Indian Journal of Chemistry
Vol. 60B, October 2021, pp. 1385-1395

Insight QSDAR models for prediction of anticancer activity on Hela cell line of new flavonoid isolating from rhizome *Zingiber zerumbet* SM in Viet Nam

Bui Thi Phuong Thuy[b], Nguyen Minh Quang[c], Nguyen Hung Huy[d] & Pham Van Tat*[a]

[a] Faculty of Health Science, Hoa Sen University, Ho Chi Minh City, Viet Nam

[b] Faculty of Basic Sciences, Van Lang University, Ho Chi Minh City, Viet Nam

[c] Faculty of Chemical Engineering, Industrial University of Ho Chi Minh City, Ho Chi Minh City,Viet Nam

[d] Department of Chemistry, Hanoi University of Science, National University of Ha Noi, Viet Nam

E-mail: vantat@gmail.com

Received 4 March 2020; accepted (revised) 11 August 2021

This research predicts the anticancer activity on the Hela cell line of new flavonoid kaempferol-3-O-methyl ether isolating from rhizome *Zingiber zerumbet* SM by using the spectrum data activity relationship (QSDAR) models. This model has been developed for a set of 3-aminoflavonoids based on the simulated-spectral data ^{13}C NMR and ^{15}O NMR resulting from the semi-empirical quantum chemical calculations TNDO/2 SCF. The atomic sites O_1, O_{11}, C_2, C_3, C_6, C_7, and $C_{2'}$ in the QSDAR models significantly contribute to anticancer activity resulting from the Genetic algorithm (GA). The best regression model $QSDAR_{MLR}$ with the values R^2_{train} of 0.9057 and R^2_{test} of 0.7137, and the neural network model $QSDAR_{ANN}$ I(7)-HL(9)-O(1) with the values R^2_{train} of 0.993 and R^2_{pred} of 0.971 have been explored to predict the anticancer activities on Hela cell line for new flavonoid kaempferol-3-O-methyl ether from rhizome *Zingiber zerumbet* SM in Viet Nam.

Keywords: $QSDAR_{MLR}$ model, $QSDAR_{ANN}$ model, quantitative spectrum data - activity relationship, chemical-shift data, anticancer activities, Hela cell line

Natural products from plants are interested in searching for new anticancer drugs and can directly affect the Hela cell line and reduce side effects. Recently, we have isolated a few flavonoids from rhizome *Zingiber zerumbet* SM proposed by Do Tat Loi 2006[1] and tested *in vitro* activities pointed out the relatively substantial impacts for cancer cells Hela[2]. Flavonoids are polyphenolic compounds in most plants[3-5]. The flavonoids from rhizome *Zingiber zerumbet* SM have also tested the biological activities in some different cancer cells[4,6]. Furthermore, the flavonoids presented their activities, and the role of food within flavonoids in cancer inhibition is widely studied[8-9].

The derivatives flavonoids are present in all higher plants and are found in many everyday vegetables and fruits[1,2]. Flavonoids are known to be a group of compounds that have antioxidant effects or prevent the oxidation of free radicals produced during metabolism, such as OH •, ROO • ...[3,5]. In addition, flavonoids in natural foods can form complexes with metal ions, which act as catalysts that inhibit oxidative reactions. Therefore, flavonoids have the effect of protecting the body, preventing atherosclerosis, stroke, aging, liver degeneration, radiation damage, and prevention: osteoporosis, hypertension, cardiovascular, high cholesterol, and some cancers[10,11].

The experimental results of this work have been studied the relationships between the structure and anticancer activity of flavonoids kaempferol-3-O-methyl ether and kaempferol-3-O-(2,4-O-diacetyl-alpha-L-rhamnopyranoside) isolating from rhizome *Zingiber zerumbet* SM in Viet Nam[1] using spectrum data ^{13}C NMR and ^{15}O NMR on carbon and oxygen atoms. In addition, the statistical techniques were supported for building $QSDAR_{MLR}$ and $QSDAR_{ANN}$ models to predict the anticancer activities of these flavonoids[10,12].

In this work, we report using semi-empirical quantum calculations TNDO/2 SCF to calculate spectrum data ^{13}C NMR and ^{15}O NMR on carbon and oxygen atoms and the construction of quantitative spectrum data and activity relationships (QSDARs) for 32 flavone and isoflavone derivatives. The multivariate regression and the artificial neural

network are used to construct the $QSDAR_{MLR}$ and $QSDAR_{ANN}$ model. In addition, the anticancer activities GI_{50}/ μ M of flavonoids in the test group and new flavonoid kaempferol-3-O-methyl ether isolating from rhizome *Zingiber zerumbet* SM in Viet Nam resulting from QSDAR models are compared with those from experimental data.

Experimental Section

Isolation process of kaempferol

Before determining the kaempferol-3-O-methyl ether structure by ^{1}H NMR and ^{13}C NMR spectrum[14-15] and crystal method[14], the chemicals and the equipment for isolating and purifying kaempferol-3-O-methyl ether were also used similarly in our work[15]. In addition, the techniques of thin-layer and column chromatography and the different spectrum techniques[15] as nuclear magnetic resonance spectrum (NMR) ^{1}H NMR (500 MHz) and ^{13}C NMR (125 MHz) implemented on Bruker AM500 FT-NMR Spectrometer were used to identify the kaempferol-3-O-methyl ether structure.

Database

The accurate capability of the QSDAR model depends on the collected data source. So the data set for building QSDAR model consisting of molecular structures with anticancer activities GI_{50}/μM for Hela cell line (GI_{50} is the concentration of 50% of maximal inhibition of cell proliferation) were taken from Wang *et al*. in the literature[11,16], as pointed out in Figure 1 and Table I. The value $logGI_{50}$ is the subsequent dependent variable that defines the biological parameter for the QSDAR model.

$$pGI_{50} = -\log GI_{50} \quad \ldots(1)$$

This research constructs the quantitative relationships between chemical-shift data τ_k of ^{13}C NMR and ^{15}O NMR spectra and anticancer activities on the Hela cell line. The 32 flavonoid structures proposed by Wang *et al*.[9,11] were optimized by the MM+ method, and they have been utilized to calculate their chemical-shift data τ_k of ^{13}C NMR and ^{15}O NMR spectra of carbon and oxygen atoms using semi-empirical quantum chemical calculation TNDO/2 in HyperChem[18]. The flavonoid skeleton is given in Figure 1.

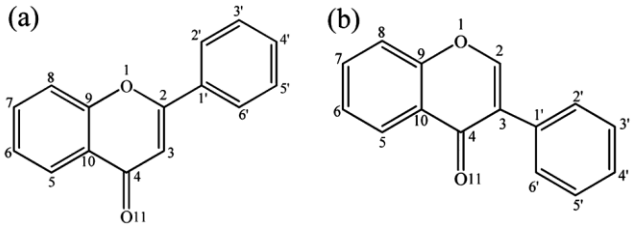


Figure 1 — Molecular skeleton: (a) flavone and (b) isoflavone [11,16]

The $QSDAR_{MLR}$ model was constructed with the multivariable regression techniques[19-21]. Artificial neural networks are intelligence systems; it is based on the natural neural network. So the artificial neural network model $QSDAR_{ANN}$ can be constructed with the program Neural Network in Mathlab[22,23].

Constructing QSDAR models

Multivariate regression model

The linear least-squares construct multivariable linear regression models $QSDAR^{MLR}$ fit on several independent variables τ_k for the activity variable *Y*.

All the above mentioned can be generalized to the case of dependence of activity variable *Y* on several independent (chemical-shift) variables τ_k {XE "Independent variables"} (x_1, x_2, …, x_{ip})[20,23]. The $QSDAR_{MLR}$ model consists of the selected coefficients, b_k ensuring the most significant adequacy to the formed linear model. The coefficients b_k are found from the minimization of the residual sum of squares {XE "Residual:sum of squares"} (*RSS*) of differences between observed and predicted values[19-24]:

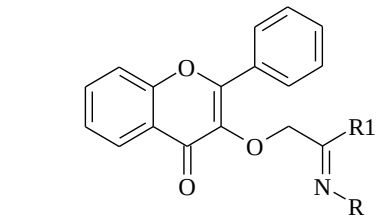
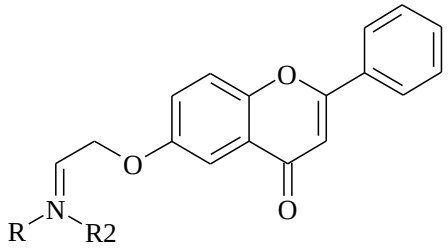
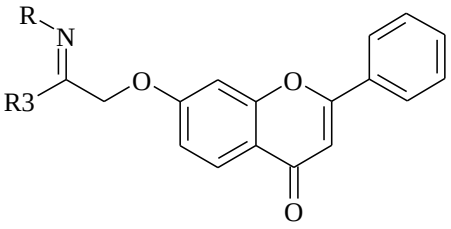
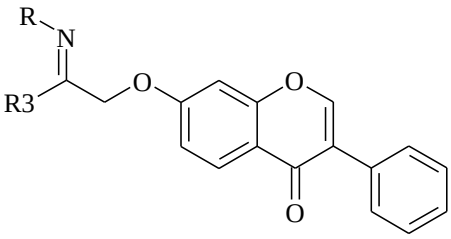
$$RSS_p = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad \ldots(2)$$

Where y_i and $\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_p x_{ip}$ are experimental and calculated activity.

Artificial neural network model

The artificial neural network (ANN) is generally presented by a neural network diagram, node characteristics, and training or learning rules. Many nodes create an ANN architecture. The backpropagation neural network (BPNN) is also known as a multi-layer feed-forward network. The training dataset trains this BPNN while tuning the network parameters using an error backpropagation mechanism[22,23]. A BPNN consists of several layers of networks; however, it is most frequently accepted as three-layer architecture type $QSDAR_{ANN}$ I(*k*)-HL(*m*)-O(*n*), as exhibited in Figure 2. The input layer I(*k*) with *k* neurons are seven chemical-shift parameters, τ_k; the hidden layer HL(*m*) with *m* hidden neurons; and the output layer O(*n*) with one neuron *n* = 1 is anticancer activity pGI_{50}. The training algorithm for

Table I — The experimental anticancer activity pGI_{50} and structures of 32 flavonoids[11,16]

Structure	Substitutes R, R_1, R_2 and R_3	pGI_{50}
	1*: R=OH, R_1 = Me	5.6990
	2: R= OH, R_1 = C_6H_5	5.7447
	3: R= OH, R_1 = *p*-F- C_6H_4	5.6778
	4: R= OH, R_1 = *p*-OMe- C_6H_4	5.6990
	5: R= Me, R_1 = Me	5.6990
	6: R= Me, R_1 = C_6H_5	5.7959
	7: R= Me, R_1 = *p*-F- C_6H_4	5.6990
	8: R= Me, R1 = *p*-OMe- C6H4	5.6990
	9: R=OH, R_2 = Me	5.9208
	10: R= OH, R_2 = C_6H_5	6.0969
	11*: R= OH, R_2 = *p*-F- C_6H_4	5.7959
	12: R= OH, R_2 = *p*-OMe- C_6H_4	6.0000
	13: R= Me, R_2 = Me	6.0458
	14: R= Me, R_2 = C_6H_5	5.6198
	15: R= Me, R_2 = *p*-F- C_6H_4	5.6383
	16: R= Me, R_2 = *p*-OMe- C_6H_4	5.6990
	17: R=OH, R_3 = Me	5.6990
	18: R= OH, R_3 = C_6H_5	5.6990
	19: R= OH, R_3 = *p*-F- C_6H_4	5.6990
	20: R= OH, R_3 = *p*-OMe- C_6H_4	5.6990
	21: R= Me, R_3 = Me	5.6576
	22: R= Me, R_3 = C_6H_5	5.1805
	23: R= Me, R_3 = *p*-F- C_6H_4	5.5686
	24*: R= Me, R_3 = *p*-OMe- C_6H_4	5.6021
	25*: R=OH, R_3 = Me	5.0088
	26*: R= OH, R_3 = C_6H_5	5.0458
	27: R= OH, R_3 = *p*-F- C_6H_4	5.1079
	28: R= OH, R_3 = *p*-OMe- C_6H_4	5.1192
	29: R= Me, R_3 = Me	5.0706
	30*: R= Me, R_3 = C_6H_5	5.0862
	31: R= Me, R_3 = *p*-F- C_6H_4	5.1938
	32: R= Me, R_3 = *p*-OMe- C_6H_4	5.1367

* has randomly selected substances in the test group

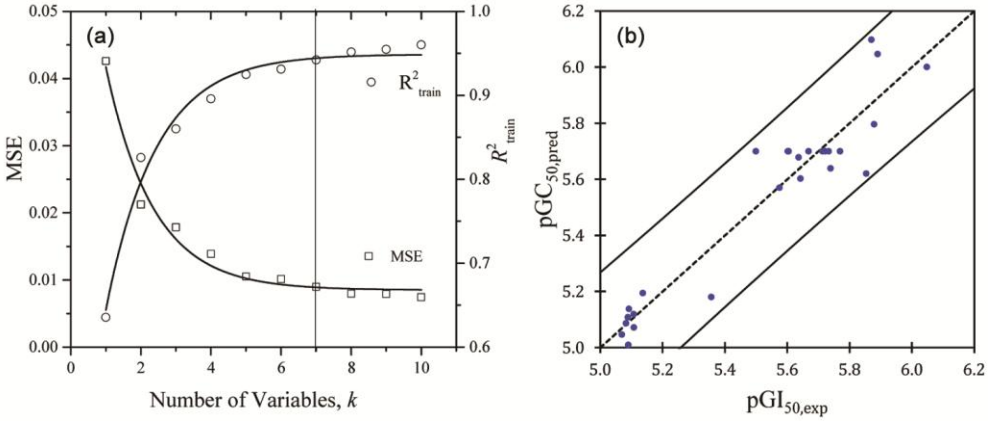


Figure 2 — (a) The change of R^2_{train} and SE by numbers *k*; (b) Correlation between $pGI_{50,exp}$ and $pGI_{50,pred}$ at $QSDAR_{ANN}$ model I(7)-HL(9)-O(1).

backpropagation neural network (BPNN) is Levenberg – Marquardt[22,23].

Performance criteria

In this study, the different errors were used to assess the performance of the $QSDAR_{MLR}$ and $QSDAR_{ANN}$ models. The root means square error (*RMSE*) and multiple correlation value R^2 were used to indicate the observed and predicted values[20,23-25]. The expressions of these errors are given as

Multiple value R^2 is designated in the following formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad \ldots(3)$$

The predicted results derived from QSDAR models are compared with experimental data using the relative errors (*ARE*,%)[12,15,20-25]:

$$ARE,\% = 100\left|(y_i - \hat{y}_i)/y_i\right| \quad \ldots(4)$$

The average value of absolute relative errors *ARE*,%[23,24] is calculated and used for assessing the global uncertainty of QSDAR models

$$MARE,\% = \frac{100}{N}\left|\frac{(y_i - \hat{y}_i)}{y_i}\right| \quad \ldots(5)$$

with N is number of activity values

Where y_i is experimental activity $pGI_{50,exp}$; $\hat{y}_i$ is calculated activity $pGI_{50,pred}$ from QSDAR models; $\bar{y}$ is the average value of $pGI_{50,exp}$.

Results

Constructing $QSDAR_{MLR}$ model

Experimental structures of flavonoids[11,16] were re-optimized by using molecular mechanics MM+ at gradient value 0.05. The chemical-shift values, τ_k on atoms, resulted from simulation-spectrum calculation ^{13}C NMR and ^{15}O NMR using the semi-empirical quantum chemistry method TNDO/2 SCF of module HyperNMR[18]. The $QSDAR_{MLR}$ and $QSDAR_{ANN}$ models were constructed from chemical-shift values, τ_k on atoms oxygen and carbon.

From experimental activity pGI_{50} and independent variables (chemical-shift values, τ_k are changed in range 2 to 10) for the atoms oxygen and carbon, the $QSDAR_{MLR}$ models are established by the equation form $pGI_{50} = \sum b_i \tau_i + b_0$[19-25]. The $QSADR_{MLR}$ models were selected by combining the backward and forward technique with the genetic algorithm, as shown in Table II. The selection process for $QSDAR_{MLR}$ models is based on the statistical values R^2_{train}, standard error SE, $R^2_{adj,}$ and R^2_{test}. The selected models $QSDAR_{MLR}$ were also cross-validated by the leave-one-out technique.

From Figure 2a and Table II, the $QSDAR_{MLR}$ model (with k = 7) including the chemical-shift values ^{13}C NMR and ^{15}O NMR of atoms O_1, O_{11}, C_2, C_3, C_6, C_7 và $C_{2'}$ are exhibited at value R^2_{train} of 0.9057, F_{stat} of 24.683 and SE of 0.1213. This model seems to be the best $QSDAR_{MLR}$ model. Those values also exhibited the significant contribution to the anticancer activity of flavonoid derivatives, as in Table III. The $QSDAR_{MLR}$ model (with k = 7) is pointed out in

$$pGI_{50} = 32.990 + 0.0156O_1 + 0.0055O_{11} - 0.045C_2 - 0.063C_3 - 0.047C_6 - 0.059C_7 - 0.037C_{2'} \quad \ldots(6)$$

with n = 26; R^2_{train} of 0.9057; R^2_{adj} of 0.8690; R^2_{test} of 0.7137; SE of 0.1213

In Table III, the average percentage of contribution $MP_m x_i$,% of important atoms O_1, O_{11}, C_2, C_3, C_4, C_6, $C_{7,}$ and $C_{2'}$ by percentage of contribution $P_m x_i$,% of each independent variable[12,24,25]. The values $MP_m x_i$,% are calculated by:

Table II — Atomic sites are chosen in $QSADR_{MLR}$ models by using the backward and forward technique

k	Atomic sites in QSDAR models	R^2_{train}	R^2_{adj}	SE	R^2_{test}
2	O_1, C_4	0.3267	0.2681	0.2866	0.0362
3	O_1, C_4, C_{11}	0.4836	0.4132	0.2566	0.2035
4	O_1, C_4, C_5, C_6	0.6356	0.5662	0.2206	0.3801
5	O_1, O_{11}, C_4, C_6, C_7	0.8016	0.7520	0.1668	0.6308
6	O_1, O_{11}, C_3, C_4, C_6, C_7	0.8606	0.8166	0.1434	0.6734
7	O_1, O_{11}, C_2, C_3, C_6, C_7, $C_{2'}$	0.9057	0.8690	0.1213	0.7137
8	O_1, O_{11}, C_2, C_3, C_4, C_5, C_6, C_7	0.9094	0.8668	0.1223	0.7129
9	O_1, O_{11}, C_2, C_3, C_4, C_5, C_6, C_7, $C_{2'}$	0.9212	0.8769	0.1175	0.6782
10	O_1, O_{11}, C_2, C_3, C_4, C_5, C_6, C_7, $C_{1'}$,$C_{2'}$	0.9214	0.8689	0.1213	0.6407

Table III — The statistical values, regression coefficients, and contribution percentage of chemical-shift values τ_k in $QSDAR_{MLR}$ models with k independent variables

Variables	$QSDAR_{MLR}$			$MP_m x_i$,%			$GMP_m x_i$,
	k = 5	k = 6	k = 7	k = 5	k = 6	k = 7	%
R^2_{train}	0.8016	0.8606	0.9057	–	–	–	–
R^2_{adj}	0.7520	0.8166	0.8690	–	–	–	–
SE	0.1668	0.1434	0.1213	–	–	–	–
R^2_{test}	0.6308	0.6734	0.7137	–	–	–	–
constant	14.4590	37.0550	32.990	–	–	–	–
O_1	–0.0001	0.0136	0.0156	0.1327	7.6489	9.9485	5.9100
O_{11}	0.0041	0.0075	0.0055	18.599	11.3859	9.5007	13.1620
C_2	–	–	–0.0450	–	–	16.4483	5.4828
C_3	–	–0.0580	–0.0630	–	14.3481	17.6032	10.6500
C_4	–0.0680	–0.0780	–	71.640	27.4712	–	33.0370
C_6	0.0062	–0.0590	–0.0470	5.1686	16.4235	14.9654	12.1860
C_7	–0.0050	–0.0750	–0.0590	4.4604	22.7224	20.3118	15.8320
$C_{2'}$	–	–	–0.0370	–	–	11.2222	3.74070

$$MP_m x_i,\% = \frac{1}{N}\sum_{i=1,m=5,6,7}^{k}\left(100.\left|b_{m,i}x_{m,i}\right|/C_{\text{total}}\right) \quad \ldots(7)$$

với $C_{\text{total}} = \sum_{i=1}^{k}\left|b_{m,i}x_{m,i}\right|$

Where n = 26 - total compounds, m QSDAR models (with m of 5 to 7); k variables in a model.

The global average percentage of contribution $GMP_m x_i$,%[12,24,25] is calculated by

$$GMP_m x_i,\% = \frac{1}{m}\sum_{i=1}^{k} MP_m x_i \quad \ldots(8)$$

with m is 3 QSDAR models

The significant contribution $GMP_m x_i$,% of atomic sites on flavonoid compounds was pointed out in 3 models $QSDAR_{MLR}$ (with k = 5, 6, 7), as given in Table III. Those can be sorted by values $GMP_m x_i$,% as $C_4 > C_7 > O_{11} > C_6 > C_3 > O_1 > C_2 > C_{2'}$. The atomic sites O_4, C_7, O_{11}, and C_6, presented the most important contribution. These seem to be the most important sites for flavonoid substances due to these atoms wear the carbonyl group $C_4 = O_{11}$. The π paired electrons of bond $C_2 = C_3$ associate with free paired electrons of atom O_1 to generate the conjugate system. The carbonyl group $C_4 = O_{11}$ presents the reactive activity of the carbonyl system. This satisfied with experimental works[11,15,16]. Furthermore, the atomic sites C_6, C_7, C_3 showed to be important sites, and these are considered to substitute the new functional group on atom sites C_3, C_6, and C_7[12,15]. Moreover, the selected sites C_3, C_6, and C_7 are still the blank sites for constructing new flavonoid derivatives with higher activity.

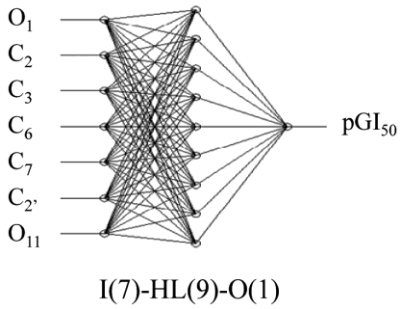


Figure 3 — Back-propagation Neural Network BPNN with architecture I(7)-HL(9)-O(1)

From the $QSDAR_{MLR}$ model (with k = 7) we have constructed the $QSDAR_{ANN}$ model with architecture I(7)-HL(9)-O(1) based on the $QSDAR_{MLR}$ model, as illustrated in Figure 3. The neural network model $QSDAR_{ANN}$ I(7)-HL(9)-O(1) was constructed by Matlab System[22,23]. Neurons in input layer I(7) involving chemical-shift values τ_k of atomic sites O_1, O_{11}, C_2, C_3, C_6, C_7, and $C_{2'}$ in $QSDAR_{MLR}$ model (6). The output layer O(1) consists of one neuron pGI_{50}; the hidden layer HL(9) has nine neurons. The error-back propagation algorithm was chosen to train the neural network. The value MSE of 2.5573×10^{-5} was obtained from the training process after training 10000 epochs. And training moment of 0.7. It was obtained from the training process with 10000 epochs. The quality of neural network model $QSDAR_{ANN}$ I(7)-HL(9)-O(1) is presented in the values R^2_{train} of 0.993 and R^2_{pred} of 0.971. This can be explored to predict the anticancer activities on the Hela cell line for new flavonoid kaempferol-3-O-methyl ether isolated from rhizome *Zingiber zerumbet* SM in Viet Nam.

Identification of new flavonoid

After isolating flavonoid from Plant rhizome *Zingiber zerumbet* SM[1] in Figure 4a, Kaempferol-3-O-methyl ether was purified to generate the yellow crystal with melting temperature in the range 249 to 250°C; the chromatography of thin layer with solvent system n-hexane – EtOAc (7:3) was colored by solution H_2SO_4 10%/ EtOH, and the yellow circle created with Rf in 0.14.

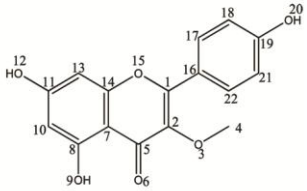
The structure kaempferol-3-O-methylether was identified by using the different spectra such as: ^{1}H NMR (500 MHz, DMSO, δ ppm): δ 6,20 (1H, d, J = 2, H_6); 6,44 (1H, d, J = 2, H_8); 7,93 (2H, d, $H_{6'}$, $H_{2'}$); 6,94 (2H, d, J = 9, $H_{5'}$,$H_{3'}$); 12,68 (1H, s, 5-OH); 3,78 (3H, s, H_{11}). ^{13}C NMR (125MHz, δ ppm): δ155,6 (C_2); 137,6 (C_3); 177,9 (C_4); 161,2 (C_5); 98,5 (C_6); 164,1 (C_7); 93,7 (C_8); 156,3 (C_9); 104,2 (C_{10}); 120,5 ($C_{1'}$); 130,1 ($C_{2'}$); 115,6 ($C_{3'}$); 160,1 ($C_{4'}$); 115,6 ($C_{5'}$); 130,1 ($C_{6'}$).

Interaction of atom C and H in heteronuclear multiple-bond correlation (HMBC) and heteronuclear single-quantum correlation spectroscopy (HSQC) were pointed out the atomic sites: H_6 - C_5 - C_7 - C_8 - C_{10}; H_8 -C_6 - C_7 - C_9 - C_{10}; H_{11} - C_3; $H_{2'}$ - C_2 - $C_{3'}$- $C_{4'}$ - $C_{6'}$; $H_{3'}$ -$C_{1'}$ - $C_{2'}$ - $C_{4'}$ - $C_{5'}$; $H_{5'}$ - $C_{1'}$ - $C_{3'}$ - $C_{4'}$- $C_{6'}$; $H_{6'}$ - C_2 - $C_{1'}$ - $C_{2'}$ - $C_{4'}$ - $C_{5'}$. The structure of kaempferol-3-O-methylether in rhizome *Zingiber zerumbet* SM are shown in Figure 4b.

Determining structure by X-ray diffraction

In this work, we use an X-ray diffraction method to determine kaempferol-3-O-methyl ether structure. X-ray diffraction is the phenomenon of diffracted X-ray beams on the crystalline surfaces due to the periodicity of the crystalline structure that produces the maxima and minimum diffraction. X-ray diffraction (often abbreviated as X-ray diffraction) is used to analyze the structure of solids, materials, *etc*. In terms of physical nature, X-ray diffraction is similar to that of electromagnetic diffraction. The difference in diffraction properties is due to the interaction between X-rays with atoms and the interaction between electrons and atoms.

(a) Plant rhizome *Zingiber zerumbet* SM



(b) kaempferol-3-O-methylether with $GI_{50,exp}$ (μM) = 15.37 ± 0.401

Figure 4 — The kaempferol-3-O-methylether isolating from rhizome *Zingiber zerumbet* SM [1]

The single crystal of structure kaempferol-3-O-methyl ether $C_{16}H_{12}O_6$ was determined[14]. A suitable crystal was selected on a Bruker D8 Quest diffractometer. The crystal was kept at 100 *K* during data collection. Using Olex2[14], the structure was solved with the structure solution program ShelXT[27] using Direct methods and refined with the refinement package ShelXL[27] using least-squares minimization.

- The length of double bond C=C of 1.33 Å is calculated by the theoretical method. The single crystal results of X-ray diffraction for bonds C_1=C_2, C_8=C_{10}, C_7=C_{14}, C_{11}=C_{13}, C_{16}=C_{22}, C_{17}=C_{18} with experimental length are in the range 1.360 to 1.399 Å.

- Bond length C-C is theoretically 1.54 Å, the experimental length of C_1-C_{16} is 1.47 Å.

- Bond length C=O is theoretically 1.20 Å, the experimental length of C_5-O_6 is 1.264 Å.

- Bond length C-O is theoretically 1.43 Å, the experimental length of C_2-O_3 is 1.3766 Å, length O_3-C_4 of 1.445 Å, length C_8-O_9 of 1.3489 Å, length C_{11}-O_{12} of 1.348 Å, length C_{19}-O_{20} of 1.362 Å are suitable for those from the theoretical calculation.

The technique of X-ray diffraction presented the kaempferol-3-O-methyl ether structure. The geometric structure is shown by the atom sites in the space, including bond lengths and different angles, as given in Table IV and Table V. The configuration properties of kaempferol-3-O-methyl ether are satisfactory for those from theoretical calculations. This is a suitable structure for substance kaempferol-3-O-methyl ether isolating from rhizome *Zingiber zerumbet* SM[1], as shown in Figure 5.

Table IV — Bond Lengths for kaempferol-3-O-methyl ether resulting from X-ray diffraction

Atom	Atom	Length/Å	Atom	Atom	Length/Å
C_1	C_2	1.360(2)	C_{10}	C_{11}	1.404(2)
C_1	O_{15}	1.3631(19)	C_{11}	O_{12}	1.348(2)
C_1	C_{16}	1.470(2)	C_{11}	C_{13}	1.396(2)
C_2	O_3	1.3766(19)	C_{13}	C_{14}	1.388(2)
C_2	C_5	1.446(2)	C_{14}	O_{15}	1.3701(17)
O_3	C_4	1.445(2)	C_{16}	C_{17}	1.402(2)
C_5	O_6	1.264(2)	C_{16}	C_{22}	1.397(2)
C_5	C_7	1.437(2)	C_{17}	C_{18}	1.390(2)
C_7	C_8	1.423(2)	C_{18}	C_{19}	1.393(3)
C_7	C_{14}	1.399(2)	C_{19}	O_{20}	1.360(2)
C_8	O_9	1.3489(19)	C_{19}	C_{21}	1.390(3)
C_8	C_{10}	1.385(2)	C_{21}	C_{22}	1.389(2)

The crystal data and experimental structure of kaempferol-3-O-methyl ether were derived from the X-ray diffraction method at the Laboratory of the National University of Science, Ha Noi, as given in Table VI.

Crystal data for $C_{16}H_{12}O_6$ (M =300.26 g/mol): orthorhombic, space group kaempferol-3-O-methylether, a = 7.0060(4) Å, b = 25.9070(14) Å, c = 7.2062(4) Å, V = 1307.96(13) $Å^3$, Z = 4, *T* = 100.0 K, μ(MoKα) = 0.118 mm-1, D_{calc} = 1.525 g/cm^3, 44181 reflections measured (5.868° ≤ 2Θ ≤ 61.67°), 4114 unique (R_{int} = 0.0465, R_{sigma} = 0.0199) which were used in all calculations. The final $R_1$was 0.0387 (I > 2σ(I)) and wR_2 was 0.0987 (all data).

Table V — The bond angle of geometric structure kaempferol-3-O-methyl ether obtained from X-ray diffraction

Atom	Atom	Atom	Angle/°	Atom	Atom	Atom	Angle/°
C_2	C_1	O_{15}	121.24(14)	O_{12}	C_{11}	C_{13}	122.20(15)
C_2	C_1	C_{16}	126.82(15)	C_{13}	C_{11}	C_{10}	121.37(15)
O_{15}	C_1	C_{16}	111.93(14)	C_{14}	C_{13}	C_{11}	117.74(15)
C_1	C_2	O_3	121.51(14)	C_{13}	C_{14}	C_7	123.04(14)
C_1	C_2	C_5	121.51(14)	O_{15}	C_{14}	C_7	120.67(13)
O_3	C_2	C_5	116.98(14)	O_{15}	C_{14}	C_{13}	116.28(14)
C_2	O_3	C_4	111.96(13)	C_1	O_{15}	C_{14}	120.71(13)
O_6	C_5	C_2	121.72(14)	C_{17}	C_{16}	C_1	119.98(15)
O_6	C_5	C_7	122.74(13)	C_{22}	C_{16}	C_1	121.14(15)
C_7	C_5	C_2	115.54(14)	C_{22}	C_{16}	C_{17}	118.86(15)
C_8	C_7	C_5	121.94(14)	C_{18}	C_{17}	C_{16}	120.71(16)
C_{14}	C_7	C_5	120.32(13)	C_{17}	C_{18}	C_{19}	119.69(17)
C_{14}	C_7	C_8	117.73(14)	O_{20}	C_{19}	C_{18}	116.81(17)
O_9	C_8	C_7	120.14(14)	O_{20}	C_{19}	C_{21}	123.10(17)
O_9	C_8	C_{10}	119.56(14)	C_{21}	C_{19}	C_{18}	120.09(16)
C_{10}	C_8	C_7	120.30(14)	C_{22}	C_{21}	C_{19}	120.20(17)
C_8	C_{10}	C_{11}	119.82(14)	C_{21}	C_{22}	C_{16}	120.44(17)
O_{12}	C_{11}	C_{10}	116.43(14)				

Prediction of *in vitro* activity of new substances

After isolating flavonoid kaempferol-3-O-methyl ether from rhizome *Zingiber zerumbet* SM[1], the biological activity value pGI_{50} was conducted to test

Table VI — Crystal data and refinement structure for kaempferol-3-O-methyl ether from X-ray diffraction

Identification code	kaempferol-3-O-methylether
Empirical formula	$C_{16}H_{12}O_6$
Formula weight	300.26
Temperature/K	100
Crystal system	orthorhombic
Space group	kaempferol-3-O-methylether
a/Å	7.0060(4)
b/Å	25.9070(14)
c/Å	7.2062(4)
α/°	90
β/°	90
γ/°	90
Volume/$Å^3$	1307.96(13)
Z	4
ρ_{calc}g/cm^3	1.525
μ/mm^{-1}	0.118
F(000)	624
Crystal size/mm^3	0.1 × 0.08 × 0.04
Radiation	MoKα (λ = 0.71073)
2Θ range for data collection/°	5.868 to 61.67
Index ranges	–10 ≤ h ≤ 10, -37 ≤ k ≤ 37, –10 ≤ l ≤ 10
Reflections collected	44181
Independent reflections	4114 [R_{int} = 0.0465, R_{sigma} = 0.0199]
Data/restraints/parameters	4114/1/203
Goodness-of-fit on F^2	1.092
Final R indexes [I >= 2σ (I)]	R_1 = 0.0387, wR_2 = 0.0959
Final R indexes [all data]	R_1 = 0.0427, wR_2 = 0.0987
Largest diff. peak/hole / e $Å^{-3}$	0.44/-0.24
Flack parameter	0.0 (2)

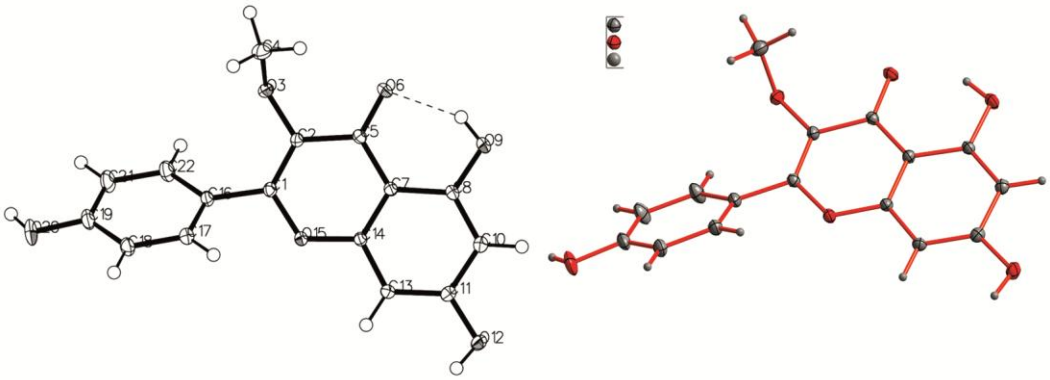


Figure 5 — The identified substance kaempferol-3-O-methyl ether from the X-ray diffraction method

Table VII — Comparison between experimental *in vitro* activities pGI_{50} of six test flavonoids and kaempferol-3-O-methyl ether with those from $QSDAR_{MLR}$ and $QSDAR_{ANN}$ model

Substance	Ref.	$pGI_{50,exp}$	$pGI_{50,pred}$		*ARE*,%	
			$QSDAR_{MLR}$	$QSDAR_{ANN}$	$QSDAR_{MLR}$	$QSDAR_{ANN}$
Fla-1	[11, 16]	5.6990	5.3879	5.7371	5.4589	0.6685
Fla-11		5.6990	5.9188	5.6681	3.8568	0.5422
Fla-24		5.6198	5.9316	5.6662	5.5482	0.8257
Fla-25		5.6383	5.8627	5.6843	3.9799	0.8158
Fla-26		5.6990	5.8841	5.7058	3.2479	0.1193
Fla-30		5.0862	4.8677	4.9365	4.2959	2.9433
Kaempferol	this work	4.8130	4.8360	4.8061	0.4779	0.1434
				MARE,%	3.3812	0.7785

in vitro toxicity on the Hela cell line in the laboratory of molecular biology, National University of Sciences, Ho Chi Minh city. The anticancer activities $pGI_{50,pred}$ of randomly six selected flavonoids from Table I[11,16] and *ARE*, % values resulting from $QSDAR_{MLR}$ model (with *k* of 7) and $QSDAR_{ANN}$ I(7)-HL(9)-O(1) given in Table VII.

The predictability of the models $QSDAR_{MLR}$ (with *k* of 7) and $QSDAR_{ANN}$ I(7)-HL(9)-O(1)[22] were evaluated carefully by the leave-one-out (LOO) technique in the predictive values R^2_{pred} [15,24,25]. The activity values pGI_{50} of six substances in the test group resulting from $QSDAR_{MLR}$ and $QSDAR_{ANN}$ models were compared with values *ARE*,%, as shown in Table VII. Due to value *MARE*,% of the $QSDAR_{MLR}$ model is higher than value *MARE,%* of $QSDAR_{ANN}$ model. So the predictability of model $QSDAR_{ANN}$ I(7)-HL(9)-O(1) is better than model $QSDAR_{MLR}$ (with *k* of 7). Furthermore, two $QSDAR_{MLR}$ and $QSDAR_{ANN}$ models were used for predicting anticancer activity pGI_{50} of six test substances in Table VII. They showed that the predicted errors are in the uncertainty range of experimental measurements. Therefore, the $QSDAR_{MLR}$ and $QSDAR_{ANN}$ model can satisfy the accurately predictability pGI_{50} for new flavonoids.

Docking simulation

To further confirm the inhibitory activity of flavonoids on the growth of the Hela cancer cell line, we investigated the mechanism of binding the kaempferol derivative to the protein receptor by docking simulation. In this section, the kaempferol derivatives insolating from plant rhizome *Zingiber zerumbet* SM are evaluated to inhibit the growth of the Hela cancer cells. Furthermore, we investigated the possibility of inhibiting the growth of these cancer cells by inhibiting tubulin polymerization[28,28]. This result may help elucidate the anticancer effects of flavonoid derivatives based on predictability from the $QSDAR_{MLR}$ and $QSDAR_{ANN}$ relationships. The ability to bind kaempferol and its derivatives to tubulin (Figure 6) was evaluated and compared. The crystal structure of tubulin used in this study was obtained from the protein data bank (PDB ID: 1TUB)[28].

The protein crystal structure (ID CODE: 1TUB) was used as the receptor model for docking simulation. The distances between all essential amino acids are limited between the ligands with a spherical radius < 1.5 Å[28]. The water and amino acid residues are removed before determining the active site of the protein. The flavonoid structure was optimized at MMFF94x force field with gradient level 0.001; the maximum epochs of 10,000 and the minimum energy of the structure is 0.001 $kcal.mol^{-1}$. The RMSD radius value of 1.408 Å is the distance between the bound and original structures that can bind the protein[29]. Simulating the interaction of flavonoid derivatives with 1TUB protein by Triangle Matching method and epochs of 1000, the number of retained conformations is 10, the number of solutions is 200/fragmentation time. The docking results are based on docking score (DS) ($kcal.mol^{-1}$), interaction distance (RMSD) between ligand and protein; The bonds and interactions evaluated include van der Waals bonds, π- bonds, ionic bonds, and hydrogen bonds[28].

Kaempferol-s interacts with 1TUB with DS =-7.959 $kcal.mol^{-1}$; RMSD = 1.009 Å, amino acid hydrogen interactions are SER 140 (2.95 Å; -1.3 $kcal.mol^{-1}$), ASN 101 (3.28 Å, -1.5 $kcal.mol^{-1}$), LYS 254 (2.96 Å, -10 kcal $.mol^{-1}$). The substance Flav-1 interacts with 1TUB with DS = -8.214; RMSD = 1.478, aminoacid hydrogen interactions are SER 140 (3.03 Å, -1.2 $kcal.mol^{-1}$), ASN 101(3.34 Å, -1.2 $kcal.mol^{-1}$), LYS 254 (2.93 Å, -10.4 kcal. mol^{-1}). The substance Flav-2 interacts with 1TUB with

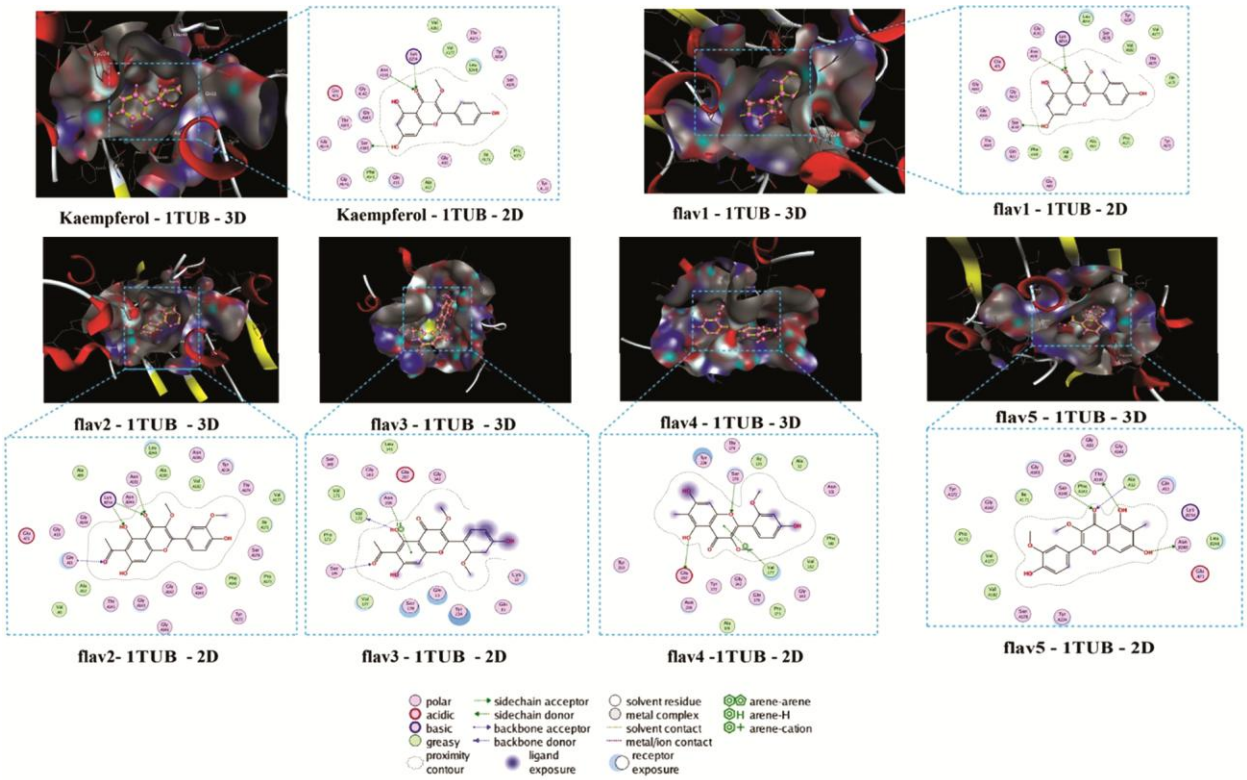


Figure 6 — Compounds kaempferol-1, flav-1, flav-2, flav-3, flav-4, and flav-5 (Table VIII) were docked at the interface of a tubulin heterodimer (PDB code: 1TUB), 2D and 3D structures.

DS =-8.402; RMSD = 1,952, aminoacid hydrogen interactions are ASN 101 (3.55 Å, -0.9 kcal.mol^{-1}), LYS 254 (3.35 Å, -7.7 kcal.mol^{-1}), LYS 254 (2.66 Å, -1.5 kcal. mol^{-1}), GLN 11 (3.27 Å, -3 kcal.mol^{-1}). The substance Flav3 interacts with 1TUB with DS = -8,735; RMSD = 0.748, aminoacid hydrogen interactions are VAL 172 (2.78 Å, -5.4 kcal.mol^{-1}), SER 174 (3.27 Å, -1.7 kcal.mol^{-1}), SER 174 (3.31 Å, -0.8 kcal. mol^{-1}), ASN 206 (3.59 Å, -0.6 kcal.mol^{-1}). The substance Flav4 interacts with 1TUB with DS = -8.868; RMSD = 1,196, aminoacid hydrogen interactions are GLU 207 (3.52 Å, -0.8 kcal.mol^{-1}), SER 178 (2.81 Å, -0.7 kcal.mol^{-1}), VAL 177 (3.75 Å, -1 kcal. mol^{-1}). The substance Flav5 interacts with 1TUB with DS= -8,965; RMSD = 1.541, aminoacid hydrogen interactions are THR 145 (2.95 Å, -1 kcal.mol^{-1}), ASN 249 (2.9 Å, -3 kcal.mol^{-1}), ALA 12 (3.28 Å, -0.7 kcal. mol^{-1}), SER 140 (2.8 Å, -0.9 kcal.mol^{-1}).

The combination of hydrophobic interactions and hydrogen bonding is responsible for the strong binding of the flavonoid derivatives to the amino acids of the 1TUB protein. The docking results showed that the newly designed substances interact well with the 1TUB protein and have a lower docking score than kaempferol-1 insolating from plant rhizome *Zingiber zerumbet* SM.

Discussion

To use ANN methods, we may need a large group of data. In the case of many scattered data, there are no definite rules. However, in this case, we found that our data is highly linear and that the isolated kaempferol-3-O-methyl ether is structurally located in the region of 32 flavonoid compounds. The experiment methods with 32 data are enough to establish the regression-correlation models. The regression model is also sufficient to predict the anticancer activity of the new compound. We add a regression ANN to provide a higher level of confidence in predictability. Therefore, this group of 32 flavonoids is reliable enough to predict the new compound kaempferol-3-O-methyl ether isolated from rhizome *Zingiber zerumbet* SM in Vietnam. We choose the QSDAR model because this model gives the highest linearity and is reliable to predict the antitumor activity for the kaempferol-3-O-methyl ether structure. We also selected the test group to test these models $QSDAR_{MLR}$ and $QSDAR_{ANN}$, as shown in Table I and Table VII. The predicted results for the

Table VIII — The anticancer activities GI_{50} (μM) of five new flavonoids resulting from the $QSDAR_{ANN}$ model

New substance	C_6	$C_{2'}$	$C_{3'}$	pGI_{50}	A method in this work
kaempferol-1	H	H	H	4.813	*in vitro* test on Hela
flav-1(new)	H	CH_3	H	5.781	$QSDAR_{ANN}$
flav-2(new)	CH_3CO^-	H	OCH_3	6.155	$QSDAR_{ANN}$
flav-3(new)	CH_3CO^-	OCH_3	H	6.158	$QSDAR_{ANN}$
flav-4(new)	CH_3	OCH_3	H	6.488	$QSDAR_{ANN}$
flav-5(new)	CH_3	H	OCH_3	6.538	$QSDAR_{ANN}$

test substances agree with the experimental measurements, as shown in Table VII. The structure kaempferol-3-O-methyl ether and *in-vitro* anticancer activity in the Hela cell line were predicted by ^{1}H NMR, ^{13}C NMR, X-ray results, and the *in-vitro* tests for kaempferol-3-O-methyl ether in the laboratory of molecular biology, National University of Sciences, Ho Chi Minh city. This could be an essential testament to our success in this project.

Therefore, the compound kaempferol-3-O-methyl ether with vacant positions C_6, $C_{2'}$ and $C_{3'}$ as the lead compound was wielded for designing five new various derivatives. The new function groups are substituted into the positions C_6, $C_{2'}$ and $C_{3'}$. The biological activities pGI_{50} of five new-designed flavonoids were predicted by using the $QSDAR_{ANN}$ I(7)-HL(9)-O(1) model, as given in Table VIII. The docking results of the newly designed derivatives from kaempferol-1 show that the biological activities of the new derivatives flav-1(new) to flav-5(new) (Table VIII) are improved in line with the predicted results from the $QSDAR_{MLR}$ and $QSDAR_{ANN}$ models built above.

The predicted values pGI_{50} for five new-designed substances are compared with the experimental activity of kaempferol-3-O-methyl ether in Table VII. Accordingly, the anticancer activities GI_{50} (μM) of five new derivatives designing from C_6, $C_{2'}$ and $C_{3'}$ sites turn out to be stronger than the lead compound kaempferol-1. Thus, herein the five newly designed compounds will promise to forward the designing plans for the new pharmaceutical products from natural products.

Conclusion

We conclude that this work successfully isolated compound kaempferol-3-O-methyl ether from plant rhizome *Zingiber zerumbet* SM in Viet Nam[1] that may prove to be helpful in guiding the rational search of new therapeutic agents for cancer diseases. The chemical-shift values, τ_k ^{13}C NMR and ^{15}O NMR on atoms oxygen and carbon from simulated spectrum were utilized to construct the QSDAR models successfully with important-contribution sites O_1, O_{11}, C_2, C_3, C_6, C_7 và $C_{2'}$ on flavonoid compound which effect an *in vitro* activity on Hela cell line.

The model $QSDAR_{ANN}$ I(7)-HL(9)-O(1) was employed to predict success *in vitro* anticancer activities of compound kaempferol-3-O-methyl ether from plant rhizome *Zingiber zerumbet* SM in Viet Nam[1] and the biological activities of five new-designed derivatives from the vacant sites C_6, $C_{2'}$, $C_{3'}$ of lead compound kaempferol-3-O-methyl ether. The predicted activities from model $QSDAR_{ANN}$ I(7)-HL(9)-O(1) turn out to be stronger. The docking results showed that the newly designed substances interact well with the 1TUB protein and have a lower docking score than kaempferol-1 insolating from plant rhizome *Zingiber zerumbet* SM.

Acknowledgment

The authors would like to thank colleagues currently working at Hanoi University of Natural Sciences and the Institute of Chemistry for creating necessary conditions for laboratory facilities for spectroscopic research and measurement techniques X-ray and bioactivity studies of the compound kaempferol-3-O-methyl ether.

Conflict of interest

The authors declare no conflict of interest.

References

1 Loi D T, *Medicinal Plants and Drugs from Vietnam* (Publisher of Medicine, Ha Noi, Viet Nam) p.1274 (2006).
2 Singh M, Kaur M & Silakari O, *Eur J Med Chem*, 84(12) (2014) 206.
3 Mahapatra, D.K., S.K. Bharti, and V. Asati, *Eur J Med Chem*, 98(15) (2015) 69.
4 R. Vidya Priyadarsini, R. Senthil Murugan, S. Maitreyi, K. Ramalingam, D. Karunagaran, S. Nagini., *Eur J Pharmacology*, 649 (1–3) (2010) 84.
5 L Ziberna, S Fornasaro, J Čvorović, F Tramer, S Passamonti, RR Watson, VR Preedy, S Zibadi, *Polyphenols in Human Health and Disease*, 1 (2014) 489.
6 Dae Sik Jang, Ah-Reum Han, Gowooni Park, Gil-Ja Jhon, and Eun-Kyoung Seo, *Arch Pharm Res*, 27(4) (2004) 386.
7 Nobuji Nakatani, Akiko Jitoe, Toshiya Masuda & Shigetomo Yonemori, *Agricultural and Biological Chemistry*, 55(2) (1991) 449.
8 Gavin, N.M. and M.J. Durako, *Experimental Marine Biology and Ecology*, 32–40 (2012) 416.
9 Lee, I.S.L., M.C. Boyce, and M.C. Breadmore, *Food Chemistry*, 133 (2012) 205.
10 Bożena Pawlikowska-Pawlęga, Halina Dziubińska, Elżbieta Król, Kazimierz Trębacz, Anna Jarosz-Wilkołazka, Roman

Paduch, Antoni Gawron, Wieslaw I. Gruszecki, *Biochim Biophys Acta - Biomembranes*, 1838(1B) (2014) 195.
11 Si Yan Liao, Jin Can Chen, Li Qian, Yong Shen, Kang Cheng Zheng, *Eur J Med Chem*, 43 (2008) 2159.
12 Tai-Chi Wang, I.-Li Chen, Pei-Jung Lu, Chui-Hei Wong, Chang-Hui Liao, Kuei-Ching Tsiao, Ken-Ming Chang, Yeh-Long Chen and Cherng-Chyi Tzeng., *Bioorg Med Chem*, 13 (2005) 6045.
13 Thuy, B.T.P. and P.V. Tat, *Viet Nam Journal of Chemistry*, 50(5A) (2012) 203.
14 Dolomanov, O.V., et al., *Appl Cryst*, 42 (2009) 339.
15 O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard and H. Puschmann, Appl Cryst, 42 (2009) 339.., *Appl Cryst*, 42 (2009) 339.
16 Thuy, B.T.P, Nhung, N.T.A, Duong, T, Trung, P.V, Quang, N.M, Dung, H.T.K, Tat, P.V, *Theoretical and Computational Chemistry*, *Cogent Chemistry*, 2(1) (2016) 12.
17 I-Li Chen, Jhy-Yih Chen, Po-Chuen Shieh, Jih-Jung Chen, Choa-Hsun Lee, Shin-Hun Juang, Tai-Chi Wang, *Bioorg Med Chem*, 16(16) (2008) 7639.
18 Allinger, N., HyperChem® Computational Chemistry., Hypercube, USA, Inc. (2002) 2170.
19 Bastien, P., V. Esposito Vinzi, and M. Tenenhaus, *Computational Statistics and Data Analysis*, 48 (2005) 17.
20 Mercader, A.G., P.R. Duchowicz, and P.M. Sivakumar, *Chemometrics applications and research QSAR in Medicinal Chemistry* (CRC Apple Academic Press Inc.) (2016) 458.
21 Pomerantsev A L, *Chemometrics in Excel* (John Wiley and Sons, Inc. Hoboken, New Jersey), 333 (2014).
22 Beale, M.H., M.T. Hagan, and H.B. Demuth, Neural Network Toolbox™ 7, User's Guide. 3 Apple Hill Drive Natick, MA (2010) 840.
23 Swayamprakash Patel, Girish Jani, Mruduka Patel., *Inventi Journals*: *PNDDS*, 865 (2014) 14.
24 Waterbeemd, H.V.D., Chemometric methods in molecular design. The Federal Republic of Germany: Inc., New York, NY (USA)., 2 (2008) 379.
25 Tat, P.V., Development of Quantitative Structure-Activity Relationships (QSARs) and Quantitative Structure-Property Relationships (QSPRs). Ha Noi, Viet Nam: Natural Science and Technology. (2009) 198.
26 Tat P V, *Development of New Anticancer Agents From Leaf of Plants in Viet Nam* (LAP Lambert Academic Publishing, GmbH & Co. KG, Germany), 80 (2017).
27 Sheldrick G M, A Short History of Shelx. *Acta Crystallographica Section A Foundations and Advances*, A64 (2008) 112.
28 D. Hadzi, J. Kidric, J. Koller And J. Mavri, *Journal of Molecular Structure*, 237 (1990) 139.
29 Yan Lu, Jianjun Chen, Min Xiao, Wei Li, and Duane D. Miller, *Pharmaceutical Research*, 29 (11) (2012) 2943.
30 Mengqi Dong, Fang Liu, Hongyu Zhou, Shumei Zhai, and Bing Yan., *Molecules*, 21(10) (2016).