Journal of Molecular Structure 1195 (2019) 95–109

Contents lists available at ScienceDirect

Journal of Molecular Structure

journal homepage: http://www.elsevier.com/locate/molstruc

Novel QSPR modeling of stability constants of metal-thiosemicarbazone complexes by hybrid multivariate technique: GA-MLR, GA-SVR and GA-ANN

Nguyen Minh Quang [c, d], Tran Xuan Mau [c], Nguyen Thi Ai Nhung [c], Tran Nguyen Minh An [d], Pham Van Tat [a, b, *]

[a] Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Viet Nam
[b] Faculty of Applied Sciences, Ton Duc Thang University, Ho Chi Minh City, Viet Nam
[c] Department of Chemistry, University of Sciences, Hue University, Hue City, Viet Nam
[d] Faculty of Chemical Engineering, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Viet Nam

ARTICLE INFO

Article history:
Received 7 March 2019
Received in revised form
29 April 2019
Accepted 14 May 2019
Available online 28 May 2019

Keywords:
QSPR models of stability constants
Metal-thiosemicarbazone complexes
Multivariate linear regression
Support vector regression
Artificial neural networks

ABSTRACT

The quantitative structural property relationship (QSPR) models of the logβ_{11} stability constants of M:L complexes of the structurally diverse thiosemicarbazones and several metal ions (M = Ag^{+}, Cd^{2+}, Co^{2+}, Cu^{2+}, Fe^{3+}, Mn^{2+}, Cr^{3+}, La^{3+}, Mg^{2+}, Mo^{6+}, Nd^{3+}, Ni^{2+}, Pb^{2+}, Zn^{2+}, Pr^{3+}, Dy^{3+}, Gd^{3+}, Ho^{3+}, Sm^{3+}, Tb^{3+}, V^{5+}) in aqueous solution have been constructed by combining the genetic algorithm with multivariate linear regression ($QSPR_{GA\text{-}MLR}$), support vector regression ($QSPR_{GA\text{-}SVR}$) and artificial neural network ($QSPR_{GA\text{-}ANN}$). The multi-levels optimization for grid search technique is used to find the best $QSPR_{GA\text{-}SVR}$ model with the optimized parameters capacity C = 1.0, Gamma, $\gamma = 1.0$ and Epsilon, $\varepsilon = 0.1$. The quality of the QSPR models presented in statistical values as training R^2 in range 0.9148–0.9815, validation Q^2 in range 0.7168–0.9669 and MSE values in range 0.2742–2.4906. The new two thiosemicarbazone reagents were designed and synthesized based on the lead thiosemicarbazone reagents. The logβ_{11} values of new complexes $Cu^{2+}L$, $Ni^{2+}L$, $Cd^{2+}L$ and $Zn^{2+}L$ derived from the $QSPR_{GA\text{-}SVR}$ and $QSPR_{GA\text{-}ANN}$ model turn out to be in a good agreement with experimental data.

1. Introduction

In recent years the thiosemicarbazones (Fig. 2) represented an important group of Schiff based substances bearing sulfur and nitrogen as donor atoms [1]. In the years 60, thiosemicarbazones appeared in significant applications in the drug areas against the dangerous disease such as tuberculosis, leprosy and smallpox [2,3]. In the decade of 60, one of the first cancer prevention activities of thiosemicarbazones have been discovered and present [4,5]. The anticancer activity of it is also very wide, but it depends very much on the characteristics of the cell. Thiosemicarbazone ligands have great biological importance as they have on display a wide range of biological activities including antibacterial, antifungal, antimalarial, against advanced, anti-inflammatory and antiviral [6,7]. The thiosemicarbazone ligand based on Schiff was synthesized by condensation reactions between primary amines and aldehydes or ketones ($R_3CR_2 = NR_1$ where R_1, R_2 and R_3 represent alkyl and/or aryl substituents) [8].

In the environmental fields, the diverse metal ions appear in nature into the coalition together in the minerals. Several metals have been used specifically for electric and steel plate. Large amounts of these metals are discharged into the environment. About half of the metal ion is released into the rivers through the weathering of rocks and some metals are released into the air through the fire woods and an active volcano. The rest of the differing metal ions is disengaged through human activities, such as production processes and the activities, etc. The amount of the metal consumption takes place primarily through the diet [9,10]. Track amounts of metal ions are important in industry [11], as a toxicant [12], and biological inessential [3], an environmental pollutant [11,12], and an occupational hazard [13]. Most of them are extremely toxic metal ions. To determine the metal ions in trace level, there are a number of methods appropriated regularly for

* Corresponding author. Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Viet Nam.
E-mail address: phamvantat@tdtu.edu.vn (P. Van Tat).

https://doi.org/10.1016/j.molstruc.2019.05.050
0022-2860/

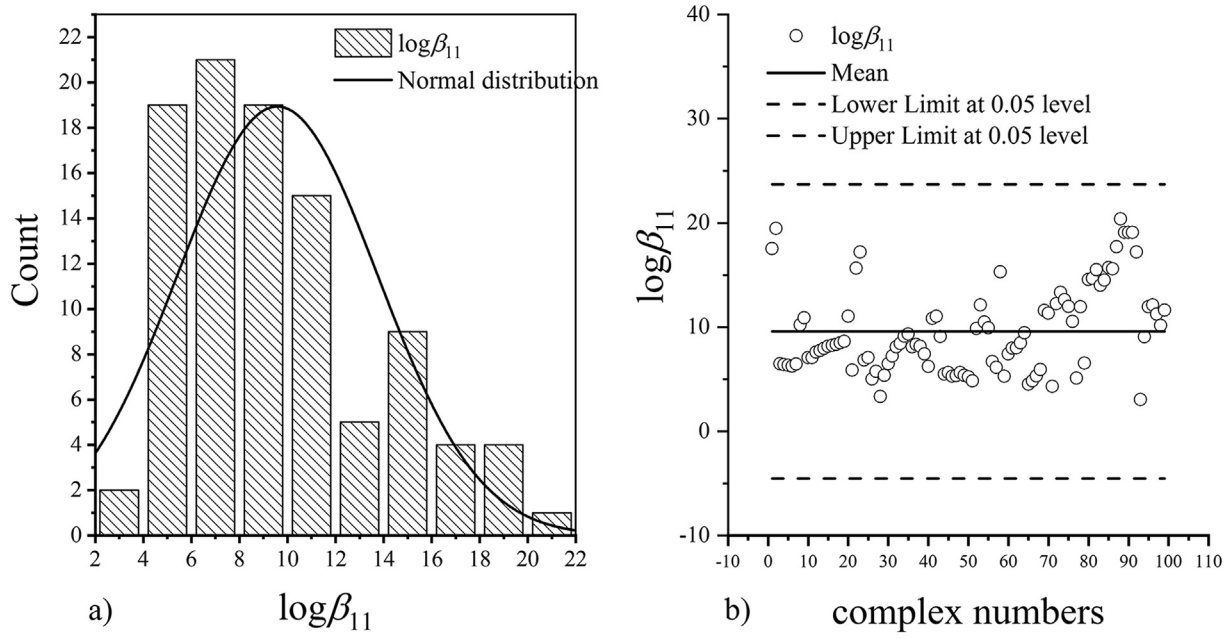


Fig. 1. The dataset behaviour of a) normal distribution of dataset b) Grubb's test used to test the outlier points of complexes at 95% confidence level.

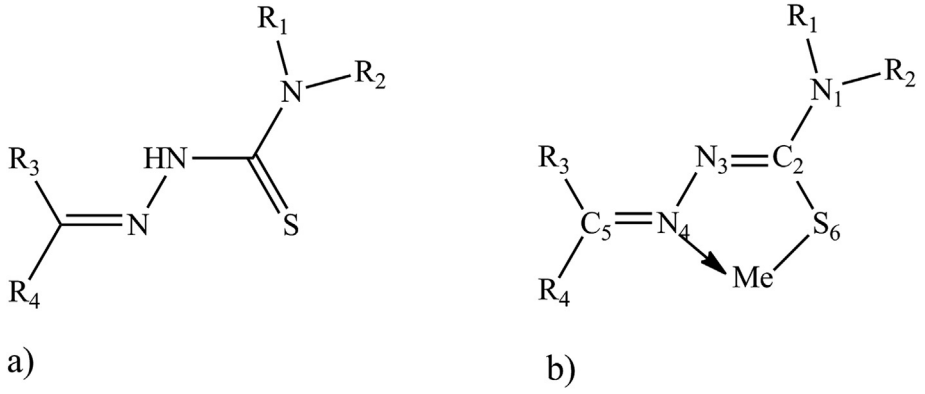


Fig. 2. Molecular skeleton: a) thiosemicarbazone ligand; b) complex of thiosemicarbazone with metal ion.

analytical techniques, such as AAS, ICP-AES, ICP-MS, X-ray fluorescence spectroscopy, spectrophotometry, and so on. Of these, the spectrophotometric method is preferred, because the it's cost is cheaper and easier to handle, and can compare the sensitivity and accuracy with others. There are many organic reagents [12,14], are used for determination of different metals by spectrophotometric method. However, they suffer from the disadvantages such as lower sensitivity and intervention from a large number of foreign ions.

Recently, the development of the sulfur-bearing ligands as thiosemicarbazones in analytical and inorganic chemistry is being interested in rapid expansion to determine the differing metal ions [11–16]. The metal complexes of reagents containing the sulfur and nitrogen donors proved the wide applicability in medicine and agriculture [2,4–6]. A survey of the literature showed a few of thiosemicarbazones employed to define the spectrophotometric database of metal ions in aqueous solution [9,10,12–14]. In the articles were published, the authors proposed the new thiosemicarbazone reagents in analytical chemistry to identify the trace amounts of metal ions by the spectrophotometer method. Those reagents also provides advantages like reliability and reproducibility as well as less interference. The development of a thiosemicarbazone ligand for the environmental and food analysis using the UV–Vis spectrophotometric method is an important task. In recent decades, the QSPR models have been developed rapidly in the field of theory chemistry to build the relationships between the metal ions with the organic ligand in the aqueous solution. Accordingly, the combination of the multivariate models and 2D and 3D molecular descriptors is also being used to develop the complexes between the thiosemicarbazone ligands with different metal ions. In many cases, the application of QSPR models is very complicated due to the statistical evaluation inadequately and the lack of modeling competence, half-finished information on the calculations of the molecular descriptors, statistical parameters, and new statistical techniques. Effective ways to overcome a large part of the problem have not been solved thoroughly.

In this work, we report the development of the hybrid QSPR modeling of logβ_{11} stability constants of the thiosemicarbazone ligands with metal ions (M = Ag^{+}, Cd^{2+}, Co^{2+}, Cu^{2+}, Fe^{3+}, Mn^{2+}, Cr^{3+}, La^{3+}, Mg^{2+}, Mo^{6+}, Nd^{3+}, Ni^{2+}, Pb^{2+}, Zn^{2+}, Pr^{3+}, Dy^{3+}, Gd^{3+}, Ho^{3+}, Sm^{3+}, Tb^{3+}, V^{5+}) in aqueous solution. The 2D and 3D

molecular descriptors of metal-thiosemicarbazone complexes are calculated to use for screening and modeling from the published database. The hybrid QSPR models are constructed by combining the genetic algorithm with multivariate linear regression methods ($QSPR_{GA\text{-}MLR}$), the support vector regression ($QSPR_{GA\text{-}SVR}$) and the artificial neural network ($QSPR_{GA\text{-}ANN}$). We could propose the new thiosemicarbazone reagents specific for the bivalent metallic cations Zn^{2+}, Cu^{2+}, Ni^{2+} and Cd^{2+}. The stability constants $\log\beta_{11}$ of the newly designed thiosemicarbazone ligands with those ions are determined by the built QSPR models.

2. Materials and methods

To implement the development of hybrid QSPR models of $\log\beta_{11}$ stability constants for the metal-thiosemicarbazone complexes we have conducted many different stages below.

2.1. Database preparation

Preparing good quality databases is a very important task that determines the success of mathematical models [17,57]. However, the preparation of experimental data in sufficient quantity and of appropriate quality for building QSPR models is a difficult screening task. The experimental $\log\beta_{11}$ stability constants of 108 M:L complexes of various thiosemicarbazones with 21 metal cations Ag^{+}, Cd^{2+}, Co^{2+}, Cu^{2+}, Fe^{3+}, Mn^{2+}, Cr^{3+}, La^{3+}, Mg^{2+}, Mo^{6+}, Nd^{3+}, Ni^{2+}, Pb^{2+}, Zn^{2+}, Pr^{3+}, Dy^{3+}, Gd^{3+}, Ho^{3+}, Sm^{3+}, Tb^{3+}, V^{5+} in aqueous solution were collected from the recent published articles [15–54]. The experimental $\log\beta_{11}$ stability constants are varied for the same complexes proposed by the different authors. The data collected were removed the outlier points with Grubb's test. This shows the data run to determine whether $\log\beta_{11}$ can be adequately modeled by a normal distribution. The Grubb's test is based upon comparing the quantiles of the fitted normal distribution to the quantiles of the data. The Grubb's test Statistic is of 2.5931; and Critical value is 3.3807. At the 95% confidence level, there is no significant outlier. The outlier points of experimental dataset were removed by the Grubb's test. The retained complexes are satisfactory for the Grubb's test and the normal distribution (Fig. 1).

The experimental $\log\beta_{11}$ stability constants are evaluated by the different ranges, as shown in Table 1. The skeleton of thiosemicarbazone ligand is chosen to form the complexes with the $\log\beta_{11}$ stability constants (Fig. 2) [59].

Most of $\log\beta_{11}$ stability constants of metal-thiosemicarbazone complexes is corrected by the temperature of 298K and at the ionic strength in the range I = 0.0 M–0.2 M to an ionic strength I = 0.1 M combining the theory of Debye-Hückel and Davies equation [55,56]. The 2D molecular structures of metal-thiosemicarbazone complexes and stability constants $\log\beta_{11}$ collected from the different materials were converted into the SDF database of 3D molecular structures in QSARIS [57,58]. The entire data set of 108 complexes for 21 metal cations with different thiosemicarbazone ligands is indicated in Table 2S.

2.2. Division of dataset

The thiosemicarbazone derivatives are different in functional groups substituting at the sites R_1, R_2, R_3 and R_4, as shown in Table 2S. The entire dataset is divided into a training set of 44 complexes, a validation set of 26 complexes and the additional test set of 30 complexes. This is an important task to construct and validate the quality of the QSPR models. The K-means clustering method [17] is used to partition randomly in the descriptors space [64,65]. In addition to the 8 lead complexes are also selected for prediction test with new metallic-thiosemicarbazone complexes, in Table 2.

Table 1
The stability constants $\log\beta_{11}$ of thiosemicarbazone ligands and metal ions are statistically based on the mean, minimum, and maximum values, respectively.

No	metal ion	M:L	mean	minimum	maximum
1	Ag^{+}	7	14.820	11.240	17.200
2	Cd^{2+}	7	6.830	4.830	10.630
3	Co^{2+}	10	8.624	5.099	11.970
4	Cu^{2+}	28	11.681	5.280	20.400
5	Fe^{3+}	6	10.649	5.496	19.480
6	Mn^{2+}	9	7.574	4.320	12.140
7	Cr^{3+}	2	7.496	4.842	10.150
8	La^{3+}	2	9.220	7.600	10.840
9	Mg^{2+}	2	3.185	3.030	3.340
10	Mo^{6+}	2	6.444	6.337	6.551
11	Nd^{3+}	2	8.520	7.950	9.090
12	Ni^{2+}	10	9.079	5.310	12.710
13	Pb^{2+}	4	6.795	5.010	8.109
14	Zn^{2+}	9	8.083	5.230	12.400
5	Pr^{3+}	2	9.400	7.760	11.040
16	Dy^{3+}	1	8.490	8.490	8.490
17	Gd^{3+}	1	8.160	8.160	8.160
18	Ho^{3+}	1	8.640	8.640	8.640
19	Sm^{3+}	1	8.260	8.260	8.260
20	Tb^{3+}	1	8.340	8.340	8.340
21	V^{5+}	1	5.322	5.322	5.322
	Sum	108			

2.3. Molecular descriptors calculation

The molecular descriptors calculation is one of the most important tasks of building process of the QSPR models [17,57]. This is an important period to quantify the structural information of the complexes used in this study [57]. The 2D experimental complexes were re-built by BIOVA Draw 2017 R2 [60] and re-optimized by the semi-empirical quantum chemistry method PM7 SCF of program MoPac2016 [61,62]. In this study, 230 molecular descriptors for each of the complexes calculated by program QSARIS [58].

2.4. Descriptors selection

In many of the current studies regarding the construction of QSPR models, one of the biggest difficulties is that the descriptor selected has a significant contribution to stability constants. In this study, we have used hybrid techniques that combine genetic algorithms with multi-parameter regression techniques. Genetic algorithms [66] are preferred to select the most important contribution descriptors to significantly reduce the number of descriptors in all 230 molecular descriptors in the entire data set. The most important meaningful molecular descriptors are chosen to be used to build QSPR models.

The parameters were used in genetic algorithm [57,58] includes the initial population size of 10, the probability for the variable to be included in the solution is 0.05, the linear ranking selection with a Toumant size of 4, the probability mating of 0.5, the one-point crossover with the number of offspring from the same parents of 2, and the probability of mutation is 0.1 with uniform mutation. In the process of selecting the descriptor, update the population with the number of all generated offspring of 6 and replace worst by 1 solution by best offspring. The fitness function uses Friedman's lack-of-fit scoring function with a parameter of 2. The Tolerance is 0.0001 and the maximum number of generations is 2000. Genetic algorithms focus on the following points: (1) Remove descriptors of the same value; (2) Remove descriptors with a standard deviation less than 0.05. (3) Remove descriptors with Pearson coefficients over 0.75. We retained 10 most significant descriptors (Table S3).

Table 2
The complexes of thiosemicarbazone ligands and metal ions with experimental and predicted stability constants logβ_{11}, respectively. The values of parentheses are the residual values from the experimental data and calculation results.

No	Thiosemicarbazone ligand				Cation	Exp.	Ref.	QSPR model		
	R_1	R_2	R_3	R_4				GA-MLR	GA-SVR	GA-ANN
1t	H	H	CH_3	C_5H_4N	Dy^{3+}	8.49	[28]	7.51(0.98)	8.275(0.22)	8.301(0.19)
2v	H	H	H	C_5H_4N	Zn^{2+}	5.23	[43]	8.59(-3.36)	4.765(3.73)	5.299(3.19)
3t	H	H	CH_3	C_5H_4N	Pr^{3+}	7.76	[28]	7.981(-0.22)	8.732(-0.24)	8.357(0.13)
4a	H	H	H	$C_6H_4NH_2$	Mn^{2+}	12.14	[47]	10.557(1.58)	11.275(-2.79)	12.125(-3.64)
5v	H	CH_3	CH_3	$C_7H_7N_2$	Cu^{2+}	12.14	[29]	11.814(0.33)	11.272(-2.78)	11.797(-3.31)
6t	H	C_6H_5	CH3	C_2H_3NOH	Cu^{2+}	6.468	[25]	6.298(0.17)	7.336(1.15)	6.03(2.46)
7t	H	H	H	$C_6H_4NO_2$	Nd^{3+}	9.09	[16]	10.185(-1.10)	9.956(-1.47)	9.13(-0.64)
8a	H	H	H	C_5H_4N	Ag^{+}	14	[32]	3.511(10.49)	14.868(-6.38)	14.58(-6.09)
9t	H	H	C_6H_4OH	C_6H_4OH	Fe^{3+}	5.496	[38]	6.898(-1.40)	6.364(2.13)	5.544(2.95)
10p	H	H	H	$C_6H_4NO_2$	Cd^{2+}	10.63	[51]	4.929(5.70)	9.759(-1.27)	10.508(-2.02)
11v	H	H	H	C_5H_4N	Co^{2+}	5.36	[43]	8.742(-3.38)	6.227(2.26)	5.392(3.10)
12t	H	H	H	$C_{13}H_{16}NO_3$	Fe^{3+}	19.48	[15]	18.885(0.59)	18.613(-10.12)	19.568(-11.08)
13a	H	H	H	$C_6H_4NH_2$	Co^{2+}	11.95	[47]	10.736(1.21)	11.189(-2.70)	12.142(-3.65)
14t	H	H	C_6H_5	C_7H_6NO	Cu^{2+}	5.748	[35]	6.216(-0.47)	6.617(1.87)	5.934(2.56)
15v	H	H	H	C_9H_5NOH	Zn^{2+}	6.68	[29]	6.388(0.29)	7.549(0.94)	6.853(1.64)
16t	CH_3	CH_3	C_5H_4N	C_5H_4N	Fe^{3+}	7.06	[27]	7.093(-0.03)	7.93(0.56)	7.395(1.10)
17a	H	H	H	C_5H_4N	Mn^{2+}	4.32	[39]	9.34(-5.02)	5.188(3.30)	3.454(5.04)
18t	H	H	H	$C_7H_7O_3$	Ni^{2+}	6.489	[21]	5.653(0.84)	6.357(2.13)	5.654(2.84)
19p	H	H	H	$CH{=}CHC_6H_5$	Cd^{2+}	5.544	[53]	10.25(-4.71)	6.413(2.08)	5.297(3.19)
20t	H	H	H	$C_6H_3OHOCH_3$	Cd^{2+}	7.07	[33]	6.849(0.22)	7.942(0.55)	6.911(1.58)
21a	H	H	H	C_6H_5	Ag^{+}	15.5	[32]	8.897(6.60)	14.633(-6.14)	15.47(-6.98)
22t	H	H	–	C_9H_8NO	Pb^{2+}	8.109	[37]	6.072(2.04)	8.979(-0.49)	7.681(0.81)
23a	H	H	–	C_9H_8NO	Cu^{2+}	9.06	[37]	9.927(-0.87)	9.659(-1.17)	8.64(-0.15)
24v	H	H	H	$C_6H_5NH_2$	Cu^{2+}	11.61	[47]	10.975(0.64)	10.741(-2.25)	11.723(-3.23)
25t	H	H	CH_3	C_5H_4N	Tb^{3+}	8.34	[28]	7.573(0.77)	9.208(-0.72)	8.35(0.14)
26t	H	H	H	$C_{10}H_6OH$	Cu^{2+}	9.34	[36]	7.308(2.03)	9.039(-0.55)	9.133(-0.64)
27a	H	H	H	C_5H_4N	Cu^{2+}	20.4	[45,46]	3.011(17.39)	19.531(-11.04)	20.846(-12.36)
28a	H	H	H	$C_6H_3OHOCH_3$	Co^{2+}	11.97	[48]	7.553(4.42)	11.104(-2.61)	12.514(-4.02)
29t	H	H	H	$C_6H_4N(CH_3)_2$	Ag^{+}	17.2	[32]	16.403(0.80)	16.328(-7.84)	17.62(-9.13)
30a	H	H	H	$C_6H_3OHOCH_3$	Ni^{2+}	12.62	[48]	7.488(5.13)	11.748(-3.26)	12.47(-3.98)
31t	H	H	H	$C_{10}H_6OH$	Co^{2+}	8.43	[36]	7.936(0.49)	9.1(-0.61)	7.523(0.97)
32t	H	H	CH_3	C_5H_4N	La^{3+}	7.6	[28]	8.529(-0.93)	8.469(0.02)	7.511(0.98)
33p	H	H	H	$C_6H_4NH_2$	Ni^{2+}	12.71	[47]	12.428(0.28)	11.841(-3.35)	12.061(-3.57)
34t	H	H	H	$C_{10}H_6OH$	Mn^{2+}	5.36	[36]	7.776(-2.42)	6.229(2.26)	5.189(3.30)
35p	H	H	–	C_9H_8NO	Ni^{2+}	8.50	[37]	2.135(6.37)	9.366(-0.88)	8.047(0.44)
36v	H	H	H	C_6H_4-N-$(CH_3)_2$	Cu^{2+}	15.3	[45,46]	15.084(0.22)	15.543(-7.05)	15.203(-6.71)
37v	H	H	H	$C_6H_3OHOCH_3$	Zn^{2+}	7.42	[33]	7.242(0.18)	8.289(0.20)	7.197(1.29)
38v	H	H	H	$C_6H_3OHOCH_3$	Cr^{3+}	4.842	[44]	7.659(-2.82)	4.937(3.55)	5.27(3.22)
39a	H	H	H	$C_6H_4NO_2$	Al^{3+}	11.24	[51]	6.391(4.85)	10.372(-1.88)	10.593(-2.10)
40t	H	H	H	$C_{10}H_6OH$	Pb^{2+}	7.23	[36]	7.023(0.21)	8.1(0.39)	6.738(1.75)
41v	H	C_2H_5	H	C_9H_5NOH	Zn^{2+}	6.13	[29]	6.928(-0.80)	6.999(1.49)	6.647(1.84)
42a	H	H	H	C_4H_3O	Co^{2+}	5.099	[49]	7.36(-2.26)	5.968(2.52)	5.089(3.40)
43t	H	H	CH_3	C_5H_4N	Ho^{3+}	8.64	[28]	7.413(1.23)	9.509(-1.02)	8.76(-0.27)
44a	H	H	CH_3	C_5H_4N	Cu^{2+}	5.491	[52]	1.625(3.87)	6.36(2.13)	5.215(3.28)
45t	H	H	–	C_9H_8NO	Zn^{2+}	8.16	[37]	8.893(-0.73)	9.029(-0.54)	7.624(0.87)
46t	H	H	H	$C_6H_3OHOCH_3$	Pb^{2+}	6.83	[33]	6.194(0.64)	7.699(0.79)	6.495(2.00)
47v	H	CH_3	CH_3	$C_7H_7N_2$	Mn^{2+}	9.91	[29]	11.277(-1.37)	9.758(-1.27)	10.074(-1.58)
48p	H	C_6H_5	H	C_9H_5NOH	Zn^{2+}	7.30	[31]	15.215(-7.92)	8.169(0.32)	7.171(1.32)
49p	H	H	H	$C_{13}H_{16}NO_3$	Zn^{2+}	12.40	[15]	18.404(-6.00)	11.532(-3.04)	13.017(-4.53)
50a	H	H	H	C_6H_4OH	Ag^{+}	15.6	[32]	4.544(11.06)	14.731(-6.24)	15.067(-6.58)
51v	H	H	CH_3	C_6H_4OH	Ni^{2+}	5.31	[34]	5.693(-0.38)	6.18(2.31)	5.183(3.31)
52v	H	H	CH_3	C_6H_4OH	Cd^{2+}	4.83	[34]	5.066(-0.24)	5.698(2.79)	6.339(2.15)
53a	H	H	CH_3	$CH{=}NNHC_6H_5$	Mo^{5+}	6.551	[30]	10.292(-3.74)	7.419(1.07)	6.602(1.89)
54a	H	H	CH_3	$CCH_3{=}N$-OH	Cu^{2+}	19.1	[45,46]	−1.59(20.69)	18.232(-9.74)	19.392(-10.90)
55a	H	H	H	C_6H_4OH	Cu^{2+}	19.1	[45,46]	5.197(13.90)	15.793(-7.30)	19.621(-11.13)
56a	H	C_2H_5	H	C_9H_5NOH	Cu^{2+}	14.67	[31]	6.696(7.97)	13.799(-5.31)	14.081(-5.59)
57t	H	H	CH_3	C_6H_4OH	Pb^{2+}	5.01	[34]	5.392(-0.38)	6.879(1.61)	5.48(3.01)
58v	H	H	CH_3	$C_7H_7N_2$	Mn^{2+}	9.87	[29]	10.606(-0.74)	9.658(-1.17)	10.062(-1.57)
59a	H	H	H	$C_{13}H_{16}NO_3$	Fe^{2+}	12.24	[15]	18.859(-6.62)	13.112(-4.62)	12.234(-3.74)
60a	H	H	H	C_6H_4OH	Cu^{2+}	17.2	[45,46]	4.383(12.82)	16.333(-7.84)	16.777(-8.29)
61t	H	H	H	$C_{10}H_6OH$	Zn^{2+}	8.11	[36]	7.48(0.63)	8.978(-0.49)	8.547(-0.06)
62v	H	H	H	$C_6H_3OHOCH_3$	Ni^{2+}	8.48	[33]	7.55(0.93)	8.843(-0.35)	7.857(0.63)
63t	H	H	CH_3	$C_7H_7N_2$	Co^{2+}	10.22	[26]	10.89(-0.67)	9.658(-1.17)	9.205(-0.72)
64a	H	H	H	$C_6H_3OHOCH_3$	Cu^{2+}	13.33	[48]	7.116(6.21)	12.461(-3.97)	10.185(-1.70)
65t	H	H	H	$C_8H_9O_3$	Cu^{2+}	6.236	[24]	6.681(-0.45)	6.104(2.39)	6.697(1.79)
66t	H	H	H	$C_{10}H_6OH$	Cd^{2+}	6.47	[36]	7.085(-0.62)	7.339(1.15)	7.114(1.38)
67p	H	H	CH_3	C_5H_4N	Cu^{2+}	5.924	[52]	10.235(-4.31)	6.792(1.70)	5.867(2.62)
68v	H	H	CH_3	C_6H_4OH	Mn^{2+}	4.51	[34]	6.167(-1.66)	5.382(3.11)	4.206(4.28)
69t	H	C_6H_5	H	C_9H_6NO	Cu^{2+}	15.65	[31]	15.25(0.40)	14.782(-6.29)	15.169(-6.68)
70v	H	H	H	$C_6H_3OHOCH_3$	Mn^{2+}	5.28	[33]	7.616(-2.34)	4.41(4.08)	5.081(3.41)
71v	H	H	H	C_6H_4OH	Cu^{2+}	5.28	[40]	5.197(0.08)	7.793(0.70)	5.621(2.87)
72t	H	H	CH_3	C_5H_4N	Nd^{3+}	7.95	[28]	8.073(-0.12)	8.667(-0.18)	8.324(0.17)

Table 2 (*continued*)

No	Thiosemicarbazone ligand				Cation	Exp.	Ref.	QSPR model		
	R_1	R_2	R_3	R_4				GA-MLR	GA-SVR	GA-ANN
73t	H	H	CH_3	C_5H_4N	Sm^{3+}	8.26	[28]	7.812(0.45)	8.883(-0.39)	8.426(0.06)
74v	H	H	H	C_5H_4N	Ni^{2+}	5.63	[39]	8.881(-3.25)	6.499(1.99)	5.011(3.48)
75a	H	H	H	$C_6H_4NO_2$	Fe^{3+}	11.63	[51]	5.664(5.97)	10.761(-2.27)	10.746(-2.26)
76a	H	H	H	C_6H_5	Cu^{2+}	17.7	[45,46]]	8.975(8.73)	16.83(-8.34)	16.807(-8.32)
77t	H	H	CH_3	C_5H_4N	Gd^{3+}	8.16	[28]	7.814(0.35)	8.934(-0.44)	8.41(0.08)
78t	H	H	CH3	$C_7H_7N_2$	Ni^{2+}	10.89	[26]	12.405(-1.52)	10.021(-1.53)	10.274(-1.78)
79t	H	H	H	$C_{10}H_6OH$	Mg^{2+}	3.34	[36]	5.328(-1.99)	4.208(4.28)	2.754(5.74)
80t	H	H	–	C_9H_8NO	Cd^{2+}	7.409	[37]	8.666(-1.26)	8.279(0.21)	7.446(1.04)
81v	H	H	H	$C_6H_3OHOCH_3$	Co^{2+}	8.02	[33]	7.641(0.38)	7.149(1.34)	8.144(0.35)
82a	H	H	H	$C_5H_3NCH_3$	Cu^{2+}	19.1	[45,46]	3.188(15.91)	18.229(-9.74)	19.152(-10.66)
83t	H	H	H	$C_6H_4NO_2$	La^{3+}	10.84	[16]	10.933(-0.09)	10.46(-1.97)	10.25(-1.76)
84v	H	H	CH_3	C_6H_4OH	Cu^{2+}	5.91	[34]	5.531(0.38)	6.302(2.19)	5.411(3.08)
85p	H	CH_3	CH_3	C_5H_4N	Cu^{2+}	6.114	[54]	2.442(3.67)	6.984(1.51)	6.382(2.11)
86t	H	H	H	$C_{14}H_{12}N$	Cd^{2+}	5.86	[30]	5.365(0.50)	6.728(1.76)	5.803(2.69)
87t	H	H	H	$C_{13}H_{16}NO_3$	Cu^{2+}	17.54	[15]	17.506(0.03)	16.67(-8.18)	17.754(-9.26)
88v	H	H	H	$C_6H_3OHOCH_3$	Cu^{2+}	9.44	[33]	7.162(2.28)	8.807(-0.32)	10.067(-1.58)
89a	H	H	H	$C_5H_3NCH_3$	Ag^{+}	14.5	[32]	3.375(11.13)	15.026(-6.54)	14.647(-6.16)
90v	H	H	H	C_6H_3BrOH	Cu^{2+}	5.633	[42]	3.593(2.04)	6.501(1.99)	6.832(1.66)
91a	H	H	H	C_6H_4OH	Ag^{+}	15.7	[32]	6.033(9.67)	14.833(-6.34)	15.36(-6.87)
92t	H	H	H	$C_7H_7O_3$	Co^{2+}	6.382	[22]	5.107(1.28)	7.25(1.24)	6.137(2.35)
93t	CH_3	CH_3	C_5H_4N	C_5H_4N	Cu^{2+}	7.08	[27]	6.659(0.42)	7.947(0.54)	7.304(1.19)
94a	H	H	H	$C_6H_3OHOCH_3$	Mn^{2+}	10.55	[48]	7.778(2.77)	9.678(-1.19)	9.688(-1.20)
95a	H	H	H	C_9H_5NOH	Cu^{2+}	14.56	[31]	6.453(8.11)	13.692(-5.20)	14.782(-6.29)
96t	H	H	H	$C_{10}H_6OH$	Ni^{2+}	9.13	[36]	7.74(1.39)	9.075(-0.58)	9.501(-1.01)
97a	H	H	CH_3	$CH{=}NNHC_6H_5$	Cu^{2+}	11.95	[26]	7.746(4.20)	11.082(-2.59)	12.105(-3.62)
98t	H	H	–	C_9H_8NO	Mn^{2+}	6.23	[37]	6.443(-0.21)	7.098(1.39)	6.243(2.25)
99v	H	CH_3	CH_3	$C_7H_7N_2$	Co^{2+}	10.47	[29]	10.982(-0.51)	9.658(-1.17)	10.407(-1.92)
100t	H	CH_3	CH_3	$C_7H_7N_2$	Ni^{2+}	11.03	[29]	11.777(-0.75)	10.161(-1.67)	10.616(-2.13)
101v	H	H	CH_3	$C_6H_5NH_2$	Zn^{2+}	11.32	[47]	10.866(0.45)	10.452(-1.96)	12.244(-3.75)
102t	H	H	H	$C_6H_4NO_2$	Pr^{3+}	11.04	[16]	10.237(0.80)	10.171(-1.68)	10.11(-1.62)
103v	H	H	H	$C_6H_3OHOCH_3$	Fe^{2+}	7.99	[33]	7.544(0.45)	8.859(-0.37)	7.638(0.85)
104t	H	H	–	C_9H_8NO	Co^{2+}	8.34	[37]	8.011(0.33)	9.206(-0.72)	7.455(1.04)
105a	H	H	CH_3	C_6H_4OH	Mg^{2+}	3.03	[34]	2.502(0.53)	3.899(4.59)	2.775(5.72)
106v	H	H	H	C_6H_4OH	V^{5+}	5.322	[41]	5.793(-0.47)	6.191(2.30)	5.354(3.14)
107t	H	H	H	$C_8H_9O_3$	Mo^{5+}	6.337	[23]	7.278(-0.94)	7.205(1.29)	5.551(2.94)
108a	H	H	H	$C_6H_4NO_2$	Cr^{3+}	10.15	[51]	6.217(3.93)	9.925(-1.44)	10.916(-2.43)

t: training set; v: validation set; a: additional test set; p: prediction lead complexes

The $QSPR_{GA\text{-}MLR}$ models were constructed by changing the number of descriptors k. Thus the descriptors are reduced by more than 95.6% of the entire descriptors in the selection step; (4) Finally, the multiple linear regression technique [63] is used to remove further descriptors that have the insignificant effect on the predictability of QSPR model. So the $QSPR_{GA\text{-}MLR}$ model with $k = 7$ seem to be most appropriate (Table 3) for development of different QSPR models.

2.5. Development of QSPR model

2.5.1. Regression model

The significant-contribution descriptors are retained by the genetic algorithm to build the $QSPR_{GA\text{-}MLR}$ model using the multivariate linear regression (MLR) technique [17–19]. For a given dataset (x_i, y_i), $i = 1, 2, \ldots n$. where x is the descriptor and y is stability constant; β_0 and β_1 are coefficients, and ε_i is a random error term with mean.

$$y_i = \beta_0 + \beta_1 x_i + \beta_i \quad (1)$$

2.5.2. Support vector regression model

The support vector regression (SVR) technique is also operated to construct the $QSPR_{GA\text{-}SVR}$ relationship models that map nonlinear input data into a high dimensional space. The account theory of support vector machine regression is presented in several materials [70–73]. In this work the training set of 44 complexes with known $\log\beta_{11}$ values y_i and selected descriptors x_i are

Table 3
The statistical parameters and the selected descriptors of $QSPR_{GA\text{-}MLR}$ models, respectively.

k	selected descriptors	R^2	R^2_{adj}	MSE	Q^2
1	Surface	0.261	0.239	9.669	0.143
2	LogP/Surface	0.445	0.412	5.675	0.260
3	LogP/Ovality/Surface	0.409	0.353	6.776	0.229
4	LogP/SaasC/Ovality/Surface	0.567	0.512	6.366	0.461
5	LogP/SaasC/Ovality/Surface/nelem	0.715	0.667	3.987	0.588
6	xp5/SaasC/ABSQ/Ovality/Surface/nrings	0.718	0.659	3.690	0.629
7	**xp3/xp5/SaasC/Ovality/Surface/nelem/nrings**	**0.915**	**0.893**	**1.290**	**0.865**
8	xp3/xp5/SaasC/ABSQ/Ovality/Surface/nelem/nrings	0.906	0.878	1.301	0.837
9	xp3/xp5/xvch8/SaasC/ABSQ/Ovality/Surface/nelem/nrings	0.949	0.931	0.870	0.899
10	LogP/xp3/xp5/xvch8/SaasC/ABSQ/Ovality/Surface/nelem/nrings	0.958	0.941	0.671	0.892

represented by the correlation $y_i = f(x_i)$. There are several kernels described non-linear transformations of higher dimensional space. Basically the radial basis function (RBF) kernel could be utilized to delve out the nonlinear input data by the following equation

$$K(x,y) = \exp\left(-\gamma\|x-y\|^2\right) \quad (2)$$

This RBF function is used for the new feature space separated out by hyperplanes which it minimizes the distance between the data set.

2.5.3. Artificial neural network model

To perform neural network construction, we proceed to process the smallest number of descriptors possible. This is a challenge regarding the selection of the number of molecular descriptors. Genetic algorithms are used to overcome this difficulty to choose the actual and the least descriptors set; the artificial neural networks are built on the basis of those.

The genetic algorithm parameters used in the selection process of input descriptors such as smoothing of 0.01; a unit penalty of 0.001; population size of 50; a crossover rate of 0.9; a mutation rate of 0.1 and generations number of 50; iterations total of 50. To avoid the overfitted models the data set was randomly divided into two subsets (85%) for the training phase and (15%) for the internal validation phase of the model [75,76]; We used the neural network style MLP-ANN [77]. A back-propagation error method and the Levenberg–Marquardt algorithm are used for training process of neural network [74–76]. The neural network architectures I(k)-HL(m)-O(1) consist of an input layer I(k) with k input neurons as input descriptors, a hidden layer HL(m) with m hidden neurons and an output layer O(1) with 1 neuron as stability constant $\log\beta_{11}$. The transfer functions such as sigmoidal function and hyperbolic tangent function in program Matlab version 2018 are used for training the neural network [74]. The number of neurons in hidden layer is determined from 2 to 6. Therefore we can use the simple rule below:

$$0.5\times(k-l) \le m \le 0.5\times(k+l) \quad (3)$$

where k is the number of input neurons; m is the number of hidden neurons; l is the number of layers in neural network.

2.6. Validation of QSPR models

In order to validate the quality of QSPR models, the statistical parameters and the coefficient of determination (R^2), the adjusted coefficient of determination (R^2_{adj}), the leave-one-out cross-validation coefficient (Q^2_{LOO}) and mean-square error (MSE) [17–20,67–69] are used to determine the predictability of the constructed QSPR models. The Q^2 value of a QSPR model is more than the stipulated value of 0.6, then the QSPR model is considered to be well predictive; the mean of the absolute-relative error (*MARE*,%) and average percentage contribution ($APC_{m,n,xi}$,%) are employed to appreciate the significant contribution descriptors [57,58] and most important QSPR models.

The predictability of the models was also validated by the mean percentage of absolute-relative error (*MARE*,%) and average percentage contribution ($APC_{m,n,xi}$,%) of molecular descriptors [57,78,79], which these are calculated by following formula

$$MARE,\% = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i-\hat{y}_i|}{y_i}100\% \quad (4)$$

$$APC_{m,n,x_i},\% = \frac{1}{n}\left(\frac{1}{m}\sum_{j=1}^{m}\frac{|b_i x_i|}{\sum_{i=1}^{k}|b_i x_i|}100\%\right) \quad (5)$$

Here, n refers to the number of complexes in the training set; x_i are descriptors ith; k are the number of selected descriptors in QSPR model; m is the number of selected models.

3. Results and discussion

3.1. The $QSPR_{GA\text{-}MLR}$ model

The $\log\beta_{11}$ values of complexes differed in the maximum range from 3.340 to 19.480 for Mg^{2+} and Fe^{3+}, in the minimum range from 3.030 to 11.240 for Mg^{2+} and Ag^{+}, in the mean range of 3.185–14.820 for Mg^{2+} and Ag^{+} (Table 1). The $QSPR_{GA\text{-}MLR}$ modeling was performed for $\log\beta_{11}$ values of the diverse ML complexes for metal cations (M = Cu^{2+}, Fe^{3+}, Ni^{2+}, Co^{2+}, Cr^{3+}, Mo^{6+}, La^{3+}, Pr^{3+}, Nd^{3+}, Gd^{3+}, Sm^{3+}, Tb^{3+}, Dy^{3+}, Ho^{3+}, Cd^{2+}, Ag^{+}, Pb^{2+}, Mg^{2+}, Mn^{2+}, Zn^{2+}, V^{5+}) and the structural descriptors (Table 2). The $QSPR_{GA\text{-}MLR}$ models are screened by fitting and cross-validation ability when the number of descriptors k changes from 1 to 10. So the statistical values R^2, R^2_{adj} and Q^2 increase and the MSE values decrease. Accordingly the most significant model seem to be the $QSPR_{GA\text{-}MLR}$ model (with $k = 7$) with an optimal subset of 7 descriptors, which involves the significant statistical values of R^2, R^2_{adj}, MSE and Q^2 (Table 3). The molecular descriptors consist of xp3, xp5, SaasC, Ovality, Surface, nelem, and nrings. The appropriate model $QSPR_{GA\text{-}MLR}$ is the following model:

$$\log\beta_{11} = 46.4335 + 5.3211\times xp3 - 9.9711\times xp5 + 2.9632\times SaasC - 32.0753\times Ovality + 0.0707\times Surface - 4.4522\times nelem + 7.2474\times nrings \quad (6)$$

$R^2 = 0.9145$; $R^2_{adj} = 0.8932$; $Q^2_{LOO} = 0.8650$; MSE. = 1.2899; RMSE = 1.1357; Durbin-Watson statistic = 1.0434.

Since the P values is less than the significant level 0.05, so those interpreted the statistically significant relationship of the descriptors. The R^2 value of 0.9145 indicates that the $QSPR_{GA\text{-}MLR}$ model (6) with $k = 7$ as fitted explains 91.45% of the variability in $\log\beta_{11}$. The R^2_{adj} statistic of 0.8932, which is more suitable for comparing models with different numbers of predictors, is 89.32%. The mean-squared error (MSE) of 1.2899 is the average value of the residuals. In determining whether the model can be simplified, notice that the highest P-value on the descriptors is 0.0000. Consequently, there is no desire to remove any descriptors from the $QSPR_{GA\text{-}MLR}$ model (6).

The statistical values of seven screened descriptors of $QSPR_{GA\text{-}MLR}$ model (6) presented the significant confidence at 95% level (Table 4). The significant average percentage contribution

Table 4
The statistical parameters of the descriptors in the $QSPR_{GA\text{-}MLR}$ model (6) with $k = 7$.

Source	Coefficient	Standard error	t-Stat	P-value	Max$APC_{m,n,xi}$,%
Intercept	46.4335	6.3224	7.3443	0.0000	
xp5	−9.9711	1.2540	−7.9517	0.0000	33.5080
Ovality	−32.0753	4.4403	−7.2236	0.0000	24.2053
xp3	5.3211	0.9068	5.8677	0.0000	18.3579
nrings	7.2474	1.2785	5.6687	0.0000	17.2673
Surface	0.0707	0.0131	5.3797	0.0000	12.7214
nelem	−4.4522	0.5877	−7.5760	0.0000	9.7382
SaasC	2.9632	0.2938	10.0845	0.0000	4.2656

($APC_{m,n,xi}$,%) to the logβ_{11} value of each descriptor is estimated by using formula (5) for the $QSPR_{GA\text{-}MLR}$ model (6).

Besides the average percentage contribution values ($APC_{m,n,xi}$,%) of 10 selected descriptors resulting from the training set using $QSPR_{GA\text{-}MLR}$ model with $k = 10$ (Table S5) are sorted descending according to the maximum percentage contribution ranging from 33.51% to 1.0% such as xp5 (33.51%) > Ovality (24.21%) > xp3(18.36%) > nrings (17.27%) > Surface (12.72%) > nelem (9.74%) > SaasC (4.27%) > ABSQ (4.07%) > logP(2.21%) > xvch8 (0.96%), as shown in Fig. 3. Herein the average percentage contribution of ABSQ (2.38%), xvch8 (1.55%) and logP (0.25%) presented an insignificant contribution for stability constant logβ_{11}, so those were not prioritized for the $QSPR_{GA\text{-}MLR}$ model (6). This information may also be useful in a new complex design. The xp5, Ovality, xp3 and nrings descriptors are utilized for new reagent design due to these exhibited the most significant contribution to the stability constant logβ_{11}.

The 2D descriptors xp5, xp3 and nrings, and the 3D descriptor Ovality are the most significant descriptors, so we found that the stability constants logβ_{11} of the complexes depend mainly on the simple 5th-order and 3rd-order path chi index level and number of rings in a molecule R = 1p - (nvx - 1) as well as 3D descriptor Ovality calculated as Surface/$4\pi R^2$. We could rely on these descriptors to collect the appropriate ligands or design the new ligands to produce more stability complexes with metal ions. So we can orient the development of new ligands towards the greatest contribution of xp5, xp3, nrings and Ovality descriptions. We can express the relationship between the stability constant logβ_{11} versus the metal-thiosemicarbazone ML complexes and the contribution $APC_{m,n,xi}$,% of the descriptors xp5, xp3, nrings and Ovality, as depicted in Fig. 4.

We found that the most complexes of $Fe^{3+}L$, $Cu^{2+}L$, $Ni^{2+}L$, $Ag^{+}L$ and $Co^{2+}L$ presented the high stability constants logβ_{11}, respectively. Thus, we could use these characteristics to develop the new thiosemicarbazone structure which it can generate more stability complexes with metal cations. And these may also be used to identify the metal ions Ni^{2+}, Cu^{2+}, Fe^{3+}, Ag^{+}, and Co^{2+} in environmental samples by UV–Vis spectrophotometric method.

3.2. The $QSPR_{GA\text{-}SVR}$ model

Along with the development of $QSPR_{GA\text{-}MLR}$ model (6), the support vector regression (SVR) method is also employed to produce the high predictable model. The predictors xp5, Ovality, xp3, nrings, Surface, nelem and SaasC were also operated to construct the $QSPR_{GA\text{-}SVR}$ model. Due to the nonlinear data, so we conducted the surveys of the radial basis function (RBF) [71–73] to construct the $QSPR_{GA\text{-}SVR}$ model. The values Capacity (C), the Gamma (γ), epsilon (ε) were searched by the intensity grid search method. An error surface is optimized by multi-level technique using the genetic algorithms. The minimum region of root error (RMSECV) values and the maximum region of the values R^2 were spanned by the 5-level parameters Capacity (C) and Gamma (g), as given in Table S6.

The optimal parameters reached out as Capacity (C) of 1.0, Gamma (γ) of 1.0 and epsilon = 0.1 with number of support vectors = 27 are selected in the optimal region. These can carry the relative importance weight of the regression error, which it found the appropriable coefficient R^2 of 0.9269 and value RMSECV of 2.0942 (Table S6). The optimal region defines the most significant parameters, as described in Fig. 5. The Q^2 value of 0.6414 is more than the stipulated value of 0.6. So this $QSPR_{GA\text{-}SVR}$ model may well predict. The logβ_{11} values of complexes of the validation and additional test set can be estimated by the $QSPR_{GA\text{-}SVR}$ model (Table 2). The correlation of the calculation results derived from the $QSPR_{GA\text{-}SVR}$ model versus those from experimental data represents in statistical values R^2, as depicted in Fig. 6. The calculated stability constants found in uncertainty range of experimental measurements at 95% confidence. The dissimilarity between the experimental and calculated stability constants of complexes is acceptable.

3.3. The $QSPR_{GA\text{-}ANN}$ model

In order to continue to develop the good predictable QSPR model for the logβ_{11} stability constants of metal-thiosemicarbazone complexes, the neural network model $QSPR_{GA\text{-}ANN}$ I(k)-HL(m)-O(1) used involves the neurons of the input layer as xp3, xp5, SaasC, Ovality, Surface, nelem, and nrings. These also are in $QSPR_{GA\text{-}MLR}$

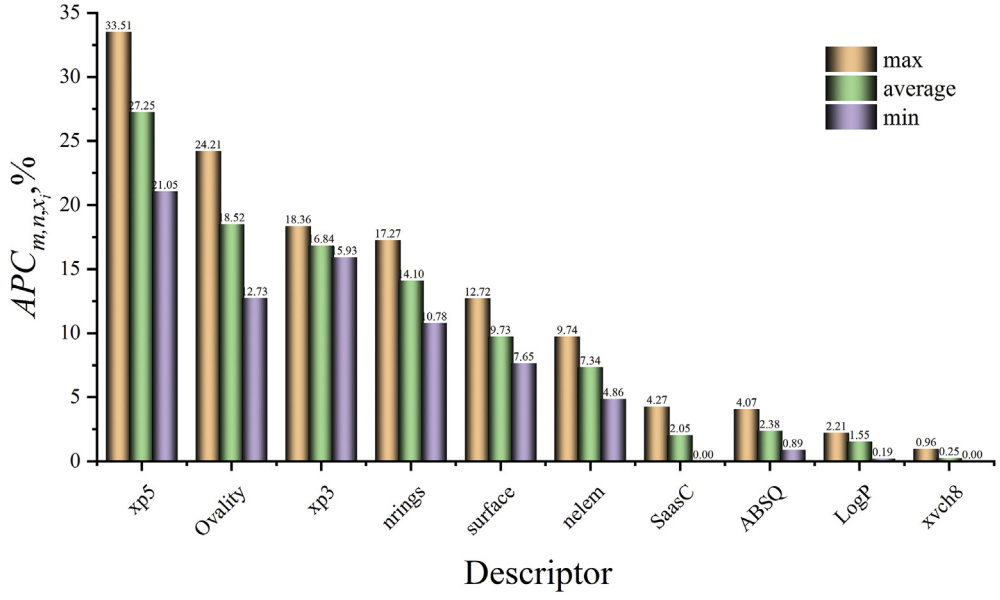


Fig. 3. Representation of minimum, maximum and average percentage contribution ($APC_{m,n,xi}$,%) for $QSPR_{GA\text{-}MLR}$ model with $k = 10$ and 44 complexes of training set.

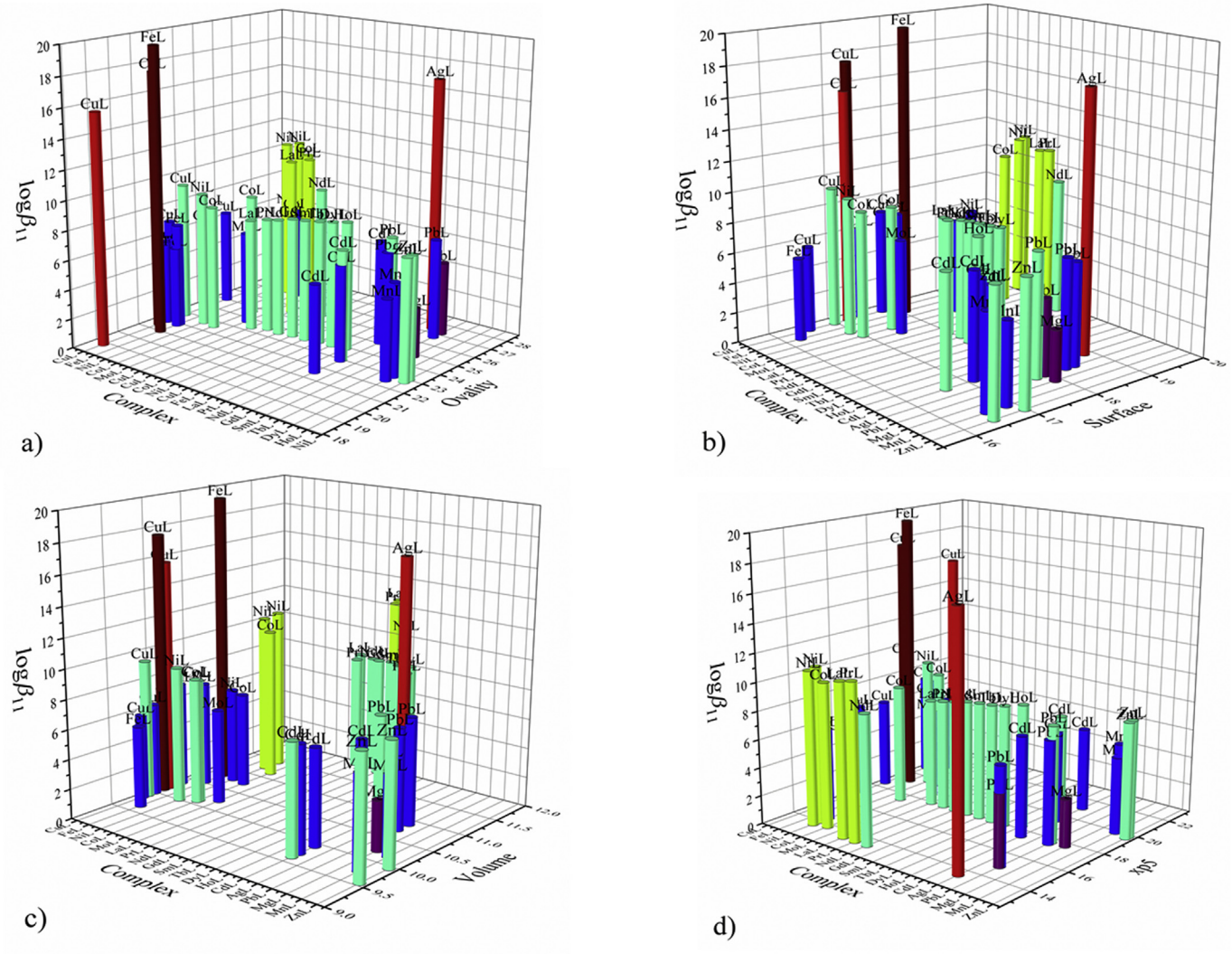


Fig. 4. The relationship between the stability constants logβ_{11} versus ML complexes and contribution $APC_{m,n,xi}$,% of descriptors: a) xp5; b) xp3; c) nrings and d) Ovality.

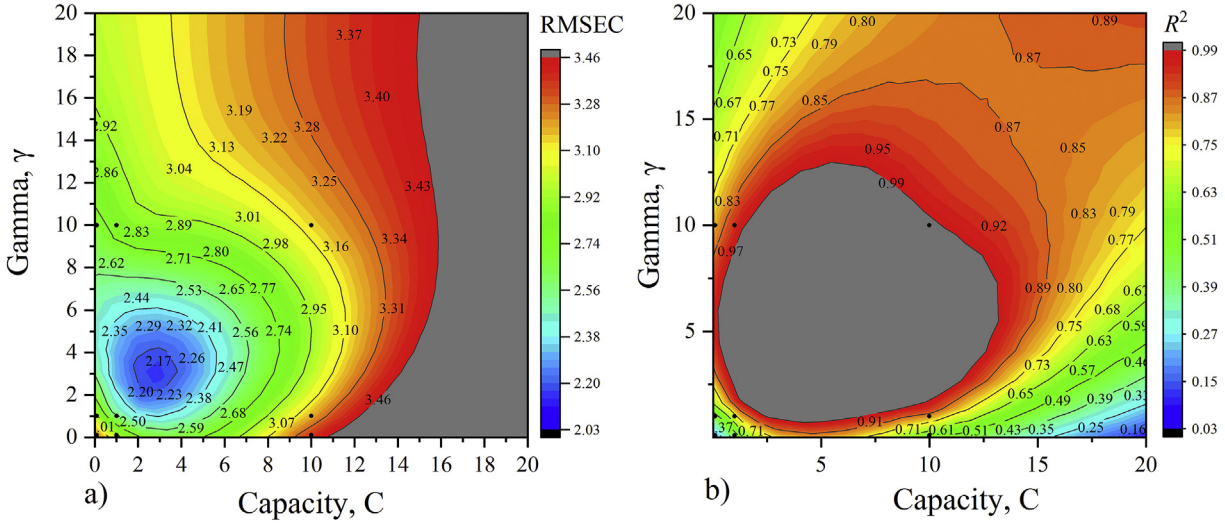


Fig. 5. Contour plots for searching 5-level parameters Gamma, γ and Capacity, C; a) The optimal area of the RMSEC values; b) The optimal area of the R^2 values.

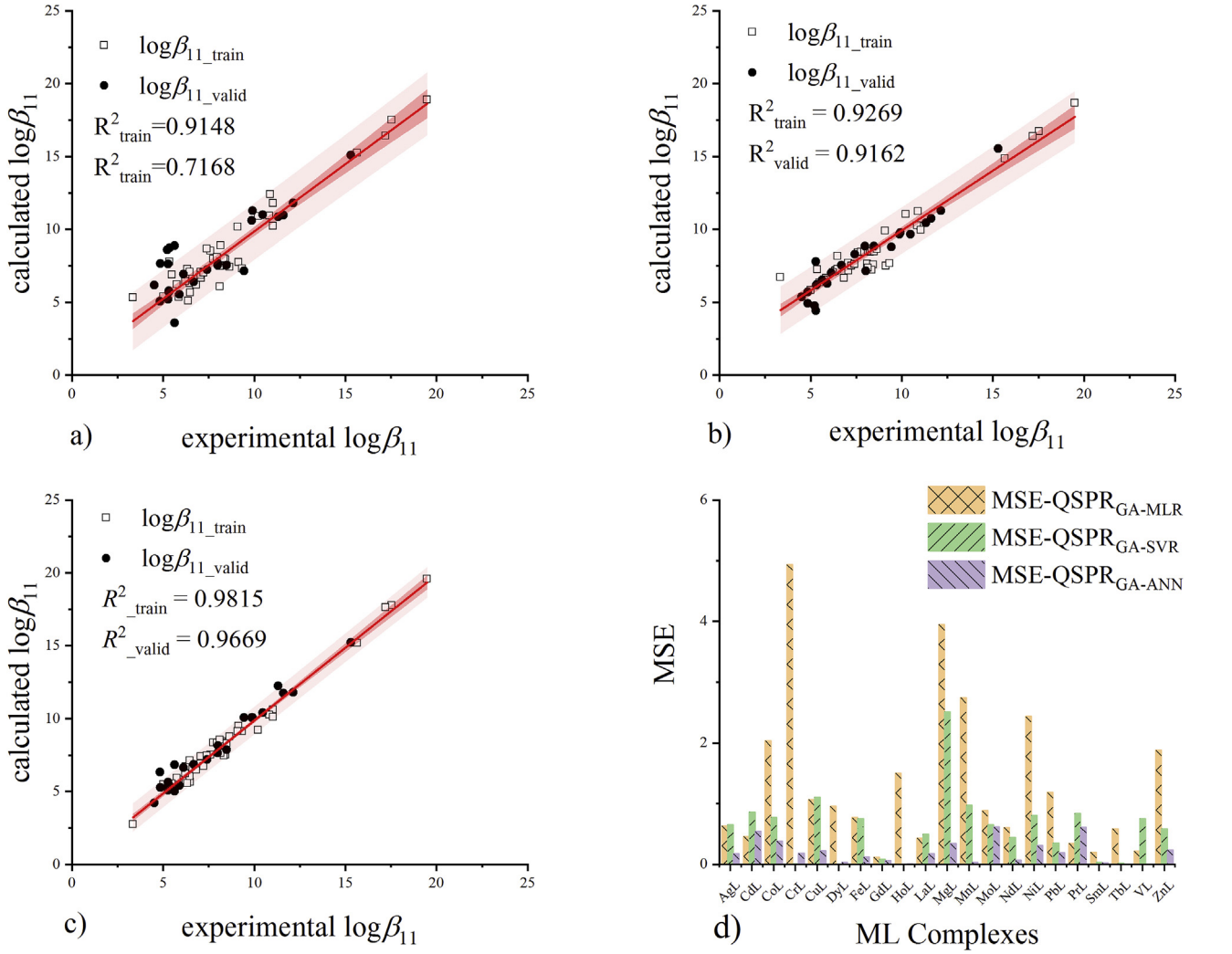


Fig. 6. The correlation between experimental versus calculated logβ_{11} stability constants of complexes of training and validation set (Table 2); a) QSPR$_{GA-MLR}$; b) QSPR$_{GA-SVR}$; c) QSPR$_{GA-ANN}$; d) MSE values for complexes from QSPR models.

model (6) with $k = 7$. The neurons of the hidden layer considered to vary from 3 to 5 according to the rule (3). The output neuron is the stability constant logβ_{11}. Every neuron on any layer is fully connected to the neurons of the next layer. Input and output data of the neural network are normalized between 0 and 1. The learning rate is set from 1 and decreases during training. The selected QSPR$_{GA-ANN}$ model with neural network architecture I(7)-HL(5)-O(1) is suitable.

The correlation between experimental and the estimated stability constants resulting from the models expressed the predictability of QSPR models with the high statistics R^2 and Q^2 (Fig. 6). It found that the calculated results are in a good agreement with the experimental data. Although, the complexes in the validation set are not used for the building process of QSPR models.

Three constructed QSPR models demonstrate the predictability with the negligible errors MSE and *MARE*, %. Thus, these QSPR models turn out the confidently applicability for predicting the stability constants logβ_{11}. The QSPR$_{GA-ANN}$ model depicted the best predictability. Contrariwise the QSPR$_{GA-MLR}$ model exhibits the lowest predictability with the largest error values. This difference can be also found by comparing the QSPR models based on the statistical values of them, (see in Table 5).

Table 5
The statistical properties of the QSPR models for stability logβ_{11} constants.

Method	Data set	*R*	R^2	Q^2	MSE	*MARE*, %
QSPR$_{GA-MLR}$	Training	0.9565	0.9148	0.8650	1.2898	10.7076
	Validation	0.8466		0.7168	2.4906	19.0119
	Test	0.8921		0.7958	4.4894	13.7829
	Prediction	0.3556		0.1264	28.2505	63.1614
QSPR$_{GA-SVR}$	Training	0.9628	0.9269	0.6414	0.9559	11.4975
	Validation	0.9572		0.9162	0.7730	11.7517
	Test	0.9853		0.9708	1.0357	8.4924
	Prediction	0.9873		0.9747	0.7547	11.0850
QSPR$_{GA-ANN}$	Training	0.9907	0.9815	0.9317	0.2209	4.9796
	Validation	0.9833		0.9669	0.2742	5.9019
	Test	0.9884		0.9769	0.5520	4.3756
	Prediction			0.9819	0.1468	3.5159

3.4. Estimate of stability constants

In order to estimate the stability constants logβ_{11} of complexes as well as to assess more the predictability of the QSPR models, we produced the stability constants logβ_{11} of 30 complexes of the additional test set from the QSPR models (Table 2). The prediction quality of the QSPR models represented in the statistical values R^2,

Q^2, MSE and *MARE*,% (Table 5). The thiosemicarbazone reagents with metal cations Mn^{2+}, Zn^{2+}, Fe^{2+}, Cd^{2+}, Cu^{2+}, Ni^{2+}, Co^{2+}, Mo^{5+}, Ag^{+}, Mg^{2+}, Al^{3+}, Cr^{3+}, Fe^{3+} of the additional test set have not also been used in the QSPR modeling process. Hereinbefore, we found that the descriptors xp5 and xp3, Ovality and nrings influenced greatly the structural properties, so the stability constants of complexes are also impacted. Thereupon, we could conduct the design and synthetic way for new thiosemicarbazone reagents based on the significant contribution of those descriptors. In this work the two new thiosemicarbazone ligands were designed by substituting the R_4 group with the larger aromatic hetero-ring groups to increase the contribution ability of the descriptors xp5, xp3, nrings, Ovality and Surface (Table 6). From this orientation, these two new thiosemicarbazone ligands as reagents were synthesized in our laboratory (Fig. S8). So the new complexes can constitute by complexation of new reagents with metal cations Cu^{2+}, Ni^{2+}, Zn^{2+} and Cd^{2+} which may be used to determine those ions in environmental samples by UV–Vis method (see also Fig. S8). We selected the eight prediction lead complexes of 4 metal cations Cu^{2+}, Ni^{2+}, Zn^{2+} and Cd^{2+} (Table 2) which these are employed to evaluate with our synthesized complexes. The lead complexes are also not used in the QSPR modeling process. The $\log\beta_{11}$ values of all those complexes for metal cations Ni^{2+}, Cu^{2+}, Cd^{2+} and Zn^{2+} were estimated by using three QSPR models (Tables 2 and 6).

The prediction $\log\beta_{11}$ values of the lead complexes from the $QSPR_{GA\text{-}SVR}$ and $QSPR_{GA\text{-}ANN}$ model are close to the experimental data. But those from the $QSPR_{GA\text{-}MLR}$ model are larger errors. Accordingly, this can be suitable way for the development of the QSPR models from the available stability constants of complexes due to it can allow to screen the metal-thiosemicarbazone complexes meaningfully.

In addition, we could also look for other ways to determine the stability constants based on the correlation between the experimental and predicted stability constants $\log\beta_{11}$ for each individual ion Cu^{2+}, Zn^{2+}, Cd^{2+} and Ni^{2+}. This can be found that the calculation results of each complex $Cu^{2+}L$, $Zn^{2+}L$, $Cd^{2+}L$ and $Ni^{2+}L$ over training, validation and additional test set resulting from the $QSPR_{GA\text{-}SVR}$ and $QSPR_{GA\text{-}ANN}$ model can be used to establish the correlation equations. In this case the values R^2 are in range 0.8933–0.9766 for the $QSPR_{GA\text{-}SVR}$ model, and in range 0.8897–0.9836 for the $QSPR_{GA\text{-}ANN}$ model (as in Fig. 7). In the similar way, the stability constants of new complexes can also be interpolated by these correlation equations of each individual ion Ni^{2+}, Cd^{2+}, Cu^{2+} and Zn^{2+} (Table 7) based on the correlation rule of predictability domain, respectively. This can also be the results of further evaluation of what has been achieved from the $QSPR_{GA\text{-}SVR}$ and $QSPR_{GA\text{-}ANN}$ models for lead and new complexes.

The interesting issue here is that we could select the complexes that can be used for designing new reagents. The stability constants of the eight lead complexes (Table 2) and the four new complexes $Cu^{2+}L$, $Zn^{2+}L$, $Cd^{2+}L$ and $Ni^{2+}L$ derived from the QSPR models are compared to each other, as given in Fig. 8.

The prediction stability constants of new complexes presented are higher than lead complexes. So we believe that the new complexes also could be satisfy the reagent demand in analytical chemistry. For the lead and new complexes the $\log\beta_{11}$ values resulting from the correlation equations turn out also to be in a good agreement with those from the $QSPR_{GA\text{-}SVR}$ and $QSPR_{GA\text{-}ANN}$ model and experimental data. This is consistent with our consideration for design of new reagents based on the significant contribution of xp5, xp3, Ovality and nrings.

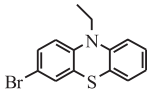
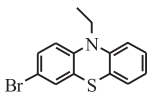
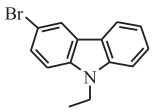
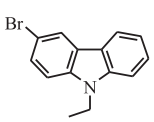
The stability constants $\log\beta_{11}$ of the new complexes found are close to the correlation line of eight lead complexes (Fig. 9). The predicted $\log\beta_{11}$ values are in a good agreement with experimental data in statistical values $Q^2_{pred} = 0.9455$ for $QSPR_{GA\text{-}SVR}$ and Q^2_{pred} for $QSPR_{GA\text{-}ANN}$. These $\log\beta_{11}$ values are in uncertainty range of experimental measurement at 95% confident level.

4. Discussion

This paper reports the novel QSPR models for the $\log\beta_{11}$ stability constants of ML complexes of several metal ions with the thiosemicarbazone reagents. Based on the survey results have been received above we could have the following discussion:

The $QSPR_{GA\text{-}MLR}$ models have been described by the correlation equations between the stability constants and the molecular descriptors. The appropriate statistical parameters R^2, R^2_{adj}, Q^2 and MSE are used effectively to select the correct correlation models $QSPR_{GA\text{-}MLR}$ including a small number to large descriptors (in Table 3). Also the regression techniques combined with support vector machine and neural networks were used to screen the descriptors of complex molecules, V. Solov et al. successfully applied

Table 6
The predicted $\log\beta_{11}$ stability constants of new complexes for metal cations Zn^{2+}, Cd^{2+}, Cu^{2+} and Ni^{2+} using the QSPR models, respectively.

No	Thiosemicarbazone ligand				cation	Ref.	QSPR model		
	R_1	R_2	R_3	R_4			MLR	SVR	ANN
109n	H	H	H		Ni^{2+}	This work	28.9072	19.7739	19.3169
110n	H	H	H		Cd^{2+}	This work	29.4325	18.6960	19.3283
111n	H	H	H		Cu^{2+}	This work	12.0367	17.9329	18.8385
112n	H	H	H		Zn^{2+}	This work	12.1104	15.0341	18.7494

n: new complexes

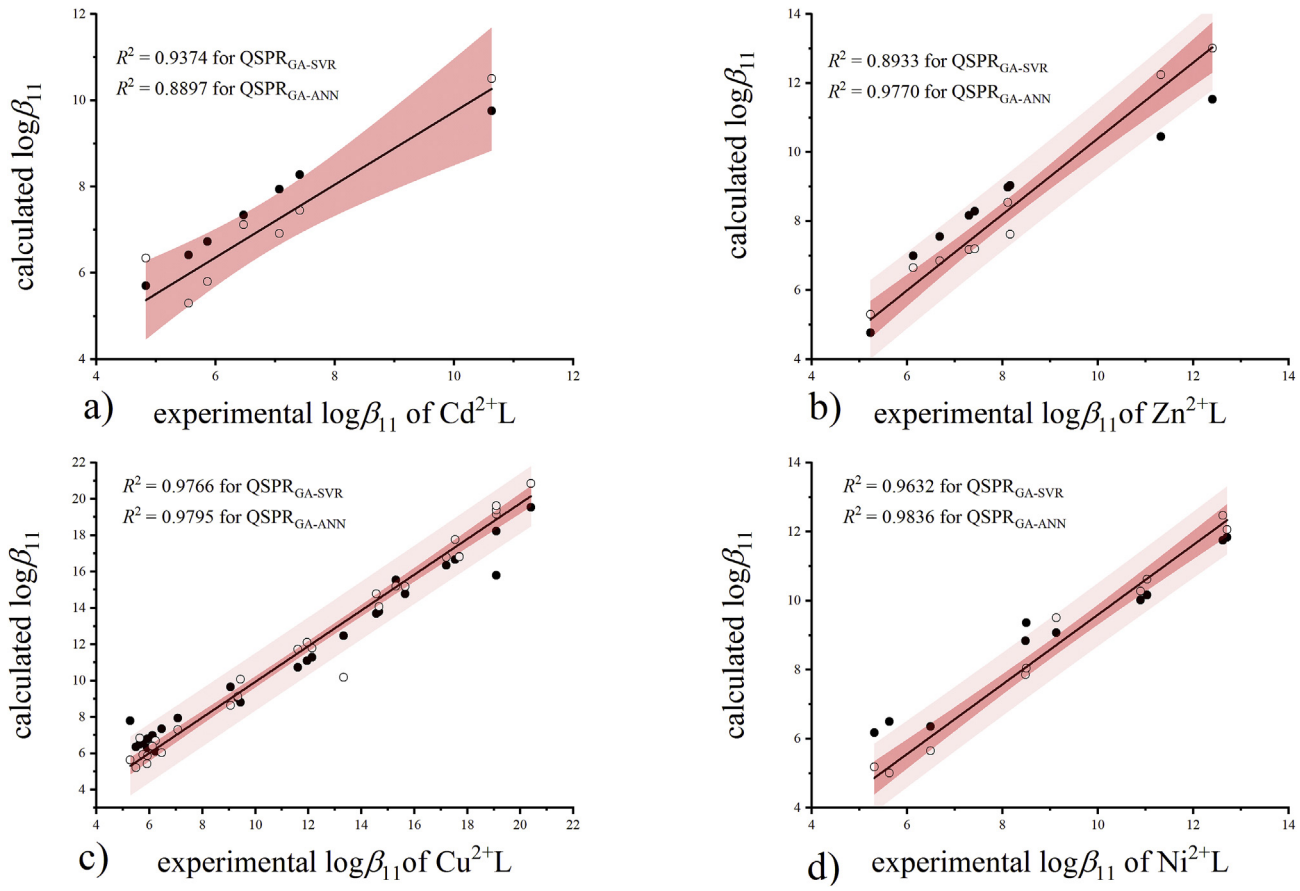


Fig. 7. The correlation between experimental and calculated $\log\beta_{11}$ values for complexes $Cu^{2+}L$, $Zn^{2+}L$, $Cd^{2+}L$ and $Ni^{2+}L$ over training, validation and additional test set (Table 2). Symbol: ●: results from $QSPR_{GA-SVR}$; ○: results from $QSPR_{GA-ANN}$.

Table 7
The predicted $\log\beta_{11}$ stability constants of lead and new complexes using the correlation equations for individual metal ion Zn^{2+}, Cd^{2+}, Cu^{2+} and Ni^{2+}, respectively.

Complex		Correlation equation	$\log\beta_{11exp}$	ref.	$\log\beta_{11cal}$	ARE%
35p	$Ni^{2+}L$	$\log\beta_{11\text{-}SVR} = 2.1438 + 0.7562 \times \log\beta_{11exp}$	8.50	[37]	8.5713	0.8386
33p			12.71	[47]	11.7548	7.5157
35p		$\log\beta_{11\text{-}ANN} = -0.4963 + 1.0093 \times \log\beta_{11exp}$	8.50	[37]	8.0831	4.9045
33p			12.71	[47]	12.3324	2.9706
19p	$Cd^{2+}L$	$\log\beta_{11\text{-}SVR} = 2.7148 + 0.6934 \times \log\beta_{11exp}$	5.544	[53]	6.5592	18.3119
10p			10.63	[51]	10.0860	5.1171
19p		$\log\beta_{11\text{-}ANN} = 1.2926 + 0.8443 \times \log\beta_{11exp}$	5.544	[53]	5.9736	7.7495
10p			10.63	[51]	10.2679	3.4060
85p	$Cu^{2+}L$	$\log\beta_{11\text{-}SVR} = 1.7142 + 0.8354 \times \log\beta_{11exp}$	6.114	[54]	6.8217	11.5751
67p			5.924	[52]	6.6630	12.4744
85p		$\log\beta_{11\text{-}ANN} = 0.1163 + 0.9823 \times \log\beta_{11exp}$	6.114	[54]	6.1219	0.1285
67p			5.924	[52]	5.9352	0.1895
49p	$Zn^{2+}L$	$\log\beta_{11\text{-}SVR} = 2.0424 + 0.7888 \times \log\beta_{11exp}$	12.40	[15]	11.8229	4.6538
48p			7.30	[31]	7.8003	6.8535
49p		$\log\beta_{11\text{-}ANN} = -0.6056 + 1.1003 \times \log\beta_{11exp}$	12.40	[15]	13.0384	5.1480
48p			7.30	[31]	7.4267	1.7360
109n	$Ni^{2+}L$	$\log\beta_{11\text{-}SVR} = 2.1438 + 0.7562 \times \log\beta_{11exp}$	23.3149	This work	19.7739	15.1877
110n	$Cd^{2+}L$	$\log\beta_{11\text{-}SVR} = 2.7148 + 0.6934 \times \log\beta_{11exp}$	23.0463	This work	18.6960	18.8763
111n	$Cu^{2+}L$	$\log\beta_{11\text{-}SVR} = 1.7142 + 0.8354 \times \log\beta_{11exp}$	19.4148	This work	17.9329	7.6328
112n	$Zn^{2+}L$	$\log\beta_{11\text{-}SVR} = 2.0424 + 0.7888 \times \log\beta_{11exp}$	16.4712	This work	15.0341	8.7249
109n	$Ni^{2+}L$	$\log\beta_{11\text{-}ANN} = -0.4963 + 1.0093 \times \log\beta_{11exp}$	19.6298	This work	19.3169	1.5940
110n	$Cd^{2+}L$	$\log\beta_{11\text{-}ANN} = 1.2926 + 0.8443 \times \log\beta_{11exp}$	21.3607	This work	19.3283	9.5147
111n	$Cu^{2+}L$	$\log\beta_{11\text{-}ANN} = 0.1163 + 0.9823 \times \log\beta_{11exp}$	19.0603	This work	18.8385	1.1637
112n	$Zn^{2+}L$	$\log\beta_{11\text{-}ANN} = -0.6056 + 1.1003 \times \log\beta_{11exp}$	17.5903	This work	18.7494	6.5894

p: prediction lead complexes; n: new complexes

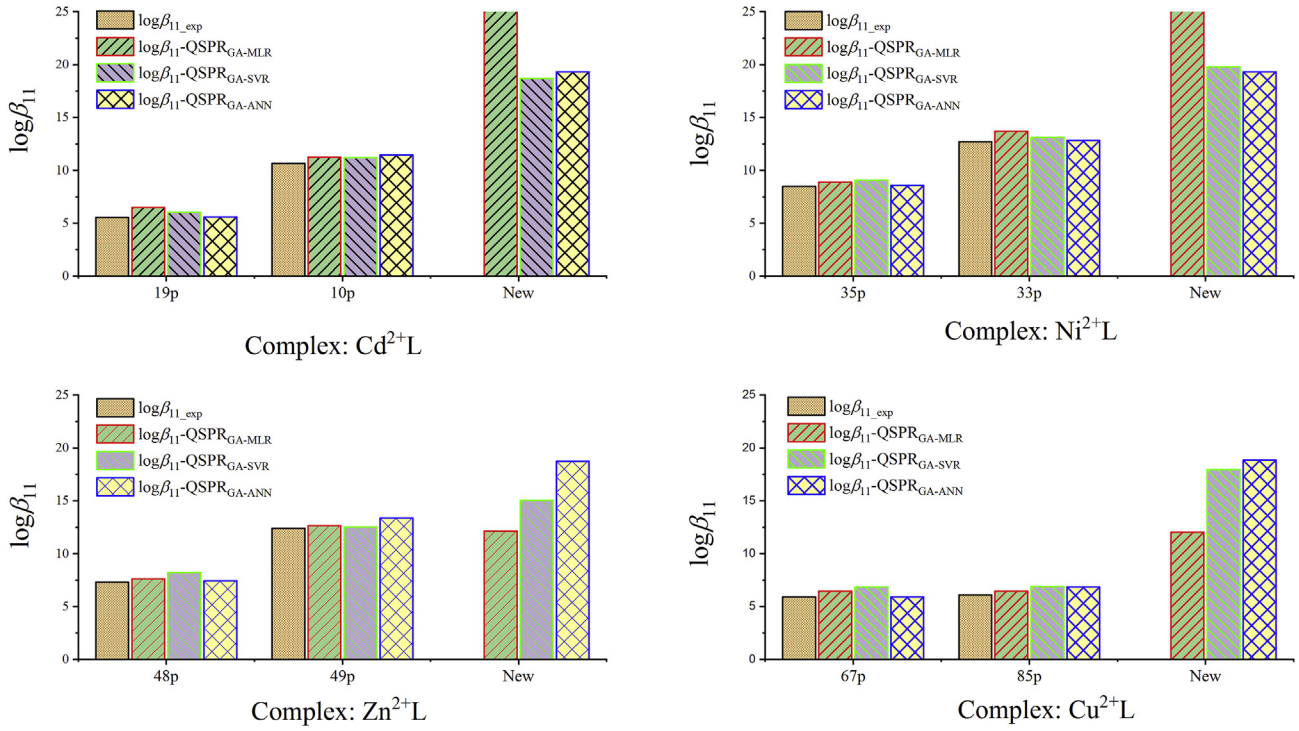


Fig. 8. The comparison of predicted $\log\beta_{11}$ values from the $QSPR_{GA\text{-}SVR}$ and $QSPR_{GA\text{-}ANN}$ models of the new complexes with the experimental data of lead complexes.

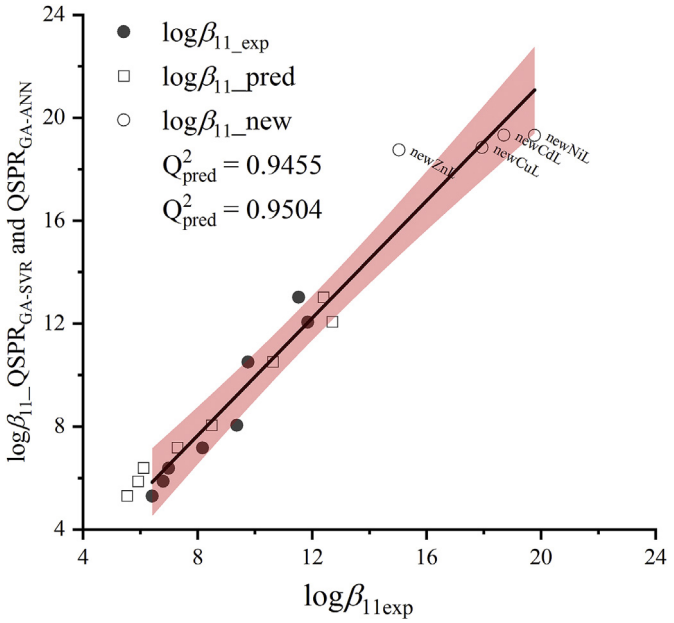


Fig. 9. The comparison of predicted $\log\beta_{11}$ values of four new complexes from the $QSPR_{GA\text{-}SVR}$ and $QSPR_{GA\text{-}ANN}$ models with the experimental data of lead complexes.

to build the QSPR models to predict the stability constants of new complexes with organic substances in different structures [80–82]. Their constructed QSPR models demonstrate effectively of these methods to predict the stability constants of metal-organic ligand complexes. In this work, we have also conducted to approach the QSPR models on the predictability for the stability constants $\log\beta_{11}$. Complex structural information is shown clearly through 2D and 3D molecular descriptors. The constructed $QSPR_{GA\text{-}MLR}$ model (6) with $k = 7$ molecular descriptors such 2D descriptors xp3, xp5, nelem, nrings, SaasC and 3D descriptors Ovality, Surface has shown predictability by statistical values $R^2 = 0.9145$, $Q^2_{LOO} = 0.8650$ and MSE = 1.290 is better than the $QSPR_{GA\text{-}MLR}$ models with $k = 6$ and $k = 8$ (in Table 3). In addition, the statistical parameters describe the significance of the coefficients using the values |t-stat| of 5.3797–10.0845 with corresponding probability values P = 0.000 at 95% confidence level (in Table 4). The molecular descriptors of this $QSPR_{GA\text{-}MLR}$ model respond well to statistics and cannot remove any descriptors from the model. The selected $QSPR_{GA\text{-}MLR}$ (6) model with $k = 7$ is reasonable because it shows a good correlation and significant predictability $Q^2_{LOO} = 0.8650 > 0.6$. In a recent study Huanhuan Dong et al. successfully also proposed similarly the QSAR models by the regression techniques to develop new thiosemicarbazone derivatives with anti-tyrosinase activity of thiosemicarbazone group [83]. Their QSAR models have been developed with 14 molecular descriptors that also include 2D and 3D molecular descriptors. These also include Volume, Surface, and Ovality descriptor. Similarly Huanhuan Dong et al. also showed that the anti-tyrosinase activity of thiosemicarbazone group is highly dependent on the 3D molecular descriptors [83]. Huanhuan Dong's QSAR models developed were cross-validated for the internal and external groups with values R^2 from 0.732 to 0.938 and Q^2 range 0.675–0.894. For our selected $QSPR_{GA\text{-}MLR}$ models in Table 3, the 3D Ovality and Surface molecular descriptors, and 2D SaasC and nrings descriptors appear in most $QSPR_{GA\text{-}MLR}$ models. These 3D descriptors are not eliminated according to statistical standards at 95% confident level when screening with genetic algorithm. So we developed the thiosemicarbazone structure to increase the magnitude of Ovality and Surface descriptors. Thus, the stability constants of the complexes also increase with the size of the new thiosemicarbazone structure. Furthermore, we assessed the global predictability of QSPR models for the $\log\beta_{11}$ constants of 108

complexes for 21 different metal ions Ag^{+}, Cd^{2+}, Co^{2+}, Cr^{3+}, Cu^{2+}, Dy^{3+}, Fe^{2+}, Fe^{3+}, Gd^{3+}, Ho^{3+}, La^{3+}, Mg^{2+}, Mn^{2+}, Mo^{5+}, Nd^{3+}, Ni^{2+}, Pb^{2+}, Pr^{3+}, Sm^{3+}, Tb^{3+}, V^{5+}, Zn^{2+} of the entire data set represented by error values *ARE*,% of 14.846% for $QSPR_{GA\text{-}MLR}$ model (6), 12.274% for $QSPR_{GA\text{-}SVR}$ model and 5.092% for $QSPR_{GA\text{-}ANN}$ model. Moreover, the QSPR models are cross-validated by the internal predictability for training set based on the Q^2_{LOO} values in the increase range 0.8650–0.9317 corresponding to $QSPR_{GA\text{-}MLR}$ and $QSPR_{GA\text{-}ANN}$ model, and based on the *MARE*,% values in the decrease range 10.7076%–4.9796% corresponding to $QSPR_{GA\text{-}MLR}$ and $QSPR_{GA\text{-}ANN}$ model. The QSPR models are also validated by the external predictability with the Q^2 values in increase range 0.7168–0.9669 and the decrease of *MARE*,% values in range 19.01119%–5.9019% corresponding to $QSPR_{GA\text{-}MLR}$ and $QSPR_{GA\text{-}ANN}$ model. Besides the QSPR models also presented the good correlation with R^2 values in the range 0.9148–0.9815 for training process. We have also carried out to assess the predictability of QSPR models for additional-external data set with the increase of Q^2 values in range 0.7958–0.9769 and with the decrease of *MARE*,% values in range 13.7829%–4.3756% corresponding to $QSPR_{GA\text{-}MLR}$ and $QSPR_{GA\text{-}ANN}$. All values $R^2 > 0.9$ and $Q^2 > 0.6$ are very significant at 95% confident level for all constructed QSPR models. The $QSPR_{GA\text{-}SVR}$ and $QSPR_{GA\text{-}ANN}$ models were selected to estimate $\log\beta_{11}$ stability constants of new complexes for metal ions Ni^{2+}, Cd^{2+}, Cu^{2+} và Zn^{2+} due to those represented the reliable predictability.

For the design of new thiosemicarbazone reagents, we found that our design of newly thiosemicarbazone reagents for complexing with metal ions Ni^{2+}, Cd^{2+}, Cu^{2+} và Zn^{2+} are very suitable for thiosemicarbazone group. The selected substituents are ring functional groups and in particular the aromatic heterocyclic groups may contribute to the activation of thiosemicarbazone derivatives by increasing the molecular Ovality value. In this study, we relied on the nature of the molecular descriptors in the constructed QSPR models and their contribution $APC_{m,nx}$,% to design the new thiosemicarbazone reagents that are capable of complexation with metals Ni^{2+}, Cd^{2+}, $Cu^{2+}$and Zn^{2+}. We here have demonstrated the orientation to design the new thiosemicarbazone reagents. The structures of thiosemicarbazone reagents and complexes with metal ions Ni^{2+}, Cd^{2+}, Cu^{2+} and Zn^{2+} were considered and screened based on the contribution capacity, as shown in Fig. 4. Most of these complexes have the stability constants $\log\beta_{11}$ that have been tested by the Grubb standard within the 95% confidence interval limit. We designed the new thiosemicarbazone molecule by replacing 3-bromo-10-ethyl-10H-phenothiazine and 3-bromo-9-ethyl-9H-carbazole group in the R_4 position, as shown in Tables 6 and 7. We chose these two groups because these are multi-ring aromatic heterocyclic substituents and after attaching one of these functional groups to the R_4 position it could cause the increase of Ovality and Surface magnitude as discussed above. These are the aromatic heterocyclic substitutes with adjacent rings. When replacing group R_4 by these groups the stability of the complexes could be increased, because of the substitute of the group 3-bromo-10-ethyl-10H-phenothiazine or 3-bromo-9-ethyl-9H-carbazole in C_5 position of the double bond $-C_5{=}N_4-$ of thiosemicarbazone molecular skeleton could form a stable conjugated system with the conjugated bonds $-C_5{=}N_4-$. This causes an increase in electron density on the N_4 atom to facilitate the N_4 atom bonding with the metal ion Me^{n+} with an empty p orbital. Complexes of thiosemicarbazone ligand with metal ions could be constituted more easily and rapidly. Furthermore, the 3-bromo-10-ethyl-10H-phenothiazine and -3-bromo-9-ethyl-9H-carbazole functional groups in position C_5 of the bond $-C_5{=}N_4-$ have also increased the magnitude of Ovality and Surface molecular descriptors as well as increasing the contribution $APC_{m,nx}$,% of these descriptors to the $\log\beta_{11}$ constants of the complexes $Ni^{2+}L$, $Cu^{2+}L$, $Cd^{2+}L$ and $Zn^{2+}L$. So the estimated results derived from the $QSPR_{GA\text{-}SVR}$ and $QSPR_{GA\text{-}ANN}$ models in accordance with the experimental study, as shown in Tables 6 and 7. This can be easily seen as follows: The complexes **33p** and **35p** for $Ni^{2+}L$, complexes **10p** and **19p** for $Cd^{2+}L$, complexes **67p** and **85p** for $Cu^{2+}L$ and complexes **48p** and **49p** for $Zn^{2+}L$ were selected from the entire data after screening by the $QSPR_{GA\text{-}MLR}$ model (as described in Fig. 4) to use them as the lead substances. The initial complexes **33p** and **35p** for $Ni^{2+}L$ gave the values Ovality = 1.612 and 1.602, Surface = 280.897 and 293.865, and the values $\log\beta_{11} = 12.710$ and $\log\beta_{11} = 8.50$, respectively; after designing the **109n** $Ni^{2+}L$ new complex, the values of Ovality = 1.6838 increased in range 4.4540%–5.05618% and Surface = 389.3638 increased in range 32.4975%–38.6144% and the obtained values $\log\beta_{11}$ in range 19.6298–23.3149; the initial complexes **10p** and **19p** for $Cd^{2+}L$ gave the values Ovality = 1.665 and 1.686, Surface = 295.792 and 308.931, the values $\log\beta_{11} = 10.630$ và $\log\beta_{11} = 5.544$, respectively; after designing the **110n** $Cd^{2+}L$ new complex, the values of Ovality = 1.7252 increased in range 2.325%–3.6156% and Surface = 414.5862 increased in range 34.2003%–40.1614%, the obtained values $\log\beta_{11}$ in range 21.3607–23.0463; the initial complexes **67p** and **85p** for $Cu^{2+}L$ gave the values of Ovality = 1.548 and 1.670, Surface = 257.546 and 299.933 the values $\log\beta_{11} = 5.924$ and $\log\beta_{11} = 6.114$, respectively; after designing **111n** $Cu^{2+}L$ new complex, the values of Ovality = 1.8207 increased in range 9.02395%–17.61628% and Surface = 436.4311 increased in range 44.8427%–69.45749%, the obtained values $\log\beta_{11}$ in range 19.0603–19.4148; the initial complexes **48p** và **49p** $Zn^{2+}L$ gave the values of Ovality = 1.787 and 1.727, Surface = 419.980 and 380.609, the values $\log\beta_{11} = 7.300$ and $\log\beta_{11} = 12.4$, respectively; after designing **112n** $Zn^{2+}L$ new complex, the values of Ovality = 1.8168 increased in range 1.6676%–5.1998% and Surface = 435.6422 increased in range 3.72927%–14.45925%, the obtained values $\log\beta_{11}$ in range 16.4712–17.5903; We find that the $\log\beta_{11}$ constants of the new complexes are higher than the $\log\beta_{11}$ constants of the lead complexes, as shown in Table 7 and Fig. 8. Molecular descriptors used in this study for our thiosemicarbazone group are similar to those used by Huanhuan Dong et al. [83] to develop the QSAR models for the set of anti-tyrosinase thiosemicarbazones using regression techniques. These obtained results are also consistent with the comments that Huanhuan Dong et al. [83] have carried out to predict the anti-tyrosinase activity of new thiosemicarbazones and design new thiosemicarbazones. We have succeeded in building QSPR models for 108 complexes of 21 metal ions and has also successfully designed and synthesized two new thiosemicarbazone reagents and synthesized new complexes of $Ni^{2+}L$, $Cu^{2+}L$, $Cd^{2+}L$ and $Zn^{2+}L$. The obtained results are also consistent with the thiosemicarbazone ligands based on condensation reactions with $R_3CR_2{=}NR_1$ where R_1, R_2 and R_3 represent aryl substituents [8].

We could conclude that the combined use of molecular descriptors with the support of genetic algorithms to screen the significant-contribution descriptors could develop accurately the QSPR models for metal-thiosemicarbazone complexes. The $QSPR_{GA\text{-}SVR}$ and $QSPR_{GA\text{-}ANN}$ models turn out to be better predictable than the $QSPR_{GA\text{-}MLR}$ model. Of the $QSPR_{GA\text{-}SVR}$ and $QSPR_{GA\text{-}ANN}$ models were found to outperform the traditional linear regression. This facilitates the screening and design of new reagents rapidly.

In addition, we also selected the significant complexes to develop the new complexes using the structural descriptors that are inexpensive using our developed QSPR models. The new thiosemicarbazone reagents designed successfully from the selected ligands. We believe that no QSPR model has been studied to predict the $\log\beta_{11}$ constants so far for the metal-thiosemicarbazone complexes. The QSPR models in this work may be useful in the design of new thiosemicarbazone ligands.

Acknowledgements

We would like to thank Prof. Dr. James JP Stewart, who provided us with free MOPAC2016. We would like to thank Industrial University of Ho Chi Minh City for their best support to complete our experiments. We also thank Ton Duc Thang University for enabling us to carry out this project. We would like to thank the software companies for providing us with a fully functional trial program. The our researched results have proven the accuracy and effectiveness of the used software.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.molstruc.2019.05.050.

References

[1] S. Kumar, D.N. Dhar, P.N. Saxena, Applications of metal complexes of Schiff bases-A review, J. Sci. Ind. Res. 68 (2009) 181–187.
[2] I. Kizilcikli, Y.D. Kurt, B. Akkurt, A.Y. Genel, S. Birteksöz, G. Ötük, B. Ülküseven, Antimicrobial activity of a series of thiosemicarbazones and their Zn^{II} and Pd^{II} complexes, Folia Microbiol. 52 (2007) 15–25.
[3] D.X. West, A.E. Liberta, S.B. Padhye, R.C. Chikate, P.B. Sonawane, A.S. Kumbhar, R.G. Yerande, Thiosemicarbazone complexes of copper (II): structural and biological studies, Coord. Chem. Rev. 123 (1993) 49–71.
[4] T. Khan, R. Ahmad, S. Joshi, A.R. Khan, Anticancer potential of metal thiosemicarbazone complexes: a review, Chem. Sin. 6 (2015) 1–11.
[5] V. Jevtović, S. Ivković, S. Kaisarević, R. Kovačević, Anticancer activity of new copper (II) complexes incorporating a pyridoxal-semicarbazone ligand, Contemp. Mater. I-2 (2010) 133–137.
[6] T. Bal-Demirci, M. Şahin, E. Kondakçi, M. Özyürek, B. Ülküseven, R. Apak, Synthesis and antioxidant activities of transition metal complexes based 3-hydroxysalicylaldehyde-S-methylthiosemicarbazone, Spectrochim. Acta Mol. Biomol. Spectrosc. 138 (2015) 866–872.
[7] M. Aljahdali, A.A. El-Sherif, Synthesis, characterization, molecular modeling and biological activity of mixed ligand complexes of Cu(II), Ni(II) and Co(II) based on 1,10-phenanthroline and novel thiosemicarbazone, Inorg. Chim. Acta 407 (2013) 58–68.
[8] M. Bakherad, A. Keivanloo, B. Bahramian, S. Jajarmi, Copper-and solvent-free Sonogashira coupling reactions of aryl halides with terminal alkynes catalyzed by 1-phenyl-1, 2-propanedione-2-oxime thiosemi-carbazone-functionalized polystyrene resin supported Pd(II) complex under aerobic conditions, Appl. Catal. Gen. 390 (2010) 135–140.
[9] A.S. El-Tabl, M.M. Abd-El Wahed, M.A. Wahba, M. Shakdofa, A. Gafer, Bimetallic transition metal complexes of 2, 3-dihydroxy-N', N'4-bis ((2-Hydroxynaphthalen-1-yl) methylene) succinohydrazide ligand as a new class of bioactive compounds; synthesis, characterization and cytotoxic evaluation, Indian J. Adv. Chem. Sci. 4 (2016) 114–129.
[10] S. Padhye, G.B. Kauffman, Transition metal complexes of semicarbazones and thiosemicarbazones, Coord. Chem. Rev. 63 (1985) 127–160.
[11] D.N. Reddy, Extractive direct and derivative spectrophotometric determination of Nickel (II) in Medicinal leaves, Soil, and Alloy samples by using Pyridoxal-3-thiosemicarbazone (PDT), J. Mater. Environ. Sci. 5 (2014) 1188–1199.
[12] S.A. Reddy, K.J. Reddy, S. Lakshminaraya, D.L. Priya, Y.S. Rao, A.V. Reddy, Extractive spectrophotometric determination of trace amounts of cadmium(II) in medicinal leaves and environmental samples using benzildithiosemicarbazone (BDTSC), J. Hazard Mater. 152 (2008) 903–909.
[13] S.V. Babu, K.H. Reddy, Spectrophotometric determination of copper (II) and palladium (II) using 3-hydroxybenzaldehyde thiosemicarbazone, J. Indian Chem. Soc. 83 (2006) 20–22.
[14] G. Bhatt, I. Patel, K. Desai, Spectrophotometric determination of Cu (II) with p-Chloro-benzaldehyde-4-(2'-Carboxy-5'-Sulphophenyl)-3-Thiosemicarbazone, J. Inst. Chem. India 65 (1993), 190-190.
[15] M.N. Milunovic, E.A. Enyedy, N.V. Nagy, T. Kiss, R. Trondl, M.A. Jakupec, B.K. Keppler, R. Krachler, G. Novitchi, V.B. Arion, L.-, D Proline Thiosemicarbazone Conjugates, Coordination behavior in solution and the effect of copper(II) coordination on their antiproliferative activity, Inorg. Chem. 51 (2012) 9309–9321.
[16] S.S. Sawhney, S.K. Chandel, Stability and thermodynamics of La(III)-, Pr(III)-, Nd(III)-, Gd(III)- and Eu(III)-p-nitrobenzaldehyde thiosemicarbazone systems, Thermochim. Acta 72 (1984) 381–385.
[17] A.G. Mercader, P.R. Duchowicz, P.M. Sivakumar, Chemometrics Applications and Research: QSAR in Medicinal Chemistry, CRC Press Taylor & Francis Group, 2016, pp. 102–120.
[18] R. Kunal, K. Supratik, D.R. Narayan, A Primer on QSAR/QSPR Modeling: Fundamental Concepts, Springer, London, 2015, pp. 37–58.
[19] I. Mitra, P.P. Roy, S. Kar, P. Ojha, K. Roy, *On further application of r^2_m as a metric for validation of QSAR models*, J. Chemom. 24 (2010) 22–33.
[20] G. Schuurmann, R.U. Ebert, J. Chen, B. Wang, R. Kuhne, External validation and prediction employing the predictive squared correlation coefficient-test-set activity mean vs training set activity mea, J. Chem. Inf. Model. 48 (2008) 2140–2145.
[21] M. Hymavathi, C. Viswanatha, N. Devanna, A study on synthesis of novel chromogenic organic reagent 3,4-dihydroxy-5-methoxy benzaldehyde thiosemicarbazone and spectrophotometric determination of nickel(II) in presences of triton X-100, Res. J. Pharmaceut. Biol. Chem. Sci. 5 (2014) 625–630.
[22] M. Hymavathi, C. Viswanatha, N. Devanna, A study on synthesis of novel chromogenic organic reagent 3,4-dihydroxy-5-methoxy benzaldehyde thiosemicarbazone and spectrophotometric determination of Cobalt (II) in presences of Triton X-100, J. Chem. Pharm. Res. 6 (2014) 2787–2791.
[23] M. Hymavathi, C. Viswanatha, N. Devanna, A sensitive and selective chromogenic reagent using 2-hydroxy 3, 5-dimethoxy benzaldehyde thiosemicarbazone (HDMBTSC) for direct and derivative spectrophotometric determination of Molybdenum (VI), Int. J. Math. and Phys. Sci. Res. 2 (2014) 43–48.
[24] M. Hymavathi, C. Viswanatha, N. Devanna, Direct and derivative spectrophotometric determination of Copper (II) using a sensitive and selective chromogenic organic reagent 2-hydroxy 3,5-dimethoxy benzaldehyde thiosemicarbazone (HDMBTSC), W. J. Pharm . Phar. Sci. 3 (2014) 1688–1695.
[25] E.A. Gomaa, K.M. Ibrahim, N.M. Hassan, *Evaluation of thermodynamic parameters (conductometrically) for the interaction of Cu(II) ion with 4-phenyl -1-diacetyl monoxime -3- thiosemicarbazone (BMPTS) in (60%V) ethanol (EtOH-H_2O) at different temperatures*, Int. J. Eng. Sci. 3 (2014) 44–51.
[26] M. Aljahdali, A.A. EL-Sherif, Synthesis, characterization, molecular modeling and biological activity of mixed ligand complexes of Cu(II), Ni(II) and Co(II) based on 1,10-phenanthroline and novel thiosemicarbazone, Inorg. Chim. Acta 407 (2013) 58–68.
[27] A. Gaál, G. Orgován, Z. Polgári, A. Réti, V.G. Mihucz, S. Bősze, N. Szoboszlai, C. Streli, 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.
[28] B.S. Garg, S.R. Singh, S.B. Basnet, R.P. Singh, Potentiometric studies on the complexation equilibria between La(III), Pr(III), Nd(III), Gd(III), Sm(III), Tb(III), Dy(III), Ho(III) and 2-acetylpyridinethiosemicarbazone (2-APT), Polyhedron 7 (1988) 147–150.
[29] A.T.A. El-Karim, A.A. El-Sherif, Potentiometric, equilibrium studies and thermodynamics of novel thiosemicarbazones and their bivalent transition metal(II) complexes, J. Mol. Liq. 219 (2016) 914–922.
[30] J.R. Koduru, K.D. Lee, 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 Chem. 150 (2014) 1–8.
[31] D. Rogolino, A. Cavazzoni, A. Gatti, M. Tegoni, G. Pelosi, V. Verdolino, C. Fumarola, D. Cretella, P.G. Petronini, M. Carcelli, Anti-proliferative effects of copper(II) complexes with Hydroxyquinoline-Thiosemicarbazone ligands, Eur. J. Med. Chem. 128 (2017) 140–153.
[32] M.A. Jiménez, M.D. Luque De Castro, M. Valcárcel, Potentiometric study of silver(I)-thiosemicarbazonates, Microchem. J. 25 (1980) 301–308.
[33] B.S. Garg, V.K. Jain, Determination of thermodynamic parameters and stability constants of complexes of biologically active o-vanillinthiosemicarbazone with bivalent metal ions, Thermochim. Acta 146 (1989) 375–379.
[34] B.S. Garg, S. Ghosh, V.K. Jain, P.K. Singh, Evaluation of thermodynamic parameters of bivalent metal complexes of 2-hydroxyacetophenonethiosemicarbazone (2-HATS), Thermochim. Acta 157 (1990) 365–368.
[35] K.H. Reddy, N.B.L. Prasad, Spectrophotometric determination of copper (II) in edible oils and seed using novel oxime-thiosemicarbazones, Indian J. Chem. 43A (2004) 111–114.
[36] Sahadev, R.K. Sharma, S.K. Sindhwani, Thermal studies on the chelation behaviour of biologically active 2-hydroxy-1-naphthaldehyde thiosemicarbazone (HNATS) towards bivalent metal ions: a potentiometric study, Thermochim. Acta 202 (1992) 291–299.
[37] K. Sarkar, B.S. Garg, Determination of thermodynamic parameters and stability constants of the complexes of p-MITSC with transition metal ions, Thermochim. Acta 113 (1987) 7–14.
[38] F. Toribio, J.M.L. Fernandez, D.P. Bendito, M. Valcárcel, 2,2'-dihydroxybenzophenone thiosemicarbazone as a spectrophotometric Reagent for the determination of copper, cobalt, nickel, and iron trace amounts in mixtures without previous separations, Microchem. J. 25 (1980) 338–347.
[39] D.N. Kenie, A. Satyanarayana, Protolitic equilibria and stability constants of Mn(II) and Ni(II) complexes of 3-formylpyridine thiosemicarbazone in sodium dodecyl sulphate (SDS)- water mixture, Sci. Technol. Arts Res. J. 4 (2015) 74–79.
[40] R. Biswas, D. Brahman, B. Sinha, Thermodynamics of the complexation between salicylaldehyde thiosemicarbazone with Cu(II) ions in methanol-1,4-dioxane binary solutions, J. Serb. Chem. Soc. 79 (2014) 565–578.
[41] N.S.R. Reddy, D.V. Reddy, Spectrophotometric determination of vanaditun(V) with salicylaldehyde thiosemicarbazon, J. Indian Inst. Sci. 64B (1983) 133–136.
[42] G. Ramanjaneyulu, P.R. Reddy, V.K. Reddy, T.S. Reddy, Direct and derivative spectrophotometric determination of copper(II) with 5-bromosalicylaldehyde thiosemicarbazone, Open Anal. Chem. J. 2 (2008) 78–82.
[43] D.N. Kenie, A. Satyanarayana, Solution equilibrium study of the complexation of Co(II) and Zn(II) with nicotinaldehyde thiosemicarbazone, Sci. Technol. Arts

Res. J. 4 (2015) 145–149.
[44] I. Sreevani, P.R. Reddy, V.K. Reddy, A rapid and simple spectrophotometric determination of traces of chromium(VI) in waste water samples and in soil samples by using 2-hydroxy, 3-methoxy benzaldehyde thiosemicarbazone (HMBATSC), IOSR J. Appl. Phys. 3 (2013) 40–45.
[45] M.A. Jiménez, M.D. Luque De Castro, M. Valcárcel, Titration of thiosemicarbazones with Cu(II) and vice versa by use of a copper selective electrode in acetone-water mixture: determination of the conditional formation constants of the cupric thiosemicarbazonates, Microchem. J. 32 (1985) 166–173.
[46] T. Atalay, E. Ozkan, Thermodynamic studies of some complexes of 4'-morpholinoacetophenone thiosemicarbazone, Thermochim. Acta 237 (1994) 369–374.
[47] S.S. Sawhney, S.K. Chandel, Solution chemistry of Cu(II)-, Co(II)-, Ni(II)-, Mn(II)- and Zn(II)-p-aminobenzaldehyde thiosemicarbazone systems, Thermochim. Acta 71 (1983) 209–214.
[48] D.K. Singh, P.K. Jha, R.K. Jha, P.M. Mishra, A. Jha, S.K. Jha, R.P. Bharti, Equilibrium studies of transition metal complexes with tridentate ligands containing N, O, S as Donor Atoms, Asian J. Chem. 21 (2009) 5055–5060.
[49] V. Veeranna, V.S. Rao, V.V. Laxmi, T.R. Varalakshmi, Simultaneous second order derivative spectrophotometric determination of cadmium and cobalt using furfuraldehyde thiosemicarbazone (FFTSC), Res. J. Pharm. Technol. 6 (2013) 577–584.
[50] D.G. Krishna, G.V.K. Mohan, A facile synthesis, characterization of cinnamaldehyde thiosemicarbazone and determination of molybdenum (VI) by spectrophotometry in presence of micellar medium, Indian J. Appl. Res. 8 (2013) 7–8.
[51] S.S. Sawhney, R.M. Sati, pH-metric studies on Cd(II)-, Pb(II)-, Al(III)-, Cr(III)- AND Fe(III)-p-nitrobenzaldehyde thiosemicarbazone systems, Thermochim. Acta 66 (1983) 351–355.
[52] D. Admasu, D.N. Reddy, K.N. Mekonne, Spectrophotometric determination of Cu(II) in soil and vegetable samples collected from Abraha Atsbeha, Tigray, Ethiopia using heterocyclic thiosemicarbazone, SpringerPlus 5 (2016) 1–8.
[53] D.G. Krishna, C.K. Devi, Determination of cadmium (II) in presence of micellar medium using cinnamaldehyde thiosemicarbazone by spectrophotometry, Int. J. Green Chem. Biopro. 5 (2015) 28–30.
[54] K.V. Reddy, D.N. Reddy, S.V. Babu, K.H. Reddy, *Spectrophometric determination of copper (II) in Biological samples by using 2-acetylpyridine 4-methyl -3-thiosemicarbazone (APMT*), Der Pharm. Sin. 2 (2011) 176–183.
[55] P. Debye, E. Hückel, The Theory of Electrolytes. I. Lowering of Freezing Point and Related Phenomena, Physikalische Zeitschrift, German, 1923.
[56] C.W. Davies, Ion Association, Butterworths, London, 1962, pp. 37–53.
[57] P.V. Tat, Development of QSAR and QSPR, Publisher of Natural sciences and Technique, Hanoi, 2009.
[58] QSARIS 1.1, Statistical Solutions Ltd., USA, 2001.
[59] J.R. Dilworth, R. Hueting, Review: metal complexes of thiosemicarbazones for imaging and therapy, Inorg. Chim. Acta 389 (2012) 3–15.
[60] BIOVA Draw 2017 R2., Version: 17.2, *NET*., Dassault Systèmes, France, 2016.
[61] J.J.P. Stewart, Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters, J. Mol. Model. 19 (2013) 1–32.
[62] J.J.P. Stewart, MOPAC2016, Version: 17.240W, Stewart Computational Chemistry, USA, 2002.
[63] XLSTAT Version 2016.02.28451, Addinsoft, USA, 2016.
[64] J.B. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1967, pp. 281–297.
[65] B.S. Everitt, S. Landau, M. Leese, Cluster Analysis, fourth ed., Arnold, London, 2001.
[66] J.H. Holland, Genetic algorithms, Sci. Am. 267 (1992) 44–50.
[67] D.D. Steppen, J. Werner, P.R. Yeater, Essential Regression And Experimental Design for Chemists And Engineers, Free Software Package, 1998. http://home.t-online.de/home/jowerner98/indexeng.html.
[68] D.C. Montgomery, E.A. Peck, C.G. Vining, Introduction to Linear Regression Analysis, third ed., Wiley-Interscience, New York, 2001.
[69] S. Weisberg, Applied Linear Regression, second ed., Wiley, New York, 1985.
[70] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge university Press, Cambridge, 2000.
[71] C. Corinna, V.N. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[72] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222.
[73] K. Samghani, M.H. Fatemi, Developing a support vector machine based QSPR model for prediction of half-life of some herbicides, Ecotoxicol. Environ. Saf. 129 (2016) 10–15.
[74] H. Beale, M.T. Hagan, H.B. Demuth, Neural Network Toolbox™ User's Guide, The MathWorks, Inc., USA, 2015.
[75] J. Gasteiger, J. Zupan, Neural networks in chemistry, Chiw. Inr. Ed. Engl. 32 (1993) 503–521.
[76] R. Rojas, Neural Networks, Springer-Verlag, Berlin, 1996.
[77] M. Hagan, H. Demuth, M. Beale, Neural Network Design, MA, PWS Publishing, Boston, USA, 1996.
[78] S. Makridakis, Accuracy measures: theoretical and practical concerns, Int. J. Forecast. 9 (1993) 527–529.
[79] C.M. Judd, G.H. McClelland, C.S. Ryan, Data Analysis: A Model Comparison Approach, Routledge, New York, 2009.
[80] Vitaly Solov'ev, Igor Sukhno, Vladimir Buzko, Aleksey Polushin, Gilles Marcou, Tsivadze Aslan, Alexandre Varnek, Stability constants of complexes of Zn^{2+}, Cd^{2+}, and Hg^{2+} with organic ligands: QSPR consensus modeling and design of new metal binders, J. Inclusion Phenom. Macrocycl. Chem. 72 (2012) 309–321.
[81] Vitaly Solov'ev, Natalia Kireeva, Svetlana Ovchinnikova, Tsivadze Aslan, The complexation of metal ions with various organic ligands in water: prediction of stability constants by QSPR ensemble modelling, J. Inclusion Phenom. Macrocycl. Chem. 83 (Issue 1–2) (2015) 89–101.
[82] Vitaly Solov'ev, Alexandre Varnek, Tsivadze Aslan, QSPR ensemble modelling of the 1:1 and 1:2 complexation of Co^{2+}, Ni^{2+}, and Cu^{2+} with organic ligands: relationships between stability constants, J. Comput. Aided Mol. Des. 28 (5) (2014) 549–564.
[83] Huanhuan Dong, Jing Liu, Xiaoru Liu, Yanying Yu, Shuwen Cao, Combining molecular docking and QSAR studies for modelling the anti-tyrosinase activity of aromatic heterocycle thiosemicarbazone analogues, J. Mol. Struct. 1151 (5) (2018) 353–365.