Design of New 2,4-Substituted Furo [3,2-B] Indole Derivatives as Anticancer Compounds Using Quantitative Structure-Activity Relationship (QSAR) and Molecular Docking

Jufrizal Syahri[1*], Nurul Hidayah[1], Rahmiwati Hilma[1], Beta Achromi Nurohmah[2], Emmy Yuanita[3]

[1]Department of Chemistry, Universitas Muhammadiyah Riau, Jalan Tuanku Tambusai, Pekanbaru, Indonesia, Tel/Fax. +62-761-35008
[2]Department of Chemistry, Universitas Gadjah Mada, Jalan Kaliurang Sekip Utara Bulaksumur 21, 55281, Yogyakarta, Indonesia, Tel/Fax: +62-274-545188 / 513339
[3]Department of Chemistry, Universitas Mataram, Jalan Majapahit 62A, Mataram, Indonesia, Telp/Fax. +62-370-646506

*Corresponding author email: jsyachri@umri.ac.id

ABSTRACT. This study aimed to propose new indole derivatives as anticancer through Quantitative Structure-Activity Relationship (QSAR) and molecular docking method. The best predicted anticancer activity of indole derivatives was recommended based on the QSAR equation. A data set consist of 18 indole derivatives from literature with anticancer activity against the A498 cell line was used to generate a QSAR model equation. The data set was divided randomly into training (14) and test (4) set compounds. The structure of indole compound was optimized first using AM1 semi-empirical methods, and the descriptors involved were analyzed using Multiple Linear Regression (MLR). The best QSAR equation obtained was Log IC_{50} = 65.596 (qC2) + 366.764 (qC6) – 92.742 (qC11) + 503.297 (HOMO) – 492.550 (LUMO) – 76.966. Based on the QSAR model, varying electron-withdrawing groups in C2 and C6 atom, as well as adding electron-donating groups in C11 were proposed could increase the anticancer activity of the indole derivatives. The QSAR analysis showed that compound **15** has the best predicted anticancer activity, supported by molecular docking results that showed hydrogen bond interaction with essential amino acids to build anticancer activity such as MET769, THR830, and THR766 residues.

Keywords: AM1 semi-empirical, Anticancer, Docking, Indole, QSAR

INTRODUCTION

Cancer is one of the leading causes of death in the world. One common sign of cancer is the continual uncontrolled cell division that can invade normal tissue and organs, then spread throughout the other parts of the body. In 2018, WHO reported 18.1 million new cancer cases and 9.6 million of mortality globally. The largest number of deaths was reported caused by lung cancer (1.8 million deaths), followed by colorectal cancer (881,000 deaths), stomach cancer (783,000 deaths), liver cancer (782,000 deaths), breast cancer (627,000 deaths), and esophageal cancer (400,000 deaths). The number of worldwide cancer deaths is expected to increase in every year (WHO, 2018). In Indonesia, the prevalence of cancer is quite high. According to the Basic Health Research (Riskesdas) data by The Health Ministry, the incidence of cancer in Indonesia was around 362,000 people in 2018. The highest rate of death was caused by failure for early detection and improper cancer treatment that requires one or more stages of treatment, such as surgery, radiotherapy, and chemotherapy.

Some studies have reported that xanthone, benzylchroman, naphthoquinone, quinolones, thiosemicarbazone, and tetrazine derivatives have anticancer activity (Yuanita et al., 2016; Simon et al., 2017; Prachayasittikul et al., 2017; Shaikh et al., 2017; Bertosa et al., 2010; Wang et al., 2017; Hu et al., 2004). The main drawback of their studies was the compounds were less potent than the current standard drug doxorubicin. Indole derivative is one of the organic compounds that has been reported to have excellent anticancer activity (Zhuang et al., 2013). Indole is one of the alkaloid groups with many biological activities such as antitubercular, antioxidant, antibacterial, anticancer, an inhibitor of hepatitis C, leukemia, and Alzheimer's (Zhuang et al., 2013; Khan et al., 2017; Estevao et al., 2010; Han et al., 2016; Karimabad et al., 2017; Sreenivasachary et al., 2017).

The Quantitative Structure-Activity Relationship (QSAR) method has been widely used in the design of new drug compounds to predict biological activity statistically, along with molecular docking. The computational chemistry methods,

such as QSAR and molecular docking, could significantly lowering the cost and time consumed during the drug development process. Those methods allow the researcher to only synthesis selected compounds with the best predicted biological activity and do further studies to the promising compounds.

In this study, an attempt to find a better anticancer activity of indole derivatives was carried out using QSAR and molecular docking. A QSAR equation was used to determine and propose what kind of substituents should be added to the active site of indole compounds to get the best predicted anticancer activity). Meanwhile, molecular docking was performed to study the intermolecular configuration complexes of indole derivatives to the active site of protein EGFR-TK (Epidermal Growth Factor Receptor-Tyrosine Kinase).
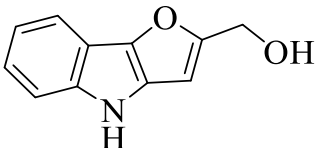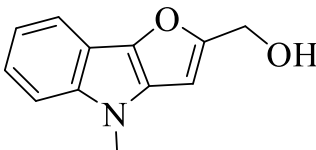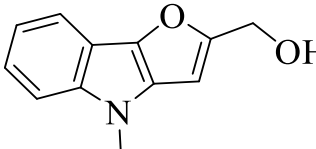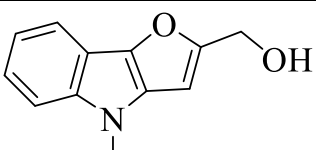
EXPERIMENTAL SECTION
Instrumentation
The computer specifications used in this study were: Intel Core i7 4.00 GHz; 4GB RAM; 64-bit system type; Operating System: Windows® 7 Ultimate. The programs used in QSAR and molecular docking were *Gaussian 09*, *ChemDraw Professional* 15, and Statistical Package for the Social Sciences (*SPSS*) *23.0* software.
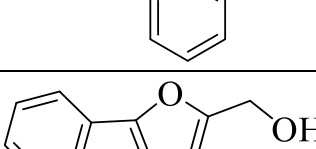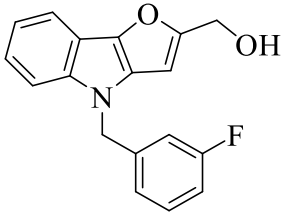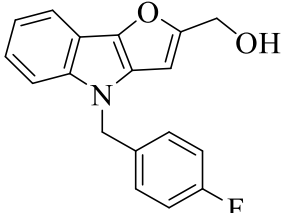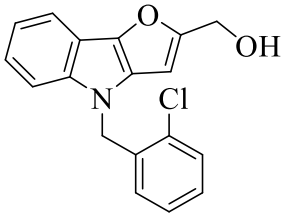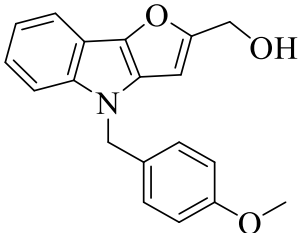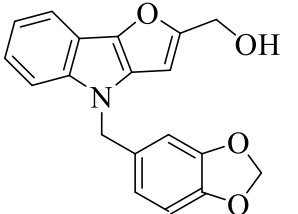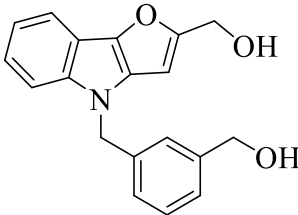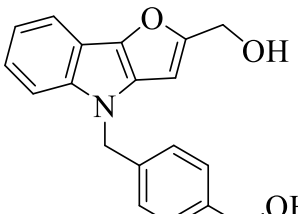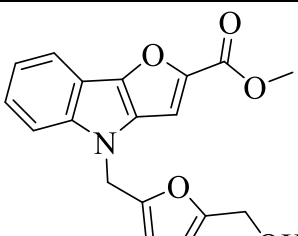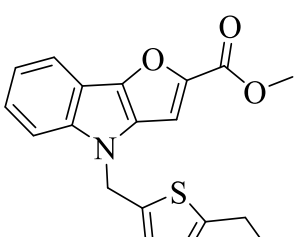
Data
The data set used in the QSAR study were 18 indole derivatives with anticancer activities against the A498 cell line (**Table 1**) (Zhuang et al., 2013). This data s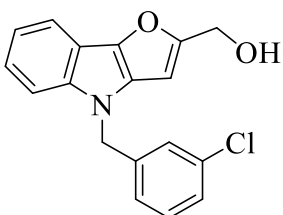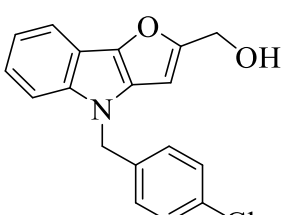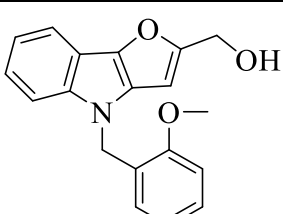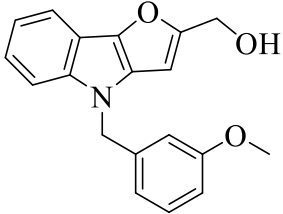et was divided into two parts, i.e., a training (14 compounds) and test (4 compounds) set. Anticancer activity that was expressed in half-maximal inhibitory concentration (IC_{50}) was converted into logarithms (log IC_{50}) before the QSAR analysis to get a better distribution of the activity val ues.

Table 1. Data set of indole derivatives and their anticancer activity against A498 cell line

No	Molecular structures	IC_{50} (μM)	log IC_{50}
2		21.8	1.338
7a[a]		4.6	0.663
7b[a]		3.7	0.568
7c		0.24	-0.620
7d		0.43	-0.367

No	Molecular structures	IC50 (µM)	log IC50
7e		0.59	-0.229
7f		0.69	-0.161
7g		5.5	0.740
7h		1.1	0.041
7i		0.79	-0.102
7j		4.3	0.633
7k		0.33	-0.481

No	Molecular structures	IC50 (µM)	log IC50
7l		0.36	-0.444
7m		0.27	-0.569
7n		0.37	-0.432
7o[a]		0.47	-0.328
10a[a]		0.21	-0.678
10b		0.69	-0.161

[a]) test set

Determination of the Best Optimization Method

One of the compounds in **Table 1** (**7h**) was optimized using the *Austin Model 1* (AM1) semi-empirical, the *Parameterized Model 3* (PM3) and the *ab initio Hartree-Fock* (HF) method in *Gaussian 09* software to calculate the ^{1}H-NMR data. The ^{1}H-NMR data from the calculations (calc.) were then compared with the experimental H-NMR data (exp.). The method that produces the smallest PRESS (*Predictive Residual Sum of Square*) value or similar result with the experimental data was used as the best method to optimize the structure geometry of indole derivatives.

Geometry Optimization

The molecular structure of the eighteen indole derivative compounds was drawn in *ChemDraw Professional 15.0* software. The initial optimization was performed using *Chem3D 15.0,* followed by re-optimization using the best optimization method in *Gaussian 09* software. The optimized structure geometry will produce electronic descriptors in the form of atomic charge, HOMO, and LUMO energy data.

Multiple Linear Regression Analysis

Multiple-Linear Regression (MLR) method is used to establish the correlation of anticancer activity (dependent variable) and electronic descriptors (independent variables) statistically. In this work, the determination of the significant descriptors that influence the anticancer activity was calculated using MLR analysis in *SPSS 23.0* program with a backward method.

Several models were produced from the MLR analysis. The validation of the best model was carried out by using predetermined statistical parameters, such as $R^2 > 0.6$; PRESS; and $F_{calc.}/F_{tab.} \geq 1$ (Frimayanti et al., 2011). The QSAR models that fit with the statistical criteria were then validated using the test set data to generate the best QSAR equation model.

Design of The Proposed Compound

The proposed compound was designed by replacing substituents in the indole derivatives structure based on the best QSAR equation model. The IC_{50} value of the new proposed anticancer compound can be calculated from the QSAR equation and referred to as the predicted IC_{50} value.

Molecular Docking

The biological activity of a compound as an anticancer can be identified by its ability to inhibit the EGFR-TK phosphorylation (Mphahlele et al., 2018; Prajapati et al., 2017). In this study, molecular docking was performed to the crystallographic protein structure of molecular target EGFR-TK with Protein Data Bank (PDB) ID: 1M17 (resolution 2.6 Å). Molecular docking was carried out using *Discovery Studio* software following the standard protocol implemented in the *Discovery Studio* (Accelrys, San Diego, USA).

RESULTS AND DISCUSSION
Determination of the Best Optimization Method

Calculation of the ^{1}H-NMR chemical shift of compound **7h** was performed using AM1, PM3, and HF methods. Based on the result in **Table 2**, the AM1 method was the best optimization method with the closest chemical shift (δ) results to the experimental data. This result was determined by comparing the PRESS value of the AM1 (1.68) that was smaller than the PM3 (3.30) and HF (5.05) method. Therefore, the AM1 method was used to optimize the structure and calculate the overall descriptors of the compound series.

Table 2. ^{1}H-NMR Chemical shifting values (δ, ppm) of **7h** from the experimental and calculations results using the AM1, PM3, and HF methods.

No H Atom	δ Experimental	δ AM1	δ PM3	δ HF
28	4.53	4.67	4.20	4.12
27	5.38	5.53	4.51	4.95
31	5.5	6.56	6.51	6.16
24	6.7	6.66	6.59	6.23
23	7.09	6.83	6.66	6.25
26	7.15	6.84	6.73	6.30
36	7.2	6.89	6.78	6.45
33	7.25	7.02	6.84	6.53
34	7.3	7.14	6.84	6.69
35	7.34	7.22	7.12	6.72
30	7.62	7.33	7.14	6..97
25	7.64	7.33	7.14	6.97
Correlation (r)		0.93	0.89	0.92
PRESS		1.68	3.30	5.05

Determination of the Best QSAR Model

Multiple-Linear Regression (MLR) statistical calculations with the backward method provide two QSAR models (**Table 3**). The two models illustrate an excellent correlation between biological activity and descriptors that can be seen from the value of the statistical parameters of each model (**Table 3**).

Correlation of the experimental log IC_{50} vs. the predicted log IC_{50} to the training set compound from both models was presented in **Figure 1**. According to **Figure 1**, Model **2** gives a better correlation value (R^2) with 0.9314 than Model **1** (0.9234). Therefore, both models were then validated using the test set data.

Validation of the QSAR Model

The validation of the best model was carried out by calculating the PRESS value from the test sets compounds. The smaller PRESS value gave a better QSAR model. The PRESS value of the test set data based on Models **1** and **2** can be seen in **Table 4**. The correlation coefficient (R^2) of the predicted and the experimental log IC_{50} for the QSAR model to the test set was displayed in **Figure 2**.

Table 4 showed that Model **1** has a smaller PRESS value (0.360) than Model **2** (0.401). However, **Figure 2** showed a predictive correlation coefficient R^2 value of Model **2** to the test set (0.9544) was better than Model **1** (0.9181). Moreover, Model **1** has more descriptors (6 descriptors) involved in the equation than Model 2 (5 descriptors).

Based on the analysis of Model **2**, the independent variables such as atomic charges on C2 (qC2), C6 (qC6), and C11 atom (qC11), as well as HOMO and LUMO can affect 95.4% of the anticancer activity. This correlation was better than Model 1 obtained with a correlation of 91.8%. Therefore, it can be concluded that Model **2** was the best model for designing new indole compounds with better anticancer activity. The complete QSAR equation of Model 2 was:

Log IC_{50} = 65.596 (qC2) + 366.764 (qC6) – 92.742 (qC11) + 503.297 (HOMO) – 492.550 (LUMO) – 76.966

Table 3. The QSAR equation models and their statistical parameters from the MLR analysis

Model	Variables	R^2	SEE	F_{calc}/F_{tab}	PRESS
1	LUMO, qC6, qC2, qC11, qO12, HOMO	0.976	0.180	6.123	0.372
2	LUMO, qC6, qC2, qC11, HOMO	0.964	0.206	5.439	0.374
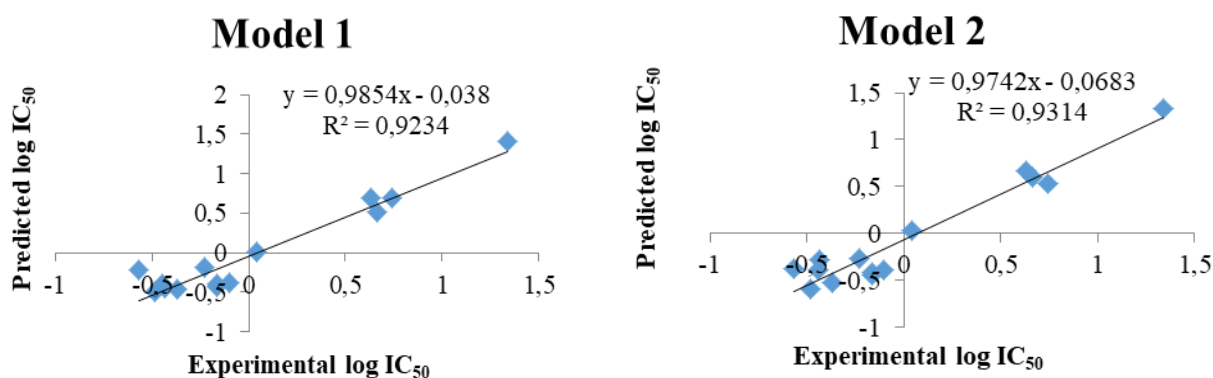
Figure 1. The plot of log IC_{50} of the training set from the experimental and prediction result using Model **1** and **2**

Table 4. The experimental and prediction of IC_{50} of the test set using Model **1** and **2**

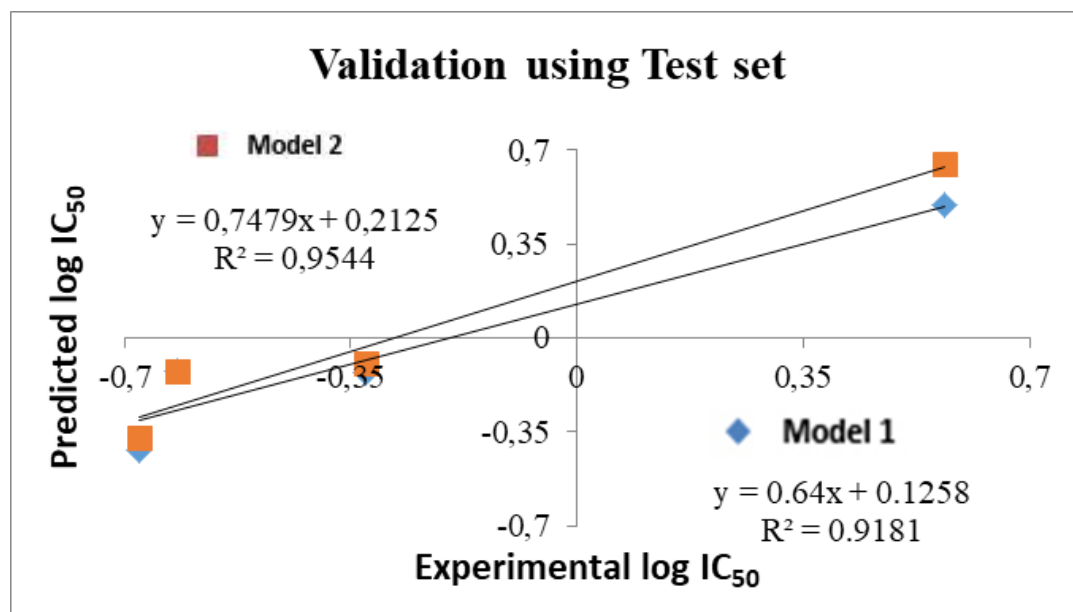| Compounds	Experimental log IC_{50}	Predicted log IC_{50}	
		Model **1**	Model **2**
7c	-0.619	-0.121	-0.122
10a	-0.677	-0.415	-0.371
7o	-0.327	-0.132	-0.095
7b	0.568	0.496	0.649
PRESS		0.360	0.401

Figure 2. The plot of log IC$_{50}$ of the test set from the experimental and prediction result using Model 1 and 2

Designing of The Proposed Compound

The new anticancer indole compounds were designed by replacing substituents in indole structures based on the QSAR equation. From this study, it is expected that the designed compounds have better anticancer activity (IC$_{50}$) than the previous reported indole derivative compounds. Modification of the compound by varying the substituents need to consider the electronic, hydrophobic, and steric properties of the atoms or substituent added. The design of the proposed structure of indole can be done by adding an electron-withdrawing or donating group.

The electron-withdrawing group has electrophilic properties that can decrease the electron density of the atom attached to it and make it more positive. Meanwhile, the electron-donating group has nucleophilic properties, which result in an increasing electron density of the atom. For example, the addition of an electron-withdrawing group to C6 atom in the indole compound instigates the resonance from C6 towards the electron-withdrawing group so that C6 acts as a nucleophile with a higher electron density.
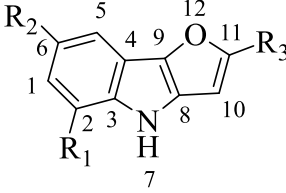
This study presented the design of new indole derivatives and their predicted anticancer activity. The designed compound with the best predicted anticancer activity was obtained by modifying the functional group in the main structure of indole according to the QSAR equation (**Table 5**).

Table 5 showed that the electron-withdrawing groups in the position R1, R2, and electron-donating group in R3 significantly could increase the anticancer activity of the compound. According to

the QSAR equation obtained, electron-withdrawing in R1 and R2, as well as electron-donating in R3, could make the coefficient value of the atomic charges of C2, C6, and C11 were negative. The more negative log IC$_{50}$ obtained, the better the predicted anticancer activity of indole derivatives. Based on the result, compounds **14** and **15** gave the best predicted anticancer activity, expected by the presence of a halide group in R3 that has high electronegativity properties.

Molecular docking was performed to support the QSAR result by predicting the interaction of compound **15** to crystal protein of Epidermal Growth Factor Receptors (EGFR) tyrosine kinases (PDB ID: 1M17) with co-crystal ligand AQ4. Protein EGFR is a large family of receptor tyrosine kinases that has a significant role in the emergence of breast, lung, esophageal, head, and neck cancer. This protein was also reported as the key to the viability of the cancer cell so it can be a target molecule in docking study. According to the docking result, the Root Mean Square Deviation (RMSD) value obtained was 1.8521 Å.

Molecular docking of compound **15** to the 1M17 protein displayed interaction with seven amino acid, and three of them were hydrogen bonds with MET769, THR830, and THR766 (Figure 3). These three amino acid residues were revealed to act as the active site of anticancer activity formed by co-crystal ligand AQ4 (Stamos et al., 2002). A molecular docking study showed that binding pocket formed by compound **15** was similar to co-crystal ligand AQ4. Thus, it was predicted that compound **15** has an excellent anticancer activity based on the QSAR and molecular docking results.

15

Table 5. The predicted anticancer activity of the designed indole derivatives against the A498 cell line

Compound	R_1	R_2	R_3	log IC_{50} (μM)	IC_{50} (μM)
11	CF_3	CH_3	OCH_3	-1.022	0.095
12	CH_3	$COCH_3$	OCH_3	-1.260	0.055
13	H	CF_3	NH_2	-2.000	0.010
14	$COCH_3$	CF_3	F	-2.301	0.005
15	COOH	CF_3	Cl	-2.523	0.003
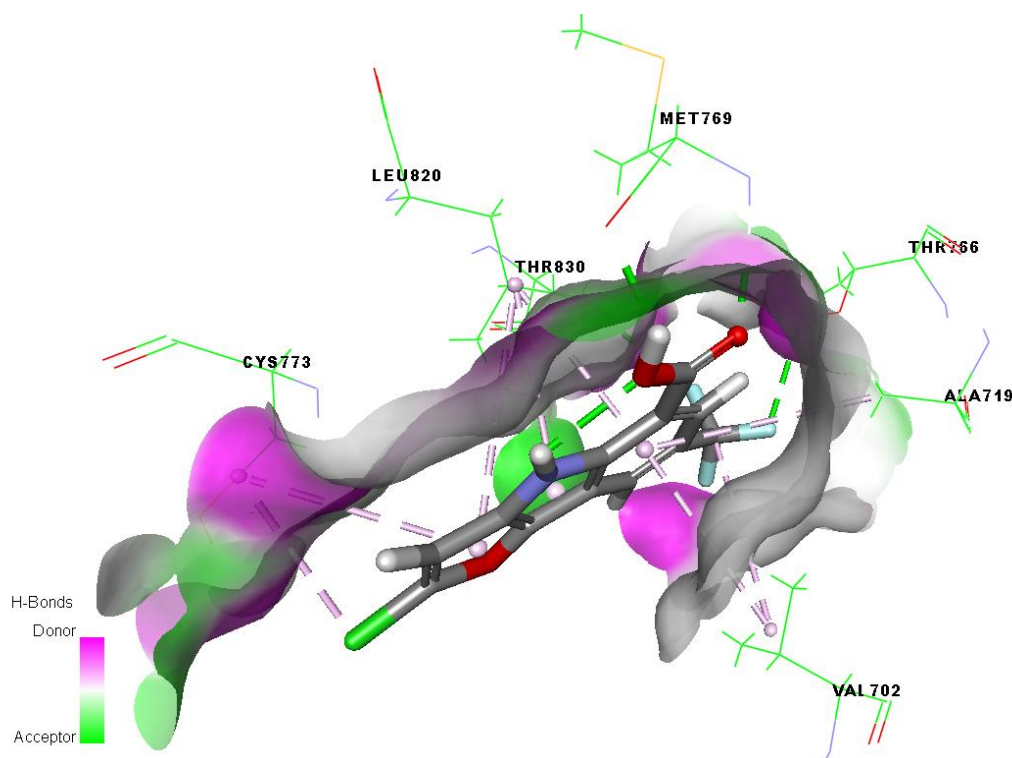
Figure 3. Binding interaction from molecular docking simulation of compound **15** into the active site of protein EGFR-TK (PDB ID: 1M17). The coloring atom for the compound is in order as follows: carbons in black, oxygen in red, nitrogen in blue, and hydrogen in white. The green line indicates hydrogen-bonding interaction with distance ascribed in angstroms, Å.

CONCLUSIONS

Based on this study, the relationship of the structure of indole derivative compounds with anticancer activities can be modeled using QSAR analysis with the best QSAR equation model obtained was:

Log IC_{50} = 65.596 (qC2) + 366.764 (qC6) – 92.742 (qC11) + 503.297 (HOMO) – 492.550 (LUMO) – 76.966

Designing of the new indole compound using QSAR and molecular docking showed the best predicted anticancer activity by compound **15** with IC_{50} of 0.003 μM and displayed hydrogen bonds interaction to the essential amino acid MET769, THR830, and THR766.

REFERENCES

Bertosa, B., Aleksic, M., Karminski-Zamola, G., & Tomic, S. (2010). QSAR analysis of antitumor active amides and quinolones from thiophene series. *International Journal of Pharmaceutics, 394*, 106-114.

Estevao, M.S., Carvalho, L.C., Ribelro, D., Couto, D., Freltas, M., Gomes, A., Ferreira, L.M., Fernandes, E., & Maerques, M.M.B. (2010). Antioxidant activity of unexplored indol derivatives: synthesis and screening. *European Journal of Medicinal Chemisty, 45*, 4869-4878.

Frimayanti, N., Yam, M.L., Lee, H.B., Othman, R., Zain, S.M., & Rahman, A.N. (2011). Validation of Quantitative Structure-Activity Relationship (QSAR) model for photosensitizer activity prediction. *International Journal of Molecular Science, 12*, 8626-8644.

Han, Z., Liang, X., Wang, Y., Qing, J., Cao, L., Shang, L., & Yin, Z. (2016). The discovery of indole derivates as novel hepatitis C virus inhibitor. *European Journal of Medicinal Chemistry, 116*, 147-155.

Hu, W.X., Rao, G.W., & Sun, Y.Q. (2004). Synthesis and antitumor activity of s-tetrazine derivates. *Bioorganic & Medicinal Chemistry Letter, 14*, 1177-1181.

Karimabad, M.N., Mahmoodi, M., Jafarzadeh, A., Darehkordi, A., Hajizadeh, M.R., Khorramdelazad, H., Sayadi, A.R., Rahmani, F., & Hassanshahi, G. (2017). Evaluating of OCT-4 and NANOG was differentially regulated by a new derivative indole in leukimia cell line. *Immunology Letterr, 190*, 7-14.

Khan, G.A., War, J.A., Kumar, A., Sheikh, I.A., Saxena, A., & Das, R. (2017). A facile synthesis of novel indole derivatives as potential antitubercular agents. *Journal of Taibah University for Science*, 11, 910-921.

Mphahlele MJ, Mmonwa MM, Aro A, McGaw LJ, & Choong YS. (2018). Synthesis, biological evaluation and molecular docking of novel indole-aminoquinazoline hybrids for anticancer properties, *International Journal of Molecular Science*, 19, 2232.

Prachayasittikul, V., Pingaew, R., Worachartcheewan, A., Nantasenamat, C., Prachayasittikul, S., Ruchirawan, S., & Prachayasittikul, V. (2014). Synthesis, anticancer activity and QSAR study of 1,4-naphthoquinone derivatives. *European Journal of Medicinal Chemistry, 84*, 247-263.

Prajapati C & Reddy MN. (2017). Molecular docking studies of canthin-6-one from simarouba glauca against EGFR tyrosine kinase, *International Journal of Pharmaceutical Science and Research, 8*, 5130-5136.

Shaikh, S.K.J., Kamble, R.R., Somagond, S.M., Devarajegowda, H.C., Dixit, S.R., & Joshi, S.D. (2017). Tetrazolymethyl quinolones: design, docking studies, synthesis, anticancer and antifungal analyses. *European Journal of Medicinal Chemistry, 128*, 258-273.

Shi, Z., Zhao, Z., Huang, M., & Fu, X. (2015). Ultrasound-assisted, one-pot, three-component synthesis and antibacterial activities of novel indole derivatives containing 1,3,4-oxadiazole and 1,2,4- triazole moieties. *Comptes Rendus Chimie, 18*, 1320-1327.

Simon, L., Salam, A.A.A., Kumar, S.M., Shilpa, T., Srinivasan, K.K., & Byiappa, K. (2017). Synthesis, anticancer, structural, and computational docking studies of 3-benzylchroman-4-one derivatives. *Bioorganic & Medicinal Chemistry Letter, 27*, 5284-5290.

Sreenivasachary, N., Khort, H., Benderitter, P., Hamel, A., Varisco, Y., Hickman, D.T., Froestl, W., Pfeifer, A., & Muhs, A. (2017). Discovery and characterization of novel indole and 7-azaindole derivatives as inhibitors of β-amyloid-42 aggregation for the treatment of Alzheimer's disease. *Bioorganic & Medicinal Chemistry Letter, 27,* 1405-1411.

Stamos J, Sliwkowski MX, & Eigenbrot C. (2002). Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *Journal of Biological Chemistry, 277*, 46265–46272.

Wang, Y., Gu, W., Liu, F., Xu, X., Yang, Y., Zhang, Q., Zhang, Y., Kuang, H., Wang, Z., & Wang, S. (2017). Design, synthesis and anticancer activity of novel nopinone-based thiosemicarbazone derivatives. *Bioorganic & Medicinal Chemistry Letter, 27*, 2360-2363.

WHO. (2018). World Cancer Report 2018: International Agency for Research on Cancer. USA.

Yuanita, E., Pranowo, H.D., Jumina, J., & Mustofa, M. (2016). Design of hydroxy xanthones derivatives as anticancer using quantitative structure activity relationship. *Asian Journal of Pharmaceutical and Clinical Research, 9*, 180-185.

Zhuang, S.H., Lin, Y.C., Chou, L.C., Hsu, M.H., Lin, H.Y., Huang, C.H., Lien, J.C., Lin, J.C., Kuo, S.C., & Huang, L.J. (2013). Synthesis and anticancer activity of 2,4- disubstituted furo[3,2-b] indole derivatives. *European Journal of Medicinal Chemistry, 66*, 466-479.