QSPR MODELING APPLICATION FOR DEVELOPMENT OF NEW METAL-THIOSEMICARBAZONE COMPLEXES

Đến tòa soạn: 21-05-2025

Nguyen Minh Quang*, Cao Thanh Nhan
Faculty of Chemical Engineering, Industrial University of Ho Chi Minh City, Ho Chi Minh City

TÓM TẮT

ỨNG DỤNG PHƯƠNG PHÁP MÔ HÌNH HÓA QSPR ĐỂ PHÁT TRIỂN CÁC PHỨC CHẤT MỚI GIỮA ION KIM LOẠI VÀ DẪN XUẤT THIOSEMICARBAZONE

Bộ mô tả của mô hình trong nghiên cứu đã được tính toán cho các phức chất giữa ion kim loại và dẫn xuất thiosemicarbazone (kim loại-thio). Các mô tả này đã được chọn để xây dựng mô hình quan hệ định lượng cấu trúc-tính chất (QSPR) dựa trên các giá trị hằng số bền ($\log\beta_{11}$) của các phức chất kim loại-thio trong nước bằng cách sử dụng phương pháp hồi quy tuyến tính đa biến và phi tuyến tính theo phương pháp máy học (ML). Các mô tả của mô hình QSPR như Dipole, xp5, SsCH3 và xch6 kèm các giá trị thống kê được sử dụng để khẳng định chất lượng của chúng. Mô hình hồi quy tuyến tính đa biến $QSPR_{MLR}$ có R^2_{train} = 0,921; Q^2_{LOO} = 0,886; MSE = 0,639 và F-stat = 143,72. Trong khi đó, mô hình ML có kết quả thống kê ấn tượng như: R^2_{train} = 0,985, R^2_{test} = 0,919 và Q^2_{CV} = 0,965. Các mô hình QSPR đã vượt qua kỹ thuật đánh giá ngoại về khả năng dự đoán trên tập dữ liệu bao gồm mười hai giá trị $\log\beta_{11}$ từ thực nghiệm. Kết quả mười sáu hợp chất kim loại mới đã được phát hiện bằng cách sử dụng mô hình QSPR và kết quả nghiên cứu có thể được sử dụng để tạo ra các phức chất kim loại-thio mới để sử dụng trong hóa học phân tích.

Keywords: *Machine learning, QSPR, stability constants $\log\beta_{11}$, metal-thiosemicarbazone complex.*

1. INTRODUCTION

Metal-thiosemicarbazone (metal-thio) complexes have recently emerged as a significant class of Schiff-based ligands with special donor atoms as S and N. They are typically produced by the condensation reaction of aldehydes or ketones with thiosemicarbazide [1]. It also has a wide range of anticancer action. However, it is highly reliant on cell characteristics. Thiosemicarbazone (Thio/thio) ligands have a broad spectrum of biological actions, including antifungal, antibacterial, antimalarial, anti-proliferative, anti-inflammatory, and antiviral [2].

Heavy metal ions naturally form alliances with one another in minerals. Heavy metals are mainly used to electroplate steel. Naturally, a significant proportion of these metals are released into the environment. Tobacco usage is one of the most common causes of metal ions accumulating in the body over time [3].

Most are highly hazardous metals, and various widely used analytical techniques are used to determine heavy metal ions in trace amounts, especially UV-VIS method [4]. The spectrophotometric approach is recommended because it is less expensive and easier to use, with equivalent sensitivity, accuracy, and good precision. Numerous organic ligands are applied for the different metals determination using this spectrophotometric method.

Sulfur-containing ligands, such as thiosemicarbazones, have recently gained prominence in analytical/inorganic chemistry for metal ion determination [4]. The metal chelates of these sulfur and nitrogen-containing chemicals have numerous applications in medicine and agriculture [2]. A literature review reveals that a few thiosemicarbazones are used

for direct spectrophotometric determination of metals in aqueous solution, but not for extractive spectrophotometric determination. In published articles, the authors introduced novel analytical reagents, thiosemicarbazones, for the spectrophotometric detection of trace metals.

The stability constant plays an essential role in analytical chemistry by controlling the complexation of thiosemicarbazone ligands with metal ions. Currently, multivariate analysis and 0-3D molecular descriptors are utilized to screen the stability constants of metal-thio complexes [5]. In recent decades, QSPR models were quickly created in the theoretical field of chemistry to establish the interactions between metal ionic bonds and organic ligands in aqueous solutions. Metal ions (M = Cu^{2+}, Fe^{2+}, Ni^{2+}, Co^{2+}, Mn^{2+}, Ag^{+}, Zn^{2+}, Pb^{2+}, Cd^{2+},...) had their stability constants stated. However, in many cases, applying QSPR models were highly complicated due to insufficient statistical appraisal, model application capability gaps, incomplete parametric calculation information, and the use of molecular descriptors, statistical parameters, and new statistical techniques. The most effective technique to solve a significant portion of the problem had yet to be fully resolved.

This study employed the quantitative relationship between structure and property (QSPR) to describe the molecular descriptors and stability constant of metal-thiosemicarbazone complexes. The structural descriptors were computed using 0-3D topologically optimized structures and the semi-empirical quantum chemistry methods PM7 [6]. The $QSPR_{MLR}$ models were built using multivariate linear regression. Furthermore, the artificial neural network $QSPR_{ANN}$ model was created by combining significant descriptors from the $QSPR_{MLR}$ model with the error back-propagation training approach. The QSPR models were externally evaluated against the test set. QSPR models' $\log\beta_{11}$ values for metal-thio complexes were compared to experimental data from the literature.

2. METHODOLODY

2.1. Experimental data

Table 1 showed the stability constants ($\log\beta_{11}$) of metal ion (M = Cu^{2+}, Fe^{2+}, Ni^{2+}, Co^{2+}, Mn^{2+}, Ag^{+}) complexes with various thiosemicarbazones in aqueous solution, based on published literature [7-13] under various circumstances. The 0-3D molecular structures of metal-thio complexes modified in the data files were used as input structures for the QSARIS program [14]. Table 1 showed how the 54 complexes from the training set were used to generate the QSPR models. Figure 1 showed the overall structure of the metal-ligand complex.

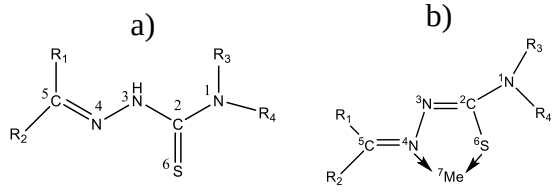


Figure 1. *(a) Thiosemicarbazone; (b) The metal-thio complex*

The production of the complexes could be understood using the Lewis acid-base theory, which stated that an acid accepted electron pairs from bases. In this work, the complexed species with the overall stability constants associated with equilibrium could be represented by the generic equation (1); charges were eliminated for simplicity [15].

$$m\,M + n\,L \rightleftharpoons M_mL_n \qquad (1)$$

The stability constant (β_{pq}) described the creation of the complex ML in a single step with the *m* value of 1 and the *n* value of 1 [15].

$$\beta_{11} = \frac{[ML]}{[M][L]} \quad (2)$$

2.2. Descriptor calculation

The experimental structures of metal-thio complexes were converted into 2D molecular structures using ChemDraw-Pro13. These were optimized using quantum mechanics (QM) on the MoPac2016 system. The topological and quantum descriptors were calculated on the QSARIS program and MoPac2016 systems using the semi-empirical quantum method PM7 [6]. Following the computation of all molecular descriptors, the procedure of descriptor selection was critical. The forward regression technique picked the relevant descriptors in the $QSPR_{MLR}$ models. The acceptable QSPR models had to meet statistically significant criteria.

2.3. QSPR modeling

In this study, the QSPR models were developed using two modelling methods: multivariate linear regression and artificial neural network. The $QSPR_{MLR}$ model's input variable was the foundation for the deep development of artificial neural networks. $QSPR_{MLR}$ modelling confirmed $\log\beta_{11}$ values as dependent variables (Y), whereas structural descriptors (X) were independent variables in the equation. When the values of the X variables correlated strongly with the values of the Y targets, the $QSPR_{MLR}$ model equation was as follows [16]:

$$Y = \beta_0 + \sum_{j=1}^{k} \beta_j X_j \quad (3)$$

where β_0, is the constant of the model, β_i is the regression coefficients and *k* is number of variables in the regression equation.

Table 1. *The 54 stability constants of complexes (n) in training data set*

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	$-C_6H_3BrOH$	H	H	Cu^{2+}	1	5.633	5.633	[7]
2	$-CH_3$	$C_7H_7N_2$	H	H	Mn^{2+}	3	9.600	9.870	[8]
3	$-CH_3$	$C_7H_7N_2$	H	H	Ni^{2+}	2	10.790	10.940	[8]
4	$-CH_3$	$C_7H_7N_2$	H	H	Co^{2+}	2	9.900	10.020	[8]
5	H	$-C_8H_{10}N$	H	H	Ag^{+}	1	17.200	17.200	[9]
6	H	$-C_8H_{10}N$	H	H	Cu^{2+}	1	15.300	15.300	[10]
7	H	$-C_7H_7O_2$	H	H	Fe^{2+}	4	7.690	8.170	[11]
8	H	$-C_7H_7O_2$	H	H	Co^{2+}	4	7.860	8.470	[11]
9	H	$-C_7H_7O_2$	H	H	Ni^{2+}	4	8.110	8.650	[11]
10	H	$-C_7H_7O_2$	H	H	Cu^{2+}	4	9.030	9.830	[11]
11	H	$-C_{10}H_7O$	H	H	Mn^{2+}	4	4.660	5.670	[12]
12	H	$-C_9H_7NO$	H	-	Ni^{2+}	8	7.709	8.500	[13]
13	H	$-C_9H_7NO$	H	-	Co^{2+}	8	7.251	8.340	[13]
14	H	$-C_9H_7NO$	H	-	Mn^{2+}	8	5.439	6.041	[13]

Because of their superior nonlinear fitting capabilities, artificial neural networks had widely used in QSPR research. Neural networks had topology, training procedures, and computational features in their constituents. Three distinct layers, I(k), HL(m), and O(n), were included. Layer 1 was the input layer, where each

neuron received structural information, corresponding to a structure descriptor used as input. The output layer's neurons corresponded to the projected output (The stability constant $\log\beta_{11}$ value). Between these two layers, a hidden layer controled the neural network's predictive power. The complexity of the dataset dictated the critical number of hidden layers [18].

In the multilayer perceptron (MLP) approach of a neural network, the input layer I(k) was connected to all neurons of the hidden layer HL(m), and the output layer O(n) got information from the hidden layer. Each link between layers 1 and 2 carried a weight value [18,19]. This work employed a feed-forward neural network with an error back-propagation method.

The creation of the ANN model occurred during two periods. First, the m values of hidden neurons were pre-screened using Neural Designer tools [20], and the best ANN model was excommunicated via external validation on a data set. The second stage was performed using Matlab2016a [20] with the "*nntool*" toolbox, and the process of ANN model training employed three transfer functions in neural network research: hyperbolic sigmoid tangent, log-sigmoid, and exponential transfer functions [24-27]. The neural network was trained until the mean square error (MSE) was minimized, after which the output was compared to the experimental data [19,20].

The MLP approach was also used to create the artificial neural network models $QSPR_{ANN}$ I(k)-HL(m)-O(n) utilizing neural network function in Matlab2016a's platform. The best $QSPR_{MLR}$ models were chosen using the regression techniques of Regress tool [16], it is add-on of MS-Excel. The predictability of QSPR models was cross-validated using the internal leave-one-out (LOO) approach and external validation. The validated results are compared with the experimental values.

2.4. Validation of models

Modelling strategies aimed to minimize the sum of squared discrepancies between observed and expected values. This minimization resulted in the estimation of the model's parameters. In general, there were two cross-validation techniques used when building a model, which were the Leave-Many-Out (LMO) and Leave-One-Out (LOO) methods. LOO cross-validation was a practical approach for generating an almost unbiased assessment of model performance, especially when working with limited datasets where data utilisation for both training and testing was critical. In this study, the LOO method was used for this goal due to the small dataset. Therefore, the models were screened using the values R^2_{train} for the training set, Q^2_{LOO} or Q^2_{CV} for the cross-validation set, R^2_{test} for an independent test of just the ANN model, and Q^2_{test} for an external validation test of all models [16, 21]. They were calculated using the same formula (4).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (4)$$

where Y_i, $\hat{Y}_i$, and $\bar{Y}$ values are the experimental, calculated and average values, respectively.

Validation was estimating model parameters using a portion of the data set and assessing the neural network's predictability with the remainder. The training set was used to determine the model parameters. The validation set validated the model's predictability. The test set provided a final, independent assessment of model predictability [22]. The mean squared error (*MSE*) employed in

QSPR was a statistical tool for assessing the performance of a predictive model. It calculated the average of the squared errors between calculated and experimental values for a set of substances. The *MSE* is defined by [16,21]

$$MSE = \frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{N-k-1} \quad (5)$$

The projected results of the QSPR models were evaluated by the absolute value of the relative error (ARE,%) [23]. The experimental and computed stability constants were $\log\beta_{11,exp}$ and $\log\beta 11,cal$, respectively, with n representing the number of test compounds. The typical absolute values of the relative error MARE,% [23] were the following equation.

$$MARE,\% = \frac{\sum_{i=1}^{n} ARE_i,\%}{n} = \frac{\frac{\left|\log\beta_{11,exp} - \log\beta_{11,cal}\right|}{\log\beta_{11,exp}}.100}{n} \quad (6)$$

3. RESULTS AND DISCUSSION

3.1. Linear regression modeling

To modelling linear regression, the data set had to be randomly split into training and test subsets, with the training subset taking up 80% of the data. In the first step, $QSPR_{MLR}$ models were built using complexes from the training group, which included 0-3D molecular descriptors and quantum parameters. These $QSPR_{MLR}$ models evaluated the model quality using statistical measures such as k number of molecular descriptors, R^2_{train}, R^2_{adj}, Q^2_{LOO}, MSE, and F-stat. The highest values of R^2_{train}, R^2_{adj}, Q^2_{LOO}, and F-stat indicated model quality, as did the lowest value of MSE. Table 2 showed the statistical values for the $QSPR_{MLR}$ models (k of 4).

The results showed that five models were built in Table 2. In the next step, the selection of the best MLR model was performed by external evaluation on an independent data set (Table 4) based on two parameters: Q^2_{ext} and MARE%. The results in this step showed that the MLR3 model (in bold in Table 2) was the best model selected for ANN model development and design of new ligands and metal-thio complexes.

Table 2. *The results of $QSPR_{MLR}$ models construction*

No	$QSPR_{MLR}$ models
1	$\log\beta_{11}$ = -9.527 - 0.925×*xp3* + 2.389×*SsCH3* - 30.45×^{1}*N* + 21.81×*xch6* R^2_{train} = 0.945; R^2_{adj} = 0.940; Q^2_{LOO} = 0.928 ; MSE = 0.535 ; F-stat = 210.31
2	$\log\beta_{11}$ = 8.081 + 0.002×*core-core repulsion* - 18.89×*xp5* + 18.45×*xp6* - 23.45×*MaxQp* R^2_{train} = 0.942; R^2_{adj} = 0.938; Q^2_{LOO} = 0.912; MSE = 0.548; F-stat = 200.02
3	**$\log\beta_{11}$ = 7.441 - 0.177×*Dipole* - 1.338×*xp5* + 2.409×*SsCH3* + 27.86×*xch6*** **R^2_{train} = 0.921; R^2_{adj} = 0.915; Q^2_{LOO} = 0.886; MSE = 0.639; F-stat = 143.72**
4	$\log\beta_{11}$ = 66.67 - 0.105×*Cosmo Area* - 3.893×*nelem* - 0.516×*ncirc* - 0.943×*numBHa* R^2_{train} = 0.951; R^2_{adj} = 0.947; Q^2_{LOO} = 0.882; MSE = 0.505; F-stat = 237.91
5	$\log\beta_{11}$ = 60.44 - 0.063×*Cosmo Volume* - 4.012×*nelem* - 0.388×*ncirc* - 5.149×*Hmax* R^2_{train} = 0.958; R^2_{adj} = 0.955; Q^2_{LOO} = 0.944; MSE = 0.465; F-stat = 282.23

Therefore, the $QSPR_{MLR3}$ model's variables included *Dipole*, *xp5*, *SsCH3*, and *xch6*, all of which had excellent statistical values. Specifically, the Dipole parameter was the 3D index that represented the molecule's dipole moment [24]. The Gasteiger-Marsili approach, which was implemented in the QSARIS tool [20], calculated the dipole moment based on the 3D structure and charges. Meanwhile, xp5 and xch6 were simple chi indices [24, 25]. The xp5 represented the simple 5th-order indices, which were part of the chi path indices of simple chi indices. The parameter was defined up to the fifth-order path of the molecule. The xch6 was a chi chain six from the chain chi indices of simple chi indices, also known as the simple 6th-order chain chi index. The atom type count index, Ssch3, counted the number of ($-CH_3$) groups in the molecule. It demonstrated that the presence of $-CH_3$ groups in the molecule had a significant impact on the complexes' stability constant values. So, critical indices would be used to create new thio and metal-thio compounds. To use these parameters for the purpose of developing new ligands, this study designed hundreds of new compounds based on a clear concept explained below. Then, optimization and descriptor calculations were performed as performed on the training compounds. Hundreds of new compounds were screened to select the ones that fit the constructed model by inputting the descriptors of these new compounds into the model and performing AD and Outliers evaluation techniques to select the suitable compounds.

3.2. Modeling non-linear ANN model

The neural network was trained with four selected descriptors as input values and $\log\beta_{11}$ stability constants as output values. The $QSPR_{ANN}$ models were created by non-linear characterization of an ANN. The MLP-ANN of I(4)-HL(*m*)-O(1) were created using the Levenberg-Marquardt algorithm. Table 1 showed the utilization of 54 metal-thiosemicarbazone complexes as a training set. The non-linear ML of ANN I(4)-HL(*m*)-O(1) model was utilized by determination of the stability constant $\log\beta_{11}$ values on the external data set, as illustrated in Table 4. Table 3 showed the initial results of *m* values neurons.

Table 3. *The initial screening of MLP-ANN I(4)-HL(m)-O(1) model*

Symbol	The *m* values of HL(*m*) layer	R^2_{train}	Q^2_{test}	Q^2_{CV}	Training error	Test error	Validation error	Training algorithm
ANN1	HL(3)	0.987	0.913	0.967	0.075	0.143	0.099	Tanh
ANN2	HL(7)	0.985	0.918	0.966	0.087	0.135	0.104	Logistic
ANN3	**HL(4)**	**0.985**	**0.920**	**0.965**	**0.088**	**0.133**	**0.124**	**Logistic**
ANN4	HL(3)	0.986	0.918	0.963	0.082	0.135	0.115	Exponential
ANN5	HL(6)	0.982	0.919	0.966	0.109	0.135	0.116	Logistic

In the next stage, an external dataset was utilized to train the optimal neural network model. This was followed by an external evaluation of the MLR model using the external validation technique, specifically employing the Q^2_{ext} index. The results revealed the ANN model with the architecture I(4)-HL(4)-O(1), as illustrated in Figure 2, which was emphasized in bold in Table 3. This model achieved the highest predictability, reflected in a Q^2_{ext} value of 0.899, as shown in Figure 3. Consequently, a

logistic transfer function was employed to identify the best network training and to determine the optimal parameters for the artificial neural network (ANN), including a learning rate of 0.01, a momentum constant of 0.05, and a convergence objective of 10^{-10}.

Therefore, the optimal neural network I(4)-HL(4)-O(1) chosen was statistically significant, with values R^2_{train} = 0.985, R^2_{test} = 0.920, and Q^2_{CV} = 0.965. Furthermore, the results in Figure 3 demonstrate that the MLP-ANN4 of I(4)-HL(4)-O(1) model outperforms the MLR model. Thus, the MLP-ANN4 model was appropriate for predicting the external dataset.

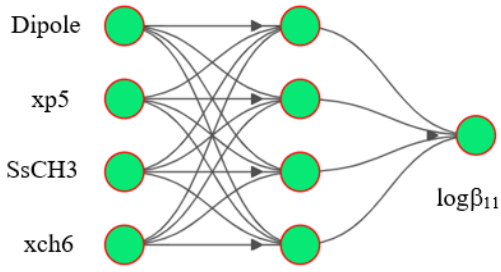


Figure 2. *Architecture of neural network I(4)-HL(4)-O(1)*

3.3. Validation of QSPR models

A comprehensive QSPR model had to be externally assessed using an independent data source [21]. It was the final phase in the model-building procedure. The other dataset for external validation containing twelve metal-thio compounds from the published articles was used to validate both the $QSPR_{MLR}$ and $QSPR_{ANN}$ models. The evaluation results were displayed in Table 4.

According to the calculable data in Table 4, the MARE values for the linear and non-linear ML I(4)-HL(4)-O(1) models were 7.898 % and 5.711 %, respectively. The non-linear ML model of ANN I(4)-HL(4)-O(1) outperformed the linear model in terms of predictability, with projected $log\beta_{11,cal}$ values closely matching experimental $log\beta_{11,exp}$ values. Based on the data analysis in Table 4 and Figure 3, the predictions from the two models are conclusive. Both the ANN4 model and the MLR3 model demonstrated a strong correlation between the predicted values and the experimental values, with Q^2ext values of 0.899 and 0.814, respectively. Besides, the MARE, % value also showed that the choice of the ANN4 model among the five initially developed ANN models (Table 3) was reasonable. At the same time, the result also shows that the prediction of the ANN4 model was better than the MLR3 model with the MARE, % values of 5.711 and 7.898, respectively.

The ANOVA approach was used to examine the differences between the values of both models in Figure 3. Consequently, the differences were minimal ($F = 0.0751 < F_{0.05} = 3.2849$).

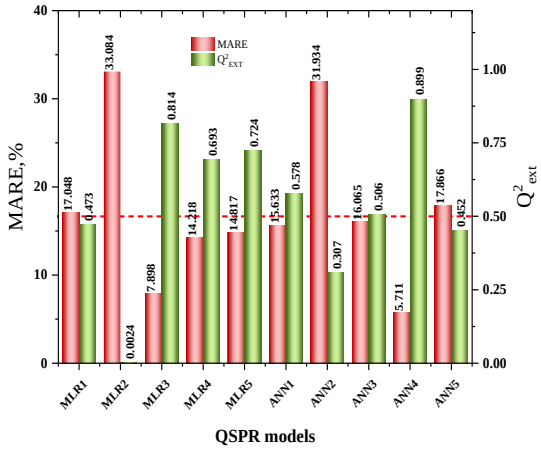


Figure 3. *The Q^2_{ext} and MARE(%) values of models*

Table 4. *The data set for validating externally the models*

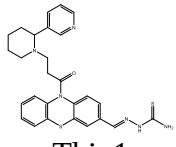
No.	Ligands				Metal ions	$\log\beta_{11.exp}$	$\log\beta_{11.cal}$		ref.
	R_1	R_2	R_3	R_4			MLR	ANN	
1	H	$-C_7H_7O_3$	H	H	Ni^{2+}	6.489	7.316	6.999	[26]
2	$-CH_3$	$C_7H_7N_2$	H	H	Ni^{2+}	11.080	10.799	11.091	[8]
3	$-CH_3$	$C_7H_7N_2$	H	H	Cu^{2+}	11.530	10.701	10.465	[8]
4	H	$-C_7H_7O_2$	H	H	Pb^{2+}	7.100	8.042	7.408	[11]
5	H	$-C_7H_7O_2$	H	H	Cd^{2+}	7.340	8.297	7.407	[11]
6	H	$-C_7H_7O_2$	H	H	Zn^{2+}	7.470	8.404	7.623	[11]
7	H	$-C_7H_7O_2$	H	H	Fe^{2+}	8.150	8.524	7.882	[11]
8	$-C_6H_4$	$-C_7H_6NO$	H	H	Cu^{2+}	5.748	5.523	5.034	[27]
9	H	$-C_9H_7NO$	H	-	Cu^{2+}	9.060	7.591	7.858	[13]
10	H	$-C_9H_7NO$	H	-	Pb^{2+}	7.307	7.362	6.914	[13]
11	H	$-C_9H_7NO$	H	-	Zn^{2+}	7.039	6.605	6.966	[13]
12	H	$-C_9H_7NO$	H	-	Cd^{2+}	6.611	6.732	7.181	[13]
						MARE, %	7.898	5.711	

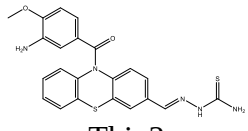
3.4. Development of novel metal-thio derivatives

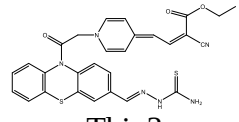
The study made the structuring new thio ligands by adding selected phenothiazine and carbazole derivatives, to form new thio and metal-thio complexes between thio ligands with metal ions, such as Cu^{2+}, Ag^{+}, Cd^{2+}, Zn^{2+}, and Ni^{2+}. The selected compounds were referred in published documents [28-30]. The design was based on the four descriptors, *Dipole, xp5, Ssch3,* and *xch6,* of QSPR models. The new thio ligands were developed by attaching the groups at the R_4 site of the structure (Figure 1a), and the other positions, such as R_1, R_2, and R_3, were hydrogen atoms.

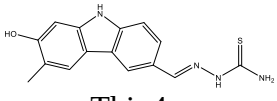
Various freshly constructed compounds were drawn, and structural variables were derived. They were carefully selected and integrated into the training data set for check AD and Outlier [21], utilizing Cook's distance value (D_{Cook}). QSPR models would forecast new derivatives inside the AD of the training dataset with $|D_{Cook}|$ values less than 1.0, whereas derivatives found in the dataset's outliers with $|D_{Cook}|$ values more than 1.0 will be rejected.

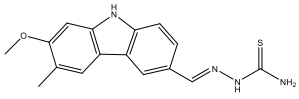
The sixteen novel complexes of eleven thiosemicarbazone ligands (Figure 4) matched the AD criteria, and their stability constant was anticipated by the two developed QSP_{MLR} and $QSPR_{ANN}$ models. Figure 6 showed the predicted $\log\beta_{11,pred}$ values for sixteen novel metal-thio complexes.

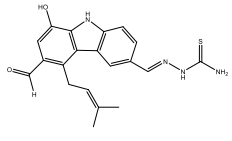

Thio1

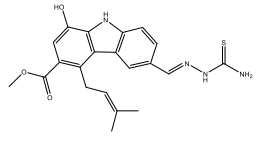

Thio2

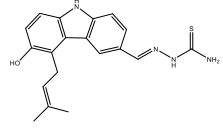

Thio3

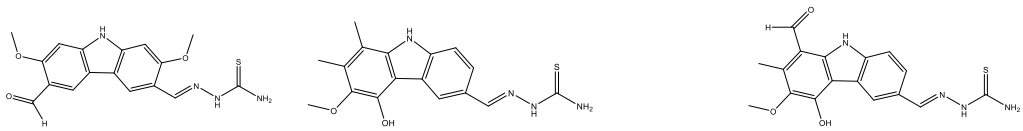

Thio4


Thio5


Thio6


Thio7


Thio8



Thio9 Thio10 Thio11

Figure 4. *Structures of eleven new thiosemicarbazone ligands*

Similar properties of newly optimized complexes were added to the training dataset to precisely manufacture and screen many novel thiosemicarbazone-ligand molecules. The stability constants that had to exceed the AD and Outliers requirements were anticipated by new complexes using the absolute values D-Cook. The findings are shown in Figure 5.

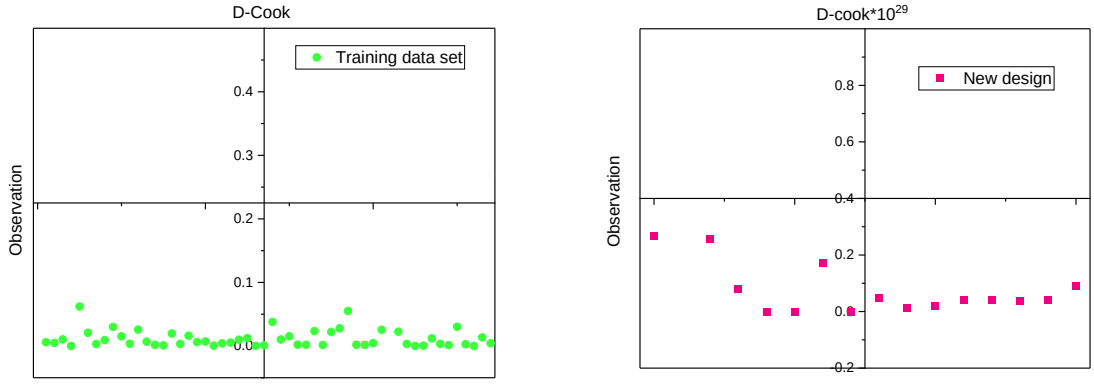


Figure 5. *The D-Cook values of the building data set and new complexes*

A comparison of calculated values from MLR and ANN models ($\log\beta_{11,pred}$) using single-factor ANOVA revealed no significant difference ($F = 0.9141 < F0.05 = 4.1709$).

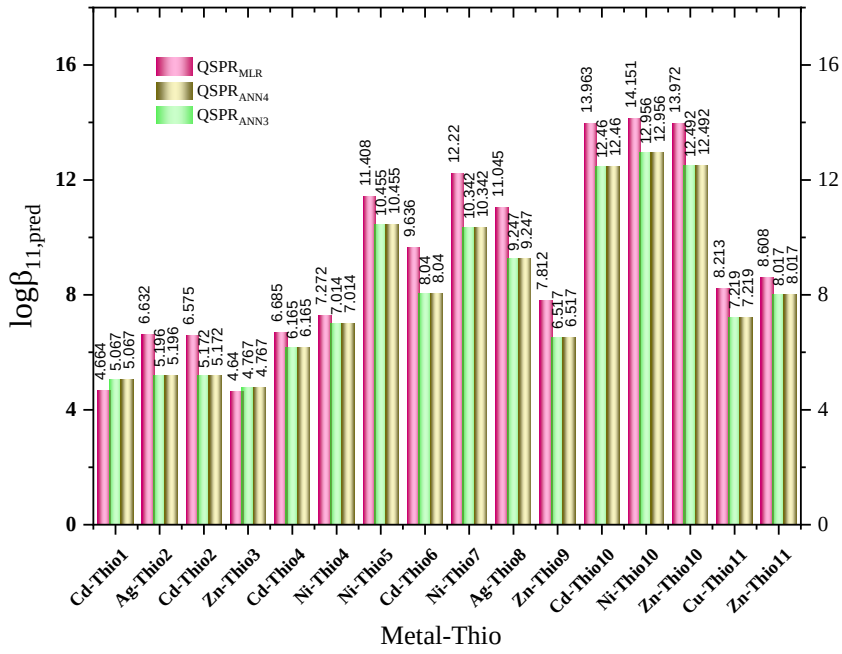


Figure 5. *The* $\log\beta_{11,pred}$ *values of sixteen metal-thio compounds*

4. CONCLUSION

The study found that the *in silico* approach effectively models for complexes between metal ions and thio ligands, with stability constants $\log\beta_{11}$ values. The $QSPR_{MLR}$ models were constructed using multivariate regression techniques. The 4-decriptors linear model effectively generated the best non-linear ML model, ANN of I(4)-HL(4)-O(1). The created QSPR models were statistically acceptable. These QSPR models could accurately estimate the $\log\beta_{11}$ stability constants of metal-thio complexes. QSPR models predicted $\log\beta_{11}$ stability constants for complexes in the external set, consistent with experimental results. QSPR models might predict logb11 stability constants for freshly designed ligand complexation with metal ions. As a result, sixteen novel metal-complexes were newly designed using QSPR models, and these results promised to bring many values to fields such as analytical chemistry, environment, and pharmacy.

REFERENCES

1. Ezhilarasi et al, (2012). Synthesis Characterization and Application of Salicylaldehyde Thiosemicarbazone and Its Metal Complexes. *Int. J. Res. Chem. Environ.*, **(2)**, pp. 130-148.
2. S. B. Zahra et al, (2025). Versatile biological activities of thiosemicarbazones and their metal complexes. *Journal of Molecular Structure*, **(1322)**, 140511.
3. M. Singh, and P. S. Verghese, (2016). Conventional and innovative techniques for removal of heavy metals from electroplating industry waste water. *Int. J. Eng. Sci. Res. Tech.*, **(5)**, 150-159.
4. A. K. Rai, S. N. Upadhyay, S. Kumar, and Y. D. Upadhyay, (1998). Heavy metal pollution and its control through a cheaper method: A review. *J. IAEM*, **(25)**, 22-51.
5. N. M. Quang, T. X. Mau, N. T. A. Nhung, T. N. M An, and P. V Tat, (2019). Novel QSPR modeling of stability constants of metal-thiosemicarbazone complexes by hybrid multivariate technique: GA-MLR, GA-SVR and GA-ANN. *Journal of Molecular Structure*, **(1195)**, 95-109, .
6. J. J. P Stewart, (2013). Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model*, **19**, 1-32.
7. G. Ramanjaneyulu, P. R. Reddy, V. K. Reddy, and T. S. Reddy, (2008). Direct and Derivative Spectrophotometric Determination of Copper(II) with 5-Bromosalicylaldehyde Thiosemicarbazone. *The Open Anal. Chem. J.*, **(2)**, 78-82.
8. A. T. A. El-Karim, and A. A. El-Sherif, (2016). Potentiometric, equilibrium studies and thermodynamics of novel thiosemicarbazones and their bivalent transition metal(II) complexes. *J. Mol. Liq.*, **(219)**, 914-922.
9. Sahadev, R. K. Sharma, and S. K. Sindhwani, (1992). Potentiometric Studies on the Complexation Equilibria Between Some Trivalent Lanthanide Metal Ions and Biologically Active 2-Hydroxy-1- Naphthaldehyde Thiosemicarbazone (HNATS). *Monatshefte fur Chemie*, **(123)**, 883-889, .
10. T. Atalay, and E. Ozkan, (1994). Thermodynamic studies of some complexes of 4'-morpholinoacetophenone thiosemicarbazone. *Thermochimica Acta*, **(237)**, 369-374.
11. B. S. Garg, and V. K. Jain, (1989). Determination of thermodynamic parameters and stability constants of complexes of biologically active o-

vanillinthiosemicarbazone with bivalent metal ions. *Thermochimica Acta*, **(146)**, 375-379, .

12. Sahadev, R. K. Sharma, and S. K. Sindhwani, (1992). Thermal studies on the chelation behaviour of biologically active 2-hydroxy-1-naphthaldehyde thiosemicarbazone (HNATS) towards bivalent metal ions: a potentiometric study. *Thermochimica Acta*, **(202)**, 299.
13. K. Sarkar, and B. S. Garg, (1987). Determination of thermodynamic parameters and stability constants of the complexes of p-MITSC with transition metal ions. *Thermochimica Acta*, **(113)**, 7-14.
14. Statistical Solutions. QSARIS 1.1., USA, (2001).
15. D. Harvey, Modern analytical Chemistry, Mc.Graw Hill, Boston, Toronto, (2000).
16. D. D. Steppan, J. Werner, and P. R. Yeater, Essential Regression and Experimental Design for Chemists and Engineers, Free Software Package, (1998).
17. J. Gasteiger, and J. Zupan, (1993). Neural Networks in Chemistry. *Chiw. Inr. Ed. EngI.*, **(32)**, 5-15.
18. R. Rojas, Neural Networks, Springer-Verlag, Berlin, (1996).
19. D. R. Baughman, and Y. A. Liu, Neural Networks in Bioprocessing and Chemical Engineering, Academic Press: San Diego, CA, (1995).
20. MathWorks. Matlab R2016a 9.0.0.341360. USA, (2016).
21. OECD, Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships Models. France: Organisation for Economic Co-operation and Development, (2007).
22. A. Golbraikh, and A. Tropsha, (2002). Beware of Q2. *J. Mol. Graphics Model*, **(20)**, 269-276.
23. P. V. Tat, Development of QSAR and QSPR. Ha Noi: Publisher of Natural sciences and Technique, (2009).
24. L. B. Kier, and L. H. Hall, Molecular Connectivity in Structure-Activity Analysis. Wiley & Sons Inc, (1986).
25. L. H. Hall, and L. B. Kier, (2007). The Molecular Connectivity Chi Indexes and Kappa shape Indexes in Structure-Property Relations. *In Review of Computational Chemistry*, **(2)**, 367-422.
26. M. Hymavathi, C. Viswanatha, and N. Devanna, (2014). A Study on Synthesis of Novel Chromogenic Organic Reagent 3,4-dihydroxy-5-methoxy benzaldehyde thiosemicarbazone and Spectrophotometric Determination of Nickel (II) in Presences of Triton X-100. *Res. J. Pharm. Bio. and Chem. Sci.*, **(5)**, 625-630.
27. K. H. Reddy, and N. B. L. Prasad, (2004). Spectrophotometric determination of copper (II) in edible oils and seed using novel oxime-thiosemicarbazones. *India J. Chem.*, **(43A)**, 111-114.
28. L. Huang, Z. L. Feng, Y. T. Wang, and L. G. Lin, (2017). Anticancer carbazole Alkaloids and coumarins from Clausena plants: A review. *Chinese Journal of Natural Medicines*, **(15)**, 881-888.
29. G. Krucaite, and S. Grigalevicius, (2019). A review on low-molar-mass carbazole- based derivatives for organic light emitting diodes. *Synthetic Metals*, **(247)**, 90-108.
30. I. J. Al-Busaidi, A. Haque N. K. Al-Rasbi, and M. S. Khan, (2019). Phenothiazine-based derivatives for optoelectronic applications: A review. *Synthetic Metals*, **(257)**, 116189.