Anal. Bioanal. Electrochem., Vol. 14, No. 6, 2022, 598-609

Analytical & Bioanalytical Electrochemistry

2022 by CEE
www.abechem.com

Full Paper

A QSPR Study for the Prediction of the Selectivity of Pb(II) Sensors by Stability Constants of Ion-Ionophore Complexes

Roya Kiani-Anbouhi,[1,*] and Mohammad Hadi Ghasemi[2]

[1]*Department of Chemistry, Faculty of Science, Imam Khomeini International University, Qazvin, Iran*

[2]*Applied Chemistry Research Group, ACECR-Tehran Organization, Tehran, Iran*

*Corresponding Author, Tel.: +982833901369

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

Received: 15 April 2022 / Received in revised form: 13 June 2022 / Accepted: 17 June 2022 / Published online: 30 June 2022

Abstract- The stability constants of ion-ionophore complexes can determine the selectivity of ion-selective electrodes. In this study, the quantitative structure-property relationship (QSPR) model was employed to predict the complex stability constants of lead ions with different ionophores. The Genetic algorithm-multiple linear regression method (GA-MLR) developed models based on calculated molecular descriptors. Y-randomization testing, cross-validation, and test set compounds were applied to evaluate the predictive ability of the built model. This built model obtained high statistical quantities (R^2_{train}= 0.899, R^2_{adj} = 0.877, Q^2_{LOO} = 0.831, Q^2_{LOO} = 0.776, and Q^2_{boot} = 0.780) and showed that GA-MLR was a promising tool to predict the complexation stability of pb^{2+} with different ionophores. The current study introduces an efficient model for testing and assessing selectophores in lead-selective sensors based on complex stability constants. Additionally, this model could guide the design of highly selective ionophores for Pb (II) sensors.

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

1. INTRODUCTION

Among the methods for low-level determination of ions in solutions, potentiometric ion-selective sensors are widely used as the conventional method, which offers an inexpensive and simple way for monitoring several cations and anions within acceptable sensitivity and selectivity. It is well known that ion-selective electrodes (ISEs) are quickly and easily used for chemical, clinical, and environmental analyses [1,2]. The highly selective interactions between the sensing material (ionophore) and the target species give these sensors selectivity [3].

As the importance of controlling pollution levels in various biological and environmental samples, researchers increasingly develop new sensors to detect heavy metals such as lead ions. Lead is the most widely used heavy metal. Its industrial use has increased due to many of its properties over the last decades. This element has been described as toxic, slow-acting, and producing a variety of symptoms [4]. Therefore, developing ion-selective electrodes for determination in various actual samples is both necessary and challenging .

An ion-selective potentiometric electrode relies entirely on its sensing material, so selecting the suitable sensing material is one of the main steps in designing ISEs. It was suggested that the mechanism for the selective recognition of different ions within those sensors is based on ion-dipole interactions that result in selective complexation between the target ions and the sensing material [3]. Thus, the stability constants of complexes between lead ion and ionophores (K) indicate the ionophore's selectivity towards the ion. In contrast, the selectivity of ISEs can be predicted fully from the stability constants of the complexes. Despite the requirement for high stability constants for a complex between an ion and ionophore to exhibit selectivity, these constants must not be so high that ions are tightly bound, thus kinetically irreversible, and the sensor shows a long response time. A helpful range of ion-selectophore complex formation constants from previous studies is typically about 104-107.

The increase in the industrial use of lead, on the one hand, and its health hazards, on the other, have caused research into neutral ionophores for the preparation of Pb(II) selective sensors to be relatively widely pursued [5-8]. It is necessary to run several experiments to determine the selectivity of newly designed ionophores, but there are time constraints and extremely high costs associated with the experiments. Consequently, it would be interesting to make a prediction model of the selectivity of ion-selective sensors based on the stability constant of ionophores with target ions before conducting any experiments [9-11]. Here, the well-known quantitative structure-property/activity relationship (QSPR or QSAR) is developed and has proven to be a valuable tool for predicting various physical, chemical, biological, and toxicological properties of compounds based on experimental data and molecular structures [12-15]. By using the QSPR/QSAR method, the considered property can be readily ascertained without requiring any experiments to synthesize and test new compounds. Therefore, this approach can quickly develop materials and molecules with desired properties.

The experimental data in QSPR models are related to physicochemical properties, used to create the model as dependent variables. The parameters are descriptors that need molecular structures. Only those in correlation with the considered property should select some of these descriptors. Therefore, one of the essential steps of the QSPR study is to employ a technique to choose the respective variables. After that, various modelling methods are used to build the model.

For the first time, we developed a QSPR model to predict ionophores' selectivity behavior toward lead(II) applied to ion-selective electrodes. To attain this goal, we used Genetic algorithm-multiple linear regression (GA-MLR) as a variable selection method for the QSPR model that predicted the stability constants of complexes between lead ions and ionophores.

2. MATERIALS AND METHODS

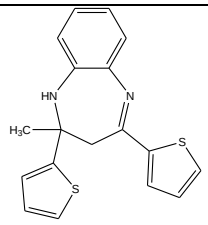
2.1. Data set

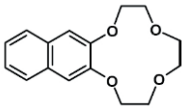
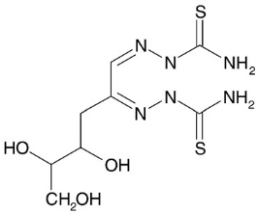
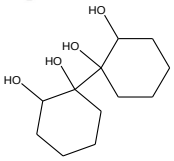
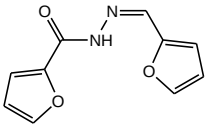
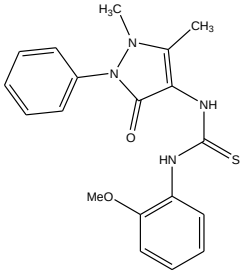
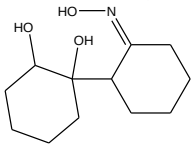
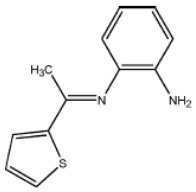
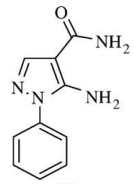
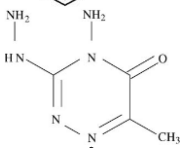
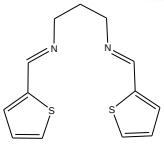
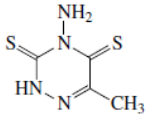
The data set used in this study consists of experimental stable constant values for 1:1 complexes of Pb^{2+}cation with 30 organic ligands, as reported by Ganjali et al. [16-45]. A logarithmic constant log K_f was used to develop the QSPR model, where K is defined as follows:

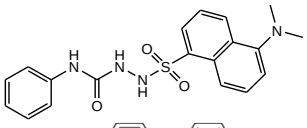
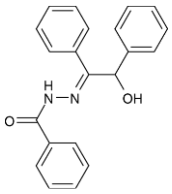
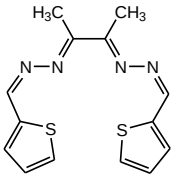
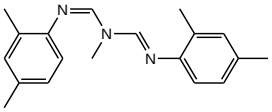
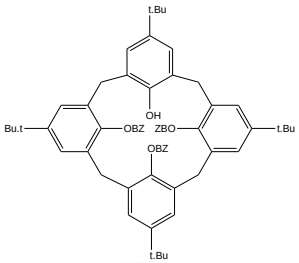
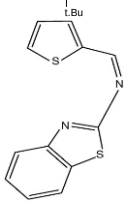
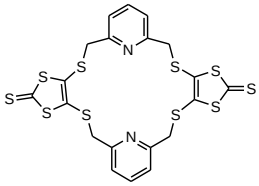
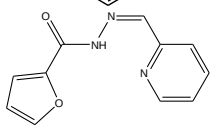
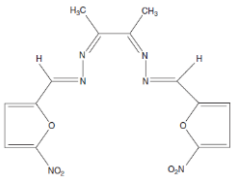
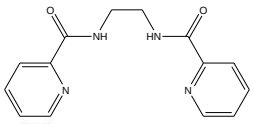
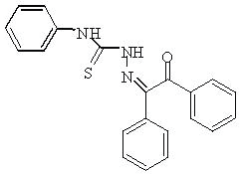
$$K_f = \frac{[PbL^{2+}]}{[Pb^{2+}][L]} \quad (1)$$

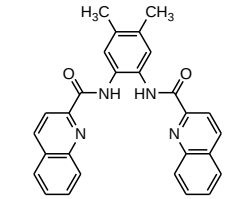
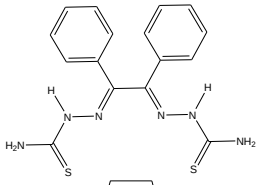
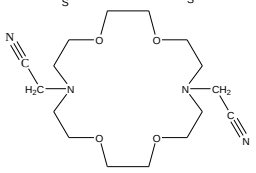
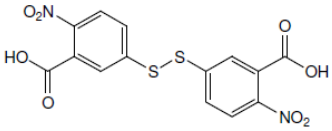
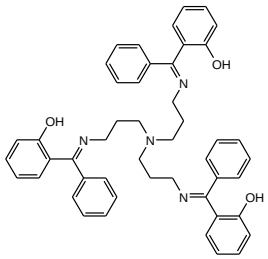
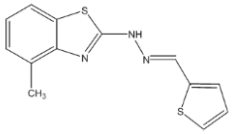
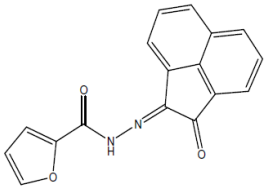
According to the conductometric method, the stability constants were obtained in acetonitrile solution at 25ºC. The chemical structures of the molecules and the corresponding values of log K_f are given in Table 1. Two subsets of a data set were randomly chosen: a training set consisting of 24 molecules (80%); and a test set consisting of 6 molecules (20%). The QSPR model was constructed from the training set compounds, whereas the test set molecules were used to evaluate the built model's ability to predict the response. The test set molecules were marked and listed in Table 1.

Table 1. Experimental and predicted values of log K_f for Pb^{2+}-ionophore complexes for training and test sets by GA-MLR model[a]

No.	Structure	Log K_f		Ref.
		Exp.	Pred.	
1[a]		2.44	1.95	[16]

2		2.29	2.54	[17]
3		2.74	3.18	[18]
4[a]		4.09	3.18	[19]
5		2.90	2.54	[20]
6		2.23	2.26	[21]
7		2.45	2.86	[22]
8		2.44	2.87	[23]
9[a]		2.74	2.86	[24]
10		2.73	3.18	[25]
11[a]		3.35	2.55	[26]
12		2.33	2.55	[27]

13		2.10	2.26	[28]
14		2.74	2.86	[29]
15		2.24	2.55	[30]
16		2.31	2.55	[31]
17[a]		2.86	2.54	[32]
18		2.20	1.95	[33]
19[a]		2.10	1.95	[34]
20		2.42	2.54	[35]
21		2.95	2.55	[36]
22		2.74	2.54	[37]
23		2.9	2.55	[38]

24		3.44	3.18	[39]
25		4.42	3.18	[40]
26		6.90	6.85	[41]
27		6.22	6.26	[42]
28		2.86	2.54	[43]
29		2.74	2.55	[44]
30		2.27	2.54	[45]

[a] The test set

2.2. Descriptor calculation

A two-dimensional (2D) chemical structure of each of the studied molecules was drawn in the Hyperchem program [46] and pre-optimized via molecular mechanics force fields (MM+). MOPAC software was used for the final optimization with a root mean square gradient of 0.01 kcal mol^{-1} using the semi-empirical AM1 method. After optimization, these optimized structures are fed into the Dragon program [47] to calculate the molecular descriptors such as geometrical, constitutional, topological, and charge descriptors. We scanned the derived descriptors and excluded any constants or near-constant variables. The correlation of the remaining descriptors to log K_f values was calculated during the next step,

and those detected as collinear descriptors (r > 0.90) were removed. Ultimately, 763 descriptors remained for further analysis.

2.3. Genetic algorithm analysis

QSPR models should be constructed with the most relevant descriptors to achieve the highest accuracy and precision. The genetic algorithm is the most common approach for selecting this best collection of variables (GA). GA begins with a chromosome, which contains several randomly chosen variables, and each chromosome contains genes [48,49]. Each chromosomal subgroup is assessed for its ability to predict property values. In this investigation, the fitness function in the genetic algorithm approach was the correlation coefficient of leave-one-out cross-validation (Q2LOO) [50]. The ultimate mutation happens after executing the incorrect subset of variables and breeding the remaining subsets, and the most relevant descriptors are picked. The GA approach was implemented as a selection tool in Matlab 6.5 [51]. This study used GA-MLR to select descriptors and develop a suitable QSPR model.

3. RESULTS AND DISCUSSION

In the modeling process, the genetic algorithm chose the most relevant descriptors. The training set chemicals and chosen descriptors were subjected to multiple linear regression analysis. According to the GA-MLR analysis, the following equation describes the prediction of log K_f complexes:

$$Log\ K_f = 2.552 + (0.317)\ nHBonds + (4.306)\ B05[C\text{-}O] + (-0.598)\ B05[N\text{-}S] + (-4.310)\ B06[C\text{-}O] \quad (2)$$

$$R^2 = 0.899,\ R^2_{adj} = 0.877,\ Q^2_{LOO} = 0.831,\ Q^2_{LGO} = 0.776,\ Q^2_{Boot} = 0.780,\ R^2_{test} = 0.714,\ F = 42.3$$

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

This higher result for Q^2_{LOO} (0.831) suggests that the QSPR model is more reliable. Leave group out (LGO) cross-validation involves removing 20% of the data points from the dataset and refitting the model, which compares the predicted with the experimental values. Each data point is removed only once during this process. The obtained regression model shows an excellent external predictive power based on cross-validation for the leave group out parameter.

The bootstrapping results (high Q^2_{Boot}) also guaranteed the robustness and predictive ability of the proposed model. During bootstrap validation, random training sets with sample repetitions are created, and then predicted responses of samples not included in the training set are assessed. Each verified model went through the bootstrapping procedure 5,000 times.

The GA–MLR method's statistical characteristics (high R^2, R^2_{adj}, and F) reflect the proposed model's predictive power. Also, the results produced by the cross-validation tests

include LOO (Q^2_{LOO}= 0.831) and LGO (Q^2_{LGO} = 0.776), and bootstrapping (Q^2_{Boot} = 0.780) illustrated the quality of the obtained model. Table 1 lists the predicted logarithmic stability constants of Pb(II) complexes, including various ionophores. In addition, the experimental log K_f values are shown versus the predicted log K_f values in Figure 1.

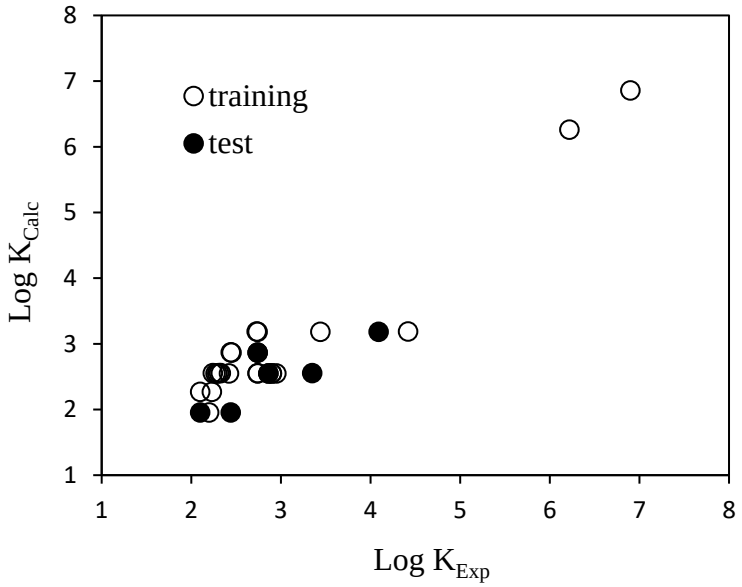


Figure 1. The plot of the calculated complex stability constant (log K_f) against the experimental values

According to equation 2, four descriptors appeared in this QSPR model consisting of nHBonds, B05[C-O], B05[N-S], and B06[C-O], with mean effects %3.09, %49.96, %2.19, and %44.74 respectively. The first selected descriptor is the number of intramolecular hydrogen bonds (nHBonds). The formation of intramolecular hydrogen bonds has a very pronounced effect on molecular structure. According to this model, it can be speculated that this effect leads to increasing the interaction between ionophore and $Pb^{2+,}$ and consequently, the complexation stability constant value increases. B05[C-O] and B06[C-O] are sub-structural descriptors that describe the presence/absence of the C-O topological fragment in 05 or 06 distance, respectively. It is reasonable to assume that the presence of C-O fragment at a proper distance describes the charge-dipole interaction between oxygen lone pairs in ionophores and lead cation. B05[N-S] descriptor is the presence/absence of N-S at topological distance 05. This descriptor has a few effects (ME% = 2.19) on the stability constant, but this descriptor's insertion in the model improves the reliability.

The developed model's robustness was assessed using Y-randomization [52]. We shuffled log K_f values for the set of chemicals before building a new model using the randomized data. In Table 2, we presented the results for the new QSPR models (after ten repetitions). According to Table 2, all new R^2 and Q^2_{LOO} values were lower than those obtained for non-shuffled data, suggesting that the excellent model was not due to chance.

Table 2. The Q^2_{LOO} and $R^2_{training}$ values after several Y-randomization tests

No.	R^2	Q^2
1	0.296	0.088
2	0.223	0.113
3	0.309	0.075
4	0.097	0.094
5	0.184	0.097
6	0.092	0.100
7	0.261	0.292
8	0.115	0.084
9	0.316	0.003
10	0.217	0.105

Williams plot, which plots residuals (δ) vs. leverage values (h), was performed to show the applicability domain (AD) to understand possible compounds as outliers [53]. In this research, the normal control values for Y outliers were set to ±3σ, and for X outliers (h* or 3h) were computed as 3p/n, where p is the number of model variables plus one and n is the number of the objects used to create the model. The William plot is shown in Fig. 2. All compounds have leverage lower than the warning h* values of 0.625, indicating that all compounds have good leverage. There are also no outlier chemicals with standard residuals over 3δ. The acceptability of the constructed model to predict the property using the GA-MLR approach is shown in the figure.

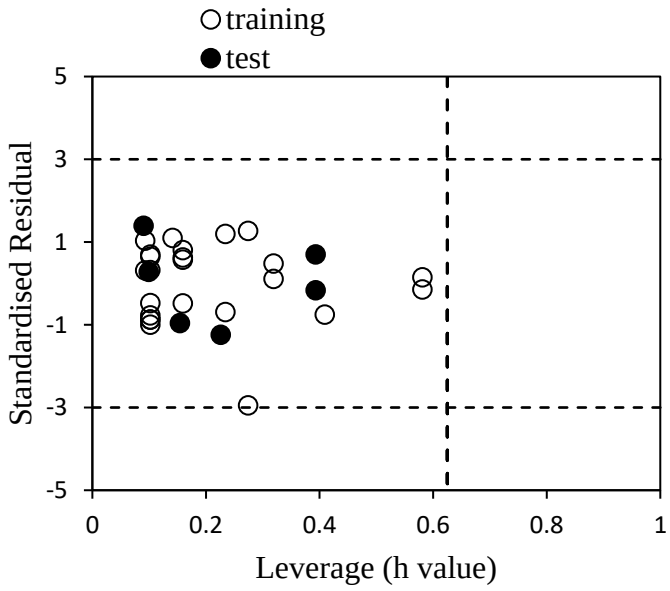


Figure 2. The Williams plot of the GA-MLR model for the training and test sets

4. CONCLUSION

As the ionophore is the most critical component of any ion-selective sensor, we have developed a new accurate and reliable QSPR model to predict ionophore selectivity toward Pb^{2+} based on stability constants of the lead-ionophore complex. According to the results obtained with the derived GA-MLR model, the model's predictive ability was acceptable. The QSPR model can also predict the properties of new compounds, and the information derived from that can be used to design highly selective ionophores for lead sensors.

REFERENCES

[1] F. Faridbod, M. R. Ganjali, R. Dinarvand, and P. Norouzi, African J. Biotechnol. 6 (2007) 2960.

[2] M. R. Ganjali, P. Norouzi, F. Faridbod, M. Rezapour, and M. R. Pourjavid, Iranian Chem. Soc. 4 (2007) 1.

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

[4] [Online] available at: https://en.wikipedia.org/wiki/Lead_poisoning

[5] W. J. Zhang, C. Y. Li, X. B. Zhang, and Z. Jin, Anal. Lett. 40 (2007) 1023.

[6] A. Alva, A. Widinugroho, M. Adrian, D. S. Khaerudini, S. E. Pratiwi, and A. S. Abd Aziz, J. Electrochem. Soc. 166 (2019) 1513.

[7] X. G. Li, H. Feng, M. R. Huang, G. L. Gu, and M. Moloney, Anal. Chem. 84 (2012) 134.

[8] M. R. Ganjali, N. Motakef-Kazami, F. Faridbod, S. Khoee, and P. Norouzi, J. Hazard. Mater. 173 (2010) 415.

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

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

[11] R. Kiani- Anbouhi, Anal. Bioanal. Electrochem. 13 (2021) 423.

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

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

[14] F. Dioury, A. Duprat, G. Dreyfus, C. Ferroud, and J. Cossy, Chem. Inf. Model. 54 (2014) 2718.

[15] V. Nadezhda, P. Valery, A. Julia, B. Vasily, L. Andrey, and K. Dmitry, Chemosensors 10 (2022) 2.

[16] M. R. Ganjali, P. Norouzi, T. Alizadeh, and M. Adib, Brazilian Chem. Soc. 17 (2006) 1217.

[17] M. R Ganjali, P. Norouzi, F. Faridbod, A. Sepehrifard, M. Ghandi, and A. Moghimi, Can. J. Anal. Sci. Spectrosc. 52 (2007) 46.
[18] H.A. Zamani, M.T. Hamed-Mosavian, E. Aminzadeh, M. R. Ganjali, M. Ghaemy, H. Behmadi, and F. Faridbod, Desalination 250 (2010) 56.
[19] M. Javanbakht, A. Shabani-Kia, M. Raouf Darvich, M. R. Ganjali, and M. Shamsipur, Anal. Chim. Acta 408 (2000) 75.
[20] H. A. Zamani, M. R. Ganjali, and M. Adib, Sens. Lett. 4 (2006) 1.
[21] H. A. Zamani, M. R. Ganjali, P. Norouzi, A. Tajarodi, and Y. Hanifehpour, J. Chil. Chem. Soc. 52 (2007) 1332.
[22] M. R. Ganjali, F. Mizani, M. Emami, M. Darjezini, M. Raouf Darvich, and M. Yousefi, Anal. Sci. 20 (2004) 531.
[23] M. R. Ganjali, P. Norouzi, F. Faridbod, M. Ghorbani, and M. Adib, Anal. Chim. Acta 569 (2006) 35.
[24] H. A. Zamani, G. Rajabzadeh, M. Masrornia, A. Dejbord, M. R. Ganjali, and N. Seifi, Desalination 249 (2009) 560.
[25] H. A. Zamani, G. Rajabzadeh, and M. R. Ganjali, Sens. Actuators B 119 (2006) 41.
[26] M. R. Ganjali, T. Poursaberi, L. Haji-agha Babaei, S. Rouhani, M. Yousefi, M. Kargar-Razi, A. Moghimi, H. Aghabozorg, and M. Shamsipur, Anal. Chim. Acta 440 (2001) 81.
[27] H. A. Zamani, G. Rajabzadeh, A. Firouz, and M. R. Ganjali, J. Anal. Chem. 62 (2007) 1080.
[28] F. Faridbod, M. R. Ganjali, B. Larijani, and P. Norouzi, Electrochim. Acta 55 (2009) 234.
[29] M. R. Ganjali, F. Faridbod, P. Norouzi, and M. Adib, Sens. Actuators B 120 (2006) 119.
[30] M. R. Ganjali, P. Norouzi, A. Daftari, F. Faridbod, M. Salavati-Niasari, Sens. Actuators B 120 (2007) 673.
[31] M. R. Ganjali, V. Akbara, M. Ghorbania, P. Norouzia, and A. Ahmadi, Anal. Chim. Acta 531 (2005) 185.
[32] M. Hosseinia, M. Rahimi, H. Bagheri Sadeghi, S. Taghvaei-Ganjali, S. Dehghan Abkenara, and M. R. Ganjali, Intern. J. Environ. Anal. Chem. 89 (2009) 407.
[33] F. Faridbod, M. R. Ganjali, B. Larijani, M. Hosseini, P. Norouzi, Mater. Sci. Eng. C 30 (2010) 555.
[34] M. R. Ganjali, R. Nemati, F. Faridbod, P. Norouzi, F. Darviche, Int. J. Electrochem. Sci. 3 (2008) 1288.
[35] H. A. Zamani, M. R. Ganjali, P. Norouzi, M. Adib, M. Aceedy, Anal. Sci. 22 (2006) 943.
[36] M. R. Ganjali, M. B. Gholivand, M. Rahimi-Nasrabadi, B. Maddah, M. Salavati-Niasari, and F. Ahmadi, Sens. Lett. 4 (2006) 1.
[37] M. R. Ganjali, P. Norouzi, A. Atrian, F. Faridbod, S. Meghdadi, and M. Giahi, Mater. Sci. Eng. C 29 (2009) 205.
[38] H. A. Zamani, M. R. Ganjali, and F. Faridbod, Serb. Chem. Soc. 76 (2011) 1295.

[39] H. A. Zamani, A. Imani, A. Arvinfar, F. Rahimi, M. R. Ganjali, F. Faridbod, and S. Meghdadi, Mater. Sci. Eng. C 31 (2011) 588.
[40] M. R. Ganjali, M. Hosseini, M. Salavati-Niasari, T. poursaberi, M. Shamsipur, M. Javanbakht, and O. R. Hashemid, Electroanalysis 14 (2002) 526.
[41] M. R. Ganjali, M. Hosseini, F. Basiripour, M. Javanbakht, O. R. Hashemi, M. Faal Rastegar, M. Shamsipur, and G. W. Buchanen, Anal. Chim. Acta 464 (2002) 181.
[42] F. Faridbod, M. R. Ganjali, B. Larijani, M. Hosseini, K. Alizadeh, and P. Norouzi, Int. J. Electrochem. Sci. 4 (2009) 1528.
[43] M. Hosseini, H. Bagheri Sadeghi, M. Rahimi, M. Salavati-Niasari, S. Dehghan Abkenar, K. Alizadeh, M. R. Ganjali, Electroanalysis 21 (2009) 859.
[44] M. R. Ganjali, P. Norouzi, B. Akbari-Adergani, Electroanalysis 19 (2007) 1145.
[45] F. Faridbod, M. R. Ganjali, M. Pirali-Hamedani, and P. Norouzi, Int. J. Electrochem. Sci. 5 (2010) 1103.
[46] HyperChem Molecular modeling system, 7.03rd Edn. Hypercube, Gainesville (2002).
[47] R. Todeschini, V. Consonni, A. Mauri, M. Pavan: DRAGON-Software for the calculation of molecular descriptors.5.3rd Edn. Talete SRL, Milan (2005).
[48] J. H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, Michigan (1975).
[49] J. Ghasemi, and S. Saaidpour, Anal. Chim. Acta 604 (2007) 99.
[50] R. Leardi, R. Boggia, and M. Terrile, Chemomet. 6 (1992) 267.
[51] M. A. Natick, and Mathworks, Genetic Algorithm and Direct Search Toolbox Users Guide, The Mathworks Inc., USA (2005).
[52] A. Tropsha, P. Gramatica, and V. K. Gombar, QSAR Comb. Sci. 22 (2003) 69.
[53] L. Eriksson, E. Johansson, M. Muller, and S. Wold, J. Chemomet. 14 (2000) 599.

Copyright © 2022 by CEE (Center of Excellence in Electrochemistry)

ANALYTICAL & BIOANALYTICAL ELECTROCHEMISTRY (http://www.abechem.com**)**

Reproduction is permitted for noncommercial purposes.