

This article was downloaded by:

On: 24 January 2011

Access details: *Access Details: Free Access*

Publisher *Taylor & Francis*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Journal of Liquid Chromatography & Related Technologies

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713597273>

### PCA, FOLLOWED BY TWO-DIMENSIONAL NONLINEAR MAPPING AND CLUSTER ANALYSIS, VERSUS MULTILINEAR REGRESSION IN QSRR

Annamaria Jakab<sup>a</sup>; Gábor Schubert<sup>b</sup>; Miklos Prodan<sup>a</sup>; Esther Forgacs<sup>a</sup>

<sup>a</sup> Institute of Chemistry, Chemical Research Center, Hungarian Academy of Sciences, Hungary <sup>b</sup> Institute of Isotope and Surface Chemistry, Chemical Research Center, Hungarian Academy of Sciences, Hungary

Online publication date: 01 November 2002

**To cite this Article** Jakab, Annamaria , Schubert, Gábor , Prodan, Miklos and Forgacs, Esther(2002) 'PCA, FOLLOWED BY TWO-DIMENSIONAL NONLINEAR MAPPING AND CLUSTER ANALYSIS, VERSUS MULTILINEAR REGRESSION IN QSRR', Journal of Liquid Chromatography & Related Technologies, 25: 1, 1 – 16

**To link to this Article:** DOI: 10.1081/JLC-100108535

**URL:** <http://dx.doi.org/10.1081/JLC-100108535>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## PCA, FOLLOWED BY TWO-DIMENSIONAL NONLINEAR MAPPING AND CLUSTER ANALYSIS, VERSUS MULTILINEAR REGRESSION IN QSRR

Annamaria Jakab,<sup>1,\*</sup> Gábor Schubert,<sup>2</sup> Miklos Prodan,<sup>1</sup> and Esther Forgacs<sup>1</sup>

<sup>1</sup>Institute of Chemistry, Chemical Research Center, Hungarian Academy of Sciences, P. O. Box 17, 1525, Hungary

<sup>2</sup>Institute of Isotope and Surface Chemistry, Chemical Research Center, Hungarian Academy of Sciences, P. O. Box 77, 1525, Hungary

### ABSTRACT

Retention parameters of 45 different barbituric acid derivatives were determined on an amide embedded RP silica column (Discovery RP-AmideC16) using non-buffered tetrahydrofuran-water mixtures as eluents. Linear correlations were calculated between the  $\lg k$  values and the tetrahydrofuran concentration in the eluent. Chromatographic parameters of barbituric acid derivatives were correlated with different conventional and quantum chemical structural descriptors in QSRR study (the different parameters were: intercept ( $\lg k_0$ ) and slope

---

\*Corresponding author. E-mail: [janna@chemres.hu](mailto:janna@chemres.hu)

(*b*) values of the linear, the combined retention parameter ( $\lg k_0/b$ ), asymmetry factor (*AF*) and theoretical plate values (*N*). Multilinear regression analysis (SRA) and principal component analysis (PCA), followed by two dimensional nonlinear mapping and cluster analysis techniques, were used to determine the retention behavior of barbituric acid derivatives. Different mathematical-statistical methods indicated that the hydrophobic characteristics of the solutes have marked influence on the retention behavior of barbituric acid derivatives on this amide embedded RP silica column in tetrahydrofuran-water eluent systems. The significant effect of the hydrophobic characteristics of the analytes in the retention behavior indicated that the effects of the interaction between the analytes and the residual silica silanol groups are negligible.

## INTRODUCTION

Multivariate mathematical-statistical techniques have been extensively applied to interpret complex data matrices in different fields, such as characterization of chromatographic systems in QSRR (Quantitative Structure-Retention Relationship) studies, (1–3) correlating the molecular structure with different activity in QSAR (Quantitative Structure-Activity Relationship) studies, (4,5) characterization of catalyzations, (6) analysis of complex spectra in spectroscopy, (7) etc. The modern multivariate mathematical-statistical methods, like stepwise regression analysis (SRA) or principal component analysis (PCA), make possible the simultaneous assessment of a practically unlimited number of variables in a reasonably short computing time.

SRA is a common multilinear regression analysis. In SRA, the presence of independent variables, which exert no significant influence on the dependent variable, lessens the significance level of those independent variables, which do significantly influence the dependent variables. To overcome this difficulty, the SRA automatically eliminates from the selected equation the insignificant independent variables.

PCA has been developed to find the relationships between the variables of any data matrices without any of them being the dependent variable. (8,9) The main advantage of PCA in chromatography are (i) the clustering of variables according to their relationship, for example, clustering chromatographic systems or solutes according to their retention behavior, (ii) the possibility of the extraction of one or more background variables having concrete physico-chemical meaning for the theory and practice of chromatography. PCA in combination with visualization methods, such as two-dimensional nonlinear



mapping (nl mapping)(10) or cluster analysis (11) considerably facilitates the evaluation. Theoretically, cluster analysis and nonlinear mapping techniques are similar: both methods calculate and visualize the relative distances between the members of the data matrix, however, the mode of calculation is different. Two-dimensional nonlinear mapping and cluster analysis can be carried out on both the correlation and the covariance matrix of the original data set.

QSRR is based on the correlation between the chromatographic parameters and the different descriptors of analytes. For understanding the retention mechanisms in HPLC, three main types of methods have been established. In the beginning, (i) correlating the logarithms of the capacity factor extrapolated to 100% water concentration ( $\log k_0$  or  $\lg k_0$ ) with the logarithms of *n*-octanol-water partition coefficients ( $\log P$  or  $\lg P$ ) has been generally applied. (12,13) Further on, (ii) correlating the  $\log k_0$  with different conventional structural descriptors has also been reported. (14,15) Conventional structural descriptors were used to describe the different characteristics of the molecules by Hanch and Leo, (16) Abraham et al., (17) and Corr and Galvez et al. (18)

Recently, (iii) correlating the  $\log k_0$  values with quantum chemical descriptors have been extensively applied. (19,20) In the recent studies, quantum chemical descriptors have been employed alone, or in combination, with the above mentioned descriptors. (21–25)

Columns with embedded amide groups were first introduced by Supelco, who recently produced a version of the original phase based on pure octadecyl coated silica. (26) Internal polar groups, (27) such as amide (28) or carbamate (29) groups can react with the residual silica silanols through electrostatic or hydrogen bonding interaction, resulting in weakening the interaction between the polar analytes and the silanol groups. Such embedded columns have an excellent resolution of polar compound and different elution profiles compared to C18 phases.

The aim of our work was to determine the retention behavior of barbituric acid derivatives on an amide embedded RP silica column in tetrahydrofuran-water eluent systems, to elucidate the relationship between the retention characteristics and physicochemical descriptors using SRA and PCA followed by two-dimensional nonlinear, modified nonlinear mapping and cluster analysis techniques, and to compare these different mathematical-statistical methods in this QSRR study.

## EXPERIMENTAL

RP silica column with embedded amide groups (Discovery RP-AmideC16, 125 × 4.6 mm I.D., particle size 5 μm) was purchased from Supelco (Bellafonte,



U.S.A.). The HPLC system consisted of a Waters LC Module I and Waters 746 integrator (Waters, Milford, MA, U.S.A.). The flow rate was  $0.8 \text{ mL} \cdot \text{min}^{-1}$  and the detection wavelength was set to 240 nm. Mixtures of non-buffered tetrahydrofuran-water were used as eluents, tetrahydrofuran concentration ranged from 35 to 55 v/v% (minimum at 5 different concentrations). The shape of the peaks was sufficient for not buffering the eluent. The experiments were carried out at room temperature (21–23°C). The barbituric acid derivatives (Figure 1) were dissolved in methanol at the concentration of  $0.1 \text{ mg} \cdot \text{mL}^{-1}$ . The retention time, asymmetric factor ( $AF$ ), and the theoretical plate values ( $N$ ) of each compound in each eluent were determined in triplicate. The asymmetry factors ( $AF$ ) were calculated according to  $AF = B/A$ , where  $A$  and  $B$  can be measured by drawing a perpendicular line from the apex of the peak to the baseline, and measure the front ( $A$ ) and back ( $B$ ) widths of the peak at 5% height. The number of the theoretical plates were calculated according to  $N = 16(t_r/w)$  formula, where  $t_r$  is the retention time and  $w$  is the band width measured at the baseline.

Linear correlation was used to describe the dependence of the  $\lg k$  value on the concentration of tetrahydrofuran.

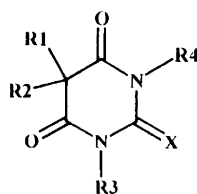
$$\lg k = \lg k_0 + b \cdot C \quad (1)$$

where  $\lg k$  is the logarithm of the capacity factor;  $C$  is the tetrahydrofuran concentration in the eluent (vol%).  $\lg k_0$  and  $b$  are constants to be determined by the least square method. The  $\lg k_0$  is the logarithm of the capacity factor extrapolated to zero concentration of the organic component in the mobile phase (related to molar lipophilicity), (30) and  $b$  is the change of the  $\lg k$  value caused by a unit change (1% vol.) in the organic mobile phase concentration (related to the specific hydrophobic surface area in contract with support). (31)

To find the physico-chemical parameters of the barbituric acid derivatives, which significantly influence their retention parameters ( $\lg k_0$ ,  $b$ ,  $AF$ ,  $N$ ,  $\lg k_0/b$ ), (32) SRA was used. The dependent variables were, separately,  $\lg k_0$ ,  $b$ ,  $\lg k_0/b$  and the average value of  $AF$  and  $N$  of each barbituric acid derivative and the independent variables were the structural descriptors of the barbituric acid derivatives. The acceptance level for the individual independent variables was set to 95% significance level. The list of both the conventional and quantum chemical structural descriptors, considered in this study, is given in Table 1. The quantum chemical descriptors of the barbiturates were calculated using the Gaussian 94 (33) software package.

The calculations were carried out on the triketo forms of the barbiturates in every case. First, geometry optimization was done using the AM1 semi-empirical method. (34) The starting geometries were predicted by using the usual bond length for organic molecules. The barbiturate ring was assumed to be planar. Single point energy and molar volume calculations were carried out on the optimized geometries at the Hartree-Fock/3-21G level of theory. From the results





No.	R1	R2	R3	R4	X
1	H	H	H	H	O
2	methyl	methyl	H	H	O
3	3-pentyl	methyl	H	H	O
4	methyl	1-methylpentyl	H	H	O
5	ethyl	ethyl	H	H	O
6	ethyl	1-methylbutyl	H	H	O
7	ethyl	3-methylbutyl	H	H	O
8	ethyl	1-methylpropyl	H	H	O
9	ethyl	<i>n</i> -pentyl	H	H	O
10	buthyl	1-methylpropyl	H	H	O
11	buthyl	1-methylbutyl	H	H	O
12	buthyl	3-methylbutyl	H	H	O
13	ethyl	<i>n</i> -octyl	H	H	O
14	ethyl	3-dimethyloctyl	H	H	O
15	allyl	<i>i</i> -propyl	H	H	O
16	allyl	<i>i</i> -butyl	H	H	O
17	allyl	1-methylbutyl	H	H	O
18	methyl	cyclohexenyl	methyl	H	O
19	allyl	cyclopentenyl	H	H	O
20	ethyl	1-cyclohexenyl	H	H	O
21	ethyl	ethyl	H	H	S
22	ethyl	1-methylbutyl	H	H	S
23	allyl	1-methylbutyl	H	H	S
24	ethyl	1,3-dimethylbutyl	H	H	O
25	ethyl	phenyl	H	H	O
26	ethyl	ethyl	phenyl	H	O
27	ethyl	ethyl	benzoyl	H	O
28	ethyl	ethyl	benzoyl	benzoyl	O
29	ethyl	ethyl	<i>p</i> -Cl-benzoyl	H	O
30	ethyl	ethyl	<i>p</i> -NO <sub>2</sub> -benzoyl	H	O
31	ethyl	phenyl	<i>p</i> -NO <sub>2</sub> -benzoyl	<i>p</i> -NO <sub>2</sub> -benzoyl	O
32	ethyl	phenyl	phenyl	H	O
33	ethyl	phenyl	benzoyl	methyl	O
34	ethyl	phenyl	<i>p</i> -NH <sub>2</sub> -benzoyl	methyl	O
35	ethyl	phenyl	<i>o</i> -NO <sub>2</sub> -benzoyl	methyl	O
36	ethyl	phenyl	<i>p</i> -NO <sub>2</sub> -benzoyl	methyl	O
37	ethyl	phenyl	<i>m</i> -NO <sub>2</sub> -benzoyl	methyl	O
38	ethyl	ethyl	<i>p</i> -NO <sub>2</sub> -benzoyl	methyl	O
39	ethyl	ethyl	benzoyl	methyl	O
40	methyl	phenyl	benzoyl	H	O
41	methyl	phenyl	benzoyl	methyl	O
42	ethyl	phenyl	benzoyl	H	O
43	ethyl	methyl	H	H	O
44	ethyl	propyl	H	H	O
45	methyl	methyl	methyl	H	O

**Figure 1.** Chemical structures of barbituric acid derivatives. Barbituric acid derivatives were synthesized by Professor J. Bojarski Academy of Medicine, Krakow, Poland and co-workers.



**Table 1.** List of the Molecular Structural Descriptor Used in SRA and PCA Studies

Molecular Structural Descriptors	
Conventional Descriptors	
$\pi$	Hansch-Fujita's constant characterizing hydrophobicity
H-Ac	Indicator variable for proton acceptor properties
H-Do	Indicator variable for proton and donor properties
M-Re	Molar refractivity
F	Electronic parameter characterizing the inductive effect
R	Electronic parameters characterizing the resonance effects
$\sigma$	Hammett's constant, characterizing the electron withdrawing power
Es	Taft's constant, characterizing steric effects
B <sub>1</sub> and B <sub>4</sub>	Sterimol's width parameters determined by distance of the molecule at their maximum point perpendicular to attachment
Quantum Chemical Descriptors	
E <sub>T</sub>	Total energies
Vol	Molar volumes
$\mu$	Dipole moments

of this second series of calculations total energies ( $E_T$ ), molar volumes ( $Vol$ ), and dipole moments ( $\mu$ ) were obtained for each of the different barbiturates.

PCA was used to find the similarities and dissimilarities between the chromatographic parameters and peak characteristics of barbituric acid derivatives. The  $\lg k_0$ ,  $b$ ,  $AF$ ,  $N$ , and  $\lg k_0/b$  values were considered as variables and the barbituric acid derivatives were the observations. PCA reduces the dimensionality of the original data matrices; the resulting matrices of PC loadings and variables are sometimes even multidimensional. As the capacity of human brain to evaluate data distributed in multidimensional space is limited, the dimensions of the matrices of PC loadings and variables can be reduced to two by nonlinear mapping and cluster analysis techniques. Both traditional nonlinear mappings take into consideration the positive and negative signs of the correlations by constructing the corresponding maps. Necessarily, the variables with strong negative correlation are far from each other on the map. Theoretically, this discrepancy can be avoided by using only the absolute values for the constructing of the map. (17)

The two-dimensional nonlinear maps of principal component (PC), variables, original, and absolute value of PC loadings were calculated. Cluster analysis was carried out, also, on the original and the absolute value of PC loadings.



## RESULTS AND DISCUSSION

The slope ( $b$ ) and the intercept ( $\lg k_0$ ) values of the barbituric acid derivatives were found different in most cases, indicating the appropriate separation of the barbituric acid derivatives on this column in tetrahydrofuran-water eluent system (Table 2). The standard deviations are low in each instance, indicating the good reproducibility of the retention time and the stability of the amide embedded RP silica column. The correlation coefficients are above 0.9900 in most cases (Table 2) confirming the applicability of the Equation 1.

Stepwise regression analysis found significant relationship between the chromatographic parameters and the structural descriptors of the barbituric acid derivatives (Table 3). The results indicate that the hydrophobicity ( $\pi$ ) has great influence on the retention, and the effect of the steric and the proton acceptor or donor properties is negligible. The  $\lg k_0$ ,  $\lg k_0/b$ , and  $b$  values can be explained using only the hydrophobicity ( $\pi$ ) parameter of the barbituric acid derivatives. In the equations of the  $N$  and  $AF$  chromatographic parameters, electronic parameters are also included ( $\sigma$ ,  $\mu$  and  $R$ ), indicating that these physico-chemical parameters also have influence on the  $N$  and  $AF$  values.

The selected structural descriptors included in the equations also account for a relatively high ratio of change in these parameters: the  $r$  values 0.6479 and 0.6181, respectively. Although, it can be seen from the  $r$  values, that these descriptors can not properly explain the change of the chromatographic parameters, this indicates that other structural descriptors, not included in the SRA calculations, may have no negligible impact on the  $\lg k_0$ ,  $\lg k_0/b$ ,  $b$ ,  $N$ , and  $AF$  values.

Four principal components explain the majority of variance, indicating that the eighteen original variables can be substituted by four background variables with only 14.83% loss of information (Table 4). Retention parameters, together with the polarity parameters, have loadings in the first PC, indicating the marked influence of these physico-chemical parameters on the retention. Unfortunately, PCA does not prove the existence of such background variables as concrete physico-chemical entities, it only indicates their mathematical possibility.

Two-dimensional nonlinear maps calculated from the original PC loadings and the absolute value of the PC loadings, are shown in Figure 2A and 2B. Maps show marked differences in the distributions of the variables, indicating the considerable impact of the modifications of the calculation mode. The  $\lg k_0$  value, together with  $N$  and  $AF$  values, form a cluster with the hydrophobicity ( $\pi$ ) parameter and this cluster is far away from the  $b$ , and the  $\lg k_0/b$  values on the map calculated from the original PC loadings. However, the data in Table 4 clearly show significant relationship between the  $\lg k_0$ ,  $N$ ,  $AF$ , and the  $b$ ,  $\lg k_0/b$  values, but with an opposite sign. This finding supports our previous theoretical conclusions that the information contained in the two-dimensional nonlinear map





**Table 2.** Parameters of Linear Regression Between the Logarithm of Capacity Factor and Tetrahydrofuran Concentration (C) in the Eluent;  $\lg k = \lg k_0 + b \cdot C$

No.	$\lg k_0 \times 10^2$	$S(\lg k_0) \times 10^3$	$b \times 10^2$	$S(b) \times 10^3$	$r^2$
1			not significant		
2	-18.53	8.11	-0.45	0.19	0.9983
3	280.10	9.69	-4.59	0.30	0.9978
4	214.51	4.00	-3.73	0.89	0.9994
5	118.11	10.20	-2.96	0.59	0.9515
6	194.05	30.20	-3.51	0.67	0.9996
7	184.98	12.76	-3.38	0.28	0.9999
8	130.35	10.03	-2.41	2.22	0.9916
9	183.28	23.39	-3.35	0.52	0.9998
10	261.13	21.82	-4.50	0.48	0.9999
11	259.78	26.93	-4.46	0.60	0.9998
12	287.79	5.55	-4.98	0.12	0.9994
13	289.47	12.84	-4.79	2.84	0.9965
14	272.74	12.13	-4.55	2.68	0.9965
15	125.12	29.20	-2.49	0.65	0.9993
16	164.45	42.45	-3.06	0.94	0.9991
17			not significant		
18	136.75	31.72	-2.62	0.70	0.9993
19	157.14	51.05	-3.00	1.13	0.9986
20	143.00	6.10	-2.77	0.13	0.9976
21	255.83	40.29	-4.96	0.89	0.9687
22	286.94	13.18	-4.67	0.29	1.0000
23	306.55	15.46	-4.94	0.34	1.0000
24	216.06	10.52	-3.82	0.23	1.0000
25	136.01	5.13	-2.75	0.11	0.9983
26	351.23	56.47	-6.96	1.40	0.9609
27	198.47	33.43	-3.67	0.74	0.9996
28	137.51	10.82	-2.69	0.24	0.9921
29	270.41	10.87	-4.52	0.24	0.9972
30			not significant		
31			not significant		
32			not significant		
33	304.18	16.11	-5.40	0.36	0.9957
34	231.51	65.80	-4.38	1.46	0.9989
35	278.88	12.28	-4.88	0.03	0.9969
36	162.71	10.35	-3.09	2.29	0.9946
37	234.67	14.80	-4.25	0.33	0.9941
38	258.88	21.22	-4.26	0.47	0.9880
39	210.29	48.20	-3.74	1.07	0.9992
40	155.32	14.77	-3.04	0.33	0.9886
41	277.98	20.78	-4.91	0.49	0.9900
42	138.62	47.75	-2.69	1.09	0.9984
43	25.62	3.16	-1.11	0.07	0.9960
44	74.67	6.22	-1.80	0.14	0.9942
45	9.12	20.52	-0.82	0.45	0.9969

$S(\lg k_0)$  and  $S(b)$  are the standard deviations of the  $\lg k_0$  and  $b$  values, respectively.

**Table 3.** Relationship Between the Structural Descriptors of Barbituric Acid Derivatives and Their Retention Behavior: Results of Stepwise Regression Analysis

$y^1$	Parameters <sup>2</sup>											$n$	
	$a$	$b_1$	$x_1$	$b_1[\%]$	$b_2$	$x_2$	$b_2[\%]$	$b_3$	$x_3$	$b_3[\%]$	$r_{99.9\%}$		$F_{\text{calc}}$
$\lg k_0$	-0.249	0.590	$\pi$	100	-	-	-	-	-	-	0.7150	-	40
$b$	-0.013	0.005	$\pi$	100	-	-	-	-	-	-	0.6489	-	40
$\lg k_0/b$	-7.386	-11.43	$\pi$	100	-	-	-	-	-	-	0.7052	-	40
$N$	316	166	$\pi$	26.7	-537	$\sigma$	30.5	162	$\mu$	42.8	-	6.6	40
$AF$	0.673	0.291	$\pi$	60	-1.027	$R$	40	-	-	-	-	8.23	40

<sup>1</sup> $y = a + b_1x_1 + b_2x_2 + b_3x_3$ .

<sup>2</sup> $a$  = Intercept;  $b_1$ ,  $b_2$  and  $b_3$  = Regression coefficients;  $b_1[\%]$ ,  $b_2[\%]$  and  $b_3[\%]$  = Path coefficients (dimensionless numbers indicating the relative impact of the individual independent variables on the dependent variable);  $r_{99.9\%}$  = Coefficient of determination (indicates the ratio of variance explained by the independent variables);  $F$  = Calculated value of the Fisher significance test;  $n$  = Number of the barbituric acid derivatives were included in calculation.

may be misleading when both negative and positive correlation occur between the variables.

The distributions of the variables on the map 2B, calculated from the absolute values of the PC loadings (Figure 2B), correspond to the data in Table 4. This distribution shows that the all chromatographic parameters ( $\lg k_0$ ,  $b$ ,  $\lg k_0/b$ ,  $N$ ,  $AF$ ) and the hydrophobicity parameter ( $\pi$ ) form a separate cluster, while the other parameters are isolated from this cluster. The cluster analysis support the conclusion drawn from the distribution of the variables on the two-dimensional nonlinear maps, both for calculation from the original and the absolute values of the PC loadings (Figure 3A, 3B). Cluster dendrogram calculated from the original PC loadings (Figure 3A) shows that the  $\pi$ ,  $\lg k_0$ ,  $N$ , and  $AF$  values form a separate cluster, while cluster dendrogram calculated from the absolute values of PC loadings (Figure 3B) shows that  $\lg k_0$ ,  $b$ ,  $\pi$ ,  $\lg k_0/b$ ,  $N$ , and  $AF$  values form a separate cluster. Few clusters can be seen on the two-dimensional nonlinear maps calculated from the PC variables (Figure 4). Barbituric acid derivatives substituted with nitro-benzoyl groups (compounds 35–38), substituted with small, short-chain groups (compounds 2, 5, 43–45), and substituted with sulfur (compounds 21–23) form separate clusters. The cluster formation suggests that the ring structure of the solutes somewhat influences the retention behavior, probably having different interactions with the support.

It can be concluded that the barbituric acid derivatives can be well separated on embedded RP silica column using tetrahydrofuran-water eluent systems. Stepwise regression analysis and principal component analysis, followed by two-



**Table 4.** Similarities and Dissimilarities Between the Molecular Descriptors and Retention Characteristics of Barbituric Acid Derivatives on Discovery RPAmidC16 Column in Tetrahydrofuran-Water Eluent

No. of Principal Component	Eigenvalue	Variance Explained (%)	Total Variance Explained (%)
1.	9.566	53.14	53.14
2.	3.339	18.55	71.69
3.	1.462	8.12	79.81
4.	0.965	5.36	85.17

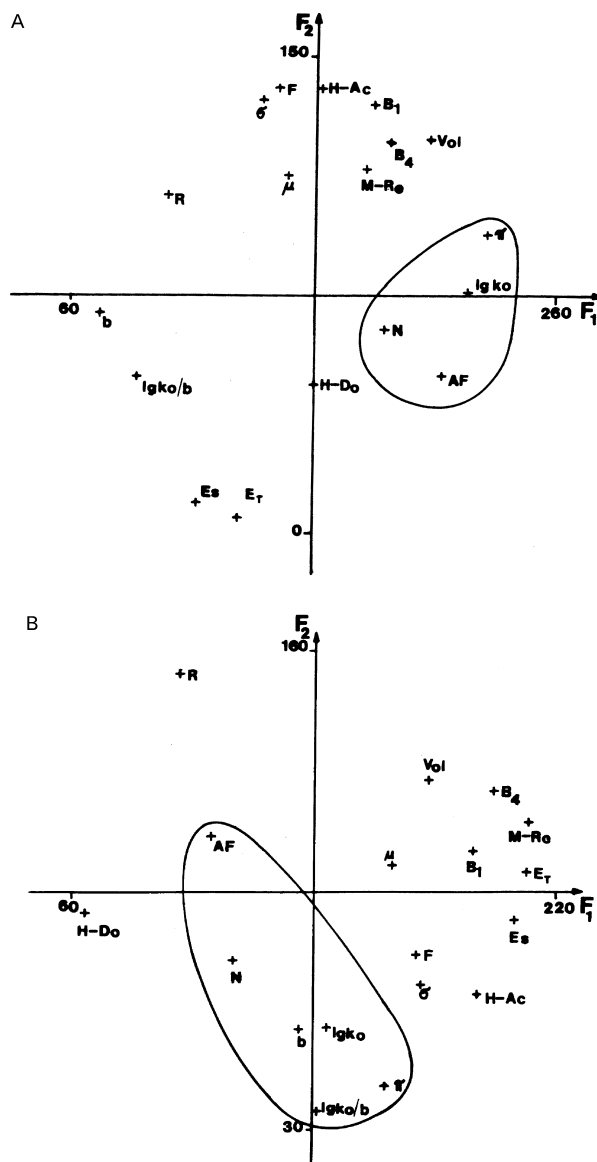
Principal Component Loadings

Parameter	No. of Principal Component			
	1.	2.	3.	4.
lg $k_0$	<u>0.581</u>	<u>-0.716</u>	0.093	-0.225
B	<u>-0.553</u>	<u>0.683</u>	-0.088	0.261
AF <sub>5</sub>	<u>0.405</u>	<u>-0.462</u>	<u>-0.513</u>	0.149
N	<u>0.476</u>	<u>-0.494</u>	<u>-0.121</u>	0.124
lg $k_0/b$	<u>-0.534</u>	<u>0.609</u>	-0.131	0.275
$\pi$	<u>0.558</u>	<u>0.682</u>	0.039	0.064
H-Ac	<u>0.843</u>	<u>0.471</u>	-0.115	0.083
H-Do	<u>0.161</u>	<u>-0.306</u>	<u>0.518</u>	<u>0.740</u>
M-Re	<u>0.967</u>	<u>0.039</u>	<u>-0.070</u>	<u>0.044</u>
F	<u>0.807</u>	<u>0.519</u>	0.014	-0.160
R	<u>0.234</u>	<u>0.088</u>	<u>0.886</u>	-0.159
$\sigma$	<u>0.792</u>	<u>0.520</u>	<u>0.133</u>	-0.202
Es	<u>-0.937</u>	<u>-0.246</u>	0.036	-0.094
B <sub>1</sub>	<u>0.924</u>	<u>0.125</u>	-0.045	0.230
B <sub>4</sub>	<u>0.944</u>	<u>-0.002</u>	-0.122	0.128
E <sub>T</sub>	<u>-0.939</u>	<u>-0.125</u>	-0.163	0.038
Vol	<u>0.876</u>	<u>-0.051</u>	-0.113	-0.042
$\mu$	<u>0.815</u>	<u>0.330</u>	-0.044	0.017

Results of principal component analysis calculated from the correlation matrix.

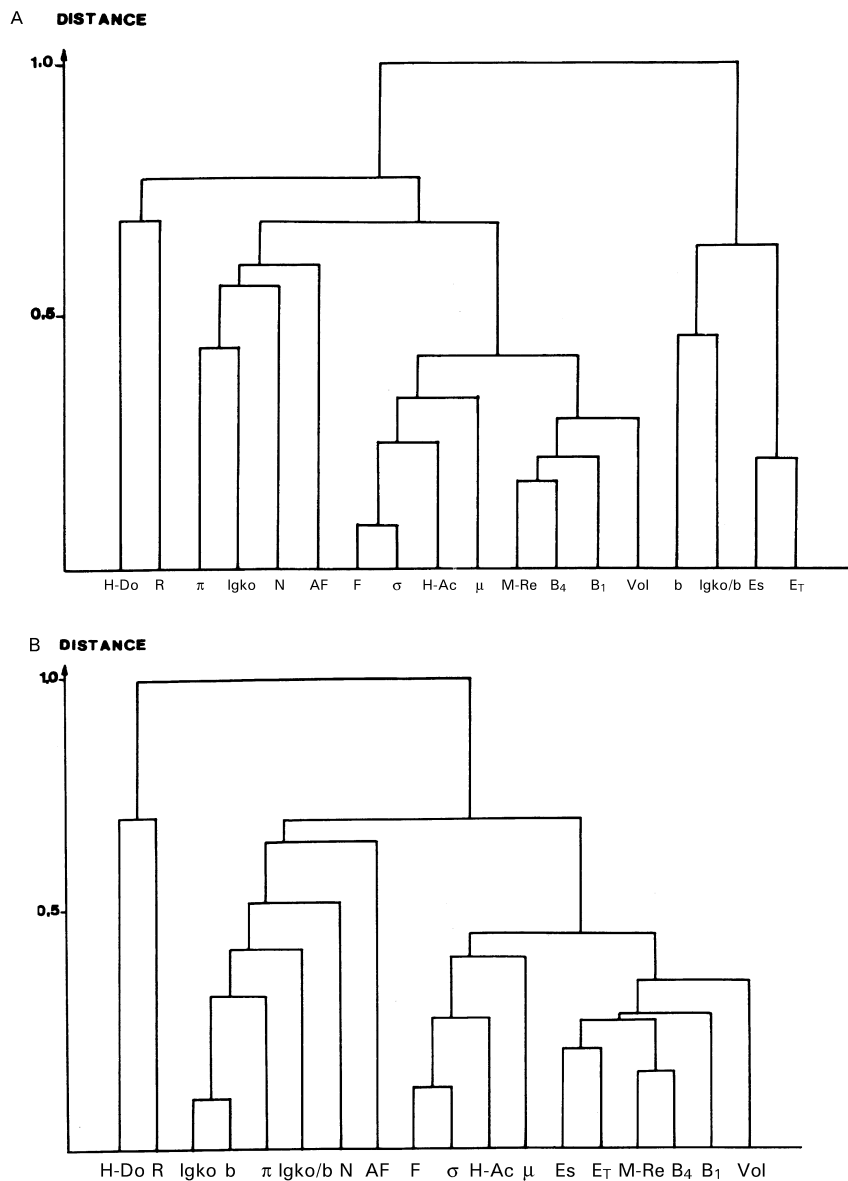
dimensional nonlinear mapping and cluster analysis calculated from the absolute values of the principal component loading, lead to the same results; that the hydrophobic characteristics of the solutes have the marked influence on the retention behavior of barbituric acid derivatives on this column in tetrahydrofuran-water eluent systems. This indicates that the effect of residual silanol groups is negligible, probably because the embedded amide groups react and cover the majority of the residual silanol groups. Use of the absolute values of principal





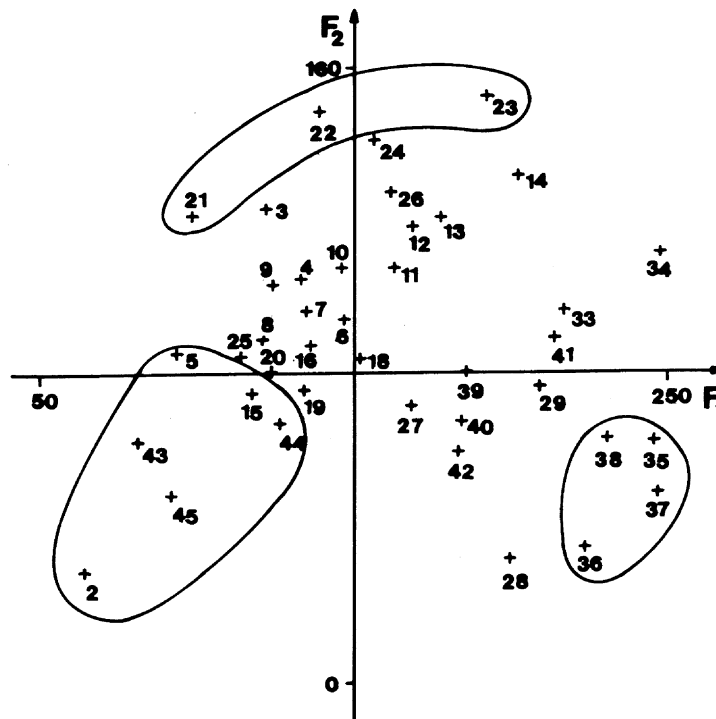
**Figure 2.** Relationship between the structural descriptors of barbituric acid derivatives and their retention behavior. Two-dimensional nonlinear maps, 2A: calculated from the original PC loadings (number of iterations: 253, maximum error:  $4.39 \cdot 10^{-2}$ ) and 2B: calculated from the absolute values of PC loadings (number of iterations: 142, maximum error:  $3.16 \cdot 10^{-2}$ ). For symbols, see Table 1 and the experimental section.





**Figure 3.** Relationship between the structural descriptors of barbituric acid derivatives and their retention behavior. Cluster dendrograms, 3A: calculated from the original PC loadings and 3/B: calculated from the absolute values of PC loadings. For symbols, see Table 1 and the experimental section.





**Figure 4.** Distribution of barbituric acid derivatives according to their retention behavior. Two-dimensional nonlinear map of principal component variables calculated from the original correlation matrices (number of iterations: 100, maximum error:  $3.05 \cdot 10^{-2}$ ). For symbols, see Figure 1.

component loadings for the calculations of two-dimensional nonlinear maps and cluster dendograms, prevents the occurrence of error originated from the different signs of the variables.

### REFERENCES

1. Kaliszan, R.; Straten, M.A.; Markuszewski, M.; Cramers, C.A.; Claessens, H.A. Molecular Mechanism of Retention in Reversed-Phase High-Performance Liquid Chromatography and Classification of Modern Stationary Phases by Using Quantitative Structure-Retention Relationships. *J. Chromatogr. A* **1999**, *855*, 455–486.



2. Körtvélyesi, T.; Görgényi, M.; Héberger, K. Correlation Between Retention Indices and Quantum-Chemical Descriptors of Ketones and Aldehydes on Stationary Phases of Different Polarity. *Anal. Chim. Acta* **2001**, *428*, 73–82.
3. Cserhati, T.; Forgacs, E.; Morasis, H.; Mota, T. Classification of Chili Powders by Thin-layer Chromatography and Principal Component Analysis. *J. Biochem. Biophys. Methods* **2000**, *45* (2), 221–229.
4. Bikker, J.A.; Kubanek, J.; Weaver D. F. Quantum Pharmacologic Studies Applicable to the Design of Anticonvulsants: Theoretical Conformational Analysis and Structure-Activity Studies of Barbiturates. *Epilepsia* **1994**, *35* (2), 411–425.
5. Micheli, A.; Sperduti, A.; Starita, A.; Bianucci, A.M. Analysis of the Internal Representations Developed by Neutral Networks for Structures Applied to Quantitative Structure-Activity Relationship Studies of Benzodiazepines. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (1), 202–218.
6. Pasadakis, N.; Yiokari, C.; Varotsis, N.; Vayenas, C. Characterization of Hydrotreating Catalysts Using the Principal Component Analysis. *Appl. Catal. A. Gen.* **2001**, *207*, 331–341.
7. Sasic, S. Eigenvalues and Principal Component Loadings of Heavily Overlapped Vibrational Spectra. *Spectrochim. Acta Part A* **2001**, *57 A* (2), 323–336.
8. Morrison, D.F. *Multivariate Statistical Methods*, 2nd Ed.; McGraw-Hill: New York, 1976.
9. Mardia, K.V.; Kent, J.T.; Bibby, J.M. *Multivariate Analysis*, Academic Press: London, 1979.
10. Cserhádi, T.; Bordás, B.; Ekiert, E.; Bojarski, J. Effect of Layer and Eluent Characteristics on the Reversed-Phase Thin-Layer Chromatographic Behaviour of Some Barbituric Acid Derivatives. *J. Chromatogr.* **1984**, *287*, 385–390.
11. Willett, P. *Similarity and Clustering in Chemical Information Systems*, Research Studies Press: New York, 1987.
12. Abraham, M.H.; Chadha, H.S.; Leitao, R.A.; Mitchell, E.R.C.; Lambert, W.J.; Kaliszan, R.; Nasal, A.; Habel, P. Determination of Solute Lipophilicity, as Log P(octanol) and Log P(alkane) using Poly(styrene-divinylbenzene) and Immobilised Artificial Membrane Stationary Phases in Reversed-Phase High-Performance Liquid Chromatography. *J. Chromatogr. A* **1997**, *766*, 35–47.
13. Herbert, B.J.; Dorsey, J.G. n-Octanol-Water Partition Coefficient Estimation by Micellar Electrokinetic Capillary Chromatography. *Anal. Chem.* **1995**, *67*, 744–749.
14. Forgács, E.; Cserhádi, T. Retention Behavior of Barbituric Acid Derivatives on a  $\beta$ -Cyclodextrin Polymer-Coated Silica Column. *Chromatogr. A* **1994**, *668*, 395–402.



15. Forgács, E.; Cserhádi, T. Dependence of the Retention of Some Barbituric Acid Derivatives on a Porous Graphitized Carbon Column. *J. Pharm. Biomed. Anal.* **1992**, *10*, 861–865.
16. Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley Sons: Chichester, 1979.
17. Forgács, E.; Kósa, A.; Csiktusnádi Kiss, G.; Cserhádi, T.; Kaliszan, R.; Haber, P.; Nasal, A. Use of Modified Nonlinear Mapping Method in Quantitative Structure Retention Relationship Study. *J. Liq. Chrom. Rel. Technol.* **1998**, *21* (16), 2523–2534.
18. Galvez, J.; Garcia-Domenech, R.; Salabert, M.T.; Soler, R. Charge Indexes. New Topological Descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520–525.
19. Toshihiko, H.; Hiroyuki, H.; Noriyuki, N.; Toshio, K. Quantitative Structure-Retention Relation (QSRR) in Liquid Chromatography Aided by Computational Chemistry. *J. Liq. Chromatogr.* **1993**, *16*, 1453–1462.
20. Karelson, M.; Lobanov, V.S.; Katritzky, A.R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027–1043.
21. Azzaoui, K.; Morin-Allory, L. Quantitative Structure-Retention Relationships for the Investigation of the Retention Mechanisms in High Performance Liquid Chromatography Using Apolar Eluent with Very Low Content of Polar Modifiers. *Chromatographia* **1995**, *40*, 690–696.
22. Montanari, M.L.C.; Cass, Q.B.; Montanari, C.A. Quantitative Structure-Retention Relationships of Antimicrobial Hydrazides Evaluated by Reverse-Phase Liquid Chromatography. *Chromatographia* **2000**, *51*, 723–726.
23. Montana, M.P.; Pappano, N.B.; Debattista, N.B.; Raba, J.; Luco, J.M. High-Performance Liquid Chromatography of Chalcones: Quantitative Structure-Retention Relationships Using Partial Least-Squares (PLS) Modeling. *Chromatographia* **2000**, *51*, 727–735.
24. Hennion, M-C.; Coquart, V.; Guenu, S.; Sella, C. Retention Behavior of Polar Compounds Using Porous Graphitic Carbon with Water Rich Mobile Phases. *J. Chromatogr. A* **1995**, *712*, 287–301.
25. Azzaoui, K.; Morin-Allory, L. Comparison and Quantification of Chromatographic Retention Mechanisms on Three Stationary Phases Using Structure-Retention Relationships. *Chromatographia* **2000**, *42*, 389–395.
26. Shieh, P.; Cooke, N.; Gant, R.; Eksteen, R. A Reliable Tool for Drug Discovery and Validation. *Amer. Lab.* **1998**, *30*, 66–70.
27. McCalley, D.V. Comparison of the Performance of Conventional C18 Phases with Others of Alternative Functionality for the Analysis of Basic Compounds by Reversed-Phase High-Performance Liquid Chromatography. *J. Chromatogr. A* **1999**, *844*, 23–38.





28. Ascah, T.L.; Kallbury, K.M.R.; Szafranski, C.A.; Corman, S.D.; Liu, F. Characterization and High Performance Liquid Chromatographic Evaluation of a New Amide-Functionalized Reversed Phase Column. *J. Liq. Chromatogr. Rel. Tech.* **1996**, *19*, 3049–3073.
29. O’Gara, J.E.; Alden, B.A.; Walter, T.H.; Peterson, J.S.; Niederlander C.L.; Neue, U.D. Simple Preparation of a C8 HPLC Stationary Phase with an Internal Polar Functional Group. *Anal. Chem.* **1995**, *67*, 3809–3813.
30. Valkó, K.; Olajos S.; Cserháti, T. Prediction of the High-Performance Liquid Chromatographic Retention Behaviour of Some Benzodiazepine Derivatives by Thin-Layer Chromatography. *J. Chromatogr.* **1990**, *499*, 361–371.
31. Horváth, C.; Melander, W.; Molnár, J. Solvophobic Interactions in Liquid Chromatography with Nonpolar Stationary Phases. *J. Chromatogr.* **1976**, *125*, 129–156.
32. Valkó, K.; Slégel, P. New Chromatographic Hydrophobicity Index  $\Phi$  Based on the Slope and the Intercept of the  $\log k'$  Versus Organic Phase Concentration Plot. *J. Chromatogr.* **1993**, *631*, 49–61.
33. *Gaussian 94* (Revision D.1), Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Gill, P.M.W.; Johnson, B.G.; Robb, M.A.; Cheeseman, J.R.; Keith, T.A.; Petersson, G.A.; Montgomery, J.A.; Raghavachari, K.; Al-Laham, M.A.; Zakrzewski, V.G.; Ortiz, J.V.; Foresman, J.B.; Peng, C.Y.; Ayala, P.Y.; Wong, M.W.; Andres, J.L.; Replogle, E.S.; Gomperts, R.; Martin, R.L.; Fox, D.J.; Binkley, J.S.; Defrees, D.J.; Baker, J.; Stewart, J.P.; Head-Gordon, M.; Gonzalez, C.; Pople, J.A. Gaussian, Inc., Pittsburgh PA, 1995.
34. Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Amer. Chem. Soc.* **1985**, *107*, 3902–3909.

Received July 20, 2001

Accepted August 15, 2001

Manuscript 5625



## **Request Permission or Order Reprints Instantly!**

Interested in copying and sharing this article? In most cases, U.S. Copyright Law requires that you get permission from the article's rightsholder before using copyrighted content.

All information and materials found in this article, including but not limited to text, trademarks, patents, logos, graphics and images (the "Materials"), are the copyrighted works and other forms of intellectual property of Marcel Dekker, Inc., or its licensors. All rights not expressly granted are reserved.

Get permission to lawfully reproduce and distribute the Materials or order reprints quickly and painlessly. Simply click on the "Request Permission/Reprints Here" link below and follow the instructions. Visit the [U.S. Copyright Office](#) for information on Fair Use limitations of U.S. copyright law. Please refer to The Association of American Publishers' (AAP) website for guidelines on [Fair Use in the Classroom](#).

The Materials are for your personal use only and cannot be reformatted, reposted, resold or distributed by electronic means or otherwise without permission from Marcel Dekker, Inc. Marcel Dekker, Inc. grants you the limited right to display the Materials only on your personal computer or personal wireless device, and to copy and download single copies of such Materials provided that any copyright, trademark or other notice appearing on such Materials is also retained by, displayed, copied or downloaded as part of the Materials and is not removed or obscured, and provided you do not edit, modify, alter or enhance the Materials. Please refer to our [Website User Agreement](#) for more details.

**[Order now!](#)**

Reprints of this article can also be ordered at

<http://www.dekker.com/servlet/product/DOI/101081JLC100108535>