

# Predictions of Retention Factors for Some Organic Nucleophiles in Complexation Gas Chromatography

Mohammad Hossein Fatemi\* and Zahra Ghorbannezhad

Laboratory of Chemometrics, Faculty of Chemistry, University of Mazandaran, Babolsar

## Abstract

In the present work, the quantitative structure–retention relationship (QSRR) was used to predict the gas chromatographic retention factors of some organic nucleophile on chemically modified stationary phase by complexes of Cu (II) with amino groups. The gravitation index, relative negative charge surface area, C component of moment of inertia and weighted negative charged partial surface area are selected as the most relevant descriptors from the pool of descriptors. These descriptors were used for developing multiple linear regression (MLR) and artificial neural network (ANN) models as linear and nonlinear feature mapping techniques. The root mean square errors (RMES) in calculation of retention factors for training, internal and external test set are 0.242, 0.295, and 0.240, respectively for MLR model, and for ANN model the RMSE for training, internal and external test set are; 0.084, 0.108, and 0.176. The ANN and MLR model were further examined by cross validation test, which obtained statistics of  $Q^2 = 0.82$  and  $SPRESS = 0.22$  for MLR model and  $Q^2 = 0.97$ ,  $SPRESS = 0.07$  for ANN model. Comparison between these results and other statistics of ANN and MLR models revealed the superiority of ANN over MLR model.

## Introduction

Chemically bonded phases comprising amino groups bound to the support surface through an alkyl chain show electron-donor–acceptor properties and, due to this fact, they can be permanently bonded with the metal cations and provide a good way to obtain selective stationary phases for complexation gas chromatography (CGC). Metals of “d” and “f” blocks can easily enter selective interactions with the analyzed nucleophilic compounds, thus forming reversible complexes of different stability. The obtained chemically bonded chelates, which are selective sorbents, have been frequently used to separate different organic compounds (e.g., hydrocarbons, alcohols, and amines) (1,2). Khuhawar et al. studied chelate complexes of Ni (II) with Schiff’s bases as stationary phase to separate aromatic hydrocarbons, heteroaromatic aldehydes, ketones, amines, and alcohols (3). Transition metals easily enter into specific interactions with elec-

tron-donor compounds forming p-complexes. Also separation of higher olefins (C5 and higher) and their isomers have been carried on transition metal complexes of bonded silicas with the carrier groups –CN, –SH, –NH, and PPh (4,5). Some factors which influence the stability of these complexes are the degree of oxidation, electronic structure, radii of the central cation and ligand, ligand basicity. These factors allow, on one hand, a control of retention to obtain the selections required and, on the other hand they permit the physical and chemical investigation of the packing and retention mechanism in complexation gas chromatography. Because empirical determination of solute retention in CGC may be expensive, time consuming and need to pure standard compounds, therefore the development of a theoretical model for estimating the solute retention seems to be interesting and necessary. Moreover the results of these studies can be used to better understand the mechanism of retention in complexation gas chromatography. One approach to calculate retention parameters from molecular structural descriptors is quantitative structure retention relationship (QSRR) methods. In QSRR approaches, the structural features of solute encoded by non-empirical numerical descriptors, and the relation among these descriptors and solute retention was examined. QSRRs have been extensively used to obtain simple models to explain separation mechanisms and predict the retention properties of various classes of organic compounds in analytical chemistry (6). For example B. Ren investigated the quantitative correlation between the Kovats retention indices of 33 aldehydes and ketones on four stationary phases (HP-1, HP-50, DB-210, and HP-Innowax) at 50°C using four separate 4–5 variables linear equations (7). Also O. Farkas and K. Heberger reported a linear model for prediction of retention indices of some aliphatic alcohols (8). They used some variable selection methods such as ridge regressions, partial least square, pair-correlation method, forward selection, and best subset selection methods. In the other report Jurs and his group correlated, the observed Kovats retention indices of sulfur vesicants by multiple linear regression techniques (MLR) by using 9 descriptors in their models on different stationary phases (9). The standard errors of the mean for their models were ranged between 2.1–2.5. Fatemi et al. have been used QSRR techniques for prediction of retention factors in supercritical fluid chromatography (10). They used a number of single bonds, a number of double bounds, and hydrophilic factors as molecular descriptors to developed some MLR and artifi-

\*Author to whom correspondence should be addressed: email mhfatemi@umz.ac.ir.

**Table I. Data Set and Corresponding Observed, MLR and ANN Calculated Values of Retention Factors**

No.	Name	Log (K) Exp	Log (K) MLR	Deviation	Log (K) ANN	Deviation
1	Pentane	0.590	0.693	0.103	0.565	-0.025
2	1-Pentene	0.710 <sub>int</sub>	0.806	0.097	0.541	-0.168
3	1-Pentyne	1.060	1.069	0.009	1.056	-0.003
4	<i>cis</i> -2-Penten	0.420	0.740	0.320	0.345	-0.075
5	<i>trans</i> -2-Pentene	0.404	0.784	0.380	0.498	0.095
6	Hexane	0.920	0.999	0.079	1.091	0.172
7	1-Hexene	1.030 <sub>ext</sub>	1.195	0.165	1.156	0.126
8	1-Hexyne	1.430 <sub>ext</sub>	1.461	0.031	1.551	0.121
9	<i>cis</i> -2-Hexene	0.722	1.138	0.416	0.835	0.114
10	<i>trans</i> -2-Hexene	0.700	1.120	0.420	0.839	0.139
11	Heptane	1.250	1.380	0.130	1.210	-0.040
12	1-Heptene	1.370	1.521	0.151	1.235	-0.135
13	1-Heptyne	1.840	1.790	-0.050	1.809	-0.031
14	<i>cis</i> -2-Heptene	1.010	1.497	0.487	1.061	0.051
15	<i>trans</i> -2-Heptene	0.990 <sub>int</sub>	1.487	0.497	1.024	0.035
16	<i>cis</i> -3-Htene	0.990	1.523	0.533	1.196	0.207
17	<i>trans</i> -3-Heptene	0.970	1.472	0.502	0.902	-0.068
18	Octane	1.580	1.674	0.094	1.594	0.015
19	1-Octene	1.670	1.882	0.212	1.697	0.027
20	1-Octyne	2.220	2.175	-0.045	2.154	-0.066
21	Nonane	1.870 <sub>int</sub>	2.042	0.172	1.954	0.085
22	1-Nonene	1.920	2.201	0.281	2.051	0.131
23	1-Nonyne	2.270	2.510	0.240	2.295	0.026
24	Decane	2.180	2.329	0.149	2.062	-0.117
25	1-Decene	2.210	2.575	0.365	2.307	0.098
26	1-Decyne	2.390 <sub>int</sub>	2.884	0.494	2.315	-0.075
27	2,2-Dimethylbutane	0.620	0.916	0.296	0.526	-0.094
28	2,3-Dimethylbutane	0.690	0.975	0.285	0.883	0.193
29	2,3-Dimethyl-1-butene	0.711	1.082	0.371	0.706	-0.005
30	2,3-Dimethyl-2-butene	0.810 <sub>ext</sub>	1.083	0.273	1.032	0.222
31	3,3-Dimethyl-1-butene	0.580 <sub>ext</sub>	0.982	0.402	0.555	-0.025
32	2-Methylpentane	1.017	0.975	-0.042	1.003	-0.014
33	3-Methylpentane	1.020	0.995	-0.025	0.837	-0.183
34	2-Methyl-1-pentene	1.140	1.148	0.008	1.089	-0.050
35	3-Methyl-1-pentene	1.070	1.125	0.055	1.037	-0.033
36	4-Methyl-1-pentene	1.090	1.145	0.055	1.065	-0.024
37	2-Methyl-2-pentene	1.150	1.111	-0.039	0.893	-0.256
38	<i>cis</i> -3-Methyl-2-pentene	1.180	1.119	-0.061	1.142	-0.038
39	<i>trans</i> -3-Methyl-2-pentene	1.170	1.100	-0.070	1.113	-0.057
40	<i>cis</i> -4-Methyl-2-pentene	1.110	1.133	0.023	1.036	-0.074
41	<i>trans</i> -4-Methyl-2-pentene	1.170 <sub>ext</sub>	1.076	-0.094	1.501	0.331
42	2,2,4-Trimethylpentane	1.500	1.706	0.206	1.496	-0.003
43	2,4,4-Trimethyl-1-pentene	1.620	1.819	0.199	1.604	-0.016
44	2,4,4-Trimethyl-2-pentene	1.590 <sub>int</sub>	1.754	0.165	1.538	-0.051
45	Cyclopentane	0.470	0.645	0.175	0.544	0.074
46	Cyclopentene	0.570	0.778	0.208	0.577	0.008

cial neural network models. Also Fatemi et al. have used quantitative structure migration relationship to relate the migration factor of some benzene derivatives to their molecular structural features (11). Other reports about QSRR prediction of retention parameters can be finding in references (12–14). Our literature survey reveals that there is not any report about QSRR prediction of gas chromatographic retention factor of solute on chemically modified silica stationary phase with complexes of Cu (II) with amino groups (complexation gas chromatography). Therefore a multiple linear regression analysis and artificial neural network was employed to develop some QSRR models to correlate the molecular descriptors and observed retention factors of some nucleophile organic compounds. The results of these models not only can be used in prediction of retention factors of other compound but also investigates the forces influences on solute retention in complexation gas chromatography.

**Table I. Data Set and Corresponding Observed, MLR and ANN Calculated Values of Retention Factors**

No.	Name	Log (K) Exp	Log (K) MLR	Deviation	Log (K) ANN	Deviation
47	Cyclohexane	0.803	0.989	0.186	0.762	-0.041
48	Cyclohexene	0.950	1.130	0.180	0.896	-0.054
49	Cycloheptane	1.190	1.343	0.153	1.160	-0.030
50	Cycloheptene	1.330	1.459	0.129	1.379	0.049
51	Cyclooctane	1.560	1.690	0.130	1.509	-0.051
52	Cyclooctene	1.670	1.839	0.169	1.635	-0.034
53	1,3-Cyclohexadiene	1.189	1.322	0.133	1.215	0.027
54	1,4-Cyclohexadiene	1.154	1.309	0.155	1.198	0.045
55	Methylcyclopentane	0.800	1.089	0.289	0.817	0.018
56	1-Methyl-1-cyclopentene	0.940	1.209	0.269	1.084	0.144
57	Methylcyclohexane	1.130 <sub>int</sub>	1.371	0.242	1.146	0.016
58	1-Methyl-1-cyclohexene	1.311	1.552	0.241	1.320	0.010
59	4-Methyl-1-cyclohexene	1.290 <sub>int</sub>	1.555	0.265	1.502	0.212
60	Ethylcyclohexane	1.460	1.785	0.325	1.509	0.049
61	1,3,5-Cycloheptatriene	1.640	1.809	0.169	1.708	0.068
62	1,3-Cyclooctadiene	1.705 <sub>ext</sub>	1.904	0.199	1.943	0.238
63	1,5-Cyclooctadiene	2.023	1.974	-0.049	1.934	-0.089
64	Cyclooctatetraene	1.860	2.163	0.303	2.023	0.164
65	Benzene	1.420	1.453	0.033	1.360	-0.060
66	Toluene	1.805	2.014	0.209	1.750	-0.054
67	Ethylbenzene	2.091	2.391	0.300	2.128	0.037
68	<i>o</i> -Xylene	2.160	2.312	0.152	2.122	-0.037
69	<i>m</i> -Xylene	2.190 <sub>int</sub>	2.287	0.098	2.121	-0.069
70	<i>p</i> -Xylene	2.150	2.352	0.202	2.129	-0.021
71	Styrene	2.257	2.595	0.338	2.187	-0.070
72	Cumene	2.276 <sub>ext</sub>	2.616	0.340	2.316	0.040
73	Propylbenzene	2.329	2.762	0.433	2.303	-0.026
74	Diethyl ether	1.725	1.735	0.010	1.702	-0.022
75	Butyl methyl ether	1.980	2.276	0.296	2.003	0.024
76	tert.-Butyl ethyl ether	2.000	2.234	0.234	2.0308	0.031
77	Dipropyl ether	2.080 <sub>ext</sub>	2.269	0.189	2.121	0.041
78	Dibutyl ether	2.520	2.848	0.328	2.423	-0.097

\* In this table, int and ext denote internal and external test set.

## Experimental

### Data set

The data set consists of retention factors for a collection of 78 aliphatic, branched and aromatic hydrocarbons and ethers, which are shown in Table I and were taken from values reported by W. Wasiak, and W. Urbaniak (15). The retention factors are varies in the range of 0.404–2.52 for *trans*-2-Pentane and dibutyl ether, respectively, in logarithmic value. All chromatographic measurements were carried out on a GCHF 18.3 gas chromatograph (Chromatron, Berlin, Germany), equipped with a flame ionization detector (FID) by using Stainless-steel columns of 1 m length and 0.3 cm i.d. silica gel (kieselgel 60) from Macherey, Nagel, and Co. (Duren, Germany). The silica stationary phase was chemically modified with the complex of Cu (II) and N-(2-aminoethyl)-3-aminopropyltrimethoxysilane. The reported application of chemically bonded complexes to gas chromatography shows a high selectivity of this packing to the wide range of organic compounds having nucleophilic properties (16–20). The data set were sorted according to their retention factor and the training, internal and external test sets were chosen from this list with desired distance from each other (*y*-ranking procedure). The training set has 62 members, which were used to adjust the parameters of model and model constructions; the internal test set (with 8 members) was used to prevent the model overfitting and external test set (with 8 members) was used to evaluate the prediction power of constructed models.

### Molecular descriptors

Retention in GC is the result of competitive distribution of the solute between the mobile and stationary phase. The molecular structure and chemical properties of the solute and stationary phase determine the type and extent of the interactions between the solute and stationary phase. The forces associated with these interactions may be related to the geometric, topological, and electric characteristics of the solute. In QSRR studies, the calculation and selection of molecular descriptors are very important steps that affected the quality of the obtained models. Descriptors are generated solely from the molecular structures and aimed to encode numerically significant features of each molecule. Various structural features of the molecules are used as the descriptors.

In the first step of calculation of molecular descriptors the structures of compounds were drawn with Hyperchem [Hyper (2002) release 7.0 for Windows, Hypercube, Gainesville, FL] and exported to the file with suitable format for MOPAC package (version 6.0) (21), on the basis of the minimum energy molecular geometries optimized by AM1 semiempirical method (22).

The Hyperchem and MOPAC output files were used by the CODESSA program (23–25) to calculate five classes of descriptors, which are; constitutional, topological, electrostatic, geometrical, and quantum chemical. Then variables, which have the constant and/or near-constant values were excluded from pool of descriptors. Also the pair correlation cutoff selection method was employed on the remaining descriptors. In this method for each pair of variables with a correlation value greater than cutoff value ( $R > 0.90$ ), one of the two correlated variables is eliminated. Subsequently, the method of stepwise MLR was performed on

the training set to select the most relevant descriptors. These descriptors were used as independent variable for developing MLR and ANN models.

### Nonlinear modeling

Besides the standard multilinear QSRR approach, an artificial neural network (ANN) was also applied to further examine nonlinear relation between selected molecular descriptors and solute retention factor. The used network implements the feed-forward back propagation algorithm (26,27) as one of the most complex and powerful realizations of the neural network. ANNs have been used for investigation of a wide variety of chemical problems such as, prediction of dielectric constants (28), selectivity coefficients of ion selective electrodes (29), mass spectral search (30), and QSPR investigations (31–33).

In the first step of developing ANN model, the program of a feed-forward neural network, which was trained by a back-propagation algorithm was written by MATLAB version 7.4.0. This network had the four nodes in the input layer and one node in the output layer. The initial weights were selected from uniform distribution of numbers that were ranged between –0.3 and 0.3. The initial bias values were set to be one. These values were optimized during the network training. The value of each input was divided into its mean value to bring them into the dynamic range of the sigmoid transfer function of ANN. Before training, the network parameters would be optimized. These parameters were: the number of nodes in the hidden layer, weights and biases learning rates, and the momentum. Then the optimized network was trained using training set for adjustment of weights and biases values. To maintain the predictive power of the network at a desirable level, training was stopped when the value of error for the internal test set started to increase. Because the test error was not a good estimate of the generation error, the prediction potential of the model was evaluated on a third set of data, named the external test set.

### Molecular diversity

The choice of an optimal metric space that represents the structural diversity of a compound population is determinant in the efficiency of the model (34). Molecular diversity analysis explores the way of molecules to cover a determined structural space and underlies many approaches for compound selection and design of combinatorial libraries. In this study, diversity analysis was performed for the data set based on decrypted algorithm by Luan et al. (35). In this way, the mean distances of one sample to the remaining ones were computed from descriptor space matrix as follows:

$$d_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad \text{Eq. 1}$$

$$d_i = \frac{\sum_{k=1}^m d_{ij}}{n-1} \quad i = 1, 2, \dots, n \quad \text{Eq. 2}$$

In the previously described equations,  $X_i$  and  $X_j$  denote the collective database involving the descriptors and  $d_{ij}$  is a distance score for two different compounds. Then the mean distances of samples in descriptors space were plotted versus the corre-

sponding experimental data (Figure 1). As shown in this figure, the structures of the compounds are diverse in both sets. The training set with a broad representation of the chemistry space was adequate to ensure models' stability and the diversity of test set can prove the predictive capability of the model.

## Results and Discussion

In this work, the quantitative relationships between the gas chromatographic retention factors of some organic nucleophiles on stationary phases, which chemically modified by complexes of Cu (II) with amino groups to their molecular descriptors were investigated by using linear and non-linear models. To obtain the optimal model at first, some theoretical molecular descriptors using the CODESSA software for the studied chemicals were calculated: constitutional, geometrical, topological, and electronic descriptors. After feature reduction, the most relevant descriptors were selected by stepwise variable subset selection procedure. These descriptors are gravitation index (GI), relative negative charge surface area (RNCG), C component of moment of inertia (IC), and weighted negative charged partial surface area (WNSA-1) (36). Inter-correlations among these descriptors are presented in Table II. As can be seen in this table, there is not any high correlation between selected descriptors. These descriptors were used as inputs for developing MLR and ANN models.

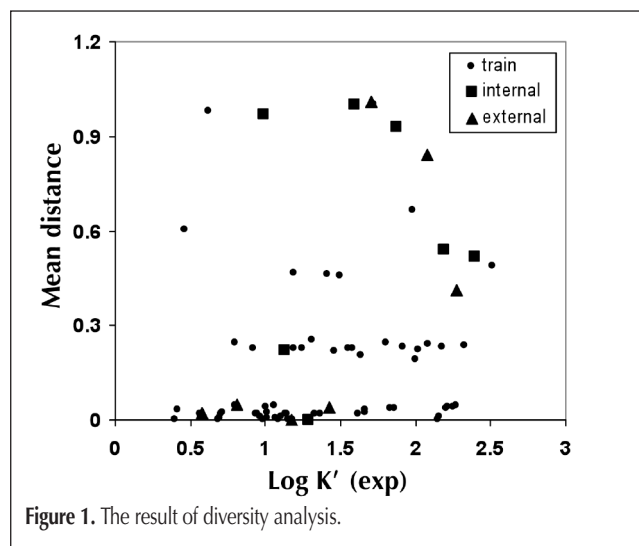


Figure 1. The result of diversity analysis.

Table II. Correlation Matrix Between Selected Descriptors\*

	$G_I$	$I_C$	WNSA-1	RNCG
$G_I$	1.000	-0.369	0.457	-0.170
$I_C$		1.000	-0.262	0.103
WNSA-1			1.000	-0.122
RNCG				1.000

\*  $G_I$  = Gravitation index  
 $I_C$  = C component of moment of inertia  
WNSA-1 = Weighted negative charged partial surface area  
RNCG = Relative negative charge surface area

For MLR, good correlations with the experimental retention factors were selected based on the correlation coefficient (R), Fisher criterion (F), and standard error of the regression (SE) of models. The specification of the selected MLR model was shown in Table III. Also the calculated value of retention factors for training and test sets by this model was shown in Table I. The standard error (SE), average error (AE), and average absolute error (AAE) in calculation of retention factor by this model for training set are SE = 0.244, AE = -0.189, and AAE = 0.201; SE = 0.315, AE = -0.253, and AAE = 0.253 for the internal test set, and SE = 0.257, AE = -0.188, and AAE = 0.211 for the external test set, respectively.

To examine any nonlinear relation among selected molecular descriptors and their retention factors, ANN was used. The previous selected molecular descriptors were used as inputs for developed ANN. Before training the network, the parameters of the number of nodes in the hidden layer, weights and biases learning rates and momentum values were optimized that are 5, 0.2, 0.5, and 0.4, respectively. Procedures for the optimization of these parameters were reported in our previous work (37–39).

This network was then trained by using the training set for optimization of the weights and biases values by back propagation algorithm. In order to prevent overtraining, an internal test set was used to evaluate the prediction power of the network during its training (48). In this method, after each 1000 training iterations, the network was used to calculate retention factors of molecules included in the internal test set. The network training was stopped when the value of errors for the internal test set started to increase. Finally, to estimate the generalization error and prediction potential of trained ANN model, it can be used to predict the retention factors of an independent data, which were named external test set. The ANN predicted values of retention factors for training, internal, and external sets were shown in Table I. The statistics of these calculations are: SE = 0.087, AE = 0.0012 and AAE = 0.066 for the training set; SE = 0.116, AE = 0.001, and AAE = 0.088 for the internal test set; and SE = 0.188, AE = -0.136, AAE = 0.143 for the external test set, respectively.

In order to evaluate the reliability of developed models, cross-validation was employed. The obtained statistical parameters of leave-5-out cross-validation test were  $Q^2 = 0.82$  and SPRESS = 0.22 for MLR and  $Q^2 = 0.97$ , SPRESS = 0.07 for ANN models. Also, a Y-scrambling procedure was performed to investigate any

Table III. Specification of the MLR Model\*

Notation	Coefficient	SE	t-value	p-value	ME
GI	0.004	±0.000	13.837	0.000	1.068
IC	-2.297	±0.992	-2.316	0.005	-0.227
WNSA-1	0.018	±0.003	7.890	0.000	0.123
RNCG	1.455	±0.171	8.496	0.000	0.037
Constant	-1.058	±0.181	-5.854	0.000	-

\*  $G_I$  = Gravitation index  
 $I_C$  = C component of moment of inertia  
WNSA-1 = Weighted negative charged partial surface area  
RNCG = Relative negative charge surface area  
where:  $n = 62$ ,  $R = 0.964$ ,  $F = 488.0$ , and  $SE = 0.1331$

chance correlation in our modeling (40). The mean value of R and SE of these models are 0.254 and 0.392, respectively. According to these results it can be concluded that there was not any chance correlation in our data matrix. The statistical results of ANN and MLR models were shown in Table IV. Comparison among these values revealed the superiority of ANN over MLR model. Figure 2 shows the plot of the ANN calculated versus experimental log K' for the training, internal and external test sets, which represent the good correlation between these values. Also the residuals of the ANN calculated values of the retention factors were plotted against their experimental values in Figure 3. The propagation of residuals in both sides of zero line indicated that no systematic error exists in the development ANN model.

By interpreting the descriptors in QSRR model, it was possible to gain some insight into factors that were likely related to the solute retention in complexation gas chromatography. For inspection of the relative importance and contribution of each descriptor in the model, the value of mean effect ( $ME_j$ ) was calculated for each descriptor from the following equation (41):

$$ME_j = \frac{\beta_j \sum_{i=1}^n D_{ij}}{\sum_j \beta_j \sum_i D_{ij}} \quad \text{Eq. 3}$$

where  $ME_j$  is the mean effect for considered descriptor  $j$ ,  $\beta_j$  is the coefficient of descriptor  $j$ ,  $D_{ij}$  is the value of interested descriptors for each molecule, and  $m$  is the number of descriptors in the model. The value of mean effect revealed the relative importance of a descriptor in comparison with other descriptors in the

model and its sign represented the direction of variation in the values of desired properties or activities resulted by increasing (or decreasing) the values of interested descriptor. The calculated values of  $ME_j$  this equation for all descriptors are shown in the last column of Table II. As can be seen in this table the relative importance order of descriptors is: GI > IC > WNSA-1 > RNCG.

The most important descriptor according to calculated mean effects is the gravitation index (GI), obtained by summation over all bonded atoms from following equation:

$$GO = \sum_{\text{bonded atoms } ij} \frac{m_i m_j}{r_{ij}^2} \quad \text{Eq. 4}$$

where  $m_i$  and  $m_j$  are the atomic masses of the bonded atoms and  $r_{ij}$  is the corresponding bond length. The gravitation index has been used effectively in recent studies to describe intermolecular dispersion forces in chromatographic retention (42,43). The positive sign for the ME (1.068) of this descriptor shows that by increasing in the value of these descriptors, the solute retention was increased. The next descriptor in the model was C component of moments of inertia (-0.227), which was a geometrical descriptor encoding the size and mass distribution in a molecule. The values of this descriptor, the solute retention were decreased due to difficulties in complex formation with stationary phases. Another descriptor is the charged partial surface area (CPSA) type descriptor (WNSA-1), which represent the charge distribution over molecular surface area, and can encode features of molecule, which responsible in polar interactions with stationary phase. The appearance of this descriptor in the model clearly indicates the importance of charge distribution of the solutes on their interactions with Cu (II) complexes of sta-

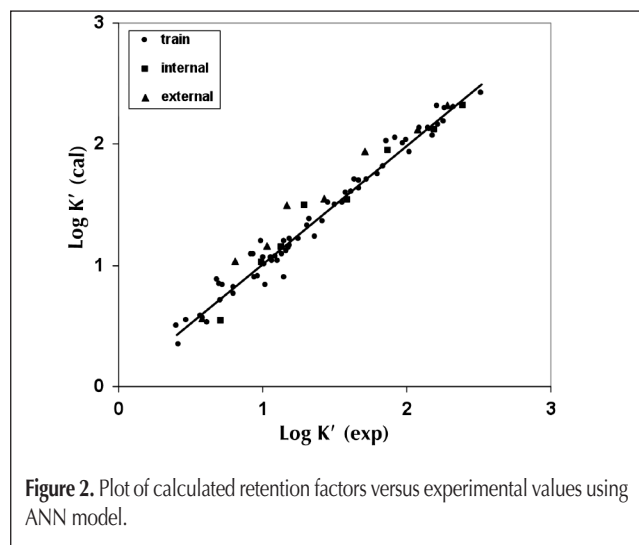


Figure 2. Plot of calculated retention factors versus experimental values using ANN model.

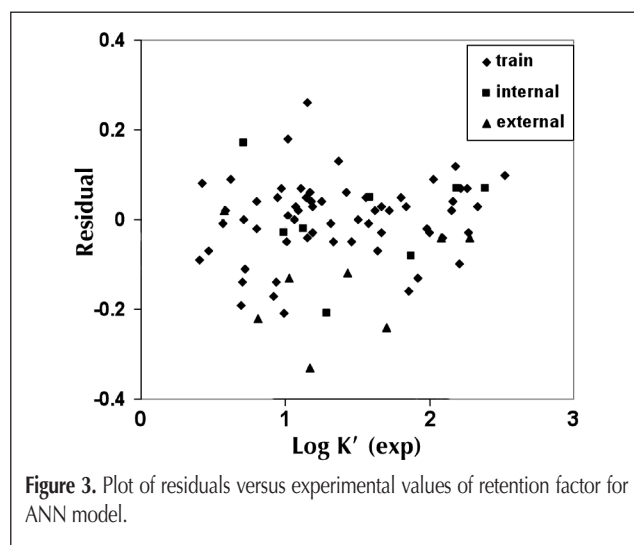


Figure 3. Plot of residuals versus experimental values of retention factor for ANN model.

Table IV. The Statistical Results of ANN and MLR Models

Set	ANN						MLR					
	R	SE	F	RMSE	AE	AAE	R	SE	F	RMSE	AE	AAE
Training set	0.990	0.087	2415	0.084	0.001	0.066	0.964	0.150	787	0.242	-0.189	0.201
Internal test set	0.981	0.124	154	0.108	0.001	0.088	0.964	0.162	87	0.295	-0.253	0.253
External test set	0.980	0.129	145	0.240	-0.188	0.211	0.960	0.169	82	0.240	-0.188	0.211

tionary phases. The positive sign for the ME of this descriptor (0.123) indicates that by increasing the value of partial charge on molecules its interactions with stationary phase increases and therefore solute retention increased due to formation of reversible complex. The last descriptor was RNCG, which is the ratio among the maximum atomic negative charge and the overall negative charge in the molecule. This descriptor can encode the negative partial charge distribution information in the molecule that can account the hydrogen-bonding formation ability of the molecules. The positive sign of the mean effect for this descriptor (0.037) indicate that the tendency of solute to stationary phases increased by increasing of relative negative charge on the molecule.

## Conclusion

In conclusion, the results of this study revealed that QSRR method can be used for prediction of retention factors of organic nucleophile in complexation gas chromatography. The descriptors which were used in this work are topological and electronic, which indicate the effect of steric and electronic interaction on the solute retention in complexation GC. The statistical parameters of ANN model were better than MLR one, which illuminate that there are some non-linear relations between calculated molecular descriptors and solute retention. The results of this study can be used to predict the solute retention of new compounds well as can be used to better understand the mechanism of retention in complexation gas chromatography.

## References

- W. Wasiak. Chemically bonded chelates as selective complexing sorbents for gas chromatography: I. Alkenes. *J. Chromatogr.* **547**: 259–268 (1991).
- W. Wasiak. Chemically bonded chelates as selective complexing sorbents for gas chromatography. II: Ketones, ethers and nitroalkanes. *J. Chromatogr. A* **653**: 63–69 (1993).
- M.Y. Kuhuawar, A.A. Memon, and M.I. Bhangar. Nickel(II) Chelates of some Tetradentate Schiff Bases as Stationary Phases. *J. Chromatogr. A* **715**: 366–371 (1995).
- W. Wasiak. Investigation of the interaction of olefin-bonded transition metal complexes by gas chromatography. II. Phosphine complexes of Cu(II). *Chromatographia* **22**: 147–152 (1986).
- W. Wasiak and W. Urbaniak. Chemically bonded chelates as selective complexing sorbents for gas chromatography V. Silica chemically modified by Cu(II) complexes via amino groups. *J. Chromatogr. A* **757**: 137–143 (1997).
- R. Kalisz. Structure and Retention in chromatography, Hardwood, Amsterdam, 1997.
- B. Ren. Atom-type-based AI topological descriptors for quantitative structure–retention index correlations of aldehydes and ketones. *Chemom. Intell. Lab. Syst.* **66**: 29–39 (2003).
- O. Farkas and K. Heberger. Comparison of ridge regression, partial least-squares, pairwise correlation, forward and best subset selection methods for prediction of retention indices for aliphatic alcohols. *J. Chem. Inf. Model.* **45**: 339–346 (2005).
- T.F. Woloszyn and P.C. Jurs. Quantitative structure-retention relationship studies of sulfur vesicants. *Anal. Chem.* **64**: 3059–3063 (1992).
- M.H. Fatemi and E. Baher. Prediction of retention factors in supercritical fluid chromatography using artificial neural network. *Anal. Chem.* **60**: 969–975 (2005).
- M.H. Fatemi, H. Shamseddin, and H.J. Malekzadeh. Quantitative structure migration relationship modeling of migration factor for some benzene derivatives in micellar electrokinetic chromatography. *Sep. Sci.* **32**: 1934–1940 (2009).
- R. Yueying, L. Huanxiang, Y. Xiaojun, and L. Mancang. An accurate QSRR model for the prediction of the GC×GC–TOFMS retention time of polychlorinated biphenyl (PCB) congeners. *Anal. Bioanal. Chem.* **388**: 165–172 (2007).
- L.B. Ledesma and M. J. Wornat. QSRR Prediction of chromatographic retention of ethynyl-substituted PAH from semiempirically computed solute descriptors. *Anal. Chem.* **72**: 5437–5443 (2000).
- M. Pompe and M.J. Novic. Prediction of gas chromatographic retention indices using topological descriptors. *Chem. Inf. Comput. Sci.* **39**: 59–67 (1999).
- W. Wasiak and W. Urbaniak. Chemically bonded chelates as selective complexing sorbents for gas chromatography V. Silica chemically modified by Cu(II) complexes via amino groups. *J. Chromatogr. A* **757**: 137–143 (1997).
- W. Wasiak. GC investigation of olefins with chemically bonded gamma mercaptopropylsilane copper(II) complexes. *Chromatographia* **23**: 423–426 (1987).
- W. Wasiak. Investigation of specific interaction between olefins and cyano-silica modified with copper(II) complexes. *Chromatographia* **23**: 427–431 (1987).
- W. Wasiak. Specific interactions of n-donor adsorbates with N-benzoylthiourea-copper(II) complexes chemically bonded to silica supports. *Chromatographia* **41**: 107–111 (1995).
- W. Wasiak. Chemically bonded chelates as selective complexing sorbents for gas chromatography III. Silica chemically modified by N-benzoylthiourea groups. *J. Chromatogr. A* **690**: 93–102 (1995).
- W. Wasiak and I. Rykowska. Chemically bonded chelates as selective complexing sorbents for gas chromatography IV. Silica surfaces modified with Co(II) and Ni(II) complexes. *J. Chromatogr. A* **723**: 313–324 (1996).
- J.P.P. Stewart. MOPAC 6.0, Quantum Chemistry Program Exchange 1990.
- M.J.S. Dewar, E.G. Zoebisch, E.F. Healy and J.J.P. Stewart. Development and use of quantum molecular models, 75. Comparative tests of theoretical procedures for studying chemical reactions. *J. Am. Chem. Soc.* **107**: 3898–3902 (1985).
- A.R. Katritzky, V.S. Lobanov and M. Karelson. Comprehensive Descriptors for Structural and Statistical Analysis, 1994.
- A. R. Katritzky, V. S. Lobanov and M. Karelson. QSPR: the correlation and quantitative prediction of chemical and physical properties from structure. *Chem. Soc. Rev.* **24**: 279–287 (1995).
- A.R. Katritzky, V.S. Lobanov, and M. Karelson. QSPR as a means of predicting and understanding chemical and physical properties in terms of structure. *Pure Appl. Chem.* **69**: 245–248 (1997).
- S. Haykin. Neural Networks. A Comprehensive Foundation. 2nd ed. Pearson, Education, New Jersey, 1999.
- T. Masters. Practical Neural Network Recipes in C++, Academic Press, New York, 1993.
- R.C. Schweitzer and J.B. Morris. The development of a quantitative structure property relationship (QSPR) for the prediction of dielectric constants using neural networks. *Anal. Chim. Acta.* **384**: 285–303 (1999).
- W.L. Xing and X.W. He. Prediction of the selectivity coefficients of a berberine selective electrode using artificial neural networks. *Anal. Chim. Acta.* **349**: 283–286 (1997).
- C.S. Tong and K.C. Cheng. Mass spectral search method using the neural network approach. *Chemom. Intell. Lab. Syst.* **49**: 135–150 (1999).
- M. Jalali-Heravi and M.H. Fatemi. Artificial neural network modeling of Kovats retention indices for noncyclic and monocyclic terpenes. *J. Chromatogr. A* **915**: 177–183 (2001).
- K.L. Peterson. Counter-propagation neural networks in the modeling and prediction of Kovats indices for substituted phenols. *Anal. Chem.* **64**: 379–386 (1992).
- M. Jalali-Heravi and F. Parastaar. Use of artificial neural networks in a QSAR Study of anti-HIV activity for a large group of HEPT derivatives. *J. Chem. Inf. Comput. Sci.* **40**: 147–154 (2000).
- A.G. Maldonado, J.P. Doucet, M. Petitjean, and B.T. Fan. Molecular similarity and diversity in chemoinformatics: From theory to applications. *Mol. Divers.* **10**: 39–79 (2006).
- F. Luan, X.Y. Zhang, H.X. Zhang, R.S. Zhang, M.C. Hu, Z.D. Liu, and B.T. Fan. QSPR study of permeability coefficients through low-density polyethylene based on radial basis function neural networks and the heuristic method. *Comput. Mater. Sci.* **37**: 454–461 (2006).
- R. Todeschini and V. Consonni. Handbook of Molecular Descriptors, Wiley-VCH, Weinheim.
- M. Jalali and M.H. Fatemi. Prediction of flame ionization detector response factors using an artificial neural network. *J. Chromatogr. A* **825**: 161–169 (1998).
- M. Jalali-Heravi and M.H. Fatemi. Prediction of thermal conductivity detection response factors using an artificial neural network. *J. Chromatogr. A* **897**: 227–235 (2000).
- M.H. Fatemi, M. Jalali-Heravi, and E. Knouze. Prediction of bioconcentration factor using genetic algorithm and artificial neural network. *Anal. Chim. Acta.* **486**: 101–108 (2003).
- S. Wold and L. Eriksson. Chemometrics Methods in Molecular Design, VCH, Weinheim 1995.
- R. Todeschini and V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH: Weinheim, Germany, 2002.
- A. R. Katritzky, L. Mu, V.S. Lobanov and M. Karelson. Correlation of boiling points with molecular structure. 1. A training set of 298 diverse organics and a test set of 9 simple inorganics. *J. Phys. Chem.* **100**: 10400–10407 (1996).
- A.R. Katritzky, L. Mu, and M. Karelson. A QSPR. Study of the Solubility of Gases and Vapors in Water. *J. Chem. Inf. Comput. Sci.* **36**: 1162–1168(1996).

Manuscript received January 19, 2010;  
revision received May 4, 2010.