Phyton-
International Journal of
Experimental Botany

Tech Science Press

DOI: 10.32604/phyton.2022.018571

ARTICLE

Dimension-Enhanced Ultra-High Performance Liquid Chromatography/Ion Mobility-Quadrupole Time-of-Flight Mass Spectrometry Combined with Intelligent Peak Annotation for the Rapid Characterization of the Multiple Components from Seeds of *Descurainia sophia*

Simiao Wang[1,#], Xue Li[1,#], Boxue Chen[1], Shitong Li[1], Jiali Wang[1], Jing Wang[2], Mingshuo Yang[3], Xiaoyan Xu[1], Hongda Wang[1] and Wenzhi Yang[1,*]

[1]State Key Laboratory of Component-Based Chinese Medicine, Tianjin Key Laboratory of TCM Chemistry and Analysis, Tianjin University of Traditional Chinese Medicine, Tianjin, 301617, China

[2]Waters Technology Co., Ltd., Beijing, 101102, China

[3]School of Computing Science, University of Glasgow, Glasgow, G12 8RZ, UK

[*]Corresponding Author: Wenzhi Yang. Email: wzyang0504@tjutcm.edu.cn

[#]Simiao Wang & Xue Li contributed equally to this work

Received: 03 August 2021 Accepted: 08 September 2021

ABSTRACT

The complex composition of herbal metabolites necessitates the development of powerful analytical techniques aimed to identify the bioactive components. The seeds of *Descurainia sophia* (SDS) are utilized in China as a cough and asthma relieving agent. Herein, a dimension-enhanced integral approach, by combining ultra-high performance liquid chromatography/ion mobility-quadrupole time-of-flight mass spectrometry (UHPLC/IM-QTOF-MS) and intelligent peak annotation, was developed to rapidly characterize the multicomponents from SDS. Good chromatographic separation was achieved within 38 min on a UPLC CSH C18 (2.1 × 100 mm, 1.7 μm) column which was eluted by 0.1% formic acid in water (water phase) and acetonitrile (organic phase). Collision-induced dissociation-MS^2 data were acquired by the data-independent high-definition MS^E ($HDMS^E$) in both the negative and positive electrospray ionization modes. A major components knockout strategy was applied to improve the characterization of those minor ingredients by enhancing the injection volume. Moreover, a self-built chemistry library was established, which could be matched by the UNIFI software enabling automatic peak annotation of the obtained $HDMS^E$ data. As a result of applying the intelligent peak annotation workflows and further confirmation process, a total of 53 compounds were identified or tentatively characterized from the SDS, including 29 flavonoids, one uridine derivative, four glucosides, one lignin, one phenolic compound, and 17 others. Notably, four-dimensional information related to the structure (e.g., retention time, collision cross section, MS^1 and MS^2 data) was obtained for each component by the developed integral approach, and the results would greatly benefit the quality control of SDS.

KEYWORDS

Descurainia sophia; multicomponent characterization; ultra-high performance liquid chromatography; ion mobility/quadrupole time-of-flight mass spectrometry; high-definition MS^E; flavonoid

1 Introduction

Traditional Chinese Medicine (TCM), mostly derived from plants, is attracting more attention from a global scope [1]. It has been included in the system of alternative therapy, and is particularly effective in treating some chronic and prevalent diseases, such as COVID-19 [2]. Systematic elucidation of the chemical substances of TCM is helpful to explore the therapeutic basis and elaborate the quality standards to promote its modernization and globalization [3]. However, there are some obstacles hindering the internationalization of TCM. As a distinctive feature, the multiple parts of plants, such as roots [4], leaves [5], flowers [6], and seeds [7], etc., can be used as a source for TCM. Generally, herbal medicines contain various chemical components, which are featured by the co-existance of primary and secondary metabolites with wide spans of acidity-base properties and molecular masses, different polarities, and sharply different contents, etc. [8]. Therefore, it is difficult to perform a quality investigation of TCM. Increasing demands on the sensitive characterization of the known components (targeted), and meanwhile, having the ability to probe into those unknown (untargeted) necessitate the elaboration of powerful analytical techniques, such as liquid chromatography/mass spectrometry (LC–MS) [9]. MS, by combining the negative and positive ionization modes or using different ionization patterns, can give good response to most of the natural components with ultra-high sensitivity. Moreover, the ability of MS^n acquisition, by the alternate fragmentation mechanisms and flexible scan approaches, provides more options to facilitate the fit-for-purpose profiling and characterization of multiple components with significantly greater coverage on the components of interest [10]. Another milestone progress in herbal components analysis in recent years is the introduction of ion mobility mass spectrometry (IM-MS). It offers an additional dimension of separation based on the size, shape, and charge state of the gas-phase ions orthogonal of MS. Impressively, IM-MS coupled with LC, such as the ultra-high-performance liquid chromatography/ion mobility-quadrupole time-of-flight mass spectrometry (UHPLC/IM-QTOF-MS), is able to provide four-dimensional data (e.g., retention time, drift time, MS information and the response) of the components. IM-derived collision cross section (CCS) has the potential to discriminate among isomeric metabolites [11,12].

Descurainia sophia (L.) Webb ex Prantl., belonging to the Brassicaceae family, is a medicinal herb mainly distributed in Shandong, Hebei, Henan, Zhejiang, Inner Mongolia, and Gansu provinces of China [13,14]. The seeds of *D. sophia* (SDS) are used as the source for the TCM Descurainiae Semen (Ting-Li-Zi). The plants are harvested when the fruits are ripe in summer, dried, and the seeds are rubbed out with the impurities removed [15]. According to the basic theory of TCM, Descurainiae Semen has a pungent, bitter taste, and a severe cold property, which can relieve lung and asthma, and is also effective in relieving swelling [16–18]. To date, a number of studies have been reported on the phytochemistry of SDS, and a variety of compounds have been isolated [19–21]. These compounds include flavonoids, glucosinolates and isothiocyanates, phenylpropanoids, coumarins, cardiac glycosides, organic acids, and volatile oils, etc. Additionally, quality control on Descurainiae Semen has been conducted to identify and quantitatively assay the components contained. For instance, a UHPLC/Q-TOF-MS/MS method was established, by Meng et al. [14] which enabled the characterization of 14 constituents from SDS, including nine flavonoids, four fatty acids, and one cardiac glycoside. Li et al. [22] established an HPLC method to determine the content of quercetin-3-*O*-*β*-D-glucopyranosyl-7-*O*-*β*-D-gentiobioside in 28 batches of SDS samples, and suggested that its content is not less than 0.075%. However, according to the report of Wang et al. [23] this flavonoid compound in SDS was unstable in water at high temperature in a long-term extraction process.

In the current work, we reported a dimension-enhanced integral approach, by combining UHPLC/IM-QTOF-MS-based data-independent high-definition MS^E ($HDMS^E$) and UNIFI™-facilitated computational peak annotation to rapidly profile and characterize the multiple components from SDS, especially those minor ones. Fig. 1 illustrates the overall technical roadmap for this strategy. For this purpose, numerous

efforts were made to achieve better performance in both the resolution and sensitivity: 1) the chromatography (e.g., stationary phase, column temperature, and gradient elution program) and key parameters of the Vion^{TM} IM-QTOF mass spectrometer (capillary voltage, cone voltage, and collision energy) were both carefully optimized; 2) a major components knockout strategy was utilized to boost the profiling and characterization of minor components by enhancing the injection volume; 3) a chemical library recording 232 known compounds was in-house established to guide the automatic peak annotation and components identification; and 4) streamlined workflows were created by applying UNIFI^{TM} to process the obtained high-resolution CID (collision-induced dissociation)-MS^2 data. As a result of these efforts, we could identify or tentatively characterize 53 compounds from SDS, with four-dimensional information with respect to each component (e.g., t_R, drift time/CCS, MS^1 and MS^2 data) conveniently provided.

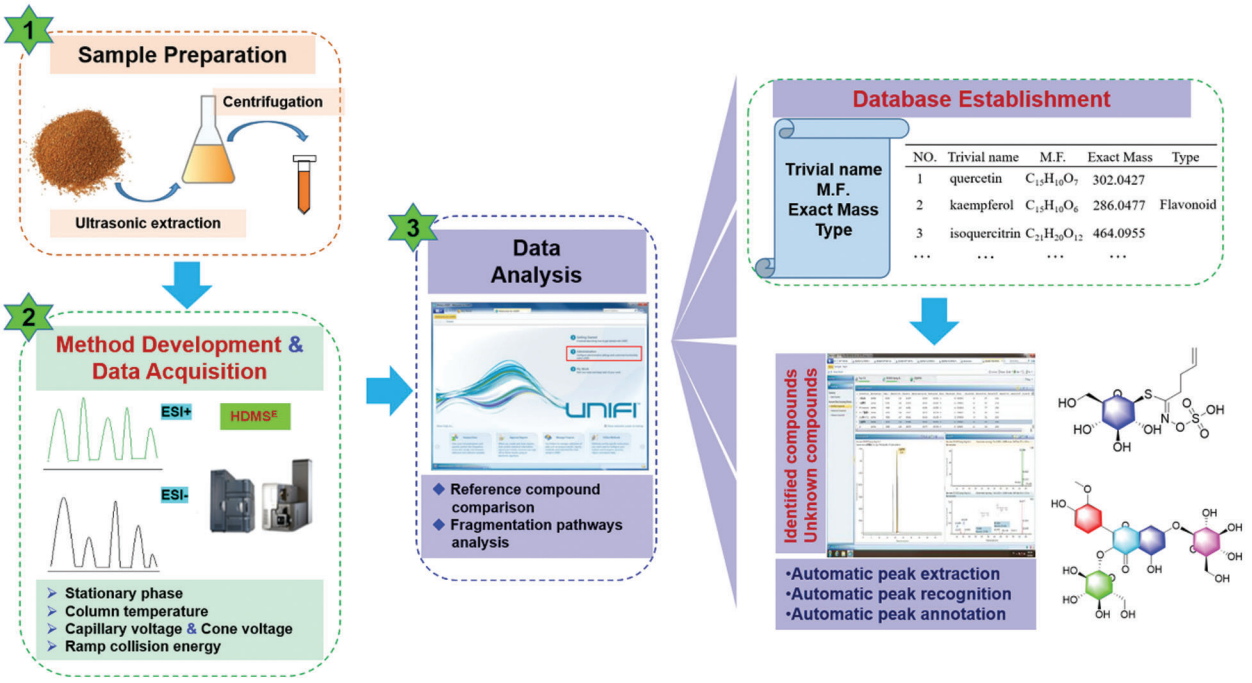

Figure 1: Schematic diagram for the rapid characterization of multiple components from the seeds of *Descurainia sophia* (SDS) by integrating dimension-enhanced UHPLC/IM-QTOF-MS and computational peak annotation because of the automatic MS information matching with the incorporated in-house library

2 Materials and Methods

2.1 Materials

Twenty compounds (Fig. 2) were used as reference compounds in this work. These compounds were isoquercitrin (**1**, $C_{21}H_{20}O_{12}$), quercetin-7-*O*-*β*-D-glucopyranoside (**2**, $C_{21}H_{20}O_{12}$), kaempferol-3-*O*-neohesperidoside (**3**, $C_{27}H_{30}O_{15}$), kaempferol-3-*O*-rutinoside (**4**, $C_{27}H_{30}O_{15}$), kaempferol (**5**, $C_{15}H_{10}O_6$), luteolin (**6**, $C_{15}H_{10}O_6$), quercetin (**7**, $C_{15}H_{10}O_7$), apigenin (**8**, $C_{15}H_{10}O_5$), isorhamnetin-3-*O*-glucoside (**9**, $C_{22}H_{22}O_{12}$), kaempferol-7-*O*-glucoside (**10**, $C_{21}H_{20}O_{11}$), kaempferol-3-*O*-*β*-D-glucuronide (**11**, $C_{21}H_{18}O_{12}$), apigenin-7-*O*-β-D-glucuronide (**12**, $C_{21}H_{20}O_{10}$), rutin (**13**, $C_{27}H_{30}O_{16}$), eleutherodide A (**14**, $C_{33}H_{40}O_{20}$), *β*-sitosterol (**15**, $C_{35}H_{60}O_6$), glucosinalbin (**16**, $C_{29}H_{50}O$), scopolin (**17**, $C_{14}H_{19}NO_{10}S$), psoralen (**18**, $C_{16}H_{18}O_9$), sinapic acid (**19**, $C_{11}H_6O_3$), and vanillic acid (**20**, $C_8H_8O_4$). They were purchased from Chengdu Desite Biotechnology Co., Ltd. (Chengdu, China) or Shanghai Standard Biotech. Co., Ltd. (Shanghai, China),. The drug material for the seeds of *D. sophia* was collected from

Hebei Province of China (Batch No. 190302). The specimens were deposited at the authors' lab in Tianjin University of Traditional Chinese Medicine (Tianjin, China). HPLC-grade acetonitrile, methanol (Fisher, Fair Lawn, NJ, USA), formic acid (Sigma-Aldrich, MO, Switzerland), and ultra-pure water [in-house prepared using a Milli-Q Integral 5 water purification system (Millipore, Bedford, MA, USA)) were used.

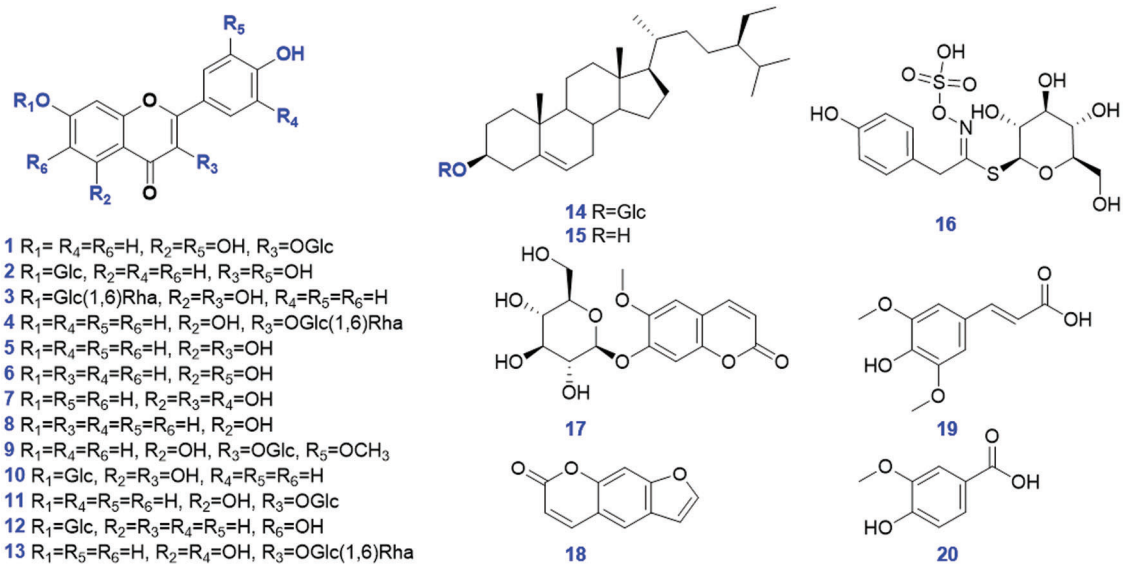

Figure 2: Chemical structures of 20 reference compounds used in the current work

2.2 Sample Preparation

An easy-to-implement ultrasonic extraction method was utilized for sample preparation from the SDS. In detail, 0.4 g of accurately weighed powder was soaked in a 15-mL centrifuge tube containing 5 mL of methanol. After being vortexed for 2 min, the sample was extracted on a water bath at 25°C with ultrasound assistance (power, 1130 W; frequency, 40 kHz) for 40 min. The extract was centrifuged for 10 min at 3,219 g (equal to 4,000 revolutions per min). The supernatant was further transferred into a 10-mL volumetric flask. The extraction process was repeated by using another 4 mL of methanol. The pooled supernatant was diluted to a constant volume (10 mL), and it was then well mixed. The solution was centrifuged at 11,481 g (14,000 revolutions per min) for 10 min, and diluted reaching a final concentration of drug material of 10 mg/mL.

2.3 UHPLC/IM-QTOF-MS

Metabolites profiling of SDS was performed on the ACQUITY UPLC I-Class/Vion^{TM} IM-QTOF system (Waters Corporation, Milford, MA, USA). A Waters Acquity UPLC CSH C18 (2.1 × 100 mm, 1.7 μm) column hyphenated with a VanGuard Pre-column (2.1 × 50 mm, 1.7 μm) maintained at 30°C was used for the UHPLC separation. A binary mobile phase, composed by 0.1% formic acid in H_2O (water phase: **A**) and CH_3CN (organic phase: **B**), ran at 0.3 mL/min in consistency with an optimal gradient program: 0−3 min, 2% (B); 3−5 min, 2%−10% (B); 5−7 min, 10% (B); 7−10 min, 10%−13% (B); 10−15 min, 13% 21% (B); 15−21 min, 21% −28% (B); 21−24 min, 28%−40% (B); 24−26 min, 40% −65% (B); 26−31 min, 65%−85% (B); 31−35 min, 85%−100% (B); and 35−38 min, 100% (B). An aliquot of 3 μL of the test solution was injected onto the column for analysis.

High-accuracy MS data for structural elucidation were acquired on a Vion^{TM} IM-QTOF mass spectrometer in both the positive and negative electrospray ionization (ESI) modes by data-independent

HDMS^E (Waters). The LockSpray ion source was equipped adopting the following parameters: capillary voltage, 1.5 kV; cone voltage, 60 V; source offset, 80 V; source temperature, 120°C; desolvation gas temperature, 500°C; desolvation gas flow (N_2), 800 L/h; and cone gas flow (N_2), 50 L/h. Default parameters were defined for the travelling wave IM separation [24,25]. The QTOF analyzer scanned over a mass range of *m/z* 100−1500 at a low collision energy of 6 eV at 0.3 s per scan (MS^1). Ramp collision energy (RCE) of 20−40 eV was set in high-energy channel for HDMS^E. MS data calibration was conducted by constantly infusing the leucine enkephalin solution (Sigma-Aldrich; 200 ng/mL) at a flow rate of 10 μL/min. CCS calibration was conducted according to the Manufacture's guidelines by using a mixture of calibrants [26].

2.4 Data Analysis

Uncorrected HDMS^E data in Continuum format were corrected and processed by using the UNIFI^{TM} 1.9.3.0 software (Waters). UNIFI could efficiently perform data correction, peak picking, and peak annotation [24,25]. Key parameters of UNIFI were set as follows. Find 4D Peaks (only set in HDMS^E): High-energy intensity threshold, 1000.0 counts; low-energy intensity threshold, 500.0 counts. Target by mass: Target match tolerance, 10.0 ppm; screen on all isotopes in a candidate, generate predicted fragments from structure, and look for in-source fragments; fragment match tolerance, 10.0 ppm. Adducts: Positive adduct was +H. Lock Mass: Combine width, 3 scans; mass window, 0.5 *m/z*; reference mass, 556.2766; reference charge, +1. Adducts: Negative adduct was −H. Lock Mass: Combine width, 3 scans; mass window, 0.5 *m/z*; reference mass, 554.2620; reference charge, −1.

2.5 Creation of an in-House Library Incorporated into UNIFI^{TM}

Considering the restricted coverage of the commercial library on TCM components, we had the in-house library as a resource, which could guide the automatic annotation by UNIFI [24,25,27]. A systematic summary on the phytochemical reports of the seeds of *D. sophia* and *Lepidium apetalum* (two official plant sources for the TCM Ting-Li-Zi) was retrieved from multiple available databases (e.g., Web of Science, SciFinder, and CNKI, etc.). A chemical library was thus established recording the trivial name, molecular formula, and chemical structure for each compound. The collected structure information was input into an EXCEL file (*.xls*) in a format compatible with UNIFI, and each structure was drawn using ChemDraw Professional (Cambridge, USA), which was subsequently saved as a *.mol* file with the file name in consistency with the trivial name. The resulting EXCEL file and all structure files were incorporated into the UNIFI software, which can be utilized to efficiently annotate the obtained HDMS^E data achieving structural elucidation for SDS.

3 Results

We integrated the dimension-enhanced HDMS^E and UNIFI-facilitated intelligent peak annotation workflows, based on the UHPLC/IM-QTOF-MS, to enable the rapid profiling and characterization of the multi-components from SDS. Meanwhile, considering the differentiated ionization rates for the different classes of SDS components, both the negative and positive ESI modes were utilized to acquire the CID-MS^2 data.

As a result, good separation of the SDS components was accomplished by a reversed-phase UPLC CSH C18 column (column temperature: 30°C) within 38 min using CH_3CN/0.1% FA-H_2O as the mobile phase. Capillary voltage at 1.5 kV and cone voltage at 60 V in ESI− could give satisfactory ion response, but seldom induce the in-source fragmentation. RCE of 20−40 eV was able to acquire more balanced MS^2 spectra than the setting of the fixed collision energy. The negative-mode CID-MS^2 data were mainly used for structural elucidation, while the data obtained by ESI+ exerted complementarity. In the ESI− mode, one major peak was predominant, which thus hindered the characterization of the other relatively minor ones (Fig. 3).

Thereby we applied the major components knockout strategy (the elutes at 7−8.8 min were automatically switched to the waste), by increasing the injection volume to 10 μL, which enabled the enrichment of a number of minor components. What's more, an in-house library of Ting-Li-Zi was established to aid the intelligent peak annotation by UNIFI. A total of 47 related documents were found, which involved 232 known compounds (including two alkaloids, two benzoic acids, three cardiac glycosides, 66 flavonoids, 35 glucosides, 9 lignin, 10 organic acids, five phenolic compounds, three phenolic acids, two phenylacetamides, one phenylacetonitrile, 3 phenylethylamines, 12 phenylpropanoids, one steroid, eight tobacco flavor anhydrides, one triterpenoid, 10 uridine derivatives, and 59 others (Table S1).

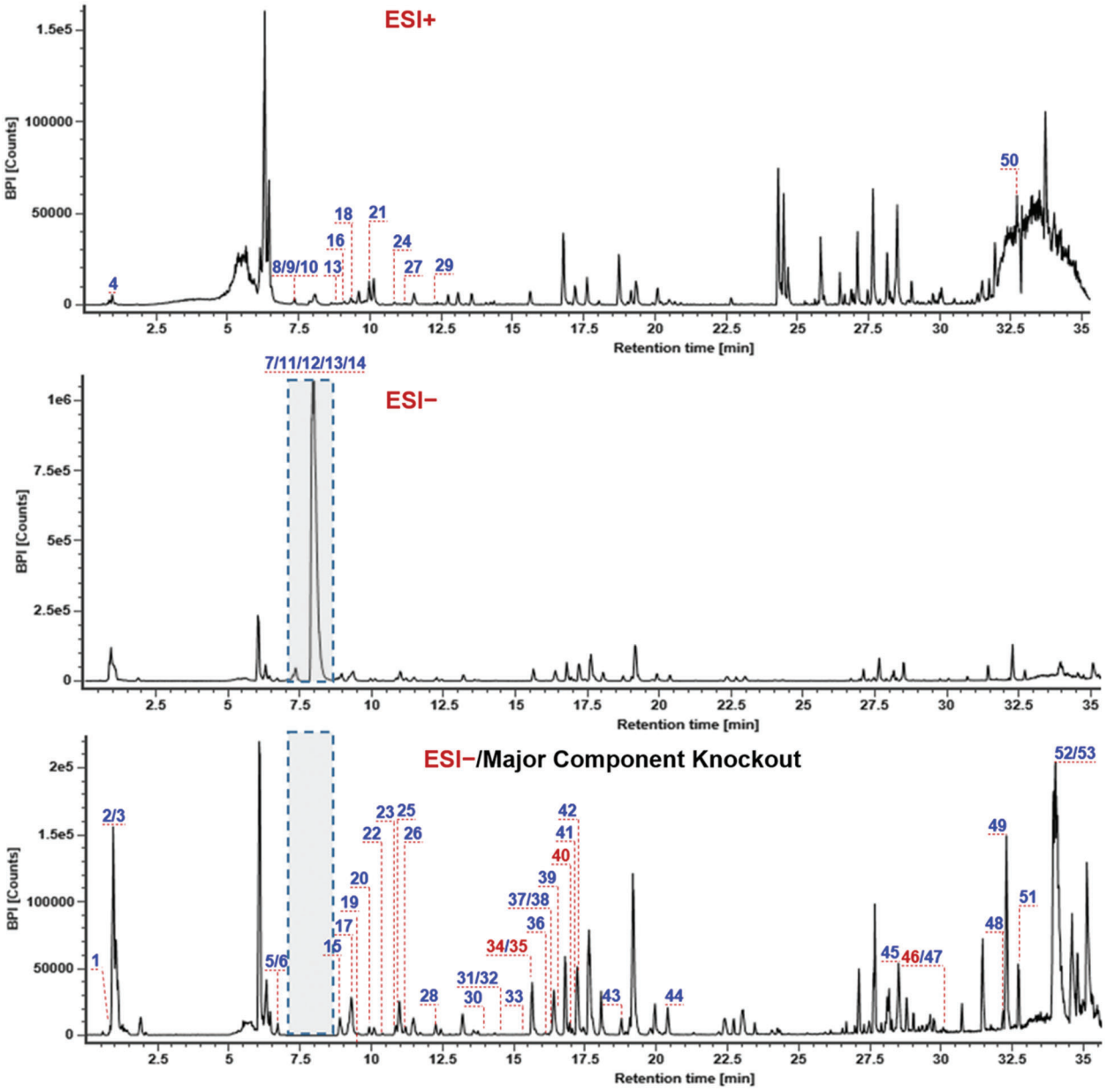

Figure 3: Base peak intensity chromatograms for the seeds of *D. sophia* obtained by HDMS^E in the positive (ESI+) and negative (ESI−) modes with the peaks annotated with the numbering consistent with Table S2. A major components knockout strategy (eluates at 7−8.8 min switched to the waste) was utilized to enhance the characterization of minor components

After the automatic matching between the determined CID-MS^2 data with the predicted MS data (based on the in-house library) using UNIFI, combined with a subsequent confirmation step, a total of 53 compounds (Table S2 and Fig. 3) were identified or tentatively characterized from SDS by analyzing both the negative and positive CID-MS^2 data. They included 29 flavonoids, one uridine derivative, four glucosides, one lignin, one phenolic compound, and 17 others. Notably, the CCS information for each characterized component is offered due to the enabling of the IM separation. The characterization of flavonoids and glucosinolates was illustrated in this work with representative compounds, and the others were outlined at the end of this section.

3.1 Characterization of Flavonoids

Flavonoids are an important class of bioactive components in SDS, and we were able to characterize 29 flavonoids based on the CID-MS^2 data in the current work. Interestingly, the sugars in these flavonoid *O*-glycosides are limited to glucose (Glc) and rhamnose (Rha), which could be characterized by the typical neutral losses of 162.05 Da and 146.06 Da, respectively. The structural differences are mainly embodied in the flavonoid aglycones, possibly as well as the glycosylation sites and the linkage patterns of multiple sugars which could not be exactly identified only by the MS information. The CID-MS/MS features of flavonoid *O*-glycosides were featured by the neutral loss of sugars and the production of deprotonated aglycone ions and their secondary fragments [28]. Here, the fragmentation pathways and characterization of compounds **18**# (t_R 9.34 min), **12**# (t_R 8.75 min) and **15**# (t_R 8.95 min) were illustrated (Fig. 4). The MS^1 spectrum of compound **18**# gave a rich protonated precursor ion ($[M+H]^+$) at *m/z* 641.1716 in the positive ESI mode, based on which its molecular formula was characterized as $C_{28}H_{32}O_{17}$. In the CID-MS^2 spectrum, a fragment, due to the neutral loss of the Glc moiety, was observed at *m/z* 479.1181 ($[M+H-Glc]^+$). In addition, the base peak ion at *m/z* 317.0645 should be the protonated aglycone after eliminating 2 × Glc ($[M+H-2Glc]^+$), which could further lose the free radical $-CH_3$ yielding a weak fragment at *m/z* 302.0414 (a homolytic fragmentation ion for the aglycone). The known flavonoid *O*-glycosides isolated from SDS only contain three flavonoid aglycones: isorhamnetin ($C_{16}H_{12}O_7$), quercetin ($C_{15}H_{10}O_7$), and kaempferol ($C_{15}H_{10}O_6$), and accordingly the aglycone moiety in compound **18**# should be isorhamnetin ($C_{16}H_{12}O_7$). The observation of free radical fragmentation of $-CH_3$ could also testify this aglycone structure. Ultimately, this compound was tentatively characterized as isorhamnetin-3,7-di-*O*-*β*-D-glucopyranoside or isomer [14]. The fragmentation pathways of compound **12**# (t_R 8.75 min) were analyzed based on the negative ESI-CID-MS^2 data, which gave a rich deprotonated precursor ion ($[M-H]^-$) at *m/z* 625.1408, based on which the molecular formula of $C_{27}H_{30}O_{17}$ could be characterized. Two major product ions, at *m/z* 463.0871 ($[M-H-Glc]^-$) and 301.0344 ($[M-H-2Glc]^-$), were generated by successively losing two Glc residues. The aglycone could match the known quercetin ($C_{15}H_{10}O_7$), and thus compound **12**# was tentatively characterized as quercetin-7-*O*-*β*-D-gentiobioside or isomer by comparison with the literature [17]. Compound **15**# (t_R 8.95 min) was characterized as a tri-glycosidic flavonoid with a rich $[M-H]^-$ precursor ion observed at *m/z* 771.1990. Three major product ions were detected in the MS/MS spectrum at *m/z* 609.1458, 447.0931, and 285.0398, which could be assigned as $[M-H-Glc]^-$, $[M-H-2Glc]^-$, and $[M-H-3Glc]^-$, respectively. The aglycone could match the known kaempferol ($C_{15}H_{10}O_6$), and thus Compound **15**# was identified as kaempferol-3-*O*-*β*-D-glucopyranosyl-7-*O*-*β*-gentiobioside or isomer [20]. To sum up, among these 29 flavonoids characterized from SDS, 12 compounds thereof (**6**#/**7**#/**8**#/**9**#/**12**#/**13**#/**20**#/**26**#/**27**#/**34**#/**37**#/**50**#) have the aglycone of quercetin, 11 (**10**#/**15**#/**16**#/**22**#/**28**#/**29**#/**35**#/**36**#/**40**#/**42**#/**43**#) with kaempferol, and 6 with isorhamnetin (/**17**#/**18**#/**23**#/**24**#/**41**#/**46**#).

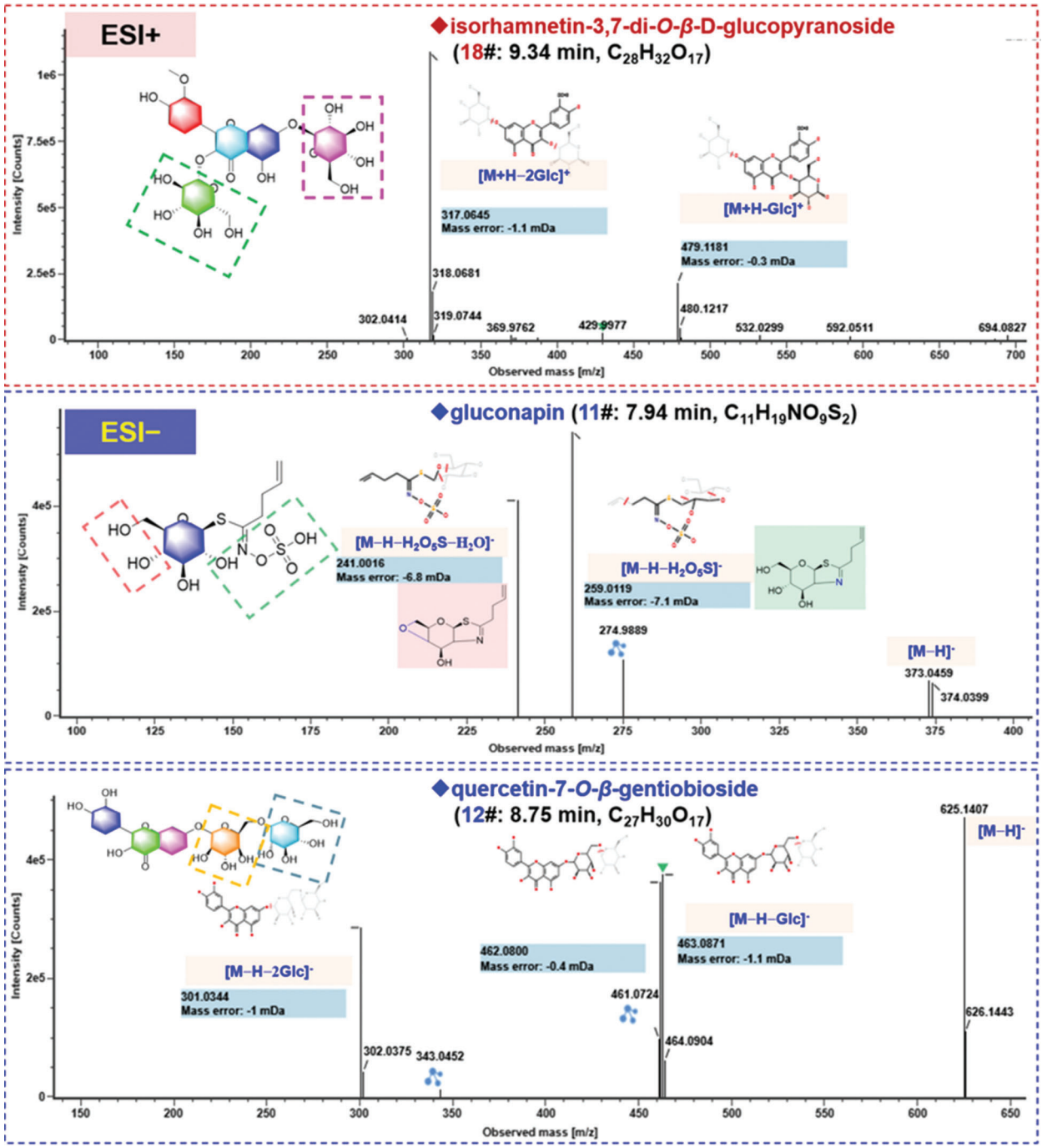

Figure 4: Continued

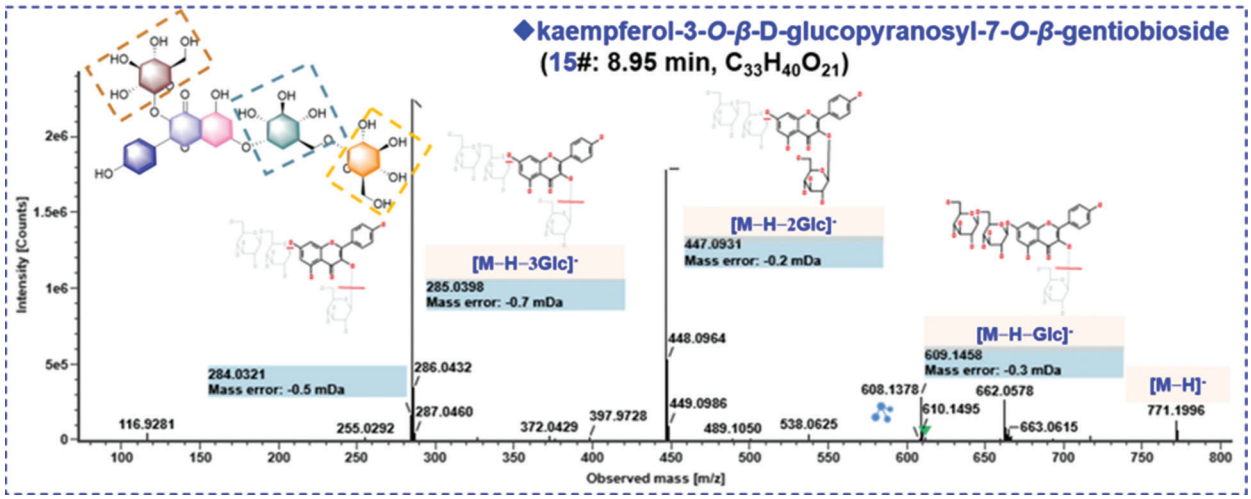

Figure 4: Annotation of the CID-MS^2 spectra of the representative compounds for the flavonoids (compounds 18#, 12#, and 15#) and the glucosinolate (compound 11#), identified from the seeds of *D. sophia* based on both the positive and negative CID-MS^2 data

3.2 Characterization of Glucosinolates

Isothiocyanates are the degradation products of glucosinolates ever isolated from SDS, including benzyl isothiocyanate, allyl isothiocyanate, 3-butenyl isothiocyanate, and 2-phenylethyl isothiocyanate, etc. [8]. Compound **11**# (t_R 7.94 min) was observed in the negative ESI mode, with a rich deprotonated precursor ion ($[M-H]^-$) at *m/z* 372.0425. In its CID-MS^2 spectrum, a fragment was observed at *m/z* 259.0119 ($[M-H-H_2O_5S]^-$), together with its secondary fragment of *m/z* 241.0016 by losing a molecule of H_2O ($[M-H-H_2O_5S-H_2O]^-$) (Fig. 4). By searching the literature, this compound was tentatively identified as gluconapin [29].

In addition to flavonoids and glucosinolates, one uridine derivative (**2**#), one lignin (**44**#), one phenolic compound (**38**#), and 17 others (**1**#/**3**#/**4**#/**5**#/**14**#/**19**#/**21**#/**25**#/**30**#/**33**#/**39**#/**47**#/**48**#/**49**#/**51**#/**52**#/**53**#), were also tentatively characterized from SDS by analyzing their high-resolution MS^1 and MS^2 data, and comparison with the literature as well. For example, the compound **3**#, was observed in the negative ESI mode, with a rich deprotonated precursor ion ($[M-H]^-$) at *m/z* 341.1083. In its CID-MS^2 spectrum, a fragment, due to the neutral loss of Glc, was observed at *m/z* 179.0555 ($[M-H-Glc]^-$). Accordingly, compound **3**# was tentatively characterized as sucrose. Information for the other compounds characterized from SDS is detailed in Table S2.

4 Discussion

4.1 Optimization of the Chromatography Condition

In order to separate and identify as many components as possible from SDS, the reversed-phase UHPLC conditions were optimized in terms of the stationary phase, column temperature, and elution gradient program. Effects of different stationary phases were evaluated, including four Waters sub-2 μm particles packed UHPLC columns, CSH C18 (2.1 × 100 mm, 1.7 μm), BEH Shield RP18 (2.1 × 100 mm, 1.7 μm), BEH C18 (2.1 × 100 mm, 1.7 μm), and HSS T3 (2.1 × 100 mm, 1.8 μm) (Fig. 5). Comparatively, most peaks could be separated on the CSH C18 column, and the peaks through the entire spectrum were more balanced than the other two columns. Therefore, we selected the CSH C18 column in this experiment. Moreover, the variations in column temperature (25−40°C) could influence the retention of SDS constituents, and most peaks were resolved under 30°C (Fig. S1). After further adjustment of the elution gradient, satisfactory separation of the SDS components was achieved.

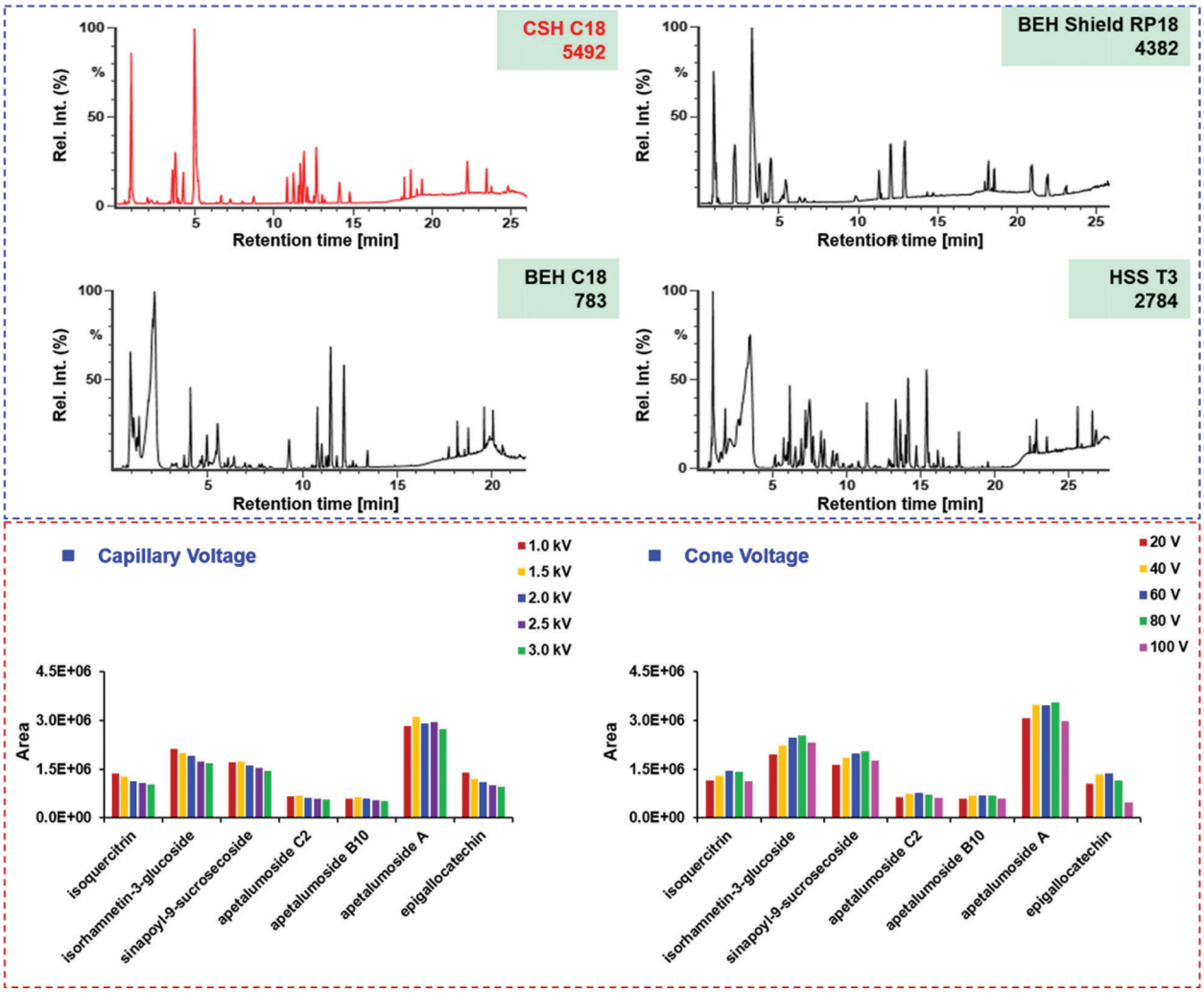

Figure 5: Screening of the candidate chromatographic columns and optimization of two key ion source parameters (capillary voltage and cone voltage) of the Vion^{TM} IM-QTOF mass spectrometer for separating and characterizing the multicomponents from the seeds of *D. sophia*

4.2 Optimization of Key Parameters on the Vion TM IM-QTOF Mass Spectrometer

Key parameters that may affect the ion response (capillary voltage and cone voltage) and the fragmentation degree (RCE) of SDS components were optimized in the negative ESI mode. Capillary voltage (1.0/1.5/2.0/2.5/3.0 kV) and cone voltage (20/40/60/80/100 V) were measured to compare the difference in ion response, using the peak areas of seven components (isoquercitrin, isorhamnetin-3-*O*-glucoside, sinapoyl-9-sucrosecoside, apetalumoside C2, apetalumoside B10, apetalumoside A, and epigallocatechin) (Fig. 5). In general, the response of seven index compounds decreased with the increase of capillary voltage. To be specific, four (sinapoyl-9-sucrosecoside, apetalumoside C2, apetalumoside B10 and apetalumoside A) of the seven index compounds were best ionized at the capillary voltage of 1.5 kV, while the other three gave the highest ion response at 1.0 kV. Hence, we selected the capillary voltage of 1.5 kV as the most appropriate. The changes of cone voltage ranging from 20 to 100 V led to inconsistent variation trends for the index compounds. Except for apetalumoside A, the response of the other six compounds increased when cone voltage ascended from 20 to 60 V. When the cone voltage

increased from 80 to100 V, the response tended to decline. The cone voltage at 60 V was finally selected in this experiment (Fig. 5). Advanced mass spectrometers can enable slope collision energy or mixed collision energy, which greatly enhanche the quality of the MS^2 spectra producing more rich product ions [1,9]. RCE was optimized at four levels (e.g., 10−30 eV, 20−40 eV, 30−50 eV, and 40−60 eV) by observing the transitions of five compounds (quercetin 3-*O*-*β*-D-glucose-7-*O*-*β*-D-gentiobioside, isoquercitrin, isorhamnetin-3-*O*-D-glucoside, gluconapin, and sinapoyl-9-sucrosecoside) from the precursor to the product ions to sufficiently fragment different classes of chemical compounds in SDS (Fig. S2). Accordingly, RCE at 20−40 eV was selected, under which rich fragments were obtained for the SDS components.

5 Conclusions

In conclusion, the developed UHPLC/IM-QTOF-$HDMS^E$ approach combined with automatic peak annotation workflows of UNIFI enable the characterization of 53 compounds from the seeds of *D. sophia*, demonstrating a great improvement compared with the literature [14]. Computational data processing-based structural elucidation was very efficient giving reproducible characterization results independent of the professional knowledge in this field. Four dimensions of structure information could be generated by this integral approach, which could improve the reliability in identifying isomers if a CCS comparison is available. The results obtained would greatly benefit the quality control of this herbal medicine.

Acknowledgement: We thank the anonymous reviewers for their helpful comments.

Author Contribution: Simiao Wang, Xue Li, Boxue Chen, Shitong Li, and Jiali Wang performed the experiment and analyzed the experimental data. Jing Wang and Wenzhi Yang designed the research. Simiao Wang, Mingshuo Yang, and Wenzhi Yang drafted the manuscript. Xiaoyan Xu and Hongda Wang polished the manuscript.

Funding Statement: This work was financially supported by the National Key Research and Development Program of China (Grant No. 2018YFC1704500), Tianjin Committee of Science and Technology of China (Grant No. 21ZYJDJC00080), and National Natural Science Foundation of China (Grant No. 81872996).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding this present study.

References

1. Qian, Y. X., Li, W. W., Wang, H. M., Hu, W. D., Wang, H. D. et al. (2021). A four-dimensional separation approach by offline 2D-LC/IM-TOF-MS in combination with database-driven computational peak annotation facilitating the in-depth characterization of the multicomponents from Atractylodis Macrocephalae Rhizoma (*Atractylodes macrocephala*). *Arabian Journal of Chemistry, 14(2),* 102957. DOI 10.1016/j.arabjc.2020.102957.
2. Luo, L., Jiang, J. W., Wang, C., Fitzgerald, M., Hu, W. F. et al. (2020). Analysis on herbal medicines utilized for treatment of COVID-19. *Acta Pharmaceutica Sinica B, 10(7),* 1192–1204. DOI 10.1016/j.apsb.2020.05.007.
3. Yang, W. Z., Zhang, Y. B., Wu, W. Y., Huang, L. Q., Guo, D. A. et al. (2017). Approaches to establish Q-markers for the quality standards of traditional Chinese medicines. *Acta Pharmaceutica Sinica B, 7(4),* 439–446. DOI 10.1016/j.apsb.2017.04.012.
4. Wang, S. M., Qian, Y. X., Sun, M. X., Jia, L., Hu, Y. et al. (2020). Holistic quality evaluation of Saposhnikoviae Radix (*Saposhnikovia divaricata*) by reversed-phase ultra-high performance liquid chromatography and hydrophilic interaction chromatography coupled with ion mobility quadrupole time-of-flight mass spectrometry-based untargeted metabolomics. *Arabian Journal of Chemistry, 13(12),* 8835–8847. DOI 10.1016/j.arabjc.2020.10.013.

5. Qiu, S., Yang, W. Z., Shi, X. J., Yao, C. L., Yang, M. et al. (2015). A green protocol for efficient discovery of novel natural compounds: Characterization of new ginsenosides from the stems and leaves of *Panax ginseng* as a case study. *Analytica Chimica Acta, 893,* 65–76. DOI 10.1016/j.aca.2015.08.048.
6. Turgumbayeva, A., Ustenova, G., Datkhayev, U., Rahimov, K., Stankevicius, E. (2019). Safflower (*Carthamus tinctorius* L.) a potential source of drugs against cryptococcal infections, malaria and leishmaniasis. *Phyton-International Journal of Experimental Botany, 88(3),* 137–146. DOI 10.32604/phyton.2020.07665.
7. Yang, W. Z., Ye, G., Meng, A. H., Sabir, G., Qiao, X. et al. (2013). Rapid characterisation of flavonoids from *Sophora alopecuroides* L. by HPLC/DAD/ESI-MS^n. *Natural Product Research, 27(4–5),* 323–330. DOI 10.1080/14786419.2012.688052.
8. Yang, W. Z., Shi, X. J., Yao, C. L., Huang, Y., Hou, J. J. et al. (2020). A novel neutral loss/product ion scan-incorporated integral approach for the untargeted characterization and comparison of the carboxyl-free ginsenosides from *Panax ginseng*, *Panax quinquefolius*, and *Panax notoginseng*. *Journal of Pharmaceutical and Biomedical Analysis, 177,* 112813. DOI 10.1016/j.jpba.2019.112813.
9. Fu, L. L., Ding, H., Han, L. F., Jia, L., Yang, W. Z. et al. (2019). Simultaneously targeted and untargeted multicomponent characterization of Erzhi Pill by offline two-dimensional liquid chromatography/quadrupole-Orbitrap mass spectrometry. *Journal of Chromatography A, 1584,* 87–96. DOI 10.1016/j.chroma.2018.11.024.
10. Feng, K. Y., Wang, S. M., Han, L. F., Qian, Y. X., Li, H. F. et al. (2021). 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. *Journal of Chromatography A, 1649,* 462237. DOI 10.1016/j.chroma.2021.462237.
11. Tu, J., Zhou, Z. W., Li, T. Z., Zhu, Z. J. (2019). The emerging role of ion mobility-mass spectrometry in lipidomics to facilitate lipid separation and identification. *TrAC Trends in Analytical Chemistry, 116(Suppl. 50),* 332–339. DOI 10.1016/j.trac.2019.03.017.
12. Zhou, Z. W., Tu, J., Xiong, X., Shen, X. T., Zhu, Z. J. (2017). LipidCCS: Prediction of collision cross-section values for lipids with high precision to support ion mobility-mass spectrometry-based lipidomics. *Analytical Chemistry, 89(17),* 9559–9566. DOI 10.1021/acs.analchem.7b02625.
13. Shi, P. P., Chao, L. P., Wang, T. X., Liu, E. W., Han, L. F. et al. (2015). New bioactive flavonoid glycosides isolated from the seeds of *Lepidium apetalum* Willd. *Fitoterapia, 103,* 197–205. DOI 10.1016/j.fitote.2015.04.007.
14. Meng, Z. H., Qiao, L., Wang, Z. Q., Ma, Z. S., Jia, J. M. et al. (2015). Analysis of the constituents in *Semen descurainiae* by UPLC/Q-TOF-MS/MS. *Journal of Chinese Pharmaceutical Sciences, 24(5),* 303–309. DOI 10.5246/jcps.2015.05.039.
15. Committee of National Pharmacopoeia (2020). *China pharmacopoeia (Part 1)*. Beijing: Chemical Industry Press.
16. Li, M., Zeng, M. N., Zhang, Z. G., Zhang, J. K., Zhang, X. K. et al. (2019). Chemical constituents from the seeds of Lepidium apetalum. *Chinese Traditional Patent Medicine, 41(1),* 115–110. DOI 10.3969/j.issn.1001-1528.2019.01.022.
17. Sun, K., Li, X. (2002). Progress in studies on chemical constituents and pharmacological effect of *Semen lepidii* and *Semen descurainiae*. *Chinese Traditional and Herbal Drugs, 33(7),* 100–102. DOI 10.7501/j.issn.0253-2670.2002.7.341.
18. Guo, W. T., Xing, S. H., Chen, Y., Yi, J. H., Zhang, Y. N. et al. (2018). Research progress of *Semen lepidii* and its herbal compound on ascites. *Chinese Medicine Modern Distance Education of China, 16(6),* 157–159. DOI 10.3969/j.issn.1672-2779.2018.06.064.
19. Sun, K., Li, X., Liu, J. M., Wang, J. H., Li, W. et al. (2005). A novel sulphur glycoside from the seeds of *Descurainia sophia* (L.). *Journal of Asian Natural Products Research, 7(6),* 853–856. DOI 10.1080/1028602042000204072.
20. Wang, A. Q., Wang, X. K., Li, J. L., Cui, X. Y. (2004). Isolation and structure identification of chemical constituents from the seeds of Descurainia sophia L. Webb ex Prantl. *Acta Pharmaceutica Sinica, 39(1),* 46–51. DOI 10.16438/j.0513-4870.2004.01.011.

21. Chen, D., Liu, X., Yang, X. H., Sun, K. (2010). Study on chemical constituents of *Descurainia sophia* (L.) Webb ex Prantl. *Special Wild Economic Animal and Plant Research, 32(2),* 62–63. DOI 10.16720/j.cnki.tcyj.2010.02.022.

22. Li, J. R., Huang, W. H., Guo, B. L., Chen, A. J. (2010). Research on chemical content standard of *Semen descurainiae*. *Chinese Journal of Pharmaceutical Analysis, 30(5),* 950–953. DOI 10.16155/j.0254-1793.2010.05.009.

23. Wang, W. P., Wu, H. M., Liu, M. T., Xu, Y. C., Yin, X. B. et al. (2020). Stability study of quercetin-3-*O*-*β*-D-glucose-7-*O*-*β*-D-diglycoside in *Descurainia sophia* seeds. *China Journal of Traditional Chinese Medicine and Pharmacy, 35(11),* 5739–5742.

24. Zhang, C. X., Zuo, T. T., Wang, X. Y., Wang, H. D., Hu, Y. et al. (2019). Integration of data-dependent acquisition (DDA) and data-independent high-definition MS^E ($HDMS^E$) for the comprehensive profiling and characterization of multicomponents from *Panax japonicus* by UHPLC/IM-QTOF-MS. *Molecules, 24(15),* 2708. DOI 10.3390/molecules24152708.

25. Zuo, T. T., Qian, Y. X., Zhang, C. X., Wei, Y. X., Wang, X. Y. et al. (2019). Data-dependent acquisition and database-driven efficient peak annotation for the comprehensive profiling and characterization of the multicomponents from compound Xueshuantong capsule by UHPLC/IM-QTOF-MS. *Molecules, 24(19),* 3431. DOI 10.3390/molecules24193431.

26. Paglia, G., Angel, P., Williams, J. P., Richardson, K., Olivos, H. J. et al. (2015). Ion mobility-derived collision cross section as an additional measure for lipid fingerprinting and identification. *Analytical Chemistry, 87(2),* 1137–1144. DOI 10.1021/ac503715v.

27. Jia, L., Zuo, T. T., Zhang, C. X., Li, W. W., Wang, H. D. et al. (2019). Simultaneous profiling and holistic comparison of the metabolomes among the flower buds of *Panax ginseng*, *Panax quinquefolius*, and *Panax notoginseng* by UHPLC/IM-QTOF-$HDMS^E$-based metabolomics analysis. *Molecules, 24(11),* 2188. DOI 10.3390/molecules24112188.

28. Yang, W. Z., Qiao, X., Bo, T., Wang, Q., Guo, D. A. et al. (2014). Low energy induced homolytic fragmentation of flavonol 3-O-glycoside by negative electrospray ionization tandem mass spectrometry. *Rapid Communications in Mass Spectrometry, 28(4),* 385–395. DOI 10.1002/rcm.6794.

29. Chen, J. M., Guan, R. Z. (2006). Extraction and identification of several glucosinolates in *Descurainia sophia* seeds. *Acta Botanica Boreali-Occidentalia Sinica, 4(6),* 1231–1235. DOI 10.3321/j.issn:1000-4025.2006.06.023.

Supporting Information

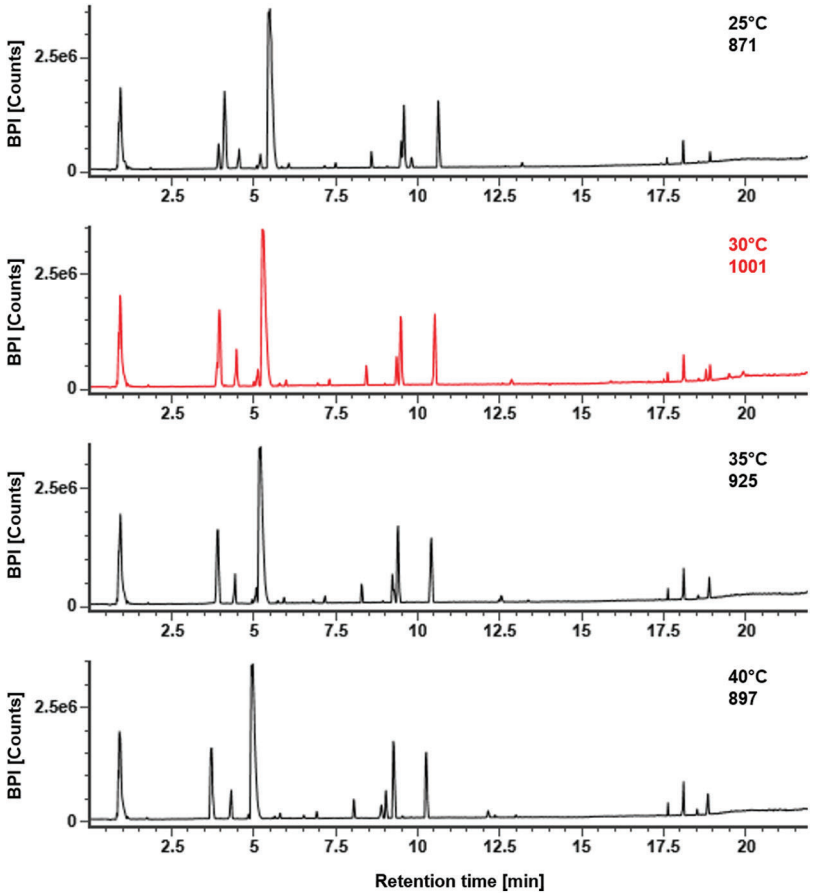

Figure S1: Comparison of column temperature (CSH C18 column) in ESI–mode for separating the multicomponents from the seeds of *Descurainia sophia* (SDS)

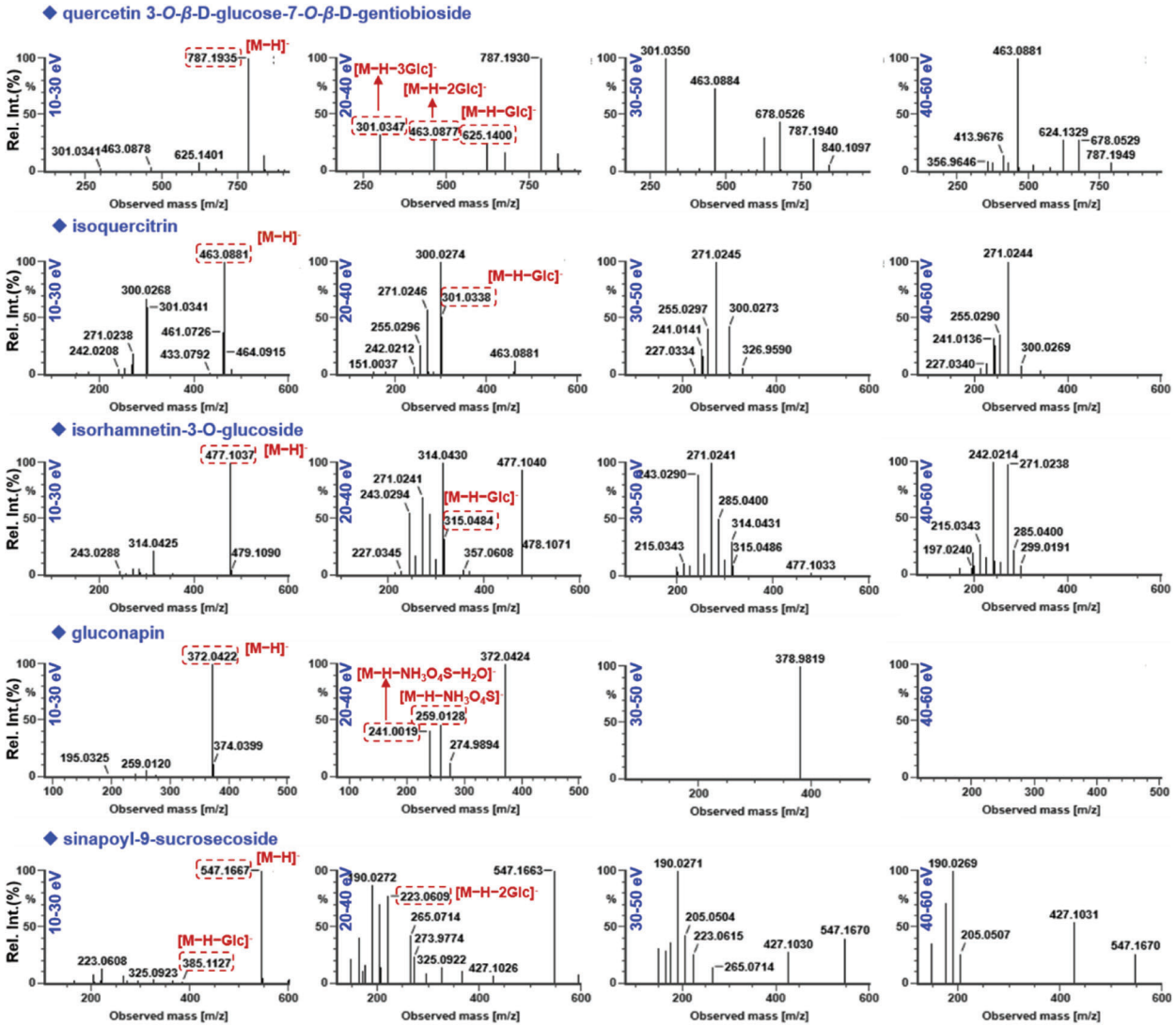

Figure S2: Optimization of the ramp collision energy (RCE) for HDMS^E using five representative compounds

Table S1: The in-house library of Semon lepidii and Semon descurainiae

No.	Name	Formula	Subclass
1	2-hydroxy-3-(1H-indol-3-yl) propanoic acid	$C_{11}H_{11}NO_3$	Alkaloid
2	2-hydroxyl-3-(1H-indol-3-yl) propanoic acid methyl ester	$C_{12}H_{13}NO_3$	
3	3,5-dihydroxy-4-methoxy-benzoic acid	$C_8H_8O_5$	Benzoic acid
4	1,2-benzenedicarboxylic acid-1,2-bis[2-(2-hydroxyethoxy) ethyl]ester	$C_7H_6O_3$	
5	strophanthidine	$C_{23}H_{32}O_6$	Cardiac glycoside
6	evomonoside	$C_{29}H_{44}O_8$	
7	3*β*,14*β*-hydroxy-5*β*-19-oxo-card-20(22)-enolide-3-*O*-*β*-D-glucopyranosyl-(1→4)-*β*-D-digitoxose	$C_{35}H_{52}O_{13}$	

(Continued)

Table S1 (continued).

No.	Name	Formula	Subclass
8	apetalumoside A1	$C_{56}H_{62}O_{30}$	Flavonoid
9	apetalumoside B8	$C_{44}H_{50}O_{26}$	
10	apetalumoside B9	$C_{55}H_{60}O_{30}$	
11	apetalumoside B10	$C_{44}H_{50}O_{26}$	
12	apetalumoside B11	$C_{50}H_{60}O_{31}$	
13	apetalumoside B12	$C_{61}H_{70}O_{35}$	
14	apetalumoside C1	$C_{55}H_{60}O_{29}$	
15	6-*O*-[E]-Sinapoyl-(α and β)-D-glucopyranoside	$C_{17}H_{22}O_{10}$	
16	(E)-2-*O*-sinapoly-D-glucopyranoside isomer	$C_{17}H_{22}O_{10}$	
17	kaempferol	$C_{15}H_{10}O_6$	
18	isorhamnetin	$C_{16}H_{12}O_7$	
19	quercetin	$C_{15}H_{10}O_7$	
20	isorhamnetin-3-*O*-β-D-glucopyranoside	$C_{22}H_{22}O_{12}$	
21	quercetin-3-*O*-β-D-glucopyranoside	$C_{21}H_{20}O_{12}$	
22	descurainin A	$C_{26}H_{24}O_{10}$	
23	quercetin-3-*O*-β-D-glucopyranosyl-7-*O*-β-gentiobioside	$C_{33}H_{40}O_{22}$	
24	kaempferol-3-*O*-β-D-glucopyranosyl-7-*O*-β-gentiobioside	$C_{33}H_{40}O_{21}$	
25	isorhamnetin-3-*O*-β-D-glucopyranosyl-7-*O*-β-gentiobioside	$C_{34}H_{42}O_{22}$	
26	quercetin-7-*O*-β-gentiobioside	$C_{27}H_{30}O_{17}$	
27	kaempferol-7-*O*-β-gentiobioside	$C_{27}H_{30}O_{16}$	
28	isorhamnetin-7-O-β-gentiobioside/isorhamnetin-3,7-di-*O*-β-D-glucopyranosyl	$C_{28}H_{32}O_{17}$	
29	quercetin-3,7-di-*O*-β-D-glucopyranoside	$C_{27}H_{30}O_{17}$	
30	kaempferol-3,7-di-*O*-β-D-glucopyranoside	$C_{27}H_{30}O_{16}$	
31	isorhamnetin-3,7-di-*O*-β-D-glucopyranoside	$C_{28}H_{32}O_{17}$	
32	kaempferol-3-*O*-β-D-glucopyranosyl-7-*O*-[(2-*O*-trans-sinnapoyl)-β-D-glucopyranosyl(1→6)]-β-D-glucopyranoside	$C_{44}H_{50}O_{25}$	
33	drabanemoroside	$C_{26}H_{28}O_{14}$	
34	quercetin-3-*O*-α-L-rhamnopyranosyl-(1→2)-α-L-arabinopyranoside	$C_{26}H_{28}O_{15}$	
35	Quercetin-3-*O*-β-D-[6-*O*-sinnapoyl-2-*O*-β-D-glucopyranosyl]-glucopyranoside	$C_{38}H_{40}O_{21}$	
36	isorhamnetin-3-*O*-[2-*O*-(6-*O*-E-sinapoyl)-B-D-glucopymosyl]-B-D-glucopyranoside	$C_{39}H_{42}O_{21}$	
37	quercetin-7-*O*-β-D-glucopyranoside	$C_{21}H_{20}O_{12}$	
38	isorhamnetin-7-*O*-β-D-glucopyranoside	$C_{22}H_{22}O_{12}$	

(Continued)

Table S1 (continued).

No.	Name	Formula	Subclass
39	kaempferol-7-*O*-*β*-D-glucopyranoside	$C_{21}H_{20}O_{11}$	
40	quercetin-7-*O*-*β*-D-glucopyranosyl(1→6)-β-D-glucopyranoside	$C_{27}H_{30}O_{17}$	
41	4'-*O*-methyl-dihydroquercetin	$C_{16}H_{14}O_{7}$	
42	kaempferol 3-*O*-*α*-L-arabinopyranoside-7-*O*-*α*-L-rhamnopyranoside	$C_{26}H_{28}O_{14}$	
43	apigenin-7-*O*-*β*-D-pyranglycuronide	$C_{21}H_{18}O_{11}$	
44	luteolin-7-*O*-*β*-D-glycuronide	$C_{21}H_{18}O_{12}$	
45	(+)-4'-*O*-methylcatechin-7-*O*-*β*-D-glucopyranoside	$C_{22}H_{26}O_{11}$	
46	kaempferol-3-*O*-*β*-D-glycuronide	$C_{21}H_{18}O_{12}$	
47	kaempferol-3-*O*-*β*-D-glucosyl(1→2)-β-D-glucoside	$C_{27}H_{30}O_{16}$	
48	kaempferol-3-*O*-*β*-D-glucosyl(1→2)-7-O-β-D-glucoside	$C_{33}H_{40}O_{21}$	
49	kaempferol 2G-glucosylgentiobioside	$C_{33}H_{40}O_{21}$	
50	quercetin-3-*O*-*β*-D-glucopyranoside	$C_{21}H_{18}O_{13}$	
51	quercetin-3-*O*-*β*-D-glucose-(1→2)-β-D-glucoside	$C_{27}H_{30}O_{17}$	
52	isorhamnetin-3-*O*-sophoroside	$C_{28}H_{32}O_{17}$	
53	quercetin-3-*O*-[(6-*O*-trans-caffeoyl)-*β*-D-glucopyranosyl]-7-*O*-*β*-D-glucopyranoside	$C_{42}H_{46}O_{25}$	
54	apetalumoside A	$C_{34}H_{42}O_{22}$	
55	apetalumoside B1	$C_{43}H_{48}O_{25}$	
56	apetalumoside B2	$C_{43}H_{48}O_{25}$	
57	apetalumoside B3	$C_{39}H_{50}O_{27}$	
58	apetalumoside B4	$C_{48}H_{56}O_{29}$	
59	apetalumoside B5	$C_{48}H_{56}O_{30}$	
60	apetalumoside B6	$C_{49}H_{58}O_{30}$	
61	apetalumoside B7	$C_{49}H_{58}O_{30}$	
62	apetalumoside C	$C_{48}H_{56}O_{29}$	
63	quercetin-3-*O*-(2,6-di-*O*-*β*-D-glucopyranosyl)-*β*-D-glucopyranoside	$C_{33}H_{40}O_{22}$	
64	isorhamnetin-7-*O*-*α*-L-rhamnopyranoside	$C_{22}H_{22}O_{11}$	
65	apetalumoside C2	$C_{44}H_{50}O_{25}$	
66	isorhamnetin-3,4'-*O*-*β*-D-diglucoside	$C_{28}H_{32}O_{17}$	
67	quercetin-3-*O*-*β*-D-galactopyranoside	$C_{21}H_{20}O_{12}$	
68	isorhamnetin-3-*O*-*β*-D-galactopyranoside	$C_{22}H_{22}O_{12}$	
69	kaempferol-3-*O*-*β*-D-galactopyranoside	$C_{21}H_{20}O_{11}$	

(Continued)

Table S1 (continued).

No.	Name	Formula	Subclass
70	kaempferol-3-*O*-*β*-D-xylopyranosyl(1→2)-β-D-glucopyranoside	$C_{26}H_{28}O_{15}$	
71	kaempferol-3-*O*-*β*-D-glucopyranoside	$C_{21}H_{20}O_{11}$	
72	quercetin-3-*O*-*β*-D-arabinopyranoside	$C_{20}H_{18}O_{11}$	
73	quercetin-3-*O*-*β*-D-xylopyranoside	$C_{20}H_{18}O_{11}$	
74	descurainoside	$C_{17}H_{20}O_9S$	Glucoside
75	gluconapin	$C_{11}H_{19}NO_9S_2$	
76	glucoiberverin	$C_{11}H_{21}NO_9S_3$	
77	glucotropaeolin	$C_{14}H_{19}NO_9S_2$	
78	glucoca-ppasalin	$C_{15}H_{27}NO_{10}S_2$	
79	3-hydroxy-5-(methylsulfinyl)pentyl	$C_{13}H_{25}NO_{11}S_3$	
80	3-hydroxy-5-(methylsulfonyl)pentyl	$C_{13}H_{25}NO_{12}S_3$	
81	3-phenylpropionitrile	C_9H_9N	
82	crotononitrile	C_4H_5N	
83	benzylisothiocyanate	C_8H_7NS	
84	allylisothiocyanate	C_4H_5NS	
85	butene isothiocyanate	C_5H_7NS	
86	2-phenylethyl isothiocyanate	C_9H_9NS	
87	4-methylthiobutyl isothiocyanate	$C_6H_{11}NS_2$	
88	diallyl disulfide	$C_6H_{10}S_2$	
89	1-cyano-3,4-epithiobutane	C_5H_7NS	
90	5-methylthiopentanitrile	$C_6H_{11}NS$	
91	1,3-di-*O*-sinapoyl-*β*-D-glucopyranoside	$C_{28}H_{32}O_{14}$	
92	1,2-di-*O*-sinapoyl-*β*-D-glucopyranoside	$C_{28}H_{32}O_{14}$	
93	1,2-disinapoylgentiobiose	$C_{33}H_{40}O_{19}$	
94	helveticoside	$C_{29}H_{42}O_9$	
95	evobioside	$C_{36}H_{56}O_{12}$	
96	erysimoside	$C_{35}H_{52}O_{14}$	
97	raphanuside C	$C_{18}H_{24}O_{10}S$	
98	apetalumoside D	$C_{22}H_{34}O_{13}S_2$	
99	1-thio-*β*-D-glucopyranosyl(1→1)-1-thio-*α*-D-glucopyranoside astragalin	$C_{12}H_{22}O_{10}S_2$	
100	raphanuside	$C_{16}H_{22}O_8S$	
101	raphanuside B	$C_{16}H_{22}O_8S$	
102	raphanuside D	$C_{16}H_{22}O_9S$	

(Continued)

Table S1 (continued).

No.	Name	Formula	Subclass
103	glucosinalbin	$C_{14}H_{19}NO_{10}S_2$	
104	sinigrin	$C_{10}H_{16}KNO_9S_2$	
105	2-phenylethy-1-*O*-*β*-D-glucopyranoside	$C_{13}H_{18}O_6$	
106	4-hydroxybenzyl cyanide	$C_{14}H_{20}O_6$	
107	trilobatin	$C_{21}H_{24}O_{10}$	
108	phloretin-2',4'-di-*O*-*β*-D-glucopyranoside	$C_{27}H_{34}O_{15}$	
109	lepidiumlignan A	$C_{22}H_{26}O_9$	Lignin
110	lepidiumlignan B	$C_{20}H_{20}O_7$	
111	erythro-1-(4-*O*-*β*-D-glucopyranosyl-3-methoxyphenyl)-2-[4-(3-hydroxypropyl)-2,6-dimethoxyphenoxy]-1,3-propanediol	$C_{27}H_{38}O_{13}$	
112	(7R,7′E,8S)-4,9-dihydroxy-3,3′,5-trimethoxy-4′,7-epoxy-8,5′-neolign-7′-en-9′-oic acid	$C_{21}H_{22}O_8$	
113	spicatolignan B	$C_{20}H_{20}O_7$	
114	pinoresinol 4-*O*-*β*-D-glucopyranoside	$C_{26}H_{32}O_{11}$	
115	(+)-isolariciresinol	$C_{20}H_{24}O_6$	
116	syringaresin-4'-*O*-*β*-D-monoglucoside	$C_{28}H_{36}O_{13}$	
117	dimethyl(E,E)-4,4'-dihydroxy-3,3',5,5'-tetramethoxylign-7,7'-dien-9,9'-dioate	$C_{24}H_{26}O_{10}$	
118	isovanillic acid	$C_8H_8O_4$	Organic acid
119	syringic acid	$C_9H_{10}O_5$	
120	p-hydroxy benzoic acid	$C_7H_6O_3$	
121	p-hydroxy benzaldehyde	$C_7H_6O_2$	
122	nicotinic acid	$C_6H_5NO_2$	
123	descurainoside B	$C_{19}H_{26}O_{10}$	
124	sinapic acid	$C_{11}H_{12}O_5$	
125	syringaldehyde	$C_9H_{10}O_4$	
126	3,4,5-trimethoxycinnamic acid	$C_{12}H_{14}O_5$	
127	sinapic acid ethyl ester	$C_{13}H_{16}O_5$	
128	4-pentenamide	C_5H_9NO	Others
129	5-hydroxymethyl furfural	$C_6H_6O_3$	
130	2,5-Dimethyl-7-hydroxy chromone	$C_{11}H_{10}O_3$	
131	*β*-sitosterol	$C_{29}H_{50}O$	
132	*β*-Amyrin	$C_{30}H_{50}O$	
133	cholesterol	$C_{27}H_{46}O$	
134	eleutheroside A	$C_{35}H_{60}O_6$	

(Continued)

Table S1 (continued).

No.	Name	Formula	Subclass
135	sinapine bisulfate	$C_{16}H_{25}NO_9S$	
136	uracil	$C_4H_4N_2O_2$	
137	thymine	$C_5H_6N_2O_2$	
138	3-methoxyinositol	$C_7H_{14}O_6$	
139	aurantiamide acetate	$C_{27}H_{28}N_2O_4$	
140	uridine	$C_9H_{12}N_2O_6$	
141	thymidine	$C_{10}H_{14}N_2O_5$	
142	tormentic acid	$C_{30}H_{48}O_5$	
143	glycerol	$C_3H_8O_3$	
144	lepidiumuridine A	$C_{15}H_{22}N_2O_{11}$	
145	methyl-5-hydroxypyridine-2-carboxlate	$C_7H_7NO_3$	
146	benzylcarbamic acid	$C_8H_9NO_2$	
147	*N*-benzylformamide	C_8H_9NO	
148	1-phenyl-1,2-ethanediol	$C_8H_{10}O_2$	
149	methyl-2,4,6-trihydroxybenzoate	$C_8H_8O_5$	
150	2-(4-hydroxyphenyl) acetonitrile	C_8H_7NO	
151	3-tert-butyl isothioisocyanate	C_5H_9NS	
152	3,4-Dihydroxybenzoic acid	$C_7H_6O_4$	
153	antithiamine factor	$C_{12}H_{14}O_5$	
154	2-hexadecenoic acid	$C_{16}H_{30}O_2$	
155	behenic acid	$C_{22}H_{44}O_2$	
156	2-*O*-(3,4-dihydroxybenzoyl)-2,4,6-trihydroxy phenylacetic acid 4-*O*-*β*-D-glucopyranoside	$C_{21}H_{22}O_{13}$	
157	4,9-di-*O*-*β*-D-glucosyl sinapoyl alcohol	$C_{23}H_{34}O_{14}$	
158	3',5'-dimethoxy-4-*O*-*β*-D-glucopyranosyl cinnamic acid	$C_{17}H_{22}O_{10}$	
159	sinapoylglucose	$C_{17}H_{22}O_{10}$	
160	sinapoyl-9-sucrosecoside	$C_{23}H_{32}O_{15}$	
161	lariciresinol 4'-*O*-*β*-D-glucopyranoside	$C_{26}H_{34}O_{11}$	
162	(7S,8R)-aegineoside	$C_{26}H_{30}O_{12}$	
163	*L*-tryptophan	$C_{11}H_{12}N_2O_2$	
164	adenosine	$C_{10}H_{13}N_5O_4$	
165	stachyose	$C_{24}H_{42}O_{21}$	
166	1,2-benzenedicarboxylic acid-1,2-bis[2-(2-hydroxyethoxy) ethyl] ester	$C_{16}H_{22}O_8$	
167	uridine	$C_{13}H_{14}N_2O_3$	

(Continued)

Table S1 (continued).

No.	Name	Formula	Subclass
168	methyl dioxindole 3-acetate	$C_{11}H_{11}NO_4$	
169	dioxindole-3-acetic acid	$C_{10}H_9NO_4$	
170	cyclo(*L*-Pro-*L*-Phe)	$C_{14}H_{16}N_2O_2$	
171	4-amino-4-carboxychroman-2-one	$C_{10}H_9NO_4$	
172	sucrose	$C_{12}H_{22}O_{11}$	
173	(S)-p-hydroxyphenyl lactate acid	$C_9H_{10}O_4$	
174	(S)-2-hydroxy-phenylpropionic acid	$C_9H_{10}O_3$	
175	linolenic acid	$C_{18}H_{30}O_2$	
176	linolic acid	$C_{18}H_{32}O_2$	
177	palmitic acid	$C_{16}H_{32}O_2$	
178	oleic acid	$C_{18}H_{34}O_2$	
179	myristic acid	$C_{14}H_{28}O_2$	
180	stearic acid	$C_{18}H_{36}O_2$	
181	arachidic acid	$C_{20}H_{40}O_2$	
182	eicosenoic acid	$C_{20}H_{38}O_2$	
183	sinapine	$C_{16}H_{24}NO_5^+$	
184	(13Z,16Z)-docosadienoic acid	$C_{22}H_{40}O_2$	
185	cis-11,14,17-eicosatrienoic acid	$C_{20}H_{34}O_2$	
186	erucic acid	$C_{22}H_{42}O_2$	
187	protocatechuic aldehyde	$C_7H_6O_3$	Phenolic
188	2-(4-hydroxy-phenyl)-ethanol	$C_8H_{10}O_2$	
189	protocatechuic acid methyl ester	$C_8H_8O_4$	
190	epigallocatechin	$C_{15}H_{14}O_7$	
191	dimethylthomasidioate	$C_{24}H_{26}O_{10}$	
192	carbamic acid	$C_8H_{15}NO_3$	Phenolic acid
193	acetamide	C_2H_5NO	
194	2-phenylacetamide	C_8H_9NO	
195	cis-desulfoglucotropaeolin	$C_{14}H_{19}NO_6S$	Phenylacetamide
196	trans-desulfoglucotropaeolin	$C_{14}H_{19}NO_6S$	
197	*N*-acetyltryptophan	C_8H_7NO	Phenylacetonitrile
198	*N*-benzyl-2-hydroxy-2-phenylacetamide	$C_{15}H_{16}N_2O$	Phenylethylamine
199	5-hydroxy-1-phenymethyl-2-pyrrolidinone	$C_{11}H_{13}NO_2$	
200	5-methoxy-1-phenymethyl-2-pyrrolidinone	$C_{12}H_{15}NO_2$	
201	descuraic acid	$C_{21}H_{22}O_8$	Phenylpropanoid
202	descurainolide A	$C_{13}H_{16}O_5$	

(Continued)

Table S1 (continued).

No.	Name	Formula	Subclass
203	descurainolide B	$C_{21}H_{22}O_8$	
204	descurainin	$C_{16}H_{18}O_6$	
205	syringaresinol	$C_{22}H_{26}O_8$	
206	scopoletin	$C_{10}H_8O_4$	
207	scopoline	$C_{16}H_{18}O_9$	
208	isoscopoletin	$C_{10}H_8O_4$	
209	xanthtoxol	$C_{11}H_6O_4$	
210	xanthtoxin	$C_{12}H_8O_4$	
211	psoralene	$C_{11}H_6O_3$	
212	bergaptane	$C_{12}H_8O_4$	
213	lanosterol	$C_{30}H_{50}O$	Steroid
214	descuraic anhydride A	$C_{20}H_{21}O_7N_2^+$	Tobacco flavor anhydride
215	descuraic anhydride B	$C_{21}H_{23}O_8N_2^+$	
216	descuraic anhydride C	$C_{19}H_{19}N_2O_6^+$	
217	descuraic cyclolign anhydride A	$C_{42}H_{44}O_{16}N_{42}^+$	
218	descuraic cyclolign anhydride B	$C_{42}H_{44}O_{16}N_{42}^+$	
219	descuraic cyclolign anhydride C	$C_{42}H_{44}O_{16}N_{42}^+$	
220	descuraic cyclolign anhyride dimer	$C_{66}H_{68}N_4O_{262}^+$	
221	descuraic cyclolign anhydride amide	$C_{32}H_{29}N_2O_{12}^+$	
222	caulophyllogenin	$C_{30}H_{48}O_5$	Triterpenoid
223	lepidiumuridine B	$C_{22}H_{26}N_2O_{13}$	Uridine derivative
224	lepidiumuridine C	$C_{22}H_{26}N_2O_{12}$	
225	lepidiumuridine D	$C_{20}H_{30}N_2O_{12}$	
226	lepidiumuridine E	$C_{20}H_{30}N_2O_{13}$	
227	lepidiumuridine F	$C_{32}H_{33}N_3O_{15}$	
228	lepidiumuridine G	$C_{32}H_{33}N_3O_{15}$	
229	lepidiumuridine H	$C_{22}H_{31}N_3O_{13}$	
230	lepidiumuridine I	$C_{25}H_{29}N_3O_{14}$	
231	lepidiumuridine J	$C_{25}H_{29}N_3O_{14}$	
232	lepidiumuridine K	$C_{16}H_{16}N_2O_7$	

Table S2: Information of the multicomponents characterized from *Descurainia sophia* in the current work

No.	Observed RT (min)	*m/z*	M.F.	Mass error (ppm)	Adducts	Observed CCS (Å²)	ESI-MS2	Identification	Subclass
1	0.87	665.2140	$C_{24}H_{42}O_{21}$	–0.8	–H	239.19	485.1514, 383.1190, 221.0662, 179.0555	stachyose	Other
2	0.92	505.1674	$C_{20}H_{30}N_2O_{13}$	–0.2	–H	208.13	221.0661	lepidiumuridine E or isomer	Uridine derivative
3	0.93	341.1083	$C_{12}H_{22}O_{11}$	–1.8	–H	172.35	179.0555	sucrose	Other
4	0.94	268.1027	$C_{10}H_{13}N_5O_4$	–4.9	+H	152.25	251.0899	adenosine	Other
5	6.60	533.1875	$C_{23}H_{34}O_{14}$	–0.1	–H	227.60	315.0504	4,9-di-*O*-*β*-D-glucosyl sinapoyl alcohol or isomer	Other
6	6.71	949.2469	$C_{39}H_{50}O_{27}$	0.3	–H	284.42	625.1404, 301.0346	apetalumoside B3	Flavonoid
7	7.35	787.1941	$C_{33}H_{40}O_{22}$	0.3	–H	271.47	625.1406, 463.0882, 343.0455, 301.0349	quercetin-3-*O*-*β*-D-glucopyranosyl-7-*O*-*β*-gentiobioside	Flavonoid
8	7.35	627.1563	$C_{27}H_{30}O_{17}$	1.2	+H	261.13	465.1029, 303.0490, 273.0383	quercetin-3,7-di-*O*-*β*-D-glucopyranoside	Flavonoid
9	7.35	465.1029	$C_{21}H_{20}O_{12}$	0.3	+H	256.88	303.0490	quercetin-7-*O*-*β*-D-glucopyranoside or isomer	Flavonoid
10	7.35	463.0872	$C_{21}H_{18}O_{12}$	0.3	+H	236.60	355.9607, 301.0333, 273.0383	kaempferol-3-*O*-*β*-D-glycuronide	Flavonoid
11	7.94	372.0425	$C_{11}H_{19}NO_9S_2$	–1.0	–H	177.24	259.0119, 241.0016	gluconapin	Glucosinolate
12	8.75	625.1408	$C_{27}H_{30}O_{17}$	–0.4	–H	251.16	463.0871, 301.0344	quercetin-7-*O*-*β*-gentiobioside	Flavonoid
13	8.76	465.1027	$C_{21}H_{20}O_{12}$	–0.1	+H	249.14	415.9821, 303.0487, 175.0379	quercetin-3-*O*-*β*-D-glucopyranoside or isomer	Flavonoid
14	8.77	385.1134	$C_{17}H_{22}O_{10}$	–1.5	–H	199.30	247.0606, 217.0136, 205.0502, 175.0029, 149.0241	sinapoylglucose or isomer	Other
15	8.95	771.1990	$C_{33}H_{40}O_{21}$	0.1	–H	273.72	609.1458, 447.0931, 285.0398		Flavonoid

(Continued)

Table S2 (continued).

No.	Observed RT (min)	*m/z*	M.F.	Mass error (ppm)	Adducts	Observed CCS (Å²)	ESI-MS2	Identification	Subclass
								kaempferol-3-*O*-*β*-D-glucopyranosyl-7-*O*-*β*-gentiobioside	
16	8.95	449.1076	$C_{21}H_{20}O_{11}$	–0.6	+H	245.05	399.9866, 287.0537	kaempferol-3-*O*-*β*-D-glucopyranoside	Flavonoid
17	9.34	801.2093	$C_{34}H_{42}O_{22}$	–0.2	–H	270.36	639.1563, 477.1036, 315.0504	apetalumoside A	Flavonoid
18	9.34	641.1716	$C_{28}H_{32}O_{17}$	0.6	+H	263.38	479.1181, 317.0645	isorhamnetin-3,7-di-*O*-*β*-D-glucopyranoside	Flavonoid
19	9.56	385.1141	$C_{17}H_{22}O_{10}$	0.2	–H	200.33	205.0503, 175.0031	(E)-2-*O*-sinapoly-D-glucopyranoside isomer	Other
20	10.04	1155.3060	$C_{50}H_{60}O_{31}$	1.3	–H	318.81	949.2506, 625.1422, 329.1397	apetalumoside B11	Flavonoid
21	10.12	387.1251	$C_{17}H_{22}O_{10}$	–8.9	+H	148.64	353.2251, 335.1960, 265.1248, 207.0637, 175.0376	3',5'-dimethoxy-4-*O*-*β*-D-glucopyranosyl cinnamic acid	Other
22	10.39	609.1460	$C_{27}H_{30}O_{16}$	–0.1	–H	251.21	447.0927, 285.0396	kaempferol-3,7-di-*O*-*β*-D-glucopyranoside	Flavonoid
23	10.86	639.1561	$C_{28}H_{32}O_{17}$	–0.9	–H	253.91	519.1147, 477.1031, 461.0725, 357.0606, 315.0502	isorhamnetin-3,4'-*O*-*β*-D-diglucoside	Flavonoid
24	10.87	479.1180	$C_{22}H_{22}O_{12}$	–0.9	+H	251.59	429.9975, 317.0644	isorhamnetin-7-*O*-*β*-D-glucopyranoside	Flavonoid
25	10.98	547.1667	$C_{23}H_{32}O_{15}$	–0.2	–H	230.24	385.1135, 367.1025, 325.0926, 273.9774, 265.0712, 223.0605, 149.0242	sinapoyl-9-sucrosecoside	Other

(Continued)

Table S2 (continued).

No.	Observed RT (min)	*m/z*	M.F.	Mass error (ppm)	Adducts	Observed CCS (Å²)	ESI-MS2	Identification	Subclass
26	11.19	993.2503	$C_{44}H_{50}O_{26}$	–1.5	–H	297.27	831.1961, 787.1929, 625.1403, 463.0874, 301.0343	apetalumoside B10	Flavonoid
27	11.20	995.2646	$C_{44}H_{50}O_{26}$	–1.7	+H	300.65	465.1022, 369.1171, 207.0639, 175.0378	apetalumoside B8	Flavonoid
28	12.25	977.2565	$C_{44}H_{50}O_{25}$	–0.3	–H	301.01	815.2031, 529.1563, 447.0928, 285.0398, 223.0608	apetalumoside C2	Flavonoid
29	12.26	979.2698	$C_{44}H_{50}O_{25}$	–1.6	+H	299.87	449.1072, 369.1172, 351.1065, 287.0537, 207.0637, 175.0377	apetalumoside C2	Flavonoid
30	14.06	521.2033	$C_{26}H_{34}O_{11}$	0.8	–H	224.41	329.1385	lariciresinol-4'-*O*-*β*-D-glucopyranoside or isomer	Other
31	14.51	597.1820	$C_{27}H_{34}O_{15}$	–0.8	–H	226.75	435.1307, 297.0763, 273.0763, 167.0347	phloretin-2',4'-di-*O*-*β*-D-glucopyranoside	Glycoside
32	14.51	435.1297	$C_{21}H_{24}O_{10}$	0.0	–H	227.03	273.0763, 167.0347	trilobatin	Glycoside
33	15.45	533.1666	$C_{26}H_{30}O_{12}$	0.2	–H	236.73	353.1024	(7S,8R)-aegineoside or isomer	Other
34[a]	15.62	463.0879	$C_{21}H_{20}O_{12}$	–0.8	–H	202.09	301.0334, 271.0239, 255.0290, 241.0136, 151.0030	quercetin-7-*O*-*β*-D-glucopyranoside	Flavonoid
35[a]	15.62	461.0725	$C_{21}H_{18}O_{12}$	–0.1	–H	266.20	301.0334, 271.0239	kaempferol-3-*O*-*β*-D-glycuronide	Flavonoid
36	16.15	579.1355	$C_{26}H_{28}O_{15}$	–0.2	–H	225.73	271.0237	kaempferol-3-*O*-*β*-D-xylopyranosyl(1→2)-*β*-*D*-glucopyranoside or isomer	Flavonoid

(Continued)

Table S2 (continued).

No.	Observed RT (min)	*m/z*	M.F.	Mass error (ppm)	Adducts	Observed CCS (Å²)	ESI-MS2	Identification	Subclass
37	16.32	433.0775	$C_{20}H_{18}O_{11}$	–0.2	–H	198.55	271.0238	quercetin-3-*O*-*β*-D-arabinopyranoside or isomer	Flavonoid
38	16.38	305.0696	$C_{15}H_{14}O_7$	9.5	–H	159.65	225.1123	epigallocatechin or isomer	Phenolic
39	16.52	465.1440	$C_{22}H_{26}O_{11}$	8.1	–H	202.21	283.1086, 161.0480	(+)-4'-*O*-methylcatechin-7-*O*-*β*-D-glucopyranoside or isomer	Other
40[a]	17.03	447.0931	$C_{21}H_{20}O_{11}$	–0.3	–H	200.77	285.0383, 255.0293, 227.0344	kaempferol-3-*O*-*β*-D-glucopyranoside	Flavonoid
41	17.21	477.1038	$C_{22}H_{22}O_{12}$	0.0	–H	208.40	357.0609, 285.0398, 271.0239, 243.0290, 215.0341	isorhamnetin-7-*O*-*β*-D-glucopyranoside	Flavonoid
42	17.39	563.1407	$C_{26}H_{28}O_{14}$	0.1	–H	224.21	284.0319	kaempferol-3-*O*-*α*-L-arabinopyranoside-7-*O*-*α*-L-rhamnopyranoside	Flavonoid
43	18.75	447.0932	$C_{21}H_{20}O_{11}$	–0.1	–H	203.87	285.0410, 271.0239	kaempferol-3-*O*-*β*-D-galactopyranoside isomer	Flavonoid
44	20.51	473.1455	$C_{24}H_{26}O_{10}$	0.3	–H	217.23	384.0853	dimethyl(E,E)-4,4'-dihydroxy-3,3',5,5'-tetramethoxylign-7,7'-dien-9,9'-dioate or isomer	Lignin
45	28.5	739.2026	$C_{33}H_{40}O_{19}$	–8.8	–H	268.30	651.2473	1,2-disinapoylgentiobiose or isomer	Glucoside
46[a]	31.44	477.0676	$C_{22}H_{22}O_{13}$	0.3	–H	206.39	279.2316, 199.8504	isorhamnetin-3-*O*-*β*-D-glucopyranoside	Flavonoid
47	31.44	277.2167	$C_{18}H_{30}O_2$	–2.3	–H	179.97	199.8504	linolenic acid or isomer	Other
48	31.91	227.2012	$C_{14}H_{28}O_2$	–1.8	–H	166.66	199.8506	myristic acid or isomer	Other
49	32.41	253.2167	$C_{16}H_{30}O_2$	–2.2	–H	172.05	199.8504	2-hexadecenoic acid or isomer	Other

(Continued)

Table S2 (continued).

No.	Observed RT (min)	*m/z*	M.F.	Mass error (ppm)	Adducts	Observed CCS (Å²)	ESI-MS2	Identification	Subclass
50	32.70	435.0922	$C_{20}H_{18}O_{11}$	0.0	+H	188.81	333.1511, 185.0753	quercetin-3-*O*-*β*-D-arabinopyranoside	Flavonoid
51	32.71	279.2325	$C_{18}H_{32}O_2$	–1.5	–H	180.42	199.8505	linolic acid or isomer	Other
52	34.22	481.0997	$C_{21}H_{22}O_{13}$	1.9	–H	210.87	297.1523, 199.8505	2-*O*-(3,4-dihydroxybenzoyl)-2,4,6-trihydroxy phenylacetic acid 4-*O*-*β*-D-glucopyranoside or isomer	Other
53	34.22	281.2485	$C_{18}H_{34}O_2$	–0.3	–H	181.33	199.8505	oleic acid or isomer	Other

Note: [a]the components identified by comparing with the reference compounds.