Anal. Bioanal. Electrochem., Vol. 13, No. 3, 2021, 423-435

Analytical & Bioanalytical Electrochemistry

2021 by CEE
www.abechem.com

Full Paper

A Quantitative Structure–Property Relationship Study on Cerium(III) Complexes with Ionophores Applied in Lanthanoid Sensors

Roya Kiani-Anbouhi*

Department of Chemistry, Faculty of Basic Sciences, Imam Khomeini International University, Qazvin, Iran

*Corresponding Author, Tel.: +982833901369

E-Mail: kiani@sci.ikiu.ac.ir

Received: 7 July 2021 / Received in revised form: 19 November 2021 / Accepted: 26 November 2021 / Published online: 31 December 2021

Abstract- Since the selectivity of an ion selective sensor is directly related to the stability constants of ion–ionophore complexes, we predicted the complexation stability (K) of cerium ions with different ionophores by the quantitative structure–property relationship model. Genetic algorithm (GA) feature selection approach was selected to choose the proper molecular descriptors which were then subjected to multiple linear regression (MLR) for prediction of the log K. The predictive ability of the built genetic algorithm-multiple linear regression (GA-MLR) model was evaluated using Leave-one-out cross-validation, Leave-group-out cross-validation, Y-randomization, and test set compounds. Statistical parameters of the model (R^2_{train}=0.852, Q^2_{LOO}= 0.813, and Q^2_{LGO}=0.777) indicated the ability of the GA-MLR model to predict the response of ionophores in cerium-selective sensors based on complex stability constants. Also, the applicability domain of the model was analyzed by the Williams plot. Based on this study, some key features are identifiable to appraise the selectivity of cerium sensors that can be used to design new selectophores.

Keywords- QSPR; Complex stability constant; Cerium-selective electrode; Genetic algorithm; Multiple linear regression

1. INTRODUCTION

Ion Selective Electrodes (ISEs) are sensors that respond selectively to ions in the presence of others. The increasing use of ion sensors in the fields of agricultural, industrial, environmental, and medicinal analysis is stimulating analytical chemists to develop new sensors for the fast, accurate, reproducible, and selective determination of various species in a wide concentration range. The sensing part of any ion selective is definitely its ion-selective membrane or selectophore. The most important mechanism has been suggested for the selective recognition of different ions in these sensors based on the ion–dipole interactions between the target ion and an ionophore [1]. Therefore, by using the stability constant of complexes between the ionophore and ions (K), the selectivity of the ion selective electrode can be predicted.

Although for the significant selectivity of a sensor, the complex stability constant must be a high enough value, but this value should not be so large as to prolong the response time of the sensor [1]. Based on previous studies, the proper range of logarithmic formation constant (log K) of the ion-selectophore complex in an ISE is usually about 4-7.

Due to the importance of detection of rare earth elements in some biological samples and extensive cerium application in metallurgical and functional material areas, selective determination of cerium (III), by various method, have always received attention [2-5]. For the design of ion selective electrodes, several experiments should be run to synthesize, and then tested to detect the selectivity behavior of the newly designed ionophores, however, these experimental methods are often time-consuming and costly. To overcome these restrictions, quantitative structure–activity relationships (QSPR) techniques emerged as a proper alternative to predict the selectivity behavior of ion selective sensors based on the stability constant of ionophores with target ions before any experimental works [6-10]. Quantitative structure-property/activity relationships (QSPR/QSAR) method have been applied as a useful tool in different disciplines of science such as chemistry, physics, drug design, medicinal chemistry, and so forth over the recent decades [11-13].

Based on the predictive QSPR model, the considered property can be obtained without any experimental efforts for synthesis and testing the novel compounds. Thus, the QSPR method can expedite the process of the development of new molecules with desired properties. One of the most important steps in constructing reliable QSAR/QSPR models is the descriptor selection that can demonstrate different aspects of the molecules by numbers.

In this work, for the first time, a QSPR model has been established for the prediction of the selectivity behavior of ionophores toward cerium(III) in ion selective electrodes. For this purpose, the proper sets of descriptors representing the best-fitted models for the regression method were selected using the GA feature selection approach and subsequently subjected to the MLR approach to the construction of the linear model for predicting the stability constants of complexes between Ce^{3+} and ionophores.

2. METHODOLOGY

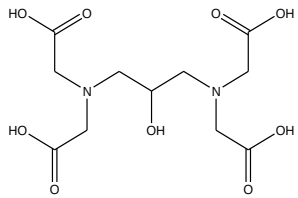
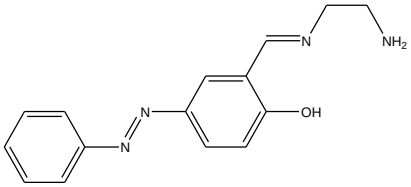
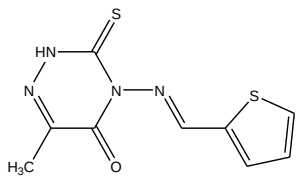
2.1. Data set

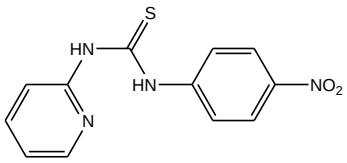
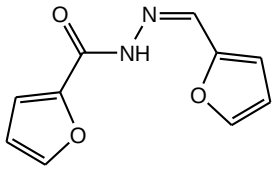
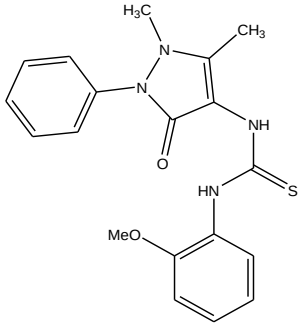
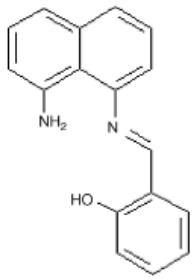
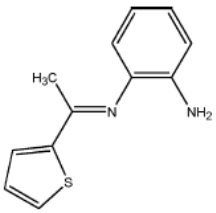
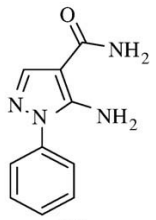
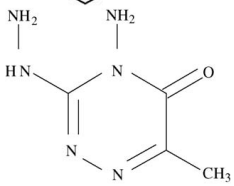
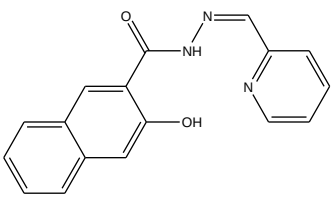
A data set including experimental stability constant values for the 1:1 complexes of Ce^{3+}cation with 28 ionophore (CeL^{3+}) was collected from the published papers by Ganjali et al. [14-41]. Since various parameters such as solvent, temperature, and measurement method affect the complex stability constants, we used only data in the same condition (conductometric method in acetonitrile solutions and temperature 25ºC). In our QSPR study, the response variable was expressed as the logarithmic scale of complex stability constants (log K), where K is defined as follows:

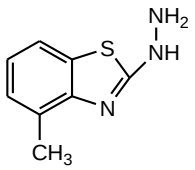
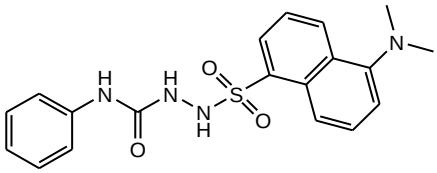
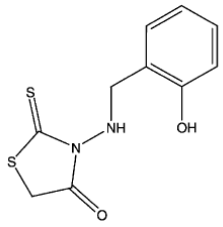
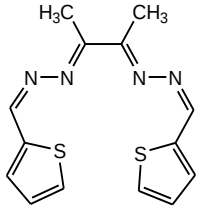
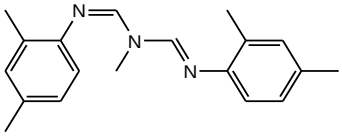
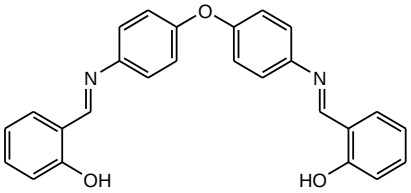
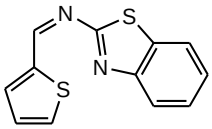
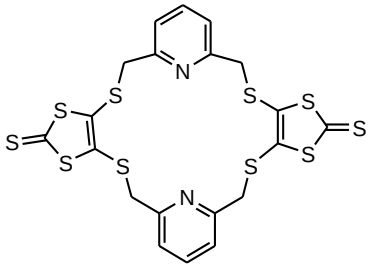
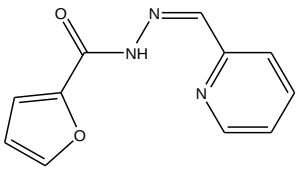
$$K = \frac{[CeL^{3+}]}{[Ce^{3+}][L]} \quad (1)$$

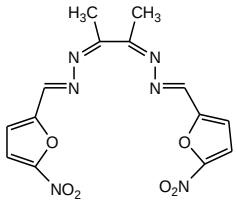
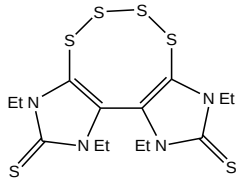
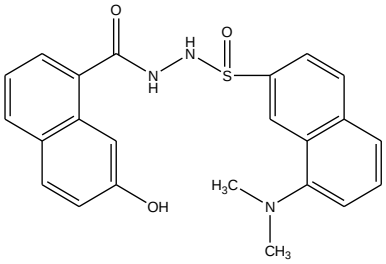
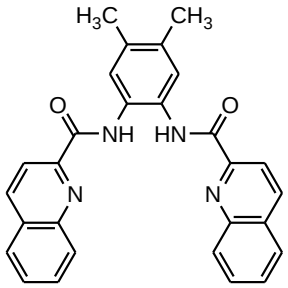
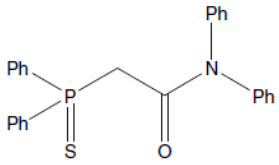
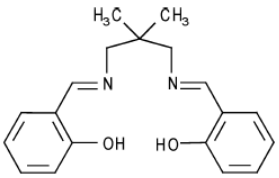
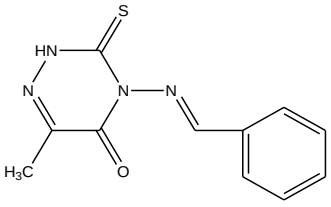
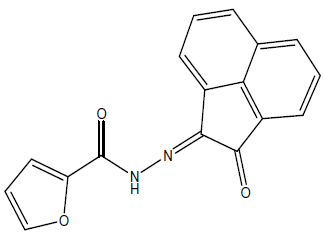
The data set was randomly divided into two sub-sets: the training set containing 23 molecules for model construction and the test set containing 5 molecules which were used to evaluate the model prediction ability. The chemical structures and the respective log K values for the compounds are listed in Table 1.

Table 1. Chemical structure and experimental and predicted values of log K for cerium-ionophore complexes by GA-MLR model[a]

No.	Structure	Log K		Ref.
		Exp.	Pred.	
1		4.26	4.27	[14]
2		2.25	2.55	[15]
3		2.93	2.94	[16]

4		2.81	2.79	[17]
5		2.36	2.51	[18]
6		4.13	4.07	[19]
7		2.77	2.54	[20]
8		2.86	2.96	[21]
9[a]		3.22	2.99	[22]
10[a]		2.76	2.86	[23]
11		2.66	2.51	[24]

12		2.95	2.95	[25]
13		2.24	2.52	[26]
14		3.16	3.02	[27]
15		3.11	2.61	[28]
16		2.95	2.98	[29]
17		2.76	2.65	[30]
18		2.80	2.52	[31]
19		2.80	2.78	[32]
20[a]		2.75	2.55	[33]

21		2.66	2.85	[34]
22		3.51	3.56	[35]
23		2.22	2.61	[36]
24		2.73	2.57	[37]
25		2.33	2.49	[38]
26		2.77	2.67	[39]
27[a]		2.32	2.51	[40]
28[a]		2.47	2.48	[41]

[a] The test set

2.2. Descriptor calculation

The 2D chemical structures of the studied molecules were first drawn and saved with the HIN extension in the Hyperchem program [42]. The geometrical optimization was performed by semi-empirical AM1 method (Austin Model-1) with the adjusted root mean square gradient of 0.01 kcal mol^{-1} in MOPAC software. In the next step, these optimized structures fed into Dragon package [43] to calculate 18 categories of the molecular descriptors such as constitutional, topological, geometrical, WHIM, radial distribution function (RDF) and charge descriptors.

After descriptors calculation, the descriptors with constant or almost constant values for all molecules were removed. Also, the descriptors having intercorrelated (correlation coefficients greater than 0.90) were detected and only one of them with the highest correlation with the response variable (log K) was considered in the development of the GA-MLR models. Therefore, the total number of 836 descriptors remained for the development of the linear models.

2.3. Genetic algorithm analysis

Since the predictive power of a model depends on the selected variables, one of the most determinative steps in QSPR analysis, is the selection step of proper descriptors. In this study, the genetic algorithm technique was employed as a selection tool to select the most relevant variables for an objective function [44-46]. Each selected group of variables would be further evaluated to predict the properties values by their fitness. In this work, the fitness function used in the genetic algorithm method was the correlation coefficient of leave-one-out cross-validation (Q^2_{LOO}) [47].

The GA-based variable selection and the other calculations were performed in the MATLAB 7.0 program [48].

3. RESULT AND DISCUSSION

After the classification of the data set (training and test sets), the genetic algorithm method was used to select the most relevant descriptors. The linear equation between the stability constants of complexes of Ce^{3+}cation with studied molecules and the calculated descriptors was obtained by the MLR. The model constructed using the GA-MLR approach consisted of four molecular descriptors as being represented in Equation (2):

$$Log\ K = 2.693 + (0.768)\ E1s + (0.546)\ nR{=}Cs + (0.480)\ N\text{-}068 + (-0.439)\ B09[C\text{-}C] \quad (2)$$

$R^2 = 0.852$, $R^2_{adj} = 0.819$, $Q^2_{LOO} = 0.813$, $Q^2_{LGO} = 0.777$, $Q^2_{Boot} = 0.766$, $R^2_{test} = 0.750$, $F = 25.99$

where R^2, R^2_{adj}, F, Q^2_{LOO}, Q^2_{LGO}, and Q^2_{Boot} are squared correlation coefficient, adjusted R^2, Fisher F statistic, squared cross-validation coefficients for leave one out, leave group out, and bootstrapping, respectively.

The obtained higher value for Q^2_{LOO} (0.813) indicates the reliability of obtained QSPR model. Cross-validation for the leave group out parameter indicates that the constructed model has a good external predictive power. The robustness of the proposed model and its predictive ability was also guaranteed by the high Q^2_{Boot} based on bootstrapping repeated 5000 times. The statistical parameters resulting from the GA–MLR method (high values of R^2, R^2_{adj}, and F) demonstrate the predictive capability of the proposed model. Also, the high values of cross-validation tests (Q^2_{LOO}= 0.813, Q^2_{LGO} = 0.777 and Q^2_{Boot} = 0.766) and R^2_{test} (0.750) confirm the high ability of the model in internal and external validation, respectively.

The predicted values of logarithmic stability constants of Ce(III) complexes with different ionophores were listed in Table 1. The prediction plot consisting of predicted versus the experimental log k values has been represented in Figure 1.

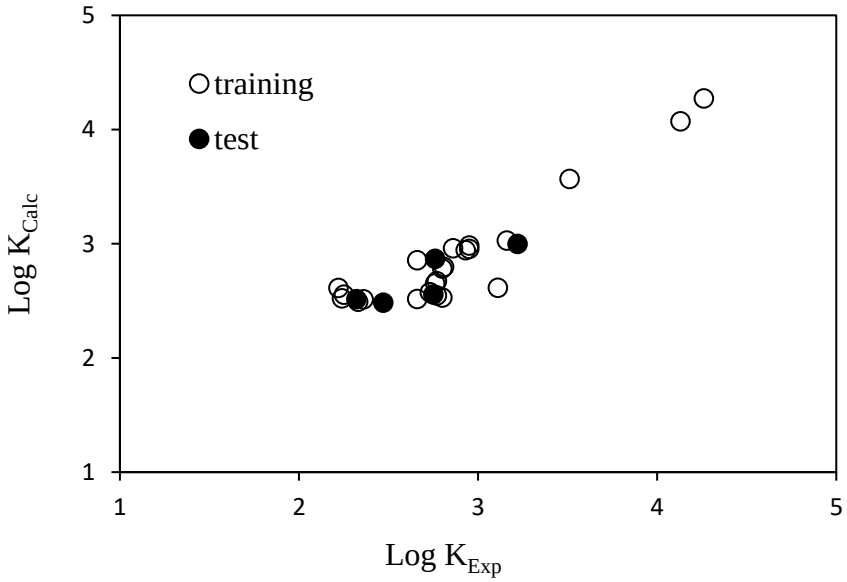


Fig. 1. The predicted versus the experimental log K values by the GA-MLR modeling

According to equation 2, four descriptors appeared in this QSPR model consist of E1s, nR=Cs, N-068, and B09[C-C]. For the evaluation of the multi-collinearity for the selected descriptors, the variation inflation factors (VIF) was usually used [49] as below:

$$VIF = \frac{1}{1-r^2} \quad (3)$$

where r is the correlation coefficient of multiple regressions between each variable and the other variables in the QSPR model. When the VIF values fall within the range 1–5, the

proposed model is acceptable and possesses enough predictive power. If the VIF value equals 1, it indicates that there is no intercorrelation for each descriptor, whereas VIF values greater than 10 stand for the inappropriateness of the constructed model. The correlation coefficient and corresponding VIF values for selected descriptors based on GA–MLR have been represented in Table 2. The tabulated data in this table reveal the selected descriptors are independent and it is no multi-collinearity between pair descriptors.

Table 2. The correlation coefficient of selected descriptors and corresponding VIF values based on GA-MLR

	E1s	nR=Cs	N-068	B09[C-C]	(VIF)[a]
E1s	1	0	0	0	1
nR=Cs	0.052	1	0	0	1
N-068	0.090	0.0506	1	0	1
B09[C-C]	0.002	0.025	0.035	1	1

[a] Variation inflation factors

According to equation 2, the selected variables are E1s, nR=Cs, N-068, and B09[C-C]. E1s (1st component accessibility directional WHIM index/weighted by atomic electrotopological states) is the first descriptor which is given in the model. This descriptor is a type of WHIM directional descriptor which is based on the statistical indices calculated on the projections of atoms along principal axes [50]. As it is clear from equation 2, the E1s plays positive effects on the log K value. Therefore, increasing the value of this descriptor1 will increase the log K value. The next descriptor is nR=Cs which describes the number of aliphatic secondary C(sp^2). This descriptor reflects the degree of unsaturation of the molecule. The positive sign of nR=Cs indicates that by increasing the degree of unsaturation of the molecule (ionophore), the interaction between ionophore and cerium ion and consequently, stability constant value increases. The third descriptor is the number of aliphatic R_3-N groups (N-068). Since this descriptor has a positive effect on the expected response, increasing the number of aliphatic R_3-N groups leads to increase of stability constant values. B09[C-C] is sub-structural descriptor that describes the presence/absence of the C-C topological fragment in 09 distance. According to this model, the negative sign of this descriptor suggests that the complexation stability constant value is inversely related to this descriptor.

The Williams plot (the plot of standardized residuals versus leverage values) was exploited to visualize the applicability domain. It can be used to obtain an immediate and simple graphical detection of both the response outliers (*Y* outliers) and the structurally influential chemicals (*X* outliers) of a QSPR model. The normal control values for Y outliers were set to

±3σ (± 3 standardized residual) and the normal control values for X outliers (h* or 3h) was calculated as 3p/n in this study, where n is the number of the calibration compound and p is the number of model variables plus one. The leverage (h) greater than the h* value suggested that the compound was very influential on the model.

According to the Williams plot (Figure 2), compound 1 has the leverage (h) more than h* value of 0.652, therefore, it can be considered as a structural outlier, but the standard residuals for this compound are within ±3σ, and therefore, there are no outlier compounds with high standard residuals.

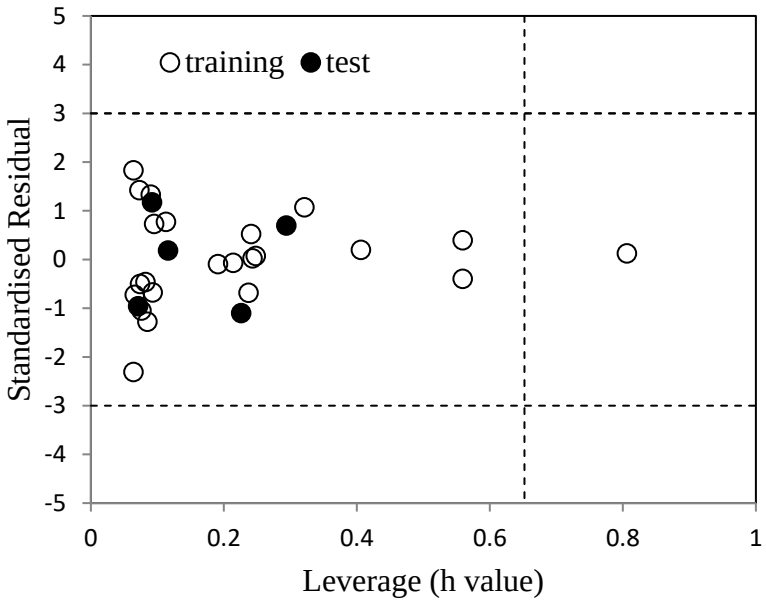


Fig. 2. The William's plot of the GA-MLR model for the training and test sets

4. CONCLUSION

A new reliable and accurate model based on GA-MLR analysis of the stability constant of cerium-ionophore complexes has been proposed, in the current report. The obtained results displayed that the GA-MLR model has a superior power to the expression of the relationship between the stability constant of cerium-ionophore complexes and the corresponding molecular descriptors. Since the response of the ISEs depends on the stability constants of ion–ionophore complexes; the derived model could be used for the prediction of the selectivity of ionophores toward Ce^{3+} ion. The developed QSPR model can be used to predict the property of new ionophores. Also, The findings of this study may provide some insights for further design of novel selectophores for Ce(III) selective sensors.

REFERENCES

[1] M.R. Ganjali, P. Norouzi, and M. Rezapour, Encyclopedia of Sensors; Potentiometric ion selective sensors. 8 (2006) 197.

[2] M. Jamaluddin Ahmed, M. Tazul Islam, and F. Farhana, RSC Adv. 9 (2019) 25609.

[3] F. Farhadinasab, G. H. Rounaghi, M. Mohajeri, and M. Esmaelpour Farkhani, Russ. J. Gen. Chem. 85 (2015) 1184.

[4] H. Bagheri, A. Afkhami, M. Saber-Tehrani, A. Shirzadmehr, S.W. Husain, H. Khoshsafar, and M. Tabatabaee, Anal. Methods 4 (2012) 1753.

[5] M. R. Ganjali, R. Kiani-Anbouhi, M. Shamsipur, T. Poursaberi, M. Salavati-Niasari, Z. Talebpour and M. Emami, Electroanalysis 16 (2004) 1002.

[6] R. Kiani-Anbouhi, M. R. Ganjali, and P. Norouzi, J. Incl. Phenom. Macrocycl. Chem. 78 (2014) 325.

[7] R. Kiani-Anbouhi, M. R. Ganjali, and P. Norouzi, J. Incl. Phenom. Macrocycl. Chem. 81 (2015) 441.

[8] E. Martynko, V. Solov'ev, A. Varnek, A.Legin, and D. Kirsanov, Electroanalysis 32 (2020) 792.

[9] V. Soloviev, A. Varnek, V. Babain, and et al., Sens. Actuators B 301 (2019) 126941.

[10] S. Riahi, M. R. Ganjali, P. Norouzi and F. Jafari, Sens. Actuators B 132 (2008) 13.

[11] J. Roy, S. Ghosh, P. Kumar Ojha, and K. Roy, Environ. Sci. Nano 6 (2019) 224.

[12] N. MinhQuang, T. XuanMau, N. ThiAi Nhung, T. NguyenMinh An, and P. Van Tat, J. Mol. Struct. 1195 (2019) 95

[13] Z. Ghomisheh, A. Ebrahimpoor Gorji, and M. A. Sobati, J. Mol. Graph. 101 (2020) 107700.

[14] H. A. Zamani, M.R. Ganjali, M. R. Abedi, and P. Norouzi, Sensor Lett. 5 (2007) 1.

[15] M. R. Ganjali, N. Davarkhah, H. Ganjali, B. Larijani, P. Norouzi, and M. Hossieni, Int. J. Electrochem. Sci. 4 (2009) 762.

[16] H. A. Zamani, G. Rajabzadeh, M.R. Ganjali, and P. Norouzi, Anal. Chim. Acta 598 (2007) 51.

[17] M. R. Ganjali, P. Norouzi, T. Alizadeh, A. Tajarodi, and Y. Hanifehpour, Sens. Actuators B Chem. 120, (2007) 487.

[18] H.A. Zamani, M. R. Ganjali, and M. Adib, Sens. Lett. 4 (2006) 1.

[19] H. A. Zamani, M. R. Ganjali, P. Norouzi, A. Tajarodi, and et.al, J. Chil. Chem. Soc, 52, (2007) 1332.

[20] M. R. Ganjali, P. Norouzi, T. Alizadeh, and M. Adib, Anal. Chim. Acta 576 (2006) 275.

[21] M. R. Ganjali, P. Norouzi, F. Faridbod, M. Ghorbani, and M. Adib, Anal. Chim. Acta 569 (2006) 35.

[22] H. A. Zamani, G. Rajabzadeh, M. Masrornia, A. Dejbord, M. R. Ganjali, and N. Seifi, Desalination 249 (2009) 560.

[23] H. A. Zamani, G. Rajabzadeh, and M. R. Ganjali, Sens. Actuators B Chem. 119 (2006) 41.

[24] M. R. Ganjali, P. Norouzi, N. Yousefian, F. Faridbod, and M. Adib, Bull. Korean Chem. Soc. 27 (2006) 1581.

[25] M. R. Ganjali, P. Norouzi, L. Shamsolahrari, and A. Ahmadi, Sens. Actuators B Chem. 114, (2006) 713.

[26] F. Faridbod, M.R. Ganjali, B. Larijani, and P. Norouzi, Electrochim. Acta 55 (2009) 234.

[27] M. R. Ganjali, Z. Memari, F. Faridbod, R. Dinarvand, and P. Norouzi, Electroanalysis 20 (2008) 2663.

[28] M. R. Ganjali, P. Norouzi, A. Daftari, F. Faridbod, and M. Salavati-Niasari, Sens. Actuators B Chem. 120 (2007) 673.

[29] M. R. Ganjali, V. Akbar, M. Ghorbani, P. Norouzi, and A. Ahmadi, Anal. Chim. Acta 531 (2005) 185.

[30] M. R. Ganjali, M. Tavakoli, F. Faridbod, S. Riahi, P. Norouzi, and M. Salavati-Niassari,. Int. J. Electrochem. Sci. 3 (2008) 1559.

[31] F. Faridbod, M.R. Ganjali, B. Larijani, M. Hosseini, and P. Norouzi, Mater. Sci. Eng. C. 30 (2010) 555.

[32] M. R. Ganjali, R. Nemati, F. Faridbod, P. Norouzi, and F. Darviche, Int. J. Electrochem. Sci. 3 (2008) 1288.

[33] H. A. Zamani, M. R. Ganjali, P. Norouzi, M. Adib, and M. Aceedy, Anal Sci. 22 (2006) 943.

[34] M. R. Ganjali, M. B. Gholivand, M. Rahimi-Nasrabadi, B. Maddah, M. Salavati-Niasari, and F. Ahmadi, Sens. Lett. 4 (2006) 1.

[35] M. Shamsipur, M. Hosseini, K. Alizadeh, Z. Talebpour, M. F. Mousavi, M. R. Ganjali, M. Arca, and V. Lippolis, Anal. Chem. 75 (2003)5680.

[36] M. R. Ganjali, H. Ganjali, M. Hosseini, and P. Norouzi, Int. J. Electrochem. Sci. 5 (2010) 967.

[37] H. A. Zamani, A. Imani, A. Arvinfar, F. Rahimi, M.R. Ganjali, F. Faridbod, and S. Meghdadi, Mater. Sci. Eng. C. 31 (2011) 588.

[38] M. R. Ganjali, P. Norouzi, F. Faridbod, S. Riahi, M. R. Yaftian, A. Zamani, and D. Matt, J. Appl. Electrochem. 37 (2007) 827.

[39] M. R. Ganjali, P. Norouzi, F. Faridbod, S. Riahi, J. Ravanshad, J. Tashkhourian, M. Salavati-Niasari, M. Javaheri, IEEE Sens. J. 7 (2007) 544.

[40] H. A. Zamani, G. Rajabzadeh, M.R. Ganjali, S. M. Khatami, Electroanalysis 17 (2005) 2260.

[41] F. Faridbod, M.R. Ganjali, M. Pirali-Hamedani, and P. Norouzi, Int. J. Electrochem. Sci. 5 (2010) 1103.

[42] HyperChem, Molecular Modeling System 7.03rd edn, Hypercube, Inc., Gainesville, FL. (2002).

[43] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, DRAGON-Software for the calculation of molecular descriptors. Version 4.0 for Windows, Talete SRL, Milan (2004).

[44] C. L. Waller, M. P. Bradley, J. Chem. Inf. Comput. Sci. 39 (1999) 345.

[45] J. Aires-de-Sousa, M.C. Hemmer, J. Gasteiger, Anal. Chem. 74 (2002) 80.

[46] S. Ahmad, M. M. Gromiha, J. Comput. Chem. 24 (2003) 1313.

[47] R. Leardi, R. Boggia, M. Terrile, J. Chemom. 6 (1992) 267.

[48] Mathworks, Genetic Algorithm and Direct Search Toolbox Users Guide, The Mathworks Inc., Natick, MA, U.S.A (2005).

[49] V. K. Agrawal, P.V. Khadikar, Bioorg. Med. Chem. 9 (2001) 3035.

[50] R. Todeschini, C. Bettiol, G. Giurin, P. Gramatica, P. Miana, E. Argese, Chemosphere. 33 (1996) 71.