Retention of Aromatic Sulfur-Containing Compounds on RP-HPLC: Correlation with Partition Coefficients and Molecular Connectivity Indices

Hui Hong[a]; Dai Zhou[a]; Shuokui Han[a]; Liansheng Wang[a]; Zheng Zhang[b]; Gongwei Zou[b]

[a] Department of Environmental Science & Engineering, Nanjing University, Nanjing, P. R. China [b] Department of Chemistry, Nanjing University, Nanjing, P. R. China

PLEASE SCROLL DOWN FOR ARTICLE

RETENTION OF AROMATIC SULFUR-CONTAINING COMPOUNDS ON RP-HPLC: CORRELATION WITH PARTITION COEFFICIENTS AND MOLECULAR CONNECTIVITY INDICES

Hui Hong,[†*] Dai Zhou,[†] Shuokui Han,[†] Liansheng Wang,[†] Zheng Zhang,[‡] Gongwei Zou[‡]

[†]Department of Environmental Science & Engineering
and
[‡]Department of Chemistry
Nanjing University
Nanjing, P. R. China 210093

ABSTRACT

The partition coefficients (log K_{ow}) in n-octanol/water, together with the HPLC capacity factors (k') measured on a C_{18} column, using methanol-water mobile phase of different compositions, were determined for 27 aromatic sulfur-containing compounds. The linear relationship between the log k' values and the percentage of the methanol in the eluent was tested for each compound. The best percentage of methanol (61%), by which the chromatographic partition system became the best model of the n-octanol/water partition system, was determined.

Various molecular connectivity indices (MCI's) were calculated for the compounds and good correlations were established between MCI's and log K_{ow} and log k'. The present correlations can be used to predict partition coefficients and HPLC retention values for other derivatives.

INTRODUCTION

With the development of industry, sulfur-containing compounds, especially aromatic sulfur-containing compounds, are being used more widely and extensively. These compounds are primarily used as intermediates in the manufacture of pesticides, herbicides and anthelmintics.[1] The n-octanol/water partition coefficient (K_{ow}) is critical, both in medical chemistry and in environmental research.[2-5]

As the determination of lipophilicity by the traditional shake-flask method has certain limitations, reverse phase, high performance liquid chromatography (RP-HPLC) has increasingly been used as an alternative method for rapidly measuring the hydrophobicity. Most RPLC methods use C_{18} modified stationary phase and methanol-water mixtures as the mobile phase.[6,7,8] However, the relationship between log k' and log K_{ow} is influenced by the mobile phase composition.[9,10]

Many investigations have reported that the log k'_w values, which can be obtained by extrapolation from the linear portion of the plot of log k' against the percentage, by volume, of methanol in the eluents to 100% water, correlate better with their log K_{ow} values than log k' in mixed eluent.[6,8] But, there are also papers showing different results. Yamagami and co-workers,[11-13] in their series studies on the estimation of heterocyclic hydrophobicity parameters, showed that good linear correlations between log K_{ow} and log k' were obtained with eluents containing 50-70% of methanol, and deviations from linearity were observed with water-rich eluents.

In order to gain more information about this problem, studies of the relationship between log k' and log K_{ow}, at different mobile phase compositions for a variety types of compounds, are required.

While interest in aromatic sulfur-containing compounds has been increasing, little has been studied comprehensively except for the investigations of Han et al.[1] and He et al.[14] In this paper, we determined the K_{ow} values by the traditional shake-flask method and examined the relationships between log K_{ow} and log k', measured under various mobile phase compositions and calculated the percentage of organic modifier in the eluent by which the chromatographic partition system became the best model of the n-octanol/water partition system.

We also investigated the quantitative relationships between molecular connectivity indices and log K_{ow} and log k'.

Table 1

Chemical Structures
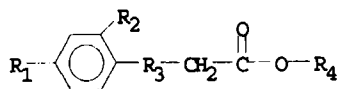
No.	R_1	R_2	R_3	R_4
1a	NO_2	H	S	CH_3
1b	NO_2	H	S	$CH(CH_3)_2$
1c	NO_2	Cl	S	CH_3
1d	NO_2	Cl	S	$CH(CH_3)_2$
1e	H	NO_2	S	CH_3
1f	H	NO_2	S	$CH(CH_3)_2$
1g	Cl	NO_2	S	CH_3
1h	Cl	NO_2	S	$CH(CH_3)_2$
1i	NO_2	NO_2	S	CH_3
1j	NO_2	NO_2	S	$CH(CH_3)_2$
2a	NO_2	H	SO	CH_3
2b	NO_2	H	SO	$CH(CH_3)_2$
2c	NO_2	Cl	SO	CH_3
2d	NO_2	Cl	SO	$CH(CH_3)_2$
2e	H	NO_2	SO	CH_3
2f	H	NO_2	SO	$CH(CH_3)_2$
2g	Cl	NO_2	SO	CH_3
2h	Cl	NO_2	SO	$CH(CH_3)_2$
2i	NO_2	NO_2	SO	CH_3
2j	NO_2	NO_2	SO	$CH(CH_3)_2$
3a	NO_2	H	SO_2	CH_3
3b	NO_2	H	SO_2	$CH(CH_3)_2$
3c	NO_2	Cl	SO_2	CH_3
3d	NO_2	Cl	SO_2	$CH(CH_3)_2$
3e	H	NO_2	SO_2	CH_3
3f	H	NO_2	SO_2	$CH(CH_3)_2$
3g	Cl	NO_2	SO_2	CH_3
3h	Cl	NO_2	SO_2	$CH(CH_3)_2$

MATERIALS AND METHODS

Instruments

The HPLC system (Shimadzu, Kyoto, Japan) consisted of a SCL-8A system monitor, a LC-8A pump, a C-R4A integrator and a SPD-6AV ultraviolet spectrophotometer as the detector. A C_{18} reverse phase Nucleosil 7 (Dalian Institute of Chemical Physics, Academic Sinica) (15 cm x 4.6 mm i.d.) column was used.

Materials

The compounds which were studied are shown in Table 1. These compounds, which are phenylthio, phenylsulfinyl and phenylsulfonyl acetates, were synthesized in our laboratory. Their purities were monitored by HPLC to assure high purity. The water used as a mobile phase component was double-distilled. Methanol was analytical grade and was redistilled before use. Sodium nitrate and n-octanol were also analytical grade.

Methods

The log k' values were determined for each compound. The mobile phases were made by mixing methanol with water in the proportions 100:0, 90:10, 85:15, 80:20, 75:25 and 70:30 (v/v). The flow rate was 0.8 mL/min. All measurements were made at least in duplicate. The average reproducibility of each determination was better than 1.0% relative. The capacity factors (k') were determined using $k' = (t_r - t_o)/t_o$, where t_r is the retention time of the compound, and t_o is the void volume or the dead time. Aqueous solution of sodium nitrate was used for the measurement of t_o.[15]

The n-octanol/water partition coefficients were determined by shake-flask method, as described by OECD Guideline for Testing of Chemicals, at 25°C, followed by centrifuging and analysis of the compound in the aqueous phase with a UV spectrophotometer against a water blank.

The molecular connectivity indices (MCI's) were calculated according to Kier and Hall.[16] The calculation was performed with an AST 386 computer.

The regression analyses were performed using the "Statgraphics" program (STSC, Inc.; 1987).

Table 2

Slope & Intercept Values of the Log k' vs OP% Straight Lines Measured in Various Compositions of Methanol

No.	Slope	Log k'_w	Corr. Coeff.
1b	-4.3866	3.5217	0.999
1c	-4.0966	3.3447	0.998
1d	-4.8403	4.1696	0.997
1e	-3.2507	2.2814	0.994
1f	-4.0279	3.1399	0.995
1g	-3.8686	3.1098	0.996
1h	-4.6978	3.9944	0.995
1i	-3.8370	2.8392	0.992
1j	-4.6216	3.6981	0.993
2a	-2.4484	1.1556	0.990
2b	-3.3513	2.1306	0.985
2c	-2.9856	1.8731	0.985
2d	-3.8653	2.8306	0.995
2e	-2.2803	1.0951	0.975
2f	-3.3452	2.2285	0.957
2g	-3.0988	1.9930	0.985
2h	-3.8492	2.8675	0.995
2i	-2.5911	1.3085	0.980
2j	-3.3090	2.1766	0.991
3a	-4.3534	2.5766	0.990
3b	-4.6980	3.1919	0.995
3c	-3.9539	2.5675	0.985
3d	-4.6605	3.3681	0.995
3e	-3.4937	1.8633	0.975
3f	-4.2896	2.8329	0.991
3g	-4.0939	2.7119	0.990
3h	-4.7195	3.4724	0.995

RESULTS AND DISCUSSION

Determination of the Best Composition of Mobile Phase for K_{ow} Estimation

There have been numerous studies on the estimation of log K_{ow} from HPLC log k' data. But, which log k' values, determined under various chromatography conditions, best correlate with the log K_{ow} values. Valko[17]

established a general approach for the estimation of K_{ow} by RP-HPLC. In his work, the following regression equation was used to obtain the best percentage of organic modifier in the eluent for modeling the n-octanol/water partition system:

$$\log K_{ow} = a \text{ 'slope'} + b \log k'_w + c \qquad (1)$$

where 'slope' and log k'_w are, respectively, the slope and intercept value of log k' vs. OP% (percentage of organic phase in the eluent) straight line. a, b and c were constants obtained by a least squares method. The quotient of the values of a and b gave the percentage of organic modifier in the eluent by which the chromatographic partition system became the best model of the n-octanol/water system.

The value of the log k' for each studied compound was correlated with OP% and yielded the values of 'slope', log k'_w and the correlation coeffcients of the compounds. These values are shown in Table 2.

Substituting the obtained values of the 'slope' and log k'_w into Eq. 1, and by applying a least square method, the following equation was obtained:

$$\log K_{ow} = 1.031 \text{ 'slope'} + 1.682 \log k'_w + 1.262 \qquad (2)$$

$$R = 0930; \, SE = 0.196; \, F = 174.62; \, n = 27$$

The relationship described by Eq. 2 is statistically significant at the level better than 0.01: $F = 174.62 > F(2,24,0.01) = 5.16$.

The quotient of the regression coefficients of the 'slope' and the log k'_w values gives 0.613. This means that the chromatographic partition system containing 61% methanol and 39% water mixture, with an RP-C_{18} column shows the greatest similarity to the n-octanol/water partition system for these aromatic sulfur-containing compounds. The observed and calculated log K_{ow} values, according to Eq. 2, are listed in Table 3. It is clear that K_{ow} values of the other derivatives of aromatic sulfur-containing compounds can be estimated by RP-HPLC at a methanol level of 61 % and RP-C_{18} column.

The observed values of log Kow were also correlated with log k'_w and the following equation was obtained:

$$\log K_{ow} = 0.817 \log k'_w - 0.354 \qquad (3)$$

$$R = 0.815; \, SE = 0.325; \, F = 110.53; \, n = 27$$

Table 3

Observed and Calculated Log K_{ow} Values for the Studied Compounds

Compound	Log K_{ow}				
	Exptl[a]	Calcd[b]	Diff[b]	Calcd[c]	Diff[c]
1b	2.79	2.66	0.13	2.81	-0.02
1c	2.26	2.66	-0.40	2.10	0.16
1d	3.20	3.29	-0.09	3.00	0.20
1e	1.88	1.75	0.13	1.80	0.08
1f	2.76	2.39	0.37	2.70	0.06
1g	2.24	2.50	-0.26	2.42	-0.18
1h	3.00	3.14	-0.14	3.32	-0.32
1i	2.05	2.08	-0.03	1.80	0.25
1j	2.81	2.72	0.09	2.69	0.12
2a	0.80	0.68	0.12	0.78	0.02
2b	1.64	1.39	0.25	1.68	-0.04
2c	1.35	1.33	0.02	1.12	0.23
2d	2.01	2.04	-0.03	2.01	0.00
2e	0.74	0.75	-0.01	0.68	0.06
2f	1.59	1.56	0.03	1.57	0.02
2g	1.29	1.42	-0.13	1.30	-0.01
2h	2.22	2.12	0.10	2.19	0.03
2i	0.63	0.79	-0.16	0.68	-0.05
2j	1.61	1.51	0.10	1.57	0.04
3a	0.78	1.11	-0.33	0.94	-0.16
3b	1.80	1.79	0.01	1.83	-0.03
3c	1.45	1.50	-0.05	1.48	-0.03
3d	2.40	2.12	0.28	2.37	0.03
3e	0.78	0.79	-0.01	0.92	-0.14
3f	1.72	1.60	0.12	1.81	-0.09
3g	1.30	1.60	-0.30	1.54	-0.24
3h	2.45	2.24	0.21	2.44	0.01

[a] Determined by shake-flask method. [b] Calculated according to Eq. 2 (RP-HPLC method). [c] Calculated according to Eq. 4 (MCI's method).

Comparing the R values of Eq. 2 and 3, it is obvious that log k'_w is less powerful for the estimation of log K_{ow} values of these studied compounds. Valko[18] also suggested that the correlation coefficients for the variable of 'slope' and log k'_w was a quantitative measure of structural similarities with regard to

the compounds' behavior in a RPLC system. The intercorrelation of the two variables in Eq. 2 is 0.94; this suggests that the studied compounds behave similarly in the RP-HPLC system. This partition similarity is important in Quantitative Structure-Activity Relationship (QSAR) investigations. It means that, although the degree of oxidation of sulfur atoms in phenylthio, phenylsulfinyl and phenylsulfonyl are different; these three sets of compounds will show similar partition behavior in the biological system.

Correlation Between MCI's and Log K_{ow} and Log k'

Molecular connectivity is a method of deriving topological indices from the hydrogen suppressed skeleton of a molecule. A detailed calculation method is provided by Kier and Hall;[19] only a brief description of the procedure used to calculate MCI's will be presented here. To calculate MCI's of a molecule, first, each non-hydrogen atom is assigned a delta value. For simple indices, each nonhydrogen atom is assigned a delta value (δ) which is equal to the number of atoms to which it is bonded. Valence indices are obtained by assigning atom δ values based on the number of valence electrons not involved in bonds to hydrogen atoms. Simple and valence indices of different orders and types can be calculated for a given molecule. The order refers to the number of bonds in the skeletal substructure or fragment used in computing the index: zero order defines individual atoms, first order uses individual bond lengths, second order uses two adjacent bond combinations, and so on. The type refers to the structural fragment (path, cluster, path/cluster or chain) used in computing the index. Only path indices are possible for orders less than 3. The symbol $^1\chi^v$ represents a first order valence index while $^2\chi$ represents a simple second order index. Each index is computed by an algorithm introduced by Randic[20] which sums the reciprocal square roots of the assigned δ values over all molecular fragments as illustrated below for zero, first, and second order indices:

$$\chi = \sum (\delta_i)^{-1/2} \text{ for all atoms}$$

$$\chi = \sum (\delta_i \delta_j)^{-1/2} \text{ for all bonded pairs of atoms}$$

$$\chi = \sum (\delta_i \delta_j \delta_k)^{-1/2} \text{ for all bonded pairs of atoms}$$

There have been many papers concerning the utility of molecular connectivity indices for describing partition coefficients and HPLC retention parameters.[21-23] Since the MCI's can be calculated quickly and accurately, only based on the structure of a chemical and with no experimental requirements, they are becoming more and more of interest. In molecular connectivity

Table 4

Selected Molecular Connectivity Indices for the Studied Compounds

No.	$^0\chi(R_4)$	$^3\chi_c^v(Ph)$	$^2\chi^v(Ph)$	$^1\chi^v(Ph)$	$^3\chi_c(Ph)$	$^5\chi_{pc}(Ph)$
1b	2.5774	0.2787	2.0824	2.8888	0.3042	0.3715
1c	1.0000	0.4185	2.5843	3.3716	0.4327	0.6487
1d	2.5774	0.4185	2.5843	3.3716	0.4327	0.6487
1e	1.0000	0.2459	2.0325	2.8948	0.2784	0.5629
1f	2.5774	0.2459	2.0325	2,8948	0.2784	0.5629
1g	1.0000	0.4347	2.6123	3.3716	0.4451	0.7399
1h	2.5774	0.4347	2.6123	3.3716	0.4451	0.7399
1i	1.0000	0.3480	2.4290	3.3429	0.4648	0.8083
1j	2.5774	0.3480	2.4290	3.3429	0.4648	0.8083
2a	1.0000	0.2466	2.1480	3.1464	0.2697	0.4081
2b	2.5774	0.2466	2.1480	3.1464	0.2697	0.4081
2c	1.0000	0.3906	2.6574	3.6292	0.4028	0.7502
2d	2.5774	0.3906	2.6574	3.6292	0.4028	0.7502
2e	1.0000	0.2181	2.1056	3.1524	0.2485	0.5727
2f	2.5774	0.2181	2.1056	3.1524	0.2485	0.5727
2g	1.0000	0.4068	2.6854	3.6292	0.4152	0.7534
2h	2.5774	0.4068	2.6854	3.6292	0.4152	0.7534
2i	1.0000	0.3202	2.5021	3.6005	0.4349	0.8219
2j	2.4774	0.3202	2.5021	3.6005	0.4349	0.8219
3a	1.0000	0.2551	2.1756	3.1645	0.3564	0.5052
3b	2.5774	0.2551	2.1756	3.1645	0.3564	0.5052
3c	1.0000	0.4048	2.6949	3.6473	0.4917	0.9403
3d	2.5774	0.4048	2.6949	3.6473	0.4917	0.9403
3e	1.0000	0.2323	2.1430	3.1705	0.3373	0.6833
3f	2.5774	0.2323	2.1430	3.1705	0.3373	0.6833
3g	1.0000	0.4210	2.7229	3.6473	0.5043	08656
3h	2.5774	0.4210	2.7229	3.6473	0.5043	08656

indices calculation, if the studied compounds have a common parent structure, molecular fragment connectivity indices are calculated instead of calculating the whole molecular connectivity indices. Kier and Hall,[19] and Kaliszan[24] had used this sort of partial molecular connectivity indices in their work.

The structures of the compounds investigated in this study are shown in Table 1. An examination of the general structures of the studied compounds suggests that the structure can be divided into two areas. One is the ester alkyl R_4 of CH_3 or $CH(CH_3)_2$, the other is the substitution pattern of R_1, R_2 and R_3 on

the benzene ring. Since the two areas occupy sites at opposite ends of the molecule, it would not be surprising if these two parts exert independent influences on their physicochemical properties and biological activities. For this reason, molecular connectivity indices were calculated for truncated portions of the molecule.

For the ester alkyl group R_4, the simple zero order index $^{\circ}\chi(R_4)$ was calculated. It was enough, by using $^{\circ}\chi(R_4)$ to discriminate CH_3 and $CH(CH_3)_2$. For the substituted phenyl part Ph, $^{\circ}\chi(Ph)$, $^{1}\chi(Ph)$, and simple path indexes $^{2}\chi(Ph)$ through $^{7}\chi_p(Ph)$, simple third-order cluster $^{3}\chi_c(Ph)$, simple fourth- and fifth-order path/cluster indexes $^{4}\chi_{pc}(Ph)$ and $^{5}\chi_{pc}(Ph)$, as well as the corresponding valence indices were calculated. All of the 23 calculated connectivity indexes were stored as variables in the data file.

By stepwise regression, the statistically significant variables were chosen for log K_{ow} and log k'_w. High quality regression equations were obtained. The selected MCI's of the studied compounds are listed in Table 4.

$$\log K_{ow} = 0.568\ (0.036)\ ^{\circ}\chi(R_4) + 22.953\ (1.275)\ ^{3}\chi^{v}_{c}(Ph)$$
$$- 6.878\ (0.479)\ ^{2}\chi^{v}(Ph) + 1.549\ (0.325)\ ^{5}\chi_{pc}(Ph)$$
$$+ 8.698\ (0.639) \tag{4}$$

r = 0 980; s = 0.146; F = 162.90; n = 27

$$\log k'_w = 0\ 547\ (0\ 053)\ ^{\circ}\chi(R4) + 6.868\ (1.079)\ ^{3}\chi^{v}_{c}(Ph)$$
$$- 3.414\ (0.293)\ ^{1}\chi^{v}(Ph) + 7.333\ (1.127)\ ^{3}\chi_{c}(Ph)$$
$$+ 7.995\ (0.740) \tag{5}$$

r = 0.964; s = 0 217; F = 87 07; n = 27

Eq. 4 and 5 are both highly significant at the level better than 0.01: F = 162.90 > F(4,22,0.01) = 4.31; F = 87.07 > F(4, 22, 0.01) = 4.31. The student t value for each coefficient also is very large: 15.897, 18.009, -14.356 and 4.761 for Eq. 4 and 15.897, 18.009, -14.357 and 4.761 for Eq. 5. The observed and calculated log K_{ow} values according to Eq. 4 are also listed in Table 3. Figure 1 shows the correlation between the extrapolated values of log k'_w and the values calculated by Eq. 5.

The results show that calculation of the connectivity indices may provide a method for the prediction of HPLC retention values and MCI's can be used successfully in QSAR and QSRR (Quantitative StructureRetention Relationship).
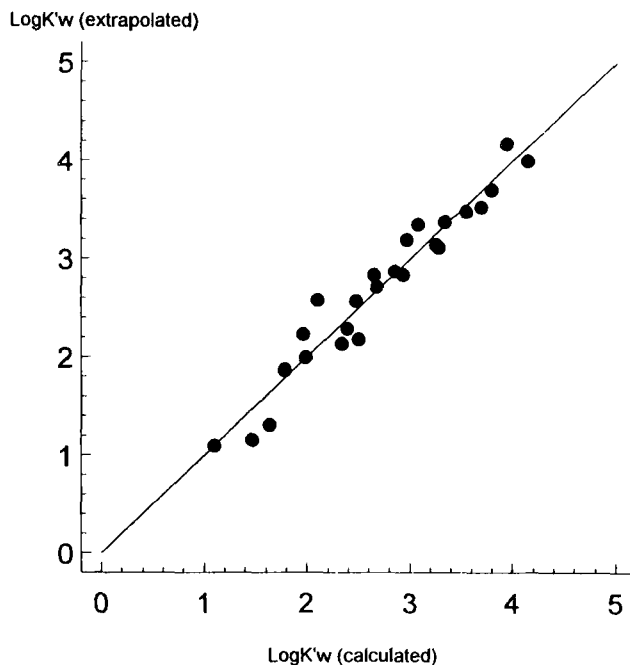
LogK'w (extrapolated)

LogK'w (calculated)

Figure 1. Plot of the extrapolated and the calculated log k'_w values according to Eq. 5.

ACKNOWLEDGEMENT

REFERENCES

1. S. K. Han, L. Q. Jiang, L. S. Wang, Z. Zhang, Chemosphere, **25**, 643-649 (1992).

2. C. Hansch, A. Leo, **Substitutent Constants for Correlation Analysis in Chemistry and Biology**, John Wiley & Sons, New York, 1979.

3. D. Mackay, A. Bobra, W. Y. Shiu, S. H. Yalkowsky, Chemosphere, **9**, 701 711 (1980).

4. G. G. Briggs, J. Agric. Food Chem., **29**, 1050-1059 (1981).

5. W. B. Neely, D. R. Branson, G. E. Blau, Environ. Sci. Technol., **8**, 1113 1115 (1974).

6. Th. Braumann, J. Chromatogr., **373**, 191-225 (1986).

7. C. Yamagami, H. Takami, K. Yamamoto, K. Miyoshi, N. Takao, Chem. Pharm. Bull., **32**, 4994-5002 (1984).

8. D. J. Minick, J. H. Frenz, M. A. Patrick, D. A. Brent, J. Med. Chem., **31**, 1923-1933 (1988).

9. J. G. Dorsey, M. G. Khaledi, J. Chromatogr. **A656**, 485-499 (1993).

10. B. McDuffie, Chemosphere, **10**, 73-83 (1981).

11. C. Yamagami, T. Ogura, N. Takao, J. Chromatogr., **514**, 123-136 (1990).

12. C. Yamagami, N. Takao, Chem. Pharm. Bull., **39**, 2924-2929 (1991).

13. C. Yamagami, N. Takao, Chem. Pharm. Bull., **40**, 925-929 (1992).

14. Y. B. He, L. S. Wang, S. K. Han, Y. H. Zhao, Chemosphere, **30**, 117-125 (1995).

15. C. Horvath, W. Melander, I. Molnar, Anal. Chem., **49**, 142-154 (1977).

16. L. B. Kier, L. H. Hall, **Molecular Connectivity in Chemistry and Drug Research**, Academic Press, New York, 1976.

17. K. Valko, J. Liq. Chromatogr., **7**, 1405-1424 (1984).

18. K. Valko, J. Liq. Chromatogr., **10**, 1663-1686 (1987).

19. L. B. Kier, L. H. Hall, **Molecular Connectivity in Structire-Activity Analysis**, Research Studies Press, John Wiley & Sons, Inc., New York, 1986.

20. M. Randic, J. Amer. Chem. Soc., **17**, 6609-6615 (1975).

21. W. G. Murray, L. H. Hall, J. Pharm. Sci., **64**, 1978-1981 (1975).

22. K. Jinno, K. Kawasaki, Chromatographia, **18**, 90-95 (1984).

23. J. Bojarski, L. Ekiert, Chromatographia, **15**, 172-176 (1982).

24. R. Kaliszan, H. Foks, Chromatographia, **10**, 346-349 (1977).