ELSEVIER

Journal of Chromatography A, 858 (1999) 187–199

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Modeling chromatographic parameters by a novel graph theoretical sub-structural approach

Ernesto Estrada*, Yaquelín Gutierrez

Department of Pharmacy, Faculty of Chemistry and Pharmacy, Universidad Central de Las Villas, Santa Clara 54830, Villa Clara, Cuba

Received 11 May 1999; received in revised form 15 July 1999; accepted 15 July 1999

Abstract

A novel approach to the study of quantitative relationships between chromatographic parameters and the chemical structure is introduced. It is based on the computation of the spectral moments of the topological bond matrix by using different weights as diagonal entries of this matrix. The main advantage of the present approach is that the quantitative contributions of the structural fragments of molecules to the chromatographic parameters studied can be obtained explicitly. By using this approach we study two data sets: one composed of 156 alkanes and the other of 81 oxygen-containing organic molecules. In both cases excellent quantitative structure–chromatographic retention relationships were obtained. The contributions of the different fragments to the chromatographic retention were generated obtaining tables of additive contributions to the properties studied. The physicochemical interpretation of the results on the basis of the retention mechanisms is also analyzed in light of this new approach.

Keywords: Structure–chromatographic parameters relationships; Topological indices; Molecular descriptors; Alkanes; Alcohols; Esters; Ketones

1. Introduction

The prediction of physicochemical and biological properties of organic molecules is one of the main objectives of the methods based on quantitative structure–property relationships (QSPRs) [1]. Among the most important parameters that have been extensively studied by using these approaches are the chromatographic ones [2]. Quantitative structure–chromatographic retention relationship (QSRR) studies have been reported with the use of structure-cryptic, structure-implicit and structure-explicit schemes [3–5]. Examples of structure-cryptic QSRR studies are those that use experimental properties as independent variables to describe chromatographic parameters [6–9]. Structure-implicit methods generally employ quantum chemical parameters in the model [10] and the structure-explicit approaches are based on the use of graph theoretical parameters in the development of the QSRR models [11–16]. Overlapping approaches using combinations of structure-implicit and structure-explicit schemes have also been reported in the literature [17–24].

The main advantage of the graph theoretical

*Corresponding author. Present address: Laboratory of Pharmaceutical Chemistry, Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15706 Santiago de Compostela, Spain. Fax: +34-981-594-912.

E-mail address: estrada66@yahoo.com (E. Estrada)

0021-9673/99/$ – see front matter
PII: S0021-9673(99)00808-0

approach to the prediction of properties is that it permits the interpretation of results in terms of structurally related concepts. In spite of that, the most important criticism of the so-called topological indices, which are graph theoretical invariants, is concerned with their physical meaning. The complexity of some of the QSPR models obtained by using this kind of molecular descriptor has been recognized as forbidding by others researchers [25]. But, the graph theory permits the definition of novel invariants that produce similar quantitative models to those obtained with ''classical'' topological indices but having more ''transparent'' structural interpretation [26].

One of the graph theoretical approaches to the prediction of physicochemical and biological properties of molecules is that based on the bond adjacency matrix of molecular graphs [27–30]. Recently, one of us (E.E.) has introduced a sub-structural approach to QSPRs that employs the spectral moments of such a matrix [31–35]. The main advantage of the novel approach consists in the possibility of expressing the spectral moments as linear combinations of structural fragments of the molecules. Consequently, we can substitute the spectral moments in the quantitative model by their expressions in terms of structural fragments of the molecules obtaining an equation that relates the property directly with the molecular structure. In this approach, the study of molecules containing any kind of heteroatoms in the structure is accounted for by using edge weights as diagonal entries of the bond matrix, such as bond distances or bond dipoles, in the molecular graphs [32,34,35].

The objective of the present work is to introduce this novel approach to the study of chromatographic parameters in simple data sets in order to test its applicability to such kind of QSPR problems. We will see that it produces significant QSRR models permitting their interpretation in terms of structural fragments of molecules. The use of this approach to describe other data sets of chromatographic parameters of more complicated molecular structures will be studied in forthcoming studies.

2. Theoretical model

The present approach is based on the representation of molecules through the so-called hydrogen depleted molecular graphs [36]. In a molecular graph $G=(V, E)$, the elements of the vertex set $V=\{v_1, v_2, ..., v_n\}$ represent the atoms in the molecule and the elements of the edge set $E=\{e_1, e_2, ..., e_m\}$ represent the covalent bonds in the molecule. Then, these molecular graphs are represented by square and symmetric matrices named their bond-adjacency matrices [27,36]. The non-diagonal elements of such matrices are ones if, and only if, bond i is adjacent to bond j. Two bonds are adjacent if they are incidents to a common atom.

Two different approaches can be distinguished in the theory of the spectral moments of the bond matrix depending on the values assigned to their diagonal entries. The first, simplest approach, uses zeros as diagonal entries in those matrices [31,33], while in the second approach the use of bond distances (or any other bond property, such as bond dipoles, bond polarizabilities, etc.) are used [32,34,35]. The first scheme is appropriated in such cases in which the presence of heteroatoms is not decisive in the explanation of the property studied. For instance, this approach is useful in the study of chromatographic or physicochemical properties of alkanes. However, in the most general case in which heteroatoms are present, the second approach is more appropriated.

The present approach to QSRR studies is based on the computation of the spectral moments of the bond matrix, i.e., the sum of diagonal entries of the powers of the bond matrix. This approach can be resumed as follows. First, the chromatographic property, P, is described in terms of the different spectral moments of the bond matrix via a linear regression model of the form:

$$P = a_0\mu_0 + a_1\mu_1 + a_2\mu_2 + a_3\mu_3 + a_4\mu_4 + \cdots + a_k\mu_k + b \quad (1)$$

This quantitative model is obtained by using any of the many different multivariate regression approaches currently described in the literature. Here we will use the multivariate linear regression analysis in order to generate models of the type of Model (1).

In the second step of the application of this approach, the spectral moments are substituted by their expression in terms of the different structural

fragments of the molecules. Two different cases can be presented here. The first is when zero-diagonal entries are used, and the other when bond distances are used as diagonal entries for the calculation of the spectral moments. In the first case, Model (1) is transformed to the following expression directly relating the chromatographic property to the molecular structure of the compounds studied:

$$P = a'_1 \cdot |F_1| + a'_2 \cdot |F_2| + a'_3 \cdot |F_3| + a'_4 \cdot |F_4| + a'_5 \cdot |F_5| + a'_6 \cdot |F_6| + \cdots + a'_p \cdot |F_p| + b \quad (2)$$

This model is obtained by substituting the spectral moments by their expressions in terms of the different structural fragments that are given in Table 1 and Fig. 1.

In the second case, when heteroatoms are considered, Model (1) is transformed to the following:

$$P = b + \sum_i (b_{AB}|AB|)_i + \sum_j (b_{A(BC)}|A(BC)|)_j + \sum_k (b_{A(BCD)}|A(BCD)|)_k + \sum_l (b_{A(BCDE)}|A(BCDE)|)_l + \sum_m (b_{A\text{-}B\text{-}C\text{-}D}|A\text{-}B\text{-}C\text{-}D|)_m + \sum_n (b_{A(BCD\text{-}E)}|A(BCD\text{-}E)|)_n + \ldots. \quad (3)$$

In this expression, the indexes *i*, *j*, *k*, *l*, *m* and *n* run over the different fragments in the molecules. The symbol *AB* represents a bond in which atoms are denoted by *A* and *B*, *A*(*BC*), *A*(*BCD*) and *A*(*BCDE*) are fragments F_2, F_3 and F_4, respectively (see Fig. 1), in which atom *A* occupies the central position. Fragments F_5 and F_6 (see Fig. 1) are denoted by *A-B-C-D* and *A*(*BCD-E*), respectively, in fragment F_6 the atom *A* is that of valence three. For instance, $(|CX|)_1 = |CC|$, $(|CX|)_2 = |C=O|$, ..., $(|C(XY)|)_1 = |C(CC)|$, $(|C(XY)|)_2 = |C(CO)|$, and so forth. The coefficients b_{AB}, $b_{A(BC)}$, etc., given in Eq. (3) are directly computed from the values of the elements used in the diagonal entries of the bond matrix, for instance, the bond distances. In the appendix we give the mathematical expressions that should be used to carry out these calculations. The standard bond distances taken from Ref. [37] are used here as the diagonal entries of the bond matrix in the case when heteroatoms are to be considered.

In the present work we will study the chromatographic retention indices of two different data series of organic compounds. The first of these data sets consisted on the experimental gas chromatographic retention indices of 156 C_2–C_{13} alkanes on squalane at 333 K measured by Chretien and Dubois [38] and on squalane at 373 K reported by Schomburg and Dielman [39]. This data set was previously used by Bosnjak et al. [40] in testing the three-dimensional Wiener index, which was published in this journal. We did not consider the methane molecule because its hydrogen-depleted graph has no bond, which makes the present approach impossible. On the other hand, 81 oxygen-containing aliphatic compounds including alcohols, ketones and esters compose the second data set. The chromatographic retention indices of these compounds were measured on OV-1 at 333 K as reported in the Sadtler catalogue [41] and they were collected from the work of Duvenbeck and Zinn [42].

Table 1
The first eight spectral moments of the bond matrix as linear combinations of the number of fragments in the acyclic graph[a]

$\mu_0 = |F_1|$
$\mu_2 = 2 \cdot |F_2|$
$\mu_3 = 6 \cdot |F_3|$
$\mu_4 = 2 \cdot |F_2| + 12 \cdot |F_3| + 24 \cdot |F_4| + 4 \cdot |F_5|$
$\mu_5 = 30 \cdot |F_3| + 120 \cdot |F_4| + 10 \cdot |F_6|$
$\mu_6 = 2 \cdot |F_2| + 60 \cdot |F_3| + 480 \cdot |F_4| + 12 \cdot |F_5| + 24 \cdot |F_6| + 6 \cdot |F_7| + 36 \cdot |F_8| + 24 \cdot |F_9|$
$\mu_7 = 126 \cdot |F_3| + 1680 \cdot |F_4| + 84 \cdot |F_6| + 210 \cdot |F_8| + 112 \cdot |F_9| + 14 \cdot |F_{10}| + 14 \cdot |F_{11}| + 84 \cdot |F_{12}|$
$\mu_8 = 2 \cdot |F_2| + 252 \cdot |F_3| + 5544 \cdot |F_4| + 28 \cdot |F_5| + 200 \cdot |F_6| + 32 \cdot |F_7| + 1008 \cdot |F_8| + 464 \cdot |F_9| + 32 \cdot |F_{10}| + 40 \cdot |F_{11}| + 672 \cdot |F_{12}| + 8 \cdot |F_{13}| + 48 \cdot |F_{14}| + 48 \cdot |F_{15}| + 32 \cdot |F_{16}| + 32 \cdot |F_{17}| + 288 \cdot |F_{18}|$

[a] The different structural fragments F_i are illustrated in Fig. 1.

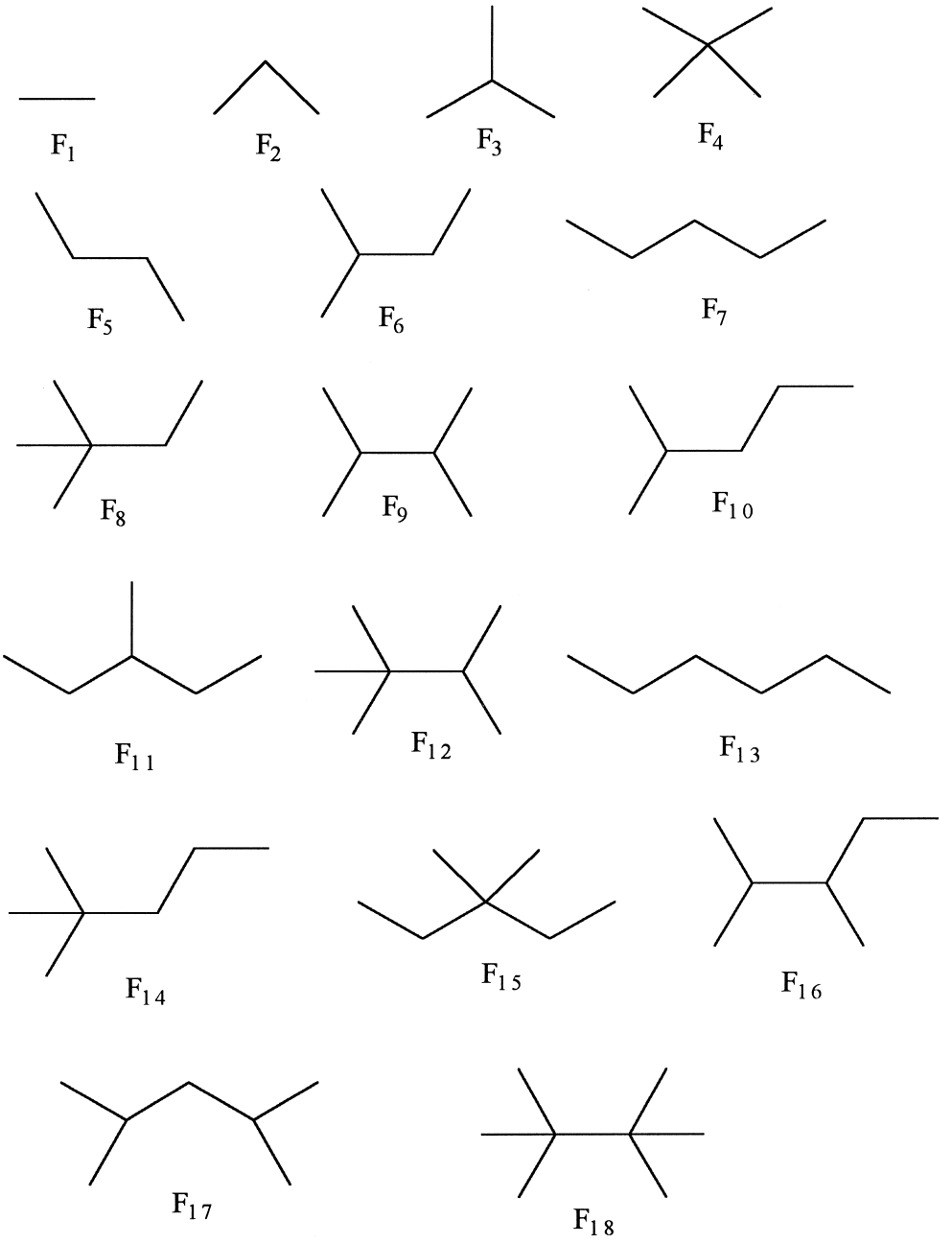

Fig. 1. Structural fragments contained in the first eight spectral moments of the bond matrix of non-weighted graphs.

3. Quantitative structure–chromatographic parameters relationships

It is normal in science to go from the simplest to the most complicated. In testing any new graph theoretical approach to QSPR studies, the first step is to prove the usability of the proposed model to describe properties of alkanes. These compounds are

the simplest ones and their properties are almost completely dependent from topological features. As a consequence, we have selected as a first data set a series of 156 alkanes as reported by Bosnjak [40]. The first 15 spectral moments of the bond matrix for the graphs representing these alkanes were calculated and correlated with their retention indexes through multivariate regression analysis.

The best linear regression model obtained by using the present approach is illustrated below together with the statistical parameters:

$$I = 31.15 + 137.94\mu_0 - 24.77\mu_3 + 6.01\mu_4 - 2.06\mu_6 + 0.23\mu_7;$$
$$n = 156, R = 0.9986, s = 9.44, s_{CV} = 10.11 \text{ and } F = 10\,718 \quad (4)$$

where I is the retention index, n is the size of the data set, R is the regression coefficient, s the standard deviation of the regression, s_{CV} is the standard deviation of the leave-one-out cross validation, and F is the Fisher ratio. The correlation between observed and calculated retention indices is illustrated in Fig. 2.

Model (4) explains more than 99.7% of the variance in the experimental retention indexes of alkanes. The standard deviation of the cross validation is only 6.6% greater than that of the regression model, which indicates the predictive power of this model. However, as we have previously stated, the most important feature of the present approach is not only its ability to obtain QSRR models that are statistically significant, but also the possibility of expressing these models in terms of structural fragments of molecules. By substituting the values of the spectral moments given in Eq. (4) by their expressions in terms of structural fragments given in Table 1, we obtain the following expression for the retention indices of alkanes:

$$I = 137.94|F_1| - 3.66|F_2| - 62.94|F_3| + 118.80|F_4| - 24.72|F_5| + 29.98|F_6| - 12.36|F_7| - 25.86|F_8| - 23.68|F_9| + 3.22|F_{10}| + 3.22|F_{11}| + 19.32|F_{12}| + 31.15 \quad (5)$$

The other example that we will study here is the description of the retention indices of a series of 81

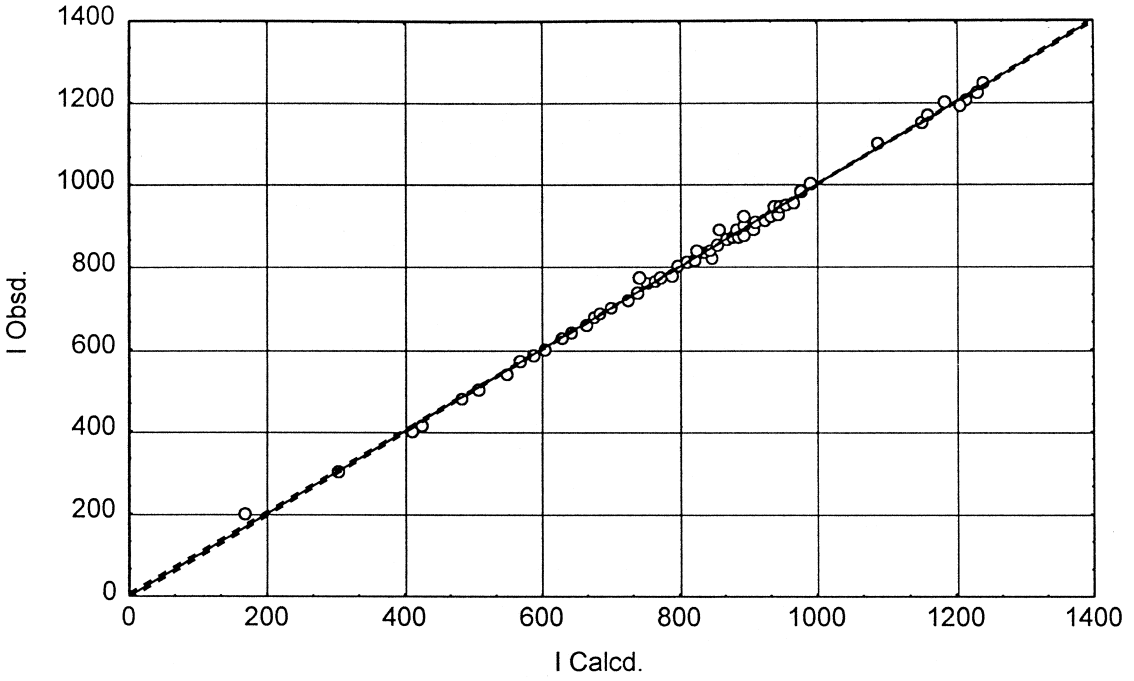

Fig. 2. Observed versus calculated retention indices of the 156 alkanes studied here.

aliphatic molecules containing oxygen in their structures on OV-1 at 333 K [41]. This data set includes alcohols, ketones and esters. We calculate the first 15 spectral moments of the bond matrix for these molecules by using the standard bond distances as weights in the diagonal entries of the matrix. The best linear regression equation obtained for this data set is illustrated below together with the statistical parameters of the model:

$$I = 273.44 - 201.52\mu_0 + 236.91\mu_1 - 38.15\mu_2 + 15.10\mu_3 - 1.41\mu_5 + 0.20\mu_6;$$
$$n = 81, R = 0.9912, s = 14.35, s_{CV} = 14.97 \text{ and } F = 692 \quad (6)$$

The observed and calculated retention indices for all compounds in this data set are illustrated in Table 2 together with the residual and the cross-validation residual. In this case, the QSRR model obtained explains more than 98% of the variance of the retention indices. This model shows a great stability to the inclusion or exclusion of data points as demonstrated by its standard deviation of the cross validation which is only 4.1% greater than that of the regression model.

The calculation of fragment contribution to the retention indices for molecules containing oxygen in their structures was carried out by the procedure explained in Section 2. The contributions coming from the 63 different fragments present in the studied molecules are illustrated in Table 3 and Fig. 3.

4. Discussion

The main objectives of any QSPR study are the prediction of the studied property with an appropriate accuracy and the interpretation of the results in terms of structural features of molecules. The first objective is concerned to the statistical quality and predictive power of the model obtained, while the second aspect is related to the structural meaning of the molecular descriptors included as independent variables in the model. The use of graph-theoretical descriptors, i.e., topological indices, in QSPRs, quantitative structure–activity relationships (QSARs) and QSRRs has been successful from the statistical point of view. Both examples studied here are not exceptions of this rule. As can be seen, the spectral moments of the bond matrix produce excellent correlations with the retention indices of alkanes and oxygen containing organic molecules, such as alcohols, ketones and esters. The good predictive abilities of the models found in the present work have been proved by using the leave-one-out technique of cross-validation. In both cases the standard deviations of the cross validations do not exceed 10% of the standard deviations of the regression models. It is obvious, that the use of other, more powerful, statistical methods for the generation of the QSRR models will produce even better quantitative models with improved predictive abilities. However, in the present work we have preferred the use of the multivariate linear regression analysis in order to show the possibilities of the present theoretical approach in a simpler and universal form.

The present approach permits an easy interpretation of the QSRR models in terms of the chemical structure. This is the principal feature of this theoretical scheme to the modeling of chromatographic properties of organic compounds. For instances, the model describing the retention indices of alkanes in terms of the molecular structure has some similarities to that obtained to describe boiling points of alkanes by using the present approach [31], proving that these properties are dependent of approximately the same structural parameters. This fact provides a structural justification to the correlations obtained between these two experimental properties, as well as to the success obtained in the description of both properties with simple topological (graph theoretical) descriptors.

The analysis of the structural influences on the chromatographic retention indices of the oxygen-containing compounds also shows interesting features. These chromatographic parameters were obtained by using the non-polar stationary phase OV-1. Consequently, it is not expected that factors related to polarity or protic nature, such as the ability to participate in hydrogen bond donation or acceptance, of the compounds studied have a great influence on the retention processes in this phase. As in the case of alkanes, the molecular ‘‘size’’ and ‘‘branching’’ can be determining factors in the chromatographic retention of the alcohols, ketones and esters studied.

Table 2
Observed and calculated retention indices of oxygen-containing compounds

Compound	Observed I[a]	Calculated I[b]	Residual[c]	CV-res.[d]
2-Methyl-2-butanol	626.20	641.56	−15.36	−18.38
1-Butanol	646.50	663.00	−16.50	−18.67
3-Methyl-2-butanol	666.00	672.23	−6.23	−6.77
2-Pentanol	682.70	713.31	−30.61	−32.52
2-Methyl-2-pentanol	717.60	736.04	−18.44	−21.08
3-Methyl-1-butanol	719.30	712.50	6.79	7.18
4-Methyl-2-pentanol	744.10	763.58	−19.48	−20.91
1-Pentanol	750.40	755.97	−5.57	−6.17
2-Methyl-3-pentanol	758.00	765.79	−7.79	−8.19
2,4-Dimethyl-2-pentanol	775.90	789.16	−13.26	−15.08
3,3-Dimethyl-1-butanol	778.80	737.53	41.26	48.72
3-Hexanol	780.40	805.96	−25.56	−26.70
2-Methyl-2-hexanol	817.30	829.17	−11.86	−13.46
2-Methyl-1-pentanol	818.40	805.19	13.21	13.76
2,4-Dimethyl-3-pentanol	821.20	821.21	−0.01	−0.01
4-Methyl-1-pentanol	821.20	805.52	15.67	16.38
2,3-Dimethyl-3-pentanol	823.70	828.59	−4.89	−5.66
2-Ethyl-1-butanol	825.90	804.99	20.91	21.80
3-Methyl-1-pentanol	828.80	805.24	23.56	24.54
5-Methyl-3-hexanol	838.20	856.34	−18.14	−19.22
3-Ethyl-3-pentanol	843.10	850.09	−6.99	−7.78
1-Hexanol	853.00	848.94	4.06	4.50
4-Heptanol	875.40	899.20	−23.80	−25.16
2,2,4-Trimethyl-3-pentanol	881.50	882.67	−1.17	−1.40
3,5-Dimethyl-3-hexanol	883.10	892.83	−9.73	−10.64
2-Methyl-2-heptanol	916.40	922.15	−5.75	−6.63
6-Methyl-2-heptanol	951.10	948.86	2.24	2.45
4-Ethyl-3-hexanol	953.30	953.76	−0.46	−0.49
4-Octanol	975.50	992.22	−16.72	−18.28
3-Octanol	982.00	991.95	−9.95	−10.88
3,6-Dimethyl-3-heptanol	986.60	982.46	4.14	4.75
Ethyl acetate	600.00	615.06	−15.06	−16.50
Methyl propionate	615.20	613.70	1.50	1.65
Methyl isobutyrate	671.00	665.76	5.24	5.61
Ethyl propionate	694.20	707.50	−13.30	−14.07
Propyl acetate	696.30	708.04	−11.74	−12.40
Methyl butyrate	705.60	706.88	−1.27	−1.35
Ethyl isobutyrate	744.60	759.64	−15.04	−15.67
sec.-Butyl acetate	743.80	758.40	−14.60	−15.27
Isobutyl acetate	757.70	757.57	0.13	0.14
Methyl isopentanoate	761.30	757.09	4.20	4.39
Ethyl butyrate	784.00	800.68	−16.68	−17.40
Propyl propionate	792.60	800.48	−7.88	−8.22
Butyl acetate	796.20	801.01	−4.81	−5.00
Isopropyl butyrate	827.60	851.42	−23.83	−24.82
Ethyl isopentanoate	838.40	850.90	−12.50	−13.01
Isobutyl propionate	852.80	850.01	2.78	2.90
Propyl butyrate	881.50	893.66	−12.16	−12.79
1,3-Dimethylbutyl acetate	885.10	902.02	−16.93	−18.95
Butyl propionate	891.40	893.45	−2.05	−2.16
Pentyl acetate	896.40	893.98	2.41	2.53
Isobutyl isobutyrate	900.00	902.17	−2.17	−2.40

(*Continued on next page*)

Table 2. (*continued*)

Compound	Observed I[a]	Calculated I[b]	Residual[c]	CV-res.[d]
Methyl hexanoate	907.00	892.86	14.14	14.92
Isobutyl butyrate	940.30	943.19	−2.89	−3.07
2-Ethylbutyl actetae	957.00	943.41	13.58	14.41
Butyl butyrate	979.40	986.63	−7.23	−7.84
Ethyl hexanoate	982.90	986.67	−3.77	−4.08
Pentyl propionate	990.50	986.42	4.07	4.41
Hexyl acetate	996.50	986.95	9.54	10.30
3-Methyl-2-butanone	640.90	623.92	16.97	18.27
2-Pentanone	663.30	665.43	−2.13	−2.23
3-Pentanone	676.40	664.81	11.59	12.13
3,3-Dimethyl-2-butanone	693.10	672.01	21.08	23.51
4-Methyl-2-pentanone	721.20	715.60	5.59	5.85
2-Methyl-3-pentanone	733.00	717.33	15.66	16.19
4-Methyl-3-pentanone	733.00	717.33	15.66	16.19
3-Methyl-2-pentanone	734.80	717.45	17.35	17.93
3-Hexanone	764.80	758.01	6.79	6.95
2-Hexanone	767.90	758.44	9.46	9.71
2,4-Dimethyl-3-pentanone	779.00	772.03	6.97	7.51
5-Methyl-3-hexanone	816.70	808.29	8.41	8.63
2-Methyl-3-hexanone	820.00	810.64	9.36	9.56
5-Methyl-2-hexanone	836.50	808.07	28.42	29.37
4-Heptanone	853.40	851.22	2.18	2.23
3-Heptanone	865.80	851.02	14.77	15.12
2-Heptanone	868.70	851.41	17.28	17.72
2,2,4,4-Tetramethyl-3-pentanone	900.00	901.58	−1.58	−4.01
2,6-Dimethyl-4-heptanone	954.70	951.79	2.91	3.28
2,2-Dimethyl-3-heptanone	964.70	955.09	9.60	10.58
3-Octanone	966.00	944.00	22.00	22.98
2-Octanone	968.80	944.38	24.41	25.54

[a] Taken from Ref. [42].
[b] Calculated from Eq. (6).
[c] Calculated minus observed retention index.
[d] Residual in the leave-one-out cross validation.

By using the values of the contributions of the different structural fragments present in such molecules given in Table 3, we can analyze the influence of branching on the retention indices for this kind of compounds. For instance, in Fig. 4 we illustrate the contribution of the C–O bond in pentanol isomers to the retention indices. These contributions are calculated by summing the contributions of the different structural fragments that include the C–O bond in their structures. Here, a principle of equal partition of the contributions between the bonds is applied. This procedure is carried out as follows: if we consider the contribution of the fragment A_2, which is equal to −12.22 (see Table 3 and Fig. 3), in the molecule of *n*-pentanol, we assign values of −6.11 to the C–O and C–C bonds forming the fragment A_2 in such molecule. As can be seen in this figure the contribution of the C–O bond decreases when the branching increases: 1-pentanol (750.4)>2-methyl-1-butanol (682.7)>2-methyl-2-butanol (626.2). The same order is maintained by the total retention indices of these compounds, see Table 2, indicating the dominant role of the branching in the chromatographic retention process for these compounds in the OV-1 stationary phase. On the other hand, we also compare the contributions to the retention indices of some structural fragments representing three different functional groups having the same topological structure. These groups are those representing the secondary hydroxyl group in 2-pentanol, the car-

Table 3
Contributions of the different structural fragments in oxygen-containing compounds to the chromatographic retention indices

Fragment[a]	Contribution	Fragment[a]	Contribution	Fragment[a]	Contribution	Fragment[a]	Contribution
F1	118.44	A9	7.24	O15	8.41	E6	−13.24
F2	−13.04	A10	7.24	O16	8.36	E7	−12.92
F3	−41.47	A11	4.82	O17	1.21	E8	−13.37
F4	60.58	O1	114.48	O18	1.21	E9	−13.75
F5	−12.48	O2	5 4.99	O19	1.21	E10	8.31
F6	9.28	O3	−12.87	O20	1.21	E11	7.74
F7	1.21	O4	−12.70	O21	7.24	E12	7.97
F8	7.24	O5	−11.51	O22	7.24	E13	7.32
F9	4.82	O6	−41.29	O23	7.24	E14	1.21
A1	96.67	O7	−37.17	O24	7.24	E15	1.21
A2	−12.22	O8	59.59	O25	4.82	E16	1.21
A3	−40.38	O9	−12.52	O26	4.82	E17	1.21
A4	55.22	O10	−12.61	E1	82.79	E18	1.21
A5	−12.71	O11	−13.14	E2	−11.32	E19	7.24
A6	9.01	O12	9.23	E3	−11.75	E20	7.24
A7	9.01	O13	9.23	E4	−11.20	E21	4.82
A8	1.21	O14	9.23	E5	−34.94	Intercept	273.44

[a] Fragments are illustrated in Figs. 1 and 3.

bonyl group of the 2-pentanone, and the carboxyl group of the ethyl acetate. The contributions of these groups are given in Fig. 4. They are ordered as follows: hydroxyl>carbonyl>carboxyl, which correspond with the order of the chromatographic retention of the corresponding compounds: 2-pentanol (682.7)>2-pentanone (666.3)>ethyl acetate (600.0).

Finally, we will show how the chromatographic retention indices can be computed by summing the bond contributions computed by the present graph theoretical approach. By considering the three alcohols illustrated in Fig. 4 we will compute the bond contributions to the retention indices according to the labeling given in this figure. For instance, the contributions of the different bonds in 1-pentanol are as follows: (1) 103.41, (2) 98.84, (3) 103.08 and (4) 108.06; for 2-pentanol: (1) 90.74, (2) 76.12, (3) 95.69 and (4) 108.36; for 2-methyl-2-butanol: (1) 74.21, (2) 61.54 and (3) 107.63. We recall that in 2-methyl-2-butanol there are two equivalent bonds corresponding to the bond labeled as 1. By summing these bond contributions plus the intercept of Model (6) we obtain the calculated values of the retention indices of these alcohols. The same can be done for the calculation of bond contribution to the retention indices of any other compound. However, these bond contributions to the chromatographic retention indices are more suitable for inter-molecular than for intra-molecular comparisons. A more detailed analysis of this topic, which is very related to the principle of equal partition of the contributions applied here, will be considered in a forthcoming paper [43].

5. Conclusions

One of the main criticisms to the use of graph-theoretical descriptors in QSPR, QSAR and QSRR studies is concerned to interpretation of the resulting models. The topological indices are obtained by algebraic manipulation of graphs representing the structural skeleton of molecules. Of course, (molecular) graphs are mathematical objects that contain important structural information on molecules. However, in many cases the manipulation of graphs in order to develop the molecular descriptors is based on very convoluted algebraic operations producing descriptors that, in many cases, are very difficult to interpret in terms of the molecular structure. We claim that this is an unnecessary and non-elegant use of graph-theory in chemistry. On the other hand, there are quantitative models obtained by using series of topological indices that are so complicated

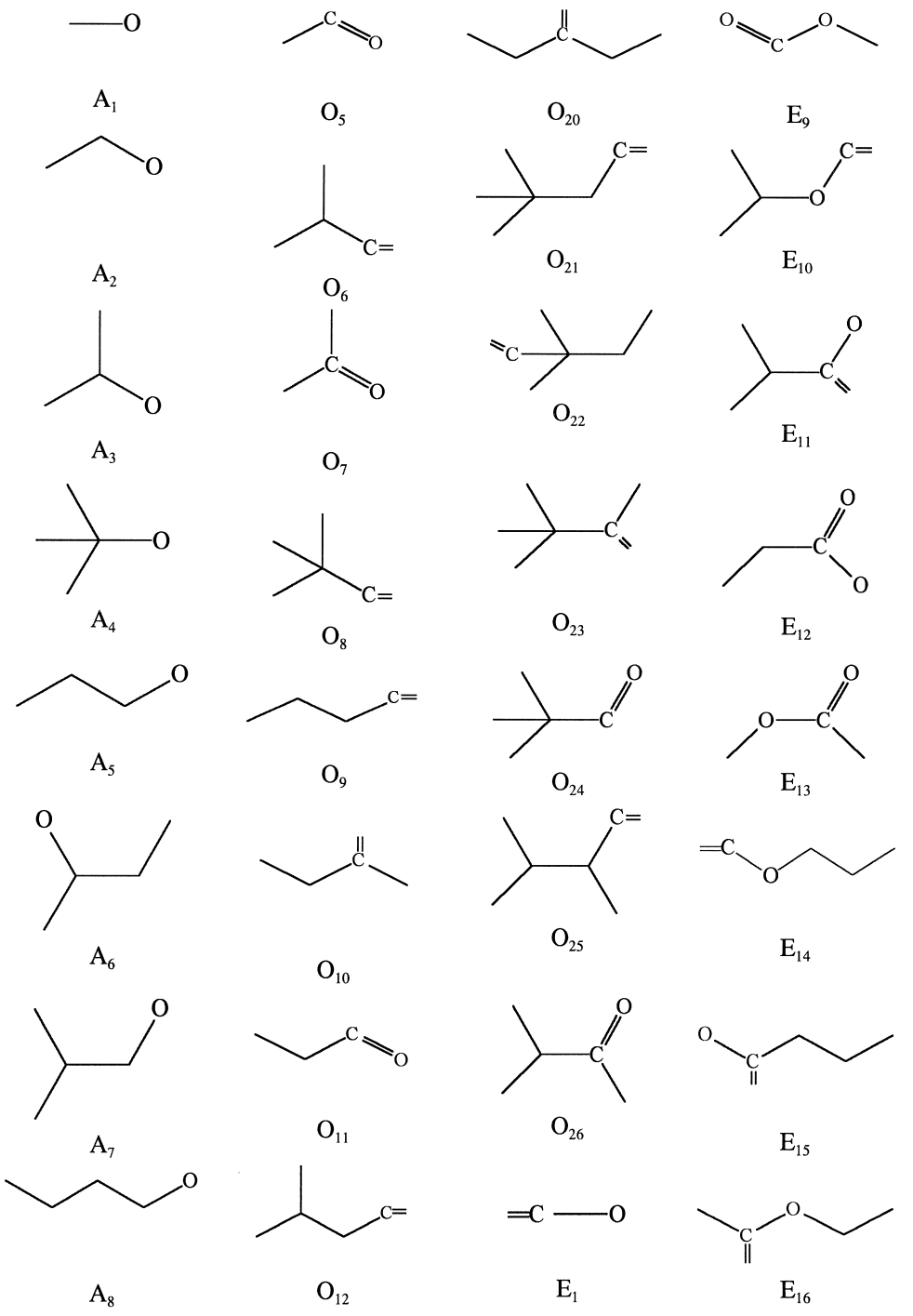

Fig. 3. Structural fragments contributing to the chromatographic retention of oxygen-containing molecules on OV-1 phase according to the model found here.

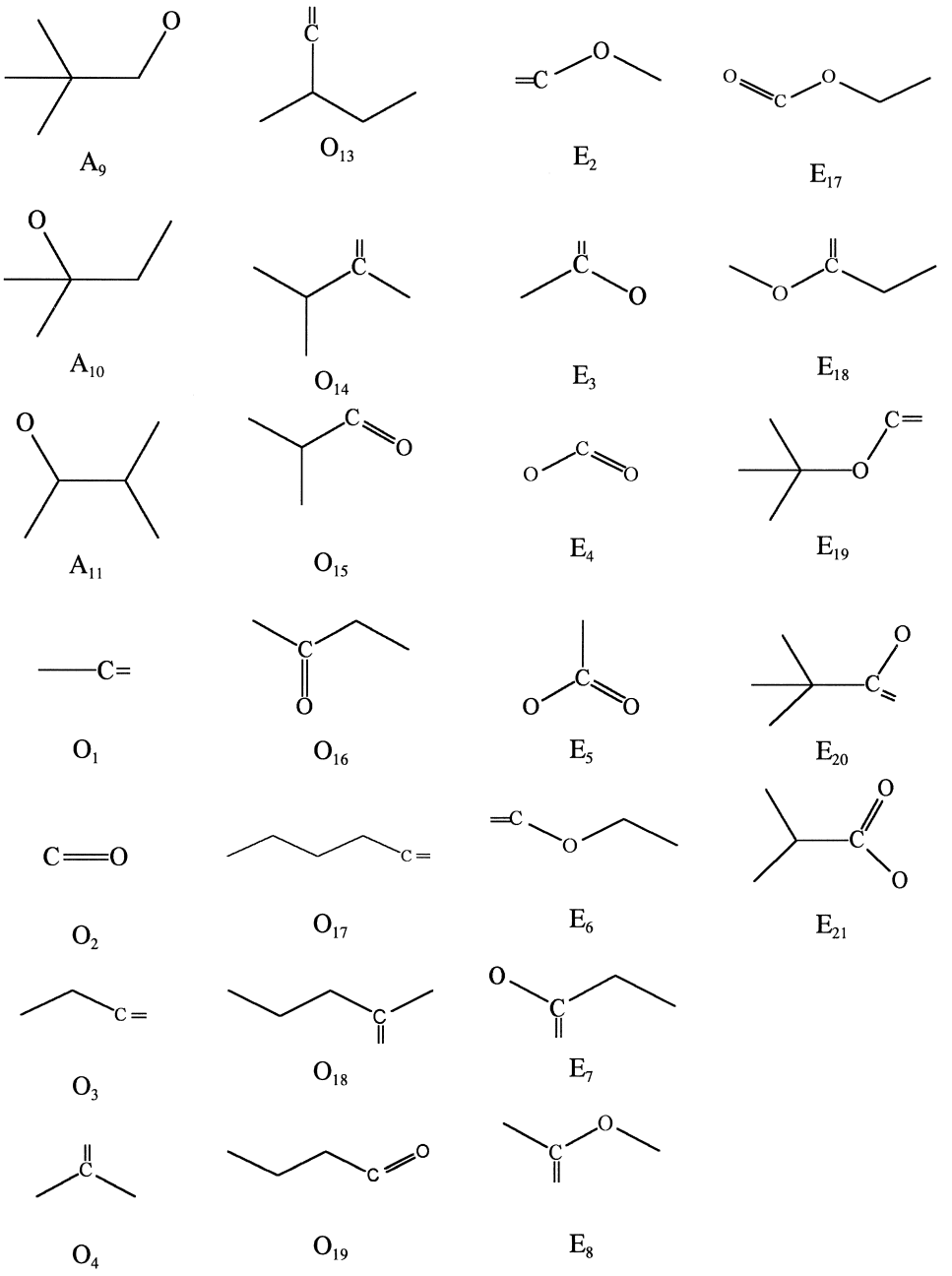

Fig. 3 (*continued*).

that we do not understand how the molecular structure determines the property. The present approach based on the calculation of the spectral moments of the bond matrix represents a step forward in the search of structurally interpretable graph-theoretical approaches in chemistry.

Appendix

The coefficients given in Eq. (3) which contain the information on the contribution of the different fragments to the chromatographic retention indices are given below:

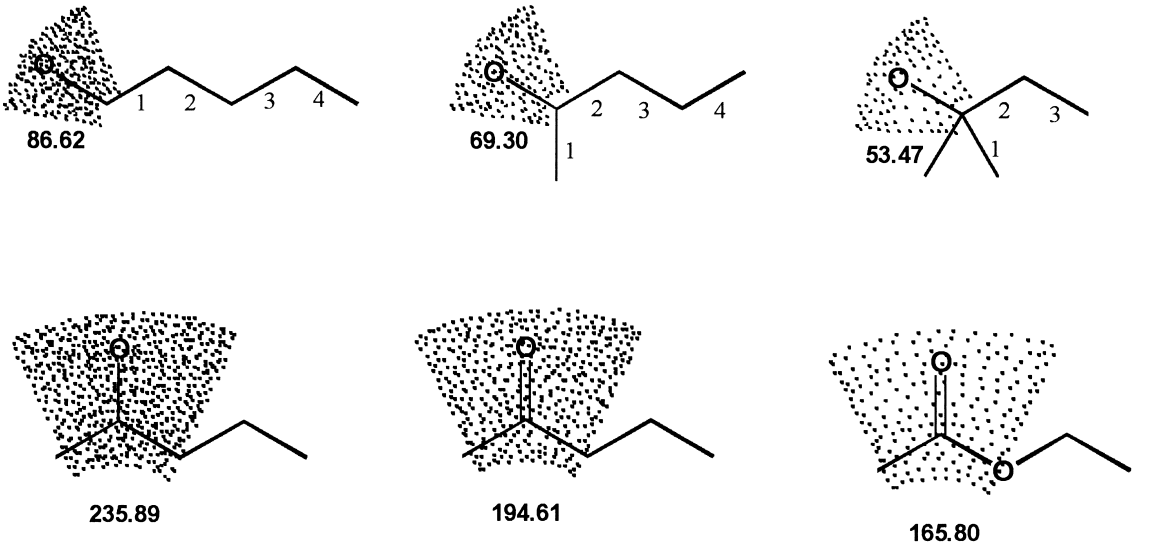

Fig. 4. Comparative visualization of the influence of branching and of the nature of functional groups on the chromatographic retention of oxygen-containing compounds on OV-1 stationary phase.

$$b_{AB} = a_0 + a_1 d_{AB} + a_2(d_{AB})^2 + a_3(d_{AB})^3 + a_4(d_{AB})^4 + a_5(d_{AB})^5 + a_6(d_{AB})^6 \quad (7)$$

$$b_{A(BC)} = 2a_2 + 3a_3(d_{AB} + d_{AC}) + a_4\{2 + 4[(d_{AB})^2 + (d_{AC})^2] + 4(d_{AB}d_{AC})\} + a_5[5(d_{AB}^3 d_{AC}^3) + 5(d_{AB}^2 d_{AC} + d_{AB}d_{AC}^2) + 5(d_{AB}d_{AC})] + a_6[6(d_{AB}^2 d_{AC}^2) + 6(d_{AB}d_{AC}^3 + d_{AB}^3 d_{AC}) + 6(d_{AB}^4 d_{AC}^4) + 9(d_{AB}^2 d_{AC}^2) + 12(d_{AB}d_{AC}) + 2] \quad (8)$$

$$b_{A(BCD)} = 6a_3 + a_4[12 + 8(d_{AB} + d_{AC} + d_{AD})] + a_5[30 + 10(d_{AB}^2 + d_{AC}^2 + d_{AD}^2) + 20(d_{AB} + d_{AC} + d_{AD}) + 10(d_{AB}d_{AC} + d_{AB}d_{AD} + d_{AC}d_{AD})] + a_6[12(d_{AB}^3 + d_{AC}^3 + d_{AD}^3) + 12(d_{AB}^2 + d_{AC} + d_{AB}d_{AC}^2 + d_{AB}d_{AD}^2 + d_{AB}^2 d_{AD} + d_{AC}d_{AD}^2 + d_{AC}^2 d_{AD}) + 30(d_{AB}^2 + d_{AC}^2 + d_{AD}^2) + 36(d_{AB}d_{AC} + d_{AB}d_{AD} + d_{AC}d_{AD}) + 60(d_{AB} + d_{AC} + d_{AD}) + 12(d_{AB}d_{AC}d_{AD}) + 60] \quad (9)$$

$$b_{A(BCDE)} = a_5[120 + 30(d_{AB} + d_{AC} + d_{AD}d_{AE})] + a_6[36(d_{AB}^2 + d_{AC}^2 + d_{AD}^2 + d_{AE}^2) + 24(d_{AB}d_{AC} + d_{AB}d_{AD} + d_{AB}d_{AE} + d_{AC}d_{AD} + d_{AC}d_{AE} + d_{AD}d_{AE}) + 180(d_{AB} + d_{AC} + d_{AD} + d_{AE}) + 480] \quad (10)$$

$$b_{A\text{-}B\text{-}C\text{-}D} = a_5[5(d_{AB} + d_{CD}) + 10d_{BC}] + a_6[6(d_{AB}^2 + d_{CD}^2) + 18d_{BC}^2 + 12(d_{AB}d_{BC} + d_{BC}d_{CD}) + 6(d_{AB}d_{CD}) + 12] \quad (11)$$

$$b_{A(BCD-E)} = 10a_5 + a_6[12(d_{AB} + d_{AC} + d_{DE}) + 24(d_{AD}) + 24] \quad (12)$$

In these expressions d_{AB} is the distance for the bond A–B and the a_i values are the coefficients in Model (1). These expressions permit us to calculate the contribution of any fragment in the molecules to the chromatographic property studied *P*. Of course, atoms symbolized by letters *A*, *B*,..., *E* can be any of the atoms that normally appear in organic molecules, permitting the calculation of the contributions of any kind of structural fragment by a simple substitution

of the corresponding bond distances into Eqs. (7)–(12).

References

[1] M. Randić, J. Chem. Inf. Comput. Sci. 37 (1997) 672–687.
[2] R. Kaliszan, Quantitative Structure–Chromatographic Retention Relationships, Wiley-Interscience, New York, 1987.
[3] R. Kaliszan, J. Chromatogr. A 656 (1993) 417–435.
[4] T.F. Woloszyn, Computer-Assisted Studies in Quantitative Structure–Retention Relationships, Report, Order No. AD-A251901, NTIS. Gov. Rep. Announce. Index (US) 1992, 92 (19).
[5] Y. Baba, J. Chromatogr. B 618 (1992) 41–55.
[6] R.D. Brauer, T.E. Bitterwolf, N.G. Smart, M.D. Burford, C.M. Wai, Anal. Chim. Acta 349 (1997) 239–244.
[7] T.J. Betts, J. Chromatogr. A 732 (1996) 408–413.
[8] Y.C. Guillaume, C. Guinchard, J. Phys. Chem. B 101 (1997) 8388–8392.
[9] L. Rohrschreider, Chromatographia 42 (1996) 489–493.
[10] G.I.C. Simpson, Y.A. Jackson, J. Chromatogr. A 719 (1996) 462–467.
[11] M. Randić, J. Chromatogr. 161 (1978) 1–14.
[12] D. Amić, D. Davidović-Amić, N. Trinajstić, J. Chem. Inf. Comput. Sci. 35 (1995) 136–139.
[13] R. Gautzsch, P. Zinn, Chromatographia 43 (1996) 163–176.
[14] G.M. Anton-Fos, F.G. García-March, F. Perez-Jimenez, M.T. Salabert-Salvador, R.A. Cecos-del-Pozo, J. Chromatogr. A 672 (1994) 203–211.
[15] N. Dimov, A. Osman, O. Mekenyan, D. Papazova, Anal. Chim. Acta 289 (1994) 303–317.
[16] J.V. de Julian-Ortiz, R. García-Domenech, J. Galvez-Alvarez, R. Soler Roca, F.J. García-March, G.M. Anton-Fos, J. Chromatogr. A 719 (1996) 37–44.
[17] L. Buydens, D.L. Massart, Anal. Chem. 55 (1983) 738–744.
[18] T.F. Woloszyn, P.C. Jurs, Anal. Chem. 64 (1992) 3059–3063.
[19] T.F. Woloszyn, P.C. Jurs, Anal. Chem. 65 (1993) 582–587.
[20] L.M. Egolf, P.C. Jurs, Anal. Chem. 65 (1993) 3119–3125.
[21] A.R. Katritzky, E.S. Ignatchenko, R.A. Barkock, V.S. Lobanov, M. Karelson, Anal. Chem. 66 (1994) 1799–1807.
[22] J.M. Sutter, T.A. Petterson, P.C. Jurs, Anal. Chim. Acta 342 (1997) 113–122.
[23] P. Payares, D. Diaz, J. Olivero, R. Vivas, I. Gomez, J. Chromatogr. A 771 (1997) 213–219.
[24] V.E.F. Heinzen, R.A. Yunes, J. Chromatogr. A 719 (1996) 462–467.
[25] Z. Mihalić, N. Trinajstić, J. Chem. Educ. 69 (1992) 701–712.
[26] M. Randić, J. Math. Chem. 9 (1992) 97–146.
[27] E. Estrada, J. Chem. Inf. Comput. Sci. 35 (1995) 31–33.
[28] E. Estrada, J. Chem. Inf. Comput. Sci. 35 (1995) 701–708.
[29] E. Estrada, A. Ramirez, J. Chem. Inf. Comput. Sci. 36 (1996) 837–843.
[30] E. Estrada, N. Guevara, I. Gutman, J. Chem. Inf. Comput. Sci. 38 (1998) 428–431.
[31] E. Estrada, J. Chem. Inf. Comput. Sci. 36 (1996) 844–849.
[32] E. Estrada, J. Chem. Inf. Comput. Sci. 37 (1997) 320–328.
[33] E. Estrada, J. Chem. Inf. Comput. Sci. 38 (1998) 23–27.
[34] E. Estrada, J. Chem. Soc., Faraday Trans. 94 (1998) 1407–1410.
[35] E. Estrada, A. Peña, R. García-Domenech, J. Comput.-Aided Mol. Des. 12 (1998) 583–595.
[36] N. Trinajstić, Chemical Graph Theory, 2nd revised ed., CRC, Boca Raton, FL, 1992.
[37] V.M. Potapov, Stereochemistry, MIR, Moscow, 1978.
[38] J.R. Chretien, J.-E. Dubois, J. Chromatogr. 126 (1976) 171.
[39] G. Schomburg, G. Dielman, J. Chromatogr. Sci. 11 (1973) 151.
[40] N. Bosnjak, Z. Mihalić, N. Trinajstić, J. Chromatogr. 540 (1991) 430–440.
[41] The Sadtler Standard Gas Chromatography Retention Index Library, Sadtler Research Laboratories, Philadelphia, PA, 1985.
[42] Ch. Duvenbeck, P. Zinn, J. Chem. Inf. Comput. Sci. 33 (1993) 211–219.
[43] E. Estrada, in preparation.