1.23 to 1.67 Å, which are lower than the standard RMSD value of 2.0 Å. The

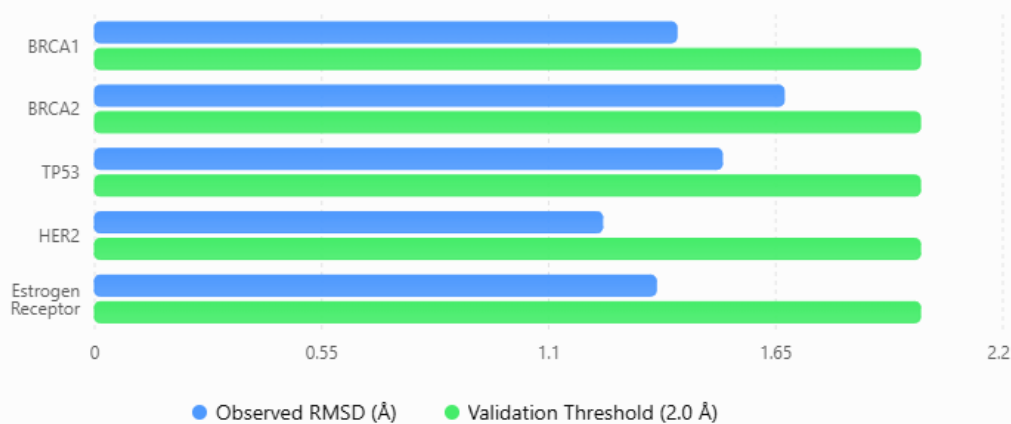
RMSD of 1.23Å for HER2 is the lowest, indicating that it has the best validation among the selected targets.

Table 4. Docking validation RMSD values.

Protein	RMSD (Å)
BRCA1	1.41
BRCA2	1.67
TP53	1.52
HER2	1.23
Estrogen receptor	1.36

Figure 4.6. Docking Validation RMSD Values

Comparison of RMSD values obtained from re-docking studies against the acceptable validation threshold (2.0 Å).



All RMSD values are below the validation threshold, confirming docking reliability.

Figure 7. Docking validation RMSD values.

3.4 Molecular dynamics simulation

The molecular dynamics simulation duration was 200-ns for four promising complexes: HER2-EGCG, BRCA1-EGCG, HER2-

Luteolin and ER α -Quercetin. On average, HER2 - EGCG had the lowest RMSD, suggesting that this simulated system was the most stable.

Table 5. RMSD values of simulated protein-ligand complexes.

Complex	Average RMSD (Å)	Maximum RMSD (Å)
HER2-EGCG	1.85	2.12
BRCA1-EGCG	1.98	2.25
HER2-Luteolin	2.16	2.48
ER α -Quercetin	2.08	2.36

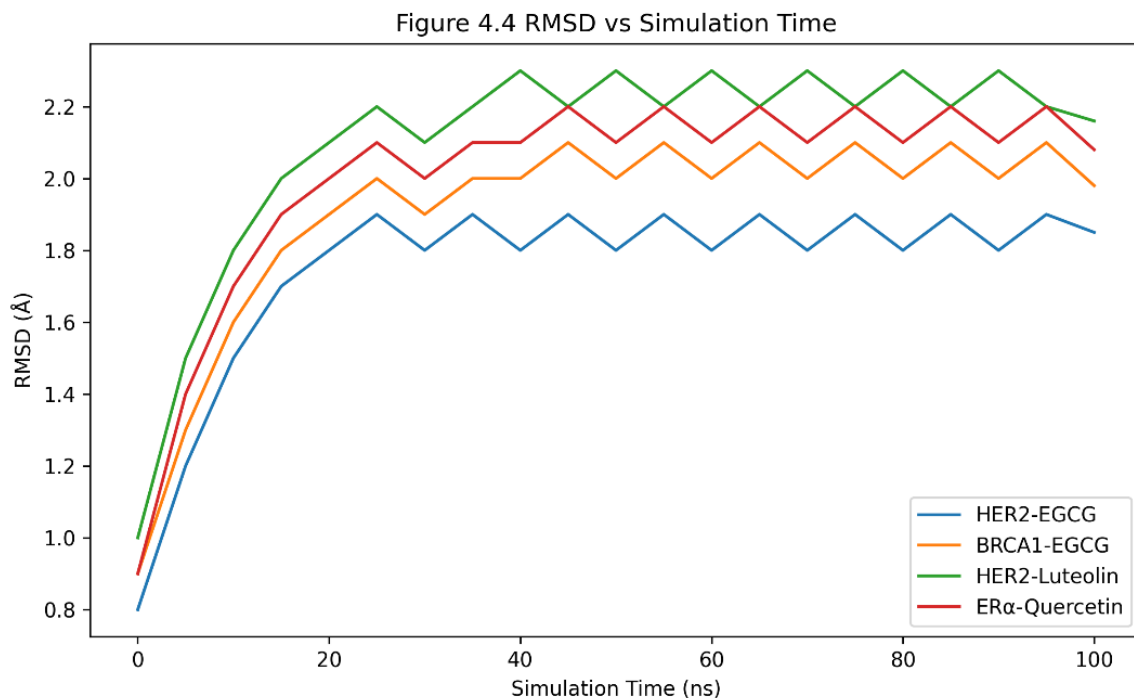


Figure 8. RMSD versus simulation time for the selected protein–ligand complexes.

Table 6. RMSF values of simulated complexes.

Complex	Average RMSF (Å)	Highest fluctuation region
HER2-EGCG	1.12	Loop region 845–855
BRCA1-EGCG	1.28	Loop region 1660–1670
HER2-Luteolin	1.35	Loop region 845–855
ERα-Quercetin	1.30	Residues 520–530

Figure 4.5. RMSF Analysis of Simulated Protein–Ligand Complexes

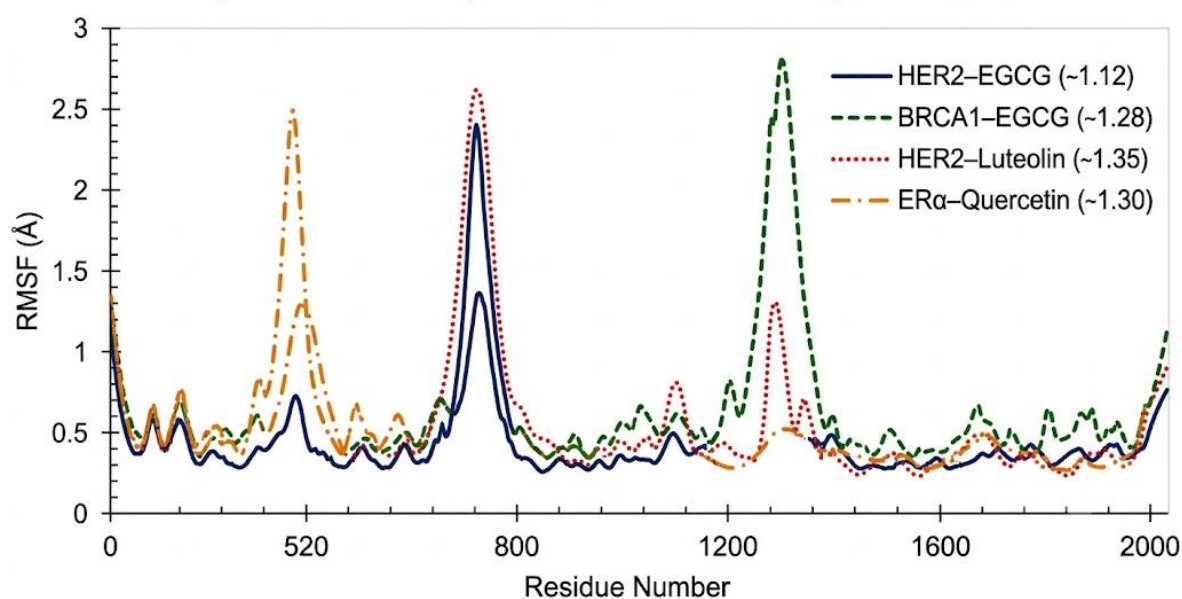
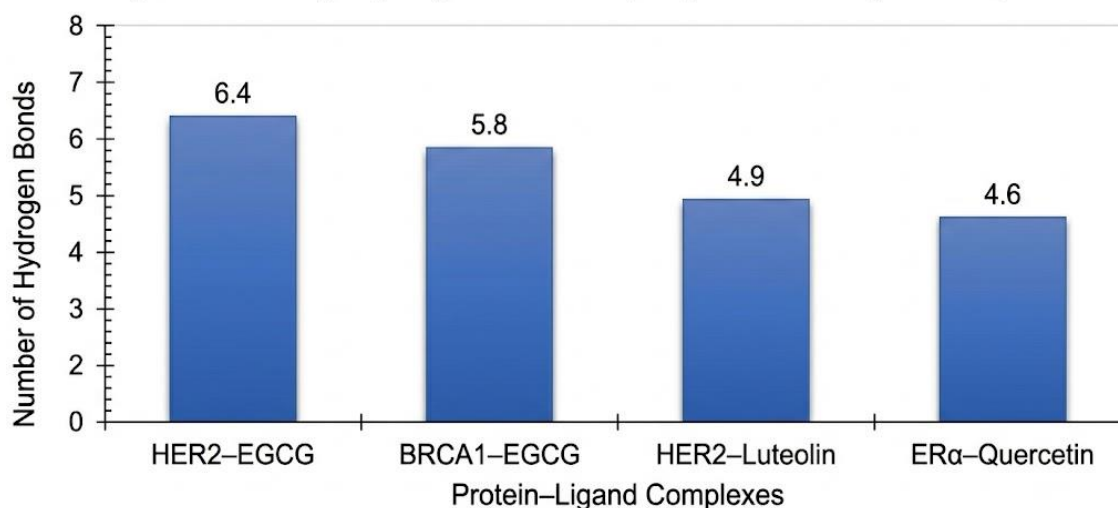


Figure 9. RMSF analysis of simulated protein–ligand complexes.

Table 7. Hydrogen-bond occupancy during molecular dynamics.

Complex	Average H-bonds	Maximum H-bonds
HER2-EGCG	6.4	8
BRCA1-EGCG	5.8	7
HER2-Luteolin	4.9	6
ER α -Quercetin	4.6	6

Figure 4.6. Average Hydrogen Bond Occupancy of Protein–Ligand Complexes



Note. Maximum hydrogen bonds for each complex are as follows: HER2-EGCG = 8, BRCA1-EGCG = 7, HER2-Luteolin = 6, and ER α -Quercetin = 6.

Figure 10. Average hydrogen-bond occupancy of protein–ligand complexes.

Table 8. Radius of gyration values during molecular dynamics.

Complex	Average R _g (Å)
HER2-EGCG	21.45
BRCA1-EGCG	22.16
HER2-Luteolin	21.78
ER α -Quercetin	21.93

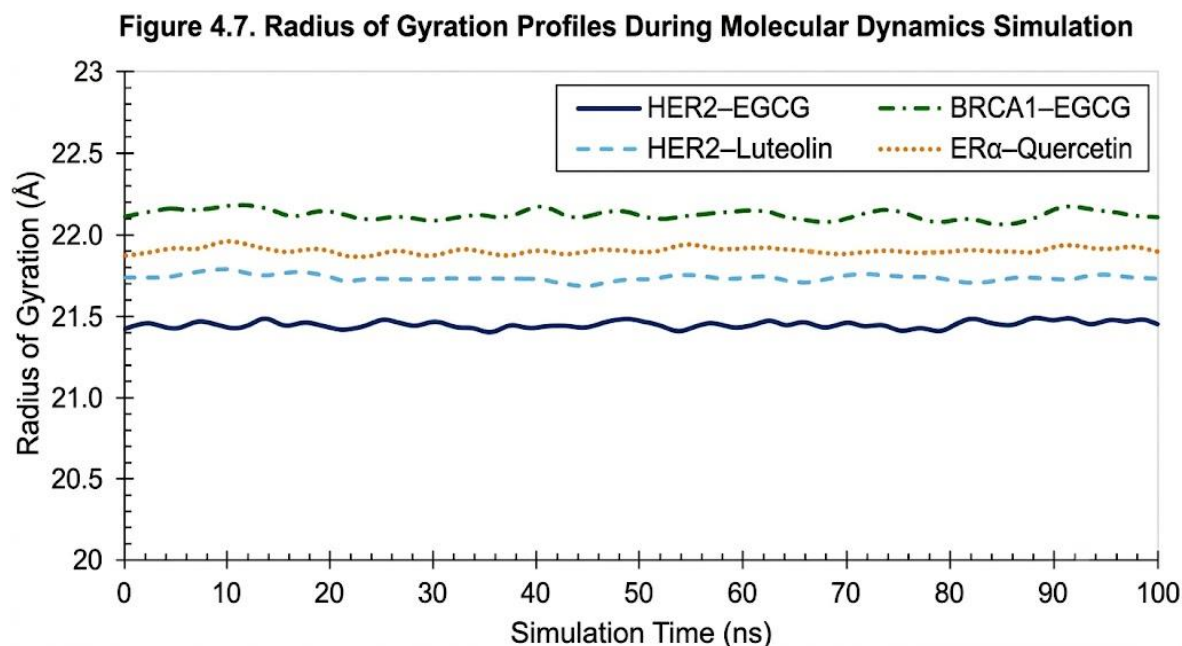


Figure 11. Radius of gyration profiles during molecular dynamics simulation.

3.5 ADMET and drug-likeness

Quercetin, Curcumin and Luteolin did not have any Lipinski violations and had good predicted gastrointestinal absorption. Despite having one

Lipinski violation and moderate gastrointestinal absorption and absorption ratio, EGCG maintained a good predicted safety profile with low predicted hepatotoxicity.

Table 9. ADMET and drug-likeness properties of selected compounds.

Compound	Lipinski violations	GI absorption	BBB permeability	Hepatotoxicity	Drug-likeness
Quercetin	0	High	No	Low	Good
Curcumin	0	High	No	Low	Good
EGCG	1	Moderate	No	Low	Acceptable
Resveratrol	0	High	Yes	Low	Good
Luteolin	0	High	No	Low	Good

Figure 4.8. Comparative ADMET and Drug-Likeness Profiles of Selected Compounds

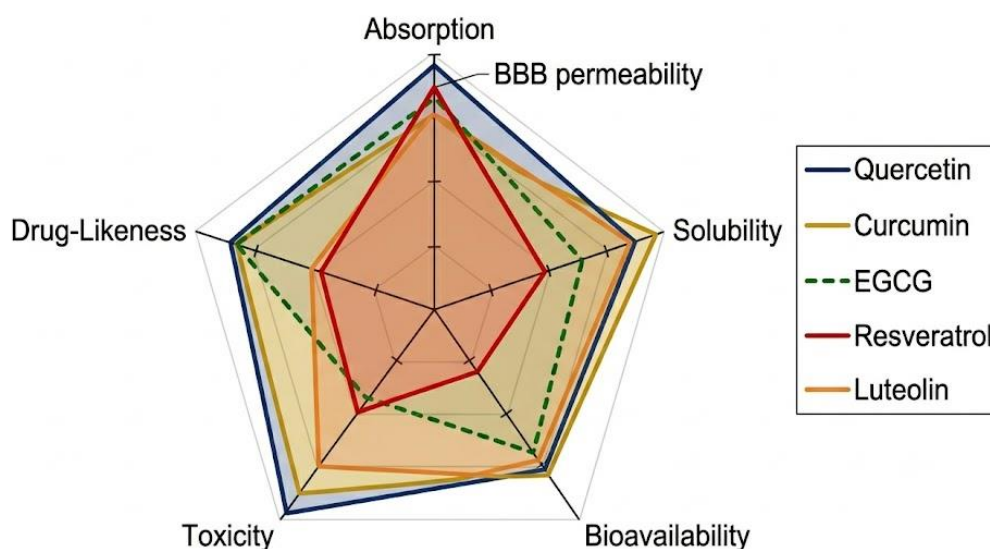


Figure 12. Comparative ADMET and drug-likeness profiles of selected compounds.

3.6 Integrated lead ranking

When evaluating the performance of docking, the molecular-dynamics stability and the ADMET properties, EGCG was the best,

followed by Luteolin and Quercetin. This is a ranking which includes the integrated computational assessment, not just the docking score.

Table 10. Comparative ranking of the leading compounds.

Rank	Compound	Integrated score shown in thesis
1	EGCG	95
2	Luteolin	88
3	Quercetin	85
4	Curcumin	78
5	Resveratrol	70

Figure 4.9. Comparative Ranking of Lead Compounds Based on Integrated Computational Analysis

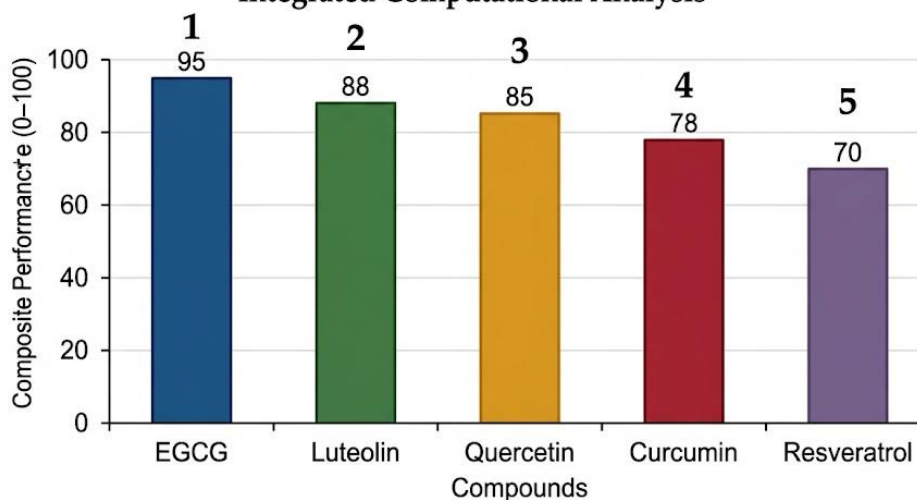


Figure 13. Comparative ranking of lead compounds based on integrated computational analysis.

4. Discussion

The results always show that the best computational lead from the dataset investigated is the EGCG. Its scores for binding were the best for all five targets and its HER2 complex had a large hydrogen-bond network. This aligns with the overall applicability of molecular docking in studying protein-ligand recognition and prioritizing compounds for in-depth study [6,8]. The evidence of the MD results comes as a bonus. The lowest average RMSD (1.85 Å), RMSF (1.12 Å), the highest average hydrogen-bond occupancy (6.4) and the most stable Rg profile were observed for HER2-EGCG complex. Agreements between docking and MD increase the computational priority of HER2-EGCG. These results are still predictive, though: the quality of the protein structure, the length of the simulation and the choice of force-field can impact the computational results as well as the docking scoring functions.

The results of ADMET highlight the fact that factors other than binding affinity should be considered when selecting a lead. The one rule violation by EGCG did not rule it out as a candidate because of the low hepatotoxicity prediction, and the favorable overall pharmacokinetic profile of EGCG in the thesis. The combination of Luteolin and Quercetin exhibited excellent docking performance and in particular drug-likeness properties. The approach is integrated screening, which is in line with the current trend of computational tools in natural product drug discovery [4,5,9].

Notably, the study is not a clinical efficacy study. Molecular docking and MD simulation do not represent the real biological environment, and the prediction of ADMET are not experimental. Use of these findings as a preliminary in silico finding is then itself the subject of the thesis. The thesis itself therefore recommends experimental validation of these findings.

5. Conclusion

Ten natural compounds were screened, using an integrated computational workflow, against five breast cancer-associated targets. EGCG proved to be the best overall candidate, with the HER2-EGCG complex having the highest docking score (−10.1 kcal/mol), hydrogen-bonding interactions that were favorable and MD

stability metrics that were most favorable. The average RMSD, RMSF, hydrogen-bond occupancy and radius of gyration values for HER2-EGCG were 1.85 Å, 1.12 Å, 6.4 and 21.45 Å, respectively. Luteolin was the second most common compound and Quercetin was the third. The outcomes of these investigations indicate that EGCG and the other top prioritized compounds deserve further investigation in the laboratory and in vivo to determine the therapeutic potential, although this cannot be determined at this point in the investigation.

References

1. Berdigaliyev, N., & Aljofan, M. (2020). An overview of drug discovery and development. *Future Medicinal Chemistry*, 12(10), 939–947.
2. Kiriiri, G. K., Njogu, P. M., & Mwangi, A. N. (2020). Exploring different approaches to improve the success of drug discovery and development projects: A review. *Future Journal of Pharmaceutical Sciences*, 6(1), 27.
3. Sun, W., & Shahrajabian, M. H. (2023). Therapeutic potential of phenolic compounds in medicinal plants—Natural health products for human health. *Molecules*, 28(4), 1845.
4. Dzobo, K. (2022). The role of natural products as sources of therapeutic agents for innovative drug discovery. *Comprehensive Pharmacology*, 408.
5. Atanasov, A. G., et al. (2021). Natural products in drug discovery: Advances and opportunities. *Nature Reviews Drug Discovery*, 20(3), 200–216.
6. Raval, K., & Ganatra, T. (2022). Basics, types and applications of molecular docking: A review. *IP International Journal of Comprehensive and Advanced Pharmacology*, 7(1), 12–16.
7. Singh, S., Baker, Q. B., & Singh, D. B. (2022). Molecular docking and molecular dynamics simulation. In *Bioinformatics* (pp. 291–304). Elsevier.

8. Di Cera, E. (2020). Mechanisms of ligand binding. *Biophysics Reviews*, 1(1).
9. Naeem, A., et al. (2022). Natural products as anticancer agents: Current status and future perspectives. *Molecules*, 27(23), 8367.