Journal of Pharmaceutical Analysis 14 (2024) 100994

Contents lists available at ScienceDirect

Journal of Pharmaceutical Analysis

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

Original article

Software-aided efficient identification of the components of compound formulae and their metabolites in rats by UHPLC/IM-QTOF-MS and an in-house high-definition MS^2 library: Sishen formula as a case

Lili Hong [a, b, 1], Wei Wang [a, b, 1], Shiyu Wang [a, b, 1], Wandi Hu [a, b], Yuyang Sha [c], Xiaoyan Xu [a, b], Xiaoying Wang [d], Kefeng Li [c], Hongda Wang [a, b, d, *], Xiumei Gao [a, b, d, **], De-an Guo [a, b, e, ***], Wenzhi Yang [a, b, d, ****]

[a] National Key Laboratory of Chinese Medicine Modernization, State Key Laboratory of Component-based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China
[b] Haihe Laboratory of Modern Chinese Medicine, Tianjin, 301617, China
[c] Centre for Artificial Intelligence Driven Drug Discovery, Faculty of Applied Sciences, Macao Polytechnic University, Macao SAR, Rua de Luís Gonzaga Gomes, Macao, 999078, China
[d] Key Laboratory of Pharmacology of Traditional Chinese Medical Formulae, Ministry of Education, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China
[e] 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

ARTICLE INFO

Article history:
Received 11 February 2024
Received in revised form
31 March 2024
Accepted 29 April 2024
Available online 3 May 2024

Keywords:
Ultra-high performance liquid chromatography/ion-mobility quadrupole time-of-flight mass spectrometry (UHPLC/IM-QTOF-MS)
Hybrid scan
High-definition MS^2 spectral library
Sishen formula

ABSTRACT

Identifying the compound formulae-related xenobiotics in bio-samples is full of challenges. Conventional strategies always exhibit the insufficiencies in overall coverage, analytical efficiency, and degree of automation, and the results highly rely on the personal knowledge and experience. The goal of this work was to establish a software-aided approach, by integrating ultra-high performance liquid chromatography/ion-mobility quadrupole time-of-flight mass spectrometry (UHPLC/IM-QTOF-MS) and in-house high-definition MS^2 library, to enhance the identification of prototypes and metabolites of the compound formulae *in vivo*, taking Sishen formula (SSF) as a template. Seven different MS^2 acquisition methods were compared, which demonstrated the potency of a hybrid scan approach (namely high-definition data-independent/data-dependent acquisition (HDDIDDA)) in the identification precision, MS^1 coverage, and MS^2 spectra quality. The HDDIDDA data for 55 reference compounds, four component drugs, and SSF, together with the rat bio-samples (e.g., plasma, urine, feces, liver, and kidney), were acquired. Based on the UNIFI™ platform (Waters), the efficient data processing workflows were established by combining mass defect filtering (MDF)-induced classification, diagnostic product ions (DPIs), and neutral loss filtering (NLF)-dominated structural confirmation. The high-definition MS^2 spectral libraries, dubbed *in vitro*-SSF and *in vivo*-SSF, were elaborated, enabling the efficient and automatic identification of SSF-associated xenobiotics in diverse rat bio-samples. Consequently, 118 prototypes and 206 metabolites of SSF were identified, with the identification rate reaching 80.51% and 79.61%, respectively. The metabolic pathways mainly involved the oxidation, reduction, hydrolysis, sulfation, methylation, demethylation, acetylation, glucuronidation, and the combined reactions. Conclusively, the

Peer review under responsibility of Xi'an Jiaotong University.
* Corresponding author. National Key Laboratory of Chinese Medicine Modernization, State Key Laboratory of Component-based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China.
** Corresponding author. National Key Laboratory of Chinese Medicine Modernization, State Key Laboratory of Component-based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China.
*** Corresponding author. 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
**** Corresponding author. National Key Laboratory of Chinese Medicine Modernization, State Key Laboratory of Component-based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China.
E-mail addresses: 17862987156@163.com (H. Wang), gaoxiumei@tjutcm.edu.cn (X. Gao), daguo@simm.ac.cn (D.-a. Guo), wzyang0504@tjutcm.edu.cn (W. Yang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jpha.2024.100994

proposed strategy can drive the identification of compound formulae-related xenobiotics *in vivo* in an intelligent manner.

1. Introduction

Metabolic profiles analysis demonstrates an effective means to explore the material basis of traditional Chinese medicine (TCM), known as the xenobiotics (involving the absorbed prototypes and the transformed metabolites) which have the potential to play a major role in the precision medicine [1,2]. Unfortunately, the delineation of the metabolic profiles of TCM *in vivo* has been an enduring obstacle to the researchers, due to the complex composition of TCM itself, the interference from the endogenous substances, and diversified biotransformation pathways [3]. Therefore, there has been an increasing demand for the accurate and efficient identification strategy that can facilitate the comprehensive identification of TCM-related xenobiotics in the bio-samples.

Liquid chromatography/high-resolution mass spectrometry (LC-HRMS), in combination with versatile post-acquisition processing vehicles, is mostly preferable for the precise analysis of drug-related xenobiotics in the bio-samples. It shows the advantages of high throughput, high sensitivity, and high coverage on the compounds' polarity [4,5]. New trends have emerged in pursuit of the fast scan speed, fit-for-purpose data acquisition and data processing, and high separation efficiency and dimension, in the multi-component characterization of TCM and the related xenobiotics in bio-samples using LC-HRMS [6–10]. In essence, the versatile MS^n ($n \geq 2$) data acquisition strategies can be divided into data-dependent acquisition (DDA) and data-independent acquisition (DIA; such as MS^E and sequential window acquisition of all theoretical mass spectra (SWATH)). Comparatively, the MS^n spectra obtained by DDA are of high quality and easy to interpret, but the coverage on the target compound analogs may be severely restricted (repeated or useless acquisitions) due to the complicated chemical matrix [11]. The inclusion of preferred precursor ions list (PIL) or the setting of iterative exclusion lists (EL) has been demonstrated powerful in improving coverage of DDA [12,13]. In addition, the use of time-staggered or mass-staggered precursor ion lists (ts-DDA and ms-DDA) displays significant improvement in the coverage of DDA [14–16]. The DIA strategy can acquire the MS^2 information for all precursor ions with the least missing of MS signals, but a pre-deconvolution step is necessary to match between the precursors and their product fragments [17–19]. Moreover, a new hybrid scan approach (namely high-definition data-independent/data-dependent acquisition (HDDIDDA)), by enabling the ion mobility separation and alternating DIA/DDA acquisitions, has been reported with remarkable merits (such as high identification precision, wide MS coverage, and high MS^2 spectra quality) over the conventional DDA or DIA in identifying the complex components from TCM [20–23].

Processing and interpretation of the massive MS data necessitate the development of artificial intelligence technologies to solve the metabolites characterization conundrum [24]. Specifically, the commercial or in-house coded software (such as the UNIFI from Waters Corporation (Manchester, UK), MassHunter from Agilent Technologies (Santa Clara, CA, USA), XCMS, and MS-DIAL, *etc.*) capable of the data pretreatment, and the accessible databases (involving METLIN, HMDB, MassBank, and PubChem, *etc.*) render the MS data processing more intelligent. Data mining technologies, such as background subtraction (BS) [25], mass defect filtering (MDF) [26], neutral loss filtering (NLF) [27], diagnostic product ions (DPI), as well as the integral strategies, can enable the identification of *in vivo* components more easily and efficiently [28]. Despite these announced advancements and merits, a strategy that can facilitate the "one-stop" identification of the *in vivo* components of TCM has been seldom reported, to date. Spectral comparison with the database, including both the MS^1 and MS^2 spectra, is a practical solution to the efficient identification of the chemical components of TCM and the related xenobiotics, while those MS databases mostly are not specific for TCM, resulting in very poor matching results [6,24]. The multi-faced bioinformatics platform, UNIFI™, has integrated multiple MS interpretation modules, which can directly connect to numerous on-line internet databases for the searching and performs the intelligent analysis for identifying the structures via isotope matching, number of citations, secondary fragments, and the other data. And thereby it can support searching of the in-house established libraries (containing the chemical structures or/and determined MS^2 spectra) to render the structures characterization in an intelligent manner [22,29]. Therefore, aimed to efficiently identify the compound formulae-related xenobiotics in the bio-samples, a potential strategy can be practical by in-house establishing the *in vitro* and *in vivo* libraries.

Sishen formula (SSF) is a classic TCM compound formula, composed of Psoraleae Fructus (*Psoralea corylifolia*; **PF**), Myristicae Semen (*Myristica fragrans*; **MyS**), Schisandrae Chinensis Fructus (*Schisandra chinensis*; **SCF**), and Euodiae Fructus (*Euodia rutaecarpa*; **EF**). SSF can exert the synergistic effects of warming the renal and dispersing dampness, and relieving diarrhea with astringents, which is widely used to treat the diarrhea of spleen and deficiency of renal Yang, chronic colitis [30,31], and irritable bowel syndromes [32]. Available researches on SSF focus on its pharmacological activities; however, the chemical basis, particularly those can be absorbed together with the transformed metabolites in animals, remains unclear [33]. Aimed at the characterization of the *in vivo* SSF metabolic profiles in an efficient manner, enhanced metabolic profiling by advanced LC-HRMS and seamless computational data processing can be integrated.

This work aimed to propose a software-aided strategy enabling the "one-stop" identification of both the prototypes and related metabolites of the compound formulae in diverse bio-samples, and to validate it by taking SSF as the template. The overall technical roadmap is described in Fig. 1. Firstly, we aimed to develop a fit-for-purpose hybrid scan method (HDDIDDA) to well resolve the complex SSF components, on the Vion™ ion-mobility quadrupole time-of-flight mass spectrometer coupled with ultra-high performance liquid chromatography (UHPLC/IM-QTOF-MS). Especially, the potential merits of HDDIDDA were exploited by comparing with six other MS^2 acquisition strategies. Secondly, the chemical compounds, the prototypes and related metabolites for four compositional drugs of SSF (**PF**/**MyS**/**SCF**/**EF**) were identified by interpreting

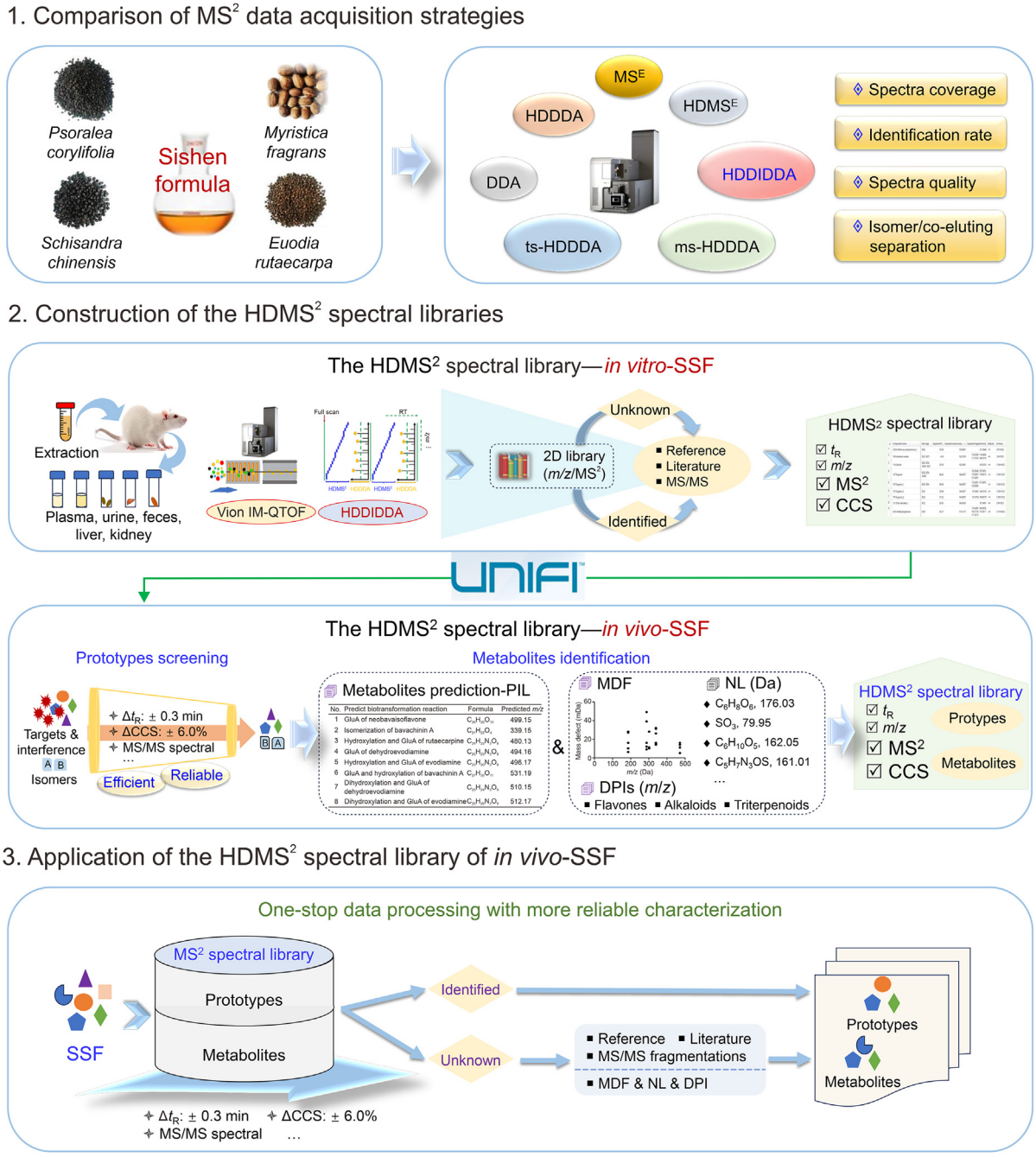

Fig. 1. The overall workflows for the construction and application of software-assisted high-definition MS^2 spectral libraries dedicated to efficient identification of the prototype compounds and metabolites of the compound formulae in rat bio-samples, using Sishen formula (SSF) as an example.

the HDDIDDA MS^2 data via the well-established UNIFI workflows (utilizing MDF, DPI, and NLF). The four-dimensional (4D) high-definition MS^2 spectral libraries (retention time (t_R), MS^1, MS^2, and collision cross section (CCS)), namely *in vitro*-SSF and *in vivo*-SSF, were established on UNIFI based on those identification results. Thirdly, the SSF-related xenobiotics in various rat bio-samples were efficiently identified by matching with the *in vitro*-SSF and *in vivo*-SSF libraries. This work can demonstrate an efficient strategy accomplishing the intelligent identification of compound formulae-related prototype compounds and metabolites in the bio-samples.

2. Experimental

2.1. Materials and reagents

Four component drugs of SSF, involving Psoraleae Fructus (Origin: Yunnan, China), Myristicae Semen (Origin: Guangdong,

China), Schisandrae Chinensis Fructus (Origin: Liaoning, China), and Euodiae Fructus (Origin: Yunnan, China), were obtained from Beijing Tongrentang (Beijing, China). Acetonitrile (ACN), methanol (Fisher, Fair lawn, NJ, USA), and formic acid (FA; Sigma-Aldrich, St. Louis, MO, USA), were of the LC-MS grade. The ultra-pure water was in-house prepared by a Milli-Q Integral 5 system (Millipore, Bed-ford, MA, USA). Fifty-five reference standards, including 20 flavonoids (**1**–**20**), two limonoids (**21** and **22**), 16 lignans (**23**–**38**), five alkaloids (**39**–**43**), four coumarins (**44**–**47**), and eight organic acids (**48**–**55**), were purchased from Desite Biotechnology Co., Ltd. (Chengdu, China) or Standard Biotechnology Co., Ltd. (Shanghai, China). Their chemical structures and information are given in Fig. 2 and Table S1.

2.2. Animal experiments

Eighteen male Sprague–Dawley (SD) rats (weighing 200–240 g), of the specific pathogen-free (SPF) grade, were fed adaptively at room temperature of 20 ± 2 °C with a 12-h light/dark cycle and 50% ± 5% relative humidity for one week. All rats were randomly divided into six groups ($n = 3$): the control group (fed with the distilled water; 32.4 mL/kg/day), the **PF** group (14.4 g/kg/day), the **MyS** group (7.2 g/kg/day), the **SCF** group (7.2 g/kg/day), the **EF** group (3.6 g/kg/day), and the SSF group (32.4 g/kg/day), and they were administered with the suspension twice a day, which lasted for one week. Urine and feces were collected from 0 to 24 h by using the metabolic cages. The blood was collected at multiple time points (e.g., 5, 10, 20, 30, 45 min, 1, 2, 4, 6, 8, 12, and 24 h), after the last intragastric administration, all rats were sacrificed by the cervical dislocation. The urine and blood were centrifuged at 2,200 *g* (10 min, 4 °C) and the supernatants were frozen at a condition under −80 °C. The liver and renal organs in each group were separated, cleared, frozen with the liquid nitrogen, and further transferred for the storage at −80 °C. All protocols were conducted by the Guidelines for Proper Conduct of Animal Experiments, and the animal experiments were approved by the Ethics Committee of Tianjin University of Chinese Medicine (Approval Number: TCM-LAEC2023073).

2.3. Sample preparation

Hot water reflux was utilized to prepare the samples of **PF**, **MyS**, **SCF**, **EF**, and **SSF** decoctions. Each drug powder, **PF** (75.6 g), **MyS** (37.8 g), **SCF** (37.8 g), and **EF** (18.9 g), was extracted to prepare the single-drug extract. These four drugs at a weight ratio 4:2:2:1 (*w/w/w/w*) were mixed together and extracted as the SSF decoction. All samples were soaked with 10-fold ultra-pure water for 30 min, and then decocted for 1 h by the hot water reflux. The extraction liquids were filtered forming the solutions at 100 mg/mL, respectively. Each of the 55 accurately weighed reference standards (1.0 mg) was firstly prepared individually as the stock solution in 1 mL of aqueous methanol (MeOH:H_2O (1:1, *V/V*)), and all the standards were split into four groups (avoiding the isomeric compounds in the same mixed standard solution) and diluted into the mixed standard solutions at the concentration of 10 μg/mL. All the samples were

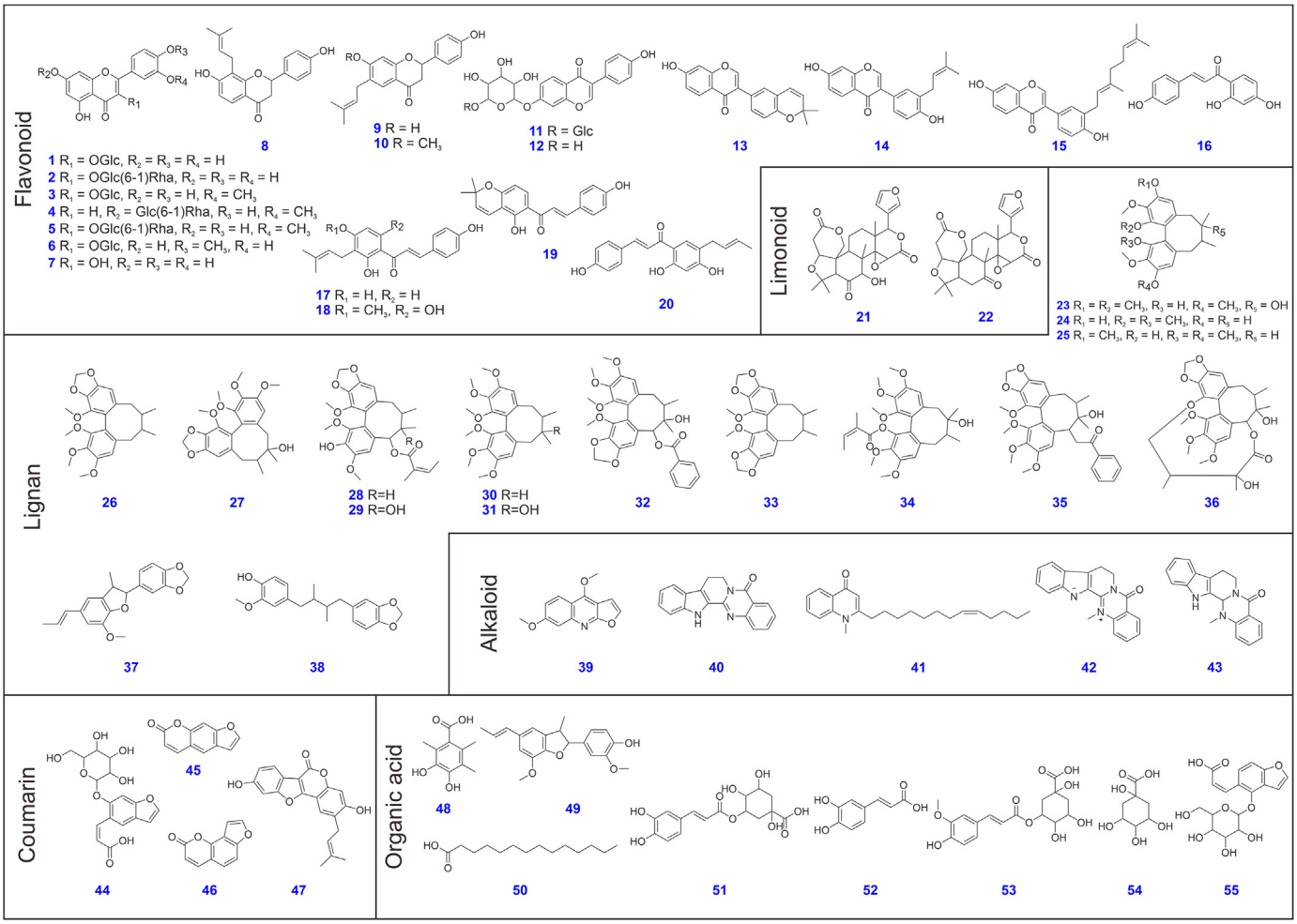

Fig. 2. Chemical structures of 55 reference standards.

centrifuged at 11,481 *g* by an Eppendorf High-Speed Centrifuge (Eppendorf China Limited, Beijing, China) for 10 min at 4 °C, and the obtained supernatant was taken as the test solution for each sample.

The plasma and urine (10 μL for each) samples, collected at different time points, were mixed to prepare the individual tissue bio-samples, respectively. After the addition with the three-fold volumes of methanol and swirling for 3 min, the mixture was centrifuged at 11,481 *g* for 10 min. The feces were dried under vacuum at 50 °C for 15 min and then crushed. The feces samples (500 mg) in each group were added with 1.5 mL of methanol, followed by the ultrasound-assisted extraction (400 W; 30 kHz) for 30 min and then centrifuged at 11,481 *g* for 10 min. The liver and kidney samples (500 mg) were weighed and homogenized with 1.5 mL of normal saline, respectively. The homogenate was transferred into a new tube, and after the centrifugation at 11,481 *g* for 10 min at 4 °C, three-fold volumes of methanol were added into the supernatant to precipitate the proteins. The samples were vortexed for 3 min, followed by another centrifugation for 10 min at 4 °C. Ultimately, the supernatants of the above plasma, urine, feces, liver, and kidney samples were dried under a steady beam of N_2, with the residues further redissolved and vortexed in 100 μL of methanol. After the centrifugation at 11,481 *g* for 10 min, the supernatant was taken as the test solution for the LC-MS analysis.

2.4. UHPLC/IM-QTOF-MS

MS data collection was performed on a Vion™ IM-QTOF mass spectrometer coupled with the ACQUITY UPLC I-Class system via a Zspray™ ESI source (Waters, Manchester, UK). Chromatographic separation relied on an HSS T3 column (2.1 mm × 100 mm, 1.8 μm) under the column temperature of 30 °C. A binary mobile phase, consisting of 0.1% FA-H_2O (A) and ACN (B), was delivered at the flow rate of 0.3 mL/min, following a gradient elution procedure: 0–13 min, 2%–10% B; 13–21 min, 10%–15% B; 21–31 min, 15%–22% B; 31–36 min, 22%–30% B; 36–38 min, 30%–36% B; 38–47 min, 36%–50% B; 47–50 min, 50%–85% B; 50–60 min, 85%–95% B; 60–65 min, 95% B; 65–65.5 min, 95%–2% B; and 65.5–70 min, 2% B. An injection volume of 3 μL was set. The ion source parameters were set as follows (positive mode/negative mode): capillary voltage, 1.5 kV/−1.0 kV; cone voltage, 60 V/−40 V; source offset voltage, 80 V; source temperature, 120 °C; desolvation temperature, 500 °C; cone gas flow rate, 50 L/h; and desolvation gas flow rate, 800 L/h. The mass scan range was *m/z* 100–1,000 under low collision energy of 6 eV, and the full-scan spectra were acquired at 0.3 s/spectrum in the MS^E/$HDMS^E$ and 0.15 s/spectrum in the DDA/HDDDA. For the MS^2 scan in $HDMS^E$, the ramp collision energy (RCE) was 10–60 eV and 20–60 eV in the positive and negative modes, respectively. In the case of MS^2 scan by HDDDA, mass-dependent RCE (MDRCE) of 15–40 eV (positive)/20–40 eV (negative) in the low mass region and 25–50 eV/30–50 eV in the high mass region were utilized. The three most intense precursor ions (top 3) were set to automatically trigger the MS/MS acquisition with the intensity threshold set at 200 counts, which stopped until the time exceeded 0.5 s. Data calibration was performed using an external reference (LockSpray™) by the constant infusion of a leucine enkephalin solution (Sigma-Aldrich; 200 ng/mL) at a flow rate of 10 μL/min. Default parameters were defined for the traveling wave IM separation, and calibration of the CCS was conducted according to the manufacture's guidelines using a mixture of calibrants [34].

2.5. Construction of the ts-HDDDA and ms-HDDDA scan approaches

Aside from the five scan approaches (involving the DDA, HDDDA, MS^E, $HDMS^E$, and HDDIDDA) that are directly available on the Vion™ IM-QTOF-MS platform, ts-HDDDA and ms-HDDDA were established for the purpose of performance comparison, by referring to the following procedures [31,32]. Step 1: the $HDMS^E$ data of SSF were collected and further processed using the UNIFI workflows followed by with subsequent manual confirmation. Step 2: the t_R or mass-staggered precursor ions lists, collected from step 1, was split into three individual ion lists depending on the detailed procedures: 1) ions were sorted according to the t_R or *m/z*; 2) the precursors were tagged from one to five and then circled repeatedly to cover all the precursor ions; 3) all precursors tagged as the same number were collected into the new ion lists as the sub-method. Step 3: five ts-HDDDA and ms-HDDDA sub-methods were established, ultimately.

2.6. Construction of the high-definition MS^2 spectral library of *in vivo*-SSF

Based on the workflows previously reported [23], three steps were applied to construct the high-definition MS^2 spectral library of *in vivo*-SSF: 1) an in-house MS^1 library containing the information of *m/z*, molecular formula, and chemical structure was created by integrating compounds obtained from the four component herbs and the multiple databases (e.g., Web of Science, SciFinder, CNKI, PubChem, ChemSpider, and Chemicalbook); 2) chemical entities of four component herbs *in vitro* and *in vivo* acquired by the HDDIDDA method were processed by the in-house MS^1 library-driven UNIFI (Waters; 1.9.3.0 version) workflows for the peak correction, extraction, and annotation. The lists of identified compounds (containing *m/z*, t_R, CCS, molecular formula, mass error, adducts, and MS^2 fragments) were finally obtained contributing to the high-definition MS^2 spectral library (*in vitro*-SSF); 3) The established *in vitro*-SSF library was then used to screen the prototype components in all bio-samples by filtering *m/z*, Δt_R (<0.3 min), and ΔCCS (<6%), where the compounds identified by the reference standards were used as the parent drugs for the prediction of possible metabolites. In detail, the computational intelligent metabolic response prediction, combined with MDF (±50 mDa) and NLF strategies was performed to track the potential metabolites. As a supplementary identification strategy, the precursor ions list (PIL) of the metabolites data collected from the literature was acquired and verified by the predicted *m/z*, precursor ions, and fragmentation behavior. Eventually, another high-definition MS^2 spectral library, *in vivo*-SSF (including the prototypes and metabolites), was established.

Key parameters of the UNIFI software for data processing were set as follows. Find 4D peaks (MS^E/$HDMS^E$): high-energy intensity threshold, 100.0 counts; low-energy intensity threshold, 200.0 counts. Find DDA masses (DDA/HDDDA): MS ion intensity threshold, 200.0 counts; MS/MS ion intensity threshold, 100.0 counts. Target by mass: target match tolerance: 10.0 ppm; select and filter one candidate of all isotopes, generate prediction fragments from the structure, and enable finding fragments in source code; fragment match tolerance, 10.0 ppm. Absolute t_R identification tolerance of 0.3 min and CCS tolerance of 6.0% were set when matching with the self-established high-definition MS^2 spectral database. Adducts: positive mode, $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$; negative mode, $[M-H]^-$, $[M+HCOO]^-$; LockMass: combine width, 3 scans; mass window: 0.5 *m/z*; reference mass: *m/z* 556.2766 (+)/554.2620 (−); reference charge: +1/−1.

2.7. Efficient identification of the prototypes and related metabolites of SSF by the in vivo-SSF high-definition MS^2 spectral library

The identification of prototype compounds and related metabolites in diverse rat bio-samples after taking the SSF decoction was

conducted by directly invocating the *in vivo*-SSF high-definition MS^2 spectral library, based on the automatic annotation workflows of the UNIFI platform (filtering condition: ΔCCS <6% and Δt_R <0.3 min). Finally, the components listed in the "Identified compounds" and "Unknown components" were characterized via the MS^2 data analysis. In addition, the other PIL, MDF, NLF, and DPIs strategies were applied to supplement the identification of those SSF metabolites.

3. Results and discussion

3.1. Establishment and performance comparison among seven MS^2 acquisition methods

The integrated high-definition MS-based data acquisition protocols that are capable of performing multi-event acquisitions can greatly drive the processing of massive MS data, thus showing advantages over the traditional methods [35,36]. In consideration of the differentiated scan criteria in the MS^2 data acquisition, by utilizing the Vion™ IM-QTOF analytical platform, we firstly compared the performance of seven different MS^2 data acquisition approaches (e.g., DDA/HDDDA/MS^E/$HDMS^E$/HDDIDDA/ts-HDDDA/ms-HDDDA) in the characterization of the multicomponents from the SSF decoction (Fig. S1 shows the base peak chromatograms). In this section, the MS/MS triggering rate (MS^2 TR) (%), which could partially embody the MS/MS coverage, was calculated as the index. Here, "annotation" is the raw characterization results obtained by using UNIFI to process the SSF decoction data (containing the repeated identifications and false-positive results), while "de-repetition" is refined from the "annotation" list by removing those repeated identifications. MS^2 TR was based on the de-repetition data by calculating the number ratio of Identified to the sum of Identified compounds and Unknown components (Table 1).

These seven MS^2 acquisition methods showed differences related to the MS^2 TR. On the one hand, the identified components by the IM-enabled acquisition modes, whatever for the annotation list or the de-repetition list, became much less compared with those without the IM separation. It was consistent with the differentiation in the intensity of precursor ions [22]. Moreover, the data-independent MS^E and $HDMS^E$, in comparison with the data-dependent DDA and HDDDA, exhibited broader coverage. Especially, the inclusion of staggered lists for the t_R and *m/z* could greatly improve the coverage of HDDDA. In the case of the hybrid scan approach (HDDIDDA), by comparing the de-repetition data, the involved $HDMS^E$ was similar to that of the single $HDMS^E$ (428 vs. 617), whilst the involved HDDDA displayed better performance than the single HDDDA (139 vs. 123). And on the other hand, assessed by the MS^2 TR, the $HDMS^E$ of the hybrid scan method was ranked the highest (47.93%), even higher than those of ms-HDDDA (40.12%) and ts-HDDDA (34.49%), while the involved HDDDA (25.88%) was also higher than DDA (18.02%) and HDDDA (23.38%).

Additionally, these seven approaches displayed differences in the MS^2 spectral quality, the ability of isomers separation, and the ion response. As illustrated in Fig. 3A, the MS^2 spectra of a reference compound, schisandrin B (lignan), acquired by HDDIDDA, DDA, and MS^E, were compared. Evidently, similar MS/MS spectra were obtained by the HDDDA involved in HDDIDDA and the DDA without IM separation, which were composed of much fewer interfering signals than that obtained by MS^E. The enabling of IM separation could well resolve some co-eluting components to differentiate some isomers. It could be evident from the enhanced separation on the co-eluted isomers at t_R 21.93, 31.42, and 41.14 min, as well as the effective separation of the isomers at *m/z* 163.07, 271.05, and 483.20, observed in the heat map of the $HDMS^E$ data (Fig. 3B).

In the last step, the identification rate (IR (%)), referring to the number ratio of compounds, identified by comparison with the reference standards or tentatively characterized by analyzing the fragmentation pathways, to those listed in the "annotation" list, was utilized to indicate the accuracy of characterization in both the ms/ts-HDDDA and HDDIDDA modes. Venn diagrams were used to illustrate the differences in the characterization profiles. As depicted in Fig. 3C, separately 240, 289, and 200 compounds were characterized based on the data of ts-HDDDA, ms-HDDDA, and $HDMS^E$ of the hybrid scan approach (HDDIDDA-$HDMS^E$), and thus the corresponding IR values were 2.12%, 3.53%, and 11.86%, respectively. The ts-HDDDA and HDDIDDA workflows separately identified 95 and 69 unique compounds, aside from the shared 145 ones, whilst the ms-HDDDA and HDDIDDA yielded 140 and 65 distinct compounds, respectively. The complex process of collecting ion lists and the cumbersome data processing were involved (Identified: 5,455/5,078, Unknown: 5,842/3,090) in ts-/ms-HDDDA, but there were no discernible advantages of ts/ms-HDDDA in terms of IR. Taking together, we finally selected the hybrid scan approach of HDDIDDA in establishment of the high-definition MS^2 spectral libraries.

3.2. Construction of the high-definition MS^2 spectral libraries: *in vitro*-SSF and *in vivo*-SSF

Based on the MS^2 data of HDDIDDA acquired for four component herbal medicines both *in vitro* (the contained multicomponents) and *in vivo* (drug-related xenobiotics in the plasma, urine, feces, liver, and kidney), the high-definition MS^2 spectral libraries, *in vitro*-SSF and *in vivo*-SSF, were elaborated.

The *in vitro*-SSF library was composed of 439 compounds identified from four component drugs (including 125 ones from **PF**, 74 from **MyS**, 64 from **SCF**, and 176 from **EF**; Table S2). We firstly characterized the multicomponents from four single drugs and then gathered them into the UNIFI platform by sending them to the

Table 1
Comparison of the performance of seven MS^2 acquisition methods in characterizing the multicomponents from Sishen formula (SSF).

Scan method	Identified (annotation)	Unknown (annotation)	Identified (de-repetition)	Unknown (de-repetition)	MS^2 TR (%)
DDA	1,444	4,091	167	760	18.02
HDDDA	953	1,907	123	403	23.38
MS^E	2,925	5,032	2,433	4,488	35.15
$HDMS^E$	1,158	1,948	617	1,125	35.42
HDDIDDA-$HDMS^E$	1,092	594	428	465	47.93
HDDIDDA-HDDDA	631	1,086	139	398	25.88
ts-HDDDA	5,455	5,842	396	752	34.49
ms-HDDDA	5,078	3,090	487	727	40.12

MS^2 triggering rate (TR) (%): the number ratio of the Identified to the sum of Identified and Unknown based on the de-repetition data.
DDA: data-dependent acquisition; HDDDA: high-definition data-dependent acquisition; $HDMS^E$: high-definition MS^E; HDDIDDA-$HDMS^E$: $HDMS^E$ in high-definition data-independent/data-dependent acquisition; HDDIDDA-HDDDA: HDDDA in high-definition data-independent/data-dependent acquisition; ts-HDDDA: HDDDA with time-staggered precursor ion lists; ms-HDDDA: HDDDA with mass-staggered precursor ion lists.

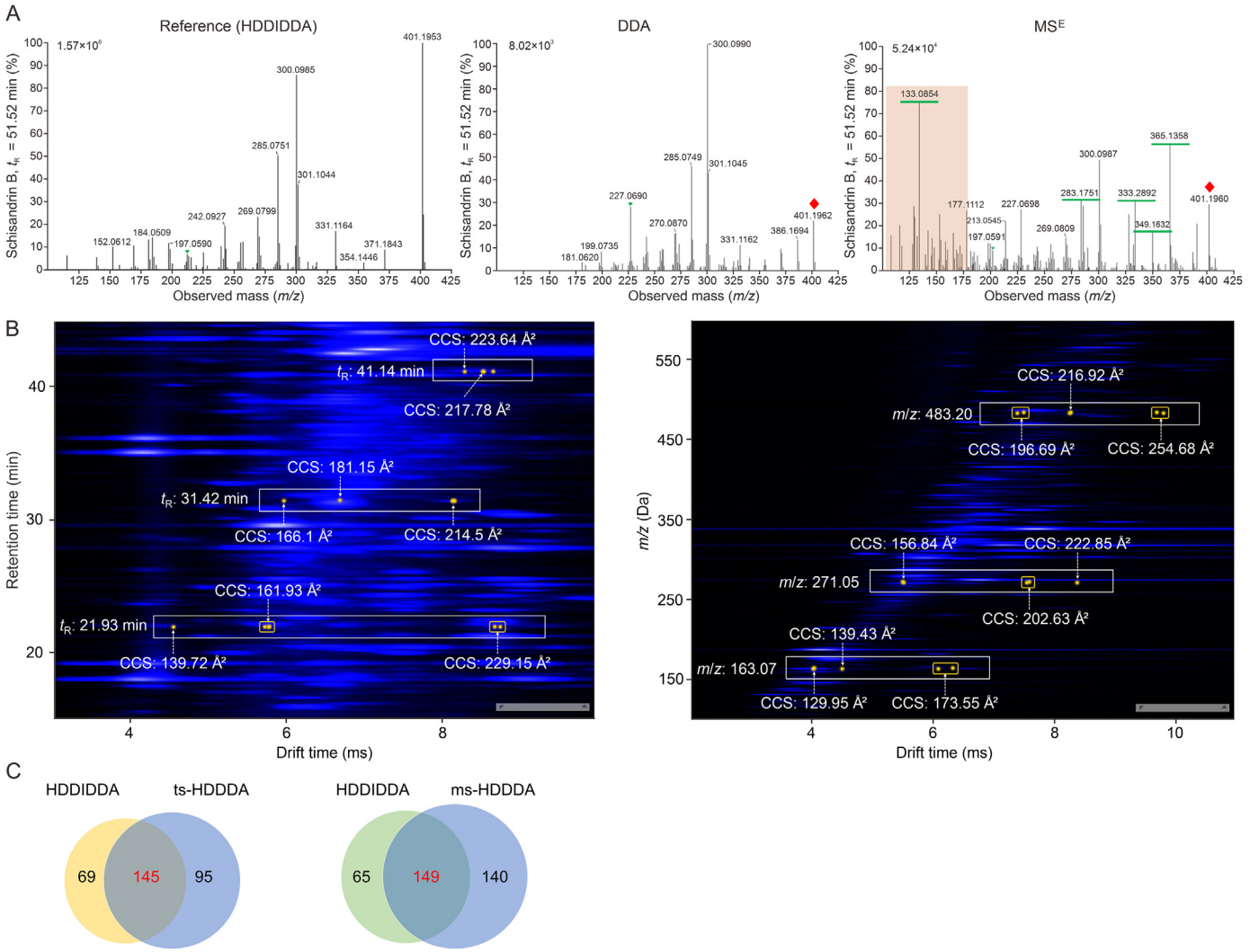

Fig. 3. Evaluation of the performance of different MS^2 acquisition approaches in characterizing the multicomponents from Sishen formula (SSF). (A) Comparison of the MS^2 spectra of schisandrin B obtained by high-definition data-independent/data-dependent acquisition (HDDIDDA), data-dependent acquisition (DDA), and MS^E. (B) Two-dimensional heat map view of the ion mobility data showing the separation of co-eluting components (retention time (t_R) vs. drift time) and isomers (*m/z* vs. drift time). (C) Comparison of the common and unique components between the HDDDA with mass-staggered precursor ion lists (ms-HDDDA)/HDDDA with time-staggered precursor ion lists (ts-HDDDA) and HDDIDDA.

"Scientific Library" to construct the high-definition MS^2 library of *vitro*-SSF. Notably, 43 compounds were identified by the reference standards comparison (Table S1). The *in vitro*-SSF library contained rich information regarding each characterized component (e.g., *m/z*, t_R, CCS, molecular formula, mass error, adducts, and MS^2 fragments), and various categories of compounds (such as the flavonoids, alkaloids, coumarins, triterpenoids, and lignans) were recorded. Generally, the identification of these compounds was based on the exact *m/z* and the MS/MS fragmentation information, featured by various NLs and DPIs. The characterized flavonoids were mainly derived from **PF** and **EF**, including flavones and chalcones. By the positive collision-induced dissociation (CID)-MS/MS, the flavones were prone to undergo the NLs of CO, H_2O, radical fragmentation ($\bullet CH_3$), and *Retro* Diels-Alder (RDA) cleavages, while chalcones could easily break the annotated I and II bonds. The typical DPIs, including *m/z* 149.02 $[C_8H_5O_3+H]^+$, 147.04 $[C_9H_7O_2+H]^+$, 137.02 $[C_7H_5O_3+H]^+$, and 119.04 $[C_8H_7O+H]^+$, were used for the rapid structural elucidation of the flavonoid compounds [37]. For instance, the characterization of brosimacutin G (**#146**: t_R 32.71 min, CCS 183.42 Å^2 for the $[M+H]^+$ at *m/z* 357.1321) and bavachinin A (**#388** identified by comparison with the reference standard: t_R 49.74 min, CCS 184.63 Å^2 for the $[M+H]^+$ at *m/z* 339.1584) was illustrated as the examples of flavonoids (Fig. S2). Lignans, such as myticaganal B (**#158**: t_R 34.54 min, CCS 171.53 Å^2 for the $[M+H]^+$ at *m/z* 313.1058) and schisandrol A (**#317** identified by comparison with the reference standard: t_R 44.10 min, CCS 225.93 Å^2 for the $[M+H-H_2O]^+$ at *m/z* 415.2107) exhibited in Fig. S2, were mostly characterized from **SCF**. Upon the fragmentation, they showed the NL of H_2O and radical cleavages of $\bullet CH_3$ and $\bullet OCH_3$, as well as the typical fragmentation rules including biphenyl bond rupture, loss of C-7, C-8, or (and) C-6/C-9 [38]. As the main subtypes of compounds in **SCF** and **EF**, the identification of triterpenoids was achieved using the information of NLs (CO, CO_2, and H_2O) and DPIs (*m/z* 161.05 $[C_{10}H_9O_2+H]^+$ and 95.01 $[M+H-C_{21}H_{28}O_6]^+$), and the characterization of limonin (**#284** identified by comparison with the reference standard: t_R 42.12 min, CCS 197.84 Å^2 for the $[M+H]^+$ at *m/z* 471.2019) was a typical case (Fig. S3). In addition, the characterized alkaloids were primarily characterized from **EF**, such as evodiamide (**#309** identified by comparison with the reference standard: t_R 43.57 min, CCS 160.05 Å^2 for the $[M+H]^+$ at *m/z* 308.1734) and rutaecarpine (**#341** identified by comparison with the reference standard: t_R 45.62 min,

CCS 157.81 Å^2 for the $[M+H]^+$ at *m/z* 288.1121) as the examples (Fig. S3). DPIs occurring to the positive-mode fragmentation of alkaloids like *m/z* 134.06 $[C_8H_8NO+H]^+$, 116.05 $[C_8H_6N+H]^+$, and 106.06 $[C_7H_8N+H]^+$, as well as the NL of CO, radical cleavage of $\bullet CH_3$, and RDA cleavages, were frequently utilized for the rapid structural characterization of alkaloids [39]. In the case of the coumarins primarily from **PF**, the common NLs of CO, H_2O, and CO_2, together with radical cleavage of $\bullet CH_3$, were the indicative evidence, using psoralidin (**#368** identified by comparison with the reference standard: t_R 48.25 min, CCS 174.74 Å^2 for the $[M+H]^+$ at *m/z* 337.1069) as an example (Fig. S3).

The *in vivo*-SSF library consisted of 145 prototype compounds (Table S3) and 247 metabolites (Table S4) identified from the bio-samples of rats after separately administrating four component drugs (**PF**/**MyS**/**SCF**/**EF**). On the one hand, the identification of 145 prototype compounds from the rat bio-samples was based on the high-definition *in vitro*-SSF MS^2 library. The compound **P108#** (t_R 44.59 min, CCS 148.02 Å^2 for the $[M+H]^+$ at *m/z* 255.0641) was here illustrated. The searching of the elaborated *in vitro*-SSF MS^2 library could result in the characterization of an isomer of daidzein (Δt_R: 0.02 min; ΔCCS: 0.13%). Consistently, its MS^2 fragment ions were similar to those observed for daidzein, involving *m/z* 237.06 $[M+H-H_2O]^+$, 227.06 $[M+H-CO]^+$, 199.07 $[M+H-C_2O_2]^+$, and 137.02 $[M+H-C_8H_6O]^+$, which might involve the hydroxyl substituents on the A and B rings different from those of daidzein [40]. On the other hand, a total of 247 metabolites were characterized from the rat bio-samples after administrating four component drugs. MDF, as a powerful data processing vehicle used in drug metabolism, was enabled by the UNIFI software to filter the possible metabolites based on the mass defect data of pre-designed metabolites. We integrated MDF processing and the NLF/DPIs strategies, which led to the identification of 176 metabolites. In addition, a metabolite knowledge-based PIL was created and incorporated into the HDDDA scan of HDDIDDA to enhance the identification of metabolites. The known *in vivo* metabolites information (including the *m/z* and adducts) of the reference compounds and four component drugs were screened from the literature via Web of Science (https://www.webofknowledge.com), CNKI (www.cnki.net), and PubMed (https://pubmed.ncbi.nlm.nih.gov/). Through these efforts, 71 metabolites were additionally characterized from all the biological samples. The main metabolic pathways involved the oxidation, reduction, hydrolysis,

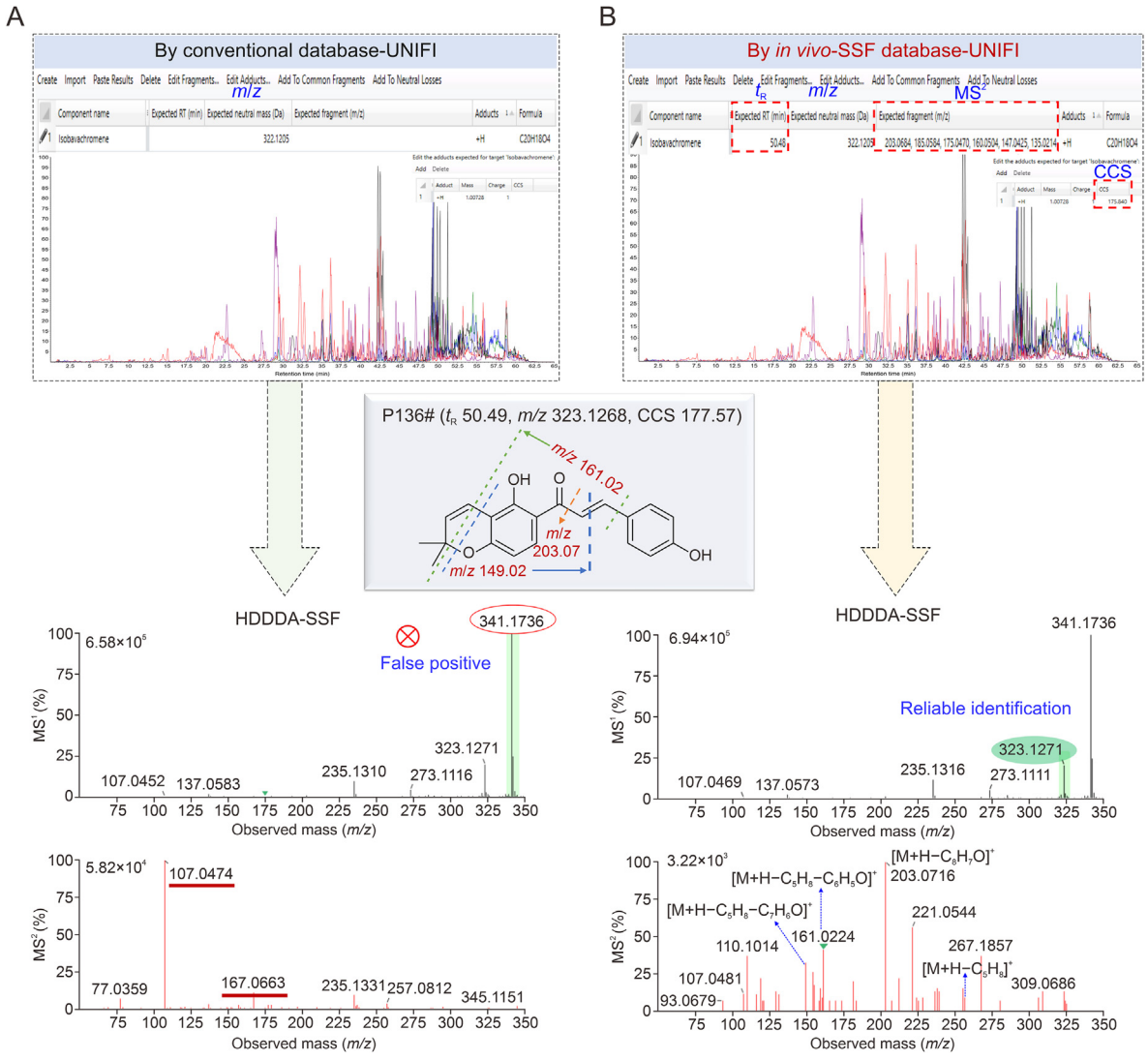

Fig. 4. The identification of isobavachromene based on (A) the conventional strategy of the UNIFI software and (B) the high-definition MS^2 spectral library approach.

demethylation, acetylation, glucuronidation, sulfation, and the related combined reactions (Table S4). Ultimately, these characterized 145 prototypes and 247 metabolites were used to establish the high-definition MS^2 library of *in vivo*-SSF. As an illustrative example, the compound **M231#** (t_R 45.56 min, CCS 161.28 Å^2 for the $[M+H]^+$ at *m/z* 304.1071) was characterized as an oxidized product of rutaecarpine because of the additional 16 Da in the molecular mass compared with the protonated rutaecarpine at *m/z* 288.1131. The MS^2 spectrum revealed the major product ions of *m/z* 261.0992 $[M+H-NCO]^+$, 169.0752 $[M+H-C_7H_4NO_2]^+$, and 115.0532 $[M+H-C_{10}H_8N_2O_2]^+$, all of which were consistent with those observed for rutaecarpine [41]. Moreover, by the *in vivo* transformation, glucosides could be catalyzed by numerous enzymes forming the aglycones, and these aglycones could further undergo the oxidation, reduction, sulfation, glucuronidation, and the other reactions [42]. As a typical case, **M18#** showed the deprotonated precursor ion ($[M-H]^-$) at *m/z* 301.0016 (t_R 13.19 min, CCS 152.29 Å^2), and was easily dissociated into the product ions of *m/z* 221.0447 $[M-H-SO_3]^-$, 203.0343 $[M-H-SO_3-H_2O]^-$, 131.0488 $[M-H-SO_3-C_2H_2O_4]^-$, and 115.0543 $[M-H-SO_3-H_2O-C_2O_4]^-$ via the CID-MS/MS fragmentation. The product ion at *m/z* 203.0343 should be the deglycosylation product (−162.05 Da) of psoralenoside, as the secondary fragments similar to those of psoralenoside were observed. Thus, we presume this metabolite was the result of psoralenoside undergoing the sulfation, hydration, and deglycosylation.

3.3. Superiority of the high-definition MS^2 spectral library strategy in identifying the SSF-related xenobiotics in vivo

Currently, commercial or in-house coded software and the accessible databases are contributing to more intelligent processing of MS data [35]. Particularly for the bio-samples, the need for automatic data assignment from the high-definition MS^2 datasets is much greater to expedite the identification of *in vivo* components. The UNIFI™ bioinformatics software from Waters can facilitate the automated peak annotation by matching with an in-house library or the commercial database [22,29,43], thus greatly improving the analytical efficiency. However, software-assisted characterization can easily bring the false-positive identifications when facing doubly charged ion clusters [22]. Furthermore, due to the challenging and ambiguous identification of isomers or co-eluted compounds, manual evaluation of the MS^2 fragments was often necessary to confirm the characterizations. Accordingly, we aimed to construct a UNIFI-based multi-dimensional information library (UNIFI-MIL), including the t_R, MS^1, MS^2, and CCS information (4D) of those characterized components. A typical case was the chemical library of Xuebijing [23]. Because of the more restrictive filtering criteria (Δt_R and ΔCCS), this UNIFI-MIL strategy could fully utilize the automatic peak extraction and annotation functions, and thus enhance the accuracy and efficiency of the analysis.

As shown in Fig. 4, **P136#** (t_R 50.49 min, CCS 177.57 Å^2 for the $[M+H]^+$ at *m/z* 323.1268) showed the Δt_R 0.09 min and ΔCCS 0.18%

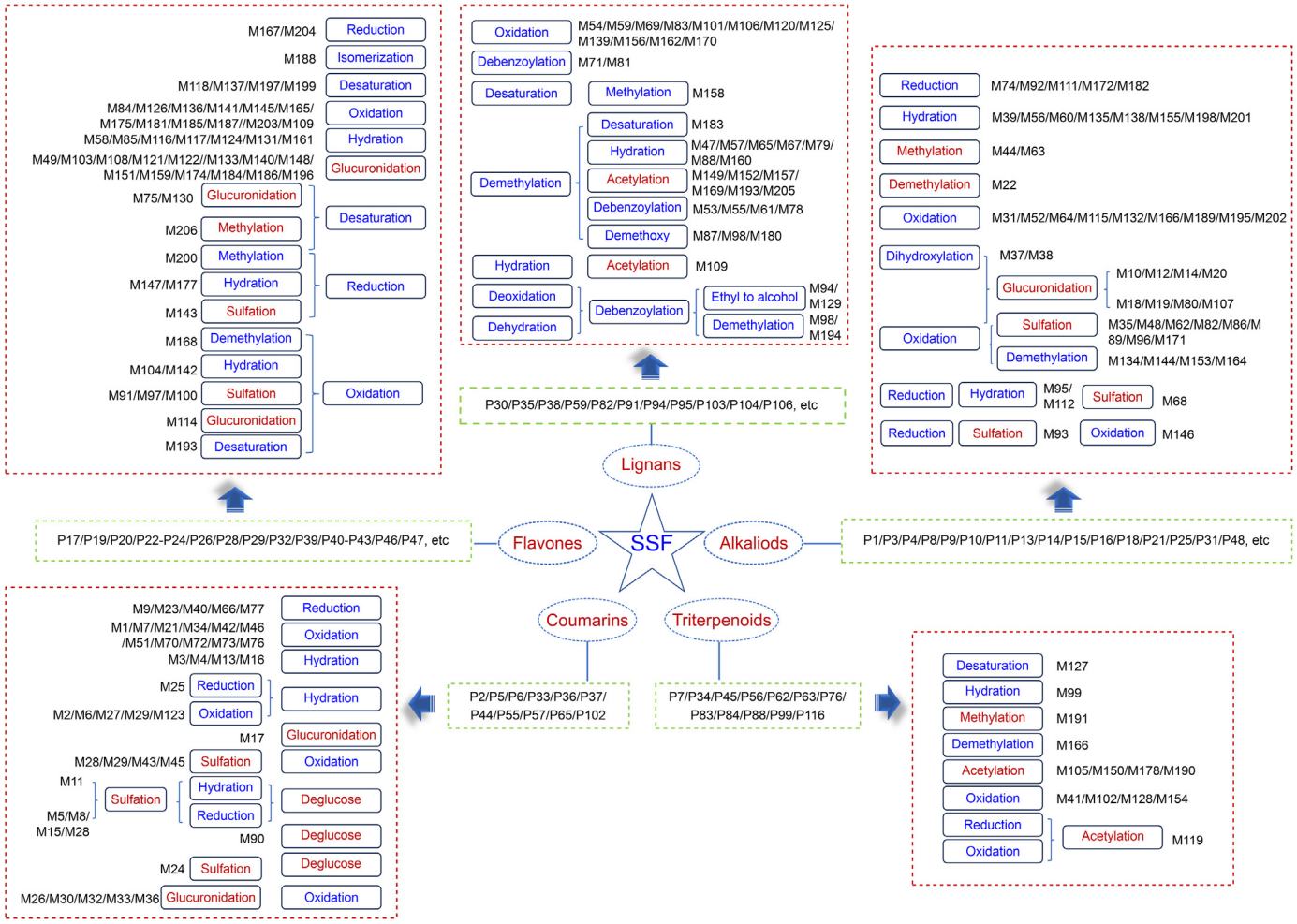

Fig. 5. Summary and illustration of the involved metabolic pathways for the metabolites identified in the plasma, urine, feces, liver, and kidney bio-samples of rats after the oral administration of Sishen formula (SSF) decoction.

variations (Table S3) using the HDDDA data, compared with isobavachromene in the library of *in vivo*-SSF. By the conventional 2D UNIFI library (with the documented components and structures input), it was incorrectly matched and listed as the "Unknown components". The characterization of a co-eluting compound at *m/z* 341.1736 should be a false-positive result. However, when invoking the high-definition MS^2 library of *in vivo*-SSF, automatic extraction and matching based on the identical fragmentation ions with those of isobavachromene were achieved, which could correctly assign **P136#** among the "Identified components". Notably, the multi-dimensional information filtering provided a favorable distinction between the isomers and those co-eluted compounds. For example, **P120#** with *m/z* 321.1113 (t_R 47.16 min, CCS 172.65 $Å^2$) was identified as corylin, while **P61#** with *m/z* 321.1113 (t_R 38.81 min, CCS 171.29 $Å^2$) as an isomer of corylin. The co-eluted compound **P24#** at t_R 32.66 min (*m/z* 300.1122, CCS 159.65 $Å^2$) was characterized as an isomer of evodiamine, while **P25#** at t_R 32.71 min (*m/z* 255.0640, CCS 148.41 $Å^2$) as daidzein.

In spite of these mentioned advantages, this strategy can be further improved by establishing the high-definition MS^2 library more accessible to the investigators.

3.4. Comprehensive and efficient characterization of the xenobiotics in the bio-samples of SSF-administrated rats by invoking the in vivo-SSF library

3.4.1. Identification of the prototype compounds

Comprehensive and efficient identification of the prototype compounds in the rat bio-samples after administrating the SSF decoction was achieved by processing the HDDIDDA data via the *in vivo*-SSF library incorporated UNIFI workflows. Consequently, 95 peaks were structurally annotated as the prototype components occurring in the plasma, urine, feces, liver, and kidney of rats (Table S5). In addition, another 23 components were identified because of the additional manual analyses with the DPIs and NLF strategies. These prototypes included 46 flavonoids, 30 alkaloids, 15 lignans, 10 coumarins, 13 triterpenoids, and four other compounds, and 33 thereof were identified by comparison with the reference standards. It was noted that the identification rate by directly invoking the high-definition MS^2 library reached 80.51%.

3.4.2. Characterization of the metabolites

The metabolites in the bio-samples of rats were efficiently identified by applying the *in vivo*-SSF high-definition MS^2 spectral library and complementary identification strategies. We could characterize 164 metabolites of SSF by directly invoking the *in vivo*-SSF library, while another 42 compounds were additionally characterized by using the MDF and NLF strategies. It could indicate an identification rate of 79.61% by accessing the high-definition MS^2 library. An overview of the metabolic pathways for the SSF components in rats is presented in Fig. 5, and the main metabolic pathways involved the oxidation, reduction, hydrolysis, methylation, demethylation, acetylation, sulfation, glucuronidation, deglycosylation, and the associated combination reactions.

Diverse subtypes of the SSF components (e.g., flavones/alkaloids/lignans/coumarins/triterpenoids/organic acids/anthraquinones) exhibited differentiated metabolic features among five tested bio-samples. The amounts of characterized prototype compounds were analyzed. Firstly, in the case of the prototypes, separately 78, 65, 58, 56, and 44 compounds were characterized, showing a decreasing trend from the feces, urine, plasma, and liver, to the kidney. Comparatively, flavones, alkaloids, lignans, coumarins, and triterpenoids, were detected in all five bio-samples, whilst organic acids and anthraquinones were almost undetectable from the urine and the liver/kidney samples (Fig. S4). Particularly, the amounts of flavones and alkaloids were the two richest subtypes in all bio-samples.

3.5. Metabolites identification driven by the CCS prediction

Despite that the multidimensional structure information can be offered by LC-MS enabling the characterization of the drug-related xenobiotics in various bio-samples, the lack of sufficient evidence by this technique often disenables the accurate identification of

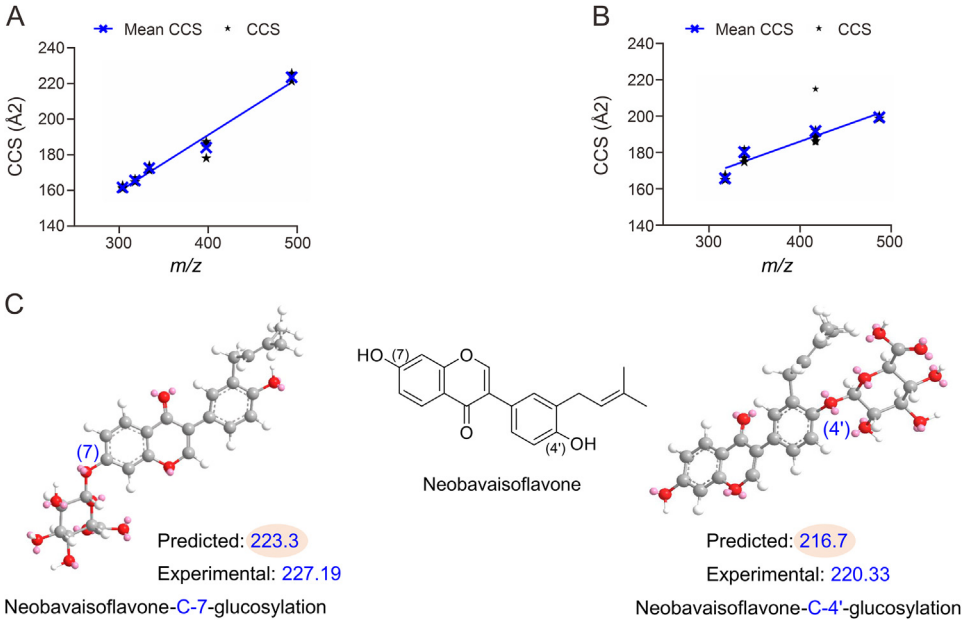

Fig. 6. The relationship between the structures of the Sishen formula (SSF) metabolites and their determined collision cross section (CCS) values. (A) The *m/z* versus CCS for different metabolic pathways of dehydroevodiamine metabolites. (B) The *m/z* versus CCS for the identified oxidation metabolites of SSF; (C) the gas-phase conformations of the positional isomers of neobavaisoflavone-*O*-glucuronides and their difference in the determined CCS values. CCS prediction was based on ALLCCS (http://allccs.zhulab.cn/).

those isomers and stereoisomers within the various metabolic pathways or the metabolic sites. IM-MS is able to provide the additional dimension of separation of the gas-phase ions that is orthogonal to the t_R and *m/z* resolution, and the derived CCS has shown the great potential in identifying the isomeric metabolites [44]. Here, the role of IM-MS in identifying dehydroevodiamine (a major alkaloid compound from **EF**) and its various metabolites was illustrated, by correlating the *m/z* with the determined CCS values (mean CCS: the average CCS value for the metabolites undergoing the same metabolic pathway; CCS: the real CCS value for each metabolite) for some metabolic pathways of metabolites (including the demethylation and oxidation, dihydroxylation, oxidation and sulfation, oxidation and glucuronidation). On the broad scale, it was found that the CCS values were highly linear to the adduct masses (R^2 0.978; Fig. 6A). In addition, when viewed in terms of different metabolites undergoing the same metabolic pathway (such as the oxidation reaction), the CCS values of different oxidized metabolites were also proportional to their *m/z* values (R^2 0.892; Fig. 6B). Nonetheless, some significantly differentiated CCS values of the metabolites undergoing the same metabolic pathway were observed. The flavonoids in SSF can have multiple reaction sites of the phase II metabolism, and the glucuronic acid adduct sites may cause the difference in CCS. It's known that the extensibility of a structure is closely related to a larger CCS [45]. An isoflavone compound, neobavaisoflavone, and its metabolites were thus further investigated. It contains two hydroxyl groups separately at C-7 and C-4′, suggesting two potential glycosylation sites. We calculated the total energy of two regio-isomers (Fig. 6C), and the C-7-glucuronidation molecule presented a total energy of 117.4927 koal/mol which was much higher than that of the C-4′-glucuronidation molecule (43.7035 koal/mol). In contrast, the latter metabolite exhibited a more compact structure than the former. Amongst the glucuronidation metabolites of neobavaisoflavone identified in the bio-samples after administrating the SSF decoction, **M49**# (t_R 26.89 min, CCS 227.19 Å^2) showed the highest CCS value and was thus speculated as the C-7-glucuronidation metabolite. In addition, the shorter elution time in the reversed-phase chromatography could indicate higher polarity of a metabolite, and consequently **M49**# was inferred as a C-7-glucuronidation metabolite of neobavaisoflavone, whereas **M121#** (t_R 37.37 min, CCS 220.33 Å^2) as a C-4′-glucuronidation metabolite. These findings agree with an earlier study in which two metabolites of neobavaisoflavone-*O*-glucuronides were discovered [46], but confirmation on these identifications is in need of more direct evidence in the future studies.

4. Conclusions

To facilitate the efficient and more reliable identification of the drug-related xenobiotics in bio-samples, the current work presented and validated a software-oriented high-definition MS^2 spectral library strategy. It enabled the “one-stop” data processing and characterization of prototypes and metabolites in various rat bio-samples, taking a four-component compound formula SSF as a case. In particular, the well-established hybrid scan approach of HDDIDDA exhibited merits in the MS/MS TR and identification rate with the enhanced separation of isomers and co-eluted components. Compared with the traditional workflows enabled by UNIFI (2D library), the strategy by invoking the high-definition MS^2 spectral library (4D library) could improve the identification efficiency and accuracy. The libraries of *in vitro*-SSF (containing 439 compounds) and *in vivo*-SSF (involving 145 prototypes and 247 metabolites) were constructed, and the application of the *in vivo*-SSF library could very efficiently characterize the prototypes (80.51%) and metabolites (79.61%) in the rat bio-samples. By combining with the PIL, NLF, and DPIs strategies, we were able to identify 118 prototypes and 206 metabolites of SSF. CCS prediction and comparison were proven as complementary in discriminating some isomeric metabolites. The major metabolic pathways for five subcategories of SSF constituents were unveiled. Conclusively, this software-oriented high-definition MS^2 spectral library strategy can be a practical solution to the identification of compound formula-related xenobiotics in the bio-samples featured by the high accuracy, reproducibility, and efficiency.

CRediT authorship contribution statement

Lili Hong: Formal analysis, Investigation, Writing – original draft. **Wei Wang:** Formal analysis, Writing – original draft. **Shiyu Wang:** Formal analysis. **Wandi Hu:** Software. **Yuyang Sha:** Software. **Xiaoyan Xu:** Software. **Xiaoying Wang:** Conceptualization. **Kefeng Li:** Conceptualization. **Hongda Wang:** Conceptualization, Writing – original draft. **Xiumei Gao:** Funding acquisition, Supervision. **De-an Guo:** Project administration, Supervision. **Wenzhi Yang:** Conceptualization, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was financially supported by National Natural Science Foundation of China (Grant No.: 82192914), Tianjin Outstanding Youth Fund (Grant No.: 23JCJQJC00030), and the Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (Grant No.: ZYYCXTD-C-202009).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpha.2024.100994.

References

[1] H. Xu, Y. Zhang, P. Wang, et al., A comprehensive review of integrative pharmacology-based investigation: A paradigm shift in traditional Chinese medicine, Acta Pharm. Sin. B 11 (2021) 1379–1399.
[2] R.D. Beger, W. Dunn, M.A. Schmidt, et al., Metabolomics enables precision medicine: ‘A white paper, community perspective.’, Metabolomics 12 (2016), 149.
[3] Y.H. Chen, J.H. Bi, M. Xie, et al., Classification-based strategies to simplify complex traditional Chinese medicine (TCM) researches through liquid chromatography-mass spectrometry in the last decade (2011–2020): Theory, technical route and difficulty, J. Chromatogr. A 1651 (2021), 462307.
[4] M.L. Huang, S.J. Yu, Q. Shao, et al., Comprehensive profiling of Lingzhihuang capsule by liquid chromatography coupled with mass spectrometry-based molecular networking and target prediction, Acupunct. Herb. Med. 2 (2022) 58–67.
[5] X. Gao, J.J. Wang, X.L. Chen, et al., Reduning Injection prevents carrageenan-induced inflammation in rats by serum and urine metabolomics analysis, Chin. Herb. Med. 14 (2022) 583–591.
[6] S. Ma, S.K. Chowdhury, Data acquisition and data mining techniques for metabolite identification using LC coupled to high-resolution MS, Bioanalysis 5 (2013) 1285–1297.
[7] X. Li, J. Liu, T.T. Zuo, et al., Advances and challenges in ginseng research from 2011 to 2020: the phytochemistry, quality control, metabolism, and biosynthesis, Nat. Prod. Rep. 39 (2022) 875–909.
[8] X.Y. Wang, M.T. Jiang, J. Lou, et al., Pseudotargeted metabolomics approach enabling the classification-induced ginsenoside characterization and differentiation of ginseng and its compound formulation products, J. Agric. Food Chem. 71 (2023) 1735–1747.
[9] X.Y. Zhang, Z.Z. Jiang, L. Zhang, et al., Identification of prototype compounds and their metabolites in rats' serum from Xuefu Zhuyu Decoction by UPLC-Q-TOF/MS, Chin. Herb. Med. 15 (2023) 139–150.

[10] K.Y. Feng, S.M. Wang, L.F. Han, et al., Configuration of the ion exchange chromatography, hydrophilic interaction chromatography, and reversed-phase chromatography as off-line three-dimensional chromatography coupled with high-resolution quadrupole-orbitrap mass spectrometry for the multicomponent characterization of *Uncaria sessilifructus*, J. Chromatogr. A 1649 (2021), 462237.
[11] H.D. Wang, C.X. Zhang, T.T. Zuo, 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. 411 (2019) 7817–7829.
[12] C.X. Zhang, M.Y. Liu, X.Y. Xu, et al., Application of large-scale molecular prediction for creating the preferred precursor ions list to enhance the identification of ginsenosides from the flower buds of *Panax ginseng*, J. Agric. Food Chem. 70 (2022) 5932–5944.
[13] M.X. Sun, X.H. Li, M.T. Jiang, et al., A practical strategy enabling more reliable identification of ginsenosides from *Panax quinquefolius* flower by dimension-enhanced liquid chromatography/mass spectrometry and quantitative structure-retention relationship-based retention behavior prediction, J. Chromatogr. A 1706 (2023), 464243.
[14] Z.X. Yan, R. Yan, Improved data-dependent acquisition for untargeted metabolomics using gas-phase fractionation with staggered mass range, Anal. Chem. 87 (2015) 2861–2868.
[15] F.Y. Zhong, M.Y. Xu, J.J. Zhu, Development and application of time staggered/mass staggered-globally optimized targeted mass spectrometry, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 1120 (2019) 80–88.
[16] Y. Wang, R.B. Feng, R.B. Wang, et al., 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 992 (2017) 67–75.
[17] D. Mehta, S. Scandola, R.G. Uhrig, BoxCar and library-free data-independent acquisition substantially improve the depth, range, and completeness of label-free quantitative proteomics, Anal. Chem. 94 (2022) 793–802.
[18] J.Z. Zhao, Y. Yang, H. Xu, et al., Data-independent acquisition boosts quantitative metaproteomics for deep characterization of gut microbiota, NPJ Biofilms Microbiomes 9 (2023), 4.
[19] N.H. Anh, Y.C. Yoon, Y.J. Min, et al., Caenorhabditis elegans deep lipidome profiling by using integrative mass spectrometry acquisitions reveals significantly altered lipid networks, J. Pharm. Anal. 12 (2022) 743–754.
[20] J. Guo, S. Shen, S.P. Xing, et al., DaDIA: Hybridizing data-dependent and data-independent acquisition modes for generating high-quality metabolomic data, Anal. Chem. 93 (2021) 2669–2677.
[21] Y.X. Qian, D.X. Zhao, H.D. Wang, et al., An ion mobility-enabled and high-efficiency hybrid scan approach in combination with ultra-high performance liquid chromatography enabling the comprehensive characterization of the multi-components from *Carthamus tinctorius*, J. Chromatogr. A 1667 (2022), 462904.
[22] H.D. Wang, H.M. Wang, X.Y. Wang, et al., A novel hybrid scan approach enabling the ion-mobility separation and the alternate data-dependent and data-independent acquisitions (HDDIDDA): Its combination with off-line two-dimensional liquid chromatography for comprehensively characterizing the multicomponents from Compound Danshen Dripping Pill, Anal. Chim. Acta 1193 (2022), 339320.
[23] W.D. Hu, X.Y. Xu, Y.X. Qian, et al., Integration of a hybrid scan approach and in-house high-resolution MS^2 spectral database for charactering the multi-components of Xuebijing Injection, Arab. J. Chem. 16 (2023), 104519.
[24] M. Giera, O. Yanes, G. Siuzdak, Metabolite discovery: Biochemistry's scientific driver, Cell Metab. 34 (2022) 21–34.
[25] H.C. Lan, S.Z. Li, K. Li, et al., *In vitro* human intestinal microbiota biotransformation of nobiletin using liquid chromatography-mass spectrometry analysis and background subtraction strategy, J. Sep. Sci. 44 (2021) 2046–2053.
[26] W.L. Wei, H.J. Li, W.Z. Yang, et al., An integrated strategy for comprehensive characterization of metabolites and metabolic profiles of bufadienolides from *Venenum Bufonis* in rats, J. Pharm. Anal. 12 (2022) 136–144.
[27] J. Wu, W.F. Gou, Z.Y. Wang, et al., Discovery of the radio-protecting effect of *Ecliptae Herba*, its constituents and targeting p53-mediated apoptosis *in vitro* and *in vivo*, Acta Pharm. Sin. B 13 (2023) 1216–1230.
[28] H.Q. Lai, Y. Ouyang, G.H. Tian, et al., Rapid characterization and identification of the chemical constituents and the metabolites of Du-Zhi pill using UHPLC coupled with quadrupole time-of-flight mass spectrometry, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 1209 (2022), 123433.
[29] C.X. Zhang, X.Y. Wang, Z.Z. Lin, et al., Highly selective monitoring of in-source fragmentation sapogenin product ions in positive mode enabling group-target ginsenosides profiling and simultaneous identification of seven *Panax* herbal medicines, J. Chromatogr. A 1618 (2020), 460850.
[30] H.Y. Wang, H.M. Zhao, Y. Wang, et al., Sishen Wan® ameliorated trinitrobenzene-sulfonic-acid-induced chronic colitis *via NEMO*/NLK signaling pathway, Front. Pharmacol. 10 (2019), 170.
[31] F. Chen, Y. Yin, H. Zhao, et al., Sishen pill treatment of DSS-induced colitis *via* regulating interaction with inflammatory dendritic cells and gut microbiota, Front. Physiol. 11 (2020), 801.
[32] X.L. Su, Y.P. Tang, J. Zhang, et al., Curative effect of warming kidney and fortifying spleen recipe on diarrhea-predominant irritable bowel syndrome, J. Tradit. Chin. Med. 33 (2013) 615–619.
[33] L. Liu, S. Wang, Q.X. Xu, et al., Poly-pharmacokinetic strategy represented the synergy effects of bioactive compounds in a traditional Chinese medicine formula, Si Shen Wan and its separated recipes to normal and colitis rats, J. Sep. Sci. 44 (2021) 2065–2077.
[34] G. Paglia, P. Angel, J.P. Williams, et al., Ion mobility-derived collision cross section as an additional measure for lipid fingerprinting and identification, Anal. Chem. 87 (2015) 1137–1144.
[35] J. Liu, Y.T. Li, Y.J. Chen, Advances in high-resolution mass spectrometric-based data-mining technologies for detecting and characterizing the components and metabolites of Chinese materia medica, Acta Pharm. Sin. 56 (2021) 113–129.
[36] F. Fenaille, P. Barbier Saint-Hilaire, K. Rousseau, et al., Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? J. Chromatogr. A 1526 (2017) 1–12.
[37] R. Frański, B. Gierczyk, T. Kozik, et al., Signals of diagnostic ions in the product ion spectra of $[M-H]^-$ ions of methoxylated flavonoids, Rapid Commun. Mass Spectrom. 33 (2019) 125–132.
[38] Y. Gao, S.M. Wu, R.H. Cong, et al., Characterization of lignans in Schisandra chinensis oil with a single analysis process by UPLC-Q/TOF-MS, Chem. Phys. Lipids 218 (2019) 158–167.
[39] Z.W. Zhang, T.Z. Fang, H.Y. Zhou, et al., Characterization of the *in vitro* metabolic profile of evodiamine in human liver microsomes and hepatocytes by UHPLC-Q exactive mass spectrometer, Front. Pharmacol. 9 (2018), 130.
[40] Y. Zhang, J. Yuan, Y. Wang, et al., LC-MS/MS determination and pharmacokinetics study of puerarin and daidzein in rat plasma after oral administration of Gegenqinlian decoction and *Radix* Puerariae extract, Pharmacogn. Mag. 10 (2014) 241–248.
[41] D.W. Lee, Y. Kang, M.J. Kang, et al., Phase I and phase II metabolite identification of rutaecarpine in freshly isolated hepatocytes from male Sprague-Dawley rats, Arch. Pharm. Res. 40 (2017) 972–979.
[42] H. Ren, S. Guo, Y.Y. Zhang, Metabolite identification and metabolic pathway analysis of Bufei Huoxue Capsules in rats, Chin. Tradit. Herb. Drugs 54 (2023) 1051–1063.
[43] H.D. Wang, L. Zhang, X.H. Li, et al., Machine learning prediction for constructing a universal multidimensional information library of *Panax* saponins (ginsenosides), Food Chem. 439 (2024), 138106.
[44] D.H. Ross, L.B. Xu, Determination of drugs and drug metabolites by ion mobility-mass spectrometry: A review, Anal. Chim. Acta 1154 (2021), 338270.
[45] C. Chalet, B. Hollebrands, H.G. Janssen, et al., Identification of phase-II metabolites of flavonoids by liquid chromatography-ion-mobility spectrometry-mass spectrometry, Anal. Bioanal. Chem. 410 (2018) 471–482.
[46] J.-J. Xu, M.-S. Li, Z.-H. Yao, et al., *In vitro* metabolic mapping of neobavaisoflavone in human cytochromes P450 and UDP-glucuronosyltransferase enzymes by ultra high-performance liquid chromatography coupled with quadrupole time-of-flight tandem mass spectrometry, J. Pharm. Biomed. Anal. 158 (2018) 351–360.