Use of Information Theory and Numerical Taxonomy Methods for Evaluating the Quality of Thin-layer Chromatographic Separations of Flavonoids and Phenolic Acids of Rhamni Cathartici Fructus

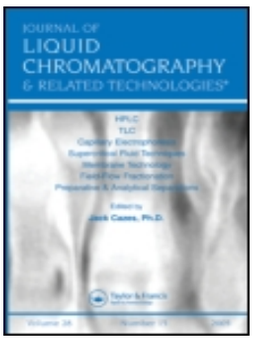
Medić-Šarić, Marica; Maleš, Željan; Šarić, Slavko; Debeljak, Željko

Source / Izvornik: **Journal of Liquid Chromatography and Related Technologies, 2006, 22, 83 - 103**

Journal article, Published version
Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

https://doi.org/10.1081/JLC-100101645

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:239:716494



Taylor & Francis
Taylor & Francis Group

Journal of Liquid Chromatography & Related Technologies

ISSN: 1082-6076 (Print) 1520-572X (Online) Journal homepage: https://www.tandfonline.com/loi/ljlc20

USE OF INFORMATION THEORY AND NUMERICAL TAXONOMY METHODS FOR EVALUATING THE QUALITY OF THIN-LAYER CHROMATOGRAPHIC SEPARATIONS OF FLAVONOIDS AND PHENOLIC ACIDS OF *RHAMNI CATHARTICI FRUCTUS*

Marica Medič-Šarić , Željan Maleš , Slavko Šarić & Željko Debeljak

To cite this article: Marica Medič-Šarić , Željan Maleš , Slavko Šarić & Željko Debeljak (1999) USE OF INFORMATION THEORY AND NUMERICAL TAXONOMY METHODS FOR EVALUATING THE QUALITY OF THIN-LAYER CHROMATOGRAPHIC SEPARATIONS OF FLAVONOIDS AND PHENOLIC ACIDS OF *RHAMNI CATHARTICI FRUCTUS* , Journal of Liquid Chromatography & Related Technologies, 22:1, 83-103, DOI: 10.1081/JLC-100101645

To link to this article: https://doi.org/10.1081/JLC-100101645

J. LIQ. CHROM. & REL. TECHNOL., 22(1), 83–103 (1999)

USE OF INFORMATION THEORY AND NUMERICAL TAXONOMY METHODS FOR EVALUATING THE QUALITY OF THIN-LAYER CHROMATOGRAPHIC SEPARATIONS OF FLAVONOIDS AND PHENOLIC ACIDS OF *RHAMNI CATHARTICI FRUCTUS*

Marica Medić-Šarić,[1,*] Željan Maleš,[2] Slavko Šarić,[3] Željko Debeljak[1]

[1] Department of Pharmaceutical Chemistry
Faculty of Pharmacy & Biochemistry
University of Zagreb
A. Kovačića 1
10000 Zagreb, Croatia

[2] Department of Pharmacognosy
Faculty of Pharmacy and Biochemistry
University of Zagreb,
Marulićev trg 20
10000 Zagreb, Croatia

[3] Faculty of Traffic Science
University of Zagreb,
Vukelićeva 4
10000 Zagreb, Croatia

ABSTRACT

A rational selection of a restricted set from fifteen available chromatographic systems for the separation of flavonoids and phenolic acids identified in the methanolic extract of *Rhamni cathartici fructus* is discussed. Series of mathematical techniques for the evaluation of solvents and solvent combinations in thin-layer chromatography of flavonoids and phenolic acids have been investigated. The chromatographic systems are classified according to their mutual resemblance by numerical taxonomy techniques. The selection criterion in the groups, obtained by numerical taxonomy classification, is the information content or discriminating power. The numerical taxonomic and information theoretical selection procedures are compared and their respective advantages and disadvantages discussed.

INTRODUCTION

Rhamnus catharticus L. (Buckthorn) is a plant belonging to the family Rhamnaceae, of widespread distribution in Europe, North Africa, and Asia. It is a bush that grows up to 3 m tall with opposite, finely serrated leaves, on branches often terminating in a spine. Small, yellowish-green, auxillary flowers are arranged in cymes.

Fruits are shiny and black when ripe.[1-3] The drug (*Rhamni cathartici fructus*) is used in constipation as a laxative acting on the large intestine. In folk medicine, the drug is also used as a diuretic.[3] The drug contains a number of constituents, including anthraquinone glycosides, flavonoids, phenolic acids, pectins, saccharides, and ascorbic acid.[3-5] Flavonoids and phenolic acids from this drug have some influence on its pharmacological effects.

Thin-layer chromatography (TLC) is an ideal technique for the screening of drugs, because of low cost, easy maintenance, and selectivity of detection reagents. A rather common problem in chromatography is to find objective criteria for the evaluation of the most efficient chromatographic system and an optimal choice of combinations to identify the group of compounds. Numerical taxonomy has been used to classify chromatographic systems according to their similarities,[6,7] but the measurement of the informing power[8-11] or discriminating power[12-15] is more useful when the selection of optimal systems is required. These approaches were compared by applying computer search programs (KT1)[15] on TLC data of the flavonoids and phenolic acids identified in the methanolic extract of *Rhamni cathartici fructus.*

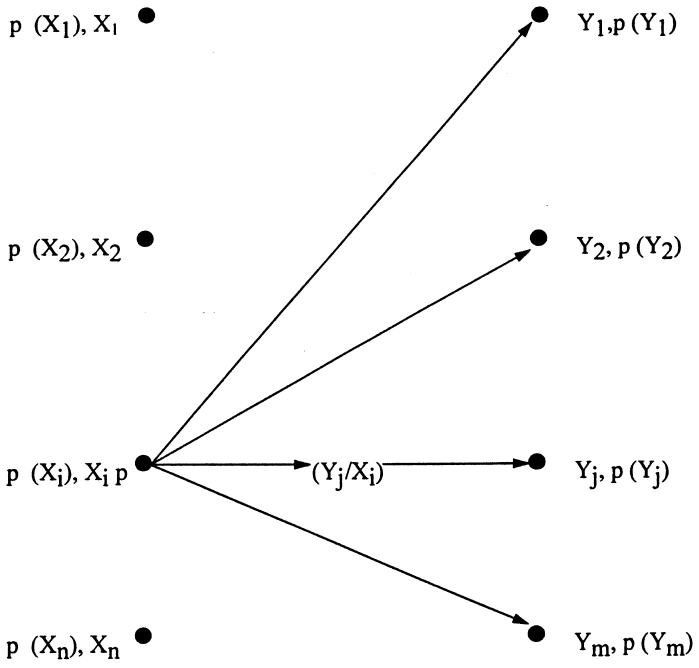


Figure 1. The relationship between output and input symbols.

MATHEMATICAL BACKGROUND

Mathematical methods described in the text below find their use in many area of science. They provide means for classification of "similar" objects (for example, classification of plants, genes, diseases, etc.). Here, we apply them to analytical methods and to the methods of identification of bioactive compounds as evaluation tools (quality measures).

Calculation of the Information Contents

The generating of information can be considered as the reduction of uncertainty with respect to the composition or identity of the sample to be analyzed.[11]

It implies that any uncertainty remaining after analysis can be treated as a parameter for evaluation of the analytical results. Let us assume that X_i (i = 1,2,...n) is a set of possible inputs, and Y_j (j = 1,2,...m) a set of outputs. The X_i

stands for analyzed compounds and the Y_j represents any chromatographic parameter (*e.g.*, R_F value in TLC). The relationship between these input and output symbols is illustrated in the Fig. 1.

The uncertainty, before analysis, with respect to the identity of the unknown compound, is described by *a priori* probability $p\ (X_i)$. Symbols X_i and Y_j are not associated one-to-one due to the possible experimental errors and/or to imperfections in the procedure of translating the measurement into analytical results. Therefore, it is more justified to talk about the probability of the simultaneous occurrence of a pair (X_i, Y_j). For a given possible output Y_j, we say that there is *a priori* probability $p\ (X_i)$ with respect to an event X_i. The uncertainty of the possible outcome is characterized by *a posteriori* probability $p\ (X_i, Y_j)$.[8,16,17]

Determination of Discriminating Power (DP)

Two compounds are chromatographically similar if the differences in their identification values do not exceed the error factor *E*. Identification values of an ideal chromatographic system are rectangularly distributed, and the probability, P_S, of finding two chromatographically similar compounds is:[12]

$$P_S = 2E - E^2 \tag{1}$$

The maximum *DP* of any chromatographic system can be derived from the relation:

$$DP = 1 - 2E + E^2 \tag{2}$$

If there is a lack of correlation between chromatographic systems, than the *DP* value for *k* systems equals:

$$DP_k = 1 - \prod_{i=1}^{k} (1 - DP_i) \tag{3}$$

and if all distributions are rectangular then:

$$DP_k = 1 - \prod_{i=1}^{k} (2E_i - E_i^{\,2}) \tag{4}$$

The effectiveness of chromatographic systems is most usefully expressed by a single number, *i.e.*, a calculated *DP*.[18,19] In case of a large number of compounds, complete identification is rather difficult. Thus, calculating and

maximizing the *DP* values can be more readily achieved by satisfying the following conditions:

a) a rectangular distribution of R_F values
b) reproducibility of results
c) no correlations between chromatographic systems.

Taxonomic Distances and Cluster Formation

Taxonomy is defined as the theoretical study of classification, including its basic principles, procedures, and rules.[6] Numerical taxonomy deals with the ways of classifying operational taxonomic units (OTU) into taxonomic groups based on the characteristic values of OTU. Input data are given in matrix form (*N* x *t*), where *N* is the number of properties (in our special case the number of compounds in consideration) and *t* is the number of OTUs (in our case the number of chromatographic systems).

Property	OTUs		
	1	2…	*t*
1	$X_{1,1}$	$X_{1,2}$	$X_{,1,t}$
2	$X_{2,1}$	$X_{2,2}$	$X_{2,t}$
3	$X_{3,1}$	$X_{3,2}$	$X_{3,t}$
N	$X_{N,1}$	$X_{N,2}$	$X_{N,t}$

Relationships between properties and OTUs can be presented in an *A*- or an *I*-space. An *A*-space is an *N*-dimensional space with the points in the space (1,2...*t*) representing OTUs. An *I*-space is a *t*-dimensional space whose points represent properties (1,2...*N*). Classification is carried out with respect to resemblances between OTUs.

Dissimilarity, expressed as the complement of similarity, is proportional to OTU-distances in the given metric space. The measure of dissimilarity, or any function which complements similarity coefficients, must satisfy the metric properties specified by four basic axioms. Assuming φ (a real non-negative number) is a function describing the measure of dissimilarity for any pair of OTUs, these axioms can be formulated as:

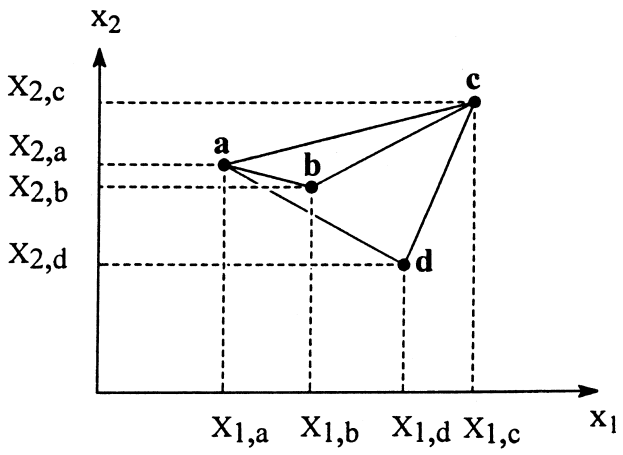


Figure 2. Representation of four OTUs (*a,b,c,* and *d*) as points on a plane determinated by their character states for two characters 1 and 2. Each character is represented by two dimensions X_1 and X_2. $a = (X_{1,a}; X_{2,a})$; $b = (X_{1,b}; X_{2,b})$; $c = (X_{1,c}; X_{2,c})$; $d = (X_{1,d}; X_{2,d})$.

1. $\varphi(a,b) \geq 0$ and $\varphi(a,a) = \varphi(b,b) = 0$ (5)
2. $\varphi(a,b) = \varphi(b,a)$ (6)
3. $\varphi(a,c) \leq \varphi(a,b) + \varphi(b,c)$ (7)
4. for $a \neq b \Rightarrow \varphi(a,b) > 0$ (8)

$$\Delta^2_{b,d} = (X_{1,b} - X_{1,d})^2 + (X_{2,b} - X_{2,d})^2 \quad (9)$$

$a = (X_{1,a}; X_{2,a})$ $\quad$ $c = (X_{1,c}; X_{2,c})$
$b = (X_{1,b}; X_{2,b})$ $\quad$ $d = (X_{1,d}; X_{2,d})$

$$\Delta^2_{b,d} = (X_{1,b} - X_{1,d})^2 + (X_{2,b} - X_{2,d})^2 + (X_{3,b} - X_{3,d})^2 \quad (10)$$

$a = (X_{1,a}; X_{2,a}; X_{3,a})$ $\quad$ $c = (X_{1,c}; X_{2,c}; X_{3,c})$
$b = (X_{1,b}; X_{2,b}; X_{3,b})$ $\quad$ $d = (X_{1,d}; X_{2,d}; X_{3,d})$

A system with properties N = 2 and OTUs t = 4 (*a, b, c, d*) in an A-space is presented in Fig. 2. Abscissa (X_1) stands for a property 1 and ordinate (X_2) stands for a property 2. Two OTUs with similar values (properties) are

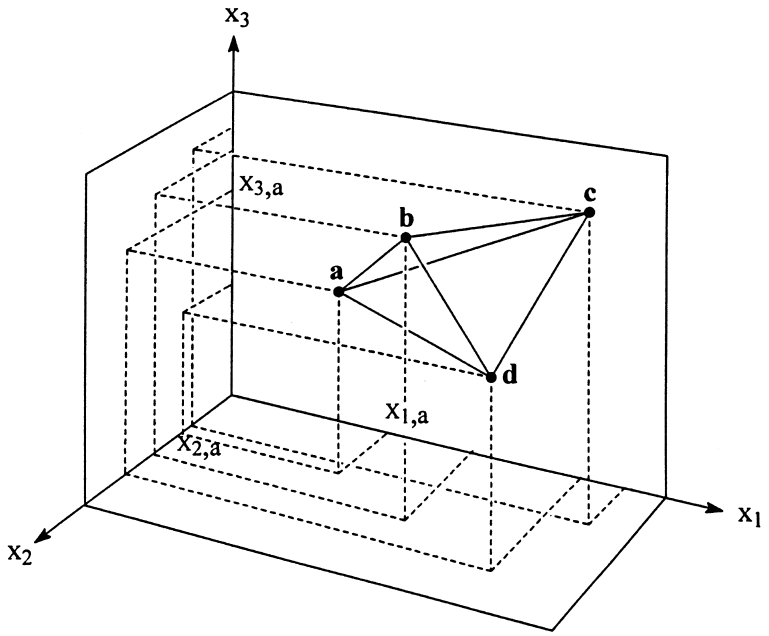


Figure 3. Representation of four OTUs (*a*,*b*,*c*, and *d*) as points on a plane in a three dimensional space. $a = (X_{1,a}; X_{2,a}; X_{3,a})$; $b = (X_{1,b}; X_{2,b}; X_{3,b})$; $c = (X_{1,c}; X_{2,c}; X_{3,c})$; $d = (X_{1,d}; X_{2,d}; X_{3,d})$.

associated with only one point in the space (taxonomic distance equals zero). The greater the differences in properties, the larger their spatial distances. In other words, taxonomic distance is inversely related to similarity. Generally, a distance $d_{j,k}$ between OTUs *j* and *k* in a *N*-dimensional space is equal to:

$$d_{j,k} = \sqrt{\sum_{i=1}^{N} (X_{i,j} - X_{i,k})^2} \qquad (11)$$

and the taxonomic distance is:

$$\Delta_{j,k} = \sqrt{\frac{d_{j,k}^{\ 2}}{N}} \qquad (12)$$

It can be easily shown that taxonomic distance satisfy equations [5] - [8]. From the initial matrix (*N* x *t*), the matrix containing similarity coefficients or their

complements showing dissimilarities proportional to taxonomic distances is obtained. Let us take $U_{j,k}$ as a general symbol for dissimilarity and $\Delta_{j,k}$ for a taxonomic distance. Let clusters *J*, *K* and *L* contain $t_J \geq 1$, $t_K \geq 1$ and $t_L \geq 1$ OTUs. If clusters *J* and *K* form a joint cluster, then we have to find dissimilarity (distances) between joint cluster and a candidate for joining *L*. The joint cluster (*J*, *K*) contains $t_{(J,K)} = t_J + t_K$ OTUs.

The coefficient $U_{(J,K)L}$ can be determined both by a recursive and a direct procedure. Applying the recursive procedure, $U_{(J,K)L}$ is calculated from the coefficients $U_{J,L}$, $U_{K,L}$ and $U_{J,K}$, which for the weighted pair group method, are expressed as:

$$U_{(J,K),L} = \frac{U_{J,L} + U_{K,L}}{2} \tag{13}$$

The major advantage of this procedure is that there is no need to store the starting matrix but only the result of the last classification step, thus reducing the computer memory needed. For the direct procedure the following equation holds:

$$U_{J,K} = \sum_{j,k} W_J W_k U_{j,k} \tag{14}$$

where $W_j = (1/2)^{Cj}$ and C_j stands for steps involved in cluster formation and OTUs. This procedure requires memorizing the starting values $U_{J,K}$. The methods mentioned above were presented in our earlier publications.[20-24]

EXPERIMENTAL

Materials

Extract solution: 1.0 g air-dried, powdered fruits of *Rhamnus catharticus L.* (*Rhamni cathartici fructus*) was refluxed with 10.0 mL methanol for 30 min, filtered, the filtrate concentrated under reduced pressure, and the residue taken up in 5.0 mL methanol.[25]

Reference solution : 10 mg kaempferol, 10 mg quercetin, 10 mg quercitrin and 10 mg hyperoside dissolved in 10.0 mL methanol. The fifteen systems used are given in Table 1.[25-33]

Table 1

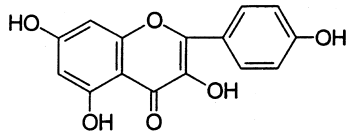
The Thin-Layer Chromatographic Systems Studied

System No.	Solvent	Ref.
1	Ethyl acetate:formic acid:acetic acid:water (100:11:11:27)	25
2	Ethyl acetate:formic acid:water (8:1:1)	26
3	Ethyl acetate:formic acid:water (65:15:20)	27
4	Ethyl acetate:formic acid:water (67:20:13)	28
5	Ethyl acetate:formic acid:water (88:6:6)	29
6	Ethyl acetate:formic acid:water (30:2:3)	30
7	Ethyl acetate:methyl ethyl ketone:formic acid:water (50:30:10:10)	31
8	Ethyl acetate:methyl ethyl ketone:formic acid:water (50:30:30:10)	27
9	Ethyl acetate:formic acid:acetic acid:methyl ethyl ketone:water (50:7:3:30:10)	25
10	Ethyl acetate:1-propanol:water:formic acid (40:40:28:2)	28
11	Ethyl acetate:methanol:water (77:13:10)	32
12	1-Butanol:acetic acid:water (66:17:17)	28
13	1-Butanol:acetic acid:water-upper phase (40:10:50)	25
14	1-Butanol:acetic acid:water (12:3:5)	33
15	1-Butanol:acetic acid:water (4:1:2)	33

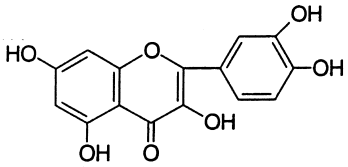
In all systems, silica gel plates (20x20 cm, 0.25 mm thickness) incorporating a fluorescent indicator, kieselgel 60 F_{254} -Alufolien (E. Merck, Darmstadt, Art. Nr. 5554) were used. Paper liners were used in all tanks, and, after addition of the appropriate solvents, the systems were allowed to equilibrate for at least 30 minutes.

Five μL of the extract solution and of the reference solution were applied to the plates, and the systems were allowed to develop for 15 cm.

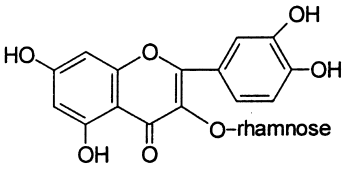
Visualization of the flavonoids and phenolic acids was obtained by spraying the sheets with 1 % methanolic diphenylboryloxyethylamine, followed by 5 % ethanolic polyethylene glycol 4000. The chromatograms were viewed in UV 366 nm light (flavonoids as orange-yellow and phenolic acids as blue fluorescent bands).[25] The structures of the flavonoids identified in the methanolic extract of *Rhamni cathartici fructus* are presented in Fig. 4.



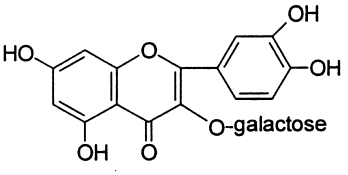
Kaempferol



Quercetin



Quercitrin (quercetin 3-O-rhamnoside)



Hyperoside (quercetin 3-O-galactoside)

Figure 4. Structures of the flavonoids identified.

Calculation of the Information Content

Extensive information has been calculated for fifteen TLC systems by Shannon's formula. Calculation of the information content will become possible if the uncertainties before and after the analysis can be expressed in a quantitative way.

Distribution of R_F values into groups with error factor E (*e.g.*, $E = 0.05$ or $E = 0.10$) with respect to R_F units and the assumption of n_k R_F values in the k^{th} group, the average information content (entropy) is given by the following Shannon equation:[11,14]

$$I(X) = H(X) = -\sum_k \frac{n_k}{n} ld \frac{n_k}{n} [bit] \qquad (15)$$

It is assumed that the compounds with R_F values within one group cannot be identified. The system with the highest informational content makes the best solution for the differentiation of compounds that were considered.

It is obvious that the entropy is at its highest level, i.e., $H_{max}(X) = ld\ n$ ($n = \Sigma n_k$) if there is only one R_F value within each group.

Determination of Discriminating Power (DP)

The *DP* of a set of chromatographic systems is defined as the probability of identifying two randomly selected compounds in at least one of the systems.[12,13,18,19] It must be possible to discriminate all pairs of *N* in order to compute the *DP* of *k* chromatographic systems in which *N* compounds are investigated. For the total number of matching pairs (*M*), the probability of a random selection of chromatographically similar pairs is $2M/N(N\text{-}1)$. Therefore, the *DP* of *k* systems is:

$$DP_k = 1 - \frac{2M}{N(N-1)} \qquad (16)$$

The average number of chromatographically similar compounds (*T*) for the chromatographic systems considered can be calculated from the following equation:[12]

$$T = 1 + (N-1)(1 - DP_k) \qquad (17)$$

Computation of Taxonomic Distances, Cluster Formation, and Dendrogram

The optimal combination of two or more chromatographic systems for the identification of a compound by TLC can be readily determined from the taxonomic distances.[7] The cluster formation is carried out in three steps:

Step 1. *Entering data*. Identification characteristics (in our case R_F values) are entered into a matrix (*N* x *ISYST*).

Step 2. *Determination and evaluation of similarities of chromatographic systems*. Similarity is determined from taxonomic distances. For two chromatographic systems, *k* and *l*, taxonomic distance is formulated as:

$$\Delta_{k,l} = \sqrt{\sum_{i=1}^{N} \frac{(R_{F_{i,k}} - R_{F_{i,l}})^2}{N}} \qquad (18)$$

Normalization (i.e., dividing the sums of squares by *N*) allows chromatographic systems with some unknown *N* values to be included. If there are *m* unknown within chromatographic systems *k* and *l*, the denominator changes its value from *N* to *N-m*. Having determined all distances $\Delta_{k,l}$ (k = 1, 2... *ISYST* - 1; l = 2,3... *ISYST*), the resemblance matrix (ISYST x *ISYST*) is constructed:

	1	2	3	4	5	
1	0	$\Delta_{1,2}$	$\Delta_{1,3}$	$\Delta_{1,4}$	$\Delta_{1,5}$	$\Delta_{1,\ ISYST}$
2		0	$\Delta_{2,3}$	$\Delta_{2,4}$	$\Delta_{2,5}$	$\Delta_{2,\ ISYST}$
3			0	$\Delta_{3,4}$	$\Delta_{3,5}$	$\Delta_{3,\ ISYST}$
4				0	$\Delta_{4,5}$	$\Delta_{4,\ ISYST}$
5					0	$\Delta_{5,\ ISYST}$
ISYST						0

Step 3. *Classification*. Chromatographic systems with high degree of resemblance are grouped into clusters. Cluster formation in this work was carried out by a weighted pair group method. First of all, one should find the smallest $\Delta_{k,l}$. Then, if three is $\Delta_{q,p}$ the chromatographic systems with $j = q$ and

$j = p$ are the most similar and they form a cluster p'. Following this step, the matrix is reduced (elimination of the q-th row and the q-th column) where $\Delta_{j,p'} = (\Delta_{j,p'} + \Delta_{j,q})/2$ (the new structure is sometimes called centroid). All other values $\Delta_{k,l}$ remain unchanged, and the reduced matrix has the following form:

	1’	2	4	5	.....	ISYST
1’	0	$\Delta_{1,2}$	$\Delta_{1',4}$	$\Delta_{1',5}$	...	$\Delta_{1'}$, ISYST
2		0	$\Delta_{2,4}$	$\Delta_{2,5}$	...	Δ_{2}, ISYST
4			0	$\Delta_{4,5}$	...	Δ_{4}, ISYST
5				0	...	Δ_{5}, ISYST
ISYST						0

Step 3 is repeated as many times as necessary for all chromatographic systems to form a common cluster. The procedure for cluster formation is presented by dendrogram in Fig. 5. For the example analyzed here, the smallest $\Delta_{k,l}$ in the resemblance matrix is $\Delta_{1,3} = 2{,}3274$. Systems 1 and 3 both fall into cluster 1'. The 3rd column and the 3rd row are omitted from the resemblance matrix and $\Delta_{1',j}$ is calculated according to the above expression; $\Delta_{mink,l}$ is determined by the same procedure.

RESULTS AND DISCUSSION

A data set of R_F values for the separation of flavonoids and phenolic acids (Fig.4.) of a methanolic extract of *Rhamni cathartici fructus* in fifteen different chromatographic systems, shown in Table 1, was analyzed. An optimal combination of two or more systems was selected using the following procedures:

a) determination and comparison of the amount of information and discriminating power for all possible combinations of chromatographic systems.
b) classification of chromatographic systems into groups with similar separation properties and selection of the most efficient chromatographic system from each group.

Table 2

Input Data: R_F Values of the Components of *Rhamni cathartici fructus*

Solvent System*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
t (min):	55	35	62	50	35	33	34	45	45	73	45	90	95	90	90
Compound	**R_F Values:**														
Kaempferol	0.99	0.98	0.99	0.99	0.97	0.98	0.99	0.99	0.99	0.97	0.94	0.87	0.87	0.92	0.90
Quercetin	0.98	0.96	0.98	0.98	0.94	0.95	0.98	0.97	0.98	0.94	0.91	0.82	0.85	0.88	0.85
Quercitrin	0.93	0.90	0.91	0.94	0.61	0.64	0.84	0.91	0.89	0.90	0.70	0.70	0.71	0.75	0.73
Phenolic acid 1	0.90	0.85	0.87	0.90	0.57	0.57	0.72	0.87	0.84	0.87	0.69	0.67	0.65	0.70	0.68
Flavonoid 1	0.82	0.72	0.83	0.88	0.47	0.49	0.69	0.87	0.81	0.87	0.65	0.67	0.62	0.69	0.68
Phenolic acid 2	0.65	0.49	0.78	0.85	0.37	0.37	0.65	0.85	0.78	0.84	0.55	0.66	0.62	0.68	0.66
Hyperoside	0.60	0.45	0.74	0.83	0.34	0.29	0.62	0.85	0.73	0.84	0.52	0.66	0.60	0.64	0.64
Flavonoid 2	0.50	0.38	0.71	0.78	0.27	0.23	0.57	0.83	0.62	0.81	0.48	0.63	0.55	0.63	0.61
Flavonoid 3	0.43	0.31	0.65	0.73	0.18	0.17	0.51	0.78	0.54	0.76	0.44	0.57	0.53	0.60	0.57
Flavonoid 4	0.33	0.19	0.54	0.61	0.11	0.11	0.38	0.69	0.38	0.66	0.34	0.51	0.44	0.52	0.49
Flavonoid 5	0.26	0.14	0.47	0.49	0.05	0.06	0.32	0.63	0.24	0.61	0.28	0.47	0.38	0.47	0.44

* Copies of chromatograms can be obtained from the authors on request.

Table 3

Output Data for DP and I in a Range of Error Factors for Each Chromatographic System

	E = 0.05		E = 0.10	
TLC-System	**DP**	**I (bit)**	**DP**	**I (bit)**
1	0.9455	3.278	0.8000	2.664
2	0.9273	3.278	0.8545	2.550
3	0.9091	3.096	0.7273	2.413
4	0.8545	2.914	0.6727	2.118
5	0.9455	3.459	0.8182	2.914
6	0.9636	3.459	0.8545	2.914
7	0.9273	3.096	0.8182	2.482
8	0.7636	2.664	0.5455	1.790
9	0.9273	3.096	0.8182	2.664
10	0.7818	2.595	0.5455	1.686
11	0.8909	2.914	0.7818	2.482
12	0.7273	2.404	0.6000	1.686
13	0.8545	2.845	0.6364	2.413
14	0.8364	2.845	0.6182	2.222
15	0.8364	2.845	0.6364	2.040

Table 2 gives the input data for the investigated compounds. Table 3 gives output data for the discriminating power and the information content for each TLC system. Table 4 gives output data for combined systems k = 2 and k = 3 in a range of error factors. The error factors were 0.05 and 0.10, respectively.

Under the conditions most frequently used in chromatographic analysis, i.e., E = 0.05, the most suitable system for separating the compounds studied is the chromatographic system 6 (ethyl acetate: formic acid : water, 30:2:3) because it showed the largest discriminating power and information content (DP = 0.9636, I = 3.459).

Chromatographic system 5 (ethyl acetate : formic acid : water, 88:6:6) is also suitable because of its slightly lower discriminating power (DP = 0.9455) and identical information content (I = 3.459) compared to system 6. At E = 0.10, chromatographic system 6 seems to be the most appropriate due to its largest discriminating power and information content (DP = 0.8545, I = 2.914). Chromatographic system 5 is also suitable because it has identical information content and slightly lower discriminating power compared to system 6 (DP = 0.8182, I = 2.914).

Table 4

Output Data for DP and T for Combined Solvent Systems

a) Combined Solvent Systems - k = 2

	E = 0.05			E = 0.10		
Combination Sequence	**Solvents**	**DP**	**T**	**Solvents**	**DP**	**T**
1.	6-15	0.9818	1.182	6-9	0.9091	1.909
2.	6-13	0.9818	1.182	2-9	0.9091	1.909
3.	6-12	0.9818	1.182	2-7	0.9091	1.909
4.	6-11	0.9818	1.182	1-7	0.9091	1.909
5.	6-10	0.9818	1.182	6-7	0.8909	2.091
6.	6-9	0.9818	1.182	5-9	0.8909	2.091
7.	6-8	0.9818	1.182	4-6	0.8909	2.091
8.	6-7	0.9818	1.182	2-15	0.8909	2.091
9.	5-6	0.9818	1.182	2-14	0.8909	2.091
10.	4-6	0.9818	1.182	2-13	0.8909	2.091

b) Combined Solvent Systems - k = 3

	E = 0.05			E = 0.10		
Combination Sequence	**Solvents**	**DP**	**T**	**Solvents**	**DP**	**T**
1.	6-12-15	1.0000	1.000	2-7-9	0.9455	1.545
2.	6-12-13	1.0000	1.000	6-7-9	0.9273	1.727
3.	6-11-12	1.0000	1.000	2-9-15	0.9273	1.727
4.	6-10-12	1.0000	1.000	2-9-14	0.9273	1.727
5.	6-9-12	1.0000	1.000	2-9-13	0.9273	1.727
6.	6-8-12	1.0000	1.000	2-9-12	0.9273	1.727
7.	6-7-12	1.0000	1.000	2-9-11	0.9273	1.727
8.	5-6-12	1.0000	1.000	2-6-9	0.9273	1.727
9.	4-6-12	1.0000	1.000	2-5-9	0.9273	1.727
10.	3-6-12	1.0000	1.000	2-4-7	0.9273	1.727

Combining two chromatographic systems with the error factor E = 0.05, all the systems have the identical discriminating power (DP = 0.9818). This is shown in Table 4a. The number of compounds with similar chromatographic properties is at a minimum (T = 1.182). At this error factor, system 6 turned out to be the best, because it was included in all of the first ten combinations. At E = 0.10 system 6 is the best again, because it is in the first combination (DP = 0.9091, T = 1.909).

Applying the combinations of three TLC systems (Table 4b) at E= 0.05, all combination sequences have the maximal discriminating power (DP = 1.0000) and minimal number of chromatographycally similar compounds (T = 1.000). Consequently, there is no need to conduct three chromatographic experiments

Table 5

Formation of Clusters

	Cluster Formation		
Cluster	**Solvent**	**Solvent**	**Distance**
1.	8	10	0.0183
2.	11	14	0.0213
3.	5	6	0.0230
4.	10	12	0.0277
5.	4	7	0.0507
6.	9	10	0.0552
7.	3	4	0.0718
8.	5	7	0.0773
9.	1	6	0.0789
10.	5	6	0.0876
11.	1	5	0.1220
12.	1	2	0.1507
13.	1	3	0.2098
14.	1	2	0.3070

because, in each of three proposed combinations, all of the compounds[11] can be positively identified. Again, solvent system 6 is in all combinations. By cluster formation, shown in Table 5, and from the dendrogram, presented in Fig. 5., the same results were obtained. In order to obtain the optimal combination of two chromatographic systems according to the dendrogram (Fig. 5.), one should choose one system from cluster 3 [(system 5 and 6); although both chromatographic systems have the same amount of information (I = 3.459), system 6 is better because it has a higher DP value (DP = 0.9636)] and one chromatographic system from cluster 6 (systems 12-15).

CONCLUSIONS

The use of information content, discriminating power, and numerical taxonomy methods have been proposed for the evaluation of single systems or combinations of systems for TLC separation of flavonoids and phenolic acids of a methanolic extract of *Rhamni cathartici fructus*.

This article investigates and compares these existing mathematical selection procedures. The question is which combination of two or three different solvents must be chosen from a large number of possibilities, so that

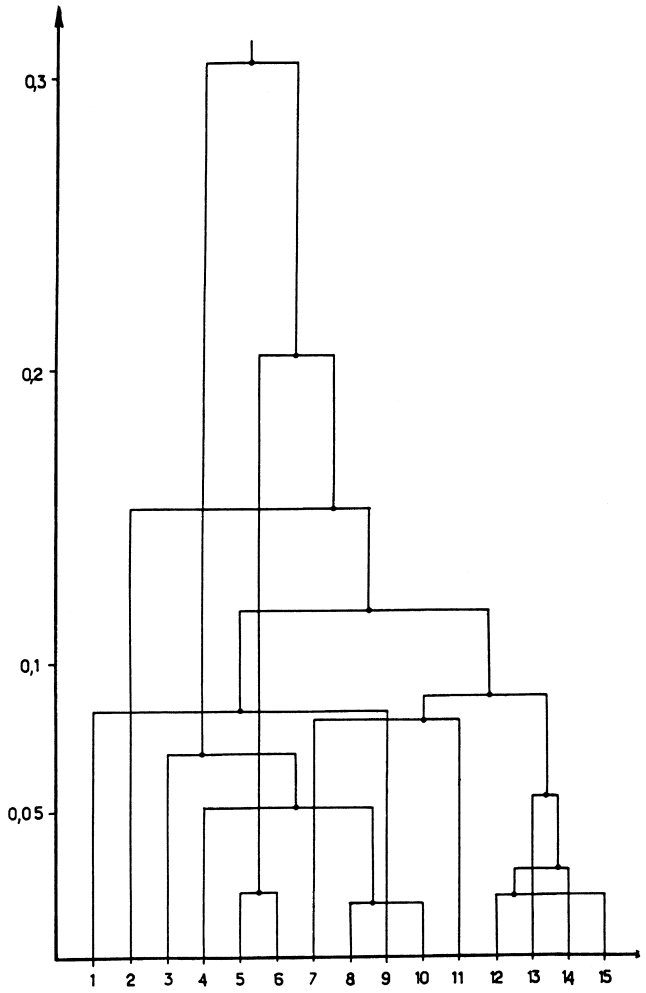


Figure 5. Dendrogram for fifteen TLC systems.

the combination yields as much information as possible. The best chromatographic system is shown to be system 6, but system 5 is almost as effective. These two systems belong also to the same cluster, which indicates their similarity, with improved results obtained by calculation of the discrimination power and information content. These mathematical techniques can be also applied to other classification problems in analytical chemistry and biomedicine.

List of Symbols

Ps - probability of finding two chromatographicaly similar compounds with rectangular distribution
DP_k - Discriminating power of analyzed systems divided into groups containing *k* elements
OTU - Operational Taxonomic Unit N-number of properties
t - number of OTUs
$X_{i,j}$ - value of i^{th} OTU in j^{th} property
φ - dissimilarity function
d_{jk} - distance between j^{th} and k^{th} OTU
$\Delta_{j,k}$ - normed distance between j^{th} and k^{th} OTU
$U_{j,k}$ - dissimilarity coefficient between j and k OTUs
C_j - the number of steps involved in cluster formation
n_k - the number of compounds whose R_F is in k^{th} R_F interval
n - total number of compounds
I(X) - average information content
H(X) - entropy
H_{max} - maximal value of entropy
T - average number of chromatographicaly similar compounds
ISYST - the number of examined chromatographic system
X_i - i^{th} analyzed compound
Y_j - j^{th} chromatographic property in examination
R_F - retention value
$p(X_i)$ - *a priori* probability of X_i
$p(X_i/Y_j)$ - *a posteriori* probabilityof X_i
E_i - error of measurement of R_F of i^{th} compound (resolution)

REFERENCES

1. G. Hegi. *Illustrierte Flora von Mittel-Europa, Band V/I.* München: Carl Hanser Verlag (1954).

2. R. Hegnauer. *Chemotaxonomie der Pflanzen, Band VI.* Basel und Stuttgart: Birkhäuser Verlag (1973).

3. M. Wichtl. *Herbal Drugs and Phytopharmaceuticals*. Stuttgart: Medpharm, Scientific Publishers (1994).

4. E. Steinegger, R. Hänsel. *Lehrbuch der Pharmakognosie und Phyto-pharmazie.* Berlin: Springer Verlag (1988).

5. H. Wagner. *Pharmazeutische Biologie-Drogen und ihre Inhaltsstoffe*. Stuttgart, New York: Gustav Fischer Verlag (1988).

6. P. H. A. Sneath, R. R. Sokal. *Numerical Taxonomy.* San Francisco: W. H. Freeman and Co. (1973).

7. D. L. Massart, H. De Clercq. *Anal. Chem*., **46**, 1988-1992 (1974).

8. P. Cleij, A. Dijkstra. *Fresenius Z. Anal. Chem*., **298**, 97-109 (1979).

9. D. L. Massart. *J. Chromatogr*., **79**, 157-163 (1973).

10. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, L. Kaufman. *Chemometrics.* Amsterdam: Elsevier (1988).

11. G. J. Chaitin. *Algorithmic Information Theory.* Cambridge: Cambridge University Press (1987).

12. A. C. Moffat, K. W. Smalldon, C. Brown. *J. Chromatogr*., **90**, 1-7 (1974).

13. A. C. Moffat, K. W. Smalldon. *J. Chromatogr*., **90**, 9-17 (1974).

14. H. de Clercq, D. L. Massart. *J. Chromatogr*., **115**, 1-7 (1975).

15. M. Medić-Šarić, S. Šarić, D. Maysinger. *Acta Pharm. Jugosl*., **39**, 1-16 (1989).

16. H. Kaiser. *Anal. Chem*., **42**, 24A (1970).

17. F. Dupuis, A. Dijkstra. *Anal. Chem*., **47**, 379-383 (1975).

18. P. Owen, A. Pendlebury, A. C. Moffat. *J. Chromatogr*., **161**, 187-193 (1978).

19. R. E. Kaiser, ed. *Planar Chromatography., Vol. 1.* Heidelberg: Alfred Huethig Verlag (1986).

20. A. Rotar, F. Kozjek, M. Medić-Šarić. *Acta Pharm*., **43**, 157-165 (1993).

21. Ž. Maleš, M. Medić-Šarić, D. Kuštrak. *Acta Pharm*., **44**, 183-191 (1994).

22. M. Medić-Šarić, A. Brantner, Ž. Maleš. *Acta Pharm*., **46**, 115-124 (1996).

23. M. Medić-Šarić, Ž. Maleš, G. Stanić, S. Šarić. *Croat. Chem. Acta*, 69, 1265-1274 (1996).

24. M. Medić-Šarić, G. Stanić, Ž. Maleš, S. Šarić. *J. Chromatogr. A*, **776**, 355-360 (1997).

25. H. Wagner, S. Bladt, E. M. Zgainski. *Drogenanalyse*. Berlin: Springer Verlag (1983).

26. M. Luckner, O. Bessler, R. Luckner. *Pharmazie*, **20**, 681-685 (1965).

27. G. Willuhn, P. M. Röttger. *Dtsch. Apoth. Ztg*., **120**, 1039-1042 (1980).

28. M. Wichtl, B. Bozek, T. Fingerhut. *Dtsch. Apoth. Ztg*., **127**, 509-510(1987).

29. M. Wichtl. *Teedrogen, Ein Handbuch für die Praxis auf Wissenschaftlicher Grundlage*. Stuttgart: Wissenschaftliche Verlagsgesellschaft GmbH (1989).

30. R. Hänsel, K. Heller, H. Rimpler, G. Schneider, eds. *Hagers Handbuch der Pharmazeutischen Praxis, Vol. 5*. Berlin: Springer Verlag (1994).

31. W. Poethke, C. Schwarz, H. Gerlach. *Planta Med.*, **19**, 177-188 (1970).

32. E. Stahl. *Chromatographische und Mikroskopische Analyse von Drogen*, Stuttgart: Gustav Fischer Verlag (1970).

33. *Deutscher Arzneimittel-Codex*. Frankfurt am Main: Govi-Verlag, Pharmazeutischer Verlag and Stuttgart: Deutscher Apotheker Verlag (1986).

Received March 21, 1998
Accepted April 27, 1998
Manuscript 4824-TLC