Analytical & Bioanalytical Electrochemistry

> 2023 by CEE www.abechem.com

Full Paper

# A Novel QSPR Approach in Modeling Selectivity Coefficients of the Lanthanum-Selective Electrode

Roya Kiani-Anbouhi,<sup>1,\*</sup>and Zeinab Mozafari<sup>2</sup>

<sup>1</sup>Department of Chemistry, Faculty of Science, Imam Khomeini International University, Qazvin, Iran <sup>2</sup>Department of Chemistry, Shahrood University of Technology, Shahrood, Semnan, Iran

\*Corresponding Author, Tel.: +982833901369 E-Mail: <u>kiani@sci.ikiu.ac.ir</u>

Received: 4 November 2023 / Received in revised form: 25 December 2023 / Accepted: 25 December 2023 / Published online: 31 December 2023

Abstract- This study presents a pioneering application of a novel quantitative structureproperty relationship (QSPR) model to predict the selectivity coefficients of a cation-selective electrode. Specifically, the selectivity coefficients of a Lanthanum (La(III)) membrane sensor utilizing 8-amino-N-(2-hydroxybenzylidene) naphthylamine (AIP) as the sensing ligand were efficiently estimated and predicted. To establish the QSPR model, calculated molecular descriptors were employed, considering the limitation of cation descriptors. A new strategy was introduced for descriptor calculation by optimizing the structure of Mn<sup>+</sup>-AIP and utilizing density functional theory (DFT) with the B3LYP functional and SBKJC basis set. Genetic algorithm (GA) and stepwise techniques were employed for descriptor selection, with the most significant descriptors identified. Following variable selection, multiple linear regression (MLR) was employed to construct linear QSPR models. Comparative analysis revealed that the GA-MLR modeling approach exhibited superior performance compared to the stepwise-MLR method. Furthermore, the predictions generated by the GA-MLR model demonstrated excellent agreement with the experimental values. The proposed strategy outlined in this study has the potential to be extended to other QSPR investigations involving cation-selective electrodes. These findings contribute to the advancement of predictive modeling in the field of cation-selective sensors and offer valuable insights for future research in this area.

**Keywords-** Selectivity coefficient; Multiple linear regression (MLR); Genetic algorithm; Molecular descriptors; QSPR; Chemometrics

# **1. INTRODUCTION**

Ion-selective electrodes (ISEs) are electrochemical sensors that exhibit a specific response to the presence of ionic species [1-4]. The selectivity of these sensors is recognized as one of their most crucial characteristics, as it plays a significant role in determining the feasibility of obtaining accurate measurements in the target sample [5-8]. This attribute holds particular importance in clinical applications of ion-selective electrodes, where maintaining a low emf deviation (error) is essential for whole blood or serum measurements, with a maximum allowable deviation of 0.1 mV [9]. The response behavior of ion-selective membrane electrodes in the presence of interfering ions is commonly characterized and quantified using selectivity coefficients, denoted as  $k_{sel}$  [10-12]. These coefficients provide a means to describe the influence of interfering ions on the overall response of the electrodes.

As laboratory-based determination of selectivity coefficients can be a time-consuming and costly process, there arises a need for the application of theoretical methods to efficiently and economically estimate the selectivity coefficients of chemicals.

Quantitative structure-property relationships (QSPRs) investigations have emerged as a significant field in chemometrics, biological chemistry, medicinal chemistry, and various other disciplines [13-18]. The QSPR methodology has proven effective in predicting diverse chemical and physical properties, including selectivity coefficients (k<sub>sel</sub>) of electrochemical sensors. To date, only a limited number of attempts have been made to develop QSPR models specifically focused on selectivity coefficients of ion-selective electrodes [9-12]. However, despite the abundance of cation sensors, studies on the prediction of selectivity coefficients for cation-selective electrodes are very little reported due to the limited descriptors available for cations.

To achieve improved prediction outcomes, the effective implementation of the QSPR method necessitates careful selection of input descriptor sets. These descriptors are associated with a diverse range of physicochemical properties and biological activities [19]. Currently, quantum chemically derived descriptors are widely utilized as the predominant descriptors in QSPR investigations [20,21].

The availability of semi-empirical, ab initio, and Density Functional Theory (DFT) methods has significantly advanced our ability to acquire comprehensive information about chemical compounds, including their geometry and charge distribution. These quantum mechanical methods are founded on the Schrödinger equation [22], which describes molecular systems by means of their respective molecular wave functions. However, solving the Schrödinger equation exactly for molecular systems is often impractical, necessitating approximate solutions. Ab initio structure methods typically treat the potential energy operator as time-independent and focus solely on solving the spatial wave function. These calculations can be performed at the Hartree-Fock level of approximation [23] or utilize various post-

Hartree-Fock theories such as configuration-interaction, multiconfiguration, and self-consistent field methods.

The B3LYP density functional is widely employed in quantum chemistry studies, primarily owing to its inclusion in commonly used quantum chemistry software packages and the extensive experience of the computational chemistry community with this functional. Consequently, there exists a substantial body of knowledge regarding the accuracy that can be attained through calculations employing the B3LYP functional. However, it is well-known that the B3LYP functional, along with several other density functional, lacks the capability to adequately describe London dispersion interactions [24].

For all calculations in this study, we employed the parallel version of Gamess software. It should be noted that commonly used general basis sets like 3-21G, 6-31G, LANL2MB, and others are not suitable for lanthanide ions. Instead, a specialized basis set such as SBKJC must be utilized. Performing quantum mechanical calculations with the SBKJC basis set is a challenging and time-consuming task. Therefore, to overcome these computational demands, all calculations were executed using a supercomputer.

In the field of QSPRs, Multiple Linear Regression (MLR) is a commonly employed modeling method [25-27]. MLR models provide a straightforward and easily interpretable approach compared to other modeling techniques. In this study, the variable selection in the MLR method was carried out using the genetic algorithm selection method (GA). GA is a stochastic optimization method that incorporates the principles of Darwinian evolution, employing fitness criteria and genetic functions such as crossover and mutation [28]. However, due to the inherent limitation of GA-MLR as a "black box" approach, the stepwise-MLR technique was also employed for modeling purposes. The results obtained from both GA-MLR and stepwise-MLR were compared to assess their respective performances.

The objective of this study is to develop a novel QSPR model for the prediction of selectivity coefficients in Lanthanum-selective electrodes. Given the limited number of descriptors available for cations, a new approach was employed to calculate these descriptors.

## 2. EXPERIMENTAL SECTION

## 2.1. Electrode preparation

The preparation of the viscous solution utilized for membrane formation involved the combination of 30 mg of powdered PVC, 61 mg of NOPE, 2 mg of KTpCIPB additive, and 7 mg of AIP in 5 ml of THF. The resulting mixture, characterized by its low viscosity, was thoroughly mixed and transferred into a 2 cm diameter glass dish. Subsequently, the solvent was gradually evaporated to yield a concentrated oily mixture. The membrane was then formed on the tip of a Pyrex tube with an outer diameter of 3-5 mm. The tube was dipped into the mixture for approximately 10 seconds, allowing for the formation of a transparent membrane with a thickness of around 0.3 mm. After removing the tube and allowing sufficient time for

drying at room temperature for approximately 10 hours, the tube was filled with an internal filling solution containing  $1.0 \times 10^{-3}$  M LaCl<sub>3</sub>. To condition the electrode, it was immersed in a  $1.0 \times 10^{-3}$  M La(NO<sub>3</sub>)<sub>3</sub> solution for 24 hours. An internal reference electrode coated with silver/silver chloride was employed for the electrode setup. The electrode under consideration was prepared following the methodology reported by Ganjali et al. serving as a reference for the experimental procedure [29].

## 2.2. Data set

The selectivity coefficients for a total of 30 cations were obtained using the Matched Potential Method (MPM) and served as the dataset for this study (see Table 1). Among the cations listed in Table 1 (1 to 19), the experimental data for 19 cations were previously reported by Ganjali et al. [29]. The selectivity coefficients for the remaining 11 cations were determined as part of this research, following the methodology described in reference [29]. Notably, each selectivity coefficient was calculated as the average of five measurements. Subsequently, the dataset was randomly partitioned into two distinct groups: a training set, encompassing 24 cations, and a prediction set, comprising 6 cations (refer to Table 1).

No.	Cation	$\operatorname{Log} \mathrm{K}_{A,B}^{MPM}$	No.	Cation	$\mathrm{Log}~\mathrm{K}_{A,B}^{MPM}$
1	Na <sup>+</sup>	-5.7	16	$Nd^{3+}$	-3.9
2	$\mathbf{K}^+$	-5.6	17	$Dy^{3+}$	-4.1
3	$\mathrm{Hg}^{2+}$	-6.0	18	Eu <sup>3+</sup>	-4.0
4	Cu <sup>2+</sup>	-5.2	19	Lu <sup>3+</sup>	-4.4
5	Co <sup>2+</sup>	-5.9	20	Ba <sup>2+</sup>	-4.8
6	$Mg^{2+}$	-5.3	21	$Cs^+$	-6.5
7	Ca <sup>2+</sup>	-5.2	22 <sup>t</sup>	Li <sup>+</sup>	-6.1
8	Ni <sup>2+</sup>	-5.4	23	$Rb^+$	-6.3
9 <sup>t</sup>	$Zn^{2+}$	-5.6	24	Sr <sup>2+</sup>	-4.9
10 <sup>t</sup>	$Pb^{2+}$	-4.8	25	$Tb^{3+}$	-3.5
11	Yb <sup>3+</sup>	-3.2	26 <sup>t</sup>	Er <sup>3+</sup>	-3.4
12 <sup>t</sup>	Gd <sup>3+</sup>	-3.9	27	$Ho^{3+}$	-3.5
13	Ce <sup>3+</sup>	-3.6	28 <sup>t</sup>	Fe3+	-3.3
14	Pr <sup>3+</sup>	-2.6	29	$Tm^{3+}$	-3.4
15	Sm <sup>3+</sup>	-4.2	30	$Cd^{2+}$	-4.3

Г	ab	le	1.	Data	set
---	----	----	----	------	-----

## 2.3. Descriptor generation

## 2.3.1. Computational calculation

Since, selectivity coefficient value of a cation depends on the amount of interaction between AIP and cation; we investigated effect of each cation on molecular structure of AIP and calculated structural descriptor of AIP in presence cation. For this purpose, structure of uncomplexed AIP and complex of M<sup>n+</sup>-AIP were optimized and finally, the structural changes of AIP as the result of interaction with cations as descriptors were calculated. Therefore, chemical structures of 30 studies molecules were drawn with the Hyperchem software then primal precise optimization is done with MM+ (molecular mechanics) in presence of cations and saved with the HIN extension. Quantum chemical descriptors and optimized geometries were obtained after optimizing by Gamess (Firefly) using DFT/B3LYP with the SBKJC basis set.

To investigate the influence of each cation on the molecular structure of AIP and calculate the corresponding structural descriptors, it was recognized that the selectivity coefficient of a cation is contingent upon the extent of interaction between AIP and the cation. To this end, both the uncomplexed AIP structure and the M<sup>n+</sup>-AIP complex were subjected to optimization procedures. This facilitated the calculation of structural changes in AIP resulting from cation interaction, which were subsequently employed as descriptors. The chemical structures of the 30 molecules of interest were constructed using Hyperchem software. Subsequently, initial optimization was performed employing MM+ (molecular mechanics) in the presence of cations, and the resulting structures were saved with the HIN extension. Quantum chemical descriptors and optimized geometries were then obtained by applying Gamess (Firefly) optimization using the DFT/B3LYP method with the SBKJC basis set.

To investigate the influence of each cation on the molecular structure of AIP and calculate the corresponding structural descriptors, it was recognized that the selectivity coefficient of a cation is contingent upon the extent of interaction between AIP and the cation. To this end, both the uncomplexed AIP structure and the Mn+-AIP complex were subjected to optimization procedures. This facilitated the calculation of structural changes in AIP resulting from cation interaction, which were subsequently employed as descriptors. The chemical structures of the 30 molecules of interest were constructed using Hyperchem software. Subsequently, initial optimization was performed employing MM+ (molecular mechanics) in the presence of cations, and the resulting structures were saved with the HIN extension. Quantum chemical descriptors and optimized geometries were then obtained by applying Gamess (Firefly) optimization using the DFT/B3LYP method with the SBKJC basis set.

#### 2.3.2. Quantum chemical descriptors

Due to the similarity among the compounds in the dataset, the relative differences in descriptor values are relatively small. Thus, it is crucial to determine the descriptors with high

precision. To achieve this, quantum chemical calculations were chosen for descriptor calculation. Quantum chemical calculations provide valuable insights into the impact of molecular geometry and conformations on chemical systems. The optimized 3D geometry of the molecules, obtained through quantum chemical calculations, can offer useful information in this regard. Figures 1 and 2 depict the optimized structures of both the uncomplexed AIP and its Cu<sup>2+-</sup>AIP complex, while Table 2 presents the quantum chemical descriptors calculated using Gamess [30].

No.	Definition	Notation
1	Mulliken charges of all atoms	МС
2	Mulliken charges of cations	MCC
3	One electron energy	OEE
4	Two electron energy	TEE
5	Nuclear repulsion energy	NRE
6	Nucleus-electron potential energy	NEPE
7	Total potential energy	TPE
8	Total kinetic energy	TKE
9	Virial ratio(V/T)	V
10	Dipol moment	DM
11	Polarizability	Р
12	Highest occupied molecular orbital	HOMO
13	Lowest unoccupied molecular orbital	LUMO
14	Total energy	TE

Table 2. List of quantum chemical descriptors calculated by Gamess software



Figure 1. The optimized structure of AIP



**Figure 2.** The optimized structure of  $Cu^{2+}$ -AIP

### 2.3.3. Dragon descriptors

The molecular descriptor calculation software, DRAGON, is recognized for its advanced capabilities. It encompasses a wide range of descriptor types, including constitutional, topological, autocorrelation, geometrical descriptors, as well as 3D-MORSE, RDF descriptors, WHIM, GETAWAY descriptors, functional groups, and atom-centered fragments. In this study, version 3 of the Dragon package [31] was employed to compute these descriptors. The output files generated by the Gamess software for each compound were utilized as input for the Dragon program to calculate the descriptors. Following the calculation, descriptors that exhibited constant or nearly constant values across all molecules were identified and subsequently eliminated. Furthermore, the generated descriptors were used to construct a matrix, with the number of rows corresponding to the number of molecules and the number of columns representing the descriptors. To mitigate redundancy within the matrix, the correlations among descriptors as well as their correlations with the selectivity coefficient were examined. In cases where the correlation coefficient (R) exceeded 0.90 between descriptor pairs, the descriptor with the higher correlation with the selectivity coefficient  $(K_s)$  was retained, while the other descriptor was removed from the matrix. Following these steps, a total of 370 descriptors were retained, while the remaining descriptors were eliminated.

#### 2.4. MLR analysis

The obtained structural descriptors and experimental property values were subjected to analysis using GA-MLR (Genetic Algorithm-Multiple Linear Regression) and stepwise-MLR (Stepwise Multiple Linear Regression) techniques. To address the issue of information overlap among descriptors and to minimize the number of descriptors required in the regression equation, the concept of non-redundant descriptors was employed in this study. While GA-MLR serves as a black-box model, stepwise multiple linear regression was also employed to select the most relevant descriptors and construct the MLR model. By combining these approaches, we aimed to identify the key structural features that are most strongly correlated with the logarithm selectivity coefficient values of the compounds. Through the utilization of the MLR method, we obtained an equation that establishes a relationship between the structural characteristics and the logarithm selectivity coefficient values of the compound.

$$\operatorname{Log} k_{\operatorname{sel}} = a_0 + a_1 x_1 + \dots + a_n x_n \tag{1}$$

The calculations, including GA-MLR, were conducted using MATLAB software (Version 7.0, MathWorks, Inc.) as part of this research endeavor.

## **3. RESULTS AND DISCUSSION**

This study utilized a dataset comprising 30 cations, as presented in Table 1. Quantum chemical calculations were performed to determine various quantum chemical parameters, including geometrical parameters, atomic charges, and dipole moments of Mn<sup>+</sup>-AIP, as outlined in Table 2. To identify the most significant descriptors, a combination of the genetic algorithm and stepwise multiple regression techniques was employed. The QSPR (Quantitative Structure-Property Relationship) models, which establish relationships between the selectivity coefficients as dependent variables and the quantum chemical descriptors as independent variables, are depicted in Table 3.

To assess the individual contributions of each descriptor to the selectivity coefficient, the mean effect (ME) percentage for each parameter was determined using the following calculation:

$$ME_{j}^{J}\% = \frac{\sum_{i=1}^{n} D_{ij} \times a_{j}}{\sum_{j=1}^{J} |\sum_{i=1}^{n} D_{ij} \times a_{j}|} \times 100$$
(2)

 Table 3. Selected MLR models based on quantum chemical descriptors

	Descriptor	Notation	Coefficient	Mean effect (%)	t-Test
	Mulliken charge of cation	MCC	1.734	18.78	8.006
	Mulliken charge of C14	14C	5.807	0.01	3.412
GA-MLR <sup>a</sup>	Mulliken charge of H34	34H	15.317	42.25	6.976
	Mulliken charge of H35	35H	14.837	38.96	3.203
	constant		-16.551		-8.933
	Mulliken charge of cation	MCC	1.096	17.19	5.096
Stepwise-	Mulliken charge of C13	13C	-1.299	13.27	-2.065
MLR <sup>b</sup>	Mulliken charge of H34	34H	17.049	69.54	6.367
	constant		-11.949		-12.931

<sup>a</sup> Statistics of the model: n = 24,  $R^2 = 0.832$ ,  $Q^2 = 0.751$ , S.E. = 0.442 and F = 23.63

<sup>b</sup> Statistics of the model: n = 24,  $R^2 = 0.816$ ,  $Q^2 = 0.726$ , S.E. = 0.507 and F = 29.63

By referring to Equation 2, it becomes possible to acquire the ME (Mean Effect) value of a specific descriptor and comprehend its impact on the selectivity coefficients ( $k_{sel}$  values), whether it is proportional or inversely proportional, without taking into account the effects of other descriptors included in the model. The ME% values of the descriptors can be found in Table 3.

Based on the findings presented in Table 3, the model includes four descriptors: Mulliken charges of cation (MCC), C14, H34, and H35. Mulliken charges are derived from Mulliken population analysis, which offers a method for estimating partial atomic charges through computational chemistry approaches, particularly those based on the linear combination of atomic orbitals molecular orbital method.

The proposed model indicates a direct relationship between the selectivity coefficient and all of the descriptors. Notably, H34 and H35 exhibit the highest mean effect values of +42.25 and +38.96, respectively, making them the most influential factors on the selectivity coefficients. The positive sign denotes a direct effect on the expected response. An increase in the values of H34 and H35 corresponds to an increase in interference. Since these descriptors represent the Mulliken charges of H34 and H35, higher Mulliken charges for H34 and H35 result in a greater interfering effect.

Furthermore, as the interference of a cation increases, there is a corresponding increase in the interaction between the cation and AIP, as well as in the partial charge values of N21, H34, and H35.

The MCC descriptor exhibits a positive impact on the anticipated response. The favorable influence of Mulliken charges of cations on the selectivity coefficient underscores the notion that higher Mulliken charges of interfering cations lead to increased interactions between the cation and AIP, consequently intensifying the interfering effect on the electrode response. As for the 14C descriptor (ME% = 0.01), its effect on selectivity coefficients is minimal. Nevertheless, incorporating this descriptor in the model enhances the reliability of the predictions.

Furthermore, Table 3 reveals that the stepwise-MLR model incorporates three descriptors. Specifically, the Mulliken charge of C13 exhibits a positive mean effect of +13.27 on the expected response. An increase in the partial charge of C13 leads to enhanced interactions between C13 and N12, subsequently influencing the interaction between the cation and AIP, and ultimately increasing interference caused by the cation.

Comparing the squared regression coefficients ( $R^2$ ) and standard error values, the GA-MLR model demonstrates superior descriptive capability with  $R^2$  and standard errors of 0.832 and 0.442, respectively. In contrast, the stepwise-MLR model yields  $R^2$  and standard errors of 0.816 and 0.507, respectively. Consequently, the GA-MLR model provides a more comprehensive description of the relationship between the selectivity coefficients and the molecular structure.

To construct a QSPR equation with enhanced predictive capability, a combination of quantum chemical and dragon descriptors was employed. Initially, the descriptors selected by the genetic algorithm (GA) were utilized to develop a predictive multiple linear regression (MLR) model. The resulting linear model consisted of four variables, and the standard error (S.E.) was calculated for all molecules included in the study.

The analysis revealed that the praseodymium cation  $(Pr^{3+})$  stood out as an outlier, with a residual value exceeding three times the standard error (1.13 > 1.00). This discrepancy could potentially be attributed to experimental errors. Consequently, this particular cation was excluded from the original dataset. Subsequently, an equation linking the selectivity coefficients and the molecular descriptors was derived as follows:

Log 
$$k_{sel} = -3.224 + (-0.253) \text{ RDF070u} + (9.201) \text{ Mor26m}$$
 (3)  
+ (-4.944) Mor13v + (-51.606) H7v

In order to prediction of the selectivity coefficients of the lanthanum-selective electrode, two distinct approaches were utilized: the genetic algorithm (GA) and stepwise regression methods. Table 4 showcases the resultant QSPR models derived from these methods. Notably, the GA-MLR model exhibited significant improvements compared to the stepwise-MLR model in terms of the squared regression coefficients ( $R^2$ ) and standard errors. The R2 value increased from 0.880 to 0.945, while the standard error decreased from 0.400 to 0.258.

	Descriptor	Notation	Coefficient	t-Test
	Radial Distribution Function-7.0	RDF070u	-0.253	-6.126
	3D-Morse-signal 26/ weighted by atomic masses	Mor26m	9.201	15.384
GA-MLR <sup>a</sup>	3D-Morse-signal 13/ weighted by atomic van der waals volumes	Mor13v	-4.944	-6.651
	H autocorrelation of lag 7	H7v	-51.606	-9.405
	constant		-3.224	-3.147
	Mulliken charge of cation	MCC	1.445	9.361
	Mulliken charge of H34	34H	23.307	7.170
Stepwise- MI R <sup>b</sup>	R maximal autocorrelation of lag 2/ weighted by atomic polarizabilities	R2p+	48.172	-3.354
MER	3D-Morse-signal 32/ weighted by atomic Sanderson electronegativities	Mor32e	-3.701	3.223
	constant		-16.485	-10.203

Table 4. Selected predictive models of multiple linear regression

<sup>a</sup> Statistics of the model: n = 23,  $R^2 = 0.945$ ,  $Q^2 = 0.914$ , S.E = 0.258 and F = 78.3

<sup>b</sup> Statistics of the model: n = 24,  $R^2 = 0.880$ ,  $Q^2 = 0.772$ , S.E = 0.400 and F = 45.98

The comparative analysis between the GA-MLR and stepwise-MLR methodologies yielded comprehensive insights into their respective performances, clearly establishing the superiority of the GA-MLR model. Consequently, the GA-MLR model was selected as the preferred predictive model for this study. As shown in Table 4, the GA-MLR model incorporated four dragon descriptors. Notably, one of the 3D descriptors incorporated in the GA-MLR model was RDF070u, belonging to the category of Radial Distribution Function (RDF) descriptors. The RDF descriptor represents the probability distribution of locating an atom within a spherical volume of radius R in an ensemble of A atoms. With the aid of simple rule sets, the RDF descriptor can be interpreted, enabling the conversion of the code into its corresponding 3D structure. In addition to providing information on interatomic distances throughout the molecule, the RDF code offers valuable insights into bond distances, ring types, planar and non-planar systems, and atom types. This attribute is particularly advantageous for computer-assisted code elucidation [32]. Notably, RDF070u exhibits a negative sign, indicating an inverse relationship between the selectivity coefficient and this specific descriptor.

Two additional descriptors, Mor26m and Mor13v, are included in the model, both of which belong to the 3D-MoRSE descriptors. The 3D-MoRSE descriptors are derived from the simulation of Infrared spectra using a generalized scattering function [32]. Mor26m represents signal 26 weighted by atomic masses, while Mor13v represents signal 13 weighted by atomic van der Waals volumes.

Mor26m exhibits a positive sign, indicating a direct relationship between the selectivity coefficient and this descriptor. Consequently, an increase in atomic mass corresponds to an increase in selectivity coefficient values. On the other hand, Mor13v displays a negative sign, indicating an inverse relationship. Hence, an increase in the van der Waals volume of the molecules results in a decrease in selectivity coefficient values.

	RDF070u	Mor26m	Mor13v	H7v
RDF070u	1			
Mor26m	0.442	1		
Mor13v	0.088	0.337	1	
H7v	-0.282	0.210	0.436	1

Table 5. Correlation matrix of the selected descriptors

Furthermore, H7v is classified as an Autocorrelation descriptor. These descriptors are calculated based on the 3D-spatial molecular geometry, utilizing interatomic distances collected in the geometry matrix and a property function defined by a set of atomic properties

[32]. The negative sign associated with H7v signifies an inverse relationship between the selectivity coefficient and this descriptor.

The correlation matrix, as presented in Table 5, illustrates the relationships between the selected descriptors. It is evident from the table that the linear correlation coefficients between each pair of descriptors are below 0.442. This indicates that there are no significant correlations observed among the selected descriptors.

To validate the reliability and efficacy of the predictive model, several techniques were employed, including bootstrap analysis, y-scrambling, and external validation [32,31]. The bootstrapping process was repeated 3000 times, while y-scrambling was performed 300 times. The statistical parameters of the predictive model are reported as follows:

$$R^{2} = 0.945; Q^{2}_{LOO} = 0.914; Q^{2}_{BOOT} = 0.890; Q^{2}_{EXT} = 0.862;$$
  
s = 0.258; a =0.091; F = 78.3 (4)

The disparities observed among  $Q^2_{LOO}$ ,  $Q^2_{BOOT}$ ,  $Q^2_{EXT}$ , and  $R^2$  affirm the validity and strong predictive capability of the obtained model. Furthermore, the intercept value of the y-scrambling analysis (a=0.091) indicates that chance correlation had minimal or negligible influence on the development of the model [33].

**Table 6.** presents a comprehensive comparison of the standard error of prediction (SEP) and the standard error of testing (SET) for the selected predictive model, alongside the test models derived using various molecules as the prediction set

Model	SET	SEP	Molecules in the prediction set <sup>a</sup>
Selected model	0.228	0.286	9, 10, 12, 22, 26 and 28
Test model I	0.247	0.329	1, 5, 15, 18, 20 and 30
Test model II	0.275	0.366	4, 6, 7, 16, 21 and 27
Test model III	0.281	0.367	3, 8, 13, 17, 24 and 29
Test model IV	0.239	0.306	2, 11, 19, 23 and 25

<sup>a</sup> Numbers refer to the number of the compounds given in Table 1. The remaining cations for each set are due to the corresponding training set.

Considering the small size of the dataset, different prediction and training sets were utilized to train the multiple linear regression (MLR) model. For each iteration, a random selection of six cations out of the 29 cations was designated as the prediction set, while the remaining cations were employed to develop the MLR model. Subsequently, the selectivity coefficients of the six cations in the prediction set were predicted using the GA-MLR model. This process was repeated three times, and the standard error of training (SET) and standard error of

prediction (SEP) are presented in Table 6. Notably, the results obtained demonstrate independence from the specific cations used in the prediction set.

The correlation between the experimental and predicted values of the selectivity coefficients is excellently depicted in Figure 3, where a plot of the calculated values against the experimental values is shown.



Figure 3. Plot of the calculated selectivity coefficients against the experimental values

Figure 4 displays the residuals of the MLR-predicted values of the selectivity coefficients plotted against the corresponding experimental values. The distribution of the calculated residuals on both sides of the zero line indicates the absence of any systematic errors in the development of the MLR model.



Figure 4. Plot of the residuals vs. experimental selectivity coefficients

# **4. CONCLUSION**

The GA-MLR modeling approach was employed to predict the selectivity coefficient values of a La(III) membrane sensor. The selectivity coefficient is a crucial parameter in electrochemical sensors, and its measurement is often time-consuming and expensive. In this

study, we propose a novel chemometrics-based method for predicting selectivity coefficients of cation sensors, specifically focusing on lanthanides sensors. This research is particularly intriguing due to the limited availability of cation descriptors and the utilization of a restricted basis set for lanthanides calculations. Consequently, QSPR investigations on selectivity coefficients of cation sensors, especially those targeting lanthanides, have garnered significant interest.

The QSPR model efficiently estimated and predicted the selectivity coefficients of the Lanthanum-selective electrode based on the AIP (Atom Invariant Polarizability) approach. Initially, the MLR model with quantum chemical descriptors was constructed, benefiting from their simplicity and significance in characterizing chemical properties. The resulting MLR model demonstrated a good descriptive ability.

To design a QSPR equation capable of predicting the selectivity coefficients of the Lanthanumselective electrode, both quantum chemical descriptors and dragon descriptors were employed. Given that GA serves as a "black box" variable selection method, the stepwise regression technique was also utilized in developing the MLR model, and its results were compared to those of the GA-MLR model. The findings clearly indicate the superior performance of the GA-MLR model over the stepwise-MLR model.

## **Declarations of interest**

The authors declare no conflict of interest in this reported work.

#### REFERENCES

- [1] N. Adil, R. Wang, and X. Wang, Anal. Chem. 95 (2023) 29.
- [2] F. Faridbod, M. R. Ganjali, M. Pirali-Hamedani, and P. Norouzi, Int. J. Electrochem. Sci. 5 (2010) 8.
- [3] M. R. Ganjali, L. Naji, T. Poursaberi, M. Shamsipur, and S. Haghgoo, Anal. Chim. Acta 475 (2003) 1-2.
- [4] M. R. Ganjali, P. Norouzi, M. Adib, and A. Ahmadalinezhad, Anal. lett. 39 (2006) 6.
- [5] M. R. Ganjali, P. Norouzi, F. S. Mirnaghi, S. Riahi, and F. Faridbod, IEEE Sensors Journal 7 (2007) 8.
- [6] M. R. Ganjali, J. Ravanshad, M. Hosseini, M. Salavati-Niasari, M.R. Pourjavid, and M. R. Baezzat, Electroanalysis 16 (2004) 21.
- [7] M. R. Ganjali, M. Tavakoli, F. Faridbod, S. Riahi, P. Norouzi, and M. Salavati-Niassari, Int. J. Electrochem. Sci. 3 (2008) 12.
- [8] H. A. Zamani, B. Feizyzadeh, F. Faridbod, and M. R. Ganjali, Sensor Lett. 9 (2011) 5.
- [9] W. Gao, E. Zdrachek, X. Xie, and E. Bakker, Angew. Chem. Int. Ed. 59 (2020) 6.
- [10] H. A. Zamani, R. Kamjoo, M. Mohammadhosseini, M. Zaferoni, Z. Rafati, M. R. Ganjali, F. Faridbod, and S. Meghdadi, Mater. Sci. Eng. C 32 (2012) 3.

- [11] H. A. Zamani, M. Mohammadhosseini, S. Haji-Mohammadrezazadeh, F. Faridbod, M. R. Ganjali, S. Meghdadi, and A. Davoodnia, Mater. Sci. Eng. C 32 (2012) 4.
- [12] H. A. Zamani, M. Rohani, M. Mohammadhosseini, M. R. Ganjali, F. Faridbod, and S. Meghdadi, Sensor Lett. 9 (2011) 5.
- [13] J. Bloxham, D. Hill, N. F. Giles, T. A. Knotts IV, and W. V. Wilding, Mol. Inform. 41 (2022) 8.
- [14] T. N. Brown, J. M. Armitage, and J. A. Arnot, Mol. Inform. 38 (2019) 8-9.
- [15] L. H. Nguyen, and T. N. Truong, ACS omega 3 (2018) 8.
- [16] T. M. Nolte, and A. M. Ragas, Environ. Sci. Process. Impacts. 19 (2017) 3.
- [17] S. Riahi, A. Beheshti, M. R. Ganjali, and P. Norouzi, Electrophoresis 29 (2008) 19.
- [18] J. F. Stults, Y. J. Choi, C. Rockwell, C. E. Schaefer, D. D. Nguyen, D. R. Knappe, T. H. Illangasekare, and C. P. Higgins, Environ. Sci. Technol. 57 (2023) 13.
- [19] M. Karelson, V. S. Lobanov, and A. R. Katritzky, Chem. Rev. 96 (1996) 3.
- [20] J. Wan, L. Zhang, G. Yang, and C.G. Zhan, J. Chem. Inf. Comput. Sci. 44 (2004) 6.
- [21] Y. Wang, and X. Chen, RSC advances 10 (2020) 70.
- [22] J. Tammiku-Taul, Quantum chemical study of the properties of Grignard reagents. Citeseer, Estonia (2003).
- [23] G. W. Richings, and S. Habershon, J. Chem. Theory Comput. 13 (2017) 9.
- [24] T. Van Mourik, and R. J. Gdanitz, J. Chem. Phys. 116 (2002) 22.
- [25] S. Ahmadi, and S. Ganji, Curr. Drug Disc. Technol. 13 (2016) 4.
- [26] S. Ahmadi, and E. Habibpour, Anticancer. Agents. Med. Chem. 17 (2017) 4.
- [27] S. Riahi, M.R. Ganjali, E. Pourbasheer, and P. Norouzi, Chromatographia 67 (2008)
- [28] E. Habibpour, and S. Ahmadi, Curr. Comput. Aided Drug Des. 13 (2017) 2.
- [29] M. Ganjali, P. Norouzi, T. Alizadeh and M. Adib, Anal. Chim. Acta 576 (2006) 2.
- [30] M. W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, and S. Su, J. Comput. Chem. 14 (1993) 11.
- [31] R. Todeschini, V. Consonni, A. Mauri, and M. Pavan, Anal. Chim. Acta 515 (2004) 199.
- [32] R. Todeschini, and V. Consonni, Handbook of molecular descriptors. John Wiley & Sons (2008).
- [33] J. G. Topliss, and R. J. Costello, J. Med. Chem. 15 (1972) 10.