Application of Ultra-Performance LC-TOF MS Metabolite Profiling Techniques to the Analysis of Medicinal *Panax* Herbs

By: Guo X. Xie, Yan Ni, Ming M. Su, Yuan Y. Zhang, Ai H. Zhao, Xian F. Gao, Zhong Liu, Pei G. Xiao, and Wei Jia

Xie, G., Ni, Y., Su, M., Zhang, Y., Zhao, A., Gao, X., Liu, Z., Long, X., & Jia, W. (2008). Application of ultra-performance LC-QTOF MS metabolite profiling techniques to the analysis of medicinal Panax herbs. *Metabolomics*, *4*, 248-260.

Abstract:

The morphological appearance and some ingredients of *Panax ginseng*, *Panax notoginseng* and *Panax japonicus* of the *Panax* genus are similar. However, their pharmacological activities are obviously different due to the significant differences in the types and quantity of saponins in each herb. In the present study, ultra-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-QTOFMS) was used to profile the abundances of metabolites in the three medicinal *Panax* herbs. Multivariate statistical analysis technique, that is, principle component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) were used to discriminate between the *Panax* samples. PCA of the analytical data showed a clear separation of compositions among the three medicinal herbs. The critical markers such as chikusetsusaponin IVa, ginsenoside R0, ginsenoside Rc, ginsenoside Rb1, ginsenoside Rb2 and ginsenoside Rg2 accountable for such variations were identified through the corresponding loading weights, and the tentative identification of biomarkers is completed by the accurate mass of TOFMS and high resolution and high retention time reproducibility performed by UPLC. The proposed analytical method coupled with multivariate statistical analysis is reliable to analyze a group of metabolites present in the herbal extracts and other natural products. This method can be further utilized to evaluate chemical components obtained from different plants and/or the plants of different geographical locations, thereby classifying the medicinal plant resources and potentially elucidating the mechanism of inherent phytochemical diversity.

Article:

INTRODUCTION

It is estimated that about 65–80% of the world's population is using traditional medicine as the primary form of healthcare (Akerele 1992). Traditional Chinese medicines (TCMs) are gaining more and more attention in many fields because of their low toxicity and good therapeutic performance. The quality and contents of herbs are highly variable depending on geographical origins, climate, cultivation, and the growth stage when harvested (Mahady et al. 2001). The

profiling of natural products requires an analytical system capable of generating an information rich data set and needs to identify the compounds of interest, compare different batches of material to be compared and contrasted and isolate the compounds of interest from bulk solution. However, because of the chemical diversity of the metabolome, which for any given multicellular species comprises a mixture of thousands of compounds differing in size, polarity etc., and varying in abundance by several orders of magnitude, the need for multiparallel analytical techniques is obvious and well-accepted (Goodacre et al. 2004; Hall 2006). A number of techniques including nuclear magnetic resonance (NMR), liquid chromatography (LC) or gas chromatography (GC) coupled with mass spectrometry (MS) have been employed for metabolite profiling (Fiehn et al. 2000; Want et al. 2005; Fukusaki et al. 2006; Nordstrom et al. 2006; Wagner et al. 2006; Farag et al. 2007). While the NMR metabolite profiling approach provides the valuable metabolite signature of a complex sample, LC-MS has higher sensitivity, resolution and resolves individual chemical components into separate peaks, enhancing the opportunity to mine and uncover novel metabolites (Chan et al. 2007). Unlike GC-MS, LC-MS is in principle a more versatile analytical tool. It covers a much wider mass range and allows one to target many compound classes not detectable by GC-MS. Furthermore, there is usually no need for complicated derivatization, and modern LC-MS setups offer superior options to structurally elucidate unknown metabolites, namely, accurate mass determination and elemental composition analysis. Among the various LC platforms, the soaring development of column particle materials has highlighted a new addition to chromatographic separation technology, the ultra-performance liquid chromatography (UPLC), which allows satisfactory separation, good resolution and sensitivity, and high-speed detection (Shen et al. 2005; Wilson et al. 2005; Nordstrom et al. 2006) with complex biological samples, such as herbal medicine (Guan et al. 2007). Coupled to quadrupole time-of-flight mass spectrometry (QTOFMS), UPLC/QTOFMS was used to achieve comprehensive and reliable metabolite profiling (Nordstrom et al. 2006) and recently the UPLC coupled with MS has been used for metabolite profiling in herbal medicine successfully both by other researchers and our group (Chan et al. 2007; Xie et al. 2008).

Panax ginseng, *Panax notoginseng* and *Panax japonicus* are all highly valuable and important tonic herbs belonging to *Panax* genus cultivated in different locations in China. While the *P. ginseng*, mainly cultivated in northeastern China, Jilin, for example, is used in traditional Chinese medicine (TCM) to enhance stamina and capacity coping with fatigue and physical stress. Brekhman, a pioneer in the experimental evaluation of ginseng, used the term adaptogen to describe these non-specific “tonic” effects of *P. ginseng* and other members of the Ariliaceae family (Brekhman and Dardymov 1969), in contrast, *P. notoginseng* is mainly cultivated in southwestern China (e.g. Yunnan Province of China) and used in TCM for its haemostatic and cardiovascular properties (National Committee of Pharmacopoeia 2005). *P. japonicus*, mainly distributed in west China, has effect of protecting on the cardiovascular system, enhancing the immune system, antiviral, relieving swelling and alleviating pain (Jiangsu New Medical College 1985). However, their traditional indications are significantly different although the three *Panax* samples all contains ginsenosides (saponins) which are usually considered to be the major active components and the types and quantity of saponins in each herb were significantly different. The level of total saponins of *P. japonicus* was higher than that of *P. ginseng* and *P. notoginseng* and in which oleanane saponins are the main active constituents (Zuo and Yuan 2005). There are dammarane-type and oleanane-type saponins in both *P. ginseng* and *P. japonicus*, but the latter saponins are absent in *P. notoginseng* (Wei et al. 1991). The content of dammarane-type

saponins in *P. ginseng* was higher than in *P. japonicus* whereas the content of oleanane saponins was lower in *P. japonicus* (Guo et al. 2004). These saponins are accountable for diverse therapeutic effects, thus understanding of the bio-distribution of these secondary metabolites in different type of *Panax* plants is paramount for the investigation of their pharmacological efficiency (Wang et al. 2005) and therefore, chemical characteristics are very important for the identification and quality control of the three herbs. The metabolite profiling of *Panax* herbs including Chinese ginseng, American ginseng and Korean ginseng has been carried out successfully with the developed UPLC-TOFMS method and multivariate statistical analysis (Xie et al. 2008). From the result we can see that ginseng samples from different locations in China and outside China (including Chinese, American and Korean) were different and can be separated with the developed method. So in this study we want to know whether the ginseng samples cultivated in different locations of China is different and can be separated with the developed method or not, and identify the differentiate saponins accountable for such variations.

Metabolomics, aims to identify and quantify the full complement of low-molecular-weight, soluble metabolites in actively metabolizing tissues (Oliver et al. 1998; Fiehn 2002; Fiehn and Weckwerth 2003; Rochfort 2005), has aroused extensive awareness and interests, especially in the botanical studies such as quality control of plants (Huhman and Sumner 2002; Chang et al. 2006; Holmes 2006; Eisenreich 2007; Tarachiwin et al. 2007) and plant engineering (Anterola and Lewis 2002; Long et al. 2006). Several methods have been developed for the analysis and differentiation of *Panax* samples, but the number and type of saponins detected was relatively low. Moreover, the comparison of among the three *Panax* herbs including *P. notoginseng*, *P. ginseng* and *P. japonicus* in detail has been not reported. Comprehensive metabolite profiling of these extracts will provide greater information to elucidate the holistic therapeutic effects of *P. notoginseng* as well as *P. ginseng* and *P. japonicus* when used as TCMs. In this study, an UPLC-QTOFMS-based analytical method was developed in the analysis of three *Panax* samples and a metabolite profiling of three *Panax* herbs was performed utilizing an UPLC-QTOFMS analytical method in association with multivariate statistical analysis, not as a form of quality control, but as a tool for the identification of the variations in bioactive components among different *Panax* herbs.

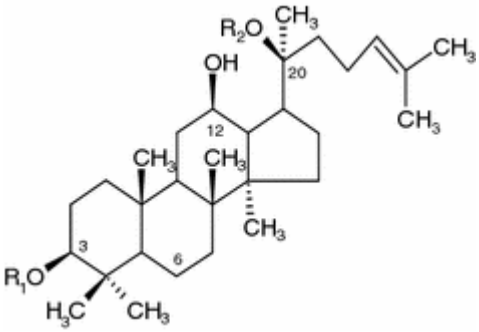
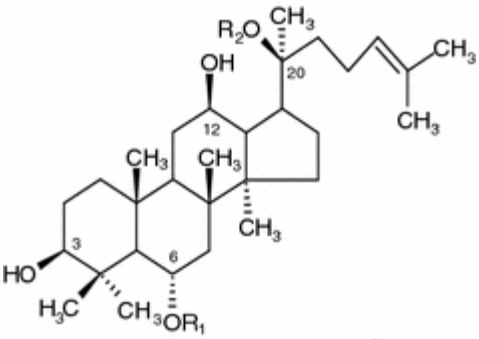
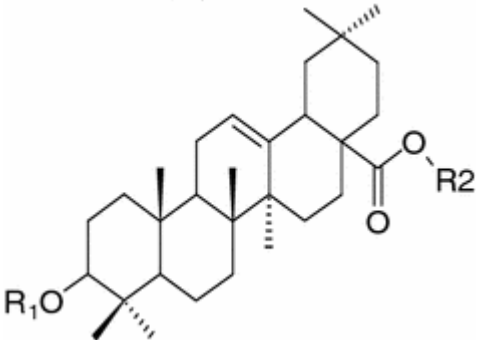
MATERIALS AND METHODS

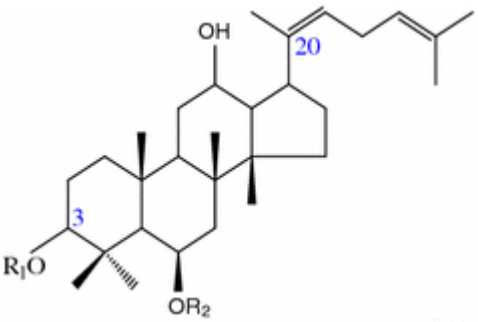
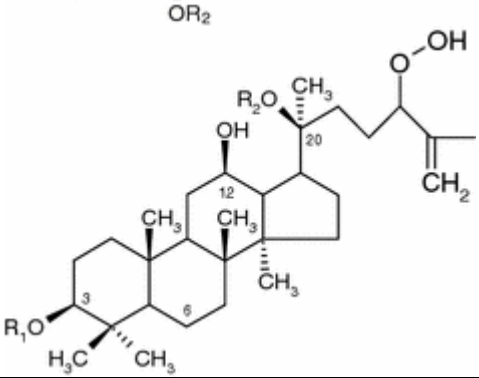
Chemicals and materials

Leucine-enkephalin and formic acid were purchased from Sigma-Aldrich (St. Louis, MO, USA). Acetonitrile and methanol of HPLC grade were obtained from Merck, Germany. Analytical-grade methanol was obtained from Shanghai Lin Feng chemical reagent Co., Ltd., P. R. China. All aqueous solutions were prepared with ultrapure water produced by Milli-Q system (18.2 MΩ, Milipore, Bedford, MA, USA). Notoginsenoside R1, ginsenoside Rb1, Rg1 and Re standards were purchased from the National Institute for the Pharmaceutical and Biological Products (Beijing, P. R. China). Ginsenoside Rb2, Rb3, Rc, Rd and F2 standards were obtained from Sanqi Research Institute of Wenshan Prefecture (Wenshan, Yunnan Province, P. R. China). Notoginsenoside R2, ginsenoside Rh_2, 20(R)-Rg3, Rg2, and F1 were obtained from Wuhu DELTA Co. Ltd (Wuhu, Anhui Province, P. R. China). The chemical structures of the investigated saponins are shown in Table 1. Eleven *Panax notoginseng* samples cultivated for 3 years were collected locally from three villages (Xinkaitian, Kanxin and Leshichong) of Wenshan Prefecture, Yunnan Province, P. R. China. Eleven authentic root samples of *Panax*

ginseng were collected locally from eleven places (Tonghua, Changbai, Huadian, Jian, Songjianghe, Dunhua, Jingyu, Lushuihe, Fusong, Linjiang and Baishan) of Jinlin Province, China and one from Shanghai Hua Yu herbs Co., Ltd, Shanghai, China. *Panax japonicus* were collected from Hubei, Sichuan and Shanxi provinces of China, respectively. They all were identified as the genuine root samples of *Panax notoginseng* (Burk.) F. H. Chen, *Panax ginseng* and *Panax japonicus* by Zhongliu, Prof. of pharmacognostics, School of Pharmacy, Shanghai Jiao Tong University of China, respectively.

Table 1: Structures of investigated saponins (Glc, β-D-glucose; Rha, α-L-rhamnose; Ara(p), α-L-arabinose (pyranose); Ara(f), α-L-arabinose (furanose); Xyl, β-D-xylose; glcU: β-D-glucuronic acid)

Chemical structures	Saponins	R1	R2
	Ginsenoside Rb1	–Glc^{2}–Glc	–Glc^{6}–Glc
	Ginsenoside Rb2	–Glc^{2}–Glc	–Glc^{6}–Ara(p)
	Ginsenoside Rc	–Glc^{2}–Glc	–Glc^{6}–Ara(f)
	Ginsenoside Rd	–Glc^{2}–Glc	–Glc
	Ginsenoside F2	–Glc	–Glc
	Ginsenoside Ra3	–Glc^{2}–Glc	–Glc^{6}–Glc
	Ginsenoside Rh2	–Glc	–H
	Protopanaxadiol	–H	–H
	Notoginsenoside R4	–Glc^{2}–Glc	–Glc^{6}–Glc^{6}–Xyl
	Notoginsenoside Fa	–Glc^{2}–Glc^{2}–Xyl	–Glc^{6}–Glc
	Chikusetsusaponin III	–Glc^{2}–Glc^{6}–Xyl	–H
	Notoginsenoside R1	–Glc^{2}–Xyl	–Glc
	Ginsenoside Rg1	–Glc	–Glc
	Ginsenoside Rg_2	–Glc^{2}–Rha	–H
	Ginsenoside Re	–Glc^{2}–Rha	–Glc
	Ginsenoside Rf	–Glc^{2}–Glc	–Glc
	Ginsenoside R0	–GlcU^{2}–Glc	–Glc
	Chikusetsusaponin IV	–GlcU^{4}–Araf	–Glc
	Chikusetsusaponin Ib	–GlcU	–Glc
	Chikusetsusaponin IVa	–GlcU^{4}–Ara(f)^{6}–Glc	–H

	Ginsenoside Rh4	–H	–Glc
	Ginsenoside II	–Glc^{2}–Glc	–Glc

Sample preparation

A standard mixture containing, ginsenoside Rb1, Rg1, Re**,** Rb2, Rb3, Rc, Rd**,** F2, Rh2, 20(R)-Rg3, Rg2, F1 and notoginsenoside R1, R2 was prepared in 70% (v/v) methanol. A volume of 2 µl of a standard mixture was used during UPLC/TOFMS method development and validation of retention time reproducibility and mass accuracy. The original root samples were pulverized with a pulverizer, and the powder was sieved through 180 µm sieves. The obtained fine powder (2 g) was weighed, and 20 ml of 70% (v/v) methanol were added. The powder was then extracted in an ultrasonic water bath at 50 °C for 60 min (Gao et al. *2008*). The extraction was repeated three times. After cooling, the solution was filtered and concentrated till dryness in vacuo; the residue was dissolved through adding little volume of 70% (v/v) methanol in succession by ultrasonication till the residue was entirely dissolved and transferred to a 50 ml volumetric flask and 70% methanol was added to scale. The fluid was filtered through a syringe filter (0.22 µm) and injected directly into the UPLC system. Blank 70% (v/v) methanol (2 µl) was injected between selected analyses to validate inter-sample cross talking effect.

UPLC-QTOFMS analysis

Panax herbs metabolite profiling was performed using a Waters ACQUITY UPLC™ system (Waters Corporation, Milford, MA, USA) which was equipped with a binary solvent delivery manager, and a sample manger coupled to Micromass Q-TOF Premier™ mass spectrometer (Waters Corporation, Milford, MA, USA) equipped with an electrospray interface. Chromatographic separations were performed on a 2.1 × 100 mm 1.7 µm ACQUITY BEH C18 chromatography column. The column was maintained at 40°C and eluted with a gradient of 10–90% acetonitrile (0.05% (v/v) formic acid)-aqueous formic acid (0.05% (v/v) formic acid) over 20 min at a flow rate of 0.35 ml/min. A 10 µl injection of this sample was made onto the column.

A metabonomics MS System Test Mixture including acetaminophen, tolbutamide, 3′-azido-3′-deoxythymidine, reserpine, verapamil and coumarin was used as chromatographic reference and mass accuracy quality control. This test mixture was injected every ten injections. The mass spectrometer was operated in positive ion mode. The desolvation temperature was set to 400°C at a flow rate of 700 l/h and source temperature of 100°C. The capillary and cone voltages were set to 3500 and 45 V, respectively. The data was collected between 400 and 1,500 *m/z* with an alternating collision energy, at 5 eV for precursor ion information generation and a collision profile from 10 to 40 eV for fragment ion information. The Q-TOF premier™ was operated in v mode with 10,000 mass resolving power. Data were centroided during acquisition using independent reference lock-mass ions via the LockSpray™ interface to ensure mass accuracy and reproducibility. Leucine-enkephalin was used as the reference compound at a concentration of 50 pg/µl and an infusion flow rate of 0.05 ml/min. The Lockspray™ was operated at a reference scan frequency of 10 or analyte to reference scan ratio of 9:1 and a reference cone voltage of 45 V. For ES+, isotopic [M+H]+ ions of leucine-enkephalin at 556.2771 Da and 557.2804 Da were used as the attenuated lock mass and lock mass, respectively. During metabolite profiling experiments, centroided data were acquired for each sample from 400 to 1,500 Da with a 0.10 s scan time and a 0.01 s interscan delay over a 20 min analysis time.

Chemometric data analysis

The UPLC-QTOFMS data of *P. ginseng*, *P. japonicus* and *P. notoginseng* samples were analyzed to identify potential discriminant variables. The ES+ raw data were analyzed by the MarkerLynx applications manager version 4.1 (Waters, Manchester, UK). The parameters used were RT range 0–20 min, mass range 400–1,500 Da, mass tolerance 0.02 Da, isotopic peaks were excluded for analysis, noise elimination level was set at 10.00, minimum intensity was set to 10% of base peak intensity, maximum masses per RT was set at 6 and, finally, RT tolerance was set at 0.01 min. After creating a suitable processing method, the next step is to process the dataset through the Create Dataset window. Selecting the method we just created and creating our dataset, we would like to (a) Detect Peaks, (b) Collect Markers, and (c) perform PCA. Select all of these options from the Processing Options panel of the Create DataSet display and at this point it is also possible to automatically Print Reports and Export data into a text file for use in third party software such as Pirouette and SIMCA-P. After data processing, a list of the intensities of the peaks detected was generated for the first sample, using retention time (RT) and *m/z* data pairs as the identifier of each peak. The resulting two-dimensional matrix of measured mass values and their intensity for each sample were further exported to SIMCA-P software 11.0 (Umetrics, Umeå, Sweden) for multivariate statistical analysis. Multivariate statistical analysis, including unsupervised principal component analysis (PCA) and supervised partial least squares-discriminant analysis (PLS-DA), were performed by SIMCA-P 11.0 software (Umetrics, Umeå, Sweden). The data sets were mean-centered, pareto-scaled in a columnwise manner before PCA and PLS-DA modeling. Mean centering subtracts the average from the data sets columnwise, thereafter resulting in a shift of the data towards the mean. Pareto scaling gives each variable a variance equal to the square root of its standard deviation. As compared to UV-scaling (scaling to unit variance) method, the advantage of using this technique lies in the fact that it enhances the contribution of lower concentration metabolites without amplifying noise and artifacts commonly present in the metabonomic data sets (Cloarec *2005*). PCA was universally used for achieving the natural interrelationship including grouping, clustering, and outliers among observations without priori knowledge of the data set. The first principal component (PC1)

contains the most variance in the data set. The second principal component, (PC2), is orthogonal to PC1, and represents maximum amount of variance not explained by PC1. The remaining components are attained in a similar manner, thereby reducing the high dimensional data sets to a two- or three-dimensional scores map without losing profound information.

Additionally, in order to maximize the variations among the *Panax* species, sophisticated supervised methods including PLS-DA was further applied. PLS-DA is a modification of PLS targeting classification and discrimination problems by pre-defining the dummy Y as a specific descriptor (e.g. 0/1). PLS is a generalized multiple regression method that can deal with multiple collinear X and Y variables and has two objectives: one is to well approximate X and Y, and the other is to model the relationship between them. R2X and R2Y represent the fraction of the variance of X matrix and Y matrix, respectively, while Q2Y suggests the predictive accuracy of the model. The cumulative values of R2X, R2Y and Q2Y close to 1 indicate an excellent model. Variable influence on projection (VIP) values greater than 1.0 are considered statistically significant for group discrimination. In SIMCA-P package, a typical cross-validation procedure was conducted by leaving 1/7th samples out in each round so as to validate the PLS-DA model against over-fitting. In order to compare the combined MS dataset from *P. notoginseng* to *P. ginseng* and from *P. japonicus* to *P. ginseng*, on a basis of the threshold of p values and fold change values from nonparametric Wilcoxon–Mann–Whitney test implemented in the Matlab statistical toolbox, the differential metabolites derived from the correlation coefficients (Cloarec *2005*) of a cross-validated PLS-DA model were validated at a univariate level.

RESULTS AND DISCUSSION

UPLC method development

The UPLC system provided a rapid, effective, and convenient analytical method for the detection of a wide range of saponins present in the three *Panax* herbs. The major benefit expected from the use of the sub-2 μm particles is the increased column efficiency and concomitant increased resolution. Under the optimum UPLC conditions mentioned in our previous work (Xie et al. *2008*), the BPI (Based Peak Intensity) chromatogram obtained from the analysis of *P. ginseng* in positive ion mode is shown in Fig. 1. The peak widths generated were on the order of 10 s at the base, giving a separation peak capacity of 130 for a 20 min separation or approximately 6 peaks per min. Furthermore, a 2D map plot of UPLC chromatogram from the three *Panax* herbs clearly shows that hundreds of peaks were detected by the UPLC system over the *m/z* scan range 400–1,500 within 20 min (Fig. 2). Based on the optimum UPLC conditions, the reproducibility of the data was studied also and the retention time variability was determined to be 0.01 s or with a relative standard deviation (RSD) value less than 0.1%. From the result we can see that the retention time reproducibility of the system was excellent.

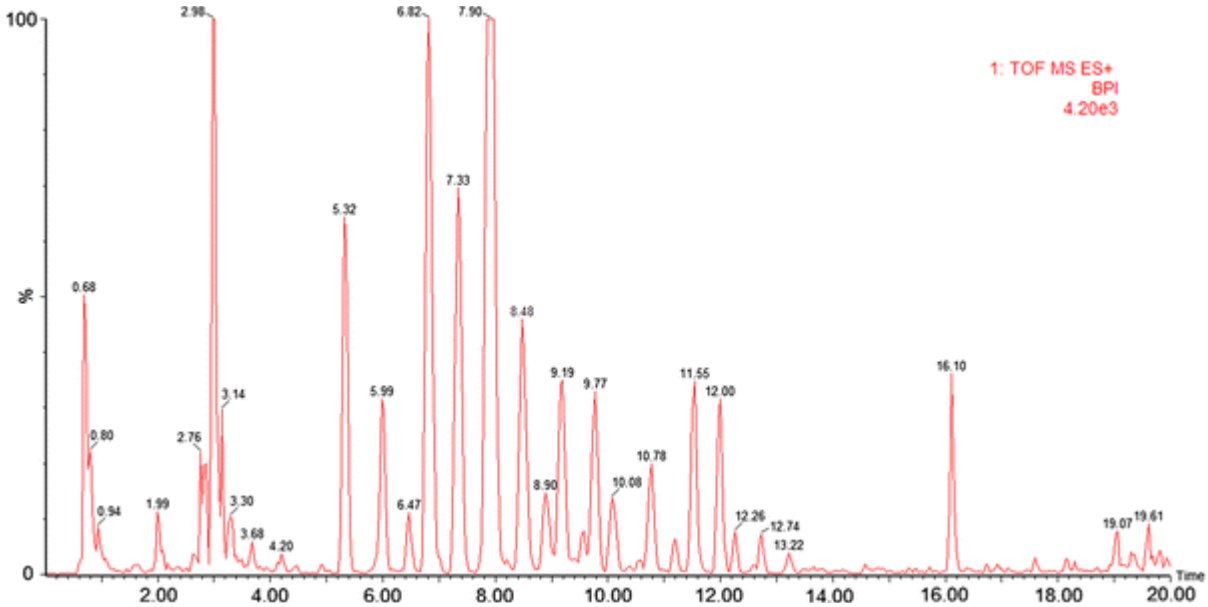


Figure 1: UPLC-TOFMS ES+ base peak intensity (BPI) chromatogram of *P. ginseng* (20 min separation on a 2.1 × 100 mm 1.7 µm ACQUITY BEH C18 UPLC column)

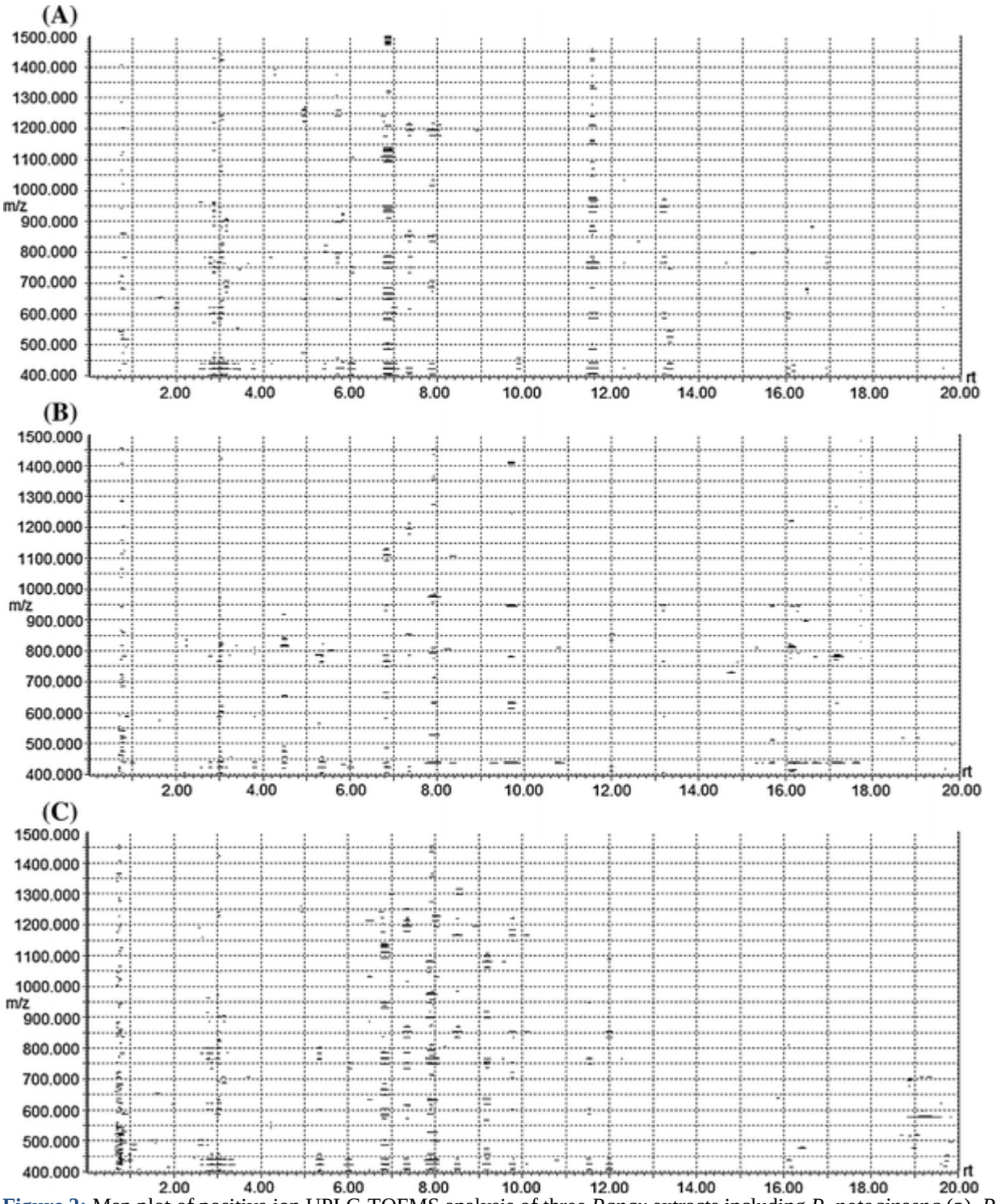


Figure 2: Map plot of positive ion UPLC-TOFMS analysis of three *Panax* extracts including *P. notoginseng* (**a**), *P. Japonicus* (**b**), and *P. ginseng* (**c**)

To illustrate the ability of the UPLC chromatography system to distinguish between the different *Panax* herbs, *P. notoginseng* (A), *P. japonicus* (B) and *P. ginseng* (C) were analyzed by the same

methodology, the results are displayed in Fig. 3 and there is a definite difference between the three samples from a visual examination of the chromatograms.

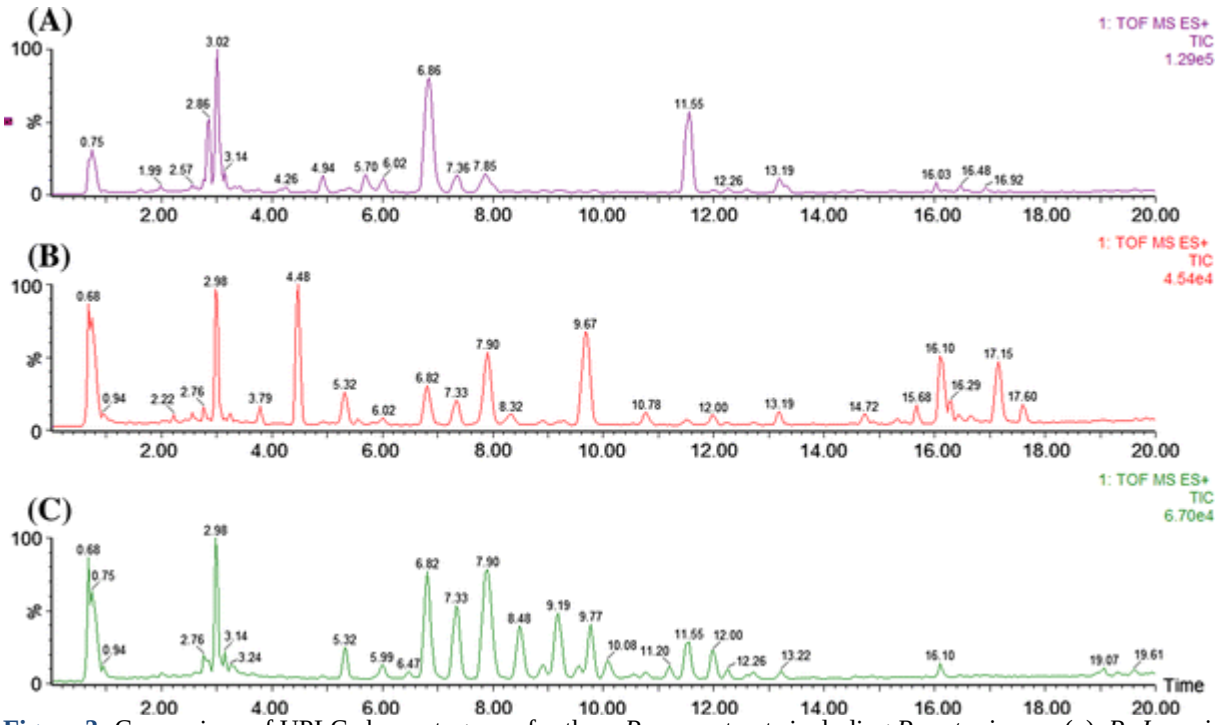


Figure 3: Comparison of UPLC chromatograms for three *Panax* extracts including *P. notoginseng* (**a**), *P. Japonicus* (**b**), and *P. ginseng* (**c**)

Mass detection conditions

More information on saponins was obtained in the ES+ mode than ES− mode due to higher degree of ion fragmentation as revealed by the representative total ion current (TIC) spectra on the UPLC-TOFMS. The mass detection conditions were optimized according to our previous work (Xie et al. 2008). The precision and reproducibility of peak integration was good for the final results and the ability to collect the MS and MS/MS data in one analytical run was clearly demonstrated (Kevin et al. 2007) and the sampling rate is sufficient to provide an accurate representation of the chromatographic separation.

The accurate mass measurement and MS/MS analysis of saponins in this study was performed at different collision energies ranging from 5 to 40 eV with MS^E (E represents low-high collision energy) approach (Xie et al. 2008) in order to provide sufficient information necessary to distinguish saponins with the same molecular weight, accurate MS and available standards were used to verify these important marker ions. Mass measurement of the TOF mass spectrometer (Chan et al. 2007) with a group of known low-molecule-weight (between 100 and 1,500 *m/z*) compounds was determined according to our previous work (Xie et al. 2008). The average mass accuracy was determined to be 1.14 ppm and from this data, using the MS/MS mode, the exact mass of a molecule can be obtained so as to be able to acquire a relevant identification. At the same time, accurate mass measurement was rendered by the simultaneous but independent acquisition of reference ions of leucine–enkephalin via the LockSpray interface to ensure the

mass accuracy and through this, problems associated with post column addition of a reference compound such as ion suppression due to reference ions, reference ionization efficiency and mass interference were avoided.

Principal component analysis

A two-component PCA score plot of UPLC-QTOFMS data was utilized to depict general variation of saponins among the three types of *Panax* samples (Fig. 4a). According to the data pre-treat methods used in our previous work (Xie et al. 2008), clear separation in the PCA space of the UPLC-TOFMS analysed *Panax* samples from the three groups (Fig. 4) was illustrated. Visual examination of the chromatograms of UPLC-TOFMS displayed clear difference between groups. The clear separation of the three different herbs was observed in the PCA scores plot where each coordinate represents a sample (Fig. 4). Two-component PCA model cumulatively accounted for 47.7% of variation. The PCA scores plot in Fig. 4(a) could be readily divided into three big clusters: one *P. notoginseng*, the other *P. ginseng*, and the third *P. japonicus*, confirming that the *P. ginseng*, *P. japonicus* and *P. notoginseng* were different in the levels and occurrence of ginsenosides. Each sample was represented as a point in a scores plot (*Panax japonicus* = □, *Panax ginseng* = ○, and *Panax notoginseng* = ◊). The *P. japonicus* and the others were clearly separated by the principal component 1 (PC1), whereas the *P. ginseng* was easily discriminated by the principal component 2 (PC2) with the others. In this work, therefore, PCA scores plot and loadings plot were achieved using mean-centered, Par-scaled data. Additionally, a two-component PLS-DA model (Fig. 4b) (R2X = 0.466, R2Y = 0.922, Q2Y = 0.914) was subsequently constructed to identify and reveal the differential metabolites in response to the different *Panax* herbs, which were partially validated with the putative standards. The preferential distribution of marker ions in the second quadrant of the loadings plot accounted primarily for the difference of *P. ginseng*. Analogously, the distribution in the third and fourth quadrants of the loadings plot indicated the variation of *P. notoginseng* and *P. japonicus*, respectively.

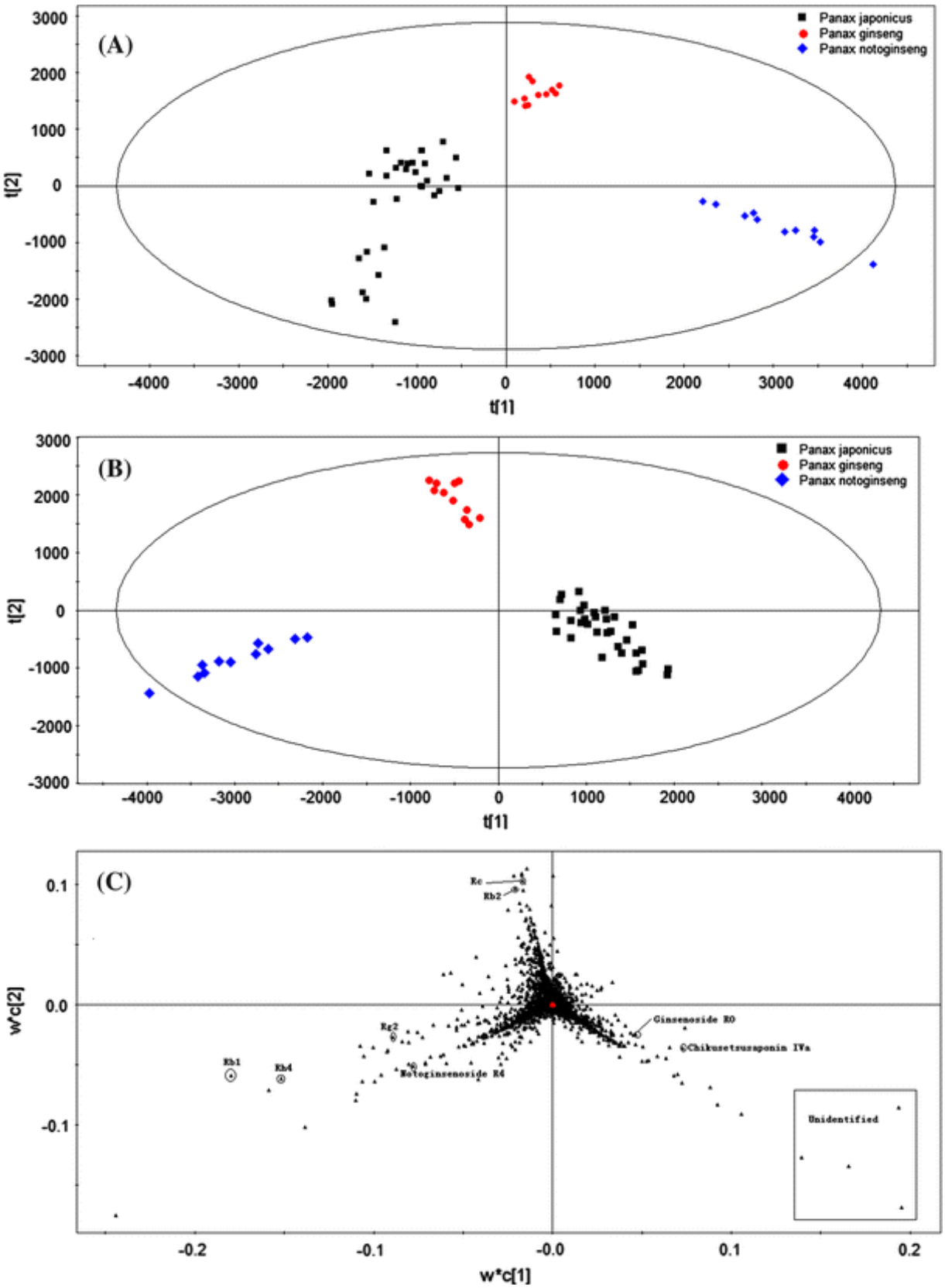


Figure 4: (**a**) PCA scores plot, (**b**) PLS-DA scores plot of (*Panax japonicus* = □, *Panax ginseng* = ○, and *Panax notoginseng* = ◊) and (**c**) PLS-DA loadings plot obtained using Pareto scaling with mean centering

The loadings plot, as presented in Fig. 4c, was obtained using Par scaling with mean-centering. Although some information of low intensity might be lost due to Par scaling, the loadings plot clearly facilitated the profiling of the marker ions in our experiments. The variation in PC2 in the loadings plot was due to the marker ions that accounted for the differences between the *P. ginseng* and the others as observed in the scores plot. These spectral variables were further confirmed using the ion intensity plot. The ion intensity plot (Fig. 5) obtained within MarkerLynx or SIMCA P 11.0 software was found to be a convenient instrument to aid the profiling of marker ions. The representative ion intensity plot (Fig. 5) illustrated the abundance of marker ions ginsenoside Rc at *m/z* 1079.6008 (RT 9.19 min), ginsenoside Rb1 at *m/z* 1109.6108 ($[M + H]^+$, Rt 6.82 min), and chikusetsusaponin IVa at *m/z* 795.4532 ($[M+H]^+$, Rt 7.90 min) over 54 samples. The ions fulfilled the criteria of marker ions because they were found only to be present or have higher level in one herb but not the others. Furthermore, variables with high relevance for explaining the differences of the three *Panax* samples were also identified from VIP (variable importance in the projection) values. Large VIP values, more than 1.0, are the most relevant for explaining the differences of the three *Panax* samples. Chikusetsusaponin IVa, ginsenoside R0, ginsenoside Rc, ginsenoside Rb1, ginsenoside Rb2 and ginsenoside Rg2 were found to be significant in creating a differentiation model for the three *Panax* herbs. Among the critical metabolites, chikusetsusaponin IVa and ginsenoside R0 are the active saponins of *P. japonicus* and has the effect of inhibiting the activity of cyclic AMP phos-phodiesterase (Guo et al. 2004), ginsenoside Rc and Rb2 have the calcium channel blockade and anti-free-radical actions (Zhong et al. 1995).

Tentative peak assignment

It is possible to determine the importance of variables by analyzing the correlation between variables in the PC1 and PC2 dimension; a list of marker ions (metabolites) of interest was therefore obtained from the PLS-DA loadings weight (Fig. 4). The experimentally determined *m/z* of the selected marker ion was used to compute the possible calculated mass, mass accuracy (mDa and ppm), double-bond equivalent (DBE, total number of rings and double bonds in a molecule), empirical formula, and i-Fit value (the likelihood that the isotopic pattern of the elemental composition matches a cluster of peaks in the spectrum) and up to 50 best isotopic matches for each mass of the potential candidate ions with elemental composition analysis. The potential calculated masses, mass accuracy, DBE, i-FIT value and elemental compositions associated with the measured mass of the marker ion were generated and studied with MarkerLynx. Using the potential marker ion in Fig. 4 as an example, the measured mass of the ion was determined to be 795.4532 Da (chikusetsusaponin IVa) and its isotopic pattern indicated it as a singly charged ion. Using a mass tolerance of 100 mDa or 10 ppm, four possible candidates were produced with the elemental composition analysis software implemented in the MarkerLynx. The calculated masses/mDa/ppm/DBE/i-FIT/formula of the four candidates were 795.4531/0.1/0.1/9.5/6.7/C42H67O14, 795.4590/-5.8/-7.3/0.5/9.1/C35H71O19, 795.4472/6.0/7.5/18.5/17.7/C49H63O9 and 795.4566/-3.4/-4.3/31.5/54.7/C60H59O, respectively. Taking mass accuracy (mDa and ppm), DBE and i-FIT value into consideration, the confidence of the tentative marker assignment to the first candidate was enhanced by the high mass accuracy obtained (0.1 mDa or 0.1 ppm), fulfillment of 9 DBE (7 rings and 2 double bonds) and low i-FIT value (6.7). Marker ions *m/z* 795.4532 ($[M+H]^+$, Rt 7.90 min) is far from the centre in the fourth quadrant of the PLS-DA loadings plot, suggesting that the compound

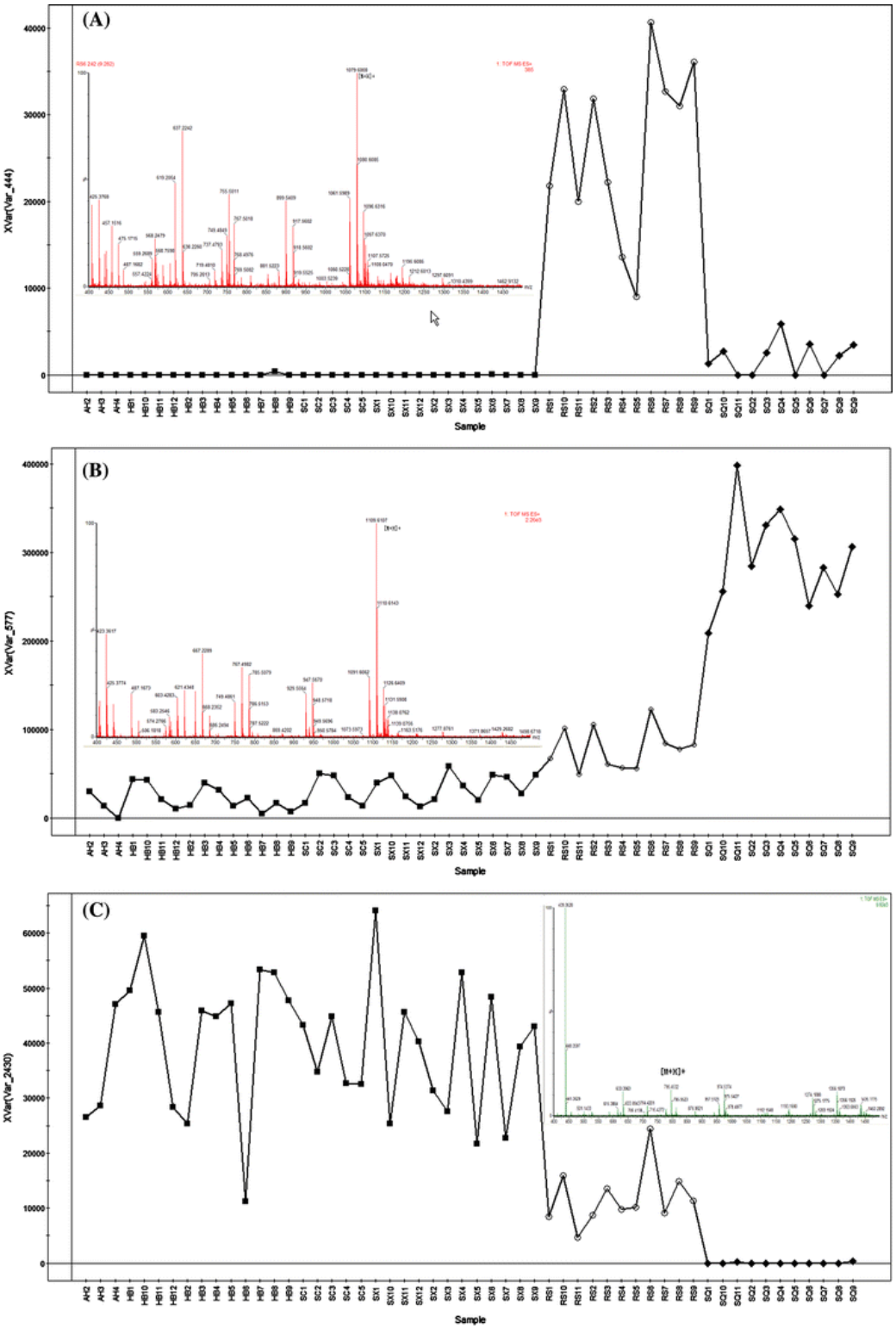


Figure 5: Representative ion intensity plot for marker ions ginsenoside Rc at *m/z* 1079.6008 (RT 9.19 min) (**a**), ginsenoside Rb1 at *m/z* 1109.6108 ([M+H]+, Rt 6.82 min) (**b**), and chikusetsusaponin IVa at *m/z* 795.4532 ([M+H]+, Rt 7.90 min) (**c**) over 54 samples

exhibits a higher importance in *P. japonicus* than that in the other two herbs. This marker was tentatively identified as chikusetsusaponin IVa and this was the characteristic saponin of *P. japonicus.* Marker ions 1109.6108 ($[M+H]^+$, Rt 6.82 min), 621.4328 ($[M+H]^+$, Rt 3.02 min), 1241.6522 ($[M+H]^+$, Rt 4.90 min), and 785.5065 ($[M+H]^+$, Rt 6.82 min) mainly present in the third quadrant are to characterize the *P. notoginseng*. Marker ions 1079.6008 ($[M+H]^+$, Rt 9.19 min) mainly present in the second quadrant are to characterize the *P. ginseng*. In order to identify these important marker ions tentatively, qualitative analysis of saponins present in the *Panax* samples was carried out on UPLC-MS/MS system. Saponins were verified by comparing the major fragment ions, retention time and accurate mass of marker ions with those of available standards. A total of eleven saponins were verified in *P. ginseng*, eight in *P. notoginseng* and sixteen in *P. japonicus* using this method (Table 2). Fold changes in relative concentration of each saponin of interest were determined for the three *Panax* herbs and the corresponding heat map plot was illustrated to envisage such alterations between groups (Fig. 6). From the figure we can see that some saponins such as ginsenoside Rb1 in *P. notoginseng* was higher than in *P. ginseng*, while in *P. japonicus* was lower than in *P. ginseng* and this was consistent with the results obtained from the loadings plot.

Table 2: Saponins observed in *Panax* samples of *P. ginseng*, *P. notoginseng* and *P. japonicus*

***Panax* sample**	**No.**	**Rentention time (min)**	**Identity**	**Formula [M+H]+**	**Actual mass**	**Exact mass**	**Mass error (ppm)**
P. ginseng	1	7.85	Ginsenoside Rb2	C53H91O22	1079.6056	1079.6002	5.0
	2	9.19	Ginsenoside Rc	C53H91 O22	1079.6008	1079.6002	0.6
	3	6.82	Ginsenoside Rb1	C54H93O23	1109.6099	1109.6108	−0.8
	4	2.98	Ginsenoside Re	C48H83O18	947.5574	947.5579	−0.5
	5	13.15	Ginsenoside Rd	C48H83O18	947.5532	947.5579	−5.0
	6	6.82	Ginsenoside Rg2	C42H73O13	785.5089	785.5051	4.8
	7	3.02	Ginsenoside Rg1	C42H73O14	801.5048	801.5000	6.0
	8	5.32	Ginsenoside Rf	C42H73O14	801.5015	801.5000	1.9
	9	4.94	Ginsenoside Ra3	C59H101O27	1241.6577	1241.6530	3.8
	10	7.94	Ginsenoside II	C48H83O20	979.5470	979.5478	−0.8
	11	3.02	Ginsenoside Rh4	C36H61O8	621.4353	621.4366	−2.1
P. notoginseng	1	2.83	Notoginsenoside R1	C47H81O18	933.5424	933.5423	0.1
	2	6.82	Ginsenoside Rg2	C42H73O13	785.5056	785.5051	0.6
	3	13.19	Ginsenoside Rd	C48H83O18	947.5560	947.5579	−2.0
	4	6.82	Ginsenoside Rb1	C54H93O23	1109.6096	1109.6108	−1.1
	5	4.90	Notoginsenoside R4	C59H101O27	1241.6539	1241.6530	0.7
	6	5.70	Notoginsenoside Fa	C59H101O27	1241.6450	1241.6530	−6.4
	7	5.41	Ginsenoside Rf	C42H73O14	801.5009	801.5000	1.1
	8	3.02	Ginsenoside Rg1	C42H73O14	801.5026	801.5000	3.2
P. japonicus	1	5.29	Chikusetsusaponin IVa methyl ester	C43H69O14	809.4669	809.4687	−2.2
	2	9.67	Chikusetsusaponin IV	C47H75O18	927.4955	927.4953	0.2
	3	8.32	Chikusetsusaponin Ib	C47H75O18	927.4983	927.4953	3.2
	4	7.90	Chikusetsusaponin IVa	C42H67O14	795.4547	795.4531	2.0
	5	4.90	Notoginsenoside R4	C59H101O27	1241.6584	1241.6530	4.3
	6	6.82	Ginsenoside Rg2	C42H73O13	785.5051	785.5051	0.0
	7	7.90	Ginsenoside R0	C48H77O19	957.5109	957.5059	5.2
	8	4.48	Protopanaxadiol	C30H53O4	477.3902	477.3944	−8.8
	9	3.02	Ginsenoside Rh2	C36H63O8	623.4554	623.4523	5.0
	10	2.76	Ginsenoside Rb2	C53H91O22	1079.6006	1079.6002	0.4
	11	6.82	Ginsenoside Rb1	C54H93O23	1109.6167	1109.6108	5.3
	12	3.02	Ginsenoside Rg1	C42H73O14	801.4996	801.5000	−0.5
	13	5.55	Ginsenoside Rf	C42H73O14	801.5030	801.5000	3.7
	14	2.98	Ginsenoside Re	C48H83O18	947.5576	947.5579	−0.3
	15	13.19	Ginsenoside Rd	C48H83O18	947.5624	947.5579	4.7
	16	7.81	Chikusetsusaponin III	C47H81O17	809.4669	809.4687	−2.2

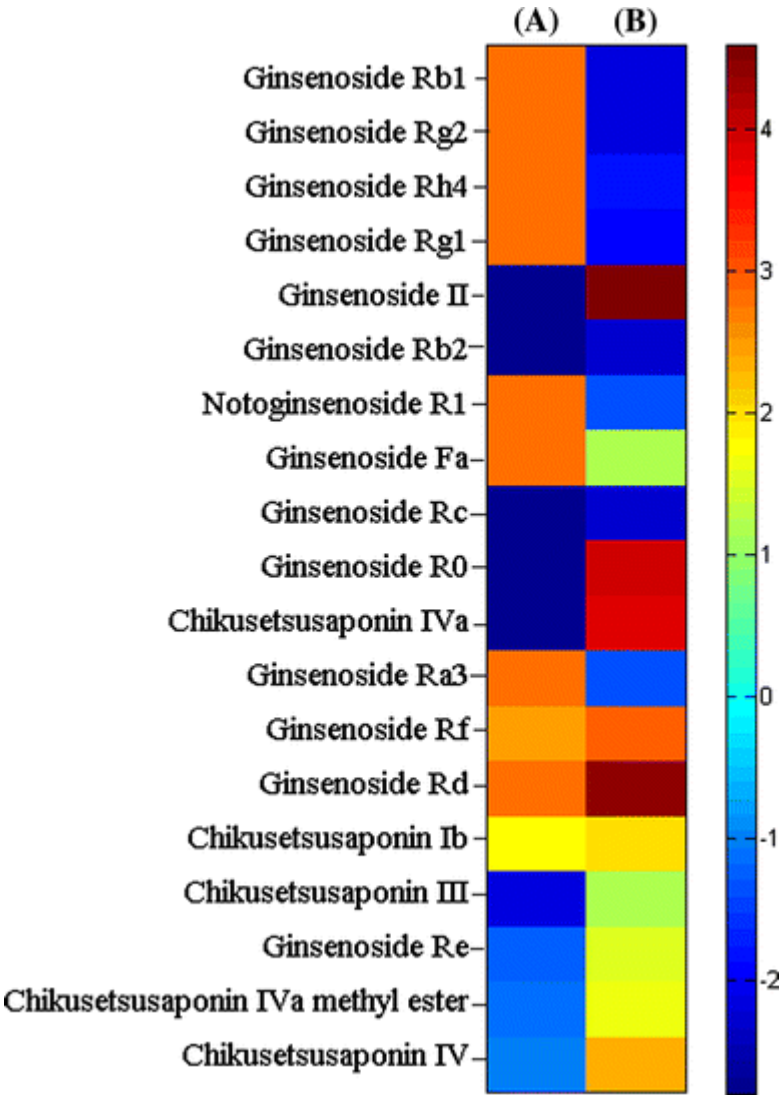


Figure 6: Fold changes of saponins in *P. notoginseng*, *P. ginseng* and *P. japonicus.* (**a**) the ratio of the level of saponins in *P. notoginseng* to *P. ginseng*; (**b**) the ratio of the level of saponins in *P. japonicus* to *P. ginseng*. Positive value means that the content of each saponin in *P. notoginsng* or *P. japonicus* was higher than in *P. ginseng*, whereas, the negative value means that the content of each saponin in *P. notoginsng* or *P. japonicus* was lower than in *P. ginseng*

CONCLUDING REMARKS

This work has developed and optimized a convenient, high-throughput, and reliable UPLC-QTOFMS method to analyze a group of saponins of *P. notoginseng*, *P. ginseng* and *P. japonicus*, which is applicable for analysis and evaluation of complex herbal medicines. This study also demonstrates that metabolomics analysis using UPLC-QTOFMS combined with multivariate statistical analysis provides some useful information in the study of *Panax* herbs and can be used as a powerful tool to profile and differentiate phytochemical compositions among different kinds of *Panax* samples. The marker compounds such as ginsenoside Rb1, Rc and chikusetsusaponin IVa accountable for the different metabolite profiles of *P. ginseng*, *P. notoginseng* and *P. japonicus* were observed and the results can provide a useful additional character in the authentication of these *Panax* herbs. Thus, such a UPLC-QTOF-MS-based metabolite profiling

strategy is promising for revealing and elucidating the metabolic outcomes as a result of geographical variations, and genetic diversity, seasonal and different cultivation methods employed as well. Hence, this work is of great importance for the evaluation of overall quality of medicinal plants, ultimately of great significance in the pharmacological and clinical investigation of drug products of plant origin.

ACKNOWLEDGEMENTS

This study was financially supported by the International Collaboration Project (project number 2006DFA02700) of the Ministry Science and Technology of the People's Republic of China. We also thank Yongbo Wang for assisting with sample collection.

REFERENCES

Akerele, O. (1992). WHO guidelines for the assessment of herbal medicines. *Fitoterapia, 63,* 99–104.

Anterola, A. M., & Lewis, N. G. (2002). Trends in lignin modification: A comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. *Phytochemistry, 61,* 221–294. doi:10.1016/S0031-9422(02)00211-X.

Brekhman, I., & Dardymov, I. (1969). New substances of plant origin which increase nonspecific resistance. *Annual Review of Pharmacology, 9,* 419–430. doi:10.1146/annurev.pa.09.040169.002223.

Chan, E. C. Y., Yap, S. L., Lau, A. J., Leow, P. C., Toh, D. F., & Koh, H. L. (2007). Ultra-performance liquid chromatography/time-of-flight mass spectrometry based metabolomics of raw and steamed *Panax notoginseng*. *Rapid Communications in Mass Spectrometry, 21,* 519–528. doi:10.1002/rcm.2864.

Chang, W. T., Thissen, U., Ehlert, K. A., Koek, M. M., Jellema, R. H., Hankemeier, T., et al. (2006). Effects of growth conditions and processing on *Rehmannia glutinosa* using fingerprint strategy. *Planta Medica, 72,* 458–467. doi:10.1055/s-2005-916241.

Cloarec, O., Dumas, M. E., Trygg, J., Craig, A., Barton, R. H., Lindon, J. C., et al. (2005). Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies. *Analytical Chemistry, 77,* 517–526. doi:10.1021/ac048803i.

Eisenreich, W., & Bacher, A. (2007). Advances of high-resolution NMR techniques in the structural and metabolic analysis of plant biochemistry. *Phytochemistry, 68,* 2799–2815. doi:10.1016/j.phytochem.2007.09.028.

Farag, M. A., Huhman, D. V., Lei, Z., & Sumner, L. W. (2007). Metabolic profiling and systematic identification of flavonoids and isoflavonoids in roots and cell suspension cultures of Medicago truncatula using HPLC-UV-ESI-MS and GC-MS. *Phytochemistry,*

68, 342–354. doi:10.1016/j.phytochem.2006.10.023.

Fiehn, O. (2002). Metabolomics-the link between genotypes and phenotypes. *Plant Molecular Biology, 48*, 155–171. doi:10.1023/A:1013713905833.

Fiehn, O., & Weckwerth, W. (2003). Deciphering metabolic networks. *European Journal of Biochemistry, 270*, 579–588. doi:10.1046/j.1432-1033.2003.03427.x.

Fiehn, O., Kopka, J., Do¨rmann, P., Altmann, T., Trethewey, R. N., & Willmitzer, L. (2000). Metabolic profiling for plant functional genomics. *Nature Biotechnology, 18*, 1157–1161. doi:10.1038/81137.

Fukusaki, E., Jumtee, K., Bamba, T., Yamaji, T., & Kobayashi, A. (2006). Metabolic fingerprinting and profiling of *Arabidopsis thaliana* leaf and its cultured cells T87 by GC/MS. *Zeitschrift Fur Naturforschung, 61*, 267–272.

Gao, X. D. M., Zhao, A., Xie, G., & Jia, W. (2008). Simultaneous determination of saponins in flower buds of *Panax notoginseng* using high performance liquid chromatography. *Biomedical Chromatography, 22*, 244–249. doi:10.1002/bmc.915.

Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G., & Kell, D. B. (2004). Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends in Biotechnology, 22*, 245–252. doi:10.1016/j.tibtech.2004.03.007.

Guan, J., Lai, C. M., & Li, S. P. (2007). A rapid method for the simultaneous determination of 11 saponins in *Panax notoginseng* using ultra performance liquid chromatography. *Journal of Pharmaceutical and Biomedical Analysis, 44*, 996–1000. doi:10.1016/j.jpba.2007.03.032.

Guo, Q., Yu, X. L., & Hu, X. X. (2004). Progress of chemical and pharmacological research of *Panax japonicus*. *Research & Information on Traditional Chinese Medicine, 6*, 17–19.

Hall, R. D. (2006). Plant metabolomics: from holistic hope, to hype, to hot topic. *The New Phytologist, 169*, 453–468. doi:10.1111/j.1469-8137.2005.01632.x.

Holmes, E., Tang, H. R., Wang, Y. L., & Seger, C. (2006). The assessment of plant metabolite profiles by NMR-based methodologies. *Planta Medica, 72*, 771–785. doi:10.1055/s-2006-946682.

Huhman, D. V., & Sumner, L. W. (2002). Metabolic profiling of saponins in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. *Phytochemistry, 59*, 347–360. doi:10.1016/S0031-9422(01)00432-0.

Jiangsu New Medical College. (1985). *Dictionary of Chinese Materia Medica*. Shanghai Science and Technology Press, Shanghai, pp. 417–420.

Kevin, P. B., Jose, C. P., Mark, W., John, P. S., Kate, Y., Renata, O., et al. (2007). MS^E with mass defect filtering for in vitro and in vivo metabolite identification. *Rapid Communications in Mass Spectrometry, 21*, 1485–1496. doi:10.1002/rcm.2996.

Long, M., Millar, D. J., Kimura, Y., Donovan, G., Rees, J., Fraser, P. D., et al. (2006). Metabolite profiling of carotenoid and phenolic pathways in mutant and transgenic lines of tomato: Identification of a high antioxidant fruit line. *Phytochemistry, 67*, 1750–1757. doi:10.1016/j.phytochem.2006.02.022.

Mahady, G. B., Fong, H. H. S., & Farnsworth, N. R. (2001). Botanical Dietary Supplements: Quality, Safety, and Efficacy. Lisse, The Netherlands: Swets & Zeitlinger Publishers, B.V.

National Committee of Pharmacopoeia. (2005). *Pharmacopoeia of the People' s Republic of China* (Vol. I, pp. 10). Beijing: Chemical Industry Press.

Nordstrom, A., O'Maille, G., Qin, C., & Siuzdak, G. (2006). Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum. *Analytical Chemistry, 78*, 3289–3295. doi:10.1021/ac060245f.

Oliver, S. G., Winson, M. K., Kell, D. B., & Baganz, F. (1998). Systematic functional analysis of the yeast genome. *Trends in Biotechnology, 16*, 373–378. doi:10.1016/S0167-7799(98)01214-1.

Rochfort, S. (2005). Metabolomics reviewed: A new "omics" platform technology for systems biology and implications for natural products research. *Journal of Natural Products, 68*, 1813–1820. doi:10.1021/np050255w.

Shen, Y. F., Zhang, R., Moore, R. J., Kim, J., Metz, T. O., Hixson, K. K., et al. (2005). Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1,000–1,500 and capabilities in proteomics and metabolomics. *Analytical Chemistry, 77*, 3090–3100. doi:10.1021/ac0483062.

Tarachiwin, L., Ute, K., Kobayashi, A., & Fukusaki, E. (2007). 1H NMR based metabolic profiling in the evaluation of Japanese green tea quality. *Journal of Agricultural and Food Chemistry, 55*, 9330–9336. doi:10.1021/jf071956x.

Wagner, S., Scholz, K., Donegan, M., Burton, L., Wingate, J., & Völkel, W. (2006). Metabonomics and biomarker discovery: LC-MS metabolic profiling and constant neutral loss scanning combined with multivariate data analysis for mercapturic acid analysis. *Analytical Chemistry, 78*, 1296–1305. doi:10.1021/ac051705s.

Wang, M., Lamers, R. J. A. N., Korthout, H. A. A. J., Van Nesselrooij, J. H. J., Witkamp, R. F., Van Der Heijden, R., et al. (2005). Metabolomics in the context of systems biology: Bridging Traditional Chinese Medicine and molecular pharmacology. *Phytotherapy*

Research, 19, 173–182. doi:10.1002/ptr.1624.

Want, E. J., Cravatt, B. F., & Siuzdak, G. (2005). The expanding role of mass spectrometry in metabolite profiling and characterization. *ChemBioChem, 6*, 1941–1951. doi:10.1002/cbic.200500151.

Wei, J. X., Chen, Y. G., & Du, Y. C. (1991). Introduction of separation method of dammarane-type saponins in Panax samples. *Academic Journal of Kunming Medical College, 12*, 56–64.

Wilson, I. D., Nicholson, J. K., Castro-Perez, J., Granger, J. H., Johnson, K. A., Smith, B. W., et al. (2005). High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *Journal of Proteome Research, 4*, 591–598. doi:10.1021/pr049769r.

Xie, G. X., Plumb, R., Su, M. M., Xu, Z. H., Zhao, A. H., Qiu, M. F., et al. (2008). Ultra-performance LC/TOF MS analysis of medicinal Panax herbs for metabolomic research. *Journal of Separation Science,* 1015–1026. doi:10.1002/jssc.200700650.

Zhong, G. G., Sun, C. W., Li, Y. Y., Qi, H., Zhao, C. Y., Jiang, Y., et al. (1995). Calcium channel blockade and anti-free-radical actions of panaxadiol saponins Rb1, Rb2, Rb3, Rc, and Rd. *Acta Pharmacologica Sinica, 16*, 255–260.

Zuo, R., Yuan, D. (2005). Research on chemical constituents and biological activities of Rhizoma of *Panax japonicus*. *Lishizhen Medicine and Materia Medica Research, 16*, 838–841.