

This article was downloaded by: [Renmin University of China]

On: 13 October 2013, At: 10:21

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Coordination Chemistry

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/gcoo20>

Simultaneous estimation of stability constants of Mg, Ba, Ca, and Sr complexes using a small subset of molecular descriptors

Zahra Garkani-Nejad^a & Mohammad Ahmadvand^a

^a Department of Chemistry, Faculty of Science, Vali-e-Asr University, Rafsanjan, Iran

Published online: 14 Jul 2011.

To cite this article: Zahra Garkani-Nejad & Mohammad Ahmadvand (2011) Simultaneous estimation of stability constants of Mg, Ba, Ca, and Sr complexes using a small subset of molecular descriptors, *Journal of Coordination Chemistry*, 64:14, 2466-2479, DOI: [10.1080/00958972.2011.599382](https://doi.org/10.1080/00958972.2011.599382)

To link to this article: <http://dx.doi.org/10.1080/00958972.2011.599382>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Simultaneous estimation of stability constants of Mg, Ba, Ca, and Sr complexes using a small subset of molecular descriptors

ZAHRA GARKANI-NEJAD* and MOHAMMAD AHMADVAND

Department of Chemistry, Faculty of Science, Vali-e-Asr University, Rafsanjan, Iran

(Received 11 February 2011; in final form 6 June 2011)

This work develops a model for the prediction of stability constants of Mg, Ba, Ca, and Sr complexes, simultaneously. Quantitative structure–property relationship models are obtained by selecting descriptors from diverse functional groups, topological, and quantum chemical descriptors. Linear and non-linear methods such as multiple linear regression, partial least squares, and artificial neural networks with Levenberg–Marquardt (ANN-LM) back propagation algorithm were used to investigate the correlation between stability constant ($\log K$) and molecular descriptors. In this work, a suitable small subset of descriptors has been selected that is able to predict the stability constants of Mg, Ba, Ca, and Sr complexes, simultaneously. Finally, a 5-4-4 ANN-LM model was designed for predicting the stability constants of complexes with high predictive ability.

Keywords: Stability constant; Alkaline earth metal complexes; Quantitative structure–property relationship; Molecular descriptors

1. Introduction

There are many ways of estimating stability constants of coordination compounds, from simple empirical rules (Irving–Williams order [1], cation classifications [2, 3], distinction of hard and soft Lewis acids and bases, HSAB [4, 5]) to quantitative models, which differ considerably both in the theoretical basis and in the level of sophistication. Models based on “interaction parameters” between atoms in the first coordination sphere [6, 7] were followed by molecular-mechanic [8, 9] and molecular-dynamic [10] approaches, along with the methods based on continuous distribution of solvents [11, 12] and the overlapping sphere models [13].

In coordination chemistry, protonation constants of N-alkylated and N,N-dialkylated glycines were successfully correlated with their Wiener and Randic indices, as well as the stability constants of their ML and ML₂ chelates [14]. Later, the method was applied to the estimation of stability constants of copper(II) chelates with nine naturally occurring amino acids, reproducing stability constants of mono-complexes with a standard error (SE) = 0.07 $\log K$ unit [15]. Stability constants of their binary and

*Corresponding author. Email: garakani@mail.vru.ac.ir

ternary complexes were consecutively reproduced with an SE of 0.13 and 0.14 units, respectively [15]. The stability constants of macrocyclic ligand–metal complexes in solutions have been measured by various experimental techniques, such as potentiometry, conductometry, and calorimetry [16].

The main aim of this work was to select a small subset of common descriptors which are able to predict the stability constants of different metal complexes including Ba, Mg, Ca, and Sr; then develop a network that can predict the stability constant of different metal complexes, simultaneously. Such models enable reliable prediction of the stability constants for unknown complexes and to elucidate the structural factors determining the stability of complexes.

2. Experimental

2.1. Data set

Stability constants of complexes of Mg, Ba, Ca, and Sr metals with 22 different organic ligands have been taken from the literature [17] and presented in table 1. As can be seen from this table, most of the ligands are carboxylic acids. The ranges of stability constant in logarithmic scale for these compounds are between 0.51 and 8.69 for Mg, 0.31 and 7.78 for Ba, 0.51 and 10.7 for Ca, and 0.43 and 8.63 for Sr.

2.2. Molecular descriptors

The 3-D structures of the studied compounds were optimized using HYPERCHEM [18] software and semi-empirical quantum-chemical method of AM1. After screening and optimization of compounds, the descriptors from HYPERCHEM, MOPAC [19], and

Table 1. The full name of 22 ligands together with their experimental values of stability constants.

No.	Ligand name	Log $K(\text{Ba})$	Log $K(\text{Ca})$	Log $K(\text{Mg})$	Log $K(\text{Sr})$
1	Acetic acid	0.39	0.53	0.51	0.43
2	ADP	2.36	2.82	3.11	2.50
3	Aspartic acid	1.14	1.16	2.43	1.48
4	ATP	3.29	3.60	4.00	3.03
5	<i>n</i> -Butyric acid	0.31	0.51	0.53	0.36
6	Citric acid	2.30	3.50	2.80	2.80
7	EDTA	7.78	10.7	8.69	8.63
8	Gluconic acid	0.95	1.21	0.70	1.00
9	Glutamic acid	1.28	1.43	1.90	1.37
10	Glutaric acid	2.04	1.06	1.08	0.60
11	Glyceric acid	0.80	1.18	0.86	0.89
12	Glycine	0.77	1.43	3.45	0.91
13	Glycolic acid	0.66	1.11	0.92	0.80
14	β -Hydroxybutyric acid	0.43	0.60	0.60	0.47
15	Lactic acid	0.55	1.07	0.93	0.70
16	Malic acid	1.30	1.80	1.55	1.45
17	NTA	4.82	6.41	5.41	4.98
18	Oxalic acid	2.31	3.00	2.55	2.54
19	Propionic acid	0.34	0.50	0.54	0.43
20	Succinic acid	1.57	1.20	1.20	0.90
21	Tartaric acid	1.95	1.80	1.36	1.94
22	Tetrametaphosphate	4.90	5.20	5.17	2.80

DRAGON [20] softwares were calculated. After submitting the calculated descriptors to the regression routine, a few suitable multiple linear regression (MLR) models were obtained. We used SPSS software package version 16 [21] for this processing and selected the best model. A full list of five calculated molecular descriptors entered in the best selected models along with their chemical meaning are given in table 2; calculated values of five descriptors for all 22 ligands are given in table 3.

2.3. QSPR modeling

2.3.1. MLR and PLS analysis. MLR is a common method used in quantitative structure–property relationship (QSPR) study. The QSPR equation was obtained by forward stepwise multiple regression techniques following the multi-linear forms $SC = b_0 + b_1D_1 + b_2D_2 + \dots + b_nD_n$, where SC was stability constant in logarithmic scale ($\log K$), D1, D2, and Dn were the descriptors. The intercept (b_0) and the regression coefficients of the descriptors (b_1, b_2, \dots, b_n) were determined by using the least squares method and n was the number of the descriptors. Statistical evaluation of the data was obtained by the software SPSS [21]. In MLR analysis, the descriptors in the regression equation must be independent; to reduce the number of the descriptors and minimize the overlapped information in the descriptors, the concept of non-redundant descriptors [22, 23] was used. The correlation coefficients between two descriptors should be < 0.9 . Table 4 shows the correlation matrix of the descriptors which proved that the selected descriptors are independent.

After calculation of molecular descriptors, stepwise regression routine was used to develop QSPR models. The best obtained linear model and the regression coefficients of the descriptors for the stability constants of metal complexes are listed in table 5. As can be seen from this table, five descriptors including mean atomic van der Waals volume (Mv), number of double bond (nDB), complementary information content (order 2) (CIC2), number of total secondary C (SP³) (nCS), and the energy of highest occupied molecular orbital (HOMO) are entered in the best models. For regression analysis, the data set was separated into two groups: training set including 14 compounds and test set including eight compounds (table 6). Obtained models were used for calculating the stability constants of different complexes in training and test sets. Calculated stability constants of Mg, Ba, Ca, and Sr complexes using obtained MLR models are shown in table 6.

Table 2. Definition of selected descriptors in the MLR models.

Descriptor	Group	Definition
1-Mv	Constitutional descriptor	Mean atomic van der Waals volume (scale on carbon atom)
2-nDB	Constitutional descriptor	Number of double bonds
3-CIC2	Topological descriptor	Complementary information content (neighborhood of symmetry of 2-order)
4-nCS	Functional groups	Number of total secondary C (SP ³)
5-HOMO	Quantum chemical	The energy of HOMO

Table 3. Calculated values of five molecular descriptors for all 22 ligands.

No.	Ligand name	Descriptor Mv	nDB	CIC2	nCS	HOMO
1	Acetic acid	0.53	1	0.59	0	-11.62
2	ADP	0.61	3	0.67	1	-8.80
3	Aspartic acid	0.55	2	0.62	1	-10.68
4	ATP	0.61	4	0.84	1	-9.09
5	<i>n</i> -Butyric acid	0.53	1	0.91	2	-11.45
6	Citric acid	0.57	3	1.41	2	-11.64
7	EDTA	0.56	4	2.42	6	-9.76
8	Gluconic acid	0.53	1	1.42	1	-11.08
9	Glutamic acid	0.55	2	0.84	2	-10.52
10	Glutaric acid	0.56	2	1.50	3	-11.56
11	Glyceric acid	0.53	1	0.52	1	-11.42
12	Glycine	0.52	1	0.4	1	-10.29
13	Glycolic acid	0.53	1	0.44	1	-11.66
14	β -Hydroxybutyric acid	0.53	1	0.58	1	-11.20
15	Lactic acid	0.53	1	0.56	0	-11.30
16	Malic acid	0.56	2	0.72	1	-11.71
17	NTA	0.57	3	1.79	3	-10.66
18	Oxalic acid	0.58	2	1.00	0	-11.96
19	Propionic acid	0.53	1	0.61	1	-11.44
20	Succinic acid	0.56	2	1.29	2	-11.59
21	Tartaric acid	0.55	2	1.25	0	-11.62
22	Tetrametaphosphate	0.60	4	2.00	0	-12.39

Table 4. Correlation matrix for the five selected descriptors.

	Mv	nDB	CIC2	nCS	HOMO
Mv	1.000	–	–	–	–
nDB	0.868	1.000	–	–	–
CIC2	0.406	0.676	1.000	–	–
nCS	0.061	0.392	0.608	1.000	–
HOMO	0.356	0.388	-0.042	0.346	1.000

Table 5. Regression coefficients of the descriptors in MLR models for Ba, Ca, Mg, and Sr complexes.

X	Constant	Mv	nDB	CIC2	nCS	HOMO
Ba	10.763(\pm 7.57)	-19.298(\pm 13.42)	1.361(\pm 0.43)	1.376(\pm 0.54)	0.055(\pm 0.18)	0.212(\pm 0.23)
Ca	25.224(\pm 13.44)	-44.639(\pm 23.82)	2.149(\pm 0.76)	1.368(\pm 0.96)	0.119(\pm 0.31)	0.36(\pm 0.41)
Mg	30.851(\pm 11.49)	-50.405(\pm 20.38)	2.348(\pm 0.65)	0.577(\pm 0.82)	-0.051(\pm 0.27)	0.529(\pm 0.35)
Sr	21.055(\pm 11.15)	-33.407(\pm 19.76)	1.572(\pm 0.63)	0.929(\pm 0.79)	0.159(\pm 0.26)	0.451(\pm 0.34)

Modeling by the partial least square (PLS) regression method was performed by using MINITAB [24]. The PLS method uses a few latent variables, so-called PLS factors, to describe a given response y . These factors are deducted from the original variables (X) and constructed to maximize the covariance between X (the matrix of molecular descriptors), and y (the log K). For regression analysis, the data set was separated into two groups: a training set including 14 compounds and a test set

Table 6. The observed and calculated log K values – training and test sets for the MLR and PLS models for Ba, Ca, Mg, and Sr complexes.

No.	Ba			Ca			Mg			Sr		
	Log $K_{(EXP)}$	Log $K_{(MLR)}$	Log $K_{(PLS)}$	Log $K_{(EXP)}$	Log $K_{(MLR)}$	Log $K_{(PLS)}$	Log $K_{(EXP)}$	Log $K_{(MLR)}$	Log $K_{(PLS)}$	Log $K_{(EXP)}$	Log $K_{(MLR)}$	Log $K_{(PLS)}$
1 ^a	0.39	0.25	0.28	0.53	0.34	0.62	0.51	0.68	1.06	0.43	0.23	0.59
2 ^a	2.36	2.19	2.56	2.82	2.31	2.95	3.11	2.83	3.44	2.50	2.21	2.55
3 ^a	1.14	1.52	1.42	1.16	2.10	1.70	2.43	2.48	1.72	1.48	1.75	1.35
4	3.29	3.72	3.15	3.60	4.59	3.69	4.00	5.12	4.12	3.03	3.80	3.14
5	0.31	0.83	0.27	0.51	1.08	0.14	0.53	0.85	0.30	0.36	0.92	0.09
6	2.30	3.42	3.18	3.50	4.20	3.83	2.80	3.72	3.25	2.80	3.10	2.92
7	7.78	6.99	7.50	10.7	9.33	9.69	8.69	7.94	8.32	8.63	7.43	7.55
8	0.95	1.56	1.06	1.21	1.79	0.80	0.70	1.39	0.61	1.00	1.40	0.52
9	1.28	1.91	1.89	1.43	2.57	2.25	1.90	2.64	2.13	1.37	2.18	1.76
10	2.04	2.46	2.40	1.06	2.77	2.90	1.08	1.92	2.53	0.60	2.15	2.22
11	0.80	0.24	-0.03	1.18	0.43	-0.25	0.86	0.69	0.03	0.89	0.41	-0.19
12	0.77	0.51	0.83	1.43	1.12	1.37	3.45	1.72	1.79	0.91	1.14	1.21
13	0.66	0.09	0.44	1.11	0.24	0.74	0.92	0.52	1.12	0.80	0.24	0.67
14 ^a	0.43	0.38	0.26	0.60	0.59	0.14	0.60	0.84	0.36	0.47	0.57	0.11
15 ^a	0.55	0.27	0.43	1.07	0.41	0.59	0.93	0.83	0.90	0.70	0.34	0.52
16	1.30	1.24	1.11	1.80	1.41	1.18	1.55	1.49	1.16	1.45	1.03	0.90
17	4.82	4.21	4.66	6.41	5.19	5.97	5.41	4.40	5.17	4.98	4.06	4.64
18	2.31	1.13	2.19	3.00	0.69	2.99	2.55	0.56	2.87	2.54	0.36	2.38
19	0.34	0.37	0.47	0.50	0.56	0.72	0.54	0.74	1.04	0.43	0.49	0.63
20 ^a	1.57	2.10	2.36	1.20	2.35	2.95	1.20	1.82	2.65	0.90	1.77	2.28
21 ^a	1.95	2.13	1.84	1.80	2.49	2.11	1.36	2.40	1.85	1.94	1.74	1.60
22 ^a	4.90	4.75	4.00	5.20	5.31	4.72	5.17	4.60	3.89	2.80	3.57	3.58

^aMolecules selected as test set

including eight compounds which are the same as MLR sets (table 6). The number of significant factors for the PLS algorithm was determined using the cross-validation method. With cross-validation, one sample was kept out (leave-one-out (LOO)) of the calibration and used for prediction. The process was repeated so that each of the samples was kept out once. The predicted values of left-out samples were then compared to the observed values using prediction error sum of squares (PRESS). The PRESS obtained in the cross-validation was calculated each time that a new principal component (PC) was added to the model. The optimum number of PLS factors is the one that minimizes PRESS. Calculated stability constants using PLS models for training and test sets are shown in table 6.

2.3.2. Artificial neural network. For artificial neural network (ANN) production, the data set was divided into three groups: training, test, and validation sets. All molecules were randomly chosen for these sets. The training, test, and validation sets consisted of 14, four, and four molecules, respectively. The training, validation, and test data sets were used to optimize the network performance. The training set was used for model generation and the test set was used to take care of overtraining. The validation set was used to evaluate the generated model. MATLAB version 7.6 [25] was used for implementing three-layer feed forward back propagation ANN with Levenberg–Marquardt algorithm. The proper number of nodes in the hidden layer was determined

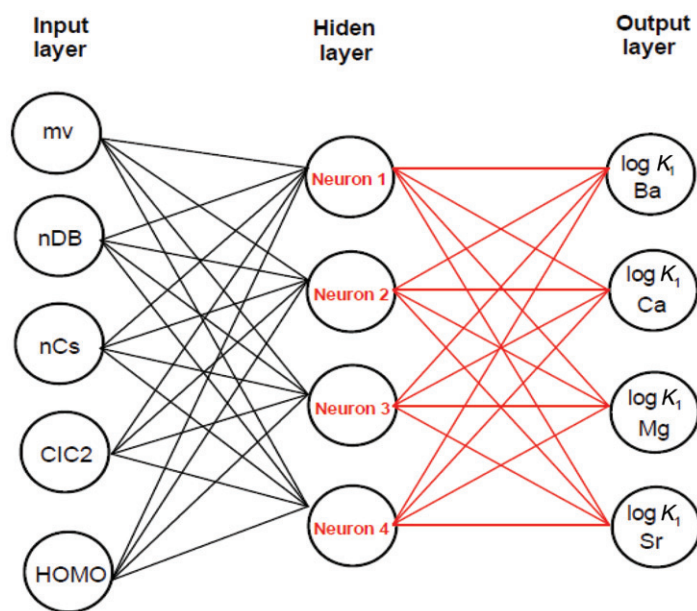


Figure 1. Structure of designed ANN.

by training the network with different numbers of nodes in the hidden layer. In this work, four neurons were selected. Five molecular descriptors in the best linear model (MLR) were used as inputs for ANN-LM and four output values were the $\log K$ (stability constant) of Ba, Ca, Mg, and Sr complexes. The structure of optimized ANN-LM is shown in figure 1. Calculated stability constants using ANN-LM model for training, test, and validation sets are shown in table 7. Statistical parameters including determination coefficient (R^2) and SE for MLR, PLS, and ANN-LM models for different sets are shown in table 8.

2.4. Model validation

2.4.1. Y-randomization. Part of validating the models is to check for the possibility of chance correlations. This can be done by performing the entire sequence of computations over but with the dependent variables scrambled. This scrambling destroys any relationship between the descriptors and the dependent variable. No model that exceeds chance performance should be found. The results obtained are compared to the results achieved with the actual computations to demonstrate that the actual results were achieved by finding relationships rather than by finding chance correlations. Results of this test are compiled in table 9.

2.4.2. Cross-validation technique. The consistency and reliability of a method can be explored using the cross-validation technique. Two different strategies of LOO and leave-multiple-out (LMO) can be carried out in this method. In LOO strategy, by

Table 7. The observed and calculated log K values – training, validation, and test sets for the ANN-LM model for Ba, Ca, Mg, and Sr complexes.

No.	Ba		Ca		Mg		Sr	
	Log $K_{(EXP)}$	Log $K_{(ANN-LM)}$	Log $K_{(EXP)}$	Log $K_{(ANN-LM)}$	Log $K_{(EXP)}$	Log $K_{(ANN-LM)}$	Log $K_{(EXP)}$	Log $K_{(ANN-LM)}$
Train.								
4	3.29	3.12	3.60	3.54	4.00	3.98	3.03	3.05
5	0.31	0.47	0.51	0.53	0.53	0.46	0.36	0.39
6	2.30	2.52	3.50	3.34	2.80	2.83	2.80	2.67
7	7.78	7.81	10.7	10.69	8.69	8.71	8.63	8.61
8	0.95	0.93	1.21	1.19	0.70	0.75	1.00	1.02
9	1.28	1.32	1.43	1.46	1.90	1.90	1.37	1.33
10	2.04	2.02	1.06	1.08	1.08	1.11	0.60	0.62
11	0.80	0.58	1.18	1.05	0.86	0.99	0.89	0.80
12	0.77	0.82	1.43	1.42	3.45	3.38	0.91	0.90
13	0.66	0.60	1.11	1.15	0.92	1.00	0.80	0.84
16	1.30	1.35	1.80	1.82	1.55	1.41	1.45	1.52
17	4.82	4.73	6.41	6.44	5.41	5.32	4.98	5.04
18	2.31	2.31	3.00	3.02	2.55	2.55	2.54	2.41
19	0.34	0.38	0.50	0.67	0.54	0.54	0.43	0.56
Valid								
2	2.36	2.58	2.82	2.62	3.11	3.45	2.50	2.41
3	1.14	1.77	1.16	2.12	2.43	2.28	1.48	1.83
15	0.55	0.54	1.07	0.78	0.93	0.56	0.70	0.73
22	4.90	4.14	5.20	5.50	5.17	4.91	2.80	4.33
Test								
1	0.39	0.35	0.53	0.54	0.51	0.23	0.43	0.54
14	0.43	0.57	0.60	0.98	0.60	1.02	0.47	0.76
20	1.57	1.36	1.20	1.48	1.20	1.16	0.90	1.18
21	1.95	1.95	1.80	2.66	1.36	2.16	1.94	2.16

Table 8. Statistical parameters for training, test, and validation sets using MLR, PLS, and ANN-LM models for Ba, Ca, Mg, and Sr complexes.

Method	Set	Ba		Ca		Mg		Sr	
		R^2	SE	R^2	SE	R^2	SE	R^2	SE
MLR	Train.	0.897	0.692	0.840	1.170	0.820	1.020	0.823	0.997
	Test	0.963	0.313	0.834	0.684	0.914	0.504	0.823	0.421
PLS	Train.	0.960	0.429	0.921	0.824	0.905	0.740	0.914	0.694
	Test	0.901	0.508	0.781	0.786	0.728	0.897	0.738	0.512
ANN-LM	Train.	0.997	0.120	0.999	0.081	0.999	0.074	0.999	0.077
	Valid	0.945	0.556	0.918	0.680	0.972	0.361	0.834	0.481
	Test	0.999	0.001	0.959	0.147	0.734	0.269	0.987	0.099

deleting each time one object from training set, a number of models will be produced. Obviously, the number of models produced by the LOO procedure is equal to the number of available examples. In the case of LMO, M represents a group of randomly selected data points, which would leave out at the beginning and would be predicted by the model, which was developed using the remaining data points. So, M molecules are considered as prediction set.

Table 9. Regression coefficient (R^2) and SE values for Y-randomization tests.

Y-randomization	Ba		Ca		Mg		Sr	
	R^2	SE	R^2	SE	R^2	SE	R^2	SE
1	0.236	1.856	0.239	2.424	0.213	2.092	0.208	1.939
2	0.264	1.822	0.254	2.401	0.255	2.035	0.262	1.872
3	0.195	1.905	0.186	2.507	0.142	2.185	0.223	1.920
4	0.215	1.882	0.165	2.540	0.242	2.053	0.183	1.970
5	0.192	1.909	0.151	2.561	0.129	2.201	0.142	2.019

3. Results and discussion

The goal of this work was developing a QSPR model to predict the stability constants of Mg, Ba, Ca, and Sr complexes using a small subset of common descriptors. Therefore, the development of a robust and interpretable QSPR model, which is able to accurately predict the stability constants, is necessary.

In general terms, the stability constant of a metal complex can be calculated as follows: $K = [ML]/[M][L]$, where K is the stability constant, M is the amount of metal ion, and L is the amount of a ligand. From the definition of K , the stability constant depends on three factors: the concentrations of free metal, free ligand, and complex. Clearly, K relies heavily upon the concentration of the complex. The total concentration of metal, C_M can be computed. The basic equation $C_M = [M] + [ML]$ with $[ML] = K [M][L]$ becomes $C_M = [M] (1 + K [L])$; hence $[M] = C_M / (1 + K [L])$ shows that the concentration of M depends on the stability constant of the complex and free concentration of the ligand which is dependent upon corresponding pK and pH values. Fortunately, the metal ions in the studied complexes are the same and this allowed us to ignore the ions, considering only the ligands.

As the first step, we developed a linear model of MLR, whose specifications are given in table 5. This model has been developed considering two purposes. First, a stepwise MLR procedure was used to select suitable variables. It can be seen from table 5 that descriptors of M_v , nDB , $CIC2$, nCS , and $HOMO$ were chosen. These descriptors can be classified as constitutional (M_v , nDB), topological ($CIC2$), functional group (nCS), and quantum chemical ($HOMO$) descriptors. The description of these descriptors is given in table 2. These descriptors indicate cation–ligand interactions (nDB), steric effects (M_v , $CIC2$, nCS), and electronic interactions ($HOMO$) of the ligands on the stability constant.

The second purpose of developing a MLR model was to assess the linear relationship between these descriptors and the stability constant parameters. We used the proposed linear model to interpret the stability constant values. This means we should investigate the variables that are most important predictors among the five descriptors appearing in the MLR models. In the case of the MLR, the mean effect of each descriptor can be considered as a measure of its role in predicting the stability constant. Mean effect is defined as:

$$MF_j = \frac{\beta_j \sum_{i=1}^{i=n} d_{ij}}{\sum_j \beta_j \sum_{i=1}^{i=n} d_{ij}}, \quad (1)$$

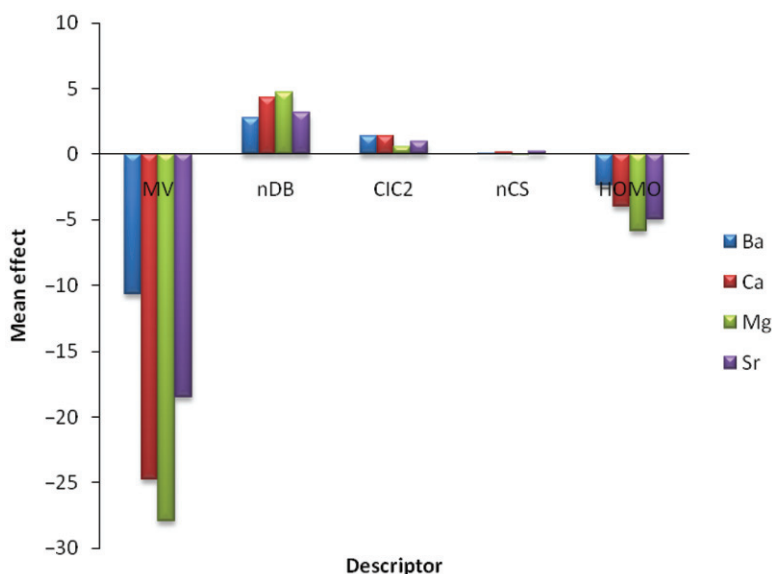


Figure 2. Mean effects of descriptors in MLR model.

where MF_j represents the mean effect for the considered descriptor j , β_j is the coefficient of the descriptor j , d_{ij} stands for the value of the target descriptors for each molecule and, eventually, m is the descriptor number in the model. The MF value indicates the relative importance of a descriptor, compared with the other descriptors in the model. Its sign exhibits the variation direction in the values of the stability constant as a result of the increase (or reduction) of these descriptor values.

Figure 2 shows the mean effect of each variable in the MLR model for all Ba, Mg, Ca, and Sr complexes. As can be seen from this figure, Mv, nDB, and HOMO are the most important parameters affecting the stability constant of the complexes; these effects for all metal complexes are the same. As can be seen from this figure, stability constant values decrease with increasing Mv values and increases with increasing nDB values. Also, higher values of HOMO could lead to decrease in log K values. Therefore, with increasing HOMO, the complex is less stable. The HOMO values are equal to the ionization potential with a negative sign; the more easily the ligand ionizes the more stable the complex will be.

In this QSPR modeling, according to selected descriptors, complexation is mainly related to: cation (acceptor)/ligand (donor) interactions (nDB), steric effects (Mv, CIC2, nCS), and electronic interactions (HOMO). Figure 3 indicates the changes of stability constants against the number of double bonds in the ligands. Topological features of the ligands involving the nature and number of donors in the ligand, and the type of substituents on the ligand, play important roles in both the selectivity and stability of the resulting complexes. The MLR and PLS calculated values of stability constants for training and test sets are shown in table 6.

As a second step, we were interested to investigate the non-linear characteristics of the stability constant and develop a network that could predict the stability constant of different metal complexes including Ba, Mg, Ca, and Sr, simultaneously. Therefore, a

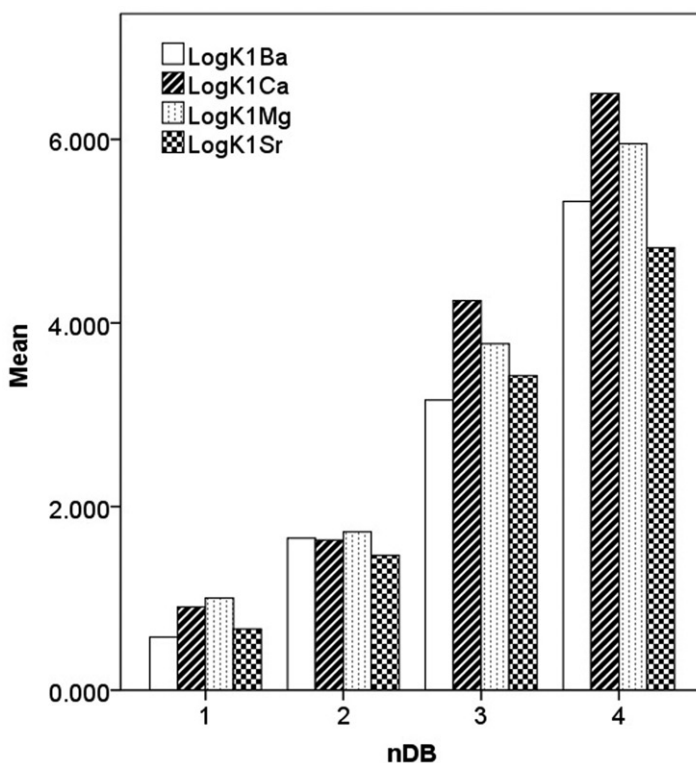


Figure 3. Plot of log K values against the descriptor values of nDB.

feed forward propagation ANN was developed using the descriptors appearing in the MLR model as its inputs. It is a common practice to optimize the parameters of number of nodes in the hidden layer, learning rate, and momentum in developing a reliable network. The procedures for optimizing these parameters are given elsewhere [26].

A back propagation network with a 5–4–4 architecture was developed using Levenberg–Marquardt algorithm (ANN-LM) to predict the stability constants of Ba, Mg, Ca, and Sr complexes. The calculated values of stability constant for training, test, and validation sets using ANN-LM are shown in table 7.

The consistency and reliability of the ANN-LM and MLR models are illustrated using the cross-validation technique. In particular, the leave-4-out (L4O) procedure was utilized in this work, which produces a number of models by deleting four objects from the training set. Calculations of R^2 and SE were based on three random selections of groups of four molecules from the data set (table 10).

The consistency of the R^2 values for L4O modified data sets indicates that the proposed model is reliable. For the sake of comparison, a PLS analysis was also performed using all variables. Table 8 shows that results of the ANN-LM model are superior compared with those of the PLS model.

In order to ensure the robustness of the ANN-LM model, the Y-randomization test was performed. The dependent variable vector (stability constant) was randomly

Table 10. Cross-validation results for ANN-LM model.

Set	Number of molecules ^a	Ba		Ca		Mg		Sr		
		R ²	SE	R ²	SE	R ²	SE	R ²	SE	
1	Test	3,11,18,19	0.934	0.314	0.914	0.339	0.993	0.081	0.945	0.137
	Valid	1,6,10,17	0.992	0.252	0.928	0.931	0.999	0.788	0.871	0.915
	Train.		0.970	0.410	0.980	0.445	0.937	0.670	0.968	0.398
2	Test	8,10,16,17	0.872	0.676	0.889	0.715	0.868	0.677	0.963	0.261
	Valid	4,5,18,19	0.933	0.421	0.992	0.160	0.996	0.120	0.761	0.695
	Train.		0.996	0.140	0.992	0.253	0.982	0.312	0.989	0.230
3	Test	4,6,20,21	0.931	0.181	0.830	0.460	0.910	0.500	0.867	0.340
	Valid	7,11,13,16	0.999	0.144	0.999	0.139	0.994	0.412	0.999	0.610
	Train.		0.991	0.152	0.983	0.234	0.968	0.297	0.969	0.205

^aNumbers refer to the number of the compounds given in table 1. The remaining molecules for each set are due to the corresponding training set.

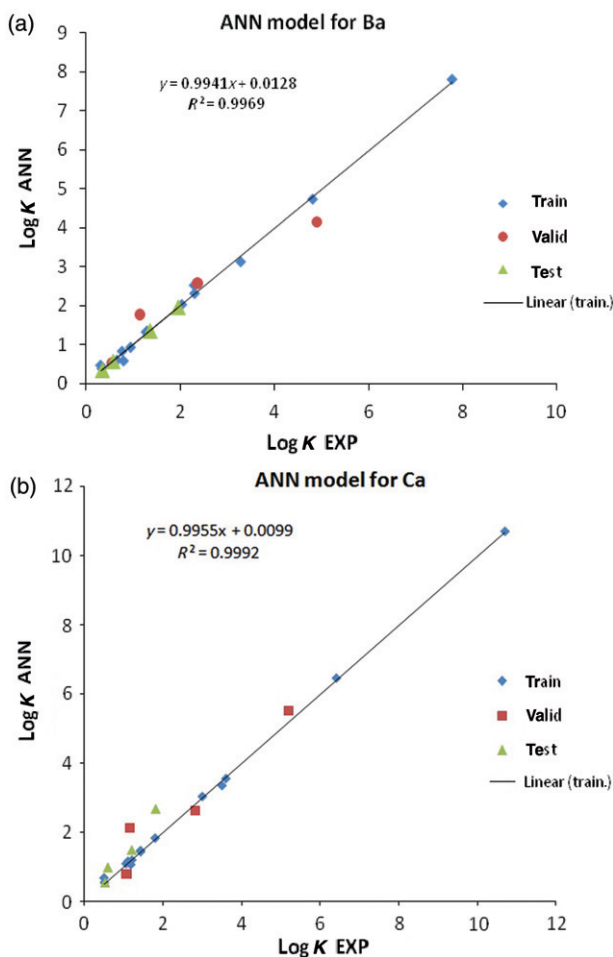


Figure 4. Plots of predicted vs experimental values of log K for ANN-LM model: (a) Ba, (b) Ca, (c) Mg, (d) Sr.

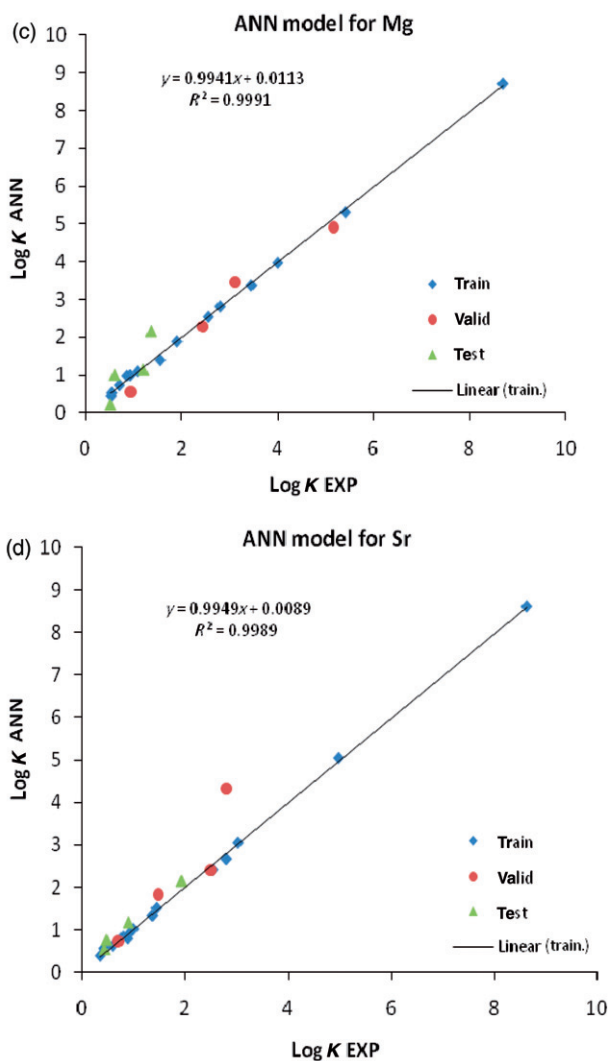


Figure 4. Continued.

shuffled and a new QSPR model was developed using the original independent variable matrix. The new QSPR model is expected to have low R^2 and high SE values. Several random shuffles of the y vector were performed and the results are shown in table 9. The R^2 and SE values indicate that the good results for the ANN-LM model are not due to a chance correlation or structural dependency of the training set.

The observed and ANN-LM predicted values of the stability constant for all compounds studied in this work are shown in table 7. Figure 4(a–d) shows the plots of the ANN-LM predicted *versus* the experimental values of the stability constant for the Ba, Ca, Mg, and Sr complexes. The correlation of this plot indicates the reliability of

the models. Of course, small outliers may be seen in some plots. In this work, an ANN with four outputs has been used, simultaneously. Training of an ANN with more than one output is more difficult than ANNs with only one output. Therefore, the SEs for some output values are important.

Statistical parameters for the MLR, PLS, and ANN-LM methods are listed in table 8. The ANN-LM represented the best results for stability constants of metal complexes. Superiority of non-linear models (ANN-LM) over the linear models (MLR, PLS) revealed that the stability constant of complexes have non-linear characteristics.

4. Conclusion

This study has shown that the stability constants of complexes ($\log K$) can be modeled in terms of structure-based descriptors solely calculated from the structure of ligands. A small subset of common descriptors is able to predict the stability constants of different metal complexes including Ba, Ca, Mg, and Sr, simultaneously. Validation of the QSPR model suggests that the model can be used to make predictions for compounds not in the original data set. The structural information encoded in the descriptors in the discussed models indicates the significant and specific structure information that may be useful for predicting the stability constants of new complexes. A back propagation network with a 5–4–4 architecture was developed using Levenberg–Marquardt algorithm (ANN-LM). Obtained results using non-linear models indicated that the stability constants of complexes have non-linear characteristics.

Acknowledgment

Support of this work by the council research of Vali-e-Asr University is acknowledged.

References

- [1] I. Irving, D.R. Williams. *J. Chem. Soc.*, 3192 (1953).
- [2] G. Schwartzenbach. *Anal. Chem.*, **32**, 6 (1960).
- [3] S. Ahrland, J. Chatt, N.R. Davies. *Q. Rev. Chem. Soc.*, **12**, 265 (1958).
- [4] R.G. Pearson. *J. Am. Chem. Soc.*, **85**, 3533 (1963).
- [5] R.G. Pearson (Ed.). In *Hard and Soft Acids and Bases*, Hutchinson and Ross, Dowden (1973).
- [6] M. Tanaka. *J. Inorg. Nucl. Chem.*, **35**, 965 (1973).
- [7] M. Tabata, M. Tanaka. *Inorg. Chem.*, **27**, 3190 (1988).
- [8] N. Raos, V. Simeon. *J. Inorg. Biochem.*, **18**, 133 (1983).
- [9] J. Sabolovic, N. Raos. *Polyhedron*, **9**, 1277 (1990).
- [10] J. Sabolovic, N. Raos. *Polyhedron*, **9**, 2419 (1990).
- [11] A. Fedorowicz, J. Mavri, P. Bala, A. Koll. *Chem. Phys. Lett.*, **289**, 457 (1998).
- [12] J. Tomasi, M. Persico. *Chem. Rev.*, **94**, 2027 (1994).
- [13] R. Bonaccorsi, P. Palla, J. Tomasi. *J. Am. Chem. Soc.*, **106**, 1945 (1984).
- [14] A. Milicevi, N. Raos. *Croat. Chem. Acta*, **78**, 525 (2005).
- [15] A. Milicevic, N. Raos. *Polyhedron*, **25**, 2800 (2006).

- [16] B.G. Cox, H. Schneider. *Coordination and Transport Properties of Macrocyclic Compounds in Solution*, Chap. 2.3, Elsevier, Amsterdam (1992).
- [17] T.E. Furia. *CRC Handbook of Food Additives*, 2nd Edn, CRC Press, Cleveland, OH (1972).
- [18] *HYPERCHEM pro 8.0.5*. Available online at: www.softarchive.net
- [19] J.J.P. Stewart, *MOPAC, A Semiempirical Molecular Orbital Program* (1990).
- [20] R. Todeschini, V. Consonni, M. Pavan. *DRAGON, software, Version 3* (2003).
- [21] *SPSS, Version 16* (2008). Available online at: www.spssscience.com
- [22] A.R. Katritzky, E. Gordeeva. *J. Chem. Inf. Comput. Sci.*, **33**, 835 (1993).
- [23] J. Olivero, T. Gracia, P. Payares, R. Vivas, D. Diaz, E. Daza, P. Geerlings. *J. Pharm. Sci.*, **86**, 625 (1997).
- [24] *MINITAB, Version 15.1* (2008). Available online at: <http://www.minitab.com>
- [25] *MATLAB, Mathworks Inc., Version 7.6*. USA (2008). Available online at: <http://www.mathworks.com>
- [26] M. Jalali-Heravi, Z. Garkani-Nejad. *J. Chromatogr. A*, **971**, 207 (2002).