DOI: https://doi.org/10.62985/j.huit_ojs.vol24.no3.83

DISCOVERY OF NOVEL METAL-THIOSEMICARBAZONE COMPLEXES USING *IN SILICO* MODELS IN ENVIRONMENTAL ANALYSIS

Huynh Ngoc Chau, Cu Huy Duc, Nguyen Minh Quang*

Industrial University of Ho Chi Minh City, Ho Chi Minh City

*Email: *nguyenminhquang@iuh.edu.vn*

Received: 16 October 2023; Accepted: 24 January 2024

ABSTRACT

Twenty novel metal-thiosemicarbazone complexes were discovered by *in silico* techniques. The stability constants ($\log\beta_{12}$) of complexes were also predicted by using the quantitative structure and property relationships (QSPR) models. The models were created using the multiple linear regression (MLR) and artificial neural network (ANN) approaches. The structure characteristics of complexes consist of molecular and quantum properties. The published literature is used to collect the stability constants with experimental parameters. The best model, MLR2-QSPR (k = 4), consisted of molecular descriptors such as S^6, *Dipole*, *xv1*, and N'^4. Statistical metrics such as R^2_{train} = 0.913, Q^2_{LOO} = 0.903, and SE = 0.408 were used to validate the quality of this MLR-QSPR. The statistical data for the ANN4-QSPR model I(4)-HL(10)-O(1) were also reported: R^2_{train} = 0.972, Q^2_{test} = 0.975, and R^2_{CV} = 0.985. In addition, the work used the results of variables from the QSPR models for developing new thiosemicarbazone ligands and based-ligand complexes. As a result, novel metal-thiosemicarbazone complexes were newly outlined and predicted the stability constants by two developed QSPR models. The results obtained from models can be applied to develop novel chemicals that can be administrated for use in analytical chemistry and environmental evaluation monitoring.

*Keywords***:** ANN, complexes of thiosemicarbazone, MLR, QSPR, stability constants $\log\beta_{12}$.

1. INTRODUCTION

Nowadays, as the industry grows, toxic metal ions are emitted into the environment from manufacturing facilities, polluting the environment. Heavy metal ion control and analysis must be quick and affordable to fulfill practical requirements. Besides, many methods for determining heavy metal concentration have been employed around the worldwide [1], with ligand-metal ion complexes commonly used [2]. In which, Photometric analysis is a very useful instrument in chemical analysis. It is a flexible technique for determining the concentration of a wide range of compounds in solution. It is also a reasonably easy and affordable approach to execute, making it broadly available. Because thiosemicarbazone is easy to high complexity and multiple research published on its application based on a simple and inexpensive spectrophotometry analysis approach, we propose to characterize the thiosemicarbazone derivative in this study.

When employing metal complexes, the stability constants are a crucial consideration. Metal-ligand bond strengths are a measure of how strongly a ligand binds to a metal ion. They can influence the complex's characteristics and reactivity. A compound with a high stability constant is less likely to dissociate and more difficult to replace with another ligand.

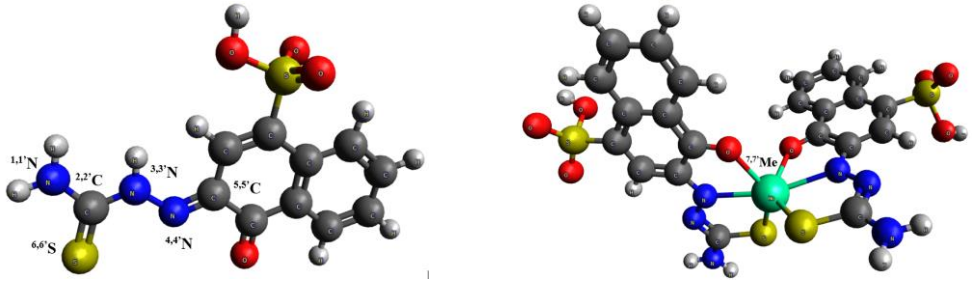
Theoretical research on using computational chemistry to tackle complex mathematical problems, as well as acceptable mathematical methodologies, is becoming more popular [3]. This is due to the ability of computational chemistry to describe and simulate interactions between metal ions and ligands, which can be difficult to examine experimentally. This study has resulted in the creation of novel prediction algorithms.

In general, the use of structural descriptors and stability constants to construct QSPR models of complexes between metal ions and thiosemicarbazone is a powerful tool for understanding and predicting the properties of these complexes. These descriptors can provide additional information about the electronic structure of the complexes, which can be useful for predicting their stability constants. In the specific case of using the semi-empirical quantum mechanics (QM) methods PM7-PM7/sparkle [4] as well as molecular mechanics and connectivity computations, these methods can be used to calculate additional structural descriptors for the complexes in the dataset. These descriptors can then be used to build more accurate and predictive QSPR models. In this work, two models of the methods of the multivariable linear regression (MLR-QSPR) and artificial neural network (ANN-QSPR) are built to discover the new complexes. Error back-propagation and MLP are used to build ANN models with structural descriptors from the best MLR-QSPR model. In addition, the $\log\beta_{12}$ values of complexes in the test set are validated and compared to experimental results.

2. METHODOLOGY

2.1. Data set

The first stage in the research should be to identify the complexes that will comprise the data set. Table 1 shows the $\log\beta_{12}$ values of metal-thiosemicarbazone (Me-ligand) complexes gathered from the literature. Thiosemicarbazone ligands and Complexes are typically represented schematically as seen in Figure 1 [1].



(a) (b)

Figure 1. Structure of the thiosemicarbazone (a) and Me-ligand (b) complexes in the work with fixed atomic sites of the structural framework

The $\log\beta_{12}$ values are log-transformed metal-thiosemicarbazone complex stability constants. The stability constant, derived from an aqueous solution reaction between a metal ion (M) and a thiosemicarbazone (L), quantifies the equilibrium between the free metal ion and the metal-ligand complex, providing crucial insights into the thermodynamics and kinetics of metal-ligand interactions [5].

$$pM^{n+} + qL^{m-} \rightleftharpoons [M_pL_q] \qquad (1)$$

The equation represents the stability constant for the production of ML_2-form complexes in solution and provides information about their stability as follows [5]:

$$\beta_{12} = \frac{[ML_2]}{[M^{n+}][L^{m-}]^2} \qquad (2)$$

The stability constants ($\log\beta_{12}$) of the complexes ML_2 of some M ions (M = Ho^{3+}, Dy^{3+}, Tb^{3+}, Y^{3+}, Eu^{3+}, Sm^{3+}, Gd^{3+}, Nd^{3+}, Pr^{3+}, Ce^{3+}, and La^{3+}) with (E)-3-(2-carbamothioylhydrazono)-4-oxo-3,4-dihydronaphthalene-1-sulfonic acid thiosemicarbazones (L) in aqueous solution at various experimental conditions (temperature, pH, and ionic strength) were selected from the published literature [6].

Table 1. The 77 $\log\beta_{12}$values of experimental complexes (n) with range of $\log\beta_{12,min}$ to $\log\beta_{12,max}$ values on training data set

Ligand				Metal ions	Number of complexes, *n*	$\log\beta_{12,min}$	$\log\beta_{12,max}$	Ref.
R_1	R_2	R_3	R_4					
H	H	-	$-C_{10}H_7O_4S$	Ho^{3+}	7	10.420	11.540	[6]
H	H	-	$-C_{10}H_7O_4S$	Dy^{3+}	7	9.660	10.420	[6]
H	H	-	$-C_{10}H_7O_4S$	Tb^{3+}	7	9.270	10.110	[6]
H	H	-	$-C_{10}H_7O_4S$	Y^{3+}	7	8.900	9.700	[6]
H	H	-	$-C_{10}H_7O_4S$	Eu^{3+}	7	8.650	9.330	[6]
H	H	-	$-C_{10}H_7O_4S$	Sm^{3+}	7	8.420	9.070	[6]
H	H	-	$-C_{10}H_7O_4S$	Gd^{3+}	7	7.670	8.550	[6]
H	H	-	$-C_{10}H_7O_4S$	Nd^{3+}	7	7.330	8.130	[6]
H	H	-	$-C_{10}H_7O_4S$	Pr^{3+}	7	6.850	7.930	[6]
H	H	-	$-C_{10}H_7O_4S$	Ce^{2+}	7	6.840	7.680	[6]
H	H	-	$-C_{10}H_7O_4S$	La^{2+}	7	6.040	6.900	[6]

2.2. Descriptors

Quantum and 0-3D molecular descriptors are all used in QSPR modeling. Using BIOVIA Draw 2017 R2 [7], the structures of experimental complexes were recreated. The complexes were then optimized using the MoPac2016 software with the method of quantum mechanics QM [8]. On QSARIS [9], the 0-3D molecular descriptors were calculated using the ideal structures. The MoPac2016 system was used to determine the quantum parameters using the quantum approach PM7-PM7/sparkle [4]. The models were constructed on descriptor parameters and stability constants ($\log\beta_{12}$) as a data set [3].

2.3. Estimation of QSPR models

2.3.1. MLR-QSPR models

MLR is a powerful tool for understanding the complex relationships between variables. It can be used to make predictions, identify causal relationships, and develop policies and interventions. It is written as following [10]:

$$Y = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + ... + \beta_k.X_k + \varepsilon \qquad (3)$$

Here, the regression coefficients (β_0, β_1, β_2 ..., β_k) measure the strength and direction of the relationship between each independent variable (X_i) and the dependent variable (Y), while holding all other independent variables constant. The error term (ε) represents the unexplained variation in the dependent variable.

MLR is used in the case to create a link of the $\log\beta_{12}$ values and the structural characteristics that affect them. Meanwhile, the dependent variable is the log-transformed stability constant ($\log\beta_{12}$), while the other parameters are independent.

MLR-QSPR models are constructed using the multiple linear regression (MLR) approach, which uses the least squares principle to select model variables. This means that the model parameters are chosen to minimize the sum of squared discrepancies between the actual

and calculated values of the dependent variable. R^2_{train} and Q^2_{LOO} values were used to screen the models [10,11]. These are computed using the same formula (4):

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

This equation uses the following notation: n is the number of observations; Y_i, $\hat{Y}_i$, and $\bar{Y}$ are the experimental, calculated, and average values of the i^{th} observation.

Another important statistic used to evaluate the performance of a multiple linear regression (MLR) model is the standard error (SE). The SE is a measure of how accurately the model can predict the dependent variable, given the values of the independent variables. The lower the SE, the more accurate the model is expected to be. This quantity is significantly related to the estimate's standard error, which is defined below [10, 11].

$$SE = \sqrt{\frac{\sum_{1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n-k-1}} \qquad (5)$$

2.3.2. ANN-QSPR models

An ANN is a mathematical model inspired by the human brain, which is made up of interconnected nodes called neurons. It is made up of interconnected nodes, called artificial neurons, that process information by passing signals to each other. ANNs are trained on data to learn to perform specific tasks, such as image recognition, natural language processing, and machine translation. They are now used in a wide range of applications, including mathematics, robotics, medicine, and chemistry [12].

Artificial neural networks (ANNs) are composed of layers of interconnected neurons. Each layer has its weights, which determine how the neurons in that layer respond to the inputs from the previous layer. The most common ANN architecture is the multilayer perceptron (MLP), which has one input layer, one hidden layer, and one output layer. The ANN used in this study is a multilayer feed-forward network with a single hidden layer. This means that the neurons in the hidden layer are only connected to the neurons in the input layer and the output layer [13].

Furthermore, the model was trained using a standard feed-forward neural network with error back-propagation. This type of neural network propagates information from the input layer to the output layer in a single direction, without any feedback loops. The following mathematical equation describes how this process works [12,13]:

$$o_j = f\left(\sum_{i=0}^{N} w_{ij}.x_i - q_j\right) \qquad (6)$$

In this equation, x_i represents the input factor, o_j represents the output factor, w_{ij} represents the weight factor between two nodes, q_j represents the internal threshold, and f represents the transfer function.

To train ANN-QSPR models, the study used the tanh and log-sigmoid transfer function. The functions are both commonly applied in neural networks and both non-linear functions that squash their inputs to a range of values, making them well-suited for modeling complex relationships between data. The following equation describes it [13]:

$$g(x) = \log sig(x) = \frac{1}{1+e^{-x}} \qquad (7)$$

$$g(x) = \tan sig(x) = \frac{1-e^{-x}}{1+e^{-x}} \qquad (8)$$

ANN-QSPR models are built until the mean value of square error (SE_{ANN}) is cut down, and then the network output is compared to the real values of the output acquired from experimental data [12]. The mean squared error between predictions of network (o) and actual values (t) is a measure of how inaccurate the predictions of the network via SE_{ANN} quantity [13]. It is written as follows:

$$SE_{\text{ANN}} = \frac{1}{n}\sum_{1}^{n}\left(t_i - o_i\right)^2 \quad (9)$$

3. RESULTS AND DISCUSSION

3.1. MLR-QSPR models

The MLR-QSPR model was built using the data set of complexes in Table 1. This database included the variables and stability constant $\log\beta_{12}$. The initial structures of the metal-thiosemicarbazone complexes were drawn in the BIOVA tool [7] and optimized using QM on the MoPac2016 system [8]. Quantum parameters were generated using the semi-empirical QM methods PM7-PM7/sparkle [7]. The optimized geometry of the molecule was then fed into the QSARIS system [9], which computed the topological descriptors.

The dataset was first split into training and test subsets, with the test subset containing about 20% of the initial data. The training subset was then applied to develop the regression models. Two types of QSPR models were built: MLR-QSPR and ANN-QSPR. MLR-QSPR models were built using the stepwise regression techniques on the Regress system [10]. ANN-QSPR models were built on the Matlab system using the multilayer training technique [13]. The forecasting ability of the QSPR models was cross-validated (CV) using the leave-one-out method (LOO) and the statistic Q^2_{LOO}.

Table 2. Detailed statistics of the resulted in MLR-QSPR models

No	Symbol	The QSPR models
1	MLR1-QSPR	$\log\beta_{12}$ = -851.951 – 0.852**Volume* + 0.135**MW* – 1.520**SpcPolarrizability* + 0.520**LUMO*. R^2_{train} = 0.892, Q^2_{LOO} = 0.880, SE = 0.455
2	**MLR2-QSPR**	**$\log\beta_{12}$ = -32.690 + 7.236*S^6 – 1.794**Dipole* – 2.122**xv1* - 9.673*N'^4. R^2_{train} = 0.913, Q^2_{LOO} = 0.903, SE = 0.408**
3	MLR3-QSPR	$\log\beta_{12}$ = 63.640 + 0.946**HOMO* – 46.594*C'^2 + 397.815**xvch8* – 0.128**Surface*. R^2_{train} = 0.829, Q^2_{LOO} = 0.807, SE = 0.573
4	MLR4-QSPR	$\log\beta_{12}$ = 57.958 – 0.241**Dipole* – 57.314*C'^2 + 460.578**xvch8* – 0.133**Surface*. R^2_{train} = 0.890, Q^2_{LOO} = 0.876, SE = 0.460
5	MLR5-QSPR	$\log\beta_{12}$ = -50.413 – 0.223**Cosmo Volume* – 10.950*Me^7 + 8.608**xv0* – 0.001*ΔH_f. R^2_{train} = 0.927, Q^2_{LOO} = 0.917, SE = 0.375

Statistical parameters such as *SE*, R^2_{train}, Q^2_{LOO}, and F_{stat} (Fischer's value) were employed to evaluate the models in those models. A good calibrating model has strong R^2, Q^2, and *F* values, as well as a low SE value with the fewest descriptors. Results of MLR-QSPR models are displayed in detail as follows (Table 2).

3.2. ANN-QSPR models

Based on the descriptors of the MLR-QSPR equation, the ANN-QSPR model is also built with the neural network technique in the following study using the "nntool" command on the Matlab system [13]. The neural network design consists of three layers: I(4)-HL(*m*)-O(1); the input layer I(4) comprises four neurons: S^6, *Dipole*, *xv*1, and N'^4; the output layer O(1) includes one neuron: logβ_{12}; and the hidden layer has *m* neurons.

The ANN-QSPR model is trained using the Levenberg-Marquardt back-propagation algorithm and two popular transfer functions as equations (7) and (8). The dataset is randomly split into three parts: training set (70 %), cross-validation set (15%), and independent test set (15%). The hyperparameters are tuned to achieve the best performance. The best model is validated using the R^2_{train}, Q^2_{CV}, and R^2_{test} metrics, and is found to have high statistical significance.

The training of ANN models is carried out in two steps. In the first step, finding the *m* value of the hidden layer (*m*) in I(4)-HL(*m*)-O(1) architecture by using the same training set of the MLR-QSPR model (Table 1). The results found ANN models as presented in Table 3.

Table 3. Initial ANN model I(4)-HL(m)-O(1) outcomes with statistical parameters

Symbol	ANN models	R^2_{train}	Q^2_{test}	Q^2_{cv}	Training algorithm	Transfer function
ANN1	I(4)-HL(8)-O(1)	0.974	0.966	0.989	BFGS 132	Log-sigmoid
ANN2	I(4)-HL(4)-O(1)	0.972	0.969	0.985	BFGS 69	Hyperbolic tangent
ANN3	I(4)-HL(10)-O(1)	0.969	0.969	0.985	BFGS 116	Log-sigmoid
ANN4	**I(4)-HL(10)-O(1)**	**0.972**	**0.975**	**0.985**	**BFGS 62**	**Hyperbolic tangent**
ANN5	I(4)-HL(4)-O(1)	0.971	0.969	0.987	BFGS 175	Log-sigmoid

In the second step, using the external data set of 16 experimental complexes (Table 4) to assess comprehensively the predictability of ANN-QSPR models through the Q^2_{EXT} value. The best ANN-QSPR model is the model in which the Q^2_{EXT} value must be greater than 0.5 and the larger the better [14]. Based on the results obtained from Figure 3, it can be seen that there are three models such as ANN1, ANN3, and ANN4 receiving values greater than 0.5. However, the ANN4 model receives the largest value (Q^2_{EXT} = 0.889). This shows that the ANN4 model has the best prediction ability compared to the remaining models.

As a whole, with the Q^2_{train} value of 0.972, the Q^2_{test} value of 0.975, and the Q^2_{CV} value of 0.985, the survey findings revealed that the ANN-QSPR model with the architecture I(4)-HL(10)-O(1) in bold, as shown in Table 3 and Figure 2, had the best predictability.

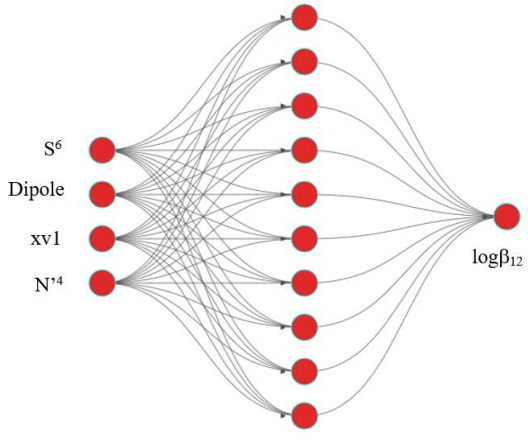


Figure 2. The architecture I(4)-HL(10)-O(1) of ANN4-QSPR model

3.3. External validation

External assessment of model prediction abilities was undertaken using 16 metal-thiosemicarbazone compounds (Table 4). According to the same equation (4), the quality of these calculations can be quantified in terms of Q^2_{EXT}. Table 4 shows the projected results that were received.

Error analysis is an essential part of QSPR research. The average absolute relative error (AARE,%), which is applied to measure the error sum of the QSPR models, is determined by equation (10) to validate the predictive ability of the built models.

$$AARE,\% = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\log\beta_{12,real} - \log\beta_{12,\mathrm{pred}}\right|}{\log\beta_{12,real}} x100 \qquad (10)$$

Here, n is the number of test substances; $\beta_{12,real}$ and $\beta_{12,pred}$ are the actual and predicted stability constants.

Table 4. Sixteen chemicals and stability constants resulting from models on EV data set

Ligand				Metal Ions	$\log\beta_{12,exp}$	Ref.	Predicted values $\log\beta_{12,pred}$ by QSPR	
R_1	R_2	R_3	R_4				MLR2	ANN4
$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Fe^{2+}	10.250	[40]	10.483	10.512
$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Co^{2+}	12.470	[40]	14.750	11.483
$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	11.680	[40]	10.568	10.484
H	H	$-C_5H_4N$	$-C_5H_4N$	Zn^{2+}	10.370	[41]	10.431	9.751
H	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	7.000	[41]	11.047	7.077
H	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	11.110	[41]	10.135	10.476
H	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	12.430	[41]	3.956	11.406
H	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Zn^{2+}	10.460	[41]	13.861	9.728
H	$-C_2H_5$	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	7.2000	[41]	13.230	6.137
H	$-C_2H_5$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	11.130	[41]	13.179	10.490
H	$-C_2H_5$	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	12.580	[41]	6.216	11.433
$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	7.760	[41]	15.887	7.068
$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	12.490	[41]	17.208	11.489
H	$-CH_3$-CH=CH_2	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	7.330	[41]	7.540	6.273
H	$-CH_3$-CH=CH_2	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	11.140	[41]	16.329	12.591
H	$-CH_3$-CH=CH_2	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	12.530	[41]	15.724	11.384

External evaluation is a critical phase in the development of regression models [2]. The evaluation process must be carried out on a separate data set [2]. This study combines external evaluation with the search for the best MLR model in Table 1 and ANN models in Table 3. In addition to the Q^2_{EXT} quantity, the evaluation method employs the additional quantity amount of AARE(%) as formula (10). When the Q^2_{EXT} value matches Tropsha's criteria ($Q^2_{EXT}>0.5$) [14], the model has strong predictive ability; if there are more than two models that meet this condition, the AARE value is used to pick the best model. Then, the better forecasting model is the model with a smaller AARE value.

Based on the results in Table 4 and Figure 3a, two models are chosen: MLR2-QSPR and ANN4-QSPR. The ANN4-QSPR model has a higher Q^2_{EXT} value (0.889) and a lower AARE

value (8.242%) than the MLR2-QSPR model (0.815 and 27.616%, respectively). Therefore, The ANN4-QSPR model demonstrates a higher degree of accuracy in its predictions than the MLR2-QSPR model. Additionally, the $\log\beta_{12}$ estimated values by the ANN4-QSPR model are closer to the experimental values.

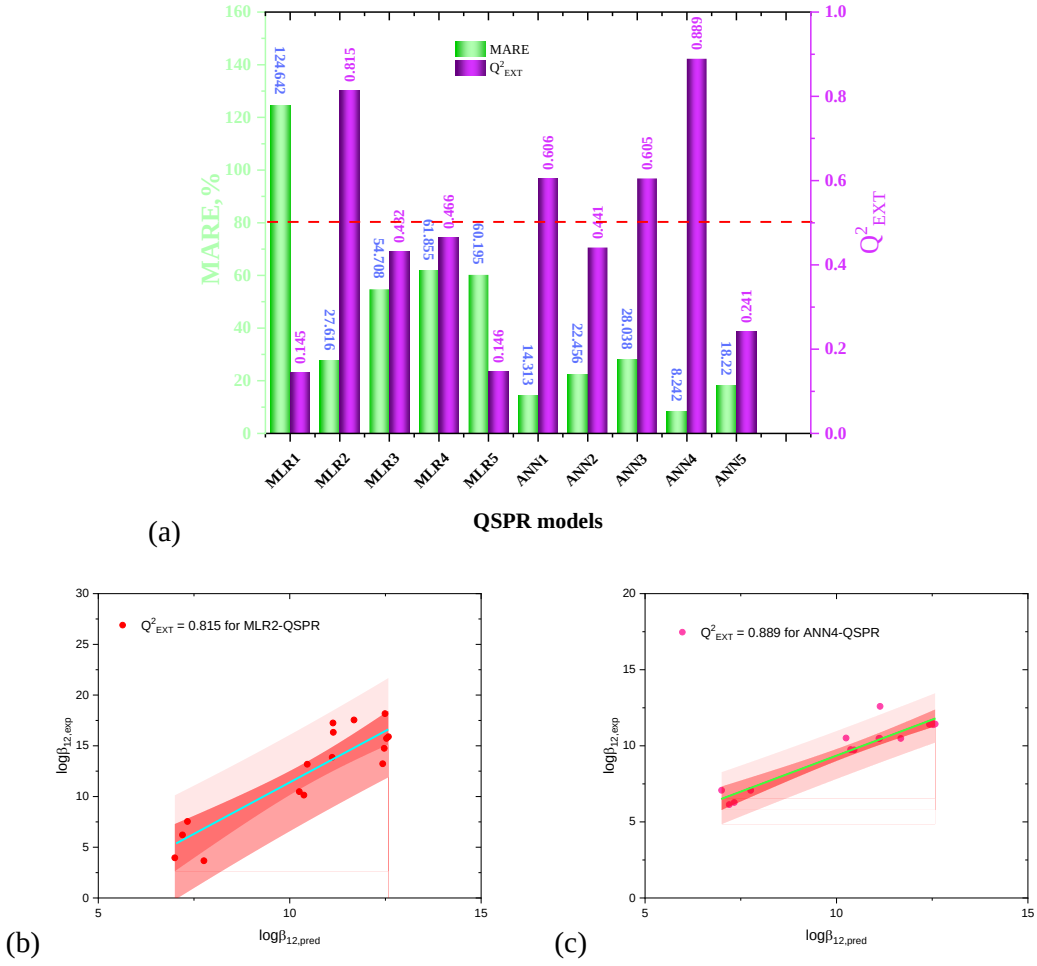


Figure 3. (a) The MARE(%) and Q^2_{EXT} values of QSPR models. (b)&(c) Relationships between experimental vs. estimated values of EV dataset from MLR2-QSPR and ANN4-QSPR models

The single-factor ANOVA approach is also utilized to analyze the disagreement between the estimated values ($\log\beta_{12,cal}$) and the initial values ($\log\beta_{12,exp}$) of both models on the EV dataset. The findings show that the difference between these outcomes is minor ($F = 3.6573 < F_{0.05} = 4.1709$).

3.4. Discovery of new complexes

Phenothiazine and carbazole derivatives (Figure 4a-b) have strong biological activity, as do thiosemicarbazone and its complexes [15, 16]. The work previously created thiosemicarbazone derivatives including phenothiazine, carbazole, and related constituents [15, 16]. Fourteen new thiosemicarbazone ligands with substituted derivatives at the R_4 site were developed in this study. The hydrogen atoms are substituted at the ligand sites R_1, R_2, and R_3 (Figure 4c).

The phenothiazine and carbazole group design principle are based on the generated model descriptors such as S^6, *Dipole*, *xv1*, and N'^4, and the derivatives were examined as a result of

the predicted model findings. The building blocks for the 20 new complexes are novel-designed thiosemicarbazones including metal ions such as Ag^{+}, Cu^{2+}, Zn^{2+}, Ni^{2+}, and Cd^{2+}.

The data set for the new complexes was created using similar calculations to those used to develop the complexes of the training and external data sets. The complexation is determined by the total energy values and structural morphology using PM7 and PM7/sparkle semi-experimental quantum calculations on MoPac2016. Table 5 shows the projected $\log\beta_{12,pred\text{-}new}$ values for the novel complexes.

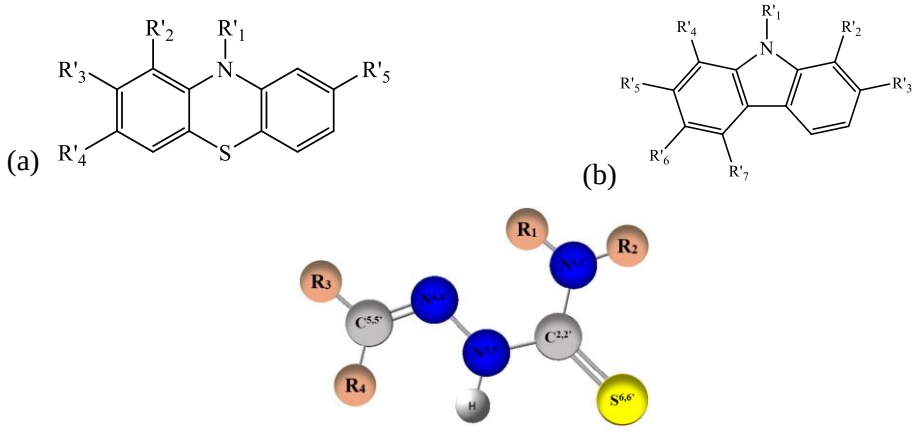
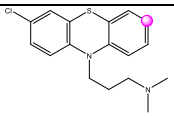
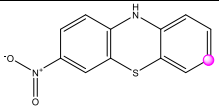
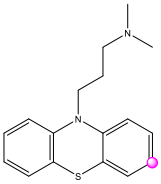
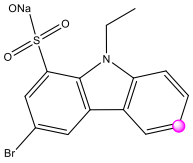
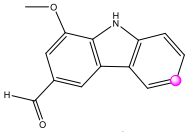
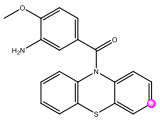
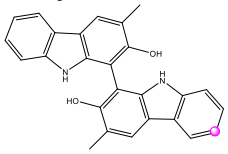
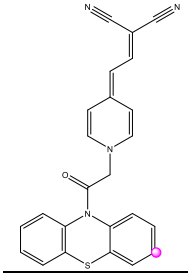
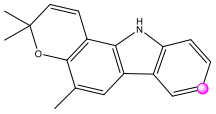
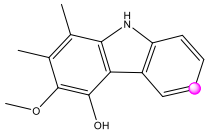
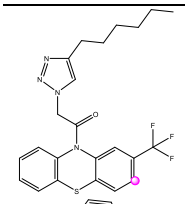
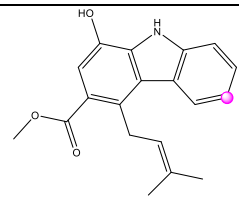
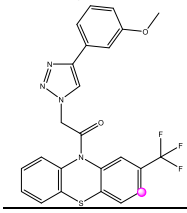
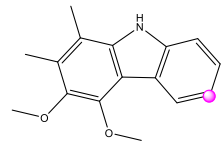


Figure 4. The structure of phenothiazine (a) and carbazole (b) derivatives and structures of ligand (c).

Table 5. Twenty new complexes with the calculated $\log\beta_{12,pred\text{-}new}$ values from the QSPR models

R_4 site	Metal ions	$\log\beta_{12,pred\text{-}new}$ MLR2	$\log\beta_{12,pred\text{-}new}$ ANN4	R_4 site	Metal ions	$\log\beta_{12,pred\text{-}new}$ MLR2	$\log\beta_{12,pred\text{-}new}$ ANN4
	Zn^{2+}	12.803	13.361		Cu^{2+}	13.201	11.534
	Cu^{2+}	15.749	16.602		Ag^{+}	16.479	17.831
	Zn^{2+}	11.968	12.091		Ni^{2+}	15.960	17.787
	Zn^{2+}	15.729	16.221		Ag^{+}	5.750	7.146
	Ag^{+}	9.177	8.136		Cu^{2+}	9.728	10.234
	Cd^{2+}	12.849	12.413		Zn^{2+}	10.598	9.630
	Cu^{2+}	15.437	14.265		Ni^{2+}	11.619	12.391
	Zn^{2+}	11.658	11.243		Zn^{2+}	12.353	13.248

	Ag^+	8.170	7.681		Zn^{2+}	14.776	15.589
	Ag^+	5.754	6.881		Zn^{2+}	10.952	9.327

The new complexes were evaluated for the application domain (AD) (|D-Cook| < 1.0) [19,20] and outliers by integrating the data set for the new complexes into the initial training data set and computing the Cook distance indicator (D-Cook). The results showed that the Cook distance values of 20 new complexes met the prediction criteria. Furthermore, when the projected $\log\beta_{12,\text{pred-new}}$ values from the MLR2-QSPR and ANN4-QSPR models were evaluated using the one-way ANOVA technique ($F = 0.0189 < F_{0.05} = 4.0915$), there was no difference between them.

4. CONCLUSION

In this investigation, new quantitative structure-property relationship (QSPR) models were developed to predict the stability constants of metal-thiosemicarbazone complexes. Two types of machine learning models were used: multivariate linear regression (MLR) and artificial neural networks (ANN). A library of molecular descriptors was generated using semi-empirical quantum calculations (PM7 and PM7/sparkle) on the optimized geometries of the complexes. The geometries were optimized using the QSARIS system and Mopac2016 software. The MLR and ANN QSPR models were statistically robust and had low prediction errors. Their performance was thoroughly evaluated using the leave-one-out method, which yielded good statistical values for Q^2_{LOO}, MARE(%), and the one-way ANOVA method. The generated QSPR models showed a high correlation coefficient for predicting the stability constants of the compounds, as well as meeting Tropsha's criteria. This knowledge can be used to develop new thiosemicarbazone derivatives with improved properties.

Conflict of Interest: We have no conflict of interest.

REFERENCES

1. Lobana T.S., Sharma R., Bawa G., and Khanna S. – Bonding and structure trends of thiosemicarbazone derivatives of metals - An overview. Coordination Chemistry Reviews **253** (7-8) (2009) 977-1055. https://doi.org/10.1016/j.ccr.2008.07.004
2. Kumar S, Dhar D.N, and Saxena P.N. – Applications of metal complexes of Schiff bases-A review. J. Sci. Ind. Res. **68** (2009) 181-187.
3. Kunal R., Supratik K., and Rudra N.D. – A Primer on QSAR/QSPR Modeling. Fundamental Concepts. Springer (2015).
4. Stewart J. J. P. – Optimization of parameters for semi-empirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. J. Mol Model **19** (2013) 1-32. https://doi.org/10.1007/s00894-012-1667-x

5. Harvey D. – Modern analytical Chemistry. V. Mc. Graw Hill Boston Toronto (2000).
6. Garg B. S., Saxena V., and R. Dixit. – Evaluation of thermodynamic functions and stability constants of lanthanon (III) complexes with 1,2- naphthoquinone-2-thiosemicarbazone-4-sulphonic acid (sodium salt) (NQTS.4S) from potentiometric data. Thermochimica Acta. **195** (1992) 169-175. https://doi.org/10.1016/0040-6031(92)80060-A
7. BIOVA Draw 2017 R2. Version: 17.2.NET. Dassault Systèmes. France (2016).
8. Stewart J. J. P. MOPAC2016, Version: 17.240W. Stewart Computational Chemistry. USA (2002).
9. QSARIS 1.1. Statistical Solutions Ltd. USA (2001).
10. Steppan D. D., Werner J., and Yeater P. R. – Essential Regression and Experimental Design for Chemists and Engineers. Germany (1998).
11. Tat P. V. – Development of QSAR, Publisher of Natural sciences and Technique, Ha Noi (2009).
12. Gasteiger J., and Zupan J. – Neural Networks in Chemistry. Chiw. Inr. Ed. EngI. **32** (1993) 503-521.
13. Matlab R2016a. MathWorks. USA (2016).
14. Golbraikh A., and Tropsha A. – Beware of Q2. J. Mol. Graphics Model. **20** (2002) 269-276. https://doi.org/10.1016/S1093-3263(01)00123-1
15. Sudeshna G., and Parimal K. – Multiple non-psychiatric effects of phenothiazines: A review. European Journal of Pharmacology **648** (1-3) (2010) 6-14. https://doi.org/10.1016/j.ejphar.2010.08.045
16. Al-Busaidi I. J., Haque A., Al Rasbi N. K., and Khan M. S. – Phenothiazine-based derivatives for optoelectronic applications: A review. Synthetic Metals **257** (2019) 116189. https://doi.org/10.1016/j.synthmet.2019.116189
17. Gaál A., Orgován G., Polgári Z., Réti A., Mihucz V. G., Bősze S., Szoboszlai N., and Streli C. – Complex forming competition and in-vitro toxicity studies on the applicability of di-2-pyridylketone-4,4,-dimethyl-3-thiosemicarbazone (Dp44mT) as a metal chelator., J. Inorg. Biochem. **130** (2014) 52–58. https://doi.org/10.1016/j.jinorgbio.2013.09.016
18. Bernhardt P. V., Sharpe P. C., Islam M., Lovejoy D. B., Kalinowski D. S., Richardson D. R. – Iron Chelators of the Dipyridylketone Thiosemicarbazone Class: Precomplexation and Transmetalation Effects on Anticancer Activity., J. Med. Chem. **52** (2) (2009) 407-415. https://doi.org/10.1021/jm801012z
19. Organisation for Economic Co-operation and Development (OECD). – Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships Models (2007).
20. Sahigara F., Mansouri K., Ballabio D., Mauri, Consonni, and Todeschini R. – Comparison of different approaches to define the applicability domain of QSAR models. Molecules **17** (2012) 4791-4810. https://doi.org/10.3390/molecules17054791

TÓM TẮT

KHÁM PHÁ CÁC PHỨC CHẤT GIỮA ION KIM LOẠI VÀ THIOSEMICARBAZONE SỬ DỤNG CÁC MÔ HÌNH *IN SILICO* ỨNG DỤNG TRONG PHÂN TÍCH MÔI TRƯỜNG

Huỳnh Ngọc Châu, Cù Huy Đức, Nguyễn Minh Quang*
Trường Đại học Công nghiệp Thành phố Hồ Chí Minh
*Email: *nguyenminhquang@iuh.edu.vn*

Trong nghiên cứu này, hai mươi phức chất giữa ion kim loại và phối tử thiosemicarbazone mới được khám phá bởi các kỹ thuật *in silico*. Các hằng số bền ($\log\beta_{12}$) của các phức cũng được dự đoán bằng cách sử dụng các mô hình quan hệ định lượng cấu trúc và tính chất (QSPR). Các mô hình này được tạo ra bằng cách sử dụng phương pháp hồi quy tuyến tính đa biến (MLR) và mạng thần kinh nhân tạo (ANN). Đặc tính cấu trúc của các phức chất bao gồm các thuộc tính phân tử và lượng tử. Các công trình nghiên cứu thực nghiệm đã xuất bản được sử dụng để thu thập các hằng số bền với các thông số thực nghiệm. Mô hình tốt nhất, MLR2-QSPR (k = 4), được tạo thành từ các mô tả phân tử như S^6, *Dipole, xv1, và* N^{-4}. Các số liệu thống kê như R^2_{train} = 0,913; Q^2_{LOO} = 0,903 và SE = 0,408 đã được sử dụng để xác nhận chất lượng của MLR-QSPR này. Dữ liệu thống kê cho mô hình ANN4-QSPR I(4)-HL(10)-O(1) cũng được tìm thấy, đó là R^2_{train} = 0,972; Q^2_{test} = 0,975 và R^2_{CV} = 0,985. Bên cạnh đó, nghiên cứu sử dụng kết quả của các biến từ các mô hình QSPR để phát triển các dẫn xuất thiosemicarbazone mới và các phức chất từ các phối tử này. Kết quả các phức chất giữa ion kim loại và thiosemicarbazone mới đã được phát triển và các hằng số bền đã được dự đoán bởi hai mô hình QSPR. Kết quả thu được từ các mô hình có thể được áp dụng để phát triển các hóa chất mới có thể sử dụng trong hóa học phân tích và giám sát đánh giá môi trường.

Từ khóa: ANN, Hằng số bền $\log\beta_{12}$, MLR, QSPR, Thiosemicarbazone.