Journal of Chromatography A, 1227 (2012) 181–193

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Virtual separation of phytochemical constituents by their adduct-ion patterns in full mass spectra

Yufeng Zhang, Jianyang Pan, Jinghua Zhong, Yi Wang, Xiaohui Fan*, Yiyu Cheng

College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

ARTICLE INFO

Article history:
Received 31 July 2011
Received in revised form
26 December 2011
Accepted 1 January 2012
Available online 11 January 2012

Keywords:
Classifier for traditional Chinese medicine
Wei-Fu-Chun tablet
Liquid chromatography/mass spectrometry

ABSTRACT

In the present study, a tool called classifier for traditional Chinese medicine (CTCM) was developed to facilitate the discrimination of phytochemical constituents in two-dimensional datasets of liquid chromatography/mass spectrometry (LC/MS). Based on the full mass spectral characteristics of components in a mixture, particularly their adduct-ion patterns, an entire LC/MS dataset can be separated into several sub-datasets, each corresponding to one or several types of natural products. CTCM has been verified using 24 standard compounds and successfully applied in two previously reported LC/MS datasets, which confirmed the capability of proposed tool to extract adduct-ion patterns from LC/MS datasets. Moreover, the LC/MS dataset of a Wei-Fu-Chun (WFC) tablet, a prescription drug consisting of three crude herbs used for the treatment of enteric diseases, was analyzed using CTCM. The analysis indicated that the compounds in WFC could be split into three groups, with the main constituents including saponins from Radix Ginseng Rubra, flavonoids from Fructus aurantii, and phenolic compounds from Isodon amethystoides. The major compounds in the three groups were either positively identified or tentatively characterized by multi-stage and high resolution MS. The proposed tool provides a novel approach for processing the LC/MS datasets of complex samples, such as traditional Chinese medicine and botanical drugs.

1. Introduction

Traditional Chinese medicine (TCM), which is usually composed of one or several herbs, is a complex mixture containing hundreds of chemical constituents that may be responsible for its therapeutic effects. Liquid chromatography/mass spectrometry (LC/MS) is widely used in the chemical identification of TCM [1,2]. The pivotal step in studying the chemical composition of TCM is to identify and annotate the number of compounds embodied in an LC/MS chromatogram. The work is usually performed by manually selecting peaks from a total ion chromatogram (TIC). Due to the two-dimensional character of LC/MS data, extracting useful data is a time- and labor-consuming process. The number of compounds that could be found relies largely on the complexity of the dataset and the experience of the operators. Moreover, compounds may be missed when co-eluted or enveloped in TIC peaks with higher abundance. Many algorithms and software have been developed to manipulate LC/MS datasets [3–6]. Most available tools, however, are specially designed for the analysis of "-omics" data, and those for TCM or herbal medicine remain scarce. Therefore, it would be desirable to develop additional tools with which to facilitate the analysis of large LC/MS datasets.

Electrospray ionization (ESI) source is the most popular interface in LC/MS configurations. The types of ions ($[M\pm X]^{+/-}$, X = H, Na, K, HCOO, NH_3, etc.) and their relative intensities in their full mass spectra are correlated with the structures of analytes, organic modifiers, and ion source parameters [7,8]. Among these factors, the intrinsic trait of a compound (structure) is the most important parameter. Ng et al. [9] indicated that ginsenosides containing different numbers of glycosyl groups generate different ginsenoside–acetate adduct anions or deprotonated ginsenosides in ESI sources with the addition of ammonium acetate. The intensity ratios of $[M+2OAc]^{2-}$ and $[M+OAc]^{-}$ depend on the chain length of the glycosyl groups. Chan et al. [10] found that most ginsenosides afford $[M+HCOO]^{-}$ and $[M+2HCOO]^{2-}$ adduct ions, and the abundance of the expected deprotonated molecule $[M-H]^{-}$ is relatively low due to the use of formic acid as a mobile-phase modifier. An overview of mass spectrometric methods used in the structure analysis of flavonoids in the last decade has been given by Cuyckens and Claeys [11], who pointed out that the peak at the highest m/z ratio is not always the molecular ion species ($[M+H]^{+}$ in the positive mode and $[M-H]^{-}$ in the negative mode), because adducts with solvent and/or acid molecules, as well as molecular complexes ($[2M+H]^{+}$ or $[2M-H]^{-}$), could be generated. An increase in cone voltage reduces the incidence of both adduct and complex formation. Zhou et al. [12] analyzed

* Corresponding author at: Room 325, College of Pharmaceutical Sciences, Zhejiang University (Zijingang Campus), Hangzhou 310058, China.
Tel.: +86 571 88208426; fax: +86 571 88208426.
E-mail address: fanxh@zju.edu.cn (X. Fan).

0021-9673/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2012.01.006

the chemical constituents of *Glycyrrhiza uralensis* using HPLC/(−) ESI-MS^n. In the negative ion mode, both flavonoids and coumarins show $[M+Cl]^-$, $[M-H]^-$, $[M+HSO_4]^-$, and $[2M-H]^-$ ions, whereas triterpene saponins exhibit abundant $[M-H]^-$ ions, but no $[M+Cl]^-$ ions. Thus, the constituents of natural products could be ionized by ESI sources by protonation/deprotonation or coordination with cations/anions, and the distribution of these ions in their mass spectrum depends mainly on their structures when analytical parameters are set. These characteristics could be used to classify compounds in a complex herbal extract and accelerate their identification.

The aim of present study was to develop a computer tool to aid in the analysis of complex TCM samples. First, two previously reported LC/MS datasets, one low and one high resolution, were utilized to investigate the parameters of this tool and confirm its ability to extract adduct-ion patterns from different LC/MS datasets. Then, this tool was applied to analyze Wei-Fu-Chun (WFC) tablet, which is composed of three herbs, Radix Ginseng Rubra, Fructus aurantii, and Isodon amethystoides. WFC tablet is a nationally protected traditional Chinese medicine used for the treatment of gastric precancerous lesions [13]. In clinical practice, it has been proven to be effective in treating enteric diseases, such as chronic atrophic gastritis, chronic superficial gastritis, and functional dyspepsia [14–16]. Obtaining a useful chemical profile containing the great majority of constituents from the WFC tablet is of significant importance in understanding its therapeutic basis; this profile may also prove useful for the quality control of this product. Owing to its chemical complexity, the analysis of WFC tablets could prove to be more challenging than that of an individual herbal drug. Although the chemical constituents of a single medicinal herb have been previously reported, to the best of our knowledge, little is known about the chemical composition of WFC tablets. After classification of the dataset of the WFC tablet, the entire dataset was simplified into three sub-datasets, in which the overlapping peaks were separated and the pure mass spectrum of each peak obtained.

2. Theory

The ions in an ESI/MS are composed of protonated/deprotonated molecules, fragment ions, adduct ions and cluster ions. The relationship between a protonated/deprotonated molecule (X) and its related ions (Y, fragment ions, adduct ions and cluster ions) could be expressed by an algebraic equation: $Y = f(X)$. The m/z values of fragment ions or cluster ions are always compound-specific or too bigger to be in the m/z scan range. As discussed in Section 1, only the adduct ions are ideal to represent the common attribute of a certain type of compounds. According to the different equations ($Yi = f(Xi)$, $i = 1, 2, 3, \ldots$), compounds having different adduct-ion patterns could be selectively extracted from a complex LC/MS dataset and the obtained sub-datasets are relatively simple and easy to be handled by analyzers.

Adduct ions are determined by their mass difference to protonated/deprotonated molecules. Background noise (or white noise) in a mass spectrum may disturb the finding of real adduct ions. It is better to remove these noise before adduct ion determination. Most compounds from nature are made up of C, H, O and N and their protonated/deprotonated molecules (or the related ions) have similar isotope patterns. Signals from background noise have a lower chance to mimic both the isotopologue ions and their intensities. Based on this characteristic, we could discriminate real signals from noise.

Classifier for traditional Chinese medicine (CTCM) was developed according to the above mentioned theory. It consisted of three key steps: (1) raw data conversion, (2) signal denoising using isotope patterns and (3) adduct-ion pattern extraction. The flow chart of CTCM to analyze a LC/MS dataset is shown in Fig. 1 and the following is a brief description of the key steps.

2.1. Raw data conversion

CTCM was developed in a Matlab environment (Version 7.11.0, MathWorks, USA), with currently supported MS data formats of netCDF and mzXML [17]. Various proprietary data formats from different manufactures of mass spectrometers can be converted to common data formats using commercial or open-source tools [18]. Mass spectrometric data must be acquired or converted to centroid mode before they can be analyzed by CTCM. In this step, different formats of raw data are converted to the same data structure, which could be used by other modules of CTCM. If more than one level of MS data exists, only the first level (full scan MS) is extracted. In this work, the raw data files acquired from the instruments of Thermo Fisher Scientific were converted to mzXML files using ReAdW (Version 4.3.1).

2.2. Signal denoising using isotope patterns

The elements of most natural products are C, H, O, and N, and the intensities of their isotopologue ions are always in descending order (M > M+1 > M+2). This characteristic was used to separate the ions of natural products from those of noise. Through the Pearson correlation (or Dice) coefficients of the isotope clusters with a representative isotope pattern (i.e., M:(M+1):(M+2) = 10:3:1), true signals with a high correlation coefficient could be sorted out from a mass spectrum.

For each mass spectrum (Spec(i), $i = 1, 2, \ldots, n$), find its true signals (tSpec) and noise signals (nSpec) using the following algorithm:

(1) Find the highest ion (principal ion) Spec(I) in Spec if there are ions in Spec.
(2) Find Spec(I)'s isotopic ions within a m/z tolerance tol.
(3) If no isotopic ions are found, this ion is moved to nSpec and goes to (1).
(4) If no isotope pattern (isoPattern) are provided by users, this ion Spec(I) and its isotopic ions (M+1, M+2, ...) are considered to be real signals and transfer them from Spec to tSpec and go to (1).
(5) Calculate this isotope cluster's correlation coefficients (r) with isoPattern. If r is above the user given limit, the ion Spec(I) and its isotopes are moved to tSpec; otherwise, move them to nSpec.
(6) Go to step (1).

Hence, each mass spectrum is denoised and the original whole dataset (WD) could be split into denoised and residual WDs. If obvious peaks are found in the total ion chromatogram (TIC) of the residual WD, other types of elements, such as Cl or Br, may exist. Then, other isotope patterns (10:1:3 for compounds with one Cl atom and 10:2:10 for compounds with one Br atom) should be applied. This process is repeated until only random noise can be found in the residual WD.

2.3. Adduct-ion pattern extraction

Similar compounds (or the same type of compounds) would have the same adduct-ion pattern in ESI-MS. Using characteristic adduct ions and their relative intensities to those of $[M+H]^+$ or $[M-H]^-$, a series of sub-datasets could be obtained from a complicated LC/MS dataset. The mass spectra in a sub-dataset are composed of only one type of adduct ions and their $[M+H]^+$ or $[M-H]^-$. This would be very helpful in the annotation and

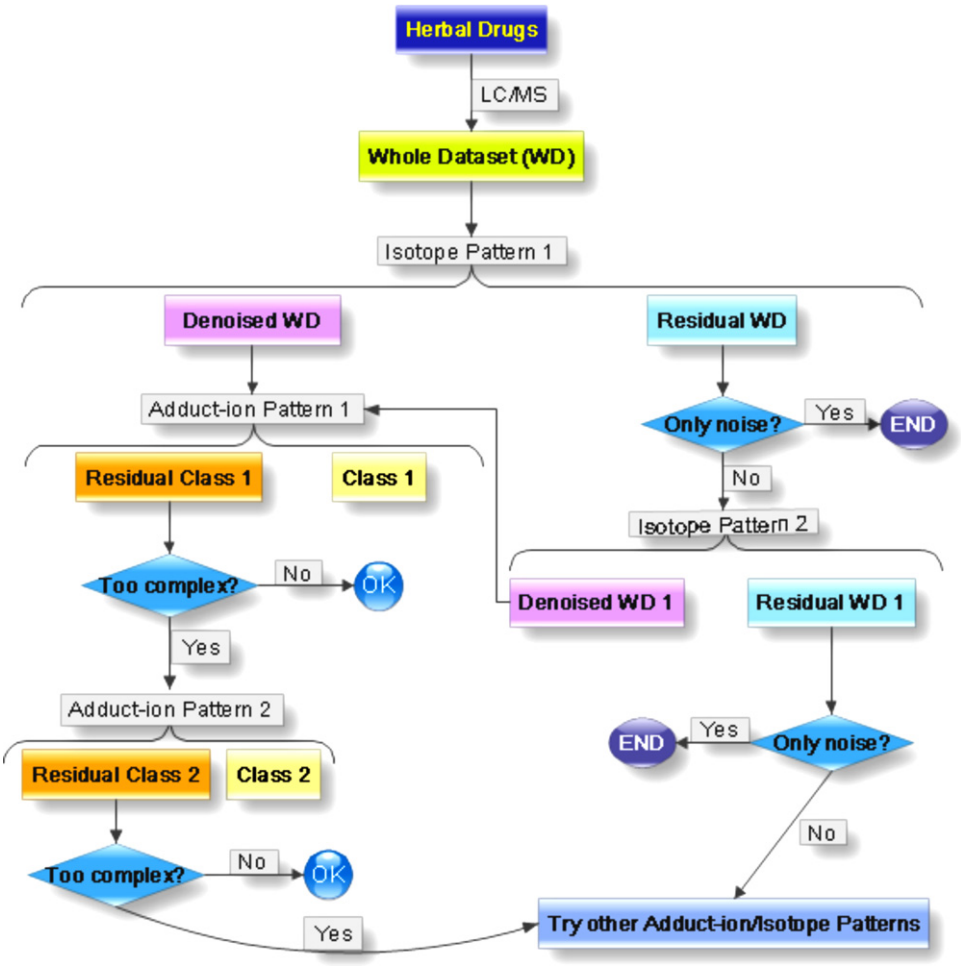


Fig. 1. Flow chart of classifier for traditional Chinese medicine (CTCM) to process a LC/MS dataset.

identification process. This task is achieved by the following algorithm for each mass spectrum:

(1) Find the isotope clusters in each mass spectrum using the algorithm described in Section 2.2. The principal ions (M) are stored in a matrix SingleIon and the corresponding isotope cluster (M, M+1, M+2, . . .) is stored in a cell frame Cluster.
(2) Find the highest ion SingleIon(I) in SingleIon if there are ions in SingleIon.
(3) According to a given adduct-ion pattern (ionPattern, $Y=f(X)$), calculate the target ions' m/z values (a and b) when SingleIon(I) is supposed to be $[M+H]^+$ or $[M-H]^-$ or the adduct ion.
(4) Find SingleIon(X) in SingleIon, whose m/z value is equal to a or b within a m/z tolerance tol.
(5) If no intensity ratio between the quasi-molecular ion and the adduct-ion is restricted or the ratio is within the pre-defined range (I), delete SingleIon(I) and SingleIon(X) from SingleIon and move Cluster(I) and Cluster(X) to tSpec. Go to (2).
(6) Remove SingleIon(I) from SingleIon and go to (2).

When all mass spectra were disposed, a sub-dataset for a given adduct-ion pattern is obtained by combing all their tSpecs. An interactive interface was also developed to show the TIC and mass spectra of original raw dataset or the extracted sub-datasets. The Matlab codes of CTCM are freely available upon request from the corresponding author.

3. Datasets

3.1. Erigeron injection

Erigeron injection (EI) was a Chinese herbal preparation from *Erigeron breviscapus*. The chemical components of *E. breviscapus* and EI have been confirmed by reference compounds or deduced using LC/MS technologies in our previous work [19,20]. The major compounds in *E. breviscapus* were flavonoids, and caffeoylquinic acids were more abundant due to their higher solubility in aqueous solution in EI. The LC/MS dataset of EI used in current work was obtained using the same analytical condition for *E. breviscapus* except for a post-column modification. A solution of 5 mmol/L tetrabutyl ammonium hydroxide (TBA-OH) was introduced to the column eluent through a T connector by a syringe pump. This was aimed to improve the compounds' response in (−)ESI/MS. After post-column modification, high abundant adduct-ions with TBA^+ ($[M-2H+TBA]^-$) were observed. The compounds' identification in EI could be found in Ref. [19].

3.2. Sarcandra glabra

The LC/MS dataset of *Sarcandra glabra* (SG) was from our recent published work [21], in which the major 26 peaks were identified by comparing with reference compounds. It was acquired on an LTQ Orbitrap XL hybrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA), connected to a Accela™ ultra-high-pressure

liquid chromatography (U-HPLC) system (Thermo Fisher Scientific, San Jose, CA, USA) via an ESI interface. The Key MS conditions were as follows: negative ion mode, mass range of 100–1000 with resolution set at 30,000 using the normal scan rate. The data-dependent MS/MS events were always performed on the most intense ions detected in full scan MS. ESI parameters were as follows: source voltage: 3.0 kV; sheath gas (nitrogen): 50 L/min; auxiliary gas flow: 10 L/min; capillary voltage: −35.0 V; capillary temperature: 300 °C; tube lens: −110.0 V.

3.3. WFC tablet

HPLC-grade acetonitrile and methanol were purchased from Merck Company (Merck, Darmstadt, Germany). HPLC-grade formic acid was supplied by Tedia (Fairfield, OH, USA). Ultra-pure water from a Milli-Q system (Millipore SAS, Molsheim, France) was used throughout the study. Other reagents and chemicals were of analytical grade. Wei-Fu-Chun tablets (batch number: 080913) and reference herbs, specifically, Radix Ginseng Rubra (Hongshen, HS), Fructus aurantii (Zhiqiao, ZQ), and Isodon amethystoides (Xiangchacai, XCC), were provided by Huqingyutang Pharmaceutical Co., Ltd. (Hangzhou, China). Reference standards used for identification, including hesperidin, naringin, neohesperidin, poncirin, and danshensu, were purchased from the National Institute for the Control of Pharmaceutical and Biological Products (Beijing, China). Ginsenosides Rb_1, Rg_1, Re, Rf, Rb_2, Rb_3, 20(R)-Rg_3, and 20(S)-Rg_3, were provided by Prof. Chen YP, College of Chemistry, Jinlin University, Changchun, China. All reference standards were dissolved in methanol to obtain a final concentration of about 10 μg/mL. The film coats of a WFC tablet were removed by scraping, after which 0.5 g of the WFC tablet powder was extracted with 10 mL methanol using an ultrasonic bath. Herbal extracts of HS were prepared in the same way as the WFC tablet. The extractions of ZQ and XCC were performed by decocting 5 g of the pulverized herbs in 100 mL water for 1–3 h. The extraction was repeated and all extracts were combined and concentrated to 50 mL. All solutions were centrifuged at 12,000 rpm for 15 min, and the supernatant filtered through a 0.45-μm nylon filter membrane before analysis.

Chromatographic experiments were carried out on an Agilent 1100 system (Agilent, Waldbronn, Germany) consisting of a binary pump, an auto-sampler, a column oven, and a diode array detector. Samples were separated on an Agilent Zorbax SB-C_{18} column (4.6 mm × 250 mm, 5 μm, Agilent, Wilmington, DE, USA) maintained at 30 °C with a guard column. The mobile phase consisted of 0.05% aqueous formic acid (v/v) (A) and acetonitrile (B), with a constant flow rate of 0.5 mL/min. The optimized elution gradient was programmed as follows: 5% B (0 min), 20% B (15 min), 20% B (30 min), 30% B (55 min), 50% B (75 min), and 95% B (90 min). UV spectra were recorded from 190 to 400 nm. Sample aliquots of 10 μL were injected for HPLC–MS^n and HPLC–Q/TOF analyses.

For online multi-stage mass spectrometric analysis, an LCQ DecaXP^{plus} mass spectrometer (Thermo Finnigan, San Jose, CA, USA) was connected to the Agilent 1100 HPLC instrument via an ESI interface in the negative ion mode. The acquisition parameters were as follows: ion spray voltage, −4.5 kV; capillary voltage, −19 kV; capillary temperature, 350 °C; collision gas, ultrahigh-purity helium; sheath gas, N_2 30 arbitrary units; auxiliary gas, N_2 10 arbitrary units; scan range, m/z 100–1500. Data-dependent or selected reaction monitoring scans were performed to obtain the MS^n spectra of all the components in herbal samples.

For MS^n analysis of standards, a syringe pump was used for direct loop injections. The flow rate was set to 10 μL/min, the capillary temperature was set to 275 °C, and the sheath gas was set to five arbitrary units. No auxiliary gas was needed.

High-resolution mass spectrometric data were acquired on a hybrid Q/TOF mass spectrometer (QSTAR XL, Applied Biosystems, Concord, Ontario, Canada) equipped with an Turbo Ion Spray source. The instrument was operated in negative ion mode and full-scan spectra were recorded in the range of m/z 100–1500. Other operation parameters were set as follows: nebulizing gas: N_2 50 psi, auxiliary gas: N_2 35 psi, curtain gas: 20 psi, turbo spray voltage: 3700 V, and source temperature: 350 °C.

3.4. Validation of CTCM using standard compounds

Compounds used for validation of CTCM were as follows: protocatechualdehyde (PCA), danshensu (DSS), rosmarinic acid (RA), caffeic acid (CA), salvianolic acid A (SA), salvianolic acid B (SB), hesperidin (HD), naringin (NG), methyl hesperidin (MH), neohesperidin (NH), quercetin (QT), ginsenosides R_1, Re, Rd, Rh_1, Rg_1, Rg_2, 20(R)-Rg_3, Rb_1, Rb_2, F_1, F_2, F_3 and Gypenoside XVII (G_{17}). PCA, RA, CA, SA, SB, MH and QT were purchased from Sigma–Aldrich (St. Louis, USA) and ginsenoside R_1, Rd, Rh_1, Rg_2, F_1, F_2, F_3 and G_{17} were provided by Shanghai Ronghe Medical Technology Co. (Shanghai, China). These standards were dissolved in methanol and mixed according to their categories to obtain a final concentration from 99.88 to 858.33 μg/mL (Table S1). Aliquots of 2 μL of these standard mixtures were injected onto the same LC–MS^n system for WFC tablet except for the gradient program. The elution program was as follows: 18% B (0 min), 21% B (15 min), 25% B (20 min), 29% B (35 min), 35% B (45 min), 90% B (52 min) and 95% B (55 min), which were delivered with a flow rate of 0.8 mL/min.

4. Results and discussion

4.1. Validation

To check the performance of CTCM, three categories of compounds, including 6 phenolic acids (PCA, DSS, RA, CA, SA and SB), 5 flavonoids (HD, NG, MH, NH and QT) and 13 saponins (R_1, Re, Rd, Rh_1, Rg_1, Rg_2, Rg_3, Rb_1, Rb_2, F_1, F_2, F_3 and G_{17}), were selected and analyzed using the LC–MS system for WFC tablet. As shown in Fig. 2, both $[M-H]^-$ and $[M+HCOO]^-$ were observed in their full mass spectra, but the intensity ratios (I) between $[M-H]^-$ and $[M+HCOO]^-$ varied greatly. Otherwise, we found $[2M-H]^-$ ions in mass spectra of phenolic acids and flavonoids. Thus, the two adduct-ion patterns (Neg_HCOOH and Neg_2M) were selected to verify the performance of CTCM. Total ion chromatograms (TICs) of sub-datasets extracted using Neg_HCOOH with and without the restriction of intensity ratios (I) were shown in Fig. 3. The green traces are TICs of sub-datasets without restriction of I, corresponding to the total intensity of $[X]^-$ and $[X+HCOOH]^-$ ions. The identity of X could be "M−H" or "2M−H" as exemplified by the mass spectrum of NG (Fig. 2A). If other type of ions existed, the peaks in TIC of a sub-dataset would be lowered (green peaks in Fig. 3A and B). The abundances of $[M-H]^-$ and $[M+HCOO]^-$ of RA were much lower than those of $[2M-H]^-$ ions (Fig. 2B), resulting in a small peak in Fig. 3B for RA. The peak of RA would be more significant if extracted using Neg_2M (Fig. S1). When restricted the value of I during extraction process, only specific compounds would be retained in the extracted sub-datasets. The red and black traces in Fig. 3 were from sub-datasets with different ranges of I ($0.1 < I < 1$ and $1 < I < 100$, respectively). It was found that NG, HD, NH, PCA and RA were in black traces and MH, CA and most of saponins (R1, Re, Rg1, Rb1, F3, Rg2, Rb2, Rd, F1, G17 and Rg3) were in red traces. If the intensity ratio between $[M-H]^-$ and $[M+HCOO]^-$ of a compound is out of the defined ranges ($0.1 < I < 1$ and $1 < I < 100$), the compound's peak would disappear and this is the cases for QT, SB, SA, Rh1 and F2, whose intensities of $[M-H]^-$ or $[M+HCOO]^-$ are extremely low. The above results show that CTCM can separate LC/MS datasets to different sub-datasets as expected.

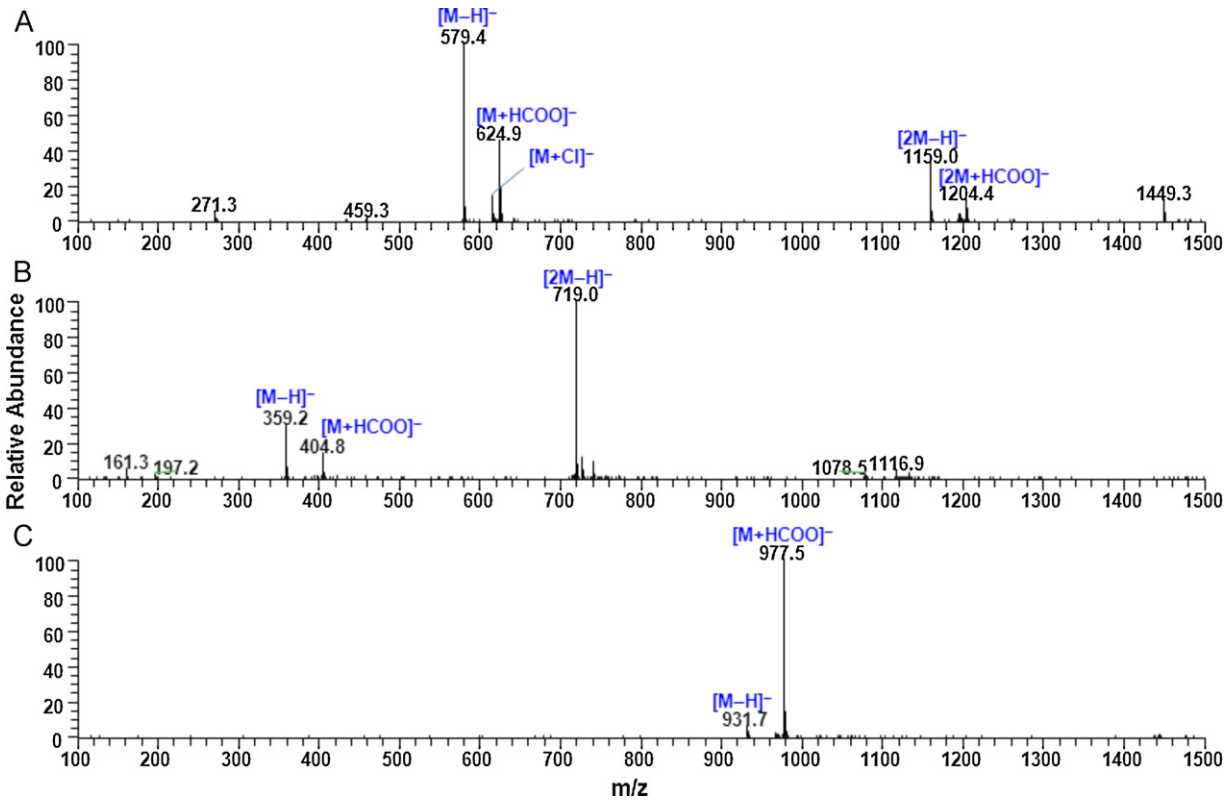


Fig. 2. Full mass spectra of NG (A), RA (B) and R_1 (C).

4.2. Parameters of CTCM

There are five parameters for CTCM: tol, *r*, *l*, isoPattern and ionPattern. In order to guide the parameter determination, two LC/MS datasets (EI and SG), in which their main components have been identified, were selected to explore the relationship between parameters and the intrinsic characteristics of LC/MS datasets. The low resolution EI dataset was obtained on an ion trap mass spectrometer with unit resolution and the high resolution SG dataset was acquired using a LTQ Orbitrap mass spectrometer.

4.2.1. tol: m/z tolerance

To investigate the *m*/*z* bias in low and high resolution datasets, we collected all the isotope clusters in EI and SG and for each cluster calculated the deviation of M+1 with its theoretical value. As shown in Fig. 4A and B, the deviation of M+1 in EI and SG is correlated with not only the resolution, but also the intensities of ions. The distributions of *m*/*z* deviations are approximately in accord with a normal distribution, and the fitted values of mean (mu) and standard deviation (δ) are attached to the figures. The value of mu and δ is equivalent to the accuracy and mass stability of the mass spectrometer, respectively. The smaller these two values are, the higher the accuracy and mass stability of the mass spectrometer is. Except for the influence from instruments, it is noticed that the deviation of M+1 from a theoretical value would expand along with the drop of ion intensities for both EI and SG datasets. This indicated that a too small tol value would tend to lose those low abundant ions; on the other hand, if the tol is too large, the denoising effect would be ruined. Thus, it is suggested that a value of 3δ to 5δ would be a good candidate for the parameter tol.

The deviation of adduct-ions in EI and SG (Fig. 4C and D) has a similar trend with that of M+1 isotope. Although the values of mu and δ have several times of fluctuation, the above suggested tol value could also be applied to the extraction of adduct-ions from both low and high resolution datasets.

4.2.2. isoPattern and r: isotope distribution and denoising

To represent the similarity between an isotope cluster ([M M+1 M+2]) and isoPattern, the default algorithm is Pearson correlation (PC) and an alternative algorithm is the Dice coefficient (DC). The isotope ratio (1:0.4:0.11) of the highest ion ($[M-H]^-$ of rosmarinic acid) in SG dataset is assigned as the isoPattern. Let's suppose that the intensity of M in an isotope cluster is equal to 1 and change the intensities of M+1 and M+2 from 0.01 to 1 with a step of 0.01. Then, the correlation coefficients for all possible combinations of M+1 and M+2 were computed using the above two algorithms. The results are presented in Fig. 5. The lines or circles in Fig. 5 are contour lines and their colors are coded by the values of correlation coefficients. At a specific *r* value, the PC algorithm defines a region by two lines radiating from the right-top corner (Fig. 5A) and the regions defined by DC are circles whose centers are the position of isoPattern (Fig. 5B). Fig. 6 is the real isotope distribution of EI (tol = 0.2) and SG (tol = 0.02). Each point (or circle) in Fig. 6 is corresponding to one isotope cluster. The red points are those clusters that are above the 5% of the maximum intensity. When *r* is equal 0.9, the regions defined by PC (red dotted line) and DC (green solid line) are also plotted in Fig. 6. For natural products (make up by C, H, O, N), the intensities of M+1 should always be bigger than that of M+2 and not more than 0.5. Thus, the isotope clusters outside of the low-left triangle are probably to be noise, especially the blue circles. As shown in Fig. 6, more noise clusters were observed in EI due to its bigger *m*/*z* tolerance. It suggested that such a denoising method is more suitable for high resolution datasets. The denoised TICs of SG and EI were given in Fig. S2.

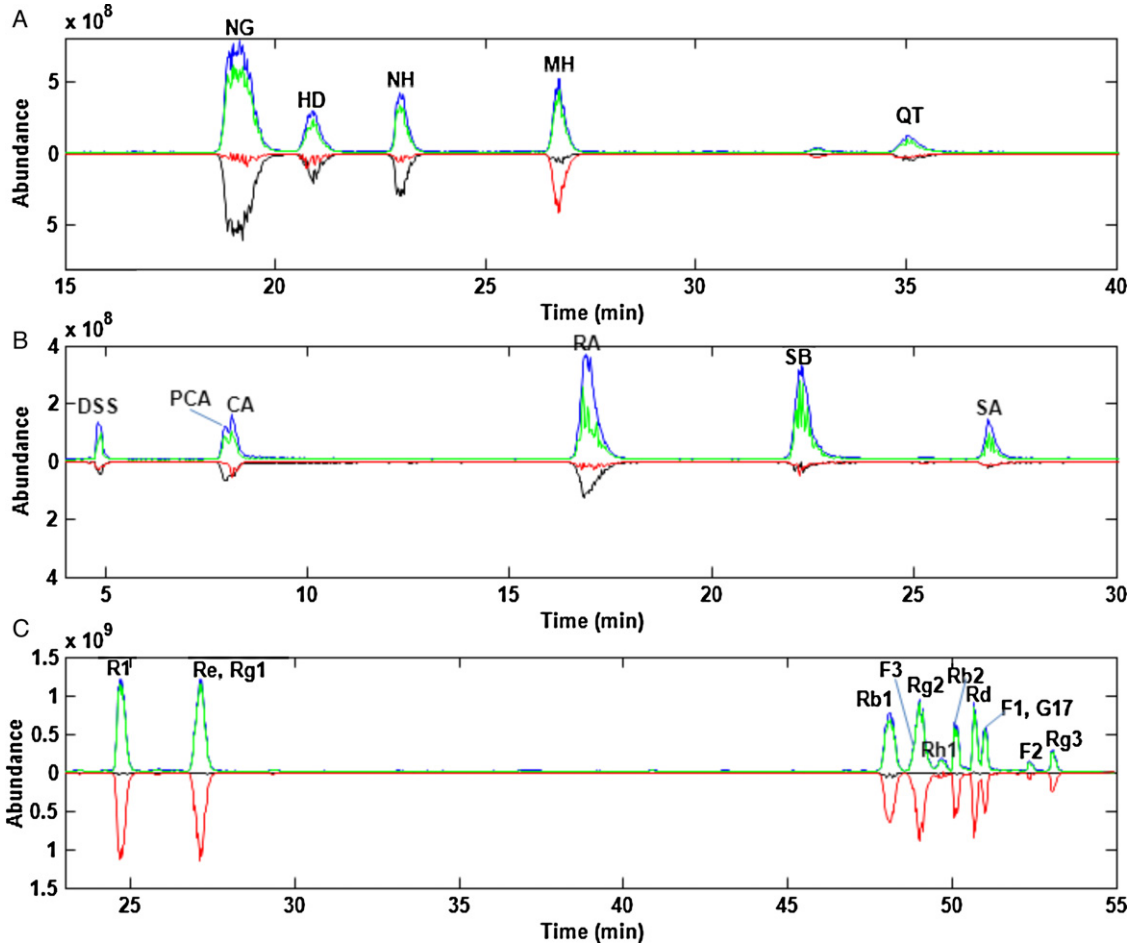


Fig. 3. Total ion chromatograms (TICs, blue traces) of flavonoids (A), phenolic acids (B) and saponins (C) and their corresponding sub-datasets extracted by CTCM: TICs of sub-datasets by Neg_HCOOH without restriction of l (green traces); TICs of sub-datasets by Neg_HCOOH with $0.01 < l < 1$ (red traces); TICs of sub-datasets by Neg_HCOOH with $1 < l < 100$ (black traces). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

4.2.3. ionPattern and l: adduct-ion extraction

The common adduct-ions have been included in CTCM, such as Neg_HCOOH, Neg_HCOONa and Neg_2M. New adduct-ions could be easily added by modifying an existing adduct-ion function. In most cases, the parameter l is not needed to provide, only when the studied LC/MS dataset is too complex and want to get more sub-datasets.

The $[2M-H]^-$ of SG and $[M-2H+TBA]^-$ of EI were extracted using the corresponding adduct ion patterns without parameter l (Fig. 7). The peak numbers in Fig. 7 were the same as those in previous references [19,21]. Peak a in Fig. 5A has not been identified in our previous work. The mass spectrum of peak 41 (RA) in SG and that of peak 38 (3-Benz-9-p-Co-OA) in EI (blue stems) were given in the right of Fig. 7. As show in Fig. 7B′, only the $[M-H]^-$ and $[M-2H+TBA]^-$ ions were retained and other adduct ions (m/z 553 $[M-H+HCOONa]^-$ and 772 $[M-2H+TBA+HCOOH]^-$) or background ions (m/z 332 and 378) were removed from the original mass spectrum. All peaks' mass spectra were also checked using CTCM's presentation module. It suggested that CTCM could thoroughly extract a certain adduct ion pattern from different LC/MS dataset, either low or high resolution. The TICs and mass spectra of the extract sub-dataset were not as complex as those of original datasets. It would be a tool for rapid and effective analysis of herbal medicines. In the following section, CTCM was used to assist the identification of components from WFC tablet.

4.3. Classification of the constituents in WFC tablet by CTCM

After LC/MS analysis with a 90-min gradient program, the TIC of a WFC tablet was obtained, as shown in Fig. 8A. The main peaks in the tablet were well resolved, but the peaks from 67 to 82 min overlapped, and minor components may be buried in this range. Hence, CTCM was applied to reduce the complexity of the TIC.

First, an isotope pattern (10:3:1) were convoluted with each mass spectrum to remove random noises in the LC/MS dataset of the WFC tablet (tol = 0.3). As shown in Fig. 8B, the baseline of the denoised TIC was almost zero and all peaks were retained. Since no obvious peaks could be observed in TIC reconstructed by the removed signals (Fig. S3), no further isotope patterns were applied. Second, the mass spectra of the peaks in Fig. 8B were manually checked for possible adduct-ion patterns. It showed that most of them had a $[M+HCOO]^-$ ion, but the ratios between $[M-H]^-$ and $[M+HCOO]^-$ varied greatly and could be divided into two groups: one is bigger than one; another is less than one. Thus, two sub-datasets (Classes 1 and 2) were obtained using the formic acid adduct-ion pattern and two l ranges (1–100 and 0.01–1), respectively. The TICs of Class 1 and Class 2 were given in Fig. 8C and D, respectively. After subtracting Classes 1 and 2 from the denoised WD, the remaining signals (residual Class) were obtained, as shown in Fig. 8E. The peaks in Fig. 8E were compounds lacking a formic acid adduct-ion (**3**, **10**, **14**, **15**, **20**, **21**, **26**, **31**, **50** and **51**) or other types of adduct-ions and fragments of the compounds in Class 1 and Class

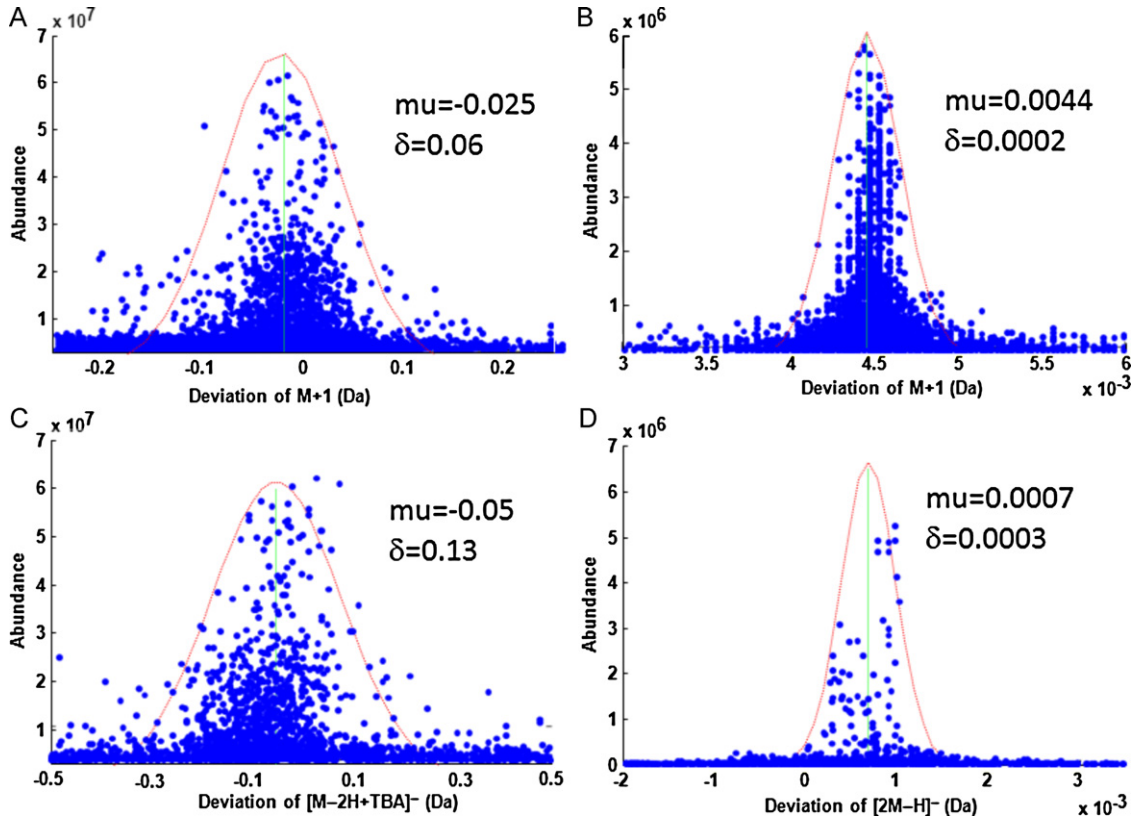


Fig. 4. Distribution of deviations between experimental and theoretical results for M+1 isotopic ions (A and B) and adduct ions (C and D) in EI (A and C) and SG (B and D). The red lines are the fitted Gaussian curves with estimated values of the mean (mu) and standard deviation (δ). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

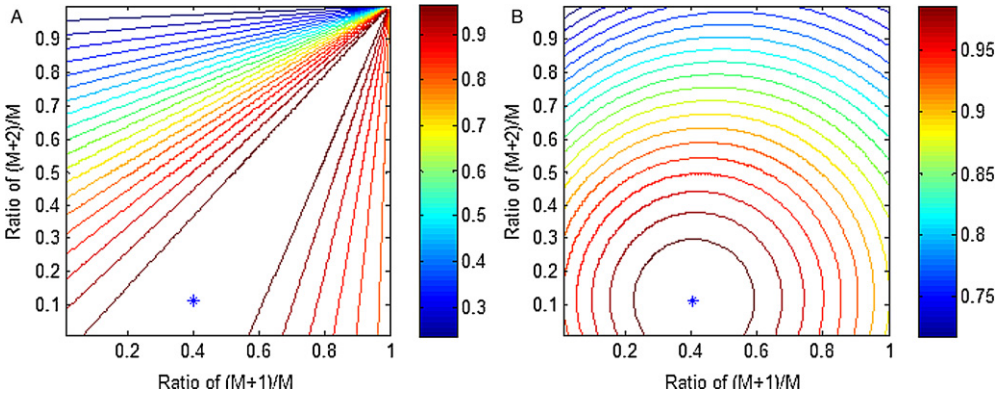


Fig. 5. Contour maps of similarity coefficients calculated using Pearson (A) and Dice correlation (B).

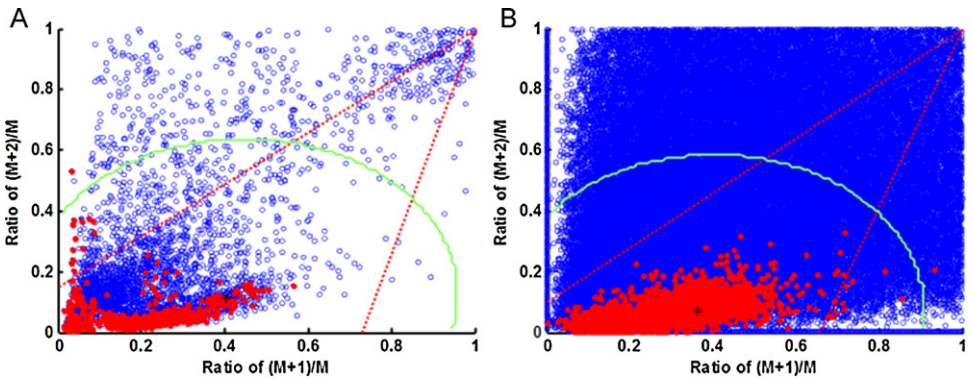


Fig. 6. Distribution of isotope clusters of SG (A) and EI (B).

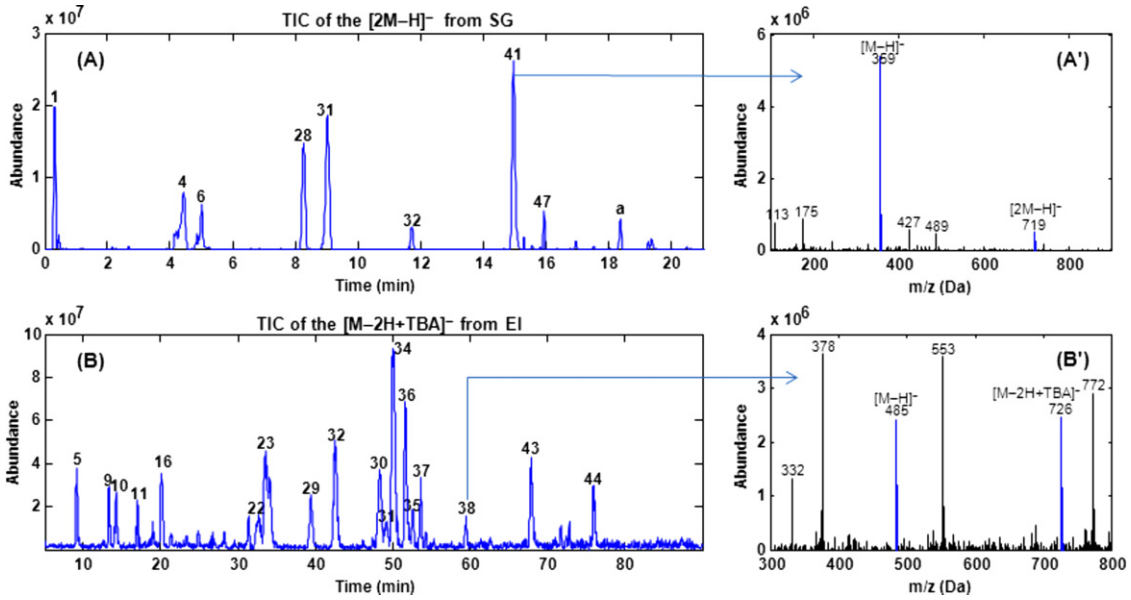


Fig. 7. Total ion chromatograms of the [2M−H]⁻ pattern from SG (A) and the [M−2H+TBA]⁻ pattern from EI (B). The peak numbers in (A) and (B) correspond to the numbers for SG and EI in Refs. [21] and [19], respectively. Typical mass spectra for one peak of SG (A′) and EI (B′) were given at the right of this figure. The extracted adduct ions were indicated in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

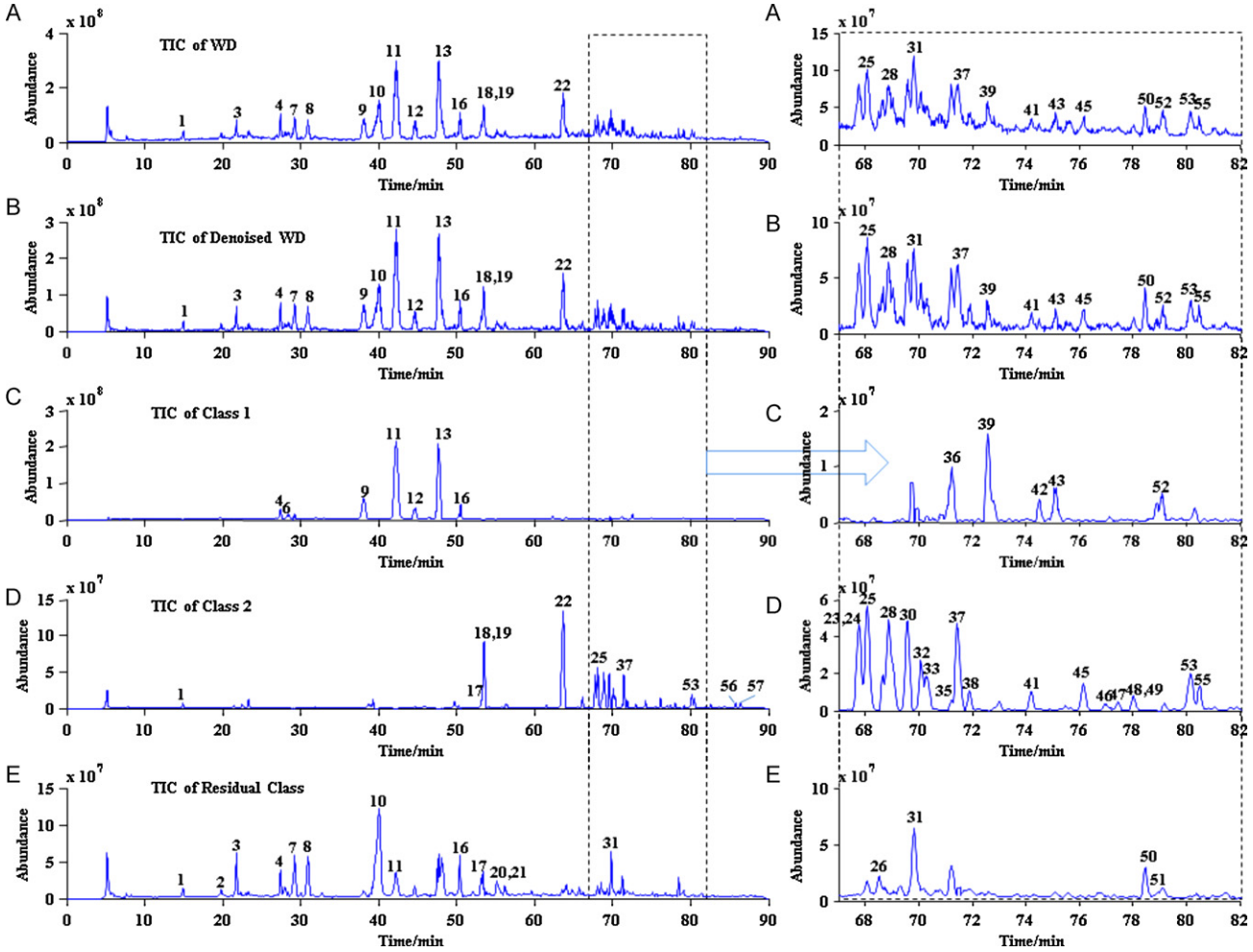


Fig. 8. Total ion chromatograms (TIC) of WFC tablet (A) and those of denoised whole dataset (B), Class 1 (C), Class 2 (D), and residual Class (E).

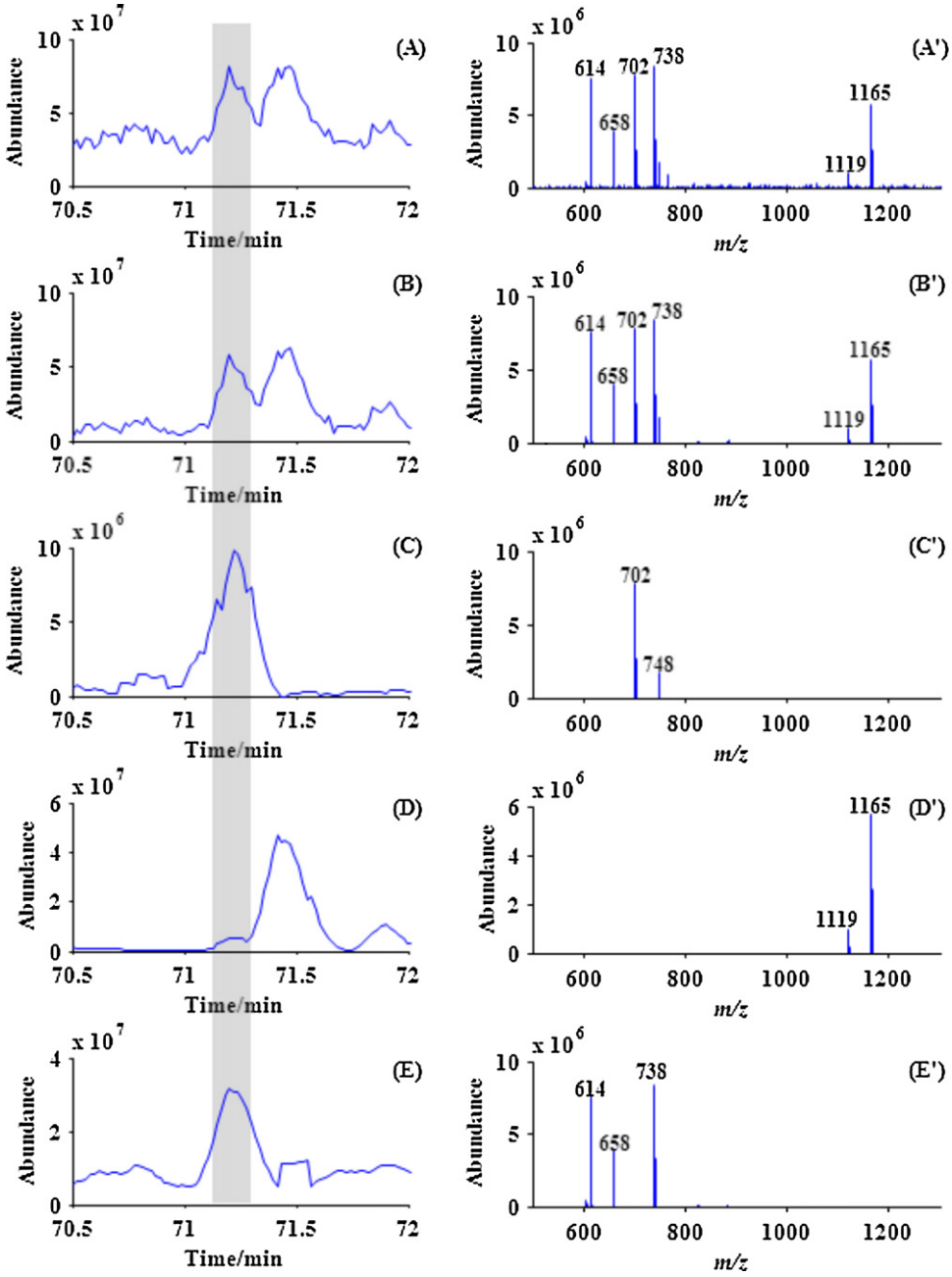


Fig. 9. Zoomed range of Fig. 8 at around 71.2 min (left: A–E) and their corresponding average mass spectra in the gray range (right: A′–E′).

2. For example, the $[M+Cl]^-$ and $[2M-H]^-$ ions of peak 11 in Class 1 formed another peak in residual Class at the same retention time. The figures (Fig. 8C–E) show that the crowded TIC of the WFC tablet was represented by three relatively simple TICs, particularly in the range of 67–82 min (enlarged figure is given to the left of Fig. 8).

To show the performance of CTCM, an example is given in Fig. 9. Only one peak at 71.2 min was observed in the TIC of the WFC tablet (Fig. 9A). Its mass spectrum was rather complex (Fig. 9A′); thus, it was difficult to determine the deprotonated molecules for this peak. After classification, at least three compounds were found in this peak (Figs. 9C–E). Based on their adduct ions ($[M+HCOO]^-$), the deprotonated molecules could be easily determined as m/z 702 and 1119 for the peaks in Fig. 9C and D. The three ions in Fig. 9E may be the fragments of a single compound or $[M-H]^-$ ions of different compounds and multi-stage mass spectrometric analysis should be performed to determine the relationship of the three ions.

When a compound has other adductions besides $[M+HCOO]^-$ (i.e., $[M+Cl]^-$, $[2M-H]^-$, . . .), peaks may be formed in residual Class at the same retention time. Such peaks could be determined by comparatively checking their full mass spectra. At last, a total of 57 peaks were determined for the WFC tablet, including 16 from Class 1, 30 from Class 2 and 11 from residual Class (Table 1).

4.4. Chemical constituents of WFC tablet

The TICs of Classes 1 and 2 were similar in shape to those of ZQ and HS (Fig. S4), respectively. This is due to the major peaks in ZQ and HS are the same types of compounds, respectively. To characterize the compounds in these classes further, they were analyzed using multi-stage and high resolution mass spectrometry. By integrating the reported information of each herb with the analytical data, the elemental compositions and candidate compounds could

Table 1
Peaks observed with corresponding retention times, assigned identity and quasi-molecular ions in negative ESI-MS of the extracts of WFC tablet.

No.	t_R (min)	$[M-H]^-$	Observed mass (Da)	Error (ppm)	Proposed formula	Identification	Class	Source
1	14.94	197	197.0483	14.21	$C_9H_{10}O_5$	Danshensu	2, 3	XCC
2	19.81	357	357.0605	−3.08	$C_{18}H_{14}O_8$	Epiphyllic acid	1, 3	XCC
3	21.77	593	593.1533	2.70	$C_{27}H_{30}O_{15}$	Apigenin 6,8-di-C-glucoside	3	ZQ
4	27.38	427	427.1598	−2.58	$C_{20}H_{28}O_{10}$	Hyuganoside II/hyuganoside V	1, 3	ZQ
5	27.94	537	537.1012	−4.84	$C_{27}H_{22}O_{12}$	Rabdosic acid	3	XCC
6	28.45	595	595.1667	−0.17	$C_{27}H_{32}O_{15}$	Eriocitrin	1, 3	ZQ
7	29.22	649	649.2494	−1.23	$C_{32}H_{42}O_{14}$	Limonin-17-β-D-glucopyranoside	1, 3	ZQ
8	30.93	595	595.1656	−2.02	$C_{27}H_{32}O_{15}$	Neoeriocitrin	1, 3	ZQ
9	38.06	579	579.1710	−1.55	$C_{27}H_{32}O_{14}$	Narirutin	1	ZQ
10	40.04	717	717.1400	−8.51	$C_{36}H_{30}O_{16}$	Rabdosiin	3	XCC
11	42.25	579	579.1687	−5.53	$C_{27}H_{33}O_{14}$	Naringin	1, 3	ZQ
12	44.64	609	609.1821	−0.66	$C_{28}H_{34}O_{15}$	Hesperidin	1	ZQ
13	47.61	609	609.1802	−3.78	$C_{28}H_{34}O_{15}$	Neohesperidin	1	ZQ
14	48.03	359	359.0761	−3.06	$C_{18}H_{16}O_8$	Rosmarinic acid	3	XCC
15	48.22	651	–		–	Deacetyl nomilin glucoside	3	ZQ
16	50.44	711	711.2839	−4.22	$C_{34}H_{48}O_{16}$	Nomilinic acid glucoside	1	ZQ
17	53.17	543.3	543.2797	−2.58	$C_{27}H_{44}O_{11}$	Unknown	2	XCC
18	53.38	799.6	845.4865	−4.61	$C_{43}H_{74}O_{16}$	Ginsenoside Rg1	2	HS
19	53.49	945.7	991.5423	−6.05	$C_{49}H_{84}O_{20}$	Ginsenoside Re	2	HS
20	55.13	633	633.2511	−6.47	$C_{32}H_{42}O_{13}$	Obacunone glucoside	3	ZQ
21	55.36	315	315.0520	3.17	$C_{16}H_{12}O_7$	Pedalitin/isorhamnetin	3	XCC
22	63.66	593	593.1887	1.85	$C_{28}H_{34}O_{14}$	Poncirin	2	ZQ
23	67.66	1209	1255.6313	−1.19	$C_{59}H_{100}O_{28}$	Ginsenoside Ra1	2	HS
24	67.77	799.6	845.4859	−5.32	$C_{43}H_{74}O_{16}$	Ginsenoside Rf	2	HS
25	68.08	1107.8	1153.5953	−5.03	$C_{55}H_{94}O_{25}$	Ginsenoside Rb1	2	HS
26	68.49	1193.3	1193.5899	−5.19	$C_{57}H_{94}O_{26}$	Malonyl-ginsenoside Rb_1	3	HS
27	68.66	1209.7	1255.6332	0.32	$C_{59}H_{100}O_{28}$	Ginsenoside Ra2	2	HS
28	68.87	1077.7	1123.5897	−0.80	$C_{54}H_{92}O_{24}$	Ginsenoside Rb2	2	HS
29	69.05	769.6	815.4823	3.07	$C_{42}H_{72}O_{15}$	Notoginsenoside R_2	2	HS
30	69.57	1077.7	1123.585	−4.98	$C_{54}H_{92}O_{24}$	Ginsenoside Rb3	2	HS
31	69.82	955.8	955.4909	0.10	$C_{48}H_{76}O_{19}$	Ginsenoside Ro	3	HS
32	70.07	783.6	829.4952	−0.36	$C_{43}H_{74}O_{15}$	20(S)-ginsenoside Rg2	2	HS
33	70.27	1149.6	1195.6146	2.43	$C_{57}H_{96}O_{26}$	Acetyl ginsenoside Rb1	2	HS
34	70.36	783.6	829.4988	3.98	$C_{43}H_{74}O_{15}$	20(R)-ginsenoside Rg2	2	HS
35	71.2	1119	–	–	–	Ginsenoside Rs1	2	HS
36	71.2	702.3	–	–	–	Unknown	1	ZQ
37	71.45	945.7	991.5438	−4.54	$C_{49}H_{84}O_{20}$	20(S)-ginsenoside Rd	2	HS
38	71.89	1119.6	1165.6038	2.32	$C_{56}H_{94}O_{25}$	Ginsenoside Rs2	2	HS
39	72.57	643.4	–	–	–	Unknown	1	XCC
40	72.63	945	991.5522	3.93	$C_{49}H_{84}O_{20}$	20(R)-ginsenoside Rd	2	HS
41	74.19	987.6	1033.5653	6.19	$C_{51}H_{86}O_{21}$	20(S)-acetyl-ginsenoside Rd	2	HS
42	74.49	643.4	–	–	–	Unknown	1	XCC
43	75.11	477.4	–	–	–	Unknown	1	XCC
44	75.49	987.6	1033.5540	−4.74	$C_{51}H_{86}O_{21}$	20(R)-acetyl-ginsenoside Rd	2	HS
45	76.15	789.3	–	–	–	Unknown	2	?[a]
46	77.03	773.3	–	–	–	Unknown	2	?[a]
47	77.43	765.7	811.4844	−0.62	$C_{43}H_{72}O_{14}$	Ginsenoside Rg5	2	HS
48	78.03	773.4	–	–	–	Unknown	2	?[a]
49	78.03	765.7	811.4800	−6.04	$C_{43}H_{72}O_{14}$	Ginsenoside F4	2	HS
50	78.47	531.3	–	–	–	Unknown	3	?[a]
51	78.89	471.3	–	–	–	Unknown	3	?[a]
52	79.11	469	–	–	–	Limonin	1	ZQ
53	80.15	783.6	829.4973	2.17	$C_{43}H_{74}O_{15}$	20(S)-ginsenoside Rg3	2	HS
54	80.34	793.5	–	–	–	Unknown	1	?[a]
55	80.49	783.6	829.4918	−4.46	$C_{43}H_{74}O_{15}$	20(R)-ginsenoside Rg3	2	HS
56	85.77	765.7	811.4834	−1.85	$C_{43}H_{72}O_{14}$	Ginsenoside Rg6	2	HS
57	86.37	765.7	811.4829	−2.46	$C_{43}H_{72}O_{14}$	Ginsenoside Rk1	2	HS

[a] The origination of this compound could not be determined.

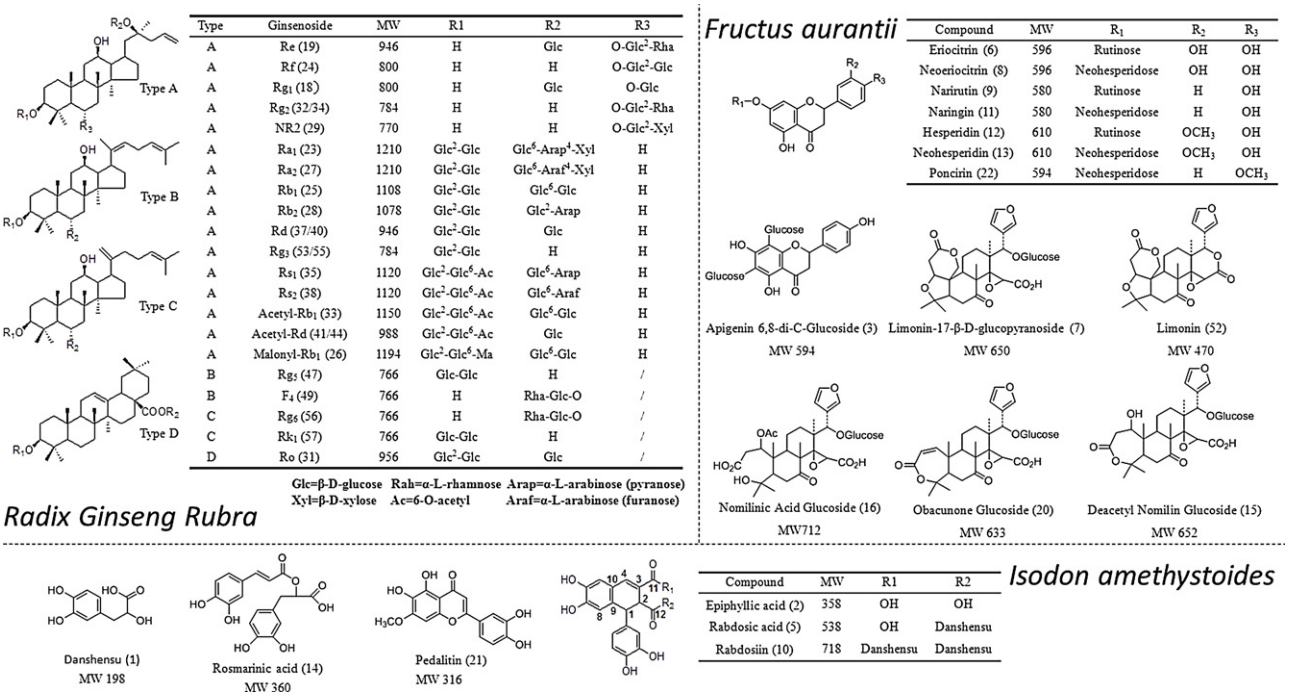
be obtained, and verified if standard compounds were available. This led to either the identification or tentative characterization of 46 compounds (Fig. 10 and Table 1). Detailed MS^n and UV data are supplied as supporting information (Table S2).

4.4.1. Identification of the compounds in Class 1

Among the 16 compounds in Class 1, 11 (**4**, **6**, **7**, **8**, **9**, **11**, **12**, **13**, **16**, **36**, and **52**) originated from ZQ, while 4 (**2**, **39**, **42** and **43**) originated from XCC. The source of **54** could not be determined and could have formed during the preparation of the WFC tablet.

The exact mass of the $[M-H]^-$ of compound **2** was found at *m*/*z* 357.0605, indicating a molecular formula (MF) $C_{18}H_{14}O_8$. In the MS^n of compound **2**, two neutral losses of 44 Da and a neutral loss of 110 Da suggested the existence of two carboxyl groups and one catechol. The possible structures with MF $C_{18}H_{14}O_8$ were retrieved using the client module of SciFinder Scholar, and 259 candidates were found. By removing those containing metal ions or mixtures and those without carboxyl groups and one catechol, only two structures, epiphyllic acid and blechnic acid, matched the fragmentation pathway of compound **2** (Fig. 11). As other compounds possessing an epiphyllic acid unit have been reported from XCC [22], we tentatively assigned compound **2** as epiphyllic acid.

The MS^n of compound **4** suggested that it was a glycoside and that it contained another carboxyl group in its structure. The MF of compound **4** was deduced to be $C_{20}H_{28}O_{10}$ and 103 possible structures were found using SciFinder. However, only two compounds



Type	Ginsenoside	MW	R1	R2	R3
A	Re (19)	946	H	Glc	O-Glc^2-Rha
A	Rf (24)	800	H	H	O-Glc^2-Glc
A	Rg$_1$ (18)	800	H	Glc	O-Glc
A	Rg$_2$ (32/34)	784	H	H	O-Glc^2-Rha
A	NR2 (29)	770	H	H	O-Glc^2-Xyl
A	Ra$_1$ (23)	1210	Glc^2-Glc	Glc^6-Arap^4-Xyl	H
A	Ra$_2$ (27)	1210	Glc^2-Glc	Glc^6-Araf^4-Xyl	H
A	Rb$_1$ (25)	1108	Glc^2-Glc	Glc^6-Glc	H
A	Rb$_2$ (28)	1078	Glc^2-Glc	Glc^2-Arap	H
A	Rd (37/40)	946	Glc^2-Glc	Glc	H
A	Rg$_3$ (53/55)	784	Glc^2-Glc	H	H
A	Rs$_1$ (35)	1120	Glc^2-Glc^6-Ac	Glc^6-Arap	H
A	Rs$_2$ (38)	1120	Glc^2-Glc^6-Ac	Glc^6-Araf	H
A	Acetyl-Rb$_1$ (33)	1150	Glc^2-Glc^6-Ac	Glc^6-Glc	H
A	Acetyl-Rd (41/44)	988	Glc^2-Glc^6-Ac	Glc	H
A	Malonyl-Rb$_1$ (26)	1194	Glc^2-Glc^6-Ma	Glc^6-Glc	H
B	Rg$_5$ (47)	766	Glc-Glc	H	/
B	F$_4$ (49)	766	H	Rha-Glc-O	/
C	Rg$_6$ (56)	766	H	Rha-Glc-O	/
C	Rk$_1$ (57)	766	Glc-Glc	H	/
D	Ro (31)	956	Glc^2-Glc	Glc	/

Compound	MW	R$_1$	R$_2$	R$_3$
Eriocitrin (6)	596	Rutinose	OH	OH
Neoeriocitrin (8)	596	Neohesperidose	OH	OH
Narirutin (9)	580	Rutinose	H	OH
Naringin (11)	580	Neohesperidose	H	OH
Hesperidin (12)	610	Rutinose	OCH_3	OH
Neohesperidin (13)	610	Neohesperidose	OCH_3	OH
Poncirin (22)	594	Neohesperidose	H	OCH_3

Compound	MW	R1	R2
Epiphyllic acid (2)	358	OH	OH
Rabdosic acid (5)	538	OH	Danshensu
Rabdosiin (10)	718	Danshensu	Danshensu

Fig. 10. Structures of identified compounds from WFC tablet.

had both a glucose unit and a CO_2H unit. The two candidates, hyuganoside II and hyuganoside V, are a pair of position isomers and could not be discriminated by their MS^n data alone (Fig. S5). Thus, compound **4** was deduced to be either hyuganoside II or hyuganoside V.

Compounds **6** and **8** exhibited identical $[M–H]^-$ ions at *m*/*z* 595, which was 16 Da higher than those of naringin and narirutin. HRMS data indicated a compound with MF $C_{27}H_{31}O_{15}$, suggesting the existence of another hydroxyl group. Under CID conditions, the $[M–H]^-$ ion of compound **6** readily eliminated a rutinose unit to produce a $[M–H–308]^-$ ion at *m*/*z* 287. Thus, compound **6** was plausibly identified as eriocitrin after comparison with the reference data [23,24]. Compound **8** was tentatively characterized as neoeriocitrin due to the occurrence of the fragment ion at *m*/*z* 459, whose MS/MS spectrum was similar to reported data [23,24]. Compound **9** showed a $[M–H]^-$ ion at *m*/*z* 579.1719 in ESI(−)-Q-TOF, corresponding to MF $C_{27}H_{31}O_{14}$. The fragment ion at *m*/*z* 271 (originating from the cleavage of an arutinose unit) was determined from the source collision-induced dissociation (CID) of the deprotonated molecules. The MS^2 and MS^3 data of compound **9** were identical to those of naringin. After comparison with the reported data [23], compound **9** was tentatively characterized as narirutin. In the same way, compounds **11**, **12**, and **13** were characterized as naringin, hesperidin, and neohesperidin, respectively.

Compounds **7**, **16**, and **52** were plausibly identified as limonin-17-β-D-glucopyranoside, nomilinicacid glucoside, and limonin, respectively, by comparing their MS^n data with those of standards or data from literature [25]. Compounds **36**, **39**, **42**, **43**, and **54** were not identified due to the absence of their MS^n and high resolution data.

4.4.2. Identification of the compounds in Class 2

Almost all compounds (**18**, **19**, **23**–**25**, **27**–**30**, **32**–**35**, **37**, **38**, **40**, **41**, **44**, **47**, **49**, **53**, and **55**–**57**) in Class 2 were ginsenosides from HS. Compounds **18**, **19**, **24**, **25**, **28**, **30**, **53**, and **55** were unambiguously attributed to ginsenosides Rg_1, Re, Rf, Rb_1,Rb_2, Rb_3,20(S)-Rg_3, and 20(R)-Rg_3, respectively, after comparisons of their retention time, UV, and mass spectra with those of authentic standards. Compounds **23**, **27**, **29**, **32**, **34**, **35**, **37**, **38**, **40**, **47**, **49**, **56**, and **57** were tentatively identified by careful study of their MS and MS^2 spectra and through comparison with literature data [26–32].

Under ESI-MS^n conditions, compound **33** yielded intense $[M–H]^-$ ions at *m*/*z* 1149 in negative mode. The fragment ion at *m*/*z* 1107 was indicative of one acetyl group. The MS^3 spectrum of *m*/*z* 1107 displayed a fragmentation pathway similar to that of the standard ginsenoside Rb1, suggesting that there is ginsenoside Rb1 in this fraction. Therefore, compound **33** was tentatively identified as Acetyl-Rb1.

Compounds **41** and **44** had *m*/*z* values identical to that of the quasi-molecular ion (987.55) in the MS spectra, indicating that they may be isomers. The MS^2 spectrum shows the presence of an acetyl group, and was similar to that of $[M–H]^-$ from CID of ginsenoside Rd. These data suggested that they were acetylated ginsenoside Rd. The two isomers were discriminated by their elution order on reversed-phase(RP)-HPLC columns [33]. Considering the long retention time of compound **44**, it was plausibly identified as 20(R)-acetyl-ginsenoside Rd, whereas compound **41** was considered as 20(S) acetyl-ginsenoside Rd.

By comparing them with authentic standards, compounds **22** from ZQ and **1** from XCC were identified as poncirin and danshensu, respectively. The molecular weights of compounds **45**, **46**, and **48** were in the range of those of ginsenosides, but they were not detected in the herbal extracts of HS, suggesting that they may be depredated products of ginsenosides. More work should be performed to elucidate their structures.

4.4.3. Identification of the compounds in residual Class

The compounds in residual Class were those from XCC (**5**, **10**, **14**, and **21**) and those of ZQ and HS, which lacked a $[M+HCOO]^-$ ion (**3**, **15**, and **20** from ZQ; **26** and **31** from HS). Triterpenes were reported to be major components in XCC; however, no triterpene was detected in the present paper. This may be due to the low response of triterpene under negative ionization mode.

Compound **3** was deduced to be apigenin 6,8-di-C-glucoside based on its MS^n data. This assumption was confirmed by high accuracy mass measurement, with an observed mass at *m*/*z* 593.1517

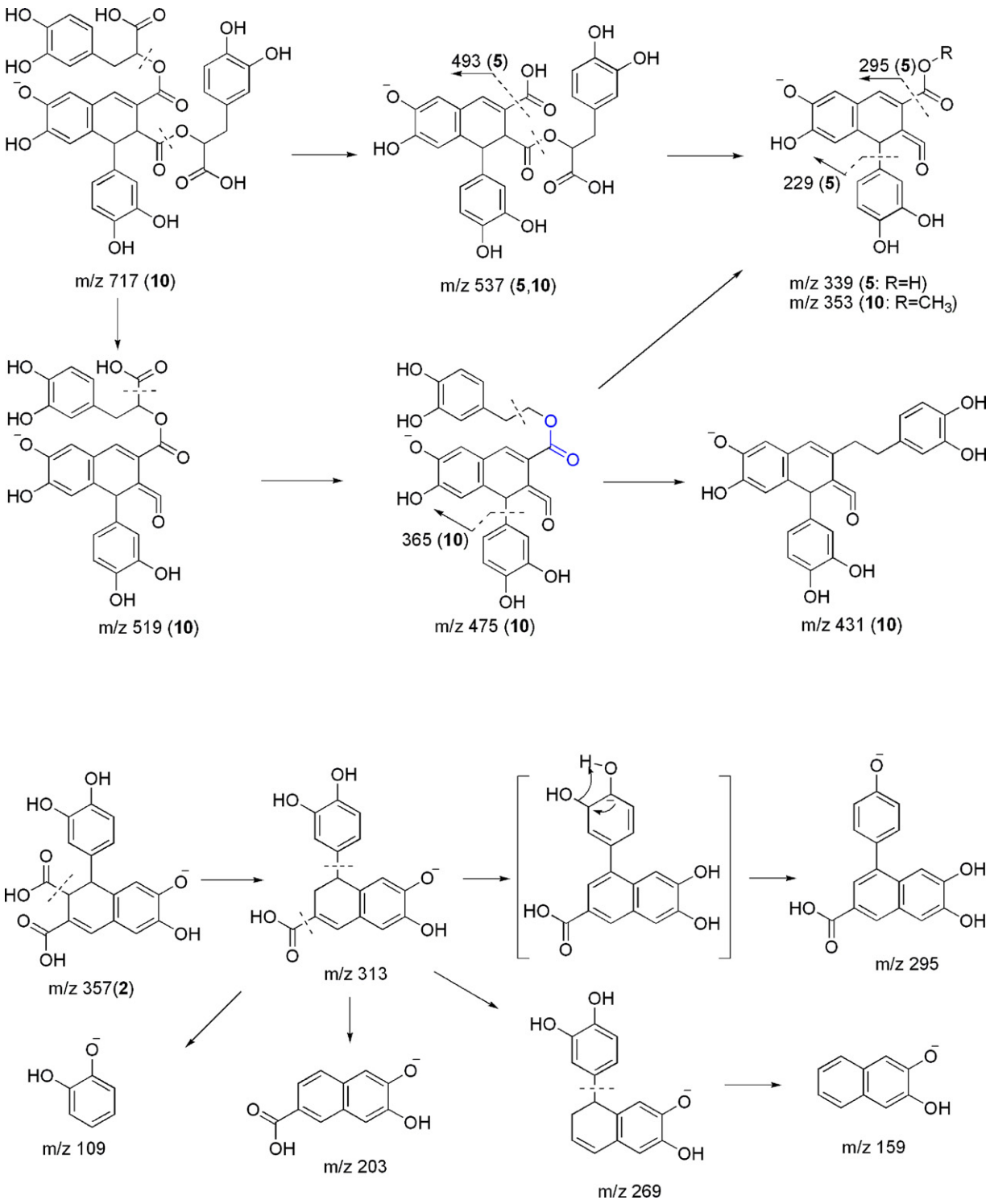


Fig. 11. Proposed fragmentation pathways for compounds **2**, **5**, and **10**.

corresponding to MF $C_{27}H_{30}O_{15}$. Further fragmentation of $[M–H]^-$ resulted in $[M–H–120]^-$ ions at *m/z* 473 and $[M–H–90]^-$ ions at *m/z* 503, which were consistent with the characteristic ions of flavonoid C-glycosides [34,35].

The MF of compound **10** was proposed to be $C_{36}H_{30}O_{16}$ according to its exact mass at *m/z* 717.1400, which was consistent with that of rabdosiin extracted from XCC [22]. The MS^n data of compound **10** also supported this assignment, and its fragmentation pathways are shown in Fig. 11. The $[M–H]^-$ of compound **5** was 198 Da (a danshensu unit) less than that of compound **10** and had similar fragmentation. Two possible structures (hydrolyzing the ester bonds at C-11 or C-12) could be obtained when danshensu is lost from rabdosiin. As shown in Fig. 11, the two danshensu groups in rabdosiin underwent different fragmentations due to the alpha-hydrogen of C-12. As a neutral loss of 198 Da was observed in the MS^2 of compound **5**, it was deduced that the danshensu unit was

connected with the C-12 carboxyl group. After thorough retrieval using SciFinder, compound **5** appeared to be a new structure, which was named rabdosic acid.

From comparisons with reported data [36], compound **14** was believed to be rosmarinic acid. The MS^n data (featuring a radical loss of 15 Da and a very stable backbone) of compound **21** suggested that it was a methylated flavonoid, such as pedalitin or isorhamnetin. This should be further confirmed by standard compounds. Compounds **15** and **20** from XCC have fragmentation patterns similar to those of limonoids; thus, they were tentatively characterized as deacetyl nomilin glucoside and obacunone glucoside, respectively [25]. In contrast with other ginsenosides, compounds **26** and **31** could not add to a molecule of formic acid and, thus, were tentatively identified as malonyl-ginsenoside Rb1 and ginsenoside Ro, respectively, based on their MS^2 data [29].

4.5. Adduct-ion patterns and types of compounds

Compounds in ZQ belong to flavonone glycosides and limonoids. The intensity ratios of their $[M+HCOO]^-/[M-H]^-$ were less than 0.5 except for compound **22**. So most of them were grouped into Class 1 and those (**3**, **15**, **20**) lacking a $[M+HCOO]^-$ were found in residual Class. By comparing the structure of **22** with other flavonone glycosides, we found that it may be the methoxyl group at C-4′ leading to its high abundance of $[M+HCOO]^-$. The triterpenoid saponins in HS can be classified into four types (Fig. 10) according to their skeletons. Similar to previous reports [10], we found that these saponins were inclined to form a base peak of $[M+HCOO]^-$ in their full scan mass spectra (Class 2) except for ginsenoside Ro (Type D, **31**) and malonyl-ginsenoside Rb1 (Type A, **26**). Compound **31** is an oleanane-type saponin and there is a malonyl group in the structure of **26**. This suggests that more compounds of such types should be tested to establish the relationship between their structures and the adduct ions in negative ESI/MS. The relative intensity of an adduct ion is also influenced by the compound's concentration and composition of mobile phases. So the adduct ion pattern could not be used as a rigorous standard for compound classification. However, such limitations of adduct ion patterns could not restrict their application to simplify a complicated LC/MS dataset.

5. Conclusion

The present study established a novel approach for processing LC/MS datasets, and applied this new method to studying the chemical composition of a TCM prescription drug. The approach utilized a Matlab-based LC/MS dataset extraction tool (CTCM) for compound classification, HPLC–ITMS for MS fragmentation pathway deduction, and HPLC–QTOFMS for MF determination. Constituents with different structural characteristics were classified automatically by their full mass spectra patterns, which smartly facilitated subsequent structural elucidation. Forty-six components in a WFC tablet, including 26 saponins, 10 flavonoids, and 10 other compounds, were either unambiguously identified or tentatively characterized. The method proposed here may potentially afford a rapid and efficient way with which to study the chemical compositions of complex samples, such as natural products and TCM prescriptions.

Acknowledgement

This study was supported by the National Science and Technology Major Project of China (No. 2011ZX09307-002-01) and the National Natural Science Foundation of China (Nos. 30830121 and 81001687).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2012.01.006.

References

[1] M. Yang, J. Sun, Z. Lu, G. Chen, S. Guan, X. Liu, B. Jiang, M. Ye, D.A. Guo, J. Chromatogr. A 1216 (2009) 2045.
[2] J. Zhou, L. Qi, P. Li, J. Chromatogr. A 1216 (2009) 7582.
[3] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, Anal. Chem. 78 (2006) 779.
[4] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, BMC Bioinformatics 11 (2010) 395.
[5] A. Lommen, Anal. Chem. 81 (2009) 3079.
[6] P.M. Palagi, D. Walther, M. Quadroni, S. Catherinet, J. Burgess, C.G. Zimmermann-Ivol, J.C. Sanchez, P.A. Binz, D.F. Hochstrasser, R.D. Appel, Proteomics 5 (2005) 2381.
[7] Y. Cai, R.B. Cole, Anal. Chem. 74 (2002) 985.
[8] A.N. Krutchinsky, B.T. Chait, J. Am. Soc. Mass Spectrom. 13 (2002) 129.
[9] K. Ng, C. Che, S. Wo, P.K. Tam, A.S. Lau, Rapid Commun. Mass Spectrom. 20 (2006) 1545.
[10] E.C.Y. Chan, S. Yap, A. Lau, P. Leow, D. Toh, H. Koh, Rapid Commun. Mass Spectrom. 21 (2007) 519.
[11] F. Cuyckens, M. Claeys, J. Mass Spectrom. 39 (2004) 1.
[12] Y. Zhou, M. Wang, X. Liao, Chin. J. Anal. Chem. 32 (2004) 174.
[13] National Commision of Chinese Pharmacopoeia, Pharmacopoeia of People's Republic of China, China Medical Science Press, Beijing, 2010, p. 883.
[14] H. Zhou, X. Xu, Z. Bian, J. Emerg. Tradit. Chin. Med. 18 (2009) 1421.
[15] Y. Chen, H. Wang, J. Yan, Zhejiang Clin. Med. J. 10 (2008) 907.
[16] Y. Lei, J. Gansu, Tradit. Chin. Med. 23 (2010) 39.
[17] P.G. Pedrioli, J.K. Eng, R. Hubley, M. Vogelzang, E.W. Deutsch, B. Raught, B. Pratt, E. Nilsson, R.H. Angeletti, R. Apweiler, K. Cheung, C.E. Costello, H. Hermjakob, S. Huang, R.K. Julian, E. Kapp, M.E. McComb, S.G. Oliver, G. Omenn, N.W. Paton, R. Simpson, R. Smith, C.F. Taylor, W. Zhu, R. Aebersold, Nat. Biotechnol. 22 (2004) 1459.
[18] http://tools.proteomecenter.org/wiki/index.php?title=Formats:mzXML, 2011.
[19] Y.F. Zhang, Q.Q. Zhao, J.F. Ma, B. Wu, X. Zeng, Chromatographia 72 (2010) 651.
[20] Y. Zhang, P. Shi, H. Qu, Y. Cheng, Rapid Commun. Mass. Spectrom. 21 (2007) 2971.
[21] X. Li, Y. Zhang, X. Zeng, L. Yang, Y. Deng, Rapid Commun. Mass Spectrom. 25 (2011) 2439.
[22] I. Agata, T. Hatano, S. Nishibe, T. Okuda, Chem. Pharm. Bull. 36 (1988) 3223.
[23] P. Shi, Q. He, Y. Song, H. Qu, Y. Cheng, Anal. Chim. Acta 598 (2007) 110.
[24] N. Fabre, I. Rustan, E. de Hoffmann, J. Quetin-Leclercq, J. Am. Soc. Mass Spectrom. 12 (2001) 707.
[25] G.K. Jayaprakasha, D.V. Dandekar, S.E. Tichy, B.S. Patil, J. Sep. Sci. 34 (2011) 2.
[26] X. Fan, Y. Wang, Y. Cheng, J. Pharm. Biomed. Anal. 40 (2006) 591.
[27] L. Li, R. Tsao, J. Dou, F. Song, Z. Liu, S. Liu, Anal. Chim. Acta 536 (2005) 21.
[28] Y. Liu, J. Yang, Z. Cai, J. Pharm. Biomed. Anal. 41 (2006) 1642.
[29] H. Zhang, Y. Wu, Y. Cheng, J. Pharm. Biomed. Anal. 31 (2003) 175.
[30] Y. Wei, P. Li, B. Shu, H. Li, Y. Peng, Y. Song, J. Chen, L. Yi, Biomed. Chromatogr. 21 (2007) 797.
[31] Y. Cheng, Q. Liang, P. Hu, Y. Wang, F.W. Jun, G. Luo, Sep. Purif. Technol. 73 (2010) 397.
[32] M. Cui, F. Song, Y. Zhou, Z. Liu, S. Liu, Rapid Commun. Mass Spectrom. 14 (2000) 1280.
[33] C. Zheng, H. Hao, X. Wang, X. Wu, G. Wang, G. Sang, Y. Liang, L. Xie, C. Xia, X. Yao, J. Mass Spectrom. 44 (2009) 230.
[34] F. Ferreres, B.M. Silva, P.B. Andrade, R.M. Seabra, M.A. Ferreira, Phytochem. Anal. 14 (2003) 352.
[35] J. Han, M. Ye, M. Xu, J. Sun, B. Wang, D. Guo, J. Chromatogr. B 848 (2007) 355.
[36] M.B. Hossain, D.K. Rai, N.P. Brunton, A.B. Martin-Diana, C. Barry-Ryan, J. Agric. Food Chem. 58 (2010) 10576.