Department of Pharmaceutical Chemistry[1] and Department of Pharmacognosy[2], Faculty of Pharmacy and Biochemistry, University of Zagreb, Croatia

The use of information theory and numerical taxonomy methods for evaluating the quality of thin-layer chromatographic separations of flavonoid constituents of Matricariae flos

M. Medić-Šarić[1], G. Stanić[2] and I. Bošnjak[2]

A methanol extract of Matricariae flos was analysed with regard to the flavonoid composition. Rational selection of a restricted set from eight chromatographic systems tested for the separation of these compounds is discussed. Series of mathematical techniques for the evaluation of solvents and solvent combinations in thin-layer chromatography of flavonoid constituents have been explored. 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 the discriminating power. The most suitable mobile phases for TLC separation of flavonoid constituents of Matricariae flos are: ethylacetate-methanol-water (75 : 15 : 0 v/v), ethylacetate-formic acid-water (80 : 10 : 10 v/v) and ethylacetate-formic acid-acetic acid-water (100 : 11 : 11 : 27 v/v).

1. Introduction

Matricariae flos (*Chamomilla recutita* (L.) Rauschert) is a well known plant remedy with antiphlogistic and spasmolytic properties. While the first property is attributed primarily to the essential oil components, its flavonoid constituents are responsible for the spasmolytic activity of chamomile flowers. By the end of the 1970s approximately thirty flavonoid glycosides or aglycones had been identified as chamomile constituents [1–7]. A number of papers have documented both the spasmolytic [8–12] and antiphlogistic activity [13, 14] of chamomile extract or of particular flavonoids. From the pharmacological point of view, apigenin derivatives predominate in comparison to those of luteolin and quercetin. The interest in flavonoids as a part of the bioactive chamomile complex has resulted in development of methods for the qualitative and quantitative analysis of these compounds in the drug and phytopreparations [15–18].

However, flavonoids are not only the active principle of chamomile but can also serve as specific analytical markers in the differential TLC analysis of various drug samples and adulterants. Therefore, the selection of the most efficient chromatographic system is of great interest.

In our previous paper we reported the results of testing the efficiency of a number of TLC mobile phases for the separation of the components of chamomile essential oil [19]. Using the same methods, i.e. information theory and numerical taxonomy methods we have now tested the efficiency of mobile phases appropriate for the thin-layer chromatography of flavonoid constituents of Matricariae flos.

2. Investigations, results and discussion

The efficiency of eight mobile phases for TLC separation of chamomile flavonoids was tested using the following procedures:

- determination and comparison of discriminating power (DP) and information content (I) for each solvent system and combinations of two and three chromatographic systems.
- division of chromatographic systems into groups with similar separation properties and selection of the most efficient chromatographic systems from each group.

Table 1: Mobile phases for TLC of chamomile flavonoids

System No.	Mobile phase	Ref.
1	Ethyl acetate-methylethylketone-formic acid-water (50 : 30 : 10 : 10 v/v)	[20]
2	ethyl acetate-methanol-water (75 :15 : 10 v/v)	*
3	ethyl acetate-formic acid-water (80 : 10 : 10 v/v)	[21]
4	ethyl acetate-formic acid-water (100 : 20 : 30 v/v)	[22]
5	ethyl acetate-formic acid-acetic acid-water (100 : 11 : 11 : 27 v/v)	[23]
6	n-butanol-acetic acid-water (66 : 17 : 17 v/v)	[24]
7	ethyl acetate-methanol-formic acid-water (75 : 10 : 5 : 10 v/v)	*
8	ethyl acetate-acetic acid-water (80 : 10 : 10 v/v)	*

* own modification

Some of the mobile phases listed in Table 1 were taken from the literature [20–24] while the others were created by modifying the polarity of known systems. Table 2 presents input data for R_F values of separated substances in eight mobile phases tested.

Table 3 gives the output data of DP and I for each mobile phase for each of two error factors, $E = 0.03$ and $E = 0.05$. The selection of the most suitable mobile phases is based on the highest values of DP and I. Thus, from Table 3, mobile phases **2** (ethylacetate-methanol-water 75 : 15 : 10 v/v) and **3** (ethylacetate-formic acid-water 80 : 10 : 10 v/v) possess the best chromatographic properties ($DP_2 = 0.969$, $I_2 = 3.585$ and $DP_3 = 0.969$, $I_3 = 3.252$).

Further, for combinations of two mobile phases (Table 4a; $K = 2$; $E = 0.03$) the highest values of DP ($DP = 1.000$) and the smallest value of T ($T = 1.000$) occur with the first three combinations. Mobile phase **2** appears in each of them as well as in another four combinations. At error factor $E = 0.05$ the best values are given by mobile phases **2** and **5** (ethylacetate-formic acid-acetic acid-water 100 : 11 : 11 : 27 v/v). In addition, six combination sequences contain mobile phase **3**.

Table 2: Input data — R_F values of chamomile flavonoids and related compounds separated in mobile phases tested (1–8)

Component (R_F values)	Mobile phase							
	1	2	3	4	5	6	7	8
Herniarin	0.94	0.85	0.93	0.94	0.94	0.90	0.91	0.92
Flavonoid 1	0,89	0.71	0.83	0.90	0.88	0.82	0.84	0.69
Apigenin-7-diacetylglucoside	0.85	0.72	0.77	0.87	0.83	0.80	0.81	0.66
Apigenin-7-acetylglucoside	0.85	0.67	0.77	0.87	0.83	0.80	0.81	0.63
Flavonoid 2	0.79	0.62	0.69	0.87	0.83	0.80	0.81	0.49
Apigenin-7-glucoside	0.67	0.58	0.56	0.72	0.68	0.74	0.66	0.42
Luteolin-7-glucoside	0.60	0.55	0.48	0.64	0.59	0.69	0.58	0.34
Chlorogenic acid	0.47	0.48	0.35	0.52	0.48	0.64	0.47	0.26
Flavonoid 3	0.44	0.40	0.32	0.49	0.46	0.60	0.46	0.24
Rutin	0.42	0.36	0.28	0.46	0.43	0.56	0.42	0.20
Flavonoid 4	0.36	0.32	0.23	0.41	0.37	0.48	0.36	0.16
Flavonoid 5	0.38	0.24	0.17	0.34	0.39	0.41	0.29	0.18

Table 3: Output data of discriminating power (DP) and average information content (I) for each mobile phase

Mobile phase	Error: E = 0.03		Error: E = 0.05	
	DP	I (bit)	DP	I (bit)
1	0.954	3.085	0.924	3.085
2	0.969	3.585	0.909	3.252
3	0.969	3.252	0.939	3.252
4	0.893	3.022	0.878	2.752
5	0.924	2.855	0.893	2.855
6	0.893	2.752	0.893	2.752
7	0.893	3.022	0.893	2.752

Table 4: Output data of DP values and T values for K = 2 and K = 3

(a) combination of two mobile phases; K = 2

Combination sequence	Error: E = 0.03			Error: E = 0.05		
	Mobile phases	DP	T	Mobile phases	DP	T
1	2, 5	1.000	1.000	2, 5	0.984	1.167
2	2, 3	1.000	1.000	7, 8	0.969	1.333
3	1, 2	1.000	1.000	6, 8	0.969	1.333
4	3, 8	0.984	1.000	4, 8	0.969	1.333
5	2, 8	0.984	1.000	3, 7	0.969	1.333
6	2, 7	0.984	1.000	3, 6	0.969	1.333
7	2, 6	0.984	1.000	3, 5	0.969	1.333
8	2, 4	0.984	1.000	3, 4	0.969	1.333
9	7, 8	0.969	1.000	2, 3	0.969	1.333
10	6, 8	0.969	1.000	1, 3	0.969	1.333

(b) combination of three mobile phases; K = 3

Combination sequence	Error: E = 0.03			Error: E = 0.05		
	Mobile phases	DP	T	Mobile phases	DP	T
1	2, 5, 8	1.000	1.000	5, 7, 8	0.985	1.167
2	2, 5, 7	1.000	1.000	5, 6, 8	0.985	1.167
3	2, 5, 6	1.000	1.000	4, 5, 8	0.985	1.167
4	2, 4, 5	1.000	1.000	3, 7, 8	0.985	1.167
5	2, 3, 8	1.000	1.000	3, 6, 8	0.985	1.167
6	2, 3, 7	1.000	1.000	3, 5, 8	0.985	1.167
7	2, 3, 6	1.000	1.000	3, 4, 8	0.985	1.167
8	2, 3, 5	1.000	1.000	2, 5, 8	0.985	1.167
9	2, 3, 4	1.000	1.000	2, 5, 7	0.985	1.167
10	1, 2, 8	1.000	1.000	2, 5, 6	0.985	1.167

Table 5: Cluster formation

Cluster	Mobile phase	Mobile phase	Distance
1	1	5	0.0163
2	1	6	0.0312
3	1	4	0.0454
4	1	4	0.0872
5	2	3	0.0886
6	2	3	0.1289
7	1	2	0.2024

The list of combinations of three mobile phases with the corresponding DP and T values is presented in Table 4b. It is seen that at error factor E = 0.03 all the combination

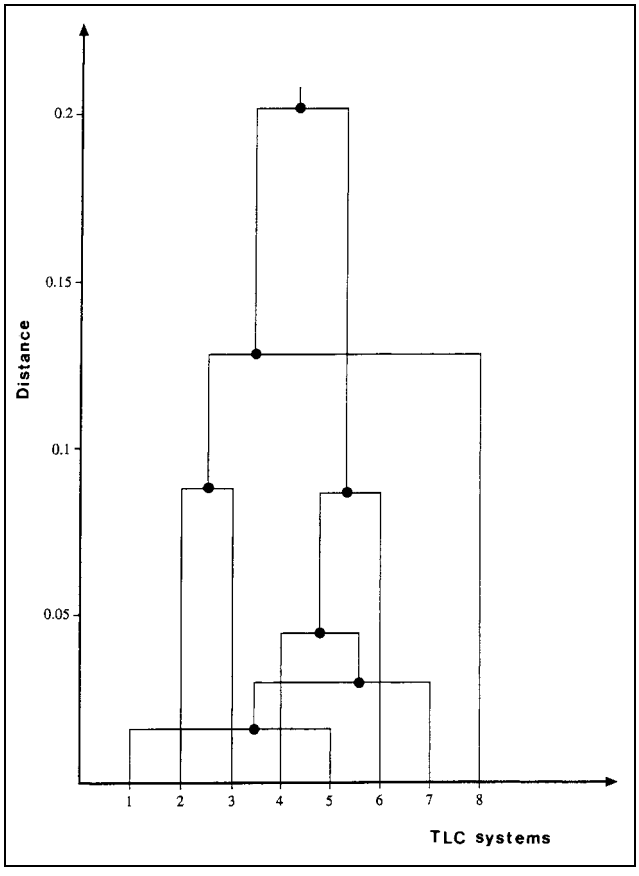

Fig.: Dendrogram for eight TLC system

sequences possess the maximum discriminating power (DP = 1.000) and the number of chromatographically similar compounds is minimal (T = 1.000). Every combination appears to contain mobile phase **2** which is almost invariably followed by mobile phases **3** or **5**. At the error factor E = 0.05, mobile phase **2** together with mobile phase **5** will be noticed in three combination sequences. Moreover, mobile phase **5** appears in another four combination sequences at K = 3.
Finally, the results obtained were confirmed by cluster analysis (Table 5) of chromatographically similar mobile phases. According to the dendrogram (Fig.) mobile phases **2** and **3** should be chosen from cluster 1 and mobile phase **5** from cluster 2 respectively.

3. Experimental

3.1. Materials

The specimen (*Matricariae flos*) originated from cultivated chamomile (PPK Kutjevo, Croatia). Air-dried and coarsely powdered plant material was refluxed with 10 ml methanol for 10 min and filtered. The filtrate was concentrated under reduced pressure and the residue dissolved in 5 ml methanol.
Reference substances were apigenin-7-glucoside, apigenin-7-(6″-O-acetyl)-glucoside, apigenin-7-(6″-O-diacetyl)-glucoside, luteolin-7-glucoside and quercetin-3-rutinoside (rutin) (C. Roth, Karlsruhe, Germany). 5–10 mg of each reference substance were dissolved in 10 ml methanol.
All solvents were of analytical grade, purchased from Chemica (Zagreb, Croatia).

3.2. Methods

3.2.1. Thin-layer chromatography

Precoated 20 × 20 cm silica gel 60 F_{254} TLC plates (Merck, Darmstadt, Germany) were used. 5 µl volumes of drug extract (test solution) and reference solutions were applied as bands. Development was in the linear ascending mode at room temperature in chambers saturated with mobile phase. The eight mobile phases tested are given in Table 1. The chromatograms were dried in a stream of air. Visualisation of the flavonoids was achieved by spraying the sheets with 1% methanolic diphenylboryloxyethylamine, oversprayed with 5% methanolic polyethylene glycol 400. The chromatograms were evaluated in UV 365 nm light. Derivatives of apigenin and quercetin appeared as yellowish and orange fluorescent bands respectively, while herniarin and chlorogenic acid gave blue fluorescent zones.

3.2.2. Mathematical methods

The mathematical methods described in the text below are applied in many areas 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 methods for identification of bioactive compounds as evaluation tools (quality measures).

3.2.2.1. Calculation of the information content

Extensive information was calculated for eight TLC systems by Shannon's formula. Calculation of the information content would become possible if the uncertainties before and after the analysis could be expressed in a quantitative way.
Distributing of R_F values into groups with error factor E (e.g., E = 0.05 or E = 0.10) with respect to units and assuming R_F values of n_k in the k^{th} group, the average information content (entropy) is given by the following Shannon equation [25, 26]

$$I(X) = H(X) = -\sum_k \frac{n_k}{n} \operatorname{ld} \frac{n_k}{n} \text{ [bit]} \quad (1)$$

It is assumed that the compounds with R_F values within one group cannot be distinguished. The system with the highest informational content gives the best solution for the differentiation of the compounds considered. It is obvious that the entropy is at its highest level, i.e., $H_{max}(X) = \operatorname{ld} n (n = \Sigma n_k)$ if there is only one R_F value within each group.

3.2.2.2. 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 [27–30]. 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 − 1). Therefore, the DP of k systems is:

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

The average number of chromatographically similar compounds (T) for the chromatographic systems considered can be calculated from the following equation [27]:

$$T = 1 + (N-1)\,(1 - DP_k) \quad (3)$$

3.2.2.3. 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 31.
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 × 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 given by the equation:

$$d_{k,l} = \sqrt{\sum_{i=1}^{N} \frac{(R_{F_{i,k}} - R_{F_{i,l}})^2}{N}} \quad (4)$$

Normalization (i.e., dividing the sums of squares by N) allows chromatographic systems with some unknown N values to be included.
Step 3: Classification. Chromatographic systems with a high degree of resemblance are grouped into clusters. Cluster formation in this work was carried out by a weighted pair group method [31, 32]. The smallest distance $d_{j,k}$ or highest correlation coefficient (r > 0.95) between solvents j and k is selected: j and k are the most similar solvent systems and are therefore considered to form one group p′. The similarity coefficient between the new group p′ and all other phases (e.g. q) is calculated, e.g. for the distance, as follows:

$$d_{j,p'} = 1/2(d_{j,p} + d_{j,q}) \quad (5)$$

The total number of rows and columns in the resemblance matrix is, therefore, reduced to one. 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 a dendrogram [33–36]. The three approaches were compared applying our computer search program KT 1 37.

List of symbols

I(X)	–	average information content
H(X)	–	entropy
H_{max}	–	maximal value of entropy
n_k	–	the number of compounds whose R_F is in k^{th} R_F interval
n	–	total number of compounds
E_i	–	error of measurement of R_F of i^{th} compound (resolution)
DP_k	–	Discriminating Power of analyzed systems divided in groups containing k elements
M	–	total number of matching pairs
T	–	average number of chromatographically similar compounds
$d_{k,l}$	–	normed distance between k^{th} and l^{th}
$d_{jp'}$	–	distance between j^{th} and p'^{th}

Acknowledgement: This work was supported by the Ministry of Science and Technology of the Republic of Croatia (Project No. 006242). The authors are grateful to Dr. Claudio Redaelli for providing apigenin glycosides.

References

1 Sorm, P.; Cekan, Z.; Herout, V.; Raskova, H.: Chem. Listy **46**, 308 (1952)
2 Lang, W.; Schwandt, K.: Dtsch. Apoth. Ztg. **97**, 149 (1957)
3 Tyihak, E.; Sarkany-Kiss, J.; Verzar-Petri, G.: Pharmazie **17**, 301 (1962)
4 Hörhammer, L.; Wagner, H.; Salfner, B.: Arzneim. Forsch. **13**, 33 (1963)
5 Kunde, R.; Isaac, O.: Planta Med. **37,** 124 (1979)
6 Exner, J.; Reichling, J.; Cole, T. H.; Becker, H.: Planta Med. **41**, 198 (1981)

7 Redaelli, C.; Formentini, L.; Santaniello, E.: Herba Hung. **18**, 323 (1979)
8 Hava, M.; Janku, J.: Rev. Czech. Med. **3**,130 (1957)
9 Achterrath-Tuckermann, U.; Kunde, R.; Flaskamp, E.; Isaac, O.; Thiemer, K.: Planta Med. **39**, 38 (1980)
10 Hörhammer, L.: Dtsch. Apoth. Ztg. **101**, 1178 (1961)
11 Hörhammer, L.; Wagner, H.: Dtsch. Apoth. Ztg. **14**, 1 (1962)
12 Jakovlev, V.; Isaac, O.; Flaskamp, E.: Planta Med. **49**, 67 (1983)
13 Della Loggia, R.: Dtsch. Apoth. Ztg., Suppl. I **125**, 9 (1985)
14 Tubaro, A.; Zilli, C.; Redaelli, C.; Della Loggia, R.: Planta Med. **51**, 35 (1984)
15 Redaelli, C.; Formentini, L.; Santaniello, E.: Planta Med. **42**, 288 (1981)
16 Redaelli, C.; Formentini, L.; Santaniello, E.: Planta Med. **43**, 412 (1981)
17 Dölle, B.; Carle, R.; Müller, W.: Dtsch. Apoth. Ztg. Suppl. I **125**, 14 (1985)
18 Stanić, G.; Katušin-Ražem, B.; Petričic, J.: Farm. Glas. **41**, 179 (1988)
19 Medić-Šarić, M.; Stanić, G.; Maleš, Ž.; Šarić, S.: J. Chromatogr. **776**, 355 (1997)
20 Stahl, E.; Schorn, P. J.: Hoppe-Seyler's Z. Physiol. Chem. **325**, 263 (1961)
21 Luckner, M.; Bessler, O.; Luckner, R.: Dtsch. Apoth. Ztg. **127**, 681 (1965)
22 Glotzbach, B.; Rimpler, H.: Planta Med. **16**, 1 (1968)
23 Wagner, H.; Bladt, S.; Zgainski, E. M.: Drogenanalyse, p. 164, Springer Verlag, Berlin Heidelberg New York 1983
24 Wichtl, M.; Bozek, B.; Fingerhut, T.: Pharmazie **42**, 509 (1987)
25 Chaitin, G. J.: Algorithmic Information Theory, Cambridge University Press, Cambridge 1987
26 De Clercq, H.; Massart, D. L.: J. Chromatogr. **115**, 1 (1975)
27 Moffat, A. C.; Smalldon, K. W.; Brown, C.: J. Chromatogr. **90**, 1 (1974)
28 Moffat, A. C.; Smalldon, K. W.: J. Chromatogr. **90**, 9 (1974)
29 Owen, P.; Pendlebury, A.; Moffat, A. C.: J. Chromatogr. **161**, 187 (1978)
30 Planar Chromatography, (R. E. Kaiser, ed.) Vol. 1, Alfred Huethig Verlag, Heidelberg 1986
31 Massart, D. L.; De Clercq, H.: Anal. Chem. **46**, 1988 (1974)
32 Sneath P. H. A.; Sokal, R. R.: Numerical Taxonomy, W. H. Freeman and Co., San Francisco 1973
33 Rotar, A.; Kozjek, F.; Medić-Šarić, M.: Acta Pharm. **43**, 157 (1993)
34 Maleš, Ž.; Medić-Šarić, M.; Kuštrak, D.: Acta Pharm. **44**, 183 (1994)
35 Medić-Šarić, M.; Brantner, A.; Maleš, Ž.: Acta Pharm. **46**, 115 (1996)
36 Medić-Šarić, M.; Maleš, Ž.; Stanić, G.; Šarić, S.: Croat. Chem. Acta **69**, 1265 (1996)
37 Medic-Šarić, M.; Šarić, S.; Maysinger, D.: Acta Pharm. Jugosl. **39**, 1 (1989)

Received February 4, 2000
Accepted July 7, 2000

Prof. Dr. sc. Marica Medić-Šarić
Department of Pharmaceutical Chemistry
University of Zagreb
A. Kovačića 1
10000 Zagreb
Croatia
bebamms@nana.pharma.hr