Yufeng Jane Tseng[1,2,3,4]
Chun-Ting Kuo[1]
San-Yuan Wang[2,4]
Hsiao-Wei Liao[1,2]
Guan-Yuan Chen[1,2]
Yuan-Ling Ku[6]
Wei-Cheng Shao[1]
Ching-Hua Kuo[1,2,5]

[1]School of Pharmacy, College of Medicine, National Taiwan University, Taiwan
[2]The Metabolomics Core Laboratory, Center of Genomic Medicine, National Taiwan University, Taiwan
[3]Graduate Institute of Biomedical Electronic and Bioinformatics, National Taiwan University, Taiwan
[4]Department of Computer Science and Information Engineering, National Taiwan University, Taiwan
[5]Department of Pharmacy, National Taiwan University Hospital, Taiwan
[6]Medical and Pharmaceutical Industry Technology and Development Center, Taiwan

Received October 24, 2012
Revised December 18, 2012
Accepted January 20, 2013

Research Article

Metabolomic characterization of rhubarb species by capillary electrophoresis and ultra-high-pressure liquid chromatography

This study developed CE and ultra-high-pressure LC (UHPLC) methods coupled with UV detectors to characterize the metabolomic profiles of different rhubarb species. The optimal CE conditions used a BGE with 15 mM sodium tetraborate, 15 mM sodium dihydrogen phosphate monohydrate, 30 mM sodium deoxycholate, and 30% ACN v/v at pH 8.3. The optimal UHPLC conditions used a mobile phase composed of 0.05% phosphate buffer and ACN with gradient elution. The gradient profile increased linearly from 10 to 21% ACN within the first 25 min, then increased to 33% ACN for the next 10 min. It took another 5 min to reach the 65% ACN, then for the next 5 min, it stayed unchanged. Sixteen samples of *Rheum officinale* and *Rheum tanguticum* collected from various locations were analyzed by CE and UHPLC methods. The metabolite profiles of CE were aligned and baseline corrected before chemometric analysis. Metabolomic signatures of rhubarb species from CE and UHPLC were clustered using principle component analysis and distance-based redundancy analysis; the clusters were not only able to discriminate different species but also different cultivation regions. Similarity measurements were performed by calculating the correlation coefficient of each sample with the authentic samples. Hybrid rhizome was clearly identified through similarity measurement of UHPLC metabolite profile and later confirmed by gene sequencing. The present study demonstrated that CE and UHPLC are efficient and effective tools to identify and authenticate herbs even coupled with simple detectors.

Keywords:
Capillary electrophoresis / Chemometric analysis / Metabolite profiling / Rhubarb / Ultra-high-pressure liquid chromatography DOI 10.1002/elps.201200580

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

Rhubarb is a group of plants that belong to the genus *Rheum* in the family Polygonaceae. Various parts of these plants have medicinal and culinary uses. Rhubarb alone or its combination with other herbs is widely used as a dietary supplement. The pharmacological effects of rhubarb include laxative, antibacterial, hemostatic, anti-inflammatory, antispasmodic, and antitumor [1–3]. Different species of rhubarb present variations in their chemical constituents. Therefore they can produce different pharmacological effects. Traditional methods for species differentiation include powder examination, morphology test, and tissue examination. However, these methods are not very specific due to the similar morphology shared among different rhubarbs. Metabolomic characterization or metabolite profiling aims at detecting metabolites of representative profiles for different species [4]. With the help of the chemometric analysis, the similarity of metabolite profile can be calculated. A similar approach, chromatographic fingerprint, has been adopted by the World Health Organization as a quality control technique for identifying medicinal herbs [5].

Various classes of components have been extracted from rhubarb [6, 7]. Several qualitative and quantitative methods for rhubarb analysis have been established like high-speed counter current chromatography [8,9], HPLC [6,10–13], CE [14, 15], and CEC [16]. However, only one or very few

Correspondence: Professor Ching-Hua Kuo, School of Pharmacy, College of Medicine, National Taiwan University, 1, Section 1, Jen-Ai Road, Taipei 100, Taiwan
E-mail: kuoch@ntu.edu.tw
Fax: +886-2-23919098

Abbreviations: dbRDA, distance-based redundancy analysis; **PC**, principal component; **PCA**, principal component analysis; **SDC**, sodium deoxycholate; **UHPLC**, ultra-high-pressure liquid chromatography

components were analyzed in these studies. VanMen et al. [17] compared the differences between using five marker compounds or 17 common peaks in rhubarb for classification of its different species. They concluded the method accuracy was reduced significantly if they only used five marker compounds for classification. One study used HPLC for fingerprint analysis of *Rheum tanguticum* [18]. Analytical time, however, was considerably long—142 min for a single analysis due to the nature of the chemical complexity of the herbal medicine. With the long analytical time, it is impractical to apply these methods as a routine quality control method for rhubarb. Therefore, there is still a strong demand to develop efficient metabolic profiling methods for species differentiation of rhubarb.

CE is increasingly recognized as an important separation technique due to its higher resolution, decreased solvent consumption, and rapid separation. Ultra-high-pressure liquid chromatography (UHPLC), a relatively new technique, provides superior advantages in improving runtime, efficiency, and sensitivity when compared to conventional HPLC. CE and UHPLC offer considerable advantages in terms of separation speed and peak efficiency, which are especially useful in analyzing complex herbal extracts. Until now, no CE or UHPLC method has been utilized for the metabolite profiling of rhubarb. Hyphenated chromatographic-mass spectrometric methods are widely used for metabolite profiling of herbal extracts [19, 20]. However, the high instrumentation cost is not affordable to many laboratories.

The roots of *R. tanguticum* Maxim. ex Balf., and *Rheum officinale* Baillon are the most widely used rhubarb species in Asia. The goal of the present study is to develop efficient methods for metabolomic characterization of rhubarb species. Instruments with high-separation efficiency such as CE and UHPLC were used to reduce analytical time while maintaining good selectivity for complicated herbal extracts. Both CE-UV and UHPLC-DAD) conditions were optimized to construct metabolic profiles of rhubarb. To resolve the peak shifts and baseline drift problems frequently encountered in CE, we adapted peak alignment and baseline correction tools for CE chromatograms. CE and UHPLC chromatograms of *R. officinale* and *R. tanguticum* were used for similarity measurements and principal component analysis (PCA). The analytical results of CE and UHPLC were discussed for their advantages in metabolic profiling of rhubarb.

2 Material and methods

2.1 Materials

Gallic acid was purchased from Wako Pure Chemical Industries (Wako Pure Chemical Industries, Tokyo, Japan). Sennoside A, sennoside B, and physcion were purchased from Extrasynthese (Paris, France). Aloe-emodin was purchased from Scientific Pharmaceutical Elite Company (Taipei, Taiwan). Chrysophanol was purchased from Fluka (Steinheim, Germany). Emodin, rhein, β-CD, sodium tetraborate, (+) catechin, (−) epicatechin, sodium dihydrogen phosphate, sodium cholate, sodium taurocholate, sodium taurodeoxycholate, SDS, and sodium deoxycholate (SDC) were purchased from Sigma-Aldrich (Sigma, Steinheim, Germany). ACN was purchased from J.T. Baker (Phillipsburg, NJ, USA). Phosphoric acid was purchased from Merck (Darmstadt, Germany). All reagents and solvents used were of analytical or chromatographic grade. DNA primers were purchased from Bio Basic (Ontario, Canada). Prime *Taq* DNA polymerase and 10× reaction buffer were purchased from Genet Bio (Cheonan, Korea). dNTP mixture was purchased from Takara Bio (Shiga, Japan). DNA-loading dye and 100 bp DNA ladder were purchased from Bioman Scientific (Taipei, Taiwan).

Table 1. The sources of the tested samples

Species	Code	Source
Rheum officinale	H694	Hubei, China
Rheum officinale	H697	Hubei, China
Rheum officinale	H705	Hubei, China
Rheum officinale	L445	Beijing, China
Rheum officinale	L446	Beijing, China
Rheum officinale	L447	Beijing, China
Rheum officinale	L457	Beijing, China
Rheum officinale	L472	Beijing, China
Rheum tanguticum	H1131	Beijing, China
Rheum tanguticum	H1203	Beijing, China
Rheum tanguticum	H1205	Beijing, China
Rheum tanguticum	H1207	Beijing, China
Rheum tanguticum	H1209	Beijing, China
Rheum tanguticum	H1214	Beijing, China
Rheum tanguticum	H1215	Beijing, China
Rheum tanguticum	H1218	Beijing, China

The dried roots of *R. officinal*e and *R. tanguticum* were collected from different regions in Hubei and Beijing, China (Table 1). All samples were provided by the Medical and Pharmaceutical Industry Technology and Development Center of Taiwan under research collaboration. The herbal samples were authenticated by the local experts through morphology test and powder examination at the sample collection site. Voucher specimens were stored at the herbarium of the Medical and Pharmaceutical Industry Technology and Development Center of Taiwan.

2.2 Apparatus and analytical conditions

2.2.1 CE

The CE experiments were carried out in a system consisting of a Prince programmable injector from Lauer Labs (Emmen, the Netherlands) and a 30 kV high-voltage supply, connected to a UV-C absorbance detector from Dynamax (Rainin, Emeryville, CA, USA). A fused-silica capillary from Polymicro Technologies (Phoenix, AZ, USA) was used. The electropherograms were recorded using an EZChrom (Scientific Software, San Ramon, CA, USA) chromatographic data system. All separations were performed on a 70 cm (62.5 cm

effective length) × 50 μm id fused-silica capillary. Detection wavelength was 254 nm and the capillary was thermostated by fan at 25°C. The new fused-silica capillary was conditioned by flushing with 1.0 M NaOH for 10 min, 0.2 M NaOH for 10 min, water for 10 min, and running solution for 10 min in sequence.

The BGE solution of MEKC comprised 15 mM of sodium tetraborate, 15 mM of sodium dihydrogen phosphate, 30 mM SDC and 30% v/v ACN. Its pH value was adjusted to 8.3 with 0.2 M sodium hydroxide. The sample was hydrodynamically injected at 50 mbar for 4.8 s. The applied voltage was 30 kV. At the beginning of each experiment, the capillary was treated with 0.1 M HCl for 2 min, DI water for 2 min, 0.1 N NaOH for 2 min, DI water for 2 min, and the BGE solution for 8 min.

2.2.2 UHPLC

The UHPLC analysis was carried out on a Waters Ultra performance liquid chromatographic system (UPLC). It consisted of a Waters Acquity UPLC autosampler, a column manager, heater/cooler, a binary solvent manager, and an Acquity UPLC-DAD (Waters, Milford, MA, USA), monitored over a range of 200–500 nm. Separations were carried out on a Waters Acquity UPLC BEH C18 column (100 × 2.1 mm, 1.7 μm). The analytical column was thermostated at 30°C. The mobile phase was composed of a mixture of solution A: buffer solution (0.05% phosphoric acid) and solution B: ACN. The gradient profile was: 0 min: 10% B, 25 min: 21% B, 35 min: 33% B, 40 min: 65% B, then for the next 5 min, it stayed unchanged. The flow rate was kept at 0.4 mL/min. The UV wavelength was set at 254 nm. Partial loop with needle overfill mode was used for sample injection, and the injection volume was 5 μL.

2.3 Preparation of standard solution

The standard stock solutions were prepared by dissolving the appropriate amount of standards in 70% ACN v/v. The standard working solutions were prepared by diluting the stock solution with 70% ACN to the following concentration: aloe-emodin 15 μg/mL, (+) catechin 100 μg/mL, chrysophanol 15 μg/mL, emodin 15 μg/mL, (−) epicatechin gallate 100 μg/mL, gallic acid 100 μg/mL, physcion 15 μg/mL, rhein 40 μg/mL, sennoside A 40 μg/mL, and sennoside B 40 μg/mL.

2.4 Preparation of sample solution

Sixteen rhubarb samples were pulverized into fine powders by a pulverizer. Each sample was accurately weighed for 200 mg and extracted with 5 mL of 70% ACN by ultrasonication for 20 min. The resulting solution was then centrifuged at 1500 × *g* for 20 min (4°C). The extraction was repeated for three times and the extracts were combined. The sample was filtered through a 0.22 μm membrane and analyzed by CE and UHPLC.

2.5 Isolation of DNA, PCR, and DNA sequencing

Sixteen rhubarb samples were pulverized into fine powders by a pulverizer. Twenty milligrams of each rhubarb sample were used for DNA extraction. The DNA extraction was completed with EasyPure Plant Genomic Spin Kit. The PCR amplification was performed using 100 ng of extracted DNA as a template in 50 μL of reaction mixture, consisting of 10 mM Tris–HCl (pH 9.0), 50 mM KCl, 0.1% Triton X-100, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.25 mM of each primer, and 1.5 U of *Taq* Polymerase (Bioman Scientific). The two pairs of primers flank the two different regions (I, II) of *mat*K gene as follows: *mat*KAF (5′- CTA TAT CCA CTT ATC TTT CAG GAG T-3′) and trnK1544R (5′-GGA TAA CCC CAG AAT GCT TAG-3′) for region I; matK780F (5′-ACT AAG CAT TCT GGG GTT ATC-3′) and *mat*K8R (5′-AAA GTT CTA GCA CAA GAA AGT CGA-3′) for region II [21]. PCR was carried out in a Master cycler personal Eppendorf® with the following cycling conditions: hot start at 94°C for 5 min followed by 38 cycles of 94°C for 1 min, 48°C for 1 min, and 72°C for 1.5 min, and final extension at 72°C for 10 min. Following PCR amplification, 20 μL of the resulting PCR product were analyzed by 1.2% agarose gel electrophoresis, under 100 kV for 30 min on Bio-Rad PowerPac™ Basic. The gel image was taken by the UVP BioDoc-It™ imaging system. The DNA sequencing was performed by Tri-I Biotech, Taiwan. The phylogenetic tree was constructed and displayed by Align X in Vector NTI Advance 10.30.

2.6 Chemometric analysis

2.6.1 Data acquisition

Each CE chromatogram with UV wavelength 254 nm was collected using 2400 data points within 20 min. Each UHPLC chromatogram with UV wavelength 254 was collected using 54 000 points within 45 min. In short, all signals of each analytic sample were used for PCA and distance-based redundancy analysis (dbRDA) in this study.

2.6.2 Baseline correction

Both baseline correction and PCA were implemented with *R* (Version 2.8.1) [22]. To correct the baseline, we calculated the SD of peak intensities with a moving window of 25 s to slide through a chromatogram and distinguish the baseline noises and signals. The SD of intensities of each window was obtained to select the third quartile of the SD as the value of SD for a noise. If the difference of the maximum and minimum intensity in each window was smaller than the SD for a noise, the central point of this window was classified as a noise. The

moving window continued to slide through a chromatogram until each point in this chromatogram was evaluated. Finally, the baseline was deduced by linear interpolation of points classified as noise.

2.6.3 Peak alignment

In a chromatography experiment, retention time shifts can result in an incorrect detection of differences between preclassified samples—especially on CE chromatograms. In order to align the chromatograms, we used a web server, Chromaligner, developed in-house to perform peak alignment on all CE chromatograms [23]. Chromaligner is a tool for chromatogram alignment that can align retention time of chromatographic methods such as HPLC and CE. Chromaligner resolves peak shifts by using a constrained chromatogram alignment. For a collection of chromatograms and a set of defined peaks, Chromaligner divides each chromatogram into $k+1$ segments according to the defined peaks, where k is the number of defined peaks, and continues to align corresponding segments using a correlation optimized warping algorithm with heuristically tuned parameters. Chromaligner is able to provide alignments based on known component peaks to reach the best result prior to further chemometric analysis. Chrysophanol, emodine, rhein, and gallic acid were selected as the defined peaks for the alignment in this study.

2.6.4 Principal component analysis (PCA)

To decompose the commonly correlated chromatographic data into a new set of uncorrelated variables, all chromatograms were analyzed with PCA. PCA projects the chromatographic data to a new set of orthogonal variables named principal components (PCs). Those PCs are related to the original chromatographic data since each PC is the linear combination of the original variables. Due to the highly correlated nature of original data variables, the reduced variable dimensions of PCs are able to describe the maximum variation within the data. PCs are ranked according to the amount of variance they can explain in the data. Hence, the first PC explains the maximum amount of the variance in the data. A scores plot from PCA is a 2D map as the function of the first two PCs results. If there are patterns or certain groupings (classes) in the data, samples of similar patterns or groupings would be close to each other in a scores plot. PCAs were performed with respect to chromatograms from CE and UHPLS on different rhubarb species in this study. The PCA scores plots of the first two PCs were generated with the prcomp function of R (Version 2.8.1).

2.6.5 Distance-based redundancy analysis (dbRDA)

dbRDA is a multivariate technique. The technique allows researchers to evaluate the degree of similarity between species.

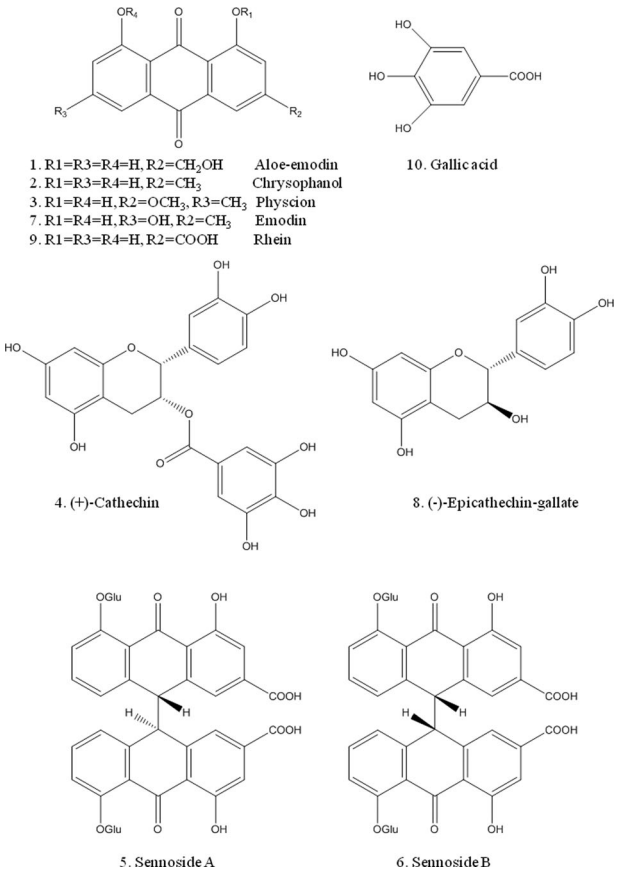

Figure 1. Structures of aloe-emodin (1), chrysophanol (2), physcion (3), (+)cathechin (4), sennoside B (5), sennoside A (6), emodin (7), (−) epicathechin gallate (8), rhein (9) and gallic acid (10).

The dbRDA analysis was performed with the vegan package of R [24]. Classification results using dbRDA ordination plots of 16 rhubarb samples from UHPLC and CE metabolite profiles were performed. One thousand permutations were performed to test the significance level of separation between species for each dbRDA model using UHPLC and CE metabolite profiles.

3 Results and discussion

3.1 Analytical method development

The purpose of this study is to develop metabolic profiling methods for rhubarb species classification. Thus, it is critical to have chromatographic conditions showing high number of peaks and good selectivity to provide good discriminating power. To achieve this goal, we systematically optimize each analytical parameter to improve the separation results. Ten standard materials including aloe-emodin, chrysophanol, emodin, physcion, rhein, sennoside A, sennoside B, (+) cathechin, (−) epicathechin, and gallic acid (Fig. 1) were used to develop the analytical method.

3.1.1 CE method development

The separation of ten rhubarb standards in CE is achieved by optimizing critical parameters. The most frequently used modes in CE for the separation of small molecules are MEKC and CZE. Previous studies revealed that MEKC mode provided much better selectivity compared to CZE for the separation of anthraquinones and bianthrones [15]. Surfactants including SDS, sodium cholate, SDC, sodium taurocholate, and sodium taurodeoxycholate were tested for their utility in MEKC separation of the tested standards. SDC and sodium taurodeoxycholate both gave better resolution than the others. For practical reasons, SDC was chosen to proceed with the experiment. Critical factors for selectivity adjustment in MEKC system include micelle concentration, pH value, and organic modifier. These parameters were all optimized in this study. SDC concentration was optimized by considering selectivity and separation speed. Thirty millimolars of SDC was found to provide the best selectivity within the shortest time, and it was selected for further method optimization.

It was found that peak shape of (+) cathechin and (−) epicathechin was very broad under certain conditions. The p*K*a of (+) cathechin and (−) epicathechin is about 9 ($pKa_1 = 8.16$, $pKa_2 = 9.20$) [25], and the variation of the pH of BGE between 8.0 and 9.0 showed significant effect on their peak shape and migration time. Lowering pH value of BGE would decrease the dissociation of their phenolic OH groups, and improve their peak shape. Although lower pH provided better peak shape, resolution between peak pair aloe-emodin (peak 1), and chrysophanol (peak 2), peak pair emodin (peak 7), and (−) epicathechin-gallate (peak 8) were sacrificed. Buffer pH at 8.3 was selected as the best compromise between resolution and peak shape.

Effects of organic modifiers include extending migration window, facilitating dissolution of hydrophobic molecules, and changing separation selectivity [26]. ACN percentage plays the most significant role in selectivity adjustment among all of the parameters, and it was further optimized. Different concentrations (24 to 32%, v/v) of ACN were added into the BGE to test their improvements on selectivity. The total retention time of the ten analytes increased as the ACN percentage increased as seen in Fig. 2 due to the decreased EOF. The resolution between peaks 1 and 2 increased, but the resolution between peaks 7 and 8 decreased with the increase in ACN percentage. Ten analytes were baseline separated when 30% ACN were added into the BGE. Thirty percent of ACN was therefore selected as the optimum percentage. The sample injection time was optimized through considering the method sensitivity and peak shape. Although increasing sample injection time improved the method sensitivity, the peak shape became too broad if the sample injection time was longer than 4.8 s. The optimal injection time was therefore selected as 4.8 s.

The metabolic profiles of *R. officinale* and *R. tanguticum* obtained under optimal CE conditions were shown in Fig. 3A and B, respectively. Further adjusting the BGE or applied voltage did not give better metabolite profile of rhubarb extract. Optimal conditions obtained by standards were therefore used for the metabolic profiling of rhubarb extract. The metabolite profiles of rhubarb could be obtained within 20 min with CE analysis.

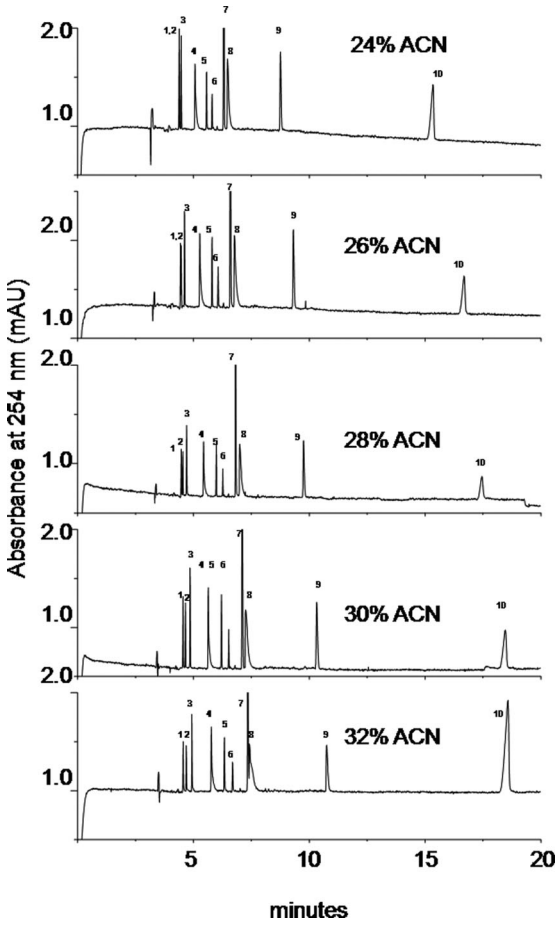

Figure 2. Effect of ACN percentage on separation of ten rhubarb standards. Separation conditions: fused-silica capillary: BGE: 15 mM $Na_2B_4O_7$ /15 mM NaH_2PO_4, 30 mM SDC, pH 8.3; applied voltage: 30 kV; temperature: 25°C; injection: 50 mbar for 4.8 s. The ACN percentages are as indicated in the figure. Compound identities are indicated in Fig. 1.

3.1.2 UHPLC method development

The major components in rhubarb are polyphenols. Dalluge et al. [27] and Lin et al. [12] reported that using endcapped stationary phase and acidic buffer in the mobile phase could eliminate peak tailing in polyphenols. To improve the peak shape of metabolite profiling for the complex herbal extracts, different concentrations of TFA and phosphoric acid were tested as the buffer solution in our mobile phase. Best peak shape was obtained with 0.05% of phosphoric acid and subsequently selected as the buffer solution for the mobile phase.

The separation of ten rhubarb standards was evaluated by changing the gradient eluting programs. The chromatogram

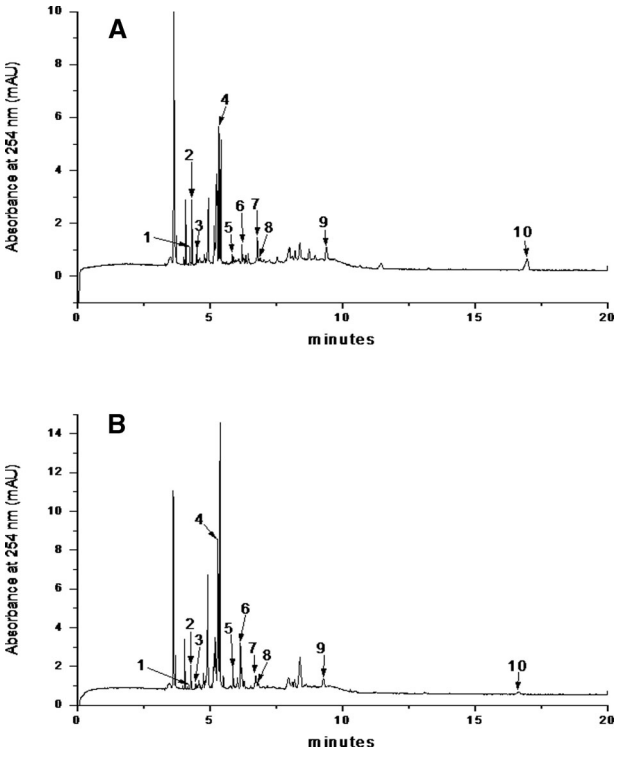

Figure 3. Electropherograms of (A) *Rheum officinale* (L446) and (B) *Rheum tanguticum* (H1205) obtained under optimal MEKC conditions. MEKC conditions as indicated in Fig. 2, except ACN percentage is 30%. Compound identities are as indicated in Fig. 1.

of rhubarb extract showed a large number of overlapping bands in the first 15 min and a small number of widely separated bands in the last 20 min when linear gradient was used (Fig. 4A). Therefore, nonlinear gradient was adapted in this study. The gradient profile increased linearly from 10 to 21% ACN within the first 25 min and then increased to 33% ACN for the next 10 min. It took another 5 min to reach the 65% ACN and then for the next 5 min, it stayed unchanged. The total analytical time was 45 min for the UHPLC method. Metabolite profiles of *R. officinale* and *R. tanguticum* and standard chromatogram obtained at optimal UHPLC conditions are shown in Fig. 4B–D. Ten components in the crude extracts were identified in terms of retention times and their respective UV spectrum. Compared to Fig. 4A, the analytical time of Fig. 4B and 4C was reduced and the peaks were evenly distributed in the chromatogram under the optimal analytical conditions.

3.1.3 Method precision

We used rhubarb extracts to test the method precision. The precision of the UHPLC and CE methods was determined for intra and interday variations in terms of retention time. The RSD values of the intraday ($n = 3$), and interday precision ($n = 3$) of the retention times of the ten chemicals used in the analytical method development obtained by the CE method were

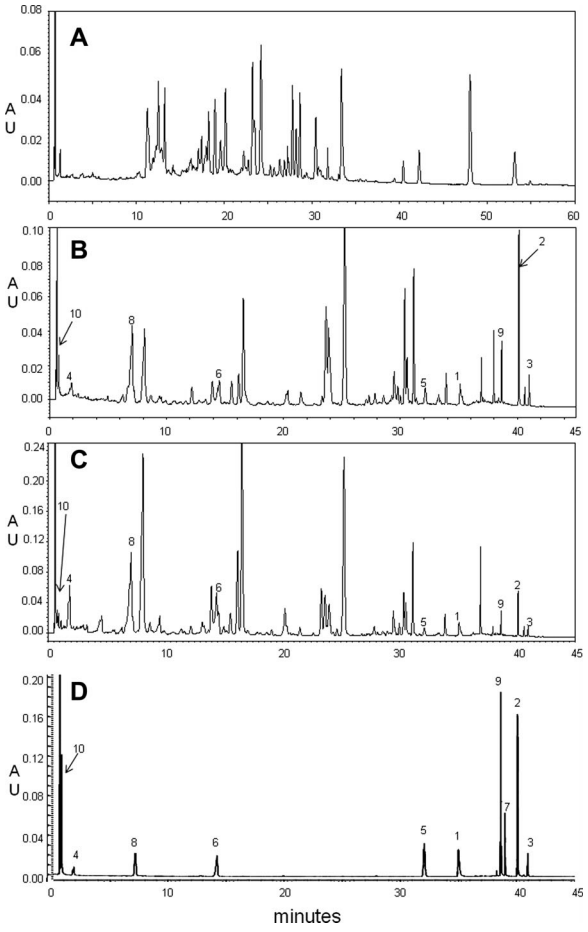

Figure 4. UHPLC chromatograms of the sample extract of (A and B) *Rheum officinale* (L446), (C) *Rheum tanguticum* (H1205), and (D) standard chromatogram obtained by different gradient. Mobile phase of UHPLC was composed of 0.05% phosphoric acid (solution A) and ACN (solution B); gradient profile: Figure (A): linear gradient from 5%B to 60%B in 60 min. Figure (b, c): 0 min: 10% B, 25 min: 21% B, 35 min: 33% B, 40 min: 65% B, 45 min: 65% B, v/v). Compound identities are indicated in Fig. 1.

found to be smaller than 2.58 and 3.11%, respectively. Their RSD values of the intraday ($n = 3$) and interday precision ($n = 3$) in terms of migration time obtained by the UHPLC method were smaller than 1.34 and 1.44%, respectively. CE chromatograms of rhubarb were aligned and baseline corrected before similarity and PCA analysis. UHPLC chromatograms of rhubarb were directly used for chemometric analysis without the preprocessing of chromatograms.

3.2 Analysis of rhubarb samples by CE and UHPLC

Sixteen crude extracts of rhubarb samples including eight *R. tanguticum* and eight *R. officinale* were analyzed by CE and UHPLC. Some of the CE electropherograms had significant baseline drift and retention time shift. Thus, all the CE chromatograms were aligned and baseline corrected before the similarity comparison.

Peak alignment is a critical step for the comparison of metabolite profiles obtained from CE. Variations of the EOF in CE lead to poor reproducibility in migration time. The crowded peaks in a chromatogram could result in incorrect similarity comparison between samples if retention time shift was occurred. To minimize these biases in the subsequent chemometric analysis, we used a chromatogram alignment tool, Chromaligner, to resolve the retention time shifts in chromatograms. Chromaligner developed by our group could perform chromatogram alignment by using a modified version of correlation optimized warping [28]. It allows users to use a set of peaks to improve the performance of alignment [23]. Chrysophanol, emodine, rhein, and gallic acid were selected for peak alignment due to their disposition in different regions in the electropherogram. The migration time shifts were corrected by Chromaligner, and the aligned chromatograms could be subjected to similarity measurement.

The analytical results of UHPLC showed high precision with a flat baseline. Metabolite profile of rhubarb tested samples obtained under optimal UHPLC conditions was directly used for PCA and similarity test.

3.3 Classification and similarity measurement of metabolite profiles of tested rhubarb samples

Metabolomic signatures of rhubarb species from CE and UHPLC were analyzed with PCA to evaluate the similarity or diversity of different rhubarb samples. PCA classification results of 16 rhubarb samples from UHPLC and CE metabolite profiles are shown in Fig. 5A and C. The metabolite profiles of *R. tanguticum* and *R. officinale* are displayed as two separate groups in the PCA scores plot. The same species were clustered in the PCA scores plot. Moreover, the same rhubarb species from the same cultivation region were also clustered close to each other in the PCA scores plot. For example, samples H694, H697, and H705 from Hubei (Table 1) are clustered in Fig. 5A and C. Since, metabolic profiles reflect the chemical constituents of herbs, this result indicates how cultivation geography affects rhubarb's constitution. dbRDA analysis was performed to determine whether the differences between rhubarb species are significant. The classification results using dbRDA ordination plots of 16 rhubarb samples from UHPLC and CE metabolite profiles are shown in Fig. 5B and D. The metabolite profiles of *R. tanguticum* and *R. officinale* are displayed as two separate groups in dbRDA ordination plots—showing the same species were clustered in the dbRDA ordination scores plot. Moreover, the same rhubarb species from the same cultivation region (such as H694, H697, and H705 from Hubei) were also clustered close to each other in the dbRDA ordination plots—similar to what we saw from the PCA scores plots. Moreover, based on 1000 permutations, both dbRDA models from profiles of UHPLC and CE results displayed significant separation between the two species ($p = 0.001$). Additional loading plot information of PCA analysis is now included in the Supporting Information Fig. 1.

The correlation coefficient was used as the similarity index between two metabolic profiles. The Pearson's correlation coefficient of two metabolic profiles, x and y, was calculated by the following equation:

$$\gamma_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{(n-1)\, S_x S_y} \quad (1)$$

where $\overline{x}$, S_x, $\overline{y}$, and S_y were the sample means and SDs of the metabolic profile x and y, respectively.

L446 and H1205 were selected as the authenticate samples of *R. officinale and R. tanguticum* confirmed by analyzing their *mat*K gene. Similarity measurements were performed with each sample and these two authentic samples. The pair wise correlation coefficients of each sample to L446 and H1205 were listed in Table 2. Among all rhubarb samples, H1209, identified as *R. tanguticum* in morphology test, showed the lowest correlation when compared to the authenticated *R. tanguticum* sample in both UHPLC and CE (0.77 and 0.8, respectively). We further calculated the H1209 correlation coefficient to the authenticated *R. officinale* sample, and found the similarity was even lower (0.7 and 0.67 in UHPLC and CE, respectively).

The low correlation values from either CE or UHPLC analysis indicated a distinct metabolite profile of H1209 to both rhubarb species, and therefore suspected to be a hybrid rhizome. In general, the high correlation of the metabolite profile is originated from their phylogenetic closeness. We therefore hypothesize that H1209 is genetically closer to *R. tanguticum* than *R. officinale*.

3.3.1 Analysis of rhubarb species based on the chloroplast matK gene sequence

The analytical results obtained from CE and UHPLC were further cross-examined with gene sequencing. Previously, Komatsu et al. concluded the *mat*K gene from chloroplast provided valuable information for identification of rhubarb species [21]. We applied the same primers from their study to check the *mat*K gene sequence of H1209, H1205 (*R. tanguticum*), L446 (*R. officinale)*, and H1214 (one selected rhubarb sample for comparison). Although H1209 was identified as *R. tanguticum* by morphology test, many nucleotides in *mat*K gene are the same as the *R. officinale* species. Other study shows rhubarb plants are self-incompatible in nature, and have been hybridized randomly since ancient times. Therefore, morphologically intermediate forms of rhubarb plants are commonly seen [29]. The phylogenetic tree of H1209, H1205, L446, and H1214 shows the *mat*K gene sequences of H1209 were somewhat different from the genus of *R. tanguticum* but even more genetically distinct from the samples of *R. officinale* (Fig. 6). The trend of genetic phylogeny is consistent with our metabolite profiling result. These results suggested that H1209 is a hybrid rhizome. Although traditional species differentiation methods such

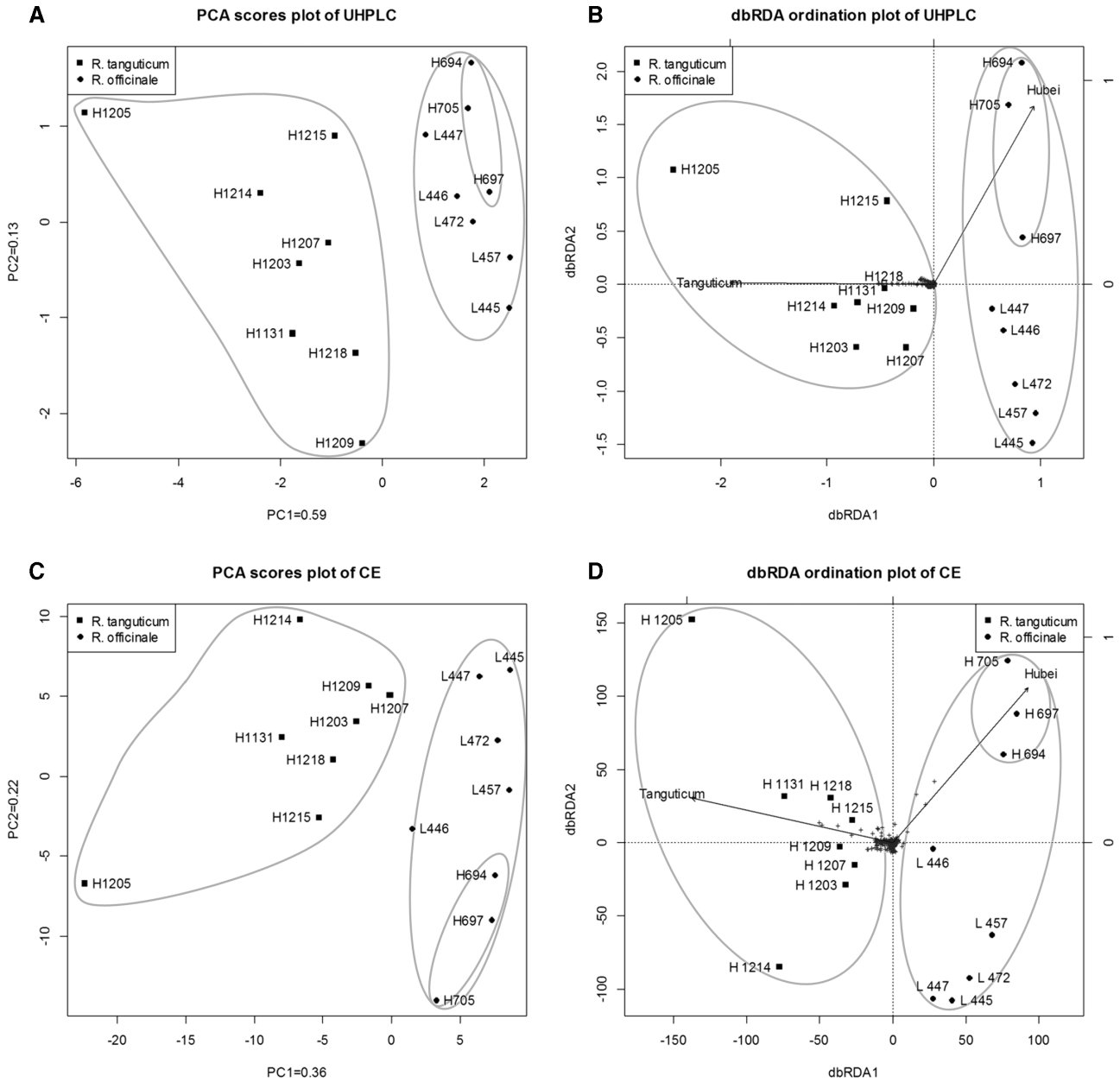

Figure 5. The PCA scores plots (A and C) and dbRDA ordination plots (B and D) of UHPLC and CE metabolite profiles from 16 rhubarb samples.

as morphology test and powder examination are able to characterize rhubarb species, the hybrid rhizome is very difficult to be identified. The present results demonstrated the developed metabolic profiling method is very sensitive, and it is able to identify the hybrid rhizome.

3.3.2 Comparison of CE and UHPLC for metabolite profiling of rhubarb

In general, CE is good in green chemistry in terms of low-solvent consumption but is less sensitive with lower reproducibility [30]. UHPLC is considered analytical robust [31]. The low reproducibility of CE can be greatly improved by various alignment tools such as the Chromaligner used in this study.

Metabolic profiles from CE and UHPLC were both able to distinguish different species. With more detected peaks (90 peaks for UHPLC and 50 peaks for CE), the developed UHPLC method was more sensitive than CE for detecting hybrid rhizome. From Table 2, the correlation coefficient between the genetically confirmed hybrid, H1209, and the authenticated *R. tanguticum* in UHPLC showed distinct difference when compared to the correlation coefficients of other

Table 2. Correlation coefficient of CE and UHPLC metabolite profiles from 16 rhubarb samples

Rheum officinale Sample ID	Correlation coefficient		*Rheum tanguticum* Sample ID	Correlation coefficient	
	CE	UHPLC		CE	UHPLC
L445	0.85	0.86	H1131	0.92	0.91
L446	1.00	1.00	H1203	0.86	0.95
L447	0.97	0.94	H1205	1.00	1.00
L457	0.82	0.82	H1207	0.91	0.97
L472	0.95	0.92	**H1209**[a)]	**0.80**	**0.77**
H694	0.87	0.85	H1214	0.86	0.94
H697	0.85	0.92	H1215	0.86	0.89
H705	0.87	0.89	H1218	0.90	0.89

a) The bold terms represent the correlation coefficient of both CE and UHPLC metabolite profiles were the lowest among all of the tested samples.

rhubarbs with either the authenticated *R. tanguticum* or *R. officinale*. However, the correlation coefficient differences of the hybrid rhizome to authentic and others to authentic in CE were not as significant. With the higher sensitivity and selectivity of the developed UHPLC profiling in conjunction with chemometric techniques, smaller chemical constituents in herbs originated from hybridization could be more clearly observed through similarity measurement of metabolic profiles.

VanMen et al. [17] compared the differences between using five marker compounds (sennoside A, rhaponticin, emodin-glucoside, emodin, and chrysophanol) versus 17 common peaks in rhubarb for species classification of rhubarb. When only five marker compounds were used for pattern analysis, only 79.5% of samples were properly classified. Instead, most samples could be classified with an accuracy of 100% if 17 common peaks were used for pattern analysis. To improve the accuracy for species differentiation, our study applied the metabolic profiling technique to obtain as much as possible information about rhubarb constituents for species differentiation. The dbRDA analysis revealed that both of the developed UHPLC (90 peaks) and CE (50 peaks) methods achieved 100% accuracy for species differentiation. Moreover, the UHPLC profiling method is able to identify the hybrid rhizome due its higher sensitivity and selectivity.

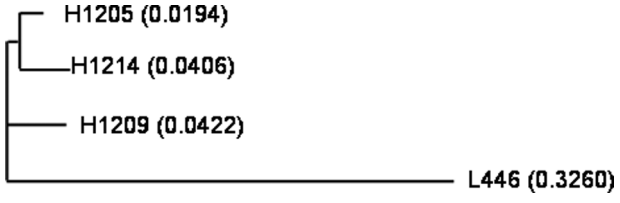

Figure 6. Phylogenetic tree of H1205, H1209, H1214, and H446.

4 Concluding remarks

The present study demonstrated that CE and UHPLC are efficient and effective tools to identify and authenticate herbs even coupled with simple detectors. The analytical times for both methods were largely reduced compared to previous HPLC method. Novel chemometric tools resolved the reproducibility problems of CE. Metabolomic signatures from CE and UHPLC could be used to differentiate rhubarb species and cultivation regions. Hybrid rhizome could be clearly identified through similarity measurement of UHPLC metabolite profiles. Both methods can be used for quality control and authentication of rhubarb samples.

This study was sponsored by College of Medicine (Aim for Top University Program) and Electronics Engineering and Computer Science (Excellent Research Project), National Taiwan University. The authors thank the NTU Integrated Core Facility for Functional Genomics of National Research Program for Genomic Medicine of Taiwan for technical assistance.

The authors have declared no conflict of interest.

5 References

[1] *WHO Monographs on Selected Medicinal Plants*, World Health Organization, Geneva 1999.

[2] Huang, K. C., *The Pharmacology of Chinese Herbs*, CRC Press, Boca Raton 1993.

[3] Tang, W., Eisenbrand, G., *Chinese Drugs of Plant Origin*, Springer, Berlin, Heidelberg 1992.

[4] Lee, E. J., Shaykhutdinov, R., Weljie, A. M., Vogel, H. J., Facchini, P. J., Park, S. U., Kim, Y. K., Yang, T. J., *J. Agr. Food Chem.* 2009, *57*, 7513–7522.

[5] Jiang, Y., David, B., Tu, P. F., Barbin, Y., *Anal. Chim. Acta* 2010, *657*, 9–18.

[6] Komatsu, K., Nagayama, Y., Tanaka, K., Ling, Y., Basnet, P., Meselhy, M. R., *Chem. Pharm. Bull.* 2006, *54*, 941–947.

[7] Wang, J., Li, H., Jin, C., Qu, Y., Xiao, X., *J. Pharm. Biomed. Anal.* 2008, *47*, 765–770.

[8] Liu, R., Li, A., Sun, A., *J. Chromatogr. A* 2004, *1052*, 217–221.

[9] Yang, F., Zhang, T., Tian, G., Cao, H., Liu, Q., Ito, Y., *J. Chromatogr. A* 1999, *858*, 103–107.

[10] Jin, W., Wang, Y.-F., Ge, R.-L., Shi, H.-M., Jia, C.-Q., Tu, P.-F., *Rapid Commun. Mass Spectrom.* 2007, *21*, 2351–2360.

[11] Koyama, J., Morita, I., Kobayashi, N., *J. Chromatogr. A* 2007, *1145*, 183–189.

[12] Lin, C.-C., Wu, C.-I., Lin, T.-C., Sheu, S.-J., *J. Sep. Sci.* 2006, *29*, 2584–2593.

[13] Ye, M., Han, J., Chen, H., Zheng, J., Guo, D., *J. Am. Soc. Mass Spectrom.* 2007, *18*, 82–91.

[14] Li, F., Cao, Q.-E., Ding, Z., *Chromatographia* 2004, *59*, 753–757.

[15] Kuo, C. H., Sun, S. W., *Anal. Chim. Acta* 2003, *482*, 47–58.

[16] Lü, H., Wang, J., Wang, X., Lin, X., Wu, X., Xie, Z., *J. Pharm. Biomed. Anal.* 2007, *43*, 352–357.

[17] VanMen, C., Jang, Y. S., Zhu, H. M., Lee, J. H., Trung, T. N., Ngoc, T. M., Kim, Y. H., Kang, J. S., *Phytochem. Analysis* 2012, *23*, 359–364.

[18] Jin, W., Ge, R. L., Wei, Q. J., Bao, T. Y., Shi, H. M., Tu, P. F., *J. Chromatogr. A* 2006, *1132*, 320–324.

[19] Urakami, K., Zangiacomi, V., Yamaguchi, K., Kusuhara, M., *Biomed. Res.-Tokyo* 2010, *31*, 161–163.

[20] Arbona, V., Iglesias, D. J., Talon, M., Gomez-Cadenas, A., *J. Agr. Food Chem.* 2009, *57*, 7338–7347.

[21] Yang, D. Y., Fushimi, H., Cai, S. Q., Komatsu, K., *Biol. Pharm. Bull.* 2004, *27*, 375–383.

[22] R Development Core Team. *R: A Language and Environment for Statistical Computing*, Vienna, Austria, http://www.R-project.org, 2010.

[23] Wang, S. Y., Ho, T. J., Kuo, C. H., Tseng, Y. F. J., *Bioinformatics* 2010, *26*, 2338–2339.

[24] Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Wagner, H. *Vegan: Community Ecology Package*, http://CRAN.R-project.org/package=vegan.

[25] Slabbert, N. P., *Tetrahedron* 1977, *33*, 821–824.

[26] Chen, N., Terabe, S., *Electrophoresis* 1995, *16*, 2100–2103.

[27] Dalluge, J. J., Nelson, B. C., Brown Thomas, J., Sander, L. C., *J. Chromatogr. A* 1998, *793*, 265–274.

[28] Nielsen, N. P. V., Carstensen, J. M., Smedsgaard, J., *J. Chromatogr. A* 1998, *805*, 17–35.

[29] Komatsu, K., Nagayama, Y., Tanaka, K., Ling, Y., Cai, S. Q., Omote, T., Meselhy, M. R., *Chem. Pharm. Bull. (Tokyo)* 2006, *54*, 1491–1499.

[30] Cancalon, P. F., *Food Technol.* 1995, *49*, 52–58.

[31] Nguyen, D. T. T., Guillarme, D., Rudaz, S., Veuthey, J. L., *J. Sep. Sci.* 2006, *29*, 1836–1848.

