DEVELOPMENT OF QSPR MODELS FOR NOVEL THIOSEMICARBAZONES AS POTENTIAL COMPLEXING LIGANDS FOR METAL ION ANALYSIS

Received 15-05-2024

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

1. Faculty of Chemical Engineering, Industrial University of Ho Chi Minh City, Ho Chi Minh City

2. Institute of Pharmaceutical Education and Research, Binh Duong University, Binh Duong

*Email: nguyenminhquang@iuh.edu.vn

TÓM TẮT

PHÁT TRIỂN CÁC DẪN XUẤT THIOSEMICARBAZONE TIỀM NĂNG TRONG VIỆC PHÂN TÍCH CÁC ION KIM LOẠI DỰA TRÊN MÔ HÌNH QSPR

Trong nghiên cứu này, hai mươi phức chất mới của các phối tử mới thiosemicarbazone đã được phát triển dựa trên các mô hình định lượng quan hệ cấu trúc-tính chất (QSPR). Các mô hình QSPR được xây dựng từ 87 giá trị hằng số bền ($\log\beta_{12}$) của các phức chất thực nghiệm bằng cách hai phương pháp hồi quy tuyến tính đa biến (MLR-QSPR) và mạng thần kinh nhân tạo (ANN-QSPR). Ngoài ra, các mô hình đã được đánh giá ngoại trên một tập dữ liệu độc lập bao gồm 18 phức chất thực nghiệm. Cấu trúc của các phức chất đã được tối ưu hóa bằng kỹ thuật tính toán lượng tử bán thực nghiệm với phiên bản mới nhất PM7 và PM7/sparkle. Nghiên cứu này cũng áp dụng nguyên tắc "miền ứng dụng (AD) và giá trị ngoại biên (Outliers)" theo hướng dẫn của tổ chức OECD để thiết kế các ligand và phức hợp mới. Kết quả mô hình MLR-QSPR bao gồm sáu biến như ^6S, xvp5, xvp7, surface, xvch9 và k0 đã được xây dựng thành công với các giá trị thống kê như R^2_{train} = 0,914; Q^2_{LOO} = 0,844; SE = 0,413 và $Q^2_{ext\text{-}MLR}$ = 0,767. Hơn nữa, mô hình ANN-QSPR với kiến trúc I(6)-HL(5)-O(1) đã được tìm thấy với các giá trị thống kê: R^2_{train} = 0,978; Q^2_{CV} = 0,975; Q^2_{test} = 0,976 và $Q^2_{ext\text{-}ANN}$ = 0,867. Kết quả thu được từ các mô hình QSPR được sử dụng để phát triển các phức chất mới và các phức chất mới này có thể được ứng dụng trong lĩnh vực hóa học phân tích.

Keywords. *MLR, MLP-ANN, QSPR, Stability constant $\log\beta_{12}$, Thiosemicarbazone.*

1. INTRODUCTION

The structural features of thiosemicarbazone, which include thiol sulphur and azomethine nitrogen, allow it to form complexes with a variety of transition metal ions [1]. Thiosemicarbazones and their metal chelates are thus the subject of extensive research worldwide. Their application as analytical agents, antibacterial, antifungal, and anticancer agents are the main uses [2]. Stabilityconstants should be a major concern of the formation of complexes. It is closely related to the structural characteristics of the ligands and the metal ions. It is an indicator of how well the ligand interacts with the metal ions to

create various complexes. On the other hand, computer science has advanced, it has become very common to conduct theoretical research on applying computational chemistry in conjunction with chemical software to solve complex mathematical problems and appropriate mathematical methods in order to quickly orient experimental research [3]. Computational chemistry applications significantly reduce the time and cost associated with complexation research. Chemists can utilize computers to simulate and predict the properties of complexes rapidly and efficiently, eliminating the need for time-consuming and expensive experiments.

The environment is becoming more and more contaminated nowadays as a result of the leakage of heavy metal ions into the environment from factories, industrial parks, and production facilities. To achieve the practical requirements, heavy metal ion control and analysis must be quick and affordable. AAS, ICP-AES, ICP-MS have all been used to analyze the presence of heavy metals in various parts of the world. However, the methods are quite expensive and strictly operating techniques. Meanwhile, spectrophotometry is quite popular, cheap, and easy to operate. It relies on the ability of metal ions and ligands to form complexes. Therefore, developing a new ligand and complex is a way to quantify metal ions that can be easily applied in practice [4]. Because of its high complexity and the numerous research that have been published on its use, we seek to characterize the thiosemicarbazone derivative in this study using a straightforward and affordable method called spectrophotometric analysis.

Table 1. Experimentally determined stability constants for 87 metal-thiosemicarbazone complexes (n) with minimum ($\log\beta_{12,min}$) and maximum ($\log\beta_{12,max}$) values

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	$-CH_3$	$-C_{10}H_{12}BrO$	Ni^{2+}	1	6.5490	6.5490	[8]
H	H	$-C_2H_5$	$-C_{10}H_{12}BrO_2$	Ni^{2+}	1	8.8070	8.8070	[9]
H	H	$-C_2H_5$	$-C_{10}H_{12}BrO_2$	Cd^{2+}	1	6.9274	6.9274	[10]
H	H	$-C_4H_9$	$-C_6H_4BrO_2$	Co^{2+}	1	10.4362	10.4362	[11]
H	$-C_6H_5$	$-C_6H_5$	$-C_5H_4N$	Cu^{2+}	16	11.0010	11.3050	[12]
H	$-C_6H_5$	$-C_6H_5$	$-C_5H_4N$	Co^{2+}	16	10.3280	10.4710	[12]
H	$-C_6H_5$	$-C_6H_5$	$-C_5H_4N$	Zn^{2+}	16	11.0610	11.2610	[12]
H	$-C_6H_5$	$-C_6H_5$	$-C_5H_4N$	Ni^{2+}	16	10.1230	10.2540	[12]
$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	1	11.2100	11.2100	[13]
$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Fe^{3+}	1	12.6700	12.6700	[14]
H	$-C_2H_5$	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	1	7.2000	7.2000	[13]
H	$-C_2H_5$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	1	11.1300	11.1300	[13]
$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	1	7.7600	7.7600	[13]
H	$-C_6H_5$	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	1	7.4100	7.4100	[13]
H	$-C_6H_5$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	1	11.3200	11.3200	[13]
H	$-C_6H_5$	$-C_5H_4N$	$-C_5H_4N$	Zn^{2+}	1	10.2100	10.2100	[13]
H	$-C_4H_7$	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	1	7.3300	7.3300	[13]
H	$-C_4H_7$	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	1	12.5300	12.5300	[13]
H	$-C_4H_7$	$-C_5H_4N$	$-C_5H_4N$	Zn^{2+}	1	10.2200	10.2200	[13]
H	H	-	$-C_{10}H_7O_4S$	Pr^{3+}	7	6.8500	7.9300	[15]
H	H	$-C_6H_5O$	$-C_6H_5O$	Cu^{2+}	1	11.3570	11.3570	[16]

Additionally, the quantitative structure-property relationship (QSPR) modelling is widely employed in the field of chemistry [3]. Through experiments and knowledge of the molecular structure, this theoretical approach creates prediction models based on a method known as quantitative structure and property relationship. The outcomes of QSPR modelling offer numerous benefits. In line with the present trend, designing chemical processes towards green chemistry aids in the development of safe, efficient, and environmentally friendly compounds [3].

In this study, we build models based on the molecular descriptors and stability constants ($\log\beta_{12}$) of complexes between metal ions and thiosemicarbazones using an *in silico* model of quantitative structure and property relationship. The electronic parameters are used in quantum chemistry with the semi-empirical approach of PM7 and PM7/sparkle new version [5-6], and the 0D-3D topological descriptors are generated using graph theoretical methods [7]. Multivariate linear regression and artificial neural networks are used to create the MLR-QSPR and ANN-QSPR model series. With the test set, the models are also subjected to external evaluation methods. We also predict the stability constants $\log\beta_{12}$ values of newly constructed complexes using the developed models.

2. METHODOLOGY

2.1. Structural selection and data set

Thiosemicarbazones often interact with transition metal ions as bidentate ligands through bonds with the sulphur and azomethine nitrogen atoms. They can sometimes link through the sulphur, nitrogen, and potential donor atoms to function as tridentate or tetradentate ligands [1,2]. Thiosemicarbazone complexes are chosen for research in this work as bidentate ligands in association with the basic metal ions. The detailed structure of ligands and complexes are shown in Fig. 1; in which, the substituents of R_1, R_2, R_3 and R_4 can change in structure to form different ligands and complexes.

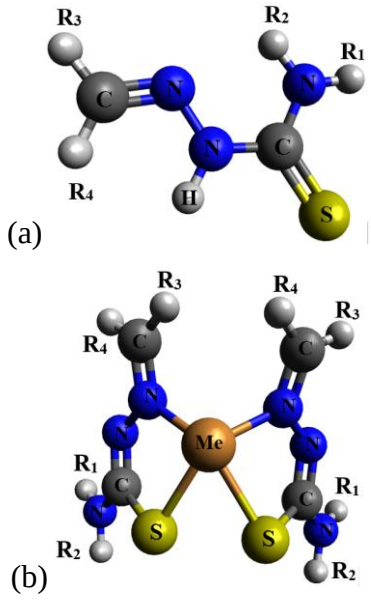


Figure 1. Structure of ligand (a) and complex (b) in the work

According to the general equilibrium equation, the complex formation in an aqueous solution can be explained as follows:

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

The following formula is used to compute the overall stability constants (β_{12}) when the complex ML_2 is formed in two phases with p = 1 and q = 2:

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

From the published literature [8-16], the stability constants ($\log\beta_{12}$) of the complexes ML_2 of some metal ions (M = Ni^{2+}, Cd^{2+}, Cu^{2+}, Co^{2+}, Mn^{2+}, Zn^{2+}, Fe^{3+}, and Pr^{3+}) with various thiosemicarbazones (L) in aqueous solution at various

experimental conditions (temperature, pH, and ionic strength) were selected.**2.2. Descriptors**

Quantum, 0D-3D molecular descriptors are all used in QSPR modelling. Using BIOVIA Draw 2017 R2, the structures of experimental complexes were recreated [17]. The complexes were then optimized using the MoPac2016 software with the method of quantum mechanics QM [6]. On QSARIS tools [7,18], the 0D-3D molecular parameters of 230 descriptors were calculated using the ideal structures. The MoPac2016 system was used to determine the quantum parameters using the semi-empirical quantum method PM7 and PM7/sparkle [5]. The QSPR models were constructed using the descriptors and stability constants ($\log\beta_{12}$) as a data set [3].

2.3. Estimation of QSPR models

2.3.1. Multivariate linear regression

The statistical technique of multivariate linear regression (MLR) is a well-liked one for QSPR modelling. The equation (3) describes the relationship between independent and dependent variables [3,19].

$$Y = \sum_{i=1}^{k} b_0 + b_i x_i \quad (3)$$

where Y is the dependent variable, b_0 represents the intercept of models, b_i represents a slope, x_i represents an independent variable, and k represents the number of variables in the equation. In this case, Y is the logarithm of the stability constant and x_i is the descriptor.

In this study, MLR is used to establish a connection between the descriptor parameters and the log-transformed stability constants ($\log\beta_{12}$). The data set must be randomly divided into a training set and a test set, with the training set containing roughly 80% of the total data set, in order to generate the MLR-QSPR models. The statistical parameters such as k number of molecular descriptors, R^2_{train}, Q^2_{LOO}, *SE*, F_{stat} (Fischer's value), and *PRESS* were utilized in these MLR-QSPR models to assess the fitness of models [3,19].

2.3.2. Artificial neural network

In specialized domains of mathematics, electrical and electronic research, medicine, chemistry, and several other real-world applications, artificial neural networks (ANN) are successfully used [20]. Generally, an input layer, one or more hidden layers, and an output layer are typically included in an ANN model. Each layer contains neurons as well as the weights that link them to one another. There are numerous alternative ANN architectures utilized in studies for various purposes, but multi-layer perceptron (MLP) networks are the most widely used for building models [20-21].

The MLP-ANN type is used with an error back-propagation technique in this work. There are three layers including the input layer I(k), the hidden layer H(m) and the output layer O(n) in the ANN architecture I(k)-HL(m)-O(n). The variables of MLR-QSPR models make up the input layer. The stability constant $\log\beta_{12}$, which is to be anticipated, is a component of the output layer. The input and output layer neurons calculate the hidden neurons. ANN-QSPR models are trained using the hyperbolic sigmoid tangent, log-sigmoid and exponential transfer function [21].

2.4. Validation of QSPR models

It is crucial to use validation techniques to strengthen a prediction of models when analyzing data sets and to help determine the number of the regression equation of test data. In this study, we employ data sets for both internal and external

methods that aid in the development of ideal QSPR models. To make sure that the models are good, statistical indicators are employed. For the training set, they contain the values R^2_{train} and Q^2_{LOO}, and for the external-validation set, Q^2_{ext} [3,19]. The Q^2_{ext} value should be greater than 0.5 and the R^2_{train} and Q^2_{LOO} values should be greater than 0.6. Additionally, the gap between the R^2_{train} and Q^2_{LOO} values should not be greater than 0.3 [3,22]. The same formula (4) was used to generate these results.

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

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

Additionally, fundamental statistics like *PRESS*, F_{stat}, and *SE* (standard error) are used to assess the MLR-QSPR models. In the case of ANN-QSPR models, modelling is carried out until the mean square error (*MSE*) is reduced, and then the network output is compared to the real values of the output acquired from experimental results [3,20]. Between the network outputs (o) and the target outputs (t), MSE is the average squared error. The following is written [21]:

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

We employed the single factor ANOVA to assess the differences between the experimental and the anticipated $\log\beta_{12}$ values from the models. Besides, the average absolute values of the relative error, MARE(%) are used to evaluate the overall error of the QSPR models. Error analysis is also an important component of QSPR study to evaluate the predictive performance of the created models. It is determined using the formula (6) [18]:

$$MARE, \% = \frac{1}{n}\frac{\left|\log\beta_{12,real} - \log\beta_{12,\text{pred}}\right|}{\log\beta_{12,real}} 100 \quad (6)$$

here n is the number of test substances; $\beta_{12,real}$ and $\beta_{12,pred}$ are the experimental and calculated stability constants.

3. RESULTS AND DISCUSSION

3.1. MLR models

The MLR-QSPR models are based on statistical mathematics, which is supported by software tools, and careful evaluation of statistical indices. First, we screened the variables in the models using XLSTAT2016 [23]; Later, utilizing the variable group, we created MLR-QSPR models by using back-elimination and forward regression on Excel platform with add-in Regression [19]. The leave-one-out (LOO) method was used for the cross-validation (CV) of the MLR-QSPR models. The construction of the 11 MLR-QSPR models is shown in Table 2. The difference between the 11 models is a few different variables, but the regression indices are still satisfactory. It is explained that the training data set represents structural diversification in building the QSPR model. This will facilitate the selection of models and variables for developing new ligands and complexes.

As a results, the study verified the 11 MLR-QSPR models as being optimum and statistically significant based on the aforementioned analyses [3,19]. The outcomes demonstrated the high R^2_{train}, Q^2_{LOO}, and F_{stat} values and the extremely low *SE* and *PRESS* values of these six-variable MLR-QSPR models, which are very good quality models. To develop complete models, however, the external evaluation of the MLR-QSPR models is required [3,22]. The next step of this

investigation will be completed and presented this validation.

3.2. ANN models

The six variables that make up the input values of ANN models are the outcomes of the construction of MLR-QSPR models. Table 2 provides an overview results of the built models from the training set of 87 $\log\beta_{12}$ values. The Matlab program is used to advance ANN-QSPR models [21]. The dataset is divided into three sections at random. These are divided as follows: 60% for the training set; 20% for cross-validation; and the remaining 20% for the test set.

Table 2. Statistical values of MLR-QSPR models and the best models are shown in bold.

No	Symbol	The models
1	MLR-QSPR1	$\log\beta_{12}$ = 2.000 – 2.528**xvp9* + 0.457**xvp5* + 0.036**SaaCH* + 2.328**Ionization potential* + 2.255**xvp10* + 2.023**HOMO*. R^2_{train} = 0.928, R^2_{adj} = 0.923, Q^2_{LOO} = 0.887, SE = 0.377, F_{stat} = 172.54, PRESS = 17.94
2	MLR-QSPR2	$\log\beta_{12}$ = -3.183 – 0.233**logP* – 0.898**xvp9* + 0.468**xvp5* + 0.077**SaaCH* + 0.313* *Ionization potential* + 0.239**SdsN*. R^2_{train} = 0.920, R^2_{adj} = 0.914, Q^2_{LOO} = 0.842, SE = 0.399, F_{stat} = 152.53, PRESS = 25.03
3	MLR-QSPR3	$\log\beta_{12}$ = 18.600 – 0.457**LOMO* + 0.709**xvp5* – 0.880**xvp7* + 0.053**SpcPolarizability* – 2.722**S'6* – 1.621**Hmin*. R^2_{train} = 0.915, R^2_{adj} = 0.909, Q^2_{LOO} = 0.869, SE = 0.410, F_{stat} = 144.03, PRESS = 20.85
4	MLR-QSPR4	$\log\beta_{12}$ = 3.218 – 0.042**Cosmo volume* – 1.210**ka1* + 1.411**xvp5* – 1.616**xvp6* + 114.37**xvch9* + 80.45**xch8*. R^2_{train} = 0.921, R^2_{adj} = 0.915, Q^2_{LOO} = 0.885, SE = 0.396, F_{stat} = 155.83, PRESS = 18.19
5	MLR-QSPR5	$\log\beta_{12}$ = -3.788 – 1.342**knotpv* – 0.994**SaasC* – 0.427**xvp9* + 35.992**xvch6* – 2.011**xS'6* – 12.092**N3*. R^2_{train} = 0.911, R^2_{adj} = 0.905, Q^2_{LOO} = 0.852, SE = 0.420, F_{stat} = 136.78, PRESS = 23.46
6	MLR-QSPR6	$\log\beta_{12}$ = -0.478 – 31.827**N3* + 4.943**Hmin* – 72.381**xch8* – 0.012**Molecular Weight* + 0.296**xvp5* – 1.184**xvpc4*. R^2_{train} = 0.922, R^2_{adj} = 0.916, Q^2_{LOO} = 0.841, SE = 0.394, F_{stat} = 157.56, PRESS = 25.25
7	MLR-QSPR7	$\log\beta_{12}$ = -9.003 – 0.599**logP* + 0.842**xvp5* – 0.983**xvp7* + 0.026**Surface* + 0.347**SdsN* – 1.511**xpc4*. R^2_{train} = 0.914, R^2_{adj} = 0.908, Q^2_{LOO} = 0.827, SE = 0.413, F_{stat} = 142.17, PRESS = 27.49
8	**MLR-QSPR8**	**$\log\beta_{12}$ = -15.223 – 1.534**S6* + 0.876**xvp5* – 1.240**xvp7* + 0.041**Surface* + 100.286**xvch9* – 0.057**k0*. R^2_{train} = 0.914, R^2_{adj} = 0.908, Q^2_{LOO} = 0.844, SE = 0.413, F_{stat} = 141.75, PRESS = 24.82**
9	MLR-QSPR9	$\log\beta_{12}$ = -1.029 + 0.0001**Core-Core Repulsion* – 3.062**knotpv* – 1.271**xvp9* + 1.055**Gmax* + 0.042**SpcPolarization* + 3.217**N'4*. R^2_{train} = 0.906, R^2_{adj} = 0.899, Q^2_{LOO} = 0.849, SE = 0.432, F_{stat} = 128.48, PRESS = 23.97
10	MLR-QSPR10	$\log\beta_{12}$ = -2.015 – 0.593**Me7* – 3.241**S'6* + 0.432**xvp5* – 0.614**xvp8* + 0.022* *Volume* – 0.007**Molecular Weight*. R^2_{train} = 0.918, R^2_{adj} = 0.911, Q^2_{LOO} = 0.884, SE = 0.405, F_{stat} = 148.31, PRESS = 18.44
11	MLR-QSPR11	$\log\beta_{12}$ = -6.761 + 0.009*ΔH_f + 1.009**xvp5* - 1.060**xvp6* + 0.590**SdssC* + 0.691**Gmax* – 4.700**N'3*. R^2_{train} = 0.908, R^2_{adj} = 0.902, Q^2_{LOO} = 0.858, SE = 0.426, F_{stat} = 132.24, PRESS = 22.49

Utilizing the Levenberg-Marquardt optimization technique and the back-propagation error approach, the ANN architectures I(6)-HL(*m*)-O(1) are created. The ANN-QSPR models are trained using the log-sigmoid, exponential and

hyperbolic tangent transfer function [21]. As shown in Table 3, the *m* neurons in hidden layer HL(*m*) are initially assessed.

On an external data set, the predictability of ANN-QSPR models is comprehensively assessed. With the Q^2_{train} value of 0.978, the Q^2_{test} value of 0.976 and the Q^2_{CV} value of 0.975, the survey findings revealed that the ANN-QSPR model with the architecture I(6)-HL(7)-O(1) in bold, as shown in Table 3 and Figure 2, had the best predictability.

Table 3. The trained ANN models I(6)-HL(m)-O(1) with statistical parameters

Code	ANN models	R^2_{train}	Q^2_{test}	Q^2_{cv}	Training algorithm	Transfer function
ANN1	I(6)-HL(6)-O(1)	0.976	0.985	0.964	BFGS 33	Hyperbolic tangent
ANN2	I(6)-HL(8)-O(1)	0.958	0.985	0.976	BFGS 31	Log-sigmoid
ANN3	**I(6)-HL(7)-O(1)**	0.978	0.976	0.975	BFGS 13	**Exponential**
ANN4	I(6)-HL(7)-O(1)	0.886	0.926	0.949	BFGS 12	Log-sigmoid
ANN5	I(6)-HL(7)-O(1)	0.986	0.998	0.943	BFGS 63	Hyperbolic tangent

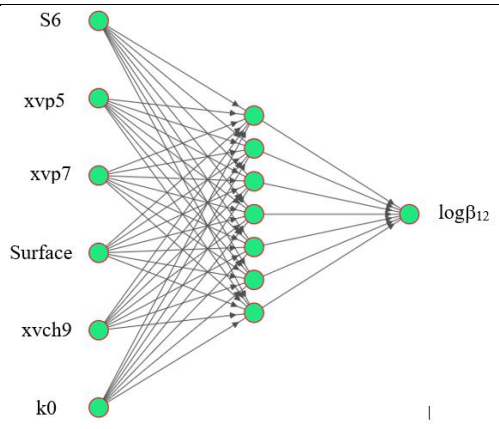


Figure 2. The architecture I(6)-HL(7)-O(1) of ANN model

Through the parameter Q^2_{ext} (>0.5) [22] and the MARE(%) values as the formula (6), the assessment outcomes are strictly monitored and are both to select the best MLR-QSPR model of 11 models in Table 2 and to train also the best ANN-QSPR model of 5 models in Table 3. The models that satisfied the statistical requirements were found in the findings (shown in bold in Table 2 and 3) [3,22,24]. Results of calculation of the Q^2_{ext} and *MARE*(%) values have been calculated as Figure 3. Based on the calculated values, it is possible to assert that the MLR-QSPR8 and ANN3-QSPR models were the good ones selected in this study, corresponding to the Q^2_{ext} values of 0.767 and 0.793, respectively. The values show the predictive power of the models is acceptable (> 0.5) [22].As another issue, the *MARE*(%) values of MLR-QSPR8 and ANN3-QSPR models are 8.772 and 7.136, respectively. The results indicate that the ANN3-QSPR model have the better predictive power than the MLR-QSPR8 model. It means that the models built on intelligent machine-learning techniques meet real-world applications.

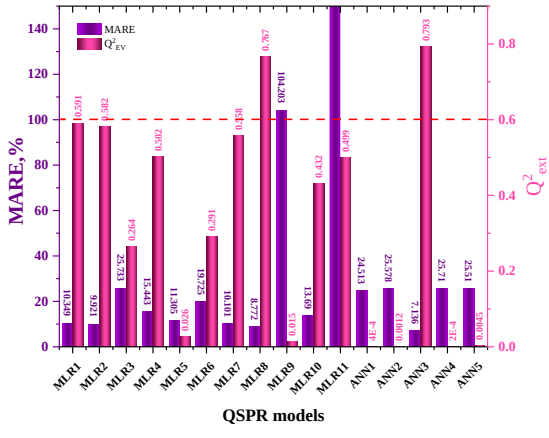


Figure 3. The MARE(%) and Q^2_{ext} values of QSPR models

Table 4. The 18 experimental $\log\beta_{12,exp}$ values in the external data set

Code	Thiosemicarbazone				Metal ions	$\log\beta_{12.exp}$	Ref.
	R_1	R_2	R_3	R_4			
thio1	H	H	$-C_2H_5$	$-C_{10}H_{12}BrO_2$	Fe^{2+}	7.086	[25]
thio2	$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Fe^{3+}	10.250	[14]
thio3	$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	11.680	[14]
thio4	$-CH_3$	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	12.490	[13]
thio5	H	H	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	11.290	[13]
thio6	H	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Mn^{2+}	7.0000	[13]
thio7	H	$-CH_3$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	11.110	[13]
thio8	H	$-C_2H_5$	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	12.580	[13]
thio9	H	$-C_2H_5$	$-C_5H_4N$	$-C_5H_4N$	Zn^{2+}	10.270	[13]
thio10	H	$-C_6H_5$	$-C_5H_4N$	$-C_5H_4N$	Cu^{2+}	12.410	[13]
thio11	H	$-C_4H_7$	$-C_5H_4N$	$-C_5H_4N$	Ni^{2+}	11.140	[13]
thio12	H	H	-	$-C_{10}H_7O_4S$	La^{3+}	6.9000	[15]
thio13	H	H	-	$-C_{10}H_7O$	Pr^{3+}	12.940	[26]
thio14	H	H	-	$-C_{10}H_7O$	Nd^{3+}	13.500	[26]
thio15	H	H	$-CH_3$	$-C_6H_5O$	Cd^{2+}	9.7900	[27]
thio16	H	H	$-CH_3$	$-C_6H_5O$	Ni^{2+}	10.380	[27]
thio17	H	H	$-CH_3$	$-C_6H_5O$	Cu^{2+}	11.960	[27]
thio18	H	H	$-C_6H_5O$	$-C_6H_5O$	Co^{2+}	12.215	[16]

The one-way ANOVA method is also used to investigate the discrepancy between the calculated values ($\log\beta_{12,cal}$) and the initial values ($\log\beta_{12,exp}$) on the EV dataset of both two models. The results indicate that the difference between these results is insignificant (F = 0.1522 < F-crit (0.05) = 4.1300).

3.4. Design and calculation of new complexes

The biological activity of phenothiazine derivatives is very high, just like that of thiosemicarbazone and its complexes [28]. We have previously produced certain thiosemicarbazones derivatives that contain phenothiazine and their compounds [28,29]. In this investigation, eight novel thiosemicarbazone ligands with substituted derivatives at the R_4 site were created. The hydrogen atoms are substituted at the positions R_1, R_2, and R_3 of the ligands (Figure 1a).

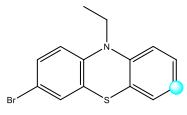
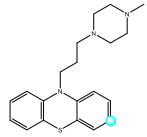
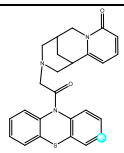
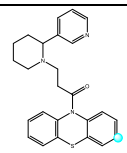
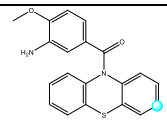
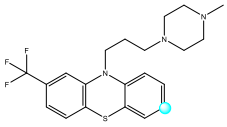
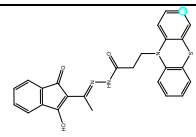
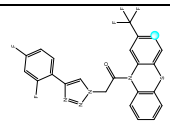
The designing principle of phenothiazine groups is based on the descriptors of the developed models such as *S6*, *xvp5*, *xvp7*, *Surface*, *xvch9* and *k0*, and the derivatives were explored as a result of the predictive model results. The novel-designed thiosemicarbazones containing metal ions such as Ag^+, Cu^{2+}, Zn^{2+}, Ni^{2+}, and Cd^{2+} are the building blocks for the 20 new complexes.

Similar calculations to those used to build the complexes of the training and external data set were applied to create the data set for the new complexes. Using PM7 and PM7/sparkle semi-experimental quantum calculations on MoPac2016, the

complexation is defined by the overall energy values and structural morphology. Table 5 lists the new complexes with the predicted $\log\beta_{12,\text{Pred-new}}$ values.

By integrating the data set for the new complexes into the initial training data set and computing the Cook distance indicator (D_{cook}), the new complexes were evaluated for the application domain (AD) ($|D_{Cook}| < 1.0$) [24,30] and outliers in order to test the calculated $\log\beta_{12,\text{Pred-new}}$ values from the MLR-QSPR8 and ANN3-QSPR models.

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

R_4 site	Metal ions	$\log\beta_{12,\text{Pred-new}}$		R_4 site	Metal ions	$\log\beta_{12,\text{Pred-new}}$	
		MLR8	ANN			MLR8	ANN
	Ag^{+}	12.714	11.880		Ag^{+}	11.840	9.604
	Cu^{2+}	10.977	10.278		Ni^{2+}	11.778	11.047
	Ni^{2+}	11.171	10.330		Cd^{2+}	9.961	10.243
	Cd^{2+}	8.790	9.314		Ag^{+}	13.244	12.202
	Cu^{2+}	9.615	9.451		Cd^{2+}	10.236	10.595
	Zn^{2+}	11.297	10.898		Zn^{2+}	13.789	12.144
	Cu^{2+}	8.492	8.271		Cu^{2+}	12.523	11.425
	Cd^{2+}	10.178	10.386		Ni^{2+}	11.673	10.713
	Ni^{2+}	11.884	11.072		Zn^{2+}	10.165	9.668
	Zn^{2+}	11.945	11.133		Ag^{+}	9.653	8.004

The outcomes demonstrated that the Cook distance values of 20 novel complexes satisfied the prediction criteria.

Additionally, the projected $\log\beta_{12,\text{Pred-new}}$ values generated from the MLR-QSPR8 and ANN3-QSPR models were also compared using the one-way ANOVA technique ($F = 2.6033 < F_{0.05} = 4.0982$) showed that there is no difference between them.

4. CONCLUSION

The development of MLR-QSPR and ANN-QSPR models to validate the complexation between metal ions and thiosemicarbazones with $\log\beta_{12}$ values in an aqueous solution has been successful in this work. The techniques of multivariate linear regression are used to create the MLR-QSPR models. The six descriptors such as *S6*, *xvp5*, *xvp7*, *Surface*, *xvch9*, and *k0* of the best MLR-QSPR8 model were found. The six variables were used as the input for effectively constructing the I(6)-HL(7)-O(1) ANN3-QSPR model. The developed QSPR models were statistically fit-valid. To assess the importance of the descriptors in the created model, the

predicted $\log\beta_{12}$ values of 20 novel complexes have been checked for applicability domain and outliers using the indicators of OECD instructions. More than that, the 20 newly created metal-thiosemicarbazone complexes have successfully been designed and calculated the $\log\beta_{12}$ values. Results of potential QSPR models could be used to identify various thiosemicarbazones and their complexes for application in the field of metal analysis.

REFERENCES

[1] Kumar, S., Dhar, D.N., Saxena, P. N., (2009). Applications of metal complexes of Schiff bases-A review. *J. Sci. Ind. Res.,* **68**, 181-187.

[2] Khan, T., Ahmad, R., Joshi, S., Khan, A. R., (2015). Anticancer potential of metal thiosemicarbazone complexes: a review. *Chem. Sin.,* **6(12)**, 1-11.

[3] Kunal, R.; Supratik, K.; Rudra, N. D., (2015). *A Primer on QSAR/QSPR Modeling, Fundamental Concepts*. New York: Springer.

[4] Hu, T.; Lai, Q.; Fan, W.; Zhang, Y.; Liu, Z., (2023). Advances in Portable Heavy Metal Ion Sensors. *Sensors*, **23(8)**, 4125.

[5] Stewart, J. J. P., (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.

[6] Stewart, J. J. P., (2002). MOPAC2016 Version: 17.240W. Stewart Computational Chemistry, USA.

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

[8] Patel, K. N., Parikh, K. S., Patel, R. M, (2010). 2-Hydroxy-4-n-butoxy-5-bromoacetophenone thiosemicarbazone as an extractive spectrophotometric reagent for nickel. *Orbital Elec. J. Chem.,* **2(4)**, 341-346.

[9] Patel, R. M.. Parikh, K. S., Patel, K. N., (2010). 2-hydroxy-4-n-butoxy-5-bromopropiophenone thiosemicarbazone as an extraction spectrophotometric reagent for Nickel (II). *Int. J. Chem Tech. Res.,* **2(2)**, 1090-1093.

[10] Parikh, K. S., Patel, R. M., Patel, K. N. New, (2009). Spectrophotometric Method for Determination of Cadmium. *E-J. Chem.,* **6(S1)**, S496-S500.

[11] Patel, N. B.; Solanki, Y. J., (2016). 2,4-Dihydroxy-5-Bromo [2'Methyl] Propiophenone Thiosemicarbazone [DHBMPT] as an Analytical Reagent: Studies on Co(II) Chelate. *J. App. Chem.,* **5(3)**, 654-660.

[12] Atalay, T.; Akgemci, E. G., (1998). Thermodynamic Studies of Some Complexes of 2-benzoylpyridine 4-phenyl-3-thiosemicarbazone. *Tr. J. Chem.,* **22**, 123-127.

[13] Bernhardt, P. V.; Sharpe, P. C.; Islam, M.; Lovejoy, D. B.; Kalinowski, D. S.; Richardson, D. R., (2009). Iron Chelators of the Dipyridylketone Thiosemicarbazone Class: Precomplexation and Transmetalation Effects on Anticancer Activity. *J. Med. Chem.,* **52(2)**, 407-415.

[14] Gaál, A.; Orgován, G.; Polgári, Z.; Réti, A.; Mihucz, G.; Bősze, S.; Szoboszlai, N.; Streli, C., (2014). 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**, 52-58.

[15] Garg, B. S., Saxena, V., Dixit, R., (1992).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**, 169-175.

[16] Toribio, F.; Fernandez, J. M.; Bendito, T.; Valcárcel, M., (1980). 2,2'-dihydroxybenzophenone thiosemicarbazone as a spectrophotometric Reagent for the determination of copper, cobalt, nickel, and iron trace amounts in mixtures without previous separations. *Microchemical Journal*, **25**, 338-347.

[17] Dassault Systèmes (2016). BIOVA Draw 2017 R2. Version: 17.2.NET, France.

[18] Tat, P. V., (2009). *Development of QSAR and QSPR*. Ha Noi: Publisher of Natural sciences and Technique.

[19] Steppan, D. D.; Werner, J.; Yeater, P. R., (1998). *Essential Regression and Experimental Design for Chemists and Engineers*. Free Software Package.

[20] Gasteiger, J.; Zupan, J., (1993). Neural Networks in Chemistry. *Chiw. Inr. Ed. EngI.* **32**, 5-15.

[21] MathWorks. Matlab R2016a 9.0.0.341360. USA (2016).

[22] Golbraikh, A.; and Tropsha, A, (2002). Beware of Q2, *J. Mol. Graphics Model.*, **20**, 269-276.

[23] Addinsoft, (2016). XLSTAT. Version 2016.02.28451, USA.

[24] Organisation for Economic Co-operation and Development (OECD), (2007). Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships Models. France: Organisation for Economic Co-operation and Development.

[25] Parikh, S.; Patel, R. M.; Patel, K. N., (2010). 2-Hydroxy-4-n-butoxy-5-bromo propiophenone Thiosemicarbazone as Spectrophotometric Reagent for Iron. *Asian J. Chem.*, **22(4)**, 2805-2810.

[26] Sahadev, R.; Sharma, R. K.; Sindhwani, S. K., (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.

[27] Garg, B. S.; Ghosh, S. Jain, V. K.; Singh, P. K., (1990). Evaluation of thermodynamic parameters of bivalent metal complexes of 2-hydroxyacetophenonethiosemicarbazone (2-HATS). *Thermochimica Acta*, **157**, 365-368.

[28] Sudeshna G.; Parimal, K., (2010).Multiple non-psychiatric effects of phenothiazines: A review. *European Journal of Pharmacology*, **648(1-3)**, 6-14.

[29] Al-Busaidi, I. J.; Haque, A.; Al-Rasbi, N. K.; Khan, M. S., (2019). Phenothiazine-based derivatives for optoelectronic applications: A review. *Synthetic Metals*, **257**, 116189

[30] Sahigara, F. Mansouri, K. Ballabio, D. Mauri, M. Consonni, K. and odeschini, R. T., (2012). Comparison of different approaches to define the applicability domain of QSAR models. *Molecules*, **17**, 4791-4810.