Journal of Chromatography, 630 (1993) 251–256
Elsevier Science Publishers B.V., Amsterdam

CHROM. 24 681

Relationship between Kováts retention indices and molecular connectivity indices of tetralones, coumarins and structurally related compounds

Ana C. Arruda, Vilma E. F. Heinzen and Rosendo A. Yunes

Departamento de Química, Universidade Federal de Santa Catarina, 88040-900 Florianópolis, Santa Catarina (Brazil)

(First received June 2nd, 1992; revised manuscript received August 18th, 1992)

ABSTRACT

A study was undertaken to test the ability of several molecular connectivity indices to predict the retention indices (I) of tetralones, coumarins and structurally related compounds determined on OV-17 and Apiezon L as stationary phases. The regression analyses with $I_{OV\text{-}17}$ showed that a two-variable linear regression equation with $^2\chi$ and $^4\chi^v_{pc}$ gives the best correlation coefficient, suggesting that retention depends basically on branching and the presence and number of adjacent atoms, and secondarily on unsaturations and the number and orientation of substituents. I_{ApL} gives the best correlation with a two-variable linear regression equation with $^1\chi$ and $^3\chi^v_p$ indicating that on this stationary phase retention depends basically on the presence and number of adjacent atoms, and secondarily on unsaturation, branching of adjacent atoms and the presence of heteroaroms. ΔI ($\Delta I = I_{OV\text{-}17} - I_{ApL}$), according to the different polarities of both phases, considered to be a measure of the polar forces in retention, does not give a good correlation.

INTRODUCTION

Over the last few years, many workers have observed a good correlation between experimental retention indices and topological indices, such as the molecular connectivity index, first introduced by Randic [1] and later developed and extensively used by Kier [2].

With the discovery of topological indices, the capacity for the prediction of the chemical properties of substances is now becoming a reality. Although still in its early stages, it can already claim substantial success in a broad range of applications.

Thus, the topological method can be used to predict physico-chemical properties [2], chromatographic retention indices [3–8] and the extent to which various pollutants might spread in the environment and the harm they might do once they have spread [9–11]. It can also be used to develop new anaesthetics and psychoactive drugs [12].

Kier [2] has shown good correlations between connectivity indices and psycho-chemical properties, such as density, boiling point and water solubility, and also molecular surface area for a series of alkanes.

Szász *et al.* [13] demonstrated the relationship between the partition data obtained by gas–liquid chromatography and molecular connectivity indices of derivatives of pyrido [1,2-*a*]pyrimidin-4-one. Szász *et al.* [14] also verified that the molecular connectivity indices can be used quantitatively to describe the gas chromatographic retention indices of a series of derivatives of pyrido[2,1-*a*]pyrimidine and pyrido[2,1-b]quinazoline. Sabljic [15] has shown that high correlation coefficients and a correct elution sequence are necessary for predicting retention indices.

This work was carried out to test the ability of the molecular connectivity method in predicting the Kováts retention indices (I) of tetralones and sub-

Correspondence to: R. A. Yunes, Departamento de Química, Universidade Federal de Santa Catarina, 88040-900 Florianópolis, Santa Catarina, Brazil.

0021-9673/93/$06.00

stituted coumarins, using both correlation coefficients and correctly predicted elution sequence as criteria of fit, and also to determine the structural factors that are important in the chromatographic behaviour of the compounds studied.

EXPERIMENTAL

Samples

The tetralones, coumarins, tetrahydronaphthalene and monocyclic compounds studied are indicated in Table I. Most of them were obtained commercially and the others were synthesized.

TABLE I

EXPERIMENTAL RETENTION INDICES AND ΔI VALUES AT 170°C FOR TETRALONES, COUMARINS AND STRUCTURALLY RELATED COMPOUNDS ON NON-POLAR (APIEZON L) AND POLAR (OV-17) STATIONARY PHASES

No.	Compound	Retention index (I)		ΔI
		OV-17	Apiezon L	
1	Cyclohexane	757	–	–
2	Methylcyclohexane	793	–	–
3	Benzene	774	–	–
4	Toluene	887	–	–
5	Cyclohexanone	1080	–	–
6	Methoxybenzene	1085	–	–
7	Tetrahydro-4*H*-pyran-4-one	1088	–	–
8	δ-Valerolactone	1107	–	–
9	2-Methylcyclohexanone	1119	–	–
10	4-Methylcyclohexanone	1144	–	–
11	3-Methylcyclohexanone	1155	–	–
12	Tetrahydronaphthalene	1348	–	–
13	2-Coumaronone	1501	1289	212
14	4-Chromanone	1621	1390	231
15	β-Tetralone	1625	1503	122
16	α-Tetralone	1645	1458	187
17	2-Methyl-1-tetralone	1667	1492	175
18	1-Methyl-2-tetralone	1669	1479	190
19	Dihydrocoumarin	1682	1434	248
20	4-Methyl-1-tetralone	1688	1514	174
21	Coumarin	1758	1521	237
22	6-Methylcoumarin	1863	1618	245
23	7-Methoxy-1-tetralone	1897	1666	232
24	7-Methoxy-2-tetralone	1903	1680	223
25	5-Methoxy-1-tetralone	1905	1690	219
26	6-Methoxy-2-tetralone	1910	1711	199
27	6-Methoxy-1-tetralone	1970	1820	162
28	7-Methoxycoumarin	2056	1798	258
29	4-Methoxycoumarin	2085	1844	241
30	7-Methoxy-4-methylcoumarin	2214	2021	193

The solutions were prepared with carbon tetrachloride at concentrations lower than 200 μg/ml, with the exception of δ-valerolactone, prepared with methyl ethyl ketone, and 7-methoxy-4-methylcoumarin, prepared with methanol. In the solutions of β-tetralone, 6-methoxy-2-tetralone and 7-methoxy-2-tetralone, ascorbic acid was utilized as an antioxidant and the solution was saturated with nitrogen.

Methods

Samples (1.0 μl) were injected into a gas chromatograph equipped with a flame ionization detector. A glass column (1.8 m × 3.2 mm I.D.) packed with 3% OV-17 on Chromosorb W AW DMCS (80–100 mesh) and two nickel columns, one (1.8 m × 3.2 mm I.D.) with 15% Apiezon L on Chromosorb W and the other (5.5 m × 5.3 mm I.D.) with 3% OV-17 on Chromosorb W, were utilized. The samples (monocyclic compounds) that have low boiling points and consequently short retention times were studied with the longer nickel column. The carrier gas was nitrogen at a flow-rate of 30 ml/min in all instances.

Conditions that gave symmetrical chromatograms were selected. The temperature range using the OV-17 columns was 170–210°C for tetralones, 210–250°C for substituted coumarins, 170–200°C for tetrahydronaphthalene, 6-methylcoumarin and un-substituted aromatic lactones and 80–140°C for different monocyclic compounds. Using Apiezon L columns the ranges were 210–230°C for tetralones, 220–280°C for some methoxytetralones and substituted coumarins and 170–232°C for un-substituted aromatic lactones and 6-methylcoumarin. The measurements were made at four different temperatures in order to decrease the errors in extrapolations. For the same reason, 170°C was chosen for extrapolation of the I values on OV-17 because it is located between the ranges of temperatures utilized to determine the volatile (80–140°C) and slightly volatile compounds (210–280°C).

Calculation

The connectivity indices ${}^{1}\chi$, ${}^{2}\chi$, ${}^{3}\chi_{p}$, ${}^{4}\chi_{p}$, ${}^{3}\chi_{pc}$, ${}^{4}\chi_{pc}$, ${}^{1}\chi^{v}$, ${}^{2}\chi^{v}{}_{p}$, ${}^{3}\chi^{v}{}_{c}$ and ${}^{4}\chi^{v}{}_{pc}$ were calculated by the method of Kier [2] utilizing a Molconn-X computer program for molecular topology analysis. They are given in Table II.

TABLE II

CONNECTIVITY INDICES OF TETRALONES, COUMARINS AND STRUCTURALLY RELATED COMPOUNDS

No.[a]	Connectivity index										
	$^{1}\chi$	$^{2}\chi$	$^{3}\chi_{p}$	$^{4}\chi_{p}$	$^{3}\chi_{pc}$	$^{4}\chi_{pc}$	$^{1}\chi^{v}$	$^{2}\chi^{v}$	$^{3}\chi_{p}^{v}$	$^{3}\chi_{c}^{v}$	$^{4}\chi_{pc}^{v}$
1	3.0000	2.1213	1.5000	1.0607	–	–	3.0000	2.1213	2.1213	–	–
2	3.3938	2.7432	1.8938	1.3067	0.2887	0.4082	3.3938	2.7432	1.8938	0.2887	0.4082
3	3.0000	2.1213	1.5000	1.0607	–	–	2.0000	1.1547	0.6667	–	–
4	3.3938	2.7432	1.8938	1.3067	0.2887	0.4082	2.4107	1.6547	0.9405	0.1667	0.1924
5	3.3938	2.7432	1.8938	1.3067	0.2887	0.4082	2.9112	2.0993	1.4111	0.1020	0.1443
6	3.9319	2.9123	2.3021	1.5954	0.2041	0.4928	2.5231	1.5172	0.9789	0.0680	0.1466
7	3.3938	2.7432	1.8938	1.3067	0.2887	0.4082	2.4886	1.6510	1.0505	0.1020	0.1443
8	3.3938	2.7432	1.8938	1.3067	0.2887	0.4082	2.5505	1.6776	1.0383	0.0589	0.0833
9	3.8045	3.2388	2.5403	1.5017	0.4714	1.1381	3.3319	2.6131	1.8600	0.2875	0.5732
10	3.7877	3.3650	2.3045	1.4267	0.5773	0.8165	3.3048	2.7212	1.8320	0.3907	0.5526
11	3.7877	3.3764	2.1986	1.7374	0.5773	0.7416	3.3051	2.7402	1.7287	0.3907	0.4796
12	4.9663	4.0891	3.4663	2.8576	0.3333	0.9428	4.0345	2.9758	2.2606	0.2041	0.5244
13	4.4663	3.7355	3.2163	2.6868	0.3333	0.9428	2.8892	1.9584	1.3616	0.1423	0.3144
14	5.3770	4.6166	3.9332	3.2600	0.5384	1.4402	3.6278	2.5178	1.7742	0.2033	0.4612
15	5.3602	4.7228	3.8019	3.1081	0.6220	1.2920	3.9457	2.9842	2.1792	0.3062	0.6264
16	5.3770	4.6166	3.9332	3.2600	0.5384	1.4402	3.9886	2.9339	2.1677	0.2464	0.5847
17	5.7877	5.1214	4.5336	3.4488	0.7309	2.0673	4.4093	3.4611	2.5771	0.4373	0.9403
18	5.7877	5.1214	4.5336	3.4488	0.7309	2.0673	4.3933	3.3868	2.6453	0.4131	1.0295
19	5.3602	4.7228	3.8019	3.1081	0.6220	1.2920	3.6468	2.5818	1.7949	0.2199	0.4420
20	5.7877	5.1441	4.4140	3.5552	0.7436	1.9455	4.4093	3.4501	2.5965	0.4318	0.9787
21	5.3602	4.7228	3.8019	3.1081	0.6220	1.2920	3.3504	2.2923	1.5285	0.1904	0.3617
22	5.7540	5.3566	4.1374	3.3585	0.9107	1.6411	3.7605	2.7957	1.7748	0.3570	0.5348
23	6.3089	5.4189	4.6993	3.7013	0.7426	1.8888	4.5116	3.2998	2.4674	0.3144	0.7204
24	6.2920	5.5257	4.5612	3.5967	0.8261	1.7367	4.4687	3.3501	2.4742	0.3742	0.7599
25	6.3257	5.3351	4.7200	3.9462	0.6745	1.9348	4.5176	3.2631	2.4840	0.2917	0.7470
26	6.2920	5.5257	4.5612	3.5918	0.8261	1.7367	4.4687	3.3501	2.4742	0.3742	0.7599
27	6.3089	5.4184	4.6925	3.7526	0.7426	1.8849	4.5116	3.2996	2.4628	0.3144	0.7181
28	6.2920	5.5257	4.5612	3.5967	0.8261	1.7367	3.8735	2.6582	1.8303	0.2584	0.4985
29	6.3089	5.4526	4.5080	3.9237	0.7581	1.7496	3.8794	2.6300	1.8319	0.2381	0.5109
30	6.7027	6.0651	4.4374	4.1981	1.0313	2.1660	4.2901	3.1105	2.1585	0.3916	0.7662

[a] See Table I.

To test the quality of the regression equation, the following statistical parameters were used: correlation coefficient (r), test of the null hypothesis (F-test) and Student's t-test. All calculations of single and multiple linear regression analyses were carried out on an IBM PC/XT computer.

The Ambrus [16] method was used to calculate the dead time for each column in order to obtain the adjusted retention indices that were used in all the correlations.

Experimental retention times are averages of three sample injections.

RESULTS AND DISCUSSION

It was observed that the relationship between the retention index of non-polar substances and temperature on a non-polar stationary phase is linear. However, on polar stationary phases and with different kinds of compounds, it is not linear. Thus Hoigné *et al.* [17] demonstrated that over large temperature ranges the retention index is a hyperbolic function. In this work, extrapolations were done in some instances assuming that the range of linearity will be large, considering the low polarity of the

compounds and of the stationary phases (OV-17, Apiezon L).

The adjusted retention indices obtained on the OV-17 column show excellent linearity with temperature, with correlation coefficients greater than 0.9700, over the range of experimental temperatures.

On the non-polar stationary phase (Apiezon L) the correlation coefficients are greater than 0.9800, with the exception of 4-methoxycoumarin (r = 0.9686), 4-chromanone (r = 0.9666) and 7-methoxy-4-methylcoumarin (r = 0.9568).

The values of the I adjusted or extrapolated to 170°C on the OV-17 and Apiezon L columns are given in Table I.

Application of the molecular connectivity method

Simple linear correlation. The best single linear regression equations between the I values on the polar OV-17 stationary phase for the 30 compounds studied and the connectivity indices were obtained with $^1\chi$ and $^2\chi$, as indicated by eqns. 1 and 2 and their statistical parameters:

$$I_{OV\text{-}17} = 341.1620\ ^1\chi - 188.8422$$
$$r = 0.9751,\ F = 541.92\ (P > 0.0001),\ r^2 = 0.9509 \quad (1)$$

$$I_{OV\text{-}17} = 347.5188\ ^2\chi + 27.6842$$
$$r = 0.9769,\ F = 586.09\ (P > 0.0001),\ r^2 = 0.9544 \quad (2)$$

It should be noted that $^1\chi$, which conveys more information about the number of atoms in a molecule, was not able to discriminate among 8 groups of compounds in a total of 23 compounds (3 groups of 2 compounds, 4 groups of 3 compounds and 1 group of 5 compounds) and $^2\chi$, which encodes more information about branching, among 7 groups of compounds in a total of 19 compounds (4 groups of 2 compounds, 2 groups of 3 compounds and 1 group of 5 compounds).

The inability of these indices to discriminate between such a number of compounds shows that they are not appropriate for the effective prediction of experimental retention indices considering Sabljic and Protic's criteria [11].

When the single linear regression equation is applied for the retention indices, $I_{OV\text{-}17}$ of bicyclic compounds numbered from 12 to 30 (19 substances), the correlation coefficients with $^1\chi$ (r = 0.8955) and with $^2\chi$ (r = 0.8999) are lower than that corresponding to the 30 compounds.

On the non-polar phase Apiezon L, a single linear regression equation between the retention indices of bicyclic compounds numbered from 13 to 30 (18 substances) and the different connectivity indices studied gives low correlation coefficients, $^1\chi$ (r = 0.9046) and $^2\chi$ (r = 0.8932).

When the ΔI values ($I_{OV\text{-}17} - I_{ApL}$), for the compounds numbered from 13 to 30 are correlated with the above-indicated molecular connectivity indices, the correlation coefficients are very low. ΔI is considered to be a measure of the polar forces in chromatographic retention. ΔI does not measure the dispersion forces because these are considered to be the same on polar and non-polar phases. Hence the low correlation with ΔI suggests that the connectivity indices basically give information about the dispersion forces that act on the chromatographic retention. This has also been observed by other workers [11].

Multiple linear correlation. Despite the high values of the correlation coefficient obtained with $^1\chi$ and $^2\chi$ through a simple linear regression equation, the lack of discrimination among many compounds makes them incapable of determining the correct elution sequence. Consequently, two-variable regression equations, which give a more complete representation of the molecules, were tested to find the best equations able to discriminate all the compounds and predict their correct elution.

A correlation matrix was applied to select the connectivity indices, and the scatterplot method was used to define the type of function that relates the retention to the connectivity indices.

The best two-variable regression equation for $I_{OV\text{-}17}$ of the 28 compounds and its statistical parameters was

$$I_{OV\text{-}17} = 425.0345\ ^2\chi - 422.5803\ ^4\chi^v_{pc} - 75.2959$$
$$r = 0.9882,\ F(^2\chi) = 633.63\ (P > 0.0001)$$
$$r^2 = 0.9765,\ F(^4\chi^v_{pc}) = 36.46\ (P > 0.0001) \quad (3)$$

The connectivity indices $^2\chi$ and $^4\chi^v_{pc}$ do not distinguish cyclohexanone from tetrahydro-4*H*-pyran-4-one because the connectivity index $^2\chi$ does not encode information about heteroatoms, and they do not distinguish 7-methoxy-2-tetralone from

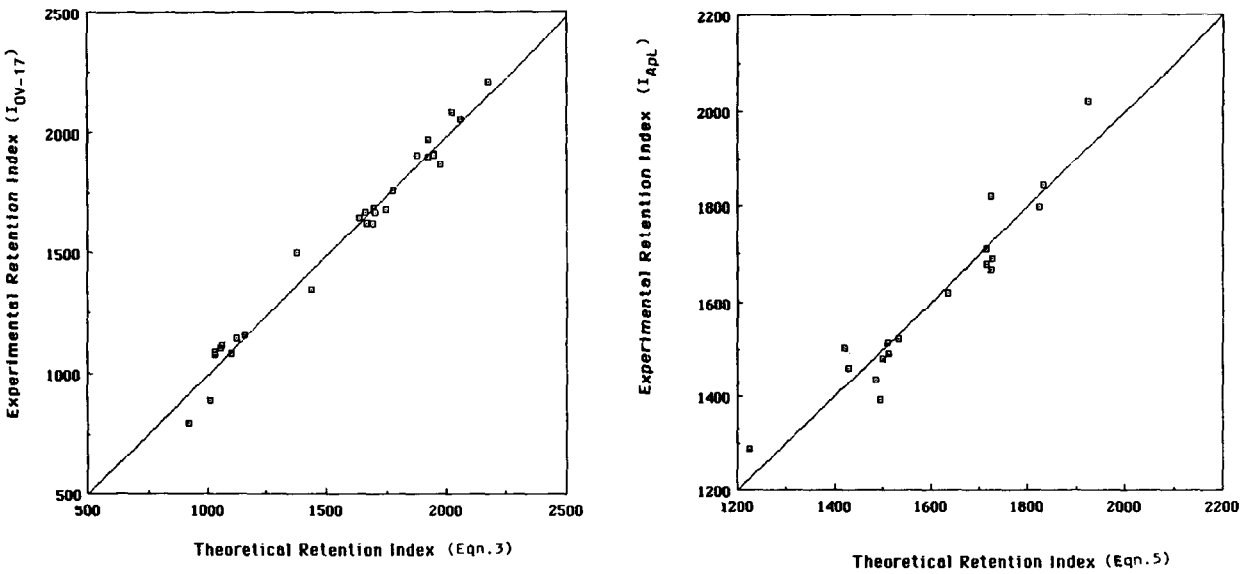

Fig. 1. Correlation between experimental I_{OV-17} and I_{ApL} values and calculated retention indices (eqns. 3 and 5) for tetralones, coumarins and structurally related compounds.

6-methoxy-2-tetralone owing to the inability of the connectivity index $^4\chi^v_{pc}$ to distinguish this kind of isomer.

For the bicyclic compounds (19 compounds) the two-variable regression equation for I_{OV-17} with $^2\chi$ and $^4\chi^v_{pc}$ is

$$I_{OV-17} = 425.6130\ ^2\chi - 431.2538\ ^4\chi^v_{PC} - 87.2267$$
$$r = 0.9651,\ F(^2\chi) = 211.05\ (P > 0.0001)$$
$$r^2 = 0.9315,\ F(^4\chi^v_{pc}) = 28.36\ (P > 0.0001) \quad (4)$$

As can be observed from eqns. 3 and 4 relative to eqns. 1 and 2, the chromatographic retention of the molecules studied depends basically on the presence and the number of adjacent atoms, and secondarily on the unsaturation and the number and orientation of substituents. The addition of these last factors gives a better differentiation of the compounds.

The best two-variable regression equation for I_{ApL} of 18 bicyclic compounds is with $^1\chi$ and $^3\chi^v_p$ (eqn. 5) considering the correlation coefficient and the discrimination of the compounds. It does not distinguish 7-methoxy-2-tetralone from 6-methoxy-2-tetralone.

$$I_{ApL} = 372.1000\ ^1\chi - 170.3856\ ^3\chi^v_p - 203.3994$$
$$r = 0.9548,\ F(^1\chi) = 144.26\ (P > 0.0001)$$
$$r^2 = 0.9263,\ F(^3\chi_p) = 15.87\ (P > 0.0012) \quad (5)$$

A graphical representation of the fit between the experimental I_{OV-17} and I_{ApL} values and those calculated by eqns. 3 and 5 is shown in Fig. 1.

The correlation between $\Delta I = (I_{OV-17} - I_{ApL})$ through the multiple linear regression equation, with two connectivity indices, did not show statistically acceptable correlation coefficients.

CONCLUSIONS

The retention indices showed good correlation coefficients through a simple linear correlation equation with some molecular connectivity indices ($^1\chi$, $^2\chi$). However, the inability of these indices to discriminate among many compounds makes these correlations inadequate for predicting the correct elution sequence of the compounds.

In order to discriminate better between the compounds studied a two-variable linear regression equation that gives a more complete representation of the molecules was applied. The correlation coefficients and the capacity for discrimination were improved in all instances. The best correlation for I_{OV-17} was found with $^2\chi$ and $^4\chi^v_{pc}$, showing that branching, unsaturation and the presence and orientation of the substituents in the aromatic rings are important factors in differentiating the retention

indices of the compounds. $^{2}\chi$ and $^{4}\chi^{v}_{pc}$ fail to distinguish only two compounds among 30 compounds analyzed.

The best correlation for I_{ApL} was found with $^{1}\chi$ and $^{3}\chi^{v}_{p}$, showing in this instance that the presence and the number of adjacent atoms are the main factors in the retention and the unsaturation and the presence of heteroatoms are secondarly factors.

The correlations of the ΔI values with the connectivity indices are not statistically acceptable with either simple or two-variable linear regression equations. As ΔI is a measure of the polar forces in the retention, it can be concluded that the connectivity indices basically give information about the dispersion forces in retention.

ACKNOWLEDGEMENTS

The authors thank CNPq (Brazil) for financial support and also Professors Pedro A. Barbetta and José Francisco Fletes of the Departamento de Ciencias Estatísticas e da Computaçao (UFSC) for useful discussions of the statistical methods.

REFERENCES

1 M. Randic, *J. Am. Chem. Soc.*, 97 (1975) 6609.
2 L. B. Kier, in S. H. Yakowsky, A. Sinkula and S. C. Valvani (Editors), *Physical Chemical Properties of Drugs*, Marcel Dekker, New York, 1980, Ch. 9, p. 277.
3 A. Robbat, Jr., G. Xyrafas and D. Marshall, *Anal. Chem.*, 60 (1988) 982.
4 K. Heberger, *Chromatographia*, 25 (1988) 725.
5 V. A. Gerasimento and V. M. Nabivach, *J. Chromatogr.*, 498 (1990) 357.
6 M. Kuchar, H. Tomková, V. Rejholec and D. Slalicka, *J. Chromatogr.*, 333 (1985) 21.
7 O. Papp, Gy. Szász, M. Farkas, G. Simon and I. Hermecz, *J. Chromatogr.*, 403 (1987) 19.
8 M. N. Hasan and P. C. Jurs, *Anal. Chem.*, 62 (1990) 2318.
9 A. Sabljic, *Z. Gesamte Hyg. Ihre Grenzgeb.*, 33 (1987) 493.
10 R. Koch, *Toxicol. Environ. Chem.*, 6 (1983) 87.
11 A. Sabljic and M. Protic, *Chem. Biol. Interact.*, 42 (1982) 301.
12 L. B. Kier and R. L. Hall, *J. Med. Chem.*, 20 (1977) 1631.
13 G. Szász, K. Valkó, O. Papp and I. Hermecz, *J. Chromatogr.*, 243 (1982) 347.
14 G. Szász, O. Papp, J. Vámos, K. Hankó-Novák and L. B. Kier, *J. Chromatogr.*, 269 (1983) 91.
15 A. Sabljic, *J. Chromatogr.*, 314 (1984) 1.
16 L. Ambrus, *J. Chromatogr.*, 294 (1984) 328.
17 J. Hoigné, J. Widmer and T. Gaumann, *J. Chromatogr.*, 11 (1963) 459.