IJPST
Indonesian Journal of Pharmaceutical Science and Technology

Research Article
Vol.12 Suppl. 3 (2025): 256-269
https://doi.org/10.24198/ijpst.v12s3.70710.

In Silico Screening of Soursop Leaf (*Annona muricata* L.) Secondary Metabolites as Potential Estrogen Receptor Alpha Inhibitors

Fachrizal D. Kautsar, John E. Sianipar, Ulaya W. Haiba, Neysa Z. Fathinabila, Rizky P. Wulandari, Halwa A. Nurdin, and Rina F. Nuwarda*

[2]Department of Pharmaceutical Analysis dan Medicinal Chemistry, Faculty of Pharmacy Universitas Padjadjaran, Jatinangor, Indonesia

Abstract

Breast cancer is the most prevalent malignancy among women worldwide. Estrogen receptor alpha (ERα) plays a critical role in regulating cancer cell proliferation, differentiation, and survival, making it an important therapeutic target. *Annona muricata* Linn. has attracted considerable attention because its leaves contain diverse secondary metabolites with reported anticancer activities, making them a promising source of potential ERα inhibitors. Therefore, this study aimed to evaluate the inhibitory potential of ten selected secondary metabolites from A. muricata leaves against ERα using an in silico approach. Drug-likeness and pharmacokinetic properties were assessed using Lipinski's Rule of Five and ADMET prediction. Ligand-based pharmacophore modeling was performed, followed by pharmacophore-based virtual screening and molecular docking. Four compounds, namely chlorogenic acid, luteolin, quercetin, and daidzein, fulfilled the pharmacophore criteria with fit scores of 47.17; 47.17; 47.17; and 46.97, respectively, and were subsequently evaluated by molecular docking. Among them, daidzein exhibited the strongest inhibitory potential, with a binding free energy of −8.56 kcal/mol and an inhibition constant of 513.26 nM. Overall, these findings suggest that daidzein is the most promising secondary metabolite of A. muricata leaves for ERα inhibition and may serve as a lead compound for further computational and experimental validation.

Keywords: breast cancer, estrogen receptor alpha, in silico, secondary metabolites, soursop leaf.

Article History:
Submitted 18 May 2026
Revised 26 June 2026
Accepted 01 July 2026
Published 12 August 2026

*Corresponding author:
rina.nuwarda@unpad.ac.id

Citation:
Kautsar FD, Sianipar JE, Haiba UW, Fathinabila NZ, Wulandari RP, Nurdin HA, et al. In Silico Screening of Soursop Leaf (*Annona muricata* L.) Secondary Metabolites as Potential Estrogen Receptor Alpha Inhibitors. Indonesian Journal of Pharmaceutical Science and Technology. 2025: 12 (3), 256-269.

Skrining *In Silico* Metabolit Sekunder Daun Sirsak (*Annona muricata* L.) sebagai Potensi Inhibitor Reseptor Estrogen Alfa

Abstrak

Kanker payudara merupakan keganasan yang paling banyak terjadi pada wanita di seluruh dunia. Estrogen Receptor Alpha (ERα) berperan penting dalam mengatur proliferasi, diferensiasi, dan kelangsungan hidup sel kanker, sehingga menjadi salah satu target utama dalam pengembangan terapi. Annona muricata Linn. telah banyak dilaporkan memiliki aktivitas antikanker karena kandungan metabolit sekundernya yang beragam, sehingga berpotensi menjadi sumber kandidat inhibitor ERα. Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi potensi penghambatan 10 metabolit sekunder terpilih dari daun sirsak (*Annona muricata* Linn.) terhadap ERα melalui pendekatan *in silico*. Kelayakan senyawa sebagai kandidat obat dan profil farmakokinetiknya dievaluasi menggunakan Aturan Lipinski dan prediksi ADMET. Selanjutnya, dilakukan pemodelan farmakofor berbasis ligan yang diikuti dengan virtual screening berbasis farmakofor. Empat senyawa, yaitu asam klorogenat, luteolin, kuersetin, dan daidzein, memenuhi kriteria farmakofor dengan nilai fit score berturut-turut sebesar 47,17; 47,17; 47,17, dan 46,97, kemudian dievaluasi lebih lanjut menggunakan molecular docking. Di antara keempat senyawa tersebut, daidzein menunjukkan potensi penghambatan paling baik terhadap ERα dengan energi bebas ikatan sebesar −8,56 kcal/mol dan konstanta inhibisi sebesar 513,26 nM. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa daidzein merupakan metabolit sekunder daun sirsak yang paling menjanjikan sebagai kandidat inhibitor ERα dan berpotensi dikembangkan lebih lanjut melalui validasi komputasional maupun eksperimental.

Kata Kunci: daun sirsak, estrogen alfa, in-silico, kanker payudara, metabolit sekunder.

1. Introduction

The World Health Organization (WHO) defines cancer as a wide group of illnesses that can originate in nearly any organ or tissue when abnormal cells proliferate uncontrollably beyond their normal state and/or start spreading to other organs (metastatic).[1] According to estimated data covering 185 countries from the International Agency for Research on Cancer (IARC), there were close to 20 million new cancer diagnoses and 9.7 million cancer-related deaths in 2022,[2] with lung, breast, and colorectal cancer being the most commonly occurring cancer worldwide, followed by prostate and stomach cancer. Breast cancer was the most prevalent in women in most countries (157 of 185).[3] In Indonesia, breast cancer is also the most common type of cancer among women, accounting for 30.8% of all female cancer cases, leading to 20.4% of female cancer-related fatalities in 2020. By 2040, there is a prediction of a 47.1% and 62.1% increase in the incidence and mortality of breast cancer in Indonesia, respectively.[4]

Several treatments have been available for breast cancer, including chemotherapy using anticancer drugs (either alone or in combination with surgery), radiotherapy, hormone therapy, and targeted therapy.[5] For example, since the estrogen receptor (ER) pathway was identified nearly a century ago, tremendous efforts have been made to develop agents that target ER as a recognized therapeutic target for treating ER-positive (ER+) breast cancers. Selective estrogen receptor modulators (SERMs), such as tamoxifen, are employed as antagonists by blocking the binding of estrogens and obstructing ER signaling.[6-7] However, these therapies do not adequately cure metastatic breast cancer and have numerous drawbacks, including cardiotoxicity from anticancer agents,[8] incidences of drug resistance,[9] and other adverse effects.[10] Consequently, it is necessary to explore alternative strategies to address these limitations. Recent research has heightened interest in less toxic alternative or complementary medicine, particularly phytotherapy, where their secondary metabolites/ bioactive compounds have been identified to have various pharmacological effects.[11]

Soursop, *Annona muricata* Linn (*A. muricata*), a member of the Annonaceae family, is found in the tropical and subtropical regions globally, including the lowlands of Eastern and Western Africa, Southeast Asia, Australia, Mesoamerica, and the Caribbean.[12] It has been long been used in traditional medicine to treat such as bacterial infections, diabetes mellitus, hypertension, and cancer.[12-14] In vitro and in vivo studies have demonstrated that *A. muricata* possesses diverse biological activities, including antibacterial, antiviral, antioxidant,[15-16] anti-ulcer,[17] anti-diabetic,[18] anti-hypertensive,[19] wound healing,[20] and anti-cancer.[5,11, 4]

Various parts of *A. muricata*, including leaves and bark, have been utilized for therapeutic application. Over 200 chemical compounds have been identified and extracted, including phenolics, acetogenins, and alkaloids.[21] The leaves of *A. muricata* are among the most extensively studied, mostly due to their common usage in medicine, while fruits are considered exotic products.[13] Numerous chemical components have been isolated from the *A. Muricata*, such as chlorogenic acid, tannic acid[22] daidzein, kaempferol, emodin, quercetin, tangeretin, luteolin,[12] hexadecanoic acid, octadecanoic acid.[23]

These secondary metabolites were selected because they have been consistently identified in previous phytochemical investigations of *A. muricata* leaves and represent diverse chemical classes, including flavonoids, phenolic acids, fatty acids, anthraquinones, and tannins.[12-13,22-23] Therefore, they were chosen for subsequent in silico screening to identify potential ERα inhibitors.

Although *A. muricata* has been reported to exhibit anticancer activity against breast cancer cell lines, including MCF-7, MDA-MB-231, and 4T114, there is still a lack of information regarding the compounds and their interactions with the ERα at the molecular level. Therefore, this research aimed to utilize an in-silico approach to investigate the potential anti-cancer activity of several compounds from *A. muricata* leaves using Lipinski's RO5 and ADMET prediction, ligand-based pharmacophore modeling, and molecular docking methods against ERα.

2. Materials and Method

The in-silico experiments were carried out using a personal computer with the specifications of a B660M DS3H DDR4 Motherboard, Intel®CoreTM i3-12100F @3.3 GHz Processor, AMD Radeon RX 6600 8 GB Graphics card, 16.00 GB 2666 MHz RAM, with a Windows 10 Home 64-bit operating system.

2.1. Materials

The 2D structure of tested ligands was downloaded from PubChem and ChemSpider websites. ChemDraw 20.1.1 and Chem3D 20.1.1 were utilized to create 3D structures of these ligands. Ligand-based pharmacophore modeling was carried out using LigandScout 4.4.5 software, and molecular docking was performed using AutoDock 4.26 and AutoDock Tools 1.5.7, downloaded from The Scripps Research Institute (http://autodock.scripps.edu/downloads).

For control docking, the crystal structure of human estrogen receptor alpha ligand-binding domain (LBD) in complex with 4-hydroxytamoxifen (PDB ID: 3ERT) was downloaded from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank website (www.rcsb.org). The visualization of docking results was performed using BIOVIA Discovery Studio 2021 Client.

2.2. Methods

2.2.1. Lipinski's Rule of Five (RO5) and ADMET Prediction

The 3D structure of 10 tested ligands was uploaded to the http://www.swissadme.ch/ website to predict Lipinski's RO5 as a criterion to evaluate the drug-likeness property of tested ligands. Meanwhile, to predict the ADMET (Absorption, Distribution, Metabolism, and Excretion–Toxicity) properties, the ligands were uploaded to the PreADMET website https://preadmet.webservice.bmdrc.org/ with chosen features were Drug-likeness Prediction, ADME Prediction, Toxicity Prediction.

2.2.2. Ligand-based Pharmacophore Modeling

LigandScout 4.4.5 software was used to identify hit compounds with pharmacological activity with certain pharmacophore groups. An active/decoy database (https://dude.docking.org/targets) was created to increase the number of active compounds in the hit list and was used as the training set. Meanwhile, all tested ligands were saved as a Test set. Ligand-based used for pharmacophore modeling was ESR1 (PDB ID: 1sj0). The Receiver Operating Characteristic (ROC) curve was generated to validate the pharmacophore models.

2.2.3. Positive Control Docking

The downloaded PDB format of the 3ERT crystal structure was divided into two separate PDB files using Biovia Discovery Studio Client, one comprising all the atoms of the ERα, and the other comprising those of 4-hydroxytamoxifen as the native ligand. The structures were then protonated using AutoDock Tools 1.5.7. by adding Gasteiger charge to 4-hydroxytamoxifen and Kollmann charges to ERα. Both 4-hydroxytamoxifen and ERα were saved in PDBQT format. 4-hydroxytamoxifen was re-docked to NA with grid box dimensions 40 x 40 x 40, and coordinate x= 30.01, y= -1.913, z= 24.207. A default grid point of 0.375 Å was used in the study. The control docking procedure was performed using the following parameters: GA run = 100, maximum number of evaluations = 2,500,000, and population size = 150. Control docking results were analyzed based on their RMSD value.

2.2.4. Molecular Docking Simulations

The 3D structure of ten of *A. Muricata* L. compounds were loaded to the BIOVIA Discovery Studio Client. After removing the water molecules, the ligands were saved in ligan.pdb format. Tested ligands were then prepared for docking simulation using AutoDock Tools 1.5.7. All hydrogen atoms were added to the molecule. Then, the "Input" option on the "Ligand" menu in AutoDockTools was used to examine all the rotatable bonds in the ligand. Gasteiger partial charges were added, and non-polar hydrogens were merged. The ligands were then saved in PDBQT format, and the molecular docking simulation was carried out to screen the potential ligands using validated parameters from control docking. The visualization of docking simulations was performed using BIOVIA Discovery Studio Client, and the results were sorted based on the free energy of binding on the most populated cluster and their intermolecular interactions with ER.

3. Result

The following are the results obtained from Lipinski's RO5 Prediction, ADMET Prediction, and Ligand-based Pharmacophore Modeling.

Molecular weight, number of hydrogen bond donors and acceptors, and LogP value of 10 compounds were calculated, and the data are shown in Table 1. All compounds, except tannic acid, exhibit molecular weight below 500 g/mol, which corresponds to the favorable criteria as per Lipinski's RO5. Among the 10 compounds, tannic acid does not meet Lipinski's RO5 criteria, with 11 hydrogen bond donors and 18 acceptors. Of the 10 compounds, octadecanoic acid is the only compound with a Log P > 5.

The ADMET prediction results for the 10 compounds are shown in Table 2, including HIA, Caco-2 permeability, PPB, BBB, mutagenicity, and carcinogenicity. Most compounds demonstrated favorable absorption profiles, with 7 compounds categorized as having high absorption and 3 moderate absorption compounds. Emodin exhibited high Caco-2 permeability, while the remaining compounds showed medium to low permeability. Six compounds showed PPB values above 90%, indicating strong plasma protein binding. Additionally, seven compounds were predicted to be unable to cross the blood–brain barrier. In terms of toxicity, octadecanoic acid, hexadecanoic acid, and tannic acid were predicted as non-mutagenic.

The validation of pharmacophore model shown in

Table 1. Lipinski's Rule of Five Predictions for Drug-likeness

No	Compound's Name	Molecular Weight (g/mol)	Log P	Hydrogen bonds		Lipinski Violation
				Donor	Acceptor	
1	Chlorogenic acid	354.31	-0.65	6	9	No
2	Quercetin	302.26	1.98	5	7	No
3	Daidzein	254.24	2.87	2	4	No
4	Kaempherol	286.23	2.28	4	6	No
5	octadecanoic acid	284.48	6.33	1	2	No
6	Hexadecanoic acid	256.42	5.55	1	2	No
7	Luteolin	286.24	2.28	4	6	No
8	Emodin	270.24	1.89	3	5	No
9	Tangeretin	372.37	3.50	0	7	No
10	Tannic acid	636.47	-0.28	11	18	Yes

Table 3. A total of 10 pharmacophore models were successfully generated. All models exhibited AUC values greater than 0.5, with the highest reaching 0.94, indicating good capability in distinguishing active compounds from inactive ones. Model 1 showed the best performance, with AUC100% and fit score values of 0.94 and 0.7718.

Overall, the pharmacophore models generated in this study demonstrated acceptable to good screening performance, as shown in Table 3. Among the evaluated pharmacophore models, Models 5 and 6 demonstrated the highest enrichment capability, with EF values of 9.568 and 9.401 and GH scores of 0.510 and 0.504, respectively. Although both models exhibited comparable screening performance, Model 5 achieved the highest EF and GH values, indicating superior enrichment of active compounds relative to random selection and the best overall virtual screening performance. Therefore, Model 5 was selected for subsequent virtual screening analysis. Scores for the pharmacophore-fit, as shown in Table 3, ranged from 46.97 to 47.17, indicating similar compatibility of all compounds with the pharmacophore model. Chlorogenic acid, luteolin, and quercetin showed the highest fit score (47.17), while daidzein showed the lowest score (46.97).

In the screening results, pharmacophore Model 5 demonstrated the best overall virtual screening performance based on the enrichment factor (EF) and Goodness of Hit (GH) score, with values of 9.57 and 0.5108, respectively. These parameters are considered more representative of pharmacophore screening performance because they evaluate not only the ability of the model to retrieve active compounds but also its selectivity in minimizing false-positive hits.[24-26] Although Model 1 exhibited the highest AUC value (0.94), indicating excellent discrimination between active compounds and decoys,[27] it retrieved a considerably larger number of false-positive compounds, resulting in lower screening selectivity. Therefore, Model 5 was

Table 2. ADMET Prediction Results

No	Compound's name	Absorption		Distribution		Toxicity	
		HIA (%)	Caco-2 (nm/sec)	PPB (%)	BBB	Muta-genicity	Carcino-genicity
1	Chlorogenic acid	20.43	18.72	41.96	0.03	Mutagenic	Mouse (+) Rat (-)
2	Quercetin	63.49	3.42	93.24	0.17	Mutagenic	Mouse (-) Rat (+)
3	Daidzein	92.65	7.72	88.70	0.19	Mutagenic	Mouse (-) Rat (+)
4	Kaempherol	79.44	9.58	89.61	0.29	Mutagenic	Mouse (-) Rat (+)
5	Octadecanoic acid	98.45	28.29	100.00	10.93	Non - mutagenic	Mouse (-) Rat (+)
6	Hexadecanoic acid	98.29	26.07	100.00	8.22	Non - mutagenic	Mouse (-) Rat (+)
7	Luteolin	79.43	4.54	99.72	0.37	Mutagenic	Mouse (-) Rat (+)
8	Emodin	90.43	90.43	100.00	0.69	Mutagenic	Mouse (-) Rat (+)
9	Tangeretin	96.81	36.48	86.66	0.014	Mutagenic	Mouse (-) Rat (+)
10	Tannic acid	0.00	11.40	100.00	0.03	Non - mutagenic	Mouse (+) Rat (+)

HIA: Human Intestinal Absorption
Caco-2: Cancer Coli-2
PPB: Plasma Protein Binding
BBB: Blood Brain Barrier

Table 3. Pharmacophore model validation and fit scores of selected compounds

Model	Fit Score	AUC	GH	Precision	Recall	F1	Compound	Fit Score
1	0.7718	0.94	0.2397	0.0420	0.92	0.0803		
2	0.7629	0.88	0.4162	0.0363	0.79	0.0693		
3	0.7558	0.90	0.4172	0.0376	0.82	0.0719		
4	0.7487	0.92	0.2900	0.0402	0.88	0.0769		
5	0.7458	0.88	0.5108	0.0354	0.77	0.0676	Daidzein	46.97
6	0.7453	0.88	0.5044	0.0354	0.77	0.0676	Chlorogenic acid Luteolin	47.17 47.17
7	0.7416	0.94	0.3797	0.0411	0.90	0.0786	Quercetin	47.17
8	0.7377	0.93	0.3639	0.0407	0.89	0.0778		
9	0.7351	0.94	0.3670	0.0415	0.91	0.0794		
10	0.7341	0.92	0.3785	0.0398	0.87	0.0761		

considered the most appropriate pharmacophore model for subsequent virtual screening analysis.

Based on the pharmacophore validation results, each model exhibits a distinct Enrichment Factor (EF) value. The EF value represents a model's ability to enrich active compounds compared to random selection. A higher EF value indicates a superior ability to distinguish between active compounds and decoys during the virtual screening process. The EF score is calculated using the following equation:

$$EF = \frac{Ha \times D}{Ht\ x\ A}$$

An EF value of 1 indicates random performance, whereas EF values greater than 1 indicate enrichment beyond random selection.[24-26]

The results show that Model 5 achieved the highest performance with an EF of 9.57, followed by Model 6 at 9.40. This indicates that both models are approximately 9 times more effective at identifying active compounds than random selection. Conversely, Model 1 yielded the lowest EF value of 2.86, reflecting a relatively lower enrichment capability compared to the other models.

Model 2 and Model 3 also demonstrated strong performance with EF values of 7.06 and 6.94, respectively. Meanwhile, Models 7, 8, 9, and 10 showed EF values ranging from 5.38 to 5.80, indicating moderate enrichment capabilities. Model 4 recorded an EF of 3.83, placing its performance below most other models.

Notably, all models maintained an EF value greater than 1.0, signifying that every model performed better than random selection. A high EF suggests that the model can produce a higher proportion of active compounds in the screening results compared to their initial distribution in the database.[25] Consequently, models with the highest EF values are considered more selective and effective for virtual screening.[26]

The goodness of hit score (GH) reflects how effectively a model can recognize and retrieve active compounds. A higher GH score indicates better model selectivity and screening efficiency. The GH score is calculated using the following equation[27]:

$$GH = \left(\frac{Ha(3A + Ht)}{4HtA}\right) \times \left(1 - \frac{Ht - Ha}{D - A}\right)$$

The GH score ranges from 0 to 1, in which a value of 0 represents a null model with no predictive capability, whereas a value approaching 1 indicates an ideal pharmacophore model with excellent discriminative capability.[24] Based on the pharmacophore models developed according to the latest literature-derived approach, the GH scores obtained ranged from 0.2397 to 0.5108, indicating differences in screening efficiency among the generated models.

Among all generated pharmacophore hypotheses, Model 5 exhibited the highest GH score (0.5108), indicating superior capability in distinguishing active compounds from inactive compounds during virtual screening. This result was further supported by its high percentage yield of actives (45.56%), the highest enrichment factor (9.57), and a relatively low number of false positives (92 compounds). Therefore, Model 5 demonstrated the best balance between sensitivity and selectivity, making it the most reliable pharmacophore model for screening purposes.

Similarly, Model 6 also demonstrated strong screening performance with a GH score of 0.5044, suggesting that this model possesses a robust discriminative ability toward active ligands. The relatively high GH values observed for Models 5 and 6 indicate that both pharmacophore hypotheses are sufficiently selective and capable of enriching active compounds from large chemical databases. Such characteristics are

essential in virtual screening studies because efficient pharmacophore models can significantly reduce the number of false-positive hits and improve hit identification efficiency.[28]

In contrast, Model 1 showed the lowest GH score (0.2397), primarily due to the excessive number of retrieved hits (676 compounds) relative to the number of active hits (92 compounds). This result indicates poor selectivity and a high false-positive rate, suggesting that the model was less efficient in differentiating active molecules from inactive molecules. Consequently, the screening performance of Model 1 can be considered inadequate for reliable virtual screening applications.

Precision measures the proportion of predicted positive results that are truly positive and is calculated as the ratio of true positives (TP) to the sum of true positives and false positives (FP). A high precision value indicates that the model produces fewer false-positive predictions and is therefore more reliable in identifying relevant positive instances. In this study, precision was used to evaluate the selectivity of the screening model toward active compounds.[29-30]

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall represents the proportion of actual positive instances that are successfully identified by the model and is calculated as the ratio of true positives (TP) to the sum of true positives and false negatives (FN). A high recall value indicates that the model is capable of detecting most active compounds while minimizing false negatives. In virtual screening studies, recall is particularly important because failing to identify potential active compounds may result in the loss of promising candidates for further development.[30-31]

$$\text{Recall} = \frac{TP}{TP + FN}$$

Based on the precision and recall parameters, Model 5 demonstrated a more balanced screening performance than the other pharmacophore models. Although it did not achieve the highest recall value, Model 5 maintained an appropriate balance between retrieving active compounds and minimizing false-positive predictions, which is essential for effective virtual screening. Precision measures the proportion of predicted positive results that are truly active, whereas recall reflects the proportion of actual active compounds successfully identified by the model.[29-31] The balanced precision and recall achieved by Model 5 are consistent with its superior enrichment factor (EF) and Goodness of Hit (GH) score, indicating that the model effectively enriches active compounds while maintaining good screening selectivity.

Although the precision values remained relatively low, this outcome is commonly observed in pharmacophore-based virtual screening because inactive compounds generally outnumber active compounds within large chemical databases.[29-31] In such applications, the primary objective is to maximize the enrichment of active compounds while keeping the number of false-positive hits at an acceptable level rather than achieving perfect classification accuracy. Therefore, the balanced precision and recall demonstrated by Model 5, together with its highest EF and GH values, support its selection as the optimal pharmacophore model for subsequent virtual screening analysis.

F1 score was used as an integrated performance metric to evaluate the balance between precision and recall in the generated pharmacophore models. The F1 score is calculated using the following equation:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

A higher F1 score indicates a better balance between the ability of the model to correctly identify active compounds and its capability to minimize incorrect positive predictions.[32,33,34] Although the F1 score provides a useful measure of the balance between precision and recall, the selection of the optimal pharmacophore model in this study was primarily based on the Enrichment Factor (EF) and Goodness of Hit (GH) score. EF is one of the most widely used validation metrics in virtual screening because it evaluates the ability of a pharmacophore model to enrich active compounds within the early-ranked fraction of a screening database relative to random selection, thereby reflecting the model's early recognition capability. Furthermore, the GH score is considered a critical metric for pharmacophore model validation, as it comprehensively assesses model reliability by integrating the recovery of active compounds, hit-list quality, and the proportion of false-positive hits. Consequently, EF and GH provide a more representative assessment of pharmacophore screening performance and were therefore used as the primary criteria for selecting the optimal pharmacophore model.[35-36] Consequently, Model 5 was selected as the optimal pharmacophore model for subsequent virtual screening analysis.

As shown in Figure 1, the pharmacophore model for ERα binding comprises four key features, including hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), hydrophobic regions, and aromatic ring (AR) features, all of which are well accommodated by the identified hit compound.

A molecular docking simulation was carried out to

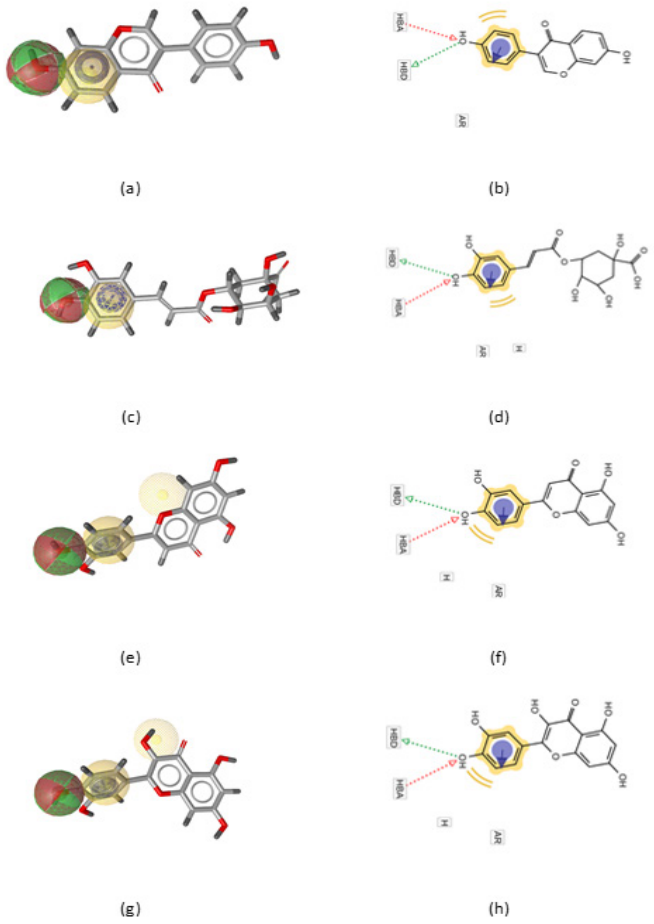

Figure 1. The 3D and 2D pharmacophore visualization of daidzein (a) & (b), Chlorogenic acid (c) & (d), Luteolin (e) & (f), Quercetin (g) & (h)

predict the interaction of the target receptor with its ligand. It comprises two components: a search algorithm to identify the potential ligand conformations at the binding site and a scoring function to estimate the ligand's binding affinity to the protein.

A molecular docking simulation was carried out to predict the interaction of the target receptor with its ligand. It comprises two components: a search algorithm to identify the potential ligand conformations at the binding site and a scoring function to estimate the ligand's binding affinity to the protein.

For the first step, control docking was performed to validate docking parameters. The process involves selecting a crystal structure of a target macromolecule complexed with a known ligand, separating the ligand from the protein, and docking it back to its "induced-fit" form. In this work, the crystal structure of the human estrogen receptor alpha ligand-binding domain in a complex with 4-hydroxytamoxifen (PDB ID: 3ERT) was selected. The RMSD value serves as a parameter for the validity of control docking, with a successful control docking giving the RMSD value less than 2 Å. The result showed that the structural deviation, as measured by RMSD value, was 1.26 Å, which suggests that the docking parameters were acceptable and can be used further for docking simulation with tested ligands.

A molecular docking simulation was carried out to screen 10 tested ligands. The free energy binding (FEB) and inhibition constant (Ki) were recorded, and the interactions between ligands with amino acid residues in the receptor binding site were identified. Table 5 demonstrates the molecular docking results of 10 ligands and tamoxifen as control ligands.

The results showed that the free energy of binding of the tested compounds varies from -4.65 kcal/mol to -8.59 kcal/mol. These values are still above the binding free energy of Tamoxifen as a controlled ligand (-11.96 kcal/mol). We analyzed and visualized the docked conformation to investigate intermolecular bonding, such as hydrogen bonds, Van der Walls, and pi interactions (Table 4).

Daidzein, emodin, and kaempferol were the three compounds with the lowest binding energy: -8.59,

Table 4. Molecular docking results for the tested ligands and their interaction with receptor binding site

No.	Compound's name	FEB (kcal/mol)	Ki	Amino acids interaction		
				H-bonds	Van der Waals	Others
1	Tamoxifen	-11.96	1.71 nM	GLU353	THR347	Alkyl dan PI-Alkyl: MET421; LEU525; ALA350; LEU387; LEU391; PHE404 Amide-Pi Stacked: LEU346 Pi-Sulfur: MET343
2	Chlorogenic acid	-7.14	5.85 uM	LEU346; GLU419; GLU353; ARG394; LEU387	-	Unfavorable Donor-Donor and Unfavorable Acceptor-Acceptor: ALA350 Pi-Sigma: MET343 Alkyl&Pi-Alkyl: LEU525 Pi-Sulfur: MET421
3	Quercetin	-7.14	5.86 uM	ARG394; GLU353; THR347; MET388	-	Pi-Alkyl: LEU346; LEU387; LEU391; LEU525; ALA350; MET343 Pi-Sigma: LEU384
4	Daidzein	-8.59	503.27 nM	GLY521; HIS524; GLU353	GLY420; ILE424; MET343; LEU346; LEU349; LEU384; MET388; ARG394	Pi-Alkyl: LEU387; ALA350; LEU391
5	Kaempferol	-7.71	2.22 uM	ARG A:394 GLY A:521 HIS A:524 LEU A:387 GLU A:353 MET A:388	-	Pi-Alkyl: LEU384; ALA350; LEU346
6	Octadecanoic acid	-5.24	143.23 uM	-	-	Alkyl: ALA350; LE346; LEU525; LEU387; MET388; LEU391
7	Hexadecanoic acid	-5.05	200.4 uM	-	-	Alkyl: LEU387; LEU391; LEU349; MET388; MET421; MET343; LEU346; ALA350
8	Luteolin	-7.35	4.12 uM	PHE A:404 GLU A:419	-	Alkyl and Pi-Alkyl: LEU346; LEU391; MET A421
9	Emodin	-7.83	1.81 uM	LEU A:346 GLU A:353 ARG A:394	-	Alkyl and Pi-Alkyl: LEU391; ALA350; LEU387; ILE424; MET388 Pi-Sigma: LEU384
10	Tangeretin	-7.25	4.88 uM	LEU A:346 GLU A:353	-	Alkyl and Pi-Alkyl: ALA350; LEU349; LEU387; PHE404; LEU391; LEU428; ILE424; LEU525 Pi-Sulfur: MET421

-7,83, and -7.71, respectively. Two-dimensional (2D) structures of the selected compounds showed that most of the compounds possessed hydroxyl functional groups Figure 2. As observed in the 3D visualization of docking results, these groups were found to interact via hydrogen bond interactions with several amino acids in the ER binding site, such as GLU353. Pi interactions in the form of Pi-Alkyl and Pi-sigma interactions have also been observed in the best docking pose. Most compounds contain aromatic rings, likely to form Pi interaction with amino acid residues in the ER binding site, such as ALA350, LEU346, LEU387, and LEU391.

Aside from free binding energy, inhibition constants (Ki) were also generated. Ki is a value that quantifies the strength of an inhibitor in blocking the activity of an enzyme. The smaller the Ki, the greater the binding affinity and the smaller the ligand needed to inhibit its binding partner's activity. The result showed that daidzein has the lowest Ki value of 503,27 nM, followed

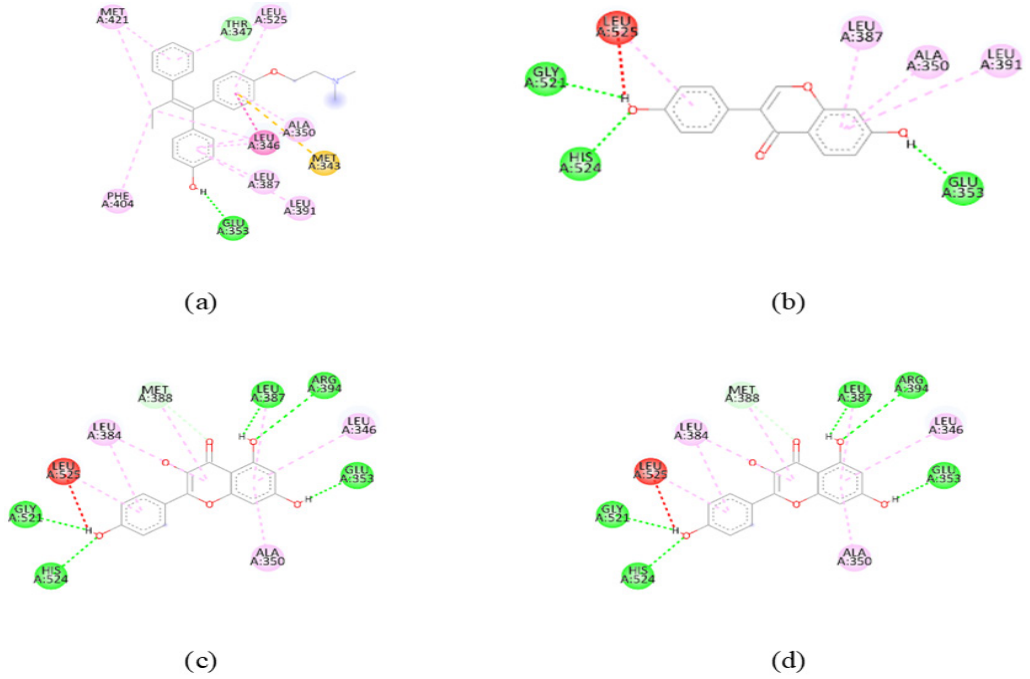

Figure 2. The 2D docking visualization of tamoxifen (a), daidzein (b), kaempferol (c) & and emodin (d). The hydrogen bond interactions were presented in dashed green lines, and the Pi interaction in dashed pink and purple lines

by emodin and kaempferol, with values of 1,81 uM and 2,22 uM, respectively. Although daidzein demonstrated a substantially higher Ki than the reference ligand (1.7 nM), this difference is expected because the reference compound has been structurally optimized to achieve high-affinity binding to the target receptor, whereas daidzein is a naturally occurring phytochemical that has not undergone medicinal chemistry optimization for this specific target. Furthermore, differences in molecular structure, functional groups, and ligand-receptor interactions may contribute to the observed disparity in binding affinity. It should also be noted that the Ki values reported in this study were estimated from molecular docking calculations and therefore represent predicted binding affinity rather than experimentally determined inhibitory potency. Nevertheless, the sub-micromolar Ki of daidzein indicates favorable binding to the target receptor, supporting its potential as a promising lead compound for further optimization and experimental validation.

The interaction between ligands and amino acid residues in the ERα ligand-binding domain (LBD) is a crucial aspect that determines the binding affinity and agonist/antagonist effects of compounds on the receptor. The ERα LBD structure contains a large and flexible binding site, but there are several conservative polar residues that are consistently involved in the formation of important hydrogen bonds, especially GLU353 and HIS524, which act as key anchor points for the ligand through the hydroxyl group on the phenolic ring of ligands such as estradiol or its analogs. GLU353 forms a strong hydrogen bond with phenolic group of the A-ring of the ligand, while HIS524 is involved in binding with the hydroxyl group at the other end of the ligand, which contributes to the orientation and stability of the LBD complex and influences the conformational changes of the H12 helix that are important in receptor activation or inactivation. Structural analysis of various ERα crystal structures also shows that in addition to the extensive hydrophobic energy contribution, hydrogen bonds with GLU353 and HIS524 appear consistently in many agonist and antagonist ligands, indicating a role for these residues in predicting the binding affinity and agonist/antagonist character of a compound.[37-38]

4. Discussion

Numerous efforts have been conducted to identify the general criteria for differentiating drug-like from nondrug-like molecules within compound databases or new drug candidates. Lipinski's RO5 has been widely utilized to help identify promising drug candidates based on their pharmacokinetic profiles, as proposed by Christopher A. Lipinski, a medicinal chemist.[35] Lipinski's approach is among the simplest and most frequently used methods for predicting permeation or absorption, relying on assessing specific physical properties of molecules. A general "rule of thumb" for the evaluation of drug-like properties predicts that poor absorption or permeation is more likely when a molecule has a molecular weight greater than 500, more than 5 H-bond donors, and 10 H-bond acceptors, and the Log P is greater than 5.39 Based on the rule,

an orally active drug should have no more than one violation of the aforementioned criteria.[40]

Hydrogen bond donors and acceptors are associated with the ability to form intermolecular interactions. The more hydrogen bonds may lead to lower bioavailability and absorption due to elevated polarity. Increased H-bond interaction might also make the drug molecules randomly bind to any receptor associated with unfavorable side effects. Thus, Lipinski's RO5 limits the number of hydrogen bonds. The LogP serves as a critical measure in assessing oral bioavailability and absorption. The higher the LogP value of a molecule, the better its ability to cross the lipid bilayer. A highly negative log P value is undesirable as it indicates the molecule's inability to cross the membrane. However, the limit of Log P is necessary as a compound with high lipophilicity will be stored in adipose tissues, resulting in an inability to sustain its plasma concentrations. It will also affect therapeutic efficacy and drug excretion, increasing systemic toxicity.[41]

Based on the results, 9 out of 10 compounds were found to be in full compliance with Lipinski's RO5; one molecule (octadecanoic acid) could still be acceptable, although it violated one rule, and another molecule (tannic acid) failed to pass for violating more than one rule. Thus, tannic acid was not used in molecular docking simulation.

The % HIA parameter and the permeability of the Caco-2 monolayer are among the descriptors used to predict drug absorption. With the growing risk and complexity of drug discovery processes, HIA prediction has become more significant. Previous studies have demonstrated the relationship between oral bioavailability and intestinal absorption, concluding that the intestinal absorption process primarily controls the bioavailability of most compounds (64%).[42] HIA, which refers to the uptake of substances in the human intestine, is categorized into three groups based on percentage: weak (0–20%), moderate (20–70%), and great absorption (70–100%). Of the 10 compounds tested, it was found that 7 compounds had great absorption, 3 had moderate, and one had poor absorption.

The Caco-2 monolayer permeability is utilized to evaluate drug permeability and measure its passage across epithelial cells in the intestine. This parameter indicating penetration through the gut barrier and high intestinal absorption are important for understanding initial bioavailability but do not cover aspects of metabolism by phase I/II enzymes or potential organ-specific side effects such as hepatotoxicity or cardiotoxicity.[43] Caco-2 cell permeability is divided into three categories: high (>70 nm/sec), medium (4-70 nm/sec) and low (4 nm/sec).[44] The result showed that emodin exhibits high Caco-2 permeability, whereas other compounds demonstrated medium to low permeability.

Drug distribution prediction utilizes the parameters of PPB and BBB. Drugs enter the blood circulation to target specific tissues. As only free, unbound drugs can diffuse between capillaries and tissues, PPB can significantly affect distribution. Drugs with a high PPB affinity (>90% bound) generally have a delayed start and longer duration of effect when compared to those with low binding affinity.[45-46] Results in Table 2 showed that six tested compounds have %PPB over 90%. The BBB is a specialized capillary network that protects the brain from the circulatory system. It can shield the brain from infections, including bacteria and viruses. The BBB permeability index determines the capacity of a compound to cross the BBB. Compounds with a BBB value of 0.3 or higher can cross the BBB and potentially cause CNS toxicity.[47] Of the 10 compounds evaluated, seven were unable to cross the BBB.

Toxicity prediction was carried out using parameters related to mutagenicity (causing mutations in the DNA of the test organism) and carcinogenicity (having the potential to cause cancer). Octadecanoic acid, hexadecanoic acid, and tannic acid showed negative results for mutagenicity. The predictions provide an initial overview of the pharmacokinetic and toxicological profiles of the compounds, but the results are estimates and not a substitute for laboratory testing.[48-49]

Using ligand-based methods, pharmacophore models can be developed for virtual screening or lead optimization when the receptor's three-dimensional structure is limited but sufficient active ligands are available. The Receiver Operating Characteristic Curve (ROC) is a way to compare screening results to sets of active and inactive molecules. The rate of getting true positives (actives) set against the rate of getting false positives makes up an ROC curve. The ROC curve's Y-coordinate shows the true-positive rate and the number of retrieved actives. Its X-coordinate shows the false-positive rate, the number of retrieved inactive or decoys. In a perfect situation, the curve would go up vertically along the Y-axis until it reached the highest true positive rate, which is one.[50]

The best-performing model was selected based on the highest ROC curve quality for each model, using the parameters area under the curve (AUC100%) and enrichment factor (EF). The AUC value ranges from 0.5 to 1.0, where 1.0 indicates perfect discrimination between active and inactive compounds.[51] Meanwhile, the EF value quantifies the proportion of active compounds identified when analyzing a specific

fraction (nX%) of the entire ranked compound library. The enrichment factor at 1% (EF1%) measures the number of active ligands successfully retrieved in the top 1% of the active/decoy database compared to the number of active compounds expected to be found randomly. A high EF value in the first 1% indicates that the scoring function is more effective in identifying true active compounds at an early stage.[49]

In virtual screening studies, the F1 score is commonly used as an integrated performance metric because the datasets are generally imbalanced, with inactive or decoy compounds substantially outnumbering active compounds. By combining precision and recall into a single value, the F1 score provides a balanced assessment of a model's ability to correctly identify active compounds while minimizing false-positive predictions. A higher F1 score indicates a better balance between sensitivity and precision; however, it should be interpreted alongside other validation metrics, particularly in pharmacophore-based virtual screening.

The F1 scores obtained in this study indicate variations in the balance between precision and recall among the generated pharmacophore models. While some models exhibited relatively higher F1 values due to greater sensitivity in retrieving active compounds, others demonstrated lower F1 values as a consequence of more stringent screening selectivity. Because the F1 score primarily reflects classification performance, it should not be used as the sole criterion for selecting the optimal pharmacophore model. Instead, comprehensive evaluation using enrichment-based metrics, such as the Enrichment Factor (EF) and Goodness of Hit (GH) score, provides a more appropriate assessment of pharmacophore screening performance by considering both active-compound enrichment and false-positive reduction.

The pharmacophore-fit score reflects the degree of compatibility between the chemical features of a compound and the receptor binding requirements. A compound is generally considered active toward its receptor when it exhibits a pharmacophore-fit score greater than 50%. Furthermore, a higher pharmacophore-fit score indicates a better correspondence with the pharmacophore model and suggests that the compound may possess stronger biological activity, thereby increasing its potential to be identified as an active hit molecule.[34,52]

The results suggest that the compounds possess partial compatibility with the essential pharmacophore features required for receptor interaction. These findings indicate that the compounds may still pass the initial virtual screening stage and can be further evaluated through molecular docking studies to assess their binding interactions with the target receptor. However, the compounds cannot yet be considered as definitive candidates without further validation of their binding affinity and biological activity.

Among the generated pharmacophore models, Model 5 was selected for virtual screening because it showed the highest enrichment factor (9.57) and GH score (0.51), together with the highest percentage yield of active compounds (45.56%) and fewer false positives hits than the other models. Although Model 1 produced a higher AUC value and recall, it also retrieved a much larger number of false positives compounds, indicating lower screening selectivity. In pharmacophore-based virtual screening, enriching active compounds while minimizing unnecessary hits is more important than simply maximizing the number of retrieved compounds. Therefore, the higher enrichment capability and better selectivity of Model 5 make it more appropriate for identifying potential hit compounds prior to molecular docking.

Virtual screening using Model 5 identified 4 compounds, namely chlorogenic acid, luteolin, quercetin, and daidzein, that fulfilled the required pharmacophore features for ERα binding. Although these compounds exhibited very similar pharmacophore fit scores, their molecular docking results showed different binding affinities toward ERα. This suggests that sharing similar pharmacophore features does not necessarily result in similar binding strength, as ligand orientation within the binding pocket and the interactions formed with surrounding amino acid residues also influence binding affinity. These findings indicate that pharmacophore modeling is useful for narrowing the number of candidate compounds, whereas molecular docking provides a more detailed evaluation of ligand receptor interactions.[53]

Among the four screened compounds, daidzein showed the greatest potential for further development as an ERα inhibitor. Besides fulfilling Lipinski's Rule of Five and showing a favorable predicted pharmacokinetic profile, daidzein produced the lowest binding free energy (-8.59 kcal/mol) and the lowest inhibition constant (503.27 nM) among all screened compounds. Although its binding affinity was still lower than that of tamoxifen, these findings suggest that daidzein has the strongest predicted interaction with ERα among the secondary metabolites evaluated in this study. Therefore, daidzein may be considered a promising lead compound for further optimization and experimental validation.

Analysis of the docking results showed that several conserved amino acid residues play important roles in

ligand binding within the ERα ligand binding domain. Among these residues, GLU353 was consistently involved in hydrogen bond formation with most of the active compounds, including daidzein, quercetin, kaempferol, emodin, chlorogenic acid, and tangeretin, indicating its important role in stabilizing ligand binding. HIS524 also interacted with daidzein and kaempferol, whereas ARG394, LEU346, LEU387, and LEU391 mainly contributed through hydrogen bond and hydrophobic interactions. Previous structural studies have identified GLU353 and HIS524 as key anchoring residues responsible for maintaining ligand orientation and stabilizing the ERα ligand complex.[38,55] The interaction pattern observed for daidzein is consistent with these findings and may explain its better docking performance than the other screened compounds.

The results of Lipinski's Rule of Five and ADMET prediction also supported the selection of potential compounds before molecular docking. Lipinski's Rule of Five excluded compounds with less favorable drug-like properties, such as tannic acid, which violated more than one criterion and may therefore have poor oral bioavailability.[41-43] ADMET prediction further provided an initial overview of the absorption, distribution, and toxicity profiles of each compound. Although these predictions cannot replace experimental evaluation, combining drug-likeness assessment, pharmacokinetic prediction, pharmacophore screening, and molecular docking provides a more comprehensive basis for selecting compounds with greater potential for further drug development.

This study has several limitations. First, all results were obtained from computational methods and therefore remain predictive. Molecular docking evaluates ligand receptor interactions under static conditions and does not fully consider protein flexibility, solvent effects, or conformational changes that may occur under physiological conditions.[55] In addition, the ADMET and toxicity profiles were generated from computational prediction models and still require experimental confirmation. Therefore, further studies involving molecular dynamics simulations, in vitro assays using ERα positive breast cancer cell lines, and in vivo studies are needed to confirm the binding stability, inhibitory activity, pharmacokinetic properties, and safety of daidzein as a potential ERα inhibitor.

5. Conclusion

This study demonstrated that an integrated in silico approach combining Lipinski's Rule of Five, ADMET prediction, ligand-based pharmacophore modeling, pharmacophore-based virtual screening, and molecular docking is useful for identifying potential estrogen receptor alpha (ERα) inhibitors from secondary metabolites of Annona muricata leaves. Most compounds exhibited favorable drug-like and predicted pharmacokinetic properties, whereas tannic acid was excluded because it violated more than one Lipinski criterion. Among the validated pharmacophore models, Model 5 showed the best virtual screening performance, achieving the highest enrichment factor (9.57) and Goodness of Hit score (0.5108). Virtual screening identified four candidate compounds: chlorogenic acid, luteolin, quercetin, and daidzein. Molecular docking demonstrated that daidzein exhibited the strongest binding affinity toward ERα, with a binding free energy of −8.59 kcal/mol, a predicted inhibition constant of 503.27 nM, and favorable interactions with the key ERα residues GLU353 and HIS524. Overall, these findings suggest that daidzein is the most promising secondary metabolite from *A. muricata* leaves as a potential ERα inhibitor and may serve as a lead compound for further ERα-targeted drug development. However, further molecular dynamics simulations and experimental validation through in vitro and in vivo studies are required to confirm its binding stability, biological activity, pharmacokinetic properties, and safety.

Conflict of Interest

The authors declare no conflicts of interest

References

1. Health Topics: Cancer. Available online: https://www.who.int/health-topics/cancer#tab=tab_1 (accessed on 2024).
2. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R.L.; Soerjomataram, I.; et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2024, 74, 229–263.
3. Global Cancer Burden Growing, Amidst Mounting Need for Services. Available online: https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services (accessed on 2024).
4. Ferlay, J.; Ervik, M.; Lam, F.; Colombet, M.; Mery, L.; Piñeros, M.; et al. Global Cancer Observatory: Cancer Today; International Agency for Research on Cancer: Lyon, France, 2020.
5. Silihe, K.K.; Mbou, W.D.; Ngo Pambe, J.C.; Kenmogne, L.V.; Maptouom, L.F.; Kemegne Sipping, M.T.; et al. Comparative anticancer effects of Annona muricata Linn (Annonaceae) leaves and fruits on DMBA-induced breast cancer in female rats. BMC Complementary Medicine and Therapies. 2023, 23, 234.
6. Haines, C.N.; Wardell, S.E.; McDonnell, D.P. Current and emerging estrogen receptor-targeted therapies for the treatment of breast cancer. Essays in Biochemistry. 2021, 65, 985–1001.
7. Rej, R.K.; Thomas, J.E., II; Acharyya, R.K.; Rae, J.M.; Wang, S. Targeting the estrogen receptor for

the treatment of breast cancer: Recent advances and challenges. Journal of Medicinal Chemistry. 2023, 66, 8339–8381.
8. Kumar, A.; Gautam, B.; Dubey, C.; Tripathi, P.K. A review: Role of doxorubicin in treatment of cancer. International Journal of Pharmaceutical Sciences and Research. 2014, 5, 4105.
9. Jayaraj, R.; Nayagam, S.G.; Kar, A.; Sathyakumar, S.; Mohammed, H.; Smiti, M.; et al. Clinical theragnostic relationship between drug-resistance specific miRNA expressions, chemotherapeutic resistance, and sensitivity in breast cancer: A systematic review and meta-analysis. Cells. 2019, 8.
10. Nurgali, K.; Jagoe, R.T.; Abalo, R. Editorial: Adverse effects of cancer chemotherapy: Anything new to improve tolerance and reduce sequelae? Frontiers in Pharmacology. 2018, 9, 245.
11. Mbaveng, A.T.; Kuete, V.; Efferth, T. Potential of Central, Eastern and Western Africa medicinal plants for cancer therapy: Spotlight on resistant cells and molecular targets. Frontiers in Pharmacology. 2017, 8, 343.
12. Zubaidi, S.N.; Nani, H.M.; Kamal, M.S.A.; Qayyum, T.A.; Maarof, S.; Afzan, A.; et al. Annona muricata: Comprehensive review on the ethnomedicinal, phytochemistry, and pharmacological aspects focusing on antidiabetic properties. Life. 2023, 13, 353.
13. Moghadamtousi, S.Z.; Fadaeinasab, M.; Nikzad, S.; Mohan, G.; Ali, H.M.; Kadir, H.A. Annona muricata (Annonaceae): A review of its traditional uses, isolated acetogenins and biological activities. International Journal of Molecular Sciences. 2015, 16, 15625–15658.
14. Najmuddin, S.U.F.S.; Romli, M.F.; Hamid, M.; Alitheen, N.B.; Rahman, N.M.A.N.A. Anti-cancer effect of Annona muricata Linn leaves crude extract (AMCE) on breast cancer cell line. BMC Complementary and Alternative Medicine. 2016, 16, 311.
15. Pinto, N.C.C.; Campos, L.M.; Evangelista, A.C.S.; Lemos, A.S.; Silva, T.P.; Melo, R.C.; et al. Antimicrobial Annona muricata L. (soursop) extract targets the cell membranes of Gram-positive and Gram-negative bacteria. Industrial Crops and Products. 2017, 107, 332–340.
16. Balderrama-Carmona, A.P.; Silva-Beltrán, N.P.; Gálvez-Ruiz, J.C.; Ruíz-Cruz, S.; Chaidez-Quiroz, C.; Morán-Palacio, E.F. Antiviral, antioxidant, and antihemolytic effect of Annona muricata L. leaves extracts. Plants. 2020, 9.
17. Bento, E.B.; Júnior, F.E.B.; de Oliveira, D.R.; Fernandes, C.N.; de Araújo Delmondes, G.; Cesário, F.; et al. Antiulcerogenic activity of the hydroalcoholic extract of leaves of Annona muricata Linnaeus in mice. Saudi Journal of Biological Sciences. 2018, 25, 609–621.
18. Agu, K.C.; Eluehike, N.; Ofeimun, R.O.; Abile, D.; Ideho, G.; Ogedengbe, M.O.; et al. Possible anti-diabetic potentials of Annona muricata (soursop): Inhibition of α-amylase and α-glucosidase activities. Clinical Phytoscience. 2019, 5, 1–13.
19. Sokpe, A.; Mensah, M.L.K.; Koffuor, G.A.; Thomford, K.P.; Arthur, R.; Jibira, Y.; et al. Hypotensive and antihypertensive properties and safety for use of Annona muricata and Persea americana and their combination products. Evidence-Based Complementary and Alternative Medicine. 2020, 2020, 8833828.
20. Moghadamtousi, S.Z.; Rouhollahi, E.; Hajrezaie, M.; Karimian, H.; Abdulla, M.A.; Kadir, H.A. Annona muricata leaves accelerate wound healing in rats via involvement of Hsp70 and antioxidant defence. International Journal of Surgery. 2015, 18, 110–117.
21. Coria-Téllez, A.V.; Montalvo-Gónzalez, E.; Yahia, E.M.; Obledo-Vázquez, E.N. Annona muricata: A comprehensive review on its traditional medicinal uses, phytochemicals, pharmacological activities, mechanisms of action and toxicity. Arabian Journal of Chemistry. 2018, 11, 662–691.
22. Kohonou, A.C.N.; Dah-Nouvlessounon, D.; Nounagnon, M.; Sognigbe, B.; Sina, H.; N'Tcha, C.; et al. Antioxidant, anti-inflammatory efficacy and HPLC analysis of Annona muricata leaves extracts from Republic of Benin. American Journal of Plant Sciences. 2020, 11, 803–818.
23. Sivanandham, V. Determination of phytocomponents in methanolic extract of Annona muricata leaf using GC-MS technique. International Journal of Pharmacognosy and Phytochemical Research. 2015, 7, 1251–1255.
24. Gleeson, M.P. ADMET for Medicinal Chemists: A Practical Guide; Wiley-VCH: Weinheim, Germany, 2019.
25. Sheridan, D. Review on computational toxicology methods applied to drug candidates. Annual Review of Pharmacology and Toxicology. 2025, 65, 123–142.
26. Smith, L.J.; Jones, R.T. Interpreting in silico ADMET predictions in drug development: Challenges and considerations. Journal of Computational Toxicology. 2023, 6, 78–92.
27. Çorbacıoğlu, Ş.K.; Aksel, G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. Turkish Journal of Emergency Medicine. 2023, 23, 195–198.
28. Opo, F.A.D.M.; Alkarim, S.; Alrefaei, G.I.; Molla, M.H.R.; Alsubhi, N.H.; Alzahrani, F.; Ahammad, F. Pharmacophore-model-based virtual-screening approaches identified novel natural molecular candidates for treating human neuroblastoma. Current Issues in Molecular Biology. 2022, 44, 4838–4858.
29. Neves, B.J.; Braga, R.C.; Melo-Filho, C.C.; Moreira-Filho, J.T.; Muratov, E.N.; Andrade, C.H. QSAR-based virtual screening: Advances and applications in drug discovery. Frontiers in Pharmacology. 2018, 9, 1275.
30. Najem, S.M.; Kadeem, S.M. A survey on fraud detection techniques in e-commerce. Tech-Knowledge. 2021, 1, 33–47.
31. Sujon, K.M.; Hassan, R.; Choi, K.; Samad, M.A. Accuracy, precision, recall, F1-score, or MCC? Empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models. Journal of Big Data. 2025, 12, 268.
32. Zeynali, T.M.; Nowkarizi, M. Three approaches to measuring recall on the Web: A systematic review. The Electronic Library. 2020, 38, 477–492.
33. Powers, D.M.W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv. 2020.
34. Saito, T.; Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE. 2015, 10, e0118432.
35. Chowdhury, R.; Saima, S.A.; Amin, M.A.; et al. In silico

discovery of novel cephalosporin antibiotic conformers via ligand-based pharmacophore modelling and de novo molecular design. Journal of Genetic Engineering and Biotechnology. 2025, 23, 100514.
36. Tran, X.T.; Phan, T.L.; To, V.T.; Tran, N.V.; Nguyen, N.N.; Nguyen, D.N.; Tran, N.T.; Truong, T.N. Integration of the Butina algorithm and ensemble learning strategies for the advancement of a pharmacophore ligand-based model: An in silico investigation of apelin agonists. Frontiers in Chemistry. 2024, 12, 1382319.
37. Ichsani, L.N.; Elvian, E.; Zahra, C.A.; Ramdani, A.R.S.; Aprilio, K.; Rusdin, A.; Mardisanutomo, H.T.; Muchtaridi, M. Studi in silico daun kemangi (Ocimum basilicum Folium) sebagai antikanker payudara terhadap ESRα. Indonesian Journal of Pharmaceutical Science and Technology. 2025, 12, 319–328.
38. Lee, S.; Barron, M.G. Structure-based understanding of binding affinity and mode of estrogen receptor α agonists and antagonists. PLoS ONE. 2017, 12, e0169607.
39. Li, J.; Liu, W.; Song, Y.; Xia, J. Improved method of structure-based virtual screening based on ensemble learning. RSC Advances. 2020, 10, 7609–7618.
40. Wójcikowski, M.; Ballester, P.J.; Siedlecki, P. Performance of machine-learning scoring functions in structure-based virtual screening. Scientific Reports. 2017, 7, 1–10.
41. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 1997, 23, 3–25.
42. Benet, L.Z.; Hosey, C.M.; Ursu, O.; Oprea, T.I. BDDCS, the rule of 5 and drugability. Advanced Drug Delivery Reviews. 2016, 101, 89–98.
43. Sreelakshmi, V.; Raj, N.; Abraham, A. Evaluation of the drug-like properties of kaempferol, chrysophanol and emodin and their interactions with EGFR tyrosine kinase—an in silico approach. Natural Product Communications. 2017, 12.
44. Understanding Lipinski's Rule of 5 and the Role of LogP Value in Drug Design and Development. Available online: https://www.sailife.com/understanding-lipinskis-rule-of-5-and-the-role-of-logp-value-in-drug-design-and-development/ (accessed on 15 July 2024).
45. Guan, L.; Yang, H.; Cai, Y.; Sun, L.; Di, P.; Li, W.; et al. ADMET-score—A comprehensive scoring function for evaluation of chemical drug-likeness. MedChemComm. 2019, 10, 148–157.
46. Wang, N.N.; Huang, C.; Dong, J.; Yao, Z.J.; Zhu, M.F.; Deng, Z.K.; et al. Predicting human intestinal absorption with modified random forest approach: A comprehensive evaluation of molecular representation, unbalanced data, and applicability domain issues. RSC Advances. 2017, 7, 19007–19018.
47. Suherman, M.; Meilia, E.; Riska, P.; Purnamasari, A.R. In silico study: Secondary metabolites from jambolan (Syzygium cumini L.) as potential breast cancer treatments. Jurnal Ilmiah Farmako Bahari. 2024, 15, 148–162.
48. Belpaire, F.M.; Bogaert, M.G. The fate of xenobiotics in living organisms. In The Practice of Medicinal Chemistry, 2nd ed.; Wermuth, C.G., Ed.; Academic Press: London, UK, 2003; pp. 501–515.
49. Afzal, M.; Kazmi, I.; Kaur, R.; Hosawi, S.B.I.; Kaleem, M.; Alzarea, S.I.; et al. Introduction to molecular pharmacology: Basic concepts. In How Synthetic Drugs Work; Kazmi, I.; Karmakar, S.; Shaharyar, M.A.; Afzal, M.; Al-Abbasi, F.A., Eds.; Academic Press: London, UK, 2023; pp. 1–25.
50. Okella, H.; Okello, E.; Mtewa, A.G.; Ikiriza, H.; Kaggwa, B.; Aber, J.; et al. ADMET profiling and molecular docking of potential antimicrobial peptides previously isolated from African catfish, Clarias gariepinus. Frontiers in Molecular Biosciences. 2022, 9.
51. Seidel, T.; Bryant, S.D.; Ibis, G.; Poli, G.; Langer, T. 3D pharmacophore modeling techniques in computer-aided molecular design using LigandScout. In Tutorials in Chemoinformatics; Springer: New York, NY, USA, 2017; pp. 279–309.
52. Susanti, N.M.P.; Laksmiani, N.P.L.; Dewi, P.P.P.; Dewi, P.Y.C. Molecular docking terpinen-4-ol pada protein IKK sebagai antiinflamasi pada aterosklerosis secara in silico. Jurnal Farmasi Udayana. 2019, 8, 44–49.
53. Giordano, D.; Biancaniello, C.; Argenio, M.A.; Facchiano, A. Drug design by pharmacophore and virtual screening approach. Molecules. 2022, 27, 646.
54. Chandrasekaran, B.; Agrawal, N.; Kaushik, S. Pharmacophore development. In Encyclopedia of Bioinformatics and Computational Biology; Elsevier: Oxford, UK, 2019; Vol. 2, pp. 677–687.
55. Suvannang, N.; Preeyanon, L.; Malik, A.A.; Schaduangrat, W.; Worachartcheewan, T.; et al. Probing the origin of estrogen receptor alpha inhibition via large-scale QSAR study. Journal of Molecular Graphics and Modelling. 2018, 82, 201–211.