

Modification of Nonlinear Mapping Technique for Quantitative Structure–Retention Relationship Studies

Tibor Cserhádi,^{a,} Esther Forgács,^a Zdenek Deyl,^b
Ivan Miksik,^b and Adam Eckhardt^b*

^a*Institute of Chemistry, Chemical Research Centre, Hungarian Academy of Sciences, P.O.Box 17, 1525 Budapest, Hungary*

^b*Institute of Physiology, Czech Academy of Sciences, Prague, Czech Republic*

Received October 30, 2000; revised February 21, 2001; accepted June 5, 2001

The retention of 19 solutes (Ala, Gly, Lys, Phe and their homopeptides) has been determined on impregnated alumina layers using various mixtures of water-methanol as mobile phases. Lipophilicity and specific hydrophobic surface area of solutes were estimated from linear correlations between the retention of solutes and the concentration of methanol in the mobile phase. The relationship between the hydrophobic retention parameters and physicochemical parameters has been elucidated by the principal component (PC) analysis followed by nonlinear mapping carried out on both the original and absolute values of PC loadings and variables. Ala, Gly, Lys and their homopeptides showed anomalous behavior, the retention increasing with increasing the concentration of methanol in the mobile phase. The use of absolute values in the calculations enhanced the reliability of the mapping technique. It was found that other than hydrophobic forces are involved in the solute – stationary phase interaction (mixed retention mechanism).

Key words: structure–retention relationship, principal component analysis, nonlinear mapping technique, amino acids and homopeptides.

* Author to whom correspondence should be addressed. (E-mail: forgacs@cric.chemres.hn)

INTRODUCTION

Principal component analysis (PCA), a versatile and easy-to-use multivariate mathematical-statistical method, has been developed for the extraction of maximal information from large data matrices containing numerous columns and rows.¹ PCA enables elucidation of the relationship between the columns and rows of any data matrix without being one of the dependent variables. It can be employed for characterization of any data matrices, whereby the heterogeneity of the matrix does not influence the calculation process.^{2,3} PCA has been frequently used in many fields of up-to date research. Thus, PCA has been employed in quantitative structure-activity relationship (QSAR) studies⁴ for exploration of molecular structure-property relationships,⁵ for the evaluation of molecular lipophilicity,^{6,7} for theoretical organic chemistry,⁸ for quantitative structure-retention studies in chromatography,⁹ for elucidation of structure-biodegradation relationships,¹⁰ for clustering of amino acids,¹¹ for assessment of solvent properties,¹² *etc.* As the resulting matrices of PC loadings and variables are generally multidimensional, they cannot be evaluated by visual methods. A nonlinear mapping technique (NLMAP) has been developed for the dimensionality reduction of such matrices.¹³ Two-dimensional nonlinear mapping projects the points of PCA loadings or variables to a plane in such a manner that the distances between the points may be similar on the plane and in multidimensional space. The theoretical background and the mathematical model underlying the technique have been described in detail in Ref. 13. NLMAP takes into account the positive or negative signs of the relationships by calculating the corresponding maps. The result of this mode of calculation is that the points showing strong negative correlation are far from each other on the map, similarly to the points that are not intercorrelated. Consequently, the similarities or dissimilarities between the members of the map cannot be assessed without preliminary knowledge of the character of relationships between the matrix elements when both positive and negative correlations occur. Use of absolute values of the coefficients of correlation in the matrices of PC loadings and variables overcomes this difficulty.¹⁴

Because of their biological activity and importance in many biochemical and biophysical processes,^{15,16} much effort has been devoted to the separation and quantitative determination of peptides in various matrices.¹⁷ It has been observed many times that peptides do not follow the general rule in reversed-phase liquid chromatography, their retention does not decrease monotonously with increasing concentration of the organic modifier in the mobile phase.^{18,19} This phenomenon has been tentatively explained by the so-called silanophilic effect: the free silanol groups not covered by the hydrophobic ligand are available for peptides in the mobile phase. The binding of peptides to adsorption centers results in anomalous retention behavior.^{20,21}

The objectives of the study were to determine the retention behavior of some homopeptides on impregnated alumina stationary phase, to elucidate the quantitative relationship between the reversed-phase retention parameters and physicochemical characteristics of peptides, and to compare the NLMAP results using the original data and their absolute values. The use of impregnated alumina layers was motivated by the finding that thin-layer chromatography (TLC) can be employed as a pilot method for high-performance liquid chromatography^{22,23} and reversed-phase alumina-based stationary phases have found application in the up-to-date HPLC practice.²⁴

EXPERIMENTAL

DC-Aluminium oxide F₂₅₄ plates (Merck, Darmstadt, Germany) were impregnated by overnight predevelopment in n-hexane/paraffin oil (volume ratio 95 : 5) as previously described.²⁵ Amino acids: Ala, Gly, Lys, Phe and their homopeptides (Ala₂, Ala₃, Ala₄, Ala₅, Gly₂, Gly₃, Gly₄, Gly₅, Lys₂, Lys₃, Lys₄, Lys₅, Phe₂, Phe₃, Phe₄) were purchased from Sigma Chemical (St. Louis, MO) and used as received. Analytes were dissolved in the mobile phases at a concentration of 1 mg/mL and 5 μ L of solutions were separately spotted into the plates. Methanol-water mixtures were used as mobile phases, the methanol volume fractions (ϕ) varying between 10 and 90% in steps of 5%. The employment of this wide range of methanol concentration was motivated by the very different retention of peptides on impregnated alumina. Developments were carried out in sandwich chambers (22 \times 22 \times 3 cm) at ambient temperature, the distance of development being about 16 cm. After development, the plates were dried at 105 °C and the spots of analytes were revealed by the ninhydrin reagent. In order to increase the sensitivity of detection, the plates were sprayed with 2 M acetic acid prior to the ninhydrin reaction. Each experiment was run in quadruplicate. The R_M value characterizing the molecular hydrophobicity in reversed-phase thin-layer chromatography was calculated for each solute in each eluent:

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

When the coefficient of variation of parallel determinations was higher than 5%, the R_M value was omitted from the following calculations. In order to determine the hydrophobicity parameters of the solutes, linear relationships were calculated between the R_M values and the concentration of methanol in the mobile phase:

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

where R_M = the value for a solute determined at the given methanol concentration; R_{M0} = R_M value extrapolated to zero methanol concentration (best estimate of molecular lipophilicity); b = decrease in the R_M value caused by a 1% increase in methanol concentration in the mobile phase (related to the specific hydrophobic surface area of the solutes);²⁶ c = methanol concentration in the mobile phase. Eq. (2) was applied separately for each solute.

The relationship between the retention parameters and physico-chemical characteristics of solutes was elucidated by PCA. Variables included the R_{M0} and b values of Eq. (2) (variables I and II), number of amino acid units in the peptide molecule (N , variable III), hydrophobicity (z_1 , variable IV), side chain bulk (z_2 , variable V) and electronic properties of amino acids (z_3 , variable VI). The z parameters are the results of multivariate parametrization including seven TLC, 3 NMR and two theoretical variables. They characterize and summarize in one number the various aspects of the hydrophobicity, side chain bulk and electronic parameters of amino acids without being identical with any concrete variables. Combined variables (z values multiplied by the number of amino acid units, $z_i \times N$, variables VII-IX) were also included in the calculation (altogether 9 variables). Molecular parameters were taken from Ref. 27. Observations were the solutes (altogether 19 observations). The variance explained by PCA was set to 95% of the total variance. Two-dimensional NLMAPs of the data matrices of PC loadings and components were calculated using the original data and the absolute values of loadings and components. Iteration was carried out to the point where the difference between the last two iterations was lower than 10^{-8} . The varimax rotation of PC loadings around two axes was also carried out and the results of varimax rotation were compared with those of NLMAPs. Inclusion of varimax rotation into the calculation process was motivated by the fact that, similarly to NLMAP, varimax rotation also reduces to two the dimensionality of the original multidimensional matrix.

Softwares for PCA and NLMAP were prepared by Dr. Barna Bordás (Plant Protection Institute, Hungarian Academy of Sciences, Budapest, Hungary).

RESULTS AND DISCUSSION

The relationship between the lipophilicity (R_M) value of some solutes and the concentration of methanol in the mobile phase is demonstrated in Figure 1. Marked differences have been observed between the retention behavior of solutes. Thus, Phe₂ showed regular retention behavior, the mobility increasing with increasing concentration of methanol in the mobile phase. Interestingly, the retention of Phe was not changed with changing the composition of the mobile phase while the retention of Ala₂ increased with higher methanol concentrations. The parameters of Eq. (2) are compiled in Table I. Except for Phe, the R_M value of every other solute depended significantly on the concentration of methanol in the mobile phase. Comparison of the calculated coefficients of correlation with the corresponding tabulated values indicated that the significance level was over 95% in each instance. Both the intercept (R_M) and slope (b) values showed high variations, indicating that the impregnated alumina stationary phase can be successfully employed for the separation of this class of solutes by modifying the composition of the mobile phase. Ala, Gly and Lys and their homopeptides showed anomalous retention behavior; their mobility decreased with increasing the concentration of methanol. This effect is similar to the silanophilic effect; however,

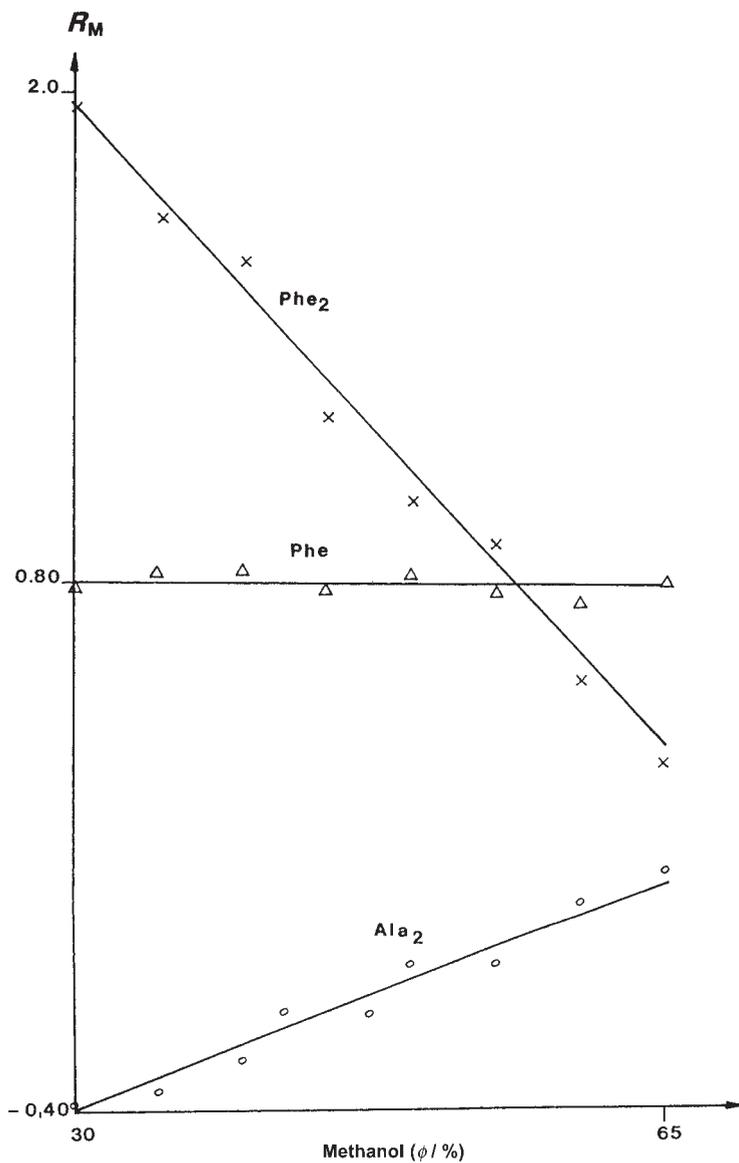


Figure 1. Relationship between the R_M value of Phe, Phe₂ and Ala₂ and the concentration of methanol in the mobile phase (vol. fraction, ϕ).

silanol groups are not present on the surface of impregnated alumina. It can be assumed that the dielectric constant of the mobile phase decreases at higher concentrations of methanol, suppressing the degree of dissociation of the polar substructures of amino acids and peptides. The less dissociated

TABLE I

Parameters of linear relationships between the R_M values of amino acids and homopeptides and the concentration of methanol in the mobile phase^a

Solute		R_{M0}	$b \times 10^2$	r
No.	Name			
1	Ala	-0.36	1.80	0.9977
2	Ala ₂	-0.85	1.50	0.9912
3	Ala ₃	-1.03	1.30	0.9790
4	Ala ₄	-1.85	2.57	0.9675
5	Ala ₅	-2.21	3.37	0.9788
6	Gly	0.07	1.61	0.9993
7	Gly ₂	-0.15	1.44	0.9944
8	Gly ₃	-0.23	1.67	0.9948
9	Gly ₄	-0.31	2.13	0.9762
10	Gly ₅	-0.10	2.09	0.9891
11	Lys	0.06	2.87	0.9857
12	Lys ₂	-0.17	3.02	0.9893
13	Lys ₃	0.19	3.48	0.9022
14	Lys ₄	0.23	4.30	0.9731
15	Lys ₅	-1.01	9.51	0.9647
16	Phe	0.85	0.00	0.1237
17	Phe ₂	3.35	-4.37	0.9750
18	Phe ₃	5.02	-6.80	0.9963
19	Phe ₄	4.73	-8.60	0.9643

^a R_{M0} – value extrapolated to zero methanol concentration; b – decrease in the R_M value caused by a 1% increase in methanol concentration in the mobile phase.

solutes display higher apparent lipophilicity, resulting in an enhanced retention under reversed-phase conditions. The fact that the Phe homopeptides show a regular retention behavior can be tentatively explained by the supposition that the effect of the more hydrophobic ring structures overshadows the effect of the suppressed dissociation of the polar amino and carboxyl groups. The data suggest that the effects discussed above are counterbalanced in Phe resulting in the anomalous retention behaviour observed.

Results of the principal component analysis are compiled in Table II. Four background (hypothetical) variables explain the overwhelming majority of total variance present in the original 9 variables with a loss of 3.96% of information. Unfortunately, PCA does not define these hypothetical variables as concrete physical or physicochemical entities; it only indicates their

TABLE II

Similarities and dissimilarities between the retention parameters and physicochemical characteristics of amino acids and homopeptides. Results of the principal component analysis^a

PC	Eigenvalue	Variance explained		Total variance explained	
		%		%	
1	3.98	44.19		44.19	
2	3.02	33.55		77.74	
3	1.21	13.41		91.16	
4	0.44	4.88		96.04	

Variables		PC			
		1	2	3	4
I	R_{M0}	<u>0.70</u>	<u>0.51</u>	0.33	-0.34
II	b	0.16	<u>0.84</u>	0.25	0.41
III	N	-0.33	<u>0.57</u>	<u>0.74</u>	0.01
IV	z_1	<u>-0.94</u>	-0.07	-0.16	0.04
V	z_2	0.34	<u>0.82</u>	-0.35	-0.15
VI	z_3	<u>0.75</u>	<u>-0.53</u>	0.12	0.32
VII	$z_1 \times N$	<u>-0.98</u>	0.08	0.01	0.04
VIII	$z_2 \times N$	0.39	<u>0.70</u>	<u>-0.57</u>	0.16
IX	$z_3 \times N$	<u>0.82</u>	<u>-0.53</u>	0.11	0.00

^a PC – principal component; R_{M0} – value extrapolated to zero methanol concentration; b – decrease in the R_M value caused by a 1% increase in methanol concentration in the mobile phase; N – number of amino acid units in peptide molecule; z_1 – hydrophobicity; z_2 – side chain bulk; z_3 – electronic properties of amino acids.

mathematical possibility. The high loading of both retention characteristics and physicochemical parameters in the two first PC components indicates that these parameters influence considerably the retention behavior of solutes. The factor loadings rotated around two axes are compiled in Table III. According to the data, the side chain bulk (z_2) and the side chain bulk multiplied by the number of amino acid units in the homopeptide ($z_2 \times N$) exert the strongest impact on the retention while the influence of other physicochemical parameters is of secondary importance.

The two-dimensional nonlinear maps of PC loadings calculated from the original loadings and from their absolute values are shown in Figures 2 and 3, respectively. The points of variables are markedly different on the maps, indicating that the use of absolute values exerts a considerable effect on the

TABLE III

Similarities and dissimilarities between the retention parameters and physicochemical characteristics of amino acids and homopeptides. Principal component loadings rotated around two axes

Variables		First axis	Second axis
I	R_{M0}	0.46	<u>0.73</u>
II	b	-0.17	<u>0.84</u>
III	N	<u>-0.53</u>	0.40
IV	z_1	<u>-0.85</u>	-0.42
V	z_2	0.01	<u>0.89</u>
VI	z_3	<u>0.89</u>	-0.21
VII	$z_1 \times N$	<u>-0.94</u>	-0.29
VIII	$z_2 \times N$	0.10	<u>0.79</u>
IX	$z_3 \times N$	<u>0.96</u>	-0.18

^aFor symbols see Table II.

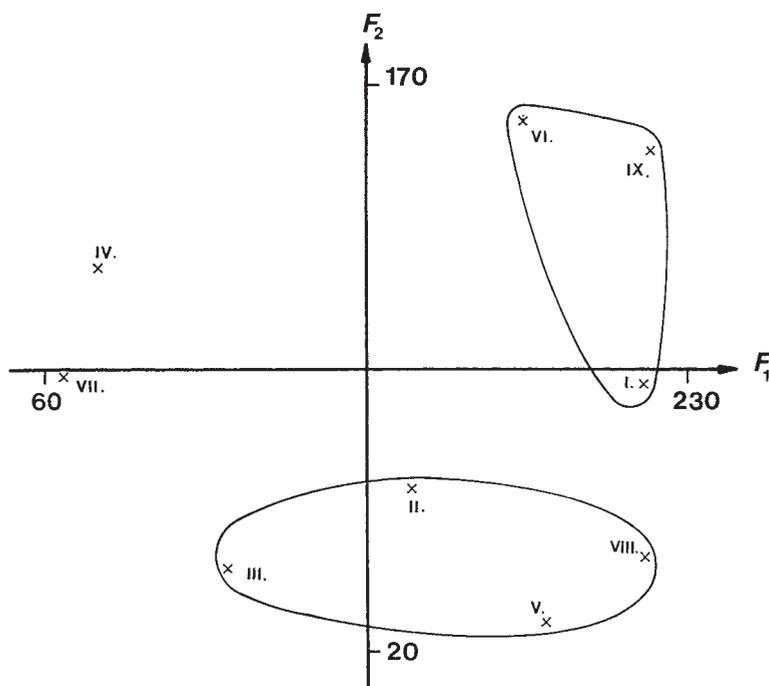


Figure 2. Similarities and dissimilarities between the retention characteristics and physicochemical parameters of solutes. Two-dimensional nonlinear map of the original PC loadings. Number of iterations: 220; maximum error: 2.85×10^{-2} . For symbols see Experimental.

distribution of variables. Points IV (z_1) and VII ($z_1 \times N$) are far removed from the other points on the map constructed from the unmodified loadings (Figure 2). However, according to Table II, they have high loadings in the first PC component together with variables I, VI and IX, indicating their similarity. Because they are represented by negative values, their position on the map is incorrect. Variables I, VI and IX having positive signs are correctly represented. The same holds for the other cluster containing variables with high positive loadings in the second PC component.

The use of absolute values for the calculation of maps overcomes this difficulty (Figure 3) in that the variables correlated are correctly represented. Retention parameters form clear-cut clusters with more than one physicochemical parameter. This result suggests that the retention of these amino acids and homopeptides is not governed only by hydrophobic interactions but that other interactive forces are involved in their binding to the impregnated alumina stationary phase.

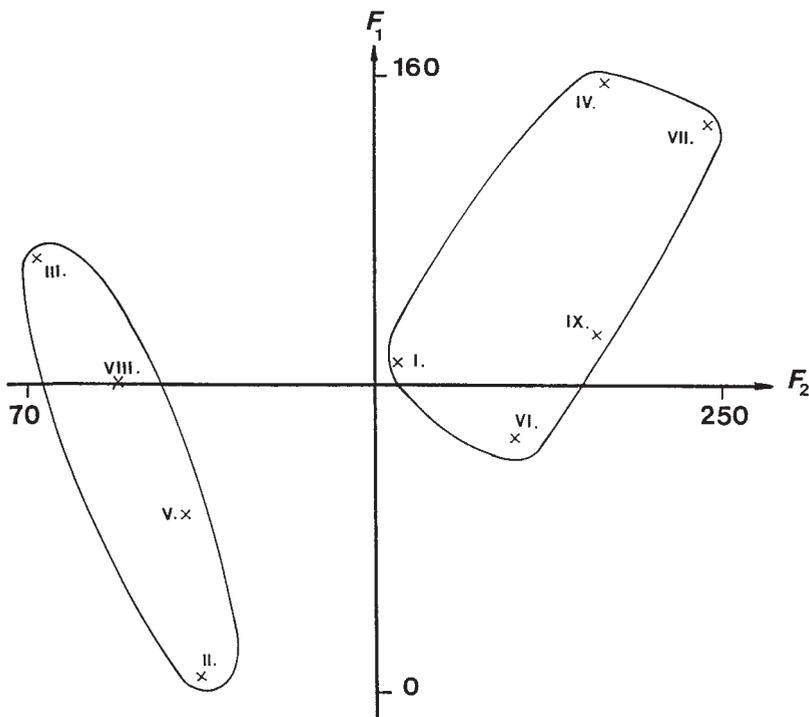


Figure 3. Similarities and dissimilarities between the retention characteristics and physicochemical parameters of solutes. Two-dimensional nonlinear map of the absolute values of PC loadings. Number of iterations: 120; maximum error: 8.08×10^{-3} . For symbols see Experimental.

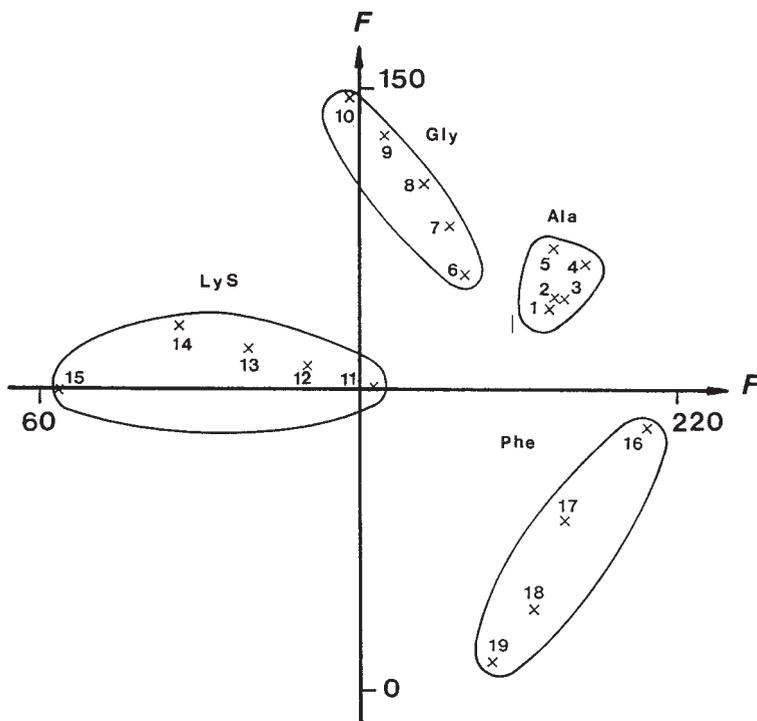


Figure 4. Similarities and dissimilarities between the solutes. Two-dimensional nonlinear map of the original PC variables. Number of iterations: 161; maximum error: 1.22×10^{-2} . For symbols see Table I.

The two-dimensional nonlinear map of the original PC variables is shown in Figure 4. Solute clusters are formed according to the character of the amino acid units and not according to the length of the homopeptide chain. This finding indicates that homopeptides containing different amino acid building blocks can be easily separated while the separation of homopeptides according to the number of amino acid units is difficult in this chromatographic system. The distribution of solutes was similar on the map prepared from the absolute values (data not shown), which means that in that case the application of this mode of calculation was not necessary.

CONCLUSIONS

It can be concluded from the data that Ala, Gly, Lys and their homopeptides show anomalous retention behavior on the impregnated alumina layer and that their mobility decreases with increasing the concentration of

methanol in the mobile phase. Suppression of the dissociation of the polar substructures of solutes by methanol may account for the anomaly. Principal component analysis and the nonlinear mapping technique using the absolute values of PC loading and variables can be successfully employed for the study of the quantitative relationship between retention characteristics and physicochemical parameters.

Acknowledgment. – This work was supported by the Czech-Hungarian cooperation programme »The role of surfactants as pseudophase in MEKC and as eluents in capillary electrochromatography«.

REFERENCES

1. K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic Press, London, 1979, pp. 213–254.
2. T. Cserháti and E. Forgács, *J. Chromatogr. A* **771** (1997) 105–109.
3. T. Cserháti and E. Forgács, *J. Chromatogr. A* **776** (1997) 31–36.
4. M. G. B. Drew, G. H. R. Wilden, W. J. Spillane, R. M. Walsh, C. A. Ryder and J. M. Simmie, *J. Agric. Food Chem.* **46** (1998) 3016–3026.
5. P. G. Seybold, *SQERED* **10** (1999) 101–115.
6. C. Sarbu and S. Todor, *J. Planar Chromatogr. Mod. TLC* **11** (1998) 123–126.
7. R. Mannhold, G. Cruciani, K. Dross, and R. Rekker, *J. Comput. Aided Mol. Design* **12** (1998) 573–581.
8. K. Héberger and A. Lopata, *J. Org. Chem.* **63** (1998) 8646–8653.
9. K. Héberger and M. Görgényi, *J. Chromatogr. A* **845** (1999) 21–31.
10. J. Damborsky, A. Berglund, M. Kutý, A. Ansorgová, I. Nagata, and M. Sjöström, *Quant. Struct. Act. Relat.* **17** (1998) 450–458.
11. A. Zaliani and E. Gancia, *J. Chem. Inf. Comput. Sci.* **39** (1999) 525–533.
12. A. R. Katritzky, T. Tamm, Y. Wang, and M. Karelson, *J. Chem. Inf. Comput. Sci.* **39** (1999) 692–698.
13. J. W. Sammon Jr., *IEEE Trans. Comput.* **C18** (1969) 401–407.
14. E. Forgács, A. Kósa, G. Csiktusnádi-Kiss, T. Cserháti, R. Kaliszan, P. Haber, and A. Nasal, *J. Liq. Chromatogr. Rel. Technol.* **21** (1998) 2523–2534.
15. J. D. Reveille, D. Overbach, R. Goldstein, R. Moreda, R. A. Isern, and F. C. Arnett, *J. Clin. Invest.* **89** (1992) 1208–1213.
16. J. Choppin, *Path. Biol.* **40** (1992) 257–264.
17. C. T. Mant and R. S. Hodges, *High Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation*. CRC Press, Boca Raton, FL, 1991.
18. T. Cserháti, *J. Chromatogr.* **553** (1991) 467–475.
19. T. Cserháti, *J. Chromatogr. A* **660** (1992) 149–156.
20. A. Nahum and C. Horváth, *J. Chromatogr.* **203** (1981) 53–64.
21. K. E. Bij, C. Horváth, W. R. Melander, and A. Nahum, *J. Chromatogr.* **203** (1981) 65–84.
22. J. K. Rozylo and M. Jaroniec, *Fresenius' J. Anal. Chem.* **321** (1985) 371–373.
23. J. K. Rozylo and M. Jaroniec, *J. Liq. Chromatogr.* **14** (1991) 3197–3212.

24. K. Cabdera, D. Lubda, and G. Jung, *Alkaline stable stationary phases for HPLC*, Part 2, Aluspher RP-select B, Kontakte, Darmstadt, 1992, pp. 32–35.
25. T. Cserhádi, B. Bordás, E. Fenyvesi, and J. Szejtli, *J. Chromatogr.* **259** (1983) 107–110.
26. C. Horváth, W. Melander, and I. Molnár, *J. Chromatogr.* **125** (1976) 129–156.
27. J. Jonsson, L. Eriksson, S. Hellberg, M. Sjöström, and S. Wold, *Quant. Struct. Act. Relat.* **8** (1989) 204–209.

SAŽETAK

Prilagodba metode nelinearnog preslikavanja za kvantitativnu analizu odnosa između strukture i retencijskog parametra

*Tibor Cserhádi, Esther Forgács, Zdenek Deyl, Ivan Miksik
i Adam Eckhardt*

Retencija 19 otopljenih tvari (Ala, Gly, Lys, Phe, i njihovi homopeptidi) praćena je na impregniranim slojevima Al_2O_3 , uporabom različitih omjera vode i metanola kao mobilne faze. Lipofilnost i specifična hidrofobna površina otopljenih tvari procijenjeni su na temelju linearnih korelacija između retencije otopljenih tvari i koncentracije metanola u mobilnoj fazi. Relacija između hidrofobnih retencijskih parametara i fizikalno-kemijskih parametara objašnjena je s pomoću metode glavnih komponenata nakon koje je primjenjena metoda nelinearnog preslikavanja koja je provedena i na izvornim i na apsolutnim vrijednostima težinskih faktora glavnih osi i na varijablama. Ala, Gly, Lys i njihovi homopeptidi ponašaju se anomalno i povećavaju retenciju s povećanjem koncentracije metanola u mobilnoj fazi. Uporaba apsolutnih vrijednosti u računima povećava pouzdanost metode preslikavanja. Opaženo je da su i drugi čimbenici osim hidrofobnih sila uključeni u međudjelovanje otopljenih tvari i stacionarne faze (mješoviti retencijski mehanizam).