QESAR STUDY OF TRIPEPTIDE ANALOGUES AS ANTIOXIDATION AGENTS

Nong Thi Hong Duyen(1) – **Pham Van Tat**(2)

(1) Hue University of Science; (2) Thu Dau Mot University

ABSTRACT

A database consisting of 23 tripeptides was used to study the quantitative relationships between electric surface potential descriptors and antioxidant activity QESARs. The important structural descriptors SaaNH_acnt, SsOH_acnt, SaaN, SaaN_acnt, SsssCH, SaaaC, SsNH3p, SdO, SdO_acnt were selected for constructing the linear models QESARs with genetic algorithm. The best 4-variable linear model $QESAR_{linear}$ *including the structural descriptors SaaN, SdO, SdO_acnt and SsOH_acnt was constructed. The quality* $QESAR_{linear}$ *was exhibited in statistical values* $R^2_{fitness}$ *of 97.5660, standard error of estimation SE of 0.0378, F-stat of 130.2731,* R^2_{test} *of 93.3851. The non-linear model as neural network model* $QESAR_{neural}$ *I(4)-HL(3)-O(1) with* $R^2_{fitness}$ *of 98.2296 was built by using structural descriptors in* $QESAR_{linear}$ *model. The antioxidation activities of tripeptides resulting from* $QESAR_{linear}$ *and* $QESAR_{neural}$ *model were pointed out in values MARE, % of 27.4282 and 20.0672, respectively.*

Keywords: *QESARs model, multiple regression,*
neural network and antioxidation tripeptides

*

1. Introduction

The antioxidation compounds prevent the biological and chemical substances from radical-induced oxidation damage [4]. The hydrolysis from various proteins, such as soybean, casein, bullfrog, royal jelly, venison, r-lactalbumin, myofibrillar, rice endosperm, have been shown to have antioxidant activities against the peroxidation of lipids or radical scavenging activities [1].

Relationships between structural descriptors (electric surface potential) and antioxidation activities QESAR may indicate quantitatively change of biological activity or physicochemical properties corresponding to composition of amino acids in peptide chain [2], [3].

This work reports the use of multivariate regression and neuro-fuzzy technique with genetic algorithm to construct the quantitative relationships between electric surface potential descriptors and antioxidation activities for tripeptides. The electric surface potential descriptors of tripeptides are calculated by incorporating molecular mechanics MM+ and semiempirical quantum chemical calculation SCF PM3. The linear model $QESAR_{linear}$ and non-linear model $QESAR_{neural}$ are founded by those structural descriptors. The antioxidant activities of tripeptides resulting from these models QESARs are compared to those from literature.

2. Methodology

2.1. Antioxidant data

The experimental data of 23 antioxidation tripeptides used in this study (AC_{exp}: antioxidant activities of peptides were measured by the ferric thiocyanate methods which are relative activities by adjusting the control 1.0) were taken from a source of Li Yao Wang [1]. The experimental data were divided into the training set as calibration group and the test set as external validation set. The validation set of 5 tripeptides was derived randomly from original data. The remaining tripeptides were constituted the training set. This set includes 18 tripeptides with values of experimental activities, as listed in Table 1. The AC_{exp} values in range 0.0441 – 0.6369 were used to fit for the adjustable parameters of QESAR models. The test set consisting of 5 tripeptides in Table 5 with AC_{exp} values in range 0.3170 – 0.6369 was used to evaluate its predictability.

Table 1. *The tripeptide structures and experimental antioxidant values AC_{exp}, respectively [1]*

N°	Tripeptide	AC_{exp}	N°	Tripeptide	AC_{exp}
1	CYY	0.4699	13	HHR	0.0635
2	HHA	0.0680	14	HHS	0.0862
3	HHC	0.1277	15	HHT	0.0862
4	HHD	0.1877	16	HKH	0.0441
5	HHE	0.1877	17	HRH	0.0441
6	HHG	0.3170	18	LWL	0.6061
7	HHI	0.0680	19	PWK	0.4066
8	HHK	0.0635	20	RWK	0.6061
9	HHL	0.0680	21	RWQ	0.6061
10	HHM	0.0817	22	RWV	0.6061
11	HHN	0.3170	23	YYC	0.6369
12	HHQ	0.3170			

2.2. Electric surface potential descriptors

The tripeptide structures were built and optimized by using MM+ molecular mechanics method and semi-empirical PM3 calculation level in package HyperChem [5]. The optimization was performed by Polak-Ribiere algorithm at gradient level 0.05. Tripeptide notation and their experimental antioxidant activities are presented in Table 1. Program QSARIS [7] was used to calculate the electric surface potential descriptors of each tripeptide, respectively. The electric surface potential descriptors with calculation techniques were pointed out in literature [9].

2.3. Regression analysis

A step-wise multiple linear regression MLR procedure was used for variable selection or model development. It is clear that MLR models can be obtained using a step-wise multiple regression procedure; among these models, the best one must be chosen [8], [9]. For this objective, it is common to consider four statistical parameters: the number of molecular descriptors, the square correlation coefficient (R^2), the standard Error (SE) and the F-stat value. A reliable MLR model is one that has high R^2 and F values, and low SE and number of descriptors. Multiple linear regression (MLR) techniques based on least-squares procedures are very often used for estimating the regression coefficients using program packages Regress [8] and QSARIS [7], [9].

2.4. Neural networks

Neural networks NNs are artificial intelligent systems. They use a large number of interrelated data-processing neurons to emulate the function of brain. Although there are several NN models in use today, the most frequently used type I(i)-HL(m)-O(n) in this research consists of three-layered back-propagation neural net. In this neural net, the neurons are arranged in an input layer I(i) with i neurons, a hidden layer HL(m) with m neurons, and an output layer O(n) with n neurons. Each neuron in any layer is fully connected with the neurons of another layer. The neural net was trained by using the parameters as sigmoid transfer function was applied to each node in the hidden layer, momentum 0.7, learning rate 0.7 and random seed 10,000 [6].

3. Results and discussion

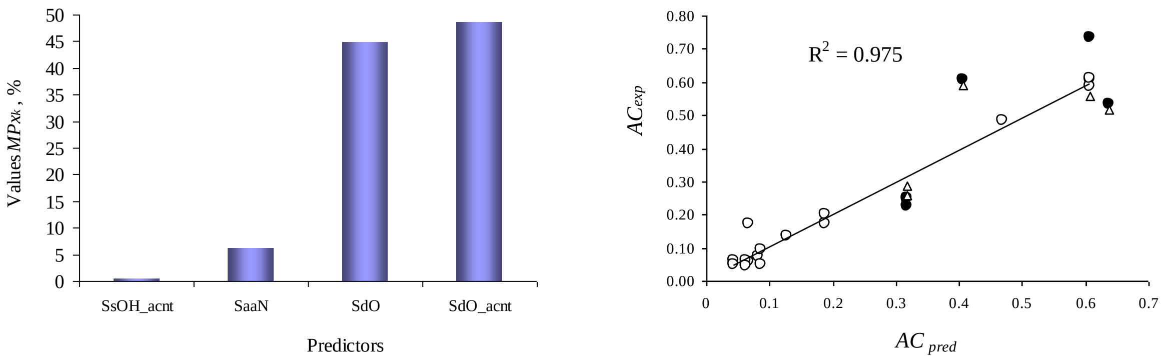
3.1. Variable selection and linear relationship

The correlation between the electric surface potential descriptors and experimental antioxidant values was first constructed based on the training set through linear regression analysis. Four descriptors SaaN, SdO, SdO_acnt and SsOH_acnt were identified and included in the $QESAR_{linear}$ model, and there was no significant correlation between the selected descriptors.

The electric surface potential descriptors were selected by using the linear regression techniques forward and back elimination. The best-suitable model QESARlinear (1) with four variables was selected to describe accurately the quantitative relationship between electric surface potential descriptors (X) and antioxidant values (Y).

$$AC = 0.4002 - 0.0753SaaN - 0.0671SdO + 0.8702SdO\_acnt - 0.0765SsOH\_acnt \quad (1)$$

The linear model $QESAR_{linear}$ (1) with $k = 4$ was adopted with statistical value R^2_{test} of 93.3851. The quality of this model $QESAR_{linear}$ was also reflected by value $R^2_{fitness}$ of 97.5660, standard error SE of of 0.0378 and F-stat of 130.2731. The t-Stat ratio values of coefficients in linear model $QESAR_{linear}$ were tested by statistical criteria at confident level a = 0.05. These turn out to be very satisfactory for statistical standards. This linear model $QESAR_{linear}$ (1) needs also to be validated by cross-validation and external validation. The cross-validation results showed that linear model $QESAR_{linear}$ (1) can be used to predict the antioxidant values of any tripeptides.



a) b)

***Figure 1**. a) Mean values of contribution percentage MPx_k,%; b) Correlation of values AC_{exp} versus AC_{pred} of training set (o) and the test set (●) for $QESAR_{linear}$ model and $QESAR_{neural}$ model (Δ)*

Moreover the important contribution of molecular descriptors in this model $QESAR_{linear}$ (1) was arranged in order SdO_acnt > SdO > SaaN > SsOH_acnt. These based on the mean values of contribution percentage MPx_k,% [9]. In this case the magnitude of regression coefficients orresponding to each descriptor was arranged in order SdO_acnt > SsOH_acnt > SaaN > SdO. The values $P_m x_k$,% and MPx_k,% for each predictor in model (1) was exhibited in Figure 1. So, the important contribution of each descriptor in this model $QESAR_{linear}$ (1) may not rely on the magnitude of the coefficient to make.

The values $P_m x_k$,% and MPx_k,% in Figure 1 were calculated by following formula [9].

$$P_m x_k,\% = 100.\left|b_{m,i} x_{m,i}\right| / \mathrm{C}_{\mathrm{total}} \quad (2)$$

$$MP_m x_k,\% = \frac{1}{N}\sum_{j=1}^{N}\left(100.\left|b_{m,i} x_{m,i}\right| / \mathrm{C}_{\mathrm{total}}\right) \quad \text{with } \mathrm{C}_{\mathrm{total}} = \sum_{i=1}^{k}\left|b_{m,k} x_{m,k}\right| \quad (3)$$

Where N of 18 is number of tripeptides in training set; and m of 4 is number of predictors in this model $QESAR_{linear}$.

3.2. Neural network model

The NN models were generated by using four descriptors appearing in linear model $QESAR_{linear}$ (1) as their inputs. One neuron, which encoded the antioxidant activity, constituted the output layer, and the hidden layer contained a variable number of neurons.

The non-linear model as a NN model $QESAR_{neural}$ was created by incorporating the neuro-fuzzy technique with genetic algorithm in INForm system [[6]]. This non-linear model type consists of three layers I(4)-HL(3)-O(1). The input layer I(4) involves four neurons SaaN, SdO, SdO_acnt and SsOH_acnt. The output layer O(1) is only neuron AC_{exp}. The hidden layer HL(3) includes three neurons. The quality of this non-linear model $QESAR_{neural}$ appeared by value $R^2_{fitness}$ of 98.2296.

3.3. Comparison of $QESAR_{linear}$ and $QESAR_{neural}$ models

Predictability of linear model $QESAR_{linear}$ and non-linear model $QESAR_{neural}$ was validated carefully by leave-one-out validation techniques. The predicted antioxidation values of 5 tripeptides in test set resulting from these models, as shown in Table 2.

Table 2. *Experimental AC_{exp} and predicted AC_{pred} antioxidant activities of 5 tripeptides.*

No	Tripeptide	AC_{exp}	linear model $QESAR_{linear}$		non-linear model $QESAR_{neural}$	
			AC_{pred}	ARE, %	AC_{pred}	ARE, %
1	HHN	0.3170	0.2491	21.4259	0.2856	9.9054
2	HHQ	0.3170	0.2255	28.8530	0.2570	18.9274
3	PWK	0.4066	0.6059	49.0205	0.5905	45.2287
4	RWQ	0.6061	0.7354	21.3278	0.5600	7.6060
5	YYC	0.6369	0.5317	16.5136	0.5180	18.6686
Value MARE, %				27.4282		20.0672

The predicted resulting from these models was judged by absolute value of the relative error ARE, % [9], [10], the medium absolute value of the relative error MARE, % [9] was used for assessing overall error of models QESAR.

The predicted values resulting from these models QSARs were judged by the absolute value of the relative error ARE, %:

$$ARE,\% = 100\left|(AC_{exp} - AC_{pred})/AC_{exp}\right| \quad (4)$$

The medium absolute values of the relative error MARE, % were used for assessing overall error for models QSARs:

$$\text{MARE},\% = \frac{100}{N}\left|\frac{(AC_{\text{exp}} - AC_{\text{pred}})}{AC_{\text{exp}}}\right| \qquad (5)$$

Where N of 5 is number of tripeptides in test set; AC_{exp} and AC_{pred} are experimental and predicted antioxidant values.

ANOVA one factor rating also pointed out that the antioxidation values resulting from linear model $QESAR_{linear}$ and non-linear model $QESAR_{neural}$ turn out to be not different ($F = 0.0494 < F_{0.05} = 5.3177$). However, model $QESAR_{neural}$ has less MARE, % value than model $QESAR_{linear}$.

4. Conclusion

This work has appeared successfully the construction of linear model $QESAR_{linear}$ and non-linear model $QESAR_{neural}$. The Genetic algorithm was used to select consistently the important descriptors from a set of molecular descriptors to establish the best-fitting model QESAR. The non-linear model $QESAR_{neural}$ turn out to be better predictable than linear model $QESAR_{linear}$. The above results obtained from this work can become a good research way and promise for prediction of antioxidant activity values for tripeptides.

*

NGHIÊN CỨU QESAR CỦA NHÓM TRIPEPTIDE NHƯ CÁC TÁC NHÂN CHỐNG OXI HÓA

Nông Thị Hồng Duyên(1) **– Phạm Văn Tất**(2)

(1) Trường Đại học Khoa học – Đại học Huế; (2) Trường Đại học Thủ Dầu Một

TÓM TẮT

Một cơ sở dữ liệu gồm 23 tripeptide được sử dụng để nghiên cứu các mối quan hệ định lượng giữa các tham số bề mặt thế tĩnh điện và hoạt tính chống oxi hóa QESAR. Các tham số cấu trúc quan trọng SaaNH_acnt, SsOH_acnt, SaaN, SaaN_acnt, SsssCH, SaaaC, SsNH3p, SdO, SdO_acnt được chọn để xây dựng các mô hình tuyến tính QESAR bằng giải thuật di truyền. Mô hình tuyến tính 4 biến số tốt nhất $QESAR_{linear}$ *bao gồm các tham số cấu trúc SaaN, SdO, SdO_acnt và SsOH_acnt được xây dựng. Chất lượng mô hình* $QESAR_{linear}$ *được thể hiện ở các giá trị thống kê* $R^2_{fitness} = 97,5660$, *sai số chuẩn ước tính SE = 0,0378, F-stat = 130,2731,* $R^2_{test} = 93,3851$. *Mô hình phi tuyến là mô hình mạng rơron* $QESAR_{neural}$ *cấu trúc I(4)-HL(3)-O(1) với* $R^2_{fitness} = 98,2296$ *đã được xây dựng bằng cách sử dụng các tham số cấu trúc trong mô hình* $QESAR_{linear}$. *Các hoạt tính chống oxi hóa của các tripeptide nhận được từ mô hình* $QESAR_{linear}$ *và* $QESAR_{neural}$ *cho thấy các giá trị MARE, % = 27,4282 và 20,0672 tương ứng.*

Từ khóa: *các mô hình QESAR, hồi qui bội,*
mạng thần kinh và các tripeptide chống oxi hóa

REFERENCES

[1] Li Yao-Wang, Li B., He J., Qian P, *J. Molecular Structure*, No. 998, P. 53–61, (2011).

[2] S. Mittermayr, M. Olajos, T. Chovan, G.K. Bonn, A. Guttman, *Trends in Analytical Chemistry*, Vol. 27, No. 5, (2008).

[3] K. Saito, J. Dong-hao, T. Ogawa, K. Muramoto, E. Hatakeyama, T. Yasuhara, and K. Nokihara, *J. Agric. Food Chem.*, No.51, 3668#3674, (2003).

[4] Zhang H. Z., Yang D. P. and Tang G. Y., Vol 11 (15/16), P. 749 – 754 (2006).

[5] HyperChem Release 8.05, Hypercube Inc., USA (2008).

[6] INForm v2.0, Intelligensys Ltd., UK (2000).

[7] QSARIS 1.1, Statistical Solutions Ltd., USA (2001).

[8] D. D. Steppan, J. Werner, P. R. Yeater, *Essential Regression and Experimental Design for Chemists and Engineers*, (2000).

[9] Pham Van Tat, *Development of Quantitative Structure-Activity Relationship and Quantitative Structure-Property Relationship*, Natural science and technology publisher, Hanoi, (2009).

[10] Pham Van Tat, Pham Thi Tra My, *Vietnamese Journal of Chemistry and Application*, P. 10-15, No. 4, (2010).