THEORETICAL & COMPUTATIONAL CHEMISTRY | RESEARCH ARTICLE

Prediction of anticancer activities of cynaroside and quercetin in leaf of plants Cynara scolymus L and Artocarpus incisa L using structure–activity relationship

Bui Thi Phuong Thuy1, Nguyen Thi Ai Nhung1, Tran Duong2, Phung Van Trung3, Nguyen Minh Quang4, Hoang Thi Kim Dung3 and Pham Van Tat5*

Abstract: Natural products from plants are an alternative resource in the search for anti-cancer drugs and can have a direct impact on eliminating cancer cells and also reduce cancer side effects. Recently, we have isolated a few flavonoid quercetin and cynaroside from leaf of Cynara scolymus L and Artocarpus incisa L in Vietnam, with cytotoxic activity relatively strong in HeLa cancer cells. The flavonoid compound is a search target, research and development of anti-cancer agents in clinical use. To clarify the important nature of the activity, the subject QSAR studies on cancer HeLa cell line use the multiple linear regression (MLR) gradually, partial least square regression (PLS) and artificial neural network. The MLR and PLS models showed good correlation values of $R^2 = 0.938$, $R_{\text{pred}}^2 = 0.903$, and $R^2 = 0.943$, $R_{\text{pred}}^2 = 0.912$, respectively. The MLR model shows the level of importance of atomic charge descriptors. Also, artificial neural network architecture I(6)-HL(4)-O(1) is built with RMSE = 0.00345, $R^2 = 0.993$, $R_{\text{pred}}^2 = 0.971$ using the atomic charge descriptors selected in the MLR model such as neurons of input layer and the anti-cancer activity such as neuron of output layer. The anti-cancer activities of the flavonoids and isoflavonoids in the test group and

ABOUT THE AUTHORS
Scientific research for us is important and necessary in order to improve and expand the knowledge. Since 1990s, our role of scientific research had become to be important in a university, so we had some projects in QSAR field as:

- Studying the process of extracting and purifying the total alkaloid extract from Dichroa febrifuga Lour and the screening process for alkaloid from the extract of leaves of Dichroa febrifuga Lour. Using artificial neural network for screening process of alkaloids from Dichroa febrifuga Lour. During this duration, we had a number of scientific works in the field of computer applications in chemistry. For example, the QSAR in design for new drugs.

- In 1990s, we had a project of Ministry of Education and Training in the field of computer applications in the chemistry with the topic: Study of quantitative relationship between the structure and activity of the group antimalarial compounds, anti-cancer, anti-HIV, anti-fungal, and anti-bacterial.

PUBLIC INTEREST STATEMENT
The studies for Cynara scolymus L showed that its ingredients from flowers, leaves, stems, roots are very effective in healing and for food. The flavonoid compounds were extracted from Cynara scolymus L working in the treatment of some diseases such as liver, bile, cardio, antioxidants and reduce cholesterol in blood, especially HIV anti-virus (Loi, 2006).

In Vietnam, Artocarpus Incisa L scattered only has been planted in the orchard of the Vietnamese family. Artocarpus Incisa L is the kind of big trees. The compound groups in leaves Artocarpus Incisa L were determined quantitatively; the results showed that the leaves of Artocarpus Incisa L contain substances: flavonoids, saponins, anthranoid, tanin, reducing sugar, acid amines and polysaccharides. The water extract from the leaves of Artocarpus Incisa L showed that the blood pressure is lowered and decreased heart rate in mice. The water extract from leaves of Artocarpus Incisa L effects on cancer cells of the pancreas (Loi, 2006).
compounds quercetin and cynaroside isolated from cynara scolymus L and artocarpus incisa L are compared with experimental data and those from references.

Subjects: Computational and Theoretical Chemistry; Medicinal & Pharmaceutical Chemistry; Organic Chemistry

Keywords: QSAR_{MLR} and QSAR_{PLS} model; neural network QSAR_{ANN} model; anticancer activities Hela

1. Introduction

Natural products from plants are of interest in searching for new anti-cancer drugs and can have a direct effect on HeLa cancer cells and reduce side effects. Recently, we have isolated a few flavonoids from leaf of cynara scolymus L and artocarpus incisa L (Loi, 2006) and tested their in vitro activities pointed out the relatively strong impacts for cancer cells HeLa (Singh, Kaur, & Silakari, 2014). These flavonoids from leaf of cynara scolymus L (Apóstolo, Brutti, & Llorente, 2005; Fritsche, Beindorff, Dachtler, Zhang, & Lammers, 2002; Zhu, Zhang, & Lo, 2004) and artocarpus incisa L (El Senousy, Farag, Al-Mahdy, & Wessjohann, 2014) were also tested biologically in the treatment of some diseases such as liver, bile, cardio, antioxidants and reduce cholesterol in blood, especially HIV anti-virus (Abbasi & Samadi, 2014; Moreira, Castelo-Branco, Monteiro, Tavares, & Beltramin, 1998). Flavonoids are polyphenolic compounds in most plants (Mahapatra, Bharti, & Asati, 2015; Priyadarsini et al., 2010; Ziberna, Fornasaro, Čvorović, Tramer, & Passamonti, 2014). The flavonoids have shown their activities and role of food within flavonoids in the cancer inhibition are widely studied (Gavin & Durako, 2012; Lee, Boyce, & Breadmore, 2012; Pawlikowska-Pawlęga et al., 2014).

In recent years, the methods of quantum chemistry calculations are widely applied to the study of chemical properties and seeking new drugs. The field of new drug design by computer tools has become an important tool nowadays. Study on quantitative relationships between structure and activity (QSAR) of natural compounds is of concern for new drug researchers and pharmaceutical manufacturing facilitators. In Vietnam, there are a number of works of scientists from universities and institutions published in the journal (Phuong Thuy & Tat, 2012a, 2012b). The previous studies of 3-aminoflavonoid substances have focused on the basis of semi-empirical calculation (Tat, 2009a). These studies have shown a way for designing new drugs efficiently with the assistance of computers. The QSAR model can be predicted the biological activity of new drugs from the atomic charges in the molecule. This method allows for the identification of an active-central location of molecule.

The set of flavones and isoflavones is known to have an important activity against cervical cancer cells (Chen, 2008; Liao et al., 2005; Liao, Chen, Qian, Shen, & Zheng, 2008). This group is currently of interest for researching in different directions such as the synthesis and metabolizing of natural products or extracting them from plants (Loi, 2006). The consideration of the quantitative relationship between the structure of flavones and isoflavones with activity against cancer is an important issue in searching for the flavone and isoflavone derivatives to be valid.

In this work, we report the use of semi-empirical quantum calculations and construction of quantative structure–activity relationship (QSAR) models using 32 flavone and isoflavone derivatives (Chen, 2008; Liao et al., 2005, 2008). The flavones and isoflavones are constructed and optimized by means of molecular mechanics MM+. The atomic charge descriptors resulting from Parametric Model number 3 (PM3) method are used to build the multivariate QSAR models such as multiple linear regression (MLR), partial least squares regression (PLS), and artificial neural network (ANN). Anti-cancer activities GI_{50}/μM of flavones and isoflavones in the test group and the new flavonoids quercetin and cynaroside isolated from the leaves of cynara scolymus L and artocarpus incisa L are predicted from QSAR models and compared with those from experimental data.
2. Computational details

2.1. Materials and means
To ensure the accurate level of QSAR models, structural data and anticancer activities GI$_{50}$/μM for Hela cells (GI$_{50}$ is the concentration for 50% of maximal inhibition of cell proliferation) for flavones and isoflavones are taken from the data source of Wang and et al. (Chen, 2008; Liao et al., 2005, 2008) as pointed out in Figure 1 and Table 1. The anti-cancer activities are transformed into following value pGI$_{50}$

\[
pGI_{50} = -\log GI_{50}
\]  

(1)

The atomic charge parameters on molecules are calculated by means of program HyperChem v8.0 (HyperChem Release 8.03, 2008). The multiple linear regression (QSAR$_{MLR}$) and the partial least squares regression QSAR$_{PLS}$ models are built with program Origin 2015 (Tat, 2009b). The artificial neural network (QSAR$_{ANN}$) models are constructed with Visual Gene Developer 1.7 (Jung & McDonald, 2011).

2.2. Isolated technology of quercetin and cynaroside

2.2.1. Chemicals and equipment
In this work, we use the chemicals and the equipment for isolating and purifying two flavonoids quercetin and cynaroside before determining the substance structures by $^1$H-NMR and $^{13}$C-NMR spectrum:

- Silica gel with the particle size in range 0.04–0.06 mm was used for ordinary and Rp18 phase chromatography.
- Thin-layer chromatography was implemented by the thin plate DC-Alufolien F254 (Merck) for the ordinary phase and Rp18 F254s (Merck) for the reverse-phase chromatography.
- Solvents were used for the isolation processes: hexane, petroleum ether, chloroform, methanol, ethyl acetate, ethanol, acetone, distilled water.
- Reagent was used to trace out the compound coloration on plate: using H$_2$SO$_4$/EtOH; FeCl$_3$/EtOH.
- UV handheld lamps, 254 and 365 nm UVITEC effect.
- Vacuum Evaporators Buchi-111.
- Water Bath cooker JULABO 461.
- Infrared heating equipment SCHOTT.
- Chromatography column with diameter range 2–5.5 cm.
- Analytical Balances AND HR 200.
### Table 1. Molecular structure and anticancer activities GI_{50} (μM) of flavones and isoflavones (Chen, 2008; Liao et al., 2005, 2008).

| Substance | Name        | Substitutive site | Substitutes R | GI_{50} (μM) |
|-----------|-------------|-------------------|---------------|-------------|
|           | Training set for establishing QSAR models |                   |               |             |
| Fla1      | 1a-1        | Flavone           | C_{3}         | -OCH_{2}CCH=NOH | 2.0        |
| Fla2      | 2a-3        | Flavone           | C_{7}         | -OCH_{2}CCH=NOH | 2.0        |
| Isofla4   | 3a-4        | Isoflavone        | C_{7}         | -OCH_{2}CCH=NOH | 9.8        |
| Fla5      | 4a          | Flavone           | C_{3}         | -OCH_{2}CCH=NOCH_{3} | 2.0        |
| Fla6      | 5a          | Flavone           | C_{3}         | -OCH_{2}CCH=NOCH_{3} | 0.9        |
| Fla7      | 6a          | Flavone           | C_{3}         | -OCH_{2}CCH=NOCH_{3} | 2.2        |
| Isofla8   | 7a          | Isoflavone        | C_{3}         | -OCH_{2}CCH=NOCH_{3} | 8.5        |
| Fla10     | 8a          | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 2.1        |
| Fla11     | 9a          | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 2.0        |
| Fla13     | 10a         | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 1.6        |
| Fla14     | 11a         | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 1.0        |
| Fla17     | 12a         | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 2.0        |
| Isofla18  | 13a         | Isoflavone        | C_{3}         | -OCH_{2}C(H\_4)=NOH | 9.0        |
| Isofla19  | 14a         | Isoflavone        | C_{3}         | -OCH_{2}C(H\_4)=NOH | 7.8        |
| Isofla20  | 15a         | Isoflavone        | C_{3}         | -OCH_{2}C(H\_4)=NOH | 7.6        |
| Fla21     | 16a         | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 1.6        |
| Fla22     | 17a         | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 2.0        |
| Fla23     | 18a         | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 2.0        |
| Fla24     | 19a         | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 2.4        |
| Fla25     | 20a         | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 2.3        |
| Fla26     | 21a         | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 2.0        |
| Fla27     | 22a         | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 2.0        |
| Fla28     | 23a         | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 2.7        |
| Fla29     | 24a         | Flavone           | C_{3}         | -OCH_{2}C(H\_4)=NOH | 2.5        |
| Isofla30  | 25a         | Isoflavone        | C_{3}         | -OCH_{2}C(H\_4)=NOH | 8.2        |
| Isofla31  | 26a         | Isoflavone        | C_{3}         | -OCH_{2}C(H\_4)=NOH | 6.4        |
|           | Test set for validating QSAR models |                   |               |             |
| Fla2      | 1b          | Flavone           | C_{3}         | -OCH_{2}CCH=NOH | 1.2        |
| Fla9      | 2b          | Flavone           | C_{3}         | -OCH_{2}CCH=NOH | 1.8        |
| Fla12     | 3b          | Flavone           | C_{3}         | -OCH_{2}CCH=NOH | 0.8        |
| Fla15     | 4b          | Flavone           | C_{3}         | -OCH_{2}CCH=NOH | 2.0        |
| Fla16     | 5b          | Flavone           | C_{3}         | -OCH_{2}CCH=NOH | 2.0        |
| Isofla32  | 6b          | Isoflavone        | C_{3}         | -OCH_{2}C(H\_4)=NOH | 7.3        |

**Figure 2.** Separate equipment for two flavonoids quercetin and cynaroside.

a) Vacuum Evaporators  
b) Column chromatography at atmospheric and high pressure  
c) Thin-layer chromatography
2.2.2. Isolation and identification of flavonoids

To isolate and purify the flavonoid compounds, we used the techniques of thin-layer and column chromatography, as exhibited in Figure 2. After isolating the compounds, they were identified with the structure using different spectrums:

- Melting temperature carried out on Electrothermal IA 9000 series, using unadjusted capillary.
- Column chromatography with silica gel for ordinary-phase, reverse-phase chromatography Rp 18 and Sephadex techniques combined with thin-layer chromatography.
- Substances were detected by ultraviolet light at wavelengths 254 and 365 nm or reagent used is liquid H₂SO₄/EtOH or FeCl₃/EtOH.
- Nuclear magnetic resonance spectrum (NMR) ¹H-NMR (500 MHz) and ¹³C-NMR (125 MHz) implemented on Bruker AM500 FT-NMR Spectrometer (Figure 3).

2.3. Constructing QSAR models

The fitness models were proved using different validations. To validate QSAR models, the method was carried out: (1) leave-one-out cross-validation technique, (2) validation was implemented by dividing randomly the 32 compounds into training and test group. The developed model should be capable enough making accurate and reliable predictions of anticancer activities of new substances. So, the QSAR models that are constructed from a training set should be validated using the divided compounds in test group for testing the predictability of the developed models. The validation methods depicted the reliability of the developed models for their applicability on a group of new compounds, and confident level of predictability can thus be resolved (Tat, 2009). For all the newly-constructed models, the multiple-determined coefficient (R²) and leave-one-out cross-validation value (R²pred) for the training set were evaluated by (OriginLab Corporation, 2015; Tat, 2009). However, additionally, test set was used for calculation of (R²pred) values. The acceptability and the predictability of the model are determined based on statistical parameters such multiple-correlation coefficient (R²) of training group and test group and cross-validation coefficient (R²pred), Fisher value (F), and standard deviation (SD). Multiple linear models were set (Tat, 2009).

\[ Y = \sum_{i=1}^{k} a_i x_i + b \]  

where Y is the biological activity pGI₅₀ (the dependent variable), \( x_i \) is the atomic charge parameters (independent variable). Check the results from the model with experimental data based on single-factor analysis of variance (ANOVA). The correlation coefficient R², absolute relative error (ARE, %) and mean of absolute relative error (MARE,%) were calculated according to the Equation (3-5) (Tat, 2009).
Here $Y$ is $pGI_{50}$ experimental value and $\hat{Y}$ is $pGI_{50}$ predicted value; $\bar{Y}$: $pGI_{50}$ is mean value; The statistical error values $ARE,\%$ are determined by

$$ARE,\% = 100|\left( \frac{pGI_{50,exp} - pGI_{50,pred}}{pGI_{50,exp}} \right)|$$

where $pGI_{50,exp}$ and $pGI_{50,pred}$ are experimental and prediction activities. The average value of absolute relative error $MARE,\%$ is used to assess the global uncertainty of QSAR model using the formula:

$$MARE,\% = \frac{100}{N} \left( \frac{\sum (pGI_{50,exp} - pGI_{50,pred})}{pGI_{50,exp}} \right)$$

$N$ as number of activity values.

3. Results and discussion

3.1. Calculation of charge parameters

The molecular structures were built and optimized by means of MM+ molecular mechanics. The atomic charge parameters on molecules were calculated by semi-empirical quantum chemistry method SCF PM3 using the optimized molecules. The program HyperChem 8.05 (HyperChem Release 8.03, 2008) was used for all these calculations. The atomic charge parameters were used to build the multiple linear regression (QSAR MLR), partial least squares regression (QSAR PLS), and artificial neural network (QSAR ANN) model.

3.2. Constructing QSAR MLR and QSAR PLS model

Before conducting the QSAR MLR and QSAR PLS modeling, the activity values $GI_{50}(\mu M)$ are transformed into the values $pGI_{50}$ to adapt the statistical properties. The activity values $pGI_{50}(\mu M)$ are the most appropriate value. The QSAR MLR models were established using the relationship of the atomic charge predictors and biological activities $pGI_{50}$ (Tat, 2009). The change of values $R^2, R^2_{pred}$ and SE (standard error) in the QSAR MLR models with the atomic charge predictors, respectively, are pointed out in Table 2.

To have those QSAR MLR models, the atomic charge descriptors were selected using stepwise regression algorithm. The selection process for atomic charge descriptors is based on the change of

| $k$ | The atomic charge predictors in models | $R^2$ | SE  | $R^2_{pred}$ |
|-----|--------------------------------------|-------|-----|--------------|
| 2   | $O_1, C_7$                           | 0.816 | 0.139 | 0.765        |
| 3   | $O_1, C_4, C_6$                      | 0.860 | 0.124 | 0.800        |
| 4   | $O_1, C_2, C_4, C_5$                 | 0.901 | 0.107 | 0.829        |
| 5   | $O_1, O_11, C_3, C_4, C_6, C_7, C_8$ | 0.924 | 0.096 | 0.873        |
| 6   | $O_1, O_11, C_3, C_4, C_5, C_6, C_7$ | 0.938 | 0.089 | 0.903        |
| 7   | $O_1, O_11, C_5, C_6, C_7, C_8, C_9$ | 0.959 | 0.074 | 0.879        |
| 8   | $O_1, C_2, C_3, C_4, C_5, C_6, C_7$ | 0.970 | 0.065 | 0.696        |
| 9   | $O_1, O_11, C_3, C_5, C_6, C_7, C_9$ | 0.978 | 0.057 | 0.563        |
| 10  | $O_1, O_11, C_3, C_4, C_5, C_6, C_7$ | 0.978 | 0.059 | 0.358        |
the statistical values $R^2$, SE, $R^2_{\text{pred}}$, and $F$-stat. The QSAR$_{\text{MLR}}$ models were cross-validated using leave-one-out (LOO) technique to determine $R^2_{\text{pred}}$. The 10 fitness models are shown in Table 2. The QSAR$_{\text{MLR}}$ models (with $k$ from 2 to 10) that are arranged in an orderly change of values $R^2$ and $R^2_{\text{pred}}$. From the models in Table 2, the QSAR$_{\text{MLR}}$ models (with $k$ from 5 to 7) are shown the greater values of $R^2_{\text{pred}}$ than others.

In particular, the QSAR$_{\text{MLR}}$ model with $k = 6$ with value $R^2$ of 0.938 gave the highest value $R^2_{\text{pred}}$ of 0.903. So we selected the best models (with $k$ of 5, 6, and 7) to determine the contribution percentage of atomic charges. The valuable contribution percentages $MP_{m^*k},\%$, $GMP_{m^*k},\%$ and the statistical values of these models (with $k$ of 5, 6, and 7), respectively, are exhibited in Table 3.

The valuable contribution percentages $MP_{m^*k},\%$ of independent variables in each model QSAR$_{\text{MLR}}$ (with $k$ of 5, 6, and 7) were determined from the contribution percentages $PX_k,\%$ of variables in each case, respectively (Tat, 2009). This value is determined by the total value of contribution $C_{\text{total}}$ of variables in a substance (Chen, 2008). So the average contribution percentage $MP_{m^*k},\%$ of each variable is defined by the formula (6) and the results are depicted in Table 3.

\[
MP_{m^*k},\% = \frac{1}{N} \sum_{j=1}^{N} \left( 100 \left| b_{m^*j}X_{m^*j} / C_{\text{total}} \right| \right) \quad \text{with} \quad C_{\text{total}} = \sum_{j=1}^{k} \left| b_{m^*j}X_{m^*j} \right|
\]

where $N$ the total number of cases, $m$ number of variables. The global average contribution percentage $GMP_{m^*k},\%$ of each independent variable for three models is determined by the formula (7):

\[
GMP_{m^*k},\% = \frac{1}{n} \sum_{n=1}^{3} MP_{m^*k}
\]

Table 3. Statistical values and valuable contribution percentages $MP_{m^*k},\%$ and $GMP_{m^*k},\%$ for atomic charges in the models QSAR$_{\text{MLR}}$ (with $k$ of 5, 6, and 7)

| Variable | QSAR$_{\text{MLR}}$ | $MP_{m^*k},\%$ | $GMP_{m^*k},\%$ |
|----------|---------------------|----------------|----------------|
| $x_1$    |                     | $k = 5$       | $k = 6$       | $k = 7$       |
| $R^2$    | 0.9243              | 0.9382        | 0.9589        |
| $R^2_{\text{adj}}$ | 0.9053          | 0.9186        | 0.9429        |
| SE       | 0.0957              | 0.0887        | 0.0743        |
| $R^2_{\text{pred}}$ | 0.873          | 0.903         | 0.879         |
| Constant | $-0.9332$           | 6.7116        | 4.714         |
| $O_1$    | $-101.2076$         | $-42.3105$    | -             | 57.6024       | 24.6289       | -             | 27.4104       |
| $O_11$   | -                   | $-8.1592$     | $-32.8026$    | -             | 18.6316       | 21.4621       | 13.3646       |
| $C_2$    | $-15.4264$          | -             | 13.4176       | -             | -             | -             | 4.4725        |
| $C_3$    | -                   | 3.0139        | -             | -             | 4.2160        | -             | -             | 4.4725        |
| $C_4$    | $-6.8735$           | $-19.0370$    | $-60.0703$    | 15.1206       | 42.4467       | 38.3868       | 31.9847       |
| $C_5$    | $-7.9686$           | -             | 2.0583        | -             | -             | -             | 0.6861        |
| $C_6$    | -                   | 6.6117        | 20.8772       | -             | 6.5716        | 5.7785        | 4.1167        |
| $C_7$    | -                   | 4.6038        | -             | 3.5052        | -             | -             | 1.1684        |
| $C_8$    | -                   | -             | 16.9016       | -             | -             | 5.6960        | 1.8987        |
| $C_9$    | -                   | -             | 95.4205       | -             | -             | 22.5970       | 7.5323        |
| $C_{10}$ | -                   | -             | $-24.4720$    | -             | -             | 2.4973        | 0.8324        |
| $C_{11}$ | $-16.1166$          | -             | 11.8011       | -             | -             | 3.9337        |
| $C_{12}$ | -                   | -             | $-25.4219$    | -             | -             | 3.5824        | 1.1941        |
The contribution percentages $GMP_{mxk}\%$ in Table 3 display the important level of atomic charges in flavones and isoflavones. For 3 QSAR$_{MRL}$ models, the important level of atomic charges are that arranged by the values $GMP_{mxk}\%$: C$_4$ $>$ O$_1$ $>$ O$_{11}$ $>$ C$_9$ $>$ C$_2$ $>$ C$_6$ $>$ C$_3$. The atom positions C$_4$, O$_1$, O$_{11}$ are considered such as the most important positions in the molecules. Besides those atoms are in carbonyl group C$_4$ = O$_{11}$ and atom O$_1$ has free electron pair conjugating with π electronic bond C$_2$ = C$_3$, and C$_4$ = O$_{11}$ to form a conjugate system. The carbonyl group C$_4$ = O$_{11}$ has fully reactive natures of carbonyl substance. So, these important atoms are demonstrated quantitatively using the $GMP_{mxk}\%$ values and this is also consistent with the verdicts from experimental evaluation (Lee et al., 2012; Liao et al., 2005, 2008). Also, the atomic position C$_6$ is also an important position and is explored for attaching the new substitutes (Chen, 2008; Liao et al., 2005, 2008). The atomic positions C$_9$ and C$_3$ also represent the important impacts for biological activities GI$_{50}$, but the C$_9$ atom is not vacant position so should not be selected for attaching the new substitutes. So the C$_9$ is vacant position can be chosen to add the new substitutes to sample flavone in Table 1 or new flavonoid. Similarly, the position C$_3$ is also empty and can be utilized to add the new substitutes. Those can hope to constitute a new compound with higher activity. From this orientation, a new flavonoid isolating from the leaf of *artocarpus incisa* L was selected such as sample substance to design new drugs with high activity. This is carried out in below discussion.

The QSAR$_{PLS}$ model is also built from the atomic charges, in which those were selected for the QSAR$_{MRL}$ model (Tat, 2009). The six variables O$_1$, O$_{11}$, C$_3$, C$_4$, C$_6$, and C$_7$ are also used to build the QSAR$_{PLS}$ models. The present results of biological activities are depicted in $R^2$ values in which those are consistent with experimental data. The partial least squares (QSAR$_{PLS}$) model exhibited in the form:

$$Y = 5.168 - 20.643 \times O_1 - 0.358 \times C_3 - 7.892 \times C_4 + 0.425 \times C_6 - 0.583 \times C_7 - 3.465 \times O_{11}$$

(8)

$$R^2 = 0.943; \text{SE} = 0.360; R^2_{\text{pred}} = 0.912.$$  

3.3. Building QSAR$_{ANN}$ model

The QSAR$_{ANN}$ model is built by the neuro-fuzzy technique with the genetic algorithms using program Visual Gene Developer v1.7 (Jung & McDonald, 2011). The artificial neural network architecture consists of three layers I(6)-HL(4)-O(1); the input layer I(6) includes six neurons as parameters O$_1$, O$_{11}$, C$_3$, C$_4$, C$_6$, and C$_7$; the neuron on output layer O(1) is biological activity pGI$_{50}$; the hidden layer HL(4) consists of four neurons. This multi-layer neural network employing backpropagation algorithm is used to train the network. The transfer function is sigmoid on each node of the network; the neural network parameters include the training rate of 0.7 and learning rate of 0.7; the goal monitoring error MSE = 0.000816 with 10,000 iteration. After training the neural network, $R^2$ value is 0.993 and $R^2_{\text{pred}}$ of 0.971 while for QSAR$_{MRL}$ model, the value $R^2$ is 0.938 and $R^2_{\text{pred}}$ of 0.903.

3.4. Prediction of biological activity for new substance

The predictability of the models QSAR$_{MRL}$, QSAR$_{PLS}$ and QSAR$_{ANN}$ are evaluated carefully using the leave-one-out (LOO) technique to determine the value $R^2_{\text{pred}}$ the flavonoids were divided randomly from the data in Table 1 into the training group of 26 compounds and the test group of six compounds. The biological activities of six flavonoids in the test group in Tables 1 and 2 new flavonoids isolated from the leaves of *cynara scolymus* L and *artocarpus incisa* L (Loi, 2006) are predicted from models QSAR$_{MRL}$, QSAR$_{PLS}$, and QSAR$_{ANN}$.

The predicted values of biological activities for those are compared with experimental values, as presented in Table 4. The substance cynaroside is isolated from the leaf of *cynara scolymus* L (Loi, 2006) and its structure is identified using the different spectra such as: $^1$H-NMR (500 MHz, DMSO, $\delta$ ppm): $\delta$ 6.39 (s, H$_4$); 6.44 (d, J = 2 Hz, H$_6$); 6.78 (d, J = 2 Hz, H$_8$); 5.08 (1H, d, J = 7.5 Hz, H$_1$); 7.41 (d, J = 8.5 Hz, H$_7$); 6.90 (d, J = 8.5 Hz); 7.44 (dd, J = 8.5 Hz, J = 2); 3.489 (1H, m, H$_2$); 3.476 (1H, m, H$_4$); 3.466(1H, m, H$_6$); 3.500(1H, m, H$_8$); 3.725(1H, dd, J = 12.3; 2.5 Hz, H$_{2'}$); 3.702(1H, dd, J = 12.4; 6.3 Hz, H$_{2''}$); 12.8 (C$_5$OH); 3.466–3.725 (7H, m, glucose protons). The associated spectrum was also used to
have more structural information such as: $^{13}$C-NMR (DMSO, $\delta$ ppm, 125 MHz), DEPT: $\delta$ 164.5(C$_2$); 105.3(C$_3$); 181.785 (C$_4$); 161.2 (C$_5$); 100.2 (C$_6$); 162.9 (C$_7$); 95 (C$_8$); 156.9 (C$_9$); 103.1 (C$_{10}$); 121.4 (C$_{1'}$); 113.6 (C$_{2'}$); 145.7 (C$_{3'}$); 149.9 (C$_{4'}$); 115.9 (C$_{5'}$); 119.2 (C$_{6'}$); 100.0 (C$_{1''}$); 76.4 (C$_{2''}$); 77.2 (C$_{3''}$); 69.5 (C$_{4''}$); 78 (C$_{5''}$); 60.6 (C$_{6''}$). Interaction of C and H in spectrum HMBC and HSQC are also pointed out: H$_{6'}$-C$_5$-C$_7$-C$_8$-C$_{10}$; H$_8$-C$_6$-C$_7$-C$_9$-C$_{10}$; H$_{2''}$-C$_2$-C$_1'$-C$_3'$-C$_4'$-C$_6'$; H$_{5''}$-C$_1'$-C$_3'$-C$_4'$-C$_6'$; H$_{6''}$-C$_2$-C$_1'$-C$_2'$-C$_4'$-C$_5'$. 

The substance quercetin is isolated from the leaf of *artocarpus incisa* L (Loi, 2006) and its structure is also identified using the spectrum 1H-NMR (DMSO-d$_{6}$, 500 MHz, $\delta$ ppm) combining with spectrum HSQC, HMBC: $\delta$ 6.26 (1H, d, $J = 1.5$ Hz, H$_6$); $\delta$ 6.52 (1H, s like t, H$_8$); $\delta$ 7.82 (1H, d, $J = 1.5$ Hz, H$_2'$); $\delta$ 7.68 (1H, dd, $J = 8.5$ và 2 Hz, H$_7$); $\delta$ 6.98 (1H, d, $J = 8.5$ Hz, H$_3'$); $\delta$ 12.16 (1H, s). Also, using spectrum $^{13}$C-NMR (DMSO-d$_{6}$, 125 Hz) combining with spectra DEPT, HSQC, HMBC: $\delta$ 146.9 (C$_2$); $\delta$ 136.6 (C$_3$); $\delta$ 176.5 (C$_4$); $\delta$ 162.2 (C$_5$); $\delta$ 99.1 (C$_6$); $\delta$ 164.9 (C$_7$); $\delta$ 94.4 (C$_8$); $\delta$ 157.7 (C$_9$); $\delta$ 104.0 (C-10); $\delta$ 121.4 (C$_{1'}$); $\delta$ 116.1 (C$_{2'}$); 145.7 (C$_{3'}$); $\delta$ 148.2 (C$_{4'}$); $\delta$ 115.6 (C$_{5'}$); $\delta$ 123.7 (C$_{6'}$).

### Table 4. Activities pGI$_{50}$ of test group resulting from models QSAR$_{MLR}$, QSAR$_{PLS}$, and QSAR$_{ANN}$

| Substance | Reference | pGI$_{50}$,exp | pGI$_{50}$,pred | ARE (%) |
|-----------|-----------|---------------|----------------|---------|
|           |           | QSAR$_{MLR}$  | QSAR$_{PLS}$  | QSAR$_{ANN}$ | QSAR$_{MLR}$ | QSAR$_{PLS}$ | QSAR$_{ANN}$ |
| Fla2      | (Lee et al., 2012; Liao et al., 2005, 2008) | 5.9208 | 6.0079 | 5.8012 | 5.8509 | 1.4719 | 2.0200 | 1.1802 |
| Fla9      | (Lee et al., 2012; Liao et al., 2005, 2008) | 5.7447 | 5.6915 | 5.6082 | 5.7407 | 0.9252 | 2.3758 | 0.0678 |
| Fla12     | (Lee et al., 2012; Liao et al., 2005, 2008) | 6.0969 | 5.7587 | 5.8416 | 5.8136 | 5.5478 | 4.1875 | 4.6463 |
| Fla15     | (Lee et al., 2012; Liao et al., 2005, 2008) | 5.6990 | 5.6511 | 5.6518 | 5.7124 | 0.8402 | 0.8275 | 0.2357 |
| Fla16     | (Lee et al., 2012; Liao et al., 2005, 2008) | 5.6990 | 5.6514 | 5.6549 | 5.7188 | 0.8350 | 0.7746 | 0.3466 |
| Isofla32  | (Lee et al., 2012; Liao et al., 2005, 2008) | 5.1367 | 5.0917 | 5.0830 | 5.1115 | 0.8767 | 1.0457 | 0.4090 |
| Cynaroside| This work | 5.3260 | 5.1910 | 5.6317 | 5.3186 | 2.5350 | 5.7393 | 0.1388 |
| Quercetin | This work | 5.3790 | 4.5858 | 5.5355 | 5.3591 | 14.7455 | 2.9094 | 3.9388 |

MARE (%) 3.4722 2.4850 1.3808

### Figure 4. The molecular structures of: (a) cynaroside and (b) quercetin.

**a) leaf of *cynara scolymus* L [1]**

**b) leaf of *artocarpus incisa* L [1]**

**Cynaroside with GI$_{50}$,exp (µM) = 4.72 ± 0.280**

**Quercetin with GI$_{50}$,exp (µM) = 4.18 ± 0.327**
The molecular structures of substances cynaroside and quercetin are shown in Figure 4.

After isolation of two new flavonoids cynaroside and quercetin, their activities pGI50 were conducted to test in vitro toxicity on Hela cells in the laboratory of molecular biology, university of natural sciences. The activity values pGI50 of two these flavonoids were also predicted from the models QSARMLR, QSARPLS, and QSARANN, as shown in Table 4. Those were compared with experimental activities and with each other based on the average value of the absolute relative error MARE,%.

The predictability of the model QSARMLR is lower than models QSARPLS and QSARANN, respectively, as given in Table 4. The QSARANN model has the valuable error MARE,% of 1.3808. This is smaller than values MARE,% of both models QSARMLR and QSARPLS. So, the predictability of QSARANN model is better than models QSARMLR and QSARPLS. After using the models QSARMLR, QSARPLS, and QSARANN to predict the biological activities pGI50 of six compounds in test group and two new flavonoids quercetin and cynaroside, the accurate level of the predicted results is exhibited in the acceptable errors within the uncertainty of experimental measurements. Thus, the models QSARMLR, QSARPLS and QSARANN are good adaptable for predicting the biological activities of new substances.

In this work, we selected the new flavonoid quercetin with the vacant positions C6 and C3' such as sample compound for designing five new compounds. The substitutes were attached to two vacant positions C6 and C3', as shown in Table 5. The new designed compounds were also predicted with the biological activities pGI50 using the QSARANN model. Then, the predictive activities pGI50 were recovered in the original form GI50 (μM), as given Table 5.

The predicted results pGI5 for new substances are transformed into values GI50 (μM) and compared with experimental activity of sample quercetin, as depicted in Figure 5. Thus, the five new compounds designing from the C6 and C3' positions on quercetin displayed stronger activity GI50 (μM) than sample quercetin. Herein, the new designed compounds will promise to forward a designing plan for the new pharmaceutical products from natural products.

### Table 5. The anti-cancer activities GI50 (μM) of five new flavonoids (n) designing from the vacant positions C6 and C3' on quercetin, resulting from QSARANN model

| New substance | Substitutes at C6 | Substitutes at C3' | GI50 (μM) | Method in this work |
|---------------|------------------|-------------------|-----------|---------------------|
| Quercetin     | -H               | -H                | 4.18 ± 0.327 | in vitro test on Hela |
| Fla-1(n)      | -OCH2CONHCH3     | -OH               | 0.1539    | QSARANN             |
| Fla-2(n)      | -OCH2CONHC6H4F   | -H                | 0.1487    | QSARANN             |
| Fla-3(n)      | -OH              | -OCH2CONHCH3      | 0.1247    | QSARANN             |
| Fla-4(n)      | -OCH2CHC=NOH     | -OH               | 0.1233    | QSARANN             |
| Fla-5(n)      | -OCH2CONHC6H4OCH3| -H                | 0.1174    | QSARANN             |
4. Conclusion

We used the quantum chemistry calculations, multivariable regression and artificial neural network to construct successfully the quantitative relationships between the partial atomic charges and anti-cancer activities \( GT_{50} \) (\( \mu M \)) of flavonoids. The models QSAR\(_{\text{MLR}}\) showed six important sites \( O_1, O_1', C_3, C_4, C_6, \) and \( C_7 \) on flavonoids.

The QSAR\(_{\text{MLR}}\) model found out the most important positions \( C_6 \) and \( C_3' \) to add the new substitutes to create five new flavonoids with higher activity of quercetin isolating from the leaf of artocarpus incisa L. The QSAR\(_{\text{ANN}}\) model with architecture I(6)-HL(4)-O(1) is better predictable for flavonoids.

**Supplementary material**

The supplementary material for this paper is available online at [http://dx.doi.org/10.1080/23312009.2016.1212452](http://dx.doi.org/10.1080/23312009.2016.1212452).

**Funding**

The authors received no direct funding for this research.

**Author details**

Bui Thi Phuong Thuy
E-mail: phuongthuy.trt@gmail.com

Nguyen Thi Ai Nhung
E-mail: ainhunguyen.chem@gmail.com

Tran Duong
E-mail: duongtran2001@gmail.com

Phung Van Trung
E-mail: trung_cnhh@yahoo.com

Nguyen Minh Quang
E-mail: nguyenminhquang@hui.edu.vn

Hoang Thi Kim Dung
E-mail: hoangthikimdung@gmail.com

Pham Van Tat
E-mail: vantat@gmail.com

1. Department of Chemistry, University of Science, Hue City, Vietnam.
2. Department of Chemistry, University of Education, Hue City, Vietnam.
3. Institute of Chemical Technology, Vietnam Academy of Science and Technology, Ho Chi Minh City, Vietnam.
4. Department of Chemical Engineering, Industrial University of Ho Chi Minh, Ho Chi Minh City, Vietnam.
5. Faculty of Science and Technology, Hoa Sen University, Ho Chi Minh City, Vietnam.

**Citation information**

Cite this article as: Prediction of anticancer activities of cynaroside and quercetin in leaf of plants Cynara scolymus L and Artocarpus incisa L using structure–activity relationship, Bui Thi Phuong Thuy, Nguyen Thi Ai Nhung, Tran Duong, Phung Van Trung, Nguyen Minh Quang, Hoang Thi Kim Dung & Pham Van Tat, Cogent Chemistry (2016), 2: 1212452.

**References**

Abbasì, F., & Samadi, F. (2014). Effect of different levels of artichoke (Cynara scolymus L.) Leaf powder on the performance and meat quality of Japanese quail. Poultry Science Journal, 2, 95–111.

Apóstolo, N. M., Bratti, C. B., & Llorente, B. E. (2005). Leaf anatomy of Cynara scolymus L. in successive micropropagation stages. In Vitro Cellular & Developmental Biology-Plant, 41, 307–313.

Chen, I. L. (2008). Synthesis and antiproliferative evaluation of amide-containing flavone and isoflavone derivatives. Bioorganic & Medicinal Chemistry, 16, 7639–7645.

El Senousy, A. S., Farag, M. A., Al-Mahdy, D. A., & Wessjohann, L. A. (2014). Developmental changes in leaf phenolics composition from three artichoke cvs. (Cynara scolymus) as determined via UHPLC-MS and chemometrics. Phytochemistry, 108, 67–76.

Fritsche, J., Beindorff, C. M., Dachtler, M., Zhang, H., & Lammers, J. G. (2003). Isolation, characterization and determination of minor artichoke (Cynara scolymus L.) leaf extract compounds. European Food Research and Technology, 215, 149–157.

Gavin, N. M., & Durako, M. J. (2012). Localization and antioxidant capacity of flavonoids in Halophila johnsonii in response to experimental light and salinity variation. Journal of Experimental Marine Biology and Ecology, 32–40, 416–417.

http://www.sciencedirect.com/science/article/pii/S0022098112000457

HyperChem Release 8.03. (2008). USA: Hypercube.

Lee, I. S., Boyce, M. C., & Breadmore, M. C. (2012). Extraction and on-line concentration of flavonoids in Brassica oleracea by capillary electrophoresis using large volume sample stacking. Food Chemistry, 133, 205–211.

http://dx.doi.org/10.1016/j.foodchem.2012.01.006

Liao, S. Y., Chen, J. C., Qian, L., Shen, Y., & Zheng, K. C. (2008). QSAR, action mechanism and molecular design of flavone and isoflavone derivatives with cytotoxicity against HeLa. European Journal of Medicinal Chemistry, 43, 2159–2170.

http://dx.doi.org/10.1016/j.ejmech.2007.10.033

Liao, S. Y., Chen, J. C., Qian, L., Shen, Y., Zheng, K. C., Wang, T. C., and coworkers. (2005). Journal of Bioorganic and Medicinal Chemistry, 13, 6045–6053.

Loi, D. T. (2006). Medicinal plants and drugs from Vietnam. Hanoi: Publisher of Medicine.

Mahapatra, D. K., Bharti, S. K., & Asati, V. (2015). Anti-cancer chalcones: Structural and molecular target perspectives. European Journal of Medicinal Chemistry, 98, 69–114.

Moheiri, R. A., Castelo-Branco, C. C., Monteiro, A. C., Tavares, R. O., & Beltramini, L. M. (1998). Isolation and partial characterization of a lectin from Artocarpus incisa L. seeds. Phytochemistry, 47, 1183–1188.

http://dx.doi.org/10.1016/S0031-9422(97)00753-X
OriginLab Corporation. (2015). Tutorials for Origin 2015 SR0.
http://cloud.originlab.com/pdfs/Origin2015_Documentation/EnglishOrigin_Tutorial_2015_E.pdf

Pawlowska-Pawełga, B., Dziubińska, H., Knöl, E., Trębacz, K., Jarosz-Wilkolazka, A., Paduch, R., & Gruszecki, W. I. (2014). Characteristics of quercetin interactions with liposomal and vacuolar membranes. Biochimica et Biophysica Acta (BBA)-Biomembranes, 1838, 254–265.
http://www.sciencedirect.com/science/article/pii/S0005273613002915

Phuong Thuy, B. T., & Tat, P. V. (2012a). QSAR models: Prediction of physicochemical property and anticancer activity of structurally similar Flavone and Isoflavone derivatives. Vietnam Journal of Chemistry, 5, 550–556.

Phuong Thuy, B. T., & Tat, P. V. (2012b). Development of QDSAR models for anticancer flavone and Isoflavone analogues using 13C-NMR and 15O-NMR. Vietnam Journal of Chemistry, 50, 203–208.

Priyadarshini, R. V., Murugan, R. S., Maitreyi, S., Ramalingam, K., Karunagaran, D., & Nagini, S. (2010). The flavonoid quercetin induces cell cycle arrest and mitochondria-mediated apoptosis in human cervical cancer (HeLa) cells through p53 induction and NF-κB inhibition. European Journal of Pharmacology, 649, 84–91.
http://dx.doi.org/10.1016/j.ejphar.2010.09.020

Singh, M., Kaur, M., & Silokan, O. (2014). Flavones: An important scaffold for medicinal chemistry. European Journal of Medicinal Chemistry, 84, 206–239.
http://dx.doi.org/10.1016/j.ejmech.2014.07.013

Tat, P. V. (2009a). QSDAR model: Quantitative spectrum Data 13C-NMR and 15O-NMR activity relationships (QDSAR) of a set of anti-cancer 3-aminoflavonoid., Vietnamese Journal of Chemistry and Application, 14, 43–46.

Tat, P. V. (2009b). Development of quantitative structure-activity relationships (QSARs) and quantitative structure-property relationships (QSPRs). Hanoi: Publisher of Natural Science and Technology.

Zhu, X., Zhang, H., & Lo, R. (2006). Phenolic compounds from the leaf extract of Artichoke (Cynara scolymus L.) and their antimicrobial activities. Journal of Agricultural and Food Chemistry, 52, 7272–7278. http://dx.doi.org/10.1021/jf0609192

Ziberna, L., Fornasaro, S., Čvorović, J., Tramer, F., & Passamonti, S. (2014). Bioavailability of flavonoids. Polyphenols in Human Health and Disease, 1, 489–511.
http://dx.doi.org/10.1016/B978-0-12-398456-2.00037-2