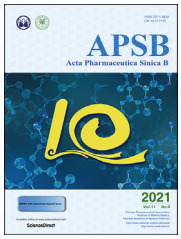
Acta Pharmaceutica Sinica B 2021;11(6):1469−1492

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com



REVIEW

Insight into chemical basis of traditional Chinese medicine based on the state-of-the-art techniques of liquid chromatography−mass spectrometry

Yang Yu[a,b,†], **Changliang Yao**[a,†], **De-an Guo**[a,b,*]

[a]*Shanghai Research Center for Modernization of Traditional Chinese Medicine, National Engineering Laboratory for TCM Standardization Technology, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China*
[b]*University of Chinese Academy of Sciences, Beijing 100049, China*

Received 7 December 2020; received in revised form 2 February 2021; accepted 22 February 2021

KEY WORDS

Liquid chromatography −mass spectrometry;
Qualitative analysis;
Traditional Chinese medicine;
Data acquisition;
Data post-processing

Abstract Traditional Chinese medicine (TCM) has been an indispensable source of drugs for curing various human diseases. However, the inherent chemical diversity and complexity of TCM restricted the safety and efficacy of its usage. Over the past few decades, the combination of liquid chromatography with mass spectrometry has contributed greatly to the TCM qualitative analysis. And novel approaches have been continuously introduced to improve the analytical performance, including both the data acquisition methods to generate a large and informative dataset, and the data post-processing tools to extract the structure-related MS information. Furthermore, the fast-developing computer techniques and big data analytics have markedly enriched the data processing tools, bringing benefits of high efficiency and accuracy. To provide an up-to-date review of the latest techniques on the TCM qualitative analysis, multiple

Abbreviations: BS, background subtraction; CCS, collision cross section; CE, collision energy; CID, collision-induced dissociation; cMRM, conventional multiple reaction monitoring; DDA, data-dependent acquisition; DE, dynamic exclusion; DIA, data-independent acquisition; DIF, diagnostic ion filtering; DM, database matching; EL, exclusion list; EMS, enhanced mass spectrum; EPI, enhanced product ion; FS, full scan; HCD, high-energy C-trap dissociation; IDA, information dependent acquisition; IM, ion mobility; IPF, isotope pattern filtering; ISCID, in-source collision-induced dissociation; LC, liquid chromatography; LTQ-Orbitrap, linear ion-trap/orbitrap; MDF, mass defect filtering; MIM, multiple ion monitoring; MN, molecular networking; MRM, multiple reaction monitoring; MS, mass spectrometry; MTSF, mass spectral trees similarity filter; NL, neutral loss; NLF, neutral loss filtering; NLS, neutral loss scan; NRF, nitrogen rule filtering; PCA, principal component analysis; PIL, precursor ion list; PIS, precursor ion scan; PLS-DA, partial least square-discriminant analysis; QqQ, triple quadrupole; QSRR, quantitative structure retention relationship; Q-TRAP, hybrid triple quadrupole-linear ion trap; RT, retention time; SA, statistical analysis; sMRM, scheduled multiple reaction monitoring; TCM, traditional Chinese medicine; UHPLC, ultra-high performance liquid chromatography.

*Corresponding author. Tel.: +86 21 20231000 2211; fax: +86 21 50272789.
E-mail address: daguo@simm.ac.cn (De-an Guo).

†These authors made equal contributions to this work.

Peer review under responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2021.02.017

data-independent acquisition methods and data-dependent acquisition methods (precursor ion list, dynamic exclusion, mass tag, precursor ion scan, neutral loss scan, and multiple reaction monitoring) and post-processing techniques (mass defect filtering, diagnostic ion filtering, neutral loss filtering, mass spectral trees similarity filter, molecular networking, statistical analysis, database matching, etc.) were summarized and categorized. Applications of each technique and integrated analytical strategies were highlighted, discussion and future perspectives were proposed as well.

1. Introduction

As the quintessence and treasure of China, traditional Chinese medicine (TCM) has taken on the responsibility to prevent and treat human diseases for millennia, and the excellent therapeutic effects of TCM have also been proved through the long historical clinical practice[1]. For this reason, TCM has always been inspiring scientists around the world and employed as an ideal library for development of drugs and dietary supplements[2]. However, disparate from mono-component Western medicines, the constituents of TCM are complex and yet undisclosed, thus leading to the vague mechanisms of action[3] and hampering the worldwide adoption of TCMs, especially in the Western world. Therefore, it is of vital importance to conduct more in-depth research for a better understanding of the chemical complexity of TCM, especially those closely related with the pharmacological activities and toxicities. However, on the one hand, a large number of minor and trace components exhibiting predominant biological activities cannot be detected with an ordinary approach; on the other, the compounds in TCM are usually one or multiple types of secondary metabolites, and the metabolites of the same type are structurally similar and hard to differentiate because of the common biosynthetic pathways. Accordingly, the number and structural variety of metabolites constitute a great analytical challenge in terms of detection and identification.

Combining the separation capacity of liquid chromatography (LC), especially the ultra-high performance liquid chromatography (UPLC), and remarkable selectivity and sensitivity of mass spectrometry (MS), the hyphenated analytical platform (LC−MS) shows significant advantages in analyzing complex chemical systems and has become one of the most powerful tools for analysis of TCM[4]. In an endeavour to explore the structural variety of metabolites, the first step is to acquire as much LC−MS information as possible, and the second step is to convert LC−MS data into chemical information. Between these two steps, except for the stationary phase development of LC[5] and hardware enhancement of MS instrumentations[6], the fast-developing MS data acquisition and data post-processing techniques also play a pivotal role[7,8].

Data acquisition methods by LC−MS can be roughly classified into data-independent acquisition (DIA) method and data-dependent acquisition (DDA) method. Typical DIA methods may provide production information in a non-selective manner without prior knowledge of the ions of interest. Featured with full coverage of MS^1 and MS^2 spectra in one injection, DIA methods maximally avoid the loss of MS information and possess satisfactory reproducibility and quantitative performance[9]. However, it is a daunting task to precisely assign the product ions generated by DIA methods to their precursor ions especially when the co-eluting compounds exist, due to the complex spectra and the lost link between MS^1 and MS^2, as well as requirements for more complicated deconvolution and data processing algorithms to further interpret the acquired MS data[10]. In a DDA method, a survey scan (usually full scan, FS) is first launched, and if certain criteria are met, further MS/MS (or MS^n) scan will be automatically triggered[7]. The diverse criteria set up in DDA methods provide customized solutions, taking advantages of the intrinsic chemical characteristics and fragmentation behaviors of the target compounds. In contrast to the DIA data, the structural information could be easily elucidated from DDA data, especially with the help of data post-processing techniques.

Even though plenty of researches and applications of LC−MS techniques in TCM qualitative analysis have been reported, it is still short of relevant reviews, with only several categorizing different compound classes and citing some representative examples of analyzing these compounds in TCM and TCM prescriptions[4,11–14]. Therefore, summary of universal and applicable techniques in TCM analysis are in urgent need. This review primarily focused on the applications of the advanced MS data acquisition and data post-processing techniques (Fig. 1), as well as integrated analytical strategies in qualitative analysis of TCM. Mechanisms of these techniques were firstly introduced, followed by representative applications. It is anticipated that this review would render a comprehensive reference for future research with regard to the metabolites identification in TCM complex systems.

2. Data acquisition techniques

2.1. *Data-independent acquisition techniques*

Typical DIA modes include AIF (all ion fragmentation) mode of Orbitrap system (Thermo Fisher Scientific, CA, USA)[15], MS^E (elevated energy MS) and SONAR mode of Q-TOF system (Waters, MA, USA)[16,17], MS^{ALL} and SWATH™ (sequential window acquisition of all theoretical mass spectra) mode of Triple TOF system (AB SCIEX, CA, USA)[18,19]. Among them, AIF acquisition mode allows all the precursor ions of defined range into the HCD (high-energy C-trap dissociation) cell for fragmentation[15]. MS^{ALL} is similar to MS^E which involves two scan functions, one is at low collision energy (CE), collecting intact molecular ion information with a wide mass range. The other one obtains product ion data using higher CE (fixed or ramped), which is useful for structural elucidation[16]. However, even though the aforementioned DIA modes avoid missing MS/MS information of minor or trace components in TCM, the defects of complex MS/

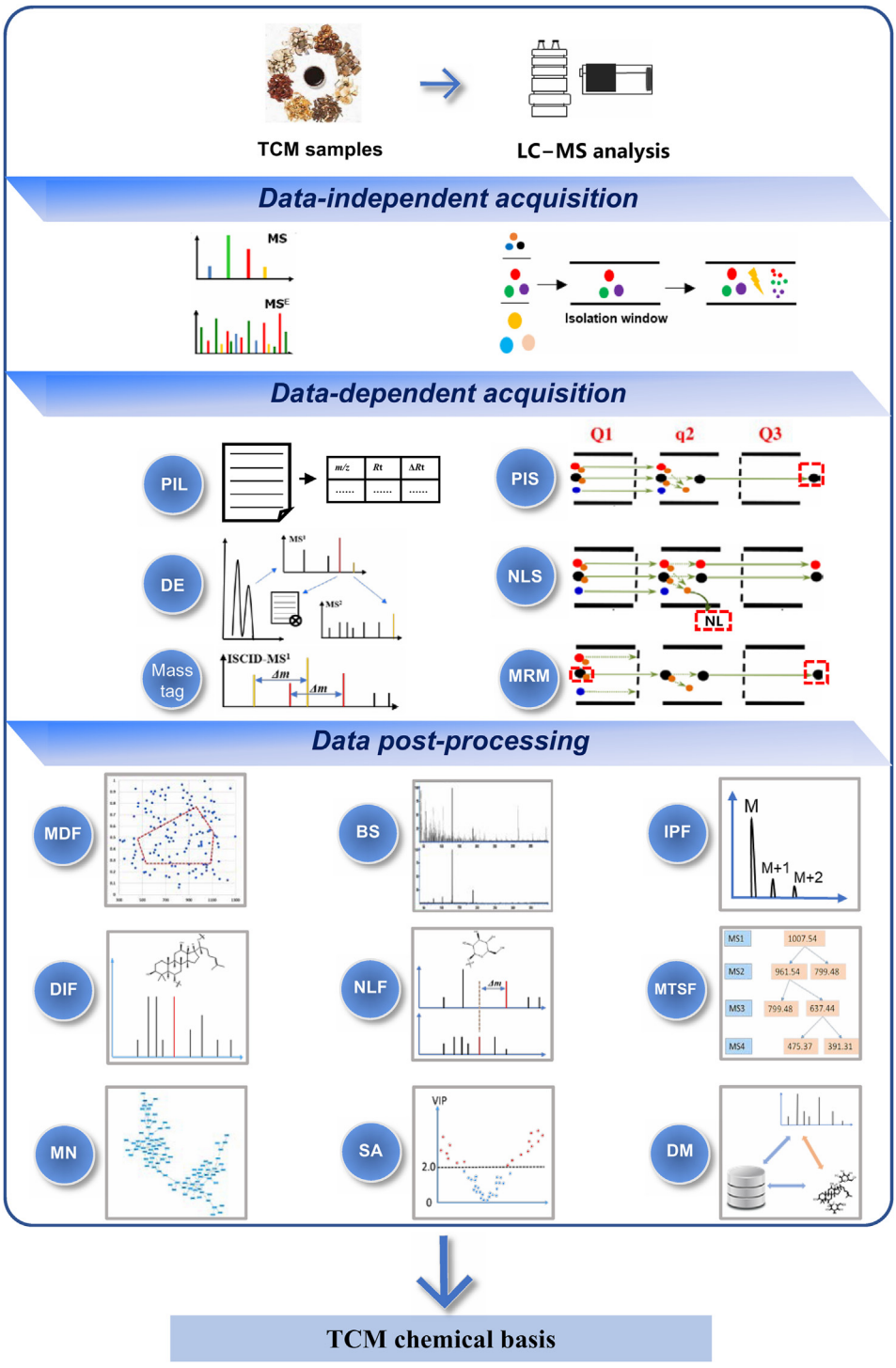


Figure 1 Data acquisition and data post-processing techniques of LC−MS in TCM qualitative analysis. PIL, precursor ion list; DE, dynamic exclusion; PIS, precursor ion scan; NLS, neutral loss scan; MRM, multiple reaction monitoring; MDF, mass defect filtering; BS, background subtraction; IPF, isotope pattern filtering; DIF, diagnostic ion filtering; NLF, neutral loss filtering; MTSF, mass spectral trees similarity filter, MN: molecular networking; SA, statistical analysis; DM, database matching.

MS spectra resulting from co-eluting ions still restricted the prevalent use of DIA techniques in TCM qualitative analysis. Luckily, different manufacturers have been devoted to improving the performance of DIA modes through instrument enhancement. For example, newly-developed SWATH™ acquisition mode allows the fragmentation of all ions across a given mass range in sequential narrow selection windows, thus retaining the benefits of both selectivity and scan speed[20]. While SONAR acquisition mode operates in a similar way, when the resolving quadrupole slides over a selected mass range during each MS scan with collision cell alternating between high and low CE from scan to scan, and the software aligns to associate the precursor and

product ions in different data channels. Besides, the application of ion mobility (IM) technique also helps obtain relatively clean and high-quality MS/MS spectra by enhanced separation of co-eluting ions based on their sizes, shapes and charge states[21].

Different DIA-based strategies have been developed for TCM analysis. For example, MS^{ALL} was applied in identification of triterpene saponins in *Acanthopanax senticosus* leaves by simultaneously collecting all the precursor ions ($[M+H]^+$, $[M+NH_4]^+$ and $[M+Na]^+$) and product ions with ramped CE on an AB SCIEX QTOF system, while SWATH mode as a complementation for co-eluting ions[22]. To characterize lanostane-type triterpene acids in *Poria cocos*, UPLC−IM−MS^E method was used to obtain multi-dimensional information including retention time, drift time and mass-to-charge ratio (*m*/*z*), where UPLC was for preliminary separation, whilst IM for identification of co-eluents by aligning precursor ions and corresponding product ions with the same drift time[23].

2.2. *Data-dependent acquisition techniques*

Traditionally, in a DDA method, precursor ions were selected for fragmentation according to their signal abundances in MS^1 scans. However, there could be a stochastic nature to data acquisition as TCM mixtures become more complex. In addition, only fragmentations of the *top-n* precursor ions were available. As a result, it often encounters obstacles in comprehensively detecting minor or trace components in TCM and systematically characterizing metabolites of the same type. Since the multiple types of constituents in a certain TCM are closely related to one or more common biosynthetic pathways, they could be structurally classified into several chemical families according to the same carbon skeletons or substructures with varieties in substitutional groups. Therefore, MS fragmentation patterns of chemically similar compounds usually displayed the same neutral losses, product ions, or other characters. Accordingly, various DDA techniques targeting one-type or multi-types of metabolites, and covering the trace metabolites have been developed.

2.2.1. *Precursor ion list*

In a precursor ion list (PIL) experiment, only (or preferred) target ions contained in the list can be scanned to trigger multistage fragmentations. PIL-based acquisition thus becomes a commonly used DDA technique with high selectivity, sensitivity and efficiency in discovery and characterization of target compounds. Unlike conventional DDA method that automatically fragments the *top-n* abundant ions, MS/MS or MS^n data are acquired regardless of their peak abundances, overcoming the drawback of neglecting minor components in TCM extracts.

In a PIL-based acquisition, the key is to establish a sample-specific PIL, based on molecular features of *m*/*z* value or a combination with retention time[24]. One of the most widespread approach is collecting previous phytochemical reports for further structural prediction. With the loss or addition of different substitutional groups to certain core substructures in accordance with the biosynthetic pathways, the molecular formulas of these derivatives are predicted, correspondingly, molecular weights and theoretical parent ions can be calculated. PIL constructed in this way encompasses both known and predicted compounds, thus greatly increasing the coverage and improving the selectivity of detecting expected compounds in TCM. Another approach, is using various data mining techniques to explore the potential metabolites, such as mass defect filtering (MDF)[25−27], neutral loss filtering (NLF)[28,29] and molecular networking (MN)[30]. Their principles and applications in qualitative analysis of TCM are discussed in depth in Section 3.

For example, some compound-specific PILs were generated by summarizing the reported natural compounds, and applied in profiling of components in *Carthamus tinctorius* L. (safflower)[31], Ligustri Lucidi Fructus[32], etc. Besides, based on compound prediction, aglycones substituted by 2−7 hydroxyl and methoxyl groups were included in a PIL, achieving targeted screening of 135 polymethoxylated flavonoids when analyzing the sample of *Citrus reticulata* Blanco on a hybrid linear ion-trap/Orbitrap (LTQ-Orbitrap)[33]. Another PIL constructed by the arrangement of different substitutional groups to tanshinol was also used to characterize polymeric phenolic acids in *Salvia miltiorrhiza*[34].

One successful application of generating PIL with data post-processing techniques is conducted by a FS-based untargeted metabolomics approach[35]. This method displays great advantages in analyzing complex systems, as no prior knowledge of the sample is necessary. Besides, separating a long PIL into several individual ion lists based on *m*/*z* value or time point can drastically avoid data loss and reduce duty cycle of the instumentation[34]. There is evidence showing that DDA triggered by time-staggered PIL exhibited superior selective advantages in activating MS/MS acquisition of co-eluting ions compared to the conventional DDA[35]. In summary, PIL-MS^n, especially the time-triggered mode, displays excellent superiority in acquiring the fragmentation information of target ions, thereby enhancing the selective detection of target components in TCM.

2.2.2. *Dynamic exclusion*

Contrary to PIL−MS/MS technique which preferentially trigger MS/MS fragmentation of predefined ions in list, exclusion list (EL)-based acquisition methods improve the selectivity and sensitivity of characterization by entering an EL of masses that will be ignored for MS^n analysis, exhibiting great advantages in removing background interfering ions and non-targeted ions, and showing potentials in exposing more novel compounds[36]. Except for abundant interfering ions, there will always be target components that completely overlap when analyzing TCM complex systems regardless of the optimized chromatographic separation conditions. Under these circumstances, only the *top-n* intense ions can be selected for fragmentation, and MS/MS information of less abundant co-eluting ions will escape the detection. Dynamic exclusion (DE) provides a solution for this problem by temporarily putting a mass over the defined threshold into an EL for a selected period of time after its MS/MS spectrum is acquired, enabling less abundant ions to be analyzed rather than repeated scans of the abundant ions[37]. For this reason, DE is frequently set to improve the detection coverage in TCM qualitative analysis, especially for detecting those less abundant co-eluting ions[25,26,28,38].

The advantages of DE were fully illustrated by screening the polymethoxylated flavonoids from *Citrus reticulata* Blanco[33]. Compared with FS, FS-PIL and FS-DE on LTQ-Orbitrap, FS-PIL-DE acquisition was found to be most powerful in triggering MS/MS fragmentations. Its superiority in acquiring the MS^n information was further confirmed by another study for screening indole alkaloids from *Uncaria sinensis*[26].

Appropriate DE parameters are also essential for MS data acquisition. Because they are closely related to the chromatographic peak width, optimal chromatographic conditions should

be investigated to minimize the differences in the peak widths of different retention times[34]. Besides, short repetition time and long exclusion time could minimize MS information loss of a complex system but potentially miss the fragmentation of the neighboring isomers. Therefore, a coordination of these two parameters is always a necessity for the better data quality.

2.2.3. *Mass tag*

In a mass tag dependent acquisition, the parent ions with certain mass differences were recognized as mass partners, and selected to trigger MS^n fragmentation. The mass partners can be produced by introducing isotopically labeled compounds, such as chrysin-d_0 and chrysin-d_5 with a mass difference of 5 Da. The technique of mass tag can largely improve signal-to-noise ratios and data quality in detection of tagged compounds, including reactive metabolites[39] and modified peptides[40]. Mass tag technique can be performed on LTQ-Orbitrap and applied to selectively detect phosphopeptides and *N*-glycopeptides with an in-source collision-induced dissociation (ISCID) energy on. After the *m/z* values of the charged ions in MS^1 spectra were determined and converted into masses, mass partners that differed in the mass of defined mass tag and with intensity above the defined threshold, were selected to trigger MS^2 acquisition with the ISCID off[40].

Recently, expanding applications of mass tag can be found in TCM qualitative analysis, especially for targeted screening of components that can easily undergo in-source fragmentations. For example, malonylginsenosides from three *Panax* species were targeted screened out by defining mass tag of 43.9898 Da with an ISCID energy of 40 V, corresponding to the elimination of CO_2[41]. In addition, by applying ISCID of 50 V and mass tag of 46.0055 Da (referring to formic acid), neutral ginsenosides from three *Panax* species were also successfully classified and characterized[42]. In summary, mass tag-oriented acquisition has unique advantages in selectively screening the target components in TCM with a high coverage and introducing less false positives.

2.2.4. *Precursor ion scan*

Precursor ion scan (PIS) can be realized in a triple quadrupole (QqQ) mass spectrometer. Ions within a defined mass range are scanned in the first quadrupole (Q1), fragmented in the collision cell (q2), and then one specific product ion is selected in the third quadrupole (Q3)[43]. This acquisition mode allows selective screening of compounds with identical product ions, and is popular in metabolites identification[44], lipidomics study[45]. Besides, PIS also shows great potential in characterization of various TCM components. Its advantage lies in requiring little prior-knowledge about the exact structure of compounds.

Facilitated with PIS technique, target compounds that are expected to produce the same product ions can be screened out, such as the esters, saponins, flavonoids, etc. For example, benzoyl substituted diterpenoid alkaloids with diagnostic benzoyl ion at *m/z* 105[46]; caffeoylquinic acid derivatives with diagnostic product ion of *m/z* 191 referring to quinic acid anion[47]; oplopane-type sesquiterpenoids with ions at *m/z* 215 and 217, and bisabolane-type sesquiterpenoids with ions at *m/z* 229 and 231[48]. Besides, glycosyl-conjugated compounds could also be easily detected using characteristic ions corresponding to their core structures and sugar moieties. As an example, triterpene saponins in *Glycyrrhiza yunnanensis* with saccharide chains of glucuronic acid (GluA) and rhamnose (Rham) were detected using ions at *m/z* 351 $[2GluA-H]^-$ and 497 $[2GluA+Rham-H]^-$[49]. Polyoxypregnane and its glycosides were identified from *Marsdenia tenacissima* (Roxb.) Wight et Arn., using characteristic ions at *m/z* 329 and *m/z* 273, corresponding to the aglycone core and sugar moiety at C-3 position, respectively[50]. And characteristic ions of $153^+/151^-$ were used to identify glycosyl-free flavonoids and then the related glycosyl-conjugated flavonoids were further searched using ions of the identified glycosyl-free flavonoids in PIS mode[51].

PIS-information dependent acquisition (IDA)-enhanced product ion (EPI) mode implemented on a hybrid triple quadrupole-linear ion trap (QTRAP) mass spectrometer is more effective in rapid detection and characterization of certain types of analytes than experiments on QqQ, due to an additional MS/MS fragmentation. It was applied to rapidly identify phenolic constituents in Danhong injection using six diagnostic ions, they are *m/z* 197 $([M-H]^-)$, 179 $([M-H-H_2O]^-)$, 135 $([M-H-H_2O-CO_2]^-)$, 123 $([M-H-H_2O-2CO]^-)$ generated from danshensu-related compounds, 137 $([M-H]^-)$ and 108 $([M-H-CHO]^-)$ generated from protocatechualdehyde-related compounds[52]. Wider applications of PIS-IDA-EPI in TCM qualitative analysis include identification of naphthalenyl glycosides and amino naphthoquinones in *Juglans cathayensis*[53], coumarins in Glehniae Radix[54], phenylethanoid glycosides, iridoids, and lignans in *Cistanche deserticola* and *C. tubulosa*[55], and so on.

2.2.5. *Neutral loss scan*

In the neutral loss scan (NLS) mode of QqQ and QTRAP, ions within a defined mass range are scanned in the first quadrupole (Q1), fragmented in the collision cell (q2), and then Q3 scan range shifts by Δm to a low mass, which corresponds to a specific neutral loss (NL) of every potential precursor ion[43], *e.g.*, 162 Da corresponding to glucose. It is widely used for specific screening of modified metabolites that can undergo neutral eliminations, and also highly selective and sensitive to profile and characterize various specific conjugated components in TCM, as well as discovery of new compounds.

Because *O*-glycosides can easily undergo the NL of sugars, NLS was used to profile glycosides with different aglycones in tobacco leaves[56]. NL-IDA-EPI scan mode was employed to screen glycosyl flavonoids in Astragali Radix with NL of 162 or 179 Da for screening of flavonoid aglycones that tended to form protonated or ammoniated glycosides, respectively[57]. Triterpene saponins in *Glycyrrhiza yunnanensis* were profiled using adducts-targeted NL-EPI by monitoring NL of 17 Da (referring to NH_3) and 46 Da (referring to HCOOH) generated from their ammonium and formic acid adducts, respectively[58]. Sulfated flavonoids in *Flaveria* were detected by a NL of 80 Da (representing SO_3) in negative mode[59]. And iridoids, lignans, and phenylethanoid glycosides in *Cistanche deserticola* and *C. tubulosa* were screened with NL of 162 Da (glucose residue or a $C_9H_6O_3$ group), 146 Da (coumaroyl group) and 176 Da (feruloyl groups)[55]. In addition, NLS was also enabled in high-resolution LTQ-Orbitrap, as the exact mass of the neutral loss can be used to screen the targets more accurately. For instance, malonylginsenosides were screened from nine Ginseng extracts with NL set at 43.9898 Da, corresponding to loss of a CO_2 unit from malonyl group[41]. Steroids substructures of dicarboxylic acid conjugated bufotoxins in *Venenum bufonis* were obtained by a NL-based MS^3 acquisition in positive ion mode with neutral masses of seven dicarboxylic acids[38].

2.2.6. *Multiple reaction monitoring*

Multiple reaction monitoring (MRM) mode is a typical scan mode operated in QqQ and QTRAP. By applying an appropriate CE, certain parent ions detected in Q1 cell can fragment in q2 cell and

generate product ions passing through Q3 cell. Under optimal MS conditions, MRM-based acquisition, by simultaneously monitoring specific parent ion and corresponding product ion with the highest intensity, can drastically improve the selectivity and sensitivity of target components in TCM.

MRM scan mode was adopted to screen out 77 flavonoid glycosides in the flower of *Carthamus tinctorius* L.[60], identify 80 triterpenes in Alismatis Rhizoma[61], etc. MRM-IDA-EPI scan on a QTRAP mass spectrometer was also utilized to identify 27 major components (18 diterpenoids, 6 phenolic acids, and 3 flavonoids) in *Isodon serra*[62], characterize 421 flavonoids in Astragali Radix[57], verify the structures of all tentative curcuminoids in turmeric[63], and screen ginsenosides in Ginseng, American Ginseng and their processed materials[64].

However, the increasing number of analytes monitored in a single run will result in lower sensitivity of detection, because either the dwell time for each transition will be reduced and unable to accumulate sufficient signal for each transition, or the cycle time for each MS scan will increase, decreasing the data points of each peak[65]. Scheduled MRM (sMRM, or dynamic MRM) offers an option to increase the number of MRM transitions without compromising the data quality, by monitoring every sMRM transition in a defined range of retention time[66]. The superiority of sMRM over conventional MRM (cMRM) was proved by detection of components in a TCM mixture made up of eight commonly used TCM[67], and more analytes could be detected using sMRM than cMRM method.

Multiple ion monitoring (MIM) mode could be considered as a special type of MRM mode, with a minimal collision energy applied in q2 cell, thus the product ion monitored in Q3 cell is identical to the parent ion in Q1 cell. In general, suitable ion transitions in an MRM method are designed based on the prior-knowledge of the fragment patterns of the parent ion, causing great challenges to screen unknown compounds. MIM method can overcome these limitations and be used to detect potential compounds regardless of their fragmentation patterns, thus playing an important role in targeted monitoring of compounds of interest in TCM, and can be served as a complementary approach for MRM[68]. For example, in order to identify naphthoquinones in *Juglans cathayensis*, MIM-EPI method was conducted based on 36 different molecular weights of all reported naphthoquinones from family Juglandaceae, stepwise MIM-EPI with a wide specified mass range (150.0−700.0 Da) was conducted for discovery of untargeted naphthoquinones, and segmented stepwise MIM-EPI with a narrow specified mass range was used for identification of certain types of naphthoquinones[53]. Another strategy integrating predefined MRM, step-wise MIM, and EPI scans was also proposed to universally screen the hydrophilic substances in Shenfu injection, and a total of 157 hydrophilic compounds were detected, 154 of which were identified as amino acids, nucleosides, organic acids, carbohydrates, etc.[69].

3. Data post-processing techniques

Conventionally, the structures of compounds were manually elucidated based on their MS fragmentation patterns, which requires laborious spectral assignments and produces somewhat experience-based results. Similar to the DDA techniques, newly-developed data post-processing techniques take full advantages of the same neutral losses, product ions, or other characters produced by MS/MS fragmentation. In addition, the MS characters in FS, such as mass defect, can also be a clue for compound recognition. In contrast to the DDA techniques, different data post-processing techniques can be applied to the same MS data for different purposes without reacquisition of the MS data. For example, the MS data of *Carthamus tinctorius* L. acquired by LTQ-Orbitrap was explored for both the quinochalcone *C*-glycosides[70] and flavonoid *O*-glycosides[71] with versatile data mining strategies, showing the advantages of data post-processing techniques on recycling of MS data. In a word, various data post-processing techniques can greatly simplify the manual interpretation process of the MS data, providing more conveniences, evidences and higher accuracy for TCM qualitative analysis.

3.1. *Mass defect filtering*

The calculation of the mass defect is conducted by the mass difference between the exact mass and nominal integer mass of different elements, in this way, each molecule can be described with a unique exact mass and mass defect based on its elemental composition[72]. Therefore, structural analogues sharing similar core substructure with various chemical substituted groups will have very similar mass defects, and the limited differences are only corresponded to the different substituted groups. Based on this fact, mass defect filtering (MDF) technique is developed and used as a specific filtering criterion to selectively pick out certain compounds or compound classes in a complex mixture. For example, a target class of endogenous substances, or parent drug and its metabolites[73,74] in a complex biological sample within an allowable mass defect range. In addition, MDF is also a powerful tool to expedite the screening process of different components contained in TCM and to generate cleaner profiles by providing more distinct and specific information[75].

Classic MDF algorithms, including fixed MDF algorithms with a rectangular coverage area and linear gradient MDF algorithms with a parallelogram coverage area[76], can be conveniently realized by some commercial software such as Peakview[77], Metabolynx XS[78], Metworks[79] and MetID[80]. And both the mass range and the mass defect range can be defined for automatic exclusion of irrelevant ions from complex matrices. For example, a mass defect filter (mass range: 367.5 ± 65.5 Da; mass defect: 186 ± 40 mDa) was established using Metworks, which not only encompassed all free indole alkaloids in *Uncaria rhynchophylla*, but also distinguished indole alkaloids from triterpenic acids (mass range: 453−649 Da, mass defect: 333−413 mDa) and other components[79]. Besides, stepwise MDF approach which divides the mass defect range or mass range into multiple mass defect windows or mass windows were proved to improve the signal to noise ratio of the target peaks, and polymethoxylated flavonoids were successfully screened from the leaves of *Citrus reticulata* Blanco[81]. The strategy combining a single MDF window (mass range: 267−418 Da; mass defect: 0.04−0.12 Da) for methoxylated flavonoids, and multiple MDF windows (mass range/mass defect: 337−398 Da/0.08−0.12 Da; mass range/mass defect: 483−604 Da/0.10−0.18 Da; mass range/mass defect: 629−810 Da/0.13−0.23 Da) for three classes of chlorogenic acids was also successfully constructed and applied in Folium Artemisiae Argyi[82].

However, because the distribution space of target compounds established by a classic MDF strategy usually covers a quite large area, increasing the possibility of false positives, several in-house polygonal MDF techniques have been developed to narrow down

the distribution space and focus more on the target ions. A modified MDF strategy, termed as "five-point screening" MDF strategy was proposed to rapidly screen saponins in *Panax notoginseng*[83]. Every selected ion of notoginsenosides was distributed in the region depicted by five points (integer mass as *x* and decimal mass as *y*), and the "IF" function of the Microsoft excel platform was used to pick out every potential ion distributed in this region in the.xls documents. Compared with the classic MDF, this modified approach effectively removed more putative spaces and thus increasing the screening efficiency. Another polygonal MDF algorithm established by a relatively compact octagonal region was also used to screen target alkaloids in *Uncaria sinensis,* as well as enable three novelty levels classification: known, unknown-but-predicted, and unexpected[26]. The region established by eight vertexes were generated by a two-dimensional distribution plot of the mass range (Da) and the mass defect range (mDa) of the known and unknown-but-predicted molecules, and multiple "IF" equations representing the octagonal region were edited in Microsoft Excel to rapidly screen the alkaloid components. This modified MDF algorithm greatly enhanced the accuracy in screening target components, and more importantly, enabled the discovery of novel compounds.

Despite of the great efforts made in exploration of different MDF algorithms, including classic rectangular MDF and modified MDF algorithms, for reducing false positives and expanding the coverage, tedious data processing procedures are required to generate sample-specific target ions. To solve this problem, a raster-MDF screening strategy was implemented by a step-wise PIL with the parent mass width of ±55 mDa, and utilized for screening and characterization of indole alkaloids from Uncariae Ramulus Cum Unicis of multiple botanical origins[84]. This strategy not only comprehensively extended the coverage of potential indole alkaloids, but also excluded interferences of other components from the genus *Uncaria*.

3.2. *Background subtraction*

Background interferences are common in TCM analysis, which could be traced back to all steps of sample preparation procedures such as plasticizers from pipette tips and centrifuge tubes[85], or just chemical noise resulting from mobile phases or buffers in LC−MS analytical systems[86]. To suppress interference of the background noise and obtain relatively clean spectra, BS technique is therefore introduced. It can be implemented manually[87], directly in a commercial software (Waters Masslynx, Thermo Scientific Metworks[26,88], etc.) by selecting a region with background noise and then subtracting from the original chromatogram, preparing blank samples in the same manner as the test samples for detection and removal of the background interferences, or utilizing some customized programs for further information[89].

BS is a useful technique in elimination of background signals, as well as exposure of chemical constituents in TCM, especially those with a low abundance. To identify unknown compounds in Er-xian decoction, signals corresponding to the background interferences were checked manually and excluded from 533 candidate compounds, and finally 240 non-background compounds were extracted[87]. To comprehensively detect and characterize the chemical constituents in Arnebiae Radix, the obtained total ion chromatogram (TIC) was first processed with BS, and then MDF to filter the background-subtracted ion chromatogram, resulting in characterization of 96 compounds, including 30 with a low abundance[88].

3.3. *Isotope pattern filtering*

Isotope pattern filtering (IPF) technique is specially developed for compounds containing elements of distinct isotopic distribution patterns in mass spectra, such as chlorine, bromine, iodine, etc. With the aid of IPF, compounds containing certain elements can be easily screened out. For instance, an ion with isotopic mass distance of 2 Da and equal isotopic abundance is most likely to correspond to a compound containing one bromine atom. Except for these natural stable isotopes, other isotope-labeled compounds such as ^{14}C, ^{15}N and ^{18}O-labeled compounds are also widely used in tracing and detecting related metabolites[90]. Compared with traditional IPF implemented by commercially available software (*e.g.*, Metworks), some newly-developed algorithms such as accurate-mass-based spectral-averaging IPF enhances the specificity by inspecting not only the M+2 isotopic ion and M molecular ion, but also the M+1 isotopic ion[91].

In recent years, sulfurous compounds with very low abundance in TCM including *Pueraria lobata* (Willd.) Ohwi (Gegen) and *Pueraria thomsonii* (Fenge) were detected and characterized by a fine IPF technique[92]. Despite of the low abundance of ^{34}S signal and the influence of $^{13}C_2+^{18}O$, this IPF strategy achieved screening of S-containing compounds by defining the accurate mass shift and relative abundance between M+2S (representing the M+2 isotopic ion of $^{12}C^{1}_{x+2}H^{16}_{y}O^{34}_{z+1}S$) and M (representing the molecular ion) as well as M+2OC (representing the M+2 isotopic ion of $^{12}C^{1}_{x}H^{16}_{y}O^{32}_{z}S^{13}C^{18}_{2}O$) and M in Compound Discoverer software.

3.4. *Diagnostic ion filtering*

Similar to PIS, diagnostic ion filtering (DIF) provides a criterion for rapid classification of the target compounds that produce the identical product ions, by extracting the relevant ions in the MS/MS or MS^n channels. Whereas PIS can only be realized by triple quadrupole like low resolution mass spectrometry, DIF can be applied to high resolution data obtained by QTOF, LTQ-Orbitrap, and Q-Orbitrap, thus leading to high selectivity. In addition to the diagnostic ions, the high-resolution data and abundant fragmentation information drastically increase the identification confidence. DIF has been applied to characterize different types of chemical components in TCM, including saccharides and glycosides[93], quinones[94], phenylpropanoids[82,95], flavonoids[96–98], triterpenes[61], triterpenoid saponins[22,95,99,100], alkaloids[101,102], other compounds such as physalins[103], as well as multiple compound classes in the same sample[104,105].

To get the diagnostic ions in abundance, different dissociation methods at different energies have been explored on LTQ-Orbitrap. For example, to selectively identify flavonoid *O*-glycosides from *Carthamus tinctorius*, combination of two fragmentation modes, CID and HCD, was used for obtaining complementary fragmentation information, including high-mass product ions produced from loss of sugars, low-mass product ions originated from aglycone moieties, as well as intensity ratios of radical aglycone ion species useful for precise elucidation of aglycones and glycosylation patterns[71]. DIF combined with enhanced separation strategy such as multidimensional LC and IM can greatly expand the peak capacity and enhance sensitivity of characterizing minor components with structural diversity in TCM. An LTQ-

Orbitrap approach using DIF of *m/z* 119.05 ($C_8H_7O^-$) was developed to characterize 163 quinochalcone *C*-glycosides in *Carthamus tinctorius* L. by combing offline two-dimensional LC separation[70]. Another multidimensional analytical strategy combining LC−IM−MS^E and LC−IM−MS/MS were established to screen polycyclic polyprenylated acylphloroglucinols using diagnostic ions *m/z* 165.0182 or/and 177.0182, leading to the successful characterization of 140 targets in *Garcinia oblongifolia*[106].

In addition, DIF has also significantly facilitated the rapid classification and structural elucidation of more complicated multi-components contained in TCM prescriptions, including ginsenosides and lignans in Shenmai injection[107], phenolic acids, saponins, diarylhepatonoids and gingerol-related compounds in Guge Fengtong tablet[108], flavonoids, ginkgolides, phenolic acids, tanshinones, and saponins in Yindan Xinnaotong soft capsule[109], etc. In summary, DIF is an option to filter and classify target ions, as well as remove interfering ions from the raw LC−MS data.

3.5. Neutral loss filtering

Neutral loss filtering (NLF) works by extracting the predicted specific neutral loss between the daughter ions and parent ions. NLF is similar to NLS, and the advantage lies in circumventing the use of multiple injections because of the combination use of multiple filters to identify multiple targets. NLF is promising in untargeted characterization of endogenous modified metabolites with specific neutral loss fragments, such as acetylation, sulfation, glucuronidation, etc.[110], and exclusive profiling of phase II metabolites such as GSH conjugates in drug metabolism studies[111].

In analysis of TCM, NLF also finds its place in characterization of glycosidic compounds such as flavonoid glycosides[71,109], alkaloid glycosides[79] and triterpenoid saponins[95], acylation compounds such as malonyl ginsenosides[99] and diester-diterpenoid alkaloids[112], as well as other compounds substituted with certain moieties such as methoxylated flavonoids[82], etc. For example, the diester-diterpenoid alkaloids in *Aconitum carmichaelii* Debx. were determined by the diagnostic neutral losses of 60 28, 32, 18 and 122 Da, corresponding to AcOH, CO, MeOH, H_2O and BzOH moieties, respectively[112]. Besides, NLF of the recognized sugars and acylation (GluA, Glc, Rha, Ace) was conducted for target characterization of flavonoid *O*-glycosides in *Carthamus tinctorius*, and some unknown ones (132.04, 86.00 and 71.98 Da for Xyl, Mal, and Oxa, respectively) were also detected and characterized[71].

A software named Neutral Loss MS Finder was developed to extract the in-source NL[110], and successfully applied in screening the target malonyl-ginsenosides from three *Panax* species[29]. Because of the in-source removal of both neutral CO_2 and $C_3H_2O_3$ groups, two NL filters (43.9898 and 86.0004 Da) were set to reduce the false positives and enhance the characterization accuracy, and ultimately, to produce the peak lists containing 156, 130 and 126 target malonyl-ginsenosides in three *Panax* species, respectively.

3.6. Mass spectral trees similarity filter

Mass spectral trees similarity filter (MTSF) is a data post-processing technique for rapid classification of the unknown compounds by comparing the similarity of mass spectral trees. Commercially available software such as Mass Frontier is usually used to construct mass spectral trees, using HRMS data as the stem of the tree, MS^2 data as the bough, and multi-stage MS^n data further extended as the branches. After importing MS^n data of the template compounds into the software to build the library of MTSF, the similarity scores between the compounds detected in the sample and those in the library can be calculated, and unknown compounds can be classified and characterized by matching with those templates to obtain candidate structures.

Because the mass spectral trees of structural analogs usually exhibit high similarity scores, MTSF technique is widely applied for rapid identification of metabolites in a complicated biological matrix, due to the structural resemblance between the metabolites and corresponding prototype[113]. Meanwhile, it also reveals great potential to significantly accelerate and simplify the process of discovery and identification of various chemical components in TCM which share the same substructures as template compounds, including flavones, phenylpropanoids, and sphingolipids in *Saussurea involucrata*[114], chlorogenic acids in *Duhaldea nervosa*[115] and Lonicerae Japonicae Flos[116], etc. In addition, for TCM prescriptions containing diverse types of compounds, MTSF can also efficiently fish potential components belonging to different compound classes, and rapidly identify both templated compounds and related compounds in Xiao-Xu-Ming decoction[117], prenylflavonoid glycosides, alkaloids and phenylpropanoids in Er-xian decoction[87]. In a word, taking full advantages of MS^n information, MTSF is a simple and efficient technique to fish unknown components in TCM complex systems, especially the structural analogs of the template compounds.

3.7. Molecular networking

Molecular networking (MN) uses a visualized computational strategy for comparison of the MS/MS spectral and calculation of the similarity degree between structural analogs. The major strength of this technique lies in requirements of no prior knowledge of the chemical structures for simultaneous exploration of up to millions (probably billions) of MS^2 data[118]. Developed from early-used MATLAB scripts installed on computers for similarity computation, Global Natural Products Social Molecular networking (GNPS) (http://gnps.ucsd.edu), an open-access online platform, allows even more analysts with no professional bioinformatics background to take advantage of this advanced technique[119]. Data visualization can be performed either directly online, or offline with Cytoscape[120] as well as other visualization tools, where similar MS/MS spectra can be grouped into clusters, and each node represents a unique MS/MS spectrum, connected with edges indicating the cosine similarity between corresponding MS/MS spectra. To help data interpretation, node size and color can also be tuned according to precursor ion intensity and sample origin. Moreover, GNPS can automatically retrieve detected MS/MS spectral against public libraries, allowing rapid identification of molecules that are similar to database records. In addition to known compounds, MN also allows for exploration and identification of potential unknown analogues or compound families based on their MS/MS spectra relatedness with any known compounds. Compared with MTSF, MN is suitable for more LC−MS platforms and attracts wider attentions.

Since the introduction of MN in metabolic profiling of live microbial colonies in 2012[121], it has been successfully applied in multiple areas[122], such as drug discovery and natural products dereplication[123], including microbials[124−127], marine organisms[128−130], fungi[131], and plants[132−136]. In recent years, MN has also been used in TCM qualitative analysis. For example, offline

two-dimensional separation integrated with MN was developed for dereplication and sorted-characterization of 229 bufadienolides in Venenum Bufonis, including two new subclasses[137]. Rapid recognition of 537 components in Lonicerae Japonicae Flos was also achieved with the help of MN, including a myriad of potential novel structures[36]. In addition, MN was also applied for identification and exploration of potential chemical markers of Aconiti Lateralis Radix Praeparata before and after processing based on both the compound clusters and node areas[138].

3.8. *Statistical analysis*

In face of complex and large-scale mass spectral data, although several data post-processing techniques such as DIF, NLF, MDF, MTSF and MN have been utilized to facilitate the data interpretation process and enhance the identification accuracy of target compounds in complex samples, it's still a challenge in the aspect of rapid recognition and extraction of potential characteristic ions. Therefore, in order to save time and manpower to dig for attractive compounds, some advanced statistical analysis (SA) techniques have been developed based on the intrinsic relationships between compound structures and MS data. Targeted data with similar characteristics can be classified into the same groups by statistical methods, while compounds of different substructures in TCM are efficiently discriminated. Combined with some diagnostic tools, such as loading plot and variable importance in projection (VIP), rapid recognition of potential characteristic ions can be easily achieved.

SA technique was explored for rapid screening of the minor compounds in Danhong injection[139]. Firstly, a total of 7157 product ions were extracted from MS/MS spectra of 22 reference standards, and 4 filtering ions specific for compounds in Dan-shen, Hong-hua, and both herbs were selected from 195 common ions. Combined with another 6 diagnostic ions, 117 compounds were finally identified. Besides, SA was also utilized for discovery and global profiling of novel compounds from *Curcuma longa*[140]. After targeted detection of 846 terpecurcumins by 12 NL/PIS, principle component analysis (PCA) was used for discrimination of different structures and recognition of potential novel compounds by clustering similar NL/PIS patterns and differentiating distinct NL/PIS patterns. In another example of profiling phenylethanoid glycosides from *Magnolia officinalis*[141], recognition models of partial least square-discriminant analysis (PLS-DA) were established to discriminate both the stereoisomers and the positional isomers using MS^2 data of the reference standards, and product ions contributing the most to isomer recognition were picked out according to the ranking order of their VIP values. Finally, PCA analysis was conducted for discrimination of isomers in samples and verification of the selected discriminant ions.

3.9. *Database matching*

Since manual analysis of massive and informative data generated by LC−MS is a laborious task, which requires a tremendous amount of time and effort, and often generates non-reproducible results, database matching (DM) can significantly improve the efficiency *via* computer-aided extensive annotation of compounds in complex matrices, and is widely applied in identification of a variety of compounds. DM was accomplished by comparing the exact mass and fragments between the detected compounds and the recorded items in the database. These databases generally include chemical databases using accurate mass or molecular formula, and MS/MS spectral databases such as MassBank[142], METLIN[143], mzCloud[144], HMDB, etc. Since the applications of some open-access online-databases in identification of compounds have already been described in several published reviews[145−148], this review mainly focuses on the in-house databases used in identification of TCM constituents.

Some intelligent software platforms from different manufacturers, including UNIFI[23,149−156], Progenesis QI[157], MetWorks[158], Compound Discoverer[159], SURIUS[160], MassHunter[161] and PCDL Manager[162], play an important part in alleviating the labor for mining structural information from large-scale datasets, by integrating various automated data processing steps. They usually provide accessibility to various online databases and allow creation of a customized database as well, facilitating automatic profiling and identification of chemical components in TCM and more complex TCM formulas[155]. For example, to rapidly elucidate structures of lanostane analogs and isomers in *Poria cocos*, LC−IM−MS^E data were post-processed with the assistance of UNIFI software, a targeted compound database and a key MS database, resulting in identification of 121 lanostanes[23]. In a bid to expand the screening coverage of novel compounds in the leaves of *P. notoginseng*, a predicted metabolites screening approach was implemented in UNIFI, by using four modified groups and one or two steps of modification, increasing the theoretical coverage by 14 times, and finally 945 ginsenosides were discovered, including 662 potential novel ginsenosides[154]. Another strategy was developed to systematically profile lipids in three congeneric *Panax* species by searching full MS and fragments information against HMDB and LIPID MAPS, and collision cross section (CCS) information against Metabolic Profiling CCS Library (incorporated in Progenesis QI) and LipidCCS Predictor software[157].

Besides, several self-developed automatic data-processing tools can also increase the chemical coverages of traditional databases by creation of in-house databases, enhancing the efficiency and reproducibility in characterization of compounds in TCM. FlavonQ was developed for characterization of flavone and flavonol glycosides by calculation and combination of the common substituted groups of the core structures, and the results obtained by FlavonQ proved to be consistent with those determined conventionally by a sophisticated chemist, but greatly facilitated the data interpretation process in flavonoid research[163]. Another custom-built software PlantMAT (Plant Metabolite Annotation Toolbox), was also developed for identification of saponins and glycosylated flavonoids by running through various glycosyl, and acyl groups in every possible combination[164].

4. Integrated analytical strategies

In majority of the cases, single analytical method is insufficient for the comprehensive characterization of the complex constituents in TCM, since each technique has its own strength and limitations. Therefore, development of integrated analytical strategies is necessary for rapid, accurate and systematic chemical analysis of TCM. The integrated strategies can be developed by integration of different MS instrumentations, or integration of various data acquisition or/and data post-processing techniques, laying the solid foundation for interpretation of the chemical diversity of TCM.

High-resolution mass spectrometers (such as QTOF) with accurate mass measurement of ions are advantageous for untargeted qualitative analysis of TCM. However, extraction of

corresponding precursor and product ions from the total ions chromatograms (TICs) could be time-consuming and labor-intensive. On the contrary, QTRAP-MS show its advantages in sensitive detection of target compounds especially co-eluted minor components in TCM by using multiple targeted scanning modes such as MRM, and an EPI scan can also be triggered to acquire fragmentations of the preferred precursor ions. Therefore, QTRAP could act as a complementary tool of QTOF despite of its low resolution. Combination of MS/MS functions of both QTOF and QTRAP have been applied in systematic characterization and targeted identification of triterpenes in Alismatis Rhizoma and the corresponding processed products[61], curcuminoids in turmeric[63], naphthoquinones in *Juglans cathayensis* dode[53], as well as minor components in TCM formulas such as Baoyuan decoction[155].

The integrated use of different DDA techniques is also powerful in identifying unknown trace compounds in TCM. Take QTRAP as an example, it possesses the scan capacities of both QqQ and linear ion trap (LIT), involving MIM, MRM, NLS, PIS, enhanced scans such as EPI and enhanced mass spectrum (EMS), as well as some combined functions, such as PIS-IDA-EPI. By the combined use of MIM-IDA-EPI and PIS-IDA-EPI modes, 41 coumarins were identified from Radix Glehniae[54]. Besides, integration of different triggered IDA-EPI scans, including those triggered by EMS, NLS, MRM, and PIS scans, lead to the detection of 513 components from *Cistanche deserticola* and *C. tubulosa*[55]. In the study of chemical profiling of Baoyuan decoction, the stepped MIM-EPI and predefined MRM-EPI data acquisition methods were established to pick out potential saponins based on the MS behaviors of authentic compounds, while MIM-EPI and PIS-EPI scanning modes were adapted to comprehensively recognize flavonoids based on various diagnostic product ions, and thus the co-eluted chromatographic peaks were easily distinguished and extracted[155].

Compared with individual filtering techniques, integration of different data post-processing techniques also exhibits higher efficiency in classifying multi-type compounds, excluding interference ions, as well as detecting the minor or trace components in TCM. It is illustrated by a strategy integrating different filtering techniques, including MDF and nitrogen rule filtering (NRF) based on MS information, and DIF and NLF based on MS/MS information, to identify two types of components in Artemisiae Argyi Folium, including chlorogenic acids and methoxylated flavonoids[82]. Results showed that none of a single technique could totally eliminate the interfering ions, whereas the integration of MS and MS/MS-based filtering techniques exerted a complementary effect to remove more interfering ions and expose target compounds. In another example for rapid identification of chemical constituents in Yindan Xinnaotong soft capsule, MDF and DIF were applied to preliminarily classify detected compounds into different chemical families, while NLF were used for characterization of flavonoid glycosides and triterpene saponins substituted with certain sugar units[109].

Intelligent DDA techniques allow targeted acquisition of MS/MS or MS^n data, improving the sensitivity and selectivity in detection of attractive compounds, while various data post-processing techniques can further simplify the data interpretation process. Therefore, combined approaches which integrate targeted DDA and various data post-processing techniques have great advantages in detection and identification of target components in TCM. A strategy based on LTQ-Orbitrap following offline two-dimensional LC separation was established to systematically analyze five botanical origins of Uncariae Ramulus Cum Unicis, and lead to the ultimate characterization of 1227 indole alkaloids[84]. In the data acquisition process, a theoretical step-wise PIL (mass range: 310−950 Da; step size: 2 Da) and an optimal parent mass width (±55 mDa), was defined to selectively trigger fragmentations of potential alkaloids, namely step-wise PIL-based raster-MDF scan method. Additionally, different data post-processing techniques were also combinedly used to facilitate structural elucidation, including elemental composition analysis to remove false positives, DIF for subtype classification, DM and fragments annotation for structure characterization. In another example, 537 components were characterized in Lonicerae Japonicae Flos, including an enormous number of new structures[36]. To reduce the redundant scans and increase the selectivity, an EL was first generated by a two-step polygonal MDF and then used for triggering data dependent-MS/MS scan. To rapidly annotate the known compounds and elucidate those unreported ones, DM and MN were both adopted for the data interpretation process.

All the above-mentioned applications of LC−MS in TCM qualitative analysis were summarized in Table 1. Besides, the selection criteria of each data acquisition and data post-processing technique were also listed in Table 2, including the advantages, limitations, and scope of applications, providing valuable reference for future research on characterization and identification of chemical constituents in TCM. In addition, a decision tree in Fig. 2 also explained how to select appropriate data acquisition/post-processing methods based on the known chemical information, which may help the readers' practical adoption of the above methods.

5. Discussion and future perspectives

Despite of the continuous developments of advanced data acquisition and post-processing techniques in TCM qualitative analysis, there still remains drastic challenges to be solved to achieve the ultimate goal of comprehensive chemical profiling and annotation. Firstly, it is virtually impossible to simultaneously detect all the components in one experiment, due to the selectivity and priority of the analytical system. Therefore, arrays of separation techniques differing from conventional reversed-phase LC were developed and combined in either on-line mode or off-line mode[165], including hydrophilic interaction liquid chromatography[36,152,154], supercritical fluid chromatography[137], etc. Besides, IM[23,106,157] was also adopted to improve the MS performance. Secondly, to make the full use of typical DIA methods featured with minimum loss of MS information and good reproducibility, more complicated deconvolution and data processing techniques should be developed to further interpret the DIA data. Thirdly, manual annotation and interpretation of the MS data is still a dominant method used, which is labor-intensive and somewhat subjective. Databases of MS spectra of TCM ingredients combined with intelligent software may be an option. Though database search is anticipated to play a pivotal role in matching the known compounds, however, the comprehensive TCM databases of the known compounds are still in serious shortage, let alone the potentially new structures. Therefore, *de novo* annotation software, especially those considering the biosynthetic pathways may help increase the identification credibility. Fourthly, ambiguous differentiation and identification of the isomers hinder the further application of the acquired data. Common isomer identification strategies involve two aspects. One is associated with the liquid chromatographic separation by comparing the difference of

Table 1 Applications of LC−MS in TCM qualitative analysis.

TCM sample	Mass analyzer	Data acquisition technique	Data post-processing technique	Characterization result	Ref.
Acanthopanax senticosus	Q-TOF	MS^{All}; SWATH	DIF	89 triterpene saponins, along with 14 sapogenins, including 33 potentially new compounds and the first report of malonyl-saponin in genus *Acanthopanax*	22
Poria cocos	Q-TOF	IMS-MS^{E}	DM; NLF	121 lanostane-type triterpene acids, including three new compounds	23
Yi-Xin-Shu capsule	Q-TOF	PIL-MS/MS	DIF	276 compounds, including 50 lignans, 30 tanshinones, 4 lactones, 18 flavones, 14 phenolicacids, 1 triterpenoid acid, 3 ophiopogonins, 7 astragalus saponins, and 149 ginsenosides	24
Erzhi pill	Q-Orbitrap	Full MS/PIL/dd-MS^2; "If idle-pick others" (IIPO); DE	MDF; NLF; DIF	146 components, including 25 triterpenes, 31 flavonoids, 46 iridoids, 16 phenylethanols, 3 coumarins, 12 phenols, and 13 others	25
Uncaria sinensis (stem, leaf, and flower)	LTQ-Orbitrap	PIL-DE-CID/MS^2-HCD/MS^3	BS; MDF; DM	158 potentially new alkaloids, including 10 unknown but predicted and 108 unexpected ones	26
Three different parts (the root, stem leaf, and flower bud) of *Panax quinquefolius* L.	Q-Orbitrap	Full MS/PIL/IIPO/dd-MS^2; DE	MDF	347 saponins (147 from root, 173 from stem leaf, and 195 from flower bud), and 157 thereof not ever-isolated from the *Panax* genus	27
Puerarin lobate (Gegen)	LTQ-Orbitrap	PIL; DE	NLF	66 compounds, including 43 malonates of isoflavone glycoside	28
Panax ginseng, *P. quinquefolius*, and *P. notoginseng*	LTQ-Orbitrap	ISCID-FS; PIL-HCD-MS/MS; DE	NLF	101 malonyl-ginsenosides, including 69 from *P. ginseng*, 52 from *P. quinquefolius*, and 44 from *P. notoginseng*	29
Eleutherococcus senticosus leaves	LTQ-Orbitrap	HCD-MS^2; PIL-CID-MS^4; FS-MS	MN; DM	106 triterpene saponins including 49 potentially new ones	30
Carthamus tinctorius L. (safflower)	Q-Orbitrap	full MS/PIL/dd-MS^2; IIPO; DE	Retention behavior; Reference compounds comparison; Elemental composition analysis; Fragmentation pathways interpretation; DM	13 primary metabolites (1 nucleoside, 2 sugars, 5 organic alkali/acids, and 5 amino acids) and 135 secondary metabolites (97 quinochalcone *C*-glycosides and 38 flavonoids)	31
Ligustri Lucidi Fructus	Q-Orbitrap	Full MS/PIL/dd-MS^2; IIPO; DE; Polarity switching	Reference compounds comparison; Elemental composition analysis; Fragmentation pathways analysis; DM	158 components, including 137 in the negative mode and 21 in the positive mode	32
Citrus reticulata Blanco (leaves)	LTQ-Orbitrap	FS-PIL-DE	DIF	135 polymethoxylated flavonoids, including 81 polymethoxyflavones, 54 polymethoxyflavanones or polymethoxychalcones	33
Salvia miltiorrhiza	LTQ-Orbitrap	PIL-MS^n; DE	Predicted compounds filtering; Total ion chromatogram filtering; Fragment ion search	190 polymeric phenolic acids, including 18 first detected ones	34
Lonicerae Japonicae Flos	Q-Obitrap	EL-dd-MS^2	DM; MN	A total of 537 compounds, including a large number of potential novel structures	36

(*continued on next page*)

Table 1 (*continued*)

TCM sample	Mass analyzer	Data acquisition technique	Data post-processing technique	Characterization result	Ref.
Venenum bufonis	LTQ-Orbitrap	FS-DE-MS^1; HCD-MS^2 (−); CID-MS^2 (−); CID-MS^3 (−); Product ion-MS^4 (−); CID-MS^2 (+); NL-MS^3 (+)	–	78 dicarboxylic acid conjugated bufotoxins, including 68 new compounds, 25 types of substructure formulas and seven dicarboxylic acid side chains	38
Roots, leaves, and flower buds of *Panax ginseng*, *P. quinquefolius*, and *P. notoginseng*	LTQ-Orbitrap	ISCID-FS-MS^1; Mass tag/CID-MS^2; NLS/HCD-MS^3; DE	–	178 malonylginsenosides	41
Panax ginseng, *Panax quinquefolius*, and *Panax notoginseng*	LTQ-Orbitrap	ISCID-FS-MS^1; Mass tag/CID-MS^2; Product ion scan/CID-MS^3; DE	NLF	216 carboxyl-free ginsenosides	42
Processed *Aconitum* roots (*Shengfuzi*, *Heishunpian*, *Baifupian*, *Zhichuanwu*, and *Zhicaowu*)	QTOF; QqQ	PIS	–	24 benzoyl-containing alkaloids	46
Cynara scolymus L.	Q-TRAP	PIS	–	10 caffeoylquinic acid derivatives	47
Tussilago farfara L. (buds)	QqQ; QTOF	PIS; Product ion scan	–	74 oplopane- and bisabolene-sesquiterpenoids, and isolation of 11 compounds	48
Glycyrrhiza yunnanensis	QqQ; QTOF	NLS; PIS	–	50 triterpene saponins	49
Marsdenia tenacissima (Roxb.) Wight et Arn.	QqQ	PIS; FS; Product ion scan	DIF; NLF	110 polyoxypregnane and its glycosides	50
Nicotiana tabacum	QqQ	Two-step PIS	–	17 flavonoids, including 9 newly identified ones	51
Danhong injection	QTRAP	EMS-IDA-EPI; PIS-IDA-EPI	DM	90 compounds, including 46 salvianolic acids and related phenolic compounds	52
Juglans cathayensis	QTRAP; QTOF	EMS-EPI; Stepwise MIM-EPI; PIS-EPI	NLF	48 naphthoquinones including 24 novel ones	53
Radix Glehniae	QTRAP	MIM-IDA-EPI; PIS-IDA-EPI	–	41 coumarins	54
Cistanche deserticola and *C. tubulosa*	QTRAP	EMS-IDA-EPI; Predefined MRM-IDA-EPI; PIS-IDA-EPI; NL-IDA-EPI	DM	513 components detected, including 379 annotated compounds	55
Tobacco leaves	QqQ	NLS; Product ion scan	DM	64 glycosides, including 39 glycosides linked with monosaccharides, 18 glycosides linked with disaccharides and 7 glycosides linked with trisaccharides	56
Astragali Radix (Huangqi)	QTRAP	MIM-EPI; PI-EPI; NL-EPI; MRM-EPI	Extracted ion chromatogram; DIF; NLF	421 flavonoids	57
Astragali Radix	QTRAP	NL/PIS-IDA-enhanced resolution-EPI	–	136 triterpenoid saponins	58
Flaveria (*F. robusta*, *F. pringlei*, *F. linearis*, *F. anomala*, *F. australasica*, and *F. bidentis*) seeds	QTRAP	NLS	–	27 sulfated flavonoids	59
Carthamus tinctorius	Ion trap	FS; MS^n; MRM	DIF; NLF	77 flavonoid glycosides	60

Table 1 (*continued*)

TCM sample	Mass analyzer	Data acquisition technique	Data post-processing technique	Characterization result	Ref.
L.					
Alismatis Rhizoma and processed Alismatis Rhizoma	QTOF; QTRAP	FS; MS/MS; MRM	DIF; NLF	80 triterpenes including 14 novel compounds; 7 more triterpenes compounds in the processed *Alismatis rhizoma*	61
Isodon serra	QTRAP	MRM-IDA-EPI	–	27 components, including eighteen diterpenoids, six phenolic acids, and three flavonoids	62
Turmeric	TripleTOF; QTRAP	CID-MS/MS; MRM-EPI	BS; DIF; Ring double bond equivalents calculation	96 curcuminoids	63
Ginseng, American ginseng, and processed products	QTRAP	Step-wise MRM-IDA-EPI; sMRM	PCA	221 ginsenosides, including 185 annotated ones	64
Shenfu injection	QTRAP	Predefined MRM-IDA-EPI; Step-wise MIM-IDA-EPI	DM	157 detected hydrophilic compounds, and 154 of which were identified as amino acids, nucleosides, organic acid, carbohydrates, etc.; 40 primary hydrophilic and hydrophobic ingredients including 11 amino acids, 9 nucleosides, 9 aconite alkaloids, and 11 ginsenosides	69
Carthamus tinctorius L.	LTQ-Orbitrap	FS-MS^1; CID-MS^2; CID-MS^3; DE-HCD-MS/MS	DIF; Ring double bond equivalent; Characteristic UV absorption; Diagnostic product ions analysis	163 quinochalcone *C*-glycosides homologs, including 149 potential new ones	70
Carthamus tinctorius L.	LTQ-Orbitrap	CID-MS^1-MS^2-MS^3; DE-HCD-MS^1-MS^2	DIF; NLF; DM	107 flavonoids, including 80 newly reported flavonoid *O*-glycosides	71
Ophiopogon japonicus	Q-TOF; IT-TOF	FS; MS^1-MS^2-MS^3	MDF	More than 50 ophiopogonins and 27 ophiopogonones	77
Yin Chen Si Ni Tang	Q-TOF	MS^E	MDF	62 *Aconitum* alkaloids	78
Uncaria rhynchophylla	LTQ-Orbitrap	DE-CID-MS^4	MDF; NLF; DIF; DM	92 alkaloids (60 free alkaloids and 32 alkaloid *O*-glycosides), 56 of which were potential new alkaloids for the *Uncaria* genus	79
Processed Semen Strychni	IT-TOF	FS; MS^2 and MS^3	MDF; DIF	24 dihydroindole-type alkaloids, including 4 that were previously not described	80
Citrus reticulata Blanco	LTQ-Orbitrap	MS and MS/MS	MDF; DIF	81 polymethoxylated flavonoids, including 50 polymethoxyflavones and 31 polymethoxyflavanones or polymethoxychalcones	81
Folium Artemisiae Argyi	Q-Orbitrap	FS; MS/MS	NRF; MDF; NLF; DIF	16 methoxylated flavonoids and 55 chlorogenic acids analogues	82
Panax notoginseng (Sanqi)	Q-TOF	MS; Auto MS/MS; targeted MS/MS	MDF; Characteristic isotopic distribution	234 ginsenosides including 67 potential new ones	83
Five botanical origins of Uncariae Ramulus Cum Unicis	LTQ-Orbitrap	DE-step-wise PIL-CID/MS^2-HCD/MS^3	Elemental composition analysis; NLF; DIF; DM	1227 indole alkaloids	84
Er-xian decoction	LTQ FT	FS; DE-CID-MS^2; CID-MS^3	MTSF; Discriminant analysis; BS	553 potential compounds and 66 candidates	87
Arnebiae Radix	Q-Orbitrap	Full MS; MS/MS	BS; DM; MDF	96 compounds, 13 of which were confirmed by the reference standards, 30 with a low abundance, and 9 unknowns	88

(*continued on next page*)

Table 1 (*continued*)

TCM sample	Mass analyzer	Data acquisition technique	Data post-processing technique	Characterization result	Ref.
Pueraria lobata (Willd.) Ohwi (Gegen) and *Pueraria thomsonii* (Fenge)	LTQ-Orbitrap	FS; DDA	IPF	12 *S*-derivatives were found, 9 of which were tentatively identified	92
Forsythia suspensa	Q-TOF	MS^E; CID-MS/MS	DIF	8 potentially new C_6–C_2 glucoside conjugates	93
Zicao	Q-TOF	MS; MS/MS	BS; DIF; Characteristic product ion filtering	58 compounds including 32 novel ones	94
Akebiae Fructus	Q-TOF	MS^E	Adduct ion filtering; DIF; Characteristic ion filtering; NLF	94 compounds (85 triterpenoid saponins and 9 chlorogenic acids), including 9 types of triterpenoid saponins and 2 types of chlorogenic acid; 50 newly discovered constituents	95
Spatholobus suberectus	Q-TOF	FS in both positive and negative modes; (−)-MS/MS	DM; DIF	38 compounds, including 36 flavonoids and 2 non-flavonoid compounds	96
Scutellaria baicalensis	Q-Orbitrap	FS; MS/MS	DIF	132 compounds, 59 of which were reported for the first time	97
Amphimas pterocarpoides	LTQ-Orbitrap	FS; MS/MS	DIF	21 molecules, 17 of which were found to belong to the targeted chemical groups, including 8 isoflavones, 7 dihydroisoflavones and 2 isoflavone glycosides	98
Roots and berries of raw and steamed American ginseng (*P. quinquefolius* L.)	TOF; Q-TOF	FS; CID-MS/MS	DIF	70 saponins	99
Fresh and processed ginseng	Q-TOF	FS; CID-MS/MS	DIF; NLF	12 malonyl ginsenosides	100
Uncaria rhynchophylla	Q-TOF	MS; MS/MS	DIF	A total of 29 compounds, comprising 18 alkaloids, 6 flavonoids, and 5 quinic acids; 4 novel tetracyclic monoterpenoid oxindole alkaloids (TMOAs)	101
Lycoris spp. (Amaryllidaceae)	Q-TOF	MS; MS/MS	DIF	37 lycorine-type alkaloids, including 16 previously undescribed compounds	102
Calyx of *Physalis alkekengi* L. var. *franchetii* (Mast.) Makino	Q-TOF	MS; MS/MS	DIF; NLF	46 physalins, including 20 novel ones	103
Glechomae Herba	Q-TOF	MS; MS/MS	DIF; NLF	120 compounds, including 10 chlorogenic acids, 10 gallic acids, 21 phenylpropionic acids and 77 flavonoids; 65 newly discovered constituents; 4 types of chlorogenic acids, 3 types of galloylglucoses, 3 types of phenylpropionic acid skeletons, and 5 types of flavonoid aglycone skeletons	104
Raw and processed pieces of *Rheum palmatum* L.	Q-TOF	(−)-MS; (−)-MS/MS	DM; DIF	73 characteristic markers	105

Table 1 (*continued*)

TCM sample	Mass analyzer	Data acquisition technique	Data post-processing technique	Characterization result	Ref.
Garcinia oblongifolia	QTOF	MS/MS; ISCID MS/MS; ISCID combined with TAP fragmentation; IM-MS^E; IM-MS/MS	DIF	140 potential polycyclic polyprenylated acylphloroglucinols (PPAPs), including 7 pairs coeluting isobaric PPAPs that were indistinguishable by conventional UHPLC−HRMS alone	106
Shengmai injection	IT-TOF	MS^1, MS^2, MS^3	DIF	More than 30 ginsenosides and 20 lignans	107
Guge Fengtong tablet	Q-TOF	FS	DIF	47 components, including 18 phenolic acids, 8 saponins, 14 gingerol-related compounds, and 7 diarylhepatonoids	108
Yindan Xinnaotong soft capsule	Q-TOF; QqQ	FS; MRM	DM; DIF; MDF; NLF	124 compounds, including 44 flavonoids, 31 phenolic acids, 20 tanshi-nones, 16 saponins, 6 ginkgolides, 5 sulfides, camphor and borneol	109
Roots of *Aconitum carmichaelii* Debx.	LTQ-Orbitrap	FS; CID-MS^2-MS^3	NLF	42 diester-diterpenoid alkaloids; 23 potential new compounds, including 16 short chain fatty acyls diester-diterpenoid alkaloids, 4 *N*-dealkyl diester-diterpenoid alkaloids and several isomers of aconitine, mesaconitine and hypaconitine	112
Saussurea involucrata	LTQ FT	DE-CID-MS^3	MTSF	38 compounds, including 19 flavones, 11 phenylpropanoids and 8 sphingolipids; among them, 7 flavonoids, 8 phenylpropanoids and 8 sphingolipids were identified for the first time in *Saussurea involucrate*	114
Duhaldea nervosa (Wallich ex Candolle) A. Anderberg	LTQ-Orbitrap	FS; CID-MS^2; DE	MTSF	47 chlorogenic acids, including 19 monoacyl-quinic acids, 22 diacyl-quinic acids, and six triacyl-quinic acids	115
Flos Lonicerae Japonicae	LTQ-Orbitrap	FS; FS-PIL; MIM	DM; MTSF; DIF	115 chlorogenic acids attributed to 18 categories, most of which were newly reported	116
Xiao-Xu-Ming decoction	LTQ FT	FS; DE-CID-MS^2; CID-MS^3	MTSF	68 compounds among 3362 detected compounds, including 14 templated compounds (reference compounds), 50 related compounds fished by MTSF technique, and 4 unrelated compounds identified by manual method	117
Venenum Bufonis	Q-TOF	Fast DDA	MN	229 bufadienolides, including two subclasses of compounds (bufogeninsconjugated with carboxylic acid and *N*-heterocyclic bufogenins) which were found in Venenum Bufonis for the first time	137
Aconiti Lateralis Radix Praeparata and its processed products	Q-TOF	MS^E	MN	145 diterpenoid alkaloids, including 78 C^{19} (19 diester-diterpenoid alkaloids, 20 monoester-diterpenoid alkaloids, 7 lipo-diterpenoid alkaloids, 32 amine-diterpenoid alkaloids), 13 C20, and 54 other diterpenoid	138

(*continued on next page*)

Table 1 (*continued*)

TCM sample	Mass analyzer	Data acquisition technique	Data post-processing technique	Characterization result	Ref.
				alkaloids	
Danhong injection	Q-Orbitrap	FS	SA	117 compounds, including 76 phenolic acids, 20 flavonoids, and 21 other compounds	139
Curcuma longa	QqQ	NL/PIS	SA	846 terpecurcumins (terpene-conjugated curcuminoids), including a number of potentially novel compounds	140
Magnolia officinalis	Q-TOF	MS; MS/MS	DIF; NLF; SA	87 phenylethanoid glycosides, including 14 isomers	141
Ziziphi Spinosae Semen and Ziziphi Mauritianae Semen	Q-TOF	MS^E	DM; DIF; NLF	60 target components in Ziziphi Spinosae Semen, including 27 flavonoids, 16 saponins, 10 alkaloids, 6 terpenes, and 1 other, and 53 nontarget components with 40 new deduced components; 132 chemical components in Ziziphi Mauritianae Semen, including 7 additional nontarget new components	149
Peanut stems and leaves	Q-TOF	MS^E	DM	283 chemical compounds, including 207 new compounds	150
White and red ginsengs	Q-TOF	IM-MS^E; IM-MS/MS	DM	201 ginsenosides, including 10 pairs of co-eluting isobaric ginseng saponins	151
Stems and leaves of *Panax ginseng*	LTQ-Orbitrap	DE-CID-MS^3; DE-HCD-MS^2	NLF; DIF; DM	646 ginsenosides, including 427 which have not been isolated from the genus of *Panax* L.	152
Five different parts (root, leaf, flower bud, berry, and seed) of *Panax ginseng*	Q-TOF	MS^E	DM	164 compounds	153
Panax notoginseng	Q-TOF	Fast DDA	DM; DIF; NLF	945 ginsenosides, including 662 potentially novel ginsenosides	154
Baoyuan decoction	Q-TOF; QTRAP	FS-MS^1; MS^E; Predefined MRM-EPI, MIM-EPI, and PIS-EPI	DM; DIF	236 compounds, including 139 saponins, 83 flavonoids, 6 procyanidins, 4 lignans, and 4 diterpenes	155
Paris polyphylla Smith var. *yunnanensis* (Franch.) Hand.-Mazz and *Paris vietnamensis* (Takht.) H. Li	Q-TOF	MS^E	DM	146 metabolites, including 42 potential new compounds	156
Panax ginseng, *P. quinquefolius*, and *P. notoginseng*	Q-TOF	MS^E; $HDMS^E$	DM	24 potential lipid markers enabling the simultaneous differentiation of three *Ginseng* species	157
Cistanche deserticola, *C. sinensis*, and *C. tubulosa*	LTQ-Orbitrap	FS-HRMS; MS/MS; DE	DM; Parent mass modification search; DIF; MTSF	69 phenylethanoid glycosides, including 17 new ones	158
Aconitum carmichaelii Debx.	Q-TOF	MS; MS/MS	DM	148 lipo-alkaloids, including 93 potential new compounds and 38 compounds with oxygenated fatty acid moieties	161
Gelsemium elegans	Q-TOF	MS/MS	DM	57 components, including 43 alkaloids, 9 iridoids, 2 steroids, 2 phenolic acids and 1 coumarin	162

Table 2 Selection criteria of data acquisition and post-processing techniques in TCM qualitative analysis.

Technique		Advantage	Limitation	Scope of application
DIA	DIA	Full coverage of MS^1 and MS^2 information	Loss link between precursor and product ions	Acquisition of MS/MS information of minor or trace compounds
DDA	PIL	Screening of only (or preferred) target ions in the list	Miss of novel compounds with new sub-structures or substituted groups	Target compounds with predictable molecular weights and/or *m/z* values
	DE	Detection of less abundant co-eluting ions	Quality loss of MS^2 information of abundant ions	MS/MS information of both abundant and less abundant compounds in complex samples
	Mass tag	Selective screening of compounds that undergo in-source fragmentations	Not applicable to relatively stable compounds that cannot generate certain in-source fragmentation patterns	Compounds that can undergo certain in-source fragmentations with ISCID energy
	PIS	Requirements of little prior-knowledge about the exact structure of target compounds	Miss of compounds without selected product ions	Compounds with known specific product ions
	NLS	Targeted screening of modified or conjugated compounds that undergo neutral eliminations	Only compounds with certain neutral eliminations	Compounds with certain neutral eliminations
	MRM	High selectivity and sensitivity in screening target compounds	Low resolution and unable to screen unknown compounds without information of parent and product ions	Target compounds with known parent and product ions
Data post-processing	MDF	Selective screening of certain compounds or compound classes and remove of interferences	Requirements of a carefully-designed appropriate MDF region	Compounds with predictable mass range and mass defect range
	BS	Elimination of background signals, and exposure of target compounds	Poor selectivity of target compounds	Severe background interferences
	IPF	Specific screening of compounds with certain elements	Relatively narrow application scope	Compounds with distinct isotope pattern
	DIF	Rapid screening of target compounds that produce identical product ions	Miss of compounds without selected product ions	Compounds with known product ion
	NLF	Compounds that undergo neutral eliminations	Only compounds with certain neutral eliminations	Compounds with specific neutral loss
	MTSF	Rapid classification and characterization of unknown compounds	Need of MS^n information for both template compounds and target compounds	Target compounds sharing the same sub-structures and similar MS^n mass spectra as template compounds
	MN	Rapid classification and characterization of compounds without prior knowledge of the chemical structures	Manual interpretation of detailed structure information of unknown compounds based on the related known compounds	Compounds with MS/MS spectra
	SA	Rapid exploration of intrinsic relationships between compound structures and MS data	Further characterization and identification of compounds	Recognition of potential characteristic ions, discrimination of different structures and isomers
	DM	Alleviation of labor for mining structural information from large-scale datasets	Limited records in databases and requirements of further manual confirmation	Automatic search and annotation of compounds through on-line or in-house databases

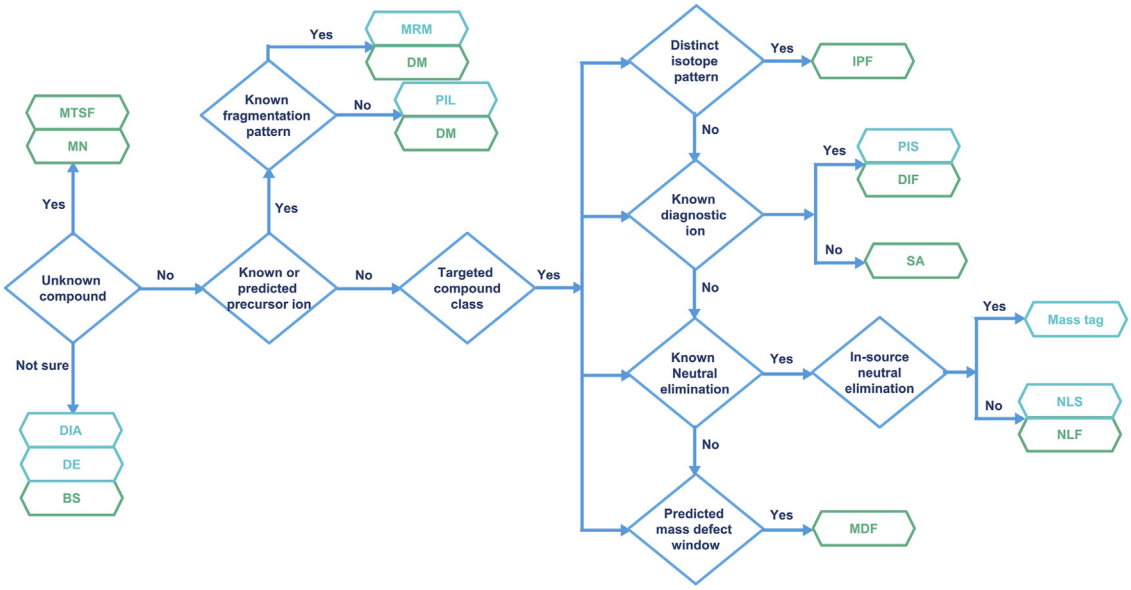


Figure 2 A decision tree that represents a possible way to select an LC−MS based data acquisition/post-processing strategy.

retention time (RT), and various retention time prediction strategies have also been proposed to improve the identification reliability, including logKow (log*P*) prediction[166], ACD/ChromGenius software prediction[167], and some quantitative structure retention relationship (QSRR) models such as linear solvation energy relationship, multiple linear regression[168], Kernel-based partial least-squares[169], artificial neural network[170], OPERA-RT model[171], etc. The other one is based on the MS information, including fragmentation patterns and intensity of product ions. Besides, SA methods have been applied for recognition of isomers[141], and some advanced techniques based on other physicochemical principles have also been explored. For example, the characteristic CE (CE_{50})[172] values and optimal CE (OCE) values which corresponded to the dissociation of chemical bonds were proved to be helpful in isomeric differentiation[173,174]. In addition, some cutting-edge instrumentations such as LC−infrared ion spectroscopy MS have been introduced and applied to distinguish closely related isomers[175−177].

In summary, based on the recent development and bottlenecks of LC−MS in TCM qualitative analysis, following suggestions are tentatively proposed for the future research. Firstly, because the bioactive components contained in TCM are usually secondary metabolites, which share similar substructures due to the common biosynthetic pathways, in order to achieve targeted characterization of different classes of compounds, more efficient and specific data acquisition and post-processing techniques based on their chemical characteristics, MS fragmentation patterns and other properties are in urgent demand. Next, resorting to the theory of big data analytics and rapid developments of computer science as well as mathematical statistics methods, exploration of high-throughput, high coverage and automatic data analysis strategies become possible. By developing computer and software-aided automatic data processing workflows, MS data interpretation efficiency of TCM complex systems will be dramatically improved. Last but not least, utilization of integrated analytical strategies, combing the advantages of various MS instrumentations, data acquisition and data post-processing techniques, are helpful for comprehensive and systematic analysis of chemical components of TCM.

6. Conclusions

TCM is gaining more and more global attentions because of its effectiveness in preventing and treating human diseases. LC−MS is one of the most powerful tools for TCM qualitative analysis. In this review, the basic principles and applications of a variety of advanced LC−MS-based data acquisition and post-processing techniques were summarized and illustrated for rapid and efficient characterization and identification of various chemical constituents in TCM. These data acquisition techniques include PIL, DE and mass tag based on MS information, PIS, NLS and MRM based on MS/MS information, and data post-processing techniques include MDF, BS, IPF, DIF, NLF, MTSF, MN, SA and DM, each one is indispensable with its unique superiority. Therefore, a brief introduction and applications of the integrated analytical strategies in TCM qualitative analysis were delivered. At last, discussions on the obstacles, possible solutions and future trends of TCM analysis were also provided, laying foundations for future research on the chemical basis of TCM.

Acknowledgments

This work was financially supported by National Key R&D Program of China (2018YFC1707900, 2019YFC1711000, and 2019YFC1711400), National Natural Science Foundation of China (82003938), State Key Program of National Natural Science Foundation of China (81530095), and Qi-Huang Scholar of National Traditional Chinese Medicine Leading Talents Support Program (2018).

Author contributions

Yang Yu wrote the paper. Changliang Yao edited the paper. De-an Guo conceived the review topic and edited the paper.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Luo L, Jiang JW, Wang C, Fitzgerald M, Hu WF, Zhou YM, et al. Analysis on herbal medicines utilized for treatment of COVID-19. *Acta Pharm Sin B* 2020;**10**:1192−204.
2. Harvey AL. Natural products in drug discovery. *Drug Discov Today* 2008;**13**:894−901.
3. Zhao J, Jiang P, Zhang WD. Molecular networks for the study of TCM pharmacology. *Briefings Bioinf* 2010;**11**:417−30.
4. Wu HF, Guo J, Chen SL, Liu X, Zhou Y, Zhang XP, et al. Recent developments in qualitative and quantitative analysis of phytochemical constituents and their metabolites using liquid chromatography−mass spectrometry. *J Pharmaceut Biomed Anal* 2013;**72**:267−91.
5. Jin HL, Liu YF, Guo ZM, Wang JX, Zhang XL, Wang CR, et al. Recent development in liquid chromatography stationary phases for separation of Traditional Chinese Medicine components. *J Pharmaceut Biomed Anal* 2016;**130**:336−46.
6. Holčapek M, Jirásko R, Lísa M. Recent developments in liquid chromatography−mass spectrometry and related techniques. *J Chromatogr A* 2012;**1259**:3−15.
7. Ma SG, Chowdhury SK. Data acquisition and data mining techniques for metabolite identification using LC coupled to high-resolution MS. *Bioanalysis* 2013;**5**:1285−97.
8. Stavrianidi A. A classification of liquid chromatography mass spectrometry techniques for evaluation of chemical composition and quality control of traditional medicines. *J Chromatogr A* 2020;**1609**: 460501.
9. Doerr A. DIA mass spectrometry. *Nat Methods* 2015;**12**. 35−35.
10. Zhou JT, Li YH, Chen X, Zhong LJ, Yin YX. Development of data-independent acquisition workflows for metabolomic analysis on a quadrupole-orbitrap platform. *Talanta* 2017;**164**:128−36.
11. Yang M, Sun JH, Lu ZQ, Chen GT, Guan SH, Liu X, et al. Phytochemical analysis of traditional Chinese medicine using liquid chromatography coupled with mass spectrometry. *J Chromatogr A* 2009;**1216**:2045−62.
12. Li M, Hou XF, Zhang J, Wang SC, Fu Q, He LC. Applications of HPLC/MS in the analysis of traditional Chinese medicines. *J Pharm Anal* 2011;**1**:81−91.
13. Steinmann D, Ganzera M. Recent advances on HPLC/MS in medicinal plant analysis. *J Pharmaceut Biomed Anal* 2011;**55**: 744−57.
14. Ganzera M, Sturm S. Recent advances on HPLC/MS in medicinal plant analysis—an update covering 2011−2016. *J Pharmaceut Biomed Anal* 2018;**147**:211−33.
15. Sentandreu E, Peris-Díaz MD, Sweeney SR, Chiou J, Muñoz N, Tiziani S. A survey of orbitrap all ion fragmentation analysis assessed by an R MetaboList package to study small-molecule metabolites. *Chromatographia* 2018;**81**:981−94.
16. Bateman KP, Castro-Perez J, Wrona M, Shockcor JP, Yu K, Oballa R, et al. MS^E with mass defect filtering for *in vitro* and *in vivo* metabolite identification. *Rapid Commun Mass Spectrom* 2007;**21**: 1485−96.
17. Gethings LA, Richardson K, Wildgoose J, Lennon S, Jarvis S, Bevan CL, et al. Lipid profiling of complex biological mixtures by liquid chromatography/mass spectrometry using a novel scanning quadrupole data-independent acquisition strategy. *Rapid Commun Mass Spectrom* 2017;**31**:1599−606.
18. Zhu XC, Chen YP, Subramanian R. Comparison of information-dependent acquisition, SWATH, and MS^{All} techniques in metabolite identification study employing ultrahigh-performance liquid chromatography−quadrupole time-of-flight mass spectrometry. *Anal Chem* 2014;**86**:1202−9.
19. van der Laan T, Boom I, Maliepaard J, Dubbelman AC, Harms AC, Hankemeier T. Data-independent acquisition for the quantification and identification of metabolites in plasma. *Metabolites* 2020;**10**: 514.
20. Hopfgartner G, Tonoli D, Varesio E. High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices. *Anal Bioanal Chem* 2012; **402**:2587−96.
21. Kaufmann A, Butcher P, Maden K, Walker S, Widmer M. Practical application of *in silico* fragmentation based residue screening with ion mobility high-resolution mass spectrometry. *Rapid Commun Mass Spectrom* 2017;**31**:1147−57.
22. Xia YG, Gong FQ, Guo XD, Song Y, Li CX, Liang J, et al. Rapid screening and characterization of triterpene saponins in *Acanthopanax senticosus* leaves *via* untargeted MS^{All} and SWATH techniques on a quadrupole time of flight mass spectrometry. *J Pharmaceut Biomed Anal* 2019;**170**:68−82.
23. Feng GF, Zheng Y, Sun Y, Liu S, Pi ZF, Song FR, et al. A targeted strategy for analyzing untargeted mass spectral data to identify lanostane-type triterpene acids in *Poria cocos* by integrating a scientific information system and liquid chromatography−tandem mass spectrometry combined with ion mobility spectrometry. *Anal Chim Acta* 2018;**1033**:87−99.
24. Wang HP, Chen C, Liu Y, Yang HJ, Wu HW, Xiao HB. Identification of the chemical constituents of Chinese medicine Yi-Xin-Shu capsule by molecular feature orientated precursor ion selection and tandem mass spectrometry structure elucidation. *J Separ Sci* 2015;**38**: 3687−95.
25. Fu LL, Ding H, Han LF, Jia L, Yang WZ, Zhang CX, et al. Simultaneously targeted and untargeted multicomponent characterization of Erzhi Pill by offline two-dimensional liquid chromatography/quadrupole-Orbitrap mass spectrometry. *J Chromatogr A* 2019;**1584**:87−96.
26. Pan HQ, Yang WZ, Yao CL, Shen Y, Zhang YB, Shi XJ, et al. Mass defect filtering-oriented classification and precursor ions list-triggered high-resolution mass spectrometry analysis for the discovery of indole alkaloids from Uncaria sinensis. *J Chromatogr A* 2017;**1516**:102−13.
27. Wang HD, Zhang CX, Zuo TT, Li WW, Jia L, Wang XY, et al. In-depth profiling, characterization, and comparison of the ginsenosides among three different parts (the root, stem leaf, and flower bud) of *Panax quinquefolius* L. by ultra-high performance liquid chromatography/quadrupole-Orbitrap mass spectrometry. *Anal Bioanal Chem* 2019;**411**:7817−29.
28. Yang M, Zhou Z, Yao S, Li SR, Yang WZ, Jiang BH, et al. Neutral loss ion mapping experiment combined with precursor mass list and dynamic exclusion for screening unstable malonyl glucoside conjugates. *J Am Soc Mass Spectrom* 2016;**27**:99−107.
29. Shi XJ, Yang WZ, Qiu S, Yao CL, Shen Y, Pan HQ, et al. An in-source multiple collision-neutral loss filtering based nontargeted metabolomics approach for the comprehensive analysis of malonyl-ginsenosides from *Panax ginseng, P. quinquefolius*, and *P. notoginseng*. *Anal Chim Acta* 2017;**952**:59−70.
30. Ge YW, Zhu S, Yoshimatsu K, Komatsu K. MS/MS similarity networking accelerated target profiling of triterpene saponins in *Eleutherococcus senticosus* leaves. *Food Chem* 2017;**227**:444−52.
31. Lu JX, Zhang CX, Hu Y, Zhang MH, Wang YN, Qian YX, et al. Application of multiple chemical and biological approaches for quality assessment of *Carthamus tinctorius* L. (safflower) by determining both the primary and secondary metabolites. *Phytomedicine* 2019;**58**:152826.
32. Li MR, Wang XY, Han LF, Jia L, Liu EW, Li Z, et al. Integration of multicomponent characterization, untargeted metabolomics and mass spectrometry imaging to unveil the holistic chemical transformations and key markers associated with wine steaming of Ligustri Lucidi Fructus. *J Chromatogr A* 2020;**1624**:461228.
33. Zhang JY, Wang ZJ, Zhang Q, Wang F, Ma Q, Lin ZZ, et al. Rapid screening and identification of target constituents using full scan-parent ions list-dynamic exclusion acquisition coupled to diagnostic product ions analysis on a hybrid LTQ-Orbitrap mass spectrometer. *Talanta* 2014;**124**:111−22.

34. Shen Y, Feng ZJ, Yang M, Zhou Z, Han SM, Hou JJ, et al. Rapid profiling of polymeric phenolic acids in *Salvia miltiorrhiza* by hybrid data-dependent/targeted multistage mass spectrometry acquisition based on expected compounds prediction and fragment ion searching. *J Separ Sci* 2018;**41**:1888−95.
35. Wang Y, Feng R, Wang R, Yang F, Li P, Wan JB. Enhanced MS/MS coverage for metabolite identification in LC−MS-based untargeted metabolomics by target-directed data dependent acquisition with time-staggered precursor ion list. *Anal Chim Acta* 2017;**992**:67−75.
36. Pan HQ, Zhou H, Miao S, Cao JY, Liu JM, Lan L, et al. An integrated approach for global profiling of multi-type constituents: comprehensive chemical characterization of Lonicerae Japonicae Flos as a case study. *J Chromatogr A* 2020;**1613**:460674.
37. Zhang ZQ. Automated precursor ion exclusion during LC−MS/MS data acquisition for optimal ion identification. *J Am Soc Mass Spectrom* 2012;**23**:1400−7.
38. Bi QR, Hou JJ, Yang M, Shen Y, Qi P, Feng RH, et al. A strategy combining higher energy C-trap dissociation with neutral loss- and product ion-based MS^n acquisition for global profiling and structure annotation of fatty acids conjugates. *J Am Soc Mass Spectrom* 2017;**28**:443−51.
39. Yan ZY, Caldwell GW, Maher N. Unbiased high-throughput screening of reactive metabolites on the linear ion trap mass spectrometer using polarity switch and mass tag triggered data-dependent acquisition. *Anal Chem* 2008;**80**:6410−22.
40. Hsiao HH, Urlaub H. Pseudo-neutral-loss scan for selective detection of phosphopeptides and *N*-glycopeptides using liquid chromatography coupled with a hybrid linear ion-trap/orbitrap mass spectrometer. *Proteomics* 2010;**10**:3916−21.
41. Shi XJ, Yang WZ, Huang Y, Hou JJ, Qiu S, Yao CL, et al. Direct screening of malonylginsenosides from nine Ginseng extracts by an untargeted profiling strategy incorporating in-source collision-induced dissociation, mass tag, and neutral loss scan on a hybrid linear ion-trap/Orbitrap mass spectrometer coupled to ultra-high performance liquid chromatography. *J Chromatogr A* 2018;**1571**:213−22.
42. Yang WZ, Shi XJ, Yao CL, Huang Y, Hou JJ, Han SM, et al. A novel neutral loss/product ion scan-incorporated integral approach for the untargeted characterization and comparison of the carboxyl-free ginsenosides from *Panax ginseng, Panax quinquefolius*, and *Panax notoginseng*. *J Pharm Biomed Anal* 2020;**177**:112813.
43. Marcos J, Pozo OJ. Current LC−MS methods and procedures applied to the identification of new steroid metabolites. *J Steroid Biochem Mol Biol* 2016;**162**:41−56.
44. Tang CM, Tang CX, Zhan W, Du J, Wang ZF, Peng XZ. Strategies for ascertaining the interference of phase II metabolites co-eluting with parent compounds using LC−MS−MS. *J Separ Sci* 2013;**36**:2584−92.
45. Chao HC, Chen GY, Hsu LC, Liao HW, Yang SY, Wang SY, et al. Using precursor ion scan of 184 with liquid chromatography−electrospray ionization-tandem mass spectrometry for concentration normalization in cellular lipidomic studies. *Anal Chim Acta* 2017;**971**:68−77.
46. Gao W, Liu XG, Liu L, Li P, Yang H. Targeted profiling and relative quantification of benzoyl diterpene alkaloids in *Aconitum* roots by using LC−MS/MS with precursor ion scan. *J Separ Sci* 2018;**41**:3515−26.
47. Shen Q, Lu YB, Dai ZY, Cheung HY. Precursor ion scan driven fast untargeted screening and semi-determination of caffeoylquinic acid derivatives in *Cynara scolymus* L. *Food Chem* 2015;**166**:442−7.
48. Song K, Ha IJ, Kim YS. A strategy for identification and structural characterization of oplopane- and bisabolane-type sesquiterpenoids from *Tussilago farfara* L. by multiple scan modes of mass spectrometry. *J Chromatogr A* 2019;**1602**:188−98.
49. Ji S, Wang Q, Qiao X, Guo HC, Yang YF, Bo T, et al. New triterpene saponins from the roots of *Glycyrrhiza yunnanensis* and their rapid screening by LC/MS/MS. *J Pharmaceut Biomed Anal* 2014;**90**:15−26.
50. Wu X, Zhu L, Ma J, Ye Y, Lin G. Adduct ion-targeted qualitative and quantitative analysis of polyoxypregnanes by ultra-high pressure liquid chromatography coupled with triple quadrupole mass spectrometry. *J Pharmaceut Biomed Anal* 2017;**145**:127−36.
51. Li Y, Pang T, Shi JL, Lu XP, Deng JH, Lin Q. Liquid chromatography with mass spectrometry method based two-step precursor ion scanning for the structural elucidation of flavonoids. *J Separ Sci* 2014;**37**:3067−73.
52. Li C, Yang JH, Tong X, Zhao C, He Y, Wan HT. Precursor ion scan enhanced rapid identification of the chemical constituents of Danhong injection by liquid chromatography−tandem mass spectrometry: an integrated strategy. *J Chromatogr A* 2019;**1602**:378−85.
53. Gan Y, Zhang Y, Li AQ, Song CW, Chen C, Xu Y, et al. A novel four-step approach for systematic identification of naphthoquinones in *Juglans cathayensis* dode using various scan functions of liquid chromatography−tandem mass spectrometry along with data mining strategies. *Phytochem Anal* 2015;**26**:413−22.
54. Yang W, Ye M, Liu M, Kong DZ, Shi R, Shi XW, et al. A practical strategy for the characterization of coumarins in Radix Glehniae by liquid chromatography coupled with triple quadrupole-linear ion trap mass spectrometry. *J Chromatogr A* 2010;**1217**:4587−600.
55. Song YL, Song QQ, Li J, Zhang N, Zhao YF, Liu X, et al. An integrated strategy to quantitatively differentiate chemome between *Cistanche deserticola* and *C. tubulosa* using high performance liquid chromatography-hybrid triple quadrupole-linear ion trap mass spectrometry. *J Chromatogr A* 2016;**1429**:238−47.
56. Ding L, Wang XY, Wang S, Yu JJ, Qin YQ, Zhang XB, et al. Systematic screening and characterization of glycosides in tobacco leaves by liquid chromatography with atmospheric pressure chemical ionization tandem mass spectrometry using neutral loss scan and product ion scan. *J Separ Sci* 2015;**38**:4029−35.
57. Yan ZX, Lin G, Ye Y, Wang YT, Yan R. A generic multiple reaction monitoring based approach for plant flavonoids profiling using a triple quadrupole linear ion trap mass spectrometry. *J Am Soc Mass Spectrom* 2014;**25**:955−65.
58. Yan ZX, Lin G, Ye Y, Wang YT, Yan R. Triterpenoid saponins profiling by adducts-targeted neutral loss triggered enhanced resolution and product ion scanning using triple quadrupole linear ion trap mass spectrometry. *Anal Chim Acta* 2014;**819**:56−64.
59. Kleinenkuhnen N, Büchel F, Gerlich SC, Kopriva S, Metzger S. A novel method for identification and quantification of sulfated flavonoids in plants by neutral loss scan mass spectrometry. *Front Plant Sci* 2019;**10**:885.
60. Jin Y, Xiao YS, Zhang FF, Xue XY, Xu Q, Liang XM. Systematic screening and characterization of flavonoid glycosides in *Carthamus tinctorius* L. by liquid chromatography/UV diode-array detection/electrospray ionization tandem mass spectrometry. *J Pharmaceut Biomed Anal* 2008;**46**:418−30.
61. Li S, Jin SN, Song CW, Jia SL, Zhang Y, Feng YL, et al. The strategy for establishment of the multiple reaction monitoring based characteristic chemical profile of triterpenes in Alismatis Rhizoma using two combined tandem mass spectrometers. *J Chromatogr A* 2017;**1524**:121−34.
62. Liu PW, Du YF, Zhang XW, Sheng XN, Shi XW, Zhao CC, et al. Rapid analysis of 27 components of *Isodon serra* by LC−ESI-MS−MS. *Chromatographia* 2010;**72**:265−73.
63. Jin SN, Song CW, Jia SL, Li S, Zhang Y, Chen C, et al. An integrated strategy for establishment of curcuminoid profile in turmeric using two LC−MS/MS platforms. *J Pharmaceut Biomed Anal* 2017;**132**:93−102.
64. Song YL, Zhang N, Shi SP, Li J, Zhao YF, Zhang Q, et al. Homolog-focused profiling of ginsenosides based on the integration of stepwise formate anion-to-deprotonated ion transition screening and scheduled multiple reaction monitoring. *J Chromatogr A* 2015;**1406**:136−44.

65. Yan ZX, Li TX, Lv P, Li X, Zhou C, Yang XH. Sensitive and reliable multianalyte quantitation of herbal medicine in rat plasma using dynamic triggered multiple reaction monitoring. *J Chromatogr B Analyt Technol Biomed Life Sci* 2013;**928**:22−31.
66. Liu Y, Uboh CE, Soma LR, Li XQ, Guan FY, You YW, et al. Efficient use of retention time for the analysis of 302 drugs in equine plasma by liquid chromatography−MS/MS with scheduled multiple reaction monitoring and instant library searching for doping control. *Anal Chem* 2011;**83**:6834−41.
67. Song QQ, Song YL, Zhang N, Li J, Jiang Y, Zhang KR, et al. Potential of hyphenated ultra-high performance liquid chromatography-scheduled multiple reaction monitoring algorithm for large-scale quantitative analysis of traditional Chinese medicines. *RSC Adv* 2015;**5**:57372−82.
68. Yao M, Ma L, Humphreys WG, Zhu M. Rapid screening and characterization of drug metabolites using a multiple ion monitoring-dependent MS/MS acquisition method on a hybrid triple quadrupole-linear ion trap mass spectrometer. *J Mass Spectrom* 2008;**43**:1364−75.
69. Song YL, Zhang N, Shi SP, Li J, Zhang Q, Zhao YF, et al. Large-scale qualitative and quantitative characterization of components in Shenfu injection by integrating hydrophilic interaction chromatography, reversed phase liquid chromatography, and tandem mass spectrometry. *J Chromatogr A* 2015;**1407**:106−18.
70. Yang WZ, Si W, Zhang JX, Yang M, Pan HQ, Wu J, et al. Selective and comprehensive characterization of the quinochalcone *C*-glycoside homologs in *Carthamus tinctorius* L. by offline comprehensive two-dimensional liquid chromatography/LTQ-Orbitrap MS coupled with versatile data mining strategies. *RSC Adv* 2016;**6**:495−506.
71. Yao CL, Yang WZ, Si W, Shen Y, Zhang NX, Chen HL, et al. An enhanced targeted identification strategy for the selective identification of flavonoid *O*-glycosides from *Carthamus tinctorius* by integrating offline two-dimensional liquid chromatography/linear ion-trap-Orbitrap mass spectrometry, high-resolution diagnostic product ions/neutral loss filtering and liquid chromatography-solid phase extraction-nuclear magnetic resonance. *J Chromatogr A* 2017;**1491**:87−97.
72. Sleno L. The use of mass defect in modern mass spectrometry. *J Mass Spectrom* 2012;**47**:226−36.
73. Zhu MS, Ma L, Zhang DL, Ray K, Zhao WP, Humphreys WG, et al. Detection and characterization of metabolites in biological matrices using mass defect filtering of liquid chromatography/high resolution mass spectrometry data. *Drug Metab Dispos* 2006;**34**:1722−33.
74. Zhang HY, Zhang DL, Ray K, Zhu MS. Mass defect filter technique and its applications to drug metabolite identification by high-resolution mass spectrometry. *J Mass Spectrom* 2009;**44**:999−1016.
75. Gu ZM, Wang LQ, Wu J. Mass defect filter—a new tool to expedite screening and dereplication of natural products and generate natural product profiles. *Nat Prod J* 2011;**1**:135−45.
76. Mortishire-Smith RJ, Castro-Perez JM, Yu K, Shockcor JP, Goshawk J, Hartshorn MJ, et al. Generic dealkylation: a tool for increasing the hit-rate of metabolite rationalization, and automatic customization of mass defect filters. *Rapid Commun Mass Spectrom* 2009;**23**:939−48.
77. Xie T, Liang Y, Hao HP, A JY, Xie L, Gong P, et al. Rapid identification of ophiopogonins and ophiopogonones in *Ophiopogon japonicus* extract with a practical technique of mass defect filtering based on high resolution mass spectrometry. *J Chromatogr A* 2012;**1227**:234−44.
78. Yan GL, Sun H, Sun WJ, Zhao L, Meng XC, Wang XJ. Rapid and global detection and characterization of *aconitum* alkaloids in Yin Chen Si Ni Tang, a traditional Chinese medical formula, by ultra performance liquid chromatography-high resolution mass spectrometry and automated data analysis. *J Pharmaceut Biomed Anal* 2010;**53**:421−31.
79. Pan HQ, Yang WZ, Zhang YB, Yang M, Feng RH, Wu WY, et al. An integrated strategy for the systematic characterization and discovery of new indole alkaloids from *Uncaria rhynchophylla* by UHPLC/-DAD/LTQ-Orbitrap-MS. *Anal Bioanal Chem* 2015;**407**:6057−70.
80. Tian JX, Tian Y, Xu L, Zhang J, Lu T, Zhang ZJ. Characterisation and identification of dihydroindole-type alkaloids from processed semen strychni by high-performance liquid chromatography coupled with electrospray ionisation ion trap time-of-flight mass spectrometry. *Phytochem Anal* 2014;**25**:36−44.
81. Zhang JY, Wang F, Zhang H, Lu JQ, Qiao YJ. Rapid identification of polymethoxylated flavonoids in traditional Chinese medicines with a practical strategy of stepwise mass defect filtering coupled to diagnostic product ions analysis based on a hybrid LTQ-orbitrap mass spectrometer. *Phytochem Anal* 2014;**25**:405−14.
82. Ren DB, Ran L, Yang C, Xu ML, Yi LZ. Integrated strategy for identifying minor components in complex samples combining mass defect, diagnostic ions and neutral loss information based on ultra-performance liquid chromatography-high resolution mass spectrometry platform: Folium Artemisiae Argyi as a case study. *J Chromatogr A* 2018;**1550**:35−44.
83. Lai CJS, Tan T, Zeng SL, Qi LW, Liu XG, Dong X, et al. An integrated high resolution mass spectrometric data acquisition method for rapid screening of saponins in *Panax notoginseng* (Sanqi). *J Pharmaceut Biomed Anal* 2015;**109**:184−91.
84. Pan HQ, Yao CL, Yang WZ, Yao S, Huang Y, Yb Zhang, et al. An enhanced strategy integrating offline two-dimensional separation and step-wise precursor ion list-based raster-mass defect filter: characterization of indole alkaloids in five botanical origins of Uncariae Ramulus Cum Unicis as an exemplary application. *J Chromatogr A* 2018;**1563**:124−34.
85. Keller BO, Sui J, Young AB, Whittal RM. Interferences and contaminants encountered in modern mass spectrometry. *Anal Chim Acta* 2008;**627**:71−81.
86. Windig W, Phalp JM, Payne AW. A noise and background reduction method for component detection in liquid chromatography−mass spectrometry. *Anal Chem* 1996;**68**:3602−6.
87. Wang CH, Zhang JL, Wu CS, Wang Z. A multiple-dimension liquid chromatography coupled with mass spectrometry data strategy for the rapid discovery and identification of unknown compounds from a Chinese herbal formula (Er-xian decoction). *J Chromatogr A* 2017;**1518**:59−69.
88. Feng JJ, Yu PF, Zhou Q, Tian ZH, Sun MJ, Li XL, et al. An integrated data filtering and identification strategy for rapid profiling of chemical constituents, with Arnebiae Radix as an example. *J Chromatogr A* 2020;**1629**:461496.
89. Zeng SL, Duan L, Chen BZ, Li P, Liu EH. Chemicalome and metabolome profiling of polymethoxylated flavonoids in Citri Reticulatae Pericarpium based on an integrated strategy combining background subtraction and modified mass defect filter in a Microsoft Excel Platform. *J Chromatogr A* 2017;**1508**:106−20.
90. Guo J, Zhang M, Elmore CS, Vishwanathan K. An integrated strategy for *in vivo* metabolite profiling using high-resolution mass spectrometry based data processing techniques. *Anal Chim Acta* 2013;**780**:55−64.
91. Zhu PJ, Tong W, Alton K, Swapan C. An accurate-mass-based spectral-averaging isotope-pattern-filtering algorithm for extraction of drug metabolites possessing a distinct isotope pattern from LC−MS data. *Anal Chem* 2009;**81**:5910−7.
92. Yang M, Zhou Z, Guo DA. A strategy for fast screening and identification of sulfur derivatives in medicinal *Pueraria* species based on the fine isotopic pattern filtering method using ultra-high-resolution mass spectrometry. *Anal Chim Acta* 2015;**894**:44−53.
93. Xia YG, Liang J, Guo XD, Sun HM, Kuang HX. UPLC-QTOF-MS^E-based diagnostic product ion filtering to unveil unstable C_6-C_2 glucoside conjugates in Forsythia suspensa. *J Mass Spectrom* 2017;**52**:848−59.
94. Liao M, Li AQ, Chen C, Ouyang H, Zhang Y, Xu Y, et al. Systematic identification of shikonins and shikonofurans in medicinal Zicao species using ultra-high performance liquid chromatography quadrupole time of flight tandem mass spectrometry combined with a data mining strategy. *J Chromatogr A* 2015;**1425**:158−72.

95. Ma BL, Yang SL, Li JM, Ouyang H, He MZ, Feng YL, et al. A four-step filtering strategy based on ultra-high-performance liquid chromatography coupled to quadrupole-time-of-flight tandem mass spectrometry for comprehensive profiling the major chemical constituents of Akebiae Fructus. *Rapid Commun Mass Spectrom* 2019;**33**:1464−74.
96. Cheng XL, Wan JY, Li P, Qi LW. Ultrasonic/microwave assisted extraction and diagnostic ion filtering strategy by liquid chromatography-quadrupole time-of-flight mass spectrometry for rapid characterization of flavonoids in Spatholobus suberectus. *J Chromatogr A* 2011;**1218**:5774−86.
97. Qiao X, Li R, Song W, Miao WJ, Liu J, Chen HB, et al. A targeted strategy to analyze untargeted mass spectral data: rapid chemical profiling of *Scutellaria baicalensis* using ultra-high performance liquid chromatography coupled with hybrid quadrupole orbitrap mass spectrometry and key ion filtering. *J Chromatogr A* 2016;**1441**:83−95.
98. Tchoumtchoua J, Njamen D, Mbanya JC, Skaltsounis AL, Halabalaki M. Structure-oriented UHPLC−LTQ Orbitrap-based approach as a dereplication strategy for the identification of isoflavonoids from *Amphimas pterocarpoides* crude extract. *J Mass Spectrom* 2013;**48**:561−75.
99. Qi LW, Wang HY, Zhang H, Wang CZ, Li P, Yuan CS. Diagnostic ion filtering to characterize ginseng saponins by rapid liquid chromatography with time-of-flight mass spectrometry. *J Chromatogr A* 2012;**1230**:93−9.
100. Wan JY, Fan Y, Yu QT, Ge YZ, Yan CP, Alolga RN, et al. Integrated evaluation of malonyl ginsenosides, amino acids and polysaccharides in fresh and processed ginseng. *J Pharmaceut Biomed Anal* 2015;**107**:89−97.
101. Xie SL, Shi YY, Wang Y, Wu CY, Liu WY, Feng F, et al. Systematic identification and quantification of tetracyclic monoterpenoid oxindole alkaloids in *Uncaria rhynchophylla* and their fragmentations in Q-TOF-MS spectra. *J Pharmaceut Biomed Anal* 2013;**81**−**82**:56−64.
102. Li AQ, Du ZF, Liao M, Feng YL, Ruan HL, Jiang HL. Discovery and characterisation of lycorine-type alkaloids in *Lycoris* spp. (Amaryllidaceae) using UHPLC−QTOF−MS. *Phytochem Anal* 2019;**30**:268−77.
103. Huang C, Xu QM, Chen C, Song CW, Xu Y, Xiang Y, et al. The rapid discovery and identification of physalins in the calyx of *Physalis alkekengi* L.var.*franchetii* (Mast.) Makino using ultra-high performance liquid chromatography-quadrupole time of flight tandem mass spectrometry together with a novel three-step data mining strategy. *J Chromatogr A* 2014;**1361**:139−52.
104. Li JM, Wen Q, Feng YL, Zhang J, Luo Y, Tan T. Characterization of the multiple chemical components of Glechomae Herba using ultra high performance liquid chromatography coupled to quadrupole-time-of-flight tandem mass spectrometry with diagnostic ion filtering strategy. *J Separ Sci* 2019;**42**:1312−22.
105. Liu Y, Li L, Xiao YQ, Yao JQ, Li PY, Yu DR, et al. Global metabolite profiling and diagnostic ion filtering strategy by LC−QTOF MS for rapid identification of raw and processed pieces of *Rheum palmatum* L. *Food Chem* 2016;**192**:531−40.
106. Zhang H, Zheng D, Li HH, Wang H, Tan HS, Xu HX. Diagnostic filtering to screen polycyclic polyprenylated acylphloroglucinols from *Garcinia* oblongifolia by ultrahigh performance liquid chromatography coupled with ion mobility quadrupole time-of-flight mass spectrometry. *Anal Chim Acta* 2016;**912**:85−96.
107. Zheng CN, Hao HP, Wang X, Wu XL, Wang GJ, Sang GW, et al. Diagnostic fragment-ion-based extension strategy for rapid screening and identification of serial components of homologous families contained in traditional Chinese medicine prescription using high-resolution LC−ESI−IT−TOF/MS: Shengmai injection as an example. *J Mass Spectrom* 2009;**44**:230−44.
108. Zeng SL, Liu XG, Lai CJS, Liu EH, Li P. Diagnostic ion filtering strategy for chemical characterization of Guge Fengtong Tablet with high-performance liquid chromatography coupled with electrospray ionization quadrupole time-of-flight tandem mass spectrometry. *Chin J Nat Med* 2015;**13**. 390−400.
109. Pang HQ, An HM, Yang H, Wu SQ, Fan JL, Mi L, et al. Comprehensive chemical profiling of Yindan Xinnaotong soft capsule and its neuroprotective activity evaluation *in vitro*. *J Chromatogr A* 2019;**1601**:288−99.
110. Dai WD, Yin PY, Zeng ZD, Kong HW, Tong HW, Xu ZL, et al. Nontargeted modification-specific metabolomics study based on liquid chromatography−high-resolution mass spectrometry. *Anal Chem* 2014;**86**:9146−53.
111. Brink A, Fontaine F, Marschmann M, Steinhuber B, Cece EN, Zamora I, et al. Post-acquisition analysis of untargeted accurate mass quadrupole time-of-flight MS^E data for multiple collision-induced neutral losses and fragment ions of glutathione conjugates. *Rapid Commun Mass Spectrom* 2014;**28**:2695−703.
112. Zhang J, Huang ZH, Qiu XH, Yang YM, Zhu DY, Xu W. Neutral fragment filtering for rapid identification of new diester-diterpenoid alkaloids in roots of *Aconitum carmichaeli* by ultra-high-pressure liquid chromatography coupled with linear ion trap-orbitrap mass spectrometry. *PLoS One* 2012;**7**:e52352.
113. Jin Y, Wu CS, Zhang JL, Li YF. A new strategy for the discovery of epimedium metabolites using high-performance liquid chromatography with high resolution mass spectrometry. *Anal Chim Acta* 2013;**768**:111−7.
114. Jia ZX, Wu CS, Jin HT, Zhang JL. Identification of the chemical components of *Saussurea involucrata* by high-resolution mass spectrometry and the mass spectral trees similarity filter technique. *Rapid Commun Mass Spectrom* 2014;**28**:2237−51.
115. Liu LH, Zhang JY, Zheng BJ, Guan Y, Wang LT, Chen L, et al. Rapid characterization of chlorogenic acids in *Duhaldea nervosa* based on ultra-high-performance liquid chromatography-linear trap quadropole-Orbitrap-mass spectrometry and mass spectral trees similarity filter technique. *J Separ Sci* 2018;**41**:1764−74.
116. Zhang JY, Wang ZJ, Li Y, Liu Y, Cai W, Li C, et al. A strategy for comprehensive identification of sequential constituents using ultra-high-performance liquid chromatography coupled with linear ion trap-Orbitrap mass spectrometer, application study on chlorogenic acids in Flos Lonicerae Japonicae. *Talanta* 2016;**147**:16−27.
117. Wang CH, Wu CS, Qin HL, Zhang JL. Rapid discovery and identification of 68 compounds in the active fraction from Xiao-Xu-Ming decoction (XXMD) by HPLC−HRMS and MTSF technique. *Chin Chem Lett* 2014;**25**:1648−52.
118. Garg N, Kapono CA, Lim YW, Koyama N, Vermeij MJA, Conrad D, et al. Mass spectral similarity for untargeted metabolomics data analysis of complex mixtures. *Int J Mass Spectrom* 2015;**377**:719−27.
119. Wang MX, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat Biotechnol* 2016;**34**:828−37.
120. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**:2498−504.
121. Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, et al. Mass spectral molecular networking of living microbial colonies. *Proc Natl Acad Sci U S A* 2012;**109**:E1743−52.
122. Quinn RA, Nothias LF, Vining O, Meehan M, Esquenazi E, Dorrestein PC. Molecular networking as a drug discovery, drug metabolism, and precision medicine strategy. *Trends Pharmacol Sci* 2017;**38**:143−54.
123. Allard PM, Péresse T, Bisson J, Gindro K, Marcourt L, Pham VC, et al. Integration of molecular networking and *in-silico* MS/MS fragmentation for natural products dereplication. *Anal Chem* 2016;**88**:3317−23.
124. Yang JY, Sanchez LM, Rath CM, Liu X, Boudreau PD, Bruns N, et al. Molecular networking as a dereplication strategy. *J Nat Prod* 2013;**76**:1686−99.

125. Crusemann M, O'Neill EC, Larson CB, Melnik AV, Floros DJ, da Silva RR, et al. Prioritizing natural product diversity in a collection of 146 bacterial strains based on growth and extraction protocols. *J Nat Prod* 2017;**80**:588−97.
126. Purves K, Macintyre L, Brennan D, Hreggviðsson Gó, Kuttner E, Ásgeirsdóttir ME, et al. Using molecular networking for microbial secondary metabolite bioprospecting. *Metabolites* 2016;**6**:2.
127. Trivella DBB, de Felicio R. The tripod for bacterial natural product discovery: genome mining, silent pathway induction, and mass spectrometry-based molecular networking. *mSystems* 2018;**3**. e00160−17.
128. Philippus AC, Zatelli GA, Wanke T, Gabriela de A, Barros M, Kami SA, Lhullier C, et al. Molecular networking prospection and characterization of terpenoids and C15-acetogenins in Brazilian seaweed extracts. *RSC Adv* 2018;**8**:29654−61.
129. Grauso L, Yegdaneh A, Sharifi M, Mangoni A, Zolfaghari B, Lanzotti V. Molecular networking-based analysis of cytotoxic saponins from sea cucumber Holothuria atra. *Mar Drugs* 2019;**17**:86.
130. Ding CYG, Pang LM, Liang ZX, Goh KKK, Glukhov E, Gerwick WH, et al. MS/MS-based molecular networking approach for the detection of aplysiatoxin-related compounds in environmental marine cyanobacteria. *Mar Drugs* 2018;**16**:505.
131. Hou XM, Li YY, Shi YW, Fang YW, Chao R, Gu YC, et al. Integrating molecular networking and ^{1}H NMR to target the isolation of chrysogeamides from a library of marine-derived *Penicillium* fungi. *J Org Chem* 2019;**84**:1228−37.
132. Olivon F, Apel C, Retailleau P, Allard PM, Wolfender JL, Touboul D, et al. Searching for original natural products by molecular networking: detection, isolation and total synthesis of chloroaustralasines. *Organic Chemistry Frontiers* 2018;**5**:2171−8.
133. Caesar LK, Kellogg JJ, Kvalheim OM, Cech RA, Cech NB. Integration of biochemometrics and molecular networking to identify antimicrobials in Angelica keiskei. *Planta Med* 2018;**84**:721−8.
134. Raheem DJ, Tawfike AF, Abdelmohsen UR, Edrada-Ebel R, Fitzsimmons-Thoss V. Application of metabolomics and molecular networking in investigating the chemical profile and anti-trypanosomal activity of British bluebells (*Hyacinthoides non-scripta*). *Sci Rep* 2019;**9**:2547.
135. de Oliveira GG, Carnevale Neto F, Demarque DP, de Sousa Pereira-Junior JA, Sampaio Peixoto Filho RC, de Melo SJ, et al. Dereplication of flavonoid glycoconjugates from *Adenocalymma imperatoris-maximilianii* by untargeted tandem mass spectrometry-based molecular networking. *Planta Med* 2017;**83**:636−46.
136. Abbas-Mohammadi M, Moridi Farimani M, Salehi P, Nejad Ebrahimi S, Sonboli A, Kelso C, et al. Acetylcholinesterase-inhibitory activity of Iranian plants: combined HPLC/bioassay-guided fractionation, molecular networking and docking strategies for the dereplication of active compounds. *J Pharmaceut Biomed Anal* 2018;**158**:471−9.
137. Wei WL, Hou JJ, Yao CL, Bi QR, Wang X, Li ZW, et al. A high-efficiency strategy integrating offline two-dimensional separation and data post-processing with dereplication: characterization of bufadienolides in *Venenum Bufonis* as a case study. *J Chromatogr A* 2019;**1603**:179−89.
138. Lei HB, Zhang YH, Ye J, Cheng TF, Liang YL, Zu XP, et al. A comprehensive quality evaluation of Fuzi and its processed product through integration of UPLC−QTOF/MS combined MS/MS-based mass spectral molecular networking with multivariate statistical analysis and HPLC−MS/MS. *J Ethnopharmacol* 2021;**266**:113455.
139. Xu LL, Shang ZP, Bo T, Sun L, Guo QL, Qiao X, et al. Rapid quantitation and identification of the chemical constituents in Danhong Injection by liquid chromatography coupled with orbitrap mass spectrometry. *J Chromatogr A* 2019;**1606**:460378.
140. Qiao X, Lin XH, Ji S, Zhang ZX, Bo T, Guo DA, et al. Global profiling and novel structure discovery using multiple neutral loss/precursor ion scanning combined with substructure recognition and statistical analysis (MNPSS): characterization of terpene-conjugated curcuminoids in *Curcuma longa* as a case study. *Anal Chem* 2016;**88**:703−10.
141. Xue ZZ, Lai CJS, Kang LP, Kotani A, Hakamata H, Jing ZX, et al. Profiling and isomer recognition of phenylethanoid glycosides from *Magnolia officinalis* based on diagnostic/holistic fragment ions analysis coupled with chemometrics. *J Chromatogr A* 2020;**1611**:460583.
142. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, et al. MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom* 2010;**45**:703−14.
143. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, et al. METLIN: a metabolite mass spectral database. *Ther Drug Monit* 2005;**27**:747−51.
144. Fabresse N, Larabi IA, Stratton T, Mistrik R, Pfau G, Lorin de la Grandmaison G, et al. Development of a sensitive untargeted liquid chromatography-high resolution mass spectrometry screening devoted to hair analysis through a shared MS*2* spectra database: a step toward early detection of new psychoactive substances. *Drug Test Anal* 2019;**11**:697−708.
145. Barlow DJ, Buriani A, Ehrman T, Bosisio E, Eberini I, Hylands PJ. *In-silico* studies in Chinese herbal medicines' research: evaluation of *in-silico* methodologies and phytochemical data sources, and a review of research to date. *J Ethnopharmacol* 2012;**140**:526−34.
146. Blazenovic I, Kind T, Ji J, Fiehn O. Software tools and approaches for compound identification of LC−MS/MS data in metabolomics. *Metabolites* 2018;**8**:31.
147. Gil de la Fuente A, Grace Armitage E, Otero A, Barbas C, Godzien J. Differentiating signals to make biological sense—a guide through databases for MS-based non-targeted metabolomics. *Electrophoresis* 2017;**38**:2242−56.
148. Johnson SR, Lange BM. Open-access metabolomics databases for natural product research: present capabilities and future potential. *Front Bioeng Biotechnol* 2015;**3**:22.
149. Zhang FX, Li M, Qiao LR, Yao ZH, Li C, Shen XY, et al. Rapid characterization of *Ziziphi Spinosae Semen* by UPLC/Qtof MS with novel informatics platform and its application in evaluation of two seeds from *Ziziphus* species. *J Pharmaceut Biomed Anal* 2016;**122**:59−80.
150. Deng L, Shi AM, Liu HZ, Meruva N, Liu L, Hu H, et al. Identification of chemical ingredients of peanut stems and leaves extracts using UPLC−QTOF−MS coupled with novel informatics UNIFI platform. *J Mass Spectrom* 2016;**51**:1157−67.
151. Zhang H, Jiang JM, Zheng D, Yuan M, Wang ZY, Zhang HM, et al. A multidimensional analytical approach based on time-decoupled online comprehensive two-dimensional liquid chromatography coupled with ion mobility quadrupole time-of-flight mass spectrometry for the analysis of ginsenosides from white and red ginsengs. *J Pharmaceut Biomed Anal* 2019;**163**:24−33.
152. Qiu S, Yang WZ, Shi XJ, Yao CL, Yang M, Liu X, et al. A green protocol for efficient discovery of novel natural compounds: characterization of new ginsenosides from the stems and leaves of *Panax ginseng* as a case study. *Anal Chim Acta* 2015;**893**:65−76.
153. Qiu S, Yang WZ, Yao CL, Qiu ZD, Shi XJ, Zhang JX, et al. Non-targeted metabolomic analysis and "commercial-homophyletic" comparison-induced biomarkers verification for the systematic chemical differentiation of five different parts of *Panax ginseng*. *J Chromatogr A* 2016;**1453**:78−87.
154. Yao CL, Pan HQ, Wang H, Yao S, Yang WZ, Hou JJ, et al. Global profiling combined with predicted metabolites screening for discovery of natural compounds: characterization of ginsenosides in the leaves of *Panax notoginseng* as a case study. *J Chromatogr A* 2018;**1538**:34−44.
155. Ma XL, Guo XY, Song YL, Qiao LR, Wang WG, Zhao MB, et al. An integrated strategy for global qualitative and quantitative profiling of traditional Chinese medicine formulas: Baoyuan Decoction as a case. *Sci Rep* 2016;**6**:38379.
156. Kang LP, Huang YY, Zhan ZL, Liu DH, Peng HS, Nan TG, et al. Structural characterization and discrimination of the *Paris polyphylla*

var. *yunnanensis* and *Paris vietnamensis* based on metabolite profiling analysis. *J Pharmaceut Biomed Anal* 2017;**142**:252−61.
157. Shi XJ, Yang WZ, Qiu S, Hou JJ, Wu WY, Guo DA. Systematic profiling and comparison of the lipidomes from *Panax ginseng, P. quinquefolius*, and *P. notoginseng* by ultrahigh performance supercritical fluid chromatography/high-resolution mass spectrometry and ion mobility-derived collision cross section measurement. *J Chromatogr A* 2018;**1548**:64−75.
158. Zhang JY, Li C, Che YY, Wu JR, Wang ZJ, Cai W, et al. LTQ-Orbitrap-based strategy for traditional Chinese medicine targeted class discovery, identification and herbomics research: a case study on phenylethanoid glycosides in three different species of Herba Cistanches. *RSC Adv* 2015;**5**:80816−28.
159. Chen XF, Wu YL, Chen C, Gu YQ, Zhu CY, Wang SP, et al. Identifying potential anti-COVID-19 pharmacological components of traditional Chinese medicine Lianhuaqingwen capsule based on human exposure and ACE2 biochromatography screening. *Acta Pharm Sin B* 2020;**11**:222−36.
160. Luo SF, Guo LB, Sheng CM, Zhao YM, Chen L, Li CF, et al. Rapid identification and isolation of neuraminidase inhibitors from mock-strawberry (*Duchesnea indica* Andr.) based on ligand fishing combined with HR-ESI-Q-TOF-MS. *Acta Pharm Sin B* 2020;**10**:1846−55.
161. Liang Y, Wu JL, Leung EL, Zhou H, Liu ZQ, Yan GY, et al. Identification of oxygenated fatty acid as a side chain of lipo-alkaloids in *Aconitum carmichaelii* by UHPLC−Q-TOF-MS and a database. *Molecules* 2016;**21**:437.
162. Liu YC, Xiao S, Yang K, Ling L, Sun ZL, Liu ZY. Comprehensive identification and structural characterization of target components from *Gelsemium elegans* by high-performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry based on accurate mass databases combined with MS/MS spectra. *J Mass Spectrom* 2017;**52**:378−96.
163. Zhang ML, Sun JH, Chen P. FlavonQ: an automated data processing tool for profiling flavone and flavonol glycosides with ultra-high-performance liquid chromatography-diode array detection-high resolution accurate mass-mass spectrometry. *Anal Chem* 2015;**87**:9974−81.
164. Qiu F, Fine DD, Wherritt DJ, Lei Z, Sumner LW. PlantMAT: a metabolomics tool for predicting the specialized metabolic potential of a system and for large-scale metabolite identifications. *Anal Chem* 2016;**88**:11373−83.
165. Zhou WJ, Liu YM, Wang JX, Guo ZM, Shen AJ, Liu YF, et al. Application of two-dimensional liquid chromatography in the separation of traditional Chinese medicine. *J Separ Sci* 2020;**43**:87−104.
166. Bade R, Bijlsma L, Sancho JV, Hernández F. Critical evaluation of a simple retention time predictor based on LogKow as a complementary tool in the identification of emerging contaminants in water. *Talanta* 2015;**139**:143−9.
167. Tyrkkö E, Pelander A, Ojanperä I. Prediction of liquid chromatographic retention for differentiation of structural isomers. *Anal Chim Acta* 2012;**720**:142−8.
168. Zhang QQ, Huo MQ, Zhang YL, Qiao YJ, Gao XY. A strategy to improve the identification reliability of the chemical constituents by high-resolution mass spectrometry-based isomer structure prediction combined with a quantitative structure retention relationship analysis: phthalide compounds in Chuanxiong as a test case. *J Chromatogr A* 2018;**1552**:17−28.
169. Falchi F, Bertozzi SM, Ottonello G, Ruda GF, Colombano G, Fiorelli C, et al. Kernel-based, partial least squares quantitative structure-retention relationship model for UPLC retention time prediction: a useful tool for metabolite identification. *Anal Chem* 2016;**88**:9510−7.
170. Munro K, Miller TH, Martins CP, Edge AM, Cowan DA, Barron LP. Artificial neural network modelling of pharmaceutical residue retention times in wastewater extracts using gradient liquid chromatography-high resolution mass spectrometry data. *J Chromatogr A* 2015;**1396**:34−44.
171. McEachran AD, Mansouri K, Newton SR, Beverly BEJ, Sobus JR, Williams AJ. A comparison of three liquid chromatography (LC) retention time prediction models. *Talanta* 2018;**182**:371−9.
172. Nagy T, Kuki Á Antal B, Nagy L, Purgel M, Sipos A, et al. Chiral differentiation of the noscapine and hydrastine stereoisomers by electrospray ionization tandem mass spectrometry. *J Mass Spectrom* 2015;**50**:240−6.
173. Song YL, Song QQ, Liu Y, Li J, Wan JB, Wang YT, et al. Integrated work-flow for quantitative metabolome profiling of plants, Peucedani Radix as a case. *Anal Chim Acta* 2017;**953**:40−7.
174. Song QQ, Li J, Huo HX, Cao Y, Wang YT, Song YL, et al. Retention time and optimal collision energy advance structural annotation relied on LC−MS/MS: an application in metabolite identification of an antidementia agent namely echinacoside. *Anal Chem* 2019;**91**:15040−8.
175. Martens J, Berden G, Gebhardt CR, Oomens J. Infrared ion spectroscopy in a modified quadrupole ion trap mass spectrometer at the FELIX free electron laser laboratory. *Rev Sci Instrum* 2016;**87**:103108.
176. Kranenburg RF, van Geenen F, Berden G, Oomens J, Martens J, van Asten AC. Mass-spectrometry-based identification of synthetic drug isomers using infrared ion spectroscopy. *Anal Chem* 2020;**92**:7282−8.
177. Martens J, van Outersterp RE, Vreeken RJ, Cuyckens F, Coene KLM, Engelke UF, et al. Infrared ion spectroscopy: new opportunities for small-molecule identification in mass spectrometry—a tutorial perspective. *Anal Chim Acta* 2020;**1093**:1−15.