sciendo

ASN, Vol 5, No 2, Pages 77-89, 2018

Acta Scientifica Naturalis
Former Annual of Konstantin Preslavsky University – Chemistry, Physics, Biology, Geography
Journal homepage: http://www.shu.bg

Received: 09.2018

Accepted: 11.2018

CHROMATOGRAPHIC RETENTION MODELING OF ALKYLAZOLES BY QSRR APPROACH

M. N. Moskovkina[1], I. P. Bangov

Faculty of Natural Sciences, Department of Chemistry, "Konstantin Preslavsky" University of Shumen, 9712 Shumen, Universitetska Str, 115, Bulgaria

m.moskovkina@shu-bg

***Abstract**: The Quantitative Structure Retention Relationship (QSRR) approach has been applied to model the gas chromatographic retention of 16 alkyloxazoles and 16 alkylthiazoles on three capillary columns with different polarities. The potential of the Charge-related Topological Index (CTI) developed by one of the authors (I.B.) was investigated as a descriptor in QSRR linear multivariate regressions. Calculated values of atomic charges and the indication of the presence of substitutions in different positions in the solute structures are used to generate regressions. Analysis of the equations derived proves their ability to describe and evaluate the participants in the chromatographic separation process. The present quantitative characterization of the chromatographic retention of alkylazoles shows the potentials of deriving QSRR models to exhibit the retention intermolecular interactions.*

Keywords*: alkyazoles, molecular indices, QSRR, retention modeling.*

Introduction

Alkylazoles are widely used in chemistry, medicine and pharmacist science and their analytical problems are the matter of a constant interest. Among the most common analytical methods for the purpose are different chromatographic techniques. Retention mechanisms of various chromatographic modes have to be quantitatively investigated by computational chemistry. Due to its rapid development in the last time, new methods of studying these intermolecular interactions often appear. Differences in retention mechanisms of hydrophobic (reversed phase) and aqueous hydrophilic interactions, as well as the jon-exchange liquid chromatographic methods are explained in the papers [1, 2]. The retention mechanism of hydrophilic-interaction liquid chromatography (HILIC)-mode liquid chromatography is reviewed [3-9].

The goal of the **Q**uantitative **S**tructure **R**etention **R**elationship (QSRR) approach is to exhibit the influence of the participants in chromatographic separation process by means of formal mathematical procedure. The models derived usually include a set of numeric molecular indices to quantitate nonspecific (dispersive) and specific (polar) interactions between the solutes and chromatographic phases. Finely the chromatographic retention data is presented in terms of chemoinformatics in the form of multivariable linear regression (MLR):

$$\mathrm{RI} = \sum a_i D_i + \sum b_i P_i + c\,, \qquad (1)$$

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023

Where RI is the experimental retention data (most often the Retention Index of *Kovats*, [10]): P_i and D_i are the molecular descriptors for polar and dispersive interactions; a_i, b_i and c are constants. Usually the quantitation of nonspecific dispersive molecular interactions can be successfully provided by global "bulk" molecular indices. In contrast, the various specific polar interactions can be quantitated less precisely by numeric local molecular indices.

The reason to choice MLR is because the regression coefficients has the physical meaning and indicates the quantity contribution of the particular parameter D_i or P_i. The statistical treatment of the regressions received makes it possibly to evaluate the significance of both the whole equation and of a single parameter**.** The possibility to refine the relationships between the retention solute characteristics and their molecular properties makes the QSRR approach inquisitive. The aim is to derive mathematical equations with adequate accuracy to experimental retention data. In this case one can use some theoretically derived values instead of experimental one in order to predict a solute retention property or for identification purposes.

The QSRR approach has been applied to model the gas chromatographic retention index of *Kovats* for the set of 16 alkyloxazoles and 16 alkylthiazoles separated on three stationary phases with different polarities. Some of our results of QSRR modeling for the same two groups of azoles have been published previously [11-13]. The present investigation continues our efforts to derive QSRR regressions with the topological index *C*harge-related *T*opological *I*ndex (CTI) [14-16] as a global molecular descriptor and been further tuned with some local molecular descriptors:

$$CTI = 1/2 \sum_i \sum_j \frac{L_i L_j}{D_{ij}}, \qquad (2)$$

where Li and L_j are local (atomic) charge-related indices in the form of:

$$Li = n_v + q - N_H. \qquad (3)$$

Here n_v is the corresponding atom valence, q is the atomic charge density and N_H is the number of the attached hydrogen atoms to this heavy atom. The D_{ij} parameters are the topological distances (the number of bonds) between the two atoms.

A comparison of the common equations accuracy and of the significance levels for each single descriptor reflects the specific features of chromatographic separation mechanism.

Experimental

Multiparametric linear regressions (MLR) equations were derived by using expression (1). The experimental data for the *Kovats* retention indices of alkylazoles separated on three stationary phases with different polarities: OV-101, Triton X-305, and PEG-40M, have been taken from literature [17] and shown in Table 1.

A common approach in QSRR is that the molecular descriptors are grouped as parameters to quantitate the ability of a solute to participate either in nonspecific dispersive or in specific polar interactions within a stationary phase. The dispersive forces can be accurately quantified by any of the global molecular descriptors – constitutional, topological or steric one. Hence, in the present investigation we checked the ability of the *CTI* to be used as a global descriptor D [18]. The polar interactions were presented by local molecular descriptors: the charges $\boldsymbol{q_i}$ of the atoms in the azole's ring and the descriptors $\boldsymbol{R_i}$, which indicate the presence of the alkyl substituent in the solute structure. In the case of presence, the indicative descriptor has the value 1 and it is 0 when there is no substituent in the structure.

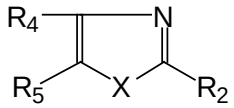
Table 1. Retention Indices of Kovats for a set of oxazoles and thiazoles separated on different stationary phases-OV-101, Triton_X-305, PEG-40M.

Compound[a]	Structure descriptors	*Kovats* Retention Indices, $\mathbf{RI_{phase}}$,	*Kovats* Retention Indices, $\mathbf{RI_{phase}}$,

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023

		R_i		Oxazoles			Thiazoles		
	X-azole	**R_4**	**R_5**	**OV-101**	**Triton-X-305**	**PEG-40M**	**OV-101**	**Triton-X-305**	**PEG-40M**
1	2,4 di-Me-	1	0	730	1043	1094	887	1255	1291
2	2,5di-Me-	0	1	765	1092	1145	922	1267	1324
3	2Et,4Me-	1	0	818	1104	1159	970	1285	1349
4	2Me,4Et-	1	0	820	1111	1171	974	1291	1356
5	2,4,5tri-Me-	1	1	843	1150	1200	997	1319	1380
6	2Et,5Me-	0	1	851	1155	1204	1004	1325	1388
7	2Me5Et-	0	1	855	1162	1222	1010	1337	1403
8	2Pr,4Me-	1	0	901	1175	1229	1053	1357	1418
9	2,4diEt-	1	0	903	1172	1228	1053	1343	1409
10	2Me,4Pr-	1	0	910	1198	1247	1064	1380	1428
11	2,4di-Me,5Et-	1	1	923	1201	1252	1072	1380	1438
12	2Et,4,5di-Me-	1	1	926	1210	1260	1077	1381	1439
13	2,5di-Et-	0	1	940	1222	1279	1090	1402	1463
14	2,4di-Me,5Pr-	1	1	1000	1269	1319	1157	1457	1515
15	2Pr,5Et-	0	1	1024	1294	1347	1175	1475	1534
16	2Et,4Me,5Pr-	1	1	1079	1327	1374	1233	1512	1567



[a] General structure for alkykoxazoles (X=O) and alkylthiazoles (X=S)

R_2 = Me, Et, Pr; **R_4** = H, Me, Et, Pr; **R_5** = H, Me, Et, Pr

The atomic charges have been calculated by the computer programs *HyperChem* (AM1 method) and the *CTI* indices being calculated by the *StrMngr* program developed by the author (I.B.). The calculated molecular indices values used as descriptors for the QSRR computations are shown in Table 2.1 (for oxazoles) and Table 2.2 (for thiazoles).

Table 2.1. Calculated molecular indices for Oxazoles

№	**Oxazole**	**CTI[a]**	**Bal J[b]**	**q_1 [c]**	**q_2**	**q_3**	**q_4**	**q_5**
1_{ox}	2,4 di-Me-Ox	112.5824	2.26	-0.128	0.004	-0.141	-0.112	-0.13
2_{ox}	2,5di-Me- Ox	112.1732	2.26	-0.13	0.001	-0.142	-0.173	-0.066
3_{ox}	2Et,4Me-Ox	120.2383	2.22	-0.13	0.005	-0.140	-0.114	-0.129
4_{ox}	2Me,4Et-Ox	120.6052	2.22	-0.14	0.003	-0.129	-0.127	-0.113
5_{ox}	2,4,5tri-Me-Ox	122.7005	2.39	-0.129	0.003	-0.139	-0.113	-0.072
6_{ox}	2Et,5Me-Ox	120.1239	2.22	-0.13	0.001	-0.141	-0.174	-0.067
7_{ox}	2Me5Et-Ox	119.7718	2.22	-0.127	0	-0.143	-0.172	-0.066
8_{ox}	2Pr,4Me-Ox	126.8265	2.14	-0.129	0.007	-0.141	-0.115	-0.129
9_{ox}	2,4di-Et-Ox	127.9865	2.21	-0.129	0.003	-0.138	-0.114	-0.128
10_{ox}	2Me,4Pr-Ox	126.6353	2.14	-0.129	0.003	-0.140	-0.113	-0.127
11_{ox}	2,4di-Me,5Et-Ox	130.2556	2.38	-0.128	0.003	-0.141	-0.112	-0.07
12_{ox}	2Et,4,5di-Me-	131.4146	2.35	-0.123	0.003	-0.144	-0.112	-0.075

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023

	Ox							
13 $_{ox}$	2,5di-Et-Ox	128.0285	2.21	-0.126	0	-0.146	-0.171	-0.068
14 $_{ox}$	2,4di-Me,5Pr-Ox	136.567	2.31	-0.134	0.003	-0.142	-0.11	-0.066
15 $_{ox}$	2Pr,5Et-Ox	134.1632	2.15	-0.131	0.001	-0.141	-0.173	-0.065
16 $_{ox}$	2Et,4Me,5Pr-Ox	144.1772	2.32	-0.129	0.003	-0.139	-0.113	-0.07

a – CTI -Charge-Relative Topology index; b- Bal J - Balaban topology index, c- q_i – atomic chareges (AM1)

Table 2.2. Calculated molecular indices for Thiazoles

№	Compound [a]	CTI[a]	Bal J [b]	q_1 [c]	q_2	q_3	q_4	q_5
1 $_{thia}$	2,4 di-Me-Th	117.746	2.26	0.471	-0.268	-0.106	-0.07	-0.468
2 $_{thia}$	2,5di-Me-Th	112.317	2.26	0.455	-0.268	-0.108	-0.138	-0.379
3 $_{thia}$	2Et,4Me-Th	120.133	2.22	0.471	-0.268	-0.104	-0.072	-0.468
4 $_{thia}$	2Me,4Et-Th	120.047	2.22	0.466	-0.266	-0.106	-0.072	-0.464
5 $_{thia}$	2,4,5tri-Me-Th	120.74	2.39	0.454	-0.266	-0.106	-0.075	-0.387
6 $_{thia}$	2Et,5Me-Th	120.306	2.22	0.455	-0.268	-0.106	-0.138	-0.379
7 $_{thia}$	2Me5Et-Th	119.876	2.22	0.460	-0.270	-0.108	-0.137	-0.38
8 $_{thia}$	2Pr,4Me -Th	126.722	2.14	0.461	-0.260	-0.106	-0.074	-0.461
9 $_{thia}$	2,4diEt-Th	129.884	2.21	0.472	-0.266	-0.109	-0.072	-0.466
10 $_{thia}$	2Me,4Pr-Th	126.575	2.14	0.466	-0.267	-0. 106	-0.071	-0.461
11 $_{thia}$	2,4di-Me,5Et -Th	130.21	2.38	0.461	-0.269	-0.105	-0.075	-0.385
12 $_{thia}$	2Et,4,5di-Me -Th	126.85	2.35	0.462	-0.266	-0.109	-0.075	-0.393
13 $_{thia}$	2,5di-Et-Th	127.778	2.21	0.456	-0.267	-0.112	-0.136	-0.381
14 $_{thia}$	2,4di-Me,5Pr -Th	136.539	2.31	0.426	-0.260	-0.112	-0.067	-0.365
15 $_{thia}$	2Pr,5Et-Th	134.095	2.15	0.448	-0.266	-0.107	-0.138	-0.375
16 $_{thia}$	2Et,4Me,5Pr -Th	144.114	2.32	0.461	-0.269	-0.104	-0.07	-0.468

a – CTI -Charge-Relative Topology index; b- Bal J - Balaban topology index, c- q_i – atomic charges (AM1).

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023

Results and discussion

The retention on nonpolar methyl silicone phase OV-101 is governed by dispersive interactions. The second phase Triton X-305 contains the alkylenoxide – derivate structure fragments H-(O-CH_2-CH_2)$_m$-O-C_6H_4-C_8H_{17} and has a middle polarity; its retention behavior is due to both dispersive and polar interactions. The polyethyleneglycol phase PEG-40M is a polar phase with HO-CH_2-CH_2-(O-CH_2-CH_2)$_n$-O-CH_2-CH_2-OH structural fragments and the dominate type to the polar intermolecular interactions are the H- bonds with O atoms.

There are three sites any alkyl substituent in the alkylazole's structure. Any substituent R_i is in an α-position with respect to one or two heteroatoms N, O (resp.S): R_4 proves α-N effects; R_5 proves α-O (or α-S) effects; R_2 exhibits both $R_{\alpha\text{-N,O}}$, resp. $R_{\alpha\text{-N,S}}$ influences. The substituent in any α-X position has a possibility to manifest inductive, resonance or steric effect; which one is preferred depends on the solvent nature.

We started the QSRR deriving with two parameter regression equations (Tables 4.1.1-7).

Tables 3. The cross-correlation matrix for Oxazole and Thiazole descriptors

	Iov-101	*IX-305*	*Ipeg40M*	*CTI*	*Bal J*	*q1*	*q2*	*q3*	*q4*	*q5*	*Rvic*	*R4*	*R5*
Iov-101	1												
IX-305	0.991	1											
Ipeg40M	0.991	0.999	1						Oxazoles				
CTI	0.975	0.948	0.945	1									
Bal J	0.049	0.070	0.050	0.179	1								
q1	0.075	0.105	0.092	0.068	0.192	1							
q2	-0.092	-0.186	-0.199	0.058	-0.018	-0.106	1						
q3	-0.173	-0.230	-0.221	-0.095	-0.071	-0.802	0.281	1					
q4	0.082	-0.002	-0.021	0.283	0.367	-0.037	0.789	0.302	1				
q5	0.389	0.485	0.483	0.299	0.465	0.162	-0.702	-0.404	-0.528	1			
Rvic	-0.036	-0.002	-0.016	0.054	0.585	0.374	0.067	-0.114	0.275	0.222	1		
R4	0.045	-0.043	-0.059	0.249	0.330	-0.133	0.801	0.415	0.991	-0.566	0.255	1	
R5	0.399	0.494	0.488	0.321	0.516	0.292	-0.652	-0.500	-0.468	0.988	0.293	-0.522	1

	Iov-101	*IX-305*	*Ipeg40M*	*CTI*	*Bal J*	*q1*	*q2*	*q3*	*q4*	*q5*	*R4*	*R5*
Iov-101	1											
IX-305	0.989	1										
Ipeg40M	0.995	0.995	1						Thiazoles			
CTI	0.952	0.938	0.930	1								
Bal J	0.051	0.058	0.051	0.115	1							
q1	-0.454	-0.503	-0.510	-0.338	-0.200	1						
q2	0.166	0.159	0.177	0.109	-0.338	-0.171	1					
q3	-0.182	-0.204	-0.217	-0.104	0.017	0.536	0.019	1				
q4	0.035	-0.015	-0.041	0.254	0.307	0.242	-0.078	0.270	1			
q5	0.422	0.476	0.491	0.247	0.442	-0.726	-0.083	-0.379	-0.569	1		
R4	0.045	-0.008	-0.033	0.260	0.330	0.267	-0.084	0.309	0.997	-0.557	1	
R5	0.403	0.452	0.467	0.235	0.516	-0.631	-0.132	-0.317	-0.542	0.988	-0.522	1

The comparison of the parametric values of coefficients in the similar equations both for oxazoles and thiazoles can be used to reflect the differences in the intermolecular interactions between the solutes and the stationary phase. We have checked all possible combinations of local molecular descriptors and

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023

the *CTI* index.

It can be seen that q_i and R_i descriptors being in different positions in the molecular cycle have a different distribution in the models, different influence in QSRR modelling, respectively different participation in chromatographic separation process.

Tables 4. The regression coefficients values $\mathbf{a_0}$, $\mathbf{b_i}$, $\mathbf{c_i}$ and statistics (**p, t**) in various equations $\mathbf{RI=a_0 + \sum b_i D_i + \sum c_i P_i}$ for oxazoles / *thiazoles*

4.1.1.	**RI=a₀ + bCTI + c₁q₁**			
Phase	**R**	**a₀**	**b**	**c₁**
OV-101 thiazoles	*0.963*	*316.71 ±356.01* *t = 0.9; p =0.39*	*10.27 ±0.91* *t = 11.3; p = 4.10⁻⁸*	*-1225.9 ±654.67* *t =-1.8; p =0.08*
X-305 thiazoles	*0.959*	*992.87 ±303.59* *t = 3.3; p = 6.10⁻³*	*8.02 ±0.77* *t = 10.3;p=1.10⁻⁷*	*-1399.2 ±558.5* *t = -2.5; p =0.02*
PEG-40 thiazoles	*0.953*	*1095.97 ±326.9* *t = 3.3; p =5.10⁻³*	*7.99 ±0.83* *t = 9.6; p=3.10⁻⁷*	*-1488.5 ±601.4* *t = -2.4; p =0.03*

The adequate QSRR models with $\mathbf{q_1}$ as a polar descriptor are meaningful just for thiazoles, but not for oxazoles (Table.4.1.1). The sulfur atom in the thiazole has a negative correlation with RI_{phase}. its contribution increases in the polar phases. The inclusion of the sulfur atom in the regression equation results in a decrease of retention, and this effect is enhanced by an increase of phase polarity. It is possible in this case to take into account the influence of the mesomeric effect of the sulfur atom which leads to a decrease in the basicity of the nitrogen atom, and hence its ability to participate in polar interactions, especially donor-acceptor.

The adequate QSRRs with participation of $\mathbf{q_2}$ were received only for oxazoles **(**Table 4.1.2). The carbon atom in C-(2) position exhibits both $R_{\alpha\text{-}N,O}$, resp. $R_{\alpha\text{-}N,S}$ influences. The sign of the c_1q_2 term in the model is negative; its inclusion in the MLR equations for oxazoles decreases the retention, especially in the polar phase.

4.1.2.	**RI=a₀ + bCTI + c₁q₂**			
Phase	**R**	**a₀**	**b**	**c₁**
OV-101 oxazoles	0.986	- 429.89 ±63.41 t = -6.8; p =1.10⁻⁵	10.67 ±0.50 t = 21.2; p = 1.10⁻¹¹	-7565.19 ±2354.14 t =-3.2; p =6.10⁻³
X-305 oxazoles	0.979	134.11 ±63.64 t = 2.1; p = 0.06	8.52 ±0.50 t = 16.9;p=3.10⁻¹⁰	-10042.81 ±2363.07 t = -4.2; p =9.10⁻⁴

The presence of alkyl substituents in o-position affects both the inductive effect towards the N-atom and the steric hindrance of the N-atom. In the structures with methyl substituent as a R_2, the effect of hyperconjugation takes place. In all the cases a substituent at the o-position decreases the partition of the polar interactions with the stationary phases. The influence is particularly sensible in the polar phases.

The regression equations with the $\mathbf{q_3}$ descriptor are shown in Tabl.4.1.3. The nitrogen atom in the third position is a nucleophilic center, which leads to take part in polar interactions and its contribution increases in polar phases. Taking into account that the $\mathbf{q_3}$ atomic charge's own value is

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023

negative, one can see that the involving of the q_3 into equation contibutes the phase retention depiction with 310-315 i.u. The accuracy of the oxazole's equations is higher than for thiazoles, likewise the individual significance of the q_3 descriptor.

4.1.3.	**RI= a_0+bCTI + c_1q_3**			
Phase	**R**	**a_0**	**b**	**c_1**
OV-101 oxazoles	0.978	-718.07 ±244.44 t = -3.3; p =5.10^{-3}	10.49±0.63 t = 16.6; p = 3.10^{-10}	-2066.8 ±1472.8 t =-1.4; p =0.18
OV-101 ***thiazoles***	*0.956*	*-645.64 ±339.11* *t = 0.9; p =0.39*	*10.74 ±0.93* *t = 11.5; p = 3.10^{-8}*	*-3171.2 ±3085.6* *t =-1.02; p =0.32*
X-305 oxazoles	0.959	-274.26 ±238.76 t = -1.1; p = 0.27	8.28 ± 0.70 t = 11.8;p=2.10^{-8}	-2934.4 ±1639.35 t = -1.8; p =0.09

A sharp improvement of the model's accuracy (R) is noticeable when entering q_4 as a descriptor for models deriving, especially for oxazoles equations (Table.4.1.4). The C-4 site is a nucleophilic center of the azole ring; the positive inductive effect of the alkyl substituent in this position enhances this role. The term q_4 is significant at p=10^{-6} for oxazoles and at p=10^{-4} for thiazoles. The contribution of c_1q_4 term increases the retention, given that the value of the q_4 atomic charge is negative.

4.1.4.	**RI= a_0 + bCTI + c_1q_4**			
Phase	**R**	**a_0**	**b**	**c_1**
OV-101 oxazoles	0.996	- 611.39 ±42.17 t = -145; p =2.10^{-9}	11.22 ±0.30 t = 37.9; p = 1.10^{-14}	-689.56 ±88.94 t = -7.71; p =3.10^{-6}
OV-101 *thiazoles*	*0.976*	*- 459.26 ±95.06* *t = -4.8; p =3.10^{-4}*	*11.48 ±0.71* *t = 16.2; p = 5.10^{-10}*	*-649.93 ±182.78* *t = -3.6; p = 3.10^{-3}*
X-305 oxazoles	0.990	-73.90 ±52.77 t = -1.4; p ≈ 0.18	9.14 ±0.37 t = 24.7;p=2.10^{-12}	-784.20 ±111.28 t = -7.0; p ≈8.10^{-6}
X-305 *thiazoles*	*0.974*	*128.09 ±80.42* *t = 1.6; p =0.13*	*9.31 ±0.60* *t = 15.5;p=9.10^{-10}*	*-645.59 ±154.63* *t = -4.2; p ≈1.10^{-3}*
PEG-40 *thiazoles*	*0.973*	*169.83 ±82.30* *t = 2.1; p=6.10^{-2}*	*9.39 ±0.61* *t = 15.3;p≈1.10^{-9}*	*-715.25 ±158.30* *t = -4.5; p≈6.10^{-4}*

When considering the effect of the alkyl substituent descriptor R_4 usage in the model (Table 4.1.5), the retention decreases both for oxazoles and thiazoles. The effect of substitution of H-atom with alkyl group in the forth position leads to a steric shielding of the N-atom and to a decrease of the polar intermolecular interactions with the stationary phases, more substantial for oxazoles.

4.1.5.	**RI= a_0 + bCTI + c_1R_4**			
Phase	**R**	**a_0**	**b**	**c_1**
OV-101 oxazoles	0.996	- 482.07 ±35.27 t = 13.7; p = 4.10^{-9}	11.14 ±0.28 t = 39.1; p = 7.10^{-15}	-40.52 ±5.06 t = -8.0; p = 2.10^{-6}
OV-101 thiazoles	*0.975*	*- 370.92 ±89.38* *t = -4.1; p = 1.10^{-3}*	*11.48 ±0.72* *t = 15.9; p = 7.10^{-10}*	*-41.43 ±12.11* *t = -3.4; p = 4.10^{-3}*

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023

X-305 oxazoles	0.991	72.34 ±41.26 t = 1.7; p = 0. 1	9.06 ±0.33 t = 27.1;p=8.10^{-13}	-46.85±5.92 t = -7.9; p =3.10^{-6}
X-305 thiazoles	*0.974*	*214.96 ±74.53 t = 2.9; p = 1.10^{-2}*	*9.33 ±0.60 t = 15.51;p=9.10^{-10}*	*-41.86 ±10.10 t = -4.1; p =1.10^{-3}*
PEG-40 oxazoels	0.992	145.02 ±36.32 t = 3.8; p =2.10^{-3}	8.91±0.31 t = 28.7;p=3.10^{-13}	-48.50±5.50 t = -8.8; p = 7.10^{-7}
PEG-40 thiazoels	*0.973*	*266.38 ±79.94 t = 3.5; p =4.10^{-3}*	*9.4 ±0.62 t = 15.1;p=3.10^{-9}*	*-46.13 ±10.43 t = -4.4; p = 7.10^{-4}*

The contribution of c_1R_4 term reflects the above tendencies: the sign of the R_4 term contribution is negative and decreases the retention in all equations. The contribution of R_4 is larger in polar phases, especially for PEG-40 phase (H-bonding case) both for two solute groups. The values of R_4 contribution and its significance (p) are larger for oxazoles then thiazoles in the case of polar phase.

Five models were derived with participation of q_5 descriptor (Table 4.1.6).

4.1.6.	**RI= a_0+bCTI + c_1q_5**			
Phase	**R**	**a_0**	**b**	**c_1**
OV-101 oxazoles	0.980	- 364.39 ±85.61 t = -4.2; p =9.10^{-4}	10.23 ±0.63 t = 16.3; p = 5.10^{-10}	337.46±183.37 t =1.8; p =0.09
X-305 oxazoles	0.971	247.97 ±83.26 t = 2.9; p = 0.01	7.81 ±0.61 t = 12.8;p=9.10^{-9}	570.73±178.35 t = 3.2; p =7.10^{-3}
X-305 thiazoles	*0.971*	*531.01 ±103.181 t = 5.1; p=2.10^{-4}*	*8.08 ±0.63 t = 12.9;p=9.10^{-9}*	*456.36 ±119.54 t = 3.8; p =2.10^{-3}*
PEG-40 oxazoles	0.968	319.77 ±87.21 t = 3.7; p =2.10^{-3}	7.65 ±0.64 t = 12.0;p=2.10^{-8}	558.59 ±186.8 t = 3.0; p =0.01
PEG-40 thiazloes	*0.968*	*609.82 ±109.38 t = 5.6; p =9.10^{-5}*	*8.05 ±0.67 t = 12.0;p=2.10^{-8}*	*494.49 ±126.73 t = 3.9; p =2.10^{-3}*

In all the cases q_5 is included with a (+) sign within the parametric score, which due to its own negative atomic charge leads to a decrease of the retention. In the process of chromatographic separation of alkylazoles, the replacement of the hydrogen atom in the fifth position with an alkyl substituent reduces the retention. Probably in this case the steric hindrance effect with respect to O-atom is estimated. The presence of steric influence leads to a decrease in the retention property. The effect is more pronounced for thiazols than the oxazoles.

4.1.7.	**RI= a_0 + bCTI + c_1R_5**			
Phase	**R**	**a_0**	**b**	**c_1**
OV-101 oxazoles	0.979	178.69 ±78.61 t = 2.3; p =4.10^{-2}	10.24 ±0.64 t = 15.8; p = 7.10^{-10}	17.70 ±11.03 t =1.6; p =0.13
OV-101 thiazoles	*0.970*	*-276.28 ±97.87 t = -2.8; p = 1. 10^{-2}*	*10.33 ±0.79 t = 13.1; p = 7.10^{-9}*	*34.63 ±12.64 t =2.7; p =1.10^{-2}*
X-305 oxazoles	0.969	178.68 ±78.61 t = 2.3; p = 4.10^{-2}	7.80 ±0.64 t = 12.2; p=2.10^{-8}	31.80 ±10.85 t = 2.9; p =1.10^{-2}
X-305 thiazoles	*0.968*	*312.20 ±82.15 t = 3.8; p = 2.10^{-3}*	*8.15 ±0.66 t = 12.3; p=1.10^{-8}*	*36.33 ±10.61 t = 3.4; p =41.10^{-3}*

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023

PEG-40 oxazoles	0.964	250.74 ±83.01 t = 3.0; p =1.10^{-2}	7.65 ±0.67 t = 11.3;p=4.10^{-8}	30.43 ±11.45 t = 2.6; 2.10^{-2}
PEG-40 thiazoles	*0.965*	*372.75 ±87.25 t = 4.3; p =9.10^{-4}*	*8.12 ±0. 70 t = 11.5;p=3.10^{-8}*	*39.38 ±11.27 t = 3.4; 4.10^{-3}*

When modeling with a structural descriptor R_5 (Table 4.1.7), its contribution to retention is positive for all phases. In the case of equations for thiazols the contribution for R_5 is almost the same for the non-polar and the middle polar phase and slightly increases in the polar phase. For oxazoles, the contribution of R_5 increases almost double in the polar phases.

The comparison of the equations accuracy (R) with a structure descriptor R_4 and with R_5, respectively, one can see that R_4 parameter seems to be more reliable parameter to quantify the ability of azole solutes to participate in interactions with stationary phases.

The next step in QSRR building was the creating of three-parametric equations in order to increase the equation's accuracy. The requirement of orthogonality of the variables in a common equation is a necessary condition for meaningful results [19]. It limits the use of the variables in the common regression only, in the case of the low level cross-correlation coefficient, ir < 0.5. It would certainly be desirable for more meaningful QSRR results to find equally significant but less collinear independent variables.

For the case to explore the CTI index only one combination with Balaban molecular index Bal J had to be possible for meaningful modeling. A new set of QSRRs have been established to model the retention data. It is shown in Tables 4.2. (1-7).

4.2.1.	**RI = a_0 + b_1CTI + b_2Bal J + c_1q_1**				
Phase	**R**	**a_0**	**b_1**	**b_2**	**c_1**
OV-101 ***thiazoles***	0.966	589.14 ±426.72 t = 1.4; p =0.19	10.32±0.89 t = 11.5; p = 8.10^{-8}	-97.83 ±86.49 t = -1.1; p =0.28	-1354.2 ±657.88 t = -2.1; p =0.06
X-305 ***thiazoles***	0.963	1219.3 ±364.8 t = 3.3; p =6.10^{-3}	8.06±0.77 t = 10.5; p = 2.10^{-7}	-81.3 ±73.95 t = -1.1; p =0.05	-10174±2091.04 t = -4.9; p =4.10^{-4}
PEG-40 ***thiazoles***	0.958	1345.87 ±391.87 t = 3.4; p =3.10^{-3}	8.57±0.9 t = 9.4; p = 3.10^{-7}	-89.74 ±79.4 t = -1.1; p =0.28	-1606.1 ±604.16 t = -2.7; p =0.02

4.2.2.	**RI = a_0+b_1CTI + b_2Bal J + c_1q_2**				
Phase	**R**	**a_0**	**b_1**	**b_2**	**c_1**
OV-101 oxazoles	0.995	-115.29 ±81.30 t = -1.4; p =0.18	10.93 ±0.32 t = 33.5; p = 3.10^{-13}	-154.29 ±34.58 t = -4.5; p =8.10^{-4}	-7761.75 ±1503.24 t = -5.1; p =2.10^{-4}
X-305 oxazoles	0.985	344.84 ±113.08 t = 3.0; p =0.01	8.70±0.45 t = 19.2; p = 2.10^{-10}	-103.34 ±48.10 t = -2.1; p =0.05	-10174 ±2091.04 t = -4.9; p =4.10^{-4}
PEG-40 oxazoles	0.987	455.16 ±103.53 t = 4.4; p =9.10^{-4}	8.56 ±0.42 t = 20.6; p = 1.10^{-10}	-120.74±44.04 t = -2.79; p =2.10^{-2}	-10554.21 ±1194.56 t = -5.5; p =1.10^{-4}

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023

4.2.3.	**RI = a_0+b_1CTI + b_2Bal J + c_1q_3**				
Phase	**R**	**a_0**	**b_1**	**b_2**	**c_1**
OV-101 oxazoles	0.987	-427.92 ±200.57 t = -2.1;p =5.10^{-2}	10.74±0.51 t =20.9;p =8.10^{-11}	-154.77 ±54.49 t = -2.8; p =1.10^{-2}	-2253.0 ±1187.24 t = -1.9; p =2.10^{-2}

4.2.4.	**RI = a_0+b_1CTI + b_2Bal J + c_1q_4**				
Phase	**R**	**a_0**	**b_1**	**b_2**	**c_1**
OV-101 oxazoles	0.998	-432.47 ±65.88 t = -6.6; p =3.10^{-5}	11.29 ±0.23 t = 49.2; p = 3.10^{-15}	-78.8 ±25.09 t = -3.1; p =9.10^{-3}	-603.18 ±72.95 t = -8.4; p =2.10^{-6}

4.2.5.	**RI = a_0+b_1CTI + b_2Bal J + c_1R_4**				
Phase	**R**	**a_0**	**b_1**	**b_2**	**c_1**
OV-101 oxazoles	0.998	-303.81 ±50.53 t = -6.0; p =6.10^{-5}	11.23 ±0.19 t = 57.5; p = 5.10^{-16}	-85.19 ±21.25 t = -4.0; p =1.10^{-3}	-36.18 ±3.61 t = -10.0; p=3.10^{7}

A combination of two global topologic indices (CTI+BalJ) used for the regression lead to formation of new more precise models. It can be seen that the improvement of the statistics of the models for oxazole retention is possible after including of the q_2, q_4 and R_4 parameters (Tables 4.2.2, 4.2.4 and 4.2.5). Contributions of the local descriptors in the new set of equations retain their character.

For both groups of azoles, the highest accuracy of the models is received when the C-5 position descriptor is included (Table 4.2.6 and 4.2.7).

4.2.6.	**RI= a_0 + b_1CTI + b_2Bal J + c_1q_5**				
Phase	**R**	**a_0**	**b_1**	**b_2**	**c_1**
OV-101 oxazoles	0.998	210.74 ±53.99 t = 3.9; p =2.10^{-3}	10.33 ±0.18 t = 56.9; p = 6.10^{-16}	-248.96 ±20.77 t = -11.9; p =5.10^{-8}	647.07 ±58.96 t = 10.9; p =1.10^{7}
OV-101 ***thiazoles***	*0.984*	*444.81 ±194.61 t = 2.3; p =4.10^{-2}*	*10.29 ±0.60 t = 17.2; p =8.10^{-10}*	*-199.01 ±64.87 t = -3.1; p =1.10^{-2}*	*597.23 ±126.31 t = 4.7; p =5.10^{-4}*
X-305 oxazoles	0.995	775.78 ±79.89 t = 9.7; p =5.10^{-7}	7.91 ±0.27 t = 29.4; p = 1.10^{-12}	-228.47 ±30.73 t = -7.4; p =8.10^{-6}	854.86 ±87.23 t = 9.8; p =4.10^{-7}
X-305 ***thiazoles***	*0.987*	*1000.71 ±142.75 t =7.0; p =1.10^{-5}*	*8.09 ±0.44 t = 18.4;p = 4.10^{-10}*	*-181.49 ±47.58 t = -3.8; p =2.10^{3}*	*608.27 ±92.65 t = 6.7; p =3.10^{-5}*
PEG-40 oxazoles	0.996	891.22 ±69.21 t = 12.9; p =2.10^{-8}	7.76 ±0.23 t = 33.3; p = 3.10^{-13}	-247.36 ±26.63 t = 9.3; p =8.10^{-7}	866.22 ±75.58 t = 11.5; p =8.10^{8}
PEG-40 ***thiazoles***	*0.987*	*1125.8 ±144.41 t =7.8 p =5.10^{-6}*	*8.06 ±0.44 t = 18.1;p = 4.10^{-10}*	*-199.38 ±48.14 t = -4.1; p =1.10^{3}*	*661.37 ±93.73 t = 7.1; p =1.10^{-5}*

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023

4.2.7.	**RI= a_0 + b_1CTI + b_2Bal J + c_1R_5**				
Phase	**R**	**a_0**	**b_1**	**b_2**	**c_1**
OV-101 oxazoles	0.999	172.57 ±47.80 t = 3.6; p =3.10^{-4}	10.23 ±0.17 t = 61.1; p = 2.10^{-16}	-265.83 ±19.74 t = -13.5; p=1.10^{-8}	39.47 ±3.29 t = 12.0; p =5.10^{-8}
OV-101 ***thiazoles***	*0.986*	*241.12 ±157.99 t = 1.5; p =0.15*	*10.32 ±0.56 t = 18.3;p = 4.10^{-10}*	*-234.50 ±64.17 t = -3.6;p =3.10^{-3}*	*54.07 ±10.50 t = 5.2; p =2.10^{-4}*
X-305 oxazoles	0.996	727.28 ±68.7 t = 10.6; p =2.10^{7}	7.83 ±0.24 t = 32.4; p = 5.10^{-13}	-251.49 ±28.40 t = -8.9; p =1.10^{-6}	52.40 ±4.73 t = 11.1; p =1.10^{-7}
X-305 ***thiazoles***	*0.988*	*785.98 ±117.72 t = 6.7; p =2.10^{-5}*	*8.13 ±0.42 t = 19.3 p = 2.10^{-10}*	*-214.73 ±47.82 t = -4.5; p =7.10^{-4}*	54.13 ±7.83 t = 6.9; p =2.10^{-5}
PEG-40 oxazoles	0.996	837.01 ±65.58 t = 12.6; p =3.10^{8}	7.69 ±0.23 t = 32.8; p = 4.10^{-13}	-268.76 ±27.51 t = -9.8; p =4.10^{-7}	52.44±4.58 t = 11.5; p =8.10^{-8}
PEG-40 ***thiazoles***	*0.988*	*892.85 ±117.67 t = 7.6; p =6.10^{-6}*	*8.10 ±0.42 t = 19.3;p = 2.10^{-10}*	*-235.72 ±47.80 t = -4.9; p =3.10^{-4}*	*58.92 ±7.82 t = 75; p =6.10^{-6}*

The sign of the c_1 in the models is positive. Taking into account the own (-) charge of the atom, the contribution of the term c_1q_5 to the retention is negative. Entering q_5 is conducive to a decrease of the retention. However, when using of R_5 in the equation, its contribution becomes positive. The models obtained are extremely accurate for both groups of azoles. This combination of two topological indices occurs to be very successful in the case. The rise of correlation coefficient value is obvious: it is changed from **R**=0.98 to **R**=0.999 for non polar **OV-101** phase. The similar tendency can be trended for both polar phases: in the case of **Triton X-305** the correlation coefficient value increases from **R**=0.971 up to **R**=0.996 for oxazoles and from **R**=0.968 to **R**=0.988 for thiazoles.

Some additional combinations of local descriptors with *CTI* index that create adequate QSRR models are shown in the Tables 4.3.(1-4). The attempt to improve the modeling with q_1 by including q_4 (resp. R_4) was not successful. It appeared that the thiazole models are adequate only. Their accuracy grew insignificantly (Table 4.3.2-2).

4.3.1.	**RI = a_0 + b_1CTI + c_1q_1+ c_2q_4**				
Phase	**R**	**a_0**	**b_1**	**c_1**	**c_2**
OV-101 ***thiazoles***	*0.978*	*-113.1 ±317.89 t = -0.36; p =0.72*	*11.1 ±0.77 t = 14.4; p = 6.10^{-9}*	*-634.79 ±556.95 t = -1.1; p =0.27*	*-570.31 ±193.74 t = -3.9; p =0.01*
X-305 ***thiazoles***	*0.980*	*585.60 ±247.08 t = 2.3; p =0.035*	*8.82 ±0.60 t = 14.6; p = 5.10^{-9}*	*-839.09 ±432.88 t = -1.9; p =0.07*	*-540.34±150.58 t = -3.6; p =4.10^{-3}*
PEG-40 ***thiazoles***	*0.980*	*638.05 ±252.83 t = 2.5; p =0.026*	*8.89 ±0.62 t = 14.4; p = 6.10^{-9}*	*-858.73 ±442.96 t = -1.9; p =7.10^{-2}*	*-607.54 ±154.09 t = -3.9; p =2.10^{-3}*

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023

4.3.2	**RI = a_0 + b_1CTI + c_1q_1+ c_2R_4**				
Phase	**R**	**a_0**	**b_1**	**c_1**	**c_2**
X-305 ***thiazoles***	*0.979*	*637.45 ±247.54 t = 2.6; p =0.02*	*8.84 ±0.62 t = 14.2; p = 7.10^{-9}*	*-798.24 ±449.16 t = -1.8; p =0.1*	*-34.80 ±10.16 t = -3.4; p =5.10^{-3}*
PEG-40 ***thiazoles***	*0.978*	*698.53 ±256.15 t = 2.7; p =0.018*	*8.91 ±0.64 t = 13.9; p = 9.10^{-9}*	*-816.51 ±464.77 t = -1.7; p =0.1*	*-38.91 ±10.52 t = -3.7; p =3.10^{-3}*

Modeling with q_3 with additionally including of q_4 was successful for oxazoles only, and just for the X-305 phase (Table 4.3.3). The accuracy increased significantly from 0.959 to 0.991.

4.3.3.	**RI = a_0 + b_1CTI + c_1q_3+ c_2q_4**				
Phase	**R**	**a_0**	**b_1**	**c_1**	**c_2**
X-305 oxazoles	0.991	-202.11 ±120.19 t = -1.7; p =0.11	9.05±0.37 t = 24.3; p = 1.10^{-11}	-1135 ±874.61 t = -1.2; p =0.26	-736.66 ±116.72 t = -6.3; p =4.10^{-5}

Models involving q_2 with additional adding of q_5 (Table 4.3.4) became significant for thiazole, but didn't describe properly oxazoles:

4.3.4.	**RI = a_0 + b_1CTI + c_1q_2 + c_2q_5**				
Phase	**R**	**a_0**	**b_1**	**c_1**	**c_2**
X-305 ***thiazoles***	*0.975*	*673.29 ±144.49 t = 4.6; p =6.10^{-4}*	*7.97 ±0.61 t = 13.0; p = 2.10^{-8}*	*461.70 ±338.68 t = 1.4; p =0.2*	*472.43 ±116.54 t = 4.1; p =3.10^{-3}*
PEG-40 ***thiazoles***	*0.975*	*788.42 ±148.34 t = 5.3; p =2.10^{-4}*	*7.91 ±0.63 t = 12.5; p = 3.10^{-8}*	*579.57 ±347.69 t = 1.7; p =0.1*	*517.17 ±119.64 t = 4.3; p =1.10^{-3}*

The search for combinations to increase the statistics can be continued adding more parameters into the model. As an example for the oxazoles the equation derived for the OV-101 phase is perfectly accurate (R=0.9986; s=5.6) :

$RI_{OV\text{-}101}$=131.9(± 89.7) +10.33(± 0.29) CTI - 245.7(± 20.81) Bal J-486.5 (± 443.9)q_3+ 617.9 (± 64.2)q_5

Conclusions

Retention modeling for two groups of azoles – 16 alkyloxazoles and 16 alkylthiazoles separated by gas chromatography on three columns with different polarities was proceeded. MLR adequate equations were established for retention data generated on unpolar OV-101**,** polar Triton X-305 and PEG-40M stationary phases.

All regressions derived vouch for the requirements for reporting the results of correlation analysis.

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023

Comparative analyses of regression coefficients values in similar regression models for different chromatographic phases have been provided.

The present quantitative characterization of the chromatographic retention of alkylazoles shows the potentials of deriving **QSRR** models by the employment of **CTI** index as global descriptor for the GC retention modeling. The high precision (R=0.998) of the models makes it possible to use them for prediction and/or control the chromatographic retention by QSSR based expert systems.

Acknowledgment: This work has been supported by grant DFNI IO1/7 of the National Fund for Scientific Investigation and under project № RD-08-104//06.02.2018, of "Scientific Research" section of Konstantin Preslavsky University of Shumen.

References

[1]. Hanai, T. (Ed.) Quantitative in Silici Chromatography: Computational Modeling of molecular Interactions; Royal Society of Chemistry: Cambridge, UK, **2014**.
[2]. Hanai, T., *J. Chromatogr. Sep.Tech.*, **2016**, *7*, 1.
[3]. McCalley, D.V., *J. Chromatogr. A*, **2010**, *1217*, 3408.
[4]. Karatapanis, A.E.; Fiamegos, Y.C.; Stalikas, C.D., *Chromatographia*, **2010**, *71*, 751.
[5]. Karatapanis, A.E.; Fiamegos, Y.C.; Stalikas, C.D., *J. Chromatogr. A*, **2011**, *1218*, 2871.
[6]. Jandera, P., *Anal. Chim. Acta*, **2011**, *692*, 1.
[7]. Jandera, P.; Janas, P., *Anal. Chim. Acta*, **2017**, *967*, 12.
[8]. Noga, S.; Bocian, S.; Buszewski, B. *J. Chromatogr. A*, **2013**, *1301*, 98.
[10]. Kovats, E. Sz., *Chimia*, **1968,** *22***,** 459.
[11]. Moskovkina, M.; Bangov, I., *Annual of Konstantin Preslavsky University, Shumen*, **2006**, *XVI, B2***,** 7.
[12]. Moskovkina, M.; I. Bangov, *Annual of Konstantin Preslavsky University, Shumen*, **2006**, *XVI, B2***,** 40.
[13]. Moskovkina, M.; I. Bangov, Annual of Konstantin Preslavsky University, Shumen, **2008**, XVIII, B2**,** 106.
[14]. Bangov, I.P., *J. Chem. Inform. Comput. Sci.*,**1990**, *30*, 277.
[15]. Demirev, P.A.; Dyulgerov, A.; Bangov, I.P., *J. Math. Chem.*, **1991**, *8*, 367.
[16]. Petrov, E.; Stoyanov, B.; Kochev, N.; Bangov I., *Commun. Math. Comput. Chem.* (MATCH), **2014,** 71, *(3)*, 645.
[17]. Golovnya, R. V., *J.Chromat.*, **1991**, *552*, 1.
[18]. Moskovkina, M.; Bangov, I., *C. R. Bulg. Acad. Sci. CHEMISTRY*, **2012**, *65 (9)*, 1199.
[19]. Charton, M.; Clementi, S.; Ehrenson, S.; Exner, O.; Shorter, J.; Wold, S., *Quant. Struct. Act.Relat.*, **1985,** *4***,** 29.

Corresponding author: m.moskovkina@shu.bg

DOI: 10.2478/asn-2018-0023