ELSEVIER

Journal of Chromatography A, 719 (1996) 462–467

JOURNAL OF CHROMATOGRAPHY A

Short communication

Using topological indices in the prediction of gas chromatographic retention indices of linear alkylbenzene isomers

Vilma E.F. Heinzen, Rosendo A. Yunes*

Departmento de Química, Universidade Federal de Santa Catarina, 88040-900, Florianópolis, Santa Catarina, Brazil

First received 21 March 1995; revised manuscript received 13 June 1995; accepted 22 June 1995

Abstract

The Balaban, Wiener, Electrotopological state and molecular shape indices were determined and correlated with the retention indices of linear alkylbenzene isomers with C_{10}–C_{14} linear alkyl chains. The Wiener index is the only one which shows a good single linear correlation with the retention indices and a correct elution sequence. Multiple linear regression analysis with the connectivity index ($^1\chi$) and electrotopological state index (S) for the C-7 position of the linear alkyl chain, which encode complementary information, also gives a good correlation coefficient and a good elution sequence (it is not correct only for four compounds). The different orders of shape indeces (kappa values), indices of molecular shape, also give a good correlation by a multiple linear regression, the best correlation coefficient being obtained with the $^1\kappa$, $^6\kappa$ indices, and the $^1\kappa$, $^7\kappa$ indices (the elution sequence is not correct for five compounds).

1. Introduction

The idea of representing molecules by means of a graph is a very old one. Wiener [1] was the first to have the idea that these graphs could be represented by numeric indices. Since that time, a large number of indices have been derived from molecular graphs. Among the best known can be mentioned those of Hosoya [2], Randic [3], Bonchev and Trinajstic [4], Balaban [5] and Kier and Hall [6].

The common characteristic in all of these graph-based approaches is the counting of one or more of the atoms of the graph. These atoms are graph vertices and graph edges are bonds. The information contained in these counts, represented at the end by a single number called the index, has been shown to relate closely to the physical property values of molecules. Generally, the index differs from one molecule to another. The representation of the molecular structure by a single number permits the correlation with the molecular properties recorded as well.

Molecular topological indices have been shown to be very important structural parameters for describing the chromatographic behaviour of chemicals [1–6].

We demonstrated [7] that with only one molecular connectivity index it is not possible to predict the retention indices and elution sequences of linear alkylbenzene isomers (LABs)

* Corresponding author.

0021-9673/96/$15.00
SSDI 0021-9673(95)00707-5

studied (with SE-54 and DB-1 stationary phases) without making corrections.

Recently, Kier and co-workers [8,9] developed an index which encodes additional information taking into account electronic and topological factors, called the electrotopological state index (S). The electrotopological state index of an atom unifies in a single index both an electronic and a topological description.

Kier [10] has also proposed an approach to shape or steric quantification through the index of molecular shape. Although it is in an early stage of development, the author feels that this is an important area of research. These graphs attempt to quantify molecular structure using shape effects and geometry-based methods, in contrast with the other graph-based methods, which consider the presence of atoms and their adjacency relationships, with no assumptions about a surface.

In this work, the Wiener index method [1], the Balaban index method [5], the recently developed electrotopological state index (S) [8,9] and the indices of molecular shape (different orders of kappa values, ${}^1\kappa$, ${}^2\kappa$, ${}^3\kappa$, etc.) [10] in order to establish the capacity of these methods to predict the retention indices of the positional isomers of LABs.

2. Experimental

Several extensive reviews have been published [1,5,8–12] which give detailed descriptions of the theory and method of calculation of all topological indices used in the present investigation.

The molecular connectivity index (χ) of different orders, the electrotopological state index (S) and the indices of molecular shape, kappa values, (${}^1\kappa$, ${}^2\kappa$, ${}^3\kappa$) were calculated by the method of Kier and Hall utilizing the Molconn-X computer program for molecular topology analysis [12]. The kappa values of different orders ${}^4\kappa$, ${}^5\kappa$, ${}^6\kappa$ and ${}^7\kappa$ were calculated according to the method of Kier and Hall applying subgraphs in accordance with the order of corresponding kappa values. The Wiener index was calculated utilizing the same program [12] and the Balaban index was calculated according to the method of the author [5].

The chromatographic retention indices (I) of linear alkylbenzene isomers with C_{10}–C_{14} alkyl chains (LABs) were reported by Takad and Ishiwatari [13] and Peng et al. [14]. The retention data of Peng et al. are based on the Kováts convention, whereas those of Takad and Ishiwatari refer to linear phenylalkanes as standards.

Single and multiple linear regression analyses were carried out using an IBM AT/486 computer program (Origin). To test the quality of the regression equation, the correlation coefficient (r), the coefficient of determination (r^2) and the test of null hypothesis (F-test) were utilized as statistical parameters.

3. Results and discussion

The values of the chromatographic retention indices (I) of isomers of LABs with C_{10}–C_{14} linear alkyl chains [13,14] and their topological indices, electrotopological state (S), Balaban (J), Wiener (W) and connectivity index of the first order (${}^1\chi$), are given in Table 1. The Wiener index distinguishes the isomers, decreasing when the position of the phenyl group goes from C_1 to C_n ($n = 2$, 3, 4, 5 and 6) in the same isomeric series, in accordance with the retention indices which also decrease due to the decrease in the interaction between the molecular surface and the stationary phase because of the molecule becoming more symmetrical and more compact.

The linear equations obtained for the calculation of the retention indices of LABs using the Wiener indices are as follows:

$$I_{\text{SE-54}} = 55.3212 + 0.0686\,W \qquad n = 29;\ r = 0.9930;\ r^2 = 0.9860 \tag{1}$$

$$I_{\text{SE-54}} = 51.7850 + 0.0739\,W \qquad n = 23;\ r = 0.9924;\ r^2 = 0.9848 \tag{2}$$

$$I_{\text{DB-1}} = 1166.23 + 0.7623\,W \qquad n = 23;\ r = 0.9911;\ r^2 = 0.9823 \tag{3}$$

Table 1
Wiener index (W), Balaban index (J), molecular connectivity index ($^1\chi$) and electrotopological state index (S) with relation to the carbon atom in position 7 in the linear alkyl chain and observed retention indices of linear alkylbenzenes with C_{10}–C_{14} linear alkyl chains with SE-54 and DB-1 stationary phases

Compound *nCm* [a]	$I_{SE\text{-}54}$ [b]	$I_{DB\text{-}1}$ [c]	W	J	$^1\chi$	S
1C10	100.00	1664	612	1.7760	7.9319	1.4397
2C10	92.84	1588	564	1.9416	7.8425	1.4268
3C10	89.37	1553	528	2.0990	7.8805	1.4184
4C10	87.54	1534	504	2.2173	7.8805	1.4063
5C10	86.70	1526	492	2.2791	7.8805	1.3887
1C11	110.00	1771	742	1.7454	8.4319	1.4509
2C11	102.71	1692	688	1.9040	8.3425	1.4381
3C11	99.15	1656	646	2.0494	8.3805	1.4296
4C11	97.15	1636	616	2.1682	8.3805	1.4175
5C11	96.20	1626	598	2.2445	8.3805	1.3999
6C11	95.92	1620	592	2.2749	8.3805	1.3766
1C12	120.00	1870	889	1.7280	8.9319	1.4570
2C12	112.63	1791	829	1.8698	8.8425	1.4442
3C12	108.95	1755	781	2.0044	8.8805	1.4357
4C12	106.90	1735	745	2.1181	8.8805	1.4237
5C12	105.76	1723	721	2.2080	8.8805	1.4060
6C12	105.31	1719	709	2.2518	8.8805	1.3827
1C13	130.00	1978	1054	1.7158	9.4319	1.4607
2C13	122.52	1894	988	1.8410	9.3425	1.4478
3C13	118.81	1854	934	1.9670	9.3805	1.4394
4C13	116.59	1833	892	2.0696	9.3805	1.4273
5C13	115.42	1821	862	2.1638	9.3805	1.4097
6C13	114.75	1814	844	2.2192	9.3805	1.3864
1C14	140.00		1238	1.7022	9.9319	1.4631
2C14	132.46		1166	1.8106	9.8425	1.4502
3C14	128.67		1106	1.9286	9.8805	1.4418
4C14	126.37		1058	2.0275	9.8805	1.4297
5C14	125.01		1022	2.1211	9.8805	1.4121
6C14	124.20		998	2.1800	9.8805	1.3887

[a] *nCm*: *n* indicates the position of the benzene ring in the alkyl chain and *m* indicates the number of carbon atoms in the alkyl chain.
[b] From Ref. [13].
[c] From Ref. [14].

The good correlation obtained with the Wiener index shows that the retention indices of LABs depend mainly on the shape and size of the molecular structure.

The Balaban index distinguishes the isomeric structure of LABs, but their values increase when the position of the aromatic ring is more internal in the molecule, in contrast with those of the retention indices. The linear correlation is unsatisfactory.

Analysis by single linear regression with connectivity indices demonstrated that this method is not sufficient to give a good correlation or correct elution sequence [7]. However, a good correlation coefficient by a multi-variable regression equation is obtained with two connectivity indices, $^1\chi$ and $^5\chi_p$, but, in spite of this, the elution sequence was wrong for eight compounds. This led us to test the correlation using the connectivity indices and electrotopological state index (S), which gives an electronic and a topological description of each atom in the molecule.

The best correlation was obtained with $^1\chi$

(first-order connectivity index) and S (electrotopological state index) for the C-7 position of the alkyl linear chain (Eqs. 4 and 5). The $^1\chi$ index conveys more information about the number of atoms in a molecule. This index decreases with displacement of the phenyl group from the end to the middle of the alkyl chain of an isomeric series according to the I values, but it does not distinguish the isomers 3C, 4C, 5C and 6C.

The electrotopological state index (S), with relation to the C-7 position of the linear chain, gives more information about the proximity of the phenyl group with relation to this atom, distinguishing all the different positions of the phenyl group.

$$I_{SE\text{-}54} = 17.8234\ ^1\chi + 180.1287\ S - 304.7002$$
$$n = 29;\ r^2 = 0.9706;\ F(^1\chi) = 626.61\ (P > 0.0001)$$
$$r = 0.9852;\ F(S) = 78.10\ (P > 0.0001) \quad (4)$$

$$I_{DB\text{-}1} = 180.2364\ ^1\chi + 1867.8413\ S - 2494.4242$$
$$n = 23;\ r^2 = 0.9536;\ F(^1\chi) = 280.76\ (P > 0.0001)$$
$$r = 0.9765;\ F(S) = 56.79\ (P > 0.0001) \quad (5)$$

These equations distinguish all the compounds studied and the elution sequence is correct.

The values of the kappa indices ($^1\kappa$, $^2\kappa$, $^3\kappa$, $^4\kappa$, $^6\kappa$, $^7\kappa$) are given in Table 2. The $^1\kappa$, $^2\kappa$, $^3\kappa$ and $^4\kappa$ do not distinguish most of the positional isomers of the aromatic ring; $^6\kappa$ and $^7\kappa$ distinguish all of them, but neither of them shows a good single

Table 2
Values of different orders of kappa indices ($^1\kappa$, $^2\kappa$, $^3\kappa$, $^4\kappa$, $^6\kappa$ and $^7\kappa$) of linear alkylbenzenes with C_{10}–C_{14} linear alkyl chains

Compound[a]	$^1\kappa$	$^2\kappa$	$^3\kappa$	$^4\kappa$	$^6\kappa$	$^7\kappa$
1C10	14.0625	10.1730	7.8642	7.0222	11.0000	12.9337
2C10	14.0625	9.0741	6.3700	5.7483	8.7716	9.9023
3C10	14.0625	9.0741	5.7778	4.7921	7.0222	7.8241
4C10	14.0625	9.0741	5.7778	4.4010	5.7483	6.3375
5C10	14.0625	9.0741	5.7778	4.4010	4.7921	5.7483
1C11	15.0588	11.1111	8.6870	7.8400	12.2500	13.9378
2C11	15.0588	9.2723	7.1111	6.4793	9.6790	10.8512
3C11	15.0588	9.2723	6.4793	5.4444	7.8400	8.6870
4C11	15.0588	9.2723	6.4793	5.0176	6.4793	7.1111
5C11	15.0588	9.2723	6.4793	5.0176	5.4444	5.9282
6C11	15.0588	9.2723	6.4793	5.0176	5.0176	5.4444
1C12	16.0555	12.0554	9.6000	8.6735	13.2353	14.9414
2C12	16.0555	10.8800	7.9339	7.2306	10.5956	11.8055
3C12	16.0555	10.8800	7.2590	6.1200	8.6735	9.5625
4C12	16.0555	10.8800	7.2590	5.6583	7.2306	7.9029
5C12	16.0555	10.8800	7.2590	5.6583	6.1200	6.6406
6C12	16.0555	10.8800	7.2590	5.6583	5.6583	5.6583
1C13	17.0526	13.0050	10.4490	9.5207	14.2222	15.9446
2C13	17.0526	11.7959	8.7108	8.0000	11.5200	12.7645
3C13	17.0526	11.7959	8.0000	6.8166	9.5207	10.4490
4C13	17.0526	11.7959	8.0000	6.3210	8.0000	8.7108
5C13	17.0526	11.7959	8.0000	6.3210	6.8166	7.3728
6C13	17.0526	11.7959	8.0000	6.3210	6.3210	6.3210
1C14	18.0550	13.9592	11.3802	10.3800	15.2105	16.9475
2C14	18.0550	12.7190	9.5625	8.7856	12.4512	13.7275
3C14	18.0550	12.7190	8.8128	7.5322	10.380	11.3450
4C14	18.0550	12.7190	8.8128	7.0038	8.7856	9.5330
5C14	18.0550	12.7190	8.8128	7.0038	7.5321	8.1228
6C14	18.0550	12.7190	8.8128	7.0038	7.0038	7.0040

[a] See Table 1.

linear correlation with the retention index. This led us to test the multiple linear correlation with two variables, obtaining very good correlations with $^{1}\kappa$, $^{6}\kappa$ ($r = 0.9976$ for $I_{SE\text{-}54}$, $r = 0.9955$ for $I_{DB\text{-}1}$) and $^{1}\kappa$, $^{7}\kappa$ ($r = 0.9975$ for $I_{SE\text{-}54}$, $r = 0.9975$ for $I_{DB\text{-}1}$) using Eqs. 6–9.

$$I_{SE\text{-}54} = 8.2363\ ^{1}\kappa + 1.9190\ ^{6}\kappa - 38.6154$$
$$n = 29;\ r^2 = 0.9952;\ r = 0.9976 \quad (6)$$
$$F = 2714.99\ (P > 0.0001)$$

$$I_{DB\text{-}1} = 84.5072\ ^{1}\kappa + 20.2527\ ^{6}\kappa - 234.8143$$
$$n = 23;\ r^2 = 0.9911;\ r = 0.9955 \quad (7)$$
$$F = 1118.29\ (P > 0.0001)$$

$$I_{SE\text{-}54} = 8.5154\ ^{1}\kappa + 1.6168\ ^{7}\kappa - 42.1312$$
$$n = 29;\ r^2 = 0.9951;\ r = 0.9975 \quad (8)$$
$$F = 2645.39\ (P > 0.0001)$$

$$I_{DB\text{-}1} = 87.7956\ ^{1}\kappa + 17.0641\ ^{7}\kappa - 194.6065$$
$$n = 23;\ r^2 = 0.9913;\ r = 0.9975 \quad (9)$$
$$F = 1135.07\ (P > 0.0001)$$

The data in Table 2 shows that $^{1}\kappa$ varies only with the total number of carbon atoms in the LAB molecule and does not depend on the position of the phenyl group along the alkyl chain. According to Kier [10], the $^{3}\kappa$ values are larger when branching is located at the extremities of a graph, encoding information about the centrality of branching. In this case, it can be observed that the $^{3}\kappa$ values are larger for the phenyl group in position 1 than for positions 2 and 3. However, this index does not distinguish the compounds containing the phenyl group in positions 4, 5 and 6. The $^{2}\kappa$ values do not distinguish the compounds containing the phenyl group in positions 3, 4, 5 and 6, $^{4}\kappa$ does not distinguish the 5- and 6-positional isomers and $^{5}\kappa$ does not distinguish the 6-positional isomers. Increases in the order of κ values imply increases in their power of distinction. Here appears the importance of indices $^{6}\kappa$ and $^{7}\kappa$, which in this case distinguish all the positional isomers. This suggests that a relationship should exist between the number of isomeric positions in the molecules and the order of κ indices distinguishing the different isomers.

The fitting of $^{1}\kappa$ and $^{6}\kappa$ or of $^{1}\kappa$ and $^{7}\kappa$ into multiple linear regressions with the retention indices of LABs, as given in Eqs. 6–9, means that $^{1}\kappa$ can code the size of the molecule and that $^{6}\kappa$ or $^{7}\kappa$ can code branching or the position of the phenyl groups along the alkyl chain.

4. Conclusions

It should be emphasized that the only method which gives a good result by single linear regression is the Wiener index. Hence, we can conclude that the Wiener index is the best approach for the characterization of this kind of compound.

With multiple linear regression, the combination of connectivity index and electrotopological state index, and also the combination of different κ indices, good results are obtained. It is interesting that the kappa indices of orders 6 and 7, which have been studied for the first time in this work, demonstrate the encoding of important information about the symmetry of these molecules, and the fact that a relationship should exist between the number of isomeric positions in the molecules and the order of the κ indices with respect to their capacity to distinguish these isomers.

Acknowledgement

The authors thank CNPq (Brazil) for financial support.

References

[1] H. Wiener, J. Am. Chem. Soc., 69 (1947) 17 and 2636.
[2] H. Hosoya, Bull. Chem. Soc. Jpn., 44 (1971) 2332.
[3] M. Randic, J. Chromatogr., 161 (1978) 1.
[4] D. Bonchev and N. Trinajstic, J. Chem. Phys., 67 (1977) 4517.
[5] A.T. Balaban, Chem. Phys. Lett., 89 (1982) 399.
[6] L.B. Kier and L.H. Hall, Molecular Connectivity in Structure–Activity Analysis, Wiley, New York, 1986.

[7] V.E.F. Heinzen and R.A. Yunes, J. Chromatogr. A, 654 (1993) 183.
[8] L.B. Kier and L.H. Hall, Pharm. Res., 7 (1990) 801.
[9] L.H. Hall, B. Mohney and L.B. Kier, J. Chem. Inf. Comput. Sci., 31 (1991) 76.
[10] L.B. Kier, Med. Res. Rev., 7 (1987) 417.
[11] L.B. Kier, in S.H. Yalkowsky, A. Sinkula and S.C. Valvani (Editors), Physical Chemical Properties of Drugs, Marcel Dekker, New York, 1980, Ch. 9, p. 277.
[12] L.H. Hall and L.B. Kier, Molconn-X, a Program for Molecular Topology Analysis, User's Guide, Quincy, MA, 1991.
[13] H. Takad and R. Ishiwatari, J. Chromatogr., 346 (1985) 281.
[14] C.T. Peng, R.L. Hua and D. Maltby, J. Chromatogr., 589 (1992) 231.