ELSEVIER

Journal of Chromatography A, 668 (1994) 485–494

JOURNAL OF
CHROMATOGRAPHY A

Reversed-phase retention characteristics of some bioactive heterocyclic compounds

Yassin Darwish, Tibor Cserháti*, Esther Forgács

Central Research Institute for Chemistry, Hungarian Academy of Sciences, P.O. Box 17, 1525 Budapest, Hungary

---

Abstract

The lipophilicities and specific hydrophobic surface areas of sixteen bioactive heterocyclic compounds were determined by reversed-phase thin-layer chromatography using various eluent systems. The data were evaluated by principal component analysis (PCA), separately calculated from the correlation and covariance matrices, and by cluster analysis. The ratio of the variances explained by the first PC component was high and was very similar for both PCA methods, suggesting that the eluents have common elution characteristics. The lipophilicities and specific hydrophobic surface areas of the compounds are well separated on the two-dimensional non-linear maps of PC variables, indicating the different information contents of the two physico-chemical parameters. The information contents of cluster analysis and two-dimensional non-linear mapping techniques were found to be similar but not identical.

---

1. Introduction

In recent years, quantitative structure–activity relationship (QSAR) studies have found growing acceptance and application in the design of new bioactive compounds [1,2]. Many molecular parameters have been correlated with the biological activity, and many of them can be readily determined by various chromatographic techniques [3]. Chromatographic methods have some advantages, *e.g.*, they are simple and rapid and require minute amounts of substances which need not necessarily be very pure.

Lipophilicity is a frequently used molecular parameter in QSAR studies [4]. Reversed-phase thin-layer chromatography (RP-TLC) has been extensively applied to determine lipophilicity [5,6]. However, in RP-TLC the apparent lipophilicity (R_M value) depends on the chromatographic conditions, mainly on the organic phase concentration in the eluent [7–9]. The support particles partially retain their original adsorptive characteristics even after coating [10], and the R_M value changes with the amount [11] and quality [12,13] of the coating substance. When the compound contains one or more dissociable polar substituents, the pH of the eluent [14,15] and the salt concentration [16–18] also modify the lipophilicity. It has been established that not only the R_M value extrapolated to zero organic phase concentration (R_{M0}) but also the slope (b) of the plot between the R_M values and the organic phase concentration of the eluent (C) characterizes the molecular lipophilicity [19,20]. For a homologous series of compounds, the slope and the R_{M0} value exhibit a significant linear correlation [20], but for a non-homologous

---

* Corresponding author.

0021-9673/94/$07.00
SSDI 0021-9673(93)E1037-Z

series of solutes both parameters are needed to describe the lipophilicity accurately [21]. The slope has been regarded as a characteristic of the specific hydrophobic surface area of the compounds [22].

The application of computer-assisted multivariate mathematical–statistical methods makes possible the simultaneous evaluation of an almost unlimited number of variables (chromatographic parameters), which greatly facilitates the solution of theoretical and practical problems. Principal component analysis (PCA) has been applied in chromatography to identify basic factors that influence solute–solvent interactions and to classify solutes and solvents into groups having similar characteristics [23]. The advantages of PCA are that it clusters the variables according to their relationship (clustering chromatographic systems or solutes on the basis of their retention behaviour) and the clusters can be easily visualized by a non-linear mapping technique [24]. In addition, PCA offers the possibility of the extraction of one or more background variables having a concrete physicochemical meaning for the theory and practice of chromatography and it decreases the number of variables to the minimum number required for the solution of a problem [25]. PCA is suitable not only for the calculation of two–two variable relationship (PC variables), but also for the study of all variables of a linear correlation system. PCA can be carried out on both the correlation and covariance matrices of the original data set but the results can sometimes be different [26].

Recently, cluster analysis has found broad application in various fields [27]. In chemistry, cluster analysis has been used for the classification of compounds based on their properties. Cluster analysis belongs to the class of procedures that seek to separate the component data into groups by uncovering the interrelationships that exist among them [28,29].

The objectives of this study were to determine the lipophilicities and specific hydrophobic surface areas of some bioactive heterocyclic compounds for future QSAR studies by reversed-phase chromatography, to compare the physicochemical parameters determined in various chromatographic systems, to evaluate the relationships between retention characteristics by various multivariate mathematical-statistical methods such as PCA, two-dimensional non-linear mapping and cluster analysis and to assess the similarities and dissimilarities between the information content of the methods.

2. Experimental

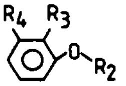
The structures of the compounds are compiled in Table 1. The imines (group A) were the starting materials for further synthetic work and the phenoxy (group B) and benzothiazole derivatives (group C) were synthesized as promising insecticides and antibacterial agents, respectively. The compounds were dissolved in acetone (0.5 mg/ml) and 2 μl of each solution were spotted on Silcoplate F-254 plates (Reanal Fine Chemicals, Budapest, Hungary) impregnated by overnight predevelopment with *n*-hexane–paraffin oil (95:5, v/v). The plates were developed in a sandwich chamber at room temperature using different organic modifiers at various concentrations, and neutral, salt-containing, basic and acidic aqueous phases. The following solvent systems were applied: I, methanol–water; II, methanol–aqueous NaCl; III, methanol–aqueous CH_3COONa; IV, methanol–aqueous CH_3COOH; V, ethanol–water; VI, ethanol–aqueous NaCl; VII, methanol–aqueous CH_3COONa; and VIII, methanol–aqueous CH_3COOH [methanol and ethanol were used in the concentration range 30–50% (v/v) in steps of 5%, and the end concentration of additives in the eluent was 0.25 *M* in each instance]; IX, acetone; X, tetrahydrofuran and XI, dioxane [20–40% (v/v) in steps of 5%]; XII, 1-propanol; and XIII, 2-propanol [10–20% (v/v) in steps of 2.5%]; XIV, acetonitrile [15–35% (v/v) in steps of 5%]; and XV, glycerol [80–95% (v/v) in steps of 2.5%].

After development, the plates were dried at 105°C and the spots were detected under a UV lamp and with iodine vapour. Four independent parallel determinations were carried out in each

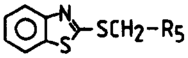
Table 1
Structures of heterocyclic compounds studied

$Cl_2-CHCO-R_1$

A

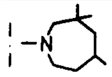
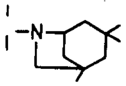
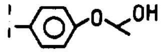
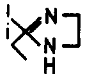
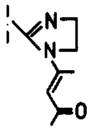
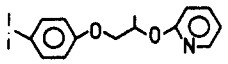
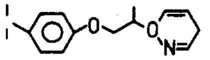
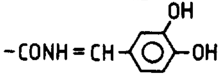
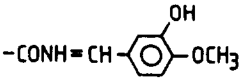
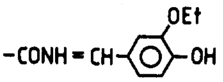
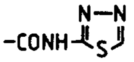
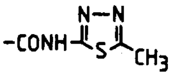
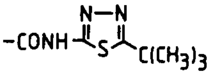


B



C

General formula

Compound No.	General formula	R_1	R_2	R_3	R_4	R_5
1	A		–	–	–	–
2			–	–	–	–
3	B	–		CH_3	CH_3	–
4		–		CH_3	CH_3	–
5		–		CH_3	CH_3	–
6		–		H	H	–
7		–		H	H	–
8	C	–	–	–	–	$-COOC_2H_5$
9		–	–	–	–	$-CONHNH_2$
10		–	–	–	–	
11		–	–	–	–	
12		–	–	–	–	
13		–	–	–	–	
14		–	–	–	–	
15		–	–	–	–	
16		–	–	–	–	

The heterocyclic compounds were synthesized by Dr. Y. Darwish (Mansoura University, Mansoura, Egypt).

instance. The data were omitted from the calculations when the spot of the solute remained at the start or was very near to the front or the relative standard deviation of the four parallel determinations was above 8%. The R_M values of the compounds were determined according to the equation

$$R_M = \log(1/R_F - 1) \quad (1)$$

To increase the accuracy of the lipophilicity determination and to determine the specific hydrophobic surface area, linear correlations were calculated between the R_M values of the compounds and the concentration of organic solvent (C) in the eluent:

$$R_M = R_{M0} + bC \quad (2)$$

where R_M is the actual R_M value of a compound determined at a concentration C of organic modifier in the eluent, R_{M0} (intercept) is the R_M value of a compound extrapolated to zero organic phase concentration in the eluent and b (slope) is the lipophilicity change of a compound caused by unit concentration change of the organic phase. The R_{M0} and b values in eqn. 2 were considered to be the best indicator of the lipophilicity and the specific hydrophobic surface area of the compounds, respectively.

The calculation was carried out separately for each compound and for each eluent system.

To compare the retention behaviour of the compounds, the data (parameters of eqn. 2) were evaluated by multivariate mathematical–statistical methods, as follows.

2.1. *Principal component analysis*

PCA has been used to assess the similarities and dissimilarities both between the RP-TLC systems and solutes. PCA was carried out on the retention data matrix consisting of the lipophilicity and specific hydrophobic surface area values of solutes. The RP-TLC systems were the variables and the solutes were the observations. To elucidate the effect of normalization required for the calculation of the correlation matrix on the results of PCA, PCA was carried both on the correlation and on the covariance matrix. The limit of the variance explained was set to 99.9% in both instances.

2.2. *Non-linear mapping technique*

For easier visualization of the distribution of RP-TLC systems and solutes, the two-dimensional non-linear maps of PC loadings and variables obtained by both PCAs were calculated. The iteration was carried out to the point where the difference between the two last iterations was lower than 10^{-8}.

2.3. *Cluster analysis*

Cluster analysis has been used to elucidate the similarities between RP-TLC systems. To study the effect of PCA on the cluster formation, calculations were carried out on the original data matrix and on the PC variables calculated both from the correlation and the covariance matrices.

2.4. *Comparison of the various multivariate methods*

To compare the information content of the two PCA methods, linear correlations were calculated between the corresponding coordinates of the two-dimensional non-linear maps. Dependent variables were always those calculated from the covariance matrix:

PC loadings:

$$Y_{\mathrm{cov1}} = a + bX_{\mathrm{corr1}} \quad (3)$$

$$Y_{\mathrm{cov2}} = a + bX_{\mathrm{corr2}} \quad (4)$$

PC variables:

$$Y_{\mathrm{cov3}} = a + bX_{\mathrm{corr3}} \quad (5)$$

$$Y_{\mathrm{cov4}} = a + bX_{\mathrm{corr4}} \quad (6)$$

where X and Y are the coordinates of the corresponding non-linear maps of PC loadings and variables calculated from the correlation and covariance matrices, respectively.

Cluster analysis and non-linear mapping techniques are theoretically similar, and therefore the distances between the units (in this instance RP-TLC systems) have to be similar. To verify the validity of the hypothesis outlined above, linear correlations were calculated between the corresponding distances on the cluster dendograms (the dependent variable being in both instances the distances on the cluster dendogram calculated from the original data matrix). Linear correlations were also calculated between the distances on the cluster dendogram (dependent variables) of the original data matrix and the distances on both two-dimensional non-linear maps. To facilitate calculations only the neighbouring distances were included.

3. Results and discussion

Each compound showed normal retention behaviour, that is, the retention decreased monotonously with increasing concentration of organic modifier and no anomalous retention behaviour was observed to make questionable the application of eqn. 2. The relationships between the R_M values and the concentration of organic phase in the eluent were in each instance highly significant, as shown on some examples in Table 2. The r values indicate that the change in the organic solvent concentration explains 82–99% of the change in compound retention, proving again the applicability of eqn. 2. It was found that the slope values for glycerol–water eluents are the lowest, a finding in good accordance with the solvent strength theory that a less lipophilic solvent exerts a weaker effect on retention. The high R_{M0} values indicate that the compounds are hydrophobic and they show very low mobility in water as eluent, that is, the use of a relatively high concentration of organic modifier is needed for their effective reversed-phase separation. The R_{M0} values differ slightly according to the character of the organic modifier, although they have to be identical as theoretical values extrapolated to zero organic phase concentration. This discrepancy can be tentatively explained by the inherent low reproducibility of RP-TLC and the possible non-linearity of extrapolation. However, R_{M0} values of similar magnitude have been successfully used in previous QSAR studies and we strongly believe that the discrepancies discussed above will not influence their further application.

3.1. Principal component analysis

Results of PCA carried out on the correlation and covariance matrices are summarized in Table 3. The first principal component explains most of the total variance, indicating that the retention characteristics of the fifteen eluent systems can be described by only one background variable. This result suggests that the eluents have common elution characteristics. In other words, a single chromatographic system should be sufficient to obtain most of the information content of all RP-TLC systems. Unfortunately, PCA does not define this system as a concrete physico-chemical unit, but only indicates its mathematical possibility. The ratio of the variance explained for both PCA methods is very similar, that is, the PCAs calculated from the correlation and covariance matrices extract similar amounts of information from the original data matrix. This finding indicates that the normalization required for the calculation of the correlation matrix does not modify markedly the variance explained by the principal components.

3.2. Non-linear mapping technique

The two-dimensional non-linear maps of PC loadings calculated from the correlation and covariance matrices are very similar (Figs. 1 and 2). Except for glycerol, all the eluent systems form one cluster. This finding supports our previous quantitative conclusion that the retention characteristics of glycerol differ considerably from those of the other organic modifiers. The eluents containing salt, acidic and alkaline additives are not separated from the corresponding neutral eluent systems, that is, the pH value and sodium chloride content have negligible effects on the retention behaviour of these heterocyclic derivatives. This finding is surprising because the compounds are alkaline and therefore the pH of the eluent should modify their retention. The

Table 2
Parameters of linear correlations between the R_M value of pesticides and the concentration of organic modifier (C) in the mobile phase

Compound No.	No. of eluent system																	
	I			II			III			IV			IX			XV		
	R_{M0}	$-b \cdot 10^2$	r	R_{M0}	$-b \cdot 10^2$	r	R_{M0}	$-b \cdot 10^2$	r	R_{M0}	$-b \cdot 10^2$	r	R_{M0}	$-b \cdot 10^2$	r	R_{M0}	$-b \cdot 10^2$	r
1	2.17	3.81	0.9844	2.13	3.72	0.9708	2.11	3.72	0.9658	2.36	4.20	0.9761	1.66	2.90	0.9764	3.04	2.24	0.9882
2	2.50	4.14	0.9883	2.35	3.82	0.9708	2.48	4.16	0.9917	2.36	3.86	0.9942	1.81	2.26	0.9598	2.38	1.41	0.9649
3	2.82	4.42	0.9766	2.93	4.70	0.9744	2.71	4.26	0.9695	3.17	5.44	0.9806	2.36	4.30	0.9593	3.21	2.37	0.9873
4	2.12	2.55	0.8511	2.73	3.32	0.9960	1.13	2.24	0.9608	1.12	2.62	0.9559	2.19	3.86	0.9647	2.32	1.45	0.9709
5	2.74	3.83	0.9117	2.26	3.28	0.9367	2.35	3.64	0.9036	1.38	3.20	0.9776	2.45	5.10	0.9729	3.41	2.51	0.9470
6	2.26	4.13	0.9672	2.09	3.62	0.9942	2.33	4.42	0.9516	2.28	4.26	0.9774	2.12	4.00	0.9651	6.47	5.68	0.9623
7	2.94	4.71	0.9878	2.92	4.64	0.9803	3.05	5.00	0.9781	3.06	5.08	0.9882	2.59	5.04	0.9837	3.14	2.19	0.9854
8	2.51	3.89	0.9797	2.51	3.88	0.9969	2.51	3.94	0.9674	2.68	4.36	0.9614	2.24	3.88	0.9598	3.00	2.18	0.9371
9	2.07	3.74	0.9874	1.94	3.50	0.9846	2.30	4.40	0.9698	1.61	3.08	0.9480	1.28	3.02	0.9942	2.23	2.07	0.9810
10	2.57	4.27	0.9931	2.73	4.58	0.9859	2.79	4.84	0.9594	2.87	5.06	0.9615	2.55	5.22	0.9210	2.32	1.75	0.9603
11	2.84	5.13	0.9930	2.69	4.56	0.9674	3.00	5.48	0.9569	3.02	5.50	0.9925	2.11	5.10	0.9835	2.42	2.04	0.9784
12	3.14	5.18	0.9649	2.34	3.34	0.9579	2.70	4.26	0.9933	3.18	5.36	0.9886	2.50	5.56	0.9686	3.56	2.81	0.9960
13	2.84	4.11	0.8209	3.39	5.42	0.9992	3.85	6.52	0.9746	3.71	6.16	0.9955	2.72	5.62	0.9847	2.83	1.72	0.9915
14	2.73	4.39	0.9905	2.38	3.45	0.9915	2.81	4.56	0.9621	2.49	3.86	0.9952	2.19	3.72	0.9955	3.42	2.54	0.9914
15	2.48	3.86	0.9155	2.57	4.04	0.9919	2.24	3.22	0.9603	2.65	3.36	0.9705	2.72	5.52	0.9749	3.03	2.10	0.9686
16	2.53	3.93	0.9914	2.50	3.80	0.9935	2.47	3.84	0.9608	2.91	4.92	0.9740	2.20	3.66	0.9690	2.82	1.95	0.9919

Roman numbers refer to the reversed-phase chromatographic systems in Experimental and arabic numbers refer to solutes in Table 1. Equation: $R_M = R_{M0} + bC$.

Table 3
Similarities and dissimilarities between the retention characteristics of RP-TLC systems

No. of principal component	Variance explained (%)		Total variance explained (%)	
	CORR	COVA	CORR	COVA
1	80.15	84.53	80.15	84.53
2	6.69	5.53	86.84	90.06
3	3.92	3.17	90.77	93.24

Results of principal component analysis using the correlation (CORR) and the covariance matrix (COVA).

elucidation of this phenomenon needs further investigation.

The two-dimensional non-linear maps of PC variables are shown in Figs. 3 and 4. As for the two-dimensional non-linear map of PC loadings, similar conclusions can be drawn from both maps: the hydrophobicity (R_{M0}) and specific hydrophobic surface area (b) of the compounds are well separated from each other, indicating that the two physico-chemical parameters have different information contents and both of them can be included in QSAR studies as separate independent variables. The compounds do not form separate clusters according to their chemical structures, which makes it probable that each substructure has a similar impact on the retention characteristics.

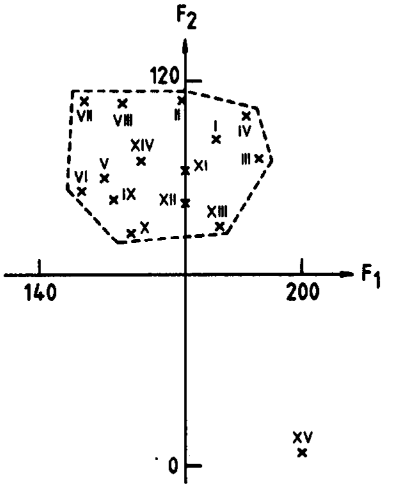


Fig. 1. Two-dimensional non-linear map of PC loadings calculated from the correlation matrix. Number of iterations, 485; maximum error, $2.87 \cdot 10^{-2}$. Numbers refer to eluent systems in Experimental.

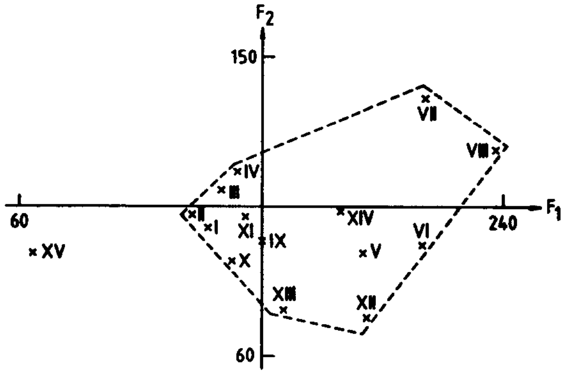


Fig. 2. Two-dimensional non-linear map of PC loadings calculated from the covariance matrix. Number of iterations, 68; maximum error, $1.67 \cdot 10^{-2}$. Numbers refer to eluent systems in Experimental.

3.3. Cluster analysis

The cluster dendogram calculated from the original data matrix is shown in Fig. 5. The distribution of RP-TLC systems represented by distances proves again that the retention characteristics of glycerol (system XV) deviate the most strongly from the retention characteristics of other systems. The dendograms of clusters calcu-

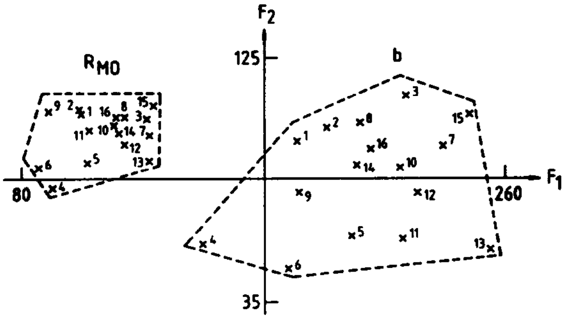


Fig. 3. Two-dimensional non-linear map of PC variables calculated from the correlation matrix. Number of iterations, 288; maximum error, $1.00 \cdot 10^{-2}$. Numbers refer to heterocyclic compounds in Table 1. R_{M0} = lipophilicity; b = specific hydrophobic surface area.

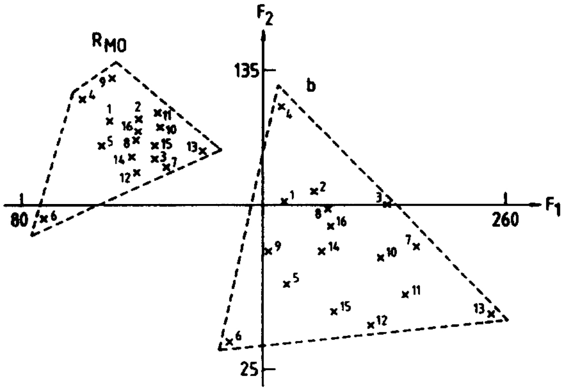


Fig. 4. Two-dimensional non-linear map of PC variables calculated from the covariance matrix. Number of iterations, 109; maximum error, $7.95 \cdot 10^{-3}$. Numbers refer to heterocyclic compounds in Table 1. R_{MO} = lipophilicity; b = specific hydrophobic surface area.

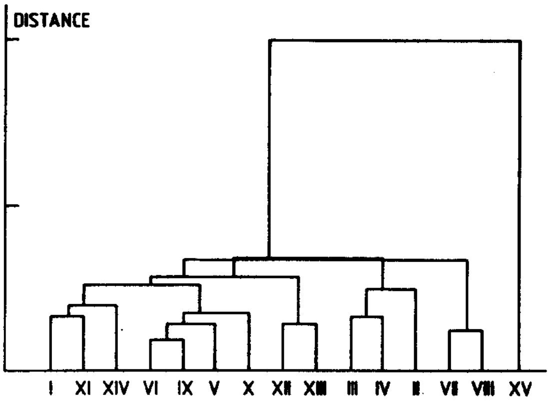


Fig. 6. Cluster dendogram of eluent systems calculated from PCA loadings (correlation matrix). Numbers refer to eluent systems in Experimental.

lated from the PC loadings are shown in Figs. 6 and 7. The clusterings of chromatographic parameters are appreciably different on both dendograms. These results indicate that the use of PCA considerably modifies the distribution of eluent systems, and the results of PCA and cluster analysis may be different. We must stress that this conclusion is based only on experimental data and is not supported by theoretical considerations. It is reasonable to assume that the conclusions are valid only for this data matrix and each generalization may lead to severe misinterpretation.

3.4. Comparison of the various multivariate methods

The parameters of linear correlations between the corresponding coordinates of the two-dimensional non-linear maps of PC loadings and variables calculated from the correlation and covariance matrices are compiled in Table 4. The significant correlations between the coordinates

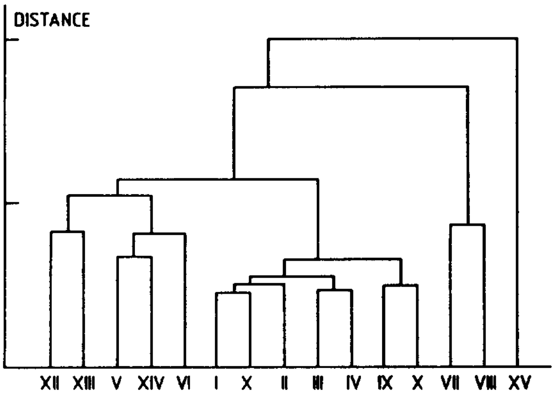


Fig. 5. Cluster dendogram of eluent systems calculated from the original data matrix. Numbers refer to eluent systems in Experimental.

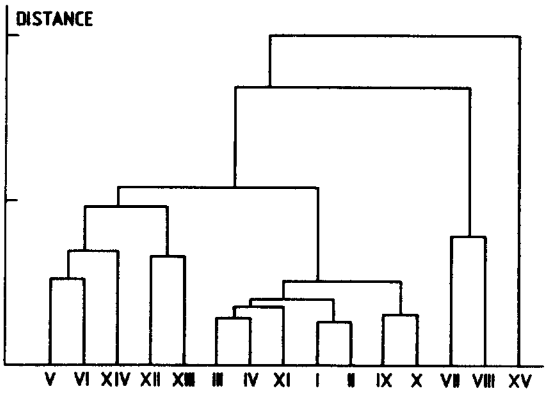


Fig. 7. Cluster dendogram of eluent systems calculated from the PCA loadings (covariance matrix). Numbers refer to eluent systems in Experimental.

Table 4
Comparison of the information content of PCAs carried out on the correlation and on the covariance matrix

Equation No.	Parameter				
	n	a	b	S_b	r
3	15	421.41	−1.63	0.38	0.7682
4	15	51.34	0.42	0.18	0.5401
5	32	−30.89	1.19	0.06	0.9659
6	32	57.46	0.35	0.10	0.5422

Parameters of the linear correlations between the corresponding coordinates of the two-dimensional non-linear maps of PC loadings and variables:

PC loadings:

$$Y_{\mathrm{cov1}} = a + bX_{\mathrm{corr1}} \quad (3)$$

$$Y_{\mathrm{cov2}} = a + bX_{\mathrm{corr2}} \quad (4)$$

PC variables:

$$Y_{\mathrm{cov3}} = a + bX_{\mathrm{corr3}} \quad (5)$$

$$Y_{\mathrm{cov4}} = a + bX_{\mathrm{corr4}} \quad (6)$$

Table 5
Comparison of the information content of non-linear mapping technique and cluster analysis both carried out on the PCA loadings

Equation No.	Parameter				
	n	a	b	S_b	r
7	7	26.88	0.54	0.11	0.9116
8	7	10.55	0.87	0.10	0.9656

Parameters of significant linear correlations[a]:

$$Y = a + bX_1 \quad (7)$$

$$Y = a + bX_2 \quad (8)$$

[a] Y = distances between the nearest neighbouring RP-TLC systems on the cluster dendogram calculated from the original data matrix; X_1 = distances between the same RP-TLC systems on the two-dimensional non-linear map of PC variables calculated from the covariance matrix; X_2 = distances between the same RP-TLC systems on the two-dimensional non-linear map of PC variables calculated from the correlation matrix.

support our previous conclusions drawn from the comparison of the variances explained that in our case the information content of PCAs computed from different matrices is similar. However, the ratio of variance explained is fairly low in the case of the second coordinates, suggesting the existence of small differences between the PCA methods.

No significant linear correlation was found between the distances of RP-TLC systems on the various cluster dendograms. This finding supports our previous qualitative conclusion that the use of PCA modifies the group formation of RP-TLC systems. The parameters of significant linear correlations between the corresponding distances on various maps are compiled in Table 5. The results suggest that the information contents of non-linear mapping and cluster analysis are similar.

It can be concluded that both PCA and CA can be successfully applied for the evaluation of the similarities in the retention behaviour of reversed-phase chromatographic systems and solutes. We assume that PCA followed by two-dimensional non-linear mapping is superior to cluster analysis, because the non-linear mapping technique has a higher dimensionality than cluster analysis.

4. Acknowledgements

Y.D. expresses his gratitude to the Ministry of Culture and Public Education of Hungary for supporting his work. This work was supported by grant OTKA T6436 from the Hungarian Academy of Sciences.

5. References

[1] C. Hansch, in C.J. Cavallito (Editor), *Structure–Activity Relationships*, Pergamon Press, Oxford, 1973, p. 57.

[2] J. Andrew, W.E. Stuper, P. Brugger and S. Jurs, *Computer Assisted Studies of Chemical Structure and Biological Function*, Wiley, New York, 1979.

[3] R. Kaliszan, *Quantitative Structure–Chromatographic Retention Relationships*, Wiley, New York, 1987.

[4] R. Franke, in J.K. Seydel (Editor), *QSAR and Strategies in the Design of Bioactive Compounds*, VCH, Weinheim, 1985, p. 59.

[5] C.B.C. Boyce and B.B. Milborrow, *Nature*, 208 (1965) 537.

[6] G.L. Biagi, A.M. Barbaro and M.C. Guerra, *J. Chromatogr.*, 41 (1969) 371.

[7] P. Jandera, J. Churacek and H. Colin, *J. Chromatogr.*, 214 (1981) 35.

[8] I.D. Wilson, C.R. Bielby and E.D. Morgan, *J. Chromatogr.*, 238 (1982) 97.

[9] E. Janos and T. Cserhati, *Acta Phytopathol. Acad. Sci. Hung.*, 17 (1982) 343.

[10] W.F. Giesen and L.H.M. Janssen, *J. Chromatogr.*, 237 (1982) 199.

[11] T. Cserhati, Y.M. Darwish and Gy. Matolcsy, *J. Chromatogr.*, 270 (1983) 97.

[12] M.C. Guerra, A.M. Barbaro, G. Cantelliforti, M.T. Folfani, G.L. Biagi and P.A. Borea, *J. Chromatogr.*, 216 (1981) 93.

[13] L. Ogierman and A. Siloviecki, *J. High. Resolut. Chromatogr. Chromatogr. Commun.*, 4 (1981) 357.

[14] B. Rittich, M. Polster and O. Kralik, *J. Chromatogr.*, 197 (1980) 43.

[15] Gy. Vigh, J. Varga-Puhony, J. Hlavay and E. Pap, *J. Chromatogr.*, 236 (1982) 51.

[16] T. Cserhati, Y.M. Darwish and Gy. Matolcsy, *J. Chromatogr.*, 241 (1982) 107.

[17] E. Pap and Gy. Vigh, *J. Chromatogr.*, 259 (1983) 49.

[18] T. Cserhati, B. Bordas, E. Fenyvesi and F. Szejtli, *J. Chromatogr.*, 259 (1983) 107.

[19] T. Cserhati, *Chromatographia*, 18 (1984) 18.

[20] T. Cserhati, *Chromatographia*, 18 (1984) 318.

[21] K. Valko, *J. Liq. Chromatogr.*, 7 (1984) 1405.

[22] C. Horvath, W. Melander and I. Molnar, *J. Chromatogr.*, 125 (1976) 129.

[23] T. Cserhati and H.E. Hauck, *J. Chromatogr.*, 514 (1990) 45.

[24] J.W. Sammon, Jr., *IEEE Trans. Comput.*, C18 (1969) 401.

[25] K.V. Mardia, J.T. Kent and J.M. Bibby, *Multivariate Analysis*, Academic Press, London, 1979, p. 213.

[26] T. Cserhati and Z. Illes, *J. Pharm. Biomed. Anal.*, 9 (1991) 685.

[27] W.R. Dillon, *Multivariate Analysis*, Wiley, New York, 1984.

[28] P. Willett, *Similarity and Clustering in Chemical Information Systems*, Research Studies Press, New York, 1987.

[29] V.S. Rose, I.F. Croall and J.H. MacFie, *Quant. Struct.-Act. Relat.*, 10 (1991) 350.