Article published online: 2022-08-15

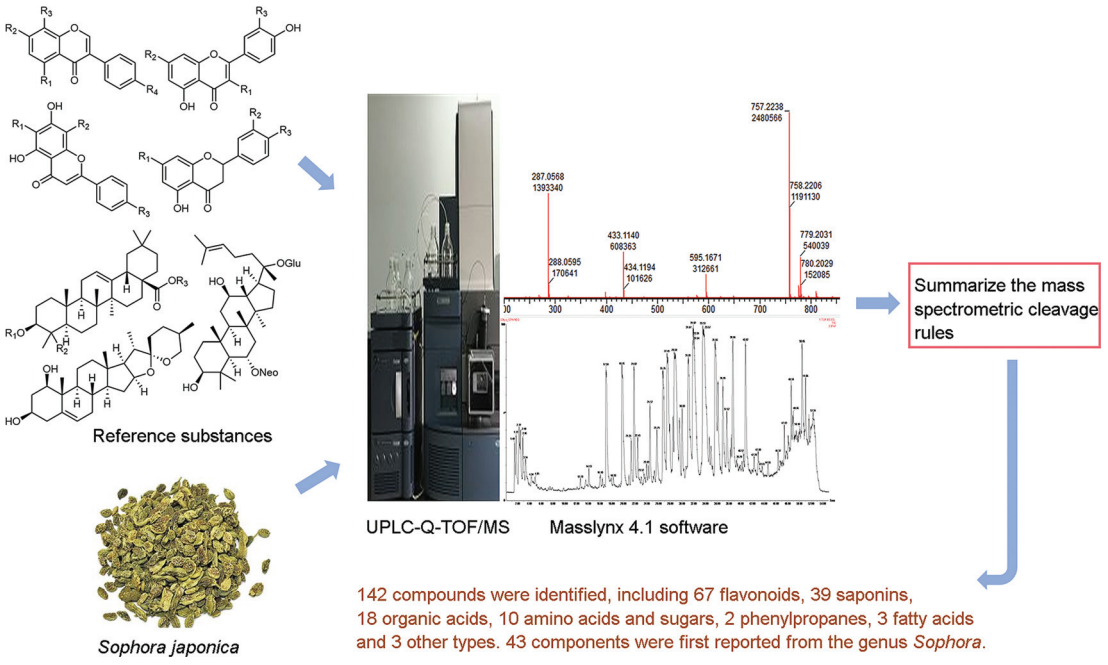
Structural Characterization of Chemical Compounds Based on Their Fragmentation Rules in *Sophorae Fructus* by UPLC-QTOF-MS/MS

Zi-Hui He[1] Mo Liu[1] Jun-Xuan Ren[1] Dan-Wei Ouyang[1,*]

[1] State Key Laboratory of New Drug and Pharmaceutical Process, Shanghai Institute of Pharmaceutical Industry Co., Ltd., China State Institute of Pharmaceutical Industry, Shanghai, People's Republic of China

Address for correspondence Dan-wei Ouyang, PhD, Shanghai Institute of Pharmaceutical Industry Co., Ltd., China State Institute of Pharmaceutical Industry, 285 Gebaini Road, Shanghai 201203, People's Republic of China (e-mail: ouyangdanwei@163.com).

Pharmaceut Fronts 2022;4:e162–e178.



Abstract

This study aims to identify the chemical components in *Sophorae Fructus*, and explore the mass spectrometric cleavage rules using the UPLC-Q-TOF-MS/MS method. The main characteristic fragments of the compounds were analyzed by electrospray ionization (ESI) ion source under positive and negative ion modes. The compounds were identified by molecular formula, multistage mass spectrometry, ultraviolet spectrum, and the fragmentation patterns of standards. A total of 142 compounds were identified, including 68 flavonoids, 39 saponins, 21 organic acids, and 14 others, of which 43

Keywords

- *Sophorae Fructus*
- flavonoid glycosides
- saponins
- UPLC-Q-TOF-MS/MS
- fragmentation

DOI https://doi.org/10.1055/s-0042-1751315.
ISSN 2628-5088.

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

components were reported from *Sophora* for the first time. Moreover, the mass spectrometric fragmentation rules of some identified species components were deduced, which are helpful for the structural analysis of flavonoid and saponins. This method provides a reference for the rapid identification of chemical components and is conducive to further study the pharmacodynamic material basis and action mechanism of *Sophorae Fructus*.

Introduction

Sophorae Fructus is the dry and mature fruit of *Sophora japonica* (L.), a leguminous plant. It has the functions of clearing away heat and toxic material, cooling blood and stopping bleeding, and is usually used for treating intestinal heat, hematochezia, nevus swelling and bleeding, dizziness, as well as red eyes.[1] It also has anticancer and estrogen-like effects, and plays a roles in prevention and treatment of cardiovascular disease, osteoporosis, and female menopause syndrome.[2] The study of the chemical components of *S. japonica* is of great significance for its quality control and clinical application. The main components of *Sophorae Fructus* are flavonoids, isoflavones, alkaloids, triterpenoid saponins, amino acids, stearic acids etc., among which isoflavones and their glycosides are the highest. However, up to now, there are few reports on the analysis of the total components of *Sophorae Fructus*. Sun et al identified and inferred 24 common compounds and 21 variance compounds in *Sophorae Fructus* from different producing areas by ultra-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-Q-TOF/MS).[3] Zhang identified 131 compounds from the total extract of *Sophorae Fructus* in positive and negative ion modes by UPLC-Q-TOF/MS.[4] The structural types include 81 flavonoids, 18 triterpenoid saponins, 5 steroids, 2 anthraquinones, 3 phenols, and 22 others. Triterpenoid saponins were less identified, and flavonoids were mainly reported.

Traditional methods for phytochemical isolation and identification are time-consuming and labor-intensive. In recent years, techniques combining the efficient separation ability of liquid chromatography and strong identification ability of MS have been widely used in the separation and qualitative and quantitative analysis of complex Chinese medicines.[5,6] In this study, UPLC-Q-TOF/MS was used to rapidly identify the chemical composition of *Sophorae Fructus*. We established a UPLC-Q-TOF/MS qualitative analysis method to analyze the constituents of *Sophorae Fructus*, which lay a foundation for the study of its pharmacodynamic substance basis and quality control.

At present, the studies on MS pyrolysis are relatively scattered. The electrospray ionization MS of saponin[7,8] and flavonoid[9,10] components has been reported. Flavonoid species can be identified according to the ultraviolet (UV) absorption characteristics of compounds and the characteristic ion fragments of the parent nuclei. But there are few literature reports on determining the connection mode between glycosyl groups in flavonoid glycosides by using the cleavage rule of MS. In this study, through the comparison of the MS data of 21 reference substances, including flavonoid oxyglycosides, flavonoid carbon glycosides, dihydro-flavonoid glycosides, isoflavone glycosides and saponins, and a large number of literature reports, we systematically deduced the cleavage characteristics of these compounds, so as to provide reference for the MS structure identification of such components.

Materials and Methods

Materials and Reagents

The prepared slices of *Sophorae Fructus* is obtained from Tianjin Darentang TCM Chinese Herbal Medicine Co., Ltd (Bozhou, China). Reference standards citric acid (97.0%), gallic acid (90.8%), protocatechuic acid (97.7%), protocatechuic aldehyde (99.6%), baicalein (97.9%), apigenin (99.4%), rutin (91.6%), kaempferol (93.2%), isoquercitrin (97.2%), genistein (98.8%), kaempferol-3-*O*-rutinoside (94.0%), kaempferol-3-*O*-gentiobioside (93.1%), isorhamnetin-3-*O*-neohesperidin (93.0%), naringin (93.5%), hesperidin (95.3%), neohesperidin (99.4%), genistin (99.9%), sophoricoside (99.6%), puerarin (95.4%), vitexin (99.1%), asperosaponin VI (94.3%), mogroside V (96.1%), ginsenoside Re (96.0%), jujuboside A (96.0%), ruscogenin (98.0%), and oleanolic acid (95.8%) were purchased from National Institute for Food and Drug Control (Beijing, China). Sophorabioside (≥98%) was purchased from Shanghai Standard Biotech Co., Ltd (Shanghai, China). Kaempferol-3-*O*-sophoroside (≥98%) was purchased from Chengdu Purifa Technology Development Co., Ltd. Kaempferol-3,7-di-*O*-glucoside (≥98%) and kaempferol-3-*O*-(2″-*O*)-β-*D*-glucosyl)-β-*D*-rutinoside (≥98%) was purchased from Chengdu Push Biotechnology Co., Ltd (Chengdu, China). Isorhamnetin-3-*O*-β-*D*-rutinoside (≥98%) and kaempferol-3-*O*-β-*D*-sophorae-7-*O*-α-*L*-rhamnoside (≥98%) were self-made in the laboratory. Liquid chromatography-MS (LC-MS)-grade acetonitrile (ThermoFisher, United States), methanol (ThermoFisher, United States), formic acid (ThermoFisher, United States), and deionized water prepared by a Millipore Alpha-Q water purification system (Millipore, United States) were used as the mobile phase for the chromatographic separation. Other reagents were of analytical grade.

Preparation of Standards and Samples

All reference materials were dissolved in methanol and each was prepared into a solution of 0.1 mg/mL. In brief, 1.0 g of *Sophorae Fructus* powder (through No. 3 sieve) was accurately weighed and ultrasonicated with 30 mL 70% methanol

(v/v) (250 W, 40 kHz) for 60 minutes. The sample solution and standard solution were filtered through 0.22 μm microporous filter membrane.

Instrumentation and Conditions

The UPLC-QTOF MS/MS analysis was performed using a Waters Acquity UPLC system coupled with a Xevo G2-XS QTOF mass spectrometer (Waters, United States) with an electrospray ionization ion source in MS^E mode.

The chromatographic separation process of flavonoids was performed on an ACQUITY CSH C18 (150 mm × 2.1 mm, 1.7 μm; Waters, United States) at 35°C, with a mobile phase consisting of methanol (B) and 0.05% formic acid aqueous solution (A). The gradient elution was as follows: 0–9 minutes, 10–20% eluent B; 9–27 minutes, 20–40% eluent B; 27–30 minutes, 40% eluent B; 30–39 minutes, 40–60% eluent B; 39–42 minutes, 60% eluent B; 42–48 minutes, 60–80% eluent B; 48–50 minutes, 80% eluent B; 50–50.1 minutes, 80–10% eluent B; 50.1–65 minutes,10% eluent B. The flow rate was 0.2 mL/min.

Saponins were separated by Hypersi1 Gold (100 mm × 2.1 mm, 1.9 μm; ThermoFisher Scientific, United States) at 35°C. The flow rate was 0.3 mL/min. The mobile phase was acetonitrile (B) and 0.1% formic acid (A) in water. The gradient elution was as follows: 0–1 minute, 99% eluent A; 1–5 minutes, 99–91% eluent A; 5–9 minutes, 91–84% eluent A; 9–12 minutes, 84% eluent A; 12–18 minutes, 84–67% eluent A; 18–23 minutes, 67–63% eluent A; 23–29 minutes, 63–49% eluent A; 29–34 minutes, 49–0% eluent A; 34–36 minutes, 0–99% eluent A; 36–37 minutes, 99% eluent A. The injection volume for all was 1 μL.

MS conditions were operated in both positive and negative ion modes and applied as the following: solvent gas temperature (nitrogen), 450°C; capillary voltage, 3.0/2.5 KV; an ion source temperature, 120°C; desolvation gas flow, 500 L/h; cone gas flow, 100 L/h; the low collision energy, 6 V; the high collision energy, 25 to 60 V.

Data Processing and Compound Identification

Masslynx 4.1 software (Waters, United States) was used to analyze the mass spectra peaks of *Sophorae Fructus* in positive and negative ion modes. According to the comparison of reference standards or references, the compounds were identified by UV spectrum, retention time, excimer ion peak, molecular formula, fragment ions, and other information combined with Scifinder database.

Results and Discussion

To systematically and qualitatively analyze the chemical components in *Sophorae Fructus*, the MS behavior of the existing reference standards was studied to summarize their chromatographic retention behavior, UV absorption, cracking rule, and characteristic fragment ions.

The Cracking Rules of the Deglycosylation Group of Flavone-*O*-diglycoside

Kaempferol-3-*O*-sophoroside (t_R = 26.28 minutes) and kaempferol-3-*O*-gentiobioside (t_R = 28.33 minutes) are isomers, their mass spectra in negative and positive ion modes are shown in ►**Fig. 1**. In the negative ion mode (►**Fig. 1A**), kaempferol-3-*O*-sophoroside can obtain fragment ions of *m/z* 429.0819 $[M - H - 162 - H_2O]^-$. In the positive ion mode (►**Fig. 1C**), it could generate fragment ions of *m/z* 449.1071 $[M + H - 162]^+$, and the relative abundance (>10%) was higher than that of kaempferol-3-*O*-gentiobioside (<10%, ►**Fig. 1D**). Similarly, isorhamnetin-3-*O*-β-*D*-neohesperidoside produces fragment ions of *m/z* 459.0932 $[M - H - 146 - H_2O]^-$ in the negative ion mode. The relative abundance of *m/z* 479.1197 $[M + H - 146]^+$ fragment ion (>50%) was higher than that of isorhamnetin-3-*O*-β-*D*-rutinoside (<50%) in the positive ion mode. The retention time and ion fragments of the four reference substances showed the following regularities: (1) the polarity of flavone-*O*-diglycoside linked to monosaccharides in 1→2 mode was greater than that of flavone-*O*-diglycoside linked to monosaccharides in 1→6 mode; (2) in the negative ion mode, when the flavone-*O*-diglycoside is linked in 1→2 mode, it can produce [M – H – monosaccharide - $H_2O]^-$ characteristic fragment ion, but when it is linked in 1→6 mode, it can only produce [M – H - monosaccharide]$^-$ fragment ion, which is the same as reported in the literature[11]; (3) the relative abundance of $[M + H$ - monosaccharide$]^+$ fragment ions produced by 1→2 linkage between glycogroups is higher than that of the same fragment ions produced by 1→6 linkage between glycogroups. It is consistent with the cleavage law of *Fructus aurantii* flavone diglycosides in the positive ion mode reported in the literature.[12] The rules can provide a basis for identifying the most common two disaccharide connection modes (1→6, 1→2) in flavonoid oxyglycosides.

The Cracking Rules of the Deglycosylation Group of Flavone-*O*-triglycoside

Kaempferol-3-*O*-β-*D*-sophoroside-7-*O*-α-*L*-rhamnoside (t_R = 19.66 minutes) and kaempferol-3-*O*-(2″-*O*-β-*D*-glucopyranosyl)-β-*D*-rutinoside (t_R = 25.33 minutes) are isomers. Their mass spectrometric cleavage pathways in positive and negative ion modes are shown in ►**Fig. 2**. In the negative ion mode, kaempferol-3-*O*-β-*D*-sophoroside-7-*O*-α-*L*-rhamnoside can obtain fragment ions of *m/z* 755.1894, 609.1498, 449.1126, and 284.0459, indicating that rhamnose on the C_7 position was lost first and then glucose groups on the C_3 position were lost successively. However, in the positive ion mode, the glycogroup at the end of C_3 site was lost first, and the fragment ion of *m/z* 595.1671 was detected. Then, after the loss of all glycogroups at the C_3 site, the rhamnose group at the C_7 site was lost, and the fragment ion with *m/z* of 433.1140 and 287.0568 appeared (►**Fig. 2A**). Kaempferol-3-*O*-(2″-*O*-β-*D*-glucopyranosyl)-β-*D*-rutinoside in the negative ion mode could generate fragment ions of *m/z* 755.1838, 593.1495, 575.1411, and 284.0424. In the positive ion mode, it could generate fragment ions of *m/z* 779.1974, 595.1671, 493.1533, 449.1071, and 287.0568. This shows that whether in the positive or the negative ion mode, the glucose connected with 1→2 at the end was lost first, then the rhamnose connected with 1→6 was lost, and finally the glucose connected with aglycone was lost (►**Fig. 2B**).

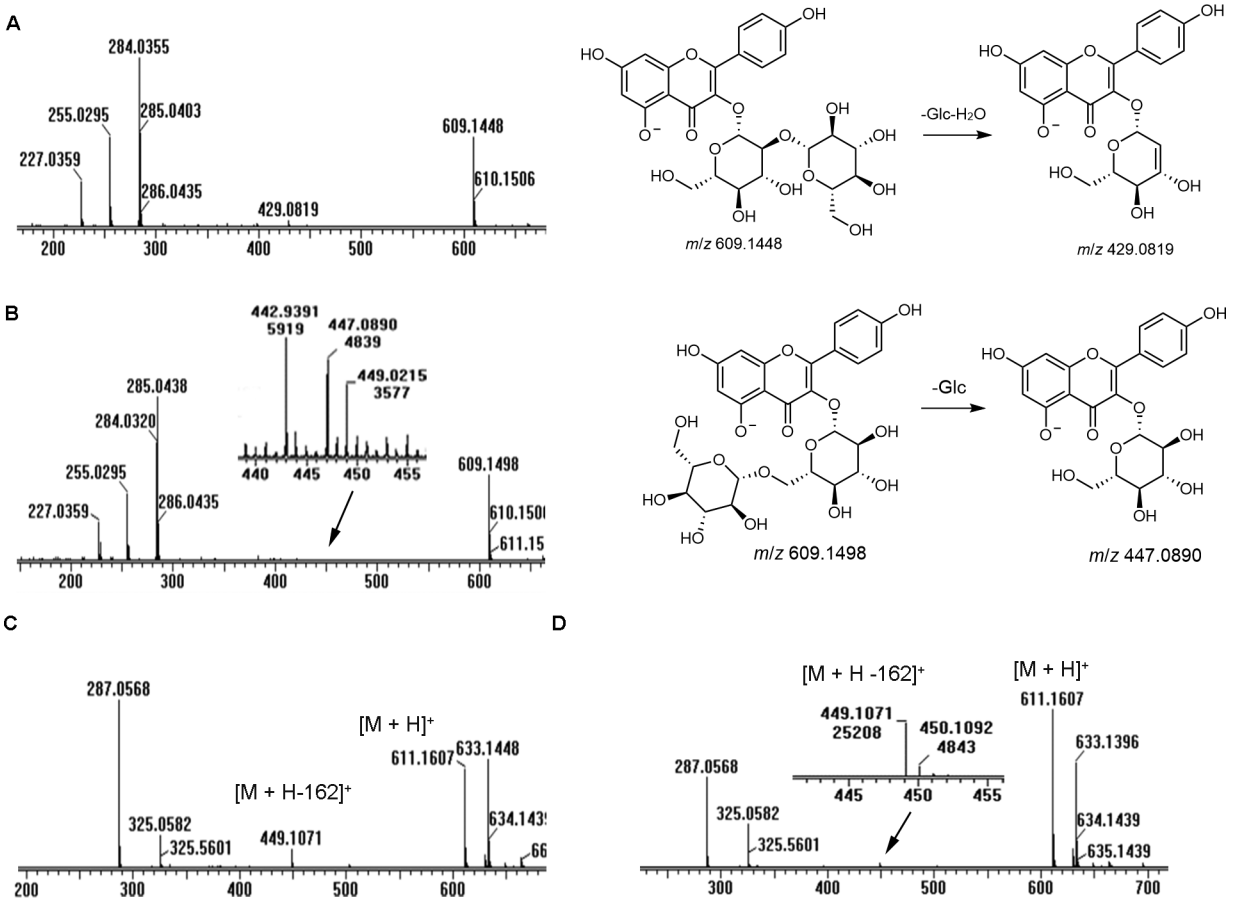


Fig. 1 Mass spectra and cleavage of (A) kaempferol-3-*O*-sophoroside and (B) kaempferol-3-*O*-gentiobioside in negative ion mode, and mass spectra of (C) kaempferol-3-*O*-sophoroside and (D) kaempferol-3-*O*-gentiobioside in positive ion mode.

Therefore, we come to the conclusion that: (1) the polarity of flavonol glycoside substitution at the C_3 and C_7 sites is greater than that of glycoside substitution at the C_3 site only. (2) Flavonol glycosides replaced by glycogroups at the C_3 and C_7 sites lose the glycogroups on the C_7 position first and then the glycogroups on the C_3 position is lost in turn in the anion mode. In the positive ion mode, the glycogroups at the end of position C_3 were lost successively, and then the glycogroups at the C_7 site were lost, which was consistent with the pyrolysis rule of flavonoids in *Herba Epimedii* in the positive ion mode described in the literature.[13] (3) The three monosaccharides in flavone-*O*-triglycoside are connected to each other. Whether in the positive or negative electrode, the glycosyl connected at the end with 1→2 is lost first, then the glycosyl connected with 1→6 is lost, and finally the loss of the glycosyl connected with aglycone.

Cleavage Rules of Flavonoid Carboglycosides

The mass spectrogram of puerarin in positive and negative ion modes showed fragment ions of *m/z* 325.0699 $[M - H - 90]^-$, 295.0578 $[M - H - 120]^-$, 297.0750 $[M + H - 120]^+$, and a series of dehydration peaks *m/z* 399.1077/381.0974/363.0857 were generated by the ionization peaks of $[M + H]^+$ *m/z* 417.1158, and $[M - H - 120]^-$ and $[M + H - 120]^+$ are the main characteristic fragments with high abundance (►**Fig. 3**). Vitexin showed the same cleavage pattern, indicating that if the fragment peak of the disaccharide group does not appear first, but there are $[M - H - 90]^-$and $[M - H - 120]^-$ion fragments and $[M - H - 120]^-$or $[M + H - 120]^+$ are the main characteristics, the fragments can basically be determined as hexacarbon flavonoid carboglycoside compounds. This is consistent with the research of Liu et al.[14] Meanwhile, according to relevant literature,[14–17] in the positive ion scanning mode, the continuous dehydration of glycosyl mainly occurred, and the negative ion scanning mode has more obvious mass spectrum characteristics than the positive ion scanning mode.

Cleavage of Dihydroflavonoid Glycosides and Isoflavone Glycosides

Through the secondary mass spectra of naringin ($t_R = 29.63$ minutes), hesperidin ($t_R = 30.23$ minutes), and neohesperidin ($t_R = 31.42$ minutes) (►**Table 1**), we found that: (1) the polarity of dihydroflavonoid glycosides connected in the way of 1→2 between the monosaccharides substituted on the C_7 position of dihydroflavonoid glycosides is less than that connected in the way of 1→6. (2) In the negative ion mode, in the secondary mass spectra of naringin and neohesperidin, in addition to the conventional ions [M – H - Rha]$^-$, [M – H – Rha - Glu]$^-$, there was also special ion [M – H

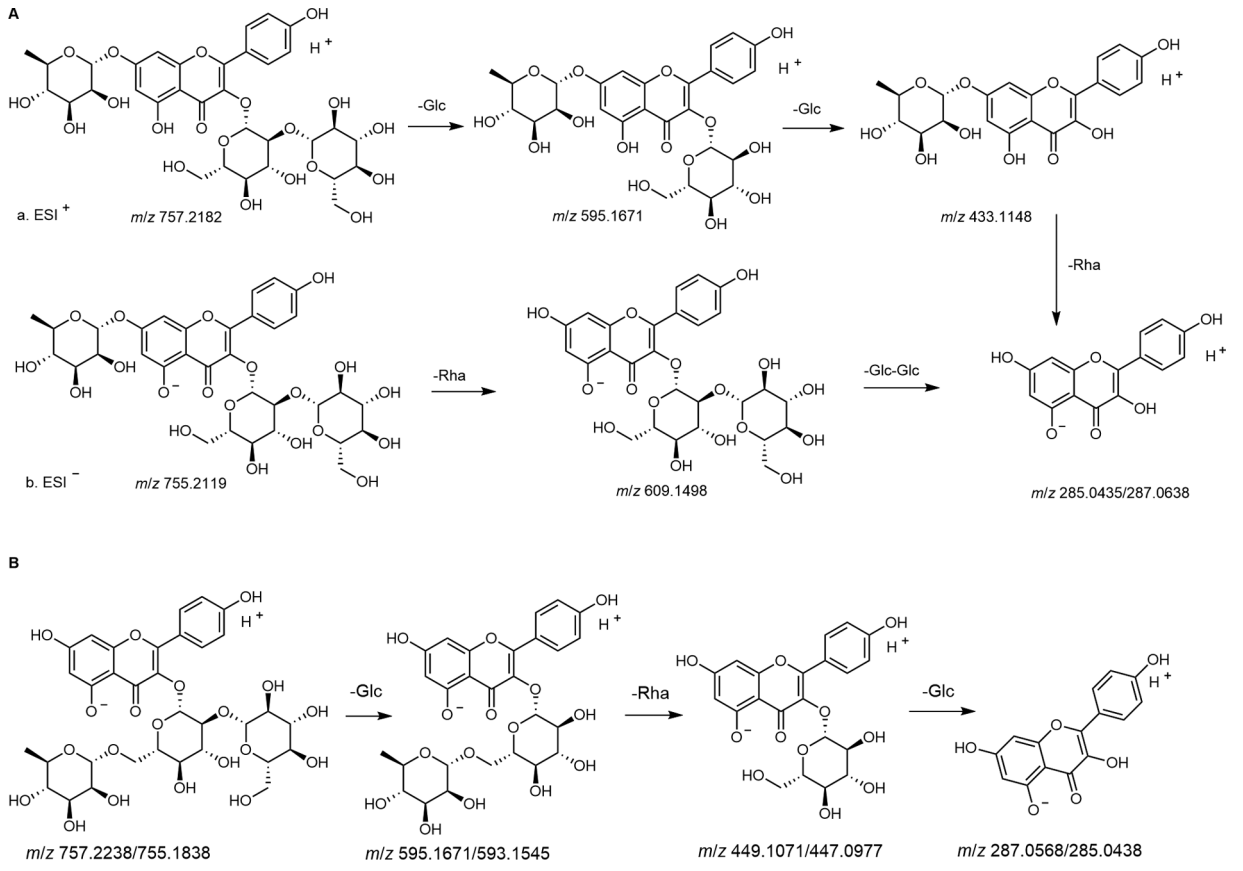


Fig. 2 Mass fragmentation pathways deduced of (A) kaempferol-3-*O*-β-*D*-sophoroside-7-*O*-α-*L*-rhamnoside and (B) kaempferol-3-*O*-(2″-*O*-β-*D*-glucopyranosyl)-β-*D*-rutinoside in positive and negative ion modes.

- 120]$^-$with strong abundance, and the fragment ion [M – H - 162]$^-$could be observed. Hesperidin did not appear as these special ions. It may be due to that the rhamnose linked to the hydroxyl in the C_2 position of the glucose at the end of aglycone has rearranged and cleaved and the ion [M – H - 120]$^-$resulted from the loss of a hexose residue in positions 0–3. The conclusion is to be proven by further experiments. (3) In the positive ion mode, the relative abundances of ions $[M + K - Rha]^+$ and $[M + K - O - Rha]^+$ were higher than those of 1→6 connected when the C_7-substituted rhamnose and glucose are 1→2 connected. In the negative ion mode, the rhamnose linked to the hydroxyl in the C_2 position of the glucose at the end of aglycone is more likely to rearrange, which may be related to the different charge distribution in the positive and negative ion modes. In the positive ion mode, the charge is mainly concentrated on the added sodium ions; and in the negative ion mode, the charge is mainly distributed in the whole sugar chain.[18]

By comparing genistin ($t_R = 24.04$ minutes), sophoricoside ($t_R = 28.72$ minutes), and sophorabioside ($t_R = 30.28$ minutes), we found that the polarity of isoflavone glycosyl substitution on the C_7 position is greater than that at position $C_{4'}$ (►**Fig. 4**). The glycosyl group of genistein substituted at position C_7 only lost 120 fragment ions at the negative electrode. Both in the positive and negative electrodes, the glycosyl groups of sophoricoside and sophorabioside substituted at position $C_{4'}$ detected the loss of 120 fragment signal, and the positive signal intensity is higher. However, whether the lost fragment signal (120 U) can be used as the diagnostic fragment of isoflavone glycosides needs further research.

Mass Spectrometric Cleavage of Saponins

Full scan and mass spectrometric cleavage analysis were performed for saponin standard under positive and negative ion modes. The analysis results of characteristic fragments are shown in ►**Table 2**, and the mass spectrometric cleavage pathway of asperosaponin VI is shown in ►**Fig. 5**. The summary rules are as follows: (1) In the negative ion mode, the saponin parent nucleus fragments are not obvious, mainly the deglycosylated fragments and $[M + Cl]^-$, $[M - H]^-$ excimer ion peaks; in the positive ion mode, a series of dehydrated fragments and $[M + Na]^+$ excimer ion peaks in the mother nucleus were mainly detected, while the response of deglycosylated fragments was weak. (2) The glycosyl group at C_3 position in asperosaponin VI, mogroside V, and ginsenoside Re is the last to fall off. Asperosaponin VI first lost the glycosylation at position C_{28}, ginsenoside Re first lost the glycosylation at position C_{20}, and mogroside V first lost the glycosylation at position C_{23}. This may be due to the ester bond and the ether bond on the straight chain is easier to break than the ether bond on the C_3 ring. (3) In the positive ion mode, the dehydration reaction of saponin parent nucleus fragments is only related to the number of hydroxyl groups carried on the mother nucleus, not related to the type of saponin, sugar chain

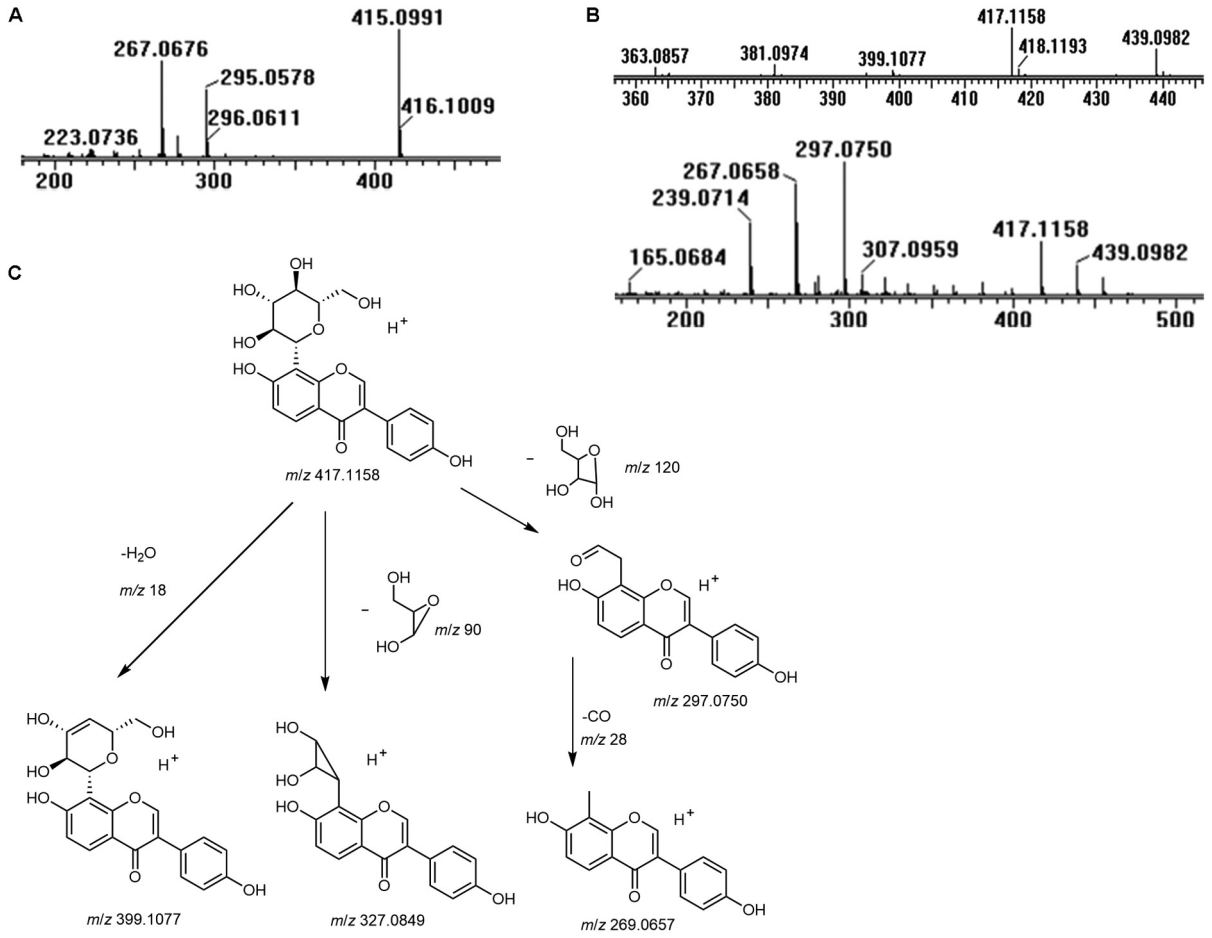
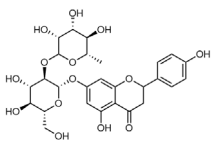
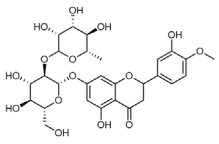
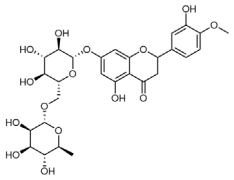


Fig. 3 Secondary mass spectra of puerarin in (A) negative ion mode, (B) positive ion mode, and (C) its possible cleavage pathways.

Table 1 MS^2 data of naringin, neohesperidin, hesperidin, and the relative abundance (%) of ions

Compound	MS^2 (ESI^-)	MS^2 (ESI^+)
 Naringin	579.1885 $[M-H]^-$ 459.1290 $[M-H-120]^-$(100) 433.1394 $[M-H-Rha]^-$ (17.9) 415.1147 $[M-H-164]^-$(24.6) 271.0694 $[M-H-Rha-Glu]^-$	619.1328 $[M+K]^+$ (23.2) 603.1697 $[M+Na]^+$ (100) 473.0746 $[M+K-Rha]^+$ (4.1) 457.1020 $[M+K-O-Rha]^+$ (3.3) 273.0744 $[M+H-Rha-Glu]^+$ (39.8)
 Neohesperidin	609.1806 $[M-H]^-$ 489.1284 $[M-H-120]^-$(100) 463.1176 $[M-H-Rha]^-$ (6.6) 449.0977 $[M-H-160]^-$(89.3) 447.0865 $[M-H-162]^-$(9.4) 301.0676 $[M-H-Rha-Glu]^-$	649.1487 $[M+K]^+$ (18.7) 633.1815 $[M+Na]^+$ (100) 503.0856 $[M+K-Rha]^+$ (6.6) 487.1154 $[M+K-O-Rha]^+$ (2.0) 303.0862 $[M+H-Rha-Glu]^+$ (34.7)
 Hesperidin	609.1806 $[M-H]^-$ 463.1176 $[M-H-Rha]^-$ 301.0711 $[M-H-Rha-Glu]^-$	649.1539 $[M+K]^+$ (21.6) 633.1815 $[M+Na]^+$ (100) 503.0810 $[M+K-Rha]^+$ (2.6) 487.1064 $[M+K-O-Rha]^+$ (1.4) 303.0862 $[M+H-Rha-Glu]^+$ (22.1)

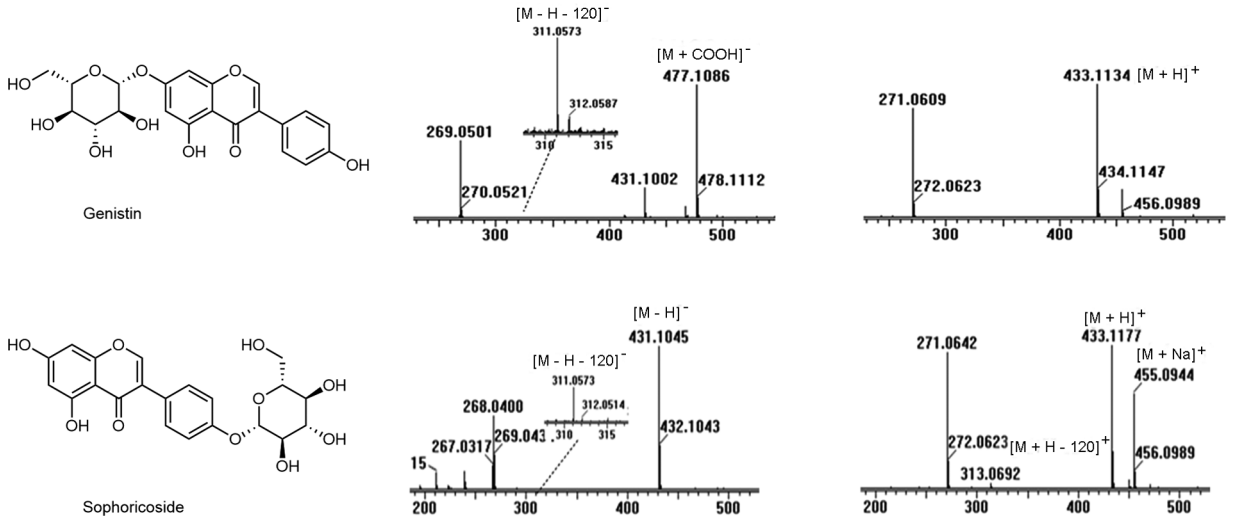


Fig. 4 Secondary mass spectra of genistein and sophoricoside.

and substitution position, and the number of dehydrated fragments is consistent with the number of hydroxyl groups on the mother nucleus, which is consistent with the literature report.[19] (4) Saponin aglycone: triterpenoid saponin oleanolic acid has a $[M - H]^-$ excimer ion peak in the negative ion mode, while steroidal saponin ruscogenin has a $[M + H]^+$ excimer ion peak in the positive ion mode.

Identification of Chemical Constituents of *Sophorae Fructus*

Identification of Flavonoids

When analyzed the test sample according to chromatographic conditions **(1)**, under the negative (►**Fig. 6A**) and positive (►**Fig. 6B**) ion scanning modes, the separation degree and ionization degree of each component of *Sophorae Fructus* meet the requirements. Identification of flavonoids: first, the types of flavonoids were speculated according to the characteristics of UV spectrum. Combined with the above MS rules and literature references, 84 compounds were finally identified in positive and negative ion modes, including 51 flavonoids, 12 isoflavones, 3 dihydroflavonoids, 8 organic acids and 10 amino acids, and sugars, as shown in ►**Table 3**.

Among the flavonoids, flavonols are the most abundant. Take compound **35** in ►**Table 3** as an example to illustrate the identification process. According to the UV maxima at 260 and 353 nm, the structural type was flavonol with substitution at the C_3 position. In the negative ion mode, the MS^2 of the aglycon-related ions were *m/z* 315.0502, 300.0277, 284.0320, 269.0450, 255.0295, 243.0678, 215.0729, 125.0258, which is identified as isorhamnetin by comparison with the literature.[20] Compound **35** displayed a $[M - H]^-$ ion at *m/z* 785.2151 and ions at *m/z* 639.1575 $[M - H - Rha]^-$, *m/z* 459.1283 $[M - H - Rha - Glu - H_2O]^-$, as well as *m/z* 315.0502 $[M - H - Rha - 2Glu]^-$. It was noted the presence of the *m/z* 459.1283 ion, which results from the loss of a hexose (180 u). This H_2O loss shows that the dihexosyl should have a 1→2 interglycosidic linkage because, as referred to above when the 1→2 bond versus the 1→6 bond was compared, the 1→6 bond is difficult to break. In the positive ion mode, compound **35** produced $[M + H]^+$ ion at *m/z* 787.2308 and in its MS^2 fragmentation of these ions (*m/z* 625.1698 $[M + H - Glu]^+ \rightarrow$ *m/z* 463.1243 $[M + H - 2Glu]^+ \rightarrow$ *m/z* 317.0664 $[M + H - 2Glu - Rha]^+$) can be observed. Combined with the above flavone cleavage rules, rhamnose was bound to a phenolic hydroxyl at position C_7, dihexosides with interglycosidic linkage 1→2 was substituted at position C_3. Therefore, compound **35** was proposed to be isorhamnetin-3-*O*-sophoroside-7-*O*-rhamnoside by comparison with the literature.[11]

Identification of Saponins

After analyzing the test sample according to chromatographic conditions (**2**), the total ion flow diagram is shown in ►**Fig. 6C** (negative ion scanning mode) and ►**Fig. 6D** (positive ion scanning mode). A total of 58 compounds were identified by using the above saponin cleavage rules in combination with relevant literature and reference standards, including 39 saponins, 10 phenolic acids, 3 fatty acids, 2 phenylpropanoids, 1 flavonol, and 3 others (►**Table 4**).

Triterpenoid saponins are mainly contained in *Sophorae Fructus*, and the structure is mostly oleanene type. The sugar chain structure in saponins is easy to be removed during cracking. If it is a branched glycosyl group and the two terminal glycosyl groups are different, the fragment peaks that lose the two terminal glycosyl groups will appear, so it is easy to distinguish between branched glycosyl groups and straight chain glycosyl groups.[21] Compound **104** in ►**Table 4** is taken as an example to derive the cracking rule of these compounds. Compound **104** was detected at *m/z* 941.5151 $[M - H]^-$ in the negative ion mode. The fragment ion at *m/z* 795.4539 indicated the loss of a deoxyhexose residue (146 u); peaks at *m/z* 615.3890 [aglycone + GluA – H_2O – $H]^-$, *m/z* 457.3663 [aglycone – $H]^-$, and *m/z* 483.1363 [Rha + Glu + GluA – $H]^-$ were presented in spectra. In the positive ion mode, the characteristic fragment ions at *m/z* 965.5106,

Table 2 ESI-MS cleavage characteristics of six saponins

Compounds	Formula	RT (min)	Mr	*m/z* (ESI^-)	*m/z* (ESI^+)
Asperosaponin VI	$C_{47}H_{76}O_{18}$	18.64	929.1000	963.4807 $[M+Cl]^-$ 927.5045 $[M-H]^-$ 603.3326 $[M-H-Glc-Glc]^-$ 471.3488 $[M-H-Glc-Glc-Ara]^-$	952.5076 $[M+Na]^+$ 455.3562 $[M+H-Glc-Glc-Ara-H_2O]^+$ 437.3417 $[M+H-Glc-Glc-Ara-2H_2O]^+$ 409.3485 $[M+H-2Glc-Ara-2H_2O-CO]^+$ 391.3341 $[M+H-2Glc-Ara-2H_2O-CO-H_2O]^+$
Mogroside V	$C_{60}H_{102}O_{29}$	15.78	1,287.4300	1,331.7566 $[M+HCOOH]^-$ 1,321.6241 $[M+Cl]^-$ 1,285.6539 $[M-H]^-$ 1,123.5906 $[M-H-Glc]^-$ 961.5374 $[M-H-2Glc]^-$ 799.4830 $[M-H-3Glc]^-$ 637.4334 $[M-H-4Glc]^-$ 485.1541 $[Glc+Glc+Glc-H]^-$	1,309.6475 $[M+Na]^+$ 1,125.6078 $[M+H-Glc]^+$ 963.5576 $[M+H-2Glc]^+$ 459.3859 $[M+H-5Glc-H_2O]^+$ 441.3770 $[M+H-5Glc-2H_2O]^+$ 423.3704 $[M+H-5Glc-3H_2O]^+$ 405.3514 $[M+H-5Glc-4H_2O]^+$
Ginsenoside Re	$C_{48}H_{82}O_{18}$	15.93	947.1500	981.5267 $[M+Cl]^-$ 945.5445 $[M-H]^-$ 783.4909 $[M-H-Glc]^-$ 765.4784 $[M-H-Glc-H_2O]^-$ 637.4334 $[M-H-Glc-Rha]^-$ 619.4224 $[M-H-Glc-Rha-H_2O]^-$ 475.3793 $[M-H-Glc-Rha-Glc]^-$	969.5428 $[M+Na]^+$ 789.4763 $[M+Na-Glc-H_2O]^+$ 459.3815 $[M+H-Glc-Ara-Glc-H_2O]^+$ 441.3770 $[M+H-Glc-Ara-Glc-2H_2O]^+$ 423.3746 $[M+H-Glc-Ara-Glc-3H_2O]^+$ 405.3556 $[M+H-Glc-Ara-Glc-4H_2O]^+$
Jujuboside A	$C_{58}H_{94}O_{26}$	19.29	1,207.3500	1,241.5747 $[M+Cl]^-$ 1,206.6047 $[M-H]^-$ 1,074.5632 $[M-H-Ara]^-$ 911.5027 $[M-H-Ara-Glc]^-$ 749.4507 $[M-H-Ara-2Glc]^-$ 603.3880 $[M-H-Ara-2Glc-Rha]^-$ 471.3444 $[M-H-Ara-2Glc-Rha-Ara]^-$	1,229.5935 $[M+Na]^+$ 455.3650 $[M+H-Ara-2Glc-Rha-Ara-H_2O]^+$ 437.3417 $[M+H-Ara-2Glc-Rha-Ara-2H_2O]^+$
Ruscogenin	$C_{27}H_{42}O_4$	30.84	430.6300	311.1677 $[M-H-118]^-$	431.3171 $[M+H]^+$ 413.3069 $[M+H-H_2O]^+$ 395.3064 $[M+H-2H_2O]^+$
Oleanolic acid	$C_{30}H_{48}O_3$	32.65	456.3600	455.7130 $[M-H]^-$	439.3678 $[M+H-H_2O]^+$ 411.3676 $[M+H-H_2O-CO]^+$ 393.3522 $[M+H-H_2O-CO-H_2O]^+$

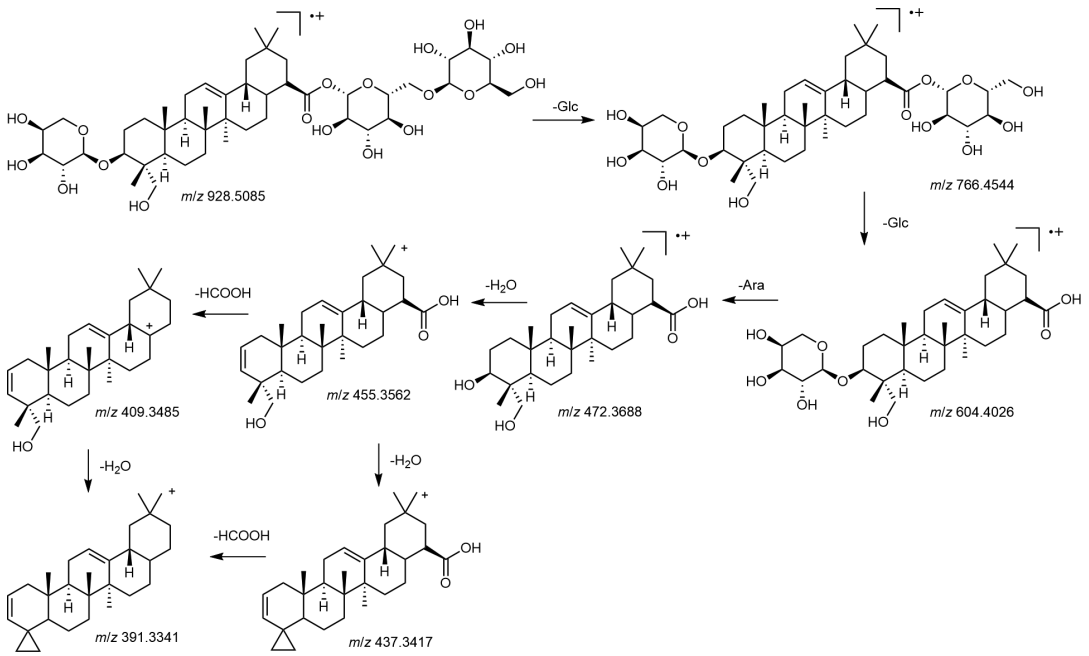


Fig. 5 Mass fragmentation pathways deduced of asperosaponin VI.

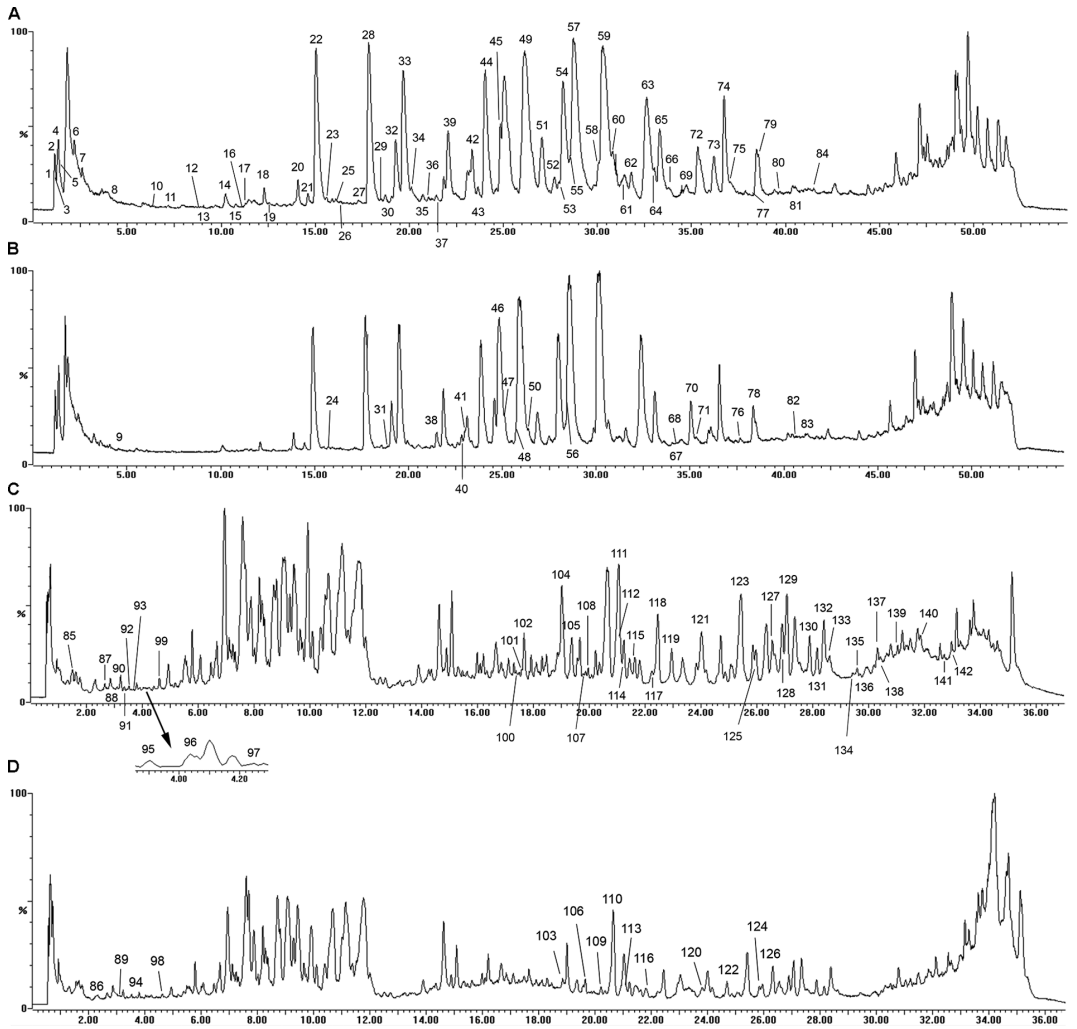


Fig. 6 Total ion flow diagram of *Sophorae Fructus* components in (A, C) negative ion mode and (B, D) positive ion mode.

Table 3 Analysis and identification of flavonoids from *Sophorae Fructus* by UPLC-QTOF mass spectrometry

No.	Component name	Formula	Calculated mass (*m*/*z*)	Measured mass (*m*/*z*)	Mass error (ppm)	RT (min)	Adducts	Fragment ions (*m*/*z*, ESI^-/ESI^+)
1	Arginine	$C_6H_{14}N_4O_2$	173.1044	173.1036	−4.62	1.02	$[M-H]^-$	173.1036
2	Alanine	$C_3H_7NO_2$	134.0459	134.0464	3.73	1.12	$[M+HCOO]^-$	134.0464
3	α–Sophora	$C_{12}H_{22}O_{11}$	387.1144	387.1134	−2.58	1.30	$[M+HCOO]^-$	387.1134
4	Aspartic acid	$C_4H_7NO_4$	132.0302	132.0296	−4.55	1.33	$[M-H]^-$	132.0296
5	Arabinose	$C_5H_{10}O_5$	195.0510	195.0514	2.05	1.37	$[M+HCOO]^-$	195.0514
6	Malic acid	$C_4H_6O_5$	133.0142	133.0137	−3.76	2.17	$[M-H]^-$	133.0137
7	Citric acid[a]	$C_6H_8O_7$	191.0197	191.0191	−3.14	3.42	$[M-H]^-$	191.0191
8	Phenylalanine	$C_9H_{11}NO_2$	164.0717	164.0710	−4.27	4.06	$[M-H]^-$	164.0170
9	Gallic acid[a]	$C_7H_6O_5$	171.0288	171.0287	−0.58	4.11	$[M+H]^-$	171.0287, 125.0300
10	γ-Glutamyltyrosine	$C_{14}H_{18}N_2O_6$	309.1092	309.1088	−1.29	6.32	$[M-H]^-$	309.1088
11	Tryptophan	$C_{11}H_{12}N_2O_2$	203.0826	203.0822	−1.97	7.16	$[M-H]^-$	203.0822
12	Kaempferol-3-*O*-sophoroside-7-*O*-glucoside[b]	$C_{33}H_{40}O_{21}$	771.1989	771.1990	0.13	8.83	$[M-H]^-$	771.1990, 593.1495, 447.089, 285.0403
13	Protocatechuic acid[a]	$C_7H_6O_4$	153.0193	153.0191	−1.31	8.97	$[M-H]^-$	153.0191
14	Methyl gallate	$C_8H_8O_5$	183.0299	183.0296	−1.64	10.13	$[M-H]^-$	183.0296, 169.0179, 125.0219
15	4-(β-*D*-Glucopyranosyloxy)-3-hydroxyphenyl caffeate[b]	$C_{21}H_{22}O_{11}$	449.1089	449.1083	−1.34	10.76	$[M-H]^-$	449.1083, 287.0553
16	Protocatechualdehyde[a]	$C_7H_6O_3$	183.0299	183.0301	1.09	10.86	$[M+HCOO]^-$	183.0301
17	1,6-di-*O*-galloyl-β-*D*-glucose	$C_{20}H_{20}O_{14}$	483.0780	483.0791	2.28	11.49	$[M-H]^-$	483.0791, 331.0651, 313.0550, 169.0130, 125.0235
18	Kaempferol-3-*O*-gentionbioside-7-*O*-glucoside[b]	$C_{33}H_{40}O_{21}$	771.1989	771.1990	0.13	12.29	$[M-H]^-$	771.1990, 609.1448, 447.0890, 285.0369
19	Kaempferol-3-*O*-(2″-*O*-β-*D*- glucopyranosyl)-β-*D*-rutinoside-7-*O*-glucoside[b]	$C_{39}H_{50}O_{25}$	917.2568	917.2574	0.65	12.47	$[M-H]^-$	917.2574, 755.2006, 577.1556, 446.0854, 284.0320
20	Lepidoside	$C_{26}H_{28}O_{14}$	563.1406	563.1374	−5.68	14.09	$[M-H]^-$	563.1374, 419.1205, 285.0403
21	Genistein-7-*O*-β-*D*-glucoside 4′-*O*-sophoroside	$C_{33}H_{40}O_{20}$	757.2186	757.2182	−0.53	14.45	$[M+H]^+$	757.2182, 595.1622, 433.1140, 271.0597
22	Genistein-7,4′-di-*O*-β-*D*-glucoside	$C_{27}H_{30}O_{15}$	595.1657	595.1664	1.18	14.91	$[M+H]^+$	595.1664, 433.1140, 271.0597
23	Quercetin-3-*O*-(3‴-*O*-β-*D*- glucopyranosyl)-β-*D*-neohesperidoside[b]	$C_{33}H_{40}O_{21}$	771.1989	771.1990	0.13	15.59	$[M-H]^-$	771.1990, 609.1448, 462.0806, 431.0984, 299.0180
24	Apigenin-7,4′-diglucoside	$C_{27}H_{30}O_{15}$	595.1657	595.1671	2.35	15.88	$[M+H]^+$	595.1671, 433.1097, 271.0630

(*Continued*)

Table 3 (*Continued*)

No.	Component name	Formula	Calculated mass (*m*/*z*)	Measured mass (*m*/*z*)	Mass error (ppm)	RT (min)	Adducts	Fragment ions (*m*/*z*, ESI^-/ESI^+)
25	Neoeriocitrin[b]	$C_{27}H_{32}O_{15}$	595.1668	595.1653	−2.52	16.08	$[M-H]^-$	595.1653, 431.0984, 287.0553
26	Diosmetin-7-*O*-sophoroside[b]	$C_{28}H_{32}O_{16}$	669.1672	669.1663	−1.35	16.33	$[M+HCOO]^-$	623.1638, 461.1086, 445.0786, 268.0387
27	Quercetin-3-*O*-(2″-*O*-β-*D*- glucopyranosyl)-β-*D*-rutinoside-7-*O*-rhamnoside[b]	$C_{39}H_{50}O_{25}$	917.2568	917.2533	−3.82	17.30	$[M-H]^-$	917.2533, 771.1990, 609.1397, 461.1306, 301.0356
28	Genistein-7-*O*-β-*D*-glucoside-4′-*O*-neohesperidoside	$C_{33}H_{40}O_{19}$	739.2091	739.2086	−0.68	17.88	$[M-H]^-$	739.2086, 577.1654, 431.0984, 415.1070, 268.0454
29	Kaempferol-3,7-diglucoside[a]	$C_{27}H_{30}O_{16}$	609.1461	609.1448	−2.13	18.15	$[M-H]^-$	609.1448, 447.0977, 285.0403
30	Kaempferol-3-*O*-(4‴-*O*-β-*D*- glucopyranosyl)-β-*D*-rutinoside[b]	$C_{33}H_{40}O_{20}$	755.2040	755.2062	2.91	18.74	$[M-H]^-$	755.2062, 593.1545, 446.0854, 285.0403
31	Apigenin-7-*O*-(2′′-*O*-sophorosyl)-β-*D*-rutinoside[b]	$C_{39}H_{50}O_{24}$	903.2765	903.2797	3.54	18.90	$[M+H]^+$	903.2797, 741.2186, 579.1738, 433.1140, 271.0597
32	Ternatumoside VIII[b]	$C_{39}H_{50}O_{24}$	901.2619	901.2601	−2.00	19.29	$[M-H]^-$	901.2601, 755.2062, 593.1495, 430.0917, 284.0355
33	Kaempferol-3-*O*-sophoroside-7-*O*-rhamnoside[a]	$C_{33}H_{40}O_{20}$	757.2186	757.2196	1.32	19.64	$[M+H]^+$	757.2196, 595.1671, 433.1140, 287.0568
34	Isorhamnetin-3-*O*-(4′′-*O*-rutinosyl)-β-*D*-rutinoside[b]	$C_{40}H_{52}O_{25}$	933.2870	933.2865	−0.54	20.12	$[M+H]^+$	933.2865, 787.2250, 625.1698, 463.1243, 317.0664
35	Isorhamnetin-3-*O*-sophoroside-7-*O*-rhamnoside[b]	$C_{34}H_{42}O_{21}$	785.2146	785.2151	0.64	20.72	$[M-H]^-$	785.2151, 639.1575, 459.1283, 314.0445
36	Kaempferol-3-*O*-glucoside-7-*O*-rutinoside[b]	$C_{33}H_{40}O_{20}$	755.2040	755.2062	2.91	21.03	$[M-H]^-$	755.2062, 635.0911, 608.1398, 447.0933, 285.0403
37	Dihydrokaempferol 3-*O*-glucoside	$C_{21}H_{22}O_{11}$	449.1089	449.1083	−1.34	21.45	$[M-H]^-$	449.1083, 287.0578
38	Genistein-7-*O*-malonylglucoside-4′-*O*-glucoside	$C_{30}H_{32}O_{18}$	681.1661	681.1641	−2.94	21.82	$[M+H]^+$	681.1641, 433.1140, 271.0597
39	Quercetin-3-*O*-β-*D*-glucopyranosyl(1→2)-[α-*L*-rhamnopyranosyl(1→6)]-β-*D*-glucopyranoside	$C_{33}H_{40}O_{21}$	773.2135	773.2154	2.46	22.12	$[M+H]^+$	773.2154, 611.1609, 465.1011, 303.0498
40	Quercetin 3-*O*-gentiobioside	$C_{27}H_{30}O_{17}$	627.1556	627.1563	1.12	23.08	$[M+H]^+$	627.1563, 465.1033, 303.0494
41	Quercetin-3-*O*-(6‴-*O*-adipoyl)-β-*D*-rutinoside[b]	$C_{33}H_{38}O_{19}$	739.2080	739.2090	1.35	23.21	$[M+H]^+$	739.2090, 465.1033, 303.0494
42	Quercetin-3-*O*-(6″-O-(3‴-*O*-arabinose)-α-*L*-rhamnosyl)-β-*D*-neohesperidoside[b]	$C_{38}H_{48}O_{24}$	887.2463	887.2480	1.92	23.46	$[M-H]^-$	887.2480, 741.1857, 609.1448, 301.0356

Table 3 (*Continued*)

No.	Component name	Formula	Calculated mass (*m*/*z*)	Measured mass (*m*/*z*)	Mass error (ppm)	RT (min)	Adducts	Fragment ions (*m*/*z*, ESI^-/ESI^+)
43	Compactin [b]	$C_{27}H_{30}O_{16}$	611.1607	611.1609	0.33	23.67	$[M+H]^+$	611.1609, 449.1079, 287.0559
44	Genistin[a]	$C_{21}H_{20}O_{10}$	431.0984	431.0984	0	24.03	$[M-H]^-$	431.0984, 269.0551
45	Apigenin-7-*O*-(3‴-*O*-acetyl)-β-*D*-rutinoside	$C_{29}H_{32}O_{15}$	619.1668	619.1690	3.55	24.85	$[M-H]^-$	619.1690, 431.0984, 268.0387
46	Kaempferol-3-*O*-(2″-*O*-β-*D*- glucopyranosyl)-β-*D*-rutinoside[a]	$C_{33}H_{40}O_{20}$	757.2186	757.2182	−0.53	25.04	$[M+H]^+$	757.2182, 595.1671, 449.1071, 287.0672
47	Rhamnocitrin-3-*O*-β-*D*-glucopyranosyl(1→2)-*D*-Apio-α-*D*-furanoside -4′-*O*-glucoside	$C_{33}H_{40}O_{20}$	757.2186	757.2182	−0.53	25.31	$[M+H]^+$	757.2182, 595.1622, 463.1243, 301.0723
48	Isorhamnetin-3-*O*-sophoroside	$C_{28}H_{32}O_{17}$	641.1712	641.1703	−1.40	25.94	$[M+H]^+$	641.1703, 479.1197, 317.0664
49	Kaempferol-3-*O*-sophoroside[a]	$C_{27}H_{30}O_{16}$	609.1461	609.1448	−2.13	26.12	$[M-H]^-$	609.1448, 429.0819, 285.0507
50	Multiflorin B	$C_{27}H_{30}O_{15}$	593.1512	593.1495	−2.87	26.58	$[M-H]^-$	593.1495, 431.0984, 284.0320
51	Naringin[a]	$C_{27}H_{32}O_{14}$	579.1719	579.1736	2.94	27.05	$[M-H]^-$	579.1736, 433.1239, 271.0601
52	Kaempferol-3-*O*-α-*L*-rhamnopyranosyl (1→4)]-β-*D*-glucopyranoside	$C_{27}H_{30}O_{15}$	593.1512	593.1495	−2.87	27.73	$[M-H]^-$	593.1495, 447.0933, 429.0776, 285.0403
53	6″-β-*D*-Xylosegenistin	$C_{26}H_{28}O_{14}$	587.1377	587.1366	−1.87	28.05	$[M+Na]^+$	587.1366, 433.1181, 271.0609
54	Rutin[a]	$C_{27}H_{30}O_{16}$	633.1432	633.1448	2.53	28.16	$[M+Na]^+$	633.1448, 465.1033, 303.0565
55	Isoquercitrin[a]	$C_{21}H_{20}O_{12}$	463.0882	463.0902	4.32	28.59	$[M-H]^-$	463.0902, 300.0323
56	Kaempferide-3-*O*-glucoside	$C_{22}H_{22}O_{11}$	463.1235	463.1243	1.73	28.66	$[M+H]^+$	463.1243, 301.0723
57	Sophoricoside[a]	$C_{21}H_{20}O_{10}$	433.1129	433.1140	2.54	28.74	$[M+H]^+$	433.1140, 271.0630
58	Helieianeoside A[b]	$C_{32}H_{38}O_{19}$	725.1935	725.1947	1.66	30.06	$[M-H]^-$	725.1947, 593.1495, 431.0984, 285.0403
59	Sophorobioside[a]	$C_{27}H_{30}O_{14}$	601.1533	601.1518	−2.50	30.28	$[M+Na]^+$	601.1518, 579.1738, 433.1140, 271.0597
60	Apigenin-7-*O*-neohesperidoside	$C_{27}H_{30}O_{14}$	577.1563	577.1556	−1.21	30.83	$[M-H]^-$	577.1654, 431.0984, 413.0867, 269.0450
61	Apigenin-7-*O*-rutinoside	$C_{27}H_{30}O_{14}$	577.1563	577.1556	−1.21	31.43	$[M-H]^-$	577.1556, 431.0984, 269.0450
62	Apigenin-4′-*O*-rutinoside[b]	$C_{27}H_{30}O_{14}$	577.1563	577.1556	−1.21	31.81	$[M-H]^-$	577.1556, 431.0984, 268.0387
63	Kaempferol-3-*O*-rutinoside[a]	$C_{27}H_{30}O_{15}$	595.1657	595.1671	2.35	32.62	$[M+H]^+$	595.1721, 449.1071, 287.0603
64	Apigenin-7-*O*-gentiobioside	$C_{27}H_{30}O_{15}$	593.1512	593.1495	−2.87	33.02	$[M-H]^-$	593.1495, 447.0933, 269.0450
65	Isorhamnetin-3-*O*-β-*D*-rutinoside[a]	$C_{28}H_{32}O_{16}$	625.1763	625.1749	−2.24	33.31	$[M+H]^+$	625.1749, 479.1197, 317.0664
66	Diosmetin-7-*O*-glucopyranosyl-(6→1)-*O*-arabinopyranoside	$C_{27}H_{30}O_{15}$	593.1512	593.1495	−2.87	33.80	$[M-H]^-$	593.1495, 461.1102, 299.0576
67	Apigenin-7-*O*-[6-*O*-acetyl-2-*O*-(6-*O*-acetyl-β-*D*-glucopyranosyl)]-β-*D*-glucoside[b]	$C_{31}H_{34}O_{17}$	677.1723	677.1702	−3.11	34.45	$[M-H]^-$	677.1702, 473.1059, 268.0372

(*Continued*)

Table 3 (*Continued*)

No.	Component name	Formula	Calculated mass (*m*/*z*)	Measured mass (*m*/*z*)	Mass error (ppm)	RT (min)	Adducts	Fragment ions (*m*/*z*, ESI^-/ESI^+)
68	Apigenin-7-*O*-(4,6-di-*O*-acetyl)-β-*D*-glucoside-4′-*O*-β-*D*-glucoside[b]	$C_{31}H_{34}O_{17}$	723.1778	723.1840	8.58	34.62	$[M+HCOO]^-$	723.1840, 677.1646, 431.0989, 268.0372
69	6″-*O*-Acetylnaringin[b]	$C_{29}H_{34}O_{15}$	621.1825	621.1818	−1.13	34.72	$[M-H]^-$	621.1826, 473.1068, 271.0601
70	Genistein-4′-*O*-malonylglucoside	$C_{24}H_{22}O_{13}$	519.1133	519.1116	−3.28	35.04	$[M+H]^+$	519.1116, 271.0597
71	Kakkanin	$C_{27}H_{30}O_{14}$	579.1708	579.1689	−3.28	35.31	$[M+H]^+$	579.1689, 447.1278, 285.0774
72	Euryanoside[b]	$C_{29}H_{32}O_{15}$	619.1668	619.1639	−4.68	35.48	$[M-H]^-$	619.1639, 473.1068, 269.0450
73	Acacetin-5-*O*-α-*L*-mannopyranosyl(1→2)-α-*L*-rhamnopyranoside[b]	$C_{28}H_{32}O_{13}$	575.1770	575.1411	−6.24	36.11	$[M-H]^-$	575.1411, 431.0984, 283.0600
74	Apigenin-5-*O*-acetyl-7-*O*-neohesperidoside[b]	$C_{29}H_{32}O_{15}$	619.1668	619.1639	−4.68	36.74	$[M-H]^-$	619.1639, 473.1068, 311.0552, 269.0450
75	Apigenin-7-*O*-rhamnoside	$C_{21}H_{20}O_9$	415.1035	415.1028	−1.69	37.22	$[M-H]^-$	415.1028, 268.0387
76	Acacetin-7-(6-malonylglucoside)	$C_{25}H_{24}O_{13}$	555.1115	555.1107	−1.44	37.72	$[M+Na]^+$	555.1107, 285.0729
77	Apigenin-7,4′-di-*O*-rhamnoside[b]	$C_{27}H_{30}O_{13}$	561.1614	561.1620	1.07	38.12	$[M-H]^-$	561.1620, 415.1063, 268.0400
78	Genistein[a]	$C_{15}H_{10}O_5$	271.0601	271.0597	−1.48	38.46	$[M+H]^+$	271.0601
79	Apigenin-7-*O*-α-*L*-mannopyranosyl (1→3)-α-*L*-rhamnopyranoside[b]	$C_{27}H_{30}O_{13}$	561.1614	561.1620	1.07	38.58	$[M-H]^-$	561.1692, 415.1063, 397.0893, 269.0484
80	Diosmetin	$C_{16}H_{12}O_6$	299.0561	299.0570	3.01	39.69	$[M-H]^-$	299.0570, 284.0320, 255.0295
81	Acacetin	$C_{16}H_{12}O_5$	283.0612	283.0600	−4.24	40.36	$[M-H]^-$	283.0600, 255.0295, 242.9433
82	Kaempferol[a]	$C_{15}H_{10}O_6$	287.0550	287.0559	3.14	40.71	$[M+H]^+$	287.0559
83	Baicalein[a]	$C_{15}H_{10}O_5$	269.0455	269.0450	−1.86	41.11	$[M-H]^-$	269.0450
84	Apigenin[a]	$C_{15}H_{10}O_5$	269.0455	269.0450	−1.86	41.41	$[M-H]^-$	269.0450

[a]Compared with reference substance.
[b]Be first found in *Sophora*.

Table 4 Analysis and identification of saponins from *Sophorae Fructus* by UPLC-Q-TOF mass spectrometry

No.	Component name	Formula	Calculated mass (*m*/*z*)	Measured mass (*m*/*z*)	Mass error (ppm)	RT (min)	Adducts	Fragment ions (*m*/*z*, ESI^-/ESI^+)
85	Gallic acid-*O*-glucoside (isomer)	$C_{13}H_6O_{10}$	331.0671	331.0687	4.83	1.51	$[M-H]^-$	331.0761, 169.0122, 125.0242
86	Gallic acid-*O*-diglucoside	$C_{19}H_{26}O_{15}$	517.1169	517.1176	1.35	2.33	$[M+Na]^+$	517.1176, 355.0602, 171.0284, 125.0230
87	*p*-Hydroxybenzoic acid glucoside	$C_{13}H_{16}O_8$	299.0772	299.0756	−5.35	2.67	$[M-H]^-$	299.0756, 137.0223
88	M-digallic acid	$C_{12}H_{16}O_6$	321.0252	321.0264	3.74	2.86	$[M-H]^-$	321.0264, 169.0122, 125.0219
89	1-p-anisate-glucopyranuronic acid[a]	$C_{14}H_{16}O_9$	328.0794	328.1377	8.23	3.11	$[M+e]^+$	328.1377, 279.0945, 141.9588
90	Gallic acid-*O*-glucoside (isomer)	$C_{13}H_6O_{10}$	331.0671	331.0687	4.83	3.23	$[M-H]^-$	331.0761, 169.0122, 125.0242
91	Digallic acid[a]	$C_{14}H_{12}O_9$	323.0409	323.0394	−4.64	3.37	$[M-H]^-$	323.0394, 169.0122, 125.0242
92	3,5-Dihydroxy-4-[(3,4,5-trihydroxybenzoyl) oxy]benzoic acid	$C_{14}H_{10}O_9$	321.0252	321.0264	3.74	3.52	$[M-H]^-$	321.0264, 178.9772, 144.0444
93	Amygdalinic acid[a]	$C_{20}H_{28}O_{13}$	475.1457	475.1428	−6.11	3.67	$[M-H]^-$	475.1428, 329.0865, 313.0917, 268.0372, 151.0381
94	2,3-Dihydroxy-3-(3,4,5-trimethoxyphenyl) propyl-β-*D*-glucopyranoside[a]	$C_{18}H_{28}O_{11}$	419.1559	419.1549	−2.39	3.77	$[M+Na]^+$	443.1170, 421.1345, 289.0913, 158.9624, 127.0396
95	Everlastoside H[a]	$C_{21}H_{28}O_{14}$	503.1406	503.1408	0.40	3.92	$[M-H]^-$	503.1408, 345.0841, 323.1351, 178.9772, 145.9299
96	2-Hydroxy-1-(hydroxymethyl)-2-(3,4,5-trimethoxyphenyl)ethyl-β-*D*-glucopyranoside[a]	$C_{18}H_{28}O_{11}$	465.1614	465.1653	8.39	4.03	$[M+HCOO]^-$	465.1225, 419.1172, 315.0724, 235.9240, 178.9772
97	Pothobanoside C[a]	$C_{26}H_{38}O_{16}$	651.2142	651.2100	−6.45	4.24	$[M+HCOO]^-$	651.2100, 443.1891, 329.0865, 153.0169, 137.0223
98	Tuberosinine D	$C_{19}H_{28}O_{12}$	471.1478	471.1494	3.40	4.43	$[M+Na]^+$	471.1494, 343.0466, 315.0544, 153.0164
99	Di-*O*-galloyl glucose	$C_{20}H_{19}O_{14}$	483.0780	483.0778	−0.41	4.61	$[M-H]^-$	483.0778, 331.0612, 271.0450, 243.0513, 169.0148
100	Isorhamnetin	$C_{16}H_{12}O_7$	315.0510	315.0506	−1.27	17.49	$[M-H]^-$	315.0506, 300.0286, 285.0383, 269.0436
101	28–(Hydroxymethyl)olean-12-en-3β-yl-*O*-β-*D*-glucopyranosyl-(1→6)-*O*-β-*D*-galactopyranosyl(1→4)-*O*-[α-*L*-rhamnopyranosyl-(1→2)]-β-*D*-galactopyranosiduronic acid[a]	$C_{55}H_{90}O_{22}$	1,101.5851	1,101.5781	−6.36	17.56	$[M-H]^-$	1101.5781, 957.5237, 795.4545, 633.3990, 458.3735, 439.3592, 421.3480
102	3-[(*O*-hexopyranosyl-(1→2)-*O*-hexopyranosyl-(1→3)-*O*-hexopyranuronosyl-(1→4)-6-deoxyhexopyranosyl)oxy]-, 21-acetate 22-(2-methylpropanoate), (3β)-Olean-12-ene-15,16,21,22,28-pentol[a]	$C_{60}H_{96}O_{28}$	1,287.5986	1,287.5944	−3.26	17.66	$[M+Na]^+$	1287.5944, 1265.6073, 1103.5602, 765.4705, 619.1520, 499.1179, 471.0606, 458.3761, 445.1229, 417.1301, 401.2097
103	Polybosaponin A	$C_{48}H_{76}O_{19}$	979.4878	979.4835	−4.39	18.86	$[M+Na]^+$	979.4835, 811.4466, 649.3914, 473.3636, 455.3542, 437.3432

(*Continued*)

Table 4 (*Continued*)

No.	Component name	Formula	Calculated mass (*m*/*z*)	Measured mass (*m*/*z*)	Mass error (ppm)	RT (min)	Adducts	Fragment ions (*m*/*z*, ESI^-/ESI^+)
104	Azukisaponin V	$C_{48}H_{78}O_{18}$	941.5115	941.5151	3.83	19.01	$[M-H]^-$	941.5151, 795.4539, 615.3890, 457.3663, 441.3744, 423.3672, 405.3513
105	Azukisaponin II	$C_{42}H_{68}O_{14}$	795.4536	795.4539	0.38	19.33	$[M-H]^-$	795.4539, 633.1541, 617.4020, 441.3744, 423.3627, 405.3471, 395.0494
106	Dehydrosoyasaponin I	$C_{48}H_{76}O_{18}$	941.5104	941.5088	−1.70	19.67	$[M+H]^+$	941.5088, 795.4539, 633.3990, 457.3667, 439.3549, 421.3438, 403.3315, 395.0746
107	Abrisaponin I	$C_{48}H_{74}O_{20}$	971.4846	971.4894	4.94	19.87	$[M+H]^+$	971.4894, 825.4259, 649.4018, 469.3304, 451.3204, 433.1133, 405.3430
108	Umbellatoside A[a]	$C_{48}H_{76}O_{19}$	979.4878	979.4835	−4.39	19.96	$[M+Na]^+$	979.4835, 819.4411, 649.3965, 473.3636, 455.3542, 437.3389, 421.3444, 391.1188
109	Astragaloside VIII	$C_{47}H_{76}O_{17}$	935.4980	935.4990	1.07	20.23	$[M+Na]^+$	935.4990, 789.4240, 633.1387, 441.3744, 423.3627, 405.3513
110	Putranoside C	$C_{47}H_{72}O_{19}$	963.4929	963.4963	3.53	20.65	$[M+Na]^+$	963.4963, 779.4532, 633.4012, 457.3692, 439.3565, 421.3480, 391.1228
111	Yunganoside D_1[a]	$C_{48}H_{74}O_{19}$	977.4722	977.4739	1.74	21.03	$[M+Na]^+$	977.4739, 831.4089, 633.4012, 453.3394, 435.3260
112	Soyasaponin Bg	$C_{47}H_{74}O_{17}$	933.4824	933..4748	−8.15	21.08	$[M+Na]^+$	933.4788, 765.4444, 633.4012, 455.3455, 439.3565
113	(3β,4β)-23-hydroxy-22-oxoolean-12-en-3-yl-*O*-6-deoxy-α-*L*-mannopyranosyl-(1→2)-*O*-β-*D*-glucopyranosyl-(1→4)-β-*D*-glucopyranosiduronic acid	$C_{48}H_{76}O_{18}$	963.4929	963.4900	−3.01	21.14	$[M+Na]^+$	963.4900, 817.4276, 633.4012, 439.3565, 421.3486
114	Soybean phenol A	$C_{42}H_{66}O_{14}$	793.4380	793.4404	3.03	21.25	$[M-H]^-$	793.4404, 631.3856, 455.0141
115	Glycyrrhetinic acid-3-*O*-Glucopyranosiduronic acid-29-*O*-glucoside	$C_{42}H_{64}O_{15}$	831.4143	831.4148	0.61	21.62	$[M+Na]^+$	831.4148, 647.4064, 471.3449, 453.3351, 435.3260, 424.1924, 407.3350
116	Uralsaponin X	$C_{50}H_{74}O_{22}$	1,049.4569	1,049.4613	4.19	21.81	$[M+Na]^+$	1049.4613, 741.3734, 704.2580, 565.3466, 525.1416, 507.1360, 481.0825, 457.2379, 439.3565, 421.3444, 403.3354
117	(3β,4β,22β)-22,23-Dihydroxy-11-oxoolean-12-en-3-yl-*O*-6-deoxy-α-L-mannopyranosyl-(1→2)-*O*-β-*D*-galactopyranosyl-(1→2)-β-*D*-glucopyranosiduronic acid	$C_{48}H_{76}O_{19}$	979.4878	979.4835	−4.39	22.23	$[M+Na]^+$	979.4835, 833.4841, 671.3387, 455.3542, 445.1306, 423.3585, 409.1622
118	Wistariasaponin D	$C_{47}H_{74}O_{17}$	909.4853	909.4829	−2.64	22.46	$[M-H]^-$	909.4829, 763.4296, 631.3805, 455.3504
119	Glycyrflavoside B	$C_{47}H_{72}O_{18}$	947.4616	947.4581	−3.70	22.95	$[M+Na]^+$	947.4581, 801.3914, 669.3496, 471.3449, 453.3351, 435.3260
120	(3β,4α,22β)-22,23-dihydroxyolean-12-en-3-yl-*O*-β-*D*-arabinofuranosyl-(1→2)-*O*-6-deoxy-	$C_{47}H_{76}O_{17}$	935.4980	935.4927	−5.67	23.84	$[M+Na]^+$	935.4927, 781.4766, 635.3337, 441.3658, 423.3585, 405.3471

Table 4 (*Continued*)

No.	Component name	Formula	Calculated mass (*m*/*z*)	Measured mass (*m*/*z*)	Mass error (ppm)	RT (min)	Adducts	Fragment ions (*m*/*z*, ESI^-/ESI^+)
	β-*D*-galactopyranosyl-(1→2)-β-*D*-glucopyranosiduronic acid							
121	Soyasaponin I	$C_{48}H_{78}O_{18}$	943.5261	943.5261	0	23.99	$[M+H]^+$	943.5261, 797.4693, 635.4162, 441.3781, 423.3627
122	Soyasaponin III	$C_{42}H_{68}O_{14}$	795.4536	795.4539	0.38	24.69	$[M-H]^-$	795.4539, 633.3983, 455.3504
123	Kaikasaponin III	$C_{48}H_{78}O_{17}$	949.5137	949.5123	−1.48	25.42	$[M+Na]^+$	949.5123, 803.4489, 601.4117, 425.3773, 407.3681
124	β-*D*-glucopyranosiduronic acid deriv-oleanane	$C_{48}H_{78}O_{17}$	949.5137	949.5123	-1.48	25.87	$[M+Na]^+$	949.5123, 803.4431, 657.3563, 633.4012, 457.3648, 437.1935, 425.3773, 407.3681
125	Pisumsaponin II	$C_{48}H_{76}O_{18}$	939.4959	939.4937	−2.34	25.97	$[M-H]^-$	939.4937, 793.4346, 631.3856, 455.3504
126	Kaikasaponin I	$C_{42}H_{68}O_{13}$	803.4558	803.4547	−1.37	26.33	$[M+Na]^+$	803.4547, 641.2833, 425.3773, 407.3681
127	Azukisaponin I	$C_{42}H_{68}O_{13}$	779.4587	779.4567	−2.57	26.65	$[M-H]^-$	779.4567, 617.4016, 441.3724
128	Kakkasaponin I	$C_{47}H_{76}O_{16}$	895.5061	895.5054	−0.78	26.92	$[M-H]^-$	895.5054, 749.4451, 599.3972, 441.3724
129	Phaseoside IV	$C_{48}H_{76}O_{17}$	947.4980	947.4958	−2.32	27.09	$[M+Na]^+$	947.4958, 803.4025, 641.4025, 617.4020, 441.3744, 423.3627, 405.3513
130	Kakkasaponin II	$C_{42}H_{66}O_{13}$	777.4431	777.4463	4.12	27.91	$[M-H]^-$	777.4463, 615.3941, 437.3435
131	Zygophyloside M[a]	$C_{40}H_{62}O_{13}$	795.4536	795.4539	0.38	28.16	$[M+HCOO]^-$	795.4539, 749.4507, 633.3983, 441.3724
132	Kakkasaponin III	$C_{47}H_{74}O_{16}$	917.4875	917.4889	1.53	28.39	$[M+Na]^+$	917.4889, 749.4507, 617.4071, 441.3701, 423.3627, 405.3471
133	Paradoxoside E[a]	$C_{37}H_{56}O_{11}$	699.3720	699.3745	3.58	28.64	$[M+Na]^+$	699.3582, 471, 453.1652, 437.2020, 407.3681
134	Presenegenin	$C_{30}H_{46}O_7$	563.3226	563.3226	0	29.44	$[M+HCOO]^-$	563.3226, 502.2917, 311.1677, 265.1462
135	Paritriside C[a]	$C_{41}H_{64}O_{12}$	771.4295	771.4313	2.33	29.59	$[M+Na]^+$	771.4313, 609.3352, 441.3701, 423.3627, 405.3471
136	Aspacochinoside O[a]	$C_{33}H_{52}O_{12}$	641.3532	641.3506	−4.06	29.95	$[M+H]^+$	641.3506, 479.2936, 317.1825, 301.0718, 279.2312
137	Coronaric acid	$C_{18}H_{32}O_3$	295.2279	295.2259	−6.78	30.31	$[M-H]^-$	295.2259, 265.1462
138	Δ^4-Pregnen-20β-ol-3-one glucoside[a]	$C_{27}H_{42}O_7$	479.3003	479.2981	−4.59	30.40	$[M+H]^+$	479.2981, 318.2985, 301.0789, 281.2932
139	Oleanonic acid	$C_{31}H_{50}O_3$	453.3374	453.3346	−6.18	31.03	$[M-H]^-$	453.3316, 325.1842, 285.1703
140	Linolenic acid	$C_{18}H_{32}O_2$	277.2173	277.2166	−2.53	32.56	$[M-H]^-$	277.2166, 251.1630, 99.9244
141	Oleanolic acid[b]	$C_{30}H_{48}O_3$	455.3531	455.3548	3.74	32.65	$[M-H]^-$	455.3548, 325.1842, 271.2271
142	Methyl 9-hexadecenoate	$C_{17}H_{32}O_2$	269.2475	269.2500	9.29	32.96	$[M+H]^+$	269.2500, 255.2637, 184.0740

[a]Be first found in *Sophora*.
[b]Compared with reference substance.

819.4528, 617.4071, 441.3744, 423.3627, 405.3513 corresponded to $[M+Na]^+$, $[M+Na-146]^+$, $[M+H-146-162-H_2O]^+$, $[M+H-146-162-176-H_2O]^+$, $[M+H-146-162-176-2H_2O]^+$, $[M+H-146-162-176-3H_2O]^+$, according to the above rules of saponins, there are three hydroxyl groups in the parent nucleus. Combined with Scifinder database and related literature,[22] it was speculated that the compound may be azukisaponin V.

Conclusion

In this experiment, the UPLC-Q-TOF-MS/MS method was used to quickly characterize the chemical components of *Sophorae Fructus* in positive and negative ion modes. The cracking rules of main flavone glycosides and saponins, which were preliminarily discussed, were helpful to improve the structural analysis efficiency and provide reference for the rapid screening and identification of flavonoids and saponins. From the data presented, 142 compounds were analyzed and inferred, including 67 flavonoids, 39 saponins, 18 organic acids, 10 amino acids and sugars, 2 phenylpropanes, 3 fatty acids, and 3 other types. A total of 43 components were first reported from the genus *Sophora*. This work will be helpful for the further study of pharmacodynamic material basis and quality evaluation of *Sophorae Fructus*.

Through the application of LC-MS technology, the repeated identification of known compounds by traditional separation and purification methods is avoided, which is conducive to saving resources, increasing the discovery probability of new compounds, and effectively improving work efficiency. It provides ideas and methods for the basic research and new drug development of traditional Chinese medicine and other complex substrates.

Supporting Information

The chemical structures of the 32 reference substances can be seen in the Supporting Information (**►Fig. S1 [online only]**).

Conflicts of Interests

None declared.

References

1 National Pharmacopoeia Commission of the PRC. Pharmacopoeia of the People's Republic of China. Part I. Beijing: Chemical Industry Press; 2020;247

2 Feng XL, Hu JY, Xiao J, Wang XL, Duan DZ. Separation and structural identification of ginsenosides from *Fructus Sophorae* in Qinling Mountains. [In Chinese].J Baoji Univ Arts & Sci: Nat Sci Ed 2021;41(01):48–51

3 Sun GD, Huo JH, Pan WJ, Wang WM. Identification and characterization of major chemical compounds in *Sophorae Fructus* based on UPLC-Q-TOF/MS. Chin Tradit Herbal Drugs 2019;50(16):3774–3783

4 Zhang QH. Studies on the Chemical Constituents and Biological Activities of *Fructus Sophorae* and *Nelumbbinis Plumula* [In Chinese]. Changchun: Jilin University; 2018

5 Shao J, Xiong YX, Li YG, et al. Analysis of chemical components of *Sarcandrae* Herba based on UPLC-Q-TOF/MS. Yaowu Fenxi Zazhi 2021;41(06):1054–1063

6 Su LL, Ding XY, Hao M, et al. Study on the potential chemical markers for the discrimination between raw and processed *Schisandrae Chinensis Fructus* using UPLC-Q-TOF/MS coupled with multivariate statistical analyses. J Herb Med 2020; 19:100311

7 Fuzzati N, Pace R, Papeo G, et al. Identification of soyasaponins by liquid chromatography–thermospray mass spectrometry. J Chromatogr A 1997;777:233–238

8 Zhao J, Qin ZX, Peng B, et al. Fragmentation pathway of ginsenosides in *Panaxnotoginseng* using electrospray ionization-quadrupole/time-of-flight mass spectrometer. J Chin Mass Spectrom Soc 2017;38(01):97–108

9 Cuyckens F, Claeys M. Determination of the glycosylation site in flavonoid mono-O-glycosides by collision-induced dissociation of electrospray-generated deprotonated and sodiated molecules. J Mass Spectrom 2005;40(03):364–372

10 Wang Y, Li XB, Li JJ, et al. Fragmentation pathway comparison of 5,6,7,4′-tetrahydroxy-flavone and 5,6,7,4′-tetramethoxy-flavone by high resolution electrospray ionization tandem mass spectroscopy. J Chin Mass Spectrom Soc 2016;37(05):385–392

11 Ferreres F, Taveira M, Pereira DM, Valentão P, Andrade PB. Tomato (Lycopersicon esculentum) seeds: new flavonols and cytotoxic effect. J Agric Food Chem 2010;58(05):2854–2861

12 Zhou DY, Xu Q, Xue XY, et al. Determination of flavonoid glycosides in *Fructus aurantii* by high performance liquid chromatography-electrospray ionization mass spectrometry. Chin J Anal Chem 2006;34:31–35

13 Zhu JB, Li DF, Xiao HB, et al. HPLC/ESI-MS analysis of flavonoids in *Epimedium* extract. [In Chinese]. Dalian Gongye Daxue Xuebao 2009;28(05):321–325

14 Liu JQ, Shu JC, Zhang R, et al. Study on the four flavone-C-glycosides by electrospray ionization tandem mass spectrometry. Chin J Exp Tradit Med Form 2013;19(08):72–76

15 Kachlicki P, Marczak L, Kerhoas L, Einhorn J, Stobiecki M. Profiling isoflavone conjugates in root extracts of lupine species with LC/ESI/MSn systems. J Mass Spectrom 2005;40(08):1088–1103

16 Zhou GF, Lv GY. Study on eight flavone C-glycosides in *Dendrobium officinale* leaves and their fragmentation pattern by HPLC-DAD-ESI-MS^n. Chung Kuo Yao Hsueh Tsa Chih 2012;47 (01):13–19

17 AlGamdi N, Mullen W, Crozier A. Tea prepared from Anastatica hirerochuntica seeds contains a diversity of antioxidant flavonoids, chlorogenic acids and phenolic compounds. Phytochemistry 2011;72(2–3):248–254

18 Li F, Ding LS, Wang MK. Study on the fragmentation pathway of raddeanin A by ESI—MS/MS. J Chin Mass Spectrom Soci 2008;29 (02):76–83

19 Yuan E, Deng MZ, Zhou LF, et al. Identification on triterpenoid saponins of *Clematis* species by UPLC-LTQ-Orbitrap MS combined with molecular network. Chin Tradit Herbal Drugs 2020;24(51): 6157–6167

20 Zhao Y, Wang L, Bao Y, Li C. A sensitive method for the detection and quantification of ginkgo flavonols from plasma. Rapid Commun Mass Spectrom 2007;21(06):971–981

21 Zhou Y, Li R, Wang XM, et al. HPLC/MS^n analysis of triterpenoid saponins from *Anemone rupestris ssp. gelida*. Youji Huaxue 2006; 26(01):116–119

22 Zhang B, Li MY, Luo XM, Wang XB, Wu T. Analysis of the chemical components of Qixianqingming granules and their metabolites in rats by UPLC-ESI-Q-TOF-MS. J Mass Spectrom 2020;55(01):e4484