Phytochemical Analysis
Phytochem. Anal. **18**: 123–132 (2007)
Published online 15 December 2006 in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/pca.960

Phytochemical Analysis

A New Method for the Authentication of Plant Samples by Analyzing Fingerprint Chromatograms

MARKO OBRADOVIČ,[1]* SIMONA STRGULC KRAJŠEK,[2] MARINA DERMASTIA[2] and SAMO KREFT[1]

[1] Faculty of Pharmacy, University of Ljubljana, Aškerčeva 7, SI-1000 Ljubljana, Slovenia
[2] Department of Biology, Biotechnical Faculty, University of Ljubljana, Večna pot 100, SI-1000 Ljubljana, Slovenia

Received 12 April 2005; Revised 24 August 2006; Accepted 30 August 2006

Abstract: Chemical analysis by high-performance liquid chromatography or capillary electrophoresis of plant pulverized samples, juices or extracts is an excellent method for the authentication of medicinal plant species and their products, particularly when morphological authentication is not possible. In the conventional procedure, chromatograms are integrated and the heights or areas of several peaks are used in a supervised pattern recognition method to confirm the authenticity of the product. We propose a new section approach in analysing chromatograms, where chromatograms are split into sections, which are described by four variables (number of peaks in the section, average retention time of peaks in the section, total area of peaks in the section and average area of peaks in the section), and these variables are then used in statistical analysis. The method is especially useful when the peaks on the chromatogram are not well separated and it is not easy to link individual peaks on one chromatogram with corresponding peaks on other chromatograms. In comparison with the standard procedure, our approach in analyzing chromatographic data of willow-herb (*Epilobium* and *Chamaenerion* spp.) extracts was more objective, gave better results and was also easier to perform. Copyright © 2006 John Wiley & Sons, Ltd.

Keywords: Capillary electrophoresis; chromatogram; plant sample authentication; multivariate analysis; *Epilobium*; *Chamaenerion*.

INTRODUCTION

A reliable pharmacological or clinical study must employ well-authenticated plant material. High-performance liquid chromatography (HPLC), capillary electrophoresis (CE) or thin-layer chromatography (TLC) are the most commonly used analytical methods for the chemical authentication of medicinal products (Barakat *et al.*, 1997; Blanquer *et al.*, 1998; Schaneberg *et al.*, 2003; Xie *et al.* 2006), food products (Goodall *et al.*, 1995; Gonzáles *et al.*, 2001; Alonso-Salces *et al.*, 2004) and in the manufacturer of an individual product (Welsh *et al.*, 1996). If the peaks on chromatograms are well separated and retention times are very reproducible, such that corresponding peaks on different chromatograms unambiguously represent the same substance, a plant sample can be authenticated by a chemical fingerprinting method based on the presence or absence of a limited number of peaks (Schaneberg *et al.*, 2003). In such cases no statistical procedure is required and it is not necessary for the substances represented by these peaks to be known. For the differentiation of a large number of similar samples by multivariate statistical analysis, discriminant analysis was found to be efficient (Goodall *et al.*, 1995; Gonzáles *et al.*, 2001). In this approach a large number of variables (peaks) are required. As with the limited peak number method, there must be no doubt that an individual peak on a chromatogram represents the same substance as the equivalent peak on another chromatogram, but again, the substances represented by these peaks need not be known. Unambiguous identification of all peaks in all chromatograms can be a difficult task, even using retention times, absorption spectra and multiple internal standards. There is always some bias involved in the selection of peaks. Frequently retention times can vary by perhaps up to 1 min, which is sometimes equal to the distance between peaks (Goodall *et al.*, 1995).

Willow-herbs (*Epilobium*), which are frequently used in folk medicine, were used in the present study as model plants. Pharmacological studies of willow-herb extracts have confirmed their antimicrobial (Battinelli *et al.*, 2001), anti-inflammatory (Hiermann *et al.*, 1991), analgesic (Tita *et al.*, 2001) and anti-tumour (Voynova *et al.*, 1991) activities. The retarding effect of willow-herbs on the growth of epithelial cells of human prostate was also proven (Ducrey *et al.*, 1997; Vitalone *et al.*, 2001).

The 16 willow-herb species growing in Slovenia are members of the genera *Chamaenerion* and *Epilobium* (Ravnik, 1999; Strgulc Krajšek and Jogan, 2006). The genus *Chamaenerion* comprises species *C. angustifolium* and *C. palustre*. The genus *Epilobium* is divided into two sections (Haussknecht, 1884): section Schizostigma comprising species *E. hirsutum*, *E. parviflorum*, *E. montanum* and *E. collinum*; and

* Correspondence to: M. Obradovič, Faculty of Pharmacy, University of Ljubljana, Aškerčeva 7, SI-1000 Ljubljana, Slovenia.
E-mail: marko.obradovic@ffa,uni-lj.si

section Systigma comprising species *E. tetragonum*, *E. lamyi*, *E. alpestre*, *E. nutans*, *E. alsinifolium*, *E. anagallidifolium*, *E. palustre*, *E. obscurum*, *E. ciliatum* and *E. roseum*.

In the present study we have investigated the reliability of a new approach for transferring chromatographic data into a form suitable for statistical analysis. In the willow-herb model, this descriptive approach provided more objective data for reliable authentication systems and better classification of species than the conventional approach involving the identification of chromatographic peaks.

EXPERIMENTAL

Apparatus. CE analysis was performed on a Hewlett-Packard (Waldbronn, Germany), HP 3D CE apparatus using HP ChemStation software. The fused-silica uncoated capillary (Agilent, Waldbronn, Germany), with a bubble cell, was 64.5 cm long with an internal diameter of 50 μm.

Samples. Forty-six herbarium sheets comprising 16 willow-herb species that grow in Slovenia were analysed. Two independent leaf samples were taken from each sheet. Voucher specimens are deposited in the Herbarium LJU of the Department of Biology, University of Ljubljana.

Analytical procedures. Optimized conditions for extraction (i.e. solvent, time and temperature) were determined and employed. Separation was optimized according to Wäetzig *et al.* (1998) and Terabe (2004). The following parameters were varied: pH of the borate buffer, concentrations of borate and of sodium dodecyl sulphate (SDS), sample injection volume, temperature of the capillary and voltage. Organic modifiers (methanol or isopropanol), complexing reagents (EDTA, Cu), cyclodextrins or a co-surfactant (sodium cholate) were added to the separation buffer in order to test for improvement of separation.

Extracts were prepared and purified as follows: 10 mg of leaf were pulverized using a drill with a modified round-top metal probe. A 1 mL aliquot of 10% aqueous methanol was added and the suspension was exposed to ultrasound for 5 min, followed by 10 min shaking at 250 rpm and a further 5 min of ultrasonic extraction. Samples were then centrifuged for 4 min at 11000 ***g*** force and the supernatants filtered through cellulose filters (0.45 μm pore size).

Samples were analysed by CE using an injection time of 20 s and pressure of 20 mbar. The separation buffer was 200 mM borate buffer, pH 9.3 and 100 mM SDS. The voltage was set to 20 kV, the capillary temperature to 30°C and the time of analysis was 40 min. Absorbance was measured at 200 nm.

Data analysis. Peaks in each chromatogram were integrated. The retention time for each peak was calculated relative to the peak that was unambiguously identified in all chromatograms (the highest peak). The area of each peak was expressed as a percentage of the total peak area. Three approaches were used to depict chromatograms.

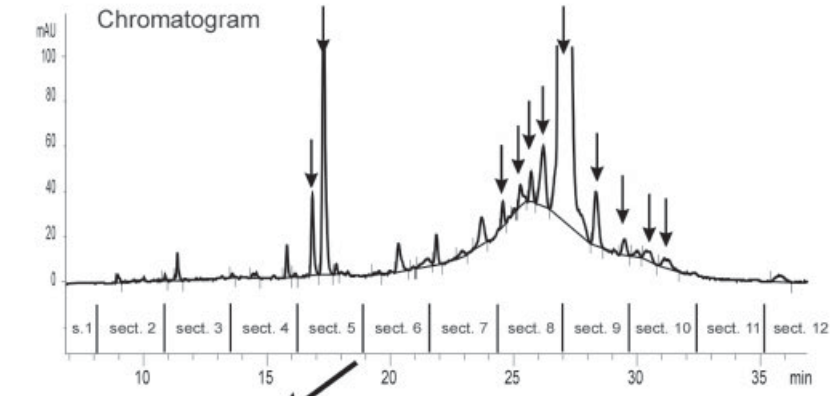
In the first approach we selected 11 peaks that were identified in all but two of the chromatograms. Peaks were identified with a relatively high accuracy by retention times (and relative retention times) and absorption spectra. The percentages of the total peak area for each of the 11 identified peaks were used for statistical analysis. This method is referred to as the 'conventional approach' hereinafter.

In the second approach additional variables for statistical analysis were created by mathematical depiction of the areas between previously determined peaks. These areas included peaks that could not have been unambiguously determined in each chromatogram, but were expected to carry additional information. Each area between two peaks was described by four variables, namely, the number of peaks, average relative retention time, sum of the percentages of the total peak area, and average percentage of total peak area. Calculated variables were added to the percentages of the total peak area for each of the 11 previously identified peaks and further used in statistical analysis. This approach is referred to as the 'additional-variables approach' hereinafter.

The third approach was to split the chromatographic area into several sections. An iterative procedure was employed whereby the number of sections was varied from 7 to 57. In the first variant all sections were of an equal width. In the second variant, the first and the last sections were broader whilst the central part was split to equal sections since most of the peaks were eluted between the 10th and 35th min. Each section was described by four variables, number of peaks, average relative retention time, sum of percentage of total peak area and average percentage of total peak area. This procedure was performed equally for every chromatogram using the set of functions available in Microsoft Excel® (the template employed is available from the authors upon request). Variables attained by this sectioning were further used in the statistical analysis. This method is referred to as the 'section approach' hereinafter.

Ninety-two chromatograms (two independent samples from each of 46 herbarium sheets) were processed as described above (Fig. 1) and variables were analysed statistically by linear discriminant analysis (LDA) using SPSS (Chicago, II, USA) software package. Variables for discriminant functions were chosen by a *stepwise method*. The accuracy of the sample classification was estimated by the *leave-one-out method* (prediction ability), i.e. the analysis was repeated omitting one of

DOI: 10.1002.pca



Integration results

Relative retention time (rRt)	Area %
0,332762345	0,141956
0,404669695	0,074008
0,423251657	0,41171
0,506069859	0,059843
0,54260073	0,163007
0,588627393	0,497946
0,601549117	0,074428
0,627057422	1,441583
0,644075371	6,671103
0,662992478	0,17114
0,727936248	0,11309
0,757093915	0,861942
0,779288002	0,051626
0,800476652	0,46571
0,814217621	0,626788
0,85357861	0,257299
0,882587324	1,013279
0,914612348	0,457885
0,931257913	0,083176
0,941200566	0,579534
0,957324793	0,65613
0,975497133	2,248775
1	78,84334
1,055112832	1,755307
1,098309377	0,503065
1,117673345	0,236843
1,136031876	0,518557
1,159380353	0,565951

Identified peaks

	Area %
Peak 1	1,44158281
Peak 2	6,671102895
Peak 3	0,45788484
Peak 4	0,579533985
Peak 5	0,656129796
Peak 6	2,248774748
Peak 7	78,84333509
Peak 8	1,755307031
Peak 9	0,503064984
Peak 10	0,236843233
Peak 11	0,518557473

depiction of chromatogram divided into section

	No. of peaks	Average rRt	Sum Area %	Average Area %
sect.1	0	0	0	0
sect.2	1	0,332762345	0,141956338	0,141956338
sect.3	2	0,413960676	0,485717674	0,242858837
sect.4	3	0,545765994	0,720795496	0,240265165
sect.5	4	0,633918597	8,358253727	2,089563432
sect.6	3	0,754772722	1,026658525	0,342219508
sect.7	4	0,837715052	2,363076488	0,590769122
sect.8	6	0,953315459	82,86883439	13,8114724
sect.9	2	1,076711104	2,258372016	1,129186008
sect.10	3	1,137695191	1,32135137	0,440450457
sect.11	0	0	0	0
sect.12	1	1,332650629	0,454983979	0,454983979

Figure 1 Schematic presentation of the conventional and section approaches for depicting chromatograms. In both approaches, chromatograms are integrated. In the conventional approach, the peaks are then identified and their areas used for further statistical analysis. In the section approach, the chromatogram is divided into sections (in this case 12 sections) and each section is described by four variables, namely, the number of peaks in the section, the average relative retention time of peaks in the section, the sum of the peak area percentages of all peaks in the section and the average peak area percentage.

the samples at a time from the learning procedure and the obtained discriminant functions used to classify this 'unknown' sample.

RESULTS AND DISCUSSION

Analysis of the obtained chromatograms of willow-herb leaf extracts was performed by conventional, additional-variables and section approaches. Schematic presentations of the sample classification obtained using all approaches are presented in Tables 1–3. The accuracy of predicting the identity of one of the 16 species by chance would be 6.3%, but the conventional approach with 11 identified peaks resulted in a 46% accuracy of classification. As expected, the additional-variables approach produced a higher accuracy (50%) resulting from the additional information carried by the unidentified peaks. In some cases, if a better resolution and a higher repeatability of retention times can be achieved,

DOI: 10.1002/pca

Table 1 Schematic presentation of the classification of extracts of *Chamaenerion* and *Epilobium* species using the conventional approach for chromatogram depiction. The number in each field represents the percentage of the total cases listed in the rightmost column. Related species of *Epilobium* are placed together according to Haussknecht (1884), Ravnik (1999) and Strgulc Krajšek and Jogan (2006)

	C. angustifolium	*C. palustre*	*E. montanum*	*E. collinum*	*E. parviflorum*	*E. hirsutum*	*E. tetragonum*	*E. lamyi*	*E. nutans*	*E. palustre*	*E. anagallidifolium*	*E. alsinifolium*	*E. obscurum*	*E. alpestre*	*E. roseum*	*E. ciliatum*	Total
C. angustifolium	**100**																**4**
C. palustre		**75**				25											**4**
E. montanum			**33.3**	33.3						16.7					16.7		**6**
E. collinum				**80**		20											**5**
E. parviflorum		10			**50**		10			10			10		10		**10**
E. hirsutum					60	**40**											**5**
E. tetragonum					14.3		**0**	85.7									**7**
E. lamyi							66.7	**0**	16.7							16.7	**6**
E. nutans					50				**0**		50						**2**
E. palustre			16.7		33.3					**33.3**						16.7	**6**
E. anagallidifolium											**100**						**2**
E. alsinifolium									11.1		22.2	**55.6**	11.1				**9**
E. obscurum										16.7	16.7	16.7	**33.3**		16.7		**6**
E. alpestre														**100**			**6**
E. roseum			33.3		33.3										**33.3**		**6**
E. ciliatum					16.7				16.7	16.7			16.7			**33.3**	**6**

DOI: 10.1002/pca

Table 2 Schematic presentation of the classification of extracts of *Chamaenerion* and *Epilobium* species using the additional variables approach for chromatogram depiction. The number in each field represents the percentage of the total cases listed in the rightmost column. Related species of *Epilobium* are placed together according to Haussknecht (1884), Ravnik (1999) and Strgulc Krajšek and Jogan (2006)

	C. angustifolium	*C. palustre*	*E. montanum*	*E. collinum*	*E. parviflorum*	*E. hirsutum*	*E. tetragonum*	*E. lamyi*	*E. nutans*	*E. palustre*	*E. anagallidifolium*	*E alsinifolium*	*E. obscurum*	*E. alpestre*	*E. roseum*	*E. ciliatum*	Total
C. angustifolium	**66.7**	33.3															**4**
C. palustre		**50**				25	25										**4**
E. montanum			**16.7**	16.7										16.7	33.3	16.7	**6**
E. collinum				**75**	25												**5**
E. parviflorum		10			**70**			10							10		**10**
E. hirsutum					60	**40**											**5**
E. tetragonum					14.3		**57.1**	14.3			14.3						**7**
E. lamyi					16.7			**66.7**	16.7								**6**
E. nutans								100	**0**								**2**
E. palustre			33.3		16.7					**16.7**	16.7					16.7	**6**
E. anagallidifolium							50				**50**						**2**
E. alsinifolium				11.1	11.1							**56.6**	11.1		11.1		**9**
E. obscurum			20										**60**		20		**6**
E. alpestre				33.3										**66.7**			**6**
E. roseum			33.3				16.7	16.7							**33.3**		**6**
E. ciliatum							16.7			33.3					16.7	**33.3**	**6**

DOI: 10.1002/pca

Table 3 Schematic presentation of the classification of extracts of *Chamaenerion* and *Epilobium* species using the section approach for chromatogram depiction. The number in each field represents the percentage of the total cases listed in the rightmost column. Related species of *Epilobium* are placed together according to Haussknecht (1884), Ravnik (1999) and Strgulc Krajšek and Jogan (2006)

	C. angustifolium	*C. palustre*	*E. montanum*	*E. collinum*	*E. parviflorum*	*E. hirsutum*	*E. tetragonum*	*E. lamyi*	*E. nutans*	*E. palustre*	*E. anagallidifolium*	*E. alsinifolium*	*E. obscurum*	*E. alpestre*	*E. roseum*	*E. ciliatum*	Total
C. angustifolium	**100**																**4**
C. palustre		**75**					25										**4**
E. montanum			**83.3**													16.7	**6**
E. collinum			16.7	**83.3**													**6**
E. parviflorum				10	**70**		10									10	**10**
E. hirsutum					28.6	**66.7**		16.7					16.7				**6**
E. tetragonum					16.7		**71.4**										**7**
E. lamyi							50	**83.3**									**6**
E. nutans									**0**							50	**2**
E. palustre					16.7					**83.3**							**6**
E. anagallidifolium											**100**						**2**
E. alsinifolium					12.5							**62.5**	12.5			12.5	**8**
E. obscurum			16.7		16.7				16.7				**16.7**			33.3	**6**
E. alpestre											16.7			**83.3**			**6**
E. roseum			33.3	33.3											**33.3**		**6**
E. ciliatum					16.7							33.3	16.7			**33.3**	**6**

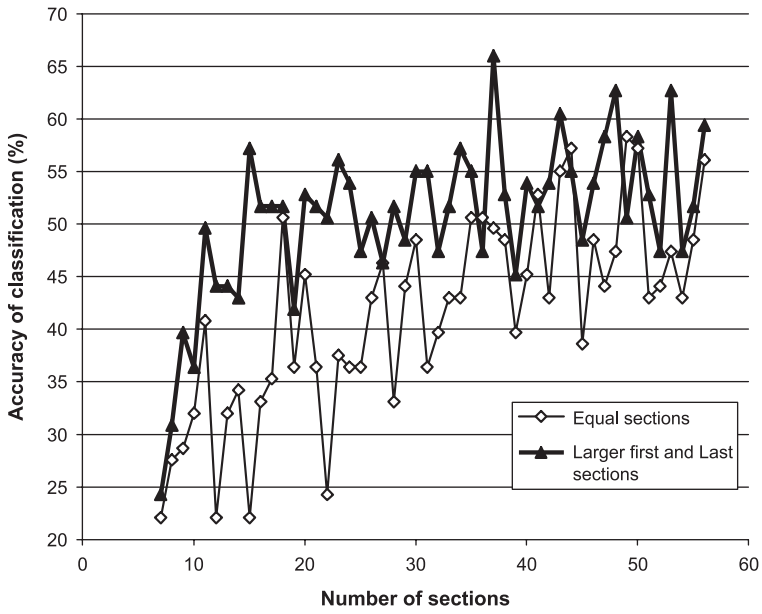


Figure 2 Accuracy of classification of species depends on the number of sections used in the depiction of the chromatogram. When whole chromatograms (0–60 min) were divided into increasing numbers of even sections, the accuracy of classification increased as the number of sections approached 55. Since most of the peaks eluted between the 10th and 35th minute, however, another approach was tested. The start and end sections of the chromatogram were considered as two large sections, and the central part of the chromatogram was divided into equal sections. In this approach, the accuracy of classification increased until the plateau was reached at 20 sections.

or if multiple internal standards are used, it is possible to determine larger numbers of corresponding peaks in each chromatogram. Our results show that the remaining undetermined peaks could provide additional information, which can be handled in the manner previously described. The accuracy of classification achieved with the section approach depended on the number of sections (Fig. 2). By considering more than 35 sections, an accuracy of >55% could have been reached with both variants of the section approach, the highest accuracy being 65%. It can be seen from Fig. 2 that the accuracy of both variant approaches becomes similar and reaches a plateau when the sections become sufficiently narrow. A larger number of sections are required in order to achieve sections that are narrow enough in the case where the sections are of equal across the whole chromatogram compared with the approach when the first and the last sections are wider. The consequence of the equal section approach was thus a slower increase in accuracy. The alternative approach using a larger first and last section was merely a way of reducing the number of variables and was valid since the majority of peaks were located between the two large sections. The border values of the two large sections could be defined arbitrarily. It is interesting to note that the accuracy of classification did not decrease if the number of sections was increased up to 57; however, in order to avoid optimization of number of sections for every situation, 57 sections can be used arbitrarily.

Linear discriminant analysis, which was used to determine species based on the chromatographic data, belongs to a group of supervised methods as it needs a training set for calibration (Hair *et al.*, 1987). This type of discriminant analysis was used in recently published papers in metabolomic studies (Bijlsma *et al.*, 2006; Lutz *et al.*, 2006). LDA was most suitable in our case owing to the relatively high number of species and small number of samples for each species. Principal component analysis and hierarchical cluster analysis, other frequently used multivariable techniques, were also tested, but proved to be less appropriate.

Willow-herb species served as model plants since it could be shown that the conventional approach of determining a limited number of peaks used as variables in statistical classification procedures was improved by using information from all of the peaks. The classification accuracy thus gained depended on how well individual peaks in one chromatogram can be linked with the corresponding peaks in other chromatograms, and this can be difficult in complex biological samples.

ANOVA was used to calculate whether the variables obtained by conventional and section chromatogram depiction methods differed among the willow-herb species. The list of variables that differed significantly ($p < 0.01$) among the species is presented in Table 4. Variables for the section approach relate to the splitting of the chromatographic area into 37 sections, with broader first and last sections. Such splitting resulted

DOI: 10.1002/pca

Table 4 List of variables that differed significantly (ANOVA, $p < 0.01$) among species. The variables that were used in the discriminant analysis are shown in bold

Conventional approach		Section approach (37 sections: first section from beginning to relative retention time 0.3; last section from relative retention time 1.3 to the end; central 35 sections all of equal width)		
Type of variable	Significance level	Type of variable	Section	Significance level
Peak 3	3.70×10^{-22}	**Average area percentage**	**25**	2.21×10^{-19}
Peak 6	7.17×10^{-20}	Sum area percentage	25	1.22×10^{-14}
Peak 2	1.90×10^{-17}	**Sum area percentage**	**26**	2.81×10^{-14}
Peak 7	3.15×10^{-14}	**Average area percentage**	**28**	2.81×10^{-9}
Peak 8	5.20×10^{-10}	**Average area percentage**	**24**	3.45×10^{-8}
Peak 1	1.17×10^{-5}	Sum area percentage	28	1.05×10^{-7}
Peak 5	1.01×10^{-4}	Average area percentage	26	1.51×10^{-7}
Peak 9	4.09×10^{-4}	**Average area percentage**	**22**	2.41×10^{-7}
Peak 10	3.31×10^{-3}	Sum area percentage	22	5.12×10^{-7}
Peak 11	4.17×10^{-3}	**Sum area percentage**	**24**	9.91×10^{-7}
		Average area percentage	**14**	2.04×10^{-6}
		Sum area percentage	**18**	1.17×10^{-5}
		Average area percentage	23	3.05×10^{-5}
		Sum area percentage	**11**	4.24×10^{-5}
		Sum area percentage	**12**	7.56×10^{-5}
		Number of peaks	12	0.000145
		Average area percentage	12	0.000183
		Sum area percentage	14	0.000294
		Average area percentage	18	0.00056
		Sum area percentage	**23**	0.000655
		Sum area percentage	**30**	0.001085
		Number of peaks	11	0.00225
		Average area percentage	**11**	0.002885
		Number of peaks	**29**	0.003006
		Number of peaks	28	0.007987
		Average rRt	12	0.009319
		Sum area percentage	29	0.009695

in the highest accuracy score. In the conventional approach, 10 peaks out of the 11 differed significantly, the *p*-values for the first three most significant peaks being less than 10^{-15}. On the other hand, as many as 27 variables obtained by the section approach differed significantly among the species, although only one variable in this approach reached a *p*-value as low as 10^{-15}. However, there were almost three times more significant variables depicted with this approach, which might be the reason for better species classification. At a significance level of 0.05, the remaining peak in the conventional approach was significant, but 24 additional variables of the section approach became significant as well.

The correct combination of variables is crucial for discriminant analysis. Discriminant analysis first chooses the most discriminant variable and the selection of further variables is based on the previously selected variables. These variables are used in discriminant functions, which are linear combinations of all selected variables. In this regard, the variables that are in strong correlation with the included variables would bring no additional information. Therefore such variables were excluded by the analysis, although some of them differed significantly among species. The variables that were included in the discriminant functions are presented in Table 5.

The separation of samples with CE or HPLC may result in complicated chromatograms, and selecting several peaks that are identifiable in different chromatograms can be a difficult task. This becomes an even bigger problem in thin-layer chromatography where the use of internal standards in this technique is limited, as they can affect the resolution of separation. The results of our study show that the conventional approach to sample authentication, depicting chromatograms by defining several peaks, provides a lower accuracy in classification of samples compared to splitting chromatograms into sections and using mathematical variables to describe each section. A higher number of sections produce better classification accuracy. Moreover, this novel section approach also gives more objective results and it is much easier to perform than the conventional approaches.

DOI: 10.1002.pca

Table 5 List of variables that were included in the discriminant functions for the classification of willow-herb species by discriminant analysis. Variables shown in bold are those that differed significantly among species according to the ANOVA test (Table 4). Variables are listed in order of inclusion in the discriminant functions: the most discriminant variables are at the top of the list

	Section approach (37 sections: first section from beginning to relative retention time 0.3; last section from relative retention time 1.3 to the end; central 35 sections all of equal width)	
Conventional approach	Type of variable	Section
Peak 3	**Average area percentage**	**25**
Peak 2	**Sum area percentage**	**26**
Peak 8	**Average area percentage**	**28**
Peak 6	**Sum area percentage**	**12**
Peak 7	Sum area percentage	16
Peak 1	**Sum area percentage**	**18**
Peak 4	Average rRt*	3
	Average area percentage	**22**
	Sum area percentage	**23**
	Sum area percentage	3
	Average rRt*	19
	Average rRt*	14
	Sum area percentage	4
	Sum area percentage	**24**
	Sum area percentage	2
	Average area percentage	19
	Sum area percentage	**30**
	Sum area percentage	**11**
	Average area percentage	20
	Average rRt*	26
	Sum area percentage	6
	Average area percentage	**14**
	Sum area percentage	15
	Sum area percentage	36
	Average area percentage	7
	Number of peaks	36
	Average area percentage	13
	Average rRt*	4
	Average rRt*	31
	Average area percentage	**11**
	Average area percentage	**24**
	Number of peaks	**29**

*rRt = relative retention time.

REFERENCES

Alonso-Salces RM, Guyot S, Herrero C, Berrueta LA, Drilleau JF, Gallo B, Vicente F. 2004. Chemometric characterisation of Basque and French ciders according to their polyphenolic profiles. *Anal Bioanal Chem* **379**: 464–475.

Barakat HH, Sahar Hussein AM, Mohamed Marzouk S, Merfort I, Linscheid M, Nawwar Mahmoud AM. 1997. Polyphenolic metabolites of *Epilobium hirsutum*. *Phytochemistry* **46**: 935–941.

Battinelli L, Tita B, Evandri MG, Mazzanti G. 2001. Antimicrobial activity of Epilobium spp. extracts. *Il Farmaco* **56**: 345–348.

Bijlsma S, Bobeldijk I, Verheij ER, Ramaker R, Kochhar S, Macdonald IA, van Ommen B, Smilde AK. 2006. Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. *Anal Chem* **78**: 567–574.

Blanquer A, Boira H, Soler V, Perez I. 1998. Variability of the essential oil of *Thymus piperella*. *Phytochemistry* **47**: 1271–1276.

Ducrey B, Marston A, Gohring S, Hartmann RW, Hostettmann K. 1997. Inhibition of 5 alpha-reductase and aromatase by the ellagitannins oenothein A and oenothein B from *Epilobium* species. *Planta Med* **63**: 111–114.

Gonzáles AG, Pablos F, Martin MJ, León-Camacho M, Valdenebro MS. 2001. HPLC analysis of tocopherols and triglycerides in coffee and their use as authentication parameters. *Food Chem* **73**: 93–101.

Goodall I, Dennis MJ, Parker I, Sharmann M. 1995. Contribution of high-performance liquid chromatographic analysis of carbohydrates to authenticity testing of honey. *J Chromatogr* **706**: 353–359.

Hair JF Jr, Anderson RE, Tatham RE. 1987. *Multivariate Data Analysis with Readings* 2nd Edn. Macmillan: Indianapolis, USA, 295-336.

Haussknecht C. 1884. *Monographie der Gattung Epilobium*. Verlag von Gustav Fischer: Jena.

Hiermann A, Reidlinger M, Juan H, Sametz W. 1991. Isolierung des antiphlogistischen Wirkprinzips von *Epilobium angustifolium*. *Planta Med* **57**: 357–360.

Lutz U, Lutz RW, Lutz WK. 2006. Metabolic profiling of glucuronides in human urine by LC-MS/MS and partial least-squares discriminant analysis for classification and prediction of gender. *Anal Chem* **78**: 4564–4571.

Ravnik V, 1999. *Epilobium* L.—vrbovec. In: *Mala Llora Slovenije*, Martinčič A (ed.). Tehniška založba Slovenije Ljubljana; 289–291.

Schaneberg BT, Crockett S, Bedir E, Khan IA. 2003. The role of chemical fingerprinting: application to *Ephedra*. *Phytochemistry* **62**: 911–918.

Strgulc Krajšek S, Jogan N. 2004. *Epilobium obscurum* Schreb., new species in the flora of Slovenia. *Natura Sloveniae* **6**: 5–10.

Terabe S. 2004. *Micellar Electrokinetic* Chromatography. www.ceandcec.com/266924%20-%20CE%20Primer%203.pdf

Tita B, Abdel-Ha H, Vitalone A, Mazzanti G, Saso L. 2001. Analgesic properties of *Epilobium angustifolium*, evaluated by the hot plate test and the writhing test. *Il Farmaco* **56**: 341–343.

Vitalone A, Bordi F, Baldazzi C, Mazzanti G, Saso L, Tita B. 2001. Antiproliferative effect on a prostatic epithelial cell line (PZ-HPV-7) by *Epilobium angustifolium* L. *Il Farmaco* **56**: 483–489.

Voynova E, Dimitrova S, Naydenova E, Karadjov P. 1991. Inhibitory action of extracts of *Maclura aurantiaca* and *Epilobium hirsutum* on tumour models in mice. *Acta Physiol Pharmac Bulg* **17**: 50–52.

Wätzig H, Degenhardt M, Kunkel A. 1998. Strategies for capillary electrophoresis: Method development and validation for pharmaceutical and biological applications. *Electrophoresis* **19**: 2695–2752.

Welsh WJ, Lin W, Tersigni SH, Collantes E, Duta R, Carey MS, Zielinski WL, Brower J, Spencer JA, Layloff TP. 1996. Pharmaceutical fingerprinting: evaluation of neural networks and chemometric techniques for distinguishing among same-product manufacturers. *Anal Chem* **68**: 3473–3482.

Xie P, Chen S, Liang YZ, Wang X, Tian R, Upton R. 2006. Chromatographic fingerprint analysis—a rational approach for quality assessment of traditional Chinese herbal medicine. *J Chromatogr A.* **1112**: 171–180.