Vietnam Journal of Catalysis and Adsorption
Tạp chí xúc tác và hấp phụ Việt Nam

https://chemeng.hust.edu.vn/jca/

Development of new metal-thiosemicarbazone complexes using visual screening methods and *in silico* models

Nguyen Minh Quang[1,*], Tran Nguyen Minh An[1], Pham Van Tat[2]

[1]*Faculty of Chemical Engineering, Industrial University of Ho Chi Minh City, Ho Chi Minh City, VIETNAM*
[2]*Department of Pharmacy, Faculty of Health Science, Hoa Sen University, Ho Chi Minh City, VIETNAM*
**Email: nguyenminhquang@iuh.edu.vn*

ARTICLE INFO

Received: 15/5/2021
Accepted: 15/7/2021
Published: 15/10/2021

Keywords:

ANN, MLR, QSPR, stability constants logβ_{11}, metal-thiosemicarbazone complex.

ABSTRACT

The stability constants (logβ_{11}) of forty-two new metal-thiosemicarbazone complexes were predicted based on the results of the quantitative structure-property relationship (QSPR). The QSPR models were developed from 88 logβ_{11} values of experimental complexes by using the multivariate linear regression ($QSPR_{MLR}$) and artificial neural network ($QSPR_{ANN}$). Four descriptors such as xch9, xv0, core-core repulsion and cosmo area were found out in the best of the linear model $QSPR_{MLR}$ which was harshly evaluated by the statistical values: R^2_{train} = 0.864, Q^2_{LOO} = 0.840, SE = 0.711, F_{stat} = 131,355 and PRESS = 49.31. Furthermore, the artificial neural network model $QSPR_{ANN}$ with architecture I(4)-HL(5)-O(1) was discovered with the same variables of the $QSPR_{MLR}$ model that the statistical results were extremely impressive as R^2_{train} = 0.970, Q^2_{CV} = 0.984 and Q^2_{test} = 0.974. Also, both of the QSPR models were externally validated on the data set of 18 logβ11 values of independently experimental complexes. As a consequence, the results from the QSPR models could be used to calculate the stability constants of other new metal-thiosemicarbazones.

Introduction

The quantitative structure and property relationship (QSPR) method is known as the popular *in silico* method and it is also used widely in many fields for predicting properties of chemical compounds based on the relationships between the structural characteristics and the properties [1]. The QSPR modeling techniques are developed from a quantitative structure and activity relationship (QSAR) approach in which the properties of the model are replaced by activity and it was first introduced by Crum Brown and Fraser in 1868 [1]. In the 1940s, the appearance of chemical graph theory and the publications of Wiener and Platt's research helped the development of QSPR modeling [1]. According to statistics of OECD, the number of published works related to QSPR models was about 11,000 projects [1]. Nowadays, the QSPR method is widely used and is seen as an effective method for finding new compounds. The QSAR/QSPR model should meet the requirements of the OECD principles [2] and the model acceptance criteria of Golbraikh and Tropsha's [1]. The basic equation of the QSPR method can be expressed mathematically as follows [1]:

$$\text{Response (property)} = f(\text{descriptors}) \quad (1)$$

https://doi.org/10.51316/jca.2021.096

Normally, there are two popular methods to establish QSPR models, they are linear regression (MLR, PLS, PCR) and machine learning methods (SVR, ANN) [1]. In this study, we use two approaches to build the QSPR models which are MLR and ANN. In addition, the diverse structure and easy complexation with many metal ions of thiosemicarbazone derivatives led to its wide applications in many fields [3]. This is the reason why thiosemicarbazone derivatives and their complexes are popularly studied in practice. In this work, we modeled the dataset of the logarithm of stability constants ($\log\beta_{11}$) of the complexes (M:L) between thiosemicarbazone ligands with some metal ions (M = Mo^{6+}, Cu^{2+}, Co^{2+}, Ni^{2+}, Mn^{2+}, Mg^{2+}, Zn^{2+}, Cd^{2+}, Pb^{2+}, Cd^{2+}) in aqueous solution. The $\log\beta_{11}$ values were mined from an experimental published database (table 1). The QSPR models were formed by using multiple linear regression ($QSPR_{MLR}$) and the artificial neural network ($QSPR_{ANN}$). Besides, a new series of thiosemicarbazone ligands and complexes were designed and calculated the stability constant by the results of the developed QSPR models.

Experimental

Data selection

This study selects the complex of the ML type that formed between a metal ion (M) and a thiosemicarbazone ligand (L). The structure of the selected complexes is shown in fig 1.

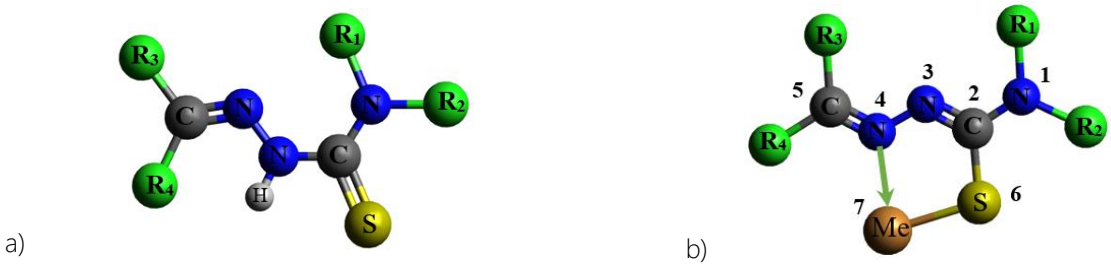


Figure 1: The structure of the thiosemicarbazone (a) and the metal-thiosemicarbazone complex (b)

The data mining is the first step in the QSPR modeling research. Firstly, a big data was mined from prestigious data spring, then the k-means method are approached to divide it into several data clusters [1]. In this study, a data set including the 88 $\log\beta_{11}$ values of the complexes between metal ions and the ligand thiosemicarbazone was used to build QSPR modeling in table 1.

Table 1: The 88 stability constants of complexes (n) in experimental dataset with minimal ($\log\beta_{11,min}$) and maximal ($\log\beta_{11,max}$) values

No	Thiosemicarbazone ligand				Metal ions	Number of complexes, n	$\log\beta_{11,min}$	$\log\beta_{11,max}$	Ref.
	R_1	R_2	R_3	R_4					
1	H	H	H	$-C_6H_2(OCH_3)_2OH$	Mo^{6+}	1	6.3365	6.3365	[4]
2	H	H	H	$-C_6H_2(OCH_3)_2OH$	Cu^{2+}	1	6.2355	6.2355	[5]
3	H	$-C_6H_5$	$-CH_3$	$-CCH_3$=N-OH	Cu^{2+}	2	7.4183	7.7559	[6]
4	H	H	$-CH_3$	$-CH=N-NHC_6H_5$	Co^{2+}	4	9.9000	10.220	[7,8]
5	H	H	$-CH_3$	$-CH=N-NHC_6H_5$	Ni^{2+}	4	10.790	11.080	[7,8]
6	H	H	$-CH_3$	$-CH=N-NHC_6H_5$	Mn^{2+}	3	9.600	9.870	[8]
7	H	$-CH_3$	$-CH_3$	$-CH=N-NHC_6H_5$	Cu^{2+}	3	11.980	12.300	[8]
8	H	$-CH_3$	$-CH_3$	$-CH=N-NHC_6H_5$	Ni^{2+}	3	10.910	11.210	[8]
9	H	H	H	$-C_{10}H_6OH$	Mg^{2+}	4	3.250	3.400	[9]
10	H	H	H	$-C_{10}H_6OH$	Mn^{2+}	1	5.670	5.670	[9]
11	H	H	H	$-C_{10}H_6OH$	Cd^{2+}	2	6.470	6.560	[9]
12	H	H	H	$-C_{10}H_6OH$	Pb^{2+}	2	6.570	6.680	[9]
13	H	H	H	$-C_{10}H_6OH$	Zn^{2+}	1	7.170	7.170	[9]
14	H	H	H	$-C_{10}H_6OH$	Co^{2+}	4	7.890	9.000	[9]
15	H	H	H	$-C_{10}H_6OH$	Ni^{2+}	4	8.370	9.600	[9]
16	H	H	H	$-C_{10}H_6OH$	Cu^{2+}	4	8.680	9.780	[9]

https://doi.org/10.51316/jca.2021.096

17	H	H	-	-C_9H_8NO	Ni^{2+}	8	7.709	8.500	[10]
18	H	H	-	-C_9H_8NO	Pb^{2+}	7	7.307	8.109	[10]
19	H	H	-	-C_9H_8NO	Co^{2+}	8	7.251	8.340	[10]
20	H	H	-	-C_9H_8NO	Zn^{2+}	8	7.039	8.160	[10]
21	H	H	-	-C_9H_8NO	Cd^{2+}	6	6.611	7.599	[10]
22	H	H	-	-C_9H_8NO	Mn^{2+}	8	5.439	6.230	[10]

Descriptors

Molecular descriptors are known as the variables in the equations of the QSPR models and they consist of physicochemical, topological indicator and quantum parameters related chemical structures. The descriptors of the metal-thiosemicarbazone complexes were generated using QSARIS package [11] and calculated by using the semi-empirical quantum method with new version PM7 and PM7/sparkle on the MoPac2016 system [12].

QSPR modeling methods

MLR method

In the multivariate linear regression (MLR) modeling methods, the equation of the $QSPR_{MLR}$ model is described according to the following equation [1,13]:

$$Y = b_0 + \sum_{j=1}^{k} b_j X_j \quad (2)$$

where b_0, is the intercept of the model, b_j is the regression coefficients and k is number of explanatory variables in the equation.

The MLR analysis is performed by stepwise regression technique on the Regress system [13] and MS-EXCEL [14] and the cross-validation (CV) method for the QSPR models is carried out by the leave-one-out process (LOO) [1,13].

ANN method

In general, an artificial neural network (ANN) model includes an input layer, one or more hidden layer, and an output layer. Neurons in each of the layers are called nodes that interconnect with each other and receive linked weights. The typical ANN architecture used in many studies is multi-layer perceptron (MLP) for the formation of the models [15].

In this study, the MLP-ANN type is used with an error back-propagation algorithm [15]. The architecture consists of three layers I(k)-HL(m)-O(n). The input layer (k) contains the variables of the MLR model, an output layer (n) is the stability constant $\log\beta_{11}$ and the number of hidden neurons (m) is specified by neurons on the input and output layer. To find out the best ANN architecture for $QSPR_{ANN}$ model, the process is operate within two steps. Firstly, the m values of hidden neurons are examined by using Neural Designer tools [16], then a data set is used to externally validate the $QSPR_{ANN}$ model from the results of surveyed models. The evaluation of the second step are run on the Matlab 2016a with Neural Network tool (nntool) toolbox [17].

In addition, two basic transfer functions in the neural network such as the hyperbolic sigmoid tangent and log-sigmoid transfer function are used in the training process of ANN models. The transfer functions are represented mathematically as follows [15-17]:

$$a = \tan sig(n) = \frac{2}{\left(1+e^{-2n}\right)^{-1}} \quad (3)$$

$$a = \log sig(n) = \frac{1}{1+e^{-n}} \quad (4)$$

Model Validation

The validation of the $QSPR_{MLR}$ model was appreciated by the statistical parameters like explained variance (R^2_{adj}), coefficient of determination (R^2_{train}), standard error (SE), Fischer's value (F_{stat}), and predicted residual sum of squares (PRESS) [1]. The MLR models were trained with internal validation by LOO statistics (Q^2_{LOO}) and validated externally on another data set using the statistic Q^2_{ex}.

The $QSPR_{ANN}$ models were controlled by three coefficients of determination such as R^2_{train} for training set, Q^2_{test} for test set and Q^2_{CV} for cross-validation set. The ANN model is trained till the mean square error (M*SE*) value is minimized followed by a difference of the output (o) and real (r) values [17]. It is described as followed [15-17]:

$$MSE = \frac{1}{n}\sum_{1}^{n}\left(r_i - o_i\right)^2 \quad (5)$$

This work uses the average absolute values of the relative errors MARE (%) where ARE (%) is the absolute value of the relative errors to compare the quality of the models [18]. Furthermore, to find out the variable contributions in the models, The average contribution

percentage ($MPx_{k,i}$) is used and it is determined according to formula (6) [18]:

$$MPx_{k,i},\% = \frac{1}{N}\sum_{m=1}^{N}\frac{100.\left|b_{k,i}.x_{m,i}\right|}{\sum_{j}^{k}\left|b_{k,j}.x_{m,j}\right|} \quad (6)$$

where N is the number of observations; m is the number of substances used to calculate $Px_{k,i}$ value; $b_{k,i}$ are the parameters of the model.

QSPR$_{MLR}$ models

As above-mentioned, the data set for the construction of QSPR$_{MLR}$ including the 88 stability constants values of complexes are divided into a training set of 71 observations (80 %) and a test set of 17 observations (20 %) [1]. The training process of QSPR$_{MLR}$ models is based on the criteria of statistical values to evaluate the quality of models [1]. The results of QSPR$_{MLR}$ models and the statistical values are shown in table 2.

Results and discussion

Table 2: Selected models QSPR$_{MLR}$ (k of 1 to 5) and statistical values

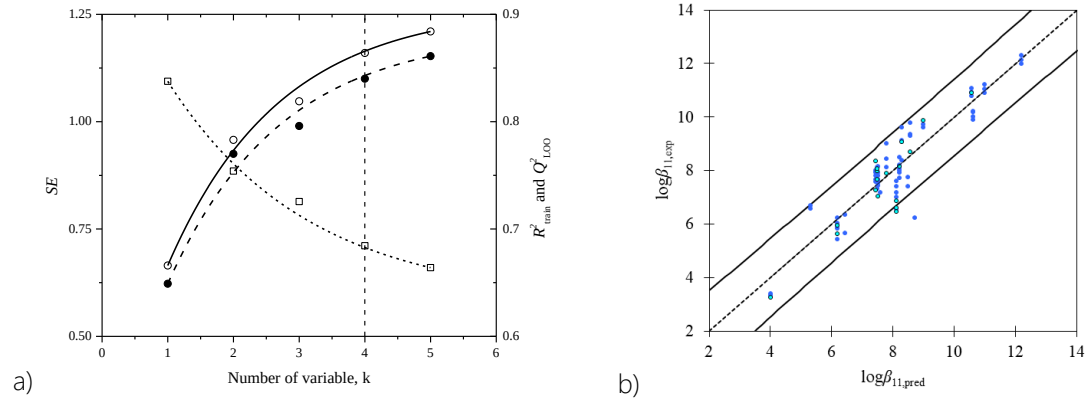
k	Variables	SE	R^2_{train}	R^2_{adj}	Q^2_{LOO}	F_{stat}	$PRESS$
1	x_1	1.094	0.666	0.662	0.649	171.2253	108.02
2	x_1/x_2	0.885	0.783	0.778	0.770	153.8012	70.74
3	$x_1/x_2/x_3$	0.814	0.819	0.813	0.796	126.7552	62.67
4	**$x_1/x_2/x_3/x_4$**	**0.711**	**0.864**	**0.857**	**0.840**	**131.3551**	**49.31**
5	$x_1/x_2/x_3/x_4/x_5$	0.660	0.884	0.877	0.861	924.8164	42.88

Notation of molecular descriptors			
xch9	x_1	Cosmo area	x_4
xv0	x_2	xvch5	x_5
core-core repulsion	x_3		

The number of the variables (k) for the best QSPR model is selected on basis of the changing the R^2_{train}, Q^2_{LOO}, F_{stat}, and $PRESS$ values for meeting statistical requirements.[1,13] Moreover, the variables from x_1 to x_5 were closely monitored on the basis of the p-value (< 0.05) and t-student characterized by the variables [1,13]. The results of fig 2a show that when the k values increase to 5, the QSPR model gets the best statistical values; however, in this case, we choose the k value of 4 because when the k value increase from 4 to 5, the statistical values increase insignificantly and this is unnecessary for the building of models. So the best QSPR$_{MLR}$ model was selected with the statistical values as follows (7):

$$\log\beta_{11} = 10.29 - 261.86\cdot x_1 + 0.763\cdot x_2 + 0.00038\cdot x_3 - 0.053\cdot x_4$$

$$n = 88;\ R^2_{train} = 0.864;\ Q^2_{LOO} = 0.840;\ SE = 0.711;\ \text{Fstat} = 131.3551;\ \text{PRESS} = 49.31 \quad (7)$$



a) b)

Figure 2: (a) Change tendency line of values SE, R^2_{train} and Q^2_{LOO} according to k values; (b) Correlation of experimental vs. predicted values $\log\beta_{11}$ of the training compounds using the QSPR$_{MLR}$ model (with k = 4)

As a consequence, this data set used to build the MLR is completely agreeable and the $QSPR_{MLR}$ model has good predictability for complex groups. Therefore, the model can be used to predict new complexes of the same type group by using the Applicability Domain (AD) and Outliers estimates [1,2].

The investigation uses three models in the vicinity of selected *k* values to validate the role of descriptors in the model according to the $GMPx_i$ values ($GMPx_i$ is the average value of $MPx_{k,i}$). The results in Table 3 show that the contribution of the variables in the order of xv0 (x_2) > Cosmo area (x_4) > core-core repulsion (x_3) > xch9 (x_1) corresponds to the values of 37.2860, 31.4096, 18.9904, and 11.7660. The xv0 parameter (xv0 = $^0\chi^v$ in Kier & Hall notation), namely Chi valence zero, which is a zero-order valence connectivity index computed over all vertices (atoms) in the entire molecules [11].

Cosmo area and core-core repulsion are the quantum parameters, in which Cosmo area is the total area of molecules [12]. The xch9 parameter, namely Chi chain 9, is the simple 9th-order chain chi index and the Chi index is the sum of the 9th-order chain of subgraph [11]. The important parameters are selected to design new complexes and predict the stability constant of these complexes.

Table 3: Statistical values and variables, $MPx_{k,i}$ and $GMPx_i$ contribution in models $QSPR_{MLR}$ with *k* of 4 to 6

Statistical values and variables	$QSPR_{MLR}$			$MPx_{k,i}$, %			$GMPx_i$, %
	k = 3	*k* = 4	*k* = 5	*k* = 3	*k* = 4	*k* = 5	
R^2_{train}	0.819	0.864	0.884	–	–	–	–
R^2_{adj}	0.813	0.857	0.877	–	–	–	–
Q^2_{LOO}	0.796	0.840	0.861	–	–	–	–
RMSE	0.814	0.711	0.660	–	–	–	–
constant	-2.603	10.29	19.13	–	–	–	–
x_1	-197.09	-261.86	-315.07	16.1301	10.3839	8.7840	11.7660
x_2	-0.695	0.763	0.931	57.0639	29.5652	25.2288	37.2860
x_3	0.000284	0.000383	0.000423	26.8059	17.0347	13.1308	18.9904
x_4	–	-0.05344	-0.09108	–	43.0162	51.2125	31.4096
x_5	–	–	-2.100	–	–	1.6439	0.54798

$QSPR_{ANN}$ models

The ANN model is built on the same variables of the MLR model, so the model is developed upon 4 variables of the $QSPR_{MLR}$ model. The data set is split into a training set of 60 observations (70 %), a test set of 14 observations (15 %), and a cross-validation set of 14 observations (15 %) [17]. The architecture of the neural network is I(4)-HL(*m*)-O(1), in which four neurons of the input layer I(4) are xv0, Cosmo area, core-core repulsion and xch9; one neuron of the output layer O(1) is the logβ_{11} values and the number of hidden layer (*m*) is scanned and the results of the *m* neurons are given in table 4.

Table 4: The developed $QSPR_{ANN}$ model I(4)-HL(*m*)-O(1) with statistical parameters

No	$QSPR_{ANN}$ model	R^2_{train}	Q^2_{test}	Q^2_{CV}	Training error	Test Error	Validation Error	Transfer Function
1	I(4)-HL(6)-O(1)	0.9840	0.9765	0.9853	0.0503	0.1216	0.0664	hyperbolic tangent
2	I(4)-HL(8)-O(1)	0.9753	0.9615	0.9802	0.0775	0.1707	0.0829	hyperbolic tangent
3	I(4)-HL(6)-O(1)	0.9640	0.9800	0.9830	0.1129	0.0981	0.0672	log-sigmoid
4	**I(4)-HL(5)-O(1)**	**0.9700**	**0.9739**	**0.9836**	**0.0951**	**0.1129**	**0.0616**	**hyperbolic tangent**
5	I(4)-HL(8)-O(1)	0.9668	0.9697	0.9848	0.1040	0.1245	0.0655	log-sigmoid

https://doi.org/10.51316/jca.2021.096

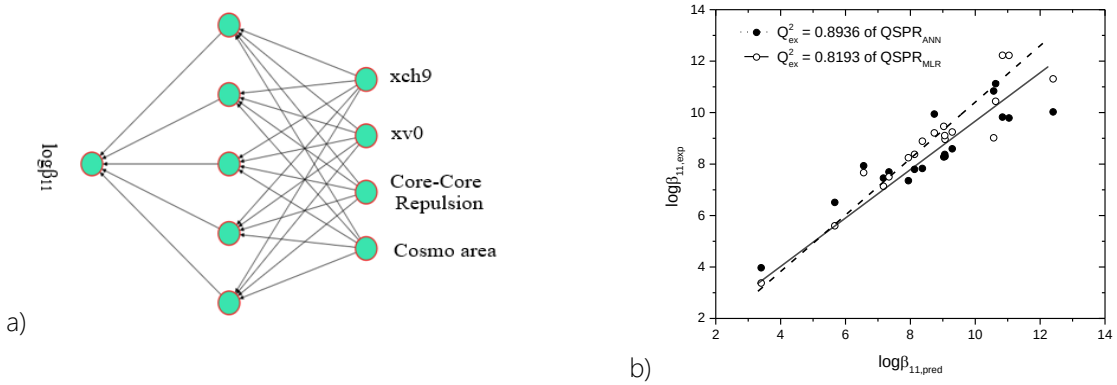


Figure 3: (a) The architecture of neural network I(4)-HL(5)-O(1); (b) The correlation between experimental vs. predicted values of external data set of QSPR models

Using a data set of 18 samples to find out the best $QSPR_{ANN}$ model by external validation technique, the results show that the $QSPR_{ANN}$ model I(4)-HL(5)-O(1) (bolded in table 5) is the best predictability with the Q^2_{ex} value of 0.8936 as in fig 3b. As a consequence, the hyperbolic tangent transfer function is used for network training and the optimum parameters of ANN model such as the momentum constant of 0.05, the learning rate of 0.01, and the convergent goal of 10^{-10}.

External Validation

To build a good model, it is essential to perform the external evaluation on a data set that is independent of the training data set.[1] This work used the external data of 18 observations from the experimental studies. The full information of 18 samples and the validated results are presented in table 5.

As observed in table 5 and fig 3b, the data analysis from the calculated results indicate the neural network and linear regression models express the correlation between the predicted values and the experimental values with the Q^2_{ex} values of 0.8936 and 0.8193, respectively. Hence, the predictability of the two models is extremely positive [1].

As a further matter, the *MARE* values of $QSPR_{MLR}$ and $QSPR_{ANN}$ I(4)-HL(5)-O(1) models are 9.578 % and 5.318 %, respectively. The results point out that the ANN model has better predictability than the MLR model. The predicted $\log\beta_{11,cal}$ values of ANN model approximate to the experimental $\log\beta_{11,exp}$ values. Furthermore, using the one-way ANOVA method to evaluate the difference between the experimental and predictive values of both models; accordingly, the differences between the QSPR models are negligible ($F = 0.1402 < F_{0.05} = 3.1788$).

Table 5: The experimental $\log\beta_{11,exp}$ and external predicted $\log\beta_{11,cal}$ values from the QSPR models

Thiosemicarbazone ligand				Metal ions	$\log\beta_{11,exp}$	$\log\beta_{11,cal}$		ref.
R_1	R_2	R_3	R_4			$QSPR_{MLR}$	$QSPR_{ANN}$	
H	H	H	$-C_{13}H_{16}NO_3$	Zn^{2+}	12.400	10.025	11.309	[19]
H	H	H	$-C_6H_3(OH)OCH_3$	Cd^{2+}	7.340	7.693	7.502	[20]
H	H	H	$-C_6H_3(OH)OCH_3$	Co^{2+}	7.940	7.354	8.245	[20]
H	H	H	$-C_6H_3(OH)OCH_3$	Ni^{2+}	8.370	7.827	8.886	[20]
H	H	H	$-C_6H_3(OH)OCH_3$	Cu^{2+}	9.030	8.276	9.464	[20]
H	H	H	$-C_{10}H_6OH$	Mg^{2+}	3.400	3.967	3.374	[9]
H	H	H	$-C_{10}H_6OH$	Mn^{2+}	5.670	6.510	5.600	[9]
H	H	H	$-C_{10}H_6OH$	Cd^{2+}	6.560	7.933	7.664	[9]
H	H	H	$-C_{10}H_6OH$	Zn^{2+}	7.170	7.454	7.136	[9]
H	H	H	$-C_{10}H_6OH$	Co^{2+}	8.130	7.792	8.376	[9]
H	H	H	$-C_{10}H_6OH$	Ni^{2+}	9.070	8.293	8.959	[9]
H	H	H	$-C_{10}H_6OH$	Cu^{2+}	9.290	8.591	9.244	[9]
H	H	-	$-C_9H_8NO$	Cu^{2+}	9.06	8.371	9.104	[10]
H	H	H	$-C_6H_4NH_2$	Cu^{2+}	10.570	10.838	9.017	[21]
H	H	H	$-C_6H_4NH_2$	Zn^{2+}	8.740	9.941	9.211	[21]

https://doi.org/10.51316/jca.2021.096

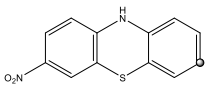
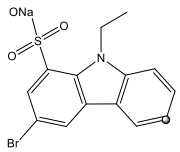
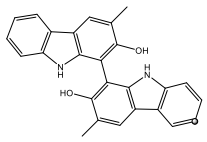
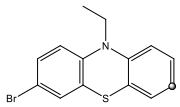
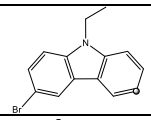
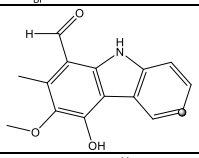
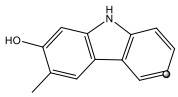
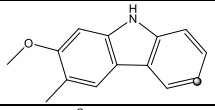
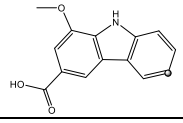
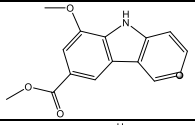
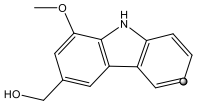
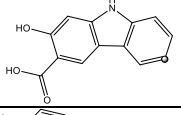
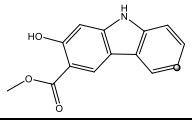
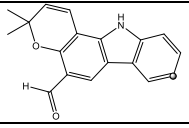
H	H	H	$-C_6H_4NO_2$	La^{3+}	10.840	9.823	12.226	[21]
H	H	H	$-C_6H_4NO_2$	Pr^{3+}	11.040	9.788	12.222	[21]
H	H	H	$-C_6H_4NO_2$	Cd^{2+}	10.630	11.123	10.439	[22]
					MARE, %	9.578	5.318	

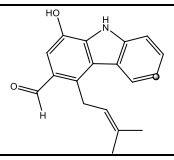
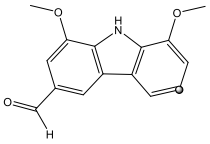
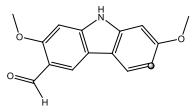
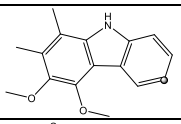
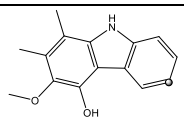
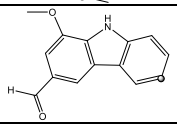
Development of new complexes

Using four descriptors such as xv0, Cosmo area, core-core repulsion, and xch9 of the resulted models and the AD and Outliers rules to develop the new complexes, the phenothiazine and carbazole derivatives are selected to design new thiosemicarbazone and the complexes among the new ligands with several popular metal ions as Ag^+, Cd^{2+}, Cu^{2+}, Ni^{2+}, and Zn^{2+}. These derivatives have been synthesized in experimental researches [23-26]. The new thiosemicarbazones are formed by attaching phenothiazine and carbazole groups at the R_4 site while the remaining positions as R_1, R_2, and R_3 of the thiosemicarbazones are hydrogen atoms.

A series of new complexes are designed and screened carefully by embedding their descriptors into the spatial data of the training data set [1,2]. The results, including 42 new complexes, meet the standard of AD and Outliers rules via D-Cook values [1,2], and the stability constant ($\log\beta_{11,new}$) are predicted by the two built $QSPR_{MLR}$ and $QSPR_{ANN}$ models (table 7).

Table 6: Forty-two new metal-thiosemicarbazone complexes with the calculated $\log\beta_{11,new}$ values from the built QSPR models

R_4 site	Metal ions	$\log\beta_{11,new}$		R_4 site	Metal ions	$\log\beta_{11,new}$	
		MLR	ANN			MLR	ANN
	Ag^+	14.0992	11.8642		Ag^+	11.8770	11.8863
	Cd^{2+}	13.1645	11.8726		Cd^{2+}	10.8608	11.8896
	Cu^{2+}	13.7560	11.3848		Cu^{2+}	11.5941	11.8514
	Ni^{2+}	13.3203	11.1200		Ni^{2+}	10.9582	11.8453
	Zn^{2+}	12.8330	11.7337		Zn^{2+}	10.2510	11.8810
	Ag^+	6.1956	5.0159		Cd^{2+}	14.2768	11.8900
	Cd^{2+}	5.1529	4.9896		Ni^{2+}	14.3241	11.8790
	Cu^{2+}	6.1128	5.1800		Zn^{2+}	13.8725	11.8877
	Ni^{2+}	5.0871	5.4205		Ni^{2+}	6.8250	8.1519
	Zn^{2+}	4.9812	5.1407		Zn^{2+}	9.0312	10.4932
	Ag^+	5.7760	6.3457		Ag^+	6.3136	6.3913
	Zn^{2+}	4.7133	5.6821		Zn^{2+}	5.2077	5.4452
	Cd^{2+}	6.3349	5.9627		Cd^{2+}	6.7341	4.9664
	Cd^{2+}	5.7434	6.4616		Cd^{2}+	5.8818	6.7292
	Cd^{2+}	6.2653	5.7119		Zn^{2+}	8.8060	7.1983

	Cd^{2+}	10.8266	11.7292		Ag^{+}	9.0288	7.5373
	Zn^{2+}	10.5792	10.4042		Zn^{2+}	7.9959	7.8082
	Ag^{+}	9.2069	8.8589		Ag^{+}	10.5251	10.2175
	Zn^{2+}	8.0592	8.9143		Zn^{2+}	9.2111	9.0465
	Ag^{+}	9.3004	9.3045		Ag^{+}	6.9281	7.9893
	Zn^{2+}	7.8424	7.7793		Zn^{2+}	5.8215	7.0230

Furthermore, the single-factor ANOVA method is used to compare the predicted $\log\beta_{11,new}$ values from the resulted $QSPR_{MLR}$ and $QSPR_{ANN}$ models. It indicates that there is no difference between the two models ($F = 0.1930 < F_{0.05} = 4.0195$).

Conclusion

In this investigation, the QSPR models based on the multivariate linear regression and artificial neural network methods were successfully formed by using the dataset of structural descriptors and the stability constant values of metal-thiosemicarbazone complexes. The structures of the complexes were optimized by semi-empirical quantum mechanics with the new version PM7. Also, the QSPR models were completely validated upon the statistical values as R^2_{train}, Q^2_{LOO}, Q^2_{CV}, *MARE* (%), and ANOVA methods. The resulted models are acknowledged as being novel models for the prediction of forty-two new designing thiosemicarbazone derivatives. Furthermore, the results of QSPR models can be effective for discovering new other complexes that can be further applied in essential fields such as analytical chemistry, environment monitoring, and drug design in pharmacology.

References

1. R. Kunal, K. Supratik, N. D. Rudra. A Primer on QSAR/QSPR Modeling, Fundamental Concepts. New York: Springer (2015). https://doi.org/ 10.1007/978-3-319-17281-1
2. OECD, France: Organisation for Economic Co-operation and Development (2007). https://doi.org/10.1787/9789264085442-en
3. J. S. Casas, M. S. García-Tasende, J. Sordo., Coordination Chemistry Reviews 209(1) (2000) 197-261. https://doi.org/10.1016/S0010-8545(00)00363-5
4. M. Hymavathi, C. Viswanatha, N. Devanna. Int. J. Math and Phys. Sci. Res. 2(1) (2014) 43-48.
5. M. Hymavathi, C. Viswanatha, N. Devanna. W. J. Pharm. Phar. Sci. 3(8) (2014) 1688-1695.
6. E. A. Gomaa, K. M. Ibrahim, N. M. Hassan., Int J. Eng. Sci. 3(1) (2014) 44-51.
7. M. Aljahdali, A. A. EL-Sherif., Chimica Acta. 407 (2014) 58-68. https://doi.org/10.1016/j.ica.2013.06.040
8. A. T. A. El-Karim, A. A. El-Sherif., J. Mol. Liq. 219 (2016) 914-922. https://doi.org/10.1016/j.molliq.2016.04.005
9. Sahadev, R. K. Sharma, S. K. Sindhwani., Thermochimica Acta 202 (1992) 291-299. https://doi.org/10.1016/0040-6031(92)85173-S
10. K. Sarkar, B. S. Garg. Thermochimica Acta 113 (1987) 7-14. https://doi.org/10.1016/0040-6031(87)88301-6
11. QSARIS 1.1, USA: Statistical Solutions Ltd, 2001.
12. J. J. P. Stewart. MOPAC2016, Version: 17.240W, Stewart Computational Chemistry, USA, 2016.
13. D. D. Steppan, J. Werner, P. R. Yeater. Essential Regression and Experimental Design for Chemists and Engineers, Free Software Package. http://home.t-online.de/home/jowerner98/indexeng.html, 1998.
14. E. J. Billo.USA: John Wiley and Sons, Inc, 2007. https://doi.org/ 10.1002/9780470126714
15. J. Gasteiger, J. Zupan. Chiw. Inr. Ed. Engl. 32 (1993) 503-521. https://doi.org/10.1002/anie.199305031
16. Artelnics. Neural Designer software, USA: Artificial Intelligence Techniques Ltd., 2020.
17. Matlab R2016a 9.0.0.341360, USA: MathWorks, 2016.

18. V. T. Pham. Development of QSAR and QSPR. Ha Noi: Publisher of Natural sciences and Technique, 2009.
19. M. N. M. Milunovic, E. A. Enyedy, N. V. Nagy, T. Kiss, R. Trondl, M. A. Jakupec, B. K. Keppler, R. Krachler, G. Novitchi, V. B. Arion. L- And D-Proline, Inorg. Chem. 51 (2012) 9309-9321. https://doi.org/10.1021/ic300967j
20. B. S. Garg, V. K. Jain., Thermochimica Acta. 146 (1989) 375-379. https://doi.org/10.1016/0040-6031(89)87108-4
21. S. S. Sawhney, S. K. Chandel, Thermochimica Acta 71 (1983) 209-214. https://doi.org/10.1016/0040-6031(83)80369-4
22. S. S. Sawhney, R. M. Sati., Thermochimica Acta 66 (1983) 351-355. https://doi.org/10.1016/0040-6031(93)85047-D
23. I. J. Al-Busaidi, A. Haque, N. K. Al Rasbi, M. S. Khan. Synthetic Metals 257 (2019) 116-189. https://doi.org/10.1016/j.synthmet.2019.116189
24. G. Sudeshna, K. Parimal., Journal of Pharmacology 648(1-3) (2010) 6-14. https://doi.org/10.1016/j.ejphar.2010.08.045
25. L. Huang, Z. L. Feng, Y. T. Wang, L. G. Lin. Chinese Journal of Natural Medicines 15(12) (2017) 881-888. https://doi.org/10.1016/S1875-5364(18)30003-7
26. G. Krucaite, S. Grigalevicius. Synthetic Metals 247 (2019) 90-108. https://doi.org/10.1016/j.synthmet.2018.11.017