Retention modeling in gas chromatography by QSRR approach

M.N. Moskovkina*, I.P. Bangov, A.Zh. Patleeva

Department of Chemistry, Faculty of Natural Sciences, Shumen University, Shumen, 9712, 115 University str., Bulgaria

The Quantitative Structure–Retention Relationship (QSRR) approach was applied to model the retention behavior of substituted phenols in Gas Chromatography (GC). The experimental retention data for a set of 42 phenol derivates, including priority pollutants, separated on GC columns with different polarities – the non–polar SE–30 phase and the polar phases OV–225 and NGA, were taken from literature. The Multiple Linear Regression (MLR) statistical method was preferred for the QSRR developing. A charge–related topological index (CTI), developed by one of the authors (I.B.) was probed as a global descriptor in order to measure and compare its potential to contribute to QSRRs. Analysis of the equations derived with different descriptors proved the ability to describe and evaluate the participants in the chromatographic retention process. Comparison of the numerical values of the regression coefficients in similar QSRR models for different stationary phases exhibits the specific features of solute–stationary phase interactions in each case. The parametric values of the regression coefficients in similar models for different stationary phases correlate with McReynolds phase selectivity.

Keywords: QSRR; substituted phenols; molecular indices; gas chromatographic Kovats retention index; phase polarity; McReynolds constants.

1. INTRODUCTION

Phenol derivates are constantly a point of interest for analytic chemists. Most of phenol substituted derivates form the list of priority pollutants [1, 2] and bring harm to human health and environmental problems. Among the widespread analytical methods for phenol detection and identification is the gas chromatography. Experimental chromatographic methods for component testing in complex organic mixtures have some limits: you need to own expensive instruments and to collect a wide range of standards subjected to rigorous analytical testing to verify identity and determine purity. The quantitative relationships between the solute structure and its retention data (QSRR), proposed in 1979 [3], are often used to solve the identification problem to predict the chromatographic separation behaviour of the solutes. One of the main reasons to deal with the QSRR approach is due to its possibility to study the relationships between the solute structure and its chromatographic retention data.

A comprehensive look over the publications [4–10] dealing with the QSRR subject reveals its importance until today.

Linear methods are widely applied in the QSAR and QSRR area [11–14].

Support vector machine (SVM), a non-linear algorithm was developed for regression and classification [15] and gained popularity in QSPR studies for drug design and biological activity [16, 17].

Another non-linear regression method used in chemometric investigations, especially in chemomatics and bioinformatics, is the neural network (NN) method [18–24]. Three different mathematical approaches (SVM, NN and MLR) were used in [25] to investigate the relationship between structure and retention index and to derive QSRRs for data sets of 174 and 132 diverse compounds. The statistical models derived by these methods revealed similar prediction ability.

The preferable statistical method for QSRR remains the Multiple Linear Regression (MLR) developing.

The modern state of art in the QSRR approach gives the possibility to generate multivariable regression equations able to reflect the chromatographic retention behavior for different solute series in different chromatographic modes. The goal of QSRR is to predict the retention characteristics and to take a peep into the mechanism of chromatographic separations. Surely the anticipated development of more precise methods for solute structure parameterization will be advanced due to QSRR studies.

---

* To whom all correspondence should be sent:
E-mail mmarinan@abv.bg

When a multiple linear regression form is preferred for the QSRR modelling, each element of the matrix of regression coefficients has certain significance. The value of the correlation coefficient **r** indicates the physical meaning of the linear relationship between the retention parameter and the derived set of molecular descriptors. The statistical significance of the parameters included in a common regression can be used to render an account of the significance of the participation of each descriptor in intermolecular interactions during the separation process. This information can be used to quantify the nonspecific (dispersive) and specific (polar) interactions between the solutes and the chromatographic phases. Whereas quantifying of the nonspecific dispersive molecular interactions can be successfully carried out by using some global topological or constitutive molecular indices, the various specific polar interactions are quantified less precisely by numeric local molecular indices.

Generally for gas chromatographic practice, all problems dealing with solute separation to be reduced to phase selectivity end inevitably to McReynolds constants [26].

Despite a serious amount of critical publications against the theoretical base of McReynolds system [27 mmarinan@abv.bg 33]］, the usage of McReynolds constants is still popular in chromatographic laboratories and literature.

In the present investigation the experimental GC retention data are related to the separation in 3 different columns: non polar phase SE–30 (methylsilicone) and two polar phases: OV–225 (3–cyanopropyl methylphenyl polysiloxane) and NGA (neopentyl glycol adypate). The phases OV–225 and NGA have almost similar polarities according to McReynolds polarity scale (1849 and 1813), but differ in donor–acceptor characteristics. Both phases exhibit electron donor properties. The NGA phase structure contains the oxygen atom in a carboxyl group as an important electron pair donor and a center for hydrogen-bonding interactions. The OV–225 phase possesses cyano-groups – centers for electron pair donor - electron pair acceptor interactions.

The chemometric analysis of the gas chromatographic retention on stationary phases with similar polarities and small variance in McReynolds constants gives a chance to expose the ability to reflect the molecular solute–stationary phase interactions. A large number of solute sets (n = 42) allows to derive regressions with a wide set of molecular descriptors to contribute significantly to the regression models and to rise the statistical accuracy of the resulting correlation.

The chromatographic phase selectivity is a complex term rendering an account for different kinds of solute–stationary phase interactions. According to Rohrschneider–McReynolds concept, the Kovats retention index difference ΔI_x of the test compound on the column (I_i) and squalane (I_{sq}) provides a measure of polarity [26, 34–35]. The polarity of squalane is defined to be 0 as a standard apolar reference phase.

The empirical McReynolds scale for stationary phase polarity/selectivity characterization deals with ten test compounds. The phase selectivity is expressed with 10 indices (McReynolds Stationary phase polarity constants).

In the present paper the Quantitative Structure–Retention Relationship (QSRR) approach was applied to model the retention behavior of a set of substituted phenols (n = 42) in gas chromatography (GC) separated on stationary phases with different polarities, to describe and evaluate the participants in the chromatographic retention process and to compare the information received from QSRRs with McReynolds Polarity Constants.

2. EXPERIMENTAL COMPUTATION METHODOLOGY

2.1. Retention data

The experimental retention data – the Kovats indices - for the set of 42 substituted phenols including priority pollutants separated on GC columns with different polarities – the nonpolar SE–30 phase and the polar phases OV–225 and NGA, were taken from the literature [37]. The McReynolds constants for the three phases of our investigation were taken from literature [36] and are shown in Table 1. The structures of the solute set are shown in Fig. 1.
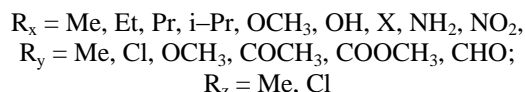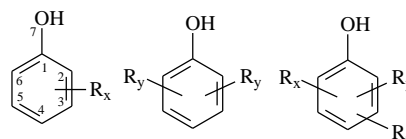

R_x = Me, Et, Pr, i–Pr, OCH$_3$, OH, X, NH$_2$, NO$_2$,
R_y = Me, Cl, OCH$_3$, COCH$_3$, COOCH$_3$, CHO;
R_z = Me, Cl

Fig. 1. General formulae of the phenol derivates.

Table1. McReynolds Stationary phase polarity constants

McR	I_1	ΔI_2	ΔI_3	ΔI_4	ΔI_5	ΔI_6	ΔI_7	ΔI_8	ΔI_9	ΔI_{10}
	X′	Y′	Z′	U′	S′	H′	I′	K′	L′	M′
SE–30	15	53	44	64	41	31	3I	22	44	2
OV–225	228	369	338	492	386	282	226	150	342	117
NGA	234	425	312	402	438	339	210	157	362	103

ΔI_1 (benzene) – **X′**; ΔI_2 (1-butanol) – **Y′**; ΔI_3 (2-pentanone) – **Z′**; ΔI_4 (1-nitropropane) – **U′**; ΔI_5 (pyridine) – **S′**; ΔI_6 (2-methyl-2-pentanol) – **H′**; ΔI_7 (1-iodobutane) – **I′**; ΔI_8 (2-octyne) – **K′**; ΔI_9 (1,4-dioxane) – **L′**, ΔI_{10} (*cis*-hidrindane) – **M′**

Table 2.1. Experimental Gas Chromatographic Retention Data (*Kovats Indices)* for substituted phenols separated on three stationary phases (I phase- SE-30; II phase –OV-225; III phase-NGA).

№	Compounds	I_{SE-30}	I_{OV-225}	I_{NGA}	№	Compounds	I_{SE-30}	I_{OV-225}	I_{NGA}
1	2-Me Ph	1035	1587	1742	22	3-NH$_2$ Ph	1335	2219	2352
2	4-Me Ph	1059	1654	1813	23	2,4,6triCl Ph	1349	1928	2067
3	3-Me Ph	1065	1648	1782	24	2,4,5-triCl Ph	1362	2039	2158
4	2,6diMePh	1098	1593	1716	25	3-OH Ph	1368	2371	2576
5	2,4diMePh	1134	1660	1825	26	3,5-Cl Ph	1391	2217	2343
6	3-Et Ph	1160	1742	1898	27	4-I Ph	1398	2230	2348
7	4-Et Ph	1162	1746	1890	28	4-CO$_2$CH$_3$Ph	1500	2376	2461
8	3,5diMe Ph	1163	1706	1877	29	2-NH$_2$ Ph	1242	2039	2196
9	2,3diMe Ph	1169	1693	1857	30	3-Br Ph	1270	2069	2214
10	2.4diCl Ph	1183	1708	1877	31	2-I-Prop,5-Me Ph	1271	1776	1932
11	4-Cl Ph	1192	1922	2058	32	2-OCH$_3$ Ph	1095	1544	1627
12	3-Cl Ph	1194	1911	2061	33	2-NO$_2$Ph	1149	1556	1703
13	2,4,6-triMe Ph	1204	1612	1778	34	2.6-(OCH$_3$)$_2$Ph	1347	1936	2014
14	2,6diCl Ph	1206	1727	1871	35	2-OCH$_3$-4-Pr Ph	1392	1810	1884
15	4-OCH$_3$ Ph	1210	1930	2050	36	2-OCH$_3$-4-CHO Ph	1447	2199	2235
16	3-OCH$_3$ Ph	1211	1940	2083	37	2.6-(OCH$_3$)$_2$-4-CH$_3$ Ph	1473	2076	2106
17	2,3,5,-triMe Ph	1260	1823	1960	38	2-OCH3-4-COCH$_3$ Ph	1531	2283	2326
18	4-Br Ph	1274	2054	2191	39	4-COCH$_3$ Ph	1578	2478	2529
19	3-Me,4-Cl Ph	1283	2025	2135	40	2.6-(OCH$_3$)2-4-Pr Ph	1624	2254	2256
20	4-NH$_2$ Ph	1314	2154	2277	41	2.6-(OCH$_3$)2-4-COCH$_3$ Ph	1849	2685	2683
21	4-OH Ph	1334	2330	2515	42	(2-OCH$_3$-4-OCH$_2$CH=CH$_2$)Ph	1367	1848	1923

Table 2.2. Structural indicative descriptors for substituted phenols.

№	Compounds	R orto	R vic	OCH₃	Cl	Br	I	X	NO₂	NH₂	R keton	R aldh	R ester	alkyl	R o-Cl	R o-Me
1	2-MePh	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1
2	4-MePh	0	0	0	0		0	0	0	0	0	0	0	1	0	0
3	3-MePh	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
4	2,6diMePh	2	0	0	0	0	0	0	0	0	0	0	0	2	0	2
5	2,4diMePh	1	0	0	0	0	0	0	0	0	0	0	0	2	0	1
6	3-EtPh	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
7	4-EtPh	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
8	3,5diMePh	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
9	2,3diMePh	1	1	0	0	0	0	0	0	0	0	0	0	2	0	1
10	2.4diCl	1	0	0	2	0	0	2	0	0	0	0	0	0	1	0
11	4-ClPh	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
12	3-ClPh	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
13	2,4,6-triMePh	2	0	0	0	0	0	0	0	0	0	0	0	3	0	2
14	2,6diCl Ph	2	0	0	2	0	0	2	0	0	0	0	0	0	2	0
15	4-OCH₃ Ph	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
16	3-OCH₃ Ph	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
17	2,3,5,-triMePh	1	1	0	0	0	0	0	0	0	0	0	0	3	0	1
18	4-Br Ph	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
19	3-Me,4-Cl Ph	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0
20	4-NH₂ Ph	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
21	4-OH Ph	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	3-NH₂ Ph	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
23	2,4,6triCl Ph	2	0	0	3	0	0	3	0	0	0	0	0	0	2	0
24	2,4,5-triCl Ph	1	0	0	3	0	0	3	0	0	0	0	0	0	1	0
25	3-OH Ph	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	3,5-Cl Ph	0	0	0	2	0	0	2	0	0	0	0	0	0	0	0
27	4-I Ph	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
28	4-CO₂CH₃ Ph	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
29	2-NH₂ Ph	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
30	3-Br Ph	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
31	2-i-Prop,5-MePh	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0
32	2-OCH₃ Ph	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
33	2-NO₂ Ph	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
34	2.6-(OCH₃)₂ Ph	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0
35	2-OCH₃-4-Pr Ph	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0
36	2-OCH₃-4-CHO Ph	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0
37	2.6-(OCH₃)₂-4-CH₃ Ph	2	0	2	0	0	0	0	0	0	0	0	0	1	0	0
38	2-OCH₃-4-COCH₃ Ph	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0
39	4-COCH₃ Ph	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
40	2.6-(OCH₃)₂-4-Pr Ph	2	0	2	0	0	0	0	0	0	0	0	0	1	0	0
41	2.6-(OCH3)₂-4-COCH₃ Ph	2	0	2	0	0	0	0	0	0	1	0	0	0	0	0
42	2-OCH₃-4-COCH₂CH=CH₂Ph	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The solute set of substituted phenols and the experimental retention data (Kovats indices I_i) for the three stationary phases –SE-30, OV–225 and NGA, are listed in Table 2.1.

2.2. Descriptor calculation

The multiple linear regressions (MLR) were derived according to equation (1):

$$I = \sum d_i D_i + \sum p_i P_i + a_0, \qquad (1)$$

where I is the experimental retention value (the Kovats Retention Index); D_i – a global structure molecular index to quantify the non–specific chromatographic interactions. We have checked some global molecule indices ("*bulk* indices") such as molecular mass (M), molar refractivity (MR), polarizability (α), and Wiener topology index (W) as a measure of solute dispersive properties (D = M, MR, α, W); for the P_i indices specific molecular descriptors should be used to quantify the polar intermolecular interactions of the solute with the stationary phase; a_o, d_i and p_i are the regression coefficients depending on the properties of the stationary phases.

A charge–related topological index (CTI) developed by one of the authors (I.B.) [38–40] was probed as a D descriptor in order to measure and compare its ability to contribute to QSRRs.

The CTI index is expressed as:

$$CTI = \sum_i \sum_j \frac{L_i L_j}{D_{ij}} \quad (2)$$

Here D_{ij} are the inter–atomic distances and L_i are local indices characterizing the separate heavy (non–hydrogen) atoms i which are expressed as follows:

$$L_i = L_o + N_H + q_i \quad (3)$$

L_o are the constant values for each atom and for each hybridization state (they can be atom valences in some cases), N_H is the number of hydrogen atoms attached to a given heavy atom, and q_i are the corresponding charge densities computed by either the empirical method of Gasteiger or by any of the most sophisticated quantum chemistry methods at semiempirical or nonempirical level. In the present investigation two kinds of CTI indices were developed - CTI_{AM1} and CTI_{DFT}, calculated either at the semiempirical AM1 or at the nonempirical DFT level.

The molecular indices M, MR, α, were calculated in Excel [41]. The CTI index and Wiener index W were developed with a Str Manager Software program.

Some kinds of solute structure descriptors were used for the P_i variables: nonempirical quantum chemical indices (dipole moment μ, atomic charges q_i; energy of the highest occupied molecular orbital (HOMO) E_{Homo} and energy of the lowest unoccupied molecular orbital (LUMO) E_{Lumo}; energy of hydration E_{hydr}; total molecular energy E_{total} on one side and a set of structure indicative descriptors to reflect the presence of different functional groups in the solute structures: P_i = R_{ketone}, $R_{aldehide}$, R_{ester}, OH, NH_2, NO_2, Hal, Cl, Br, I, OCH_3, alkyl, on the other. Some additional indicative descriptors were used to reflect the presence of the substitutes in *ortho*-position towards the OH-group of the phenol ring: R_{orto}, $R_{o–OH}$, $R_{o–NO2}$, $R_{o–NH2}$, $R_{o–Cl}$, especially the presence of methyl substituents in vicinal position towards the OH-group of the phenol ring – R_{vic}. The values of the indicative structural descriptors used are listed in Table 2.2.

The geometries of the molecules were fully optimized at the semiempirical AM1 or DFT B3LYP/6–31+G(d,p) level of theory and were characterized as minima (no imaginary frequencies) at the potential energy surface (PES) by frequency calculations at the same computational level. All reported total energies were corrected by zero-point energy (unscaled), estimated from the harmonic frequency calculations at temperature 273 K and atmospheric pressure 1.01325×10^5 Pa. All calculations were performed using the PC GAMESS version 7.1 (Tornado) [42] of the GAMESS (US) QC package [43].

The molecular quantum indices calculated with AM1 and entered into the developed QSRR were: atomic charges Q_i ; (Q_1–Q_6 are the charges of the C-atoms in the benzene ring; Q_7 is the atomic charge of the O–atom in the phenyl group); the energy of hydration – E_{hydr}, the energy of the HOMO-orbitals –E_{homo}. The procedure of CTI index generation permits to use the atomic charge values calculated with different quantum methods. Since we have used both the semi-empirical AM1 and the DFT approach for atom charges calculation, two kinds of CTI indices were obtained – the CTI_{AM1} and the CTI_{DFT} one. The numeric values of the quantum indices developed with the AM1 approach, along with the global dispersive molecular descriptors, are presented in Table 2.3.

The similar quantum descriptors calculated with DFT and used for QSRR modeling were: atomic charges q_i (Lowdin); (q_1–q_6 are the charges of the C-atoms in the benzene ring; q_7 is the atomic charge of the O–atom in the phenyl group); dipole moment μ (Debye); E_{tot} – the total molecular energy (Hartree), the energy of HOMO-orbitals – $E_{Homo (DFT)}$ and CTI_{DFT} values, which are presented in Table 2.4.

Table 2.3. The Global dispersive molecular descriptors calculated for QSRR models (M – molecular mass; MR – molar refractivity; α – polarizability, W – Wiener topologic index), and quantum molecular descriptors, calculated at AM1 level: charge-relative topology index – CTI (AM1); HOMO-energy – E_{Homo}; E_{hydr} – hydratation energy; Q_i – atomic charges of resp. atoms.

№	Compounds	M	MR	a	W	CTI (AM1)	E_{Homo} (AM1)	E_{hydr} (AM1)	Q_1 (AM1)	Q_2 (AM1)	Q_6 (AM1)	Q_7 (AM1)
1	2-MePh	108.14	2.79	12.91	162	118.90	-8.997	-7.38	0.080	-0.151	-0.155	-0.255
2	4-MePh	108.14	32.79	12.91	182	117.94	-8.880	-7.66	0.073	-0.209	-0.154	-0.253
3	3-MePh	108.14	32.79	12.91	173	118.18	-9.012	-7.58	0.081	-0.213	-0.162	-0.253
4	2,6diMePh	122.17	37.83	14.74	207	128.26	-8.891	-5.01	0.084	-0.148	-0.089	-0.256
5	2,4diMePh	122.17	37.83	14.74	224	127.43	-8.786	-6.15	0.076	-0.148	-0.153	-0.254
6	3-EtPh	122.17	37.39	14.74	217	125.49	-9.019	-7.05	0.080	-0.211	-0.161	-0.253
7	4-EtPh	122.17	37.39	14.74	235	125.16	-8.910	-7.18	0.074	-0.155	-0.210	-0.253
8	3,5diMePh	122.17	37.83	14.74	229	126.94	-8.970	-6.19	0.085	-0.218	-0.165	-0.254
9	2,3diMePh	122.17	37.83	14.74	211	127.55	-8.916	-6.21	0.085	-0.146	-0.160	-0.257
10	2.4diCl	163	37.36	14.93	224	153.81	-9.271	-7.87	0.094	-0.144	-0.144	-0.241
11	4-ClPh	128.56	32.56	13	182	131.13	-9.125	-8.58	0.081	-0.207	-0.149	-0.248
12	3-ClPh	128.56	32.56	13	173	131.76	-9.300	-8.49	0.089	-0.212	-0.156	-0.248
13	2,4,6-triMePh	136.19	42.88	16.58	264	136.67	-8.698	-3.93	0.077	-0.145	-0.085	-0.256
14	2,6diCl Ph	163	37.36	14.93	207	155.12	-9.374	-6.85	0.100	-0.145	-0.087	-0.230
15	4-OCH$_3$ Ph	124.14	32.22	13.54	235	134.24	-8.636	-10.63	0.044	-0.180	-0.124	-0.253
16	3-OCH$_3$ Ph	124.14	32.22	13.54	217	134.77	-8.966	-10.43	0.110	-0.246	-0.190	-0.250
17	2,3,5,-triMePh	136.19	42.88	16.58	269	136.48	-8.850	-4.87	0.084	-0.147	-0.164	-0.258
18	4-Br Ph	173.01	35.37	13.7	182	130.93	-9.189	-8.56	0.093	-0.158	-0.216	-0.247
19	3-Me,4-Cl Ph	142.59	37.6	14.83	238	140.35	-9.035	-7.31	0.085	-0.210	-0.155	-0.249
20	4-NH$_2$ Ph	109.13	32.45	12.42	182	121.88	-7.957	-13	0.016	-0.159	-0.104	-0.256
21	4-OH Ph	110.11	29.45	11.71	182	126.43	-8.725	-15.89	0.044	-0.178	-0.119	-0.253
22	3-NH$_2$ Ph	109.12	32.45	12.42	187	121.67	-8.281	-12.78	0.126	-0.306	-0.217	-0.253
23	2,4,6triCl Ph	197.45	42.17	16.85	264	174.72	-9.390	-6.5	0.101	-0.139	-0.081	-0.226
24	2,4,5-triCl Ph	197.45	42.17	16.85	264	175.34	-9.388	-7.39	0.103	-0.146	-0.145	-0.237
25	3-OH Ph	110.4	29.45	11.71	187	126.78	-8.982	-15.7	0.117	-0.301	-0.192	-0.249
26	3,5-Cl Ph	163	37.36	14.93	229	153.35	-9.537	-7.95	0.099	-0.214	-0.155	-0.243
27	4-I Ph	220.01	40.16	16.1	182	130.80	-9.243	-8.55	0.097	-0.220	-0.162	-0.247
28	4-CO$_2$CH$_3$ Ph	152.15	39.28	15.46	355	166.90	-9.536	-9.16	0.114	-0.233	-0.178	-0.245
29	2-NH$_2$ Ph	109.13	32.45	12.42	169	122.41	-8.204	-12.3	-0.030	0.058	-0.110	-0.255
30	3-Br Ph	173.01	35.37	13.7	173	131.67	-9.337	-8.46	0.081	-0.191	-0.144	-0.248
31	2-i-Prop,5-MePh	150.22	46.98	18.41	321	141.79	-8.962	-5.32	0.085	-0.148	-0.160	-0.258
32	2-OCH$_3$ Ph	124.14	34.22	13.54	80	136.02	-8.783	-9.63	0.066	-0.004	-0.137	-0.250
33	2-NO$_2$ Ph	139.11	35.08	12.91	251	165.39	-9.911	-12.62	0.177	-0.229	-0.180	-0.245
34	2.6-(OCH$_3$)$_2$ Ph	154.17	40.46	16.02	301	161.14	-8.745	-8.5	-0.003	0.030	0.083	-0.261
35	2-OCH$_3$-4-Pr Ph	166.22	48.46	19.05	438	156.52	-8.621	-7.46	0.061	-0.250	0.002	-0.135
36	2-OCH$_3$-4-CHO Ph	152.15	40.81	15.46	351	166.37	-9.126	3	0.100	-0.019	-0.156	-0.243
37	2.6-(OCH$_3$)$_2$-4-CH$_3$ Ph	168.19	45.72	17.85	372	169.49	-8.619	-7.31	-0.050	-0.261	0.032	0.085
38	2-OCH$_3$-4-COCH$_3$ Ph	166.18	43.62	17.3	389	174.23	-9.151	-8.62	0.050	-0.009	-0.143	-0.252
39	4-COCH$_3$ Ph	36.15	37.16	14.83	290	175.04	-9.428	-7.83	0.109	-0.230	-0.175	-0.246
40	2.6-(OCH$_3$)$_2$-4-Pr Ph	196.25	54.92	21.52	551	181.57	-8.636	-6.37	-0.003	0.030	0.083	-0.260
41	2.6-(OCH3)$_2$-4-COCH$_3$ Ph	196.2	50.09	19.77	540	198.12	-9.013	-7.42	0.015	0.026	-0.082	-0.265
42	2-OCH$_3$-4-COCH$_2$CH=CH$_2$ Ph	164.25	48.5	18.86	438	162.93	-8.657	-9.51	0.063	-0.002	-0.135	-0.250

14

Table 2.4. The quantum descriptors for QSRR models calculated at DFT level (CTI$_{DFT}$ –charge-relative topology index; E$_{Homo}$ –the Homo- energy; q$_i$ –atomic charges (Lowdin); E$_{total}$ – total energy (Hartree)

№	Compounds	E$_{Homo}$ (DFT)	CTI (DFT)	E$_{total}$ (Hartree)	q$_1$ (DFT)	q$_2$ (DFT)	q$_6$ (DFT)	q$_7$ (DFT)
1	2-MePh	-0.0063	128.2227	-346.66	0.113	-0.033	-0.107	-0.233
2	4-MePh	-0.0002	117.1808	-346.66	0.119	-0.140	-0.112	-0.234
3	3-MePh	-0.0022	117.8407	-346.66	0.122	-0.150	-0.117	-0.233
4	2,6diMePh	-0.0066	139.563	-385.95	0.108	-0.031	-0.008	-0.230
5	2,4diMePh	-0.0058	138.6126	-385.96	0.111	-0.029	-0.112	-0.234
6	3-EtPh	-0.0009	138.615	-385.95	0.122	-0.144	-0.117	-0.230
7	4-EtPh	-0.0018	136.6195	-385.95	0.120	-0.112	-0.139	-0.233
8	3,5diMePh	-0.002	136.1877	-385.96	0.125	-0.130	-0.158	-0.234
9	2,3diMePh	-0.0074	140.1745	-385.95	0.115	-0.040	-0.117	-0.228
10	2.4diCl	-0.0085	147.1098	-1226.6	0.111	-0.104	-0.096	-0.216
11	4-ClPh	-0.0099	120.8116	-766.97	0.123	-0.132	-0.104	-0.227
12	3-ClPh	-0.0078	121.9637	-766.9	0.127	-0.146	-0.112	-0.222
13	2,4,6-triMePh	-0.0061	161.3894	-425.25	0.103	-0.006	-0.031	-0.232
14	2,6diCl Ph	-0.0107	147.8288	-1226.6	0.101	-0.105	-0.090	-0.206
15	4-OCH$_3$ Ph	-0.0006	138.0317	-421.86	0.106	-0.129	-0.105	-0.240
16	3-OCH$_3$ Ph	-0.2131	139.9949	-421.86	0.355	-0.200	-0.130	-0.565
17	2,3,5,-triMePh	-0.0068	159.0513	-425.25	0.116	-0.046	-0.127	-0.228
18	4-Br Ph	-0.0007	120.9104	-2878.5	0.124	-0.133	-0.105	-0.227
19	3-Me,4-Cl Ph	-0.0083	139.9538	-806.27	0.123	-0.143	-0.112	-0.227
20	4-NH$_2$ Ph	-0.1853	115.4051	-362.71	0.101	-0.103	-0.130	-0.243
21	4-OH Ph	-0.0004	117.448	-382.58	0.106	-0.127	-0.103	-0.240
22	3-NH$_2$ Ph	-0.0043	113.891	-362.71	0.130	-0.181	-0.143	-0.230
23	2,4,6triCl Ph	-0.0028	174.7997	-1686.17	0.098	-0.086	-0.101	-0.203
24	2,4,5-triCl Ph	-0.0039	174.3801	-1686.17	0.113	-0.108	-0.109	-0.209
25	3-OH Ph	-0.0097	115.7902	-382.59	0.131	-0.193	-0.137	-0.226
26	3,5-Cl Ph	-0.0013	144.7492	-1226.58	0.361	-0.103	-0.131	-0.553
27	4-I Ph	-0.0034	120.6733	-7195.74	0.270	-0.141	-0.112	-0.475
28	4-CO$_2$CH$_3$ Ph	-0.0086	209.4104	-535.21	0.136	-0.140	-0.114	-0.219
29	2-NH$_2$ Ph	-0.1896	117.0761	-362.71	0.086	0.053	-0.129	-0.245
30	3-Br Ph	-0.0087	122.0323	-2878.53	0.127	-0.149	-0.110	-0.223
31	2-i-Prop,5-MePh	-0.007	183.2764	-464.53	0.116	-0.032	-0.122	-0.228
32	2-OCH$_3$ Ph	-0.0018	141.3289	-421.86	0.098	0.078	-0.113	-0.227
33	2-NO$_2$ Ph	-0.0014	195.7651	-511.88	0.141	-0.034	-0.108	-0.194
34	2.6-(OCH$_3$)$_2$ Ph	-0.0028	186.7031	-536.35	0.073	0.089	0.079	-0.227
35	2-OCH$_3$-4-Pr Ph	-0.0034	199.7381	-539.73	0.095	0.077	-0.110	-0.232
36	2-OCH$_3$-4-CHO Ph	-0.0051	213.9548	-535.18	0.118	0.077	-0.113	-0.210
37	2.6-(OCH$_3$)$_2$-4-CH$_3$ Ph	-0.0037	210.6052	-575.64	0.068	0.089	0.081	-0.230
38	2-OCH$_3$-4-COCH$_3$ Ph	-0.0019	240.3078	-574.48	0.114	0.075	-0.112	-0.214
39	4-COCH$_3$ Ph	-0.0102	186.8544	-459.98	0.350	-0.122	-0.143	-0.552
40	2.6-(OCH$_3$)$_2$-4-Pr Ph	-0.004	254.1766	-654.22	0.069	0.090	0.081	-0.227
41	2.6-(OCH3)$_2$-4-COCH$_3$ Ph	-0.0056	298.1189	-688.96	0.088	0.085	0.078	-0.212
42	2-OCH$_3$-4-COCH$_2$CH=CH$_2$ Ph	-0.0031	221.6187	-538.52	0.095	0.076	-0.107	-0.231

2.3. Calculation method

The chemometric approach for QSRRs development was executed with Excel program [41]. The regression linear models (MLR) were obtained by using forward stepwise multiple regression techniques.

The requirements for statistical accuracy [44] for all developed MLR models were fulfilled. The best regression model was selected on the basis of the highest values of the correlation coefficient (r), the square correlation coefficient (r^2), the value of the F-test, (a statistic for assessing the overall significance), the lowest standard error of estimation (S) and the maximum residual value between the experimental retention data and those calculated with equation derived Δ_{max}. The cross–correlation coefficients r_{ij} between the independent variables in the equation are presented in Table 3–1 for AM1-calculated descriptors and in Table 3–2 for DFT-calculated descriptors. In order to derive meaningful results, the independence of the

variables (their orthogonality) was checked. The occasionally observed colinearity of the structural parameters used in the same equation eliminates its informative value. When the number of cases (experimental data) analyzed is limited, high correlations can be obtained by including large numbers of independent variables in a process of regressions deriving. The correlations thus obtained can be statistically insignificant unless the F-test value for a given number of degrees of freedom is lower than the value calculated for the respective significance level. The sequential F-test allows one to decide whether an introduction of an individual independent variable into the regression equation is statistically justified [45].

3. RESULTS AND DISCUSSION

The QSRRs derived for three phases and their statistics are shown as three parts of Table 4 as a set of descriptors, grouped as follows: each of the first five equations includes a global descriptor and is tuned just with indicative structural descriptors; the regressions in the next two groups are developed with entering the local indicative descriptors and quantum descriptors calculated with the AM1 method (eq. 6–9) and DFT approach (eq. 10–13), respectively.

The set of regressions 1-5 developed for each stationary phase, displays good statistics (R = 0.955–0.975 for SE–30; R = 0.931–0.947 for OV–225 and R = 0.925–0.940 for NGA). Each of the checked global molecular indices (M, MR, α, W, CTI) can be successfully used for QSRR modeling as a descriptor with almost similar statistic accuracy. One of the main factors monitored in chromatographic laboratory practice is the value of Δ_{max}, which indicates the deviation between experimental and calculated retention data. The experimental interlaboratory uncertainty of retention data on polar phases is up to 25 index units [5]. Proceeding to these requirements, the Δ_{max} values corresponding to our equations were unsatisfactorily larger than the experimental error generally assumed, even though the statistic parameters for the equations 1–5 (Table 4) seem to be high: (Δ_{max} = 84–124 i.u. for SE-30; Δ_{max} = 201–269 i.u. for OV–225 and Δ_{max} = 249-299 i.u. for NGA).

The addition of calculative descriptors into QSRRS improves their statistics. In the case of AM1 approach for QSRR modeling, the entering of the quantum descriptors E_{Homo} and Q_i (eq.6-9) rise the R value (R = 0.974-0.983 for SE-30; R = 0.961-0.981 for OV–225 and R = 0.961–0.982 for NGA).

The values of Δ_{max} change as follows for the equations 6-9 (Table 4): (Δ_{max} = 77–117 i.u. for SE–30; Δ_{max} = 103–188 i.u. for OV–225 and Δ_{max} = 89–186 i.u. for NGA).

The entering of the quantum descriptors E_{total} and q_i calculated at DFT level (eq. 10–13, Table 4) was meaningful and led to an increase in R (R=0.955–0.981 for SE–30; R= 0.965–0.975 for OV–225 and R= 0.944–0.974 for NGA). The changes for Δ_{max} in the equations 10–13 are (Table 4): (Δ_{max} = 74–103 i.u. for SE–30; Δ_{max} = 133–209 i.u. for OV–225 and Δ_{max} = 121–189 i.u. for NGA).

The retention modeling in both polar phases benefits from the addition of quantum descriptors, calculated with either AM1, or DFT approach. The entering of the E_{Homo} parameter into the models improves the statistics, particularly for the polar phases OV–225 and NGA (Table 4, eq.6). This descriptor E_{Homo} quantifies electron pair donor - electron pair acceptor (EPD–EPA) interactions and meaningfully parameterizes the polar intermolecular interactions with the stationary phase. The role of the energies E_{Homo} and E_{Lumo} as an expression of Lewis basicity and acidity, respectively, is known and accounted for [46].

The full forms of the equations 6 from Table 4 with descriptor E_{Homo} for the three phases are presented in Table 5–1. It seems interesting to track the influence of entering the index E_{Homo} to QSRRs for phases with different polarities. The numeric values of the regression coefficients for the E_{Homo} descriptor increase according to phase selectivity towards the EPD-EPA interactions: (104.5 ± 43.3) for SE–30; (286.6 ± 153.9) for OV–225 and (321.2 ± 143.4) for NGA.

Comparing the statistics for the equations from Table 4 derived with different calculation methods allows claiming that the accuracy in the case of semiempirical AM1 approach is sufficient and it seems pointless to execute DFT calculations for QSRR developing.

Table 3-2. The cross-correlation matrix (DFT)

	M	MR	α	CTI_{dft}	E_{tot}	E_{homo}	q_1	q_2	q_6	q_7	OH	NH_2	NO_2	X	OCH_3	R_{orto}	$R_{o\text{-}Cl}$	R_{aldh}	R_{keton}	R_{ester}
M	1																			
MR	0.70	1																		
α	0.72	0.99	1																	
CTI_{dft}	0.58	0.83	0.82	1																
E_{tot}	-0.63	-0.06	-0.08	0.15	1															
E_{homo}	0.27	0.28	0.27	0.22	-0.12	1														
q_1	0.04	-0.22	-0.17	-0.16	-0.28	-0.23	1													
q_2	0.29	0.66	0.62	0.69	0.22	0.08	-0.43	1												
q_6	0.40	0.55	0.55	0.57	0.05	0.16	-0.36	0.58	1											
q_7	-0.08	0.13	0.07	0.12	0.30	0.31	-0.96	0.28	0.20	1										
OH	-0.25	-0.34	-0.33	-0.22	0.09	0.06	-0.04	-0.24	-0.10	0.05	1									
NH_2	-0.33	-0.28	-0.32	-0.28	0.12	-0.61	-0.10	-0.04	-0.18	0.05	-0.06	1								
NO_2	-0.03	-0.09	-0.15	0.14	0.05	0.05	0.02	0.05	-0.04	0.10	-0.03	-0.04	1							
X	0.53	0.01	0.05	-0.12	-0.43	0.14	0.13	-0.28	-0.11	-0.06	-0.12	-0.14	-0.08	1						
OCH_3	0.34	0.50	0.53	0.65	0.14	0.00	-0.19	0.64	0.76	0.04	-0.12	-0.14	-0.08	-0.27	1					
R_{orto}	0.43	0.65	0.64	0.60	0.12	0.24	-0.43	0.70	0.72	0.33	-0.20	-0.25	0.06	0.10	0.50	1				
$R_{o\text{-}Cl}$	0.38	0.08	0.11	0.02	-0.16	0.07	-0.12	-0.12	-0.02	0.16	-0.07	-0.08	-0.05	0.77	-0.16	0.39	1			
R_{aldh}	0.04	0.07	0.03	0.20	0.04	0.04	-0.03	0.24	-0.05	0.08	-0.03	-0.04	-0.02	-0.08	0.16	0.06	-0.05	1		
R_{keton}	0.21	0.26	0.28	0.54	0.07	0.07	0.21	0.23	0.16	-0.21	-0.06	-0.08	-0.04	-0.14	0.29	0.11	-0.08	-0.04	1	
R_{ester}	0.04	0.03	0.03	0.19	0.04	0.03	0.01	-0.13	-0.05	0.06	-0.03	-0.04	-0.02	-0.08	-0.08	-0.14	-0.05	-0.02	-0.04	1

Table 3-2. The cross-correlation matrix (DFT)

	M	MR	α	CTI_{dft}	E_{tot}	E_{homo}	q_1	q_2	q_6	q_7	OH	NH_2	NO_2	X	OCH_3	R_{orto}	R_{o-Cl}	R_{aldh}	R_{keton}	R_{ester}
M	1																			
MR	0.70	1																		
α	0.72	0.99	1																	
CTI_{dft}	0.58	0.83	0.82	1																
E_{tot}	-0.63	-0.06	-0.08	0.15	1															
E_{homo}	0.27	0.28	0.27	0.22	-0.12	1														
q_1	0.04	-0.22	-0.17	-0.16	-0.28	-0.23	1													
q_2	0.29	0.66	0.62	0.69	0.22	0.08	-0.43	1												
q_6	0.40	0.55	0.55	0.57	0.05	0.16	-0.36	0.58	1											
q_7	-0.08	0.13	0.07	0.12	0.30	0.31	-0.96	0.28	0.20	1										
OH	-0.25	-0.34	-0.33	-0.22	0.09	0.06	-0.04	-0.24	-0.10	0.05	1									
NH_2	-0.33	-0.28	-0.32	-0.28	0.12	-0.61	-0.10	-0.04	-0.18	0.05	-0.06	1								
NO_2	-0.03	-0.09	-0.15	0.14	0.05	0.05	0.02	0.05	-0.04	0.10	-0.03	-0.04	1							
X	0.53	0.01	0.05	-0.12	-0.43	0.14	0.13	-0.28	-0.11	-0.06	-0.12	-0.14	-0.08	1						
OCH_3	0.34	0.50	0.53	0.65	0.14	0.00	-0.19	0.64	0.76	0.04	-0.12	-0.14	-0.08	-0.27	1					
R_{orto}	0.43	0.65	0.64	0.60	0.12	0.24	-0.43	0.70	0.72	0.33	-0.20	-0.25	0.06	0.10	0.50	1				
R_{o-Cl}	0.38	0.08	0.11	0.02	-0.16	0.07	-0.12	-0.12	-0.02	0.16	-0.07	-0.08	-0.05	0.77	-0.16	0.39	1			
R_{aldh}	0.04	0.07	0.03	0.20	0.04	0.04	-0.03	0.24	-0.05	0.08	-0.03	-0.04	-0.02	-0.08	0.16	0.06	-0.05	1		
R_{keton}	0.21	0.26	0.28	0.54	0.07	0.07	0.21	0.23	0.16	-0.21	-0.06	-0.08	-0.04	-0.14	0.29	0.11	-0.08	-0.04	1	
R_{ester}	0.04	0.03	0.03	0.19	0.04	0.03	0.01	-0.13	-0.05	0.06	-0.03	-0.04	-0.02	-0.08	-0.08	-0.14	-0.05	-0.02	-0.04	1

Table 4. QSRR models derived for a set of substituted phenols (n=42) in GC and statistics: the correlation coefficient **R**; its square **R^2**; the standard deviation **s**; Fisher ratio value **F**; maximal residual value of deviation between the experimental and calculated retention data for corresp.solute, **Δmax$_i$** (**i.u.**).

	Calc. Method	Descriptors	R	R^2	s	F	Δmax$_i$
		I phase=SE-30					
1		CTI, OH, X, Cl, Br, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, R$_{ketone}$	0.971	0.942	45.3	58	115$_{32}$
2		CTI, OH, X, Cl, NH$_2$, NO$_2$, alk, R$_{o\text{-}Cl}$, R$_{o\text{-}Me}$, R$_{ketone}$, R$_{ester}$, R$_{o\text{-}NH2}$	0.975	0.950	44.0	46	91$_{32}$
3		α, OH, X, Cl, NH$_2$, R$_{o\text{-}Cl}$, R$_{o\text{-}Me}$, R$_{ketone}$, R$_{ester}$, R$_{aldeh}$	0.952	0.907	58.5	30	124$_{31}$
4		MR , OH, X, NH$_2$, OCH$_3$, R$_{orto}$, R$_{ketone}$, R$_{ald}$, R$_{ester}$	0.962	0.926	51.5	44	84$_{39}$
5		M, OH, X, Cl, NH$_2$, OCH$_3$, R$_{orto}$, R$_{ald}$, R$_{ketone}$,R$_{estr}$	0.961	0.923	53.2	37	92$_{26,.32}$
6	AM1	MR, E$_{homo}$,OH, X, Cl, NH$_2$, OCH$_3$, R$_{orto}$, R$_{ketone}$, R$_{ester}$,R$_{aldeh}$, R$_{o\text{-}Me}$	0.974	0.949	44.8	45	117$_{38}$
7	AM1	CTI $_{AM1}$, E$_{hydrat}$, OH, X, Cl, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, Q$_1$, Q$_2$, Q$_6$, Q$_7$	0.983	0.967	36.1	71	77$_{27}$
8	AM	CTI $_{AM1}$,OH, X, Cl, NH$_2$, R$_{o\text{-}Cl}$, R$_{o\text{-}NO2}$, Q$_1$, Q$_2$, Q$_6$, Q$_7$	0.980	0.960	38.9	66	78$_{27}$
9	AM1	CTI $_{AM1}$, OH, X, Cl, NH$_2$, R$_{o\text{-}Cl}$, R$_{o\text{-}NO2}$, R$_{vic}$, Q$_1$,Q$_2$,Q$_6$,Q$_7$	0.982	0.964	37.7	64	77$_{27}$
10$_1$	DFT	CTI$_{DFT}$,OH, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, E$_{tot}$, q$_2$	0.973	0.946	42.3	86	103$_{41}$
10$_2$	DFT	CTI$_{DFT}$,OH, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, E$_{tot}$, q$_2$	0.955	0.912	54.5	50	105$_{42}$
10$_3$	DFT	CTI$_{DFT}$, E$_{tot}$, OH, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, R$_{ketone}$, q$_1$,q$_2$, q$_6$,	0.968	0.937	48.2	46	97$_{16}$
11	DFT	MR, E$_{tot}$, OH, X, NH$_2$, OCH$_3$, R$_{orto}$, R$_{ketone}$, R$_{ald}$, R$_{ester}$, q$_2$, q$_6$,q$_7$	0.981	0.963	39.1	56	74$_{33}$
12	DFT	M, E$_{tot}$, OH, X, NH$_2$, OCH$_3$, R$_{orto}$, R$_{ketone}$, R$_{ald}$, R$_{ester}$, q$_2$, q$_6$,	0.973	0.947	45.6	43	101$_{26}$
13	DFT	α,OH,X,OCH$_3$,NH$_2$,R$_{orto}$, R$_{ketone}$,R$_{ester}$,R$_{aldeh}$,q$_6$	0.967	0.936	48.5	45	102$_{33}$
		II phase=OV-225					
1		CTI,OH, X, Cl, Br, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, R$_{o\text{-}Me}$, R$_{keton}$, R$_{ester}$,	0.947	0.898	105	24	261$_{32}$
2		CTI, OH, X, Cl, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, R$_{o\text{-}Me}$, R$_{keton}$, R$_{ester}$,	0.944	0.892	107	26	261$_{32}$
3		α, OH, X, Cl, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, R$_{o\text{-}Me}$, R$_{keton}$, R$_{ester}$, R$_{aldeh}$	0.934	0.873	118	18	201$_{16}$
4		MR,OH, X, NH$_2$, OCH$_3$, R$_{orto}$, R$_{keton}$, R$_{ester}$, R$_{aldeh}$,	0.931	0.867	117	23	229$_{35}$
5		M, OCH$_3$, X, OH, Cl, NH$_2$, R$_{orto}$, R$_{ald}$, R$_{keton}$, R$_{estr}$	0.943	0.889	108	25	237$_{32}$
6	AM1	MR,E$_{homo}$,OH,X,Cl,NH$_2$, NO$_2$,OCH$_3$, R$_{orto}$, R$_{ketone}$,R$_{ester}$, R$_{aldeh}$, R$_{o\text{-}Me}$	0.961	0.923	95	26	188$_{35}$
7	AM1	CTI $_{AM1}$, E$_{hydrat}$, OH, X, Cl, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, Q$_1$, Q$_2$, Q$_6$, Q$_7$	0.964	0.930	89	32	145$_{24}$
8	AM1	CTI $_{AM1}$, E$_{homo}$, OH, X, Cl, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, R$_{aldeh}$, Q$_1$, Q$_2$, Q$_6$, Q$_7$	0.980	0.960	68	52	104$_{27}$
9	AM1	CTI $_{AM1}$,E$_{homo}$,OH,X,Cl,NH$_2$,NO$_2$, alk, R$_{o\text{-}Cl}$, R$_{aldeh}$, Q$_1$, Q$_2$, Q$_6$, Q$_7$	0.981	0.962	67	49	103$_{27}$
10$_1$	DFT	CTI$_{DFT}$,OH, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, R$_{keton}$, E$_{tot}$, q$_1$, q$_2$, q$_6$, q$_7$	0.969	0.939	81	42	149$_{24}$
10$_2$	DFT	CTI$_{DFT}$, OH, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, R$_{keton}$, E$_{tot}$, q$_1$, q$_2$, q$_6$, q$_7$	0.940	0.883	112	21	209$_{13}$
11	DFT	MR,OH, X, NH$_2$, OCH$_3$, R$_{orto}$, R$_{keton}$, R$_{ald}$, R$_{ester}$, E$_{tot}$, q$_2$, q$_6$,q$_7$	0.975	0.951	75	42	185$_{10}$
12	DFT	M,OH, NH$_2$, R$_{orto}$, R$_{keton}$, R$_{ald}$, R$_{ester}$, E$_{tot}$, q$_2$, q$_6$,	0.965	0.931	85	42	200$_{10}$
13	DFT	α, OH, X, OCH$_3$, NH$_2$, R$_{orto}$, R$_{keton}$, R$_{ester}$, R$_{aldeh}$, E$_{tot}$, q$_2$, q$_6$,q$_7$	0.974	0.948	78	39	133$_{17}$
		III phase=NGA					
1		CTI,OH,X,Cl,Br,NH$_2$,NO$_2$, R$_{o\text{-}Cl}$,R$_{o\text{-}Me}$, R$_{keton}$,R$_{ester}$,	0.940	0.884	104	21	299$_{32}$
2		CTI,OH,X,Cl,NH$_2$,NO$_2$,alk,R$_{o\text{-}Cl}$,R$_{o\text{-}Me}$, R$_{keton}$,R$_{ester}$,R$_{o\text{-}alk}$, R$_{o\text{-}NH2}$	0.940	0.884	108	16	285$_{32}$
3		α, OH, X, Cl,NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, R$_{o\text{-}Me}$, R$_{keton}$, R$_{ester}$, R$_{ald}$	0.935	0.875	109	19	249$_{32}$
4		MR,OH, X, NH$_2$, R$_{orto}$,O-CH$_3$, R$_{keton}$, R$_{ester}$, R$_{aldeh}$	0.925	0.855	113	21	255$_{32}$
5		M, OH, X, Cl, NH$_2$, OCH$_3$, R$_{orto}$, R$_{ald}$, R$_{keton}$, R$_{estr}$	0.938	0.88	104	23	262$_{32}$
6	AM1	MR,E$_{homo}$,OH,X,Cl,NH$_2$, NO$_2$,OCH$_3$, R$_{orto}$, R$_{keton}$, R$_{ester}$, R$_{ald}$, R$_{o\text{-}Me}$	0.961	0.923	88	26	186$_{35}$
7	AM1	CTI$_{AM1}$, E$_{hydrat}$, OH, X, Cl, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, Q$_1$, Q$_2$, Q$_6$, Q$_7$	0.967	0.935	78	35	141$_{24}$
8	AM1	CTI$_{AM1}$, E$_{homo}$, OH, X, Cl, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, R$_{aldeh}$, Q$_1$, Q$_2$, Q$_6$, Q$_7$	0.982	0.965	59	60	89$_{27}$
9	AM1	CTI$_{AM1}$,E$_{homo}$, OH, X, Cl, NH$_2$, R$_{o\text{-}NO2}$, R$_{o\text{-}Cl}$, R$_{aldeh}$, Q$_1$, Q$_2$, Q$_6$, Q$_7$	0.982	0.965	59	60	115$_{10}$
10$_1$	DFT	CTI$_{DFT}$, E$_{tot}$, OH, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, R$_{keton}$, q$_1$, q$_2$, q$_6$, q$_7$	0.968	0.938	77	41	128$_{24}$
10$_2$	DFT	CTI$_{DFT}$, E$_{tot}$, OH, NH$_2$, NO$_2$, R$_{o\text{-}Cl}$, R$_{keton}$, q$_1$, q$_2$, q$_6$, q$_7$	0.944	0.892	101	22	186$_4$
11	DFT	MR, E$_{tot}$, OH, X, NH$_2$, OCH$_3$, R$_{orto}$, R$_{keton}$, R$_{aldeh}$, R$_{ester}$, q$_2$, q$_6$, q$_7$	0.973	0.948	73	39	121$_{3,17}$
12	DFT	M, E$_{tot}$, OH, NH$_2$, R$_{orto}$, R$_{keton}$, R$_{ald}$, R$_{ester}$, q$_2$, q$_6$,	0.964	0.930	80	40	189$_{26}$
13	DFT	α, E$_{tot}$, OH, X, OCH$_3$, NH$_2$, R$_{orto}$, R$_{keton}$, R$_{ester}$, R$_{aldeh}$, q$_2$, q$_6$, q$_7$	0.972	0.945	75	37	161$_{10}$

The pool of regressions with CTI index as a global descriptor was derived for the three phases and in the majority cases these regressions exhibit the best (equal) statistics. This was the reason to mainly list the models with the CTI index (AM1 calculations) in Table 4.

The set of equations (1–5) from Table 4 can be easily developed because it needs the simplest molecular parameterization just with additive global molecular indices (M, MR, etc.) and adding of structural substituents. The QSRRs of this type have good statistics and provide interesting information about stationary phase selectivity.

The full form for the QSRR equations № 4 from Table 4 with MR global descriptor can be seen in Table 5–2. The numeric values of the regression coefficients for the indicative structural descriptors in the similar equations from Table 5–2 valid for the three phases are graphically compared on Fig. 2. It is clearly seen on Fig. 2 that the contribution to the retention on the unipolar phase SE–30 is certainly smaller than that on polar phases for all structural substituents. The McReynolds Y' and H' indices are used to quantitate the phase selectivity towards the presence of alcohol, acidic and amide fragments in solute structures. The distribution of the values of McReynolds Y' and H' indices and their dependence on phase polarity on Figure 3 has the similar profile on varying the sets of regression coefficients for $-OH$, $-NH_2$ and $-R_{o-OH}$ descriptors from Figure 2.
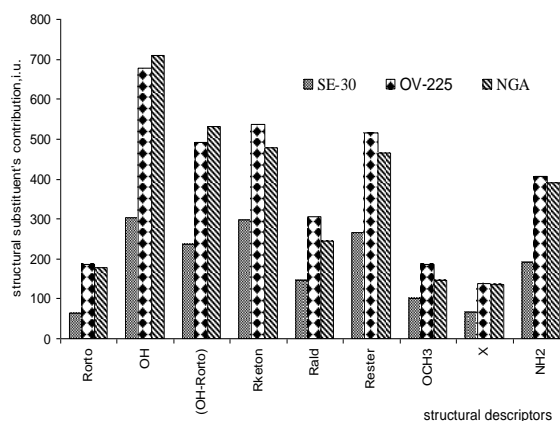
Figure 2. The contribution to the retention of structural substituents in QSRR models (equations 4, Table 5) developed for GC separation of substituted phenols on SE-30, OV-225 and NGA stationary phases.

Table 5-1. The QSRR models for a set of substituted phenols (n = 42) and statistics

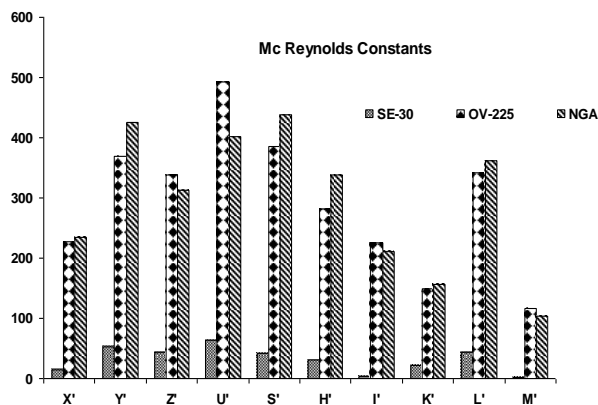| St. Phase	Equations 6 (Table 4)	Statistics
SE-30	$RI = -546.2 (\pm 427.1) + (21.2 \pm 2.1)\mathbf{MR} - (104.5 \pm 43.3)\mathbf{E_{Homo}} - ----(85.2 \pm 20.3)\mathbf{R_{orto}} + (261.7 \pm 34.3)\mathbf{R_{keton}} + (123.0 \pm 48.2)\mathbf{R_{ald}} + (218.4 \pm 53.7)\mathbf{R_{ester}} + (137.9 \pm 21.3)\mathbf{OCH_3} + (110.5 \pm 31.4)\mathbf{X} + (348.7 \pm 38.5)\mathbf{OH} + (305.0 \pm 50.4)\mathbf{NH_2} - (52.5 \pm 30.3)\mathbf{Cl} + (48.6 \pm 24.6)\mathbf{R_{o-Me}}$	$R = 0.97$ $R^2 = 0.95$ $F = 44.9$ $S = 44.8$ $\Delta max_i = 117_{38}$
OV-225	$RI = -1673.3 (\pm 1439.8) + (23.7 \pm 4.4)\mathbf{MR} - (286.6 \pm 153.9)\mathbf{E_{Homo}} - (230.7 \pm 44.6)\mathbf{R_{orto}} + (433.5 \pm 91.6)\mathbf{R_{keton}} + (237.5 \pm 111.5)\mathbf{R_{ald}} + (385.6 \pm 138.6)\mathbf{R_{ester}} + (282.8 \pm 45.1)\mathbf{OCH_3} + (262.0 \pm 45.1)\mathbf{X} + (788.4 \pm 81.9)\mathbf{OH} + (706.6 \pm 143.1)\mathbf{NH_2} - (149.6 \pm 65.4)\mathbf{Cl} + (137.2 \pm 52.5)\mathbf{R_{o-Me}} - (211.8 \pm 184.8)\mathbf{NO_2}$	$R = 0.96$ $R^2 = 0.92$ $F = 25.9$ $S = 94.6$ $\Delta max_i = 188_{35}$
NGA	$RI = -1749.3 (\pm 1341.6) + (21.2 \pm 4.1)\mathbf{MR} - (321.2 \pm 143.4)\mathbf{E_{Homo}} - (224.6 \pm 441.6)\mathbf{R_{orto}} + (360.1 \pm 85.4)\mathbf{R_{keton}} + (165.0 \pm 103.9)\mathbf{R_{ald}} + (314.1 \pm 129.1)\mathbf{R_{ester}} + (246.7 \pm 42.0)\mathbf{OCH_3} + (242.0 \pm 74.1)\mathbf{X} + (826.1 \pm 76.3)\mathbf{OH} + (718.9 \pm 133.4)\mathbf{NH_2} - (139.9 \pm 60.9)\mathbf{Cl} + (142.4 \pm 48.9)\mathbf{R_{o-Me}} - (250.9 \pm 172.2)\mathbf{NO_2}$	$R = 0.96$ $R^2 = 0.92$ $F = 25.9$ $S = 88.1$ $\Delta max_i = 186_{35}$

Table 5-2. The QSRR models for a set of substituted phenols (n = 42) and statistics

St. Phase	Equations 4 (Table 4)	Statistics
SE-30	$RI = 487.5(\pm74.1) + (19.0\pm2.1)\mathbf{MR} - (65.0\pm14.8)\mathbf{R_{orto}} + (298.9\pm33.2)\mathbf{R_{keton}} + (146.8\pm53.2)\mathbf{R_{ald}} + (264.5\pm53.9)\mathbf{R_{ester}} + (100.6\pm16.4)\mathbf{OCH_3} + (65.3\pm10.8)\ \mathbf{X} + (302.7\pm40.8)\mathbf{OH} + (191.6\pm33.4)\mathbf{NH_2}$	R = 0.96 R^2 = 0.93 F = 44.2 S = 51.5 $\Delta max_i = 84_{39}$
OV-225	$RI = 1109.2(\pm167.8) + (19.2\pm4.7)\mathbf{MR} - (185.4\pm33.5)\mathbf{R_{orto}} + (535.6\pm75.1)\mathbf{R_{keton}} + (306.1\pm120.5)\mathbf{R_{ald}} + (514.3\pm122.2)\mathbf{R_{ester}} + (186.9\pm37.1)\mathbf{OCH_3} + (138.8\pm24.48)\ \mathbf{X} + (677.1\pm92.5)\mathbf{OH} + (406.5\pm75.5)\mathbf{NH_2}$	R = 0.93 R^2 = 0.87 F = 23.1 S = 116.7 $\Delta max_i = 229_{35}$
NGA	$RI = 1358.3(\pm162.8) + (16.2\pm4.5)\mathbf{MR} - (176.8\pm32.7)\mathbf{R_{orto}} + (477.7\pm72.9)\mathbf{R_{keton}} + (245.7\pm116.9)\mathbf{R_{ald}} + (465.1\pm118.6)\mathbf{R_{ester}} + (145.4\pm36.0)\mathbf{OCH_3} + (134.8\pm23.7)\ \mathbf{X} + (709.1\pm89.7)\mathbf{OH} + (389.9\pm73.3)\mathbf{NH_2}$	R = 0.92, R^2 = 0.85 F = 20.9 S = 113.2 $\Delta max_i = 255_{32}$

The variation of McReynolds parameters for the same stationary phases and the dependence of these contributions on the polarity of the stationary phase are presented on Figure 3.

The third McReynolds index Z′ is used for phase selectivity evaluation towards ketone, aldehyde, amine and ester fragments in solute structures. In the case of our models from Table 5-2, the ketone, aldehyde, OCH_3, $-NH_2$ and ester fragments are included as separate descriptors, but for all these descriptors the similar profile of retention contribution is created from QSRRs regression coefficients values.

The phase selectivity towards the halogen substitutes is quantified with McReynolds seventh I′ index. The character of the changes in the I′ index on Figure 3 and our descriptor (X) on Figure 2 is similar. The maximum values for the retention contribution of iodine–containing structures are typical for a OV–225 phase.

The tendency in contribution of structural indicative descriptors to the retention expressed by McReynolds indices and the variation in regression coefficients values derived from the QSRR models depends on stationary phase polarity and seems to be influenced in the similar way. The comparison of the parametric (numeric) values of the regression coefficients for structural substituents used as indicative descriptors in similar QSRR models derived for different stationary phases provide information about the polarity of the phases used in the case, similar to McReynolds constants.

Apparently in the case when the MLR form for QSRR modeling is preferred to quantify the retention behavior for a set of diverse noncongeneric solutes, separation on stationary phases with different polarities, the parametric values for these indicative descriptors reveal the stationary phase polarity in a similar way as McReynolds constants.

4. CONCLUSIONS

QSRR approach was applied for gas chromatographic retention modeling for a set of phenol derivates (n = 42) separated on three stationary phases with different polarity. The molecular indices, used as descriptors in MLRs were calculated at AM1 and DFT level. The charge–related topological index CTI, probed as a global descriptor for QSRRs deriving, revealed its ability to provide models with good statistics.

It was shown that the descriptors contribution analyses can be used to reveal the intermolecular solute – phase interactions. The numeric values of

Figure 3. The *McReynolds* constants for SE-30, OV-225 and NGA stationary phases.

the regression coefficients were compared with McReynolds constants. The comparison of the regression coefficient contribution of the structural descriptors in similar QSRR models for different stationary phases revealed the stationary phase selectivity.

REFERENCES

1. Sampling and Analysis Procedures for Screening of Industrial Effluents for Priority Pollutants, USEPA., Envir. Monitoring and Support Labor., Cincinnati, (1977).
2. WHO, World Health Organization Environmental Health Criteria, Geneva, No.202, 1998.
3. B.K. Chen, Cs. Horvath, *J. Chromatogr.,* 171 (1979), 15.
4. R. Kaliszan, *Chem. Rev.*, 107, 3212 (2007).
5. K. Heberger, *J. Chromatogr.* A, 1158, 273 (2007).
6. R. Kaliszan, QSRR: Quantitative Structure-Chromatographic Retention Relationships, Willey, New Jork, 1987.
7. R. Kaliszan, Structure, Retention in Chromatography. A Chemometric Approach, Harwood Academic, Amsterdam, 1997.
8. S. Eric, M. Pavlovic, G. Popovic, D. Agbaba, *J. Chromatogr. Sci.*, 45, 140 (2007).
9. W.L. Fitch, M. McGregor, A.R. Katritzky, A. Lomaka, R. Petruhin, M. Karelson, *J. Chem. Inf. Comput. Sci.,* 42, 830 (2002).
10. A.R. Katritzky, R. Petruhin, D. Tatham, S. Basak, E. Benfenati, M. Karelson, U. Maran, *J. Chem. Inf. Comput. Sci.,* 41, 675 (2001).
11. P. Gramatica, P. Pilutti, E. Papa, *SAR QSAR Environ. Res.* 13, 743 (2002).
12. M.R. Hadjmohammadi, M.H. Fatemi, K. Kamel, *J. Chromatogr.Sci.*, 45, 400 (2007).
13. C.K. Roy, I. Sanyal, P.P. Roy, *SAR QSAR Environ. Res.* 17, 563 (2006).
14. D.K. Tuppurainen, S.P.Korhonen, J. Ruuskanen, *SAR QSAR Environ. Res.* 17, 549 (2006).
15. V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
16. R. Burbidge, M. Buxton, S. Holden, *Comput. Chem.* 26, 5 (2001).
17. R. Cherminski, A. Yasri, D. Hartsourgt, *Quant. Struct.-Act.Relat.* 20, 227 (2001).
18. V.C. Viterbo, J.P. Braga, A.P. Braga, M.B. de Almeida, *J. Chem. Inf. Comput. Sci,* 41, 309 (2001).
19. G.A. Bakken, P.C. Jurs, *J. Med. Chem.*, 43, 4534 (2000).
20. J. Sorich, J.O. Miners, R.A. McKinnon, D.A. Winkler, F.R. Burden, P.A. Smith, *J.Chem. Inf. Comput.Sci.* 43, 2019 (2003).
21. A.R. Katritzky, R. Petruhin, R. Jain, M. Karelson, *J. Chem. Inf. Comput. Sci,* 41, 1521 (2001).
22. S. Dietmann, C.Frőmmel, *Bioinformatics,* 18, 167 (2002).
23. S. Sremac, B. Skrbic, A. Onjia, *J. Serb. Chem. Soc.,* 70, 1291 (2005).
24. S. Sremac, A. Popovic, Z. Todorovic, D. Cokesa, A. Onjia, *Talanta*, 76, 66 (2008).
25. H.F. Chen, *Anal. Chim. Acta,* 609, 24 (2008).
26. W.O. McReynolds, *J. Chromatogr. Sci.* 8, 685 (1970).
27. L. Rohrschneider, *J. Chromatogr.* 48, 728 (1998).
28. M.H. Abraham, C.F. Poole, S.K. Poole, *J. Chromatogr.* A, 842, 79, (1999).
29. C.F. Poole, S.K. Poole, *J. Chromatogr.* A, 1184, 254,(2008).
30. C.F. Poole, The Essence of Chromatography, Elsevier, Amsterdam, 2003.
31. S.K. Poole, B.R. Kersten, C.F. Poole, *J. Chromatogr.,* 471 (1989), 91.
32. C.F. Poole, S.K. Poole, *Chem. Rev.*, 89, 377 (1989).
33. L. Rohrschneider, *J. Sep. Sci.* 24, 3 (2001).
34. L. Rohrschneider, *J. Chromatogr.* 17, 1 (1965).
35. L. Rohrschneider, *J. Chromatogr.* 22, 6 (1966).
36. H. Rotzsche, Stationary Phases in Gas Chromatography, Elsevier, Amsterdam, 1991.
37. J. Grzibowski, H. Lamparczyk, A. Nasal, A. Radecki, *J. Chromatogr.*, 196, 217 (1980).
38. I.P. Bangov, *J. Chem. Inf. Comput. Sci.*, 30, 277 (1990).
39. P.A. Demirev, A.S. Dyulgerov, I.P. Bangov, *J. Math. Chem.,* 8, 367 (1991).
40. I.P. Bangov, M. Moskovkina, A. Patleeva, *Bulg. Chem. Commun.*, 42 (4), 338 (2010).
41. MS Excel, Version 2003, Microsoft Corp., Redmond, WA, 2003.
42. Alex A. Granovsky, PC GAMESS version 7.1(Tornado), http://classic.chem.msu.su/gran/gamess/index.htm
43. 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, S. Su, T.L. Windus, M. Dupuis, J.A. Montgomery. *J.Comput. Chem.* 14, 1347 (1993).
44. M. Charton, S.Clementi, S.Ehrenson, O.Exner, J. Shorter and S.Wold, *Quant. Struct. Act. Relat.*, 4, 29 (1985).
45. R. Kaliszan, *CRC Critical Reviews in Analytical Chemistry*, 16(4), 323 (2005).
46. G.R. Faminy, L. I. Wilson, *Rev. Comp. Chem.*, 211 (2002).

МОДЕЛИРАНЕ НА ГАЗХРОМАТОГРАФСКО ЗАДЪРЖАНЕ С ПОМОЩТА НА ХЕМОМЕТРИЧЕН ПОДХОД

М.Н. Московкина, И.П. Бангов, А.Ж. Патлеева

Катедра „Обща химия", Факултет по природни науки, Шуменски Университет „Еп. К. Преславски", Шумен,, 9712, ул. Университетска 115, България

(Резюме)

Хемометричният подход за установяване на количествени връзки от типа „структура-ретентно хроматографско свойство" (QSRR) е приложен за група от заместени феноли с цел създаване на модели за задържането им в газхроматографска колона. Експерименталните данни за група от 42 заместени фенола, включваща и приоритетни природни замърсители, разделени в три колони с различна хроматографска полярност (неполярна фаза SE–30 и полярни фази OV–225 и NGA), са взети от литературата. За създаването на моделите е използван статистическия метод на многопараметричната линейна регресия (MLR). Тополого-електронният индекс CTI, създаден от един от авторите (И.Б.) е проверен и оценен като глобален дескриптор. Хемометричния анализ на създадените модели ясно показва факторите, описващи междумолекулните взаимодействия (ММВ) с неподвижната фаза и обуславящи разделянето в хроматографската колона. Сравнението на приносите на отделните дескриптори в регресионните уравнения за различни неподвижни фази, показва, че те корелират добре с константите на МакРейнолдс за фазова селективност.