Walisongo Journal of Chemistry
Walisongo Journal of Chemistry Vol. 8 Issue 1 (2025), 45-56
ISSN: 2621-5985 (online); 2549-385X (print)
DOI: https://doi.org/10.21580/wjc.v8i1.23867

MODELING THE RELATIONSHIP BETWEEN NET ATOMIC CHARGE AND THE ACTIVITY OF 4-HYDRAZINYL-6-PHENYLPYRIMIDINE-5-CARBONITRILE DERIVATIVES AS ANTI-BREAST CANCER AGENTS

Fitria Eka Putri[1], Fajriah Azra[1*]

[1]Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, Indonesia

*Corresponding author: bunda_syasfa@yahoo.com

Abstract

A quantitative structure-activity relationship (QSAR) analysis was conducted on 15 derivatives of 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile to investigate the relationship between atomic charge and anti-breast cancer activity (MCF-7 cell line) based on experimental log IC_{50} values. Descriptor data were obtained through AM1 semi-empirical quantum mechanical calculations using HyperChem. A multiple linear regression model with the backward elimination method was employed to establish the quantitative relationship between net atomic charge and log IC_{50}, yielding the following QSAR model: log IC_{50} = -415.573 – 224.759(qC2) + 175.860(qC5) + 1307.672(qC6) + 1251.123(qC8) + 1142.590(qC9) – 3606.800(qC10) + 3.840(qC13); with n = 15, R = 0.899, R^2 = 0.808, F_o/F_t = 1.002, and PRESS = 0.043. The results indicate a statistically significant correlation between atomic charge and biological activity.

Keywords: *AM1, anticancer, breast cancer, QSAR*

Introduction

Cancer, or malignant tumors, remains a major global health challenge (Badawi et al., 2023). It arises from uncontrolled cell division caused by genetic mutations that induce abnormal cellular behavior (Thurston & Pysz, 2021). Among various types, breast cancer ranks as the most commonly diagnosed cancer worldwide, following lung cancer. In 2020, there were approximately 19.3 million new cancer cases, with breast cancer accounting for 11.7% of these diagnoses (Chhikara & Parang, 2023).

Breast cancer primarily affects the mammary glands, milk ducts, and surrounding supportive tissues (Jabeen et al., 2022). Several factors contribute to its development, including age, unhealthy lifestyle habits (such as excessive alcohol consumption, obesity, and smoking), genetic predispositions, and the timing of menopause and menarche (Subia et al., 2021). Current treatment modalities encompass chemotherapy, surgery, and radiotherapy. However, surgical and radiotherapeutic approaches are generally limited to localized tumors and are less effective once the cancer has metastasized (Pratama et al., 2022).

The search for more effective breast cancer therapies remains a priority, as no current drug or drug combination is entirely free from side effects. While breast cancer is the most prevalent cancer among women globally, only 39 of the 206 anticancer drugs approved by the U.S. Food and Drug

Administration (FDA) specifically target breast cancer (Chaurasia et al., 2023).

One such drug is 5-Fluorouracil (5-FU), a chemotherapy agent used extensively against breast, colon, head, and neck cancers. However, its clinical effectiveness is often compromised by drug resistance and significant side effects, including leukopenia—a reduction in white blood cell count (VanderVeen et al., 2020). The emergence of resistance to existing therapies underscores the urgent need for the discovery of novel anticancer compounds with improved efficacy and safety profiles.

Recent research has explored 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile, a pyrimidine derivative known for its broad spectrum of biological activities, including anticancer, analgesic, anticonvulsant, antimalarial, antifungal, antitubercular, anti-inflammatory, anti-HIV, antibacterial, and antidepressant properties (Natarajan et al., 2023).

The development of new drug candidates typically involves several laboratory stages, including design, synthesis, characterization, purification, and biological evaluation. This process is time-consuming, costly, energy-intensive, and can generate substantial chemical waste (Zain et al., 2020). Moreover, the potential for ineffective outcomes emphasizes the importance of early-stage screening methods, such as molecular modeling, which can predict compound activity prior to synthesis.

Computational chemistry plays a vital role in drug discovery, enabling the prediction of compounds' physicochemical properties and biological activities. One widely used technique is Quantitative Structure-Activity Relationship (QSAR) analysis, which examines correlations between molecular structures and their experimentally determined biological activities (Iswanto et al., 2023). A key step in QSAR involves structural modeling and geometry optimization, often performed using the AM1 semi-empirical quantum mechanical method. AM1 offers a computational balance that is faster than ab initio methods yet more accurate than molecular mechanics approaches. QSAR models typically consider steric, electronic, and lipophilic parameters. Steric parameters relate to molecular interactions with receptors; lipophilic parameters reflect membrane permeability; and electronic parameters, such as net atomic charge, influence receptor-ligand binding interactions (Aini et al., 2022; Jufri & Azra, 2023).

The electronic interactions between bonded atoms in a compound are influenced by the net atomic charge, which, in turn, is affected by the electron-binding characteristics of each atom. In this study, QSAR equations were developed for 15 derivatives of 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile, and the best-fitting model was selected based on correlation analysis for validation purposes (Jufri & Azra, 2023). Previous research by Badawi et al. (2023) involved the design, synthesis, characterization, and molecular docking studies of these compounds, highlighting several new pyrimidine candidates with promising activity against breast cancer cell lines, particularly MCF-7 and MDA-MB-231. Furthermore, their study demonstrated the inhibitory potential of these compounds on EGFR and aromatase (ARO) enzymes, both of which are critical targets in breast cancer therapy. However, the study did not include a QSAR analysis to quantitatively assess the relationship between molecular structure and biological activity. Therefore, further investigation using QSAR modeling is warranted. Accordingly, the present study examined the relationship between the net atomic charge and the biological activity of the selected derivatives against breast cancer cells. The net atomic charge was calculated using the AM1 semi-empirical method via HyperChem software, and multiple linear regression analysis was applied to construct the QSAR equation.

Methodology

Materials and Tools

This study utilized 15 derivative compounds of 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile, along with biological activity data (IC_{50} values) obtained from experimental results reported by Badawi et al. (2023). The tools employed included hardware and software components. The hardware consisted of a computer with an Intel® Core™ i3-8100 CPU @ 3.60 GHz and 3.87 GB usable RAM, available at the Information and Communication Technology (ICT) Laboratory of the Chemistry Department, Universitas Negeri Padang (UNP). Software used in this research included HyperChem Professional version 8.0 (Hypercube, 2007) for three-dimensional structure modeling, geometry optimization, and calculation of physicochemical properties. Microsoft Office Excel 2013 (Microsoft, 2013) and IBM SPSS Statistics for Windows (IBM, 2016) were utilized for data processing and statistical analysis.

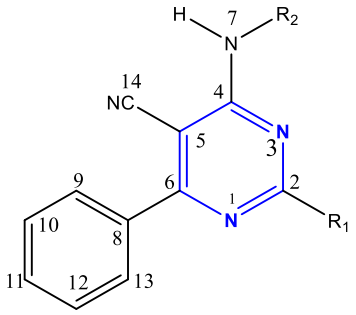
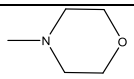
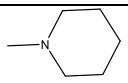
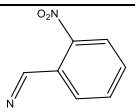
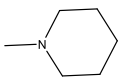
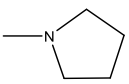
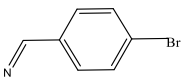
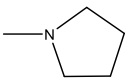
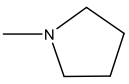
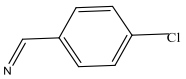
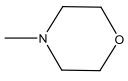
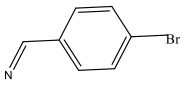
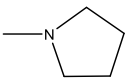
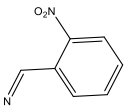
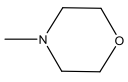
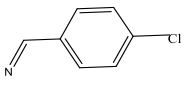
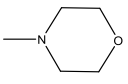
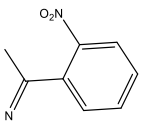
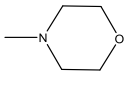
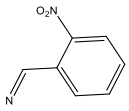
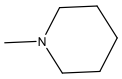
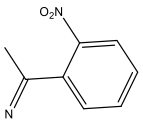
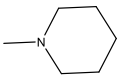
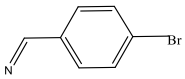
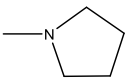
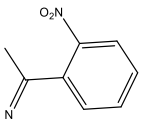
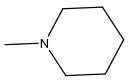
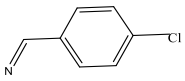


Figure 1. Structure of 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile.

Table 1. Structure and IC_{50} values of 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile derivatives

Code	R_1	R_2	IC_{50} (µg/ml)	Code	R_1	R_2	IC_{50} (µg/ml)
S1.		-NH_2	69.36 ± 3.07	S9			6.44 ± 0.29
S2.		-NH_2	18.26 ± 0.81	S10			12.91 ± 0.57
S3.		-NH_2	47.23 ± 2.09	S11			6.86 ± 0.3
S4.			42.72 ± 1.89	S12			2.13 ± 0.09
S5.			34.22 ± 1.52	S13			26.31 ± 1.17

H6.			2.19 ± 0.09	S14			0.87 ± 0.04
H7.			22.72 ± 1.01	S15			22.99 ± 1.02
S8.			2.74 ± 0.12				

Procedure

The molecular structures of the 15 derivative compounds were drawn and optimized using the HyperChem software, applying the Austin Model 1 (AM1) semi-empirical quantum chemical method. The optimization was carried out using the Polak-Ribiere algorithm with a convergence limit of 0.001, iteration limit of 32,767, and root-mean-square (RMS) gradient of 0.001. Multiple linear regression analysis was employed to determine the quantitative relationship between net atomic charge and biological activity (log IC_{50}) using the Backward elimination method. The model selection for the QSAR equation was based on criteria including high R, R^2, and adjusted R^2 values approaching 1 (Vikaliana et al., 2022), a low standard error (SE) value (Nindita & Sanjaya, 2014), an f-score/f-table ratio greater than 1, and a significance value below 0.05 (Slamet & Aglis, 2020). The final QSAR model was further validated by selecting the equation with the lowest Predicted Residual Sum of Squares (PRESS) value (Toni et al., 2022; Zain et al., 2020).

Results and Discussion

To investigate the relationship between the net atomic charge and the activity of 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile derivative compounds as potential anti-breast cancer agents, modeling was conducted using the AM1 semi-empirical quantum mechanical method.

Compound structure modeling and geometry optimization

The geometry optimization in this study aimed to obtain stable molecular structures characterized by minimum potential energy (Evita et al., 2022).

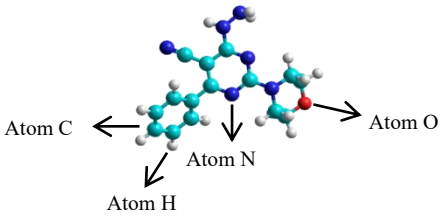


Figure 2. Structure of 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile derivative (S1) generated using the HyperChem application.

Geometry optimization is an iterative process whereby the molecular conformation is adjusted, and potential energy is recalculated until a convergence criterion is met. In this study, the convergence limit was set at 0.001 kcal/mol. Once this threshold was achieved, the conformation with the lowest potential energy was considered optimal (Rakhman et al., 2019)..

Table 2. Total energy values before and after geometry optimization of 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile derivatives

Compound Code	Total energy before optimization (kcal/mol)	Total energy after optimization (kcal/mol)
S1	-85256.09114	-85326.53026
S2	-81465.45408	-81536.80898
S3	-77873.77899	-77942.54270
S4	-114998.10004	-115066.27286
S5	-115470.06080	-115539.12302
S6	-126559.67694	-126707.28527
S7	-111207.57280	-111276.40499
S8	-111679.52970	-111749.24238
S9	-122769.15117	-122917.36405
S10	-107615.92479	-107681.95844
S11	-108087.88019	-108154.72266
S12	-119177.50410	-119322.91650
S13	-129207.36505	-130300.51195
S14	-125416.85220	-126510.81573
S15	-121825.20818	-122917.04220

The calculation of the potential energy for the 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile derivative compound (S1) showed a lower value after optimization (-85326 kJ/mol) compared to the value before optimization (-85256 kJ/mol), as presented in Table 2. This decrease in potential energy indicates that geometry optimization affected the bond angles and interatomic distances, leading to changes in energy until the lowest potential energy was achieved. This signifies that the molecular structure has reached a stable state (Rakhman et al., 2019).

Calculation results of net atomic charge descriptors

In this context, descriptors refer to physicochemical properties used to model the relationship between molecular structure and biological activity. Net atomic charge values of 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile derivatives were calculated as descriptors to assess their impact on anticancer activity. The charges analyzed included qN_1, qC_2, qN_3, qC_4, qC_5, qC_6, qN_7, qC_8, qC_9, qC_{10}, qC_{11}, qC_{12}, and qC_{13}. These net atomic charge values were treated as independent variables, as they significantly affected intermolecular interactions. Variations in the substituent groups affected the polarity of these atomic charges, thereby impacting their role in molecular activity (Ananto et al., 2020).

 | ISSN 2621-5985 (online) | ISSN 2549-385X (print)
Volume 7, Nomor 2, 2024

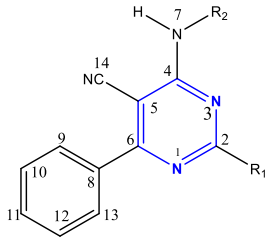


Figure 3. Structure of 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile compounds (R_1-R_2).

Table 3. Net atomic charge values of constituent atoms in 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile derivatives

Code	Net Atomic Charge												
	qN_1	qC_2	qN_3	qC_4	qC_5	qC_6	qN_7	qC_8	qC_9	qC_{10}	qC_{11}	qC_{12}	qC_{13}
S1	-0.242119	0.199524	-0.295979	0.165674	-0.208634	0.165536	-0.202998	-0.064279	-0.095572	-0.130942	-0.110439	-0.138182	-0.08641
S2	-0.240392	0.202018	-0.295341	0.163592	-0.211289	0.163547	-0.202326	-0.062672	-0.09595	-0.131399	-0.11146	-0.138714	-0.08634
S3	-0.244991	0.221809	-0.300584	0.166588	-0.223411	0.166716	-0.202661	-0.062764	-0.096379	-0.132331	-0.111754	-0.139331	-0.08514
S4	-0.240339	0.193871	-0.24445	0.249560	-0.229818	0.167424	-0.236957	-0.065328	-0.097576	-0.131808	-0.110252	-0.13799	-0.08417
S5	-0.240468	0.193521	-0.244464	0.249276	-0.229988	0.167173	-0.236244	-0.065134	-0.097596	-0.131855	-0.110375	-0.138046	-0.08422
S6	-0.240793	0.193437	-0.244654	0.250105	-0.230941	0.166706	-0.236704	-0.064221	-0.098352	-0.131712	-0.11067	-0.137659	-0.08400
S7	-0.239311	0.196642	-0.243128	0.247825	-0.232491	0.165659	-0.236893	-0.063573	-0.09786	-0.132162	-0.11126	-0.138312	-0.08436
S8	-0.239442	0.196324	-0.243136	0.247524	-0.232657	0.165408	-0.236166	-0.06338	-0.09788	-0.132207	-0.111385	-0.138368	-0.08441
S9	-0.239837	0.196267	-0.243146	0.248106	-0.233528	0.164958	-0.236283	-0.062454	-0.098644	-0.132053	-0.11167	-0.137963	0.08418
S10	-0.246159	0.218954	-0.249792	0.252826	-0.247020	0.169736	-0.23828	-0.062235	-0.097426	-0.132153	-0.112033	-0.13858	-0.08510
S11	-0.24766	0.220566	-0.252299	0.259251	-0.250364	0.170357	-0.245962	-0.061349	-0.098117	-0.132108	-0.112355	-0.138252	-0.08484
S12	-0.246397	0.218384	-0.247649	0.24792	-0.246173	0.168658	-0.231633	-0.061858	-0.097492	-0.132222	-0.112257	-0.138686	-0.08523
S13	-0.242215	0.192925	-0.244173	0.250557	-0.233184	0.166421	-0.238448	-0.064128	-0.097752	-0.131971	-0.110943	-0.138222	-0.08466
S14	-0.240216	0.19551	-0.241095	0.245356	-0.235151	0.163901	-0.23241	-0.061712	-0.098937	-0.132205	-0.112114	-0.138044	-0.08432
S15	-0.247334	0.215295	-0.243230	0.244325	-0.243189	0.16522	-0.220966	-0.059597	-0.097912	-0.132651	-0.113584	-0.139027	-0.08551

Table 3 illustrates that the N atoms (N_1, N_3, and N_7) possessed a negative charge due to their higher electronegativity than other atoms. This resulted in bonding electron pairs being drawn towards the N atoms. Conversely, the C atoms in heterocyclic rings (C_2, C_4, and C_6) carried a positive charge, while others exhibited a negative charge. For instance, in this study, the C_2 atom demonstrated a positive charge because it formed a double bond with the electronegative N_1 atom and a single bond with the N_3 atom, both of which strongly attracted electron density. Meanwhile, all C atoms in the phenyl ring (C_8, C_9, C_{10}, C_{11}, C_{12}, and C_{13}) exhibited a negative charge, as the C atoms are more electronegative than the H atoms, causing the bonding electrons to be pulled toward the C atoms (Asmara & Dwi, 2015).

Statistical analysis

Statistical analysis was conducted to determine the relationship between net atomic charge and the biological activity of 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile derivatives. This analysis was performed using SPSS statistical software. Correlation analysis and multiple linear regression were applied to develop a QSAR model. The correlation analysis aimed to assess whether a significant relationship exists between net atomic charge descriptors and the biological activity, expressed as log IC_{50} from experimental data. The independent variables included qN_1, qC_2, qN_3, qC_4, qC_5, qC_6, qN_7, qC_8, qC_9, qC_{10}, qC_{11}, qC_{12}, qC_{13}, while the dependent variable was the log of the experimental IC_{50}.

Multiple linear regression analysis yielded several candidate QSAR models using combinations of atomic charge descriptors. The selection of the best model was based on key statistical parameters, including the correlation coefficient (R), coefficient of determination (R^2), standard error (SE), significance level (sig), F-statistic (F_{score}/F_{table}), and Predictive Residual Sum of Squares (PRESS).

Table 4. QSAR Model Equations

Model	Equation	Descriptor
1	Log IC_{50} = (-460.325) + (60.816 * qN_1) - (286.999 * qC_2) + (10.72 * qN_3) + (186.354 * qC_5) + (1752.148 * qC_6) + (54.977 * qN_7) + (1585.045 * qC_8) + (1313.229 * qC_9) - (4539.889 * qC_{10}) + (501.643 * qC_{12}) + (4.385 * qC_{13})	qN_1, qC_2, qN_3, qC_5, qC_6, qN_7, qC_8, qC_9, qC_{10}, qC_{12}, qC_{13}
2	Log IC_{50} = (-413.646) - (238.521 * qC_2) + (15.516 * qN_3) + (198.427 * qC_5) + (1497.383 * qC_6) + (31.575 * qN_7) + (1376.432 * qC_8) + (1288.505 * qC_9) - (4291.912 * qC_{10}) + (595.909 * qC_{12}) + (4.224 * qC_{13})	qC_2, qN_3, qC_5, qC_6, qN_7, qC_8, qC_9, qC_{10}, qC_{12}, qC_{13}
3	Log IC_{50} = (-459.511) - (235.056 * qC_2) + (21.558 * qN_3) + (219.215 * qC_5) + (1548.494 * qC_6) + (28.805 * qN_7) + (1387.989 * qC_8) + (1010.196 * qC_9) - (3788.183 * qC_{10}) + (4.178 * qC_{13})	qC_2, qN_3, qC_5, qC_6, qN_7, qC_8, qC_9, qC_{10}, qC_{13}
4	Log IC_{50} = (-460.77) - (273.97 * qC_2) + (152.142 * qC_5) + (1515.646 * qC_6) + (40.876 * qN_7) + (1336.426 * qC_8) + (1082.989 * qC_9) - (3789.648 * qC_{10}) + (4.527 * qC_{13})	qC_2, qC_5, qC_6, qN_7, qC_8, qC_9, qC_{10}, qC_{13}
5	Log IC_{50} = (-415.573) - (224.759 * qC_2) + (175.486 * qC_5) + (1307.672 * qC_6) + (1251.123 * qC_8) + (1142.59 * qC_9) - (3606.8 * qC_{10}) + (3.84 * qC_{13})	qC_2, qC_5, qC_6, qC_8, qC_9, qC_{10}, qC_{13}
6	Log IC_{50} = (-399.917) - (207.338 * qC_2)+(176.922 * qC_5) + (1241.053 * qC_6) + (1190.187 * qC_8) + (926.211 * qC_9) - (3356.699 * qC_{10})	qC_2, qC_5, qC_6, qC_8, qC_9, qC_{10}

Table 5. Statistical Parameters of the QSAR Models

Model	R	R^2	Sig	F_{score}	F_{table}	F_{score}/ F_{table}	PRESS
1	0.919	0.844	0.415	1.480	8.763	0.168	48.6458
2	0.919	0.844	0.237	2.166	5.964	0.363	0.7324
3	0.918	0.843	0.120	2.992	4.772	0.627	0.7357
4	0.914	0.835	0.061	3.791	4.147	0.958	0.7760
5	0.899	0.808	0.039	4.196	3.787	1.108	0.9043
6	0.864	0.746	0.040	3.910	3.5805	1.092	1.1948

The analysis obtained from multiple linear regression using the backward method resulted in six QSAR equation models that met the specified criteria, each with R^2 value ≥ 0.746. This indicates that the independent variables (atomic charges) had a significant effect on breast anticancer activity, explaining more than 74.6% of the variance in activity (J. La Kilo & Kilo, 2019). The R-value represents the correlation between the dependent variable (log IC_{50}) and the independent variables (atomic charges),

while the R^2 value quantifies how well the variability in the dependent variable can be explained by the model (Ananto et al., 2020). The threshold of significance value (p-value) was set at 5% (0.05). Models 5 and 6 were statistically acceptable, with significance values of 0.039 and 0.040, respectively (A. La Kilo et al., 2019). Another evaluation parameter was the F_{score}/F_{table}. which assessed the overall effect of descriptors on the biological activity of the compound. Models 1 through 4 were deemed insignificant due to F_{score}/F_{table} values < 1, indicating that they did not meet the statistical criteria. Conversely, models 5 and 6 fulfilled this requirement, having F_{score}/F_{table} values ≥ 1. As a result, these two models were selected for further validation to identify the most optimal QSAR model (J. La Kilo & Kilo, 2019).

The PRESS value represents the sum of squared deviations between the experimental IC_{50} values and the values predicted by the model. A lower PRESS value indicates higher predictive accuracy, as the model's predictions closely approximate the actual biological activity (Fitriani et al., 2022). Models 5 and 6 exhibited the lowest PRESS values, 0.9043 and 1.1948, respectively. Based on these statistical parameters, model 5 was considered the most robust QSAR model, incorporating descriptors qC_2, qC_5, qC_6, qC_8, qC_9, qC_{10}, and qC_{13}.

Table 6. Experimental and Predicted Log IC_{50} Values with PRESS for QSAR Model 5

Code	Log IC_{50} (Experiment)	Log IC_{50} (Prediction)	Y-Y'	$(Y-Y')^2$
S1	1.841109084	1.765852715	0.075256369	0.005663521
S2	1.261500773	1.365553352	-0.104052579	0.010826939
S3	1.674217946	1.695090273	-0.020872327	0.000435654
S4	1.630631244	1.317716667	0.312914577	0.097915533
S5	1.534280005	1.427513847	0.106766158	0.011399013
S6	0.340444115	0.43204916	-0.091605045	0.008391484
S7	1.356408327	1.065116425	0.291291902	0.084850972
S8	0.437750563	1.159947018	-0.722196455	0.52156772
S9	0.808885867	0.809023993	-0.000138126	1.90787E-08
S10	1.110926242	0.966620101	0.144306141	0.020824262
S11	0.836324116	0.987197729	-0.150873613	0.022762847
S12	0.328379603	0.478335228	-0.149955625	0.022486689
S13	1.420120848	1.514332509	-0.094211661	0.008875837
S14	-0.060480747	-0.19311621	0.132635463	0.017592166
S15	1.361538971	1.09559841	0.265940561	0.070724382
			PRESS	0.904317039

Additionally, validation was conducted by analyzing the correlation between the experimental log IC_{50} values and the predicted log IC_{50} values. This correlation was illustrated through a curve with an R^2 value of 0.8075, which was close to 1. This high correlation confirms the model's ability to reliably predict the biological activity based on the selected descriptors (Fitriani et al., 2022).

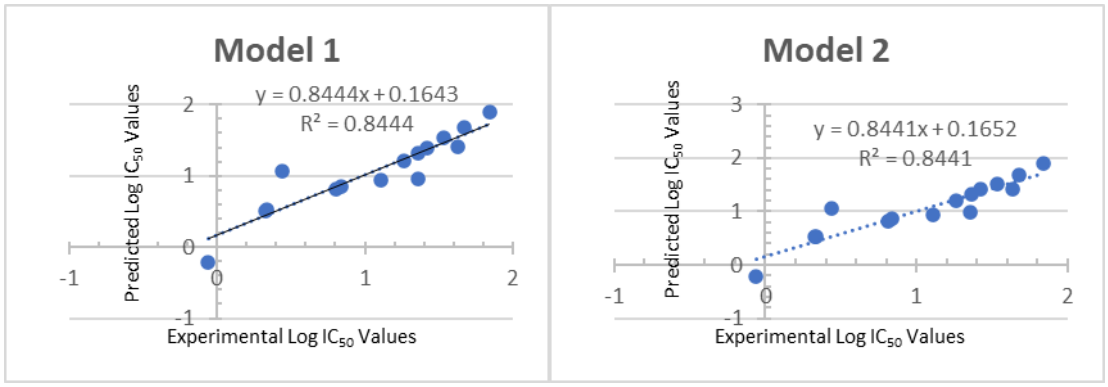


Figure 4. Correlation curve between experimental and predicted log IC_{50} values for models 1-2.

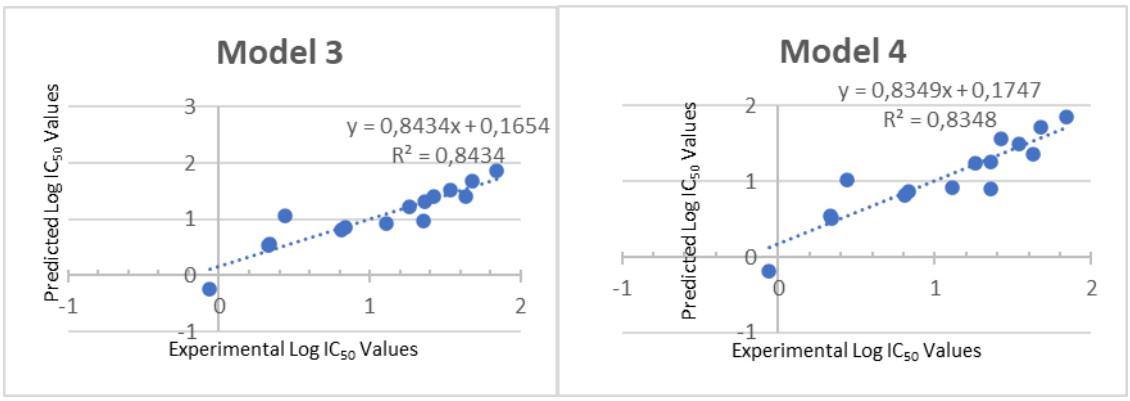


Figure 5. Correlation curve between experimental and predicted log IC_{50} values for models 3-4.

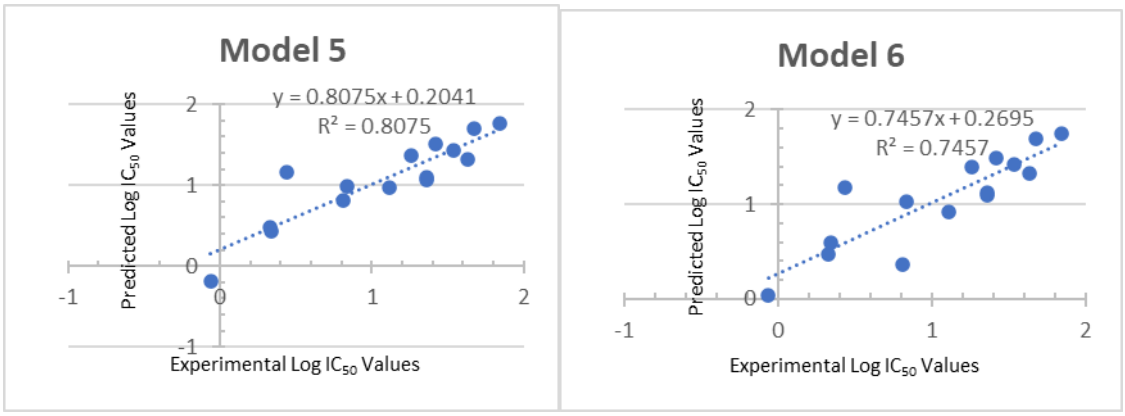


Figure 6. Correlation curve between experimental and predicted log IC_{50} values for models 5-6.

The graph in Figures 4-6 demonstrates that the 15 compounds had data points closely aligned with a linear trend line, indicating a strong agreement between the predicted and experimental biological activity values (Fitriani et al., 2022). Therefore, Model 5 was considered the most optimal QSAR equation model, as it possessed the lowest PRESS value and a correlation curve that fulfilled the necessary criteria. The equation was as follows: Log IC_{50} = (-415,573) - (224,759 * qC_2) + (175,486 * qC_5) + (1307,672 * qC_6) + (1251,123 * qC_8) + (1142,59 * qC_9) - (3606,8 * qC_{10}) + (3,84 * qC_{13}); with n = 15, R = 0.89, R^2 = 0.808, F_{score}/F_{table} = 1.002, and PRESS = 0.9043.

Based on the coefficient values in Model 5, the order of impact was as follows: $qC_{10} > qC_6 > qC_8 > qC_9 > qC_2 > qC_5 > qC_{13}$. The largest coefficients were associated with qC_{10} (3606.8), qC_6 (1307.672), qC_8 (1251.123), and qC_9 (1142.59), indicating that slight variations in atomic charge at these positions had a significant effect on the anticancer activity of the compound.

Furthermore, the QSAR model allows for the prediction of the extent to which each descriptor contributes to increasing or decreasing biological activity (IC_{50}). A positive coefficient suggests a higher IC_{50} value, indicating lower biological activity, whereas a negative coefficient implies a reduction in log IC_{50}, corresponding to improved predicted activity (Ngoc et al., 2023). The large negative coefficient of qC_{10} in Model 5 highlights its significant role in decreasing log IC_{50} and thus improving anticancer activity (Ngoc et al., 2023).

Based on the research conducted by Badawi et al. (2023), compound 14 was identified as having the highest anticancer activity among the tested compounds. Although its activity represented only 39% of erlotinib's ability to inhibit EGFR, it was recognized as the most effective EGFR inhibitor within the series. Consequently, compound 14 was selected for molecular docking with erlotinib to the EGFR crystal structure to examine their interactions with specific amino acid residues within the receptor. The crystal structure of EGFR kinase was retrieved from the Protein Data Bank (PDB ID: 1M17), which contains the tyrosine kinase domain of EGFR complexed with the inhibitor erlotinib.

Docking results revealed that compound 14 had a binding score of –6.71 kcal/mol and an RMSD value of 1.86 Å. Key interactions included a hydrogen bond between the oxygen atom of the 4-nitrophenylhydrazone moiety and the Lys721 residue, with an interaction distance of 3.27 Å. Additionally, a hydrogen bond was observed between the nitrile group at the 5-position of the pyrimidine ring and the Thr766 residue, with a distance of 3.50 Å. These interactions demonstrate how the compound engages with the EGFR active site, although its inhibitory efficacy remains lower compared to erlotinib.

5-Fluorouracil (5-FU) is a pyrimidine analog used as an anticancer agent, designed to inhibit the conversion of uridine to thymidine by substituting the hydrogen at the 5-position of uracil with a fluorine atom. This modification produces an antimetabolite that forms a stable covalent complex with thymidylate synthase (TS) and its cofactor tetrahydrofolate, thereby inhibiting DNA synthesis, as thymidine is not a constituent of RNA. TS catalyzes the methylation of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP), and 5-FU disrupts this process by forming ternary complexes with TS and the cofactor. Only about 20% of administered 5-FU is metabolically activated to 5-fluoro-2′-deoxyuridylate, the active form that inhibits TS, while the majority is degraded by dihydropyrimidine dehydrogenase (DPD), an enzyme responsible for 5-FU catabolism. Variations in DPD activity significantly affect the drug's bioavailability (Beale & Block, 2015).

In drug design, various factors must be considered when modifying molecular structures, as biological activity is closely related to the molecule's three-dimensional configuration and electronic properties. The introduction of substituents can significantly alter the physical and chemical properties of a compound. A well-constructed QSAR model can be valuable in designing new drug candidates with enhanced biological activity (Rodrigues et al., 2020).

Conclusion

The best equation in the QSAR analysis, conducted using 15 compounds of 4-hydrazinyl-6-phenylpyrimidine-5-carbonitrile derivatives and the AM1 semi-empirical method, was as follows: Log IC50 = (-415.573) - (224.759 * qC2) + (175.486 * qC5) + (1307.672 * qC6) + (1251.123 * qC8) + (1142.59 * qC9) - (3606.8 * qC10) + (3.84 * qC13); with n = 15, R = 0.899, R2 = 0.808, Fscore/Ftable = 1.002, and PRESS = 0.9043. Hence, the best descriptors of breast

anticancer activity were found to be the net atomic charges at positions C2, C5, C6, C8, C9, C10, and C13 (qC2, qC5, qC6, qC8, qC9, qC10, and qC13), which represented key electronic parameters.

References

Aini, N. S., Kharisma, V. D., Widyannada, Mu. H., Murtadlo, A. A. A., Probojati, R. T., Turisa, D. D. R., Tamam, M. B., Jakhmola, V., Yuniarti, E., Al-aziz, S., Ghifari, M. R., Purnamasari, D., Oktavia, B., Lubis, A. P., Azra, F., Fitri, F., Ansori, A., Rebezov, M., & Zainul, R. (2022). In silico screening of bioactivecompounds from Garcinia mongostana L.against SARS-CoV-2 via tetra inhibitors. *Pharmacogn*.

Ananto, A. D., Muliasari, H., & Hadisaputra, S. (2020). Design New Compound of Meisoindigo Derivative as Anti Breast Cancer Based on QSAR Approach. *Jurnal Kimia Sains Dan Aplikasi*, *23*(9), 305–311. https://doi.org/10.14710/jksa.23.9.305-311

Asmara, A. P., & Dwi. (2015). Analisis Hubungan Kuantitatif Struktur Dan Aktivitas Senyawa Turunan Triazolopiperazin Amida Menggunakan Metode Semiempirik AM1. *Elkawnie*, *1*(2), 125–138.

Badawi, W. A., Samir, M., Fathy, H. M., Okda, T. M., Noureldin, M. H., Atwa, G. M. K., & AboulWafa, O. M. (2023). Design, synthesis and molecular docking study of new pyrimidine-based hydrazones with selective anti-proliferative activity against MCF-7 and MDA-MB-231 human breast cancer cell lines. *Bioorganic Chemistry*, *138*(May), 106610. https://doi.org/10.1016/j.bioorg.2023.106610

Chaurasia, M., Singh, R., Sur, S., & Flora, S. J. S. (2023). A review of FDA approved drugs and their formulations for the treatment of breast cancer. *Frontiers in Pharmacology*, *14*(July). https://doi.org/10.3389/fphar.2023.1184472

Chhikara, B. S., & Parang, K. (2023). Global Cancer Statistics 2022: the trends projection analysis. *Chemical Biology Letters*, *10*(1), 1–16.

Evita, I. D., Banon, C., Yudha, S., & Adfa, M. (2022). Optimasi Hubungan Kuantitatif Struktur Senyawa Turunan Quercetin Yang Berpotensi Terhadap Aktivitas Antivirus Dengue dengan Metode Austin Model 1. *Chimica et Natura Acta*, *10*(2), 53–59. https://doi.org/10.24198/cna.v10.n2.39530

Fitriani, I. N., Sholahuddin, M., & Zikri, A. T. (2022). Analysis of Quantitative Structure-Activity Relationship (QSAR) Of 1 , 8- Napthalymide-4-Aminoquinoline Derivatives as Antimalarial Compounds. *Walisongo Journal of Chemistry*, *5*(2), 194–201.

Hypercube, I. (2007). *HyperChem Release 7*. Hypercube, Inc.

IBM. (2016). *IBM SPSS Statistics for Windows, Version 16.0*. IBM Corp.

Iswanto, P., Firdaus, I. M., Dafaulhaq, A. F., B, A. G. R., A, M. P. S., & C, H. E. (2023). Quantitative Structure-Activity Relationship of 3-Thiocyanate-1H-Indoles Derived Compounds as Antileukemia by AM1, PM3, and RM1 Methods. *Jurnal Kimia Sains Dan Aplikasi*, *26*(3), 109–117. https://doi.org/https://doi.org/10.14710/jksa.26.3.109-1178

Jabeen, K., Khan, M. A., Alhaisoni, M., Tariq, U., Zhang, Y. D., Hamza, A., Mickus, A., & Damaševičius, R. (2022). Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion. *Sensors*, *22*(3). https://doi.org/10.3390/s22030807

John M. Beale, & Block, J. H. (2015). *Wilson and Gisvold's Textbook of Organic Medicinal And Pharmaceutical Chemistry*. Lippincott Williams and Wilkins.

Jufri, Q. S., & Azra, F. (2023). Pemodelan Hubungan Muatan Atom Bersih dengan Aktivitas Senyawa Turunan Metronidazol Ariloksi sebagai Antikanker Payudara dengan Metode AM1. *Periodic*, *12*(2), 63–70.

Kilo, J. La, & Kilo, A. La. (2019). Kajian HKSA

Antimalaria Senyawa Turunan Quinolon-4(1H)-imines Menggunakan Metode MLR-ANN. *Jambura Journal of Chemistry*, *1*(1), 21–26. https://doi.org/10.34312/jambchem.v1i1.2104

La Kilo, A., Aman, L. O., Sabihi, I., & La Kilo, J. (2019). Studi Potensi Pirazolin Tersubstitusi 1-N dari Thiosemicarbazone sebagai Agen Antiamuba melalui Uji In Silico. *Indo. J. Chem. Res.*, *7*(1), 9–24. https://doi.org/10.30598//ijcr.2019.7-akr

Microsoft. (2013). *Microsoft Office Excel 2013*. Microsoft Corporation.

Natarajan, R., Anthoni Samy, H. N., Sivaperuman, A., & Subramani, A. (2023). Structure-Activity Relationships of Pyrimidine Derivatives and their Biological Activity - A Review. *Medicinal Chemistry*. https://doi.org/https://doi.org/10.2174/1573406418666220509100356

Ngoc, H., Nguyen, B., & Yudistira, M. (2023). Results in Chemistry Quantitative Structure-Activity Relationship (QSAR) modelling of the activity of anti-colorectal cancer agents featuring quantum chemical predictors and interaction terms. *Results in Chemistry*, *5*(September 2022), 100888. https://doi.org/10.1016/j.rechem.2023.100888

Pratama, P. R., Isman, F., & Fadlan, A. (2022). Penyelidikan Aktivitas Antikanker Payudara Oleh Minyak Atsiri Bunga Michelia ALba Secara In Silico. *Al-Kimiya : Jurnal Ilmu KImia Dan Terapan*, *9*(1), 1.

Rakhman, K. A., Limatahu, N. A., Karim, H. B., & Abdjan, M. I. (2019). Kajian Senyawa Turunan Benzopirazin sebagai Antimalaria Menggunakan Metode HKSA dan MLR. *EduChemia (Jurnal Kimia Dan Pendidikan)*, *4*(2), 112. https://doi.org/10.30870/educhemia.v4i2.4989

Rodrigues, G. C. S., Maia, M. D. S., Muratov, E. N., Scotti, L., & Scotti, M. T. (2020). Quantitative Structure-Activity Relationship Modeling and Docking of Monoterpenes with Insecticidal Activity against Reticulitermes chinensis Snyder and Drosophila melanogaster. *Journal Agric Food Chem*, *68*.

Subia, B., Dahiya, U. R., Mishra, S., Ayache, J., Casquillas, G. V., Caballero, D., Reis, R. L., & Kundu, S. C. (2021). Breast tumor-on-chip models: From disease modeling to personalized drug screening. *Journal of Controlled Release*, *331*(September 2020), 103–120. https://doi.org/10.1016/j.jconrel.2020.12.057

Thurston, D. E., & Pysz, I. (2021). Chemistry and Pharmacology of Anticancer Drugs. In *Jurnal Penelitian Pendidikan Guru Sekolah Dasar* (Vol. 6, Issue August). CRC Press.

Toni, N. P. D., Azra, F., & Maahury, M. F. (2022). Modeling the Relationship of Net Atomic Charge with the Activity of 5-Aminopyrazole Derivative Compounds as Antioxidants with AM1 Method. *EKSAKTA: Berkala Ilmiah Bidang MIPA*, *23*(04), 242–254.

VanderVeen, B. N., Sougiannis, A. T., Velazquez, K. T., Carson, J. A., Fan, D., & Murphy, E. A. (2020). The Acute Effects of 5 Fluorouracil on Skeletal Muscle Resident and Infiltrating Immune Cells in Mice. *Frontiers in Physiology*, *11*(December), 1–15. https://doi.org/10.3389/fphys.2020.593468

Zain, D. N., Amin, S., & Indra. (2020). Hubungan Kuantitatif Struktur-Aktivitas Fungisida Turunan 1,2,4-Thiadiazolin Menggunakan Metode Austin Model 1 (AM1). *Journal of Pharmacopolium*, *53*(9), 1689–1699.