Iranian Journal of Mathematical Chemistry

Novel Atom–Type–Based Topological Descriptors for Simultaneous Prediction of Gas Chromatographic Retention Indices of Saturated Alcohols on Different Stationary Phases

FARIBA SAFA•

Department of Chemistry, Rasht Branch, Islamic Azad University, Rasht, Iran

ARTICLE INFO	ABSTRACT
	In this work, novel atom-type-based topological indices, named *AT* indices, were presented as descriptors to encode structural information of a molecule at the atomic level. The descriptors were successfully used for simultaneous quantitative structure-retention relationship (QSRR) modeling of saturated alcohols on different stationary phases (SE-30, OV-3, OV-7, OV-11, OV-17 and OV-25). At first, multiple linear regression models for Kovats retention index (*RI*) of alcohols on each stationary phase were separately developed using *AT* and Randic's first-order molecular connectivity ($^1\chi$) indices. Adjusted correlation coefficient (R^2_{adj}) and standard error (*SE*) for the models were in the range of 0.994-0.999 and 4.40-8.90, respectively. Statistical validity of the models were verified by leave-one-out cross validation ($R^2_{cv} > 0.99$). In the next step, whole *RI* values on the stationary phases were combined to generate a new data set. Then, a unified model, added McReynolds polarity term as a descriptor, was developed for the new data set and the results were satisfactory (R^2_{adj}=0.995 and *SE*=8.55). External validation of the model resulted in the average values of 8.29 and 8.69 for standard errors of calibration and prediction, respectively. The topological indices well covered the molecular properties known to be relevant for retention indices of the model compounds.

1. INTRODUCTION

QSRR modeling as a branch of quantitative structure–property relationship (*QSPR*) studies, is one of the most effective approaches to provide significant information on retention

---

mechanism as well as the effect of molecular structure on retention behavior of various classes of compounds [1,2]. The first important step in QSPR studies is to quantify chemical structures of the molecules by descriptors that can show structural similarity and diversity of the molecules. Among the different types of descriptors, **topological indices** have found major popularity in QSRR studies since they can be rapidly obtained using only two dimensional structure of molecules. These descriptors are exact numbers without uncertainty that offer a simple way of measuring molecular branching, shape, size, cyclicity, symmetry, centricity and complexity [3].

A current trend in QSPR studies is the use of the **atomic level topological indices** that unlike classical topological indices that characterize a molecule as a whole, code the structural environment of each atom type in the molecule and describe the structural information of a molecule at the atomic level. Ren [4] proposed a set of **atomic-based AI topological indices** that along with modified **Xu** index were successfully used in estimation of different physical and chemical properties. The topological indices showed good correlations with normal boiling points, molar volumes, molar refractions, and molecular total surface areas of alcohols [5]. Combination of *Xu* and *AI* indices was also used to develop high quality QSPR models for physical properties of alkanes including normal boiling points, heats of vaporization, molar volumes, molar refractions, van der Waals' constants, and Pitzer's eccentric factors [6]. Panneerselvam *et al*. [7] developed a linear regression model based on *AI* indices for normal boiling point of trialkyl phosphates. In a previous paper, QSPR study of standard formation enthalpies of acyclic alkanes using atom-type-based *AI* topological indices was reported by our group [8]. Nevertheless, *AI* indices have not found widespread application in QSRR studies and there is only one report in which gas chromatographic retention index (**RI**) values of a data set consisting of 33 aldehydes and ketones on four different polar stationary phases were separately correlated with *AI* atom-type-based indices [9].

The main aim of this study was to introduce novel atom-type-based topological indices that can be effectively used in QSRR modeling. The other goal was to demonstrate the role of structural features on the molecular mechanism of chromatographic retention on different stationary phases. In this paper, **novel atom-type-based AT topological indices** along with the most-used **Randić's first-order molecular connectivity index** as a classical topological index were used for simultaneous prediction of gas chromatographic Kovats retention index of saturated alcohols on low to medium polar stationary phases (SE-30, OV-3, OV-7, OV-11, OV-17 and OV-25). The model satisfactorily accounted for the influence of molecular size and each atom-type or group on retention indices of alcohols on different stationary phases. As far as I am aware, this is the first time use of the atomic level topological indices for simultaneous prediction of *RI* data on the stationary phases of different polarity.

2. COMPUTATIONAL DETAILS

2.1 DATA SET

The data sets of Kovats retention indices of 25 saturated alcohols on six stationary phases of different polarity (SE-30, OV-3, OV-7, OV-11, OV-17 and OV-25) were taken from the report of Pias and Gasco [10]. The *RI* values of the model compounds fall in the range of 544-1156 on different stationary phases. Table 1 shows the data sets of Kovats retention indices of 25 model compounds on six stationary phases of different polarity.

Table 1. Kovats retention indices of saturated alcohols on different stationary phases.

No	Compound	Stationary phases					
		SE-30 (53)[a]	OV-3 (86)[a]	OV-7 (113)[a]	OV-11 (142)[a]	OV-17 (158)[a]	OV-25 (204)[a]
1	1-Butanol	650	672	702	725	748	792
2	1-Hexanol	856	881	907	935	959	1003
3	1-Heptanol	960	985	1010	1038	1062	1104
4	2-Butanol	586	607	633	656	675	711
5	2-Pentanol	689	711	735	756	777	811
6	3-Pentanol	689	708	733	756	777	808
7	3-Hexanol	785	807	830	853	878	904
8	3-Heptanol	886	909	929	955	975	1008
9	4-Heptanol	880	904	924	946	968	999
10	4-Methyl-2-Butanol	628	652	674	692	709	738
11	2-Methyl-2-Hexanol	822	848	862	884	904	930
12	2-Methyl-2-Heptanol	920	944	961	982	1001	1026
13	2-Methyl-3-Hexanol	858	876	897	920	939	969
14	3-Methyl-1-Butanol	725	747	771	798	817	855
15	4-Methyl-1-Pentanol	827	849	876	902	923	960
16	2-Ethyl-1-Hexanol	1019	1046	1067	1092	1116	1156
17	3-Ethyl-3-Pentanol	853	876	898	920	939	974
18	2,2-Dimethyl-3-Pentanol	814	834	855	874	890	919
19	2,2-Dimethyl-3-Hexanol	906	926	944	962	977	1004
20	1-Propanol	544	574	-	-	-	-
21	1-Pentanol	751	777	-	806	856	900
22	2-Hexanol	787	811	835	-	878	914
23	2-Methyl-1-Propanol	612	641	654	-	680	740
24	2-Methyl-2-Pentanol	726	748	767	-	801	827
25	2-Ethyl-1-Butanol	834	857	-	907	928	-

[a] McReynolds polarity

2.2 DEFINITION OF *AT* TOPOLOGICAL INDEX

Novel atom-type-based *AT* topological indices were derived from the molecular topological graph in the same manner of Ren [11] with some modifications. As known, a molecular graph G = {*V, E*} consists of the vertex V(*G*) and edge E(*G*) sets, where vertices correspond to individual atoms in the graph and the edges correspond to chemical bonds between them. The graph G can be characterized by the distance matrix, $\mathbf{D} = [d_{ij}]_{n \times n}$, whose elements are the length of the shortest distance between the vertices *i* and *j* in a molecular graph with n vertices. Clearly, the sum over the row *i* (or column *j*) of the distance matrix gives the distance sum vector, $\mathbf{S} = [s_i]_{n \times 1}$, for the graph. The molecular graph can also be coded by a vertex degree vector $\mathbf{V} = [v_i]_{n \times 1}$ whose elements are the number of connections (edges) of the vertex i in the graph. For heteroatoms and multiple bonds in the graph, the new degree of vertex introduced by Ren [6], v^m, can be used to replace the v_i values. This novel degree of vertex is derived from the valence connectivity δ^v of Kier-Hall [12] and is defined as:

$$v^m = \delta + k \tag{1}$$

$$k = 1 / [(2/N)^2 \delta^v + 1] \tag{2}$$

where δ is the number of connections of that atom and parameter *k* is a perturbing term reflecting the effect of heteroatom. *N* is the principal quantum number of the valence shell.

According to the above definitions, for any atom *i* that belongs to the *j*-th atom-type in the graph, the AT_i (*j*) is defined as follows:

$$AT_i(j) = n \times \frac{s_i v_i}{\sum_{i=1}^{n} s_i v_i} \tag{3}$$

Clearly, for *j*-th atom-type in a graph, the corresponding *AT(j)* index is the sum of all $AT_i(j)$ values of the same atom type.

$$AT(j) = \sum_{i}^{n} AT_i(j) \tag{4}$$

As an illustration, Figure 1 depicts the labeled molecular graph of 3-methyl-1-butanol.
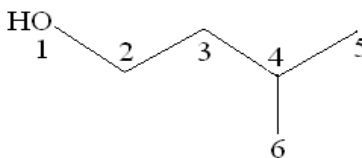

Figure 1. Labeled molecular graph of 3-methyl-1-butanol.

The distance matrix **D** is expressed as follows:

$$D = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 4 \\ 1 & 0 & 1 & 2 & 3 & 3 \\ 2 & 1 & 0 & 1 & 2 & 2 \\ 3 & 2 & 1 & 0 & 1 & 1 \\ 4 & 3 & 2 & 1 & 0 & 2 \\ 4 & 3 & 2 & 1 & 2 & 0 \end{bmatrix}$$

The distance sum vector and vertex degree vector are readily obtained as

$$\mathbf{S} = [\, 14 \quad 10 \quad 8 \quad 8 \quad 12 \quad 12 \,]$$

$$\mathbf{V} = [\, 1.167 \quad 2 \quad 2 \quad 3 \quad 1 \quad 1 \,]$$

where the elements of the vector **V**, are the modified vertex-degree v^m. According to the above definitions, the *AT* indices are calculated as follows:

$$AT\,(-CH_3) = AT\,(5) + AT\,(6) = \left(6 \times \frac{1 \times 12}{100.338} \right) + \left(6 \times \frac{1 \times 12}{100.338} \right) = 1.4351.$$

$$AT\,(-CH_2-) = AT\,(2) + AT\,(3) = \left(6 \times \frac{2 \times 10}{100.338} \right) + \left(6 \times \frac{2 \times 8}{100.338} \right) = 2.1527.$$

$$AT\,(-CH-) = AT\,(4) = \left(6 \times \frac{3 \times 8}{100.338} \right) = 1.4351.$$

$$AT\,(-OH) = AT\,(1) = \left(6 \times \frac{14 \times 1.167}{100.338} \right) = 0.9770.$$

2.3 REGRESSION ANALYSIS AND MODEL VALIDATION

Multiple linear regression (MLR) analyses were performed using SPSS/PC software package [13]. Criteria for selection of the best models were the statistics of coefficient of multiple determination (R^2), adjusted correlation coefficient, Fisher-ratio and *SE*. Validity and stability of the individual models obtained for the stationary phases was tested using leave-one-out cross-validation (LOO-CV) method [14]. External stability of the unified model generated for simultaneous prediction of *RI* data of saturated alcohols on different stationary phases was tested by external validation method. Standard error of calibration (*SEC*) and standard error of prediction (*SEP*) were used for evaluating quality of the unified MLR model [15].

3.   RESULTS AND DISCUSSIONS

3.1 INDIVIDUAL AND COMBINED LINEAR REGRESSION MODELS FOR DIFFERENT STATIONARY PHASES

In the first step of the study, a combination of $^1\chi$ and AT topological indices were used to develop individual QSRR models for saturated alcohols on different stationary phases. Specifications of the best models found for describing the RI values of saturated alcohols on the stationary phases are given in Table 2. It can be seen that the equations represent excellent QSRR models judging from R^2_{adj} and SE values in the range of 0.994-0.999 and 4.40-8.90, respectively. Also, F values show a high degree of statistical credibility and are indicative of an excellent fit of the models to experimental retention data. Table 3 shows the values of topological descriptors entered in the individual MLR models.

To validate the models, LOO-CV test was performed and cross-validated correlation coefficient (R^2_{cv}) values were in the range of 0.993-0.999 (Table 2) which indicate that the models are robust and reliable over the sample space. As shown, in all cases, cross-validated correlation coefficient is only slightly less than the corresponding value of the full model indicating the models are robust and reliable over the sample space. Table 4 gives the cross-validated predicted values of RI and corresponding relative errors resulted from the individual models developed for the stationary phases. It can be easily seen that relative errors for predicted retention indices are less than 3% and only one case with a slightly large error was found on OV-17 column (4.3% for 2-Methyl-1-Propanol).

In the next step of the study, to generate a unified QSRR model for simultaneous prediction of Kovats retention indices of saturated alcohols on the stationary phases of different polarity, whole retention data for saturated alcohols were combined and used as a new data set including 140 RI data points. As known, generation of a single QSRR model for simultaneous prediction of retention data on different stationary phases requires a parameter that reflects the contribution of the stationary phase to chromatographic retention. The contribution may be reflected by the polarity which is the most representative chromatographic property of the stationary phase. In this work, system of the stationary phase selectivity constant of McReynolds was used and the polarity value compared to butanol (y') was employed [16] as additional parameter in QSRR modeling. McReynolds polarity parameter values (M) are given in Table 1.

Table 2. Characteristics of the best individual QSRR models found for saturated alcohols on different stationary phases.

Parameter	Coefficients (standardized coefficients)					
	SE-30	OV-3	OV-7	OV-11	OV-17	OV-25
Constant	-78.658	-61.020	-24.207	-32.319	23.379	23.473
$^1\chi$	212.539 (1.109)	212.954 (1.114)	211.921 (1.094)	213.848 (1.096)	216.928 (1.098)	216.654 (1.113)
AT(–OH)	236.651 (0.172)	244.899 (0.179)	237.995 (0.176)	266.654 (0.204)	233.514 (0.175)	283.435 (0.213)
AT(–CH₃)	-13.586 (-0.068)	-15.149 (-0.076)	-18.715 (-0.094)	-18.006 (-0.099)	-28.620 (-0.147)	-32.403 (-0.169)
Statistics						
R^2	0.998	0.999	0.998	0.995	0.995	0.995
R^2_{adj}	o.998	0.999	0.997	0.995	0.994	0.994
R^2_{cv}	0.998	0.999	0.997	0.994	0.993	0.993
SE	5.39	4.40	6.20	8.34	8.83	8.90
F	4079.5	6082.8	2514.8	1225.3	1311.3	1251.9
n^a	25	25	22	21	24	23

[a] Number of molecules in the data set.

Table 3. Values of the topological descriptors entered in the individual QSRR models.

No	$^1\chi$	AT (–CH₃)	AT (–OH)	No	$^1\chi$	AT (–CH₃)	AT (–OH)
1	2.4142	0.8108	0.9462	14	2.770	1.4351	0.9770
2	3.4142	0.7924	0.9248	15	3.270	1.4055	0.9648
3	3.9142	0.7869	0.9183	16	4.346	1.5273	0.8459
4	2.2701	1.5937	0.8752	17	3.682	2.3684	0.737
5	2.7701	1.5599	0.8402	18	3.481	2.8968	0.7982
6	2.8081	1.6278	0.8037	19	3.981	2.8547	0.7364
7	3.3081	1.6137	0.7634	20	1.914	0.8275	0.9657
8	3.8081	1.5915	0.7429	21	2.914	0.8000	0.9336
9	3.8081	1.6183	0.7177	22	3.270	1.5340	0.8225
10	2.5607	2.2947	0.8369	23	2.270	1.4952	0.9815
11	3.5607	2.1802	0.7673	24	3.063	2.2309	0.7923
12	4.0607	2.1455	0.7541	25	3.346	1.5572	0.9086
13	3.6807	2.2554	0.7403				

Table 4. Prediction results of the individual QSRR models for saturated alcohols on six stationary phases using leave-one-out cross validation method.

No	Predicted retention indices (relative error%)					
	SE-30	OV-3	OV-7	OV-11	OV-17	OV-25
1	647 (-0.5)	673 (0.1)	696 (-0.9)	721 (-0.6)	745 (-0.4)	788 (-0.5)
2	855 (-0.1)	880 (-0.1)	904 (-0.3)	929 (-0.6)	957 (-0.2)	999 (-0.4)
3	960 (0.0)	986 (0.1)	1009 (-0.1)	1034 (-0.4)	1065 (0.3)	1107 (0.3)
4	590 (0.7)	614 (1.2)	636 (0.5)	658 (0.3)	674 (-0.1)	712 (0.1)
5	688 (-0.1)	711 (0.0)	733 (-0.3)	756 (0.0)	776 (-0.1)	811 (0.0)
6	686 (-0.4)	709 (0.1)	731 (-0.3)	753 (-0.4)	773 (-0.5)	807 (-0.1)
7	783 (-0.3)	806 (-0.1)	828 (-0.2)	849 (-0.5)	872 (-0.7)	904 (0.0)
8	885 (-0.1)	908 (-0.1)	930 (0.1)	951 (-0.4)	978 (0.3)	1007 (-0.1)
9	878 (-0.2)	900 (-0.4)	923 (-0.1)	944 (-0.2)	972 (0.4)	1000 (0.1)
10	633 (0.8)	655 (0.5)	675 (0.1)	699 (1.0)	709 (0.0)	742 (0.5)
11	831 (1.1)	852 (0.5)	873 (1.3)	895 (1.2)	913 (1.0)	943 (1.4)
12	936 (1.7)	957 (1.4)	978 (1.8)	1001 (1.9)	1022 (2.1)	1051 (2.4)
13	847 (-1.3)	869 (-0.8)	889 (-0.9)	911 (-1.0)	929 (-1.1)	956 (-1.3)
14	721 (-0.6)	746 (-0.1)	768 (-0.4)	794 (-0.5)	810 (-0.9)	854 (-0.1)
15	825 (-0.2)	851 (0.2)	871 (-0.6)	898 (-0.4)	917 (-0.7)	960 (0.0)
16	1028 (0.9)	1049 (0.3)	1071 (0.4)	1096 (0.4)	1121 (0.4)	1155 (-0.1)
17	845 (-0.9)	867 (-1.0)	886 (-1.3)	907 (-1.4)	925 (-1.5)	953 (-2.2)
18	810 (-0.5)	831 (-0.4)	847 (-0.9)	872 (-0.2)	879 (-1.2)	906 (-1.4)
19	902 (-0.4)	923 (-0.3)	940 (-0.4)	965 (0.3)	977 (0.0)	1002 (-0.2)
20	546 (0.4)	569 (-0.9)	-	-	-	-
21	751 (0.0)	776 (-0.1)	-	829 (2.9)	850 (-0.7)	892 (-0.9)
22	790 (0.4)	814 (0.4)	836 (0.1)	-	881 (0.3)	915 (0.1)
23	617 (0.8)	640 (-0.2)	665 (1.7)	-	709 (4.3)	747 (0.9)
24	730 (0.6)	752 (0.5)	772 (0.7)	-	810 (1.1)	841 (1.7)
25	826 (-1.0)	850 (-0.8)	-	896 (-1.2)	916 (-1.3)	-

Specifications of the unified QSRR model found for simultaneous prediction of Kovats retention indices of saturated alcohols on the six stationary phases are given in Equation 5. As can be easily found, addition of the McReynolds polarity parameter to the descriptors entered in the individual models resulted in a unified model for all *RI* data with R^2_{adj} value of 0.995 and *SE* of 8.55. Based on the results, regression explained by the model is significant at 99% confidence level and the combined model explains 99.5% of the variance in the retention data.

$$RI = -130.088 + 214.334\ {}^{1}\chi + 250.597\ AT(-OH) - 35.750\ AT(-CH_3) + 0.828\ M \tag{5}$$

$$n = 140 \qquad R^2 = 0.995 \qquad R^2_{adj} = 0.995 \qquad SE = 8.55 \qquad F = 7334.2$$

To prove the external stability of the combined model and to further demonstrate the absence of chance correlation, external validation method was employed. In the method, the whole data set divided into five subsets and each subset was predicted by the other four subsets as the training set. In this procedure, the same descriptors were retained in the correlation equation but the coefficients were allowed to vary. Calculated values of *SEC* and *SEP* for different subsets are given in Table 5. Good agreement between average values of 8.29 for *SEC* and 8.69 for *SEP* with standard error of the full model shows good efficiency of the topological indices in simultaneous modeling of *RI* data for saturated alcohols on different stationary phases. Average values of training and predicting qualities ($R^2_{train} = R^2_{pred} = 0.995$) also indicate high statistical stability and validity of the combined model. Graphical indication of the quality of the combined model can be seen in Figure 2 through the correlation between the experimental and calculated retention indices for different prediction sets. Figure 3 is the plot of residuals versus experimental retention indices for the prediction sets. As shown, all residuals values fall within a horizontal band centered around zero showing absence of systematic error in development of the combined QSRR model.

Table 5. Verification of statistical validity of the combined model based on the external validation method.

Training set[a]	Prediction set[a]	R^2_{train}	SEC	R^2_{pred}	SEP
2-5	1	0.995	8.35	0.993	8.84
1 & 3-5	2	0.996	8.07	0.994	8.45
1,2 & 4,5	3	0.996	8.31	0.995	8.67
1-3 & 5	4	0.995	8.28	0.995	8.79
1-4	5	0.995	8.38	0.997	8.70
Average		0.995	8.29	0.995	8.69

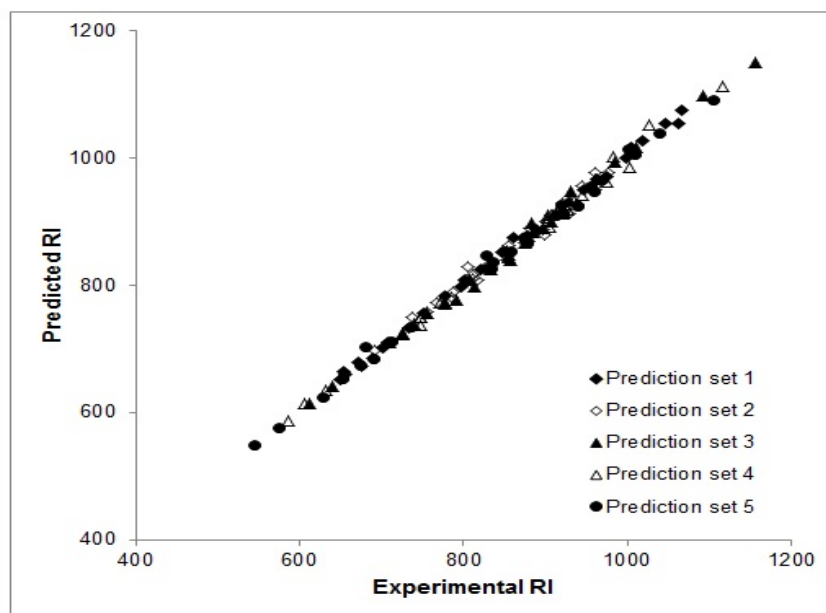[a] Number of molecules in the training and prediction sets are 112 and 28, respectively.

Figure 2. Plot of the predicted versus experimental retention indices for different prediction sets used in the external validation test.
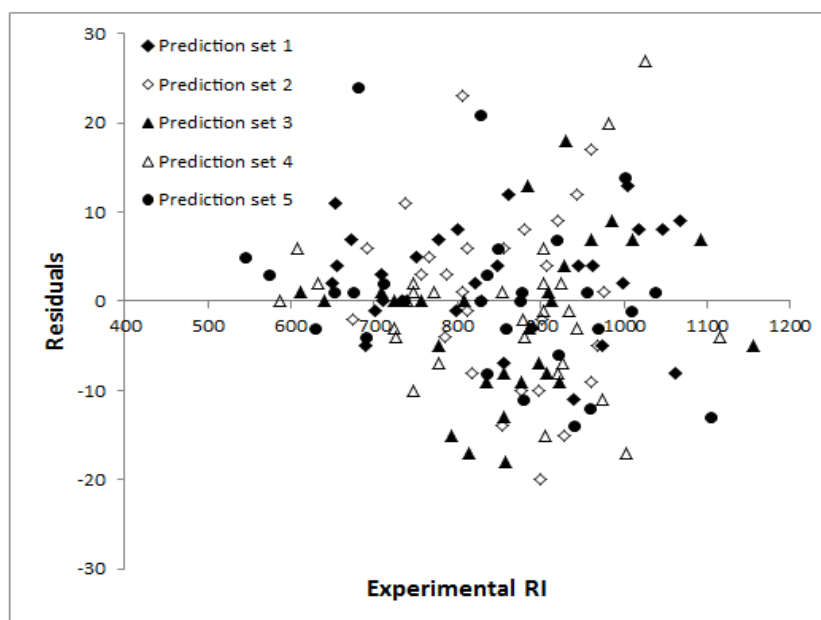

Figure 3. Plot of residuals versus experimental retention indices for different prediction sets used in the external validation test.

To assess the performance of the *AT* topological indices in retention prediction of saturated alcohols, the obtained results were compared with previously reported topological based QSRR models developed for the same data set. Figure 4 shows a comparison between standard errors of the present MLR models and those of the models reported by Liu et al. [17] and Guo et al. [18] for predicting retention indices of saturated alcohols. Clearly, the *AT* based individual models are significantly better than the other two models. As shown, standard errors of the *AT* based individual models are 35.1 to 60.7% less than the model developed by Liu et al. and are 16.9 to 51.3% less than the other model. Moreover, a decrease of about 18% is observed in the standard error of the combined QSRR model presented in this work relative to the model generated by Guo et al. Better results obtained in the present study provide good evidence for high potential of the *AT* topological indices in QSRR modeling of saturated alcohols.
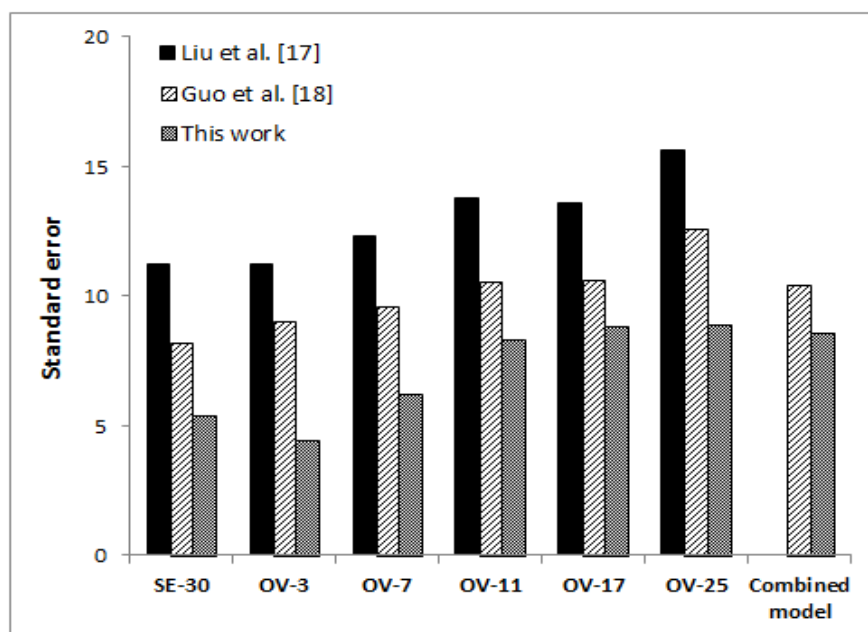

Figure 4. Comparison between standard errors of the presented and previous linear models for prediction of retention indices of saturated alcohols on different stationary phases.

3.2 CONTRIBUTION OF THE INDIVIDUAL INDICES TO RETENTION INDEX OF SATURATED ALCOHOLS

Gas chromatographic retention indices strongly depend on the solute-stationary phase interactions and QSRR studies provide useful information on the mechanism of these interactions [19]. To explore the role of the topological features of the model compounds in

determining their retention indices on each stationary phase, the standardized regression coefficients for the individual QSRR models were used (Table 2). Furthermore, precise contributions of the parameters entered in the combined QSRR model to *RI* values were determined using calculation of the relative contribution (Ψ_r) and fraction contribution (Ψ_f) as follows [20]:

$$\Psi_r(i) = c_i \overline{TI}_i \tag{6}$$

$$\Psi_f(i) = [\, R^2 \times |\Psi_r(i)| \,/ \sum_i |\Psi_r(i)| \,] \times 100\% \tag{7}$$

where c_i and $\overline{TI}_i$ are the coefficient and the average value of the *i*-th topological index entered in the model. The sum is over all indices in the model.

According to the results, standardized regression coefficients of the descriptors decreased in the order of $AT(^1\chi) > AT(–OH) > AT(–CH_3)$ for all the stationary phases. The $^1\chi$ index encodes information about the bulkiness and branching of molecules and its value increases with molecular size but decreases with molecular branching [21]. Positive standardized coefficient for $^1\chi$ index indicates that the molecule with higher $^1\chi$ value will have higher *RI* value due to larger molecular size or less branching. This relationship suggests that the main interaction force contributing to the chromatographic behavior of saturated alcohols on the stationary phases is dispersive. As known, the magnitude of this type of intermolecular solute-stationary phase interaction is related to molecular size as well as degree of branching [22].

The topological index $AT(–OH)$ had smaller standardized coefficient than $^1\chi$ for all the stationary phases. The descriptor that is a measure of solute polarity showed positive standardized coefficients for all the stationary phases as the hydroxyl group of the alcohol molecule makes the intermolecular interaction between solute and stationary phase stronger and increases the *RI* value. The $AT(–CH_3)$ topological index with minimum value of the standardized coefficient indicates the role of branching in determining *RI* of saturated alcohols, because $AI(–CH_3)$ index is clearly related to the number of methyl groups which is a crude measure of branching [20]. As shown, standardized coefficient for the descriptor is negative that indicates the larger the number of the branches of molecule is, the smaller its *RI* value because branching prevents solutes from close contact with stationary phase and reduces the interactions between the solutes and the stationary phase. Consequently, retention indices of the solutes decrease.

Relative contributions of the topological indices entered in the combined model to *RI* data showed the same trend as the standardized regression coefficients of the individual QSRR models. The obtained Ψ_r values for $^1\chi$, $AT(–OH)$ and $AT(–CH_3)$ were 701.58, 210.07 and -35.75, respectively. Moreover, McReynold polarity had much larger contribution (Ψ_f=103.07) than $AT(–CH_3)$ indicating the important role of the stationary phase polarity in determining retention indices of the model compounds. Positive sign of the parameter

shows that *RI* of saturated alcohols increases with increasing polarity of the stationary phase. Based on the obtained results, fraction contributions of bulkiness and polarity of the solutes to the retention indices were 66.45% and 19.90%, respectively. Furthermore, stationary phase polarity ($\Psi_f = 9.76\%$) and methyl groups of the alcohol molecules ($\Psi_f = 3.39\%$) showed smaller contributions to *RI* values. Results of the study showed that the topological descriptors included in the regression models provide useful information about structural features important in determining retention indices of saturated alcohols on the stationary phases of different polarity.

4. CONCLUSION

In the present study, preliminary individual and combined QSRR models for predicting Kovats retention indices of saturated alcohols on the stationary phases of different polarity were developed using a combination of $^{1}\chi$ index and the novel atom-type-based *AT* topological indices. According to the results, $^{1}\chi$ and *AT(–OH)* indices were the most important descriptors affecting *RI* of saturated alcohols showing important role of size, branching and polarity of the solutes molecules in determining their retention indices. In addition, polarity of the stationary phases showed relatively large contribution in the unified QSRR model developed for simultaneous prediction of retention indices of saturated alcohols on the stationary phases. Satisfactory prediction results evidently suggest efficiency of the *AT* based QSRR models for accurate estimation of the retention indices for similar compounds using only two dimensional structures of the molecules. This work can be a good starting point to provide a simple procedure for QSRR study of other heteroatom containing aliphatic compounds.

REFERENCES

1. R. Kaliszan, *Structure and Retention in Chromatography: A Chemometric Approach*, Harwood, Amsterdam, 1997.
2. E. Forgacs and T. Cserhati, *Molecular Bases of Chromatographic Separations*, CRC Press, Boca Raton, FL, 1997.
3. Q. N. Hu, Y. Z. Liang and K. T. Fang, The matrix expression, topological index and atomic attribute of molecular topological structure, *J. Data Sci.* **1** (2003) 361−389.
4. B. Ren, A new topological index for QSPR of alkanes, *J. Chem. Inf. Comput. Sci.* **39** (1999) 139−143.
5. B. Ren, Novel atom-type AI indices for QSPR studies of alcohols, *Comput. Chem.* **26** (2002) 223−235.

6. B. Ren, Application of novel atom-type AI topological indices to QSPR studies of alkanes, *Comput. Chem.* **26** (2002) 357−369.

7. K. Panneerselvam, M. P. Antony and T. G. Srinivasan, P. R. Vasudeva Rao, Estimation of normal boiling points of trialkyl phosphates using retention indices by gas chromatography, *Thermochim. Acta* **511** (2010) 107−111.

8. F. Safa and M. Yekta, Quantitative structure–property relationship study of standard formation enthalpies of acyclic alkanes using atom-type-based AI topological indices, *Arabian J. Chem.,* in press.

9. B. Ren, Atom-type-based AI topological descriptors for quantitative structure–retention index correlations of aldehydes and ketones, *Chemom. Intell. Lab. Sys.* **66** (2003) 29−39.

10. J. B. Pias and L. Gasco, GC retention data of alcohols and benzoyl derivatives of alcohols, *J. Chromatogr.-Chrome Data* **104** (1975) d14−d16.

11. B. Ren, Application of novel atom-type AI indices in the structure–property correlations, *J. Mol. Struct.* (*Theochem*) **586** (2002) 137−148.

12. L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure–Activity Studies. Letchworth*, Research Studies Press, 1986.

13. SPSS for windows, *The statistical package for IBMPC*, SPSS Inc, http://www.spss.com.

14. 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** (1995) 279−287.

15. R. Kramer, *Chemometric Techniques for Quantitative Analysis*, Marcel Dekker, New York, 1998.

16. T. O. Kollie and C. F. Poole, Influence of solute size and the non-polar interaction term on the selection of test solutes for the classification of stationary phase selectivity in gas chromatography, *J. Chromatogr. A* **556** (1991) 457−484.

17. F. Liu, C. Cao and B. Cheng, A Quantitative structure-property relationship (QSPR) study of aliphatic alcohols by the method of dividing the molecular structure into substructure, *Int. J. Mol. Sci.* **12** (2011) 2448−2462.

18. W. Guo, Y. Lu and X. M. Zheng, The predicting study for chromatographic retention index of saturated alcohols by MLR and ANN, *Talanta* **51** (2000) 479−488.

19. R. Kaliszan, *Quantitative Structure-Chromatographic Retention Relationships*, John Wiley, & Sons, New York, 1987.

20. D. E. Needham, I. C. Wei and P. G. Seybold, Molecular modeling of the physical properties of alkanes, *J. Am. Chem. Soc.* **110** (1988) 4186−4194.

21. M. Randić, The structural origin of chromatographic retention data, *J. Chromatogr.* **161** (1978) 1–14.
22. M. Randić, Characterization of molecular branching, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.