Hue University Journal of Science: Natural Science; ISSN 1859–1388

Vol. 127, No. 1A, 2018, P. 67–82; DOI: 10.26459/hueuni-jns.v127i1A.4791

IN SILICO MODEL QSPR FOR PREDICTION OF STABILITY CONSTANTS OF METAL-THIOSEMICARBAZONE COMPLEXES

Nguyen Minh Quang[1,3], Tran Xuan Mau[1], Pham Van Tat[2*], Tran Nguyen Minh An[3], Vo Thanh Cong[3]

[1]University of Sciences, Hue University, 77 Nguyen Hue St., Hue, Vietnam
[2]Faculty of Science and Technology, Hoa Sen University, 93 Cao Thang St., Ho Chi Minh City, Vietnam
[3]Faculty of Chemical Engineering, Industrial University of Ho Chi Minh City, 12 Nguyen Van Bao St., Ho Chi Minh City, Vietnam

Abstract. In the present work, the stability constants logβ_{11} and the concentration of metal ion and thiosemicarbazone in the solutions of their complex were determined by using *in silico* models. The 2D, 3D, physicochemical and quantum descriptors of the complexes were generated from the molecular geometric structure and semi-empirical quantum calculation PM7 and PM7/sparkle. The quantitative structure and property relationships (QSPRs) were constructed by using the ordinary linear regression (OLR) and artificial neural network (ANN). The best linear model QSPR$_{OLR}$ (with k of 6) involved descriptors k0, core-core repulsion, xp5, xch5, valence, and SHHBd. The quality of model QSPR$_{OLR}$ had the statistical values: R^2_{train} = 0.898, R^2_{adj} = 0.889, Q^2_{LOO} = 0.846, MSE = 1.136, and F_{stat} = 91.348. The neural network model QSPR$_{ANN}$ with architecture I(6)-HL(6)-O(1) had the statistical values: R^2_{train} = 0.9768, and Q^2_{LOO} = 0.8687. The predictability of QSPR models for the complexes of the test group turned out to be in good agreement with those from the experimental data in the literature.

Keywords: *in silico* models, stability constants logβ_{11}, QSPRs, ordinary linear regression, artificial neural network, thiosemicarbazone

1 Introduction

Thiosemicarbazone compounds and its metal complexes have many practical applications. Thiosemicarbazones are known as analytical reagents [1, 2] and have biological activities [3]. The complexes of thiosemicarbazones and metal ions have biological applications and great medicinal activities including antibacterial, antifungal, antimalarial, antitumor, and antiviral activity [4–6]. They are also used as a catalyst in chemical reactions [7].

For complexes, the stability constant is an important parameter. This is used to identify the complex stability in solutions. It is also a measure of the strength of the interaction between the ligand and the metal ions to form different complexes. In addition, the stability constant of

** Corresponding: vantat@gmail.com*

Submitted: 13–4–2018; Revised: 29–4–2018; Accepted: 3–5–2018

complexes is the basic factor to explain such phenomena as reaction mechanism and various properties of biological systems. We can calculate the equilibrium concentration of ingredients in a solution based on the stability constant. The changes of the complex structure in solutions can be forecast by using the initial concentration of the metal ion and the ligand. Recently, the stability constant of the complexes has been estimated by incorporating the UV/VIS spectrophotometric method and the computational techniques [8, 9]. Furthermore, the theoretical methods are also used for predicting the stability constant of complexes based on the relationships between the structural descriptors and the properties [9]. A few complex descriptors between the metal ions and thiosemicarbazone were determined by using quantum mechanics methods [12, 13, 20].

In recent years, computers have been becoming a helpful tool and an effective means of strong calculation in different areas of chemistry, such as inorganic chemistry, analytical chemistry, organic chemistry, physical chemistry, material simulation, and data mining [14–16]. The molecular design by means of a computer is also a tool to accelerate the discovery process for resulting knowledge of material properties. This is also a tendency to reduce the classical trial-and-error approach [17]. In this case, the development of molecular models, such as quantitative structure and property relationship (QSPR) and conformational search methodologies has also contributed greatly to the discovery and development of new molecules [18–20]. In this way, the multivariate analysis methods have been becoming a convenient and easy tool for supporting empirical and theoretical models. The multivariable linear relationships can be used to assess different characteristics of the systems.

In this work, we report the construction of the quantitative structure and property relationships using the structural descriptors and stability constant of complexes between the metal ions and thiosemicarbazone. The QSPR$_{OLR}$ and QSPR$_{ANN}$ models were successfully built based on the regression technique and neural network. The stability constant logβ_{11} of the complexes between the metal ions and thiosemicarbazone in the test set resulting from the QSPR models was validated and compared with those from experimental data in the literature.

2 Computational methods

2.1 Formation of complex

In an aqueous solution, the formation of a complex between a metal ion (M) and a thiosemicarbazone ligand (L) is, in fact, usually an addition reaction [15]. The general equilibrium equation is as follows

$$p\,\mathrm{M} + q\,\mathrm{L} \;\rightleftharpoons\; \mathrm{M}_p\mathrm{L}_q \tag{1}$$

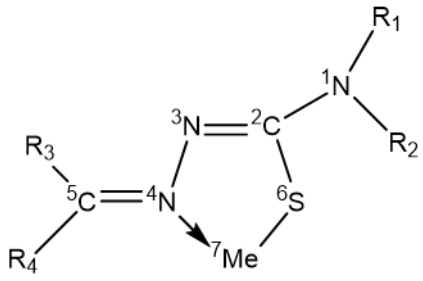
The overall or stability constant, given the symbol β, is the constant for the formation of the complex from the reagents. The stability constant for the formation of $\mathrm{M}_p\mathrm{L}_q$ is given by

$$\beta_{pq} = \frac{[\mathrm{M}_p\mathrm{L}_q]}{[\mathrm{M}]^p[\mathrm{L}]^q} \tag{2}$$

The stability constant β refers to the formation of the complex ML in one step with $p = 1$ and $q = 1$

$$\beta_{11} = \frac{[\mathrm{ML}]}{[\mathrm{M}][\mathrm{L}]} \tag{3}$$

a)



b)

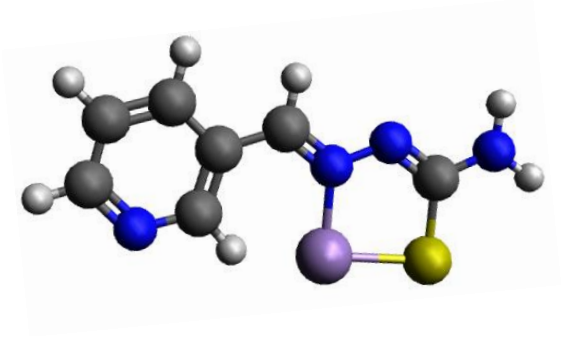


Figure 1. Structure of the complex between metal ions and thiosemicarbazone: a) General complex structure; b) Complex between Mn^{2+} and 3-formylpyridine thiosemicarbazone [21]

2.2 Data and computational details

The values $\log\beta_{11}$ of complexes between metal ions and the ligand thiosemicarbazone were taken from the literature [20–29] (Table 1).

The complexes of metal ions and ligand thiosemicarbazone were re-built and optimized by means of quantum mechanics on the MoPac 2016 system [30]. The quantum descriptors were calculated by using the semi-empirical quantum method with new version PM7 and PM7/sparkle for lanthanides [31]. The 2D and 3D topological descriptors were calculated by using the QSARIS system [10, 32]. The construction of $\mathrm{QSPR_{OLR}}$ models was performed using the back-elimination and forward regression technique on the Regress system [33] and MS-Excel [10, 14, 34]. The artificial neural network model $\mathrm{QSPR_{ANN}}$ was constructed using the

multilayer training technique on the Visual Gene Developer system [35]. The predictability of the QSPRs models was cross-validated by means of the leave-one-out method (LOO) using the statistic Q^2_{LOO}.

Table 1. Complexes of metal ions and thiosemicarbazone and stability constant [20–29]

Ligand				Metal ions	$\log\beta_{11}$
R_1	R_2	R_3	R_4		
H	H	H	$-C_5H_4N$	Ni(II)	5.630
H	H	H	$-C_5H_4N$	Mn(II)	4.320
H	H	H	$-C_5H_4N$	Co(II)	5.360
H	H	H	$-C_5H_4N$	Zn(II)	5.230
H	H	H	$-C_6H_4OH$	V(V)	5.322
H	H	H	$-C_4H_3O$	Co(II)	5.099
H	H	$-CH_3$	$-C_5H_4N$	La(III)	7.600
H	H	$-CH_3$	$-C_5H_4N$	Pr(III)	7.760
H	H	$-CH_3$	$-C_5H_4N$	Nd(III)	7.950
H	H	$-CH_3$	$-C_5H_4N$	Gd(III)	8.160
H	H	$-CH_3$	$-C_5H_4N$	Sm(III)	8.260
H	H	$-CH_3$	$-C_5H_4N$	Tb(III)	8.340
H	H	$-CH_3$	$-C_5H_4N$	Dy(III)	8.490
H	H	$-CH_3$	$-C_5H_4N$	Ho(III)	8.640
H	H	H	$-C_6H_5$	Ag(I)	15.500
H	H	H	$-C_5H_4N$	Ag(I)	14.000
H	H	H	$-C_6H_4OH$	Ag(I)	15.600
H	H	H	$-C_6H_5$	Cu(II)	17.700
H	H	H	$-C_5H_4N$	Cu(II)	20.400
H	H	$-CH_3$	$-C_2H_4NO$	Cu(II)	19.100
H	H	$-CH_3$	$-C_6H_4OH$	Mg(II)	3.300
H	H	$-CH_3$	$-C_6H_4OH$	Mg(II)	3.030
H	H	$-CH_3$	$-C_6H_4OH$	Mg(II)	2.920

Ligand				Metal ions	$\log\beta_{11}$
R_1	R_2	R_3	R_4		
H	H	$-CH_3$	$-C_6H_4OH$	Cd(II)	5.590
H	H	$-CH_3$	$-C_6H_4OH$	Cd(II)	4.830
H	H	$-CH_3$	$-C_6H_4OH$	Cd(II)	4.740
H	H	$-CH_3$	$-C_6H_4OH$	Pb(II)	5.740
H	H	$-CH_3$	$-C_6H_4OH$	Pb(II)	5.010
H	H	$-CH_3$	$-C_6H_4OH$	Pb(II)	4.900
H	H	H	$-C_6H_4NH_2$	Cu(II)	10.570
H	H	H	$-C_6H_4NH_2$	Ni(II)	12.710
H	H	H	$-C_6H_4NH_2$	Ni(II)	11.210
H	H	H	$-C_6H_4NH_2$	Co(II)	11.950
H	H	H	$-C_6H_4NH_2$	Co(II)	9.870
H	H	H	$-C_6H_4NH_2$	Mn(II)	12.140
H	H	H	$-C_6H_4NH_2$	Mn(II)	9.990
H	H	H	$-C_6H_4NH_2$	Zn(II)	11.320

2.3 Ordinary least square regression

The ordinary least square regression (OLR) was used to model and predict the values of one or more dependent quantitative or qualitative variables by means of a linear combination of one or more explanatory quantitative and/or qualitative variables. This technique did not face the constraints of ordinary least square regression (OLR) on the number of variables versus the number of observations.

The ordinary least square regression or ordinary linear regression is more commonly named linear regression [33, 34]. In this case, the regression model with *k* explanatory variables writes

$$Y = \beta_0 + \sum_{j=1}^{k} \beta_j \cdot X_j + \varepsilon \quad (4)$$

where Y is the dependent variable, β_0 is the intercept of the model, β_j is the coefficient of the j^{th} explanatory variable, X_j corresponds to the j^{th} explanatory variable (with j = 1 to k), and ε is the random error with mean 0 and variance σ^2.

In the case of k observations, the estimation of the predicted value of the dependent variable Y is given by expression (5) [36–40]

$$\hat{Y} = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j \cdot X_j \tag{5}$$

The OLR method corresponds to minimizing the sum of squared differences between the observed and predicted values. This minimization leads to the following estimators of the parameters of the model. The models were screened by using the values R^2_{train} and Q^2_{LOO} [10, 33–40]. These were assessed by the same formula (6)

$$R^2 = 1 - \frac{SS_E}{SS_T} = 1 - \frac{\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2} \tag{6}$$

where Y_i, $\hat{Y}_i$, and $\bar{Y}$ are the experimental, predicted and average value of the response, respectively; n is the total number of observations.

Adjusted R^2 (R^2_{adj}) is the adjusted determination coefficient for the model. The value of R^2_{adj} can be negative if the R^2 is close to zero. This coefficient is only calculated if the constant of the model has not been fixed by the user. R^2_{adj} is defined by

$$R^2_{adj} = 1 - \frac{MS_E}{MS_T} = 1 - \frac{n-1}{n-k-1} \cdot (1 - R^2) \tag{7}$$

R^2_{adj} is a correction to R^2, which takes into account the number of variables used in the model. The error mean square (MS_E) is defined by

$$MS_E = \frac{\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2}{n-k-1} \tag{8}$$

2.4. Artificial neural network

A neural network as a function of a set of derived inputs is called hidden nodes. The hidden nodes are nonlinear functions of the original inputs. The neural network can specify many layers of hidden nodes [41, 42].

The functions applied at the nodes of the hidden layers are called activation functions. The activation function is a transformation of a linear combination of the X variables. The function applied at the response is a linear combination of continuous responses, or a logistic transformation for nominal or ordinal responses [43, 44]. There are three transfer functions, namely sigmoid, hyperbolic tangent, and Gaussian transfer function.

The main advantage of the neural network is that it can efficiently model different response surfaces. Neural networks are very snappy models and tend to overfit data. When that happens, the forecast of the model is very good but predicts future observations poorly. The weakness of the neural network model is that the results are not easily explainable, since there are intermediate layers rather than a direct path from the X variables to the Y variables, as in the case of regular regression [45, 46]. To alleviate overtraining, the neural network is validated by use of an independent data set to evaluate the predictive ability of the model [41].

Validation is a process of using a part of the data set to estimate the model parameters and using the other part to assess the predictability of the neural network. The first part is the training set used to estimate the model parameters. The second part is the validation set used to validate the predictability of the model. The test set is the final, independent assessment of the model predictability [42].

In this work, we used a typical feed-forward neural network, which was trained by using an error back-propagation learning algorithm. This neural network style propagates information in the feed-forward direction using equation (9) [41, 42]

$$b_j = f\left(\sum_{i=0}^{N} w_{i,j} \cdot a_i - T_j\right) \tag{9}$$

where a_i is the input factor, b_j is the output factor, w_{ij} is the weight factor between two nodes, T_j is the internal threshold, and f is the transfer function.

There exist many transfer functions that are used in neural networks such as hyperbolic tangent, Gaussian, sigmoid… In this study, we used the hyperbolic tangent function. The hyperbolic tangent learning algorithm is based on a generalized delta rule accelerated by a momentum term. To increase the efficiency of the neural network, both the weight factors and the internal threshold values were adjusted using equations (10) and (11) [41, 42]

$$W_{i,j}^{new} = w_{i,j}^{old} + \eta \cdot \sum_{k} \delta_{k,j} \cdot O_{k,i} + \alpha \cdot \Delta W_{i,j}^{old} \tag{10}$$

$$T_j^{new} = T_j^{old} + \eta \cdot \sum_k \delta_{k,j} + \alpha \cdot \Delta T_j^{old} \quad (11)$$

where η is the learning rate; α is the momentum coefficient; ΔW is the previous weight factor change; ΔT is the previous threshold value change; O is the output – the gradient-descent correction term; and k stands for the pattern. The performance of the trained network was verified by determining the error between the predicted value and the real value. All the data of the patterns were normalized to be less than 1 before training the neural network; the initial weight factors were randomly generated from –0.2 to 0.2, and the initial internal threshold values were set to zero.

3 Results and discussion

3.1 Constructing models QSPR$_{OLR}$

The QSPR$_{OLR}$ model was constructed from the database of complexes between metal ions and the ligands including the 2D and 3D molecular descriptors, and the quantum parameters. The general complex structure is shown in Fig. 1a and 1b, and the stability constant logβ_{11} is given in Table 1.

The linear regression model was constructed based on the training set and the test set, in which the portion of the test set is 20 %. The model quality was evaluated by means of statistical values R^2_{train}, R^2_{adj}, Q^2_{LOO} and F_{stat} (Fischer's value). The QSPR$_{OLR}$ models and the statistical values are shown in Table 2.

The best linear models QSPR$_{OLR}$ were selected using the back-elimination and forward method with the critical value α = 0.05; the important descriptors selected were based on the changes of the statistical parameters: standard error – SE, R^2_{train}, R^2_{adj}, Q^2_{LOO}, and F_{stat}. The number of descriptors k was selected in range 2 to 10. The change of the amount of structural parameter leads to the change of the values SE, R^2_{train} and Q^2_{LOO} (Figure 2a).

Table 2. Selected model QSPR$_{OLR}$ (k of 2 to 10) and statistical values

k	Variables	SE	R^2_{train}	R^2_{adj}	Q^2_{LOO}	F_{stat}
2	x_1/x_2	2.136	0.617	0.606	0.550	53.234
3	$x_1/x_2/x_3$	1.649	0.775	0.765	0.705	74.789
4	$x_1/x_2/x_3/x_4$	1.504	0.816	0.806	0.755	71.012

k	Variables	SE	R^2_{train}	R^2_{adj}	Q^2_{LOO}	F_{stat}
5	$x_1/x_2/x_3/x_4/x_5$	**1.347**	**0.855**	**0.843**	**0.799**	**74.173**
6	$x_1/x_2/x_3/x_4/x_5/x_6$	**1.136**	**0.898**	**0.889**	**0.846**	**91.348**
7	$x_1/x_2/x_3/x_4/x_5/x_6/x_7$	**1.024**	**0.919**	**0.909**	**0.786**	**98.462**
8	$x_1/x_2/x_3/x_4/x_5/x_6/x_7/x_8$	0.925	0.935	0.926	0.829	107.373
9	$x_1/x_2/x_3/x_4/x_5/x_6/x_7/x_8/x_9$	0.871	0.943	0.934	0.850	108.659
10	$x_1/x_2/x_3/x_4/x_5/x_6/x_7/x_8/x_9/x_{10}$	0.802	0.953	0.944	0.862	116.588

Notation of molecular descriptors			
k0	x_1	SHHBd	x_6
core-core repulsion	x_2	xp4	x_7
xp5	x_3	HOMO	x_8
xch5	x_4	LUMO	x_9
valence	x_5	xvc3	x_{10}

The average contribution percentage, $MPx_{k,i}$, is the percentage of each independent variable in the selected models QSPRs (with i of 1 to k), is determined according to formula (12) by the contribution for C_{total} value [10, 32]

$$MPx_{k,i}, \% = \frac{1}{N}\sum_{m=1}^{N}\frac{100.\left|b_{k,i}.x_{m,i}\right|}{\sum_{j=1}^{k}\left|b_{k,j}.x_{m,j}\right|} = \frac{1}{N}\sum_{m=1}^{N}\frac{100.\left|b_{k,i}.x_{m,i}\right|}{C_{total}} \quad (12)$$

where N is the number of substances (N = 69); m is the number of substances used to calculate $Px_{k,i}$ value; $b_{k,i}$ are the parameters of the model. The important contribution of molecular descriptors in each complex is arranged in the order based on $GMPx_i$ values ($GMPx_i$ is the average value of $MPx_{k,i}$): k0 > xp5 > core-core repulsion > xch5 > valence > SHHBd (Table 3).

In the surveyed models, the QSPR$_{OLR}$ model (with k = 6) has the best Q^2_{LOO} value although it changes when k increases. Thus, this QSPR$_{OLR}$ model is the best match in all the models. The quality of the QSPR$_{OLR}$ model is shown with the R^2_{train} value of 0.898; the standard error SE of 1.136; the F_{stat} value of 91.348 and the Q^2_{LOO} value of 0.846. The linear regression equation of the QSPR$_{OLR}$ model is as follows

$$\log\beta_{11} = 66.01 - 5.861 \cdot x_1 + 0.00137 \cdot x_2 + 7.246 \cdot x_3 - 39.35 \cdot x_4 - 1.745 \cdot x_5 + 2.07 \cdot x_6 \quad (13)$$

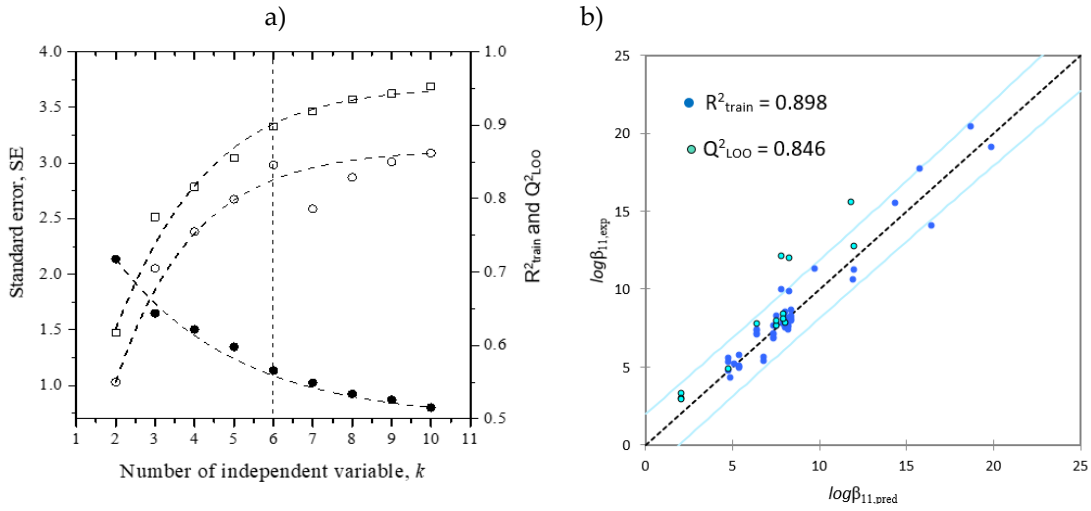


Figure 2. a) Change trend line of values *SE*, R^2_{train} and Q^2_{LOO} according to *k* descriptors; b) Correlation of experimental versus predicted values logβ11 of the test compounds using the QSPROLR model (with *k* = 6)

The importance of each descriptor in QSPROLR model (13) or rather the contribution of the descriptors to the stability constant of the complexes is assessed according to the $GMPx_i$, % values. The k0 parameter (x_1) with the $GMPx_1$ value of 55.5680 influences the stability constant of complexes most. The k0 parameter is called Kappa zero index, i.e., Shannon information index based on atom classes. Next, the xp5 parameter is called Chi path 5, the simple 5th-order path Chi index (x_3) with the $GMPx_3$ value of 14.6137. The last parameter that strongly affects the stability constant is core-core repulsion (x_2) with the $GMPx_2$ value of 10.7750.

Table 3. Statistical values and variables, and $MPx_{k,i}$ and $GMPx_i$ contribution in models QSPROLR with *k* of 5 to 7

Statistical values and variables	QSPROLR			$MPx_{k,i}$, %			$GMPx_i$, %
	k = 5	*k* = 6	*k* = 7	*k* = 5	*k* = 6	*k* = 7	
R^2_{train}	0.855	0.898	0.919	–	–	–	–
R^2_{adj}	0.843	0.889	0.909	–	–	–	–
Q^2_{LOO}	0.799	0.846	0.786	–	–	–	–
SE	1.347	1.136	1.024	–	–	–	–
Constant	51.11	66.01	68.9	–	–	–	–

Statistical values and variables	QSPR$_{OLR}$			$MPx_{k,i}$, %			$GMPx_i$, %
	k = 5	k = 6	k = 7	k = 5	k = 6	k = 7	
x_1	–5.08	–5.861	–8.094	59.4123	63.0614	44.2303	55.5680
x_2	0.00263	0.00137	0.00095	19.6377	9.4153	3.1953	10.7750
x_3	5.37	7.246	–9.05	14.5019	17.9917	11.3473	14.6137
x_4	–33.93	–39.35	–87.73	3.3895	3.6193	4.0628	3.6905
x_5	–1.613	–1.745	0.996	3.0585	3.0427	0.8777	2.3263
x_6	–	2.07	4.494	–	2.8696	3.3202	2.0633
x_7	–	–	20.76	–	–	32.9662	10.9887

As such, the training data set is good, and the application of QSPR$_{OLR}$ model is statistically very meaningful. The cross-validated technique shows that the QSPR$_{OLR}$ model can be used to predict the logβ_{11} values. The statistical values were used to check the meaning of the coefficients in the QSPR$_{OLR}$ models, as given in Table 3.

3.2 Constructing models QSPR$_{ANN}$

In addition to model QSPR$_{OLR}$, the QSPR$_{ANN}$ model was also developed with the neural network technique on the Visual Gene Developer system [35] upon the molecular descriptors of model QSPR$_{OLR}$. The architecture of the neural network comprising three layers is I(6)-HL(6)-O(1) (Fig. 3a); the input layer I(6) includes 6 neurons (k0, core-core repulsion, xp5, xch5, valence, and SHHBd); the output layer O(1) includes 1 neuron, that is, logβ_{11}; the hidden layer includes 6 neurons.

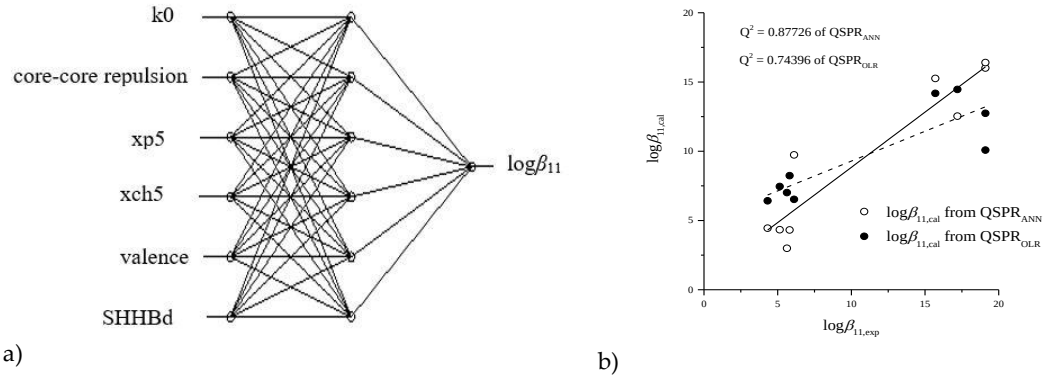


Fig. 3. a) Architecture of neural network I(6)-HL(6)-O(1);
b) Correlation of experimental vs. predicted values of test set from QSPR$_{OLR}$ and QSPR$_{ANN}$ model

The error back-propagation algorithm was used to train the network. The hyperbolic tangent transfer function was set on each node of the layers; the training network parameters included the learning rate of 0.01, the momentum coefficient of 0.1, and the sum of error of 0.000016 with 1,000,000 loops. The results of the training process are given in Table 4.

Table 4. Training quality of neural network $QSPR_{ANN}$ I(6)-HL(6)-O(1)

Data set	Regression coefficient	Slope	y-intercept
Training	0.9768	0.9770	0.00253
Validation	0.8687	1.4432	–0.1397

As can be seen from Fig. 3b, the neural model $QSPR_{ANN}$ based on the architecture of neural network I(6)-HL(6)-O(1) adapted better than the $QSPR_{OLR}$ model. In fact, the neural model $QSPR_{ANN}$ exhibited a better fit and correlation between the predicted values and the experimental values than the $QSPR_{OLR}$ model with Q^2 of 0.8773 and 0.7440, respectively.

3.3 Predictability of QSPR models

The predictability of the $QSPR_{OLR}$ and $QSPR_{ANN}$ model was carefully evaluated by means of the phasing-each-case technique. The predicted results received for 9 randomly chosen substances with the experimental values [27–29, 45, 46] are presented in Table 5.

The average absolute values of the relative error *MARE* used to assess the overall error of the QSPR models were calculated according to formula (14)

$$MARE, \% = \frac{\sum_{i=1}^{n} ARE_i, \%}{n} \quad (14)$$

where $ARE, \% = \frac{|\log \beta_{11,exp} - \log \beta_{11,cal}|}{\log \beta_{11,exp}}$, n is the number of test substances; and $\beta_{11,exp}$ and $\beta_{11,cal}$ are the experimental and calculated stability constants.

Table 5. Stability constant of 9 test substances resulting from $QSPR_{OLR}$ model and model $QSPR_{ANN}$

n	Ligand				Metal Ions	$\log\beta_{11,exp}$	The linear model $QSPR_{OLR}$		The neural model $QSPR_{ANN}$	
	R_1	R_2	R_3	R_4			$\log\beta_{11,cal}$	*ARE*, %	$\log\beta_{11,cal}$	*ARE*, %
1	H	CH_3	CH_3	$–C_5H_4N$	Cu(II)	6.114	6.5280	6.7721	9.7353	59.2289
2	H	H	H	$–C_6H_4BrO$	Cu(II)	5.633	7.0116	24.4729	2.9897	46.9245

n	Ligand				Metal Ions	$\log\beta_{11,exp}$	The linear model QSPR$_{OLR}$		The neural model QSPR$_{ANN}$	
	R_1	R_2	R_3	R_4			$\log\beta_{11,cal}$	*ARE*, %	$\log\beta_{11,cal}$	*ARE*, %
3	H	H	H	$-C_6H_4OH$	Ag(I)	15.700	14.1818	9.6700	15.2536	2.8435
4	H	H	H	$-C_5H_6N$	Cu(II)	19.100	10.0826	47.2116	16.0061	16.1984
5	H	H	H	$-C_6H_4OH$	Cu(II)	19.100	12.7401	33.2981	16.4007	14.1325
6	H	H	H	$-C_6H_4OH$	Cu(II)	17.200	14.4635	15.9098	12.5361	27.1158
7	H	H	CH_3	$-C_6H_4OH$	Mn(II)	4.320	6.4178	48.5592	4.4445	2.8828
8	H	H	CH_3	$-C_6H_4OH$	Ni(II)	5.140	7.4541	45.0205	4.3317	15.7250
9	H	H	CH_3	$-C_6H_4OH$	Cu(II)	5.810	8.2349	41.7361	4.3133	25.7603
							MARE, %	30.295	*MARE*, %	23.424

One-way ANOVA was used to evaluate the difference between the experimental and predicted $\log\beta_{11}$ values from the QSPR$_{OLR}$ and QSPR$_{ANN}$ model. Accordingly, the discrepancies between the experimental and calculated values of stability constants $\log\beta_{11}$ resulting from the QSPR$_{OLR}$ model and the QSPR$_{ANN}$ model I(6)-HL(6)-O(1) were insignificant ($F = 0.1728 < F_{0.05} = 3.4028$). Therefore, the predictability of both QSPR models turns out to be in good agreement with the experimental data.

The *MARE* values of models QSPR$_{OLR}$ and QSPR$_{ANN}$ I(6)-HL(6)-O(1) were 30.295 % and 23.424 %, respectively (Table 5), indicating that model QSPR$_{ANN}$ showed higher predictability than model QSPR$_{OLR}$, and the $\log\beta_{11}$ values resulting from model QSPR$_{ANN}$ were closer to the experimental values.

4 Conclusion

This work successfully built the quantitative structure and property relationships incorporating ordinary linear regression and artificial neural network. The QSPR models were constructed by using the data set of structural descriptors resulting from the semi-empirical quantum calculation and molecular mechanics. The models were cross-validated carefully using the leave-one-out method upon statistical values R^2_{train}, Q^2_{LOO}, *MARE*, %, and the single factor ANOVA method. The QSPR$_{ANN}$ model I(6)-HL(6)-O(1) turned out to be satisfactory for actual applicability. The results from this work could serve for designing new

thiosemicarbazone derivatives that are helpful in the fields of analytical chemistry, pharmacy, and environment.

References

1. Singh R. B., Garg B. S., Singh. R. P. (1978), Analytical applications of thiosemicarbazones and semicarbazones: A review, *Talanta,* 25(11–12), 619–632.
2. Patel B. H., Shah J. R., Patel R. P. (1976), Stability constants of complexes of 2-hydroxy-5-methylacetophenone-thiosemicarbazone with Cu(II), Ni(II), Co(II), Zn(II) and Mn(II), J. *Ind. Chem. Soc,* 53, 9–10
3. E. B. Seena, R. Bessy, M. R. Prathapachandra Kurup, I. E. Suresh, A crystallographic study of 2-hydroxyacetophenone *N (4)* cyclohexyl thiosemicarbazone, *J. Chem. Crystallogr,* 36, 189 (2006)
4. Ezhilarasi et al., Synthesis Characterization and Application of Salicylaldehyde Thiosemicarbazone and Its Metal Complexes, *Int. J. Res. Chem. Environ.,* 2(4), 130–148 (2012)
5. Giorgio Pelosi, Thiosemicarbazone Metal Complexes: From Structure to Activity, J. *Open Crystallogr.*, 3, 16–28 (2010)
6. A. Nagajothi, A. Kiruthika, S. Chitra, K. Parameswari, Fe(III) Complexes with Schiff base Ligands: Synthesis, Characterization, Antimicrobial Studies, *Res. J. chem. Sci*., 3(2), 35–43 (2013)
7. M. Rajendran, A. Panneerselvam, V. Periasamy, M. J. Grzegorz, Palladium(II) pyridoxal thiosemicarbazone complexes as efficient and recyclable catalyst for the synthesis of propargylamines by a three-component coupling reactions in ionic liquids, *Polyhedron*, 119, 300–306 (2016)
8. B. S. Chandra, M. Tyagi, Spectroscopic and biochemical studies of chromium (III) and manganese (II) complexes with *p*-vanillin containing thiosemicarbazone and semicarbazone ligands, *J. Indian Chem. Soc*., 85, 42–47 (2008)
9. R. Chaudhary and Shelly, Synthesis, Spectral and Pharmacological Study of Cu (II), Ni (II) and Co (II) Coordination Complexes, *Res. J. chem. Sci*., 1(5), 1–5 (2011)
10. Pham Van Tat, Development of QSAR and QSPR, Publisher of Natural sciences and Technique, Ha Noi (2009)
11. M. Ante and N. Raos, Estimation of Stability Constants of Mixed Copper(II) Chelates Using Valence Connectivity Index of the 3rd Order Derived from Two Molecular Graph Representations, *Acta. Chim. Slov*., 56, 373–378 (2009)
12. M. Ante and N. Raos, Estimation of Stability Constants of Copper(II) Bis-chelates by the Overlapping Spheres Method, *Croatica Chemica Acta*., 79(2), 281–290 (2006)
13. S. Nikolic and N. Raos, Estimation of Stablity Constantsof Mixed Amino Acid Complexes with Copper(II) from Topological Indices, *Croatica Chemica Acta*, 74(3), 621–631 (2001)
14. E. J. Billo, Excel For Scientists And Engineers: Numerical Methods, John Wiley and Sons, Inc., Hoboken, NJ, USA (2007)
15. D. Harvey, Modern analytical Chemistry, Mc.Graw Hill, Boston, Toronto (2000)
16. K. Roy, S. Kar, R.N. Das, Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment, Academic Press, Amsterdam (2015)

17. A. Speck-Planche, V. V. Kleandrova, L. Feng, M. Natália, D. S. Cordeiro, Rational drug design for anti-cancer chemotherapy: multi-target QSAR models for the in silico discovery of anti-colorectal cancer agents, *Bioorg Med Chem*, 20(15), 4848–4855 (2012)
18. A. Speck-Planche, V. V. Kleandrova, L. Feng, M. Natália, D. S. Cordeiro, Chemoinformatics in anti-cancer chemotherapy: Multi-target QSAR model for the in silico discovery of anti-breast cancer agents, *Eur J Pharm Sci.*, 47(1), 273–27 (2012)
19. R. Sabet, M. mohammadpour, A. Sadeghi, A. Fassihi, QSAR study of isatin analogues as in vitro anti-cancer agents, *Eur J Med Chem.*, 45(3), 1113–1118 (2010)
20. N. S. R. Reddy, D. V. Reddy, Spectrophotometric determination of vanaditun(V) with salicylaldehyde thiosemicarbazone, *J. Indian Inst. Sci.*, 64(B), 133–136 (1982)
21. D. N. Kenie, A. Satyanarayana, Protolitic Equilibria and Stability Constants of Mn (II) and Ni (II) Complexes of 3-formylpyridine Thiosemicarbazone in Sodium Dodecyl Sulphate (SDS)-Water Mixture, *J. Technol Arts Sci Res.*, 4(1), 74–79 (2015)
22. D. N. Kenie, A. Satyanarayana, Solution Equilibrium Study of the Complexation of Co(II) and Zn(II) with Nicotinaldehyde Thiosemicarbazone, *J. Technol Arts Sci Res.*, 4(3), 145–149 (2015)
23. V. Veeranna, V. S. Rao, V. V. Laxmi, T.R. Varalakshmi, Simultaneous Second Order Derivative Spectrophotometric Determination of Cadmium and Cobalt using Furfuraldehyde Thiosemicarbazone (FFTSC), *J. Pharm. and Tech Res.*, 6(5) (2013)
24. B. S. Garg, S. R. Singh, R. B. Basnet, R. P. Singh, Potentiometric Studies On The Complexation Equilibria Between La(III), Pr(III), Nd(III), Gd(III), Sm(III), Tb(III), Dy(III, Ho(III) And 2-acetylpyridinethiosemicarbazone (2-APT), *Polyhehon*, 7 (2), 147–150 (1988)
25. A. T. A. El-Karim, A. Ahmed, El-Sherif, Potentiometric, equilibrium studies and thermodynamics of novel thiosemicarbazones and their bivalent transition metal(II) complexes, *J. Mol Liq.*, 219, 914–922 (2016)
26. M. A. Jiménez, M. D. Luque De Castro, M. Valcárcel, Potentiometric Study of Silver(l)-Thiosemicarbazonates, J. *Microchemical*, 25, 301–308 (1980)
27. M. A. Jiménez, M. D. Luque De Castro, M. Valcárcel, Titration of Thiosemicarbazones with Cu(ll) and Vice Versa by Use of a Copper Selective Electrode in Acetone-Water Mixture: Determination of the Conditional Formation Constants of the Cupric Thiosemicarbazonates, *J. Microchemical,* 32, 166–173 (1985)
28. T. Atalay, E. Ozkan, Thermodynamic studies of some complexes of 4'-morpholino-acetophenone thiosemicarbazone, *Thermochimica Acta*, 237, 369–374 (1994)
29. B. S. Garg, S. Ghosh, V. K. Jain, P. K. Singh, Evaluation of thermodynamic parameters of bivalent metal complexes of 2-hydroxyacetophenone thiosemicarbazone (2-HATS), *Thermochimica Acta*, 157, 365–368 (1990)
30. MOPAC2016, Version: 17.240W, J. J. P. Stewart, Stewart Computational Chemistry, USA (2002)
31. James J. P. Stewart, Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters, *J Mol Model*., 19, 1–32 (2013)
32. QSARIS 1.1, Statistical Solutions Ltd., USA (2001)
33. D. D. Steppan, J. Werner, P. R. Yeater, Essential Regression and Experimental Design for Chemists and Engineers, Germany (1998)

34. E. J. Billo, Excel for chemists, Wiley-VCH, Weinheim (1997)
35. S. K. Jung, K. McDonald, Visual Gene Developer: a fully programmable bioinformatics software for synthetic gene optimization, *BMC Bioinformatics,* 12(1): 340 (2011)
36. T. Amemiya, Selection of regressors, *International Economic Review*, 21, 331–354 (1980)
37. A. P. Dempster, Elements of Continuous Multivariate Analysis, Addison-Wesley, Reading, MA, (1969)
38. L. Eriksson, E. Johansson, N. Kettaneh-Wold N, S. Kettaneh-Wold, Multi- and Megavariate Data Analysis: Principles and Applications, *Journal of Chemometrics*, 16(5), 261–262 (2001)
39. S. Kotz, N. L. Johnson, Breakthroughs in Statistics, Vol. 1. Foundations and Basic Theory, 610–624, New York: Springer (1992)
40. G. Schwarz, Estimating the dimension of a model, *Annals of Statistics,* 6, 461–464 (1978)
41. J. Gasteiger, J. Zupan, Neural Networks in Chemistry, *Chiw. Inr. Ed. EngI.*, 32, 503–521 (1993)
42. R. Rojas, Neural Networks, Springer-Verlag, Berlin (1996)
43. M. N. M. Milunovic, Éva A. Enyedy, Nóra V. Nagy, T. Kiss, R. Trondl, M. A. Jakupec, B. K. Keppler, R. Krachler, G. Novitchi, V. B. Arion, L- and D-Proline Thiosemicarbazone Conjugates: Coordination Behavior in Solution and the Effect of Copper(II) Coordination on Their Antiproliferative Activity, *J. Inorg. Chem.*, 51, 9309–9321 (2012)
44. M. Hymavathi, N. Devanna, C. Viswanatha, A sensitive and selective chromogenic reagent using 2-hydroxy 3,5-dimethoxy benzaldehyde thiosemicarbazone (HDMBTSC) for direct and derivative spectrophotometric determination of Molybdenum (VI), *J. Inter. Math Phy Sci. Res.,* 2(1), 43–48 (2014)
45. K. V. Reddy, D. N. Reddy, S. V. Babu and K. H. Reddy, Spectrophometric determination of copper (II) in Biological samples by using 2–acetylpyridine 4–methyl–3-thiosemicarbazone (APMT), *Der Pharmacia Sinica*, 2(4), 176–183 (2011)
46. G. Ramanjaneyulu, P. R. Reddy, V. K. Reddy, T. S. Reddy, Direct and Derivative Spectrophotometric Determination of Copper(II) with 5-Bromosalicylaldehyde Thiosemicarbazone, *J. Open Anal. Chem .*, 2, 78–82 (2008)