Application of information theory and numerical taxonomy methods to thin-layer chromatographic investigations of Hypericum perforatum L.

Medić-Šarić, Marica; Brantner, Adelheid; Maleš, Željan

Source / Izvornik: **Acta Pharmaceutica, 1996, 46, 115 - 124**

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

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:163:452067

Acta Pharm. 46 (1996) 115–124 Original scientific paper

Application of information theory and numerical taxonomy methods to thin-layer chromatographic investigations of *Hypericum perforatum* L.

MARICA MEDIĆ-ŠARIĆ

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

ADELHEID BRANTNER*

Institute of Pharmacognosy
University of Graz, Universitätsplatz 4/I
A-8010 Graz, Austria

ŽELJAN MALEŠ

Department of Pharmacognosy
Faculty of Pharmacy and Biochemistry
University of Zagreb
Marulićev trg 20/II
10000 Zagreb, Croatia

The efficiency of nine TLC systems for separating six main components of a methanolic extract of *Hypericum perforatum* L. (St. John's wort) was investigated. For this purpose, three mathematical approaches have been applied. The first approach was the calculation of the information content for each chromatographic system. Secondly, the discriminating power of the systems was measured individually and in diverse combinations. As a third approach the chromatographic systems were classified into dissimilar groups and of each group with similar separation characteristics the best system was selected according to the information content. The classification was carried out by the numerical taxonomy methods.

Keywords: *Hypericum perforatum* L., thin-layer chromatography, information content, discriminating power, numerical taxonomy

Received January 3, 1996

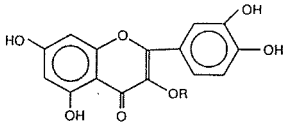
In this study we used numerical taxonomy methods in order to evaluate the separating power of TLC systems for splitting a methanolic extract of *Hypericum perforatum* L. (St. John's wort) into six main compounds (flavonoid glycosides, chlorogenic acid, hypericin).

Hypericum perforatum is an aromatic perennial herb belonging to the family *Hypericaceae*, of widespread distribution in Europe, western Asia and throughout the United States. It is a ca. 60 cm tall herbaceous plant producing five-merous yellow flowers with many and prominent, long stamens and opposite, translucently dotted leaves. The drug

* Correspondence

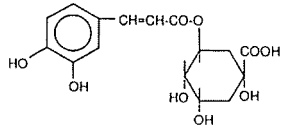
(*Hyperici* herba) consists of the aerial parts (leaves and flowering tops). The herb is used internally for the treatment of psychogenic disturbances, depressive states and/or nervous excitements, and externally as oily preparations for the treatment and after-treatment of incised and contused wounds and first-degree burns (1).

Chemical investigations have detected a number of constituents including hypericin and hypericin-like substances (2), flavonoids (3, 4), the hyperforin antibiotic substance (2, 5, 6), essential oil, small amounts of procyanidins (7) and tannins. The structures of the flavonoid glycosides rutin, hyperoside, quercitrin, *iso*quercitrin, of chlorogenic acid and hypericin analysed in this paper are presented in Fig. 1.

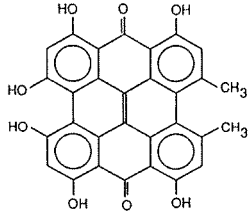


R =

rhamnose	
+ glucose	rutin
galactose	hyperoside
glucose	isoquercitrin
rhamnose	quercitrin



chlorogenic acid



hypericin

Fig. 1. Structural formulas of the investigated constituents of a methanolic *Hypericum* extract.

In our times, St. John's wort is widely used in Europe for its antidepressive effects as an alternative to medical synthetic drugs. For this reason numerous reports dealing with the drug and its preparations can be found in literature, *e.g.* a monograph edited by the German Commission E. Several authors propose analytical methods for investigating *Hypericum perforatum*. In this paper the efficiency of different TLC systems described in literature and own modifications of them are compared by numerical taxonomy methods.

The information content or equivocation can serve – with certain restrictions – as an important criterion for the evaluation, selection or optimization of analytical procedures (8). Information theory can be used to compare the quality of chromatographic systems as well as to optimize such systems (9). Shannon's formula was applied to calculate the information contents and the unit of information (bit), as it is used in binary-coded systems. The information theory approach is based on the combination of all chromatographic systems studied with an estimation for the average amount of information of the selected set of compounds (10, 11). The discriminating power (DP) is used

as a measure of the effectiveness of chromatographic systems. The DP of chromatographic system is the probability of separating two compounds selected at random from a specific substance population (12). The calculation of the DP values and their maximization can be more easily attained by the following conditions: an even distribution of hR_f values, reproducibility of results and no correlations between the chromatographic systems.

The methods of numerical taxonomy classify the chromatographic systems according to clusters (13, 14). The chromatographic systems are divided into groups with similar separation properties. The selection of the most efficient chromatographic system from each group is carried out by numerical taxonomy methods. The results obtained by applying those techniques were compared by using our computer search program KT 1 (10).

EXPERIMENTAL

Materials

Test solution. – 1.0 g air-dried, coarsely powdered drug was refluxed with 10.0 mL methanol for 10 minutes, filtered, the filtrate concentrated under reduced pressure, and the residue taken up in 5.0 mL methanol.

Reference solution. – 10 mg rutin and 5 mg hyperoside were dissolved in 10 mL methanol.

Precoated 20×20 cm TLC silica gel 60 F_{254} sheets (thickness of layer 0.2 mm) were obtained from Merck (Darmstadt, Germany). The foils were prewashed twice with methanol, desiccated in stream of air and kept overnight at relative humidity of 40–45%. Sample application was performed with a Linomat IV apparatus (Camag, Muttenz, Switzerland). 5 μL of the extract solutions were applied 20 mm from the bottom of the sheet as 10 mm wide bands at a delivery rate of 4 s μL^{-1}. Developments were performed by the linear ascending mode at 20.5–22.0 °C room temperature and 40–45% relative humidity in a saturated (eluent system 1–6, 9) or unsaturated (eluent system 7, 8) twin-through 20×20 chamber (Camag). The solvent volume was 30 mL and the development distance 150 mm (single development). All solvents were of analytical grade, purchased from Merck. The foils were dried at room temperature in a stream of cold air. Visualization of the flavonoids was attained by spraying the sheets with 1% methanolic diphenylborylethylamine, followed by 5% ethanolic polyethylene glycol 4000. The chromatograms were evaluated in UV 366 nm light (flavonoids as orange-yellow, chlorogenic acid as blue and hypericin as red fluorescent bands).

Methods

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

The distribution of hR_f values into groups with an error factor E (*e.g.*, E = 0.05 or E = 0.10) with respect to hR_f units and the assumption of $n_k hR_f$ values in the k-th groups, the average information content (entropy) is given by the following Shannon's equation (15–17):

$$I(X) = H(X) = -\sum_k [n_k/n] \text{ ld } [n_k/n][\text{bit}]$$

It is also assumed that the compounds with hR_f values within one group cannot be identified. It is obvious that the entropy is at its highest level if there is only one hR_f value, *i.e.* $H_m(X) = \text{ld } n$ 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 (18– 22). 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:

$$DP_k = 1 - 2M/N(N - 1)$$

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

$$T = 1 + (N - 1)(1 - DP_k)$$

Calculation of taxonomic distances, cluster formation and dendrogram. – Taxonomy is defined as the theoretical study of classification including its elementary principles, procedures and rules (13). Numerical taxonomy deals with ways of classifying chromatographic systems into taxonomic groups based on characteristic values (hR_f values). The mathematical principle of this procedure is established on the formation of a matrix where the columns represent the solvent systems and the lines of the substances. Classification is carried out with respect to resemblances between the solvent systems. Dissimilarity expressed as the complement of similarity is proportional to the distances of the solvent systems in the given metric space. The greater the differences in properties of the solvent systems, the larger their spatial distances are. Taxonomic distance is inversely related to similarity. The distance $d_{j,k}$ between the solvent systems j and k is defined as:

$$d_{j,k} = \left[\sum_{i=1}^{N} (X_{i,j} - X_{i,k})^2\right]^{\frac{1}{2}}$$

where $X_{i,j}$ and $X_{i,k}$ are the hR_f values of investigated compound i in the solvent systems j and k, and N is the number of hR_f values taken into account.

In the classification by taxonomy a resemblance matrix containing either the correlation coefficient or the taxonomic distance is constructed. The reduction of this matrix is carried out by a weighted pair group method using the arithmetic average (13). The smallest distance $d_{i,k}$ or highest correlation coefficient is selected: i 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:

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

The total number of rows and columns in the resemblance matrix is therefore reduced to one. This process is repeated until all chromatographic systems are comprised in one non-overlapping hierarchic system of groups and subgroups (clusters). The procedure for cluster formation is presented by a dendrogram (10, 23, 24).

RESULTS AND DISCUSSION

Optimization procedures are widely used in the development of different analytical methods. Chromatographic methods, particularly TLC and HPLC, are procedures in which different chemometrical methods are used most frequently (25, 26).

The different times of development of the TLC sheets and the hR_f values of the compounds on which the calculations are based are listed in Table I. A data set of hR_f values for the separation of the main components of a methanolic extract of *Hypericum perforatum* into nine different chromatographic systems was analysed. An optimal combination of two or more systems was selected using two procedures:

a) determination and comparison of the information content and DP for all possible combinations of chromatographic systems,

b) division of chromatographic systems into groups with similar separation properties and selection of the most efficient chromatographic system from each group.

Table II gives the input and output data for the information content and the DP for each TLC system and for combined systems in a range of error factors. The error factors were 0.05 and 0.10, respectively.

Under the conditions frequently used in chromatographic analysis, *i.e.* E = 0.05, the most suitable systems for separating the compounds studied are the chromatographic systems 2 and 4 because they showed the largest DP values (DP = 1.000) and a high information content (I = 2.585). Systems 1, 3 and 5 are also suitable because of their slightly lower DP value (DP = 0.933) and identical information content as systems 2 and 4.

The results made evident that at E = 0.10 the most appropriate systems for separating the compounds are the systems 1 and 4 because of their high DP value and information content (DP = 0.933, I = 2.585). The systems 2 and 3 are also acceptable. Combining two chromatographic systems with the error E = 0.05 or 0.10 all systems show very high DP values (DP = 1.000 and 0.933, respectively) and the number of compounds with similar chromatographic properties is at a minimum (T = 1.000 and 1.333, respectively).

Table I. TLC-investigation of components of Hypericum perforatum *in different eluents*

No	Eluent system	Time of development (min)	Literature	hR_f values of components					
				Rutin	Hyperoside	*Iso*-quercitrin	Quercitrin	Chlorogenic acid	Hypericin
1	Etac-EMK-HCOOH–H_2O 50:30:10:10	100	(27)	47	70	74	85	58	98
2	Etac-HCOOH–H_2O 30:2:3	105	(28)	12	37	43	63	30	82
3	Etac-HCOOH–CH_3COOH–H_2O 100:11:11:27	141	(29)	46	70	74	86	54	95
4	Etac-HCOOH–CH_3COOH–EMK–H_2O 50:7:3:30:10	108	(30)	44	69	75	87	55	99
5	Ethylformate-acetone-H_2O–CH_3COOH 5:4:1:0.4	85	(31)	36	52	56	69	31	99
6	n-BuOH–CH_3COOH–H_2O 12:3:5	380	(32)	57	69	69	75	48	83
7	Toluene-Etac-HCOOH 50:40:10	58	(33)	0	4	6	13	2	62
8	Toluene-Etac-HCOOH 15:75:10	52	(33), var.	6	18	25	45	18	78
9	Acetone-tert. butylmethylether-HCOOH 40:50:10	60	(34), var.	24	48	57	78	47	97

System 4 can be found in the first ten combinations. In a series of three systems, regardless of the error factor E, the TLC system 4 is also found in the first ten combinations.

In order to obtain the optimal combination of systems according to the dendrogram (Fig. 2), irrespective of the error factor E and number of systems (K = 2 or K = 3), the

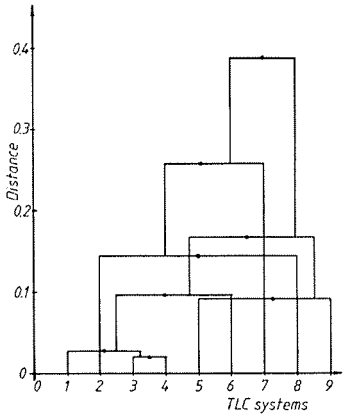


Fig. 2. Dendrogram for nine TLC systems.

Table II. Input and output data for the DP and cluster formation

Input data (hR_f-values of the main components of *Hypericum perforatum*)

Compound	Solvent system								
	1	2	3	4	5	6	7	8	9
1. Rutin	47	12	46	44	36	57	0	6	24
2. Hyperoside	70	37	70	69	52	69	4	18	48
3. *Iso*quercitrin	74	43	74	75	56	69	6	25	57
4. Quercitrin	85	63	86	87	69	75	13	45	78
5. Chlorogenic acid	58	30	54	55	31	48	2	18	47
6. Hypericin	98	82	95	99	99	83	62	78	97

Systems

TLC-system	E = 0.05		E = 0.10	
	DP	I(bit)	DP	I(bit)
1	0.933	2.585	0.933	2.585
2	1.000	2.585	0.867	2.585
3	0.933	2.585	0.800	2.585
4	1.000	2.585	0.933	2.585
5	0.933	2.585	0.867	1.918
6	0.933	2.252	0.667	2.252
7	0.667	1.792	0.467	1.252
8	0.933	2.252	0.800	2.252
9	0.933	2.252	0.800	2.252

Table II. Continued

Combined systems

	Combination sequence		E = 0.05		E = 0.10	
			DP	T	DP	T
K = 2	1. system	6 – 9	1.000	1.000	0.933	1.333
	2. system	6 – 8	1.000	1.000	0.933	1.333
	3. system	5 – 9	1.000	1.000	0.933	1.333
	4. system	5 – 8	1.000	1.000	0.933	1.333
	5. system	4 – 9	1.000	1.000	0.933	1.333
	6. system	4 – 8	1.000	1.000	0.933	1.333
	7. system	4 – 7	1.000	1.000	0.933	1.333
	8. system	4 – 6	1.000	1.000	0.933	1.333
	9. system	4 – 5	1.000	1.000	0.933	1.333
	10. system	3 – 9	1.000	1.000	0.933	1.333
K = 3	1. system	6 – 8 – 9	1.000	1.000	0.933	1.333
	2. system	6 – 7 – 9	1.000	1.000	0.933	1.333
	3. system	6 – 7 – 8	1.000	1.000	0.933	1.333
	4. system	5 – 8 – 9	1.000	1.000	0.933	1.333
	5. system	5 – 7 – 9	1.000	1.000	0.933	1.333
	6. system	5 – 7 – 8	1.000	1.000	0.933	1.333
	7. system	5 – 6 – 9	1.000	1.000	0.933	1.333
	8. system	5 – 6 – 8	1.000	1.000	0.933	1.333
	9. system	4 – 8 – 9	1.000	1.000	0.933	1.333
	10. system	4 – 7 – 9	1.000	1.000	0.933	1.333

Cluster formation			
Cluster	Solvent	Solvent	Distance
1	3	4	0.0200
2	1	3	0.0208
3	3	7	0.0915
4	1	4	0.0956
5	2	5	0.1417
6	1	3	0.1636
7	2	3	0.2560
8	1	2	0.3889

best systems in cluster 1 are the systems 4, 5 or 6, and in cluster 2 the systems 2, 7, or 8. Considering the best system to be the one with the largest DP and information content I, it is obvious that the most suitable system of cluster 1 is system 4 due to its high DP value (DP = 1.000 or 0.933) and the information content (I = 2.585).

The results obtained by applying these techniques are useful mathematical tools in the classification and combination of chromatographic systems. As far as the methods are concerned, it is clear that the combined use of mathematical tools such as information theory, pattern recognition, principal components analysis and numerical taxonomy permits the classification and combination of chromatographic techniques. They should

therefore be of value in comparative physicochemical studies of these systems and in the selection of sets of preferred solvent systems.

The proposed calculations point to the conclusion that for the TLC analysis of six main components of *Hypericum perforatum* chromatographic system 4, which is described by Wagner *et al.* (30) for *Crataegi* folium and flos and system 2 reported in Hager's Handbuch der Pharmazeutischen Praxis (28) are the most useful methods. The suggested solvent combinations proved an excellent reproducibility of the results and an even distribution of the hR_f values. Some of the TLC systems described, especially the one for *Hypericum perforatum* in papers or monographis (*e.g.* Deutscher Arzneimittel-Codex), did not give sufficient results.

REFERENCES

1. German Commission E monograph 1984, *BAnz* No. 228, dated 05. 12. 1984.
2. J. Hölzl and E. Ostrowski, *Dtsch. Apoth. Ztg.* **127** (1987) 1227.
3. A. Brantner and Ž. Maleš, *Planta Med.* **56** (1990) 582.
4. D. Kuštrak, Ž. Maleš, A. Brantner, and I. Pitarević, *Acta Pharm. Jugosl.* **40** (1990) 551.
5. J. Hölzl and E. Ostrowski, *Dtsch. Apoth. Ztg.* **127** (1987) 127.
6. P. Maisenbacher and K.-A. Kovar, *Planta Med.* **58** (1992) 351.
7. J. Hölzl and H. Münker, *Acta Agron.* **34** (1985) 52.
8. D. E. Clegg and D. L. Massart, *J. Chem. Educ.* **70** (1993) 19.
9. P. Cleij and A. Dijkstra, *Fresenius' Z. Anal. Chem.* **298** (1979) 97.
10. M. Medić-Šarić, S. Šarić, and D. Maysinger, *Acta Pharm. Jugosl.* **39** (1989) 1.
11. P. F. Dupuis and A. Dijkstra, *Anal. Chem.* **47** (1975) 379.
12. A. C. Moffat, A. H. Stead, and K. W. Smalldon, *J. Chromatogr.* **90** (1974) 19.
13. P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy*, W. H. Freeman and Co., San Francisco 1973.
14. D. L. Massart, M. Lanwereys, and P. Lenders, *J. Chromatogr. Sci.* **12** (1974) 617.
15. C. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana 1949.
16. H. De Clercq and D. L. Massart, *J. Chromatogr.* **115** (1975) 1.
17. G. J. Chaitin, *Algorithmic Information Theory*, Cambridge University Press, Cambridge 1987.
18. A. C. Moffat, K. W. Smalldon, and C. Brown, *J. Chromatogr.* **90** (1974) 1.
19. A. C. Moffat and K. W. Smalldon, *ibid.* **90** (1974) 9.
20. P. Owen, A. Pendlebury, and A. C. Moffat, *ibid.* **161** (1978) 195.
21. P. Owen, A. Pendlebury, and A. C. Moffat, *ibid.* **161** (1978) 187.
22. R. E. Kaiser, *Planar Chromatography*, Alfred Huethig Verlag, Heidelberg 1986, Vol. I, 22.
23. A. Rotar, F. Kozjek, and M. Medić-Šarić, *Acta Pharm.* **43** (1993) 157.
24. Ž. Maleš, M. Medić-Šarić, and D. Kuštrak, *ibid.* **44** (1994) 183.
25. R. G. Brereton, *Anal. Proc.* **28** (1991) 145.
26. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, and L. Kaufman, *Chemometrics*, Elsevier, Amsterdam 1988.
27. E. Stahl, *Dünnschichtchromatographie*, Springer Verlag, Berlin–New York 1967, 658.
28. R. Hänsel, K. Keller, H. Rimpler, and G. Schneider (Eds.), *Hagers Handbuch der Pharmazeutischen Praxis*, Springer Verlag, Berlin-New York 1994, Vol. 5, 474.
29. H. Wagner, S. Bladt, and E. M. Zgainski, *Drogenanalyse*, Springer Verlag, Berlin–New York 1983, 110.
30. H. Wagner, S. Bladt, and E. M. Zgainski, *ibid.*, 164.

31. L. Xu and A. Liu, *Yaowu Fenxi Zachi* **7** (1987) 280.
32. M. D'Agostino, V. DeFeo, and C. Pizza, *Filoterapia* **65** (1994) 472.
33. *Deutscher Arzneimittel-Codex*, Govi-Verlag, Pharmazeutischer Verlag GmbH, Frankfurt am Main and Deutscher Apotheker Verlag, Stuttgart 1986, Vol. II, 3rd Suppl. 1991, J010.
34. M. Wichtl, in *Herbal Drugs* (Ed. N. G. Bisset), Medpharm Scientific Publ., Stuttgart and CRC Press, Boca Raton–Ann Arbor–London–Tokyo 1994, 273.

SAŽETAK

Primjena teorije informacije i metoda numeričke taksonomije u istraživanju gospine trave (*Hypericum perforatum* L.) tankoslojnom kromatografijom

MARICA MEDIĆ-ŠARIĆ, ADELHEID BRANTNER i ŽELJAN MALEŠ

Metodom tankoslojne kromatografije istražen je metanolni ekstrakt gospine trave (*Hypericum perforatum* L.). Uspješnost devet kromatografskih razvijača ispitana je matematičkim metodama: određivanjem srednjega vlastitog sadržaja informacije, koeficijenta DP, te formiranjem klastera i dendrograma. Svrstavanje kromatografski sličnih razvijača u klastere provedeno je metodama numeričke taksonomije.

Zavod za farmaceutsku kemiju i Zavod za farmakognoziju
Farmaceutsko-biokemijski fakultet Sveučilišta u Zagrebu, Zagreb

i

Institut für Pharmakognosie
der Karl-Franzens-Universität Graz, Graz, Austria