ELSEVIER

Journal of Chromatography A, 822 (1998) 263–269

JOURNAL OF CHROMATOGRAPHY A

Determination of lipophilicity of some non-steroidal anti-inflammatory agents and their relationships by using principal component analysis based on thin-layer chromatographic retention data

Costel Sârbu*, Sorina Todor

"Babeş-Bolyai" University, Faculty of Chemistry and Chemical Engineering, Arany Janos 11, 3400 Cluj-Napoca, Romania

Received 2 February 1998; received in revised form 1 May 1998; accepted 7 July 1998

Abstract

The relative lipophilicity of ten non-steroidal anti-inflammatory agents have been determined by reversed-phase thin layer chromatography using different reversed-phase high-performance thin-layer chromatography plates and water–methanol mixtures as eluents. The compounds studied showed regular retention behavior, their R_M values decreasing linearly with increasing concentration of methanol in the eluent. Principal component analysis allowed a more rational and objective estimation and comparison of lipophilicity determined by reversed-phase thin-layer chromatography. It also affords a useful graphical tool, since scatterplots of the scores onto the plane described by the first two components will have the effect of separating compounds from each other most effectively, thus obtaining a ''congeneric lipophilicity chart''.

Keywords: Lipophilicity; Principal component analysis; Congeneric lipophilicity chart; Non-steroidal anti-inflammatory drugs

1. Introduction

The use of multivariate methods in developing models that relate molecular structure to a physical, chemical, or biological property has found growing acceptance and application not only in the design of drugs, pesticides, etc., but also in the study of biochemical and biophysical processes [1,2]. To obtain a significant correlation, it is crucial that appropriate descriptors be employed, whether they are theoretical, empirical, or derived from readily available experimental characteristics of the structures. Many descriptors reflect simple molecular properties and thus can provide insight into the physicochemical nature of the activity/property under consideration [3–5].

In this order, however, the lipophilicity seems to be the most important molecular parameter used in quantitative structure–activity relationship studies because the biological activity of a molecule can generally be correlated with its ability to penetrate the different hydrophobic barriers.

*Corresponding author.

0021-9673/98/$ – see front matter
PII: S0021-9673(98)00571-8

Lipophilic character has been defined in many ways, the best known of which is probably the partition coefficient, *P*, which represents the tendency of a molecule to partition itself between organic and aqueous phases:

$$P = C_o / C_{aq} \quad (1)$$

where C_o and C_{aq} represents concentrations in the organic and aqueous phases, respectively.

Usually the octanol–water partition coefficient (log *P*) estimated by direct equilibration method or by calculation according to different mathematical models is used for correlation with lipophilicity [4–18].

The use of R_M values obtained from various types of paper and thin layer chromatography as hydrophobic parameters for correlation with biological activity appears to be more suitable as a support of the lipophilicity scale [19–21]:

$$R_M = \log(1/R_f - 1) \quad (2)$$

The advantages of chromatographic methods are the small quantity of compounds needed for the determination and the less strict requirement of purity because the impurities separate during the chromatographic process. They are rapid and relatively simple. In addition, we have to stress the dynamic aspect of the chromatographic process. The R_M value (related to the molecular lipophilicity), determined by use of RPTLC, generally depends linearly on the concentration of the organic component of the mobile phase [22]. To increase the accuracy of lipophilicity determination, the R_M value extrapolated to zero organic component concentration (R_{Mo}) has been calculated from the linear correlation between the R_M value and the concentration of the organic component of the mobile phase [23]:

$$R_M = R_{Mo} + bc \quad (3)$$

where R_M values were calculated using Eq. (2).

The intercept (R_{Mo}) in the Eq. (3) represents the extrapolated R_M values, i.e., the theoretical R_M values at 0% organic solvent. In other words the intercept determined using this TLC equation can be considered as an estimation of the partitioning of compounds between nonpolar stationary phase and the aqueous system, and hence all the compounds studied can be compared on the basis of their lipophylicity determined in this way [11–35].

Concerning the reversed-phase thin-layer chromatography (RPTLC), it has been additionally stated that not only the R_M value extrapolated to zero organic solvent content, R_{Mo}, but also the slope, *b*, of the regression line (3) can be used as an estimation of lipophilicity [24]. For a homologous series of compounds there was a significant linear correlation between slope and R_{Mo} values [25], but for a nonhomologous series both parameters were needed to describe the lipophilicity accurately [26]. The slope has been regarded as a characteristic of the specific hydrophobic surface area of the compounds [27]. Although these results were also systematically observed within the large groups of very different chemical compounds Cserháti et al. [11,13,14], who were very active in this field, have suggested that the grade of the linear correlation could indicate the presence or absence of a congeneric class.

Recently, we addressed the question if a high linear correlation between the intercept and slope can be an objective criterion to find the real (natural) congeneric classes within the chemical compounds series. Moreover, using an algorithm based on fuzzy regression we demonstrated that within the families of straight lines there is always a high linear correlation between the slope and intercept [34,35].

The purpose of this paper is to investigate the feasibility of the scores, obtained by principal component analysis using RPTLC retention data, as a measure of lipophilicity in correlation with biological activity of non-steroidal anti-inflammatory agents. PCA also affords a useful graphical tool, since scatterplots of the scores onto a plane described by the first two components will have the effect of separating the compounds one from each other most effectively, obtaining in this way a ‘‘congeneric lipophilicity chart’’.

2. Principal components analysis (PCA)

PCA is also known as eigenvector analysis, eigenvector decomposition or Karhunen–Loéve expansion. Many problems from analytical chemistry are strongly related to PCA. The main purpose of PCA is to represent in an economic way the location of

the objects in a reduced coordinate system where instead of n-axes (corresponding to n variables) only p ($p<n$) can usually be used to describe the data set with maximum possible information. The new variables are called principal components and they are given by the linear combination of the n real variables,

$$\boldsymbol{P}_j = a_{1j}\boldsymbol{x}_1 + a_{2j}\boldsymbol{x}_2 + \cdots + a_{nj}\boldsymbol{x}_n \quad (4)$$

where x_i represents the original variables, while the coefficients a_{1j}, a_{2j}, ..., a_{nj} give an indication of the relative importance of the corresponding raw variable in the factor. These coefficients are often called loadings and the new values corresponding to each principal component for every object are called factor scores or scores. The new principal components are also called latent factors and their exact interpretation is an important task.

Moreover, it may well turn out that usually two or three principal components provide a good summary of all the original variables. Scores plots are very useful as a display tool for examining the relationships between objects, looking for trends, grouping or outliers.

3. Experimental

The non-steroidal anti-inflammatory agents aspirin, indomethacin, ibuprofen, ketoprofen, naproxen, diclofenac, piroxicam, tenoxicam, phenylbutazone and niflumic acid were provided by Terapia (Cluj–Napoca, Romania) and were used as supplied.

The chromatographic behavior of these compounds was studied on the C_{18} and CN silica gel bonded plates and also on paraffin-oil coated plates. HPTLC plates (10×10 cm) RP-18W/UV_{254} and Nano–Sil CN/UV_{254} were obtained as a gift from Machery–Nagel (Düren, Germany). Paraffin-oil coated plates were prepared as described previously [36,37] using also as support HPTLC Machery–Nagel plates precoated with silica gel (Nano–DURASIL-20 UV254). Methanol for chromatography was obtained from Reactivul (Bucharest, Romania).

Solution of sample (0.1%) were prepared using ethanol for aspirin methanol for diclofenac and ethanol–acetone (1:1) for the rest of the studied compounds; 3 μl of each solution was spotted to origin of the plates by hand. Chromatography was performed in a normal developing chamber at room temperature, the developing distance being 8 cm. Methanol was used as the organic component of the mobile phase in the concentration range of 40–80% (v/v) for the C_{18} plates and only of 40–60% (v/v) for CN plates in steps of 10%, as the studied compounds differed considerably in their retention. After being developed, the dried plates were observed under ultraviolet light at 254 nm with a Camag lamp.

Principal components analysis was performed on the retention data by the use of a computer program discussed in [38]. It display objects (non-steroidal anti-inflammatory agents) in a reduced space by finding a direction (first principal component) that best preserves the scatter of the observations (100×R_f values) in the multidimensional space. PCA gives both the coordinates (or scores) of the studied compounds and the loadings of variables (solvents) on the principal components.

The results obtained from the initial matrix (dimensions: 10×5 for C_{18} and 10×4 for CN bonded plates) using covariance matrix (without autoscaling) can presented as usually in three panels, although typically there are only two. The first panel shows the table of data statistics, the second is the table of components and the third panel displays the eigenvectors associated with each of the components. Table 1 lists the eigenvalues of the covariance matrix, ordered from largest to smallest, the third column of this table shows the difference between each eigenvalues and the next smallest eigenvalue and the fourth column shows the proportion. These results suggest a significant two component model which explained 99.33% of the total variance (information) in the case of C_{18} plates and 99.76% in the case of CN plates, respectively, considering only the eigenvalues higher than one in each case. The first component explains 94.45% of the total variance and the second only 4.88% in the first case (C_{18} plates) and 95.73% and 4.03%, in the second case (CN plates), respectively. In both situations the subsequent eigenvalues are just sampling noise. When the significance of the component model retained was tested applying the Bartlet's statistics

Table 1
The eigenvalue of the five components in the case of C_{18} plates and of four components in the case of CN plates, respectively using covariance matrix

Component	Eigenvalue	Difference	Proportion, %	Cumulative
1	0.162206	0.153828	94.45	94.45
2	0.008378	0.007790	4.88	99.33
3	0.000588	0.000080	0.34	99.67
4	0.000505	0.000439	0.29	99.96
5	0.000066		0.04	100.00
1	0.141240	0.135301	95.73	95.73
2	0.005939	0.005694	4.03	99.76
3	0.000245	0.000122	0.17	99.93
4	0.000123		0.08	100.00

[38], testing the hypothesis that (p–k) eigenvalues in the variance–covariance matrix are equal, a model with two components was also selected in both cases, i.e. C_{18} and CN bonded plates, respectively. Table 2 displays the eigenvectors (loadings) associated with the first two principal components in Table 1 for each type of chromatographic plate investigated. The first eigenvector goes with the first eigenvalue, and so on. The results confirm that the five (four) solvents are strongly related to each other and so could be reduced. This shows that for most compounds there was generally a regular increase in R_f values with increasing methanol content; the loadings corresponding to the four solvent systems on the first principal component in the case of CN plates are practically the same.

Here we have to mention that the results obtained on the paraffin-oil coated plates were very similar to those obtained on CN bonded plates and their presentation appears to be superfluous.

Table 2
Loadings associated with the first two principal components in Table 1, in the case of C_{18} and CN plates, respectively

Variable (Solvent)	C_{18}		CN	
	P_1	P_2	P_1	P_2
40–60	0.2951	−0.5686	0.4398	−0.7562
50–50	0.3736	−0.3761	0.5130	−0.1268
60–40	0.4813	−0.2816	0.5362	0.1535
70–30	0.5156	0.2058	0.5059	0.6233
80–20	0.5251	0.6431		

4. Results and discussion

The results of regression analysis using Eq. (3) are compiled in Tables 3 and 4. The statistics obtained (see also Tables 3 and 4) illustrate that the linear equation fits excellently to the experimental data, the

Table 3
Regression data, scores on the first two principal components for the investigated non-steroidal anti-inflammatory agents on RP-18W/UV_{254} plates

Compound	R_{Mo}	b	r	R^2	P_1	P_2
Aspirin(**1**)	1.89	−4.26	−0.9588	0.9193	1.716	0.088
Indomethacin(**2**)	2.62	−3.53	−0.9949	0.9898	0.717	0.313
Ibuprofen(**3**)	2.14	−2.93	−0.9995	0.9990	0.791	0.279
Ketoprofen(**4**)	1.74	−2.73	−0.9986	0.9972	1.078	0.250
Naproxen(**5**)	2.07	−3.02	−0.9985	0.9970	0.920	0.292
Diclofenac(**6**)	2.40	−3.29	−0.9999	0.9998	0.771	0.318
Piroxicam(**7**)	1.90	−2.09	−0.9944	0.9888	1.029	0.269
Tenoxicam(**8**)	1.40	−2.53	−0.9893	0.9787	1.314	0.172
Phenylbutazone(**9**)	2.71	−1.98	−0.9902	0.9805	0.085	0.036
Niflumic acid(**10**)	1.60	−2.53	−0.9862	0.9726	1.087	0.236

Table 4
Regression data, scores on the first two principal components for the investigated non-steroidal anti-inflammatory agents on Nano–Sil CN/UV_{254} plates

Compound	R_{Mo}	b	r	R^2	P_1	P_2
Aspirin(**1**)	0.56	−2.50	−0.9995	0.9990	1.705	0.064
Indomethacin(**2**)	1.50	−2.83	−0.9870	0.9742	1.075	0.284
Ibuprofen(**3**)	0.75	−1.78	−0.9875	0.9690	1.250	0.144
Ketoprofen(**4**)	0.52	−1.67	−0.9540	0.9101	1.420	0.110
Naproxen(**5**)	0.70	−1.70	−0.9870	0.9742	1.270	0.142
Diclofenac(**6**)	1.33	−2.56	−0.9870	0.9742	1.110	0.271
Piroxicam(**7**)	0.72	−1.58	−0.9722	0.9448	1.170	0.137
Tenoxicam(**8**)	0.52	−1.34	−0.9541	0.9101	1.260	0.103
Phenylbutazone(**9**)	2.31	−2.14	−0.9750	0.9510	0.158	0.070
Niflumic acid(**10**)	1.30	−2.64	−0.9924	0.9841	1.170	0.248

linear model explaining approximately 99% of the total variance (see R^2 values) in the majority of cases. Despite of a large difference concerning the chemical structure of the compounds studied, a good correlation was also found between the R_{Mo} and b values of Eq. (3) as is shown in Figs. 1 and 2 . The exceptions to this were the compounds (**1**) and (**9**), namely aspirin and phenylbutazone. This finding could indicate that the intercept, R_{Mo}, (lipophilicity) and slope, b, (specific hydrophobic surface area) for the majority of these drugs are intercorrelated and, in that case, they might form a homologous series of compounds as has been suggested by some authors [11,13,24]. The position of outliers of the aspirin and phenylbutazone is in good agreement with their pharmatherapeutical effects [39] and this could be explained, for example, by comparing their quite different chemical structure. Moreover, a higher correlation was observed between R_{Mo} values and the scores of the same non-steroidal anti-inflammatory agents on the first principal component as is shown in Figs. 3 and 4 for each case. On the basis of these findings and as can be seen from data provided in Tables 3 and 4, the scores on the first principal component can replace efficiently the R_{Mo} values in the estimation experiments of the lipophilicity. It is

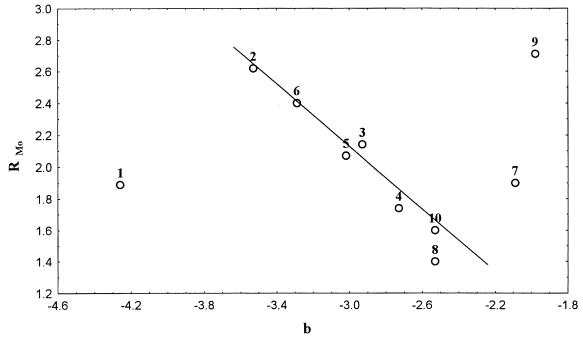

Fig. 1. Relationship between R_{Mo} and b (C_{18} plates).

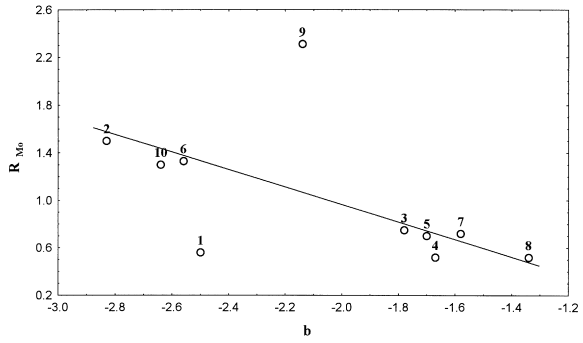

Fig. 2. Relationship between R_{Mo} and b (CN plates).

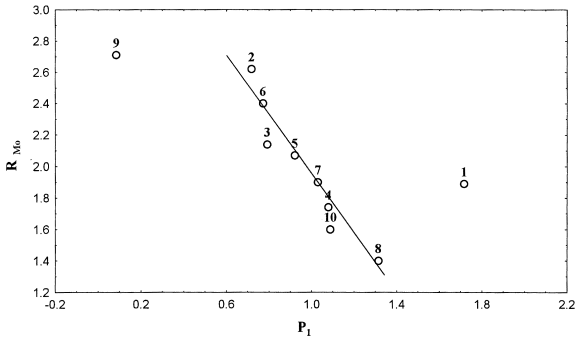

Fig. 3. Relationship between R_{Mo} and P_1 (C_{18} plates).

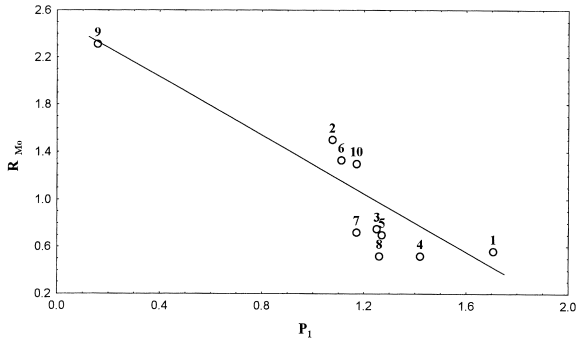

Fig. 4. Relationship between R_{Mo} and P_1 (CN plates).

easy to observe that for the compounds with practically the same R_{Mo}, the corresponding scores on the first principal component are quite different.

Moreover, as is shown in Figs. 5 and 6, score plots are very useful as a display tool for examining the relationships between compounds, looking for trends, groupings or outliers. Hence, graphing scores onto the plane described by P_1 and P_2 we obtain what one could be call the ''congeneric lipophilicity chart''. It appears clearly that the compounds studied in this paper form practically two congeneric classes (groups) in each case, two diffuse classes in the first case (C_{18} plates) and more well defined in the second case (CN plates), in a very good agreement with the selected eigenvectors when the Barlett test was applied. Again, phenylbutazone appears as an outlier and this reveals its very different chemical structure compared with the rest of the drugs. This surprising fact contradicts the results obtained using linear correlation between the intercept (lipophilicity) and slope (specific hydrophobic area) and suggests that the last approach is incapable of finding the real number of congeneric classes. Taking into account these results and others yet unpublished, we appreciate that some similar discrepancies, already observed in the literature [11], could be explained in this way.

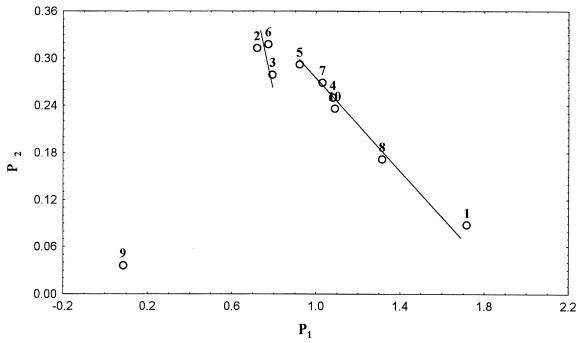

Fig. 5. Congeneric lipophilicity chart in the case of C_{18} plates.

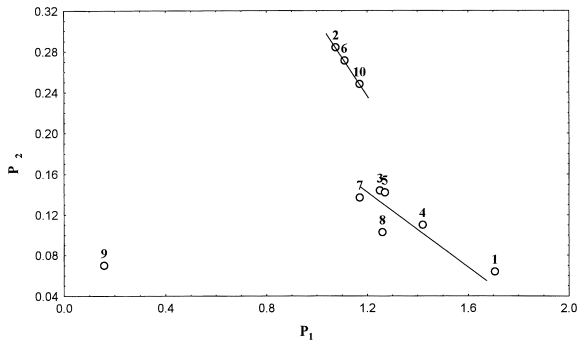

Fig. 6. Congeneric lipophilicity chart in the case of CN plates.

Additionally, comparing the results obtained on the three types of reversed-phase TLC we have to note that the Nano–Sil CN/UV254 plates appear to be more suitable for these kind of studies.

5. Conclusions

The results presented herein using principal component analysis allow a more rational and objective estimation and comparison of lipophilicity determined by RPTLC. Scatterplots of the scores onto planes described by the most important components will have the effect of separating compounds one from each other most effectively into ''congeneric lipophilicity charts'' and the scores described by first principal component can be assessed as a relative estimation of the lipophilicity. It should, also, be remembered that R_M is a logarithmic quantity, so that it will be more sensitive (in a certain way than the score) to the errors of R_f. Moreover, in some cases for the same R_{Mo} values, the corresponding score values on the first principal component are quite different. It is important to remark also that the number of congeneric classes corresponds to the number of the significant principal components (eigenvalues) and in this way we have a much more objective and rational method of finding the number

of congeneric classes. This main conclusion is clearly supported by the results reported in this paper and others yet unpublished. Additionally, we appreciate that using principal component analysis it should be possible to explain some of the discrepancies observed in the literature concerning the congeneric series of compounds.

References

[1] S.D. Brown, T.B. Blank, S.T. Sum, L.G. Weyer, Anal. Chem. 66 (1994) 337R.
[2] S.D. Brown, S.T. Sum, F. Despagne, Anal. Chem. 68 (1996) 42R.
[3] M. Karelson, V.S. Lobanov, A.R. Katritzky, Chem. Rev. 96 (1996) 1027.
[4] R.F. Rekker, R. Mannhold, Calculation of Drug Lipophilicity, Weinheim, 1992.
[5] M.V. Diudea, O. Ivanciuc, Molecular Topology, Comprex Publishing House, Cluj, 1995.
[6] N. Okano, M. Murakami, Y. Miyamoto, K. Koizumi, H. Ogawa, K. Wakabayaski, J. Pestic. Sci. 18 (1993) 361.
[7] A. Tanaka, K. Nakamura, I. Nakanishi, H. Fujiwara, J. Med. Chem. 37 (1994) 4563.
[8] K. Nakamura, K. Hayashi, I. Ueda, H. Fujiwara, Chem. Pharm. Bull. 43 (1995) 369.
[9] S. Hatrik, J. Lehotay, J. Cizmarik, Collect. Czech. Chem. Commun. 60 (1995) 960.
[10] C. Hansch, S.M. Anderson, J. Org. Chem. 32 (1967) 2583.
[11] Y. Darwish, T. Cserháti, E. Forgacs, J. Planar Chromatogr. 6 (1993) 458.
[12] M.L. Bieganovska, A. Doraczynska-Szopa, A. Petruczynik, J. Planar Chromatogr. 8 (1995) 122.
[13] T. Cserháti, J. Biochem. Biophys. Methods 27 (1993) 133.
[14] T. Cserháti, J. Liq. Chromatogr. 16 (1993) 1805.
[15] G. Trapani, C. Altomare, M. Franco, A. Latrofa, G. Liso, Int. J. Pharm. 116 (1995) 95.
[16] J.M. McCall, J. Med. Chem. 18 (1975) 549.
[17] D.R. Clifford, D.A.M. Watkins, Pestic. Sci. 2 (1971) 41.
[18] B. Kocjan, J. Sliwiok, J. Planar Chromatogr. 7 (1994) 327.
[19] C.B.C. Boyce, B.V. Milborrow, Nature 208 (1965) 537.
[20] E. Tomlinson, J. Chromatogr. 113 (1975) 1.
[21] T. Braumann, J. Chromatogr. 373 (1986) 191.
[22] L. Ekiert, Z. Grodzinska-Zachwieja, J. Bojarski, Chromatographia 13 (1980) 471.
[23] T. Cserháti, B. Bordás, Gy. Ösapay, Chromatographia 18 (1987) 184.
[24] T. Cserháti, Chromatographia 18 (1984) 18.
[25] T. Cserháti, Chromatographia 18 (1984) 318.
[26] K. Valkó, J. Liq. Chromatogr. 7 (1984) 1405.
[27] C. Horváth, W. Melander, I. Molnar, J. Chromatogr. 125 (1976) 129.
[28] A.J.P. Martin, R.L.M. Synge, Biochem. J. 35 (1941) 1358.
[29] K. Dross, C. Sonntag, R. Mannhold, J. Chromatogr. A 673 (1994) 113.
[30] K. Dross, C. Sonntag, J. Chromatogr. 639 (1993) 287.
[31] K. Kovács-Hadady, E. Balázs, J. Planar Chromatogr. 6 (1993) 463.
[32] G.L. Biagi, A.M. Barbaro, A. Sapone, M. Recanatini, J. Chromatogr. A 662 (1994) 341.
[33] I.D. Wilson, J. Chromatogr. 291 (1984) 241.
[34] C. Sârbu, H. Pop, Anal. Chem. 68 (1996) 771.
[35] C. Sârbu, H. Pop, Anal. Chem., summited.
[36] C. Sârbu, M. Coman, C. Măruţoiu, J. Planar Chromatogr. 4 (1991) 325.
[37] C. Sârbu, S. Todor, J. Planar Chromatogr. 11 (1998) 123.
[38] A. Mackiewicz, W. Ratajczak, Comput. Geosci. 19 (1993) 303.
[39] V. Stroescu, Pharmaceutical Fundamentals in Medical Practice, Editura Medicalã, Bucureşti, 1995, p. 1028.