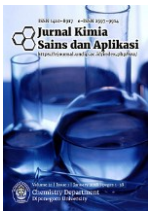
ISSN: 1410-8917
Jurnal Kimia Sains & Aplikasi
e-ISSN: 2597-9914

Jurnal Kimia Sains dan Aplikasi 28 (6) (2025): 316-326

Jurnal Kimia Sains dan Aplikasi
Journal of Scientific and Applied Chemistry

Journal homepage: http://ejournal.undip.ac.id/index.php/ksa



Quantitative Structure-Activity Relationship of O-methyl Quercetin Analogs, Structure Modification, and Molecular Docking as Lung Anticancer EGFR-TK Inhibitor

Bina Lohita Sari [1,*], Lusi Agus Setiani [1], Shafana Zahra Aulia [1]

[1] Pharmacy Study Program, Faculty of Mathematics and Natural Sciences, Universitas Pakuan, Bogor, Indonesia

* Corresponding author: binalohitasari@unpak.ac.id

https://doi.org/10.14710/jksa.28.6.316-326

Article Info

Article history:

Received: 21st April 2025
Revised: 07th July 2025
Accepted: 14th July 2025
Online: 05th August 2025

Keywords:
QSAR; Quercetin Analogs; Molecular Docking; EGFR-TK; Lung Cancer

Abstract

Cancer arises from the uncontrolled proliferation of cells. Lung cancer stands as one example among the diverse array of cancer types. The main cause of the development of lung cancer is the activation of epidermal growth factor receptor (EGFR)-tyrosine kinases (TK). O-methyl quercetin analogs, as one of the quercetin derivatives, can be potential drug candidates for treating lung cancer. In this study, we disclose our findings that O-methyl quercetin analogs and their modified forms, O-methylamino analogs, predicted EGFR-TK inhibitors as lung anticancer. The O-methylated quercetin analogs can be predicted using a Quantitative Structure-Activity Relationship (QSAR) model. The structures were optimized using the parameterized method 3 (PM3) and analyzed through multiple linear regression (MLR). Lower PRESS QSAR values are used for structural modification of O-methylamino as new compounds. Structures of O-methyl quercetin and O-methylamino analogs were docked to the EGFR-TK receptor using molecular docking. The best QSAR model of IC_{50} predicted result is expressed as log IC_{50} = 23.059 + (7.397 × log P) + (0.273 × dipole moment) – (0.005 × heat of formation) – (0.733 × E_{LUMO}) – (0.501 × E_{HOMO}) with statistical parameters: R = 0.966; R^2 = 0.933; F_{count}/F_{table} = 3.829853; and Q^2 = 0.752226. The O-methyl quercetin analog QC14 (quercetin 5,3',4'-trimethyl ether) and the modified derivative QC6_8 (3,5-dihydroxy-2-(3-hydroxy-4-((methylamino)methoxy)phenyl)-7-methoxy-4H-chromen-4-one) exhibited the lowest docking scores. Both compounds interact with the key residue Met769 of the EGFR-TK receptor, suggesting their potential as drugs for lung cancer.

1. Introduction

Cancer encompasses a collection of illnesses marked by abnormal cell proliferation and the dissemination of cells from their site to other areas of the body [1]. In 2020, around 2.2 million new cases and 1.8 million confirmed cancers as the leading cause of mortality associated with the illness [2]. Non-small cell lung cancer (NSCLC) accounts for approximately 80-85% of all lung cancer cases [3]. The epidermal growth factor receptor (EGFR), a tyrosine kinase receptor, plays a crucial role by activating signaling cascades, including PI3K/Akt and MAPK, implicated in the pathogenesis of carcinomas [4]. This signaling can trigger cell proliferation. Pharmacophores of quercetin, one of the flavonoid family, possess wide biological activity, such as anticancer. Quercetin has been found to suppress the expression of EGFR/Akt/β-catenin signaling molecules in the A549 lung cancer cell line [5].

EGFR is commonly overexpressed in various types of cancer, including NSCLC. Structurally, EGFR is a transmembrane protein with distinct domains: the N- terminal extracellular ligand-binding domain, the transmembrane lipophilic domain, and the C-terminal intracellular tyrosine kinase domain [6]. In lung cancer, the primary external signal molecules are those that attach to the EGF ligand. The section of EGFR that spans the cell membrane links the part that binds to the ligand with the inner tyrosine kinase portion. When the ligand binds, EGFR forms pairs either with itself or with other

HER/ErbB tyrosine kinase families, a crucial step for initiating EGFR signaling and its targeted function [7].

In asymmetric dimers, the C-lobe kinase domain (of activator kinase) supports the N-lobe kinase domain (of receiver kinase), leading to catalytic activation and phosphorylation of intracellular tyrosine residues in the activator kinase. This phosphorylation process results in accelerated cell proliferation and migration, enhanced cell survival (avoiding apoptosis), and the promotion of angiogenesis [8]. There are two classes of EGFR inhibitors used clinically for NSCLC: monoclonal antibodies (e.g., Cetuximab) and tyrosine kinase inhibitors (e.g., Erlotinib and Gefitinib). Monoclonal antibodies target EGFR by binding to its extracellular domain and inhibiting receptor activation, while EGFR-TKIs reversibly compete with adenosine triphosphate to bind to the catalytic domain of the intracellular kinase domain to inhibit cell growth signals [9]. EGFR-TKIs such as erlotinib, gefitinib, and afatinib have proven their efficacy in the treatment of patients with advanced-stage NSCLC with tumors harboring the EGFR sensitizing mutations (such as exon 19 deletion and the L858R point mutation in exon 21) [10].

Quercetin (3,3',4',5,7-pentahydroxyflavone) is a flavonoid compound in the flavonols group, found in apples, red grapes, onions, raspberries, and green leafy vegetables. It exhibits various biological effects, including antioxidant, anticancer, cell cycle modulator, and inhibition of angiogenesis [11]. The anticancer effectiveness of methylated metabolites from quercetin is notably more potent than that of quercetin alone, regardless of whether the methyl group is positioned at the quercetin B-ring. The addition of a methyl group at the 3'- or 4'-position notably enhances the anticancer activity against lung adenocarcinoma cells, making them potential candidates in the fight against lung cancer [12].

In the research conducted by Shi *et al.* [13] on the structure-activity relationship of methylated quercetin as a cancer proliferation inhibitor, it was shown that the methylation of quercetin's OH groups at the 4'- or/and 7-positions is crucial in maintaining or even enhancing its inhibitory activity against 16 cancer cell lines, including A549 lung cancer cells line. One effective strategy in designing drug candidates involves the use of the Quantitative Structure-Activity Relationship (QSAR) approach. The QSAR model developed on quercetin derivatives can provide specific activity as new chemotherapeutic agents.

QSAR is defined as a mathematical equation that correlates the biological activities of compounds to their structural features [14]. QSAR modeling serves as an invaluable tool to prioritize a large number of chemicals associated with expected biological activity, employing an in silico approach that significantly reduces the number of candidate chemicals requiring in vivo experiments [15]. The primary objective of the QSAR modeling approach is to construct a model based on the most relevant descriptors that exhibit the highest correlation to the endpoint value. A highly effective model is characterized by the utilization of the minimum number of descriptors capable of encompassing most of the activity data involved for the investigated compounds. The descriptors utilized reflect a number of Hansch QSAR parameters, which are categorized into three categories according to their hydrophobic, electronic, and steric properties. The Hansch model, which was developed in 1964, allowed medicinal chemists to express their hypotheses of structure-activity relationships in quantitative terms and to test them using statistical approaches.

To measure hydrophobicity bioavailability, the octanol-water partition coefficient (log P) and steric properties used heat of formation (HF), surface area (surface area approximation (SAA) and surface area grid (SAG)), molar refractivity, and molar volume [16]. Statistical analysis in QSAR is commonly performed using the multiple linear regression (MLR) method, which represents the simplest form of linear regression involving the equation between the dependent variable and multiple independent variables. The leave-one-out (LOO) cross-validation method is often employed to assess QSAR models. This method involves systematically removing one compound at a time from the entire set, followed by training the model using the remaining molecules to predict all activity values [17].

The molecular structure will be minimized using the semi-empirical parameterized method 3 (PM3), which is based on the Hartree-Fock theory. Semi-empirical quantum chemical methods, such as PM3, significantly reduce computational cost by ignoring and parameterizing the integral part of electrons [16]. In this study, PM3 was selected as the fundamental semi-empirical geometric optimization method due to its widespread use and features that leverage the self-consistent field procedure.

Molecular docking is a computational approach based on the structure of molecules, allowing the discovery of new compounds by forecasting interactions between ligand and targets on a molecular scale, or elucidating structure-activity relationship (SAR) [18]. In this study, molecular docking was conducted on both reference compounds and new compounds, whose activities were predicted using a validated QSAR model. The aim was to identify O-methyl quercetin analogs with improved activity and docking scores compared to the reference compounds, as potential EGFR-TK inhibitors for anticancer therapy. Two compounds of the analogs were modified by adding O-alkylamino as a new compound design. Molecular docking served as an in silico approach to evaluate and identify promising compound candidates.

2. Experimental

This research aimed to derive a QSAR model for predicting novel designed compounds. The dataset for model development comprised compounds that had undergone experimental antiproliferative activity testing via high-throughput screening on the A549 cancer cell line. These compounds were optimized, and their descriptors were calculated to obtain values for statistical analysis, ultimately leading to the development of the

QSAR equation model. Novel designed compounds exhibiting favorable activity values based on the validated QSAR equation model were subsequently subjected to molecular docking with the EGFR-TK receptor. This docking study aimed to determine if binding occurs between the drug candidates and the target protein of the receptor.

2.1. Equipment

The hardware used was a laptop equipped with an Intel Core i5-5200U processor (2.20 GHz), 8 GB of Random Access Memory (RAM), and 250 GB of Solid State Drive (SSD). The software used included ChemDraw Professional 16.0 for building two-dimensional molecular structures. HyperChem 8.0.7 for optimizing structures and calculating physicochemical properties as descriptors. SPSS Statistics 24 for data analysis. DOCK 6.9 for molecular docking. Chimera v.1.16 for displaying docking results scores, and Discovery Studio Visualizer for visualizing docking results.

2.2. Materials

The inhibitory activity of 16 O-methyl quercetin analogs and quercetin against the proliferation of the human lung cancer cell line A549 was assessed using the high-throughput screening approach employed to construct the QSAR model. This data was sourced from the research conducted by Shi *et al.* [13]. For molecular docking, the EGFR-TK complex, bound with the native ligand 4-anilinoquinazoline, was utilized. The structural information for the receptor was retrieved from the Protein Data Bank (https://www.rcsb.org/structure/1M17) with PDB ID: 1M17.

2.3. Experiment

2.3.1. Geometry Optimization

The molecular structure was minimized using the semi-empirical Parameterized Method 3 (PM3), which is based on the Hartree-Fock theory. PM3 offers better accuracy compared to AM1 for hydrogen bond angles. Furthermore, its overall predictions for HF, bond energies, and bond lengths are more accurate than those from MNDO or AM1. The primary advantage of semi-empirical calculations like PM3 is their speed, which surpasses that of ab initio calculations such as DFT. The PM3 approach uses parameterization from experimental data and simpler integral calculations, in contrast to the DFT approach, which requires a large computational database because its optimization and calculation processes are time-consuming [19, 20]. In this study, PM3 was selected as the fundamental semi-empirical geometry optimization method due to its widespread use and features that leverage the self-consistent field procedure [17].

The two-dimensional structures of the O-methyl quercetin analogs, drawn using ChemDraw Professional 16.0, were optimized in the HyperChem 8.0.7 program. The optimization was conducted using the semi-empirical PM3 method, and hydrogen atoms were added using the 'Add H & model build' menu. The optimization algorithm employed was Polak-Ribiere, with the convergence limit set at 0.1 kcal/Å·mol.

Physicochemical properties, including surface area (grid and approximation), polarizability, hydration energy, volume, log P, and refractivity, were used as descriptors and were obtained from the 'QSAR Properties' menu. Additionally, properties such as total energy, binding energy, electronic energy, HF, and dipole moment (from the 'Molecular Properties' menu), as well as E_{HOMO} (highest occupied molecular orbital energy) and E_{LUMO} (lowest unoccupied molecular orbital energy) (from the 'Orbital' menu), were also considered.

2.3.2. Multiple Linear Regression (MLR) Analysis

QSAR models were developed using MLR with the backward elimination method in SPSS Statistics 24. The dependent variable was the logarithm of the compound's activity (log IC_{50}). Independent variables were initially chosen based on their potential to form regression models with an R-value > 0.9.

The backward elimination method began by including all potential independent variables in the model. Subsequently, variables were systematically removed one by one, starting with the least statistically significant, until only significant descriptors remained. The resulting final model was considered the recommended model. This final model was used to construct a QSAR equation based on the linear regression formula shown in Equation (1). To facilitate the formulation of the QSAR equation, the selected descriptors were further combined into subsets of 3, 4, or 5 descriptors. If a model yielded an R-value less than 0.9, data selection was performed to improve its performance. This involved removing 1 to 3 compound data. A linear regression graph was then plotted, comparing experimental log IC_{50} values with those predicted by the model. If this plot indicated an improved R^2 (approaching 0.9 or higher), the refined dataset was re-analyzed in SPSS and subjected to LOO cross-validation.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (1)$$

In the regression model, β_0 represents the constant; Y is the dependent variable; and X_1, X_2,...X_n are independent variables that represent molecular parameters with their corresponding coefficients β_1, β_2,...β_n (1). The QSAR model was then validated using LOO cross-validation, in which compound data were systematically removed one by one, and coefficients from the linear regression were entered. A good model is indicated by a Q^2 value ≥ 0.5. The Q^2 calculation is defined by Equation (2).

$$Q^2 = 1 - \frac{\sum(Y_{obs(train)} - Y_{pred(train)})^2}{\sum(Y_{obs(train)} - \bar{Y}_{(train)})^2} \quad (2)$$

Where, $Y_{obs(train)}$ represents the activity observation value, $Y_{pred(train)}$ is the activity prediction value, and $\bar{Y}_{(train)}$ is the average activity value.

2.3.3. New Compounds Design

The design of new compounds involved modifying the substituent locations in the reference compounds,

which were altered with O-alkylamino. The antiproliferative activity (IC_{50}) of these modified compounds was then predicted using the validated QSAR model.

2.3.4. Molecular Docking

The structure of the EGFR-TK receptor complexed with the native ligand, 4-anilinoquinazoline (erlotinib), was obtained from X-ray diffraction with a resolution of 2.60 Å and downloaded from the Protein Data Bank with the code 1M17. Using Discovery Studio Visualizer, water molecules were removed, and the native ligand was separated from the receptor, the saved in *.pdb* format. Subsequently, the ligand was separated and prepared in the Chimera program by adding hydrogen and applying the AMBER ff14SB force field charges. The receptor, with hydrogen removed, was prepared for sphere identification based on the native ligand location, with an RMSD of 5 Å.

After obtaining the sphere output, the Showbox command (in the Ubuntu command prompt) was used to generate a box around the ligand, and the output from the network was obtained. Before performing molecular docking with DOCK 6.9, energy minimization was carried out. Subsequently, flexible molecular docking was conducted, allowing the ligand to be completely free-rotational. The grid score was shown using the View Dock menu in the Chimera program. The molecular docking method was validated by redocking the native ligand, with a parameter value of RMSD < 2 Å. The chemical interactions from the molecular docking results were visualized with Discovery Studio Visualizer.

3. Results and Discussion

The modifiable positions on the basic structure are denoted as R1, R2, R3, R4, and R5, as illustrated in Figure 1. The basic structure (Figure 1) was modified by introducing methyl groups (CH_3) as substituents at various positions, as listed in Table 1, which also presents their corresponding biological activity values.

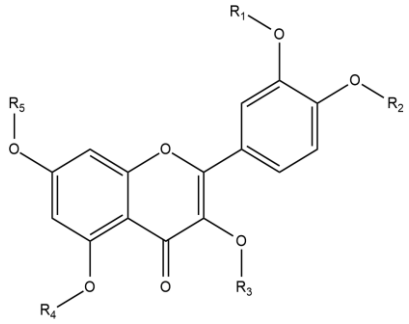


Figure 1. The basic structure

Table 1. IC_{50} value* (µM) antiproliferative activity of quercetin analog compounds against A549 lung cancer cells

No.	Code	Compound name	Substituent					IC_{50} (µM)	log IC_{50}
			R1	R2	R3	R4	R5		
1.	QC1	Quercetin	H	H	H	H	H	6.20	0.7924
2.	QC2	Isorhamnetin	CH_3	H	H	H	H	10.78	1.0326
3.	QC3	Tamarixetin	H	CH_3	H	H	H	2.63	0.4200
4.	QC4	3-O-methylquercetin	H	H	CH_3	H	H	8.14	0.9106
5.	QC5	Azaleatin	H	H	H	CH_3	H	6.26	0.7966
6.	QC6	Rhamnetin	H	H	H	H	CH_3	3.08	0.4886
7.	QC7	3',5-O-dimethylquercetin	CH_3	H	H	CH_3	H	3.14	0.4969
8.	QC8	3,7-O-dimethylquercetin	H	H	CH_3	H	CH_3	5.40	0.7324
9.	QC9	Dillenetin	CH_3	CH_3	H	H	H	6.29	0.7987
10.	QC10	3,5-O-dimethylquercetin	H	H	CH_3	CH_3	H	10.32	1.0137
11.	QC11	5,7-O-dimethylquercetin	H	H	H	CH_3	CH_3	6.67	0.8241
12.	QC12	Ombuin	H	CH_3	H	H	CH_3	3.07	0.4871
13.	QC13	Ayanin	H	CH_3	CH_3	H	CH_3	20.79	1.3179
14.	QC14	3',4',5-tri-O-methylquercetin	CH_3	CH_3	H	CH_3	H	4.66	0.6684
15.	QC15	Quercetin 3',4',7-trimethyl ether	CH_3	CH_3	H	H	CH_3	5.90	0.7709
16.	QC16	Retusin	CH_3	CH_3	CH_3	H	CH_3	> 50•	-
17.	QC17	Quercetin 3',4',3,5,7-pentamethyl ether	CH_3	CH_3	CH_3	CH_3	CH_3	23.72	1.3751

*IC_{50} values of compounds to inhibit human lung cancer cell lines A549 are shown as the mean of three determinations.
•IC_{50} value > 50 means that the data were not applicable.

QSAR is a chemoinformatic technique that utilizes data mining to expedite drug design and discovery. Its objective is to predict the biological activity of compounds and facilitate the design of new active compounds. Seventeen compounds were assessed for their proliferation inhibitory activity using high-throughput screening on the A549 lung cancer cell line. Of these, sixteen were O-methyl quercetin analogs, and one was quercetin itself. The results revealed that 15 out of these compounds exhibit an IC_{50} value < 50 µM, making them eligible for inclusion in the QSAR analysis. These details are summarized in Figure 1 and Table 1.

3.1. Geometry Optimization and Descriptors Calculation

In constructing a QSAR model, it is crucial to select descriptors that effectively capture the physicochemical properties of molecules. For this study, descriptors were specifically chosen to mirror the parameters outlined in the Hansch equation, encompassing hydrophobic, electronic, and steric parameters. The hydrophobicity parameter is represented by log P. Electronic parameters include descriptors such as total energy, electronic energy, binding energy, hydration energy, dipole moment, polarizability, E_{HOMO}, and E_{LUMO}. The steric parameters encompass HF, SAA, SAG, refractivity, and molar volume (Supplementary Table).

Rouane *et al.* [21] conducted a study using seven physicochemical properties of eighteen quercetin derivatives as new chemotherapeutic agents. The QSAR properties calculated from HyperChem 8.0.3 were: molar weight, surface area grid, molar volume, molar polarizability, molar refractivity, hydration energy, and partition coefficient octanol/water (log P). Various physicochemical parameters, especially the log P, could be successfully used to model the chemotherapeutic activity of quercetin derivatives.

Hydrophobicity plays a crucial role in regulating passive membrane transport, and calculated hydrophobicity metrics are often employed to predict drug absorption. Log P represents the molecular log P between the water and hydrophobic phase, with parameters like water and octanol commonly used to define hydrophobicity. Atom-based log P prediction considers the contribution of each atom to the log P [22]. In Table 2, the 15 compounds exhibited log P values ranging from -3.091 to -3.854, which were higher than the log P value of quercetin (-4.013). This indicates that as the number of methyl substituents increases, the compound's polarity decreases.

The difference in electronegativity between two atoms forming a covalent bond is represented by the dipole moment value [23]. A substantial difference in electronegativity values indicates correspondingly large dipole moment values [24]. Molecules with a significant electronegative difference exhibit polarity due to the uneven distribution of electrons between bonded atoms, resulting in a dipole moment close to 0 (zero). Based on the dipole moment values presented in Table 2, the largest and smallest dipole moments are observed in compounds QC8, with a value of 4.0299 Debye, and QC14, with a value of 1.2120 Debye, respectively.

Table 2. Recapitulation of descriptors (5 of 14)

No.	Code	Descriptors						IC_{50} (µM)	Predicted IC_{50} (µM)
		log P	Dipole moment (D)	HF (kcal/mol)	E_{HOMO} (eV)	E_{LUMO} (eV)	GAP (eV)		
1	QC2	-3.981	2.2016	-219.6104	-8.9417	-0.9713	7.9705	10.78	3.18
2	QC3	-3.981	1.4872	-217.0653	-9.0379	-0.9910	8.0469	2.63	2.28
3	QC4	-3.981	3.6956	-169.2159	-9.1968	-1.1141	8.0827	8.14	7.78
4	QC5	-3.981	1.7248	-208.1270	-8.6723	-1.0051	7.6672	6.26	1.60
5	QC6	-3.981	1.7618	-217.8947	-8.7235	-1.0721	7.6514	3.08	2.18
6	QC7	-3.949	1.8568	-202.5358	-8.8920	-0.8845	8.0075	3.14	2.96
7	QC8	-3.949	4.0299	-162.0914	-9.1799	-1.0783	8.1016	5.40	14.07
8	QC9	-3.949	2.0807	-209.2738	-8.9104	-0.9853	7.9251	6.29	4.46
9	QC10	-3.949	3.8894	-151.6564	-9.1106	-0.9933	8.1173	10.32	9.14
10	QC11	-3.949	2.6890	-200.8729	-8.6577	-0.9617	7.6960	6.67	4.26
11	QC12	-3.949	2.2266	-207.8881	-8.6579	-1.0517	7.6062	3.07	4.02
12	QC13	-3.918	3.4176	-152.5533	-9.2124	-1.0995	8.1129	20.79	15.66
13	QC14	-3.918	1.2120	-191.7244	-8.8310	-1.0443	7.7868	4.66	3.60
14	QC15	-3.918	1.4628	-203.0709	-9.0300	-0.9503	8.0798	5.90	5.16
15	QC17	-3.854	3.2063	-128.3908	-9.1191	-09751	8.1440	23.72	22.48

The calculated results of electronic descriptors such as E_{HOMO}, E_{LUMO}, and the E_{LUMO}-E_{HOMO} gap (GAP) provide insights into electron transfer [23]. E_{HOMO} denotes the ability to donate electrons to an acceptor. Higher E_{HOMO} value indicates a tendency to donate electrons to a low-energy acceptor (which lacks electrons). E_{LUMO} represents the lowest energy level capable of accepting electrons [25]. The GAP serves as an index of molecular stability and reactivity. A higher GAP value suggests greater molecular stability. Molecules with large GAPs tend to have lower reactivity due to the increased energy required for electronic transitions. Conversely, molecules with a smaller GAP are typically more reactive as they undergo electronic transitions more readily. Therefore, the compound QC17, with the highest GAP value, exhibits the lowest reactivity and optimal stability.

The HF is the alteration in enthalpy occurring when 1 mole of a compound is created from its constituent elements in their standard states at 25°C and 1 atmosphere pressure [21]. Among the 15 O-methyl analogs of quercetin compounds, it was observed that compound QC17 has the highest HF value (-128.3908 kcal/mol), while compound QC2 has the lowest (-219.6104 kcal/mol). This indicates that compound QC2 requires the least energy to form under standard conditions (25°C, 1 atm). Compound QC2, QC6, and QC5 share the top three positions based on the same log P descriptors value (-3.981), along with other parameters such as HF, SAA, SAG, molar volume, total energy, binding energy, electronic energy, hydration energy, and the lowest GAP.

In this research, the geometry optimization method employed is Parameterized Method 3 (PM3), processed using the HyperChem 8.0.7 program for Windows. PM3 is a semi-empirical quantum mechanical calculation method widely used for energy minimization. This method is based on Hartree-Fock theory, well-documented with features, and recognized for its simplicity utilizing self-consistent field procedures [17]. Geometry optimization was performed after converting the two-dimensional compound structure into a three-dimensional form. A stable conformation was achieved through an iterative process, repeated until the convergence limit of 0.1 kcal/mol was met. During the iteration process, energy calculations were conducted to track the compound's conformational changes.

3.2. Multiple Linear Regression (MLR) Analysis for QSAR Model Development

The determination of the QSAR model involved statistical analysis using SPSS Statistics 24 with MLR. The dependent variable was the logarithm of the proliferation inhibition variable was the logarithm of the proliferation inhibition activity value (log IC_{50}), while the independent variables comprised the values of fourteen molecular descriptors for the fifteen compounds in the dataset, utilizing the backward elimination method. MLR is a statistical process aimed at identifying correlations between descriptors as independent variables and activities as the dependent variable. The correlation in MLR was modeled with a linear function, attempting to predict the unknown independent variables [26].

The results of the MLR analysis using the backward elimination method, presented in Table 3, reveal three models. The first model, comprising ten descriptors, produced a correlation coefficient (R-value) of 0.973. The backward elimination method worked by eliminating descriptors with significance values > 0.1, thus reducing descriptors in the first model due to the significance values of binding energy and hydration energy being 0.484 and 0.532, respectively. In the second model, the binding energy descriptor was removed, resulting in nine descriptors. Subsequently, the backward elimination method further reduced one descriptor due to the significance value of hydration energy being 0.290. In the third model, the significance values of the remaining eight descriptors met the criteria (< 0.1).

In model 3, a combination of eight descriptors, including log P, HF, molar volume, surface area (SAA and SAG), dipole moment, E_{HOMO}, and E_{LUMO}, was created by grouping them into 3, 4, and 5 descriptors. This combined set was reanalyzed using the enter method. The purpose of combining descriptors was to simplify the formulation of the QSAR model. Subsequently, eighteen models with an R-value < 0.9 were generated. To improve the R-value, a selection process for the compound was initiated. This involved iteratively removing one to three compounds and plotting the experimental log IC_{50} values against the predicted log IC_{50} values using the QSAR model, resulting in fifteen equations with an R-value > 0.7. Microsoft Excel aided in predicting R^2-values by creating linear regression graphs of experimental log IC_{50} and predicted log IC_{50}.

Table 3. Result of MLR with the backward method

Model	Descriptor	n	m	R	R^2	SEE
1.	Molar volume, E_{LUMO}, dipole moment, E_{HOMO}, HF, binding energy, SAA, SAG, hydration energy, log P	15	10	0.973	0.947	0.124546
2.	Molar volume, E_{LUMO}, dipole moment, E_{HOMO}, HF, binding energy, SAA, SAG, log P	15	9	0.970	0.940	0.117726
3.	Molar volume, E_{LUMO}, dipole moment, E_{HOMO}, HF, SAA, SAG, log P	15	8	0.961	0.924	0.121603

Note: Model is a linear regression representation; n is the number of compound data; R is the correlation coefficient; R^2 is the coefficient determination; SEE (Standard Error of the Estimation) is an estimate of the accuracy of predictions; SAA is a surface area (approximation); SAG is a surface area (grid).

Table 4. QSAR models validation

Model	Descriptor	n	R	R^2	SEE	F_{count}/F_{table}	Q^2	PRESS
7	log P, HF, SAG, molar volume	12	0.927	0.860	0.148990	2.608	0.046	1.0579
8	log P, dipole moment, HF, E_{LUMO}	12	0.946	0.895	0.128636	3.634	0.964	0.3102
18	log P, dipole moment, HF, E_{LUMO}, E_{HOMO}	12	0.966	0.933	0.110881	3.830	0.752	0.2742

Note: Q^2 is the LOO cross-validation coefficient; PRESS (Predicted Residual Error Sum of Squares) is a form of cross-validation in regression analysis to provide a summary measure of model fit to a sample of observations not used to estimate the model.

Validation of the QSAR model aimed to assess its goodness and demonstrate repeatability. In linear regression analysis, the quality of the LOO cross-validation, represented by the Q^2-value, served as a crucial indicator. Q^2 was utilized to evaluate the model's predictive ability in the LOO method. Among the three QSAR models presented in Table 4, only QSAR model 7 (Equation 3) exhibited a positive Q^2-value of 0.0465. However, it is essential to note that the minimum acceptable Q^2-value is 0.5 [11]. As QSAR model 7 shows below this threshold, indicating a lower predictive ability, it was not chosen for predicting the activity values of new compounds.

$$\log IC_{50} = -101.998 - (22.694) \log P + (0.000093) HF - (0.011) SAG + (0.022) \text{molar volume} \quad (3)$$

The QSAR model 8 (Equation 4) showed a Q^2-value of 0.9645. Despite meeting the Q^2-value criteria, the QSAR model 8 had an R^2-value of 0.895, which is less than 0.9.

$$\log IC_{50} = -19.786 - (5.036) \log P + (0.189) \text{dipole moment} - (0.001) HF - (0.570) E_{LUMO} \quad (4)$$

The R^2-value of QSAR model 8 was lower than that of model 18 (Equation 5), which exhibited a Q^2-value of 0.7522 and an R^2-value of 0.933. In addition to a superior R^2-value, the PRESS value of QSAR value for model 18 at 0.2742, compared to model 8, is 0.3192. PRESS, a type of cross-validation in regression analysis, offers a summary measure of model fitness using observation samples not employed in model estimation. A lower PRESS value signifies a better model, indicating decreased potential errors [27]. Consequently, QSAR model 18 was selected as the best model for prediction.

$$\log IC_{50} = 23.059 + (7.397) \log P + (0.273) \text{dipole moment} - (0.005) HF - (0.733) E_{LUMO} - (0.501) E_{HOMO} \quad (5)$$

To validate a QSAR model, most researchers apply the LOO or leave-many-out (LMO) cross-validation procedure. Q^2 is frequently used as a criterion of both robustness and predictive ability of the model. They do not test the models for their ability to predict the activity of compounds in an external test set. Hawkins *et al.* [28], as leading advocates of internal validation, believe that cross-validation is an effective method for evaluating model fit and determining whether the predictions can be applied to new, unused data. They argue that when the sample size is small, holding a portion of it back for testing is wasteful, and it is much better to use 'computationally more burdensome' LOO cross-validation [29].

Assessment of prediction capability and applicability of a QSAR model to predict newly designed or untested molecules is done using external validation metrics. In most cases, some compounds from the original dataset are used for validation purposes when true external data points are limited or not available [30]. Rouane *et al.* [21] tested the validity of their model using the LOO cross-validation technique. A positive Q^2 value indicated good predictive power. The PRESS value was also an important parameter in the validation, as it is a good estimate of the model's true prediction error.

3.2.1. New Compounds Design

Based on QSAR model 18 (Equation 5), the activity prediction results in Table 2 indicated that compounds tamarixetin (QC3) and rhamnetin (QC6) exhibited both low experimental IC_{50} values (2.63 µM and 3.08 µM, respectively) and low predicted IC_{50} values (2.28 µM and 2.18 µM, respectively). Notably, azaleatin (QC5) had the lowest predicted IC_{50} value (1.60 µM), even though its experimental IC_{50} value was relatively high (6.26 µM).

On the other hand, ombuin (QC12) had a low experimental IC_{50} value (3.07 µM), but its predicted IC_{50} value was comparatively higher (4.02 µM). Considering these factors, tamarixetin and rhamnetin were selected as reference compounds for structural modification. Apigenin, a flavonoid compound structurally similar to quercetin, has previously been modified by adding an O-alkylamino group at the 4'-position, resulting in an increased inhibitory effect compared to apigenin itself [31].

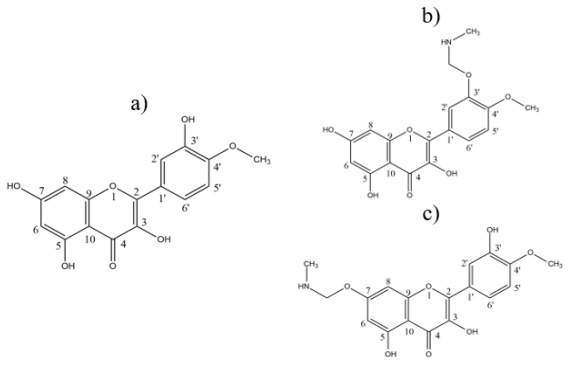


Figure 2. The structures of a) tamarixetin (QC3), the tamarixetin that was modified with b) O-alkylamino in 3'-position (QC3_8), and c) O-alkylamino in 7- position (QC3_9)

In this research, O-methyl analogs at the 3'- and/or 4'-positions were modified with amino groups to synthesize O-methylamino-tethered quercetin derivatives, and their activities were predicted using QSAR model 18. Additionally, the hydroxyl groups of the reference compounds were modified with O-alkylamino groups at the 3'- and/or 4'-positions, and their activities were predicted using the QSAR model, favoring secondary or primary amines with free amino groups at the terminal of the side chains.

The catechol groups (3'- and 4'-OH from ring B) and the 3-OH group from ring C of quercetin are known to be more reactive sites. The 7-OH group participates as a reaction orientor, not as a direct H donor in radical scavenging. It is believed that the loss of the H atom from the 4'-OH group gives rise to the semiquinone quercetin radical, considered the most stable radical that promotes deactivation of radicals when proton and/or electron transfers occur [32]. Elevated levels of reactive oxygen species (ROS) can lead to oxidative stress, causing an overstimulation of signal transduction pathways and heightened cell proliferation. High ROS levels also lead to metabolic adaptation to the tumor microenvironment, promoting tumorigenesis. Quercetin modulates both internal and external pathways within protein kinase-C (PKC) signaling. PKC, a crucial regulator of cell growth and differentiation in mammals, depends partially on ROS activation. PKC inhibits cell proliferation and apoptosis in cancer cells. Quercetin acts as a preventive agent against cancer by improving the regulation of p53, the frequently inactive tumor suppressor [32].

Quercetin's bioavailability is intricately linked to its bioaccessibility. However, the effectiveness of quercetin is constrained by factors such as its hydrophobicity, inadequate gastrointestinal absorption, instability in physiological fluids, and significant xenobiotic metabolism in the liver and intestines, primarily through glucuronidation or sulfation. These factors collectively contribute to the notably low oral bioavailability of quercetin [33].

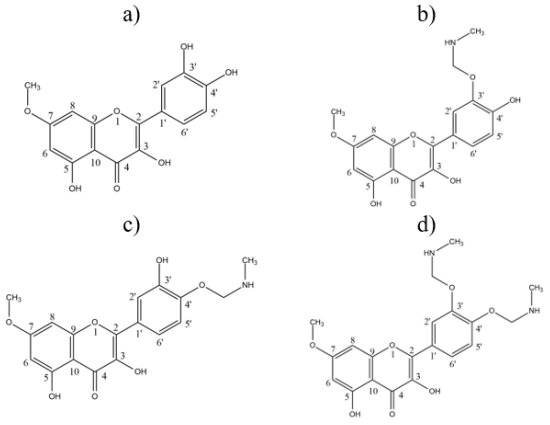


Figure 3. The structures of a) rhamnetin (QC6); the rhamnetin that is modified with b) O-alkylamino in 3'-position (QC6_7), c) O-alkylamino in 4'-position (QC6_8), and d) O-alkylamino in 3'- and 4'-positions (QC6_9)

3.2.2. Molecular Docking

Before initiating molecular docking, the EGFR-TK receptor complexed with the native ligand 4- anilinoquinazole (erlotinib) from PDB with PDB ID: 1M17 was prepared. This involved removing water molecules and separating the receptor from the native ligand. The purpose of removing water molecules was to prevent interference during the docking process.

Molecular docking was conducted using the DOCK 6.9 program, employing an incremental construction algorithm. In this approach, the ligand was fragmented from rotatable bonds into several segments. The anchor segment, considered a fragment demonstrating maximum interaction with the receptor surface, possessing a minimum number of alternative conformations, and exhibiting relative rigidity, such as a ring system, was prepared first. Subsequently, each additional fragment was added step by step. Fragments with a higher fragment were added step by step. Fragments with a higher probability, indicating interactions like hydrogen bonds, were ideally first to account for the specificity and ensure accurate geometric prediction. The algorithm evaluated the poses with the least energy after adding each fragment for the next iteration, making it robust and fast [34].

The classical force field-based scoring function calculated bond energies by accumulating van der Waals and electrostatic interactions between atom-ligand pairs [35]. However, a limitation of this function was its failure to consider charged groups, leading to a preference for a polar environment. On the contrary, no polar groups, typically associated with the desolvation effect, tended to favor a non-polar environment.

The validation assessment aimed to assess the simulation's capability to generate the pose and structure of the native ligand. The validation process was based on the RMSD, which quantifies the deviation in pose distance between redocking compared to before docking. The molecular docking validation using DOCK 6.9 was carried out on the Ubuntu operating system through the terminal to execute commands. The flexible molecular docking command was executed with the *flex.in* file setting until a qualified RMSD value was achieved. Redocking of native ligand resulted in an RMSD value of 1.0987 Å, meeting the specified requirements (< 2 Å).

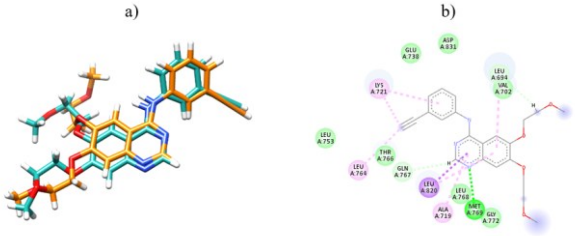


Figure 4. Visualization of redocking results with a) UCSF Chimera and the RMSD value is 1.0987 Å, and with b) Discovery Studio Visualizer for showing the interaction bond between ligand and receptor

Table 5. Molecular docking of new compounds results

No.	Code	IUPAC name	Predicted IC_{50} (μM)	Grid score (kcal/mol)	Amino acid residues	
					Hydrogen bond	Hydrophobic bond
1.	QC3_8	3,5,7-trihydroxy-2-(4-methoxy-3-((methylamino)methoxy) phenyl-4H-chromen-4-one	4.39	-68.464	Lys721; Met769; Asp831	Leu694; Val702; Ala719; Cys751; Met742; Leu820
2.	QC3_9	3,5-dihydroxy-2-(3-hydroxy-4-methoxyphenyl)-7-((methyl amino) methoxy)-4H-chromen-4-one	4.50	-70.847	Gln767; Met769; Cys773	Leu694; Val702; Ala719; Glu738; Gly772; Leu820
3.	QC6_7	3,5-dihydroxy-2-(4-hydroxy-3-((methylamino)methoxy) phenyl)-7-methoxy-4H-chromen-4-one)	1.76	-77.699	Met742; Met769	Leu694; Val702; Ala719; Lys721; Glu738; Cys773; Leu820
4.	QC6_8	3,5-dihydroxy-2-(3-hydroxy-4-((methylamino)methoxy) phenyl)-7-methoxy-4H-chromen-4-one	1.98	-79.458	Glu738; Met769; Gly772; Thr830; Asp831	Leu694; Val702; Ala719; Lys721; Leu820
5.	QC6_9	2-(3,4-bis(methylamino)methoxy) phenyl)-3,4-dihydroxy-7-methoxy-4H-chromen-4-one	2.81	-96.153	Ala719; Lys721; Glu738; Leu764; Thr766; Met769; Gly772; Thr830	Leu694; Val702; Leu820; Asp831

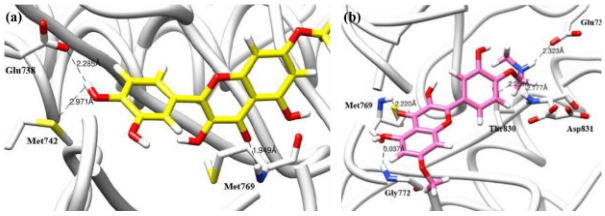


Figure 5. Binding interactions of (a) QC6 (rhamnetin as O-methyl quercetin analog) and (b) QC6_9 (modification of rhamnetin with O-alkylamino) with Met769 as the active amino acid of EGFR-TK

The results were obtained by setting the parameters in the *flex.in* file, such as: `pruning_max_orients` (maximum orientations produced before pruning), set to 900; `pruning_clustering_cutoff` (maximum number of clusters retained from pruning), set to 90; `pruning_conformer_score_cutoff` (maximum score allowed for the conformer), set to 10; `max_orientations` (maximum number of orientations), set to 2000; and `simplex_grow_max_orientations` (maximum number of iterations per cycle per receptor), set to 471.

The redocking results were visualized with the Chimera program (Figure 4a), revealing that the native ligand acted as a hydrogen bond acceptor with the active amino acid residue Met769 in the EGFR-TK receptor. This interaction occurred through the nitrogen atom in the heterocyclic structure, forming various bonds. These include conventional hydrogen bonds (N...HN) with residue Met769, carbon-hydrogen bonds with residues Leu694 and Gln767, pi-sigma (symbol) bonds with residues Leu820, alkyl bonds with residue Leu764, and pi-alkyl (symbol) bonds with residues Leu694, Ala719, Lys721, and Leu820. Van der Waals interactions occurred with residues Glu738, Leu753, Leu763, Gly772, and Asp831 (Figure 4b).

A conventional hydrogen bond was formed between the new O-alkylamino-modified compound and the catalytic residue Met769, with the carbonyl group on ring C acting as the hydrogen bond acceptor (O...HN). Among the compounds, QC3_8 (Figure 2b) had the longest bond distance (2.407 Å), while QC3_9 (Figure 2c) formed the strongest and most stable bond with Met769 (1.910 Å). Consistent with Zubair *et al.* [4] study, in the molecular docking of quercetin compounds with EGFR-TK, quercetin was found to interact with the catalytic residue Met769 via a hydrogen bond and exhibit hydrophobic interactions with several amino acids, namely Leu820, Leu694, Ala719, and Lys721 within the hydrophobic pocket of the EGFR-TK protein. Met769 is a key residue in inhibitor binding to the EGFR-TK receptor, similar to how erlotinib binds to the EGFR-TK receptor [36].

The docking results for five newly designed compounds (Table 5) showed grid scores ranging from -68.4645 to -96.1534 kcal/mol, which were more negative than those of quercetin (-47.2471 kcal/mol) and the reference compound rhamnetin (-55.9166 kcal/mol). Notably, QC6_7 (Figure 3b) and QC6_8 (Figure 3c) exhibited better predicted IC_{50} values at 1.76 μM and 1.98 μM, respectively, with corresponding grid scores of -77.6986 kcal/mol and -79.4575 kcal/mol. The determination of the most stable drug and receptor complex was based on the smallest docking score [30], indicating that QC6_7 and QC6_8 formed more stable complexes with EGFR-TK than quercetin and rhamnetin. Considering the predicted IC_{50} value and docking scores indicate that QC6_7 and QC6_8 have strong potential as more potent lung cancer drug candidates than quercetin.

Compared to molecular docking, QSAR provides a statistical model that correlates structural features with activity. A total of seventeen O-methyl quercetin analogs, as small molecules, were screened and optimized to predict active compounds. Derivatization through structural modification with O-alkylamino groups can potentially enhance biological activity and reduce

toxicity. Molecular docking shows how and where the compounds might bind to the active amino acid of EGFR-TK, giving hit compounds a mechanistic understanding. The bioactive compounds of *Glycyrrhiza glabra* as EGFR inhibitors have been screened against the EGFR catalytic site and binding energies to the EGFR. These active compounds were found to engage with essential residues of the EGFR, indicating their possible role as receptor inhibitors. Moreover, these hits exhibit favorable drug-like characteristics and warrant additional investigation for their possible use in cancer treatment [37].

4. Conclusion

From this computational research, the optimal QSAR model was established using five validated descriptors through the LOO method: log P, dipole moment, heat of formation (HF), E_{HOMO}, and E_{LUMO}. The statistical criteria for model accuracy include R = 0.966; R^2 = 0.933; F_{count}/F_{table} = 3.829853; and Q^2 = 0.752226. Among five new O-alkylamino-modified compounds predicted, QC6_7 and QC6_8 had the lowest IC_{50} values. Both are rhamnetin derivatives with O-alkylamino substitutions at the 3'- and 4'-positions of ring B, and molecular docking indicated more negative grid scores than quercetin, rhamnetin, and the native ligand. Consequently, QC6_7 and QC6_8 emerge as promising candidates for cancer drugs, specifically as EGFR-TK receptor inhibitors, based on their favorable QSAR predictions and molecular docking outcomes. To determine whether the O-methyl quercetin and O-methyl quercetin-O-alkylamino are able to inhibit EGFR-TK as an anticancer, it is necessary to conduct an in vitro test on the EGFR enzyme and A549 human cancer cells. To support the in vitro test, additional expensive in vivo research is needed to verify the molecule's safety and activity in a biological system.

Acknowledgment

This research is supported by the Faculty of Mathematics and Natural Sciences of Universitas Pakuan for funding the research grant program.

References

[1] L. Pecorino, *Molecular Biology of Cancer: Mechanisms, Targets, and Therapeutics*, Oxford University Press, 2012,

[2] Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, Freddie Bray, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, *CA: A Cancer Journal for Clinicians*, 71, 3, (2021), 209-249 https://doi.org/10.3322/caac.21660

[3] Haiying Jiang, Mei Zhu, Yanfang Li, Qian Li, Association between EGFR exon 19 or exon 21 mutations and survival rates after first-line EGFR-TKI treatment in patients with non-small cell lung cancer, *Molecular and Clinical Oncology*, 11, 3, (2019), 301-308 https://doi.org/10.3892/mco.2019.1881

[4] Muhammad Sulaiman Zubair, Syariful Anam, Saipul Maulana, Muhammad Arba, In Vitro and In Silico Studies of Quercetin and Daidzin as Selective Anticancer Agents, *Indonesian Journal of Chemistry*, 21, 2, (2021), 310-317 https://doi.org/10.22146/ijc.53552

[5] Perumal Elumalai, Devaraj Ezhilarasan, Subramanian Raghunandhakumar, Quercetin Inhibits the Epithelial to Mesenchymal Transition through Suppressing Akt Mediated Nuclear Translocation of β -Catenin in Lung Cancer Cell Line, *Nutrition and Cancer*, 74, 5, (2022), 1894-1906 https://doi.org/10.1080/01635581.2021.1957487

[6] Vladimir Jurišić, Jasmina Obradovic, Sonja Pavlović, Nataša Djordjevic, Epidermal Growth Factor Receptor Gene in Non-Small-Cell Lung Cancer: The Importance of Promoter Polymorphism Investigation, *Analytical Cellular Pathology*, 2018, 1, (2018), 6192187 https://doi.org/10.1155/2018/6192187

[7] Tie-Cheng Liu, Xin Jin, Yan Wang, Ke Wang, Role of epidermal growth factor receptor in lung cancer and targeted therapies, *American journal of cancer research*, 7, 2, (2017), 187-202

[8] Kazue Yoneda, Naoko Imanishi, Yoshinobu Ichiki, Fumihiro Tanaka, Treatment of Non-small Cell Lung Cancer with *EGFR*-mutations, *Journal of UOEH*, 41, 2, (2019), 153-163 https://doi.org/10.7888/juoeh.41.153

[9] Adi F. Gazdar, Epidermal growth factor receptor inhibition in lung cancer: the evolving role of individualized therapy, *Cancer and Metastasis Reviews*, 29, 1, (2010), 37-48 https://doi.org/10.1007/s10555-010-9201-z

[10] Misako Nagasaka, Shirish M. Gadgeel, Role of chemotherapy and targeted therapy in early-stage non-small cell lung cancer, *Expert Review of Anticancer Therapy*, 18, 1, (2018), 63-70 https://doi.org/10.1080/14737140.2018.1409624

[11] Mahmoud Hashemzaei, Amin Delarami Far, Arezoo Yari, Reza Entezari Heravi, Kaveh Tabrizian, Seyed Mohammad Taghdisi, Sarvenaz Ekhtiari Sadegh, Konstantinos Tsarouhas, Dimitrios Kouretas, George Tzanakakis, Dragana Nikitovic, Nikita Yurevich Anisimov, Demetrios A. Spandidos, Aristides M. Tsatsakis, Ramin Rezaee, Anticancer and apoptosis-inducing effects of quercetin in vitro and in vivo, *Oncology Reports*, 38, 2, (2017), 819-828 https://doi.org/10.3892/or.2017.5766

[12] Katrin Sak, Helen Lust, Marju Kase, Jana Jaal, Cytotoxic action of methylquercetins in human lung adenocarcinoma cells, *Oncology Letters*, 15, 2, (2018), 1973-1978 https://doi.org/10.3892/ol.2017.7466

[13] Zhi-Hao Shi, Nian-Guang Li, Yu-Ping Tang, Qian-Ping Shi, Hao Tang, Wei Li, Xu Zhang, Hai-An Fu, Jin-Ao Duan, Biological Evaluation and SAR Analysis of O-Methylated Analogs of Quercetin as Inhibitors of Cancer Cell Proliferation, *Drug Development Research*, 75, 7, (2014), 455-462 https://doi.org/10.1002/ddr.21181

[14] Siavoush Dastmalchi, Maryam Hamzeh-Mivehroud, Babak Sokouti, *Quantitative Structure–Activity Relationship: A Practical Approach*, 1st ed., CRC Press, Boca Raton, 2018, ^https://doi.org/10.1201/9781351113076

[15] Sunyoung Kwon, Ho Bae, Jeonghee Jo, Sungroh Yoon, Comprehensive ensemble in QSAR prediction for drug discovery, *BMC Bioinformatics*, 20, 1, (2019), 521 https://doi.org/10.1186/s12859-019-3135-4

[16] Bharat Jhanwar, Vandana Sharma, Rajeev K. Singla, Birendra Shrivastava, QSAR-Hansch analysis and related approaches in drug design, *Pharmacologyonline*, 1, (2011), 306-344

[17] Guoqing Zhou, Nicholas Lubbers, Kipton Barros, Sergei Tretiak, Benjamin Nebgen, Deep learning of dynamically responsive chemical Hamiltonians with semiempirical quantum mechanics, *Proceedings of the National Academy of Sciences of the United States of America*, 119, 27, (2022), e2120333119 https://doi.org/10.1073/pnas.2120333119

[18] Luca Pinzi, Giulio Rastelli, Molecular Docking: Shifting Paradigms in Drug Discovery, *International Journal of Molecular Sciences*, 20, 18, (2019), 4331 https://doi.org/10.3390/ijms20184331

[19] David C. Young, *Computational chemistry: a practical guide for applying techniques to real world problems*, John Wiley & Sons, Inc., 2004, ^https://doi.org/10.1002/0471220655

[20] Mutista Hafshah, Irvan Maulana Firdaus, Ika Nur Fitriani, Lutfiah Rahmawati, QSAR, Molecular Docking, and Molecular Dynamic of Novel Coumarin Derivatives as α -Glucosidase Inhibitor, *Jurnal Kimia Sains dan Aplikasi*, 27, 7, (2024), 316-327 https://doi.org/10.14710/jksa.27.7.316-327

[21] Abderrahmane Rouane, Noureddine Tchouar, Aicha Kerassa, Mehmet Cinar, Salah Belaidi, Qualitative and quantitative structure-activity relationships studies of quercetin derivatives as chemotherapeutic activity, *Journal of Bionanoscience*, 12, 2, (2018), 278-283

[22] Brian J. Bennion, Nicholas A. Be, M. Windy McNerney, Victoria Lao, Emma M. Carlson, Carlos A. Valdez, Michael A. Malfatti, Heather A. Enright, Tuan H. Nguyen, Felice C. Lightstone, Timothy S. Carpenter, Predicting a Drug's Membrane Permeability: A Computational Model Validated With *in Vitro* Permeability Assay Data, *The Journal of Physical Chemistry B*, 121, 20, (2017), 5228-5237 https://doi.org/10.1021/acs.jpcb.7b02914

[23] Grandys Perwira, Kasmui Kasmui, Subiyanto Hadisaputro, Analisis Hubungan Kuantitatif Struktur Aktivitas Antioksidan Senyawa Turunan Apigenin, *Indonesian Journal of Chemical Science*, 4, 3, (2015), 212-216

[24] Yusthinus Male, I Wayan Sutapa, Yulian A. D. Pusung, Prediksi Potensi Antikanker Senyawa Turunan Xanthon Menggunakan Hubungan Kuantitatif Struktur dan Aktivitas (HKSA), *Chemistry Progress*, 11, 1, (2019),

[25] Ahmed H. Radhi, Ennas A. B. Du, Fatma A. Khazaal, Zaid M. Abbas, Oday H. Aljelawi, Salam D. Hamadan, Haider A. Almashhadani, Mustafa M. Kadhim, HOMO-LUMO energies and geometrical structures effecton corrosion inhibition for organic compounds predict by DFT and PM3 methods, *NeuroQuantology*, 18, 1, (2020), 37-45

[26] Mohammad Reza Keyvanpour, Mehrnoush Barani Shirzad, An Analysis of QSAR Research Based on Machine Learning Concepts, *Current Drug Discovery Technologies*, 18, 1, (2021), 17-30 http://dx.doi.org/10.2174/1570163817666200316104404

[27] Sari Paramita, Maylani Permata S., Eva Vaulina Y. D., Nasrokhah Nasrokhah, Ponco Iswanto, Pemilihan Metode Perhitungan Kimia Komputasi Semi-empiris untuk Pengembangan 1, 3, 4-Thiadiazole, *Indonesian Journal of Chemical Research*, 8, 1, (2020), 51-56 https://doi.org/10.30598/ijcr.2020.8-pon

[28] Douglas M. Hawkins, Subhash C. Basak, Denise Mills, Assessing Model Fit by Cross-Validation, *Journal of Chemical Information and Computer Sciences*, 43, 2, (2003), 579-586 https://doi.org/10.1021/ci025626i

[29] Douglas M. Hawkins, The Problem of Overfitting, *Journal of Chemical Information and Computer Sciences*, 44, 1, (2004), 1-12 https://doi.org/10.1021/ci0342472

[30] Priyanka De, Supratik Kar, Pravin Ambure, Kunal Roy, Prediction reliability of QSAR models: an overview of various validation tools, *Archives of Toxicology*, 96, 5, (2022), 1279-1295 https://doi.org/10.1007/s00204-022-03252-y

[31] Haijun Chen, Amy A. Mrazek, Xiaofu Wang, Chunyong Ding, Ye Ding, Laura J. Porro, Huiling Liu, Celia Chao, Mark R. Hellmich, Jia Zhou, Design, synthesis, and characterization of novel apigenin analogues that suppress pancreatic stellate cell proliferation in vitro and associated pancreatic fibrosis in vivo, *Bioorganic & Medicinal Chemistry*, 22, 13, (2014), 3393-3404 https://doi.org/10.1016/j.bmc.2014.04.043

[32] Artem G. Veiko, Elena A. Lapshina, Ilya B. Zavodnik, Comparative analysis of molecular properties and reactions with oxidants for quercetin, catechin, and naringenin, *Molecular and Cellular Biochemistry*, 476, 12, (2021), 4287-4299 https://doi.org/10.1007/s11010-021-04243-w

[33] Dong Xu, Meng-Jiao Hu, Yan-Qiu Wang, Yuan-Lu Cui, Antioxidant Activities of Quercetin and Its Complexes for Medicinal Application, *Molecules*, 24, 6, (2019), 1123 https://doi.org/10.3390/molecules24061123

[34] Aaftaab Sethi, Khusbhoo Joshi, Sasikala K., Mallika Alvala, Molecular Docking in Modern Drug Discovery: Principles and Recent Applications, in: V. Gaitonde, P. Karmakar, A. Trivedi (Eds.) *Drug Discovery and Development - New Advances*, IntechOpen, Rijeka, 2019, https://doi.org/10.5772/intechopen.85991

[35] Qiyao Luo, Liang Zhao, Jianxing Hu, Hongwei Jin, Zhenming Liu, Liangren Zhang, The scoring bias in reverse docking and the score normalization strategy to improve success rate of target fishing, *PLOS ONE*, 12, 2, (2017), e0171433 https://doi.org/10.1371/journal.pone.0171433

[36] Frangky Sangande, Jonly Piere Uneputty, Identifikasi Senyawa Bahan Alam Sebagai Inhbitor Tirosin Kinase EGFR: Skrining In Silico Berbasis Farmakofor dan *Molecular Docking*, *Jurnal Fitofarmaka Indonesia*, 8, 1, (2021), 1-6 https://doi.org/10.33096/jffi.v8i1.539

[37] Mohammad Azhar Kamal, Hanadi M. Baeissa, Israa J. Hakeem, Reem S. Alazragi, Mohannad SHazzazi, Tahani Bakhsh, Akhmed Aslam, Bassem Refaat, Elshiekh B. Khidir, Kadhem Juma Alkhenaizi, Insights from the molecular docking analysis of EGFR antagonists, *Bioinformation*, 19, 3, (2023), 260-265 https://doi.org/10.6026/97320630019260