ELSEVIER

Journal of Chromatography A, 728 (1996) 75–81

JOURNAL OF CHROMATOGRAPHY A

Relationship between the retention characteristics and physico-chemical parameters of some steroid drugs in reversed-phase high-performance liquid chromatography

Esther Forgács*, Tibor Cserháti

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

Abstract

The retention times of thirteen steroidal drugs were determined on an ODS column using methanol–0.025 *M* K_2HPO_4 mixtures as eluent. A linear correlation was calculated between the capacity factor (log k') and the concentration of methanol in the eluent. Principal component analysis was applied to find the physico-chemical parameters of solutes influencing their retention behaviour. The visualization of results was carried out by non-linear mapping, cluster analysis and varimax rotation. Each steroidal drug showed regular retention behaviour on ODS, the retention time decreasing monotonously with increasing concentration of methanol in the eluent. Principal component analysis demonstrated that five background components contain most (90.67%) of the information present in the eleven original variables. The chromatographic parameters form a loose cluster with the Taft's steric constant indicating that these parameters may also have a considerable influence on the reversed-phase chromatographic retention of these drugs.

Keywords: Principal component analysis; Retention behaviour; Hydrophobicity; Steroid drugs

1. Introduction

Reversed-phase (RP) chromatography has become the dominant branch of high-performance liquid chromatography (HPLC) over the last few decades. The term RP chromatography was a rational choice at a time when chromatography was practised almost exclusively by using a polar stationary phase and a non-polar eluent. However, today 80–90% of chromatographic systems used in HPLC work consist of non-polar stationary phases and polar eluents. The popularity of the technique rests with the reproducibility offered by the use of hydrocarbon-bonded phases and with the fact that the aqueous eluents have high optical transparency at low UV wavelengths and are cheaper, less toxic and less flammable.

Many HPLC methods have been developed for the separation of biactive steroids [1] using ion-pair HPLC [2], porous graphitic carbon columns [3], cyclodextrin-bonded phases [4], etc.

Quantitative structure–activity relationship (QSAR) methods play an important role in contemporary drug design [5,6]. Lipophilicity is one of the most important molecular properties applied in QSAR studies [7] because the biological activity of a molecule can generally be correlated with its ability to penetrate the differ-

* Corresponding author.

0021-9673/96/$15.00
SSDI 0021-9673(95)01085-8

ent hydrophobic barriers (membranes) of the target organs or organisms [8]. In addition to the classical partition method [9], lipophilicity can be determined by RP thin-layer chromatography (RP-TLC) [10,11], by RP-HPLC [12,13] and by gas–liquid chromatography [14].

Multivariate mathematical–statistical methods such as principal component analysis (PCA) [15] and cluster analysis [16] have been developed to extract maximum information from large data matrices. Both methods have been successfully used for the evaluation of data structure in HPLC [17,18].

The objectives of this study were the determination of the retention behaviour of some steroidal drugs on an octadecylsilica support, to calculate the hydrophobicity and specific hydrophobic surface area of the drugs, to find the physico-chemical parameters of solutes influencing the retention, to evaluate the retention data with principal component analysis followed by non-linear mapping, cluster analysis and varimax rotation and to compare the results of the various mathematical statistical methods.

2. Experimental

The HPLC system consisted of a Gilson (Villiers-le-Bel, France) Model 307 pump, a Cecil Instruments (Cambridge, UK) CE-212 variable-wavelength UV detector, a Valco (Houston, TX, USA) injector with a 20-μl sample loop and a Waters (Milford, MA, USA) Model 740 integrator. The reversed-phase column was Hypersil ODS (250 × 4 mm I.D., particle diameter 5 μm) (Phenomenex, Torrance, CA, USA). The flow-rate was 1 ml/min and the detection wavelength was set at 220 nm. Mixtures of aqueous 0.025 *M* KH_2PO_4 and methanol were used as eluents. The methanol concentration ranged from 60 to 85 vol.%.

The structures of the steroid drugs (a gift from Professor S. Görög, Gedeon Richter, Budapest, Hungary) are given in Table 1. The drugs were dissolved in methanol at a concentration of 0.05 mg/ml. The retention time of each compound in each eluent was determined with three consecutive determinations. Linear correlations was calculated between the log k' value of the drugs and the methanol concentration in the eluent:

$$\log k' = \log k'_0 + bC \quad (1)$$

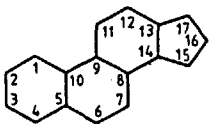
where log k' = logarithm of the capacity factor, log k'_0 = logarithm of the capacity factor extrapolated to zero methanol concentration in the eluent, b = change in log k' value caused by a unit change (1 vol.%) in methanol concentration and C = methanol concentration (vol.%). The intercept and slope values were considered as the best estimation of the hydrophobicity and specific surface area [19] of the drugs.

Principal component analysis (PCA) was used to find the physico-chemical parameters of drugs influencing their retention behaviour in RP-HPLC. The parameters of Eq. 1 (slope = log k'_0 and intercept = b) and various physico-chemical characteristics of the drugs (total twelve) were considered as variables and the steroidal drugs were the observations. The physico-chemical parameters included in the calculation were π = Hansch–Fujita substituent constant characterizing hydrophobicity; *H-Do* = indicator variable for proton donor properties; *M-RE* = molar refractivity; F and R = Swain–Lupton electronic parameters characterizing the inductive and resonance effect, respectively; σ = Hammett's constant characterizing the electron-withdrawing power of the substituent; E_s = Taft's constant characterizing steric effects of the substituent; and B_1 and B_4 = Sterimol width parameters. The limit of the variance explained was set to 99.9%.

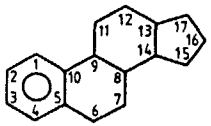
To facilitate the evaluation of the results of PCA, both two-dimensional non-linear mapping [20] and cluster analysis were carried out on the PC loadings and variables. Varimax rotation around two axes [21] was carried out only on the PC loadings. To elucidate the influence of PCA on the data evaluation, cluster analysis was also applied to the original data matrix. Cluster analysis and non-linear mapping are theoretically similar: both methods calculate and visualize the relative distances between the members of a data matrix (in our case physico-chemical and chromatographic parameters of drugs). To compare the information content of non-linear mapping

Table 1
Structures of steroid drugs

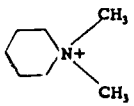
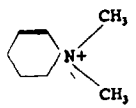
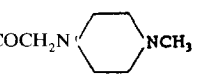
General structures



A (nor form)



B (nor form)

Compound No.	General structure	Substituent position						
		2	3	7	11	13[d]	16	17
I	A		$-OCOCH_3$	–	–			$-OCOCH_3$
II	A[a]	–	=O	–	–	–	–	–OH $-COCH_3$
III	A[a]	–	=O	–	–	–	–	–OH –C≡CH
IV	A[a]	–	=O	–	–	–	–	–OH $-COCH_2OH$
V	A[a]	–	–	–	–	–	–	–OH
VI	A[a]	–	=O	–	–	–	–	$-OCOCH_2C_6H_5$
VII	A[a]	–	$-OCOCH_3$	–	–	–	–C≡CH	$-OCOCH_3$
VIII	A[b]	–	=O	–	–OH	–	–	$-COCH_3$ C-16–$OC(CH_3)_2O$–C-17
IX	A[b]	–	=O	–	–OH	–	–	
X	A[c]	C^{-2}–CH=NO–C – 3		–	–	–	–C≡CH	–OH
XI	B	–	–OH	–	–	–	–	–OH
XII	B	–	$-OCOCH_3$	–	–	–	–	=O
XIII	B	–	–OH	–	–	–	–C≡CH	–OH

[a] Double bond between C-4 and C-5.
[b] Double bonds between C-1–C-2 and C-4–C-5.
[c] Double bonds between C-2–C-3 and C-4–C-5.
[d] Substitutents in position 13 are in nor form.

and varimax rotation techniques, linear correlations were calculated between their corresponding coordinates:

$$Y_{1-2} = a + bX_{1-2} \quad (2)$$

where Y_{1-2} = first and second coordinates of varimax rotation and X_{1-2} = first and second coordinates of the two-dimensional non-linear map.

3. Results and discussion

The parameters of Eq. 1 are given in Table 2. In each instance the relationship between the logarithm of the capacity factor and the methanol concentration in the eluent was linear. The value of regression coefficient was over 0.98, demonstrating the applicability of Eq. 1. Both the slope and intercept values show large differences, indicating that this set of steroidal drugs can be successfully separated by RP-HPLC by using an adequate eluent system.

The results of PCA are summarized in Table 3. Five principal components explain more than 90% of the total variance. This result indicates that the eleven physico-chemical and chromatographic parameters can be substituted by five background variables with only a 10% loss of information. Unfortunately, PCA does not prove the existence of such background variables as concrete physico-chemical entities but only indicates its mathematical possibility. The chromatographic parameters together with most of the physico-chemical characteristics of the drugs have high loadings in more than one principal component. This result indicates that not only the hydrophobicity of the drugs but also their other physico-chemical characteristics can influence their retention on an ODS column. This finding can be explained by the assumption that the free silanol groups not covered by the hydrophobic ligand can interact with the polar substructures of drugs exerting a marked influence on the retention. Both steric (availability of silanol groups) and electronic forces (hydrogen bonding between the solute and support surface) can be involved in the interaction.

The two-dimensional non-linear map of PCA

Table 2
Parameters of linear correlations between the logarithm of capacity factor (log k') and the methanol concentration (C) in the eluent: log $k' = \log k'_0 + bC$

Compound No.	log k'_0	$-b \times 10^2$	$s_b \times 10^3$	r
I	1.51	4.77	1.0	0.9975
II	4.33	6.72	0.9	0.9987
III	5.75	8.17	2.0	0.9992
IV	4.23	5.90	0.7	0.9987
V	4.16	6.16	0.7	0.9954
VI	4.41	6.40	1.2	0.9943
VII	4.51	5.68	1.3	0.9990
VIII	3.84	5.21	2.1	0.9897
IX	5.68	6.23	0.9	0.9889
X	3.82	5.39	1.4	0.9994
XI	5.05	7.15	1.0	0.9981
XII	3.92	4.93	0.5	0.9944
XIII	5.64	6.33	1.1	0.9961

Table 3
Relationship between retention characteristics and physico-chemical parameters for steroidal drugs: results of principal component analysis

Eigenvalue	Total variance explained (%)
4.05	38.83
2.55	60.04
1.69	75.46
1.05	85.02
0.62	90.67

Principal component loadings

Parameter	No. of principal components				
	1	2	3	4	5
π	0.56	0.51	0.20	0.48	0.06
H-Do	0.03	−0.85	0.36	−0.09	−0.09
M-RE	0.86	0.30	0.26	−0.07	0.16
F	0.56	0.09	0.50	0.32	−0.51
R	0.56	0.62	0.18	−0.24	0.28
σ	0.06	0.86	0.10	0.01	−0.15
E_s	−0.39	−0.28	0.44	0.57	0.39
B_1	0.94	−0.18	0.20	−0.01	0.01
B_4	0.84	−0.26	0.20	0.15	0.19
Log k'_0	0.47	0.17	−0.69	0.50	0.05
b	0.61	−0.35	−0.62	0.18	−0.10

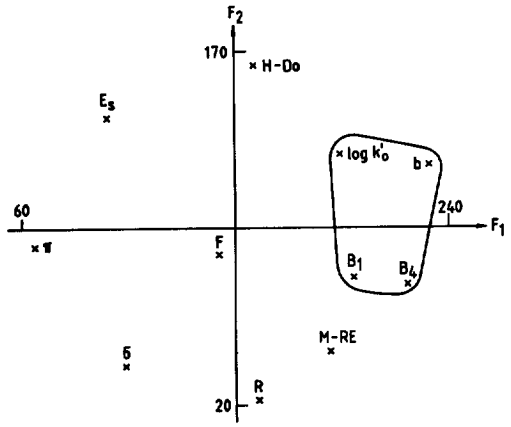


Fig. 1. Similarities and differences between the retention characteristics and physico-chemical parameters of steroidal drugs. Two-dimensional non-linear map of principal component loadings. Number of iterations, 177; maximum error, $4.61 \cdot 10^{-2}$. For symbols, see Experimental.

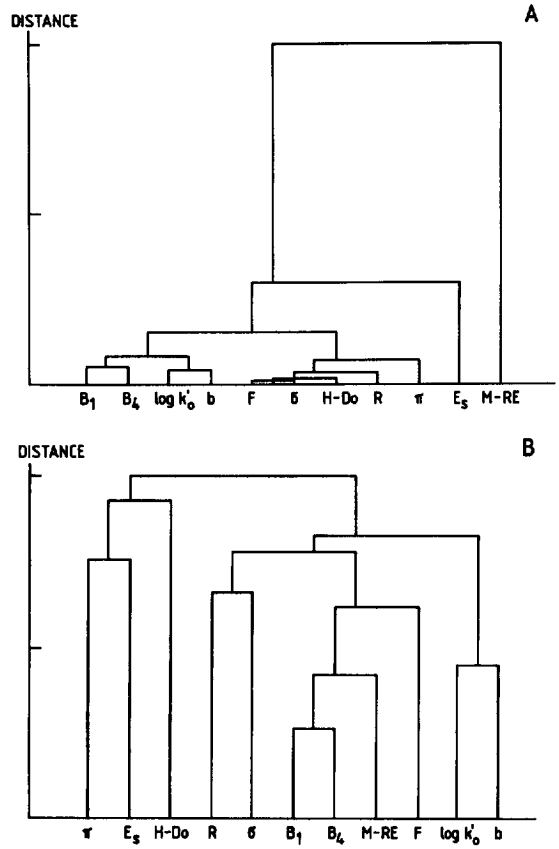


Fig. 2. Similarities and differences between the retention characteristics and physico-chemical parameters of steroidal drugs. Cluster dendograms calculated (A) from the original data matrix and (B) from the principal component loadings. For symbols, see Experimental.

loadings (distribution of RP-HPLC characteristics of drugs and their physico-chemical parameters in a plane) is shown in Fig. 1. The chromatographic parameters of the drugs (log k'_0, b) and their Sterimol width parameters (B_1, B_4) form a distinct cluster. This distribution of parameters suggests that retention behaviour of this set of steroidal drugs on the ODS column is mainly governed by their bulkiness. We assume that the marked impact of the molecular size on the RP-HPLC retention is due to the steric hindrance to large solute molecules contacting the hydrophobic octadecyl chains. We stress that the other physico-chemical parameters not included in the calculations may also have some influence on the retention of steroidal drugs, and our conclusion refers only to the parameter set applied. The distribution of the retention characteristics and physico-chemical parameters of drugs on the two-dimensional non-linear map PC loadings (Fig. 1) and on the clusters calculated from the original data matrix (Fig. 2A) and from PC loadings (Fig. 2B) show considerable differences. These discrepancies may be due to the different methods of calculation and to the effect of PCA on the structure of the original data matrix.

The distribution of steroidal drugs according to their retention characteristics and physico-chemical parameters (two-dimensional non-linear map of PC variables) is shown in Fig. 3. The drugs containing a polar OH group form a well defined cluster, well separated from other drugs. This finding suggests again that electronic forces are involved in the retention mechanism also in the RP-HPLC separation mode. The cluster dendograms of drugs calculated from the original data matrix and from the PC variables are shown in Fig. 4A and B, respectively. The dendograms are different, supporting our previous conclusion that PCA can cause data distortion which may influence the distribution in cluster analysis. Although cluster analysis and

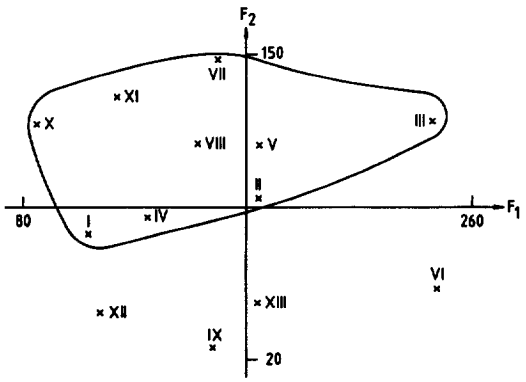


Fig. 3. Distribution of steroidal drugs according to their retention characteristics and physico-chemical parameters. Two-dimensional non-linear map of principal component variables. Number of iterations, 238; maximum error, $4.68 \cdot 10^{-2}$. Numbers refer to steroidal drugs in Table 1.

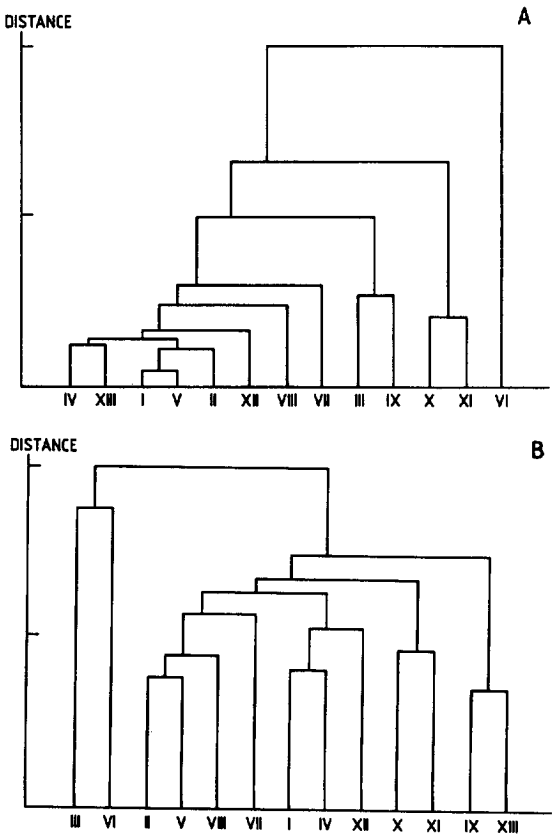


Fig. 4. Distribution of steroidal drugs according to their retention characteristics and physico-chemical parameters. Cluster dendograms calculated (A) from the original data matrix and (B) from the principal component variables. Numbers refer to steroidal drugs in Table 1.

non-linear mapping give similar results, we strongly advocate the application of two-dimensional non-linear mapping instead of cluster analysis cause of its higher dimensionality. We assume that the two-dimensional non-linear map may contain more information than the one-dimensional structure of clusters.

Significant linear correlations were found between the rotated PC loadings and the coordinates of the two-dimensional non-linear map of PC loadings. The correlation coefficients were 0.9124 and 0.8468 for the first and the second coordinates, respectively ($n = 11$). This finding indicates that the results of the varimax rotation and non-linear mapping techniques are similar and both can be successfully used to reduce the dimensionality of retention data matrices.

Acknowledgement

This work was sponsored by grant OTKA 2670.

References

[1] A. Lagana and A. Marino, J. Chromatogr., 588 (1991) 89.
[2] V. Legrand-Defretin, C. Juste, R. Henry and T. Corting, Lipids, 26 (1991) 578.
[3] B.J. Clark, in S. Görög and E. Heftmann (Editors), Advances in Steroid Analysis 90, Proceedings of 4th Symposium on the Analysis of Steroids, 1990, Akadémiai Kiadó, Budapest, 1991, p. 129.
[4] A.-H. Ahmed, S.M. El-Gizany and N.M. Omar, Anal. Lett., 24 (1991) 2207.
[5] J. Andrew, W.E. Stuper, P. Brugger and S. Surs, Computer Assisted Studies of Chemical Structure and Biological Function, Wiley, New York, 1979.
[6] C.J. Cavallito, Structure–Activity Relationships, Vol. 1, Pergamon Press, New York, 1973.
[7] C. Hansch and J.M. Clayton, in C.J. Cavallito (Editor), Structure–Activity Relationships, Pergamon Press, Oxford, 1973, p. 75.
[8] T. Fujita, J. Iwasa and C. Hansch, QSAR and Strategies in the Design of Bioactive Compounds, VCH, Weinheim, 1964, p. 59.
[9] C. Hansch and S.M. Anderson, J. Org. Chem., 32 (1967) 2583.

[10] G.L. Biagi, M.F. Gamba, A.M. Barbaro and M.C. Guerra, J. Chromatogr., 41 (1969) 371.
[11] G.L. Biagi, M.C. Guerra, A.M. Barbaro and M.F. Gamba, J. Med. Chem., 13 (1970) 511.
[12] W.J. Haggerty and E.A. Murril, Res. Dev., 25 (1974) 30.
[13] J.M. McCall, J. Med. Chem., 18 (1975) 549.
[14] D.R. Clifford and D.M. Watkins, Pestic. Sci., 2 (1971) 41.
[15] K.V. Mardia, J.T. Kent and J.M. Bibby, Multivariate Analysis, Academic Press, London, 1979.
[16] P. Willett, Similarity and Clustering in Chemical Information System, Research Studies Press, New York, 1987.
[17] E. Forgács, T. Cserháti and B. Bordás, Chromatographia, 36 (1993) 19.
[18] E. Forgács, T. Cserháti and B. Bordás, Anal. Chim. Acta, 279 (1993) 115.
[19] C. Horváth, W. Melander and I. Molnár, J. Chromatogr., 125 (1976) 129.
[20] J.W. Sammon, Jr., IEEE Trans. Comput., C18 (1969) 401.
[21] J.R. Llinas and J.M. Ruiz, in G. Vernin (Editor), Computer Aids to Chemistry, Ellis Horwood, Chichester, 1986.