

CHROM. 17,333

CALCULATION OF RETENTION INDICES BY MOLECULAR TOPOLOGY CHLORINATED BENZENES

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. Box 1016, 41001 Zagreb, Croatia (Yugoslavia)*

(First received August 29th, 1984; revised manuscript received October 27th, 1984)

SUMMARY

A comparative study was undertaken to test the ability of several different topological indices to predict the retention indices of chlorinated benzenes on polar and non-polar stationary phases using both correlation coefficients and correctly predicted elution sequences as criteria of fit. The test was performed on three topological indices: connectivity indices, Wiener numbers, and Balaban indices. The regression analyses showed that the molecular connectivity model predicted the retention indices of chlorinated benzenes more successfully than either Wiener numbers or Balaban indices. The results also demonstrated that the major structural property controlling chromatographic behavior was the size of the chlorinated benzene. In addition, the use of the new non-empirical heteroatom parameterization scheme in the calculation of Wiener numbers and Balaban indices was successfully tested for the first time.

INTRODUCTION

Molecular topology was shown to be a very important structural parameter for describing chromatographic¹⁻⁸ and environmental⁹⁻¹² behavior of chemicals. In my recent article¹ it was demonstrated that the molecular connectivity model¹³⁻¹⁵ successfully predicts retention indices and elution sequences of chlorinated alkanes on non-polar and polar stationary phases and performs much better than the empirical additive scheme. Now I want to test the ability of the molecular connectivity model to describe correctly the chromatographic behavior of cyclic, aromatic compounds. Chlorinated benzenes were selected for this test because they, and other chlorinated hydrocarbons, constitute the major group of environmental hazards and a detailed understanding of their chromatographic behavior will help us obtain better insight into their environmental fate. In addition to being important ecologically, chlorinated benzenes provide a challenge to the molecular connectivity model because the effect of chlorine substitution was shown by Haken and Korhonen¹⁶ to be non-

linear (*i.e.* the incremental effect of chlorine substitution shows strong increase with the level of chlorine substitution). This result indicates that the empirical additive scheme will not be able to reproduce adequately the retention indices of chlorinated benzenes unless a large number of parameters are employed.

The second objective of this investigation is to test the ability of two other topological indices, Wiener numbers¹⁷ and Balaban indices¹⁸, to describe correctly the chromatographic behavior of chlorinated benzenes and to correlate their results with those obtained for the connectivity indices. The Wiener number method is selected because of its initial successful application in predicting retention indices of mono- and dialkylbenzenes¹⁹. Strangely enough it was never used again. The Balaban index method is selected because it has the highest discriminatory power of all proposed topological indices, *i.e.* it can discriminate between all isomeric alkanes having eleven or fewer carbon atoms. The comparative study proposed here was not possible until recently, when the first non-empirical parameterization scheme was proposed²⁰ that can be used in the calculation of Wiener numbers and Balaban indices for molecules containing heteroatoms. In addition to the above objectives, the current study will also test the proposed heteroatoms parameterization scheme for the first time.

To achieve these objectives the molecular connectivity indices, Wiener numbers, and Balaban indices are calculated for benzene and its chlorinated derivatives. For the former, the single- and multi-variable linear regression models of increased complexity are screened to find the simplest molecular connectivity model that generates the experimental elution sequence. The Wiener numbers and Balaban indices, calculated with and without the heteroatom parameterization scheme, are also correlated with the experimental retention indices. The results of all these studies are then compared, using correlation coefficients and the number of compounds whose elution sequence is incorrectly predicted as criteria of fit. This will enable us to assess directly the performances of each topological index as well as to evaluate the quality of the non-empirical heteroatoms parameterization scheme for Wiener number and Balaban index.

METHOD OF CALCULATION

Several extensive reviews have been published recently^{13-15,21} which give detailed descriptions of the theory and method of calculation of all topological indices used in the present investigation. Thus, only a brief summary thereof will be given in this section.

The following molecular connectivity indices were found useful in correlations with the experimental retention indices of chlorinated benzenes: the first- and third-order molecular connectivity indices ($^1\chi$ and $^3\chi$), the fourth-order path/cluster molecular connectivity index ($^4\chi_{PC}$), and the seventh-order chain molecular connectivity index ($^7\chi_{CH}$). The $^1\chi$ index is calculated as the sum of contributions from all bonds. $^3\chi$ is calculated from all possible combinations of three consecutive bonds. The $^4\chi_{PC}$ index corresponds to isobutane graphs and the $^7\chi_{CH}$ index corresponds to monosubstituted benzene graphs. Each of the latter indices ($^4\chi_{PC}$ and $^7\chi_{CH}$) is the first member of its respective series of connectivity indices.

The Wiener number (W) was defined¹⁷ as the number of bonds between all

pairs of non-hydrogen atoms in a molecule. It can be easily calculated from the distance matrix of the molecular graph. In its original scheme¹⁷ it is equal to half the sum of the off-diagonal elements of the distance matrix, while in the new version²⁰ (with the non-empirical parameterization for heteroatoms) the sum of all diagonal elements has to be added.

The Balaban index (J)¹⁸, known also as the average distance sum connectivity index, is also based on the distance matrix of the molecular graph. Each non-hydrogen atom is described by its numerical value which is equal to sum of the elements of the corresponding row of the distance matrix. The J index is then calculated from these sums and factors dependent on the molecule size and number of rings.

The experimental retention indices (I) used in this investigation are from the recent study of Haken and Korhonen¹⁶. They were obtained with the non-polar stationary phase SE-30 at 120°C and with the polar stationary phase Carbowax 20M (CW20M) at 140°C. The sets of experimental data selected for our investigation are those in which the best separations are obtained.

Molecular connectivity indices were calculated by a CFUNC computer program²². Single and multiple linear regression analyses were performed using the modeling laboratory (MLAB) components developed by the NIH Division of Computer Research and Technology²³. 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 Student's t -test. All calculations were carried out on a DECsystem10 computer at the National Institutes of Health, Bethesda, Maryland.

RESULTS AND DISCUSSION

The topological indices and observed retention indices of benzene and its twelve chlorinated derivatives examined in the present study are shown in Tables I and II, respectively.

TABLE I
TOPOLOGICAL INDICES OF BENZENE AND ITS TWELVE CHLORINATED DERIVATIVES

Compound	Connectivity indices				Wiener number		Balaban index	
	${}^1\chi$	${}^3\chi$	${}^4\chi_{PC}$	${}^7\chi_{CH}$	W	$W(het)^*$	J	$J(het)^*$
Benzene	3.000	1.500	0.0	0.0	27	18.0	3.0000	3.0000
Chlorobenzene	3.394	1.894	0.408	0.102	42	26.8	3.0215	3.3438
1,3-Dichlorobenzene	3.788	2.199	0.742	0.167	61	37.6	3.0777	3.6286
1,4-Dichlorobenzene	3.788	2.305	0.816	0.167	62	38.2	3.0325	3.3572
1,2-Dichlorobenzene	3.805	2.540	1.138	0.167	60	36.9	3.1349	3.7205
1,3,5-Trichlorobenzene	4.182	2.414	1.000	0.204	84	50.4	3.1657	3.8935
1,2,4-Trichlorobenzene	4.198	2.860	1.478	0.204	84	50.4	3.1717	3.9044
1,2,3-Trichlorobenzene	4.215	3.114	1.734	0.204	82	49.1	3.2478	4.0365
1,2,3,5-Tetrachlorobenzene	4.609	3.343	2.006	0.222	110	64.6	3.3288	4.2148
1,2,4,5-Tetrachlorobenzene	4.609	3.430	2.154	0.222	110	64.6	3.2824	4.1634
1,2,3,4-Tetrachlorobenzene	4.626	3.702	2.341	0.222	109	64.0	3.3422	4.2730
Pentachlorobenzene	5.037	4.197	2.887	0.227	142	82.2	3.4555	4.5096
Hexachlorobenzene	5.464	4.976	3.643	0.222	174	99.2	3.6035	4.6286

* Wiener numbers and Balaban indices calculated with the heteroatoms parameterization scheme²⁰.

TABLE II

OBSERVED RETENTION INDICES OF BENZENE AND ITS CHLORINATED DERIVATIVES ON NON-POLAR (SE-30) AND POLAR (CARBOWAX 20M) STATIONARY PHASE

Compound	Retention indices*	
	SE-30	Carbowax 20M
Benzene	667**	988***
Chlorobenzene	832	1257
1,3-Dichlorobenzene	964	1415
1,4-Dichlorobenzene	970	1438
1,2-Dichlorobenzene	1005	1447
1,3,5-Trichlorobenzene	1131	1515
1,2,4-Trichlorobenzene	1177	1630
1,2,3-Trichlorobenzene	1211	1705
1,2,3,5-Tetrachlorobenzene	1326	1754
1,2,4,5-Tetrachlorobenzene	1326	1764
1,2,3,4-Tetrachlorobenzene	1366	1871
Pentachlorobenzene	1496	1956
Hexachlorobenzene	1656	2124

* Observed retention indices of chlorobenzenes are those reported in ref. 16.

** Extrapolated from ref. 26.

*** Extrapolated from ref. 27.

The best one-variable equations were obtained with the first-order molecular connectivity index, ${}^1\chi$, for both retention indices, I^{SE-30} and I^{CW20M} . Although very high correlation coefficients were obtained (0.997 and 0.985), the ${}^1\chi$ index alone was not able to discriminate between one (1,3- and 1,4-dichlorobenzene) and two (1,3- and 1,4-dichlorobenzene plus 1,2,3,5- and 1,2,4,5-tetrachlorobenzene) pairs of isomers, respectively. Consequently, two-variable regression equations were screened to find the best equations that can both generate the experimental elution sequence and discriminate between chlorobenzene isomers. The best two-variable regression equation for I^{SE-30} (non-polar stationary phase) and its statistical parameters are as follows:

$$I^{SE-30} = 337.0 \cdot {}^1\chi + 47.9 \cdot {}^4\chi_{PC} - 329.9 \quad (1)$$

$$n = 13 \quad R = 0.998 \quad F^{2,10} = 1051 \quad EV = 99.5\%$$

Statistically, eqn. 1 is significant above the 99.9% level, while the ${}^1\chi$ and ${}^4\chi_{PC}$ indices are significant above the 99.95% and 95% level, respectively. Eqn. 1 accounts for 99.5% of the variation in the I^{SE-30} data. The size of chlorinated benzenes, best described and quantified by the ${}^1\chi$ index whose numerical value is directly proportional to the number of bonds in a molecule, is thus found to be the major structural factor that governs the magnitude of the I^{SE-30} values. The second most significant structural factor influencing the I^{SE-30} data is the substitution pattern of the chlorine atoms, given by the ${}^4\chi_{PC}$ index. This structural parameter has already been found to be significant for the other polysubstituted benzenes²⁴, branched alkyl alcohols²⁵, and branched alkyl chlorides¹. From Table I we see that within isomeric classes the value of the ${}^4\chi_{PC}$ index increases with increasing proximity of the chlorine substituents. Its

positive regression coefficient is consistent with the observation that the isomers which have all chlorine substituents together have the highest retention index in their class (Table II). Thus we conclude that the resulting I^{SE-30} is proportional to the size of chlorinated benzenes and the substitution pattern only slightly modifies this dependence. A graphical representation of the fit between observed I^{SE-30} values and those calculated by eqn. 1 is given in Fig. 1.

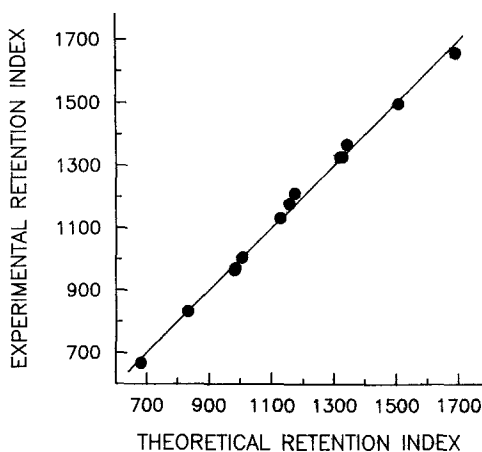


Fig. 1. Correlation between experimental (non-polar stationary phase, SE-30) and calculated (regression eqn. 1) retention indices of benzene and its chlorinated derivatives.

For the polar stationary phase, higher-order molecular connectivity indices were needed to describe correctly the chromatographic behavior of the chlorinated benzenes and to obtain the same level of precision in predicting retention indices as with equation 1. (The same was also found for monochlorides with linear alkyl chain¹.) The best two-variable regression equation for I^{CW20M} (a polar stationary phase) and its statistical parameters are as follows:

$$I^{CW20M} = 226.8 \cdot {}^3\chi + 1588.0 \cdot {}^7\chi_{CH} + 649.1 \quad (2)$$

$n = 13 \quad R = 0.998 \quad F^{2,10} = 1347 \quad EV = 99.6\%$

Eqn. 2 is statistically significant above the 99.9% level and accounts for 99.6% of the variance in the I^{CW20M} data. The ${}^3\chi$ and ${}^7\chi_{CH}$ indices are significant above the 99.95% level. These results are identical with those obtained for regression eqn. 1. Again, the size of chlorinated benzenes is found to be the main structural factor that determines their chromatographic behavior on polar stationary phase. This structural property is best described by the ${}^3\chi$ index, which correlates strongly with the ${}^1\chi$ index but has higher discriminatory power (*i.e.* is more sensitive to structural differences). The substitution pattern of the chlorine atoms is again the next most significant structural factor determining the magnitudes of retention indices on the polar stationary phase. Here its quantitative contribution is far more important than for the non-polar stationary phase. This contribution is most adequately quantified by the ${}^7\chi_{CH}$ index, which is calculated from the monosubstituted benzene type graphs. A

graphical representation of the fit between observed I^{CW20M} values and those calculated by eqn. 2 is given in Fig. 2. Valence type molecular connectivity indices were also tried but did not yield any improvement in the calculation of retention indices on either stationary phase.

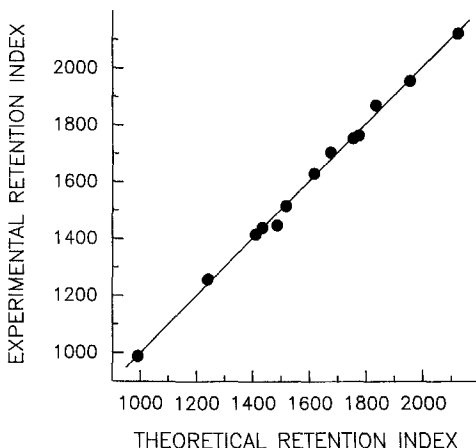


Fig. 2. Correlation between experimental (polar stationary phase Carbowax 20M) and calculated (regression eqn. 2) retention indices of benzene and its chlorinated derivatives.

The correlation analyses presented here demonstrated that the connectivity model is capable of predicting the retention indices of chlorinated benzenes. It is worth noting here that both regression equations reproduce the elution patterns of isomeric chlorobenzenes very nicely. However, the one-variable molecular connectivity models are shown to be insufficient to predict correctly elution sequences of chlorinated benzenes.

Since the connectivity model successfully reproduces the retention indices of chlorinated benzenes, the next step is to correlate its results with those obtained by the Wiener number and Balaban index methods. Correlations between the I^{SE-30} data and the ${}^1\chi$, W , and J indices are presented in Table III, which compares both the correlation coefficients and the number of compounds whose elution sequence is incorrectly predicted.

TABLE III

EVALUATION OF THE RESULTS OBTAINED BY DIFFERENT TOPOLOGICAL INDICES FOR THE CHROMATOGRAPHIC BEHAVIOR OF CHLOROBENZENES ON NON-POLAR STATIONARY PHASE SE-30

Topological index	${}^1\chi$	W	$W(het)^*$	J	$J(het)^*$
Correlation coefficient	0.997	0.983	0.984	0.967	0.985
Number of compounds whose elution sequence is incorrectly predicted	0	3	3	2	2

* Wiener numbers and Balaban indices calculated with the heteroatoms parameterization scheme²⁰.

An examination of Table III clearly demonstrates several important results: the first-order molecular connectivity index performs much better than either Wiener number or Balaban index. We might ask why the topological index with the lower structural information content and lower discriminatory power correlates better with the retention indices of chlorinated benzenes? The most plausible explanation is that the large amount of structural informations contained in the single W or J index may obscure those factors that are significant for the chromatographic behavior of chlorinated benzenes. In addition, the molecular connectivity model is more flexible, allowing quantification of general as well as local structural properties of molecules. Thus I conclude that, at least in the case of chlorinated benzenes, the connectivity model describes their chromatographic behavior much better than Wiener number or Balaban index.

In addition to the above, I conclude that the heteroatoms parameterization scheme performed²⁰ satisfactorily, quantitatively improving correlation with the experimental retention indices, particularly with the Balaban index. Thus, it should be used in future calculations of Wiener numbers and Balaban indices of molecules containing heteroatoms.

CONCLUSION

In this investigation the molecular connectivity model has been shown to predict successfully the retention indices of chlorinated benzenes on polar and non-polar stationary phases. To reproduce the experimental elution sequences the two-variable models were necessary. Statistical analyses show that the size of chlorinated benzene is the major structural factor that controls its chromatographic behavior. A comparison of the results obtained with the molecular connectivity model and the Wiener number or Balaban index reveals that the former method predicts the retention indices of chlorinated benzenes more reliably. In addition, the heteroatoms parameterization scheme, for topological indices calculated from the distance matrix, performed very satisfactory. Its use in future investigations is recommended.

REFERENCES

- 1 A. Sabljic, *J. Chromatogr.*, 314 (1984) 1; and references cited therein.
- 2 M. Randić, *J. Chromatogr.*, 161 (1978) 1.
- 3 R. Kalisz and H. Lamparczyk, *J. Chromatogr. Sci.*, 16 (1978) 246.
- 4 M. Gassiot-Matas and G. Firpo-Pamies, *J. Chromatogr.*, 187 (1980) 1.
- 5 M. J. M. Wells, C. R. Clark and R. M. Patterson, *J. Chromatogr. Sci.*, 19 (1981) 573.
- 6 Gy. Szász, O. Papp, J. Vámos, K. Hankó-Novák and L. B. Kier, *J. Chromatogr.*, 269 (1983) 91.
- 7 J. R. Chrétien and J. E. Dubois, *Anal. Chem.*, 49 (1977) 747.
- 8 J. R. Chrétien and J. E. Dubois, *J. Chromatogr.*, 158 (1978) 43.
- 9 A. Sabljic and M. Protić, *Bull. Environm. Contam. Toxicol.*, 28 (1982) 162.
- 10 A. Sabljic and M. Protić, *Chem.-Biol. Interact.*, 42 (1982) 301.
- 11 A. Sabljic, *J. Agric. Food Chem.*, 32 (1984) 243.
- 12 R. Koch, *Toxicol. Environ. Chem.*, 6 (1983) 87.
- 13 L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
- 14 A. Sabljic and N. Trinajstić, *Acta Pharm. Jugosl.*, 31 (1981) 189.
- 15 N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, FL, 1983.
- 16 J. K. Haken and I. O. O. Korhonen, *J. Chromatogr.*, 265 (1983) 323.

- 17 H. Wiener, *J. Amer. Chem. Soc.*, 69 (1947) 17 and 2636.
- 18 A. T. Balaban, *Chem. Phys. Lett.*, 89 (1982) 399.
- 19 D. Bonchev, Ov. Mekenjan, G. Protić and N. Trinajstić, *J. Chromatogr.*, 176 (1979) 149.
- 20 M. Barysz, G. Jashari, R. S. Lall, V. K. Srivastava and N. Trinajstić, in R. B. King (Editor), *Chemical Applications of Topology and Graph Theory*, Elsevier, Amsterdam, 1983, p. 222.
- 21 A. T. Balaban, A. Chiriac, I. Motoc and Z. Simon, *Steric Fit in Quantitative Structure-Activity Relationships. Lecture Notes in Chemistry No. 15*, Springer, Berlin, 1980.
- 22 L. H. Hall and L. B. Kier, *Eur. J. Med. Chem.*, 16 (1981) 399.
- 23 G. D. Knott, *Comput. Prog. Biomed.*, 10 (1979) 271.
- 24 L. B. Kier, *J. Pharm. Sci.*, 69 (1980) 1034.
- 25 A. Sabljic and M. Protić-Sabljić, *Mol. Pharmacol.*, 23 (1983) 213.
- 26 L. E. Cook and F. M. Raushel, *J. Chromatogr.*, 65 (1972) 556.
- 27 F. Vernon and J. B. Suratman, *Chromatographia*, 17 (1983) 600.