ROYAL SOCIETY OF CHEMISTRY

PAPER

Check for updates

Improving natural products identification through molecular features-oriented precursor ion selection and targeted MS/MS analysis: a case study of Zhi-Ke-Yang-Yin capsule†

Hongping Wang, [ID] [a] Qiong Yin,[a] ZiJian Wang,[a] Ping Peng,[a] Chunlan Fan,[c] Run Zhang,[b] Chen Zhao[a] and Zhaozhou Lin[c]

Chemical substance identification is an indispensable step in research on therapeutic materials based on traditional Chinese medicine and its formulas. The successful characterization of chemical substances mainly relies on high-quality MS/MS spectra. However, to date, relatively few studies have specifically addressed the issues of improving the acquisition of MS/MS spectra of compounds for characterization. The current auto-MS/MS mode, where the precursor ions are selected depending on their signal intensity, encounters a drawback when the sample contains many overlapping signals, leading to compounds with a lower or much lower abundance missing identification. To solve this problem, a strategy in which molecular features oriented precursor ion selection was followed by targeted MS/MS analysis for structure elucidation was proposed. The precursor ions were selected according to their first and second molecular features, namely m/z and retention time, irrespective of their intensities. By performing targeted MS/MS analysis, the MS/MS spectra of many more compounds of interest can be obtained, leading to an improvement in natural product identification. As an example, the chemical substances in the Zhi-Ke-Yang-Yin extract were analyzed using this strategy, and as a result, 431 ingredients were tentatively characterized, including both known and unknown or new compounds.

1. Introduction

The traditional Chinese medicine (TCM) formula (TCMF), which is typically composed of several herbal medicines, contains hundreds or even thousands of chemical substances. Studying these chemical substances plays an extremely important role in clarifying the therapeutic basis of TCMF. Many researchers have attempted to clarify the chemical substances in TCMF using LC-MS, and a series of results have been reported;[1–7] however, no new analytical approaches have been proposed in these reports, and only a small part of the constituents has been identified. To clarify the holistic chemical substances in TCMF, some approaches, such as diagnostic ion-guided network bridging,[8] diagnostic fragment-ion-based extension strategy,[9] mass defect filtering,[10] characteristic components dataset (CCDS),[11] and utilizing the UNIFI informatics platform for computer-assisted UPLC/QTOF MS data analyses[12] have been developed. Using these strategies, the number of compounds identified from TCMF has increased, and up to 190 ingredients were identified from the Niuhuang Shangqing pill.[11] However, this is not sufficient because hundreds of compounds are usually present in TCMF. The successful identification of the constituents mentioned above critically relied on high-quality MS/MS spectra, and some compounds were not characterized mainly because of the lack of their MS/MS spectra. However, relatively few studies have specifically addressed the issues of improving the acquisition and characterization of MS/MS spectra. The current auto-MS/MS mode is a type of data-dependent acquisition, where precursor ions are usually selected according to signal intensity, which implies that the MS/MS spectra of ingredients with lower or much lower abundance will not be obtained when samples contain many overlapping signals. Therefore, this study aimed to improve the acquisition of MS/MS spectra for many more compounds and characterize them during TCMF UPLC-MS analysis.

Each compound has the features of mass to charge ratio (m/z), retention time, and intensity in a full scan. The m/z value, which is the most valuable information in a full scan, namely, the first molecular feature, was used to determine the molecular weight and deduce the molecular formula of a compound when detected by high-resolution mass spectrometry. Usually, the

[a] Scientific Research Institute of Beijing Tongrentang Co., Ltd., Beijing 100011, China. E-mail: sungirl9626@163.com; Fax: +86 10 87632655; Tel: +86 10 87632655

[b] Beijing Tongrentang Technology Development Co., Ltd., Beijing 100079, China

[c] Beijing Zhongyan Tongrentang Pharmaceutical R & D Co., Ltd., Beijing 100000, China

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2ra06063a

first molecular features of compounds in the same chemical cluster have a general range, in which those of known and unknown compounds are included. Except for the first molecular feature, retention time, namely the second molecular feature, is another important piece of information in a full scan, which is used to determine when to perform MS/MS analysis. Intensity, namely the third molecular feature, which is above a set threshold, cannot be considered sampling noise, but related to a compound. Compared to the third molecular feature, the first and second molecular features are much more important and are crucial for constructing the precursor ion list. Once they are obtained, targeted MS/MS can be performed for structure elucidation.

Therefore, in our study, a strategy involving molecular feature-oriented precursor ion selection and targeted MS/MS analysis for structure elucidation was proposed. For example, the ingredients of the Zhi-Ke-Yang-Yin (ZKYY) capsule, which is composed of 7 TCMs: *Renshen*, *Huangqi*, *Shengdihuang*, *Shanzhuyu*, *Shanyao*, *Tianhuafen*, and *Zhuyizang*, were analyzed using the proposed method, and 431 compounds were successfully identified.

2. Methodology

2.1 Reagents, reference standards, and materials

LC-MS-grade formic acid was purchased from Fisher Scientific (Fair Lawn, USA), and LC-MS-grade acetonitrile and methanol were obtained from Merck (Germany). Distilled water was obtained from Watson.

A total of 49 reference standards (see Fig. S1†) including ginsenoside Re$_3$ (**R2**), ginsenoside Re$_4$ (**R3**), 20-*gluco*-ginsenoside Rf (**R5**), ginsenoside Re$_1$ (**R7**), notoginsenoside R$_1$ (**R9**), ginsenoside Re$_2$ (**R10**), ginsenoside Rg$_1$ (**R16**), ginsenoside Re (**R17**), ginsenoside Rf (**R42**), notoginsenoside R$_2$ (**R47**), ginsenoside Ra$_2$ (**R55**), ginsenoside Ra$_3$ (**R58**), ginsenoside Rb$_1$ (**R59**), 20(*S*)-ginsenoside Rg$_2$ (**R60**), 20(*S*)-ginsenoside Rh$_1$ (**R61**), 20(*R*)-ginsenoside Rg$_2$ (**R68**), ginsenoside Rc (**R70**), 20(*R*)-ginsenoside Rh$_1$ (**R71**), ginsenoside Ra$_1$ (**R72**), ginsenoside Ro (**R74**), ginsenoside Rb$_2$ (**R81**), ginsenoside Rb$_3$ (**R82**), quinquenoside R$_1$ (**R86**), ginsenoside Rs$_2$ (**R91**), ginsenoside Rd (**R92**), ginsenoside Ro methyl ester (**R113**), ginsenoside F4 (**R129**), ginsenoside Rg$_6$ (**R133**), 20(*S*)-ginsenoside Rg$_3$ (**R139**), 20(*R*)-ginsenoside Rg$_3$ (**R141**), 20(*S*)-ginsenoside Rh$_2$ (**R147**), 20(*R*)-ginsenoside Rh$_2$ (**R148**), astragaloside IV (**H13**), astragaloside II (**H18**), loganic acid (**S8**), morroniside (**S16**), cornine (**S24**), sweroside (**S27**), cornuside I (**S42**), naringenin (**S87**), catalpol (**D1**), geniposidic acid (**D11**), 8-epiloganic acid (**D16**), genipin 1-gentiobioside (**D18**), geniposide (**D20**), rhamnocitrin (**H62**), quercetin (**H63**), oleanolic acid (**H93**), and ursolic acid (**H94**) were either purchased from Shanghai Yuanye Bio-Technology Co., Ltd. (Shanghai, China), as well as the National Institute for the Control of Pharmaceutical and Biological Products (Beijing, China), or gifts from the State Key Laboratory of Natural and Biomimetic Drugs, Department of Natural Medicines, School of Pharmaceutical Sciences, Peking University. The purity of all the reference standards was >98%. ZKYY capsule powder was

supplied by the Scientific Research Institute of Beijing Tongrentang Co., Ltd.

2.2 Standard solution and sample preparation

The 49 reference standards were used only to confirm the accuracy of characterization, rather than quantification. Each reference standard (1.0 mg) was separately dissolved in 1 mL of 70% methanol to obtain its stock solution (1.0 mg mL^{-1}). Ten to twenty microliters of each stock solution were mixed to obtain a mixture stock solution. The samples were then stored at 4 °C until analysis. ZKYY capsule powder (1.0 g) was ultrasonically extracted for 30 min with 20 mL of 70% methanol at 25 °C. The extracted solution was then filtered through a filter paper. This extraction was repeated twice. The filtrate was combined and evaporated to dryness using a rotary evaporator at 40 °C. The residue (0.3 g) was then dissolved in 5 mL of 70% methanol and filtered through a 0.22 μm nylon filter membrane before analysis.

2.3 Chromatographic and mass spectrometric conditions

A Vanquish™ Flex UHPLC system (Thermo Scientific, USA) equipped with a binary pump and a thermostated column compartment was used to separate the multiple components. The optimized chromatographic conditions were detailed as follows: A Waters ACQUITY UPLC® BEH C$_{18}$ column (2.1 × 100 mm, 1.7 μm) coupled with a ACQUITY UPLC® BEH C$_{18}$ VanGuard™ Pre-Column (2.1 × 5 mm, 1.7 μm) was employed to perform the chromatographic separation of ZKYY extract using mobile phase A (0.1% formic acid/water, v/v) and mobile phase B (acetonitrile) by the following gradient elution program: 0–7 min, 2–20% B; 7–10 min, 20–25% B; 10–20 min, 25–40% B; 20–25 min, 40–65% B; 25–30 min, 65–95% B; 30–31 min, 95–95% B. The flow rate was 0.3 mL min^{-1}, and the temperature was set at 35 °C. The injection volume was 2 μL.

High-accuracy mass spectrometry data were recorded on an Orbitrap Exploris 240 mass spectrometer (Thermo Scientific, USA) equipped with a Heated ESI source. The instrument was operated in the negative-ion mode. The optimized MS parameters were as follows: ion spray voltage: 2500 V, sheath gas: 5.08 L min^{-1}, auxiliary gas: 9.37 L min^{-1}, ion transfer tube temperature: 320 °C, vaporizer temperature: 350 °C, scan range (*m/z*):150–2000, and collision-energy voltage: 35 V. The full scan was operated at a mass resolution of 60 000 whereas the MS^2 scan was operated at a mass resolution of 15 000. An internal calibration source, Thermo Scientific EASY-IC™, was adopted to calibrate the entire mass range.

2.4 Data acquisition and data mining

Thermo Xcalibur software (version 4.5) was used for data acquisition, whereas Free Style™ 1.8 SP1 and Compound Discoverer™ softwares (Thermo Scientific™, version 3.2.0.421) were used for data processing and mining. The negative-ion mode was adopted for data acquisition because the compounds in the ZKYY extract exhibited high ion responses in this mode. The ZKYY extract was first analyzed in full scan mode to minimize signal loss, then targeted MS/MS analysis was
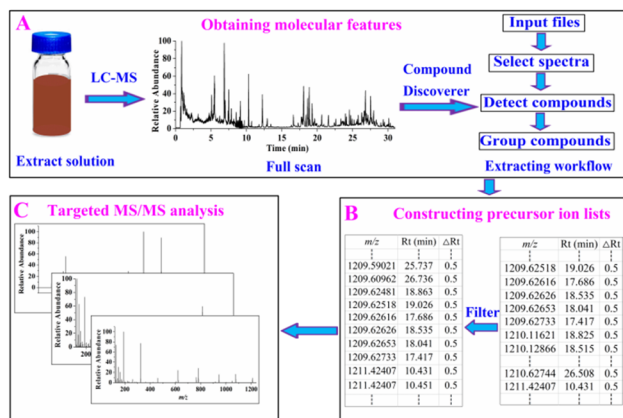
Fig. 1 The workflow of molecular features orientated precursor ions selection and targeted MS/MS analysis for structure elucidation: (A) obtaining molecular features; (B) constructing precursor ion lists; (C) targeted MS/MS analysis.

performed after constructing the precursor ion lists. The proposed data mining strategy for molecular feature-oriented precursor ion selection and targeted MS/MS analysis for structure elucidation mainly involved three steps (as shown in Fig. 1):

2.4.1 Obtaining molecular features. In this strategy, all compounds in a full scan were extracted using Compound Discoverer™ software, and the effective extracting workflow mainly included "input files → select spectra → detect compounds → group compounds". In the step of "input files," the spectra file to be processed was selected, whereas in the step of "select spectra," the spectra properties, such as the lower and upper retention time limit as well as polarity mode, were selected for further processing. All compounds in the data were extracted using component elucidator algorithm in the "detect compounds" step. In this step, a series of parameters were set: mass tolerance: 5 ppm, intensity tolerance: 30, S/N threshold: 3, minimum peak intensity: 100 000, extracted ions: $[M - H]^-$, $[M - H + HAc]^-$, $[2M - H]^-$, $[2M - H + HAc]^-$, minimum element composition: CHO, and maximum element composition: C90H190O90. The same substances detected in different addition methods were grouped by molecular weight (mass tolerance: 5 ppm) as well as retention time (RT tolerance: 0.1 min) across all files in the "group compounds" step. By running this workflow, all compounds with intensities above the set threshold were extracted and presented as m/z (the first molecular feature), retention time (the second molecular feature), and peak area. The first and second molecular features are essential for constructing a precursor ion list.

2.4.2 Construction of precursor ion lists according to molecular features. The obtained molecular features of the compounds included all necessary information to construct the precursor ion lists. However, before constructing the precursor ion lists, the ions of uninterest could be manually eliminated. The successfully constructed lists of precursor ions included m/z, retention time (R_t), and retention time window (ΔR_t).

2.4.3 Targeted MS/MS analysis based on precursor ion lists and structure elucidation. With these precursor ion lists in hand, we tested the sample using a precursor ion-directed LC-

MS/MS method. The ingredients were characterized according to the MS/MS spectra obtained.

3. Results and discussion

3.1 Obtaining molecular features

In our experiment, the full scan data of the ZKYY extract were selected for processing, and the entire run time of the spectra as well as the negative-ion mode were selected for further processing. When detecting compounds, $[M - H]^-$, $[M - H + Hac]^-$, $[2M - H]^-$, and $[2M - H + HAc]^-$ ions were selected as extracted ions and detected by a mass tolerance of 5 ppm as well as a threshold intensity of 100 000. Compounds with different addition methods were consolidated by a mass tolerance of 5 ppm and retention time tolerance of 0.1 min. Running the extracted workflow produced 35 131 ingredients presented by m/z, retention time, and peak area.

3.2 Constructing precursor ion lists for compounds of interest

The constituents of TCMF mainly originate from a single herb. Theoretically, the components of ZKYY mainly include ginsenosides from *Renshen*,[13] astragalus saponins from *Huangqi*,[14] iridoids, ionones, phenylethanoid glycosides, phenylpropanoids, organic acids from *Shengdihuang* and *Shanzhuyu*,[15,16] polyphenols from *Shanyao*,[17] and triterpenoids from *Tianhuafen*.[18] Flavonoids, which are widely present in nature, are also found in most of the herbal medicines mentioned above.[14,16–18] The above chemical clusters have been reported to have some biological activities; therefore, we were more interested in them during the clarification of the chemical substances in ZKYY.

Ideally, we can construct different precursor ion lists of different chemical clusters for convenient characterization. However, by searching the PubChem database and literature, we found that some different chemical clusters shared some common molecular features, not only in m/z ranges, but also in retention time ranges. We found that the molecular weights of ginsenosides, astragalus saponins, and triterpene glycosides were all greater than 600, whereas those of iridoids, ionones, flavonoids, phenylpropanoids, organic acids, polyphenols, and triterpenoids were all lower than 600. However, for phenylethanoid glycosides, the molecular weights were greater than 600 or lower than 600. Therefore, two types of precursor ion lists were constructed: compounds with molecular weights lower than 600 were selected in the first type of the precursor ion list to characterize iridoids, ionones, flavonoids, phenylpropanoids, organic acids, polyphenols, triterpenoids, and phenylethanoid glycosides, whereas the constituents with molecular weights greater than 600 were assigned to the second type of precursor ion list to include ginsenosides, astragalus saponins, triterpene glycosides, and phenylethanoid glycosides.

It was worth mentioning that the precursor ions of known and unknown compounds belonging to the same chemical cluster were all included in the constructed lists. If an individual compound belongs to the first list according to its

chemical cluster but is assigned to the second list on the basis of its molecular weight, it will be characterized and *vice versa*.

In addition, from reported literature,[13–18] we found the molecular formulas of our concerned compounds are composed of only three elements, namely C, H, and O. In theory, because no nitrogen atom was involved in the molecular formulas, the precursor ions of our compounds had odd nominal masses in the negative-ion mode. Therefore, ions with even normal masses were regarded as redundant data and eliminated to obtain a high-quality dataset.

3.3 Performing targeted MS/MS acquisition

After constructing the precursor ion lists, we performed targeted MS/MS analysis. The first and second types of precursor ion lists were separately subjected to MS/MS acquisition. Auto-MS/MS acquisition was performed under the same chromatographic and mass spectrometric conditions, except for the precursor selection. Note that the auto-MS/MS mode always selects the three most intensity m/z of a full scan, whereas our targeted mode was precursor ion-directed analysis.

3.4 Evaluation of the efficacy of the proposed strategy

To evaluate the efficacy of the proposed strategy, the number of compounds characterized using this method and the auto-MS/MS mode were compared. We found that, in any chemical cluster, many more compounds were identified by the molecular features-oriented precursor ion selection mode (shown in Fig. 2), and auto-MS/MS-derived ingredients were only part of them. Furthermore, we believe that the inclusion of low-intensity precursor ions resulted in the difference in the number of identified compounds, which was further confirmed by Fig. 5C and D, as well as Fig. 6C and D.

3.5 Identification of the components in ZKYY

After obtaining the MS/MS spectra, we identified the chemical substances in crude ZKYY extract. The constituents belonging to different chemical clusters could be easily differentiated
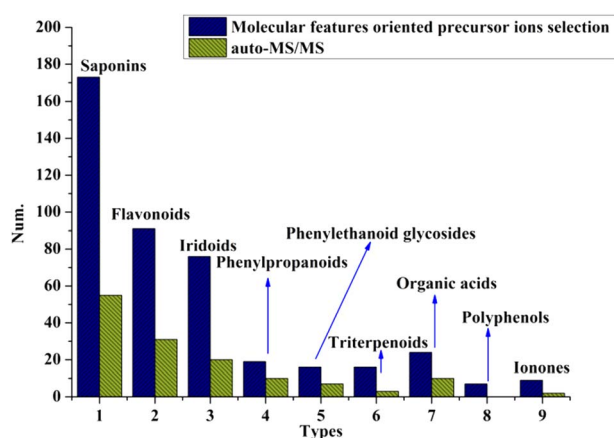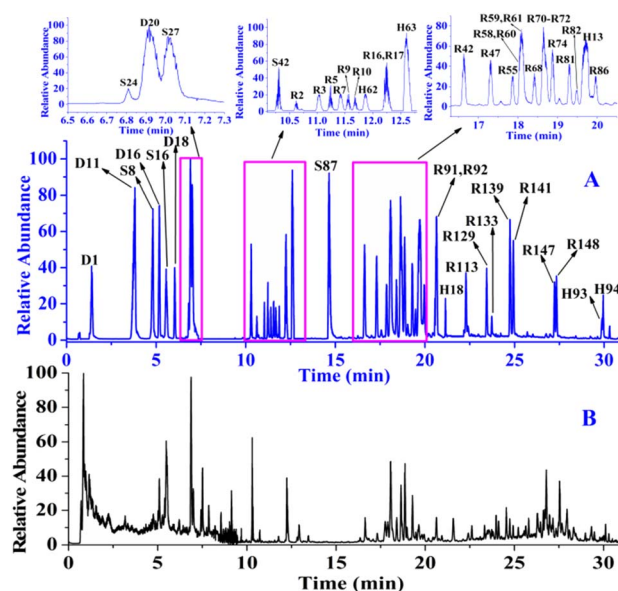

Fig. 3 Total ion chromatograms of the 49 reference standards (A) and ZKYY extract (B) in negative-ion mode.

because of their different fragmentation pathways as well as their different diagnostic ions, whereas the components belonging to the same chemical clusters usually have similar fragmentation patterns and the same diagnostic ions, and they could be successfully characterized either by comparison with the reference standards or the elution order information in the reversed-phase column described in literature and the polar size of the compounds. In addition, it is generally believed that a compound with high abundance in a full scan stands for a relatively high amount in herbal medicine, which is relatively easier to isolate than the compound with much lower abundance or amount, according to which the isolated known compounds and their isomers were tentatively identified. The total ion chromatograms of the reference standards and the ZKYY extract are shown in Fig. 3A and B, respectively. Based on the proposed strategy, a total of 431 compounds, including 173 saponins, 91 flavonoids, 76 iridoids, 19 phenylpropanoids, 16 phenylethanoid glycosides, 16 triterpenes and glycosides, 9 ionones, 24 organic acids, and 7 polyphenols, were characterized, and 49 of these compounds were confirmed by comparison with reference standards.

3.5.1 Identification of saponins. Protopanaxatriol, protopanaxadiol, and oleanane types of ginsenosides from *Renshen* yielded diagnostic aglycone ions at m/z 475.3787, 459.3848, and 455.3510 in their MS/MS spectra, respectively, by successive losses of a series of saccharide moieties, such as the glucosyl group (Glc, 162 Da), rhamnosyl (Rha, 146 Da), and xylosyl/ arabinoyl (Xyl/Ara, 132 Da) from the $[M - H]^-$ ion, whereas saponins in *Huangqi* produced the diagnostic ion at m/z 489.3580. Diagnostic ions and fragmentation pathways were adopted to deduce not only known but also unknown saponins. A total of 148 ginsenosides [59 known ginsenosides and 89 unknown ginsenosides, including 50 isomers, 14 acetylation, 16 malonylation, 4 acetylation + malonylation of known
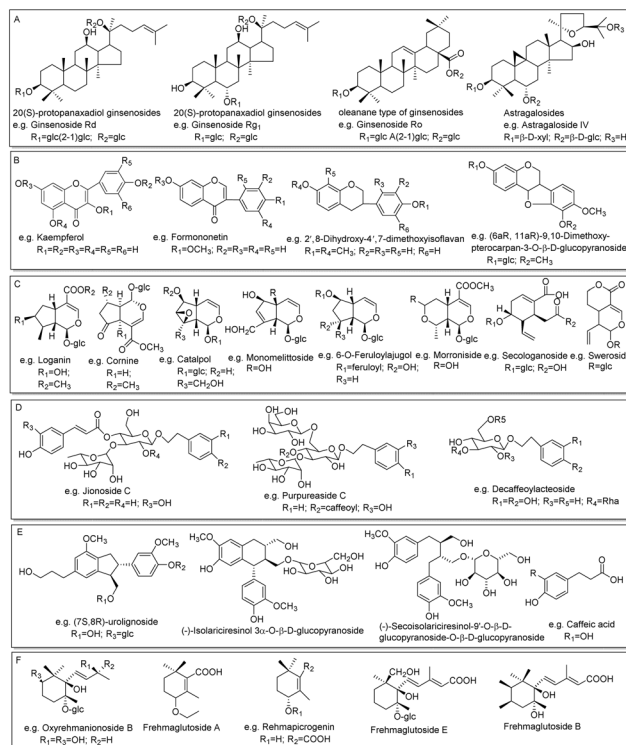

Fig. 2 Numbers of compounds identified based on molecular features oriented precursor ions selection and auto-MS/MS mode.

Fig. 4 The main structure types of compounds identified in this work: sapoinins (A), flavonoids (B), iridoids (C), phenylethanoid glycosides (D), phenylpropanoids (E), and ionones (F).

Fig. 5 An example of identification ginsenosides with low abundance. (A) In the constructed precursor ion list, 6 compounds (R49, R55, R63, R72, R76, and R78) showed the precursor ion at m/z 1209.6248–1209.6273; (B) performing targeted MS/MS analysis to obtain the fragmentation ions, such as R55 and R78; (C) the 6 compounds were covered by other high intensity compounds and their relative abundances were no more than 1.0% in a full scan; (D) the relative abundances of these 6 ingredients were much lower than those of others in MS spectra, such as R78.

ginsenosides, and 2 glycosylation of protopanaxadiol (**R37** and **R131**) as well as 3 glycosylation protopanaxatriol components (**R6**, **R12** and **R14**)] and 25 astragalus saponins (13 known astragalus saponins, 11 isomers, and 1 acetylation of known astragalus saponins) were characterized. The LC-MS traces are shown in Fig. S3 and S4.† The main structural types of saponins are shown in Fig. 4A, and the details of all the identified saponins are shown in ESI Table S1† with mass deviations of no greater than 2.01 ppm. Among them, 33 ginsenosides and an astragalus saponin (marked in ESI Table S1†) matched well with the reference standards in terms of retention time, fragmentation ions, and pathways.

For instance, 6 constituents (**R49**, **R55**, **R63**, **R72**, **R76** and **R78**) constructed in the list showed their precursor ions at m/z 1209.6248–1209.6273 (shown in Fig. 5A) with mass deviations of −1.65–0.41 ppm indicating that their molecular formula was $C_{58}H_{98}O_{26}$. Targeted MS/MS analysis showed the same fragmentation ions at m/z 1077.5858, 945.5461, 783.4919, 621.4398, and 459.3864, yielded by successive losses of Xyl, Ara, Glc, Glc, and Glc from the $[M - H]^-$ ion (shown in Fig. 5B). **R55** and **R72** were unambiguously identified as ginsenosides Ra$_2$ and Ra$_1$, respectively, by comparison with the reference standards, and the remaining four compounds **R49**, **R63**, **R76** and **R78** were identified as their isomers. As shown in Fig. 5C, we found that the relative abundances of the extracted ion chromatograms were no more than 1.0% in the full scan, and the signals were covered by other relatively high-intensity components. In this case, we deduced that the intensities of these 6 components
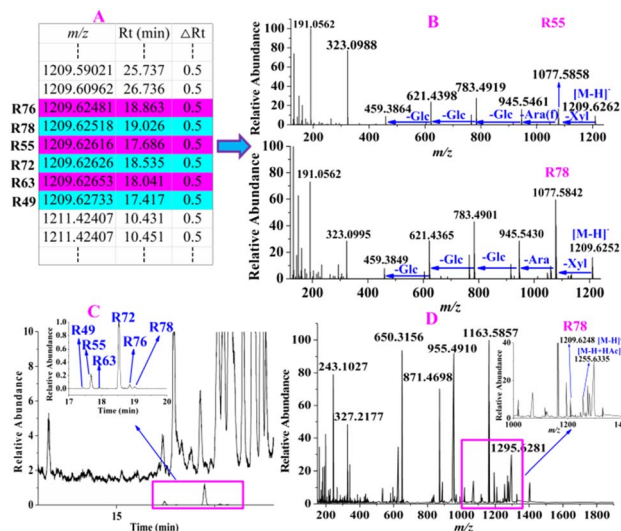
were lower or much lower than those of the others in the MS spectrum, as shown in Fig. 5D, using **R78** as an example. As shown in Fig. 5D, the relative abundance of the precursor ion of **R78** was approximately 10%, whereas those of many compounds were greater than 10%, leading to the precursor ion of **R78** not being selected in the auto-MS/MS mode because of the much lower intensity, but it was selected in our experiment. **R78** was determined based on its MS/MS spectrum. Five compounds (**R6**, **R12**, **R14**, **R37**, and **R131**), which were regarded as potential new compounds, were characterized. **R6**, **R12** and **R14**, which were eluted at 11.31, 11.89, and 12.08 min, respectively, showed not only the same fragmentation pathway but also the same fragmentation ions at m/z 945.5485, 783.4889, 637.4330, and 475.3810, respectively, by successively losses of Glc, Glc, Glc and Rha from $[M - H]^-$ ion. The diagnostic ion at m/z 475.3810 indicated that the compound was a protopanaxatriol-type ginsenoside. However, no known compounds matched them well; thus, **R6**, **R12**, and **R14** were tentatively identified as potential new compounds, protopanaxatriol + 3Glc + Rha. Similarly, **R37** was identified as protopanaxadiol + 5Glc, whereas **R131** was characterized as protopanaxadiol + Rha + 2Glc. In addition to the ginsenosides mentioned above, other known compounds and their isomerized, acetylated, and malonylated compounds were tentatively identified.

3.5.2 Identification of flavonoids. As reported in literature,[19] except for retro Diels–Alder fragmentation, flavonoids always produce fragmentation ions by the loss of a series of small molecular units, such as H_2O (18 Da), CO (28 Da), and CO_2 (44 Da) in their MS/MS spectra in negative-ion mode. The ions yielded by losses of CH_3 (15 Da) and OCH_2 (30 Da) are

found when there is a substituent of OCH_3 in their structures, whereas the ion produced by the loss of neutral molecular C_3O_2 (68 Da) is found when there is 5,7-dihydroxy in the A-ring in the MS/MS spectrum. However, the C-ring always loses CHO (29 Da) and C_2H_2O (42 Da). Flavonoids existing in the form of glycosides always lose their sugar moiety first, followed by a series of neutral losses of small molecules in their MS/MS spectrum. These fragmentation pathways were used to deduce the flavonoids in ZKYY extract, and a total of 91 flavonoids (40 known compounds and their 51 isomers) were characterized. The LC-MS traces are shown in Fig. S5.† The main structure types of the flavonoids are shown in Fig. 4B, and the details of all the identified flavonoids are shown in ESI Table S2.† The measured masses were all within a mass deviation of 2.21 ppm, which made the characterization more reliable. Three deduced flavonoids, **H62**, **H63**, and **S87**, were confirmed by comparison with the reference standards.

For instance, the precursor ions of 12 constituents (**H40**, **H43**, **H50**, **H52**, **H62**, **H67**, **H71**, **H75**, **H83**, **H86**, **H89**, and **H91**) were extracted at m/z 299.0556–99.0562 (shown in Fig. 6A) with a mass deviation of no more than 2.01 ppm, indicating that their molecular formula was $C_{16}H_{12}O_4$. By performing target MS/MS analysis, they shared the same fragmentation ions at m/z 284.0328 and 269.0454, which were yielded by the loss of CH_3 and OCH_2 from the $[M - H]^-$ ion, respectively. In addition, m/z 256.0379 was produced by the elimination of CO from m/z 284.0328 (Fig. 6B). **H62** was unambiguously identified as rhamnocitrin when compared with the reference standard. Therefore, the remaining 11 compounds were tentatively characterized as isomers. When extracting their ion
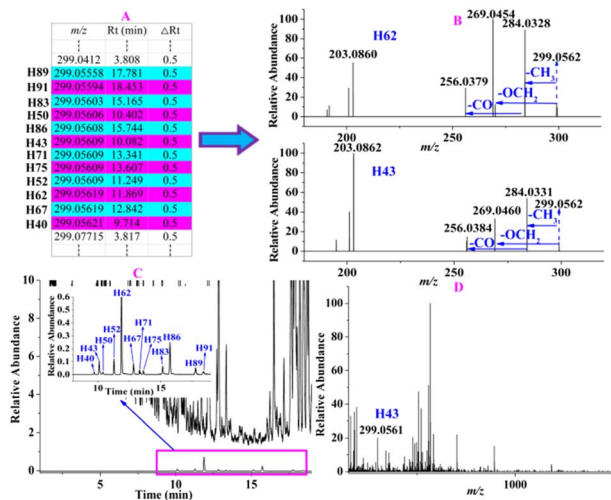

Fig. 6 An example of characterization flavonoids with low intensity. (A) In the constructed precursor ion list, twelve compounds (**H40**, **H43**, **H50**, **H52**, **H62**, **H67**, **H71**, **H75**, **H83**, **H86**, **H89**, and **H91**) were extracted at m/z 299.0556–299.0562; (B) performing targeted MS/MS analysis to obtain the fragmentation ions, such as **H43** and **H62**; (C) the twelve compounds were covered by other high intensity ingredients and exhibited relative abundances of not more than 0.7% in a full scan; (D) the relative abundances of the precursor ions of these twelve ingredients were lower or much lower than those of others in MS spectra, such as **H43**.

chromatography in the full scan of ZKYY extract, we found that the 12 compounds were of much lower abundance and were overlapped by other high-intensity compounds (shown in Fig. 6C), leading to a much lower relative abundance than others in the MS spectra, taking **H43** as an example (shown in Fig. 6D). Their precursor ions were not selected in the auto-MS/MS mode but were selected in our experiment. The precursor ions of **H44**, **H54**, **H56**, and **H74** were extracted at m/z 461.1086–461.1090 with a mass deviation of 1.30 ppm, indicating that their molecular formula was $C_{22}H_{22}O_{11}$. In their MS/MS spectrum, the ion observed at m/z 299.0558 indicates that the Glc unit was lost from the $[M - H]^-$ ion. In addition, the remaining fragmentation ions, produced from m/z 299.0558 and observed at m/z 284.0324, 269.0450, and 256.0392, were the same as those of rhamnocitrin. When extracting their ion chromatograms, the relative abundance of **H74** was much higher than those of **H44**, **H54**, and **H56**. Therefore, **H74** was tentatively identified as rhamnocitin-3-O-β-D-glucopyranoside, a known compound isolated from *Huangqi*,[14] whereas **H44**, **H54**, and **H56** were identified as its isomers. Similarly, other known compounds and their isomers were tentatively assigned.

3.5.3 Identification of iridoids. The iridoids produced fragmentation ions through a series of losses of small molecules of CH_3 (15 Da), H_2O (18 Da), CO_2 (44 Da), and Glc (162 Da) in their MS/MS spectra in the negative-ion mode.[20] When a substituent of CH_2OH is present in their structures, the ion yielded by the loss of CH_2O (30 Da) can also be observed. All fragmentation pathways were used to deduce the iridoids in the ZKYY extract. Seventy-six iridoids, including 51 known and 25 isomers, were characterized. The LC-MS traces are shown in Fig. S6.† The main structural types of iridoids are shown in Fig. 4C, and the details of all identified iridoids are shown in ESI Table S3.† Ten deduced iridoids, including **D1**, **D11**, **D16**, **D18**, **D20**, **S8**, **S16**, **S24**, **S27**, and **S42** were confirmed by comparison with reference standards.

For instance, the precursor ions of **D1** and **D3** were extracted at m/z 361.1136 and 361.1137, respectively, and their mass deviations were no more than 0.55 ppm, suggesting that their molecular formula was $C_{15}H_{22}O_{10}$. The ions observed at m/z 199.0604 and 169.0508 indicate that Glc and CH_2O were successively eliminated from the $[M - H]^-$ ion. **D1** was deduced and confirmed to be catalpol, whereas **D3** was characterized as monomelittoside. The precursor ions of **D7** and **D11** were extracted at m/z 373.1140 and 373.1138, respectively, within a mass deviation of 1.34 ppm indicating that their molecular formula was $C_{16}H_{22}O_{10}$. The fragmentation ions at m/z 329.0871 and 167.0349 indicate that CO_2 and Glc were successively eliminated from the $[M - H]^-$ ion. **D11** was deduced and confirmed as geniposidic acid, whereas **D7** was tentatively assigned as a known compound, gardoside.[15] **D14** and **D16** shared not only the same precursor ions at m/z 375.1291, but also the same fragmentation ions at m/z 213.0776 and 169.0872, which were attributed to successive losses of Glc and CO_2 from the $[M - H]^-$ ion. **D16** was eluted and confirmed as 8-epilloganic acid, whereas **D14** was deduced to be mussaenosidic acid, a known compound isolated from *Shengdihuang*.[15] The extracted precursor ions of **S1**, **S9**, **S27**, and **S37** were at m/z 357.1188–

357.1189 within a mass deviation of 0.84 ppm, indicating that their molecular formula was $C_{16}H_{22}O_9$. Targeted MS/MS analysis revealed fragmentation ions at m/z 195.0663 and 177.0559, suggesting that Glc and H_2O were successively lost from their $[M - H]^-$ ions. **S27** was characterized and confirmed as sweroside, and the remaining compounds, **S1**, **S9**, and **S37** were tentatively assigned as isomers. Similarly, other known compounds and their isomers were tentatively characterized.

3.5.4 Identification of phenylethanoid glycosides, phenylpropanoids, and ionones. Phenylethanoid glycosides produce different fragmentation pathways owing to the presence of different substituents in their structures. For instance, caffeoyl phenylethanoid glycosides always generate a series of fragmentation ions correlated with caffeoyl moiety ($C_9H_6O_3$, 162 Da), whereas coumaroyl phenylethanoid glycosides usually produce a series of fragmentation ions related to feruloyl moiety ($C_{10}H_8O_3$, 176 Da) in the negative-ion mode.[21] These lost characteristic ions provide important structural information for the effective characterization of phenylethanoid glycosides. In addition, neutral losses of Glc and H_2O were always observed in their MS/MS spectra. According to the fragmentation pathways mentioned above, 16 phenylethanoid glycosides, including 14 known compounds and two isomers, were identified in the ZKYY extract with mass deviations of 1.30 ppm. The main structural types of phenylethanoid glycosides are shown in Fig. 4D, and the details of all identified phenylethanoid glycosides are shown in ESI Table S4.† The precursor ions of **D49** and **D51** were extracted at m/z 623.1960 and 623.1971, respectively, indicating that they had the same molecular formula, $C_{29}H_{36}O_{15}$. Targeted analysis yielded the same fragmentation ions at m/z 461.1664 and 315.1078, suggesting that caffeoyl and Rha were successively eliminated from the $[M - H]^-$ ion. In addition, the ion at m/z 161.0245 confirmed that caffeoyl moiety was present in their structures and was eliminated in the MS/MS analysis. The retention time of verbascoside was in front of that of isoverbascoside on a reversed phase column,[22] therefore, we concluded that **D49** was verbascoside and **D51** was isoverbascoside. The precursor ion of **D55** was extracted at m/z 651.2289, indicating that its molecular formula is $C_{31}H_{40}O_{15}$. In the MS/MS spectrum, the fragmentation ions at m/z 475.1840 and 329.1229 were produced by successive losses of feruloyl and Rha from $[M - H]^-$ ions, respectively, which was confirmed by the ion at m/z 175.0407. Therefore, **D55** was tentatively deduced to be a known compound martynoside.[15,16] The precursor and fragmentation ions of **D56** were the same as those of **D55**, and **D56** was tentatively assigned as the martynoside isomer, but with a much lower abundance than that of **D55**. Similarly, other known phenylethanoid glycosides and their isomers were characterized.

The fragmentation pathways of phenylpropanoids and ionones are mainly losses of a series of small molecules, such as CH_3, CO, CO_2, H_2O, OCH_2, OCH_3, and Glc,[23] which were used to deduce the phenylpropanoids and ionones in the ZKYY extract. For example, the precursor ions of **S55**, **S59**, and **S60** were extracted at m/z 523.2184–523.2187, with mass deviations of no more than 1.53 ppm, indicating that their molecular formula was $C_{26}H_{36}O_{11}$. In the MS/MS spectrum, they all produced the same fragmentation ions at m/z 361.1659, 346.1414, and 315.1249 by successive loss of Glc, CH_3 and OCH_3, respectively. Thus, **S60** was tentatively assigned as (−)-secoisolariciresinol-9′-O-β-D-glucopyranoside, a known compound isolated from *Shanzhuyu*,[16] whereas the other two compounds, **S55** and **S59**, were characterized as isomers with much lower abundances than that of **S60**. Similarly, other 17 known phenylpropanoids and ionones, as well as 8 of their isomers, were tentatively characterized (shown in ESI Tables S5 and S6†). The main structural types of phenylpropanoids and ionones are shown in Fig. 4E and F, respectively.

In addition to the compounds mentioned above, 16 triterpenes and glycosides (shown in ESI Table S7†), 24 organic acids, and 7 polyphenols (shown in ESI Table S8†) were tentatively characterized. The LC-MS traces are shown in Fig. S7.†

4. Conclusions

To improve structure elucidation, a strategy in which molecular features-oriented precursor ion selection and targeted MS/MS analysis was proposed in our study. Using this strategy, a total of 431 ingredients, including known and unknown compounds, irrespective of their intensities were tentatively characterized. Our results showed that more compounds were identified by our strategy than those elucidated by the auto-MS/MS mode. The auto-MS/MS mode, where the precursor ions are selected depending on their signal intensity, plays a substantial role in the chemical characterization of TCM or TCMF. However, it encounters a drawback when many signals overlap and the intense signals do not represent the ingredients of interest. In contrast, our strategy overcomes this drawback and has two advantages relative to the auto-MS/MS mode. First, the precursor ions of the constituents were selected irrespective of their intensity so that the concerned compounds with not only high but also low abundance could be identified. However, in the auto-MS/MS mode, compounds of interest with lower or much lower abundance were not identified because they were covered or overlapped by some compounds with high abundance. The second advantage was to avoid interference from unwanted compounds. In auto-MS/MS mode, the three most intense ions are usually selected; however, when the ion intensities of unconcerned compounds meet the requirements (among the three most intense), unconcerned compounds will be selected for MS/MS analysis. To some extent, this leads to missing compounds being selected and unidentified. Nevertheless, in our strategy, the constructed precursor ion lists were composed of the ions of the concerned compounds, and when performing targeted MS/MS analysis, no interference was encountered from the unconcerned compounds, and we could obtain the fragmentation ions of more concerned compounds. Therefore, more compounds were identified through our strategy, and even more potential new compounds, such as isomers, acetylations, and malonylations of known compounds were characterized.

Conflicts of interest

The authors have declared no conflict of interest.

Acknowledgements

References

1 L. L. Chen, J. Qi, Y. X. Chang, D. N. Zhu and B. Y. Yu, *J. Pharm. Biomed. Anal.*, 2009, **50**, 127.

2 Y. Y. Wang, S. He, X. C. Cheng, Y. X. Lu, Y. P. Zou and Q. L. Zhang, *J. Pharm. Biomed. Anal.*, 2013, **80**, 24.

3 Z. H. Su, G. A. Zou, A. Preiss, H. W. Zhang and Z. M. Zou, *J. Pharm. Biomed. Anal.*, 2010, **53**, 454.

4 Y. Qi, S. Z. Li, Z. F. Pi, F. R. Song, N. Lin, S. Liu and Z. Q. Liu, *Talanta*, 2014, **118**, 21.

5 M. H. Liu, X. Tong, J. X. Wang, W. Zou, H. Cao and W. W. Su, *J. Pharm. Biomed. Anal.*, 2013, **74**, 141.

6 Y. H. Wang, C. Qiu, D. W. Wang, Z. F. Hu, B. Y. Yu and D. N. Zhu, *J. Pharm. Biomed. Anal.*, 2011, **54**, 1110.

7 W. J. Miao, Q. Wang, T. Bo, M. Ye, X. Qiao, W. Z. Yang, C. Xiang, X. Y. Guan and D. A. Guo, *J. Pharm. Biomed. Anal.*, 2013, **72**, 99.

8 H. Hao, N. Cui, G. Wang, B. Xiang, Y. Liang, X. Xu, H. Zhang, J. Yang, C. Zheng, L. Wu, P. Gong and W. Wang, *Anal. Chem.*, 2008, **80**, 8187.

9 C. Zheng, H. Hao, X. Wang, X. Wu, G. Wang, G. Sang, Y. Liang, L. Xie, C. Xia and X. Yao, *J. Mass Spectrom.*, 2009, **44**, 230.

10 T. Xie, Y. Liang, H. Hao, A. Jiye, L. Xie, P. Gong, C. Dai, L. Liu, A. Kang, X. Zheng and G. Wang, *J. Chromatogr. A*, 2012, **1227**, 234.

11 D. D. Wang, J. Liang, W. Z. Yang, J. J. Hou, M. Yang, J. Da, Y. Wang, B. H. Jiang, X. Liu, W. Y. Wu and D. A. Guo, *J. Pharm. Biomed. Anal.*, 2014, **89**, 130.

12 F. X. Zhang, M. Li, L. R. Qiao, Z. H. Yao, C. Li, X. Y. Shen, Y. Wang, K. Yu, X. S. Yao and Y. Dai, *J. Pharm. Biomed. Anal.*, 2016, **122**, 59.

13 X. B. Yang, X. W. Yang and J. X. Liu, *Mod. Chin. Med.*, 2013, **15**, 349.

14 N. N. Hu and X. J. Zhang, *Inf. Tradit. Chin. Med.*, 2021, **38**, 76.

15 J. P. Chen, K. X. Zhang, Y. Liu, X. H. Gai, T. Ren, S. X. Liu and C. W. Tian, *Chin. Tradit. Herb. Drugs*, 2021, **52**, 1772.

16 Q. Fan, X. B. Chen, L. Rong and C. X. Zhang, *Nat. Pro. Res. Dev.*, 2020, **32**, 1244.

17 J. J. Bi, Q. Zhou and P. Zhang, *Jiangsu Agr. Sci.*, 2018, **46**, 1.

18 H. L. Xu, Y. C. Li, Y. Q. Zhang, H. Feng, X. F. Li and L. L. Zhang, *Asia-Pacific Tradi. Med.*, 2018, **14**, 120.

19 W. Li, Y. L. Feng, T. E. Li, H. Wu, Y. Li, G. Y. Zhong, B. Wu and M. Z. He, *J. Chin. Mass Spectr. Soc.*, 2016, **37**, 504.

20 C. M. Li, Y. W. Luo and B. Y. Tian, *J. Heibi Norm. Univ.*, 2015, **39**, 522.

21 L. Li, C. M. Liu, Z. Q. Liu, L. Lv and S. Y. Liu, *J. Chin. Mass Spectr. Soc.*, 2006, **27**, 504.

22 B. B. Xu, B. T. Huang, J. X. Zeng, M. Li, J. Wang, Y. Y. Zhu, G. Ren, J. X. Zhu, G. M. Luo and X. Y. Wang, *J. Exper. Tradit. Med. Formu.*, 2016, **22**, 64.

23 J. Huang, Q. Shao, Y. H. Xiang, Z. W. Ge and X. H. Fan, *Chin. J. of Chin. Mater. Med.*, 2014, **39**, 2513.