CALCULATION OF STABILITY CONSTANTS OF NEW METAL-THIOSEMICARBAZONE COMPLEXES BASED ON THE QSPR MODELING USING MLR AND ANN METHODS

Nguyen Minh Quang[1*], Tran Nguyen Minh An[1], Pham Van Tat[2], Bui Thi Phuong Thuy[3], and Nguyen Thanh Duoc[4]

[1]*Faculty of Chemical Engineering, Industrial University of Ho Chi Minh City*

[2]*Institute of Development and Applied Economics, Hoa Sen University*

[3]*Faculty of Basic Sciences, Van Lang University*

[4]*Faculty of Pharmacy, Hong Bang International University*

[*]*Corresponding author: nguyenminhquang@iuh.edu.vn*

Article history

Received: 21/01/2021; Received in revised form: 23/02/2021; Accepted: 22/04/2021

Abstract

In this study, the stability constants ($log\beta_{11}$) of twenty-eight new complexes between several ion metals and thiosemicarbazone ligands were predicted on the basis of the quantitative structure property relationship (QSPR) modeling. The stability constants were calculated from the results of the QSPR models. The QSPR models were built by using the multivariate least regression ($QSPR_{MLR}$) and artificial neural network ($QSPR_{ANN}$). The molecular descriptors, physicochemical and quantum descriptors of complexes were generated from molecular geometric structure and semi-empirical quantum calculation PM7 and PM7/sparkle. The best linear model $QSPR_{MLR}$ involves five descriptors, namely Total energy, xch6, xp10, SdsN, and Maxneg. The quality of the $QSPR_{MLR}$ model was validated by the statistical values that were R^2_{train} = 0.860, Q^2_{LOO} = 0.799, SE = 1.242, F_{stat} = 54.14 and PRESS = 97.46. The neural network model $QSPR_{ANN}$ with architecture I(5)-HL(9)-O(1) was presented with the statistical values: R^2_{train} = 0.8322, Q^2_{CV} = 0.9935 and Q^2_{test} = 0.9105. Also, the QSPR models were evaluated externally and achieved good performance results with those from the experimental literature. In addition, the results from the QSPR models could be used to predict the stability constants of other new metal-thiosemicarbazones.

Keywords: *Artificial neural network, multivariate least regression, QSPR, stability constants $log\beta_{11}$, thiosemicarbazone.*

DOI: https://doi.org/10.52714/dthu.10.5.2021.893

Cite: Nguyen Minh Quang, Tran Nguyen Minh An, Pham Van Tat, Bui Thi Phuong Thuy, and Nguyen Thanh Duoc. (2021). Calculation of stability constants of new metal-thiosemicarbazone complexes based on the QSPR modeling using MLR and ANN methods. *Dong Thap University Journal of Science*, *10*(5), 31-45.

TÍNH TOÁN HẰNG SỐ BỀN CỦA CÁC PHỨC CHẤT MỚI GIỮA THIOSEMICARBAZONE VÀ ION KIM LOẠI DỰA TRÊN MÔ HÌNH HÓA QSPR SỬ DỤNG PHƯƠNG PHÁP MLR VÀ ANN

Nguyễn Minh Quang[1*], Trần Nguyễn Minh Ân[1], Phạm Văn Tất[2], Bùi Thị Phương Thúy[3] và Nguyễn Thành Được[4]

[1]Khoa Công nghệ Hóa học, Trường Đại học Công nghiệp Thành phố Hồ Chí Minh

[2]Viện Phát triển và Công nghệ ứng dụng, Trường Đại học Hoa Sen

[3]Khoa Khoa học Cơ bản, Trường Đại học Văn Lang

[4]Khoa Dược, Trường Đại học Quốc tế Hồng Bàng

[*]Tác giả liên hệ: nguyenminhquang@iuh.edu.vn

Lịch sử bài báo

Ngày nhận: 21/01/2021; Ngày nhận chỉnh sửa: 23/02/2021; Ngày duyệt đăng: 22/04/2021

Tóm tắt

Trong nghiên cứu này, hằng số bền ($\log\beta_{11}$) của 28 phức chất mới giữa một số ion kim loại và phối tử thiosemicarbazone được dự đoán dựa trên mô hình hóa mối quan hệ định lượng giữa tính chất-cấu trúc (QSPR). Hằng số bền được tính toán từ kết quả các mô hình QSPR. Các mô hình QSPR được xây dựng bằng cách sử dụng phương pháp hồi quy đa biến ($QSPR_{MLR}$) và mạng thần kinh nhân tạo ($QSPR_{ANN}$). Các mô tả phân tử, hóa lý và lượng tử của các phức chất được tính toán từ cấu trúc hình học phân tử và phương pháp lượng tử bán thực nghiệm PM7 và PM7/sparkle. Mô hình tuyến tính tốt nhất $QSPR_{MLR}$ bao gồm năm mô tả: Total energy, xch6, xp10, SdsN và Maxneg. Chất lượng của mô hình $QSPR_{MLR}$ được đánh giá qua các giá trị thống kê như R^2_{train} = 0,860, Q^2_{LOO} = 0,799, SE = 1,242, F_{stat} = 54,14 và PRESS = 97,46. Mô hình mạng thần kinh $QSPR_{ANN}$ với kiến trúc I(5)-HL(9)-O(1) được tìm thấy với các giá trị thống kê: R^2_{train} = 0,8322, Q^2_{CV} = 0,9935 và Q^2_{test} = 0,9105. Ngoài ra, các mô hình QSPR này đã được đánh giá ngoại và cho kết quả tốt so với các giá trị thực nghiệm. Hơn nữa, kết quả từ các mô hình QSPR có thể được sử dụng để dự đoán hằng số bền của các phức chất giữa ion kim loại và thiosemicarbazone mới khác.

Từ khóa: *Mạng thần kinh nhân tạo, hồi quy đa biến, QSPR, hằng số bền $\log\beta_{11}$, thiosemicarbazone.*

1. Introduction

The diverse structure and easy complexation with many metal ions of thiosemicarbazone derivatives led to its wide applications in many fields (Casas *et al*., 2000). In the field of chemistry, thiosemicarbazones are used as analytical reagents (Reddy *et al*., 2011), they are also used as a catalyst in chemical reactions (Eğlencea *et al*., 2018). Besides, they also have application in biology (Nagajothi *et al*., 2013), environment (Pyrzynska, 2007) and medicine (Ezhilarasi, 2012). This is the reason why thiosemicarbazone derivatives and their complexes are popularly studied in practice. Recently, the stability constant of the complexes regarding thiosemicarbazone ligands has been explored for related applications like analytical chemistry with the UV/VIS spectrophotometric method or drug design via good pharmaceutical activity (Nagajothi *et al*., 2013; Ezhilarasi, 2012).

On the flip side with continuous efforts of scientists, new mathematical methods have been discovered and the powerful development of computer science has led to the emergence of many chemometric tools applied widely in computational chemistry (Yee and Wei, 2012). Therefore, we combined mathematical methods, chemistry and software in order to find an exact direction in theoretical research for a new substance group. This method was called the modeling of the quantitative structure property relationships (QSPR) applied on the complexes of thiosemicarbazone and metal ions in the work (Yee and Wei, 2012).

In this work, we approached the QSPR modeling methods for the construction QSPR models with the logarithm of stability constants ($\log\beta_{11}$) of the complexes (M:L) between thiosemicarbazone ligands with several metal ions (M = Cu^{2+}, Zn^{2+}, Fe^{2+}, Fe^{3+}, Cd^{2+}, Ag^{+}, Mo^{6+}, Mn^{2+}, La^{3+}, Pr^{3+}, Nd^{3+}) in aqueous solution. The $\log\beta_{11}$ values were selected from an experimental published database. The 2D and 3D-descriptors of metal-complexes are taken from the results of calculation on the structure optimization of complexes by means of semi-empirical quantum mechanics (Kunal *et al*., 2015) and QSARIS package (QSARIS 1.1, 2001). The two kinds of QSPR models were constructed by using multiple linear regressions ($QSPR_{MLR}$) and the artificial neural network ($QSPR_{ANN}$). These QSPR models were evaluated fully by combining cross and external validation procedures. Besides, a new series of thiosemicarbazone ligands and complexes were designed and predicted the stability constant by the outcome of the developed QSPR models.

2. The QSPR modeling method

Obviously, the quantitative structure and property relationship (QSPR) method is known as in the silico method used widely in many fields for predicting properties of chemical compounds based on the relationships between the structural characteristics and the properties (Yee and Wei, 2012). Also, the QSPR is known to derive from a quantitative structure and activity relationship (QSAR), in which the properties of the model are replaced by activity, first introduced by Crum Brown and Fraser (Kunal *et al*., 2015) in 1868.

In the 1940s, the appearance of chemical graph theory and the publications of Wiener and Platt's research helped the development of QSPR modeling (Kunal *et al*., 2015). According to the statistics up to 2016, the number of published works related to QSPR models was about 11,000 projects (Kunal *et al*., 2015). Nowadays, the QSPR method is widely used and deemed as an effective method for finding new compounds.

The QSAR/QSPR model should meet the requirements of the OECD principles (OECD, 2007) as follows:

- A determined response;
- A clearly algorithm;
- A detailed applicability domain;

- Statistical response;
- Explaining the mechanism, if possible.

The development of QSPR model consists of the following main steps (Kunal *et al*., 2015):

- Data mining;
- Structural compounds design and optimization;
- Calculating the molecular descriptors;

Standardized data sets;

- Building models;
- Testing and evaluation model;
- Application of the models.

The basic equation of the QSPR method can be expressed mathematically as follows (Kunal *et al*., 2015):

$$Response\ (property) = f(descriptors). \quad (1)$$

There are two popular approaches to establish QSPR models, that is linear regression (MLR, PLS, PCR) and machine learning method (SVR, ANN) (Kunal *et al*., 2015). In this work, we use two approaches to build the QSPR models of MLR and ANN.

3. Data and Computational methods

For a QSPR model, the standardized steps that must be carried out (Kunal *et al*., 2015), which are clearly indicated in the following subsections.

3.1. Stability constant of complex and structure selection

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

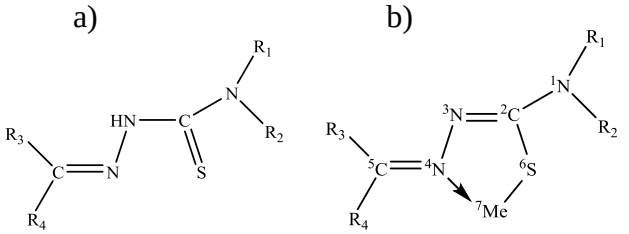


Figure 1. Structure of the thiosemicarbazone ligand (a) and the metal-thiosemicarbazone complex (b)

Therefore, the formation of the complex is the general equilibrium reaction (Harvey, 2000)

$$p\,M + q\,L \rightleftharpoons MpLq.$$

In which, in one step with p = 1 and q = 1, the stability constant (β_{11}) is calculated on the concentrations of the reagents and complexes at the equilibrium time. It is given by

$$\beta_{11} = \frac{[ML]}{[M][L]}.$$

3.2. Data selection

The data mining is the first step in the QSPR modeling research. Firstly, a great amount of related data was mined from prestigious data source, then two methods such as AHC and k-means are used to divide it into several data sets (Kunal *et al*., 2015). In this study, a data set comprising the 50 values $\log\beta_{11}$ of complexes between metal ions and the ligand thiosemicarbazone was used to build QSPR modeling on Table 1.

3.3. Descriptors calculation

Molecular descriptors are understood as the variables in the equations of the QSPR models. They can be specified as basic numerical characteristics related chemical structures. So, the metal-thiosemicarbazone complexes were drawn molecular structure with Avogadro 1.2.0 (Jekyll and Minimal, 2017) and optimized by using the semi-empirical quantum method with new version PM7 and PM7/sparkle on the MoPac2016 system (Stewart, 2002). The variable descriptors in the data set were determined by means of the QSARIS package (QSARIS 1.1, 2001; Pham Van Tat, 2009). The quantum descriptors were collected fully from the results of quantum mechanics (Kunal *et al*., 2015).

3.4. The QSPR models development

The two modeling methods were used to develop the QSPR regression models in this study, namely the multivariate linear regression (MLR) and artificial neural network (ANN). The $QSPR_{ANN}$ models are established on the basis of the initial variable form the result of the $QSPR_{MLR}$ model.

Table 1. The 50 stability constants of complexes (n) in experimental dataset with minimal ($log\beta_{11,min}$) and maximal ($log\beta_{11,max}$) values

No	Thiosemicarbazone ligand				Metal ions	Number of complexes, *n*	$log\beta_{11,min}$	$log\beta_{11,max}$	Ref.
	R_1	R_2	R_3	R_4					
1	H	H	H	- C_6H_4OH	Cu^{2+}	12	4.750	5,280	Biswas *et al.*, 2014
2	H	H	H	-$C_{13}H_{16}NO_3$	Cu^{2+}	1	17.540	17.540	Milunovic *et al.*, 2012
3	H	H	H	-$C_{13}H_{16}NO_3$	Zn^{2+}	1	12.400	12.400	Milunovic *et al.*, 2012
4	H	H	H	-$C_{13}H_{16}NO_3$	Fe^{2+}	1	12.240	12.240	Milunovic *et al.*, 2012
5	H	H	H	-CH=CHC_6H_5	Cd^{2+}	1	5.544	5.544	Krishna and Devi, 2015
6	H	H	H	-CH=CHC_6H_5	Mo^{6+}	1	6.5514	6.5514	Krishna and Mohan, 2013
7	-CH_3	-CH_3	-C_5H_4N	-C_5H_4N	Cu^{2+}	1	7.080	7.080	Gaál *et al*., 2014
8	-CH_3	-CH_3	-C_5H_4N	-C_5H_4N	Fe^{3+}	1	7.060	7.060	Gaál *et al*., 2014
9	H	H	H	-$C_{14}H_{12}N$	Cd^{2+}	1	5.860	5.860	Koduru and Lee, 2014
10	H	-C_2H_5	H	-C9H5NOH	Cu^{2+}	1	14.670	14.670	Rogolino *et al.*, 2017
11	H	-C_6H_5	H	-C_9H_5NOH	Zn^{2+}	1	7.300	7.300	Rogolino *et al.*, 2017
12	H	H	H	-$C_5H_3NCH_3$	Ag^{+}	1	14.500	14.500	Jiménez *et al.*, 1980
13	H	H	H	-$C_6H_3(OH)OCH_3$	Cd^{2+}	4	6.790	7.340	Garg and Jain, 1989
14	H	H	H	-$C_6H_3(OH)OCH_3$	Zn^{2+}	4	7.110	7.470	Garg, B. S., and Jain, V. K., 1989
15	H	H	-CH_3	-C_6H_4OH	Mn^{2+}	3	4.320	5.000	Garg *et al.*, 1990
16	H	H	-C_6H_5	-$C(C_6H_5)$=N-OH	Cu^{2+}	1	5.7482	5.7482	Reddy and Prasad, 2004
17	H	H	H	-$C_6H_4NH_2$	Cu^{2+}	2	11.570	11.610	Sawhney and Chandel, 1983
18	H	H	H	-$C_6H_4NO_2$	La^{3+}	2	9.450	10.840	Sawhney and Chandel 1984
19	H	H	H	-$C_6H_4NO_2$	Pr^{3+}	2	10.420	11.040	Sawhney and Chandel, 1984
20	H	H	H	-$C_6H_4NO_2$	Nd^{3+}	2	8.410	9.090	Sawhney and Chandel,1984
21	H	H	H	-$C_6H_4NO_2$	Cd^{2+}	2	10.630	10.950	Sawhney and Sati, 1983

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					
22	H	H	H	$-C_6H_4NO_2$	Al^{3+}	2	10.980	11.240	Sawhney and Sati, 1983
23	H	H	$-C_6H_4OH$	$-C_6H_4OH$	Fe^{3+}	1	5.496	5.496	Toribio *et al.*, 1980
24	H	H	$-CH_3$	$-C_5H_4N$	Cu^{2+}	1	5.491	5.491	Admasu *et al.*, 2016
25	H	H	$-CH_3$	$-C_5H_4N$	Cu^{2+}	1	5.924	5.924	Admasu *et al.*, 2016

3.4.1. MLR method

In QSPR modeling methods, the values $\log\beta_{11}$ are considered as the target values and in this case, they are dependent variables (*Y*) while the independent variables are quantitative variables as structural descriptors (X). If they are well correlated, the model is represented by a multivariate linear regression (MLR) model according to the following equation: (Kunal *et al*., 2015; XLSTAT, 2016)

$$Y = \beta_0 + \sum_{j=1}^{k} \beta_j X_j, \qquad (4)$$

where β_0, is the intercept of the model, β_j is the regression coefficients and *k* is number of explanatory variables in the equation.

3.4.2. Artificial neural network

In its nature, the artificial neural network (ANN) is a non-linear regression method that exerts to facsimile the operation of human neural networks. Nowadays, ANN is used widely in many fields such as mathematics, electronic research, medicine, chemistry and several other practical applications (Gasteiger and Zupan, 1993); particularly, it is applied successfully in the field of drug design and searching for new chemical compounds.

Generally, an ANN model includes an input layer, one or more hidden layer, and an output layer. Neurons in each of the layers are called nodes interconnecting with one another and receiving linked weights. The typical ANN architecture used in many studies is multi-layer perceptron (MLP) for the formation of models (Gasteiger and Zupan, 1993).

In this study, the MLP-ANN type is used with an error back-propagation algorithm (Vogl *et al*., 1988). The architecture consists of three layers I(*k*)-HL(*m*)-O(*n*). The input layer (*k*) put out from the variables of the MLR model. A quantitative output layer (*n*) is the stability constant $\log\beta_{11}$ and the number of hidden neurons (*m*) is determined by neurons on the input and output layer. So, there are two steps to find out the best ANN architecture for $QSPR_{ANN}$ model. In the first step, the *m* values of hidden neurons are surveyed by using Neural Designer tools (Artelnics, 2020), then we use data sets to build and externally validate the $QSPR_{ANN}$ model from the results of surveyed models. These calculations of the second step are run on the Matlab system (Matlab 2016a 9.0.0.341360, 2016) with Neural Network tool (nntool) toolbox.

In addition, to investigate the *m* values of hidden neurons, the training of ANN models uses two basic transfer functions in the neural network that are the hyperbolic sigmoid tangent and log-sigmoid transfer function. These transfer functions are represented mathematically as follows (Vogl *et al*., 1988)

$$a = \tan sig(n) = \frac{2}{\left(1 + e^{-2n}\right)^{-1}}. \qquad (5)$$

$$a = \log sig(n) = \frac{1}{1 + e^{-n}}. \qquad (6)$$

3.5. Model validation

The validation of the models is an important period in QSPR research. Normally, the models were validated internally and externally by two different data sets. Because the models were constructed based on statistics methods, they were checked by using the values R^2_{train} for internal set, Q^2_{LOO} or Q^2_{EV} for external-validation set (Kunal, 2015; Steppan *et al.*, 1998). These were calibrated by the same formula

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}, \qquad (7)$$

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

In addition, the R^2_{adj} is an adjustment to R^2, which takes into account the number of variables used in the model R^2_{adj} is defined by (Steppan *et al.*, 1998)

$$R^2_{adj} = R^2 - \frac{k-1}{N-1}\left(1 - R^2\right). \qquad (8)$$

The standard errors (*SE*) is the square root of the mean squared error (*MSE*) and it is defined by (Steppan *et al.*, 1998)

$$SE = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{N-k-1}}, \qquad (9)$$

where N and k are the number of variables training set and the models, respectively.

The building of ANN model is trained until the mean square error (MSE_{ANN}) is minimized followed by a discrepancy of the output and real values (Matlab 2016a 9.0.0.341360, 2016). MSE_{ANN} is the average squared error between the networks outputs (o) and the target outputs (t). It is described as follows (Gasteiger and Zupan, 1993; Rojas, 1996).

$$MSE_{ANN} = \frac{1}{n}\sum_{1}^{n}\left(t_i - o_i\right)^2. \qquad (10)$$

This work uses the average absolute values of the relative errors MARE (%), where ARE (%) is the absolute value of the relative errors to compare the quality of the models. These are represented as follows (Pham Van Tat, 2009).

$$MARE, \% = \frac{\sum_{i=1}^{n} ARE_i, \%}{n}, \qquad (11)$$

$$ARE, \% = \frac{\left|\log \beta_{11,exp} - \log \beta_{11,cal}\right|}{\log \beta_{11,exp}} 100, \qquad (12)$$

where n is the number of test substances; $\beta_{11,exp}$ and $\beta_{11,cal}$ are the experimental and calculated stability constants, respectively

To evaluate the variable contributions in the models, we used a quantity which is the average contribution percentage, $MPx_{k,i}$. It is determined according to formula (13) (Pham Van Tat, 2009)

$$MPx_{k,i}, \% = \frac{1}{N}\sum_{m=1}^{N}\frac{100.\left|b_{k,i}.x_{m,i}\right|}{\sum_{j}^{k}\left|b_{k,j}.x_{m,j}\right|}, \qquad (13)$$

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

4. Results and discussion

4.1. $QSPR_{MLR}$ modeling

The multiple linear regression analysis was accomplished by stepwise regression technique on the Regress system (Steppan *et al.*, 1998) and MS-EXCEL (Billo, 2007). The cross validation for QSPR models was carried out by the leave-one-out process (LOO) using the statistic Q^2_{LOO} (Kunal *et al.*, 2015; Steppan *et al.*, 1998).

The data set for the building of $QSPR_{MLR}$ including the 50 stability constants values of complexes are divided into the training set and the test set. The criteria of statistical values such as R^2_{train}, R^2_{adj}, Q^2_{LOO} and F_{stat} (Fischer's value) are used to evaluate the quality of models (Kunal *et al.*, 2015). The $QSPR_{MLR}$ models and the statistical values are shown in Table 2.

Table 2. The results of $QSPR_{MLR}$ models (*k* of 1 to 5) and statistical values

k	Variables	*SE*	R^2_{train}	R^2_{adj}	Q^2_{LOO}	F_{stat}	*PRESS*
1	x_1	2.941	0.144	0.126	0.029	8.085	470.89
2	x_1/x_2	1.723	0.712	0.700	0.649	58.23	170.12
3	$x_1/x_2/x_3$	1.533	0.777	0.763	0.692	53.50	149.33
4	$x_1/x_2/x_3/x_4$	1.366	0.827	0.811	0.768	53.70	112.72
5	$x_1/x_2/x_3/x_4/x_5$	1.242	0.860	0.844	0.799	54.14	97.46

Notation of molecular descriptors			
Total energy	x_1	SdsN	x_4
xch6	x_2	Maxneg	x_5
xp10	x_3		

The choice of the variables for the best QSPR model is based on the results of Table 2. The variables are selected on basis of the changing the R^2_{train}, Q^2_{LOO} and F_{stat} values and *k* variables for meeting statistical requirements. The results showed that when k values increased to 5, the QSPR model recieved the best statistical values, so it leads to the selection of the model with the *k* of 5 for the tendency of variation. Specifically, the variables from x_1 to x_5 were closely monitored on the basis of the p-value (< 0.05) and t-student characterized the variables (Kunal *et al*., 2015; Steppan *et al.*, 1998). The best linear $QSPR_{MLR}$ models were selected with the statistical values follows:

$$\log\beta_{11} = -2.896 - 0.00231 \cdot x_1 - 95.81 \cdot x_2 + 3.810 \cdot x_3 + 1.558 \cdot x_4 - 20.76 \cdot x_5 \quad (14)$$

$$n = 50;\ R^2_{train} = 0.827;\ Q^2_{LOO} = 0.769;\ SE = 1.242.$$

As a consequence, the data used to develop the MLR is completely consistent and has a good predictability. The predictive ability of the $QSPR_{MLR}$ model is very suitable for complex groups. Therefore, this model can be used to predict new complexes of the same type group based on the AD and Outliers estimates (XLSTAT, 2016).

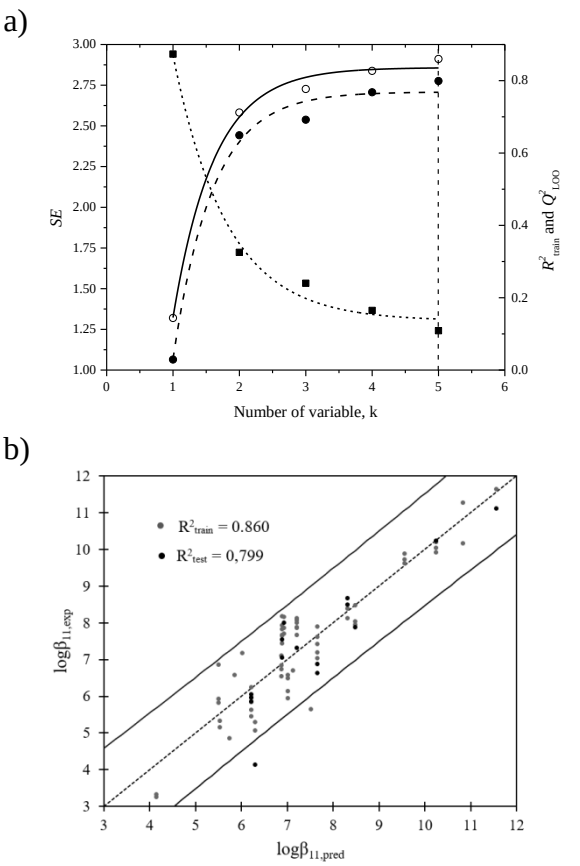


Figure 2. a) Change tendency line of values *SE*, R^2_{train} and Q^2_{LOO} according to *k* variables; b) Correlation of experimental vs. predicted values $\log\beta_{11}$ of the training compounds using the $QSPR_{MLR}$ model (with *k* = 5)

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

Statistical values and variables	$QSPR_{MLR}$			$MPx_{k,i}$, %			$GMPx_i$, %
	k = 3	k = 4	k = 5	k = 3	k = 4	k = 5	
R^2_{train}	0.777	0.827	0.860	–	–	–	–
R^2_{adj}	0.763	0.811	0.844	–	–	–	–
Q^2_{LOO}	0.692	0.768	0.799	–	–	–	–
SE	1.533	1.366	1.242	–	–	–	–
constant	-6.375	3.569	-2.896	–	–	–	–
x_1	-0.00331	-0.00269	-0.00231	47.1322	28.9843	19.4208	31.8458
x_2	-77.41	-92.54	-95.81	46.2050	41.6942	33.8984	40.5992
x_3	2.563	3.074	3.810	6.6628	6.1032	6.0056	6.2572
x_4	–	1.391	1.558	–	23.2184	20.3011	14.5065
x_5	–	–	-20.76	–	–	20.3740	6.7913

The study used three vicinity models to find out the effect of descriptors in the model according to the $GMPx_i$ values ($GMPx_i$ is the average value of $MPx_{k,i}$). The results in Table 3 showed that the contribution of the variables in the order of xch6 (x_2) > Total energy (x_1) > SdsN (x_3) corresponds to the values of 40.5992, 31.8458 and 14.5065. The xch6 parameter namely Chi chain 6 is the simple 6^{th}-order chain chi index which appreciates the important role of six cycle (QSARIS 1.1, 2001). The total energy is the quantum parameter which is the sum of kinetic and potential energies of the formed complexes while SdsN parameter is the sum of all (= N –) E-State values in a molecule (QSARIS 1.1, 2001). The values indicate the role of nitrogen bonds type in the complexes. The two remaining parameters (xp10 and Magnex) do not significantly affect the model. The important parameters will be selected to design new complexes and predict the stability constant of these complexes.

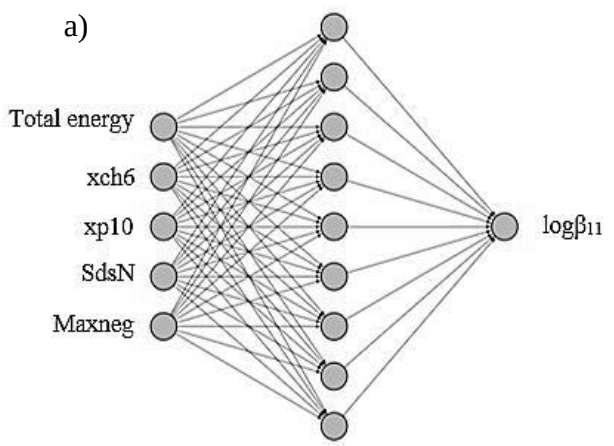
4.2. $QSPR_{ANN}$ modeling

The ANN model is built on the same data set and the resulting variables of the MLR model. So, the model is developed upon 5 variables of $QSPR_{MLR}$ model and the architecture of the neural network consist of three layers I(5)-HL(m)-O(1). The input layer I(3) holds 5 neurons of Total energy, xch6, xp10, SdsN and Maxneg, while the output layer O(1) contains 1 neuron of the $\log\beta_{11}$ values. The number of hidden layer (m) will be scanned to search for several good models with the training data set by using Design Neural tools (Artelnics, 2020). The ANN models are trained with Levenberg-Marquardt algorithm and back-propagation error approach (Gasteiger and Zupan, 1993; Vogl *et al.*, 1988). The transfer functions used in the training process are log-sigmoid and hyperbolic tangent function (Vogl *et al.*, 1988). The results of the m neurons are given in Table 4.

Table 4. The scanned $QSPR_{ANN}$ models I(5)-HL(*m*)-O(1) with statistical parameters

No	$QSPR_{ANN}$ models	R^2_{train}	Q^2_{test}	Q^2_{CV}	Training error	Test Error	Validation Error	Transfer Function
1	I(5)-HL(3)-O(1)	0.9288	0.9518	0.9876	0.6046	0.2869	1.3679	hyperbolic tangent
2	I(5)-HL(5)-O(1)	0.9579	0.9536	0.9748	0.3615	0.3210	1.5054	hyperbolic tangent
3	I(5)-HL(6)-O(1)	0.9276	0.9528	0.9910	0.6470	0.2946	0.8261	log-sigmoid
4	I(5)-HL(7)-O(1)	0.8614	0.9292	0.9703	1.3253	0.3097	1.2169	hyperbolic tangent
5	**I(5)-HL(9)-O(1)**	**0.8322**	**0.9105**	**0.9935**	**1.6650**	**0.3839**	**1.8190**	**log-sigmoid**
6	I(5)-HL(10)-O(1)	0.8758	0.9264	0.9935	1.4146	0.3425	0.1477	log-sigmoid
7	I(5)-HL(10)-O(1)	0.8858	0.9375	0.9933	0.9978	0.3336	0.4616	hyperbolic tangent

Next, we use the external data set to find out the best $QSPR_{ANN}$ model by external validation technique. The obtained results show that the $QSPR_{ANN}$ model I(5)-HL(9)-O(1) is shown in bold (Table 5) with the best predictability attached to the Q^2_{EV} value of 0.9404 as in Fig. 3. As a consequence, the log-sigmoid transfer function is used for network training and the optimized parameters of ANN such as the momentum constant of 0.05, the learning rate of 0.01, and the convergent goal of 10^{-7}.



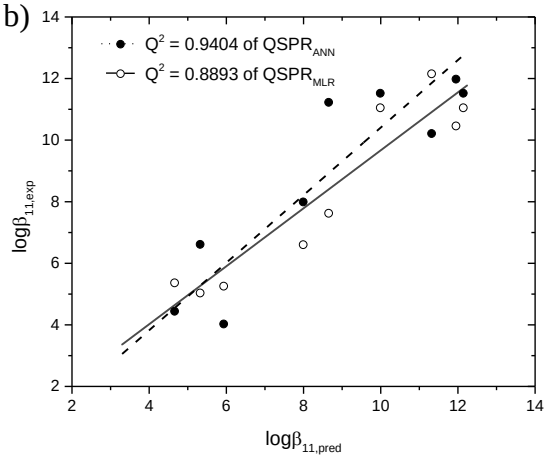


Figure 3. (a) Architecture of neural network I(5)-HL(9)-O(1); (b) The correlation between experimental vs. predicted values of external data set of QSPR models

4.3. The external validation of QSPR models

To create a good model, it is necessary to perform external evaluation on a data set independent of the training data set. This work used the external data with twelve complexes from the experimental studies. The validated results are described in Table 5.

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

Thiosemicarbazone ligand				Metal ions	$\log\beta_{11.exp}$	$\log\beta_{11.cal}$		ref.
R_1	R_2	R_3	R_4			$QSPR_{MLR}$	$QSPR_{ANN}$	
H	H	H	- C_6H_4OH	V^{5+}	5.3222	6.6116	5.0338	Reddy and Reddy, 1983
H	H	H	-C_9H_5NOH	Cu^{2+}	14.5600	16.3493	13.5943	Rogolino, *et al*., 2017
H	H	H	-C_6H_4OH	Ag^{+}	15.6000	13.0625	12.8733	Jiménez, *et al*., 1980
H	H	-CH_3	-CCH_3=N-OH	Cu^{2+}	19.1000	20.3499	17.3796	Atalay, T., and Ozkan, E., 1994
H	H	H	-$C_6H_3(OH)OCH_3$	Ni^{2+}	8.6500	11.2270	7.6249	Atalay, and Ozkan, 1994
H	H	H	-$C_{10}H_6OH$	Mn^{2+}	4.6600	4.4417	5.3646	Sahadev *et al*., 1992
H	H	H	-$C_{10}H_6OH$	Cd^{2+}	5.9300	4.0276	5.2577	Sahadev *et al*., 1992
H	H	-	-C_9H_8NO	Pb^{2+}	7.9920	7.9901	6.6023	Sarkar and Garg, 1987
H	H	H	-$C_6H_4NH_2$	Co^{2+}	11.9500	11.9804	10.4610	Sawhney and Chandel, 1983
H	H	H	-$C_6H_4NH_2$	Mn^{2+}	12.1400	11.5222	11.0507	Sawhney and Chandel, 1983
H	H	H	-$C_6H_4NH_2$	Mn^{2+}	9.9900	11.5222	11.0507	Sawhney and Chandel, 1983
H	H	H	-$C_6H_4NH_2$	Zn^{2+}	11.3200	10.2128	12.1537	Sawhney and Chandel, 1983
				MARE, %		13.0307	11.1375	

As observed by Table 6, the *MARE* values of $QSPR_{MLR}$ and $QSPR_{ANN}$ I(5)-HL(9)-O(1) models are 13.0307% and 11.1375%, respectively. The results pointed out that the ANN model has better predictability than the MLR model. Furthermore, the predicted $\log\beta_{11,cal}$ values of ANN model are approximate to the experimental $\log\beta_{11,exp}$ values.

In addition, the results of data analysis in Table 5 are indicated Fig. 4b, it can add to that point that the predictability of the two models is extremely positive (Kunal *et al*., 2015). With this, the neural network and linear regression models express the correlation between the predicted values and the experimental values with Q^2_{EV} values of 0.8893 and 0.9404, respectively.

Using the ANOVA method to evaluate the difference between the experimental and predictive values of the both models; accordingly, the differences between the QSPR models are negligible ($F = 0.1462 < F_{0.05} = 3.2849$).

4.4. Designing and Prediction of new complexes

The phenothiazine and carbazole derivatives are selected to design new

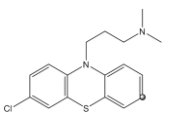
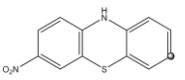
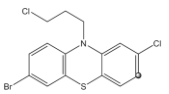
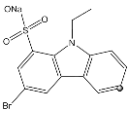
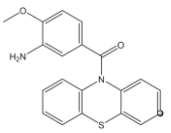
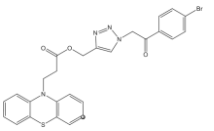
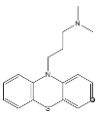
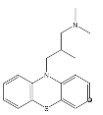
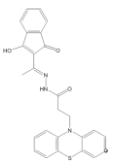
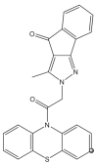
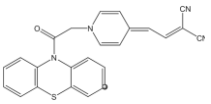
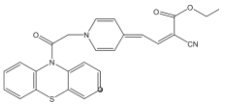
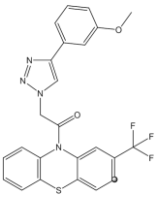
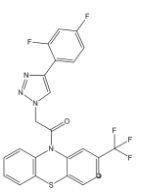
thiosemicarbazone and the complexes among the new ligands with several popular metal ions as Cd^{2+}, Ni^{2+}, Cu^{2+}, Ag^{+} and Zn^{2+} regarding the five descriptors Total energy, xch6, xp10, SdsN and Maxneg of the built models.

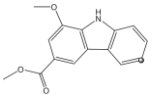
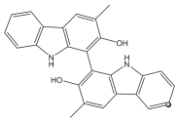
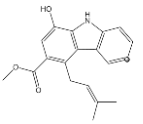
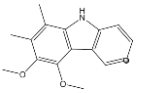
We choose these derivatives because they have their wide applications in many fields and they have been synthesized in previous experimental studies (Al-Busaidi *et al*., 2019; Sudeshna and Parimal, 2010; Huang *et al*., 2017; Krucaite and Grigalevicius, 2019). The new thiosemicarbazones are formed by attached phenothiazine and carbazole groups at the R_4 site while the remaining positions as R_1, R_2 and R_3 of the thiosemicarbazones are hydrogen atoms.

A series of new complexes are carefully screened and they are embedded in the spatial space data of the training set to test for AD and Outlier (XLSTAT, 2016). The results have 28 complexes that meet the standards of AD and they are predicted the stability constant from the two built QSP_{RMLR} and $QSPR_{ANN}$ models. The prediction of new complexes ($\log\beta_{11,new}$) is given in Table 6.

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

Table 6. Twenty-eight metal-thiosemicarbazone complexes with the predicted $\log\beta_{11,new}$ values from the constructed QSPR models

R_4 site	Metal ions	$\log\beta_{11,new}$		R_4 site	Metal ions	$\log\beta_{11,new}$	
		MLR	ANN			MLR	ANN
	Cd^{2+}	5.6901	6.7180		Cd^{2+}	7.6102	6.6247
	Ni^{2+}	6.4982	7.0743		Cu^{2+}	9.0953	7.2410
	Cu^{2+}	7.1884	8.0819		Zn^{2+}	7.5959	6.6284
	Ag^{+}	11.0823	11.6565		Ag^{+}	19.5092	17.0444
	Ni^{2+}	7.6813	7.2907		Cu^{2+}	11.8261	9.5105
	Ag+	14.5477	17.4063		Cd^{2+}	16.9690	17.5344
	Zn^{2+}	8.9921	9.2583		Cu^{2+}	18.4662	17.5347
	Ag^{+}	11.4791	8.4823		Ag^{+}	11.7851	9.2863
	Ag^{+}	19.1357	17.5366		Ag^{+}	17.7735	17.5328
	Ag^{+}	15.0392	17.5287		Ag^{+}	17.0232	17.5345
	Zn^{2+}	18.8756	17.4754		Zn^{2+}	19.2865	17.4970

R_4 site	Metal ions	$\log\beta_{11,new}$		R_4 site	Metal ions	$\log\beta_{11,new}$	
		MLR	ANN			MLR	ANN
	Ag^+	18.3002	17.3404		Ni^{2+}	19.4625	15.5944
	Cd^{2+}	10.5710	8.8755		Cu^{2+}	16.4397	17.1392
	Ni^{2+}	10.6671	12.7539		Ag^+	20.8683	17.0888

5. Conclusion

In this study, the quantitative structure-property relationship (QSPR) models based on the multivariate linear regression ($QSPR_{MLR}$) and artificial neural network ($QSPR_{ANN}$) have been successfully formed by using the dataset of structural descriptors and the stability constant value of complexes. The complexes structures were optimized by new version semi-empirical quantum mechanics PM7 and PM7/sparkle. The QSPR models were validated completely upon statistical values as R^2_{train}, Q^2_{LOO}, *MARE*, %, and ANOVA methods. The obtained results confirm these models as being the new ones for the prediction of twenty-eight new designing thiosemicarbazone derivatives. Furthermore, the results of QSPR models can be useful to discover new complexes that can be applied further in the essential fields such as analytical chemistry, environment monitoring and drug designing in pharmacology.

References

Admasu, D., Reddy, D. N., and Mekonnen, K. N. (2016). Spectrophotometric determination of Cu(II) in soil and vegetable samples collected from Abraha Atsbeha, Tigray, Ethiopia using heterocyclic thiosemicarbazone. *SpringerPlus, 5*(1169), 1-8.

Al-Busaidi, I. J., Haque, A., Al Rasbi, N. K., and Khan, M. S. (2019). Phenothiazine-based derivatives for optoelectronic applications: A review. *Synthetic Metals, 257*, 116189. doi:10.1016/j.synthmet.2019.116189.

Artelnics. (2020). *Neural Designer software*. USA: Artificial Intelligence Techniques, Ltd.

Atalay, T., and Ozkan, E. (1994). Thermodynamic studies of some complexes of 4'-morpholinoacetophenone thiosemicarbazone. *Thermochimica Acta., 237*, 369-374.

Billo, E. J. (2007). *Excel For Scientists and Engineers: Numerical Methods*. USA: John Wiley and Sons, Inc.

Biswas, R., Brahman, D., and Sinha. B. (2014). Thermodynamics of the complexation between salicylaldehyde thiosemicarbazone with Cu(II) ions in methanol–1,4-dioxane binary solutions. *J. Serb. Chem. Soc., 79*(5), 565-578.

Casas, J. S., García-Tasende, M. S., and Sordo, J. (2000). Main group metal complexes of semicarbazones and thiosemicarbazones. A structural review. *Coordination Chemistry Reviews, 209*(1), 197-261.

Eg˘lencea, S., Sahin, M. Özyürek, M., Apak, R. and Ülküseven, B. (2018). Dioxomolybdenum (VI) complexes of S-methyl-5-bromosalicylidene-Nalkyl substituted thiosemicarbazones: Synthesis,

catalase inhibition and antioxidant activities. *Inor. Chimi. Acta, 469,* 495-502.

Ezhilarasi. (2012). Synthesis Characterization and Application of Salicylaldehyde Thiosemicarbazone and Its Metal Complexes. *Int. J. Res. Chem. Environ, 2*(4*),* 130-148.

Gaál, A., Orgován, G., Polgári, Z., Réti, A., Mihucz, V. G., Bősze, S., Szoboszlai, N., and 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.

Garg, B. S., and Jain, V. K. (1989). Determination of thermodynamic parameters and stability constants of complexes of biologically active o-vanillinthiosemicarbazone with bivalent metal ions. *Thermochimica Acta., 146,* 375-379.

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

Gasteiger, J., and Zupan, J. (1993). Neural Networks in Chemistry. *Chiw. Inr. Ed. EngI., 32,* 503-521.

Harvey, D. (2000). *Modern analytical Chemistry.* Toronto: Mc.Graw Hill.

Huang, L., Feng, Z. L., Wang, Y. T., and Lin, L. G. (2017). Anticancer carbazole Alkaloids and coumarins from Clausena plants: A review. *Chinese Journal of Natural Medicines, 15*(12), 881-888.

Jekyll and Minimal M. (2017). *Avogadro 1.2.0*. Avogadro Chemistry, USA.

Jiménez, M. A., Luque De Castro, M. D., Valcárcel, M. (1980). Potentiometric Study of Silver (I)-Thiosemicarbazonates. *Microchemical Journal, 25,* 301-308.

Koduru, J. R., and Lee, K. D. (2014). Evaluation of thiosemicarbazone derivative as chelating agent for the simultaneous removal and trace determination of Cd(II) and Pb(II) in food and water samples. *Food Chemistry, 150,* 1-8.

Krishna, D. G., and Devi, C. K. (2015). Determination of cadmium (II) in presence of micellar medium using cinnamaldehyde thiosemicarbazone by spectrophotometry. *Int. J. Green Chem. Biopro., 5*(2), 28-30.

Krishna, D. G., and Mohan, G. V. K. (2013). A Facile Synthesis, Characterization of Cinnamaldehyde Thiosemicarbazone and Determination of Molybdenum (VI) by Spectrophotometry in Presence of Micellar Medium. *Indian J. Appl. Res., 3(8),* 7-8.

Krucaite, G., and Grigalevicius, S. (2019). A review on low-molar-mass carbazole- based derivatives for organic light emitting diodes. *Synthetic Metals, 247,* 90-108.

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

Matlab R2016a 9.0.0.341360. (2016), USA: MathWorks.

Milunovic, M., Enyedy, E. A., Nagy, N. V., Kiss, T., Trondl, R., Jakupec, M. A., Keppler, B. K., Krachler, R., Novitchi, G., and Arion, V. B. (2012). L- and D-Proline Thiosemicarbazone Conjugates: Coordination Behavior in Solution and the Effect of Copper(II) Coordination on Their Antiproliferative Activity. *Inorg. Chem., 51,* 9309-9321.

Nagajothi, A., Kiruthika, A., Chitra, S., and Parameswari, K. (2013). Fe(III) Complexes with Schiff base Ligands: Synthesis, Characterization, Antimicrobial Studies. *Res. J. chem. Sci., 3*(2), 35-43.

OECD. (2007). *Guidance Document on the Validation of (Quantitative) Structure–Activity Relationships Models*, France: Organisation for Economic Cooperation and Development.

Pham Van Tat. (2009). *Development of QSAR and QSPR*. Ha Noi: Publisher of Natural Sciences and Technique.

Pyrzynska, K. (2007). Determination of molybdenum in environmental samples. *Anal. Chim. Acta, 590*, 40-48.

QSARIS 1.1. (2001). USA: Statistical Solutions Ltd.

Reddy, K. H., and Prasad, N. B. L. (2004). Spectrophotometric determination of copper (II) in edible oils and seed using novel oxime-thiosemicarbazones. *India J. Chem., 43A*, 111-114.

Reddy, K. V., Babu, S. V. and Reddy, K. H. (2011). Spectrophotometric Determination of Copper(II) in Biological Samples by Using 2-Acetylfuran Thiosemicarbazone as Chelating Reagent. *Asia. J. Chem., 23*(10), 4425-4429.

Reddy, N. S. R., and Reddy, D. V. (1983). Spectrophotometric determination of vanadium (V) with salicylaldehyde thiosemicarbazone. *J. Indian. Inst. Sci., 64*(B), 133-136.

Rogolino, D., Cavazzoni, A., Gatti, A., Tegoni, M., Pelosi, G., Verdolino, V., Fumarola, C., Cretella, D., Petronini, P.G., and Carcelli, M. (2017). Anti-proliferative effects of copper (II) complexes with Hydroxyquinoline-Thiosemicarbazone ligands. *Eu. J. Med. Chem., 128*, 140-153.

Rojas, R. (1996). *Neural Networks*. Berlin: Springer-Verlag.

Sahadev, M., Sharma, R. K., and Sindhwani, S. K. (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*, 291-299.

Sarkar K., and Garg, B. S. (1987). Determination of thermodynamic parameters and stability constants of the complexes of p-MITSC with transition metal ions. *Thermochimica Acta, 113*, 7-14.

Sawhney, S. S., and Chandel, S. K. (1983). Solution chemistry of Cu(II)-, Co(II)-, Ni(II)-, Mn(II)- and Zn(II)-p-aminobenzaldehyde thiosemicarbazone systems. *Thermochimica Acta, 71*, 209-214.

Sawhney, S. S., and Chandel, S. K. (1984). Stability and thermodynamics of La(III)-, Pr(III)-, Nd(III)- Gd(III)- and Eu(III)-p-nitrobenzaldehyde thiosemicarbazone systems. *Thermochimica Acta, 72*, 381-385.

Sawhney, S. S., and Sati, R. M. (1983). pH-metric studies on Cd(II)-, Pb(II)-, AI(III)-, Cr(III)- and Fe(III)-p-nitrobenzaldehyde thiosemicarbazone systems. *Thermochimica Acta, 66*, 351-355.

Steppan, D. D., Werner, J., and Yeater, P. R. (1998). *Essential Regression and Experimental Design for Chemists and Engineers*, Free Software Package. http://home.t-online.de/home/jowerner98/indexeng.html.

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

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

Toribio, F., Fernandez, J. M. L., Bendito, D. P., and 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.

Vogl, T. P., Mangis, J. K., Rigler, A. K., Zink, W. T., and Alkon, D. L. (1988). Accelerating the convergence of the backpropagation method. *Biological Cybernetics, 59*, 257-263.

XLSTAT Version 2016.02.28451 (2016). USA: Addinsoft.

Yee, L. C. and Wei, Y. C. (2012). *Statistical Modelling of Molecular Descriptors in QSAR/QSPR, chapter 1: Current Modeling Methods Used in QSAR/QSPR*. German: Wiley-VCH Verlag GmbH & Co. KgaA.