

CHROM. 17,104

## CALCULATION OF RETENTION INDICES BY MOLECULAR TOPOLOGY: CHLORINATED ALKANES

ALEKSANDAR SABLJIĆ\*

*\*Laboratory of Chemical Physics, National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20205 (U.S.A.) and Theoretical Chemistry Group, Department of Physical Chemistry, Institute Rudjer Bošković, P.O.B. 1016, 41001 Zagreb, Croatia (Yugoslavia)*

(First received March 19th, 1984; revised manuscript received August 2nd, 1984)

---

### SUMMARY

This study was undertaken to test the ability of the molecular connectivity model to predict retention indices using both statistical correlation coefficients and correctly predicted elution sequences as criteria of fit. The test was performed on three groups of chloroalkanes. Regression analyses show that the molecular connectivity model successfully predicts the retention indices of chlorinated alkanes on polar and non-polar stationary phases. However, first-order molecular connectivity indices alone are not sufficient, higher order indices are demonstrated to be necessary. The results also indicate that different structural features determine the retention index values of mono- and dichlorides. For monochlorides the major factor is the size of the alkyl chain, while for dichlorides the major factor is the topological relation between the two chlorine atoms. The comparison of the results obtained with the molecular connectivity model and the empirical additive scheme reveals several important advantages of the molecular connectivity approach.

---

### INTRODUCTION

During the last five years numerous investigators have noted<sup>1-17</sup> a very good correlation between the experimental retention indices and topological indices such as molecular connectivity indices<sup>18-21</sup> and Wiener numbers<sup>22</sup>. In addition, Dubois and co-workers have extensively analysed experimental retention indices of alkanes and alkenes<sup>23-26</sup> using the DARC topological system<sup>27,28</sup>. Above correlations suggest that topological indices may be an alternative to the empirical additive schemes developed to predict retention indices. An extensive reduction in the number of parameters necessary to obtain a good correlation with the experimental retention indices makes the topological approach far more favorable. Unfortunately, the only criterion used to test the fit between the observed and calculated retention indices in these investigations was the statistical correlation coefficient. This criterion is insufficient for testing the usefulness of the topological approach for predicting retention indices

since a high correlation coefficient does not necessarily imply a correct elution sequence. An excellent illustration of this failure is the study by Kaliszan and Lamparczyk<sup>4</sup> of 48 polycyclic aromatic hydrocarbons (PAHs). An extremely high correlation coefficient (0.9944) was obtained between first-order molecular connectivity indices (originally named branching indices<sup>1</sup>) and retention indices for the OV-101 stationary phase. However, an incorrect elution sequence was predicted for 20 of the compounds and it was not possible to discriminate between the ten pairs and three triplets of isomers. The latter is not surprising, since this index, which was the foundation of the latter developed connectivity model<sup>18-21</sup>, is only able to discriminate between alkanes having eight or fewer carbon atoms. In the case of unsubstituted PAHs, the first-order molecular connectivity index can only discriminate between isomers with up to three benzene rings. It thus appears that both a high correlation coefficient and a correctly predicted elution sequence are necessary prerequisites before any approach, empirical or theoretical, can be usefully and widely applied to predict the retention indices.

In the present investigation we examined two questions. First, can the connectivity model reliably predict retention indices? Second, if affirmative, how sophisticated or complex must be the model to fulfill both aforementioned criteria? To achieve these goals single and multi-variable linear regression models of various complexity were screened to find the simplest molecular connectivity model that generates the experimental elution sequence. The molecular connectivity indices were used as structural descriptors. In this study the molecular connectivity approach was tested on the three groups of compounds: (a) nineteen monochlorides with linear alkyl chains, (b) fifteen monochlorides with branched alkyl chains, and (c) fifteen dichlorides with linear alkyl chains. Selected groups of compounds have already been studied<sup>28</sup> with the empirical additive scheme. Thus a correlation of the results of both investigations will enable us to directly compare both the complexity and the precision of each model and to learn more about the advantages and limitations of each.

#### METHOD OF CALCULATION

All valence and non-valence molecular connectivity indices used in this study are displayed in Table I.

In the case of non-valence molecular connectivity indices all non-hydrogen atoms and bonds between them are considered to be identical. The valence type indices were introduced to account for the nature of non-hydrogen atoms and for differences in bond orders. Several extensive reviews have been published recently<sup>18-21</sup> which can be consulted for a detailed description of the theory and method of calculating these molecular connectivity indices. Only the valence subgraph index ( $\chi_s^v$ ) deserves more attention, since it is not commonly used in quantitative structure-property or structure-activity relationship studies. It refers to the topological relation between the two chlorine atoms in dichlorides. In the present study it is used to describe the path (subgraph) between the two chlorine atoms. Each non-hydrogen atom on this path is described by its valence delta value,  $\delta^v$ , which is calculated for second row atoms by the expression  $\delta^v = Z^v - h$ , where  $Z^v$  is the number of valence electrons in the atom and  $h$  is the number of hydrogen atoms bound to the same atom. For the atoms beyond the second row in the periodic chart, the correct expres-

TABLE I

LIST OF SYMBOLS FOR ALL MOLECULAR CONNECTIVITY INDICES\* USED IN THE REGRESSION ANALYSES

Type of index	Index order				
	1st	3rd	4th	5th	6th
Path	$^1\chi^v$		$^4\chi$	$^5\chi$	
Cluster**		$^3\chi_C$ $^3\chi_C^c$	$^4\chi_C$		
Path/Cluster***			$^4\chi_{PC}$ $^4\chi_{PC}^v$	$^5\chi_{PC}$ $^5\chi_{PC}^v$	$^6\chi_{PC}^v$

\* The left-side superscript designates the order of connectivity index. The right-side superscript differentiates between valence and non-valence types of indices. The subscript specifies the subclass of the molecular connectivity index which in the case of chlorinated alkanes may be path, cluster, and/or path/cluster type index. If no subscript is indicated it is assumed to be a path index.

\*\* The third-order cluster molecular connectivity index is the first (simplest) member of the cluster type indices.

\*\*\* The fourth-order path/cluster molecular connectivity index is the first (simplest) member of the path/cluster type indices.

sion is:

$$\delta^v = \frac{Z^v - h}{Z - Z^v} \quad (1)$$

where  $Z$  is their atomic number. The valence subgraph index is then calculated in accordance with the following expression:

$$\chi_s^v = (\delta_1^v \cdot \delta_2^v \cdot \delta_3^v \cdot \dots \cdot \delta_n^v)^{-0.5} \quad (2)$$

where  $n$  is equal to the number of non-hydrogen atoms, including both chlorine atoms, in a particular path (subgraph).

Molecular connectivity indices were calculated by a CFUNC computer program<sup>30</sup>. Single and multiple linear regression analyses were performed using the modeling laboratory (MLAB) components developed by the National Institutes of Health Division of Computer Research and Technology<sup>31</sup>. To test the quality of the regression equations, the following statistical parameters were used: the multiple correlation coefficient ( $r$ ), a test of the null-hypothesis ( $F$ -test), the amount of explained variance (EV), and the Student's  $t$ -test. All calculations were carried out on a DEC System-10 computer at the National Institutes of Health, Bethesda, Maryland.

The retention index (RI) values used in this study, reported by Morishita *et al.*<sup>29</sup>, were obtained with non-polar stationary phase Apiezon L (ApL) and with polar stationary phase polyethylene glycol 20M (PEG). The reproducibility of the RI values was 0.5 and 1.0 index units, respectively.

## RESULTS AND DISCUSSION

Molecular connectivity indices and observed retention indices of nineteen mono-

TABLE II

MOLECULAR CONNECTIVITY INDICES AND OBSERVED RETENTION INDICES OF MONOCHLORIDES WITH LINEAR ALKYL CHAINS USED IN REGRESSION EQNS. 3, 4 and 8

Compound	$^1\chi^v$	$^4\chi$	$^5\chi$	$^6\chi$	$^4\chi_{PC}^v$	$^5\chi_{PC}$	$^5\chi_{PC}^v$	$^6\chi_{PC}^v$	RI	
									ApL	PEG
1-Chloropentane	3.052	0.500	0.250	0.0	0.0	0.0	0.0	0.0	762	953
1-Chlorohexane	3.552	0.677	0.345	0.177	0.0	0.0	0.0	0.0	857	1050
1-Chloroheptane	4.052	0.854	0.479	0.250	0.0	0.0	0.0	0.0	958	1151
1-Chlorooctane	4.552	1.030	0.604	0.338	0.0	0.0	0.0	0.0	1059	1251
1-Chlorononane	5.052	1.207	0.729	0.427	0.0	0.0	0.0	0.0	1159	1350
2-Chloropentane	2.883	0.577	0.0	0.0	0.345	0.289	0.345	0.0	712	884
2-Chlorohexane	3.383	0.612	0.408	0.0	0.345	0.204	0.244	0.244	809	978
2-Chloroheptane	3.883	0.803	0.433	0.289	0.345	0.204	0.244	0.173	908	1077
2-Chlorooctane	4.383	0.979	0.568	0.306	0.345	0.204	0.244	0.173	1006	1177
2-Chlorononane	4.883	1.156	0.693	0.401	0.345	0.204	0.244	0.173	1106	1275
3-Chloropentane	2.921	0.289	0.0	0.0	0.690	0.289	0.345	0.0	721	892
3-Chlorohexane	3.421	0.697	0.204	0.0	0.589	0.408	0.488	0.244	812	974
3-Chloroheptane	3.921	0.757	0.493	0.144	0.589	0.348	0.416	0.345	906	1069
3-Chlorooctane	4.421	0.947	0.535	0.348	0.589	0.348	0.416	0.295	1006	1166
3-Chlorononane	4.921	1.124	0.670	0.378	0.589	0.348	0.416	0.295	1106	1264
4-Chloroheptane	3.921	1.130	0.289	0.144	0.488	0.553	0.660	0.345	900	1054
4-Chlorooctane	4.421	1.190	0.595	0.204	0.488	0.493	0.589	0.467	997	1149
4-Chlorononane	4.921	1.380	0.637	0.421	0.488	0.493	0.589	0.416	1094	1245
5-Chlorononane	4.921	1.250	0.901	0.276	0.488	0.433	0.518	0.589	1092	1243

chlorides with linear alkyl chains examined in the present study are shown in Table II.

The best one-variable equations were obtained with the valence first-order molecular connectivity index,  $^1\chi^v$ , for both retention indices,  $RI^{ApL}$  and  $RI^{PEG}$ . (The correlation coefficients are 0.996 and 0.980, respectively.) Although very high correlation coefficients were obtained (especially with RI from the non-polar stationary phase), the  $^1\chi^v$  index alone was not able to correctly predict elution sequences for the nineteen monochlorides with linear alkyl chains. Consequently, multi-variable regression equations were screened to find the simplest equation that generated the experimental elution sequence. Both the order and number of connectivity indices were varied. For  $RI^{ApL}$  (non-polar stationary phase), a three-variable equation (eqn. 3) is necessary, while for  $RI^{PEG}$  (polar stationary phase) a four-variable equation is needed (eqn. 4).

$$RI^{ApL} = 204.4 \cdot ^1\chi^v - 9.4 \cdot ^4\chi - 69.7 \cdot ^6\chi_{PC}^v + 134.8 \quad (3)$$

$$N = 19 \quad r^2 = 0.999 \quad F^{3,15} = 5150 \quad EV = 99.9\%$$

$$RI^{PEG} = 233.5 \cdot ^1\chi^v - 83.6 \cdot ^5\chi - 71.1 \cdot ^6\chi - 138.7 \cdot ^5\chi_{PC}^v + 260.2 \quad (4)$$

$$N = 19 \quad r^2 = 1.000 \quad F^{4,14} = 8866 \quad EV = 99.9\%$$

The statistical analyses demonstrate that excellent correlations between the

selected topological indices and the retention indices are obtained for both polar and non-polar stationary phases. Both regression equations are statistically significant above the 99.9% level. All variables are statistically significant above the 99.95% level (Student's *t*-test), except the  ${}^4\chi$  index, which is significant above the 78% level. The regression analyses show that the size of alkyl chain is the major factor that governs the magnitude of the RIs, which is described best by the  ${}^1\chi^v$  index, since its numerical value is directly proportional to the number of bonds in a molecule and it accounts for the most variation in the RI data. The other factors that control the magnitudes of the RIs are the position of the chlorine atom and flexibility of the alkyl chain. Compared with the main factor they can be viewed as a fine tuning elements for the RIs, especially for the non-polar stationary phase. The flexibility of the alkyl chains is well described by the higher-order path type indices, which is consistent with the recently proposed<sup>32</sup> physical meaning of the third-order molecular connectivity index. This index was found to relate closely to the number of possible *gauche-trans* rearrangements and the density of the alkanes. The position of the chlorine atom along the alkyl chain is quantified by the path/cluster type indices, which are highly sensitive to substitutional variations along an alkyl chain<sup>33,34</sup>. The negative regression coefficients for all higher-order indices are consistent with the observation that all secondary chlorides have lower RI values than their primary isomers. Thus we conclude that the resulting RI is proportional to the size of the alkyl chain and slightly dependent on the position of the chlorine atom and flexibility of the alkyl chain. A graphical representation of the fit between observed RIs and those calculated by eqns. 3 and 4, are given in Figs. 1 and 2, respectively.

Molecular connectivity indices and observed retention indices of fifteen monochlorides with branched alkyl chains investigated in the present study are shown in Table III. As in the case of monochlorides with linear alkyl chains, the  ${}^1\chi^v$  index accounts for most variations in the RI data of monochlorides with branched alkyl chains, but a three-variable equation (eqn. 5) was needed to generate the experimental elution sequence.

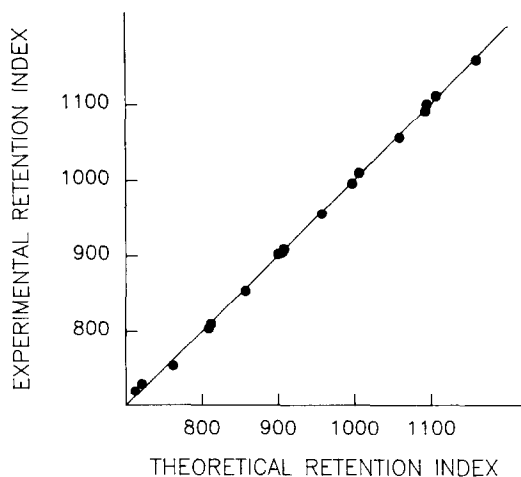


Fig. 1. Correlation between experimental (non-polar stationary phase, Apiezon L) and calculated (eqn. 3) retention indices of nineteen monochlorides with linear alkyl chains.

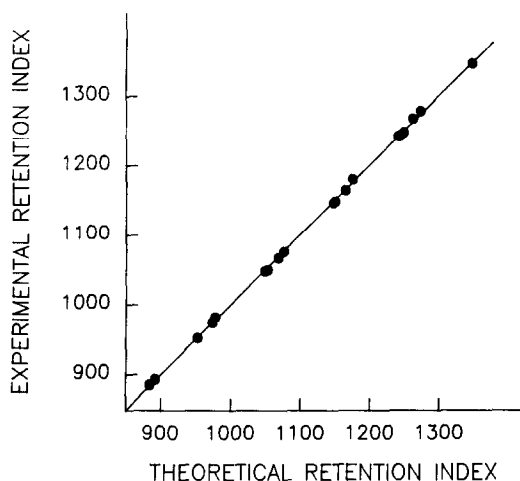


Fig. 2. Correlation between experimental (polar stationary phase, polyethylene glycol 20M) and calculated (eqn. 4) retention indices of nineteen monochlorides with linear alkyl chains.

$$RI^{ApL} = 188.2 \cdot {}^1\chi^v - 56.6 \cdot {}^4\chi_{PC} + 11.9 \cdot {}^5\chi_{PC} + 200.4 \quad (5)$$

$$N = 15 \quad r^2 = 0.998 \quad F^{3,11} = 1972 \quad EV = 99.8\%$$

Statistically, eqn. 5 is significant above the 99.9% level, while the  ${}^1\chi^v$  and  ${}^4\chi_{PC}$  indices are significant above the 99.95% level and the  ${}^5\chi_{PC}$  index above the 97.5% level. The size of alkyl chain is again the main factor that determines the magnitude of the RI values. The other factors influencing the RI data are the absolute and the relative positions of the chlorine atom and the methyl group. In our model, the fine balance between these three secondary effects is well described by the difference between the

TABLE III

MOLECULAR CONNECTIVITY INDICES AND OBSERVED RETENTION INDICES OF MONOCHLORIDES WITH BRANCHED ALKYL CHAINS USED IN REGRESSION EQNS. 5 and 8

Compound	${}^1\chi^v$	${}^4\chi$	${}^4\chi_{PC}$	${}^4\chi_{PC}^v$	${}^5\chi_{PC}$	${}^6\chi_{PC}^v$	$RI^{ApL}$
1-Chloro-2-methylpentane	3.446	0.697	0.493	0.549	0.408	0.244	823
1-Chloro-2-methylhexane	3.946	0.757	0.493	0.549	0.348	0.317	921
1-Chloro-2-methylheptane	4.446	0.947	0.493	0.549	0.348	0.275	1091
1-Chloro-4-methylpentane	3.408	0.612	0.289	0.289	0.204	0.244	823
1-Chloro-6-methylheptane	4.408	0.979	0.289	0.289	0.204	0.144	1021
2-Chloro-2-methylpentane	3.158	0.750	0.750	0.848	1.000	0.299	763
2-Chloro-2-methylhexane	3.658	0.707	0.750	0.848	0.780	0.811	855
2-Chloro-2-methylheptane	4.158	0.905	0.750	0.848	0.780	0.635	949
2-Chloro-4-methylpentane	3.239	0.943	0.471	0.517	0.943	0.282	800
4-Chloro-2-methylheptane	4.277	1.415	0.606	0.679	1.058	0.740	981
2-Chloro-6-methylheptane	4.239	0.934	0.577	0.634	0.408	0.259	966
3-Chloro-2-methylpentane	3.293	0.471	1.276	1.479	0.707	0.282	763
3-Chloro-2-methylhexane	3.793	0.789	1.207	1.397	0.667	0.764	850
3-Chloro-2-methylheptane	4.293	0.859	1.207	1.397	0.618	0.740	947
3-Chloro-6-methylheptane	4.277	0.822	0.781	0.878	0.489	0.740	966

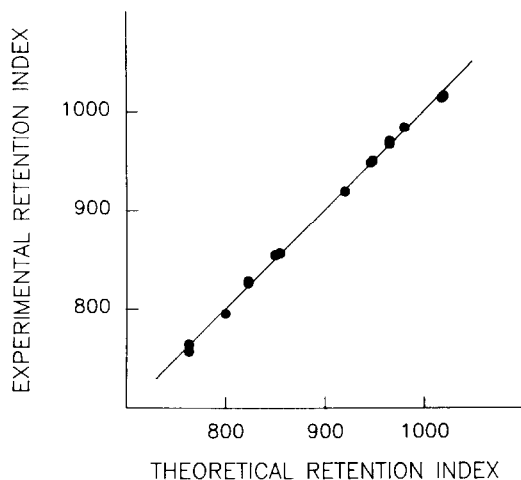


Fig. 3. Correlation between experimental (non-polar stationary phase, Apiezon L) and calculated (eqn. 5) retention indices of fifteen monochlorides with branched alkyl chains.

fourth- and fifth-order path/cluster molecular connectivity indices. Eqn. 5 also suggests that the chlorine and methyl group have opposite effects on the RI values. A graphical representation of the fit between observed and calculated RI values is shown in Fig. 3.

Molecular connectivity indices and observed retention indices of fifteen dichlorides with linear alkyl chains, are presented in Table IV. *Erythro* and *threo* isomers were excluded from the analysis since the molecular connectivity indices can

TABLE IV

MOLECULAR CONNECTIVITY INDICES AND OBSERVED RETENTION INDICES OF DICHLORIDES WITH LINEAR ALKYL CHAINS USED IN REGRESSION EQNS. 6 AND 7

Compound	${}^1\chi^v$	$\chi_s^v$	${}^3\chi_c$	${}^4\chi_c$	${}^3\chi_c^v$	RI	
						ApL	PEG
1,1-Dichloropentane	3.495	0.825	0.408	0.0	0.583	889	1144
1,2-Dichloropentane	3.559	0.583	0.289	0.0	0.345	909	1218
1,3-Dichloropentane	3.559	0.412	0.289	0.0	0.345	940	1272
1,4-Dichloropentane	3.521	0.292	0.408	0.0	0.488	959	1337
1,5-Dichloropentane	3.690	0.253	0.0	0.0	0.0	1028	1444
2,2-Dichloropentane	3.256	0.714	1.561	0.354	2.065	831	1043
3,3-Dichloropentane	3.317	0.714	1.207	0.250	1.608	857	1074
1,1-Dichlorohexane	3.995	0.825	0.408	0.0	0.583	989	1240
1,2-Dichlorohexane	4.059	0.583	0.289	0.0	0.345	1005	1311
1,3-Dichlorohexane	4.059	0.412	0.289	0.0	0.345	1028	1347
1,4-Dichlorohexane	4.059	0.292	0.289	0.0	0.345	1058	1419
1,5-Dichlorohexane	4.021	0.206	0.408	0.0	0.488	1072	1456
1,6-Dichlorohexane	4.190	0.179	0.0	0.0	0.0	1129	1536
2,2-Dichlorohexane	3.756	0.714	1.561	0.354	2.065	925	1136
3,3-Dichlorohexane	3.817	0.714	1.207	0.250	1.608	944	1144

not discriminate between these isomers. To our surprise, the relation between the two chlorine atoms, not the size of the alkyl chain, was found to be the dominant topological factor that determines the RI values of dichlorides on both non-polar and polar phases. However, to generate the experimental elution sequences, four-variable (eqn. 6) and three-variable (eqn. 7) regression equations were necessary for the non-polar and polar phases, respectively.

$$RI^{ApL} = 174.3 \cdot {}^1\chi^v + 24.5/\chi_s^v - 45.0 \cdot {}^3\chi_C + 89.4 \cdot {}^4\chi_C + 273.5 \quad (6)$$

$$N = 15 \quad r^2 = 0.992 \quad F^{4,10} = 309 \quad EV = 98.9\%$$

$$RI^{PEG} = 137.5 \cdot {}^1\chi^v + 59.0/\chi_s^v - 73.2 \cdot {}^3\chi_C^v + 665.0 \quad (7)$$

$$N = 15 \quad r^2 = 0.982 \quad F^{3,11} = 197 \quad EV = 97.7\%$$

Both equations are statistically significant above 99.9%, while all variables are statistically significant above the 99.95% level except the  ${}^4\chi_C$  index, which is significant above the 85% level. As discussed above, the RI values of dichlorides are directly proportional to the separation between the two chlorine atoms. The size of alkyl chain is the next most significant factor influencing the magnitude of the RI values, while the topological environment of the chlorine atoms, described by the third- and fourth-order cluster indices, is found to account for only a small part of the variation of the RI values. Figs. 4 and 5 show the correlation between the observed retention indices and those calculated by eqns. 6 and 7, respectively.

Similarly as dichlorides, chlorinated esters<sup>35-39</sup> are eluted in order of increased distance of the chlorine substituent from the carbonyl group. However, this result is only valid within the isomeric series of chlorinated esters.

When the order of the indices necessary to reproduce the experimental elution sequence for each group of chlorinated alkanes is compared with the size of corresponding alkyl chains an interesting relation emerges. Sixth-order indices are seen to be needed for linear monochlorides with up to nine carbon atoms, fifth-order indices

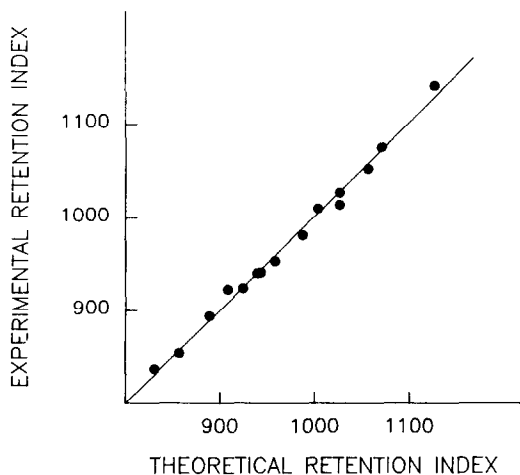


Fig. 4. Correlation between experimental (non-polar stationary phase, Apiezon L) and calculated (eqn. 6) retention indices of fifteen dichlorides with linear alkyl chains.



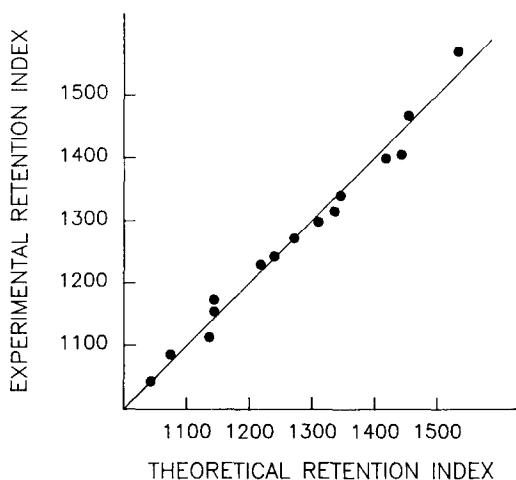


Fig. 5. Correlation between experimental (polar stationary phase, PEG) and calculated (eqn. 7) retention indices of fifteen dichlorides with linear alkyl chains.

are needed for branched monochlorides having up to eight carbon atoms, while fourth-order indices are found to be sufficient for dichlorides with up to six carbon atoms. Thus, it appears that the highest order connectivity index needed in each regression equation is closely related to the alkyl chain length of the described group. This study clearly demonstrates relation between the order of the connectivity indices and the size of molecules for the first time, although it was intuitively perceived that this relation must exist. This result can be used as a guideline in future searches for a minimal set of molecular connectivity indices to adequately describe chromatographic behavior of molecules.

The correlation analyses presented here demonstrate that the connectivity model is highly suitable for the prediction of retention indices. It also shows that the first-order molecular connectivity index is insufficient where effects other than molecular size affect the magnitudes of the retention indices. Generally a three or four variables model is necessary to both generate experimental elution sequences and obtain a high degree of correlation.

Since the connectivity model successfully reproduces the experimental retention indices of chlorinated alkanes, our next step is to correlate its results with those obtained by the empirical additive scheme<sup>29</sup>. The criteria used to compare the major features of these models are multiple correlation coefficients, the number of parameters used in each regression analysis ( $N_p$ ), and the number of compounds ( $N_c$ ) whose elution sequence is incorrectly calculated. They are all presented in Table V.

An examination of Table V clearly demonstrates two important results: (1) both models give almost identical correlation coefficients, and (2) the empirical additive scheme requires two to four times as many parameters to attain the same correlation level as the connectivity model. The second result emphasizes the most serious limitation of the empirical additive scheme: the number of parameters needed to generate experimental retention indices increases rapidly with the number of functional groups or any other structural variation in any class of studied compounds. Thus experimental retention indices have to be measured for a large number of com-

TABLE V

COMPARISON OF THE RESULTS OBTAINED BY THE TOPOLOGICAL APPROACH AND THE EMPIRICAL ADDITIVE SCHEME FOR THE RETENTION INDICES OF CHLORINATED ALKANES

Multiple correlation coefficients ( $r$ ), number of parameters used in each regression analysis ( $N_P$ ), and number of compounds whose elution sequence is incorrectly predicted ( $N_C$ ) are used to compare and evaluate the models.

Test sample	Stationary phase	Topological model			Empirical additive scheme		
		$r$	$N_P$	$N_C$	$r$	$N_P$	$N_C$
Nineteen monochlorides with the linear alkyl chains	ApL*	1.000	3	0	1.000	7	0
	PEG**	1.000	4	0	1.000	7	0
Fifteen monochlorides with the branched alkyl chains	ApL	0.999	3	0	0.987	14	3***
Fifteen dichloroalkanes with the linear alkyl chains	ApL	0.996	4	0	0.998	13	1 <sup>§</sup>
	PEG	0.991	3	0	0.999	13	1 <sup>§</sup>

\* ApL = Non-polar stationary phase Apiezon L.

\*\* PEG = Polar stationary phase polyethylene glycol 20M.

\*\*\* 2-Chloro-4-methylpentane, 2-chloro-2-methylhexane and 4-chloro-2-methylheptane.

§ 3,3-Dichlorohexane

pounds to provide a useful and meaningful empirical additive scheme. Note that the number of additive parameters used to calculate the RIs of dichlorides and monochlorides with branched alkyl chains approximates the number of experimental data (RI values) from which they are derived. It is thus possible that some correlation between additive parameters and experimental RI values could emerge by chance<sup>40</sup>. Despite the large number of parameters, the empirical additive scheme can not reproduce the experimental elution sequences for dichlorides and monochlorides with branched alkyl chains, for which three out of fifteen compounds were misplaced, and the same RIs were calculated for several groups of compounds with different observed RIs.

In addition to this practical advantage, the molecular connectivity approach also provides physical insight into the relation between different structural features to their chromatographic behaviour. For example, the connectivity approach demonstrated that the relation between the two chlorine atoms is more important than the size of their alkyl chains for the chromatographic behavior of the dichlorides, which the additive scheme can not reveal.

The highly satisfactory performance of the connectivity model obtained for chlorinated alkanes prompted us to try combining two groups of studied molecules into a single regression model. The obvious choice was the 34 monochlorides and their retention indices obtained with non-polar stationary phase. The best result was obtained with the five-variable regression equation.

$$\begin{aligned} \text{RI}^{\text{APL}} = & 215.0 \cdot {}^1\chi^{\text{v}} - 57.2 \cdot {}^4\chi - 55.8 \cdot {}^4\chi_{\text{PC}}^{\text{v}} + 54.2 \cdot {}^5\chi_{\text{PC}} \\ & - 25.9 \cdot {}^6\chi_{\text{PC}}^{\text{v}} + 133.6 \quad (8) \\ N = & 34 \quad r^2 = 0.998 \quad F^{5,28} = 3601.0 \quad \text{EV} = 99.8\% \end{aligned}$$

The connectivity indices from eqn. 8 are the same as those used as structural descriptors in the individual regression eqns. 3 and 5. Eqn. 8 is statistically significant above 99.9%, while all its variables are significant above the 99.95% level. The unified regression model successfully predicted the experimental elution sequence of 34 monochlorides. Only one compound (3-chloroheptane) was misplaced. Such a unification of monochlorinated alkanes can not be achieved within the framework of the empirical additive scheme, since different values for the same structural subunit are generated in the two classes of monochloroalkanes, discussed above.

A further attempt to describe all the chloroalkanes by one regression model was unsuccessful. The major reason for this is probably the difference in interaction patterns between the mono- and dichlorides with the stationary phase molecules.

## CONCLUSION

In this investigation the molecular connectivity model has been shown to successfully predict the retention indices of chlorinated alkanes on polar and non-polar stationary phases. However, to obtain reliable results a more complex connectivity model than was originally suggested<sup>2-4</sup> was found to be necessary. Our results also demonstrate that different structural features determine the RI values of mono- and dichlorides. For monochlorides the major factor is the size of alkyl chain, for dichlorides the major factor is the topological relation between the two chlorine atoms. This result strongly suggests that different interaction patterns occur between mono- and dichlorides and the stationary phase molecules. It is also in contrast with the result for chlorinated esters<sup>3,5-38,41</sup> whose chromatographic behavior on non-polar columns is predominately determined by their boiling points (*i.e.* the size of molecules). A comparison of the results obtained with the molecular connectivity model and the empirical additive scheme reveals several important advantages of the former, which are discussed in detail in the previous section. However, the molecular connectivity model also has limitations. At present, it can not cope with three-dimensional problems, like isomerism around single bonds, optical activity, or *cis-trans* isomerism. Its second weakness is the limited number of applications carried out thus far. This obstacle will be overridden as its use expands. Although the molecular connectivity model performed better than the empirical additive scheme for chlorinated alkanes, the two calculations should be regarded as complementary rather than opposing approaches. Thus, the most desirable approach would be to apply both methods to the same problem whenever possible. The advantages of both models, empirical and non-empirical, can thus be combined, thereby increasing the probability of creating a reliable, predictive model, and learning more about the relations between molecular structure and chromatographic behavior.

## REFERENCES

- 1 M. Randić, *J. Amer. Chem. Soc.*, 97 (1975) 6609.
- 2 R. Kaliszan and H. Foks, *Chromatographia*, 10 (1977) 346.
- 3 R. Kaliszan, *Chromatographia*, 10 (1977) 529.
- 4 R. Kaliszan and H. Lamparczyk, *J. Chromatogr. Sci.*, 16 (1978) 246.
- 5 M. Randić, *J. Chromatogr.*, 161 (1978) 1.
- 6 J. S. Millership and A. D. Woolfson, *J. Pharm. Pharmacol.*, 30 (1978) 483.
- 7 D. Bonchev, Ov. Mekenjan, G. Protić and N. Trinajstić, *J. Chromatogr.*, 176 (1979) 149.
- 8 A. Radecki, H. Lamparczyk and R. Kaliszan, *Chromatographia*, 12 (1979) 595.
- 9 T. R. McGregor, *J. Chromatogr. Sci.*, 17 (1979) 314.
- 10 L. B. Kier and L. H. Hall, *J. Pharm. Sci.*, 68 (1979) 120.
- 11 M. Gassiot-Matas and G. Firpo-Pamies, *J. Chromatogr.*, 187 (1980) 1.
- 12 K. D. Bartle, M. L. Lee and S. A. Wise, *Chromatographia*, 14 (1981) 69.
- 13 M. J. M. Walls, C. R. Clark and R. M. Patterson, *J. Chromatogr. Sci.*, 19 (1981) 573.
- 14 J. Bojarski and L. Ekiert, *Chromatographia*, 15 (1982) 172.
- 15 R. J. Hurtubise, T. W. Allen and H. F. Silver, *J. Chromatogr.*, 235 (1982) 517.
- 16 Gy. Szász, K. Valkó, O. Papp and I. Hermeecz, *J. Chromatogr.*, 243 (1982) 347.
- 17 Gy. Szász, O. Papp, J. Vámos, K. Hanko-Novák and L. B. Kier, *J. Chromatogr.*, 269 (1983) 91.
- 18 L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
- 19 A. T. Balaban, A. Chiriac, I. Motoc and Z. Simon, *Streic Fit in Quantitative Structure-Activity Relationships. Lecture Notes in Chemistry No. 15*, Springer Verlag, Berlin, 1980.
- 20 A. Sabljčić and N. Trinajstić, *Acta Pharm. Jugoslav.*, 31 (1981) 189.
- 21 N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, FL, 1983.
- 22 H. Wiener, *J. Amer. Chem. Soc.*, 69 (1947) 17 and 2636.
- 23 J. É. Dubois and J. R. Chrétien, *J. Chromatogr. Sci.*, 12 (1974) 811.
- 24 J. R. Chrétien and J. É. Dubois, *J. Chromatogr.*, 126 (1976) 171.
- 25 J. R. Chrétien and J. É. Dubois, *Anal. Chem.*, 49 (1977) 747.
- 26 J. É. Dubois, J. R. Chrétien, L. Soják and J. A. Rijks, *J. Chromatogr.*, 194 (1980) 121.
- 27 J. É. Dubois, D. Laurent and A. Aranda, *J. Chim. Phys.*, 70 (1973) 1608.
- 28 J. É. Dubois, D. Laurent and A. Aronda, *J. Chim. Phys.*, 70 (1973) 1616.
- 29 F. Morishita, Y. Terashima, M. Ichise and T. Kojima, *J. Chromatogr. Sci.*, 21 (1983) 209.
- 30 L. H. Hall and L. B. Kier, *Eur. J. Med. Chem.-Chim. Ther.*, 16 (1981) 399.
- 31 G. D. Knott, *Comput. Prog. in Biomedicine*, 10 (1979) 271.
- 32 L. B. Kier and L. H. Hall, *Quant. Struct.-Act. Relat.*, 2 (1983) 55.
- 33 L. B. Kier, *J. Pharm. Sci.*, 69 (1980) 1034.
- 34 A. Sabljčić and M. Protić-Sabljić, *Mol. Pharmacol.*, 23 (1983) 213.
- 35 I. O. O. Korhonen, *J. Chromatogr.*, 219 (1981) 306.
- 36 I. O. O. Korhonen, *J. Chromatogr.*, 246 (1982) 241.
- 37 I. O. O. Korhonen, *J. Chromatogr.*, 248 (1982) 69.
- 38 K. Komárek, L. Hornová and J. Churáček, *J. Chromatogr.*, 244 (1982) 142.
- 39 J. K. Haken, *J. Chromatogr.*, 250 (1982) 96.
- 40 J. G. Topliss and R. P. Edwards, *J. Med. Chem.*, 22 (1979) 1238.
- 41 J. K. Haken and I. O. O. Korhonen, *J. Chromatogr.*, 268 (1983) 511.