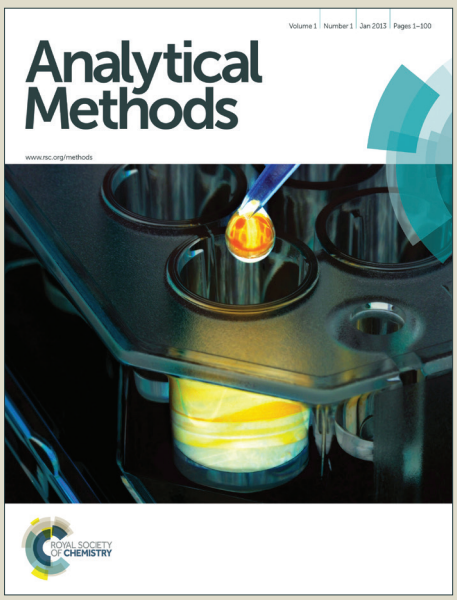
Analytical Methods

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Open column, reversed-phase high-performance liquid chromatography with diode array detection and chemometric strategy for investigation of metabolic fingerprints of complex systems

Raquel Justo[a], Sabrina Afonso[a], Roy E. Bruns[b], Ieda S. Scarminio[a,*]

[a]Laboratório de Quimiometria em Ciências Naturais, Departamento de Química, Universidade Estadual de Londrina, Caixa Postal 6001, CEP 86051-970, Londrina, Paraná, Brazil

[b]Instituto de Química, Universidade Estadual de Campinas, Caixa Postal 6154, CEP 13083-862, Campinas, São Paulo, Brazil.

Correspondence: Dra. Ieda Spacino Scarminio, *Laboratório de Quimiometria em Ciências Naturais, Departamento de Química, Universidade Estadual de Londrina,CP6001, CEP 86051-990, Londrina, Paraná, Brazil*

E-mail: ieda@qui.uel.br

Tel: +55-043-3371-4811

Fax: +55-043-3371-4286

Abstract

A new strategy for investigation of metabolic fingerprints in complex systems was developed based on a combination of open column chromatography, reversed-phase high-performance liquid chromatography and K-means clustering. A simplex centroid mixture design with four solvents, hexane, dichloromethane, ethyl acetate and methanol, was applied to crude extract of the *Bauhinia variegata* L to gradually change the mobile phase polarity of the open column. K-means clustering was used to screen the profiles of the 3033 spectra from high-performance liquid chromatography with diode array detection data. The seventeen k-means clusters were discriminated into three groups. The largest one containing 2617 (86.3%) spectra has a strong absorbance band around 210 nm that is commonly found in many naturally abundant compounds. The smallest group contains only 134 (4.4%) spectra and has strong absorbance bands at wavelengths well above 250 nm and are potentially interesting for further metabolic studies and fingerprinting. The other group judged to have spectral profiles intermediate between those of the other two contains 282 (9.3%) spectra. Ternary and quartenary mixtures of the statistical design provided potentially useful information for fingerprinting *Bauhinia variegata* L. extracts relative to what could be obtained using only pure solvents and binary mixtures that are commonly employed in open column chromatography.

1. Introduction

In recent years, the chromatographic fingerprinting technique has been used as a tool for metabolic profiling of natural products, as well as for identification and quality control of herbal medicines. This technique is recommended by both the Food and Drug Administration (FDA)[1] and the European Medicines Agency (EMEA)[2] to assess the consistency of botanical drugs. The World Health Organization (WHO) accepts fingerprint chromatography for identification and quality evaluation techniques of herbal medicines[3]. The quality of a chromatographic fingerprint is highly dependent on the degree of chromatographic separation and the concentration distribution of all the chemical components in herbal medicines. Multivariate optimization results in improved resolution of these concentration distributions while chemometric data analysis facilitates searching among many spectral profiles.

Although thin layer chromatography (TLC), high performance liquid chromatography (HPLC), and high performance capillary electrophoresis[4], have been widely used for screening active components, little attention has been paid to open column chromatography. Open column chromatography is still one of the most important operations carried out in natural products laboratories for isolating or searching for a given constituent. The separation normally involves four solvents and starts with a nonpolar solvent (or a minimum polarity solvent) with the polarity being gradually increased[5] using pure solvents or binary mixtures. Considering that the separation of sample components is governed to a large extent by the composition of the mobile phase, and that a single eluent or binary mixture may not be sufficient to separate and elute each compound of a given natural product, a specific separation strategy is necessary to

achieve this objective. Thus the choice of the appropriate mobile phases as well as the stationary phase is crucial for obtaining optimum separation conditions[6].

In our previous work[7-10], a statistical mixture design approach has been used to efficiently optimize mobile phases in reverse-phase liquid chromatography (RP-HPLC). Results showed that changes in the mobile phase proportions of the organic modifier, led to significant changes in separation selectivity. Recently, we used a statistical mixture design involving four solvents commonly applied in separation techniques in open column chromatography to investigate the influence of solvent proportions on chromatographic fractionation[11].

Here a strategy is developed based on a combination of open column chromatography, reversed-phase high-performance liquid chromatography and chemometric techniques for investigation of the metabolic fingerprint of crude extract of *Bauhinia variegata* L. Initially open column chromatography is used for the fractionation of its crude extract changing mobile phase polarity according to a four component simplex centroid mixture design. The separation started with the nonpolar hexane solvent with the polarity being gradually increased until the last mobile phase, pure methanol. Besides investigating the effect of polarity on system resolution the statistical design is efficient for optimizing the solvent proportions of the mixture providing a maximum of spectral information for the metabolic fingerprint.

Then k-means clustering is applied to the UV spectra of the open column fractions obtained from reversed phase high performance liquid chromatography coupled with photodiode array detection (RP-HPLC-DAD). Since such a huge number of spectra are produced complicating analysis, the k-means clustering method effectively screens them

finding the subsets most promising for detailed study. In this way the diversity of chemical compounds is assessed while exploring the effects of each open column chromatographic mobile phase for fingerprint analysis in natural products. The method is not restricted to natural product fingerprint and compound identification but should be useful for screening other complex systems for quality control and other purposes.

2. Materials and Methods

2.1 Chemicals

HPLC grade acetonitrile and methanol were purchased from VETEC Química Fina (Rio de Janeiro, Brazil). Mobile phase mixture preparations were made using water prepared with the Millipore Milli-Q purification system (São Paulo, Brazil). Hexane, dichloromethane, ethyl acetate and methanol were also purchased from VETEC and were of analytical grade.

2.2 Plant material

Leaves of *Bauhinia variegata* were collected from Londrina State University. The botanical identity was authenticated in its Animal and Vegetal Department by the biologist Manuel R. C. Paiva. A voucher specimen has been preserved in the UEL Herbarium, under No. 35133. Collection was performed from April to October 2008.

2.3 Extract preparations and open column chromatography

B. variegata extract leaves were prepared by weighing about 20.00 g of dried leaves of a mixture obtained from a six month collection period and placed in contact with 100 mL

of a 65:35 (V:V) dichloromethane - acetone solvent mixture. This proportion was chosen based on the results of our previous work[7]. Then the mixture was solubilized in ultrasound (Unique, model USC 1400, São Paulo, Brazil) for 30 min and filtered through cotton. This process was repeated six times, resulting in a total volume of 600 mL, which was taken to evaporation in a rotary evaporator (Fisatom, São Paulo, Brasil) at 50 °C ± 3 °C. After evaporation of the solvent mixture, the extracts were placed in amber vials and stored at room temperature.

A 2 g extract, was placed on top of a (30×2.5 cm) silica gel chromatographic column. The extract was eluted using mixtures of the statistical design[12] given in Figure 1 for the hexane, dichloromethane, ethyl acetate and methanol components. 20 mL portions of the eluent were collected sequentially in labeled tubes and analyzed by thin layer chromatography to determine the different fractions. TLC was carried out with glass plates (10×10 cm) coated with silica gel (VETEC, Rio de Janeiro, Brazil). The separated spots were visualized with ultraviolet light. A minimum volume of 250 mL was passed through the columns for each solvent mixture. Column chromatography was carried out on silica gel 60, E. Merck, mesh size 35-70 ASTM 200-500 μm.

2.4 HPLC analysis

Extract solutions of each of the discriminated TLC fractions were subjected to RP-HPLC analysis. For this the extracts of each fraction were dissolved in about 4.00 mL of a hexane:dichloromethane:ethyl acetate:methanol (1:1:1:1,v/v/v/v) mixture. After 1 hour of storage in the refrigerator, each solution was filtered through common filter paper (Labstore, Londrina, Brazil) and stored in a freezer. Then 200 μL aliquots of the filtrate

were mixed with 200 μL (1:1v/v) of the mobile phase (HPLC). Diluted samples were filtered two more times through 0.22 μm Millipore Millex filters purchased from Labstore Equipamentos para Laboratórios (Curitiba, Brasil) and injected into the HPLC system.

Elution was performed isocratically and monitored at 210, 240 and 254 nm. RP-HPLC analysis was carried out on a Thermo Electron Corporation Finnigan Surveyor Plus instrument with a PDA plus detector using a 5 μm Phenomenex Gemini C18 (250×4.6 mm) column (São Paulo, Brazil) purchased from Labstore Equipamentos para Laboratórios. The flow rate was 1 mL/min. All solvents were degassed in an ultrasonic, Unique, Model USC 800 bath (Labstore Equipamentos para Laboratórios). Satisfactory separation was achieved in 30 min. Two mobile phases with different chromatographic strengths were used. The first phase was prepared in methanol:acetonitrile:water at a ratio of 35:35:30 (v/v/v) and the second in 17.5:17.5:65 (v/v/v) methanol:acetonitrile:water.

2.5 Chemometric methods

Chemometric methods comprise a group of statistical, mathematical and graphical techniques that analyze many variables simultaneously. The discussion below summarizes the methods used here.

In Hierarchical Cluster Analysis (HCA) the goal is to find an optimal grouping for which the observations or objects within each cluster are similar, but different clusters are characterized by dissimilar objects[13]. Hierarchical clustering, in which clusters are formed sequentially, is the most common approach. The most similar objects are grouped first and these initial clusters are merged according to their similarities. In cluster

analysis, neither the number of clusters nor the clusters themselves are known in advance. The similarity or dissimilarity between objects is usually represented in a dendrogram for ease of interpretation. Different approaches to measuring distance between clusters give rise to different hierarchical methods. In this work Ward's procedure[13] was used to establish clusters.

The K-means method is a non-hierarchical clustering technique that partitions n observations into k clusters so that each observation belongs to the cluster with the nearest mean[13]. This can be done by minimizing the within-cluster sum of squares. It requires the user to pre-specify the number of clusters present in the data set although there is no a priori mathematical procedure to determine this number. Hence the requirement for pre-specifying the number of clusters is its major drawback. However the user can compare results obtained assuming different numbers of clusters, choosing the one providing the most useful information for his or her purposes.

3. Results and discussion

The twenty eight mixtures (open column chromatography mobile phases) were prepared according to the statistical mixture design in Figure 1 and applied in the elution order of increasing polarity from pure hexane to pure methanol as shown in Table 1. During the open column chromatographic process 603 small portions were collected and then combined into 95 different fractions according to their TLC similarities. Table 1 contains the number of different fractions for each solvent mixture.

RP-HPLC-DAD mobile phase and spectral data were investigated to determine conditions providing significant information about the metabolite classes in each fraction.

The effect of mobile phase composition on chromatographic separation was first investigated. For this experiment, two mobile phases were investigated, (35:35:30, v/v/v) and (17.5:17.5:65, v/v/v) MeOH:ACN:water. Three wavelengths at 210, 240 and 254 nm were also investigated to compare the number of peaks eluted. The largest number of peaks was observed with the (35:35:30, v/v/v) MeOH:ACN:water mobile phase at 240 nm and these conditions were selected for obtaining the RP-HPLC-DAD data. The second to last column of Table 1 contains the total number of chromatographic peaks observed for each one of the 28 open column mobile phase mixtures.

One can notice that the (75:25, v/v) hexane:ethyl acetate mixture yielded eleven different open column fractions with 379 RP-HPLC peaks, the (20:60:20, v/v/v) dichoromethane: ethyl acetate:methanol mixture yielded eight different open column fractions with 280 peaks and the (25:25:25:25, v/v/v/v) hexane:dichoromethane:ethyl acetate:methanol mixture yielded seven different fractions and 157 peaks. It is interesting to observe that the solvent proportions significantly affect the number of fractions produced. For example, the (33:33:33, v/v/v) and (60:20:20, v/v/v) ternary mixtures of dichloromethane: ethyl acetate: methanol produced only one fraction each. By varying the proportions of the (33:33:33 v/v/v) dichloromethane: ethyl acetate: methanol mixture to (20:60:20, v/v/v), the number of fractions changes from one to eight and the number of HPLC peaks increases from 40 to 280. These mixtures have almost identical polarities (4.2 and 4.3) yet appear to elute very different substances. Increasing the methanol proportion at the expense of dichloromethane and acetate also increases the number of open column fractions to five and the number of HPLC peaks to 178. In contrast pure hexane, dichoromethane and methanol each produced three different fractions with 105,

120 and 95 peaks, respectively, while ethyl acetate produced five different fractions and 212 peaks. These results show that ternary and quaternary mixtures furnish new information not provided by pure solvents and their binary mixtures normally used in the classical elution procedure[5]. Furthermore solvent proportion variations can significantly affect the number of fractions and peaks obtained in each mobile phase even though the polarity of the mixture hardly changes.

According to the literature, UV spectral data can be used in an attempt to identify and deduce the possible structures of plant metabolites. Although it is possible to visually differentiate spectra, it is very difficult to sort and compare large numbers of spectra. In our study ten of the open column mobile phase mixtures each extract more than a hundred peaks. Furthermore the spectra of all the open column mobile phases must be examined. In this case chemometric methods can be useful to discriminate among different spectra and reveal similar ones. Therefore, several of these methods were applied to the spectral data of the chromatograms of the 95 fractions. The data matrix was comprised of a number of rows equal to the number of peaks of the 95 fractions (3033) and 211 columns, corresponding to the 190-400 nm wavelength range.

Hierarchical cluster analysis (HCA) performed with Ward's agglomerative method was first used to group similar UV spectra and separate dissimilar ones. The clustering results indicated the existence of four to at least six different types of characteristic spectra depending on the selected linkage distance, with each cluster consisting of many small subclusters. Analyzing some of the spectra in each subcluster revealed the existence of many more than six different spectral groups complicating data analysis.

Considering the difficulty of directly observing the profiles of the spectra of each group, the k-means clustering method was finally used.

K-means clustering was executed initially assuming 20 different clusters. Some of these clusters have very similar spectra and further clustering attempts permitted reducing the number of clusters to 17 different sets of spectral profiles for the 95 fractions. This method is most convenient in this study, because it provides the average spectral profile of each cluster furnishing important information for those with knowledge in the field of natural products[14].

Table 1 shows the influence of mobile phases in open column chromatography for the separation of different spectral profiles of metabolites obtained by k-means clustering. The average spectra for each of the 17 clusters are presented in Figures 2a – 2c since they can be classified into three main types. The spectra in Figure 2a are characterized by strong absorption bands in the 200 - 210 nm region with little or no absorption at higher wavelengths. The clusters in Figure 2a contain a large majority of all the RP-HPLC-DAD spectra, 2617 spectra or 86.3%. These clusters, especially clusters II, VII, XI and XII, contain spectra that are eluted by almost all the open column mobile phases, apparently independent of solvent polarity. A majority of naturally occurring compounds strongly absorb in the 195-210 nm region. Furthemore absorptions in this range often indicate the presence of a $n \rightarrow \sigma^*$ transition, as mentioned earlier, indicating molecules containing atoms with nonbonding electrons, such as N or O. Cluster II, by far the one with the largest number of spectra, shows spectral profiles in Figure 2a that are indicative of the alkaloids, terpenoids[15] or alkamides[16].

The average spectra of the clusters shown in Figure 2b are discriminated from those in Figure 2a by strong absorbance bands in the 300-400 nm region. Strong absorption above 300 nm indicates the existence of a extended conjugated system such as condensed aromatics and condensed aromatics heterocyclic or their derivatives. These averages represent 134 individual spectra in clusters III, IV, VI, VIII, X, XIII and XVII or only 4.4% of all the DAD spectra. The last column in Table 1 contains the number of HPLC peaks provided by each open column mixture resulting in the spectra shown in Figure 2. These 134 spectra correspond to seven k-means clusters, the same number of clusters as found for the 2617 spectra represented in Fig. 2a, and indicate greater spectral diversity suggesting priority focussing on their more detailed study. Note that these spectra are obtained in almost all the open column mobile phases. In fact the (33:33:34 v/v/v) dichloromethane-acetone-methanol ternary mixture results in the highest number of peaks, 19 of which 16 belong to cluster III. On the other hand a more methanol-rich ternary mixture, (20:20:60), provides almost all the spectral information about cluster X extracting eight of its nine peaks.

Clusters III, IV, VIII and X (total of 66 spectra) have strong absorptions in the 400 nm region that could indicate the presence of chalcones. Their spectral profiles are also indicative of phenolic substances including acids and catechins that are extracted with a variety of mixtures[17]. Also according to the literature[18] most flavonoids show a band in the 219 to 290 nm region (commonly called band II) and another band between 320-380 nm (band I). Also these results are consistent with literature data for these plant species for which mainly flavones and flavonoids are found[19-24].

The III, VIII and XVII clusters, containing 95 spectra, have absorbance bands in the 240-300 nm region that could indicate the presence of flavonol 3-O-glycosides. These bands are strongest in the seventeenth cluster and can be seen in Fig. 3.

Clusters VI and XIII, correspond to only 19 spectra but have two important spectral absorptions in the 250-285 nm (band II) and 330-385 nm (band I) regions indicating the presence of flavones or flavanols. Figure 4 contains the spectral profiles of the sixth cluster.

Figure 2c contains the three average spectra for clusters I, XV and XVI that show significant absorption bands in the 200-250 nm interval but have little or no absorption above 300 nm. Strong absorption within 220-250 nm indicates the existence of two conjugated unsaturated bonds such as a conjugated diene or an α,β-unsaturated aldehyde or ketone group in the molecular structure. Their spectral profiles are indicative of alkaloids, polyketides or terpenoids[15].

Further study of the compounds and spectral profiles indicated by these clusters is warranted. Appropriate solvent mixture compositions for open column elution of the material of each cluster can be expected to be obtained using a weighted average of all the mobile phase compositions in the experimental design where the weighting factors are the number of peaks obtained with each mobile phase. For the third cluster this weighted average is (2, 27, 33 and 38), essentially a mixture of dichloromethane, ethyl acetate and methanol in similar proportions. Examination of the results in Table 1 shows this composition is reasonble because 29 of the 32 spectra of cluster III were eluted by mobile phases without the presence of hexane. Furthermore the ternary mixture in equal proportions of these compounds extracts half of the peaks of cluster III.

This weighted average for cluster VIII is (30, 0, 44, 26). All its 14 peaks are eluted by mixtures without dichloromethane. The seventeenth cluster has a weighted average composition of (32, 15, 27, 26), with appreciable presence of all solvents, since its peaks are eluted with almost all the mobile phases. This composition is close to the average of the (33, 0, 33, 33) and (25, 25, 25, 25) mixtures that provide the largest number of peaks for cluster XVII.

4. Conclusions

K-means clustering discriminated 17 characteristic spectral profiles from a data set containing 3033 spectra. Of these, seven clusters (II, V, VII, IX, XI, XII and XIV), having 2617 spectra or 86.3% of the total, have little or no absorbance bands at wavelengths greater than 250 nm. These spectra are characterized by a very strong absorption band around 210 nm that is common to many naturally occurring compounds and shoulders or much weaker bands around 225 and 250 nm. In spite of containing only 134 spectra, or 4.4 % of the total, seven clusters (III, IV, VI, VIII, X, XIII and XVII) are clearly discriminated by k-means clustering and merit more detailed examination for metabolite presence. Finally three groups (I, XV and XVI) have spectral profiles with absorbances beyond 250 nm that are intermediate between those of the other two. As such k-means clustering techniques appear to provide a promising way to focus on interesting features in large spectral data sets.

Mixture design elution in column chromatography has been shown to be an attractive alternative strategy to the classical method of using pure solvents and their binary mixtures in order of increasing polarity. Ternary and quartenary mixtures,especially the (33:33:33 v/v/v) and (20:20:60 v/v/v) ones, provided new and potentially interesting

information for fingerprinting *Bauhinia variegata* L. extracts relative to what could be obtained using only pure solvents and binary mixtures. Furthermore mixtures with different compositions but essentially equal polarities can elute very different substances. As such the application of mixture designs and k-means clustering should be useful in the separation of the chemical constituents of any complex mixture.

The data set treated here (matrix with 3033 x 211 dimensions) uses only spectral information to determine charcteristic spectral groups classified by k-means clustering and the number of chromatographic peaks belonging to each of these groups. In fact a more complete data set would include all the chromatographic elution information and not just the number of peaks. As such a 95x211x611 data cube, where each of the 95 chromatograms has m digitized peak heights as a function of retention time along with the 211 spectral intensities as a function of wavelength, could be analyzed using higher order techniques such as multivariate curve resolution-alternating least squares[25], PARAFAC and Tucker3 methods[26,27] providing potentially more valuable chemical information than that obtained here.

Conflict of interest

The authors declare no competing financial interests.

Acknowledgements

This work was support by the Brazilian granting agencies CNPq, CAPES/PROAP, and Fundação Araucária.

References

1 FDA Guidance for Industry Botanical Drug Products (Draft Guidance), US Food and Drug Administration, Rockville, 2000, pp. 18-22.

2 Note for Guidance on Quality, of Herbal Medicinal Products, European Medicines Agency, London, 2001, pp. 6.

3 WHO, Guidelines for the Assessment of herbal Medicines, World Health Organization, Munich, 1991.

4 D. W. Choi, J. H. Kim, S. Y. Cho, D. H. Kim, S. Y. Chang, Regulation and quality control of herbal drugs in Korea, *Toxicology*, 2002, **181-182**, 581-586.

5 P. Vallat, W. Fan, N. El Tayar, P. A. Carrupt, B. Testa, Solvatochromic analysis of the retention mechanism of two novel stationary phases used for measuring lipophilicity by RP-HPLC, *J. of Liq. Chromatogr.*, 1992, **15**, 2133–2151.

6 P. J. Schoenmkakers, A. C. J. H. Drouen, H. A. H. Billiet, L. De Galan, A simple procedure for the rapid optimization of reversed-phase separations with ternary mobile phase mixtures, *Chromatographia.*, 1982, **15**, 688-696.

7 F. Delaroza, I. S. Scarminio, Mixture design optimization of extraction and mobile phase media for fingerprint analysis of *Bauhinia variegata* L., *J. Sep. Sci.*, 2008, **31**, 1034-1041.

8 A. A. de Almeida, I. S. Scarminio, Statistical mixture design optimization of extraction media and mobile phase compositions for the characterization of green tea, *J. Sep. Sci.*, 2007, **30**, 414-420.

9 C. N. Borges, M. C. Breitkreitz, R. E. Bruns, L. M. C. Silva, I. S. Scarminio, Unreplicated split-plot mixture designs and statistical models for optimizing mobile

chromatographic phase and extraction solutions for fingerprint searches, *Chemom. Intell. Lab Syst*., 2007, **89**, 82-89.

10 C. N. Borges, R. E. Bruns, A. A. de Almeida, I. S. Scarminio, Mixture-mixture design for the fingerprint optimization of chromatographic mobile phases and extraction solutions for *Camellia sinensis*, *Anal. Chim. Acta*, 2007, **595**, 28-37.

11 D. X. Soares, I. S. Scarminio, R. E. Bruns, Mixture designs for exploring class diversity and metabolite fingerprinting: An efficient column chromatographic strategy, *Anal. Chim. Acta.*, 2011, **702**, 288-294.

12 R. E. Bruns, I. S. Scarminio, B. B. Neto, Statistical Design-Chemometrics, Elsevier, Amsterdam, 2006, pp. 313-364.

13 A. C. Rencher, Methods of Multivariate Analysis, second ed., Edition, John Wiley & Sons, Inc, NY, 2002, pp. 451-497.

14 W. J. H Liun, Traditional Herbal Medicine Research Methods: Identification, Analysis, Bioassay and Pharmaceutical and Clinical Study, New Jersey, USA, 2011.

15 S. M. Colegate, R. J. Molyneux, Bioactive Natural Products: Detection, Isolation, and Structural Determination. New York, USA, 2008.

16 M. W. Hajnos, J. Sherma, High Performance Liquid Chromatography in Phytochemical Analysis, New York, 2011, pp. 907.

17 J. M. Peñarrieta, J. A. Alvarado, B. Akesson and B. Bergenståhl, Separation of phenolic compounds from foods by reverse-phase high performnce liquid chromatography, *Rev. Boliviana de Quím.*, 2007, **24**, 1-4.

18 X. Yang, L. Yang, A. Xiong, D. Li, Z. Wang, Authentication of *Senecio scandens* and *S. vulgaris* based on the comprehensive secondary metabolic patterns gained by UPLC–DAD/ESI-MS, *J. Pharm. Biomed. Anal.*, 2011, **56**, 165-172.

19 M. V. B. Reddy, M. K. Reddy, D. Gunasekar, C. Caux, B. Bodo, A flavanone and a dihydrodibenzoxepin from *Bauhinia variegata*, *Phytochem.*, 2003, **64**, 879-882.

20 M. F. Ramadan, G. Sharanabasappa, Y. N. Seetharam, M. Seshagiri, J-T Moersel, Characterisation of fatty acids and bioactive compounds of kachnar (*Bauhinia purpurea* L.) seed oil, *Food Chem.*, 2006, **98**, 359-365.

21 H. P. Bhartiya, P. C. Gupta, A chalcone glycoside from the seeds of *Bauhinia purpurea*, *Phytochem*., 1981, **20**, 2051.

22 R. N. Yadava, P. Tripathi, A novel flavone glycoside from the stem of *Bauhinia purpurea*, *Fitoterapia*, 2000, **71**, 88-90.

23 H. Achenbach, M. Stöcker, M. A. Constenla, Flavonoid and other constituents of *Bauhinia manca*, *Phytochem.*, 1988, **27**, 1835-1841.

24 Y-H Kuo, M-H Yeh, S-L Huang, A novel 6-butyl-3-hydroxyflavanone from heartwood of *Bauhinia purpurea*, *Phytochem.*, 1998, **49**, 2529-2530.

25 A. Olivieri,G. M. Escandar, Practical Three-Way Calibration, Elsevier, USA, 2014, 127-156.

26 P.M. Kroonenberg, Applied Multiway Analysis, John Wiley & Sons Ltd., New York, 2008, 43-106

27 A.K. Smilde, R. Bro, P. Geladi, Multi-way Analysis, John Wiley & Sons Ltd., New York, 2005, 89-144

Figure captions

Figure 1. Mixture design with pure, binary, ternary and quaternary mixtures for mobile phases for open column chromatography.

Figure 2. Average spectra of each of the 17 clusters using k-means clustering: a) II, V, VII, IX, XI, XII and XIV, b) III, IV, VI, VIII, X, XIII and XVII, c) I, XV and XVI. The spectra were obtained by RP-HPLC-DAD with the (30:35:30, v/v/v) water: MeOH: ACN mobile phase.

Figure 3. Spectra of the seventeenth (XVII) cluster obtained by k-means clustering of the RP-HPLC-DAD data.

Figure 4. Spectra of the sixth (VI) cluster obtained by k-means clustering of the RP-HPLC-DAD data.

Table 1- Mobile phases, number of fractions and number of spectra in the each mobile phase in open column chromatography distributed in seventeen different groups

Mobile phases				NF[b]	Clusters																Total Peaks	Peaks Fig 2
h[a]	d	a	m		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII		
1	0	0	0	3	1	73			1		3				13	9		1		2	2	105	2
0.75	0.25	0	0	2		30					2				14	8	1	3		2	1	61	2
0.75	0	0.25	0	11	5	194		1	12	1	14	2	4		58	82				4	2	379	6
0.50	0.50	0	0	2		17					2				16	17					2	54	2
0.60	0.20	0	0.20	4	1	49			8		3		1		20	9	1			1	2	95	3
0.60	0	0.20	0.20	4	1	63			10		4		1		28	19		3	2	2	1	134	1
0.50	0	0.50	0	3		54			1		3	4	3		21	15		1		5	2	109	6
0.33	0.33	0.34	0	5	1	52					4				15	10		3		5	3	93	3
0.50	0	0	0.50	3	11	27			4		8		3		12	19	2	1	1	3	4	95	6
0.33	0.33	0	0.34	2	7	13			2		18				6	3			3	2		54	0
0.20	0.60	0.20	0	1		9	3		2		1				5	6	2		2	4		34	5
0.20	0.60	0	0.20	5	5	41			4		19				8	8	1	2		4	2	94	3
0	1	0	0	3		59			4		6				9	19	1	7	3	12		120	1
0.33	0	0.33	0.34	5	26	56		4	9		20	2	2		12	32	1	2	3	5	8	182	15
0.25	0.25	0.25	0.25	7	4	73			11		23				18	16	1			6	5	157	6
0.20	0.20	0.60	0	1	2	5			4		1				13	1		1	1		2	30	2
0.20	0	0.60	0.20	1		9			2		1				5	5	2		2	4		30	2
0.20	0.20	0	0.60	1	1	17					1				4	9	3	1	2	3		41	3
0	0.50	0.50	0	1		23			1		1				5	4				1		35	0
0	0.60	0.20	0.20	1		16	1	4			1		3		7	7				1	1	41	6
0.20	0	0.20	0.60	1		29					1				4	2				1		37	0
0	0.50	0	0.50	1	1	15			3		1				8	10				2		40	0
0	0.33	0.33	0.34	1	5		16		4	2	12			1								40	19
0	0.20	0.60	0.20	8	18	123			7		8		3		36	54	1		4	24	2	280	3
0	0	1	0	5	10	81			5		39				31	29		4	1	10	2	212	2
0	0.20	0.20	0.60	5	9	68	5	1	7		6		2	8	25	27		5	4	8	3	178	17
0	0	0.50	0.50	6	18	75	7	1	4		35	6	4		21	19		3	2	10	3	208	17
0	0	0	1	3	0	33					4		20		12	19				5	2	95	2
Total				**95**	**126**	**1304**	**32**	**11**	**105**	**3**	**241**	**14**	**46**	**9**	**426**	**458**	**16**	**37**	**30**	**126**	**49**	**3033**	**134**

[a] h=hexane, d=dichloromethane, a=ethyl acetate and m=methanol

[b]Number of fractions

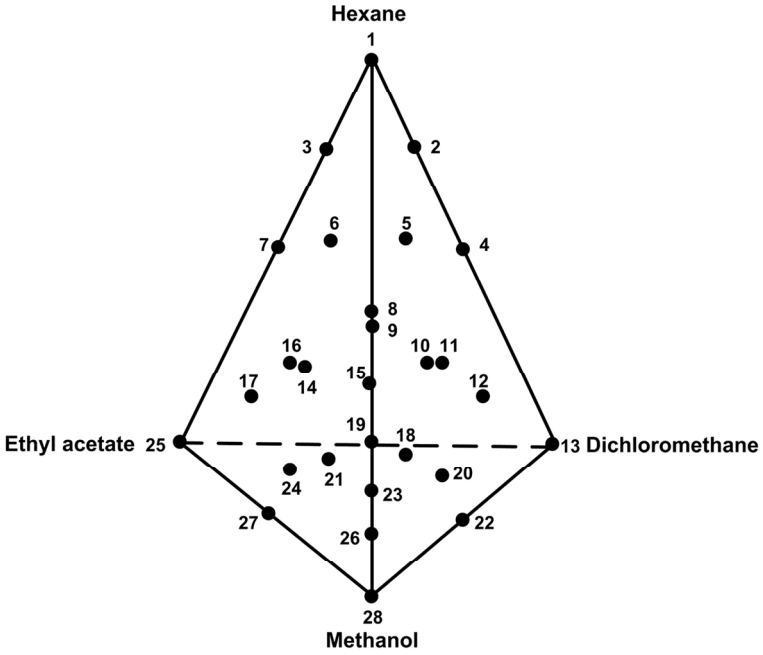


Figure 1. Mixture design with pure, binary, ternary and quaternary mixtures for mobile phases for open column chromatography.
152x100mm (300 x 300 DPI)

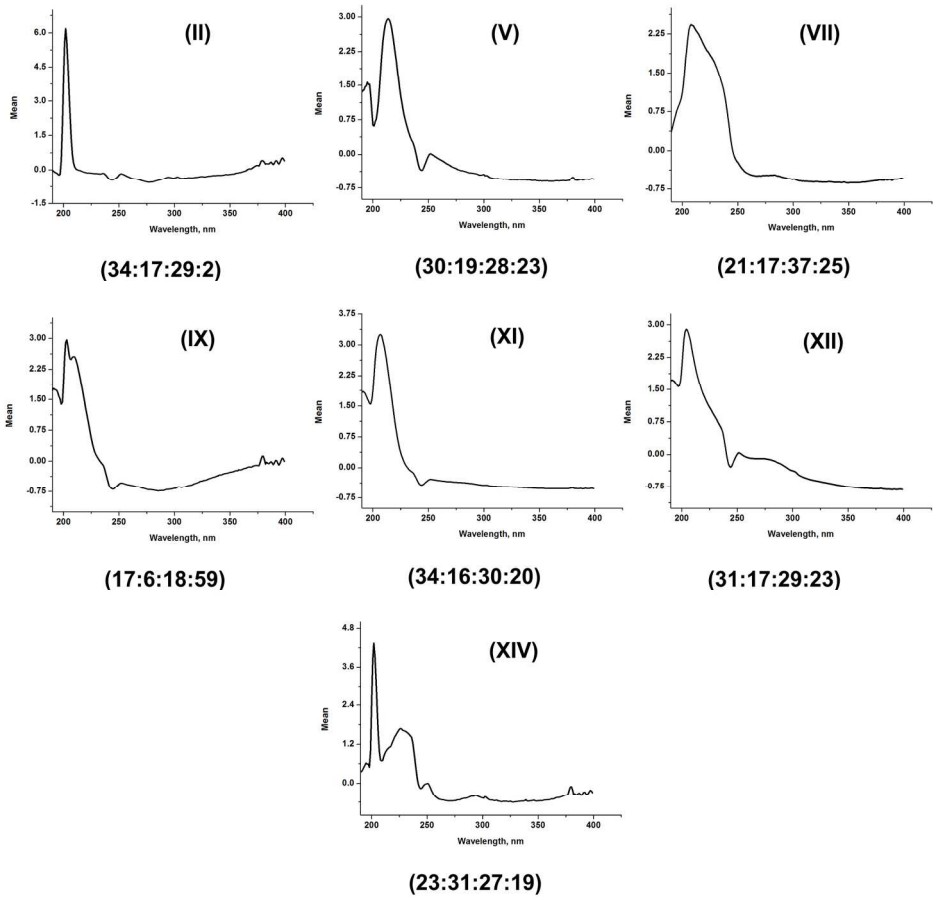


Figure 2. Average spectra of each of the 17 clusters using k-means clustering: a) II, V, VII, IX, XI, XII and XIV, b) III, IV, VI, VIII, X, XIII and XVII, c) I, XV and XVI. The spectra were obtained by RP-HPLC-DAD with the (30:35:30, v/v/v) water: MeOH: ACN mobile phase.
179x165mm (300 x 300 DPI)

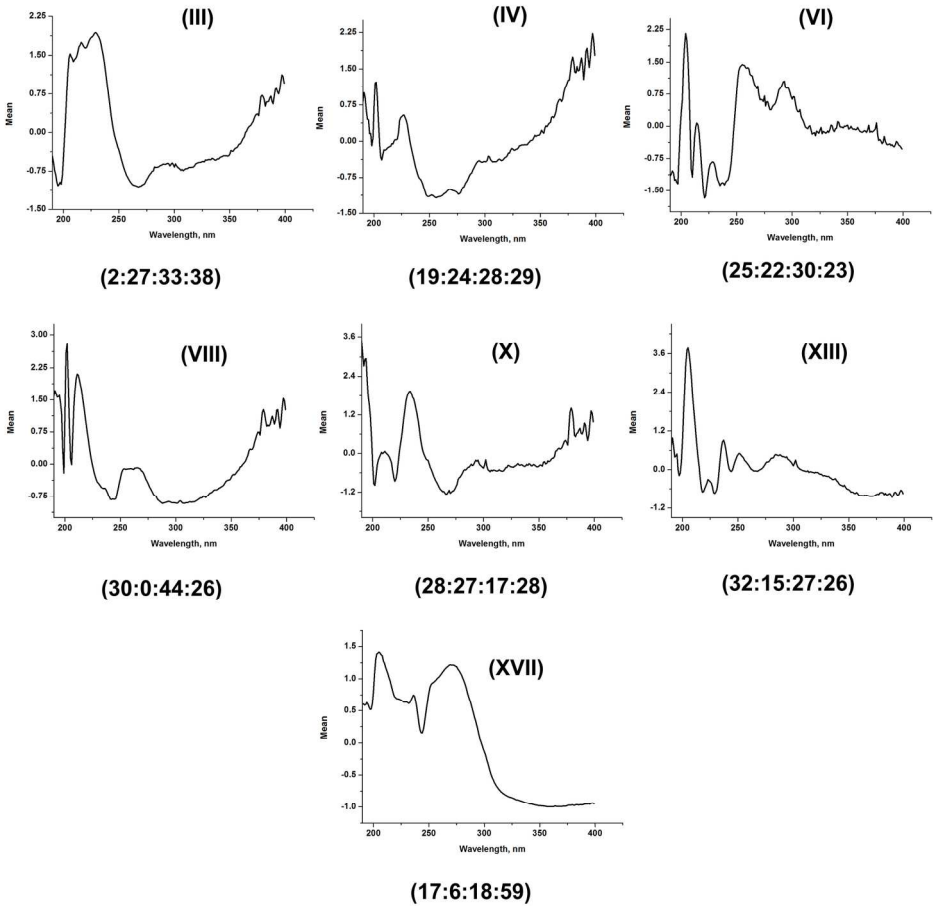


Figure 2b
179x164mm (300 x 300 DPI)

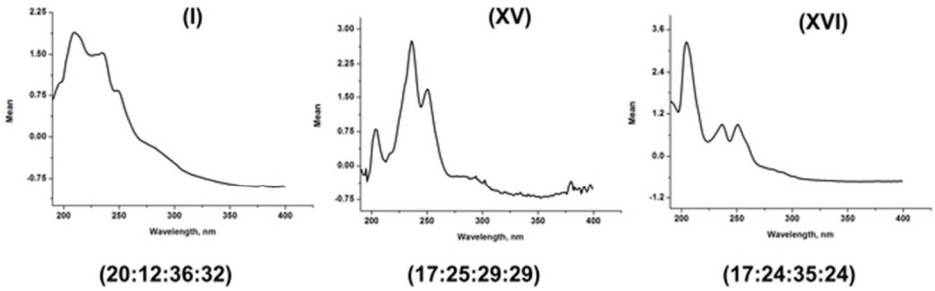


Figure 2c
59x18mm (300 x 300 DPI)

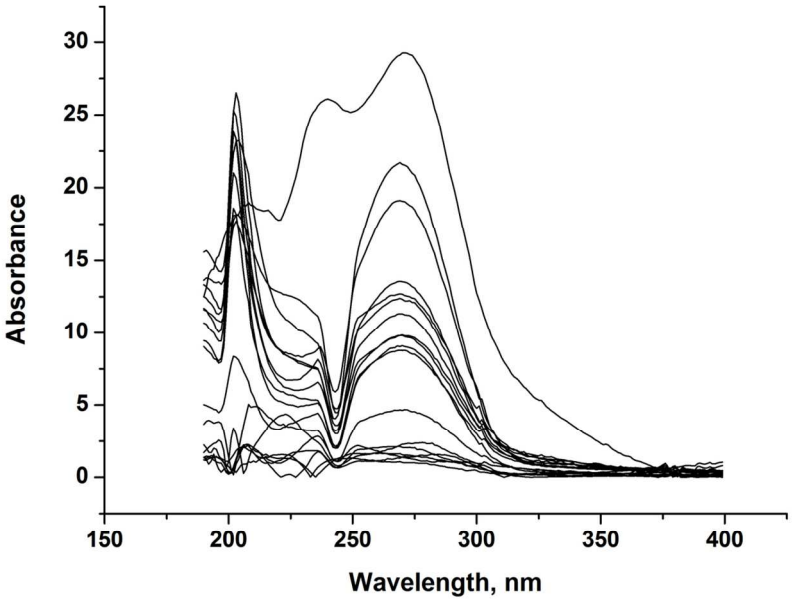


Figure 3. Spectra of the seventeenth (XVII) cluster obtained by k-means clustering of the RP-HPLC-DAD data.
125x88mm (300 x 300 DPI)

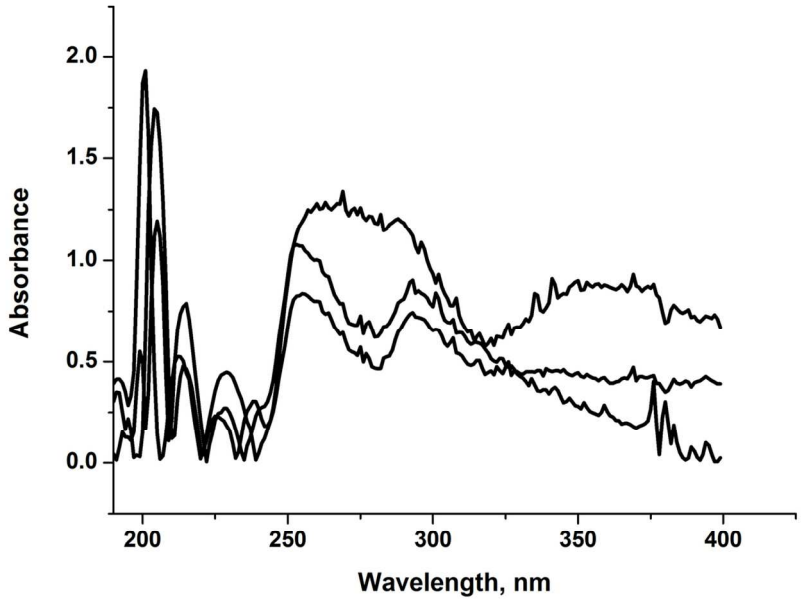


Figure 4. Spectra of the sixth (VI) cluster obtained by k-means clustering of the RP-HPLC-DAD data.
125x88mm (300 x 300 DPI)